Mini Review

Genomic Basis for Cell-Wall Diversity in Plants.
A Comparative Approach to Gene Families in Rice and *Arabidopsis*

Ryusuke Yokoyama and Kazuhiko Nishitani [1]

Department of Developmental Biology and Neurosciences, Graduate School of Life Sciences, Tohoku University, Sendai, 980-8578 Japan

        Monocotyledons and dicotyledons are distinct, not only in their body plans and developmental patterns, but also in the structural features of their cell walls. The recent completion of the rice (*Oryza sativa*) genomic sequence and publication of the sequence data, together with the completed database of the *Arabidopsis thaliana* genome, provide the first opportunity to compare the full complement of cell-wall-related genes from the two distinct classes of flowering plants. We made this comparison by exploiting the fact that *Arabidopsis* and rice have type I and type II walls, respectively, and therefore represent the two extremes in terms of the structural features of plant cell walls. In this review article, we classify all cell-wall-related genes into 32 gene families, and generate their phylogenetic trees. Using these data, we can phylogenetically compare individual genes of particular interest between *Arabidopsis* and rice. This comparative genome approach shows that the differences in wall architecture in the two plant groups actually mirror the diversity of the individual gene families involved in the cell-wall dynamics of the respective plant species. This study also identifies putative rice orthologs of genes with well-defined functions in *Arabidopsis* and other plant species.

**Keywords**: Chitinase — Monolignol — Pectin — Polysaccharide — XTH — Xyloglucan.

Introduction

        The morphological diversity of flowering plants mirrors the diversity of cell types, each of which adopts a specific shape and plays a role that is peculiar to plant species. The cell wall plays several critical roles in determining the cell type. Moreover, it plays a key role in regulating cell growth and differentiation and cell-wall dynamics are reflected in the developmental pattern (Freshour et al. 1996, and reviewed by Martin et al. 2001, Nishitani 2002, Fry 2003). The primary cell walls of flowering plants are classified into two major groups, type I walls and type II walls, with respect to the chemical structures of components, wall architecture and their biosynthetic proc-

esses (Carpita 1996). Cells of dicotyledonous plants and the non-commelinoid monocotyledonous plants are composed of type I cell walls, which are characterized by a cellulose–xyloglucan framework with approximately equal amounts of cellulose microfibrils and xyloglucans (reviewed by Nishitani 1997, Fry 2003). The cellulose–xyloglucan framework is typically embedded in a network of abundant pectic polysaccharides, which comprise principally homogalacturonans (HGA), rhamnogalacturonan I (RG-I) and rhamnogalacturonan II (RG-II). Current models suggest that these three components are covalently linked to one another, thereby forming the pectic network (reviewed by Willats et al. 2001, Ridley et al. 2001). Neutral polymers composed of arabinose or galactose residues are usually attached, as branches, to the rhamnosyl residues of RG-I in the pectic polysaccharides. Some of these side chains are further cross-linked by ester linkages to other pectic components or to non-pectic polymers through feruloyl and coumaroyl residues. In the pectic network, calcium ions serve as cross-links between the de-esterified carboxylic acid groups, particularly in the HGA and RG-I domains, whereas borate diester bridges cross-link the RG-II domains (Kobayashi and Matoh 1996, Ishii et al. 1999, O'Neill et al. 2001).

        Type II walls are found only in commelinoid monocotyledons, which include cereals such as rice (*Oryza sativa*), oats and barley. Unlike the type I walls, the type II walls have less xyloglucan than cellulose. The predominant glycans that cross-link the cellulose microfibrils in cereals are glucuronoarabinoxylan (Nishitani and Nevins 1991, Carpita and Gibeaut 1993) and $\beta 1,3:\beta 1,4$ mixed glucans (Kato et al. 1982). Compared with the pectin-abundant type I wall, the type II wall contains less pectin and higher amounts of phenylpropanoids, which form extensive interconnecting networks primarily when cells stop expanding (Iiyama et al. 1990). The principal hydroxycinnamate compound found in the non-lignified type II wall is ferulic acid, residues of which are esterified to the C5 of the arabinosyl side chains of arabinoxylans (Nishitani and Nevins 1989, Ishii 1997). Ferulate residues are thought to undergo oxidative dimerization, thereby forming arabinoxylan networks (Lam et al. 1994, Marry et al. 2001, Yoshida-Shimokawa et al. 2001). Fig. 1 diagrammatically illustrates the structural differences between the two cell-wall types, as represented by *Arabidopsis* and rice.

[1] Corresponding author: E-mail, nishitan@mail.tains.tohoku.ac.jp; Fax, +81-22-217-6700.
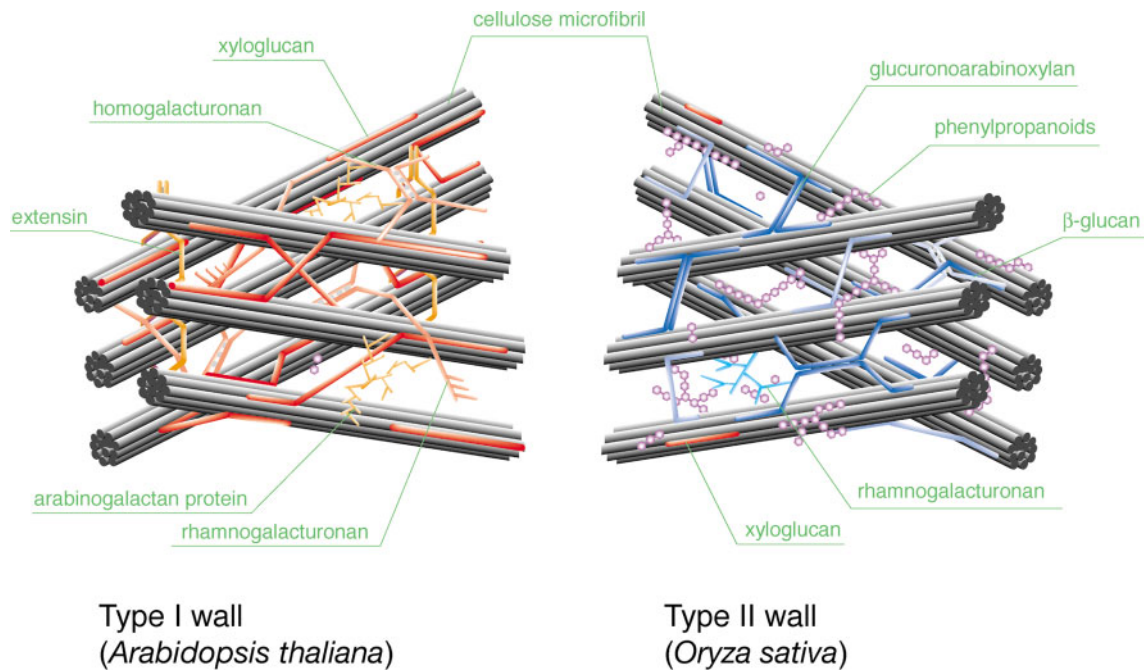
Fig. 1 Schematic structural models of type I and type II walls as represented by *Arabidopsis* and rice cell walls, respectively. The schemes are based on the model of Carpita and McCann (2000).

