

# In Silico Characterization and Functional Validation of Cell Wall Modification Genes Imparting Waterlogging Tolerance in Maize

Bioinformatics and Biology Insights  
Volume 11: 1–13  
© The Author(s) 2017  
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/1177932217747277



Kanika Arora<sup>1,2</sup>, Kusuma Kumari Panda<sup>2</sup>, Shikha Mittal<sup>1</sup>, Mallana Gowdra Mallikarjuna<sup>1</sup> and Nepolean Thirunavukkarasu<sup>1,3</sup>

<sup>1</sup>Division of Genetics, ICAR—Indian Agricultural Research Institute, New Delhi, India.

<sup>2</sup>Amity Institute of Biotechnology, Amity University, Noida, Uttar Pradesh, India. <sup>3</sup>Maize Research Lab, Division of Genetics, ICAR—Indian Agricultural Research Institute, New Delhi, India.

**ABSTRACT:** Cell wall modification (CWM) promotes the formation of aerenchyma in roots under waterlogging conditions as an adaptive mechanism. Lysigenous aerenchyma formation in roots improves oxygen transfer in plants, which highlights the importance of CWM as a focal point in waterlogging stress tolerance. We investigated the structural and functional compositions of CWM genes and their expression patterns under waterlogging conditions in maize. Cell wall modification genes were identified for 3 known waterlogging-responsive *cis*-acting regulatory elements, namely, GC motif, anaerobic response elements, and G-box, and 2 unnamed elements. Structural motifs mapped in CWM genes were represented in genes regulating waterlogging stress-tolerant pathways, including fermentation, glycolysis, programmed cell death, and reactive oxygen species signaling. The highly aligned regions of characterized and uncharacterized CWM proteins revealed common structural domains amongst them. Membrane spanning regions present in the protein structures revealed transmembrane activity of CWM proteins in the plant cell wall. Cell wall modification proteins had interacted with ethylene-responsive pathway regulating genes (E3 ubiquitin ligases RNG finger and F-box) in a maize protein-protein interaction network. Cell wall modification genes had also coexpressed with energy metabolism, programmed cell death, and reactive oxygen species signaling, regulating genes in a single coexpression cluster. These configurations of CWM genes can be used to modify the protein expression in maize under waterlogging stress condition. Our study established the importance of CWM genes in waterlogging tolerance, and these genes can be used as candidates in introgression breeding and genome editing experiments to impart tolerance in maize hybrids.

**KEYWORDS:** Candidate genes, cell wall modification, maize, plant stress, waterlogging tolerance

**RECEIVED:** August 5, 2017. **ACCEPTED:** November 11, 2017.

**TYPE:** Original Research

**FUNDING:** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by the ICAR Network Projects on Transgenics in Crop Plants (Maize Functional Genomics Component: 21-22). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHOR:** Nepolean Thirunavukkarasu, Maize Research Lab, Division of Genetics, ICAR—Indian Agricultural Research Institute, Pusa, New Delhi 110012, India. Email: tnepolean@gmail.com

## Introduction

Cell wall modification (CWM) is an adaptive strategy during waterlogging stress in crop plants. The cell wall is degraded in the process of programmed cell death (PCD),<sup>1</sup> caused due to the formation of reactive oxygen species (ROS) that degrades cell wall polysaccharides.<sup>2</sup> Such changes in the plant occur in early phases of waterlogging stress. Under waterlogging conditions, oxygen deficiency occurs in the root zone and results in no-diffusion zone of oxygen from roots to shoots. This explains the prerequisite of decomposition of cell wall components in root cortical cells, which leads to PCD and subsequent formation of lysigenous aerenchyma to overcome oxygen-deficient conditions.<sup>3,4</sup>

Cell wall modification in plants is regulated by loosening and expansion of the cell wall.<sup>5</sup> This development in plants involves depolymerization, hydrolyzation, and deacetylation of cell wall components. First, depolymerization of pectin network is regulated by methyl de-esterification which is influenced by the activities of pectin methylesterases and polygalacturonases.<sup>6</sup> Pectin polysaccharides are linked by  $\alpha$ -1,4 glycosidic bonds in the plant cell wall and are hydrolyzed by polygalacturonases.<sup>7</sup> Second, deacetylation of pectin polysaccharides is catalyzed by

pectin acylesterases (PAE), resulting in loosening of the primary cell wall.<sup>8</sup> Xyloglucans and linked glucans are among the hemicellulosic polysaccharides present in the plant cell wall,<sup>9</sup> which are degraded by the enzymes xyloglucan endo-transglycosylase or hydrolase (XTH) and endoglucanases. XTH can cut and re-ligate the xyloglucan chains that assist in cell wall loosening and subsequent expansion,<sup>10</sup> whereas endoglucanases can hydrolyze the  $\beta$ -1,4 linkages between linear glucans that degrade the cell wall. Finally, loosening of cell wall is catalyzed by expansins that induce cell wall extension and stress relaxation.<sup>11</sup> Under waterlogging conditions, transcript abundance of expansins is associated with cell wall loosening in the elongation zone of the roots.<sup>12</sup> Plasma membrane and cell wall also comprise of structurally complex protein arabinogalactan, which play an important role in signaling events.

Cell wall modification proteins are researched for catalytic sites and are characterized by conserved amino acids and motifs. A pattern of conservation is specific to the type of proteins. Polygalacturonases possess 4 highly conserved motifs NTD, G/QDD, G/SHG, and RIK in catalytic and substrate-binding regions.<sup>13</sup> *Poplar* PAE comprised a conserved GX SXG motif



which is a characteristic of hydrolytic enzymes.<sup>8</sup> *DkXTH* proteins revealed a conserved motif in the catalytic site, N-linked glycosylation site, and 2 cysteine residues in the C-terminal region, suggesting a similarity with XTHs in other plants.<sup>14</sup> Active sites of endoglucanases revealed membrane-spanning domains that anchor protein to the plasma membrane for degradation and re-ligation of cell wall components.<sup>15,16</sup> Expansins expressed in deep water rice were characterized by a series of cysteine residues in N terminus, HFD motif in the central region, and a series of tryptophan residues near C terminus.<sup>17</sup> Arabinogalactan proteins are hydroxyproline-rich glycoproteins which include glycosylation motifs that are repetitive in nature.<sup>18</sup> These motifs include contiguous hydroxyproline residues which are characterized as sites of HRGP arabinosylation, (Ala-Pro-Ala-Pro)<sub>n</sub>, (Thr-Pro-Thr-Pro)<sub>n</sub>, and (Val-Pro-Val-Pro)<sub>n</sub>.<sup>18</sup>

Earlier studies have characterized the crystalline structures and active site of PAE, XTH, and  $\beta$ -expansins in *Poplar* and maize.<sup>19,20</sup> Stress-specific proteins have different amino acid conservation patterns that are still uncovered for CWM proteins in waterlogged maize. Therefore, the present investigation aims to study the conservation pattern of motifs, catalytic residues, and phylogenetic analysis of CWM proteins in waterlogged maize. In addition, selected CWM genes were validated in the waterlogging-tolerant and waterlogging-susceptible genotypes using gene expression assay. Sequential and structural information generated from this experiment would be useful in designing introgression breeding programs to improve waterlogging tolerance in maize hybrids.

## Materials and Methods

### Selection of genes

Twelve CWM genes from 7 gene families *arabinogalactan*, *XTH*, *polygalacturonases*, *expansins*, *endoglucanases*, *PAE*, and *pectin esterases* were selected from previous published data<sup>21,22</sup> (genome-wide expression studies; Table 1), and sequences were retrieved from *Cell Wall Genomics* database (<https://cellwall.genomics.purdue.edu>). The genes were subjected to structural and functional analysis through in silico tools and expression assay.

### Plant material and stress conditions

A pair of contrasting maize genotypes SKV239 (tolerant) and CML22 (susceptible) was screened for their response to waterlogging stress (Figure 1; Supplementary Table S1). These contrasting genotypes represent maize inbred lines. Plants were sowed in cups maintained under optimal conditions. At 3-leaf stage, waterlogging stress was imposed by maintaining the water table up to 3 cm above the soil level to create adequate stress condition. Twenty seedlings were replicated each for control and stressed plants. Phenotypic values of chlorophyll content (Chlorophyll Concentration Index), root dry weight, and shoot dry weight were recorded. Chlorophyll content was

measured using SPAD meter. Leaf and root samples were isolated separately after 5 days of stress treatment.

