https://doi.org/10.1093/plphys/kiad267 PLANT PHYSIOLOGY 2023: 1–22

Plant Physiology®

Research Article

Highly differentiated genomic properties underpin the different cell walls of Poaceae and eudicots

Francesco Pancaldi [1], Michael Eric Schranz [2], Eibertus N. van Loo [1] and Luisa M. Trindade [1,*]

1 Plant Breeding, Wageningen University & Research, Droevendaalsesteeg 1, 6708PB, Wageningen, The Netherlands
2 Biosystematics group, Wageningen University & Research, Droevendaalsesteeg 1, 6708PB, Wageningen, The Netherlands

*Author for correspondence: luisa.trindade@wur.nl
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/General-Instructions) is Luisa M. Trindade.

Abstract

Plant cell walls of Poaceae and eudicots differ substantially, both in the content and composition of their components. However, the genomic and genetic basis underlying these differences is not fully resolved. In this research, we analyzed multiple genomic properties of 150 cell wall gene families across 169 angiosperm genomes. The properties analyzed include gene presence/absence, copy number, synteny, occurrence of tandem gene clusters, and phylogenetic gene diversity. Results revealed a profound genomic differentiation of cell wall genes between Poaceae and eudicots, often associated with the cell wall diversity between these plant groups. For example, overall patterns of gene copy number variation and synteny were clearly divergent between Poaceae and eudicot species. Moreover, differential Poaceae–eudicot copy number and genomic contexts were observed for all the genes within the *BEL1-like HOMEODOMAIN 6* regulatory pathway, which respectively induces and represses secondary cell wall synthesis in Poaceae and eudicots. Similarly, divergent synteny, copy number, and phylogenetic gene diversification were observed for the major biosynthetic genes of xyloglucans, mannans, and xylans, potentially contributing to the differences in content and types of hemicellulosic polysaccharides differences in Poaceae and eudicot cell walls. Additionally, the Poaceae-specific tandem clusters and/or higher copy number of *PHENYLALANINE AMMONIA-LYASE*, *CAFFEIC ACID O-METHYLTRANSFERASE*, or *PEROXIDASE* genes may underly the higher content and larger variety of phenylpropanoid compounds observed in Poaceae cell walls. All these patterns are discussed in detail in this study, along with their evolutionary and biological relevance for cell wall (genomic) diversification between Poaceae and eudicots.

Introduction

All plant cells are surrounded by a cell wall, which mechanically supports plant growth and mediates plant–environment interactions (Somerville et al. 2004). The general plant cell wall architecture is conserved across angiosperms. Typically, primary cell walls are formed during plant cell expansion and are composed of cellulose, different hemicellulosic polysaccharides, pectins, and structural proteins (Somerville et al. 2004; Sarkar et al. 2009). When cell growth ceases, secondary cell walls can be synthesized, in which pectins are absent or present at very low levels, while lignin is a major component (Zhong et al. 2019).

Despite the general cell wall patterns, extensive cell wall compositional and structural variation exists between plant taxa (Vogel 2008; Sarkar et al. 2009). A major differentiation is between type I cell walls, which are specific to eudicots and noncommelinoid monocots, and type II cell walls, which are found in grasses (Poaceae) (Vogel 2008). Type I cell walls contain xyloglucan (XyG) as the major hemicellulose, along with relatively large amounts of (gluco)mannans. Moreover, they display large quantities of pectins and structural proteins

Received February 09, 2023. Accepted April 03, 2023. Advance access publication May 4, 2023

Open Access

(Vogel 2008; Penning et al. 2019), while the total amount of phenylpropanoids in type I (secondary) cell walls is typically lower than that in the type II ones (Vogel 2008; Guerriero et al. 2016). Conversely, type II cell walls contain xylans as the major hemicellulose, while XyGs and (gluco)mannans are only found in trace amounts. Additionally, type II cell walls contain large quantities of (1,3; 1,4)-β-glucans (also termed mixed linkage glucans—MLGs), a hemicellulose polysaccharide mainly restricted to grasses (Fincher and Stone 1986; Vogel 2008). Regarding nonhemicellulosic polysaccharides, type II cell walls contain much lower amounts of pectins and structural proteins than type I cell walls (Vogel 2008; Penning et al. 2019). Finally, the total content of phenylpropanoid compounds of type II (secondary) cell walls is usually higher than that of type I cell walls (Vogel 2008; Guerriero et al. 2016). These phenylpropanoid molecules include lignin, ferulic acid, *p*-coumaric acid, and tricin, which overall provide rigidity and mechanical strength to cell walls, cross-link cell wall polysaccharides, and mediate plant defense to (a)biotic stresses (Ralph 2010; De Oliveira et al. 2015; Hatfield et al. 2017; Chandrakanth et al. 2023).

The differences between type I and type II cell walls are not absolute, and cell wall compositional variation exists within both Poaceae and eudicots (Vogel 2008; Burton and Fincher 2014). Nevertheless, type I and type II cell walls represent established valuable models to describe the marked differentiation of cell walls observed between these plant groups (Carpita and Gibeaut 1993; Vogel 2008; Carpita and McCann 2020). This aspect, along with the agricultural relevance of Poaceae species and the importance of cell wall composition for the industrial utilization of plant biomass, makes the understanding of the genetics underlying type I and type II cell wall differences a valuable research target (Burton and Fincher 2012; Pancaldi and Trindade 2020). This latter aspect is however far from being resolved. For MLGs, it was possible to associate their mostly Poaceae-specific occurrence with the grass-specific presence of *Cellulose synthase-like F* (*CslF*) genes (Burton et al. 2006). Notwithstanding, the complexity of cell wall biosynthesis, which relies on a multitude of genes with pleiotropic effects, hampers the elucidation of the genetic basis of type I and type II cell wall differentiation (Burton and Fincher 2012; Yokoyama 2020). In this context, Penning et al. (2019) analyzed the occurrence of different carbohydrate-active enzymes in the genomes of Arabidopsis (*Arabidopsis thaliana*), maize (*Zea mays*), and rice (*Oryza sativa*), showing their ubiquitous presence in all the genomes, irrespectively of the species' cell wall type. This result indicates that, with few exceptions such as *CslF*, gene presence–absence variation is not sufficient to explain type I and type II cell wall differentiation (Penning et al. 2019). Other researchers studied the regulatory differences between Poaceae and eudicot cell walls, showing that the cell wall regulatory machinery is overall conserved across these clades, even if few but relevant differences were observed [see Rao and Dixon (2018) for a review on this topic]. For example, the master transcription factor *BEL1-like HOMEODOMAIN 6* (*BLH6*) has opposite function in eudicots and Poaceae, by repressing and inducing secondary cell wall deposition, respectively (Hirano et al. 2013; Liu et al. 2014; Rao and Dixon 2018). This observation highlights the importance of comparative genetic research to understand the cell wall differentiation between Poaceae and eudicots. However, these types of studies are scarce, and the question of what is the genetic–evolutionary basis of type I and type II cell walls is currently largely unresolved (Yokoyama 2020).

In this research, the study of the genetics underlying type I and type II cell walls was tackled from the perspective of the genomic properties of cell wall genes across Poaceae and eudicots. This was performed by analyzing patterns of gene copy number, synteny, tandem gene clusters, and phylogenetic relatedness of 150 different cell wall gene families across 169 angiosperm genomes representing plant cell wall diversity. This approach is on purpose large-scale and genomic-oriented, since recent research showed that such methodologies are very powerful to investigate complex genetic patterns at the basis of plant diversity (Zhao et al. 2017; Kerstens et al. 2020). To conclude, the data produced were on purpose analyzed in comparisons between grasses and eudicots, as they are the taxonomic groups representing type I–type II cell wall differences.

Results

Extensive copy number variation within the cell wall gene families of Poaceae and eudicots cell walls

Gene copy number was quantified across the 150 target cell wall gene families in each of the 169 angiosperm genomes (Supplemental Tables S1 and S2 list the genes and genomes used). The data obtained were plotted onto a heatmap (Fig. 1A) and analyzed by principal component analysis (PCA; Figs. 1, B and C). Results revealed extensive copy number variation (CNV), across both gene families and plant species. Specifically, CNV between gene families ranged from singleton families in the majority of the species surveyed (e.g. the master transcription factor *EARLY2 FACTOR c*, *E2Fc*; *KATANIN* genes involved in xylem development, *KTN*; or the homologs of Arabidopsis *ALTERED XYLOGLUCAN 9*, *AXY9*) to families containing 100+ gene copies per genome on average (e.g. peroxidases, *PRX*; polygalacturonases, *PG*; and pectin esterases, *PE*) (Fig. 1A and Supplemental Table S3). Moreover, CNV across plants highlighted species displaying deviations of gene copy number for several cell wall gene families. On the one hand, this was observed for specific plant clades known to share taxon-specific genome duplications, as the Salicaceae and Cucurbitaceae families or the *Gossypium* and *Brassica* genera (dashed boxes in Fig. 1A). On the other hand, copy number differences were also revealed between Poaceae and several eudicot species for multiple cell wall gene families (solid black boxes in Fig. 1A and Supplemental Table S4). Remarkably, some of

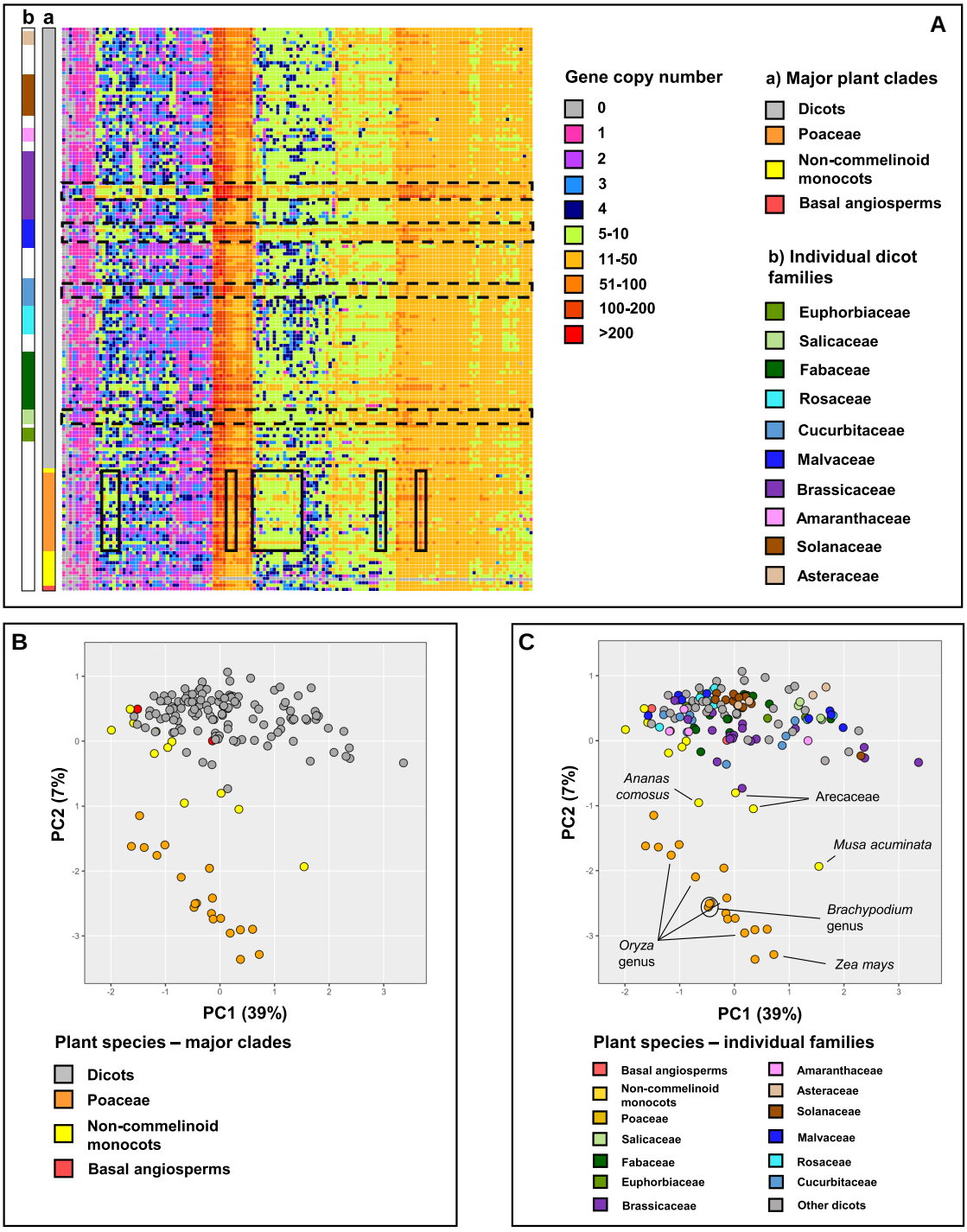

Figure 1. Copy number properties of cell wall genes. **A)** Heatmap showing the large CNV of cell wall gene families (columns) across the 169 genomes of the study (rows). Colors of heatmap cells represent the copy number of each species–gene combination. Left to the heatmap, genomes are categorized based on taxonomic clades. Dashed boxes indicate species displaying CNV for most cell wall gene families compared with other taxonomic groups. Solid boxes indicate groups of gene families showing CNV between Poaceae and other angiosperms (mostly eudicots). **B)** and **C)** PCA plots of the 169 genomes of the study based on CNV patterns. The 2 plots display PCA results at the level of general plant clades **(B)** and of individual plant families **(C)**, respectively.

these gene families are particularly relevant for type I to II cell wall differentiation. For example, they include the *IRREGULAR XYLEM* (*IRX*) *9* and *14* genes involved in xylan synthesis (CAZy GT43 family); the *PECTIN METHYL-ESTERASES* and associated inhibitors (*PME* and *PMEI*), which affect the pectin content of cell walls; or the *PHENYLALANINE AMMONIA-LYASE* genes (*PAL*), which catalyze the first step of the phenylpropanoid/lignin pathway.

