Genetic mapping and annotation of genomic microsatellites isolated from globe artichoke

A. Acquadro ☉S. Lanteri ☉D. Scaglione ☉P. Arens ☉B. Vosman ☉E. Portis

Genetic mapping and annotation of genomic microsatellites isolated from globe artichoke

A. Acquadro ☉S. Lanteri ☉D. Scaglione ☉E. Portis*

Di.Va.P.R.A. Plant Genetics and Breeding, University of Turin, via L. da Vinci 44, I-10095 Grugliasco (Turin), Italy

* corresponding author (e-mail: ezio.portis@unito.it Fax: +390112368807)

P. Arens ☉B. Vosman

Plant Research International, P.O.Box 16, 6700 AA Wageningen (NL)

Abstract

Cynara cardunculus includes the three taxa, the globe artichoke (subsp. *scolymus* L. Hegi), the cultivated cardoon (var. *altilis*) and their progenitor, the wild cardoon (var. *sylvestris*). Globe artichoke is an important component of the Mediterranean rural economy, but its improvement through breeding has been rather limited and its genome organization remains largely unexplored. We report here the isolation of 61 new microsatellite loci which amplified a total of 208 alleles in a panel of 22 *C. cardunculus* genotypes. Of these, 51 were informative for linkage analysis, and 39 were used to increase marker density in the available globe artichoke genetic maps. Sequence analysis of the 22 loci associated with genes showed that nine are located within coding sequence, with the repetitive domain probably being involved in DNA binding or in protein-protein interactions. The expression of the genes associated with nine of the 22 microsatellite loci was demonstrated by RT PCR.

Introduction

Cynara cardunculus L. (*Asteraceae*, 2n=2x=34) contains the three taxa: subsp. *scolymus* L. Hegi (the cultivated globe artichoke), var. *altilis* DC. (the cultivated cardoon) and var. *sylvestris* (Lamk) Fiori (the wild cardoon). Globe artichoke is an allogamous plant, with an estimated genome size of 1078Mbp (Marie and Brown 1993). The immature inflorescences (capitula or heads) provide the edible part of the plant, and are used fresh, canned or frozen for the preparation of a variety of dishes; its leaves have been exploited as hepatoprotectants, and either choleretic or diuretic agents in traditional medicine since Ancient Roman times. In modern times, leaf extracts have been identified as containing cellular protectants against oxidative damage, HIV integrase inhibitors, and bile-expelling and lipid-lowering agents (Gebhardt 1997, 1998; Kraft 1997; Llorach et al. 2002; McDougall et al. 1998; Wang et al. 2003), while roots and seeds have been used to extract inulin (Raccuia and Melilli 2004), with high degree of polymerization, and oil (Maccarone et al. 1999, Raccuia and Melilli 2007). The crop is grown across the Middle East, North Africa, South America, China, the USA, and particularly in the Mediterranean region, where it has a significant impact on the rural economy. Italy is the leading global producer (http://faostat.fao.org/). Despite its economic, pharmacological and nutritional value, its improvement through breeding has been rather limited, while, unlike other crop species belonging to the same botanical family (such as sunflower, lettuce and chicory), its genome organization remains largely unexplored.

The first molecular maps of globe artichoke have only recently been published (Lanteri et al. 2006). These were largely based on dominant DNA fingerprinting platforms, although a small number of microsatellite (SSR) markers was included (Acquadro et al. 2003; Acquadro et al. 2005a,b). Although SSRs are widely favoured as a marker platform for genetic mapping and biodiversity studies on account of their allelic variability, it has become clear that some can also act as regulatory elements (Iglesias et al. 2004; Martin et al. 2005). The 5' untranslated region of many monocot and dicot genes contains highly conserved (both with respect to motif and genomic position) SSR sequences (Guo and Moose 2003; Yang et al. 1999), and this has been taken to imply that SSRs can also play a role in gene regulation. The region upstream of the transcription initiation sites in both *Arabidopsis thaliana* and rice genes has been shown to be characterized by a gradient of pyrimidine-rich SSR density (Zhang et al. 2006). SSRs also occur within exons, and their translation products (typically G_x, N_x or P_x) may provide a domain for DNA binding or protein-protein interaction. Such repetitive polypeptide stretches are known to be involved in the activation/de-activation of transcription (Berger et al. 2001; Gerber et al. 1994; Kolaczkowska et al. 2002; Perutz et al. 1994; Toth et al. 2000), and allelic variants in such genic SSRs have been implicated as the genetic determinants of a number of human diseases (Leroy et al. 2000).

Here we report the development of a set of globe artichoke SSR assays, extracted from enriched genomic libraries. We describe their informativeness for diversity analysis and taxonomic discrimination, their genetic map location as well as their annotation and Gene Ontology (GO) categorisation.

Materials and methods

Plant materials and genomic DNA isolation

DNA was extracted from young globe artichoke leaves following Lanteri et al. (2001). The primers developed were applied to DNA of i) the parents of three established mapping populations, specifically the two diverse globe artichoke genotypes [‘Romanesco C3’ (C3) and ‘Spinoso di Palermo’ (Sp-9A)], one cultivated cardoon (A41) and one wild cardoon (‘Creta 4’) genotype; ii) four F1 individuals from each of the segregating populations C3 x Sp-9A, C3 x A41 and C3 x Creta-4; and iii) six globe artichoke genotypes, demonstrated to be representative of Mediterranean Basin germplasm (Lanteri et al. 2004). Linkage analysis was performed on 94 C3 x SP-9A progeny. Full genotype details are reported in Table 1.

Enriched libraries

SSR-containing sequences were isolated from ten enriched small-insert genomic libraries following van de Wiel et al. (1999), with minor modifications. *Alu*I, *Rsa*I or *Hae*III (5U) was used to digest 500ng genomic DNA in the presence of 50pmol of both 5’-GTTTCAGATCTGGCTCATCGC-3’ (Ada +) and 3’-ACACCAAAGTCTAGACCGAGTAGCG-5’ (Ada -), in a 50µl reaction containing restriction-ligation buffer [10mM Tris-HCl pH 7.5, 10mM MgAc, 50mM KAc, 5mM DTT], 1mM ATP and 5U T4 DNA ligase. Restriction fragments in the size range 300-1000bp were selected by gel electrophoresis extraction and purified from agarose using the NucleoSpin Extract II kit (Machery-Nagel). These were then amplified using 1µl of the restricted ligated DNA as template in a 20µl PCR containing 50pmol adapter primer (Ada +), 1.5mM $MgCl_2$, 1mM dNTP and 1U Taq polymerase (Invitrogen) in the manufacturer's buffer. The amplification programme was 94°C / 120s, followed by 25 cycles of 94°C / 30s, 50°C / 30s and 72°C / 120s and ending with a 10min incubation at 72°C. The size fractioned PCR product was denatured and hybridized to a Nylon+ (Amersham) filter carrying 1.5µg single-stranded, UV bound $(GT)_{12}$, $(GA)_{12}$, $(TCT)_{10}$, $(TGT)_9$, $(GAG)_8$, $(GTG)_8$, $(TGA)_9$, $(AGT)_{10}$, $(GCT)_8$, and $(GCC)_7$ for 48h at 37°C in 5x SSC, 50mM Na-phosphate (pH7), 7% w/v SDS and 50% v/v formamide. The filters were consecutively washed in stepwise reducing concentrations of SSC (1.5X, 0.5X, 0.2X, 0X w/v) and 1% w/v SDS at 62°C. The DNA dissolved in each wash fraction was precipitated by an overnight incubation in 20µg glycogen, 0.8 M LiCl and 600µl isopropanol and then resuspended in 0.1x TE. This DNA was re-amplified in a 20µl PCR, as above. Each PCR product was ligated into pGEM-T (Promega) and introduced into *E. coli* JM109 (Promega). Insert-containing clones were bound to Hybond N+ (Amersham) membranes, which were hybridized with a mixture of the appropriate ^{32}P end-labeled oligonucleotides to select SSR containing clones, which were sequenced by Greenomics™ (Wageningen, NL). From these sequences, primer pairs were designed by Primer 3.0 (Rozen and Skaletsky 2000), http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi), adopting default parameter settings. A tailing primer strategy was used, as described by Oetting et al. (1995). The newly developed SSR markers were identified by a number, prefixed by CELMS (*Cynara* Enriched Library MicroSatellite) (Table 2).

SSR genotyping

The SSRs were tested for their informativeness on the 22 genotypes reported in Table 1. PCRs were performed and the resulting products analysed as reported by Acquadro et al. (2005b). Briefly, amplification products were mixed with 5-50µl of formamide dye, denatured and quenched, and then electrophoresed on a DNA analyser Gene ReadIR 4200 (LI-COR). The PCR products were scored as band presence (1) and absence (0), thus generating a binary data matrix. From this, the Polymorphic Information Content (PIC) was calculated for each locus using as described by Anderson et al. (1993) using Microsoft Office Excel software.

Linkage analysis

The segregation of alleles for those SSR markers informative between C3 and Sp-9A was followed in the C3 x Sp-9A population developed by Lanteri et al. (2006). Separate linkage maps were constructed for each parent using the double pseudo-testcross mapping strategy (Weeden 1994), incorporating previously scored genotypic data. Markers were separated into three types: maternal testcross markers, segregating only in C3 (expected segregation ratio 1:1); paternal testcross markers, segregating only in Sp-9A (1:1); and intercross markers, segregating within both parents (either 1:2:1 or 1:1:1:1). The goodness-of-fit between observed and expected segregation data was tested by χ^2, and only markers fitting or deviating slightly from expectation ($\chi^2_{\alpha=0.1} < \chi^2 \leq \chi^2_{\alpha=0.01}$) were used for map construction, using JoinMap v2.0 (Stam and Van Ooijen 1995). For both maps, linkage groups (LGs) were accepted on the basis of a LOD threshold of >4.0. To determine marker order within a LG, the parameter settings were Rec = 0.40, LOD = 1.0, Jump = 5. Map distances were converted to centiMorgans (cM) using the Kosambi mapping function (Kosambi 1944). Linkage maps were drawn using MapChart v2.1 (Voorrips 2002). A method-of-moments type estimator (Hulbert et al. 1988), as proposed in ‘method 3’ by Chakravarti et al. (1991), was used to estimate the genome length (G) of each parent.

