Research review

Genomics of cellulose biosynthesis in poplars

Author for correspondence:
Chandrashekhar P. Joshi
Tel: +1 906 4873480
Fax: +1 906 4872915
Email: cpjoshi@mtu.edu

Received: *4 April 2004*
Accepted: *1 June 2004*

Chandrashekhar P. Joshi, Suchita Bhandari, Priya Ranjan, Udaya C. Kalluri, Xiaoe Liang, Takeshi Fujino and Anita Samuga

Plant Biotechnology Research Center, School of Forest Resources and Environmental Science, Michigan Technological University, Houghton, Michigan 49931, USA

Key words: cellulose biosynthesis, cellulose synthase, korrigan cellulase, poplars, sucrose synthase, wood development.

Summary

Genetic improvement of cellulose production in commercially important trees is one of the formidable goals of current forest biotechnology research. To achieve this goal, we must first decipher the enigmatic and complex process of cellulose biosynthesis in trees. The recent availability of rich genomic resources in poplars make *Populus* the first tree genus for which genetic augmentation of cellulose may soon become possible. Fortunately, because of the structural conservation of key cellulose biosynthesis genes between *Arabidopsis* and poplar genomes, the lessons learned from exploring the functions of *Arabidopsis* genes may be applied directly to poplars. However, regulation of these genes will most likely be distinct in these two-model systems because of their inherent biological differences. This research review covers the current state of knowledge about the three major cellulose biosynthesis-related gene families from poplar genomes: cellulose synthases, sucrose synthases and korrigan cellulases. Furthermore, we also suggest some future research directions that may have significant economical impacts on global forest product industries.

New Phytologist (2004) **164**: 53–61

 doi: 10.1111/j.1469-8137.2004.01155.x

Introduction

Cellulose, an unbranched polymer of β-1,4 linked glucose residues, is a deceptively simple – though abundant – macromolecule with myriad practical applications. Although cellulose biosynthesis is one of the ancient biological processes that distinguish plants from animals, the precise molecular mechanism of cellulose biosynthetic process in plants is still not clearly understood (Read & Bacic, 2002). It is paradoxical that, despite the ubiquitousness, simple molecular structure and ancient origin of cellulose, we still know only very little about how plants synthesize it. This knowledge can have many significant economical and ecological repercussions (Delmer, 1999).

A number of plant systems, each with its own unique features, have been used to decipher the intricacies of this important bioprocess of carbon sequestration. For example, developing cotton fibers are the single cell extensions that contain nearly pure cellulose in their secondary cell walls. This is achieved by massive reprogramming of cellulose biosynthetic pathway. The cotton system was recently used to successfully clone the first cellulose synthase (*CesA*) gene from plants (Pear *et al.*, 1996). This delay in cloning of a first plant *CesA* gene in the modern genomics era may have stemmed

from extremely labile nature of the cellulose synthase complexes which, even today, resist successful isolation.

The availability of abundant genomic resources in *Arabidopsis* provided another model system for identification of a family of up to 10 *CesA* genes and assisted in clarifying the functions of at least six *CesA* genes by mutant analyses (Richmond, 2000). Interestingly, three distinct *CesA* genes appear to be involved in primary cell wall development (*AtCesA1*, *AtCesA3* and *AtCesA6*), and the other three *CesA* genes are associated with secondary cell wall development (*AtCesA4*, *AtCesA7* and *AtCesA8*) in *Arabidopsis* (for review see: Delmer, 1999; Doblin *et al.*, 2002; Joshi, 2003a). Taylor *et al.* (2003) recently showed that AtCESA4, AtCESA7 and AtCESA8 proteins interact with each other, thereby supporting a number of hypotheses about the roles of cellulose synthase (CESA) proteins in the formation of cellulose synthesizing complexes (i.e. rosettes with sixfold symmetry; Perrin, 2001; Doblin *et al.*, 2002; Joshi, 2003b).

Using polyclonal antibodies raised against catalytic domain of cotton CESA proteins, Kimura *et al.* (1999) recently immunolocalized them at the rosette complex in plasma membranes of another plant system, that of the mung bean. In so doing, they provided critical physical evidence of CESA and rosette involvement in cellulose biosynthesis. A number of recent reviews have covered developments in cellulose biosynthesis in model and crop plants (e.g. Delmer, 1999; Dhugga, 2001; Haigler *et al.*, 2001; Doblin *et al.*, 2002; Williamson *et al.*, 2002; Joshi, 2003a,b). Therefore, the focus of this review is on the genomics of cellulose biosynthesis in poplar tree systems.

Genetic improvement of cellulose biosynthesis in trees is a major research priority. A prerequisite for genetic augmentation of cellulose biosynthesis in forest trees is a clear understanding of the cellulose biosynthetic processes in trees, including the functional elucidation of key genes and a determination of their spatiotemporal regulation during tree development. Although trees provide major raw materials for forest product industries, tree systems are rarely used to understand the molecular mechanisms of cellulose biosynthesis. Among other reasons, this is because of the paucity of *CesA* and cellulose biosynthesis related genes from trees. With the recent advances in sequencing of the poplar genome (Wullschleger *et al.*, 2002), this situation should change rapidly. The purpose of this review is to summarize the current state of knowledge regarding cellulose biosynthesis in poplar trees. The term 'poplar' encompasses species of the genus *Populus* and includes aspens, cottonwoods and poplars (Bradshaw *et al.*, 2000). Moreover, cellulose is not merely some inert material banished outside the cell wall but, rather, is a dynamic medium that changes its constitution depending on environmental cues and changes. Therefore, explorations of the cellulose biosynthesis of trees offer some novel opportunities and difficult challenges compared with herbaceous plants.

Current models envision plant cellulose biosynthesis to be at least a three step process: (1) Plasma membrane-associated sucrose synthase (SUSY) directly channel the UDP-glucose substrate to cellulose synthesizing machinery; (2) coordinately expressed multiple CESA, organized in the form of hexagonal rosettes, polymerize glucose monomers into glucan chains while recycling liberated UDP back to SUSY; and (3) a membrane-associated cellulase, KORRIGAN (KOR), acts as an editor or monitor of glucan chain conversion to cellulose microfibril, cleaving the defective glucan chains (Delmer & Haigler, 2002; Molhoj *et al.*, 2002) (Fig. 1). Thus, interplay of at least CESA, SUSY and KOR proteins appears to regulate cellulose biosynthesis in plants (Read & Bacic, 2002). Although several candidate genes such as callose synthase, cellulose synthase-like (CSL), α-glucosidase, RAC13, actin and tubulin have been indirectly implicated in cellulose biosynthetic process in plants, their role in the formation of cellulose synthase complex is still being explored and debated (Doblin *et al.*, 2002). Therefore, this review will focus only on *CesA*, *SuSY* and *Kor* gene families from poplars.

