Candidate Gene Strategy for the Study of the Chlorogenic Acid Biosynthesis

C. CAMPA[1], M. VENKATARAMAIAH[2], A. de KOCHKO[1], L. LE GAL[1], M. BOURGEOIS[1], C. MOREAU[1], S. HAMON[1], M. NOIROT[1]

[1]IRD, Génomique des Caféiers, UMR DGPC, 911 Avenue Agropolis, BP 64501, 34304 Montpellier Cédex 5, France
[2]Avestha gengraine Technologies, "Discoverer", 9th floor, ITPL, Bangalore, India

SUMMARY

Chlorogenic acid and other quinic esters are soluble phenolics that accumulate to substantial levels in green beans of some coffee species. If their diversity between species and during plant growth have been studied, very little is known about their actual biosynthetic pathway. A candidate gene strategy involving some genes of the phenylpropanoid pathway in *Coffea canephora* has been initiated. This study will be helpful to describe the metabolic pathway in coffee trees and highlight the major genes which regulated the CGA biosynthesis.

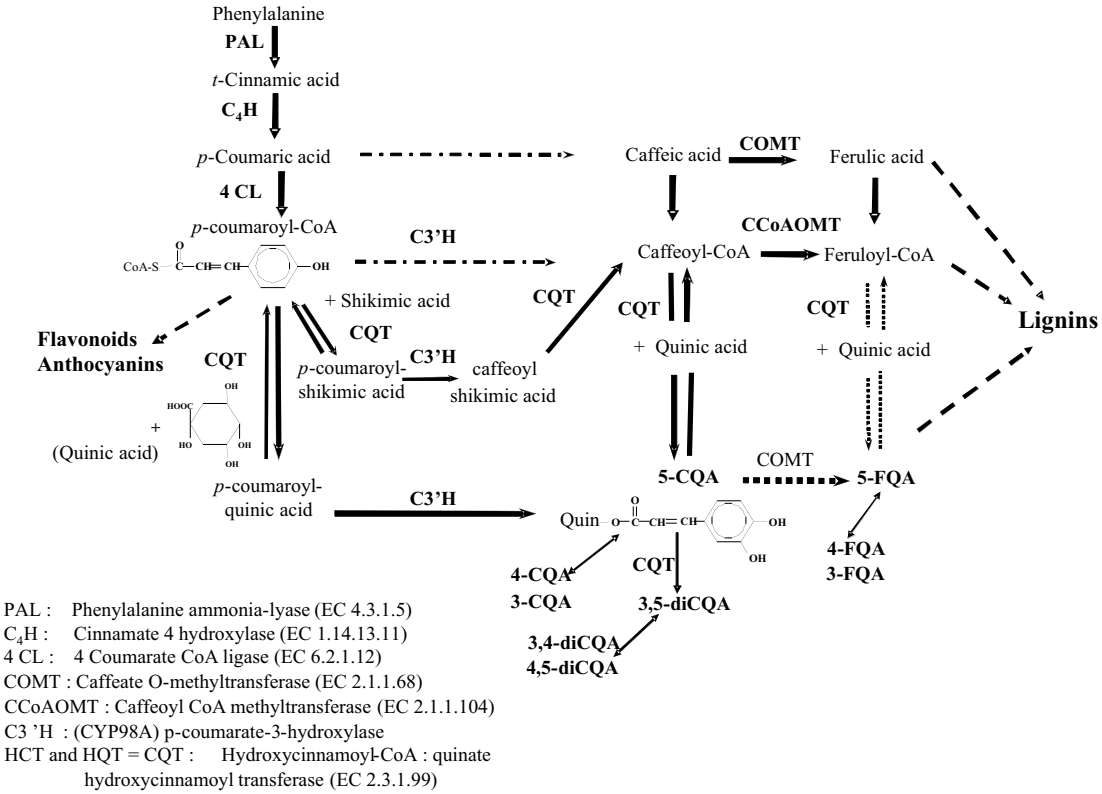
RESUME

L'acide chlorogénique ainsi que d'autres esters de l'acide quinique sont des composés phénoliques solubles pouvant être accumulés de façon importante dans les grains verts de certaines espèces de caféier. Si la diversité entre espèces et durant la croissance des plantes a été étudiée, très peu d'informations sont disponibles sur la voie de biosynthèse complète de ces composés. Pour cette raison, une stratégie gène-candidat a été initiée, prenant en compte certains gènes de la voie des phénylpropanoïdes. Elle permettra de décrire la voie métabolique chez les caféiers ainsi que de mettre en valeur les gènes majeurs qui régulent la biosynthèse du CGA et des autres esters de l'acide quinique.