Sequence annotation

Sequences were analysed with the BlastX or BlastN algorithm (Altschul et al. 1997). The non-default BlastX parameters applied were: database = reference proteins; organism = Viridiplantae; max target sequences = 50; matrix = BLOSUM62; filter = low complexity regions. No threshold was set. The BlastN parameters applied were: database = reference mRNA sequences; organism = Viridiplantae; optimize for = highly similar sequences (“Megablast”); filter = low complexity, species-specific repeats for *Arabidopsis*. The target database contained all the available Viridiplantae sequences (3,592,723 entries for BlastX, 32,825,875 for MegaBlast, March 2008). MegaBlast analyses covering six *Asteraceae* genera (*Lactuca*, *Helianthus*, *Chicorium*, *Taraxacum*, *Centaurea*, *Carthamus*) were executed using the same parameters, with a threshold of 1.0 e^{-8}. In some cases, local alignment hits with an e-value below the threshold were considered, where their annotation was interpretable. A second annotation was performed with the Blast search tool of AmiGO (http://amigo.geneontology.org/cgi-bin/gost/gost.cgi) using default parameter settings. GO annotations terms were reported for each CELMS locus (Table 3), considering the biological process (P), the cellular component (C) and molecular function (F). The gene structure prediction system Gene Builder (http://l25.itba.mi.cnr.it/%7Ewebgene/genebuilder.html) was used to confirm the presence of significant (>45 residues) open reading frames (ORFs), using parameters derived from *A. thaliana*. The CELMS loci which did not align with any GenBank entry were analysed using CENSOR (Jurka et al. 1996), applied in genome projects to identify and mask repetitive elements. Loci which contained an SSR motif within an ORF were designated as coding SSRs.

Experimental confirmation of expressed SSRs

The transcription of each CELMS locus was assayed by RT-PCR. Total RNA was extracted from eight week old leaves of Sp-9A using the Nucleospin RNA plant extraction kit (Macherey Nagel), and 2µg of this RNA were denatured at 70°C for 5min and then reverse-transcribed at 42°C for 1h in a 20µl reaction containing 100U M-MuLV reverse transcriptase (Fermentas), 0.5mM dNTP and 0.8µg dT_{15}, in the buffer supplied by the manufacturer; 5µl of a 1:10 dilution of this reaction were provided as template for a 20µl PCR containing 10pmol of each CELMS primer (Table 2), 1.5mM $MgCl_2$, 0.2mM dNTP, 1U GoTaq (Promega) in the buffer supplied by the manufacturer. The cycling conditions consisted of a denaturation of 94°C / 60s, followed by 27 cycles of 94°C / 30s, 55°C / 30s and 72°C / 60s, terminated with a 10min incubation at 72°C. Control reactions were derived from template produced in the absence of reverse transcriptase. In some cases, primers had to be re-designed (Table 4). RT-PCR products were separated by agarose gel electrophoresis and visualized by EtBr staining.

Results

SSR development and evaluation of marker polymorphism

A total of 279 positive clones was selected, producing 179 unique sequences. Of these, 99 were amenable to primer design, the remaining 80 were discarded because they either contained too little flanking sequence, or because the sequences were refractory to primer design. In all, 61 primer pairs (Table 2) reproducibly amplified a product, which consisted of two alleles per template; the remainder amplified poorly, or generated complex profiles. The recovery efficiency was thus 22% (61 out of 279).

Of the 61 CELMS loci, 51 were informative in one of the mapping populations; specifically 39 in C3 and Sp-9A, 43 in C3 and A-41 and, 50 in C3 and Creta4, respectively. The germplasm panel showed variation at 49 loci (Fig. 1A), allowing for the identification of 208 alleles (2-7 alleles per locus, mean 3.8). The PIC values varied from 0.23 to 0.77 (mean 0.52±0.02); CELMS-05 had the highest PIC, and CELMS-18 the lowest. Each genotype was uniquely distinguished by its combined SSR profile.

Linkage analysis

Of the 39 informative loci between C3 and Sp-9A, 12 segregated only within the female parent C3, six only within the male parent Sp-9A, and 21 (15 as 1:1:1:1 and 6 as 1:2:1) within both (Fig. 1B). Four loci suffered from mild segregation distortion, but only CELMS-33 showed a severely distorted segregation and was therefore excluded from the mapping exercise. Markers which segregated with only a minor deviation from the expected ratio are identified with one ($\chi^2_{\alpha=0.1} < \chi^2 \leq \chi^2_{\alpha=0.05}$) or two ($\chi^2_{\alpha=0.05} < \chi^2 \leq \chi^2_{\alpha=0.01}$) asterisks in Fig. 2. In all, 29 SSR loci were placed on the C3 map, distributed across 11 of the 18 major (containing a minimum of four markers) LGs described by Lanteri et al. (2006). Seven mapped to LG1. CELM-60 was linked to a previously orphan AFLP marker, thus generating a new LG (LG 19, Fig 2). On the Sp-9A map, 25 loci were placed on 11 of the 17 major LGs, including six on LG1; two loci allowed the definition of new LGs (LG18 and LG19, Fig 2). One intercross (CELMS-23) and two female-testcross (CELMS-25 and CELMS-45) loci remained unlinked.

As a result of the integration of the SSR loci, the 19 maternal LGs now comprise 239 markers, spanning 1373.0cM with a mean inter-marker distance of 5.7cM, and cover 53.2% of the estimated G. Similarly, the 19 paternal LGs comprise 212 markers, spanning 1294.9cM (54.3% of G) with a mean inter-marker distance of 6.1cM. The maternal and paternal maps share all 19 mapped SSR intercross markers, allowing for the definition of homologous LGs. In summary, 35 SSR loci were added to the genetic map, covering 12 of the 16 homologous LGs in addition to three non-aligned groups (Fig. 2).

Sequence analysis and annotation

The annotation pipeline resulted in 39 non-genic CELMS loci and 22 genic CELMS loci that contained at least one ORF (Table 3). Of the 39 non-genic loci, 15 were related to transposon-like elements, and 24 showed no similarity to any existing sequence. Of the 22 which shared sequence homology with database entries, five matched a transcription factor, five a transport protein, four a gene encoding a specific enzyme, four a protein involved in the signal transduction cascade, three a protein involved in the DNA repair processes and one in chromatin assembly (Fig. 3A, Table 3). Nine of the genes contained a protein-protein interaction domain associated with protein/DNA binding. The majority possess polyglutamine/asparagine, or polyproline tracts, known to be involved in protein-protein interactions (Berger et al. 2001). MegaBlast analysis within the *Asteraceae* produced nine high e-value hits in which the CELMS sequence aligned with an EST (http://cgpdb.ucdavis.edu/database/Database_Description.html, Table 3). In 12 loci, the repeat motif was present within an ORF. Of these, CELMS-18, CELMS-20 and CELMS-48 had conserved polyglutamine stretches, matching, respectively, auxin response factor 16 (ARF-16), a transcriptional co-repressor (LUG, Fig. 3B) and phytochrome 1 (PFT1). CELMS-33 and CELMS-37 carried polyproline stretches, matching, respectively, the cytosolic factor family protein 14 (SEC14) and a leucine-rich repeat-like protein. CELMS-38 had a polyhistidine stretch, homologous to a WRKY DNA binding protein (Table 3).

In the ten remaining CELMS, the SSR motifs were located either up- or downstream of the ORFs, or within an intron. Only four loci, out of the 22 genic CELMS, showed evidence of transcriptional activity in leaf tissue (CELMS-5, -38, -47, -49), but when new primer pairs were designed targeted to the coding sequence (Table 4), a further five such loci (CELMS-18, -20, -37, -48, -52) were identified.

Discussion

SSR development and evaluation of marker polymorphism

Until March 2008, only 173 *Cynara* spp. DNA sequences were present in the GenBank database; these included 32 SSR-containing sequences previously developed (Acquadro et al. 2003; Acquadro et al. 2005a,b) of which twelve were mapped in the globe artichoke genetic map (Lanteri et al. 2006). The main objective of the present work was to develop additional informative SSR markers from enriched genomic libraries, to improve the genetic maps of *Cynara cardunculus.* At the same time, their usefulness for genotype identification and phylogenetic studies was assessed.

In conventional methods for SSR isolation from genomic libraries, the efficiency of recovery is rather low, varying from 0.045% to 12% (Zane et al. 2002). The necessary procedures tend to be time and labour intensive, and thus costly. As a result, a number of library enrichment methods has been proposed (Acquadro et al. 2005b; Squirrell et al. 2003). Oligo hybridization capture techniques, based on either probe immobilization on filters or on streptavidin coated magnetic beads improves the recovery rate of SSR-containing sequences to 20-90% across a variety of taxa (Zane et al. 2002). The enrichment protocol used here was based on the targeting of ten repetitive di- or trinucleotide motifs known to occur frequently in the coding regions of plant genomes (Morgante et al. 2002). A surprisingly high level of redundancy was encountered, resulting in the loss of 100 out of the original 279 positive clones. Duplication of clones was assumed to have occurred during the enrichment phase, and may be associated with the two step PCR procedure, each comprising 25 cycles. Our subsequent experience has indicated that 15-20 cycles per PCR does lessen the extent of clone redundancy.

