Identification and mapping of genes related to caffeoylquinic acid synthesis in Cynara cardunculus L.

Identification and mapping of genes related to caffeoylquinic acid synthesis in *Cynara cardunculus* L.

Barbara Menin, Cinzia Comino[§], Andrea Moglia, Yulia Dolzhenko, Ezio Portis, Sergio Lanteri

DIVAPRA Plant Genetics and Breeding, University of Torino 10095, Grugliasco (Torino), Italy

[§]Corresponding author:

Cinzia Comino

DIVAPRA, Plant Genetics and Breeding,

University of Torino,

Via L. da Vinci 44,

10095 Grugliasco (Torino), Italy.

Tel: +39-0116708811; Fax +39-0112368811

cinzia.comino@unito.it

Abstract

The caffeoylquinic acid (CQA) biosynthesis pathway in plants is not fully understood, but the BAHD superfamily, a large class of acyl-CoA-dependent acyltransferases, may represent an interesting group for the isolation of enzymes playing a key role. Globe artichoke (*Cynara cardunculus* var. *scolymus*) represents a relevant model species for studying CQA biosynthesis, because the plant accumulates high amounts and a diverse spectrum of CQAs in its leaves. An *in silico* scan of globe artichoke ESTs was based on a search string targeting the conserved motifs of BAHD proteins and complete BAHD sequences present in a range of species. EST hits were assembled into 32 unigenes and their phylogeny characterized. Among which were three putative BAHD acyltransferases showing a high level of similarity to globe artichoke HCT and HQT enzymes. The full-length cDNAs of these three proteins were isolated; two of these acyltransferases were able to synthesize both chlorogenic acid and *p*-coumaroylquinate, assessing their involvement in CQA biosynthesis. Upon exposure to UV-C radiation, which is known to induce CQA biosynthesis, expression of two of the enzyme transcripts was up-regulated while that of the third was unaffected. By analysing the globe artichoke unigene consensus sequences, further two genes (i.e. C4H, 4CL), related not only to flavonoids but also to chlorogenic acid (a mono-CQA) biosynthesis, were also identified and were genetically mapped along with the three acyltransferase genes. The research has improved the knowledge base surrounding globe artichoke BAHD members. The mapping of the genes related to the chlorogenic acid biosynthesis pathway may contribute to a greater understanding of the genetic basis of the synthesis of these compounds of nutraceutical and pharmaceutical value.

Keywords: acyltransferases, chlorogenic acid, gene mapping, globe artichoke

1. Introduction

Approximately two-thirds of the medicinal drugs developed over the last 25 years have been derived from, or inspired by natural products, in particular plant secondary metabolites [1, 2]. Plants produce at least $2x10^5$ distinct secondary metabolites, many of which are restricted to a group of species or genera, and some of these are of major nutraceutical and/or pharmaceutical importance [3, 4]. In particular, a number of plant phenolics can, synergistically or additively, provide protection against the damage induced by free radicals during oxidative stress, and reduce the risk of certain chronic diseases in human beings [5-7]. The health benefits of the class of phenolic compounds referred to as caffeoylquinic acids (CQAs) have attracted particular interest in recent years. Plant extracts containing a substantial CQA content exhibit hepatoprotective, anti-carcinogenic, anti-oxidative, anti-bacterial, anti-HIV, bile-expelling and urinative activity, as well as showing an ability to inhibit cholesterol biosynthesis and LDL oxidation [6, 8-11]. Polyphenolic extracts have also shown an *in vitro* apoptotic effect on human liver cancer cells [12]. Despite a significant research effort, however, the molecular basis of CQA biosynthesis remains only partially understood [13-15] (Fig. 1). The first steps of the pathway involve the enzymes phenylalanine ammonia-lyase (PAL), cinnamate-4-hydroxylase (C4H) and 4-coumarate: coenzyme A ligase (4CL), which together are responsible for the synthesis of *p*-coumaroyl-CoA. This compound is the substrate for CQA synthesis, mediated by shikimate (or quinate) acyltransferase and *p*-coumaroyl ester 3'-hydroxylase (C3'H) [13]. No enzyme able to add a second caffeoyl group to the chlorogenic acid structure leading to synthesis dicaffeoylquinic acid has yet been identified in plants rich in CQAs compounds.

Cynara cardunculus, a species within the *Asteraceae* (*Compositae*) family, includes the three taxa globe artichoke (var. *scolymus* L.), cultivated cardoon (var. *altilis* DC), and wild cardoon [var. *sylvestris* (Lamk) Fiori]. Globe artichoke produces notable quantities of CQAs; up to 6% of weight of its dry leaves can be composed of monocaffeoylquinic (such as chlorogenic acid) and dicaffeoylquinic acids [10, 16]. Some of the genes underlying the biosynthesis of CQA in globe

artichoke have been identified and characterized. These include three PAL genes [17], two hydroxycinnamoyltransferases (HCT and HQT) and one hydroxylase (C3'H). Enzymatic assays have confirmed the involvement of all these genes in CQA synthesis [18-20]. HCT and HQT belong to the BAHD superfamily, a large class of acyl-CoA-dependent acyltransferases, which displays a range of specificities for acyl acceptors [21-25]. A growing number of BAHD family members across a range of plant species has now been characterized [25]. The primary nucleotide sequence of BAHD genes seems to be variable, with homology at the peptide level lying in the range 15-30%. However, all appear to encode a monomeric enzyme of molecular mass between 48 and 55 kDa. BAHD genes are assumed to have evolved from a common ancestral sequence, since most share the two highly conserved motifs HXXXD and DFGWG [22, 24]. The HXXXD domain lies in the central portion of the protein. It has been shown experimentally that the histidine residue is necessary for full enzymatic activity, and that a mutation to the asparagine residue results in complete loss of activity [21]. The DFGWG motif, located near the C-terminus of the protein, is important for catalytic activity or binding to coenzyme A [26, 27]. In *Medicago truncatula*, analysis of a truncated malonyltransferase gene has suggested that the DFGWG motif acts to stabilize protein conformation under physiological conditions, and may also be responsible for ensuring correct sub-cellular localization [28].

The aim of the present study was to gather more information on genetic determinism (mapping analysis) as well as biosynthetic pathway (gene isolation) of caffeoylquinic acid synthesis in globe artichoke. In order to reach these goals we have exploited the conservation of the DFGWG motif and several known acyltransferase sequences from other species to identify BAHD family genes within a set of 19,000 globe artichoke unigenes (made available by the Compositae Genome Project, http://compgenomics.ucdavis.edu/). The full length versions of three of these sequences were determined in order to both test their enzymatic activity and to assess whether their expression can be induced by UV-C radiation, which has been shown to trigger the accumulation of CQAs in globe artichoke leaves [30]. We have used the unigene consensus sequences to identify C4H and

4CL genes related to chlorogenic acid biosynthesis, and have localized all these genes on the emerging *C. cardunculus* genetic maps [30].

2. Materials and Methods

2.1. In silico analysis of ESTs

The conserved BAHD superfamily DFGWG motif, along with the sequences of known acyl-CoA-dependent acyltransferases, was used as a search string to identify putative acyltransferases among a 19,055 unigene set [31]. The BLAST matrix was set as pam100 (with the expectation at 100) for searching gene fragments, and as blosum62 (with the expectation at 10) for full-length sequences. The InterProScan tool [32] (available within the Blast2GO software package [33]), which scans given protein sequences for the presence of particular domains, regions, repeats and other identifiable features of protein families available in database, was used.

Moreover, several known genes from other species involved in chlorogenic acid biosynthesis (i.e. C4H, 4CL) were used to investigate the set of globe artichoke unigene consensus assembled by Scaglione *et al.* [31].

Phylogenetic analysis was conducted by using all putative BAHD acyltransferases of *C. cardunculus* available in our unigene set (reported in Table 1) together with the function-known BAHD sequences from other species that were characterized biochemically or genetically to guide our inference of phylogenetic relationships. The acyltransferases sequences used are listed in Figure 2 legend. The phylogenetic analysis was conducted using MEGA v3.0 software [34].

2.2. Plant material, RNA extraction and cDNA synthesis

Leaves of globe artichoke were collected from the field at Scalenghe, Torino (Italy). Total RNA was extracted from 100 mg fresh tissue using the "Trizol" reagent (Invitrogen, USA), following the manufacturer's instructions. RNA concentration was determined by spectrophotometry, and its integrity assessed by electrophoresis through 1% (w/v) formaldehyde-agarose gels [35]. Reverse transcription was primed with poly(dT) and achieved using M-MuLV RNaseH-RT (Finnzymes, Finland), following the manufacturer's instructions.

2.3. Isolation, heterologous expression in E. coli and the assessment of in vitro enzymatic activity

Primers based on the partial cDNA sequence were designed and used to obtain full-length cDNA sequences by means of 5' and 3' RACE PCR approach as described elsewhere [18]. The ORF of the putative acyltransferases was amplified from cDNA template using primers designed both to anneal to the 5'- and 3'- ends of the transcript, and to generate restriction sites for subsequent cloning. Primer sequences are detailed in Additional file 1. The cDNA amplicons were cloned into a pCR®2.1 vector (Invitrogen, USA) and sequenced by BMR genomics (http://www.bmr-genomics.it/). After restriction digestion, the fragments were ligated into the cloning site of a digested pET28b plasmid (Novagen, USA), and the resulting recombinant plasmid was transformed into *E. coli* strain BL21(DE) pLysE. Selection of transformants was achieved by growing the cells in Luria-Bertani (LB) broth agar supplemented with 50 mg/l kanamycin, 34 mg/l chloramphenicol and 1% w/v glucose. Individual recombinant colonies were inoculated into 4ml overnight culture in the same medium, and 2ml of this culture was then inoculated into 50 ml LB liquid medium and incubated for 3 h at 28°C with shaking (200 rpm). Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1mM and the cultures were grown for further 18h at 22°C. After centrifugation (5000 g for 10 min), the pellet was re-suspended in 1ml phosphate-buffered saline pH 7.5 and lysed by three cycles of freezing in liquid nitrogen and thawing at 37°C, followed by three 30 s sonication bursts on ice. The sonicated cells were centrifuged (14,000 g for 5 min at 4°C), and the protein content of both the supernatant and pellet profiled by SDS-PAGE (8% resolving gel, 5% stacking gel) using Coomassie brilliant blue staining [35]. Negative controls used comparable preparations harbouring an empty vector.