### RNA isolation and quantitative real-time polymerase chain reaction expression analysis

Total RNA was isolated from leaf and root tissue samples using RNeasy kit (Qiagen, Hilden, Germany) as per manufacturer's protocol. Quality and quantity were checked using agarose gel electrophoresis and NanoDrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). Messenger RNA (mRNA) was transformed into first-strand complementary DNA (cDNA) using cDNA-synthesis kit (Thermo Fisher Scientific, Waltham, MA, USA) at 44°C for 60 minutes, followed by 92°C for 10 minutes. First-strand cDNA was investigated for expression using quantitative real-time polymerase chain reaction (qRT-PCR) (Agilent Technologies, Santa Clara, CA, USA). Primer designing of the 12 CWM genes was done by using the IDT software (<http://eu.idtdna.com>; Supplementary Table S2). The PCR reaction was carried out at 95°C for 4 minutes, followed by 40 cycles of 95°C for 15 seconds, 60°C for 30 seconds, and 72°C for 1 minute. The expression values (CT values) measured in root and leaf samples of tolerant and susceptible genotypes were statistically validated using paired *t*-test. The CT values measured using qRT-PCR were transformed to fold change. Fold change is defined as the expression difference of a gene in stress sample over nonstress sample. A threshold level of 1-fold change between stress and nonstress samples was used to designate a gene as differentially expressed in waterlogging stress conditions.

### In Silico analysis of CWM genes

The gene structure of CWM genes was revealed using Gene Structure Display Server<sup>23</sup> (GSDS). The gene structure information was added as an input to GSDS server in a BED format file. Sequence motifs present in CWM genes were scanned over their nucleotide sequences using MEME suite.<sup>24</sup> The cDNA sequences of CWM genes were added as an input to the MEME suite for sequence motifs search. Parameters of MEME suite were optimized as maximum number of motifs set to 5 and minimum 1 occurrence of motif site per sequence. *Cis*-acting regulatory elements (CREs) present in CWM genes were scanned using Plant CARE database.<sup>25</sup> The cDNA sequences of CWM genes were searched for CREs in Plant CARE database. Neighborhood-joining phylogenetic trees of maize (*Zea mays* L.), *Arabidopsis*, barley (*Hordeum vulgare* L.), rice (*Oryza sativa* L.), and sorghum (*Sorghum bicolor* L.) CWM genes were generated using *R* packages phangorn<sup>26</sup> and ape.<sup>27</sup> Protein structures of these CWM genes were predicted using IntFOLD server.<sup>28</sup> Amino acid sequences of CWM proteins were added as an input to the IntFOLD server. These predicted protein structures were investigated for structural motifs using SA-Mot server<sup>29</sup> and analyzed based on their sequences

**Table 1.** Characteristics of selected candidate genes involved in cell wall modification.

S. NO.	GENE	GENE MODEL	CHROMOSOME	GENE START (BP)	GENE END (BP)	GENE LENGTH (BP)	ANNOTATION	REFERENCES
1	<i>Arabinogalactan</i>	GRMZM2G003165	7	80831434	80834726	3292	Fasciclin-like arabinogalactan protein 8	22
2	<i>XTH</i>	GRMZM2G004699	4	191756828	191759110	2282	XTH protein	22
3	<i>XTH</i>	GRMZM2G039919	9	144146526	144149307	2781	Uncharacterized protein	33
4	<i>Polygalacturonases</i>	GRMZM2G037431	3	216010535	216013992	3457	Polygalacturonase	21
5	<i>Polygalacturonases</i>	AC231180.2_FG006	8	75646247	75647874	1627	Uncharacterized protein	33
6	<i>Expansins</i>	GRMZM2G105844	5	206888609	206890560	1951	Uncharacterized protein	22
7	<i>Expansins</i>	GRMZM2G094523	3	34036069	34037197	1128	Uncharacterized protein	33
8	<i>Endoglucanases</i>	GRMZM2G141911	5	71930286	71932471	2185	Endoglucanase	21
9	<i>Endoglucanases</i>	AC199765.4_FG008	2	27015679	27017169	1490	Uncharacterized protein	33
10	<i>Pectin acetylsterases</i>	GRMZM2G156365	2	12031024	12034766	3742	Pectin acetylsterase	22,33
11	<i>Pectin esterases</i>	GRMZM2G162333	2	19888815	19891284	2469	Pectin esterase	22
12	<i>Pectin esterases</i>	GRMZM2G175499	4	62359341	62364922	5581	LOC100281178	33

Abbreviation: XTH, xyloglucan endo-transglycosylase or hydrolase.



**Figure 1.** Phenotypic response of SKV 239 (tolerant) and CML 22 (susceptible) maize genotypes to waterlogging stress.

using Phyre2 server.<sup>30</sup> Default parameters of SA-Mot and Phyre2 servers were used. The normal mode of Phyre 2 server that models structures based on a single template was implemented. Multiple structural alignment was investigated through ProCKSI server where different tools were used to analyze the predicted protein structures.<sup>31</sup> Parameters of ProCKSI server were set to default settings. These structures were then searched for similarity with the submitted Protein Data Bank (PDB) structures using an online PDBeFold server.<sup>32</sup> Predicted structures were searched for similarity against the whole PDB archive as a target in PDBeFold server.

#### *Protein-protein interaction network of CWM proteins and waterlogging-responsive maize proteins*

Waterlogging-responsive maize genes were searched in in-house waterlogging genome-wide expression<sup>22</sup> and transcriptomics data.<sup>33</sup> The waterlogging-responsive proteins were searched for interactions with CWM proteins *via* protein-protein interaction network constructed through a cytoscape plugin stringApp.<sup>34</sup> Protein-protein interaction network of CWM proteins and waterlogging-responsive proteins was imported from STRING database using stringApp plugin. Network was analyzed using NetworkAnalyzer in cytoscape.

## Results

### *Phenotyping of maize genotypes*

Waterlogging-tolerant SKV239 and waterlogging-susceptible CML22 genotypes were grown up to 3-leaf stage. After that, the plants were maintained under normal conditions and subjected to waterlogging stress conditions (Figure 1). During the stress period, the plants were phenotyped for chlorophyll content (Chlorophyll Concentration Index), root dry weight (gram per plant), and shoot dry weight (gram per plant) (Supplementary Table S1). Chlorophyll content had significantly decreased in susceptible genotype (27.7–21.6), whereas a slight reduction was found in tolerant genotype (28.4–26.8) from normal to stress. Root dry weight of waterlogging-susceptible genotype had significantly decreased from 0.32 to 0.12 g/plant in normal and stress conditions, respectively. On the contrary, root dry weight of waterlogging-tolerant genotype had slightly increased from 0.36 to 0.37 g/plant in normal

and stress conditions, respectively. Shoot dry weight of waterlogging-tolerant genotype had significantly decreased (0.55–0.19 g/plant) compared with the waterlogging-susceptible genotype (0.51–0.46 g/plant).

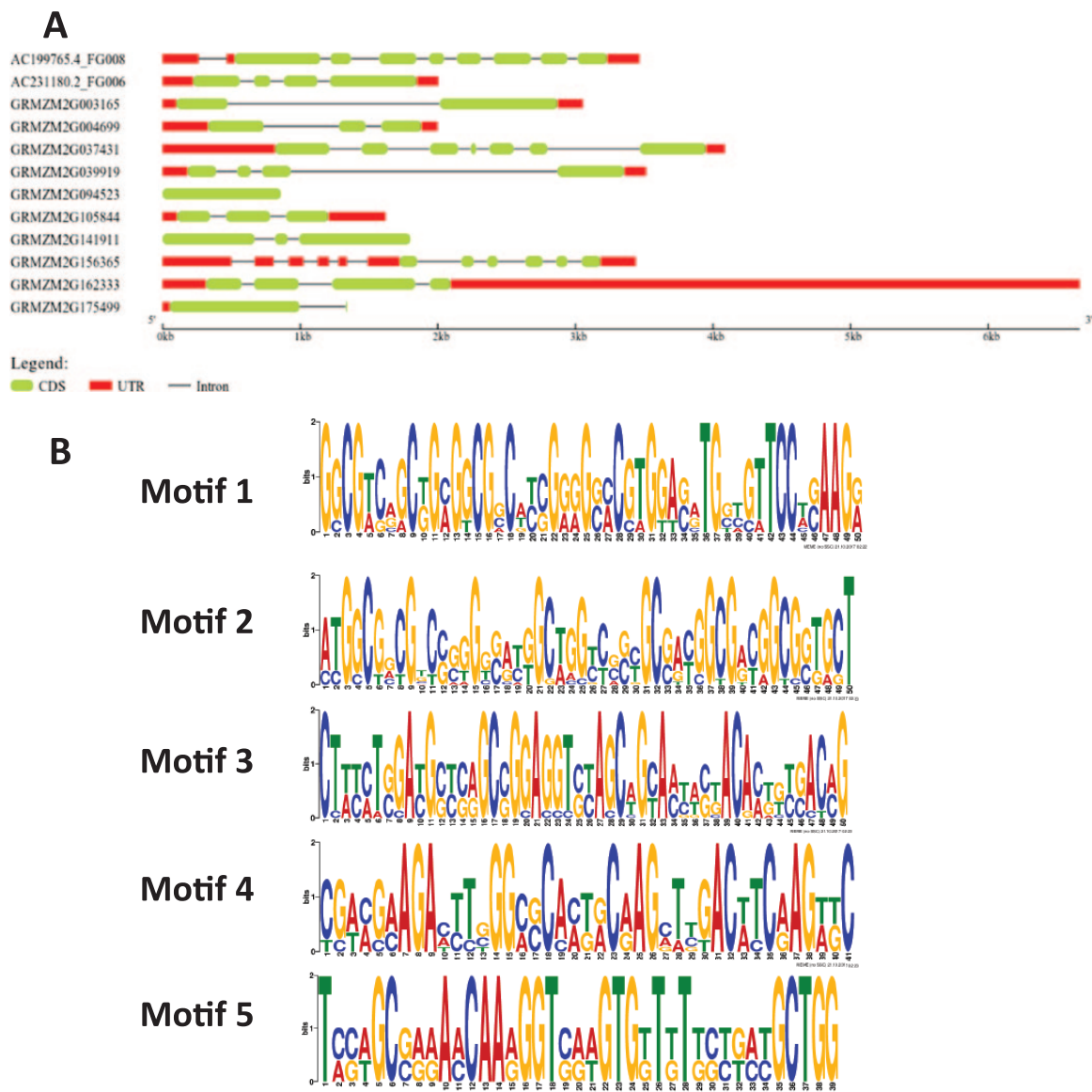
### *Gene structure analysis and identification of sequence motifs*

