



Gene Expression Patterns for Proteins With Lectin Domains in Flax Stem Tissues Are Related to Deposition of Distinct Cell Wall Types

Natalia Petrova¹, Alsu Nazipova², Oleg Gorshkov², Natalia Mokshina¹, Olga Patova³ and Tatyana Gorshkova^{2*}

¹ Laboratory of Plant Glycobiology, Kazan Institute of Biochemistry and Biophysics, FRC Kazan Scientific Center of RAS, Kazan, Russia, ² Laboratory of Plant Cell Growth Mechanisms, Kazan Institute of Biochemistry and Biophysics, FRC Kazan Scientific Center of RAS, Kazan, Russia, ³ Institute of Physiology, FRC Komi Science Centre of Ural Branch of Russian Academy of Sciences, Syktyvkar, Russia

OPEN ACCESS

Edited by:

Els J. M. Van Damme,
Ghent University, Belgium

Reviewed by:

Hervé Canut,
UMR 5546 Laboratoire de Recherche
en Sciences Vegetales (LRSV), France
Yokoyama Ryusuke,
Tohoku University, Japan

*Correspondence:

Tatyana Gorshkova
gorshkova@kibb.knc.ru

Specialty section:

This article was submitted to
Plant Physiology,
a section of the journal
Frontiers in Plant Science

Received: 28 November 2020

Accepted: 16 March 2021

Published: 26 April 2021

Citation:

Petrova N, Nazipova A,
Gorshkov O, Mokshina N, Patova O
and Gorshkova T (2021) Gene
Expression Patterns for Proteins With
Lectin Domains in Flax Stem Tissues
Are Related to Deposition of Distinct
Cell Wall Types.
Front. Plant Sci. 12:634594.
doi: 10.3389/fpls.2021.634594

The genomes of higher plants encode a variety of proteins with lectin domains that are able to specifically recognize certain carbohydrates. Plants are enriched in a variety of potentially complementary glycans, many of which are located in the cell wall. We performed a genome-wide search for flax proteins with lectin domains and compared the expression of the encoding genes in different stem tissues that have distinct cell wall types with different sets of major polysaccharides. Over 400 genes encoding proteins with lectin domains that belong to different families were revealed in the flax genome; three quarters of these genes were expressed in stem tissues. Hierarchical clustering of the data for all expressed lectins grouped the analyzed samples according to their characteristic cell wall type. Most lectins differentially expressed in tissues with primary, secondary, and tertiary cell walls were predicted to localize at the plasma membrane or cell wall. These lectins were from different families and had various architectural types. Three out of four flax genes for proteins with jacalin-like domains were highly upregulated in bast fibers at the stage of tertiary cell wall deposition. The dynamic changes in transcript level of many genes for lectins from various families were detected in stem tissue over the course of gravitropic response induced by plant gravistimulation. The data obtained in this study indicate a large number of lectin-mediated events in plants and provide insight into the proteins that take part in tissue specialization and reaction to abiotic stress.

Keywords: plant lectins, flax, jacalin, malectin, amaranthin, gene expression, cell wall, gravitropic reaction

INTRODUCTION

The term lectins joins a number of protein families capable of selectively recognizing various types of carbohydrates, while not interacting enzymatically with the recognized targets (Peumans and Van Damme, 1995). Lectins are widespread and present in all biological kingdoms (Sharon, 2008; Tsaneva and Van Damme, 2020). Initially, lectins were identified by the agglutination reaction; therefore, this group of proteins was often referred to as agglutinins. With the development of

proteomics and genomics methods, the concept of lectins has moved to a fundamentally different qualitative level. It has been shown that not all lectins found during sequencing of plant genomes exhibit the property of agglutination (Van Damme et al., 2008).

The genomes of higher plants encode a variety of lectin proteins, which are classified into families according to their conserved carbohydrate-recognition domains (Van Damme et al., 2008; Jiang et al., 2010). Sequence bioinformatics studies across all available databases have shown that some of the lectin motifs are widespread (from plants to animals, fungi, and bacteria), while others are present only in certain plant families. Several informative papers with generalizations on plant lectin families have been published (Bellande et al., 2017; Van Holle et al., 2017; Tsaneva and Van Damme, 2020).

The absence of enzymatic interaction poses a number of unresolved questions about the functional role of lectins. As revealed by an inventory of plant lectins in several species with fully sequenced genomes, most plant proteins with lectin domains are multi-domain proteins in which one or more lectin domains are linked to other protein domains such as a protein kinase domain, an F-box domain, or a glycosyl hydrolase domain (Van Holle and Van Damme, 2019). The description of a plant protein as a lectin is often based only on the identification of a lectin motif in the encoding sequence; however, there is a lack of evidence on the ability to actually bind a carbohydrate, as well as for the characterization of the exact ligand specificity. Lectin specificity is often named due to a single monosaccharide, like mannose-binding or galactose-binding; it is clear though that lectins recognize more extended carbohydrate chains (Barre et al., 2019).

Proteins with lectin domains may have different types of architectures (Bellande et al., 2017). Lectin domains can be appended to intracellular kinase and trans-membrane domains to constitute receptor-like kinases; architecture of such proteins is designated as LecRLK type. Proteins with similar architecture but without kinase domains are designated LecRLP. Finally, the soluble proteins with neither kinase nor transmembrane domains are designated LecP. Plants are highly enriched in proteins with LecRLK architecture (Bellande et al., 2017). Proteins with lectin domains can be located in different cell compartments, including the vacuole, cytoplasm, nucleus, plasma membrane, and cell wall (Lannoo and Van Damme, 2010). Thus, the functions of different lectins may be diverse.

Some lectins accumulate in vacuoles as the defensive compounds against herbivorous animals and are quite abundant in seeds and vegetative storage tissues (Van Damme et al., 1998). However, in non-storage tissues, such as leaves, stems, roots, and flowers, the concentrations of lectins are low. Lectins were suggested to be involved in signaling reactions in plant cells or between plants and other organisms (Lannoo and Van Damme, 2010). The participation of lectins in a plant's response to pathogens is supported by the numerous descriptions of the induced expression of lectins upon pathogen attack (Yan et al., 2005; Wakefield et al., 2006; Hwang and Hwang, 2011; Vandenborre et al., 2011). These inducible lectins are typically localized to the nucleus and/or cytoplasm of plant cells. Based on

these observations, it was suggested that lectin-mediated protein-carbohydrate interactions in the cytoplasm and nucleus play an important role in the physiology of plant cell stress (Van Damme et al., 2004, 2008). In addition, lectins may also be involved in normal plant development, though the information on this topic is scarce. Several proteins with lectin domains were demonstrated to be involved in the regulation of plant growth and morphogenesis (Lee et al., 2003; Jiang et al., 2007; Duan et al., 2010). Research on plant lectins has mainly focused on the proteins of LecRLK type (Boisson-Dernier et al., 2011; Vaid et al., 2012; Bellande et al., 2017; He et al., 2018; Van Holle and Van Damme, 2018), since the presence of kinase domain allows for the direct coupling of these proteins with signaling pathways.

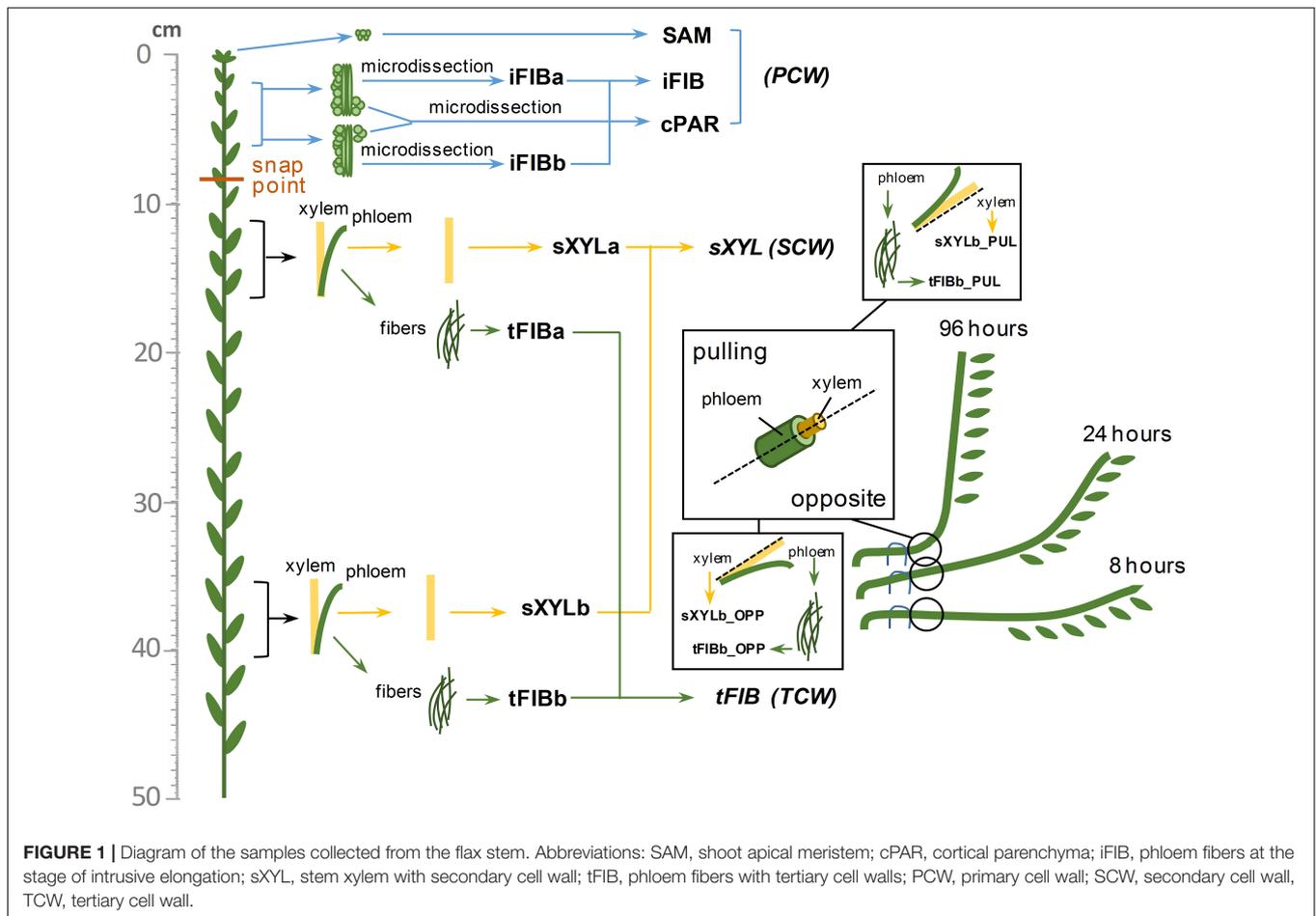
The bulk of plant polymeric carbohydrates are deposited in cell walls. The complex carbohydrate diversity in plant cell walls actively changes over the course of plant development, suggesting the presence of mechanisms capable of selectively recognizing the structural parameters of polysaccharides. Plants develop several types of cell walls that are based on different non-cellulosic polysaccharides. In dicots, primary cell walls are characterized by high proportions of xyloglucan and pectins (Carpita and Gibeaut, 1993), secondary cell walls are enriched with xylans (Zhong et al., 2019), and tertiary cell walls contain rhamnogalacturonans I with specific structure (Gorshkova et al., 2018a). The profiles of cell wall glycoproteins may also differ between the different types of cell walls, as suggested from proteomics (Nilsson et al., 2010; Bygdell et al., 2017) and transcriptomics studies (Gorshkov et al., 2018). Lectins that are secreted into the apoplast or are anchored to the plasma membrane have direct access to cell wall polysaccharides. If there are lectins specialized for the carbohydrate motifs of certain cell wall polymers, their expression could differ at deposition of various cell wall types.

To determine that we have used the developing flax (*Linum usitatissimum*) stem as an established model system that has tissues with three basic types of cell wall. Since proteomics is quite complicated for cell wall research and yields only a limited proportion of the expected proteins (Albenne et al., 2014), we conducted a transcriptomics analysis to compare the mRNA levels for all flax genes encoding proteins with lectin domains in various stem tissues of developing plants and during gravitropic response that exemplifies the effect of an abiotic stressor. The pronounced differential expression of lectins demonstrated the involvement of lectins in various aspects of plant physiology, and relation of some proteins with lectin domains to distinct cell wall types.

MATERIALS AND METHODS

Input Data and Processing

The 40 flax transcriptome libraries considered in this study were previously obtained by us and deposited in the Sequence Read Archive (SRA) as BioProjects (PRJNA475325, PRJNA631357). The four RNA-Seq datasets from the stem apex (PRJNA229810, Zhang and Deyholos, 2016) were downloaded from the European



Nucleotide Archive¹ as fastq-files of raw data. The listed datasets include samples from various tissues of the flax stem consisting of cells with different types of cell wall. The detailed localization of samples on the flax stem is shown in **Figure 1**. The SAM (shoot apical meristem) and cPAR (cortical parenchyma) samples consist of cells with primary cell wall (PCW), same as the iFIBa and iFIBb samples that contain phloem fibers at the stage of intrusive growth; iFIBa, iFIBb and cPAR samples were isolated by cryosectioning and laser microdissection (Gorshkova et al., 2018b). The samples of xylem stem part (sXYLa, sXYLb) were enriched in cells with secondary cell walls (SCW) and isolated phloem fibers at the stage of tertiary cell wall (TCW) deposition constituted the samples tFIBa and tFIBb (**Figure 1**). Additionally, the set of SCW and TCW samples was obtained from the region of stem curvature formed in the course of gravitropic response to return the inclined plants to vertical position (Ibragimova et al., 2017; Gorshkov et al., 2018). Each segment of gravibending stems was cut along into halves and separated into pulling (PUL) and opposite (OPP) sides both in the xylem (inner part) and in the phloem (outer part); the outer peel of each stem side was used to isolate phloem fibers. Segments of the flax stem were collected 8 h, 24h and 96h after plant inclination (**Figure 1**).

After processing of raw reads by the BBDuk utility of BBTtools v 37.02² the clean reads were mapped onto the flax genome downloaded from Phytozome v 12 (Goodstein et al., 2012; Wang et al., 2012) using HISAT2 v2.1.083 (Kim et al., 2015). Transcript abundance was determined by StringTie v2.0 (Pertea et al., 2016) and the number of all reads for each gene were calculated as total gene read (TGR) counts. DESeq2 v.1.28.1 as R package was used to normalize the counts per gene by the estimateSizeFactors function and to perform pairwise differential expression analysis (Love et al., 2015).

Identification of Genes Encoding Lectins of Different Families

The plant genes for proteins with lectin domains of different families were recognized by the name search of characteristic pfam domains (Pfam 33.1 database³; El-Gebali et al., 2019) in the Phytozome v12.1.6 database (⁴Goodstein et al., 2012). The pfam domain names that were used to identify genes are listed in **Table 1**. Members of CRA lectin family were identified by the presence of cd02879 (GH18_plant_chitinase_class_V)

²<https://jgi.doe.gov/data-and-tools/bbttools/bb-tools-user-guide/bbdduk-guide/>

³<http://pfam.xfam.org/>

⁴<https://phytozome.jgi.doe.gov/pz/portal.html>

¹<https://www.ebi.ac.uk/ena>

domain using the CD-search tool (⁵Lu et al., 2020). The protein sequences of plant genes were downloaded from the Phytozome v12.1.6 and from the Uniprot (release 2020_05) (⁶The UniProt Consortium, 2019) databases. For each protein with an unreliable amino acid sequence (truncated, or significantly different from orthologous genes, or including not typical domains that were not in other members, etc.) based on genome sequence from Phytozome, the open reading frame as well as the intron-exon structure and protein sequence were re-predicted using Augustus program⁷, some sequences were additionally checked by FGENESH (⁸Solovyev et al., 2006); if for any reason a reliable sequence for such proteins was not established (errors or gaps in the sequencing, SNPs, etc.) the sequence was used as taken from Phytozome. Revised sequences are presented in **Supplementary File 1**.

The prediction of a signal peptide was performed using number of bioinformatics resources: SignalP-5.0 (⁹Armenteros et al., 2019), PrediSi¹⁰, iPSORT (¹¹Bannai et al., 2002). The presence of transmembrane domain was performed using TMHMM v2.0 (¹²Krogh et al., 2001) (**Supplementary File 1**). Domain organization of flax members of lectin families were resolved using the InterProScan tool implemented in the InterPro database (¹³Mitchell et al., 2019) and was visualized in Adobe Illustrator CC 2017 software.

The putative subcellular localization of flax lectins was revealed as predicted for their *A. thaliana* homologs using SUBA4 (¹⁴Hooper et al., 2017) and using LocTree3 service (¹⁵Goldberg et al., 2014).

Quantification and Tissue Specificity Analysis of Gene Expression

In total, 32,870 from 43,486 genes were considered as expressed according to the cut-off TGR ≥ 16 at least in one sample (SEQC/MAQC-III Consortium, 2014) and were used for the analysis of differential expression. The regularized-logarithm transformation or rlog (Love et al., 2015) of the raw count data was used for hierarchical clustering of genes encoding proteins with lectins domain and the visualization by heatmap using the hclust function in R (R Core Team, 2014). A dendrogram and a heatmap were generated by the R function heatmap.2.

To characterize the differential expression of lectin genes depending on the type of cell wall at the input to DESeq2 we used datasets designated as PCW, SCW, TCW. In particular, dataset of PCW samples consisted of SAM, cPAR, iFIBa and iFIBb samples as biological replicates, SCW set consisted of sXYLa and sXYLb,

whereas the tFIBa and tFIBb samples were used as biological replicates for TCW dataset. To be considered as differentially expressed, genes identified by DESeq2 were required to have at least the 2-fold change and $\text{padj} < 0.01$. The data of RNA-Seq were verified by qRT-PCR analysis for the selected 10 genes (**Supplementary File 2**).

To analyze the changes in gene expression in the course of graviresponse, the rlog-dataset of samples obtained from gravistimulated stems was used as input for clustering based not on absolute expression value but rather on the amount by which each gene deviates in a specific sample from the gene's average across all samples (Love et al., 2015). Hence, in this case we centered the values of each gene in the samples and built a heatmap where color corresponds to the amount by which a gene expression variance deviates from the gene's mean variance across all samples.

Additionally, on the basis of normalized TGR values for the lectin genes we calculated the index tau as tissue specificity score of a gene (Yanai et al., 2005; Kryuchkova-Mostacci and Robinson-Rechavi, 2016) using the roonysgalbi/tispec package in R¹⁶. The total set of samples to calculate the index tau included SAM, cPAR, the combined sample of iFIB as an average of iFIBa and iFIBb; sXYLa and sXYLb were averaged with the name sXYL, whereas tFIB was represented by averaging values of expression in tFIBa and tFIBb from plants grown in normal condition (**Figure 1**).

