


Flax rhamnogalacturonan lyases: phylogeny, differential expression and modeling of protein structure

Natalia Mokshina* , Olga Makshakova, Alsu Nazipova, Oleg Gorshkov and Tatyana Gorshkova

Kazan Institute of Biochemistry and Biophysics, Federal Research Center 'Kazan Scientific Center of RAS', Kazan, 420111, Russian Federation

Correspondence

*Corresponding author,
e-mail: ne.mokshina@gmail.com

Received 10 September 2018;
revised 9 November 2018

doi:10.1111/ppl.12880

Rhamnogalacturonan lyases (RGLs; EC 4.2.2.23) degrade the rhamnogalacturonan I (RG-I) backbone of pectins present in the plant cell wall. These enzymes belong to polysaccharide lyase family 4, members of which are mainly from plants and plant pathogens. RGLs are investigated, as a rule, as pathogen 'weapons' for plant cell wall degradation and subsequent infection. Despite the presence of genes annotated as *RGLs* in plant genomes and the presence of substrates for enzyme activity in plant cells, evidence supporting the involvement of this enzyme in certain processes is limited. The differential expression of some *RGL* genes in flax (*Linum usitatissimum* L.) tissues, revealed in our previous work, prompted us to carry out a total revision (phylogenetic analysis, analysis of expression and protein structure modeling) of all the sequences of flax predicted as coding for RGLs. Comparison of the expressions of *LusRGL* in various tissues of flax stem revealed that *LusRGLs* belong to distinct phylogenetic clades, which correspond to two co-expression groups. One of these groups comprised *LusRGL6-A* and *LusRGL6-B* genes and was specifically upregulated in flax fibers during deposition of the tertiary cell wall, which has complex RG-I as a key noncellulosic component. The results of homology modeling and docking demonstrated that the topology of the *LusRGL6-A* catalytic site allowed binding to the RG-I ligand. These findings lead us to suggest the presence of RGL activity in planta and the involvement of special isoforms of RGLs in the modification of RG-I of the tertiary cell wall in plant fibers.

Introduction

Rhamnogalacturonan I (RG-I) is a pectin that is a principal component of the plant cell wall and has a backbone of repeating disaccharide units of $[\rightarrow 2)\text{-}\alpha\text{-L-Rhap-(1}\rightarrow 4)\text{-}\alpha\text{-D-GalpA-(1}\rightarrow]$ (McNeil et al. 1980, Lau et al. 1985). Side chains of RG-I can be built from neutral galactans, arabinogalactan I and arabinans (Ridley et al. 2001, Vincken et al. 2003). RG-I is a crucial component of the primary (Carpita

and Gibeaut 1993, Ridley et al. 2001) and tertiary cell walls (TCWs; Gorshkova et al. 2018a). The side chains of such polysaccharides, after their incorporation into the cell walls, may be subjected to postsynthetic modifications, including degradation; for example, expression of β -galactosidases and α -arabinofuranosidases that trim off RG-I side chains is reported (Arsovski et al. 2009, Roach et al. 2011). Theoretically, degradation of the RG-I backbone could be caused by rhamnogalacturonan hydrolase or rhamnogalacturonan lyase

Abbreviations – CAZy, database of carbohydrate-active enzymes; FIB, isolated fibers with tertiary cell wall; GDP, guanosine diphosphate; iFIB, intrusively elongating fibers with primary cell wall; MID, segment of the whole flax stem taken below snap point; NADP(H), nicotinamide adenine dinucleotide phosphate (reduced); PL, polysaccharide lyase; PL4, pectin lyase family 4; qRT-PCR, quantitative real-time polymerase chain reaction; RG-I, rhamnogalacturonan I; RGL, rhamnogalacturonan lyase; SYBR Green, N',N'-dimethyl-N-[4-[(E)-(3-methyl-1,3-benzothiazol-2-ylidene)methyl]-1-phenylquinolin-1-ium-2-yl]-N-propylpropane-1,3-diamine; TCW, tertiary cell wall; TOP, segment of the whole flax stem taken above snap point.

(RGL) activity. Rhamnogalacturonan hydrolases (EC 3.2.1.171–3.2.1.174) are not found in plants; they are described as part of the degradation system for RG-I in *Aspergillus aculeatus* and *Bacillus subtilis*. However, genes encoding RGLs (EC 4.2.2.) are annotated in plant genomes. In carbohydrate-active enzymes (CAZy) database (www.cazy.org), RGLs are classified into three polysaccharide lyase (PL) families: PL11 contains essential bacterial sequences and the PL4 family mainly comprises eukaryotic enzymes of fungal and plant origin (Garron and Cygler 2010, Lombard et al. 2010); bacterial and fungal *exo*-RGLs belonging to a new family, PL26, were recently described (Kunishige et al. 2018). As demonstrated in nonplant organisms, the family 4 polysaccharide lyase cleaves the α -1,4 backbone of RG-I through a β -elimination mechanism, generating oligomers with unsaturated uronic acid, such as uGalA and a double bond between C-4 and C-5 at the newly formed nonreducing end (Mutter et al. 1998, McDonough et al. 2004).

The only eukaryotic RGL from PL4 with a resolved 3-D structure is the enzyme from *A. aculeatus* (AaRGL4; McDonough et al. 2004). It contains three domains packed tightly together. The N-terminal domain (PF06045) is the largest and is predicted to contain the catalytic site. The analysis of the related sequences indicates that Lys150, Tyr203, Tyr205 and His210 in the N-terminal domain are strictly conserved residues, where Lys150 and His210 are the main candidates for the involvement in catalysis (Garron and Cygler 2010). The central (PF14686) and C-terminal (PF14683) domains are presumed to be involved in binding to the polysaccharide substrate. These domains were identified as a fibronectin type III-like domain and a carbohydrate-binding module-like domain (McDonough et al. 2004). The smallest substrate that AaRGL4 can cleave is a deacetylated 12-residue oligomer, suggesting the presence of a binding site consisting of a minimum of 12 subsites (Mutter et al. 1998). Experimental evidence showing the binding of RG-I hexasaccharide to the active site of AaRGL4 has been previously presented (Jensen et al. 2010).

N-terminal domains (domain I) of fungal and *Arabidopsis thaliana* (*Arabidopsis*) RGL sequences share low homology (only 4–9% identity); this may indicate the evolutionary divergence of a subfamily in which the catalytic machinery has been maintained, while the substrate recognition has undergone major modifications (McDonough et al. 2004). Sequence diversity in PL4 may reflect variations in the patterns of arabinan and galactan side chains of RG-Is from different species (McDonough et al. 2004). Moreover, molecular evolution could affect the catalytic activity of the enzyme such that its function

might be substantially modified or even lost (Kozlova et al. 2017), as has occurred with chitinase-like proteins (Patil et al. 2013).

No structural characterizations of plant RGLs have been reported to date. Moreover, despite the presence of RG-I in virtually all plant tissues and the wide distribution of RGL genes in plant genomes, the enzymatic activity of RGLs has been demonstrated only in a couple of studies. The application of exogenous RG oligomers labeled with a fluorescent tag to cotton cotyledons facilitated the detection of products from RGL activity (Naran et al. 2007). A possible dimeric product of RGL reaction – lepidimoic acid – was detected in seed exudates of cress (*Lepidium sativum*; Iqbal et al. 2016). However, in both cases, the observed activity was not attributed to the particular protein. One of the reasons for the limited number of reports on RGL activity in plants could be the difficulty to confirm the presence of unsaturated products in nanomole quantities (Naran et al. 2007).

Arguments supporting the importance of the proteins annotated as RGL in plants come from the analysis of gene expression. Aspen (*Populus tremula* L. \times *tremuloides* Michx.) tension wood (Andersson-Gunnerås et al. 2006), and flax bast fibers appear to be enriched with RGL transcripts (Roach and Deyholos 2007, 2008, Hobson et al. 2010). These genes were identified using the microarray method and predicted based on their homology with the known sequences; distinct genes were not identified, as whole genome sequences for *Linum* and *Populus* were not published yet. Recent transcriptome analysis using next-generation sequencing has revealed differential expression of RGLs in the flax fibers forming TCW (Gorshkov et al. 2017a). These data gave us a reason to suggest that particular RGLs could be specific for TCW formation. The TCW (also called the G-layer, considered by some researchers to be one of the layers of secondary cell wall; Clair et al. 2018) is a very distinct type of cell wall that is deposited in many plant fibers. As it is formed after primary and secondary cell walls and is very distinct in composition, architecture (Mikshina et al. 2013), and molecular machinery of deposition and modification (Gorshkov et al. 2017a), it was designated as the TCW (Gorshkova et al. 2018a). TCW has a high content of cellulose (up to 90%) and is characterized by axial orientation of all cellulose microfibrils. Due to the usual absence of xylan and lignin (Gorshkova et al. 2010, 2018a), TCW gained more and more attention from the scientific community and biotechnologists. In our study, we used flax not only because it is one of the important crops and a bast fiber resource, but also as a convenient model for investigation of the processes of

TCW formation, since all the stages of fiber development in flax have been well described (Mokshina et al. 2018, Gorshkova et al. 2018b). The presence of pectic rhamnogalacturonan I with long galactan side chains together with the available genome sequence of flax (Wang et al. 2012) demanded a revision and characterization of all RGL sequences from flax genome to provide further information for revealing the role of these proteins in plant cell functioning. Analysis of the plant RGL is an interesting point per se, for the development of fundamental knowledge about plant enzymes, their substrates, and mechanisms of interaction, as well as their evolution.

