PHYSIOLOGIA PLANTARUM 120: 631–641. 2004
Printed in Denmark – all rights reserved

Cloning and characterization of cellulose synthase-like gene, *PtrCSLD2* from developing xylem of aspen trees

Anita Samuga and Chandrashekhar P. Joshi*

Plant Biotechnology Research Center, School of Forest Resources and Environmental Science, Michigan Technological University, Houghton, MI 49931, USA
**Corresponding author, e-mail: cpjoshi@mtu.edu*

Received 5 March 2003; revised 4 September 2003

Genetic improvement of cell wall polymer synthesis in forest trees is one of the major goals of forest biotechnology that could possibly impact their end product utilization. Identification of genes involved in cell wall polymer biogenesis is essential for achieving this goal. Among various candidate cell wall-related genes, cellulose synthase-like D (*CSLD*) genes are intriguing due to their hitherto unknown functions in cell wall polymer synthesis but strong structural similarity with cellulose synthases (CesAs) involved in cellulose deposition. Little is known about CSLD genes from trees. In the present article *PtrCSLD2*, a first CSLD gene from an economically important tree, aspen (*Populus tremuloides*) is reported. *PtrCSLD2* cDNA was isolated from an aspen xylem cDNA library and encodes a protein that shares 90% similarity with *Arabidopsis* AtCSLD3 protein involved in root hair tip growth. It is possible that xylem fibers that also grow by intrusive tip growth may need expression of *PtrCSLD2* for controlling the length of xylem fibers, a wood quality trait of great economical importance. PtrCSLD2 protein has a N-terminal cysteine-rich putative zinc-binding domain; eight transmembrane domains; alternating conserved and hypervariable domains; and a processive glycosyltransferases signature, D, D, D, QXXRW; all similar to aspen CesA proteins. However, PtrCSLD2 shares only 43–48% overall identity with the known aspen CesAs suggesting its distinct functional role in cell wall polymer synthesis perhaps other than cellulose biosynthesis. Based on Southern analysis, the aspen CSLD gene family consists of at least three genes and this gene copy estimate is supported by phylogenetic analysis of available CSLDs from plants. Moreover, gene expression studies using RT-PCR and in situ mRNA hybridization showed that *PtrCSLD2* is expressed at a low level in all aspen tissues examined with a slightly higher expression level in secondary cell wall-enriched aspen xylem as compared to primary cell wall enriched tissues. Together, these observations suggest that *PtrCSLD2* gene may be involved in the synthesis of matrix polysaccharides that are dominant in secondary cell walls of poplar xylem. Future molecular genetic analyses will clarify the functional significance of CSLD genes in the development of woody trees.

Introduction

Secondary cell walls of tree xylem provide an abundant source of renewable and valuable raw materials for a variety of forest products. Biotechnological improvement in quantity and quality of cell wall components of trees therefore holds an enormous promise for forest product industries. We are interested in understanding the biosynthesis of major cell wall polymers in trees in order to genetically improve them for better wood utilization. We have discovered a number of tree genes involved in cellulose biosynthesis (e.g. Wu et al. 2000, Samuga and Joshi 2002, Joshi 2003a) and lignin biosynthesis (e.g. Li et al. 1997, Osakabe et al. 1999). Recently, we began searching for genes involved in hemicellulose or matrix polysaccharide biosynthesis in order to develop a global picture of cell wall polymer synthesis in trees.

Genes involved in polysaccharide synthesis have been notoriously difficult to isolate. The first plant cellulose synthase (*CesA*) gene from cotton was reported only in 1996 (Pear et al. 1996). Following that discovery,

Abbreviations – CesA, cellulose synthase; CSL, cellulose synthase-like; EST, expressed sequence tag; GFP-green fluorescent protein; pfu, plaque forming units; RT-PCR, reverse transcription-PCR; 5′RACE, Rapid amplification of 5′cDNA ends.

Cutler and Somerville (1997) used in-silico analysis for identification of a seven-member family of cellulose synthase-like (*CSL*) genes in *Arabidopsis*. Systematic analysis of the recently sequenced *Arabidopsis* genome suggested that it hosts at least six distinct *CSL* gene families, *CSLA, CSLB, CSLC, CSLD, CSLE,* and *CSLG* sharing only limited identity with *CesA* genes (Richmond and Somerville 2000). It is also suggested that unlike *CesA* genes, *CSL*s might be involved in biosynthesis of matrix polysaccharides such as xylans, xyloglucans, galactans and mannans. However, this view has yet to be experimentally supported even in *Arabidopsis* (Bonetta et al. 2002). All other plant species examined so far also have homologues of *Arabidopsis CSL* genes. Although we still do not clearly understand the functions of these *CSL* genes in polysaccharide synthesis, it is possible that each *CSL* gene family may perform a completely different function based on their extensive amino acid sequence diversity among them. CSL-GFP fusions have been localized at the internal membrane systems such as Golgi complexes and endoplasmic reticulum (ER) where hemicellulose synthesis occurs (Richmond and Somerville 2000, Favery et al. 2001). A systematic molecular genetic approach is therefore essential for deciphering functions of each of the *CSL* genes in polysaccharide synthesis (Richmond and Somerville 2001).

Recently, the first such insight became possible through detailed molecular genetic characterization of an *Arabidopsis CSLD* gene family member, *AtCSLD3* (Favery et al. 2001, Wang et al. 2001). In *Arabidopsis,* six distinct *CSLD* genes (*AtCSLD-1* to *AtCSLD-6*) were identified (Richmond and Somerville 2000). Favery et al. (2001) described the molecular genetic analysis of the *kojak* (*kjk*) mutant in *Arabidopsis* that lacks intact root hair tips resulting in a bald or hairless root phenotype. In this mutant, root hairs are initiated normally but they rupture at the tips soon after initiation, most probably due to defective cross-linking of cell wall polymers at the root hair tip. Usual chromosomal walk resulted in identification of the defective gene to be *AtCSLD3* and complementation of *kjk* mutants with wild-type *AtCSLD3* restored the normal hairy root phenotype. The majority of the AtCSLD3-GFP4 fusion proteins were also localized at the ER suggesting their involvement in the synthesis of non-cellulosic wall polysaccharides. However, experimental data on non-cellulosic polysaccharide alteration in the *kjk* mutant lacking the ability to produce intact AtCSLD3 protein are still unavailable. It was recently suggested that this lack of information might be due to low-level expression of these genes in *Arabidopsis* tissues (Bonetta et al. 2002). Albeit useful for molecular genetic analysis, *Arabidopsis* plants are typically small. Combined with a low level of CSLD expression, alternative plants such as robust trees may be explored to dissect the roles of *CSLD* genes in cell wall polysaccharide synthesis. As poplar xylem is enriched in hemicelluloses (Mellerowicz et al. 2001) and large quantities of such developing xylem tissues can easily be isolated, it would be an ideal experimental material to examine if CSLDs are indeed involved in hemicellulose synthesis. Isolation of *CSLD* genes from poplar xylem tissues is the first step towards examining such a possibility. With this objective, in the present study we cloned and characterized the first ortholog of AtCSLD3 gene from aspen xylem cDNA library, *PtrCSLD2*. Higher expression of *PtrCSLD2* gene in secondary cell wall-enriched aspen xylem compared with aspen tissues enriched in primary cell walls suggests that this gene is a likely candidate gene involved in hemicellulose biosynthesis in xylem secondary walls.

