Survey and genetic diversity of wild Brassica oleracea L. germplasm on the northern coast of Spain

Lorenzo Maggioni

l.maggioni@cgiar.org

European Cooperative Programme for Plant Genetic Resources (ECPGR), Alliance of Bioversity International and CIAT

Therése Bengtsson

Swedish Univ. of Agricultural Sciences (SLU)

Gert Bundgaard Poulsen

Danish Seedsavers

Roland von Bothmer

Swedish Univ. of Agricultural Sciences (SLU)

Research Article

Keywords: Brassica oleracea, crop wild relatives, domestication, ferality, genetic diversity

Posted Date: April 10th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4217963/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Genetic Resources and Crop Evolution on July 16th, 2024. See the published version at https://doi.org/10.1007/s10722-024-02070-8.

Abstract

This study follows from an expedition undertaken in 2011 to collect and monitor populations of *Brassica oleracea* var. *oleracea* growing on the Atlantic cliffs in the north of Spain. The genetic diversity and structure of 17 of these populations were analysed with Single Nucleotide Polymorphisms (SNP) revealed through Diversity Arrays Technology (DArT). These were compared with eight similar populations from the Atlantic coasts of northern Europe, with a representation of six Mediterranean wild taxa belonging to the *Brassica oleracea* complex and with six leafy kale landraces from Italy and Spain. The analysis reveals that the largest *B. oleracea* genetic diversity is present in the Asturias region. Another evidence is the presence of a north-south gradient of diversity along the Atlantic coasts, and the distinction between two types of accessions, one confined to northwest Spain ('AsCan' type) and another comprising most northern populations and the cultivated leafy kales, with an intermediate area of admixed types. The genetic distance among all the *B. oleracea* var. *oleracea* populations and between these and the leafy kales is minimal. This brings further weight to the interpretation of the wild-growing Atlantic populations of *B. oleracea* as a feral group. A specific pivotal position is assigned to the population of Laredo, Cantabria, which is genetically very close to 63% of the other accessions. In this study, *Brassica incana* is genetically the closest wild taxon to cultivated leafy kales and feral *B. oleracea*, and thus is indicated as a possible source of domestication of *B. oleracea*.

Introduction

The GRIN Taxonomy database (Germplasm Resource Information Network; USDA, Agricultural Research Service, National Plant Germplasm System 2023) indicates *Brassica oleracea* L. var. *oleracea* as the wild form of the cultivated cabbages. This taxon (with the alternative ranking of *B. oleracea* L. subsp. *oleracea*) is described by Snogerup et al. (1990) as a suffrutescent perennial with lignified stems, usually glabrous undivided leaves with a greyish, waxy surface usually growing on limestone and chalk cliffs along the Atlantic coasts of the British Isles and Ireland, northwest France, northern Spain and Helgoland. Subsequently, populations have been reported also from the Portuguese islands of Berlengas, the German island Rügen, and one location in Denmark (Rødvig) (Maggioni et al. 2020 and references therein). Recent evidence shows that this taxon has been observed mainly after 1990 also on the Dutch coast (NDFF & FLORON, 2023) and in Ireland (National Biodiversity Data Centre, 2023). The significance of these populations rests on at least two elements. On one hand, they belong to the same gene pool of the commercially important group of *B. oleracea* vegetable crops (cole crops), and thus their diversity can be easily used for breeding. At the same time, they constitute an object of investigation related to the domestication history of *B. oleracea*. This taxon has been variably interpreted either as the closest wild relative of the domesticated crop (Song et al.1990; Hodgkin 1995; Drenckhahn 2017), or rather as a group of feral populations, escaped from cultivations and re-naturalized (Mitchell 1976; Maggioni 2015; Mittel et al. 2020; Mabry et al. 2021). In a previous paper, we reported about a survey of French populations and provided an analysis of their genetic diversity and structure, based on Amplified fragment length polymorphism (AFLP) markers (Maggioni et al. 2020). Besides providing an updated mapping of the existing populations and their sizes, our conclusions showed a low level of genetic differentiation of these populations among each other, also characterized by a similar or lower level of genetic diversity (mean Nei's genetic diversity value H = 0.247) compared to that of local cultivated varieties (H value between 0.244 and 0.355). The limited genetic structure and the absence of a domestication bottleneck in the crop accessions, which were even more diverse than the most diverse wild-growing populations, led us us to exclude these populations as the likely source of original domestication events and rather interpret them as feral populations. Twelve populations from the UK and eight from northern Spain were recently analysed by Mittel et al. (2021) with double digest restriction-site associated DNA sequencing, finding little genetic differentiation, and no indication of isolation-by-distance from northern Scotland to Spain, with some signals of local adaptation to different climates. These authors concluded that such a pattern is not compatible with a natural range expansion followed by isolation and thus they supported the feral nature of these populations, considering them as

originating from escapes from different cultivars. Another recent work of Mabry et al. (2021), was aiming to understand the evolutionary history of wild, domesticated and feral *B. oleracea*. In this study, several accessions of crop brassicas, three putative wild *B. oleracea* accessions and other wild species from the Mediterranean area were tested using RNA-sequencing data. The conclusions supported an origin of domestication in the Eastern Mediterranean region, pointing to *Brassica cretica* Lam. as the closest living relative of the *B. oleracea* crops. All the three putative wild *B. oleracea* genebank accessions (from Helgoland, Germany; Rødvig, Denmark; and a third one obtained from Canada with unknown origin) shared their entire ancestry with cultivated types and thus they were classified as certainly feral. It seems that a consensus is emerging regarding the ferality of the Atlantic populations of *B. oleracea* and feralisation as a widespread phenomenon (Mabry et al. 2021, Mabry et al. 2023).

The corresponding search for the closest wild relative of the *B. oleracea* crops is pointing at *B. cretica* as the main suspect, as a result of molecular analysis and historical, literary and linguistic evidence (Maggioni et al. 2017; Mabry et al. 2021). At the same time, conclusive evidence regarding both the ferality of all *B. oleracea* populations and the identification of its closest wild relative will only be reached when a comprehensive panel of accessions representing the entire geographic range of the wild and putative wild species will be investigated with modern molecular markers techniques.

In 2011, we organized a collecting mission targeting *B. oleracea* var. *oleracea* growing in northern Spain, within the autonomous regions of the Basque Country, Cantabria and Asturias. We collected fresh seed samples from many individuals from most populations in that area, to verify how the pattern of genetic diversity and structure of these populations would be comparable to populations of the same taxon from other areas, particularly France, where we also extensively collected with a similar approach (Maggioni et al. 2020). The expedition to Spain has also been a case study of a collecting mission in compliance with the Convention of Biological Diversity, the FAO International Treaty on Plant Genetic Resources for Food and Agriculture and the national laws and their interpretations. The results of this experience from the legal/administrative point of view were published by Maggioni et al. (2015). Our expedition enabled us to survey the geographic distribution and size of the Spanish populations of *B. oleracea* growing in the wild, thus complementing and updating a previous survey carried out in 2002 by Gómez-Campo et al. (2005). It was useful to monitor trends in the demography of these populations and to formulate recommendations about conservation and use. In this study, we analysed with DArTseq (1.0) a panel of 25 putative wild *B. oleracea* populations, including 17 from Spain, five from France, one from the UK, one from Helgoland (Germany) and one from Rødvig (Denmark). Additionally, we added accessions from other Mediterranean wild species, i.e. *Brassica incana* Ten. (7), *Brassica montana* Pourr. (3), *Brassica cretica* Lam. (2), *Brassica rupestris* Rafin. (2), *Brassica drepanensis* (Caruel) Damanti (1) and *Brassica macrocarpa* Guss. (1). The panel was completed by three landraces of leafy kale (*Brassica oleracea* L. var. *viridis* L.) from Spain and three from Italy, for a total of 47 accessions investigated. We aimed to analyse the genetic diversity between and within populations and the structure of the available Spanish populations and put these in comparison with French populations, selected to cover their entire distribution range and the different levels of diversity previously determined with AFLPs (Maggioni et al. 2020). Other Atlantic and Mediterranean accessions were added to widen the context and improve knowledge of the genetic relationship between taxa of the *Brassica oleracea* complex, including also some leafy kale landraces, i.e. the crop type that is considered closest to the first domesticated forms. Thus, we expected to add new elements to the discussion about the hypothesized feralisation of *B. oleracea* (and other species) and the origin and closest wild relative of cole crops.

Differently from our previous studies with AFLPs, we have used the Diversity Array Technology (DArT) for genotyping. This technology enables the detection of a much larger number of Single Nucleotide Polymorphisms (SNPs), ranging, in our case, between 30 and 100 times more markers than in our previous studies. This aspect provides more reliable genetic distance, genetic diversity and genetic structure data.

Some specific questions that we aimed to address were: 1) What is the level of genetic diversity and genetic structure of the Spanish putative wild *B. oleracea* populations and how do they compare with the French populations? 2) What is the level of genetic distance/similarity between putatively wild *B. oleracea* populations from the entire geographic range, as well as between Atlantic and Mediterranean populations? 3) How do the genetic parameters of the wild and putative wild populations compare with landraces from Spain and Italy? 4) Is inter-crossing and further introgression detectable between *B. oleracea* growing in the wild and landraces in nearby home gardens? 5) Can the genetic data and clustering pattern of our analysed panel give any insight regarding the ferality of the populations and the closest wild relative of the *B. oleracea* crops?

Materials and Methods

Plant material and sampling

The accessions analysed in this study were obtained either from personal collecting missions or from genebanks (Tables 1 and S1, Online Resource 1 for details). We use the terms ‘accession’ and ‘population’ almost interchangeably, assuming that each genebank accession or collected sample represents a given population living in the wild. In the cultivated crops the six samples collected from different home gardens generated ‘accessions’ representing individual landrace populations. For the collecting mission organized in July 2011 in northern Spain, we tracked the populations mainly based on descriptions of locations in Snogerup et al. (1990)d mez-Campo et al. (2005). Consent from local authorities was obtained for the collecting missions, where necessary. Permission was requested from the owners of the home gardens where we collected the landraces, which were in the same distribution area as the populations growing in the wild. Seeds were collected keeping the offspring of each mother plant separate. Duplicates of the collected accessions were deposited in their country of origin, respectively at Centro Nacional de Recursos Fitogeneticos (CRF) of the National Institute for Agricultural Research (INIA), Madrid, Spain, the National Institute for Agricultural Research (INRA), Ploudaniel, France, Orto Botanico di Roma and/or University of Catania, Italy and the Greek Genebank, Thessaloniki, Greece. All samples are also maintained as a separate collection at the Swedish University of Agricultural Sciences, Alnarp, Sweden,

Table 1

Analysed populations and genetic diversity data (NI = number of individuals analysed). Means with different superscripts are significantly different at p < .01. Accessions preceded by * were removed from further analysis due to their dubious taxonomic status and they are excluded from the sub-total counts.