Material and methods

Sequence alignment, phylogenetic analysis and analysis of domain organization

Predicted full amino acid sequences of plant RGLs (Pfam domains: PF06045, PF14683, PF14686) were obtained from the Phytozome database v.12.0 (phytozome.jgi.doe.gov/pz/portal.html). The amino acid sequences of fungal and bacterial enzymes were selected from the CAZy database (www.cazy.org). Only sequences that had three RGL domains were chosen from the databases and aligned using PRALINE (www.ibi.vu.nl/programs/pralinewww/; Simossis and Heringa 2005) with the default parameters of homology-extended alignment strategy; Simossis et al. 2005). Multiple alignments of deduced amino acid sequences of proteins were also generated using PRALINE and phylogenetic trees were constructed using IQ-TREE web server (<http://iqtree.cibiv.univie.ac.at> [Trifinopoulos et al. 2016]). The Maximum Likelihood method (Kumar et al. 1994) based on auto searching for the best fit substitution model, ultrafast bootstrap analysis (Minh et al. 2013) and Shimodaira-hasegawa an approximate likelihood-ratio branch test (SH-aLRT) (Guindon et al. 2010; bootstrap analysis was used with 1000 replicates) were applied. Visualization of the consensus trees (Letunic and Bork 2016) was obtained from interactive tree of life (iTOL) web service (<http://itol.embl.de>). Domain organization of RGL proteins was predicted using the Conserved Domains Database (www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), followed by MyDomains, an image creator tool (<https://prosite.expasy.org/mydomains/>). All numbers and amino acid sequences of RGLs from fungi, bacteria, and plants that were used for phylogenetic tree construction are listed in Table S1, supporting information.

Bioinformatic analysis of RNA-Seq data

Transcriptomic data from the following samples were used: segment of the whole flax stem taken above snap

point (TOP), segment of the whole flax stem taken below snap point (MID), and isolated fibers with TCW (FIB; Gorshkov et al. 2017a). Snap point is the marker of developmental transition of flax bast fibers from intrusive elongation to cell wall thickening (Gorshkova et al. 2003). Additionally, samples from the experiment on flax stem gravitropic reaction, i.e., FIB at an advanced stage of development from noninclined plants (FIB2), and FIB from pulling side (FIB2_PUL) and the opposite side of gravistimulated stems (FIB2_OPP; Gorshkov et al. 2017b) were analyzed. Similarly, apical stem part (Apex), intrusively elongating fibers with primary cell wall (iFIB) and the samples from the xylem part of stem (5 cm long, at the distance 3 cm above cotyledon leaves), taken from the noninclined plants (XYL), from pulling side (XYL_PUL), and from the opposite side of the gravistimulated stems (XYL_OPP) were also analyzed (Gorshkova et al. unpublished data, two biological replicates).

Extraction of RNA from all samples was performed according to the protocol published earlier (Gorshkov et al. 2017a, 2017b). Briefly, the total RNA from plant samples was isolated using a Trizol-extraction method combined with an RNeasy Plant Mini Kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. Residual DNA was eliminated by treatment with DNase I using the TURBO DNA-free kit (Ambion, Carlsbad, CA). RNA quality was analyzed using a Qubit 2.0 fluorometer (Invitrogen, Waltham, MA). Each sample was treated using a TruSeq Sample Prep Kit v2 (Illumina, San Diego, CA) according to manufacturer's instruction.

After preprocessing the previously obtained FASTA formatted sequence and its quality data (FASTQ) files, the clean reads were remapped to the flax reference genome (Wang et al. 2012) using the aligner HISAT2 v2.1.0 (Baltimore, USA) (Kim et al. 2015) with default parameters and subsequently processed using the software Cufflinks (Trapnell et al. 2012). Fragments per kilobase of exon per million was used as the unit of measurement to estimate transcript abundance.

The gene co-expression network was constructed using the expression profiles from the above-described samples and the software CoExpNetViz (Tzfadia et al. 2016). We chose this algorithm for two reasons: (1) CoExpNetViz takes input as a set of 'bait genes', e.g. genes involved in the same biological process, and determines the genes co-expressed with these bait genes; (2) it uses Pearson correlation coefficient, which provides more accurate results because it is a parametric measure. The RGL genes were used as bait genes to query the gene co-expression network. Genes are considered to be co-expressed if their correlation does not lie between the lower (5th) and upper (95th) percentiles

of the distribution of correlations between samples of genes per gene expression matrix. The network was further visualized and analyzed using Cytoscape version 3.5.1 (New York, USA) (Shannon et al. 2003).

Validation of transcriptome experiments using qRT-PCR

All identified *LusRGL* genes (excepted three truncated genes *LusRGL-s2*, *-s3*, *-s4*) from TOP, MID and FIB samples were validated using real-time polymerase chain reaction (RT-PCR) carried out on a CFX96 Touch RT-PCR Detection System (Bio-Rad, Hercules, CA). Gene-specific primers for the analyzed genes were designed using the Universal ProbeLibrary Assay Design Centre (<http://lifescience.roche.com/>; Table S2). The efficiency of the primers was calculated by performing RT-PCR on several dilutions of first-strand cDNAs. Efficiencies of the different primer sets were similar. A cDNA dilution of 2.5 μl (1/15) was used for amplification. The PCR mixture (10 μl) was constituted by 0.4 μM of each forward and reverse gene-specific primer, 0.2 mM dNTPs, 1 \times SYBR Green (Sigma, St. Louis, MO), and 0.1 μl of 5 U μl^{-1} heat-stable Taq Polymerase (Evrogen, Moscow, Russia). The thermal cycling conditions were 95°C for 3 min, 40 cycles each at 95°C for 15 s and 60°C for 1 min. A 60-to-95°C melting curve was constructed to confirm specificity of the products. For each of the three biologically independent cDNA samples, two independent technical replications were performed and averaged for further calculations. Relative transcript abundance calculations were performed using the $2^{-\Delta\Delta\text{Ct}}$ method (Livak and Schmittgen 2001). The genes of *EUKARYOTIC TRANSLATION INITIATION FACTORS 1A, 5A* (*LusETIF1*, *LusETIF5A*) and *GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE* (*LusGAPDH*) were used as the housekeeping genes (Table S2; Huis et al. 2010). $\Delta\Delta\text{Ct}$ values were generated using TOP sample as a reference.

Sequence and template for homology modeling

The most highly expressed flax lyase, *LusRGL6-A* (Lus10004281) was chosen for structure homology modeling. Iterative threading assembly refinement (I-TASSER) web service was used to construct the 3D-structure of *LusRGL6-A* (Zhang 2008). RGL from *A. aculeatus* (GenBank AAA64368), the only dataset available from PL4 family in Cambridge Protein Data Bank (Berman et al. 2000, www.rcsb.org), was used as a structural template (pdb: 1nkg). The model with the least root-mean-square deviation was considered for further analysis, followed by structural quality control

with PROCHECK (Laskowski et al. 1996) and Verif3D services (Lüthy et al. 1992).

Substrate docking

RG-I backbone fragment containing three repeating units of [\rightarrow 2)- α -L-Rhap-(1 \rightarrow 4)- α -D-GalpA(1 \rightarrow)] was used as a ligand by analogy with the bound part of the substrate in the structure of RGL from *A. aculeatus* (pdb: 3njv). In the docking procedure, the side chains of the residues composing the substrate-binding site were iteratively allowed to rotate, to saturate the hydrogen bonds; rest of the protein was kept rigid. Docking poses were filtered, and those with the α -L-Rhap-(1 \rightarrow 4)- α -D-GalpA bond placed next to catalytic residues were taken for further analysis. Out of the cluster, the geometry of the most energetically favorable pose was analyzed. Molecular docking was carried out using Autodock4.2 (California, USA) (Morris et al. 2009). Resultant structures were visualized and analyzed in the visual molecular dynamics (VMD) program (Humphrey et al. 1996).

Results and discussion

Gene annotation, domain organization and phylogeny of flax RGLs

We searched the flax genome assembly in Phytozome database for predicted genes with homology to Pfam domains PF06045, PF14686, and PF14683, which are characteristic to RGLs of the pectin lyase family 4 (PL4). This search identified 19 predicted *LusRGLs* (Fig. 1). Six of them possessed only two (*Lus10010738*), or one (*Lus10010739*, *Lus10002666*, *Lus10010628*, *Lus10000111*, *Lus10000129*) domains from the three attributed domains. Besides, two of these sequences (*Lus10010739* and *Lus10010628*) contained additional C-terminal domain of M14N/E carboxypeptidase. These shortened sequences, designated as *LusRGL-s1*–*LusRGL-s6* (s from ‘shortened’), were not used for further phylogenetic analysis (Fig. 1).