Maize CWM genes showed the presence of exons and introns in the genomic sequence. *PAE* (GRMZM2G156365) had the largest set of introns (12) and exons (13) among all CWM genes. This gene consists of an upstream sequence which preceded the first exon. *Pectin esterases* (GRMZM2G175499) and *expansins* (GRMZM2G094523) had the least number of intron (1) and exon (1) in their genomic sequence. However, the exon (1.3 kb) of *arabinogalactan* (GRMZM2G003165) and intron (1.8 kb) of *XTH* (GRMZM2G039919) were longest in comparison with the other CWM genes. *XTH* (GRMZM2G004699), *expansins* (GRMZM2G105844), and *endoglucanases* (GRMZM2G141911) had noted 2 introns and 3 exons in their gene structures (Figure 2A).

Five sequence motifs were observed in 12 CWM genes (Figure 2B). FIMO tool in MEME suite was used to analyze the occurrences of motifs in a set of genes. Motif 3 had the highest number of occurrences (3188) across all CWM genes. This motif was associated with gene ontology (GO) annotations of molecular functions of monooxygenase activity, RNA binding, nucleotide binding, peptidyl prolyl *cis-trans* isomerase activity, and structural constituent of ribosome; biological processes of translation and nitrogen compound metabolic process; and cellular component of mitochondrion, chloroplast, and chloroplast thylakoid membrane. Motif 1 had occurred 56 times, whereas motifs 4 and 5 occurred 29 times in CWM genes. Gene ontology terms revealed the association of motif 1 with cellular component of endomembrane system and motif 4 with transcription factor activity. Motif 2 had the least number of occurrences (27), and no significant GO annotation was found to be associated with it.

### *Identification of CREs*

Many CREs were identified in CWM genes. These CREs were responsive to light, abscisic acid, meristem-specific activation, methyl jasmonate, gibberellins, low temperature, seed-specific



**Figure 2.** Gene structure and sequence motifs of cell wall modification genes. (A) Number and length of introns and exons varied in each gene. Gene models in the figure are explained in the following order from top to bottom: GRMZM2G003165 (arabinogalactan), GRMZM2G004699 (xyloglucan transglycosylase or hydrolase), GRMZM2G037431 (polygalacturonases), GRMZM2G105844 (expansins), GRMZM2G141911 (endoglucanases), GRMZM2G156365 (pectin acetylsterases), and GRMZM2G162333 (pectin esterases). (B) Best 5 sequence motifs were found in waterlogging-responsive cell wall modification genes. These motifs had different occurrences in each gene.

regulation, core promoter elements, root-specific expression, circadian control, and anoxia. Four common CREs were identified in CWM genes: GC motif, anaerobic response elements (ARE), unnamed-1 and unnamed-4. GC motif and ARE were involved in anaerobic induction, where the former was represented in *arabinogalactan*, *XTH*, *polygalacturonases*, *expansins*, *endoglucanases*, and *pectin esterases* and the latter was represented in *XTH*, *expansins*, *PAE*, and *pectin esterases* (Supplementary Table S3).

#### Phylogenetic analysis

Phylogenetic trees were generated for 7 CWM gene families, namely, *arabinogalactan*, *XTH*, *polygalacturonases*, *expansins*, *endoglucanases*, *PAE*, and *pectin esterases*, to understand the

evolutionary pattern (Supplementary Figures S1 and S2). Phylogenetic tree was constructed for each of the CWM gene family. The analysis revealed a mixed grouping pattern where no species-specific genes were grouped together. *PAE* phylogenetic tree revealed a cluster with maximum representation of rice genes in clade VIII and mixed clades of maize, sorghum, barley, and *Arabidopsis PAE* genes. Among the 7 CWM gene families, phylogenetic trees of *expansins*, *pectin esterases*, and *polygalacturonases* showed appreciable number of polytomies. The phylogenetic tree of *expansins* included a large set of genes grouped into several clusters where the evolutionary distance between maize, sorghum, and barley genes was closer. *Pectin esterases* and *polygalacturonases* were also among the largest gene families grouped into several clusters.

### Protein structure prediction of characterized CWM proteins

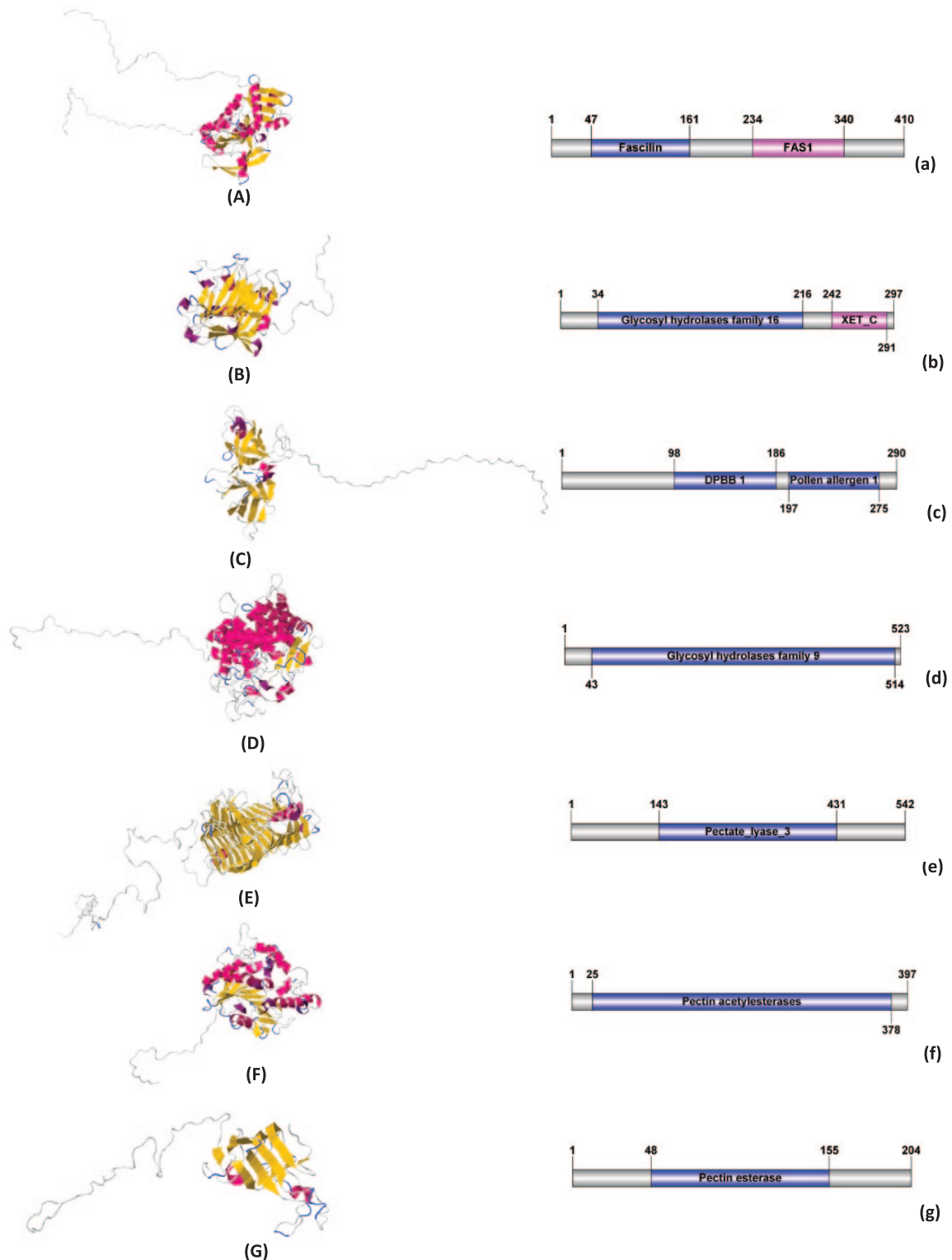
Protein sequences were used as an input to the IntFOLD server, and best 5 structure models were selected. The top ranked model was selected that had the maximum target coverage, root mean square deviation (RMSD), and template modeling score (TM-score). XTH protein (GRMZM2G004699) was among the best structure predicted with the highest confidence (Figure 3B). It was modeled with a single template of *Populus* XTH protein. The structure had a maximum target coverage of 89%, RMSD of 0.25, and highest TM-score of 0.89.  $\beta$ -D-glucose and calcium ions were predicted as the most likely ligands at the binding site. Glutamic acid, glutamine, tryptophan, arginine, and glycine were among the ligand-binding residues for XTH protein. The presence of disorders in the predicted protein structure of XTH was less. Predicted protein structures of endoglucanases (GRMZM2G141911) and expansins (GRMZM2G105844) were based on a single template. Endoglucanases protein structure had a high confidence with a  $P$ -value of  $1.29 \times 10^{-4}$  and a model quality score of 0.85 (Figure 3D). Target coverage of the structure model was computed to 84%, RMSD to 0.39, and high TM-score to 0.84. Expansins (GRMZM2G105844) also had high confidence with a  $P$ -value of  $5.8 \times 10^{-4}$  and a model quality score of 0.70 (Figure 3C). The protein structure had a target coverage of 78%, RMSD of 0.82, and TM-score of 0.76. Protein structures of polygalacturonases (GRMZM2G037431), PAE (GRMZM2G156365), and pectin esterases (GRMZM2G175499) were modeled using multiple templates (Figure 3E to G). Each template had different target coverage, RMSD, and TM-score. High confidence and low  $P$ -values were also noted for 3 structures: pectin esterases (GRMZM2G175499) had a  $P$ -value of  $8.63 \times 10^{-4}$ ; PAE (GRMZM2G156365),  $4 \times 10^{-4}$ , and polygalacturonases (GRMZM2G037431),  $2.9 \times 10^{-3}$ . Least confidence and higher  $P$ -value of  $2 \times 10^{-2}$  were observed for the arabinogalactan protein structure (GRMZM2G003165). Besides being a single template model, the arabinogalactan protein structure had a target coverage of 67%.