The informativeness of the CELMS SSRs was comparable with what has been demonstrated for an earlier set of both artichoke (Acquadro et al. 2005b), sunflower (Tang et al. 2003; Paniego et al. 2002) and lettuce (van de Wiel et al. 1999) SSRs. Furthermore, the application of three CELMS (-9, -14 ,-40) markers for addressing the pattern of genetic diversity of a collection of Sicilian globe artichoke landraces from small-holdings, made it possible to gather information on the evolution and domestication of the species (Mauro et al. 2008).

Linkage analysis and marker distribution

About 10% of the SSR loci suffered from segregation distortion, consistent with the level found for the markers used by Lanteri et al. (2006) to construct the first artichoke genetic maps. Segregation distortion has been associated with statistical bias or errors in genotyping and scoring, but stems mainly from a number of biological phenomena affecting meiosis, fertilization and embryogenesis (Bradshaw and Stettler 1994). The presence of null alleles, which is not uncommon in the context of SSR loci (due to failure of one or both primers to anneal), can also contribute to apparent skewing, as homozygotes become indistinguishable from non-null allele containing heterozygotes (Pekkinen et al. 2005). In the present work, we have chosen to include markers deviating at 1% level and above; although the inclusion of distorted loci into the map increases the chance of type I errors of false linkage, these loci can be useful in increasing our knowledge on specific regions of the genome. The newly developed SSR set has increased the number of mapped SSRs from nine to 39 in the C3 map, and from five to 34 in the Sp-9A map. The number of intercross SSRs, which serve as bridge markers between the two maps, was increased from seven to 26, resulting in the identification of 12 homologous and 3 not aligned LGs covered by one to seven SSR markers.

The new female map spanned 1373.0cM with a mean inter-marker distance of 5.7cM, representing only a 3% increase in the total length of the map, but in a ~12% decrease in the mean inter-marker distance. Similarly, the male map was increased by ~5% in length, with a ~12% decrease in the mean inter marker distance.

Since the CELMS markers appear to be well distributed along the LGs, it is likely that SSR loci are dispersed throughout the artichoke genome. Some clustering of SSRs has been observed around the putative centromeric region of LG1, 2, 3 and 10, a pattern which is not unusual (Arens et al. 1995; Bhattramakki et al. 2000; Gill et al. 2006; Jones et al. 2002; McCouch et al. 2002; Ramsay et al. 2000). The addition of new markers has allowed the filling of some of the gaps in the base maps, especially on LGs 4, 8 and 14; and the addition of a second bridge marker to LG 9. However, the distal region of LG 13 remains sparsely populated, and the gaps in LG 6, 14 and 15 remain unfilled. Increasing marker density, and the addition of genes underlying phenotypic traits to a map, requires the creation of mapping populations from parents which segregate for the latter, but retain common sets of markers (Hayes et al. 1996; Weeden et al. 2000). Examples of such consensus maps have been reported for several crops (Ellis et al. 1992; Kleinhofs et al. 1993; Tanksley et al. 1992). Markers in common across populations can serve as anchors to locate important genes to a particular LG, thereby allowing the location of genes underlying phenotype even in populations where these do not segregate. We are currently constructing genetic maps based on crosses between Romanesco C3 and cultivated or wild cardoon, which are genotypically/phenotipically highly divergent, to facilitate comparative QTL mapping. A high proportion (49 out of the 61) of the CELMS markers were suitable for mapping in multiple populations, and so represent a set of robust and informative anchor points in *C. cardunculus* populations.

By blasting all the CELMS loci against the *Asteraceae* dbESTs, we found 9 hits, putative orthologues loci from lettuce, sunflower and chicory. Four of them (CELMS-5, -16, -52 in LG-1 and CELMS-48 in LG2) were placed on the globe artichoke linkage maps and might be used as anchor markers for map alignment within the *Asteraceae* family.

Sequence annotation

We have annotated the CELMS loci in an attempt to convert anonymous markers to those associated with specific biological functions. The sequence of the SSR loci provides a handle on putative function, provided that it shares homology with already characterized orthologous sequences. This approach led to the assigning of putative function to about one third of the CELMS sequences. Most of these (20 out of 22) were among the trinucleotidic motif sequences; the two dinucleotidic types (CELMS-05 and CELMS-60) were both “coding SSRs” carrying GA_n/CT_n as stretches of glutamate-arginine or serine-leucine. The dominance of trinucleotidic motifs in genic SSRs has been reported in both *A. thaliana* and soybean (Morgante et al. 2002).

The sequences of CELMS-18, -20, -33, -37, -38, -47, -48 are likely to be orthologues of genes with known function, as they both show a high level of sequence similarity and retain the SSR sequence in the equivalent position. In CELMS-18 and -20 the orthologous sequences are conserved in the flanking regions, but not in the SSR itself (CAA in globe artichoke and CAG in *A. thaliana*), a pattern which has been previously noted in comparisons between rice and *A. thaliana* (Zhang et al. 2006).

As previously performed in the *Solanaceae* (Wu et al. 2006; Wu et al. 2009), *Fabaceae* (Phan et al. 2007; Hougaard et al. 2008; Ellwood et al. 2008) and *Asteraceae* (Chapman et al. 2007) families, a COS marker approach may represent an effective mean for generating molecular markers. Our comparative analysis among the *Asteraceae* species showed similarity values up to 100% between sequences from globe artichoke and those from the yellow starthistle (*Centaurea solstitialis*) or safflower (*Carthamus tinctorius*); accordingly, an exploration of the *Asteraceae* dbEST seems very promising for new microsatellite markers mining, as well as for synteny studies.

Conclusions

We have developed, annotated and mapped a set of 61 new genomic globe artichoke SSR markers, with the aim to extend the limited number of co-dominant markers currently available; these markers represent valuable tools for genetic analysis of the species. The new SSRs were uniformly distributed in the already developed globe artichoke maps, thus improving their coverage and contributing in future alignment of the new maps under development.

Acknowledgements

We are very grateful to Prof. Giovanni Mauromicale and Dr. Rosario Mauro, form D.A.C.P.A. – Agronomical Sciences, University of Catania, for having provided most of the *C. Cardunculus* genotypes in study. We also thank Dr. Robert Koebner (www.smartenglish.co.uk) for manuscript editing.

References

Acquadro A, Portis E, Lanteri S (2003) Isolation of microsatellite loci in artichoke (*Cynara cardunculus* L. var. *scolymus*). Mol Ecol Notes 3:37-39

Acquadro A, Portis E, Albertini E, Lanteri S (2005a) M-AFLP-based protocol for microsatellite loci isolation in *Cynara cardunculus* L. (Asteraceae). Mol Ecol Notes 5:272-274

Acquadro A, Portis E, Lee D, Donini P, Lanteri S (2005b) Development and characterization of microsatellite markers in *Cynara cardunculus* L. Genome 48:217-225

Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman D (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402

Anderson J, Churchill G, Autrique J, Tanksley S, Sorrells M (1993) Optimizing parental selection for genetic-linkage maps. Genome 36:181-186

Arens P, Odinot P, Vanheusden A, Lindhout P, Vosman B (1995) GATA-repeats and GACA-repeats are not evenly distributed throughout the tomato genome. Genome 38:84-90

Berger M, Sionov R, Levine A, Haupt Y (2001) A role for the polyproline domain of p53 in its regulation by Mdm2. J Biol Chem 276:3785-3790

Bhattramakki D, Dong J, Chhabra A, Hart G (2000) An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. Genome 43:988-1002

Bradshaw H, Stettler R (1994) Molecular-genetics of growth and development in *Populus* .2. Segregation distortion due to genetic load. Theor Appl Genet 89:551-558

Chakravarti A, Lasher L, Reefer J (1991) A maximum-likelihood method for estimating genome length using genetic-linkage data. Genetics 128:175-182

Chapman M, Chang J, Weisman D, Kesseli R, Burke J (2007) Universal markers for comparative mapping and phylogenetic analysis in the Asteraceae (Compositae). Theor Appl Genet 115:747-755

Ellis T, Turner L, Hellens R, Lee D, Harker C, Enard C, Domoney C, Davies D (1992) Linkage maps in pea. Genetics 130:649-663

Ellwood SR, Phan HT, Jordan M, Hane J, Torres AM, Avila CM, Cruz-Izquierdo S, Oliver RP. (2008) Construction of a comparative genetic map in faba bean (*Vicia faba* L.); conservation of genome structure with *Lens culinaris*. BMC Genomics 9;9:380

Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

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. Toxicol Appl Pharm 144:279-286

Gebhardt R (1998) Inhibition of cholesterol biosynthesis in primary cultured rat hepatocytes by artichoke (*Cynara scolymus* L.) extracts. J Pharmacol Exp Ther 286:1122-1128

Gerber H, Seipel K, Georgiev O, Hofferer M, Hug M, Rusconi S, Schaffner W (1994) Transcriptional activation modulated by homopolymeric glutamine and proline stretches. Science 263:808-811

Gill G, Wilcox P, Whittaker D, Winz R, Bickerstaff P, Echt C, Kent J, Humphreys M, Elborough K, Gardner R (2006) A framework linkage map of perennial ryegrass based on SSR markers. Genome 49:354-364

Guo H, Moose S (2003) Conserved noncoding sequences among cultivated cereal genomes identify candidate regulatory sequence elements and patterns of promoter evolution. Plant Cell 15:1143-1158

