ANNALS OF
BOTANY
Founded 1887

Population structure of *Cynara cardunculus* complex and the origin of the conspecific crops artichoke and cardoon

Angela Gatto[1], Domenico De Paola[1], Francesca Bagnoli[2], Giovanni Giuseppe Vendramin[3] and Gabriella Sonnante[1],*

[1]*Institute of Plant Genetics (IGV), National Research Council (CNR), Via Amendola 165/A, 70126 Bari, Italy,*
[2]*Institute for Plant Protection and* [3]*Institute of Plant Genetics, National Research Council, Via Madonna del Piano 10, 50019 Sesto Fiorentino, Firenze, Italy*
* *For correspondence. E-mail gabriella.sonnante@igv.cnr.it*

• *Background and Aims* Globe artichoke and leafy cardoon, two crops within the same species *Cynara cardunculus*, are traditionally cultivated in the Mediterranean region and play a significant role in the agricultural economy of this area. The two cultigens have different reproductive systems: artichoke is generally vegetatively propagated, while leafy cardoon is seed propagated. The domestication events underlying the origin of both artichoke and cultivated cardoon from their wild relative and the area of occurrence are not yet fully understood. The aim of this study was to investigate population structure in wild cardoon, globe artichoke and leafy cardoon material and infer domestication events.
• *Methods* Thirty-five microsatellite (simple sequence repeat) markers, distributed in the *C. cardunculus* genome, and a large geographical and numerical sampling in southern Europe and North Africa were used to assess population structure and diversity.
• *Key Results* The results suggest the presence of two distinct domestication events for artichoke and leafy cardoon, and also suggest a new possible scenario, with western wild cardoon having originated from cultivated cardoon escaped from cultivation. Evidence was found for a demographic bottleneck in the past history of globe artichoke.
• *Conclusions* The results shed new light on the relationships between the three taxa of *C. cardunculus* and highlight relevant aspects on the evolution of domestication of two crops with a different reproductive system within the same species. It is proposed that the probable centre of origin of artichoke is located in southern Italy, probably Sicily.

Key words: *Cynara cardunculus*, SSR markers, population structure, multiple domestication events, clonal propagation, bottleneck, reproductive system.

INTRODUCTION

Cynara cardunculus L. of the family Asteraceae is a diploid ($2n = 2x = 34$), cross-pollinated species complex that originated in the Mediterranean Basin area (Sonnante *et al.*, 2007*a*, *b*). This species contains three different taxa: the wild perennial cardoon [var. *sylvestris* (Lam.) Fiori], which has been recognized as the ancestor of both the globe artichoke [var. *scolymus* (L.) Fiori] and the leafy or cultivated cardoon (var. *altilis* DC) (Rottenberg and Zohary, 1996; Sonnante *et al.*, 2003, 2007*b*).

Globe artichoke is traditionally widely cultivated in southern Europe and North Africa, where it represents an important component of the agricultural economy; it is also widespread in California, South America (mainly Peru and Argentina) and in China (FAO, http://faostat.fao.org, updated to 25 February 2013). For centuries, artichoke has been vegetatively propagated, but in recent years seed-propagated varieties have been released on the market (Gil and Villa, 2004; Calabrese *et al.*, 2011). This crop is mainly grown for the production of its large immature inflorescences, or capitula, consumed as vegetables, but also for the extraction of bioactive compounds from its leaves. *C. cardunculus* plants, in fact, are rich in antioxidant compounds, polyphenols such as flavonoids, caffeic acid, chlorogenic acid and its derivatives dicaffeoylquinic acids,

including cynarin (Gebhardt, 1997; Fratianni *et al.*, 2007; Sonnante *et al.*, 2010; Negro *et al.*, 2012). Moreover, roots contain inulin, an oligosaccharide showing a positive effect on the human intestinal microbial community (Kraft, 1997; Raccuia and Melilli, 2004). Cultivated cardoon is grown on a smaller scale in northern Italy, southern France and in Spain and has been selected for the gigantism of leaf stalks, which are consumed as vegetables. The conspecific wild cardoon is considered the wild progenitor of both crops (Foury, 1989; Bazniski and Zohary, 1994; Sonnante *et al.*, 2003; Pignone and Sonnante, 2004; Sonnante *et al.*, 2008), and is distributed across the Mediterranean Region, from Cyprus and the Black Sea to Atlantic Spain, Portugal and the Canary Islands (Wiklund, 1992). Its small flower heads are traditionally gathered from the wild and used as a food in certain areas of southern Italy (Pignone and Sonnante, 2004). However, distinct morphological types of the wild cardoon have been identified. According to head morphology, Foury (1989) recognized three forms – the Sicilian, the Tunisian and the Catalan types – whereas Wiklund (1992) distinguished the eastern type with smaller plants, leaves and heads, and longer spines on head bracts (*C. cardunculus* L. subspecies *cardunculus*), from the western wild cardoon, *C. cardunculus* L. subspecies *flavescens* Wikl., distributed in Spain and Portugal, characterized by more robust plants, and

fewer and smaller spines. Molecular data provided evidence that the western wild cardoon is genetically closely related to the cultivated cardoon, while the eastern wild cardoon might be the progenitor of the globe artichoke (Sonnante *et al.*, 2007*a*, *b*, 2008). The two crops have possibly been derived from human selection pressure for either large, non-spiny heads on one side, or non-spiny, large stalked tender leaves on the other side (Bazniski and Zohary, 1994; Sonnante *et al.*, 2007*b*).

Investigating the ancestry of *Cynara* crops has been a matter of debate in recent decades. On the basis of morphological characters on a large set of specimens, and using a cladistic method, Wiklund (1992) suggested the inclusion of cultivated artichoke, leafy cardoon and wild cardoon in a single species: *C. cardunculus* L. Her results also indicated that *C. auranitica*, *C. baetica* and *C. syriaca* were close relatives of *C. cardunculus*. However, based on hybridization experiments, it was demonstrated that wild *C. cardunculus* and the two crops are completely inter-fertile and therefore they belong to the same primary gene-pool (Bazniski and Zohary, 1994; Rottenberg and Zohary, 1996, 2005). The other wild species of the genus *Cynara* show only limited or no capacity to set seeds and produce viable hybrids when crossed with the cultigen (Rottenberg and Zohary, 1996). Recently, on the basis of molecular data and using the calibration of fossil remains, it has been suggested that the genus *Cynara* might be older than believed (around 23 Myr old), while the speciation of some species could date back to about 12 Mya (Barres *et al.*, 2013).

The aims of this study were to: (1) assess genetic variation in wild cardoon populations from Mediterranean countries, including five southern European countries and Tunisia and in the two cultigens; and (2) acquire deeper information on the domestication of globe artichoke and leafy cardoon. For this, 801 individuals were genotyped at 35 microsatellite (simple sequence repeat, SSR) loci distributed on all the linkage groups of an artichoke × wild cardoon genetic map (Sonnante *et al.*, 2011).

