

RESEARCH ARTICLE

Genotyping-by-sequencing highlights patterns of genetic structure and domestication in artichoke and cardoon

Stefano Pavan^{1,2}, Pasquale L. Curci^{3a,b}, Diana L. Zuluaga³, Emanuela Blanco³, Gabriella Sonnante^{3*}

1 Department of Soil, Plant and Food Science, University of Bari "Aldo Moro", Bari, Italy, **2** Institute of Biomedical Technologies, National Research Council (CNR), Bari, Italy, **3** Institute of Biosciences and Bioresources, CNR, Bari, Italy

^a Current address: Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium

^b Current address: VIB Center for Plant Systems Biology, Ghent, Belgium

* gabriella.sonnante@ibbr.cnr.it



OPEN ACCESS

Citation: Pavan S, Curci PL, Zuluaga DL, Blanco E, Sonnante G (2018) Genotyping-by-sequencing highlights patterns of genetic structure and domestication in artichoke and cardoon. PLoS ONE 13(10): e0205988. <https://doi.org/10.1371/journal.pone.0205988>

Editor: Kenneth M. Olsen, Washington University, UNITED STATES

Received: April 12, 2018

Accepted: October 4, 2018

Published: October 23, 2018

Copyright: © 2018 Pavan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. SNP genotype files are available at the European Variation Archive (EVA, <https://www.ebi.ac.uk/eva>), under the project PRJEB28606, accession numbers ERZ701172 (whole dataset), ERZ701171 (globe artichoke dataset), ERZ701173 (cultivated and wild cardoon dataset).

Funding: This work was funded by project BiodiverSO – PSR Puglia 2007-2013 Mis. 214/4 subaz. a) to GS, and project BiodiverSO – PSR

Abstract

Exploiting the biodiversity of crops and their wild relatives is fundamental for maintaining and increasing food security. The species *Cynara cardunculus* includes three taxa: the globe artichoke, one of the most important Mediterranean vegetables, the leafy cardoon, and the wild cardoon. In this study, genotyping by sequencing (GBS) was successfully applied to reveal thousands of polymorphisms in a *C. cardunculus* germplasm collection, including 65 globe artichoke, 9 leafy cardoon, and 21 wild cardoon samples. The collection showed a strong population structure at $K = 2$, separating the globe artichoke from the leafy and wild cardoon. At higher K values, further substructures were observed, in which the wild cardoon was separated from the leafy cardoon, and the latter included the Spanish wild cardoons, while the wild sample from Portugal was admixed. Moreover, subpopulations within the globe artichoke set were highlighted. Structure analysis restricted to the globe artichoke dataset pointed out genetic differentiation between the "Catanesi" typology and all the other samples ($K = 2$). At higher values of K , the separation of the "Catanesi" group still held true, and green headed landraces from Apulia region, Italy ("Green Apulian") formed a distinct subpopulation. "Romaneschi" artichoke types fell in a variable group with admixed samples, indicating that they should not be considered as a genetically uniform typology. The results of principal component analysis and Neighbor-Joining hierarchical clustering were consistent with structure results, and in addition provided a measure of genetic relationships among individual genotypes. Both analyses attributed the wild material from Spain and Portugal to the cultivated cardoon group, supporting the idea that this might be indeed a feral form of the leafy cardoon. Different reproductive habit and possibly selective pressure led to a slower LD decay in artichoke compared to cardoon. Genotyping by sequencing has proven a reliable methodology to obtain valuable SNPs and assess population genetics in *C. cardunculus*.

Puglia 2014-2020 Mis.10.2 to GS. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: GBS, Genotyping by sequencing; MCMC, Markov Chain Monte Carlo; LD, linkage disequilibrium; SVS, SNP & Variation Suite; MAF, Minor allele frequency; PCA, Principal component analysis; EM, Expectation-maximization; F_{ST} , Fixation index; UPOV, International Union for the Protection of New Varieties of Plants.

Introduction

The analysis of genetic variation of crop species and their wild relatives is a crucial aspect for biodiversity conservation and its exploitation to maintain and increase food security [1]. Nowadays, the exploration of plant biodiversity is boosted by advanced sequencing technologies, which provide the opportunity to simultaneously discover and test a high number of molecular markers at a relatively low cost. In particular, genotyping by sequencing (GBS), based on high-throughput sequencing of a reduced complexity genome library generated by restriction enzymes, proved to be cost-effective in the development and genotyping of thousands of single nucleotide polymorphism (SNP) markers. So far, GBS has been extensively used in species with or without an available reference genome [2, 3] for linkage map construction [4, 5], genomic selection [6], and the exploration of plant genetic diversity on a genome-wide scale [3, 7–9].

The globe artichoke [*Cynara cardunculus* var. *scolymus* (L.) Fiori] is conspecific to and interfertile with the cultivated leafy cardoon (*C. cardunculus* var. *altalis* DC), and the wild progenitor of the two crops (*C. cardunculus* var. *sylvestris* Lam.). Therefore, the latter two entities belong to the artichoke primary gene pool [10, 11, 12, 13]. The globe artichoke, a diploid plant ($2n = 2x = 34$) with an estimated genome size of 1.07 Gb [14], is a typical Mediterranean vegetable crop, mainly consumed for its immature flower heads and traditionally cultivated in southern Europe and Northern Africa, with a more recent diffusion in California, South America, and China. Both heads and leaves are rich in antioxidant phenolic compounds such as flavonoids, caffeic acid, chlorogenic acid and cynarin, and for this reason, artichoke plants are used in herbal and medicinal preparations [15–17]. Moreover, artichoke organs, including roots, contain a high amount of inulin molecules [18] with a chain length of up to 200 [19], having a marked and well tolerated prebiotic effect in humans and useful for industrial applications, e.g. for inhibiting the growth of ice crystals in frozen foodstuffs [20].

In general, the globe artichoke is clonally reproduced to ensure commercial uniformity [21], although seed-propagated varieties have been released in recent years [22]. The morphological diversity in head shape and colour, coupled with flowering time, led to the distinction of four main groups, namely “Catanesi”, “Romaneschi”, “Spinosi”, and “Violetti” [23]. However, not all the traditional varieties fall in these groups, and molecular markers have often failed to attribute a specific membership for some of them [21, 24]. Italy possesses the richest germplasm diversity of globe artichoke, resulting in the local cultivation of many landraces, well adapted to local climatic conditions [12, 23, 25]. Leafy cardoon is grown on a small scale in northern Italy, southern France and in Spain for its large leaf stalks [11].

All the genus *Cynara*, and the species *C. cardunculus* in particular, originated in the Mediterranean area. Probably, the globe artichoke was domesticated in Sicily, while the cultivated cardoon (CC) followed a distinct domestication pathway possibly in the Iberian Peninsula and the South of France [10, 11]. In fact, the leafy cardoon was selected for the gigantism of leaf stalks, which are used as vegetables to prepare traditional dishes [11, 21]. The wild cardoon, the progenitor of both *C. cardunculus* crops, is distributed across the Mediterranean Region, from Cyprus and the Black Sea to Atlantic Spain, Portugal and the Canary Islands [26]. Its small flower heads are traditionally gathered from the wild and used as a food in southern Italy [27]. Although two wild forms with distinct morphology and distribution (eastern and western-Mediterranean) were distinguished [11, 24, 26], SSR analysis and population structure suggested that the western wild cardoon could be a feral form and not a real wild type [21].

Recently, a draft of the globe artichoke genome was obtained by Illumina sequencing, producing 13,588 scaffolds covering 725 Mb of the genome [28], and resequencing was performed on four artichoke genotypes and one cultivated cardoon [29]. The availability of a reference

genome allows the mapping of short reads obtained from GBS analysis to the genome, allowing the application of accurate SNP calling pipelines [30].

In the present study, we used for the first time the GBS approach to explore genome-wide SNP variation in a *C. cardunculus* collection including globe artichoke traditional varieties, cultivated leafy cardoon and wild cardoon. The aim of this work was to detect population genetic structure, patterns of genetic diversity, and relationships in artichoke and its conspecific allies.

Material and methods

Plant material

A group of 95 *C. cardunculus* samples was prepared for this study, including 65 artichokes, 9 cultivated cardoons, and 21 wild cardoons (S1 Table). Samples (artichoke and wild material) were obtained from the *Cynara* collection held at the Institute of Biosciences and BioResources (CNR, Bari, Italy) or from exchange with other institutions (GEVES, Cavaillon France; Botanic Institute, CSIC, Barcelona, Spain; ITGA, Navarra, Spain; University of Tunis, Tunisia). For the artichoke samples, the four main morpho-agronomic groups were considered: the small headed early types ("Catanesi"), the spiny early types ("Spinosi"), the violet late types ("Violetti"), and the large headed late types ("Romaneschi") [23]. Several other accessions, belonging to none of the above typologies, were analysed [12, 31]. The cultivated cardoon germplasm included material originating from the main regions where this vegetable is grown, i.e. Spain, France, and Italy. The wild material was selected in order to encompass the geographical distribution of the taxon.

