Planta (2005) 221: 747–751
DOI 10.1007/s00425-005-1563-z

PROGRESS REPORT

**Weidong Yong · Bruce Link · Ronan O'Malley
Jagdish Tewari · Charles T. Hunter · Chung-An Lu
Xuemei Li · Anthony B. Bleecker · Karen E. Koch
Maureen C. McCann · Donald R. McCarty
Sara E. Patterson · Wolf-Dieter Reiter · Chris Staiger
Steven R. Thomas · Wilfred Vermerris
Nicholas C. Carpita**

Genomics of plant cell wall biogenesis

Received: 15 July 2004 / Accepted: 1 April 2005 / Published online: 25 June 2005

Keywords *Arabidopsis thaliana* · *Zea mays* · Cell wall · Mutants · Polysaccharides · Lignin · Biosynthesis · Genomics · Infrared spectroscopy · Chemometrics · Artificial neural networks

Plant cell walls are composed of independent but interacting networks of carbohydrates, proteins, and aromatic substances. Interacting with this complex matrix are several hundred enzymes and other proteins, some with known functions in wall assembly but many of unknown function. Tentative annotations of the Arabidopsis genome reveal more than 1000 genes encoding cell-wall-related proteins encompassing precursor-generating enzymes, synthases and glycosyl transferases, structural proteins, and a host of enzymes involved in polysaccharide modification and depolymerization (Carpita et al. 2001). An equal number of unannotated genes are suspected of encoding proteins that function in wall assembly or disassembly. Proteomics approaches that select for polypeptides secreted to the exterior of the plasma membrane reveal sequences of many proteins whose function is completely unknown (Borderies et al. 2003). Complete genome sequences of model fungi, animals and plants provide a powerful means to identify proteins that are unique to plants (Gutiérrez et al. 2004). Other strategies to identify unannoted genes as wall-related include database profiling of the expression of genes closely associated with genes of known function (Perrin et al. 2001) and global expression of genes associated with secondary wall induction during trans-differentiation of Zinnia mesophyll cells into tracheary elements in vitro (Milioni et al. 2002).

Two types of walls are found among angiosperms. For extensive review of the differences between grass species and all other angiosperms with respect to polymer structure, architecture, and biosynthesis, see McCann and Roberts (1991) and Carpita and Gibeaut (1993). Briefly, "Type I" walls, which are found in dicots and the non-commelinoid monocots, contain about equal amounts of cellulose and cross-linking xyloglucans, with minor amounts of arabinoxylans, glucomannans, and galacto-glucomannans. This strong framework is embedded in a pectin matrix of homogalacturonans and rhamnogalacturonan I that controls several physiological properties, such as wall porosity, charge density, and microfibril spacing. A variant of homogalacturonan, called rhamnogalacturonan II, is decorated with side-groups containing many rare sugars and linkage structures and forms boron di-diester dimers that function in

W. Yong · N. C. Carpita (✉)
Department of Botany and Plant Pathology,
Purdue University, 915 West State Street,
West Lafayette, Indiana 47907-2054, USA
E-mail: carpita@purdue.edu
Tel.: +1-765-4944653
Fax: +1-765-4940363

B. Link · X. Li · W.-D. Reiter
Department of Molecular and Cell Biology,
University of Connecticut, Storrs, Connecticut, 06269–3125

R. O'Malley · A. B. Bleecker
Department of Botany, University of Wisconsin, Madison,
Wisconsin, 53706

S. E. Patterson
Department of Horticulture, University of Wisconsin,
Madison, Wisconsin, 53706

J. Tewari · M. C. McCann · C. Staiger
Department of Biological Sciences, Purdue University,
West Lafayette, Indiana, 47907–2054

C. T. Hunter · C.-A. Lu · K. E. Koch · D. R. McCarty
Horticultural Sciences Department, University of Florida,
Gainesville, Florida, 32611

S. R. Thomas
National Renewable Energy Laboratory,
National Bioenergy Center, Golden,
Colorado, 80401

W. Vermerris
Departments of Agronomy and Agricultural
and Biological Engineering, Purdue University,
West Lafayette, Indiana, 47907–2054

control of porosity and tensile strength. Grasses and other commelinoid monocots possess a "Type II" wall (Carpita and Gibeaut 1993). They contain cellulose microfibrils of the same structure as those of the Type I wall, but glucuronoarabinoxylans are the major microfibril tethering molecules. In the Poales mixed-linkage (1 → 3),(1 → 4)-β-D-glucans are synthesized specifically during cell expansion. In general, grass walls are pectin-poor and have very little structural protein compared to Type I walls. Instead, they possess extensive interconnecting networks of phenylpropanoids that form primarily when cells stop expanding.