Four membrane-spanning proteins were identified in CWM genes: XTH, arabinogalactan, and polygalacturonases as single-spanning proteins and endoglucanases as a double-spanning protein. XTH protein (GRMZM2G004699) was modeled with 90% sequence coverage at 100% confidence. Arginine of XTH protein was predicted as the ligand-binding site residue with highest contact numbers. Active site analysis of XTH revealed a mutation-sensitive residue (tryptophan) and a rotamer (lysine) in its binding pocket region. Glutamate and aspartate were predicted as the catalytic residues in the XTH protein. All the proteins were modeled with 100% confidence except arabinogalactan that had >90% confidence. Polygalacturonases (GRMZM2G037431) protein was modeled with 65% sequence coverage with a conserved residue leucine and 2 mutation-sensitive residues proline and

serine in its binding pocket. Aspartate was predicted as a catalytic residue in polygalacturonases protein. Endoglucanases protein (GRMZM2G141911) was modeled using 84% residues. Binding pocket residues of endoglucanases protein had high alignment confidence to the template. Aspartate and glutamate were predicted as the catalytic residues in the endoglucanases protein. Arabinogalactan protein (GRMZM2G003165) was modeled using 67% confidence with mutation-sensitive residues proline and leucine in its binding pocket region. Expansins protein (GRMZM2G105844) was modeled using 77% residues with mutation-sensitive and high alignment confidence residues in its binding pocket. PAE protein (GRMZM2G156365) was modeled using 84% residues. Here, binding pocket residues were noted to be unconserved and mutation-sensitive. However, pectin esterase protein was modeled using least number of residues (63%). This protein noted less mutation-sensitive amino acid residues in its binding pocket. The stereochemical quality of the predicted protein structures was assessed using RAMPAGE.<sup>35</sup> It was noted that >90% residues of the CWM protein structures had fallen in favorable and allowed regions of Ramachandran Plot (Supplementary Figures S3 to S6), indicating the accuracy and reliability of modeled structures for further studies.<sup>35</sup>

### Protein structure prediction of uncharacterized CWM proteins

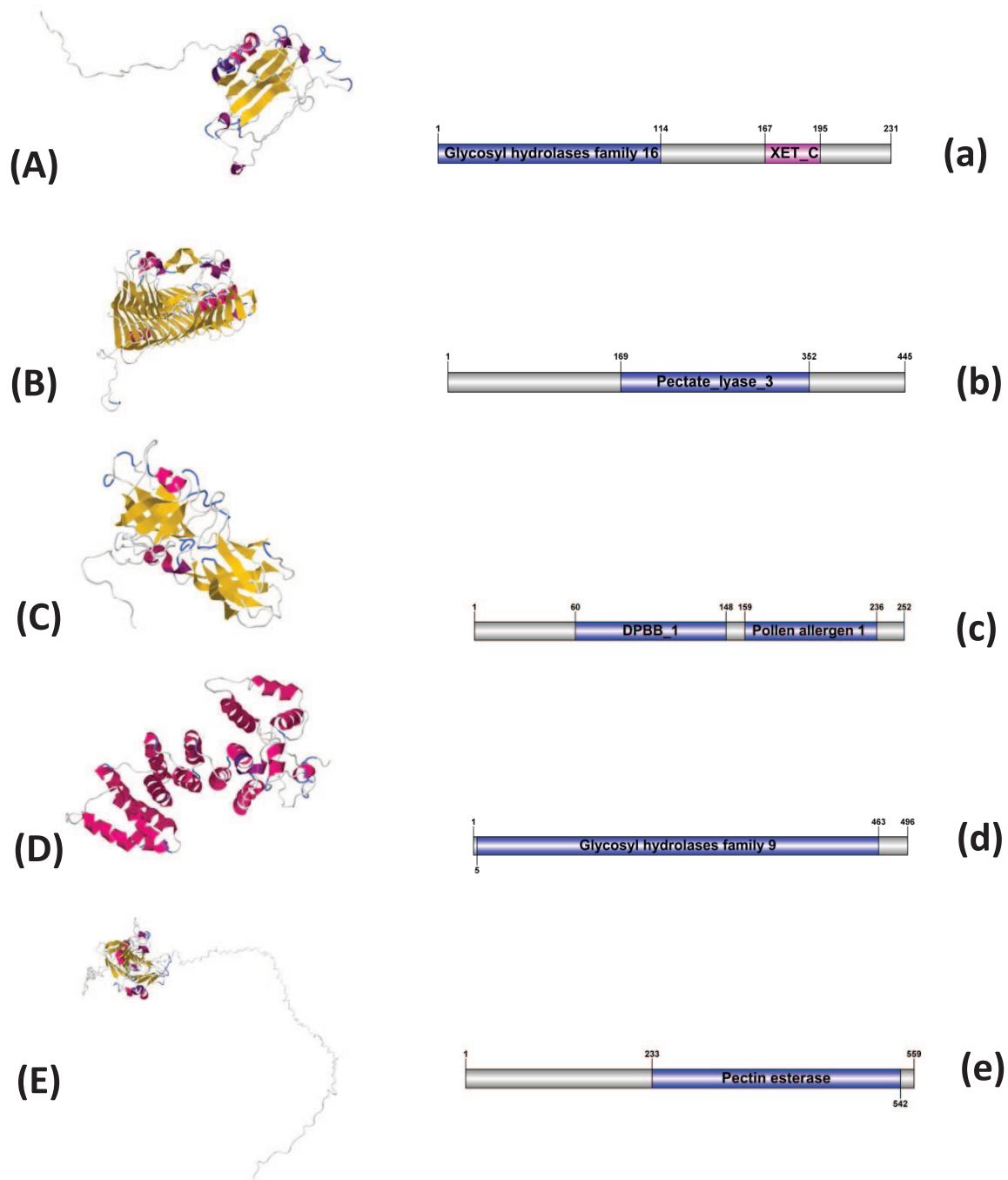
Protein structures of uncharacterized CWM proteins were predicted using IntFOLD server. The best model having maximum target coverage, RMSD, and TM-score was used for analysis. The protein structures predicted for uncharacterized CWM proteins had their target coverage above 80% and TM-score greater than 0.8 (Figure 4). The protein structure of uncharacterized maize endoglucanases (AC199765.4\_FG008) had aligned 96% of its structure to the *Arabidopsis* hydrolase protein (Figure 4D). The alignment of maize endoglucanases and *Arabidopsis* hydrolase had RMSD and TM-score of 2.3 and 0.89, respectively. Maize XTH (GRMZM2G039919) had aligned 85% of its structure to the *Populus* XTH protein (Figure 4A). The strong alignment of an uncharacterized maize XTH protein to *Populus* XTH protein had RMSD and TM-score of 1.01 and 0.83, respectively. The protein structure of maize polygalacturonases (AC231180.2\_FG006) shared RMSD of 1.2, TM-score of 0.87, and 89% coverage to the *Thermotoga* polygalacturonases protein (Figure 4B). Maize expansins protein (GRMZM2G094523) shared 90% of its structure to the maize  $\beta$ -expansins structure (Figure 4C). The expansins structural alignment had RMSD and TM-score of 0.82 and 0.88, respectively. Maize pectin esterases (GRMZM2G175499) shared 57% of its structure with pectin methylesterase protein from carrot (Figure 4E). This alignment of 2 pectin esterase proteins had RMSD and TM-score of 0.51 and 0.56, respectively. Predictions of these