Populations/Accessions	Accession code	NI	Mean genetic diversity Shannon Index		% polymorphic loci
			Mean	SE	
B. oleracea* var. *oleracea					
Denmark					
Rødvig	Rødvig_w	20	0.058^{d}	0.002	14.65
Germany					
Helgoland	Helgo_w	19	0.102^{s}	0.003	25.97
France					
Granville – Manche	Granw_w	19	0.094^{pq}	0.003	24.27
Mers-les-Bains – Somme	MLbain_w	20	0.080^{kl}	0.003	21.51
Mortagne – Charente Maritime	Mortag_w	19	0.083^{m}	0.003	
Petites Dalles – Seine-Maritime	Petdal_w	20	0.101^{rs}	0.003	24.97
Le Tréport – Seine Martime	Trepo_w	20	0.095^{q}	0.003	24.10
Sub-total France		*98*			
Spain					
Cabo Peñas – Asturias	Cpena_w	18	0.095^{q}	0.003	24.79
Cuchia – Cantabria	Cuchia_w	20	0.074^{i}	0.003	18.70
Cudillero – Asturias	Cudil_w	19	0.093^{p}	0.003	23.05
El Musel – Asturias	Elmus_w	20	0.083^{m}	0.003	21.35
El Pedrero – Cantabria	Elped_w	18	0.089^{o}	0.003	22.14
Gaztelugatxe – Basque	Gazte_w	19	0.080^{kl}	0.003	20.77
Getaria road – Basque	Getar_w	20	0.090^{o}	0.003	23.27
Getaria west – Basque	GetarW_w	5	0.088^{no}	0.003	17.61
Laredo – Cantabria	Laredo_w	20	0.102^{s}	0.002	26.70
Ondarroa – Basque	Ondar_w	19	0.069^{g}	0.003	16.98
Oyambre – Cantabria	Oyam_w	20	0.070^{g}	0.003	16.37
Pendueles – Asturias	Pendu_w	20	0.084^{m}	0.003	21.51

Populations/Accessions	Accession code	NI	Mean genetic diversity Shannon Index		% polymorphic loci
			Mean	SE	
B. oleracea* var. *oleracea					
Tazones – Asturias	Taz_w	18	0.086^{n}	0.003	24.55
Torimbia – Asturias	Torim_w	20	0.085^{mn}	0.003	22.89
Torres – Asturias	Torr_w	20	0.100^{r}	0.003	23.84
Urgull – Basque	Urgull_w	17	0.129^{w}	0.003	42.78
Xago – Asturias	Xago_w	20	0.092^{p}	0.003	22.55
Sub-total Spain		*313*			
Wales, UK					
Tenby cliffs	Tenby_w	20	0.075^{i}	0.03	19.57
Sub-total B. oleracea *var.* oleracea		*470*			
Brassica oleracea var. viridis (leafy kales)					
Tazones – El Pison – Asturias, ESP	Taz_EP_lr	17	0.077^{j}	0.003	20.46
Tazones Les Mestes – Asturias, ESP	Taz_LM_lr	17	0.067^{f}	0.002	18.13
Torimbia CULT – Asturias, ESP	Torim_lr	10	0.112^{u}	0.003	29.59
De Baudo's neighbour – Sicily, ITA	Debau_lr	17	0.122^{v}	0.003	33.35
Latassa – Calabria, ITA	Latas_lr	20	0.075^{i}	0.003	18.96
Second garden-Caltavuturo – Sicily, ITA	garCal_lr	20	0.063^{e}	0.002	17.51
Sub-total Brassica oleracea *var.* viridis		*101*			
Brassica cretica					
Crete - GRE	BcreCr_w	19	0.043^{a}	0.001	16.25
Limni – Corfu - GRE	BcreLi_w	18	0.056^{c}	0.002	16.64
Sub-total B. cretica		37			
Brassica drepanensis					
Erice – Sicily - ITA	BdreEr_w	19	0.093^{p}	0.003	26.15
Brassica incana					
*Angelokastros – Corfu - GRE	BincAn_w	-			

Populations/Accessions	Accession code	NI	Mean genetic diversity Shannon Index		% polymorphic loci
			Mean	SE	
B. oleracea* var. *oleracea					
Capri – Campania - ITA	BincCp_w	19	0.106^{t}	0.003	33.84
Castelmola – Sicily - ITA	BincCa_w	20	0.079^{k}	0.002	28.29
Ischia – Campania - ITA	BincIs_w	20	0.080^{l}	0.002	29.35
Monte Leano – Lazio - ITA	BincML_w	18	0.049^{b}	0.002	14.07
Sorrento – Campania - ITA	BincSo_w	17	0.083^{m}	0.003	24.91
Tindari – Sicily - ITA	BincTi_w	14	0.087^{n}	0.003	24.91
Sub-total B. incana		*108*			
B. macrocarpa					
Favignana – Sicily - ITA	BmacFa_w	20	0.078^{j}	0.002	25.42
Brassica montana					
*San Marino - SMR	BMoSM_w	-			
*Savona – Liguria - ITA	BMoSav_w	-			
Stazzema – Tuscany - ITA	BMoSta_w	18	0.073^{h}	0.002	21.33
Sub-total B. montana		*18*			
Brassica rupestris					
Pazzano – Calabria - ITA	BrupPA_w	20	0.049^{b}	0.002	14.86
Taureana – Calabria - ITA	BrupTa_w	4	0.263^{x}	0.004	46.55
Sub-total B. rupestris		24			

From each collected population or genebank accession, the target was to analyse 20 individuals. Regarding *B. oleracea* var. *oleracea*, we analysed 17 populations from Spain, five from France obtained in a collecting mission carried out in 2012, one from Rødvig (Denmark) and one genebank accession each from Helgoland (Germany) and Tenby (Wales, UK). We added to the analysis three leafy kale landraces from Spain and three from south Italy, all collected by us. The panel was completed by a few samples representing Mediterranean wild species: seven accessions of *B. incana*, three of *B. montana*, two of *B. cretica*, two of *B. rupestris*, one of *B. drepanensis* and one of *B. macrocarpa*. All collected populations and genebank accessions are represented on the maps in Fig. 1A and 1B.

DArTseqLD genotyping

Seeds were germinated in soil trays under greenhouse conditions at NordGen, Alnarp, Sweden, and leaf material was harvested at the 4–6 leaf stage. The leaf samples were powdered in a mixer mill (Merck Retsch mm 300) with a steel ball and DNA extraction was carried out with the same procedure described in Maggioni et al. (2020).

DNA samples were sent to Diversity Arrays Technology (DArT) Pty Ltd., Bruce, Australia, for genotyping utilizing DArTseqLD™ technology (Jaccoud et al. 2001, Sansaloni et al. 2011). Subsequently, the obtained data underwent filtration using the 'dartR' pipeline (Gruber et al. 2018) in R v. 4.1.1 (R core team 2021). The data was divided into five different subsets for comparison within 1) all wild and cultivated *B. oleracea*; 2) all wild *B. oleracea*; 3) landraces from Italy and Spain; 4) wild-growing *B. oleracea* of origin in France and Spain; and 5) all populations/accessions analysed. Each subset was filtered based on: locus call rate, set at ≥ 80%; random removal of all but one of the multiple SNPs in the same fragment to mitigate potential linkage effects (only for STRUCTURE analyses); individual call rate, set at ≥ 60%; removal of monomorphic loci; and minor allele frequency (MAF) ≤ 1%. Thus, comparisons of results from the following analyses cannot be made across subsets since the analysis for each subset is based on both different markers and numbers of markers. The 'dartR' functions gl2genalex and gl2faststructure were used to generate files for further diversity analysis using GenAlEx 6.5 (Peakall and Smouse, 2012) and population structure analysis using STRUCTURE v.2.3.4 (Pritchard et al. 2000), respectively.

Diversity analysis

Diversity analysis was performed for each of the five subsets. As an indicator of genetic diversity, we measured the Shannon's Information Index $I = -1 * Sum\left(pi * Ln(\ pi\ )\right)$ .Shannon's gene diversity was averaged over loci and individuals for each population/accession to calculate the mean genetic diversity per population/accession (I). Student's t-test was performed to determine the level of significance of the differences among the means. Distance matrices were generated from Nei's genetic distances (Nei 1972) between pairs of populations and between pairs of individuals. Based on these distance matrices, principal coordinate analysis (PCoA) was used to visualize the differences between populations or between individuals. Total genetic variation was partitioned into two levels by an analysis of molecular variance (AMOVA) between and within populations/accessions. The significance was tested using 1,000 permutations. PhiPT values are calculated as the proportion of the variance among populations, relative to the total variance. The fixation index (Fst) was used as measure of population differentiation due to genetic structure: $Fst = (Ht - MeanHe)/Ht.\ Ht = TotalExpectedHeterozygosity = 1 - Sumtpi^2$ where tpi is the frequency of the ith allele for the total and $Sumtpi^2$ is the sum of the squared total allele frequencies. $He = ExpectedHeterozygosity = 1 - Sumpi^2$. All tests were computed in GenAlEx 6.5.

STRUCTURE analysis

The Bayesian model-based clustering approach implemented in STRUCTURE v.2.3.4 (Pritchard et al. 2000) was used to explore genetic sub-population and admixture patterns within the five subsets. The analysis consisted of ten independent runs with a burn-in period of 50,000 and 100,000 Markov Chain Monte Carlo (MCMC) replications, ranging from K = 1 to K = 10. No prior information on clusters was given. The most likely number of clusters was determined by calculating the average likelihood values using the delta K method (Evanno et al. 2005) in Structure Harvester (Earl and von Holdt 2012). The results were visualized with bar plots created using 'ggplot2' (Wickham 2016, Villanueva and Chen 2019) in R.

Results

DArTseqLD

The comprehensive dataset provided by DArT Pty Ltd. comprised 864 genotypes, featuring 15,833 codominant binary SNP markers and 33% missing data. The final number of SNP markers and genotypes for each subset following filtering are provided in Table S2 (Online Resource 2).

Monitoring of northern Spain populations

We collected wild populations from 19 locations (six in the Basque Country, four in Cantabria and nine in the Asturias) and leafy kales from three home gardens in Asturias, one of which was placed few meters away from the site of the wild population of Torimbia. All the wild populations were found in the locations indicated by Gómez-Campo et al. (2005), except one new site, which was found at El Musel, near Gijon. A few differences were observed in eight locations regarding the status of the populations, compared to the notes in Gomez-Campo et al. (2005). In Cudillero, Cabo de Peñas, Cabo Torres, Torimbia and Llanes, the populations were larger and, except in Llanes, more accessible than ten years before. In Tazones, Llanes and Pendueles, they were no longer easily accessible, and at El Pedrero, the population was much smaller (Online Resource 3). At this time of the year, a large part of the seeds had been already shed, but it was still possible to find remaining siliquas with some seeds in most plants.

