The Evolutionary History of Wild, Domesticated, and Feral *Brassica oleracea* (Brassicaceae)

Makenzie E. Mabry[1], Sarah D. Turner-Hissong[2], Evan Y. Gallagher[1], Alex C. McAlvay[3], Hong An[1], Patrick P. Edger[4], Jonathan D. Moore[5], David A. C. Pink[6], Graham R. Teakle[7], Chris J. Stevens[8,9], Guy Barker[7], Joanne Labate[10], Dorian Q. Fuller[9,11,12], Robin G. Allaby[7], Timothy Beissinger[13], Jared E. Decker[14], Michael A. Gore[15], & J. Chris Pires[1]

[1]Division of Biological Sciences and Bond Life Sciences Center, University of Missouri, Columbia, MO, U.S.A.

[2]Department of Evolution and Ecology University of California Davis, Davis, CA, U.S.A.

[3]Institute of Economic Botany, The New York Botanical Garden, Bronx, NY, U.S.A.

[4]Department of Horticulture, Michigan State University, MI, USA

[5]Systems Biology Centre, University of Warwick, UK

[6]Agriculture and Environment Department, Harper Adams University, UK

[7]School of Life Science. University of Warwick, UK

[8]School of Archaeology and Museology. Peking University, Beijing, China

[9]Institute of Archaeology, University College London, London, UK

[10]USDA, ARS Plant Genetic Resources Unit, Cornell AgriTech, Geneva, NY USA

[11]School of Cultural Heritage, Northwest University, Xi'an, Shaanxi, China

[12]Department of Archaeology, Max Planck Institute for the Science of Human History, Jena, Germany

[13]Division of Plant Breeding Methodology, Department of Crop Sciences, University of Goettingen, Goettingen, Germany

[14]Division of Animal Sciences, University of Missouri, Columbia, USA.

[15]Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, U.S.A.

CORRESPONDING AUTHORS

Makenzie E. Mabry; Email: mmabry44@gmail.com

J. Chris Pires; Email: piresjc@missouri.edu

ABSTRACT

Understanding the evolutionary history of crops, including identifying wild relatives, helps to provide insight for conservation and crop breeding efforts. Cultivated *Brassica oleracea* has intrigued researchers for centuries due to its wide diversity in forms, which include cabbage, broccoli, cauliflower, kale, kohlrabi, and Brussels sprouts. Yet, the evolutionary history of this species remains understudied. With such different vegetables produced from a single species, *B. oleracea* is a model organism for understanding the power of artificial selection. Persistent challenges in the study of *B. oleracea* include conflicting hypotheses regarding domestication and the identity of the closest living wild relative. Using newly generated RNA-seq data for a diversity panel of 224 accessions, which represents 14 different *B. oleracea* crop types and nine potential wild progenitor species, we integrate phylogenetic and population genetic techniques with ecological niche modeling, archaeological, and literary evidence to examine relationships among cultivars and wild relatives to clarify the origin of this horticulturally important species. Our analyses point to the Aegean endemic *B. cretica* as the closest living relative of cultivated *B. oleracea*, supporting an origin of cultivation in the Eastern Mediterranean region. Additionally, we identify several feral lineages, suggesting that cultivated plants of this species can revert to a wild-like state with relative ease. By expanding our understanding of the evolutionary history in *B. oleracea*, these results contribute to a growing body of knowledge on crop domestication that will facilitate continued breeding efforts including adaptation to changing environmental conditions.

KEYWORDS cabbage, domestication, crop wild relatives, Mediterranean, origin, ecological niche

INTRODUCTION

"Greek legend has it that the cabbage sprung from where Zeus' sweat hit the ground."

-N. D. Mitchell (1976)

A key tenet of evolutionary and plant biology is understanding how plants respond and adapt to changes in environmental conditions, which can be better understood by leveraging genotypic diversity and investigating the connections between genotype and phenotype. Crop wild relatives (CWRs) provide pools of allelic diversity that at one time were shared through a common ancestor with cultivated relatives. Although Vavilov recognized the potential of CWRs in the early 1900s (Vavilov 1926), advances in genomics and genome editing techniques have enabled scientists to better realize the potential of CWRs as a source of diversity and novel traits for the improvement of cultivated populations (Prohens et al. 2017; Li et al. 2018; Fernie and Yan 2019; Khoury et al. 2020; Turner-Hissong et al. 2020). Yet these scientific advancements are hindered in that we still have not identified the CWRs of many important crop species. While cabbage may not have exactly formed from Zeus' sweat, its evolutionary history, including

identifying the closest living wild relative and origin of domestication, is still left unclear due to taxonomic confusion and the lack of genetic and archaeological evidence.

The horticultural crop *Brassica oleracea* L. has played an important role in global food systems for centuries, providing a source of leaf and root vegetables, fodder, and forage (Shyam et al. 2012). When first introduced to the species, Darwin drew many parallels between his theory of natural selection and the cultivation practices that led to the varied forms of this plant (Darwin and Gray 1868). While many people may recognize that various dog breeds are all part of the same species, they are often surprised to learn that the domesticated forms of *B. oleracea*, broccoli (var. *italica*), Brussels sprouts (var. *gemmifera*), cabbage (var. *capitata*), cauliflower (var. *botrytis*), kale (var. *acephala*), and kohlrabi (var. *gongylodes*) are all one species as well. The global market for *B. oleracea* crops was around 70.1 million metric tons, in terms of production for 2019 (The Food and Agriculture Organization; www.fao.org). Although just six major crop types comprise the majority of the U.S. market (Agricultural Marketing Service, Market News Reports; www.ams.usda.gov), outside these six major cultivars there exists at least 12 additional cultivated crop types (***SI Appendix*, Table S1**). These include lesser known varieties such as Chinese white kale or Cantonese gai-lan (Mandarian Mandarin *Jiè lán* 芥蓝; var. *alboglabra*), a leafy vegetable with florets, romanesco (var. *botrytis*) with unique fractal patterned curds, and walking stick kale (var. *longata*), which grows 6-12 feet (1.8 - 3.7 meters) in height.

Compared to other crops, surprisingly little is known about the progenitor species and origin of domesticated *B. oleracea*. Primary challenges in identifying the progenitor species include the number of wild species that share a single cytodeme and are interfertile with *B. oleracea* (2n = 18 chromosomes; similar genomic organization; referred to as the "C genome"), the corresponding confusion surrounding taxonomic relationships, and conflicting evidence regarding the center of origin. Wild relatives that share the C genome with domesticated *B. oleracea* include *Brassica bourgeaui*, *Brassica cretica*, *Brassica hilarionis*, *Brassica incana*, *Brassica insularis*, *Brassica macrocarpa*, *Brassica montana*, *Brassica rupestris*, and *Brassica villosa*. Throughout the literature, many of these species have been referred to by alternative names or have multiple subspecies. For example, *B. cretica* is described as having either three subspecies (subsp. *aegea, cretica,* and *laconica*; Snogerup et al. 1990) or only two (subsp. *cretica* and *nivea*; Gustafsson et al. 1976). The taxonomic confusion is perhaps best highlighted by L. H. Bailey, who stated that "*Some of these plants appear to be more confused in literature than in nature*" (Bailey 1930). The progenitor species of *B. oleracea* is further obscured by the presence of weedy, cabbage-like plants along the coastline of western Europe (England, France, and Spain), which have also been referred to as *B.*

sylvestris (Mitchell 1976) or *B. oleracea* var. *sylvestris* (Gladis and Hammer 2001). The role of these weedy populations in the domestication of *B. oleracea* is unclear, with some studies suggesting these coastal wild populations represent the progenitor species (Snogerup et al. 1990; Song et al. 1990), and others identifying these wild forms as plants that escaped cultivation (Mitchell 1976; Mitchell and Richards 1979).

Given the uncertainty surrounding wild relatives and weedy populations, researchers have proposed numerous hypotheses for the progenitor species of *B. oleracea* (**Table 1**). Hypotheses range from a single domestication with a single progenitor species (Song et al. 1990; Allender et al. 2007) to multiple domestications arising from multiple progenitor species (de Candolle 1855; Neutrofal 1927; Lizgunova 1959; Helm 1963; Snogerup 1980; Heaney et al. 1987; Song et al. 1988; Swarup and Brahmi 2005). Findings that point to a single origin of domestication have proposed different wild species as the progenitor (Snogerup et al. 1990; Song et al. 1990; Hodgkin 1995; Maggioni et al. 2018). For instance, Neutrofal (1927) suggested that *B. montana* was the progenitor of cabbages and that *B. rupestris* was the progenitor of kohlrabi, while Schulz (1936) identified *B. cretica* as the progenitor of only cauliflower and broccoli. Helm (1963) proposed a triple origin in which a single progenitor species gave rise to cauliflower, broccoli, and sprouting broccoli, while kale and Brussels sprouts were derived from another unknown wild species, and that all other crop forms were derived from a third unknown wild species. Snogerup (1980) proposed that cabbages were derived from wild *B. oleracea*, kales were derived from both *B. rupestris* and *B. incana*, and that Chinese white kale was derived specifically from *B. cretica* subsp. *nivea*.

Due to the lack of consensus on the progenitor species, the center of origin for *B. oleracea* has also remained obscure. One hypothesis is that domesticated *B. oleracea* originated in England from weedy *B. oleracea* populations, with early cultivated forms brought to the Mediterranean, where selection for many of the early crop types occurred (Hodgkin 1995). Other studies point specifically to Sicily, which boasts a large diversity of wild relatives, as the center of domestication (Schiemann 1932; Lizgunova 1959). This conforms with the observations of Nikolai Vavilov (1951) that plants tended to be domesticated in a finite number of global centers of diversity, which includes the Mediterranean. Most recently, linguistic and literary evidence provided support for domestication in the Eastern Mediterranean, where there is a rich history of expressions related to the usage and cultivation of *B. oleracea* crop types in early Greek and Latin literature (Maggioni et al. 2010; Maggioni et al. 2018).

Using newly generated RNA-seq data for a diversity panel of 224 accessions that includes 14 cultivar types and nine wild relatives, representing the largest and most diverse collection of this species and its wild relatives to date, we integrate phylogenomics, population genomics, ecological niche modeling,

archaeological, and literary evidence to clarify the taxonomy, identify the closest living wild relative, and provide insight on the origin of domestication for *B. oleracea*.

RESULTS

Sequencing depth and SNP identification. RNA sequencing of 224 samples resulted in an average of 88,598,754 reads per sample, with a range of 59,543,560 - 151,814,032 reads. The minimum per-sample sequencing depth recovered was 9X, with a maximum depth of 12X. After mapping reads to the *B. oleracea* TO1000 genome (Parkin et al. 2014), SNPs were filtered to exclude those with a Fisher strand (FS) value greater than 30 and quality depth (QD) less than 2.0. This recovered 942,357 variants in total, with 879,865 variants on chromosomes 1-9 and 62,492 variants on remaining scaffolds. Chromosomal SNPs were then filtered to exclude sites with greater than 60% missing data, sites with mean per-sample depth values less than 5, and indels, resulting in a total of 103,525 SNPs. After a final filtering step for linkage disequilibrium (LD), a conservative final dataset of 36,750 SNPs was generated. For all samples, no mapping bias was detected when comparing the percentage of uniquely mapped reads across cultivar groups, species, and sequencing lane (***SI Appendix*, Fig. S1**).

Phylogeny and population clustering distinguish wild and feral populations. Sampling of *B. oleracea* cultivars included eight types of kales, five types of cabbages, Brussels sprouts, broccoli, cauliflower, Romanesco (var. *botrytis*), and kohlrabi (***SI Appendix*, Table S2**). Together, these cultivated types accounted for 188 of the 224 total samples. The remaining 36 samples included previously identified wild relatives: putatively wild *B. oleracea*, *B. cretica*, *B. incana*, *B. montana*, *B. hilarionis*, *B. insularis*, *B. macrocarpa*, *B. rupestris*, and *B. villosa*. The phylogenetic reconstruction of all 224 samples using SNPhylo (Lee et al. 2014) recovered several well-supported clades with greater than 70% bootstrap support, although overall support was generally poor (less than 70% bootstrap support), especially along the backbone. Chinese white kale, broccoli, cauliflower, romanesco, kohlrabi, curly kale, Brussels sprouts, *B. rupestris*, *B. macrocarpa*, and *B. insularis* were all recovered as monophyletic. Aside from red cabbages, cabbages were also monophyletic, but with only 55% bootstrap support. Seven cultivars (collards, tronchuda kale, savoy cabbage, perpetual kale, red cabbage, and marrow cabbage) were found throughout the tree as polyphyletic assemblages. Several wild samples were recovered within the cultivar clade, including two samples of *B. cretica* (196, 199), one sample of *B. montana* (222), and all samples of putatively wild *B. oleracea* (175, 176, 177; sample names in bold text; **Fig. 1**). We also recovered a group in the cultivar clade consisting of five samples of three wild species, *B. incana* (205, 208, 209), *B. villosa* (233), and *B. cretica* (195), labeled ‘WildC-2’ (for wild samples with the C genome). Many of these “wild” samples also share most or all of their ancestry with cultivars. At K = 2, in our fastSTRUCTURE analyses (Raj et al. 2014),

samples clustered as either cultivars or wild (**Fig. 1**). We find that two samples of *B. incana* (204, 207; likely both from Crimea), which are sister to all cultivated samples, share 100% of their ancestry with cultivated types, as do two samples of *B. cretica* (196, 199), one sample of *B. montana* (222), and all three samples of putatively wild *B. oleracea* (175, 176, 177). Together with the placement in the phylogeny, these analyses indicate that these are not truly wild samples, but represent feral types, defining feral here as either exoferal (a domesticated population derived from admixture with either a divergent population, a wild conspecific, another domesticated species, or another wild species) or endoferal (a population of domesticated plants that has escaped from cultivation without the aid of introgression/hybridization with wild conspecifics; Gering et al. 2019). Our newly identified WildC-2 shows mixed wild and cultivar ancestry, which was also observed for one sample of tronchuda kale (30). The marginal likelihood was maximized at K = 3, in which a cluster comprised of broccoli, cauliflower, and Chinese white kale separated from other cultivated types. At K = 4, Chinese white kale was distinct from broccoli and cauliflower. The structure in the data was best explained by K = 5, in which the clade comprised of *B. insularis* and *B. macrocarpa* was separated and had shared ancestry with *Brassica cretica* (198), *B. hilarionis*, *B. montana* (224), and one sample of tronchuda kale (30). Additional K values showed similar patterns (***SI Appendix*, Fig. S2**).

Principal component analysis (PCA) also separated cultivars from most wild samples (***SI Appendix*, Fig. S3D, E & F**). The PC1 axis distinguishes wild species from cultivars and the PC2 axis separates WildC-2 from all other wild species (triangles with black outlines). While one sample of *B. cretica* (198) clusters closest to cultivated types, samples of *B. incana*, which were not in WildC-2, along with one sample of *B. montana* (222), two samples of *B. cretica* (196, 199), and all three samples of *B. oleracea* (175, 176, 177) cluster with the cultivars, corroborating the phylogenetic analyses. To further investigate the clustering patterns of *B. cretica* to cultivars, we included four additional wild-collected samples of two *B. cretica* subspecies (A and B = subsp. *nivea*, C and D = subsp. *cretica*; **Fig. 2A & B**; ***SI Appendix*, Fig. S3A;** labeled SRA in figure legend; Kioukis et al. 2020). Adding these samples supports the results of other studies that *B. cretica*, as a species, is very diverse. While sample C does not group with other *B. cretica* samples using the PC1 axis, the PC2, PC3, and PC4 axes show much tighter clustering among the four wild-collected samples and one of our samples of *B. cretica* (198), indicating that our *B. cretica* (198) sample is an informative representative of wild-collected *B. cretica* (**Fig. 2A & B; *SI Appendix*, Fig. S3A)**.

For crop samples, estimates of inbreeding coefficients from PCAngsd (Meisner and Albrechtsen 2018) roughly matched expectations for the frequency of heterozygotes under Hardy-Weinberg equilibrium, while inbreeding coefficients for wild species suggest excess homozygosity (***SI Appendix*, Fig. S4**), possibly

reflecting cultivation practices for germplasm management and the relative isolation of wild populations (i.e. small effective population size), respectively. Feral samples, those which were identified as wild taxa, but were found more closely related to cultivars than to wild taxa in our phylogeny and clustered with cultivated samples in our PCA (*B. cretica* -196, 199; *B. incana* - 204, 207; *B. montana* -222; and wild *B. oleracea* -175, 176, 177), show patterns of heterozygosity that are similar to crop samples, as do the four samples of *B. cretica* from Kioukis et al. (2020). Our WildC-2 exhibited patterns of excess homozygosity more similar to other wild taxa.

Domestication is also reflected in the transcriptome. Using expression profiles (transcript abundances) of 51,438 genes for our original 224 samples, we tested if cultivars and wild samples would still cluster separately based on the transcriptome. Overall, results and clustering patterns were similar to analyses using SNPs, with the axes of PC1 and PC2 separating most wild species from cultivars (**Fig. 2C; *SI Appendix*, Fig. S3B & C**). We again found the same samples of *B. incana* (204, 207), *B. cretica* (196, 199), *B. montana* (222), and *B. oleracea* (175, 176, 177) clustering with the cultivars, but in expression analyses WildC-2 clustered with the other wild samples, rather than separately as in our SNP based PCA. Hierarchical clustering of the expression profiles recovered similar patterns with two major groups: wild and cultivated, again with WildC-2 clustering with the other wild samples (***SI Appendix*, Fig. S5**). While most cultivar groups were not recovered as unique clusters, there were a few exceptions. Brussels sprouts, Chinese white kale, and curly kale all formed distinct clades, which corresponds to what we know about their growth habit. Since RNA was collected at the 7th leaf-stage, before substantial morphological differentiation occurs between cultivars, it is not too surprising that they do not cluster distinctively by cultivar. Curly kale is almost immediately visually distinguishable from other cultivars in that the first true leaves have margins which are already undulate and/or frilled, in contrast to the more lanceolate (i.e., long, widest in the middle, with tapered tips) leaves observed in most cultivars. Brussels sprouts are also easily identifiable at this early growing stage as they have short, oblong to nearly circular leaves. While Chinese white kale leaves look more similar to the lanceolate shape of other cultivars, they grow more rapidly and plants in this group are annual instead of biennial, which may explain why these accessions cluster separately from other cultivars.

To identify modules of genes that might be driving the observed clustering patterns, we used weighted correlation network analysis (WGCNA; Langfelder and Horvath 2008). We found that 48 modules, ranging in size from 34 to 35,981, provided the best fit for the data (***SI Appendix*, Table S3**). To assess what types of biological processes were overrepresented in these modules, we used syntenic *Arabidopsis thaliana* genes and performed a GO analysis through PANTHER v. 16.0 (Mi et al. 2021). Overlap of *B. oleracea* with *A. thaliana* genes ranged from 17% to 98.3%, perhaps indicating that some modules are more

conserved while others are unique to *B. oleracea*. Modules which were more conserved between the two species included genes related to herbivory defense compound production (secondary metabolite biosynthetic process, phenylpropanoid biosynthetic and metabolic processes), wound formation (suberin biosynthetic processes), and wax formation (wax biosynthetic and metabolic processes), the latter of which may be correlated to the characteristic glaucous leaves of cultivated *B. oleracea* (***SI Appendix*, Table S4).** Within the top five conserved modules, the transcript abundance (TPM) was significantly different among the different groups (P-value = < 2 e-16 for modules 7, 13, 31, & 34; P-value = 2.19 e-11 for module 30). Post-hoc comparisons using Tukey's honestly significant difference (HSD) revealed that transcript abundance in cultivars was significantly different compared to that of wild relatives across conserved modules, except for *B. hilarionis*, which was not recovered as significantly different from cultivars for any module. WildC-2 along with other identified feral samples had significantly different transcript abundance compared to cultivars for modules 7, 13, 31, and 34, but not for module 30. Significant differences were also found between WildC-2 and feral samples compared to wild relatives for several modules, with no obvious patterns across modules (***SI Appendix*, Fig. S6; Table S5)**.

Species tree and admixture inference indicate *Brassica cretica* is the closest living wild relative. Given the results of population clustering using both SNPs and expression profiles, we further interrogated the species level relationships between wild relatives and cultivar groups by resolving the backbone of the phylogeny. Using the PoMo model (Schrempf et al. 2016) as implemented in IQ-Tree (Nguyen et al. 2015) and only including samples representing monophyletic groups as determined in the sample-level phylogeny, we found strong support for *B. cretica* as the closest living wild relative to cultivated *B. oleracea* (**Fig. 3A**). Notably, for our species tree analyses, we included only one sample of *B. cretica* (198). This sample was used for species reconstruction due to its placement near other wild taxa in the sample level phylogeny and its clustering with wild collected *B. cretica* from Kioukis et al. (2020) in the PCA. The current distribution of *B. cretica* occurs throughout the Eastern Mediterranean, primarily in Greece, highlighting a potential origin of domestication (**Fig. 3B**). Another suggested wild relative, *B. incana*, is strongly supported as belonging to the cultivar clade, sister to lacinato kale. While our sampling is limited in regard to the distribution of *B. incana* as a whole, this result supports our other findings that *B. incana* is not a completely wild assemblage, but that at least some populations are feral. Within cultivars, several expected relationships were recovered: collards and cabbage as sister lineages (Song et al. 1988; Farnham 1996), with Brussels sprouts sister to both; cauliflower and broccoli as sister clades (Song et al. 1988; Stansell et al. 2018), with romanesco sister to both; and Chinese white kale as sister to all other cultivars, agreeing with recent literature (Cheng et al. 2016; Stansell et al. 2018).

With the overall species relationships resolved, we aimed to tease apart the evolutionary history of the wild samples that clustered within the cultivar clade. Specifically, we asked if any of the identified feral samples were the products of admixture using TreeMix (Pickrell and Pritchard 2012). While the tree model without any migration edges explained 87.3% of the variance in the dataset, sequentially adding migration events to the tree resulted in five migrations events explaining 92% of the variation (**Fig. 4A; *SI Appendix*, Fig. S7**). Adding a single migration edge resulted in an admixture event from *B. cretica* (198) to a clade of [Chinese white kale + tronchuda cabbage]. To further test this event, we used four-population (*f*4) tests for treeness as implemented in TreeMix, where a significant non-zero value indicates the presence of gene flow (Reich et al. 2009; Pickrell and Pritchard 2012; **Fig. 4B**). While the tree [[tronchuda cabbage, kohlrabi],[*B. cretica* (198), *B. hilarionis*]] showed no significant evidence of gene flow (*f*4 = 0.0008 , Z = 1.094), replacing tronchuda cabbage with Chinese white kale indicated significant gene flow from *B. cretica* (198) to Chinese white kale (*f*4 = -0.0055, Z = -5.113). This result was further verified when adding a second migration edge, as the migration edge only included Chinese white kale, but the direction was reversed (from Chinese white kale to *B. cretica* (198)). The second event, from kohlrabi to a presumably feral sample of *B. cretica* (199), was supported by *f4* tests, with the tree [[kohlrabi, *B. cretica* (196)],[*B. cretica* (199), marrow cabbage]] indicating significant evidence of gene flow from kohlrabi to *B. cretica* (199) (*f4* = 0.012, Z = 10.5). This migration event is also seen phenotypically, as *B. cretica* (199) has a swollen stem when grown to maturity. No significant evidence of gene flow was found when substituting *B. cretica* (199) with *B. oleracea* (175), which is not expected to be involved in the admixture event (*f4* = 0.00023, Z = 2.68). Two admixture events provide evidence of potential exoferal origins for at least two samples, *B. oleracea* (175) and *B. cretica* (199). The four-population tree of [[*B. montana* (222), curly kale],[*B. oleracea* (175), broccoli]] suggests significant gene flow from *B. montana* (222) to *B. oleracea* (175) (*f4* = 0.315, Z = 15.77), as does the tree of [[tronchuda cabbage, Chinese white kale],[*B. cretica* (199), broccoli] for gene flow from Chinese white kale to *B. cretica* (199) (*f4* = -0.009, Z = -7.98). The fifth added migration edge from *B. rupestris* to WildC-2 explains the shared ancestry recovered in the fastSTUCTURE results. The test for treeness with [[curly kale, WildC-2],[*B. rupestris, B. macrocarpa*]] indicated significant admixture from *B. rupestris* to WildC-2 (*f4* = -0.006, Z = -6.50), but was non-significant when substituting WildC-2 with cauliflower (*f4* = -0.0003, Z = -0.338). In general, these analyses highlight that the evolutionary history of *B. oleracea* is characterized by many admixture events and lineages of exoferal origins.

Archaeological and literary evidence point to a late-Holocene domestication. To further investigate the origins of domesticated *B. oleracea*, we surveyed archaeological, literary, and artistic evidence (***SI Appendix*, Table S6 & S7**). The earliest reported claim of *B. oleracea* comes from an archaeological collection from the Austrian Alps. This collection comprises three seeds dated to the Middle Bronze Age

(ca. 3550-3350 years before present or BP; Schmidl and Oeggl 2005). However, the lack of illustrations and discussion of separation criteria from other *Brassica* species makes us question the reliability of this species-level identification, as seeds of *Brassica* species are difficult to tell apart. The only other find of similar antiquity is *B. oleracea* seeds from the Late Bronze Age/Early Iron Age, identified by scanning electron microscopy and radiocarbon dated directly between ca. 3250-2970 BP (Kaniewski et al. 2011). These finds are associated with destruction levels at Gibala, Tell Tweini in western Syria on the Mediterranean coast. While most of the archaeological finds are of seeds (***SI Appendix*, Table S6**), there is at least one documentation of pottery residues where lipids of *Brassica* leaf waxes were identified and dated to 850-750 BP (Evershed et al. 1992; Evershed et al. 1994). The authors attribute this to the boiling of leaves of *B. oleracea*, and given the lack of evidence for other commonly eaten *Brassica* leaves in England at this time, this would appear a likely identification.

The earliest literary references to *B. oleracea* date to Greek scholars 2500-2000 BP (***SI Appendix*, Table S7**). Hipponax's writing refers to a seven-leaf cabbage in an iambic verse (West 2011), while Hippocrates *On the Nature of Women*, written around 2410-2320 BP, refers to the use of cabbage, or krambe, in a few recipes (Totelin 2009). As early as 2320 BP, there is evidence for cultivar diversity. Theophrastus refers to three varieties: a curly-leaved type, a smooth-leaved type, and a wild type with a bitter taste, many branches, and many small round leaves (Yonge 1854). Pliny in his *Natural History* writing some 200 years later describes at least ten varieties in addition to those seen in the previous classical works (Pliny, the Elder and Rackham 1950). However, while most scholars accept that the Greek or Latin translations of 'cabbage' refer to *B. oleracea*, it is important to note that 'cabbage' is not a Greek word and that the word 'raphanos' is translated as both cabbage and *B. cretica* in the Greek-English Lexicon (Liddell and Scott 1940) and in Hort's (1916) translation of Theophrastus' *Historia Plantarum*. Certainly, there are differences between the subspecies of *B. cretica* that might be reflective of the varieties described by Theophrastus and Pliny, and which may explain the diversity we observed among *B. cretica* samples in our PCA results. Further, the description by Nicander (quoted by Athenaeus; Yonge 1854; p. 582) indicates that wild or perhaps feral forms of *B. cretica* were known in Ionia, the western coast of present-day Turkey, ca. 2150-2050 BP.

Late-Holocene environmental niche modeling highlights wild relatives' ranges. Based on archaeological information, the oldest relatively reliable occurrence for *B. oleracea* cultivation is dated 3250-2970 BP in Gibala NW Syria (Kaniewski et al. 2011). To predict what would be a suitable habitat for the wild relatives during the late-Holocene, we compiled occurrence records from GBIF (www.gbif.org) and (Snogerup et al. 1990), along with environmental data, to perform environmental niche modeling using MaxEnt 3.4.1 (Phillips et al. 2017). Notably, we find that *B. cretica* has an expanded Eastern Mediterranean

habitat suitability (**Fig. 3C**) that includes Cyprus. Presently, only *B. hilarionis* is known to occur in Cyprus (**Fig. 3B**), however modeling predicts that in the late-Holocene it would have had an expanded habitat suitability in the surrounding mainland coastal regions (**Fig. 3D**). Since most of these wild species are narrow island endemics (Snogerup et al. 1990), species are generally estimated to have little change from current day distributions (***SI Appendix*, Fig. S8 and Table S8**).