To better elucidate the relationship between CNV and angiosperm cell wall diversity, CNV data were also analyzed by PCA (Figs. 1, B and C, and Supplemental Table S5). The first 2 components capture 46% of the total CNV across all genes and species and clearly separate Poaceae from the other angiosperms. This highlights the relevance of cell wall gene CNV as a major genomic property underlying Poaceae and eudicot cell wall diversity. Moreover, PCA highlighted a generally large level of CNV also within Poaceae and eudicots themselves, as these species are extensively spread within the plots in Fig. 1, B and C. This is in line with both the relatively large Poaceae cell wall diversity (within the framework of the type II cell wall) (Vogel 2008; Burton and Fincher 2014) and with the numerous (segmental) duplications and genomic translocations experienced by the different Poaceae species during grass evolution (Wang et al. 2015; Lee et al. 2020). Besides Poaceae, the majority of eudicots and noncommelinoid monocots form a single large cluster in the plots in Fig. 1, B and C. The different plant families within this cluster do not display further clear grouping patterns, suggesting that intrafamily CNV in eudicots can be large and of similar magnitude as for eudicots as a whole group.

Given the clear association between cell wall gene CNV and type I to II cell wall differentiation, statistical tests were performed to assess which gene families display significant copy number differences between Poaceae and eudicots. The results showed that 70 of the 150 cell wall gene families analyzed (47%) display significantly different copy number levels between these groups of plants (*t* test, alpha = 0.05; Supplemental Table S4). These 70 families were analyzed for the magnitude of significant differences, as well as for their relevance for cell wall differentiation between Poaceae and eudicots based on scientific literature. In this way, 20 particularly relevant families were identified (Table 1). These genes mediate critical steps in the biosynthesis of cell wall components that are variable between Poaceae and eudicot cell walls and display CNV patterns in line with such differences, mainly in a perspective of gene dosage variability. The detailed explanations of the patterns found are reported in Table 1.

Cell wall gene synteny reveals conserved and divergent genomic gene contexts between Poaceae and eudicots

Recent studies highlighted how differential gene synteny across plants might underlie trait variability and evolutionary adaptations, through changes of genomic gene contexts that can impact gene functions (Dewey 2011; Zhao et al. 2017; Kerstens et al. 2020; Pancaldi et al. 2022a). To test if this is also the case for cell wall genes and type I and type II cell walls, the syntenic conservation of the 150 cell wall gene families across the 169 genomes of the study was examined. Gene synteny was analyzed using the network approach developed by Zhao and Schranz (2017), which organizes large sets of syntenic genes from diverse genomes into networks where nodes represent genes and edges intergenic synteny. This way, synteny patterns can be dissected with statistical methods for networks analysis, including network decomposition into communities of nodes displaying significantly higher synteny within than between communities. Such syntenic communities represent independent gene positional configurations—or genomic contexts—occurring in specific groups of species.

The synteny analysis of the 320,005 cell wall genes (from the 150 gene families and 169 genomes) yielded a synteny network with 258,316 different nodes: 80.7% of the initial cell wall genes (Supplemental Data Set 1). The large number of cell wall genes retained in the network as nodes indicates a very high level of syntenic conservation of cell wall genes. Specifically, such percentage is higher than that detected in other studies of unrelated gene families (Zhao et al. 2017; Kerstens et al. 2020), but is in line with the level of synteny observed for the *Cellulose synthase* gene superfamily, a smaller distinct set of cell wall genes, in previous studies (Schwerdt et al. 2015; Pancaldi et al. 2022a). Moreover, each cell wall gene in the synteny network is on average syntenic with >50 other genes, irrespectively of it being a Poaceae or eudicot gene. This is remarkable, as while extensive gene synteny can be expected for any gene type in Poaceae (Gale and Devos 1998), this property is much less common in eudicots (Zhao and Schranz 2019).

Synteny network decomposition yielded 7,634 different syntenic communities of at least 4 nodes, each representing a specific genomic context of a specific cell wall gene type, conserved in a specific group of species (Fig. 2). These syntenic communities were taxonomically and functionally profiled, revealing 3 main community groups. The first group contains 597 communities consisting of 87,905 total cell wall genes whose positional genomic organization is conserved across all or most of the angiosperms analyzed, including Poaceae, eudicots, noncommelinoid monocots, and sister species such as *Amborella trichopoda* (black box in Fig. 2). Interestingly, functional profiling showed that these widely conserved syntenic communities contain a large proportion of transcription factor genes and genes synthesizing lignin and other cell wall phenylpropanoids (66% and 52% of all the syntenic transcription factors and lignin/phenylpropanoid genes, respectively). Conversely, the fraction of cellulose and hemicellulose genes within these communities is much smaller (14% and 33%, respectively). The second group includes 1,678 communities containing 98,538 total genes that display completely divergent syntenic conservation between eudicots, Poaceae, and noncommelinoid monocots (green boxes in Fig. 2). Of these communities, 718 are largely conserved across eudicots and noncommelinoid monocots, but not in Poaceae; 492 communities are conserved within Poaceae but not in eudicots and noncommelinoid monocots; 468 communities are conserved within Poaceae and most noncommelinoid monocots, but not in eudicots. Concerning functional profiling, hemicellulose-related gene functions are the ones with the highest proportion of genes

Table 1. Gene families displaying significant copy number differences between Poaceae and eudicots that appear particularly relevant for the differentiation of type I and type II cell walls in view of their known cell wall function and the direction of the copy number patterns observed

Gene family	Eudicot mean copy number	Poaceae mean copy number	Functional relevance of copy number pattern for type I and type II cell walls	References
Mannanases (MAN)	4	1	Mannanases are associated with mannan synthesis and remodeling. Mannans are common in eudicots but seldom in Poaceae.	Zhong et al. (2019), Vogel (2008)
Irregular xylem 9 (IRX9/9L; GT43)	6	13	Central gene for xylan synthesis. Complexes with IRX10/10L and IRX14/14L. Xylans are much more abundant in Poaceae than eudicots.	Zhong et al. (2019), Vogel (2008)
Irregular xylem 10 (IRX10/10L; GT47)	33	26	Central gene for xylan synthesis. Complexes with IRX10/10L and IRX14/14L. Xylans are much more abundant in Poaceae than eudicots.	Zhong et al. (2019), Vogel (2008)
Irregular xylem 14 (IRX14/14L; GT43)	5	7	Central gene for xylan synthesis. Complexes with IRX10/10L and IRX14/14L. Xylans are much more abundant in Poaceae than eudicots.	Zhong et al. (2019), Vogel (2008)
Cellulose synthase-like G (CslG)	3	1	At least 1 CslG transfers glucuronic acid (GlcA) during the synthesis of saponins, and CslGs could therefore act as GlcA transferases, also in the context of cell wall. GlcA substitutions of xylans highly differ between eudicots (highly substituted xylans) and grass (poorly substituted xyland).	Jozwiak et al. (2020), Pena et al. (2016)
PARVUS	26	16	Involved in GlcA substitution of xylans. GlcA-enriched xylans are more abundant in eudicots.	Pena et al. (2016), Vogel (2008), Zhu et al. (2017)
BEAT/AHCT/HCBT/DAT acyl-transferase (BAHD)	3	11	BAHD are involved in ferulic acid and *p*-coumaric acid substitution of several cell wall components. These molecules are specific to grass cell walls.	Vogel (2008), Bartley et al. (2013), Molinari et al. (2013), Chandrakanth et al. (2023)
Glycosyl-transferase 61 (GT61)	5	28	Genes mainly involved in both arabinosylation (grasses and dicots) and feruloylation (grasses) of xylans.	Feijao et al. (2022), Cenci et al. (2018), Anders et al. (2012)
Xylogalacturonan-deficient (XGD)	22	9	Central gene for the synthesis of backbone xylogalacturonan backbones. Pectins are much higher in eudicot primary cell walls compared with grass ones.	Atmodjo et al. (2013), Vogel (2008)
GAUT-like proteins (GATL)	11	5	Gene involved in pectin synthesis, even if with unclear role. Pectins are much higher in eudicot primary cell walls compared with grass ones.	Atmodjo et al. (2013), Vogel (2008)
Phenylalanine ammonia-lyase (PAL)	5	10	First gene of the phenylpropanoid pathway, affecting the total amount of substrates streamed to phenylpropanoid synthesis. The content of cell wall phenylpropanoids is higher in Poaceae cell walls.	Vogel (2008), Zhong et al. (2019)
Peroxidase (PRX)	95	152	Mediate *in muro* deposition of lignin, ferulic acid, and extensins. The content of cell wall phenylpropanoids, and specifically of ferulic acid, is higher in Poaceae cell walls.	Vogel (2008), De Oliveira et al. (2015), Zhong et al. (2019), Mishler-Elmore et al. (2021)
Cinnamoyl CoA reductase (CCR)	14	21	Central lignin gene shared by the branches of the lignin pathway leading to all monolignols. It can influence the final amounts of lignin and other phenylpropanoids in cell walls. The content of cell wall phenylpropanoids is higher in Poaceae cell walls.	Tamasloukht et al. (2011), Zhong et al. (2019)
Caffeic acid *O*-methyltransferase (COMT)	10	5	COMT genes push the lignin pathway towards synthesis of S- and G-lignin subunits instead of H-lignin. Poaceae have a higher amount of H-subunits.	Vogel (2008), Zhong et al. (2019)
Pectin lyase (PLY)	25	9	Central gene for pectin metabolism. Pectins are much higher in eudicot primary cell walls compared with grass ones.	Atmodjo et al. (2013), Vogel (2008)
Pectin methylesterase (PME)	75	41	Central gene for pectin metabolism. Pectins are	Atmodjo et al. (2013), Vogel (2008)

(continued)

Table 1. (continued)

Gene family	Eudicot mean copy number	Poaceae mean copy number	Functional relevance of copy number pattern for type I and type II cell walls	References
			much higher in eudicot primary cell walls compared with grass ones.	
Pectin methylesterase inhibitor (PMEI)	73	38	Central gene for pectin metabolism. Pectins are much higher in eudicot primary cell walls compared with grass ones.	Atmodjo et al. (2013), Vogel (2008)
Dynamin-related protein (DRP/ADL)	16	12	Genes likely involved in pectin trafficking to cell wall. Pectins are much higher in eudicot primary cell walls compared with grass ones.	Atmodjo et al. (2013), Vogel (2008)
BEL-like homeodeomain 6 (BLH6)	2	4	Transcription factor displaying divergent function in grasses (inducer of secondary cell wall synthesis) and eudicots (repressor of secondary cell wall synthesis).	Rao and Dixon (2018)
Expansins (EXP)	39	70	Involved in cell wall expansion by remodeling hemicellulose polysaccharides and pectins, as well as the interaction between hemicellulose and cellulose. Grass-specific β-expansins likely co-evolved with the specific features of grass xylans.	Vogel (2008), Atmodjo et al. (2013), Sampedro et al. (2015)

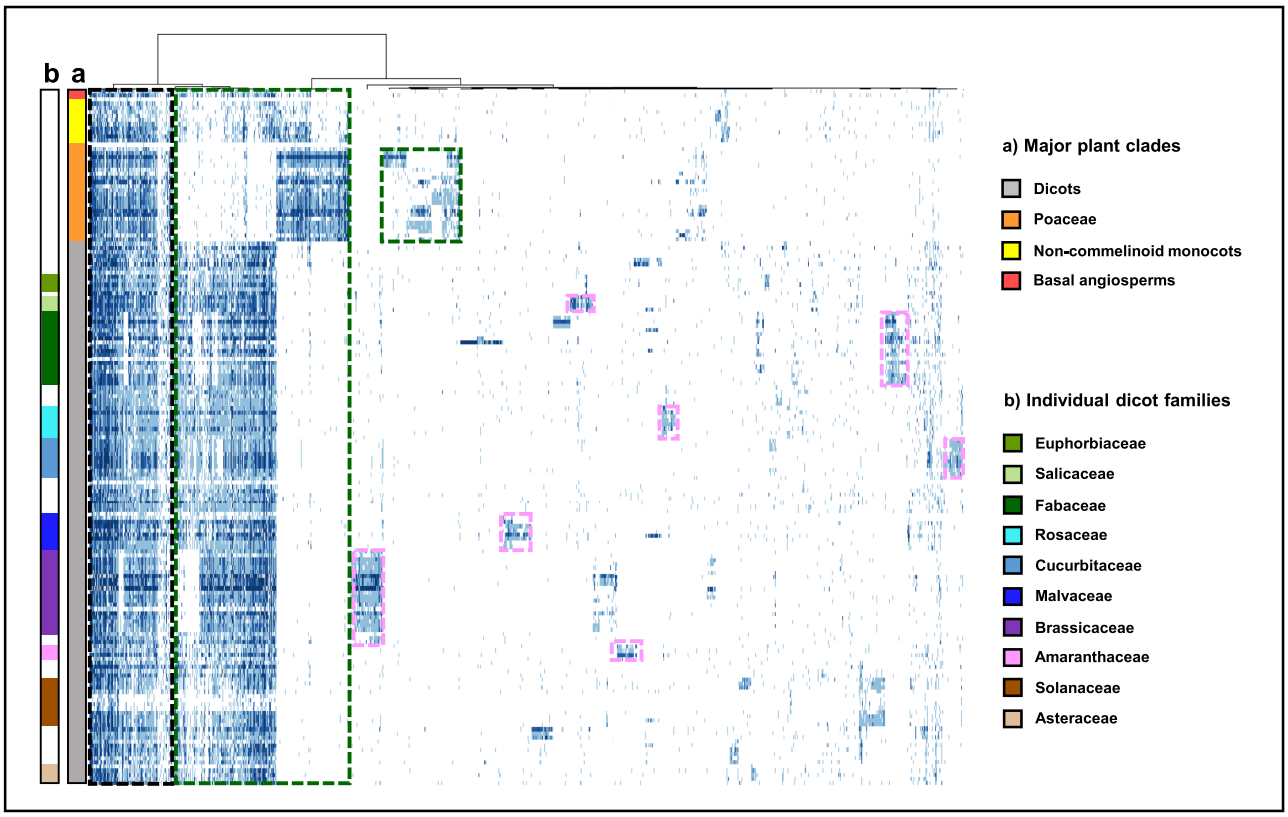

Figure 2. Heatmap displaying the taxonomic profiling of the 7,820 syntenic communities detected across the 150 cell wall gene families and 169 plant genomes of this study. Heatmap cells are colored based on gene copy number of each community (columns) and species (rows) combination. Colors range from white (copy number = 0) to dark blue (max copy number). Communities are clustered based on patterns of taxonomic profiling. Dashed rectangles group communities representing genomic contexts that (i) are conserved across all (or most) angiosperms (black rectangle), (ii) are differentially conserved between Poaceae and (most of) eudicots (green rectangles), or (iii) display lineage-specific patterns of conservation (pink rectangles). Boxes on the left side, indicated with "a" and "b," display the taxonomy of species over which syntenic communities were taxonomically profiled (see legend in the figure).

represented within the communities of group (ii) in Fig. 2 (46% of all the hemicellulose-related syntenic genes), followed by cellulose-related genes (41%), lignin/phenylpropanoid-related genes (35%), and transcription factors (29%). Finally, the third group of syntenic communities contains the remaining 5,359 communities, which include 34,138 total genes and display mostly lineage- or species-specific patterns of gene synteny (pink boxes in Fig. 2). Among these are 144 communities specific to Brassicaceae, 87 communities specific to Fabaceae, and other minor groups of communities specific to Salicaceae, Rosaceae, and Cucurbitaceae.

