

Sugarcane expressed sequence tags (ESTs) encoding enzymes involved in lignin biosynthesis pathways

Rose Lucia Braz Ramos¹, Francisco Javier Tovar², Ricardo Magrani Junqueira¹,
Fabiane Borges Lino¹ and Gilberto Sachetto-Martins^{1*}

Abstract

Lignins are phenolic polymers found in the secondary wall of plant conductive systems where they play an important role by reducing the permeability of the cell wall to water. Lignins are also responsible for the rigidity of the cell wall and are involved in mechanisms of resistance to pathogens. The metabolic routes and enzymes involved in synthesis of lignins have been largely characterized and representative genes that encode enzymes involved in these processes have been cloned from several plant species. The synthesis of lignins is linked to the general metabolism of the phenylpropanoids in plants, having enzymes (*e.g.* phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H) and caffeic acid O-methyltransferase (COMT)) common to other processes as well as specific enzymes such as cinnamoyl-CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD). Some maize and sorghum mutants, shown to have defective in CAD and/or COMT activity, are easier to digest because they have a reduced lignin content, something which has motivated different research groups to alter the lignin content and composition of model plants by genetic engineering try to improve, for example, the efficiency of paper pulping and digestibility. In the work reported in this paper, we have made an inventory of the sugarcane expressed sequence tag (EST) coding for enzymes involved in lignin metabolism which are present in the sugarcane EST genome project (SUCEST) database. Our analysis focused on the key enzymes ferulate-5-hydroxylase (F5H), caffeic acid O-methyltransferase (COMT), caffeoyl CoA O-methyltransferase (CCoAOMT), hydroxycinnamate CoA ligase (4CL), cinnamoyl-CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD). The comparative analysis of these genes with those described in other species could be used as molecular markers for breeding as well as for the manipulation of lignin metabolism in sugarcane.

INTRODUCTION

Lignins are complex cell wall phenolic heteropolymers associated with both polysaccharides and protein. These complex polymers impart important strengthening and waterproofing properties to plant cell walls, and, therefore, lignin plays a fundamental role in mechanical support, solute conductance and disease resistance in higher plants (Barber and Mitchell, 1997). Lignins form a major component of the terrestrial biomass and represent the second most abundant plant polymer after cellulose. They occur in greatest quantity in the secondary cell walls of specialized cells, which form tissues such as sclerenchyma, xylem and periderm (Boudet, 1998). These heteropolymers result from the oxidative coupling of three cinnamyl alcohol (monolignols), *p*-coumaryl, coniferyl and sinapyl alcohol, giving rise within the lignin polymer to H (hydroxyphenyl), G (guaiacyl) and S (syringyl) units respectively (Barceló, 1997). The relative proportions of these monolignols vary between species and due to the response of the plant to the environment (Boudet, 1998). Generally, Gymnosperms produce guaiacyl lignins (with a small proportion of H

units) whereas Angiosperms produce mainly guaiacyl and syringyl lignins. An exception is found in grasses, which contain lignin derived from all three monolignols. Lignin heterogeneity in grasses is regulated during secondary cell wall deposition to form layers of lignins that can differ both in the amount and in the average monomer composition. Lignin also varies in its composition and quantity between different cell types and between tissues within the same plant (Whetten *et al.*, 1998).

Lignins are one of the major problems in the production of pulp and paper because it needs to be separated from the cellulose fibers. This process consumes large quantities of energy and uses hazardous chemicals. Lignin also limits the digestibility of forage in ruminants because lignin cannot be degraded by any of the anaerobic microorganisms in the rumen, and if digestibility could be controlled fodder could be used more efficiently by livestock.

For these reasons various research groups have tried to alter lignin content and composition of model plants (trees and crops) by genetic engineering with the aim of improving the efficiency of paper pulping and the digestibility

¹Laboratório de Genética Molecular Vegetal, Departamento de Genética, Instituto de Biologia, CCS, Bloco A, sala A2-076 - Universidade Federal do Rio de Janeiro, Ilha do Fundão, 21944-970 Rio de Janeiro, RJ, Brazil.

²Laboratório de Genética Molecular de Eucariontes - Departamento de Genética, Instituto de Biologia, CCS, Universidade Federal do Rio de Janeiro, Ilha do Fundão, 21944-970 Rio de Janeiro, RJ, Brazil.

Send correspondence to G. Sachetto-Martins. E-mail: sachetto@biologia.ufrj.br.

of lignocellulose in ruminants. Few countries have surplus tracts of land that can be regarded as fertile enough to produce high quality fodder, and, in general, human population pressures and the need to supply crops for human consumption ensures that the production of forage for livestock is restricted to poor quality pasture land and crops residue or agroindustrial by-products. The feeds that are available to ruminants in developing countries are fibrous and relatively high in lignocellulose and are usually of low digestibility and deficient in critical nutrients. For these reasons, improving the quality and balance of nutrients as well as the digestibility of fodder are worthwhile and promising objectives (reviewed in Lintu, 1990; Boudet, 1998).