INTRODUCTION

Chlorogenic acid (CGA, caffeoyl quinic acid, 5-CQA) is the major soluble quinic ester accumulated in plants. It is commonly considered as a storage form of cinnamic acid derivatives and is certainly involved in lignification in plants (Schoch et al., 2001). All the quinic esters, together with the 5-CQA, are often designated as chlorogenic acids (CGAs). These compounds are putatively involved in the protection of plant tissue from damages by oxidative stress, pathogen infection and wounding. They can also intervene in animal health: CGA, by its antioxidant activity, may prevent carcinogenesis (Niggeweg et al., 2004) and analogues of CGAs demonstrated a potent anti-viral activity (King et al., 1999). CGAs are biosynthetically derived from phenylalanine through the phenylpropanoid pathway (Figure 1) which leads to the synthesis of a wide range of compounds including flavonoids, isoflavonoid phytoalexins, coumarins and lignin (Hahlbrock and Scheel, 1989). From the p-coumaroyl-CoA, three possible pathways are proposed. Each involves the same types of enzymatic reactions: esterification and hydroxylation. In cultivated coffee trees, CGAs (CQA but also dicaffeoyl- and feruloyl- quinic acids) accumulate in beans and are mainly responsible for coffee bitterness by their degradation into phenols during roasting (Leloup et al., 1995). Their nature and their diversity between species as well as during plant growth have been widely studied (Anthony et al., 1993; Ky et al., 2001) but their synthesis, transport and accumulation

mechanisms are only partially described (Aerts and Baumann, 1994; Colonna, 1986). In particular, noting is known about the synthesis of the feruloyl quinic acids derived from caffeoyl quinic acids by a methylation. Coffee species can be a useful tool to study CGAs metabolism, particularly *C. canephora* and some wild species as C.*sp* N'Koumbala which largely accumulated CGAs. In this work, we used a candidate gene strategy to isolate some genes encoding enzymes from the phenylpropanoid metabolism. The analysis of their expression pattern and their possible co-location with already identified QTLs associated with CGAs accumulation in green beans will allow us to better describe the CGAs biosynthetic pathway in coffee trees.



PAL : Phenylalanine ammonia-lyase (EC 4.3.1.5)
C_4H : Cinnamate 4 hydroxylase (EC 1.14.13.11)
4 CL : 4 Coumarate CoA ligase (EC 6.2.1.12)
COMT : Caffeate O-methyltransferase (EC 2.1.1.68)
CCoAOMT : Caffeoyl CoA methyltransferase (EC 2.1.1.104)
C3 'H : (CYP98A) p-coumarate-3-hydroxylase
HCT and HQT = CQT : Hydroxycinnamoyl-CoA : quinate hydroxycinnamoyl transferase (EC 2.3.1.99)

Figure 1. Proposed pathways for the synthesis of chlorogenic acids in Coffee.

ABBREVIATIONS

PAL: phenylalanine ammonia-lyase, C3'H: CYP98A, *p*-coumarate hydroxylase; CQT: coumaroyl quinate transferase, COMT: caffeate methyltransferase, CCoAOMT: caffeoyl-CoA methyltransferase

MATERIALS AND METHODS

Total DNA was purified according to the protocol of Ky et al. (2000), from leaves recollected on three trees of *Coffea canephora* Pierre (CAN) maintained in tropical greenhouses at the IRD research centre in Montpellier (France), and from trees of *C. liberica* var. Dewevrei (DEW), *C. pseudozanguebariae* (PSE), their F1 hybrids and the backcross progeny (BCDEW) maintained at the IRD research station in Man (Côte d'Ivoire).

CAN fruits were harvested on trees at the IRD research station. They were divided into three batches according to the husk colour: green (stage 1: F1), yellow (stage 2: F2) and red (stage 3: F3), immediately frozen in liquid nitrogen and maintained at –80°C until RNA extraction.