GBS assay and SNP filtering

Genomic DNA was isolated from young leaves as previously described [32]. A *Pst*I library was prepared at the Institute for Genomic Diversity at Cornell University (<http://www.biotech.cornell.edu/>), according to the method of Elshire et al. [2], except for the use of *Pst*I enzyme and corresponding adapters. The library was sequenced at the Cornell University (Ithaca, U.S.A.), with the Illumina technology using a HiSeq2500 in high output mode (single-end 100 bp reads), as a single lane including an empty, negative control well. The sequencing reads were searched for 100% matching barcode with the expected bases remnant of the enzyme cut site. The reads containing the barcode were sorted, de-multiplexed and trimmed to first 64 bases starting from enzyme cut site. After this step, the reads containing "N" within the first 64 bases were rejected. The TASSEL-GBS pipeline [33], implementing the globe artichoke genome assembly (NCBI SRA project number PRJNA238069) and the Burrows-Wheeler Aligner (version 0.7.8-r455) algorithm [34], was used to call SNPs and generate a vcf variant file. Minor allele frequency (MAF) higher than 5%, call rate higher than 50%, and Hardy-Weinberg equilibrium p-value $>10^{-6}$ [35] was applied for further filtering biallelic SNPs, using TASSEL v5.2.20 [36] and SNP & Variation Suite (SVS) software v8.4.0 (Golden Helix Inc., Bozeman, MT, U.S.A.). Genotypes with more than 30% missing data were removed from further analysis. The filtering procedure was applied to the whole germplasm collection and to separate subsets including either globe artichokes or cardoons. The VCFtools package [37] and TASSEL5.2.20 [36] were used to derive the distribution of SNP substitution types and the proportion of heterozygous loci.

Population structure analysis, genetic diversity and relationships

A structure analysis was carried out using the Bayesian clustering approach and the admixture model implemented in the software STRUCTURE (ver. 2.3.4) [38]. For a number of hypothetical subpopulations (the K parameter) varying from 2 to 10, ten independent runs were performed

applying 50,000 burn-in period and 100,000 Markov Chain Monte Carlo (MCMC) repetitions. Prior to analysis, the SNP dataset was pruned based on pairwise linkage disequilibrium (LD) between adjacent markers, using SVS v8.4.0 and setting the threshold for r^2 equal to 0.5.

The most probable K value was determined by *ad hoc* ΔK statistics [39], using the software Structure Harvester [40]. Individual samples were assigned to each subpopulation when the value of the corresponding membership coefficient (q) was higher than 0.75, otherwise they were considered admixed.

Expected heterozygosity between individuals in the same cluster and allele frequency divergence, "net nucleotide distance" between clusters, were obtained from STRUCTURE analysis.

SNPs selected as described above were also used to evaluate genetic relationships among genotypes within the whole dataset and the subsets of artichokes and cardoons. In particular, principal component analyses (PCA) were carried out using SVS v8.4.0., whereas Neighbor-Joining (NJ) [41] trees, based on the Tamura-Nei genetic distance model [42], were built after performing five hundred bootstrap replicates using MEGA7 package (<https://www.megasoftware.net/>).

LD decay

The extent of LD decay was assessed for the whole collection, for the globe artichoke dataset, and for the cardoon group, separately. Pairwise r^2 values were estimated using the expectation-maximization (EM) algorithm implemented in SVS v8.4.0 and then plotted against the distance (kb) between adjacent SNP loci.

Results

GBS experiment and SNP calling

Sequencing of a 95-plex GBS library of *C. cardunculus* yielded about 259 million reads. Of these, about 230 million were considered of good quality, exhibiting both a full barcode and the expected remnant of the restriction cut site [2]. Tags (unique sequences of 64 bp following the barcode) occurring at least three times were over 2.3 million, of which 27.6% were successfully mapped (25.9% uniquely and 1.7% multiply), while 72.4% remained unaligned. A master variant calling file with 22,883 SNPs was produced with the TASSEL-GBS pipeline. After filtering SNPs and samples as described in the Methods section, a final dataset of 92 samples and 3,762 SNPs was obtained (S2 Table). Filtering based on Hardy-Weinberg equilibrium ($p > 10^{-6}$) successfully eliminated highly heterozygous loci, which are expected in case of sequencing errors or paralogous loci (S1 Fig). The observed transition/transversion rate was 1.24. The most abundant substitution type was A/G with 28.7%, while the lowest represented class was C/G with 4.9% (S2 Fig).

The filtering procedure was repeated on sample subsets containing just the globe artichokes or the wild cardoons, leading to the identification of 4,711 (62 genotypes) and 6,668 (30 genotypes) SNPs markers, respectively (S2 Table).

Mean observed heterozygosity was 31.2% for the globe artichoke collection, and 15.2% for the cardoon dataset.

Genetic structure

The genetic structure of the *C. cardunculus* collection was investigated by means of the admixture model implemented in the software STRUCTURE. The Evanno delta K method identified the best model as the one containing two subpopulations ($K = 2$), followed by those with $K = 3$ and $K = 4$ (S3 Fig). Based on the $K = 2$ model, genotypes cluster according to their taxonomic distribution, with groups corresponding to globe artichoke samples on one side, and wild and

cultivated cardoons on the other side (Fig 1A). The individual membership coefficient to each population (q) is very high, most cardoons and artichokes showing a “q” above 0.9 and 0.85, respectively. However, the Portuguese cardoon sample (W_LK_8886) and seven artichoke samples display a “q” value below 0.75, indicating an admixed ancestry. The artichoke sample with the lowest “q” value (0.63) corresponds to the landrace “Di spine Ostuni”. Although this genotype is considered as a globe artichoke by the donor farmer, it has a flower head morphology resembling cultivated cardoons, and might therefore have derived from a hybridization event between the two taxa.

Since the Evanno method can underestimate K when there is hierarchical population structure [43], we also considered STRUCTURE models with K>2. At K = 3 (Fig 1B), wild and cultivated cardoons are still joined in the same group, and the Portuguese cardoon appears admixed (q = 0.62). The globe artichokes are split into two main groups, one including the “Catanesi” types (q = 0.99–1.00), and the other containing green headed artichokes from Apulia region in Southern Italy (“Green Apulian” artichokes, with q = 1.00) and other genotypes

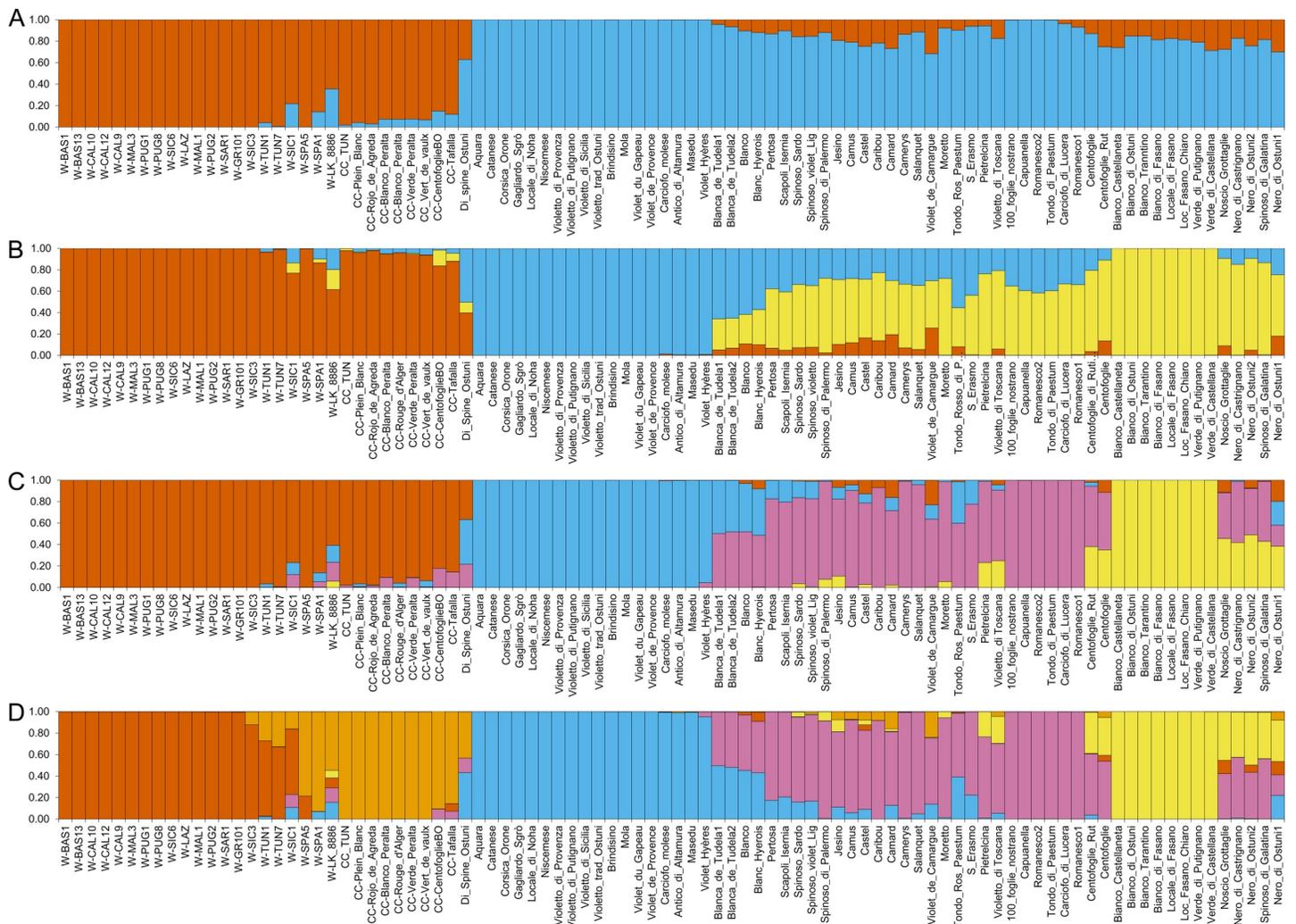


Fig 1. Structure analysis of *Cynara cardunculus* genotypes. Clusters inferred from the population structure analysis at (A) K = 2, (B) K = 3, (C) K = 4, (D) K = 5. Numbers on the y-axis show the subgroup membership. Genotype names are reported on the x axis. The colours of the bars indicate the groups identified through the STRUCTURE program.

