Tree Physiology 24, 543–550

Molecular cloning of ten distinct hypervariable regions from the cellulose synthase gene superfamily in aspen trees

XIAOE LIANG[1,2] and CHANDRASHEKHAR P. JOSHI[1,3]

[1] *Plant Biotechnology Research Center, School of Forest Resources and Environmental Science, Michigan Technological University, Houghton, MI 49931, USA*

[2] *Present address: Department of Forestry, North Carolina State University, Raleigh, NC 27695, USA*

[3] *Corresponding author (cpjoshi@mtu.edu)*

Received June 27, 2003; accepted November 9, 2003; published online March 1, 2004

Summary Recent molecular genetic data suggest that cellulose synthase (*CesA*) genes coding for the enzymes that catalyze cellulose biosynthesis (CESAs) in *Arabidopsis* and other herbaceous plants belong to a large gene family. Much less is known about *CesA* genes from forest trees. To isolate new *CesA* genes from tree species, discriminative but easily obtainable homologous DNA probes are required. Hypervariable regions (HVRII) of *CesA* genes represent highly divergent DNA sequences that can be used to examine structural, expressional and functional relationships among *CesA* genes. We used a reverse transcriptase-polymerase chain reaction (RT-PCR)-based technique to identify HVRII regions from eight types of *CesA* genes and two types of *CesA*-like *D* (*CslD*) genes in quaking aspen (*Populus tremuloides* Michx.). Comparison of these aspen CESA/CSLD HVRII regions with the predicted proteins from eight full-length *CesA*/*CslD* cDNAs available in our laboratory and with searches for aspen *CesA*/*CslD* homologs in the recently released *Populus trichocarpa* Torr. & A. Gray. genome confirmed the utility of this approach in identifying several *CesA*/*CslD* gene members from the *Populus* genome. Phylogenetic analysis of 56 HVRII domains from a variety of plant species suggested that at least six distinct classes of CESAs exist in plants, supporting a previous proposal for renaming HVRII regions as class-specific regions (CSR). This method of CSR cloning could be applied to other crop plants and tree species, especially softwoods, for which the whole genome sequence is unlikely to become available in the near future because of the large size of these genomes.

Keywords: CESA, CSLD, CSR, HVR, Populus, RT-PCR.

Introduction

Precise regulation of cell wall biogenesis is important for normal plant growth and development (Carpita and McCann 2000). In primary cell walls, cellulose is synthesized at the plasma membrane, whereas hemicellulose and pectins are synthesized in the endomembrane system but are delivered to the cell surface by Golgi-derived vesicles. In secondary cell walls, in addition to cellulose and hemicellulose, a significant amount of lignin is also deposited, imparting rigidity and strength to the cell wall. Cellulose deposition thus acts as a foundation step for proper cell wall formation, growth and development.

At least two types of cellulose synthases (CESA) are believed to be necessary for cellulose synthesis in primary and secondary cell walls of plants (Haigler and Blanton 1996). However, because of the labile nature of CESA enzyme complexes, the first plant gene encoding the catalytic subunit of CESA enzymes was isolated only recently (Pear et al. 1996) and the molecular genetic proof of their involvement in cellulose biosynthesis was unavailable until 1998 (Arioli et al. 1998). Several advances in our understanding of *CesA* gene structure, expression and function have been made recently and reviewed extensively (e.g., Delmer 1999, Dhugga 2001, Perrin 2001, Doblin et al. 2002, Joshi 2003*a*, 2003*b*, 2004).

The *Arabidopsis* genome hosts a large *CesA* gene family consisting of at least 10 distinct members (*AtCesA1* to *AtCesA10*) (Richmond 2000). Based on studies involving complementation of cellulose-deficient mutants, at least three distinct *CesA* genes (*AtCesA1*, *AtCesA3* and *AtCesA6*) have been associated with primary cell wall development, and another three distinct *CesA* genes (*AtCesA4*, *AtCesA7* and *AtCesA8*) are associated with secondary cell wall development in *Arabidopsis* (Joshi 2003*a*). However, little is known about *CesA* genes from trees. Thus far, only four full-length *CesA* cDNAs from poplar trees have been reported (Wang and Loopstra 1998, Wu et al. 2000, Samuga and Joshi 2002, Kalluri and Joshi 2003). Therefore, isolation of new *CesA* genes from trees is a major research priority to determine if *CesA* genes from trees are structurally and functionally similar to *Arabidopsis CesA* genes.

All known plant CESA proteins consist of highly diverged (hypervariable) and highly conserved domains and have two N-terminal and six C-terminal transmembrane domains (Joshi 2004). Of these, a second hypervariable domain, also known as HVRII, is situated in the central globular region containing all processive glycosyltransferase signature motifs (Saxena et

al. 1995). Vergara and Carpita (2001) have recently proposed renaming the HVRII regions as class-specific regions (CSR), because although these regions are variable among CESA paralogs from the same plant species, they appear to be highly conserved among CESA orthologs from various plants. The HVRII regions are flanked by short oligopeptide motifs that are highly conserved in most CESA proteins. Such oligopeptide motifs may also be used for designing universal and degenerate PCR-primers to amplify the intervening highly divergent HVRII regions corresponding to the various *CesA* paralogs from a particular plant species. Such primers are also most likely to amplify HVRII regions from some members of a closely related family of cellulose synthase-like D (*CslD*) genes (Richmond 2000, Favery et al. 2001). These amplified HVRII regions may assist in identification of new *CesA*/*CslD* genes by cDNA/genomic library screening under highly stringent hybridization conditions. We tested this hypothesis by using aspen (*Populus tremuloides* Michx.) as a model tree system.