New genetic resources of Arabidopsis and maize cell-wall mutants

A powerful way to probe function of the individual components of the cell wall is through identification and characterization of mutants (Reiter et al. 1997). Given the two distinct types of cell walls that angiosperms make, two distinct genetic models are required, a dicot with Type I walls, such as Arabidopsis, and a grass species with Type II walls. The goal of our "Cell Wall Genomics" project is to provide a characterization of the cell wall phenotypes of homozygous cell-wall mutants of Arabidopsis and maize for use by the cell wall community and other interested researchers (http://cellwall.genomics.purdue.edu). Large-scale insertional DNA lines for both species are now publicly available. The Salk Institute resource of T-DNA insertional mutants was developed with the goal of providing a mutant in every single gene of Arabidopsis (Alonso et al. 2003). A robotics-assisted selection of homozygous, single lines of the Salk Institute's SIGnAL collection is being developed for widespread distribution. Additional insertional lines have become or will soon be available in which flanking sequences are used to guide the researcher to the gene of interest.

The Uniform *Mu* maize population developed at the University of Florida provides a comparable resource of tagged mutants in genes of interest in a highly uniform background (http://uniformmu.org). Robertson's *Mutator* elements introgressed into a consistent inbred genetic background, W22, gives a mutagenized population of maize for which forward screens and reverse genetics approaches can be used to identify cell-wall-related genes. Random sequencing of *Mu*-tagged genes provides a growing list of these genes. Furthermore, DNA-grids representing nearly 20,000 mutated lines are being screened by PCR for potential maize cell-wall genes. This approach has identified mutations in genes representing many families of enzymes associated with substrate generation, polysaccharide synthesis, and wall remodeling.

Mutations in known cell-wall-related genes give a library of spectrotypes

Cell wall biogenesis encompasses many distinct but interactive biochemical pathways and mechanisms. We have grouped genes that function in the synthesis, assembly, and disassembly of the plant cell wall into six categories, but because these stages are tightly integrated, some proteins and enzymes may function in several of them. We are currently establishing infrared spectral phenotypes, called "spectrotypes", for cell walls of mutants representing most of the following.

Substrate generation

The substrates for polysaccharide synthesis are nucleotide-sugars, which are made de novo through the nucleotide-sugar interconversion pathways (Reiter and Vanzin 2001). Mutations that affect the synthesis of a monosaccharide are expected to give more complicated spectrotypes than mutations in glycosyltransferase genes because every polysaccharide containing that sugar will be altered.

Polysaccharide synthesis

Genes encoding synthases and glycosyl transferases that polymerize monosaccharides into complex polysaccharides have been grouped into several families (Coutinho et al. 2003). Interactions between three different isoforms of CESA polypeptides appears to be required for primary and secondary wall cellulose synthesis in Arabidopsis, rice, and other species (Taylor et al. 2003). Several dozen related genes, called 'cellulose-synthase-like' (*CSL*) genes, likely function in the synthesis of other cross-linking glycans (Richmond and Somerville 2001; Hazen et al. 2002). Those related to the Arabidopsis *CslA* class have been shown to encode (1 → 4)-β-mannan and glucomannan synthases (Dhugga et al. 2004; Liepmann et al. 2005). The glycosyl transferases (GTs) involved in the synthesis of branched polysaccharides fall into numerous classes, families and clades (Coutinho et al. 2003). Through mutants, GTs involved in decoration of xyloglucans and complex pectins have been characterized.

The secretory pathway

The cytoskeleton plays a major role in the delivery mechanism of polysaccharide-laden vesicles from the Golgi membranes to particular sites at the plasma membrane. Several mutants in secretory pathway-related genes result in cell wall abnormalities. The primary defect in these mutants is in motor proteins and other cytoskeleton-associated proteins (Zhong et al. 2002), in *N*-glycosylation and trimming pathways (Gillmor et al. 2002), or in vesicle targeting (Lukowitz et al. 1996).