Molecular Cloning of *CesA* Genes from Poplars

Owing to its immense economic importance, we are interested in understanding the process of cellulose biosynthesis in trees with special reference to wood development. Most tree cellulose

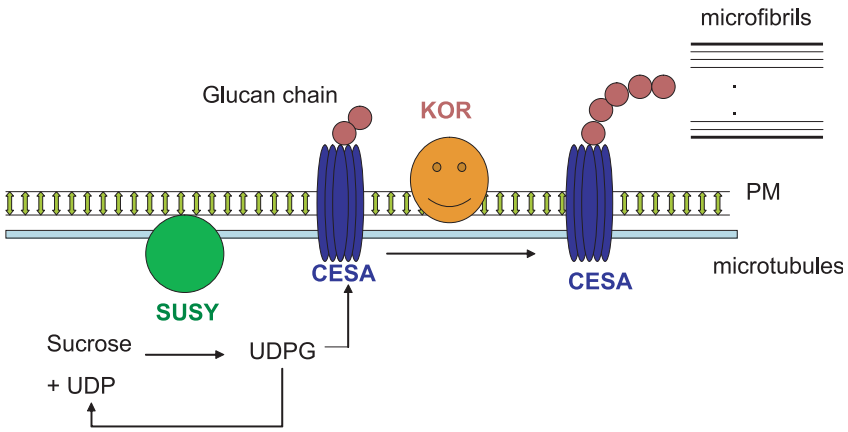

Fig. 1 Current model of cellulose biosynthesis in plants. Plasma membrane (PM)-associated form of sucrose synthase (SUSY) directly channels Uridine diphosphate glucose (UDPG) substrate to cellulose synthase (CESA) rosette complex that aid in glucan chain formation while recycling UDP back to SUSY. The CESA complexes may be moving across the PM with the help of microtubules. Glucan chains self assemble into microfibrils and KORRIGAN cellulase (KOR) acts as an editor/monitor of this process.

research has been concerned with the describing the end-product of cellulose biosynthesis rather than understanding the mechanism of cellulose biosynthesis. For example, the nature of cellulose in primary and secondary cell walls is highly distinct and heterogeneous in poplars (Mellerowicz *et al.*, 2001). Secondary walls in poplar xylem have a higher percentage of cellulose, a higher degree of polymerization, a higher crystallinity, and a lower microfibril angle than do xylem primary walls. These secondary wall cellulose qualities are the most desirable wood quality traits for forest product industries.

These variations in the quantity and quality of primary and secondary wall cellulose in plants were suspected to be a result of enzymatic activities of different types of CESAs (Haigler & Blanton, 1996) but direct confirmation of this relationship in trees was not possible because of the lack of information about the structural features and functional assignments of tree *CesA* genes. We therefore began with the isolation of *CesA* genes from aspen (*Populus tremuloides*) xylem to learn more about the cellulose biosynthesis in this economically important woody tree.

We first cloned a xylem-specific and tension stress responsive *CesA*, *PtCesA* (now known as *PtrCesA1*) gene from quaking aspen (Wu *et al.*, 2000). The only other previously reported full-length poplar *CesA* cDNA, *Pxcel1* (now known as *PcCesA1*) from hybrid poplar (*Populus tremula* × *Populus alba*) was distinctly different from *PtrCesA1* cDNA (64% identity). Although *PcCesA1* was the first full-length *CesA* cDNA reported from any poplar species, very little information about *PcCesA1* beyond cDNA sequence is available even now (Wang & Loopstra, 1998).

We subsequently cloned another six new and distinct full-length aspen *CesA* cDNAs that ranged from *c.* 3.3–3.8 kb in size, and designated them as *PtrCesA2* to *PtrCesA7* (Table 1). Of these, *PtrCesA1*, *PtrCesA2*, and *PtrCesA3* appeared to be secondary wall *CesA*s because of their high degree of amino acid sequence similarity (over 85%) with proteins encoded by *AtCesA8*, *AtCesA7* and *AtCesA4*, respectively, which are associated with secondary wall development in *Arabidopsis*. *PtrCesA3* is highly homologous to *PcCesA1* (98% identity) from hybrid poplar (Wang & Loopstra, 1998). By contrast, *PtrCesA4*, *PtrCesA5*, *PtrCesA6*, and *PtrCesA7* shared a high degree of identity with the primary wall-related *CesAs* from *Arabidopsis*, *AtCesA1*, *AtCesA3* and *AtCesA6*. Because of the importance of secondary cell walls in determining wood quality traits in aspen wood, we have first focused our attention on aspen secondary *CesA*s. The detailed descriptions of molecular characteristics of *PtrCesA1* to *PtrCesA7* cDNAs have been recently published (Wu *et al.*, 2000; Samuga & Joshi, 2002; Joshi, 2003b; Kalluri & Joshi, 2003; Kalluri, 2003; Samuga & Joshi, 2004a).

Table 1 shows the comparison of seven aspen *CesA* cDNAs, which confirms their distinctness from each other. It also illustrates the high level of identity/similarity between aspen CESAs and their corresponding *Arabidopsis* CESA orthologs. The predicted proteins of these seven aspen *CesA*s have similar protein structure: two transmembrane domains (TMDs) near the *N*-terminus and six TMDs towards the *C*-terminus. This structure is similar to all other known CESA proteins (Fig. 2). Comparison of PtrCESA proteins also indicated that the first highly diverged (hypervariable) region, HVRI resides in the cytoplasmic *N*-terminal region immediately following the putative zinc-binding domain that has recently been shown to be involved in CESA protein–protein interactions (Kurek *et al.*, 2002). The putative catalytic domain is present between TMD 2 and TMD 3 and is predicted to be on the cytoplasmic side.