Identity of the accessions

After a first data analysis of all samples together, three different approaches offered by the construction of a NJ Tree (Figure S1, Online Resource 2), Principal Component Analysis and STRUCTURE (not shown), there were three accessions that evidently did not cluster within the expected group. One of these was *B. incana* from Angelokastros, Corfu, which clustered completely separate from other *B. incana* populations and actually overlapped with *B. cretica* from Corfu. Two samples of *B. montana*, from San Marino and from Savona, clustered in the group of *B. oleracea* differently from *B. montana* from Stazzema. Only the latter formed a well-identified autonomous cluster, both in STRUCTURE and PCoA, and clustered together with the Mediterranean species *B. cretica*, *B. drepanensis*, *B. macrocarpa* and *B. rupestris* in the NJ Tree. Because of this doubtful taxonomic attribution, the three above-mentioned accessions were removed and the dataset re-filtered as described in material and methods before further analysis.

Relationship between species

Samples were analysed by species, where the cultivated *B. oleracea* group separately from the wild-growing populations. Each species forms a well distinct cluster (Fig. 2A) with the partial exception of *B. rupestris*. The cultivated types almost completely overlap with the wild *B. oleracea* and partly with *B. incana*, indicating a very close relationship between these three groups. *B. rupestris* and *B. drepanensis* are very close to each other and well separated from all other species, except four individuals of *B. rupestris* from Taureana, which indicate probable intercrossing with cultivated crops. *Brassica cretica* from Crete and Corfu form two very distinct clusters.

Looking at the Shannon index of genetic diversity (Table 2), the values are not fully comparable, since each group has a different number of individuals and populations. The case of *B. rupestris* (highest value with 0.207) can be explained by the fact that it is a combination of a truly *B. rupestris* population from Pazzano and a population from Taureana, the latter likely introgressed by cultivated crops. The two populations of *B. cretica* have a rather high value of 0.149, but when considered separately (Table 1) their respective values are only 0.043 and 0.056.

Table 2
Genetic diversity indexes for accessions grouped by taxon. NP/A = number of populations/accessions; NI = Number of individuals. Means with different superscripts are significantly different at p < .01.

Taxon	NP/A	NI	Mean genetic diversity Shannon Index		% polymorphic loci
			Mean	SE	
B. montana	1	18	0.073[a]	0.002	21.33
B. macrocarpa	1	20	0.078[b]	0.002	25.42
B. drepanensis	1	19	0.093[c]	0.003	26.15
B. oleracea var. *viridis* (leafy kales)	6	101	0.119[d]	0.003	54.54
B. oleracea var. *oleracea*	25	470	0.128[e]	0.003	73.89
B. incana	6	108	0.142[f]	0.003	62.17
B. cretica	2	37	0.149[g]	0.004	32.75
B. rupestris	2	24	0.207[h]	0.003	59.26
Total (Mean)	44	797	0.124	0.001	44.44
Fst 0.579					

The Analysis of Molecular Variance (AMOVA) shows that the variation is equally distributed among (48%) and within (52%) populations (Table 3). This pattern is quantified in terms of genetic differentiation by the relatively high overall mean Fst value = 0.579. However, looking at the pairwise values of each population (Table 4), it appears that genetic differentiation is minimal (Fst = 0.012) between wild *oleracea* and cultivated leafy kales, indicating that these two groups are genetically almost identical. The closest wild species to these two groups is *B. incana* (Fst = 0.049 vs. wild *oleracea*; Fst = 0.060 vs. leafy kales). In comparison, *B. cretica* is more differentiated from *B. oleracea* (Fst = 0.172 vs. wild *oleracea*; Fst = 0.206 vs. leafy kales). Nei's genetic identities offer a similar pattern, with wild *oleracea* virtually identical to the leafy kales (0.995) and *B. incana* as the closest wild relative (0.968), followed by *B. cretica* (0.868) at a certain distance (Table S3, Online Resource 2).

Table 3
Analysis of Molecular Variance (AMOVA) (NP/A = number of populations/accessions, NI = number of individuals, Fst = coefficient of gene differentiation, SE = Standard Error of Fst, P = probability)

Populations/Accessions	NP/A	NI	Fst	SE	PhiPT	P	Variance within populations	Variance among populations
7 species groups + leafy kales group	8	797	0.579	0.004	0.482	0.001	52%	48%
7 species groups (no cultivated)	7	696	0.578	0.004	0.548	0.001	45%	55%
Wild oleracea vs. cultivated oleracea	2	571	0.024	0.001	0.054	0.001	95%	5%
Wild oleracea from France vs. Spain	2	411	0.025	0.001	0.056	0.001	94%	6%
Wild oleracea by France and Spain sub-regions	7	411	0.124	0.002	0.109	0.001	89%	11%
Wild oleracea by country	5	470	0.127	0.003	0.117	0.001	88%	12%
Landraces from Spain vs. Italy	2	101	0.046	0.001	0.127	0.001	87%	13%
Six landraces from Spain and Italy	6	101	0.186	0.003	0.293	0.001	71%	29%

Table 4
Pairwise population Fst values

B. cretica	B. drepanensis	B. incana	B. macrocarpa	B. montana	B. oleracea	B. rupestris	Leafy kales	
0.000								B. cretica
0.515	0.000							B. drepanensis
0.214	0.370	0.000						B. incana
0.503	0.538	0.319	0.000					B. macrocarpa
0.462	0.595	0.250	0.583	0.000				B. montana
0.172	0.352	0.049	0.303	0.192	0.000			B. oleracea
0.342	0.320	0.225	0.377	0.374	0.222	0.000		B. rupestris
0.206	0.415	0.060	0.344	0.241	0.012	0.248	0.000	Leafy kales

A STRUCTURE analysis was run for all accessions and the results visualized in a bar plot grouped by species (Fig. 2B). DeltaK indicated that the accessions can be grouped into two groups (K = 2). One accession, LM18_02_04, the *B. cretica* accession from Corfu, was equally assigned to both groups. The distinction in two groups (K1 and K2) clearly separates all the cultivated accessions, together with all the Atlantic *oleracea*, from the Mediterranean wild species *B. montana, B. rupestris*, *B. drepanensis, B. macrocarpa* and *B. cretica*, with the exception of *B. incana*.

The only *B. montana* population analysed here ('Stazzema'), clusters together with the truly wild species, but also shows a certain percentage of 'domesticated' alleles (K1 type). The other two *B. montana* populations ('Savona' and 'San Marino') seemed to be heavily admixed with cultivated crops and were not included in this analysis.

In the case of *B. cretica*, the two populations from Crete and Corfu are identified as truly wild, but with a significant percentage of domesticated alleles, whereby one individual is considered an admixture.

The other wild species from South Italy (*B. macrocarpa*, *B. drepanensis* and *B. rupestris*) firmly plot among the truly wild.

Relationship between wild-growing and cultivated B. oleracea

The STRUCTURE results run for all wild-growing *B. oleracea* (n = 470) together with the cultivated *B. oleracea* (n = 101) distinguish two groups (Fig. 3B). All the cultivated accessions fall into the K2 group, while the wild-growing fall partly into the same group as the cultivated, and partly into a different group (K1), with various degrees of admixture. The 'purest' K1 populations are mainly originated from the western part of the northern coast of Spain, i.e. from Cantabria ('Cuchia', 'El Pedrero') and Asturias ('Cabo Peñas', 'Cudillero', 'El Musel', 'Torimbia', 'Torres' and 'Xago'). All the wild-growing northern populations from the UK, Denmark, Germany and France are indistinguishable from the cultivated ones. Considerable admixture is present in the geographical middle between these two extremes, i.e. the Basque Country ('Getaria west' and 'Urgull') and western France ('Mortagne'), but also in other populations from Asturias ('Pendueles') and Cantabria ('Laredo').

The corresponding PCoA (Fig. 3A) shows the majority of *B. oleracea* var. *oleracea* populations plotting on a wide area in the upper and left quadrants, while a few others (from northern Europe) are mixed with the cultivated populations in the bottom right quadrant. AMOVA indicates a variation of 5% among populations and 95% within populations (Table 3). In addition, the Nei genetic identity is very high (0.986), corresponding to a very low coefficient of gene differentiation between cultivated and wild (Fst = 0.024). Shannon index and allele polymorphism values are higher for the wild than for the cultivated accessions (Table S4, Online Resource 2).

Comparison between all wild-growing B. oleracea grouped by country of origin

Populations from each country cluster together and form a gradient of diversity with few overlaps between French and Spanish populations and between French and German ones (Fig. 4A).

Based on Shannon Index, the Spanish group is the most diverse (0.324), followed by the French (0.313), German (0.275), UK (0.193) and Danish (0.154) (Table S5, Online Resource 2). Molecular variance among populations is limited to 12% (Table 3). In fact, the level of Nei genetic distance among these groups is very low for all. The lowest and largest pairwise Nei distance is between French and Spanish (0.015) populations and Danish and British (0.096) populations, respectively. The second lowest Nei distance is between German and French (0.028) populations (Table S6, Online Resource 2). Neither the STRUCTURE analysis of the same grouping of populations shows any particular structuring associated with the country of origin (Fig. 4B). The main structuring remains between two groups (K1 and K2) that are mainly splitting into two categories the Spanish populations, while in the other countries K2 type prevails. The K2

pattern is in common with about half of the Spanish populations, mainly those from the central-eastern part of the northern coast of Spain.

Comparison between wild-growing B. oleracea from France and Spain

A comparison between French and Spanish populations of *B. oleracea* var. *oleracea* indicates a higher genetic diversity expressed in Spain. The differentiation between these two groups is very low, indicated by Fst = 0.025 (Table 3). To investigate the level of clustering of the French and Spanish populations grouped by region, the PCoA of Fig. 5 shows French populations from Charante-Maritime, Seine Maritime and Somme clustering together, reconfirming the pattern of Fig. 4A. These populations cover the bottom right quadrant. Populations from Manche cluster in the area overlapping with Basque Country and part of the Cantabrian and Asturian populations. The Asturian populations show the largest diversity (Table S7, Online Resource 2), also uniquely covering the left bottom quadrant of the PCoA. Pairwise Fst values of genetic differentiation range from 0.021 (between Asturias and Cantabria) and 0.124 (between Somme and Charente-Maritime) (Table S8, Online Resource 2). All Fst values within this analysis are low (< 0.15). The lowest values are between Spanish regions, ranging from 0.021 to 0.030. Differentiation between regional Spanish and French populations is a bit higher, ranging from 0.043 (Asturias vs. Seine-Maritime) to 0.076 (Cantabria vs. Somme). At the same time, differentiation between French populations is slightly higher, ranging from a minimum of 0.066 (Seine-Maritime vs. Somme) to a maximum of 0.124 (Charente-Maritime vs. Somme).

Nei's genetic distance shows exactly the same pattern, ranging from the lowest distance of 0.013 between Asturias and Cantabria, to the largest distance of 0.075 between Somme and Charente-Maritime (Table S9, Online Resource 2).