We used a reverse transcriptase-polymerase chain reaction (RT-PCR) based strategy to identify 10 distinct HVRII regions of aspen *CesA*/*CslD* members. These HVRII sequences were further used to search the recently released raw *Populus* genome data to validate the existence of such genes in the *Populus* genome (http://genome.jgi-psf.org/poplar0/poplar0.info.html). Next, we compared these HVRII regions to corresponding regions from eight full-length aspen *CesA*/*CslD* cDNAs available in our laboratory, confirming the success of this rapid and simple RT-PCR-based approach. Finally, we compiled currently available data about HVRII regions and confirmed the existence of at least six distinct classes of CESAs in a variety of higher plant species, thereby strengthening a recent proposal by Vergara and Carpita (2001) to rename HVRII as class-specific regions.

Materials and methods

Nomenclature

All *CesA* genes and cDNAs are indicated in italic letters and their encoded proteins are shown in capital letters (CESA). Each *CesA* begins with 2–3 letters indicating the genus and species.

Plant materials

Young leaves and developing xylem tissues were collected from quaking aspen (*Populus tremuloides* Michx., Clone 271) plants grown in a greenhouse at the School of Forest Resources and Environmental Sciences, Michigan Technological University. All plant samples were fixed in liquid nitrogen and stored at –80 °C until used.

Total RNA isolation and RT-PCR-mediated amplification of HVRII regions

Total RNA from aspen tissues was isolated using RNeasy Plant mini kit (Qiagen, Germantown, MD) as described by Samuga and Joshi (2002). First-strand cDNA synthesis was performed with 10 ng of total RNA from aspen xylem or leaf tissues and oligo-dT_{16} primer according to the procedure described in the GeneAMP Gold RNA PCR kit (Applied Biosystems, Foster City, CA). The degenerate primers used for the PCR step of the RT-PCR were: HVR2F (5′-TGYTATGTY CAGTTYCCWC-3′) and HVR2R (5′-GANCCRTARATCCA YCC-3′). The cDNA was amplified under the following PCR conditions: 95 °C for 10 min, followed by two cycles of 94 °C for 60 s, 41 °C for 90 s and 72 °C for 120 s, followed by another 28 cycles of 94 °C for 60 s, 55 or 45 °C for 90 s and 72 °C for 120 s. The amplified cDNA was diluted 50× with water and amplified once more under the same PCR conditions. The resultant products were separated by electrophoresis on 1% agarose gel and the amplified 600 bp band was purified using QIAquick gel extraction kit (Qiagen) and cloned in either TOPO pCR 2.1 vector (Invitrogen, Carlsbad, CA) or pGEM T-easy vector (Promega, Madison, WI). The recombinant plasmids were isolated and sequenced using Big Dye (dRhodamine) terminator cycle sequencing-ready reaction kit (Applied Biosystems, Foster City, CA). The cloned PCR products were subjected to automated DNA sequencing (Applied Biosystems ABI310 Genetic Analyzer). A total of 78 clones were sequenced from one direction using M13F primers (StrataGene, La Jolla, CA). The nomenclature for clones selected for sequencing was as follows. When xylem RNA was used as a template for RT and the reannealing during PCR step was done at 45 and 55 °C, such clones were denoted by the prefix, 45X and 55X, respectively. Similarly, when leaf RNA was used as a template and the reannealing was done at 45 and 55 °C, the samples had the prefix 45L and 55L, respectively. The ten selected plasmids representing the longest sequences in each *CesA*/*CslD* group were sequenced from the other direction using M13R primers (StrataGene). Because the HVR2F and HVR2R primers were made from conserved regions that do not extend much beyond the primer sequences, the predicted amino acid sequences from each of these ten sequences between the highly conserved ALYG and VISCG motifs were considered as HVRII regions and were used for further sequence comparison.

Sequence analysis

The DNA and protein sequences were analyzed using various program routines from GCG (Genetics Computer Group, Madison, WI) package Version 10.2. Multiple sequence alignment of various CESA HVRII domains was made with the PILEUP program from the GCG package. Comparison of the 78 single-pass sequences with each other was made with a modified GAP program, DOUBLEGAP, from the GCG package that reiteratively compares each sequence with the remaining sequences using the GAP routine and which grouped the sequences into 10 groups. All DNA sequences sharing > 90% identity were considered to belong to one group. These results were confirmed with a CAP3 sequence assembly program (Huang and Madan 1999) that produces sets of contiguous sequences (contigs) by searching for overlapping sequences of a particular window size. We used a window size of 25 with overlapping pieces showing > 90% identity. Phylo-

grams were developed using PAUP Version 4.0b10 program (Phylogenetic Analysis Using Parsimony, Sinaur Associates, Sunderland, MA) with parsimony analysis and a heuristic search algorithm. Bootstrap analysis with 1000 replicates and a value of over 70% were used for development of the phylogenetic tree.

Results

Sequence analysis of available CESA proteins from plants

To design a suitable pair of degenerate primers for HVRII amplification from most aspen *CesA* genes, we first performed a systematic analysis of ten predicted CESA proteins from *Arabidopsis* (Richmond 2000). Multiple sequence alignment of these CESA proteins revealed alternate arrangement of highly diverged domains (HVRI and HVRII) and highly conserved domains (A and B). Two N-terminal transmembrane domains were interposed between the HVRI and A domain and six transmembrane domains followed the B domain (Figure 1). Between the two HVR domains, only HVRII is flanked by the conserved domains. Therefore a pair of PCR primers was designed on the basis of conserved amino acids at the end of sub-domain A and at the beginning of sub-domain B. Such primers will amplify HVRII regions from several members of *CesA* genes from the aspen genome, provided CESA proteins of *Arabidopsis* and aspen share similar conserved regions flanking HVRII. To verify CESA structural conservation across various higher plants, we included in our analysis 13 additional CESA protein sequences from cotton, rice, aspen, hybrid poplar, tobacco and corn (Pear et al. 1996, Wang and Loopstra 1998, Holland et al. 2000, Laosinchai et al. 2000, Wu et al. 2000, Doblin et al. 2001). These additional CESA sequences also shared the same overall structural features with *Arabidopsis* CESAs and showed a high degree of conservation at the end of the A domain and at the beginning of the B domain. Therefore, for the isolation of HVRII regions from unknown aspen *CesA* members, we designed degenerate PCR primers on the basis of the highly conserved end of sub-domain A and the beginning of sub-domain B. Such primers may also amplify HVRII regions from another *CesA*-like (*Csl*) group of genes designated as *CslD*s that share only ~45% overall identity with *CesA*s (Joshi 2004).

