Genet Resour Crop Evol (2008) 55:1029–1046
DOI 10.1007/s10722-008-9310-5

RESEARCH ARTICLE

Identification of artichoke SSR markers: molecular variation and patterns of diversity in genetically cohesive taxa and wild allies

Gabriella Sonnante · Anna Vittoria Carluccio · Angelo De Paolis · Domenico Pignone

Received: 5 November 2007 / Accepted: 28 January 2008 / Published online: 29 February 2008

Abstract A set of 24 microsatellite markers was identified in the artichoke genome, using various approaches. A genomic library allowed the development of 14 SSR markers, whereas the other 10 were obtained from gene intron/UTR regions or from other species. Allelic variation was scored in *C. cardunculus* (artichoke, cultivated cardoon, and wild cardoon) samples, and in other wild *Cynara* allies. For the 23 polymorphic loci, a total of 165 alleles were scored, 135 of which in the artichoke primary genepool, and the remaining ones in the other *Cynara* species. Some allele combinations were able to identify artichoke varietal types, and some alleles were unique to specific groups. This makes these markers potentially useful in product traceability and in contributing to the saturation of genetic maps. The percentage of shared alleles between *C. cardunculus* taxonomic groups, and Nei's genetic distances indicated that wild cardoons from the Eastern Mediterranean were more closely related to artichoke and less to cultivated cardoon in comparison to wild cardoons from the Western Mediterranean, and the genetic distance between the two wild cardoon genepools was rather high. The UPGMA dendrogram based on Nei's genetic distances revealed that artichokes formed a fairly defined cluster, whereas Eastern wild cardoons occupied another branch, and Western wild cardoons were clustered together with cultivated cardoons. The transferability of microsatellite markers to other *Cynara* wild species was quite good. Sequencing alleles at three loci showed that, apart from microsatellite length variation, point mutations and insertion/deletions were quite abundant especially when comparing *C. cardunculus* to the other *Cynara* species. In the sequenced regions, some SNPs were identified which distinguished artichoke on one side, and cultivated and wild cardoon on the other, while other SNPs were apportioned according to the geographic distribution of *Cynara* wild species.

Keywords Artichoke · Cardoon · *Cynara* · Genetic diversity · Microsatellites · Wild relatives

G. Sonnante (✉) · A. V. Carluccio · D. Pignone
National Research Council (CNR) – Institute of Plant Genetics, Via Amendola, 165/A, 70126 Bari, Italy
e-mail: gabriella.sonnante@igv.cnr.it

A. De Paolis
CNR – Institute of Sciences of Food Production, Sez. di Lecce, Via Prov.le Lecce Monteroni, 73100 Lecce, Italy

Introduction

Globe artichoke (*Cynara cardunculus* var. *sativa* Moris, syn. *C. cardunculus* L. var. *scolymus* (L.) Fiori) is a diploid (2n = 2x = 34) outcrossing species, originating in the Mediterranean Basin and mostly cultivated for its edible immature flower heads. It is a polyannual crop, mainly based on vegetative reproduction by means of offshoots, even though in the latest years seed reproduced artichokes

are gaining growing attention (Basnizki and Zohary 1994; López Anido et al. 1998; Calabrese and Bianco 2000). Its cultivation is largely diffused in Europe, where Italy represents the main producer, and to a lesser extent in other Mediterranean countries (Egypt, Tunisia, Morocco, Turkey, Israel), in South America, in California, and more recently also in China (Bianco 2005; FAO http://faostat.fao.org/).

The species *C. cardunculus* also includes another crop, the cultivated leafy cardoon (*C. cardunculus* L. var. *Cardunculus*, syn. var. *altilis* DC.), used for its large and fleshy leaf stalks and midrips, and a wild entity, the wild cardoon (*C. cardunculus* L. var. s*ylvestris* Lam.), the presumed wild progenitor of the two crops (Zohary and Basnizki 1975; Foury 1989; Wiklund 1992; Rottenberg and Zohary 1996; Rottenberg et al. 1996; Sonnante et al. 2007a, b). These three taxa are fully interfertile (Basnizki and Zohary 1994; Rottenberg and Zohary 1996, 2005), and therefore cultivated and wild cardoons belong to the primary genepool of artichoke. Recent studies have suggested that a high level of differentiation is present in the wild cardoon genepool, and that samples from the Western Mediterranean range more closely resemble the cultivated cardoon than the wild samples from the Eastern Mediterranean one (Wiklund 1992; Sonnante et al. 2007a, b).

Four main morpho-agronomic groups have been identified in artichoke: the spiny early types ('Spinosi'), the violet late types ('Violetti'), the small headed early types ('Catanesi') and the large headed late types ('Romaneschi') (Porceddu et al. 1976). Apart from these major varietal groups, many landraces are grown, some of which not falling into the main classification. This occurrence is especially frequent in Italy, possibly because this is the place where artichoke domestication was first attempted (Pignone and Sonnante 2004; Sonnante et al. 2007a, b). Italy therefore represents the largest reservoir of artichoke germplasm and rare types (Porceddu et al. 1976; Elia and Miccolis 1996; Hammer et al. 1999; Sonnante et al. 2004a, b).

No intensive breeding programs have been carried out for artichoke, in contrast to what done for other horticultural crops. The only real novelty in this sense is the establishment of few seed propagated hybrid lines (Basnizki and Zohary 1994). For this reason, a better knowledge of artichoke genetics becomes mandatory in order to plan new strategies for breeding this crop. As a consequence, there has been a growing interest in artichoke genetics, as testified by the increasing number of scientific publications produced in this field in the latest years (Sonnante et al. 2007b, and references therein).

Microsatellites, also known as simple sequence repeats (SSRs), are a class of codominant molecular markers constituted by tandem repeats of a short (1–6 bp) core sequence. They are considered one of the most useful classes of markers in plant breeding, widely exploited for cultivar fingerprinting, germplasm characterization, and genome mapping (Röder et al. 1998; Jain et al. 2004; This et al. 2004; Viruel et al. 2005; Lacape et al. 2007). Moreover, SSR transferability across related species makes these markers very powerful for comparative genetic studies and also for the understanding of SSR evolution (Kahru et al. 2000; Matsuoka et al. 2002; Rohrer et al. 2004; Sethy et al. 2006).

So far, few microsatellite sequences have been isolated from artichoke, and only 12 of them were informative in an artichoke map based upon a pseudo-testcross approach (Lanteri et al. 2006, and references therein); in this strategy, codominant markers are particularly useful for bridging between intraparental maps (Grattapaglia and Sederoff 1994). For the above reasons, the development of new SSR loci is of great importance.

The objectives of the present study were to develop new microsatellites from artichoke, using various approaches, to characterize them within the species *C. cardunculus*, to verify their possible applicability to variety identification, to estimate the distribution of genetic diversity, and to assess genetic relationships within the artichoke primary genepool. Moreover, we evaluated the feasibility of transferring the detected microsatellites to other species of the genus *Cynara*.

Material and methods

Plant material and DNA extraction

The plant material analysed included 29 artichoke varieties or landraces, 6 samples of cultivated cardoons, 10 wild cardoon accessions, 6 samples representing 4 wild *Cynara* species, namely *C. baetica* (Spreng.) Pau, *C. cornigera* Lindley, *C. humilis* L., and *C. syriaca* Boiss (Table 1). In the text, when referring to wild material we used the terms 'wild cardoons' for *C. cardunculus* var. *sylvestris*, and 'wild *Cynara* species' for all the wild species not belonging to *C. cardunculus*.

Portions of young leaves were collected, frozen in liquid N_2 and stored at −80°C; DNA was extracted from this tissue according to Sonnante et al. (2002). A high amount of DNA was obtained by using 5 g of frozen leaves of the variety 'Mola' for library construction.

Microsatellite isolation

Various approaches were used to identify new SSR markers in artichoke, one of which was based on the use of a genomic library.

Library approach. Genomic DNA from the variety 'Mola' was partially digested with the enzyme *Sau*3AI and subjected to a continuous sucrose gradient. DNA fragments of approximately 4/6 kb long were identified and ligated into the *Bam*HI site of the ZAP Express vector and packaged according to manifacturer's instruction kit (Stratagene, La Jolla, CA). Approximately 1,000 plaques were plated, and filters of the library were lifted, processed, and fixed by standard methods (Sambrook and Russell 2001). Two replica filters were prepared and hybridized with digoxigenin labelled sinthetic oligonucleotides $(GATA)_4$ and $(CAT)_8$, respectively (Roche Diagnostic, Milan, Italy). Hybridization of the filters was carried out at 5° below the oligo-Tm. After hybridization, filters were washed twice at room temperature in 2× SSC (0.3 M NaCl, 0.03 M sodium citrate), 0.1% SDS for 5 min, and twice at the hybridization temperature with 0.5× SSC, 0.1% SDS for 20 min. Signal detection was performed according to the DIG chemiluminescent detection protocol (Roche Diagnostic). Putative positive clones were subjected to three rounds of screening, and the resulting phagemids excised by using ExAssist helper phage (Sambrook and Russell 2001). Plasmid DNA was isolated from the correspondent pBK-CMV recombinant clones using the QIA Miniprep kit (Qiagen, Milan, Italy). Sequencing of the inserts was performed using an automated sequencer (CEQ 8800, Beckman Coulter, Fullerton, USA).

