ELSEVIER

Plant Science 143 (1999) 163–171

PLANT SCIENCE

Molecular cloning of cDNAs coding for three sugarcane enzymes involved in lignification☆

Guillermo Selman-Housein [a], María A. López [a], Doramys Hernández [a], Laura Civardi [b], Frank Miranda [a], Joan Rigau [b], Pere Puigdomènech [b,*]

[a] *Plant Metabolism Engineering Laboratory, Plant Division, Centre for Genetic Engineering and Biotechnology, P.O. Box 6162, 10600 La Habana, Cuba*

[b] *Departament de Genètica Molecular, IBMB-CID-CSIC, 08034 Barcelona, Spain*

Received 2 November 1998; received in revised form 22 February 1999; accepted 15 March 1999

Abstract

Full-length cDNAs encoding the monolignol biosynthetic enzymes caffeic acid 3-*O*-methyltransferase (COMT, EC 2.1.1.6), cinnamoyl-CoA reductase (CCR, EC 1.2.1.44) and cinnamyl alcohol dehydrogenase (CAD, EC 1.1.1.195) were cloned from sugarcane (*Saccharum officinarum* L.). The encoded proteins (39.6, 40.1 and 38.7 kDa for COMT, CCR and CAD, respectively) were identified based on their sequence identities with the corresponding enzymes from other plant species. Pairwise comparisons of deduced amino acid sequences with known plant lignification proteins allowed the identification of important conserved domains and specific functional motifs within these enzymes. Two new conserved domains, probably involved in substrate specificity, are described for COMTs. Phylogenetic analysis showed a very close evolutionary relationship between sugarcane and maize sequences. Southern blot analyses are consistent with the presence of at least two copies of each studied gene in sugarcane genome. The *comt*, *ccr* and *cad* transcripts appear to happen in a parallel way in different sugarcane tissues. The mRNA accumulation patterns suggest a transcriptional regulation of these genes dependent on their specific role in lignin synthesis. The cloning and characterisation of sugarcane genes involved in lignification opens up the possibility of producing plants with lower and/or modified lignin by genetic engineering means.

Keywords: CAD; CCR; Cinnamoyl CoA reductase; Cinnamyl alcohol dehydrogenase; COMT; *O*-Methyltransferase; *Saccharum*

1. Introduction

Sugarcane (*Saccharum officinarum* L.) is a member of the Andropogoneae tribe of the Poaceae (grass) family. The *Saccharum* genus is complex and is characterised by high polyploidy and frequent aneuploidy [1,2]. Modern sugarcane cultivars have chromosome numbers between 100 and 130, and they are derived essentially from interspecific hybridisations involving different *Saccharum* species [3]. Sugarcane is widely spread and economically very important, in many regions around the world, as a source of sugar and many by-products. The bagasse is the main residue of the sugarcane industry representing, by weight, almost 30% (hundreds of millions of tons per year worldwide) of the sugarcane agricultural product so the bagasse utilisation is important for both economical and environmental considerations. Sugarcane bagasse is a good low cost raw material for paper production or animal feed but, as for other plants, its cell wall structural polymer, lignin, rich in *p*-coumaryl subunits, has a negative effect on digestibility [4] and paper pulping properties [5].

Lignins are synthesised by the dehydrogenative polymerisation of monolignols. The synthesis of

☆ The nucleotide sequence data of sugarcane COMT, CCR and CAD cDNAs are registered at the EMBL Nucleotide Sequence Database with the accession numbers: AJ231133, AJ231134 and AJ231135, respectively.

* Corresponding author. Tel.: +34-3-4006100; fax: +34-3-2045904.

E-mail address: pprgmp@cid.csic.es (P. Puigdomènech)

0168-9452/99/$ - see front matter
PII: S0168-9452(99)00041-2

monolignols is bound to the general metabolism of the phenylpropanoids in plants, having enzymes common with other processes (phenylalanine ammonia-lyase (PAL), caffeic acid 3-*O*-methyltransferase (COMT) and 4-coumarate:CoA ligase (4CL), for instance), as well as specific enzymes (cinnamoyl-CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD)) [6].

Nowadays, great interest has emerged in the possibility of modifying the content and/or composition of the lignin polymer for improving the industrial behaviour of many economically important crops. Using genetic engineering approaches the main attempts have been concentrated on down-regulating the levels of enzymes involved in the lignification process by means of sense or antisense expression of homologous or heterologous genes in transgenic plants (for review see Ref. [7]). Efforts have been mainly focused on the modification of expression levels of COMT (EC 2.1.1.6) [8–12], CAD (EC 1.1.1.195) [13–15] and, more recently, CCR (EC 1.2.1.44) [16] genes.

Within a project directed to improve, by genetic engineering means, the quality of sugarcane fibre for its use as raw material for paper and forage production, we report here the cloning, molecular characterisation and phylogenetic relationships of the COMT, CCR and CAD cDNAs from sugarcane. The obtained cDNA fragments will also serve as probes in molecular marker-assisted breeding programmes for the production of new sugarcane varieties as sources of renewable fuel, paper pulp or forage.