For most of the genes, consensus primers were designed after aligning corresponding sequences downloaded from GenBank. For *PAL* gene, primers were obtained according to the partial sequence of a *C. arabica PAL* gene (accession n° AF218454). For *C3'H*, semi-degenerated primers were designed according to the CODEHOP strategy (Rose et al., 1998). After amplification, the resulting fragments were purified using the EZNA gel extraction kit (Omega BioTek) and cloned into a pCR4-TOPO plasmid using the "TOPO TA for sequencing" kit (Invitrogen, Groningen, The Netherlands). Cloned fragments were sequenced by MWG-Biotech (Ebersberg, Germany).

Specific primers were designed when partial sequences were obtained from CAN. Genomic DNA from the other species was then amplified to determine gene organisation (exon/intron position) and interspecific polymorphism for genetic mapping.

Two *C. canephora* cDNA libraries were constructed, one from young leaves and the other from fruit (pool of immature and mature fruits), constructed with the "Zap Express cDNA synthesis kit" (Stratagene).

A leaf and a fruit EST libraries, derived from the previous cDNA libraries, were used to isolate the full length cDNAs of the candidates genes.

Total RNA was extracted from samples ground in liquid nitrogen. Extraction buffer (Corre et al., 1996) was added immediately to avoid thawing of the samples. After centrifugation (4.800 rpm, 30 min., 4°C), 7 mL of supernatant was ultracentrifuged at 20°C for 20h at 32.000 rpm (Beckman L7, SW41T1 rotor) in 3 ml of 5.7 M CsCl. Classical purification was then realised and the RNA was dissolved in water and stored at -80°C.

For RT-PCR analysis, 2 µg of total RNA from the different fruit samples were reverse-transcribed by AMV Reverse Transcriptase, with oligo dT primer (0.5 µg/µg of total RNA) using the kit Universal Riboclone cDNA synthesis-Promega. A second strand synthesis allowed to enrich the transcripts in the cDNA pool. A part (1/10) of the total cDNA was subjected to 35 cycles of PCR amplification.

Sequence similarities were investigated using the Blastx program maintained at NCBI. When a similarity greater than 90% was obtained for a protein sequence, it was considered that there was no ambiguity on the predicted gene function. Determination of ORFs and the derivation of predicted protein sequences, homology searches, multiple alignments, deduction of consensus sequences and primer design, were carried out using the DNASTAR package (Version 5.06 DNASTAR Incorporated, Madison, Wis., USA). Statistical analysis was done on Statistica.

Substrate specificity of C3'H was tested expressing the different *C3'H* genes in yeast. Activity analysis of the recombinant protein was done in presence of p-coumaric acid, p-coumaroyl shikimate or p-coumaroyl quinate and evaluated measuring by LS/MS the final product formed.

RESULTS

The first enzyme of the phenylpropanoid pathway is the phenylalanine ammonia-lyase (PAL). Analyses of EST sequences indicated that almost three genes encoding PAL are expressed in CAN fruits. The full length of two of them has been isolated: *PAL1* gene was obtained by screening the cDNA fruit library (*pF6*, accession n° AF460203) and *PAL2* by screening EST sequences. The open reading frame of *PAL1* and *PAL2* encode a polypeptide of 717 and 711

amino acids, respectively. The predicted molecular mass and theoretical pI (77.9 and 6.27; 77.2 kDa and 6.14, respectively) are consistent with the size determined for PAL polypeptides from other plants (72-83 kDa). The derived AA sequence showed the conserved motifs that are found in all PAL sequences, in particular the PAL active site: G-[STG]-[LIVM]-[STG]-[AC]-S-G-[DH]-L-x-P-L-[SA]-x(2)-[SAV]. *PAL1* and *PAL2* AA sequences shared only 84.6% similarity and the corresponding genomic sequences also displayed some differences. The two *PAL* genes differed by the length of their UTRs (3' and 5') and that of their intron (1800 and 1000 bp respectively).

The expression pattern of the two PAL genes was quite different. *PAL1* was expressed at all the observed fruit development and maturation stages and *PAL 2* was only expressed at the latest stage of fruit maturation (Figure 2).