Designing primers and PCR-mediated amplification of aspen CesA HVRII regions

Toward the end of the sub-domain A, a short oligopeptide, CYVQFPQ is highly conserved in CESA proteins. Moreover, at the beginning of sub-domain B, GWIYGS is also highly conserved. Based on the DNA sequences encoding these two oligopeptides from all known plant CESA proteins, we designed two degenerate primers: HVR2F 5′-TGYTATGT YCAGTTYCCWC-3′ (16× degeneracy and HVR2R 5′-GAN CCRTARATCCAYCC-3′ (32× degeneracy).

We hypothesized that, if the general *CesA* gene structure is conserved between aspen and *Arabidopsis*, the HVRII regions amplified from aspen *CesA* genes using HVR2F and HVR2R primers will have lengths of ~0.8–1.3 kb. Similarly, the HVRII regions from *CslD* genes flanked by HVR2F and HVR2R primers are expected to be about 0.6 kb. Therefore, we first performed PCR reactions using aspen genomic DNA with HVR2F and HVR2R primers (Joshi 2004). Two fragments, one corresponding to ~1.3 kb and the second to ~0.6 kb were reproducibly amplified. However, cloning and sequencing of at least 30 randomly selected clones containing these fragments revealed that only a single HVRII region from a new *CesA* gene of aspen and two HVRII regions corresponding to two new *CslD* genes from aspen were amplified. Contrary to our expectation, HVRII regions from a previously isolated *PtrCesA1* gene from aspen (Wu et al. 2000), or any other new *CesA*/*CslD* genes, could not be isolated even after extensively changing the PCR conditions (Joshi et al., unpublished observations).

Use of total RNA as an RT-PCR template instead of genomic DNA may alleviate the problems associated with *CesA* introns during PCR amplifications. Therefore, we used the RT-PCR approach on two distinct total RNA templates, one from developing xylem (enriched with secondary cell-wall-forming cells) and the other from leaf tissues (enriched with primary cell-wall-forming cells). We also used two re-annealing temperatures of 45 and 55 °C during the PCR stage so that different types of HVRII regions may be amplified because of the change in re-annealing temperatures. An amplified 600-bp band, in each case, presumably consisting of several HVRII regions from *CesA*/*CslD* cDNAs from aspen, was cloned in a suitable plasmid vector and a total of 78 clones (19–20 clones from each of the four treatments) were randomly selected for single-pass sequencing, similar to the pro-

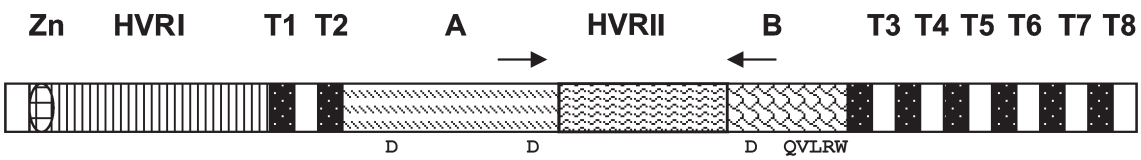

Figure 1. Diagrammatic representation of the general structure of cellulose synthase proteins (CESA). The T1 to T8 blocks represent the predicted transmembrane domains. The N-terminal region shows the presence of a putative zinc-binding motif (Zn), followed by the first hypervariable region (HVRI). The second hypervariable region (HVRII) is flanked by the highly conserved sub-domains A and B. The position of the processive glycosyltransferase signature proposed by Saxena et al. (1995) is indicated below sub-domains A and B. The positions of HVR2F and HVR2R primers used for RT-PCR are indicated by horizontal arrows flanking the HVRII regions.

cedures used with expressed sequence tags (ESTs) (Adams et al. 1993).

Sequence analyses of potential CESA/CSLD HVRII regions from aspen

Comparison of the 78 sequences with each other using the DOUBLEGAP program resulted in 10 groups as shown in Figure 2. All sequences in each group shared 90% or more identity. These results were confirmed by using a CAP3 sequence assembly program (Huang and Madan 1999). In our laboratory, seven full-length *CesA* cDNAs and one full-length *CslD* cDNA from aspen xylem have so far been isolated (Wu et al. 2000, Samuga and Joshi 2002, 2004, Kalluri and Joshi 2003, U. Kalluri, A. Samuga and C.P. Joshi, Michigan Technological University, unpublished data). Comparison of 10 representative HVRII sequences described above with corresponding HVRII regions from the eight aspen full-length *CesA*/ *CslD* cDNAs (*PtrCESA1–PtrCESA7* and *PtrCSLD2*) allowed us to propose new names for each of these groups based on their identity with known aspen *CesA*/*CslD* genes as shown in Figure 2 (P1–P9 to indicate their PCR origin). Sequences within each group were amplified from different templates and at different re-annealing temperatures. For example, groups P5 and P6 consisted of sequences from 45L, 45X and 55L samples, and 45L, 55L and 55X samples, respectively. However, group P3 was amplified only from xylem templates, whereas group P4 resulted only from leaf samples. Group P5A originated only from 45 °C re-annealing temperatures. Groups P1, P7, P7A and P9 were represented by a single clone, but group P6 had a total of 28 clones. This indicated that our RT-PCR strategy successfully yielded a variety of HVRII regions from both plant tissues and at two re-annealing temperatures. At least one of the longest sequences from each group was further selected and sequenced from the other end using an M13R primer.