Phylogenetic Analysis

The obtained sequences of plant lectin-domain containing proteins were subjected to multiple alignments using the web-based service ClustalW (¹⁷Madeira et al., 2019). The alignments were further exposed to a maximum likelihood phylogenetic analysis in IQTREE1.6.9 software (Nguyen et al., 2015). The best-fit models of sequence evolution were automatically computed in ModelFinder (IQTREE1.6.9) (Kalyaanamoorthy et al., 2017). Candidate models were selected according to Bayesian Information Criterion (BIC). The ultrafast bootstrap branch support (Minh et al., 2013) with 10,000 replicates was used to construct each dendrogram (values less than 95 are not significant). Unrooted trees were visualized using the web-based service iTOL 5.3 (¹⁸Letunic and Bork, 2019) and corrected in Adobe Illustrator CC 2017.

RESULTS

Lectin Genes in the Flax Genome

The search for genes encoding lectin domain-containing proteins was performed using the Pfam database¹⁹ and the lists of lectin domains present in literature (Van Damme et al., 2008; Jiang et al., 2010; Bellande et al., 2017; Van Holle and Van Damme, 2019). It allowed for the identification of 407 genes

⁵<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>

⁶<https://www.uniprot.org/>

⁷<http://bioinf.uni-greifswald.de/augustus/>

⁸<http://www.softberry.com/berry.phtml?topic=fgenes&group=programs&subgroup=gfind>

⁹<http://www.cbs.dtu.dk/services/SignalP/>

¹⁰<http://www.predisi.de/home.html>

¹¹<http://ipsort.hgc.jp/index.html>

¹²<http://www.cbs.dtu.dk/services/TMHMM/>

¹³<https://www.ebi.ac.uk/interpro/>

¹⁴<https://suba.live/>

¹⁶<https://rdr.io/github/roonysgalbi/tispec>

¹⁷<https://www.ebi.ac.uk/Tools/msa/clustalo/>

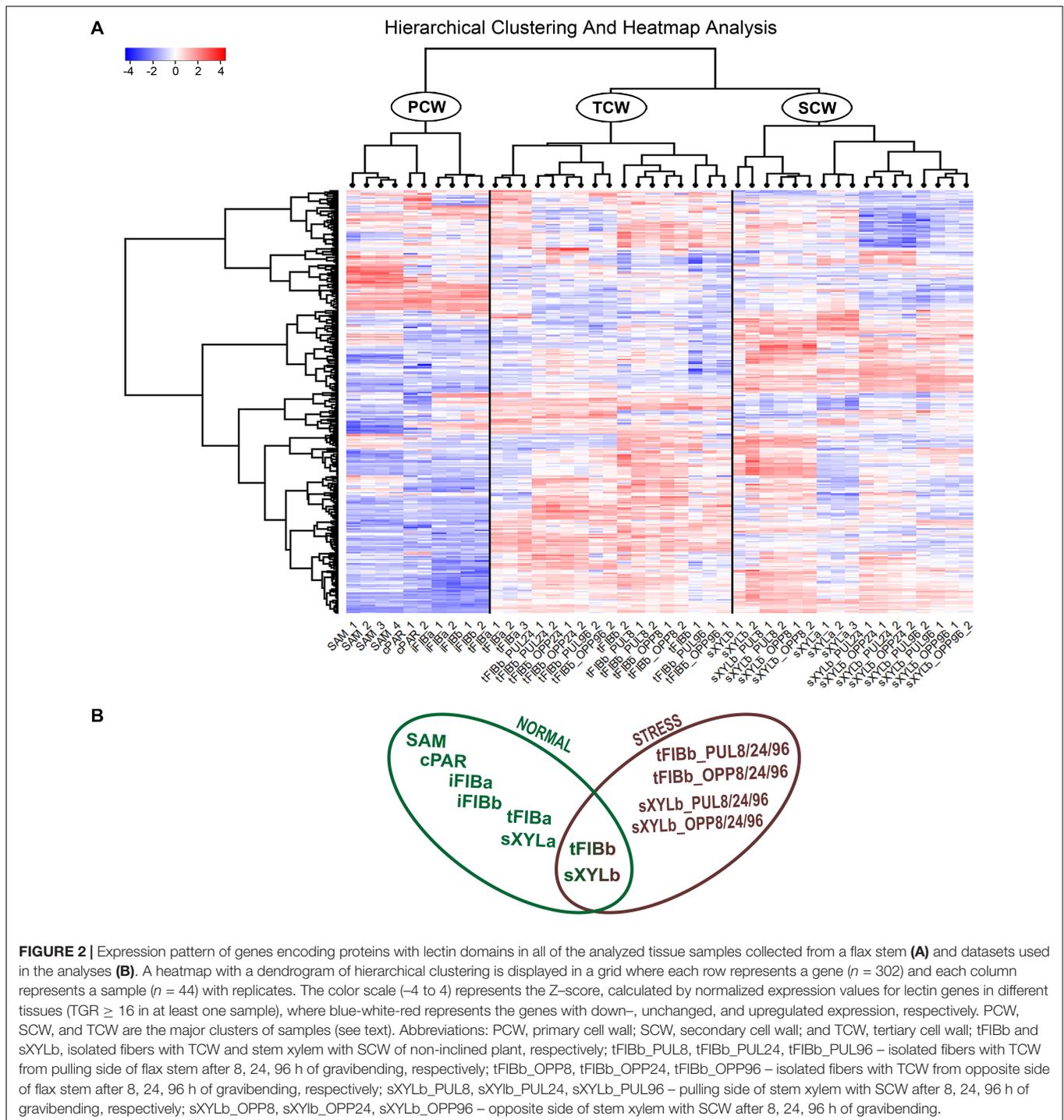
¹⁸<https://itol.embl.de/>

¹⁹<https://pfam.xfam.org/>

TABLE 1 | Distribution of proteins with lectin domains in the flax genome by families.

Family (according to Phytozome)	Tsaneva and Van Damme, 2020	Jiang et al., 2010	Designation used in this article	Pfam ID	Number of family members in the flax genome	Number of family members expressed in flax stem	Number of family members with SP	Number of family members with TM/*	Description (according to Phytozome)
B_lectin	GNA	B_lectin	GNA	PF01453	112	88	104	97/87	D-mannose binding lectin
Lectin_legB	Legume lectin domain	Lectin_legB	Legume	PF00139	72	52	63	61/58	Legume lectin domain
Malectin	-	-	Malectin	PF11721	39	37	32	33/32	Di-glucose binding within endoplasmic reticulum
PP2	Nictaba	Phloem	Nictaba	PF14299	37	33	0	0	Phloem protein 2
LysM	LysM domain	LysM	LysM	PF01476	36	21	29	25/18	LysM domain
Gal_Lectin	-	Gal_lectin	Galactose-binding lectin	PF02140	23	15	17	8/0	Galactose binding lectin domain
Malectin_like	-	-	Malectin-like	PF12819	22	18	17	17/15	Carbohydrate-binding protein of the ER
Agglutinin	Amaranthin domain	-	Amaranthin	PF07468	19	7	0	0	Agglutinin domain
Chitin_bind_1	Hevein domain	Chitin_bind_1	Hevein	PF00187	17	4	16	8/0	Chitin recognition protein (Hevein)
Gal-bind_lectin	-	Gal_binding_Lectin	Galectin-like	PF00337	11	11	0	11/11	Galactoside-binding lectin (galectin)
Calreticulin	-	Calreticulin	Calreticulin	PF00262	8	8	8	5/2	Calreticulin family
Jacalin	Jacalin-related domain	Jacalin	Jacalin	PF01419	4	4	0	0	Jacalin-like lectin domain
CRA	CRA	-	CRA	-	4	1	4	4/0	class V chitinase-related agglutinin
Lectin_C	-	Lectin_C	C-type	PF00059	2	2	2	2/2	Lectin C-type domain
EUL	EUL domain	EEA	EUL	PF14200	1	1	0	0	Ricin-type beta-trefoil lectin domain-like
Total					407	302	292	271/225	

Families are listed according to the number of members in the flax genome in the descending order. Abbreviations: SP – signal peptide (the prediction was performed using SignalP-5.0 (<http://www.cbs.dtu.dk/services/SignalP/>); Armenteros et al., 2019), PrediSi (<http://www.predisi.de/home.html>), IPSORT (<http://ipsort.hgc.jp/index.html>; Bannai et al., 2002), TM – transmembrane domain (the prediction was performed using TMHMM v2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>); Krogh et al., 2001), * – the number of family members with TM, except those with only one N-terminal TM, which according to TMHMM could be a possible signal sequence (signal peptide for this sequences was also predicted by SignalP).



in the flax genome (Table 1), which is approximately 1% of the total number of genes in the flax genome. We found representatives of 15 out of 18 described families of plant proteins with lectin domains (Table 1). Genes for members of 11 classical lectin families were recognized: amarantins, calreticulins, homologs of class V chitinases (CRA), C-type lectins that require calcium ions for carbohydrate binding, the *Euonymus europaeus* lectin (EUL) family, the *Galanthus*

nivalis agglutinin (GNA) family, the hevein family, the jacalin-related lectin family, the legume lectin family, the lysin motif (LysM) family, and the *Nicotiana tabacum* agglutinin (Nictaba) family. Lectins of the *Agaricus bisporus* agglutinin (ABA) and Cyanovirin-N (CV-N) families were not detected and have not been previously identified in angiosperms (Van Holle and Van Damme, 2019). Genes for members of the ricin B lectin family were not reliably identified in the current version of the

flax genome, though they are present in other higher plants (Van Holle and Van Damme, 2019).

Representatives of two protein families, malectin and malectin-like, that are often, but not always considered among plant proteins with lectin domains (Bellande et al., 2017) were also added to the list (**Table 1**). Malectin and malectin-like proteins are included in the CAZy database (²⁰Henrissat and Davies, 1997) because in bacteria, malectin domains (PF11721 and structurally similar PF12819) are attached to various glycosidase domains (GH2, GH16). However, in plant proteins, malectin and malectin-like domains are not combined with the domains that are enzymatically active on carbohydrates (Bellande et al., 2017; Franck et al., 2018).

Two other protein families, which in addition to a lectin domain have another domain that indicates enzymatic activity with respect to the bound carbohydrate, were included in this study. Galactose-binding lectin and galectin-like proteins have domains characteristic for β -galactosidase and for galactosyltransferase, respectively. These proteins are present in the classification of plant lectins by Jiang et al. (2010), but are absent in the lists of lectins given by other researchers (Van Damme et al., 2008; Bellande et al., 2017). Based on the presence of the domains characteristic for lectins, we included these groups of proteins in our study. In the flax genome, galactose-binding lectins and galectin-like families contained 23 and 11 genes, respectively (**Table 1**).

Lectin families may have different names in the literature. For example, the GNA-lectins (*Galanthus nivalis* agglutinin; named according to the plant from which the first representative with a similar domain in plants was isolated) (Van Damme et al., 1987) are also named B-type lectins (Jiang et al., 2010), D-mannose-binding lectins [Phytozome v12.1.6 database ((See text footnote 4) Goodstein et al., 2012)], and G-type lectins (Bellande et al., 2017). Many LecRLKs belonging to this family are also known as S-locus protein kinases due to the presence of the S-locus domain, which is involved in pollen self-incompatibility (Van Holle et al., 2017). All abbreviations of lectin family names used in this text are given in **Table 1**.

Each lectin sequence was analyzed for the presence of a signal peptide and transmembrane domains. Members of several lectin families, including Nictaba, amaranthin, jacalin, and EUL, have neither signal nor transmembrane peptides (**Table 1, Supplementary File 1**). Such proteins are translated on the free ribosomes in the cytoplasm and later remain in the cytoplasm or can be translocated into the nucleus (Lannoo and Van Damme, 2010). Signal peptides were detected in the vast majority of the proteins from most of the lectin families, suggesting that they are synthesized on the ribosomes attached to the endoplasmic reticulum and can be retained in this compartment or further transported through the Golgi apparatus, exposed to the extracellular space, anchored to the plasma membrane, or deposited into the cell wall; another possibility is transportation to the vacuoles. Altogether, two thirds of flax lectins are membrane-bound, as indicated by the presence of transmembrane domains (**Table 1**).

²⁰<http://www.cazy.org/>

General Characteristics of Lectin Expression in Flax Stem Tissues

To analyze the expression of genes for lectins in different parts of the flax stem, previously published RNA-Seq data (Zhang and Deyholos, 2016; Gorshkov et al., 2018, 2019; Gorshkova et al., 2018b; Mokshina et al., 2020) were used. The shoot apical meristem (SAM) sample contains young cells of various stem tissues at the beginning of their development. Cell division and elongation coupled with the formation of the primary cell wall (PCW) are characteristic for this stem zone. Primary cell wall deposition is also characteristic for fibers at the stage of intrusive elongation (iFIB) and for young cortical parenchyma (cPAR). Several cell types at a considerably more advanced stage of development constituted sXYL samples. Vessels, xylem fibers, and parenchyma of these samples have secondary cell walls (SCW) and are mainly involved in water transport and plant mechanical support (Esau, 1965). Finally, tFIB samples consist of the only cell type at a certain stage of development – phloem fibers that deposit tertiary cell walls (TCW) of specific composition, architecture, and function (Gorshkova et al., 2018a).

The expression levels of all lectin genes in the analyzed stem tissues of developing flax plants are given in **Supplementary File 3**. Approximately 75% of the genes for proteins with lectin domains present in the flax genome and constituting the families listed in **Table 1** were expressed in at least one analyzed stem sample (threshold of TGR value ≥ 16).

To better understand the variation of lectin gene expression across all samples, including stem tissues of plants grown under normal conditions and after plant inclination and development of gravitropic response, we performed the hierarchical clustering of 302 expressed genes (TGR ≥ 16 in at least one sample) encoding lectins of various families. According to the resulting heatmap (**Figure 2A**) all samples were subdivided into three clusters that corresponded to the cell wall type. The PCW cluster that contained SAM, cPAR, and iFIB samples ($n = 10$) consisting of cells with primary cell wall, was separated from SCW and TCW clusters. The SCW and TCW clusters were closer in distance to each other than to PCW, and formed two separate clusters. Cluster SCW contained xylem tissues ($n = 17$) that included mainly cells with secondary walls, whereas the TCW cluster consisted of samples represented by phloem fibers at the stage of tertiary cell wall deposition ($n = 17$).

Collectively, the visualization of expression patterns by heatmap demonstrated the presence of gene expression signatures associated with the distinct cell wall type both within the course of plant development and during gravitropic response induced by stem inclination. The findings herein determined the downstream analyses, which were conducted using two datasets: flax stem tissues under normal growth conditions and under conditions inducing a gravitropic response (**Figure 2B**). Further we considered the representatives of lectin families with certain expression characteristics starting from the dataset for samples collected from plants grown under normal conditions.

Genes for Proteins With Lectin Domains That Have Stable Expression Levels in All Analyzed Flax Tissues

Some genes encoding proteins with lectin domains exhibited similar levels of mRNA abundance in all analyzed stem parts. To identify the pool of stably expressed lectin genes, we calculated the tau-score for each lectin transcript in all samples of the dataset from the stem tissues under normal growth conditions (**Figure 2B**; **Supplementary File 3**). The tau-score may range from 0 to 1, with 0 and 1 indicating ubiquitously and specifically expressed genes, respectively (Yanai et al., 2005; Kryuchkova-Mostacci and Robinson-Rechavi, 2016). Based on the observed tau-scores (**Supplementary File 3**), the pool of flax lectin genes was extracted with the scores less than 0.15 that indicated an expression profile similar to house-keeping genes (**Table 2**).

For the calreticulin (PF00262) family, high gene expression levels were found for all eight members in all analyzed tissues (**Table 2**), belonging to putative calreticulins (CALR - *Lus10010923*, *Lus10010924*, *Lus10031409*, *Lus10031410*; CALRETICULIN-3 - *Lus10020222*, *Lus10026849*) and calnexins (CANX - *Lus10032521*, *Lus10043021*). The *A. thaliana* homologs of these genes encode glucose-binding lectins located in the endoplasmic reticulum (Lannoo and Van Damme, 2014). All flax lectins from the calreticulin family were predicted by the LocTree3 program to be localized in the endoplasmic reticulum (**Table 2**).

Genes for plasma membrane-localized proteins with lectin domains that had similar expression levels in all analyzed tissues included a noticeable proportion of genes with malectin (PF11721) and structurally related malectin-like (PF12819) domains (**Table 2**). Out of 61 flax genes that contain PF11721 or PF12819, one third was expressed at relatively constant levels (**Figure 3**, red dots). Most of these flax genes had LecRLK architecture, including *Lus10028140* and *Lus10042844*, which are both homologous to AT3G46290 for HERKULES1 (HERK1) reported to take part in the regulation of plant development (Li et al., 2016). *Lus10037887* (**Figure 3**) is homologous to AT5G54380 for THESEUS1 (THE1), a receptor kinase that mediates the response of growing plant cells to the perturbation of cellulose synthesis and may act as a cell-wall-integrity sensor (Hématy et al., 2007; Guo et al., 2009). In addition, constant expression levels (**Table 2**) were detected for genes encoding transmembrane proteins that have malectin domains but no kinase domains, like *Lus10015320*, *Lus10034275*, and *Lus10041488*; these were grouped together with RLP4 (AT1G28340) on the phylogenetic tree (**Figure 3**).

Low level of expression specificity between various stem tissues of flax was also detected for the genes encoding plasma membrane-localized proteins from several other lectin families, like legume lectins, GNA lectins, C-type, and LysM (**Table 2**). C-type (PF00059) lectins are scarce in plant genomes, but are widely present in vertebrates (Tsaneva and Van Damme, 2020); flax has only two C-type lectins (**Table 1**) and both were expressed in all analyzed stem tissues (**Table 2**). Legume lectins (PF00139) (such as concanavalin A from jack beans (Leguminosae family), found in abundance in seeds of this taxonomic group) are quite

numerous in the flax genome (**Figure 4**), which is similar to findings in other plant species (Bellande et al., 2017; Van Holle et al., 2017). Representatives of legume lectins have very divergent expression patterns (**Figure 4**, dots of different colors).

Low tau scores were also found for several genes for proteins with PF14299 (**Table 2**). According to Phytozome, this domain is named PP2 (phloem protein 2); however, its carbohydrate-recognizing part is designated as Nictaba sequence (Van Holle and Van Damme, 2018). Within the framework of used classification, proteins with the PP2 domain belong to the Nictaba lectin family. The Nictaba domain specifically recognizes high-mannose *N*-glycans, complex *N*-glycans, and, to a lesser extent, –GlcNAc oligomers (Tsaneva and Van Damme, 2020). Proteins with a Nictaba domain are considered nucleocytoplasmic and were confirmed to interact with O-GlcNAc-modified histones in the nucleus (Delporte et al., 2014).