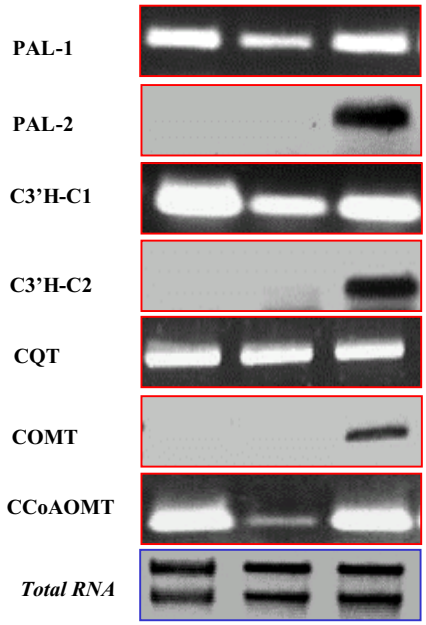


Figure 2. Semi-quantitaive RT-PCR on C. canephora fruits at different stages of maturity.

The esterification that leads to chlorogenic acids is performed by the reversible hydroxycinnamoyl-CoA: quinate hydroxycinnamoyl transferase (CQT). It catalyses the esterification of p-coumaroyl-CoA with quinate or shikimate and of caffeoyl-CoA with quinate only. Little is known about this enzyme and only some gene sequences are available in data banks. One full length of the coding part of a *CQT* gene was found in the fruit EST library. It possessed a 1299 nucleotides long ORF, which encodes a 433 AA long protein.

The sequence contained the HXXXD and DFGWG motifs that are thought to comprise the active site of the acyl-CoA transferases. Analysis of its genomic sequence showed that this gene had no intron.

The expression pattern showed that this gene is transcribed at the same level at each stage of fruit maturation.

Hydroxylation, the other step leading to chlorogenic acids, is catalysed by the *p*-coumarate-2-hydroxylase (C3'H). This cytochrome P450 (group of the CYP98A) is a microsomal enzyme which hydroxylates coumaroyl-quinate or -shikimate, but also *p*-coumaric acid or p-coumaroyl-CoA. Two full lengths *C3'H* cDNA genes have been isolated using the

CODEHOP strategy and PCR screening of the fruit cDNA library. Both *C3'H-C1* and *C3'H-C2* had an ORF of 1524 bp. Each gene encodes a protein of 508 AA sharing only 75% identity between them. At the amino acid level, *C3'H-C1* and *C3'H-C2* showed 86% and 89% identity, respectively, with CYP98A13 from *Ocimum basilicum*. Genomic studies showed that these two genes had two introns of equal length (3000 and 100 bp).

Expression of each of these genes in yeast conducted to the synthesis of a recombinant protein. C3'H-C1 metabolised preferentially *p*-coumaroyl quinate instead of *p*-coumaroyl shikimate. The converse was observed with C3'H-C2. Both did not take coumaric acid as substrate. These data indicated that C3'H-C1 can be the enzyme that directly gives 5-CQA from the *p*-coumaroyl quinic acid.