Materials and methods

Plant material

Various fresh tissue samples from quaking aspen (*Populus tremuloides*) trees were collected from the greenhouse of the Plant Biotechnology Research Center, School of Forest Resources and Environmental Science, Michigan Technological University. The plant materials were immediately frozen in liquid nitrogen and stored at −80°C till use.

Screening of aspen xylem cDNA library

We first screened about 100 000 pfu of the aspen xylem cDNA library described earlier by Samuga and Joshi (2002) with *Arabidopsis* CSLD EST H10C11T7. We obtained a partial cDNA highly similar to AtCSLD3, but the encoded protein potentially lacked the first approximately 200 amino acids. A 400-bp PstI fragment from the 5′ end of this partial cDNA was again used to screen 100 000 pfu of the same aspen xylem library and a much longer cDNA, L13ii was obtained, which was renamed as *PtrCSLD2*. DNA/protein and phylogenetic analyses were performed as described earlier (Joshi 2003a). The *PtrCSLD2* clone still lacked the DNA sequence corresponding to approximately 40 amino acids at the N-terminus.

Obtaining information about the missing N-terminal sequence of *PtrCSLD2* by genomic and 5′RACE PCR

Two highly similar genomic DNA fragments (XXI18668. ×1 and XXI589290.b1) corresponding to 5′-end of *PtrCSLD2* clone were identified from the raw poplar genome sequence data available at http://genome.jgi-psf.org/poplar0/poplar0.info.html (Wullschleger et al. 2002). These fragments contained the sequences of the possible missing 5′ end of the *PtrCSLD2* sequence, but the poplar genome is being sequenced from *Populus trichocarpa* which is different from *Populus tremuloides* used here for isolation of *PtrCSLD2* cDNA. In order to validate the presence of such genes in the aspen genome, two sets of primers were designed based on two poplar genomic fragments, respectively, to amplify the missing N-terminal region of *PtrCSLD2* from aspen.

Set I

GCSLD1 5′ TGCAGGATCTGAAGAGAAGTC 3′

GCSLD3 5′ CAACCTTTTGTGAGATTGATGG 3

Set II

GCSLD4 5′ TGCACAGTACTATAATTGCAGG 3′

GCSLD6 5′ AAATCACTACTCCCAAGTTCAC 3′

Aspen genomic DNA (20 ng per reaction^{-1}) was used as a template for PCR with 35 cycles of denaturation at 94°C for 2 min, re-annealing at 58°C for 1 min, extension at 72°C for 3 min with a final extension at 72°C for 7 min. Amplified products were cloned into pCR 2.1 TOPO vector (pCR 2.1 TOPO Cloning kit, Invitrogen, Carlsbad, CA, USA) and sequenced with ABI Prism 310 Sequencer. The 5′RACE PCR using xylem RNA as template amplified approximately 250 bp product only using the Set I primers but not with the set II primers.

Genomic DNA isolation and Southern analysis

Genomic DNA was isolated from aspen leaves and Southern blotting analysis was performed using the hypervariable (HVRI) region as well as the complete sequence of *PtrCSLD2* cDNA as described earlier (Samuga and Joshi 2002).

RT-PCR experiments

We performed RT-PCR using RNA from various aspen tissues including young shoot apices, 1–3 internodes, xylem, phloem and 8–10 internodes as described earlier by Samuga and Joshi (2002). We designed the HVRI-specific primer pair to study the expression pattern of *PtrCSLD2* gene as follows:

CSLD2ForwardHVRI 5′- GCAGTGAAAATTG-GAGGTGG - 3′

CSLD2ReverseHVRI 5′- TCCGAACAAAAAT-GAGAAGCC - 3′

Amplification reactions using 5.8S rRNA-specific primers (used as internal control) and PtrCSLD2-HVRI-specific primers for RT-PCR were performed as described earlier (Samuga and Joshi 2002).

In situ mRNA hybridization

PtrCSLD2 transcripts were detected by in situ hybridization with antisense transcripts from HVRI region of *PtrCSLD2* cDNA following the procedure described earlier (Wu et al. 2000). The HVRI region was amplified from the *PtrCSLD2* cDNA with HVRI-specific primers as described above and subcloned into pGEM-T Easy plasmid vector (pGEM-T Easy Vector System II; Promega, Madison, WI, USA). This was used to generate sense and antisense digoxygenin-labelled transcripts by in vitro transcription with T7 and SP6 RNA polymerases, respectively (DIG RNA Labeling Kit; Roche, Indianapolis, IN, USA) that were used as probes for in situ localization of *PtrCSLD2* mRNA.

Briefly, sections of various aspen tissues including young shoot apices, leaf, root and stems were fixed in 4% (w/v) phosphate buffered paraformaldehyde (PFA in 3× phosphate-buffered saline PBS, pH 7.2) at 4°C overnight, dehydrated through an ethanol series from 30 to 100%, gradually substituted with xylene and finally embedded in paraffin (Paraplast; Sigma, St. Louis, MO, USA). Ten micrometre sections were obtained on a Biocut 2030 Microtome (Global Medical Instrumentation, Albertville, MN, USA), mounted on Superfrost Plus slides (Fisher, Chicago, IL, USA) and incubated at 42°C overnight. Sections were subjected to de-waxing with xylene and rehydration through an ethanol series from 100 to 50%. Slides were then subjected to 70°C treatment for denaturing proteins followed by protein digestion with proteinase K (10 μg ml^{-1} in 1× PBS) for 1 h at room temperature. The sections were serially post-fixed with 4% PFA for 10 min, 3× PBS for 5 min, 1× PBS for 5 min, acetylated with 0.33% (v/v) acetic anhydride in 100 m*M* Triethanolamine (pH 8.1), and passed through an ascending ethanol series from 50 to 100%.

Hybridization was done with DIG-labelled and denatured (70°C for 5 min in Formamide) T7 and SP6 probes (approximately 0.6 ng/slide), respectively, in 2× hybridization mixture [10× SSC, 2% Blocking Reagent (Roche) with 0.3% Tween 20, 2.5% SDS, 0.1 mg ml^{-1} tRNA and 0.2 mg ml^{-1} heparin] at 50°C for 16 h in a humidity chamber (2× SSC and 50% formamide). Post-hybridization washes were done with 2× SSC and 50% formamide twice for 30 min each followed by maleic acid wash buffer wash (0.1 *M* maleic acid, 0.15 *M* NaCl, 0.3% Tween20, pH 7.5). Immunodetection of DIG probes with anti-DIG antibody (DIG Nucleic Acid Detection Kit; Roche) was performed at 4°C overnight. Post-antibody washes with maleic acid wash buffer and 1% blocking reagent for 3 h at room temperature was followed by colour detection with NBT/BCIP in colour development buffer (100 m*M* Tris pH 9.5, 150 m*M* NaCl, 50 m*M* $MgCl_2$) by incubation up to 6 h in dark (DIG Nucleic Acid Detection Kit; Roche). The colour development was stopped by incubation of slides in 10 m*M* EDTA and sections were mounted in glycerol. Slides were viewed with a Nikon Eclipse 400 microscope (Nikon Corporation, Tokyo, Japan) and images taken using a Sony DKC-5000 digital camera (Sony, Tokyo, Japan).