Most of the flax genes (e.g., *Lus10032650*, *Lus10037238*, and *Lus10036154*) for proteins with the galectin-like domain (PF00337) were actively expressed in all analyzed tissues (**Table 2**). In plants, this domain is present in several galactosyltransferases that form a small gene family (GALT1-6 in *A. thaliana*) within family 31 of glycosyltransferases (GT) according to CAZy (²⁰Henrissat and Davies, 1997). GALT1 (AT1G26810) is a β -1,3-galactosyltransferase that adds galactose to a terminal β -*N*-acetylglucosamine during the formation of Lewis structures in *N*-glycans (Strasser et al., 2007; Showalter and Basu, 2016). The group of GALT1 homologs in flax was extended to four genes (**Figure 5**, **Supplementary File 3**). Out of 11 members of the galectin-like family in flax, all were predicted to possess transmembrane domains (**Figure 5** and **Table 1**). Based on the LocTree3 (**Table 2**) prediction, flax galactosyltransferases with PF00337 domains are localized to membranes of the Golgi apparatus.

Some proteins with lectin domains from several families were predicted to be secreted into the cell wall (**Table 2**). Several β -galactosidases of the galactose-binding lectin family (PF02140) (*Lus10000701*, *Lus10041798*, and *Lus10028348*) were among them (**Table 2**). The *in silico* characterization of flax enzymes from this family was performed earlier (Hobson and Deyholos, 2013); lectin domains were found in 22 of the 43 flax β -galactosidases and were always appended to the protein C-terminus.

The only member of the EUL family (PF14200) in the flax genome, *Lus10006551*, was expressed in all analyzed tissues (**Table 2**). ArathEULS3 protein has been shown to interact with specific *N*-glycans using glycan microarrays representing major glycan structures of glycoproteins and glycolipids (Van Hove et al., 2011).

Differential Expression of Lectin Genes in Various Stem Tissues

The results from the hierarchical clustering and heatmap analyses of normalized expression values for 302 lectin-encoding genes indicated three groups of samples differing based on the type of cell wall: PCW, SCW, and TCW (**Figure 2A**). In the current study, the threshold of fold changes ≥ 2 in pairwise

TABLE 2 | *Linum usitatissimum* genes encoding proteins with lectin domains with relatively constant expression values (the mean values of tau-score < 0.15 across all samples).

Gene ID	Lectin family	Pfam	Description (according to Phytozome)	Close <i>A.thaliana</i> homolog	SUBA	LocTree3	SAM	cPAR	iFIB	tFIB	sXYL
Lus10000579	GNA	PF01453	D-mannose binding lectin with Apple-like carbohydrate-binding domain	AT1G78850	extracellular	secreted	926	231	855	335	872
Lus10034637		PF01453	S-locus lectin protein kinase	AT1G11340	PM	PM	73	48	39	96	61
Lus10041680		PF01453	pleiotropic drug resistance 12	AT1G15520	PM	PM	74	84	42	129	92
Lus10043391		PF01453	S-domain-2 5	AT4G32300	PM	PM	172	176	132	362	413
Lus10016833	Legume	PF00139	Concanavalin A-like lectin protein kinase	AT5G10530	PM	PM	70	84	71	138	172
Lus10017194		PF00139	Concanavalin A-like lectin	AT5G01090	extracellular	PM	1473	283	542	286	520
Lus10019923		PF00139	Concanavalin A-like lectin protein kinase	AT3G53380	PM	PM	348	304	284	148	673
Lus10035542		PF00139	Concanavalin A-like lectin	AT5G01090	extracellular	PM	539	211	268	254	629
Lus10036986		PF00139	Concanavalin A-like lectin protein kinase	AT5G06740	PM	PM	78	77	69	66	75
Lus10005418	Malectin	PF11721	receptor-like kinase in flowers 1	AT1G29750	PM	PM	176	176	164	230	184
Lus10005550		PF11721	Leucine-rich repeat transmembrane protein kinase	AT1G53440	PM	PM	214	1252	746	1880	732
Lus10010951		PF11721	Leucine-rich repeat transmembrane protein kinase	AT1G07650	PM	PM	458	633	526	752	489
Lus10014185		PF11721	Malectin/receptor-like protein kinase	AT3G51550	PM	PM	661	781	1150	816	1482
Lus10015239		PF11721	receptor-like kinase in flowers 1	AT1G29750	PM	PM	372	304	290	301	229
Lus10015320		PF11721	receptor like protein 4	AT1G28340	PM	cytoplasm	432	448	394	419	327
Lus10017478		PF11721	Malectin/receptor-like protein kinase	AT4G39110	extracellular	PM	121	123	165	180	116
Lus10027509		PF11721	Protein kinase	AT5G24010	PM	PM	208	251	201	182	383
Lus10028140		PF11721	hercules receptor kinase 1	AT3G46290	PM	PM	637	435	765	666	1209
Lus10028805		PF11721	Malectin/receptor-like protein kinase	AT4G39110	extracellular	PM	75	142	155	141	142
Lus10031374		PF11721	Leucine-rich repeat transmembrane	AT1G07650	PM	PM	615	618	699	684	684
Lus10034275		PF11721	Di-glucose binding protein with Leucine-rich repeat domain	AT1G25570	golgi	PM	228	284	181	175	182
Lus10037887		PF11721	protein kinase	AT5G54380	PM	PM	349	570	1103	1562	1863
Lus10041488		PF11721	Di-glucose binding protein with Leucine-rich repeat domain	AT1G25570	PM	PM	360	243	301	359	326
Lus10042844		PF11721	hercules receptor kinase 1	AT3G46290	PM	PM	533	300	469	393	865
Lus10005769	Nictaba	PF14299	phloem protein 2-A13	AT3G61060	cytosol	chloroplast	312	1445	2140	537	407
Lus10007926		PF14299	phloem protein 2-B10	AT2G02360	nucleus	chloroplast	131	150	199	218	204
Lus10008642		PF14299	phloem protein 2-A15	AT3G53000	cytosol	chloroplast	120	329	232	228	261
Lus10013463		PF14299	phloem protein 2-B10	AT2G02360	nucleus	nucleus	188	130	148	192	217
Lus10024494		PF14299	phloem protein 2-A12	AT1G12710	cytosol	chloroplast	53	63	72	56	61
Lus10006841	LysM	PF01476	protein kinase/peptidoglycan-binding LysM domain-containing protein	AT1G51940	PM	PM	565	531	489	258	273
Lus10029925		PF01476	peptidoglycan-binding LysM domain-containing protein	AT1G55000	PM	nucleus	45	83	60	130	97
Lus10037586		PF01476	protein kinase/peptidoglycan-binding LysM domain-containing protein	AT1G51940	PM	PM	433	659	572	311	302

(Continued)

TABLE 2 | Continued

Gene ID	Lectin family	Pfam	Description (according to Phytozome)	Close <i>A.thaliana</i> homolog	SUBA	LocTree3	SAM	cPAR	iFIB	tFIB	sXYL	
Lus10000701	Galactose-binding lectin	PF02140	beta galactosidase 1	AT3G13750	extracellular	secreted	1186	1643	1695	909	1339	
Lus10003343		PF02140	beta-galactosidase 10	AT5G63810	extracellular	secreted	1033	1370	600	291	389	
Lus10016655		PF02140	beta galactosidase 9	AT2G32810	vacuole	cytoplasm	690	1432	1741	1087	3698	
Lus10022645		PF02140	beta-galactosidase 10	AT5G63810	extracellular	secreted	1556	2192	764	368	452	
Lus10025980		PF02140	beta-galactosidase 3	AT4G36360	extracellular	PM	931	1361	898	323	328	
Lus10028348		PF02140	beta-galactosidase 3	AT4G36360	extracellular	PM	300	2325	2677	1625	1890	
Lus10041798		PF02140	beta-galactosidase 3	AT4G36360	extracellular	PM	602	2905	3039	2691	2966	
Lus10002456		Malectin-like	PF12819	Leucine-rich repeat (LRR) protein	AT3G05990	extracellular	PM	569	198	332	376	413
Lus10010528	PF12819		Leucine-rich repeat (LRR) protein	AT3G19230	extracellular	PM	232	121	193	232	194	
Lus10015057	PF12819		Leucine-rich repeat protein kinase	AT2G37050		PM	PM	475	551	757	394	313
Lus10022728	PF12819		Malectin/receptor-like protein kinase	AT3G51550		PM	PM	541	748	1124	1328	2497
Lus10023166	PF12819		Leucine-rich repeat protein kinase	AT2G37050		PM	PM	523	694	1009	486	455
Lus10037051	PF12819		Leucine-rich repeat protein kinase	AT1G67720		PM	PM	491	276	2590	1479	309
Lus10004256	Galactin-like	PF00337	Galactosyltransferase	AT5G62620	golgi	GM	83	89	94	124	131	
Lus10030467		PF00337	Galactosyltransferase	AT1G26810	extracellular	GM	50	37	78	112	88	
Lus10032650		PF00337	Galactosyltransferase	AT3G06440	mitochondrion	GM	446	526	670	626	951	
Lus10036154		PF00337	Galactosyltransferase	AT4G21060	ER, golgi	GM	1046	764	623	687	635	
Lus10036686		PF00337	Galactosyltransferase	AT5G62620	golgi	GM	248	265	256	358	192	
Lus10037238		PF00337	Galactosyltransferase	AT5G62620	golgi	GM	238	388	464	524	509	
Lus10042164		PF00337	Galactosyltransferase	AT5G62620	golgi	GM	162	155	208	205	279	
Lus10010923		Calreticulin	PF00262	calreticulin 1a	AT1G56340	ER	ER	5129	2968	3250	1983	3511
Lus10010924	PF00262		calreticulin 1a	AT1G56340	ER	ER	3093	2355	1600	924	1234	
Lus10020222	PF00262		calreticulin 3	AT1G08450	ER	ER	652	627	573	2277	1139	
Lus10026849	PF00262		calreticulin 3	AT1G08450	ER	ER	2226	1639	1520	3154	2079	
Lus10031409	PF00262		calreticulin 1a	AT1G56340	ER	ER	1624	1544	1218	729	947	
Lus10031410	PF00262		calreticulin 1a	AT1G56340	ER	ER	4498	3176	3474	1559	3009	
Lus10032521	PF00262		calnexin 1	AT5G61790	ER	ER	6660	3071	3797	2991	5115	
Lus10043021	PF00262		calnexin 1	AT5G61790	ER	ER	4879	3778	3477	3471	5338	
Lus10006866	Jacalin		PF01419	Mannose-binding lectin	AT1G73040	cytosol	cytoplasm	323	502	406	603	255
Lus10023500	C-type	PF00059	kinase/C-type lectin domain-containing protein	AT1G52310	PM	PM	120	97	75	119	95	
Lus10040381		PF00059	kinase/C-type lectin domain-containing protein	AT1G52310	PM	PM	285	298	191	159	126	
Lus10006551	EUL	PF14200	hydroxyproline-rich glycoprotein	AT2G39050	nucleus	nucleus	700	241	659	964	1486	

Only genes with TGR ≥ 50 in at least one sample are given and the rest are presented in **Supplementary File 3**. The gene description includes the gene ID according to Phytozome v12.1.6 (<https://phytozome.jgi.doe.gov/pz/portal.html>; Goodstein et al., 2012); the lectin family name that was used in present study; the Pfam domain by the presence of which lectin family members were distinguished; the gene description according to Phytozome v12.1.6; *Arabidopsis thaliana* homolog for a flax gene according to flax genome annotation (Wang et al., 2012); subcellular localization of *A. thaliana* homologs for flax lectins according to the SUBA4 database (<https://suba.live/>; Hooper et al., 2017); and subcellular localization of flax lectins (corrected protein sequences presented in **Supplementary File 1** were used) according to the LocTree3 prediction tool (<https://rostlab.org/services/loctree3/>; Goldberg et al., 2014). The red-blue heat map indicates that the mRNA abundance expressed in TGR (total gene read) values and the color gradient reflects TGR values of 1 to 500. Abbreviations: SAM, shoot apical meristem; cPAR, cortical parenchyma; iFIB, phloem fibers at the stage of intrusive elongation; sXYL, stem xylem with secondary cell wall; tFIB, phloem fibers depositing tertiary cell walls; PM, plasma membrane; ER, endoplasmic reticulum; GM, Golgi membrane.

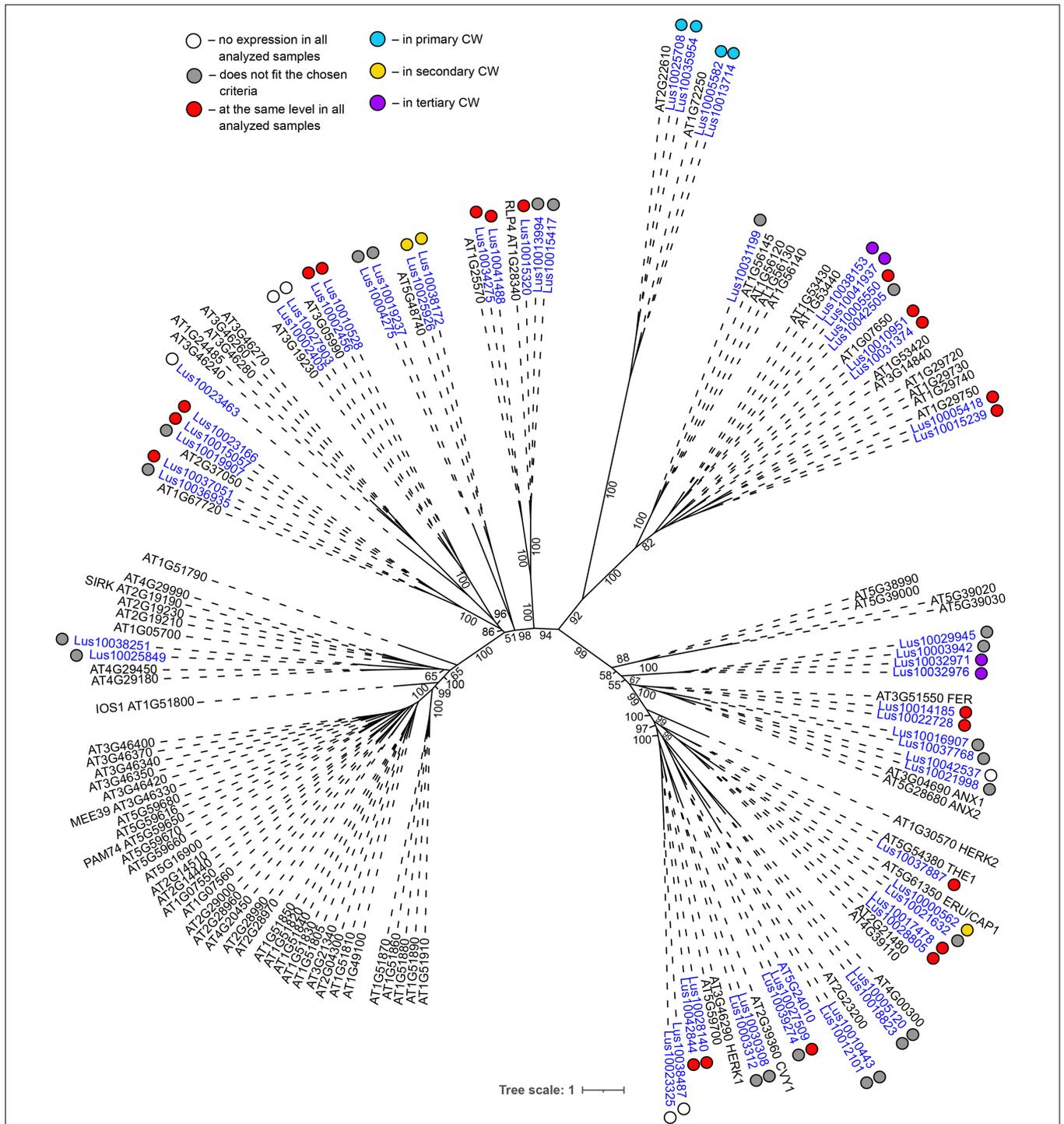
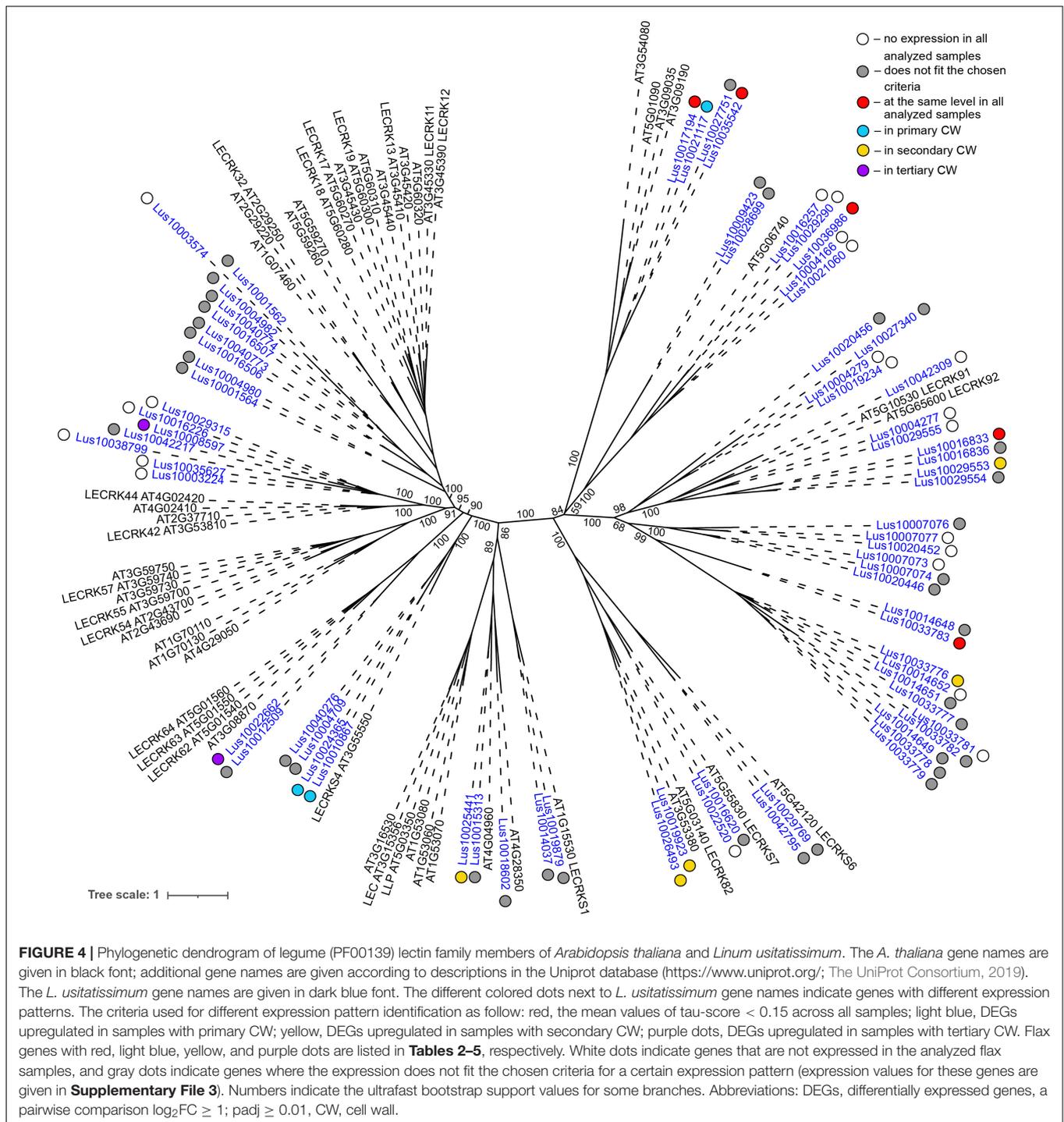


FIGURE 3 | Phylogenetic dendrogram of malectin (PF11721) and malectin-like (PF12819) family members of *Arabidopsis thaliana* and *Linum usitatissimum*. The *A. thaliana* gene names are given in black font; additional gene names are given according to descriptions in the Uniprot database (<https://www.uniprot.org/>; The UniProt Consortium, 2019). The *L. usitatissimum* gene names are given in dark blue font. The different colored dots next to *L. usitatissimum* gene names indicate genes with different expression patterns. The criteria used for different expression pattern identification as follow: red, the mean values of tau-score < 0.15 across all samples; light blue, DEGs upregulated in samples with primary CW; yellow, DEGs upregulated in samples with secondary CW; purple dots, DEGs upregulated in samples with tertiary CW. Flax genes with red, light blue, yellow, and purple dots are listed in **Tables 2–5**, respectively. White dots indicate genes that are not expressed in the analyzed flax samples, and gray dots indicate genes where the expression does not fit the chosen criteria for a certain expression pattern (expression values for these genes are given in **Supplementary File 3**). Numbers indicate the ultrafast bootstrap support values for some branches. Abbreviations: DEGs, differentially expressed genes, a pairwise comparison $\log_2FC \geq 1$; $\text{padj} \geq 0.01$, CW, cell wall.

