Isolation and molecular characterization of cinnamate 4-hydroxylase from apricot and plum

A. PINA[1]*, T. ZHEBENTYAYEVA[2], P. ERREA[1] and A. ABBOTT[2]

Unidad de Fruticultura, CITA de Aragón, Avda Montañana 930, E-50059, Zaragoza, Spain[1]
Department of Genetics and Biochemistry, Clemson University, Clemson, SC 29634 USA[2]

Abstract

Cinnamate 4-hydroxylase (C4H) is the second enzyme in the phenylpropanoid pathway which participates in the synthesis of numerous phenylpropanoid compounds such as flavonoids, lignins, suberins and others. We identified a gene putatively coding for Class I C4H in apricot and plum and we analyzed the expression pattern of this gene under different apricot/plum graft combinations with different degree of compatibility. The full-length cDNA is 1 739 bp with a 1 515 bp open reading frame encoding a protein of 504 amino acids. Like other C4Hs, the predicted C4H polypeptides included conserved domains of cytochrome P_{450}. The genomic sequence of the apricot *C4H* gene was interrupted by two introns 335 bp and 904 bp long. Several regulatory motifs including P-, A-, L- and H-boxes, which were conserved across phenylpropanoid metabolism-related genes in higher plants, were found in a 1 300 bp upstream promoter region of the apricot *C4H* gene. A phylogenetic analysis showed that all *Prunus* sequences clustered together and were closely related to *Malus* and *Rubus C4H* genes. The transcription of Class I *PruC4H* was detected in all the examined graft combinations, which suggested its rather constitutive character.

Additional key words: fruit trees, full length cDNA, graft compatibility, phenylpropanoid pathway, promoter analysis.

Introduction

Phenylpropanoid compounds provide beneficial effects in the plants as physiologically active compounds, as stress protecting agents, as attractants and by their significant role in plant resistance (Ehlting *et al*. 2006, Treutter 2006, Sullivan 2009, Vogt 2010). Phenolic compounds are derived from phenylalanine *via* the core phenylpropanoid pathway by action of phenylalanine ammonia-lyase (PAL) and cinnamate-4-hydroxylase (C4H) which is responsible for hydroxylation of cinnamic acid to *p*-coumaric acid. The product of this core phenylpropanoid reaction is then activated to its thioester by 4-coumarate coenzyme A (CoA)-ligase (4-CL) and is funneled into one of the branched pathways leading to the production of cell wall constituents (lignins), pigments (flavonoids), UV protectants (coumarins and flavonoids) and plant defense compounds (isoflavonoids, furanocoumarins). Most of the biosynthetic genes of the general phenylpropanoid metabolism have been cloned and characterized from many plant species (Ehlting *et al*. 1999, Kumar and Ellis 2001, 2003, Yang *et al*. 2005).

C4H defines a cytochrome P_{450}-linked monooxygenase (CYP73) and was one of the first P_{450} to be characterized in plants (see review by Ehlting *et al*. 2006). C4H activity and gene expression are modulated by a wide array of factors, such as light, wounding and pathogen infection indicating that it plays diverse roles in phenylpropanoid metabolism (see review by Chapple 1998). In addition, C4H function is critical to the normal biochemistry and development in *Arabidopsis* since a reduction in C4H activity results in a strong reduction in phenylpropanoid content and perturbations in growth and development (Schilmiller *et al.* 2009). C4H cDNAs and genomic clones have been isolated in several plant species such as *Arabidopsis* (Bell-Lelong *et al.* 1997, Mizutani *et al*. 1997, Raes *et al*. 2003), bean (Mizutani *et al*. 1993), *Populus* (Ge and Chiang 1996, Kawai *et al.* 1996, Lu *et al*. 2006), alfalfa (Fahrendorf and Dixon 1993), *Brassica napus* (Chen *et al*. 2007), *Salvia*

Received 21 May 2010, *accepted* 10 June 2011.
Abbreviations: BAC - bacterial artificial chromosome; C4H - cinnamate 4-hydroxylase; DAG - days after grafting; RACE - rapid amplification of cDNA ends; RT-PCR - reverse transcriptase-polymerase chain reaction.
Acknowledgements: This research was supported by the grant No. RTA2009-00128 from the Instituto Nacional de Investigacion y Tecnologia Agraria y Alimentaria (INIA). The authors highly appreciate the comments and suggestions of two anonymous reviewers and the editor which have clearly contributed to improve the quality of the original version of this manuscript.
* Corresponding author; fax: (34) 976716335, e-mail: apina@aragon.es

miltiorrhiza (Huang *et al*. 2008), rice (Yang *et al*. 2005), *Helianthus tuberosus* (Teutsch *et al*. 1993), *Citrus* spp. (Betz *et al*. 2001), and others.

Recently, the general phenylpropanoid pathway was identified as one of the important metabolic pathways responsible for physiological failure in graft-incompatible rootstock-scion combinations in apricot (Pina and Errea 2008). PAL is overexpressed in graft unions of incompatible partners, where a lack of adaptation between stock and scion takes place. Lack of information on molecular genetic organization of phenylpropanoid pathway in *Prunus* species hampers further progress in understanding the effects of the graft incompatibility reaction in the phenylpropanoid metabolism. The cloning and characterization of the genes involved in the phenylpropanoid biosynthetic pathway would greatly help to extend our knowledge on the regulation of all of the core phenylpropanoid key enzymes during graft union development and the graft incompatibility phenomenon associated with phenol accumulation and lignin biosynthesis (Errea *et al*. 1992). So far, only a few P_{450} genes involved in the phenylpropanoid pathway have been isolated in woody species. Therefore we, initiated, molecular cloning and characterization of genes involved in general phenylpropanoid metabolism in apricot and other *Prunus* species. Here, we report the first results on molecular cloning and characterization of C4H from apricot and plum, and identified putative regulatory sequences in the promoter region upstream of the gene. A full length C4H gene was isolated based on gene sequence information, and its transcription profile under different apricot/plum graft combinations was also analyzed.

Materials and methods

Plants and culture conditions: Apricot (*Prunus armeniaca* L. cvs. Paviot, Moniqui and Goldrich) and plum (*Prunus cerasifera* × *Prunus munsoniana* cv. Marianna 2624) callus tissues were obtained according to Pina and Errea (2008). Graft combinations from compatible homografts (Moniqui grafted on Moniqui, Paviot grafted on Paviot and Marianna 2624 grafted on Marianna 2624), compatible heterograft (Paviot grafted on Marianna 2624) and incompatible heterograft (Moniqui grafted on Marianna 2624) were established by placing two callus pieces upright on the agar layer under sterile conditions in the same culture medium and conditions as used for callus initiation. The graft combinations were sampled 10 d after grafting and ungrafted samples were used as controls. The samples were immediately frozen in liquid nitrogen after harvest and maintained at -80 ºC until use.

