Plant Molecular Biology **31:** 327–340, 2003.

4-Coumarate:CoA ligase gene family in *Rubus idaeus*: cDNA structures, evolution, and expression

Amrita Kumar[1] and Brian E. Ellis*
*Biotechnology Laboratory and Faculty of Agricultural Sciences, University of British Columbia, Bioscience Building Room 3508, 6270 University Boulevard, Vancouver V6T 1Z4, Canada (*author for correspondence, e-mail bee@unixg.ubc.ca); [1]Present address: Department of Microbiology and Immunology, Emory University School of Medicine, 1510 Clifton Road, Atlanta, GA 30322, USA*

Received 11 December 2001; accepted in revised form 6 June 2002

Key words: 4CL, fruits, gene family, raspberry, *Rubus*

Abstract

The enzyme 4-coumarate:CoA ligase (4CL) activates cinnamic acid and its hydroxylated derivatives by forming the corresponding CoA thioesters. These serve as substrates for biosynthesis of phenylpropanoid-derived end-products that are important determinants of fruit quality in raspberry (*Rubus idaeus* L.). In higher plants, 4CL is typically encoded by a gene family. To investigate the participation of distinct *4CL* genes in the process of fruit ripening, we have characterized this gene family in raspberry. By complementing a PCR-based homology search with low-stringency cDNA library screening, we have isolated three classes of raspberry *4CL* cDNAs (*Ri4CL1*, *Ri4CL2*, and *Ri4CL3*). Phylogenetic analysis places the three raspberry *4CL* gene family members into two distinct groups, a pattern consistent with an ancient divergence from an ancestral progenitor. Quantitative RT-PCR assay reveals a differential pattern of transcription of each of the three genes in various organs, as well as distinct temporal patterns of expression during flower and fruit development. The regulatory elements thus appear to have evolved independently of the genes themselves. Based on phylogenetic classification, expression patterns and recombinant protein activities the different *Ri4CL* genes are likely to participate in different biosynthetic pathways leading to the various phenylpropanoid-derived metabolites that help create flavor and color in raspberry fruit.

Abbreviations: 4CL, 4-coumarate:CoA ligase; PAL, phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; CHS, chalone synthase (CHS); *Ri4cl1*, *Rubus idaeus* 4-coumarate:CoA ligase 1; *Ri4cl2*: *Rubus idaeus* 4-coumarate:CoA ligase 2; *Ri4cl3*, *Rubus idaeus* 4-coumarate: CoA ligase 3; *RiHisH3*, *Rubus idaeus* histone H3

Introduction

Channeling of photosynthetically fixed carbon through the phenylpropanoid pathway in plants requires the sequential action of three enzymes: phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), and 4-coumarate:CoA ligase (4CL; EC 6.2.1.12), which together constitute the general phenylpropanoid pathway (Hahlbrock and Grisebach, 1979). As the last enzyme of this pathway, 4CL is responsible for the activation of cinnamic acid and its derivatives to their corresponding thioesters (Figure 1). These esters are central intermediates in the synthesis of more highly modified phenylpropanoid compounds that are required for various physiological functions and for adaptation to environmental perturbations (Dixon and Paiva, 1995).

In all plants examined so far, the *4 CL* occurs as gene families consisting of two to three members. In some species, such as *Arabidopsis* (Ehlting *et al.*, 1999), aspen (Hu *et al.*, 1998), and soybean (Uhlmann and Ebel, 1993), specific *4CL* genes appear to be associated with the formation of unique phenylpropanoid end-products. In others, multiple forms of 4CL appear to possess identical or nearly identical catalytic properties, for example potato (Becker-André *et al.*, 1991),

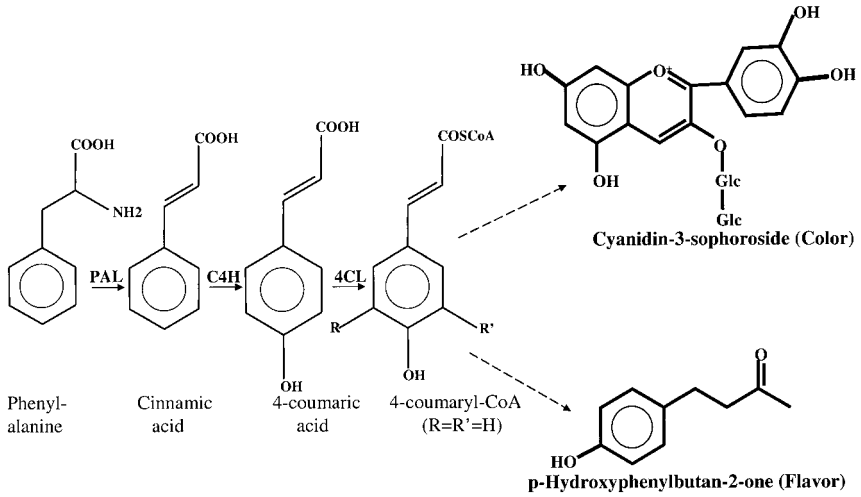

Figure 1. Phenylpropanoids that contribute to quality traits in raspberry fruits. Dashed arrows indicate a branch pathway emanating from the general phenylpropanoid pathway. The enzymes of general phenylpropanoid metabolism consist of PAL, phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; and 4CL, 4-coumarate:CoA ligase. The R and the R′ groups on hydroxycinnamic acids may be hydroxyl and/or methoxy groups.

parsley (Lozoya *et al.*, 1988), pine (Zhang and Chiang, 1997), and hybrid poplar (Alina *et al.*, 1998). In aspen, the *Pt4CL1* and *Pt4CL2* gene products are both structurally and functionally distinct, and the two genes are expressed in a compartmentalized manner (Hu *et al.*, 1998). *Pt4CL1* is expressed in lignifying xylem and the corresponding recombinant protein prefers substrates such as ferulic acid and 5-hydroxyferulic acid, suggesting that it plays a specific role in lignification (Hu *et al.*, 1998). This proposed role for *Pt4CL1* is corroborated by the results of antisense-mediated down-regulation of *Pt4CL1* in aspen, which led to accumulation of substantially reduced levels of lignin (Hu *et al.*, 1999). Aspen Pt4CL2, on the other hand, shows highest activity towards 4-coumaric acid, with little activity towards ferulic acid, and is expressed in leaf and stem epidermis but not in developing xylem tissue, suggesting that it may play a role in the biosynthesis of non-lignin phenylpropanoids (Hu *et al.*, 1998).

In *Arabidopsis*, recombinant At4CL1 has a high specificity for 4-coumaric acid and caffeic acid, and is expressed in stems, roots (Lee *et al.*, 1995), siliques and, to a lesser extent, in leaves (Ehlting *et al.*, 1999). At4CL2 is unusual in being able to convert caffeic acid with a higher efficiency than 4-coumaric acid while not acting on ferulic acid at all (Ehlting *et al.*, 1999). *At4CL2* is highly expressed in roots and siliques. At4CL3, on the other hand, displays a strong preference for 4-coumaric acid as a substrate relative to caffeic acid and ferulic acid, and is expressed at high levels in flowers but not in lignified organs. This pattern suggests that the primary function of the At4CL3 isoform is to provide activated 4-coumaric acid for the chalcone synthase (CHS) reaction that feeds the flavonoid-specific branch pathways (Ehlting *et al.*, 1999). In soybean, the two discrete *4CL* genes characterized, *Gm4CL14* and *Gm4CL16*, have been proposed to correspond to two distinct 4CL enzyme isoforms identified in that species, consistent with the idea that discrete *4CL* genes can encode function-specific isoforms (Knobloch and Hahlbrock, 1977; Uhlmann and Ebel, 1993).