Both types of wall are highly organized supermolecules, assembled from many different types of polysaccharides, phenylpropanoids, and structural proteins. For detailed information about the structural proteins, which we will not deal with in this article, the reader is referred to a comprehensive review by Johnson et al. (2003). The structural complexity of the cell wall within a single cell and its diversity among cell types are mirrored by the vast array of enzymes and structural proteins required for the synthesis, assembly and disassembly of the cell wall. Therefore, some genome-wide approach is required to clarify the molecular processes that underlie cell-wall dynamics.

The rice (*Oryza sativa*) genomic sequence has recently been completed. These newly published data, together with the completed database of the *Arabidopsis* genome sequence (Arabidopsis Genome Initiative 2000), allow comparative phylogenetic analyses (Sasaki and Burr 2000, Buell 2003) of the whole complement of cell-wall-related genes in a dicotyledon and a commelinoid monocotyledon.

A comparative genomic analysis of rice and *Arabidopsis* is extremely informative in that the two plant species represent two extremes in terms of the structural features of their cell walls. In this review, we start by defining the members of 32 families of genes that encode proteins involved in the synthesis, modification, assembly and disassembly of the cell walls of both *Arabidopsis* and rice plants. Using sequence data for the whole complement of defined gene family members, we compare family genes with particular physiological roles in the cell wall between *Arabidopsis* and rice in the context of the structural differences between type I and type II cell walls.

Classification of Cell-wall-related Genes of Arabidopsis and Rice

Based on the wide array of information deposited in database resources, and on conventional biological knowledge accumulated by many cell-wall studies during the last century, we selected 32 families of proteins as representing the proteins implicated in the construction and modification processes of plant cell walls (Table 1–3). In this review, we focus on all the members of these gene families in both *Arabidopsis* and rice.

Arabidopsis gene families—Table 1 summarizes the number of *Arabidopsis* genes defined as members of individual families based on previous references and the CAZy database, which describes families of structurally related catalytic and carbohydrate-binding modules of enzymes involved in glycosidic linkages (Coutinho and Henrissat 1999, http://afmb.cnrs-mrs.fr/~pedro/CAZY). To compare the size and diversity of a given gene family quantitatively between *Arabidopsis* and rice, the boundary of the gene family must be defined precisely based on an objective criterion. Therefore, we re-examined the criteria considered sufficient to define cell-wall-related gene families. Among the proteins encoded by members of a given gene family, a common protein motif involved in molecular function that is characteristic of the gene family is often conserved. Such a motif is likely to be the intrinsic structural feature of the family. In *Arabidopsis*, the classification of the glycosyltransferases and glycoside hydrolases based on their conserved motifs (http://afmb.cnrs-mrs.fr/~pedro/CAZY, http://www.sanger.ac.uk/Software/Pfam/) is consistent with the classification found in published references (cf. Table 1 and 2). Therefore, in the present study, we adopted such a conserved

Table 1 Cell-wall-related gene families of *Arabidopsis* that have been reported in references or defined in the CAZy database

Family names [a]	References [b]	Gene family names defined in CAZy database [c]
Cellulose synthase	Holland et al. 2000 (40)	
Callose/glucan synthase	Hong et al. 2001 (12)	GT48 (12)
Glucosyltransferase	–	GT8 (42)
α-xylosyltransferase	Faik et al. 2002 (8)	GT34 (8)
β-galactosyltransferase	–	GT47 (39)
α-fucosyltransferase	Sarria et al. 2001 (13)	GT37 (10)
Xyloglucan endotransglucosylase/hydrolase	Yokoyama and Nishitani 2001 (33)	GH16 (33)
Expansin	Li et al. 2002 (35)	
β-1,4-glucanase	Ohmiya et al. 2003 (25)	GH9 (25)
β-1,3-glucanase	–	GH17 (49)
α-fucosidase	de la Torre et al. 2002 (1)	GH29 (1)
β-galactosidase	–	GH35 (18)
α-xylosidase	Sampedro et al. 2001 (2)	GH31 (5)
β-xylosidase	Goujon et al. 2003 (15)	GH3 (15)
α-mannosidase	–	GH38 (4)
Pectate lyase	–	PL1 (27)
Polygalacturonase	–	GH28 (69)
Pectin methylesterase	–	CE8 (67)
Pectin acetylesterase	–	CE13 (5)
Chitinase	Zhong et al. 2002 (24)	GH18,19 (24)
Cellulase (Mannan-hydrolase)	–	GH5 (13)
Xylanase	Suzuki et al. 2002 (12)	GH10 (12)
Laccase	–	
Phenylalanine ammonia lyase (PAL)	Raes et al. 2003 (4)	
tans Cinnamate 4-hydroxylase (C4H)	Raes et al. 2003 (1)	
4-coumarate:CoA ligase (4CL)	Raes et al. 2003 (14)	
Coumarate 3-hydroxylase (C3H)	Raes et al. 2003 (3)	
Caffeoyl-CoA *O*-methyltransferase	Raes et al. 2003 (7)	
Cinnamoyl-CoA reductase (CCR)	Raes et al. 2003 (7)	
Ferulate 5-hydroxylase (F5H)	Raes et al. 2003 (2)	
Caffeic acid *O*-methytransferase (COMT)	Raes et al. 2003 (14)	
Cinnamyl alcoholdehydrogenase (CAD)	Raes et al. 2003 (9)	

[a] Family names were defined based on references published thus far, the *Arabidopsis* genome database (http://www.Arabidopsis.org/), and the CAZy database (http://afmb.cnrs-mrs.fr/˜pedro/CAZY).

[b] The number of genes in each family defined in individual references is shown in parentheses after the respective reference.

[c] The number of genes in each gene family defined in the CAZy database is shown in parentheses. GT, glycosyltransferase family; GH, glycoside hydrolase family; PL, polysaccharide lyase family; CE, carbohydrate esterase family.

motif as the criterion for a gene family. A series of motif data for individual gene families were acquired from http://www.Arabidopsis.org/. We defined the members of a given gene family using both the conserved motifs characteristic of each family and the Pfam database (Table 2, http://www.Arabidopsis.org/tools/bulk/protein/index.jsp), except for the cellulose synthases, expansins, xyloglucan endotransglucosylase/hydrolases (XTHs) and monolignol biosynthesis gene families. For these gene families, we adopted the criteria used in recently published studies (Holland et al. 2000, Li et al. 2002, Yokoyama and Nishitani 2001, Raes et al. 2003). Based on these criteria, we retrieved 675 *Arabidopsis* genes and assigned them each to one of the 32 gene families (Table 3). Sequence

accession numbers for all the members of the 32 cell-wall-related gene families thus classified are shown in Table S1 of the supplementary table, which is available at the journal website www.pcp.oupjournals.org. Note that the member genes of individual families, as defined based on the new criteria in the present study, are different from those in the previously defined families (cf. Table 1–3).

