

A Survey of Databases for Analysis of Plant Cell Wall-Related Enzymes

Peijian Cao · Ki-Hong Jung · Pamela C. Ronald

Published online: 24 February 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Biofuels derived from plant cell wall lignocellulose have the potential to serve as an alternative source of energy, relieving dependence on finite petroleum reserves and reducing production of climate-changing greenhouse gases. To better elucidate cell wall structure, the plant research community has developed databases to host the accumulated information on plant cell wall-related enzymes. The goal of this review is to provide a comprehensive catalog of these databases, as well as to describe the data and tools that they make available. A Directory of Databases for Plant Cell Wall-Related Enzymes (plantcellwalls.ucdavis.edu) was also developed to provide links to all the databases reviewed in this paper and relevant publications.

Keywords Biofuel · Plant cell wall · Database

P. Cao · K.-H. Jung · P. C. Ronald
Department of Plant Pathology, University of California,
Davis, CA 95616, USA

P. Cao · K.-H. Jung · P. C. Ronald (✉)
Department of Grass Genetics, Feedstocks Division,
Joint BioEnergy Institute,
Emeryville, CA 94710, USA
e-mail: pcronald@ucdavis.edu

P. Cao
Institute of Bioinformatics, Zhejiang University,
Hangzhou, Zhejiang 310029, China

K.-H. Jung
Department of Plant Molecular Systems Biotechnology and Crop
Biotech Institute, Kyung Hee University,
Yongin 446-701, Korea

Introduction

With the increasing demand for energy and the decrease in petroleum reserves, biofuels have arisen as a potential alternative source of energy. Ethanol and other fuels derived from the fermentation of sugars extracted from plant biomass are expected to become useful biofuels, thanks to its low toxicity, ease of biodegradation, and the lower levels of airborne pollutants [27]. Lignocellulosic biomass, mainly from plant cell walls, has advantages as a biofuel feedstock, compared with the currently used starch and cane sugars, largely due to its abundance and high production rate. However, the advantages of lignocellulosic biomass are overcome by the current limitations in the harvest of fermentable sugars from cellulose, a result of the inherent recalcitrance of plant cell walls. Thus, the characterization of the enzymes involved in the synthesis and decomposition of plant cell wall constituents, as well as a better understanding of the mechanisms of plant cell wall assembly are important goals in the field of biofuel research.

Plant cell walls are a complex and dynamic extracellular matrix that regulate cell growth, provide plants with mechanical support and protect against pathogens [4]. The basic structure of the primary cell wall consists of cellulose microfibrils embedded in a semi-structured matrix of non-cellulosic polysaccharides [5, 25]. There are two classes of primary cell walls in plants, which differ in their architecture, chemical composition and in their associated biosynthetic processes [4]. Type I walls, in which cellulose microfibrils are interwoven with glucans and xyloglucans and embedded in a matrix of pectin polysaccharides and

glycoproteins, are found in dicotyledonous plants [16]. Type II walls are characteristic of cereals and other grasses. Glucuronarabinoxylans are the main cross-linking glycans that surround cellulose microfibrils in type II walls. Additionally, type II walls contain a lesser amount of pectin polysaccharides, xyloglucan, and structural proteins, as compared to type I walls [4]. Plant secondary cell walls are composed of cellulose, hemicellulose and lignin. The cellulose matrix of the secondary cell wall is typically cross-linked by a phenyl propanoid-derived lignin mesh-work form [20]. Lignins are complex racemic aromatic heteropolymers derived mainly from three hydroxycinnamyl alcohol monolignols, which are synthesized through the shikimate and phenylpropanoid pathways [1]. Lignin is the main obstacle to fermentation of sugars to ethanol, as it prevents cell wall hydrolysis enzymes from accessing polysaccharides [17, 26].

