



# Transcripts for possible capsaicinoid biosynthetic genes are differentially accumulated in pungent and non-pungent *Capsicum* spp

Jeanne Curry, Maneesha Aluru, Marcus Mendoza, Jacob Nevarez, Martin Melendrez, Mary A. O'Connell \*

Molecular Biology Program, Department of Agronomy and Horticulture, New Mexico State University, PO Box 30003, Las Cruces, NM 88003-8003, USA

Received 13 April 1999; received in revised form 1 June 1999; accepted 18 June 1999

## Abstract

Capsaicinoids, the alkaloids responsible for pungency in chile fruit, are synthesized from phenylpropanoid intermediates and short-chain branched-fatty acids. Transcript levels of enzymes on the capsaicinoid pathway were monitored in *Capsicum annuum* and *C. chinense* fruit as a function of development, tissue type and genotype. Clones for *Pal*, *Ca4h*, and *Comt* were isolated from a cDNA library of habañero (*C. chinense*) placenta. These cDNA clones were used to measure transcript levels in different fruit tissues throughout development in six cultivars differing in pungency. Transcript levels for all three genes were positively correlated with degree of pungency in placental tissue; habañero, the most pungent chile fruit, had the highest transcript levels, CalWonder, a non-pungent fruit, had the lowest levels. Using the transcript accumulation pattern of the phenylpropanoid genes as a screening criterion, other cDNA clones have been selected. Clones for an aminotransferase, predicted to synthesize vanillylamine, and for a 3-keto-acyl ACP synthase, predicted to elongate branched-chain fatty acids, were identified. These genes are expressed in a placental-specific manner, and transcript levels are positively correlated with fruit pungency. © 1999 Elsevier Science Ireland Ltd. All rights reserved.

**Keywords:** Branched-chain fatty acid; Pepper; Phenylpropanoid pathway; Pungency; Secondary metabolism

## 1. Introduction

The burning sensation one gets from eating 'hot' or pungent peppers is caused by alkaloids called capsaicinoids. These compounds are uniquely produced in the fruit of members of the genus *Capsicum*. The word chile will be used in this report to indicate plants in the genus *Capsicum*. Capsaicinoids all share a common aromatic moiety, vanillylamine, and differ in the length and degree of unsaturation of a fatty acid side chain (Fig. 1) [1,2]. The two most common capsaicinoids, capsaicin and dihydrocapsaicin differ in

the degree of unsaturation of a 9-carbon fatty acid side chain; five other naturally occurring capsaicinoids: nordihydrocapsaicin, homodihydrocapsaicin, homocapsaicin, norcapsaicin and nornorcapsaicin, differ in chain length ( $n = 7-10$ ) as well as degree of unsaturation.

Non-pungency is a recessive trait; pungency is inherited as a single major gene at locus *C* [3]. However, virtually nothing is known about the genes that control the synthesis of individual capsaicinoids, or that control the abundance of total capsaicinoids. Capsaicinoids start to accumulate approximately 20 days post anthesis, and usually persist through fruit development [4,5]. The site of synthesis and accumulation of the capsaicinoids is the epidermal cells of the placenta [4]. Within the cells, capsaicin-synthesizing activity has been

\* Corresponding author. Tel.: +1-505-646-5172; fax: +1-505-646-6041.

E-mail address: moconnel@nmsu.edu (M.A. O'Connell)

demonstrated in the vacuolar fraction and capsaicinoids have been demonstrated to accumulate in vacuoles [4,6]. Ultimately the capsaicinoids are secreted extra-cellularly into receptacles between the cuticle layer and the epidermal layer of the placenta [7]. These filled receptacles of capsaicinoids often appear as pale yellow to orange droplets on the placenta of the most pungent chiles.

Capsaicinoids are synthesized by the condensation of vanillylamine with a short-chain branched-fatty acid. A possible biosynthetic pathway is presented in Fig. 1. The production of vanillylamine is via the phenylpropanoid pathway and the branched-chain fatty acid is synthesized from a branched-chain amino acid, e.g. valine. Evidence

to support this pathway includes radiotracer studies, determination of enzyme activities, and the abundance of intermediates as a function of fruit development [1,2,5,6,8,9].

Capsaicinoids are the most important spice in the world, based on scale of production [10]. We are interested in understanding and manipulating the accumulation of capsaicinoids in chile. The capsaicinoid content of the fruit is the most important fruit quality trait. Expression of this trait is responsive to environmental stress [10,11]. To understand the basis for the regulation of this character, we have begun to isolate and characterize the genes on the capsaicinoid biosynthetic pathway. The early steps on the phenylpropanoid branch of the capsaicinoid pathway are expressed