Decreased lignin-content or altered composition would lower the price of pulp by saving chemicals and energy while increasing yield. Moreover, decreased lignin content or alterations in its composition could result in improved extraction of this polymer and the production of better quality and cheaper papers at less cost to the environment. The same applies to forage, because a decrease in lignin would increase the energy content of feed due to the relative increase in the carbohydrate content of the forage fibers and would allow such fibers to be used in feeds at a higher percentage, thus contributing to better livestock management. (Boudet, 1998).

In this paper we present an overview of the expressed sequence tag (EST) present in the sugarcane EST database (SUCEST) that encode the major enzymes in the lignin biosynthesis pathway. The analysis focused on the key enzymes ferulate-5-hydroxylase (F5H), caffeic acid O-methyltransferase (COMT), caffeoyl CoA O-methyltransferase (CCoAOMT), hydroxycinnamate CoA ligase (4CL), cinnamoyl-CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD). The comparative analysis of these genes with those described in other species could result in their being used as molecular markers for breeding as well as the manipulation of lignin metabolism in sugarcane.

MATERIAL AND METHODS

Sequence data, alignment and phylogenetic analysis

A basic local alignment search tool T-Blast-n search (Altschul *et al.*, 1997) was performed using the deduced protein sequence of enzymes involved in the lignin biosynthesis pathway (4CL, F5H, COMT, CCoAOMT, CCR, CAD) as bait sequences against the full SUCEST data bank, using the fragment assembly program Phrap for clustering. The Clustal W program (Thompson *et al.*, 1994) was used to align the different protein standards with the proteins deduced from the SUCEST clusters (only cluster consensus sequences were used).

For the phylogenetic analysis we only used complete protein sequences inferred from alignments with other

known sequences. The analysis were performed using the Molecular Evolutionary Genetics Analysis (MEGA) software (Kumar *et al.*, 1994). The Neighbor-Joining distance method was used and the complete deletion option was adopted on the treatment of amino acid gaps of the multiple alignments. In construction of the phylogenetic tree the confidence levels assigned at various nodes were determined after 1000 replications using the bootstrap test.

RESULTS AND DISCUSSION

Lignin metabolic pathway

Lignin biosynthesis involves the coordinated regulation of three biosynthetic pathways: the shikimate pathway, the general phenylpropanoid pathway and the lignin branch pathway (Figure 1). The Shikimate pathway is a primary metabolic pathway which leads to the biosynthesis of the amino acids phenylalanine, tyrosine or tryptophan which are subsequently incorporated into a variety of plant products (Herrmann, 1995).

The general phenylpropanoid pathway begins with the deamination of L-phenylalanine to cinnamic acid. This step represents a switch from primary to secondary metabolism in the plant (Barber and Mitchell, 1997), the first step in this pathway being the PAL-catalyzed elimination of NH₃ from L- phenylalanine to give cinnamic acid, (Figure 1). In plants PAL is located in both the cytoplasm and

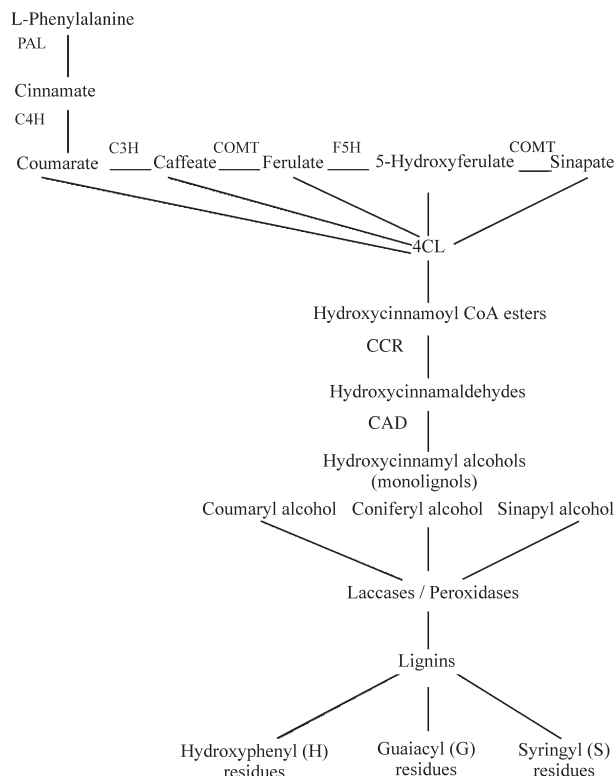


Figure 1 - Conventional lignin biosynthesis pathway, adapted from Boudet (1998).

organelles, this enzyme being encoded in Angiosperms by a multigene family but apparently by a single gene in Gymnosperms. In bean plants DNA analysis and genomic cloning points to the existence of three PAL genes. The next enzyme in the pathway is C4H, which catalyzes the addition of a hydroxyl group at the 4- carbon of cinnamic acid to form *p*-coumaric acid. Some evidence suggest that C4H may play a central role in the regulation of the phenylpropanoid pathway because its substrate, cinnamic acid, may modulate the activity of PAL (Boudet, 1998).