Due to the complex structure of plant cell walls and the variety of monosaccharide and polysaccharide cell wall components, plants encode a large variety of enzymes to synthesize, modify and degrade cell walls. For instance, *Arabidopsis* devotes approximately 10% of its genome to the metabolism of cell walls [31]. Among of them, enzymes involved in carbohydrate metabolism are designated carbohydrate-active enzymes (CAZy), and include glycosyltransferases (GTs), glycoside hydrolases (GHs), polysaccharide lyases (PLs), and carbohydrate esterases (CEs). The GTs are a large group of enzymes that are centrally involved in the biosynthesis of plant cell walls, catalyzing the transfer of sugar moieties from activated donor molecules to acceptor molecules to form glycosidic bonds. In contrast to GTs, the PLs, and GHs are enzymes primarily involved in the breakdown of carbohydrates. PLs cleave polysaccharide chains via a beta-elimination mechanism that results in the formation of a double bond at the newly formed non-reducing end [15]. GHs hydrolyze the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety [8]. The modification of carbohydrates is performed by CEs, which catalyze the de-*O* or de-*N*-acetylation of substituted saccharides. Lignin is formed through oxidative coupling of monolignols within the apoplasmic space, which are synthesized from phenylalanine by a large variety of enzymes, including hydroxylases, methyltransferases, and dehydrogenases [1].

The importance of cell wall-related research to the development of lignocellulosic-derived biofuels has resulted in enormous community efforts aimed at gathering as much information as possible about cell wall-related enzymes. These data have been organized and stored in several online databases. In this review, we provide a comprehensive account of the databases currently available for the analysis of plant cell wall-related enzymes (Table 1). The databases

discussed are organized into three sections: general databases, species-specific databases and family-specific databases. For each database, we describe the covered species, primary content, literature availability, data collected and the tools available through the database. We also set up a Directory of Databases for Plant Cell Wall-Related Enzymes (plantcellwalls.ucdavis.edu) to list all the databases reviewed in this paper and provide links and relevant publications, which will be updated periodically.

General Databases

In this section, we describe three main general databases that catalog cell wall-related genes from different species.

CAZy

The carbohydrate-active enzymes (CAZy, <http://www.cazy.org/>) database is the most comprehensive knowledge-based database specializing in cell wall-related enzymes, and has been available on the web since September 1998. The CAZy database contains information about enzymes that build and breakdown complex carbohydrates and glycoconjugates [2]. The proteins in this database are classified into different families, primarily based on amino acid sequence similarity, a comparison that can reflect the structural features of these enzymes and help to reveal evolutionary relationships. As of May 2009, CAZy contains five superfamilies with the following enzyme activities: GTs, containing 91 families; GHs, 115; PLs, 21; CEs, 16; and carbohydrate-binding modules, 54. Family-associated information and content, such as known activities, catalytic reaction mechanisms and additional statistics, are provided by the database and these entries are continuously updated. New families are created based on surveys of new publications. Additionally, re-analysis of previously released genomes and sequences available from public databases are performed to integrate new family members and assure complete coverage. The sequence annotation of every entry is available through links to public databases (e.g., National Center for Biotechnology Information (NCBI) and UniProt). Biochemical information about the protein entries, based on structural information and continuous curation of the available literature, is also provided for some enzymes in the database. For example, of the 112,398 total proteins contained in the database as of May 2009, 8,245 have been assigned EC numbers and 3,919 PDB 3D structures are available for 947 of these proteins. CAZy also provides a list of the families present in each of the 952 species for which data is available in the database. Users can search the CAZy database using Protein Accessions (GenBank/GenPept Accession, UniProt Accession and

Table 1 Currently available databases on plant cell wall-related enzymes and their attributes