MATERIALS AND METHODS

Plant material and genotyping

In total, 801 individuals belonging to the three taxa of *Cynara cardunculus* were used (Supplementary data Table S1). Wild cardoon material was obtained from direct collection of the CNR, Institute of Plant Genetics, teams or from exchange with other institutions (i.e. Botanic Institute, CSIC, Barcelona, Spain; IPK Gatersleben, Germany; University of Tunis, Tunisia). Cultivated cardoon varieties were obtained from seed companies, Spanish varieties from Instituto Técnico y de Gestión Agrícola (Cadreita, Spain), for a total of 54 individuals belonging to 10 varieties (Supplementary data Table S1). As for artichoke, 16 varietal types or landraces were used, mostly belonging to the living collection held at the Institute of Plant Genetics, with 87 individuals in total (Supplementary data Table S1). The artichoke material represented the four main morphological groups (Sonnante *et al.*, 2003), and within each group, varietal types were collected from various fields. Two landraces ('Bianco di Ostuni', 'Nero di Castrignano') were off-types and were chosen on the basis of head morphology and colour.

As for the seed-propagated material, i.e. wild and cultivated cardoons, seeds were germinated in Petri dishes on moist filter paper. Subsequently, seedlings were planted in pots and grown in a greenhouse. Leaf material was collected from wild and cultivated cardoon plants, and from globe artichoke plants in the living collection, and immediately frozen at −80 °C. DNA was isolated from each individual plant using a CTAB extraction protocol (Sonnante *et al.*, 2003).

All 801 DNA samples were genotyped using 35 SSRs, mostly expressed sequence tag SSRs, already mapped on a genetic map artichoke × wild cardoon, choosing two SSR markers for each linkage group (LG), and one for LG Mola-18 which was not integrated in the map of both parents (Sonnante *et al.*, 2011). For SSR amplification, a three-primer protocol including an M13 primer was used (Schuelke, 2000) in a total volume of 10 µL, containing 2·5 ng template DNA, 0·07 µM forward primer including an 18-bp M13-tail, 0·2 µM reverse primer, 0·2 µM of 18-bp M13-labelled primer (Sigma-Aldrich, Milan, Italy), 0·2 mM of each dNTP, 1 µL 10× buffer, 0·4 U *Taq* DNA polymerase (Invitrogen, Carlsbad, CA, USA) and 1·5 mM MgCl₂. Amplification was achieved using 3 min initial denaturation at 94 °C, 38 cycles of 30 s at 94 °C, 30 s annealing at optimal primer temperature (Sonnante *et al.*, 2011) and 45 s synthesis at 72 °C, followed by a final 10 min incubation at 72 °C. PCR products were analysed on a CEQ 8800 automated sequencer (Beckman Coulter, Carlsbad, CA, USA), and peaks were identified by comparison with an internal size standard, using the proprietary fragment analysis software of the sequencer.

Population genetic analyses

All microsatellite markers were first evaluated using the Ewens-Watterson test for neutrality (Yeh and Boyle, 1997). Genetic diversity parameters, including polymorphic information content (PIC), total number of alleles (T_A), mean number of alleles (N_a), and observed (H_o) and expected (H_e) heterozygosity, were calculated using GenAlEx v.6·1 (Peakall and Smouse, 2006). Allelic richness (A_r) was computed by means of FSTAT (Goudet, 1995). We checked raw data for the potential effect of null alleles on genetic differentiation, and calculated corrected fixation index ^{c}Fst values using the excluding null allele (ENA) method in FREENA (Chapuis and Estoup, 2007). Furthermore, corrected inbreeding coefficients ^{c}Fis were recalculated taking into account the possible presence of null alleles by using the program INEST (Chybicki and Burczyk, 2009), running the individual inbreeding model (IIM) with a Gibbs sampler of 105 iterations. Significance of null allele frequencies different from 0 was calculated for $P < 0.01$.

Population structure

Population structure was investigated using the Bayesian, model-based clustering approach as implemented in the software STRUCTURE ver. 2.3.4 (Pritchard *et al.*, 2000). We performed four main different analyses: one on the whole dataset, one for the wild material only, the third one for cultivated cardoon material and the fourth one for artichoke varietal types. The wild material was also analysed without the western (Spain and Portugal) wild samples, and the cultivated cardoon accessions were also examined together with the wild cardoon from Spain and Portugal. For each analysis, STRUCTURE was run with different values of the number of clusters (K), varying from 2 to 8 under the admixture model, with no prior population information, and, in the case of

wild material, also by using the LOCPRIOR option. To verify the robustness of the results, we performed 20 independent runs per *K* value with 50 000 burn-in period and 100 000 Markov chain Monte Carlo iterations. The maximum-likelihood run was used to assign the most probable number of groups, using the statistic described by Evanno *et al.* (2005). CLUMPP v.1·1.2 (Jakobsson and Rosenberg, 2007) was used to summarize parameters across the 20 runs and the corresponding graphical output was visualized using STRUCTURE.

A clustering algorithm for spatial population genetic studies was applied using the software TESS version 2·3 (François *et al.*, 2006; Chen *et al.*, 2007). TESS can perform both individual geographical assignment and admixture analysis and it is designed for seeking genetic discontinuities in continuous populations and estimating spatially varying individual admixture proportions. In TESS analysis both admixture and without admixture models were used. Five independent runs were settled for each *K* value ranging from 2 to 8, with a total number of sweeps of 5×10^4 and a burn-in number of sweeps of 10^4. TESS analysis was performed on the wild material only.

Bottleneck

The software BOTTLENECK version 1.2.02 (Cornuet and Luikart, 1996) was used to evaluate the hypothesis of a recent severe reduction on effective population size for artichoke, wild cardoon and cultivated cardoon. BOTTLENECK is based on the evidence that recently bottlenecked populations exhibit an excess of gene diversity among polymorphic loci. The observed heterozygosity is compared with the expected heterozygosity at mutation-drift equilibrium, given the number of alleles and the population sample size. The two-phased model (TPM) of mutation was applied setting 5 % of multistep changes and a variance among multiple step of 12 as recommended for microsatellite loci (Piry *et al.*, 1999). Two thousand iterations were used and the Wilcoxon singed rank test implemented in the software was employed to test the significance of bottlenecks. Furthermore, the qualitative descriptor of allele frequency distribution 'mode shift' was used to test the allele distribution and discriminate bottlenecked populations (Luikart and Cornuet, 1998).

Genetic relationships

Genetic relationships among wild and cultivated samples were assessed by means of Nei's D_{ST} genetic distance (Nei, 1972), and the obtained distance matrix was employed to construct a UPGMA tree using POPTREE2 (Takezaki *et al.*, 2010), with a bootstrap of 100 replicates. The consensus tree was visualized by means of the program Phylogeny.fr (Dereeper *et al.*, 2008). A principal co-ordinate analysis (PCA) was carried out on the basis of Nei's genetic distance (Nei, 1972), by means of the program GenAlex v.6·1 (Peakall and Smouse, 2006).