The substrates for the enzyme assays were chlorogenic acid (Sigma-Aldrich), quinic acid (Fluka, Switzerland), cynarin (1,3-dicaffeoylquinic acid, from Carl Roth, Karlsruhe, Germany) and 1,5-dicaffeoylquinic acid (Chengdu Biopurify Phytochemicals Ltd, Chine) while the necessary CoA esters were obtained from TransMIT (Marburg, Germany). Each 30 µl reaction mixture contained

100 mM phosphate buffer (pH7.5), 1mM dithiothreitol, 50-1,000 ng protein, and the substrates (p-coumaroyl-CoA, caffeoyl-CoA, quinic acid) at concentrations ranging from 0.1-5 mM. The conversion of chlorogenic acid to di-CQAs was tested by incubating 50-1,000 ng protein in the presence of 100 μM chlorogenic acid and 100 μM caffeoyl CoA. For assay with di-CQAs, 500 μM cynarin or 500 μM 1, 5 dicaffeoylquinic acid and 500 μM CoA were used. Reactions were incubated at 30°C for 30 min, followed by the addition of 30 μl acetonitrile:HCl (99:1). The products were analysed by reverse-phase HPLC. HPLC was performed using Varian 920-LC system and chromatograms were electronically stored and evaluated using Galaxie Software. The separation was done using an analytical Luna C18 (2) (2mm x 150mm, 100 Å, particle size 3 μM) column, along with a 2mm x 4mm pre-column (Phenomenex). The two solvents were 90% water, 9.9% CH3CN, 0.1% HCOOH and 80% CH_3CN, 19.9% water, 0.1% CH_3COOH. The percentage of the latter solvent reached 60% over a 15min run time, and 100% after 28 min. The flow rate was 0.5 mL/min, the injection volume was 10 μL, and the detection wavelength was recorded at 240-600 nm. Each reaction product was identified by comparing its retention time and absorbance spectrum with authentic samples or isolated compounds previously characterized.

2.4. UV-C stress induced expression of acyltransferase genes

Three globe artichoke foliar discs were exposed to UV-C treatment (16 W germicidal lamp, 20 min) as described elsewhere [29] and ground in liquid nitrogen to a fine powder. RNA extraction was performed from 100 mg of each powdered foliar disc, as described above Gene-specific primers (listed in Additional file 1) for qRT-PCR were designed using Primer 3 software (http://frodo.wi.mit.edu/primer3/), and tested for their specificity by means of BLAST analysis. Actin was chosen as the constitutively expressed reference gene, and was amplified using primers RT Act For and RT Act Rev (see Additional file 1) designed from the globe artichoke actin sequence (ACT, AM744951). The first-strand cDNA was synthesised using an iScript cDNA Synthesis kit (Bio-Rad), following the manufacturer's instructions. The cDNA was diluted to obtain

a threshold cycle (CT) value of between 25 and 35. Each 20 µl qRT-PCR, performed in triplicate for each sample, contained 1x iQ Supermix, 1x SYBR-Green I (iQTM, SYBR® GreenSupermix), 300 nM gene-specific primers and 3 µl diluted cDNA. The reactions were carried out in 48-well optical plates using the iCycler Real-time PCR Detection System (Bio-Rad Laboratories, USA). The PCR conditions comprised an initial incubation of 95°C for 5 min, followed by 35 cycles of 95°C for 15 s and 60°C for 60 s. Melting curve analysis was performed at the end of amplification process. Standard curves were analysed using iCycler iQ software, and amplicons by the comparative threshold cycle method, in which ΔΔCt was calculated from ΔCtI- ΔCtM. The last values were, respectively, the Ct values for, the target genes in UV-C exposed leaves (I) and in the non irradiated control leaves, (calibrator, M), both normalized against actin.

2.5. SNP detection and linkage analysis

Sequences of each target gene were amplified from 'Romanesco C3' and 'Altilis 41', parental genotypes previously used for maps development [30], and the amplicons directly sequenced. Single nucleotide polymorphism genotyping was carried out using the tetra-primers ARMS-PCR method [36] (primer sequences are listed in Additional File 1), designed using the software made available on-line (http://cedar.genetics.soton.ac.uk/public_html/primer1.html). Marker segregation data were combined with those used to construct the base genetic maps [30]. The markers were grouped as either maternal (segregating only within 'Romanesco C3') or paternal (segregating only within 'Altilis 41'), with an expected segregation ratio of 1:1; or intercross markers (segregating in both parents), with an expected segregation ratio of 1:2:1. Independent framework linkage maps were constructed on the basis of the double pseudo-testcross mapping strategy using JoinMap v4.0 [37]. Goodness-of-fit between observed and expected segregation was assessed using the χ^2 test. For both maps, linkage groups (LGs) were established on the basis of an initial threshold LOD of 6.0, applying the parameters Rec=0.40, LOD=1.0, Jump=5. Map distances were converted to cM using the Kosambi mapping function [38].

3. Results

3.1. The in silico identification of BAHD acyltransferases

A set of 51 ESTs containing the conserved motifs of BAHD family protein was identified using full-length sequences of acyltransferases with known function, and assembled in 32 unigenes (Table 1). Thirteen of these unigenes belonged to tentative consensus sequences (TCs, each containing 2-7 ESTs), and the remaining 19 were singletons (Table 1). Applying the InterProScan tool to the overall set of ESTs, the presence of the Pfam 02458 domain, which is a characteristic of the BAHD family of plant CoA-dependent acyltransferases, was highlighted in all these sequences [22]. The sequence of the CCPU1541 singleton showed 99% of similarity to globe artichoke HCT [GenBank: AAZ80046], while that of TC CL4475 was 98% similar to globe artichoke HQT [GenBank: ABK79689]. A guided phylogenetic tree was constructed to study the relationship of BAHD superfamily proteins in *C. cardunculus* together with biochemically characterized acyltransferases from other species. On the whole six clades were identified (Fig. 2). Clade A contained five TC and two singleton unigenes, and can be further divided into sub-clades: A-1, which captures five sequencing clustering with enzymes that use hydroxycinnamoyl/benzoyl CoA as acyl donor [13, 18, 19, 39], and A-2, grouping the two remaining with the enzymes involved in volatile ester biosynthesis [40]. Clade B comprised one unigene related to barley agmatine coumaroyltransferase (ACT), an enzyme which catalyzes the first step in the biosynthesis of anti-fungal hydroxycinnamoylagmatine derivatives [41]. Clade C comprised unigenes CL4742 and CCPU1786 along with acyltransferases which utilize a range of alcohol substrates (mainly acetyl-CoA as acyl donor) [24]. Ten unigenes clustered into clade D, defined by *Glossy2* and *CER2* homologues in *Zea mays* [42] and *A. thaliana* [43, 44]. These enzymes are involved in the extension of long chain epicuticular waxes, which are important both for restricting water loss and for defence against pathogens. Clade E comprised three unigenes along with enzymes involved in the modification of phenolic glucosides, predominantly anthocyanins [45]. The remaining seven unigenes were grouped together in Clade F which lacked any functionally defined homologue.

3.2. Full length isolation, heterologous expression in E. coli and in vitro enzymatic activity of BAHD acyltransferases

The clade A unigenes CL7000, CL7001, CL3797 and CCPX9719 were targeted for full-length cDNA isolation, since they were the sequences most closely related to other transferases, like HCT and HQT, involved in CQA biosynthesis. Their predicted translation products [based on the sequence of their open reading frames (ORFs)] were named Acyltransf_1, Acyltransf_2, Acyltransf_3 and Acyltransf_4, respectively. Three had a predicted molecular weight of ~50 kDa and contained the HXXXD motif, while the DFGWG motif was present in only Acyltransf_2 and _3; in Acyltransf_1, the variant DLGWG was present (Fig. 3). The Acyltransf_4 polypeptide was only ~18 kDa in size, and lacked both characteristic BAHD domains; for this reason it was not considered in any of the subsequent analyses. Unexpectedly short polypeptides have also arisen in similar *in silico* scans of *A. thaliana* ESTs [25]. Amino acid sequence alignment revealed that Acyltransf_1 [GenBank: GU248357] and _2 [GenBank: GU248358] shared ~86% similarity and were closely related to globe artichoke HQT (~70% identity and 80% similarity). In contrast, Acyltransf_3 [GenBank: GU248359] was only ~50% similar to either Acyltransf_1 or _2, and instead was closely related to a member of the *A. thaliana* transferase family (AT5G41040).

When the Acyltransf_1, _2 and _3 full-length cDNAs were transformed into *E. coli*, SDS PAGE analysis was able to detect their heterologous expression as proteins of size ~50 kDa (Fig. 4a).

The recombinant enzymes were incubated in the presence of *p*-coumaroyl-CoA or caffeoyl-CoA and quinic acid or shikimic acid as substrates, and the products of the reactions were analyzed by HPLC. In the presence of Acyltransf_1 and _2, new products (*p*-coumaroylquinate and caffeoylquinate) were identified in the reaction mixtures containing *p*-coumaroyl-CoA or caffeoyl-CoA as acyl donor and quinic acid as acceptor (Fig. 4b). A very low activity was detected after addition of shikimic acid instead of quinic acid in the reaction mixture. No significant peaks were identified when these reactions were performed with Acyltransf_3. To investigate whether

Acyltransf_1, _2 and_3 were also involved in di-CQAs production, the enzymes were assayed in hypothetical direct (chlorogenic acid and Caffeoyl-CoA) and reverse (both 1,3-dicaffeoylquinic acid or 1,5-dicaffeoylquinic acid with CoA) reactions, but no activity with these substrates was detected in our experimental conditions.