4CL gene expression is transcriptionally regulated and can be activated both during development and by external stimuli such as pathogen infection, elicitor treatment, wounding, methyl jasmonate treatment, and UV irradiation (Douglas *et al.*, 1987; Schmelzer *et al.*, 1989; Wu and Hahlbrock, 1992; Lee *et al.*, 1995). This activation of *4CL* upon elicitor treatment or UV irradiation has been shown to occur coordinately with the activation of *PAL*, the first enzyme of general phenylpropanoid metabolism (Chappell and Hahlbrock, 1984; Logemann *et al.*, 1995). This, together with other indirect evidence, has led to the suggestion that labile multi-enzyme complexes within the core phenylpropanoid pathway may provide integrated control of the temporal and cell-type-specific accumulation of discrete phenylpropanoid end-products. If so, distinct *4CL* genes might be employed to

divert hydroxycinnamoyl-CoA thioesters to specific-branch phenylpropanoid pathways in plant tissues. The *4CL* gene family in raspberry is of particular interest since maturing raspberry fruits accumulate a novel phenylpropanoid derivative, p-hydroxyphenylbutan-2-one ('raspberry ketone'), which is a primary determinant of raspberry flavor (Borejsza-Wysocki and Hrazdina, 1994 and references therein). Biosynthesis of 'raspberry ketone' has been shown to proceed via a two-step pathway branching from the general phenylpropanoid pathway (Figure 1). The first enzyme in this branch, benzalacetone synthase, catalyzes a reaction analogous to that catalyzed by CHS, with 4-coumaroyl-CoA and malonyl-CoA as substrates (Borejsza-Wysocki and Hrazdina, 1994). Raspberry fruits also accumulate other phenylpropanoid-derived metabolites, notably anthocyanin pigments (Goiffon *et al.*, 1991) whose biosynthesis is also dependent on activated 4-coumaric acid as a precursor. This combination of different 4-coumarate-derived end-products in one tissue raises the possibility that distinct *4CL* genes might be associated with the biosynthesis of individual metabolic end-products in ripening raspberry fruits. To address this question, it was first necessary to characterize the *4CL* gene family in this species.

Materials and methods

Plant materials

Raspberry (*Rubus idaeus* L. cv. Meeker) plants were grown in the experimental plots of the Agriculture and Agri-Food Canada Research Station at Abbotsford, BC, or in the greenhouse (Faculty of Agricultural Sciences, UBC) under ambient conditions. All harvested plant tissues were immediately frozen in liquid nitrogen and stored at −80 °C.

Raspberry leaf, shoot and root tissues were collected from mature greenhouse-grown plants. Flowers and fruits at different developmental stages were collected from a pool of field-grown plants. Flowers were collected at three developmental stages. Flowers I consisted of closed inflorescence buds. Flowers II consisted of fully open flowers, and Flowers III consisted of fertilized flowers. Fruits were collected at five different developmental stages. Fruits I were green, hard and still undergoing cell expansion; Fruits II were still green but had almost reached mature size; Fruits III were yellow, starting to 'bush', and had reached full size; Fruits IV were fully ripe, with the color and aroma fully developed; Fruits V were slightly overripe and somewhat dehydrated.

Design of degenerate PCR primers for 4CL *genes*

The following sequences were aligned to generate a 4CL amino acid consensus (numbers in parenthesis denote the GenBank accession numbers): *Arabidopsis* 4CL1 (S57784), soybean 4CL14 (X69954), soybean 4CL16 (X69955), *Lithospermum erythrorhizon* 4CL1 (D49367), *L. erythrorhizon* 4CL2 (D49366), rice 4CL1 (L43362), rice 4CL2 (X52623), parsley 4CL1 (X13324), parsley 4CL2 (X13325), loblolly pine 4CL1 (U12012), loblolly pine 4CL2 (U12013), potato 4CL1 (M62755), and potato 4CL2 (AF150686). An amino acid alignment of these sequences indicated that, although the N-terminal region of the 4CL protein is highly variable, there are small stretches within this region that are partially conserved and thus could be used as primer targets.

Primers C1 (5′-cggaattcG A/G TCIAA G/A CTIC-CIGA T/C AT) and C3 (5′-cggaattcG A/G TCIAA A/G TT A/G CCIGA C/T AT) target a conserved seven amino acid sequence (RSKLPDI) located near the N-terminus of the protein. To decrease primer complexity, degenerate codons were split between the two primers, C1 and C3. C1 and C3 are thus sister primers, where C1 accommodates the leucine codons CUU, CUC, CUA, and CUG and C3 uses codons UUA and UUG. Primers C2 (5′-gctctagaGA G/A TA A/G/T AT A/G TG A/G TG AG AAIAG A/G GG) and C4 (5′gctctagaGA G/A TA A/G/T AT G/A TG A/G AA T/C A A A/G G) target a seven amino acid region (P L/M FHI Y/F S/A) corresponding to residues 255–261 in *Arabidopsis* 4CL1. Alternative amino acids were found in this region in different 4CL proteins, such as leucine/methionine, tyrosine/phenylalanine, serine/alanine. Sister primers C2 and C4 thus target the same position but collectively they account for all the codons of the variant amino acids. Similarly, sister primers C6 (5′-gctctagaAGTIAGCAT C/A ACICC C/T TT) and C8 (5′-gctctagaGCT T/A AACAT C/A ACICC C/T TT) target the peptide KGVMLT, and together they account for the six leucine codons.

Restriction enzyme recognition sites (lower case letters within the primers) were incorporated at the 5′ end of each primer to facilitate cloning of the amplified products into pUC19. Forward primers C1 and C3 (odd-numbered/sense-strand primers) have an *Eco*RI site while reverse primers C2, C4, C6 and

C8 (even-numbered/antisense primers) incorporated a recognition site for *Xba*I.

Gene amplification and characterization

Genomic DNA was isolated using the method described by Doyle and Doyle (1990). Amplification reactions contained 100 ng genomic DNA, 1× Appligene buffer (providing a final concentration of 1.5 mM $MgCl_2$), 0.5 μM each dNTP, 2 μM each primer and 2.5 U *Taq* DNA polymerase. Reaction mixtures were incubated at 95 °C for 10 min and then subjected to 35 cycles of amplification (95 °C for 50 s, 55 °C for 50 s, 72 °C for 1 min), and completed by a final 10 min extension at 72 °C in a Techne PHC-3 thermal cycler (Mandel Scientific, Guelph, Ontario, Canada). PCR reaction products were analyzed by 1% w/v Tris-acetate-EDTA (TAE)-agarose gel electrophoresis. Amplified products of two independent PCR reactions with each of the eight primer combinations tested were subcloned into *Eco*RI- and *Xba*I-digested pUC19. Multiple clones from each subcloning experiment were amplified in a PCR reaction with vector-specific primers M13R and M13F to confirm the presence of inserts. Between 300 and 500 clones from each primer combination were subjected to RFLP analysis using eight different restriction enzymes (*Bam*HI, *Bgl*I, *Hinc*II, *Hin*dIII, *Eco*RI, *Eco*RV, *Sal*I, *Xba*I).

Sequencing and sequence analysis of PCR products

Plasmid DNA from selected clones was isolated for sequencing following a mini-alkaline lysis/PEG precipitation procedure (Ausubel *et al.*, 1995). Both strands of the insert were sequenced with the M13 universal primer and/or synthetic oligonucleotide primers as needed to extend the sequence. Sequencing reactions were carried out at the Nucleic Acid Protein Service Unit (NAPS, UBC) in an Applied Biosystems ABI 373 DNA sequencer.

All sequences were edited and analyzed with PC/GENE Software (Intelligenetics, Mountain View, CA). Database searches for sequence homology and comparisons were performed with various web-based analytical tools compiled at the website http://www.sdsc.edu/ResTools/.

Construction and screening of the raspberry cDNA library

Total RNA was isolated from raspberry fruits (stage III) with the RNeasy Maxi Kit (Qiagen, Missisauga, Ontario, Canada) according to the manufacturer's protocol. Poly(A)$^+$ RNA was isolated from 1.5 mg total RNA with Dynabeads Oligo (dT_{25}) (Dynal, Lake Success, NY) following the manufacturer's instructions. A cDNA library was constructed from 5 μg poly(A)$^+$ RNA with a Uni-ZAP XR Library Construction Kit (Stratagene, La Jolla, CA). The cDNA library, consisting of ca. 10^7 independent clones, was amplified once to obtain high-titer stock.