Results

Isolation of a CSLD cDNA from aspen developing xylem

To identify cDNAs encoding CSLD proteins from aspen, we first screened about 100 000 pfu of aspen xylem cDNA library with *Arabidopsis CSLD* EST (H10C11T7) probe and obtained 11 positives. End sequencing of a clone with the largest insert size, B15 showed that it contained a partial *CSLD* sequence and the predicted protein was potentially missing approximately 200 amino acids towards the N-terminus. Re-screening of same cDNA library with 400 bp 5′end *Pst*I fragment of B15 clone and end-sequencing of the eight rescued positive clones resulted in an almost full-length clone of 3570 bp, L13ii but still potentially missing approximately 40 amino acids at the N-terminus. Complete sequencing of L13ii indicated that we obtained an ortholog of the *AtCSLD3* gene from *Arabidopsis* (Favery et al. 2001). Earlier, we cloned aspen genomic fragments from three distinct CSLD genes that were designated as *PtrCSLD1*, *PtrCSLD2* and *PtrCSLD3* (GenBank accession no. AF417486, AF417487 and AY162183). Protein sequences of corresponding regions of PtrCSLD2 and L13ii are 100% identical suggesting that *PtrCSLD2* is part of L13ii. Therefore, the L13ii clone sequence was renamed as *PtrCSLD2* (GenBank accession no. AY162184).

Obtaining the missing N-terminal sequence of *PtrCSLD2*

Further attempts of re-screening the xylem cDNA library to obtain full-length cDNA as well as 5′RACE PCR experiments failed to yield additional information about the 5′ end of *PtrCSLD2* cDNA. This may be the result of low-level expression of this gene as indicated by the low number of positives obtained during cDNA library screening as compared with when we used CesA probes that yielded over 150 positives earlier (Samuga and Joshi 2002). We therefore used bioinformatics tools to predict the missing N-terminal amino acids using recently released, raw poplar genome data (http://genome.jgi-psf.org/poplar1/poplar1.home.html). Search of this database (2468 819 genomic DNA fragments) with the help of the BLASTN program (Altschul et al. 1997) and 3570 bp of *PtrCSLD2* with the *E*-value of $1\,E^{-20}$ resulted in a total of 70 hits. Seven of these fragments contained the DNA sequences highly similar to the available 5′-end of *PtrCSLD2* cDNA as well as sequences immediately upstream of the available *PtrCSLD2* cDNA. Prediction of encoded protein sequences from these seven sequences and their comparison with *PtrCSLD2* sequence yielded information about the possible missing first 39 amino acids in aspen *PtrCSLD2* cDNA. Moreover, among these seven hits, two very similar CSLD genes were present in the poplar genome sharing 87/90% identity/similarity in the missing 39 amino acids with each other (data not shown). However, in order to validate the existence of such genes in the aspen genome and also to see if data from *P. trichocarpa*, species used for sequencing the poplar genome can be directly applied to *P. tremuloides*, our species of interest, we carried out a PCR-amplification of both these genes using aspen genomic DNA with the two sets of primers that resulted in two approximately 250 bp products (*PtrCSLD2G1* and *PtrCSLD2G2*). Sequencing of these two PCR-products revealed that they were 98–99% identical to two corresponding poplar genome sequences we found earlier and thus resulted in the identification of two different CSLD sequences as shown in Fig. 1B. Thus the information from *P. trichocarpa* could directly be used for getting the sequence data about the missing 5′ end of *PtrCSLD2* from *P. tremuloides*; but which of these two aspen genomic sequences represent the real 5′ end of PtrCSLD2? In order to resolve this issue, we performed 5′RACE PCR using xylem RNA and both the sets of primers but the amplified PCR-product was produced only when we used set I primers corresponding to *PtrCSLDG1* and not when using set II primers corresponding to *PtrCSLDG2*. Therefore the *PtrCSLD2G1* sequence matching 5′RACE product was used along with the existing sequence data from *PtrCSLD2* to generate a full-length sequence as shown in Fig. 1A.

Sequence analysis of *PtrCSLD2* cDNA and its predicted protein

The *PtrCSLD2* cDNA as shown in Fig. 1A is 3723 bp long with a coding region of 3429 bp. This is the first report of a *CSLD* gene from any tree species. Comparison of *PtrCSLD2* cDNA with *Arabidopsis CSLD* gene sequences indicated the maximum DNA sequence identity of 76% with *AtCSLD3* and *AtCSLD2* genes that in turn shared 83% identity with each other, which suggests the possibility of *PtrCSLD2* being an ortholog of the *AtCSLD3*/*AtCSLD2*. Recently the *AtCSLD3* gene has been shown to be associated with root hair tip growth (Favery et al. 2001). Moreover, *PtrCSLD2* cDNA also shares 67% identity with *NaCSLD1* cDNA expressed in tobacco pollen tubes (Doblin et al. 2001).

The predicted protein sequence of *PtrCSLD2* is of 1143 amino acids with a calculated molecular mass of 128 182 Da and predicted pI of 7.1 (Fig. 1A). In our laboratory, we have so far isolated seven full-length CesA cDNAs from aspen and PtrCSLD2 shares only about 45% identity with the predicted proteins of these seven CesA cDNAs suggesting that in aspen, similar to in *Arabidopsis*, CSLD proteins are distinctly different to CesA proteins (Richmond and Somerville 2000). PtrCSLD2 shares 87% identity (90% similarity) with *Arabidopsis* AtCSLD3 protein but only 72% identity (80% similarity) with tobacco NaCSLD1 protein. Although in addition to PtrCSLD2, these are the only two full-length cDNAs encoding CSLD proteins that are currently available, the predicted protein sequences from the genomic sequences of an additional nine complete CSLD proteins from *Arabidopsis*, tobacco, alfalfa, rice are listed at http://cellwall.stanford.edu/. Comparison