**Figure 3.** Structure of characterized maize cell wall modification (CWM) proteins. (A-G) Protein structure of arabinogalactan, xyloglucan endo-transglycosylase or hydrolase, expansins, endoglucanases, polygalacturonases, pectin acetyl esterases, and pectin esterases. Blue color depicts maximum accuracy, followed by green, yellow, and orange, and red indicates least accuracy. (a-g) Domain structure of arabinogalactan, xyloglucan endo-transglycosylase or hydrolase, expansins, endoglucanases, polygalacturonases, pectin acetyl esterases, and pectin esterases.

CWM protein structures noted high confidence and low *P*-values—endoglucanases had a *P*-value of  $3.5 \times 10^{-4}$ ; XTH,  $3.6 \times 10^{-4}$ ; polygalacturonases,  $1.6 \times 10^{-4}$ ; expansins,  $2.8 \times 10^{-4}$ ; and pectin esterases,  $1.3 \times 10^{-3}$ .

The uncharacterized expansins protein was predicted for a single transmembrane helix. The protein sequence of the uncharacterized maize expansins protein (GRMZM2G094523) had 85% sequence coverage to the crystal structure of  $\beta$ -expansins



**Figure 4.** Structure of uncharacterized maize cell wall modification (CWM) proteins. (A- E) Protein structure of xyloglucan endo-transglycosylase or hydrolase, polygalacturonases, expansins, endoglucanases, and pectin esterases. Blue color depicts maximum accuracy, followed by green, yellow, and orange, and red indicates least accuracy. (a-g) Domain structure of xyloglucan endo-transglycosylase or hydrolase, polygalacturonases, expansins, endoglucanases, and pectin esterases.

from maize. Ligand-binding site residues of the uncharacterized expansin protein included threonine and asparagine. Also, binding pocket residues of the uncharacterized maize expansin protein had high alignment confidence to the template. The protein sequence of the uncharacterized maize pectin esterases (GRMZM2G175499) had 56% sequence coverage and 59% identity to the crystal structure of the complex of pectin methyl-esterase and its inhibitor protein. Catalytic residues of

pectin esterase protein included aspartic acid and glutamic acid. The uncharacterized maize endoglucanases protein (AC199765.4\_FG008) noted 98% sequence coverage and 53% identity to the crystal structure of a hydrolase protein from *Arabidopsis*. The stereochemical quality of these predicted protein structure was validated using RAMPAGE.<sup>35</sup> It was found that >90% residues of the uncharacterized maize CWM protein structures had fallen in favorable and allowed regions of



Ramachandran Plot (Supplementary Figures S7 to S9), suggesting the reliability of their protein structures for further analysis.

### Structural motif search

Each CWM protein had a set of different structural motifs. These CWM proteins shared different set of structural motifs (Supplementary Table S4). GZGS motif was shared by arabinogalactan and pectin esterases (Supplementary Table S4). This motif was represented as a functional candidate that encodes protein kinase-like superfamily. GZGS motif is found in alcohol dehydrogenase and protein kinases. FQLG and QLGI motifs were shared by XTH and expansins. These motifs were ubiquitous, namely, over-represented in several superfamilies (Supplementary Table S4). FQLG motif represents metallo-dependent hydrolases, MHC antigen-recognition domain, acyl-CoA *N*-acyltransferases (Nat), concanavalin A-like lectins/glucanases,  $\alpha$ -D-mannose-specific plant lectins, thioesterase/thiol ester dehydrase-isomerase, and trypsin-like serine protease superfamily. QLGI motif represents metallo-dependent hydrolases, MHC antigen-recognition domain, thioesterase/thiol ester dehydrase-isomerase, acyl-CoA *N*-acyltransferases (Nat), concanavalin A-like lectins/glucanases,  $\alpha$ -D-mannose-specific plant lectins, and CO dehydrogenase flavoprotein C-terminal domain-like superfamily. FQLG and QLGI motifs were found in 3-phosphoglycerate kinase, cytochrome c6, polygalacturonases, phosphoglycerate kinase,  $\beta$ -galactosidase, endo- $\beta$ -1,4-glucanase, programmed cell death protein 8, rhamnogalacturonase, XTH, acyl-coA oxidase, and other proteins. FFFI motif was shared by endoglucanases and PAE (Supplementary Table S4). This motif was found to be ubiquitous, namely, over-represented in UBC-like, *trans*-glycosidases, cystine-knot cytokines, PapD-like, MoeA N-terminal region-like, rubisco, and C-terminal domain superfamily (Supplementary Table S4). FFFI motif is found in glyceraldehydes-3-phosphate dehydrogenase, superoxide dismutase, polygalacturonases,  $\beta$ -glucosidase,  $\beta$ -galactosidase, endopolygalacturonases, cytochrome c oxidase, programmed cell death protein 8, plant arginine *N*-methyltransferase, rubisco, and glucose-6-phosphate isomerase. HBDS motif was shared by 3 CWM proteins expansins, endoglucanases, and pectin esterases (Supplementary Table S4). This motif was ubiquitous, namely, over-represented in WD40 repeat-like, ricin B-like lectins, N-terminal nucleophile aminohydrolases,  $\alpha$ -D-mannose-specific plant lectins, six-hairpin glycosidases, thioredoxin-like, cytokine, *trans*-glycosidases, FAD/NAD(P)-binding domain, concanavalin A-like lectins/glucanases, c-type lectin-like, lipocalins, trypsin-like serine proteases, calcium-dependent phosphotriesterase, polo-box domain, TATA-box binding protein-like, terpenoidcyclases/protein prenyltransferases, quinoprotein alcohol dehydrogenase-like, molybdenum cofactor-binding domain, Sm-like ribonucleoproteins, nudix superfamily, and PUA domain-like superfamily.