Table 1 shows a comparison of the amino acids among 10 types of aspen HVRII regions. Overall, representatives of Groups P8 and P9 shared limited similarity with each other (53%) and were distinctly different from Group P1 to P7A (17–35%). Comparison of all these sequences with aspen CESA and CSLD HVRII regions also confirmed that Groups P1 to P7A belonged to various CESA proteins and Groups P8 and P9 belonged to CSLD proteins. Moreover, Groups P1 to P7A shared a limited identity of 37–65% with each other, but groups P5 and P5A shared 81% identity. Thus, this study yielded at least eight types of CESA and two types of CSLD HVRII regions that could further be used for cDNA library screening to obtain new aspen *CesA*/*CslD* genes.

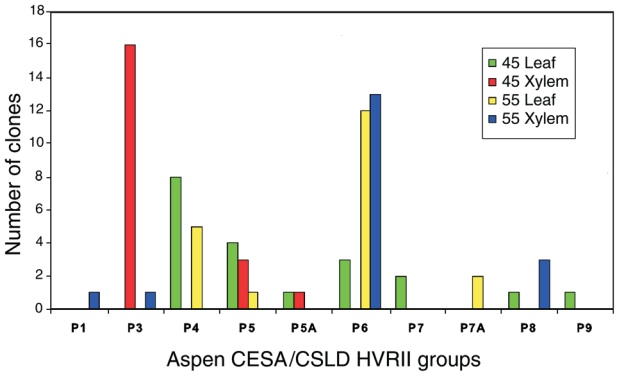

Figure 2. Classification of 78 single-pass sequences from RT-PCR experiments into 10 groups that have been renamed according to their predicted protein's identity (> 90%) with available predicted proteins from aspen full-length cellulose synthase/cellulose synthase-like D (*CesA*/*CslD*) cDNAs (P1–P9). Groups P8 and P9 did not closely match *PtrCesA*s, but showed higher identity with *PtrCSLD2*. Groups P5A and P7A shared 81 and 65% identity with *PtrCESA5* and *PtrCESA7*, respectively, and were designated as separate groups. The number of clones in each group, originating from one of the four treatments as shown on the right, is indicated by the height of bars.

The relationship between aspen HVRII isolated here using RT-PCR with the corresponding HVRII domains from other known aspen and *Arabidopsis* CESA/CSLD proteins is shown in Figure 3. Seven out of 10 isolated HVRII regions (faint yellow circle in Figure 3) shared a high percentage of identity (94–100% as indicated in green circle of Figure 3) with eight aspen CESA/CSLDs (Figure 3, central faint blue circle) as indicated by the black letters. While validating our method, we isolated three new HVRII regions by RT-PCR, namely P5A, P7A and P9 (indicated in red letters in faint yellow circle of Figure 3) that could be used in the future to isolate corresponding full-length *CesA*/*CslD* cDNAs by high-stringency cDNA library screening. The HVRII region of *PtrCesA2* cDNA re-

Table 1. Comparison of amino acid sequence identity (%) among 10 aspen HVRII domains shown in Figure 2.

	P1	P3	P4	P5	P5A	P6	P7	P7A	P8	P9
P1	–	53	44	48	48	37	38	43	24	21
P3		–	57	52	59	50	55	59	26	17
P4			–	59	57	44	48	49	35	28
P5				–	81	48	56	53	22	28
P5A					–	50	53	54	25	31
P6						–	59	63	25	34
P7							–	65	26	25
P7A								–	20	31
P8									–	53
P9										–

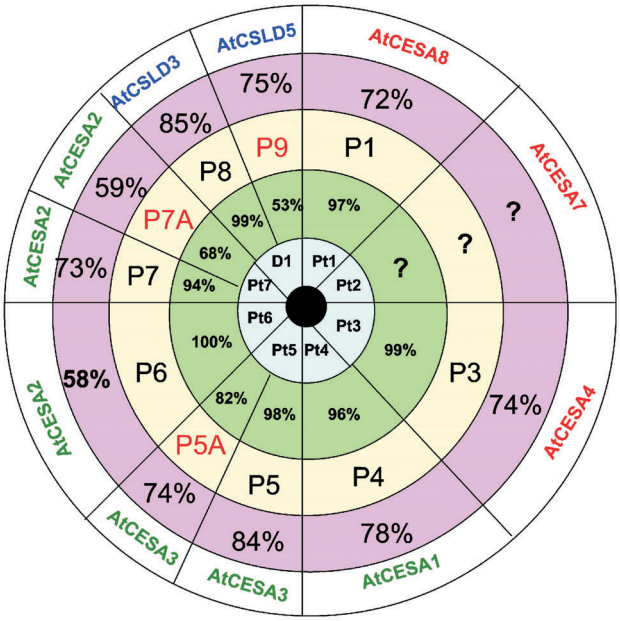

Figure 3. Relationship between P1 to P9 hypervariable II (HVRII) domains from Figure 2 with various known aspen and *Arabidopsis* cellulose synthase/cellulose synthase-like D (CESA/CSLD) proteins. The predicted proteins from aspen *CesA*/*CslD* HVRII regions isolated here are shown in the middle yellow circle. The new sequences, P5A, P7A and P9 that could be used for cDNA library screening in the future, are shown in red. The eight complete aspen CESA/CSLD proteins are shown in the central faint blue circle. Abbreviations: Pt1 = PtrCESA1; Pt2 = PtrCESA2; Pt3 = PtrCESA3; Pt4 = PtrCESA4; Pt5 = PtrCESA5; Pt6 = PtrCESA6; Pt7 = PtrCESA7; and D1 = PtrCSLD2 (GenBank accession no. AY162184). GenBank accession numbers for all aspen *CesA*s are given in the legend of Figure 4. The % identity between each of the aspen HVRIIs isolated here and the corresponding region from full-length aspen CESA/CSLD proteins is shown in the green circle. The outermost white circle shows *Arabidopsis* CESA/CSLDs sharing maximum identity with P1–P9 HVRII regions as shown in the pink circle. PtrCESA2 sharing high identity with *Arabidopsis* AtCESA7 is not represented in our current collection and is indicated here by a question mark.