2. Methods

2.1. RNA isolation, cDNA library construction and screening

Total RNA was extracted following standard published methods [17] from the elongation zone of young roots collected from sugarcane variety Jaronu 60-5 internodes germinated in pots under greenhouse conditions. mRNAs were isolated using the PolyATract Purification System (Promega) and cDNA synthesis and cloning in Lambda Uni-Zap XR Vector were performed using Stratagene kits. The following DNA fragments were labelled with [α-^{32}P]dATP, using a random primer kit from Boehringer Mannheim, and were used as probes: the 1.2-kb *Xho*I–*Xba*I cDNA fragment from plasmid pMC1 (a maize COMT clone) [18]; the 960-bp *Sty*I–*Pst*I fragment from a poplar CCR cDNA (POPCCR 2.1) [19]; and the 910-bp *Eco*RV–*Hin*dIII fragment from poplar CAD cDNA (PGEMPOPCAD1) [20]; both poplar cDNA clones were kindly supplied by Dr Wout Boerjan from the University of Gent, Belgium. The cDNA library was screened with each of the probes, blotting half a million recombinant phages onto Hybond-N membrane filters (Amersham), following the manufacturers instructions with modifications: membranes were pre-hybridised for 4 h and hybridised at 60°C for 18 h in 5 × SSPE, 5 × Denhardt's, 0.5% SDS and 100 μg/ml of sonicated salmon sperm DNA; washes were performed twice in 2 × SSC, 0.1% SDS and 0.5 × SSC, 0.1% SDS for 15 min at 60°C and finally 15 min at 42°C in 0.1 × SSC, 0.1% SDS.

2.2. Sequencing and phylogenetic analyses

Nucleotide sequencing of two isolated full-length clones for each cloned cDNA was performed automatically, in both directions, using an ABI model 377 sequencer. Sequence analysis and alignments were done using GCG [21] and Clustal W [22] software packages. The phylogenetic trees were drawn using the WET software version 1.31 (J. Dopazo, TDI, Madrid, Spain. http://www.tdi.es/), where the dendograms were obtained by the neighbour-joining method [23] using the Poisson correction for determining the amino acid distance and removing from the alignments all sites with deletions.

2.3. Southern and Northern analyses

Genomic DNA from sugarcane culms was isolated following a published procedure [24]. The DNA was digested with restriction enzymes and separated on a 0.8% agarose TAE gel [25]. A total of 20 μg of digested DNA or 10μg of total RNA per lane were fractionated on agarose gels and transferred to a Hybond-N nylon membrane (Amersham). Southern and Northern blots were performed following the manufacturers instructions and washed under high stringent conditions (0.1 × SSC, 0.1% SDS, 65°C, 15 min). The probes were the full-length COMT, CCR and CAD cDNAs from sugarcane. For Northern blot analyses,

membranes were not re-used; this procedure was completely repeated for each gene we studied.

3. Results

3.1. Cloning and nucleotide sequence of sugarcane COMT cDNA

The sugarcane full-length COMT cDNA consists of a 1486-bp fragment with the translational start site of the major open reading frame (ORF) at nucleotide 105 and the TAA stop site at nucleotide 1191. As in the case of the maize clone [18], this ORF does not have a typical plant consensus sequence around the translational start, having a G at positions −3 and +5. A possible polyadenylation signal, AATAAC, is located at positions 1444–1449. The G + C content of the coding region is typical for a graminaceous monocot, 66.85% [26], and it codes for a 362-amino acid polypeptide with a calculated molecular mass of 39.6 kDa and predicted isoelectric point of 5.12. Comparison of this protein with maize COMT [18] gives an identity of 91.16% over the entire amino acid sequence.

3.2. Comparison of sugarcane COMT with other methyltransferases: identification of conserved motifs and evolutionary relationships

Alignment of the sugarcane COMT with other methyltransferases shows that it has the three regions previously identified as conserved in enzymes requiring *S*-adenosyl-L-methionine, as a substrate (positions 203–219, 266–279 and 285–296), as well as the ATP binding site motif GXGXXG between amino acid residues 340 and 345 [18,27,28]. In addition, two other regions conserved in all COMT sequences are present in the sugarcane enzyme (positions 226–235 and 318–329). Recently, Joshi and Chiang [29] described a group of features that allow the classification of plant *O*-methyltransferases in two superfamilies. Based on its sequence signatures, the sugarcane COMT belongs to the plant OMT-II superfamily (Table 1). Computer comparison between COMT and other methylases from different organisms (data not shown) reveals a region of low similarity (positions 120–176 in the sugarcane sequence) which is extremely well conserved among COMTs (Table 2). In this region we identified two sequence motifs (S1 and S2) highly conserved in all enzymes able to catalyse the methylation of hydroxycinnamic acids. The consensus sequence for S1 motif is GVS(V/M/I/L)(A/S)(P/A)(L/I)XLMN(Q/H)(D/G) and it is always located 13 amino acids before the S2 motif, which has the consensus (V/I)L(D/E)GG(I/V)PFNKAYGM. This region could have a biological significance, perhaps containing those residues responsible for the specificity of these enzymes to different hydroxycinnamic substrates. In Table 3, we show some plant *O*-methyltransferases non-classified as COMT that have regions of high sequence homology to these putative substrate-binding motifs. It could be interesting to test whether these enzymes can also methylate hydroxycinnamic acids.