### Protein structural alignment of characterized CWM proteins

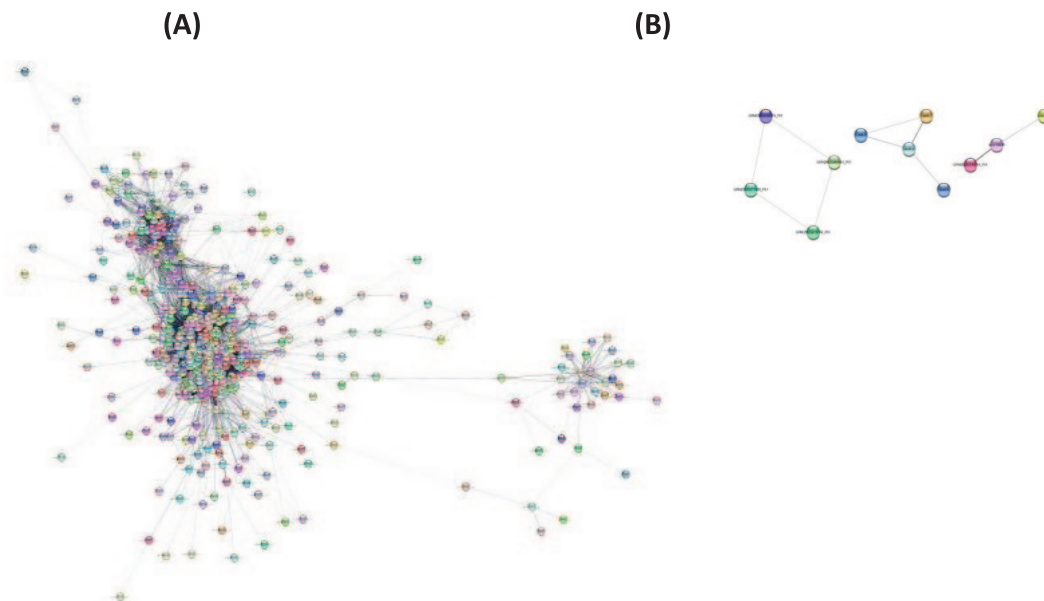
Protein structural alignment of polygalacturonases (GRMZM2G037431) and pectin esterases (GRMZM2G162333) had a highest TM-score of 0.56. ProCKSI server uses multiple tools for the assessment of protein structural alignments.<sup>31</sup> Contact numbers is a measuring parameter for the prediction of protein structures. Polygalacturonases (GRMZM2G037431) and endoglucanases (GRMZM2G141911) were observed for similar contact maps. Clustering results of ProcCKSI tools were used to assess the multiple protein structural alignments. Vorolign and URMS revealed a best clustering of arabinogalactan (GRMZM2G003165) and expansins (GRMZM2G105844). The *z*-score of combinatorial extension revealed the clusters of arabinogalactan (GRMZM2G003165), pectin esterases (GRMZM2G162333), and endoglucanases (GRMZM2G141911). The *z*-score of above threshold value of 3.5 is considered as good alignment. The same was observed for XTH-endoglucanases and endoglucanases-pectin esterases.

### Protein structural alignment of uncharacterized CWM proteins

Pairwise structural alignment of characterized and uncharacterized CWM proteins revealed the characterization of proteins as CWM proteins. TM-align of ProcCKSI tools was used to assess the pairwise structural alignment of CWM proteins. The pairwise alignment of characterized and uncharacterized XTH protein structures noted a TM-score of 0.64. This accounts for a common CATH structural domain known as probable XTH1 between both the proteins. The pairwise structural alignment of characterized and uncharacterized expansins protein structures noted a TM-score of 0.79. Both expansins proteins co-mapped CATH structural domains expansins A1 and expansins B9. Alignment of characterized and uncharacterized pectin esterases protein structures had a TM-score of 0.7. Both protein structures of pectin esterases co-mapped a CATH structural domain pectin esterases. Combinatorial extension revealed the best cluster of uncharacterized expansins (GRMZM2G094523) and polygalacturonases proteins (AC231180.2\_FG006). A high *z*-score of 5.6 was noted for this cluster in combinatorial extension.

### Similarity of predicted protein structures to PDB submitted structures

The protein structure of arabinogalactan (GRMZM2G003165) shared 86% of secondary structure with an insect cell adhesion protein fasciclin I. The similarity search revealed a novel fold of FAS1 domain pair in the arabinogalactan protein. Predicted ligands for this molecule were sulfate ion (SO<sub>4</sub>) and *N*-acetylglucosamine (NAG).



**Figure 5.** Protein-protein interaction network of cell wall modification genes and waterlogging-responsive maize genes. Protein-protein interaction network revealed 4 clusters: (A) cluster 1 and (B) clusters 2, 3, and 4. Refer Table S5 for details.

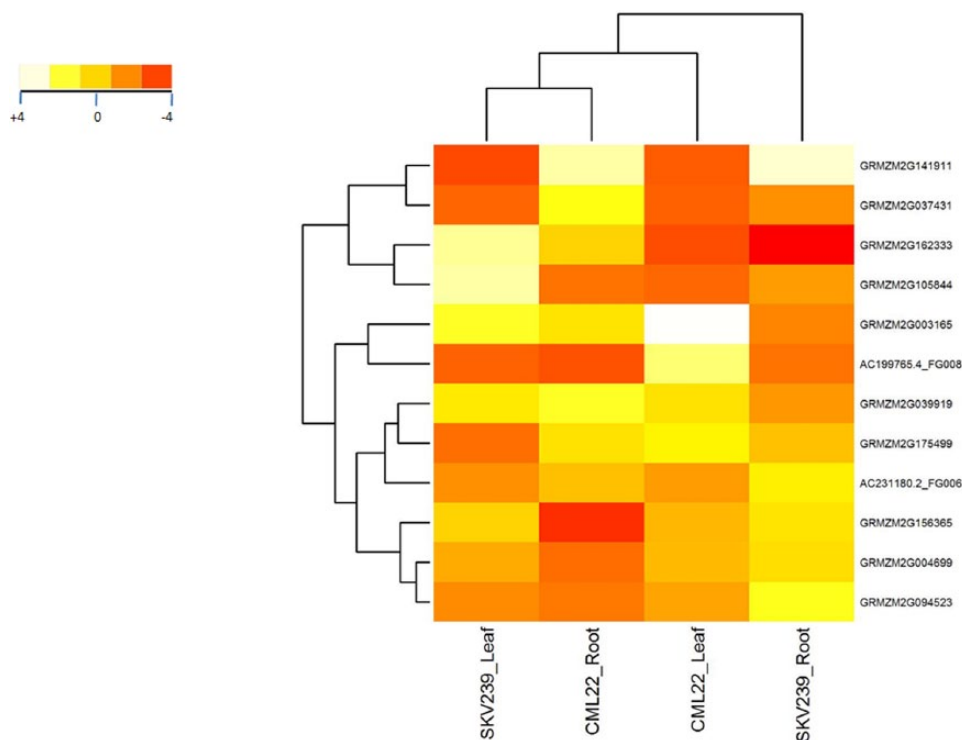
Maize XTH (GRMZM2G004699) protein structure shared 82% sequence similarity and 100% of secondary structure with *Populus* XTH protein that had transglycosylation acceptor binding sites. Ligands predicted for this PDB molecule were NAG,  $\beta$ -D-mannose,  $\beta$ -D-glucose, xylopyranose, and  $\beta$ -D-galactose. Maize polygalacturonases protein (GRMZM2G037431) shared 73% of secondary structure with an endo-xylogalacturonan hydrolase from *Aspergillus*. *N*-acetylglucosamine,  $\alpha$ -D-mannose, and sulfate ion were among the predicted ligands for the *Aspergillus* molecule. Another maize polygalacturonases protein (AC231180.2\_FG006) shared 87% of secondary structure with an exopolygalacturonase protein from *Thermotoga*. Expansins protein (GRMZM2G105844 and GRMZM2G094523) was investigated for 94% of secondary structure in  $\beta$ -expansins in maize. *N*-acetylglucosamine,  $\alpha$ -D-mannose, xylopyranose, and  $\alpha$ -D-fucose were identified as ligands for maize  $\beta$ -expansin protein. Maize endoglucanases protein (GRMZM2G141911) shared 100% secondary structure with *Perinereis* endoglucanases protein. The *Perinereis* protein had a 6-hairpin glycosidase-like sequence domain. Four ligand molecules calcium ( $\text{Ca}^{2+}$ ), sodium ( $\text{Na}^+$ ), chloride ( $\text{Cl}^-$ ), and  $\beta$ -D-glucose were predicted for the *Perinereis* endoglucanase protein. Maize PAE protein (GRMZM2G156365) shared 91% of secondary structure with *Drosophila* Wnt deacylase notum. *N*-acetylglucosamine was noted as a predicted ligand for Wnt protein. Maize pectin esterases (GRMZM2G162333) shared 85% secondary structure with *Azotobacter* mannuronan C-5 epimerase.  $\text{Ca}^{2+}$  and  $\alpha$ -D-mannopyranuronic acid were the 2 ligands predicted for the epimerase protein. Another maize pectin esterases (GRMZM2G175499) shared 84% secondary structure with a pectin methylesterase from rice weevil.

#### *CWM proteins interacting with ethylene-responsive pathway proteins*

Characterized and uncharacterized CWM proteins were searched for their interactions with waterlogging-responsive maize proteins *via* a protein-protein interaction network (Figure 5; Supplementary Table S5). The maize protein-protein interaction network had 794 nodes, 13 555 edges, and a clustering coefficient of 0.3. The interaction network was grouped into 4 clusters (Figure 5; Supplementary Table S5). The first cluster (783) was the largest and major cluster followed by second (4), third (4), and fourth (3) cluster (Figure 4). The first cluster included the interactions of CWM proteins and waterlogging-responsive genes. CWM proteins interacted with ethylene-responsive pathway proteins, E3 ubiquitin ligases RING Finger and F-box. The protein-protein interaction network revealed 4 major interactions—XTH (GRMZM2G039919 and GRMZM2G319798) interacting with E3 ubiquitin ligase RING Finger (GRMZM2G065893 and GRMZM2G077809), polygalacturonase (GRMZM2G119494) interacting with E3 ubiquitin ligase RING Finger (GRMZM2G119930), pectin acylesterase (GRMZM2G156365) interacting with E3 ubiquitin ligase RING Finger (GRMZM2G081965), and expansin (GRMZM2G445169) interacting with E3 ubiquitin ligase F-Box (GRMZM2G459166).