The next step in the phenylpropanoid pathway is the hydroxylation of *p*-coumaric acid at the 3-carbon to form caffeic acid (Figure 1) catalyzed by C3H. The hydroxylation of ferulic acid at the 5-carbon to form 5- hydroxyferulic acid is catalyzed by F5H, an enzyme which has been reported in microsomal fractions from poplars (*Populus* sp.) where a cytochrome P450-dependent mixed function oxygenase (requiring oxygen and NADPH) was identified. Comparison of F5H and C4H from poplars suggests that the activity of these enzymes is dependent on two distinct cytochrome P450 systems (Schuler, 1996). In the SUCEST database we initially found a total of 46 cluster consensus sequences related to F5H (<http://sucest.lad.dcc.unicamp.br/private/mining-reports/RG/F5H.htm>), with significant *e*-values of between $9e-89$ and 0.0. However, 35 from these cluster consensus sequences were more related to cytochrome P450. Of the other 11 cluster consensus sequences, five were related to unnamed proteins, one to cinnamate 4-hydroxylase, two to flavonoid 3- hydroxylase, one to aldehyde 5-hydroxylase and two to F5H from *Arabidopsis thaliana* (FAH1). Of the 46 clusters consensus sequences, only three contained a complete coding region, and only SCJLFL4100D06.g proved to be closely related to F5H from *Oryza sativa* (AC073867) (Figure 2). Sequences SCCST3002C11.g and SCQSRT1036E09.g appear to be only distantly related even to the F5H coding region from poplars (AJ010324) and *Arabidopsis* (AF068574), suggesting that these sequences belong to the cytochrome P450 family but are not F5H genes.

The next step in the phenylpropanoid metabolic pathway is the methylation performed by COMT, which controls the production of ferulic acid in Gymnosperms and ferns, or both ferulic acid and sinapic acid in Angiosperms (Figure 1). COMT catalyses the *o*-methylation of caffeic acid and hydroxyferulic acid. Several COMTs have been purified from plants including poplars, spinach, soybean, *Thuja orientalis*, tobacco and alfalfa, two forms having been detected in poplars and three forms in alfalfa and tobacco (Barber and Mitchell, 1997). In the SUCEST database, 31 different cluster consensus sequences related to COMT were identified (<http://sucest.lad.dcc.unicamp.br/private/mining-reports/RG/COMT.htm>), with *e*-values between $9e-13$ and 0.0. Only five of these corresponded to ESTs containing complete coding sequences (SCCAD1004B03.g, SCCCLR1066B08.g, SCCCRT

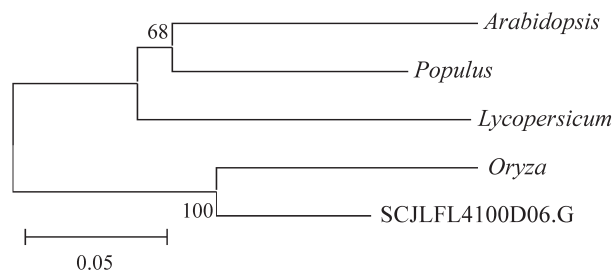


Figure 2 - Unrooted dendrogram of ferulate-5-hydroxylase (F5H) sequences. The relationships were calculated using MEGA (p distance, neighbor-joining method and the bootstrap test with 1000 replications, complete deletions). *Arabidopsis thaliana* (AF068574); *Populus balsamifera* subsp. *trichocarpa* (AJ010324); *Lycopersicum esculentum* x *L. peruvianum* (AF150881) and *Oryza sativa* (AC073867). The sugarcane sequence uses SUCEST nomenclature.

1003A12.g, SCJFRT1058D02.g and SCVPRT2073 D10.g), and were compared to other COMT sequences. As can be seen in Figure 3, these five ESTs are more related to COMT from *Saccharum officinarum* Jaronu (GenBank AJ231133 and CAA13175, 362 amino acids long). Their predicted protein sequences differs by only one amino acid, suggesting that these sequences could represent different allelic forms or recently duplicated genes, but due to the polyploid nature of the sugarcane genome it is difficult to differentiate between the two forms.