Hayes P, Schmitt K, Jones H, Gyapay G, Weissenbach J, Goodfellow P (1996) Regional assignment of human ESTs by whole-genome radiation hybrid mapping. Mamm Genome 7:446-450

Hougaard BK, Madsen LH, Sandal N, de Carvalho Moretzsohn M, Fredslund J, Schauser L, Nielsen AM, Rohde T, Sato S, Tabata S, Bertioli DJ, Stougaard J (2008) Legume anchor markers link syntenic regions between *Phaseolus vulgaris*, *Lotus japonicus*, *Medicago truncatula* and *Arachis*. Genetics 179:2299-312

Hulbert S, Ilott T, Legg E, Lincoln S, Lander E, Michelmore R (1988) Genetic analysis of the fungus, *Bremia lactucae*, using restriction fragment length polymorphisms. Genetics 120:947–958

Iglesias A, Kindlund E, Tammi M, Wadelius C (2004) Some microsatellites may act as novel polymorphic cis-regulatory elements through transcription factor binding. Gene 341:149-165

Jones E, Dupal M, Dumsday J, Hughes L, Forster J (2002) An SSR-based genetic linkage map for perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 105:577-584

Jurka J, Klonowski P, Dagman V, Pelton P (1996) Censor - A program for identification and elimination of repetitive elements from DNA sequences. Comp Chem 20:119-121

Kleinhofs A, Kilian A, Maroof M, Biyashev R, Hayes P, Chen F, Lapitan N, Fenwick A, Blake T, Kanazin V, Ananiev E, Dahleen L, Kudrna D, Bollinger J, Knapp S, Liu B, Sorrells M, Heun M, Franckowiak J, Hoffman D, Skadsen R, Steffenson B (1993) A molecular, isozyme and morphological map of the barley (*Hordeum-vulgare*) genome. Theor Appl Genet 86:705-712

Kolaczkowska A, Kolaczkowski M, Delahodde A, Goffeau A (2002) Functional dissection of Pdr1p, a regulator of multidrug resistance in *Saccharomyces cerevisiae*. Mol Ecol Notes 267:96-106

Kosambi D (1944) The estimation of map distances from recombination values. Ann Eugen, 12: 172–175

Kraft K. (1997) Artichoke leaf extract. Recent findings reflecting effects on lipid metabolism, liver and gastrointestinal tracts. Phytomedicine 4:369-378

Lanteri S, di Leo I, Ledda L, Mameli M, Portis E (2001) RAPD variation within and among populations of globe artichoke cultivar 'Spinoso sardo'. Plant Breeding 120:243-246

Lanteri S, Saba E, Cadinu M, Mallica G, Baghino L, Portis E (2004) Amplified fragment length polymorphism for genetic diversity assessment in globe artichoke. Theor Appl Genet 108:1534-1544

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 112:1532-1542

Leroy XJ, Leon KY, Branchard M (2000) Plant genomic instability detected by microsatellite-primers. Electron J Biotechn 3(2): 5-10

Llorach R, Espin J, Tomas-Barberan F, Ferreres F (2002) Artichoke (*Cynara scolymus* L.) bio-products as a potential source of health-promoting antioxidant phenolics. J Agr Food Chem 50:3458-3464

Maccarone E, Fallico B, Fanella F, Mauromicale G, Raccuia SA, Foti S (1999) Possible alternative utilization of *Cynara* spp. II. Chemical characterization of their grain oil. Ind Crop Prod 10(3): 229-237

Marie D, Brown S (1993) A cytometric exercise in plant DNA histograms, with 2C-values for 70 species Biol Cell 78:41-51

Martin P, Makepeace K, Hill S, Hood D, Moxon E (2005) Microsatellite instability regulates transcription factor binding and gene expression. P Natl Acad Sci USA 102:3800-3804

Mauro R, Portis E, Acquadro A, Lombardo S, Mauromicale G, Lanteri S (2008) Genetic diversity of globe artichoke landraces from Sicilian small-holdings: implications for evolution and domestication of the species. Conserv Genet DOI: 10.1007/s10592-008-9621-2

McCouch S, Teytelman L, Xu Y, Lobos K, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y, Zhang Q, Kono I, Yano M, Fjellstrom R, DeClerck G, Schneider D, Cartinhour S, Ware D, Stein L (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). Dna Res 9:199-207

McDougall B, King P, Wu B, Hostomsky Z, Reinecke M, Robinson W (1998) Dicaffeoylquinic and dicaffeoyltartaric acids are selective inhibitors of human immunodeficiency virus type 1 integrase. Antimicrob Agents Ch 42:140-146

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nature Genet 30:194-200

Oetting W, Lee H, Flanders D, Wiesner G, Sellers T, King R (1995) Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. Genomics 30:450-458

Paniego N, Echaide M, Munoz M, Fernandez L, Torales S, Faccio P, Fuxan I, Carrera M, Zandomeni R, Suarez E, Hopp H (2002) Microsatellite isolation and characterization in sunflower (*Helianthus annuus* L.). Genome 45:34-43

Pekkinen M, Varvio S, Kulju K, Karkkainen H, Smolander S, Vihera-Aarnio A, Koski V, Sillanpaa M (2005) Linkage map of birch, *Betula pendula* Roth, based on microsatellites and amplified fragment length polymorphisms. Genome 48:619-625

Perutz M, Johnson T, Suzuki M, Finch J (1994) Glutamine repeats as polar zippers - their possible role in inherited neurodegenerative diseases. P Natl Acad Sci USA 91:5355-5358

Phan HT, Ellwood SR, Hane JK, Ford R, Materne M, Oliver RP (2007) Extensive macrosynteny between *Medicago truncatula* and *Lens culinaris* ssp. *culinaris*. Theor Appl Genet 114:549-58

Raccuia SA, Melilli MG 2004. *Cynara cardunculus* L., a potential source of inulin in Mediterranean environment: screening of genetic variability. Aust J Agr Res 55: 693-698

Raccuia, SA., Melilli, M.G., 2007 Biomass and grain oil yields in *Cynara cardunculus* L. genotypes grown in a Mediterranean environment. Field Crop Res 101: 187-197

Ramsay L, Macaulay M, Ivanissevich S, MacLean K, Cardle L, Fuller J, Edwards K, Tuvesson S, Morgante M, Massari A, Maestri E, Marmiroli N, Sjakste T, Ganal M, Powell W, Waugh R (2000) A simple sequence repeat-based linkage map of barley. Genetics 56:1997-2005

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Humana Press, Totowa, NJ, Bioinformatics Methods and Protocols: Methods in Molecular Biology. 132:365-86

Squirrell J, Hollingsworth P, Woodhead M, Russell J, Lowe A, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? Mol Ecol 12:1339-1348

Stam P, Van Ooijen J (1995) JoinMap version 2.0: software for the calculation of genetic linkage maps. CPRO-DLO, Wageningen, The Netherlands

Tang, S, Kishore, VK, Knapp, SJ (2003). PCR-multiplexes for a genome-wide framework of simple sequence repeat marker loci in cultivated sunflower. Theor Appl Genet 107: 6-19.

Tanksley S, Ganal M, Prince J, Devicente M, Bonierbale M, Broun P, Fulton T, Giovannoni J, Grandillo S, Martin G, Messeguer R, Miller J, Miller L, Paterson A, Pineda O, Roder M, Wing R, Wu W, Young N (1992) High-density molecular linkage maps of the tomato and potato genomes. Genetics132:1141-1160

Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: Survey and analysis. Genome Res 10:967-981

van de Wiel C, Arens P, Vosman B (1999) Microsatellite retrieval in lettuce (*Lactuca sativa* L.). Genome 42:139-149

Voorrips R (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. J Hered 93:77-78

Wang M, Simon J, Aviles I, He K, Zheng Q, Tadmor Y (2003) Analysis of antioxidative phenolic compounds in artichoke (*Cynara scolymus* L.). J Agr Food Chem 51:601-608

Weeden N (1994) Approaches to mapping in horticultural crops. In: Gressho. P (ed) Plant genome analysis. CRC Press Boca Raton, pp 51–60

Weeden N, Ellis T, Timmerman-Vaughan G, Simon C, Torres A, Wolko B (2000) How similar are the genomes of the cool season food legumes. Kluwer Academic Publishers, Dordrecht, The Netherlands, Knight R (ed) Linking research and marketing opportunities for pulses in the 21st Century pp 397-410

Wu F, Mueller LA, Crouzillat D, Petiard V, Tanksley SD (2006) Combining Bioinformatics and Phylogenetics to Identify Large Sets of Single Copy, Orthologous Genes (COSII) for Comparative, Evolutionary and Systematics Studies: A Test Case in the Euasterid Plant Clade. Genetics 174:1407-20

Wu F, Eannetta NT, Xu Y, Tanksley SD (2009) A detailed synteny map of the eggplant genome based on conserved ortholog set II (COSII) markers. Theor Appl Genet. Doi: 10.1007/s00122-008-0950-9

Yang Y, Lai K, Tai P, Li W (1999) Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages. J Mol Evol 48:597-604

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. Mol Ecol 11:1-16

Zhang L, Yu S, Cao Y, Wang J, Zuo K, Qin J, Tang K (2006) Distributional gradient of amino acid repeats in plant proteins. Genome 49:900-905

Table 1. The 22 *C. cardunculus* genotypes assayed for genotypic variation.