Comparison between landraces from Italy and Spain

Italian and Spanish landraces plot in separate clusters in the PCoA (Figure S2, Online Resource 2). However, variance among populations is low (13%) (Table 3), as well as the pairwise population Fst value (0.046) and the Nei genetic distance (0.033). The Shannon diversity is similar, 0.268 (Italy) vs. 0.244 (Spain) (Table S10). Running a PCoA by populations, it is noticeable that each landrace clusters into a separate group, except 'Tazones El Pison' (Taz_EP_lr) and 'Torimbia' (Torim_lr), which are fully overlapping (Fig. 6A). In fact, the gene differentiation among these six landraces is sufficiently high (0.186) (Table 3) to recognize that each landrace tends to form a separate cluster while there is no distinct separation between Italian and Spanish landraces as such.

A STRUCTURE analysis of the same group of six Italian and Spanish landraces identifies five groups (K = 5). Also in this case, each population shows a distinct predominant genetic character, except two landraces (one from Sicily ('De Baudo neighbour'), and the other from Asturias ('Torimbia'), falling into the same category K1 (Fig. 6B). This pattern confirms that landraces maintain their own identity, but various levels of admixtures between the genetically homogeneous groups identified as five clusters (K1 to K5), emerge more clearly here (in particular, two individuals of 'Torimbia' are an admixture between K1 and K4). While the PCoAs seemed to visually indicate a rather clear-cut separation between Italian and Spanish landraces, in the STRUCTURE analysis the separation is not equally evident, as indicated by landraces 'Torimbia' and 'De Baudo neighbour', plotting in the same category K1. The landrace with less admixtures and therefore with the purest identity is 'Latassa' from Calabria, Italy.

A PCoA in which Italian and Spanish landraces are highlighted, together with all the wild *oleracea* (Fig. 2A) shows that they do not position in separate clusters, when compared to the wild *oleracea*, and that both Spanish and Italian landraces have areas of admixture with the wild populations, confirming the same pattern of Fig. 3A.

Relationship between all accessions

When keeping each population separate, information on the individual populations can be obtained. Table S11 (Online Resource 2) shows the min-max range values of all the populations within each species/group for Shannon diversity. There is no clear pattern distinguishing the genetic diversity of various species, or the wild from the cultivated. Thus, in all groups there are populations with a very low level of diversity and others with a relatively high level (> 0.1), with the exceptions of those categories where only one or two populations were tested. Data from each population (Table 1) mostly show a rather low level of diversity (I < 0.1). Only few populations have I > 0.1. The top three are *B. rupestris* – 'Taureano', Calabria (0.263), *B. oleracea* var. *oleracea* – 'Urgull' – Basque Country (0.129) and landrace 'De Baudo's neighbour' – Sicily (0.122), also with the respective top three polymorphism levels (46.6%; 42.8% and 33.3%).

Table S12 (Online Resource 2) shows the closest relationships (two or more accessions) to each accession, based on the lowest values of pairwise genetic distance (also see the full matrix of Nei Unbiased genetic distances as Table S13 in Online Resource 1). 'Laredo' is the population that most often (63% of the cases) appears as genetically close to other accessions, which includes not only other Spanish *B. oleracea* var. *oleracea* populations, but also all the populations from France, 'Helgoland' (Germany), 'Rødvig' (Denmark) and 'Tenby' (Wales), and all the landraces from both Spain and Italy. Even in the case of *B. cretica* from Corfu, 'Laredo' is the second closest of all populations. This pattern makes 'Laredo' one exemplary 'pivot' population. The genetic distance between 'Laredo' and all the *B. oleracea* populations is never higher than 0.020 (ranging between 0.009 and 0.020). 'Laredo' is also the population with the lowest values of genetic distance from *B. incana* populations (0.037 from BincCp_w and 0.040 from BincSo_w), together with 'Helgoland' (0.037 and 0.039, respectively). The Shannon diversity index I of 'Laredo' is relatively high (0.102) compared to other accessions in this study, although not the highest.

'DeBaudo's neighbour' landrace is the second most represented population (21% of the cases) among those genetically closer to other accessions. It is the closest accession to both *B. cretica* populations and one of the closest to *B. rupestris* from Sicily and Calabria. It shows a short genetic distance from wild *oleracea* populations from north Spain, Helgoland, as well as from other landraces from Italy and Spain. It also has a high level of diversity (I = 0.122). The wild population in Tazones, Asturias, is genetically closer to the nearby (ca. 50m) 'Tazones' landrace Taz_EP_lr (0.013) than to any other wild population. This may confirm inter-crossing at this location. On the other hand, this pattern is not repeated in the nearby location of Torimbia, where Tor_w is genetically closer to other more distant wild populations ('Cudillero': 0.012) than to the landrace in Torimbia (0.022) which was growing at a distance of ca. 0.5 km and thus possibly not inter-crossing with it. The 'Torimbia' landrace is genetically closer to the 'Laredo' wild population (0.012), which is > 100 km away.

Looking at the genetically most distant accessions, *B. rupestris* from Palermo and *B. drepanensis* are the most distant to all the other accessions (in 98% and 93% of the cases, respectively). The most distant to *B. drepanensis* and *B. rupestris* from Palermo are the two *B. cretica* populations. Also for *B. macrocarpa*, *B. rupestris* from Palermo is genetically the most distant, even though they are geographically rather close.

The closest population to B. *rupestris*-'Palermo' is B. *rupestris*-'Tauerana' (0.332), but the opposite is not true, since *B. rupestris*-'Taureana' is closer to any other accession than to *B. rupestris*-'Palermo'. The Shannon Index and allele polymorphism of *B. rupestris*-'Taureana' (0.263 and 46.55%, respectively) are the highest values of the entire series.

Comparison between B. oleracea and B. incana

A specific comparison was made between *B. incana* (108 individuals from six populations) and *B. oleracea* (470 wild and 101 cultivated individuals together).

The PCoA and STRUCTURE analysis (K = 2) separates neatly these two groups (Figures S3A & S3B, Online Resource 2), with no overlaps between them. The genetic distance is 0.075. Variance within populations is 74% and among

populations is 26%. The Fst value of gene differentiation is 0.092. The same analyses were run keeping each accession separate and this enabled us to distinguish, within the cluster of *B. incana*, that the populations from Monte Leano, Lazio, and from Capri, Campania, are forming sub-clusters. The *oleracea* are almost all well grouped together, with only a few individuals slightly divergent (including landraces or wild populations).

Discussion

Monitoring of Spanish populations

Our exploration of the coasts of north Spain, guided by previous surveys from Gómez-Campo et al. (2005), visited only a fraction of the known populations. By comparing the data of Gómez-Campo with our observations, we noted that most of the populations were well thriving and actually expanding, with only a few cases of endangered populations showing reductions of individuals. These observations, together with the casual discovery of a previously unidentified site (El Musel), confirmed the impression that these populations tend to expand and colonize new sites.

Identity of the accessions

We interpreted that *B. incana* from Corfu, which was removed from the overall analysis, might have been heavily introgressed by *B. cretica*, which is also growing on the island, although it maintained the typical hairy pattern of *B. incana*. The samples of *B. montana* from San Marino and Savona that were also removed from the analysis, might actually be either escapes from gardens that have become feral or introgressed populations. The case of San Marino had been described with similar conclusions of ferality by Maggioni (2015), while risks of crossing with cultivated forms in Savona were mentioned before by Snogerup et al. (1990).

Another case of dubious identity is *B. rupestris*-'Taureana'. Considering that the genetically closest populations are two landraces (one from Spain, the other from Sicily), this overall pattern indicates that *B. rupestris*-'Taureana' is probably heavily introgressed with alleles from cultivated crops and thus does not represent a pure *B. rupestris* population. This pattern is confirmed by the position on the PCoA of the four individuals analysed of *B. rupestris*-'Tauerana', which is intermediate between *B. rupestris*-'Palermo' and the *oleracea* group. The highest level of genetic diversity in the entire series is probably the effect of the promiscuity expressed by the population in Taureana.

Comparison between all the B. oleracea populations and the issue of feralisation

Our collecting mission in the north of Spain aimed to analyse the genetic structure and diversity of populations of *B. oleracea* var. *oleracea* growing in the Basque Country, Cantabria and Asturias, extending our previous study on French populations, using AFLPs (Maggioni et al. 2020). In the latter study, the French populations showed a low level of genetic differentiation from each other. They were also indistinguishable at a molecular level from *B. oleracea* crops grown in the same area and were characterized by a similar or lower level of genetic diversity than the cultivated crops. We considered this pattern in line with the theory that *B. oleracea* var. *oleracea* is not a truly wild taxon but rather consists of re-naturalized (feral) populations. This theory was originally proposed by Mitchell (1976), further elaborated by Maggioni (2015), and recently received additional support from the studies of Mittel et al. (2021) and Mabry et al. (2021). Analysis of the Spanish and French populations together in this study, also compared with individual populations from Denmark, Germany and the UK, offers a wider perspective and shows a slightly different pattern than looking at the French populations alone. This pattern indicates a larger variability of the wild-growing populations and specifically the Spanish ones. It also shows an apparent geographical gradient, with country clusters orderly organised on the plot from south to north of Europe, with few overlaps between bordering countries. These results can be interpreted by admitting the existence of two main types within the *B. oleracea* panel, together with admixtures between the two. One type is represented by some of the wild-growing populations from Asturias and Cantabria

(tentatively called the 'AsCan' type). The other one includes all the remaining wild-growing populations and the landraces. Despite the clear differentiation in STRUCTURE between K1 and K2, and the geographical clustering of the populations shown by the PCoA, the genetic distance between all the *B. oleracea* populations remains very low, both considering country groupings (the largest distance is as low as 0.096, between Danish and British populations, while the closest populations are the French and Spanish with 0.015) and individual populations (the largest distance between populations of the 'AsCan' type and other accessions is in the range of maximum 0.035, Table S14). These observations, indicating a very low genetic distance between all the *B. oleracea* populations, are compatible with the assumption that all *B. oleracea* var. *oleracea* populations are feral populations derived from escaped crops. This pattern is particularly evident, looking at cultivated crops and populations located in part of Cantabria, Basque Country and northern Europe, since all these belong to the same type. Something different is present in Cantabria and Asturias ('AsCan' type), but not so different as to hypothesize the presence of a different wild taxon from which the crops might have been domesticated. Most likely, the differentiation is the result of an isolation effect in this western part of northern Spain. It is also possible that feralisation events are here older than in other locations and/or that less intercrossing events with modern crops have taken place in this area after feralisation. The recurrent presence of the accession 'Laredo' as the closest or second closest to any of the tested landraces (either from Spain or Italy), gives a remarkable indication that this population represents some type of pivotal link between the cultivated and the feralised populations. This pattern reinforces the idea that all the investigated landraces and North Atlantic 'wild' populations belong to the same primary gene pool. One hypothesis could suggest that 'Laredo' was one of the primary sites of feralisation, also giving rise to other populations that could have been spread elsewhere by wind or other means and adapted to other areas over time.