DISCUSSION

Multiple lines of evidence support a single Eastern Mediterranean origin. Our evidence from genome-scale, multilocus data along with archeology, literature, and environmental niche modeling best supports a single Eastern Mediterranean domestication origin for *B. oleracea*, corroborating the conclusions of (Maggioni et al. 2018) based on literary sources and (Maggioni et al. 2010) using linguistics. When modeling phylogeny and population structure, two Eastern Mediterranean species, *B. cretica* and *B. hilarionis*, are found as sister species to cultivars and are assigned ancestry from all cultivar populations for values of K from 2 to 5 (**Fig. 1**), consistent with these species being likely progenitor species of *B. oleracea* cultivars. In our species tree reconstructions, we find just *B. cretica* as sister to all cultivars, specifically sample 198, which clusters with wild-collected *B. cretica* samples from Kioukis et al. (2020) in our PCA (**Fig. 2A-B**), lending further support for *B. cretica* as the progenitor species. This same sample of *B. cretica* (198) as well as our sample of *B. hilarionis* are recovered as fairly homozygous, therefore they would likely be good starting material for future research related to *de novo* domestication via selective breeding or gene editing.

While we do recover evidence of admixture between *B. cretica* (198) and both wild and cultivated taxa, the placement of *B. cretica* (198) as the closest living wild relative does not change. However, an inferred admixture event from *B. cretica* (198) to *B. hilarionis* does result in a topological change in the placement of *B. hilarionis* as sister to *B. montana* (224; originally collected in Spain) (***SI Appendix*, Fig. S7**). This novel relationship has not been identified before and warrants additional study with greater taxon sampling. The second migration event involving *B. cretica* (198) is from Chinese white kale. This event lends further evidence of admixture with wild germplasm during the domestication process, consistent with other examples demonstrating that domestication is not a single event, but a series of events characterized by continuous gene flow between wild and cultivated populations (Beebe et al. 1997; Wang et al. 2017). Together with the phylogeographic discontinuity of wild *B. oleracea* samples and their Eastern Mediterranean progenitors **(Fig. 3B)**, the more distant phylogenetic placement of *B. insularis*, *B. macrocarpa*, and *B. villosa* (**Fig. 3A**), and strong patterns of shared ancestry between *B. incana* and

cultivars **(Fig. 1**), these results lead us to support the hypothesis of domestication in the Eastern Mediterranean with *B. cretica* as the closest living wild relative.

The role of ferality in the domestication of *Brassica oleracea*. Multiple lines of evidence highlight the role of wild and feral populations as pools of diversity that contributed to crop diversification during domestication (Beebe et al. 1997; Allaby 2010; Fuller et al. 2014; Wang et al. 2017). Our data supports a similar phenomenon in the domestication of *B. oleracea*: it appears that introgression from wild or feral populations contributed to the genetic composition of particular crops, and vice versa, which is revealed by in-depth analyses of admixture using population structure and tree-based methods (**Fig. 1; Fig. 4; *SI Appendix*, Fig. S2 & S7**). Several samples of wild relatives, including *B. cretica,* as well as wild *B. oleracea*, *B. incana*, *B. montana*, and *B. villosa*, are recovered as feral in all analyses.

While we find one sample of *B. cretica* (198) as the closest living wild relative, we also identify two samples of *B. cretica* (196 and 199) are likely feral and fall within the cultivar clade (**Fig. 1; see *SI Appendix*, Fig. S9 for photos**). Interestingly, Song et al. (1988) also recovered a polyphyletic *B. cretica* using RFLPs. Results presented here support previous findings that *B. cretica* was at one point at least partially domesticated. Snogerup et al. (1990) state that wild *B. cretica* was consumed as late as 1962 and, as noted in our literary results, some early references to *B. oleracea* in the literature could be translated as *B. cretica*, meaning the vast amount of described morphology in these works, which may be the result of cultivation, could now be reflected in the multiple named subspecies and described genetic diversity of modern *B. cretica* (Snogerup et al. 1990; Widén et al. 2002; Allender et al. 2007; Edh et al. 2007). Further, *B. cretica* was known to occur in Ionia (western coast of present day Turkey) ca. 2150-2050 BP and the evidence of *B. cretica* populations today in Lebanon, which are morphologically similar to *B. cretica* subsp. *nivea*, suggests widespread trade of these species by the earliest Mediterranean civilizations (Dixon 2006). However, these plants may have been introduced into these localities without cultivation as was proposed by Snogerup et al. (1990). Previous researchers have noted that *B. cretica* populations are typically found in coastal locations associated with ancient seaports, occupying their preferred ecological niche on chalk cliffs undisturbed by grazing (Mitchell 1976; Snogerup et al. 1990). We believe that these early forms of *B. cretica* may have played underappreciated roles in the domestication of *B. oleracea* crops and to fully understand the evolutionary history of *B. oleracea*, the demographic history and domestication story of *B. cretica* must be resolved.

Sources have hypothesized that wild populations of *B. oleracea* in England are the progenitor(s) for modern cultivars (Snogerup et al. 1990; Song et al. 1990), while others have proposed that these are escaped

cultivars (Mitchell 1976; Mitchell and Richards 1979). Consistent with these hypotheses, we find that the three wild *B. oleracea* samples in our study cluster with cultivars both phylogenetically and in PCA for both SNP data and expression profiles. Although these samples are from Canada (175), Denmark (176), and Germany (177), well outside the natural distribution range for *B. oleracea* - notably not from England, one of the hypothesized geographic origins - we suggest that an origin in England is unlikely given the archeological and literary data. Although the oldest archaeobotanical record for *B. oleracea* (Middle Bronze Age; ca. 3550 -3350 BP) is from Austria, we regard this evidence with caution as wild populations of *B. oleracea* are not presently found in Austria and the major *Brassica* crops in this region include *B. nigra* (Tutin 1964) or potentially cultivated turnip (*B. rapa*). Additionally, there is no compelling archaeological evidence to suggest the possible cultivation of cabbages in Europe prior to the Late Iron Age (2350-2050 BP) and Roman periods (1950-1650 BP), but there is evidence for knowledge of *B. oleracea* in Greece during this time (Maggioni et al. 2018; ***SI Appendix*, Tables S6 and S7)**. Overall, there are no records for *B. oleracea* from before this period within databases relating to the Eastern Mediterranean (Reihl 2014), Europe (Kroll 2001; Kroll 2005), Britain (Tomlinson and Hall 1996), the Czech Republic (Kreuz and Schäfer 2002), or within pre-dynastic and Pharaonic Egypt (Murray 2000), despite having documentation for other *Brassica* species. Evidence for *B. oleracea* in Europe does not start appearing until ca. 1850 BP, when the appearance of seeds increased and can be attributed to the spread of crops both within and on the periphery of the Roman Empire (Van der Veen 2011). Additionally, several studies that sampled wild *B. oleracea* populations in the British Isles (Mitchell 1976; Mitchell and Richards 1979), South West England (Raybould et al. 1999), Atlantic coasts of western Europe (Mittell et al. 2020), and Atlantic coast of France (Maggioni et al. 2020) support that these wild *B. oleracea* populations are feral populations, typically with low levels of genetic diversity and some degree of isolation from other populations. Lanner-Herrera et al. (1996) sampled populations across Spain, France, and Great Britain, concluding that each population evolved independently, while more recently Mittell et al. (2020) found that geographically close populations were more genetically different than distant populations. Our results provide additional evidence that feralization is commonplace for *B. oleracea* crops and that references to wild *B. oleracea* likely represent multiple, independent feralization events. Additional sampling of wild populations will enable opportunities to further investigate the relationships among these feral populations and cultivated crops.

Brassica incana, another suggested progenitor species (Snogerup 1980), is also supported as feral for the samples included in our analyses. Two of our five samples (204 and 207) are recovered as sister to all cultivars in our individual level phylogeny but are found to share 100% of their ancestry with cultivars rather than other wild taxa using fastSTRUCTURE when K = 2 (**Fig. 1**). Further, these two samples were

resolved as sister to lacinato kale in our species tree analysis, providing additional evidence that these samples represent a feral lineage, possibly of lacinato kale. This result may lend insight into why previous studies have found *B. incana* as sister to *B. oleracea* (Lázaro and Aguinagalde 1998; Mei et al. 2010; Arias and Pires 2012) and the observation by Snogerup et al. (1990) that samples of *B. incana* from the Crimea are more interfertile with cultivated *B. oleracea* than others. Although Snogerup et al. (1990) suggested that *B. incana* was more interfertile due to historical introgression, we do not find evidence for this for samples 204 and 207. However, the three other samples of *B. incana* (205, 208, 209), which belong to WildC-2, do show evidence of admixture with *B. rupestris*, likely explaining their clustering together both in the PCAs and phylogeny with *B. cretica* (195) and *B. villosa* (233) which also show admixture with *B. rupestris* (see ***SI Appendix*, Fig. S10 for photos)**. All three *B. incana* were collected in Italy from two locations and therefore do not well represent the known *B. incana* range (**Fig. 3B**), while the two other samples found in this clade, *B. cretica* (195) and *B. villosa* (233), were collected in Greece and Italy, respectively. While all five WildC-2 samples share an introgression event from *B. rupestris* (**Fig. 1; Fig. 4; *SI Appendix*, Fig. S7)**, they are from different germplasm collections (IPK-gatersleben and USDA National Plant Germplasm System), ruling out the inferred migration being the result of current cultivation practices. It is possible that at least three of these samples (*B. incana* 205, 208, 209) are related to the wild kale of Crimea, which is posited as a *B. rupestris–incana* hybrid that was transferred to the Crimea via trade (Dixon 2006). This suggests that there was early widespread cultivation of these *B. rupestris–incana* types (Dixon 2006) and provides a plausible explanation for why *B. incana* and *B. rupestris* are closely related in previous studies (Lannér et al. 1997; Mei et al. 2010). The other two samples in WildC-2 (*B. cretica* 195 and *B. villosa* 233), possibly represent misidentifications, which is supported by their intermediate phenotypes (i.e., *B. rupestris* margins with varying amounts of trichomes; ***SI Appendix*, Fig. S10**).

The last feral identification is that of *B. montana*, for which we find one sample as more closely related to wild taxa (224) and one more closely related to cultivars (222). The feral sample (222) is of unknown origin, but again the literature indicates that this may not be a surprising result. Many studies have previously indicated a close relationship between *B. montana* and *B. oleracea*. For example, Panda et al. (2003) concluded that *B. montana* may be a subspecies of *B. oleracea*, while Lannér et al. (1997) found that *B. montana* and *B. oleracea* clustered together using chloroplast data. Furthermore, several authors have suggested that some populations of *B. montana* were feral *B. oleracea* (Paolucci 1890; Onno 1933; Snogerup et al. 1990), which may be reflected in the overlapping ranges produced by our niche modeling of these two species (***SI Appendix*, Fig. S8)**. Therefore, in combination with results from previous studies, our results support that at least some *B. montana* populations are of feral origin.

Taken together, it is clear that the current taxonomy of *B. oleracea* and its wild relatives is confounded by gene flow between wild and cultivated populations, resulting in confusion between wild and feral lineages and obscuring the true evolutionary history of this species. Additionally, while there is much interest in crop improvement using CWRs (Meyer et al. 2012; Khoury et al. 2020), feral lineages offer another, potentially more direct route to reintroducing genetic diversity into cultivated populations, as gene flow is less likely to be impeded by barriers such as reproductive isolation (Mabry et al. 2021). These feral populations may also provide additional avenues to explore the evolutionary capacity for range expansion and phenotypic plasticity.

Post-domestication cultivar relationships. While our knowledge of the spread and diversification of *B. oleracea* crops after domestication is confounded by both the difficulties of identifying seeds of individual crop types and frequent introgression between crop types, we can infer some patterns using the species phylogeny. Like other studies (Cheng et al. 2016; Stansell et al. 2018), we find Chinese white kale sister to all other cultivars, representing the only Asian clade of crop types (**Fig. 3A**). While the spread of *B. oleracea* to eastern Asia is still undocumented archaeologically, recent pollen analysis has provided evidence for cultivation of other *Brassica* species, including *B. rapa*, in the Yangtze valley 3250 - 3350 BP, likely corresponding to movement across “Silk Road” trade routes (Zhang 2009). However, this only provides identification criteria, not archaeological evidence (Yang et al. 2018). A review of Chinese historical sources concluded that *B. oleracea* may have been introduced to China 1450 -1350 BP and had evolved into Chinese white kale in Southern China by the period of the Tang Dynasty (1350 - 1250 BP; Zhang 2009). Due to its position as sister to all other cultivars and as the only Asian *B. oleracea* crop type, as well as its annual growth habit, this taxon warrants additional study to understand its own unique domestication story.

The dispersal of *B. oleracea* by human translocation westward, ultimately to the Atlantic coast of Europe, appears to have established both regional feral populations and the variety of modern crop types. Archaeological evidence suggests that this process may have begun with Late Bronze Age seafaring (3000-3300 years ago), when the whole Mediterranean became linked in trade perhaps for the first time (Broodbank 2015), and continued to provide a corridor for introgression and varietal diversification through the Iron Age (up to 2000 years ago). Trade links along the Atlantic seaboard from North Africa and Iberia through Britain and Ireland are clearly indicated in archaeology (Cunliffe 2004), and are associated with the first peopling of the Canary Islands from the north, where walking stick kale is endemic. Notably, many cultivars do not form monophyletic groups in our sample level phylogeny, likely indicative of admixture between crop types. This is supported by previous findings that broccoli is paraphyletic (Song et al. 1988;

Stansell et al. 2018), as well as collards (Pelc et al. 2015), and by our findings that kale types such as tronchuda kale and perpetual kale are highly polyphyletic, suggesting that the kale morphotype has been selected for multiple times independently.

In conclusion, we confirm a single Eastern Mediterranean origin for *B. oleracea* and find *B. cretica* as the closest living wild relative. We highlight several feral samples that are not reflected by the current taxonomy but likely reflect important aspects of the domestication history for *B. oleracea*. Moving forward, it will be important to identify, collect, study, and preserve these feral samples as pools of allelic diversity, which may play an important role in future crop improvement, e.g. as a source of potential pest and pathogen resistance (Mithen et al. 1987; Mithen and Magrath 1992; Mohammed et al. 2010). In clarifying the evolutionary history of *B. oleracea* and its wild relatives, we hope to enable this model system for additional studies on evolutionary phenomena such as parallel selection, polyploidy, and ferality. Additionally, since many of these wild species are very narrow endemics and are valuable for both crop improvement and for nature conservation, their identification and preservation is urgent. We hope this study can serve as a steppingstone, as the work before us has, for those who, like Darwin was, are intrigued by this group of plants and wish to further its study.

MATERIALS AND METHODS

Taxon sampling. Samples from cultivars accounted for 188 of the 224 total samples with the remaining 36 samples included being previously identified wild relatives (***SI Appendix*, Table S2**). These include accessions from the United States Department of Agriculture, Agriculture Research Service (USDA-ARS) Plant Genetic Resources Unit (PGRU; 114 accessions), The Leibniz Institute of Plant Genetics and Crop Plant Research (IPK; 71 accessions), Universidad Politécnica de Madrid (UPM; 4 accessions), The Nordic Genetic Resource Centre (NordGen; 2 accessions), Gomez Campo Collection (2 accessions), John Innes Center (1 accession), doubled haploid lines (17 samples, some accessions sampled twice), or from the Pires' personal collection (13 accessions). Four replicates of each accession were grown from seed in a sterile growth chamber at the University of Missouri (MU; Columbia, MO) Bond Life Sciences Center in a randomized complete block design across two independent outgrowths. At the seventh leaf stage, leaf four was collected from each plant and immediately flash-frozen in liquid nitrogen for RNA extraction. Morphotype identity was validated in mature plants by growing all accessions twice over the span of two years (***SI Appendix*, Table S2**).

Whole-genome resequencing data for an additional four samples from Kioukis et al. (2020) of two varieties of *B. cretica* (var. *cretica* and var. *nivea*) was downloaded from the National Center for Biotechnology

Information (NCBI) Sequence Read Archive (SRA) to supplement our sampling of *B. cretica*. These samples are under the SRA accession as follows: A = SRR9331103, B = SRR9331104, C = SRR9331105, and D = SRR9331106). Samples of A and B are *B. cretica* var. *nivea* from mainland Greece and C and D are *B. cretica* var. *cretica*, one from the mainland (C) and one from the island of Crete (D).

RNA isolation and sequencing. RNA was isolated using the ThermoFisher Invitrogen PureLink RNA mini kit (Invitrogen, Carlsbad, CA, USA) followed by TruSeq library preparation (Illumina, San Diego, CA, USA) and sequencing on the NextSeq platform (Illumina, San Diego, CA, USA) for 2 X 75 bp reads. Library preparation and sequencing were performed through the MU DNA Core Facility. For eight flow cells, 24 samples were multiplexed and sequenced in a single flow-cell, followed by a ninth flow cell with 17 samples, and a tenth flow-cell with 16 samples.

Mapping and SNP calling. Short reads were mapped to the *B. oleracea* TO1000 genome (Chinese white kale; Parkin et al. 2014; release-41) by first using the STAR v. 2.5.2 (Dobin et al. 2013) two-pass alignment to identify splice junctions, which were then used in the second pass to improve mapping (Engström et al. 2013). The TO1000 genome of Chinese white kale was chosen due to wild relatives having a more kale-like phenotype and its placement as sister to the other cultivars in recent studies (Cheng et al. 2016; Stansell et al. 2018). Mapped reads (BAM format) were then processed following the GATK v. 3.8 best practices for RNA-seq reads (McKenna et al. 2010; Van der Auwera et al. 2013; Poplin et al. 2017). To ensure that reads were mapping correctly, the GATK 'Split'N'Trim' function was used to split reads into exon segments and trim any overhanging reads in intron segments. In total, 7,564,168 variants were called before any filtering was performed. The resulting variants were filtered to exclude those with a Fisher strand (FS) value greater than 30 and quality depth (QD) less than 2.0. The remaining 879,865 chromosomal variants were then filtered using vcftools v. 0.1.17 (Danecek et al. 2011) to exclude sites with greater than 60% missing data (*--max-missing 0.4*), sites with mean sample depth values less than 5 (*--min-meanDP 5*), and indels (*--remove-indels*;) resulting in a total of 103,525 SNPs. Finally, SNPs were filtered for linkage disequilibrium (LD) using using PLINK v. 1.90 with a window size of 80 Kb, or about two times the estimated length for 80% LD decay (Cheng et al. 2016), a step size of 5 kb, and a variance inflation factor of 2 (*--indep 80kb 5 2*; Purcell et al. 2007), for a final dataset of 36,750 SNPs. The four *B. cretica* genome resequencing samples (Kioukis et al. 2020) were also mapped to the *B. oleracea* TO1000 genome (Chinese white kale; Parkin et al. 2014; release-41), using BWA (Li and Durbin 2009).

Phylogenetic and Introgression Inference. To test how the different populations are related to one another and which wild relative is most closely related to the cultivated types, we used three different phylogenetic

programs; SNPhylo v. 20160204 (Lee et al. 2014) to assess individual sample relationships, IQ-Tree v. 1.6 (Nguyen et al. 2015) to test species level relationships, and TreeMix v. 1.13 (Pickrell and Pritchard 2012) to assess introgression. For SNPhylo (Lee et al. 2014), we ran analyses using an r^2 cutoff of 0.1 for LD, minor allele frequency ≥ 0.01, proportion of missing sites ≤ 0.4, 1000 bootstrap replicates, and rooted with sample 238 (*B. villosa*). For IQ-Tree, we used the Polymorphism-aware phylogenetic Models (PoMo) software (Schrempf et al. 2016; *-m GTR+P*) to perform phylogenetic comparisons using population genetic data, using 1000 bootstrap replicates via the ultrafast bootstrap approximation method (Hoang et al. 2018) and *B. villosa* to root the tree. For our IQ-Tree analysis, we subsampled data to include only those samples which were recovered as monophyletic in our SNPhylo tree (***SI Appendix*, Table S2; samples with asterisks).** To test both the topology of relationships and for gene flow between populations, we used TreeMix with the following parameters: no sample size correction (*-noss*), rooted with *B. villosa* (*-root villosa*), bootstrapping over blocks of 500 SNPs (*-bootstrap -k 500*), and to incorporate between 2-10 migration events (*-m*). TreeMix (Pickrell and Pritchard 2012) was run with samples of *B. cretica*, *B. incana*, *B. montana*, and *B. oleracea* as individuals, but used samples found in WildC-2, cultivars, and wild relatives as populations. Four-population (*f4*) tests for treeness (Reich et al. 2009; Pickrell and Pritchard 2012) were used to test the support of the inferred migration edges from Treemix (Pickrell and Pritchard 2012) via the fourpop method.

Population Structure and Variation. To test ancestry proportions and identify the likely genetic structure of described populations we used fastSTRUCTURE v. 1.0 (Raj et al. 2014). We tested K values from 2 to 8 using default convergence criteria and priors followed by the *chooseK.py* script to determine the appropriate number of model components that best explain structure in the dataset.

ANGSD v. 0.925 (Korneliussen et al. 2014) was used to calculate genotype likelihoods for all samples, plus the four additional *B. cretica* samples from Kioukis et al. (2020), using the parameters *-doGlf 2 -doMajorMinor 1 -doMaf 2 -minMapQ 30 -SNP_pval 1e-6,* followed by analysis with PCAangsd v. 0.97 (Meisner and Albrechtsen 2018) to visualize population structure, estimate allele frequencies, and calculate individual inbreeding coefficients using the parameters *-admix -selection 1 -inbreed 2*.

Clustering based on Expression Profiles. First, Salmon v. 1.2.1 (Patro et al. 2015) was used to acquire transcript abundances for each sample and the estimated number of reads originating from transcripts. The input for expression profile analysis was prepared using tximport (Soneson et al. 2015) with design = ~plantout + cultivar type. Correction for library size (*estimateSizeFactors*) and variance-stabilizing transformation (*vst*) was performed in DESeq2 v. 1.28.1 (Love et al. 2014). To test for clustering based on

expression profiles, we ran a PCA on the normalized expression values and performed clustering based on Euclidean distance using the 'prcomp' and 'hclust' functions, respectively, in the 'stats' v. 3.6.2 package for R v. 3.6.0 (R Core Team 2018). To assess networks of genes driving differences observed in the PCA, we used WGCNA v. 1.68 (Langfelder and Horvath 2008). Following (Zhang and Horvath 2005), we found that a soft-thresholding power of nine was best as it was the lowest power that satisfied the approximate scale-free topology criterion, resulting in 48 modules of genes.

To determine biological processes which were overrepresented in the resulting modules, *Arabidopsis thaliana* orthologs of *B. oleracea* were determined using both synteny and BLAST. Synteny-based annotations were extracted from Table S7 in (Parkin et al. 2014) while the BLAST annotation was performed using blastn in BLAST v. 2.10.0+ (Camacho et al. 2009). The *B. oleracea* CDS database was downloaded from https://plants.ensembl.org/Brassica_oleracea/Info/Index, and the *A. thaliana* CDS database from Araport11_genes.201606.cds.fasta from https://www.arabidopsis.org/. The blastn parameters were *-evalue 1E-6 -max_target_seqs 1*. Genes determined using synteny were then used to perform a GO analysis through PANTHER v. 16.0 (Mi et al. 2021). ANOVAs were used to test for differences in transcript abundance among cultivars, ferals (including WildC-2), and wild relatives by using the 'aov' function in R v. 3.6.0 (R Core Team 2018) followed by multiple comparisons with Tukey's HSD using the function 'glht'.

Environmental niche modeling. We compiled occurrence records for wild relatives from the Global Biodiversity Information Facility (GBIF, www.gbif.org) data portal and data from (Snogerup et al. 1990). From the GBIF data, we omitted records that were duplicated, lacked location data and/or vouchers, were collected from the grounds of botanical gardens, and that were clearly outside of the native range. From the Snogerup et al. (1990) data, we omitted records that could not be georeferenced to <5km spatial uncertainty. Populations of *B. cretica* in Lebanon and Israel and of *B. incana* in Crimea are thought to be likely early human introductions (Snogerup et al. 1990) and records from these areas were omitted. Occurrences above 1200m altitude were also omitted, as these species rarely occur above 1000m and observations above these altitudes may represent anthropogenic dispersals to disturbed areas or misidentifications. To minimize sampling bias due to clustered observations (Beck et al. 2014; Boria et al. 2014), we thinned the filtered occurrences to records greater than or equal to 10km apart using the 'spThin' package in R (Aiello-Lammens et al. 2015). After filtering and thinning, 172 records remained for *B. cretica*, 65 for *B. incana*, 57 for *B. insularis*, 101 *B. montana*, 15 for *B. villosa*, and 7 and 6 for the narrow endemics *B. macrocarpa* and *B. hilarionis* respectively. Next, we obtained rasters for 19 bioclimatic variables at 2.5 minutes resolution based on contemporary climate data from WorldClim v. 2.0 (Fick and Hijmans 2017) and rasters

for 19 bioclimatic variables at 2.5 minutes resolution based on late-Holocene climate projections using data derived from PaleoClim (Fordham et al. 2017; Brown et al. 2018). Rasters were clipped using QGIS v. 3.83 (Open Source Geospatial Foundation Project) to constrain the geographical background to windows slightly larger than the area circumscribed by contemporary observational data (Phillips et al. 2009; Acevedo et al. 2012). While it is common practice to eliminate collinear environmental variables to avoid overfitting (Braunisch et al. 2013), recent simulations have shown that removing highly collinear variables has an insignificant impact on maximum entropy model performance (Feng et al. 2019) so all original variables were included. Projections for late-Holocene habitat suitability were generated using MaxEnt v. 3.4.1 (Phillips et al. 2017). Linear, quadratic, product, and hinge features and jackknife resampling were used to measure variable importance. Relative model performance was evaluated with the adjusted area under receiver operating characteristic (ROC) curve (AUC; DeLong et al. 1988). While optimal performance cannot be determined with this approach using presence-only data, relative performance can still be assessed (Phillips et al. 2006).

DATA AVAILABILITY

The sequences reported in this paper have been deposited in the Sequence Read Archive database (accession no. PRJNA544934; https://www.ncbi.nlm.nih.gov/bioproject/PRJNA544934). The resulting transcript abundances and VCF file are deposited at DRYAD (https://doi.org/10.5061/dryad.5mkkwh763). Scripts used can be found on github (https://github.com/mmabry/Brassica-oleracea-Population-and-Phylogenetics).

AUTHOR CONTRIBUTIONS

MEM, SDTH, ACM, HA, PPE, JDM, DACP, GRT, CJS, GB, JL, DQF, TB, RGA, JED, MAG, and JCP designed the project. MEM, EYG, HA, and SDTH grew plants and collected tissue. MEM and EYG extracted and isolated RNA. MEM analyzed the genetic data. ACM produced the species distribution models. CS, DQF, and RGA researched archeology and written data. SDTH, HA, and JED assisted with processing and analyzing the data. MEM wrote the original manuscript. SDTH, ACM, HA, PPE, JDM, DACP, GRT, CJS, GB, JL, DQF, TB, RGA, JED, MAG, and JCP provided critical feedback on manuscript drafts.