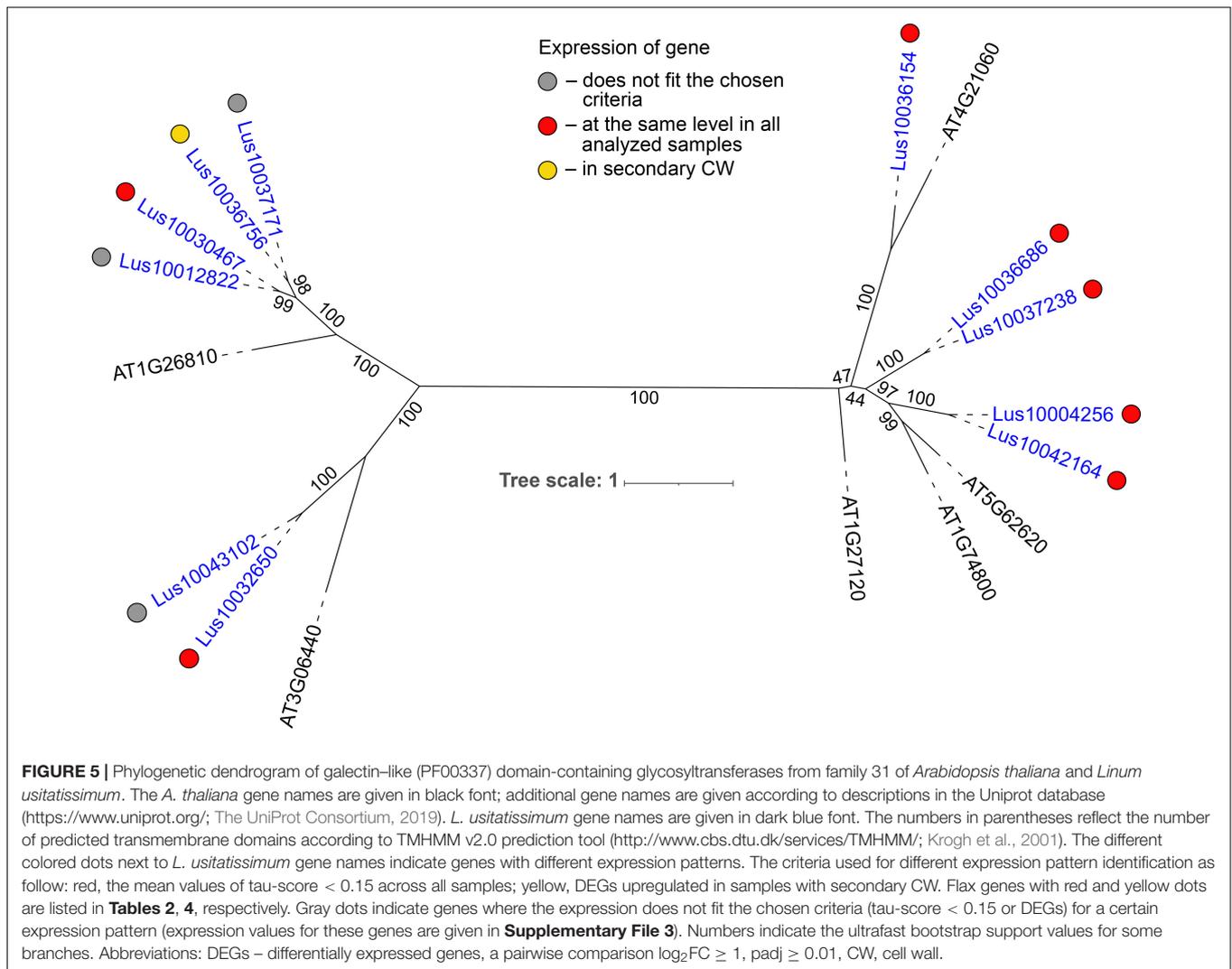


comparisons with other sample types ($p \leq 0.01$, $TGR \geq 16$ in at least one sample) was used to consider a gene as differentially expressed.

Genes for Proteins With Lectin Domains Upregulated in Tissues With Primary Cell Walls

Several lectin genes were predominantly expressed in SAM, together with other tissues with primary cell walls –

cPAR and iFIB (**Table 3**). Among these tissues, genes for four flax malectins, homologs of *A. thaliana* genes for Di-glucose binding protein with kinesin motor domains were detected: *Lus10025708* and *Lus10035954*, both homologous to *AT2G22610*; *Lus10013714* and *Lus10005582*, both homologous to *AT1G72250*. The latter grouped a separate clade in the malectin family (**Figure 3**). Both *A. thaliana* homologs, named *MDKIN1* (*At1g72250*)



and *MDKIN2* (*At2g22610*) (MALECTIN DOMAIN KINESIN) are expressed in cell division zones and in vasculature; their experimentally established intracellular localization is predominantly associated with nuclei (Galindo-Trigo et al., 2020).

Three lectin genes were detected with the established thresholds as upregulated predominantly in SAM (**Table 3**): *Lus10021117* (homolog of *AT5G01090* encoding legume lectin family protein with PF00139 domain), *Lus10018191* and *Lus10025643*, both homologous to *AT1G21880* with LysM domain (PF01476). The latter encodes a GPI-anchored protein named LYM1 that belongs to LecP and was characterized as participating in the recognition of bacterial peptidoglycan (Willmann et al., 2011).

Most of the lectins with upregulated gene expression in the samples with primary cell walls were localized in the plasma membrane according to the LocTree3 prediction. The examples included *Lus10010867* and *Lus10024365*, which are homologous to *AT3G55550*, encoding legume-type lectin receptor kinase *LECRKS4* (Bouwmeester and Govers, 2009) (**Table 3**). Some

flax proteins with lectin domains with upregulated expression in cells depositing primary cell walls were predicted to be secreted; these included amaranthin (*Lus10016109*) and two β -galactosidases (*Lus10003343* and *Lus10022645*, both homologous to *AT5G63810*) (**Table 3**). Nineteen flax genes for amaranthin-like lectins and their variable expression in flax tissues were previously characterized (Faruque et al., 2015), *Lus10016109* was designated as *LuALL7*.

Genes for Proteins With Lectin Domains Upregulated in Samples Depositing Secondary Cell Walls

Numerous lectins were upregulated in sXYL samples (**Table 4**) that contain several cell types, all depositing secondary cell wall: vessels, xylem fibers, and xylem parenchyma. sXYL samples were collected in two locations of the stem (**Figure 1**); the expression data for sXYLa and sXYLb samples were combined and averaged to reveal genes upregulated during secondary cell wall deposition.

The most common groups identified were lectins from GNA, legume, LysM, malectin, and Nictaba (PP2) families. Of these, only the latter family was not predicted to be

TABLE 3 | *Linum usitatissimum* genes encoding proteins with lectin domains upregulated in flax stem tissue samples with the primary cell wall.

Gene ID	Lectin family	Pfam	Description (according to Phytozome)	Close <i>A.thaliana</i> homologs	SUBA	LocTree3	SAM	cPAR	iFIB	tFIB	sXYL
Lus10020080	GNA	PF01453	S-locus lectin protein kinase family protein	AT4G21390	PM	PM	50	36	20	9	10
Lus10010867	Legume	PF00139	Concanavalin A-like lectin protein kinase family protein	AT3G55550	PM	PM	89	42	138	22	22
Lus10021117		PF00139	Concanavalin A-like lectin family protein	AT5G01090	extracellular	PM	630	123	134	95	156
Lus10024365		PF00139	Concanavalin A-like lectin protein kinase family protein	AT3G55550	PM	PM	166	13	140	15	24
Lus10005582	Malectin	PF11721	Di-glucose binding protein with Kinesin motor domain	AT1G72250	nucleus	ER	572	334	413	47	84
Lus10013714		PF11721	Di-glucose binding protein with Kinesin motor domain	AT1G72250	nucleus	cytoplasm	647	450	426	47	95
Lus10025708		PF11721	Di-glucose binding protein with Kinesin motor domain	AT2G22610	nucleus	cytoplasm	824	1007	878	119	163
Lus10035954		PF11721	Di-glucose binding protein with Kinesin motor domain	AT2G22610	nucleus	cytoplasm	715	927	758	157	179
Lus10005769	Nictaba	PF14299	phloem protein 2-A13	AT3G61060	cytosol	chloroplast	312	1445	2140	537	407
Lus10025643	LysM	PF01476	lysm domain GPI-anchored protein 1 precursor	AT1G21880	PM	PM	1600	362	720	277	389
Lus10018191		PF01476	lysm domain GPI-anchored protein 1 precursor	AT1G21880	PM	PM	1687	116	264	105	152
Lus10022345		PF01476	Protein kinase superfamily protein	AT3G01840	PM	PM	96	68	113	20	30
Lus10003343	Galactose-binding lectin	PF02140	beta-galactosidase 10	AT5G63810	extracellular	secreted	1033	1370	600	291	389
Lus10022645		PF02140	beta-galactosidase 10	AT5G63810	extracellular	secreted	1556	2192	764	368	452
Lus10025980		PF02140	beta-galactosidase 3	AT4G36360	extracellular	PM	931	1361	898	323	328
Lus10014278		PF02140	beta galactosidase 9	AT2G32810	vacuole	PM	379	890	716	183	141
Lus10016109	Amaranthin	PF07468	PF07468 - Agglutinin domain (Agglutinin)	-	-	secreted	217	98	205	35	11

A pairwise comparison ($\log_2FC \geq 1$, $padj \geq 0.01$) was made with tissue samples that deposit secondary and tertiary cell walls. The gene description includes the gene ID according to Phytozome v12.1.6 (<https://phytozome.jgi.doe.gov/pz/portal.html>; Goodstein et al., 2012); the lectin family name that was used in present study; the Pfam domain by the presence of which lectin family members were distinguished; the gene description according to Phytozome v12.1.6; *Arabidopsis thaliana* homolog for a flax gene according to flax genome annotation (Wang et al., 2012); subcellular localization of *A. thaliana* homologs for flax lectins (corrected protein sequences presented in **Supplementary File 1** were used) according to the SUBA4 database (<https://suba.live/>; Hooper et al., 2017); and subcellular localization of flax lectins according to the LocTree3 prediction tool (<https://roslab.org/services/loctree3/>; Goldberg et al., 2014). The red-blue heat map indicates that the mRNA abundance expressed in TGR (total gene reads) values and the color gradient reflects TGR values of 1 to 1000. Abbreviations: SAM, shoot apical meristem; cPAR, cortical parenchyma; iFIB, phloem fibers at the stage of intrusive elongation; sXYL, stem xylem with secondary cell wall; tFIB, phloem fibers depositing tertiary cell walls; PM, plasma membrane; ERM, endoplasmic reticulum membrane.

TABLE 4 | *Linum usitatissimum* genes encoding proteins with lectin domains upregulated in flax stem tissue samples with the secondary cell wall.

Gene ID	Lectin family	Pfam	Description (according to Phytozome)	Close <i>A. thaliana</i> homologs	SUBA	LocTree3	SAM	cPAR	iFIB	tFIB	sXYL
Lus10036639	GNA	PF01453	receptor-like protein kinase 1	AT5G60900	PM	PM	80	21	37	46	664
Lus10042940		PF01453	receptor-like protein kinase 1	AT5G60900	PM	PM	5	0	0	82	624
Lus10000249		PF01453	S-domain-2 5	AT4G32300	PM	PM	138	161	81	177	416
Lus10018405		PF01453	receptor kinase 3	AT4G21380	PM	PM	122	34	23	91	315
Lus10014810		PF01453	S-locus lectin protein kinase family protein	AT4G27290	PM	PM	8	0	4	77	258
Lus10029802		PF01453	S-locus lectin protein kinase family protein	AT2G19130	PM	PM	26	0	0	20	243
Lus10014808		PF01453	S-locus lectin protein kinase family protein	AT4G27290	PM	PM	0	0	0	11	85
Lus10019923		Legume	PF00139	Concanavalin A-like lectin protein kinase family protein	AT3G53380	PM	PM	348	304	284	148
Lus10026493	PF00139		Concanavalin A-like lectin protein kinase family protein	AT3G53380	PM	PM	200	151	148	118	374
Lus10025441	PF00139		Concanavalin A-like lectin protein kinase family protein	AT4G04960	extracellular	PM	29	46	15	40	309
Lus10029553	PF00139		Concanavalin A-like lectin protein kinase family protein	AT5G10530	PM	PM	26	109	71	68	221
Lus10033776	PF00139		receptor lectin kinase	AT2G37710	PM	PM	0	0	0	9	123
Lus10031473	Nictaba		PF14299	phloem protein 2-B15	AT1G09155	PM	chloroplast	2	1	7	35
Lus10031484		PF14299	phloem protein 2-B15	AT1G09155	PM	secreted	1	0	5	39	645
Lus10031472		PF14299	phloem protein 2-B15	AT1G09155	PM	chloroplast	1	1	2	17	261
Lus10015208		PF14299	phloem protein 2-B15	AT1G09155	PM	chloroplast	1	1	1	10	175
Lus10015209		PF14299	phloem protein 2-B15	AT1G09155	PM	chloroplast	0	0	0	8	133
Lus10008586	LysM	PF01476	Protein kinase superfamily protein	AT2G33580	PM	PM	150	94	100	311	792
Lus10031662		PF01476	peptidoglycan-binding LysM domain-containing protein	AT3G52790	extracellular	secreted	0	0	1	41	700
Lus10023945		PF01476	lysm domain GPI-anchored protein 2 precursor	AT2G17120	PM	PM	130	258	169	119	367
Lus10027407		PF01476	peptidoglycan-binding LysM domain-containing protein	AT3G52790	extracellular	secreted	0	0	0	6	223
Lus10015011		PF01476	peptidoglycan-binding LysM domain-containing protein	AT5G62150	extracellular	secreted	0	0	0	1	31
Lus10016655		Galactose-binding lectin	PF02140	beta galactosidase 9	AT2G32810	vacuole	cytoplasm	690	1432	1741	1087
Lus10000271	PF02140		beta-galactosidase 8	AT2G28470	extracellular	cytoplasm	142	233	373	331	1087
Lus10036108	PF02140		beta-galactosidase 8	AT2G28470	extracellular	cytoplasm	104	174	254	270	830
Lus10025926	Malectin-like	PF12819	Leucine-rich repeat protein kinase family protein	AT5G48740	PM	PM	29	4	637	136	2133
Lus10000562		PF12819	Protein kinase superfamily protein	AT5G61350	PM	PM	0	0	1	1	128
Lus10038172		PF12819	Leucine-rich repeat protein kinase family protein	AT5G48740	PM	PM	6	1	30	7	112
Lus10036756	Galectin-like	PF00337	galactosyltransferase1	AT1G26810	extracellular	GM	220	190	148	169	614

A pairwise comparison ($\log_2FC \geq 1$, $padj \geq 0.01$) was made with tissue samples that deposit primary and tertiary cell walls. The gene description includes the gene ID according to Phytozome v12.1.6 (<https://phytozome.jgi.doe.gov/pz/portal.html>; Goodstein et al., 2012); the lectin family name that was used in present study; the Pfam domain by the presence of which lectin family members were distinguished; the gene description according to Phytozome v12.1.6; *Arabidopsis thaliana* homolog for a flax gene according to flax genome annotation (Wang et al., 2012); subcellular localization of *A. thaliana* homologs for flax lectins according to the SUBA4 database (<https://suba.live/>; Hooper et al., 2017); and subcellular localization of flax lectins (corrected protein sequences presented in **Supplementary File 1** were used) according to the LocTree3 prediction tool (<https://roslab.org/services/loctree3/>; Goldberg et al., 2014). The red-blue heat map indicates that the mRNA abundance expressed in TGR (total gene reads) values and the color gradient reflects TGR values of 1 to 1000. Abbreviations: SAM, shoot apical meristem; cPAR, cortical parenchyma; iFIB, phloem fibers at the stage of intrusive elongation; sXYL, stem xylem with secondary cell wall; tFIB, phloem fibers depositing tertiary cell walls; PM, plasma membrane; GM, Golgi membrane.

sXYL samples, the most pronounced expression was observed for *Lus10036639* and *Lus10042940*, which are both homologous to *AT5G60900*, encoding RLK1a, a receptor like kinase with leucine-rich repeats (Table 4). Six genes of legume lectin receptor kinases were activated in xylem tissues [flax homologs of *LECRK-VIII.1* (*AT3G53380*), *LECRK-VII.1* (*AT4G04960*), *LECRK-IX.1* (*AT5G10530*), and *LECRK-IV.1* (*AT2G37710*)] (Table 4).