The results obtained by Varner and colleagues support the hypothesis of an alternative pathway for the methylation of caffeic acid, the candidate enzyme being CCoAOMT which catalyses the methylation of caffeoyl-CoA into feruloyl-CoA (Ye *et al.*, 1994). It has been suggested that CCoAOMT is specific for methylation of caffeoyl-CoA esters in both tobacco (Atanassova *et al.*, 1995) and poplar (Van Doorselaere *et al.*, 1995). In the SUCEST data bank 30 cluster consensus sequences related to CCoAOMT were found (I), showing *e*-values between $9e-54$ and 0.0 (<http://sucest.lad.dcc.unicamp.br/private/mining-reports/RG/CCoAOMT.htm>). Of these clusters, 17 were related to CCoAOMT from *Zea mays*, 11 cluster consensus sequences related to EST AU030740 and 2 related to EST C98431 from *O. sativa*. Of all these sequences, only six contained complete coding sequences for CCoAOMT (I), comparison of these six cluster consensus sequences with CCoAOMT from other species showed that the sugarcane ESTs were clearly split into two groups (Figure 4). One group included clusters SCCCLR1069B09.g, SCEQRT2092E02.g and SCVPRT2073H7.g (257 amino acids each) which could be allelic forms or recently duplicated genes, since all these clusters produced the same predicted protein. The second group of cluster consensus sequences included SCCCRT1001C01.g (247 amino acids), SCVPRT2080G04.g (247 amino acids) and SCCST1004G05.g (248 amino acids). These cluster consensus sequences showed about 55% homology to the first group but had a large number of substitutions, insertions and deletions between them, strongly suggesting that they

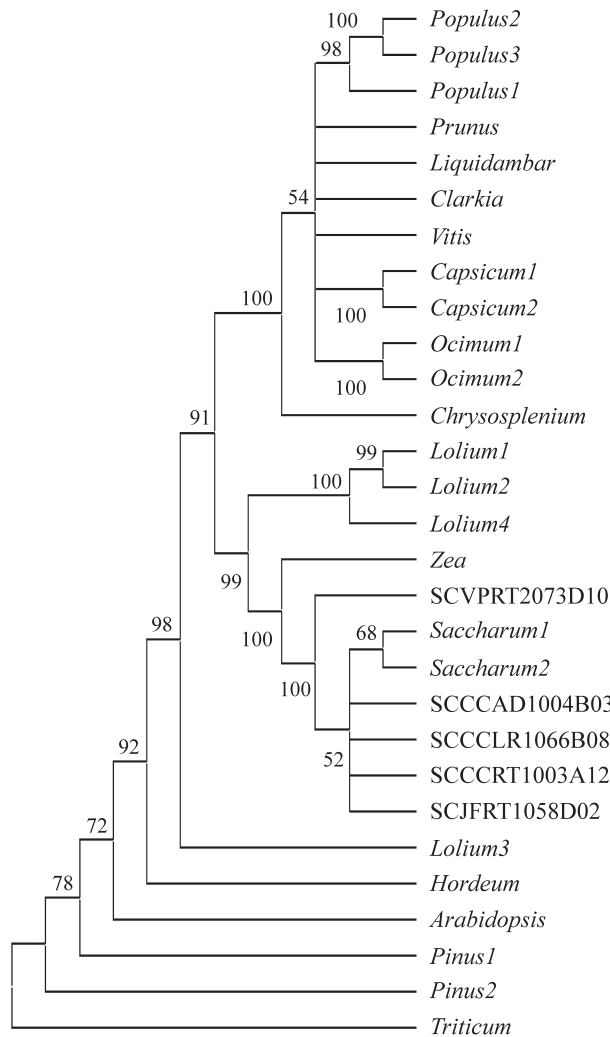


Figure 3 - Unrooted dendrogram of caffeic acid O-methyltransferase (COMT) sequences. Bootstrap consensus tree was constructed using MEGA (p distance, neighbor-joining method and bootstrap test with 1000 replications, complete deletions). *Lolium perenne* (Lolium 1 = AF010291, Lolium 2 = AF033538, Lolium 3 = AF033539, Lolium 4 = AF033540); *Hordeum vulgare* (U54767); *Saccharum officinarum* (Saccharum 1 = AJ231133, Saccharum 2 = CAA13175); *Populus kitakamiensis* (*Populus sieboldii* X *Populus grandidentata*) (Populus 1 = D49710, Populus 2 = D49711); *Populus tremuloides* (Populus 3 = U50522); *Prunus amygdalus* (X83217); *Zea mays* (Q06509); *Arabidopsis thaliana* (CAB64217); *Triticum aestivum* (AAD10485); *Clarkia breweri* (AF006009); *Chrysosplenium americanum* (U16793); *Ocimum basilicum* (Ocimum 1 = AF154917, Ocimum 2 = AF154918); *Capsicum chinense* (Capsicum 1 = AF081214); *Capsicum annuum* (Capsicum 2 = AF212316); *Vitis vinifera* (AF239740); *Liquidambar styraciflua* (AF139533); *Pinnus radiata* (Pinnus 1 = AF119225); *Pinnus pinaster* (Pinnus 2 = AJ297297). Sugarcane sequences are represented using SUCEST nomenclature.

are duplicated genes. Interestingly, both groups have counterparts in the rice and maize genomes, which suggests an ancestral duplication event in monocotyledons (Figure 4).