ACKNOWLEDGMENTS

We thank Drs. Bob Schnabel, Troy Rowan, Harly Durbin, and Paul Blischak for their assistance with computational analyses, the Mizzou DNA core, Nathan Bivens, Ming-Yi Zhou, and Karen Bromert, for their assistance in getting quality data for sequencing, and our computing resources, specifically the

Research Computing Support Services (RCSS) and Informatics Research Core Facility (IRCF) at the University of Missouri. We are also grateful to Andi Kur for providing botanical illustrations of *B. oleracea* cultivar types and Elizabeth Gjieli, the Geographical Information Manager at the New York Botanical Garden GIS Laboratory for providing the wild species range map. We thank Sarah Unruh for valuable feedback on early versions of this manuscript, Dr. Jeff Ross-Ibarra for his help with interpreting admixture statistics, and two anonymous reviewers whose comments were extremely helpful in improving the manuscript. Finally, we thank our funding USDA-ARS Project No. 8060-21000-024-00D and the National Science Foundation Postdoctoral Fellowship in Biology (Award No. 1711347, S.T-H).

TABLE LEGENDS

Table 1. Wild species which have been proposed as progenitor species for *B. oleracea* crop types. Specific location is included in parentheses if indicated by the author. *Brassica oleracea* sometimes referred to as *B. oleracea* var. *sylvestris*.

FIGURE LEGENDS

Figure 1: Demographics and population structure for 224 samples of cultivated Brassica oleracea (n=188) and wild C genome species (n = 36). (Left) Individual sample phylogeny with putatively wild samples labeled in bold and black dots indicating bootstrap values less than 70%. (Middle) Ancestry proportions for K = 2 to K = 5 as inferred from fastSTRUCTURE; K=3 maximizes marginal likelihood (+) and K=5 best explains structure in the data (++). (Right) Monophyletic clades indicated by a solid line, largest cluster of paraphyletic groups indicated by dashed lined. Illustrations of corresponding crop types by Andi Kur.

Figure 2: Principal Component Analysis (PCA) of SNPs and expression profiles. A) Genetic variation PCA of PC1 vs PC2, **B)** Genetic variation PC2 vs PC3, and **C)** Expression profile PCA for PC1 vs PC2 of wild and cultivar samples. Triangles = wild samples, circles = cultivars. Triangles with black outlines = WildC-2 samples with species identification indicated by color. Wild-collected *B. cretica* samples from (Kioukis et al. 2020) indicated by asterisks, labeled as SRA.

Figure 3: Species tree with current distribution and historical environmental niche modeling. **A)** Species tree of wild and cultivar samples. Bootstrap support indicated above branches. **B)** Current species distribution of wild relatives. **C)** Suitable habitat for *B. cretica* and **D)** *B. hilarionis* during the late-Holocene. Map provided by Elizabeth Gjieli, the Geographical Information Manager at the New York Botanical Garden GIS Laboratory.

Figure 4: Inferred Admixture events. **A)** Phylogeny five migrations labeled a-f. **B)** Corresponding four-population tests for treeness.

SUPPLEMENTARY MATERIAL

Table S1: *Brassica oleracea* crop types with common name, species name, Kew cultivar group, other used names, and phenotypic descriptions modified from [1]Bailey (1922) or [2]http://www.missouribotanicalgarden.org/ (last accessed 5/18/21). Illustrations by Andi Kur.

Table S2: Sample information with species, variety, cultivar, germplasm, accession, sample #, location of original collection (year collected), and SRA #. Asterisks (*) next to sample # indicate those samples that

were recovered as monophyletic and used in species tree reconstruction. USDA-ARS PGRU = United States Department of Agriculture, Agriculture Research Service, Plant Genetic Resources Unit, IPK = The Leibniz Institute of Plant Genetics and Crop Plant Research, UPM = Universidad Politécnica de Madrid, NordGen = The Nordic Genetic Resource Centre, GC = Gomez Campo Collection, Innes = John Innes Center, DH = doubled haploid lines, Pires = J.C. Pires personal collection.

Table S3: WGCNA predicted gene modules with number of genes in each module, the number of annotated *Arabidopsis thaliana* genes using blast and synteny, and the percent of syntenic genes represented in the module. Module number with asterisks (*) represent the five modules with the largest percent of syntenic genes in the module.

Table S4: Top five WGCNA modules with largest percent of syntenic genes represented in the module with corresponding annotated GO biological process with the largest fold enrichment, and p-value.

Table S5: P-value results from ANOVA tests for differences in transcriptome abundance (Transcripts Per Million; TPM) between wild relatives, feral samples, and cultivars with Post-hoc Tukey Tests for top five WGCNA modules with largest syntenic overlap with *Arabidopsis thaliana*. Feral = *B. cretica* (196, 199); *B. incana* (204, 207); *B. montana* (222); and wild *B. oleracea* (175, 176, 177) + WildC-2 (*B. incana* - 205, 208, 209; *B. villosa* - 233, and *B. cretica* - 195). Significance indicated as * < 0.05, ** < 0.01, *** < 0.001.

Table S6: Archaeological *Brassica* reports from Europe and the Eastern Mediterranean.

Table S7: Literary and artistic sources covering the Classical Greek, Roman, and medieval and post-medieval sources.

Table S8: Area under the receiver operating characteristic curve (AUC) values for Maxent environmental niche model runs for putative wild relatives of Brassica oleracea crops.

Figure S1: Mapping percentage of unique reads for **A)** wild and cultivated samples, **B)** cultivar groups, and **C)** sequencing lane.

Figure S2: Demographics and population structure for 224 samples of cultivated Brassica oleracea (n=188) and wild C genome species (n = 36). (Left) Individual sample phylogeny with putatively wild samples labeled in bold and black dots indicating bootstrap values less than 70%. (Middle) Ancestry proportions for K = 2 to K = 8 as inferred from fastSTRUCTURE; K=3 maximizes marginal likelihood (+) and K=5 best explains structure in the data (++). (Right) Monophyletic clades indicated by a solid line, largest cluster of paraphyletic groups indicated by dashed lined. Illustrations of corresponding crop types by Andi Kur.

Figure S3: PCAs of SNPs and Expression profiles. **A)** Genetic variation PCA of PC3 vs PC4, **B)** Expression profile PCA for PC2 vs PC3, and **C)** Expression profile PCA for PC3 vs PC4 of wild and cultivar samples. **D - F)** PCAs of Genetic variation without *B. cretica* samples from (SRA; Kioukis et al. 2020). Triangles = wild samples, circles = cultivars. Triangles with black outlines = WildC-2 samples with species identification indicated by color. Wild-collected *B. cretica* samples from (Kioukis et al. 2020) indicated by asterisks, labeled as SRA.

Figure S4: Inbreeding coefficients for wild, cultivar, and feral samples. Wild-collected *B. cretica* samples from (Kioukis et al. 2020) indicated by asterisks.

Figure S5: *Brassica oleracea* cultivars and wild relative dendrogram based on expression profiles. Wild species indicated by color below. Cultivars in grey below. WildC-2 indicated by black outlines for the corresponding bar chart below. Sample names in bold = putatively wild samples.

Figure S6: Comparisons of Transcript abundance between wild relatives, ferals, cultivars for top 5 conserved modules between *Brassica oleracea* and *Arabidopsis thaliana*. **A)** module 7, **B)** module 13, **C)** module 30, **D)** module 31, **E)** module 34.

Figure S7: TreeMix analysis of wild, cultivar, and feral samples. **A-F)** Phylogeny with 0-5 migrations indicated.

Figure S8: The current wild relatives distribution and modeled late-Holocene suitable habitat. Middle - current species distribution of wild relatives. Suitable habitat for **A)** *B. montana*, **B)** *B. insularis*, **C)** *B. macrocarpa*, **D)** *B. rupestris*, **E)** *B. villosa*, **F)** *B. incana*, and **G)** *B. oleracea*, during the late-Holocene

Figure S9: Leaf scans of feral samples. Leaf used for RNA collection with biological replicate indicated.

Figure S10: Leaf scans of WildC-2 samples. Leaf used for RNA collection with biological replicates indicated.

REFERENCES

Acevedo P, Jiménez-Valverde A, Lobo JM, Real R. 2012. Delimiting the geographical background in species distribution modelling. *J. Biogeogr.* 39:1383–1390. Available from: http://dx.doi.org/10.1111/j.1365-2699.2012.02713.x

Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP. 2015. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* 38:541–545. Available from: http://dx.doi.org/10.1111/ecog.01132

Allaby R. 2010. Integrating the processes in the evolutionary system of domestication. *J. Exp. Bot.* 61:935–944. Available from: http://dx.doi.org/10.1093/jxb/erp382

Allender CJ, Allainguillaume J, Lynn J, King GJ. 2007. Simple sequence repeats reveal uneven distribution of genetic diversity in chloroplast genomes of Brassica oleracea L. and (n = 9) wild relatives. *Theor. Appl. Genet.* 114:609–618. Available from: http://dx.doi.org/10.1007/s00122-006-0461-5

Arias T, Pires JC. 2012. A fully resolved chloroplast phylogeny of the brassica crops and wild relatives (Brassicaceae: Brassiceae): Novel clades and potential taxonomic implications. *Taxon* 61:980–988. Available from: http://doi.wiley.com/10.1002/tax.615005

Bailey LH. 1930. The Cultivated Brassicas Second Paper. *Gentes Herbarum*.

Beck J, Böller M, Erhardt A, Schwanghart W. 2014. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecol. Inform.* 19:10–15. Available from: http://www.sciencedirect.com/science/article/pii/S1574954113001155

Beebe S, Toro Ch O, Gonza´lez AV, Chaco´n MI, Debouck DG. 1997. Wild-weed-crop complexes of common bean (Phaseolus vulgaris L., Fabaceae) in the Andes of Peru and Colombia, and their implications for conservation and breeding. *Genet. Resour. Crop Evol.* 44:73–91. Available from: https://doi.org/10.1023/A:1008621632680

Boria RA, Olson LE, Goodman SM, Anderson RP. 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecol. Modell.* 275:73–77. Available from: http://www.sciencedirect.com/science/article/pii/S0304380013005917

Braunisch V, Coppes J, Arlettaz R, Suchant R, Schmid H, Bollmann K. 2013. Selecting from correlated climate variables: a major source of uncertainty for predicting species distributions under climate change. *Ecography* 36:971–983. Available from: http://doi.wiley.com/10.1111/j.1600-0587.2013.00138.x

Broodbank C. 2015. The Making of the Middle Sea: A History of the Mediterranean from the Beginning to the Emergence of the Classical World. Thames & Hudson Available from: https://play.google.com/store/books/details?id=tjw-rgEACAAJ

Brown JL, Hill DJ, Dolan AM, Carnaval AC, Haywood AM. 2018. PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Sci Data* 5:180254. Available from: http://dx.doi.org/10.1038/sdata.2018.254

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. Available from: https://doi.org/10.1186/1471-2105-10-421

de Candolle A. 1855. Géographie botanique raisonnée ou exposition des faits principaux et des lois concernant la distribution géographique des plantes de l'époque actuelle. V. Masson Available from: https://play.google.com/store/books/details?id=TPETAAAAQAAJ

Cheng F, Sun R, Hou X, Zheng H, Zhang F, Zhang Y, Liu B, Liang J, Zhuang M, Liu Y, et al. 2016. Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in Brassica rapa and Brassica oleracea. *Nat. Genet.* 48:1218–1224. Available from: http://dx.doi.org/10.1038/ng.3634

Cunliffe B. 2004. Facing the Ocean: The Atlantic and Its Peoples, 8000 BC-AD 1500. Oxford University Press Available from: https://play.google.com/store/books/details?id=EkefQgAACAAJ

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. Available from: http://dx.doi.org/10.1093/bioinformatics/btr330

Darwin C, Gray A. 1868. The variation of animals and plants under domestication / by Charles Darwin ; authorized edition, with a preface by Asa Gray. Available from: http://dx.doi.org/10.5962/bhl.title.37659

DeLong ER, DeLong DM, Clarke-Pearson DL. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44:837–845. Available from: https://www.ncbi.nlm.nih.gov/pubmed/3203132

Dixon GR. 2006. Origins and diversity of Brassica and its relatives. *Vegetable brassicas and related crucifers* :1–33. Available from: https://www.cabdirect.org/cabdirect/abstract/20073006206

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21. Available from: http://dx.doi.org/10.1093/bioinformatics/bts635

Edh K, Widén B, Ceplitis A. 2007. Nuclear and chloroplast microsatellites reveal extreme population differentiation and limited gene flow in the Aegean endemic Brassica cretica (Brassicaceae). *Mol. Ecol.* 16:4972–4983. Available from: http://dx.doi.org/10.1111/j.1365-294X.2007.03585.x

Engström PG, Steijger T, Sipos B, Grant GR, Kahles A, Rätsch G, Goldman N, Hubbard TJ, Harrow J, Guigó R, et al. 2013. Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat. Methods* 10:1185–1191. Available from: http://dx.doi.org/10.1038/nmeth.2722

Evershed RP, Arnot KI, Collister J, Eglinton G, Charters S. 1994. Application of isotope ratio monitoring gas chromatography–mass spectrometry to the analysis of organic residues of archaeological origin. *Analyst* 119:909–914. Available from: http://dx.doi.org/10.1039/an9941900909

Evershed RP, Heron C, Charters S, Goad LJ. 1992. The survival of food residues: new methods of analysis, interpretation and application. In: Proceedings of the British Academy. Vol. 77. p. 2. Available from: https://www.thebritishacademy.ac.uk/pubs/proc/files/77p187.pdf

Farnham MW. 1996. Genetic variation among and within United States collard cultivars and landraces as determined by randomly amplified polymorphic DNA markers. *J. Am. Soc. Hortic. Sci.* 121:374–379. Available from: https://journals.ashs.org/jashs/view/journals/jashs/121/3/article-p374.xml

Feng X, Park DS, Liang Y, Pandey R, Papeş M. 2019. Collinearity in ecological niche modeling: Confusions and challenges. *Ecol. Evol.* 9:10365–10376. Available from: http://dx.doi.org/10.1002/ece3.5555

Fernie AR, Yan J. 2019. De Novo Domestication: An Alternative Route toward New Crops for the Future. *Mol. Plant* 12:615–631. Available from: http://dx.doi.org/10.1016/j.molp.2019.03.016

Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* . Available from: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5086

Fordham DA, Saltré F, Haythorne S, Wigley TML, Otto-Bliesner BL, Chan KC, Brook BW. 2017. PaleoView: a tool for generating continuous climate projections spanning the last 21 000 years at regional and global scales. *Ecography* 40:1348–1358. Available from: http://doi.wiley.com/10.1111/ecog.03031

Fuller DQ, Denham T, Arroyo-Kalin M, Lucas L, Stevens CJ, Qin L, Allaby RG, Purugganan MD. 2014. Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. *Proceedings of the National Academy of Sciences* 111:6147–6152. Available from: http://dx.doi.org/10.1073/pnas.1308937110

Gering E, Incorvaia D, Henriksen R, Conner J, Getty T, Wright D. 2019. Getting Back to Nature: Feralization in Animals and Plants. *Trends Ecol. Evol.* 34:1137–1151. Available from: http://dx.doi.org/10.1016/j.tree.2019.07.018

Gladis T, Hammer K. 2001. Nomenclatural notes on the Brassica oleracea-group. *Genet. Resour. Crop Evol.* 48:7–11. Available from: http://link.springer.com/10.1023/A:1011201118844

Gustafsson M, Bentzer B, Von Bothmer B, Snogerup. 1976. Meiosis in Greek Brassica of the oleracea group. Available from: https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=PASCAL7637010832

Heaney RK, Roger Fenwick G, Mithen RF, Lewis BG. 1987. Glucosinolates of wild and cultivated Brassica species. *Phytochemistry* 26:1969–1973. Available from: http://dx.doi.org/10.1016/s0031-9422(00)81740-9

Helm J. 1963. Morphologisch-taxonomische Gliederung der Kultursippen vonBrassica oleracea L. *Kulturpflanze* 11:92–210. Available from: http://dx.doi.org/10.1007/bf02136113

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35:518–522. Available from: http://dx.doi.org/10.1093/molbev/msx281

Hodgkin T. 1995. Cabbages, kales, etc. *Brassica oleracea*:76–82.

Hort A. 1916. Theophrastus: enquiry into plants. Cambridge (Mass.) and.

Kaniewski D, Van Campo E, Van Lerberghe K, Boiy T, Vansteenhuyse K, Jans G, Nys K, Weiss H, Morhange C, Otto T, et al. 2011. The Sea Peoples, from cuneiform tablets to carbon dating. *PLoS One* 6:e20232. Available from: http://dx.doi.org/10.1371/journal.pone.0020232

Khoury CK, Carver D, Greene SL, Williams KA, Achicanoy HA, Schori M, León B, Wiersema JH, Frances A. 2020. Crop wild relatives of the United States require urgent conservation action. *Proc. Natl. Acad. Sci. U. S. A.* 117:33351–33357. Available from: http://dx.doi.org/10.1073/pnas.2007029117

Kioukis A, Michalopoulou VA, Briers L, Pirintsos S, Studholme DJ, Pavlidis P, Sarris PF. 2020. Intraspecific diversification of the crop wild relative Brassica cretica Lam. using demographic model selection. *BMC Genomics* 21:48. Available from: http://dx.doi.org/10.1186/s12864-019-6439-x

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15:356. Available from: http://dx.doi.org/10.1186/s12859-014-0356-4

Kreuz A, Schäfer E. 2002. A new archaeobotanical database program. *Veg. Hist. Archaeobot.* 11:177–180. Available from: https://doi.org/10.1007/s003340200019

Kroll H. 2001. Literature on archaeological remains of cultivated plants (1999/2000). *Veg. Hist. Archaeobot.* 10:33–60. Available from: http://dx.doi.org/10.1007/pl00013368

Kroll H. 2005. Literature on archaeological remains of cultivated plants 1981--2004. *Available at archaeobotany. de/database. html. Accessed January* 2:2016.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. Available from: http://dx.doi.org/10.1186/1471-2105-9-559

Lannér C, Bryngelsson T, Gustafsson M. 1997. Relationships of wild Brassica species with chromosome number 2 n= 18, based on RFLP studies. *Genome* 40:302–308.

Lanner-Herrera C, Gustafeson M, Filt AS, Bryngelsson T. 1996. Diversity in natural populations of wild Brassica oleracea as estimated by isozyme and RAPD analysis. *Genet. Resour. Crop Evol.* 43:13–23. Available from: http://link.springer.com/10.1007/BF00126936

Lázaro A, Aguinagalde I. 1998. Genetic Diversity in Brassica oleracea L. (Cruciferae) and Wild Relatives (2 n= 18) using Isozymes. *Ann. Bot.* 82:821–828. Available from: https://academic.oup.com/aob/article-abstract/82/6/821/2587950

Lee T-H, Guo H, Wang X, Kim C, Paterson AH. 2014. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* 15:162. Available from: http://dx.doi.org/10.1186/1471-2164-15-162

Liddell HG, Scott R. 1940. A Greek-English Lexicon Perseus. United Kingdom: Oxford University Press

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. Available from: http://dx.doi.org/10.1093/bioinformatics/btp324

Li T, Yang X, Yu Y, Si X, Zhai X, Zhang H, Dong W, Gao C, Xu C. 2018. Domestication of wild tomato is accelerated by genome editing. *Nat. Biotechnol.* 36:1160–1163. Available from: http://dx.doi.org/10.1038/nbt.4273

Lizgunova TV. 1959. The history of botanical studies of the cabbage, Brassica oleracea L. *Bulletin of Applied Botany, Genetics and Plant Breeding* 32:37–70.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. Available from: http://dx.doi.org/10.1186/s13059-014-0550-8

Mabry ME, Rowan TN, Pires JC, Decker JE. 2021. Feralization: Confronting the Complexity of Domestication and Evolution. *Trends Genet.* . Available from: http://dx.doi.org/10.1016/j.tig.2021.01.005

Maggioni L, von Bothmer R, Poulsen G, Aloisi KH. 2020. Survey and genetic diversity of wild Brassica oleracea L. germplasm on the Atlantic coast of France. *Genet. Resour. Crop Evol.* 67:1853–1866. Available from: http://dx.doi.org/10.1007/s10722-020-00945-0

Maggioni L, von Bothmer R, Poulsen G, Branca F. 2010. Origin and Domestication of Cole Crops (Brassica oleracea L.): Linguistic and Literary Considerations1. *Econ. Bot.* 64:109–123. Available from: https://doi.org/10.1007/s12231-010-9115-2

Maggioni L, von Bothmer R, Poulsen G, Lipman E. 2018. Domestication, diversity and use of Brassica oleracea L., based on ancient Greek and Latin texts. *Genet. Resour. Crop Evol.* 65:137–159. Available from: http://dx.doi.org/10.1007/s10722-017-0516-2

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–1303. Available from: http://dx.doi.org/10.1101/gr.107524.110

Mei J, Li Q, Yang X, Qian L, Liu L, Yin J, Frauen M, Li J, Qian W. 2010. Genomic relationships between wild and cultivated Brassica oleracea L. with emphasis on the origination of cultivated crops. *Genet. Resour. Crop Evol.* 57:687–692. Available from: http://link.springer.com/10.1007/s10722-009-9504-5

Meisner J, Albrechtsen A. 2018. Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics* 210:719–731. Available from: http://dx.doi.org/10.1534/genetics.118.301336

Meyer RS, DuVal AE, Jensen HR. 2012. Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196:29–48. Available from: http://dx.doi.org/10.1111/j.1469-8137.2012.04253.x

Mi H, Ebert D, Muruganujan A, Mills C, Albou L-P, Mushayamaha T, Thomas PD. 2021. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* 49:D394–D403. Available from: http://dx.doi.org/10.1093/nar/gkaa1106

Mitchell ND. 1976. The status of Brassica oleracea L. subsp. oleracea (wild cabbage) in the British Isles. *Watsonia* 11:97–103. Available from: http://archive.bsbi.org.uk/Wats11p97.pdf

Mitchell ND, Richards AJ. 1979. Brassica Oleracea L. ssp. Oleracea (B. sylvestris (L.) Miller). *J. Ecol.* 67:1087–1096. Available from: http://www.jstor.org/stable/2259229

Mithen RF, Lewis BG, Heaney RK, Fenwick GR. 1987. Resistance of leaves of Brassica species to Leptosphaeria maculans. *Trans. Br. Mycol. Soc.* 88:525–531. Available from: http://www.sciencedirect.com/science/article/pii/S0007153687800360

Mithen RF, Magrath R. 1992. Glucosinolates and Resistance to Leptosphaeria maculans in Wild and Cultivated Brassica Species. *Plant Breed.* 108:60–68. Available from: http://dx.doi.org/10.1111/j.1439-0523.1992.tb00100.x

Mittell EA, Cobbold CA, Ijaz UZ, Kilbride EA, Moore KA, Mable BK. 2020. Feral populations of Brassica oleracea along Atlantic coasts in western Europe. *Ecol. Evol.* 10:11810–11825. Available from: http://dx.doi.org/10.1002/ece3.6821

Mohammed A, Addo-Quaye A, Asare-bediako E. 2010. Control of Diamond back Moth (Plutella xylostella) on Cabbage (Brassica oleracea var capitata) using Intercropping with Non-Host Crops“ E. Asare-Bediako,” AA Addo-Quaye and“ A. Mohammed” Department of Crop Science, University of Cape Coast, Cape Coast, Ghana. *Am. J. Food Technol.* 5:269–274. Available from: https://www.researchgate.net/profile/Elvis_Asare-Bediako2/publication/43047405_Control_of_Diamondback_Moth_Plutella_xylostella_on_Cabbage_Brassica_oleracea_var_capitata_using_Intercropping_with_Non-Host_Crops/links/551968db0cf2d241f3566ca3/Control-of-Diamondback-Moth-Plutella-xylostella-on-Cabbage-Brassica-oleracea-var-capitata-using-Intercropping-with-Non-Host-Crops.pdf

Murray MA. 2000. Fruits, vegetables, pulses and condiments. In: Cambridge University Press.

Neutrofal F. 1927. Zytologische Studien über die Kulturrassen von Brassica oleracea. Oesterr. *Bot. Z.* 76:105–115.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32:268–274. Available from: http://dx.doi.org/10.1093/molbev/msu300

Onno M. 1933. Die Wildformen aus dem Verwandtschaftskreis „Brassica oleracea L.“. *Österreichische Botanische Zeitschrift* 82:309–334. Available from: http://www.jstor.org/stable/43335985

Open Source Geospatial Foundation Project. QGIS Geographic Information System. *QGIS.org* . Available from: http://qgis.org

Panda S, Martín J, Aguinagalde I. 2003. Chloroplast and nuclear DNA studies in a few members of the Brassica oleracea L. group using PCR-RFLP and ISSR-PCR markers: a population genetic analysis. *Theor. Appl. Genet.* 106:1122–1128. Available from: https://link.springer.com/article/10.1007/s00122-002-1134-7

Paolucci L. 1890. Flora marchigiana. Stab. tipo-lit. Federici Available from: https://play.google.com/store/books/details?id=KjvZ5tPzbdsC

Parkin IAP, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE, Town CD, Nixon J, Krishnakumar V, Bidwell SL, et al. 2014. Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. *Genome Biol.* 15:R77. Available from: http://dx.doi.org/10.1186/gb-2014-15-6-r77

Patro R, Duggal G, Kingsford C. 2015. Salmon: Accurate, Versatile and Ultrafast Quantification from RNA-seq Data using Lightweight-Alignment. *Cold Spring Harbor Laboratory* :021592. Available from: https://www.biorxiv.org/content/10.1101/021592v1.abstract

Pelc SE, Couillard DM, Stansell ZJ, Farnham MW. 2015. Genetic Diversity and Population Structure of Collard Landraces and their Relationship to Other Brassica oleracea Crops. *Plant Genome* 8:eplantgenome2015.04.0023. Available from: http://dx.doi.org/10.3835/plantgenome2015.04.0023

Phillips SJ, Anderson RP, Dudík M, Schapire RE, Blair ME. 2017. Opening the black box: an open-source release of Maxent. *Ecography* 40:887–893. Available from: http://dx.doi.org/10.1111/ecog.03049

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* 190:231–259. Available from: http://www.sciencedirect.com/science/article/pii/S030438000500267X

Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* 19:181–197. Available from: https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/07-2153.1?casa_token=r_0Rk0ZrCD4AAAAA:KDoFWRFjcNWbBAPnsHO88FRMtU1Za-r7cxjfKnmtVISaeHnpuVE2BQL08FyGysXudTPp9VtSB-7J_1cF

Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8:e1002967. Available from: http://dx.doi.org/10.1371/journal.pgen.1002967

Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, et al. 2017. Scaling accurate genetic variant discovery to tens of thousands of samples. *Cold Spring Harbor Laboratory* :201178. Available from: https://www.biorxiv.org/content/10.1101/201178v2.abstract

Prohens J, Gramazio P, Plazas M, Dempewolf H, Kilian B, Díez MJ, Fita A, Herraiz FJ, Rodríguez-Burruezo A, Soler S, et al. 2017. Introgressiomics: a new approach for using crop wild relatives in breeding for adaptation to climate change. *Euphytica* 213. Available from: http://dx.doi.org/10.1007/s10681-017-1938-9

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–575. Available from: http://dx.doi.org/10.1086/519795

Raj A, Stephens M, Pritchard JK. 2014. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197:573–589. Available from:

http://dx.doi.org/10.1534/genetics.114.164350

Raybould AF, Mogg RJ, Clarke RT, Gliddon CJ, Gray AJ. 1999. Variation and population structure at microsatellite and isozyme loci in wild cabbage (Brassica oleracea L.) in Dorset (UK). *Genet. Resour. Crop Evol.* 46:351–360. Available from: https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/article/10.1023/A:1008658630440&casa_token=2SJA3eCXNa0AAAAA:vkaqsajxBe_TRfJOxuZQ9Zku45gi07Toq5PFbjoz1_VuKDSiEL6-xUE3bzL8qmGHH-fu4fe7SwsDf5VMKQ

R Core Team. 2018. R: A language and environment for statistical computing. Available from: https://www.R-project.org/.