Several genes for lectins with a LysM motif that localized to the plasma membrane were upregulated in sXYL samples (Table 4). These included *Lus10008586*, which is homologous to *A. thaliana* LYK5 (*AT2G33580*) – a protein with LecRLK type architecture, and *Lus10023945*, which is homologous to LYM2 (*AT2G17120*), a GPI-anchored LecP type protein. LYK5 and LYM2 belong to distant clades (Figure 7). Three LecRLP type lectins predicted to be secreted into the cell wall showed higher gene expression levels, namely *Lus10031662*, *Lus10027407* (Table 4), and *Lus10015011* (Supplementary File 3). These three LecRLP type lectins clustered together on the phylogenetic tree (Figure 7). The functions of their closest *A. thaliana* homologs, *AT3G52790*, *AT4G25433* and *AT5G62150* are unknown.

Three genes for protein kinases with a malectin-like domain (PF12819), *Lus10000562* (homologous to *AT5G61350*), *Lus10025926*, and *Lus10038172* (both homologous to *AT5G48740*), showed an increased mRNA abundance in sXYL tissues compared to other samples (Table 4). Galactosyltransferase (*Lus10036756*) with a galectin-like domain that is homologous to GALT1 from *A. thaliana* also showed pronounced expression. GALT1 is a β -d-1,3-galactosyltransferase that is involved in the formation of *N*-glycans (Strasser et al., 2007). Expression of the specific isoform indicates some peculiarities of *N*-glycans in xylem tissue; however, nothing is currently known about *N*-glycoproteins that are specific to secondary cell walls.

Genes for Proteins With Lectin Domains Upregulated in Samples Depositing Tertiary Cell Wall

Tissue- and stage-specific expression of lectin genes was exemplified by flax phloem fibers isolated from the stem at an advanced stage of specialization when they deposited tertiary cell wall. Fibers were isolated at two locations in the stem, both located below the snap point (Figure 1; Gorshkova et al., 2003); the expression data for tFIBa and tFIBb samples were combined and averaged to reveal genes upregulated at the stage of tertiary cell wall deposition. Based on the expression profile, lectins from several families were considerably upregulated in fibers at this stage of development (Table 5).

Three genes encoding proteins with jacalin-like domains (PF01419), *Lus10024290*, *Lus10024291* (both homologs of *AT1G19715*), and *Lus10037605* (homolog of *AT1G73040*) were notably upregulated in tFIB. The flax genome contains only four jacalin genes, three of which were detected as highly activated in fibers depositing tertiary cell walls. No sequences for signal peptides and transmembrane domains were detected in flax jacalins (Table 1), however, the jacalin encoded by *Lus10024290* with the highest expression level was predicted to be secreted into the cell wall by LocTree3 program (Table 5). The extracellular

jacalin Horcolin devoid of a signal peptide was biochemically isolated from barley coleoptiles (Grunwald et al., 2007).

On the dendrogram, all flax jacalins localized within a small cluster that combined several sequences of rice and *A. thaliana* jacalins (Figure 8A). The jacalin domain is often present in proteins as tandem repeats, which can be combined with other domains (Eggermont et al., 2017). In accordance, two flax lectins belonging to this protein family had three jacalin domains in a row, while two others had only one (Figure 8B). The jacalin with a stable character of expression (*Lus10006866*, Table 2) had a jacalin domain combined with the F-box, while those upregulated in fibers depositing tertiary cell wall have only jacalin domains (Figure 8B).

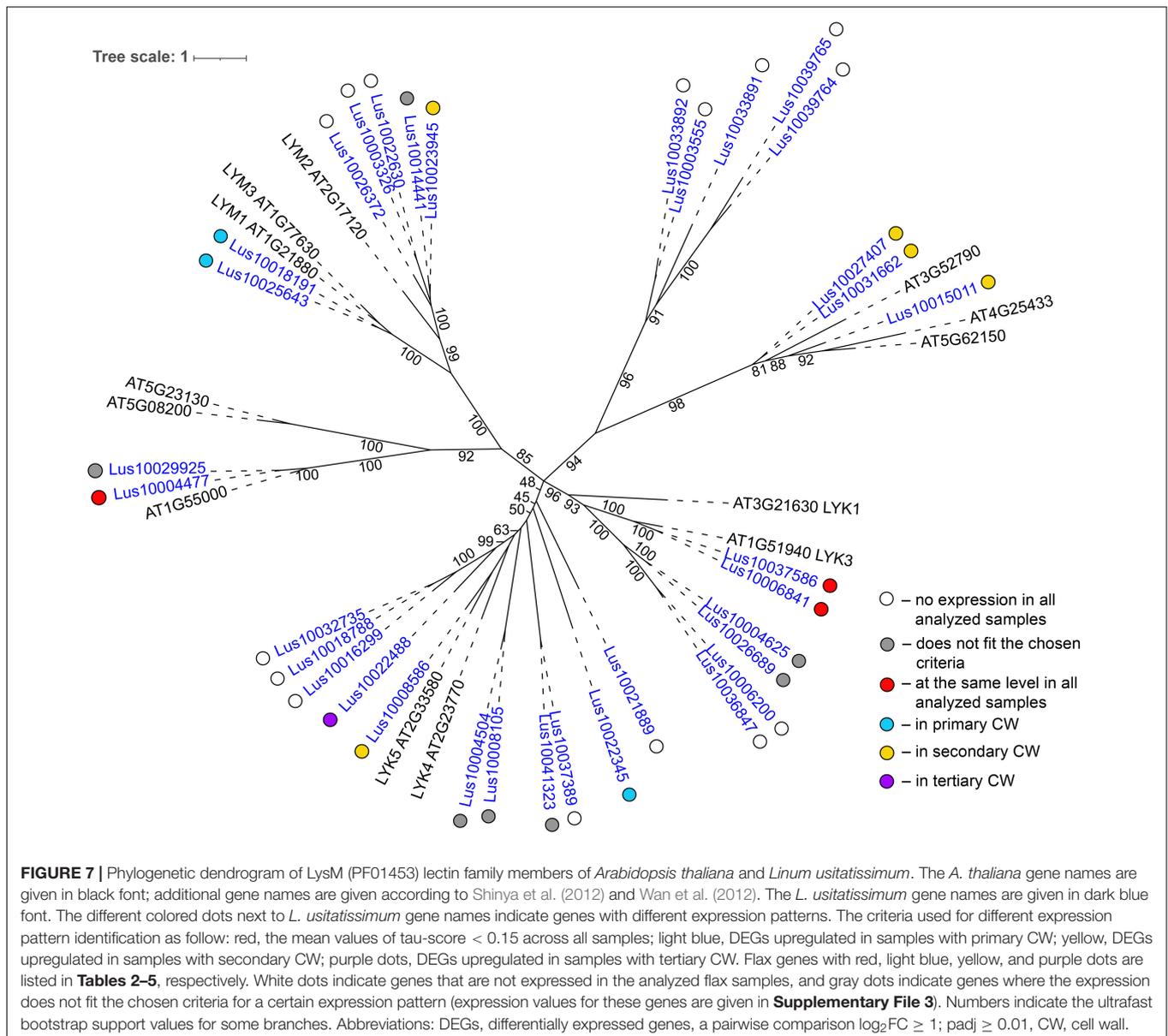
The *LuALL4* (*Lus10029186*) and *LuALL11* (*Lus10010708*) members of the amaranthin family (PF07468) (Faruque et al., 2015) were upregulated in the fibers depositing tertiary cell wall. These lectins, which are distinct from *LuALL7* mainly expressed in tissues with primary cell wall, have a pathogenesis-related protein Bet v I domain. The four flax genes belonging to the Nictaba family were also upregulated at the tertiary cell wall formation. Two of them are homologs to the *A. thaliana* phloem protein PP2-A1 (*AT4G19840*), which contains only a Nictaba domain and is part of the phloem protein bodies in the sieve elements. Recombinant protein production and glycan array analysis demonstrated the binding of PP2-A1 to *N*-acetylglucosamine oligomers, high-mannose *N*-glycans, and 9-acyl-*N*-acetylneuraminic sialic acid (Beneteau et al., 2010).

Over a half of lectins upregulated in fibers with tertiary cell walls were predicted to localize at the plasma membrane (Table 5). Among them, the most numerous were LecRLKs with a GNA domain (PF01453) and genes belonging to the malectin (PF11721) family. *Lus10009582* encoding the lectin kinase with the highest expression among GNA-type lectins, was specifically expressed in tFIB samples (Table 5). Two genes from the malectin family, *Lus10038153* and *Lus10041937*, which are both homologous to *AT1G53440* that encodes leucine-rich repeat transmembrane protein kinases, were additionally highly expressed in cortical parenchyma. Two other malectins, *Lus10032971* and *Lus10032976* are homologous to FERONIA (*AT3G51550*), which is the receptor-like kinase in *A. thaliana* demonstrated to interact with cell wall pectins and to be involved in numerous developmental processes (Li et al., 2018).

Changes in Lectin Gene Expression Upon Plant Gravistimulation

To characterize the expression of lectin genes during induced gravitropic response, we used a dataset including samples from the control non-inclined plants (sXYLb and tFIBb) and samples from the region of the formation of stem curvature developed to return the inclined plants to vertical position (Ibragimova et al., 2017; Gorshkov et al., 2018). Each stem segment was separated into pulling (PUL) and opposite (OPP) sides both in the xylem and in the phloem (Figure 1); the outer peel of each stem side was used to isolate phloem fibers.

To identify the most prominent representatives of lectins expressed during gravitropic response, we performed gene



clustering based not on the absolute expression strength but rather on the amount by which each gene deviates in a specific sample from the gene's average across all samples. Hence, we centered the values of each gene in the samples and built a heatmap. The top 30 genes with the highest variance of expression levels across samples are shown in the **Figure 9** and **Table 6**.

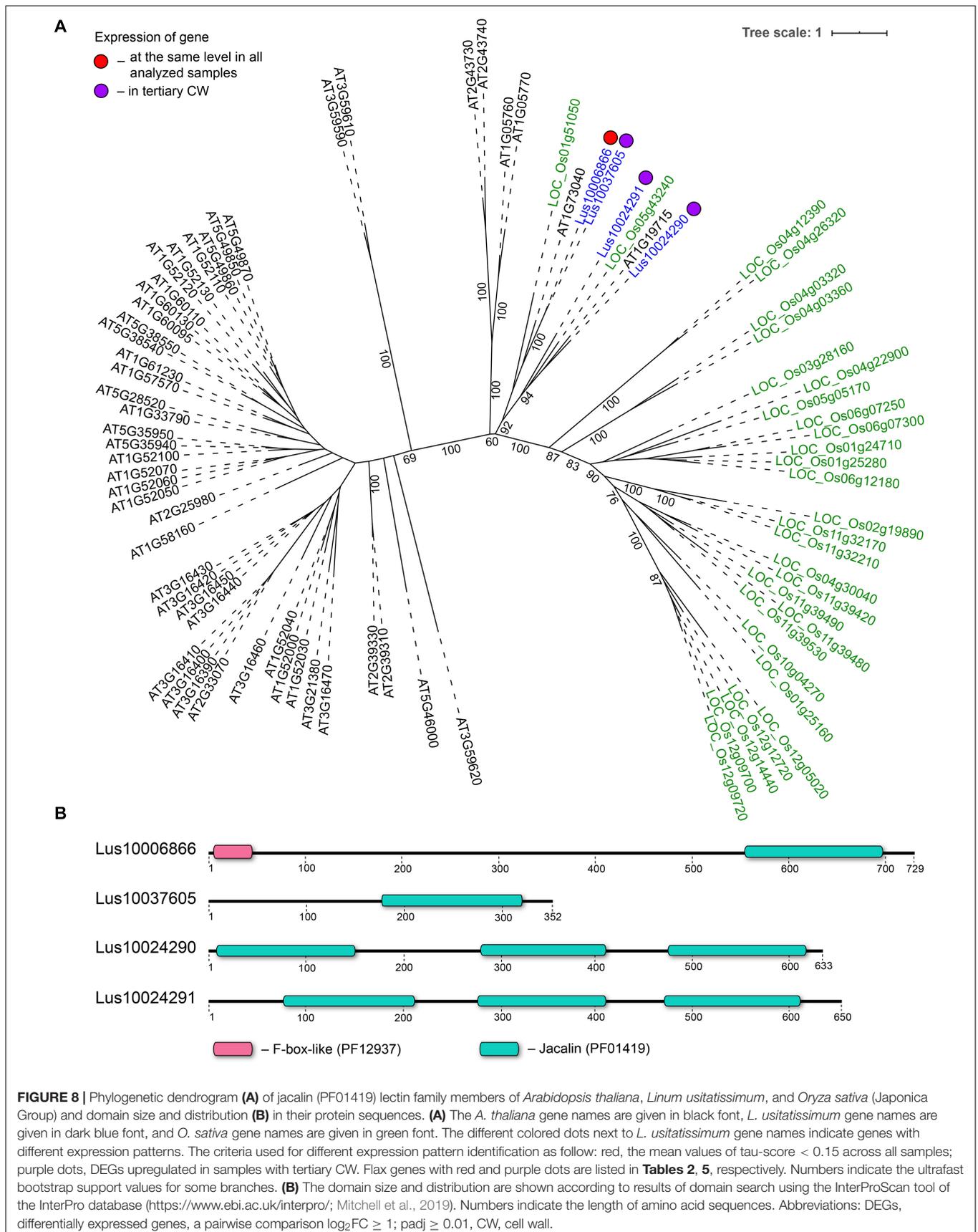
Results of the clustering analysis showed three expression patterns of lectin genes in plants subjected to gravistimulation (**Figure 9** and **Table 6**). Cluster 1 was dominated by transcripts abundant in phloem fibers (**Figure 9** and **Table 6**, Cluster 1), moreover, 3 out of 7 genes (genes for GNA lectin *Lus10009582*, for amarantin *Lus10029186*, and for hevein *Lus10006552*) of this cluster were already mentioned as genes specifically expressed in tFIB samples (see section “Genes for Proteins With Lectin Domains Upregulated in Samples Depositing Tertiary Cell Wall,”

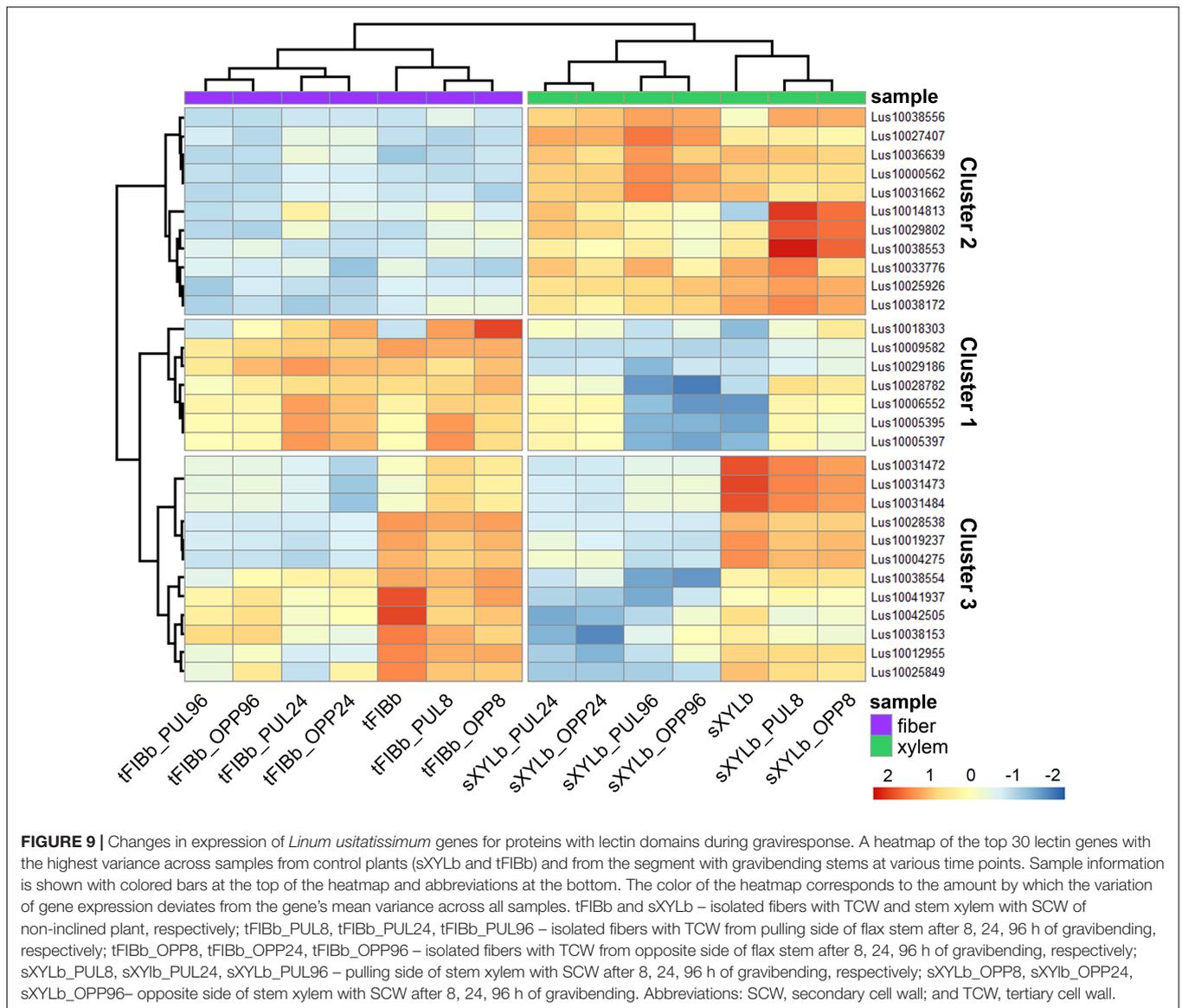
Table 5). Cluster 2 was represented by lectin genes, which were up-regulated in xylem (**Figure 9** and **Table 6**, Cluster 2), and most of genes in this cluster were designated as DEGs upregulated in sXYL samples fibers (see section “Genes for Proteins With Lectin Domains Upregulated in Samples Depositing Secondary Cell Walls”). Cluster 3 grouped transcripts that had high expression in analyzed tissues of both non-inclined plants and gravistimulated plants (8 h after inclination) and further were drastically decreased (**Figure 9** and **Table 6**, Cluster 3); this cluster included 3 genes for lectins of Nictaba family (*Lus10031472*, *Lus10031473*, *Lus10031484*) that were upregulated in sXYL samples (see section “Genes for Proteins With Lectin Domains Upregulated in Samples Depositing Secondary Cell Walls”) and 2 genes for lectins of Malectin family (*Lus10038153* and *Lus10041937*) that were upregulated in tFIB samples (see

TABLE 5 | *Linum usitatissimum* genes encoding proteins with lectin domains upregulated in flax stem tissue samples with the tertiary cell wall.