ported by Samuga and Joshi (2002) was not found in the current collection of HVRII regions. It is possible that additional sequencing of randomly selected clones or use of additional tissues or PCR conditions will yield information about the missing aspen *CesA*/*CslD* members in this collection. We also compared the aspen HVRII regions isolated here with the corresponding regions from *Arabidopsis* CESA/CSLD proteins. The *Arabidopsis* sequence showing maximum identity with a particular aspen HVRII region is shown in the white colored circle of Figure 3 with percentage identity indicated in the pink colored circle. Thus, *Arabidopsis* and corresponding aspen CESA/CSLD HVRII regions share 58 to 85% identity.

Search for aspen CesA/CslD homologs in the Populus genome

In June 2003, the Joint Genome Institute released 2.3 Gbases of raw genome sequence data from black cottonwood, *Populus trichocarpa*, Torr. & A. Gray. (http://genome.jgi-psf.org/poplar0/poplar0.info.html). By September 2003, the released data had grown to over 4 Gbases with ~5.5 million sequences. The fully annotated genome information with more than 6× coverage may be released by early 2004 (Wullschleger et al. 2002*a*, 2002*b*). Although *P. trichocarpa*, the species used for genome sequencing, and *P. tremuloides*, our study species, differ, we expect that both *Populus* species will share similar genes because of their close relationship and ability to hybridize (Bradshaw et al. 2000). To test this hypothesis, we searched for the genomic counterparts of the 10 aspen *CesA*/*CslD* HVRII regions reported here in the *P. trichocarpa* genome data. Moreover, we included the HVRII region from our previously reported *PtrCesA2* cDNA (Samuga and Joshi 2002), which is not represented in the current RT-PCR experiment.

Assuming that *CesA*/*CslD* gene structure is conserved between *Arabidopsis* and *Populus*, a single intron is likely present within the HVRII region of most *CesA* genes but not in the *CslD* HVRII regions from *P. trichocarpa*. Therefore, we searched the *P. trichocarpa* genome trace files (each with 500–700 bp of good quality data) with *CesA*/*CslD* cDNA se-

Table 2. Search for the introns in poplar genomic fragments that are homologous to aspen cellulose synthase/cellulose synthase-like D (*CesA*/*CslD*) regions corresponding to HVRII isolated in this study (with the exception of P2 that was isolated by Samuga and Joshi (2002)).

Group	Length of query (bp)	Genomic match	Intron length (bp)	*Arabidopsis* ortholog and length of intron in HVRII
P1	324	XXI155316.b1	91	*AtCesA8*, no intron
P2	264	XXI394296.g1	97	*AtCesA7*, 95 bp
P3	381	TRE149118.b2	no intron	*AtCesA4*, no intron
P4	333	XXI47215.g1	363	*AtCesA1*, 87 bp
P5	339	XXI690544.b1	365	*AtCesA3*, 163 bp
P5A	336	XXI910686.b2	376	*AtCesA3*, 163 bp
P6	345	XXI587362.g1 & XWN135647.y1	483	*AtCesA2*, 82 bp
P7	339	XWN10986.x1	94	*AtCesA2*, 82 bp
P7A	339	XXI243433.b1	245	*AtCesA2*, 82 bp
P8	324	XXI646988.b1	no intron	*AtCslD3*, no intron
P9	342	TRE209658.b2	no intron	*AtCslD5*, no intron

quences corresponding with HVRII regions from *P. tremuloides*. Out of several positive hits for each HVRII region, at least one genomic sequence encompassed substantial portions of HVRII regions sharing > 90% identity in the coding region and yielded information about the presence or absence of the intron as well as intron length as presented in Table 2. For P6, two genomic sequences were necessary to assemble information about the intron because the intron length in this case is 483 bp.

Table 2 shows the group name and the length of the aspen HVRII cDNA probe used, raw genome sequence match considered for determining the intron presence, length of the intron if present and identity of the *Arabidopsis* ortholog with the intron length, if present. Except for the P3 fragment that has no intron in the HVRII region, the remaining eight *CesA* genes have introns ranging from 91 to 483 bp and, in each case, introns followed the canonical GT-AG rule of intron ends (Brown 1986). *Arabidopsis AtCesA4*, corresponding to aspen P3 HVRII, also lacks an intron in that region. *Arabidopsis AtCesA8* has no intron in the HVRII region, but its poplar ortholog, P1 HVRII region, shows the presence of a 91 bp intron. Most of the other *Arabidopsis CesA* genes show the presence of one intron in the HVRII region with sizes ranging from 77 to 163 bp. *Populus* introns in the HVRII region are, however, larger (91–483 bp) than *Arabidopsis* introns. The two *Populus CslD* HVRII regions had no introns similar to their *Arabidopsis CslD* counterparts. This comparison suggests that both *Populus* and *Arabidopsis* are similar in their *CesA*/*CslD* genomic structures corresponding to HVRII regions.

Universality of HVRII domains in CesA proteins as a class-defining feature

Earlier, Vergara and Carpita (2001) proposed that HVRII regions from CESA proteins should be renamed as class-specific regions (CSR). Based on 26 HVRII regions from various plant species available at that time, they observed class-specific sequence conservation among these regions. They proposed that HVRII is not a hypervariable region as initially proposed by Pear et al. (1996), but that each type of HVRII actually defines a specific class of CESAs in a plant species and HVRII/CSRs contain conserved motifs important for catalysis.