Reich D, Thangaraj K, Patterson N, Price AL, Singh L. 2009. Reconstructing Indian population history. *Nature* 461:489–494. Available from: https://doi.org/10.1038/nature08365

Reihl S. 2014. Archaeobotanical database of Eastern Mediterranean and Near Eastern sites. Available from: http://www.ademnes.de/

Schiemann E. 1932. Entstehung der kulturpflanzen. Borntraeger

Schmidl A, Oeggl K. 2005. Subsistence strategies of two Bronze Age hill-top settlements in the eastern Alps—Friaga/Bartholomäberg (Vorarlberg, Austria) and Ganglegg/Schluderns (South Tyrol, Italy). *Veg. Hist. Archaeobot.* 14:303–312. Available from: https://doi.org/10.1007/s00334-005-0088-5

Schrempf D, Minh BQ, De Maio N, von Haeseler A, Kosiol C. 2016. Reversible polymorphism-aware phylogenetic models and their application to tree inference. *J. Theor. Biol.* 407:362–370. Available from: http://dx.doi.org/10.1016/j.jtbi.2016.07.042

Schulz OE. 1936. Cruciferae. In: Engler A, Harms H, editors. Die natuÈrlichen P¯anzenfamilien, 2nd edn. p. 176: 227±658.

Shyam P, Wu XM, Bhat SR. 2012. History, evolution, and domestication of Brassica crops. *Plant Breed. Rev.* . Available from: https://www.cabdirect.org/cabdirect/abstract/20123077074

Snogerup S. 1980. The wild forms of the Brassica oleracea group (2n= 18) and their possible relations to the cultivated ones. *Brassica crops and wild allies. [I].* :121–132. Available from: https://www.cabdirect.org/cabdirect/abstract/19801689308

Snogerup S, Gustafsson M, Von Bothmer R. 1990. Brassica sect. Brassica (Brassicaceae) I. Taxonomy and Variation. *Willdenowia* 19:271–365. Available from: http://www.jstor.org/stable/3996645

Soneson C, Love MI, Robinson MD. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* 4:1521. Available from: http://dx.doi.org/10.12688/f1000research.7563.2

Song KM, Osborn TC, Williams PH. 1988. Brassica taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). *Theor. Appl. Genet.* 75:784–794.

Song K, Osborn TC, Williams PH. 1990. Brassica taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). *Theor. Appl. Genet.* 79:497–506. Available from: https://doi.org/10.1007/BF00226159

Stansell Z, Hyma K, Fresnedo-Ramírez J, Sun Q, Mitchell S, Björkman T, Hua J. 2018. Genotyping-by-

sequencing of Brassica oleracea vegetables reveals unique phylogenetic patterns, population structure and domestication footprints. *Hortic Res* 5:38. Available from: http://dx.doi.org/10.1038/s41438-018-0040-3

Swarup V, Brahmi P. 2005. Cole crops. *Plant Genetic Resources: Horticultural Crops* :75–88. Available from: https://books.google.com/books?hl=en&lr=&id=jPc5k5w3GjoC&oi=fnd&pg=PA75&dq=Swarup+Brahmi+2005+Brassica&ots=u-Y3MqbHSC&sig=1EzMSXN8KerlW1-2Gb4WXWwPjeE

the Elder. P, Rackham H. 1950. Natural History with an English Translation: Vol 5: Libri XVII-XIX. Harvard University Press

Tomlinson P, Hall A. 1996. A review of the archaeological evidence for food plants from the British Isles: an example of the use of the Archaeobotanical Computer Database (ABCD). *Internet Archaeology* 1.

Totelin LMV. 2009. Hippocratic Recipes: Oral and Written Transmission of Pharmacological Knowledge in Fifth- And Fourth-Century Greece. BRILL Available from: https://play.google.com/store/books/details?id=q-TiIMgrLCgC

Turner-Hissong SD, Mabry ME, Beissinger TM, Ross-Ibarra J, Chris Pires J. 2020. Evolutionary insights into plant breeding. *Curr. Opin. Plant Biol.* 54:93–100. Available from: http://dx.doi.org/10.1016/j.pbi.2020.03.003

Tutin TG. 1964. Flora Europaea: Lycopodiaceae to Platanaceae. University Press Available from: https://play.google.com/store/books/details?id=3IHryvX9G-UC

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* 43:11–10. Available from: https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/0471250953.bi1110s43

Van der Veen M. 2011. Consumption, Trade and Innovation. Africa Magna Verlag Available from: https://play.google.com/store/books/details?id=wruovzCWdBsC

Vavilov. 1926. On the Origin of Cultivated Plants. *Food History: Critical and Primary Sources* . Available from: http://dx.doi.org/10.5040/9781474220095.ch-004

Vavilov. 1951. The Origin, Variation, Immunity and Breeding of Cultivated Plants. *Soil Science* 72:482. Available from: http://dx.doi.org/10.1097/00010694-195112000-00018

Wang H, Vieira FG, Crawford JE, Chu C, Nielsen R. 2017. Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice. *Genome Res.* 27:1029–1038. Available from: http://dx.doi.org/10.1101/gr.204800.116

West ML. 2011. Studies in Greek Elegy and Iambus. Walter de Gruyter Available from: https://play.google.com/store/books/details?id=tBPdEL189PsC

Widén B, Andersson S, Rao G-Y, Widén M. 2002. Population divergence of genetic (co)variance matrices in a subdivided plant species,Brassica cretica: G matrix variation inBrassica cretica. *J. Evol. Biol.* 15:961–970. Available from: http://doi.wiley.com/10.1046/j.1420-9101.2002.00465.x

Yang S, Zheng Z, Mao L, Li J, Chen B. 2018. Pollen morphology of selected crop plants from southern

China and testing pollen morphological data in an archaeobotanical study. *Veg. Hist. Archaeobot.* 27:781–799. Available from: http://link.springer.com/10.1007/s00334-018-0696-5

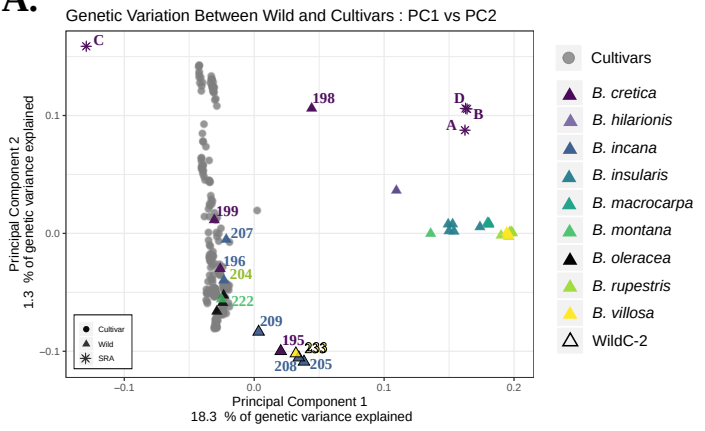
Yonge CD. 1854. Athenaeus of Naucratis. *The deipnosophists, or, Banquet of the learned of Athenæus. London: Henry G. Bohn*.

Zhang B, Horvath S. 2005. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4:Article17. Available from: http://dx.doi.org/10.2202/1544-6115.1128

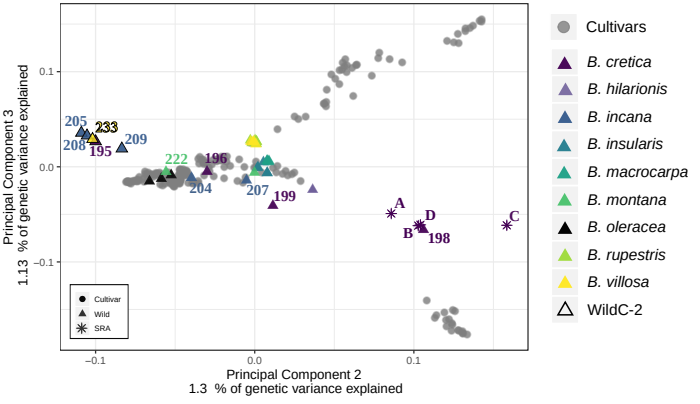
Zhang P. 2009. Studies on the origin of Brassica alboglabra Bailey. *China Vegetables* :62–65. Available from: https://www.cabdirect.org/cabdirect/abstract/20103012579

K=2
K=3++
K=4
K=5+
cabbage
savoy
collards
Brussels sprouts
marrow cabbage
perpetual kale
curly kale
kohlrabi
cauliflower
broccoli
romanesco
Chinese white kale
tronchuda kale
WildC-2
lacinato kale

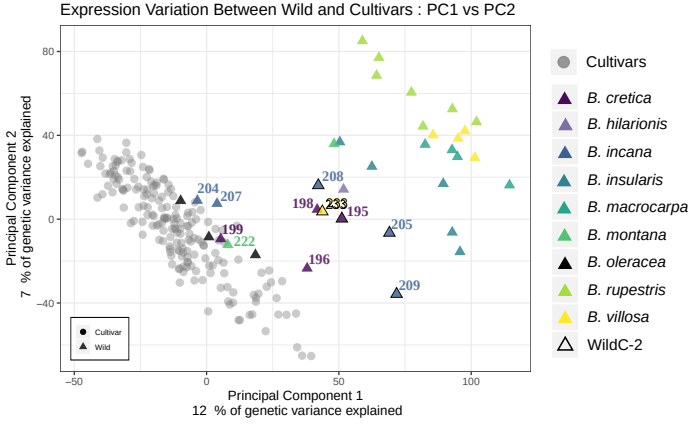
A.


Genetic Variation Between Wild and Cultivars : PC1 vs PC2
Principal Component 2
1.3 % of genetic variance explained
Principal Component 1
18.3 % of genetic variance explained
Cultivar
Wild
SRA
C
198
D
B
A
199
207
196
204
222
209
195
233
208
205
Cultivars
B. cretica
B. hilarionis
B. incana
B. insularis
B. macrocarpa
B. montana
B. oleracea
B. rupestris
B. villosa
WildC-2

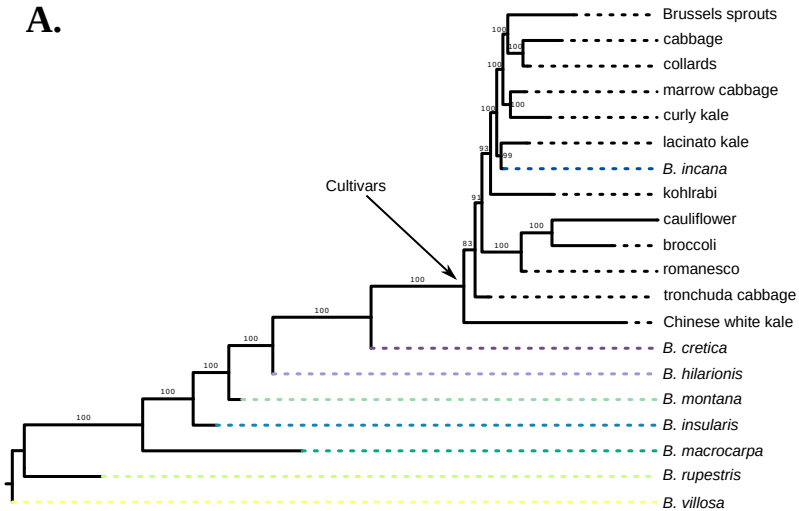
B.


Genetic Variation Between Wild and Cultivars : PC2 vs PC3
Principal Component 3
1.13 % of genetic variance explained
Principal Component 2
1.3 % of genetic variance explained
Cultivar
Wild
SRA
205
233
208
195
209
222
196
204
207
199
A
D
B
198
C
Cultivars
B. cretica
B. hilarionis
B. incana
B. insularis
B. macrocarpa
B. montana
B. oleracea
B. rupestris
B. villosa
WildC-2

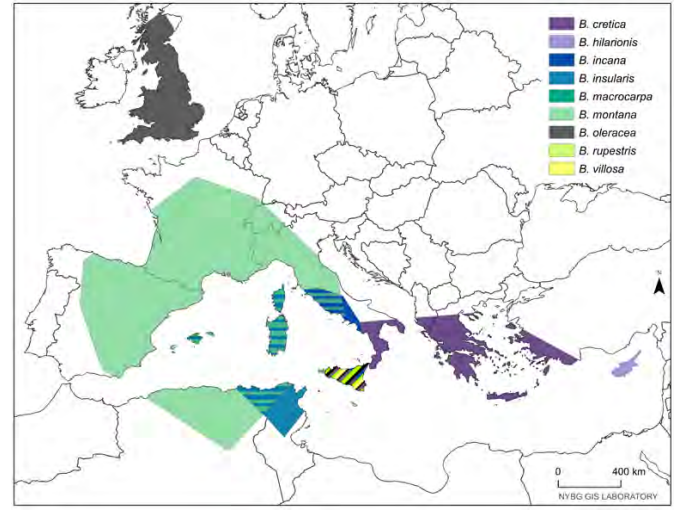
C.


Expression Variation Between Wild and Cultivars : PC1 vs PC2
Principal Component 2
7 % of genetic variance explained
Principal Component 1
12 % of genetic variance explained
Cultivar
Wild
204
207
208
198
233
195
205
199
222
196
209
Cultivars
B. cretica
B. hilarionis
B. incana
B. insularis
B. macrocarpa
B. montana
B. oleracea
B. rupestris
B. villosa
WildC-2

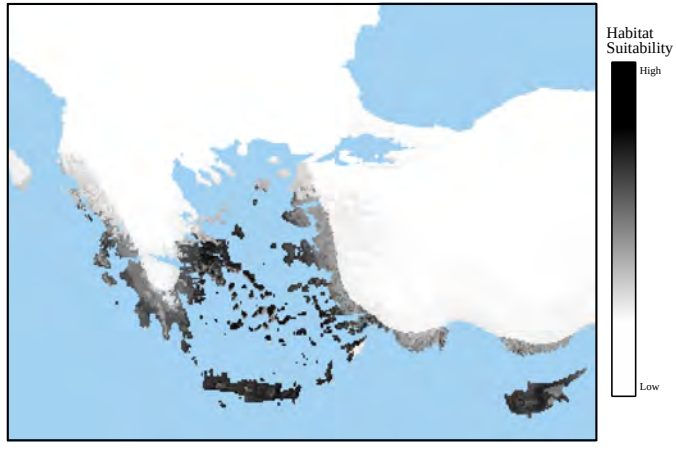
A.


Brussels sprouts
cabbage
collards
marrow cabbage
curly kale
lacinato kale
B. incana
kohlrabi
cauliflower
broccoli
romanesco
tronchuda cabbage
Chinese white kale
B. cretica
B. hilarionis
B. montana
B. insularis
B. macrocarpa
B. rupestris
B. villosa
Cultivars

B.

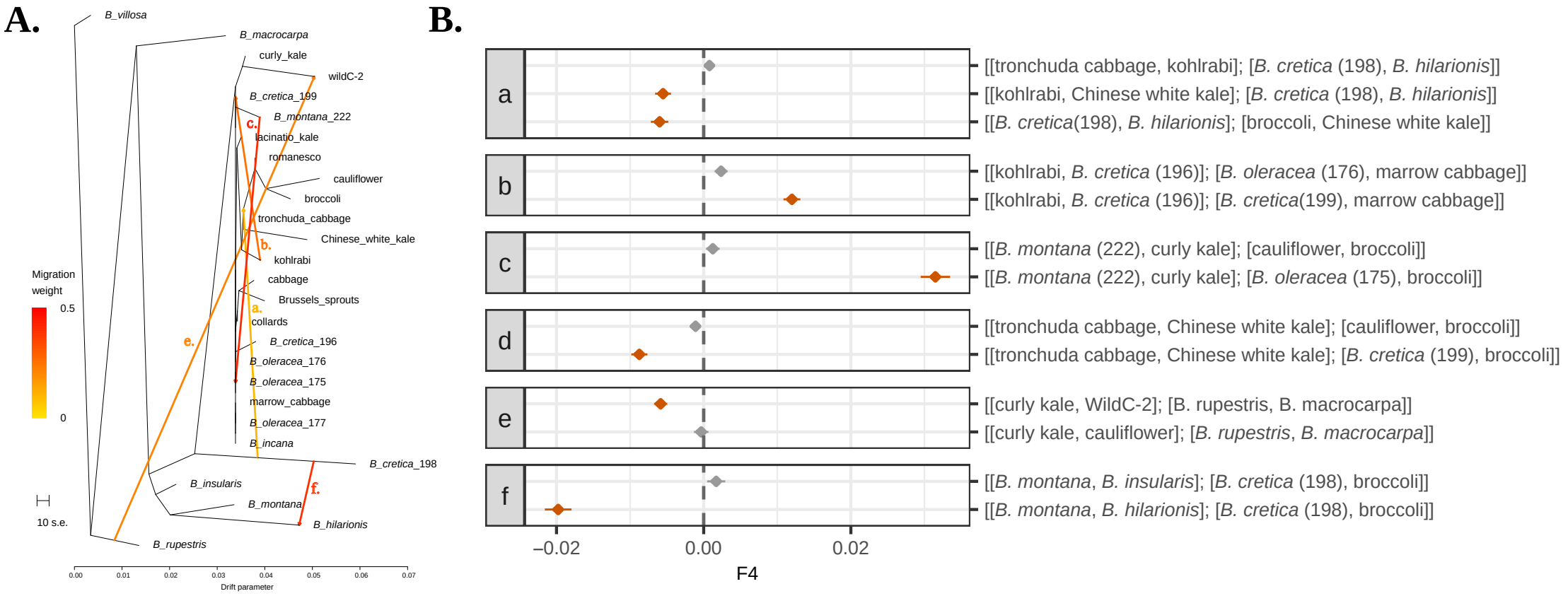

B. cretica
B. hilarionis
B. incana
B. insularis
B. macrocarpa
B. montana
B. oleracea
B. rupestris
B. villosa
0
400 km
NYBG GIS LABORATORY

C.

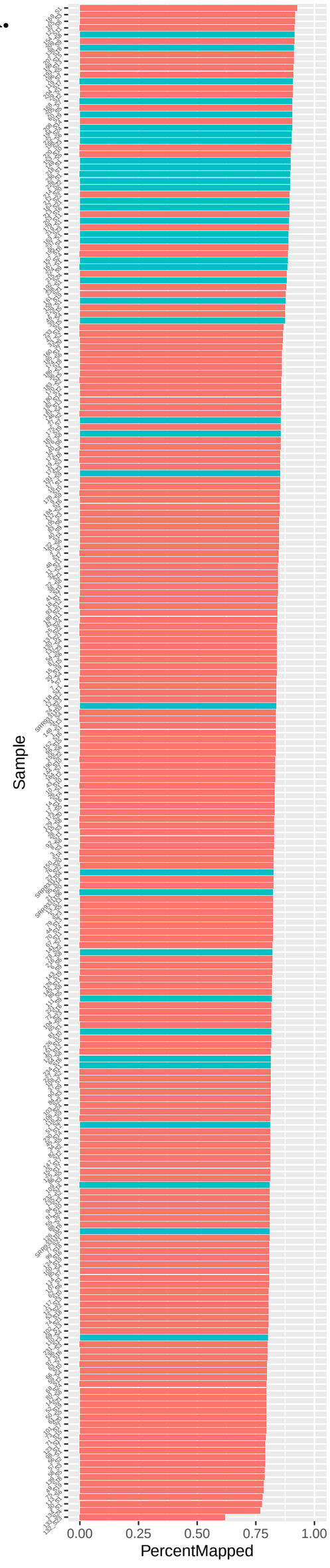

Habitat Suitability
High
Low

D.

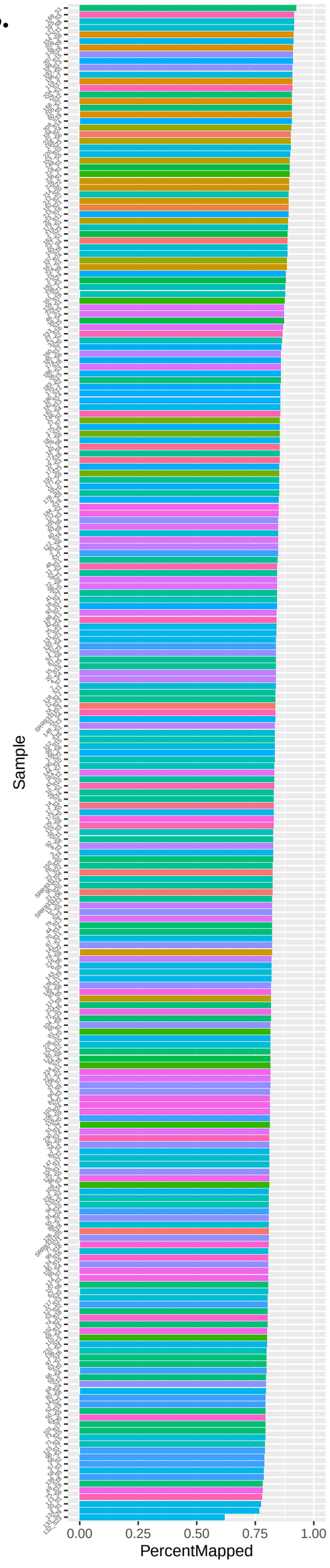

Habitat Suitability
High
Low


A.
B_villosa
B_macrocarpa
curly_kale
wildC-2
B_cretica_199
B_montana_222
lacinato_kale
romanesco
cauliflower
broccoli
tronchuda_cabbage
Chinese_white_kale
kohlrabi
cabbage
Brussels_sprouts
collards
B_cretica_196
B_oleracea_176
B_oleracea_175
marrow_cabbage
B_oleracea_177
B_incana
B_cretica_198
B_insularis
B_montana
B_hilarionis
B_rupestris
a.
b.
c.
e.
f.
Migration weight
0.5
0
10 s.e.
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
Drift parameter
B.
a
b
c
d
e
f
[[tronchuda cabbage, kohlrabi]; [B. cretica (198), B. hilarionis]]
[[kohlrabi, Chinese white kale]; [B. cretica (198), B. hilarionis]]
[[B. cretica(198), B. hilarionis]; [broccoli, Chinese white kale]]
[[kohlrabi, B. cretica (196)]; [B. oleracea (176), marrow cabbage]]
[[kohlrabi, B. cretica (196)]; [B. cretica(199), marrow cabbage]]
[[B. montana (222), curly kale]; [cauliflower, broccoli]]
[[B. montana (222), curly kale]; [B. oleracea (175), broccoli]]
[[tronchuda cabbage, Chinese white kale]; [cauliflower, broccoli]]
[[tronchuda cabbage, Chinese white kale]; [B. cretica (199), broccoli]]
[[curly kale, WildC-2]; [B. rupestris, B. macrocarpa]]
[[curly kale, cauliflower]; [B. rupestris, B. macrocarpa]]
[[B. montana, B. insularis]; [B. cretica (198), broccoli]]
[[B. montana, B. hilarionis]; [B. cretica (198), broccoli]]
–0.02
0.00
0.02
F4

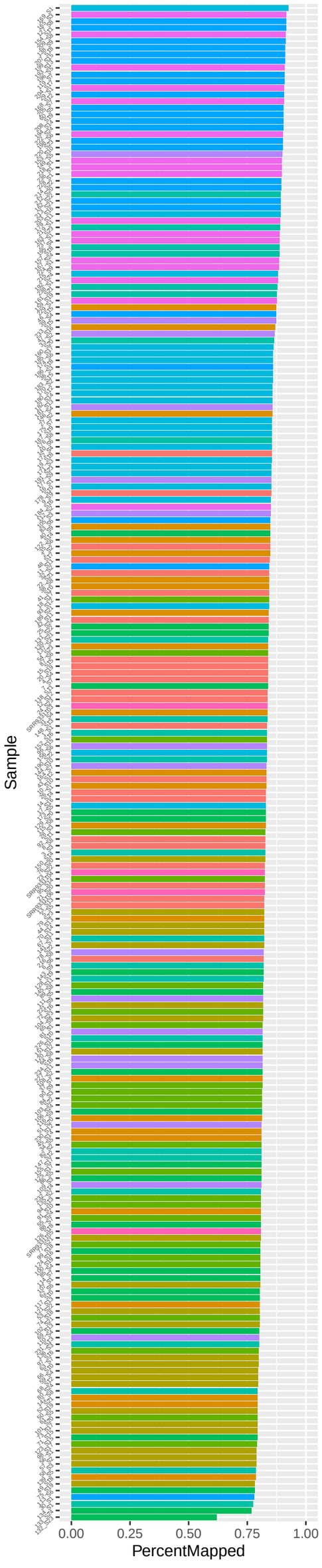
A.

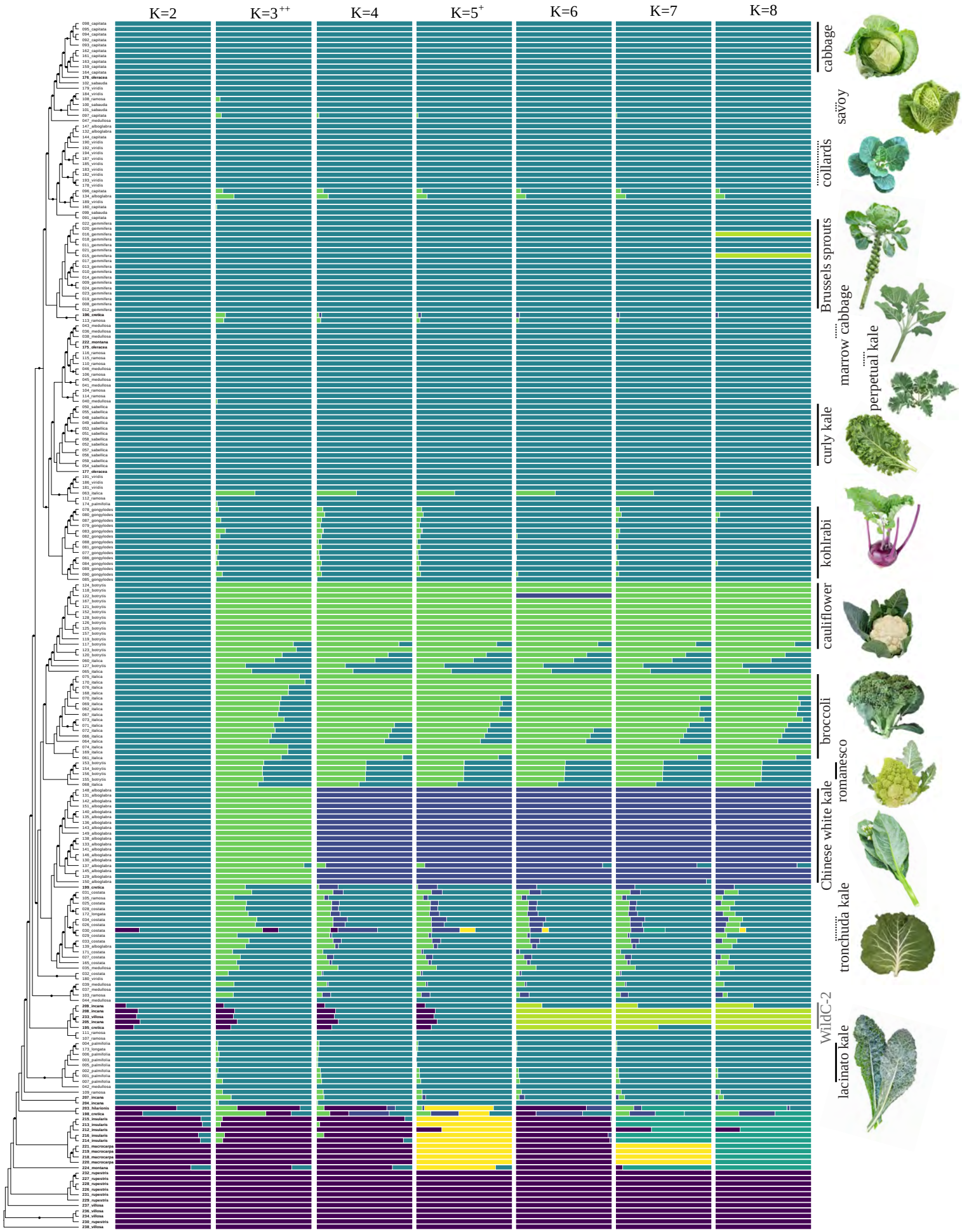

Sample
0.00
0.25
0.50
0.75
1.00
PercentMapped
Type
domesticate
wild

B.