Genotypes	***C. cardunculus* taxa**	**Cluster**[1]
Romanesco C3 (C3)	*scolymus*	**A2**
Spinoso di Palermo (Sp-9A)	*scolymus*	**B1**
A-41	*altilis*	
Creta-4	*sylvestris*	
Four F_1 genotypes from C3 x Sp-9A	*scolymus*	
Four F_1 genotypes from C3 x A-41	*scolymus* x *altilis*.	
Four F_1 genotypes from C3 x Creta-4	*scolymus* x *sylvestris*	
Gross Camus	*scolymus*	**A1**
Hyerois	*scolymus*	**A1**
Tonda di Paestum	*scolymus*	**A2**
Violet de Campagne	*scolymus*	**B1**
Empolese	*scolymus*	**B2**
Locale di Chioggia Fano	*scolymus*	**B2**

[1] Globe artichoke clusters defined in Lanteri et al. (2004)

Table 2. Primer sequences of the 61 CELMS markers and their level of polymorphism; PIC = Polymorphic Information Content; Na =Number of alleles; LG = Linkage group; S x R = progeny Sp-9A x C3; A x R = progeny A41 x C3; Sy x R = progeny Creta4 x C3; Acc. = cultivated globe artichoke genotypes; ‘+’ and ‘-’ denote, respectively, a polymorphic or a monomorphic locus.

Locus	Forward primer (5'-3')	Reverse primer (5'-3')	Repeats	Size	PIC	N_A	LG	SxR	AxR	SyxR	Acc.	GenBank
C-ELMS-01	ACAACACAGAAGCGAGGTCA	GAATGAGCCGGATTAGCATT	$(AG)_{17}$	356	0,45	3	11	+	+	+	+	EU744917
C-ELMS-02	TCCCTCAAGTCAAGCGAGTT	GGAGGGAGGGTTCAAGCTAC	$(TG)_6(GA)_{20}$	307	0,66	5	-	-	+	+	+	EU744918
C-ELMS-03	GATCAATACGTCGTCGCAGA	CTGAGGCTACCAAGGGTTTG	$(TC)_{18}(AC)_7$	392	0,71	5	3	+	+	+	+	EU744919
C-ELMS-04	TTTGTCAACCCATACGCAAC	AATCCAATCATATTACCATGTAATA	$(GA)_{19}$	274	0,67	5	10	+	+	+	+	EU744920
C-ELMS-05	CCCACCTTCTTATCCCATCA	TGGACGTCTGTTTCCTCCTC	$(CT)_{22}(CAG)_4$	309	0,77	7	1	+	+	+	+	EU744921
C-ELMS-06	CTCCATTCTGTGATGCAGTGA	TGTATCAACCTTGGCCTTCC	$(TC)_{19}(CA)_{10}(AC)_5$	231	0,35	3	-	-	-	+	+	EU744922
C-ELMS-07	AAGGCAGGGTTAGAGTGACAAC	AGACTCCATGCTTCACACAGAT	$(AG)_{22}$	196	0,30	4	8	+	-	+	+	EU744923
C-ELMS-08	TTTACAAACTTCCCCTTTCCAC	ACAATACAGATACACGCTCTCCA	$(TC)_{22}(TC)_8(TC)_{10}$	247	0,72	6	1	+	+	+	+	EU744924
C-ELMS-09	TCATCAATTGATCTGATAAGC	TTCGGTGCTAGGTAGACTT	$(CT)_{19}\ldots(TCT)_7$	195	0,65	4	-	-	+	+	+	EU744925
C-ELMS-10	TCAGACTTCAGCACCACCTC	GTCGTTCTGGATTCCCACAT	$(AG)_{25}$	315	0,70	4	9	+	+	+	+	EU744926
C-ELMS-11	GCGAATCAATCCCTTGTCTC	AAGCCATGGATGAAGCAGAG	$(TC)_{21}(TTTG)_3$	258	0,66	6	18	+	+	+	+	EU744927
C-ELMS-12	TTGATGAATTTTGATCACTA	ACCATTATCCTTTTGCTC	$(GT)_9(GA)_9(TG)_7$	326	0,55	4	16	+	+	+	+	EU744928
C-ELMS-13	ATGGGACCTTCCTCCAAAATAC	TCCATCATCACCTCACACGTA	$(TA)_7(AC)_{15}$	400	0,73	5	3	+	+	+	+	EU744929
C-ELMS-14	TCCAGCCATGCAAGAAAAGTAT	CCATCCTGAATCCATAACCAGT	$(AC)_{13}(TC)_7(AC)_{10}(TC)_9$	210	0,61	5	8	+	+	+	+	EU744930
C-ELMS-15	TGGATGGAAACACTCTTCACAG	TACAGTCCCGATGTGGGTATTT	$(CA)_{15}(TA)_5(ATGT)_{10}(TG)_5$	350	0,62	5	3	+	+	+	+	EU744931
C-ELMS-16	CTCTCTTTACCCTACTCATAA	CTTTTGGGGTTTCTATACC	$(AC)_{15}(AC)_{14}$	257	0,50	3	1	+	+	+	+	EU744932
C-ELMS-17	CCCGGATAATAGTCGATGAAGT	CCATGTGAAGATTGGGTGATT	$(GTT)_{32}$	305	0,32	4	-	-	+	+	+	EU744933
C-ELMS-18	TCCCTCCCATTGTTTCTTCTAA	CTGTTGCTGTTGCTGTAGCTG	$(CAA)_7(CAA)_4$	344	0,23	3	-	-	+	+	-	EU744934
C-ELMS-19	GATGGTGCTTCTTTCTTTTCCT	TAATATCCCAACCGTCCCC	$(TTG)_5(TTG)_6(TTG)_6(TTG)_6$	297	0,76	5	14	+	+	+	+	EU744935
C-ELMS-20	TTTTATAATTGCAGACTCAAT	TTCATTTCCAACAAGCCT	$(CAG)_5(CAA)_5(CAA)_8(CAA)_{12}$	218	0,52	3	10	+	+	+	+	EU744936
C-ELMS-21	TGTCATCAACCCCTACTCAGG	TTCAGATTTACTAACCCAAATGCTT	$(TCT)_4(TTC)_5(TTC)_4(TCT)_{12}$	388	0,56	4	14	+	+	+	+	EU744937
C-ELMS-22	TTTTCATCATCTCCTTCATGG	GCTTAGAGAAAGGGGAAAGAGG	$(CTT)_{22}(CTC)_6(TTG)_6$	392	0,74	5	-	-	-	+	+	EU744938
C-ELMS-23	GGCCCTACCTTAAAATGTCTCC	GACGGTGATTGTTGTAGTGGAA	$(CCA)_5(CCA)_5$	241	0,50	2	-	+	+	+	+	EU744939
C-ELMS-24	ACCAAACTCTGTCGACCACC	GGTTGTGGAGGACCTGGATA	$(CAC)_4(CCA)_{12}$	242	0,61	4	5	+	+	+	+	EU744940
C-ELMS-25	TTATCAGCCACCTCCACCTC	GACGGGCAATGGTAGTCAAT	$(CCA)_4(CAC)_7$	288	0,69	5	-	+	+	+	+	EU744941
C-ELMS-26	ACCATGTCACAACAAACCGA	TGATTCTCGTAGGTGGAGGG	$(CCA)_9(CAC)_8$	388	0,55	4	1	+	+	+	+	EU744942
C-ELMS-27	ACTGTTGTTGCTGGTAAGGGTT	AGAAAGGAGGAGGAAAGCATCT	$(ACC)_6$	367	0,57	4	1	+	-	+	+	EU744943
C-ELMS-28	GAAAGAAGATGCATAGACCAGGA	CCTCCAGCTGCTGCCTAATA	$(CCA)_4(CCT)_4(CAC)_4$	195	0,24	3	-	-	-	+	+	EU744944
C-ELMS-29	ATCCCCAAATCCAGCAATTT	TCAATGTGCATGGAAAGAACA	$(CCA)_5(CCT)_4$	296	0,48	2	2	+	+	+	+	EU744945