Community data were also cross-referenced with gene copy number results, showing that large CNV exists both between and within syntenic communities for several gene families. Specifically, CNV is often taxon- and/or community-specific. As an example, Fig. 3, A and B, shows that 2 genes at the basis of sucrose supply for cellulose synthesis—*SUCROSE PHOSPHATE SYNTHASE* (*SPS*) and *SUCROSE SYNTHASE* (*SUS*)—display a Poaceae-specific increase in gene copy number (>3 gene copies per species on average) relative to eudicots (mostly singleton or double-gene copies). Alternatively, Fig. 3C displays that *PECTIN LYASE* (*PLY*) genes show both copy number and presence–absence variation between Poaceae and eudicots between and within several genomic contexts. Interestingly, the different genes reported in Fig. 3 are all known to affect the content of cellulose and pectins (Zhong et al. 2019), which both tend to vary considerably between type I and type II cell walls (Vogel 2008).

To conclude, like CNV data, community data of each gene family were also analyzed together with available information on gene function in the context of type I and type II cell walls from scientific literature. The next paragraphs report the results of this analysis for different gene families that are critical for secondary cell wall deposition and for hemicellulose biosynthesis. Moreover, Table 2 reports a list of gene families that display either conservation or differentiation of genomic contexts across Poaceae and eudicots and appear relevant for type I to II cell wall differentiation.

Conserved and divergent genomic contexts for the pathways controlled by the functionally different BLH6/BLH9 transcription factors

BLH6 and *BLH9* are phylogenetically close cell wall transcription factors (Fig. 4A) that are present in all angiosperms but display diverse functionalization patterns, with *BLH9* being a repressor of lignification across all plants and *BLH6* acting as repressor and inducer of secondary cell wall in eudicots and grasses, respectively (Rao and Dixon 2018). This makes these genes a particularly interesting case to study the genetic factors underlying Poaceae–eudicot cell wall differences. Phylogenomic analysis revealed 3 distinct *BLH6* genomic contexts and 2 separate *BLH9* genomic contexts (Fig. 4A). These genomic contexts correspond to distinct phylogenetic gene clades supported by high bootstrap (95 to 100) (Fig. 4A). Interestingly, while both *BLH9* genomic contexts are conserved across all the angiosperms, of the 3 *BLH6* contexts, 1 is specific to eudicots and the other 2 are restricted to grasses (Fig. 4A). Moreover, the overall *BLH6* copy number is also different between grasses and eudicots, with Poaceae having 3.6 *BLH6* copies per species on average, compared with 2.2 of eudicots (*t* test's $P = 0.000$). This difference is not observed for *BLH9* (2.3 copies per grass species vs 1.9 in eudicots; *t* test's $P = 0.119$). Remarkably, the "surplus" *BLH6* grass copies are not equally spread across the 2 Poaceae-specific *BLH6* genomic contexts. In fact, the light blue community in Fig. 4A contains 2 *BLH6* copies per grass species on average, compared with only 1 copy for the green community in Fig. 4A. Overall, these results show a clear association between the (phylo)genomic organization of *BLH6* and *BLH9* and their functional diversification in the context of type I and type II cell walls.

To further investigate the association between the genomic organization and functional specialization of *BLH6*/*BLH9* genes, phylogenomic analyses were extended to the other genes within the *BLH6* regulatory pathway. These include 2 downstream transcription factors—*OVATE FAMILY PROTEIN 4* (*OFP4*) and *KNOX TALE 7* (*KNAT7*)—and a major lignin structural gene, *FERULATE 5-HYDROXYLASE* (*F5H*) (Rao and Dixon 2018; Qin et al. 2020). Divergent phylogenomic patterns between Poaceae and eudicots were revealed for all these genes. Specifically, *OFP4* genes turned out to be organized into 2 main syntenic communities (Fig. 4B). The largest one groups nearly all the eudicot *OFP4* copies, plus the majority of noncommelinoid monocots *OFP4* and only 2 Poaceae copies. Conversely, the smaller community includes nearly all the Poaceae *OFP4* and corresponds to an independent phylogenetic clade (bootstrap 99). Regarding *KNAT7*, its phylogenomic analysis also revealed the presence of 1 Poaceae-specific syntenic community and 1 eudicot-specific syntenic community (Fig. 4C). Moreover, *KNAT7* genes were phylogenomically analyzed together with the members of the *KNAT3* family (Fig. 4C), which groups homologs involved in lignin deposition and regulated by the master cell wall transcription factors *NST1* and *NST2* (Qin et al. 2020). As for *BLH9*, the function of *NST* transcription factors and of *KNAT3* genes is largely conserved across grasses and eudicots (Rao and Dixon 2018), and phylogenomic analyses showed that both *KNAT3* and *NST* genes display widespread syntenic conservation across both grasses and eudicots (Figs. 4, C and D). Finally, phylogenomic analysis of the last gene within the *BLH6* regulatory pathway—*F5H*—remarkably displayed differential syntenic and phylogenetic organization between Poaceae and eudicots (Fig. 4E).

As a final step in the analysis of the genes belonging to the *BLH6* regulatory pathway, it was tested whether the gene clades corresponding to differential genomic contexts between Poaceae and eudicots displayed differential selection pressures between grass and eudicot clades. Selection pressure is defined by the ratio of nonsynonymous (dN) over synonymous (dS) substitutions between different gene nucleotide sequences (Schwerdt et al. 2015). Thus,

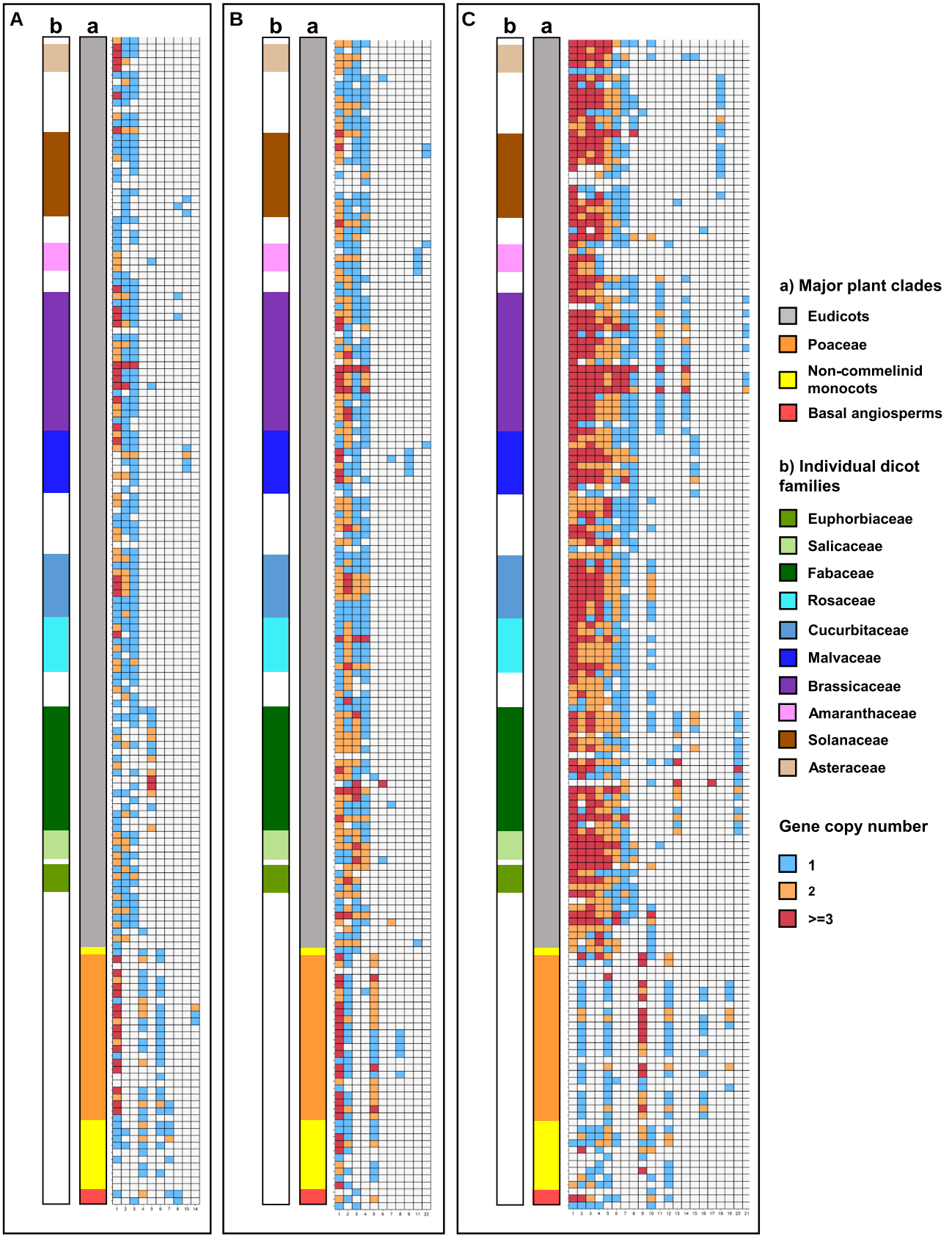

Figure 3. Examples of genomic context-specific patterns of gene CNV across 3 cell wall gene families. In all the heatmaps, columns represent syntenic communities detected for a specific gene family, while rows represent species, ordered taxonomically (see legend). Heatmap cells are colored according to the gene copy number for every genomic context–species combination (see legend). **A)** *SPS*. **B)** *SUS*. **C)** *PLY*.

Table 2. Gene families displaying either Poaceae-/eudicot-specific or angiosperm-wide synteny that appear particularly relevant for the differentiation of type I and type II cell walls in view of their known cell wall function

Cell wall gene function	Genes in Poaceae- and/or eudicot-specific communities (%)	Genes in angiosperm-wide communities (%)	Genes in other lineage-specific communities (%)	Relevance in the context of type I and type II cell walls	References
F5H	93.9	0.0	6.1	Central gene in lignin synthesis, regulating total lignin production and monolignol ratio	Zhong et al. (2019), Vogel (2008)
CslA	75.1	0.3	24.7	Binds mannan residues during (gluco) mannan synthesis	Zhong et al. (2019), Vogel (2008)
CslC	61.5	29.8	8.7	Binds glucose residues of XyG molecules	Zhong et al. (2019), Vogel (2008)
AXY9	89.8	0.0	10.2	Mediates XyG *O*-acetylation	Zhong et al. (2019), Vogel (2008)
XXT	64.6	31.0	4.4	Binds xylose residues over forming XyG backbones	Zhong et al. (2019), (Vogel 2008)
IRX9/9L (GT43)	73.0	15.0	12.1	Forms a biosynthetic complex together with IRX14/14L and IRX15/15L to bind xylose residues during xylan synthesis	Zhong et al. (2019), Vogel (2008)
IRX14/14L (GT43)	64.0	23.0	13.0	Forms a biosynthetic complex together with IRX9/9L and IRX15/15L to bind xylose residues during xylan synthesis	Zhong et al. (2019), Vogel (2008)
IRX15/15L	62.1	0.0	37.9	Forms a biosynthetic complex together with IRX9/9L and IRX14/14L to bind xylose residues during xylan synthesis	Zhong et al. (2019), Vogel (2008)
XTH	60.7	0.0	39.3	Mediates the extension of XyG molecules	Zhong et al. (2019), Vogel (2008)
RGXT	61.4	0.2	38.4	Central gene for the synthesis of RGII	Atmodjo et al. (2013), Vogel (2008)
GALS	63.5	0.0	36.5	Involved in the synthesis of RGI side chains	Atmodjo et al. (2013), Vogel (2008)
XGD	61.8	0.5	37.8	Transfers xylose during xylogalacturonan synthesis	Atmodjo et al. (2013), Vogel (2008)
BLH6	97.6	2.3	0.0	Inducer/repressor of secondary cell wall development	Rao and Dixon (2018)
E2FC	0.0	72.7	27.3	Key upstream regulator of cell wall biosynthesis, which activates several first-layer cell wall transcription factors	Taylor-Teeples et al. (2015), Rao and Dixon (2018)
VND	19.2	69.9	10.9	First-layer cell wall transcription factors regulating ectopic secondary cell wall deposition in vessels. Their function is widely conserved across angiosperms.	Taylor-Teeples et al. (2015), Rao and Dixon (2018)
SND	3.2	87.7	9.0	First-layer cell wall transcription factors required for normal secondary cell wall biosynthesis. Their function is widely conserved across angiosperms.	Taylor-Teeples et al. (2015), Rao and Dixon (2018)
NST1/2/3	0.0	98.4	1.6	First-layer cell wall transcription factors required for normal secondary cell wall biosynthesis. Their function is widely conserved across angiosperms.	Taylor-Teeples et al. (2015), Rao and Dixon (2018)
C2H2	0.0	85.1	14.9	Repressor of secondary cell wall development.	Taylor-Teeples et al. (2015), Rao and Dixon (2018)

Several structural genes involved in critical steps of the synthesis of polysaccharides whose content differs substantially between Poaceae and eudicots have their genes mostly contained in Poaceae- or eudicot-specific syntenic communities. Conversely, several transcription factors with conserved function across angiosperms are retained in syntenic communities displaying wide conservation across angiosperms.