All signature motifs present in most of the typical processive glycosyltransferases (Saxena *et al.*, 1995) are also present in this globular catalytic domain. A second hypervariable region,

Table 1 Comparison of molecular characteristics of seven full-length *CesA* cDNAs from aspen

	1	2	3	4	5	6	7
Length (bp)	3232	3277	3401	3640	3532	3773	3809
Coding region (bp)	2934	3096	3123	3249	3234	3261	3288
Length of protein (aa)	978	1032	1041	1083	1078	1087	1096
Molecular weight	110 279	116 958	118 262	121 953	119 940	122 614	123 914
Highly similar to	*AtCesA8*	*AtCesA7*	*AtCesA4*	*AtCesA1*	*AtCesA3*	*AtCesA6*	*AtCesA6*
(% aa similarity)	(87)	(91)	(86)	(91)	(89)	(81)	(91)
Arabidopsis mutant	irx1	irx3	irx5	rsw1	ixr1	prc1	prc1

1, PtrCesA1; 2, PtrCesA2; 3, PtrCesA3; 4, PtrCesA4; 5, PtrCesA5; 6, PtrCesA6; 7, PtrCesA7; aa, amino acid; bp, basepair.

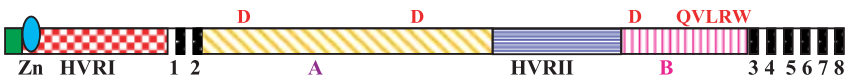

Fig. 2 A diagrammatic representation of PtrCESA proteins. Various domains are indicated by a brief description below the diagram. Zn, Zinc-binding domain; HVRI, hypervariable region near *N*-terminus; 1–8, transmembrane domains; Subdomains A and B, highly conserved (80–90%) part of catalytic domains in relation to other CESA proteins; HVRII, central hypervariable region. Location of the processive glycosyltransferase motif D, D, D; QVLRW is shown above.

	Overall							HVRI						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	–	64	68	65	64	66	64	–	27	25	24	28	23	25
2		–	69	70	71	72	70		–	36	43	46	40	47
3			–	68	67	69	68			–	42	37	42	43
4				–	71	70	67				–	49	46	39
5					–	70	68					–	43	40
6						–	76						–	57
7							–		–					–

1, PtrCesA1; 2, PtrCesA2; 3, PtrCesA3; 4, PtrCesA4; 5, PtrCesA5; 6, PtrCesA6; 7, PtrCesA7.

Table 2 Homology relationship (in percentage identity) among seven aspen CESA proteins and their various domains

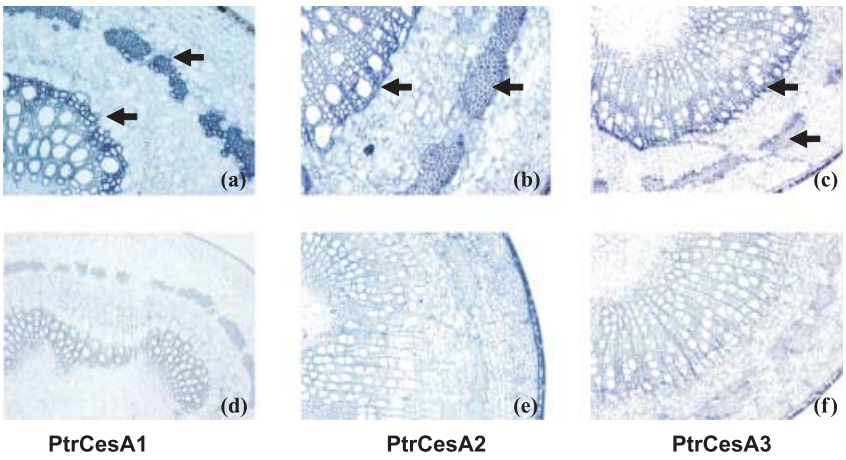

Fig. 3 *In situ* mRNA hybridization with *PtrCesA1*, *PtrCesA2*, and *PtrCesA3*-specific HVRI digoxygenin (DIG) labeled riboprobes. (a–c) Expression of *PtrCesA1*, *PtrCesA2* and *PtrCesA3* mRNA in developing xylem and phloem fibers (indicated by arrows) visualized by hybridization with antisense riboprobes, respectively. (d–f) Absence of hybridization using sense riboprobes for *PtrCesA1*, *PtrCesA2* and *PtrCesA3* HVRI regions, respectively (as expected).

HVRII, further interrupts the cytoplasmic catalytic domain into highly conserved subdomains A and B (Pear *et al.*, 1996; Wu *et al.*, 2000; Joshi, 2003b). Moreover, comparison of seven full-length aspen CESA proteins and their HVRI domains also clearly suggested that these proteins were distinct from each other and the cDNA regions corresponding to HVRI domains might be suitable probes for studying the differential expression of these seven *CesA* genes in aspen tissues (Table 2).

The high level of similarity among three secondary CESAs from *Arabidopsis* and aspen also strongly suggests that three aspen CESAs could be coexpressed in same cells and might even be the partners in secondary cell wall biogenesis in aspen xylem. This is similar to the available data from *Arabidopsis* (Taylor *et al.*, 1999, 2000, 2003). Similarly, high homology of primary aspen CESAs with the known primary CESAs from *Arabidopsis* also suggests that these four aspen CESAs could also be essential partners in primary cell wall synthesis. We therefore began exploring the spatiotemporal expression patterns of these seven aspen genes in various aspen tissues using a variety of molecular techniques.

HVRI Domain Based *in situ* Hybridization of *CesAs* in Aspen Tissues

All three aspen secondary *CesA* genes appear to be members of different small multigene families, as judged by Southern blot analysis using HVRI regions, and are expressed in secondary wall-enriched xylem tissues as evidenced by Northern blot analysis and reverse transcriptase polymerase chain reaction (RT-PCR). They are expressed in xylem tissues in young stem internodes (< 4), and in developing xylem and phloem fiber tissues in older internodes (> 8) as observed by *in situ* mRNA hybridization using gene-specific cDNA regions corresponding to HVRI regions (Fig. 3a–c). The sense probes did not show such hybridization patterns in all three cases as expected (Fig. 3d–f). These three secondary *CesA* genes thus appear to be coordinately expressed in same developing xylem and phloem fiber tissues in the internodes that are undergoing secondary growth.