3´and 5´cDNA end amplification: Total RNA was extracted from callus tissue using *Trizol* reagent (*Invitrogen Life Technologies*, USA) according to the manufacturer's recommendation. 10 µg of total RNA from callus tissue was used as template to synthesize first strand cDNA using *GeneRacer* kit (*Invitrogen Life Technologies*). Based on multiple sequence alignments from *Malus domestica* (DQ075002.1), *Fragaria chiloensis* (DQ898278), *Humulus lupulus* (FJ617541.1) and *Rubus occidentalis* (FJ554629), a degenerate forward primer PruF1 5´-CAGCT(G/T)ATGATGTACAACAAC ATGTAC-3´ was synthesized for 3´ rapid amplification of cDNA ends (RACE) from apricot C4H. This primer was paired with *GeneRacer 3´*-primer in a standard 50 mm^3 high fidelity PCR system reaction. Amplification was performed by touchdown PCR as follow: 94 ºC for 2 min, 94 ºC for 30 s, 72 ºC - 55 ºC for 2 min at the ramp of 2 ºC per cycle, 68 ºC for 1.30 min; then 29 cycles of 94 ºC for 30 s, 55 ºC for 45 s, 68 ºC for 2 min, and 68 ºC for 10 min. After agarose gel electrophoresis, the target band was cloned in *TOPO-TA* cloning (*Invitrogen Life Technologies*) and positive colonies sequenced.

In 5´RACE, antisense primer PruR1 5´-GTACAT GTTGTTGTACATCAT(A/C)AGCTG-3´ was synthesized to pair with *GeneRacer 5´*primer. The RT-PCR conditions were the same as those for 3´RACE PCR.

Amplification of full-length cDNAs and genomic sequences: Based on sequencing results of the 3´and 5´RACE products, specific sense primer PruF2 5´-GAA ATGGACCTACTCCTCCTG-3´ and antisense primer PruR2 5´-CTATGACCTTGGCTTCAACAC-3´ were designed. Full length cDNAs were amplified from different *Prunus* genotypes: plum cv. Marianna 2624, apricot cvs. Moniqui, Goldrich and Paviot. The temperature program was 3 min at 94 ºC followed by 34 cycles of 45 s at 94 ºC, 1 min at 55 ºC, 1.30 min at 68 ºC and 10 min at 68 ºC in a 50 mm^3 standard high fidelity Taq PCR system. Gel recovery, *TOPO-TA* cloning and sequencing were performed.

Apricot *C4H* gene has been sequenced directly from C4H-positive clones identified in a *Goldrich* BAC DNA library (Vilanova *et al.* 2003). Briefly, from an apricot Goldrich callus cDNA library we amplified a C4H fragment of 800 bp using the primers designed in conserved regions of C4H, PruF1 and PruR3 (Table 1). This fragment was double labelled with [α-^{32}P] dCTP and ^{32}PdATP using a *Random Hexamer* labelling mix and hybridized against high-density colony filters as described in Horn *et al*. (2005). BAC DNA preparation was performed by a modified alkaline-lysis procedure according to Vilanova *et al*. (2003). Sequencing using *Big Dye*® chemistry was done in Clemson Genomic Institute using gene specific primers. A 1 300 bp upstream promoter region was directly sequenced from BAC DNA by iterative primer walking (Table 1). Sequence data from this article have been deposited at the GenBank under acc. Nos. HM204477 (PaC4H) and HM204478 (PmC4H).

Southern blot analysis: Total genomic DNA was isolated from young leaves from dihaploid peach

cv. Lovell, diploid apricot cvs. Moniqui and Paviot, diploid plum cvs. Marianna 2624 and Myrobolan 29C, hexaploid plum cv. Reine Claude and sour cherry cv. Montmorency (tetraploid) using a CTAB-based method (Hormaza 2002). Aliquots (20 µg) of genomic DNA were digested overnight at 37 ºC with *EcoR*I and *Hind*III, then fractionated in 0.8 % (m/v) agarose gel in TBE buffer according to standard protocols (Sambrook and Russel 2001). Each gel was blotted onto a nylon membrane (*Hybon N*$^+$; *Amershan*, UK) which was then baked at 80 ºC for 2 h. The membrane was hybridized with ^{32}P-labeled cDNA fragments that included the 0.8 kb C4H open reading frame (ORF) as a probe. Hybridization was performed at 60 ºC in 7 % SDS-0.5 M phosphate buffer, pH 7.2, overnight. The blots were washed with 2× SSC, 0.5 % SDS for 15 min at 60 ºC; then with 2× SSC, 0.1 % SDS for 15 min at 60 ºC; 1× SSC, 0.1 % SDS for 15 min at 60 ºC; and finally 0.5× SSC, 0.1 % SDS for 15 min at 60 ºC. Finally, blots were exposed to X-ray film for 5 - 7 d.

Semiquantitative RT-PCR analysis of C4H under different graft combinations: RNA was isolated from different graft combinations at the graft interface and from ungrafted callus tissue using *Trizol* reagent. All RNA samples were digested with turbo DNase (Ambion) to remove possible DNA contamination. 2.5µg of total RNA was reverse transcribed in 21 mm^3 for 1 h at 42 ºC using *oligo(dT)*$_{18}$ primer, *RnaseOut*, and *SuperScript III* reverse transcriptase (*Invitrogen Life Technologies*) according to the manufacturer's instructions. For cDNA amplification the primers PruF1 and PruR3 were used and the amplification program was performed at 94 ºC for 3 min, followed by 29 cycles at 94 ºC for 45 s, 55 ºC for 1 min, 72 ºC for 1.30 min, and 72 ºC for 10 min. The expression level of *C4H* gene was calculated after normalization to the level of the *actin* gene PCR product from the same sample. The *actin* gene was amplified by using the 5´ primer (5´-CTTTAATGTGCCTGCCATG-3´) and 3´ primer (5-´TGATGGAGTTGTAGGTAGTC-3´) designed from the sequence of *P. persica* actin cDNA (GenBank accession No. AB046952). The number of cycles used for amplification with each primer pair was adjusted to ensure that the amplification was in the linear range. The intensities of the PCR products were measured with the *Gel Doc2000* (*Bio Rad*, USA) by using the *Quantity One v. 4.4* quantification software.

Sequence data analysis: Sequence alignments were carried out with *CLUSTALW* (www.ebi.ac.uk). Basic *BLAST* was done using the NCBI (National Center for Biotechnology Information, www.ncbi.nlm.nih.go). Calculation of predicted protein, molecular mass, ORF translation and PI was done using the *ExPASy* (www.expasy.ch). Phylogenetic relationships among sequences were conducted with *MEGA v. 4.1* (Kumar *et al*. 2004) using the Neighbour-Joining method. The accession numbers of the C4H sequences used to construct the phylogenetic tree are *Malus domestica* (AAY87450.1), *Rubus occidentalis* (ACM17896.1), *Populus trichocarpa* (ACC63873.1), *Populus tremuloides* (ABF69101.1), *Arabidopsis thaliana* (AAB58356.1), *Solanum tuberosum* (ABC69046.1), *Medicago sativa* (P37114.1), *Pisum sativum* (Q43067.2), *Agastache rugosa* (AAU09021.1), *Capsicum annuum* (AAG43824.1), *Nicotiana tabacum* (ABC69411.1), C4H1 *Brassica napus* (ABF17874.1), C4H2 *Brassica napus* (ABF17875.1), C4H1 *Citrus sinensis* (AF255013_1), C4H2 *Citrus sinensis* (AF255014_1.), *Sorghum bicolor* (AAK54447), *Zea mays* (NP_001151365.1), *Phaseolus vulgaris* (CAA70595.1), and *Oryza sativa* (AAV44089.1). The gymnosperm C4H sequence from *Pinus taeda* (AAD23378.1) was selected as the out-group because gymnosperms are considered, on the basis of both morphological characters and 18S RNA sequences (Chaw *et al*. 1997), to be ancestral to the angiosperms. A bootstrap analysis was performed to test the reliability of the tree, with 1000 replications.