Other approaches. Additional SSR sequences were obtained from other sources: (i) SSR repeats were detected in a *pal* gene isolated from artichoke (De Paolis et al. 2008); (ii) primers amplifying SSR, isolated from *Cirsium acaule* (L.) Scop. (Jump et al. 2002), were tested in artichoke; (iii) one more SSR sequence was detected in the artichoke 1-SST gene (EMBL accession N. Y09662); (iv) three additional SSR sequences were detected in artichoke ESTs isolated at the Institute of Plant Genetics, Bari.

Primer design, PCR reactions and detection of amplified fragments

Obtained sequences were searched for microsatellite repeats using the programme Sputnik (http://cbi.labri.fr/outils/Pise/sputnik.html). Primers for SSR amplification were designed on the flanking regions of microsatellite repeats using Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi) or PrimerQuest (http://eu.idtdna.com/Scitools/Applications/Primerquest) programmes, so that amplification products were in the range of 140–420 bp. For each pair, the forward or reverse primer was labeled at its 5′ end using a fluorescent dye. Labelled primers were synthesized by Sigma-Proligo (Milan, Italy), or Merigen® Research (Naples, Italy).

Polymerase chain reactions were performed in a 25 µl final volume containing 25 ng DNA, 5 pM of forward and reverse primer, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP and 1 U of Taq DNA polymerase (AmpliTaq Gold™, Applied Biosystem for touchdown PCRs; JumpStart Taq, Sigma-Aldrich for the remaining reactions) in the manufacturer supplied buffer. All PCR reactions were performed in a 9700 Thermal Cycler (Applied Biosystems). Hot start was carried out using the following profile: 94°C for 10 min, followed by 2 long cycles (1 min at 94°C, 1 min at optimal annealing temperature—AT, Table 2—and 1 min at 72°C), then by 33 cycles (30 s at 94°C, 30 s at AT, and 45 s at 72°C), and ending with 30 min extension at 72°C. In some cases, a touchdown PCR protocol was used: 94°C for 10 min, followed by 10 cycles of 30 s at 94°C, 30 s at AT (decreasing of 0.5°C every cycle from initial to final value), and 45 s at 72°C, then 25 cycles of 30 s at 94°C, 30 s at final AT and 45 s at 72°C, with 30 min final extension at 72°C.

Capillary electrophoresis and detection of PCR products were carried out on an automated fragment analyzer CEQ 8800 (Beckman Coulter).

Cloning and sequencing of length variants

Some allele variants at three different loci were sequenced. Amplificates were purified using the Qiaquick PCR purification kit (Qiagen), cloned in

Table 1 List of the material analysed

No.	Code	Variety/ecotype/species	Origin
1	AR-CAT	Catanese	Italy
2	AR-CAT	Mola 1	Italy
3	AR-CAT	Mola 2	Italy
4	AR-CAT	Masedu	Italy
5	AR-CAT	Violetto di Provenza	Italy
6	AR-CAT	Banfsigi	Egypt
7	AR-CAT	Baladi	Egypt
8	AR-VIO	Precoce Violetto di Chioggia	Italy
9	AR-VIO	Moretto	Italy
10	AR-VIO	Violetto di Toscana	Italy
11	AR-VIO	Violetto di Maremma	Italy
12	AR-SPI	Spinoso Violetto di Liguria	Italy
13	AR-SPI	Spinoso Sardo	Italy
14	AR-SPI	Spinoso di Palermo	Italy
15	AR-SPI	Spinoso Sciacca	Italy
16	AR-ROM	Romanesco	Italy
17	AR-ROM	Tondo di Paestum	Italy
18	AR-ROM	Mazzaferrata	Italy
19	AR-ROM	Centofoglie	Italy
20	AR-ROMl	Bayrampasa	Turkey
21	AR-ROMl	Macau	Italy
22	AR-ROMl	Ñato	Argentina
23	AR-ROMl	Green Globe	California (USA)
24	AR-OUT	Tudela	Spain
25	AR-OUT	Verde di Pesaro	Italy
26	AR-OUT	Sakiz	Turkey
27	AR-OUT	Testa di Ferro	Italy
28	AR-OUT	Verde di Putignano	Italy
29	AR-OUT	Bianco Tarantino	Italy
30	CC	Bianco Avorio	Italy
31	CC	Minerbio	Italy
32	CC	Gigante di Romagna	Italy
33	CC	Blanco Peralta	Spain
34	CC	Verde Peralta	Spain
35	CC	Tafalla	Spain
36	WC	Agrigento	Italy
37	WC	Calopezzati	Italy
38	WC	Gannano	Italy
39	WC	Porto Cesareo	Italy
40	WC	Tolfa	Italy
41	WC	Oristano	Italy
42	WC	V311A	Spain
43	WC	Corfu	Greece
44	WC	Girona	Spain
45	WC	Lleida Avellanet	Spain

Table 1 continued

No.	Code	Variety/ecotype/species	Origin
46	Spp	*C. syriaca* 1	Israel, Mount Tabor
47	Spp	*C. syriaca* 2	Israel, Tivon
48	Spp	*C. humilis* 1 (CYN 11/94)	Spain
49	Spp	*C. humilis* 2 (GR63)	Portugal
50	Spp	*C. baetica*	Spain
51	Spp	*C. cornigera*	Egypt

AR, artichoke; CC, cultivated cardoon; WC, wild cardoon; Spp, other *Cynara* species. Within artichoke, CAT, 'Catanesi'; VIO, 'Violetti'; SPI, 'Spinosi'; ROM, 'Romaneschi'; ROMl, 'Romaneschi' like; OUT, not falling in any morpho-agronomic group. Variety refers to crops; ecotype to wild cardoons; species to the other *Cynara* species. Samples were from CNR-Institute of Plant Genetics collection, except for No. 33–35 (Dr. Macua, ITGA, Tudela, Spain); 44, 45, 50, 51 (Dr. Vilatersana, CSIC, Barcelona, Spain); 46–47 (Prof. Lev-Yadun, University of Haifa, Israel); 48, (IPK, Gatersleben, Germany), and 49 (Botanic garden, Barcelona, Spain)

TOPO TA cloning vector (Invitrogen, San Giuliano Milanese, Italy) and more than one clone per sample was sequenced using the CEQ 8800 sequencer (Beckman Coulter) with the appropriate programme for clone sequencing. Subsequently, sequences were aligned using the program CLUSTAL W available on line (http://www.ebi.ac.uk/clustalw/).

Data analysis

The resulting electropherograms were analyzed using the software Genetic Analysis System v. 9.0 (Beckman Coulter) and manually checked. Observed (Ho) and Nei's (1978) expected (He) heterozygosity, and Shannon's Information index (I, Lewontin 1972) were calculated only for the *C. cardunculus* samples using the software POPGENE 1.31 (Yeh et al. 1997). Pairwise genetic distances (Nei 1972) were calculated by means of the same software using groupings related to taxonomy (Wiklund 1992) or to the morph-groups following Porceddu et al. (1976). The pairwise genetic distance matrix was used to generate a UPGMA dendrogram.

Results

Microsatellite identification, amplification, and detection

The artichoke microsatellite regions here presented (Table 2) were identified using various approaches.

An artichoke genomic library was used to identify microsatellite containing fragments. Screening of about 1,000 plaques with the $(CAT)_8$ probe brought to the identification of 6 positive plaques, two of which, showing a very strong signal were further processed to obtain isolated plaques. Southern blot analysis of these two clones digested with different restriction enzymes and hybridised with $(CAT)_8$ probe (data not shown) identified a single strong signal on an *Rsa*I fragment of about 1,200 bp for the first clone, and an *Apa*I fragment of about 7 kb for the second one. Similarly, by using $(GATA)_4$ as a probe other positive signals were recorded and two of them further processed to get isolated plaques. These two clones were sequenced and revealed to contain inserts of about 1,200 bp and about 1,500 bp, respectively. All the obtained sequences were analysed for microsatellite repeats. This approach allowed the identification of 14 microsatellite regions (CsLib01-CsLib14 loci).

Other 10 SSRs were obtained from different approaches. Within the sequence of an artichoke gene coding for phenylalanine ammonia-lyase (*CSpal3*, De Paolis et al. 2008), four microsatellites were detected, two of which in the intron (CsPal01 and CsPal02 loci), and two in the 5′ UTR region (CsPal03 and CsPal04 loci). Three additional SSRs were identified in RACE products obtained at IGV during gene isolation experiments in *Cynara* (CsEST01-CsEST03 loci). Three microsatellite sequences were retrieved from the EMBL sequence database: one (Cs1-SST locus) was obtained from artichoke 1-SST coding sequence (EMBL accession N. Y09662) and two were derived

Table 2 Locus name, core repeat, primer sequences (5′–3′), optimal annealing temperature (AT), expected size (ES), observed size range (OSR), number (No) of alleles, and EMBL accession number for the 24 artichoke microsatellite loci described