All these reactions were also performed with the control crude extract (*E. coli* extract obtained from empty vector transformed cells) and no products could be detected.

3.3. UV-C stress inducible expression of acyltransferase genes

The expression profile in the globe artichoke leaf of Acyltransf_1, _2 and _3 in response to UV-C exposure was assessed using real time quantitative PCR (qRT-PCR). Comparison between the standard curves for each enzyme revealed a correlation coefficient of >0.98 and an efficiency (slope of the curve) >0.90 (data not shown). This experiment showed that Acyltransf_1 and _2 transcription was about 2 fold higher upon the UV-C treatment, while that of Acyltransf_3 was slightly decreased (Fig. 5).

3.4. Mapping of genes related to chlorogenic pathway

The set of globe artichoke unigene consensus assembled by Scaglione *et al*. [31] were also investigated for C4H, and 4CL homologues and three sequences showed a high degree of similarity to these genes related to chlorogenic acid biosynthesis. Unigene CL1082 contains a full length cDNA of a cinnamate-4-hydroxylase (C4H) which shows an high level of similarity (92% identity and 96% similarity) with the C4H gene of *Heliantus tuberosus* [46]. CL3632 and CL5650 were both partial sequences homologous to, respectively, 4CL of *Nicotiana tabacum* [GenBank: AAB18637].

The three previously described acyltransferases together with the C4H and 4CL homologues of the parents of our mapping population (globe artichoke ‘Romanesco C3’ and cultivated cardoon ‘Altilis 41’) [30] were re-sequenced for assessing their polymorphisms. This resulted in the identification of

five single nucleotide polymorphisms (Acyltransf_1-snp, Acyltransf_2-snp, Acyltransf_3-snp, C4H-snp and 4CL-snp). Both parents were heterozygous for Acyltransf_2-snp, and this generated a 1:2:1 segregation in the population (Table 2 and Fig. 6), allowing the gene to be incorporated into both the female and male maps (LG C3_2 / Alt_4, Fig. 7). 'Romanesco C3' (but not 'Altilis 41') was heterozygous for both Acyltransf_1-snp and C4H-snp, delivering a segregation ratio of 1:1 (Table 2, Fig. 6) and locating the two genes on the maternal map on LGs C3_8 and C3_2, respectively (Fig. 7). Acyltransf_3-snp and 4CL-snp were all heterozygous in 'Altilis 41' (but not in 'Romanesco C3; Table 2 and Fig. 6). As a result, Acyltranf_3 was mapped to a distal region of LG Alt_3, while 4CL united a number of previously unlinked markers to form a new linkage group (LG Alt_18) (Fig. 7). Figure 7 also records the LGs within which other genes (PAL, HCT, HQT and C3'H) in the CQA biosynthesis pathways (from phenylalanine to chlorogenic acid) have already been mapped [30].

4. Discussion

The optimization of plant secondary metabolite production will require a firm understanding of the relevant biosynthetic pathways. Thanks to its exceptional content of CQAs, the globe artichoke represents a particularly good model species for this purpose. We have previously characterized the two key hydroxycinnamoyl-CoA transferases (HCT and HQT) involved in the synthesis of chlorogenic acid [18, 19] in globe artichoke, and the identification of critical acyltransferases along with other genes related to the biosynthesis of CQAs pathway adds to our ability to better understand the mechanism of CQA synthesis. It is not as yet certain whether the di-CQAs are derived directly from the mono-CQAs (such as chlorogenic acid), but their structural similarity provides good *a priori* evidence for this. Our current understanding of this metabolic pathway suggests the presence of an enzyme able to add a second caffeoyl group to the chlorogenic acid structure, but no such enzyme has as yet been identified in either globe artichoke or in any other plants rich in these compounds, such as tomato and coffee [47]. The important role the acyltransferases play in the CQAs biosynthetic pathway prompted us to focus our analysis on this class of enzymes.

Acyltrasferase enzymes belong to the BAHD superfamily [21, 22], whose members are distinguished by the presence of the two highly conserved HXXXD and DFGWG motifs. Our search of the globe artichoke unigene set, assembled from >36,000 EST sequences (Compositae Genome Project), allowed us to putatively identify a number of BAHD family members, all of which shared the Pfam 02458 domain typical for BAHD family CoA-dependent acyltransferases [22].

The BAHD acyltransferases of several plant species have been phylogenetically arranged into five major clades, based on the type of the substrate involved or the conditions under which the enzymes are expressed [24]. Since phylogenies of peptide sequences have been used to predict likely substrates for enzymes which have not been biochemically characterized [42-44], we chose a set of 16 acyltransferases belonging to each of the five clades reported by D'Auria *et al.* [24] to

perform a phylogenetic analysis of the globe artichoke unigenes. The topology of the guided phylogenetic tree produced by comparing the globe artichoke acyltransferases with one another and with a representative set of heterologous acyltransferases is reminiscent of that of the tree constructed using characterized acyltransferases from a range of plant species [24, 48]. Since Clade A1 contains enzymes using hydroxycinnamoyl/benzoyl CoA as acyl donors and shikimic and quinic acids as acceptors, this represents the most interesting group of enzymes involved in the CQA biosynthetic pathway in globe artichoke. Three subclades A1 members have been chosen for full-length isolation, mainly on the basis of their high similarity with the already functionally characterized hydroxycinnamoyltransferase sequences (HCT and HQT). The resulting ORFs encode respectively for three proteins characterized by a molecular weight of about 50 kDa (Fig. 4a); as previously reported by D'Auria *et al*. [24], in agreement with the molecular mass calculated for the BAHD members. Although phylogenies can give clues to function, in families of proteins like BAHD, with very versatile catalytic specificities, function needs to be established on a biochemical demonstration of activity. Of the three candidate sequences taken forward for characterization, two (Acyltransf_1 and _2) proved able to use either *p*-coumaroyl-CoA or caffeoyl-CoA as an acyl donor and quinic acid as an acceptor, showing that they are most likely involved in CQA biosynthesis, while the third (Acyltransf_3) did not recognize any of these molecules, and is therefore unlikely to be a component of CQA synthesis. The three isolated acyltransferases were also tested for their ability to add or remove a second caffeoyl group in the reactions (1) chlorogenic acid and caffeoylCoA and (2) cynarin (1,3-dicaffeoylquinic acid) or 1,5-dicaffeoylquinic acid with CoA. Since Acyltransf_1, _2 and _3 did not show any activity in these reactions we assumed that they have not role in the di-CQAs biosynthesis.

Leaf CQA content, as for most phenylpropanoid compounds, rises when the plant encounters abiotic stresses [49]. We found out that exposure to UV-C radiation leads to large increase of di-CQAs present in the leaves of different globe artichoke genotypes [29], and induces the expression of the genes encoding C3′H, HCT and HQT [19, 20]. In the present work we

observed that the UV-C treatment differentially affects the newly isolated acyltransferases, being the transcription of Acyltransf_1 and _2 up-regulated, while that of Acyltransf_3 unresponsive.

The UV-C induced expression of Acyltransf_1 and Acyltransf_2 together with the detected synthesis of quinate esters provides a strong indication of their putative involvement in CQAs biosynthesis. Side by side, the lack of a significant increase in chlorogenic acid accumulation [29] might be a consequence of its rapid conversion to di-CQAs through an unknown downstream enzymatic step.

The presence of three genes encoding HQT like proteins in globe artichoke might be due to the various independent separations of them during plant evolution; this can ensure a high accumulation of chlorogenic acid through the action of different biosynthetic enzymes.

Genetic mapping of the genes active within a particular biosynthetic pathway can help elucidate the complexity of plant secondary metabolism. In previous studies genes (i.e. PAL, HCT, HQT and C3'H) involved in CQAs biosynthesis pathway (from phenylalanine to chlorogenic acid) have already been mapped [30]. The development of DNA-based markers for the isolated genes responsible for chlorogenic acid synthesis has allowed us to locate them within the developing genetic maps of globe artichoke [30], to the extent that the entire biosynthetic pathway leading to the production of chlorogenic acid has now been genetically mapped. The integration of these gene-based markers has improved the precision of marker order and reduced inter-marker distances on some LGs. Their incorporation has changed the ordering of some loci slightly and caused some small shifts in intermarker distances; it has also succeeded in filling some of the pre-existing gaps in the map (especially on LGs Alt_3 and Alt_4). In addition, we have been able to increase the number of bridging markers by one, associating LG C3_2 with LG Alt_4, and to define new LGs in the male map by linking 4CL-snp with a set of previously unlinked loci. We are currently performing linkage analyses based on a large set of recently developed EST-SSRs [31] to build a functional genetic map.

In conclusion, our research targeting the globe artichoke acyltransferases has identified a number of BAHD family members, selected candidate acyltransferase genes involved in chlorogenic acid biosynthesis, and has genetically positioned all the actually known genes involved in the CQA biosynthetic pathway. Future research activities will be focused on the analysis of the "*in vivo*" role of the acyltransferases by means of forward genetic approaches. Furthermore the genetic placement of characterized gene-derived markers, such as the ones described here, may contribute in identifying QTLs involved in CQAs accumulation through the analyses of appropriate *C. cardunculus* segregating population, we already developed.

*This manuscript was under revision when the paper by Sonnante *et al*. ('Novel hydroxycinnamoyl-CoA quinate transferase (hqt) genes from artichoke are involved in the synthesis of chlorogenic acid') has been published on-line on April 29, 2010 in the journal 'Plant Physiology'.

Acknowledgements

We are grateful to Prof. Giovanni Mauromicale and Dr. Rosario Mauro (D.A.C.P.A. Agronomical Sciences, University of Catania) for providing the *C. cardunculus* parental accessions and the F_1 individuals.

This work was financially supported by Italian Ministry of Education, University and Research.

References

[1] F. Bourgaud, A. Gravot, S. Milesi, E. Gontier, Production of plant secondary metabolites: a historical perspective, Plant Science 161 (2001) 839-851.