	Database	Species	No. of sequences	Literature	Phylogenetic tree	Expression data	Functional genomic data	Last update
General	CAZy	Archeae, bacteria, viruses, and eukaryota	112,398 ^a	N/A	N/A	N/A	N/A	Semi-monthly
	Cell Wall Genomics	<i>Arabidopsis</i> , rice, and maize	2,381	Family level	Interactive trees	N/A	Mutant information and FTIR spectra data	Unknown
	Cell Wall Navigator	Organism-unrestricted	4,591	Family level	Available	External links available for <i>Arabidopsis</i> genes	Several links to external databases	Unknown
Species-specific	Rice GT Database	Rice	793	N/A	Available	EST, MPSS and microarray data	Orthologs, protein topology and mutants	August, 09
	MAIZEWALL	Maize	753	Links to PubMed	N/A	Microarray data for development	Homologs and protein domains	June 6, 2006
	Wheat GTIdb	Wheat	8,300	N/A	Available	N/A	N/A	January, 2008
Family-specific	GHDB	<i>Arabidopsis</i> , rice, etc.	319	N/A	Available	N/A	Functionally relevant residues	June, 2004
	XTH World	<i>Arabidopsis</i> , rice, and tomato	87	1962–2006	Available for <i>Arabidopsis</i> and rice	N/A	N/A	July, 2006
	Expansin Database	Plants	226	Available	Available for <i>Arabidopsis</i>	N/A	N/A	March 22, 2006

^aData for this figure was gathered on May 7, 2009

PDB ID), CAZy family names, NCBI Taxonomy IDs and Enzyme Commission (EC) numbers. Further, database searching can also be complemented by Google-based searches available on the site.

Cell Wall Genomics

Cell Wall Genomics (<http://cellwall.genomics.purdue.edu/>) is a useful online resource database that supports genetic analyses of cell wall-related genes in *Arabidopsis* and maize [19]. This database indexes the homozygous T-DNA insertion mutants for more than 1,100 cell wall-related genes that are currently being generated at the University of Wisconsin. This database divides cell wall biogenesis into six distinct stages, involving 44 gene families that contain 1,029 *Arabidopsis*, 859 rice, and 493 maize cell wall-related genes. As of this writing, 317 homozygous T-DNA insertion mutants for 280 *Arabidopsis* genes have been generated and their spectroscopic phenotypes or “spectrotypes” obtained via Fourier Transform Infrared (FTIR)

spectroscopy have been determined. In addition, 39 maize spectral mutants identified by screening the UniformMu population with near infrared spectroscopy (NIR) have been identified [29, 31]. The FTIR and NIR spectral data can be downloaded from this database. A brief description and the family structure are provided for each individual gene family. Phylogenetic trees are available for *Arabidopsis*, rice and maize, either alone or combined, making it easier to identify possible orthologs between the three species. Each tree can be clicked to open an interactive FLASH page that contains links to other sequence databases, such as TAIR and Gramene, for each gene. The sequences used to make the tree are available on the link “View the protein sequence file” under the trees. There are also four tutorials on the front page detailing how the data analysis of tree construction was performed and how the users can print and reproduce these trees. Tutorial of the whole database and FTIR data analysis are also included. Relevant publications for individual gene families are summarized when they are available.

Cell Wall Navigator

Cell Wall Navigator (CWN; <http://bioweb.ucr.edu/Cellwall/>) is an integrated database and mining tool used to search for protein families involved in plant cell wall metabolism [11]. The sequences used by CWN are retrieved from three different resources: the completed genome sequences of *Arabidopsis* and *O. sativa* spp. *Japonica*, the UniProt database and the expressed sequence tags (EST) division of NCBI. As of May 2009, CWN contains 4,591 sequences that code for enzymes and structural proteins known to be involved in sugar substrate generation and primary cell wall metabolism. These sequences are classified into five stages of cell wall biosynthesis, which are subdivided into 16 groups and 35 families based on sequence similarity, classes of polysaccharides, structural proteins, stage of assembly/disassembly, and CAZy families.