<https://doi.org/10.1371/journal.pone.0205988.g001>

(q ranging from 0.75 to 0.86). Moreover, many admixed genotypes can be observed, including "Romaneschi", "Spinosi", most of the French artichokes, and Blanca de Tudela. In the $K = 4$ solution (Fig 1C), besides the clusters observed at $K = 3$, a distinct artichoke group arises, composed of "Romaneschi", "Spinosi" and "Violetti" varieties (dark pink in Fig 1C). The $K = 5$ outcome separates the wild (dark orange in Fig 1D) from the cultivated (light orange in Fig 1D) cardoons, with the wild cardoons from Spain included in the group of cultivated cardoons. Some wild genotypes are admixed, namely W_LK_8886 (mostly between CC and artichoke groups), one cardoon from Sicily (with a higher " q " value for the wild group and lower contributions of CC and artichoke groups), and two cardoons from Tunisia (with a higher " q " value for the wild group and some CC contribution). The artichoke samples are organized in groups resembling those for $K = 4$, however several genotypes are admixed. Some of them share a higher " q " value for the Romaneschi/Spinosi/Violetti group (e.g. Camard, Castel, Jesino, Violetto di Toscana), whereas others display higher degree of admixture, such as Blanca de Tudela, sharing approximately equal ancestry between the "Catanesi" and the "Romaneschi" groups. At $K = 6$ (S4 Fig), some "Romaneschi" genotypes (100_foglie, Capuanella, Carciofo di Lucera, Romanesco, Tondo di Paestum) form a cluster distinct from other "Romaneschi" types (e.g. Camus, Castel, Jesino, etc.), which are grouped with the "Spinosi" types and the landrace Violetto di Toscana. Increasing the K value to 7 (S4 Fig) leads to the separation of some artichoke genotypes originating from the Apulia region, Italy (Carciofo noscio locale, Nero di Ostuni, Nero di Castrignano, Spinoso di Galatina) which are all dark in colour, but with two different head shapes. In particular, Carciofo noscio locale and Nero di Ostuni have a cylindrical flower head, while the others show a very particular flower head with everted bracts. We have attributed the generic name of "Nero del Salento" ("Black from Salento") to this last typology (<https://biodiversitapuglia.it/carciofo-nero-del-salento-varieta-locale-tutto-particolare/>), Salento being a geographical area in the south of Apulia region. To our best knowledge, this local variety is present only in Salento, where it is quite diffused, although it is mainly grown in family gardens for local consumption, and is difficult to find in stores or markets.

In order to achieve a better picture of the globe artichoke genetic structure, the analysis was also run on a subset of the original samples containing only the globe artichoke genotypes. In this case, the best fitting model was represented again by $K = 2$ (Fig 2A), which separates the "Catanesi" artichokes on one side (light blue), and all the other typologies on the other side (yellow), with several admixed samples. The second and the third best K values were $K = 3$ and $K = 4$, respectively. At $K = 3$ (Fig 2B), the first group is composed of the "Catanesi" types (light blue). The second group (mainly dark pink) includes the "Romaneschi" *sensu lato* (except for Tondo Rosso di Paestum and the two Centofoglie genotypes, which are admixed), violet and spiny types. The two Blanca de Tudela genotypes, Blanc Hyerois and Blanco are admixed between the first two groups. The third group includes the "Green Apulian" artichokes (yellow). Some other local varieties are admixed between groups 2 and 3. At $K = 4$ (Fig 2C), the "Catanesi" (light blue) and the "Green Apulian" (yellow) groups of the previous clustering at $K = 3$ are generally maintained, while the other large $K = 3$ group is split into two smaller sets with a higher number of admixed samples. Of these, one (dark pink) includes the "Romaneschi" *sensu stricto* (Romanesco1, Romanesco2, Capuanella, Tondo di Paestum, Carciofo di Lucera, 100_foglie_nostrano), while the second group (dark blue) is composed of some other "Romaneschi" *sensu lato* genotypes (Jesino, Pertosa, Scapoli-Isernia, Camus, Castel, Camard), together with "Spinosi" types and Violetto di Toscana. The other two "Violetti", namely Morretto and S. Erasmo are admixed, with a membership coefficient for this group of 0.70 and 0.56, respectively. Several other admixed genotypes are observed. At $K = 5$ (Fig 2D), the group of dark artichokes from Apulia separates (green). In the $K = 6$ solution (S4B Fig), besides the previous groups, the "Romaneschi" types are further divided and form three subsets, one of which

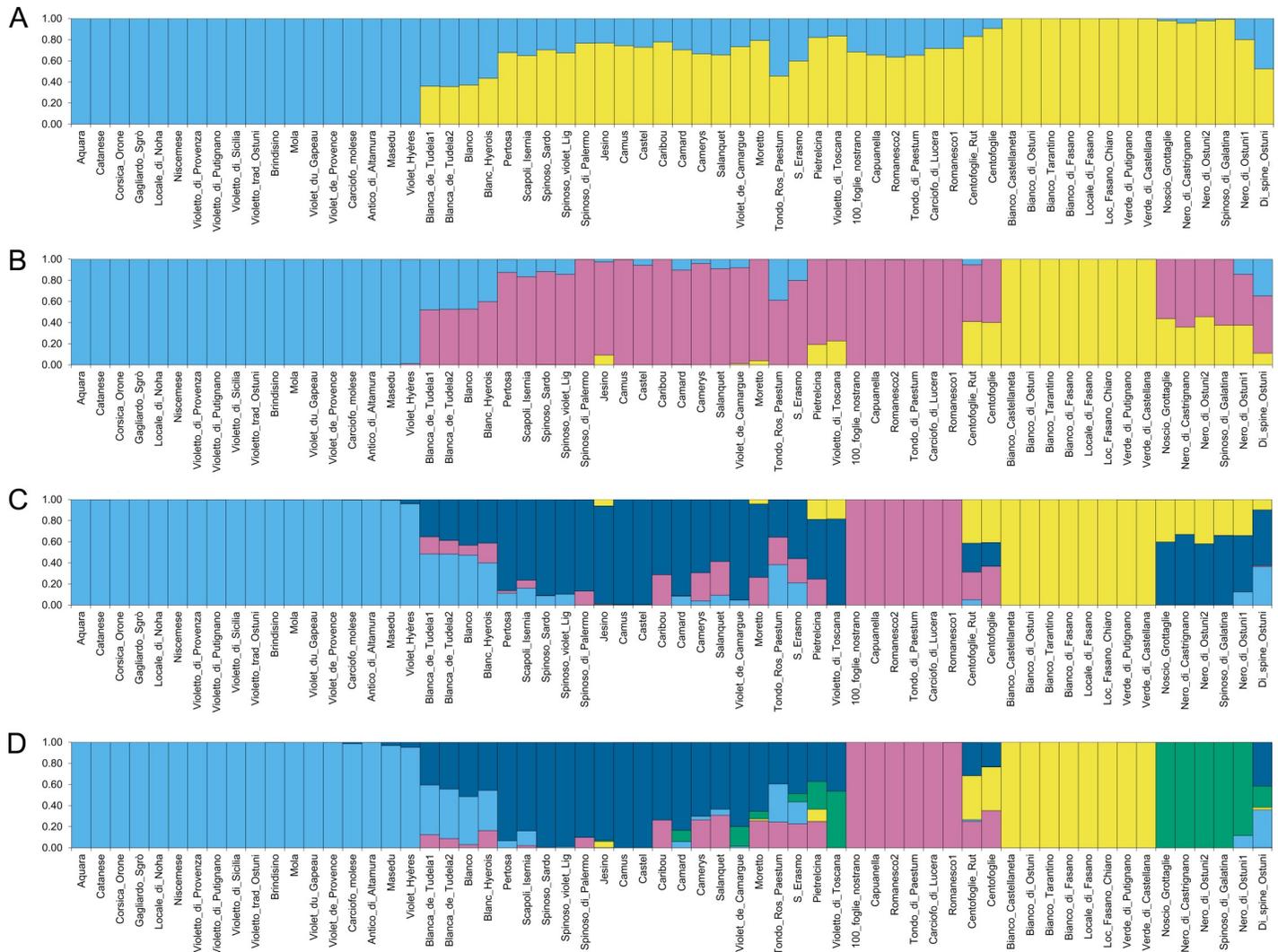


Fig 2. Structure analysis of globe artichoke genotypes. Clusters inferred from the population structure analysis at (A) $K = 2$, (B) $K = 3$ and (C) $K = 4$, (D) $K = 5$. Numbers on the y-axis show the subgroup membership. Genotype names are reported on the x axis. The colours of the bars indicate the groups identified through the STRUCTURE program.