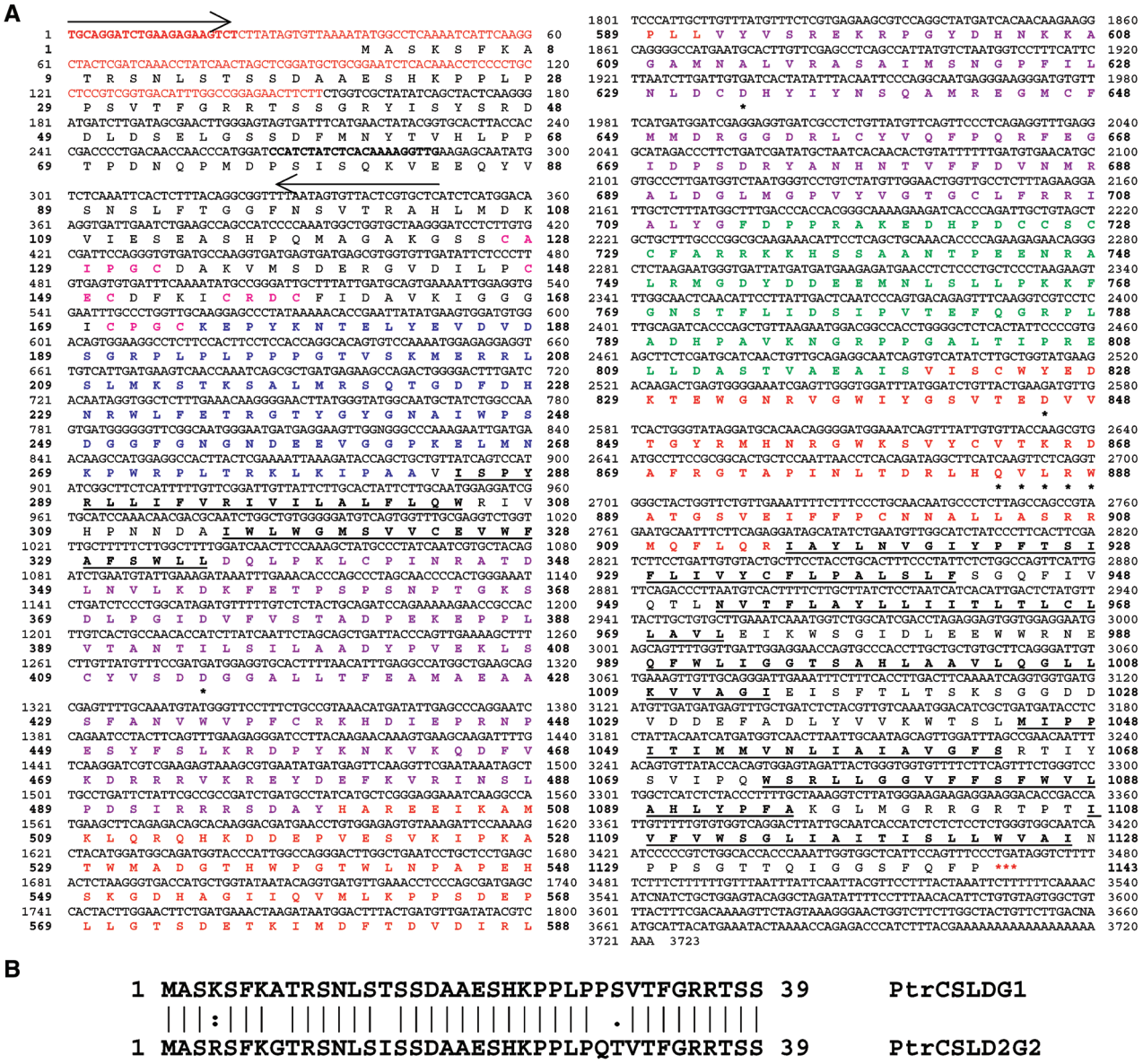

Fig. 1. (A) DNA sequence of aspen *PtrCSLD2* cDNA is shown at the top with the predicted single-letter translation below the DNA sequence indicated by bold numbers. The 5′ genomic DNA/RT-PCR sequence used for creating this hybrid molecule is shown in red colour and primers used for genomic fragment isolation are shown in bold and indicated by arrows. In the predicted protein sequence of PtrCSLD2, four cysteine-rich motifs of the N-terminal zinc-binding domain are shown in pink, HVRI in blue, subdomain A in violet, HVRII in green, HVRIII in orange and subdomain B in red, all indicated as bold letters. The two N-terminal and six C-terminal transmembrane domains are shown in bold, black and underlined letters. The processive glycosyltransferases signature D, D, D…QXXRW is indicated by black, bold asterisks below such residues and the stop codon is indicated by three bold, red stars. (B) Comparison of predicted amino acid sequences from two possible aspen isoforms of PtrCSLD2, PtrCSLD2G1 and PtrCSLD2G2 in the missing N-terminal region.

among 12 CSLD proteins shows that PtrCSLD2 shares the highest identity of 87% with AtCSLD2 and AtCSLD3 followed by rice OsCSLD2 and OSCSLD1 sequences sharing 84 and 80% identity, respectively (Table 1, upper half). Interestingly, remaining CSLD paralogs either from *Arabidopsis* or rice share much less identity with each other. This situation is similar to *CesA* genes in which orthologs are more similar to each other than paralogs (Holland et al. 2000) suggesting that the CSLD gene structural divergence preceded the species divergence.

Towards the N-terminal of PtrCSLD2 protein, there is a 47 amino acid-long, cysteine (Cys) rich region that is conserved in most of the CSLD proteins known to date with the exception of AtCSLD1 and AtCSLD6. Interestingly, similar 46 amino acid-long Cys-rich regions are also present in all known CesAs (Joshi 2003b). CSLDs are the only CSLs that possess such regions as suggested

Table 1. Comparison of 12 angiosperm CSLD proteins using GCG program 'GAP'. The percentages of amino acid sequence identities between the entire CSLD proteins are shown in the upper half and comparison of only HVRI regions are shown in the lower half separated by dashes. Identity percentages above 80% are shown in bold letters. Higher percentage identities between HVRI regions of PtrCSLD2 and four other CSLD proteins from *Arabidopsis* and rice are highlighted by underlining. The nomenclature used is the same as suggested at http://cellwall.stanford.edu.

	AtCSLD1	AtCSLD2	AtCSLD3	AtCSLD4	AtCSLD5	AtCSLD6	MtCSLD3	NaCSLD1	OsCSLD1	OsCSLD2	OsCSLD3	PtrCSLD2
AtCSLD1	–	69	67	70	64	62	62	70	66	66	68	70
AtCSLD2	51	–	**90**	73	67	66	66	72	**80**	**82**	70	**87**
AtCSLD3	48	**80**	–	73	67	65	67	71	**80**	**83**	69	**87**
AtCSLD4	44	49	49	–	64	63	65	**81**	71	73	76	74
AtCSLD5	40	54	51	47	–	63	**81**	63	66	66	63	67
AtCSLD6	25	26	0	17	26	–	62	63	64	67	61	66
MtCSLD3	37	51	48	40	76	19	–	64	65	67	63	68
NaCSLD1	45	50	50	67	41	20	41	–	70	72	76	72
OsCSLD1	50	60	59	48	49	43	48	52	–	**80**	68	**80**
OsCSLD2	47	61	59	45	49	33	44	51	60	–	71	**84**
OsCSLD3	49	51	53	58	54	24	49	58	56	52	–	70
PtrCSLD2	48	68	68	55	54	26	50	46	67	70	53	–

by Doblin et al. (2001). The consensus sequence of zinc-binding domain in CesA proteins is two tandem repeats of $CX_2CX_{10-15}CX_2C$ whereas CSLDs have two tandem repeats of $CX_{2-4}CX_{8-15}CX_{1-2}C$ (where X is any amino acid). Thus both types of Cys-rich domains have four pairs of Cys residues separated by only two amino acids in CesAs but by 1–4 amino acids in CSLDs (Fig. 1A).