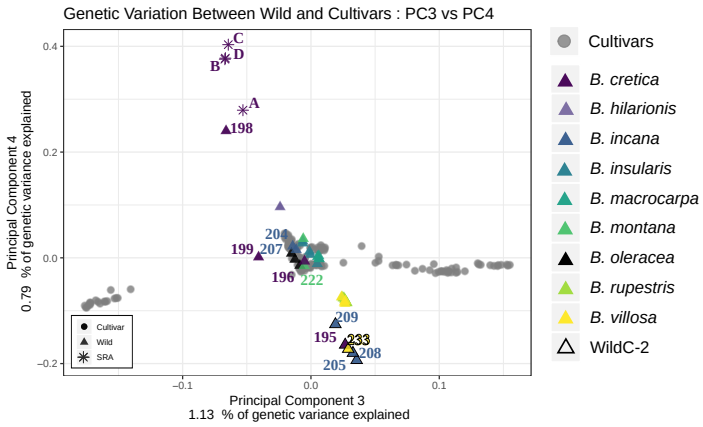

Sample
0.00
0.25
0.50
0.75
1.00
PercentMapped
Cultivar
B. cretica
B. hilarionis
B. incana
B. insularis
B. macrocarpa
B. montana
B. oleracea
B. rupestris
B. villosa
broccoli
Brussels sprouts
cabbage
cauliflower
Chinese white Kale
collards
curly kale
kolhrabi
lacinato kale
marrow cabbage
perpetual kale
savoy cabbage
tronchuda kale
walking stick kale

C.

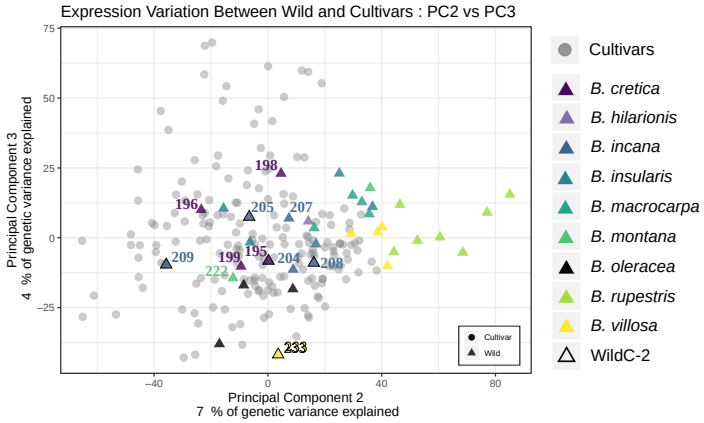

Sample
0.00
0.25
0.50
0.75
1.00
PercentMapped
Lane
1 (24)
2 (24)
3 (24)
4 (24)
5 (24)
6 (24)
7 (24)
8 (24)
9 (16)
10 (17)
SRA (4)


K=2
K=3++
K=4
K=5+
K=6
K=7
K=8
cabbage
savoy
collards
Brussels sprouts
marrow cabbage
perpetual kale
curly kale
kohlrabi
cauliflower
broccoli
romanesco
Chinese white kale
tronchuda kale
WildC-2
lacinato kale

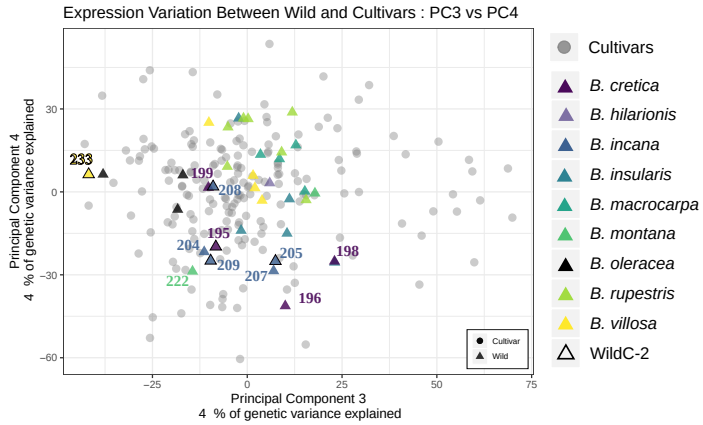
A.


Genetic Variation Between Wild and Cultivars : PC3 vs PC4
Principal Component 4
0.79 % of genetic variance explained
Principal Component 3
1.13 % of genetic variance explained
Cultivar
Wild
SRA
Cultivars
B. cretica
B. hilarionis
B. incana
B. insularis
B. macrocarpa
B. montana
B. oleracea
B. rupestris
B. villosa
WildC-2

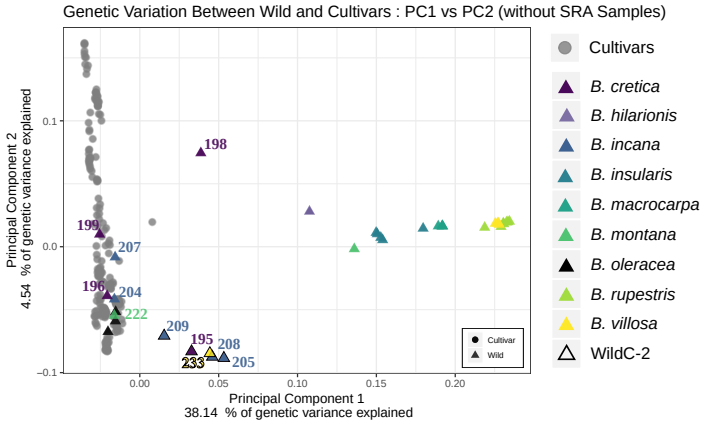
B.


Expression Variation Between Wild and Cultivars : PC2 vs PC3
Principal Component 3
4 % of genetic variance explained
Principal Component 2
7 % of genetic variance explained
Cultivar
Wild
Cultivars
B. cretica
B. hilarionis
B. incana
B. insularis
B. macrocarpa
B. montana
B. oleracea
B. rupestris
B. villosa
WildC-2

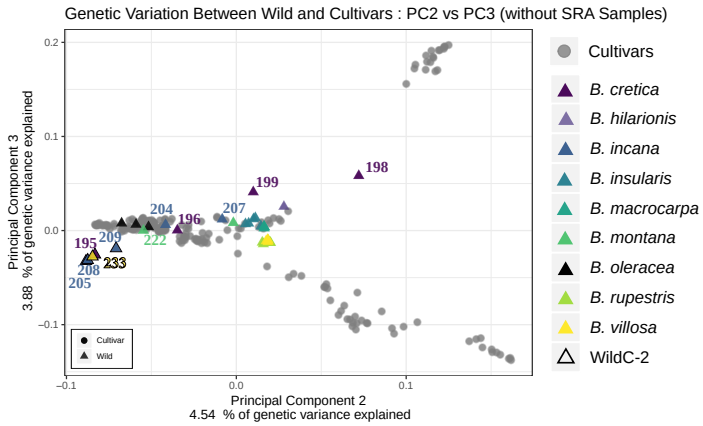
C.


Expression Variation Between Wild and Cultivars : PC3 vs PC4
Principal Component 4
4 % of genetic variance explained
Principal Component 3
4 % of genetic variance explained
Cultivar
Wild
Cultivars
B. cretica
B. hilarionis
B. incana
B. insularis
B. macrocarpa
B. montana
B. oleracea
B. rupestris
B. villosa
WildC-2

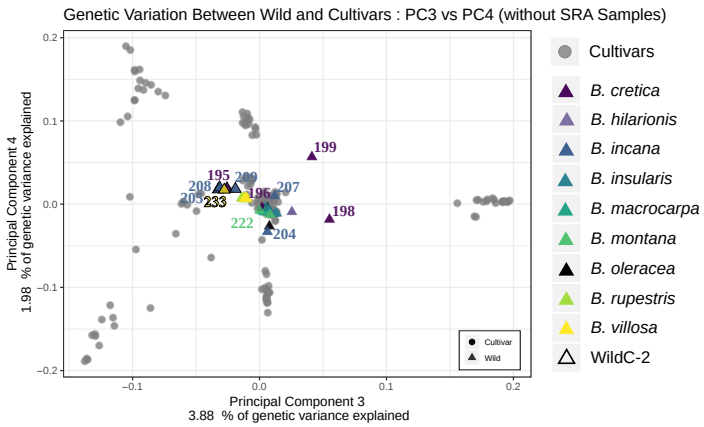
D.

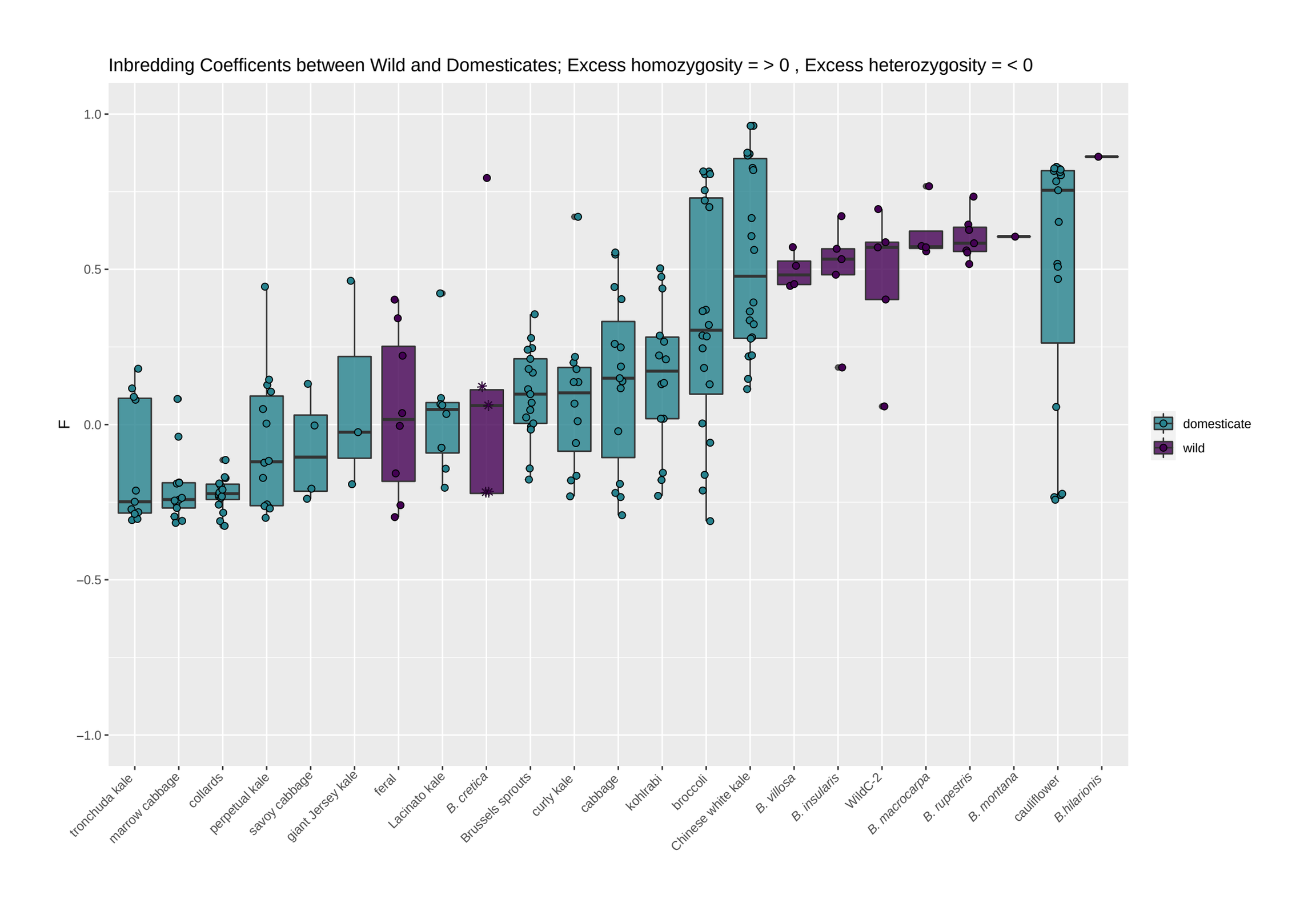

Genetic Variation Between Wild and Cultivars : PC1 vs PC2 (without SRA Samples)
Principal Component 2
4.54 % of genetic variance explained
Principal Component 1
38.14 % of genetic variance explained
Cultivar
Wild
Cultivars
B. cretica
B. hilarionis
B. incana
B. insularis
B. macrocarpa
B. montana
B. oleracea
B. rupestris
B. villosa
WildC-2

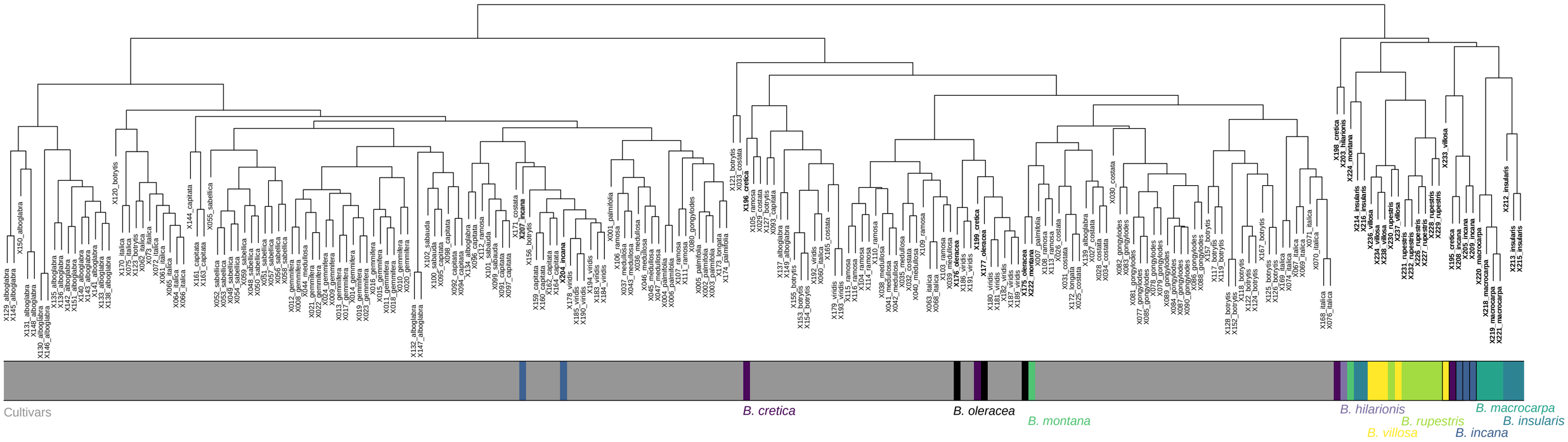
E.

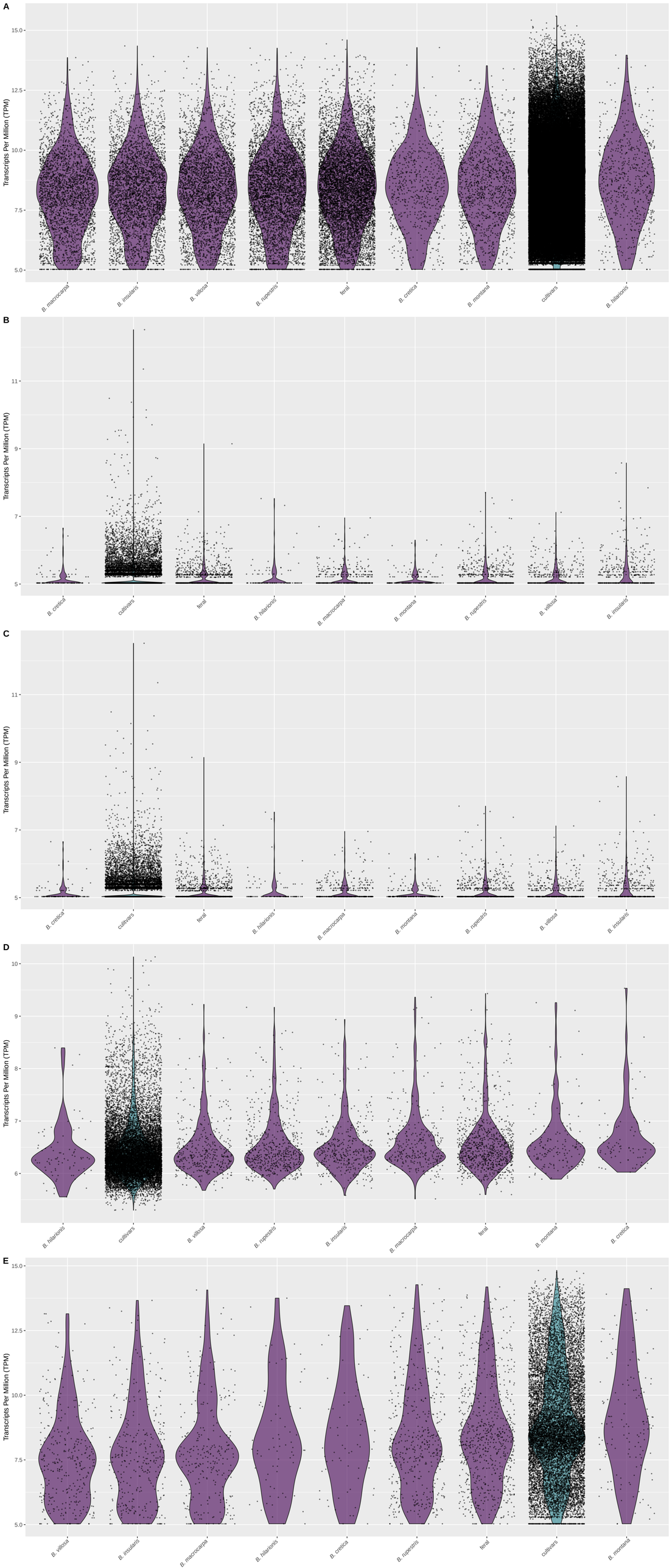

Genetic Variation Between Wild and Cultivars : PC2 vs PC3 (without SRA Samples)
Principal Component 3
3.88 % of genetic variance explained
Principal Component 2
4.54 % of genetic variance explained
Cultivar
Wild
Cultivars
B. cretica
B. hilarionis
B. incana
B. insularis
B. macrocarpa
B. montana
B. oleracea
B. rupestris
B. villosa
WildC-2

F.

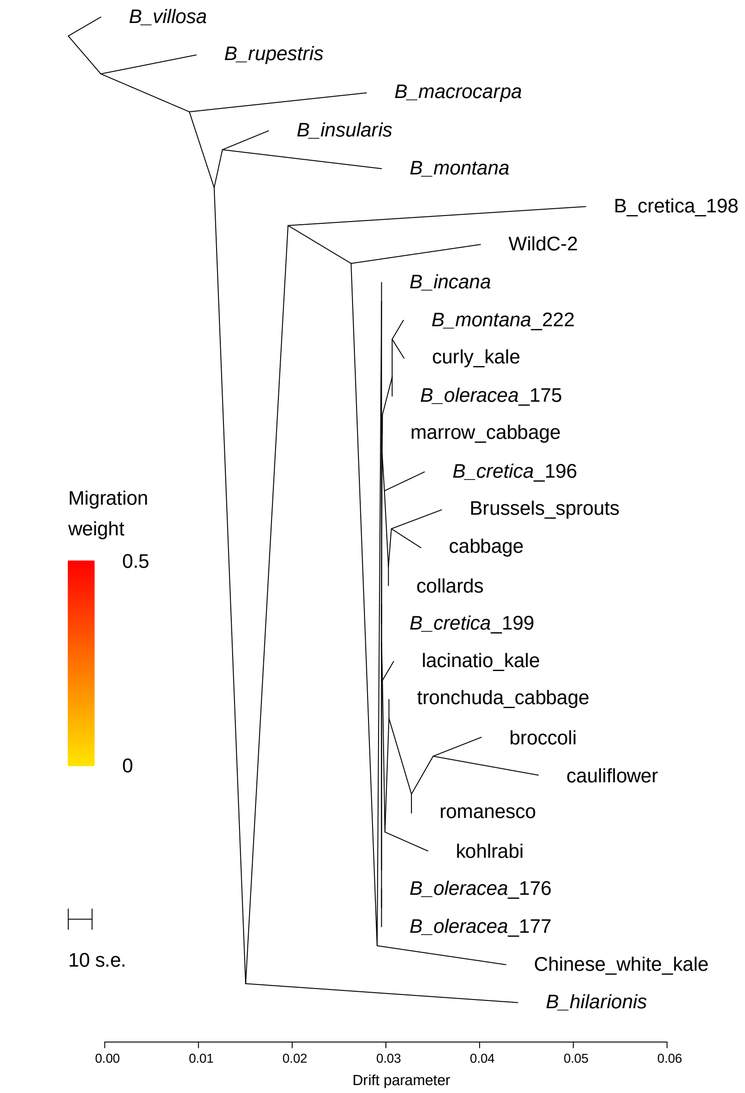

Genetic Variation Between Wild and Cultivars : PC3 vs PC4 (without SRA Samples)
Principal Component 4
1.98 % of genetic variance explained
Principal Component 3
3.88 % of genetic variance explained
Cultivar
Wild
Cultivars
B. cretica
B. hilarionis
B. incana
B. insularis
B. macrocarpa
B. montana
B. oleracea
B. rupestris
B. villosa
WildC-2


Inbredding Coefficents between Wild and Domesticates; Excess homozygosity = > 0 , Excess heterozygosity = < 0
F
1.0
0.5
0.0
−0.5
−1.0
tronchuda kale
marrow cabbage
collards
perpetual kale
savoy cabbage
giant Jersey kale
feral
Lacinato kale
B. cretica
Brussels sprouts
curly kale
cabbage
kohlrabi
broccoli
Chinese white kale
B. villosa
B. insularis
WildC-2
B. macrocarpa
B. rupestris
B. montana
cauliflower
B.hilarionis
domesticate
wild

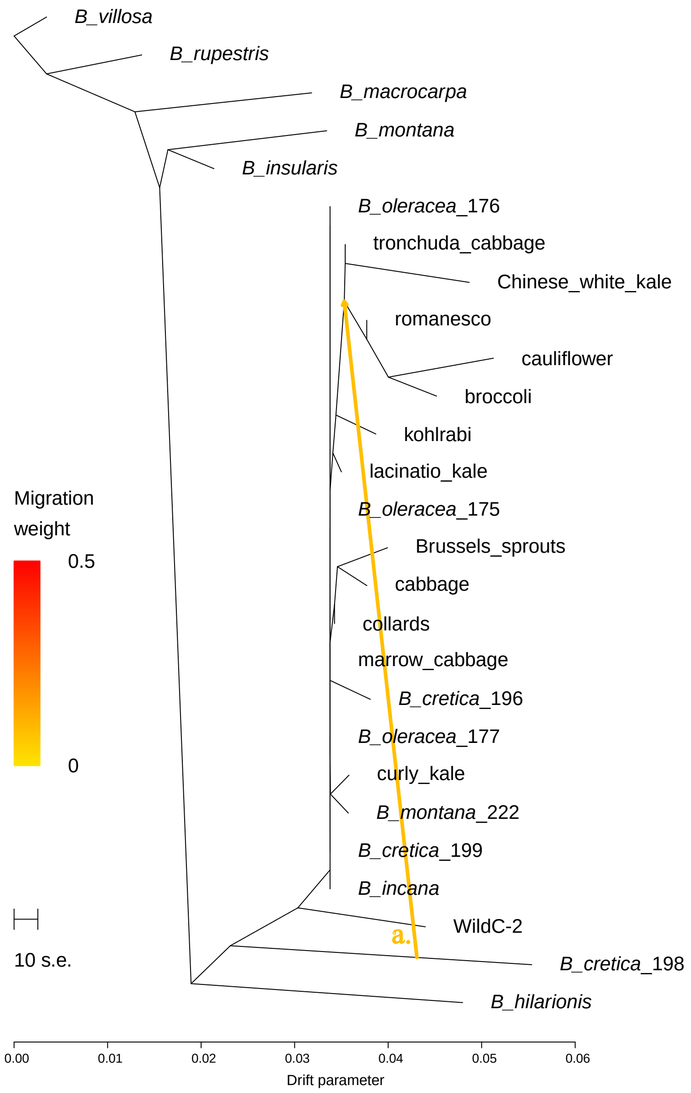

Brassica oleracea Expression Profile Clustering
X129_alboglabra
X145_alboglabra
X150_alboglabra
X131_alboglabra
X148_alboglabra
X130_alboglabra
X146_alboglabra
X135_alboglabra
X136_alboglabra
X142_alboglabra
X133_alboglabra
X140_alboglabra
X143_alboglabra
X141_alboglabra
X138_alboglabra
X139_alboglabra
X120_botrytis
X170_italica
X075_italica
X123_botrytis
X062_italica
X073_italica
X077_italica
X072_italica
X063_italica
X065_italica
X064_italica
X066_italica
X144_capitata
X161_capitata
X163_capitata
X055_sabellica
X052_sabellica
X053_sabellica
X049_sabellica
X054_sabellica
X059_sabellica
X048_sabellica
X050_sabellica
X051_sabellica
X057_sabellica
X060_sabellica
X058_sabellica
X012_gemmifera
X008_gemmifera
X044_medullosa
X021_gemmifera
X022_gemmifera
X024_gemmifera
X009_gemmifera
X013_gemmifera
X017_gemmifera
X014_gemmifera
X019_gemmifera
X023_gemmifera
X016_gemmifera
X015_gemmifera
X011_gemmifera
X018_gemmifera
X010_gemmifera
X020_gemmifera
X132_alboglabra
X147_alboglabra
X102_sabauda
X100_sabauda
X095_capitata
X098_capitata
X092_capitata
X094_capitata
X134_alboglabra
X096_capitata
X101_sabauda
X099_sabauda
X091_capitata
X097_capitata
X171_costata
X207_incana
X156_botrytis
X159_capitata
X160_capitata
X162_capitata
X164_capitata
X204_incana
X178_viridis
X185_viridis
X190_viridis
X194_viridis
X183_viridis
X184_viridis
X001_palmifolia
X106_ramosa
X037_medullosa
X043_medullosa
X036_medullosa
X046_medullosa
X045_medullosa
X047_medullosa
X004_palmifolia
X006_palmifolia
X177_ramosa
X111_ramosa
X080_gongylodes
X005_palmifolia
X002_palmifolia
X085_palmifolia
X173_longata
X174_palmifolia
X121_botrytis
X123_costata
X196_cretica
X105_ramosa
X029_costata
X172_longata
X093_capitata
X137_alboglabra
X149_alboglabra
X155_botrytis
X153_botrytis
X154_botrytis
X192_viridis
X060_italica
X165_costata
X179_viridis
X193_viridis
X115_ramosa
X116_ramosa
X104_ramosa
X114_ramosa
X110_ramosa
X038_medullosa
X041_medullosa
X042_medullosa
X035_medullosa
X032_costata
X040_medullosa
X109_ramosa
X063_italica
X088_italica
X103_ramosa
X039_medullosa
X186_viridis
X191_viridis
X199_cretica
X177_oleracea
X180_viridis
X181_viridis
X182_viridis
X187_viridis
X189_viridis
X176_oleracea
X222_montana
X007_palmifolia
X108_ramosa
X113_ramosa
X033_costata
X031_costata
X172_longata
X025_costata
X139_alboglabra
X027_costata
X028_costata
X034_costata
X030_costata
X082_gongylodes
X083_gongylodes
X081_gongylodes
X077_gongylodes
X085_gongylodes
X078_gongylodes
X079_gongylodes
X089_gongylodes
X087_gongylodes
X090_gongylodes
X086_gongylodes
X088_gongylodes
X157_botrytis
X117_botrytis
X119_botrytis
X128_botrytis
X152_botrytis
X118_botrytis
X122_botrytis
X124_botrytis
X167_botrytis
X125_botrytis
X126_botrytis
X074_italica
X067_italica
X069_italica
X071_italica
X070_italica
X168_italica
X076_italica
X198_cretica
X211_hilarionis
X224_montana
X214_insularis
X216_insularis
X236_villosa
X234_villosa
X238_villosa
X230_rupestris
X237_villosa
X231_rupestris
X232_rupestris
X226_rupestris
X227_rupestris
X228_rupestris
X229_rupestris
X233_villosa
X195_cretica
X206_incana
X205_incana
X209_incana
X220_macrocarpa
X218_macrocarpa
X219_macrocarpa
X221_macrocarpa
X212_insularis
X213_insularis
X215_insularis
Cultivars
B. cretica
B. oleracea
B. montana
B. hilarionis
B. rupestris
B. villosa
B. macrocarpa
B. insularis
B. incana