Locus	Forward primer (5'-3')	Reverse primer (5'-3')	Repeats	Size	PIC	N_A	LG	SxR	AxR	SyxR	Acc.	GenBank
C-ELMS-30	TCAGGCACCTCAAACTCCTT	CAGGTGCATGACCACCTAGT	$(ACC)_7$	294	0,56	3	19	+	-	+	+	EU744946
C-ELMS-31	AAATGGATATTGGAACACCTCC	TATTTGAGGAATGTCTGCTGCT	$(CCA)_4(CCA)_{10}(CCA)_4$	140	0,65	3	4	+	+	+	+	EU744947
C-ELMS-32	ACCTCCACCACCTTGTCCTC	CATGTAGTGCCTGGATATGG	$(ACC)_5(CAC)_5(CAC)_7$	177	0,46	2	6	+	+	+	+	EU744948
C-ELMS-33	GATGCACCACTTTCCTCTCAC	ATATGGGCTTTTCTGGTTGTTC	$(CAC)_4(CCA)_{11}$	190	0,65	4	-	+	+	+	+	EU744949
C-ELMS-34	ACCGCCCGTCGTTGCC	CGCCTAGCAGTTGTGGAAGTGG	$(ACC)_{11}$	169	0,00	1	-	-	-	-	-	EU744950
C-ELMS-35	CTCCCCCTCCGGTTCAAT	GAACCGATGTGGGGTGGTA	$(CCA)_4(CAC)_6(CCA)_5(CAC)_4$	299	0,42	3	-	-	-	+	+	EU744951
C-ELMS-36	CACCACTAGTACAATTAACCAT	AGTAGTGGTAGTTGATGTTAGA	$(CAC)_4(ACC)_5(CAC)_5(ACC)_6(CCA)_4$	241	0,63	4	5	+	+	+	+	EU744952
C-ELMS-37	CGCCGGAATATCAAGATTGT	TACCATCAACTCGGAGAGGG	$(CCA)_8$	300	0,68	5	14	+	+	+	+	EU744953
C-ELMS-38	ACTGGGGTTTACAAGCTGTGAT	CTCCTGATGTGTTGTTCTTGATG	$(ATC)_9(TCC)_4(TCA)_9$	409	0,40	3	-	-	-	+	+	EU744954
C-ELMS-39	ATTCCAATCACCTCTGTGGC	ACTGTATGGTGAAGTCGTTA	$(GAT)_5(GAT)_{14}$	182	0,42	3	10	+	+	+	+	EU744955
C-ELMS-40	TGGATTAAGGCACACACTGAAC	TGATGATAACAAAGGAGGGGAT	$(ATC)_4(ATC)_{16}(TCT)_9$	388	0,75	6	1	+	+	+	+	EU744956
C-ELMS-41	CCAAAGCCTTCAGAGCATTC	GGAATGATGTATGGATCGCC	$(ATG)_{11}(GAT)_9$	271	0,59	4	2	+	+	+	+	EU744957
C-ELMS-42	AAGCTGAAGTCGAGGAACCA	TGGGATGAAGATTCCCAGAG	$(TGA)_4(TGA)_8(GAT)_5$	358	0,66	3	3	+	+	+	+	EU744958
C-ELMS-43	CCTTCACCCCTGTCTACAAGAT	GGGGAGGCACGATGAG	$(ATG)_{10}$	288	0,00	1	-	-	-	-	-	EU744959
C-ELMS-44	GTTCCACGTTTGAAGCGAGT	TTTGCTATTGTCCATAAAAGATTGA	$(CTA)_{16}(ACT)_4(ACT)_4(TAC)_4(ACC)_4$	249	0,50	2	5	+	+	+	+	EU744960
C-ELMS-45	TTCTGTGGAGAGTTTCATCCAA	TAGCTTGCTCACGCTCAGTG	$(TCT)_{103}$	426	0,48	5	-	+	+	+	+	EU744961
C-ELMS-46	CATTAGCGTATCTAGTGGAGAAAGACT	GCCATCTTCTTCTTCTACTCAGG	$(AGA)_{52}(AGA)_4$	250	0,00	1	-	-	-	-	-	EU744962
C-ELMS-47	TGGAAAGGGGAGAGAAACAA	CTGGTGATCAAGGCCAGAGT	$(AGA)_{28}(AAG)_6$	222	0,00	1	-	-	-	-	-	EU744963
C-ELMS-48	ATAACAGGACGAGGTGTGGAAG	CTACAGTTGCTTATTGGTCCCC	$(CTG)_5(CTG)_7$	321	0,57	3	2	+	+	+	+	EU744964
C-ELMS-49	AGCAACAGCCACAACAACTTC	TGGACCTTGAACATAACCTTGA	$(CAG)_6(CAG)_4$	215	0,29	3	-	-	+	-	+	EU744965
C-ELMS-50	AACAGCAGCAGCAACAAATAAG	GGACGAAAGAAAAGGAACACAG	$(CAG)_5(AGC)_5$	190	0,00	1	-	-	-	-	-	EU744966
C-ELMS-51	CTTGTTGATGCTGTTGTCGAGT	TAGGGCTGTGTTTTGACCTTTT	$(CTG)_6(TGC)_4$	226	0,00	1	-	-	-	-	-	EU744967
C-ELMS-52	TGCAGCAAATTCTTTTGTGG	TGTGGGAACCTCTATAATCTCTTTG	$(CT)_{18}$	301	0,53	4	1	+	+	+	+	EU744968
C-ELMS-53	TTTGTTCACGGAATTCAACG	GCCCTGTCCTCGATAAGATG	$(GA)_{18}$	235	0,00	1	-	-	-	-	-	EU744969
C-ELMS-54	CGAAAAGAGTTCAAGAGGGAAA	GCACCTGAAGCATCTGAGG	$(GAA)_n$	180	0,00	1	-	-	-	-	-	EU744970
C-ELMS-55	CTCTAGTCGCAGAGGATGGA	TGCCACATTTAAAGCAACCA	$(GAGAAG)_2$	318	0,00	1	-	-	-	-	-	EU744971
C-ELMS-56	CCTAGGGATGATGCCCATAC	ATGGAGTCGATTCACCTTGC	$(TGA)_6(GAT)_4$	250	0,00	1	-	-	-	-	-	EU744972
C-ELMS-57	GTTGGGGTGTCAAAACGAAT	CCAAGGGGATGACTAAGAGC	$(TCT)_{10}$	243	0,41	3	-	-	+	+	+	EU744973
C-ELMS-58	GGATTCCATTGGACTTACAGG	GGTTTGCCTATCTCTGTCTTTCTT	$(AG)_{18}(AGAA)_3$	259	0,66	4	1	+	+	+	+	EU744974
C-ELMS-59	TCCGTTATTTCTTGCGGTTA	TACCTCTCCGGTTGGAATTG	$(CT)_{16}(TC)_8$	399	0,29	3	2	+	+	+	+	EU744975
C-ELMS-60	TGGTGGGAAAAGGAGTGTTT	CATACCCACCCTGCAAGTTA	$(GA)_5(GA)_{10}(GA)_{12}(GA)_{14}(GA)_6(AG)_5$	381	0,59	3	19	+	+	+	+	EU744976
C-ELMS-61	TGCAAACCAGAAACTGCTTG	TGCAGACTTTACCTCCACCA	$(CT)_{18}(GT)_8$	170	0,32	3	-	-	+	+	-	EU744977

Table 3. Annotation of the CELMS sequences and their putative function.