Gene ID	Lectin family	Pfam	Description (according Phytozome)	Close <i>A. thaliana</i> homologs	SUBA	LocTree3	SAM	cPAR	iFIB	tFIB	sXYL
Lus10003099	GNA	PF01453	Curculin-like (mannose-binding) lectin family protein	AT5G18470	extracellular	PM	0	0	0	37	1
Lus10003156		PF01453	lectin protein kinase family protein	AT3G16030	PM	PM	0	0	0	101	9
Lus10009582		PF01453	lectin protein kinase family protein	AT3G16030	PM	PM	0	0	0	730	5
Lus10030766		PF01453	S-locus lectin protein kinase family protein	AT1G11340	PM	PM	3	4	5	209	61
Lus10030768		PF01453	S-locus lectin protein kinase family protein	AT1G11340	PM	PM	1	6	11	45	5
Lus10031591		PF01453	S-locus lectin protein kinase family protein	AT4G03230	PM	PM	9	43	9	706	159
Lus10039762		PF01453	S-locus lectin protein kinase family protein	AT2G19130	PM	PM	0	4	1	80	8
Lus10008597	Legume	PF00139	Concanavalin A-like lectin protein kinase family protein	AT3G53810	PM	PM	2	0	0	53	5
Lus10022662		PF00139	lectin receptor kinase a4.1	AT5G01550	PM	PM	12	14	3	33	12
Lus10032971	Malectin	PF11721	Malectin/receptor-like protein kinase family protein	AT3G51550	PM	PM	159	347	306	1078	355
Lus10032976		PF11721	Malectin/receptor-like protein kinase family protein	AT3G51550	PM	PM	103	92	50	609	165
Lus10038153		PF11721	Leucine-rich repeat transmembrane protein kinase	AT1G53440	PM	PM	44	2056	243	2937	262
Lus10041937		PF11721	Leucine-rich repeat transmembrane protein kinase	AT1G53440	PM	PM	56	1459	64	2222	216
Lus10018304	Nictaba	PF14299	phloem protein 2-A1	AT4G19840	cytosol	chloroplast	29	44	169	372	19
Lus10040602		PF14299	phloem protein 2-A1	AT4G19840	cytosol	nucleus	12	0	34	63	5
Lus10022488	LysM	PF01476	protein kinase family protein/peptidoglycan-binding LysM domain-containing protein	AT2G23770	PM	PM	7	4	5	71	15
Lus10010702	Amaranthin	PF07468	MLP-like protein 423	-	-	cytoplasm	12	1	0	122	13
Lus10029186		PF07468	MLP-like protein 423	-	-	cytoplasm	7	0	0	988	40
Lus10006552	Hevein	PF00187	pathogenesis-related 4	AT3G04720	extracellular	vacuole	0	0	0	180	7
Lus10024290	Jacalin	PF01419	Mannose-binding lectin superfamily protein	AT1G19715	cytosol	secreted	3	8	207	5297	424
Lus10024291		PF01419	Mannose-binding lectin superfamily protein	AT1G19715	cytosol	cytoplasm	0	12	311	967	85
Lus10037605		PF01419	Mannose-binding lectin superfamily protein	AT1G73040	cytosol	cytoplasm	23	14	74	356	35

A pairwise comparison ($\log_2FC \geq 1$, $padj \geq 0.01$) was made with tissue samples that deposit primary and secondary cell walls. The gene description includes the gene ID according to Phytozome v12.1.6 (<https://phytozome.jgi.doe.gov/pz/portal.html>; Goodstein et al., 2012); the lectin family name that was used in present study; the Pfam domain by the presence of which lectin family members were distinguished; the gene description according to Phytozome v12.1.6; *Arabidopsis thaliana* homolog for a flax gene according to flax genome annotation (Wang et al., 2012); subcellular localization of *A. thaliana* homologs for flax lectins according to the SUBA4 database (<https://suba.live/>; Hooper et al., 2017); and subcellular localization of flax lectins (corrected protein sequences presented in **Supplementary File 1** were used) according to the LocTree3 prediction tool (<https://roslab.org/services/loctree3/>; Goldberg et al., 2014). The red-blue heat map indicates that the mRNA abundance expressed in TGR (total gene reads) values and the color gradient reflects TGR values of 1 to 1000. Abbreviations: SAM, shoot apical meristem; cPAR, cortical parenchyma; iFIB, phloem fibers at the stage of intrusive elongation; sXYL, stem xylem with secondary cell wall; tFIB, phloem fibers depositing tertiary cell walls; PM, plasma membrane.





section “Genes for Proteins With Lectin Domains Upregulated in Samples Depositing Tertiary Cell Wall,” **Table 5**).

The expression levels of genes encoding proteins with lectin domains from different families changed substantially after gravistimulation (**Figure 9** and **Table 6**). The expression patterns of lectin genes that differed between various tissues (tFIB and sXYL) and between various time points in the course of gravitropic response development (8 h, 24 h, and 96 h) were revealed.

DISCUSSION

The Set of Proteins Within Some Lectin Families Varies Between Plant Species

The total number of lectin genes in the flax genome, over 400 genes, is among the highest described for a plant species

(Van Holle et al., 2017). The proportion of the number of genes in different lectin families is similar to that described in the literature for other higher plant species (Jiang et al., 2010; Bellande et al., 2017; Van Holle et al., 2017). The most numerous lectin families are GNA-type (112 representatives) and legume-type lectins (72 representatives) that are characteristic for plants and were historically the earliest described (**Table 1**). Some lectin families have quite variable numbers of members between plant species. An extreme example is the amaranthin-like lectin family, which is not ubiquitous among angiosperms. Many taxonomically distant plant species (e.g., *A. thaliana*, soybean, poplar, and rice) do not have amaranthin-like lectins in their genomes, while others (apple, hemp, maize, cucumber, etc.) have amaranthin-like lectins that are represented by multigene families that account for up to 24 members (Dang et al., 2017); flax has one of the largest families (Faruque et al., 2015; **Table 1**). On the other hand, the group of flax jacalin-like lectins consists of only four genes, in contrast

TABLE 6 | Changes in expression of *Linum usitatissimum* genes for proteins with lectin domains during graviresponse.

Gene ID	Description (according Phytosome)	Close <i>A.thaliana</i> homologs	tFIBb	tFIBb_PUL8	tFIBb_OPP8	tFIBb_PUL24	tFIBb_OPP24	tFIBb_PUL96	tFIBb_OPP96	sXYLb	sXYLbPUL8	sXYLbOPP8	sXYLbPUL24	sXYLbOPP24	sXYLbPUL96	sXYLbOPP96	cluster
GNA lectin family (PF01453)																	cluster1
Lus10009582	lectin protein kinase	AT3G16030	637	441	517	295	211	133	221	0	8	8	0	0	0	0	
Lus10028782	receptor kinase 3	AT4G21380	168	181	262	148	177	58	102	18	146	123	55	44	4	3	
Nictaba lectin family (PF14299)																	cluster1
Lus10018303	phloem protein 2 A5	AT1G65390	2	91	318	42	67	3	12	0	8	23	10	8	2	5	
Amaranthin lectin family (PF07468)																	cluster1
Lus10005395		-	189	1408	695	1214	793	190	358	10	229	118	316	181	12	14	
Lus10005397		-	71	495	227	426	293	71	130	7	91	47	122	70	5	5	
Lus10029186	MLP-like protein 423	-	568	343	618	962	647	317	652	40	57	75	39	46	18	45	
Hevein lectin family (PF00187)																	cluster1
Lus10006552	pathogenesis-related 4	AT3G04720	346	686	607	1575	907	194	189	0	163	157	132	136	5	0	
GNA lectin family (PF01453)																	cluster2
Lus10014813	S-locus lectin protein kinase	AT4G27290	10	20	6	36	9	3	9	2	453	242	103	56	36	21	
Lus10029802	S-locus lectin protein kinase	AT2G19130	13	27	41	42	13	11	8	139	1118	889	393	301	122	50	
Lus10036639	receptor-like protein kinase 1	AT5G60900	39	56	66	126	94	54	57	1020	628	538	688	448	1009	557	
Lus10038553	S-locus lectin protein kinase	AT4G27290	0	3	2	0	0	2	2	22	159	109	21	18	21	5	
Lus10038556	receptor kinase 3	AT4G21380	0	4	1	1	1	0	0	15	201	177	88	132	227	186	
Legume lectin family (PF00139)																	cluster2
Lus10033776	receptor lectin kinase	AT2G37710	18	7	5	14	4	12	10	198	244	88	119	65	146	41	
LysM lectin family (PF01476)																	cluster2
Lus10027407	peptidoglycan-binding LysM domain-containing protein	AT3G52790	7	5	8	22	23	17	5	192	171	113	1071	1011	2673	1519	
Lus10031662	peptidoglycan-binding LysM domain-containing protein	AT3G52790	58	54	34	65	66	38	42	893	370	489	649	694	1615	961	
Malectin-like family (PF12819)																	cluster2
Lus10000562	Protein kinase	AT5G61350	1	2	2	3	3	1	1	176	133	118	187	189	613	419	
Lus10025926	Leucine-rich repeat protein kinase	AT5G48740	58	53	53	40	33	27	57	908	1173	981	475	537	594	759	
Lus10038172	Leucine-rich repeat protein kinase	AT5G48740	2	7	6	0	1	0	1	76	106	70	30	18	42	43	

(Continued)

TABLE 6 | Continued

Gene ID	Description (according to Phytozome)	Close <i>A.thaliana</i> homologs	tFIBb	tFIBb_PUL8	tFIBb_opp8	tFIBb_PUL24	tFIBb_opp24	tFIBb_PUL96	tFIBb_opp96	sXYLb	sXYLbPUL8	sXYLbopp8	sXYLbPUL24	sXYLbopp24	sXYLbPUL96	sXYLbopp96	cluster
GNA lectin family (PF01453)																	
Lus10012955	S-locus lectin protein kinase	AT2G19130	240	155	175	6	7	16	19	94	88	83	2	0	5	19	cluster 3
Lus10038554	S-locus lectin protein kinase	AT4G27290	620	362	552	98	123	41	113	90	201	159	15	26	1	1	
Malectin family (PF11721)																	
Lus10004275	Leucine-rich repeat protein kinase	AT2G37050	82	48	63	0	2	1	1	131	77	77	7	6	0	1	cluster 3
Lus10019237	Leucine-rich repeat protein kinase	AT2G37050	83	47	68	0	1	1	1	106	55	62	3	1	0	0	
Lus10038153	Leucine-rich repeat transmembrane protein kinase	AT1G53440	4024	1752	988	206	135	952	1037	546	286	164	20	11	177	289	
Lus10041937	Leucine-rich repeat transmembrane protein kinase	AT1G53440	3884	1140	1837	302	425	498	789	338	408	307	72	66	39	100	
Lus10042505	Leucine-rich repeat transmembrane protein kinase	AT1G53440	3075	372	455	54	81	163	272	277	42	49	0	2	7	44	
Lus10031472	phloem protein 2-B15	AT1G09155	39	118	74	16	9	24	20	649	286	222	12	14	19	20	
Lus10031473	phloem protein 2-B15	AT1G09155	82	272	161	44	21	56	64	1760	764	589	43	41	68	73	
Lus10031484	phloem protein 2-B15	AT1G09155	89	299	176	42	20	58	59	1603	706	548	41	35	67	65	
Malectin-like family (PF12819)																	
Lus10025849	Leucine-rich repeat transmembrane protein kinase	AT4G29990	90	44	42	2	17	8	23	46	34	23	0	0	0	1	
Galactose-binding lectin family (PF02140)																	
Lus10028538	beta-galactosidase 16	AT1G77410	64	49	60	0	0	0	0	46	31	31	0	0	0	0	

The top 30 genes ($TGR \geq 16$ in at least one sample) according to heatmap on **Figure 9**. Gene description includes: the gene ID according to the Phytozome v12.1.6 (<https://phytozome.jgi.doe.gov/pz/portal.html>; Goodstein et al., 2012); the lectin family name that was used in the present study; the Pfam domain by the presence of which lectin family members were distinguished; the gene description according to the Phytozome v12.1.6; *Arabidopsis thaliana* homologs for genes according to flax genome annotation (Wang et al., 2012). The red-blue heat map indicates the mRNA abundance expressed in TGR values and the color gradient reflects TGR (total gene reads) values from 1 to 1000. Abbreviations: SCW, secondary cell wall; TCW, tertiary cell wall; PM, plasma membrane. tFIBb and sXYLb – isolated fibers with TCW and stem xylem with SCW of non-inclined plant, respectively; tFIBb_PUL8, tFIBb_PUL24, tFIBb_PUL96 – isolated fibers with TCW from pulling side of flax stem after 8, 24, 96 h of gravibending, respectively; tFIBb_opp8, tFIBb_opp24, tFIBb_opp96 – isolated fibers with TCW from opposite side of flax stem after 8, 24, 96 h of gravibending, respectively; sXYLb_PUL8, sXYLb_PUL24, sXYLb_PUL96 – pulling side of stem xylem with SCW after 8, 24, 96 h of gravibending, respectively; sXYLb_opp8, sXYLb_opp24, sXYLb_opp96 – opposite side of stem xylem with SCW after 8, 24, 96 h of gravibending. Genes that were considered as differentially expressed (Tables 4, 5) in the tissues under normal condition are given on a gray background.

to the several dozen representatives in the genomes of rice and *A. thaliana*. In *A. thaliana*, the jacalin family is among the largest, comprising almost a quarter of all lectins (Van Holle et al., 2017).

In the phylogenetic trees of lectin families, large groups of members from different plant species may be fully separated. The large clade of 35 malectin-like proteins in *A. thaliana*, which includes MEE39 and PAM74, does not have a flax counterpart (Figure 3). Similarly, flax is lacking some large clades of legume-type lectins (Figure 4). The distribution of jacalins present in *A. thaliana* and rice genomes on the phylogenetic tree is almost mutually exclusive (Van Holle et al., 2017, Figures 8A,B); flax jacalins are localized only within a small group that is mutual for all three species. The divergent sets of proteins within several lectin families indicate that these proteins are among those determining the specificity of development and reaction to biotic and abiotic stressors between plant species.

Proteins With Lectin Domains Are Widely Involved in the Life of Plants

The expression of proteins with lectin domains in different tissues of the same organ was analyzed using flax stem. Three quarters of proteins with lectin domains present in the flax genome were expressed in stem tissues (TGR > 50), indicating the wide involvement of lectins in the everyday life of plant organisms (Table 1). In several lectin families that had a small number of members, such as calreticulins and proteins with C-type and EUL domains, all members were expressed in all analyzed tissues, suggesting the involvement in basic processes. Calreticulins, for example, are located in endoplasmic reticulum and ensure proper folding and quality control of the synthesized secretory and membrane glycoproteins before exiting the endoplasmic reticulum (Lannoo and Van Damme, 2014). Unlike classical chaperones, which interact with the peptide part of the substrates, CANX and CALR bind to oligosaccharide chains of glycoproteins. A high proportion of the expressed genes but with variable expression levels in different samples was revealed for most of the other lectin families (Supplementary File 3), including proteins with GNA-type, legume-type, LysM, malectin, Nictaba, galactose-binding lectin, and galectin-like domains.

The function of lectins is often associated with a plant's reaction to pathogens (De Hoff et al., 2009; Vandenborre et al., 2011). However, plant immunity and development often rely on similar or overlapping cellular mechanisms for signal recognition and transduction (Jamieson et al., 2018). A genome-wide analysis of all flax proteins with lectin domains demonstrates that lectins are actively and differentially expressed both under normal growth conditions and under abiotic stress conditions including during gravitropic response. The expression of flax lectins is specific for various tissues and also for the different stages of cell development such as in phloem fibers at intrusive elongation and tertiary cell wall deposition stages (Supplementary File 3). Though the flax plants sampled in this study were grown under non-sterile conditions and it cannot be excluded that the expression of some lectins was associated with the reaction to microorganisms, the high proportion of transcribed genes for

lectins, their differential expression in various tissues and at various stages of the distinct cell type development, together with the dynamic changes of expression under the influence of abiotic factors suggest that proteins with lectin domains are widely involved in plant development, tissue specialization, and reaction to the effect of abiotic stressors.

Proteins with the same lectin domain can be localized within different subcellular structures and fulfill different functions. Most of the lectins are chimerolectins, i.e., proteins with additional domains (Bellande et al., 2017; Eggermont et al., 2017). The combinations of various domains in a protein are variable, especially considering proteins from organisms in different kingdoms. For example, there are vast differences in the proteins with galectin-like domains (PF00337). In animals, the galectins are small, mainly extracellular proteins secreted by a non-classical exocytic pathway; they contribute to cell-cell and cell-matrix interactions due to the processes of multivalent carbohydrate recognition combined with galectin di- and oligomerization (Dodd and Drickamer, 2001; Nabi et al., 2015). In plants, the same characteristic domain is present within several galactosyltransferases, like in *A. thaliana* and flax (Figure 5). According to the current definitions, if taken strictly, the same domain should be considered lectin in animals and CBM in plants. Moreover, differences can be detected even within a family. In *A. thaliana*, the protein AT4G19810 from the class V chitinase-related agglutinin (CRA) family possesses chitinase activity and cannot be referred to as a lectin by definition (Eggermont et al., 2017). Lectins are primarily searched for based on corresponding protein domains, especially in full genomic studies (Bellande et al., 2017; Eggermont et al., 2017, current study); however, this method cannot guarantee the absence of enzymatic activity in protein. This complicates the identification and definition of plant lectins, and the lists of plant lectin families in classifications suggested by various authors do not fully match (Table 1). In our study, we have considered all proteins with lectin domains, since the presence of a carbohydrate-recognizing domain in a protein may add some important features to its function.