We calculated the percentages of identity at the amino acid level after a pairwise comparison between COMTs from different plant taxa (data not shown). The high level of identity of sugarcane

Table 1
The seven characteristic motifs of plant OMT-II *S*-adenosyl-L-methionine-dependent methyltransferases and the distances between them are well conserved in the sugarcane COMT[a]

	Consensus	Positions and sequences in sugarcane COMT	Identity (%)
Motif A	LVDVGGGXG	203-LVDVGGGIG-211	100
Spacing I	52	52	–
Motif B	VPXXDAXXMKW	255-VPAGDAILMKW-265	100
Spacing II	30	30	–
Motif C	ALPXXGKVIXXEXILP	285-ALPENGKVIIVEC**V**LP-294	90.9
Motif I	LDRXLRLL	76-**V**DRMLRLL-84	85.7
Motif J	IKGINFDLPHVI	224-IKGINFDLPHVI-235	100
Motif K	PGVEHVGGDMF	242-PGV**Q**HVGGDMF-252	90.9
Motif L[b]	GGKERTXXEFXXLA	325-GG**R**ER**Y**EREFHDLA-338	80.0

[a] The non-conserved amino acid residue is given in bold; conserved amino acids are underlined.
[b] This motif has a mistake in Ref. [29].

Table 2
Alignment of sugarcane amino acid sequence to other COMTs[a]

Gene	Alignment of sugarcane COMT (120–176 aa region)to other COMTs
COMT-Sug	120-GVSMAALTLMNQDKVLMESWYYLKDAVLDGGIPFNKAYGMTAFEYHGTDPRFNRVFN-176
COMT-Zea	120-GVSMAALALMNQDKVLMESWYYLKDAVLDGGIPFNKAYGMTAFEYHGTDARFNRVFN-176
COMT-Pop2	124-GVSVSPLCLMNQDKVLMESWYYLKDAILDGGIPFNKAYGMTAFEYHGTDPRFNKVFN-180
COMT-Euc	125-GVSIAALNLMNQDKILMESWYYLKDAVLEGGIPFNKAYGMTAFEYHGTDPRFNKIFN-181
COMT-Alf	124-GVSISALNLMNQDKVLMESWYHLKDAVLDGGIPFNKAYGMTAFEYHGTDPRFNKVFN-180
COMT-Sty	124-GVSLSALNLMNHDKVLMESWYYLKETVLEGGIPFNKAYGMTAFEYHGTDPRFNKVFN-180
COMT-Chr	102-GVSIAALCLMNQDKVLMESWYHLKDAVLDGGIPFNKAYGMSSFEYHGTDPRFNKVFN-158
COMT-Zin	106-GVSLAPLLLMNQDKVLMESWYYLKDPVLDGGIPFNKAYGMSAFEYHGKDQRFNKVFN-162
COMT-Pru	124-GVSIAPLCLMNQDKVLLESWYHLKDAVLEGGIPFNKAYGMTAFEYHGTDPRFNKVFN-180
COMT-Tob	124-GVSVAPLLLMNQDKVLMESWYHLKDAVLDGGIPFNKAYGMTAFEYHGTDPRFNKVFN-180
COMT-Asp1	124-GVSVSPLCLMNQGKVLMESWYYLKDAILDGGIPFNKAYGMTAFEYHGTDPRFNKVFN-180
COMT-Asp2	124-GVSVSPLCLMNQDKVLMESWYYLKDAILDGGIPFNKAYGMTAFEYHGTDPRFNKVFN-180
COMT-Pop1	124-GVSVSPLCLMNQDKVLMESWYYLKDAILEGGIPFNKAYGMTAFEYHGTDPRFNKVFN-180
COMT-Pop3	124-GVSVSPLCLMNQDKVLMESWYHLKDAILEGGIPFNKAYGMTAFEYHGTDPRFNKVFN-180
COMT-Lol2	121-GVSMAALALMNQDKVLMESWYYLKDAVLDGGIPFNKAYGMSAFEYHGTDPRFNRVFN-177
COMT-Lol1	121-GVSMAALALMNQDKVLMESWYYLKDAVLDGGIPFNKAYGMSAFEYHGTDPRFNRVFN-177
COMT-Lol3	121-GVSMAALALMNQDKVLMESWYYLKDAVLDGGIPFNKAYGMSAFEYHGTDPRFNRVFN-177
COMT-Cap	120-GVSIAPILLMNQDKVLMESWYHLTDAVLDGGVPFNKAYGMTTFEYHGTDPRFNKVFN-176
COMT-Cla	126-GVSLAPLCLMNQDKVLMESWYYLKDAILDGGIPFNKAYGMTAFEYHGTDPRFNKVFN-182
Homology	***.... ***..*.*.**** *..*.**.********..***** * ***..**
Consensus	GVSVAPLXLMNQD------13-----VLDGGIPFNKAYGM

[a] Alignment of sugarcane amino acid sequence (positions 120–176) to other COMTs shows the S1 and S2 conserved motifs. COMT-Sug, sugarcane (AJ231133); COMT-Zea, maize (M73235); COMT-Pop1, poplar (D49710); COMT-Pop2, poplar (M73431); COMT-Pop3, poplar (D49711); COMT-Euc, *Eucalyptus* (X74814); COMT-Alf, *Alfalfa* (M63853); COMT-Sty, *Stylosanthes* (L36109); COMT-Chr, *Chrysoplenium* (U16793); COMT-Zin, *Zinnia* (U19911); COMT-Pru, almond (X83217); COMT-Tob, tobacco (A26487); COMT-Asp1, aspen (U13171); COMT-Asp2, aspen (X62096); COMT-Lol1, *Lolium* (AF033538); COMT-Lol2, *Lolium* (AF010291); COMT-Lol3, *Lolium* (AF033540); COMT-Cap, *Capsicum* (AF081214); COMT-Cla, *Clarkia* (O23760). The consensus line is the proposed for S1 and S2 motifs.

and maize proteins (91.2%) points to a close evolutionary relationship between these two plant species belonging to the family Poaceae, whereas their comparison with dicot angiosperms ranged from 59.3 to 65.5%. We established a phylogenetic tree using the percentages of identity found between the different COMT proteins, where a clear evolutionary distinction among plants belonging to different taxonomic classes can be seen (Fig. 1A).