In order to defend the alternative scenario, whereby *B. oleracea* var. *oleracea* would truly be wild populations, it should be assumed that once upon a time, these populations were domesticated on the Atlantic coasts of Europe, perhaps around Laredo. These remained genetically indistinguishable from the crops and/or went through enough intercrossing with the crops to become indistinguishable across their distribution range, with only some slight differences maintained in the Asturias and Cantabria areas. This latter scenario is unlikely, based on the following elements: 1) The history of domestication based on literary and linguistic indications points at a northeastern Mediterranean origin, with the introduction of cultivation in the ancient Greek-speaking part of the Mediterranean (including southern Italy) and further domestication and radiation of many different *oleracea* crops on ancient Roman territory, following an east-west gradient of increased complexity of diversification (Maggioni et al. 2017); 2) Data on similar population sets obtained by Mittel et al. (2020) and Mabry et al. (2021) were also interpreted with the recognition of wild Atlantic populations as being feral; 3) Feralisation of *B. oleracea* is the most plausible type of event to explain recent occurrences of populations on the cliffs of Denmark, Germany, the Dutch dykes, California, New Zealand and other parts of the world where there has never been a known presence of truly wild species of the *B. oleracea* complex before.

Genetic diversity and structure of the analysed populations

In terms of genetic diversity, the Shannon Index for *B. oleracea* var. *oleracea* is higher than for the cultivated brassicas. This observation does not alter the consideration that the two groups belong to the same primary gene pool, comprising crops and their feral escapes.

B. oleracea var. *oleracea*, shows higher genetic diversity in Spain than in France. Looking at accessions with a similar sample size, Helgoland, Germany, shows higher diversity than Tenby, UK, while Rødvig, Denmark, is the least diverse population. At sub-regional level, a gradient appears of increasing diversity from north (Somme) to south-west (Asturias). Concerning the genetic diversity of individual populations, more reliable comparisons are possible since sample sizes are very similar. Most populations, independently from their geographic origin, are included between the values I = 0.070–0.100. Only very few stand out for their lower ('Rødvig' with I = 0.058) or higher diversity ('Petites

Dalles' in France with I = 0.101, 'Laredo' with I = 0.102, and 'Urgull' with I = 0.129 in Spain). While the population in Rødvig is known to be the result of a rather recent feralisation from the disposal of residuals of cultivation (Niels Jacobsen, pers. comm), 'Petites Dalles', 'Laredo' and 'Urgull' can be indicated as preferential populations to be conserved and used when looking for sources of diversity. The high diversity of 'Petites Dalles' is a confirmation of what was found in the earlier study by Maggioni et al. (2020).

A similar variability of genetic diversity can also be found by looking at populations/accessions of cultivated crops and other wild species. A few examples that stand out for higher diversity are the landraces 'Torimbia', Asturias (0.112) and 'De Baudo's neighbour', Sicily (0.122). Also in this case, 'De Baudo's neighbour' is a confirmation, since in a study by Maggioni et al. (2014), it was scored as the most diverse in a group of 16 kale landraces from Sicily. Among wild species, the highest diversity is shown by a population of *B. incana* from Capri, Campania (0.106) and by *B. rupestris* from Taureana, Calabria, showing the highest score (0.263) of the entire study, which places doubts about the identity of this population.

In the Analysis of Molecular Variance, the populations grouped in various combinations reconfirm the conclusions obtained through PCoA and STRUCTURE analyses. Namely, variance is balanced within and among species groups, confirming the existing differentiation among species. On the other hand, variance among populations is minimal (5%) between cultivated and wild-growing *B. oleracea*, a further indication that we are dealing with the same primary gene pool. Variance among populations is a bit higher (11–13%) in the case of *B. oleracea* (wild or cultivated) grouped by sub-regions or by countries, indicating the existence of some structuring and even a geographical gradient, although the Fst value lower than 0.15 and other parameters advise that there is a perceptible structuring within a rather homogeneous genetic background. The structuring between most of the landraces is more sensible (29% variation among populations), indicating that landraces cultivated in given localities tend to develop a stronger genetic identity than wild-growing populations in isolated cliffs. Evidently, human selection over landraces under cultivation has stronger effects in these cases than environmental factors under natural selection.

The genetic similarity between populations in France and Spain is very high between all of the regions, with a slightly higher differentiation between French regions, while the highest identity values are recorded between Asturias and Cantabria. This result is re-proposing the existence of an 'AsCan' type (see above) but also indicates that this 'type' remains very close to all the other analysed populations of France and Spain (maximum distance is 0.054 between Cantabria and Somme). In comparison to the conclusions of Maggioni et al. (2020) based on AFLP studies, the SNP data analysed here give us a better resolution. The four French populations analysed here still have a low genetic differentiation (Fst lower than 0.15). However, it is possible to recognize a better clustering and a higher genetic differentiation among each other than is the case of the Spanish populations. The French populations also differ from each other more than they differ from the Spanish populations.

Comparison between landraces from Italy and Spain has shown that the key differentiation is between landraces themselves and does not depend on their geographical origin. Certainly, individual landraces acquire their specific identity under human selection in their given environment. However, the genetic background from which they differentiate is rather homogeneous and thus the Spanish vs. Italian origin is not a major differential factor. It is also striking that both the Spanish and the Italian landraces have 'Laredo' as one of the populations genetically closest to them. One possible interpretation is that the Laredo population better represents ancestral genetic material from which both landraces and other feral populations have derived. 'Laredo' could have been one of the initial sites of feralisation of an ancient batch of sufficiently diverse kales imported from the Mediterranean area. As the oldest possible age of feralisation, we could speculate around 2500 years, assuming the arrival of domesticated coles from the Eastern Mediterranean during classical antiquity.

Relationship between species and hypothesis of domestication

The results presented here correspond well to an interpretation where *B. incana* is the closest relative and likely wild ancestor of the domesticated *B. oleracea*, while all the Atlantic wild-growing *B. oleracea*, being genetically indistinguishable from the cultivated *B. oleracea*, are most likely feral populations. *B. incana* is indicated as the closest taxon to *B. oleracea* both by the PcoA and the STRUCTURE analysis. In the latter analysis, it is striking how clearly *B. incana* is clustered together with *B. oleracea*, as opposed to all the other taxa. Also, the values of genetic identity confirm *B. incana* as the closest to *B. oleracea* of all the wild Mediterranean taxa.

An alternative hypothesis might interpret the Atlantic *B. oleracea* as the closest wild ancestor and *B. incana* as a group of feral populations. This interpretation is much less likely considering that: 1) Feral individuals are expected to be genetically closer to the crop from which they escape than the ancestor crop wild relative that underwent a domestication event; 2) Feralisation of *B. oleracea* is more likely to generate glabrous populations, corresponding to the domesticated types, as all the *B. oleracea* populations of the Atlantic and other regions are. It would be difficult to explain feralisations generating only hairy populations, as it would be the case with the alternative interpretation of *B. incana* as feral, as opposed to the Atlantic *B. oleracea* as feral.

The studied samples of *B. incana* cover three sites in Campania, two in Sicily and one in Lazio. Therefore, they give a good representation of the Italian distribution range of *B. incana*. The preliminary results obtained with AFLPs by Prekalj et al. (2022) on a wider representation of the *B. incana* distribution range, indicate that all the populations from Italy and Croatia cluster together, offering a homogeneous genetic pattern, with the exception of a few different populations in the southern Adriatic area. These results confirm that our sample of *B. incana* is representative of most of the diversity of this taxon. Our hypothesis that *B. incana* could be considered as a potential ancestor of the *B. oleracea* crops contrasts with the findings by Mabry et al. (2021) who indicate *B. cretica* as the likely ancestor. *B. incana* has a distribution range that is largely overlapping with the ancient Greek colonies of southern Italy and thus it is compatible with the earliest ancient Greek sources of literary and linguistic references to cultivated coles, similarly to *B. cretica*. In terms of genetic analysis, a wide enough representative sample of the entire *B. cretica* and *B. incana* distribution range, in comparison with *B. oleracea* crops, has not been investigated yet. Such an analysis could offer more solid evidence about the most likely ancestor of the cole crops.

Other considerations are that each wild species confirms its own identity through a clear clustering. This refers to *B. macrocarpa*, *B. drepanensis*, *B. rupestris*, *B. montana* and *B. cretica*. However, there are individuals or even entire specific populations (e.g. in the case of *B. montana* and *B. rupestris*) that are not clustering with their specific group, most likely due to intercrossing with cultivated crops. An apparent area of overlap between a few cultivated, wild *B. oleracea*, *B. incana* and *B. montana* individuals could also be explained with occasions of intercrossing.

However, the analysis of *B. oleracea* and *B. incana* together excludes areas of overlap between these groups, reinforcing the idea that these groups are two close but separate taxa, probably the first derived from the other through domestication.

Finally, in the case of the two populations of *B. cretica*, we note that they form two very distinct clusters. This pattern may indicate that our population from Limni, Corfu is not a true *B. cretica*. Also, the *B. incana* population from Angelokastros, Corfu, was showing an anomalous pattern, which led us to exclude it from further analysis. It might be worthwhile to investigate more closely the so far poorly known populations growing in the area of overlap between the distribution ranges of *B. incana* and *B. cretica* (i.e. the Albanian coasts, the Greek Ionian islands, etc.) to better understand the taxonomic identity of these populations and possible origin of the crop.

Conclusions

We can summarize the conclusions from this study in the following points:

- The Spanish populations of *B. oleracea* var. *oleracea* are generally well-thriving and sometimes expanding or colonising new sites.
- The specific comparison between Spanish and French *B. oleracea* var. *oleracea* shows that Spanish populations hold higher genetic diversity, with the highest levels recorded in Asturias. Overall, the genetic differentiation between Spanish and French populations remains low. Within a context of low levels of differentiation between all these populations, clustering is more evident among the French populations.
- All the Atlantic populations of *B. oleracea* var. *oleracea* are genetically very similar to each other. They are also very similar to all the cultivated *B. oleracea* landraces from Italy and Spain. It is however possible to distinguish two main types, one that is confined to parts of Asturias and Cantabria in north-west Spain (K1 or 'AsCan' Type), and the other (K2) which is extended to the rest of the Atlantic coast and also includes the cultivated crops. The purest K2 populations are situated in the northern part of Europe. A number of intermediate, admixed, populations are present especially in western France and northeast Spain. This pattern determines a distinguishable gradient of diversity that is structured from north to south of the Atlantic coasts, although the genetic distance between any of these populations is minimal.
- At the middle of this geographical gradient from K1 to K2 lies the population of Laredo, which is itself a K2 type, with some degree of admixture. This population appears to be a pivotal one, since it stands out as being the genetically closest to more than 60% of all the populations and accessions tested in this study. It could be hypothesized that an initial event of feralisation took place in this geographical area, with new populations colonising towards the east and the north maintaining a K2 type, while populations colonising in the west direction evolved the K1 type. This interpretation does not exclude other independent events of feralisation that may complicate the overall pattern, as identified in the study by Mittel et al. (2020). In addition, intercrossing between home garden crops and nearby wild-growing populations have been confirmed in this study for the case of 'Tazones', adding complexity to the overall pattern.
- The Spanish populations of *B. oleracea* var. *oleracea* display higher genetic diversity than those from other countries. Looking at individual populations, the genetic diversity is similar throughout the Atlantic range, with a few populations standing out with higher values ('Urgull', 'Laredo' and 'Petites Dalles').
- The six tested cultivated landraces from Spain and Italy mostly maintain a specific identity, showing a clustered pattern. They belong to the K2 type and thus they are genetically closer to the *B. oleracea* var. *oleracea* populations from the north side of the gradient. The clustering of Spanish and Italian landraces seems to be more an effect of the clustering of individual landraces, than dependent on their wider geographic separation. Their overall distance from any of the *B. oleracea* var. *oleracea* populations remains very low. Genetic distance and genetic differentiation from *B. incana* is also rather low, while they remain quite distant and different with respect to all the other Mediterranean wild species.
- The observed pattern is a confirmation of the feral nature of all the wild-growing populations of *B. oleracea* var. *oleracea*. This is due to the impossibility of distinctly separating the cultivated crops from the wild-growing populations. The labile distinction between K1 and K2 types is not sufficient to hypothesise the existence of a separate wild species growing on the northwest coast of Spain from which domestication might have occurred. It is not sufficient, due to the minimal genetic distance between K1 and K2, and to other types of evidence accumulated by various authors about the tendency of *B. oleracea* to easily become feral as well as based on historical and geographical considerations.
- The wild species form distinct clusters, indicating the validity of their taxonomic interpretation. There are a few exceptions of individual populations in *B. montana* or *B. rupestris* that do not cluster as expected, and these are interpreted as subject to intercrossing with cultivated crops. Also, the populations from Corfu have shown an unexpected pattern, with two putative populations of *B. incana* and *B. cretica* that are clustering separately from