Rice gene families—Comparative phylogenetic studies of rice and *Arabidopsis* have been carried out extensively on the cellulose synthase, XTH and expansin gene families. Therefore, we adopted the classifications arrived at in those previous studies. For the sequence data retrieved from these references, the reader is referred to the references (Hazen et al. 2002, Lee

Table 2 Sets of protein-domain families automatically generated based on the Pfam database

Family names [a]	Pfam ID [b]	Pfam AC [b]	Number of genes in each family [c]	
			Arabidopsis	Oryza
Cellulose synthase	–	–		
Callose/glucan synthase	Glucan_synthase	PF02364	12	10
Glucosyltransferase	Glyco_transf_8	PF01501	45	41
α-xylosyltransferase	Glyco_transf_34	PF05637	8	6
β-galactosyltransferase	Exostosin	PF03016	39	30
α-fucosyltransferase	XG_FTase	PF03254	9	19
Xyloglucan endotransglucosylase/hydrolase	Glyco_hydro_16	PF00722	33	29
Expansin	–	–		
β-1,4-glucanase	Glyco_hydro_9	PF00759	25	21
β-1,3-glucanase	Glyco_hydro_17	PF00332	52	57
α-fucosidase	Alpha_L_fucos	PF01120	1	2
β-galactosidase	Glyco_hydro_35	PF01301	19	15
α-xylosidase	Glyco_hydro_31	PF01055	5	7
β-xylosidase	Glyco_hydro_3	PF00933	15	14
α-mannosidase	Glyco_hydro_38	PF01074	4	3
Pectate lyase	Pec_lyase	PF00544	30	13
Polygalacturonase	Glyco_hydro_28	PF00295	67	40
Pectin methylesterase	Pectinesterase	PF01095	68	55
Pectin acetylesterase	Pec_acetylest	PF03283	11	11
Chitinase	Glyco_hydro_18	PF00704	11	16
Chitinase [d]	Glyco_hydro_19	PF00182	14	28
Cellulase (Mannan-hydrolase)	Cellulase	PF00150	16	24
Xylanase	Glyco_hydro_10	PF00331	12	10
Laccase	Cu-oxidase	PF00394	41	43
Phenylalanine ammonia lyase (PAL)	PAL	PF00221	4	10
trans Cinnamate 4-hydroxylase (C4H)	P450	PF00067	255	266
4-coumarate:CoA ligase (4CL)	AMP-binding	PF00501	38	44
Coumarate 3-hydroxylase (C3H)	P450	PF00067	255	266
Caffeoyl-CoA *O*-methyltransferase	Methyltransf_3	PF01596	10	9
Cinnamoyl-CoA reductase (CCR)	adh_short,	PF00106	116	59
Cinnamoyl-CoA reductase (CCR) [e]	3Beta_HSD	PF01073	56	50
Ferulate 5-hydroxylase (F5H)	P450	PF00067	255	266
Caffeic acid *O*-methyltransferase (COMT)	Methyltransf_2	PF00891	16	40
Cinnamyl alcoholdehydrogenase (CAD)	ADH_zinc_N	PF00107	38	58

[a] Family names in this table are defined on the basis of the Pfam database (http://www.sanger.ac.uk/Software/Pfam/).
[b] Pfam ID and Pfam AC are from the Pfam database.
[c] Numbers of genes that possess the protein domain defined based on the Pfam database are shown.
[d] The chitinase family consists of two different protein groups characterized by different motifs.
[e] The cinnamoyl-CoA reductase family consists of two different protein groups characterized by different motifs.

and Kende 2002, Choi et al. 2003, Yokoyama et al. 2004) as well as the supplementary Table S2.

For the other gene families, we searched the rice genome database with the BLAST algorithm using the sequences of *Arabidopsis* genes as queries (http://riceblast.dna.affrc.go.jp/). Alternatively, we performed keyword searches of the rice databases using the conserved motif name for each family (http://RiceGAAS.dna.affrc.go.jp/rgadb/). Because the rice genome database is not fully classified at present, large numbers of

highly duplicated nucleotide sequences were often obtained with the BLAST search. From the vast array of retrieved sequences, we manually removed duplications, retaining only unduplicated sequences. This procedure was repeated several times manually, whenever the database was updated. The present study is based on the March 2004 version of the rice genome sequence.

To classify each of the gene families, we adopted a two-step procedure based on exactly the same criterion as that

Table 3 Comparison of the sizes of the cell-wall-related gene families defined in the present work in *Arabidopsis* and rice

Family names [a]	Number of genes in each family [b]	
	Arabidopsis thaliana	*Oryza sativa*
Cellulose synthase	40	52
Callose/glucan synthase	12	10
Glycosyltransferase	45	41
α-xylosyltransferase	8	6
β-galactosyltransferase	39	30
α-fucosyltransferase	9	19
Xyloglucan endotransglucosylase/hydrolase	33	29
Expansin	35	48
β-1,4-glucanase	25	21
β-1,3-glucanase	52	57
α-fucosidase	1	2
β-galactosidase	19	15
α-xylosidase	5	7
β-xylosidase	15	14
α-mannosidase	4	3
Pectate lyase	30	13
Polygalacturonase	67	40
Pectin methylesterase	68	55
Pectin acetylesterase	11	11
Chitinase	25	44
Cellulase (Mannan-hydrolase)	16	24
Xylanase	12	10
Laccase	41	43
Phenylalanine ammonia lyase (PAL)	4	10
Cinnamate 4-hydroxylase (C4H)	1	4
4-coumarate:CoA ligase (4CL)	13	13
Coumarate 3-hydroxylase (C3H)	3	1
Caffeoyl-CoA *O*-methyltransferase	10	9
Cinnamoyl-CoA reductase (CCR)	7	12
Ferulate 5-hydroxylase (F5H)	2	3
Caffeic acid *O*-methyltransferase(COMT)	14	8
Cinnamyl alcoholdehydrogenase (CAD)	9	11

[a] Family names and genes within individual families were defined on the basis of the genome databases of *Arabidopsis* (http://www.Arabidopsis.org/) and rice (http://riceblast.dna.affrc.go.jp/, http://RiceGAAS.dna. affrc.go.jp/rgadb/), references (Holland et al. 2000, Yokoyama et al., 2001, Li et al. 2002), and the CaZy database (Coutinho and Henrissat 1999, http://afmb.cnrs-mrs.fr/˜pedro/CAZY).
[b] Numbers of genes in individual gene families of *Arabidopsis* and rice defined based on the criteria used in the present study are shown.

adopted for the *Arabidopsis* gene families. The first step in the classification was based on the conserved motif, whereby genes were classified by examining whether the putative protein contained the conserved motif found in the corresponding *Arabidopsis* protein encoded by a gene of a defined gene family (Table 2). Using this procedure, 465 rice genes were identified and classified into 20 families.