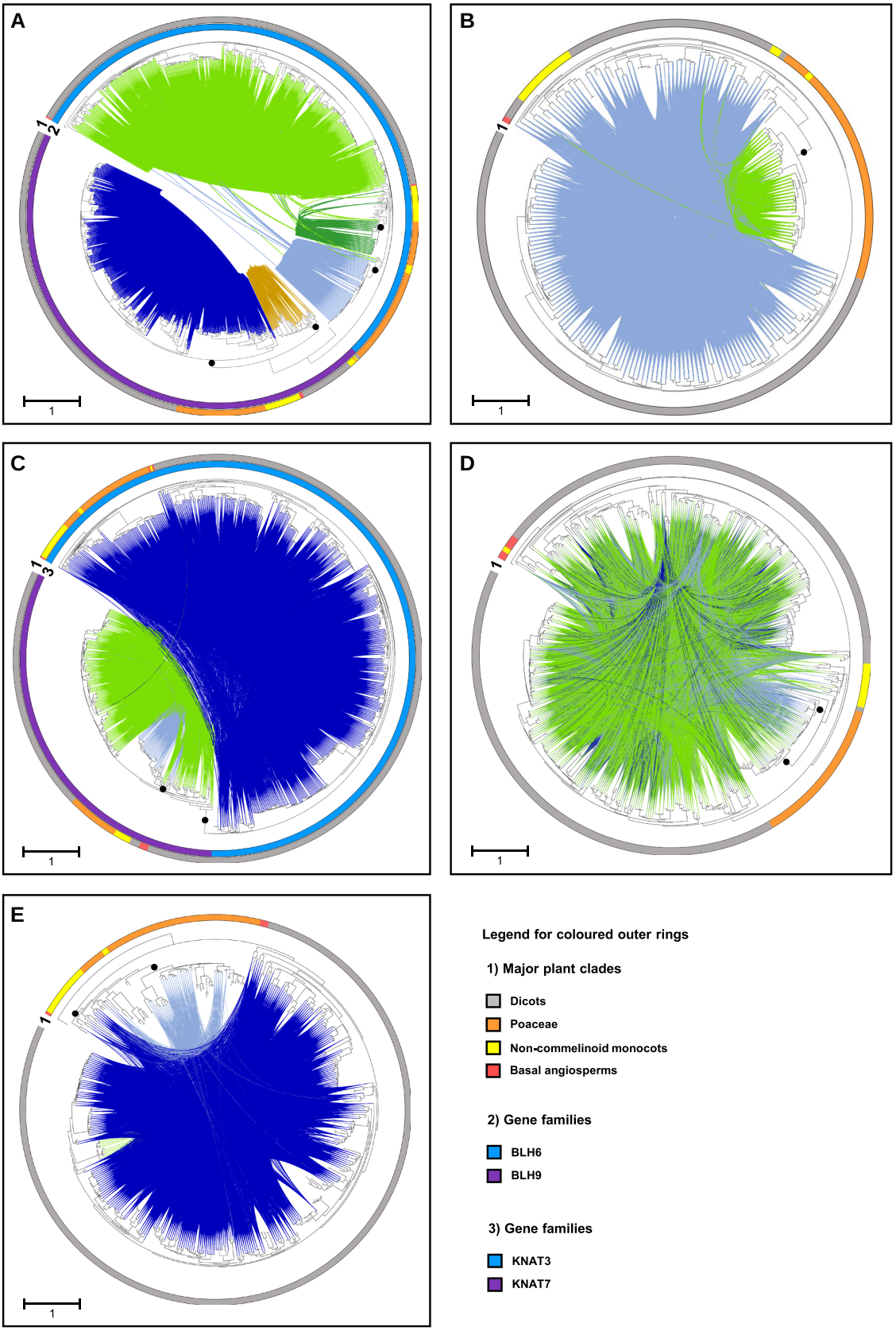

Figure 4. Phylogenetic trees displaying phylogenetic and syntenic relationships of the gene families analyzed in relation to the *BLH6*/*BLH9* study case. Overall, the trees display that genes known to concur to the different *BLH6* functions in Poaceae and eudicots are all differentially genomically

(continued)

differential dN:dS ratios between genes belonging to different taxonomic and/or phylogenomic clades highlight differential evolutionary rates associated with different genetic/genomic configurations. The analysis of selection pressure with the use of the branch model from the CodeML program (Yang 2007) revealed that Poaceae genes organized in independent genomic contexts are systematically under significantly different (and positive—dN:dS > 1) selection pressure as compared with their eudicot counterparts [likelihood ratio test's (LRT's) $P < 0.01$; Supplemental Table S6)]. This holds true for all the genes displaying differential syntenic conservation in Poaceae and eudicots, except for *KNAT7*. Vice versa, the Poaceae genes organized in syntenic communities shared with eudicot species (as, for example, for the 2 *BLH9* communities of the *KNAT3* genes) did not display significant differences in selection pressure as compared with eudicot genes within the same communities in all the tests performed (LRT's $\alpha = 0.01$; Supplemental Table S6).

To conclude, all the data displayed in this paragraph highlight a striking association between the conservation/diversification of the positional organization and the conservation/diversification of gene function for all the genes within the *BLH6* pathway across Poaceae and eudicots, which is ultimately associated with similar or divergent effects of these genes on plant cell walls, respectively. Remarkably, this diversification is associated with a phylogenetic diversification of genes and extensive patterns of differential—and often positive—selection pressures, highlighting that nucleotide diversification of genes is likely favored by and intimately associated with the observed large-scale genomic gene rearrangements.

Divergent genomic contexts between Poaceae and eudicots for multiple important hemicellulose-related genes

As the content of several hemicellulosic molecules differs substantially between type I and type II cell walls, the genes synthesizing the backbone of the most differing hemicellulosic polysaccharides between Poaceae and eudicots are also relevant targets for phylogenomic analyses. One such polysaccharide is XyG, whose synthesis depends on *Cellulose synthase-like C* (*CslC*) genes—which bond the glucose residues of XyG backbones—and on *Alfa-1,6-Xylosyltransferases* (*XXT*), which add the xylose moieties (Zabotina 2012). For both these genes, phylogenomic analysis revealed distinct phylogenetic and syntenic patterns associated with their functional diversification between Poaceae and eudicots (Fig. 5). Specifically, for the *XXT* family, we found a total of 6 phylogenetic clades corresponding to 6 independent genomic contexts (Fig. 5A). Three clades/communities grouped only eudicot genes and corresponded to the homologs of Arabidopsis *XXT1*, *XXT2*, and *XXT3/5* genes, respectively. These are the most important *XXT* copies for XyG synthesis in Arabidopsis, forming an active biosynthetic complex (Zabotina 2012). Conversely, the other 3 clades/communities were specific to Poaceae and noncommelinoid monocot genes and included the maize homologs of *AtXXT1*, *AtXXT2*, and *AtXXT3/5* genes, respectively. To conclude, the eudicot *XXT1* and *XXT2* clades displayed a relatively large extent of interclade synteny and tree branches of similar sizes, highlighting relatively little phylogenetic differentiation. Conversely, grass *XXT1* and *XXT2* are independently syntenically organized and phylogenetically more distant (Fig. 5A). Regarding *CslC* genes, phylogenomic analysis identified 8 distinct syntenic communities, 4 of which were specific to either grasses or eudicots, while the other 4 included both eudicot and Poaceae genes (Fig. 5B). The syntenic differentiation of *CslC* genes between Poaceae and eudicots is thus not absolute. However, BLAST analyses showed that the largest eudicot-specific *CslC* community included *AtCslC4*. This is the most active Arabidopsis *CslC* gene and the only one highly expressed in all Arabidopsis tissues (Zabotina 2012). The 2 maize homologs most similar to *AtCslC4* (>60% sequence identity)—XP_008662691.2 and XP_008657194.1—were included in the 2 grass-specific *CslC* communities. To conclude, copy number community data indicated that eudicot- and grass-specific *CslC* communities comprise the majority of *CslC* members from all angiosperms (342 out of 596 total genes).

In addition to XyGs, (gluco)mannans content is very different between Poaceae and eudicot cell walls. (Gluco)mannans synthesis depends largely on *Cellulose synthase-like A* (*CslA*) genes, which bind the mannose residues (Zhong et al. 2019). Remarkably, phylogenomic analysis showed that *CslA* genes are genomically very differently organized between Poaceae and eudicots (Fig. 6). Specifically, *CslA* members are divided into 10 different syntenic communities, of which 5 are specific to eudicots (2 restricted to Brassicaceae) and 5 specific to Poaceae and, partly, noncommelinoid monocots. Of the 5 Poaceae-specific communities, 2 contain genes which are phylogenetically closer to eudicot *CslA*, while the members of the other 3 form a monocot-specific *CslA* phylogenetic clade (bootstrap = 97). Interestingly, Poaceae *CslA* copy number is highest in the largest community corresponding to such monocot-specific phylogenetic clade, with >3 *CslA* genes per species on average. Moreover, the 3 phylogenetically distinct

Figure 4. (Continued)
(syntenically) organized between these species. Conversely, genes related to the *BLH6*/*BLH9* pathway but with conserved function across Poaceae and eudicots display conserved synteny across most or all angiosperms. In each plot, lines connecting tree leaves represent syntenic relationships between genes, classified according to the detected syntenic communities (different line colors indicate different syntenic communities). Rings around trees display the taxonomic profiling of the genes within each tree or the gene family to which each gene in a tree belongs to (in the case multiple gene families were aligned together into 1 tree) (see legends). Black dots on tree nodes indicate tree branches that are relevant for the phylogenomic diversification of gene sequences that are supported by bootstrap ≥90. **A)** *BLH6*/*BLH9* phylogenetic tree. **B)** *OFP4*. **C)** *KNAT3*/*KNAT7*. **D)** *NST*. **E)** *F5H*.

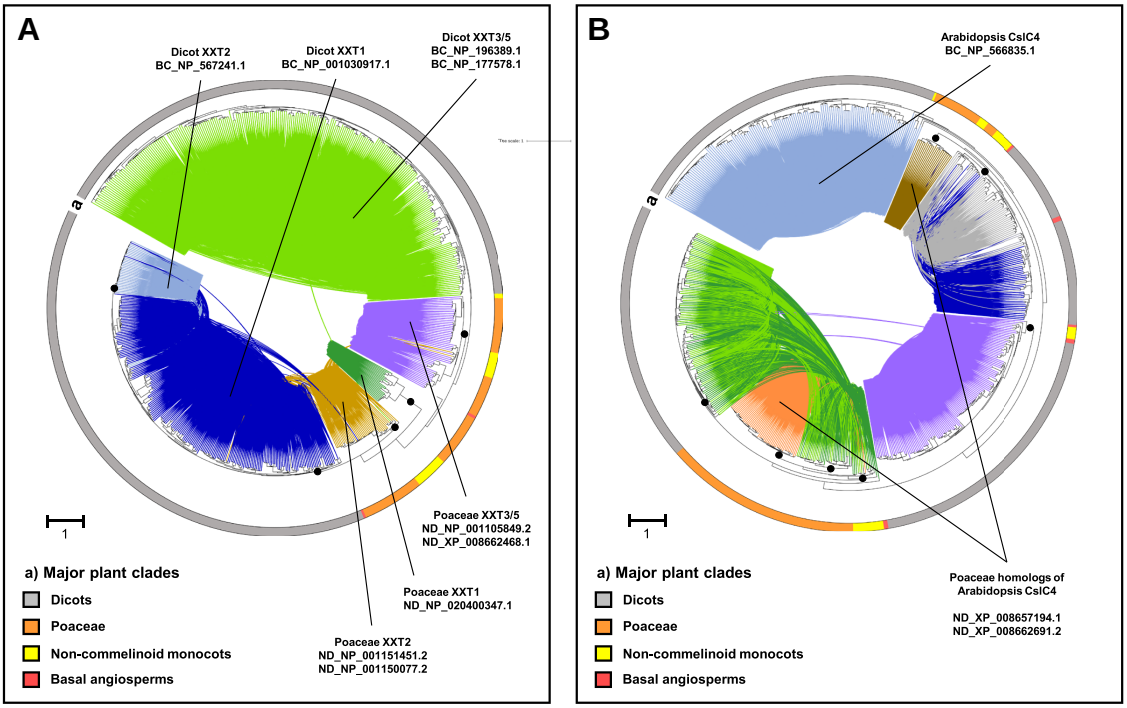

Figure 5. Phylogenetic trees displaying phylogenetic and syntenic relationships of the 2 main gene families involved in XyG biosynthesis: *XXT* **(A)** and *CslC* **(B)**. The trees show that these genes tend to be differentially organized in Poaceae and eudicots from a genomic (synteny) point of view. Specifically, the functionally most important *XXT* and *CslC* members known in *Arabidopsis* and their respective homologs in Poaceae (maize) are organized in completely different genomic contexts. In each plot, lines connecting tree leaves represent syntenic relationships between genes, classified according to the detected syntenic communities (different line colors indicate different syntenic communities). Rings around trees display the taxonomic profiling of the genes within each tree (see legends). Black dots on tree nodes indicate tree branches that are relevant for the phylogenomic diversification of gene sequences that are supported by bootstrap ≥90. The syntenic communities containing Arabidopsis and maize genes of interest are indicated (black arrows). Within gene IDs, "BC" indicates Arabidopsis, while "ND" indicates maize.

Poaceae-specific communities group the majority of Poaceae *CslA* genes (140 out of 207 copies).

Xylans represent a final relevant group of hemicellulosic molecules for type I and type II cell walls. Their most important biosynthetic genes are 3 different *IRX* families—*IRX9/9*-like (CAZy GT43 family), *IRX10/10*-like (CAZy GT47 family), and *IRX14/14*-like (CAZy GT43 family)—that form a xylan biosynthetic complex (Zhong et al. 2019). As for the other hemicellulose-related genes above, phylogenomic analyses revealed substantial divergence in the positional organization, phylogenetic diversification, and copy number dynamics of *IRX* genes between Poaceae and eudicots (Fig. 7). While 2 syntenic communities conserved across Poaceae, noncommelinoid monocots, and eudicots were detected within each of the *IRX9/9L*, *IRX10/10L*, and *IRX14/14L* clades, 54% of all the *IRX* genes studied (1,636 out of 3030) were included within 25 different syntenic communities either eudicot- or Poaceae-specific. Moreover, differential copy number representation of the *IRX* genes between Poaceae and eudicots was observed for all the *IRX* families within the syntenic communities conserved across both Poaceae and eudicots (Fig. 8), showing that the relative representation of shared *IRX* genomic contexts can differ substantially between Poaceae and eudicot genomes. To conclude, all the *IRX* communities identified are highly phylogenetically differentiated, corresponding to distinct phylogenetic clades supported by high bootstrap.

Cell wall genes are often organized in tandem gene clusters which are different between Poaceae and eudicots

To further characterize the cell wall genomic properties of Poaceae and eudicots, the (differential) occurrence and conservation of tandem cell wall gene clusters was also studied for the 150 gene families and the 169 genomes. The presence and evolution of tandem gene clusters affect plant traits in several ways, for example, by influencing gene dosage or by facilitating gene sub- or neofunctionalization (Kono et al. 2018). Therefore, genomic variability of tandem gene arrays can be relevant also for type I to II cell wall differentiation.