Results

A full length C4H cDNA was isolated from apricot cv. Goldrich *via* RT-PCR and RACE. 3´ and 5´ RACE results revealed a band about 1 000 and 650 bp, respectively. Five sequenced clones resulted in one 3´cDNA and one 5´cDNA ends. The bands were in consensus with homology based prediction, so they were cloned and sequenced. The results showed that the full length cDNA in apricot was 1 739 bp long with a 1 515 bp open reading frame encoding a protein of 504 amino acids, a 102 bp 5´ non-coding region and a 122 bp 3´untranslated region. The 1 515 bp long coding region was isolated from other apricot cultivars (Paviot and Moniqui) and the plum cv. Marianna 2624. These related species all encode a protein of 504 amino acids with a predicted molecular mass of 57.8 kDa and a predicted isoelectric point of 9.1. Sequence homology analysis indicated that they exhibit high homology to known Class I C4H sequences with high identities of 85 % with *Rubus occidentalis* (FJ554629) and *Malus domestica* (DQ075002) and 81 % with *Populus trichocarpa* (XM_002319939). Comparison of the coding sequences revealed four nucleotide differences between Goldrich and Moniqui, 1 nucleotide different between Goldrich and Paviot, and 2 nucleotide differences between Paviot and Moniqui. All of these changes are silent with regard to the encoded protein. The homology between apricot *vs*. plum is 98 % at the nucleotide level and 99 % at the amino acid level, respectively. In all species the predicted amino acid sequences contained the conserved domains and active residues that are highly conserved among C4H proteins and are required for catalytic functions (Fig. 1).

Additionally, we conducted a phylogenetic analysis to

Table 1. Reverse (R) and forward (F) primer sequences.

Primer name	Primer sequence	Use
PruR1	(R) 5′-GTACATGTTGTTGTACATCAT(A/C)AGCTG-3′	5′RACE
PruF1	(F) 5′-CAGCT(G/T)ATGATGTACAACAACATGTAC-3′	3′RACE and RT-PCR
PruF2	(F) 5′-GAAATGGACCTACTCCTCCTG-3′	full length cDNA
PruR2	(R) 5′-CTATGACCTTGGCTTCAACAC-3	full length cDNA
PruR3	(R) 5′-GCCTAAACTCCTCAGGCTTCTTCCA-3′	RT-PCR
ext1	(F) 5′-GGACTGAAGTGCGCCATCG-3′	promoter
ext2	(F) 5′-GGAAGAAGCCTGAGGAGTTTA-3′	promoter
ext3	(R) 5′-CAGATCTTCAAGTAGCCTCTC-3′	promoter
ext4	(R) 5′-GTACATGTTGTTGTACATCAT(A/C)AGCTG-3′	promoter
ext5	(F) 5′-CAGCT(G/T)ATGATGTACAACAACATGTAC-3′	promoter
ext6	(R) 5′-GCCTAAACTCCTCAGGCTTCTTCCA-3′	promoter
ext7	(R) 5′-ATGGTCGATGGCGCACTTCAGTCC-3′	promoter
ext8	(R) 5′-TTGTGTGGGTGTGGCGACTG-3′	promoter
ext9	(F) 5′-CAGTCGCCACACCCACACAA-3′	promoter
ext10	(F) 5′-GAGAGCAAGATCCTGGTGAAT-3′	promoter
ext11	(R) 5′-GTTTCCGAAAACTGGTACGGG-3′	promoter
ext12	(F) 5′-CCAGAAGCTTCCCTACCTTC-3′	promoter
ext13	(R) 5′-CAGGAGGAGTAGGTCCATTTC-3′	promoter
ext14	(R) 5′-GCTTCAGGATCACAATGCTATG-3′	promoter
Actin	(F) 5′-CTTTAATGTGCCTGCCATG-3′	normalization
Actin	(R) 5′-TGATGGAGTTGTAGGTAGTC-3′	normalization