In some cases, a certain motif was conserved among several distinct families. For example, the AMP-binding domain is found not only in members of the 4-coumarate:CoA ligases

(4CL) family but in apparently unrelated protein families, including the long-chain fatty acid Co-A ligases and acetyl-CoA synthetases. In rice, 44 proteins were found to possess this motif (cf. Table 2 and Supplementary Fig. S1a, website: www.pcp.oupjournals.org). To circumvent this problem, we adopted a second-step classification that used full amino-acid sequence similarities as the criterion with which these family members were defined. This classification, together with the first-step motif-based classification, revealed the presence of 13 *Arabidopsis* genes and 13 rice genes encoding 4CL proteins

(Table 3 and Supplementary Fig. S1a). Furthermore, this classification revealed that the 4CL protein family can be classified into two subgroups based on phylogenetic trees (Fig. S1a). Cinnamyl alcohol dehydrogenase (CAD), cinnamoyl-CoA reductase (CCR), *trans*-cinnnamate 4-hydroxylase (C4H), *p*-coumarate 3-hydroxylase (C3H) and ferulate 5-hydroxylase (F5H) were also classified using the two-step classification system, as shown in supplementary Fig. S1, S2. We finally identified a total of 665 rice genes, classified in the 32 gene families by the two-step classification procedure. The numbers of genes defined as belonging to individual gene families are shown in Table 3. The classified amino-acid sequence data for the 665 genes of the 32 families are given in Table S2 of the supplementary table.

Comparison of Cell-wall-related Genes of Rice and Arabidopsis

Since the divergence of the monocotyledons and dicotyledons about 180–240 million years ago, both groups have independently undergone one or more polyploidization events during their evolution (Wolfe et al. 1989, Goremykin et al. 1997, Soltis et al. 2002). Their different evolutionary paths via genome shuffling have been attributed to variations in genome size and numbers of genes, which are estimated to be 28,000 for *Arabidopsis* and 32,000–62,000 for rice (Sasaki and Sederof 2003). Given that rice contains about two-fold more genes than *Arabidopsis*, it is noteworthy that the number of members in most of the cell-wall-related gene families defined in the present studies in *Arabidopsis* is similar to the number in rice (Table 3). This implies the presence of some common structures conserved in both the type I and type II walls, and that similar numbers of genes are required for the construction and maintenance of those common structures. In this context, it is interesting that putative orthologs with highly similar amino-acid sequences are often found in both plant genomes (cf. Supplementary Fig. S3). It is likely that the approximate number and organization of cell-wall-related genes might have remained unchanged in both evolutionary pathways after the monocotyledons and dicotyledons diverged.

Pectic enzymes and phenylpropanoid metabolism—In some gene families, the family size differs markedly between *Arabidopsis* and rice (Table 3 and Supplementary Fig. S3). *Arabidopsis* contains more genes identified as members of the pectate lyase (30 genes), polygalacturonase (67 genes) and pectin methylesterase (68 genes) families than rice does (13 genes, 40 genes and 55 genes, respectively) (Table 3). These enzymes are considered to be involved in the metabolism of pectin, which is abundant in type I cell walls. If the mode of enzymic action and the regulatory system of a multi-gene family encoding enzymes have diversified, then the size of the gene family should reflect the diversity or complexity of the biological processes in which the gene family is involved. The larger pectic-enzyme families in *Arabidopsis* relative to those of rice might mirror their diversified roles in type I walls.
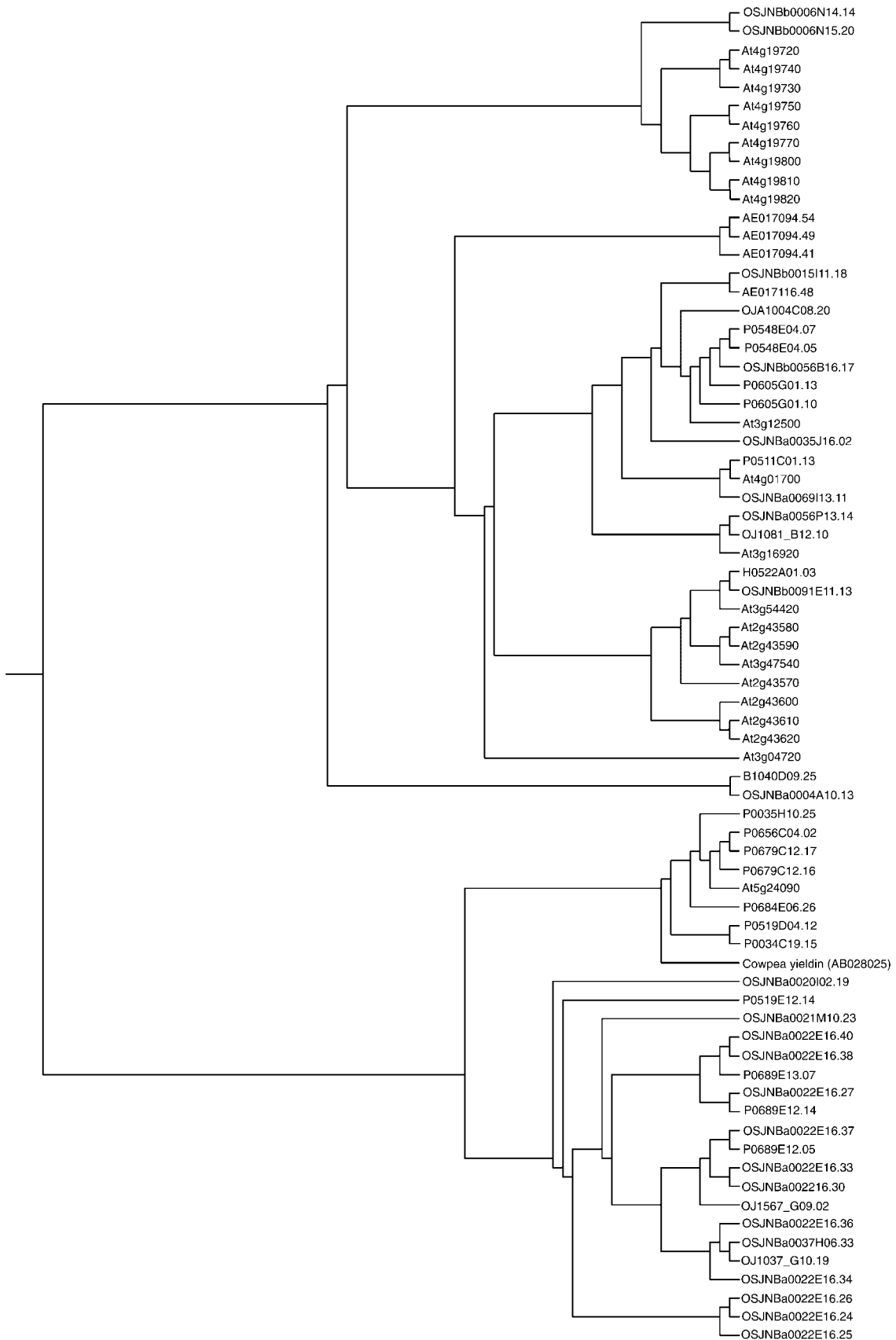
Conversely, rice has more genes for phenylalanine ammonia lyase (PAL) (10 genes), C4H (four genes) and CCR (12 genes) than does *Arabidopsis* (four genes, one gene and seven genes, respectively) (Table 3 and Supplementary Fig. S1, S2). These enzyme families are considered to be essential for the conversion of phenylalanine to hydroxycinnamic acids, most of which will be bound to lignin (Lam et al. 2001), or esterified to the arabinosyl residues of glucuronoarabinoxylans in type II walls (Nishitani and Nevins 1989). Esterified glucuronoarabinoxylans function as major cross-linking glycans in type II walls. Thus, the larger family of these enzymes in rice seems to reflect structural features of the rice cell wall.