#### *Expression pattern of CWM genes*

Expression pattern of 12 CWM genes, namely, *arabinogalactan* (GRMZM2G003165), *XTH* (GRMZM2G004699 and GRMZM2G039919), *polygalacturonases* (GRMZM2G037431 and AC231180.2\_FG006), *expansins* (GRMZM2G105844 and GRMZM2G094523), *endoglucanases* (GRMZM2G141911



**Figure 6.** Heat map representation of differentially expressed cell wall modification genes in leaf and root tissues of tolerant (SKV239) and susceptible (CML22) maize genotypes. Log<sub>2</sub>-transformed fold change values of cell wall modification genes under waterlogging conditions. The x-axis denotes the root and leaf samples of SKV239 and CML22 genotypes. The y-axis denotes the gene models of cell wall modification genes.

and AC199765.4\_FG008), *PAE* (GRMZM2G156365), and *pectin esterases* (GRMZM2G162333 and GRMZM2G175499), were measured in a contrasting set of maize inbreds under waterlogging stress in both root and leaf tissues. Paired *t*-test of the expression values (CT values) in root and leaf samples provided a significant *P*-value of less than 0.01. Expression of CWM genes in root samples was found to be contrasting in tolerant and susceptible genotypes (Figure 6). *Endoglucanases* (AC199765.4\_FG008) was differentially expressed and highly upregulated at 4.6-fold change and 2.3-fold change in the root of tolerant and susceptible genotypes, respectively, under waterlogging conditions. *Arabinogalactan* (GRMZM2G003165), *XTH* (GRMZM2G004699), *expansins* (GRMZM2G105844), *PAE* (GRMZM2G156365), and *pectin esterases* (GRMZM2G162333) were differentially expressed at a contrasting fold change with upregulation in tolerant roots and downregulation in susceptible roots under waterlogging conditions.

Similarly, under waterlogging stress, all the 12 CWM genes showed differential expression in leaf. Except *arabinogalactan* (GRMZM2G003165) (fold change: 2.85), all the genes were highly downregulated in the leaf of susceptible genotype. *Pectin esterases* (GRMZM2G162333) was highly downregulated in leaf at 1.2-fold change and 73-fold change in tolerant and susceptible genotype, respectively.

## Discussion

We had analyzed CWM proteins on the basis of their nucleic acid and amino acid sequence composition. The gene structure

of *PAE* noted the maximum number of introns and exons that explained the basis of alternative splicing events where each transcript had its own expression level. *PAE* encoding gene model GRMZM2G156365 has 6 transcripts where the third transcript GRMZM2G156365\_T03 was upregulated in the roots of waterlogging-tolerant genotype (Figure 6). Phylogenetic analysis of CWM genes revealed the evolutionary distances among CWM genes of maize, *Arabidopsis*, sorghum, barley, and rice. Furthermore, mixed grouping pattern of CWM genes suggested the evolution of CWM genes in lineages of these 5 species. The presence of polytomies in the phylogenetic trees of *expansins*, *pectin esterases*, and *polygalacturonases* suggested that these genes could have evolved rapidly under multiple speciation events from common ancestral genes ([http://evolution.berkeley.edu/evolibrary/article/starbursts\\_01](http://evolution.berkeley.edu/evolibrary/article/starbursts_01)).

*Cis*-acting regulatory elements are present in the promoter regions that are associated with stress regulation and plant development.<sup>36</sup> Du et al<sup>37</sup> reported a waterlogging-induced promoter, including multiple CREs responsive to waterlogging conditions—GC motif, ARE, G-box, and GT motif. In this study, these CREs were identified in waterlogging-responsive CWM genes (Supplementary Table S3). In addition, CREs are the non-coding regions of DNA which regulate transcription of neighboring genes.<sup>38</sup> It could be possible that these CREs also regulated the expression of waterlogging-responsive CWM genes in tolerant and susceptible genotypes.

Amino acid composition of CWM proteins revealed transmembrane helices and highly mutated and conserved residues in their active sites. Transmembrane helices found in arabinogalactan, XTH, and polygalacturonases allow them to act as transporters in the cell that maintain the osmotic balance and stability of the cell. Amino acid composition of CWM proteins revealed highly mutation-sensitive residues in their binding pockets, suggesting a likely change in phenotypic or functional effect. These mutations swap the substrate specificity and ligand binding in their active sites.<sup>39,40</sup> However, less mutational sensitive residues in the binding pocket region signified strong substrate specificity to cell wall polysaccharides. Conserved residues are essential for a proper structure and function of the protein. But the conserved residues in active sites are most important for a reaction. The same pattern of conserved residues in the binding pocket was found in polygalacturonases in this study.

Structural motif search in CWM proteins identified possible associations with hypoxia-responsive genes, including fermentation, glycolysis, PCD, and ROS signaling genes. This indicated the specificity of structural motifs to stress-responsive genes and connectivity between their secondary structural elements (SSEs). Secondary structures of CWM proteins were aligned to each other for determining structural similarity between them. High alignment score of polygalacturonases and pectin esterases explained the possibility of same structural domains among them. Same CATH and SCOP domain superfamily of pectin lyase-like single-stranded right-handed  $\beta$ -helix was identified between these 2 proteins.

The maize protein-protein interaction network (Figure 5) revealed the interactions of characterized and uncharacterized CWM proteins and ethylene-responsive pathway proteins, E3 ubiquitin ligases RING finger and F-box.<sup>33</sup> Evidently, the characterized and uncharacterized CWM proteins encoding genes were validated in waterlogging-tolerant and waterlogging-susceptible genotypes of maize (Figure 6), and E3 ubiquitin ligases were expressed in contrasting genotypes in genome-wide expression data<sup>22</sup> and waterlogging-tolerant genotype in high-throughput sequencing data.<sup>33</sup> The interacting E3 ubiquitin ligases are those proteins that catalyze the transfer of ubiquitin to the substrate protein where ubiquitin-protein complexes are the degradable proteins. However, these degradable proteins are studied as hypoxia-responsive proteins in *Arabidopsis*.<sup>41</sup> These hypoxic proteins can probably be the maize CWM proteins interacting with E3 ubiquitin ligases in the protein-protein interaction network (Figure 5). This suggests a novel fact that the CWM proteins-E3 ubiquitin ligases are responsive to ethylene under waterlogging conditions in maize.

A maize co-expression network generated in a waterlogging experiment<sup>22</sup> revealed important co-expression of CWM genes and waterlogging-responsive genes. XTH and polygalacturonases had co-expressed with genes regulating glycolytic pathway, energy metabolism, PCD, and ROS scavenging in a single

cluster. First, the co-expression of CWM genes and glycolytic pathway ( $\beta$ -glucosidase and glyceraldehydes-3-phosphate dehydrogenase) suggested that the modification of cell wall is a highly energy-consuming process. Second, the CWM genes had also coexpressed with a PCD gene (plant aspartic protease 3 [PASPA3]). In support of the finding, PASPA3 is also studied to be highly upregulated in waterlogging-tolerant genotype of maize.<sup>33</sup> The process of modification of cell wall is followed by PCD and aerenchyma formation under waterlogging conditions.<sup>1,4,21</sup> Reactive oxygen species (ROS) scavenging is another important process of waterlogging tolerance pathways. In the co-expression network, these ROS scavenging genes (glutathione-S-transferase and peroxidase) had also coexpressed with maize CWM genes. The CWM, PCD, and ROS scavenging genes co-expressed in the network tend to be waterlogging-responsive in maize.<sup>22,33</sup> These waterlogging-responsive pathways are interconnecting pathways in maize. In addition, the co-expressed CWM genes had expressed in waterlogging-tolerant and waterlogging-susceptible genotypes of maize (Figure 6), and other clustered waterlogging-responsive genes had also expressed in waterlogging-tolerant and waterlogging-susceptible maize genotypes investigated in a genome-wide expression data.<sup>22</sup>

Transcript expression of the CWM genes *arabinogalactan*, *XTH*, *polygalacturonases*, *expansins*, *endoglucanases*, *PAE*, and *pectin esterases* was investigated in roots and leaf tissues of tolerant and susceptible genotypes of maize under waterlogging stress. Differential expression of *arabinogalactan*, *XTH*, *expansins*, *PAE*, and *pectin esterases* was contrasting in the tolerant and susceptible genotypes. These genes were highly upregulated in the roots of tolerant genotype (Figure 6). It was studied that CWM genes were expressed at a high fold change in maize aerenchyma tissue under waterlogging conditions.<sup>21</sup> This suggests the regulation of CWM genes in root cortical cells during aerenchyma formation under waterlogging conditions.