We based the designation of flax RGLs, on the related sequences from Arabidopsis genome (Phytozome v12, *A. thaliana* TAIR10). The latter has eight sequences of RGLs, one of which has a truncated sequence (AT1G65210) and was not used for further alignment. AtRGLs were annotated according to their phylogeny (Fig. 1). Multiple alignment of RGL sequences from Arabidopsis and flax permitted subsequent tree construction to designate flax RGLs according to their grouping with AtRGLs. The presence of duplicates of *LusRGLs* can be associated either with local duplications or with a whole genome duplication event that probably occurred in ancestors of flax approximately 5 to 9 million

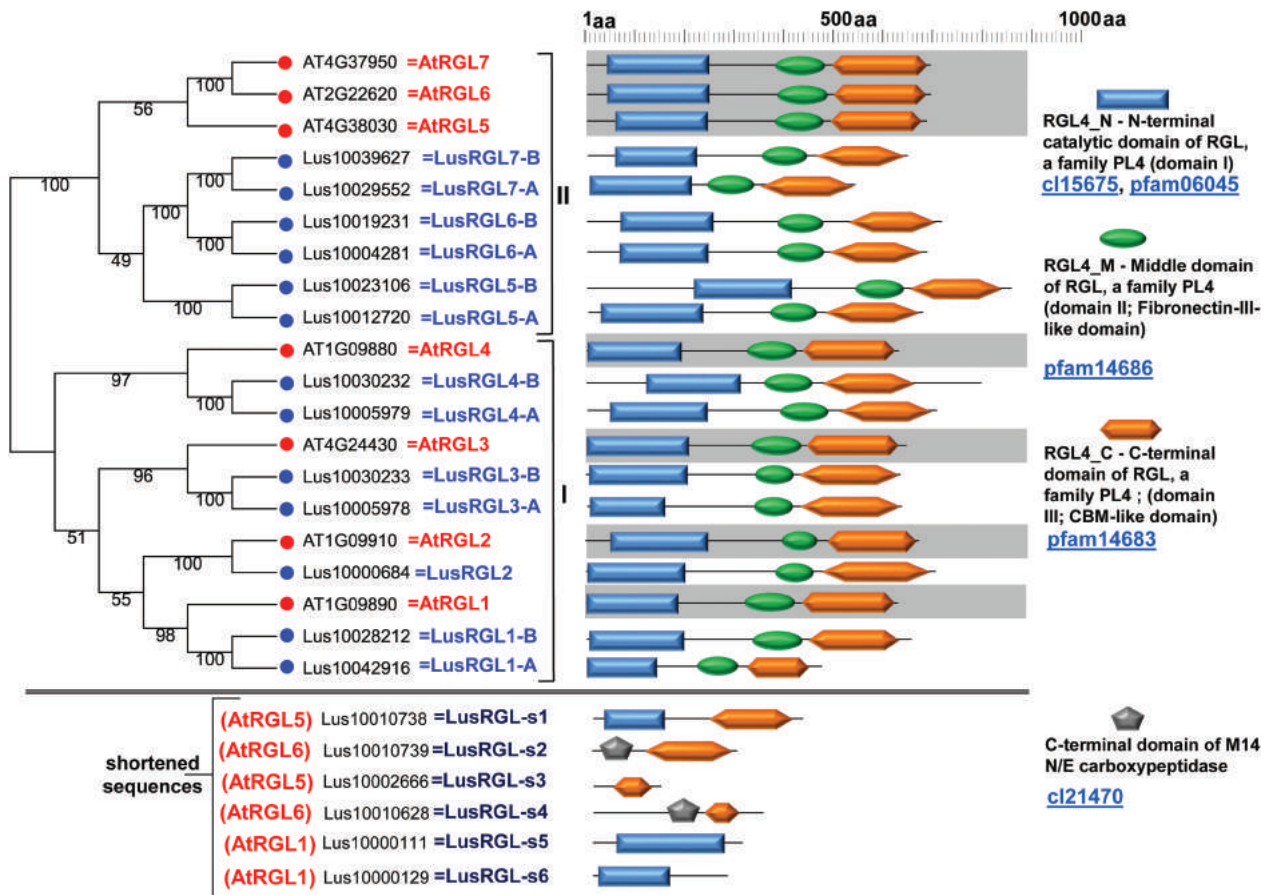


Fig. 1. Phylogenetic analysis and predicted domain organization of flax and Arabidopsis RGLs. On the left side: All predicted AtRGL (red) and LusRGL (blue) amino acid sequences were aligned using their deduced full-length peptide sequences generated using PRALINE web server. The evolutionary history of AtRGLs was inferred using maximum likelihood method based on whelan and goldman (WAG) substitution model and gamma with four categories model of rate heterogeneity, ultrafast bootstrap 1000 replicates (IQ-TREE). Length of branches was ignored. Shortened sequences that do not contain all three domains are listed below the tree (locus name in Phytozome, labels, and the closest homolog in Arabidopsis are given in brackets). On the right side: domain organization of AtRGLs and LusRGLs, predicted by the conserved domains database.

years ago (Wang et al. 2012). This is consistent with a large number of duplicated genes (9920) predicted for *Linum usitatissimum*, including gene families such as those encoding β -galactosidases, cellulose synthases, chitinase-like proteins and β -tubulins (Roach et al. 2011, Mokshina et al. 2014, Gavazzi et al. 2017), which may reflect lineage-specific genome duplications. The local duplicate has a higher chance of being removed through accumulation of degenerative mutations (Wang 2013), which probably explains the presence of shortened variants of RGLs. However, during the accumulation of degenerative mutations, the local duplicate may occasionally evolve modified/new functions (Wang 2013) that could be a reason why the truncated RGLs were preserved in flax genome.

The phylogenetic tree of flax RGLs, along with the known AtRGLs, showed two distinct groups: the first

group (I) included RGL1–RGL4 and the second one (II) included RGL5–RGL7 (Fig. 1). Segregation of plant RGLs into two groups was confirmed by construction of a tree that included RGLs from various plant species (Fig. 2) as well as those from fungi and bacteria (Fig. 3). According to the tree, plant RGLs have low similarity with fungal and bacterial RGLs that have documented properties and activity. McDonough et al. (2004) reported that plant RGLs share higher identity with some RGLs from phytopathogenic bacteria belonging to *Enterobacteriaceae* (18–22%), than with those from fungi (7%). *Dickeya* (CAD27359) in turn shares only 12% of identity with *Aspergillus* RGL (AAA6436; data not shown). However, DALI search reveals that despite considerable distance between corresponding domains of homologs, the similarity of structural domains is obvious (McDonough et al. 2004).

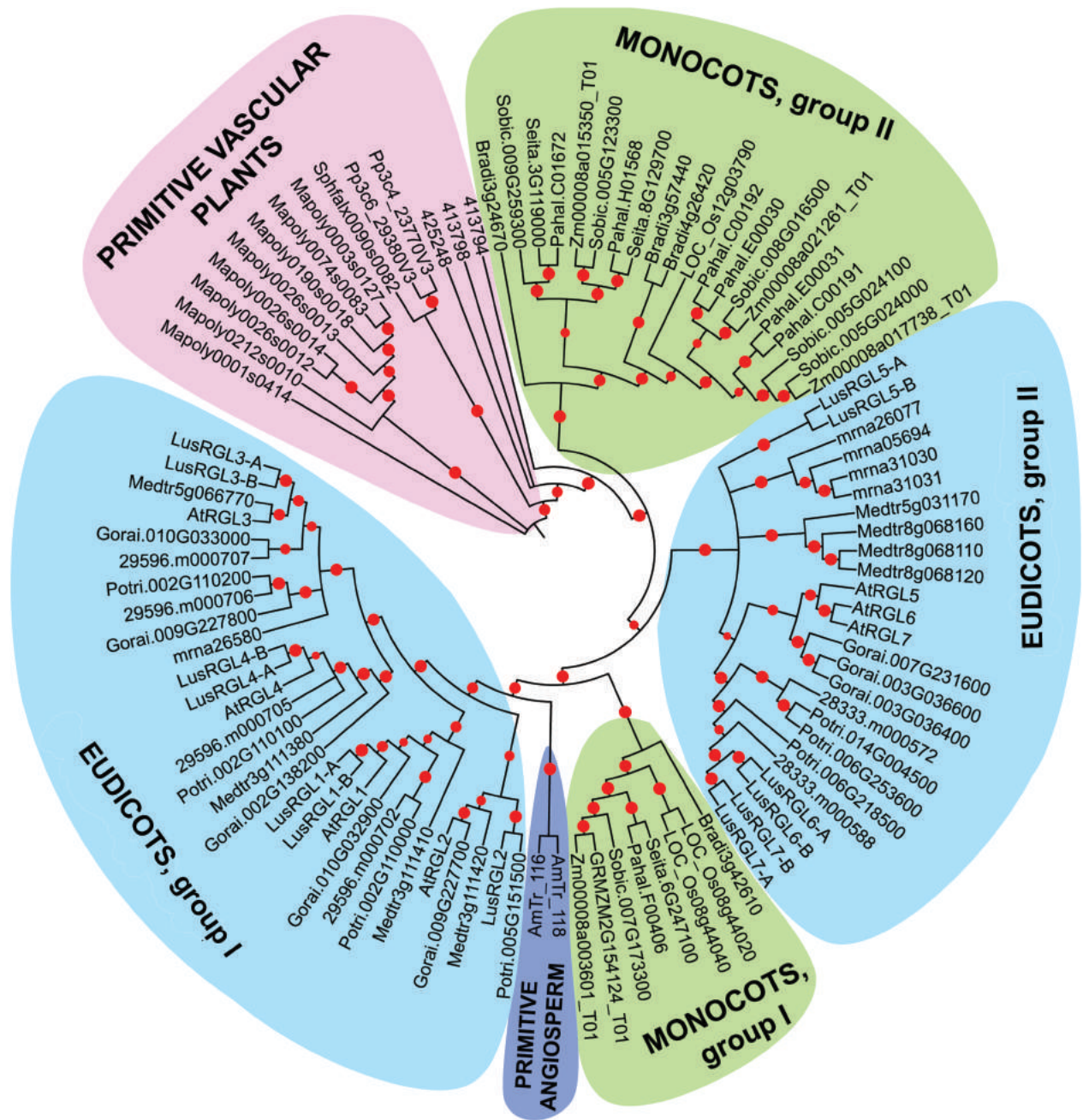


Fig. 2. Phylogenetic analysis of plant RGLs. Predicted plant RGL amino acid sequences (chosen from Phytozome) were aligned using their deduced full-length peptide sequences generated using PRALINE. The evolutionary history was inferred using maximum likelihood method based on WAG model using FreeRate with six categories model of rate heterogeneity, ultrafast bootstrap 1000 (IQ-TREE). AtRGLs and LusRGLs are annotated in accordance with Fig. 1. Branches corresponding to partitions, reproduced in less than 70% bootstrap replicates were deleted; branch length was ignored. Size of red circles indicates bootstrap support of branches: the smallest ones indicate 70% of bootstrap support and the biggest ones, 100%. List of RGL sequences used for tree construction is given in Table S1.