In this context, the relatively larger caffeic acid *O*-methyltransferase (COMT) family in *Arabidopsis* than that in rice is an apparent paradox, because enzymes encoded by this gene family are also involved in the metabolism of phenylpropanoids, which are abundant in rice. One possible explanation for this is that most COMT family genes encode enzymes that do not take part in the metabolism of phenylpropanoids. Because the substrates or modes of enzymic action of the whole complement of putative COMT proteins are not fully understood, their physiological functions have not actually been demonstrated. To address this issue, biochemical characterization of the individual proteins is required.

Chitinases and expansins—Whereas rice contains 44 chitinase family genes, *Arabidopsis* contains only 25 chitinase genes. Our phylogenetic analysis reveals that the chitinase proteins fall into two distinct subfamilies (Fig. 2). Some members of plant chitinases are involved in the defense response against invading fungi and other pathogens, the cell walls of which contain chitin (Meins et al. 1992). It is quite likely that some members of the rice chitinase gene family are involved in a wide spectrum of defense responses.

In addition to the defense response, some members of the chitinase gene family encodes proteins highly homologous to 'yieldin', a special protein which is implicated in the modulation of the mechanical properties of the cell wall in cow pea plants (Okamoto-Nakazawa et al. 2000a, Okamoto-Nakazawa et al. 2000b). Yieldin belongs to the second subfamily, which contains 27 rice genes and a single *Arabidopsis* gene (Fig. 2). Judging from full amino-acid sequence similarities, it is probable that at least some of the seven rice genes found in the second subfamily encode proteins with yieldin-like structural features, and hence yieldin-like protein functions. Given that only a single *Arabidopsis* gene belongs to this subfamily, this implies that the putative yieldin protein has diversified extensively in rice. On the other hand, the other chi-

Fig. 2 Phylogenetic comparison of chitinase family genes in rice and *Arabidopsis*. *Arabidopsis* genes are indicated by AGI numbers and rice genes are indicated by gene numbers assigned by RiceGAAS. Cowpea yieldin gene (AB028025) is included in the tree.

OSJNBb0006N14.14
OSJNBb0006N15.20
At4g19720
At4g19740
At4g19730
At4g19750
At4g19760
At4g19770
At4g19800
At4g19810
At4g19820
AE017094.54
AE017094.49
AE017094.41
OSJNBb0015I11.18
AE017116.48
OJA1004C08.20
P0548E04.07
P0548E04.05
OSJNBb0056B16.17
P0605G01.13
P0605G01.10
At3g12500
OSJNBa0035J16.02
P0511C01.13
At4g01700
OSJNBa0069I13.11
OSJNBa0056P13.14
OJ1081_B12.10
At3g16920
H0522A01.03
OSJNBb0091E11.13
At3g54420
At2g43580
At2g43590
At3g47540
At2g43570
At2g43600
At2g43610
At2g43620
At3g04720
B1040D09.25
OSJNBa0004A10.13
P0035H10.25
P0656C04.02
P0679C12.17
P0679C12.16
At5g24090
P0684E06.26
P0519D04.12
P0034C19.15
Cowpea yieldin (AB028025)
OSJNBa0020I02.19
P0519E12.14
OSJNBa0021M10.23
OSJNBa0022E16.40
OSJNBa0022E16.38
P0689E13.07
OSJNBa0022E16.27
P0689E12.14
OSJNBa0022E16.37
P0689E12.05
OSJNBa0022E16.33
OSJNBa002216.30
OJ1567_G09.02
OSJNBa0022E16.36
OSJNBa0037H06.33
OJ1037_G10.19
OSJNBa0022E16.34
OSJNBa0022E16.26
OSJNBa0022E16.24
OSJNBa0022E16.25

tinase subfamily contains almost equal numbers of rice and *Arabidopsis* genes. The remarkable diversity of the yieldin-containing subfamily in rice might suggest a diversification of the roles of the 'putative yieldin' functions in rice. Biochemical studies of these putative yieldins in rice would be of great interest, considering the mode of cell-wall expansion in monocotyledonous plants.

Expansins are another category of cell-wall proteins that are implicated in cell-wall loosening. This class of proteins is thought to interact with the cellulose-matrix framework of the cell wall, thereby modulating the stress–relaxation process of the wall under acidic conditions, although the target of the proteins is not specified (reviewed by Cosgrove 2000, Cosgrove et al. 2002). Interestingly, expansin genes are more abundant in rice than in *Arabidopsis* (Table 3).

This differential organization of the genes encoding expansins and yieldins in the two plant species might mirror different modes of cell-wall expansion in type I and II cell walls. It is also possible that the differences in family size may reflect divergence in the morphology or cell type of the two plant species. The fact that expansin genes exhibit organ-, tissue- and cell-type-specific expression profiles supports this proposition (Cosgrove et al. 2002).

Xyloglucan-related genes—In the present comparative analysis of the genes involved in the fundamental framework of the cell wall, we took advantage of the fact that the structural features differ greatly in rice and *Arabidopsis*, as does the abundance of xyloglucan. Xyloglucan in the type I cell wall is composed of a β(1,4)-glucan backbone that is substituted at position 6 with xylosyl residues in a regular pattern. Some xylosyl residues are substituted with galactosyl residues, some of which are further substituted with fucosyl residues in *Arabidopsis* (Zablackis et al. 1995). In rice cell walls, xylosyl residues are less substituted with galactosyl residues and are never substituted with fucosyl residues (Kato et al. 1982). Several classes of cell-wall-related enzymes have been identified as being involved in the construction and disassembly of xyloglucans. These gene-family-encoded enzymes include xyloglucan α-xylosyltransferase, xyloglucan β-galactosyltransferase, xyloglucan α-fucosyltransferase, XTH, α-fucosylase, β-galactosidase and α-xylosidase.