<https://doi.org/10.1371/journal.pone.0205988.g002>

includes the two French Camus and Castel varieties (orange), while the other French “Romaneschi” are admixed (“q” value ranging from 0.67 to 0.72). The second “Romaneschi” group (dark pink) is composed of the “Romaneschi” *sensu stricto*, while the third one (dark blue) includes two large-headed green genotypes (Pertosa and Scapoli_Isernia) together with “Spinosi” types. At $K = 7$ (S4B Fig), the two genotypes of Blanca de Tudela and Blanc Hyerois, which were previously admixed, form a separate group together with Blanco (dark orange).

Concerning the cardoon dataset, the most probable model was obtained for $K = 2$ (S3 Fig), where the two groups correspond to cultivated and wild cardoons. The leafy cardoon group also includes the wild cardoons from the Iberian Peninsula. Two genotypes of wild cardoon from Tunisia, and one from Sicily are admixed, with a higher “q” value for the wild group. Increasing the number of K , the wild cardoon group splits into smaller clusters according to the geographical origin of the samples, while the cultivated cardoons remain as a single group, and the number of admixed samples increases (S5 Fig). It is interesting to notice that, at $K = 8$,

the second most probable K (S3 Fig), wild genotypes from the different countries are separated from one another, although each geographical area is represented by a limited number of genotypes. Among the Italian wild cardoons, a regional subdivision can be observed, with genotypes from Apulia and Basilicata placed in the same group, samples from Sardinia and Latium in another small group, and genotypes from Sicily and Calabria sharing admixed ancestry. Moreover, the wild cardoon from Portugal and one of the wild cardoons from Spain separate from the group of cultivated cardoons.

In *C. cardunculus*, the expected heterozygosity obtained within STRUCTURE clusters at $K = 2$ revealed a lower value in cardoons (0.160), compared to the globe artichoke, which showed an expected heterozygosity of 0.216. Allele frequency divergence, or "Net nucleotide distance" between clusters was obtained for the whole dataset at $K = 5$, where wild and cultivated cardoons were separated, and the globe artichoke samples formed three groups. Net represents the average of pairwise difference between alleles from different groups, excluding the amount of variation located within each group. Similar groups have distances approaching to 0. In our study, the lowest divergence (0.086) was observed between wild and cultivated cardoons, while the highest Net was detected between two groups of artichokes, "Catanesi" and "Green Apulian" and between "Catanesi" and wild cardoons, with similar values, 0.202 and 0.201, respectively.

Genetic relationships

In order to assess genetic relationships among *C. cardunculus* genotypes, PCA and Neighbour-Joining clustering were performed. For both analyses, three datasets, referring to i) the complete *C. cardunculus* collection, ii) artichokes, and iii) cardoons, were used separately. In the 3D PCA plot for the whole dataset (Fig 3), wild cardoon, cultivated cardoon and globe artichoke samples are differently positioned. Globe artichoke samples (blue) form a quite dispersed cloud. Some samples, namely the wild cardoon from Portugal and the two wild cardoons from Spain, appear quite distant from the other genotypes belonging to the same taxonomic entity. In particular, the latter ones are closer to the leafy cardoons than to the other wild cardoons, while the wild cardoon from Portugal is placed between cultivated cardoons and globe artichokes. The eastern wild material shows a quite sparse distribution, indicating a degree of differentiation among genotypes originating from various geographical areas. The PCA plot restricted to the globe artichoke germplasm (Fig 4A) shows two main distinct groups: the "Catanesi" (bottom right), and the "Green Apulian" (bottom left) artichokes. The remaining genotypes are scattered in the central-upper part of the plot, and no clear clustering can be observed for them based on the variety, except for some of the "Romaneschi" types in the upper part. The PCA for the cardoons (Fig 4B) separates the quite compact group of cultivated cardoons (bottom right) from wild (left) cardoons. Exceptions are represented by wild cardoons from the Iberian Peninsula, placed together with the cultivated cardoons. The eastern wild cardoons are spread in the left part of the graph, indicating a quite high differentiation of this group originating from three European countries and Tunisia.

The Neighbor-Joining tree obtained from the whole dataset (Fig 5) highlights two main branches, one including the globe artichokes, and the other the rest of the material. This second branch is subdivided into two forks: wild cardoons on one side and cultivated cardoons on the other, this latter group also comprising the wild cardoons from Spain and Portugal. The land-race Di spine Ostuni is placed between the cardoon and the artichoke groups, thus substantiating the idea that this is a hybrid type. Within the artichoke cluster, subgroups can be identified, three of which are compact, i.e. the "Catanesi", the "Green Apulian", and some of the "Romaneschi" types (Romanesco1, Romanesco2, Capuanella, 100_Foglie_Nostrano, Tondo di Paestum, Carciofo_di_Lucera, and Caribou). The admixed "Tondo Rosso di Paestum" artichoke is

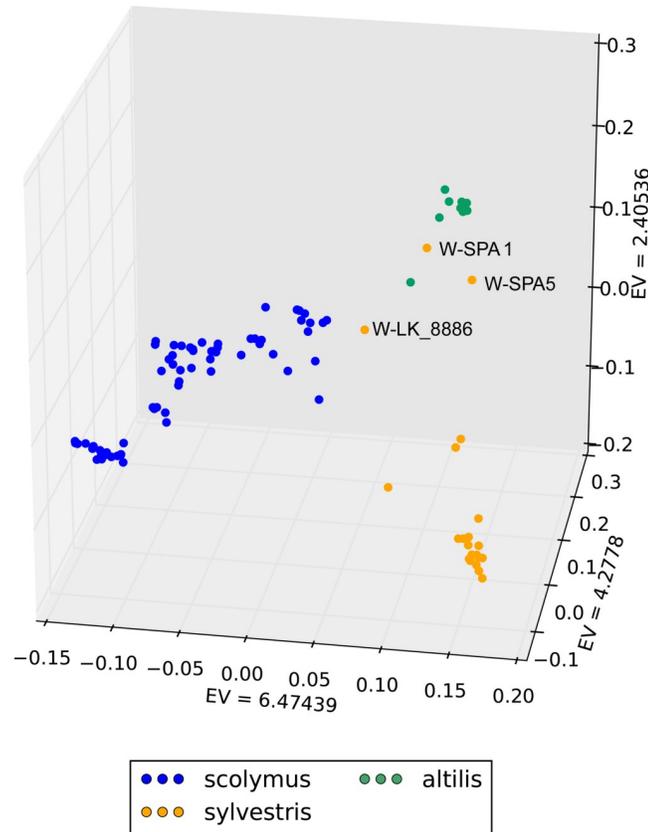


Fig 3. Graph of the first three axes from a Principal component analysis (PCA) of the 92 *Cynara cardunculus* genotypes analysed. EV: eigenvalue. Blue: globe artichoke; green: cultivated cardoon; orange: wild cardoon.

<https://doi.org/10.1371/journal.pone.0205988.g003>

included in the “Catanesi” group, although at a high distance. The other genotypes are quite scattered in the artichoke branch, even though small subsets can be recognised: the Apulian dark-headed artichokes (Noscio Grottaglie, Nero di Ostuni, Spinoso di Galatina, Nero di Castignano), the “Spinosi” types (Spinoso di Palermo, Spinoso Violetto di Liguria, Spinoso Sardo). The two Blanca de Tudela genotypes, together with Blanco, are positioned between “Catanesi” and the other artichoke types. When performing the clustering analysis only with the globe artichoke dataset (Fig 6), the main groups observed are again “Catanesi” and smaller clusters, such as the “Green Apulian”, some of the “Romaneschi”, “Spinosi”, and the previously identified Apulian dark-headed artichokes. The phylogenetic tree constructed for the cardoons (S6 Fig) is divided into two clusters: the wild and the cultivated cardoons. Also in this case, the Iberian wild cardoons are included in the group of cultivated cardoons.

LD decay

To assess the extent of LD decay, the estimate of r^2 for all pairs of SNP loci linked on the same genome scaffold, was calculated. The LD decay was evaluated for the whole dataset (including wild and cultivated material), for the globe artichoke group, and for the cardoon collection separately. The LD decay observed for the whole *C. cardunculus* array was quite rapid, with $r^2 = 0.2$ after 0.35 kb (Fig 7). When the analysis was carried out on the two subsets of the germplasm collection, a different behaviour was observed, as LD decay in globe artichokes ($r^2 = 0.2$ after 0.92 kb) was considerably slower compared to that of cardoons ($r^2 = 0.2$ after 0.05kb).