The CWN interface provides comprehensive query and visualization functions for mining sequence features, exploring evolutionary relationships, and retrieving biological information within and between families [11]. The interface is composed of three hierarchical levels: the family selection or index page, family browser pages and gene annotation pages. First, users can select a gene family of interest from the index page. Next, information about family members, sequence alignment, gene structures, sequences and publications on the chosen family is available on the family browser pages. Moreover, interactive trees with user-collapsible and user-expandable branches are also provided. In the last level, information including protein/nucleotide sequences and several feature viewing options, as well as many links to external resources for accessing additional information about individual family members, can be retrieved from the gene annotation pages. All sequences, alignments, and trees can be downloaded in different formats. In addition to a public database, CWN is also an annotation forum that allows registered users to upload important information about sequences, mutants, phenotypes, antibodies, protein functions and other valuable data, sharing it with the broader cell wall community. In contrast to the CAZy and Cell Wall Genomics databases, CWN provides several kinds of functional genomic data and visualization tools, making it a very useful database for plant cell wall-related gene research.

Species-Specific Databases

The previously discussed general databases contain data on cell wall-related enzymes from a variety of organisms (eukaryotes, prokaryotes and viruses), but the majority of entries in these databases come from the completely

sequenced genomes released by the NCBI as regular GenBank entries. In addition, not all CAZy are included in the general databases [3, 9]. Species-specific databases, which provide deeper coverage and more comprehensive analysis for their dedicated species, are a good complement to the general databases.

The Rice GT Database

The Rice GT Database (<http://ricephylogenomics.ucdavis.edu/cellwalls/gt/>) is a phylogenomic database that integrates, hosts, and displays functional genomic data in a phylogenetic context, with the goal of facilitating functional analysis of the large GT family of enzymes in the rice reference monocotyledonous species [3]. This database provides information about 617 potential GT genes (loci), corresponding to 793 transcripts (gene models) in rice.

The Rice GT Database can display several classes of functional genomic data, including mutant lines and gene expression data, for each rice GT gene in the context of a phylogenetic tree, allowing for comparative analysis both within and between GT families. On the tree viewer page, functional genomic data fields can be selected by checking their corresponding boxes. Pressing the submit button displays the selected data adjacent to the GT phylogenetic tree. The spreadsheet format allows data to be readily transferred into any database or software (such as Excel) for further analysis. Clicking on a gene model ID (12XXX.mXXXX) link provides a summary webpage for that gene model that shows all available data (excluding microarray data), including histogram representations of expression patterns of EST and MPSS tag counts. Links to the MSU/TIGR rice database, Rice Annotation Project Database (RAP-DB), CAZy database, and NCBI basic local alignment search tool (BLAST) search are provided for ease of navigation. Mutant line identification numbers are provided as hyperlinks to the corresponding libraries when phenotypic information is available for a particular mutant. Users may toggle between displaying numerical values for each replicate or averages for each sample in the display of microarray data. In addition, red/green heatmaps can be generated, providing for easy examination of each microarray dataset. The chromosome distribution map is color coded according to the different CAZy families and rice GT loci are represented as colored boxes. Placing the mouse cursor over each box generates a pop-up window that shows the ID of each rice GT locus. Clicking on the box directs the user back to the tree viewer page with the selected rice GT at the top of the view window. A search function is also available, enabling users to search the database with a locus ID or protein sequence.

The functional genomic data in the Rice GT Database allowed for the identification of 33 rice-diverged GT genes

(45 gene models) that are highly expressed in above-ground, vegetative tissues. These particular genes are good candidates for further characterization and possible application to biofuel development [3]. Specifically, these genes are expected to have important roles in the biosynthesis of grass-specific cell wall components, making them prime candidates for further functional analysis. Functional redundancy within gene families makes it difficult to determine the function of specific family members, especially in the case of large gene families [13]. Using the GT database, users can easily predict functional redundancy or choose predominant gene family members by evaluating the integrated transcriptomic data. Further, users can easily assess phenotypic data about the interesting gene family member or members if that data is available.