Our analyses revealed that tandem gene clusters are relatively common for cell wall genes, as 44 of the 150 gene families analyzed have (part of) their genes organized in tandem arrays of at least 2 members in 60% or more of the species surveyed. Conversely, 85 gene families (57% of the total) are mainly organized as (distinct) singleton loci (tandem arrays found in <40% of the species surveyed). The remaining

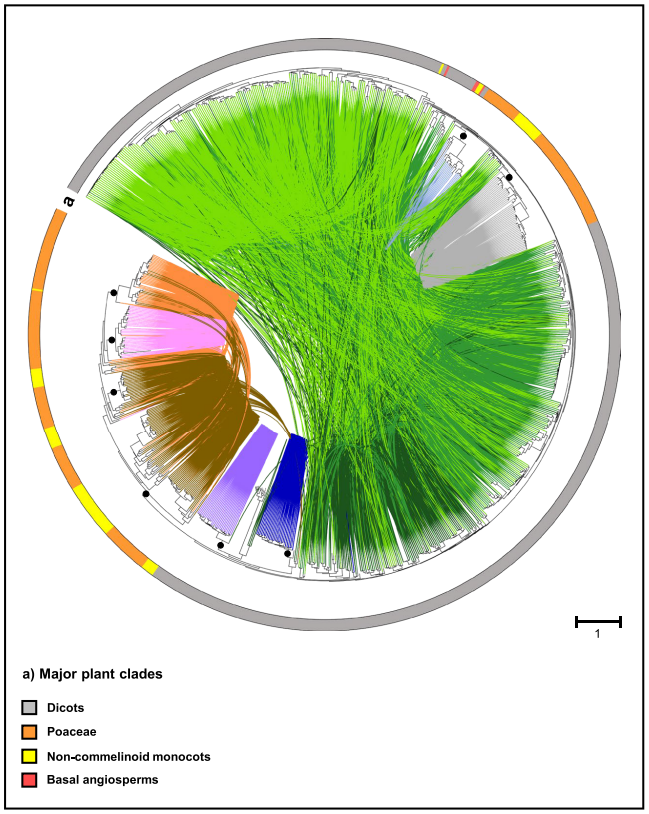

Figure 6. Phylogenetic tree displaying the phylogenetic and syntenic relationships of the main gene family involved in mannan biosynthesis: *CslA*. The tree shows that these genes are organized in (multiple) different genomic contexts in Poaceae and eudicots. Within the tree, lines connecting the leaves represent syntenic relationships between genes, classified according to the detected syntenic communities (different line colors indicate different syntenic communities). The ring around the tree displays the taxonomic profiling of the tree genes (see legends). Black dots on tree nodes indicate tree branches that are relevant for the phylogenomic diversification of gene sequences that are supported by bootstrap ≥90.

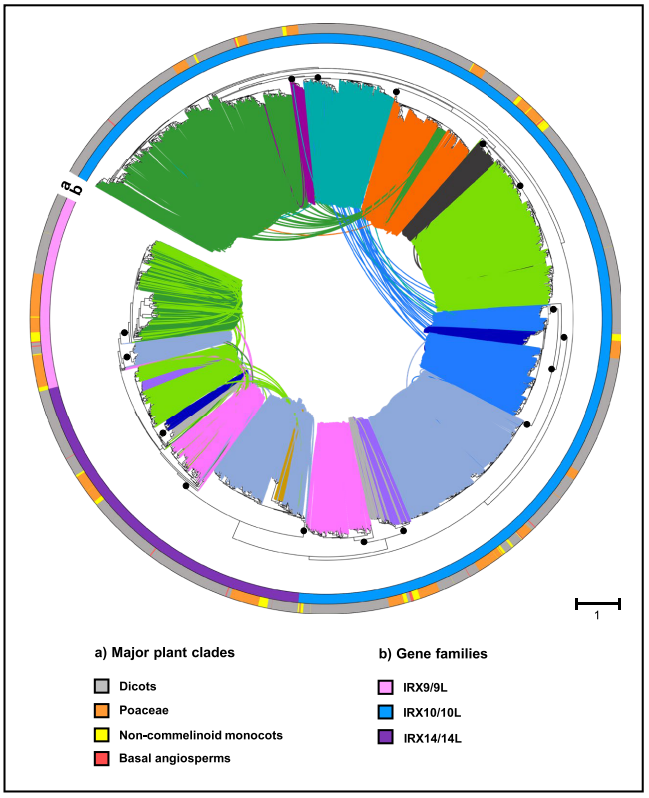

Figure 7. Phylogenetic tree displaying the phylogenetic and syntenic relationships of the 3 main gene families involved in xylan biosynthesis: *IRX9/9L*, *IRX10/10L*, and *IRX14/14L*. The tree shows that these genes are organized in (multiple) different genomic contexts in Poaceae and eudicots. Within the tree, lines connecting the leaves represent syntenic relationships between genes, classified according to the detected syntenic communities (different line colors indicate different syntenic communities). The rings around the tree display the taxonomic profiling of the tree genes and the genes belonging to each *IRX* gene family (see legends). Black dots on tree nodes indicate tree branches that are relevant for the phylogenomic diversification of gene sequences that are supported by bootstrap ≥90.

21 gene families displayed mixed singleton/tandem patterns, depending on the species (Supplemental Table S7 and Data Set 2). Overall, genes involved in cellulose, lignin/phenylpropanoid, and pectin biosynthesis are the ones most often organized as tandem clusters. Conversely, cell wall transcription factors and genes involved in sugar supply for cell wall biosynthesis are the classes mostly arranged as singleton loci (Table 3).

Given our focus on the genetics underlying Poaceae and eudicot cell wall differences, the variability of cell wall gene tandem clusters between these 2 groups of species was studied more in detail. This analysis revealed that 20 of the 44 cell wall gene families with high occurrence of tandem gene clusters display marked differences in the properties of such clusters between Poaceae and eudicots (Table 4). Specifically, such differences entail either the presence/absence variability of tandem clusters between Poaceae and eudicots (11 gene families) or the number of genes included in tandem arrays between Poaceae and eudicots (9 gene families). Remarkably, some of the gene families displaying variability of gene tandem arrays between Poaceae and eudicots are particularly relevant for the differences between type I and type II cell walls. For example, they include 2 critical families affecting the content of lignin and of cell wall hydroxycinnamates: *PAL*—displaying 1 tandem cluster of 5 genes per species on average in Poaceae, but no clusters in eudicots—and *PRX*, which displays 17 tandem arrays of 4 genes per species on average in Poaceae and 9 clusters of 3 genes each on average per eudicot species. Moreover, 4 important pectin-related gene families—*XYLOGALACTURONAN XYLOSYLTRANSFERASE* (*XGD*), *PME/PMEI*, *PE*, and *PG*—all display a higher occurrence of tandem gene clusters (with also more genes per cluster) in eudicots compared with Poaceae (Table 4). Interestingly, for several gene families, the variability of tandem gene clusters between Poaceae and eudicots goes in parallel with the differences in gene

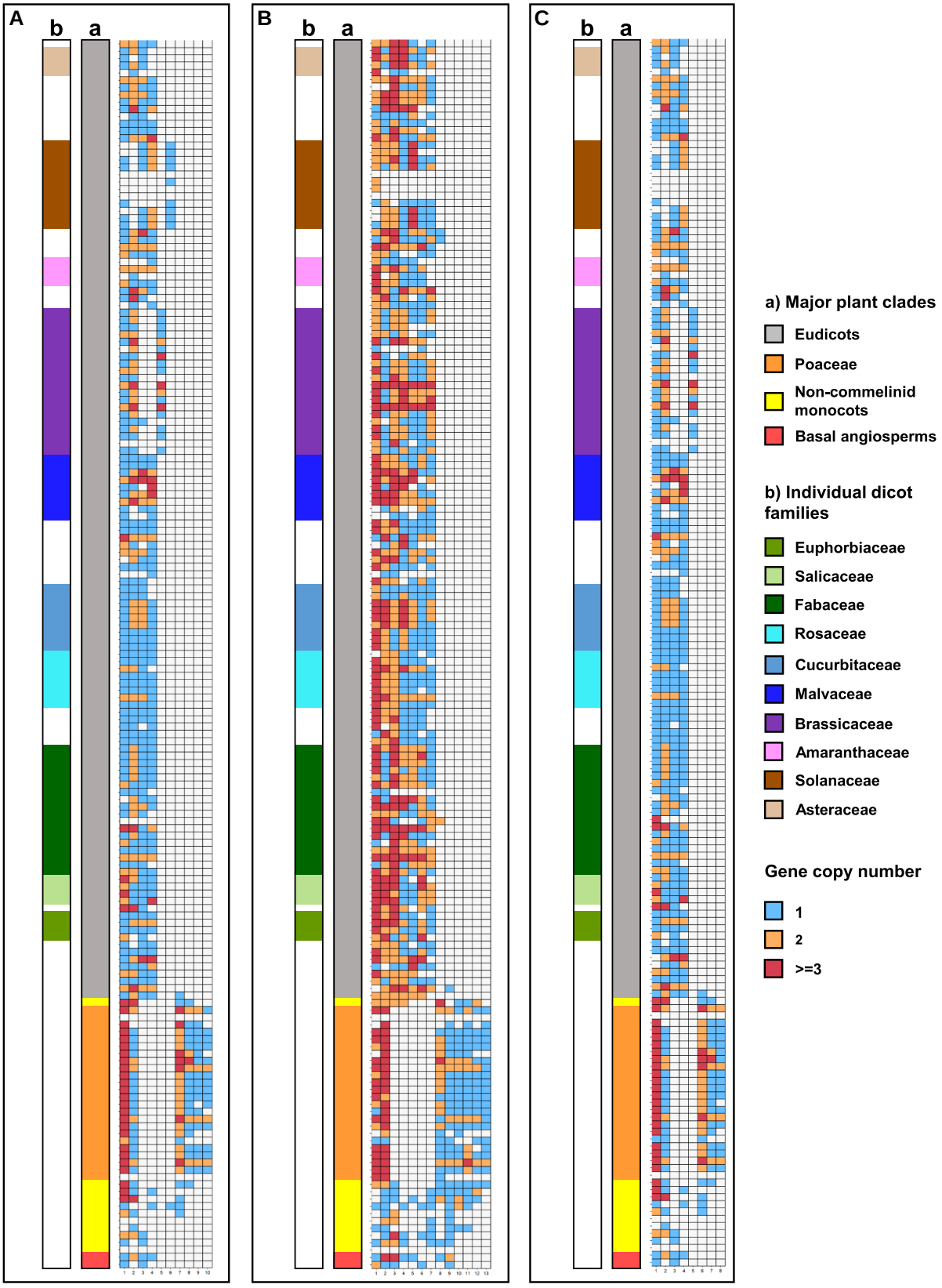

Figure 8. Heatmaps showing the genomic context- and taxonomic clade-specific patterns of gene CNV across the 3 *IRX* gene families involved in xylan synthesis. In all the heatmaps, columns represent syntenic communities detected for a specific gene family, while rows represent species, ordered taxonomically (see legend). Heatmap cells are colored according to the gene copy number for every genomic context–species combination (see legend). **A)** *IRX9/9L*. **B)** *IRX10/10L*. **C)** *IRX14/14L*.

Table 3. Statistics of gene tandem cluster presence across broad categories of cell wall gene functions

Cell wall broad process	Percentage of gene families organized in tandem clusters[a]	Number of gene families organized in tandem clusters[a]	Percentage of gene families not organized in tandem clusters[b]	Number of gene families not organized in tandem clusters[b]	Percentage of gene families displaying unclear pattern[c]	Number of gene families displaying unclear pattern[c]
Callose synthesis	0.0	0	0.0	0	100.0	1
Cellulose synthesis	46.2	6	38.5	5	15.4	2
Glucose supply to polysaccharide synthesis	0.0	0	100.0	8	0.0	0
CW_other_proteins	38.1	8	57.1	12	4.8	1
Hemicellulose metabolism	30.0	15	60.0	30	10.0	5
Lignin/ phenylpropanoid synthesis	42.1	8	31.6	6	26.3	5
Pectin metabolism	36.8	7	42.1	8	21.1	4
Transcription factors	0.0	0	94.1	16	5.9	1

[a]Tandem clusters occur in >60% of the species analyzed. [b]Tandem clusters occurring in <40% of the species analyzed. [c]Pattern observed does not fall in any of the 2 previous categories ([a] and [b]).

Table 4. Gene families displaying large variability in cell wall gene tandem clusters between Poaceae and eudicots as either differential number of tandem clusters or the presence/absence of tandem clusters

Cell wall gene function	Average number of gene tandem clusters per species (Poaceae)	Average number of genes per cluster (Poaceae)	Average number of gene tandem clusters per species (eudicots)	Average number of genes per cluster (eudicots)
ARAD	3	3	1	2
MUR2/3/4	3	3	1	2
FUT	4	3	2	2
EXP/EXPL	8	4	2	3
PME/PMEI	2	2	9	3
PRX	17	4	9	3
PE	4	2	11	3
PG	7	3	11	3
BBE_like	1	5	3	7
GALT	2	3	0	2
ERF	1	5	0	2
XXT	1	3	0	2
PAL	1	5	0	3
HCT	0	…	1	3
DRP/ADL	0	…	1	3
FLA11/12	0	…	1	4
ESB1	0	…	1	3
PLY	0	…	1	3
THE1	0	…	1	3
XGD	0	…	2	3

copy number variation (see commonalities between Tables 1 and 4). This highlights once again the extensive interaction between the genomic factors considered in this study in shaping the genomic contexts of Poaceae and eudicot cell walls.