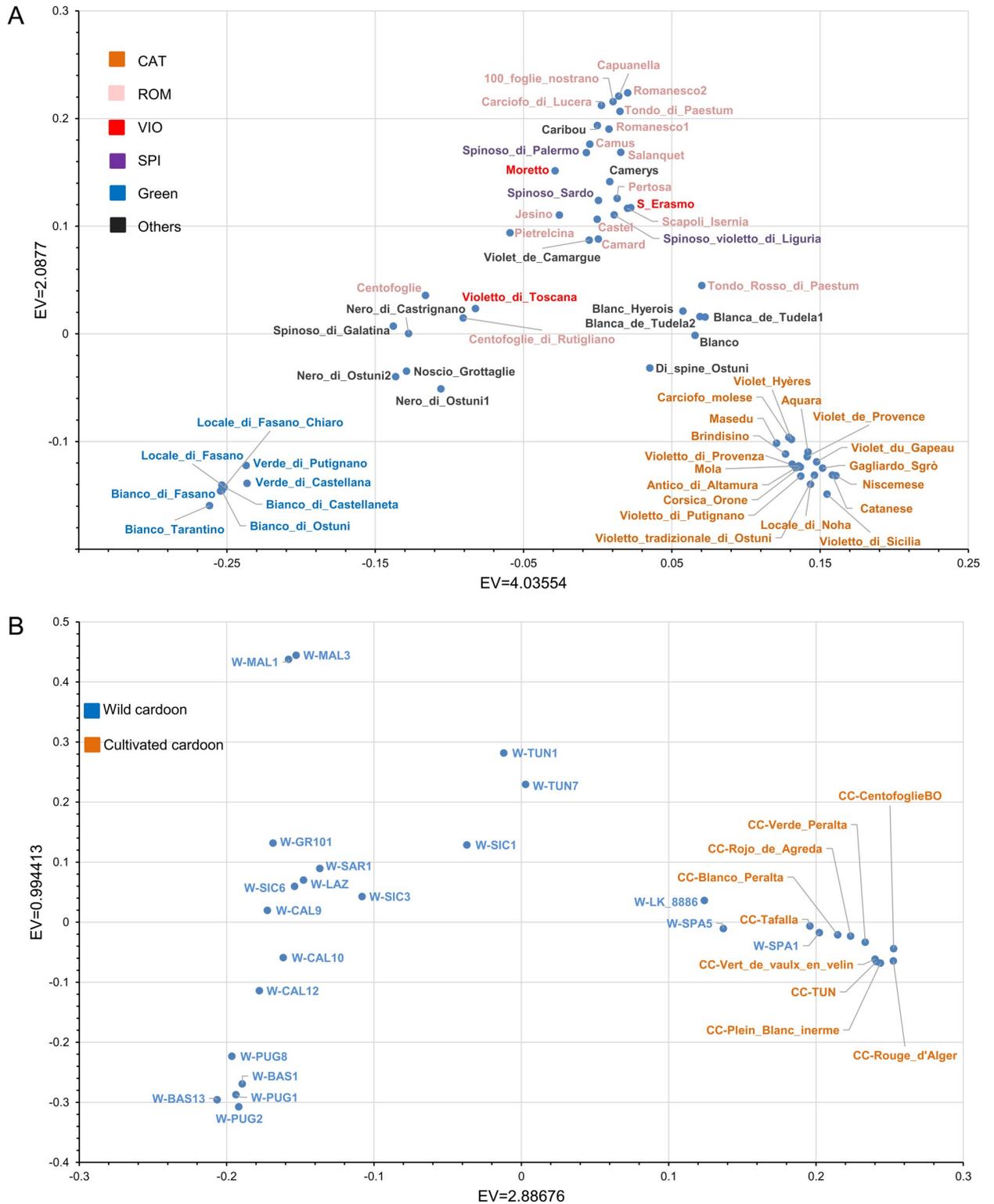


Fig 4. Graph of the first two axes from a Principal component analysis (PCA) of (A) 62 globe artichoke and (B) 30 wild and cultivated cardoon genotypes. EV: eigenvalue. Colors indicate globe artichoke types (A), or wild or cultivated cardoon (B). CAT: Catanesi; ROM: Romaneschi; VIO: Violetti; SPI: Spinosi; Green: Green Apulian artichokes; Others: other types.

<https://doi.org/10.1371/journal.pone.0205988.g004>

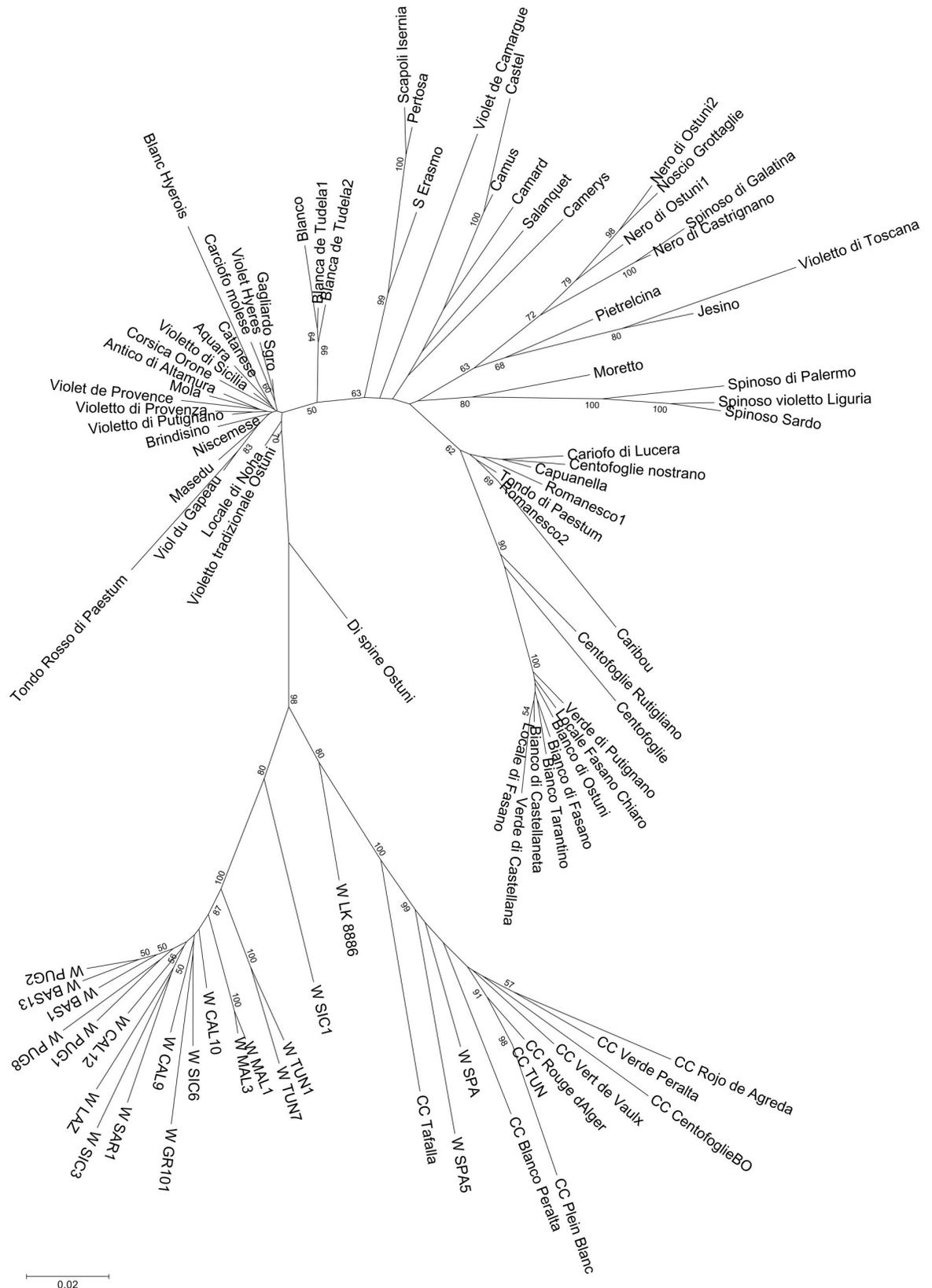


Fig 5. Neighbor-Joining tree obtained from SNP data on the whole dataset of *Cynara cardunculus* genotypes. Number on tree branches indicate bootstrap values (≥ 50).