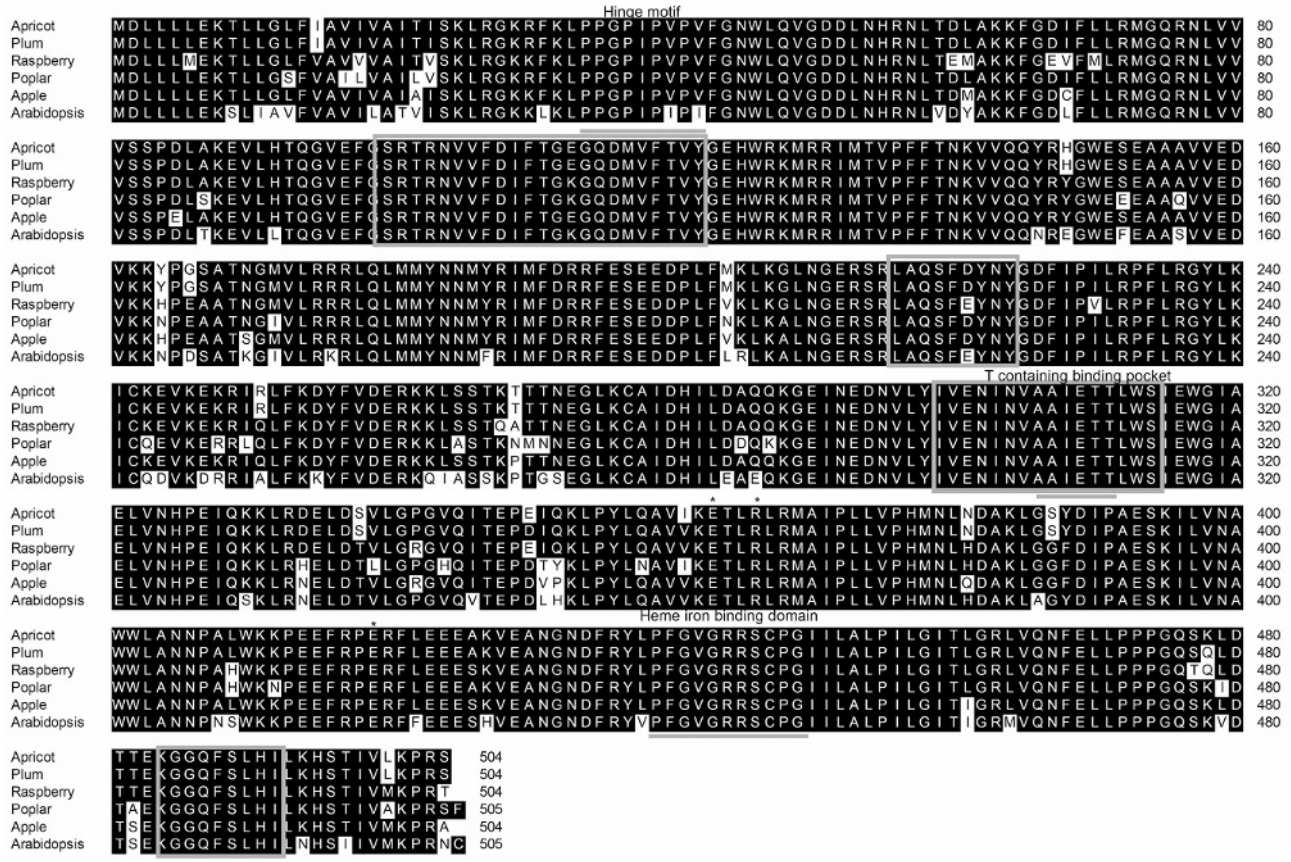

Fig. 1. Aligment of apricot and plum C4H amino acid sequences with C4Hs from other plants. The gene accession numbers are raspberry (ACM17896.1), poplar (ACC63873.1), apple (AAY87450.1) and *Arabidopsis* (AAB58356.1). The completely identical amino acids are indicated with a *black background*. Conserved domains of cytochrome P_{450} proteins such as the proline hinge, T containing binding pocket, the ERR triad (amino acids represented with *asterisks*) and haem binding domains are *gray underlined*. The substrate recognition sites (SRS) are shown with *gray rectangles*.

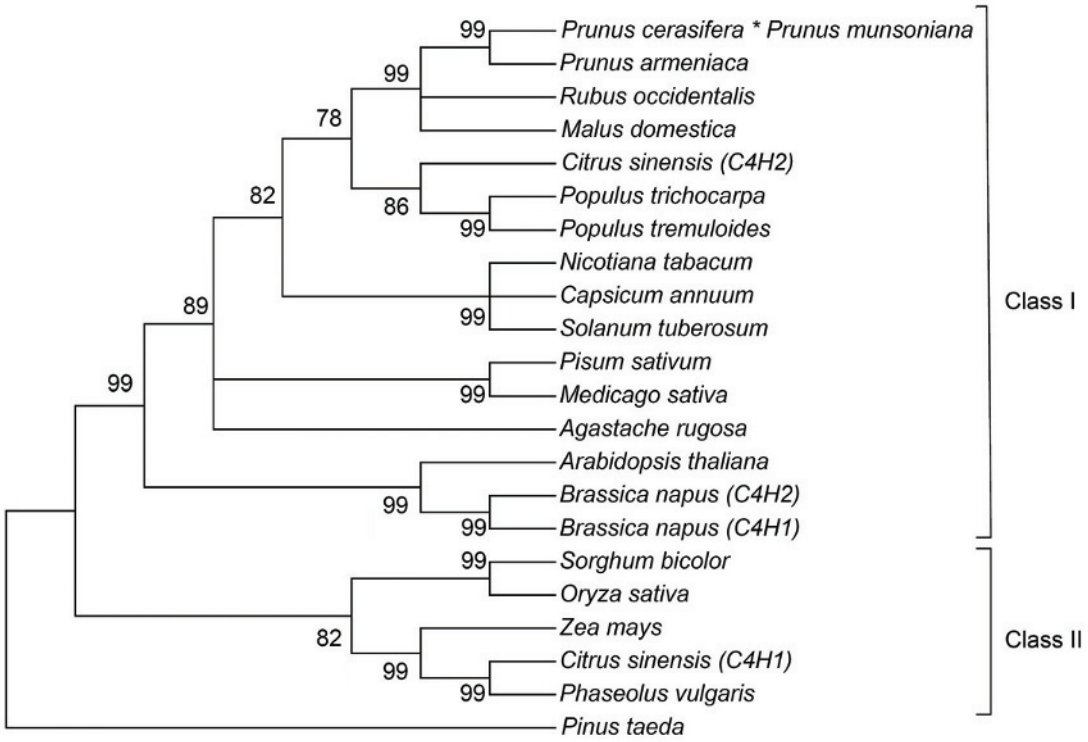

Fig. 2. Phylogenetic relationship among C4H genes from *Prunus* and other plants. The full length amino acid sequences were aligned and a Neighbour-Joining tree was constructed using the program *MEGA version 4.1*. Bootstrap values (based on 1 000 replicates) are shown. Accession numbers are described in material and methods. The *Pinus taeda* C4H sequence (AF096998.1) was used to root the tree.

study the evolutionary relationship among C4H genes from other plant species available in the *GenBank* database, including apricot and plum. The phylogenetic tree had several interesting features, with high bootstrap values placing C4H sequences in separated groups (Fig. 2). Notably, the plant C4Hs were subdivided into two branches. Most biochemically characterized proteins fall into the first branch named class I. Within this branch C4Hs were grouped into families. Among the *Prunus* C4Hs, apricot C4H and plum C4H formed the same subgroup. All *Prunus* sequences clustered with the *Malus* and *Rubus* C4H sequences. Likewise, all C4Hs belonging to the *Solanaceae* family grouped together, as do C4H gene characterized from *Nicotiana tabacum*, *Solanum tuberosum* and *Helianthus annuum*. A divergent clade of CYP73s included C4Hs from class II, such as *Citrus sinensis* C4H1 (Betz *et al*. 2001), *Phaseolus vulgaris* and monocots (*Oryza sativa*, *Shorgum bicolor*).

An apricot genomic library was screened to identify BAC clones containing C4H genes using a ^{32}P-labeled apricot C4H cDNA as a probe. Six positive BAC clones were detected in the library and checked by PCR using PruF1 and PruR3 primers. After sequencing the genomic sequence of the apricot C4H gene (1 957 bp), we found that its coding sequence is interrupted by two introns, intron I 335 bp and intron II 904 bp, which varied in length from C4H introns of other plants, as expected.

Likewise, a 1 300 bp upstream promoter region responsible for the expression of class I C4H was analyzed (Fig. 3). The transcription start site of C4H was found to a position 101 bp upstream of the translation initiation codon ATG. A putative TATA box was located 134 bp upstream from the translation start codon and a putative CCAAA at -109. The ACCTA-containing sequence most proximal to the C4H transcription initiation site overlaps the sequence CACGTA, which is similar to the G-box sequence CACGTG, previously identified in other light inducible promoters (Loake *et al*. 1992). In addition, two similar sequence motifs for A-box CCGTTA were found at -148 and -272 from the ATG codon and a wound responsive element WRE3 (CCACCT) as reported by Whitbred and Schuler (2000).
To determine the complexity of class I C4H genes in *Prunus* genome, genomic DNA was extracted from different cultivars representing various ploidy levels (Fig. 4). Three or four bands were observed in all species, consistent with the presence of internal restriction sites for these enzymes within C4H sequence. *i.e.,* 2 sites in peach, 3 sites in plum and apricot genomic DNA digested with *Eco*R1. The simplest interpretation of our data is that there is only one Class I C4H gene in *Prunus spp*. More evidence that this gene is encoded by a single copy is obtained of the draft assembled and annotated peach genome released by the *International Peach Genome Initiative* (*IPGI*) on Genome database for *Rosaceae* (*GDR*).