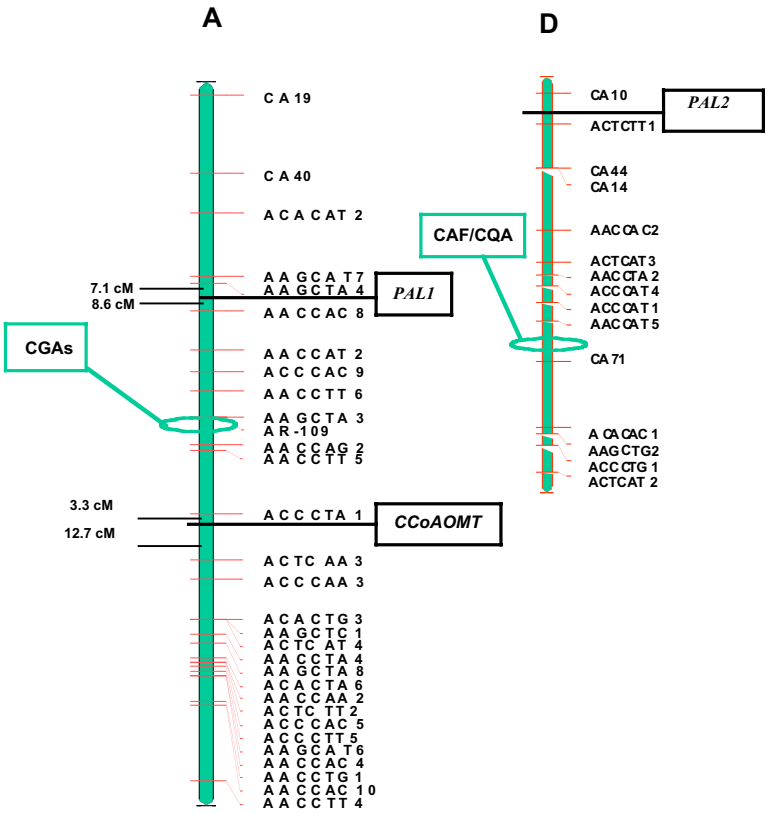


Figure 3. Linkage groups A and D of a (PSE X DEW) W DEW genetic map showing the location of the CCoAOMT, PAL1 and PAL2 genes ant two QTL related to fruit content (CGAs: total CGAs content; CAF/CQA: ratio of caffeine.

The pattern of expression of the two genes in fruits was quite different, *C3'H-C1* being always expressed and *C3'H-C2*, as *PAL2*, only clearly expressed in the last stage of maturation.

Biosynthesis of 5-FQA necessitates a methylation, which can occur on caffeic acid, catalyzed by the COMT, and on caffeoyl-CoA, catalyzed by the CCoAOMT (classical ways of lignification) or perhaps on the 5-CQA, catalyzed probably by a COMT. By screening the fruit cDNA library with specific PCR primers, a full length cDNA of COMT was isolated. The ORF was 1053 bp long, encoding a protein of 350 AA. The presence of an intron of about 700 bp near the 5'-end has been shown by genomic DNA analysis.

This gene was exclusively expressed in mature fruits, as observed for *PAL2* and *C3'H-C2*.

For the methylation of caffeoyl-CoA, analyses of EST sequences indicated that at least one gene encoded CCoAOMT and was expressed in CAN fruits. The coding part of the *CCoAOMT* gene was 744 bp long and encoded a protein of 247 AA. Analysis of genomic DNA showed the presence of a 500 bp intron.

Using specific PCR primers of *CCoAOMT*, it has been noticed that this gene is expressed preferentially in very young and mature fruits, similarly to *PAL1* and *C3'H-C1*.

A preliminary genetic mapping, using an offspring population from a backcross between PSE and DEW has been done with *PAL1*, *PAL2* and *CCoAOMT* genes. It showed that *PAL1* and *CCoAOMT* genes are located on the same linkage group (group A), in the vicinity of a QTL related to the fruit CGA content (Figure 3). *PAL2* is located on the linkage group D. As backcross heterozygous offspring displaying the PSE allele for *PAL1* and *CCoAOMT* showed a decrease of 15 and 11%, respectively for their CQA content in the fruits compared to the homozygous progenies for the DEW allele, the effect of *PAL1* and *CCoAOMT* genes on CQA content can be considered as very low.

DISCUSSION

PAL1, *C3'H-C1* and *CCoAOMT* are three genes showing the same pattern of expression in CAN fruits. Interestingly, *PAL1* and *CCoAOMT* are located in the same region of the same linkage group and seemed to intervene weakly on the fruit CGA content. *PAL2*, *C3'H-C2* and *COMT* seemed to be exclusively expressed in mature fruits, and this expression cannot be related to the CGA level in fruits. The particular expression pattern of *CQT*, which is always expressed at the same level, can be explained by the number of different esterifications/de-esterifications that can be accomplished by the enzyme it encodes.

CGA biosynthesis *via* the phenylpropanoid pathway seemed to begin through the activity of a PAL encoded by the gene *PAL1* and to continue through p-coumaroyl quinic acid synthesis, by the activity of a CQT and then a C3'H encoded by *C3'H -C1*. These genes only explain 11% of the fruit CGA content. Other paths of biosynthesis can explain this weak effect of the genes on the CGA content. As suggested by Niggeweg et al. (2004), coffee plants, as other CGA accumulating species, can synthesised CGA in majority through caffeoyl-CoA, using the CQT esterase function between caffeoyl-CoA and CGA. Another reason could be that CGA content in fruit is not only due to their local synthesis. Accumulation mechanisms of CGA synthesized in other organs or modifications in catabolism can also be a source of increased content in fruits, without any relation to neosynthesis (Matsuda et al., 2003).