[2] A. Boudet, Evolution and current status of research in phenolic compounds, Phytochem. 68 (2007) 2722-2735.

[3] E. Pichersky, D. Gang, Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective, Trends Plant Science 5 (2000) 439-445.

[4] N. Dudareva, E. Pichersky, Metabolic engineering of plant volatiles, Curr. Opinion Biotech. 19 (2008) 181-189.

[5] C. RiceEvans, J. Miller, G Paganga, Antioxidant properties of phenolic compounds, Trends Plant Science 2 (1997) 152-159.

[6] R. Gebhardt, Antioxidative and protective properties of extracts from leaves of the artichoke (*Cynara scolymus* L) against hydroperoxide-induced oxidative stress in cultured rat hepatocytes, Toxicol. Appl. Pharmacol. 144 (1997) 279-286.

[7] S. La Camera, G. Gouzerh, S. Dhondt, L. Hoffmann, B. Fritig, M. Legrand, T. Heitz, Metabolic reprogramming in plant innate immunity: the contributions of phenylpropanoid and oxylipin pathways, Immunol. Reviews 198 (2004) 267-284.

[8] T. Adzet, J. Camarassa, C.J. Laguna, Hepatoprotective activity of polyphenolic compounds from *Cynara scolymu*s against CCl4 toxicity in isolated rat hepatocytes, J. Natural Products 50 (1987) 612–617.

[9] J. Brown, C. Rice-Evans, Luteolin-rich artichoke extract protects low density lipoprotein from oxidation in vitro, Free Rad. Research 29 (1998) 247-255.

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

[11] B. Yang, Z. Meng, J. Dong, L. Yan, L. Zou, Z. Tang, G. Dou, Metabolic profile of 1,5-dicaffeoylquinic acid in rats, an in vivo and in vitro study, Drug Metab.Dispos. 33 (2005) 930-936.

[12] S. Miccadei, D. Di Venere, A. Cardinali, F. Romano, A. Durazzo, M. Foddai, R. Fraioli, S. Mobarhan, G. Maiani, Antioxidative and apoptotic properties of polyphenolic extracts from edible part of artichoke (*Cynara scolymus* L.) on cultured rat hepatocytes and on human hepatoma cells, Nutr. Cancer-An Intern. J. 60 (2008) 276-283.

[13] L. Hoffmann, S. Maury, F. Martz, P. Geoffroy, M. Legrand, Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism, J. Biol. Chem. 278 (2003) 95-103.

[14] R. Villegas, M. Kojima, Purification and characterization of Hydroxycinnamoyl D-Glucose: quinate hydroxycinnamoyl transferase in the root of sweet potato, *Ipomoea batatas* lam, J. Biol. Chem. 261 (1986) 8729-8733.

[15] R. Niggeweg, A. Michael, C. Martin, Engineering plants with increased levels of the antioxidant chlorogenic acid, Nat. Biotech. 22 (2004) 746-754.

[16] V. Lattanzio, A. Cardinali, D. Di Venere, V. Linsalata, S. Palmieri, Browning phenomena in stored artichoke (*Cynara scolymus* L.) heads: enzymic or chemical reactions?, Food Chem. 50 (1994) 1-7.

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

[18] C. Comino, S. Lanteri, E. Portis, A. Acquadro, A. Romani, A. Hehn, R. Larbat, F. Bourgaud, Isolation and functional characterization of a cDNA coding a

hydroxycinnamoyltransferase involved in phenylpropanoid biosynthesis in *Cynara cardunculus* L, BMC Plant Biol. *7 (*2007) *14*.

[19] C. Comino, A. Hehn, A. Moglia, B. Menin, F. Bourgaud, S. Lanteri, E. Portis, The isolation and mapping of a novel hydroxycinnamoyltransferase in the globe artichoke chlorogenic acid pathway, BMC Plant Biol. 9 (2009) 30.

[20] A. Moglia, C. Comino, E. Portis, A. Acquadro, R. De Vos, J. Beekwilder, S. Lanteri, Isolation and mapping of a C3'H gene (CYP98A49) from globe artichoke, and its expression upon UV-C stress, Plant Cell Rep. 28 (2009) 963-974.

[21] B. St.Pierre, P. Laflamme, A. Alarco, V. De Luca, The terminal O-acetyltransferase involved in vindoline biosynthesis defines a new class of proteins responsible for Coenzyme A dependent acyl transfer, Plant J. 14 (1998) 703-713.

[22] B. St-Pierre, V. De Luca, Evolution of acyltransferase genes: origin and diversification of the BAHD superfamily of acyltransferases involved in secondary metabolism, in: J. Romeo, R. Ibrahim, L. Varin, V. De Luca (Ed.), Recent Advances in Phytochemistry, Elsevier, Amsterdam, 2000, pp. 285-315.

[23] T. Nakayama, H. Suzuki, T. Nishino, Anthocyanin acyltransferases: specificities, mechanism, phylogenetics, and applications, *J. Mol. Catal. B-Enzym.* 23 (2003) 117-132.

[24] J. D'Auria, Acyltransferases in plants: a good time to be BAHD, Curr. Opin. Plant Biol. 9 (2006) 331-340.

[25] X. Yu, J. Gou, C. Liu, BAHD superfamily of acyl-CoA dependent acyltransferases in Populus and Arabidopsis: bioinformatics and gene expression, Plant Mol. Biol. 70 (2009) 421-442.

[26] H. Suzuki, T. Nakayama, T. Nishino, Proposed mechanism and functional amino acid residues of malonyl-CoA : anthocyanin 5-O-glucoside-6 "'-O-malonyltransferase from flowers of *Salvia splendens*, a member of the versatile plant acyltransferase family, Biochem. 42 (2003) 1764-1771.

[27] A. Bayer, X. Ma, J. Stockigt, Acetyltransfer in natural product biosynthesis - functional cloning and molecular analysis of vinorine synthase, Bioorg. Med. Chem. 12 (2004) 2787-2795.

[28] X.H. Yu, M.H. Chen, C.J. Liu, Nucleocytoplasmic-localized acyltransferases catalyze the malonylation of 7-O-glycosidic (iso)flavones in *Medicago truncatula*, Plant J. 55 (2008) 382-396.

[29] A. Moglia, S. Lanteri, C. Comino, A. Acquadro, R. de Vos, J. Beekwilder, Stress-induced biosynthesis of dicaffeoylquinic acids in globe artichoke, J. Agric. Food Chem. 56 (2008) 8641-8649.

[30] E. Portis, G. Mauromicale, R. Mauro, A. Acquadro, D. Scaglione, S. Lanteri, Construction of a reference molecular linkage map of globe artichoke (*Cynara cardunculus* var. *scolymus*), Theor. Appl. Genetics 2009 Online first, DOI 10.1007/s00122-009-1159-2.

[31] D. Scaglione, A. Acquadro, E. Portis, C.A. Taylor, S. Lanteri, S.J. Knapp, Ontology and diversity of transcript-associated microsatellites mined from a globe artichoke EST database, BMC Genomics 10 (2009) 454.

[32] R.D. Finn, J. Tate, J. Mistry, P.C. Coggill, S.J. Sammut, H.R. Hotz, G. Ceric, K. Forslund, S.R. Eddy, E.L.L. Sonnhammer, A. Bateman, The Pfam protein families database, Nucl. Acids Res. 36 (2007) D281-D282.

[33] A. Conesa, S. Gotz, J. Garcia-Gomez, J. Terol, M. Talon, M. Robles, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, Bioinform. 21 (2005) 3674-3676.

[34] S. Kumar, K. Tamura, M. Nei, MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment, Brief.Bioinform. 5 (2004) 150-163.

[35] J. Sambrook, E. F. Fritsch, T. Maniatis, Molecular cloning: a laboratory manual. Second edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA, 1989.

[36] S. Ye, S. Dhillon, X. Ke, A. Collins, I. Day, An efficient procedure for genotyping single nucleotide polymorphisms, Nucl. Acids Res. 29 (2001) e88.
[37] J. VO, JoinMap® 4: software for the calculation of genetic linkage maps in experimental populations, B.V. Kyazma, Wageningen, 2006.
[38] D. Kosambi, The estimation of map distances from recombination values, Ann Eugen 12 (1944) 172–175.
[39] Q. Yang, K. Reinhard, E. Schiltz, U. Matern, Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA : anthranilate N-hydroxycinnamoyl/benzoyltransferase from elicited cell cultures of carnation, *Dianthus caryophyllus* L, Plant Mol. Biol. 35 (1997) 777-789.
[40] J. D'Auria, F. Chen, E. Pichersky, Characterization of an acyltransferase capable of synthesizing benzylbenzoate and other volatile esters in flowers and damaged leaves of *Clarkia breweri*, Plant Physiol. 130 (2002) 466-476.
[41] K. Burhenne, B. Kristensen, S. Rasmussen, A new class of N-hydroxycinnamoyltransferases - Purification, cloning, and expression of a barley agmatine coumaroyltransferase (Ec 2.3.1.64), J. Biol.Chem. 278 (2003) 13919-13927.
[42] E. Tacke, C. Korfhage, D. Michel, M. Maddaloni, M. Motto, S. Lanzini, F. Salamini, H.P. Doring, Transposon tagging of the maize Glossy2 locus with the transposable element En/Spm, Plant J. 8 (1995) 907-917.
[43] V. Negruk, P. Yang, M. Subramanian, J.P. McNevin, B. Lemeux, Molecular cloning and characterization of the CER2 gene of *Arabidopsis thaliana*, Plant J. 9 (1996) 137-145.
[44] Y. Xia, B.J. Nikolau, P.S. Schnable, Cloning and characterization of CER2, an *Arabidopsis* gene that affects cuticular wax accumulation, Plant Cell 8(1996) 1291-1304.
[45] H. Suzuki, T. Nakayama, K. Yonekura-Sakakibara, Y. Fukui, N. Nakamura, M. Yamaguchi, Y. Tanaka, T. Kusumi, T, Nishino, cDNA cloning, heterologous expressions, and functional characterization of malonyl-coenzyme A : anthocyanidin 3-O-glucoside-6 "-O-malonyltransferase from dahlia flowers, Plant Physiol. 130 (2002) 2142-2151.
[46] H. Teutsch, M. Hasenfratz, A. Lesot, C. Stoltz, J. Garnier, J. Jeltsch, F. Durst, D. Werckreichhart, Isolation and sequence of a cdna-encoding the jerusalem-artichoke cinnamate 4-hydroxylase, a major plant cytochrome-p450 involved in the general phenylpropanoid pathway, PNAS 90 (1993) 4102-4106.
[47] M. Clifford, Chlorogenic acids and other cinnamates - nature, occurrence and dietary burden, J. Science Food Agric. 79 (1999) 362-372.
[48] C. Stewart, B. Kang, K. Liu, M. Mazourek, S. Moore, E. Yoo, B. Kim, I. Paran, M. Jahn, The Pun1 gene for pungency in pepper encodes a putative acyltransferase, Plant J. 42 (2005) 675-688.
[49] R. Dixon, N. Paiva, Stress-induced phenylpropanoid metabolism, Plant Cell 7 (1995) 1085-1097.