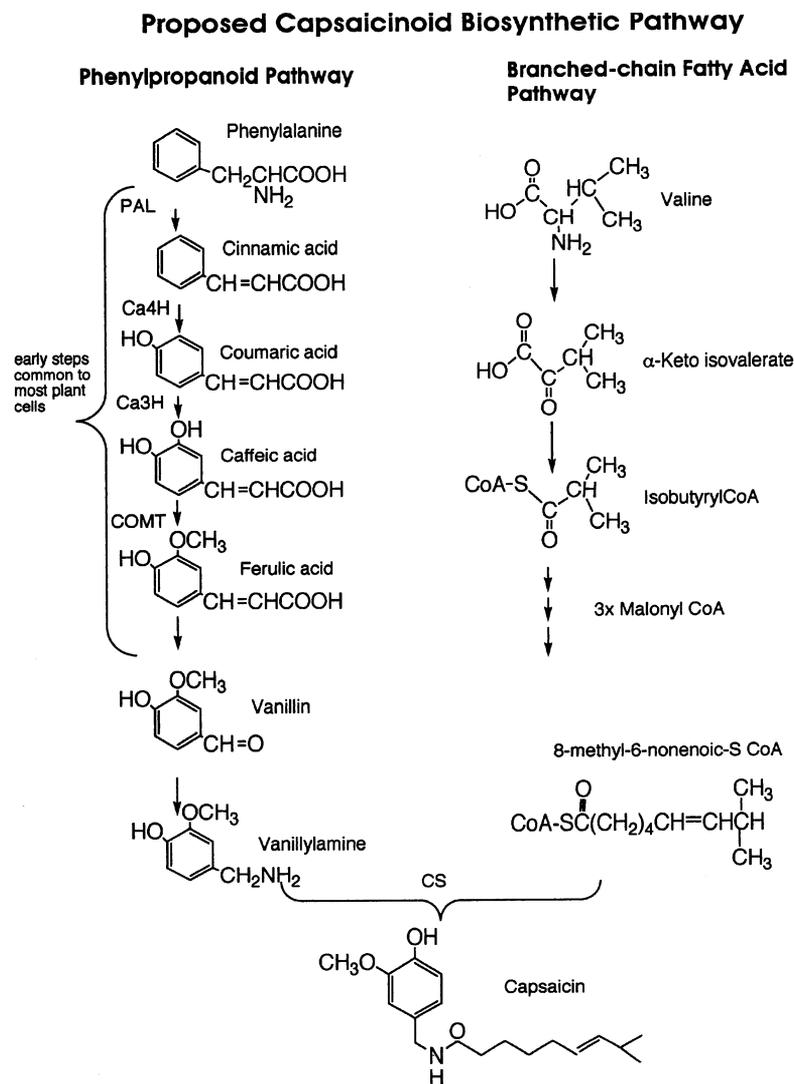


Fig. 1. Proposed pathway for capsaicinoid biosynthesis. The enzymes indicated on the pathway are: phenylalanine ammonia lyase (PAL), cinnamic acid 4-hydroxylase (Ca4H), coumaric acid 3-hydroxylase (Ca3H), caffeic acid *O*-methyltransferase (COMT), and capsaicinoid synthetase (CS).

Table 1  
Typical capsaicinoid or pungency levels in chile varieties

Species	Cultivar	Capsaicinoid <sup>a</sup>	Pungency (SHU <sup>b</sup> )
<i>Capsicum annuum</i>	CalWonder	0	0
	NuMex Conquistador	0	0
	NuMex 6-4	0.8	1200
	NuMex Joe E. Parker	0.8	1200
	Jalapeño	6.7–20	10–30 000
<i>C. chinense</i>	Habañero (PI 1720)	134–200	200–300 000
	Habañero (PI 1678)	134–200	200–300 000
	PI 1721	0	0

<sup>a</sup> Capsaicinoid concentration in mg/g dry weight of fruit [3,25].

<sup>b</sup> SHU, Scoville Heat Units; 1 ppm capsaicin = 15 SHU [25].

in many plants and in many plant cell types, as the products of these reactions are intermediates for a wide range of plant secondary products. cDNA and in some cases, genomic clones have been characterized for many plants for three of these enzymes: phenylalanine ammonia lyase (PAL) [12–16], cinnamate 4-hydroxylase (Ca4H) [17–21] and caffeic *O*-methyl transferase, (COMT) [22–24]. In this report we describe the transcript abundance for *Pal*, *Ca4h*, and *Comt* as a function of fruit developmental stage, fruit tissue type, and degree of pungency in *Capsicum spp.* cDNA clones for *Pal*, *Ca4h*, and *Comt* were isolated from a library of *C. chinense* c.v. habañero placental transcripts using heterologous probes. The habañero cDNA clones were then used to monitor transcript levels in chile fruit. Based on the differential patterns of expression of these three genes, we devised a screening strategy to isolate cDNAs for additional capsaicinoid biosynthetic enzymes.

## 2. Materials and methods

### 2.1. Plant materials

Plants were grown from seed in Metro-Mix 360<sup>®</sup>, watered daily and fertilized with a controlled release fertilizer, Osmocote<sup>®</sup> (14-14-14). The plants were maintained in a greenhouse on the NMSU main campus. The RNA for the cDNA library construction was isolated from placental fruit tissue of plants grown in the field at the Fabian Garcia Research Center, NMSU, Las Cruces. Typical capsaicinoid content of the chile germplasm used in this report is listed in Table 1.

Capsaicinoid content of chile fruit was determined using an HPLC method [25]. Essentially fruit is dried, ground to a powder and extracted with acetonitrile. An aliquot is then injected into a Waters HPLC equipped with a fluorescence detector and the samples are resolved on C-18 reverse phase columns.

### 2.2. Nucleic acid isolation and blot analyses

Total RNA and genomic DNA were prepared as described earlier [26]. Plant tissues to be extracted for RNA were collected directly into liquid N<sub>2</sub>. RNA blots were prepared after electrophoresis of ethidium bromide stained RNA samples [27]. Genomic DNA was digested with restriction endonucleases following the supplier's buffer recommendations, for at least 4 h at 37°C. Pre-hybridization, hybridization and washing conditions were as described earlier [26]. In one experiment, transcript levels were quantified using a Molecular Dynamics Storm 860 phosphorimager. In this case, the blots were re-probed with a fragment of the chile gene for 25S ribosomal RNA to normalize transcript abundance. All northern displays were replicated at least once with independently isolated RNA preparations.

### 2.3. cDNA library construction

Fruit were collected from *C. chinense* c.v. habañero grown at the Fabian Garcia Research Center, Las Cruces, NM. The fruit were sorted into three different developmental stages based on size and color development: (1) green fruit; (2) fruit with developing orange color, and (3) fully orange

fruit. The placental tissue was dissected from the fruit and RNA isolated from each of the three stages. PolyA<sup>+</sup> RNA was prepared from the placental tissue of stage 1 using a batch oligo-dT process and a cDNA library of these transcripts was generated in Lambda ZapII (Stratagene). The primary library was titered at  $2.5 \times 10^5$  recombinant pfu/ml, and then amplified to a titer of  $1 \times 10^{10}$  recombinant pfu/ml. This library was differentially screened using radiolabeled first strand cDNA from placental transcripts from either immature habañero fruit or from immature non-pungent *C. chinense* (PI 1721) fruit. Plaques which hybridized strongly to the habañero probe and not to the non-pungent *C. chinense* were purified and characterized.