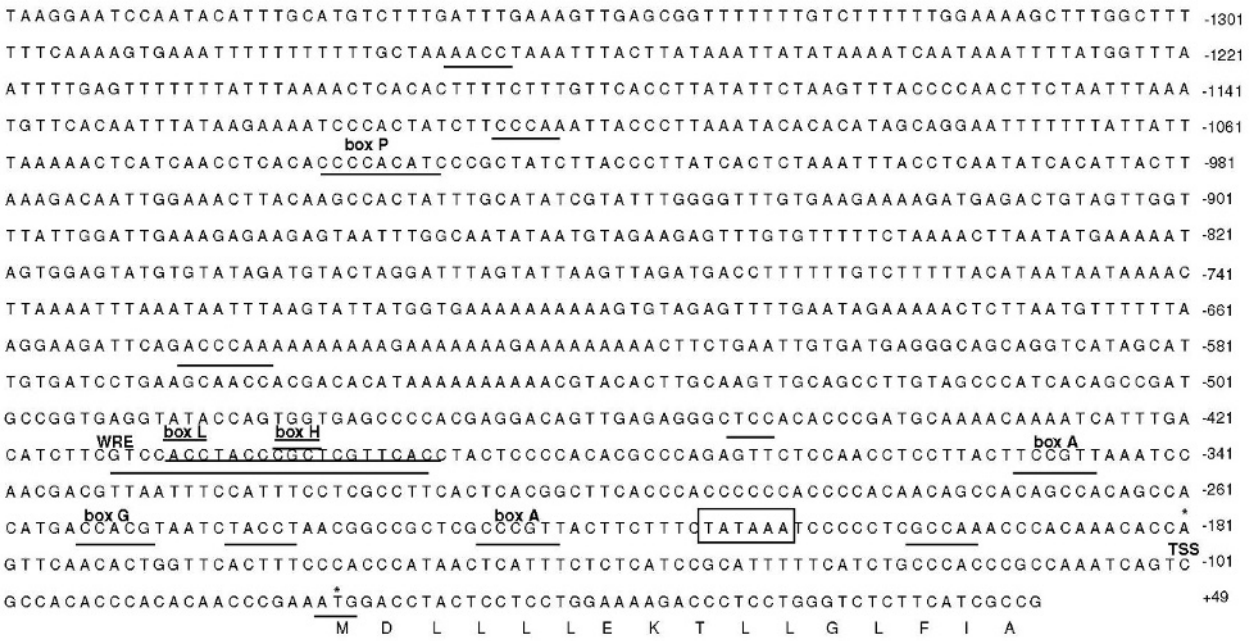

Fig. 3. Putative promoter region in the apricot cv. Goldrich C4H BAC clone. A putative TATA box is *boxed* and putative *cis*-acting elements homologous to sequence motifs for boxes P, A, H, G and L are *underlined*. One putative wound response element is indicated as WRE. The transcription start site (TSS) and the initiator ATG are indicated with *asterisks*.

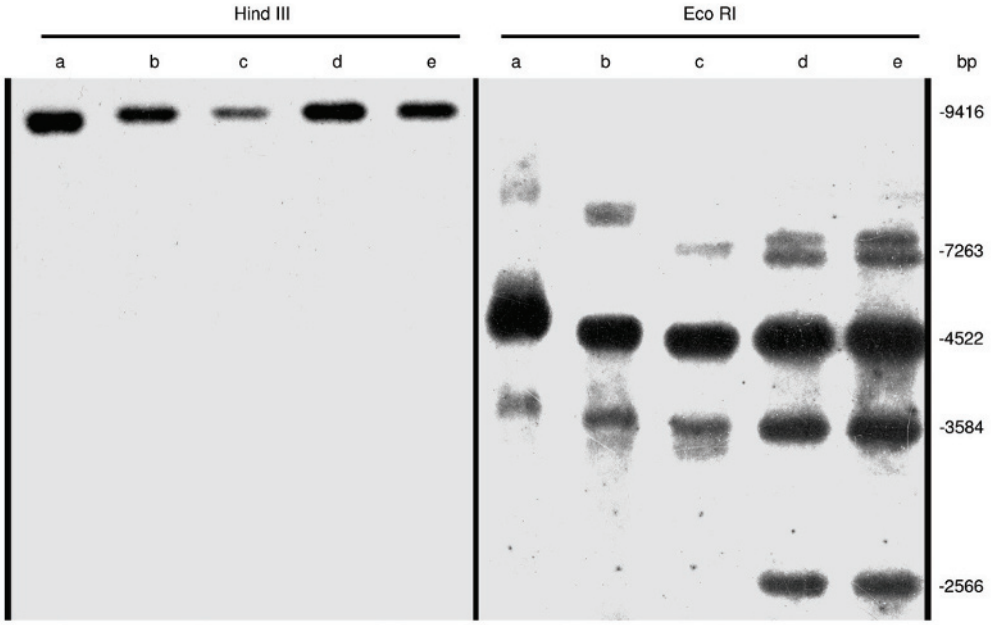

Fig. 4. Southern blot with genomic DNA from *Prunus* species with different polyploidy. Peach cv. Lovell (*lane a*), sour cherry cv. Montmorency (*lane b*), plum cvs. Reina Claudia (*lane c*) and Marianna 2624 (*lane d*) and apricot cv. Moniqui (*lane e*). Genomic DNA was digested with *Eco*RI and *Hind*III, fractioned on 0.8 %-agarose gel and blotted to a *Hybon* N^+ membrane, then hybridized with ^{32}P-labeled C4H cDNA.

Tetraploid sour cherry (cv. Montmorency) and hexaploid plum (cv. Reine Claude) did not show additional restriction fragments compared with diploid *Prunus* species (peach, apricot and plum). This result indicated that all *Prunus* species have only one copy of Class I PruC4H gene independently of ploidy level.

To examine the effect of grafting on Class I C4H transcripts, equal amounts of total RNA from homografts and heterografts during two weeks after grafting were subjected to semi-quantitative RT-PCR. The results demonstrated that C4H is expressed in both apricot cultivars Paviot and Moniqui under different conditions:

ungrafted controls, wounding, homografts and heterografts at 10 d after grafting (DAG). The results indicated that PruC4H has a similar steady state level in the combinations formed by the compatible cv. Paviot (Fig. 5*A*). Likewise, expression analysis showed that C4H transcript accumulation is not influenced by incompatibility reaction (Fig. 5*B*). Although there was a slight increase in band intensity in the incompatible heterograft (Moniqui grafted on Marianna) it was not significant compared with the ungrafted control or the compatible homograft (Moniqui grafted on Moniqui) at 10 d after grafting.