A
Transcripts Per Million (TPM)
15.0
12.5
10.0
7.5
5.0
B. macrocarpa
B. insularis
B. villosa
B. rupestris
feral
B. cretica
B. montana
cultivars
B. hilarionis
B
Transcripts Per Million (TPM)
11
9
7
5
B. cretica
cultivars
feral
B. hilarionis
B. macrocarpa
B. montana
B. rupestris
B. villosa
B. insularis
C
Transcripts Per Million (TPM)
11
9
7
5
B. cretica
cultivars
feral
B. hilarionis
B. macrocarpa
B. montana
B. rupestris
B. villosa
B. insularis
D
Transcripts Per Million (TPM)
10
9
8
7
6
B. hilarionis
cultivars
B. villosa
B. rupestris
B. insularis
B. macrocarpa
feral
B. montana
B. cretica
E
Transcripts Per Million (TPM)
15.0
12.5
10.0
7.5
5.0
B. villosa
B. insularis
B. macrocarpa
B. hilarionis
B. cretica
B. rupestris
feral
cultivars
B. montana

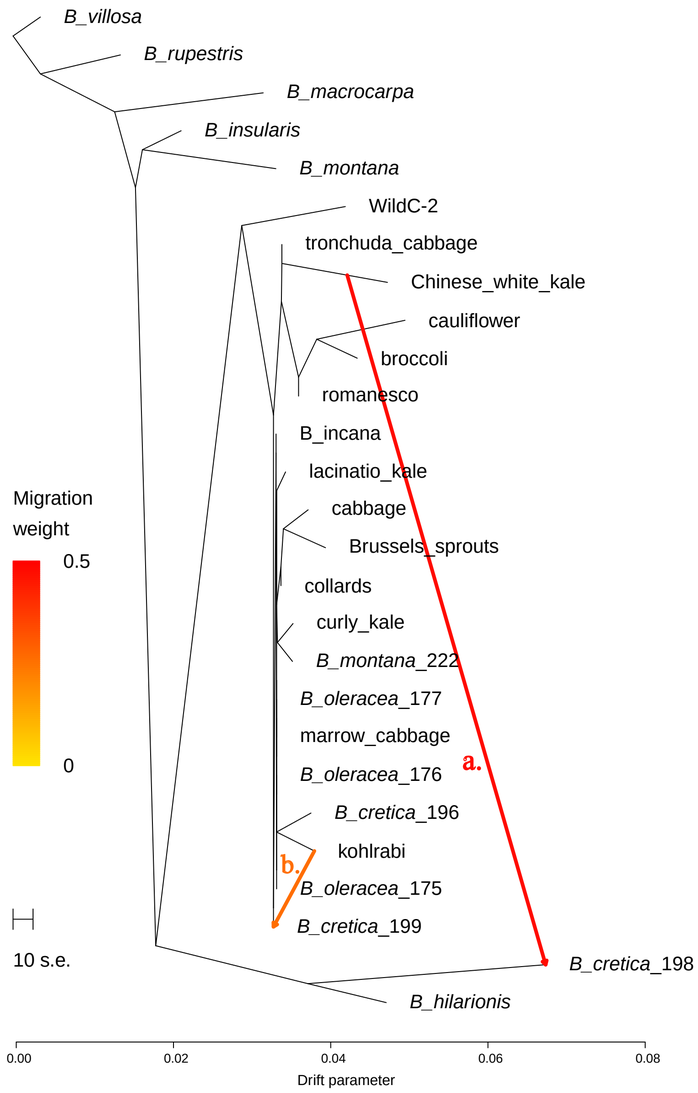
A. No migration inferred



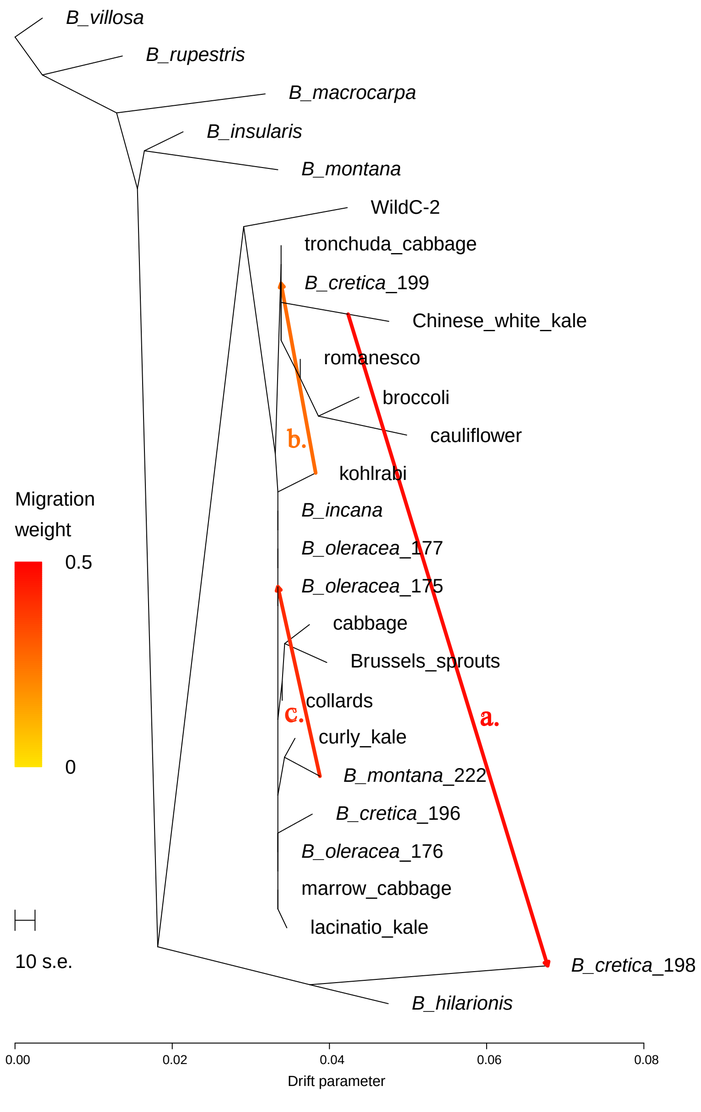
B. One migration event inferred



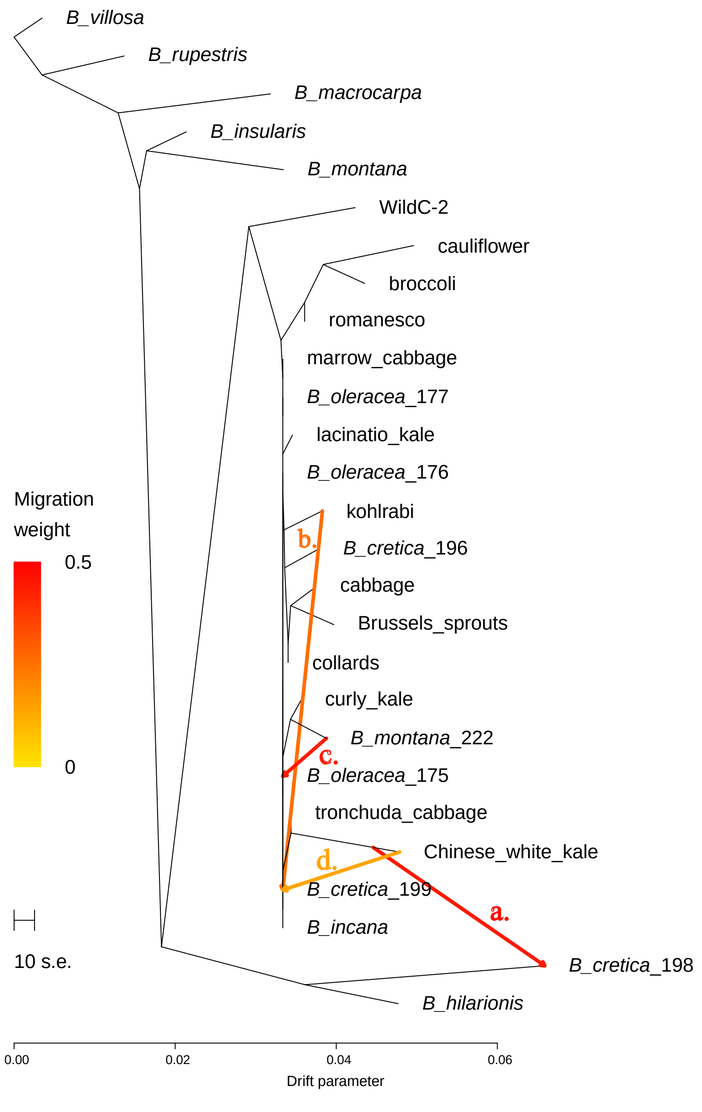
C. Two migration events inferred



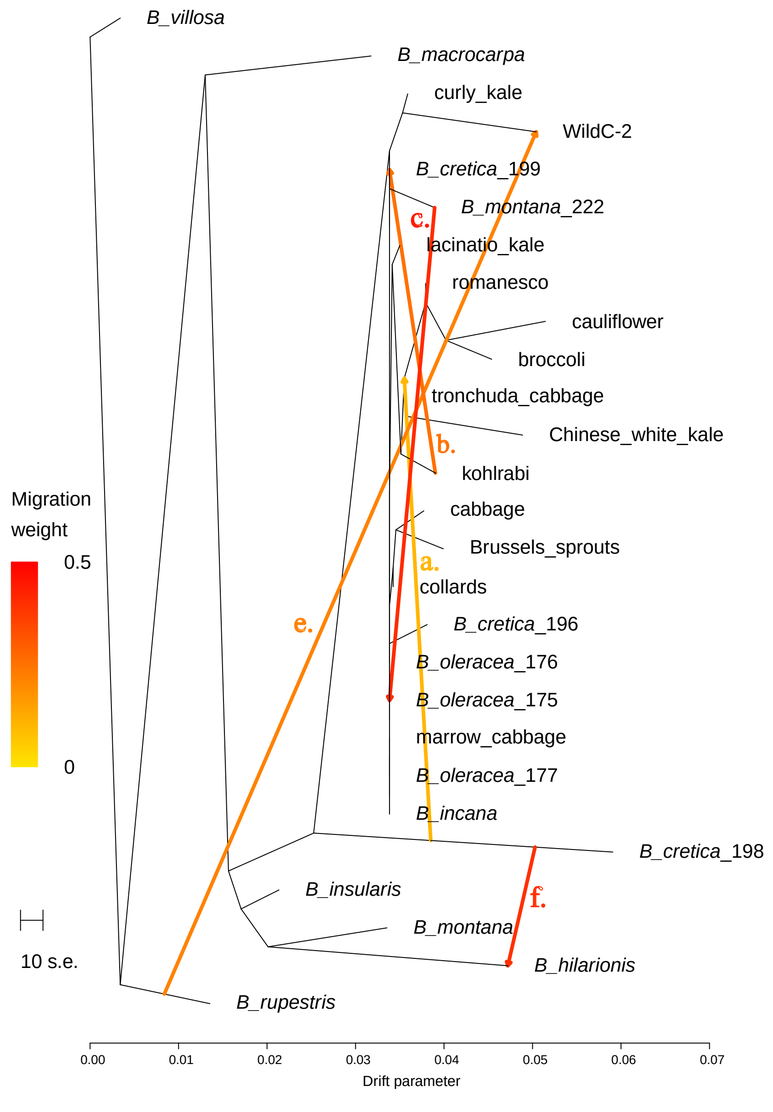
D. Three migration events inferred

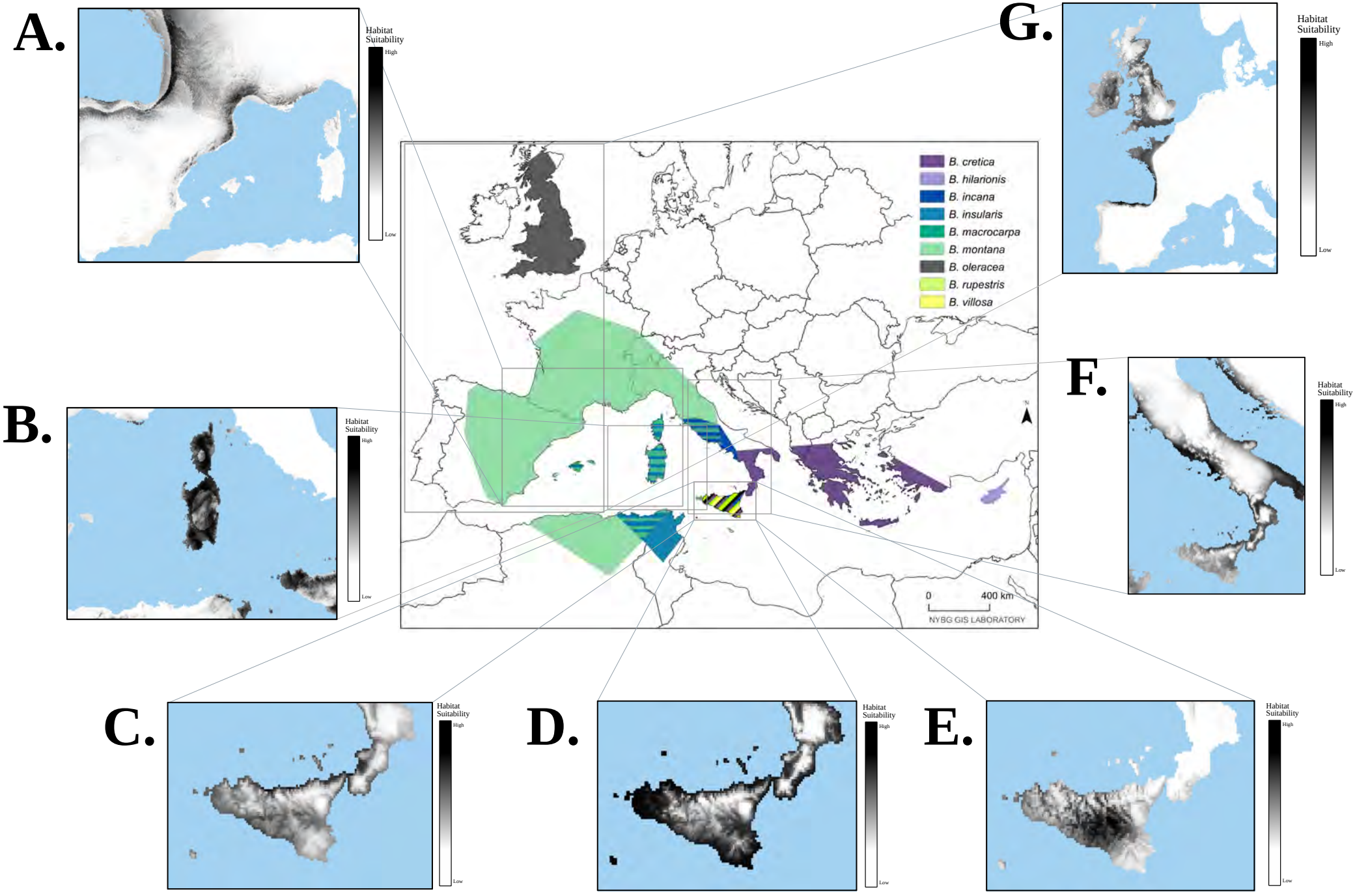


E. Four migration events inferred



F. Five migration events inferred




A.
B.
C.
D.
E.
F.
G.
Habitat Suitability
High
Low
B. cretica
B. hilarionis
B. incana
B. insularis
B. macrocarpa
B. montana
B. oleracea
B. rupestris
B. villosa
0
400 km
NYBG GIS LABORATORY

A. *B. oleracea* -175 Rep A, B, C, D

B. *B. oleracea* -176 Rep A, B, C, D

C. *B. oleracea* -177 Rep A, B, C, D

D. *B. cretica* -196 Rep A, B, C

E. *B. cretica* -199 Rep A, B, C, D

F. *B. montana* -222 Rep A, B, C, D

G. *B. incana* -204 Rep A, B, C, D

H. *B. incana* -207 Rep A, B, C, D

A. *B. cretica* -195 Rep A, B, C, D

B. *B. incana* -205 Rep A, B, C, D

C. *B. incana* -208 Rep A, B, C, D

D. *B. incana* -209 Rep A, B, C, D

E. *B. villosa* -233 Rep A, B, C, D

Table S1. *Brassica oleracea* crop types with common name, species name, Kew cultivar group, other used names, and phenotypic descriptions modified from [1]Bailey (1922) or [2]http://www.missouribotanicalgarden.org/ (last accessed 5/18/21). Illustrations by Andi Kur.

Common Name	Species Name	Kew Cultivar Group	Other Used Names	Phenotypic Description	Phenotype
Broccoli	*Brassica oleracea var.italica*	Botrytis Group		"Stout erect plant: grown for the thickened flower-shoots that are produced in the crown and also from the leaf axils, making a more or less open short large panicle: these shoots may bear flowers that are not abortive and that are taken for eating before the flowers open, or they may be less or elongated branches on which heads or knobs of abortive flowers are borne; the shoots are usually produced in the spring (plants 1 year old); fertile inflorescene open and branching as in other forms of the species, the bracts large and leafy, prominently toothed and sometimes laciniate." [1]	
Brussels spouts	*Brassica oleracea var. gemmifera*	Gemmifera Group		"Stout, simple, erect, 5-1m with large soft buds 2-3 cm. diam. Borne in the axils of the leaves practically throughout the length of the stem: leaves short and broad, blade 12-30 cm diam and about as long, short-oblong to nearly circular, sometimes unlobed but commonly with one or two rounded ear-like lobes near the base, the margins entire or only repand; petiole not winged; leaves of young plants often exhibit the characteristic shape and continuous margins; leaves of flowering stems obovate to oblong, obtuse, tapering to a petiole, margins usually obscurely shallowly toothed: flowers produced in the second year." [1]	
Cauliflower	*Brassica oleracea var. botrytis*	Botrytis Group		"Stout erect plant, leaves usually standing higher, bearing a single head of transformed flower-clusters: leaves long-oblong or elliptic, 15-24 cm broad and usually twice or more as long, with long flat mostly winged base or petiole, usually undivided but sometimes obscurely and irregularly lobed below, margins entire or minutely denticulate, strong ascending and overtopping the head or center: head composed of a condensed mass of short thick decolored peduncles and pedicles and thickened undeveloped flowers together with the attendant bracts; the little consolidate inflorescences are combined into a larger globular-topped clusters, forming the 'curds' of the head: flowering panicle arising in the second year, usually bearing in its lower part many little clusters of undeveloped flowers, but in the upper part producing the regular functional cabbage-like whitish flowers and elongated fruits; bracts small and mostly few, entire or nearly so." [1]	
Chinese White Kale	*Brassica oleracea var. alboglabra*	Alboglabra Group	Chinese broccoli, Gai laan, Kai lan	"A distinct-looking annual plant with its glaucous character, glabrous foliage, oval leaves which are petioled or not clasping, elongated inflorescence, and large while flowers." [1]	

Collard Greens	*Brassica oleracea var. viridis*	Acephala Group		“Biennial or potential perennial, with a simple stout stock 1m high in developed old plants: leaves very glaucous, in a heavy crown but not forming a solid head, the larger ones with blades 40-50cm long and nearly or quite as broad, rounded or oblong in outline, margin mostly sinuate or shallowly lobed, which more or less small lobing at the base; young plants many be harvested entire, but from old over-wintered plants the leaves may be taken as wanted.” [1]	
Curly Kale	*Brassica oleracea var. sabellica*	Acephala Group	*Brassica oleracea* var. *acephala*	“Plant mostly under 1m tall. Blooming in the second year. Stem scarcely woody. No solid heads, enlarged axillary buds. or condensed head-like thickened flower-clusters producd; grown for the foliage, the leaves. Leaves with wavy to frilled margins.” [1]	
Giant Jersey Kale	*Brassica oleracea var. longata*	Acephala Group	Couve Galega, Cow cabbage, Jersey Cabbage, Portuguese tree kale, Walking stick kale	“Perennial, usually or often not blooming till third year, the stem simple and mostly straight, becoming 2-4 m tall: leaves large, heavy and glaucous, harvested for food or mostly for fooder for cattle and also for pigs, rabbits and chickens as the plant grows, or falling naturally, the rosette therefore at the top, in some forms lobed at the base and in other undivided, sometimes fringed or curled.” [1]	
Kohlrabi	*Brassica oleracea var. gongylodes*	Gongylodes Group		“Biennial, low and stout, erect, 30-60 cm over all when not in bloom, glaucous and glabrous throughout: stem short, beginning to swell at 2-4 cm above the ground and producing a solid oblong globular or depressed globular leaf bearing edible tuber 5-10cm and more in diameter and which may continue to enlarge in the upper part producing a misshapen object: leaves thin and relatively small, 20-30 or 40 cm long of which about one-third to one-half is slender marginless petiole with an expanded base; blades oval to round-oval or oval-oblong, margings prominently irregularly dentate or notched, base of blade with one lobe on one side or one on either side; petiole usually bearing a very few little loves or leaf-fragments; leaves of main flowering stem much like the radical ones but smaller, the petiole distinct and slender; upper leaves of flowering stems as in B. oleracea: fruit as in B. oleracea, but beak usually very short and often swollen at base: seed small, 1-1.5-2 mm diameter, minutely alveolate, mostly strongly angled.” [1]	

Lacinato kale	*Brassica oleracea var. palmifolia*	Acephala Group	Black Tuscan palm, Dinosaur kale, Italian kale, Palm tree kale, Tuscan kale	“Plant mostly under 1m tall. Blooming in the second year. Stem scarcely woody. No solid heads, enlarged axillary buds. or condensed head-like thickened flower-clusters producd; grown for the foliage, the leaves. Leaves dark, almost blue-green with bumpy texture similar to Savoy cabbage.” [1]	
Marrow Cabbage	*Brassica oleracea var. medullosa*	Acephala Group	*Brassica oleracea* var. *acephala*	“Much like the Tree kales, but the stem begins to form a thick trunk with a tender flesh or marrow much relished by cattle.” [1]	
Ornamental cabbage	*Brassica oleracea var. acephala*	Acephala Group	Ornamental kale	“Typically develops large rosettes of broad flat leaves or curly, ruffled leaves in a tight rosette. Leaf colors are usually quite showy, including white/cream, pink, rose, red and purple.” [2]	
Perpetual Kale	*Brassica oleracea var. ramosa*	Acephala Group	*Brassica oleracea* var. *acephala,* Borecole, Branching kale, Thousand-headed kale	“Erect strong branching plants making a bush head standing 1-2m high and nearly as broad, perennial or potentially so: the stem usually begins to branch at 20-50 cm from the ground.” [1]	

Romanesco	*Brassica oleracea var. botrytis*	Botrytis Group	Broccolo Romanesco, Roman cauliflower, Romanesco broccoli	“Spiraling cone-shaped heads of light green florets forming a fractal pattern.” [2]	
Savoy Cabbage	*Brassica oleracea var. capitata f. sabauda*	Capitata Group		“Similar to White Cabbage but the leaves are bullate or blistered.” [1]	
Tronchuda Kale	*Brassica oleracea var. costata*	Tronchuda Group	Chou tronchuda, Coure tronchuda Portuguese cabbage, Portuguese kale, Seakale cabbage	“Biennial or annual: plant low with a simple stout stock and cabbage-like aspect but making no head although sometimes producing a loose tuft or rosette in the center, very glaucous: main leaves large and cabbage-like, with prominently developed whitish rib and veins, margin often finely serrate, basal lobes about 2 and small.” [1]	
White Cabbage	*Brassica oleracea var. capitata f. alba*	Capitata Group		“Plant low and stout: stem simple, producing one terminal head or giant leaf bud 10-30 cm or more diam, which in height and breadth usually exceeds the stem or stump: leaves large (large ones reaching 40 cm long and broad), spreading, at maturity more or less cucullate or ‘full’, oblong-obovate to nearly circular, the larger ones mostly unlobed and tapering to a very short broad winged petiole, some of them slightly lobed at base, margins undulate and variously obscurely toothed: leaves of flowering stems mostly prominently toothed, tapering to short semi-clasping base or the uppermost becoming nearly linear and much narrowed below: flower-stems arising from the axils in the heads (heads usually being split to let them out) or sometimes from the stump, produced in second year. Many forms include red cabbage which has red-purple foliage.” [1]	

Table S2. Sample information with species, variety, cultivar, germplasm, accession, sample #, location of original collection (year collected), and SRA #. Asterisks (*) next to sample # indicate those samples that were recovered as monophyletic and used in species tree reconstruction. USDA-ARS PGRU = United States Department of Agriculture, Agriculture Research Service, Plant Genetic Resources Unit, IPK = The Leibniz Institute of Plant Genetics and Crop Plant Research, UPM = Universidad Politécnica de Madrid, NordGen = The Nordic Genetic Resource Centre, GC = Gomez Campo Collection, Innes = John Innes Center, DH = doubled haploid lines, Pires = J.C. Pires personal collection.