											GO annotation		
Locus	**ORF**	**RT-PCR**	**SSR position**	**Stretch**	**dbEST, *Asteraceae***	**Similarity (Protein)**	**Identity (DNA)**	**e-value** [(2)]	**Algorithm**	**Putative Function**	**Function (F)**	**Process (P)**	**Component (C)**
Similarities with genes													
C-ELMS-03	188-367	-	SSR after ORF	-	-	NP_199800	NM_124367.3	2 e^{-08}	BlastX	CLC-C (chloride channel C)	voltage-gated chloride channel	chloride transport	membrane
C-ELMS-04	41-165	-	SSR after ORF	-	-	NP_181138	NM_129153.2	1 e^{-08}	BlastX	BLH1 (BLH1)	transcription factor	regulation of transcription	nucleus
C-ELMS-05	146-466	+	Coding SSR	(SL)$_n$	EL399062, *C. tinctorius*, 9 e^{-62}	NP_680740	NM_148374.2	2 e^{-18}	BlastX	Peptidoglycan-binding LysM domain-containing protein	-	cell wall catabolic process	-
C-ELMS-06	364-783	-	SSR before ORF	-	-	NP_191366	NM_115669.2	7 e^{-09}	BlastX	GIS (Glabrous inflorescence stems)	nucleic acid binding	regulation of transcription	intracellular
C-ELMS-10	-	-	-	-	-	NP_180289	NM_128279.3	9 e^{-06}	BlastX	Kelch repeat-containing ser/thr phosphoesterase	phosphoprotein phosphatase	-	intracellular
C-ELMS-15	395-587	-	SSR after ORF	-	-	NP_192095	NM_116416.3	4 e^{-14}	BlastX	Transducin/WD-40 repeat family protein	-	-	-
C-ELMS-16	169-215 [(1)]	-	SSR before ORF	-	AB274889, *L. sativa*, 9 e^{-40}	-	EV203634	3 e^{-39}	MegaBlast	Transcribed locus, *B. napus*	-	electron transport/photosynthesis	chloropl./thylak./PSI
C-ELMS-17	60-144; 357-456	-	Coding SSR	-	-	NP_200631	NM_125208.3	9 e^{-17}	BlastX	MSI1 (Multicopy suppressor of IRA1)	MSI type nucleosome	chromatin assembly	-
C-ELMS-18	123-436	+	Coding SSR	(Q)$_n$	-	NP_173356	NM_101780.2	3 e^{-13}	BlastX	ARF19 (Auxin Response Factor 11)	transcription factor	-	-
C-ELMS-19	90-460	-	Coding SSR	(N)$_n$	-	NP_197569	NM_122076.1	2 e^{-04}	BlastX	LRR transmembrane protein kinase	Protein kinase activity/receptor	Phosphorylation	-
C-ELMS-20	59-487	+	Coding SSR	(Q)$_n$	-	NP_567896	NM_119407.2	1 e^{-05}	BlastX	LUG (LEUNIG); Transcriptional corepressor	DNA binding	Development/cell differentiation	nucleus
C-ELMS-33	1-644	-	Coding SSR	(P)n	EL456618, *H. tuberosus*, 5 e^{-93}	NP_177361	NM_105874.3	1 e^{-29}	BlastX	SEC14 cytosolic factor/phosphoglyceride transfer family	transporter activity	transport	intracellular/membrane
C-ELMS-37	132-497	+	Coding SSR	(P)$_n$(LXXL)$_n$	-	NP_001053250	NM_001059785.1	1 e^{-20}	BlastX	LRR plant protein	Protein binding	-	-
C-ELMS-38	184-660	+	Coding SSR	(H)$_n$	DW060955, *L. sativa*, 3 e^{-40}	-	DW060955	1 e^{-38}	MegaBlast	*L. saligna* cDNA/WRKY DNA binding protein like	ATP-dependent DNA helicase	DNA repair	-
C-ELMS-44	1-138; 241-531	-	SSR between 2 ORF	-	-	NP_850630	NM_180299.1	3 e^{-18}	BlastX	MSH7 (Muts Homolog 6-2)	damaged/mism. DNA binding	mismatch repair	-
C-ELMS-47	1-802	+	Coding SSR	(K)$_n$	-	NP_001053494	NM_001060029	3 e^{-08}	BlastX	MIP, Major intrinsic protein family protein, aquaporin like	receptor activity/H$^+$ ion transporter	ATP synthesis	chloropl./thylak./PSI
C-ELMS-48	199-8	+	Coding SSR	(Q)$_n$	EH757759, *C. solstitialis*, 3 e^{-71}	-	EH757759.1	3 e^{-71}	MegaBlast	*C. solstitialis* cDNA clone	DNA topoisomerase	DNA change during replication	chromosome
C-ELMS-49	241-555	+	Coding SSR	(L)$_n$	CX944987, *H. annuus*, 1 e^{-29}	-	EV203711	4 e^{-28}	MegaBlast	Transcribed locus, *B. napus*	transcription factor	-	phloem
C-ELMS-52	212-24	+	SSR before ORF	-	EH739359, *C. maculosa*, 4 e^{-75}	NP_973993	NM_202264.2	1 e^{-31}	BlastX	Putative NMT2	methyltransferase	metabolic process	-
C-ELMS-57	82-182; 232-364	-	SSR after ORF	-	-	NP_177784	NM_106308.3	7 e^{-03}	BlastX	TF (Squamosa promoter-binding-like protein 16 - SPL16)	oxidoreductase	electron transport	intracellular
C-ELMS-60	426-565; 598-702	-	Coding SSR	(ER)$_n$	-	NP_171911	NM_100296.1	4 e^{-17}	BlastX	C2 domain containing protein	glycosyl transferase activity	-	-
C-ELMS-61	134-181; 309-381	-	SSR between 2 ORF	-	EH767288, *C. solstitialis*, 4 e^{-15}	-	EH767288	4 e^{-15}	MegaBlast	*C. solstitialis* cDNA clone	receptor	cell-matrix adhesion/signalling	integrin complex
Similarities with mobile elements													
C-ELMS-21	-	-	-	-	-	-	ATENSPM12	80%	Censor	EnSpm like Element (DNA transposon)	-	-	-
C-ELMS-23	-	-	-	-	-	-	RIRE7_I	71%	Censor	Ty3-Gypsy like Element (LTR Retrotransposon)	-	-	-
C-ELMS-24	-	-	-	-	-	-	CEREBA_I	69%	Censor	Ty3-Gypsy like Element (LTR Retrotransposon)	-	-	-
C-ELMS-25	-	-	-	-	-	-	SZ-6IN	66%	Censor	Ty1-Copia like Element (LTR Retrotransposon)	-	-	-
C-ELMS-26	-	-	-	-	-	NP_001046979	NM_001053514.1	4 e^{-10}	BlastX	Retrotransposon gag protein	RNA-directed DNA polymerase	integration	nucleus
C-ELMS-30	-	-	-	-	-	-	ATGP8	78%	Censor	Gypsy like Element (LTR Retrotransposon)	-	-	-
C-ELMS-34	-	-	-	-	-	-	TEMPINDAS	70%	Censor	hAT-like (DNA transposon)	-	-	-
C-ELMS-35	-	-	-	-	-	-	EnSpm5_OS	67%	Censor	EnSpm like Element (DNA transposon)	-	-	-
C-ELMS-36	-	-	-	-	-	-	NonLTR-5_CR	61%	Censor	Non LTR Retrotransposon like	-	-	-
C-ELMS-39	-	-	-	-	-	-	SHACOP23_MT	78%	Censor	LTR Retrotransposon like	-	-	-
C-ELMS-40	-	-	-	-	-	-	Copia40-PTR_I	90%	Censor	LTR Retrotransposon like	-	-	-
C-ELMS-41	-	-	-	-	-	NP_001061216	NM_001067751	1 e^{-10}	BlastX	Oryza sativa Copia protein	nucleic acid/zinc ion binding	DNA integration	-
C-ELMS-54	-	-	-	-	-	NP_194886	NM_119307.3	2 e^{-02}	BlastX	SRZ-22 (serine/arginine-rich 22)	nucleic acid/zinc ion binding	DNA integration/RNA splicing	-
C-ELMS-55	-	-	-	-	-	NP_001043197	NM_001049732.1	5 e^{-03}	BlastX	PDR-like ABC transporter (PDR3 ABC transporter)	nucleic acid/zinc ion binding	DNA integration/viral reprod.	-
C-ELMS-56	-	-	-	-	-	-	SHAMUDRAV_MT	70%	Censor	MuDr like DNA transposon	-	-	-

[(1)] The ORF found is less than 45 aminoacids but is present at the end of the CELMS locus

[(2)] When the CENSOR algorithm was used, a similarity value is reported

Table 4. Sequence of RT PCR primers targeted to ORF sequences.

Locus		Primer sequence (5'-3')[1]		Primer sequence (5'-3')[1]
C-ELMS-03	RT-for:	ATGGATGGTAACGTTGATATAGAATG	RT-rev:	CTCGTAATCAAGAGATTCGATTGG
C-ELMS-04	RT-for:	AATGGACAGATGACGGTGGT	RT-rev:	AGTCACCAGCAGCAGGCATA
C-ELMS-06	RT-for:	TCCTAATCAGGTGGCTGGAC	RT-rev:	CTTTGCCTCTTGGCAAACTC
C-ELMS-15	for		RT-rev:	TGGGGATCTTCGTGGTAATC
CELMS 18	RT-for:	AACCAAACAAACCAACTTGTGA	rev	
CELMS 20	RT-for:	CAACAGCTCATGTTGCAG	rev	
CELMS 33	RT-for:	TCACAACAAAAATCGCCTCA	RT-rev:	GACGACGTCGGTTCTTTCAT
CELMS 37	for		RT-rev:	TCAAACGCAAAGTGAAATCG
CELMS-48	RT-for:	AGAATGCGAAGGCGTCAAC	RT-rev:	TGACTTCAACCATGGTATCTTTG
C-ELMS-52	RT-for:	CCGCTCAAGAGCAAAAGAGA	RT-rev:	AATTTGCTGCACAGCTGGAT
C-ELMS-57	RT-for:	GCAGATGCGACCTCTGGT	RT-rev:	ATTCGTTTTGACACCCCAAC
C-ELMS-60	RT-for:	GGTGGGTATGGAAAGAAGACA	RT-rev:	GAAGAACGCGTGTGTTTCAC

[1] When the primer is omitted the original primer as reported in Table 2 has been used.

Figure 1 A) CELMS-14 profile of the germplasm panel; B) Segregation patterns for three CELMS loci among progeny of the cross C3 x Sp-9A

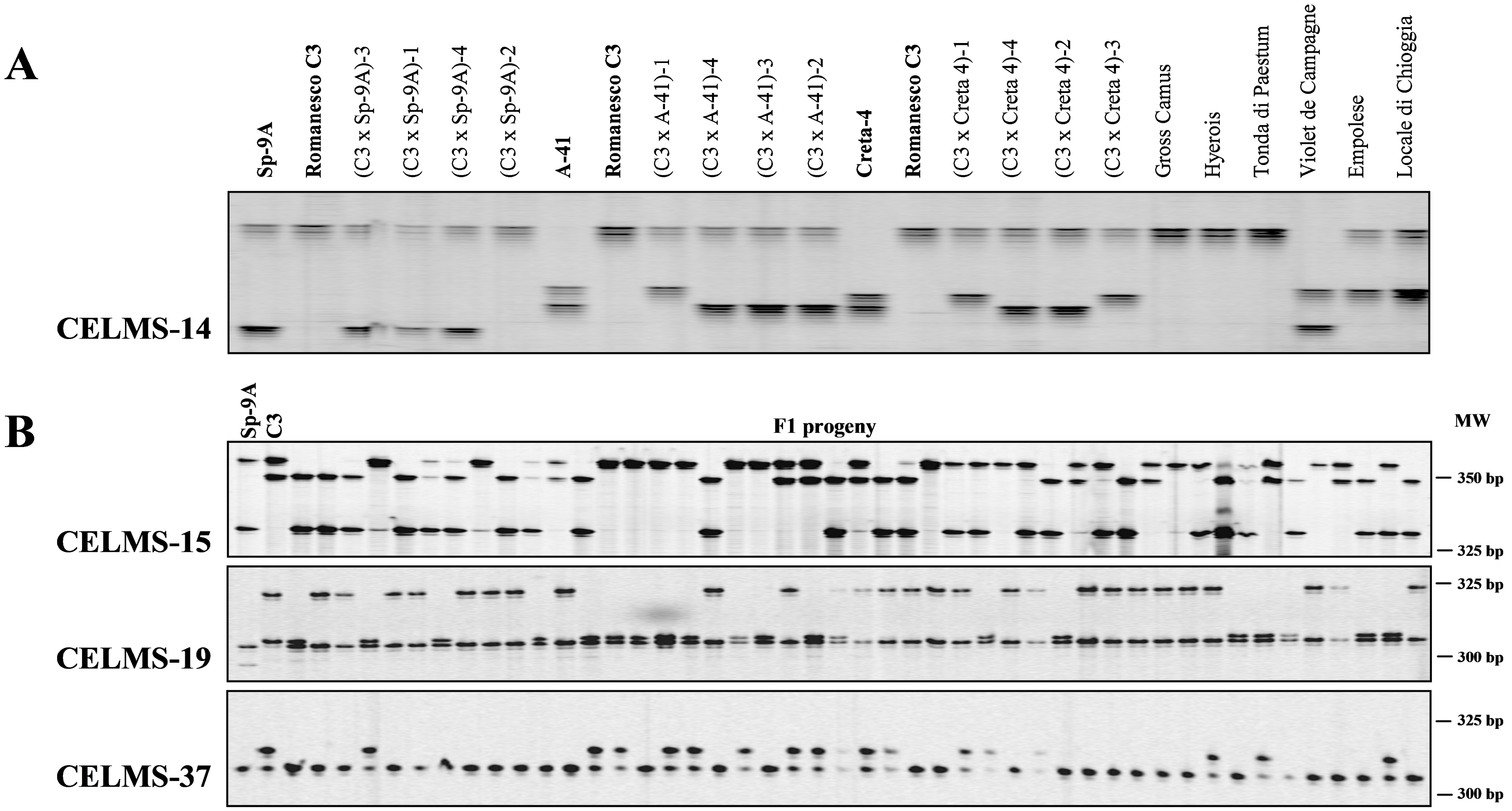


Figure 2 Genetic maps of C3 (female parent of mapping population, white LGs on the left) and Sp-9A (male parent, grey LGs on the right). The 35 mapped SSR loci are shaded light grey. Intercross markers are shown in *italics* and in bold; Aligned LGs are presented side-by-side. LG-7, -12, -13, -15, and -17 are not reported since they are not covered by CELMS markers. Markers showing significant levels of segregation distortion are indicated by *asterisks* (*: 0.1 >*P* ≥ 0.05, **: 0.05 >*P* ≥ 0.01).