3.3. Cloning and nucleotide sequence of sugarcane CCR cDNA

The full-length cDNA of the sugarcane *ccr* gene is 1511 bp long with the translational start of the major ORF at nucleotide 172 and the TAA stop at nucleotide 1288. This ORF also has a high G + C content of 69.7% and it codes for a polypeptide of 372 amino acids with an estimated molecular mass of 40.1 kDa and p*I* of 5.67. The translational start of the sugarcane CCR cDNA matches the plant consensus sequence, having an A at position −3 and C at position +5. The putative polyA+ signal also matches the consensus AATAAA sequence for eukaryote organisms, 31 bp before the polyA tail (positions 1451–1456). The amino acid identity between poplar [19] and sugarcane CCR proteins is 74.3%.

3.4. Comparison of sugarcane CCR with other oxidoreductases: identification of important motifs and evolutionary relationships

As for the maize CCR (acc. no. Y13734) [30], the amino and carboxy terminal regions of the sugarcane protein are larger than for the eucalyptus enzyme and between amino acid residues 30 and 59 there is the characteristic secondary structure corresponding to the βαβ-dinucleotide binding fold of NADP(H) and NAD(H)-dependent reductases and dehydrogenases [31]. In this region the residues V33, T34, G35, G38, and L48, conserved among CCRs and putatively involved in cofactor binding, together with the residue S146 are present. The KNWYCYGK motif, a distinctive feature of all known CCR proteins, is present at positions 177–184 in the sugarcane sequence but, as in the maize CCR, the first amino acid, also positively charged, is arginine. Lacombe et al. based on enzyme-inhibition studies [31], sug-

gested that in this conserved box at least one of the lysine residues might be involved in substrate binding. Therefore, the change of the first amino acid of this motif in maize and sugarcane enzymes could indicate a different substrate specificity and/or that this first amino acid is not directly involved in substrate binding.

Amino acid sequence comparison among six angiosperm CCRs (three dicots and three monocots including sugarcane) is shown in the resolved phylogenetic tree in Fig. 1B. The pairwise comparisons (data not shown) within the monocot angiosperms are characterised by higher levels of identity (84–90%) than between dicot plants (79–82%), being lower the amino acids identities between dicots and monocots (67–75%). These data point to a close evolutionary relationship among the monocot angiosperms, as can be seen in the tree, where branch lengths for dicot plants suggest an increased divergence when compared with monocot angiosperms.

3.5. Cloning and nucleotide sequence of sugarcane CAD cDNA

The full-length cDNA of the sugarcane *cad* gene is 1593 bp long with a major ORF of 1098 bp (start site at position 184 and TGA stop site at nucleotide 1279). This cDNA has a G + C content of 66.3% and codes for a polypeptide of 365 amino acids with a calculated molecular mass of 38.7 kDa and theoretical isoelectric point of 6.06. There is the plant consensus sequence around the initiation codon having A at position −3 and G at +5. A possible poly(A)+ signal (AATAAC) is 108 bp before the polyA tail at positions 1462–1467. Alignment of poplar [20] and sugarcane CAD proteins reveals an identity of 76.6%.

3.6. Comparison of sugarcane CAD with other alcohol dehydrogenases: identification of important motifs and evolutionary relationships

Comparison of the sugarcane CAD with other alcohol dehydrogenases shows that the sugarcane enzyme has all the typical features of this protein superfamily [20,32–35]: the Zn-1 catalytical GHEVVGXVXE(V/I)GXXV (positions 68–82) and the Zn-2 structural GDXVGXGXXVG(C/S)C(R/K)XCXXCXXXXEQYC (positions 88–114) domains, as well as the NADP co-factor binding region GXGXXG (positions 188–193). The residues I-96, Y-113, W-119 and F-299, identified by McKie et al. [36] by computer molecular modelling as possibly being responsible for substrate specificity of aromatic alcohols in the eucalyptus CAD, are also conserved in the sugarcane protein. Carboxy-terminal regions of the sugarcane and maize (acc. no. Y13733) [30] CADs are larger than for other CAD proteins and do not contain an SKL sequence for microbody targeting [35]. Therefore, a different targeting signal for subcellular localisation of sugarcane CAD should be present in this protein if the in vivo functionality of this signal is finally demonstrated for CAD proteins.

Table 3
Occurrence and conservation of motifs S1 and S2 in different plant methyltransferases[a]

Gene	Enzyme	Alignment of S1 and S2 motif regions	H-S1	H-S2
MT-Tob1	Catechol *O*-MT	121-GVSVAPLLLMNQD-13-VLDGGIPFNKAYGM-160	100	100
MT-Tob2	Catechol *O*-MT	121-GVSVAPLLLMNQD-13-VLDGGIPFNKAYGM-160	100	100
MT-Ara	O-MT	119-GVSIAALCLMNQD-13-ILDGGIPFNKAYGM-158	100	100
MT-Cap	*O*-Diphenol *O*-MT	117-GVSVAPLLLMNQD-13-VLDGGVPFNKAYGM-156	100	100
MT-Pin1[b]	*O*-MT	130-**E**LSMAPMLLMQND-13-VLEGGVAFQK**A**N**G**A-169	91.7	85.7
MT-Pin2	*O*-MT	131-**E**LSMAPMLLMQND-13-VLEGGVAFQK**A**N**G**A-170	91.7	85.7
MT-Mes	*myo*-Inositol *O*-MT	122-**QG**SLGPLLVLHHD-13-ILEGGVPF**K**RAHGM-161	83.3	92.9
MT-Chr	3′-Flavonoid *O*-MT	97-GVSIAALCV**AA**QD-13-VLDGGIPFNKAYGM-136	83.3	100
MT-Tob3	Catechol *O*-MT	123-G**A**SMGPLL**A**L**L**QD-13-VLEGGVPFDR**V**HGV-162	75.0	92.9
MT-Cla	(iso)Eugenol *O*-MT	124-GVSLAP**F**LL**TAT**D-13-ILEGGIPFNKAYGM-163	66.7	100