RESULTS

Population genetic diversity

The 35 SSR loci were chosen according to their position on a map obtained by crossing an artichoke and wild cardoon (Sonnante

et al., 2011). The amplification and analysis of genetic markers produced a total of 639 alleles, with a mean of 18·3 alleles per locus, from a minimum of eight (CyEM072) to a maximum of 38 (CELMS07) alleles, and a PIC varying from 0·42 (CyEM002 and CyEM210) to 0·91 (CyEM138, CELMS14 and CELMS58). The level of H_o per locus ranged between 0·270 (CyEM002) and 0·780 (CyEM138), with a mean of 0·462, while the level of H_e was between 0·264 for CyEM210 and 0·670 for CyEM138, with a mean of 0·473 (Supplementary data Table S2). Microsatellite markers were evaluated using the Ewens–Watterson test for neutrality (Yeh and Boyle, 1997) and none of them deviated significantly from a neutral equilibrium model, as indicated by the ObsF value ranking within the confidence interval of 95 %, while some null alleles showed a frequency above 0 (Supplementary data Table S3). Genetic diversity was assessed for the whole germplasm data set and for four main groups: (1) eastern wild cardoon from Italy, Greece, Tunisia and Malta; (2) western wild cardoon from Spain and Portugal; (3) cultivated cardoon; and (4) artichoke (Table 1). A more detailed analysis was carried out for eastern wild cardoon samples, which constituted the most numerous group, by subdividing this set into subgroups (Supplementary data Table S4). The highest value of T_A (607), A_r (16·99) and N_a (4·31) was observed for the eastern wild cardoon samples, and, within this set, for the Italian (530 T_A, 14·72 A_r) and Tunisian (5·63 N_a) germplasm, respectively, whereas the artichoke set had the lowest T_A (157), A_r (4·45) and N_a (2·23) values. Artichoke samples also showed the highest values of H_o (0·721), much higher than H_e (0·424), as expected for a very heterozygous crop; ^{c}Fst and ^{c}Fis values were 0·276 and 0·036, respectively. The other cultigen, cultivated cardoon, showed a T_A equal to 244; N_a (2·67) and ^{c}Fst (0·313) were slightly lower, while ^{c}Fis (0·034) was slightly lower than those of artichoke. Conversely, H_o in cultivated cardoon was much lower (0·386) than in artichoke. In general, eastern wild material showed a higher value of the inbreeding coefficient ^{c}Fis and a lower fixation index ^{c}Fst compared with the western wild material, indicating that eastern wild populations retain a slightly higher level of inbreeding and, at the same time, are less differentiated among themselves, compared with the Iberian wild germplasm. When considering smaller groups of wild populations (see Supplementary data Table S4), the differentiation between populations in limited geographical areas is often reduced.

TABLE 1. *Genetic diversity parameters for the five main groups of the germplasm analysed*

	WHOLE	WCAR-E	WCAR-W	CC	ART
T_A	639	607	269	244	157
P_A	113	93	10	7	3
N_a^*	3·68	4·31	2·94	2·67	2·23
A_r^*	–	16·99	7·53	6·79	4·45
H_o^*	0·462	0·452	0·320	0·386	0·721
H_e^*	0·473	0·519	0·370	0·383	0·424
^{c}Fis*	0·042	0·048	0·024	0·034	0·036
^{c}Fst*	0·360	0·283	0·445	0·313	0·276

T_A, total number of alleles; P_A, private alleles; N_a, number of alleles; A_r, allelic richness; H_o/H_e, observed/expected heterozygosity; ^{c}Fis, corrected inbreeding coefficient; ^{c}Fst, corrected fixation index; WCAR, wild cardoon; E, east; W, west; CC, cultivated cardoon; ART, artichoke.

* Averaged across all loci.

Cynara cardunculus population structure

When performing STRUCTURE on the whole dataset composed of the two cultigens and the wild cardoon samples, a maximum value of the rate of change in the log probability of data using the Evanno method (Evanno *et al.*, 2005) was observed at $K = 3$, although the value of ΔK (10·88) was quite low, meaning that this structuring of populations was not strong. Therefore, we decided to consider also other groupings for other K (Fig. 1). According to these results, at $K = 3$, Mediterranean wild populations, except for samples from Spain and Portugal, formed two groups, while cultivated cardoon and artichoke were positioned in the third group, also including wild material from Spain and Portugal (Fig. 1). In most cases, group 1 (red) included Italian populations from the Apulia and Basilicata regions, while populations from Calabria were admixed between the first and the second group; group 2 (green) was composed of populations from other Italian regions (Sicily, Sardinia, Latium) together with the material from Greece and Malta, with some admixture with group 1, and from Tunisian populations. Wild cardoons from Spain and from Portugal were included in group 3 (blue), with some admixture especially in one population from Spain, together with cultivated cardoon samples and artichoke. Note that two populations from Sicily showed a strong admixture with cultivated material, and some admixture with the cultivated material was also observed for a population from Malta and populations from Tunisia. In 15 % of cases, the artichoke group (green) was separated from the group of cultivated cardoon samples; a high genetic component of this group was also found in two populations from Sicily, one from Malta and all of the Tunisian samples.

The second higher value of ΔK was observed at $K = 4$, which provided a better picture of the groupings. In most cases (80 %) the artichoke samples were separated from the rest of the sample set, and therefore the main difference between this STRUCTURE output and the $K = 3$ output is the sharp separation between the group of Spanish and Portuguese wild cardoon together with cultivated cardoon (blue) on one side, and the artichoke cluster on the other (yellow, Fig. 1). Also in this case, the same wild populations indicated above, particularly from Sicily, but also from Malta and Tunisia, showed some admixture with the cultivated material, this time in particular with the artichoke germplasm. Also, a wild population from Spain shared some background with artichoke. The Tunisian populations were in most cases (65 %) grouped with other wild populations, while in 35 % of the cases formed a separate group. At higher values of K (5–8), the second part of the plots was almost always subdivided into three groups: (1) Tunisia, (2) wild Spain and Portugal together with cultivated cardoon and (3) artichoke; the other subdivisions concerned the first part of the plots (Fig. 1, only for $K = 5$).

To get a better idea of the groupings within wild material, STRUCTURE was also run on wild cardoon samples only. When populations from Spain and Portugal were included, a stronger structure at $K = 2$ ($\Delta K = 191$) was observed, where some Italian (Apulia, Basilicata and Calabria) and Greek populations formed group 1 (red), and Tunisian, Spanish and Portuguese populations belonged to group 2 (green), while other populations from Italy (Sicily, Sardinia, Latium) and Malta were mostly admixed (Fig. 2A). The same situation was also remarked when wild populations were identified by means of their geographical origin, using the LOCPRIOR option in STRUCTURE (data not shown). As wild populations from Spain and Portugal are genetically close to cultivated cardoon samples, a further structure analysis of wild cardoon was accomplished by excluding these wild samples. The results obtained revealed a strong structure ($\Delta K = 619$) at $K = 3$ (Fig. 2B). In this case, Italian populations either belonged to group 1 (blue, Apulia and Basilicata, partly Calabria), or to group 2 (green, partly Calabria and Sicily), or were admixed between groups 1 and 2, or 1, 2 and 3 (red). Greek and two Malta populations belonged to group 2, while one population from Malta was admixed between groups 2 and 3. The Tunisian populations formed a compact group (3).