Numerous Lectins Have the Potential to Interact With Cell Wall Glycans

Many plant lectins function in the cytoplasm and nucleus, whereas others are exposed to the cell surface because they are secreted into the cell wall or anchored to the plasma membrane and have an extracellular part of the protein. The interaction of some lectins with the cell wall or with certain cell wall glycans was demonstrated by various approaches. The direct association with polygalacturonic acid could be identified for plasma membrane receptor-like kinase with the malectin domain (Feng et al., 2018). The association with cell wall components of another transmembrane protein that has a malectin domain but no kinase domain was demonstrated at the analysis of plasmolyzed cells in lateral roots of *A. thaliana* (Schürholz, 2019). Similar experiments showed that GFP-tagged THE1 and FER are also tightly bound to the cell wall (Hématy et al., 2007; Li et al., 2018). The lectin receptor kinase with a legume-type domain

is involved in maintaining cell wall-plasma membrane adhesion through protein–protein interactions with the participation of the tripeptide motifs Arg-Gly-Asp (RGD) (Gouget et al., 2006). Lectins of various architectural types, such as LecP, LecRLPs, and LecRLKs, are readily found by proteomic approaches both in the cell wall and plasma membrane (Jamet et al., 2008; Bellande et al., 2017). The secretion of some lectins without signal peptides into the cell wall via non-classical way was demonstrated for animal galectins (Delacour et al., 2009) and plant proteins with EUL (Jamet et al., 2008; Dubiel et al., 2020) and jacalin-like (Pinedo et al., 2012, 2015) domains.

The hierarchical clustering grouped all analyzed samples according to the cell wall type (**Figure 2A**). Phloem fibers at the intrusive elongation stage were grouped together with other tissues depositing primary cell wall rather than together with phloem fibers at the later stage of development when they deposit tertiary cell wall. The majority of genes upregulated in tissues with distinct cell wall types encode lectins localized at the plasma membrane or secreted into the apoplast as predicted by LocTree3 service (**Tables 3–5**). Thus, the encoded proteins have the potential to directly interact with cell wall glycans.

A characteristic feature of lectins is di- or tetramerization that may involve homo- and hetero-interactions (Müller et al., 2016). Lectins of various architectural types can take part in such interactions. For example, RLPs were suggested to function as a specificity switch for ligand-receptor recognition (Jamieson et al., 2018). In addition, interactions of proteins with lectin domains depend on ligand binding. Glycan sequences characteristic to certain cell wall types and lectins that are specifically upregulated and/or are present in tissues with various cell wall types may provide the platform on which to arrange specific signaling complexes. This is indicated by the comprehensive analysis of all proteins with lectin domains that reveals the specific expression patterns coupled to cell wall type in various flax stem tissues (**Tables 3–5**).

Differentially Expressed Flax Lectins Between Samples With Distinct Cell Wall Types

Proteins with lectin domains that are differentially expressed in tissues with distinct cell wall types belong to various lectin families (**Tables 3–5**). GNA-type lectins comprise the largest lectin family in flax (**Table 1**), same as in many other plant species (Van Holle et al., 2017). GNA-type lectins have a carbohydrate-binding domain of approximately 150 amino acids, which binds D-mannose and has a conserved mannose-binding motif Q-X-D-X-N-X-V-X-Y (Van Damme et al., 2008; Shimokawa et al., 2012). Genes for GNA-type lectins were actively expressed in flax stem tissues (**Table 1**) but were rather poorly related to cell wall type, since their numbers among the differentially expressed tissues were low (**Tables 3–5**). Most of the GNA-type lectins have expression patterns that do not overcome the thresholds used in the current paper and are marked by gray dots in **Supplementary File 4**. No representative of this family was upregulated in tissues depositing primary cell wall. However, several lectins were detected for the tertiary cell wall. *Lus10009582* and *Lus10031591*,

which encode the LecRLKs with the highest expression in fibers among GNA-type lectins, were quite specifically expressed in tFIB samples (**Table 5**). Several lectins with the LysM motif were upregulated in sXYL samples in the secondary cell wall deposition stage (**Table 4**). These lectins include proteins with the LecRLK type of architecture that were localized to the plasma membrane, namely, *Lus10023945* (homologous to AT2G17120 (LYM2) and *Lus10008586* (AT2G33580, LYK5), which belong to distant clades, and three LecRLP type lectins predicted to be secreted into the cell wall, *Lus10031662*, *Lus10027407*, and *Lus10015011*, which all cluster together in the phylogenetic tree (**Figure 7**).

The legume-type lectin family had representatives upregulated in tissues with each cell wall type. LecRK from this family was differentially expressed in samples with primary, secondary, and tertiary cell walls. For example, *Lus10010867* and *Lus10024365*, which are homologous to AT3G555530, encoding the legume-type lectin receptor kinase LECRKS7 (Bouwmeester and Govers, 2009), were upregulated in tissues with primary cell walls (**Table 3**). The set of genes for legume-type lectin receptor kinases were activated in sXYL samples with secondary cell wall (**Table 4**). All of these kinases were predicted to be localized at the plasma membrane. However, information on the different properties of distinct family representatives is scarce, which is also true for many proteins with lectin domains from other families. This limits the discussion of the functional relevance of the changes in gene expression among lectins.

The most differential expression patterns among the proteins with lectin domains in samples with distinct cell wall types were observed for proteins with jacalin-like and malectin (together with malectin-like) domains, as well as for β -galactosidases with lectin domains. Out of four genes for proteins with jacalin-like domain, three were specifically upregulated in fibers depositing tertiary cell wall (**Table 5**). The fourth gene, distinguished by the presence of F-box (**Figure 8B**), was expressed in all analyzed tissues (**Table 2**). The encoded proteins do not have signal peptides and are considered nucleocytoplasmic mannose-binding proteins (Van Holle and Van Damme, 2019); however, research supports cell wall localization and secretion of such proteins via a non-classical pathway (Grunwald et al., 2007; Pinedo et al., 2012, 2015). Flax jacalin encoded by *Lus10024290* was predicted to be secreted into the cell wall (**Table 5**).

Both malectin and malectin-like domains are found in proteins of organisms from various kingdoms; however, plants have many more malectin, malectin-like, legume, and GNA-type domains compared with organisms from other kingdoms (Bellande et al., 2017). In animals, malectin is a well-characterized membrane-anchored endoplasmic reticulum protein that recognizes and binds Glc2-*N*-glycan, thus playing a role in the early stages of protein *N*-glycosylation (Schallus et al., 2008). In plants, this domain is found in a number of receptor kinases localized to the plasma membrane (Franck et al., 2018), indicating that malectin domains have different functions in different kingdoms. The expression patterns of various genes for malectin and malectin-like domains in flax tissues are very diverse (**Tables 2–5**). Genes for four flax malectins, which are homologs of *A. thaliana* genes for di-glucose binding protein

with a kinesin motor domain, grouped as a separate clade in the malectin family (**Figure 3**). All of them were upregulated in tissues with primary cell wall. Two other malectins, Lus10032971 and Lus10032976 are homologous to FERONIA (AT3G51550) – a receptor-like kinase that was demonstrated to interact with cell wall pectins and to be involved in numerous developmental processes (Li et al., 2016). In *A. thaliana* seedlings, FERONIA was identified as a key regulator in mechano-sensing (Shih et al., 2014; Doblás et al., 2018). In flax, the expression of FERONIA homologs was highly up-regulated in fibers depositing tertiary cell wall. This cell wall type is deposited only in fibers and is considered to have a special mechanical function by providing tension in cellulose microfibrils (Gorshkova et al., 2018a; Almeras et al., 2020). The presence of specific types of pectins is characteristic for tertiary cell walls (Mikshina et al., 2013). Thus, the enhanced expression of FERONIA-like mechano-sensors is very relevant to the composition and function of fibers with tertiary cell wall.

Several β -galactosidases with a lectin domain had expression patterns coupled to the cell wall type; for example, four of them were upregulated in tissues with primary cell wall. All plant β -galactosidases belong to GH35 in CAZy database (Chandrasekar and van der Hoorn, 2016). In flax, among the proteins with the Glyco_hydro_35 (PF01301) domain, the main catalytic domain of GH-35 β -galactosidases, only a portion also had the PF02140 domain (24 out of 43). Similarly, this was observed in jute (only six out of 11 β -galactosidases had galactose-binding lectin domains; Satya et al., 2018), tomato, and *A. thaliana* (12 out of 17 and 10 out of 17, respectively; Chandrasekar and van der Hoorn, 2016), while the lectin domain was present in all 17 β -galactosidases detected in peach (Guo et al., 2018). The β -galactosidases (*Lus10028848* and *Lus10008974*) that are known as key players in the tertiary cell wall structure of flax fibers (Roach et al., 2011; Gorshkova et al., 2018a) are devoid of the PF02140 domain. This domain, if present, is a C-terminal domain that is homologous to galactose- and rhamnose-binding animal lectins, which is designated a SUEL (Sea Urchin Egg Lectin)-type carbohydrate-binding domain. The structure and function of this domain in rice β -galactosidase, OsBGal1, was studied in detail (Rimlumduan et al., 2016). Although the binding of galactose and rhamnose was predicted for OsBGal1 based on amino acid sequence homology with SUEL lectin, binding to rhamnose, galactose, glucose, β -1,4-galactobiose, and raffinose was not observed in NMR experiments. Thus, experimental identification of carbohydrate specificity of the PF02140 domain is still required, as is the identification of carbohydrate specificity for the vast majority of plant lectins. Moreover, some lectins, like Nictaba, have many promiscuous carbohydrate binding sites and are capable of interacting with different carbohydrate motifs (Delporte et al., 2015).

Altogether, the expression of many proteins with lectin domains is related to the formation of the carbohydrate-enriched compartment of plant cell – the wall. Moreover, distinct cell wall types have both shared and “personal” cell wall-related proteins with lectin domains, the expression of the latter is highly upregulated at the formation of PCW, SCW, or TCW. This can

be coupled to the peculiarities of glycan composition in various cell wall types. Thus, in-depth further studies of lectin specificity in relation to plant cell wall polymers are highly demanded. The distinct combinations of lectins and complex carbohydrates may give rise to specific regulatory modes characteristic for certain cell wall types. The diversity of the specific recognition systems on the cell surface of various tissues may enrich the complex regulation of plant organism development and reaction to abiotic stress.

DATA AVAILABILITY STATEMENT

The 40 flax transcriptome libraries considered in this study were previously obtained by us and deposited in the Sequence Read Archive (SRA) as BioProjects (PRJNA475325 and PRJNA631357). The four RNA-seq datasets from the stem apex (PRJNA229810, Zhang and Deyholos, 2016) were downloaded from the European Nucleotide Archive (<https://www.ebi.ac.uk/ena>) as fastq-files of raw data.

AUTHOR CONTRIBUTIONS

NP, OP, and TG thought up and designed the study. OG, AN, NP, NM, and TG performed the experiments analyzed the obtained data and jointly discussed them. TG and NP wrote the manuscript. The illustrations and tables were prepared by NP (**Figure 1** and **Tables 1–6, Supplementary Files 1, 2**), OG (**Figures 2, 9, Tables 2–6, Supplementary File 3**), AN (**Figures 4–8, Supplementary Files 1, 4**), and NM (**Supplementary Files 1, 2**). All authors reviewed and approved the manuscripts.

FUNDING

The study was partially funded by the Russian Scientific Foundation, project # 20-64-47036 (TG, OP, transcriptome analysis). We also acknowledge financial support from the government assignment for FRC Kazan Scientific Center of RAS (NP, AN, OG, NM, and TG, genome analysis and phylogenetic analysis).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.634594/full#supplementary-material>

Supplementary File 1 | Amino acid sequences of analyzed flax proteins with lectin domains.

Supplementary File 2 | Validation of RNA-Seq data by qRT-PCR analysis.

Supplementary File 3 | RNA-Seq data for flax genes encoding proteins with lectin domains.

Supplementary File 4 | Phylogenetic dendrogram of GNA (PF01453) lectin family members of *Arabidopsis thaliana* and *Linum usitatissimum*.

REFERENCES

- Albenne, C., Canut, H., Hoffmann, L., and Jamet, E. (2014). Plant cell wall proteins: a large body of data, but what about runaways? *Proteomes* 2, 224–242. doi: 10.3390/proteomes2020224
- Almeras, T., Petrova, A., Kozlova, L., Gril, J., and Gorshkova, T. (2020). Evidence and quantitative evaluation of tensile maturation strain in flax phloem through longitudinal splitting. *Botany* 98, 9–19. doi: 10.1139/cjb-2019-0021
- Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z
- Bannai, H., Tamada, Y., Maruyama, O., Nakai, K., and Miyano, S. (2002). Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics* 18, 298–305. doi: 10.1093/bioinformatics/18.2.298
- Barre, A., Bourne, Y., Van Damme, E. J. M., and Rougé, P. (2019). Overview of the structure-function relationships of mannose-specific lectins from plants, algae and fungi. *Int. J. Mol. Sci.* 20:254. doi: 10.3390/ijms20020254
- Bellande, K., Bono, J.-J., Savelli, B., Jamet, E., and Canut, H. (2017). Plant lectins and lectin receptor-like kinases: how do they sense the outside? *Int. J. Mol. Sci.* 18:1164. doi: 10.3390/ijms18061164
- Beneteau, J., Renard, D., Marché, L., Douville, E., Rahbé, Y., et al. (2010). Binding properties of the N-acetylglucosamine and high-mannose N-glycan PP2-A1 phloem lectin in *Arabidopsis*. *Plant Physiol.* 153, 1345–1361. doi: 10.1104/pp.110.153882
- Boisson-Dernier, A., Kessler, S. A., and Grossniklaus, U. (2011). The walls have ears: the role of plant CrRLK1Ls in sensing and transducing extracellular signals. *J. Exp. Bot.* 62, 1581–1591. doi: 10.1093/jxb/erq445
- Bouwmeester, K., and Govers, F. (2009). *Arabidopsis* L-type lectin receptor kinases: phylogeny, classification, and expression profiles. *J. Exp. Bot.* 60, 4383–4396. doi: 10.1093/jxb/erp277
- Bygdell, J., Srivastava, V., Obudulu, O., Srivastava, M. K., Nilsson, R., Sundberg, B., et al. (2017). Protein expression in tension wood formation monitored at high tissue resolution in *Populus*. *J. Exp. Bot.* 68, 3405–3417. doi: 10.1093/jxb/erx186
- Carpita, N. C., and Gibeau, D. M. (1993). Structural models of primary-cell walls in flowering plants - consistency of molecular-structure with the physical-properties of the walls during growth. *Plant J.* 3, 1–30. doi: 10.1111/j.1365-313x.1993.tb00007.x
- Chandrasekar, B., and van der Hoorn, R. A. L. (2016). Beta galactosidases in *Arabidopsis* and tomato-a mini review. *Biochem. Soc. Trans.* 44, 150–158. doi: 10.1042/BST20150217
- Chen, Y., Peumans, W. J., Hause, B., Bras, J., Kumar, M., Proost, P., et al. (2002). Jasmonate methyl ester induces the synthesis of a cytoplasmic/nuclear chito oligosaccharide-binding lectin in tobacco leaves. *FASEB J.* 16, 905–907. doi: 10.1096/fj.01-0598fje
- Dang, L., Rougé, P., and Van Damme, E. J. M. (2017). Amaranthin-like proteins with aerolysin domains in plants. *Front. Plant Sci.* 8:1368. doi: 10.3389/fpls.2017.01368
- De Hoff, P. L., Brill, L. M., and Hirsch, A. M. (2009). Plant lectins: the ties that bind in root symbiosis and plant defense. *Mol. Genet. Genom.* 282, 1–15. doi: 10.1007/s00438-009-0460-8
- Delacour, D., Koch, A., and Jacob, R. (2009). The role of galectins in protein trafficking. *Traffic* 10, 1405–1413. doi: 10.1111/j.1600-0854.2009.00960.x
- Delporte, A., De Zaeytijd, J., De Storme, N., Azmi, A., Geelen, D., Smaghe, G., et al. (2014). Cell cycle-dependent O-GlcNAc modification of tobacco histones and their interaction with the tobacco lectin. *Plant Physiol. Biochem.* 83, 151–158. doi: 10.1016/j.plaphy.2014.07.021
- Delporte, A., Van Holle, S., Lannoo, N., and Van Damme, J. M. E. (2015). The tobacco lectin, prototype of the family of Nictaba-related proteins. *Curr. Protein Pept. Sci.* 16, 5–16. doi: 10.2174/1389203716666150213154107
- Doblas, V. G., Gonneau, M., and Höfte, H. (2018). Cell wall integrity signaling in plants: malectin-domain kinases and lessons from other kingdoms. *Cell Surf.* 3:1. doi: 10.1016/j.tcsu.2018.06.001
- Dodd, R. B., and Drickamer, K. (2001). Lectin-like proteins in model organisms: implications for evolution of carbohydrate-binding activity. *Glycobiology* 11, 71R–79R.
- Duan, Q., Kita, D., Li, C., Cheung, A. Y., and Wu, H.-M. (2010). FERONIA receptor-like kinase regulates RHO GTPase signaling of root hair development. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17821–17826. doi: 10.1073/pnas.1005366107
- Dubiel, M., De Coninck, T., Osterne, V. J. S., Verbeke, I., Van Damme, D., Smaghe, G., et al. (2020). The ArathEULS3 Lectin ends up in stress granules and can follow an unconventional route for secretion. *Int. J. Mol. Sci.* 21:1659. doi: 10.3390/ijms21051659
- Eggermont, L., Verstraeten, B., and Van Damme, E. J. M. (2017). Genome-wide screening for lectin motifs in *Arabidopsis thaliana*. *Plant Genome* 10, 1–17.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi: 10.1093/nar/gky995
- Esau, K. (1965). *Plant Anatomy*. Sydney: John Wiley & Sons, Inc.
- Faruque, K., Begam, R., and Deyholos, M. K. (2015). The amaranthin-like lectin (LuALL) genes of flax: a unique gene family with members inducible by defence hormones. *Plant Mol. Biol. Rep.* 33, 731–741. doi: 10.1007/s11105-014-0791-4
- Feng, W., Kita, D., Peaucelle, A., Cartwright, H. N., Doan, V., Duan, Q., et al. (2018). The FERONIA receptor kinase maintains cell-wall integrity during salt stress through Ca²⁺ signaling. *Curr. Biol.* 28, 666–675. doi: 10.1016/j.cub.2018.01.023
- Franck, C. M., Westermann, J., and Boisson-Dernier, A. (2018). Plant malectin-like receptor kinases: from cell wall integrity to immunity and beyond. *Annu. Rev. Plant Biol.* 69, 301–328.
- Galindo-Trigo, S., Grand, T. M., Voigt, C. A., and Smith, L. M. (2020). A malectin domain kinesin functions in pollen and seed development in *Arabidopsis*. *J. Exp. Bot.* 71, 1828–1841. doi: 10.1093/jxb/era023
- Goldberg, T., Hecht, M., Hamp, T., Karl, T., Yachdav, G., Ahmed, N., et al. (2014). LocTree3 prediction of localization. *Nucleic Acids Res.* 42, W350–W355.
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944
- Gorshkov, O., Chernova, T., Mokshina, N., Gogoleva, N., Suslov, D., Tkachenko, A., et al. (2019). Intrusive growth of phloem fibers in flax stem: integrated analysis of miRNA and mRNA expression profiles. *Plants* 8:47. doi: 10.3390/plants8020047
- Gorshkov, O., Mokshina, N., Ibragimova, N., Ageeva, M., Gogoleva, N., and Gorshkova, T. (2018). Phloem fibres as motors of gravitropic behaviour of flax plants: level of transcriptome. *Funct. Plant Biol.* 45, 203–214. doi: 10.1071/fp16348
- Gorshkova, T., Chernova, T., Mokshina, N., Ageeva, M., and Mikshina, P. (2018a). Plant ‘muscles’: fibers with a tertiary cell wall. *New Phytol.* 218, 66–72. doi: 10.1111/nph.14997
- Gorshkova, T., Chernova, T., Mokshina, N., Gorshkov, V., Kozlova, L., and Gorshkov, O. (2018b). Transcriptome analysis of intrusively growing flax fibers isolated by laser microdissection. *Sci. Rep.* 8, 1–17.
- Gorshkova, T. A., Sal’nikov, V. V., Chemiksova, S. B., Ageeva, M. V., Pavlencheva, N. V., and van Dam, J. E. G. (2003). The snap point: a transition point in *Linum usitatissimum* bast fiber development. *Ind. Crops Prod.* 18, 213–221. doi: 10.1016/S0926-6690(03)00043-8
- Gouget, A., Senchou, V., Govers, F., Sanson, A., Barre, A., Rougé, P., et al. (2006). Lectin receptor kinases participate in protein-protein interactions to mediate plasma membrane-cell wall adhesions in *Arabidopsis*. *Plant Physiol.* 140, 81–90. doi: 10.1104/pp.105.066464
- Grunwald, I., Heinig, I., Thole, H. H., Neumann, D., Kahmann, U., Kloppstech, K., et al. (2007). Purification and characterization of a jacalin-related, coleoptile specific lectin from *Hordeum vulgare*. *Planta* 226:225. doi: 10.1007/s00425-006-0467-x
- Guo, H., Li, L., Ye, H., Yu, X., Algreen, A., and Yin, Y. (2009). Three related receptor-like kinases are required for optimal cell elongation in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7648–7653. doi: 10.1073/pnas.0812346106
- Guo, S., Song, J., Zhang, B., Jiang, H., Ma, R., and Yu, M. (2018). Genome-wide identification and expression analysis of beta-galactosidase family members during fruit softening of peach [*Prunus persica* (L.) Batsch]. *Postharvest Biol. Technol.* 136, 111–123. doi: 10.1016/j.postharvbio.2017.10.005
- He, Y., Zhou, J., Shan, L., and Meng, X. (2018). Plant cell surface receptor-mediated signaling-a common theme amid diversity. *J. Cell Sci.* 131, jcs209353. doi: 10.1242/jcs.209353