The last step of the phenylpropanoid pathway is the formation of hydroxycinnamic acid CoA esters catalyzed by 4CL (Figure 1). Several 4CLs have been purified from forsythia, spruce, loblolly pine, and maize and this enzyme

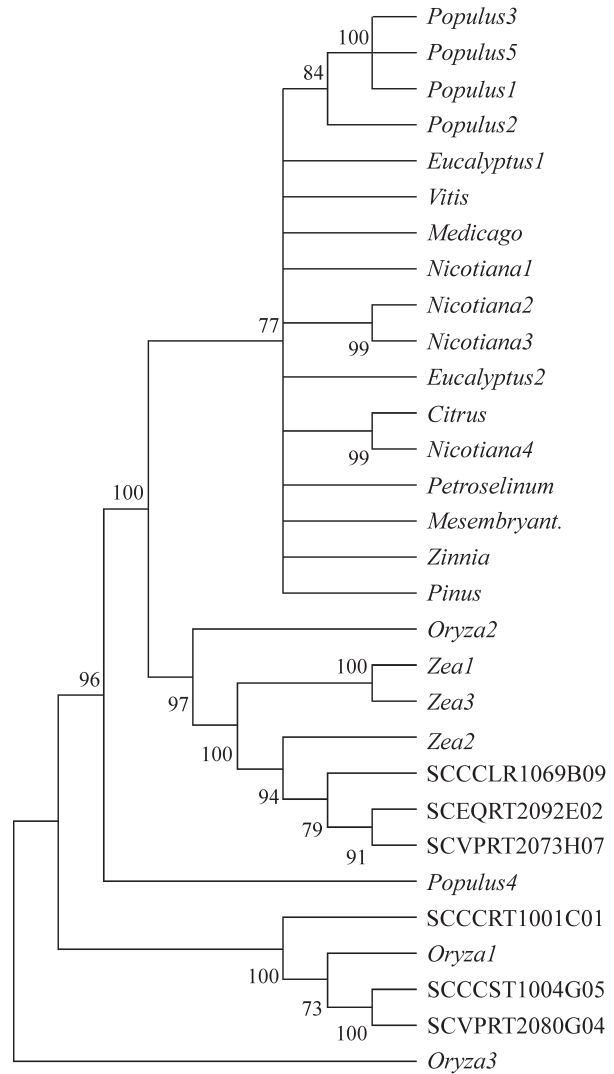


Figure 4 - Unrooted dendrogram of Caffeoyl CoA - methyltransferase (CCoAOMT) sequences. Bootstrap consensus tree was constructed using the Molecular Evolutionary Genetics Analysis (MEGA) software (p distance, neighbor-joining method and bootstrap test with 1000 replications, complete deletions). *Populus balsamifera* subsp. *trichocarpa* (Populus 1 = AJ224894, Populus 2 = AJ224895, Populus 3 = CAA12198); *Populus kitakamiensis* (Populus 4 = AB000408); *Populus tremuloides* (Populus 5 = Q43095); *Eucalyptus globulus* (Eucalyptus 1 = AF046122, Eucalyptus 2 = AF168780); *Vitis vinifera* (Z54233); *Nicotiana tabacum* (Nicotiana 1 = Z56282, Nicotiana 2 = U62734, Nicotiana 3 = U62736, Nicotiana 4 = Z82982); *Medicago sativa* (U20736); *Citrus natsudaoidai* (AB035144); *Petroselinum crispum* (A22706); *Mesembryanthemum crystallinum* (AF053553); *Pinus taeda* (AF036095); *Zinnia elegans* (U13151); *Oryza sativa* (Oryza 1 = BAA81774, Oryza 2 = BAA78733, Oryza 3 = BAA81776); *Zea mays* (Zea 1 = AJ242981, Zea 2 = AJ242980, Zea 3 = CAB45150). Sugarcane sequences are represented using SUCEST nomenclature.

also occurs in multiple forms in several plants such as soybean, pea, oat, poplar, parsley and potato (Barber and Mitchell, 1997). We were unable to find complete coding regions for 4CL genes in the SUCEST database. From a total of 30 incomplete cluster consensus sequences (Table I and <http://sucest.lad.dcc.unicamp.br/private/mining-reports/RG/4CL.htm>), we found two major partial sequences

Table I - Sugarcane expressed sequence tag (EST) from the SUCEST database related to enzymes involved in lignin biosynthesis pathways.