#### 2.4. DNA sequencing and alignments

DNA sequences were determined on recombinant plasmids using dideoxy termination methods, SequiTherm EXCEL kit (Epicentre Technol) and a LI-COR automated DNA sequencer. DNASTAR software was used to assemble and analyze DNA sequences; DNA and predicted amino acid sequences were searched against DNA and protein databases using BLAST 2.0 [28] at the NCBI web site ([www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)); in some cases against the BLOCKS database ([29]; [www.blocks.fhrc.org/](http://www.blocks.fhrc.org/)) or analyzed by PSORT to predict sub-cellular sites of accumulation ([30]; [psort.nibb.ac.jp:8800/](http://psort.nibb.ac.jp:8800/)). The nucleotide sequence data are in the GenBank Nucleotide Sequence Database under accession numbers: *Pal* AF081215 (phenylalanine ammonia lyase); *Ca4h* AF088847 (cinnamate 4-hydroxylase); *Comt* AF081214 (caffeic acid 3-*O*-methyl transferase); *Kas* AF085148 (3-keto-acyl acyl carrier protein (ACP) synthase); pAmt AF085149 (putative amino transferase); *rrn25* AF088849 (25S ribosomal RNA).

### 3. Results

#### 3.1. Isolation of cDNAs for phenylpropanoid pathway genes

Chile cDNA forms of *Pal*, *Ca4h* and *Comt* were isolated from the habañero cDNA library using heterologous cDNA clones from alfalfa and soybean [12,17,22]. DNA sequences from each of the

chile clones were determined and BLAST searches were then conducted. In some cases the library was re-screened to obtain longer forms of the genes. Alignments were made of the sequences obtained from the chile cDNA clones to cDNA and genomic sequences of other plant genes. Based on these alignments the clones were named *Pal*, *Ca4h* or *Comt*.

##### 3.1.1. *Pal*

The largest chile cDNA sequence for *Pal* was 1.8 kb, with a 1599 nucleotide ORF encoding 532 amino acids and a 256 nucleotide 3'UTR (AF081215). The chile ORF had greater than 90% amino acid sequence identity to *Pal* genes from potato, tobacco and tomato (M84466; M83314; X63103 [13]; M90692 [14]; D17467 [15]; X78269 [16]). Based on these alignments and sequence analysis, the chile clone is not full-length, the habañero cDNA clone for *Pal* is missing all of the 5'UTR and approximately 549 bp corresponding to the first 183 amino acids. The entire cDNA sequence was used as a probe in the subsequent genomic Southern analyses and RNA hybridizations.

##### 3.1.2. *Ca4h*

The largest chile cDNA clone for *Ca4h* had a 1.8 kb insert, with a 1410 nucleotide ORF encoding 469 amino acids, followed by a 236 nucleotide 3'UTR (AF088847). The chile ORF had a greater than 88% amino acid sequence identity to *Ca4h* genes from *Catharanthus*, *Populus*, *Glycyrrhiza*, and *Glycine* (Z32563 [18]; D82815 [19]; D87520 [20]; X92437 [21]). Based on these alignments and sequence analysis, the chile clone is not full-length, the habañero cDNA clone for *Ca4h* is missing all of the 5'UTR and approximately 108 bp corresponding to the first 36 amino acids. The entire cDNA sequence was used as a probe in the subsequent genomic Southern analyses and RNA hybridizations.

##### 3.1.3. *Comt*

The chile cDNA clone for *Comt* had a 1.4 kb insert, corresponding to a full-length transcript (AF081214, [31]). The chile ORF had 85% amino acid sequence identity to *Comt* genes from *C. annuum* or tobacco (U83789 [24]; X74452, X74453 [23]). The entire cDNA sequence was used as a probe in the subsequent genomic Southern analyses and RNA hybridizations.

### 3.2. Small gene families in chile encode *Pal*, *Ca4h*, and *Comt*

Genomic Southern blots containing DNA from three chile lines, two different types of pungent *C. chinense* habañero, PI 1720 (pendant fruit) and PI 1678 (erect fruit), and a non-pungent mutant *C. chinense*, PI 1721 were probed with each of the three chile cDNA clones. All three cDNA probes detected multiple fragments in the genomic digests of all chile lines (Fig. 2). The cDNA sequences for these genes have internal HindIII sites, one in *Pal*, one in *Comt* and four in *Ca4h*. Some of the genomic fragments must represent internal HindIII sites within the gene sequences, and some of the fragments represent multiple copies of each gene in the genome. Based on the complexity of the hybridization pattern, we predict that *Pal*, *Ca4h*, and *Comt* are each comprised of small gene families. While there were polymorphic bands that

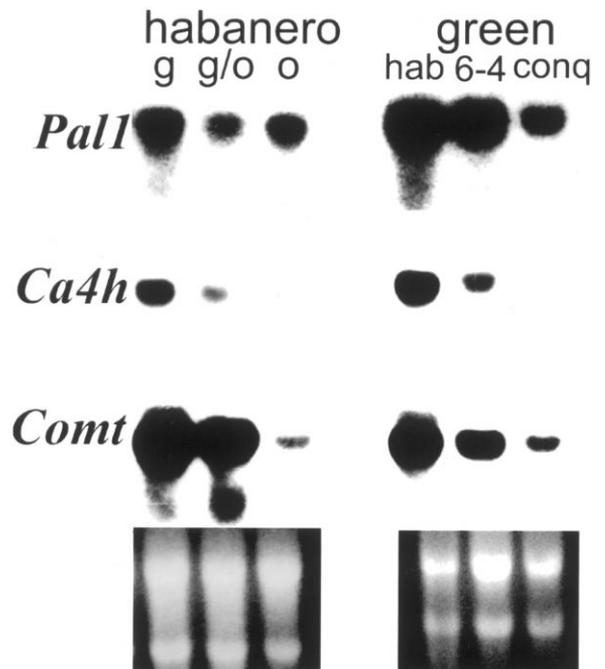


Fig. 3. Transcript accumulations of *Pal*, *Ca4h*, and *Comt* in chile placental tissues. Total RNA (7.5  $\mu$ g) from habañero placental tissue from immature green (g), green/orange (g/o) or mature orange (o) fruit or from placental tissue from immature green habañero (hab), NuMex 6-4 (6-4) or NuMex Conquistador (conq) fruit were electrophoresed, blotted, and probed with the habañero cDNA clones for *Pal*, *Ca4h*, and *Comt*. A photograph of the ribosomal RNA is shown in the bottom panel for load comparison.

distinguished each *C. chinense* line for each gene probe, the hybridization pattern for each gene probe was similar in the three chile lines.