To obtain additional information, STRUCTURE analysis was also carried out on other subsets of samples. In one case, wild material from Tunisia, Spain and Portugal was run together with cultivated cardoon, and artichoke. A reliable structure ($\Delta K = 144$) was obtained for $K = 3$, in which the Tunisian wild material formed group 1 (blue, Supplementary data Fig. S1), the second cluster was composed of wild cardoon from Spain and Portugal together with cultivated cardoon (green), and the third group was represented by the artichoke samples (red). A wild Spanish population shared some genetic background with the artichoke material.

If we consider only the cultivated cardoon varieties, a strong structure was observed for $K = 2$ ($\Delta K = 351$). Group 1 (green, Supplementary data Fig. S2) was composed of leafy cardoon varieties from Italy, France and Tunisia, while group 2 (red) included Spanish varieties. We also tried to uncover a possible structure between cultivated cardoon and wild cardoon populations from Spain and Portugal; in this case, the most probable structure at $K = 5$ was somewhat weak ($\Delta K = 12·59$) and mainly distinguished the two groups of cultivated cardoon previously described and three further groups of wild cardoon (data not shown). The wild cardoon population WSPA1 seemed to share a high genetic background with the Spanish cultivated cardoon material.

STRUCTURE analysis on artichoke samples revealed two major groups ($\Delta K = 73$): the first cluster comprised samples of the 'Catanese' type, which looks very homogeneous, while the second group included all the other artichoke varieties/landraces. 'Blanca de Tudela' possessed an admixed structure (Supplementary data Fig. S3). At a lower value of ΔK (8·44), four groups could be recognized: group 1 (yellow) including all the 'Catanese' type, as for $K = 2$; group 2 (green) containing 'Romanesco' and 'Paestum' varieties. 'Blanca de Tudela' was admixed between these two groups. The third group (red) included 'Camus de Bretagne', 'Bianco di Pertosa' and 'Spinosi' types, while the fourth group (blue) contained the two off-types ('Nero di Castrignano' and 'Bianco di Ostuni'). The 'Violetti' types were admixed between groups 3 and 4.

TESS analysis, performed on the wild populations only, confirmed the results obtained using STRUCTURE, thus strengthening the inferences made on the structure of diversity in this species across its distribution range (data not shown).

Bottleneck

Under the TPM, for both wild and cultivated cardoon the probability value obtained for each population revealed no evidence for bottleneck and a normal L-shaped form for the allele frequency distribution was observed. Conversely, the Wilcoxon
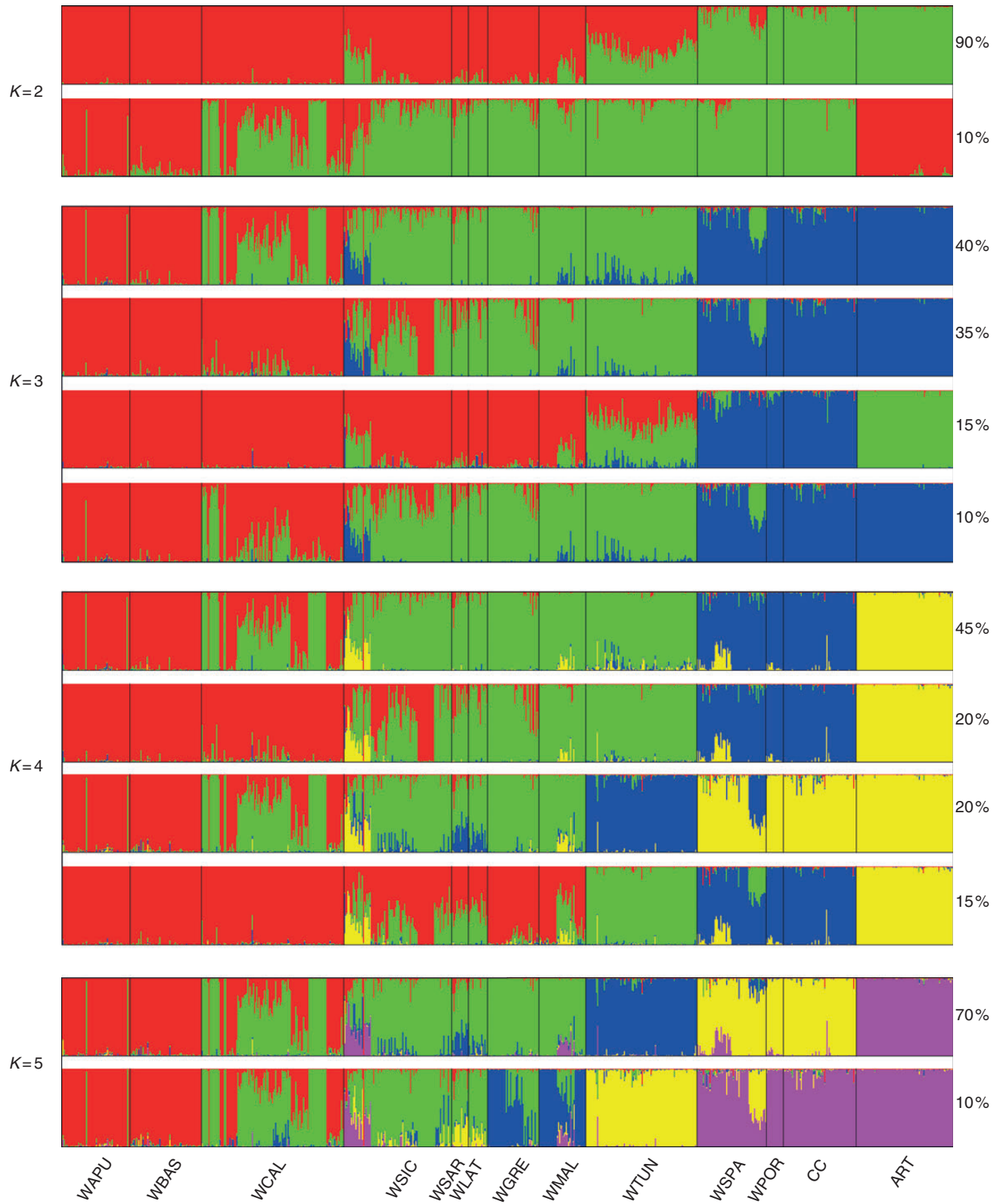
FIG. 1. Clusters inferred with the STRUCTURE program for *Cynara cardunculus* genotypes at *K* values from 2 to 5. For each *K* the percentage of the obtained clusters is shown. Clusters with a percentage <10 % are not reported. Colours indicate individual estimated membership fraction.
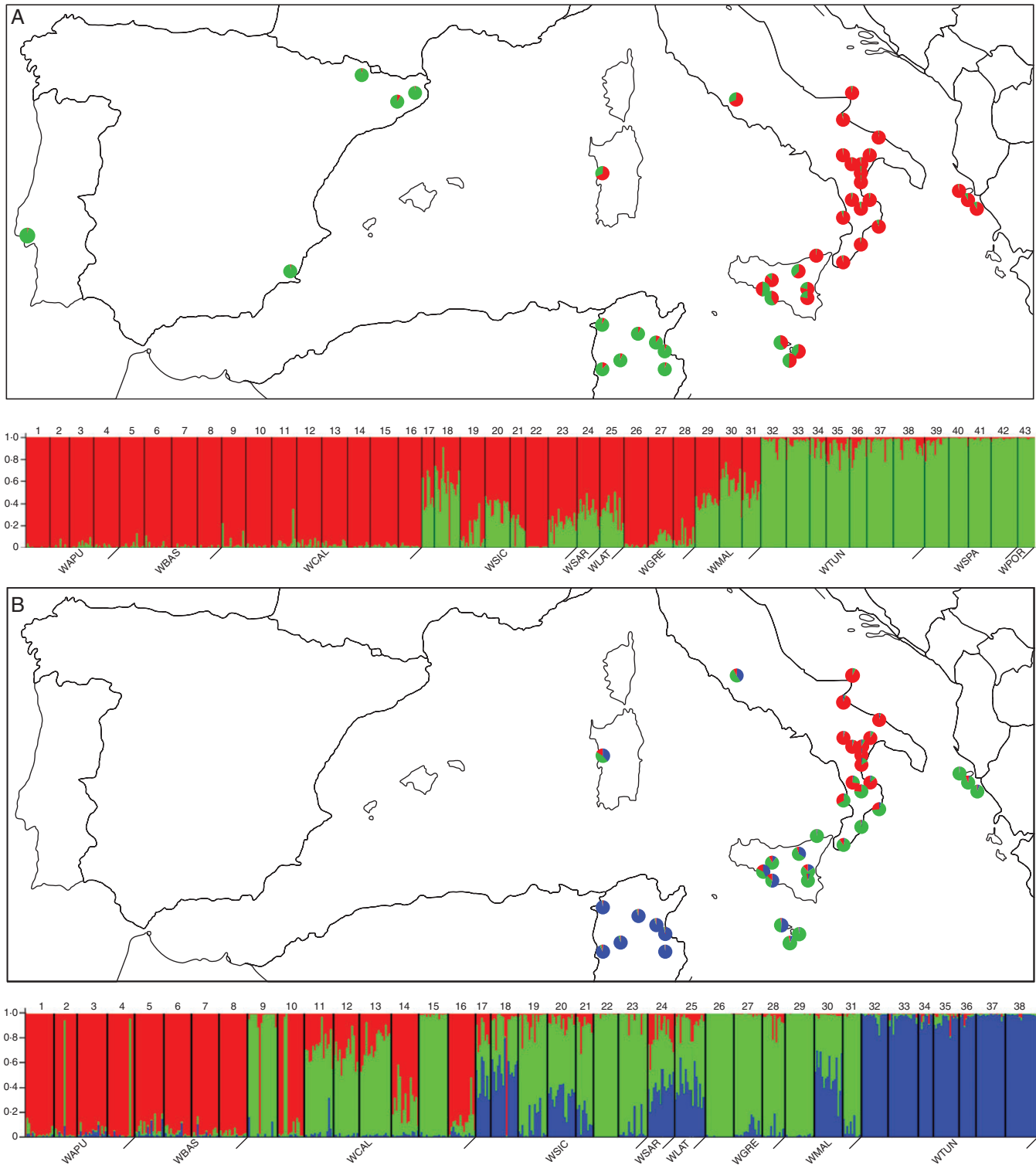
FIG. 2. Map of geographical distribution showing the average proportion of the Q value inferred by STRUCTURE and its relative plot for (A) wild cardoon at $K = 2$, and (B) wild cardoon without WSPA and WPOR populations at $K = 3$. For $K = 2$ group 1 (red) includes populations from Italy and Greece; Tunisian, Spanish and Portuguese populations are included in group 2 (green). For $K = 3$, group 1 (blue), group 2 (green) and group 3 (red) can be distinguished. Populations are separated by black vertical lines and indicated with a progressive number (see Supplementary Data Table S1 for population details). Colours indicate individual estimated membership fraction.