MAIZEWALL

The MAIZEWALL database (<http://www.polebio.scsv.upstlse.fr/MAIZEWALL/>) provides a bioinformatic analysis and gene expression data repository of cell wall biosynthesis and assembly in maize [12]. This database contains 735 contigs belonging to 174 putative gene functions that were identified and further classified into 18 functional categories based on BLAST searches against the un-annotated maize G noPlanteInfo contig database [23], using known cell wall biosynthesis and assembly genes in plants obtained from NCBI by nearly 100 keywords defined based on current knowledge. In addition to providing a gene catalog, the MAIZEWALL database allows users to perform multiple sequence alignments and identify predicted protein domains and sub-cellular localizations of target sequences by using an assortment of user-friendly bioinformatics software. MAIZEWALL also contains a full set of developmental microarray gene expression data for 651 of the 735 contigs, complete with hierarchical gene clustering analysis. Literature references for each family are provided as links to PubMed. A versatile sequence search engine that provides predefined keywords and categories is also available.

Wheat GTIdb

The Wheat GlycosylTransferase Inventory database (GTIdb; <http://www.appli.nantes.inra.fr:8180/GTIDB/>) is a searchable database that can be used to survey wheat genes for a role in a particular biological process [22]. The database consists of two parts: the wheat section and the core database. The core database contains 476 *Arabidopsis* and 929 rice sequences extracted from the CAZy database and classified into 41 plant GT families. Based on the core database, 912,573 wheat ESTs, extracted from 220 libraries from various origins and sizes, were used to characterize 833 contigs

and 2,296 singletons as GT sequences. For each family, the database provides a list of the members from wheat contigs, wheat singletons, *Arabidopsis*, rice IRGSP, and TIGR entries, and all sequences are downloadable from the interface. Phylogenetic trees that integrate all three species' members for each family are available in PDF format. This website is implemented with Java applet and servlet technologies, an architecture that provides a convenient method to manipulate and download data. In addition, this implementation allows for further graphical improvements.

Family-Specific Databases

Similar to species-specific databases, family-specific databases focus on a specific cell wall-related gene family and provide additional broader functional information. This type of database is very useful to the research community of the corresponding gene family.

GHDB

The Glycoside Hydrolases Database (GHDB; <http://www.ghdb.uni-stuttgart.de/>) is a structured database dedicated to the CAZy database family GH16 glycoside hydrolases [28]. The database is composed of 260 proteins and 319 sequences from 128 species. These database entries are assigned to three superfamilies (GH16a, GH16b, and GH16c) and SIX homologous families. In addition to the catalog of family members, the database provides experimentally determined 3D protein structures, phylogenetic trees and multiple sequence alignments of five homologous families of the GH16a and GH16b, and GH16c superfamily. Functionally relevant residues, such as catalytic amino acid residues and aromatic substrate-binding residues, among others, are highlighted in the sequence alignments provided by GHDB. A BLAST search is also integrated into the database, allowing for comprehensive analysis of the stored data. This database has been successfully used for homology modeling of seven glycoside hydrolases with varying substrate specificities from the GH16 family, enzymes for which no structural information is available [28].

XTH World

Xyloglucan endotransglucosylase-hydrolases (XTH) World (<http://labs.plantbio.cornell.edu/XTH/>) is a database specifically targeted toward the nomenclature and systematic identification of XTH, which are enzymes capable of splitting and reconnecting xyloglucan molecules in rapidly growing plant tissues [7]. Due to the rapid proliferation of

research in this area, many overlapping and contradictory nomenclatures and classifications for this type of gene have been adopted. Xyloglucan endotransglycosylase [24] and endoxyloglucan transferase (EXT, later re-designated EXGT) [18] are two examples of classifications that have been later re-designated. Thus, a standardized nomenclature for these genes is critical for sharing of information within the research community. In 2002, following a meeting of several research groups working with these enzymes/genes, a new unifying nomenclature was adopted [21] and made available online as the XTH World database. This database contains information about systematic identification, nomenclature and gene structure of *XTH* gene family members from *Arabidopsis* (33 genes), tomato (25), and rice (29). Phylogenetic trees are available for *Arabidopsis* and rice XTHs. Additionally, literature references from 1962 to 2006 for studies on *XTH* gene family members have been collected and are listed in this database.