About 5×10^5 plaques of the amplified cDNA library were blotted in duplicate onto Hybond N$^+$ nylon membrane (Amersham-Pharmacia, Mississauga, Ontario, Canada). The membranes were screened with a mixed population of genomic amplicon fragments *Ri4cl1* and *Ri4cl2*, radiolabeled to a high specific activity with [α-^{32}P]dATP using a Random Primer Labeling kit (Life Technologies, La Jolla, CA). The nylon membranes were hybridized at 50 °C for 16 h in hybridization buffer consisting of 6× SSC buffer, 20 mM NaH_2PO_4, 0.1% v/v SDS, 5× Denhardt's reagent, and 500 μg/ml denatured salmon sperm DNA. The membranes were washed three times (20 min each) at 50 °C with 6× SSC buffer, 0.1% v/v SDS. After the tertiary screen, inserts in positive clones were amplified with vector-specific primer T3 and a pool of *4CL* gene-specific primers (C6 and C8) to verify the presence and the size of the inserts. Representative plaques of each cDNA class were generated as pBluescript SK$^-$ phagemids via *in vivo* excision using the helper phage strain ExAssist (Stratagene).

Plasmids from each class were isolated using a commercial plasmid purification kit (Qiagen). Sequence accuracy was confirmed by the complementarity between the sense and the antisense strand of the DNA sequences and by translation of the sequences into amino acid sequences. Similar sequences in the databases were identified using BLAST (Altschul *et al.*, 1990).

QRT-PCR

Gene-specific primers capable of amplifying a segment from 201 to 452 of *Ri4CL1* cDNA were sense primer C15 (5′-CGACATCCACACCTACGCC) and antisense primer C13 (5′-ACTTCACTTCGTCGCATGAT). Gene-specific primers capable of amplifying a

segment from 266 to 531 of *Ri4CL2* cDNA were sense primer C25 (5′-TTCGCGAAGCTCAACGACG) and antisense primer C23 (5′-TCGTTGAGCTTCCGCGAA). Gene-specific primers capable of amplifying a segment from 375 to 625 of *Ri4CL3* cDNA were sense primer C35 (5′-CCTACACCTTCTCCGAAAC) and antisense primer C33 (5′-CCAACTTGAAAGTGCTGGCC). The specificity of each primer set was confirmed by their ability to amplify DNA from their cognate plasmid *Ri4CL* cDNAs (Figure 5). Gene-specific primers for amplification of a fragment of the raspberry *HistoneH3* (*RiHisH3*, GenBank AF304365) gene had the nucleotide sequences H15 (5′-ATGGCGCGGACGAAGGA-3′) and H13 (5′-GCCTACGCCGCCCGCTCAACCTA-3′).

A fixed amount of total RNA (100 ng) from various raspberry tissues was reverse-transcribed into cDNA with Omniscript Reverse Transcriptase (Qiagen) in a reaction volume (20 μl) containing 1× RT buffer, 0.5 mM each dNTP, 1 μM oligo-dT primer and 10 units RNase inhibitor (Amersham-Pharmacia). The reaction mixture was incubated at 37 °C for 1 h followed by 5 min incubation at 95 °C to destroy the RT enzyme. The first-strand cDNA reaction (1 μl) was amplified in a 20 μl volume containing 200 nM each PCR primer, 200 μM each dNTP, and 2.5 U *Taq* DNA polymerase (Qiagen) in 1× PCR buffer and 1× Q solution. The thermal cycling conditions were 94 °C for 5 min followed by 25 cycles for *RiHisH3* or 30 cycles of 94 °C for 20 s, 59 °C for 50 s, and 72 °C for 50 s and a final extension of 5 min at 72 °C for *Ri4CLs*. The PCR product (10 μl) was analyzed on a 3% w/v TAE-agarose gel and then stained with ethidium bromide (EtB, 10 mg/ml). Staining intensity was digitally quantified with Scion Image software (Scion Corporations, Frederick, MD). Relative intensities of the target bands were expressed as arbitrary units (AU). To normalize for equal amounts of total RNA and efficiency of cDNA synthesis from various tissue samples, the intensities of the bands were normalized with the average intensity of the *RiHisH3* product across the samples investigated. This analysis was repeated twice with consistent results and representative data from one analysis have been presented.

Cloning of cDNAs into bacterial expression vectors

Ri4CL3 was expressed in *Escherichia coli* as a His_6-tag fusion protein with vector pQE30 (Qiagen), or without the His_6 tag with vector pQE50 (Qiagen). The open reading frame of *Ri4CL3* cDNA was amplified with sense primer (5′-ACATGCATGCATGATATCCATTGCCTATAAT-3′) and antisense primer (5′-GGGTACCGGGCCCCCCCTCGAGGTC-3′) to introduce a unique *Sph*I site upstream of the start site and a *Xho*I site downstream of the stop codon. The *Sph*I/*Xho*I fragment containing the full-length coding region of *Ri4CL3* was subcloned into the *Sph*I/*Sal*I site of pQE50/pQE30 to create plasmids pQE50-4CL3/pQE30-4CL3, respectively.

Ri4CL1 and Ri4CL2 were over-expressed as recombinant protein using the streptag protein expression and purification system (IBA, Göttingen, Germany). The open reading frame of *Ri4CL1* was amplified with sense primer (5′-CGGGGTACCATGGCGGTCCAAACACCTCAAC) and antisense primer (5′-CCGCTCGAGCCTCAATTTGCAAAGCCAGCCAGCAG) to introduce unique *Kpn*I and *Xho*I sites at the 5′ and 3′ ends of the cDNA, respectively. The *Kpn*I/*Xho*I fragment was subcloned into the *Kpn*I/*Xho*I sites of pASK-IBA7 to create plasmid IBA7-4CL1. This, construct, however, contained an internal stop codon that expressed truncated Ri4CL1 recombinant proteins. To remove the internal stop codon, nucleotide T at position 1541 in the *Ri4CL1* cDNA was changed to C with the sense primer (5′-GCAATTTATTTCAAAACAGGTTGTATTCTAC) and antisense primer (5′-GTAGAATACAACCTGTTTTTGAAATAAATTGC) in the Quick- Change Site-Directed Mutagenesis Kit (Stratagene). This new expression plasmid expressed the full-length Ri4CL1 recombinant protein and was renamed IBA7-4CL1-1. The *Ri4CL2* cDNA was amplified with sense primer (5′-CGGGGTACCATGGCGAACAAACATCAAGACGAC) and antisense primer (5′-CGCGGATCCTCACTTGCTGGAAGTCTAATTGG) to create full-length *Ri4CL2* cDNA with *Kpn*I and *Bam*HI sites at the 5′ and 3′ ends, respectively. The *Kpn*I/*Xho*I fragment was cloned into *Kpn*I/*Bam*HI site of plasmid pASK-IBA7 to create plasmid IBA-4CL2.

All new constructs were fully sequenced to ensure the fidelity of the PCR product and to confirm that the cDNAs were in-frame with the start codon (ATG) of the expression vectors.

Protein expression and enzymatic assay

To produce samples for enzymatic assays, the pQE-based expression plasmids were transformed into *E. coli* strain M15 (Qiagen) and the pASK-IBA7-based plasmids were transformed into strain DH5α.

Expression vectors without any insert served as the negative control. Ri4CL3 recombinant protein was obtained by inducing 50 ml cultures in the logarithmic-phase (grown at 37 °C) with 1.0 mM isopropyl β-D-thiogalactopyranoside (IPTG) for 4 h, while Ri4CL1 and Ri4CL2 recombinant proteins were obtained by inducing 50 ml cultures in the logarithmic-phase (grown at 37 °C) with 2 mg/ml of anhydrotetetracycline for 3 h. After induction, the cultures were centrifuged and the bacterial pellet resuspended in 5 ml of 200 mM Tris pH 7.5, 1.5 mM 2-mercaptoethanol. Cells were disrupted in a French press at 7.6 MPa and the extract clarified by centrifugation at 10 000 $\times$ g for 20 min. The resulting supernatant containing the soluble protein was immediately used for further analysis.

Enzyme activity was assayed in a standard assay buffer containing 5 mM ATP, 5 mM $MgCl_2$, 0.33 mM Coenzyme A (CoA) and 0.2 mM various cinnamic acid or benzoic acid derivatives. Incubations were started by addition of CoA and carried out at 25 °C. The formation of the CoA esters was determined spectrophotometrically by monitoring the absorbance change relative to a control (no CoA), at wavelengths appropriate for cinnamoyl, hydroxycinnamoyl, benzoyl and hydroxybenzoyl-CoA derivatives (Webster *et al.*, 1974; Stöckigt and Zenk, 1975).