Discussion

The biological basis of the profound differences between the type I cell walls of eudicots and the type II cell walls of Poaceae is not yet fully understood. Such differences entail nearly all the cell wall components and represent the likely product of the combined evolution of the different angiosperm lineages and of cell walls themselves as plant structures (Vogel 2008; Sarkar et al. 2009). Typically, such complex evolutionary trajectories leave traces in plant genes and genomes (Zhao et al. 2017; Kerstens et al. 2020). However, the occurrence, extent, and relevance of a genetic differentiation underlying type I and type II cell walls are still debated (Yokoyama and Nishitani 2004; Penning et al. 2019; Yokoyama 2020; Kozlova et al. 2020). This study revealed major differences between Poaceae and eudicots in the genomic organization of several cell wall genes. This differentiation involves multiple genomic properties, including the presence/

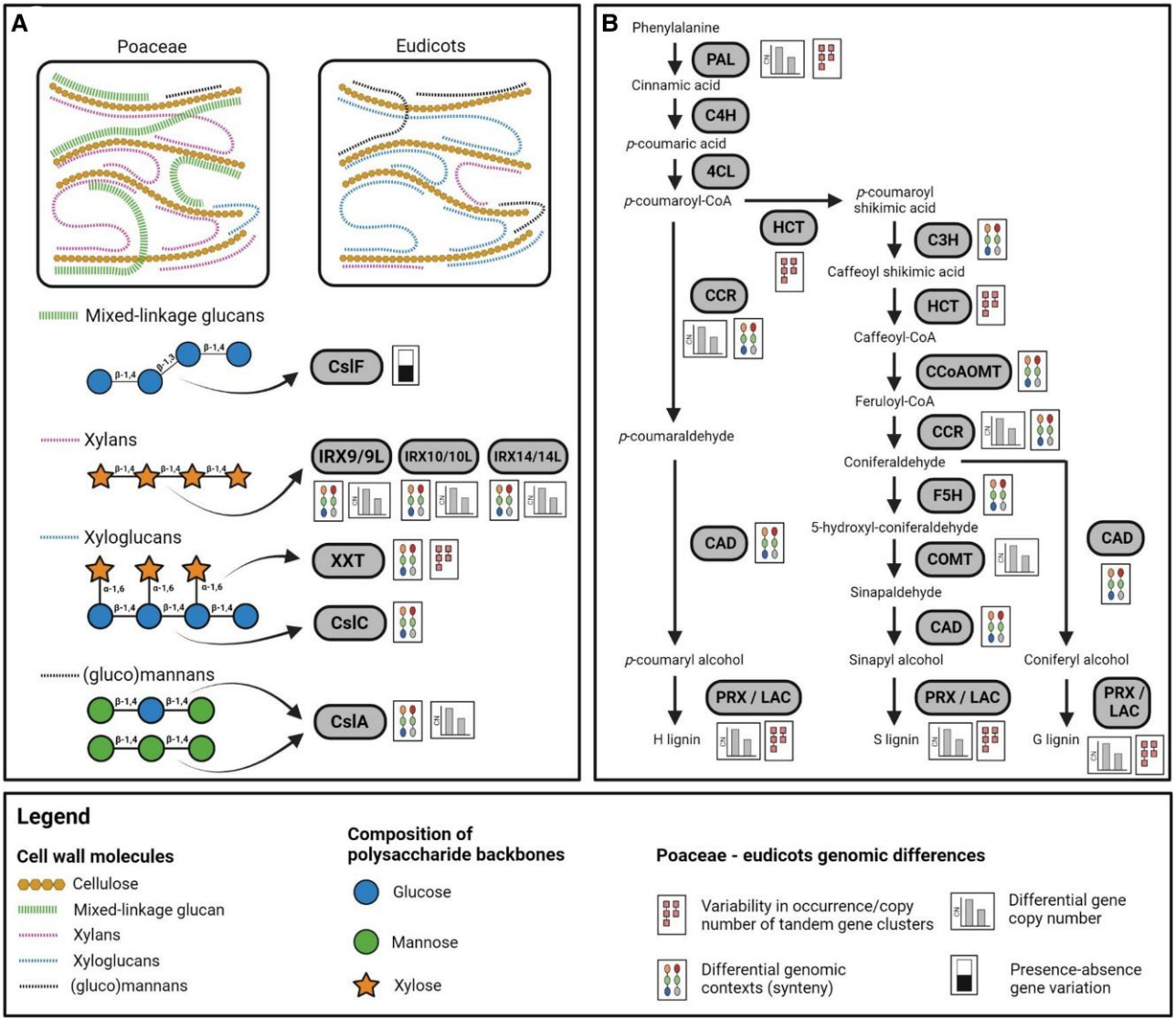

Figure 9. Summary of the differential genomic patterns between Poaceae and eudicots found for major hemicellulose and lignin genes. **A)** Differences in hemicellulose composition between Poaceae and eudicot cell walls (upper part), with schematic representation of the backbones of the polysaccharides responsible for these differences, along with the indication of the major genes synthesizing polysaccharide backbones and the differential genomic patterns found for those genes between Poaceae and eudicots. **B)** Representation of the lignin/phenylpropanoid pathway, with the indication of genes and differential genomic patterns found between Poaceae and eudicots.

absence and copy number variation of genes and gene tandem arrays, syntenic gene conservation, and gene sequence diversification. Moreover, the magnitude of such differentiation is exceptional, as the vast majority of the gene families analyzed displays at least 1 pattern of genomic diversification between Poaceae and eudicots. Finally, for several gene families for which functional characterization is available, differential genomic patterns correspond strikingly with the cell wall differences between Poaceae and eudicots (Fig. 9). Considering that genomic properties as differential gene synteny and CNV are known sources of functional gene diversification and phenotypic innovations (Flagel and Wendel 2009; Zhao et al. 2017; Artur et al. 2019; Lye and Purugganan 2019; Kerstens et al. 2020), the profoundly different cell wall genomic properties between Poaceae and eudicots are a potential major driver of the cell wall differentiation between these 2 plant clades (Fig. 9).

In light of what was just discussed, it is relevant to understand how the different cell wall "genomic landscapes" of Poaceae and eudicots got shaped and how they can lead to phenotypic cell wall diversification. In this regard, the differentiation of syntenic genomic contexts seemed to have played a major role. The fact that differentially organized genomic gene contexts can facilitate the functional diversification of plant genes, leading to phenotypic adaptations, has been amply discussed, as for the *MADS-box* (Zhao et al. 2017), *APETALA2* (Kerstens et al. 2020), *LEA* (Artur et al. 2019), or *NRT* plant gene families (Zoghbi-Rodríguez et al.

2021). Typically, different genomic contexts can differentiate the modes of gene function, for example, by determining different patterns of gene expression or by favoring gene diversification and subfunctionalization (Dewey 2011). In turn, differentiated functional modes can affect plant phenotypes and can be stabilized if selected over evolution (Dewey 2011). In this study, the *BLH* case (Section 2.2.1) clearly demonstrated that this could be the case also for cell wall genes. In fact, the major genomic difference between the *BLH9* family, whose function is conserved in Poaceae and eudicots, and the *BLH6* family—which has opposite function in Poaceae and eudicots—was found to be the occurrence of differential *BLH6* genomic contexts between Poaceae and eudicots. Moreover, differential genomic contexts were also observed for all the genes directly controlled by *BLH6* and that concur to determine its differential function in Poaceae and eudicots, as *OFP4*, *KNAT7*, and *F5H*. Conversely, all the *BLH*-related genes whose function is believed to be conserved across all angiosperms—as *NST* transcription factors and *KNAT3* homologs (Rao and Dixon 2018)—displayed conserved positional genomic organization. Finally, both CNV and phylogenetic differentiation supported by differential selection pressures were observed between the gene clades corresponding to distinct Poaceae and eudicot genomic contexts for the majority of the genes within the *BLH6* pathway. Overall, these findings suggest that the differential *BLH* genomic contexts may facilitate the different functionality of the *BLH6* pathway in Poaceae and eudicots, in ways similar to that reported in the studies referenced above. The fact that analogous patterns were also found for the major hemicellulose genes, as well as for many other cell wall structural genes—even if at a more general level—further corroborates this hypothesis and extends it to several cell wall gene families.

Interestingly, the differentiation of genomic gene contexts—even if highly extensive across and within the 150 gene families analyzed—was not absolute, especially for some of the hemicellulose-related genes that were analyzed in Section 2.2.2. In light of the likely cell wall functional relevance of differentiated genomic landscapes, the presence of shared genomic gene configurations between Poaceae and eudicots, next to highly divergent genomic contexts, for some of the genes participating in the synthesis of polysaccharides that differ between grasses and the rest of angiosperms (as the *CslC* and *IRX* genes of Section 2.2.2) could indicate that not all the genes of these families perform the same function. In this regard, it is noteworthy that Arabidopsis mutants at some of the *CslC* and *IRX* genes have been shown to display alterations of plant growth (Kim et al. 2020) and of seed viability (Voiniciuc et al. 2015), suggesting the possibility for their direct or indirect involvement in multiple plant functions (Little et al. 2018). In this sense, the presence of conserved genomic gene contexts between Poaceae and eudicots could be constrained by the involvement of some cell wall genes, particularly vital plant processes, next to their strict relationship with cell wall biosynthesis. In turn, this would once again further corroborate the hypothesis that diversification of genomic gene contexts facilitates the evolution of novel gene functions, as already proposed for other gene families (Dewey 2011; Zhao et al. 2017; Kerstens et al. 2020). However, it has to be highlighted that this remains currently only a hypothesis based on our results and final functional proof to support it should be provided in future research.

In addition to differential syntenic gene organization, gene CNV emerged as another major force shaping the differential genomic landscapes of Poaceae and eudicots. As genomic context variability, gene CNV also represents a well-known source of functional variation that can lead to phenotypic innovations (Jiao et al. 2011; Kondrashov 2012; Lye and Purugganan 2019). While duplicated variants are usually deleterious and commonly undergo purifying selection (Lye and Purugganan 2019), the extensive CNV between Poaceae and eudicots for a large part of cell wall gene families highlights the relevance of this process for the differentiation of the cell wall genomic landscape between these plant clades. Typically, CNV impacts gene function through gene dosage (Kondrashov 2012). The combined analysis of CNV and gene functionality for the genes of this study suggests that this could be the case for several cell wall gene families. For example, several important genes to determine the total cell wall content of phenylpropanoid compounds were found in higher copy number in Poaceae—which have larger amounts of phenylpropanoids in their cell walls—than in eudicots (Table 1). One of these is *PAL*, which initiates the phenylpropanoid pathway and determines the efficiency of phenylalanine conversion into precursors of lignin, ferulic acid, and *p*-coumaric acid (Zhong et al. 2019). A second example is *CYNNAMOYL-CoA REDUCTASE* (*CCR*), whose activity affects the cell wall content of lignin, ferulic acid, and *p*-coumaric acid (Tamasloukht et al. 2011; Tu et al. 2010; Smith et al. 2017). Finally, *PRX* genes, which mediate the *in muro* deposition of monolignols and cell wall hydroxycinnamates (De Oliveira et al. 2015; Zhong et al. 2019), were also found in higher copy number in Poaceae compared with eudicots. All together, these results suggest that a higher copy number of phenylpropanoid-related genes might reasonably "boost" the lignin/phenylpropanoid pathway in Poaceae, leading to the wider variety and higher content of cell wall phenylpropanoids observed in these plants compared with eudicots. This hypothesis is further supported by the CNV patterns observed for other genes related to cell wall phenylpropanoid content that are not strictly part of the lignin/phenylpropanoid pathway. These include the *BAHD* and *GT61* genes (see Table 1), which respectively transfer ferulic acid onto lignin monomers and form feruloylated cell wall oligosaccharides (de Souza et al. 2018; Cenci et al. 2018; Feijao et al. 2022; Chandrakanth et al. 2023). Remarkably, both *BAHD* and *GT61* genes were found in much higher copy number in grasses (Table 1), which could explain the higher prevalence of ferulic acid in Poaceae cell walls through higher gene dosage. Moreover, since the feruloylation of cell wall molecules in grasses increases biomass recalcitrance to

industrial processing (de Souza et al. 2018), the silencing of certain *BAHD* or *GT61* members, or the analysis of natural CNV at these genes in Poaceae populations, could open possibilities for modifying this trait. To conclude, all these genes, along with all the examples reported in Table 1, suggest that cell wall variation due to differential gene dosage dependent on CNV represents a likely important mechanism for the differentiation of Poaceae and eudicot cell walls.

Apart from the patterns reported in Table 1, CNV was also commonly observed between differentially conserved genomic contexts, different species clades within specific genomic contexts, or distinct phylogenetic clades. These patterns reflect an interaction between positional gene organization, CNV, and phylogenetic sequence diversification that took place during the evolution of cell wall genomic landscapes. Most likely, the combination of all these factors during the genomic differentiation of Poaceae and eudicots has contributed to deepen and stabilize the cell wall gene functional diversification between these plant clades. This appears reasonable as such "evolutionary boost" effect of the parallel differentiation of multiple genomic properties across diverse gene clades and families was already hypothesized for other genes, as for the different *LEA* proteins (Artur et al. 2019). Moreover, such multiple differential genomic patterns were found in all the study cases of this research involving genes that are most likely behaving differently in Poaceae and eudicots, as the genes within the *BLH6* regulatory pathway or the *XXT*, *CslC*, and *IRX* genes at the basis of hemicellulose synthesis.

A final genomic property analyzed in this study is the occurrence and conservation of tandem cell wall gene arrays. To a certain extent, the variability of these configurations can have similar effects to CNV in terms of gene dosage and can even overlap with CNV itself (Kono et al. 2018). However, the occurrence of gene tandem clusters can also facilitate gene diversification and trait variability (Picart-Picolo et al. 2020; Xu et al. 2020). Our results suggest that this could be the case for several cell wall genes between Poaceae and eudicots, especially considering that variability in tandem gene arrays goes in parallel with differentiation of genomic contexts and/or basic CNV. As an example, in the case of the previously mentioned *PAL* and *PRX* genes, the finding of a higher occurrence of gene tandem arrays in Poaceae may facilitate the combined expression of clustered genes in these species, in turn making it easier to modulate the dosage of gene products as previously hypothesized. Moreover, the combination of tandem arrays variability and the differentiation of the other genomic properties studied may further amplify and stabilize the differentiation of Poaceae and eudicot cell walls, as previously discussed for genomic contexts, CNV, and phylogenetic differentiation.

In conclusion, this study clearly showed that major genomic differences underlie the divergent cell walls of Poaceae and eudicots. Such differences involve several major genomic properties, and hypotheses have been discussed regarding their evolutionary origin and the biological modes by which they can translate into different cell walls. At this moment, the scarce knowledge about the specific function of several of the genes considered in this research in other species than Arabidopsis hampers a further interpretation of the patterns found beyond what is discussed above. Therefore, we foresee that a further detailed characterization of cell wall genes in several species, together with the results reported in this study, will advance the investigation of the genetic basis of the different cell walls of Poaceae and eudicots, within the context of the genomic patterns found in this research. Moreover, the data of this study can offer opportunities for novel approaches of fundamental cell wall research in Poaceae and eudicots (e.g. genomic context engineering), as well as for the identification of gene targets to modify cell wall composition.