## Conclusions

Degradation and expansion of cell wall allow aerenchyma formation which increases oxygen transfer from roots to shoots under waterlogging conditions. Phylogenetic analysis revealed CWM genes evolved in the lineage of maize evolution. The identified common motifs in CWM genes could be the key players in imparting waterlogging tolerance in maize. Furthermore, contrasting expression pattern of CWM genes in waterlogging-tolerant and waterlogging-susceptible genotypes suggested their possible role as candidate markers in maize breeding for waterlogging tolerance. The identified regulatory and structural elements allow proper folding of the proteins that can be used to modify the proteomic expression to enhance waterlogging tolerance in maize. The identified candidate genes could be used in introgression breeding program to develop waterlogging-tolerant maize hybrids.

## Author Contributions

KA, NT, and KKP conceived and designed the experiments; KA, MGM, and SM analyzed the data; and KA, MGM, and TN drafted the manuscript. All authors read and approved the final manuscript.

## REFERENCES

- Gunawardena AH, Greenwood JS, Dengler NG. Cell wall degradation and modification during programmed cell death in lace plant, *Aponogeton madagascariensis* (Aponogetonaceae). *Am J Bot.* 2007;94:1116–1128.
- Fry SC, Dumville JC, Miller JG. Fingerprinting of polysaccharides attacked by hydroxyl radicals in vitro and in the cell walls of ripening pear fruit. *Biochem J.* 2001;357:729–737.
- Colmer TD. Long-distance transport of gases in plants: a perspective on internal aeration and radial oxygen loss from roots. *Plant Cell Environ.* 2003;26:17–36.
- Schussler EE, Longstreth DJ. Changes in cell structure during the formation of root aerenchyma in *Sagittaria lancifolia* (Alismataceae). *Am J Bot.* 2000;87:12–19.
- Imoto K, Yokoyama R, Nishitani K. Comprehensive approach to genes involved in cell wall modifications in *Arabidopsis thaliana*. *Plant Mol Biol.* 2005;58:177–192.
- Wakabayashi K, Hoson T, Huber DJ. Methyl de-esterification as a major factor regulating the extent of pectin depolymerization during fruit ripening: a comparison of the action of avocado (*Persea americana*) and tomato (*Lycopersicon esculentum*) polygalacturonases. *J Plant Physiol.* 2003;160:667–673.
- Hadfield KA, Bennett AB. Polygalacturonases: many genes in search of a function. *Plant Physiol.* 1998;117:337–343.
- Gou J-Y, Miller LM, Hou G, Yu X-H, Chen X-Y, Liu C-J. Acetyltransferase-mediated deacetylation of pectin impairs cell elongation, pollen germination, and plant reproduction. *Plant Cell.* 2012;24:50–65.
- Keegstra K. Plant cell walls. *Plant Physiol.* 2010;154:483–486.
- Darley CP, Forrester AM, McQueen-Mason SJ. The molecular basis of plant cell wall extension. *Plant Mol Biol.* 2001;47:179–195.
- McQueen-Mason S, Durachko DM, Cosgrove DJ. Two endogenous proteins that induce cell wall extension in plants. *Plant Cell.* 1992;4:1425–1433.
- Colmer TD, Peeters AJM, Wagemaker CAM, Vriezen WH, Ammerlaan A, Voeseek LA. Expression of alpha-expansin genes during root acclimations to O<sub>2</sub> deficiency in *Rumex palustris*. *Plant Mol Biol.* 2004;56:423–437.
- Palanivelu P. Polygalacturonases: active site analyses and mechanism of action. *Indian J Biotechnol.* 2006;5:148–162.
- Han Y, Zhu Q, Zhang Z, et al. Analysis of xyloglucan endotransglycosylase/hydrolase (XTH) genes and diverse roles of isoenzymes during persimmon fruit development and postharvest softening. *PLoS ONE.* 2015;10:1–20.
- Brummell D, Catala C, Lashbrook CC, Bennett B. A membrane-anchored E-type endo-1,4-β-glucanase is localized on Golgi and plasma membranes of higher plants. *Proc Natl Acad Sci U S A.* 1997;94:4794–4799.
- Mølhøj M, Pagant S, Höfte H. Towards understanding the role of membrane-bound endo-1,4-glucanases in cellulose biosynthesis. *Plant Cell Physiol.* 2002;43:1399–1406.
- Lee Y, Kende H. Expression of alpha-expansin and expansin-like genes in deep-water rice. *Plant Physiol.* 2002;130:1396–1405.
- Tan L, Leykam JF, Kieliszewski MJ. Glycosylation motifs that direct arabinogalactan addition to arabinogalactan-proteins. *Plant Physiol.* 2003;132:1362–1369.
- Johansson P. Crystal structures of a poplar xyloglucan endotransglycosylase reveal details of transglycosylation acceptor binding. *Plant Cell.* 2004;16:874–886.
- Yennawar NH, Li L-C, Dudzinski DM, Tabuchi A, Cosgrove DJ. Crystal structure and activities of EXPB1 (Zea m 1), a beta-expansin and group-1 pollen allergen from maize. *Proc Natl Acad Sci U S A.* 2006;103:14664–14671.
- Rajhi I, Yamauchi T, Takahashi H, et al. Identification of genes expressed in maize root cortical cells during lysigenous aerenchyma formation using laser microdissection and microarray analyses. *New Phytol.* 2011;190:351–368.
- Thirunavukkarasu N, Hossain F, Mohan S, et al. Genome-wide expression of transcriptomes and their co-expression pattern in subtropical maize (*Zea mays* L.) under waterlogging stress. *PLoS ONE.* 2013;8:e70433.
- Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics.* 2015;31:1296–1297.
- Bailey TL, Boden M, Buske FA, et al. MEME suite: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37:W202–W208.
- Lescot M, Déhais P, Thijs G, et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 2002;30:325–327.
- Schliep KP. phangorn: phylogenetic analysis in R. *Bioinformatics.* 2011;27:592–593.
- Paradis E, Claude J, Strimmer K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics.* 2004;20:289–290.
- McGuffin LJ, Atkins JD, Salehe BR, Shuid AN, Roche DB. IntFOLD: an integrated server for modelling protein structures and functions from amino acid sequences. *Nucleic Acids Res.* 2015;43:W169–W173.
- Regad L, Saladin A, Maupetit J, Geneix C, Camproux AC. SA-Mot: a web server for the identification of motifs of interest extracted from protein loops. *Nucleic Acids Res.* 2011;39:203–209.
- Kelly LA, Mezulis S, Yates C, Wass M, Sternberg M. The Phyre2 web portal for protein modelling, prediction, and analysis. *Nat Protoc.* 2015;10:845–858.
- Barthel D, Hirst JD, Błazewicz J, Burke EK, Krasnogor N. ProCKSI: a decision support system for protein (structure) comparison, knowledge, similarity and information. *BMC Bioinformatics.* 2007;8:416.
- Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr.* 2004;60:2256–2268.
- Arora K, Panda KK, Mittal S, et al. RNAseq revealed the important gene pathways controlling adaptive mechanisms under waterlogged stress in maize. *Sci Rep.* 2017;7:10950.
- Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* 2017;45:D362–D368.
- Lovell SC, Davis IW, Arendall WB, et al. Structure validation by Cα geometry: φ, ψ and Cβ deviation. *Proteins.* 2003;50:437–450.
- Passricha N, Saifi S, Ansari MW, Tuteja N. Prediction and validation of cis-regulatory elements in 5' upstream regulatory regions of lectin receptor-like kinase gene family in rice. *Protoplasma.* 2017;254:669–684.
- Du H, Zhang Z, Li J. Isolation and functional characterization of a waterlogging-induced promoter from maize. *Plant Cell Rep.* 2010;29:1269–1275.
- Bilas R, Szafrań K, Hnatuszko-Konka K, Kononowicz AK. Cis-regulatory elements used to control gene expression in plants. *Plant Cell Tiss Org.* 2016;127:269–287.
- Kádas J, Weber IT, Bagossi P, et al. Narrow substrate specificity and sensitivity toward ligand-binding site mutations of human T-cell Leukemia virus type 1 protease. *J Biol Chem.* 2004;279:27148–27157.
- Stuible HP, Büttner D, Ehrling J, Hahlbrock K, Kombrink E. Mutational analysis of 4-coumarate:CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes. *FEBS Lett.* 2000;467:117–122.
- Gibbs DJ, Lee SC, Isa NM, et al. Homeostatic response to hypoxia is regulated by the N-end rule pathway in plants. *Nature.* 2011;479:415–418.