Phylogenetic analysis

Protein sequences were donwnloaded from the GenBank database and aligned with Clustal W (Altschul *et al.*, 1990). The alignment was manually optimized and used for finding the most parsimonious tree, employing the heuristic search option algorithm in the program PAUP 4.0b2 (Sinauer, Sunderland, MA). For statistical analysis, 1000 bootstrap replications (Felsenstein, 1985) were analyzed.

Results

PCR-based search for the raspberry 4CL *gene family*

Degenerate primers targeting several conserved regions within the *4CL* gene were designed as described in Materials and methods and presented in Figure 2. Genomic DNA amplifications with combinations of forward primer C1/C3 with reverse primer (C2/C4)/(C6/C8) amplified fragment sizes that matched the estimates for a *4CL* gene without intron sequences within the amplified regions. To differentiate among the PCR-amplified products obtained with each primer set, the amplified products were cloned, and multiple clones were fingerprinted with restriction digests. This led to the identification of two classes of clones based on restriction enzyme polymorphism. Class I clones did not have recognition sites for *Bgl*I, *Hin*cII and *Sal*I, while class II clones had internal recognition sites for each of these enzymes. Both classes showed a similar fingerprint pattern in *Bam*HI, *Eco*RI, *Eco*RV, *Hin*dIII and *Xba*I digests (Figure 2).

As an additional screen for new classes of clones, we next sequenced multiple clones derived from each primer combinations and compared a region of 438 nt (146 amino acids). All but one of the 24 class II clones were identical, and this sequence was therefore designated as the *Ri4cl1* gene (GenBank AF270933). All 24 of the class II-type clones were identical and different from Class I, and this sequence was therefore designated as the *Ri4cl2* gene (GenBank AF270934).

The sequenced regions of *Ri4cl1* and *Ri4cl2* consisted in both cases of a portion of an open reading frame whose predicted amino acid sequence showed high homology to 4CL proteins in the database. Neither fragment contained an intron within the sequenced regions, consistent with the size of the PCR-amplified products.

Isolation and characterization of ripening-related 4CL *cDNAs*

To complement the PCR-based screening and to specifically identify *4CL* gene(s) that might be expressed during fruit ripening, a cDNA library representing partially ripe fruit mRNA of *Rubus idaeus* was screened using a mixed population of the two partial *Ri4cl* fragments. After three rounds of screening 18 positive plaques were subjected to further analysis. Amplification of inserts of these 18 plaques with vector-specific forward primer T3 and *4CL* gene-specific reverse primers C2+C4, confirmed that only 12 plaques had potentially full-length *4CL* cDNAs (data not shown), which RFLP analysis separated into three distinct groups.

Based on the restriction digest patterns, it was apparent that group 1 clones possessed a digest pattern similar to that of PCR fragment *Ri4cl1*, and that group 2 clones were likewise related to *Ri4cl2*. This relationship was further confirmed by sequencing one representative clone from each group. Sequence analysis revealed that the group 1 clone shared 100% amino

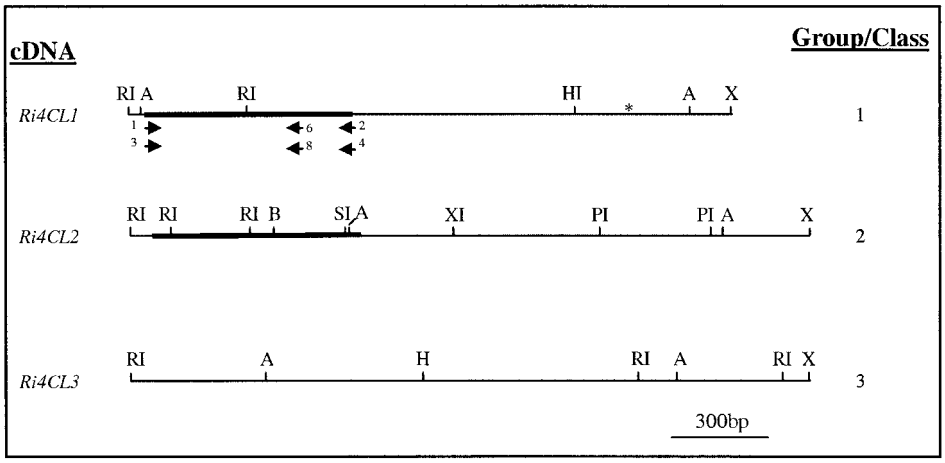

Figure 2. Restriction maps of *Rubus idaeus ECL* clones. The dark lines indicate the regions of *Ri4CL1* and *Ri4CL2* also sequenced and characterized in the PCR-based homology search. Closed arrows indicate the target locations of the degenerate PCR primers, described in detail in Materials and methods. The asterisk represents the approximate position of that stop codon. A, *Acc*I; HI, *Bam*HI; B, *Bgl*I; RI, *Eco*RI; H, *Hin*dIII; PI, *Pst*I; SI, *Sal*I; X, *Xho*I; Xi, *Xba*I.

acid identity (three nucleotide mismatches) with the predicted amino acid sequence of *Ri4cl1*, while the group 2 clone shared 100% nucleotide identity to *Ri4cl2*. However, the group 3 clone represented a new class of raspberry *4CL* gene that had not been detected by the PCR-based homology search. Detailed restriction maps of these clones are shown in Figure 2.

The *Ri4CL1* (GenBank AF239687) cDNA contained 1848 bp, consisting of 20 bp of 5′-untranslated region, a 199 bp 3′-untranslated region and a 1629 bp open reading frame. The *Ri4CL2* (GenBank AF239686) cDNA contained 2051 bp, consisting of an open reading frame of 1632 bp, 92 bp of 5′-untranslated region and 327 bp of 3′-untranslated region. The *Ri4CL3* (GenBank AF239685) cDNA contained 2083 bp, consisting of 70 bp of 5′-untranslated region, a 1773 bp open reading frame and 240 bp of 3′-untranslated region. The calculated molecular mass for the three raspberry 4CLs (59–65 kD) were within the range (55–75 kD) observed for native 4CLs from other plants. The predicted amino acid sequences contain several potential eukaryotic post-translational modification sites, but the biological relevance of these sites is uncertain given that (1) active preparations of 4CL enzymes have been obtained as recombinant proteins in *E. coli* cells and (2) the experimentally determined molecular mass for purified 4CL proteins closely match the predicted mass in species such as parsley (Lozoya *et al.*, 1988) and poplar (Alina *et al.* 1998).

The coding regions of *Ri4CL1* showed 67% and 73% nucleotide sequence identity to *Ri4CL2* and *Ri4CL3*, respectively. The predicted amino acid sequences of *Ri4CL1* and *Ri4CL2* were more similar to each other (72% amino acid sequence identity) than to the predicted *Ri4CL3* sequence (62% and 63% amino acid sequence identity, respectively). In general, the amino acid sequence identity with other 4CLs ranged from a low value of 51% amino acid identity between Ri4CL1 and rice 4CL1, to high value of 83% between Ri4CL1 and soybean 4CL14 (83% amino acid sequence identity). In a comparison with other adenylate-forming enzymes, the three cDNAs showed greatest homology to plant 4CLs, with decreasing homology to plant and bacterial *4CL*-like genes, luciferase, long-chain-fatty-acid-CoA ligase, and peptide synthetases (data not shown).

The raspberry 4CLs also contained the two sequence elements that are highly conserved among 4CL proteins and are required for catalytic functions (Stuible *et al.*, 2000). In the N-terminus is the conserved motif I (LP Y/F SSGTTGPKG), and towards the C-terminus is the second conserved sequence, GEICIRG (motif II). Interestingly, while the seven amino acids of motif II are fully conserved in Ri4CL1 and Ri4CL2, in Ri4CL3 this motif has a slightly variant amino acid sequence, GEICVRG. Such a variation has also been recently reported in the GenBank-deposited sequence from *Lolium perenne* 4CL3 (GenBank AF052223). Two mutations introduced into the GEICIRG motif of At4CL2 (GenBank AAD47193) (Glu-401 Gln, Cys-

403/Ala) reduced the specific activity of the mutant enzymes to 21% and 45% of the wild-type levels, respectively (Stuible *et al.*, 2000), suggesting an important role for this highly conserved motif in enzymatic activity. Ri4CL3 was also different from other 4CLs characterized in having an extended N-terminus which includes a string of six asparagine residues.