Wall assembly and dynamics

Gene families involved in the assembly and architectural rearrangements of the wall include α- and β-expansins

(Cosgrove et al. 2002) and xyloglucan endotransglucosylase/hydrolases (XTHs, Rose et al. 2002). Structural proteins are integrated with polysaccharide components during wall assembly. The phenotype of the *root-shoot-hypocotyl (rsh)* mutant, which has a defective structural extensin gene, results in grossly altered plantlet form that is ultimately lethal (Hall and Cannon 2002). Plants with Type I and II walls have also developed different repertories of hydrolases that function in disassembly of their different cell wall polysaccharides (Hrmova and Fincher 2001; Coutinho et al. 2003).

Lignin biosynthesis

Monolignol precursors of lignin are generated from phenylalanine by a large variety of enzymes, including hydroxylases, methyltransferases, and dehydrogenases (Boerjan et al. 2003). Lignin is then formed through oxidative coupling of monolignols within the apoplastic space. The availability of whole-genome sequences (Raes et al. 2003), along with detailed chemical analyses in Arabidopsis (Franke et al. 2002) and maize (Marita et al. 1999), have contributed to a better understanding of lignin biosynthesis.

Signaling and response systems in the cell wall

Several membrane surface proteins and proteoglycans, including the glycosylphosphatidylinositol (GPI)-anchored AGPs (Gaspar et al. 2001) and a large family of transmembrane receptor-like kinases (Shiu and Bleecker 2001), interact with the cell wall. Curiously, broad classes of secreted proteins are predicted to be GPI-anchored (Eisenhaber et al. 2003).

High through-put screening for cell wall mutants

Fourier transform infrared (FTIR) microspectroscopy is an extremely rapid, non-destructive vibrational spectroscopy that can quantitatively detect a range of functional groups that provides a composite "fingerprint" of carbohydrate constituents and their organization (McCann et al. 2001). Mid-infrared spectra of cell walls identify particular functional groups, including esters, amides, hydroxyls, and carboxylates, and a carbohydrate fingerprint region (Fig. 1). The spectrum of a cell wall also contains information that reflects the local environment of molecular bonds, such as their hydration state, the conformation of the molecule, and interactions with other molecules (McCann et al. 2001). Therefore, the spectrum is characteristic of the architecture as well as the composition of the wall, providing the basis of a spectrotype characteristic for that mutant.

We have made the cell wall mutant screen high through-put. Purified walls from two-week-old Arabidopsis plantlets and etiolated 5-day-old maize coleoptiles are placed in the wells of 48-well gold-coated microscope slides, mounted on a computer-driven stage of an infrared microscope. Five infrared spectra from each of the wells of the slide are collected by an automated mapping program (Fig. 1). Acquisition of each spectrum requires less than one minute. We also use near-infrared (NIR) reflectance spectroscopy, with a spectral window between 400 nm and 2500 nm, as a high through-put screen for mutations in secondary walls of maize under field growth conditions (Fig. 1). As we generate homozygous lines, the seed stocks will become available to the plant research community at ABRC and the Maize Genetics Cooperation Stock Center.

We have shown how statistical algorithms, such as Principal Components Analysis (PCA) and Linear Discriminant Analysis (Kemsley 1998), can be applied to FTIR spectra of cell walls as broadly applicable methods to identify otherwise "invisible" defects in wall structure and architecture (Chen et al. 1998). This allows the natural variation within a population to be accounted for while identifying sources of variance arising as a consequence of mutation (Fig. 1). While PCA can identify and classify spectrotypes that are characteristic of cellulose deficiencies or alterations in xyloglucan or pectin structure, artificial neural networks are used in order to refine the classification of spectrotypes in large sample sets (Almeida 2002). A neural network comprises three layers: an input layer, which comprises the absorbance values of infrared spectra, a "hidden" layer of about 50 neurons, and an output layer, which are the discrete mutant classes to which each individual might belong. The weightings of each input are adjusted for each neuron as the network is "trained" by mutant and wild-type standards until it makes the minimum number of errors in classification to a specific output. When dealing with mutants for which we do not know if a spectrotype exists, "unsupervised" Kohonen networks detect clustering structure in the data set without having to assign specific output classes and provide probability values for class membership. The probability of being classified into particular classes establishes similarity relationships between different mutants, sometimes to a mutant with a well-characterized cell-wall phenotype. These data indicate the biochemical defect in mutants and target further chemical characterization.