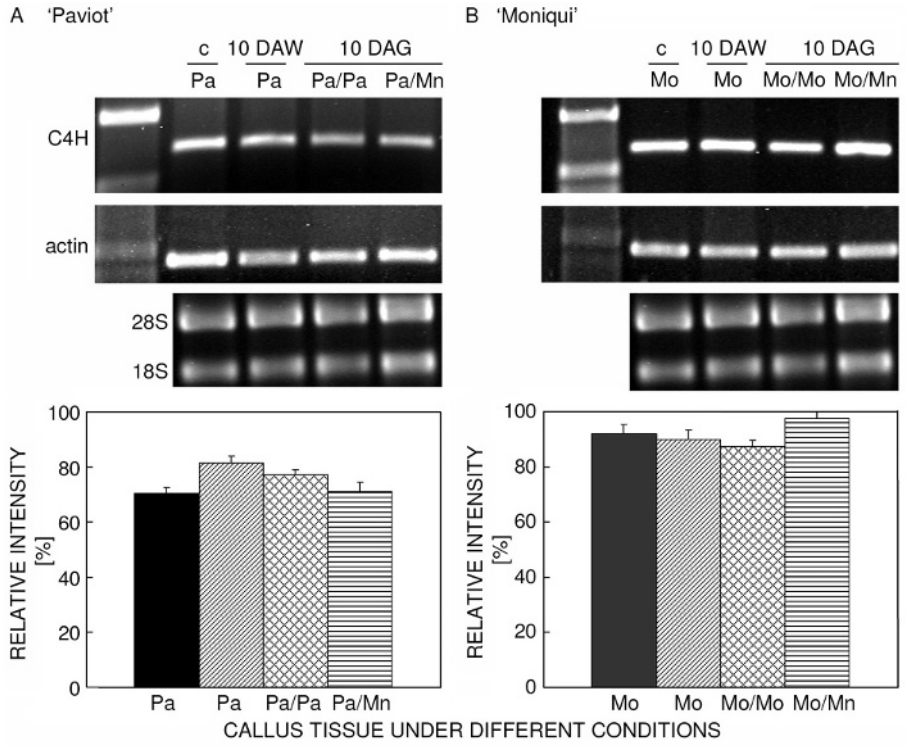

Fig. 5. Semiquantitative RT-PCR analysis of Class I C4H transcripts from two different apricot cultivars under different conditions: ungrafted control callus, calli tissue 10 d after wounding (DAW), graft interface from compatible homografts and heterografts with different degree of compatibility 10 d after grafting (DAG). *A* - apricot cv. Paviot (Pa), compatible with the plum rootstock Marianna 2624 (Mn). *B* - apricot cv. Moniqui (Mo), incompatible with the plum rootstock Marianna 2624 (Mn). Actin was used in RT-PCR amplification as the control. The intensities of RT-PCR products separated by agarose gel were measured by densitometry, as described in text. The intensities expressed are relative to the highest value, which was set to 100 %. Ribosomal RNA served for control amplifications to document the equivalent amounts of template.

Discussion

In this work, we have isolated a gene encoding C4H from *Prunus* species, apricot and plum. As it is shown in *ClustalW* alignment (Fig. 1), PruC4Hs are extremely conserved on protein and nucleotide level. The deduced amino acid sequences of PruC4H share the highest identity with C4H from *Malus domestica* (95 %) and *Rubus occidentalis* (94 %). Manual inspection of sequence alignment for presence of functional domains indicated that PruC4H might have the same catalytic activity as other C4H proteins. Indeed, PruC4H have all the P_{450} featured motifs, such as the hinge motif P_{34}PGPIPVP at the N terminus, the threonine-containing binding pocket motif A_{306}AIETT, the E_{363}-R_{366}-R_{421} triad and the heme-binding domain P_{439}FGVGRRSCPG near the C terminus which is essential for catalysis (see reviews Chapple 1998, Chen *et al*. 2007). Likewise, the multiple sequence alignment on Fig. 1 demonstrates presence of five substrate-specific recognition sites (SRS): SRS1 (S_{100}RTRNVVFDIFTGEGQDMVFTVY), SRS2 (L_{216}AQSFDYNY), SRS4 (I_{299}VENINVAAIE TTLWS), SRS5 (R_{368}MAIPLLVP) and SRS6 (K_{484}GGQFSLHI). Residues involved into interaction with aromatic ring are distributed in SRS5, SRS6 and the C-terminal portion of SRS4. Whereas, SRS1, SRS2 regions and the N-terminal end of the SRS4 region are involved in interaction with the aliphatic regions of the substrates (Ruspanhinge *et al*. 2003). C4H *Prunus* sequences have all the five SRS-signature motifs with 100 % identities to typical residue. Furthermore, they also contain the residues N302, I371 and K484, which are essential in substrate docking and orientation (Schoch *et al*. 2003).

Based on sequence similarity, C4H proteins have been classified into two classes, class I and class II according to different N- and C-terminus (Nedelkina *et al*. 1999, Betz *et al*. 2001). Phylogenetic analysis on Fig. 2 revealed that PruC4H shows identity rather with the C4Hs of Class I than with those of Class II. Based on this result, we assigned PruC4H to Class I, which is common in plants.

Most of known ClassI C4H genes have conserved exon-intron structure with variation in intron size from 85 and 220 bp for introns I and II in *Arabidopsis* (Bell-Lelong *et al*. 1997) to 205 and 1 172 bp for introns I and II in PhC4H from *Parthenocissus henryana* (Lu *et al*. 2006). Intron-exon structure of PruC4H is in agreement with that in other species. The apricot *C4H* gene was interrupted by two introns 335 and 904 bp long. Moreover, both introns have the standard GT/AG splicing sites and enriched A+T content, which is typical for plant C4H genes.

Presence of conserved motifs in promoter regions is a molecular basis for coordinated expression of the genes involved into general phenylpropanoid pathway (Logemann *et al*. 1995). Promoter analysis of *PruC4H* ClassI gene (Fig. 3) identified P, G, A and L boxes conserved in early phenylpropanoid promoters (Logemann *et al*. 1995). Apparently, the expression of *PruC4H* is coordinated with accumulation of PAL and 4CL transcripts. At -329 from ATG start-codon, the C4H promoter contains conserved motif CCTACC$(N)_{11}$CT as in *Arabidopsis* (Bell-Lelong *et al*. 1997). This motif is similar to the H-box sequence CCTACC$(N)_7$CT (Loake *et al*. 1992), essential for both light regulation and elicitor induction. However, neither TCA motif nor AT-rich elements conserved in various stress-induced genes (Goldsbrough *et al*. 1993, Kato *et al*. 1995) were identified as being responsible for the elicitor-induced expression of C4H.

Independently of ploidy level we found only one copy of Class I PruC4H gene in all *Prunus* species (Fig. 4). Seemingly, whole-genome duplications events in both dicot and monocot plant lineages are often accompanied by convergent loss of protein functional domain-containing genes (Rong *et al*. 2010). Conserved domain for C4H monooxygenase (PF00067) was listed among 95 protein conserved domains which have a tendency to be restored to a 'singleton' state after independent genome duplication events (Paterson *et al*. 2006). Molecular cloning and characterization of PAL and 4CL from *Prunus* species, which is currently in progress in our lab, will give more insight on number of individual genes in general phenypropanoid network in *Prunus* species.