- Hématy, K., Sado, P.-E., Van Tuinen, A., Rochange, S., Desnos, T., Balzergue, S., et al. (2007). A receptor-like kinase mediates the response of *Arabidopsis* cells to the inhibition of cellulose synthesis. *Curr. Biol.* 17, 922–931. doi: 10.1016/j.cub.2007.05.018
- Henrissat, B., and Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* 7, 637–644. doi: 10.1016/s0959-440x(97)80072-3
- Hobson, N., and Deyholos, M. K. (2013). Genomic and expression analysis of the flax (*Linum usitatissimum*) family of glycosyl hydrolase 35 genes. *BMC Genom.* 14:344. doi: 10.1186/1471-2164-14-344
- Hooper, C. M., Castleden, I. R., Tanz, S. K., Aryamanesh, N., and Millar, A. H. (2017). SUBA4: the interactive data analysis centre for *Arabidopsis* subcellular protein locations. *Nucleic Acids Res.* 45, D1064–D1074.
- Hwang, I. S., and Hwang, B. K. (2011). The pepper mannosyl-binding lectin gene CaMBL1 is required to regulate cell death and defense responses to microbial pathogens. *Plant Physiol.* 155, 447–463. doi: 10.1104/pp.110.164848
- Ibragimova, N. N., Ageeva, M. V., and Gorshkova, T. A. (2017). Development of gravitropic response: unusual behavior of flax phloem G-fibers. *Protoplasma* 254, 749–762. doi: 10.1007/s00709-016-0985-8
- Jamet, E., Albenne, C., Boudart, G., Irshad, M., Canut, H., and Pont-Lezica, R. (2008). Recent advances in plant cell wall proteomics. *Proteomics* 8, 893–908. doi: 10.1002/pmic.200700938
- Jamieson, P. A., Shan, L., and He, P. (2018). Plant cell surface molecular cypher: receptor-like proteins and their roles in immunity and development. *Plant Sci.* 274, 242–251. doi: 10.1016/j.plantsci.2018.05.030
- Jiang, J., Xu, Y., and Chong, K. (2007). Overexpression of OsJAC1, a lectin gene, suppresses the coleoptile and stem elongation in rice. *J. Integr. Plant Biol.* 49, 230–237. doi: 10.1111/j.1744-7909.2007.00428.x
- Jiang, S.-Y., Ma, Z., and Ramachandran, S. (2010). Evolutionary history and stress regulation of the lectin superfamily in higher plants. *BMC Evol. Biol.* 10:79. doi: 10.1186/1471-2148-10-79
- Kalyanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Kim, D., Landmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–U121. doi: 10.1038/Nmeth.3317
- Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315
- Kryuchkova-Mostacci, N., and Robinson-Rechavi, M. (2016). Tissue-specificity of gene expression diverges slowly between orthologs, and rapidly between paralogs. *PLoS Comput. Biol.* 12:e1005274. doi: 10.1371/journal.pcbi.1005274
- Lannoo, N., and Van Damme, E. J. M. (2010). Nucleocytoplasmic plant lectins. *Biochim. Biophys. Acta Gen. Subj.* 1800, 190–201. doi: 10.1016/j.bbagen.2009.07.021
- Lannoo, N., and Van Damme, E. J. M. (2014). Lectin domains at the frontiers of plant defense. *Front. Plant Sci.* 5:397. doi: 10.3389/fpls.2014.00397
- Lee, J. H., Takei, K., Sakakibara, H., Cho, H. S., Kim, Y. S., Min, S. R., et al. (2003). CHRK1, a chitinase-related receptor-like kinase, plays a role in plant development and cytokinin homeostasis in tobacco. *Plant Mol. Biol.* 53, 877–890. doi: 10.1023/b:plan.0000023668.34205.a8
- Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239
- Li, C., Liu, X., Qiang, X., Li, X., Li, X., Zhu, S., et al. (2018). EBPI nuclear accumulation negatively feeds back on FERONIA-mediated RALF1 signaling. *PLoS Biol.* 16:e2006340. doi: 10.1371/journal.pbio.2006340
- Li, C., Wu, H.-M., and Cheung, A. Y. (2016). FERONIA and her pals: functions and mechanisms. *Plant Physiol.* 171, 2379–2392. doi: 10.1104/pp.16.00667
- Love, M. I., Anders, S., Kim, V., and Huber, W. (2015). RNA-Seq workflow: gene-level exploratory analysis and differential expression. *F1000Research* 4:1070. doi: 10.12688/f1000research.7035.2
- Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., et al. (2020). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* 48, D265–D268.
- Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 47, W636–W641.
- Mikshina, P., Chernova, T., Chemiksova, S., Ibragimova, N., Mokshina, N., and Gorshkova, T. (2013). “Cellulosic fibers: role of matrix polysaccharides in structure and function,” in *Cellulose - Fundamental Aspects*, eds T. van de Ven, and L. Godbout, (London: IntechOpen), 91–112.
- Minh, B. Q., Nguyen, M. A. T., and von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195. doi: 10.1093/molbev/mst024
- Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., et al. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* 47, D351–D360.
- Mokshina, N., Gorshkov, O., Galinovsky, D., and Gorshkova, T. (2020). Genes with bast fiber-specific expression in flax plants-Molecular keys for targeted fiber crop improvement. *Ind. Crops Prod.* 152:112549. doi: 10.1016/j.indcrop.2020.112549
- Müller, C., Despras, G., and Lindhorst, T. K. (2016). Organizing multivalency in carbohydrate recognition. *Chem. Soc. Rev.* 45, 3275–3302. doi: 10.1039/c6cs00165c
- Nabi, I. R., Shankar, J., and Dennis, J. W. (2015). The galectin lattice at a glance. *J. Cell Sci.* 128, 2213–2219. doi: 10.1242/jcs.151159
- Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Nilsson, R., Bernfur, K., Gustavsson, N., Bygdell, J., Wingsle, G., and Larsson, C. (2010). Proteomics of plasma membranes from poplar trees reveals tissue distribution of transporters, receptors, and proteins in cell wall formation. *Mol. Cell. Proteom.* 9, 368–387. doi: 10.1074/mcp.m900289-mcp200
- Perteau, M., Kim, D., Perteau, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 11, 1650–1667. doi: 10.1038/nprot.2016.095
- Peumans, W. J., and Van Damme, E. J. (1995). Lectins as plant defense proteins. *Plant Physiol.* 109:347. doi: 10.1104/pp.109.2.347
- Pinedo, M., Orts, F., de Oliveira Carvalho, A., Regente, M., Soares, J. R., Gomes, V. M., et al. (2015). Molecular characterization of Helja, an extracellular jacalin-related protein from *Helianthus annuus*: insights into the relationship of this protein with unconventionally secreted lectins. *J. Plant Physiol.* 183, 144–153. doi: 10.1016/j.jplph.2015.06.004
- Pinedo, M., Regente, M., Elizalde, M., Quiroga, Y., Pagnussat, I. A., Jorrín-Novo, J. L., et al. (2012). Extracellular sunflower proteins: evidence on non-classical secretion of a jacalin-related lectin. *Protein Pept. Lett.* 19, 270–276. doi: 10.2174/092986612799363163
- R Core Team, (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rimluduan, T., Hua, Y., Tanaka, T., and Cairns, J. R. K. (2016). Structure of a plant β -galactosidase C-terminal domain. *Biochim. Biophys. Acta Proteom.* 1864, 1411–1418. doi: 10.1016/j.bbapap.2016.07.005
- Roach, M. J., Mokshina, N. Y., Badhan, A., Snegireva, A. V., Hobson, N., Deyholos, M. K., et al. (2011). Development of cellulosic secondary walls in flax fibers requires β -galactosidase. *Plant Physiol.* 156, 1351–1363. doi: 10.1104/pp.111.172676
- Satya, P., Chakraborty, A., Sarkar, D., Karan, M., Das, D., Mandal, N. A., et al. (2018). Transcriptome profiling uncovers beta-galactosidases of diverse domain classes influencing hypocotyl development in jute (*Corchorus capsularis* L.). *Phytochemistry* 156, 20–32. doi: 10.1016/j.phytochem.2018.08.017
- Schallus, T., Jaeckh, C., Fehér, K., Palma, A. S., Liu, Y., Simpson, J. C., et al. (2008). Malectin: a novel carbohydrate-binding protein of the endoplasmic reticulum and a candidate player in the early steps of protein N-glycosylation. *Mol. Biol. Cell* 19, 3404–3414. doi: 10.1091/mbc.e08-04-0354
- Schürholz, A.-K. (2019). *Spatio-Temporal Control of Cell Wall Properties and Signaling Networks in Arabidopsis Meristems*. Master's thesis, Ruperto Carola University Heidelberg, Heidelberg.
- SEQC/MAQC-III Consortium, (2014). A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by thesequencing quality control consortium. *Nat. Biotechnol.* 32, 903–914. doi: 10.1038/nbt.2957
- Sharon, N. (2008). Lectins: past, present and future. *Biochem. Soc. Trans.* 36, 1457–1460. doi: 10.1042/bst0361457

- Shih, H.-W., Miller, N. D., Dai, C., Spalding, E. P., and Monshausen, G. B. (2014). The receptor-like kinase FERONIA is required for mechanical signal transduction in *Arabidopsis* seedlings. *Curr. Biol.* 24, 1887–1892. doi: 10.1016/j.cub.2014.06.064
- Shinya, T., Motoyama, N., Ikeda, A., Wada, M., Kamiya, K., Hayafune, M., et al. (2012). Functional characterization of CEBiP and CERK1 homologs in *Arabidopsis* and rice reveals the presence of different chitin receptor systems in plants. *Plant Cell Physiol.* 53, 1696–1706. doi: 10.1093/pcp/pcs113
- Shimokawa, M., Fukudome, A., Yamashita, R., Minami, Y., Yagi, F., Tateno, H., et al. (2012). Characterization and cloning of GNA-like lectin from the mushroom *Marasmius oreades*. *Glycoconj. J.* 29, 457–465. doi: 10.1007/s10719-012-9401-6
- Showalter, A. M., and Basu, D. (2016). Extensin and arabinogalactan-protein biosynthesis: glycosyltransferases, research challenges, and biosensors. *Front. Plant Sci.* 7:814. doi: 10.3389/fpls.2016.00814
- Solovyyev, V., Kosarev, P., Seledsov, I., and Vorobyev, D. (2006). Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* 7(Suppl. 1), 10.1–10.12.
- Strasser, R., Bondili, J. S., Vavra, U., Schoberer, J., Svoboda, B., Glössl, J., et al. (2007). A unique β 1, 3-galactosyltransferase is indispensable for the biosynthesis of N-glycans containing Lewis structures in *Arabidopsis thaliana*. *Plant Cell* 19, 2278–2292. doi: 10.1105/tpc.107.052985
- The UniProt Consortium, (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. doi: 10.1093/nar/gky1049
- Tsaneva, M., and Van Damme, E. J. M. (2020). 130 years of Plant Lectin Research. *Glycoconj. J.* 37, 533–551. doi: 10.1007/s10719-020-09942-y
- Vaid, N., Pandey, P. K., and Tuteja, N. (2012). Genome-wide analysis of lectin receptor-like kinase family from *Arabidopsis* and rice. *Plant Mol. Biol.* 80, 365–388. doi: 10.1007/s11103-012-9952-8
- Van Damme, E. J. M., Allen, A. K., and Peumans, W. J. (1987). Isolation and characterization of a lectin with exclusive specificity towards mannose from snowdrop (*Galanthus nivalis*) bulbs. *FEBS Lett.* 215, 140–144. doi: 10.1016/0014-5793(87)80129-1
- Van Damme, E. J. M., Barre, A., Rougé, P., and Peumans, W. J. (2004). Cytoplasmic/nuclear plant lectins: a new story. *Trends Plant Sci.* 9, 484–489. doi: 10.1016/j.tplants.2004.08.003
- Van Damme, E. J. M., Lannoo, N., and Peumans, W. J. (2008). Plant lectins. *Adv. Botanic. Res.* 48, 107–209. doi: 10.1016/s0065-2296(08)00403-5
- Van Damme, E. J. M., Peumans, W. J., Barre, A., and Rougé, P. (1998). Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *CRC Crit. Rev. Plant Sci.* 17, 575–692. doi: 10.1016/s0735-2689(98)00365-7
- Van Holle, S., De Schutter, K., Eggermont, L., Tsaneva, M., Dang, L., and Van Damme, E. J. M. (2017). Comparative study of lectin domains in model species: new insights into evolutionary dynamics. *Int. J. Mol. Sci.* 18:1136. doi: 10.3390/ijms18061136
- Van Holle, S., and Van Damme, E. J. M. (2018). Signaling through plant lectins: modulation of plant immunity and beyond. *Biochem. Soc. Trans.* 46, 217–233. doi: 10.1042/bst20170371
- Van Holle, S., and Van Damme, E. J. M. (2019). Messages from the past: new insights in plant lectin evolution. *Front. Plant Sci.* 10:36. doi: 10.3389/fpls.2019.00036
- Van Hove, J., Fouquaert, E., Smith, D. F., Proost, P., and Van Damme, E. J. M. (2011). Lectin activity of the nucleocytoplasmic EUL protein from *Arabidopsis thaliana*. *Biochem. Biophys. Res. Commun.* 414, 101–105. doi: 10.1016/j.bbrc.2011.09.031
- Vandenborre, G., Smaghe, G., and Van Damme, E. J. M. (2011). Plant lectins as defense proteins against phytophagous insects. *Phytochemistry* 72, 1538–1550. doi: 10.1016/j.phytochem.2011.02.024
- Wakefield, M. E., Bell, H. A., Fitches, E. C., Edwards, J. P., and Gatehouse, A. M. R. (2006). Effects of *Galanthus nivalis* agglutinin (GNA) expressed in tomato leaves on larvae of the tomato moth *Lacanobia oleracea* (Lepidoptera: Noctuidae) and the effect of GNA on the development of the endoparasitoid *Meteorus gyrtator* (Hymenoptera: Braconidae). *Bull. Entomol. Res.* 96:43. doi: 10.1079/ber2005396
- Wan, J., Tanaka, K., Zhang, X.-C., Son, G. H., Brechenmacher, L., Nguyen, T. H. N., et al. (2012). LYK4, a lysin motif receptor-like kinase, is important for chitin signaling and plant innate immunity in *Arabidopsis*. *Plant Physiol.* 160, 396–406. doi: 10.1104/pp.112.201699
- Wang, Z., Hobson, N., Galindo, L., Zhu, S., Shi, D., McDill, J., et al. (2012). The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. *Plant J.* 72, 461–473. doi: 10.1111/j.1365-3113.2012.05093.x
- Willmann, R., Lajunen, H. M., Erbs, G., Newman, M.-A., Kolb, D., Tsuda, K., et al. (2011). *Arabidopsis* lysin-motif proteins LYM1 LYM3 CERK1 mediate bacterial peptidoglycan sensing and immunity to bacterial infection. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19824–19829. doi: 10.1073/pnas.1112862108
- Yan, Q., Jiang, Z., Yang, S., Deng, W., and Han, L. (2005). A novel homodimeric lectin from *Astragalus mongholicus* with antifungal activity. *Arch. Biochem. Biophys.* 442, 72–81. doi: 10.1016/j.abb.2005.07.019
- Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., et al. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* 21, 650–659. doi: 10.1093/bioinformatics/bti042
- Zhang, N., and Deyholos, M. K. (2016). RNASeq analysis of the shoot apex of flax (*Linum usitatissimum*) to identify phloem fiber specification genes. *Front. Plant Sci.* 7:950. doi: 10.3389/fpls.2016.00950
- Zhong, R., Cui, D., and Ye, Z. (2019). Secondary cell wall biosynthesis. *New Phytol.* 221, 1703–1723. doi: 10.1111/nph.15537

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Petrova, Nazipova, Gorshkov, Mokshina, Patova and Gorshkova. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.