Materials and methods

Collection of plant genomes

All the angiosperm genomes sequenced and published by the end of 2018 and available with at least a scaffold-level assembly were searched for in several online databases. For each genome, a BED file indicating gene positions and FASTA files reporting protein and nucleotide sequences coded of all the annotated protein-coding genes were retrieved. Genomes were checked for assembly completeness by using the BUSCO Viridiplantae gene set (Seppey et al. 2019) and for assembly fragmentation by assessing the number of scaffolds and the N50 statistics. Genomes with <75% BUSCO genes were excluded from further analyses. Through these criteria, a total of 169 genomes were collected (Supplemental Table S2). This set includes the genomes of 2 angiosperm species appeared before the monocot–eudicot divergence, 11 noncommelinid monocots, 24 Poaceae, and 132 eudicots from 39 different eudicot families.

Identification of cell wall genes in all the genomes used

The identification of cell wall genes within the 169 collected genomes was performed by following the approach published by Pancaldi et al. (2022b). In brief, the detailed gene annotation and the extensive cell wall research available for Arabidopsis (*A. thaliana*) in scientific literature and online databases was used to create an initial list of 1,313 genes proven to be involved in cell wall biosynthesis within this species (Supplemental Table S8). This list was further integrated with some Poaceae-specific cell wall genes for which functional information was available for maize (*Z. mays*) and/or rice (*O. sativa*). The collected genes were classified into 150 different cell wall–related functions and were annotated for functional domain composition using HMMER3 (default parameters) (Mistry et al. 2013) and the hidden Markov models of all the protein domains available at the PFAM database (El-Gebali et al. 2019). Subsequently, the collected genes were aligned against the PEP files of all the 169 plant genomes

of the study using BLAST ($E = 1E^{-3}$) (Altschul et al. 1990). This search led to the identification of all the potential homologs of the initial cell wall genes across all the collected genomes. The identified genes were also annotated for PFAM composition, and the BLAST outputs were then further filtered based on equal domain composition between BLAST queries and subjects. Finally, very large gene families for which it is known that not all the genes are involved in cell wall biosynthesis (e.g. *BAHD*) were further filtered by building phylogenetic trees with RAxML (Stamatakis 2014) and identifying clades containing genes from Arabidopsis, maize, or rice for which cell wall functional validation is available in scientific literature. For gene filtering, RAxML trees was run with 100 bootstraps and by using the PROTCATBLOSUM62 substitution matrix. At the end, the search for cell wall gene homologs yielded a list of 320,005 genes across the 169 genomes of the study and the 150 cell wall functions mentioned above (Supplemental Table S1).

Analysis of gene CNV

The number of gene copies present in each of the 169 genomes of the study was quantified for each of the 150 cell wall gene functions (custom R script, available at https://github.com/Francesco1994WUR/Cell_wall_phylogenomics). The average number of genes belonging to each cell wall function was determined for each genome, and *t* tests were computed to assess significant differences in average copy number between Poaceae and eudicots, as well as between multiple other angiosperm families at a time. In addition, heatmaps were created to visualize patterns of CNV across both cell wall gene families and plant species. PCA was also performed to assess the contribution of cell wall gene CNV to the differentiation of Poaceae, noncommelinoid monocots, and eudicots (custom R script, available at https://github.com/Francesco1994WUR/Cell_wall_phylogenomics). Finally, quantitative data from all these analyses were crossed with literature information on gene function to identify classes of genes whose copy number patterns appear particularly relevant in the context of differentiation between type I and type II cell walls.

Synteny analysis

The syntenic conservation of the ~320,000 cell wall genes of the study across the 169 angiosperm genomes was analyzed by following the methodology developed by Zhao and Schranz (2017) for large-scale network synteny analysis. Specifically, Diamond (Buchfink et al. 2015) was used to align all the proteins of each genome against all the other proteins of that genome and all the proteins of every other genome (default parameters; $E = 1E^{-3}$). The outputs of Diamond were processed with MCScanX (Wang et al. 2012) to detect synteny (i.e. conserved gene order across multiple genomes) by evaluating the relative genomic position of pairs of homologous genes from each genome comparison. MCScanX was run with default parameters, except -s (number of colinear genes to claim a syntenic block) set to 3. The outputs of MCScanX were organized in a synteny network, in which each node is a gene and edges represent syntenic connections between genes. The synteny network was then filtered to retain only pairs of nodes linking cell wall genes (Supplemental Table S4) and decomposed into syntenic communities (i.e. groups of genes that display significantly higher synteny with each other than with the rest of genes in the network) of at least 4 nodes ($k = 4$), by using the Infomap algorithm (Rosvall and Bergstrom 2007; Rosvall et al. 2009). Syntenic communities of cell wall genes were taxonomically profiled, and the copy number of syntenic genes across the species contained within each community was also assessed. Finally, the taxonomic and copy number data from different syntenic communities were used to analyze the occurrence of a divergent genomic organization for each cell wall gene function between Poaceae, noncommelinoid monocots, and eudicots.

Analysis of tandem gene clusters

The ordinal position of cell wall genes along genomic BED files was assessed for each genome of the study, to identify the occurrence of tandem gene clusters of homologous genes along chromosomes (custom R script, available at https://github.com/Francesco1994WUR/Cell_wall_phylogenomics). The results of this analysis were used to evaluate the proportion of clustered and singleton genes out of the total genes belonging to a certain cell wall function and the mean size of the tandem clusters found in every species. Moreover, the occurrence of clustered genes within syntenic communities detected in the synteny analysis was also assessed.

Phylogenetic analyses

A group of gene families that displayed genomic patterns particularly relevant in the context of the differences between type I and type II cell walls were selected for a more detailed genetic study encompassing phylogenetic analysis (see Section 2). For this purpose, the protein sequences of the genes belonging to these gene families were aligned with MAFFT v7.453 (FFT-NS-2 algorithm) (Katoh and Standley 2013), with default parameters except gap opening penalty, set to 1.0. MAFFT alignments were trimmed using TrimAl v1.2 (Capella-Gutiérrez et al. 2009), with default parameters. Finally, RAxML v8.2.9 (Stamatakis 2014) was used to build phylogenetic trees out of trimmed alignments (PROTCATBLOSUM62 substitution matrix; 100 bootstraps). Phylogenetic trees were plotted and annotated using iTOL (Letunic and Bork 2019). The TAIR database (Huala et al. 2001) was used to localize critical Arabidopsis genes of the families analyzed within each tree, while BLAST (Altschul et al. 1990) was used to find and localize relevant grass homologs of those trees within each tree ($E = 1E^{-3}$).

Analysis of selection pressure

Differences in the rates of selection pressure between the Poaceae and eudicot genes contained in both differentiated and shared genomic contexts of the genes included in the

BLH6 pathway were evaluated by using the EasyCodeML implementation (Gao et al. 2019) of the CodeML program from the PAML4.0 package (Yang 2007). For each phylogenetic tree corresponding to a gene family within the *BLH6* pathway (see Fig. 4), Poaceae clade(s) (corresponding to distinct syntenic communities from eudicots or contained in shared syntenic communities with eudicots, depending on the gene family) have been set as foreground in a branch model (model, 2; NSsites, 0) to estimate dN:dS ratios specifically for those branches. A basic model was also run (model, 0; NSsites, 0) to estimate dN:dS ratios at the whole-tree level (background), and a lLRT (also implemented in EasyCodeML) was performed to test for significantly different selection pressure between foreground and background tree branches. For computational reasons, a random set of 80 leaves was selected from the set foreground and background branches in order to run CodeML models and tests.

Accession numbers

Accession numbers of the sequence data from this article can be found in Supplemental Tables S1 (all cell wall genes) and S2 (genome assemblies).

Author contributions

F.P. designed and conducted this research and wrote the article, with inputs and supervision from L.M.T., M.E.S., and E.N.v.L. L.M.T., M.E.S., and E.N.v.L. corrected the manuscript. L.M.T., M.E.S., and EN.v.L. approved the final manuscript.

Supplemental data

The following materials are available in the online version of this article.

Supplemental Table S1. The cell wall genes considered in this study for all the 169 genomes, classified into 150 gene families.

Supplemental Table S2. The 169 angiosperm genomes used in the study.

Supplemental Table S3. Complete gene copy number data for the 169 genomes and the 150 cell wall gene families studied.

Supplemental Table S4. Copy number data of the 150 cell wall gene families of the study, highlighting differences between Poaceae and dicots.

Supplemental Table S5. Loadings of the PCA.

Supplemental Table S6. Results of PAML tests for selection pressure on the genes within the BLH6 pathway.

Supplemental Table S7. Tandem cell wall gene cluster statistics.

Supplemental Table S8. The 1,313 Arabidopsis cell wall genes used as seeds in the search for all the cell wall genes in all the 169 angiosperm genomes.

Supplemental Data Set 1. Synteny network of all the cell wall genes from the 150 gene families and 169 plant genomes analyzed (available at 4TU.researchdata, DOI: https://doi.org/10.4121/21564756).

Supplemental Data Set 2. Annotation of tandem gene clusters for all the cell wall genes from the 150 gene families and 169 plant genomes analyzed (available at 4TU.researchdata, DOI: https://doi.org/10.4121/22068791).

Funding

This research is part of a project that received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 727698.

Conflict of interest statement. None declared.

Data availability

All the data supporting the research presented in this study are included in the article or in the Supplemental data.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990:**215**(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anders N, Wilkinson MD, Lovegrove A, Freeman J, Tryfona T, Pellny TK, Weimar T, Mortimer JC, Stott K, Baker JM. Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses. Proc Natl Acad Sci U S A. 2012:**109**(3):989–993. https://doi.org/10.1073/pnas.1115858109

Artur MAS, Zhao T, Ligterink W, Schranz E, Hilhorst HW. Dissecting the genomic diversification of late embryogenesis abundant (LEA) protein gene families in plants. Genome Biol Evol. 2019:**11**(2):459–471. https://doi.org/10.1093/gbe/evy248

Atmodjo MA, Hao Z, Mohnen D. Evolving views of pectin biosynthesis. Annu Rev Plant Biol. 2013:**64**(1):747–779. https://doi.org/10.1146/annurev-arplant-042811-105534

Bartley LE, Peck ML, Kim S-R, Ebert B, Manisseri C, Chiniquy DM, Sykes R, Gao L, Rautengarten C, Vega-Sánchez ME. Overexpression of a BAHD acyltransferase, OsAt10, alters rice cell wall hydroxycinnamic acid content and saccharification. Plant Physiol. 2013:**161**(4):1615–1633. https://doi.org/10.1104/pp.112.208694

Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015:**12**(1):59–60. https://doi.org/10.1038/nmeth.3176

Burton RA, Fincher GB. Current challenges in cell wall biology in the cereals and grasses. Front Plant Sci. 2012:**3**:130. https://doi.org/10.3389/fpls.2012.00130

Burton RA, Fincher GB. Evolution and development of cell walls in cereal grains. Front Plant Sci. 2014:**5**:456. https://doi.org/10.3389/fpls.2014.00456

Burton RA, Wilson SM, Hrmova M, Harvey AJ, Shirley NJ, Medhurst A, Stone BA, Newbigin EJ, Bacic A, Fincher GB. Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1, 3; 1, 4)-ß-D-glucans. Science 2006:**311**(5769):1940–1942. https://doi.org/10.1126/science.1122975

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 2009:**25**(15):1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Carpita NC, Gibeaut DM. Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical

properties of the walls during growth. Plant J. 1993:**3**(1):1–30. https://doi.org/10.1111/j.1365-313X.1993.tb00007.x

Carpita NC, Mccann MC. Redesigning plant cell walls for the biomass-based bioeconomy. J Biol Chem. 2020:**295**(44):15144–15157. https://doi.org/10.1074/jbc.REV120.014561

Cenci A, Chantret N, Rouard M. Glycosyltransferase family 61 in liliopsida (monocot): the story of a gene family expansion. Front Plant Sci. 2018:**9**:1843. https://doi.org/10.3389/fpls.2018.01843

Chandrakanth NN, Zhang C, Freeman J, De Souza WR, Bartley LE, Mitchell RA. Modification of plant cell walls with hydroxycinnamic acids by BAHD acyltransferases. Front Plant Sci. 2023:**13**:1088879. https://doi.org/10.3389/fpls.2022.1088879

De Oliveira DM, Finger-Teixeira A, Rodrigues Mota T, Salvador VH, Moreira-Vilar FC, Correa Molinari HB, Craig Mitchell RA, Marchiosi R, Ferrarese-Filho O, Dantas Dos Santos W. Ferulic acid: a key component in grass lignocellulose recalcitrance to hydrolysis. Plant Biotechnol J. 2015:**13**(9):1224–1232. https://doi.org/10.1111/pbi.12292

De Souza WR, Martins PK, Freeman J, Pellny TK, Michaelson LV, Sampaio BL, Vinecky F, Ribeiro AP, Da Cunha BA, Kobayashi AK. Suppression of a single BAHD gene in *Setaria viridis* causes large, stable decreases in cell wall feruloylation and increases biomass digestibility. New Phytologist 2018:**218**(1):81–93. https://doi.org/10.1111/nph.14970

Dewey CN. Positional orthology: putting genomic evolutionary relationships into context. Brief Bioinform. 2011:**12**(5):401–412. https://doi.org/10.1093/bib/bbr040

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A. The Pfam protein families database in 2019. Nucleic Acids Res. 2019:**47**(D1):D427–D432. https://doi.org/10.1093/nar/gky995

Feijao C, Morreel K, Anders N, Tryfona T, Busse-Wicher M, Kotake T, Boerjan W, Dupree P. Hydroxycinnamic acid-modified xylan side chains and their cross-linking products in rice cell walls are reduced in the xylosyl arabinosyl substitution of xylan 1 mutant. Plant J. 2022:**109**(5):1152–1167. https://doi.org/10.1111/tpj.15620

Fincher G, Stone B. Cell walls and their components in cereal grain technology. Adv Cereal Sci Technol. 1986:**8**:207–295.