Available sequence information about HVRII regions, including the current work, has now been obtained for 56 HVRII regions, and a compilation and analysis of these HVRII regions provides an opportunity to reaffirm the suggestion of renaming HVRII as CSR. In addition, the first algal HVRII sequence has become available (Roberts et al. 2002) for rooting the phylogenetic tree. Vergara and Carpita (2001) showed that the topology of the phylogenetic tree is mainly determined by HVRII regions and such topology largely remained unchanged when full-length CESA proteins were included in the construction of the tree. We, therefore, developed a rooted phylogenetic tree of all plant CESAs based on amino acid sequences of HVRII regions with a bootstrap value of 1000 and a strong support of > 70% as shown in Figure 4. The HVRII from the green algae, *Mesotaenium caldariorum* CESA (Roberts et al. 2002) was used as the outgroup.

The most distinctive feature of this tree is the presence of only six classes of CESAs that are represented in various plant species studied so far. Aspen is the only non-*Arabidopsis* species that has representatives of all six classes of CESAs that are currently available for further research of their functionality. Some members of three CESA clades (shown by blue lines in Figure 4) are associated with primary cell wall development in *Arabidopsis* (Arioli et al. 1998, Fagard et al. 2000, Burn et al. 2002), whereas some other members of the remaining three clades (shown by red lines in Figure 4) are associated with sec-

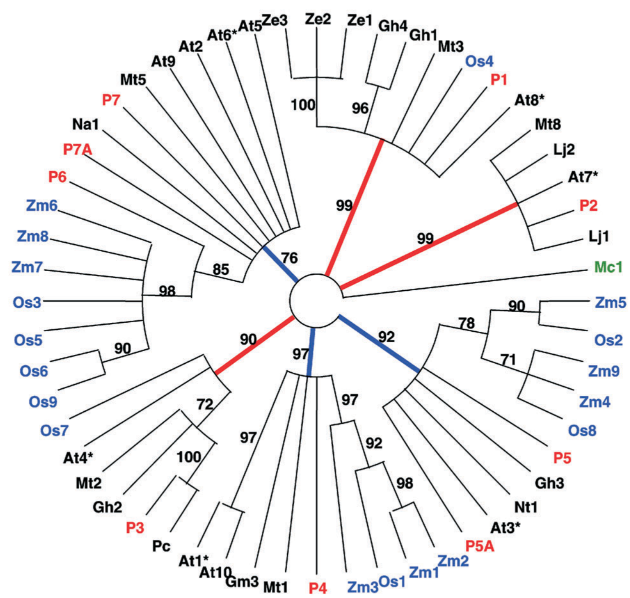

Figure 4. Phylogenetic tree derived with PAUP program based on 56 cellulose synthase (CESA) HVRII regions from plants. Bootstrap analysis was conducted with 1000 replicates and the bootstrap values of > 70 were considered for the development of the rooted tree using green algal CESA from *Mesotaenium caldariorum* HVRII as an outgroup (shown in green). HVRII domains from all CESA proteins used here were downloaded from the Stanford site http://cellwall.stanford.edu and were renamed by eliminating their CESA extension to simplify the figure. Abbreviations: At = *Arabidopsis thaliana* (L.) Heynh.; Gh = *Gossypium hirsutum* L.; Gm, *Glycine max* (L.) Merill; Mc = *Mesotaenium caldariorum* (Lagerh.) Hansg.; Na = *Nicotiana alata* Link & Otto; Mt = *Medicago truncatula* Gaertn.; Nt, *Nicotiana tabacum* L.; Os = *Oryza sativa* L.; Pc = *Populus canescens* (Ait.) Sm.; Ze = *Zinnia elegans* Jacq.; and Zm = *Zea mays* L.. The following GenBank accession numbers for aspen or some *CesA* genes that are currently missing in the protein collection at the Stanford site were used to deduce the protein sequences included in this figure: P1 = *PtrCesA1*, AF072131; P2 = *PtrCesA2*, AY095297; P3=*PtrCesA3*, AF527387; P4 = *PtrCesA4*, AY162181; P5 = *PtrCesA5*, AY055724; P5A = *PtrCesA5*-like AY330165; P6 = *PtrCesA6,* AY196961; P7 = *PtrCesA7*, AY162180; P7A = *PtrCesA7-like* AY330166; Nt1 = *NtCesA1*, AF233892; and Mc1 = *McCesA1*, AF525360. Red lines indicate the clades where some *Arabidopsis* CESAs (marked by asterisks) are implicated in secondary cell wall synthesis and blue lines indicate the clades where some *Arabidopsis* CESAs (marked by asterisks) are implicated in primary cell wall synthesis. All aspen (*Populus tremuloides*) CESA HVRIIs are indicated in red (P1–P7A) and monocot CESA HVRIIs are shown in blue.

ondary cell wall development in *Arabidopsis* (Taylor et al. 1999, 2000, 2003). Moreover, gene expression analyses have confirmed the CESA associations with primary and secondary cell wall development in non-*Arabidopsis* species (Pear et al. 1996, Holland et al. 2000, Wu et al. 2000, Samuga and Joshi 2002, Kalluri and Joshi 2003). Thus, HVRII regions (which could be renamed CSRs as suggested by Vergara and Carpita 2001) could be conveniently used for diagnosing primary and secondary cell wall-associated CESAs. In each clade, monocot CESAs, when available (shown in blue), form a separate subgroup suggesting that structural evolution of CESAs has continued after divergence of monocots from dicots. It has been suggested that some of these cereal CESAs may also be functionally associated with mixed-linkage β-glucan synthesis (Dhugga 2001, Vergara and Carpita 2001).