What about the expression patterns of primary *CesA* genes? In *Arabidopsis*, seven out of 10 *CesA* genes are suggested to be primary wall-related *CesA*s (for review see Doblin *et al.*, 2002; Joshi, 2003b). A systematic comparative expression analysis of these seven genes in various cells/tissues/organs of *Arabidopsis* is currently missing but highly desirable. Thus, we do not clearly know how many of these seven CESAs are coexpressed and therefore, possibly necessary for cellulose biosynthesis in primary walls. Study of *CesA* mutants has suggested that *AtCesA1*, *AtCesA3* and *AtCesA6* are required for cellulose biosynthesis but we do not know which of the remaining primary CESAs can functionally replace the defective CESAs in these *Arabidopsis* mutants. It is possible that some of these CESAs

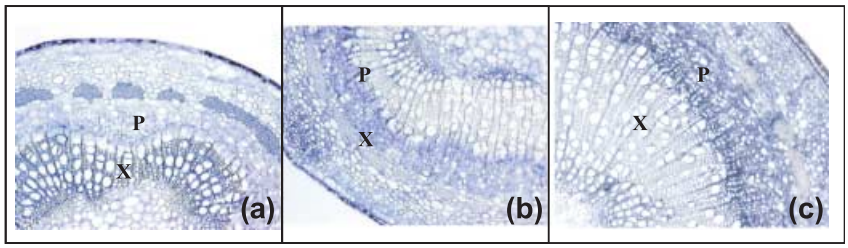

Fig. 4 *In situ* mRNA hybridization with *PtrCesA4*, *PtrCesA5* and *PtrCesA7*-specific HVRI DIG labeled riboprobes. (a–c) Expression of *PtrCesA4*, *PtrCesA5* and *PtrCesA7* mRNA as visualized by hybridization with antisense riboprobes, respectively. X, xylem; P, phloem.

perform redundant functions while others are either uniquely important or not important at all (Burn *et al.*, 2002; Williamson *et al.*, 2002; Beeckman *et al.*, 2003).

Recently, we have also started working with primary *CesA*s in aspen using *in situ* hybridization with HVRI-specific RNA probes. Using similar stem tissue sections as described above for secondary *CesA*s, our research showed that primary *CesA*s were expressed in all living expanding cells (Fig. 4). We used antisense RNA probes from the regions corresponding to HVRI of three primary *CesA*s, *PtrCesA4*, *PtrCesA5*, and *PtrCesA7* (Fig. 4). Interestingly, *PtrCesA4* showed greater expression in xylem ray cells, *PtrCesA5* transcripts were more abundant in phloem cells and *PtrCesA7* was expressed both in developing cambium and phloem cells. As expected, no hybridization was detected using the sense control probes from these primary *CesA* genes (data not shown). Thus, different aspen primary *CesA*s are expressed differentially in the stem tissues compared with expression of secondary *CesA*s. These expression patterns also suggest that primary *CesA*s may need more attention in the future and we should trace their complex expression patterns throughout the life cycle of aspen trees to determine which primary *CesA*s are uniquely important and which ones perform redundant functions during various phases of life cycle of aspen.

Exon/Intron Structure of *CesA* Genes from the Poplar Genome

A recently completed *Arabidopsis* genome sequence provided significant information about the exon/intron structure of its 10 *CesA* genes. All *Arabidopsis CesA* genes are split, have *c.* 3.5–5.5 kb long coding regions and contain 10–13 small introns (Richmond, 2000). Thus, efficient and accurate intron splicing is essential for successful production of CESA proteins.

Genome sequencing of black cottonwood (*Populus trichocarpa*) is in progress (Wullschleger *et al.*, 2002). Although the fully annotated genome information is not yet available, the Joint Genome Institute (JGI) recently released approx. 4.0 gigabases of raw genome sequence data (http://genome.jgi-psf.org/poplar0/poplar0.info.html). Even though *P. trichocarpa* (the species used for genome sequencing) and *P. tremuloides* (the species we are studying) are different, we expect that both these species will share similar genes because of their close relationship (Bradshaw *et al.*, 2000). For example, we cloned 5′ ends of aspen *CSLD* genes based on the sequence information available from the black cottonwood genome (Samuga & Joshi, 2004b) and we also predicted the exon/intron structure of HVRII regions of black cottonwood *CesA* genes using aspen HVRII regions (Liang & Joshi, 2004). Using the BLASTN program (Altschul *et al.*, 1997), we searched over approx. 5.5 million trace files (each with *c.* 400–500 bp of useful sequence data released by the JGI in September 2003) with each of the seven full-length aspen *CesA* cDNAs that are currently available in our laboratory. We assembled the resulting hundreds of hits into complete genes by using a CAP3 sequence assembly program (Huang & Madan, 1999), which produces a set of contigs by searching for the perfect overlapping sequences. The final assembled contigs were used as input for standard gene prediction algorithms (http://genes.mit.edu/GENSCAN.html), and the output was manually examined for the accuracy of computer predictions. Figure 5 shows the most probable gene structure of seven poplar genes that are homologous to aspen *CesA* cDNAs. These predictions will be reconfirmed once fully annotated poplar genome becomes available. Poplar *CesA* genes range from 4.8 to 7.2 kb in size and possess 11–14 introns. Their *Arabidopsis CesA* counterparts also have similar gene structure with similar length of the coding regions and similar number of introns in each gene. However, in most cases, poplar *CesA* introns are much longer than *Arabidopsis CesA* introns. If this feature is also conserved in other non-*CesA* genes, it may partly explain why poplar genomes are five times larger than the *Arabidopsis* genome.