Whereas all the *Arabidopsis* XTH family members thus far examined exhibit substrate specificity for xyloglucans (Nishitani and Tominaga 1992, Campbell and Braam 1999), other families, such as the α-xylosyltransferase, β-galactosyltransferase and α-fucosyltransferase families, contain only one gene encoding a xyloglucan-specific transferase (Faik et al. 2002, Madson et al. 2003, Perrin et al. 1999, Sarria et al. 2001).

Therefore, phylogenetic analysis of these gene families can be used to identify potential rice orthologs for each of the *Arabidopsis* β-galactosyltransferase and α-xylosyltransferase genes (Fig. 3A, 3B).

On the other hand, no rice α-fucosyltransferase is more similar to *Arabidopsis* xyloglucan fucosyltransferase1 (AtFUT1) than to AtFUT3-5, which exhibits no fucosylation activity to xyloglucan (Sarria et al. 2001) (Fig. 3C). Therefore, we predict that rice does not contain a gene encoding xyloglucan-specific α-fucosyltransferase, the enzyme that adds the terminal fucosyl residue to the galacotosyl residue on xyloglucans. This result is consistent with the structural features of rice xyloglucans, which lack fucosylation of the galactosyl side chains (Kato et al. 1982).

Xyloglucan endotransglucosylase/hydrolases (XTH) are a class of enzymes capable of mediating the splitting and reconnection of xyloglucan cross-links (reviewed by Nishitani 1997, Rose et al. 2002). They are considered essential for cell-wall dynamics, which encompass the construction, modification and maintenance of the cell-wall architecture (Fry et al. 1992, Nishitani and Tominaga 1992). Despite the considerably less abundant xyloglucan in the rice cell wall relative to that of *Arabidopsis* (Yokoyama and Nishitani 2001), the number of rice *XTH* genes is very similar to that of *Arabidopsis*. For a discussion on this surprising fact, and of their comprehensive expression data for rice, the reader is referred to Yokoyama et al. (2004) for fuller information.
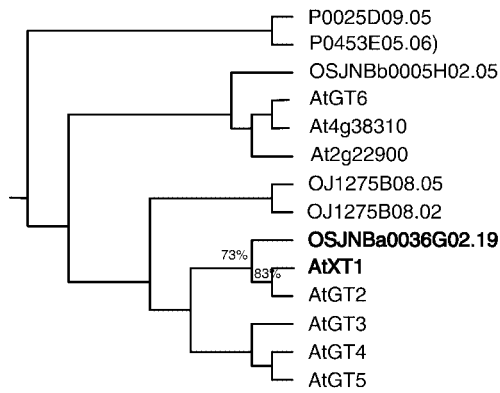
Future perspectives

Cell shape and organ morphology in plants are ultimately expressed through the construction and restructuring of cell walls. Because rice and *Arabidopsis* represent the two extremes of the flowering plants (Carpita and McCann 2000), with distinct cell-wall types, it is reasonable to focus on the genes involved in cell-wall dynamics as a first step in exploring the molecular basis of their morphological differences. Comparative analysis has revealed a set of gene families in which the sizes of the families differ greatly between rice and *Arabidopsis* (Table 3 and Supplemental Fig. S3). The present comparative genomic approach has revealed that the size and diversity of the cell-wall-related gene family mirrors the structural diversity or complexity of the cell-wall components that are the substrates or targets of the proteins encoded by that family of genes. This approach has also revealed the genomic basis of the differences between type I and type II walls, which are represented by *Arabidopsis* and rice, respectively. Furthermore, this comparative genome-based approach allowed us to identify putative rice orthologs simply by superimposing the phyloge-
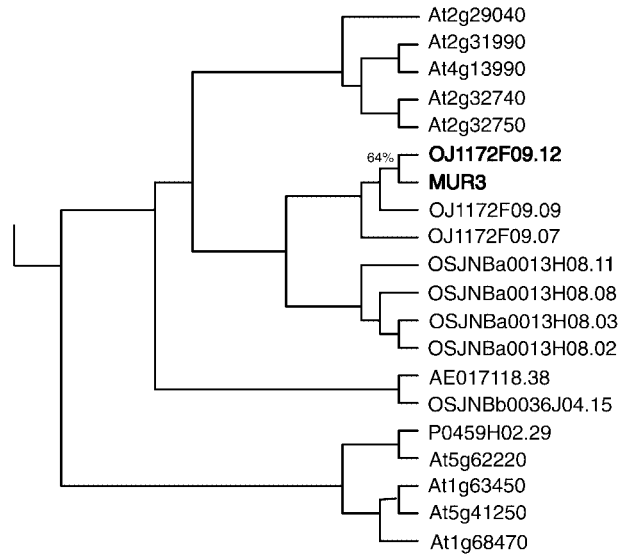
Fig. 3 Phylogenetic trees of xyloglucan-related gene families of *Arabidopsis* and rice. (A) Phylogenetic tree of the xylosyltransferase family in *Arabidopsis* and rice. (B) Phylogenetic tree of the galactosyltransferase family in *Arabidopsis* and rice. (C) Phylogenetic tree of the fucosyltransferase family in *Arabidopsis* and rice. Percentages indicate the percentage homology. *Arabidopsis* xylosyltransferase, galactosyltransferase and fucosyltransferase names are as previously published, and the other *Arabidopsis* genes are indicated by AGI numbers. Rice gene names are indicated by gene numbers assigned by RiceGAAS.
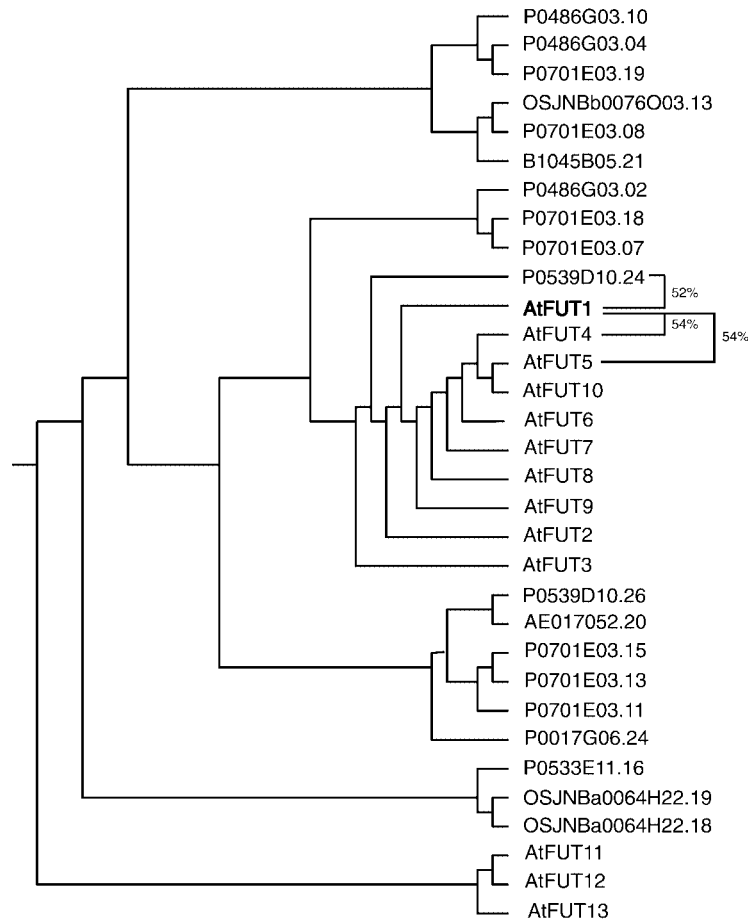
(A) Xylosyltransferase

```
┌─ P0025D09.05
│  └─ P0453E05.06)
│     ┌─ OSJNBb0005H02.05
│     │  ┌─ AtGT6
│     └──┤  At4g38310
│        └─ At2g22900
│  ┌─ OJ1275B08.05
│  └─ OJ1275B08.02
│     73% ┌─ OSJNBa0036G02.19
│     83% ├─ AtXT1
│         └─ AtGT2
│     ┌─ AtGT3
│     ├─ AtGT4
│     └─ AtGT5
```