Locus	Repeat motif	Primer pair sequence (5′–3′)	AT	ES (bp)[a]	OSR (bp)	No alleles[b]	EMBL Acc. No.
CsLib01	$(CA)_5c(CA)_4$	F: GAT CCA GAG CGA CAC R: AGA ATC CAA TGC TCA GGG A	58/52TD	193	196–215[b]	4/0	AM495260
CsLib02	$(ATC)_4acc(ATC)acc(ATC)$	F: GCC TAG CCA GGG CAA ACA AAT GAA R: TAT TGG CAA CAA ACG GTG GTG GTG	70/65 TD	206	170–224	8/4	AM495261
CsLib03	$(TGCC)_3tgct(TGCC)_2...(TG)_5...$ $(TAA)_3tta(TAA)_2...(GA)_5$	F: TTT ATG CGT GTA AGT CCG CT R: ATC CCT CTT CGT CAT TCG CCT	58	173	168–186	9/1	AM495262
CsLib04	$(AGG)_6...(CCA)_8$	F: AGC TTG GCC ACC GTT AAT TGA G R: ATT ATG CCG GGC CAG ATT CAA C	70/65TD	161	146–167	8/0	AM495263
CsLib05	$(CCA)_3…(GCC)_2gct(GCC)_4$	F: GTT GAA TCT GGC CCG GCA TAA T R: TCT TAT GCG TCA TCA TCG CCA C	67/62TD	167	158–176	6/1	AM495264
CsLib06	$(A)_6...(TG)_{11}...(T)_7$	F: GCC AGA TTG TGT GCT TGG TCA T R: CGA CGT CAA TCG AAT GAT TC	55	292	286–304	10/0	AM495265
CsLib07	$(AC)_9at(AC)_4$	F: CGA CGT CAA TCG AAT GAT TC R: CTC GTT CCC TGA TTA GCA AAG TG	50	144	142–146	3/0	AM495266
CsLib08	$(GA)_8$	F: CAC GCT CAC TAA TAG ACA TGC AC R: AGA TCA CAA CCA CAC AGT TGC G	59	171	167–175	4/0	AM497817
CsLib09	$(GAG)_9$	F: GGG CGA AGC GAA ATA GGC ACA TAA R: ACA TTT CTC GCT GTC TGC CTG A	65/60 TD	162	159–162	2/0	AM497818
CsLib10	$(GAT)_{10}...(GAGT)_3$	F: TGT TCA GGC AGA CAG CGA GAA A R: ACA ACG ACA GCG ACT ACG ATG A	65/60 TD	244	235–253	8/1	AM497819
CsLib11	$(CTAA)_5(CTCA)_4$	F: ACA ACA GTC CCG TTC AAG CCA T R: AAT AGT GTA TGC CGG ACG GTG A	60	245	241–254	5/0	AM497820
CsLib12	$(GT)_5tt(GT)_2...(GT)_3t(G)_{10}$	F: ATA GCC AGC AAT GAA ATC TT R: ATG AAC GCT GAA TCA CGT G	50	282	232–316	9/3	AM497821
CsLib13	$(GTG)_4...(CTG)_7$	F: ACA TGG TTG GGC ATG GCT TGT T R: ACG CAG GCA ACC ACC AAT GTT	65/60 TD	226	224–252	6/0	AM497822
CsLib14	$(TCA)_2tcc(TCA)_3...(TCA)_9$	F: GAT GGT AAA TTC ATC TTC ATC R: TCT GCT CCA GCT GAG GAT GTT T	58	193	151–200	6/1	AM497823
CsPal01	$(TA)_{16}$	F: TTC GGA GCT ACC TCT CAC C R: AAT TTA GAT TGA AAT CGT ATG AGG A	55	227	186–234	10/3	AM497824
CsPal02	$(TA)_{20}$	F: GAA TTG GGT TTA GGT AGA TTG AGT G R: TGG CTG CTC TTG TTG CTG AAT GTG	60	365	314–379	17/4	AM497825

Table 2 continued

Locus	Repeat motif	Primer pair sequence (5′–3′)	AT	ES (bp)[a]	OSR (bp)	No alleles[b]	EMBL Acc. No.
CsPal03	$(TA)_5CA(TA)_2(CATA)_4(CA)_6$	F: ACT TGC CTT CTC TTG TGC CTA CCT	62	334	329–349	9/2	AM497826
		R: TCC GCA ACC ATT CTC TTC ACC TCA					
CsPal04	$(TA)_5...(TA)_{26}$	F: CAA GAG TGG TAG TTG GTG GGT GTT	57/52 TD	414	349–426	18/4	AM497830
		R: GGT GCT TAT GCA GCC AAC CAA TCA					
CsEST01	$(GA)_5$	F: CAA GAA CAA CCC ACA ACC GCA GAA	58[c]	180	182[b]	1/0	AM497827
		R: ATT GAA CGC CGG CCA CTC AAT ATG					
CsEST02	$(GAA)_2...(TC)_5$	F: ACA AAC TGT GGT CAA TGG TCC CTC	58[c]	301	300–301	2/0	AM497828
		R: CAT GTT TAG CAC GAC AAC TCA CAG					
CsEST03	$(CCA)_4$	F: AAA TCA CAC AAC ACC ACC ACC ACC	62	184	184–206	3/0	AM497829
		R: AGC CAG TTT CCA AAG ATG GGT ACG					
Cs1-SST	$(GAT)_6...(GAT)_4...(GAT)_5...(TGG)_4$	F: AAG CAC AAC TGG ATC CAT TC	55	257	257–315*	11/3	AM4978438
		R: AAA TAT AAT CTC ACA AGT GGA					
CsCiCcaca01	$(AC)_3$	F: TTT GAA GTG GAT CTT CGC ACG	58[c]	196	185–200	3/2	AM497869
		R: CAT GGG AGA CGA ACT AAC AGA TGC					
CsCiCaca05	$(AC)_3(TC)_2$	F: ACC CAA CCC TCG ATC TGA A	55[c]	146	130–148	4/1	AM497854
		R: GAG GAT ACC GGC GAT TGT TA					

TD, touch down PCR protocol

[a] Expected size of the amplification products obtained in the variety 'Locale di Mola'

[b] Observed allele size not corresponding to the expected one. See text

[c] The first number refers to the total number of alleles in the whole set of samples, the second one refers to the alleles found only in wild *Cynara* species

from microsatellites described in *Cirsium acaule*, a species belonging, like *Cynara*, to the Cardueae tribe of the Asteraceae family (Caca01 AJ457836, and Caca05 AJ457840; Jump et al. 2002). We decided to rename these two latter loci as CsCiCaca01 and CsCiCaca05 respectively, since the artichoke sequences differed from those of *Cirsium*.

A total of 24 primer pairs were designed and used for the amplification of the above described microsatellite loci (Table 2).

Microsatellite polymorphism

All the designed primers produced an amplification in the artichoke samples. However, no product was obtained for wild cardoons and the wild *Cynara* species at the loci CsLib11 and CsLib06, for *C. humilis* at the locus CsLib05, for *C. baetica* at the loci CsLib03, CsLib10, and CsLib14, and for *C. cornigera* at the loci CsLib01, CsLib03, CsLib10, and CsLib12. The lack of amplification was confirmed by repeating the PCR experiments at least three times.

The locus CsEST01 was monomorphic for all samples, and the locus CsCiCaca01 was monomorphic only in *C. cardunculus*, but polymorphic when considering the wild *Cynara* species. A total of 165 alleles were scored at the 23 polymorphic loci analysed; 135 of them were found in *C. cardunculus*, 37 of which unique to artichoke, 6 to cultivated cardoon, and 12 to wild cardoon; the remaining 30 alleles were present only in the wild *Cynara* species. The number of alleles per locus ranged from 1 to 18 (average 6.9) when all samples were taken into account (Table 2).

Parameters of genetic diversity for *C. cardunculus* as a whole, and for artichoke (Table 3) were calculated for polymorphic loci using the POPGENE package (Yeh et al. 1997). The number of alleles per locus ranged from 2 to 10 (average 4.68) in artichoke, and from 2 to 14 (average 6.10) in *C. cardunculus*. The observed heterozygosity (Ho) ranged from 0 to 1, with a mean of 0.41 and 0.39 in artichoke and all *C. cardunculus* samples, respectively. Shannon's Information index (I) ranged between 0.09 and 2.08 in artichoke, and between 0.50 and 2.29 in all *C. cardunculus* samples.

Some allelic combinations appeared to specifically relate to a varietal or taxonomic group. For instance, the alleles at the locus Cs1-SST seemed to be informative both within artichoke varietal types, and between artichoke and cardoon. In fact, at this locus the 'Catanesi' genotypes were all heterozygous 266/296, the 'Violetti' artichokes were homozygous 263/263, the 'Spinosi' genotypes were either homozygous 275/275 or heterozygous 263/275, the 'Romaneschi' typical artichokes were heterozygous 266/275; the 'Romaneschi-like' artichokes and the 'Out' group possessed also other alleles. The allele 261 was observed only in cultivated and wild cardoons: the cultivated cardoons were always heterozygous 261/296, the Italian and Greek samples of wild cardoons were all homozygous 261/261, while the Spanish wild cardoon samples were more variable, one being

Table 3 Number of alleles (na), observed (Ho) and Nei's expected (He) heterozygosity, and Shannon index (I) calculated for the *C. cardunculus* samples