the rest of their taxonomic group. As the Ionian islands are in an area of overlap between the distribution range of *B. incana* and *B. cretica*, closer investigation into the taxonomic identity of populations in this area is recommended.

- The *B. incana* populations represent the closest wild species to the *B. oleracea* crops and thus the closest wild relative and possible source of domestication of *B. oleracea*.

Declarations

Acknowledgements

The authors wish to thank the Royal Physiographic Society of Lund, Sweden, for economic support to the molecular analysis and Nora Capozio, Bioversity International, for her critical proofreading and English language editing.

Funding

The Royal Physiographic Society of Lund, Sweden provided economic support for the molecular analysis.

Data availability statement

DArTseq data used in this study can be accessed at DOI https://doi.org/281/zenodo.10877768.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

LM, GBP and RvB contributed to the study conception and design, and material preparation. DNA extraction and genotyping were outsourced. Data analysis was performed by TB and LM. The first draft of the manuscript was written by LM and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

1. Drenckhahn D (2017) Vorkommen des Atlantischen Wildkohls (*Brassica oleracea* L. subsp. *oleracea*) an den Kreidefelsen der Ostseeinsel Rügen, Deutschland. Forum geobotanicum 7:18-26 https://doi.org/10.3264/FG.2017.0116
2. Earl DA, von Holdt BM (2012) STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method. - Conservation Genet Resour 4:359–361. https://doi.org/10.1007/s12686-011-9548-7
3. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 14(8):2611-20. doi: 10.1111/j.1365-294X.2005.02553.x
4. Gómez-Campo C, Aguinagalde I, Ceresuela JL, Lázaro A, Martinez-Laborde JB, Parra-Quijano M, Simonetti E, Torres E, Tortosa ME (2005) An exploration of wild *Brassica oleracea* L. germplasm in Northern Spain. Genet Resour Crop Evol 52:7-13. https://doi.org/10.1007/s10722-005-4677-z
5. Gruber B, Unmack PJ, Berry OF, Georges A (2018) dartR: An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Molecular Ecology Resources, 18, 691–699. doi: 10.1111/1755-0998.12745

6. Hodgkin T (1995) Cabbages, kales, etc. In: Smartt J, Simmonds NW (eds) Evolution of crop plants. 2nd edn. Longman, London, pp 76-82
7. Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res. 29:E25. doi: 10.1093/nar/29.4.e25
8. Mabry ME, Turner-Hissong SD, Gallagher EY et al (2021) The evolutionary history of wild, domesticated, and feral *Brassica oleracea* (Brassicaceae). Mol Biol Evol 38(10):4419–4434. https://doi:10.1093/molbev/msab183
9. Mabry ME, Bagavathiannan MV, Bullock JM et al (2023) Building a feral future: Open questions in crop ferality. Plants, People, Planet 1–15. https://doi.org/10.1002/ppp3.10367
10. Maggioni L (2015) Domestication of *Brassica oleracea* L. Dissertation, Swedish University of Agricultural Sciences. http://pub.epsilon.slu.se/12424/
11. Maggioni L, von Bothmer R, Poulsen G, Branca F, Bagger Jørgensen R (2014) Genetic diversity and population structure of leafy kale and *Brassica rupestris* Raf. in South Italy. Hereditas 151:145-158 https://doi.org/10.1111/hrd2.00058
12. Maggioni L, López Noriega I, Lapeña I, Holubec V and Engels JMM (2015) Collecting plant genetic resources in Europe: a survey of legal requirements and practical experiences. In: Coolsaet, B. et al. (eds.) Implementing the Nagoya Protocol: comparing Access and Benefit-Sharing regimes in Europe. Leiden/Boston: Brill Nijhoff. pp. 327-362.
13. Maggioni L, von Bothmer R, Poulsen G, Lipman E (2017) Domestication, diversity and use of *Brassica oleracea* L., based on ancient Greek and Latin texts. Genet Resour Crop Evol 65:1-23. https://doi.org/10.1007/s10722-017-0516-2
14. Maggioni L, von Bothmer R, Poulsen G et al (2020) Survey and genetic diversity of wild *Brassica oleracea* L. germplasm on the Atlantic coast of France. Genet Resour Crop Evol 67, 1853–1866. https://doi.org/10.1007/s10722-020-00945-0
15. Mitchell ND (1976) The Status of *Brassica oleracea* L. subsp. *oleracea* (wild cabbage) in the British Isles. Watsonia 11:97-103
16. Mittel EA, Cobbold CA, Ijaz UZ, Kilbride EA, Moore KA, Mable BK (2020) Feral populations of *Brassica oleracea* along Atlantic coasts in western Europe. Ecology and evolution 10 (20): 11810-11825. https://doi: 10.1002/ece3.6821
17. National Biodiversity Data Centre, Ireland, Wild Cabbage (Brassica oleracea), accessed 27 July 2023, <https://maps.biodiversityireland.ie/Species/28500
18. NDFF & FLORON 2023. Verspreidingsatlas Vaatplanten 27 July 2023, https://www.verspreidingsatlas.nl/5481
19. Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28, 2537-2539. http://dx.doi.org/10.1093/bioinformatics/bts460
20. Prekalj, B, Šatović Z, Liber Z, Goreta Ban S, Ban D, Bogdanović S (2022). Genetic structure and diversity of the *Brassica incana* L. (Brassicaceae) complex in the Mediterranean. In: 7th Croatian Botanical Symposium with international participation, Zagreb, 12-14 September 2022, Book of Abstracts, pp 26-27. Knjiga-sazetaka_7HBS.pdf (hbod.hr)
21. Pritchard, JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2):945-59. doi: 10.1093/genetics/155.2.945
22. R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org
23. Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D, Kilian A (2011). Diversity arrays technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping

for molecular breeding of Eucalyptus. BMC Proc. 5:P54. BMC Proceedings 5(Suppl 7):P54. doi:10.1186/1753-6561-5-S7-P54

24. Snogerup S, Gustafsson M, von Bothmer R (1990) *Brassica* sect. *Brassica* (Brassicaceae). I. Taxonomy and variation. Willdenowia 19:271-365
25. Song K, Osborn TC, Williams PH (1990) *Brassica* taxonomy based on nuclear restriction fragment length polymorphism (RFLPs). 3. Genome relationships in *Brassica* and related genera and the origin of *B. oleracea* and *B. rapa* (syn. *campestris*). Theor Appl Genet 79:497-506
26. USDA, Agricultural Research Service, National Plant Germplasm System. 2023. Germplasm Resources Information Network (GRIN Taxonomy). National Germplasm Resources Laboratory, Beltsville, Maryland. URL: https://npgsweb.ars-grin.gov/gringlobal/taxon/taxonomydetail?id=317370. Accessed 25 July 2023.
27. Villanueva RAM, Chen ZJ (2019) ggplot2: Elegant Graphics for Data Analysis (2nd ed.). - Meas. Interdiscip. Res. Perspect. 17: 160–167. https://doi.org/10.1080/15366367.2019.1565254
28. Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. - Springer-Verlag New York.