Eudicot RGLs (Fig. 2) were separated into two groups, forming the same clades as *Arabidopsis* and *Linum* (Fig. 1). The first group included AtRGL1-4 and homology sequences that were split in accordance with AtRGL

isoforms (eudicots, group I, Fig. 2). The second group included AtRGL5-7 and homology sequences that formed distinct species-specific groups (eudicots, group II, Fig. 2). RGLs belonging to primitive vascular

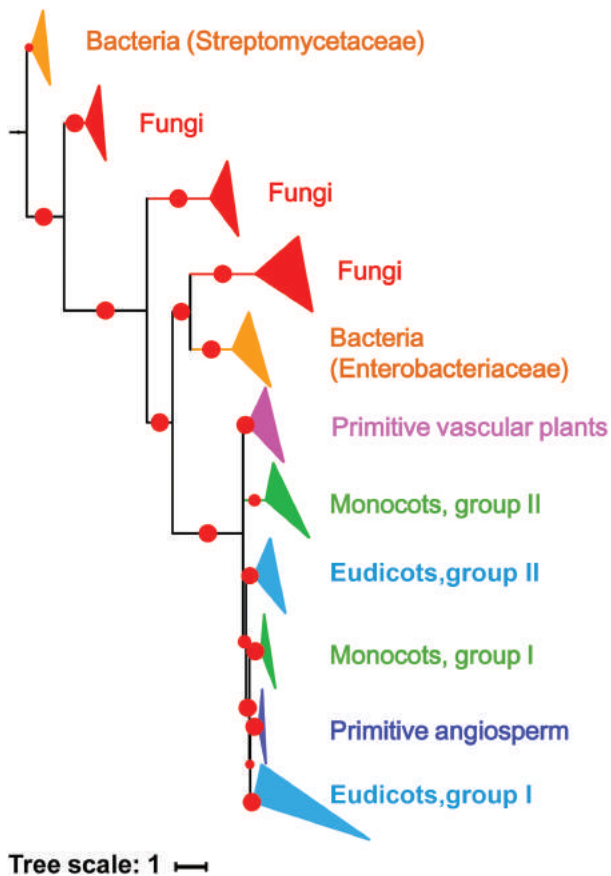


Fig. 3. Phylogenetic analysis of fungal, bacterial and plant RGLs. Predicted full-length amino acid sequences of plant, fungal and bacterial RGLs that have three RGL domains were chosen from Phytozome and CAZy databases, and subsequently, aligned in PRALINE. The evolutionary history was inferred using the maximum likelihood method based on WAG model using FreeRate with seven categories model of rate heterogeneity, ultrafast bootstrap 1000 (IQ-TREE). Branches corresponding to partitions reproduced in less than 70% bootstrap replicates were deleted. Clades are collapsed. Tree is unrooted, but clade with *Streptomyces* bacteria is drawn as root. Size of red circles indicates bootstrap support of branches and it ranged from 70% (small ones) to 100% (big ones). Sequence AAA64368 corresponding to *Aspergillus aculeatus* matched the closest to *Streptomyces* fungi clade. List of RGL sequences used for tree construction is given in Table S1.

formed a separate clade and were followed by a branch for monocots; the other group of monocots' RGLs was located close to group I of eudicots' RGLs. Two RGLs of *Amborella trichopoda*, sharing the most basal lineage in the clade of angiosperms (The Angiosperm Phylogeny Group 2016), were located at the branch close to group I of eudicots' RGLs (Fig. 2).

Generally, predicted RGLs had less distribution in monocot genomes, compared to eudicots (about 57 sequences of 14 monocot plant species vs 332 sequences of 37 eudicot plant species according to

Phytozome database). This can be explained by the difference in cell wall composition of monocots and eudicots: primary cell walls of maize and other comelinid monocots contain less pectin substances in comparison with eudicots (Carpita 1996); thus, potential substrates for these enzymes (e.g., RG-I) show a wider distribution in cell walls of eudicots. This fact can explain the variability of RGLs among eudicots and their separation from monocots in the course of evolution.

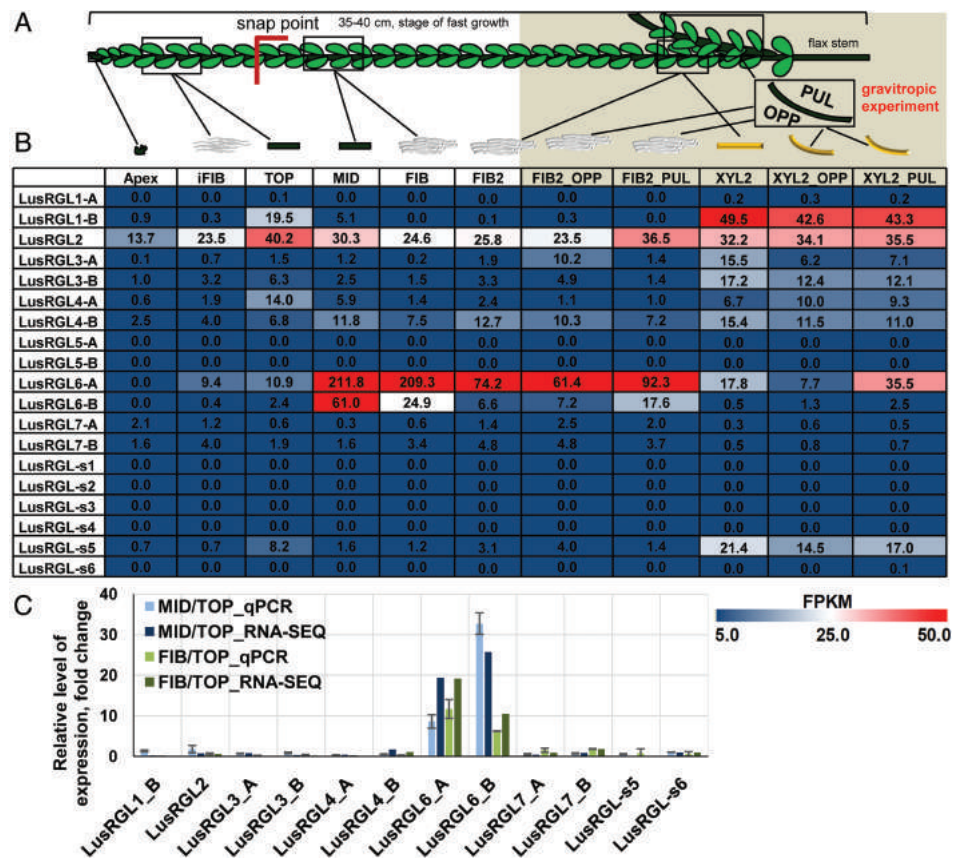
Differential expression of RGL genes (*LusRGLs*) in different flax tissues

To characterize the expression of various *LusRGLs*, we gathered transcriptomic data for different flax tissues obtained in our previous studies (Gorshkov et al. 2017a, 2017b) and data from the similar type of experiments for several other samples (Gorshkova et al. unpublished; Fig. 4A, B). To verify a subset of RNA-Seq data by an additional independent means, qRT-PCR analyses were carried out. Transcript levels for 16 *LusRGL* genes (except three truncated genes *LusRGL-s2*, *-s3*, *-s4*) were measured in TOP, MID and FIB samples. Both RNA-Seq and qRT-PCR analyses showed similar relative expression level of analyzed genes in MID and FIB samples, in comparison with TOP (MID/TOP, FIB/TOP; Fig. 4C), thus validating the RNA-Seq data. The expression of *LusRGL1-A*, *LusRGL5-A, B*, and *LusRGL-s1* in analyzed samples was not revealed by either RNA-Seq, or qRT-PCR; this fact can also be considered as RNA-Seq data confirmation.

The expression of various *LusRGLs* differed markedly. *LusRGL2* was expressed in a relatively constitutive manner in all analyzed samples (Fig. 4), similar to *AtRGL2* (AT1G09910, known as *AtMYST4*; Liu et al. 2012). This isoform of RGL might be associated with the remodeling of RG-I localized in primary cell walls or middle lamellae that are present in all plant cells. Transcripts of *LusRGL1-B* were more abundant in xylem, in comparison to other tissues, though the level of expression was not high. A similar circumstance was observed in the case of *LusRGL-s5*, which shares high homology with *AtRGL1*. Thus, the products of *LusRGL1-B* and *LusRGL-s5* genes could be involved in the modification of polysaccharides of cell wall in xylem tissues. Gravistimulation did not change transcript abundance of these xylem-specific genes. Several *LusRGLs* showed very low expression levels, if any, in analyzed tissues.