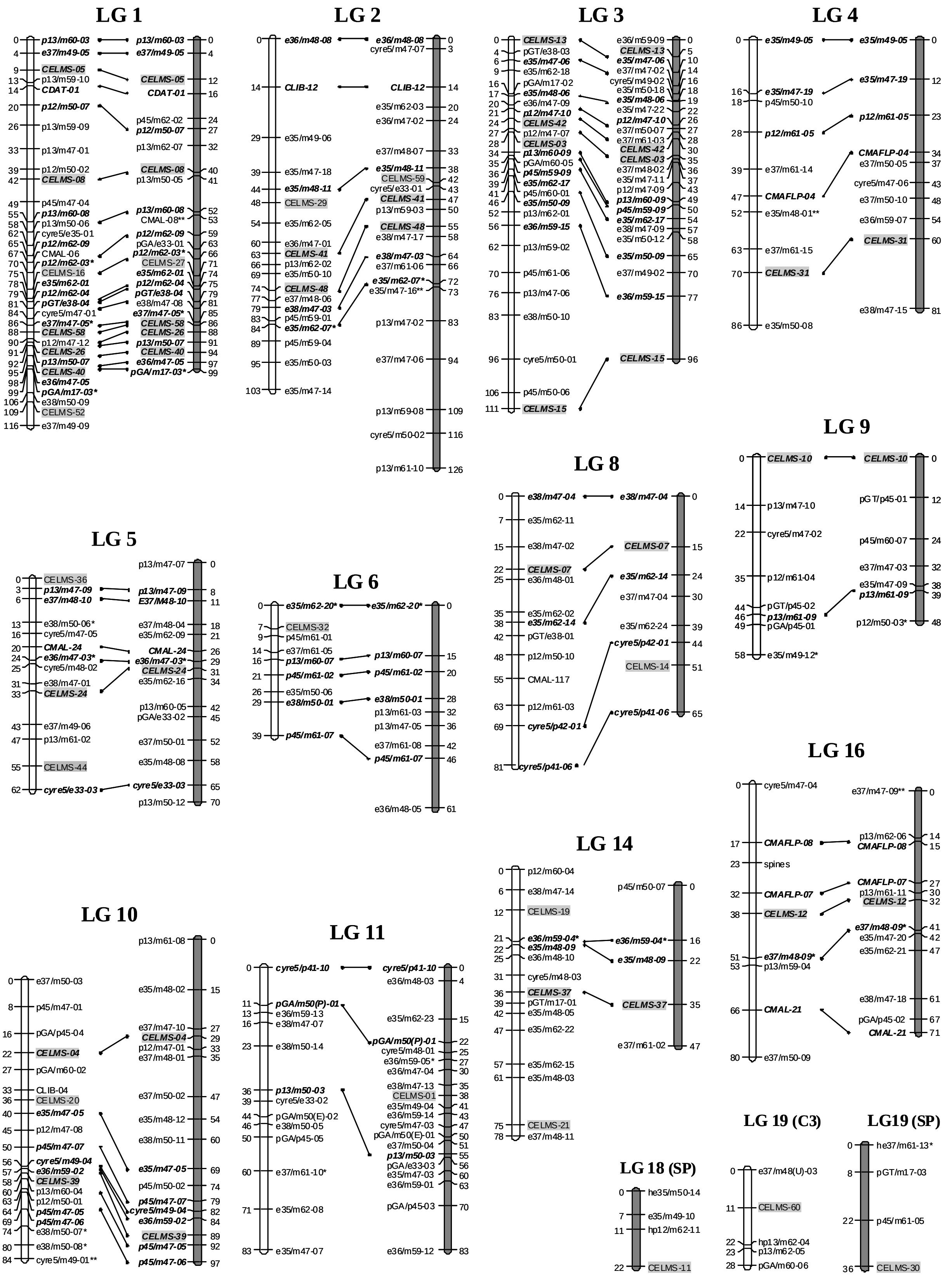
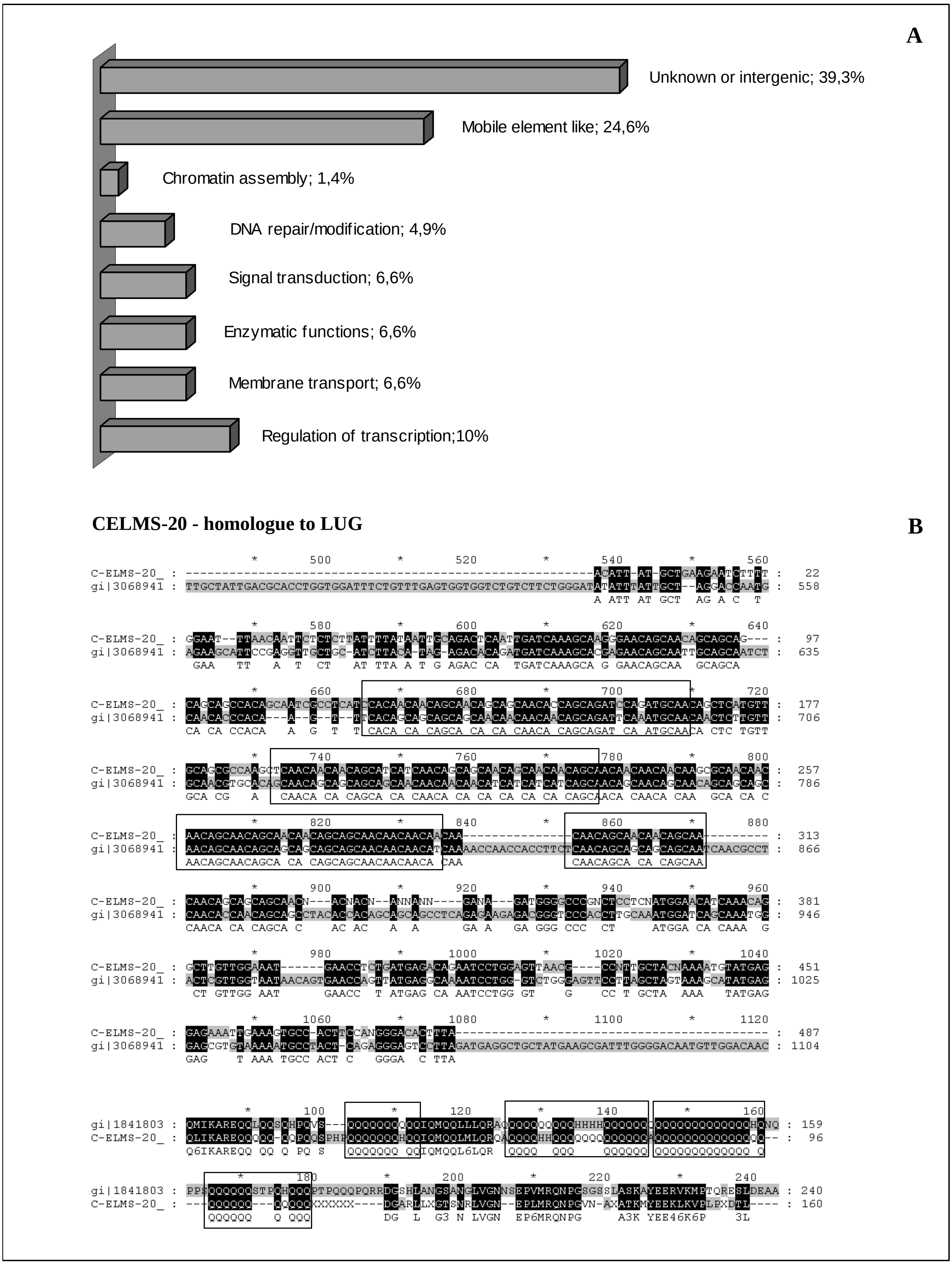


Figure 3 A) Gene ontology of CELMS loci, B) The relationship between a CELMS sequence and its *A. thaliana* homologue based on the alignment of both nucleotides and amino acids. Boxes show conserved motifs.


A
Unknown or intergenic; 39,3%
Mobile element like; 24,6%
Chromatin assembly; 1,4%
DNA repair/modification; 4,9%
Signal transduction; 6,6%
Enzymatic functions; 6,6%
Membrane transport; 6,6%
Regulation of transcription;10%
CELMS-20 - homologue to LUG
B
C-ELMS-20_
gi|3068941
gi|1841803