Expansin Database

Expansins are a group of proteins with unique “loosening” effects on plant cell walls and are thought to function in plant cell growth, cell wall disassembly, cell separation, pollen tube penetration and leaf primordium initiation [6, 10, 14, 30]. The Expansin Database (<http://www.bio.psu.edu/expansins/>) is a central database that focuses mainly on plant expansin genes, providing comprehensive information about protein structure, mechanism of action, nomenclature, experimental protocols and literature references for this type of protein. Thirty-six *Arabidopsis* expansin genes have been identified and are included in this database. These genes are further grouped into four classes based on a phylogenetic tree. Moreover, 190 sequences from other plant species, including rice (58), poplar (36), and papaya (19), are also available in this database. A large number of literature references about this gene family are also provided, with abstracts and hot-links to PDF files when available.

Summary and Prospects

We have reviewed a wide spectrum of online databases of plant cell wall-related enzymes. Because of the proliferation of research into biofuels and plant cell walls, several plant cell wall enzyme databases have been developed. However, our overall understanding of this important field of research remains incomplete. For example, the CAZy database is the most comprehensive gene catalog of cell wall-related enzymes, but it lacks functional genomic data, which is important for characterization and prioritization of target genes for further analysis. On the other hand, the Rice GT

phylogenomic database, which incorporates several types of functional genomic data, has proven useful for selection of high priority target genes in grass-specific cell wall research but coverage is limited to rice GT gene family members. In the near future, we expect the establishment of a general cell wall-related gene database that contains integrated functional genomic data.

Acknowledgements We thank M. E. Vega-Sanchez for comments and discussions. This work was part of the DOE Joint BioEnergy Institute (<http://www.jbei.org>) supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the US Department of Energy.

Conflict of Interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. *Annu Rev Plant Biol* 54:519–546
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* 37:D233–D238
- Cao PJ, Bartley LE, Jung KH, Ronald PC (2008) Construction of a rice glycosyltransferase phylogenomic database and identification of rice-diverged glycosyltransferases. *Mol Plant* 1:858–877
- Carpita NC (1996) Structure and biogenesis of the cell walls of grasses. *Annu Rev Plant Physiol Plant Mol Biol* 47:445–476
- Carpita NC, Defernez M, Findlay K, Wells B, Shoue DA, Catchpole G et al (2001) Cell wall architecture of the elongating maize coleoptile. *Plant Physiol* 127:551–565
- Cho HT, Kende H (1997) Expansins in deepwater rice internodes. *Plant Physiol* 113:1137–1143
- Cosgrove DJ (2005) Growth of the plant cell wall. *Nat Rev Mol Cell Biol* 6:850–861
- Davies G, Henrissat B (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* 3:853–859
- Egelund J, Skjot M, Geshi N, Ulvskov P, Petersen BL (2004) A complementary bioinformatics approach to identify potential plant cell wall glycosyltransferase-encoding genes. *Plant Physiol* 136:2609–2620
- Fleming AJ, McQueenMason S, Mandel T, Kuhlemeier C (1997) Induction of leaf primordia by the cell wall protein expansion. *Science* 276:1415–1418
- Gerke T, Lauricha J, Tran H, Keegstra K, Raikhel N (2004) The cell wall navigator database. A systems-based approach to organism-unrestricted mining of protein families involved in cell wall metabolism. *Plant Physiol* 136:3003–3008, discussion 3001
- Guillaumie S, San-Clemente H, Deswarte C, Martinez Y, Lapierre C, Murigneux A et al (2007) MAIZEWALL. Database and developmental gene expression profiling of cell wall biosynthesis and assembly in maize. *Plant Physiol* 143:339–363