Phylogenetic analysis

To study the relationship of the more divergent *Ri4CL3* to *Ri4CL1*, *Ri4CL2* and other *4CLs*, we constructed a phylogenetic tree based on full-length amino acid sequences of the different plant *4CLs* available in the database. The *Arabidopsis* 4CL-like gene on chromosome 1 (GenBank; AC011000) was used to root the phylogenetic trees, since in a parsimonious unrooted tree this sequence formed an independent clade closest to the bottom of the clade consisting of *bona fide* 4CL sequences (Cukovic *et al.*, 2001). Furthermore, recombinant protein expressed from this gene lacks any detectable level of 4CL activity (Stuible and Kombrink, 2001), suggesting that this sequence, while related, represents a suitable outgroup to the *bona fide* plant 4CL sequences.

As seen previously, the *4CL* genes fall into two distinct phylogenetic classes: class I and class II (Figure 3). Ri4CL1 and RiCL2 were associated with class I 4CLs, with Ri4CL1 clustering with soybean 4CL14 and aspen 4CL1, and Ri4CL2 showing a stronger relationship to hybrid poplar 4CL1 and 4CL2. Ri4CL3, however, clustered with the class II genes (Figure 3). A novel outcome of this analysis was the independent clustering of three monocot 4CL sequences, represented by ryegrass 4CL2, ryegrass 4CL3 and rice 4CL1, which appear to form a third class of 4CL genes (class III). However, this class was supported by a low bootstrap value (<50%), indicating that additional monocot 4CL sequences will be required before their placement can be confidently resolved. Interestingly, while Ri4CL3 and ryegrass 4CL3 are the only two sequences displaying a variant catalytic motif II (GEICVRG), they did not cluster together.

Enzymatic activity of the recombinant raspberry 4CL proteins

Ri4CL3 is the most divergent of the three members of the raspberry *4CL* gene family detected in this study. To compare the enzymatic properties of this divergent member to other 4CLs, the three raspberry cDNAs were each expressed as recombinant proteins bearing N-terminal His_6/Strep tags. The recombinant proteins generated in bacteria were tested for their relative abilities to utilize differently substituted hydroxycinnamic and benzoic acids as substrates. The relative substrate utilization profiles of the three raspberry recombinant proteins using 0.2 mM substrate concentrations are shown in Figure 4. Full-length Ri4CL-1 exhibited relative activities of 100%, 27%, 22% and 11% towards 4-coumarate, ferulate, caffeate, and cinnamate. Truncated Ri4CL1 (internal stop codon) exhibited no activity against any of the substrates tested, indicating that the C-terminal 40 amino acid residue stretch is important for enzymatic activity and/or proper folding of the protein. Recombinant Ri4CL2 showed a strong preference for cinnamate (153%) and somewhat lower but still prominent activities towards 4-coumarate (100%), ferulate (96%) and caffeate (59%). Ri4CL3 recombinant protein was most active with 4-coumarate and displayed some activity towards caffeate (29%), but was inactive with either cinnamate or ferulate. None of the recombinant proteins had detectable activities against sinapate, or against benzoic acid and its derivatives.

Developmental regulation of the Ri4CL genes

Quantitative RT-PCR (QRT-PCR) assays that cleanly distinguished between the three 4CL transcripts (Figure 5) showed that all three genes are expressed in each tissue examined (Figure 6). *Ri4CL1* was constitutively expressed at low levels in all organs investigated. In contrast, *Ri4CL2* and *Ri4CL3* showed a developmentally regulated pattern of gene expression (Figure 6). *Ri4CL1* and *Ri4CL2* were expressed mostly in vegetative tissue, with *Ri4CL1* being the most highly expressed form in leaves and *Ri4CL2* accounting for the majority of the *4CL* transcripts in shoots. The abundance of the three transcripts in leaves was distributed in a ratio of ca. 10:2:1 (*Ri4CL1: Ri4CL2: Ri4CL3*), while in shoots the ratio was ca. 1:3:3 (*Ri4CL1:Ri4CL21:Ri4CL3*). High levels of *Ri4CL3* transcripts were observed in the various developmental stages of flowers and fruits, while its expression was comparatively lower in vegetative tissues. In general, the expression pattern of *Ri4CL2* was reproducibly similar to that of *Ri4CL3*, although the relative amounts of *Ri4CL2* mRNA were lower than those of *Ri4CL3*. Overall, our analysis suggests that *Ri4CL1* and *Ri4CL2* are the dominant 4CL forms expressed in vegetative tissue, while expression of *Ri4CL3* is correlated with development of reproductive tissue.

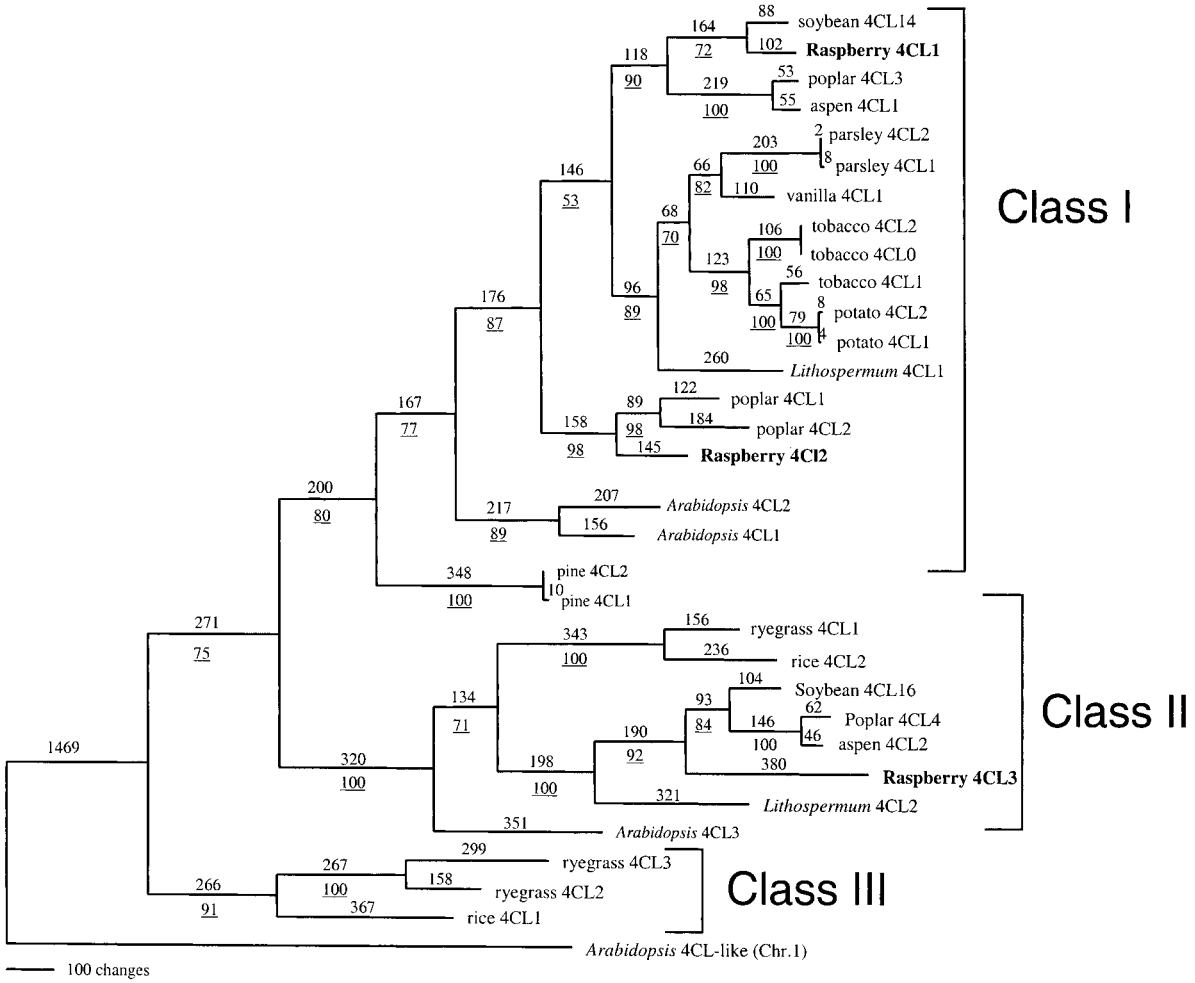

Figure 3. Phylogenetic relationships among plant 4CL and 4CL-like proteins. The most parsimonious rooted phylogenetic tree was constructed from an alignment of protein sequences using a heuristic search within PAUP 4.02b. Characters were re-weighted for maximum value of retention indices. The *Arabidopsis* 4CL-like sequence (GenBank AC011000) was used as the outgroup. Branch lengths are indicated above the branch lines and clustering percent support values derived from 1000 bootstrap replications are underlined. The tree has a consistency index of 0.764.