(B) Galactosyltransferase

```
          ┌─ At2g29040
          ├─ At2g31990
          ├─ At4g13990
          ├─ At2g32740
          └─ At2g32750
     64% ┌─ OJ1172F09.12
          ├─ MUR3
          ├─ OJ1172F09.09
          └─ OJ1172F09.07
          ┌─ OSJNBa0013H08.11
          ├─ OSJNBa0013H08.08
          ├─ OSJNBa0013H08.03
          └─ OSJNBa0013H08.02
          ┌─ AE017118.38
          └─ OSJNBb0036J04.15
          ┌─ P0459H02.29
          ├─ At5g62220
          ├─ At1g63450
          ├─ At5g41250
          └─ At1g68470
```

(C) Fucosyltransferase

```
          ┌─ P0486G03.10
          ├─ P0486G03.04
          └─ P0701E03.19
          ┌─ OSJNBb0076O03.13
          ├─ P0701E03.08
          └─ B1045B05.21
          ┌─ P0486G03.02
          ├─ P0701E03.18
          └─ P0701E03.07
          ┌─ P0539D10.24
          │  52%
          ├─ AtFUT1
          │  54%          54%
          ├─ AtFUT4
          ├─ AtFUT5
          ├─ AtFUT10
          ├─ AtFUT6
          ├─ AtFUT7
          ├─ AtFUT8
          ├─ AtFUT9
          ├─ AtFUT2
          └─ AtFUT3
          ┌─ P0539D10.26
          ├─ AE017052.20
          ┌─ P0701E03.15
          ├─ P0701E03.13
          ├─ P0701E03.11
          └─ P0017G06.24
          ┌─ P0533E11.16
          ├─ OSJNBa0064H22.19
          └─ OSJNBa0064H22.18
          ┌─ AtFUT11
          ├─ AtFUT12
          └─ AtFUT13
```

netic trees of rice gene families on those of *Arabidopsis* gene families with well-defined biological and biochemical information (Fig. 2, 3 and Supplementary Fig. S3). The rapid and precise prediction of rice orthologs, in turn, will not only facilitate the elucidation of the individual functions of proteins in rice, but will allow insight into the evolutionary processes that brought about the biochemical differences in type I and type II walls. If there exist cell-wall genes with species-specific functions that make some cell walls different from others, then the phylogenetic analyses described in this article should help to identify those genes.

Although we have defined members of individual families, the enzymic activities or biological functions have not been fully demonstrated for most of the products of these genes. Biochemical characterization of cell-wall-related enzymes and proteins has been hampered by the difficulty of establishing an assay system that uses substrates specific for individual enzymic functions. The comparative genomic approach presented here obviously helps to narrow the possible enzymic functions of individual members of gene families, and should facilitate targeted well-planned biochemical experiments using predicted substrates and enzyme reaction systems. Such an approach, in combination with the metabolome of wall components, offers an opportunity to explore new aspects of cell walls, the elucidation of which might break fresh ground in the field of plant biology.

Supplementary Material—Supplementary material mentioned in the article is available to online subscribers at the journal website www.pcp.oupjournals.org.

Acknowledgments

References

Arabidopsis Genome Initiative (2000) *Nature* 408: 796–815.

Buell, C.R. (2003) *Plant Physiol.* 130: 1585–1586.

Campbell, P. and Braam, J. (1999) *Plant J.* 34: 327–338.

Carpita, N.C. (1996) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47: 445–476.

Carpita, N.C. and Gibeaut, D.M. (1993) *Plant J.* 3: 1–30.

Carpita, N.C. and McCann, M. (2000) *In* Biochemistry and Molecular Biology of Plants. Edited by Buchanan, B.B., Gruissem, W. and Jones, R.J. pp.25–109. American Society of Plant Biologists, Rockville, MA.

Choi, D., Le, Y., Cho, H-T. and Kende, H. (2003) *Plant Cell* 15: 1386–1398.

Cosgrove, D.J. (2000) *Nature* 407: 321–326.

Cosgrove, D.J., Li, L.C., Cho, H.-T., Hoffmann-Benning, S., More, R.C. and Blecker, D. (2002) *Plant Cell Physiol.* 43: 1436–1444.

Coutinho, P.M. and Henrissat, B. (1999) *In* Recent Advances in Carbohydrate Bioengineering. Edited by Gilbert, H.J., Davies, G., Henrissat, B. and Svensson, B. pp. 3–12. Royal Society of Chemistry, Cambridge, U.K.

de la Torre, F., Sampedro, J., Zarra, I. and Revilla, G. (2002) *Plant Physiol.* 128: 247–255.

Freshour, G., Clay, R.P., Fuller, M.S., Albersheim, P., Darvill, A.G. and Hahn, M.G. (1996) *Plant Physiol.* 110: 1413–1429.

Faik, A., Price, N.J., Raikhel, N.V. and Keegstra, K. (2002) *Proc. Natl Acad. Sci. USA* 99: 7797–7802.

Fry, S.C., Smith, R. C., Renwick, K. F., Martin, D. J., Dge, S.K. and Matthews, K. J. (1992) *Biochem. J.* 22: 821–828.

Fry, S.C. (2003) *New Phytologist* 161: 641–675.