test for artichoke highlighted that this population has experienced a bottleneck in the past. Nonetheless, the mode-shift indicator test revealed also for artichoke a normal L-shaped form, as expected for non-bottlenecked populations that are near to mutation-drift equilibrium. However, when analysing allele frequency distributions to test for bottlenecks, it should be assumed that the population is random mating. As a crop, and especially being clonally propagated, artichoke does not satisfy this requirement. Using a χ^2 test we observed a highly significant deviation from Hardy−Weinberg equilibrium ($P < 0.001$) in artichoke material, supporting the hypothesis that a bottleneck occurred in the past history of this population.

Genetic relationships

The UPGMA tree obtained from Nei's genetic distances is shown in Fig. 3. Three main groups can be distinguished: one including the eastern wild cardoon samples, the second group formed by cultivated cardoon and wild cardoon from Spain and Portugal, and finally the artichoke cluster. When observing the eastern wild cardoon group, note that most Italian samples appear to be closely related, even though some populations, especially from Sicily, seem to be more distantly related. The Greek populations are in between Italian populations (from Calabria and Sicily) and populations from Malta. The group of Tunisian populations is well compacted and separated from the

FIG. 3. UPGMA tree obtained from Nei's genetic distances. Three main clusters can be observed: eastern wild cardoon (blue), cultivated cardoon and western wild cardoon (light red and green, respectively), and artichoke (red).

remaining samples. Within the second group in Fig. 3, two sub-groups can be identified: one including the cultivated cardoon from Italy and France, the other with the leafy cardoon from Spain and two wild cardoon populations from Spain. The other wild cardoons from Spain and Portugal, although belonging to this major second group, are somewhat more distantly related. The third main group within Fig. 3 is made up of artichoke varietal types. Within this cluster, the 'Catanesi' types are grouped with 'Blanca de Tudela', while the other typologies, 'Romaneschi', 'Violetti' and 'Spinosi', form another group. The two landraces 'Nero di Castrignano' and 'Bianco di Ostuni' do not cluster with any of the two subgroups.

To get a better idea of the relationships among the three taxa of *C. cardunculus* and among populations, we also performed a PCA based on Nei's genetic distances. The two first PCs explain 64·53 % of the total variation. Graphical presentation of PCA results provides a clear picture of the distribution of wild and cultivated material (Fig. 4). Three main groups can be distinguished: one (I) including the artichoke material, in the upper left side of the graph; the second group (II) in the left lower side, including cultivated cardoon and wild cardoon from Spain and Portugal; and finally the eastern wild cardoon samples forming a wide cluster (III), which can be subdivided into two subgroups (a and b). In group III, in fact, most Italian populations are close to samples from Greece and most from Malta; on the other hand, all populations from Tunisia, two from Sicily, and one from Malta form a distinct subgroup (a) and are placed in the middle of the plot, between the bigger subgroup of eastern wild cardoons (b) and the artichoke varieties.