Genus	Species	Variety	Cultivar	Germplasm	Accession	Sample #	Location	SRA #
Brassica	*oleracea*	*viridis*	Collards	USDA-ARS PGRU	PI662796	180	United States	SRR13508900
Brassica	*oleracea*	*viridis*	Collards	USDA-ARS PGRU	PI662798	190*	United States	SRR13508890
Brassica	*oleracea*	*viridis*	Collards	USDA-ARS PGRU	PI662804	184	United States	SRR13508895
Brassica	*oleracea*	*viridis*	Collards	USDA-ARS PGRU	PI662805	189	United States	SRR13508891
Brassica	*oleracea*	*viridis*	Collards	USDA-ARS PGRU	PI662806	181	United States	SRR13508899
Brassica	*oleracea*	*viridis*	Collards	USDA-ARS PGRU	PI662807	186	United States	SRR13508893
Brassica	*oleracea*	*viridis*	Collards	USDA-ARS PGRU	PI6628080	191	United States	SRR13508889
Brassica	*oleracea*	*viridis*	Collards	USDA-ARS PGRU	PI662809	183*	United States	SRR13508896
Brassica	*oleracea*	*viridis*	Collards	USDA-ARS PGRU	PI662814	182*	United States	SRR13508898
Brassica	*oleracea*	*viridis*	Collards	USDA-ARS PGRU	PI662815	178*	United States	SRR13508902
Brassica	*oleracea*	*viridis*	Collards	USDA-ARS PGRU	PI6628160	193*	United States	SRR13508887
Brassica	*oleracea*	*viridis*	Collards	USDA-ARS PGRU	PI662817	179	United States	SRR13508901
Brassica	*oleracea*	*viridis*	Collards	USDA-ARS PGRU	PI662828	192*	United States	SRR13508888
Brassica	*oleracea*	*viridis*	Collards	USDA-ARS PGRU	PI662831	185*	United States	SRR13508894
Brassica	*oleracea*	*viridis*	Collards	USDA-ARS PGRU	PI662833	187*	United States	SRR13508892
Brassica	*oleracea*	*viridis*	Collards	USDA-ARS PGRU	PI6628410	194*	United States	SRR13508885
Brassica	*oleracea*	*sabellica*	Curly Kale	IPK	BRA1491	54*	Netherlands	SRR13508803
Brassica	*oleracea*	*sabellica*	Curly Kale	IPK	BRA1841	53*	Poland	SRR13508804
Brassica	*oleracea*	*sabellica*	Curly Kale	IPK	BRA1899	55*	Sweden	SRR13508802
Brassica	*oleracea*	*sabellica*	Curly Kale	IPK	BRA2116	59*	Germany	SRR13508798
Brassica	*oleracea*	*sabellica*	Curly Kale	IPK	BRA2187	50*	Germany	SRR13508808
Brassica	*oleracea*	*sabellica*	Curly Kale	IPK	BRA2239	51*	Japan	SRR13508806
Brassica	*oleracea*	*sabellica*	Curly Kale	IPK	BRA2286	52*	Germany	SRR13508805
Brassica	*oleracea*	*sabellica*	Curly Kale	IPK	BRA2449	49*	Germany	SRR13508809
Brassica	*oleracea*	*sabellica*	Curly Kale	IPK	BRA2456	48*	Germany	SRR13508810
Brassica	*oleracea*	*sabellica*	Curly Kale	IPK	BRA960	56*	Denmark	SRR13508801
Brassica	*oleracea*	*sabellica*	Curly Kale	IPK	BRA963	57*	France	SRR13508800
Brassica	*oleracea*	*sabellica*	Curly Kale	IPK	BRA964	58*	Germany	SRR13508799
Brassica	*oleracea*	*sabauda*	Savoy Cabbage	Pires	Pires Collection 1	102	Unknown	SRR13508972
Brassica	*oleracea*	*sabauda*	Savoy Cabbage	USDA-ARS PGRU	PI246117	100	Netherlands	SRR13508753
Brassica	*oleracea*	*sabauda*	Savoy Cabbage	USDA-ARS PGRU	PI343603	99	Russian Federation	SRR13508754
Brassica	*oleracea*	*sabauda*	Savoy Cabbage	USDA-ARS PGRU	PI507856	101	Hungary	SRR13508973
Brassica	*oleracea*	*ramosa*	Perpetual Kale	USDA-ARS PGRU	G30718	115	United Kingdom	SRR13508958
Brassica	*oleracea*	*ramosa*	Perpetual Kale	USDA-ARS PGRU	G30724	116	New Zealand	SRR13508957
Brassica	*oleracea*	*ramosa*	Perpetual Kale	IPK	BRA1213	103	Republic of Georgia	SRR13508971
Brassica	*oleracea*	*ramosa*	Perpetual Kale	IPK	BRA1259	109	Italy	SRR13508965
Brassica	*oleracea*	*ramosa*	Perpetual Kale	IPK	BRA1891	105	Portugal	SRR13508969
Brassica	*oleracea*	*ramosa*	Perpetual Kale	IPK	BRA2382	112	France	SRR13508961
Brassica	*oleracea*	*ramosa*	Perpetual Kale	IPK	BRA2782	108	Colombia	SRR13508966
Brassica	*oleracea*	*ramosa*	Perpetual Kale	IPK	BRA2784	113	Serbia	SRR13508960
Brassica	*oleracea*	*ramosa*	Perpetual Kale	IPK	IBRA2905	111	Germany	SRR13508962

Brassica	*oleracea*	*ramosa*	Perpetual Kale	IPK	BRA2952	104	United Kingdom	SRR13508970
Brassica	*oleracea*	*ramosa*	Perpetual Kale	IPK	BRA2956	114	United Kingdom	SRR13508959
Brassica	*oleracea*	*ramosa*	Perpetual Kale	IPK	BRA3138	106	United Kingdom	SRR13508968
Brassica	*oleracea*	*ramosa*	Perpetual Kale	IPK	BRA67	110	Unknown	SRR13508964
Brassica	*oleracea*	*ramosa*	Perpetual Kale	IPK	K9829	107	United Kingdom	SRR13508967
Brassica	*oleracea*	*palmifolia*	Lacinato kale	Pires	Arias Collection 1	1*	Unknown	SRR13508976
Brassica	*oleracea*	*palmifolia*	Lacinato kale	Pires	Pires Collection 2	2*	Unknown	SRR13508975
Brassica	*oleracea*	*palmifolia*	Lacinato kale	IPK	BRA1905	5*	Argentina	SRR13508818
Brassica	*oleracea*	*palmifolia*	Lacinato kale	IPK	BRA1906	4*	Unknown	SRR13508829
Brassica	*oleracea*	*palmifolia*	Lacinato kale	IPK	BRA2846	6*	Italy	SRR13508807
Brassica	*oleracea*	*palmifolia*	Lacinato kale	IPK	BRA3073	7*	Unknown	SRR13508796
Brassica	*oleracea*	*palmifolia*	Lacinato kale	Pires	Pires Collection 3	3*	Unknown	SRR13508864
Brassica	*oleracea*	*palmifolia*	Giant Jersey Kale	USDA-ARS PGRU	G30723	174	United Kingdom	SRR13508906
Brassica	*oleracea*	*oleracea*	Wild oleracea	USDA-ARS PGRU	G30186	175	Canada	SRR13508905
Brassica	*oleracea*	*oleracea*	Wild oleracea	NordGen	NGB16241	177	Germany (Helgoland) - 2004	SRR13508903
Brassica	*oleracea*	*oleracea*	Wild oleracea	NordGen	NGB21657	176	Denmark (Rødvig) - 2004	SRR13508904
Brassica	*oleracea*	*medullosa*	Marrow Cabbage	USDA-ARS PGRU	G30717	35	United Kingdom	SRR13508824
Brassica	*oleracea*	*medullosa*	Marrow Cabbage	USDA-ARS PGRU	G30859	36*	United Kingdom	SRR13508823
Brassica	*oleracea*	*medullosa*	Marrow Cabbage	USDA-ARS PGRU	G30862	37	United Kingdom	SRR13508822
Brassica	*oleracea*	*medullosa*	Marrow Cabbage	IPK	CR1235	44	United Kingdom	SRR13508814
Brassica	*oleracea*	*medullosa*	Marrow Cabbage	IPK	CR1238	47	France	SRR13508811
Brassica	*oleracea*	*medullosa*	Marrow Cabbage	IPK	CR1256	43*	United Kingdom	SRR13508815
Brassica	*oleracea*	*medullosa*	Marrow Cabbage	IPK	CR1295	46*	Netherlands	SRR13508812
Brassica	*oleracea*	*medullosa*	Marrow Cabbage	IPK	CR1317	42	United Kingdom	SRR13508816
Brassica	*oleracea*	*medullosa*	Marrow Cabbage	IPK	CR1410	45*	Poland	SRR13508813
Brassica	*oleracea*	*medullosa*	Marrow Cabbage	IPK	CR1431	40*	Germany	SRR13508819
Brassica	*oleracea*	*medullosa*	Marrow Cabbage	IPK	CR1454	41*	Poland	SRR13508817
Brassica	*oleracea*	*medullosa*	Marrow Cabbage	IPK	CR2545	39	Spain	SRR13508820
Brassica	*oleracea*	*medullosa*	Marrow Cabbage	USDA-ARS PGRU	PI662703	38*	United Kingdom	SRR13508821
Brassica	*oleracea*	*longata*	walking stick kale	Pires	Pires Collection 4	172	Unknown	SRR13508909
Brassica	*oleracea*	*longata*	walking stick kale	Pires	Pires Collection 5	173	Unknown	SRR13508907
Brassica	*oleracea*	*longata*	walking stick kale	Pires	Pires Collection 6	171*	Unknown	SRR13508910
Brassica	*oleracea*	*italica*	Broccoli	DH	Early-Big	168*	Unknown	SRR13508839
Brassica	*oleracea*	*italica*	Broccoli	DH	Early-Big	76*	Unknown	SRR13508779
Brassica	*oleracea*	*italica*	Broccoli	USDA-ARS PGRU	G21111	72*	United States	SRR13508783
Brassica	*oleracea*	*italica*	Broccoli	USDA-ARS PGRU	G28865	71*	Italy	SRR13508784
Brassica	*oleracea*	*italica*	Broccoli	USDA-ARS PGRU	G30416	70*	United States	SRR13508786
Brassica	*oleracea*	*italica*	Broccoli	USDA-ARS PGRU	G30778	69*	Taiwan	SRR13508787
Brassica	*oleracea*	*italica*	Broccoli	USDA-ARS PGRU	G30779	73*	United States	SRR13508782
Brassica	*oleracea*	*italica*	Broccoli	DH	GD33-DH	170*	Unknown	SRR13508837
Brassica	*oleracea*	*italica*	Broccoli	DH	GD33-DH	75*	Unknown	SRR13508780
Brassica	*oleracea*	*italica*	Broccoli	DH	Mar34-DH	169*	Unknown	SRR13508838
Brassica	*oleracea*	*italica*	Broccoli	DH	Mar34-DH	74*	Unknown	SRR13508781
Brassica	*oleracea*	*italica*	Broccoli	USDA-ARS PGRU	PI231210	60	Italy	SRR13508797
Brassica	*oleracea*	*italica*	Broccoli	USDA-ARS PGRU	PI662524	63	United Kingdom	SRR13508793
Brassica	*oleracea*	*italica*	Broccoli	USDA-ARS PGRU	PI662525	68	United Kingdom	SRR13508788
Brassica	*oleracea*	*italica*	Broccoli	USDA-ARS PGRU	PI662529	65	United Kingdom	SRR13508791
Brassica	*oleracea*	*italica*	Broccoli	USDA-ARS PGRU	PI662596	62*	United States	SRR13508794
Brassica	*oleracea*	*italica*	Broccoli	USDA-ARS PGRU	PI662597	61*	United States	SRR13508795
Brassica	*oleracea*	*italica*	Broccoli	USDA-ARS PGRU	PI662674	64*	United States	SRR13508792
Brassica	*oleracea*	*italica*	Broccoli	USDA-ARS PGRU	PI662712	66*	Italy	SRR13508790

Brassica	*oleracea*	*italica*	Broccoli	USDA-ARS PGRU	PI662786	67*	United States	SRR13508789
Brassica	*oleracea*	*gongylodes*	Kohlrabi	USDA-ARS PGRU	G30762	90*	Italy	SRR13508764
Brassica	*oleracea*	*gongylodes*	Kohlrabi	USDA-ARS PGRU	G30935	84*	United Kingdom	SRR13508770
Brassica	*oleracea*	*gongylodes*	Kohlrabi	DH	HRIGRU005389	167*	Unknown	SRR13508911
Brassica	*oleracea*	*gongylodes*	Kohlrabi	IPK	BRA1889	85*	Italy	SRR13508769
Brassica	*oleracea*	*gongylodes*	Kohlrabi	IPK	BRA2349	89*	Netherlands	SRR13508765
Brassica	*oleracea*	*gongylodes*	Kohlrabi	IPK	BRA42	88*	Czech Republic	SRR13508766
Brassica	*oleracea*	*gongylodes*	Kohlrabi	USDA-ARS PGRU	PI096969	86*	Unknown	SRR13508768
Brassica	*oleracea*	*gongylodes*	Kohlrabi	USDA-ARS PGRU	PI188610	87*	Switzerland	SRR13508767
Brassica	*oleracea*	*gongylodes*	Kohlrabi	USDA-ARS PGRU	PI662560	80*	Poland	SRR13508775
Brassica	*oleracea*	*gongylodes*	Kohlrabi	USDA-ARS PGRU	PI662563	77*	Netherlands	SRR13508778
Brassica	*oleracea*	*gongylodes*	Kohlrabi	USDA-ARS PGRU	PI662623	79*	United States	SRR13508776
Brassica	*oleracea*	*gongylodes*	Kohlrabi	USDA-ARS PGRU	PI662624	82*	United States	SRR13508772
Brassica	*oleracea*	*gongylodes*	Kohlrabi	USDA-ARS PGRU	PI662625	83*	United States	SRR13508771
Brassica	*oleracea*	*gongylodes*	Kohlrabi	USDA-ARS PGRU	PI662671	81*	United States	SRR13508773
Brassica	*oleracea*	*gongylodes*	Kohlrabi	USDA-ARS PGRU	PI662704	78*	Netherlands	SRR13508777
Brassica	*oleracea*	*gemmifera*	Brussels Sprouts	USDA-ARS PGRU	G30601	9*	Kazakhstan	SRR13508774
Brassica	*oleracea*	*gemmifera*	Brussels Sprouts	USDA-ARS PGRU	G30867	12*	United Kingdom	SRR13508963
Brassica	*oleracea*	*gemmifera*	Brussels Sprouts	USDA-ARS PGRU	G30869	24*	Netherlands	SRR13508836
Brassica	*oleracea*	*gemmifera*	Brussels Sprouts	USDA-ARS PGRU	G30870	11*	Denmark	SRR13508974
Brassica	*oleracea*	*gemmifera*	Brussels Sprouts	USDA-ARS PGRU	G30872	10*	Germany	SRR13508763
Brassica	*oleracea*	*gemmifera*	Brussels Sprouts	USDA-ARS PGRU	PI209942	14*	Australia	SRR13508941
Brassica	*oleracea*	*gemmifera*	Brussels Sprouts	USDA-ARS PGRU	PI243050	15*	Ireland	SRR13508930
Brassica	*oleracea*	*gemmifera*	Brussels Sprouts	USDA-ARS PGRU	PI244839	16*	United States	SRR13508919
Brassica	*oleracea*	*gemmifera*	Brussels Sprouts	USDA-ARS PGRU	PI249534	19*	Spain	SRR13508886
Brassica	*oleracea*	*gemmifera*	Brussels Sprouts	USDA-ARS PGRU	PI261775	21*	United States	SRR13508863
Brassica	*oleracea*	*gemmifera*	Brussels Sprouts	USDA-ARS PGRU	PI312902	17*	Netherlands	SRR13508908
Brassica	*oleracea*	*gemmifera*	Brussels Sprouts	USDA-ARS PGRU	PI343669	22*	Russian Federation	SRR13508852
Brassica	*oleracea*	*gemmifera*	Brussels Sprouts	USDA-ARS PGRU	PI343670	18*	Russian Federation	SRR13508897
Brassica	*oleracea*	*gemmifera*	Brussels Sprouts	USDA-ARS PGRU	PI343671	20*	Russian Federation	SRR13508875
Brassica	*oleracea*	*gemmifera*	Brussels Sprouts	USDA-ARS PGRU	PI343673	23*	Russian Federation	SRR13508841
Brassica	*oleracea*	*gemmifera*	Brussels Sprouts	USDA-ARS PGRU	PI365170	13*	United Kingdom	SRR13508952
Brassica	*oleracea*	*gemmifera*	Brussels Sprouts	USDA-ARS PGRU	PI385957	8*	Kenya	SRR13508785
Brassica	*oleracea*	*costata*	Tronchuda Kale	USDA-ARS PGRU	PI662564	28*	Portugal	SRR13508832
Brassica	*oleracea*	*costata*	Tronchuda Kale	USDA-ARS PGRU	PI662565	27*	Portugal	SRR13508833
Brassica	*oleracea*	*costata*	Tronchuda Kale	USDA-ARS PGRU	PI662566	31*	Portugal	SRR13508828
Brassica	*oleracea*	*costata*	Tronchuda Kale	USDA-ARS PGRU	PI662567	32	Unknown	SRR13508827
Brassica	*oleracea*	*costata*	Tronchuda Kale	USDA-ARS PGRU	PI662568	33*	Portugal	SRR13508826
Brassica	*oleracea*	*costata*	Tronchuda Kale	USDA-ARS PGRU	PI662569	34*	Portugal	SRR13508825
Brassica	*oleracea*	*costata*	Tronchuda Kale	USDA-ARS PGRU	PI662659	30*	United Kingdom	SRR13508830
Brassica	*oleracea*	*costata*	Tronchuda Kale	USDA-ARS PGRU	PI662668	165*	Portugal	SRR13508840
Brassica	*oleracea*	*costata*	Tronchuda Kale	USDA-ARS PGRU	PI662668	29*	Portugal	SRR13508831
Brassica	*oleracea*	*costata*	Tronchuda Kale	USDA-ARS PGRU	PI662702	25*	Spain	SRR13508835
Brassica	*oleracea*	*costata*	Tronchuda Kale	Pires	Pires Collection 7	26*	Unknown	SRR13508834
Brassica	*oleracea*	*capitata*	Red Cabbage	USDA-ARS PGRU	PI187232	159*	Belgium	SRR13508917
Brassica	*oleracea*	*capitata*	Red Cabbage	USDA-ARS PGRU	PI245012	164*	France	SRR13508912
Brassica	*oleracea*	*capitata*	Red Cabbage	USDA-ARS PGRU	PI246046	163*	Netherlands	SRR13508913
Brassica	*oleracea*	*capitata*	Red Cabbage	USDA-ARS PGRU	PI246059	162*	Netherlands	SRR13508914
Brassica	*oleracea*	*capitata*	Red Cabbage	USDA-ARS PGRU	PI246060	161*	Netherlands	SRR13508915
Brassica	*oleracea*	*capitata*	Red Cabbage	USDA-ARS PGRU	PI329197	160	Netherlands	SRR13508916
Brassica	*oleracea*	*capitata*	Ornamental cabbage	USDA-ARS PGRU	PI662701	144	Italy	SRR13508932

Brassica	*oleracea*	*capitata*	Cabbage	USDA-ARS PGRU	PI171529	97	Turkey	SRR13508756
Brassica	*oleracea*	*capitata*	Cabbage	USDA-ARS PGRU	PI182149	96	Turkey	SRR13508757
Brassica	*oleracea*	*capitata*	Cabbage	USDA-ARS PGRU	PI225859	98*	Denmark	SRR13508755
Brassica	*oleracea*	*capitata*	Cabbage	USDA-ARS PGRU	PI263061	94*	Russian Federation	SRR13508759
Brassica	*oleracea*	*capitata*	Cabbage	USDA-ARS PGRU	PI320913	91	Hungary	SRR13508762
Brassica	*oleracea*	*capitata*	Cabbage	USDA-ARS PGRU	PI324236	92*	Sweden	SRR13508761
Brassica	*oleracea*	*capitata*	Cabbage	USDA-ARS PGRU	PI343514	93*	Russian Federation	SRR13508760
Brassica	*oleracea*	*capitata*	Cabbage	USDA-ARS PGRU	PI518837	95*	China	SRR13508758
Brassica	*oleracea*	*botrytis*	Romanesco	Pires	Pires Collection 8	153*	Unknown	SRR13508922
Brassica	*oleracea*	*botrytis*	Romanesco	Pires	Pires Collection 9	155*	Unknown	SRR13508920
Brassica	*oleracea*	*botrytis*	Romanesco	Pires	Pires Collection 10	156*	Unknown	SRR13508918
Brassica	*oleracea*	*botrytis*	Romanesco	Pires	Pires Collection 11	154*	Unknown	SRR13508921
Brassica	*oleracea*	*botrytis*	Cauliflower	USDA-ARS PGRU	G28888	122*	Germany	SRR13508950
Brassica	*oleracea*	*botrytis*	Cauliflower	USDA-ARS PGRU	G30435	126*	United States	SRR13508946
Brassica	*oleracea*	*botrytis*	Cauliflower	USDA-ARS PGRU	G30436	125*	United States	SRR13508947
Brassica	*oleracea*	*botrytis*	Cauliflower	USDA-ARS PGRU	G30438	124*	United States	SRR13508948
Brassica	*oleracea*	*botrytis*	Cauliflower	USDA-ARS PGRU	G30769	123*	Taiwan	SRR13508949
Brassica	*oleracea*	*botrytis*	Cauliflower	USDA-ARS PGRU	G30857	121*	United Kingdom	SRR13508951
Brassica	*oleracea*	*botrytis*	Cauliflower	DH	HRIGRU007826	128*	Unknown	SRR13508847
Brassica	*oleracea*	*botrytis*	Cauliflower	DH	HRIGRU007826	152*	Unknown	SRR13508923
Brassica	*oleracea*	*botrytis*	Cauliflower	USDA-ARS PGRU	PI231208	119*	Italy	SRR13508954
Brassica	*oleracea*	*botrytis*	Cauliflower	USDA-ARS PGRU	PI267724	127	United States	SRR13508945
Brassica	*oleracea*	*botrytis*	Cauliflower	USDA-ARS PGRU	PI267724	157*	United States	SRR13508842
Brassica	*oleracea*	*botrytis*	Cauliflower	USDA-ARS PGRU	PI269312	117*	Sweden	SRR13508956
Brassica	*oleracea*	*botrytis*	Cauliflower	USDA-ARS PGRU	PI271445	120*	India	SRR13508953
Brassica	*oleracea*	*botrytis*	Cauliflower	USDA-ARS PGRU	PI291996	118*	Israel	SRR13508955
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	DH	A12	131*	Unknown	SRR13508844
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	DH	A12	148*	Unknown	SRR13508927
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	Pires	Pires Collection 12	143*	Unknown	SRR13508933
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	DH	HRIGRU007543	130*	Unknown	SRR13508845
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	DH	HRIGRU007543	146*	Unknown	SRR13508929
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	DH	HRIGRU013023	132	Unknown	SRR13508843
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	DH	HRIGRU013023	147	Unknown	SRR13508928
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	IPK	BRA1006	134	Egypt	SRR13508943
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	IPK	BRA1143	139	Spain	SRR13508937
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	IPK	BRA1210	133*	China	SRR13508944
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	IPK	BRA1211	141*	China	SRR13508935
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	IPK	BRA1244	142*	China	SRR13508934
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	IPK	BRA1658	138*	Thailand	SRR13508938
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	IPK	BRA169	137*	China	SRR13508939
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	IPK	BRA1747	140*	Netherlands	SRR13508936
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	IPK	BRA1909	136*	Netherlands	SRR13508940
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	IPK	BRA990	135*	Unknown	SRR13508942
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	USDA-ARS PGRU	PI249556	149*	Thailand	SRR13508926
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	USDA-ARS PGRU	PI435900	150*	Spain	SRR13508925
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	USDA-ARS PGRU	PI662520	151*	China	SRR13508924
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	DH	TO1000	129*	Unknown	SRR13508846
Brassica	*oleracea*	*alboglabra*	Chinese White Kale	DH	TO1000	145*	Unknown	SRR13508931
Brassica	*cretica*	-	-	IPK	BRA3053	198*	Turkey (Kalamaki-Aydin) -1997	SRR13508882
Brassica	*cretica*	-	-	IPK	BRA3092	199	Greece	SRR13508881
Brassica	*cretica*	-	-	USDA-ARS PGRU	PI662588	195	Greece -1991	SRR13508884

Brassica	*cretica*	-	-	UPM	UPM6346	196	Turkey (Kushadasi, Millipark)	SRR13508883
Brassica	*hilarionis*	-	-	Innes	HRIGRU011463	203*	Cyprus (Kyrenia Range)	SRR13508880
Brassica	*incana*	-	-	IPK	BRA1262	204*	Unknown (likely Crimea)–1983	SRR13508879
Brassica	*incana*	-	-	IPK	BRA2918	205	Italy (Messina, Sicily)-1989	SRR13508878
Brassica	*incana*	-	-	USDA-ARS PGRU	PI662584	208	Italy (Tindari, Sicily)-1975	SRR13508876
Brassica	*incana*	-	-	USDA-ARS PGRU	PI662591	209	Italy (Tindari, Sicily)- 1975	SRR13508874
Brassica	*incana*	-	-	UPM	UPM5974	207*	Russia (Crimea, Cape Aju-Dag, Russia)	SRR13508877
Brassica	*insularis*	-	-	IPK	BRA2996	213*	Italy (Sardinia) -1993	SRR13508872
Brassica	*insularis*	-	-	IPK	BRA3050	214*	Italy (Monte Tuttavista, Sardinia) -1993	SRR13508871
Brassica	*insularis*	-	-	IPK	BRA3051	215*	Italy (Isola Rossa, Sardinia) -1993	SRR13508870
Brassica	*insularis*	-	-	IPK	BRA3052	216*	Italy (Isola di Figarolo, Sardinia) - 1993	SRR13508869
Brassica	*insularis*	-	-	USDA-ARS PGRU	PI662587	212*	Greece - 1971	SRR13508873
Brassica	*macrocarpa*	-	-	IPK	BRA2854	218*	Italy (Favignana)-2002	SRR13508868
Brassica	*macrocarpa*	-	-	IPK	BRA2944	219*	Unknown	SRR13508867
Brassica	*macrocarpa*	-	-	USDA-ARS PGRU	CO7031	220*	Unknown	SRR13508866
Brassica	*macrocarpa*	-	-	USDA-ARS PGRU	PI662585	221*	Italy (Egadi, Sicily) - 1975	SRR13508865
Brassica	*montana*	-	-	IPK	BRA1644	222	Unknown - 1983	SRR13508862
Brassica	*montana*	-	-	GC	3607-75	224*	Spain (Gerona)	SRR13508861
Brassica	*rupestris*	-	-	IPK	BRA2851	226*	Italy (Palermo, Sicily) -1998	SRR13508860
Brassica	*rupestris*	-	-	IPK	BRA2945	227*	Italy (Palermo, Sicily) -1988	SRR13508859
Brassica	*rupestris*	-	-	IPK	BRA2992	228*	Italy (Isnello, Sicily) - 1989	SRR13508858
Brassica	*rupestris*	-	-	GC	3822-75	229*	Unknown	SRR13508857
Brassica	*rupestris*	-	-	IPK	K10259	230	Italy (Trapani, Sicily) - 1998	SRR13508856
Brassica	*rupestris*	-	-	IPK	K10260	231*	Italy (Palermo, Sicily) - 1998	SRR13508855
Brassica	*rupestris*	-	-	UPM	UPM6575	232*	Italy (Caccamo, Sicily)	SRR13508854
Brassica	*villosa*	-	-	IPK	BRA1896	233	Italy (Messina, Sicily) - 1991	SRR13508853
Brassica	*villosa*	-	-	IPK	BRA2853	234*	Italy (Trapani, Sicily) - 1998	SRR13508851
Brassica	*villosa*	-	-	IPK	K10263	236*	Italy (Caltanissetta, Sicily) - 1998	SRR13508850
Brassica	*villosa*	-	-	IPK	K10264	237	Italy (Palermo, Sicily) - 1998	SRR13508849
Brassica	*villosa*	-	-	UPM	UPM6581	238*	Italy (Castelmare di Golfo, Sicily)	SRR13508848

Table S3. WGCNA predicted gene modules with number of genes in each module, the number of annotated *Arabidopsis thaliana* genes using blast and synteny, and the percent of syntenic genes represented in the module. Module number with asterisks (*) represent the five modules with the largest percent of syntenic genes in the module.

Module	Gene Number	# of *Arabidopsis* hits (Blast)	# of *Arabidopsis* hits (Synteny)	Percent of syntenic genes in module
1	35981	20982	18178	50.5
2	2268	1263	954	42.1
3	2241	2015	1991	88.8
4	1051	868	805	76.6
5	1007	860	849	84.3
6	726	454	308	42.4
7*	717	629	648	90.4
8	675	541	586	86.8
9	597	243	170	28.5
10	558	336	227	40.7
11	464	427	397	85.6
12	424	131	102	24.1
13*	420	385	378	90.0
14	382	327	306	80.1
15	304	240	269	88.5
16	297	99	75	25.3
17	291	267	260	89.3
18	270	92	74	27.4
19	267	92	71	26.6
20	224	65	46	20.5
21	178	161	153	86.0
22	177	154	159	89.8
23	173	71	47	27.2
24	133	37	24	18.0
25	128	124	105	82.0
26	111	73	45	40.5
27	109	31	20	18.3
28	108	33	20	18.5
29	106	25	18	17.0
30*	91	84	85	93.4
31*	80	78	75	93.8
32	75	71	67	89.3
33	71	17	13	18.3
34*	70	66	63	90.0
35	69	17	17	24.6
36	61	27	12	19.7
37	57	15	12	21.1
38	54	24	14	25.9
39	49	13	12	24.5
40	47	20	11	23.4
41	45	14	9	20.0
42	44	14	8	18.2
43	44	7	8	18.2
44	44	12	9	20.5
45	40	39	33	82.5
46	39	11	7	17.9
47	37	14	9	24.3
48	34	8	7	20.6

Table S4. Top five WGCNA modules with largest percent of syntenic genes represented in the module with corresponding annotated GO biological process with the largest fold enrichment, and p-value.