The most prominent changes in expression were detected for *LusRGL6* (Fig. 4), and its Arabidopsis homolog, *AtRGL6* (AT2G22620), also known as

Fig. 4. Relative expression of *LusRGLs* in different tissues of the flax stem. (A) Schematic representation of sample collection. (B) Transcriptomic data published previously were used (TOP, MID, FIB – Gorshkov et al. 2017a; FIB2, FIB2_OPP, FIB2_PUL – Gorshkov et al. 2017b, same as apex, iFIB – intrusively growing fibers, XYL2, XYL2_OPP, XYL2_PUL – xylem tissues in gravitropic experiments – Gorshkova et al. unpublished data). (C) Comparison of gene expression values obtained by qRT-PCR and RNA-Seq analyses. Ratios of MID/TOP and FIB/TOP samples were analyzed. Error bars indicate standard deviations. The gene list and sequences of corresponding primers are given in Table S2.



AtMYST6 (Buuck 2012, Liu et al. 2012). The orthologous gene of *AtMYST6* in strawberry (*Fragaria × ananassa*), *FaRGLyase1*, was expressed in the berry receptacle at the late phases of ripening (Molina-Hidalgo et al. 2013), and was regulated positively by abscisic acid and negatively by auxins. Analysis of plants with a silenced gene revealed that the product of this gene is involved in the degradation of pectins that are present in the middle lamellar region between parenchymatous cells of fruits. It was shown that *FaRGLyase1* is linked to a quantitative trait loci linkage group related to fruit hardness and firmness. Transcripts encoding poplar *RGLs*, similar to the Arabidopsis *AtMYST6* gene, increased considerably in tension wood (Andersson-Gunnerås et al. 2006).

In flax, abundance of transcripts of *LusRGL6-A* and *LusRGL6-B* was about 20-fold higher in samples containing phloem fibers at the stage of TCW deposition (MID, FIB), compared to samples with phloem fibers forming primary cell wall (TOP; Gorshkov et al. 2017a). At an advanced stage of phloem fiber development (FIB2), expression of both *LusRGL6s* dropped, but increased slightly in fibers from pulling side (FIB2_PUL) after gravistimulation, being always present in phloem fibers depositing TCW (Fig. 4).

Moreover, we detected a constitutive upregulation of *LusRGL6s* in phloem fibers forming TCW and also an upregulation in the xylem of stem after gravistimulation. Increase of transcript abundance was detected only in the xylem part on the pulling stem side – XYL2_PUL vs XYL2 (vertical control) and XYL2_OPP (xylem from the opposite stem side; Fig. 4). Notably, it is on the pulling stem side that some xylem fibers form G-layer (TCW) in the course of the gravitropic response (Ibragimova et al. 2017). These results once again confirm the involvement of *LusRGL6s* in the TCW formation.

Expression of *LusRGL6s* coincided with the presence of RG-I, which plays a significant role in TCW development (Gorshkova et al. 2010). Earlier, it was demonstrated that flax RG-I is subjected to modification by a specific β -(1,4)-galactosidase, which hydrolyzes galactan side chains of RG-I, providing proper function to this polymer in TCW (Roach et al. 2011, Mokshina et al. 2012). β -(1,4)-galactosidase might not be the only enzyme that modifies RG-I of flax fiber cell wall. The constructed co-expression networks for *LusRGLs* revealed that two distinct clades identified on the phylogenetic tree formed separate co-expression groups (Fig. 5). *LusRGL6s* had high co-expression level

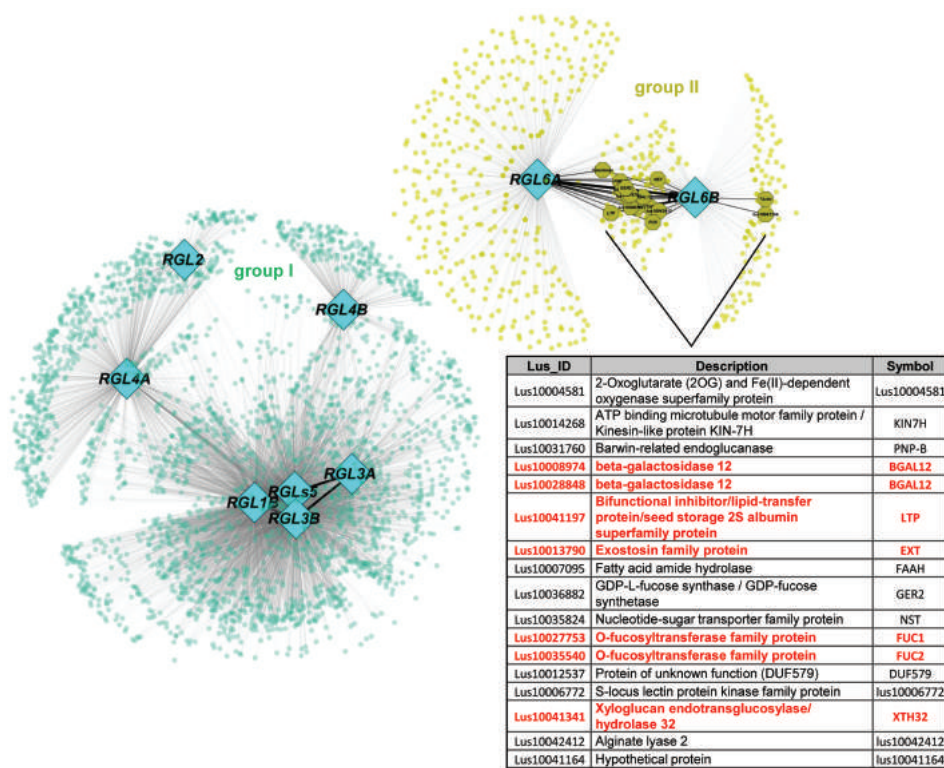


Fig. 5. Co-expression network of flax genes generated using the genes of flax RGLs (*LusRGLs*) as baits. The network comprises 3330 genes (nodes) with significant positive correlation (r -value above 0.786); cyan diamonds represent the bait genes (*LusRGLs*). The lines connecting two genes indicate that the two genes are co-expressed. Light cyan and yellow green nodes represent the genes associated with the bait genes. The big yellow green nodes represent the genes that were highly co-expressed (r -value above 0.95) with *LusRGL6* genes listed in the table below; genes that have been reported to be upregulated in the course of TCW formation in flax fibers (Gorshkov et al. 2017a) are designated in the table by red color.

with the genes that were specifically upregulated in fibers during TCW formation (Gorshkov et al. 2017a): β -galactosidase 12 (Lus10008974, Lus10028848), xyloglucan endotransglucosylase/hydrolase 32 (Lus10041341), O-fucosyltransferase family proteins (Lus10027753, Lus10035540; recently some members of this family [GT106] were proved to be RG-I rhamnosyltransferases; Takenaka et al. 2018), GDP-L-fucose synthase/GDP-fucose synthetase (Lus10036882), exostosin (Lus10013790), and others (Fig. 5).

Thus, based on *RGL* expression data, the presence of functionally distinct RGLs could be assumed: the RGLs that use RG-I of the primary cell walls (Yapo 2011) as potential substrate, and the RGLs that modify RG-I present in the gelatinous (tertiary) cell walls of plant fibers (Gorshkova et al. 2010, 2015, Guedes et al. 2017). The subfunctionalization of *RGL* genes probably occurred on the basis of *RGL* genes from group II of eudicots (Fig. 2): AtRGL6 (AtMYST6) could be involved in primary cell wall modification, while RGL6 of flax and aspen, which genes are closer related to each other than to Arabidopsis genes, are involved in TCW. Based on expression and co-expression data, we suggest that LusRGL6-A and LusRGL6-B can potentially take part in the modification of RG-I, a key stage-specific polymer of TCW (Gorshkova et al. 2018a, 2018b). To check whether the LusRGL6s are able to adopt a substrate and fulfill its lyase function,

3D-structure of LusRGL6-A was constructed based on homology modeling and docking of the substrate to the binding site.

3D model of LusRGL6-A indicates the possible enzyme catalytic action

Multiple alignments of amino acid sequences of RGLs belonging to PL4 family allowed deduction of the degree of conservation of amino acid sequences that are essential for catalysis and substrate binding (Fig. 6). In AaRGL sequences (GenBank AAA64368), the presence of four blocks of conserved residues was shown based on sequence alignment analysis. Three of these conserved blocks are in domain I and one in domain III (McDonough et al. 2004). The first conserved block is [G-E-L-R-x-A-R-L]. This conserved block was found in the sequences of fungal RGL, but this motif was changed in RGL sequences of bacteria and plants (Fig. 6). While the first arginine (R) is strictly conserved among different taxa, the second arginine is changed to lysine (K) in most of plant sequences. In a few plant sequences, the second arginine is substituted by methionine (M) or glutamine (Q; data not shown). In the conserved block [S-K-F-Y-S], only catalytic lysine was saved in all sequences of fungal, bacterial and plant RGLs. [S-K-F-Y-S] block is changed to the conserved block, [G-E-V-D-D-K-Y-x-Y], in plants by some substitutions in sequences with similar