Figures

Figure 1 - Proposed pathways for the biosynthesis of CQAs in plants

PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-hydroxycinnamoyl-CoA ligase; HCT, hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase; HQT, hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferase; C3'H, *p*-coumaroyl ester 3'-hydroxylase.

Figure 2 - Phylogenetic analysis

The guided phylogenetic tree was constructed by the neighbour-joining method with 10,000 bootstrap replicates. The length of the lines indicates the relative distances between nodes. Globe artichoke protein sequences used for the alignment are the TCs and singletons reported in Table 1 together with the HCT (HCT_Cc, AAZ80046, shikimate/quinate hydroxycinnamoyl-transferase) and HQT (HQT_Cc, ABK79689, quinate hydroxycinnamoyl-transferase), previously isolated.

Acyltransferases from other species are: HCT_Nt (CAD47830, shikimate/quinate hydroxycinnamoyl-transferase from *Nicotiana tabacum*); HCBT_Dc (CAB06430, anthranilate N-hydroxycinnamoyl/benzoyltransferase from *Dianthus caryophyllus*); TAT_Tc (AAF34254, taxa-4(20),11(12)-dien-5 alpha-ol-O-acetyl transferase from *Taxus cuspidata*); CHAT_At (AAN09797, (Z)-3-hexen-1-ol O-acetyltransferase from *Arabidopsis thaliana*); AMAT_Vl (AAW22989, anthraniloyl-CoA:methanol acyltransferase from *Vitis labrusca*); BEBT_Nt (AAN09798, benzoyl-CoA: benzylalcohol O-benzoyltransferase from *Nicotiana tabacum*); ACT_Hv (AAO73071, agmatine coumaroyltransferase from *Hordeum vulgare*); Ss5Mat2_Ss (AAR26385, malonyl-CoA:anthocyanin 5-O-glucoside-6'''-O-malonyltransferase from *Salvia splendens*); SalAT_Ps (AAK73661, salutaridinol 7-O-acetyltransferase from Papaver somniferum); SAAT_.Fa (AAG13130, strawberry alcohol acyltransferase from *Fragariaxananassa*); Glossy2_Zm (CAA61258, from *Zea mays*), CER2_At (AAM64817, from *Arabidopsis thaliana*), Gt5AT_Gt (BAA74428, anthocyanin 5-aromatic acyltransferase from *Genthiana triflora*); Dm3MAT1_Cm (AAQ63615, anthocyanidin 3-O-glucoside-6''-O-malonyltransferase from *Dendranthema x morifolium*); Pf3AT_Pf (BAA93475, hydroxycinnamoyl-CoA:anthocyanin 3-O-glucoside-6''-O-acyltransferase from *Perilla frutescens*); Ss5MaT1_.Ss (AAL50566, malonyl-CoA:anthocyanin 5-O-glucoside-6'''-O-malonyltransferase from *Salvia splendens*).

Figure 3 - Sequence alignment of the newly isolated globe artichoke acyltransferases Cc_HCT and Cc_HQT (AAZ80046 and ABK7968 respectively) from *Cynara cardunculus*. Gray boxes indicate structural motifs conserved in the acyltransferase family.

Figure 4 - Heterologous expression in *E. coli* and *in vitro* enzymatic assays.

a) SDS-Page analysis of protein extract obtained from *E. coli* transformed with Acyltransf_2 sequence (on the left, the black arrow indicates protein product of expected size) and from empty vector transformed cells as control (on the right); b)- An example of HPLC analysis on reaction products with *p*-coumaroyl-CoA and quinate as substrates. An aliquot of the incubation reaction without (CTR) or with (Acyltransf_2) recombinant enzyme was analysed. 1: *p*-coumaroyl-CoA; 2 *p*-coumaroylquinate.

Figure 5 - qRT-PCR assays

The relative expression of the three selected genes in response to UV-C exposure of globe artichoke leaves was measured by qRT-PCR, using actin to normalize the expression level.

Figure 6 - SNP segregation profiles

Segregation of Acyltransf_1-snp (a), Acyltransf_2-snp (b) and Acyltransf_3-snp (c) in the mapping population, as detected by tetra-primers ARMS-PCR on 2% agarose gel; m = male parent and f = female parent.

Figure 7 - Consensus linkage groups showing the location of the genes belonging to CQAs pathway

Consensus LGs of 'Romanesco C3' (female parent, white LGs on the left) and 'Altilis 41' (male parent, gray LGs on the right), incorporating the CQA biosynthesis pathway genes marked by gray boxes. LGs on the female map were named LG_C3, and those on the male map LG_Alt, with both numbered serially in descending order of genetic length Aligned LGs are presented side-by-side. Marker names for the female LGs are shown on the right, with map distances (in cM) on the left; and *vice versa* for the male LGs. Intercross markers, shown in italics and in bold, are connected by a line.

Additional files

Additional file 1 – Oligonucleotide sequences used to study *C. cardunculus* genes from CQA pathway

Prefices indicate primers used for (i) full-length cDNA isolation and expression (Expr); (ii) single nucleotide polymorphism genotyping (SNP) and (iii) qRT-PCR analysis (RT). The numbers 1, 2 and 3 in the primer names refer to Acyltransf_1, Acyltransf_2 and Acyltransf_3, respectively.

Tables

Unigenes	ESTs number
CL806Contig1	GE613212, GE610766, GE613279, GE607940, GE581663, GE580713, GE603921
CL3797Contig1	GE582803, GE608988, GE601268
CL5755Contig1	GE580659, GE610099, GE583515
CL6182Contig1	GE607747, GE594571
CL5920Contig1	GE595995, GE600409
CL3094Contig1	GE610401, GE598627
CL4475Contig1	GE580690, GE608990
CL3654Contig1	GE584754, GE585361
CL5838Contig1	GE580719, GE612028
CL6332Contig1	GE594462, GE579713
CL4742Contig1	GE584222, GE604507
CL7000Contig1*	GE595377 and GE605258 (CL5348Contig1), GE609292 (CCPX6248)
CL7001Contig1*	GE599983 (CCPW8349), GE607289 (CCPX4364)
CCPW6362	GE597850
CCPW12442	GE588278
CCPW16540	GE592661
CCPU8587	GE584230
CCPX9719	GE612996
CCPU1541	GE577484
CCPU1786	GE577718
CCPX2825	GE605747
CCPX6902	GE609991
CCPU3833	GE579662
CCPX6333	GE609384
CCPU8298	GE583926
CCPX5567	GE608569
CCPX3611	GE606515
CCPX4145	GE607055
CCPU7105	GE582738
CCPU6045	GE581746
CCPX3803	GE606699
CCPU2493	GE578391

Table 1 - Set of globe artichoke unigenes [32] showing the conserved motifs of BAHD family protein. Unigenes tagged with * originate from further alignments of correspondent extended ESTs.

Marker	Parental genotypes (female x male)	Expected ratios and F_1 plant genotypes		Observed ratios				χ^2
				aa	ab	bb	total	
Acyltransf_1-snp	ab x aa	1:1	(aa:ab)	43	51		94	0.05 ns
Acyltransf_2-snp	ab x ab	1:2:1	(aa:ab:bb)	27	44	21	92	0.06 ns
Acyltransf_3-snp	aa x ab	1:1	(aa:ab)	40	52		92	1.57 ns
C4H-snp	ab x aa	1:1	(aa:ab)	43	50		93	0.53 ns
4CL-snp	aa x ab	1:1	(aa:ab)	45	48		93	0.10 ns

Table 2 - Model, expected and observed segregation ratios in the F_1 progeny of the developed markers