Discussion

The structural details of plant *CesA* genes and cDNAs became available only after 1996 and we have already accumulated a great deal of information (see recent reviews by Delmer 1999, Dhugga 2001, Joshi 2003*a*, 2003*b*, 2004). The *in silico* or computer-assisted search of the available plant genome/EST databases with cotton and other plant *CesA*s has resulted in the prediction of many putative *CesA* genes and ESTs *(*http://cellwall.stanford.edu/). However, detailed information is available for only a few full-length *CesA* cDNAs from trees. Thus, isolation of full-length *CesA* cDNAs from economically important trees remains a major future research goal. Isolation of full-length tree *CesA* cDNAs spanning ~3.5 kb is still a difficult task, requiring refined techniques to isolate and characterize them.

Here we report on the use of a simple RT-PCR-based technique that is followed by single-pass sequencing of a randomly selected but limited number of clones (19–20 in each of the four cases) for identification of several HVRII regions from *CesA*/*CslD* genes in the *P. tremuloides* genome. We believe that use of such degenerate HVRII primers designed on the basis of a large number of *CesA* genes will also allow application of this method to other uncharacterized plant species of interest. These techniques are especially necessary for softwood trees such as loblolly pine, radiata pine and spruce that have great economic value. However, because of the large size of these genomes, it is unlikely that the complete genome sequence will be available in the near future. Similarly, many hardwood trees such as eucalypts, acacia and other poplar species are also good candidates for genetic improvement of cellulose biosynthesis where detailed understanding of the functionality of the entire complement of *CesA*s is currently lacking. All plants studied so far host a large *CesA* gene family and the status of the *CesA* gene family should be upgraded to superfamily if it includes *Csl* genes encoding evolutionarily and possibly functionally related proteins (Richmond 2000). Systematic dissection of functions of these genes in any tree species and understanding the key differences among *CesA*s of different trees will help answer the question why plants have so many *CesA*/*Csl* genes.

A well-annotated genome is currently available only for *Arabidopsis*, but will soon become available for *P. trichocarpa* (Wullschleger et al. 2002*a* and 2002*b*), providing an unprecedented opportunity for a direct comparison of the members of the *CesA* superfamily in these diverse species, which will help decipher the contributions of *CesA*/*CslD* sequence variations to synthesis of polysaccharides in cell walls. Our initial survey suggests that *Arabidopsis* and aspen share similar classes of *CesA* and *CslD* genes, but amino acid sequences of orthologs of these species differ by > 20% amino acids (Figure 3). In *Arabidopsis*, even a single base pair mutation in the coding region of the *CesA* gene impacts the process of cellulose biosynthesis (Joshi 2003*a*). Therefore, a > 20% difference in the amino acid sequences between *Arabidopsis* and aspen CESA orthologs raises the possiblity of many differences in the process of cellulose synthesis between these species. Use of the RT-PCR technique followed by single-pass sequencing of HVRII regions as described here may facilitate such comparisons in the near future.

There are certain shortcomings to the RT-PCR approach used here. In aspen, not all *CesA* and *CslD* genes could be isolated using RNA from two types of tissues. For example, HVRII region from *PtrCesA2* that we reported earlier from an aspen xylem cDNA library (Samuga and Joshi 2002) was not represented even though we used xylem RNA as a template for RT-PCR. It may be necessary to sequence more randomly selected HVRII clones to obtain the entire coverage of *CesA*/*CslD* HVRII regions in these tissues. Furthermore, isolating from additional tissues may be necessary to obtain all tissue-specific *CesA*/*CslD* members from the aspen genome. We also believe that the numbers of sequences we found in each group as shown in Figure 2 do not reflect the actual transcript abundances in the respective tissue, but rather the PCR conditions used (e.g. re-annealing temperatures). The transcript of *PtrCesA1* is one of the most abundant in aspen xylem (Wu et al. 2000), but was represented here by only one clone, whereas another abundantly expressed gene, *PtrCesA3* was represented by 17 clones, all from the xylem (U. Kalluri and C.P. Joshi, unpublished data). Similarly, *PtrCesA7* is represented by 28 clones here, but our in situ mRNA localization data for this gene suggested that it is weakly expressed in primary-cell-wall-enriched cells, tissues and organs (A. Samuga and C.P. Joshi, unpublished data).

Why are there only six classes of CESA in higher plants? This information fits well with what is known in *Arabidopsis* where three classes of CESA have been associated with primary cell wall development and the other three classes with secondary cell wall development (Joshi 2003*b*). The sixfold symmetry of cellulose synthesizing rosette complexes in plants also suggests that six CESA subunits may somehow form the basis of qualitative and quantitative differences in cellulose produced in primary and secondary cell walls (Doblin et al. 2002). Previously, we proposed that three primary CESAs may form heteromeric rosettes in primary-cell-wall-forming cells, whereas the other three secondary CESAs may form another type of rosette in secondary walls of tissues such as xylem and sclerenchyma (Joshi 2004). This hypothe-

sis, if correct, could have an enormous impact on the design of future genetic improvement strategies of cellulose biosynthesis in trees.

Acknowledgments

We thank Drs. Vincent L. Chiang and Glenn D. Mroz for their constant support and encouragement. We are grateful to Ms. Anita Samuga, Ms. Udaya Kalluri and Mr. Priya Ranjan for their assistance. We also thank Dr. Dana L. Richter for critical reading of this manuscript. This research was partially supported by NSF-CAREER award (IBN-0236492), USDA-NRI grant (99-35103-7986), the USDA-McIntire Stennis Forestry Research Program and Research Excellence Fund from the State of Michigan to CPJ.

References

Adams M.D., M.B. Soares, A.R. Kerlavage, C. Fields and J.C. Venter. 1993. Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. Nature Genet. 4:373–380.