[a] Occurrence and conservation of motifs S1 and S2 in different plant methyltransferases (non-conserved amino acids are given in bold). H-S1, percent overall homology (identical+conserved aa) to consensus S1 motif; H-S2, percent overall homology to consensus S2 motif; *O*-MT, *O*-methyltransferase; –, unclassified enzyme; MT-Tob1, tobacco (X74452); MT-Tob2, tobacco (X74453); MT-Tob3, tobacco (X71430); MT-Ara, *Arabidopsis* (U70424); MT-Cap, *Capsicum* (U83789); MT-Pin1, *Pinus taeda* (U39301); MT-Pin2, *Pinus radiata* (U70873); MT-Mes, ice plant (M87340); MT-Chr, *Chrysoplenium* (U16794); MT-Cla, *Clarkia* (U85760).

[b] This enzyme also catalyzes the methylation of caffeic and 5-hydroxyferulic acids [40].

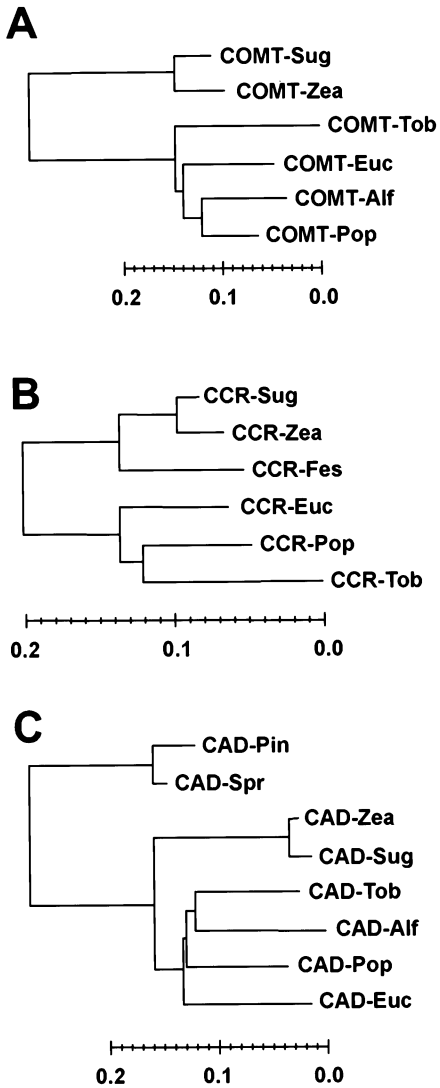

Fig. 1. Phylogenetic tree of sugarcane and other homologous sequences of COMTs (A), CCRs (B) and CADs (C). Sug, sugarcane; Zea, maize; Tob, tobacco; Euc, eucalyptus; Alf, alfalfa; Pop, populus; Fes, festuca; Pin, pinus; and Spr, spruce. The accession numbers of the sequences used in the comparisons are shown in brackets.

When we compared sugarcane to other plant CADs (data not shown), a remarkable homology with maize protein (96.9 and 95.5% identity at amino acid and nucleotide levels, respectively) was found. This high degree of homology clearly indicates the close evolutionary relationship between these two species. The identities at amino acid level between monocot and dicot plants ranged from 72 to 77%, while both types of plants had 65–69% identity with the gymnosperms pine and spruce, respectively. Within the dicot angiosperms the level of amino acid identity was 77–81%. In the phylogenetic tree based on the pairwise comparisons between different CADs (Fig. 1C) there is a distinction between each one of the three phylogenetic groups we analysed, clearly reflecting their degree of genetic divergence. These results support those previously obtained for gymnosperm and dicot angiosperm CADs [20,35].

3.7. Genomic complexity and mRNA accumulation corresponding to the cloned sugarcane cDNAs

For Southern blot analyses, genomic DNA of sugarcane commercial variety Jaronu 60-5, the source of the cDNA library, was digested with enzymes that have either one recognition site within the cloned COMT (*Bam*HI) and CCR (*Hin*dIII) cDNAs or no sites (*Eco*RI). In view of the complexity of the sugarcane genome (a high degree of ploidy and chromosomic mosaicism) [1,2], we expected complex Southern blot hybridisation patterns; however, only a small number of bands hybridised to the used probes (Fig. 2). Southern blot analysis with COMT probe gives one (*Hin*dIII and *Bam*HI) or two (*Eco*RI) strong hybridisation signals and some additional bands of variable intensities (Fig. 2A). The hybridisation pattern for the CAD probe shows also few strong signals (three bands for *Eco*RI and *Bam*HI digestions, and one for *Hin*dIII) and some faint bands (Fig. 2C). On this basis, we suspect that at least two copies of *comt* and *cad* genes are present in this sugarcane variety. The hybridisation pattern for the CCR probe is slightly more complex (Fig. 2B) and it could be attributed, as it is for maize CCR [37], to the presence of more than one coding gene.