13. Jung KH, An G, Ronald PC (2008) Towards a better bowl of rice: assigning function to tens of thousands of rice genes. *Nat Rev Genet* 9:91–101
 14. Li Y, Jones L, McQueen-Mason S (2003) Expansins and cell growth. *Curr Opin Plant Biol* 6:603–610
 15. Linhardt RJ, Galliher PM, Cooney CL (1986) Polysaccharide lyases. *Appl Biochem Biotechnol* 12:135–176
 16. 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, New York, pp 109–129
 17. Mosier N, Wyman C, Dale B, Elander R, Lee YY, Holtzapple M et al (2005) Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresour Technol* 96:673–686
 18. Nishitani K, Tominaga R (1992) Endo-xyloglucan transferase, a novel class of glycosyltransferase that catalyzes transfer of a segment of xyloglucan molecule to another xyloglucan molecule. *J Biol Chem* 267:21058–21064
 19. Penning BW, Hunter CT 3rd, Tayengwa R, Eveland AL, Dugard CK, Olek AT et al (2009) Genetic resources for maize cell wall biology. *Plant Physiol* 151:1703–1728
 20. Popper ZA (2008) Evolution and diversity of green plant cell walls. *Curr Opin Plant Biol* 11:286–292
 21. Rose JK, 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
 22. Sado PE, Tessier D, Vasseur M, Elmorjani K, Guillon F, Saulnier L (2009) Integrating genes and phenotype: a wheat-Arabidopsis-rice glycosyltransferase database for candidate gene analyses. *Funct Integr Genomics* 9:43–58
 23. Samson D, Legeai F, Karsenty E, Reboux S, Veyrieras JB, Just J et al (2003) GenoPlante-Info (GPI): a collection of databases and bioinformatics resources for plant genomics. *Nucleic Acids Res* 31:179–182
 24. Smith RC, Fry SC (1991) Endotransglycosylation of xyloglucans in plant cell suspension cultures. *Biochem J* 279(Pt 2):529–535
 25. Somerville C, Bauer S, Brininstool G, Facette M, Hamann T, Milne J et al (2004) Toward a systems approach to understanding plant cell walls. *Science* 306:2206–2211
 26. Somerville C (2006) The billion-ton biofuels vision. *Science* 312:1277
 27. Sticklen MB (2008) Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. *Nat Rev Genet* 9:433–443
 28. Strohmeier M, Hrmova M, Fischer M, Harvey AJ, Fincher GB, Pleiss J (2004) Molecular modeling of family GH16 glycoside hydrolases: potential roles for xyloglucan transglucosylases/hydrolases in cell wall modification in the poaceae. *Protein Sci* 13:3200–3213
 29. Vermerris W, Saballos A, Ejeta G, Mosier NS, Ladisch MR, Carpita NC (2007) Molecular breeding to enhance ethanol production from corn and sorghum stover. *Crop Sci* 47:S142–S153
 30. Yennawar NH, Li LC, Dudzinski DM, Tabuchi A, Cosgrove DJ (2006) Crystal structure and activities of EXPB1 (*Zea m 1*), a beta-expansin and group-1 pollen allergen from maize. *Proc Natl Acad Sci USA* 103:14664–14671
 31. Yong W, Link B, O'Malley R, Tewari J, Hunter CT, Lu CA et al (2005) Genomics of plant cell wall biogenesis. *Planta* 221:747–751
- ### Further Readings
32. NCBI: <http://www.ncbi.nlm.nih.gov/>
 33. UniProt: <http://www.uniprot.org/>
 34. TAIR: <http://www.arabidopsis.org/>
 35. MSU/TIGR Rice Genome Annotation: <http://rice.plantbiology.msu.edu/>
 36. RAP-DB: <http://rapdb.dna.affrc.go.jp/>
 37. Gramene: <http://www.gramene.org/>