PtrCSLD2 has eight transmembrane domains, two towards the N-terminus and six towards the C-terminus similar to all known CesA and CSLD proteins (Joshi 2003b) Similar to CesAs, CSLDs also have the highly diverged first hypervariable region (HVRI). Table 1 (lower half) shows the comparison of HVRI domains from 12 different CSLD proteins. Due to lack of zinc-binding domains in AtCSLD1 and AtCSLD6, the variability data for these two proteins is not reliable. PtrCSLD2 HVRI shares 67–70% identity with AtCSLD2, AtCSLD3, OsCSLD1 and OsCSLD2 but HVRI regions of all remaining CSLD proteins share only 26–55% identity with the PtrCSLD2 HVR I region. Thus this region from *PtrCSLD2* cDNA corresponding to HVRI domain has been used for gene copy number determination and gene expression analysis as described below.

A region similar to the putative catalytic domain of CesA proteins also exists in PtrCSLD2 and all other predicted CSLD proteins although no substrate binding assays have yet been performed for CSLD proteins similar to CesAs (Pear et al. 1996). Multiple sequence alignment of putative catalytic domain of 12 CSLDs indicated that they also have similar but shorter HVRII domains. However, A subdomain of CSLDs have at least one more additional highly variable, namely the HVRIII domain that is absent in CesA proteins (Fig. 1A). The D, D, D, QXXRW, a canonical sequence signature present in all processive glycosyltransferases (Saxena et al. 1995) is also present and conserved in all CSLDs suggesting that these proteins might also be involved in biogenesis of long-chain polysaccharides.

Gene copy number determination by Southern blot analysis

To estimate the size of *PtrCSLD2* gene family in the aspen genome, we performed Southern blot analysis using aspen genomic DNA and *PtrCSLD2* region corresponding to HVRI domain. Only two hybridizing bands were visible in each of the three lanes suggesting that *PtrCSLD2* belongs to a small gene family of about two genes (Fig. 2A) in agreement with our bioinformatics analysis presented earlier. It is also possible that two bands are originating from the same gene. None of the restriction enzymes used are present in *PtrCSLD2* cDNA. There is also no intron expected to be present in the HVRI region if gene structure is conserved between *Arabidopsis* and aspen. Moreover, our *PtrCSLD2* gene structure analysis from the *Populus* genome also showed the absence of introns in this region (Joshi, unpublished observations). To determine the extent of the *CSLD* gene family in aspen, Southern

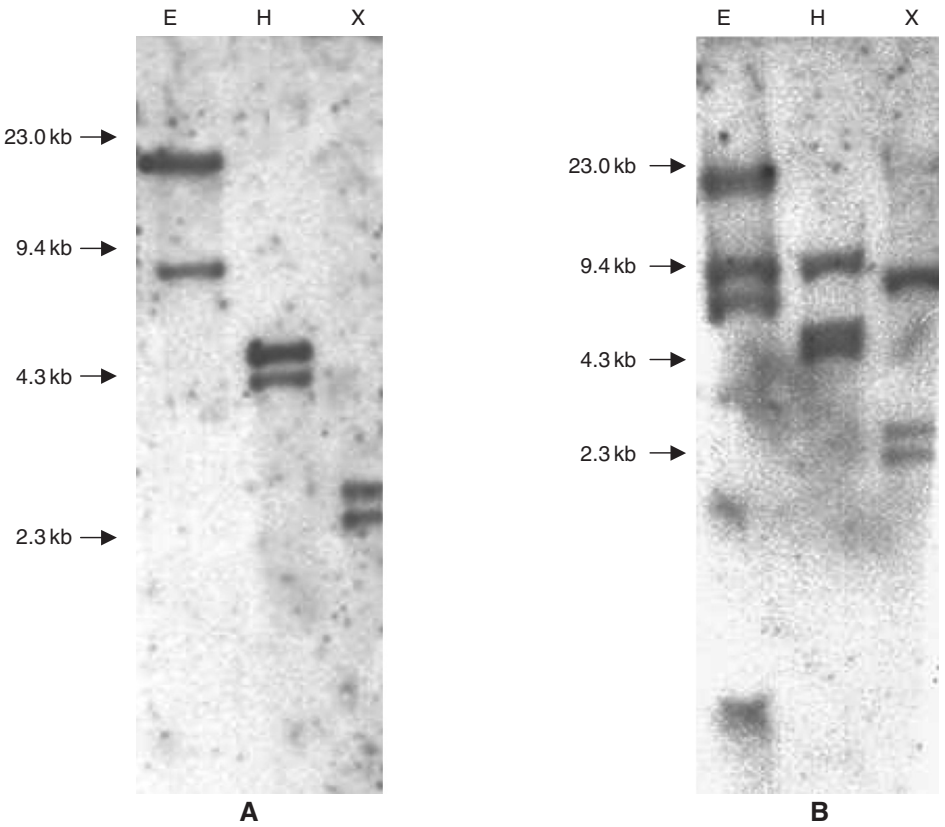

Fig. 2. Southern blot analysis of restriction digested aspen genomic DNA hybridized with *PtrCSLD2* HVRI-specific region (A) and entire *PtrCSLD2* cDNA (B) as radioactive probes. Lanes E, H and X indicate genomic DNA digested with restriction enzymes EcoR I, Hind III and Xba I, respectively. Molecular weight markers are indicated on the left.

analysis was performed using the entire available *PtrCSLD2* cDNA as a probe. Three bands were observed suggesting that the *CSLD* gene family consists of at least three members in the aspen genome (Fig. 2B). In an independent work, we have isolated two additional and distinct HVRII regions from *CSLD* gene family of the aspen genome (Joshi et al. unpublished observations). This suggests that the CSLD gene family in aspen consist of at least three genes.

PtrCSLD2 gene expression analysis by RT-PCR and in situ mRNA hybridization

Since *PtrCSLD2* was isolated from the developing xylem library, we were interested to see if it is highly expressed in xylem in comparison with other tissues. We first carried out RT-PCR experiments using a pair of primers corresponding to *PtrCSLD2* HVRI domain (Fig. 3). The *PtrCSLD2* transcripts were amplified from all seven RNA samples comprising of young internodes, xylem, apex, leaf, phloem, and 8–10 internodes. This suggests that *CSLD* genes are expressed all over the aspen plant. The slightly higher level of PtrCSLD2 expression in developing xylem tissues is an interesting and novel observation. Cloning of PCR products and sequencing of at least three clones from these amplified bands revealed the identity of the amplification products to be PtrCSLD2 HVR I region and only one species of transcript, 100% identical to *PtrCSLD2* HVRI region, was always amplified. The amplified bands in the internal control lanes using 5.8S rRNA primers were of equal intensity in all lanes suggest that the equal amount of RNA was used as a template for RT-PCR from all samples (Fig. 3).