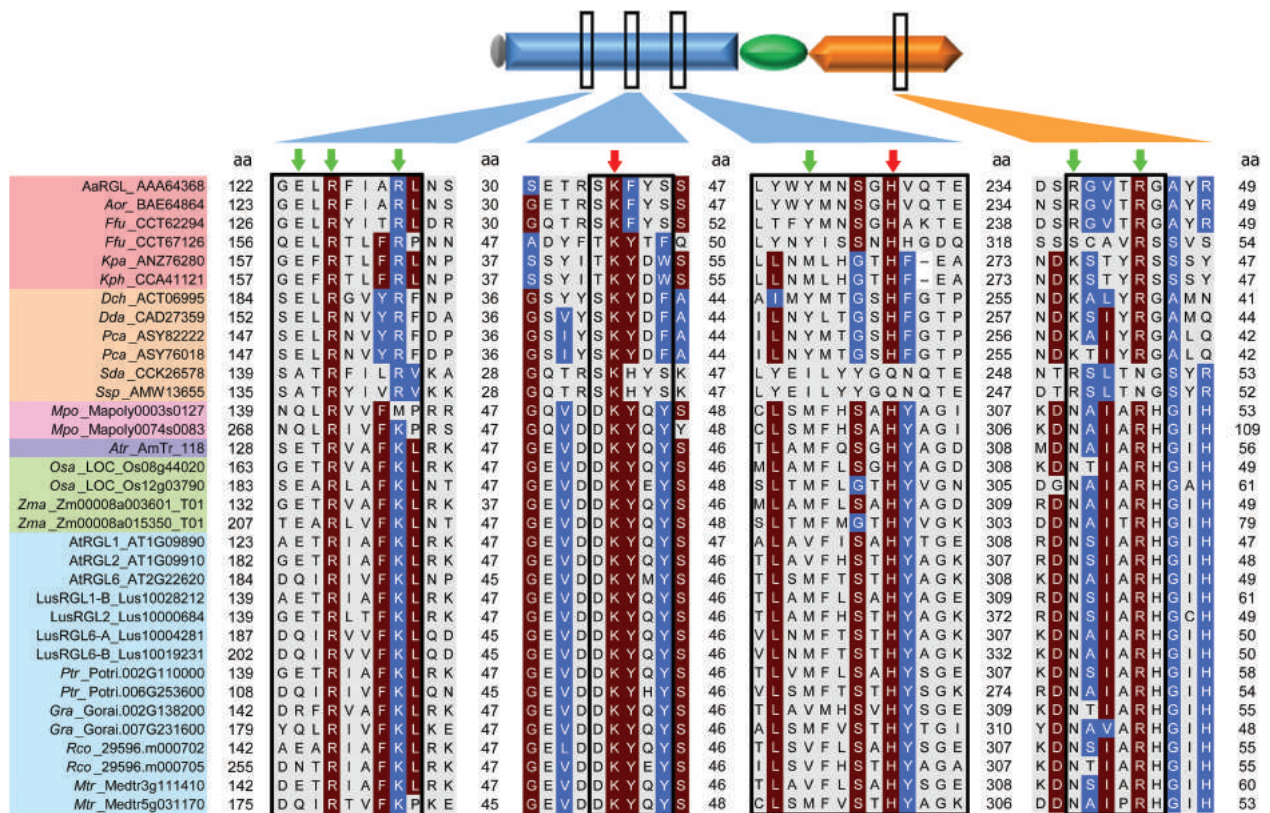


Fig. 6. Multiple amino acid sequence alignment of fungal, bacterial and plant RGLs from family 4 of polysaccharide lyases. Residues are colored by identity and similarity: Brown color indicates an identity of 70% or higher, blue color indicates 70% or higher similarity. Blocks conserved in fungal and bacterial RGLs (McDonough et al. 2004) are indicated by black frames. Columns indicated by light green arrows show substrate-binding amino acids of catalytic site, and those with red arrows show conserved catalytic amino acids, according to McDonough et al. (2004). Columns indicated by 'aa' show the number of left out amino acids in sequences. Background colors of organism protein names correspond to colors of clades on phylogenetic tree (Fig. 3). Fungal and bacterial protein sequences were obtained from CAZy, and plant sequences from Phytozome. Multiple alignments were performed in PRALINE. Domain organization of AaRGL at the top of this figure, from left to right, is as follows: Gray color indicates signal peptide, blue color, RGL4_N (pfam06045); green, RGL4_M (pfam14686); and orange color, RGL4_C (pfam14683). *Aac*, *Aspergillus aculeatus* KSM 510; *Aor*, *Aspergillus oryzae* RIB40; *Atr*, *Amborella trichopoda*; *AtRGL1*, *AtRGL2*, *AtRGL6*, RGLs of *Arabidopsis thaliana*; *Dch*, *Dickeya chrysanthemi* Ech1591; *Dda*, *Dickeya dadantii* 3937; *Ffu*, *Fusarium fujikuroi* IMI 58289; *Gra*, *Gossypium raimondii*; *Kpa*, *Komagataella pastoris* ATCC 28485; *Kph*, *Komagataella phaffii* CBS 7435; *LusRGL1-B*, *LusRGL2*, *LusRGL6-A*; *LusRGL6-B*, RGLs of *Linum usitatissimum*; *Mpo*, *Marchantia polymorpha*; *Mtr*, *Medicago truncatula*; *Osa*, *Oryza sativa*; *Pca*, *Pectobacterium carotovorum* SK-992-20-13; *P. carotovorum* Polaris; *Ptr*, *Populus trichocarpa*; *Rco*, *Ricinus communis*; *Sda*, *Streptomyces davawensis* JCM 4913; *Ssp*, *Streptomyces* sp. S10 (2016); *Zma*, *Zea mays* PH207.

amino acids. In [L-Y-x-Y-M-x-S-x-H-x-Q-x-E] block belonging to domain I, only catalytic histidine (H) is conserved in all taxa, except *Streptomyces* bacteria protein sequences, in which histidine is substituted by glutamine (Q). In [R-G-x-T-R-G] block in domain III, the following substitutions were observed: the first arginine (R) was changed to serine (S) in some fungal species, to lysine (K) in *Enterobacteriaceae* – a bacterial family – and in some fungal species, and to asparagine (N) in all the plant sequences. The second arginine maintained its position in conserved block in all analyzed RGLs barring *Streptomyces* bacteria (data not shown). Thus, RGL protein sequences of fungi, bacteria, and plants are distinguishable. Two of four arginines (which

take part in the formation of binding site, and enable the substrate capture due to interaction with residues of galacturonic acid and rhamnose in positions +3, +2, and – 2) are strictly conserved, whereas the other two are substituted by lysine, glutamine, asparagine and some others residues in different sequences. Various substitutions for tyrosine and glutamic acid (substrate-binding residues) were also observed among RGLs sequences.

For the construction of *LusRGL6-A* 3D-model, the RGL from *A. aculeatus* (AaRGL, pdb code: 1nkg) was used as a template. The sequence of *LusRGL6-A* used for homology modeling shares 20.8% identity with AaRGL. The resultant spatial structure of the model is shown in Fig. 7A. According to PROCHECK evaluation,

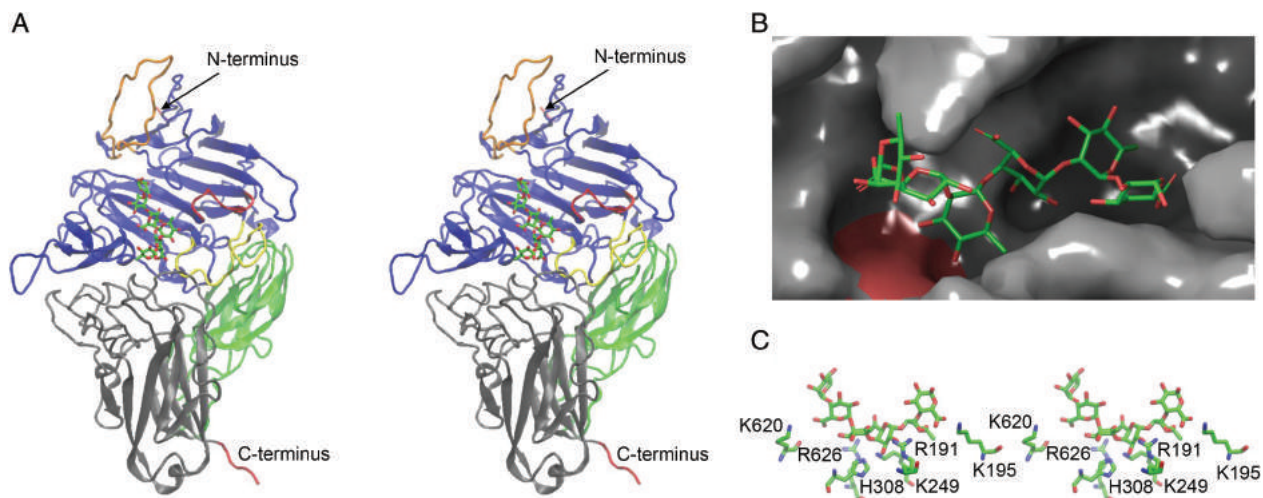


Fig. 7. (A) Three domains of the computed model of flax RGL (LusRGL6-A) are given in cartoon representation, and are highlighted in blue for N-terminal catalytic domain, green for fibronectin-III-like domain, and gray for C-terminal carbohydrate-binding module-like domain. The ligand, a hexasaccharide of RG-I backbone, is shown in stick representation, indicating C atoms in green and O atoms in red color. The loop L1 is highlighted in orange, L2 in red, and L3 in yellow. The structure is shown in side-by-side stereo view. (B) Surface representation of the LusRGL6-A-binding site with the substrate inside. The color coding of the substrate is as described above. The position of T307 is highlighted in pink. (C) The residues of the LusRGL6-A-binding site, with the substrate inside, are shown in stick representation. The color coding of C and O atoms in the substrate and residues is as described above; N are given in blue. The structure is shown in side-by-side stereo view.

most of the backbone torsions of the model structure lay in allowed regions, i.e. 61.4% in the most favored regions, 31.7% in the additional allowed regions, 5.5% in generally allowed regions and 1.3% in disallowed regions. The corresponding indicators identified in the template AaRGL structure were similar: 88.3, 10.8, 0.2 and 0.7%, respectively. The G-factors for covalent bonds were 0.57 and 0.37, and the overall G-factors were -0.86 and 0.12 for the model and template structures, respectively.