Enzyme ^a	Consensus sequence clusters	Consensus sequence clusters with complete open reading frames
F5H	46	03
4CL	30	-
COMT	31	06
CCoAOMT	30	06
CCR	33	01
CAD	27	03

^aF5H = ferulate-5-hydroxylase hydroxycinnamate, 4CL = hydroxycinnamate CoA ligase, COMT = caffeic acid O-methyltransferase, CCoAOMT = caffeoyl CoA O-methyltransferase, CCR = cinnamoyl-CoA reductase, CAD = cinnamyl alcohol dehydrogenase.

with no overlapping, SCMCRT2102F02.g (309 amino acids long) contains the N-terminal region and SCVPFL1138B04.g (253 amino acids long) the C-terminal region, both highly related to the *O. sativa* 4CL coding region (GenBank L43362). A third cluster (SCVPAM2068F01.g) represents a central segment of a 4CL coding region (222 amino acids) but interestingly did not seem to be closely related to the other sequences. These sequences can be used as a tool to help to isolate the complete sugarcane 4CL coding regions.

When hydroxycinnamic acid CoA esters are formed they are then reduced by enzymes of the lignin branch pathway (Figure 1) which contains two enzymes: CCR and CAD (Barber and Mitchell, 1997).

In the lignin branch pathway, CCR is the first enzyme and is responsible for the conversion of hydroxycinnamic acid CoA esters to their corresponding hydroxycinnamaldehydes (Figure 1). It has been suggested that CCR has a regulatory role in lignin biosynthesis, because it is able to divert general phenylpropanoid products to the accumulation of monolignols. CCR has been identified in several plants such as soybean, eucalyptus, spruce and poplar (Barceló, 1997; Barber and Mitchell, 1997). In the SUCEST data bank 33 CCR-related cluster consensus sequences were found after searching with bait sequences (<http://sucest.lad.dcc.unicamp.br/private/mining-reports/RG/CCR.htm>) (Table I), the cluster consensus sequences found presenting e-values of between 9e-50 and 0.0. However, only cluster SCMCRT2107D05.g (372 amino acids long) showed the complete coding region for the CCR protein. It seems that this sequence may be an allelic form or a recently duplicated gene related to sugarcane sequences CAA13176 and AJ231134 (Figure 5), since both are 372 amino acids long with only five amino acid substitutions.

The next enzyme in the lignin branch pathway is CAD, which is responsible for the formation of monolignols (Figure 1). CAD is a polymorphic enzyme in Angiosperms but is apparently encoded by a single gene in Gymnosperms. CAD reduces hydroxycinnamaldehydes to

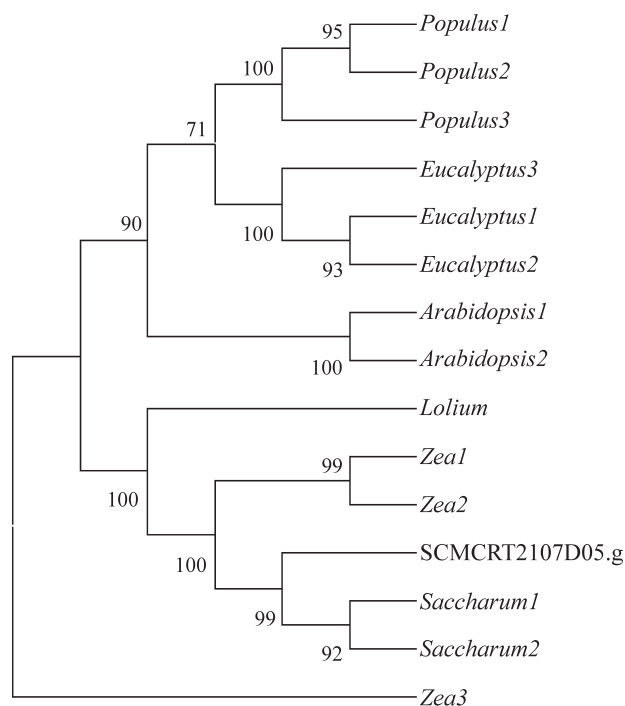


Figure 5 - Unrooted dendrogram of Cinnamoyl-CoA reductase (CCR) sequences. The relationships were calculated using Molecular Evolutionary Genetics Analysis (MEGA) software (p distance, neighbor-joining method and bootstrap test with 1000 replications, complete deletions). *Populus balsamifera* subsp. *trichocarpa* (Populus 1 = AJ224986, Populus 2 = AJ295838); *Populus tremuloides* (Populus 3 = AF217958); *Eucalyptus saligna* (Eucalyptus 1 = AF297877); *Eucalyptus gunnii* (Eucalyptus 2 = X97433, Eucalyptus 3 = X79566); *Arabidopsis thaliana* (Arabidopsis 1 = AF320623, Arabidopsis 2 = AF320624); *Saccharum officinarum* (Saccharum 1 = AJ231134, Saccharum 2 = CAA13176); *Lolium perenne* (AF278698); *Zea mays* (Zea 1 = X98083, Zea 2 = Y13734, Zea 3 = Y15069). Sugarcane sequences are represented using SUCEST nomenclature.