### 3.3. Transcript accumulations of *Pal*, *Ca4h*, and *Comt* in placental tissue

Transcript accumulations of *Pal*, *Ca4h*, and *Comt* were compared in RNA samples isolated from placental tissue in habañero fruit at three different maturity stages. Northern blots demonstrated that generally, transcript levels declined during fruit development (Fig. 3). This developmental trend was the most pronounced for *Ca4h* and *Comt* transcripts. The same trend in transcript accumulation was observed in placental samples at three different maturity stages from a mild chile type *C. annuum* c.v. NuMex 6-4 (data not shown).

Total capsaicinoid levels had been determined on a subset of the fruit used for the RNA isolation shown in Fig. 3. Dried fruit were ground and extracted with acetonitrile and the total capsaic-

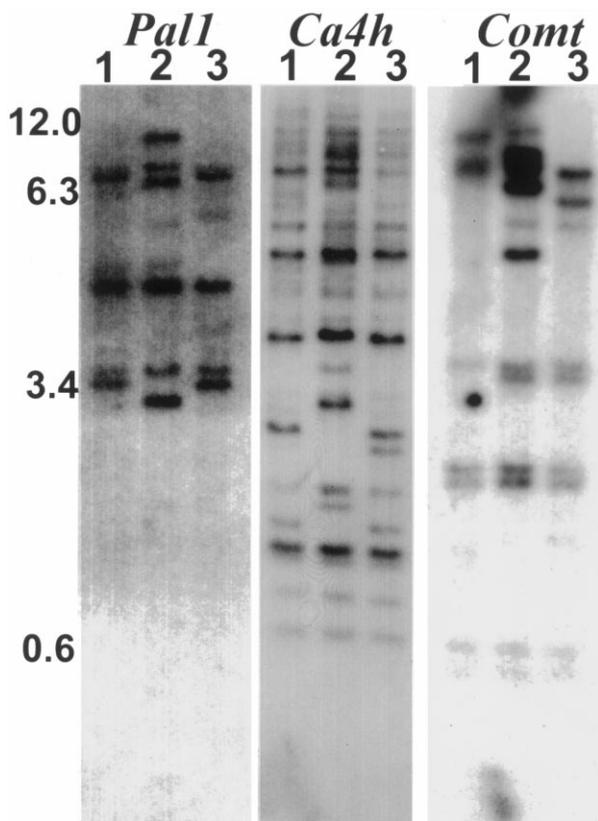


Fig. 2. Genomic Southern blot analysis. DNA (4  $\mu$ g) from non-pungent *C. chinense* PI 1721 (lane 1), or pungent *C. chinense*, habañero PI 1720 (lane 2), and PI 1678 (lane 3) were digested with HindIII, electrophoresed, blotted, and probed with the habañero cDNA clones for *Pal*, *Ca4h*, and *Comt* as indicated. The sizes of selected markers are indicated in kb on the left.

cinoid content was determined by HPLC. Habañero fruit at green, green/orange and orange stages had 115, 154 and 161 mg capsaicinoid/g dry weight fruit respectively. Transcript levels for *Pal*, *Ca4h*, and *Comt* were maximal in placental tissues prior to the maximal accumulation of the capsaicinoids.

Placental transcript accumulations of *Pal*, *Ca4h*, and *Comt* in immature fruit were compared in RNA samples isolated from habañero, NuMex 6-4, and NuMex Conquistador (Fig. 3). Transcript levels for *Pal*, *Ca4h*, and *Comt* were highest in habañero, intermediate in NuMex 6-4 and lowest in NuMex Conquistador. Of these chiles, habañero accumulates the highest levels of capsaicinoids and is the most pungent, NuMex 6-4 is mild, and NuMex Conquistador is a non-pungent chile with the same fruit pod shape and size as NuMex 6-4. Again, the most striking reduction in transcript accumulation is observed for *Ca4h* and *Comt*.

### 3.4. *Pal*, *Ca4h*, and *Comt* expression in fruit wall, placenta, and seed: effect of fruit pungency

Transcript accumulations for *Pal*, *Ca4h*, and *Comt* were determined in four different chile fruit types: habañero, jalapeño, NuMex Joe E. Parker, and CalWonder. The pungency levels for these four chiles are listed in Table 1. This larger collection of chile fruit types, representing different pod shapes, was assayed to test the correlation between placental transcript accumulations of phenylpropanoid genes and fruit pungency levels. Transcript levels were compared at two different stages in fruit development, immature and mature, and in three different tissues: fruit wall, placenta, and seed. Transcript levels were normalized for RNA load and expressed as relative intensities (Fig. 4).

In general, transcript levels for *Pal*, *Ca4h*, and *Comt* were barely detectable in mature fruit relative to immature fruit. This was observed in all four chile types and in all three fruit tissues: fruit wall, placenta, and seed (Fig. 4). Little to no relationship between fruit pungency and transcript level could be observed in samples from mature fruit.

In immature fruit tissues, several trends in *Pal*, *Ca4h*, and *Comt* expression were identified (Fig. 4). (1) For *Pal* and *Ca4h*, within a chile genotype, transcripts were most abundant in seeds, followed