Figures

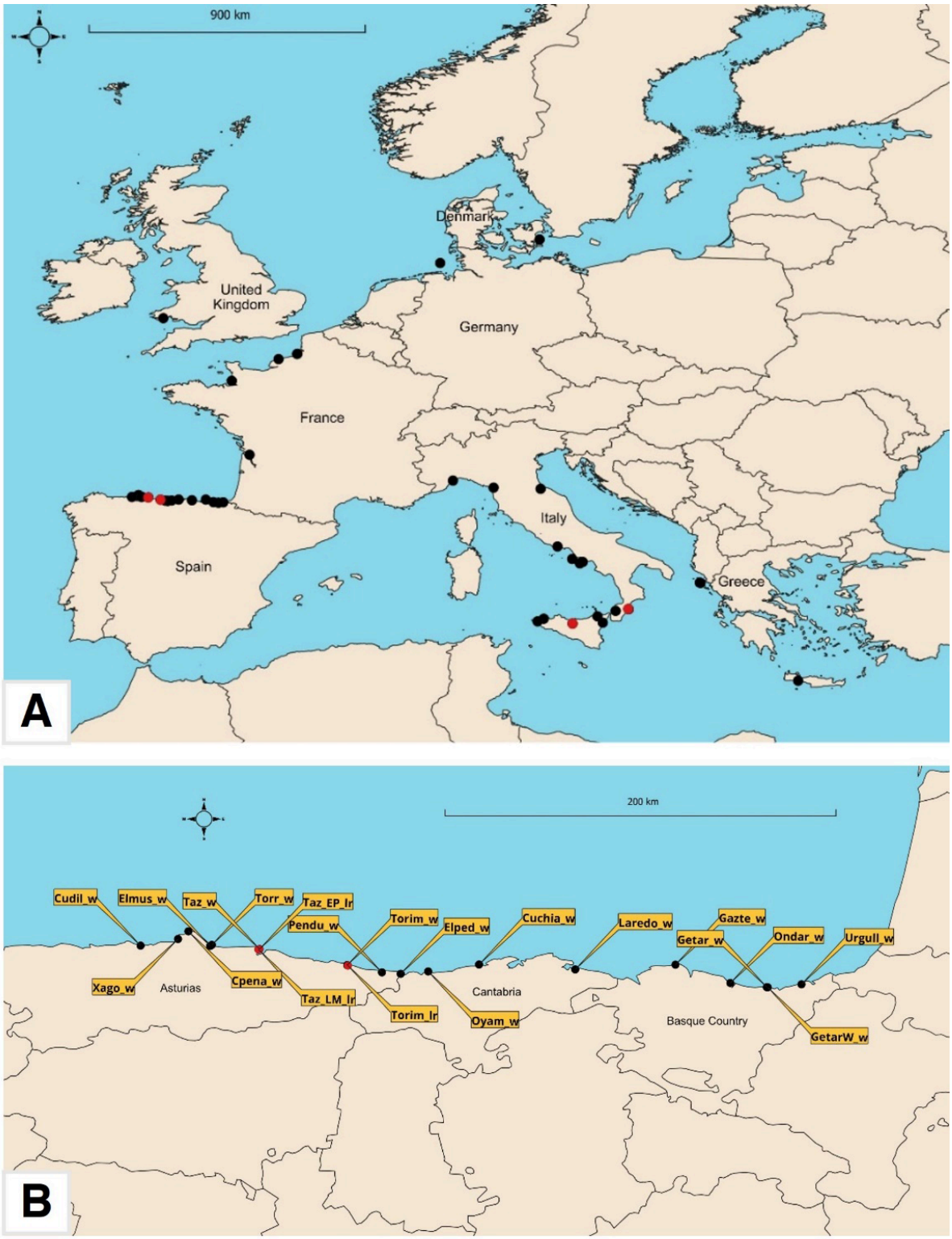

Figure 1

A) Original locations of all accessions included in this study and B) an enlarged map showing the original locations of the Spanish accessions included in this study. Collection sites for landraces and wild accessions are shown as red and black circles, respectively. Maps were created using the free and open source QGIS and free vector and raster map data @naturalearthdata.com.

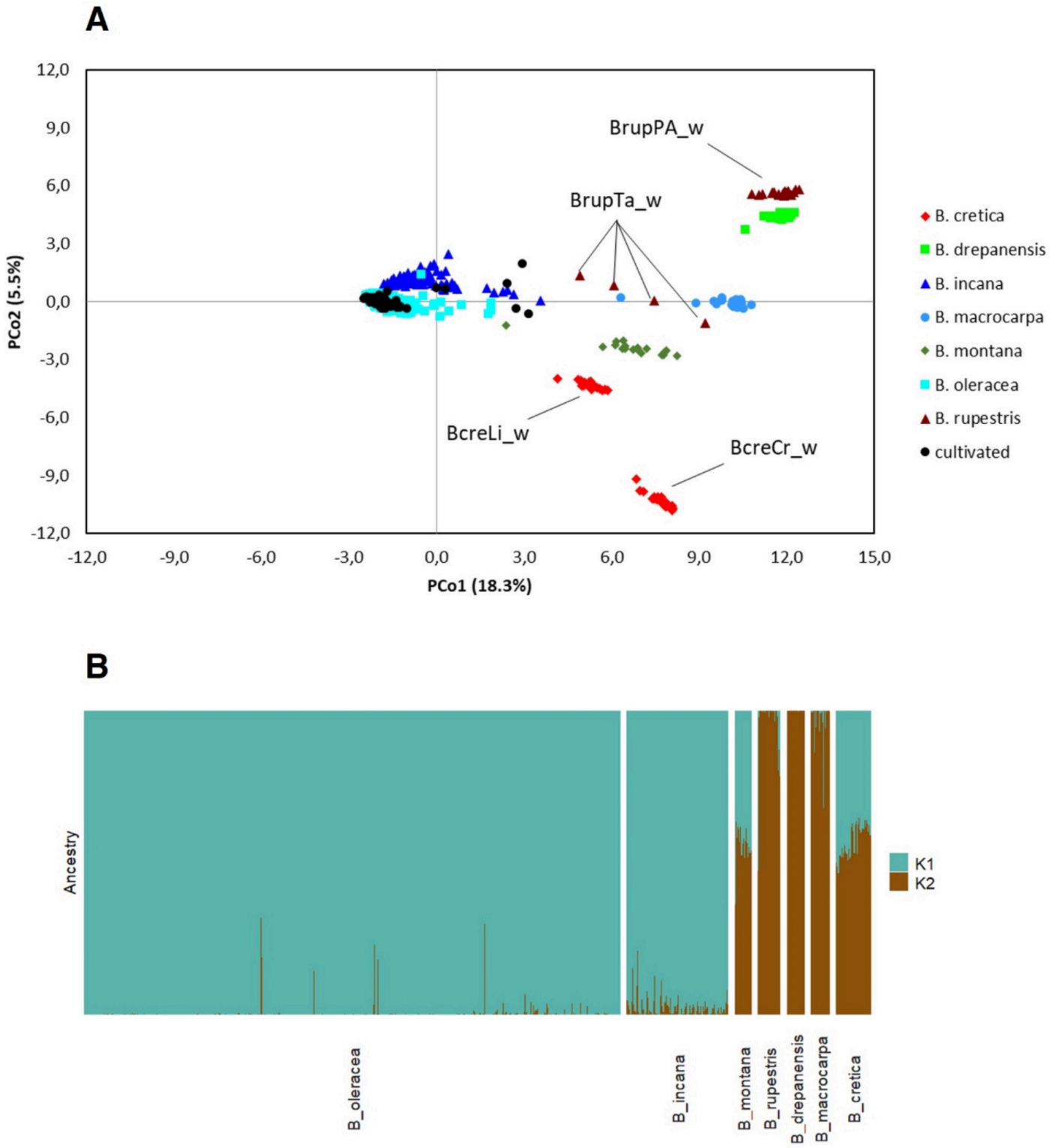

Figure 2

PCoA of all populations grouped by species. Symbols represent single individuals. The cumulated percentage of variation explained by the first two axes = 23.8%. For the number of individuals in each population/accession see Table 1. B) STRUCTURE bar plot of all accessions grouped by species for K=2.

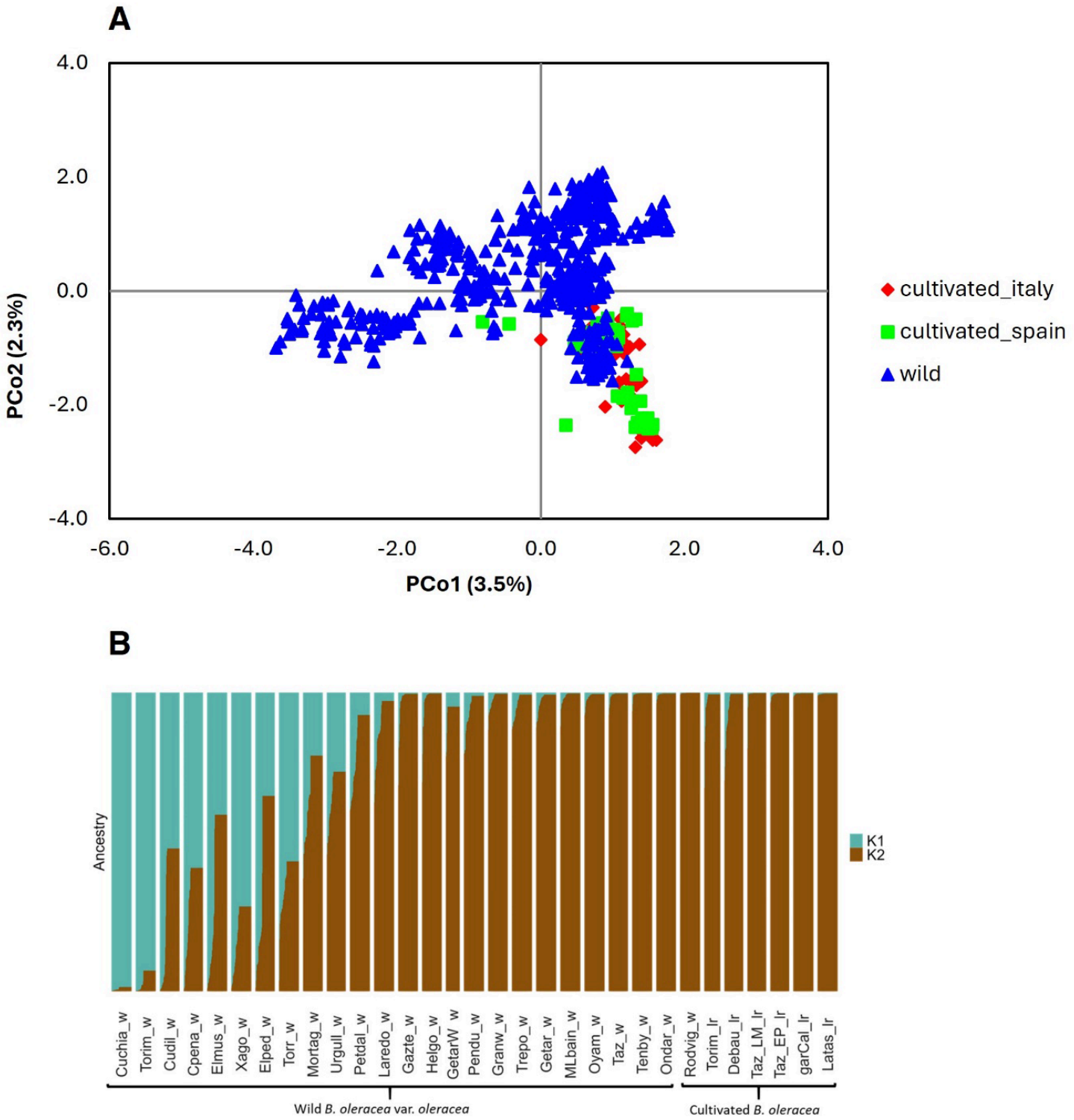

Figure 3

A) PCoA of the wild B. oleracea var. oleracea(470 individuals), cultivated B. oleraceafrom Italy (57 individuals) and cultivated B. oleracea from Spain (44 individuals). Symbols represent single individuals. The cumulated percentage of variation explained by the first two axes= 5.8%. B) STRUCTURE bar plot of wild B. oleracea var. oleracea and cultivated B. oleracea for K=2