<https://doi.org/10.1371/journal.pone.0205988.g005>

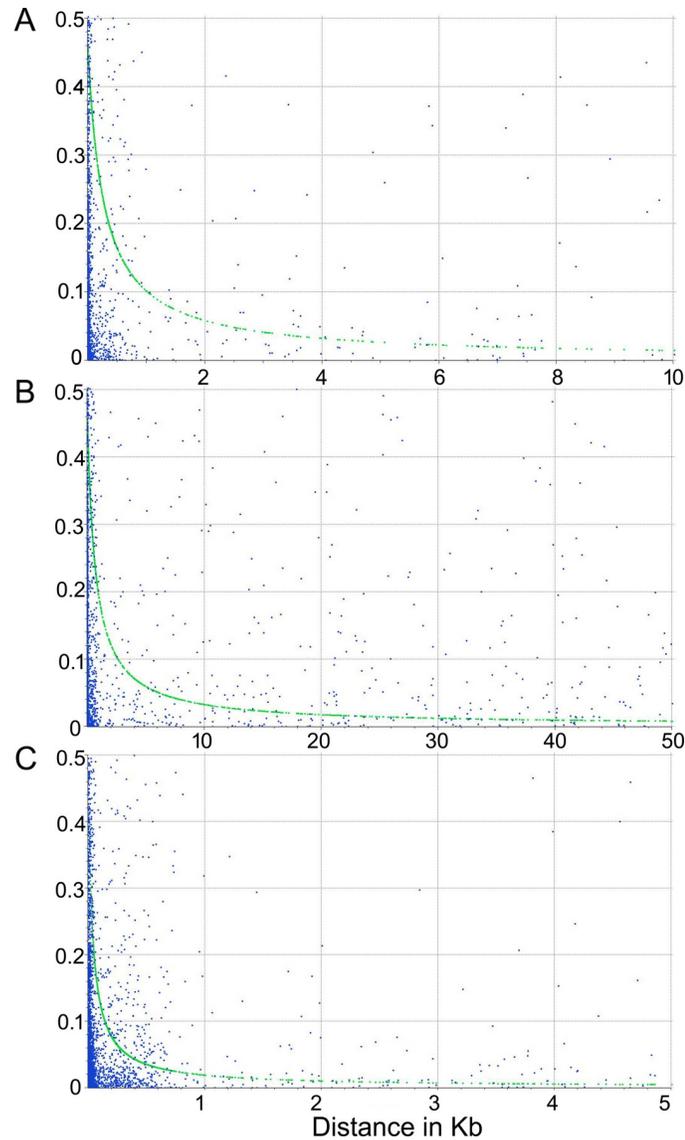


Fig 7. LD decay for (A) the whole dataset of *Cynara cardunculus* genotypes, (B) the globe artichoke subset, (C) cardoon samples. LD decay determined by squared correlations of allele frequencies (r^2) against distance between polymorphic sites.

<https://doi.org/10.1371/journal.pone.0205988.g007>

sequencing was focused on parents of mapping populations [44]. GBS performed with rare-cutter enzymes, such as *Pst*I used in this study, are generally useful to increase sequence depth [45]. This fits with the need of our study, requiring high depth for efficiently calling heterozygous genotypes, frequently occurring in an outcrossing species such as *C. cardunculus*, at multiplexed level. The number of tags mapped on the globe artichoke genome was not high, and this might be due to the draft-quality of the reference genome available, which was organized in scaffolds. However, even applying severe filtering parameters (mostly $MAF > 0.05$ and p -value for HWE $> 10^{-6}$), we detected thousands of SNP polymorphisms for downstream analyses.

Analysis of SNP data was performed on three separate datasets: i) the whole *C. cardunculus* collection; ii) globe artichoke samples; iii) wild and leafy cardoon samples. STRUCTURE

analysis provided $K = 2$ as the most probable number of populations for the whole dataset, splitting the cardoons (wild and cultivated) from the globe artichokes. For K values from 2 to 4, cultivated cardoon grouped together with wild cardoon, suggesting that cultivated cardoons retain a higher proportion of wild alleles. According to Dempewolf et al. [46], both *Cynara* crops can be regarded as semi-domesticated plants; indeed globe artichoke and leafy cardoon derive from two domestication events leading to larger flower heads or leaf stalks, respectively [11]. However, differently from globe artichoke, leafy cardoon shows the persistence of ancient wild traits, related to domestication, such as a higher number of plant branches, small capitula, etc. [11, 21].

Based on the present SNP analysis, the cultivated cardoon cluster also includes wild cardoon genotypes collected in Spain, corroborating the idea that this material might indeed represent a feral form of leafy cardoons [21]. This is also supported by phenotypic traits of these plants, which resemble cultivated cardoons as they show few or no thorns, large and tall size, and a habit and morphology different from that of other wild material (Sonnante, personal observations).

At $K = 5$, the highest divergence among groups is observed between the "Catanesi" and the "Green Apulian" artichoke and between the "Catanesi" and the wild material. The separation of the "Catanesi" group is also highlighted by Neighbor-Joining hierarchical clustering and PCA analysis. "Catanesi" possess a typical trait, in common with a few other varieties (e.g. Blanca de Tudela), that is earliness or re-flowering (they bloom in autumn and in spring). This important agronomic trait might have been selected in more recent times, leading to a sharp separation of "Catanesi" from the other artichokes and from the wild cardoons, which are also late-flowering plants since they bloom in spring. Additionally, the high homogeneity observed within the "Catanesi" group suggests that these artichokes may derive from the same genetic material, which probably, as the name indicates (from Catania, a town in Sicily), originated in Sicily. Later on, this variety possibly diffused, especially in the Apulia region (Italy) and in other areas (e.g. France), where it has given rise to a number of similar local varieties (e.g. Brindisino, Mola, Violet de Provence, etc.).

The "Romaneschi" types appear quite a heterogeneous group, and this can be perceived from the results of STRUCTURE, NJ clustering and PCA analyses. Only some of the "Romaneschi" genotypes are grouped together, while the others are scattered. This output is consistent with previous results obtained using different molecular markers [21, 24, 47] and suggests that the artichoke morpho-groups are not always substantiated by the genetic composition and relationships between genotypes. The difficulty in recognizing a precise definition for the four traditional morpho-groups, apart from the "Catanesi" and a few "Romaneschi", might suggest that just a few genes control those morphological traits determining the differences among groups. Especially for "Romaneschi" types, the circular or large transverse elliptical forms as described in the UPOV (International Union for the Protection of New Varieties of Plants) descriptors for artichoke (<http://www.upov.int/edocs/tgdocs/en/tg184.pdf>), do not identify a homogeneous group (usually called "Romaneschi"), but diverse artichokes with a variable genetic background.

Blanca de Tudela is a traditional widespread Spanish artichoke variety; it is green-coloured and, similarly to the "Catanesi" artichokes, early flowering. Following STRUCTURE analysis, Blanca de Tudela genotypes appear admixed between the "Catanesi" types and other typologies. Accordingly, in the NJ tree as well as in the PCA plot, this artichoke type is placed between "Catanesi" and the other artichokes. With both dominant and SSR markers, Blanca de Tudela artichoke had already shown this intermediate position [21, 47]. Therefore, we suppose that Blanca de Tudela might have derived from a cross between a "Catanesi" type artichoke with another typology.

The “Green Apulian” genotypes represent a compact group that is quite distinguishable from the other globe artichokes. To our best knowledge, these artichokes, as well as the “Nero del Salento” typology (see above) are grown just in the Apulia region (Southern Italy). It has been noted that Southern Italy, and especially the Apulia region, represents a centre of diversity for many crops, where still traditional farming systems can be found and ancient and diversified crops are cultivated [8, 9, 12, 48, 49].

The SNP analyses highlighted that globe artichoke retains a higher level of heterozygosity compared to wild and cultivated cardoon. This result is in agreement with previous studies carried out with SSR markers [12, 21, 47] and with a recent genome resequencing study [29]. Clonally propagated crops are generally outcrossing plants, and inbred individuals usually show lower vigour deriving from inbreeding depression [50]. It is possible that highly heterozygous globe artichoke genotypes were selected to maximize heterotic effects, which have been then maintained by farmers by clonal propagation (indeed, all the globe artichoke germplasm analysed in this study is clonally propagated). In fact, sexual recombination can knock down advantageous genetic combinations, while clonal multiplication preserves them [51]. Another, not mutually exclusive, hypothesis for high heterozygosity levels is that clonal lineages are likely to accumulate slightly deleterious mutations, differently from sexually propagated cardoons in which such mutational events are eliminated by recombination or selection.

Linkage disequilibrium is the non-random association of alleles at different loci, and is influenced by various factors. For instance, domestication, population subdivision, selection can enhance LD in the genome [52]. Generally, LD decays faster in outcrossing species than in self-fertilizing plants, but outcrossing vegetatively propagated plants are an exception [53]. The difference in LD decay between the globe artichoke group and the cardoon cluster might be related to the different reproductive and propagation system of the taxa within *C. cardunculus*, and to domestication and selective pressure preserving specific haplotype blocks. Although all three taxa are outcrossing, the clonally propagated globe artichoke shows a slower LD decay compared to wild and cultivated cardoon, probably due to the few recombination events happened during its long breeding cycle. A relatively slow LD decay has been observed in other perennial or clonally propagated crops, such as sugarcane and potato [54, 55]. Moreover, domestication events and selective pressure on improved material have been shown to decrease LD decay passing from wild relatives to landraces to modern cultivars in some species such as soybean [56], and in the development of elite populations in maize [57] and barley [58].