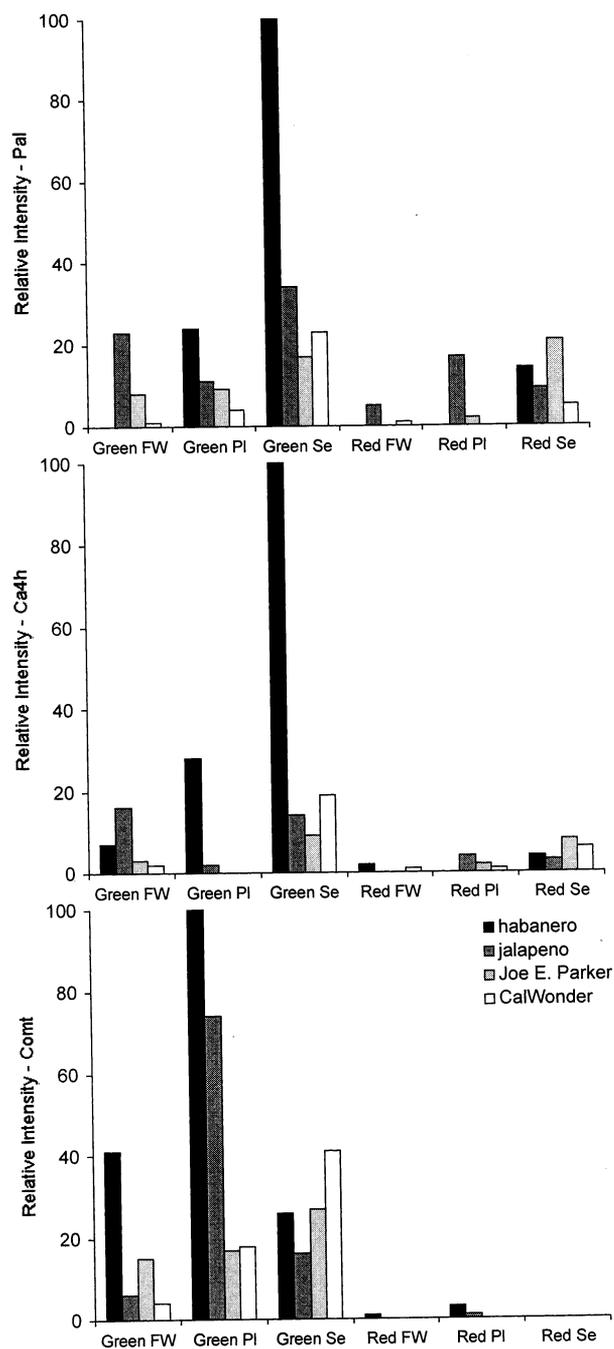


Fig. 4. Relative transcript accumulations of *Pal*, *Ca4h*, and *Comt* in chile fruit. RNA samples from seed (Se), placental tissue (PI) or fruit wall (FW), isolated from habañero, jalapeño, Joe E. Parker, and CalWonder at two developmental stages, immature (green) or mature (red). Total RNA (~7.5 µg) were electrophoresed and probed with the habañero cDNA clones for *Pal*, *Ca4h*, *Comt* and 25S ribosomal RNA. The hybridization on the blots was detected and quantified using a phosphorimager. Transcript levels for 25S ribosomal RNA normalized the transcript levels for each gene probe, and then the tissue with the maximal transcript level for each phenylpropanoid gene was set to 100%. The relative transcript levels for each sample are represented.

by placenta followed by fruit wall. This trend was also true for *Comt* for the two mildest chiles, but not for habañero and jalapeño. Placental samples had the highest transcript levels for *Comt*, for these two hottest chiles. (2) Among chile genotypes, and across fruit tissues, a sample from habañero had the highest transcript levels for *Pal*, *Ca4h*, and *Comt*. For each gene, the maximal relative intensity was in a habañero tissue. For *Pal* and *Ca4h* maximal accumulation was in habañero seed, for *Comt* maximal accumulation was in habañero placenta. (3) In placental samples from immature fruit there was a correlation between fruit pungency level and transcript abundance for all three genes. Transcript levels for *Pal*, *Ca4h*, and *Comt* were the highest in habañero, followed by jalapeño, NuMex Joe E. Parker and then Cal Wonder. The germplasm rank order, for transcript abundance for each of the three genes in the placenta, is the same rank order for capsaicinoid accumulation in the placenta.

### 3.5. Differential screening of habañero placental library

The cDNA library of habañero placental transcripts was screened to identify transcripts that were abundant in habañero samples and undetectable in non-pungent samples. These transcripts are predicted to encode genes related to capsaicinoid biosynthesis. To date, there were relatively few cDNA clones identified, ~10 unique clones, as abundant in habañero placental tissue and low abundance in non-pungent placental tissue. Two of these clones have been well characterized and appear to encode genes for enzymes on the capsaicinoid biosynthetic pathway. Analysis of the other clones is underway.

#### 3.5.1. *Kas*

This cDNA clone (AF085148, [32]) contains a complete open reading frame, 488 amino acids. The predicted translation product is 75% identical and 88% similar to 3-keto-acyl-ACP synthases from plants based on BLASTP analysis (AF026148; L13242; U24177; M60410, [33]). There is a plastid targeting sequence present in the amino-terminus of the predicted translation product, based on PSORT analysis. A *Kas1* activity is predicted to extend isobutyryl CoA to make the branched chain fatty acid (Fig. 1).

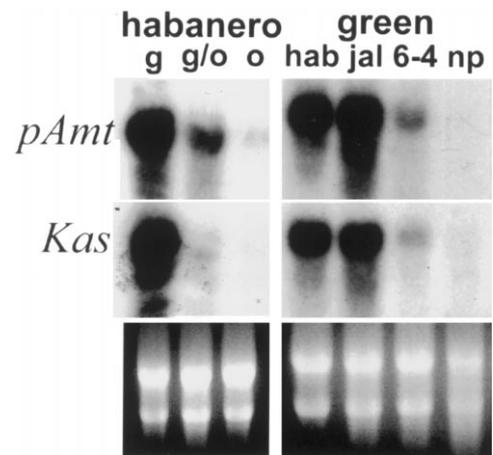


Fig. 5. Transcript accumulations of pAmt and *Kas* in chile placental tissues. Total RNA (15 µg) from habañero placental tissue from immature green (g), green/orange (g/o) or mature orange (o) fruit or from placental tissue from immature green habañero (hab), jalapeño (jal), NuMex 6-4 (6-4) or PI 1721 (np) fruit were electrophoresed, blotted, and probed with the habañero cDNA clones for pAmt, and *Kas*. A photograph of the ribosomal RNA is shown in the bottom panel for load comparison.

#### 3.5.2. *pAmt* a putative aminotransferase

This cDNA clone (AF085149, [34]) contains a complete open reading frame, 459 amino acids. The predicted translation product is 40% identical and 60% similar to a bacterial class III pyridoxal-phosphate-dependent aminotransferase, (Q53196, [35]). All seven conserved domains in class III pyridoxal-phosphate-dependent aminotransferases are present in the chile cDNA clone, based on a BLOCKS analysis [29]. The conversion of vanillin to vanillylamine on the capsaicinoid pathway would require an aminotransferase activity (Fig. 1).