In almost all plant tissues, activity of phenylpropanoid pathway is required for biosynthesis of lignin, seed coat pigments, and soluble metabolites such as sinapate esters and flavonoids (Chapple 1998). Therefore, C4H is constitutively expressed in many tissues and moderately induced by a wide range of stimuli (Mizutani *et al*. 1997). Besides, C4H may play important roles in non-lignification processes, *e g.*, flavonoid biosynthesis, due to the high level expression of C4H in lowly lignified organs (Chen *et al.* 2007). In this report, the C4H expression was detected in control callus tissue that supports involvement of C4H in flavonoid biosynthesis. The pattern of C4H gene expression showed that the level of control non-grafted tissue is detectable and maintained 10 d after wounding and after grafting. As it is displayed in Fig. 5, the C4H transcript accumulation is similar in compatible homografts and (in-)compatible heterografts formed by both apricot cultivars, Paviot and Moniqui. Both genotypes show good and poor graft compatibility, respectively, when are grafted on plum rootstocks (Errea and Felipe 1993, Errea *et al*. 2001). Apparently, Class I C4H expression pattern was not significantly coordinated with PAL expression under graft incompatibility reactions at 10 DAG in *Prunus* spp (Pina and Errea 2008). In all graft combinations, transcription of *Prunus* class I *C4H* was detected using semi-quantitative RT-PCR technique. However, we failed to detect C4H induction in incompatible heterograft. This result supports the constitutive character of C4H of Class I in *Prunus* species. Most likely, its expression is not regulated under the stock-scion interactions between graft partners of different degrees of compatibility. Elevated production of phenylpropanoid metabolites is a well documented fact in graft-incompatible combinations (Errea *et al.* 1992, Usenik *et al.* 2006, Mngómba *et al.* 2008, Pina and Errea 2005, 2008). So, two potential candidates for key regulation of an entrance point into phenyl-propanoid pathway, PAL and 4CL, should be characterized in future investigations. Further research is needed to elucidate control expression of the phenyl-propanoid genes during the graft union development processes and through the graft incompatibility reactions that potentially affect the fate of the graft.

Conclusion

Molecular cloning of C4H confirmed that the gene is well conserved in *Prunus* species and that it contains the pfam domain (PF00067), found as a 'duplication-resistant' gene (Paterson *et al*. 2006). Our results have revealed that a copy number of Class I C4H in *Prunus* is not increased with increased ploidy level. This fact can be used in identification of homology groups in polyploid cherries and plums. Population genetic models predict that the majority of gene losses after whole genome duplication should be relatively rapid. With the increasing knowledge of phenylpropanoid genes in *Prunus*, it will be possible to have a greater understanding of their regulation in relation to different physiological phenomenon, such as graft incompatibility or resistance to abiotic and biotic stress in *Prunus*.

References

Bell-Lelong, D.A., Cusumano, J.C., Meyer, K., Chapple, C.: Cinnamate-4-hydroxylase expression in *Arabidopsis* - regulation in response to development and the environment. - Plant Physiol. **113**: 729-738, 1997.

Betz, C., McCollum, T.G., Mayer, R.T.: Differential expression of two cinnamate 4-hydroxylases in 'Valencia' orange (*Citrus sinensis* Osbeck). - Plant. mol. Biol. **46**: 741-748, 2001.

Chapple, C.: Molecular genetic analysis of plant cytochrome P450-dependent monooxygenases. - Annu. Rev. Plant Physiol. Plant mol. Biol. **49**: 311-343, 1998.

Chaw, S.M., Zharkikh, A., Sung, H.M., Lau, T.C., Li, W.H.: Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. - Mol. Biol. Evol. **14**: 56-68, 1997.

Chen, A.H., Chai, Y.R., Li, J.N., Chen, L.: Molecular cloning of two genes encoding cinnamate 4-hydroxylase (C4H) from oilseed rape (*Brassica napus*). - J. Biochem. mol. Biol. **40**: 247-260, 2007.

Ehlting, J., Buttner, D., Wang, Q., Douglas, C.J., Somssich, I.E., Kombrink, E.: Three 4-coumarate : coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. - Plant J. **19**: 9-20, 1999.

Ehlting, J., Hamberger, B., Million-Rousseau, R., Werck-Reichhart, D.: Cytochromes P450 in phenolic metabolism. - Phytochem. Rev. **5**: 239-270, 2006.

Errea, P., Felipe, A.: Graft compatibility in apricot (*Prunus armeniaca* L.). - Invest Agr. Ser. Prod. Veg. **8**: 67-77, 1993.

Errea, P., Garay, L., Marin, J.A.: Early detection of graft incompatibility in apricot (*Prunus armeniaca*) using *in vitro* techniques. - Physiol Plant. **112**: 135-141, 2001.

Errea, P., Treutter, D., Feucht, W.: Scion-rootstock effects on the content of flavan-3-ols in the union of heterografts consisting of apricots and diverse *Prunus* rootstocks. - Gartenbauwissenschaft **57**: 134-138, 1992.

Fahrendorf, T., Dixon, R.A.: Stress responses in alfalfa *(Medicago sativa* L.) XVIII: molecular cloning and expression of the elicitor-inducible cinnamic acid 4-hydroxylase cytochrome P450. - Arch. Biochem. Biophys. **305**: 509-515, 1993.

Ge, L., Chiang, V.L.: A full length cDNA encoding trans-cinnamate 4-hydroxylase from developing xylem of *Populus tremuloides*. - Plant Physiol. **112**: 861-864, 1996.

Goldsbrough, A.P., Albrecht, H., Stratford, R.: Salicylicacid-inducible binding of a tobacco nuclear protein to a 10 bp sequence which is highly conserved among stress inducible genes. - Plant J. **3**: 563-571, 1993.

Hormaza, J.I.: Molecular characterization and similarity relationships among apricot (*Prunus armeniaca* L.) genotypes using simple sequence repeats. - Theor. appl. Genet. **104**: 321-328, 2002.

Horn, R., Lecouls, A.C., Callahan, A., Dandekar, A., Garay, L., McCord, P., Howad, W., Chan, H., Verde, I., Ramaswamy, K., Main, D., Jung, S., Georgi, L., Forrest, S., Mook J., Zhebentyayeva, T.N., Yu, Y., Kim, H.R., Jesudurai, C., Sosinski, B.A., Arús, P., Baird, V., Parfitt, D., Reighard, G., Scorza, R., Tomkins, J., Wing, R., Abbott, A.G.: Candidate gene database and transcript map for peach, a model species for fruit trees. - Theor. appl. Genet. **110**: 1419-1428, 2005.

Huang, B., Duan, Y., Yi. B., Sun. L., Lu. B., Yu. X., Sun. H., Zhang, H, Chen, W.: Characterization and expression profiling of cinnamate 4-hydroxylase gene from *Salvia miltiorrhiza* in rosmarinic acid biosynthesis pathway. - Russ. J. Plant Physiol. **55**: 390-399, 2008.

Kato, H., Wada, M., Muraya, K., Malik, K., Shiraishi, T., Ichinose,Y., Yamada, T.: Characterization of nuclear factors for the elicitor-mediated activation of the promoter of the pea phenylalanine ammonia-lyase gene 1. - Plant Physio1. **108**: 129-139, 1995.