Conclusion

The analysis of SNPs derived from GBS technology in *C. cardunculus* highlighted a strong structure separating the globe artichoke from the cardoon material, both wild and cultivated. The wild material from Spain was included in the cultivated cardoon group, supporting the idea that this might represent a feral form. Other wild genotypes might have derived from hybridization between wild and cultivated material. Structures were also observed within the globe artichoke collection, with genotypes belonging to the “Catanesi” or “Green Apulian” types forming well structured groups. Moreover, the late flowering large round-shaped “Romaneschi” type cannot be considered as a single, genetically uniform group. The globe artichokes displayed a slower LD decay compared to cardoons possibly due to the different reproductive system and to domestication and selective pressure.

Supporting information

S1 Table. List of the material analysed. All “varieties/ecotypes” tagged with “CC” represent cultivated cardoons, the ones tagged with “W” indicate wild cardoons, and the ones not tagged

in this way are the globe artichokes. Within globe artichokes, CAT: "Catanesi"; VIO: "Violetti"; SPI: "Spinosi"; ROM: "Romaneschi"; OFF: off types;?: uncertain attribution. In red, artichoke samples not used for diversity and STRUCTURE analyses.

(DOCX)

S2 Table. Genomic position and ID number of the *Cynara cardunculus* SNPs after filtering. Each worksheet contains the SNPs for a single dataset: 1. Total *C. cardunculus*; 2. Globe artichoke; 3. Cardoon.

(XLSX)

S1 Fig. Bar charts describing the proportion of SNP heterozygous distribution in *C. cardunculus* before (A) and after (B) HWE filtering (p -value $> 10^{-6}$)

(PDF)

S2 Fig. Pie chart showing the frequency of substitution types of the identified SNPs.

(PDF)

S3 Fig. Determination of the most probable K by means of ΔK statistics. The number of subpopulations (K) was identified based on maximum likelihood and ΔK values. (A) Whole dataset; (B) Globe artichoke dataset; (C) Cardoon dataset.

(PDF)

S4 Fig. STRUCTURE analysis at K = 6 and K = 7. (A): complete *C. cardunculus* dataset; (B): globe artichoke dataset. Numbers on the y-axis show the subgroup membership. Genotype names are reported on the x axis. The colours of the bars indicate the groups identified through the STRUCTURE program.

(TIF)

S5 Fig. STRUCTURE analysis of cardoon dataset. Numbers on the y-axis show the subgroup membership. Genotype names are reported on the x axis. The colours of the bars indicate the groups identified through the STRUCTURE program.

(PDF)

S6 Fig. Neighbor-Joining tree obtained from SNP data on the wild and cultivated cardoon collection. Number on tree branches indicate bootstrap values (≥ 50).

(TIF)

Acknowledgments

Authors thank Francesco Losavio, Anita Morgese and Giuseppe Sonnante for field (FL), laboratory (AM) and graphic (GS) technical assistance, respectively. This research was funded by the project BiodiverSO-PSR Puglia 2007–2013 Mis. 214/4 subaz. a) and by the project BiodiverSO-PSR Puglia 2014–2020 Mis.10.2.

Author Contributions

Conceptualization: Gabriella Sonnante.

Formal analysis: Stefano Pavan, Pasquale L. Curci, Diana L. Zuluaga, Gabriella Sonnante.

Funding acquisition: Gabriella Sonnante.

Investigation: Stefano Pavan, Pasquale L. Curci, Diana L. Zuluaga, Emanuela Blanco, Gabriella Sonnante.

Resources: Gabriella Sonnante.

Supervision: Gabriella Sonnante.

Validation: Stefano Pavan, Pasquale L. Curci, Diana L. Zuluaga, Gabriella Sonnante.

Writing – original draft: Stefano Pavan, Pasquale L. Curci, Diana L. Zuluaga, Gabriella Sonnante.

Writing – review & editing: Stefano Pavan, Gabriella Sonnante.

References

1. McCouch S, Baute GJ, Bradeen J, Bramel P, Bretting PK, Buckler E, et al. Agriculture: feeding the future. *Nature*. 2013; 499: 23–24. <https://doi.org/10.1038/499023a> PMID: 23823779
2. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*. 2011; 6: e19379. <https://doi.org/10.1371/journal.pone.0019379> PMID: 21573248
3. Fu YB, Peterson GW. Genetic diversity analysis with 454 pyrosequencing and genomic reduction confirmed the eastern and western division in the cultivated barley gene pool. *Plant Genome*. 2011; 4: 226–237.
4. Verma S, Gupta S, Bandhiwal N, Kumar T, Bharadwaj C, Bhatia S. High-density linkage map construction and mapping of seed trait QTLs in chickpea (*Cicer arietinum* L.) using Genotyping-by-Sequencing (GBS). *Sci Rep*. 2015; 5: 17512. <https://doi.org/10.1038/srep17512> PMID: 26631981
5. Saxena RK, Singh VK, Kale SM, Tathineni R, Parupalli S, Kumar V, et al. Construction of genotyping-by-sequencing based high-density genetic maps and QTL mapping for fusarium wilt resistance in pigeonpea. *Sci Rep*. 2017; 7: 1911. <https://doi.org/10.1038/s41598-017-01537-2> PMID: 28507291
6. Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, et al. Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome*. 2012; 5: 103–113.
7. Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, et al. Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genet*. 2013; 9: e1003215. <https://doi.org/10.1371/journal.pgen.1003215> PMID: 23349638
8. Pavan S, Lotti C, Marcotrigiano AR, Mazzeo R, Bardaro N, Bracuto V, et al. A distinct genetic cluster in cultivated chickpea as revealed by genome-wide marker discovery and genotyping. *The Plant Genome*, 2017; 10 <https://doi.org/10.3835/plantgenome2016.11.0115> PMID: 28724069
9. Pavan S, Marcotrigiano AR, Ciani E, Mazzeo R, Zonno V, Ruggieri V, et al. Genotyping-by-sequencing of a melon (*Cucumis melo* L.) germplasm collection from a secondary center of diversity highlights patterns of genetic variation and genomic features of different gene pools. *BMC Genomics*. 2017; 18: 59. <https://doi.org/10.1186/s12864-016-3429-0> PMID: 28068911
10. Foury C. Ressources génétiques et diversification de l'artichaut (*Cynara scolymus* L.). *Acta Hort*. 1989; 242: 155–166.
11. Sonnante G, Pignone D, Hammer K. The domestication of artichoke and cardoon: from Roman times to the genomic age. *Ann Bot*. 2007; 100: 1095–1100. <https://doi.org/10.1093/aob/mcm127> PMID: 17611191
12. Sonnante G, Carluccio AV, De Paolis A, Pignone D. Identification of artichoke SSR markers: molecular variation and patterns of diversity in genetically cohesive taxa and wild allies. *Genet Resour Crop Ev*. 2008; 55: 1029–1046.
13. Rottenberg A, Zohary D. The wild ancestry of the cultivated artichoke. *Genet Resour Crop Evol*. 1996; 43: 53–58.
14. Marie D, Brown SC. Cytometric exercise in plant DNA histograms, with 2C-values for 70 species. *Biol Cell*. 1993; 78: 41–51. PMID: 8220226
15. Gebhardt R. Antioxidative and protective properties of extracts from leaves of the artichoke (*Cynara scolymus* L.) against hydroperoxide-induced oxidative stress in cultured rat hepatocytes. *Toxicol Appl Pharmacol*. 1997; 144: 279–286. <https://doi.org/10.1006/taap.1997.8130> PMID: 9194411
16. Sonnante G, D'Amore R, Blanco E, Pierri CL, De Palma M, Luo J, et al. Novel hydroxycinnamoyl-Coenzyme A transferase genes from artichoke are involved in the synthesis of chlorogenic acid. *Plant Physiol*. 2010; 153: 1–15.