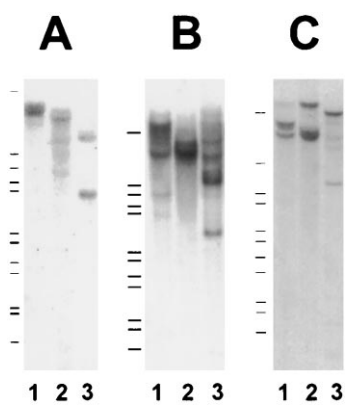

Fig. 2. Genomic Southern blot analysis of sugarcane commercial variety Jaronu 60-5 with the cloned full-length COMT (A), CCR (B) and CAD (C) cDNAs. Lanes 1, 2 and 3, total DNA digested with *Eco*RI, *Bam*HI and *Hin*dIII, respectively. Molecular weight marker is λ-phage DNA *Eco*RI–*Hin*dIII digested.

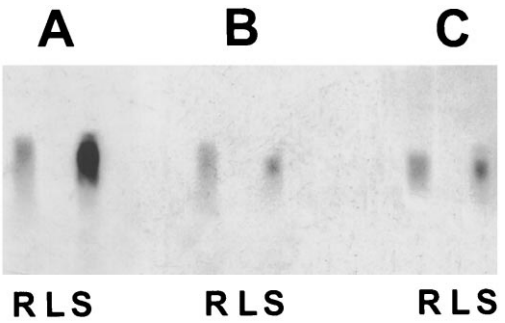

Fig. 3. Northern blot analysis of total RNAs from different tissues of sugarcane plants (C, culms; L, leaves; and R, roots). Blots were probed with the cloned full-length COMT (A), CCR (B) and CAD (C) cDNAs.

Northern blot analyses reveal single transcripts of size approximately the same as the cloned cDNAs (Fig. 3). The expression of these lignification genes is high in young tissues that are developing vascular structures [38,39]; however, in contrast to the situation in eucalyptus [31] and maize [18,32], it was not possible to detect any signal in organs such as leaves, having a low level of lignification. We found a correlation between the signal intensities in Northern blots and the number of putative positive clones (38 for COMT, 22 for CCR and 15 for CAD) during the cDNA library screening for each of the probes we used. Considering the fact that Northern blot analyses were performed independently for each studied gene using the same total RNA stock and under identical experimental conditions, this finding points to a transcriptional regulation of the genes involved in lignification, the expression level of a gene being lower when its product is more specific to lignin biosynthesis. However, more experimental evidence is required to validate this hypothesis.

4. Discussion

Full-length cDNAs corresponding to three sugarcane genes that play a fundamental role in lignification have been cloned and molecularly characterised. In all three cases the percentage of amino acid sequence similarity between different plant species is very high, indicating that the degree of sequence conservation needed for the activity of the three proteins is very similar. The sequence features that have been proposed to be essential for these proteins in other species are also conserved in the sugarcane enzymes. Moreover, we identified two sequence motifs highly conserved in plant enzymes capable of performing SAM-dependent methylation of caffeic acid. Indeed, the S1 and S2 motifs, as well as their spatial arrangements, are well conserved in loblolly pine *O*-methyltransferase AEOMT (acc. no. U39301), a multifunctional enzyme with activity on caffeic and 5-hydroxyferulic acids and low overall homology to COMTs [40]. Based on this observation, it will be interesting to test the methylation activity on hydroxycinnamic acids for *O*-methyltransferases that show high level homology to the new motifs reported here.

Recently, Pichon et al. [37] reported the isolation of two CCRs (ZmCCR1 and ZmCCR2) from maize. Cloned sugarcane CCR cDNA is highly homologous to maize ZmCCR1 (acc. no. X98083), which is over 98% identical to Y13734 sequence. We did not find any recombinant clones related to ZmCCR2 during the cDNA library screening. However, due to reported very low expression level of the ZmCCR2 gene in roots and/or to its stress response nature [37], we cannot exclude the existence of a ZmCCR2 homologous gene in sugarcane genome.

From the patterns of mRNA accumulation, the genes corresponding to the three enzymes appear to be expressed in a similar way. It is not known whether the whole pathway is present in the same cells or in related subcellular compartments, but the results presented here point towards a coordinated expression of the genes coding for these enzymes and suggest a transcriptional regulation of lignification genes dependent on their commitments to lignin synthesis. Nevertheless, new quantitative experiments using internal controls of gene expression have to be done in order to corroborate this hypothesis.

The complexity of the genes coding for these three enzymes in the sugarcane genome appears to be limited. This is somewhat surprising if the complexity of the genome of this species is considered. The Southern blot for genes such as *comt* and *cad*, which in maize are single copy genes [18,32], appears to have at least two copies probably due to the high level of ploidy present in this species. This result indicates that the allelic variability between the homologous gene in the sugarcane genome is not very high. Possible explanations may be that the different genome copies have the same origin and they have not

diverged over the limited time during which they have formed the present species, or an efficient recombination machinery tends to homogenise the homologous sequences in a complex genome. Whatever the explanation, the low variability existing in this species indicates that approaches such as antisense technologies may be applied although different alleles of the gene exist in the genome.