### 3.6. Expression of pungency-related cDNA clones

The differential pattern of expression predicted for pAmt and *Kas* based on the screening of the placental library was confirmed with northern blots (Fig. 5). Transcripts for both pAmt and *Kas* were very abundant in placental tissue early in fruit development and dropped below detectable limits as the fruit matured. Transcripts for both pAmt and *Kas* were very abundant in habañero and jalapeño placental tissue, reduced in NuMex 6-4, and not detectable in non-pungent *C. chinense* placental tissue. These patterns of expression were very similar to those determined for *Pal*, *Ca4h* and *Comt*.

Transcript accumulations for pAmt and *Kas* in other organs of the habañero chile plant were measured (Fig. 6). Transcripts for pAmt and *Kas* appear to be placental-specific. Significant levels of transcripts were detected only in immature placental tissue, low to undetectable levels of transcripts were observed in RNA samples from fruit tissues of seed or fruit wall, flowers, leaves, stem or roots. It should be pointed out that gene-specific probes for these genes were not used, the entire cDNA clone was used as a probe on the RNA blots. The organ-specific pattern of expression and the correlation of transcript abundance with fruit pungency support the hypothesis, that the habañero cDNA clones for *Kas* and pAmt encode the capsaicinoid biosynthetic activities for elongation of isobutyryl CoA and transamination of vanillin respectively.

#### 4. Discussion

The phenylpropanoid pathway is a key pathway for plant secondary metabolism. Plants synthesize a wide array of products from intermediates on this pathway: lignin, flavonoids, phytoalexins, pigments, [36]. One phenylpropanoid intermediate, ferulic acid, is the precursor for capsaicinoids. In

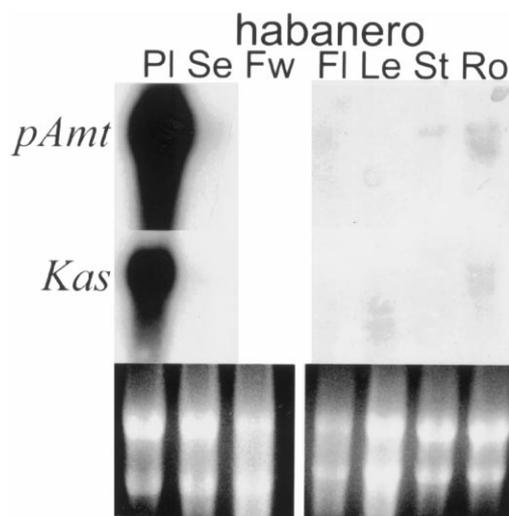


Fig. 6. Transcript accumulations of pAmt and *Kas* in habañero tissues. Total RNA (15  $\mu$ g) from habañero tissue: placenta from immature green fruit (PI), seed from immature green fruit (Se) fruit wall from immature green fruit (FW), flowers (Fl), leaves (Le), stems (St) and roots (Ro) were electrophoresed, blotted, and probed with the habañero cDNA clones for pAmt, and *Kas*. A photograph of the ribosomal RNA is shown in the bottom panel for load comparison.

this report we demonstrated that transcript levels for three enzymes on the phenylpropanoid branch of the capsaicinoid pathway were positively correlated with the accumulation of capsaicinoids in chile placental tissue. These results suggest that the degree of pungency in the fruit may be controlled in part by transcription of the structural genes on this pathway. Using differential expression as a screening criterion, two novel genes predicted to encode additional enzymatic steps on the capsaicinoid pathway were identified.

To monitor expression of the phenylpropanoid pathway, cDNA forms of *Pal*, *Ca4h* and *Comt* were isolated from a library of placental transcripts from immature habañero fruit. DNA sequence analysis demonstrated that the clones selected by hybridization to heterologous probes did in fact encode genes for *Pal*, *Ca4h* and *Comt*. There are many plant *Pal* sequences, *Ca4h* sequences [37], and *Comt* sequences [38] deposited in GenBank. The predicted translation products for the chile cDNA clones all had greater than 85% sequence identity with other plant genes for these sequences.

Genomic Southern analyses suggest that *Pal*, *Ca4h*, and *Comt* are represented as small gene families in *Capsicum*. The *Pal* gene family has been described in a number of Solanaceous plants. In tobacco and tomato, *Pal* is encoded by approximately four genes [14,16]; in potato, over 40 genes encode *Pal* [13]. *Ca4h* is a cytochrome P450-dependent monooxygenase, and as such may be considered a member of a superfamily of genes; its subfamily is designated CYP73A [37]. The specific member of the superfamily, *Ca4h* is a single gene in *Arabidopsis thaliana* [39], but a small gene family in alfalfa and *Populus* [17,19]. *Comt* is also a member of a superfamily of genes, the *O*-methyltransferase superfamily [38]. *Comt* belongs to the sub-group of *o*-diphenol-*O*-methyltransferases (EC 2.1.1.6). *Comt* is encoded by a small gene family, two members in tobacco and possibly in *Capsicum annuum* [23,24]. Based on our Southern blot analysis we predict that *Pal*, *Ca4h*, and *Comt* are encoded by small gene families, approximately three members in *Capsicum chinense*. These gene family sizes are similar to the sizes of the gene families for these genes in other plants.

Using the chile cDNA clones as probes, expression patterns for *Pal*, *Ca4h*, and *Comt* were measured in chile fruit of differing pungency class and

at different developmental stages. Expression of phenylpropanoid genes in placental tissue varied both developmentally and as a function of pungency. For all fruit types, and for all three genes, transcript levels were higher in the immature fruit tissues: fruit wall, placental or seed. In placenta from immature fruit, transcript levels for all three genes were positively correlated with fruit pungency levels. These patterns of expression are consistent with the proposed capsaicinoid pathway. If capsaicinoids are synthesized as proposed in Fig. 1, then transcript levels for *Pal*, *Ca4h* and *Comt* should be elevated early in placental development. Capsaicinoids start to accumulate early in fruit maturation, 20 days after pollination or approximately during the last 2/3 of fruit development [4,5,9]. Maximal transcript accumulation should precede maximal capsaicinoid accumulation. Similarly, those fruit types that accumulate large quantities of capsaicinoids should have maximal transcript levels for *Pal*, *Ca4h*, and *Comt*. The patterns of transcript accumulation observed for *Pal*, *Ca4h*, and *Comt* in chile placental tissue were essentially as predicted, maximal accumulations early in development and maximal accumulations in the most pungent fruit (Figs. 3 and 4). The high level of expression of *Pal* and *Ca4h* in immature seed (Fig. 4) is consistent with the role of these enzymes in the extensive lignification that occurs in seed development. The differences in transcript levels for these two genes in immature seed from different cultivars, however is not understood. Small amounts of placental tissue contamination in the dissected immature seeds may account for some of the high levels of expression in the habanero seed preparations.