Module	GO Biological Process	Fold Enrichment	p-value
7	branched-chain amino acid catabolic process	17.86	1.32E-03
	phototropism	16.58	1.29E-02
	detection of abiotic stimulus	14.21	2.79E-02
	detection of external stimulus	14.21	2.79E-02
	fatty acid beta-oxidation	12.55	1.03E-02
13	mitochondrial transcription	93.62	4.82E-02
	snoRNA 3'-end processing	53.5	2.37E-05
	polyadenylation-dependent snoRNA 3'-end processing	49.93	1.04E-02
	endonucleolytic cleavage involved in rRNA processing	48.01	6.80E-04
	U4 snRNA 3'-end processing	45.39	1.41E-02
30	suberin biosynthetic process	> 100	3.52E-08
	phenylpropanoid biosynthetic process	54.01	5.32E-10
	phenylpropanoid metabolic process	43.61	3.23E-09
	secondary metabolite biosynthetic process	33.01	3.43E-08
	pectin catabolic process	26.94	4.29E-03
31	cytidine to uridine editing	75.81	1.25E-03
	mitochondrial RNA modification	60.3	1.14E-04
	base conversion or substitution editing	58.96	3.08E-03
	mitochondrial mRNA modification	56.86	3.51E-03
	RNA modification	42.34	6.27E-37
34	wax biosynthetic process	> 100	3.59E-12
	wax metabolic process	> 100	4.61E-12
	fatty acid derivative biosynthetic process	> 100	7.43E-12
	cuticle development	> 100	1.45E-11
	very long-chain fatty acid biosynthetic process	> 100	2.62E-04

Table S5: P-value results from ANOVA tests for differences in transcriptome abundance (Transcripts Per Million; TPM) between wild relatives, feral samples, and cultivars with Post-hoc Tukey Tests for top five WGCNA modules with largest syntenic overlap with *Arabidopsis thaliana*. Feral = *B. cretica* (196, 199); *B. incana* (204, 207); *B. montana* (222); and wild *B. oleracea* (175, 176, 177) + wild C - clade 2 (*B. incana* - 205, 208, 209; *B. villosa* - 233, and *B. cretica* - 195). Significance indicated as * < 0.05, ** < 0.01, *** < 0.001.

	Module 7	**Module 13**	**Module 30**	**Module 31**	**Module 34**
ANOVA overall F-test	< 2e-16***	< 2e-16***	2.19e-11***	< 2e-16***	< 2e-16***
cultivar - *B. cretica*	0.00274**	< 0.001***	0.8777	< 0.001***	0.52584
feral - *B. cretica*	0.98915	0.00148**	0.6153	0.64813	0.99979
B. hilarionis - *B. cretica*	0.04872*	< 0.001***	0.9772	0.00786**	0.99999
B. insularis - *B. cretica*	0.46058	< 0.001***	<0.001***	0.16945	0.14638
B. macrocarpa - *B. cretica*	0.0025**	0.89108	0.9357	0.99181	0.35236
B. montana - *B. cretica*	0.96827	0.00372**	1	0.95002	0.48072
B. rupestris - *B. cretica*	0.85884	< 0.001***	0.5024	0.25341	1
B. villosa - *B. cretica*	0.79588	< 0.001***	0.6899	0.06089	0.02287*
feral - cultivar	< 0.001***	< 0.001***	0.7272	< 0.001***	< 0.001***
B. hilarionis - cultivar	0.99998	0.75125	1	0.99255	0.23947
B. insularis - cultivar	< 0.001***	0.02175*	<0.001***	0.03896*	< 0.001***
B. macrocarpa - cultivar	< 0.001***	< 0.001***	1	< 0.001***	< 0.001***
B. montana - cultivar	0.00728**	0.00941**	0.2518	0.00227**	0.9979
B. rupestris - cultivar	< 0.001***	1	0.5648	< 0.001***	< 0.001***
B. villosa - cultivar	< 0.001***	0.99898	0.9689	0.38616	< 0.001***
B. hilarionis - feral	< 0.001***	< 0.001***	1	0.04125*	0.98463
B. insularis - feral	0.25387	0.11857	<0.001***	0.65499	< 0.001***
B. macrocarpa - feral	< 0.001***	< 0.001***	0.9892	0.82098	< 0.001***
B. montana - feral	0.0593	0.99961	0.0777	0.99971	0.19953
B. rupestris - feral	0.94733	< 0.001***	0.9999	0.8622	0.8588
B. villosa - feral	0.88344	< 0.001***	1	0.16468	< 0.001***
B. insularis - *B. hilarionis*	< 0.001***	0.08348	0.0643	0.43415	0.34999
B. macrocarpa - *B. hilarionis*	< 0.001***	< 0.001***	1	0.00598**	0.63699
B. montana - *B. hilarionis*	0.23313	0.01619*	0.8558	0.05712	0.24936
B. rupestris - *B. hilarionis*	< 0.001***	0.85892	0.9992	0.25267	1
B. villosa - *B. hilarionis*	< 0.001***	0.93863	1	0.71374	0.08142
B. macrocarpa - *B. insularis*	0.02963*	< 0.001***	<0.001***	0.11338	0.99929
B. montana - *B. insularis*	< 0.001***	0.93591	<0.001***	0.73906	< 0.001***
B. rupestris - *B. insularis*	0.93961	0.11517	<0.001***	0.99993	< 0.001***
B. villosa - *B. insularis*	0.993	0.08769	<0.001***	0.9984	0.95976
B. montana - *B. macrocarpa*	< 0.001***	0.00737**	0.5367	0.99978	< 0.001***
B. rupestris - *B. macrocarpa*	< 0.001***	< 0.001***	0.9449	0.19816	0.00529**
B. villosa - *B. macrocarpa*	0.00151**	< 0.001***	0.9969	0.01668*	0.71258
B. rupestris - *B. montana*	0.00909**	0.02415*	0.0511	0.87996	0.0253*
B. villosa - *B. montana*	0.00738**	0.01758*	0.1528	0.37787	< 0.001***
B. villosa - *B. rupestris*	1	0.99999	1	0.94283	< 0.001***

Table S6: Archaeological *Brassica* reports from Europe and the Eastern Mediterranean. BP = years before the present (1950).

Date	Countries	Notes	References
Middle Bronze Age ca. 3550 - 3350 BP	Austria, Friaga	3 charred seeds, hilltop settlement in the southern Alps. Unusually early date, but no reason to suspect contamination from later occupation etc.	[1]
ca. 3250 - 2970 BP	Gibala, NW Syria	C14 dated seeds from pottery vessels associated with a destruction layer. Seeds have been SEM; Seeds could be *Brassica cretica*.	[2]
La Tène 2400 - 2050 BP	Switzerland/France	Preservation not specified but is absent in Roman samples. Refers to identification problems.	[3]
Late Iron Age 2350 - 2050 BP	Germany	Review of various finds including *Brassica oleracea* from the German Late Iron Age	[4]
Roman-Iron Age 1950 - 1650 BP	Egypt, Mons Claudianus	charred/desiccated seeds resembling *Brassica oleracea*	[5] p.200
1950 – 1550 BP	Italy, Vado Ligure, Liguria	possible seeds found in Roman well	[6]
1907-1540 BP	England	recorded from 3 military and 2 town sites	[7]
Roman 1850-1650 BP	Germany, Otterbach, Kaiserslautern	11 waterlogged seeds of probable *Brassica* cf. *oleracea* from a well in Otterbach	[8]
1550-850 BP	Sweden, Denmark, Netherlands	seeds of *Brassica oleracea*	[9, 10]
1150-700 BP	Czech Republic, Žatec	Finds of waterlogged *Brassica* cf. *oleracea*	[11]

1050-850 BP	Great Britain, York	Finds of waterlogged *Brassica* cf. *oleracea*	[12]
1050-850 BP	Great Britain, York	Finds of waterlogged *Brassica* cf. *oleracea*	[13]
850 BP	Denmark	One to four chance finds from cultural layers	[14]
Early Medieval 850-750 BP	England, Raunds, Northants	Residues of *Brassica* leaves identified within pottery sherds.	[15, 16]
727-650 BP	Montgomery, Powys, Wales	Single waterlogged seed comparable to *Brassica oleracea* from pit fill associated with castle	[17]
750 - 650 BP	Germany, Einbeck / Greifswald	Archaeobotanical material including mid. 14th century sewers. [19] p.150	[18, 19]
Late Middle Ages 650 - 450 BP	Denmark, København	Archaeobotanical finds	[20]
Medieval 550 - 450 BP	Hungrary, Budapest	Waterlogged seeds in well	[21] (Table 6.1)
650-350 BP	Germany, Stralsund, Kiel; Lüneburg; Greifswald	Mainly waterlogged seeds of *Brassica* cf. *oleracea*	[22, 23]
450- 350 BP	Netherlands, Haarlem	Uncertain identification of mineralised seed of *Brassica* cf. *oleracea*	[24]
450- 350 BP	Germany, Rostock	archaeobotanical finds	[25]
350-250 BP	Belgium, Antwerpen	Seeds found in contents of 17th Century waste pit at Antwerpen	[26]

350-250 BP	Hungary, Buda	Probable charred seed of cabbage (*Brassica* cf. *oleracea*) found within waste pit	[27]

Supplementary References

1. Schmidl, A., Oeggl, K. Subsistence strategies of two hilltop settlements in the Eastern Alps - Friaga/Bartholomäberg (Vorarlberg, Austria) and Ganglegg/Schluderns (South Tyrol, Italy). *Veget. Hist. Archaeobot.* 14, 303–312 (2005).

2. Kaniewski, D. *et al*., The Sea Peoples, from cuneiform tablets to carbon dating**.** *PLoS ONE* **6**(6), e20232 (2011).

3. Jacomet, S., Vandorpe, (Trans. Martinoli, D.) Plantes anciennes et nouvelles. La région du Rhin supérieur et l'Allemagne du Sud-Ouest, in Reddé *et al*., Eds., Aspects de la Romanisation dans l'Est de la Gaule. (Bibracte, 2011). pp. 345-360.

4. Willerding, U. Obstarten und Nüsse; Garen-, Obst- und Weinbau; Gartenbau, Obstbau, Weinbau in Benecke, N., Donat, P., Gringmuth-Dallmer, E., Willerding, U. Eds., *Frühgeschichte der Landwirtschaft in Deutschland Beitr Ur- u Frühgesch Mitteleuropa 14. Langenweißbach*, (2003), pp.30-31, pp. 162-172, pp. 258-262.

5. van der Veen, M. The botanical evidence. in Maxfield, V.A., Peacock, D.P.S.. *Mons Claudianus. Survey and excavation 1987-1993. Vol 2. Excavations: Part 1. Fouilles* (IFAO [Inst Franç Archéol Orientale] 43. Paris, 2001), pp. 174-222; appendix 1-2; fig. 8.1-8.13.

6. Arobba, D., Bulgarelli, F., Siniscalco, C., Caramiello, R. Roman landscape and agriculture on the Ligurian coast through macro and microremains from a Vada Sabatia well (Vado Ligure, Italy). *Environmental Archaeology* **18**(2), 114-131 (2013).

7. van der Veen, M., Livarda, A., Hill, A. New food plants in Roman Britain – dispersal and social access. *Environmental Archaeology* **13**(1), 11-36 (2008).

8. Weitholdd, J. Hirse, Hanf und Hohldotter: Pflanzenfunde aus einem römischen Brunnen in Otterbach, Kr. Kaiserslautern, in Stobbe, A., Tegtmeier, U. Eds., *Verzweigungen: Eine Würdigung für A. J. Kalis und J. Meurers-Balke*. (Verlag Dr. Rudolf Habelt GmbH, Bonn, 2012).

9. Hansson, A.-M. Vikingabotanik. in Pettersson, B. *et al*., Eds., *Människan och naturen.Etnobiologi i Sverige* 1. (Stockholm, 2001).

10. Heimdahl, J. Barbariska trädgårdsmästare.Nya perspektiv på hortikulturen i Sverige fram till 1200-talets slut. *Fornvännen* **105**, 265-280 (2010).

11. Kočár, P., Čech, P., Kozáková, R., Kočárová, R. Environment and economy of the Early Medieval settlement in Žatec. *Interdiciplinaria Archaeologica, Natural Sciences in Archaeology (IANSA)* **1**(1-2), 45-60 (2010).

12. Godwin, H., Bachem, K. Appendix III Plant Materials. 109-13. in Richardson, K.M. Excavations in Hungate, York. *Archaeol. J*. **116** (for 1959), 51-114 (1961).
13. Hall, A.R., Williams, D., Greig, J.R.A. [plant remains]. 157-225 & fiche. in Hall, A.R., Kenward, H.K., Williams, D., Greig, J.R.A. Environment and living conditions at two Anglo-Scandinavian sites. The Archaeology of York AY **14**(4). London: CBA. 157-240 and fiche 1 (1983).

14. Jensen, H.-A. Catalogue of late- and post-glacial macrofossils of Spermatophyta from Denmark, Schleswig, Scania, Halland, and Blekinge dated 13,000 B.P. to 1536 A.D. *Danmarks Geologiske Undersøgelser Serie A*, nr. 6 1-95. (København, Reitzels forlag, 1985).

15. Evershed, R.P., Heron, C., Charters, S., Goad, L.J. The survival of food residues: new methods of analysis, interpretation and application. *Proc. Brit. Acad.* **77**, 189–208 (1992).

16. Evershed, R.P., Amot, K.L., Collister, J., Eglinton, G., Charters, S. Application of isotope ratio monitoring gas chromatography-mass spectrometry to the analysis of organic residues of archaeological origin. *Analyst* **119**, 909–14 (1994).

17. Greig, J., Girling, M., Skidmore, P. The plant and insect remains. 60-71 in Barker, P., Higham, R. *Hen Domen Montgomery. A timber castle on the English-Welsh Border.* (I. Roy. Archaeol. Inst., London, 1982).

18. Wiethold, J. Giff in de schottele. Strowe dar peper up …Botanische Funde als Quellen zur mittel-alterlichen Ernährungs- und Umweltgeschichte in Einbeck. in Heege A. (Hrsg.), *Einbeck im Mittelalter : eine archäologisch-historische Spurensuche*. Studien zur Einbecker Geschichte 17 (Oldenburg 2002), pp. 240-246.

19. Ansorge, J., Igel, K., Schäfer, H., Wiethold, J. Ein Holzschacht aus der Baderstr. 1 in Greifswald. Aus der materiellen Alltagskultur der sozialen Oberschicht einer Hansestadt in der 2. Hälfte des 14. Jahrhunderts. Bodendenkmalpflege in Mecklenburg-Vorpommern, Jahrbuch 50, 2002, 119-157 (2003).

20. Moltsen, A.S.A. Arkæobotaniske undersøgelser fra middelalderbyen København - metodik og udvalgte eksempler in Viklund, K. Ed., *Nordic archaeobotany - NAG 2000 in Umeå*. Archaeology and Environment 15. (Univ. Umeå, Umeå, 2002). pp. 173-180.

21. Zoltán, A., Sárgadinnye, A. (Cucumis melo) Archeogenetikája, ITS- és SSR – Heterogenitása Egy 15. századi Lelettől a Mai Fajtákig, Unpublished Doctoral Thesis. Gödöllő. (2006).

22. Wiethold, J. Archäobotanische Ergebnisse. in Fries, H., Wiethold, J. *Bemerkenswertes aus Stralsunds Altstadt die Grabung Apollonienmarkt 6 und ihre Ergebnisse*. (Archäol Ber Mecklenburg-Vorpommern 10, 2003), pp. 220-247.

23. Wiethold, J. Archäobotanische Untersuchungen zur Ernährungs- und Wirtschaftsgeschichte des Mittelalters und der frühen Neuzeit. in Noël, R., Paquay, I., Sosson, J.-P. Eds., *Au-delà de l'écrit. Les hommes et leurs vécus matériels au Moyen Âge à la lumière des sciences et des techniques. Nouvelles perspectives*. (Louvain-la-Neuve, 2003), pp. 461-499.

24. Brinkkemper, O. Plantenresten uit beerputten aangetroffen op een bouwlocatie aan het Spaarne te Haarlem. Een ‚rijke' informatiebron! *Haarlems Bodemonderzoek* **36**, 104-132 (2002).

25. Wiethold, J. Ernährung und Umwelt im spätmittelalterlichen Rostock. Archäobotaniscje Ergebnisse der Analyse zweier Kloaken in der Kröpeliner Straße 55-56, Kuhstraße. *Bodendenkmalpflege Mecklenburg-Vorpommern, Jahrbuch 1999* **47**, 351-378 (2000).

26. Cooremans, B. Het macrobotanisch onderzoek: een speurtocht naar plantaardige (voedsel) resten. in Veeckman, J., van Hoof, W., Cooremans, B., Ervynck, A., van Neer, W. Eds., *De inhoud van de afvalput van de Groote Schalien Loove: speuren naar de 17de eeuwsde bewoners.* BRABOM Berichten en rapporten over het Antwerps. Bodemonderzoek en Monumentenzorg 3. (Antwerpen, Stad Antwerpen, 2000), pp. 131-139.

27. Tóth, A.J., Daróczi-Szabó, L., Kovács, Zs.E., Gál, E., Bartosiewicz, L. In the light of the crescent moon: reconstructing environment and diet from an Ottoman Period deposit in Sixteenth to Seventeenth Century Hungary. in Peres, T., Van Derwarker, A. Eds., *Integrating Zooarchaeology and Paleoethnobotany*. (Springer, New York, 2010), pp. 245-280.

Table S7: Literary and artistic sources covering the Classical Greek, Roman, and medieval and post-medieval sources. BP = years before the present (1950).

Date	Countries	Notes	References
2500-2000 BP	Ionian Greece	Earliest textual reference. Refers to a "seven-leaved" cabbage in iambic verse. Perhaps as an oath or offering.	Hipponax, Fr.104 [1] p. 145
2410-2320 BP	Classical Greece	Refers to cabbage in a few medicinal recipes	Hippocrates, *Nat Mul.* 32, *Mul 1.78* [2] p. 75
ca. 2320-2238/2235 BP	Classical Greece	3 varieties: smooth & seedless (capita), larger leafed, sweeter parsley/curly-leaved, and a wild type bitter taste with many branches and smaller leaves may be *B. cretica* (see [3] p. 20)	Theophrastus *Historia Plantarum* VII. iv .4-6 [4] pp. 84-85 Athenaeus The Deipnosophists 9.9 [5] pp. 582-3
ca. 2150-2250 BP	Cyclades, Greece	States the finest cabbages grow in Cyme, but bitter in Alexandria. Seed brought from Rhodes produces a sweet cabbage than after a year taste degenerates.	Diphilus of Siphnus (the Siphnian). Athenaeus The Deipnosophists 9.9 [5] pp. 582-3
ca. 2150-2050 BP	Classical Greece	3 varieties smooth, parsley-leaved and salty. Last has delicate taste and grows in Eretria, Cyme and Rhodes (and in Cnidos and Ephesus).	Eudemus f. 85, Quoted in Athenaeus *The Deipnosophists* 9.9 [5] pp. 582-3
ca. 2150-2050 BP	Greece +	Smooth leaved, sometimes found wild. Also a curly leaved and one of reddish color.	Nicander Georgics. f.85. Athenaeus The Deipnosophists 9.9 [5] pp. 582-3
ca. 2110 BP	Classical Roman Italy	Lists smooth-leaved, curly leaved (apiaca), a small stalked, pungent tender variety and wild type	Cato, Marcus[6] De Agricultura. Capitula CLVI-CLVII:156-157, 1 pp. 144-151
1873-1871 BP	Classical Roman Italy	Lists many varieties.	[7] 41.19.
1700-1500 BP	Egypt	List of Monastic sources from Egyptian papyri	Oxyrhynchus Papyri XIV 1656 (BL), [8]
1600-1500 BP	Roman	Cited in nine recipes attributed to Apicius: dealing with Cimæ (cymae) & Coliculi (refers to sprouts, young cabbage and soft cabbage)	[9] Apicus III, ix.87-92, x.94, xii.99 & xv.103
1180-1150 BP	France/ Germany	Capitularies of Charlemagne Caulos & Ravacaulos (Kohlrabi)	Capitulare de Villis [10] p. 90
Medieval 1150-650 BP	Poland, Krakow	archaeobotanical and literary sources. See [11] for identification criteria (also [12])	[11, 12]
Medieval Spain ca. 990-600 BP	Spain	Translations of qannabit & kurunb would indicate cauliflower & cabbage from *c.*960 AD	[13]
750-550 BP	France/ Switzerland	Le Viandier de Taillevent: Collection of medieval recipes. Mentioned in one recipe.	[14]

Late Middle Ages 650-450 BP	Czech Republic	Reference to cabbage, garlic and onion in cooking and medicine	15
Late Middle Ages 650-450 BP	Estonia, Tartu	Written and artistic depictions of *Brassica oleracea*	16, 17
450-250 BP	Flanders and Netherlands	Depictions of cauliflower and cabbages in art.	18
450-350 BP	Russia		Sil'vestr's Domostroi
396 BP	Flemish	lists 5 types of cabbage - white, red, savoy, Roosken (pale red) and curly kale	19
370 BP	England	Refers to time of sowing and Ionians holding them as sacred	20
303 BP	Denmark	*Brassica oleracea capita* referred to as garden plant in *Horticultura Danica*	21

Supplementary References

1. West, M.L. *Studies in Greek Elegy and Iambus*. (De Gruyter, 1974).

2. Totelin, L.M.V. Ed., *Hippocratic recipes: oral and written transmission of pharmacological knowledge in Fifth- and Fourth-century Greece*. (Brill., 2009).

3. Willis, R.J. *The History of Allelopathy*. (Springer, 2007).

4. Theophrastus, *Historum Plantarum*. Enquiry into Plants, Volume II: Books 6-9: On Odours. Weather Signs. Trans. Hort, A.F. (*De Causis Plantarum*). (Loeb Classical Library 79, 1916).

5. Athenaeus of Naucratis *c.*3rd C AD, *The Deipnosophists: Banquet of the Learned of Athnaeus Volume II.* (Trans. Yonge). (Henry G. Bohn, London, 1854).

6. Hooper, A.D., Ash, H.B. *Cato and Varro: On Agriculture.* Loeb Classical Library, No. 283. (Harvard Univ. Press, 1934).

7. Pliny, the Elder, AD 77-79 *Natural History*, Volume V: Books 17-19. (Trans. Rackham, H.). (Loeb Classical Library 371, 1950).

8. Harlow, M., Smith, W. Between fasting and feasting: the literary and archaeobotanical evidence for monastic diet in Late Antique Egypt, *Antiquity* **75**(290), 758–768 (2001).

9. Vehling, J.D. *Apicius. Cookery and Dining in Imperial Rome*. Trans. Joseph Dommers Vehling. (Walter M. Hill, Chicago, 1936).

10. Boretius, A. Ed., *Capitularia regum Francorum* I, MGH Legum Sectio II (Hanover, 1883), no. 32, pp. 82-91.

11. Wieserowa, A. Plant remains from the early and late Middle Ages found in the settlement layers of the main market square in Cracow. *Acta Palaeobotanica* **20**(2), 137-212 (1979).

12. Zemanek, A., Wasylikowa, K. Historia botaniki i archeobotanika w poszukiwaniu danycho użytkowaniu roślin w średniowiecznym Krakowie. *Analecta, Studia i Materiały z Dziejów Nauki*, V **1**(9). 123–138 (1996).

13. Harvey, J. Plants of Moorish Spain: a fresh look. *Garden History* **20**(1), 71-82 (1992).

14. Prescott, J. [1315-1395] *le Viandier de Taillevent: 14th Century Cookery, Based on the Vatican Library Manuscript Translated into English by James Prescott*. (Hypatia Press, Eugene, ed. 2, 1989).

15. Beranová, Zelí, cibule a česnek v kuchyni a v medicíně do konce 16. Století. *Sborník Západčeského muz Plzni [Plizen] řada History* **16**, 185-194 (2002).

16. Sillasoo, I. Gardens and garden products in Medieval Turku, Estonia, in Viklund, K. Ed., *Nordic archaeobotany* - Archaeology and Environment 15. (Univ. of Umeå, Umeå, 2002), pp. 181-192.

17. Sillasoo, I. Plant depictions in late medieval religious art, in Laszlovszky, J., Szabó, P. Eds., *People and Nature in Historical Perspective*. (Central European Univ., Budapest, 2003), pp. 377-393.

18. Zeven, A.C., Brandenburg, W.A. Use of paintings from the 16th to 19th centuries to study the history of domesticated plants. *Economic Botany* **40**(4), 397-408 (1986).

19. Dodoens, R. *Cruydeboeck*. (Antwerpen, 1554).

20. Tusser, T. (1580). *Five Hundred Pointes of Good Husbandrie*. (English Dialect Society, London, 1878).

21. Block, H.R. (1647) *Ældste danske Havebog, Horticultura Danica*. Copenhagen (Wormianum, Aarhus, 1984).

Table S8. Area under the receiver operating characteristic curve (AUC) values for Maxent environmental niche model runs for putative wild relatives of *Brassica oleracea* crops.

Species	AUC
B. hilarionis	0.988
B. insularis	0.954
B. macrocarpa	0.982
B. montana	0.972
B. oleracea	0.989
B. rupestris	0.979
B. villosa	0.977
B. incana	0.967
B. cretica	0.988

Table 1. Wild species which have been proposed as progenitor species for *B. oleracea* crop types. Specific location in parentheses if indicated by the author.

Cultivar	Wild relative	Author
Broccoli	*B. oleracea*	Linnaeus
	B. oleracea	Hedrick (1919) *
	B. oleracea	Giles (1941) **
	B. montana	Hegi (1919)
	B. oleracea (from Italy)	Giles (1941)
	B. cretica	Gates (1953)
	B. oleracea and *B. alboglabra*	Song et al. (1990)
Brussels sprouts	*B. oleracea*	Linnaeus
	B. oleracea (western Europe)	Gates (1953)
	B. oleracea (western Europe)	Snogerup (1980)
	B. oleracea and *B. alboglabra*	Song et al. (1990)
Cabbage	*B. oleracea*	Linnaeus
	B. oleracea	de Candolle (1824)
	B. oleracea	Hedrick (1919) *
	B. oleracea	Bailey (1930)
	B. montana	Hegi (1919)
	B. oleracea (western Europe)	Gates (1953)
	B. oleracea (western Europe)	Snogerup (1980)
	B. oleracea and *B. alboglabra*	Song et al. (1990)
Cauliflower	*B. oleracea*	Linnaeus
	B. oleracea	de Candolle (1824)
	B. oleracea	Bailey (1930)
	B. montana	Hegi (1919)
	B. cretica	Schulz (1936)
	B. oleracea (from Cypus)	Giles (1941)
	B. cretica	Gates (1953)
	B. oleracea and *B. alboglabra*	Song et al. (1990)
	B. cretica	Tutin et al. (1964)
Kale	*B. oleracea*	Linnaeus
	B. oleracea	Hedrick (1919) *
	B. oleracea	Bailey (1930)
	B. montana	Hegi (1919)
	B. montana	Netroufal (1927)
	B. oleracea (western Europe)	Gates (1953)
	B. cretica, B. incana, B. rupestris	Snogerup (1980)
	B. incana and *B. insularia*	Hosaka et al. (1990)
	B. oleracea and *B. alboglabra*	Song et al. (1990)
Kohlrabi	*B. oleracea*	Linnaeus
	B. rupestris	Netroufal (1927)
	Unknown Mediterranean species	Gates (1953)
	B. oleracea and *B. alboglabra*	Song et al. (1990)

* edited observations by Sturtevant in the late 19th century, ** referring to Prof Buckman's experiment