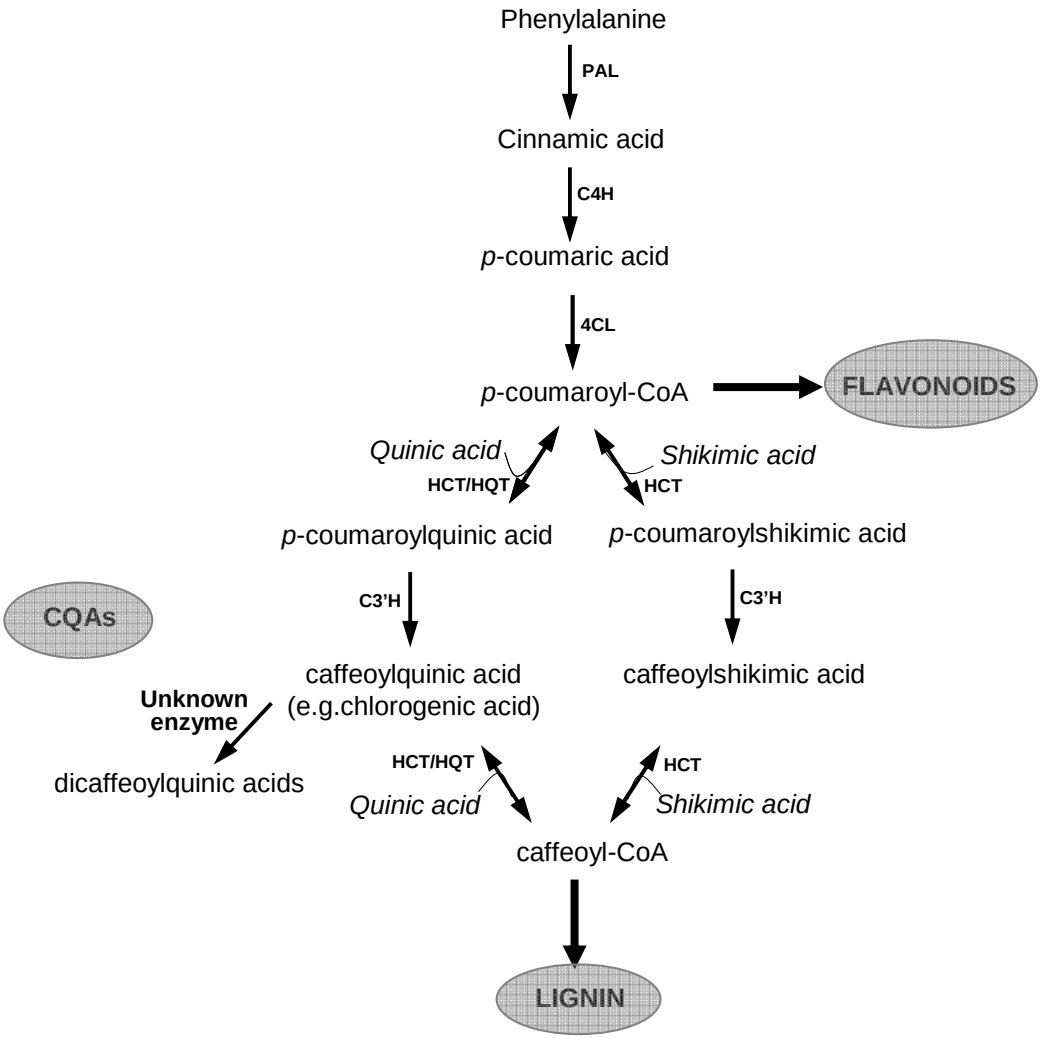

Phenylalanine
PAL
Cinnamic acid
C4H
p-coumaric acid
4CL
p-coumaroyl-CoA
FLAVONOIDS
Quinic acid
HCT/HQT
Shikimic acid
HCT
p-coumaroylquinic acid
p-coumaroylshikimic acid
CQAs
C3'H
C3'H
caffeoylquinic acid
(e.g.chlorogenic acid)
caffeoylshikimic acid
Unknown
enzyme
dicaffeoylquinic acids
HCT/HQT
Quinic acid
HCT
Shikimic acid
caffeoyl-CoA
LIGNIN

Figure 1

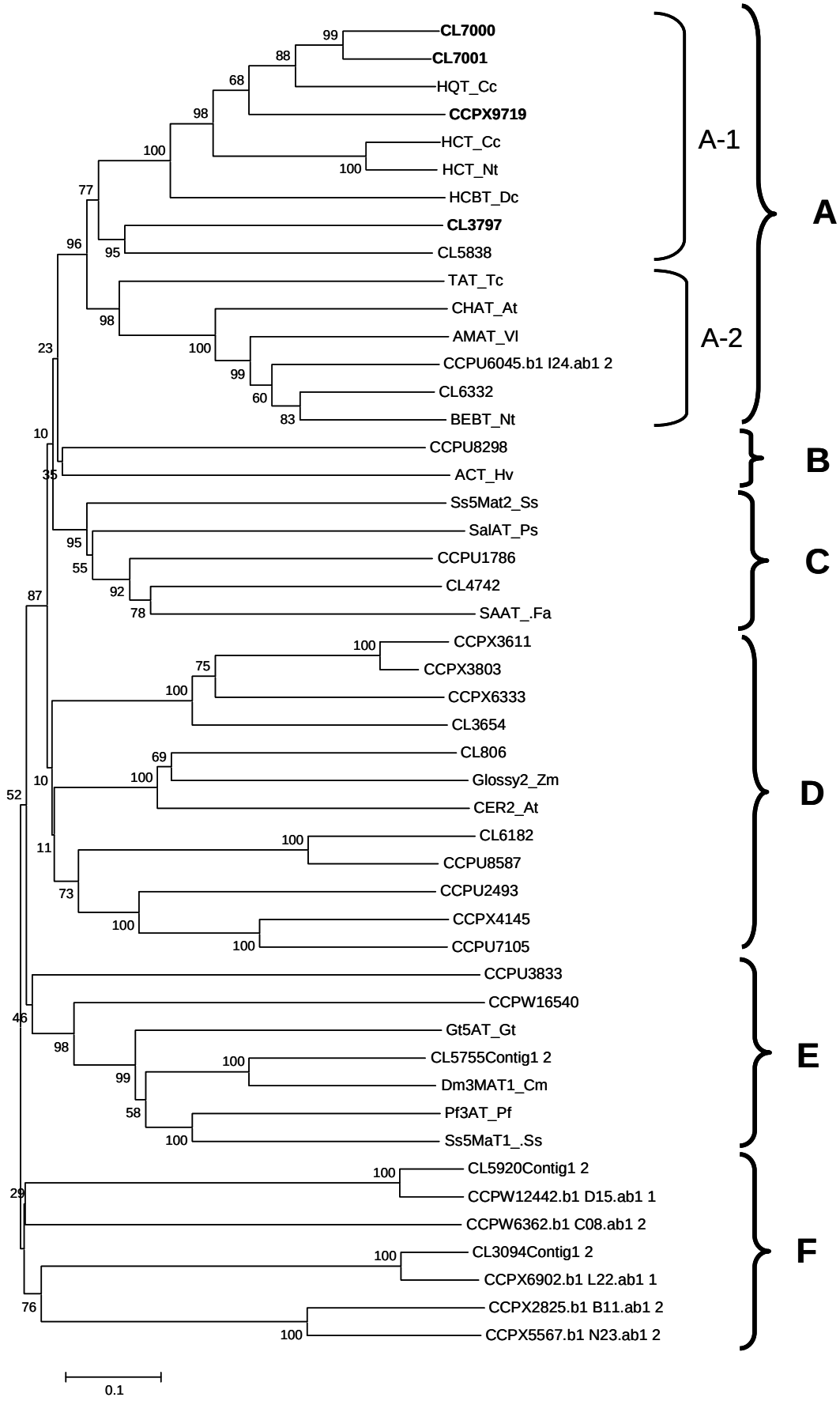
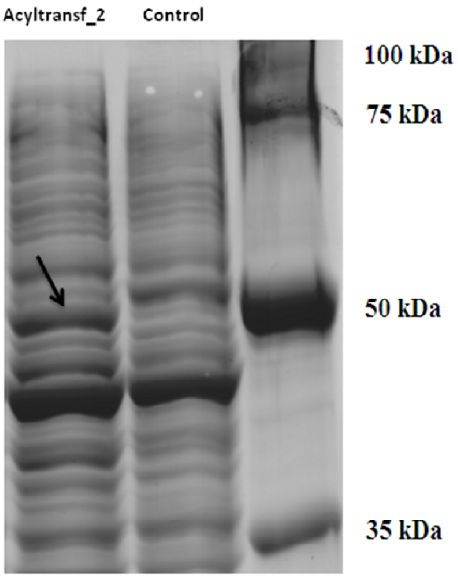

CL7000
CL7001
HQT_Cc
CCPX9719
HCT_Cc
HCT_Nt
HCBT_Dc
CL3797
CL5838
TAT_Tc
CHAT_At
AMAT_Vl
CCPU6045.b1 I24.ab1 2
CL6332
BEBT_Nt
CCPU8298
ACT_Hv
Ss5Mat2_Ss
SalAT_Ps
CCPU1786
CL4742
SAAT_.Fa
CCPX3611
CCPX3803
CCPX6333
CL3654
CL806
Glossy2_Zm
CER2_At
CL6182
CCPU8587
CCPU2493
CCPX4145
CCPU7105
CCPU3833
CCPW16540
Gt5AT_Gt
CL5755Contig1 2
Dm3MAT1_Cm
Pf3AT_Pf
Ss5MaT1_.Ss
CL5920Contig1 2
CCPW12442.b1 D15.ab1 1
CCPW6362.b1 C08.ab1 2
CL3094Contig1 2
CCPX6902.b1 L22.ab1 1
CCPX2825.b1 B11.ab1 2
CCPX5567.b1 N23.ab1 2
A-1
A-2
A
B
C
D
E
F
0.1