Sugarcane is an important crop for many countries around the world, and a better understanding of its biology will undoubtedly benefit crop breeding through classical and/or molecular approaches. The close evolutionary relationship found between sugarcane and maize genes, for instance, opens up the possibility of increasing the use of many agriculture by-products by modulating the quality and quantity of lignin not only in these crops, but in other related species such as sorghum, rice, wheat and barley, which represent a great fraction of the plant biomass produced by man.

Acknowledgements

This work has been partially supported by the EEC INCO project No. ERB3514PL962073, and by Plan Nacional de Investigación Científica y Técnica (grand BIO97-0729). The authors also want to acknowledge Osmany Ramos, Ana Pons, Dra. Salome Prat and Dra. Virginia Stiefel for the support they gave to this work.

References

[1] T.V. Sreenivasan, B.S. Ahloowalia, D.J. Heinz, Cytogenetics, in: D.J. Heinz (Ed.), Sugarcane Improvement Through Breeding, Elsevier, Amsterdam, 1987, pp. 211–253.

[2] G. Bremer, Problems in breeding and cytology of sugar cane, Euphytica 10 (1961) 59–78.

[3] J. Daniels, B.T. Roach, Taxonomy and evolution in sugarcane, in: D.J. Heinz (Ed.), Sugarcane Improvement Through Breeding, Elsevier, Amsterdam, 1987, pp. 7–84.

[4] R.A. Leng, Application of Biotechnology to nutrition of animals in developing countries. FAO Technical Papers, Animal Production and Health. Rome, 1991.

[5] V.L. Chiang, R.J. Puumala, H. Takeuchi, R.E. Eckert, Comparison of softwood and hardwood kraft pulping, Tappi J. 71 (1988) 173–176.

[6] R. Whetten, R. Sederoff, Lignin biosynthesis, Plant Cell 7 (1995) 1001–1013.

[7] A.M. Boudet, J. Grima-Pettenati, Lignin genetic engineering, Mol. Breeding 2 (1996) 25–39.

[8] U.N. Dwivedi, W.H. Campbell, J. Yu, R.S.S. Datla, R.C. Bugos, V.L. Chiang, G.K. Podila, Modification of lignin biosynthesis in transgenic *Nicotiana* through expression of an antisense *O*-methyltransferase gene from *Populus*, Plant Mol. Biol. 26 (1994) 61–71.

[9] W. Ni, N.L. Paiva, R. Dixon, Reduced lignin in transgenic plants containing a caffeic acid *O*-methyltransferase antisense gene, Transgenic Res. 3 (1994) 120–126.

[10] R. Atanassova, N. Favet, F. Martz, B. Chabbert, M.-T. Tollier, B. Monties, B. Fritig, M. Legrand, Altered lignin composition in transgenic tobacco expressing *O*-methyltransferase sequences in sense and antisense orientation, Plant J. 8 (1995) 465–477.

[11] J. Van Doorsselaere, M. Baucher, E. Chognot, B. Chabbert, M.-T. Tollier, M. Petit-Conil, J.-C. Leple, G. Pilate, D. Cornu, B. Monties, et al., A novel lignin in poplar trees with a reduced caffeic acid/5-hydroxyferulic acid *O*-methyltransferase activity, Plant J. 8 (1995) 855–864.

[12] C.-J. Tsai, J.L. Popko, M.R. Mielke, W.-J. Hu, G.K. Podila, V.L. Chiang, Suppression of *O*-methyltransferase gene by homologous sense transgene in quaking aspen causes red-brown wood phenotypes, Plant Physiol. 117 (1998) 101–112.

[13] C. Halpin, M.E. Knight, G.A. Foxon, M.M. Campbell, A.M. Boudet, J.J. Boon, B. Chabbert, M.-T. Tollier, W. Schuch, Manipulation of lignin quality by down-regulation of cinnamyl alcohol dehydrogenase, Plant J. 6 (1994) 339–350.

[14] T. Higuchi, T. Ito, T. Umezawa, T. Hibino, D. Shibata, Red-brown color of lignified tissues of transgenic plants with antisense CAD gene: wine-red lignin from coniferyl aldehyde, J. Biotechnol. 37 (1994) 151–158.

[15] M. Baucher, B. Chabbert, G. Pilate, J. Van Doorsselaere, M.-T. Tollier, M. Petit-Conil, D. Cornu, D. Monties, M. Van Montagu, D. Inzé, et al., Red xylem and higher lignin extractability by down-regulating a cinnamyl alcohol dehydrogenase in poplar, Plant Physiol. 112 (1996) 1479–1490.

[16] J. Piquemal, C. Lapierre, K. Myton, A. O'Connell, W. Schuch, J. Grima-Pettenati, A.M. Boudet, Down-regulation of cinnamoyl CoA reductase induces significant changes of lignin profiles in transgenic tobacco plants, Plant J. 13 (1998) 71–83.

[17] S. de Vries, H. Hoge, T. Bisseling, Isolation of total and polysomal RNA from plant tissues, in: S.B. Gelvin, R.A. Schilperoort, D.P.S. Verma (Eds.), Plant Molecular Biology Manual, Kluwer, Dordrecht, 1993, pp. B6/1–B6/12.

[18] P. Collazo, L. Montoliu, P. Puigdoménech, J. Rigau, Structure and expression of the lignin *o*-methyltransferase gene from *Zea mays* L., Plant Mol. Biol. 20 (1992) 857–867.

[19] J.-C. Leplé, J. Grima-Pettenati, M.V. Montagu, W. Boerjan, A cDNA encoding cinnamoyl-CoA reductase from *Populus trichocarpa* (Accession No. AJ224986), Plant Physiol. 117 (1998) 1126.