The patterns of expression of *Pal*, *Ca4h* and *Comt* were used as predictors of expression patterns for unknown capsaicinoid biosynthetic genes. A differential screen of the habañero library was performed and two differentially expressed clones have been examined in detail so far. Transcript accumulation for these genes was determined to be both placental-specific and correlated with the degree of pungency in the fruit. DNA sequence analysis predicted the gene products to be a 3-keto-acyl-ACP synthase [32] and an aminotransferase [34]. Capsaicinoid biosynthesis is a placental-unique pathway and transcripts for enzymes involved in capsaicinoid biosynthesis can be expected to accumulate only in this organ.

Biosynthesis of capsaicinoids is likely to require aminotransferase activities to aminate vanillin to vanillylamine and to deaminate the branched-chain amino acid used for fatty acid synthesis (Fig. 1). No sequence similarity was found between *pAmt* and over 20 sequences for branched-chain aminotransferases in GenBank. Instead, sequence similarity with class III pyridoxal phosphate dependent aminotransferases was observed. These sequence similarity results coupled with the placental-specific pattern of expression are the basis for the hypothesis that *pAmt* encodes a vanillin aminotransferase. This hypothesis needs to be tested with biochemical assays.

Biosynthesis of capsaicinoids is likely to require a 3-keto-acyl-ACP synthase activity to elongate branched-chain fatty acids (Fig. 1). Among the different classes of *Kas* (I–IV) [40], the chile clone has the highest degree of amino acid sequence similarity to KAS class I proteins. This class of 3-keto-acyl-ACP synthase is specific for elongation of short to medium chain (4–16 carbons) fatty acids [40], the size range relevant for capsaicinoid fatty acids. Again, these sequence similarity results coupled with the placental-specific pattern of expression support the hypothesis that *Kas* encodes an activity to elongate branched-chain fatty acids. Again, this hypothesis needs to be tested with biochemical assays.

Differential screening of a specialized cDNA library has produced very promising results. Using mutants unable to accumulate and/or synthesize capsaicinoids as the indicators of the differentially accumulated transcripts, we were able to isolate and characterize genes for capsaicinoid biosynthetic activities. We will continue to pursue this approach as a way to rapidly identify genes on this pathway.

## Acknowledgements

We thank Robert Gonzales, (Noble Foundation, Ardmore, OK) for the alfalfa cDNA clones for *Ca4h* and *Comt*; Champa Sengupta-Gopalan, for the soybean cDNA clone for *Pal* and Paul Bosland for capsaicinoid analyses and the chile germplasm. This research was supported in part by the New Mexico Agricultural Experiment Station and NIH grants S06 GM08136 and GM07667-21.