We also examined the occurrence of poplar expressed sequence tags (ESTs) homologous to aspen *CesA* genes. Currently, a total of 155 608 DNA sequences from 32 poplar species are available in the GenBank. The majority of these sequences are contributed by ongoing EST projects from all over the world. These projects perform single-pass sequencing of each of the randomly selected clones. Although EST sequences from a variety of species, tissues, and treatments contribute to this large number, a search of these EST databases with available full-length *CesAs* will provide some general idea of transcript abundance of *CesA* genes in poplars. Search of EST databases with *PtrCesA1-PtrCesA7* yielded a total of 198 hits from poplars. Of these, 69% ESTs are highly similar to secondary wall-related *PtrCesA1-PtrCesA3*, and 31% ESTs are highly similar to Ptr*CesA4-PtrCesA7*. As expected from this expression analysis, primary wall-related *CesA*s are expressed at a much lower level than secondary wall-related *CesA*s. After the poplar genome sequencing and annotation is completed,

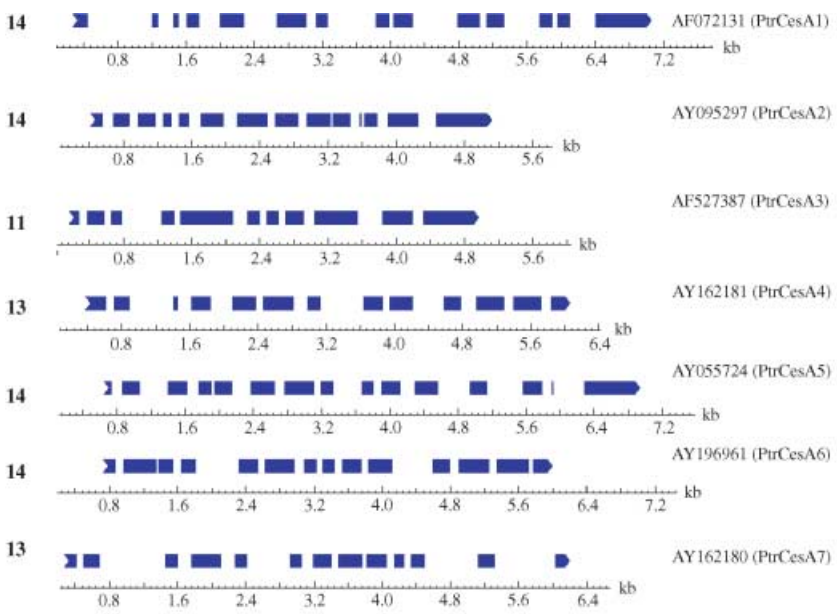

Fig. 5 *CesA* gene structure predictions based on currently available poplar genome data. The aspen *CesA* cDNA used as input sequence for searching poplar genome data is shown on the right in the parenthesis following its respective GenBank numbers. The number of predicted exons is shown on the left side of each gene in bold letters. Each gene begins with an initiator exon, followed by internal exons, and ends with a terminator exon. The blank spaces in between exons represent introns. The scale of the gene length in kilobase pairs (Kb) is shown below each gene.

a careful quantitative expression study of individual primary *CesA* genes must be undertaken to understand the relative expression levels and possible importance of these genes in cellulose biosynthesis in various poplar cells.

Very recently, during the review of this manuscript, five full-length *CesA* cDNAs from hybrid poplar (*Populus tremula* × *Populus tremuloides*) became publicly available (GenBank accession #AY573571–AY573575). The encoded proteins of these hybrid poplar *CesA* cDNA sequences, designated as PttCESAs, are 65–98% identical to seven aspen *CESA*s described above. The PttCESA3-1 and PttCESA3-2 are 96% identical to each other and are 94–97% identical to aspen PtrCESA1. Of the remaining three hybrid poplar CESA proteins, PttCESA1 is 98% identical to aspen PtrCESA3, PttCESA2 is 84% identical to aspen PtrCESA7 and PttCESA4 is 98% identical to aspen PtrCESA7. It is expected that more information about the expression patterns of hybrid poplar *CesA*s will become available in the near future.

Molecular Cloning of Sucrose Synthase Genes from Poplars

Cellulose biosynthesis is a highly regulated biological process by which plants irreversibly sequester carbon in their expanding primary and thickening secondary cell walls. Such cellulose in plant cell walls is a major sink for carbon derived from photosynthetic processes using the greenhouse gas, CO_2 (Delmer & Haigler, 2002). Sucrose, the main carbon source for cellulose synthesis, is translocated to the sink cells or drawn away from photosynthetic cells. Carbon present in sucrose is made available for cellulose biosynthesis via the sucrolytic phase of enzymes SUSY (EC 2.4.1.13; sucrose + UDP ↔ UDP-glucose + fructose) and invertase (EC 3.2.1.26; sucrose + H_2O ↔ glucose + fructose). Of these two, SUSY conserves energy in the glycosidic bonds without the use of an additional ATP molecule to generate UDP-glucose (Haigler *et al.*, 2001).

A key player in regulating this flux of incoming carbon into the irreversible sink of cellulose thus appears to be SUSY. The detection of a membrane associated form of SUSY, P-SUSY, from developing cotton fibers gave rise to a revolutionary idea that the membrane association of SUSY facilitates the direct channeling of carbon to cellulose synthesis (Amor *et al.*, 1995). It was thus hypothesized that the P-SUSY could be part of the cellulose synthase complexes (i.e. rosettes) such that the enzymatic product of SUSY activity, UDP-glucose, which is a substrate of cellulose synthesis, is made available efficiently to rosettes (Amor *et al.*, 1995). Such an association would have the advantages of avoiding competition from the cellular metabolic pool of UDPG and overall less ATP utilization as stated before (Amor *et al.*, 1995).

Downregulation of *SuSy* genes has also been associated with the reduction in crystalline cellulose synthesis in transgenic plants and such outcomes confirm that it could be one of the limiting steps of cellulose biosynthesis (Tang & Strum, 1999; Haigler *et al.*, 2001). Interestingly, these plants showed evidence of compensatory regulation between cellulose and pectin synthesis similar to tobacco *CesA* downregulation, as reported by Burton *et al.* (2000). Thus, a complex feedback mechanism in carbon partitioning is most likely present in plants. A number of recent reviews have covered the details of the role of SUSY in cellulose biosynthesis (Amor *et al.*, 1995; Haigler *et al.*, 2001; Delmer & Haigler, 2002). We will therefore focus only on what is known about *Populus SuSy* genes.

We recently cloned the first full-length aspen *SuSy* sequence (*PtrSuSy1*) from a xylem cDNA library (Kalluri, 2003)

(GenBank accession #AY341026). This 2792-bp long sequence encodes a protein of 805 amino acids. PtrSUSY1 protein showed the highest sequence identity with the citrus SUSY (87%) and cotton SUSY (86%). This suggests that these genes might be structurally and possibly functionally conserved. Both citrus and cotton SUSY have been shown to play a pivotal role in cellulose biosynthesis during secondary cell wall formation (Amor *et al.*, 1995; Komatsu *et al.*, 2002).