Similar to the template structure, organization of three domains is essential for the model. The 52-residue long fragment placed on the N-terminus of LusRGL6-A is missing in the template and remains structurally uncharacterized. Loss of this part did not seem to significantly influence the binding of the substrate RG-I, because of its spatial distance from the active site. The substrate-binding site, which is formed mainly by the residues of the first domain and partially by the third domain, has the topology of a deep groove (Fig. 7B). The binding cleft of flax RGL demonstrates a number of amino acid substitutions compared to the proteins of the PL4 family. Conserved residues R191 and R626 along with K195 and catalytic K249 provide electropositive spots on the protein surface, which entrap negatively charged fragments of the substrate and orient the molecule in the cleft. Numerous hydrogen bonds and van der Waals contacts facilitate efficient binding of the substrate to the binding pocket; some of these bonds

are formed by Q189, Y228, Y250, K120 and catalytic K249. Two residues, R191 and Y250, form contact with the galacturonyl residue, which bears the glycosidic linkage to be cleaved. The residue Q189 interacts with the adjacent rhamnosyl moiety. In the complex, glycosidic oxygen atom of rhamnose is in close contact with H308, while the catalytic K249 is in proximity of galacturonyl's C5 atom (Fig. 7C). In the β -elimination reaction, according to a previously proposed scheme (Gacesa 1987), catalytic histidine donates a proton to glycosidic oxygen, while the lysine accepts a proton from the CH5 group of galacturonyl residue during double bond C4=C5 formation.

Although the catalytic dyad is intact and the binding site can adopt the substrate molecule, the efficacy of catalysis can be tuned. The binding energy of LusRGL6-A was found to be lower than that of *A. aculeatus* calculated from X-ray structure (-7 kcal mol $^{-1}$ against -10 kcal mol $^{-1}$), suggesting a weaker affinity toward the substrate in the plant protein than the fungal one. Another important feature of LusRGL6-A is a sequence much larger than that of the fungal protein. This peculiarity gives rise to the longer loops in the model, compared to the fungal template (Fig. 7A). Two of them (here and after L1 and L2) are located next to the active site (Fig. 7A). The loops L1 and L2 comprise 24 (N91 to A114) and 13 residues (L176 to D188), respectively. In our static model, the loops are in 'open' state and do not interfere with substrate

binding. Nevertheless, we suspect that the dynamics of the loops will regulate the availability of the site for substrate, and hence, will tune the enzymatic activity. The presence of extended loops was shown, e.g. in Arabidopsis GDP-mannose dehydratase (MUR1) complexes, with GDP and GDP-D-rhamnose (Mulichak et al. 2002). The loop formed by residues 60–75, which makes interactions important for both NADP(H) binding and tetramer formation in the MUR1 structures, was completely disordered in the *E. coli* apoGMD. The other two loops, which close over the bound ligand in GDP and GDP-sugar complexes, serve as flexible flaps that may open to allow access of substrate to the catalytic site. Once the site is occupied, these loops in the closed conformation contribute to important binding interactions and isolate the substrate from the surrounding solvent. In cystathionine β -lyase from Arabidopsis, the importance of extended loop structure of the N-terminal domain in tetramer stabilization was demonstrated, as well as the involvement of another C-terminal loop in building the substrate-binding pocket (Breitinger et al. 2001).

Since RG-I usually bears arabinan, galactan or arabinogalactan side chains at O4 position of rhamnose, the possible level of substitution is an important question. The analysis of binding site topology reveals that the groove can adopt a ligand with branched rhamnosyl residues, except those bearing the target glycosidic linkage. The residue T307 belonging to the loop L3 (305–325) is in close proximity of O4 moiety of rhamnose. Based on our static model, we expect that this feature will demand rhamnose to be linear.

Conclusions

Based on the data of expression and co-expression in flax tissues, we suggest that some isoforms of these enzymes are involved in the modification of RG-I of thickened TCW of plant fibers. Homology modeling of LusRGL6-A and natural ligand docking demonstrated that the topology of the catalytic site allowed binding of the ligand in the 'correct' position, but the topology of the substrate-binding site was changed. Thus, despite differences in the structure of fungal and plant RGLs and their low similarity, there is a chance that the plant RGLs could act as true lyases. However, without strong evidence of their enzymatic activity (using recombinant proteins, e.g.), functions of other nature cannot be excluded. Further investigation of these enzymes will clarify at least two fundamental points: evolutionary origin and fate of carbohydrate-active enzymes and metabolism of one of the most important pectic substances of plant cell walls, RG-I.

Author contributions

N.M. collected flax samples, prepared cDNA libraries for RNA-Seq, performed phylogeny analysis; O.M. carried out 3D modeling and ligand docking; A.N. performed phylogenetic analysis, and took part in 3D modeling; O.G. conducted bioinformatics analysis of RNA-Seq data and performed co-expression analysis; T.G. conceived the study topic, wrote and reviewed the manuscript. N.M., O.M., A.N., O.G., and T.G. wrote the manuscript. All authors read and approved the manuscript.

Acknowledgements – This work was partially supported by grants of the Russian Science Foundation (#16-14-10256, RNA-Seq data; #17-76-20049, bioinformatics work), and by project no. MK-8014.2016.4 of the President of the Russian Federation (theoretical work, computer work on phylogeny). We are grateful to Dr. N. N. Ibragimova (Kazan Institute of Biochemistry and Biophysics) for design of experiments on gravitropism and help in sample collection.

References

- Andersson-Gunnerås S, Mellerowicz EJ, Love J, Segerman B, Ohmiya Y, Coutinho PM, Nilsson P, Henrissat B, Moritz T, Sundberg B (2006) Biosynthesis of cellulose-enriched tension wood in Populus: global analysis of transcripts and metabolites identifies biochemical and developmental regulators in secondary wall biosynthesis. *Plant J* 45: 144–165
- Angiosperm Phylogeny Group (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot J Linn Soc* 181: 1–20
- Arsovski AA, Popma TM, Haughn GW, Carpita NC, McCann MC, Western TL (2009) AtBXL1 encodes a bifunctional β -d-xylosidase/ α -L-arabinofuranosidase required for pectic arabinan modification in Arabidopsis mucilage secretory cells. *Plant Physiol* 150: 1219–1234
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* 28: 235–242
- Breitinger U, Clausen T, Ehlert S, Huber R, Laber B, Schmidt F, Pohl E, Messerschmidt A (2001) The three-dimensional structure of cystathionine β -lyase from Arabidopsis and its substrate specificity. *Plant Physiol* 126: 631–642
- Buuck R (2012) Mapping genomes: a novel gene family in plants may encode pectin-modifying proteins. *J Purdue Undergrad Res* 2: 93
- Carpita NC (1996) Structure and biogenesis of the cell walls of grasses. *Annu Rev Plant Physiol Plant Mol Biol* 47: 445–476