DISCUSSION

The *C. cardunculus* complex is particularly interesting because, although corresponding to a single biological species, it includes two wild taxa and two crops; one of the cultigens, the artichoke, has been traditionally clonally propagated, by means of off-shoots, and the other one, the cultivated leafy cardoon, is propagated by seeds.
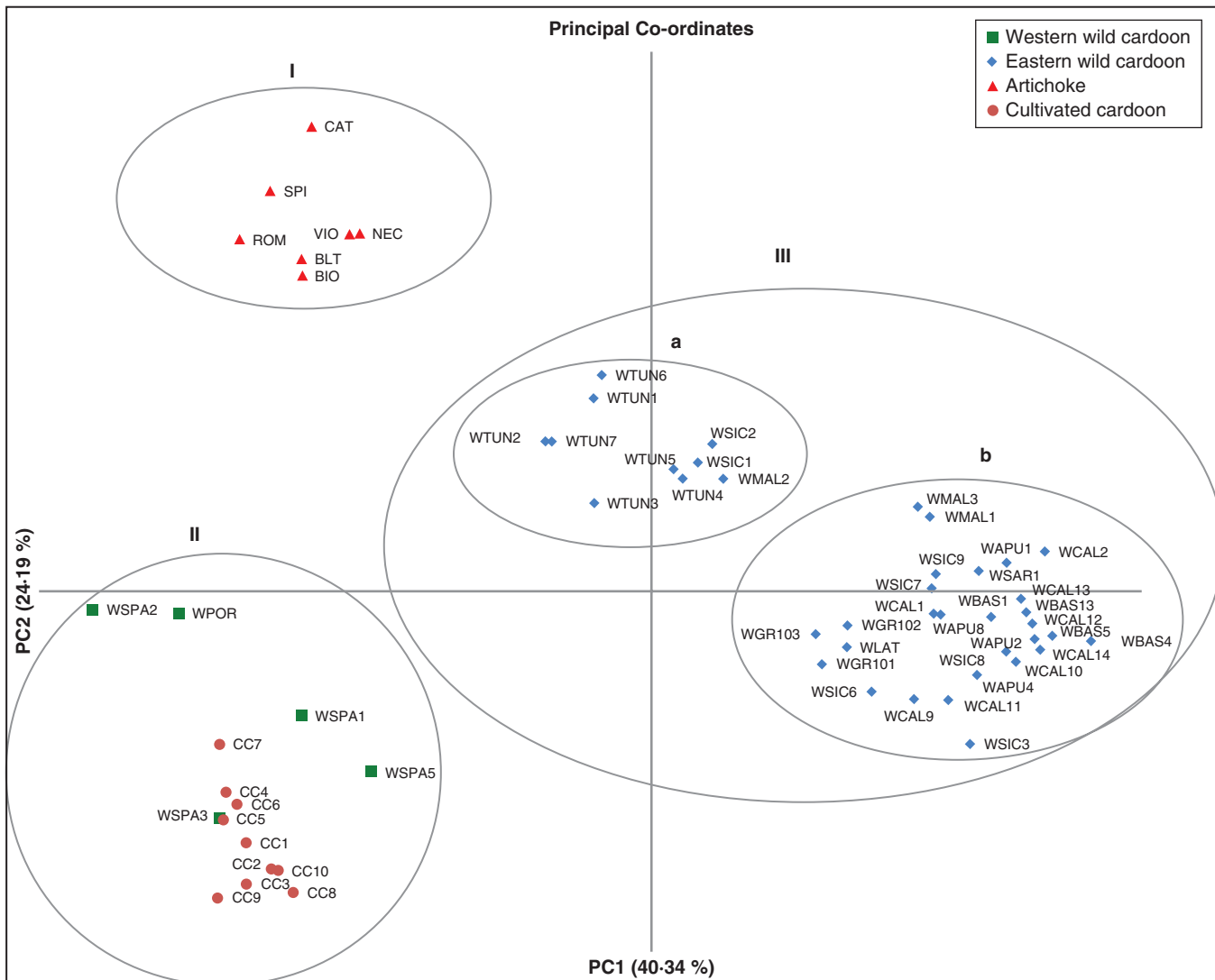

FIG. 4. Graphical representation of principal co-ordinate analysis based on Nei's genetic distances showing the distribution of *Cynara cardunculus* germplasm.

Here STRUCTURE analyses showed that, when considering the whole dataset, the highest value of ΔK obtained was not much higher than the other ΔK, indicating that the structure of populations was not well defined. Regardless, both at $K = 2$ and at $K = 3$, the western wild cardoons were grouped with the leafy cardoons, and not with the other wild cardoons. The difference between these two structures was that, in the former, the group of western wild cardoon/leafy cardoon also included the artichoke, while, in the latter, the artichoke samples formed a distinct group. Both morphologically (Wiklund, 1992; our pers. observ.) and genetically (Foury, 1989; Sonnante *et al.*, 2008; this study), wild cardoons from Spain and Portugal can be easily distinguished from the other wild cardoons and are quite similar to cultivated cardoon. However, it is possible that at least some of the wild populations present in Spain might have been derived from naturalization of the cultivated forms. In fact, wild cardoon, also referred to as artichoke thistle, is considered an invasive perennial forb in other continents, such as Australia and the Americas, where it probably escaped from cultivation (Thomsen *et al.*, 1986; Ariza-Espinar and Delucchi, 1998; Barker *et al.*, 2005; White and Holt, 2005). It has been observed that, in general, members of the family Asteraceae have a preference for open and disturbed habitats, thus facilitating weed formation (Sauer, 1993), and also in *Helianthus*, the evolutionary transition to weediness has occurred several times (Kane and Rieseberg, 2007).

The presence of the Spanish and Portuguese material in the STRUCTURE analysis of the wild populations somehow hid a very strong structure of the remaining material. In fact, when excluding the wild western germplasm, the Italian material also appeared to be structured, with a clear separation of the Tunisian group from the rest of the eastern populations. That Tunisian germplasm proved to be differentiated also from the western wild material is in agreement with the morphological observations of Foury (1989).

From the UPGMA tree and the PCA graph it appears that the wild Tunisian populations and some material from Sicily and Malta are more closely related to artichoke and therefore the area comprising these three territories could be where artichoke was domesticated. However, the centre of origin of a crop often corresponds to its centre of diversity, i.e. the area where the plant has existed the longest and has accumulated genetic variation over time, this high degree of variation being exhibited both at the population and at the genetic level (Gepts, 2004). Note that Italy represents a centre of diversity for the globe artichoke (Porceddu *et al.*, 1976; Sonnante *et al.*, 2003), and as we have shown here, also for the wild cardoon populations. This might suggest that the probable centre of origin of artichoke is southern Italy, particularly Sicily.

In our study, artichoke displayed a very high level of observed heterozygosity, confirming an excess in heterozygotes in this genetic material. Propagation material might have been selected by farmers on the basis of heterotic traits associated with a high level of heterozygosity (Sonnante *et al.*, 2007b, 2008); heterotic effects have also been noticed in hybrid F_1 individuals derived from crossing artichoke with its wild progenitor, leading to bigger and more robust plants (G. Sonnante, pers. observ.). On the other hand, lower levels of observed heterozygosity, compared with expected heterozygosity, were observed in the seed-propagated leafy cardoon and in the wild cardoon, indicating

that mating among relatives is allowed in wild populations. Indeed, in an F_1 map obtained by crossing artichoke × wild cardoon, by using a pseudo-test cross strategy, most co-dominant markers were mapped on the artichoke map as the parental wild cardoon showed a much lower level of heterozygosity (Sonnante *et al.*, 2011).