SuSy gene structure analysis using raw sequence data from the poplar genome, as described earlier in this review, suggested that poplar *SuSy* gene contains 14 exons and 13 introns and the first AUG codon occurs in the exon 2. The length of each exon is identical to *Arabidopsis AtSus1* gene but the intron length seems to be much longer in poplars than in *Arabidopsis* (Baud *et al.* 2004). Using a highly conserved region from *PtrSuSy1*, we performed *in situ* hybridization with various aspen tissues where HVRI region from xylem-specific *PtrCesA1* was used for comparison (Kalluri, 2003). While *PtrCesA1* transcripts were specifically localized in the developing xylem tissues, *PtrSuSy1* probe hybridized with the same xylem tissues as well as other cells with developing primary walls. The GenBank currently contains 208 poplar ESTs similar to *PtrSuSy1* cDNA. Thus, *SuSy* appears to be universally required and available for cellulose biosynthesis in all developing and growing cells in poplars. Recent global expression analysis using ESTs from wood-forming tissues of poplars (Hertzberg *et al.*, 2001) also showed a strong expression of *SuSy* ESTs in the secondary wall development zone of woody tissues. Thus, *SuSy* expression appears to be intimately associated with cellulose synthesis at least during wood formation in *Populus* and genetic upregulation of the *SuSy* gene, in addition to simultaneous upregulation of three secondary wall-associated *CesA*s, may enhance cellulose production in woody tissues of trees.

A Special Cellulase may be Involved in Cellulose Biosynthesis in Poplars

Recently, the occurrence of a strange cooperation between cellulose synthases that synthesize cellulose and cellulases that normally digest cellulose was discovered that appears to be necessary for successful cellulose biosynthesis (Molhoj *et al.*, 2002). Read and Bacic (2002) have even suggested that a special type of cellulase, KOR, could be a part of the rosettes that make cellulose in plants, although recent experimental data by Szyjanowicz *et al.* (2004) does not support such speculation.

Nicol *et al.* (1998) first discovered a special cellulase or endo-1,4-β-glucanase (EGase; EC 3.2.1.4) from *Arabidopsis* that is essential for correct assembly of walls in elongating cells. A *korrigan* (*kor*) mutation occurring in this special EGase gene, designated as *Kor*, resulted in extremely dwarf phenotype with defective primary cell walls. Zuo *et al.* (2000) further confirmed the role of KOR proteins during cell plate formation. The *Arabidopsis Kor* gene was cloned by chromosome walking and successfully used for complementation of *kor* mutant restoring the wild-type phenotype. *Arabidopsis Kor* gene encoded a unique membrane-anchored KOR protein of 622 residues containing eight potential *N*-glycosylation sites. The *N*-terminal stretch of 100 amino acids is uniquely conserved only in this subfamily of EGases from *Arabidopsis*, rice, *Brassica* and tomato (Molhoj *et al.*, 2002). In a phylogenic tree of EGases, KOR enzymes form a separate clade that lacks cleavable signal peptide needed for targeting to ER but possess a transmembrane domain at the *N*-terminus, suggesting that it might be an integral membrane protein. Sato *et al.* (2001), Lane *et al.* (2001), and His *et al.* (2001) independently showed that, in three allelic mutants (*acw*, *rsw2* and *kor*), the cellulose content of primary walls was severely reduced. Thus, the role of *Kor* gene in cellulose synthesis of at least primary wall synthesis was confirmed.

Sato *et al.* (2001) further noted that their *acw* mutant showed accumulation of short lipid-linked β-1,4 glucans that could be intermediate products of cellulose biosynthesis. Finally, Peng *et al.* (2002) deserve the credit for connecting all these facts and proposing a coherent picture of cellulose biosynthesis in plants. They provided evidence for the role of sitosterol glucoside (SG) as a primer for glucan chain elongation. When labeled SG was added to membrane fractions from cotton fibers, it was directly incorporated into newly formed cellulose. Since expression levels of a *Kor* homolog and its encoded proteins present in the cotton fiber membranes were also correlated with the cellulose synthesis activity, they proposed that the role of KOR may be to cleave the SG primer from the growing glucan chains. The only problem with this proposal is that the catalytic domain of KOR is predicted to be outside the plasma membranes while catalytic domains of CESA proteins are located towards cytoplasmic side. Thus, more experimental proof is required to confirm this model.

Molhoj *et al.* (2002) proposed an alternative scenario (depicted in Fig. 1) in which KOR functions as an editor of elongating glucan chain outside the plasma membrane by removing defective glucan chains from the microfibril assembly. In either case, KOR does play some important role in cellulose biosynthesis, although the exact mechanism needs to be deciphered. Doblin *et al.* (2002) have also discussed at length the speculative nature of SG primer model. Therefore, how KOR participates in cellulose biosynthesis is still an open question. All *Arabidopsis Kor* mutations studied to date were associated with primary wall formation but, from the wood industry perspective, secondary wall is more important. Szyjanowicz *et al.* (2004) recently associated mutated *Kor* gene as a cause of *irx2* mutation in *Arabidopsis* that exhibits cellulose deficiency in the secondary wall of xylem cells, resulting in irregular xylem cells. Although a number of *Kor* genes from dicotyledon species have so far been reported, none from tree species have yet been available.

Therefore, we recently cloned the first *Kor* gene from aspen xylem cDNA library (S. Bhandari & C. P. Joshi, unpubl. data).

The aspen *Kor*, *PtrKor1* cDNA is 2563-bp long with 263 bp of 5′ leader and 423 bp of 3′ untranslated region (GenBank accession #AY535003). The coding region of 1860 bp encodes a protein of 620 amino acids which shares 74–86% identity and 79–90% similarity with other KOR proteins known to date. It has one predicted transmembrane domain near the *N*-terminus and many conserved motifs such as two putative polarized targeting sequences, LL and YXXΦ in the cytoplasmic tail and two cellulase signature motifs near the *C*-terminus (Zuo *et al.*, 2000).