Locus	na (AR/All)	H_O (AR/All)[a]	H_E (AR/All)[a]	I (AR/All)[a]
CsLib01	4/4	0.64/0.49	0.48/0.62	0.86/1.11
CsLib02	3/4	0.38/0.37	0.32/0.34	0.55/0.64
CsLib03	8/8	0.52/0.51	0.78/0.80	1.67/1.74
CsLib04	6/8	0.93/0.77	0.67/0.74	1.37/1.58
CsLib05	5/5	0.69/0.69	0.57/0.69	1.06/1.36
CsLib06	10/10	0.56/0.41	0.86/0.86	2.08/2.14
CsLib07	3/3	0.12/0.18	0.64/0.65	1.05/1.07
CsLib08	3/4	0.38/0.31	0.64/0.58	1.05/1.01
CsLib09	2/2	1.00/1.00	0.50/0.50	0.69/0.69
CsLib10	7/7	0.86/0.83	0.65/0.70	1.28/1.42
CsLib11	4/5	0.21/0.20	0.20/0.51	0.45/0.98
CsLib12	5/6	0.39/0.28	0.71/0.73	1.43/1.49
CsLib13	6/6	0.52/0.45	0.46/0.40	0.97/0.85
CsLib14	2/5	0.07/0.31	0.38/0.67	0.57/1.32
CsPal01	4/7	0.41/0.33	0.39/0.54	0.74/1.15
CsPal02	4/13	0.34/0.38	0.47/0.63	0.86/1.49
CsPal03	4/7	0.83/0.67	0.65/0.68	1.18/1.42
CsPal04	10/14	0.33/0.30	0.82/0.88	1.93/2.29
CsEST01	1/1	0.00/0.00	0.00/0.00	0.00/0.00
CsEST02	2/2	0.00/0.00	0.45/0.44	0.64/0.63
CsEST03	2/3	0.00/0.07	0.45/0.50	0.64/0.84
Cs1-SST	7/8	0.69/0.59	0.78/0.83	1.61/1.84
CsCiCaca01	1/1	0.00/0.00	0.00/0.00	0.00/0.00
CsCiCaca05	2/3	0.03/0.18	0.04/0.27	0.09/0.50
Mean	4.68/6.10	0.45/0.42	0.54/0.62	1.04/1.26
St. Dev.	2.38/3.11	0.30/0.25	0.20/0.17	0.49/0.48

[a] AR, artichoke samples; All, artichoke, wild and cultivated cardoon samples

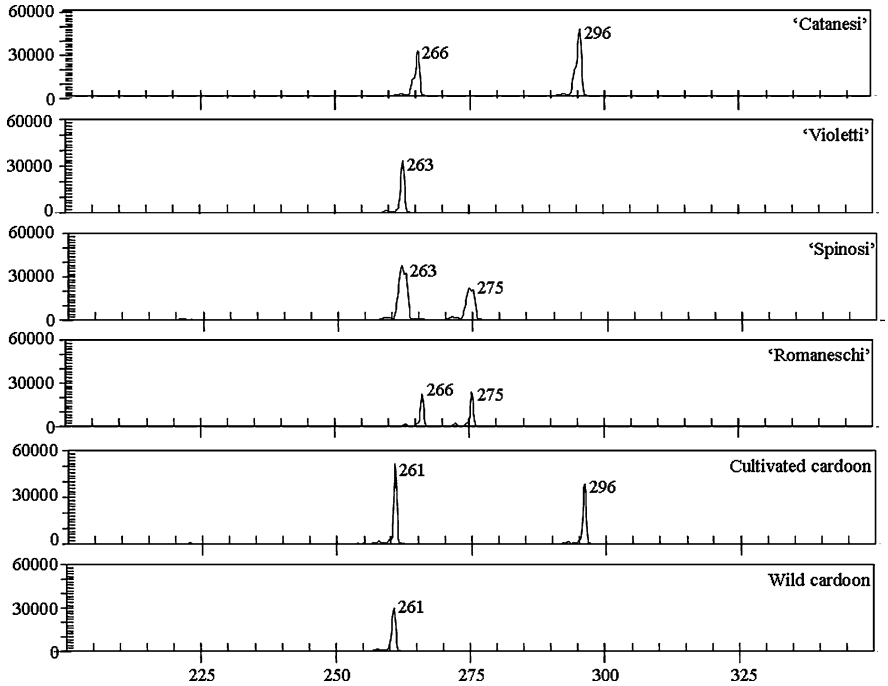

Fig. 1 Example of electropherograms showing different alleles at the locus Cs1-SST

heterozygous 261/266, the remaining being homozygous 266/266 or 296/296. Example electropherograms of these alleles are provided in Fig. 1.

At the locus CsPal02, the 'Catanesi' genotypes showed a unique allele combination (345/365), which was not found in any of the other artichoke genotypes. A similar situation was also observed at the locus CsPal03, where all the 'Catanesi' artichokes possessed the allele combination 339/343, not found elsewhere, with the sole exception of the variety 'Green Globe' ('Romaneschi-like') which does not belong to the 'Catanesi' group.

Genetic relationships among geographic and taxonomic groups

The number and percentage of shared alleles were calculated between taxonomic groups of *Cynara cardunculus* (Table 4). Wild cardoon samples were split into two groups based on the hypothesis that wild cardoon genepool is differentiated according to its geographical distribution (Wiklund 1992; Sonnante et al. 2007a). The two groups sharing the highest proportion of alleles were cultivated cardoons and Spanish wild cardoons (47%), whereas the lowest one was between artichoke and Spanish wild cardoons (26%). It is worth noticing that the percentage of shared alleles between wild cardoons of the two geographic ranges was as low as 30%.

To further go deep into the genetic relationships among the analysed *C. cardunculus* material, groups were formed considering taxonomic relationships, morpho-productive characters, or geographical origin. A first grouping was made for artichoke using the morpho-productive characters (Porceddu et al. 1976) which identify the four main varietal groups previously

Table 4 Total, shared alleles, and percentage of shared alleles between groups within *C. cardunculus*

	AR/CC	AR/WC-IG	AR/WC-S	CC/WC-IG	CC/WC-S	WC-IG/WC-S
Sum (shared + non shared alleles)	120	122	106	89	73	73
Shared alleles	52	39	28	37	34	22
Shared alleles (%)	43	32	26	42	47	30

AR, artichoke; CC, cultivated cardoon; WC-IG, Italian-Greek wild cardoon; WC-S, Spanish wild cardoon

Table 5 Nei's (1972) genetic distance among *C. cardunculus* groups

	CAT	VIO	SPI	ROM	OUT	CC	WC-IG	WC-S
CAT	–							
VIO	0.28	–						
SPI	0.21	0.26	–					
ROM	0.15	0.24	0.20	–				
OUT	0.43	0.30	0.33	0.24	–			
CC	0.42	0.48	0.45	0.47	0.50	–		
WC-IG	0.71	0.78	0.67	0.70	0.79	0.45	–	
WC-S	0.83	0.89	0.88	0.90	0.77	0.43	0.57	–

CAT, 'Catanesi'; VIO, 'Violetti'; SPI, 'Spinosi'; ROM, 'Romaneschi'; OUT, 'Out' group; CC, cultivated cardoon; WC-IG, Italian-Greek wild cardoon; WC-S, Spanish wild cardoon

described; the 'Romaneschi-like' were included in the 'Romaneschi' class, the remaining artichokes formed the 'Out' group. Wild and cultivated cardoons followed the division previously described. Therefore, Nei's (1972) genetic distances were calculated on 8 groupings: the above 5 artichoke groups, plus cultivated cardoons, Spanish wild cardoons, and Italian-Greek wild cardoons (Table 5). If the 'Out' group was not considered, genetic distances among the four main artichoke morpho-productive groups (Porceddu et al. 1976) were below 0.28, with 'Romaneschi' and 'Catanesi' showing the lowest distance (0.15). As regards the distance from the main artichoke groups, the Spanish wild cardoons displayed a value higher than 0.83, while the Italian/Greek wild cardoons showed a distance ranging from 0.67 to 0.78. The two wild cardoon genepools revealed a rather high distance value (0.57) from each other, consistently with the pattern outlined by allele sharing. The matrix of genetic distances was used to generate the dendrogram illustrated in Fig. 2, constituted by three main branches. The first one included all artichoke samples, which were quite differentiated from the rest of the *C. cardunculus* gene pool. Within the artichoke cluster, 'Romaneschi' and 'Catanesi' were grouped together, in the same sub-branch as 'Spinosi' and 'Violetti', while the 'Out' group was more differentiated. Another branch of the dendrogram included the Italian-Greek wild cardoons, and the third branch clustered the cultivated cardoons and the Spanish wild cardoons.

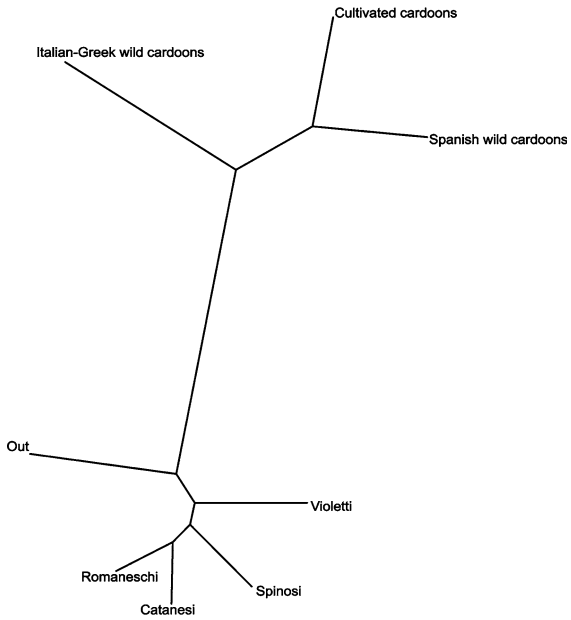

Fig. 2 UPGMA dendrogram based on Nei's (1972) genetic distance matrix calculated on the SSR loci analysed

Sequence analysis of microsatellite alleles

Some alleles at three microsatellite loci, namely CsPal04, Cs1-SST, and CsCICaca05, were sequenced. For each locus, samples belonging to the species *C. cardunculus* were chosen, together with at least one representative of each *Cynara* wild species.