Kawai, S., Mori, A., Shiokawa, T., Kajita, S., Katayama, Y., Morohoshi, N.: Isolation and analysis of cinnamic acid 4-hydroxylase homologous genes from a hybrid aspen, *Populus kitatamiensis*. - Biosci. Biotechnol. Biochem. **60**: 1586-1597, 1996.

Kumar, A., Ellis, B.E.: The phenylalanine ammonia-lyase gene family in raspberry. Structure, expression, and evolution. - Plant Physiol. **127**: 230-239, 2001.

Kumar, A., Ellis, B.E.: 4-Coumarate : CoA ligase gene family in *Rubus idaeus*: cDNA structures, evolution, and expression. - Plant mol. Biol. **51**: 327-340, 2003.

Kumar, K., Tamura, N.M.: MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. - Briefings Bioinform. **5**: 150-163, 2004.

Loake, G.J., Faktor, O., Lamb, C.J., Dixon, R.A.: Combination of H-box [CCTACC(N)7CT] and G-box (CACGTG) cis elements is necessary for feed-forward stimulation of a chalcone synthase promoter by the phenypropanoidpathway intermediate p-coumaric acid. - Proc. nat. Acad. Sci. USA **89**: 9230-9234, 1992.

Logemann, E., Parniske, M., Hahlbrock, K.: Modes of expression and common structural features of the complete phenylalanine ammonia-lyase gene family in parsley. - Proc. nat. Acad. Sci. USA **92**: 5905-5909, 1995.

Lu, S.F., Zhou, Y.H., Li, L.G., Chiang, V.L.: Distinct roles of cinnamate 4-hydroxylase genes in *Populus*. - Plant Cell Physiol. **47**: 905-914, 2006.

Mizutani, M., Ward, E., Dimao, J., Ohta, D., Ryals, J., Sato, R.: Molecular cloning and sequncing of a cDNA encoding mung bean cytochrome-P450 (P450 C4H) possessing cinnamate 4-hydroxylase activity. - Biochem. biophys. Res. Comm. **190**: 875-880, 1993.

Mizutani, M., Ohta, D., Sato, R.: Isolation of a cDNA and a genomic clone encoding cinnamate 4-hydroxylase from *Arabidopsis* and its expression manner *in planta*. - Plant Physiol. **113**: 755-763, 1997.

Mngómba, S.A, Toit, E., Akinnifesi, F.: The relationship between graft incompatibility and phenols in *Uapaca kirkiana* Müell Arg. - Sci Hort **117**: 212-218, 2008.

Nedelkina, S., Jupe, S.C., Blee, K.A., Schalk, M., Werck-Reichhart, D., Bolwell. G.P.: Novel characteristics and regulation of a divergent cinnamate 4-hydroxylase (CYP73A15) from French bean: engineering expression in yeast. - Plant mol. Biol. **39**: 1079-1090, 1999.

Paterson, A.H., Chapman, B.A., Kissinger, J.C., Bowers, J.E., Feltus, F.A., Estill, J.C.: Many genes and domain families have convergent fates following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces* and *Tetraodon* - Trends Genet. **22**: 597-602, 2006.

Pina, A., Errea, P.: A review of new advances in mechanism of graft compatibility-incompatibility. - Sci. Hort. **106**: 1-11, 2005.

Pina, A., Errea, P.: Differential induction of phenylalanine ammonia-lyase gene expression in response to *in vitro* callus unions of *Prunus* spp. - J. Plant Physiol. **165**: 705-714, 2008.

Raes, J., Rohde, A., Christensen, J.H., Van de Peer, Y., Boerjan,

W.: Genome-wide characterization of the lignification toolbox in *Arabidopsis*. - Plant Physiol. **133**: 1051-1071, 2003.

Rong, J., Feltus, F.A., Liu, L., Lin, L., Paterson, A.H.: Gene copy number evolution during tetraploid cotton radiation. - Heredity **105**: 463-472, 2010.

Rupasinghe, S., Baudry, J., Schuler, M.A.: Common active site architecture and binding strategy of four phenylpropanoid P450s from *Arabidopsis thaliana* as revealed by molecular modeling. - Protein Eng. **16**: 721-731, 2003.

Sambrook, J., Russell, D.: Molecular Cloning: a Laboratory Manual. - Cold Spring Harbour Laboratory Press, Cold Spring Harbour - New York 2001.

Schilmiller, A.L., Stout, J., Weng, J.K., Humphreys, J., Ruegger, M.O., Chapple, C.: Mutations in the cinnamate 4-hydroxylase gene impact metabolism, growth and development in *Arabidopsis*. - Plant J. **60**: 771-782, 2009.

Schoch, G.A., Attias, R., Le Ret, M., Werck-Reichhart, D.: Key substrate recognition residues in the active site of a plant cytochrome P450, CYP73A1 - homology model guided site-directed mutagenesis. - Eur. J. Biochem. **270**: 3684-3695, 2003.

Sullivan, M.L.: Phenylalanine ammonia lyase genes in red clover: expression in whole plants and in response to yeast fungal elicitor. - Biol. Plant. **53**: 301-306, 2009.

Teutsch, H.G., Hasenfratz, M.P., Lesot, A., Stoltz, C., Garnier, J.M., Jeltsch, J.M., Durst, F., Werckreichhart, D.: Isolation and sequence of a cdna-encoding the jerusalem-artichoke cinnamate 4-hydroxylase, a major plant cytochrome P_{450} involved in the general phenylpropanoid pathway. - Proc. nat. Acad. Sci. USA **90**: 4102-4106, 1993.

Treutter, D.: Significance of flavonoids in plant resistance: a review. - Environ. Chem. Lett. **4**: 147-157, 2006.

Usenik, V., Krška B., Vičan, M., Štampar, F.: Early detection of graft incompatibility in apricot (*Prunus armeniaca* L.) using phenol analyses. - Sci. Hort. **109**: 332-338, 2006.

Vilanova, S., Romero, C., Abernathy, D., Abbott, A.G., Burgos, L., Llacer, G., Badenes, M.L.: Construction and application of a bacterial artificial chromosome (BAC) library of *Prunus armeniaca* L. for the identification of clones linked to the self-incompatibility locus. - Mol. Genet. Genomics **269**: 685-691, 2003.

Vogt, T.: Phenylpropanoid biosynthesis. - Mol. Plant. **3**: 2-20, 2010.

Whitbred, J.M., Schuler, M.A.: Molecular characterization of *CYP73A9* and *CYP82A1* P450 genes involved in plant defense in pea. - Plant Physio1. **124**: 47-58, 2000.

Yang, D.H., Chung, B.Y., Kim, J.S., Kim, J.H., Yun, P.Y., Lee, Y.K., Lim, Y.P., Lee, M.C.: cDNA cloning and sequence analysis of the rice cinnamate-4-hydroxylase gene, a cytochrome P450-dependent monooxygenase involved in the general phenylpropanoid pathway. - J. Plant Biol. **48**: 311-318, 2005.