## References

- [1] D.J. Bennett, G.W. Kirby, Constitution and biosynthesis of capsaicin, *J. Chem. Soc. C* (1968) 442–446.
- [2] E. Leete, M. Loudon, Biosynthesis of capsaicin and dihydrocapsaicin in *Capsicum frutescens*, *J. Amer. Chem. Soc.* 90 (1968) 6837–6841.
- [3] J. Andrews, Peppers, The Domesticated *Capsicums*, University of Texas Press, Austin, 1995.
- [4] T. Suzuki, H. Fujiwake, K. Iwai, Intracellular localization of capsaicin and its analogues, capsaicinoid, in *Capsicum* fruit. 1. Microscopic investigation of the structure of the placenta of *Capsicum annuum* var. *annuum* cv. Karayatsubusa, *Plant Cell Physiol.* 21 (1980) 839–853.
- [5] N. Sukrasno, M.M. Yeoman, Phenylpropanoid metabolism during growth and development of *Capsicum frutescens* fruits, *Phytochemistry* 32 (1993) 839–844.
- [6] H. Fujiwake, T. Suzuki, K. Iwai, Intracellular distributions of enzymes and intermediates involved in biosynthesis of capsaicin and its analogues in *Capsicum* fruits, *Agric. Biol. Chem.* 46 (1982) 2685–2689.
- [7] Y. Ohta, Physiological and genetical studies on the pungency of *Capsicum* IV. Secretory organ, receptacles and distribution of capsaicin in the *Capsicum* fruits, *Jap. J. Breeding* 12 (1963) 179–183.
- [8] H. Fujiwake, T. Suzuki, K. Iwai, Capsaicinoid formation in the protoplast from the placenta of *Capsicum* fruits, *Agric. Biol. Chem.* 46 (1982) 2591–2592.
- [9] N. Ochoa-Alejo, J.E. Gomez-Peralta, Activity of enzymes involved in capsaicin biosynthesis in callus tissue and fruits of chili pepper (*Capsicum annuum* L.), *J. Plant Physiol.* 141 (1993) 147–152.
- [10] J. Purseglove, E. Brown, C. Green, S. Robbins, Spices, vol. 1, Longman Inc, London, 1981, pp. 331–439.
- [11] K. Harvell, P.W. Bosland, The environment produces a significant effect on pungency of chiles, *HortSci.* 32 (1997) 1292.
- [12] E.M. Estabrook, C. Sengupta-Gopalan, Differential expression of phenylalanine ammonia-lyase and chalcone synthase during soybean nodule development, *Plant Cell* 3 (1991) 299–308.
- [13] H.-J. Joos, K. Hahlbrock, Phenylalanine ammonia-lyase in potato (*Solanum tuberosum* L) Genomic complexity, structural comparison of two selected genes and modes of expression, *Eur. J. Biochem.* 204 (1992) 621–629.
- [14] S.-W. Lee, J. Robb, R. Nazar, Truncated phenylalanine ammonia-lyase expression in tomato (*Lycopersicon esculentum*), *J. Biol. Chem.* 267 (1992) 11824–11830.
- [15] N. Nagai, F. Kitauchi, M. Shimosaka, M. Okazaki, Cloning and sequencing of a full-length cDNA coding for phenylalanine ammonia lyase from tobacco cell culture, *Plant Physiol.* 104 (1994) 1091–1092.
- [16] L. Pellegrini, O. Rohfritsch, B. Fritig, M. Legrand, Phenylalanine ammonia-lyase in tobacco: molecular cloning and gene expression during the hypersensitive reaction to tobacco mosaic virus and the response to a fungal elicitor, *Plant Physiol.* 106 (1994) 877–886.
- [17] T. Fahrendorf, R.A. Dixon, 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 (1993) 509–515.
- [18] M. Hotze, G. Schroder, J. Schroder, Cinnamate 4-hydroxylase from *Catharanthus roseus*, and a strategy for the functional expression of plant cytochrome P450 proteins as translational fusions with P450 reductase in *Escherichia coli*, *FEBS Lett.* 374 (1995) 345–350.
- [19] S. Kawai, A. Mori, T. Shiokawa, S. Kajita, Y. Katayama, N. Morohoshi, Isolation and analysis of cinnamic acid 4-hydroxylase homologous genes from a hybrid aspen, *Populus kitakamiensis*, *Biosci. Biotechnol. Biochem.* 60 (1996) 1586–1597.
- [20] T. Akashi, T. Aoki, T. Takahashi, N. Kameya, I. Nakamura, S. Ayabe, Cloning of cytochrome P450 cDNAs from cultured *Glycyrrhiza echinata* L. cells and their transcriptional activation by elicitor-treatment, *Plant Sci.* 126 (1997) 39–47.
- [21] C.R. Schopfer, J. Ebel, Identification of elicitor-induced cytochrome P450s of soybean (*Glycine max* L.) using differential display of mRNA, *Mol. Gen. Genet.* 258 (1998) 315–322.
- [22] G. Gowri, R.C. Bugos, W.H. Campbell, C.A. Maxwell, R.A. Dixon, Molecular cloning and expression of alfalfa *S*-adenosyl-L-methionine: caffeic acid 3-*O*-methyltransferase, a key enzyme of lignin biosynthesis, *Plant Physiol.* 97 (1991) 7–14.
- [23] E. Jaeck, F. Martz, V. Stiefel, F. Fritig, M. Legrand, Expression of class I *O*-methyltransferase in healthy and TMV-infected tobacco, *Mol. Plant-Microbe Interact.* 9 (1996) 681–688.
- [24] B. Lee, D. Choi, K.-W. Lee, Isolation and characterization of *o*-diphenol-*O*-methyltransferase cDNA clone in hot pepper (*Capsicum annuum* L), *J. Plant Biol.* 41 (1998) 9–15.
- [25] M.D. Collins, L. Mayer-Wasmund, P.W. Bosland, Improved method for quantifying capsaicinoids in *Capsicum* using high-performance liquid chromatography, *HortSci.* 30 (1995) 137–139.
- [26] T.L. Kahn, S.E. Fender, E.A. Bray, M.A. O'Connell, Characterization of expression of drought- and ABA-regulated tomato genes in the drought-resistant species *Lycopersicon pennellii*, *Plant Physiol.* 103 (1993) 597–605.
- [27] R. Pelle, N.B. Murphy, Northern hybridization: rapid and simple electrophoretic conditions, *Nucl. Acids Res.* 21 (1993) 2783–2784.
- [28] S.F. Altschul, T.L. Madden, A.A. Schaffer, et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucl. Acids Res.* 25 (1997) 3389–3402.
- [29] S. Henikoff, J.G. Henikoff, Protein family classification based on searching a database of blocks, *Genomics* 19 (1994) 97–107.
- [30] K. Nakai, M. Kanehisa, A knowledge base for predicting protein localization sites in eukaryotic cells, *Genomics* 14 (1992) 897–911.
- [31] J. Curry, M. Mendoza, M. O'Connell, Nucleotide sequence of a caffeic acid 3-*O*-methyltransferase gene (Accession No. AF081214) from habañero chile PGR98-170, *Plant Physiol.* 118 (1998) 711.

- [32] M. Aluru, J. Curry, M. O'Connell, Nucleotide sequence of a 3-oxoacyl-[Acyl-Carrier-Protein] synthase (beta-ketoacyl-ACP Synthase) gene (accession No. AF085148) from habañoero chile PGR98-181, *Plant Physiol.* 118 (1998) 1102.
- [33] M. Siggaard-Anderson, S. Kauppinen, P. Von Wettstein-Knowles, Primary structure of a cerulenin-binding beta-ketoacyl-[acyl carrier protein] synthase from barley chloroplasts, *Proc. Natl. Acad. Sci. USA* 88 (1991) 4114–4118.
- [34] M. Aluru, J. Curry, M. O'Connell, Nucleotide sequence of a probable aminotransferase gene (accession No. AF085149) from habañoero chile PGR98-182, *Plant Physiol.* 118 (1998) 1102.
- [35] C. Freiberg, X. Perret, W.J. Broughton, A. Rosenthal, Sequencing the 500Kb GC-rich symbiotic replicon of *Rhizobium* sp. NGR234 using dye terminators and a thermostable 'sequanase': a beginning, *Genome Res.* 6 (1996) 590–600.
- [36] R. Dixon, N. Paiva, Stress-induced phenylpropanoid metabolism, *Plant Cell* 7 (1995) 1085–1097.
- [37] C. Chapple, Molecular-genetic analysis of plant cytochrome P450-dependent monooxygenases, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49 (1998) 311–343.
- [38] R.K. Ibrahim, A. Bruneau, B. Bantignies, Plant *O*-methyltransferases: molecular analysis, common signature and classification, *Plant Mol. Biol.* 36 (1998) 1–10.
- [39] D.A. Bell-Lelong, J.C. Cusumano, K. Meyer, C. Chapple, Cinnamate-4-hydroxylase expression in *Arabidopsis*: regulation in response to development and the environment, *Plant Physiol.* 113 (1997) 729–738.
- [40] S. Kauppinen, Structure and expression of the KAS12 gene encoding a beta-ketoacyl-acyl carrier protein synthase I isozyme from barley, *J. Biol. Chem.* 267 (1992) 23999–24006.