Discussion

4CL is encoded by a family of genes in plants, and the recombinant enzymes encoded by some of these gene-family members within a species often demonstrate distinct substrate utilization profiles (Hu et al., 1998; Ehlting *et al.*, 1999). It has been suggested that one role for multiple forms of 4CL might be in channeling carbon through various sub-branches within the phenylpropanoid pathway through specific interactions amongst different isoforms of the three enzymes that make up the core phenylpropanoid pathway. However, direct biochemical evidence for such interactions is limited, and most efforts have focused on PAL and C4H rather than 4CL (Rasmussen and Dixon, 1999). Examination of the possible roles of 4CL isoforms in controlling the accumulation of specific phenylpropanoid end-products requires characterization of this gene family from one species. Raspberry is of particular interest in this regard since maturing fruits accumulate several phenylpropanoid-derived end-products (Figure 1) whose biosynthesis requires hydroxycinnamoyl-CoA esters as precursors.

Identification of multiple gene family members has generally relied on screening of cDNA/DNA libraries, or on use of PCR-based approaches. The likelihood of detection of a given gene family member by these approaches is dependent on the temporal and spatial

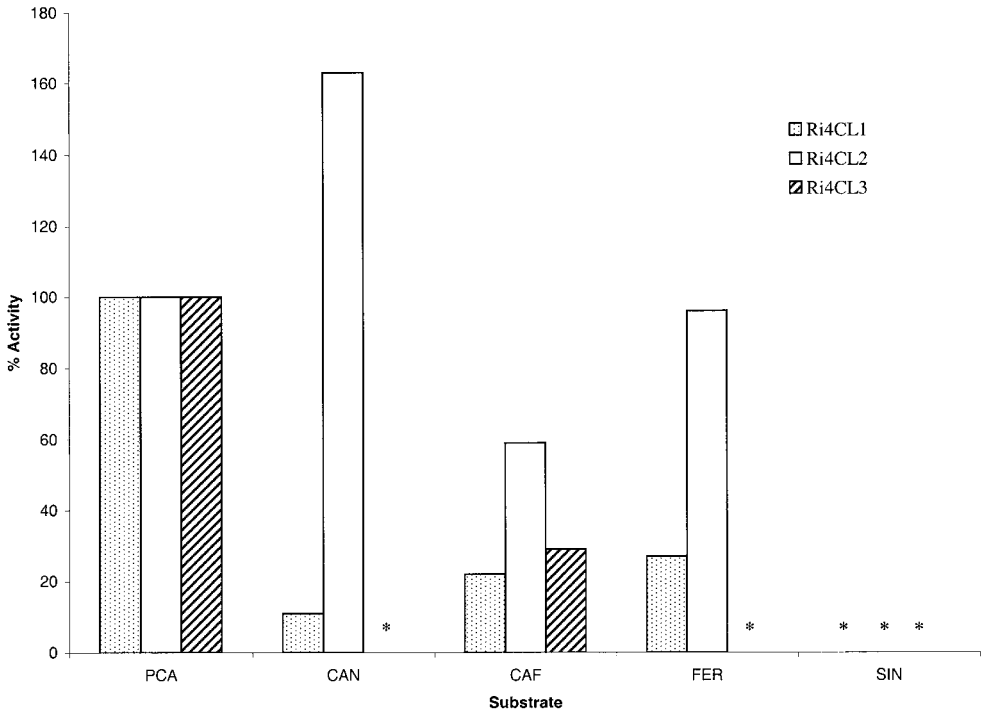

Figure 4. Substrate utilization profile of *Rubus idaeus* 4CL recombinant proteins. 4CL substrate specificity was measured in crude bacterial extracts expressing recombinant proteins, using 0.2 mM concentrations of each hydroxycinnamic acid. Results are averages of three independent assays, and are expressed as a percent of the activity obtained using 4-coumarate as a substrate. One hundred percent activity represents a specific activity of 3 nkat/mg for Ri4CL1, 3.5 nkat/mg for Ri4CL2, and 10 nkat/mg for Ri4CL3 with 4-coumarate as substrate. *, no conversion detected. PCA, 4-coumarate; CAN, cinnamate; CAF, caffeate; FER, ferulate; SIN, sinapate.

pattern of gene expression and/or on the sequence of the degenerate primers. For example, *4CL* was initially suggested to exist as a single-copy gene in *Arabidopsis*, based on cDNA library screening (Lee *et al.*, 1995), but was later found to be encoded by multiple genes (Ehlting *et al.*, 1999). Similarly, *PAL* was reported to exist as a single-copy gene in pine, based on screening of a xylem cDNA library (Whetton and Sederoff, 1992), but was later shown to consist of at least eight divergent loci based on a PCR-based homology search (Butland *et al.*, 1998). We therefore combined these two approaches to analyze and characterize the *4CL* gene family from raspberry.

Careful analysis of the coding sequences of known *4CL* genes revealed a limited number of regions that displayed appreciable sequence conservation favorable for designing degenerate PCR primers. Survey of the raspberry genomic DNA for *4CL* sequences with sets of these primers allowed us to identify two putative *4CL* genes. In addition to the PCR screen, we were able, through use of a cDNA library screen, to retrieve a third, more divergent, sequence. Translation of the three cDNA sequences revealed that *Ri4CL3* has amino acid sequence FSKLPDL at the positions targeted by the forward PCR primers C1/C3, which had been designed to target the amino acid sequence RSKLPDI. The associated nucleotide changes in the primary sequence of *Ri4CL3* include a mismatch at the 3′ end of the primers. Similarly, the nucleotide sequence of primers C2/C4 diverged at three or four positions from the primary sequence of the *Ri4CL3* cDNA. It is likely that these differences resulted in very low, or no, amplification of this gene family

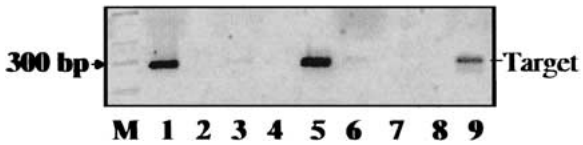

Figure 5. Specificity of the gene-specific primers used for QRT-PCR analysis. Lanes 1, 4 and 7, *Ri4CL1* cDNA; lanes 2, 5 and 8, *Ri4CL2* cDNA; lanes 3, 6 and 9, *Ri4CL3* cDNA; lane 5, *Ri4CL1* cDNA. Lanes 1–3, amplified products with gene-specific primers for *Ri4CL1*; lanes 4–6, amplified products of gene-specific primers for *Ri4CL2*; lanes 7–9, amplified products of gene-specific primers for *Ri4CL3*. M, 100 bp molecular weight marker (Life Technologies, Burlington, Canada).