CsPal04. Sequences were obtained for various alleles at the locus CsPal04 (Fig. 3, EMBL accession N. AM497830-AM497842). The sequence initially isolated from the artichoke gene *CSpal3* (De Paolis et al. 2008) contained two main TA stretches, one shorter, $(TA)_5$, at the beginning of the fragment, and one longer, $(TA)_{26}$. The $(TA)_5$ stretch showed length variation only in the *C. cornigera* sample, and a nucleotide substitution (A > T) with consequent elongation of the TA stretch in two samples, which did not alter the length of the region. The allelic variation observed was to be attributed mostly to length variation in the longer stretch $(TA)_{26}$. It is interesting to note that at the 3′ side of this stretch, a A > G substitution altered the initial position of the stretch in artichoke samples; moreover, in most wild species, the microsatellite was imperfect in the same region, for a A > G substitution. Other short deletions or insertions could be observed in the sequence of various samples, while the wild *C. syriaca* and *C. cornigera* displayed longer

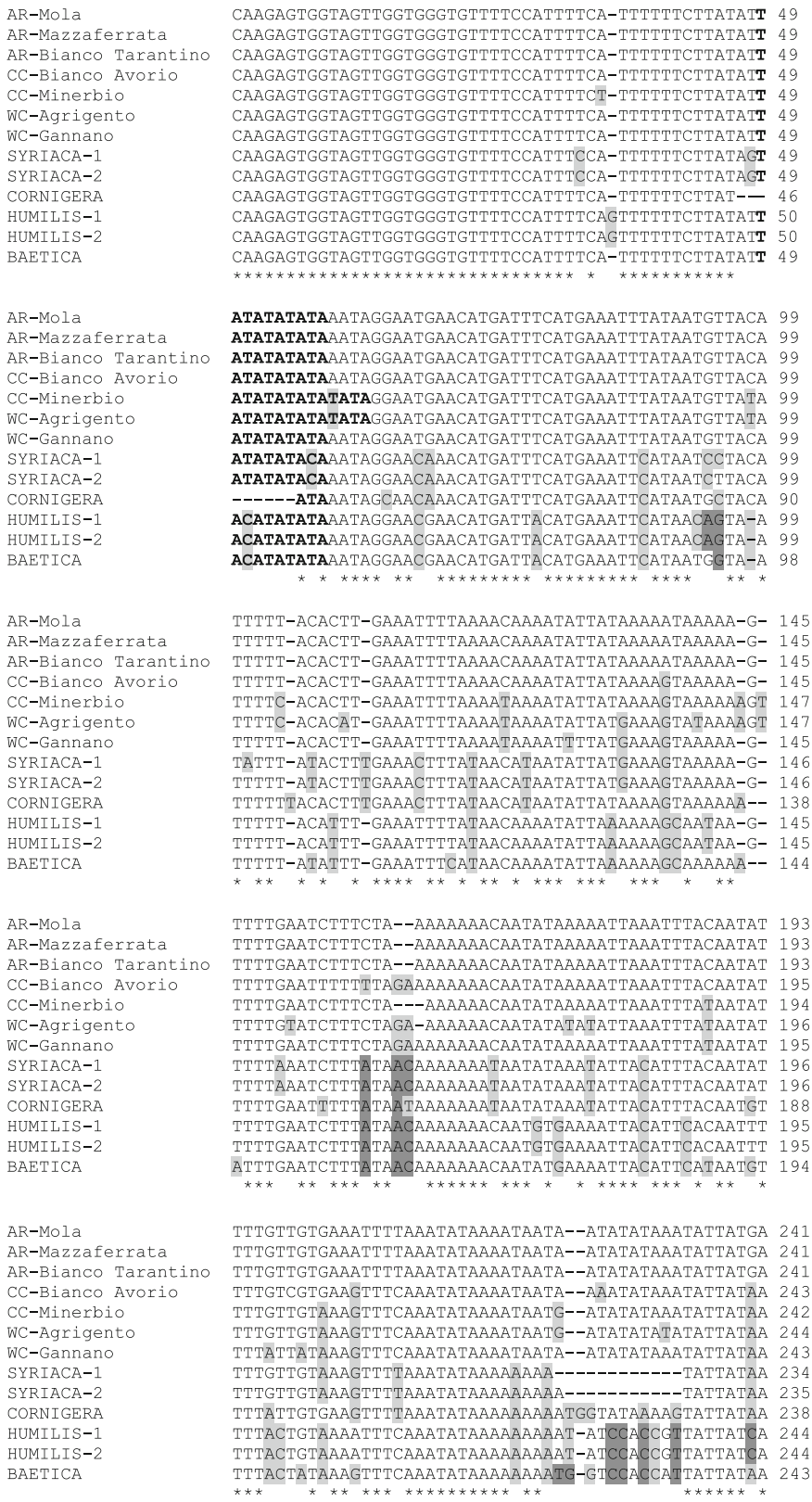

Fig. 3 Alignment of allele variants in *Cynara* samples at the locus CsPal04. Microsatellite repeats are in bold, polymorphisms are shaded. AR: artichoke; CC: cultivated cardoon; WC: wild cardoon

Fig. 3 continued

```
AR-Mola              AAATAAGTTGATGATCTCAAGTGTAT-ATATATATATATATATATATATA 290
AR-Mazzaferrata      AAATAAGTTGATGATCTCAAGTGTAT-ATATATATATATATATATATATA 290
AR-Bianco Tarantino  AAATAAGTTGATGATCTCAAGTGTAT-ATATATATATATATATATATATA 290
CC-Bianco Avorio     AAATAAGTTGATGATCTCAAGTATAT-ATATATATATATATATATATATA 292
CC-Minerbio          AAATAAGTTGATGATCTCAAGTATAT-ATATATATATA------------ 279
WC-Agrigento         AAATAAGTTGATGATATCTAGTATAT-ATATATATATATATATATATATA 293
WC-Gannano           AAATAAGTTGATGATCTCAAGTATAT-ATATA------------------ 274
SYRIACA-1            AAATAAGTTGATGATCTCAAGTATATGATATATATATA------------ 272
SYRIACA-2            AAATAAGTTGATGATCTCAAGTATATG-TATATATATA------------ 272
CORNIGERA            AAATAAGTTGATGATCTCAATATATG---TATATATA------------ 273
HUMILIS-1            AAATAAGTTGATAATCTCAAGTATATG-TATATATATA------------ 281
HUMILIS-2            AAATAAGTTGATAATCTCAAGTATATG-TATATATATA------------ 281
BAETICA              AAATAAGTTGATGATCTCAAGTATATA----------------------- 270
                     ************ ** ** * * ***

AR-Mola              TATATATATATATATATATATATATAATTGAGATAGGTAGGGCATATAAT 340
AR-Mazzaferrata      TATATATATATATATATATATA----ATTGAGATAGGTAGGGCATATAAT 336
AR-Bianco Tarantino  TATATA--------------------ATTGAGATAGGTAGGGCATATAAT 320
CC-Bianco Avorio     TA-----------------------ATTGAGATAGGTAGGGCATATAAT 318
CC-Minerbio          -------------------------ATTGAGATAGGTAGGGCATATAAT 303
WC-Agrigento         TATATATATATATATATA--------ATTGAGATAGGTAGGGCATATAAT 335
WC-Gannano           -------------------------ATTGAACTAGGTAGGGCATATAAT 298
SYRIACA-1            -------------------------ATTGAGATAGGTAGGGCA------ 290
SYRIACA-2            -------------------------ATTGAGATAGGTAGGGCA------ 290
CORNIGERA            -------------------------ATTGAGATA--------------- 282
HUMILIS-1            -------------------------ATTGAGATAGGTAGGGCATATAAT 305
HUMILIS-2            -------------------------ATTGAGATAGGTAGGGCATATAAT 305
BAETICA              -------------------------ATTGAGATAGGTAGGGCATATAAT 294
                                               *****  **

AR-Mola              AAGAGATATATAGTTGGAATTAGTGGTCAAAGTTCAAGAATTCGAGTA-T 389
AR-Mazzaferrata      AAGAGATATATAGTTGGAATTAGTGGTCAAAGTTCAAGAATTCGAGTA-T 385
AR-Bianco Tarantino  AAGAGATATATAGTTGGAATTAGTGGTCAAAGTTCAAGAATTCGAGTA-T 369
CC-Bianco Avorio     AAGAGATATATAGTTGGTATTAGTGGTCAAAGTTCAAGAATTCGAGTA-T 367
CC-Minerbio          AAGAGATATATAGTTGGAATTAGTGGTCAAAGTTCAAGAATTCGAGTA-T 352
WC-Agrigento         AAGAGATATATAGTTGGAATTAGTGGTCAAAGTTCAAGAATTCGAGTAAT 385
WC-Gannano           AAGAGATATATAATTGGAATTAGTGGTCAAAGTTCAAGAATTCGAGTA-T 347
SYRIACA-1            --------TATAGTTGGAATTAGTGGTCAAAGTCCAAGAATTAGAGTA-T 331
SYRIACA-2            --------TATAGTTGGAATTAGTGGTCAAAGTCCAAGAATTAGAGTA-T 331
CORNIGERA            --------TATAGTTGGAATTAGTGGTCAAAGTCCAAGAATTCGAGTA-T 323
HUMILIS-1            AAGAGATATATAGCTGGAATTAGTGGTCAAAGTCCAAGAATTCGAGTA-T 354
HUMILIS-2            AAGAGATATATAGCTGGAATTAGTGGTCAAAGTCCAAGAATTCGAGTA-T 354
BAETICA              AAGAGATATATAGTTGGAATTAGTGGTCAAAGTCCAAGAATTCGAGTA-T 343
                             ****  *** *************** ******** ***** *

AR-Mola              GTGATTGGTTGGCTGCATAAGCACC 414
AR-Mazzaferrata      GTGATTGGTTGGCTGCATAAGCACC 410
AR-Bianco Tarantino  GTGATTGGTTGGCTGCATAAGCACC 394
CC-Bianco Avorio     GTGATTGGTTGGCTGCATAAGCACC 392
CC-Minerbio          GTGATGGGTGGGCTGCATAAGCACC 377
WC-Agrigento         GTGATTGGTTGGCTGCATAAGCACC 410
WC-Gannano           GTGATTGGTTGGCTGCATAAGCACC 372
SYRIACA-1            GTGATTGGTTGGCTGCATAAGCACC 356
SYRIACA-2            GTGATTGGTTGGCTGCATAAGCACC 356
CORNIGERA            GTGATTGGTTGGCTGCATAAGCACC 348
HUMILIS-1            GTGATTGGTTGGCTGCATAAGCACC 379
HUMILIS-2            GTGATTGGTTGGCTGCATAAGCACC 379
BAETICA              GTGATTGGTTGGCTGCATAAGCACC 368
                     ***** *** ***************
```