Dempewolf *et al.* (2008) suggested that globe artichoke and leafy cardoon be regarded as semi-domesticated crops, as both cultigens are not extensively cultivated and do not present major genetic alteration, although showing a considerable phenotypic differentiation and improvement through major breeding. However, note that selection occurred for gigantism of leaf stalks during domestication of leafy cardoon, and gigantism for flower head in globe artichoke. Plant architecture is also quite different between the two crops, the former showing a higher number of branches, with bigger plants, and the latter showing reduced branching and smaller plants. These differences could be the result of a lower domestication pressure for cultivated cardoon, while a stronger selection could have been operated regarding artichoke.

Based on the Roman writers Pliny the Elder (23–79 AD, in *Naturalis historia*) and Columella (1st century AD, in *De re rustica*), Foury (1989) assumed that the cultivation of artichoke started around the 1st century AD; on the basis of genetic analyses, Sonnante *et al.* (2007b) suggested that by that time the domestication of artichoke was an ongoing process. It has been suggested that the domestication of artichoke and cardoon occurred separately in time and space (Pignone and Sonnante, 2004; Sonnante *et al.*, 2007b). The difference between the domestication events of artichoke on the one side and leafy cardoon on the other is also evidenced by the BOTTLENECK results. In fact, the artichoke group was the only one to exhibit a bottleneck, even though it did not show a mode-shifted allele distribution, possibly because the bottleneck was not recent or small enough to be detectable (Luikart and Cornuet, 1998). Such a divergence between Wilcoxon test results and the mode-shift analysis was also observed in other plant and animal species (Besnard *et al.*, 2007; Shahsavarani and Rahimi-Mianji, 2010).

Although some authors (e.g. Foury, 1989) suggested that domestication of artichoke could have occurred later than that of leafy cardoon, from the literature and historical data (Sonnante *et al.*, 2007b), by analysing some typical traits of domestication, such as branching (see above), and also from the results obtained in this study, we consider that artichoke was domesticated before leafy cardoon. The clonal propagation of artichoke would have then fixed some specific traits and somehow slowed down its evolution. Vegetatively propagated crops are usually outcrossing plants, and inbred individuals suffer from inbreeding depression, often expressed in lower vigour. The highly heterozygous clones of landraces/varieties were presumably selectively chosen and multiplied by cultivators because they were vigorous. Sexual recombination would often break down favourable genetic combinations, while clonal multiplication preserves them (Rival and McKey, 2008). Therefore, in clonally propagated crops, genetic recombination usually does not contribute to the diversity of these crops, unless specifically used by some farmers to create new varieties; only somatic mutation can be the source of new allelic combinations (McKey *et al.*, 2010), although also transgenerational epigenetic inheritance can be the cause of heritable phenotypic variation (Jablonka and Raz, 2009).

Most studies on crop domestication have been carried out in seed-propagated plants (Gepts, 2004; Glémin and Bataillon, 2009; Sonnante *et al.*, 2009; McKey *et al.*, 2010). It can be assumed that as for other clonally propagated crops, sexual reproduction in artichoke has played a major role under domestication, and the selective pressure acting on sex was altered after the farmers had started propagating clonally (McKey *et al.*, 2010).

The artichoke material used in this study is comprehensive of a wide range of artichoke germplasm, as within each group, genotypes from different varietal types and originating from different farmers' fields were analysed, and some off-type material was also considered. Notwithstanding this wide range of artichoke samples, the average number of alleles for this cultigen was not very high. During the domestication process, a loss of genetic diversity is usually observed in crops, especially in modern varieties; in clonally propagated crops, the absence of sexual recombination leads to the loss of some components of diversity. This is why for some crops, such as cassava, volunteer plants reproduced by seed are sometimes grown together with clonal plants (Pujol *et al.*, 2005). In artichoke, new varietal types are also obtained by sexual reproduction, by crossing clonally propagated genotypes, and in recent years new hybrid varieties propagated by seeds are spreading (Calabrese *et al.*, 2011). Moreover, it is interesting to note that within the artichoke germplasm, the small-headed early flowering 'Catanesi' types show little variation within the group and are quite separated from the other artichoke varieties, except for 'Blanca de Tudela'. Conversely, the late, large-headed types considered as 'Romaneschi' are quite heterogeneous, and some of these artichoke types share some genetic background with the 'Spinosi' types, as previously observed (Sonnante *et al.*, 2003).

In conclusion, two hypotheses can be formulated on the domestication of *C. cardunculus*. In the first, which is more conservative, the globe artichoke was domesticated in more ancient times from wild material in Sicily/northern Africa, and the leafy cardoon was derived from western wild cardoon, following a low-level domestication process. In the second scenario, the eastern wild cardoon represents the only original wild form, from which both the globe artichoke and the cultivated cardoon originated; the leafy cardoon would have reverted to feral forms, giving rise to the so-called western wild cardoon. In our opinion, this latter hypothesis, which is the most parsimonious one, seems the most plausible, also given the geographical distribution of the species. *Cynara* species are prone to endemicity (Wiklund, 1992), while *C. cardunculus* has a wide distribution, probably due to an extension by cultivation, especially in the Iberian peninsula. According to this new hypothesis, the taxonomic entity of the western wild cardoon (*C. cardunculus* subsp. *flavescens*) proposed by Wiklund (1992) should be revised.

SUPPLEMENTARY DATA

Supplementary data are available online at www.aob.osfordjournals.org and consist of the following. Table S1: description of the *Cynara cardunculus* material analysed. Table S2: SSR markers screened and relative number of alleles; observed and expected heterozygosity; polymorphic information content. Table S3: null allele frequencies for each SSR locus in all populations. Table S4: genetic diversity parameters for the eastern wild cardoon populations. Figure S1: STRUCTURE analysis of Tunisian and western wild cardoon, together with leafy cardoon and artichoke at $K = 3$. Figure S2: STRUCTURE plot for cultivated cardoon at $K = 2$. Figure S3: STRUCTURE analysis on artichoke samples at $K = 2$ and $K = 4$.

LITERATURE CITED

Ariza-Espinar L, Delucchi G. 1998. 280 Asteraceae, parte 11. Tribu XI. Cardueae. In: Ariza-Espinar L.ed. *Flora Fanerogámica*. Argentina: Córdoba Pro Flora, 24–25.

Barker B, Barker R, Jessop J, Vonow H. 2005. *Census of South Australian Vascular Plants*, 5th edn. Adelaide: The Botanic Gardens of Adelaide and State Herbarium.

Barres L, Sanmartín I, Anderson CL, *et al.* 2013. Reconstructing the evolution and biogeographic history of tribe Cardueae (Compositae). *American Journal of Botany* **100**: 1–16.

Bazniski J, Zohary D. 1994. Breeding of seed planted artichoke. *Plant Breeding Reviews* **12**: 253–269.

Besnard G, Henry P, Wille L, Cooke D, Chapuis E. 2007. On the origin of the invasive olives (*Olea europaea* L., Oleaceae). *Heredity* **99**: 608–19.

Calabrese N, Carito A, Boari F, Cantore V, De Palma E, Damato G. 2011. Agronomic evaluation of artichoke cultivar propagated by seed. *Acta Horticulturae* **942**: 153–158.

Chapuis MP, Estoup A. 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* **24**: 621–631.