Using currently available poplar genome data, we also predicted the gene structure of *Kor* gene from *P. trichocarpa*. The poplar *Kor* gene contains six exons that are almost identical in size to *Arabidopsis Kor1* gene (Zuo *et al.*, 2000) but the five poplar introns are about four times larger than *Arabidopsis Kor* introns. Thus, the translatable information content appears to be conserved between these two species. Gene expression analysis with RT-PCR and *in situ* hybridization has suggested that this gene is upregulated in xylem tissues compared with young leaf tissues. Search for poplar ESTs with aspen *Kor* cDNA as query also resulted in 73 hits, suggesting that this gene is commonly expressed in a variety of tissues from various poplar species. Park *et al.* (2003) recently overexpressed another poplar EGase (not a *Kor*) gene in *Arabidopsis* that resulted in larger transgenic plants. We are now overexpressing aspen *Kor* cDNA under constitutive promoter and xylem-specific promoter in transgenic aspen and tobacco to study the effects of such genetic manipulation on biomass production as well as cellulose biosynthesis. In view of the importance of KOR in cellulose biosynthesis, simultaneous upregulation of *Kor*, *SuSy* and *CesA* genes in transgenic poplar trees may result in increased biomass as well as improved production of wood with better quality cellulose that is typically present in the secondary walls of xylem in woody tissues.

Future Perspectives

Over the past six or seven years remarkable progress has been made towards the cloning and characterization of many key genes involved in cellulose biosynthesis of poplars. Despite the major differences in the biology and life cycle of *Arabidopsis* and poplars, both these groups share similar genes in their genomes although the patterns of expression of these genes are most likely to be different. For example, poplars develop tension wood on the upper side of their bent stem (Wu *et al.*, 2000). Tension wood fibers are enriched in almost pure cellulose (98.5%). It will be interesting to see which of these cellulose biosynthesis genes from poplars are upregulated during tension wood formation. Such experiments with the poplar system are not possible with the *Arabidopsis* system. Secondary growth and wood development are two other natural traits of economic importance that can be better explored using the poplar system. Finally, knowledge gained about the roles of *SuSy*, *CesA* and *Kor* genes in the morphogenesis and development of *Arabidopsis* plants could be extrapolated to tree systems.

Poplars would be the first targets of such genetic augmentation of cellulose biosynthesis. Simultaneous upregulation of key cellulose biosynthesis related genes may result in genetically improved trees of great significance to forest product industries. These trees would sequester even more carbon in the form of cellulose than is naturally occurring today and that would simultaneously help the environment by reducing greenhouse gases that threaten our future.

Acknowledgements

We are grateful to our colleagues Drs Glenn Mroz, Vincent Chiang, Laigeng Li, Chung-Jui Tsai, Shawn Mansfield and Scott Harding for their help and support in our research related to cellulose biosynthesis in aspen. We also thank Dr Valorie Troesch for critical reading of the manuscript. Research in the Joshi laboratory is partly 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. Finally, we would like to dedicate this article to Professor Debby Delmer on receiving the 2003 Anselme Payen award of the American Chemical Society for her pioneering research in the field of cellulose biosynthesis in plants.

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. *Nucleic Acids Research* **25**: 3389–3402.

Amor Y, Haigler CH, Johnson S, Wainscott M, Delmer DP. 1995. A membrane-associated form of sucrose synthase and its potential role in synthesis of cellulose and callose in plants. *Proceedings of the National Academy of Sciences, USA* **92**: 9353–9357.

Baud S, Vaultier M-N, Rochat C. 2004. Structure and expression profile of the sucrose synthase multigene family in *Arabidopsis*. *Journal of Experimental Botany* **55**: 397–409.

Beeckman T, Przemeck GKH, Stamatiou G, Lau R, Terryn N, de Rycke R, Inzé D, Berleth T. 2003. Genetic complexity of CesA gene function in *Arabidopsis* embryogenesis. *Plant Physiology* **130**: 1883–1893.

Bradshaw HD, Ceulemans R, Davis J, Stettler RF. 2000. Emerging model systems: poplar (*Populus*) as model forest tree. *Journal of Plant Growth Regulation* **19**: 306–313.

Burn JE, Hocart CH, Birch RJ, Cork AC, Williamson RE. 2002. Functional analysis of the cellulose synthase genes CesA1, CesA2 and CesA3 in *Arabidopsis*. *Plant Physiology* **129**: 797–807.

Burton RA, Gibeaut DM, Bacic A, Findlay K, Roberts K, Hamilton A, Baulcombe D, Fincher GB. 2000. Virus-induced silencing of a plant cellulose synthase gene. *Plant Cell* **12**: 691–706.

Delmer DP. 1999. Cellulose biosynthesis: exciting times for a difficult field of study. *Annual Reviews in Plant Physiology and Plant Molecular Biology* **50**: 245–276.

Delmer DP, Haigler CH. 2002. The regulation of metabolic flux to cellulose, a major sink for carbon in plants. *Metabolic Engineering* **4**: 22–28.

Dhugga KS. 2001. Building the wall: genes and enzyme complexes for polysaccharide synthesis. *Current Opinions in Plant Biology* **4**: 488–493.

Doblin M, Kurek I, Jacob-Wilk D, Delmer DP. 2002. Cellulose biosynthesis in plants: from genes to rosettes. *Plant Cell Physiology* **43**: 1407–1420.

Haigler C, Blanton RL. 1996. New hopes for old dreams: Evidence that plant cellulose synthase genes have finally been cloned. *Proceedings of National Academy of Sciences, USA* **93**: 12082–12085.

Haigler CH, Ivanova-Datcheva M, Hogan PS, Salnikov VV, Hwang S, Martin K, Delmer DP. 2001. Carbon partitioning to cellulose synthesis. *Plant Molecular Biology* **47**: 29–51.

Hertzberg M, Aspeborg H, Schrader J, Andersson A, Erlandsson R, Blomqvist K, Bhalerao R, Uhlén M, Teeri TT, Lundeberg J, Sundberg B, Nilsson P, Sandberg G. 2001. A transcriptional roadmap to wood formation. *Proceedings of the National Academy of Sciences, USA* **98**: 14732–14737.

His I, Driouich A, Nicol F, Jannuau A, Hofte H. 2001. Altered pectin composition in primary cell walls of *korrigan* a dwarf mutant of *Arabidopsis*. *Planta* **212**: 348–358.