The effect of regulatory genes might be also more important to explain CGA content variation than the structural genes encoding the enzymes.

REFERENCES

Aerts RJ., Baumann TW. (1994). Distribution and utilization of chlorogenic acid in *Coffea* seedlings, *J. Exp. Bot*. **45**: 497-503.

Altschul SF., Gish W., Miller W., Myers EW., Lipman DJ. (1990). Basic local alignement search tool. *J. Mol. Biol*. **215**: 403-410.

Anthony F., Noirot M., Clifford MN. (1993). Biochemical diversity in the genus Coffea L.: chlorogenic acids, caffeine. and mozambioside contents. *Gen. Res. Crop Evol.* **40**: 61-70.

Campa C., Noirot M., Bourgeois M., Pervent M., Ky CL., Chrestin H., Hamon S., de Kochko A. (2003). Genetic mapping of a caffeoyl-coenzyme A 3-O-methyltransferase gene in coffee trees. Impact on chlorogenic acid content. *Theor Appl Genet*. **107**: 751-756.

Colonna JP. (1986). Biosynthèse et renouvellement de l'acide chlorogénique et des depsides voisins dans le genre *Coffea*. II. Incorporation de la radioactivité de la L-phénylalanine-^{14}C dans l'acide chlorogénique des feuilles de caféier, en présence ou non de compétiteurs isotopiques. *Café Cacao Thé*, **4**: 247-258.

Corre F., Henry Y., Rode A., Hartmann C. (1996). Em gene expression during somatic embryogenesis in the monocot *Triticum aestivum* L. *Plant Sci.* **117**: 139-149.

Distantly Related Sequences. *Nucleic Acids Research.* **26**(7):1628-1635.

Hahlbrock K., Scheel D. (1989). Physiology and molecular biology of phenylpropanoid metabolism. *Annu. Rev. Plant Physiol. Plant Mol. Biol*. **40**: 347-369.

King PJ., Ma G., Miao W., Jia Q., McDougall BR., Reinecke MG., Cornell C., Kuan J., Kim TR., Robinson WE. Jr. (1999). Structure-activity relationships: analogues of the dicaffeoylquinic and dicaffeoyltartaric acids as potent inhibitors of human immunodeficiency virus type 1 integrase and replication *J Med Chem* **42**(3): 497-509.

Ky CL, Barre P., Lorieux M., Trouslot P., Akaffou S., Louarn J., Charrier A., Hamon S., Noirot M. (2000). Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (Coffea Sp.). *Theor. Appl. Genet*. **11**: 669-676

Ky CL., Louarn J., Dussert S., Guyot B., Hamon S., Noirot M. (2001). Caffeine, trigonelline, chlorogenic acids and sucrose diversity in wild *Coffea arabica* L. and *C. canephora* P. accessions. *Food Chem*. **75**: 223-230.

Leloup V., Louvrier A., Liardon R. (1995). Degradation mechanisms of chlorogenic acids during roasting. *Proc. Int. Congr. ASIC 16*: 192–198.

Matsuda F., Morino K., Miyashita M., Miyagawa H. (2003). Metabolic flux analysis of the phenylpropanoid pathway in wound-healing potato tuber tissue using stable isotope-labeled tracer and LC-MS spectroscopy. *Plant Cell Physiol*. **44**: 510-517.

Niggeweg R., Michael AJ., Martin C. (2004). Engineering plants with increased levels of the antioxidant chlorogenic acid. *Nature Biotechnology*. **22**(6): 746-754.

Rose T., Schultz E., Henikoff J., Pietrokovski S., McCallum C., Henikoff S. (1998).Consensus-Degenerate Hybrid Oligonucleotide Primers for Amplification of

Schoch G., Goepfert S., Morant M., Hehn A., Meyer D., Ullmann P., Werck-Reichhart D. (2001). CYP98A3 from Arabidopsis thaliana is a 3'-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. J. Biol. Chem. 276: 36566-36574.