A direct relationship between genotype and spectrotype is not expected

Characterizing and classifying the functions of previously unannotated genes is a formidable task. Quite diverse gene products can alter wall architecture in similar ways, for example, cellulose deficiency can result from mutations in an endoglucanase, the cobra protein, and a mannose-1-phosphate guanylyltransferase, as well as from mutations in the catalytic subunits of cellulose synthase (McCann and Carpita 2005). Plants are known to alter gene expression in several pathways to compensate for defects in cell wall structure caused by mutation. Alterations in cell walls can also be an indirect

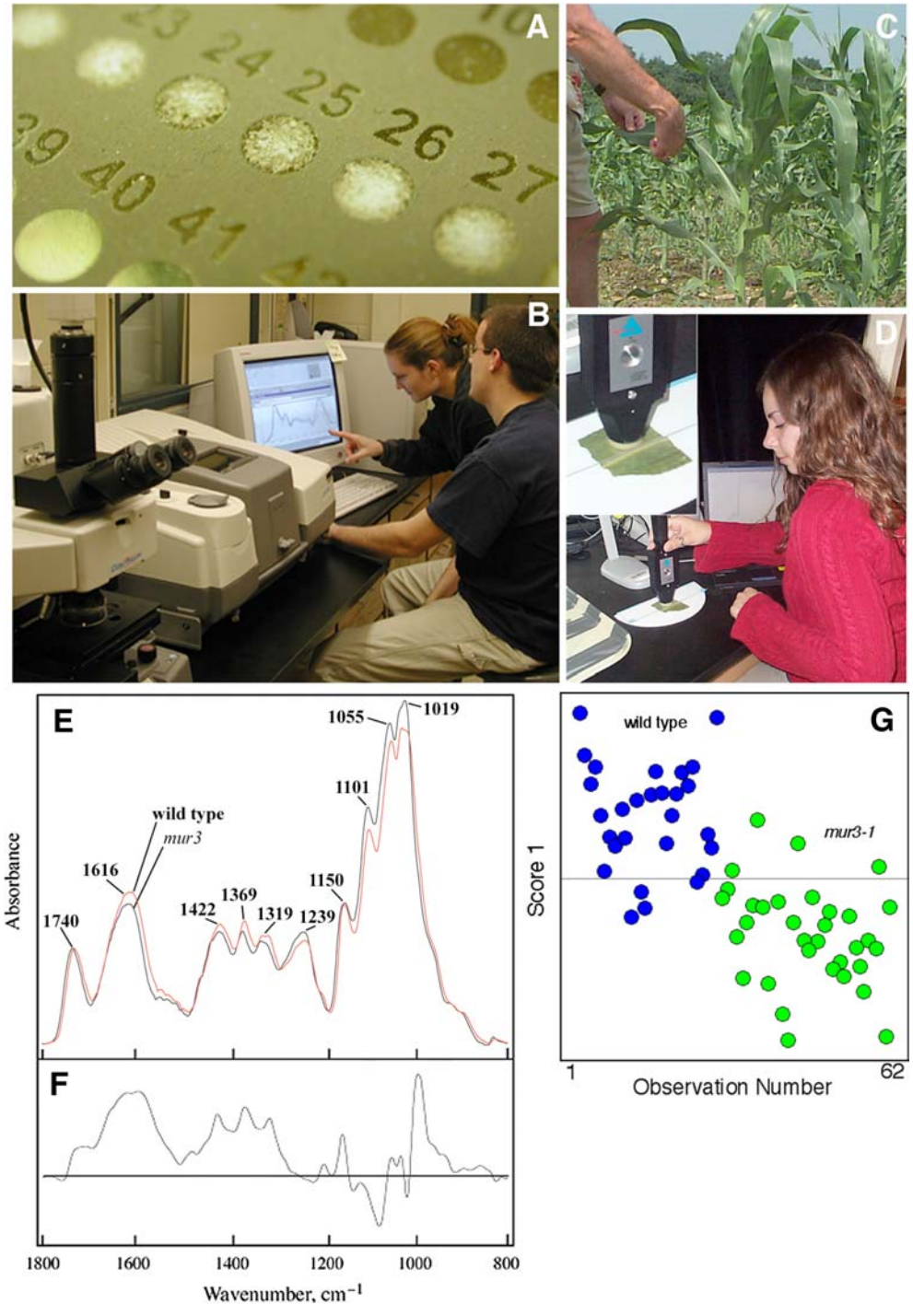

Fig. 1 Screening for cell-wall defective mutants by Fourier transform mid-range and near infrared spectroscopy. **a**. Isolated cell walls from an individual plant are placed in one of 48 wells of a gold-plated, IR-reflective slide. **b**. A computer-driven stage of the FTIR Microscope collects several mid-range IR spectra from each wall sample. **c**. Leaves from up to twenty individuals from 1000 segregating Uniform *Mu* lines are collected from field-grown plants, and air-dried at 50°C, **d**. Thirty near IR spectra are collected from a 1 cm diameter spot on the adaxial surface of a leaf blade, averaged and recorded. **e**. Analysis of infrared spectral signatures begins with digital subtraction. Average spectra generated from 30 spectra of wild-type and 30 spectra of the xyloglucan galactosyl transferase mutant, *mur3*, are very similar, but digital subtraction of the averaged *mur3* spectrum from that of the wild-type reveals some differences (**f**). **g**. Spectra obtained from populations of wild-type and *mur3* cell walls are subjected to data compression using Principal Components Analysis, followed by Canonical Variate Analysis (CVA). CVA is a widely used method for analyzing group structure in multivariate data (Kemsley 1998) and is mathematically equivalent to a one-way multivariate analysis of variance. Wild-type and *mur3* cell walls can be discriminated from each other using one CV score, with a correct classification of 87%, reflecting the probability of correctly identifying an unknown on the basis of its spectral signature. Similar methods are applied to the NIR data to select potential mutant individuals from each of the segregating Uniform *Mu* lines

consequence of mutations in genetic elements controlling developmental programs rather than in mutation of a gene directly involved in wall biogenesis. Therefore, a simple or direct relationship is unlikely to exist between a mutant genotype and its impact on wall architecture as revealed by a mutant spectrotype. However, establishing a broad library of distinct spectrotypes will provide essential information on the chemical and physical assays needed to dissect gene function.