deletions in other regions of the sequence. A lot of single nucleotide polymorphisms (SNPs) were observed particularly in the wild *Cynara* species compared to *C. cardunculus*. However, four of these SNPs were useful to distinguish artichoke samples from one side and wild and cultivated cardoon material from the other side (e.g.: A > G @ 138 bp, A > G @ 205 bp, G > A @ 240 bp, G > A @ 266 bp, referring to 'Mola' sequence).

Cs1-SST. Sequencing of Cs1-SST alleles revealed that the size scored after fragment separation did not correspond to the number of bases determined by sequencing for the same allele. Precisely, each sequence showed a 6 bp reduction in size as compared to fragment analysis. Two main repeats were found in this sequence (Fig. 4, EMBL accession N. AM497843-AM497853). The first one was a long region composed of a GAT microsatellite interleaved with a minisatellite

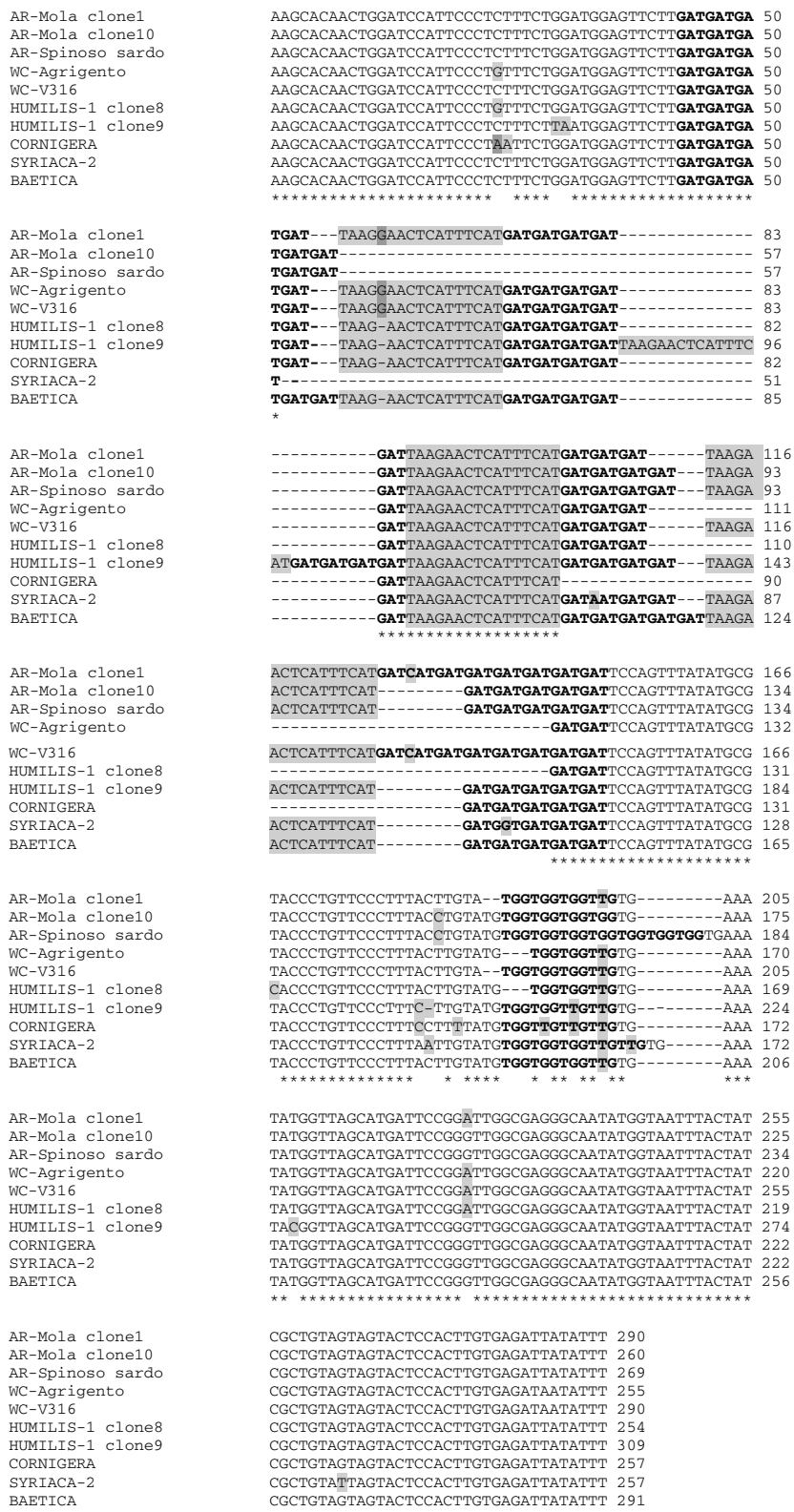

Fig. 4 Alignment of allele variants in *Cynara* samples at the locus Cs1-SST. Microsatellite repeats are in bold, minisatellite and polymorphisms are shaded. AR: artichoke; WC: wild cardoon

of 16 bp (17 bp only in three samples in the first stretch). This minisatellite region, always embedded in GAT repeats of variable length, was repeated from two to four times, greatly contributing to allelic length variation. The second SSR was a TGG repeat, sometimes becoming imperfect (TTG). A few SNPs were also observed in this sequence, and length variation was due to variation in micro- and minisatellite number of repeats.

CsCiCaca01 and CsCiCaca05. As for microsatellites isolated from *Cirsium*, a total of 5 primer pairs were initially chosen, which had proven to amplify simple SSRs from this thistle (Jump et al. 2002). In artichoke, three of these primer combinations produced a multiband amplificate when run on agarose gel, whereas Caca01 and Caca05 primers amplified a single band. The corresponding artichoke loci were named CsCiCaca01 and CsCiCaca05 respectively. The former revealed to be monomorphic within the *C. cardunculus* complex, and showed two additional alleles in *C. humilis*; on the other hand, the latter was polymorphic also within the cultigroup. Sequencing of these regions in artichoke revealed that a reduction of the (AC) repeat number occurred both for CsCiCaca01 (from 10 in *Cirsium* to 3 in *C. cardunculus,* EMBL accession N. AM497869, AM497870) and CsCiCaca05 (from 12 in *Cirsium* to 3 or 4 in *Cynara*, Fig. 5). Moreover, for the CsCiCaca01 locus, some nucleotide substitutions and deletions, among which two 8 bp deletions, were observed in artichoke compared to *Cirsium*. For CsCiCaca05 locus, besides variations in the main microsatellite stretch, nucleotide