Flagel LE, Wendel JF. Gene duplication and evolutionary novelty in plants. New Phytologist 2009:**183**(3):557–564. https://doi.org/10.1111/j.1469-8137.2009.02923.x

Gale MD, Devos KM. Comparative genetics in the grasses. Proc Natl Acad Sci U S A. 1998:**95**(5):1971–1974. https://doi.org/10.1073/pnas.95.5.1971

Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW. EasyCodeML: a visual tool for analysis of selection using CodeML. Ecol Evol. 2019:**9**(7):3891–3898. https://doi.org/10.1002/ece3.5015

Guerriero G, Hausman J-F, Legay S. Silicon and the plant extracellular matrix. Front Plant Sci. 2016:**7**:463. https://doi.org/10.3389/fpls.2016.00463

Hatfield RD, Rancour DM, Marita JM. Grass cell walls: a story of cross-linking. Front Plant Sci. 2017:**7**:2056. https://doi.org/10.3389/fpls.2016.02056

Hirano K, Kondo M, Aya K, Miyao A, Sato Y, Antonio BA, Namiki N, Nagamura Y, Matsuoka M. Identification of transcription factors involved in rice secondary cell wall formation. Plant Cell Physiol. 2013:**54**(11):1791–1802. https://doi.org/10.1093/pcp/pct122

Huala E, Dickerman AW, Garcia-Hernandez M, Weems D, Reiser L, Lafond F, Hanley D, Kiphart D, Zhuang M, Huang W. The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Res. 2001:**29**(1):102–105. https://doi.org/10.1093/nar/29.1.102

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS. Ancestral polyploidy in seed plants and angiosperms. Nature 2011:**473**(7345):97–100. https://doi.org/10.1038/nature09916

Jozwiak A, Sonawane PD, Panda S, Garagounis C, Papadopoulou KK, Abebie B, Massalha H, Almekias-Siegl E, Scherf T, Aharoni A. Plant terpenoid metabolism co-opts a component of the cell wall biosynthesis machinery. Nat Chem Biol. 2020:**16**(7):740–748. https://doi.org/10.1038/s41589-020-0541-x

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013:**30**(4):772–780. https://doi.org/10.1093/molbev/mst010

Kerstens MH, Schranz ME, Bouwmeester K. Phylogenomic analysis of the APETALA2 transcription factor subfamily across angiosperms reveals both deep conservation and lineage-specific patterns. Plant J. 2020:**103**(4):1516–1524. https://doi.org/10.1111/tpj.14843

Kim S-J, Chandrasekar B, Rea AC, Danhof L, Zemelis-Durfee S, Thrower N, Shepard ZS, Pauly M, Brandizzi F, Keegstra K. The synthesis of xyloglucan, an abundant plant cell wall polysaccharide, requires CSLC function. Proc Natl Acad Sci U S A. 2020:**117**(33):20316–20324. https://doi.org/10.1073/pnas.2007245117

Kondrashov FA. Gene duplication as a mechanism of genomic adaptation to a changing environment. Proc R Soc B: Biol Sci. 2012:**279**(1749):5048–5057. https://doi.org/10.1098/rspb.2012.1108

Kono TJ, Brohammer AB, Mcgaugh SE, Hirsch CN. Tandem duplicate genes in maize are abundant and date to two distinct periods of time. G3 (Bethesda) 2018:**8**(9):3049–3058. https://doi.org/10.1534/g3.118.200580

Kozlova LV, Nazipova AR, Gorshkov OV, Petrova AA, Gorshkova TA. Elongating maize root: zone-specific combinations of polysaccharides from type I and type II primary cell walls. Sci Rep. 2020:**10**(1):1–20. https://doi.org/10.1038/s41598-020-67782-0

Lee S, Choi S, Jeon D, Kang Y, Kim C. Evolutionary impact of whole genome duplication in Poaceae family. J Crop Sci Biotechnol. 2020:**23**:413–425. https://doi.org/10.1007/s12892-020-00049-2

Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019:**47**(W1):W256–W259. https://doi.org/10.1093/nar/gkz239

Little A, Schwerdt JG, Shirley NJ, Khor SF, Neumann K, O'donovan LA, Lahnstein J, Collins HM, Henderson M, Fincher GB. Revised phylogeny of the cellulose synthase gene superfamily: insights into cell wall evolution. Plant Physiol. 2018:**177**(3):1124–1141. https://doi.org/10.1104/pp.17.01718

Liu Y, You S, Taylor-Teeples M, Li WL, Schuetz M, Brady SM, Douglas CJ. BEL1-LIKE HOMEODOMAIN6 and KNOTTED ARABIDOPSIS THALIANA7 interact and regulate secondary cell wall formation via repression of REVOLUTA. Plant Cell 2014:**26**-(12):4843–4861. https://doi.org/10.1105/tpc.114.128322

Lye ZN, Purugganan MD. Copy number variation in domestication. Trends Plant Sci. 2019:**24**(4):352–365. https://doi.org/10.1016/j.tplants.2019.01.003

Mishler-Elmore JW, Zhou Y, Sukul A, Oblak M, Tan L, Faik A, Held MA. Extensins: self-assembly, crosslinking, and the role of peroxidases. Front Plant Sci. 2021:**12**:664738. https://doi.org/10.3389/fpls.2021.664738

Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 2013:**41**(12):e121. https://doi.org/10.1093/nar/gkt263

Molinari HB, Pellny TK, Freeman J, Shewry PR, Mitchell RA. Grass cell wall feruloylation: distribution of bound ferulate and candidate gene expression in *Brachypodium distachyon*. Front Plant Sci. 2013:**4**:50. https://doi.org/10.3389/fpls.2013.00050

Pancaldi F, Trindade LM. Marginal lands to grow novel bio-based crops: a plant breeding perspective. Front Plant Sci. 2020:**11**:227. https://doi.org/10.3389/fpls.2020.00227

Pancaldi F, Van Loo EN, Schranz ME, Trindade LM. Genomic architecture and evolution of the cellulose synthase gene superfamily as revealed by phylogenomic analysis. Front Plant Sci. 2022a:**13**:870818. https://doi.org/10.3389/fpls.2022.870818

Pancaldi F, Vlegels D, Rijken H, Van Loo EN, Trindade LM. Detection and analysis of syntenic quantitative trait loci controlling cell wall quality in angiosperms. Front Plant Sci. 2022b:**13**:855093. https://doi.org/10.3389/fpls.2022.855093

Pena MJ, Kulkarni AR, Backe J, Boyd M, O'neill MA, York WS. Structural diversity of xylans in the cell walls of monocots. Planta 2016:**244**(3):589–606. https://doi.org/10.1007/s00425-016-2527-1

Penning BW, Mccann MC, Carpita NC. Evolution of the cell wall gene families of grasses. Front Plant Sci. 2019:**10**:1205. https://doi.org/10.3389/fpls.2019.01205

Picart-Picolo A, Grob S, Picault N, Franek M, Llauro C, Halter T, Maier TR, Jobet E, Descombin J, Zhang P. Large tandem duplications affect gene expression, 3D organization, and plant–pathogen response. Genome Res. 2020:**30**(11):1583–1592. https://doi.org/10.1101/gr.261586.120

Qin W, Yin Q, Chen J, Zhao X, Yue F, He J, Yang L, Liu L, Zeng Q, Lu F. The class II KNOX transcription factors KNAT3 and KNAT7 synergistically regulate monolignol biosynthesis in Arabidopsis. J Exp Bot. 2020:**71**(18):5469–5483. https://doi.org/10.1093/jxb/eraa266

Ralph J. Hydroxycinnamates in lignification. Phytochem Rev. 2010:**9**(1):65–83. https://doi.org/10.1007/s11101-009-9141-9

Rao X, Dixon RA. Current models for transcriptional regulation of secondary cell wall biosynthesis in grasses. Front Plant Sci. 2018:**9**:399. https://doi.org/10.3389/fpls.2018.00399

Rosvall M, Axelsson D, Bergstrom CT. The map equation. Euro Phys J Special Top. 2009:**178**(1):13–23. https://doi.org/10.1140/epjst/e2010-01179-1

Rosvall M, Bergstrom CT. An information–theoretic framework for resolving community structure in complex networks. Proc Natl Acad Sci U S A. 2007:**104**(18):7327–7331. https://doi.org/10.1073/pnas.0611034104

Sampedro J, Guttman M, L LI, & Cosgrove C, J D. Evolutionary divergence of β–expansin structure and function in grasses parallels emergence of distinctive primary cell wall traits. Plant J. 2015:**81**(1):108–120. https://doi.org/10.1111/tpj.12715

Sarkar P, Bosneaga E, Auer M. Plant cell walls throughout evolution: towards a molecular understanding of their design principles. J Exp Bot. 2009:**60**(13):3615–3635. https://doi.org/10.1093/jxb/erp245

Schwerdt JG, Mackenzie K, Wright F, Oehme D, Wagner JM, Harvey AJ, Shirley NJ, Burton RA, Schreiber M, Halpin C. Evolutionary dynamics of the cellulose synthase gene superfamily in grasses. Plant Physiol. 2015:**168**(3):968–983. https://doi.org/10.1104/pp.15.00140

Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. Methods Mol Biol. 2019:**1962**:227–245. https://doi.org/10.1007/978-1-4939-9173-0_14

Smith RA, Cass CL, Mazaheri M, Sekhon RS, Heckwolf M, Kaeppler H, De Leon N, Mansfield SD, Kaeppler SM, Sedbrook JC. Suppression of CINNAMOYL-CoA REDUCTASE increases the level of monolignol ferulates incorporated into maize lignins. Biotechnol Biofuels. 2017:**10**(1):1–10. https://doi.org/10.1186/s13068-017-0793-1

Somerville C, Bauer S, Brininstool G, Facette M, Hamann T, Milne J, Osborne E, Paredez A, Persson S, Raab T. Toward a systems approach to understanding plant cell walls. Science 2004:**306**(5705):2206–2211. https://doi.org/10.1126/science.1102765

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014:**30**(9):1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Tamasloukht B, Wong Quai Lam MS-J, Martinez Y, Tozo K, Barbier O, Jourda C, Jauneau A, Borderies G, Balzergue S, & Renou J-P. Characterization of a cinnamoyl-CoA reductase 1 (CCR1) mutant in maize: effects on lignification, fibre development, and global gene expression. J Exp Bot. 2011:**62**(11):3837–3848. https://doi.org/10.1093/jxb/err077

Taylor-Teeples M, Lin L, De Lucas M, Turco G, Toal T, Gaudinier A, Young N, Trabucco G, Veling M, Lamothe R. An Arabidopsis gene regulatory network for secondary cell wall synthesis. Nature 2015:**517**(7536):571–575. https://doi.org/10.1038/nature14099

Tu Y, Rochfort S, Liu Z, Ran Y, Griffith M, Badenhorst P, Louie GV, Bowman ME, Smith KF, Noel JP. Functional analyses of caffeic acid O-methyltransferase and cinnamoyl-CoA-reductase genes from perennial ryegrass (*Lolium perenne*). Plant Cell 2010:**22**(10):3357–3373. https://doi.org/10.1105/tpc.109.072827

Vogel J. Unique aspects of the grass cell wall. Curr Opin Plant Biol. 2008:**11**(3):301–307. https://doi.org/10.1016/j.pbi.2008.03.002

Voiniciuc C, Günl M, Schmidt MH-W, Usadel B. Highly branched xylan made by IRREGULAR XYLEM14 and MUCILAGE-RELATED21 links mucilage to Arabidopsis seeds. Plant Physiol. 2015:**169**(4):2481–2495. https://doi.org/10.1104/pp.15.01441

Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee T-H, Jin H, Marler B, Guo H. MCScanx: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012:**40**(7):e49. https://doi.org/10.1093/nar/gkr1293

Wang X, Wang J, Jin D, Guo H, Lee T-H, Liu T, Paterson AH. Genome alignment spanning major Poaceae lineages reveals heterogeneous evolutionary rates and alters inferred dates for key evolutionary events. Mol Plant. 2015:**8**(6):885–898. https://doi.org/10.1016/j.molp.2015.04.004

Xu Z, Pu X, Gao R, Demurtas OC, Fleck SJ, Richter M, He C, Ji A, Sun W, Kong J. Tandem gene duplications drive divergent evolution of caffeine and crocin biosynthetic pathways in plants. BMC Biol. 2020:**18**(1):1–14. https://doi.org/10.1186/s12915-020-00795-3

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007:**24**(8):1586–1591. https://doi.org/10.1093/molbev/msm088

Yokoyama R. A genomic perspective on the evolutionary diversity of the plant cell wall. Plants 2020:**9**(9):1195. https://doi.org/10.3390/plants9091195

Yokoyama R, Nishitani K. Genomic basis for cell-wall diversity in plants. A comparative approach to gene families in rice and Arabidopsis. Plant Cell Physiol. 2004:**45**(9):1111–1121. https://doi.org/10.1093/pcp/pch151

Zabotina OA. Xyloglucan and its biosynthesis. Front Plant Sci. 2012:**3**:134. https://doi.org/10.3389/fpls.2012.00134

Zhao T, Holmer R, De Bruijn S, Angenent GC, Van Den Burg HA, Schranz ME. Phylogenomic synteny network analysis of MADS-box transcription factor genes reveals lineage-specific transpositions, ancient tandem duplications, and deep positional conservation. Plant Cell 2017:**29**(6):1278–1292. https://doi.org/10.1105/tpc.17.00312

Zhao T, Schranz ME. Network approaches for plant phylogenomic synteny analysis. Curr Opin Plant Biol. 2017:**36**:129–134. https://doi.org/10.1016/j.pbi.2017.03.001

Zhao T, Schranz ME. Network-based microsynteny analysis identifies major differences and genomic outliers in mammalian and angiosperm genomes. Proc Natl Acad Sci U S A. 2019:**116**(6):2165–2174. https://doi.org/10.1073/pnas.1801757116

Zhong R, Cui D, Ye ZH. Secondary cell wall biosynthesis. New Phytologist 2019:**221**(4):1703–1723. https://doi.org/10.1111/nph.15537

Zhu XF, Wan JX, Wu Q, Zhao XS, Zheng SJ, Shen RF. PARVUS affects aluminium sensitivity by modulating the structure of glucuronoxylan in *Arabidopsis thaliana*. Plant Cell Environ. 2017:**40**(9):1916–1925. https://doi.org/10.1111/pce.12999

Zoghbi-Rodríguez NM, Gamboa-Tuz SD, Pereira-Santana A, Rodríguez-Zapata LC, Sánchez-Teyer LF, Echevarría-Machado I. Phylogenomic and microsynteny analysis provides evidence of genome arrangements of high-affinity nitrate transporter gene families of plants. Int J Mol Sci. 2021:**22**(23):13036. https://doi.org/10.3390/ijms222313036