Arioli, T., L. Peng, A.S. Betzner et al. 1998. Molecular analysis of cellulose biosynthesis in *Arabidopsis*. Science 279:717–720.

Bradshaw, H.D., R. Ceulemans, J. Davis and R.F. Stettler. 2000. Emerging model systems: Poplar (*Populus*) as a model forest tree. J. Plant Growth Reg. 19:306–313.

Brown, J.W.S. 1986. A catalogue of splice junction and putative branch point sequences from plant introns. Nucleic Acids Res. 14: 9549–9559.

Burn, J., C.H. Hocart, R.J. Birch, A.C. Cork and R.E. Williamson. 2002. Functional analysis of the cellulose synthase genes *CesA1*, *CesA2* and *CesA3* in *Arabidopsis*. Plant Physiol. 129:1–11.

Carpita, N. and M. McCann. 2000. The cell wall. *In* Biochemistry and Molecular Biology of Plants. Eds. B.B. Buchanan, W. Gruissem and R.L. Jones, ASPP Press, Rockville, pp 52–108.

Delmer, D.P. 1999. Cellulose biosynthesis: exciting times for a difficult field of study. Annu. Rev. Plant Physiol. Plant Mol. Biol. 50: 245–276.

Dhugga, K.S. 2001. Building the wall: genes and enzyme complexes for polysaccharide synthases. Curr. Opin. Plant Biol. 4:488–493.

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

Doblin, M.S., I. Kurek, D. Jacob-Wilk and D.P. Delmer. 2002. Cellulose biosynthesis in plants: from genes to rosettes. Plant Cell Physiol. 43:1407–1420.

Fagard, M., T. Desnos, T. Desprez et al. 2000. *PROCUSTE1* encodes a cellulose synthase required for normal cell elongation specifically in roots and dark-grown hypocotyls of *Arabidopsis*. Plant Cell 12:2409–2424.

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

Haigler, C. and R.L. Blanton. 1996. New hopes for old dreams: evidence that plant cellulose synthase genes have finally been cloned. Proc. Natl. Acad. Sci. 93:12,082–12,085.

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

Huang, X. and A. Madan. 1999. CAP3: a DNA sequence assembly program. Genome Res. 9:868–877.

Joshi, C.P. 2003*a*. Molecular biology of cellulose biosynthesis in plants. *In* Recent Research Developments in Plant Molecular Biology. Ed. S. Pandalai. Research Signpost Press, Kerala, India, pp 19–38.

Joshi, C.P. 2003*b*. Xylem-specific and tension stress responsive expression of cellulose synthase genes from aspen trees. Appl. Biochem. Biotech. 105:17–26.

Joshi, C.P. 2004. Molecular genetics of cellulose biosynthesis in trees. *In* Molecular Genetics and Breeding of Forest Trees. Eds. S. Kumar and M. Fladung, Haworth Press, New York. In press.

Kalluri, U. and C.P. Joshi. 2003. Isolation and characterization of a new, full-length cellulose synthase cDNA from developing xylem of aspen trees. J. Exp. Bot. 54:2187–2188.

Laosinchai, W., X. Cui and R.M. Brown. 2000. A full length cDNA of cotton cellulose synthase has high homology with the *Arabidopsis RSW1* gene and the cotton *CelA1* gene (accession co. AF200453) (PGR00-002). Plant Physiol. 122:291.

Pear, J.R., Y. Kawagoe, W.E. Schreckengost, D.P. Delmer and D.M. Stalker. 1996. Higher plants contain homologs of the bacterial *CelA* genes encoding the catalytic subunit of cellulose synthase. Proc. Natl. Acad. Sci. 93:12,637–12,642.

Perrin, R.M. 2001. Cellulose: how many cellulose synthases to make a plant? Current Biol. 11:R213–R216.

Richmond, T.A. 2000. Higher plant cellulose synthases. Genome Biol. 1:3001.1–3001.6.

Roberts, A.W., E.M. Roberts and D.P. Delmer. 2002. Cellulose synthase (*CesA*) genes in the green alga *Mesotaenium caldariorum*. Eukaryot. Cell 1:847–855.

Samuga, A. and C.P. Joshi. 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.

Samuga A. and C.P. Joshi. 2004. Cloning and characterization of cellulose synthase-like gene, *PtrCSLD2* from developing xylem of aspen trees. Physiol. Plant. In press.

Saxena, I.M., R.M. Brown, M. Fevre, R.A. Geremia and B. Henrissat. 1995. Multidomain architecture of β-glycosyltransferases: implications for mechanism of action. J. Bacteriol. 177:1419–1424.

Taylor, N.G., W.R. Scheible, S. Cutler, C.R. Somerville and S.R. Turner. 1999. The *irregular xylem3* locus of *Arabidopsis* encodes a cellulose synthase required for secondary cell wall synthesis. Plant Cell 11:769–779.

Taylor, N.G., S. Laurie and S.R. Turner. 2000. Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in *Arabidopsis*. Plant Cell 12:2529–2540.

Taylor N.G., R.M. Howells, A.K. Huttly, K. Vickers and S.R. Turner. 2003. Interactions among three distinct CesA proteins essential for cellulose synthesis. Proc. Natl. Acad. Sci. 100:1450–1455.

Vergara, C.E. and N.C. Carpita. 2001. β-D-glycan synthases and the *CesA* gene family. Plant Mol. Biol. 47:145–160.

Wang, H. and C. Loopstra. 1998. Cloning and characterization of a cellulose synthase cDNA (accession no. AF081534) from xylem of hybrid poplar (*Populus tremula* × *Populus alba*). (PGR98-179) Plant Physiol. 118:1101.

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

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

Wullschleger, S.D, G.A. Tuskan and S.P. DiFazio. 2002*b*. Editorial: genomics and the tree physiologist. Tree Physiol. 22:1273–1276.