Goremykin, V.V., Hansmann, S. and Martin, W.F. (1997) *Plant Syst. Evol.* 206: 337–351.

Goujon, T., Minic, Z., El Amrami, A., Lerouxel, O., Aletti, E., Lapierre, C., Joseleau, J. and Jouanin, L. (2003) *Plant J.* 33: 677–690.

Hazen, S.P., Scott-Craig, J.S. and Walton, J.D. (2002) *Plant Physiol.* 128: 336–340.

Holland, N., Holland, D., Helentjaris, T., Dhugga, K.S., Xoconostle-Cazares, B. and Delmer, D.P. (2000) *Plant Physiol.* 123: 1313–1323.

Hong, Z.L., Zhang, Z.M., Olson, J.M. and Verma, D.P.S. (2001) *Plant Cell* 13: 769–779.

Iiyama, K., Lam, T.B.T. and Stone, B.A. (1990) *Phytochemistry* 29: 733–737.

Ishii, T. (1997) *Plant Sci.* 127: 111–127.

Ishii, T., Matsunaga, T., Pellerin, P., O'Neill, M.A., Darvill, A. and Albersheim, P. (1999) *J. Biol. Chem.* 274: 13098–13104.

Johnson, K.L., Jones, B.J., Schultz, S.J., and Bacic, A. (2003) *In* The Plant Cell Wall. Annual Plant Reviews. Vol. 8. Edited by Rose, J.K.C. pp. 111–154. CRC Press LLC, Boca Raton, FL.

Kato, Y., Ito, S., Iki, K. and Matsuda, K. (1982) *Plant Cell Physiol.* 23: 351–364.

Kobayashi, M. and Matoh, T. (1996) *Plant Physiol.* 110: 1017–1020.

Lam, T.B.T., Iiyama, K. and Stone, B.A. (1994) *Phytochemistry* 37: 327–333.

Lam, T.B.T., Kadoya, K. and Iiyama, K. (2001) *Phytochemistry* 57: 987–992.

Lee, Y. and Kende, H. (2002) *Plant Physiol.* 130: 1396–1405.

Li, O.Y., Darley, C.P., Ongaro, V., Fleming, A., Schipper, O., Baldauf, S.L. and McQueen-Mason, S.J. (2002) *Plant Physiol.* 128: 854–864.

Madson, M., Dunand, C., Li, X., Verma, R., Vanzin, G.F., Caplan, J., Shoue, D.A., Carpita, N.C. and Reiter, W.-D. (2003) *Plant Cell* 15: 1662–1670.

Martin, C., Bhatt, K. and Baumann, K. (2001) *Curr. Opin. Plant Biol.* 4: 540–549.

Marry, M., McCann, M.C., Kolpak, F., White, A.R., Stacey, N.J. and Roberts, K. (2001) *J. Science Food Agric.* 80: 17–28.

Meins, F.J., Neuhaus, J.-M., Sperisen, C. and Ryals, J. (1992) *In* Genes Involved in Plant Defense. Edited by Boller, T. and Meins, F. pp. 245–282. Springer-Verlag, Berlin.

Nishitani, K. (1997) *Int. Rev. Cytol.* 173: 157–206.

Nishitani, K. (2002) *J. Plant Res.* 115: 303–307.

Nishitani, K. and Nevins, D.J. (1989) *Plant Physiol.* 91: 242–248.

Nishitani, K. and Nevins, D.J. (1991) *J. Biol. Chem.* 266: 6539–6543.

Nishitani, K. and Tominaga, R. (1992) *J. Biol. Chem.* 267: 21058–21064.

Ohmiya, Y., Nakai, T., Park, Y.W., Aoyama, T., Oka, A., Sakai, F. and Hayashi, T. (2003) *Plant J.* 33: 1087–1097.

Okamoto-Nakazawa, A., Nakamura, T. and Okamoto, H. (2000a) *Plant Cell Environ.* 23: 145–154.

Okamoto-Nakazawa, A., Takahashi, K., Kido, N., Owaribe, K.T. and Okamoto, H. (2000b) *Plant Cell Environ.* 23: 155–164.

O'Neill, M.A., Eberhard, S., Albersheim, P. and Darvill, A.G. (2001) *Science* 294: 795–797.

Perrin, R.M., DeRocher, A.E., Bar-Peled, M., Zeng, W.Q., Norambuena, L., Orellana, A., Raikhel, N.V. and Keegstra, K. (1999) *Science* 284: 1976–1979.

Raes, J., Rohde, A., Christensen, J.H., Van de Peer, Y. and Boerjan, W. (2003) *Plant Physiol.* 133: 1051–1071.

Ridley, B.L., O'Neill, M.A. and Mohnen, D. (2001) *Phytochemistry* 57: 929–967.

Rose, J.K.C., Braam, J., Fry, S.C. and Nishitani, K. (2002) *Plant Cell Physiol.* 43: 1421–1435.

Sampedro, J., Siero, C., Revilla, G., Gonáez-Villa, T. and Zarra, I. (2001) *Plant Physiol.* 126: 910–920.

Sarria, R., Wagner, T.A., O'Neill, M.A., Faik, A., Wilkerson, C.G., Keegstra, K. and Raikhel, N.V. (2001) *Plant Physiol.* 127: 1595–1606.

Sasaki, T. and Burr, B. (2000) *Curr. Opin. Plant Biol.* 3: 138–141.

Sasaki, T. and Sederof, R.R. (2003) *Curr. Opin. Plant Biol.* 6: 97–100.

Soltis, P.S., Soltis, D.E., Savolainen, V., Crane, P.R. and Barraclough, T.G. (2002) *Proc. Natl Acad. Sci. USA* 99: 4430–4435.

Suzuki, M., Kato, A., Nagata, N. and Komeda, Y. (2002) *Plant Cell Phisiol.* 43: 759–767.

Willats, W.G.T., McCartney, L., Mackie, W. and Knox, J.P. (2001) *Plant Mol. Biol.* 47: 9–27.

Wolfe, K.H., Gouy, M.L., Yang, Y.W., Sharp, P.M. and Li, W.H. (1989) *Proc. Natl Acad. Sci. USA* 86: 6201–6205.

Yoshida-Shimokawa, T., Yoshida, S., Kakegawa, K., Ishii, T. (2001) *Planta* 212: 470–474.

Yokoyama, R. and Nishitani, K. (2001) *Plant Cell Physiol.* 42: 1025–1033.

Yokoyama, R., Rose, J.K.C. and Nishitani, K. (2004) *Plant Physiol.* 134: 1088–1099.

Zablackis, E., Huang, J., Muller, B., Darvill, A.G. and Albersheim, P. (1995) *Plant Physiol.* 107: 1129–1138.

Zhong, R., Kays, S.J., Schroeder, B.P. and Ye, Z-H. (2002) *Plant Cell* 14: 165–179.