The RT-PCR results can not always yield accurate information about the tissue or cell-specific patterns of gene expression since the starting plant material is derived from a mixture of heterogeneous cells. To determine which cells express *PtrCSLD2* genes in various aspen tissues, we performed in situ mRNA hybridization with sense (T7) and antisense (SP6) probes generated by in vitro transcription of the HVRI region of *PtrCSLD2* cDNA. As AtCSLD3 has been genetically associated with root development (Favery et al. 2001), we were interested in knowing if PtrCSLD2 would show similar expression patterns. The antisense *PtrCSLD2* transcripts were specifically localized to the vascular tissues in roots (Fig. 4A). The probe-specificity was confirmed by the absence of signal in the sections hybridized with the sense probe (Fig. 4B). The *PtrCSLD2* expression was also seen at low levels all over the root tip, shoot apex and in the xylem and phloem cells of the leaf (Fig. 4C, E and G) whereas hybridizations with sense PtrCSLD2 transcripts to similar sections did not show any specific staining pattern (Fig. 4D, F and H). To see if the expression was restricted to the young tissues alone, we also used stem sections from the slightly mature eighth internode

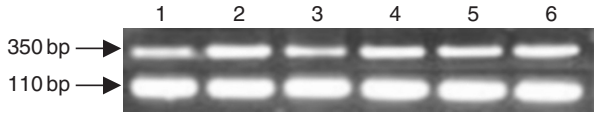

Fig. 3. Gene expression analysis in various tissues of aspen trees using RT-PCR. Total RNA from young 1–3 internodes (Lane 1), developing xylem from 3-year-old-aspen tree (Lane 2), young shoot apices (Lane 3), young leaves (Lane 4), phloem (Lane 5), and 8–10 internodes (Lane 6) were used for RT-PCR. The top panel shows RT-PCR products using *PtrCSLD2* HVRI-specific primers. Lower panel shows RT-PCR products using 5.8S rRNA-specific primers as internal control. Molecular weights of amplified bands are indicated on the left.

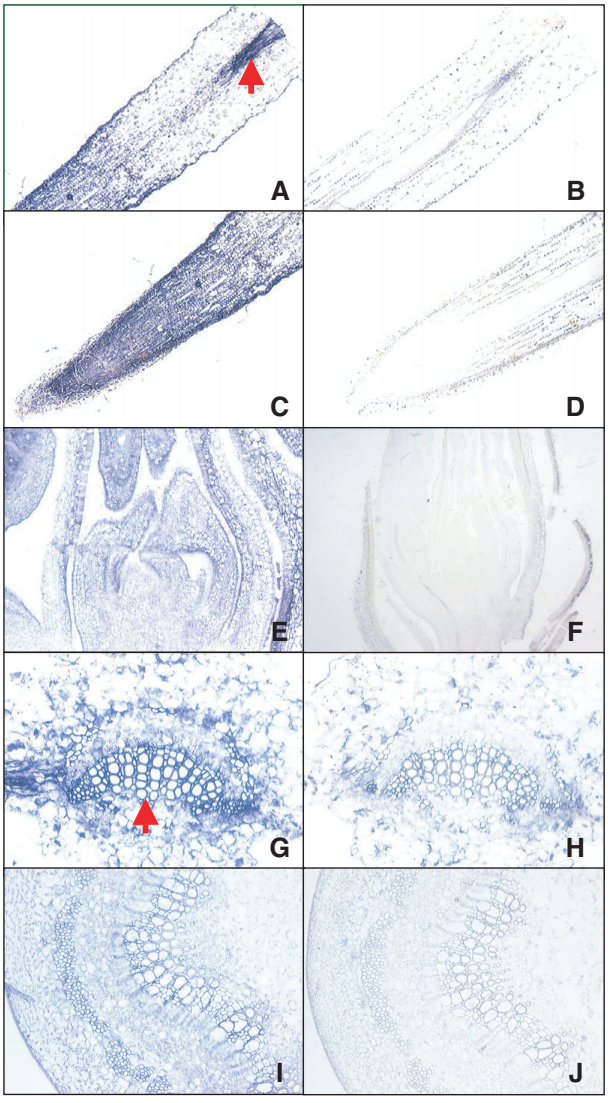

Fig. 4. In situ mRNA hybridization of PtrCSLD2 transcripts in root vascular tissue (A), root tip (C), shoot apex (E), leaf (G), and internode 8 of stem (I). All these sections represent hybridization with antisense PtrCSLD2 HVRI-specific RNA probe and the corresponding sections shown in B, D, F, H and J represent hybridization with sense PtrCSLD2 HVRI-specific RNA probe. Note the vascular tissues expressing higher levels of PtrCSLD2 transcripts are marked with red arrows.

from the top which showed overall low-level expression of *PtrCSLD2* transcripts (Fig. 4I) that was absent in control hybridization with sense probe (Fig. 4J). Thus the *PtrCSLD2* gene is expressed at a low level in all tissues examined which agrees with our RT-PCR results described above.

Phylogenetic analysis

A phylogenetic tree developed using the CLUSTALW program (http://www.ebi.ac.uk/clustalw/), visualized by TREEVIEW software (http://taxonomy.zoology.gla.uk/rod/rod.html) for 12 CSLD proteins from various plants suggests that three major groups of CSLDs are present in all plant genomes examined so far (Fig. 5). This is in agreement with our Southern blot analysis results as shown in Fig. 2B. Aspen PtrCSLD2 belongs to CSLD group II, in which AtCSLD3 has been associated with root hair tip development in *Arabidopsis* (Favery et al. 2001). NaCSLD1 protein involved in pollen and pollen tube development in tobacco belongs to another distinct clade, the CSLD group I. Interestingly, mutation in AtCSLD3 did not affect pollen tube development (Wang et al. 2001) and NaCSLD1 transcripts have been detected only at a low level in roots (Doblin et al. 2001). Thus different CSLDs from different clades may be involved in polysaccharide synthesis in different tissues and organs. However, *AtCSLD3* and *PtrCSLD2* transcripts are expressed in all organs and tissues examined whereas *PtrCSLD2* showed slightly higher expression in developing xylem tissues than the other tissues examined.

Discussion

Subsequent to prediction of *CSLD* genes from *Arabidopsis* genome (Richmond and Somerville 2000), there have been only three publications dealing with the *CSLD* genes, two describing *Arabidopsis AtCSLD3* and one describing tobacco *NaCSLD1* (Doblin et al. 2001, Favery et al. 2001, Wang et al. 2001). This is the first detailed report of the expression profile of a *CSLD* gene, *PtrCSLD2* from an economically important tree species, aspen. *PtrCSLD2* is expressed at a low level in all tissues examined with slightly higher expression levels in xylem tissues using RT-PCR and in situ hybridization. Isolation of PtrCSLD2 cDNA from the xylem library also corroborates these observations. The low-level expression of CSLD genes is in agreement with Favery et al. (2001) who could not detect *AtCSLD3* transcripts using northern blot analysis and had to resort to RT-PCR to detect the *AtCSLD3* transcripts. Similarly, Wang et al. (2001) also had to perform Southern blot of RT-PCR