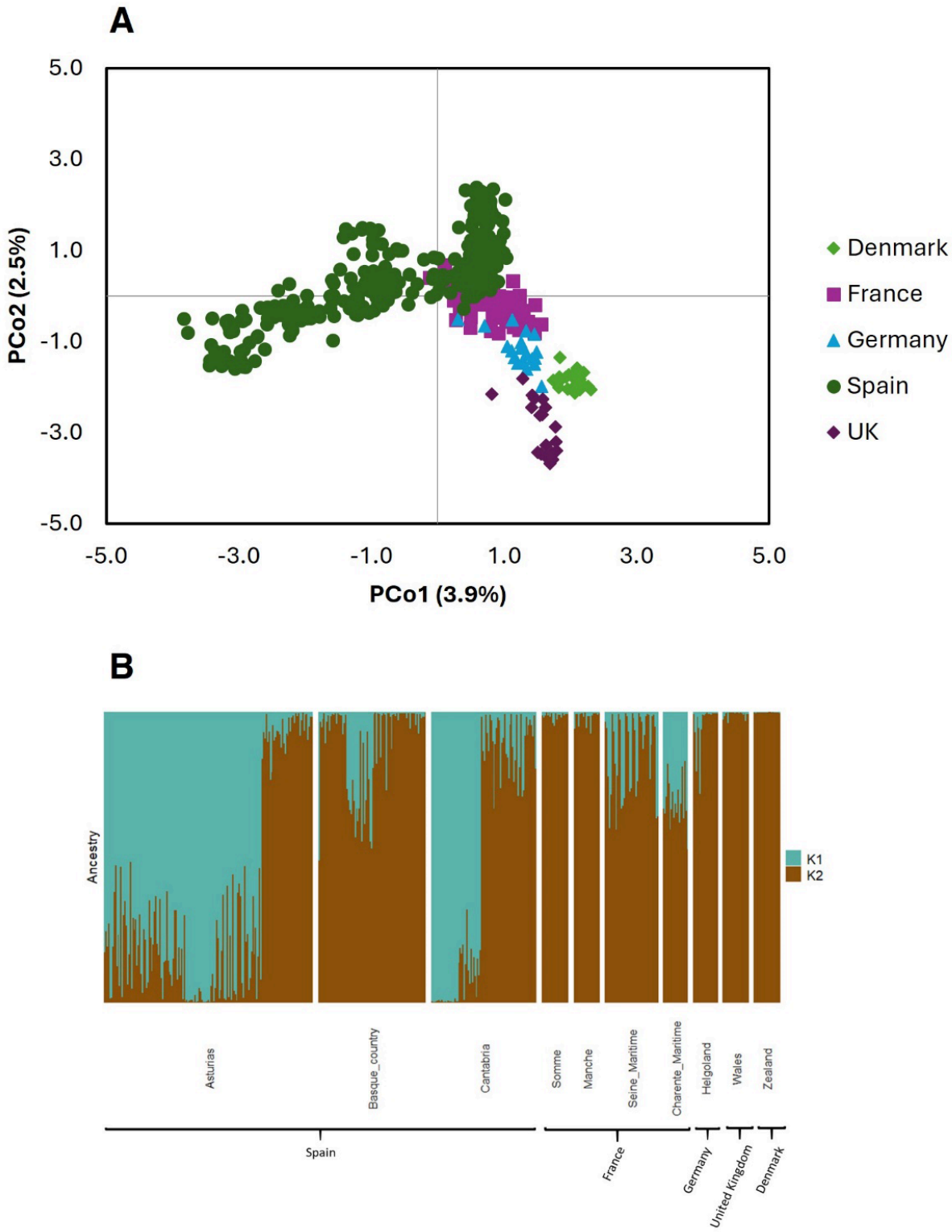

Figure 4

A) PCoA of the wild B. oleracea var. oleracea(470 individuals), grouped by country. Number of individuals: Spain (313); France (98); Germany (19); UK (20); Denmark (20). Symbols represent single individuals. The cumulated percentage of variation explained by the first two axes = 6.43%. B) STRUCTURE bar plot of B. oleracea var. oleraceapopulations for K=2 grouped by country of origin.

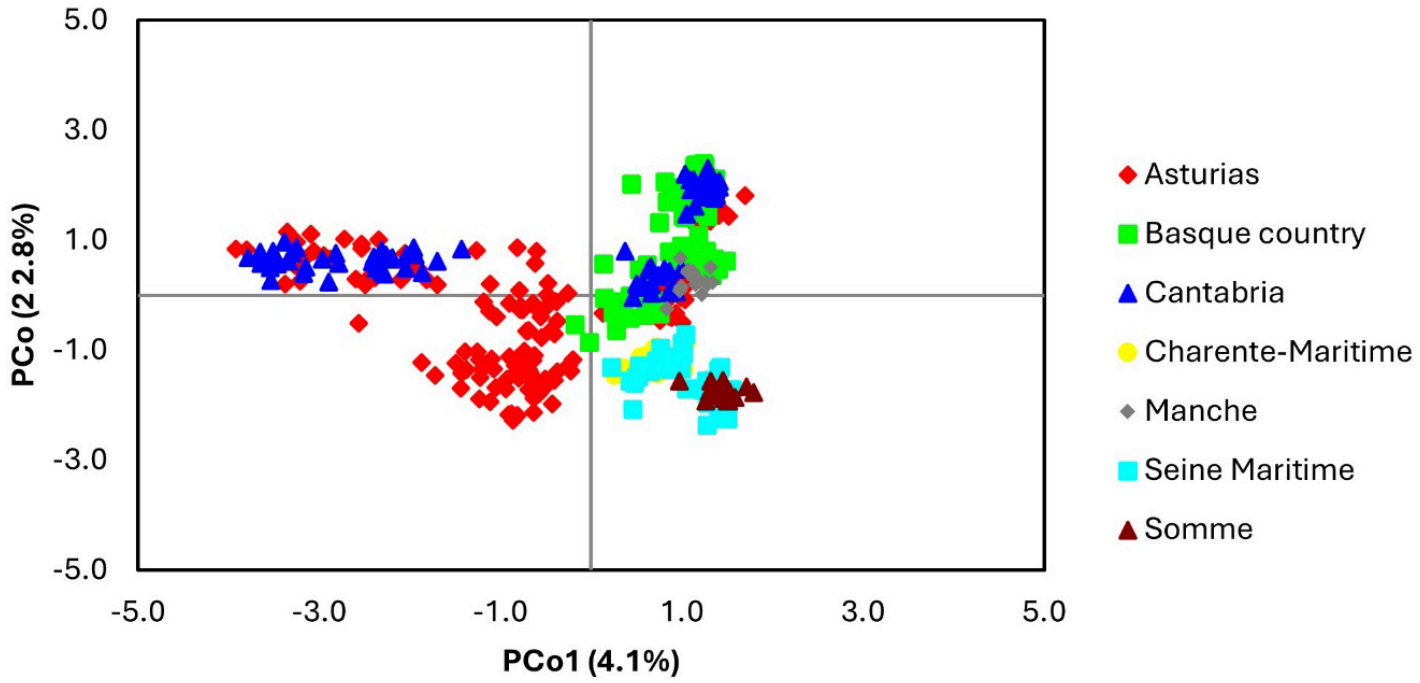

Figure 5

PCoA of the Spanish and French B. oleracea var. oleracea(411 individuals), grouped by region. Number of individuals: Asturias (155); Basque Country (80); Cantabria (78); Charente-Maritime (19); Manche (19); Seine-Maritime (40); Somme (20). Symbols represent single individuals. The cumulated percentage of variation explained by the first two axes = 6.9%.

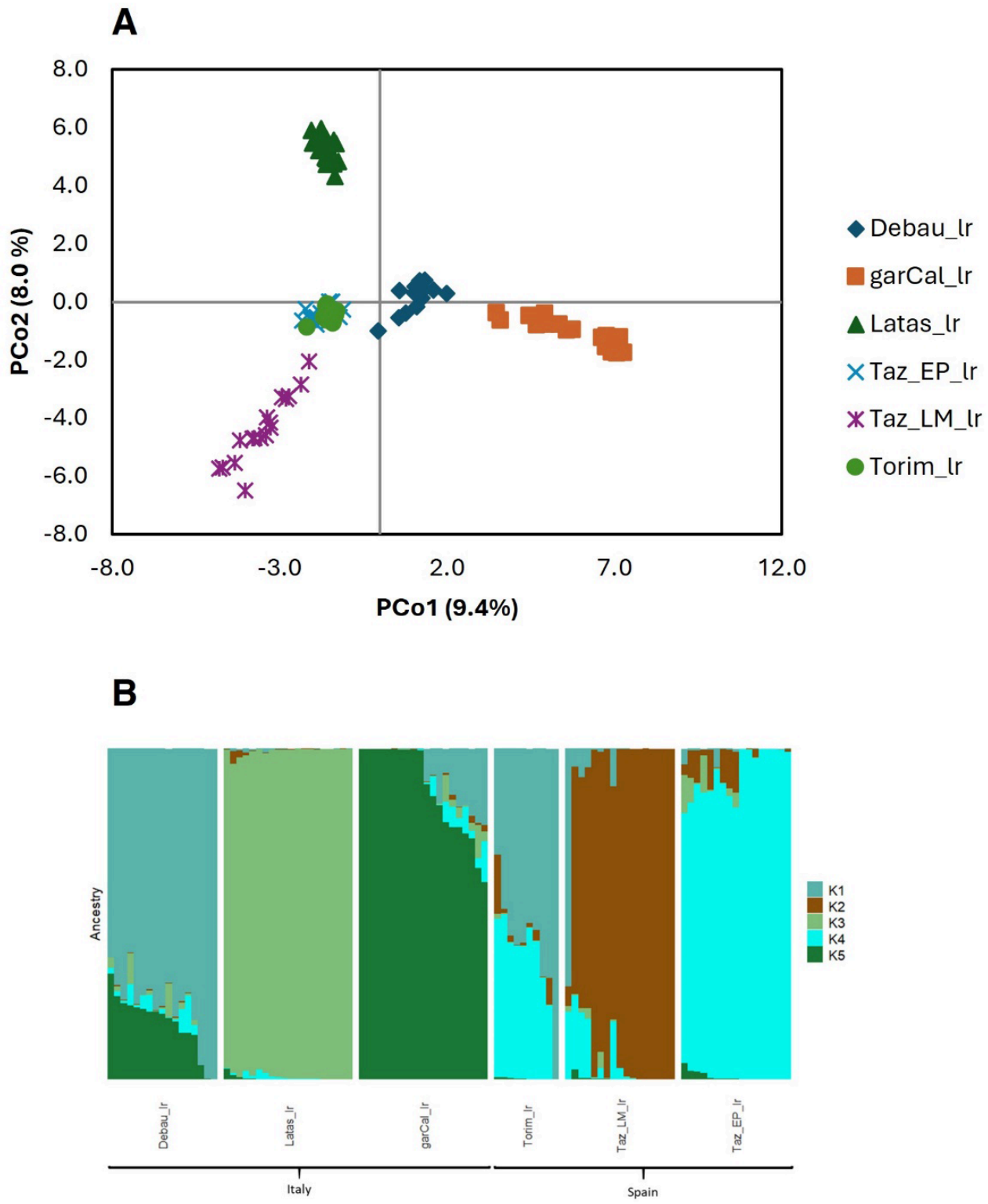

Figure 6

A) PCoA of the Italian and Spanish landraces grouped by individual populations. Symbols represent single individuals. The cumulated percentage of variation explained by the first two axes = 17.4%. For number of individuals in each population/accession see Table 1. B) STRUCTURE bar plot of Italian and Spanish landraces grouped by individual populations for K=5.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- OnlineResource1.xlsx
- OnlineResource2.docx