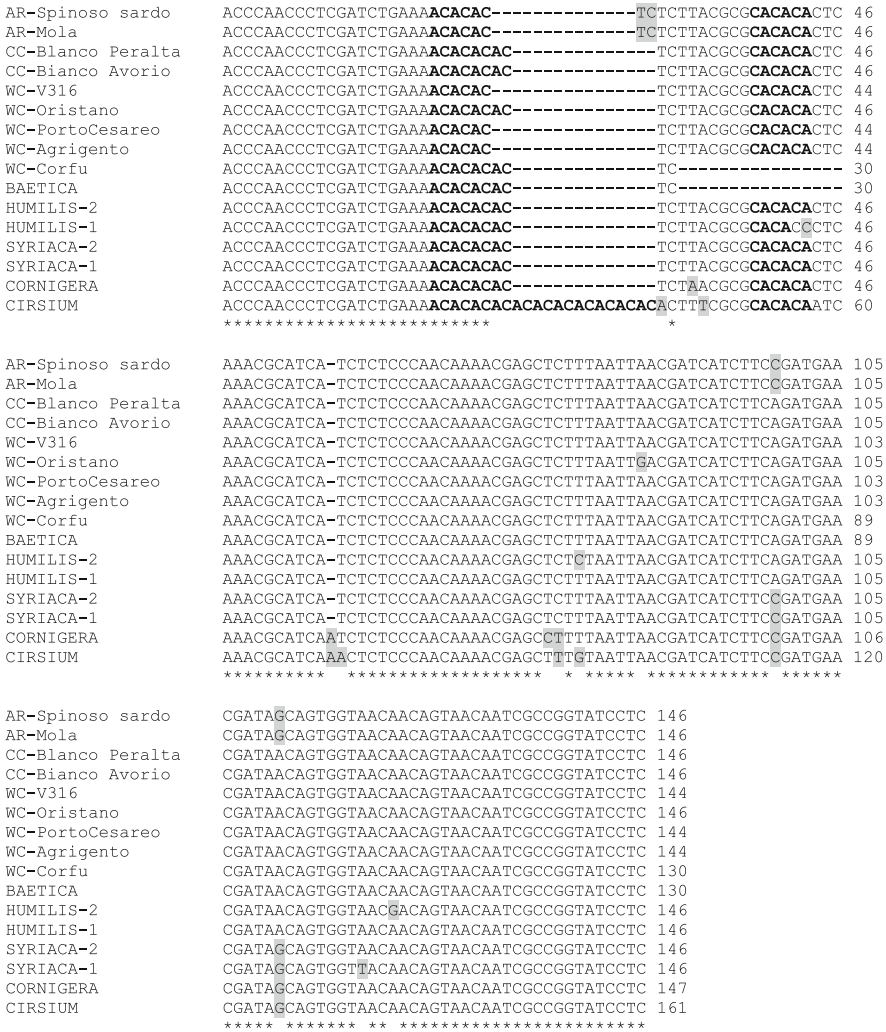

Fig. 5 Alignment of allele variants in *Cynara* samples and *Cirsium* at the locus CsCiCaca05. Microsatellite repeats are in bold, polymorphisms are shaded. Artichoke samples 'Romanesco' and 'Violetto di Toscana' showed the same sequenze as the other artichokes. AR: artichoke; CC: cultivated cardoon; WC: wild cardoon

substitutions were observed between *Cirsium* and *Cynara*, and some polymorphisms were also found within *Cynara* (Fig. 5, EMBL accession N. AM497854-AM497868, AM497871-AM497874). As for *C. cardunculus*, all the artichoke samples, 'Mola' and 'Spinoso Sardo' (plus 'Romanesco' and 'Violetto di Toscana', not shown) possessed an additional TC repeat after the AC stretch and two SNPs (A > C @ 99 bp, and A > G @ 111 bp in 'Mola' sequence), compared to cultivated and wild cardoon (shaded in Fig. 5). Some wild cardoons and the sample of *C. baetica* showed a smaller allele of 130 bp. The sequence of this allele revealed the deletion of a 16 bp sequence not belonging to the main SSR stretch.

Discussion

The present study reports on the detection, by means of different approaches, of 24 microsatellite markers in *Cynara*, one of which was monomorphic, through different approaches. Some alleles at three different loci were also cloned and sequenced in order to verify if length variation was attributable to microsatellite core variation only, or if other mutations could have occurred. Sequencing revealed in addition to SSR polymorphism a number of SNPs, some of which of diagnostic interest.

At the locus Cs-1SST we also observed lack of correspondence between the fragment size obtained by fragment separation and the number of bases determined by sequencing for the same allele. In two additional cases (CsLib01 and CsEST01) we also found this incongruence and scored fragments of higher size than the original sequences from which SSR were isolated. This lack of correspondence could be possibly due to the presence of the fluorescent dye which can bias length determination in fragment analysis. Macaulay et al. (2001) found inconsistencies when assessing SSR fragment size on an automated sequencer or on autoradiographs derived from sequencing gels. We decided to report the fragment size read at the CEQ 8800 instrument, since this is what we could observe in the experimental conditions we used.

The high number of alleles per locus in *C. cardunculus*, combined with the average observed heterozygosity value, suggests that considerable polymorphism is present at these microsatellite loci in the artichoke primary gene pool. It is interesting to notice that artichoke, a vegetatively propagated crop, retains a level of observed heterozygosity higher, although slightly, than the average of the whole *C. cardunculus*. It has been suggested that, like in other species, farmers might have selected propagation material on the basis of the heterotic traits associated to high level of heterozygosity (Sonnante et al. 2007b). In cassava, farmers use to introduce into clonal fields also few seed propagated plants that show heterotic traits (Pujol et al. 2005). A similar behaviour for artichoke might not be excluded during the domestication and differentiation of this crop. It has been observed that, even nowadays, in some Mediterranean areas under traditional agricultural systems, farmers use to cultivate seed propagated artichokes not belonging to modern seed propagated cultivars and maintain vegetatively only those plants possessing the desired traits (Pignone, personal observation; Laghetti et al. in preparation).

Even though the developed SSRs do not allow the unique identification of single varieties or landraces, some allelic combinations appear to exclusively relate to specific varietal groups. For instance, the locus Cs1-SST is informative to distinguish among 'Catanesi', 'Violetti', 'Spinosi', and 'Romaneschi' which possess, respectively, specific genotypes at this locus. This information can be extremely useful at least to identify the varietal group in derived products, such as frozen or canned artichokes, because the production process eliminates the parts possessing distinctive morphological characters. Furthermore, the possibility to get alleles specific to varietal groups, allows the bridging of different cross combinations in genetic map constructions. The observation that in artichoke some loci show variation in relation to morpho-productive groups, might indicate that these loci are associated to some Mendelian or QTL traits particularly subjected to man selective pressure during the domestication/differentiation of this crop.

Moreover, some specific alleles identified in wild and/or cultivated cardoon can also be used in analysing crosses between artichoke and cultivated or wild cardoon. This kind of information might prove useful in breeding programmes aimed at producing hybrid seed propagated artichokes, which presently appears as the future of artichoke breeding (Basnizki and Zohary 1994; Calabrese and Bianco 2000) especially referring to the incorporation of

biotic or abiotic stress resistance traits, presumably present in the wild cardoon gene pool. In this respect, the present analysis has also allowed the identification of 6 SNPs (four at the locus CsPal04, and two at the locus CsCiCaca05) and a CT insertion distinguishing artichoke on one side and cardoons on the other, which could be used for the same purpose.

Congruently with previous observations (Wiklund 1992; Sonnante et al. 2007a, b), wild cardoons from the Western and the Eastern Mediterranean range were quite differentiated from each other, and therefore were considered as separate groups. On the other hand, wild Spanish cardoons shared a higher number of alleles with cultivated cardoons. The relationships among groups here presented supports previous observations based on ITS and ETS sequence analysis (Sonnante et al. 2007a) and the hypothesis that artichokes were domesticated starting from the Eastern Mediterranean gene pool, while cultivated cardoons originated from the Western one (Sonnante et al. 2007b).

A previous work (Whitton et al. 1997) suggested that the transferability of microsatellites from *Helianthus* to other Asteraceae might not prove successful; the results, though, could be justified by the fact that a wide range of species of the Asteraceae family was used in that analysis. In the present study, two microsatellites from *Cirsium* were successfully transferred to *Cynara*; in this case the success could be attributed to the close affinity of *Cirsium* to the genus *Cynara*, since both genera belong to the Cardueae tribe. However, when sequences obtained from artichoke were compared with those of *Cirsium*, we scored a reduction of the microsatellite core repeats at both these loci. On the other hand, at the locus CsPal04, an expansion of the longer TA stretch was noticed in *C. cardunculus* in comparison to the wild *Cynara* species (Fig. 3). In other studies, microsatellite expansions have been observed, when comparing related species, the expansion being based on a low initial number of repeats (Kahru et al. 2000). It has also been described that perfect microsatellites are more easily conserved between species (Kutil and Williams 2001), while compound repeat motifs are less persistent, since they represent the last stages of microsatellite death and degradation (Taylor et al. 1999). Directional evolution of microsatellites is controversial though, and SSR evolution appears in balance between growth by slippage and degradation by mutation accumulation (Zhu et al. 2000).

In addition, it was possible to transfer most of the isolated microsatellites to other wild species of the genus *Cynara*. For some loci, these species displayed specific alleles. However, fragment length alone was not enough to express variation and relationships between different species. This was the case, for example, for the locus CsPal04, where a lot of nucleotide substitutions were observed when sequencing fragments from the different species (Fig. 3). A high number of these substitutions were related to geographical distribution of the species: in fact, 14 were shared between Western wild *Cynara* species (*C. humilis* and *C. baetica*), whereas 6 were shared between the Eastern distributed *C. cornigera* and *C. syriaca*. This pattern is consistent with the hypothesis that the whole genus *Cynara* is quite recent, having differentiated into the nowadays species during the last 20 millennia (Sonnante et al. 2007a, b). Probably, the distinct alleles were already present in the common progenitor and were disseminated in different proportions in the derived species distributed in the Western and Eastern range.