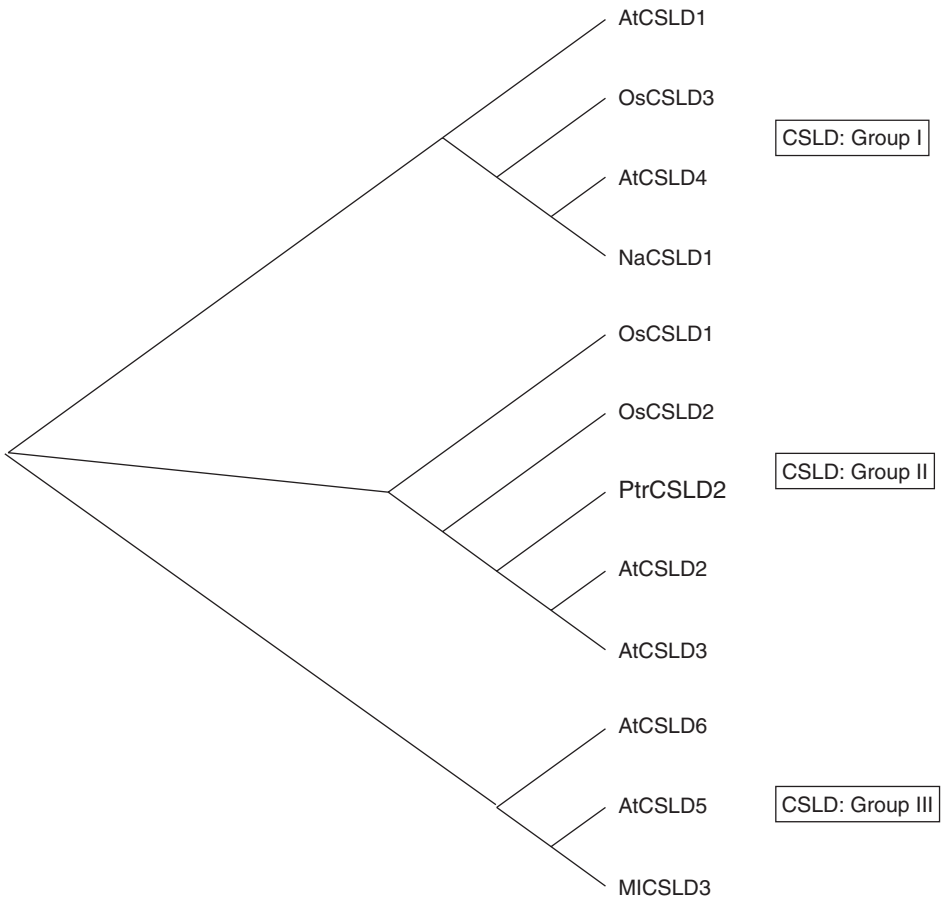

Fig. 5. Phylogenetic tree of 12 plant CSLD proteins using 'CLUSTALW' program and visualized with 'TREEVIEW' software as described in Materials and methods section. Nomenclature of CSLDs used here is same as described at http://cellwall.Stanford.edu (At, *Arabidopsis thaliana*; Mt, *Medicago truncatus*; Na, *Nicotiana alata*; Os, *Oryza sativa*; Ptr, *Populus tremuloides*). Three distinct clades of CSLDs are indicated on the right side in the open boxes.

products to examine the expression of *AtCSLD3* gene in various tissues suggesting a low level of CSLD expression. Using northern blots; Doblin et al. (2001) also could not detect any *CSLD* transcripts in vegetative tissues. They did not employ a parallel RT-PCR approach. This situation is in contrast with the occurrence of *CesA* transcripts that generally accumulate at a much higher level and are easily detected with northern blots (e.g. Taylor et al. 2000, Wu et al. 2000, Doblin et al. 2001). Thus the levels of *CSLD* transcripts are significantly lower in most of the plant organs and plants studied than *CesA* mRNAs. We believe that low-level expression of CSLD genes combined with their cell/tissue specificity might be the reason why impact of the *CSLD* mutations could not so far be related to cell wall polysaccharide alteration in the mutant *Arabidopsis* plants (Bonetta et al. 2002).

Genetic complementation of AtCSLD3-deficient mutants with wild-type *AtCSLD3* gene strongly suggests that the *AtCSLD3* gene is associated with primary cell wall development of root hair tip (Favery et al. 2001, Wang et al. 2001). However, pollen tube development was not affected in the *AtCSLD3* mutant (Wang et al. 2001) suggesting that not all cells growing by tip growth were affected by *AtCSLD3* mutation and possibly some other CSLD isoform may be involved in pollen tube growth. In support of such a theory, Doblin et al. (2001) indeed found that *NaCSLD1*, an *AtCSLD4* ortholog in tobacco, is exclusively expressed in pollens and pollen tubes but not expressed in other organs and tissues examined. The puzzling fact, however, remains that unlike *NaCSLD1*; the *AtCSLD3* and *PtrCSLD2* genes are expressed in all organs/tissues examined including xylem in aspen. However, the lack of the AtCSLD3 functional protein did not apparently produce any other phenotypic effect on xylem development in the *AtCSLD3* mutant. It is important to point out that both root hair tips and pollen tubes are single cell extensions but most of the previous RT-PCR or northern blot analysis experiments used RNA from a large group of heterogeneous cells thus not reflecting the gene expression profiles at the single-cell level and in situ mRNA hybridizations should clarify these issues.

Our in situ mRNA hybridization analyses, done for the first time using a plant CSLD probe, suggest that *PtrCSLD2* transcripts are localized at the growing tips of roots and shoots. Furthermore, localization of *PtrCSLD2* transcripts to the xylem cells in leaf, stem and root also suggests a probable role of CSLD in xylem development. These observations support a view that CSLD genes are expressed in all expanding or extending cells. Since xylem fibers elongate by intrusive tip growth (Haigler 1985, Mellerowicz et al. 2001), it is possible that *PtrCSLD2* is somehow involved in controlling the length of xylem fibers which is an economically important trait. Possible involvement of *CSLD* genes in regulation of fiber length in an internal tissue like xylem, a low level of CSLD gene expression and small growth habit of *Arabidopsis* plants may also explain the lack of a visible phenotype in xylem of *AtCSLD3* mutant, *kjk*. In future, the *PtrCSLD*2 gene will be down-regulated in transgenic aspen to shed more light on this puzzle. The availability of large quantities of xylem from such transgenic trees may also facilitate a better understanding of the functional roles of CSLDs in cell wall polymer synthesis during wood formation.

A large body of evidence has now accumulated for *CesA* genes playing pivotal roles in cellulose biosynthesis (Delmer 1999, Joshi 2003a, b, c). Due to the low-level expression of *CSLD* genes that potentially encode a protein sharing the same domain structure as CesA proteins, it is still unclear whether they play a role in cellulose biosynthesis or the synthesis of other matrix polysaccharides. Favery et al. (2001) localized AtCSLD3-GFP fusion proteins at the endomembrane systems suggesting that AtCSLD3 is involved in matrix polysaccharide synthesis rather than cellulose synthesis. However, localization of such fusion proteins at the plasma membranes (the major sites of cellulose biosynthesis) can not be completely ruled out as suggested by Reiter (2002). Doblin et al. (2001) have taken another view and based on the domain similarities of CesAs and CSLDs they proposed that similar to CesAs, CSLDs are also involved in β-linked polysaccharide or cellulose biosynthesis. One of the ways to resolve this controversy is to use trees as experimental materials. As trees are much more robust species than *Arabidopsis,* this may make them more suitable for detailed cell wall polymer analysis especially when the gene under consideration is expressed at such a low level.