Chen C, Durand E, Forbes F, François O. 2007. Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes* **7**: 747–756.

Chybicki IJ, Burczyk J. 2009. Simultaneous estimation of null alleles and inbreeding coefficients. *Journal of Heredity* **100**: 106–113.

Cornuet JM, Luikart G. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**: 2001–2014.

Dempewolf H, Rieseberg LH, Cronk QC. 2008. Crop domestication in the Compositae: a family-wide trait assessment. *Genetic Resources and Crop Evolution* **55**: 1141–1157.

Dereeper A, Guignon V, Blanc G, *et al.* 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research* **36**: W465–469.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611–2620.

Foury C. 1989. Ressources génétiques et diversification de l'artichaut (*Cynara scolymus* L.). *Acta Horticulturae* **242**: 155–166.

François O, Ancelet S, Guillot G. 2006. Bayesian clustering using hidden Markov random fields in spatial population genetics. *Genetics* **174**: 805–816.

Fratianni F, Tucci M, De Palma M, Pepe R, Nazzaro F. 2007. Polyphenolic composition in different parts of some cultivars of globe artichoke (*Cynara cardunculus* L. var. *scolymus* (L.) Fiori). *Food Chemistry* **104**: 1282–1286.

Gebhardt R. 1997. Antioxidative and protective properties of extracts from leaves of the artichoke (*Cynara scolymus* L.) against hydroperoxide-induced oxidative stress in cultured rat hepatocytes. *Toxicology and Applied Pharmacology* **144**: 279–286.

Gepts P. 2004. Crop domestication as a long-term selection experiment. *Plant Breeding Reviews* **24**: 1–44.

Gil R, Villa F. 2004. Breeding for earliness on seed propagated globe artichoke. *Acta Horticulturae* **660**: 35–37.

Glémin S, Bataillon T. 2009. A comparative view of the evolution of grasses under domestication. *New Phytologist* **183**: 273–290.

Goudet J. 1995. FSTAT Version 1·2: a computer program to calculate *F*-statistics. *Journal of Heredity* **86**: 485–486.

Jablonka E, Raz G. 2009. Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Quarterly Review of Biology* **84**: 131–176.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**: 1801–1806.

Kane N, Rieseberg LH. 2007. Genetics and evolution of weedy *Helianthus annuus* populations: evidence of multiple origins of an agricultural weed. *Molecular Ecology* **17**: 384–394.

Kraft K. 1997. Artichoke leaf extract: recent findings reflecting effects on lipid metabolism, liver and gastrointestinal tracts. *Phytomedicine* **4**: 369–378.

Luikart G, Cornuet JM. 1998. Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology* **12**: 228–237.

McKey D, Elias M, Pujol B, Duputié A. 2010. The evolutionary ecology of clonally propagated domesticated plants. *New Phytologist* **186**: 318–332.

Negro D, Montesano V, Grieco S, *et al*. 2012. Polyphenol compounds in artichoke plant tissues and varieties. *Journal of Food Science* **77**: C244–252.

Nei M. 1972. Genetic distance between populations. *American Naturalist* **106**: 283–291.

Peakall R, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**: 288–295.

Pignone D, Sonnante G. 2004. Wild artichokes of south Italy: did the story begin here? *Genetic Resources and Crop Evolution* **51**: 577–580.

Piry S, Luikart G, Cornuet JM. 1999. BOTTLENECK: a computer program for detecting recent reductions in effective population using allele frequency data. *Journal of Heredity* **90**: 502–503.

Porceddu E, Dellacecca V, Bianco VV. 1976. Classificazione numerica di cultivar di carciofo. In: *Proceedings of the II International Congress on Artichoke, Bari 1973*. Torino: Minerva Medica (in Italian), 1105–1119.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Pujol B, David P, McKey D. 2005. Microevolution in agricultural environments: how a traditional Amerindian farming practice favours heterozygosity in cassava (*Manihot esculenta* Crantz, Euphorbiaceae). *Ecology Letters* **8**: 138–147.

Raccuia SA, Melilli MG. 2004. *Cynara cardunculus* L., a potential source of inulin in the Mediterranean environment: screening of genetic variability. *Australian Journal of Agricultural Research* **55**: 693–698.

Rival L, McKey D. 2008. Domestication and diversity in Manioc (*Manihot esculenta* Crantz ssp. *esculenta*, Euphorbiaceae). *Current Anthropology* **49**: 1119–1128.

Rottenberg A, Zohary D. 1996. The wild ancestry of the cultivated artichoke. *Genetic Resources and Crop Evolution* **43**: 53–58.

Rottenberg A, Zohary D. 2005. Wild genetic resources of cultivated artichoke. *Acta Horticulturae* **681**: 307–311.

Sauer JD. 1993. *Historical geography of crop plants*. Boca Raton, FL: CRC Press.

Schuelke M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* **18**: 233–234.

Shahsavarani H, Rahimi-Mianji G. 2010. Analysis of genetic diversity and estimation of inbreeding coefficient within Caspian horse population using microsatellite markers. *African Journal of Biotechnology* **9**: 293–299.

Sonnante G, De Paolis A, Pignone D. 2003. Relationships among artichoke cultivars and some related wild taxa based on AFLP markers. *Plant Genetic Resources: Characterization and Utilization* **1**: 125–133.

Sonnante G, Carluccio AV, Vilatersana R, Pignone D. 2007*a*. On the origin of artichoke and cardoon from the *Cynara* gene pool as revealed by rDNA sequence variation. *Genetic Resources and Crop Evolution* **54**: 483–495.

Sonnante G, Pignone D, Hammer K. 2007*b*. The domestication of artichoke and cardoon: from Roman times to the genomic age. *Annals of Botany* **100**: 1095–1100.

Sonnante G, Carluccio A, De Paolis A, Pignone D. 2008. Identification of artichoke SSR markers: molecular variation and patterns of diversity in genetically cohesive taxa and wild allies. *Genetic Resources and Crop Evolution* **55**: 1029–1046.

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

Sonnante G, D'Amore R, Blanco E, *et al*. 2010. Novel hydroxycinnamoyl-Coenzyme A quinate transferase genes from artichoke are involved in the synthesis of chlorogenic acid. *Plant Physiology* **153**: 1–15.

Sonnante G, Gatto A, Morgese A, *et al*. 2011. Genetic map of artichoke × wild cardoon: toward a consensus map for *Cynara cardunculus*. *Theoretical and Applied Genetics* **123**: 1215–1229.

Takezaki N, Nei M, Tamura K. 2010. POPTREE2: software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. *Molecular Biology and Evolution* **27**: 747–52.

Thomsen CD, Barbe G, Williams W, George M. 1986. Escaped artichokes are troublesome pests. *California Agriculture* **40**: 7–9.

White VA, Holt JS. 2005. Competition of artichoke thistle (*Cynara cardunculus*) with native and exotic grassland species. *Weed Science* **53**: 826–833.

Wiklund A. 1992. The genus *Cynara* L. (Asteraceae-Cardueae). *Botanical Journal of the Linnean Society* **109**: 75–123.

Yeh FC, Boyle TJB. 1997. Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belgian Journal of Botany* **129**: 157.