their corresponding hydroxycinnamyl alcohol and has been identified in soybean, wheat, eucalyptus, tobacco and bean (Barceló, 1997; Barber and Mitchell, 1997). Table I shows the 27 CAD-related cluster consensus sequences which we found in the SUCEST data bank, their e-values were between 9e-56 and e-106 (<http://sucest.lad.dcc.unicamp.br/private/mining-reports/RG/CAD.htm>) and only 3 clusters contained complete coding regions. Since clusters SCEQLR1029E05.g and SCJFRZ2025C03.g had only three substitutions between them in 359 amino acids they could be allelic variants. Cluster SCCCLB1001F10.g was 371 amino acids long and seems to be another CAD gene. Interestingly, these clusters are only slightly different in length and sequence from a previously identified CADs from *Zea mays* and *Saccharum officinarum* (Figure 6) and it seems that these SUCEST clusters could represent new CAD isoforms related to dicotyledons. This strongly suggest that searches for more CAD genes must be made.

Numerous genes have recently been identified by random nucleotide sequencing of cDNA clones in higher plants. Such methods have proven to be valuable for the rapid screening of putative genes from cDNA libraries.

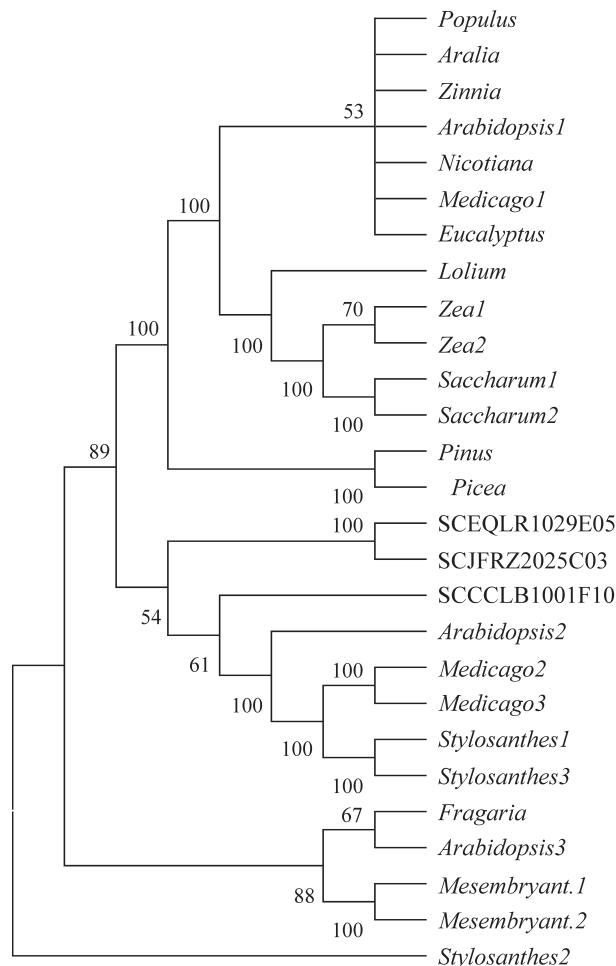


Figure 6 - Unrooted dendrogram of Cinnamyl alcohol dehydrogenase (CAD) sequences. Bootstrap consensus tree was constructed using MEGA (p distance, neighbor-joining method and bootstrap test with 1000 replications, complete deletions). *Populus tremuloides* (AF217957); *Eucalyptus saligna* (AF294793); *Aralia cordata* (D13991); *Nicotiana tabacum* (X62343); *Medicago sativa* (Medicago 1 = AF083332, Medicago 2 = AF083333, Medicago 3 = AAC35846); *Zinnia elegans* (D86590); *Arabidopsis thaliana* (Arabidopsis 1 = Z31715, Arabidopsis 2 = L37883, Arabidopsis 3 = T05625); *Lolium perenne* (AF101290); *Saccharum officinarum* (Saccharum 1 = AJ231135, Saccharum 2 = CAA13177); *Zea mays* (Zea 1 = AJ005702, Zea 2 = Y13733); *Pinus radiata* (AF060491); *Picea abies* (AJ001925); *Stylosanthes humilis* (Stylosanthes 1 = L36823, Stylosanthes 2 = L36456, Stylosanthes 3 = Q43137); *Fragaria x ananassa* (U63534) and *Mesembryanthemum crystallinum* 1 (Mesembryant. 1 = U79770), Mesembryant. 2 = T12571). Sugarcane sequences are represented using SUCEST nomenclature.

Genes identified as ESTs can be useful for functional genomic analysis, the production of restriction fragment length polymorphism (RFLP) maps and the isolation of full-length cDNA clones. The identification of sugarcane EST sequences coding for the key enzymes of lignin metabolism constitute an initial effort to study this metabolic pathway in this plant species. The further characterization of these genes will help in clarifying lignin metabolism in monocotyledonous species as well as in the development of initiatives in the manipulation of lignin content and composition in order to improve the use of sugarcane biomass.