Mutant complementation studies as an evidence for functional assignment, as carried out for AtCSLD3, cannot yet be undertaken in slow growing trees where mutants are much more difficult to obtain than in *Arabidopsis*. Molecular cloning of an important *CSLD* gene from an economically important tree, aspen is the first step towards devising such experiments using RNA interference (*RNAi*) technology of gene silencing (Fire et al. 1998). The highly variable HVRI regions from *PtrCSLD2* genes will be ideal targets for *RNAi*-mediated inhibition of expression of a specific *CSLD* gene that apparently belongs to a multigene family. Finally, when even a single amino acid change in *CesA* or *CSLD* genes from *Arabidopsis* results in mutant plants with altered phenotype and defective cell wall polymer synthesis (review: Joshi 2003a), over 100 amino acid differences between *Arabidopsis* and aspen CSLD proteins may account for a number of functional differences between aspen and *Arabidopsis* CSLD proteins. It is therefore of a great interest to examine whether the aspen *PtrCSLD2* gene can complement *AtCSLD3* mutation. Successful complementation will indicate that the non-conserved residues are not important in determining CSLD functions. Alternatively, a failed experiment may be due to amino acid differences between these two proteins. Domain swapping and site-directed mutagenesis may further assist in defining the functionally important amino acids in CSLD proteins.

Finally, tree biology is distinctly different from that of *Arabidopsis* biology and it is yet to be determined whether the aspen genome has structurally distinct genes responsible for large amounts of cell wall polymer production or cell wall biogenesis is just a passive process that goes hand-in-hand with other growth habit-related genes. Differential regulation of *Arabidopsis*-like cell wall biosynthesis-related genes is also another possibility. In conclusion, first cloning and detailed characterization of a *CSLD* gene from developing xylem of aspen trees as reported here will open up many new avenues of exploration that will one day enable us to translate the lessons learnt from *Arabidopsis* to economically important trees such as aspen.

Acknowledgements – We thank Drs Vincent Chiang, Laigeng Scott Harding and Glenn Mroz, School of Forest Resources and Environmental Science for their help with this research. We also wish to thank Mr Priya Ranjan for his help in searching CSLD homologues in raw poplar genome data. We are thankful to the Arabidopsis Biological Resource Center for providing the EST clone H10C11T7 and USDA McIntire-Stennis fund for partial financial support to C.P.J.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res 25: 3389–3402

Bonetta DT, Facette M, Raab TK, Somerville CR (2002) Genetic dissection of plant cell-wall biosynthesis. Biochem Soc Trans 30: 298–301

Cutler S, Somerville CR (1997) Cellulose synthesis: cloning *in silico*. Current Biol 7: R108–R111

Delmer DP (1999) Cellulose biosynthesis: Exciting times for a difficult field of study. Ann Rev Plant Physiol Plant Mol Biol 50: 245–276

Doblin MS, De Melis L, Newbigin E, Bacic A, Read SM (2001) Pollen tubes of *Nicotiana alata* express two genes from different beta glucan synthase families. Plant Physiol 125: 2040–2052

Favery B, Ryan E, Foreman J, Linstead P, Boudonck K, Steer M, Shaw P, Dolan L (2001) KOJAK encodes a cellulose synthase-like protein required for root hair cell morphogenesis in *Arabidopsis*. Genes Dev 15: 79–89

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double stranded RNA in C. Elgans Nature 391: 806–811

Haigler C (1985) The functions and biogenesis of native cellulose. In: Nevell TP, Zoronian SH (eds) Cellulose Chemistry and Applications. Ellis Horwood, Chichester, pp 30–83

Holland N, Holland D, Helentjaris T, Dhugga K, Xoconostle-Cazares B, Delmer DP (2000) A comparative analysis of the cellulose synthase (CesA) gene family in plants. Plant Physiol 123: 1313–1323

Joshi CP (2003a) Molecular genetics of cellulose biosynthesis in trees. In: Kumar S, Fladung M (eds) Molecular Genetics and Breeding of Forest Trees. Haworth Press, New York, in press

Joshi CP (2003b) Molecular biology of cellulose biosynthesis in plants. In: Pandalai S (ed) Recent Research Developments in Plant Molecular Biology. Research Signpost Press, Kerala, India, pp 19–38

Joshi CP (2003c) Xylem-specific and tension stress responsive expression of cellulose synthase genes from aspen trees. Appl Biochem Biotech 105: 17–26

Li L, Popko JL, Zhang X-H, Osakabe K, Tsai C-J, Joshi CP, Chiang VL (1997) A novel multifunctional *O*-methyltransferase implicated in a dual methylation pathway associated with lignin biosynthesis in loblolly pine. Proc Natl Acad Sci USA 94: 5431–5466

Mellerowicz EJ, Baucher M, Sundberg B, Boerjan W (2001) Unravelling cell wall formation in the woody dicot stem. Plant Mol Biol 47: 239–274

Osakabe K, Tsao C-C, Li L, Popko JL, Umezawa T, Carraway DT, Smeltzer RH, Joshi CP, Chiang VL (1999) Coniferyl aldehyde 5-hydroxylation and methylation direct syringyl lignin biosynthesis in angiosperms. Proc Natl Acad Sci USA 96: 8955–8960

Pear JR, Kawagoe Y, Schreckengost WE, Delmer DP, Stalker DM (1996) Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. Proc Natl Acad Sci USA 93: 12637–12642

Reiter W-D (2002) Biosynthesis and properties of the plant cell wall. Curr Opinion Plant Biol 5: 536–542

Richmond TA, Somerville CR (2000) The cellulose synthase superfamily. Plant Physiol 124: 495–498

Richmond TA, Somerville CR (2001) Integrative approaches to determining CSL function. Plant Mol Biol 47: 131–143

Samuga A, Joshi CP (2002) A new cellulose synthase gene (*PtrCesA2*) from aspen xylem is orthologous to *Arabidopsis AtCesA7* (irx3) gene associated with secondary cell wall synthesis. Gene 296: 37–44

Saxena IM, Brown RM, Fevre M, Geremia RA, Henrissat B (1995) Multidomain architecture of β-glycosyltransferases: implications for mechanism of action. J Bacteriol 177: 1419–1424

Taylor NG, Laurie S, Turner SR (2000) Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in Arabidopsis. Plant Cell 12: 2529–2540

Wang X, Cnops G, Vanderhaeghen R, De Block S, Montague MV, Van Lijsebettens M (2001) AtCSLD3, a cellulose synthase-like gene important for root hair growth in Arabidopsis. Plant Physiol 126: 575–586

Wu L, Joshi CP, Chiang VL (2000) A xylem-specific cellulose synthase gene from aspen (*Populus tremuloides*) is responsive to mechanical stress. Plant J 22: 495–502

Wullschleger SD, Jansson S, Taylor G (2002) Genomics and forest biology – *Populus* emerges as the perennial favorite. Plant Cell 14: 2651–2655

Edited by D. Van Der Straeten