[20] J. Van Doorsselaere, M. Baucher, C. Feuillet, A.M. Boudet, M. Van Montagu, D. Inzé, Isolation of cinnamyl alcohol dehydrogenase cDNAs from two important economic species: alfalfa and poplar. Demonstration of high homology of the gene within angiosperms, Plant Physiol. Biochem. 33 (1995) 105–109.

[21] J. Devereux, P. Haeberli, O. Smithies, A comprehensive set of sequence analysis programs for the VAX, Nucleic Acids Res. 12 (1984) 387–395.

[22] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.

[23] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.

[24] S.O. Rogers, A.J. Bendich, Extraction of DNA from plant tissues, in: S.B. Gelvin, R.A. Schilperoort, D.P.S. Verma (Eds.), Plant Molecular Biology Manual, Kluwer, Dordrecht, 1993, pp. A6/1–6/11.

[25] J. Sambrook, E.F. Fristch, T. Maniatis, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
[26] W.H. Campbell, G. Gowri, Codon usage in higher plants, green algae and cyanobacteria, Plant Physiol. 92 (1990) 1–11.
[27] R.C. Bugos, V.L. Chiang, W.H. Campbell, cDNA cloning, sequence analysis and seasonal expression of lignin-bispecific caffeic acid/5-hydroxyferulic acid *O*-methyltransferase of aspen, Plant Mol. Biol. 17 (1991) 1203–1215.
[28] O. Poeydomenge, A.M. Boudet, J. Grima-Pettenati, A cDNA encoding *S*-adenosyl methionine caffeic acid 3-*O*-methyltransferase from *Eucalyptus*, Plant Physiol. 105 (1994) 749–750.
[29] C.P. Joshi, V.L. Chiang, Conserved sequence motifs in plant *S*-adenosyl-L-methionine-dependent methyltransferases, Plant Mol. Biol. 37 (1998) 663–674.
[30] L. Civardi, A. Murigneux, P. Tatout, P. Puigdomenech, J. Rigau, Molecular cloning and characterization of two cDNAs encoding enzymes required for secondary cell wall biosynthesis in maize, in: F.L. Schiavo, R.L. Last, G. Morelli, N.V. Raikhel (Eds.), Cellular Integration of Signalling Pathways in Plant Development. NATO ASI Series, vol. H104, Springer, Berlin, Heidelberg, 1998, pp. 135–146.
[31] E. Lacombe, S. Hawkins, J. Van Doorsselaere, J. Piquemal, D. Goffner, O. Poeydomenge, A.M. Boudet, J. Grima-Pettenati, Cinnamoyl-CoA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: cloning, expression and phylogenetic relationships, Plant J. 11 (1997) 429–442.
[32] C. Halpin, K. Holt, J. Chojecki, D. Oliver, B. Chabbert, B. Monties, K. Edwards, A. Barakate, G.A. Foxon, Brown-midrib maize (bm1)—a mutation affecting the cinnamyl alcohol dehydrogenase gene, Plant J. 14 (1998) 545–553.
[33] H. Jörnvall, B. Persson, J. Jeffery, Characteristics of alcohol/polyol dehydrogenases. The zinc-containing long-chain alcohol dehydrogenase, Eur. J. Biochem. 167 (1987) 195–201.
[34] J. Grima-Pettenati, C. Feuillet, D. Goffner, G. Borderies, A.M. Boudet, Molecular cloning and expression of a *Eucalyptus gunnii* cDNA clone encoding cinnamyl alcohol dehydrogenase, Plant Mol. Biol. 21 (1993) 1085–1095.
[35] J.J. MacKay, W. Liu, R. Whetten, R.R. Sederoff, D.M. O'Malley, Genetic analysis of cinnamyl alcohol dehydrogenase in loblolly pine: single gene inheritance, molecular characterization and evolution, Mol. Gen. Genet. 247 (1995) 537–545.
[36] J.H. McKie, R. Jaouhari, K.T. Douglas, D. Goffner, J. Grima-Pettenati, A.M. Boudet, M. Baltas, L. Corrichon, A molecular model for cinnamyl alcohol dehydrogenase, a plant aromatic alcohol dehydrogenase involved in lignification, Biochim. Biophys. Acta 1202 (1993) 61–69.
[37] M. Pichon, I. Courbou, M. Beckert, A-M. Boudet, J. Grima-Pettenati, Cloning and characterization of two maize cDNAs encoding cinnamoyl-CoA reductase (CCR) and differential expression of the corresponding genes, Plant. Mol. Biol. 38 (1998) 671–676.
[38] C. Feuillet, V. Lauvergeat, C. Deswarte, G. Pilate, A.M. Boudet, J. Grima-Pettenati, Tissue- and cell-specific expression of a cinnamyl alcohol dehydrogenase promoter in transgenic poplar plants, Plant Mol. Biol. 27 (1995) 651–657.
[39] S. Hawkins, J. Samaj, V. Lauvergeat, A.M. Boudet, J. Grima-Pettenati, Cinnamyl alcohol dehydrogenase: Identification of new sites of promoter activity in transgenic poplar, Plant Physiol. 113 (1997) 321–325.
[40] L. Li, J.L. Popko, X.-H. Zhang, K. Osakabe, C.-J. Tsai, C.P. Joshi, V.L. Chiang, A novel multifunctional *O*-methyltransferase implicated in a dual methylation pathway associated with lignin biosynthesis in loblolly pine, Proc. Natl. Acad. Sci. USA 94 (1997) 5461–5466.