Acknowledgments All of the authors are supported by a grant from the US National Science Foundation Plant Genome Research Program (DBI-0217552), with additional support by the National Science Foundation (IBN-0215535) to X.L. and W-D.R and the Department of Energy's Biomass Program to S.R.T. We thank Adam Santone for the image of the cell-wall "microarray" depicted in Fig. 1. Journal paper No. 17,698 of the Purdue University Agricultural Experiment Station.

References

Almeida JS (2002) Predictive non-linear modeling of complex data by artificial neural networks. Curr Opin Biotech 13:72–76

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen HM, Shinn P et al. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science 301:653–657

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. Annu Rev Plant Biol 54:519–546

Borderies G, Jamet E, Lafitte C, Rossignol M, Jauneau A, Boudart G, Monsarrat B, Esquerre-Tugaye MT, Boudet A, Pont-Lezica R (2003) Proteomics of loosely bound cell wall proteins of *Arabidopsis thaliana* cell suspension cultures: A critical analysis. Electrophoresis 24:3421–3432

Carpita NC, Gibeaut DM (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. Plant J 3:1–30

Carpita NC, Tierney M, Campbell M (2001) Molecular biology of the plant cell wall: searching for the genes that define structure, architecture, and dynamics. Plant Mol Biol 47:1–5

Chen L-M, Carpita NC, Reiter W-D, Wilson RH, Jeffries C, McCann MC (1998) A rapid method to screen for cell-wall mutants using discriminant analysis of Fourier transform infrared spectra. Plant J 16:385–392

Cosgrove DJ, Li LC, Cho HT, Hoffmann-Benning S, Moore RC, Blecker D (2002) The growing world of expansins. Plant Cell Physiol 43:1436–1444

Coutinho PM, Stam M, Blanc E, Henrissat B (2003) Why are there so many carbohydrate-active enzyme-related genes in plants? Trends Plant Sci 8:563–565

Dhugga KS, Barreiro R, Whitten B, Stecca K, Hazebroek J, Randhawa GS, Dolan M, Kinney AJ, Tomes D, Nichols S, Anderson P (2004) Guar seed β-mannan synthase is a member of the cellulose synthase supergene family. Science 303:363–366

Eisenhaber B, Wildpaner M, Schultz CJ, Borner GHH, Dupree P, Eisenhaber F (2003) Glycosylphosphatidylinositol lipid anchoring proteins. Sensitive prediction from sequence- and genome-wide studies for Arabidopsis and rice. Plant Physiol 133:1691–1701

Franke R, Hemm MR, Denault JW, Ruegger MO, Humphreys JM, Chapple C (2002) Changes in the secondary metabolism and deposition of an unusual lignin in the *ref8* mutant of Arabidopsis. Plant J 30:47–59