RESUMO

Ligninas são compostos fenólicos encontrados nas paredes secundárias do sistema vascular vegetal e desempenham importantes papéis biológicos, reduzindo a permeabilidade da parede celular em relação à água, estando também envolvidas em mecanismos de defesa contra patógenos. A via metabólica da lignina e as enzimas envolvidas na sua síntese vem sendo caracterizadas nos últimos anos. Uma série de genes que codificam diferentes enzimas envolvidas na biossíntese da lignina foram identificados em diferentes espécies de plantas. A biossíntese de lignina está acoplada ao metabolismo dos fenilpropanóides, apresentando enzimas compartilhadas com outros processos metabólicos tais como fenilalanina amônio liase (PAL), cinnamato 4-hidroxilase (C4H) and ácido caféico O-metiltransferase (COMT) assim como enzimas específicas como cinamoil-CoA redutase (CCR) e cinamil álcool dehidrogenase (CAD). Em certos tipos de mutantes de milho e sorgo, um aumento na digestibilidade foi associado a uma redução no teor de lignina. Uma redução na atividade de CAD e COMT, importantes enzimas envolvidas na biossíntese de lignina vem sendo demonstrada. Estas observações tem motivado diferentes grupos de pesquisa a alterarem o conteúdo ou a composição da lignina em plantas modelo, assim como culturas de interesse econômico, através de engenharia genética. O principal objetivo prático destes projetos é uma otimização da utilização destas plantas levando a uma maior produção de papel e digestibilidade da ração animal. No presente trabalho foi realizado um inventário das seqüências de ESTs, que codificam enzimas envolvidas no metabolismo de lignina, presentes no Projeto Genoma de Cana-de-açúcar (SUCEST). A análise foi realizada com as enzimas chaves ferulato-5-hidroxilase (F5H), ácido caféico O-metiltransferase (COMT), cafeoil CoA O-metiltransferase (CCoAOMT), hidroxycinamato CoA ligase (4CL), cinamoil-CoA redutase (CCR) and cinamil álcool dehidrogenase (CAD). A análise comparativa destes genes com os descritos em outras espécies poderá servir para a identificação de marcadores moleculares em programas de melhoramento genético, assim como em projetos que visem a manipulação do metabolismo de lignina em cana-de-açúcar.

ACKNOWLEDGMENTS

R.L.B. Ramos is in receipt of a doctoral fellowship from the Brazilian agency CNPq. R.M. Junqueira and F.B. Lino were supported by CNPq and a CNPq-PIBIC fellowship. The authors thank Adriana Fusaro for critical reading of the manuscript and the Brazilian agency FAPESP for its support in the developed of the SUCEST project. The Young Scientists Program ('Cientista Jovem do Estado') of the Brazilian agency FAPERJ supported this work.

REFERENCES

- Atanassova, R., Favet, N., Martz, F., Chabbert, B., Toiller, M.T., Monties, B., Fritig, B. and Legrand, M.** (1995). Altered lignin composition in transgenic tobacco expressing *O*-methyltransferase sequences in sense and antisense orientation. *Plant J.* 8: 465-477.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J.** (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25: 3389-3402.
- Barber, M.S. and Mitchell, H.J.** (1997). Regulation of phenylpropanoid metabolism in relation to lignin biosynthesis in plants. *International Review of Cytology* 172: 243-293.
- Barceló, A.R.** (1997). Lignification in plant cell walls. *International Review of Cytology* 176: 87-132.
- Boudet, A. M.** (1998). A new view of lignification. *Trends in Plant Science* 3: 67-71.
- Herrmann, K.M.** (1995). The shikimate pathway: Early steps in the biosynthesis of aromatic compounds. *Plant Cell* 7: 907-919.
- Kumar, S., Tamura, K. and Nei, M.** (1994). MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Computer Appl. Biosci.* 10: 189-191.
- Lintu, R.** (1990). UNIDO international seminar of newsprint from bagasse, UNIDO, Havana Cuba.
- Schuler, M.A.** (1996). Plant cytochrome P450 mono-oxygenases. *Crit. Rev. Plant Sci.* 15: 235-284.
- Thompson, J.D., Higgins, D.G. and Gibson, T.J.** (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673-4680.
- Van Doorselaere, J., Baucher, M., Chognot, E., Chabbert, B., Tollier, M.T., Petit-Conil, M., Leplé, J.C., Pilate, G., Cornu, D., Monties, B., Van Montagu, M., Inzé, D., Boerjan, W. and Jouanin, L.** (1995). A novel lignin in poplar trees with reduced caffeic acid/5 hydroxyferulic acid *O*-methyltransferase activity. *Plant J.* 8: 855-864.
- Whetten, R.W., Mackay, J.J. and Sederoff, R.R.** (1998). Recent advances in understanding lignin biosynthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49: 585-609
- Ye, Z.H., Kneusel, R.E., Matern, U. and Varner, J.E.** (1994). An alternative methylation pathway in lignin biosynthesis in *Zinnia*. *Plant Cell* 6: 1427-1439.