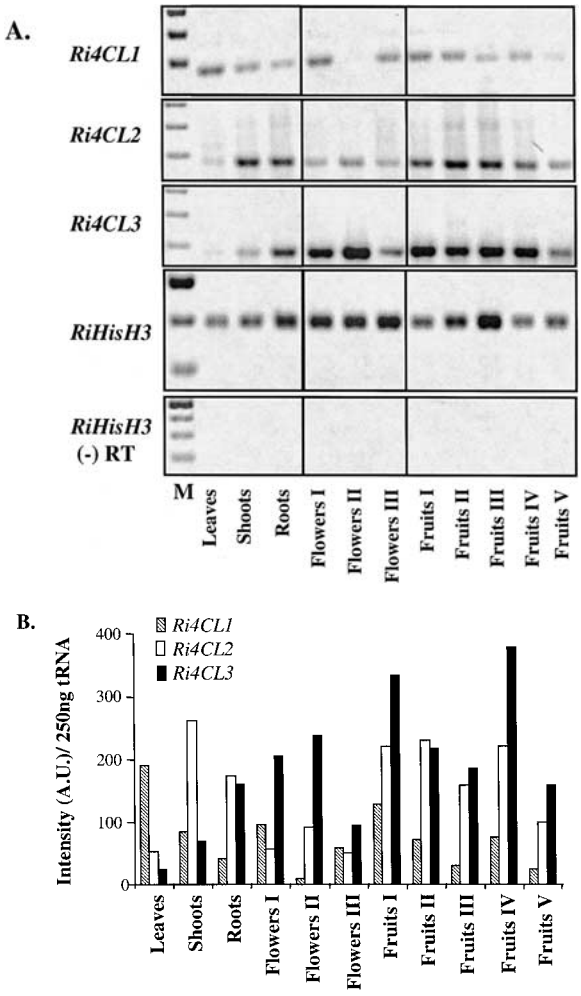

Figure 6. Expression of specific *Ri4CL* transcripts in different organs of raspberry plants as estimated by quantitative RT-PCR analysis. A. QRT-PCR analysis was performed using 250 ng total RNA (tRNA) isolated from young leaves, shoots, roots and different developmental stages of flowers and fruits. After 30 cycles of amplification, the PCR products were resolved in a 3% TAE-agarose/ethidium bromide gel. B. The relative expression levels of the *Ri4CL* transcripts (expressed in arbitrary units (AU) per 250 ng of total RNA) were estimated by measuring the ethidium bromide fluorescence of the PCR products resolved by gel electrophoresis in A. Similar results were obtained in two independent experiments. The intensity of the *Ri4CL* bands was normalized to the average intensity of the *RiHisH3* products as a control for equal amounts of starting total RNA. Since a constant amount of cDNA was used for each of three gene-specific primer pairs, the expression levels of the three *Ri4CL* transcripts in each organ are directly comparable in this graph.

member during the PCR screen. Using the same set of primers as designed in this study, Cukovic *et al.* (2001) observed similar results, where they successfully amplified three new *4CL* clones (*Ptd4CL6*, *Ptd4CL10*, *Ptd4CL14*) from the *Populus* genome but failed to identify a previously characterized *Populus* cDNA clone, *Ptd4CL1* (Alina *et al.*, 1998). Thus, while the primers designed in this study are capable of amplifying divergent *4CL* genes, they do not capture the full range of 4CL gene sequences, and the possibility remains that other, even more divergent, 4CL genes occur in the *Rubus* genome.

The three raspberry *4CL* cDNA sequences are relatively divergent, sharing about 64–83% amino acid sequence identity, and thus likely represent three different *4CL* genes. Phylogenetic analysis of the full-length Ri4CL amino acid sequences place these sequences in two classes, as previously observed (Ehlting *et al.* 1999; Cukovic *et al.* 2001), while a possible additional sequence class consisting only of the monocot sequences could also be identified (Figure 4).

Based on the phylogenetic clustering, it was apparent that all dicot taxa have multiple class I and single class II-type *4CL* genes. Within the Class I cluster are found two of the 4CL isoforms from *Arabidopsis*, as well as genes from hybrid poplar, tobacco, parsley, and potato. Ri4CL1 and Ri4CL2 are both found in Class I, but Ri4CL1 grouped with pathogen non-inducible soybean 4CL14, and was well separated from Ri4CL2, which grouped with hybrid poplar 4CLs.

While other class I *4CL* genes appear to have evolved by recent gene duplication events within respective plant lineages, as illustrated by the existence of closely related pairs of *4CL* genes in *Arabidopsis*, hybrid poplar, potato, tobacco and parsley, the raspberry class I clones might have evolved from multiple lineages. Alternatively, a sequence of duplication/deletion cycles may have resulted in the loss of paralogous loci between genera, as has been proposed for *4CL* genes in the Pinaceae (Wang *et al.*, 2000). Important factors contributing to this lack of phylogenetic resolution may be the relatively narrow range of taxa for which data are available, and the non-comprehensive isolation methods that have typically been used for detection of gene family members. From an evolutionary perspective, such gene duplication and divergence processes probably play an important role in the development of functional diversity by creating altered substrate specificities, as well as new interactions with effectors and/or activation processes.

In contrast to Ri4CL1 and 2, Ri4CL3 grouped with the class II angiosperm 4CL sequences and was most closely related to poplar 4CL4, aspen 4CL2, and soybean *4CL16*. The distinctiveness of class II *4CL* genes must reflect an ancient divergence from class I genes, since specific 4CL isoforms from highly diverged plant taxa are found grouped together in class II rather than with other 4CL sequences from the same species. *4CL* genes belonging to class II also share structural and functional differences from those in class I. For example, *Arabidopsis* 4CL3, *Lithospermum* 4CL2, rice 4CL2, aspen 4CL2, and Ri4CL3, all in Class II, have N-terminal extensions of 19–40 amino acids, which are absent from Class I sequences from the same species. It has been proposed that these variant N-terminal regions might be involved in phenolic substrate binding specificity (Hu *et al.*, 1998) but this has yet to be confirmed experimentally. The expression patterns of class I *4CL* genes have also been shown to differ from those of class II genes in *Arabidopsis* (Ehlting *et al.*, 1999), soybean (Uhlmann and Ebel, 1995) and aspen (Hu *et al.*, 1998), suggesting that these genes might have acquired additional or unique functions that distinguish them from their progenitors. While class III consisted only of monocot 4CL sequences, this class was not supported by a high bootstrap value and, as such, might represent a subset of either class I or class II type sequences. Based on their expression pattern, and the phylogenetic clustering observed by us and previous investigators, we propose that these sequences represent divergent class I-type sequences. Sequencing of additional monocot *4CL* sequences should allow further phylogenetic refinement of these sets of sequences. It is interesting that within the putative 'class III', cluster, ryegrass 4CL3 possesses a divergent motif II (GEICVRG), like class II Ri4CL3. This suggests that even highly conserved domains within the 4CL sequences have undergone independent selective changes.

Quantitative RT-PCR analysis demonstrated that all of the *Ri4CL* genes are differentially expressed. *Ri4CL1* is expressed primarily in leaves, with lower levels observed in shoots, roots and developing fruits and flower. Because of its relatively low expression in developing flowers and fruits, and its substrate utilization profile, *Ri4CL1* is unlikely to play a major role in the biosynthesis of 'raspberry ketone' or flavonoid derivatives in fruits and flowers. Instead, *Ri4CL1* may be important for the biosynthesis of phenolic derivatives in leaves. *Ri4CL2* was expressed most highly in shoots, with moderate levels of expression in roots and low levels in leaves and developing flowers, which might imply a role in lignification of the raspberry canes. The substrate utilization profiles of recombinant Ri4CL1 and Ri4CL2 (Figure 4) are consistent with such a hypothesis. While all three recombinant proteins were able to use 4-coumarate as a substrate, demonstrating that they are *bona fide* 4CLs, the only other hydroxycinnamic acid substrate used by recombinant Ri4CL3 was caffeate. Gene expression analysis showed that *Ri4CL3* was expressed most strongly in roots, relative to shoots and leaves. During the process of flower development, it was expressed in fully mature flowers (stage II), while during fruit development it was expressed most highly in fully mature red fruits (stage IV). The substrate utilization and expression profile of Ri4CL3 suggest that its primary function may be to provide activated 4-coumaroyl CoA esters for branch pathways of phenylpropanoid metabolism that only require activated 4-coumarate, as opposed to other activated hydroxycinnamic acid derivatives. In raspberry, this could support either the flavonoid or the flavor pathway, or both. In a parallel study involving raspberry *PAL* genes, we observed a similar differential pattern of expression in which *RiPAL1* was expressed in vegetative tissues while *RiPAL2* was induced during fruit development (Kumar and Ellis, 2001). This leads us to speculate that RiPAL2 and Ri4CL3 might be involved in formation of metabolites in the flavonoid and/or flavor pathway in raspberry fruit, whereas RiPAL1 and Ri4CL1/2 might be primarily committed to synthesis of other phenylpropanoid products. This hypothesis obviously needs to be corroborated with further biochemical studies.