Figure 2

```
                      *        20         *        40         *        60         *        80         *
Acyltransf_1 : -------MGSDQKMMMNIDIMKSSIVPPSELIADCPKQLWTSNLDLVVGRIHILTVYFYRPNGSSKFFDPNVMKKALADVLVSFYPMAGRL :  84
Acyltransf_2 : ---------------MNINIKHSSFVQPSQPTPSS--TIWTSNLDLVVGRIHILTVYFYRPNGASNFFDADVMKKALADVLVSFYPMAGRI :  74
Cc_HQT       : ---------------MELTVKESLMVKPSKPTPNQ--RLWNSNLDLVVGRIHILTVYFYRPNGSSNFFDSGVLKKALADVLVSFFPMAGRL :  74
Cc_HCT       : ---------------MKIEVRESTMVRPAEETPRIN--LWNSNVDLVVPNFHTPSVYFYRPNGAANFFDPKVMKDALSRALVPFYPMGGRL :  74
Acyltransf_3 : MVTTGTYSDKSGVELIVTRMGEPTLVQPFEETEKG--LYFLSNLDQNIA-VIVRTIYCFKSEEKGNEMAAEVIKDALSKVLAHYHFAAGRL :  88

                     100         *       120         *       140         *       160         *       180
Acyltransf_1 : GRDETDRIVINCNNEGVLFVEAESDSTLDDFGELKPS--PVFRQLTPSVDYSGDISSYPLLFAQVTHFKCGGVALGCGVHHTLSDGLSLLH : 173
Acyltransf_2 : SRDRNGRLEINCNGEGVLFVEAELDSTLDDFGEFTPS--PELRRLTPTVDYSGDISSYPLFFAQVTHFKCGGVALGCGVFHALADGLSSIH : 163
Cc_HQT       : GNDGDGRVEINCNGEGVLFVEAEADCSIDDFGEITPS--PELRKLAPTVDYSDQVSSYPLCITQVTRFNCGGVSLGCGLHHTLSDGLSSLH : 163
Cc_HCT       : KRDEDGRIEIDCQGQGVLFVEAESDGVIDDFGDFAPT--LELRKLIPAVDYTLGIESYSLLVLQVTYFKCGGVSLGVGMQHHAADGASGLH : 163
Acyltransf_3 : TISSEGKLIVDCTNEGAVFVEAEANGNIEDIGDHTKPDPMTLGKLVYDVPGAKNILEIPPLVVQVTKFKCGGFVLGLGMNHNLFDGIAAME : 179

                    *       200         *       220         *       240         *       260         *
Acyltransf_1 : FINTWSDMARGLSVAIPPFIERTLLRAREPPTPTYDHVEYHSPPSMNTT-AQKP-GSGSLSKSSTTMLKLTLDQLNSLKAKAKSE-SGSTH : 261
Acyltransf_2 : FINTWSDMARGLSIAIPPFIDRTLLRAREPPTPTFDHIEYHAPPSMKTI-SQNP-ESS--RKPSTTVLKLTLDQLNVLKASTKNDGSNTTY : 250
Cc_HQT       : FINTWSDKARGLSVAIPPFIDRSLLRARDPPTAMFEHLEYHSPPSLIAP-SQNQNFTSHPKLASTAMLRLTLDQINGLKSKAKGD-GSVYH : 252
Cc_HCT       : FINTWSDLARGLDLAVPPFIDRTLLRSRDPPQPAFDHIEYQPAPPMKTAPTPTPTDDESVPETTVSIFKLTRDQVNALKGKSKEDGNTVNY : 254
Acyltransf_3 : FISSWSRTARGLPLEVPPFLDRTILNARNPPLVEFPHDEFAEIEDVSNT------VDLYKEELAYRSFCFSPDDIQRLKIKATADGDMPTC : 264

                   280         *       300         *       320         *       340         *       360
Acyltransf_1 : STYEILAAHIWRCACKARGLPDDQLSKLYVATDGRSRLSPRLPPGYLGNVVFTATPVAKSGDLTSKSLSNTAKLIHTTLTKMDDDYLRSAI : 352
Acyltransf_2 : STYEILAAHLWRCACKARGLPDDQLTKLYVATDGRSRLSPQLPPGYLGNVVFTTTPVAKSGDLTTQSLSNAASLIRTTLTKMDNNYLRSAI : 341
Cc_HQT       : STYEILAAHLWRCACEARGLSDDQPTKLYVATDGRSRLNPPLPPGYLGNVVFTATPIAKSGEFKSESLADTARRIHSELAKMDDQYLRSAI : 343
Cc_HCT       : SSYEMLSGHVWRCVCKARGLPDDQDTKLYIATDGRARLRPSLPRGYFGNVIFTTTPIAVAGDLQSKPTWYAASKIHDALARMDDDYLKSAL : 345
Acyltransf_3 : TSFEALSAFVWKARTEALQMKPDQKTKLLFAVDGRSRFEPPLPEGYSGNGIVLTNSICKAGEQIENPLSFTVKLVHEAVKMTTDGYMRSAI : 355

                    *       380         *       400         *       420         *       440         *
Acyltransf_1 : DYLESQPDLSALIRGPSYFASPNLNINAWTRLPVYDADLGWGRPIFMGPACILYEGTIYVLPSPNNDRSVSLAVCLDANEQPLFEKFLYEF : 443
Acyltransf_2 : DYPEVQPDLSALIRGPSYFASPNLNINTWTRLPVHDADFGWGRPVFMGPACILYEGTIYVLPSPNNDRSMSLAVCLDADEQPLFEKFLYDF : 432
Cc_HQT       : DYLELQPDLTALVRGPTYFASPNLNINSWTRLPIYESDFGWGRPIFMGPASILYEGTIYIIPSPSGDRSVSLAVCLDPDHMSLFRKCLYDF : 434
Cc_HCT       : DYLELQPDLKALVRGAHTFKCPNLGITSWARLPIHDADFGWGRPIFMGPGGIAYEGLSFVLPSPINDGSLSIVISLQAEHMKLFSKFLYDI : 436
Acyltransf_3 : DYFEVTGARPSLAS--------TLLITTWSKLSFHAQDFGWGEPIMSGPVALPEKEVILFLSHGKQRKSVNVLLGLPVSAMKTFEELMKHI : 438
```

Figure 3

(a)


Acyltransf_2
Control
100 kDa
75 kDa
50 kDa
35 kDa

(b)

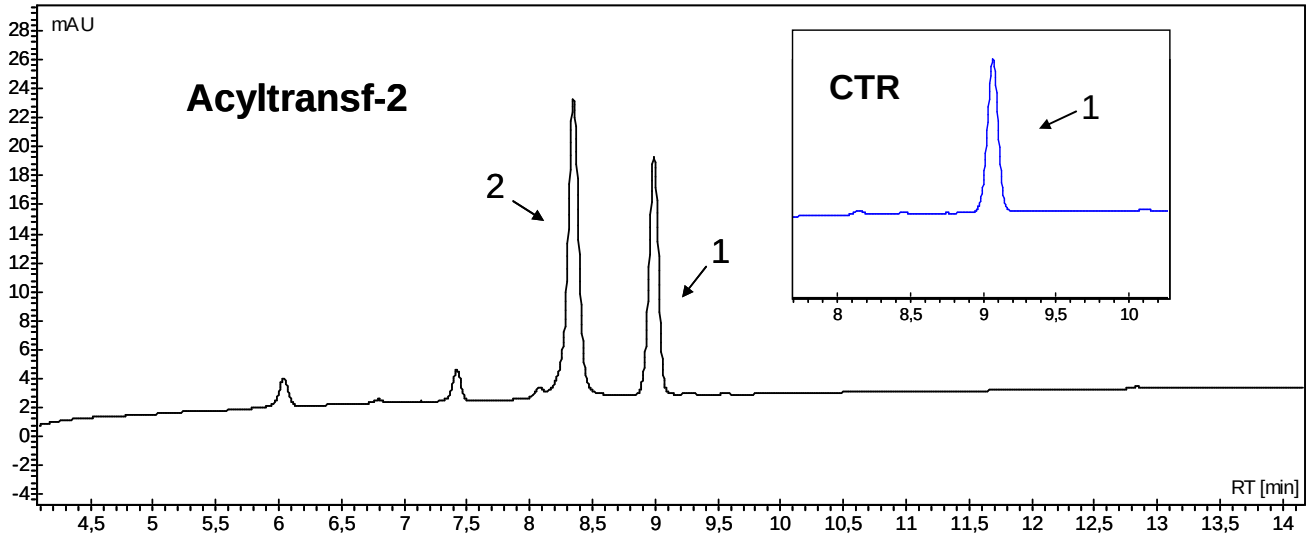

mAU
Acyltransf-2
CTR
1
2
1
RT [min]