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

Joshi CP. 2003a. Molecular biology of cellulose biosynthesis in plants. In: Pandalai R, ed. *Recent research developments in plant molecular biology.*, vol. 1. Kerala, India: Research Signpost, 19–38.

Joshi CP. 2003b. Xylem-specific and tension stress responsive expression of cellulose synthase genes from aspen trees. *Applied Biochemistry and Biotechnology* **105**: 17–26.

Kalluri UC. 2003. *Comparative expression analysis of cellulose biosynthesis related genes from aspen trees.* PhD Thesis. Houghton, MI, USA: Michigan Technological University.

Kalluri UC, Joshi CP. 2003. Isolation and characterization of a new, full-length cellulose synthase cDNA, *PtrCesA5* from developing xylem of aspen trees. *Journal of Experimental Botany* **54**: 2187–2188.

Kimura S, Laosinchai W, Itoh T, Cui X, Linder CR, Brown RM. 1999. Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant *Vigna angularis*. *Plant Cell* **11**: 2075–2085.

Komatsu A, Moriguchi T, Koyama K, Omura M, Akihama T. 2002. Analysis of sucrose synthase genes in citrus suggests different roles and phylogenetic relationships. *Journal of Experimental Botany* **53**: 61–71.

Kurek I, Kawagoe Y, Jacob-Wilk D, Doblin M, Delmer D. 2002. Dimerization of cotton fiber cellulose synthase catalytic subunits occurs via oxidation of the zinc-binding domains. *Proceedings of the National Academy of Sciences, USA* **99**: 11109–11114.

Lane D, Wiedemeier A, Peng L, Hofte H, Vernhettes S, Desprez T, Hocart C, Birch R, Baskin T, Burn J, Arioli T, Betzner A, Williamson R. 2001. Temperature sensitive alleles of *rsw2* link the Korrigan glucanase to cellulose synthesis and cytokinesis in *Arabidopsis*. *Plant Physiology* **126**: 278–288.

Liang X, Joshi CP. 2004. Molecular cloning of ten distinct hypervariable regions from cellulose synthase gene superfamily in aspen trees. *Tree Physiology* **24**: 543–550.

Mellerowicz EJ, Baucher M, Sundberg B, Boerjan W. 2001. Unraveling cell wall formation in the woody dicot stem. *Plant Molecular Biology* **47**: 239–274.

Molhoj M, Pagant S, Hofte H. 2002. Towards understanding the role of membrane bound endo-β-glucanases in cellulose biosynthesis. *Plant Cell Physiology* **43**: 1399–1406.

Nicol F, His I, Jauneau A, Vernhettes S, Canut H, Hofte H. 1998. A plasma membrane-bound putative endoglucanase is required for normal cell wall assembly and cell elongation. *EMBO Journal* **17**: 5563–5576.

Park YW, Tominaga R, Sugiyama J, Furuta Y, Tanimoto E, Samejima M, Sakai F, Hayashi T. 2003. Enhancement of growth by expression of poplar cellulase in *Arabidopsis thaliana*. *Plant Journal* **33**: 1099–1106.

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. *Proceedings of the National Academy of Sciences, USA* **93**: 12637–12642.

Peng L, Kawagoe Y, Hogan P, Delmer D. 2002. Sitosterol-beta-glucoside as primer for cellulose synthesis in plants. *Science* **295**: 147–150.

Perrin RM. 2001. Cellulose: how many cellulose synthases to make a plant? *Current Biology* **11**: R213–R216.

Read SM, Bacic T. 2002. Prime time for cellulose. *Science* **295**: 59–60.

Richmond T. 2000. Higher plant cellulose synthases. *Genome Biology* **1**: 1–6.

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.

Samuga A, Joshi CP. 2004a. Expression patterns of two primary cell wall-related cellulose synthase cDNAs, *PtrCesA6* and *PtrCesA7* from aspen. *Gene* **334**: 71–80.

Samuga A, Joshi CP. 2004b. Cloning and characterization of cellulose synthase-like gene, *PtrCSLD2* from developing xylem of aspen trees. *Physiologia Plantarum* **120**: 631–641.

Sato S, Kato T, Kakegawa K, Ishii T, Liu Y, Awano T, Takabe K, Nishiyama Y, Kuga S, Sato S, Nakamura Y, Tabata S, Shibata D. 2001. Role of putative membrane bound endo 1-4 glucanase korrigan in cell elongation. *Plant Cell Physiology* **42**: 251–263.

Saxena IM, Brown RM, Fevre M, Geremia RA, Henrissat B. 1995. Multidomain architecture of β-glycosyltransferases: Implications for mechanism of action. *Journal of Bacteriology* **177**: 1419–1424.

Szyjanowicz PMJ, McKinnon I, Taylor N, Gardiner J, Jarvis MC, Turner SR. 2004. The irregular xylem 2 mutant is an allele of korrigan that affects the secondary cell wall of *Arabidopsis thaliana*. *Plant Journal* **37**: 730–740.

Tang GQ, Strum A. 1999. Antisense repression of sucrose synthase in carrot (*Daucus carota* L.) affects growth rather than sucrose partitioning. *Plant Molecular Biology* **41**: 465–479.

Taylor NG, Howells RM, Huttly AK, Vickers K, Turner SR. 2003. Interactions among three distinct CesA proteins essential for cellulose synthesis. *Proceedings of the National Academy of Sciences, USA* **100**: 1450–1455.

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

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 H, Loopstra C. 1998. Cloning and characterization of a cellulose synthase cDNA (accession, AF081534) from xylem of hybrid poplar (*Populus tremula* × *Populus alba*). *Plant Physiology* **118**: 1101.

Williamson RE, Burn JE, Hocart CH. 2002. Towards the mechanism of cellulose synthesis. *Trends in Plant Science* **7**: 461–467.

Wu L, Joshi CP, Chiang V. 2000. A xylem-specific cellulose synthase gene from aspen (*Populus tremuloides*) is responsive to mechanical stress. *Plant Journal* **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.

Zuo J, Niu Q, Nishizawa N, Wu Y, Kost B, Chua N-H. 2000. Korrigan, an *Arabidopsis* endo-1,4-β-glucanase, localizes to the cell plate by polarized targeting and is essential for cytokinesis. *Plant Cell* **12**: 1137–1152.