It is notable that recombinant 4CL2 is able to utilize cinnamate so efficiently as a substrate, since most 4CLs analyzed at the enzyme level show little or no activity against this unsubstituted phenylpropanoid. There is no record of *Rubus* spp. accumulating B-ring unsubstituted flavonoids or benzoate derivatives whose biosynthesis might be predicted to require cinnamoyl-CoA as a precursor. A similar apparently anomalous activity against cinnamate was also reported with one of two recombinant tobacco 4CLs (Lee and Douglas, 1996), which suggests that some members of the 4CL family may be susceptible to artifactual acceptance of cinnamate as a substrate when expressed as recombinant forms.

While the individual *Ri4CL* gene family members are differentially expressed, their expression pattern does not readily map to their observed phylogenetic class. The expression pattern of *Ri4CL2*, which be-

longs to phylogenetic class I, most closely reflects the expression pattern of *Ri4CL3*, which belongs to phylogenetic class II. This suggests that the signaling mechanisms controlling their developmental expression have evolved independently of the genes themselves. Overall, the variation in expression patterns, combined with the substrate utilization profiles of the three recombinant proteins, strongly suggests that the different *Ri4CL* gene family members have different functional roles during raspberry growth and development. Defining these roles will require creation and characterization of transgenic raspberry plants in which each gene has specifically been silenced.

References

Alina, S.M., Pri-Hadash, A., Theilmann, D.A., Ellis, B.E. and Douglas, C.J. 1998. 4-coumarate:Coenzyme A ligase in hybrid poplar. Plant Physiol. 116: 743–754.

Altschul, S.F., Fish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403–410.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. 1995. Current Protocols in Molecular Biology. John Wiley, New York.

Becker-André, M., Schulze-Lefert, P. and Hahlbrock, K. 1991. Structural comparison, modes of expression, and putative *cis*-acting elements of the two 4-coumarate:CoA ligase genes in potato. J. Biol. Chem. 266: 8551–8559.

Borejsza-Wysocki, W. and Hrazdina, G. 1994. Biosynthesis of p-hydroxyphenylbutan-2-one in raspberry fruits and tissue cultures. Phytochemistry 35: 623–628.

Butland, S.L., Chow, M.L. and Ellis, B.E. 1998. A diverse family of phenylalanine ammonialyase genes in pine tree and cell cultures. Plant Mol. Biol. 37: 15–24.

Chappel, J. and Hahlbrock, K. 1984. Transcription of plant defense genes in response to UV light or fungal elicitor. Nature 311: 76–78.

Cukovic D, Ehlting J, VanZiffle JA, Douglas CJ 2001. Structure and evolution of 4-coumarate: coenzyme A ligase (4CL) gene families. Biol. Chem. 382: 645–654.

Dixon, R.A. 1999. Plant natural products: the molecular genetic basis of biosynthetic diversity. Curr. Opin. Biotechnol. 10: 192–197.

Dixon, R.A. and Paiva, N.L. 1995. Stress-induced phenylpropanoids in plants. Plant Cell 7: 1085–1097.

Douglas, C., Hoffmann, H., Schulz, W. and Hahlbrock, K. 1987. Structure and elicitor or u.v.-light-stimulated expression of two 4-coumarate:CoA ligase gene in parsley. EMBO J. 6: 1189–1195.

Doyle, J.J. and Doyle, J.L. 1990. Isolation of plant DNA from fresh tissue. Focus 12: 13–15.

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

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783–791.

Goiffon, J.P., Brun, M. and Bourrier, M.J. 1991 Fruit anthocyanin pigments of red raspberry cultivars. J. Chromatogr. 55: 101–127.

Hahlbrock, K. and Grisebach, H. 1979. Enzymic controls in the biosynthesis of lignin and flavonoids. Annu. Rev. Plant Physiol. 30: 105–130.

Hrazdina, G. and Wagner, G. 1985. Metabolic pathways as enzyme complexes: evidence for the synthesis of phenylpropanoids and flavonoids in membrane-associated enzyme complexes. Arch. Biochem. Biophys. 237: 88–100.

Hu, W.-J., Kawaoks, A., Tsai, C., Lung, J., Osakabe, K., Ebinuma, H. and Chiang, V. 1998. Compartmentalized expression of two structurally and functionally distinct 4-coumarate:CoA ligase genes in aspen (*Populus tremuloides*). Proc. Natl. Acad. Sci. USA 95: 5407–5412.

Knobloch, K.-H. and Hahlbrock, K. 1977. 4-Coumarate:CoA ligase from cell suspension cultures of *Petroselinum hortense* Hoffm. Partial purification, substrate specificity, and further properties. Arch. Biochem. Biophys. 184: 237–248.

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

Lee, D. and Douglas, C. 1996. Two divergent members of a tobacco 4-coumarate: Coenzyme A ligase gene family. Plant Physiol. 112: 193–205.

Lee, D., Ellard, M., Wanner, L.A., Davis, K.R. and Douglas, C.J. 1995. The *Arabidopsis thaliana* 4-coumarate: CoA ligase (*4CL*) gene: stress and developmentally regulated expression and nucleotide sequence of its cDNA. Plant Mol. Biol. 28: 871–884.

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

Lozoya, E., Hoffmann, H., Douglas, C., Schulz, W., Scheel, D. and Hahlbrock, K. 1988. Primary structures and catalytic properties of isoenzymes encoded by the two 4-coumarate:CoA ligase genes in parsley. Eur. J. Biochem. 176: 661–667.

Rasmussen, S. and Dixon, R.A. 1999. Transgene-mediated and elicitor-induced perturbation of metabolic channeling at the entry point into the phenylpropanoid pathway. Plant Cell 11: 1537–1552.

Schmelzer, E., Kruger-Lebus, S. and Hahlbrock, K. 1989. Temporal and spatial patterns of gene expression around sites of attempted fungal infection in parsley leaves. Plant Cell 1: 933–1001.

Stöckigt, J. and Zenk, M.H. 1975. Chemical synthesis and properties of hydroxycinnamoyl-Coenzyme A derivatives. Z. Naturforsch. 30c: 352–358.

Stuible, H.P. and Kombrink, E. 2001. Identification of the substrate specificity-conferring amino acid residues of 4-coumarate:Coenzyme A ligase allows the rational design of mutant enzymes with new catalytic properties. J. Biol. Chem. 276: 26893–26897.

Stuible, H., Buttner, D., Ehlting, J., Hahlbrock, K. and Kombrink, E. 2000. Mutational analysis of 4-coumarate:CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes. FEBS Lett. 467: 117–122.

Uhlmann, A. and Ebel, J. 1993. Molecular cloning and expression of 4-coumarate: coenzyme A ligase, an enzyme involved in the resistance response of soybean (*Glycine max* L.) against pathogen attack. Plant Physiol. 102: 1147–1156.

Wang, X.-Q., Tank, D.C. and Sang, T. 2000. Phylogeny and divergence times in *Pinaceae*: evidence from three genomes. Mol. Biol. Evol. 17: 773–781.

Webster, L.T. Jr., Mieyal, J.J. and Siddiqui, U.A. 1974. Benzoyl and hydroxybenzoyl esters of coenzyme A. J. Biol. Chem. 249: 2641–2645.

Whetten, R.W. and Sederoff, R.R. 1992. Phenylalanine ammonia-lyase from loblolly pine: purification of the enzyme and isolation of complementary DNA clones. Plant Physiol. 98: 380–386.

Wu, S.C. and Hahlbrock, K. 1992. *In situ* localization of phenyl-propanoid related gene expression in different tissues of light-grown and dark-grown parsley seedlings. Z. Naturforsch. 47c: 591–600.

Zhang, X.H. and Chiang, V.L. 1997. Molecular cloning of 4-coumarate:Coenzyme A ligase in loblolly pine and the roles of this enzyme in the biosynthesis of lignin in compression wood. Plant Physiol. 113: 65–74.