Figure 4

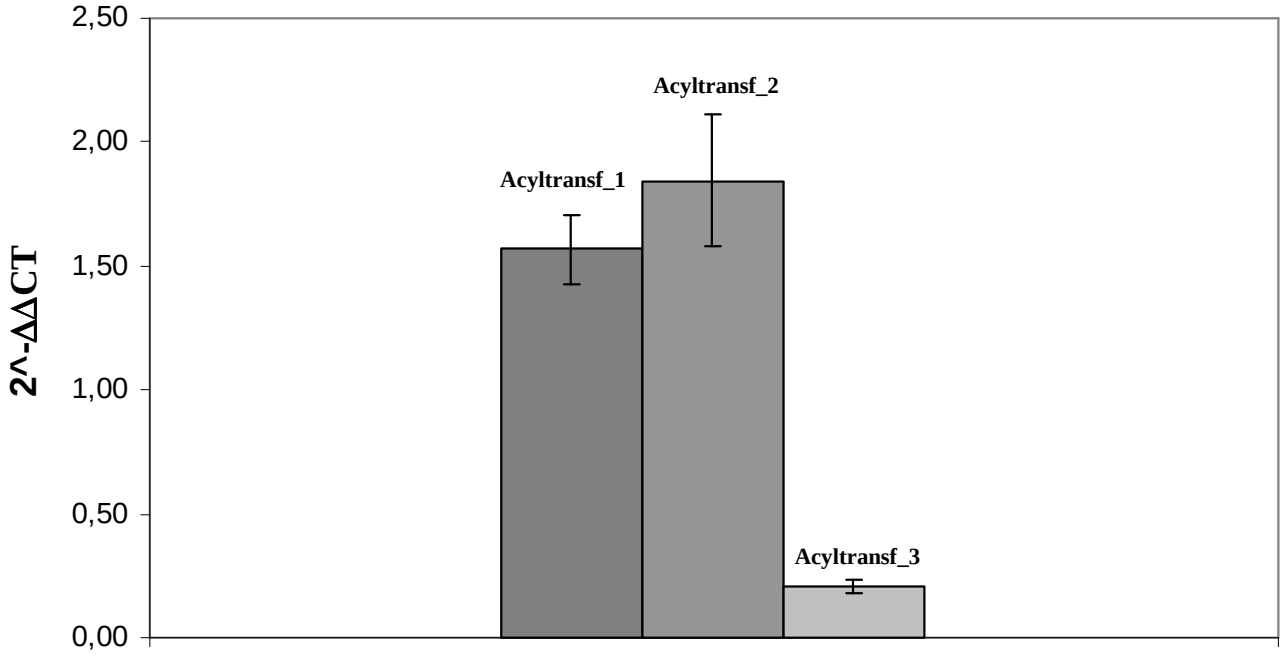

2,50
2,00
1,50
1,00
0,50
0,00
2^-ΔΔCT
Acyltransf_1
Acyltransf_2
Acyltransf_3

Figure 5

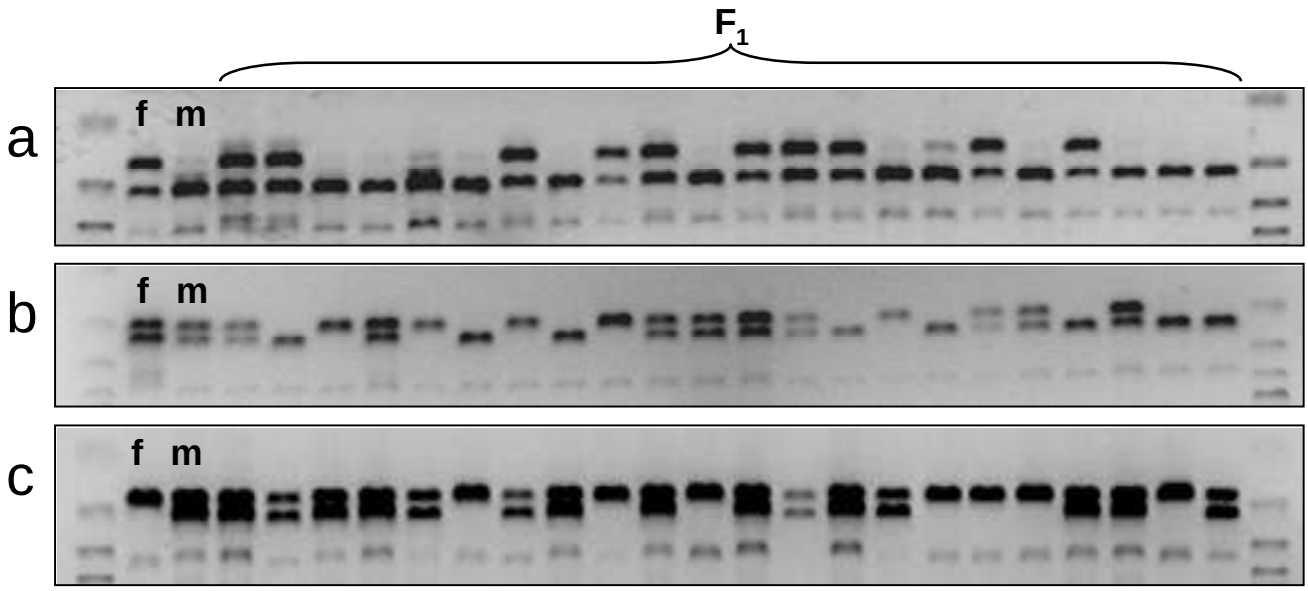


Figure 6

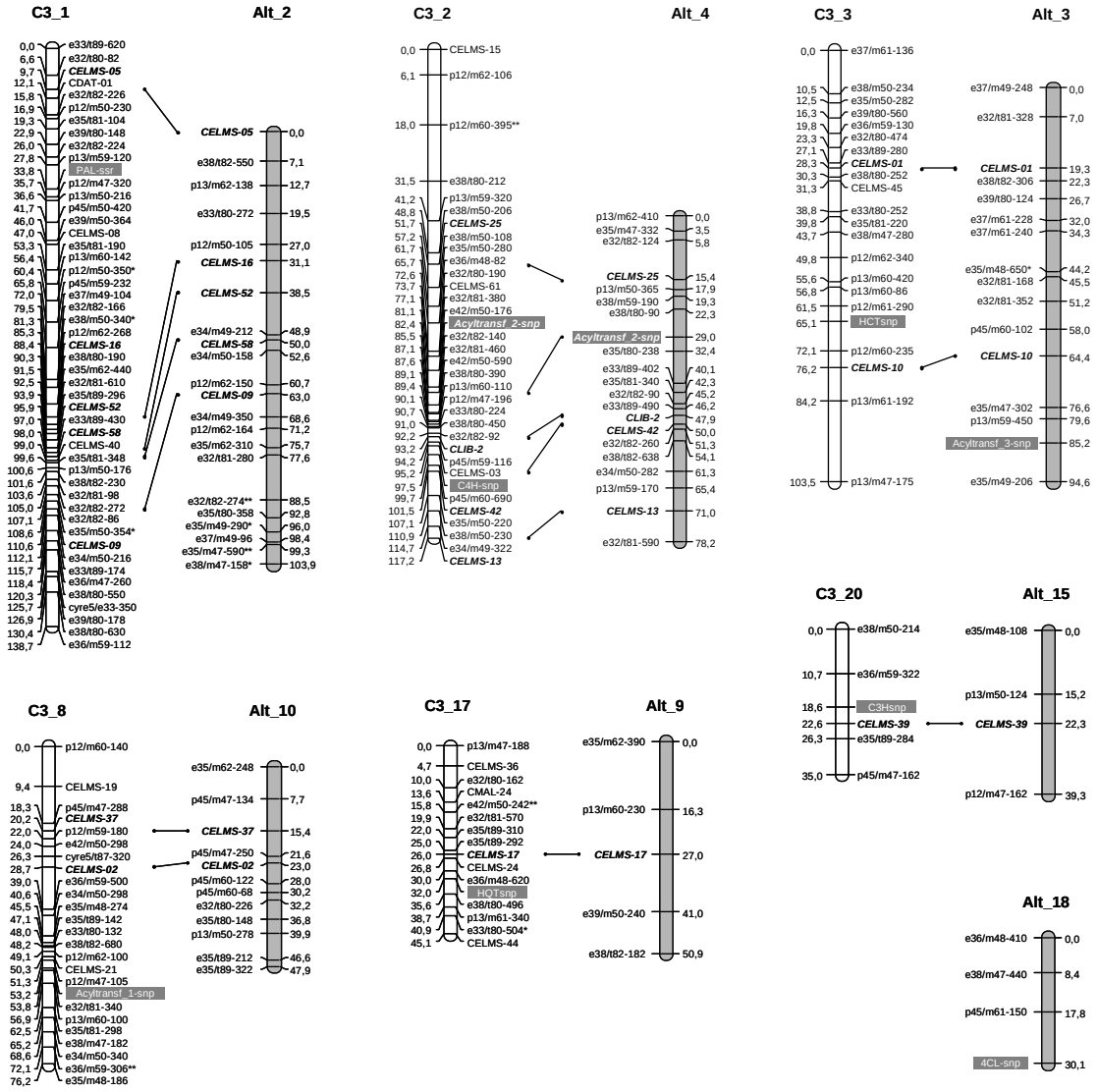

C3_1
Alt_2
C3_2
Alt_4
C3_3
Alt_3
C3_20
Alt_15
C3_8
Alt_10
C3_17
Alt_9
Alt_18
PAL-ssr
Acyltransf_1-snp
Acyltransf_2-snp
C4H-snp
C3Hsnp
HCTsnp
4CL-snp

Figure 7

Primer	Sequence (5'-3')
Expr1_For	GGGTTTCATATGGGAAGTGATCAAAAGATGATG
Expr1_Rev	ATAAGAATGCGGCCGCAAACTCGTACAAGAACTTTTCG
Expr2_For	GGGTTTCATATGAAYATTAATATAAAACAT
Expr2_Rev	CGGGATCCTTAAAATCATATAAGAACTTCTCA
Expr3_For	GGGTTTCATATGGTGACTACCGGAACCTAC
Expr3_Rev	CGGGATCCCTGATGTGCTTCATGAGTTCTTC
Snp1OutFor	TGCTCAAAAGCCAGGTTCTGGTTCCCTTTC
Snp1OutRev	TGGCGGTAAAGACCACGTTGCCTAGGTAGC
Snp1InnFor	TGGCGGCTCATATTTGGCGTTGTTCA
Snp1InnRev	GGTCATCTGGGAGTCCTCGAGCCTTGAAC
Snp2OutFor	CCCCACTTTTGACCATATCGAGTACCA
Snp2OutRev	ATCTCGTATGTGCTGTAGGTGGTGTTGC
Snp2InnFor	ATCTCCCAAAACCCGGAATCACGT
Snp2InnRev	CTTTAAAACGGTGGTGGAAGGTTTCAGG
Snp3OutFor	CCTCTTTTGAAGCTCTATCGGCTTTCGT
Snp3OutRev	CACATTCACACTTTTCCTCTGCTTTCCA
Snp3InnFor	ACGAGGCGGTGAAAATGATTCCC
Snp3InnRev	ATCAATCGCAGATCTCATGTAACCAGCA
SnpC4HOutFor	CAATGTTGCTGGTACGCATCTTTT
SnpC4HOutRev	TTGACCACTGCTTGGAGATATGGA
SnpC4HInnFor	CGGGATCTCCTACTTTTAATCTGGG
SnpC4HInnRev	AGGGTGGTCTCGATTGCTGGTT
Snp4CLOutFor	CATCCTCGCTGTGAATCCATAGATTCGG
Snp4CLOutRev	ATTACTCAATCGGAACATGCGGCGAAG
Snp4CLInnFor	TCCGACGAGATCTCAACCTCAGGAAGATC
Snp4CLInnRev	GTTGATTTCCGGCGATGAGGCGA
RT Seq1 For	CTATGACCATGTTGAATACC
RT Seq1 Rev	GCTTTGAGACTATTGAGTTG
RT Seq2 For	ACATACGAGATCCTAGCTG
RT Seq2 Rev	ACGTACAATTTGGTTAGTTG
RT Seq3 For	CATGGTCAAAACTATCATTC
RT Seq3 Rev	ACAGGATTACTTCTTTCTCC
RT Act For	TACTTTCTACAACGAGCTTC
RT Act Rev	ACATGATTTGAGTCATCTTC

Additional file 1