As stated before, the patterns of microsatellite evolution are not fully understood yet (Karhu et al. 2000). Moreover, indels and base substitutions, which are known to influence subsequent evolutionary rates, can be common even between closely related species (van Treuren et al. 1997), as we could observe in the present study. For these reasons, comparison of allele distribution among species, although accomplished in several taxa (e.g. *Olea*, Rallo et al. 2003; *Cicer*, Sethy et al. 2006), can be inappropriate in some cases, unless a low evolutionary rate allows the maintenance of the microsatellite structure also in other species related to that from which SSRs were isolated (Scott et al. 2003). As a consequence, we decided not to include the wild *Cynara* species in the analyses of genetic distances based on allele frequencies. Moreover, some point mutations probably occurred in the primer regions thus accounting for the observed lack of amplification at few loci in some wild samples.

In conclusion, a set of 23 polymorphic microsatellite markers was identified and characterized from the artichoke genome. Variation was scored within artichoke, between this crop and wild and cultivated cardoon, and between *C. cardunculus* and other wild

Cynara allies. Some allelic combinations were able to identify artichoke varietal types, and some alleles were unique to specific taxonomic groups. This makes these markers potentially useful in product traceability and in saturating genetic maps. The transferability to other *Cynara* wild allies was quite good. Sequencing alleles at three loci showed that, apart from microsatellite length variation, point mutations were quite abundant especially when comparing *C. cardunculus* to the other *Cynara* species. Some SNPs distinguished artichoke from cultivated and wild cardoon, while some other SNPs were distributed according to the geographic range of *Cynara* wild species.

Acknowledgements The authors are grateful to the colleagues who donated seeds, and to Ms. Anita Morgese for DNA sequencing. This study was funded by: CNR, Project 'AG.P02.004, Banca del DNA vegetale'; CRA-MiPAF, Project 'FAO-RGV: Implementazione Nazionale del Trattato Internazionale FAO Risorse Genetiche Vegetali'; and EU Project AGRI GEN RES 063, CYNARES.

References

Basnizki J, Zohary D (1994) Breeding of seed planted artichoke. Plant Breed Rev 12:253–269

Bianco VV (2005) Present situation and future potential of artichoke in the Mediterranean basin. Acta Hortic 681: 39–55

Calabrese N, Bianco VV (2000) Effect of gibberellic acid on yield and quality of seed grown artichoke (*Cynara cardunculus* L. var. *scolymus* (L.) Fiori). Acta Hortic 514: 25–32

De Paolis A, Pignone D, Morgese A, Sonnante G (2008) Characterization and differential expression analysis of artichoke phenylalanine ammonia-lyase coding sequences. Physiol Plant 132:33–43

Elia A, Miccolis V (1996) Relationships among 104 artichoke (*Cynara scolymus* L.) accessions using cluster analysis. Adv Hortic Sci 10:158–162

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

Grattapaglia D, Sederoff VR (1994) Genetic-linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross mapping strategy and RAPD markers. Genetics 137:1121–1137

Hammer K, Knüpffer H, Laghetti G, Perrino P (1999) Seeds from the past. A catalogue of germplasm in Central and North Italy. CNR, Bari

Jain S, Jain RK, McCouch SR (2004) Genetic analysis of Indian aromatic and quality rice (*Oryza sativa* L.) germplasm using panels of fluorescently-labeled microsatellite markers. Theor Appl Genet 109:965–977

Jump AS, Dawson DA, James CM, Woodward FI, Burke T (2002) Isolation of polymorphic microsatellites in the stemless thistle (*Cirsium acaule*) and their utility in other *Cirsium* species. Mol Ecol Notes 2:589

Kahru A, Dieterich JH, Savolainen O (2000) Rapid expansion of microsatellite sequences in pines. Mol Biol Evol 17:259–265

Kutil BL, Williams CG (2001) Triplet-repeat microsatellites shared among hard and soft pines. Am Genet Assoc 92:327–332

Lacape JM, Dessauw D, Rajab M, Noyer JL, Hau B (2007) Microsatellite diversity in tetraploid *Gossypium* germplasm: assembling a highly informative genotyping set of cotton SSRs. Mol Breed 19:45–58

Lanteri S, Acquadro A, Comino C, Mauro R, Mauromicale G, Portis E (2006) A first linkage map of globe artichoke (*Cynara cardunculus* var. *scolymus* L.) based on AFLP, S-SAP, M-AFLP and microsatellite markers. Theor Appl Genet 112:347–357

Lewontin RC (1972) The apportionment of human diversity. Evol Biol 6:381–398

López Anido FS, Firpo IT, García SM, Cointry EL (1998) Estimation of genetic parameters for yield traits in globe artichoke (*Cynara scolymus* L.). Euphytica 103:61–66

Macaulay M, Ramsay L, Powell W, Waugh R (2001) A representative, highly informative 'genotyping set' of barley SSRs. Theor Appl Genet 102:801–809

Matsuoka Y, Mitchell SE, Kresovich S, Goodman M, Doebley J (2002) Microsatellites in *Zea* – variability, patterns of mutations, and use for evolutionary studies. Theor Appl Genet 104:436–450

Nei M (1972) Genetic distance between populations. Am Nat 106:283–292

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590

Pignone D, Sonnante G (2004) Wild artichokes of south Italy: did the story begin here? Genet Resour Crop Evol 51:577–580

Porceddu E, Dellacecca V, Bianco VV (1976) Classificazione numerica di cultivar di carciofo. In: Proceed. of the 2nd international congress on Artichoke, Bari 1973. Minerva Medica, Turin, Italy, pp 1105–1119

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). Ecol Lett 8:138–147

Rallo P, Tenzer I, Gessler C, Baldoni L, Dorado G, Martìn A (2003) Transferability of olive microsatellite loci across the genus *Olea*. Theor Appl Genet 107:940–946

Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier MH, Leroy P, Ganal MW (1998) A microsatellite map of wheat. Genetics 149:2007–2023

Rohrer JR, Ahmad R, Southwick SM, Potter D (2004) Microsatellite analysis of relationships among North American plums (*Prunus* sect. *Prunocerasus*, Rosaceae). Plant Syst Evol 244:69–75

Rottenberg A, Zohary D (1996) The wild ancestry of the cultivated artichoke. Genet Resour Crop Evol 43:53–58

Rottenberg A, Zohary D (2005) Wild genetic resources of cultivated artichoke. Acta Hortic 681:307–311

Rottenberg A, Zohary D, Nevo E (1996) Isozyme relationships between cultivated artichoke and the wild relatives. Genet Resour Crop Evol 43:59–62

Sambrook J, Russell RW (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring, New York

Scott LJ, Shepherd M, Henry RJ (2003) Characterization of highly conserved microsatellite loci in *Araucaria cunninghamii* and related species. Plant Syst Evol 236:115–123

Sethy NK, Chouldhary S, Shokeen B, Bhatia S (2006) Identification of microsatellite markers from *Cicer reticulatum*: molecular variation and phylogenetic analysis. Theor Appl Genet 112:347–357

Sonnante G, De Paolis A, Lattanzio V, Perrino P (2002) Genetic variation in wild and cultivated artichoke revealed by RAPD markers. Genet Resour Crop Evol 49:247–252

Sonnante G, De Paolis A, Pignone D (2004a) Relationships among artichoke cultivars and some related wild taxa based on AFLP markers. Plant Genet Resour 1:125–133

Sonnante G, Ippedico M, De Paolis A (2004b) Microsatellite and AFLP markers in a world artichoke collection. Acta Hortic 660:61–68

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

Sonnante G, Pignone D, Hammer K (2007b) The domestication of artichoke and cardoon: from Roman times to genomics age. Ann Bot 100:1095–1100

Taylor JS, Durkin JMH, Breden F (1999) The death of a microsatellite: a phylogenetic perspective on microsatellite interruptions. Mol Biol Evol 16:567–572

This P, Jung A, Boccacci P, Borrego J, Botta R, Costantini L, Crespan M, Dangl GS, Eisenheld C, Ferreira-Monteiro F, Grando S, Ibáñez J, Lacombe T, Laucou V, Magalhães R, Meredith CP, Milani M, Peterlunger E, Regner F, Zulini L, Maul E (2004) Development of a standard set of microsatellite reference alleles for identification of grape cultivars. Theor Appl Genet 109:1448–1458

Van Treuren R, Kuittinen H, Kärkkäinen K, Baena-Gonzales E, Savolainen A (1997) Evolution of microsatellites in *Arabis lyrata*, outcrossing relative of *Arabidopsis thaliana*. Mol Biol Evol 14:220–229

Viruel MA, Escribano P, Barbieri M, Ferri M, Hormaza JI (2005) Fingerprinting, embryo type and geographic differentiation in mango (*Mangifera indica* L., Anacardiaceae) with microsatellites. Mol Breed 15:383–393

Wiklund A (1992) The genus *Cynara* L. (Asteraceae-Cardueae). Bot J Linn Soc 109:75–123

Whitton J, Rieseberg LH, Ungerer MC (1997) Microsatellite loci are not conserved across Asteraceae. Mol Biol Evol 14:202–209

Yeh FC, Yang RC, Boyle TBJ, Ye ZH, Mao JX (1997) POPGENE, the user-friendly shareware for population genetic analysis. Molecular Biology and Biotechnology Centre, University of Alberta, Canada (program available from: http://www.ualberta.ca/~fyeh/)

Zhu Y, Queller DC, Strassmann JE (2000) A phylogenetic perspective on sequence evolution in microsatellite loci. J Mol Evol 50:324–338

Zohary D, Basnizki J (1975) The cultivated artichoke *Cynara scolymus*. Its probable wild ancestors. Econ Bot 29: 233–235