17. Negro D, Montesano V, Grieco S, Crupi P, Sarli G, De Lisi A, et al. Polyphenol compounds in artichoke plant tissues and varieties. *J Food Sci.* 2012; 77: C244–252. <https://doi.org/10.1111/j.1750-3841.2011.02531.x> PMID: 22251096
18. Christaki E, Bonos E, Florou-Paneria P. Nutritional and functional properties of *Cynara* crops (globe artichoke and cardoon) and their potential applications: A review. *International Journal of Applied Science and Technology.* 2012; 2: 64–70.
19. Hellwege EM, Czaplá S, Jahnke S, Willmitzer L, Heyer AG. Transgenic potato (*Solanum tuberosum*) tubers synthesize the full spectrum of inulin molecules naturally occurring in globe artichoke (*Cynara scolymus*) roots. *PNAS.* 2000; 97: 8699–8704. <https://doi.org/10.1073/pnas.150043797> PMID: 10890908
20. Costabile A, Kolida S, Klinder A, Gietl E, Bäuerlein M, Froberg C, et al. A double-blind, placebo-controlled, cross-over study to establish the bifidogenic effect of a very-long-chain inulin extracted from globe artichoke (*Cynara scolymus*) in healthy human subjects. *Brit J Nutr.* 2010; 104: 1007–1017. <https://doi.org/10.1017/S0007114510001571> PMID: 20591206
21. Gatto A, De Paola D, De Bagnoli F, Vendramin GG, Sonnante G. Population structure of *Cynara cardunculus* complex and the origin of the conspecific crops artichoke and cardoon. *Ann Bot.* 2013; 112: 855–865. <https://doi.org/10.1093/aob/mct150> PMID: 23877076
22. Calabrese N, Carito A, Boari F, Cantore V, De Palma E, Damato G. Agronomical evaluation of artichoke cultivar propagated by seed. *Acta Hort.* 2011; 942: 153–158.
23. Porceddu E, Dellacecca V, Bianco VV. Classificazione numerica di cultivar di carciofo. In: *Proceedings of the II International Congress on Artichoke, Bari 1973.* Torino: Minerva Medica (in Italian). 1976; 1105–1119.
24. Sonnante G, De Paolis A, Pignone D. Relationships among artichoke cultivars and some related wild taxa based on AFLP markers. *Plant Genet Resour Charact Util.* 2004; 1: 125–133.
25. Pagnotta MA, Noorani A. Genetic Diversity Assessment in European *Cynara* Collections. In: *Tuberosa R, Graner A, Frison E, editors. Genomics of Plant Genetic Resources.* Springer Science+Business Media; 2014. pp. 559–584.
26. Wiklund A. The genus *Cynara* L. (Asteraceae-Cardueae). *Bot J Linn Soc.* 1992; 109: 75–123.
27. Pignone D, Sonnante G. Wild artichokes of south Italy: did the story begin here? *Gen Res Crop Evol.* 2004; 51: 577–80.
28. Scaglione D, Reyes-Chin-Wo S, Acquadro A, Froenicke L, Portis E, Beitel C, et al. The genome sequence of the outbreeding globe artichoke constructed de novo incorporating a phase-aware low-pass sequencing strategy of F1 progeny. *Sci Rep.* 2016; 6: 19427. <https://doi.org/10.1038/srep19427> PMID: 26786968
29. Acquadro A, Barchi L, Portis E, Mangino G, Valentino D, Mauromicale G, et al. Genome reconstruction in *Cynara cardunculus* taxa gains access to chromosome-scale DNA variation. *Sci Rep.* 2017; 7: 5617 <https://doi.org/10.1038/s41598-017-05085-7> PMID: 28717205
30. Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. *Nature Rev. Genet.* 2011; 12: 443–451. <https://doi.org/10.1038/nrg2986> PMID: 21587300
31. Curci PL, De Paola D, Sonnante G. Development of chloroplast genomic resources for *Cynara*. *Mol Ecol Resour.* 2015b; 16: 562–573. <https://doi.org/10.1111/1755-0998.12457> PMID: 26354522
32. Curci PL, De Paola D, Danzi D, Vendramin GG, Sonnante G. Complete chloroplast genome of the multi-functional crop globe artichoke and comparison with other Asteraceae. *Plos One.* 2015a; 10: e0120589. <https://doi.org/10.1371/journal.pone.0120589> PMID: 25774672
33. Glaubitz JC, Casstevens TM, Fei L, Harriman J, Elshire RJ, Sun Q. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS One.* 2014; 9: e90346. <https://doi.org/10.1371/journal.pone.0090346> PMID: 24587335
34. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009; 25: 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
35. Nimmakayala P, Levi A, Abburi L, Abburi VL, Tomason YR, Saminathan T, et al. Single nucleotide polymorphisms generated by genotyping by sequencing to characterize genome-wide diversity, linkage disequilibrium, and selective sweeps in cultivated watermelon. *BMC Genomics.* 2014; 15:767. <https://doi.org/10.1186/1471-2164-15-767> PMID: 25196513
36. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics.* 2007; 23: 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308> PMID: 17586829
37. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics.* 2011 Aug 1; 27(15): 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330> PMID: 21653522

38. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155: 945–59. PMID: [10835412](https://pubmed.ncbi.nlm.nih.gov/10835412/)
39. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. 2005; 14: 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x> PMID: [15969739](https://pubmed.ncbi.nlm.nih.gov/15969739/)
40. Earl DA, Vonholdt BM. Structure Harvester: a website and program for visualizing Structure output and implementing the Evanno method. *Conserv Genet Resour*. 2012; 4: 359–361.
41. Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987; 4: 406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454> PMID: [3447015](https://pubmed.ncbi.nlm.nih.gov/3447015/)
42. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol*. 1993; 10: 512–526. <https://doi.org/10.1093/oxfordjournals.molbev.a040023> PMID: [8336541](https://pubmed.ncbi.nlm.nih.gov/8336541/)
43. Waples R, Gaggiotti O. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molec Ecol*. 2006; 15: 1419–1439.
44. Scaglione D, Acquadro A, Portis E, Tirone M, Knapp SJ, Lanteri S. RAD tag sequencing as a source of SNP markers in *Cynara cardunculus* L. *BMC Genomics*. 2012; 13: 3. <https://doi.org/10.1186/1471-2164-13-3> PMID: [22214349](https://pubmed.ncbi.nlm.nih.gov/22214349/)
45. Hamblin MT, Rabbi IY. The effects of restriction-enzyme choice on properties of genotyping-by-sequencing libraries: a study in cassava (*Manihot esculenta*). *Crop Sci*. 2014; 54: 2603–2608.
46. Dempewolf H, Rieseberg LH, Cronk QC. Crop domestication in the Compositae: a family-wide trait assessment. *Genet Resour and Crop Evol*. 2008; 55: 1141–1157.
47. Pagnotta MA, Fernández JA, Sonnante G, Egea-Gilabert C. Genetic diversity and accession structure in European *Cynara cardunculus* collections. *PLoS One*. 2017; 12: e0178770. <https://doi.org/10.1371/journal.pone.0178770> PMID: [28570688](https://pubmed.ncbi.nlm.nih.gov/28570688/)
48. Hammer K, Knupfer H, Laghetti G, Perrino P. Seeds from the past. A catalogue of crop germplasm in Southern Italy and Sicily. CNR, Istituto del Germoplasma, Bari, Italy. 1992. p. 173.
49. Laghetti G, Pignone D, De Lisi A, Cifarelli S, Faslia N, Hammer K. Collecting crop genetic resources in Italian towns of Albanian origin across the Molise, Calabria and Sicily regions. *Genet Resour and Crop Evol*. 2011; 58: 139–152.
50. McKey D, Elias M, Pujol B, Duputié A. The evolutionary ecology of clonally propagated domesticated plants. *New Phytol*. 2010; 186: 318–332. <https://doi.org/10.1111/j.1469-8137.2010.03210.x> PMID: [20202131](https://pubmed.ncbi.nlm.nih.gov/20202131/)
51. Rival L, McKey D. Domestication and diversity in Manioc (*Manihot esculenta* Crantz ssp. *esculenta*, Euphorbiaceae). *Curr Anthropol*. 2008; 49: 1119–1128.
52. Rafalski A, Morgante M. Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends Genet*. 2004; 20: 103–111. <https://doi.org/10.1016/j.tig.2003.12.002> PMID: [14746992](https://pubmed.ncbi.nlm.nih.gov/14746992/)
53. Vos PG, Paulo MJ, Voorrips RE, Visser RGF, van Eck HJ, van Eeuwijk FA. Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. *Theor Appl Genet*. 2017; 130: 123–35. <https://doi.org/10.1007/s00122-016-2798-8> PMID: [27699464](https://pubmed.ncbi.nlm.nih.gov/27699464/)
54. Raboin LM, Pauquet J, Butterfield M, D'Hont A, Glaszmann JC. Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane. *Theor Appl Genet*. 2008; 116: 701–714. <https://doi.org/10.1007/s00122-007-0703-1> PMID: [18196216](https://pubmed.ncbi.nlm.nih.gov/18196216/)
55. D'hoop BB, Paulo MJ, Kowitwanich K, Sengers M, Visser RG, van Eck HJ, et al. Population structure and linkage disequilibrium unravelled in tetraploid potato. *Theor Appl Genet*. 2010; 121: 1151–1170. <https://doi.org/10.1007/s00122-010-1379-5> PMID: [20563789](https://pubmed.ncbi.nlm.nih.gov/20563789/)
56. Hyten DL, Choi IY, Song Q, Shoemaker RC, Nelson RL, Costa JM, et al. Highly variable patterns of linkage disequilibrium in multiple soybean populations. *Genetics*. 2007; 175: 1937–1944. <https://doi.org/10.1534/genetics.106.069740> PMID: [17287533](https://pubmed.ncbi.nlm.nih.gov/17287533/)
57. Tenailon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc Natl Acad Sci U S A*. 2001; 98: 9161–6. <https://doi.org/10.1073/pnas.151244298> PMID: [11470895](https://pubmed.ncbi.nlm.nih.gov/11470895/)
58. Caldwell KS, Russell J, Langridge P, Powell W. Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics*. 2006; 172: 557–567. <https://doi.org/10.1534/genetics.104.038489> PMID: [16219791](https://pubmed.ncbi.nlm.nih.gov/16219791/)