Gaspar Y, Johnson KL, McKenna JA, Bacic A, Schultz CJ (2001) The complex structures of arabinogalactan-proteins and the journey towards understanding function. Plant Mol Biol 47:161–176

Gillmor CS, Poindexter P, Lorieau J, Palcic MM, Somerville C (2002) α-Glucosidase I is required for cellulose biosynthesis and morphogenesis in Arabidopsis. J Cell Biol 156:1003–1013

Gutiérrez RA, Green PJ, Keegstra K, Ohlrogge JB (2004) Phylogenetic profiling of the *Arabidopsis thaliana* proteome: what proteins distinguish plants from other organisms? Genome Biology 5:R53

Hall Q, Cannon MC (2002) The cell wall hydroxyproline-rich glycoprotein RSH is essential for normal embryo development in Arabidopsis. Plant Cell 14:1161–1172

Hazen SP, Scott-Craig JS, Walton JD (2002) Cellulose synthase-like genes of rice. Plant Physiol 128:336–340

Hrmova M, Fincher GB (2001) Structure-function relationships of β-D-glucan endo- and exohydrolases from higher plants. Plant Mol Biol 47:73–91

Kemsley EK (1998) Discriminant Analysis of Spectroscopic Data. Wiley, Chichester, UK, p 193

Liepmann AH, Wilkerson CG, Keegstra K (2005) Expression of cellulose-synthase-like (Csl) genes in insect cells reveals that CslA family members encode mannan synthases. Proc Natl Acad Sci USA 102:2221–2226

Lukowitz W, Mayer U, Jurgens G (1996) Cytokinesis in the Arabidopsis embryo involves the syntaxin-related KNOLLE gene product. Cell 84:61–71

Marita JM, Ralph J, Hatfield RD, Chapple C (1999) NMR characterization of lignins in *Arabidopsis* altered in the activity of ferulate 5-hydroxylase. Proc Natl Acad Sci USA 96:12328–12332

McCann MC, Carpita NC (2005) Looking for invisible phenotypes in cell-wall mutants of *Arabidopsis thaliana*. Plant Biosystems (in press)

McCann MC, Roberts K (1991) Architecture of the primary cell wall. In: Lloyd CW (ed) The Cytoskeletal basis of plant growth and form. Academic Press, London, pp 109–129

McCann MC, Bush M, Milioni D, Sado P, Stacey NJ, Catchpole G, Defernez M, Carpita NC, Hofte H, Ulvskov P, Wilson RH, Roberts K (2001) Approaches to understanding the functional architecture of the plant cell wall. Phytochemistry 57:811–821

Milioni D, Sado PE, Stacey NJ, Roberts K, McCann MC (2002) Early gene expression associated with the commitment and differentiation of a plant tracheary element is revealed by cDNA-amplified fragment length polymorphism analysis. Plant Cell 14:2813–2824

Perrin R, Wilkerson C, Keegstra K (2001) Golgi enzymes that synthesize plant cell wall polysaccharides: finding and evaluating candidates in the genomic era. Plant Mol Biol 47:115–130

Raes J, Rohde A, Christensen JH, Van de Peer Y, Boerjan W (2003) Genome-wide characterization of the lignification toolbox in Arabidopsis. Plant Physiol 133:1051–1071

Reiter W-D, Vanzin G (2001) Molecular genetics of nucleotide-sugar interconversion pathways in plants. Plant Mol Biol 47:95–113

Reiter W-D, Chapple C, Somerville CR (1997) Mutants of *Arabidopsis thaliana* with altered cell wall polysaccharide composition. Plant J 12:335–345

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

Rose JKC, Braam J, Fry SC, Nishitani K (2002) The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: Current perspectives and a new unifying nomenclature. Plant Cell Physiol 43:1421–1435

Shiu SH, Bleecker AB (2001) Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. Proc Natl Acad Sci USA 98:10763–10768

Taylor NG, Howells RM, Huttly AK, Vickers K, Turner SR (2003) Interactions among three distinct CesA proteins essential for cellulose synthesis. Proc Natl Acad Sci USA 100:1450–1455

Zhong R, Burk DH, Morrison III WH, Ye Z-H (2002) A kinesin-like protein is essential for oriented deposition of cellulose microfibrils and cell wall strength. Plant Cell 14:3101–3117