- Carpita NC, Gibeaut DM (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
- Clair B, Déjardin A, Pilate G, Alméras T (2018) Is the G-layer a tertiary cell wall? *Front Plant Sci* 9: 623
- Gacesa P (1987) Alginate-modifying enzymes: A proposed unified mechanism of action for the lyases and epimerases. *FEBS Lett* 212: 199–202
- Garron ML, Cygler M (2010) Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. *Glycobiol* 20: 1547–1573
- Gavazzi F, Pigna G, Braglia L, Gianì S, Breviario D, Morello L (2017) Evolutionary characterization and transcript profiling of β -tubulin genes in flax (*Linum usitatissimum* L.) during plant development. *BMC Plant Biol* 17: 237
- Gorshkov O, Mokshina N, Gorshkov V, Chemikosova S, Gogolev Y, Gorshkova T (2017a) Transcriptome portrait of cellulose-enriched flax fibres at advanced stage of specialization. *Plant Mol Biol* 93: 431–449
- Gorshkov O, Mokshina N, Ibragimova N, Ageeva M, Gogoleva N, Gorshkova TA (2017b) Phloem fibers as motors of gravitropic behaviour of flax plants: level of transcriptome. *Funct Plant Biol* 45: 203–215
- Gorshkova TA, Sal'nikov VV, Chemikosova SB, Ageeva MV, Pavlencheva NV, van Dam JEG (2003) The snap point: a transition point in *Linum usitatissimum* bast fiber development. *Ind Crops Prod* 18: 213–221
- Gorshkova TA, Gurjanov OP, Mikshina PV, Ibragimova NN, Mokshina NE, Salnikov VV, Ageeva MV, Amenitskii SI, Chernova TE, Chemikosova SB (2010) Specific type of secondary cell wall formed by plant fibers. *Russ J Plant Physiol* 57: 328–341
- Gorshkova T, Mokshina N, Chernova T IN, Salnikov V, Mikshina P, Tryfona T, Banasiak A, Immerzeel P, Dupree P, Mellerowicz EJ (2015) Aspen tension wood fibers contain β -(1→4)-galactans and acidic arabinogalactans retained by cellulose microfibrils in gelatinous walls. *Plant Physiol* 169: 2048–2063
- Gorshkova T, Chernova T, Mokshina N, Ageeva M, Mikshina P (2018a) Plant 'muscles': fibers with a tertiary cell wall. *New Phytol* 218: 66–72
- Gorshkova T, Mikshina P, Petrova A, Chernova T, Mokshina N, Gorshkov O (2018b) Plants at bodybuilding: development of plant "muscles". In: Geitmann A, Gril J (eds) *Plant Biomechanics*. Springer, Cham, pp 141–164
- Guedes FTP, Laurans F, Quemener B AC, Lainé-Prade V, Boizot N, Vigouroux J, Lesage-Descauses MC, Leplé JC, Déjardin A, Pilate G (2017) Non-cellulosic polysaccharide distribution during G-layer formation in poplar tension wood fibers: abundance of rhamnogalacturonan I and arabinogalactan proteins but no evidence of xyloglucan. *Planta* 246: 857–878
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of phyml 3.0. *Syst Biol* 59: 307–321
- Hobson N, Roach MJ, Deyholos MK (2010) Gene expression in tension wood and bast fibres. *Rus J Plant Physiol* 57: 321–327
- Huis R, Hawkins S, Neutelings G (2010) Selection of reference genes for quantitative gene expression normalization in flax (*Linum usitatissimum* L.). *BMC Plant Biol* 10: 71
- Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14: 33–38
- Ibragimova NN, Ageeva MV, Gorshkova TA (2017) Development of gravitropic response: unusual behavior of flax phloem G-fibers. *Protoplasma* 254: 749–762
- Iqbal A, Miller JG, Murray L, Sadler IH, Fry SC (2016) The pectic disaccharides lepidimic acid and β -d-xylopyranosyl-(1→3)-d-galacturonic acid occur in cress-seed exudate but lack allelochemical activity. *Ann Bot* 117: 607–623
- Jensen MH, Otten H, Christensen U, Borchert TV, Christensen LL, Larsen S, Leggio LL (2010) Structural and biochemical studies elucidate the mechanism of rhamnogalacturonan lyase from *Aspergillus aculeatus*. *J Mol Biol* 404: 100–111
- Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12: 357–360
- Kozlova LV, Mokshina NE, Nazipova AR, Gorshkova TA (2017) Systemic use of "limping" enzymes in plant cell walls. *Rus J Plant Physiol* 64: 808–821
- Kumar S, Tamura K, Nei M (1994) MEGA: molecular evolutionary genetics analysis software for microcomputers. *Comput Appl Biosci* 10: 189–191
- Kunishige Y, Iwai M, Nakazawa M, Ueda M, Tada T, Nishimura S, Sakamoto T (2018) Crystal structure of exo-rhamnogalacturonan lyase from *Penicillium chrysogenum* as a member of polysaccharide lyase family 26. *FEBS Lett* 592: 1378–1388
- Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8: 477–486
- Lau JM, McNeil M, Darvill AG, Albersheim P (1985) Structure of the backbone of rhamnogalacturonan I, a pectic polysaccharide in the primary cell walls of plants. *Carbohydr Res* 137: 111–125
- Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44: W242–W245
- Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, Arenas-Huertero C, Chua NH (2012) Genome-wide

- analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell* 24: 4333–4345
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)⁻(Delta Delta C) method. *Methods* 25: 402–408
- Lombard V, Bernard T, Rancurel C, Brumer H, Coutinho PM, Henrissat B (2010) A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochem J* 432: 437–444
- Lüthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature* 356: 83–85
- McDonough MA, Kadirvelraj R, Harris P, Poulsen JCN, Larsen S (2004) Rhamnogalacturonan lyase reveals a unique three-domain modular structure for polysaccharide lyase family 4. *FEBS Lett* 565: 188–194
- McNeil M, Darvill AG, Albersheim P (1980) Structure of plant cell walls: X. Rhamnogalacturonan I, a structurally complex pectic polysaccharide in the walls of suspension-cultured sycamore cells. *Plant Physiol* 66: 1128–1134
- Mikshina P, Chernova T, Chemikosova S, Ibragimova N, Mokshina N, Gorshkova T (2013) Cellulosic fibers: role of matrix polysaccharides in structure and function. In: Van De Ven TGM (ed) *Cellulose – Fundamental Aspects*. InTech, Rijeka, pp 91–112
- Minh BQ, Nguyen MA, von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* 30: 1188–1195
- Mokshina NE, Ibragimova NN, Salnikov VV, Amenitskii SI, Gorshkova TA (2012) Galactosidase of plant fibers with gelatinous cell wall: identification and localization. *Russ J Plant Physiol* 59: 246–254
- Mokshina N, Gorshkova T, Deyholos MK (2014) Chitinase-like (*CTL*) and cellulose synthase (*CESA*) gene expression in gelatinous-type cellulosic walls of flax (*Linum usitatissimum* L.) bast fibers. *PLoS One* 9: e97949
- Mokshina N, Chernova T, Galinovsky D, Gorshkov O, Gorshkova T (2018) Key stages of fiber development as determinants of bast fiber yield and quality. *Fibers* 6: 20
- Molina-Hidalgo FJ, Franco AR, Villatoro C, Medina-Puche L, Mercado JA, Hidalgo MA, Monfort A, Caballero JL, Muñoz-Blanco J, Blanco-Portales R (2013) The strawberry (*Fragaria xananassa*) fruit-specific rhamnogalacturonate lyase 1 (*FaRGLyase 1*) gene encodes an enzyme involved in the degradation of cell-wall middle lamellae. *J Exp Bot* 64: 1471–1483
- Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 30: 2785–2791
- Mulichak AM, Bonin CP, Reiter W-D, Garavito RM (2002) Structure of MUR1 GDP-mannose 4,6-dehydratase from *Arabidopsis thaliana*: implications for ligand binding specificity. *Biochemistry* 41: 15578–15589
- Mutter M, Colquhoun IJ, Beldman G, Schols HA BEJ, Voragen AG (1998) Characterization of recombinant rhamnogalacturonan α -l-rhamnopyranosyl-(1,4)- α -d-galactopyranosyluronide lyase from *Aspergillus aculeatus* an enzyme that fragments rhamnogalacturonan I regions of pectin. *Plant Physiol* 117: 141–152
- Naran R, Pierce ML, Mort AJ (2007) Detection and identification of rhamnogalacturonan lyase activity in intercellular spaces of expanding cotton cotyledons. *Plant J* 50: 95–107
- Patil DN, Datta M, Dev A, Dhindwal S, Singh N, Dasauni P, Kundu S, Sharma AK, Tomar S, Kumar P (2013) Structural investigation of a novel N-acetyl glucosamine binding chi-lectin which reveals evolutionary relationship with class III chitinases. *PLoS One* 8: e63779
- Ridley BL, O'Neill MA, Mohnen D (2001) Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry* 57: 929–967
- Roach MJ, Deyholos MK (2007) Microarray analysis of flax (*Linum usitatissimum* L.) stems identifies transcripts enriched in fibre-bearing phloem tissues. *Mol Genet Genomics* 278: 149–165
- Roach MJ, Deyholos MK (2008) Microarray analysis of developing flax hypocotyls identifies novel transcripts correlated with specific stages of phloem fibre differentiation. *Ann Bot* 102: 317–330
- Roach MJ, Mokshina NY, Badhan A, Snegireva AV, Hobson N, Deyholos MK, Gorshkova TA (2011) Development of cellulosic secondary walls in flax fibers requires β -galactosidase. *Plant Physiol* 156: 1351–1363
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498–2504
- Simossis VA, Heringa J (2005) PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res* 33: W289–W294
- Simossis VA, Kleinjung J, Heringa J (2005) Homolog-extended sequence alignment. *Nucleic Acid Res* 33: 816–824
- Takenaka Y, Kato K, Ogawa-Ohnishi M, Tsuruhama K, Kajiura H, Yagyū K, Takeda A, Takeda Y, Kunieda T, Hara-Nishimura I, Kuroha T, Nishitani K, Matsubayashi Y, Ishimizu T (2018) Pectin RG-I rhamnosyltransferases represent a novel plant-specific glycosyltransferase family. *Nat Plants* 4: 669–676. <https://doi.org/10.1038/s41477-018-0217-7>
- Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012)

- Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7: 562–578
- Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res* 44: W232–W235
- Tzfadia O, Diels T, De Meyer S, Vandepoele K, Aharoni A, Van de Peer Y (2016) CoExpNetViz: comparative co-expression networks construction and visualization tool. *Frontiers in Plant Sci* 6: 1194
- Vincken JP, Schols HA, Oomen RJ, McCann MC, Ulvskov P, Voragen AG, Visser RG (2003) If homogalacturonan were a side chain of rhamnogalacturonan I. Implications for cell wall architecture. *Plant Physiol* 132: 1781–1789
- Wang Y (2013) Locally duplicated ohnologs evolve faster than nonlocally duplicated ohnologs in *Arabidopsis* and rice. *Genome Biol Evol* 5: 362–369
- Wang ZW, Hobson N, Galindo L, Zhu S, Shi D, McDill J, Yang L, Hawkins S, Neutelings G, Datla R, Lambert G, Galbraith DW, Grassa CJ, Gerald A, Cronk QC, Cullis C, Dash PK, Kumar PA, Cloutier S, Sharpe AG, Wong GK, Wang J, Deyholos MK (2012) The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. *Plant J* 72: 461–473
- Yapo BM (2011) Pectic substances: from simple pectic polysaccharides to complex pectins. A new hypothetical model. *Carbohydr Polym* 86: 373–385
- Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9: 40

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. List of RGL sequences from some species of fungi, bacteria, and plants that were used for phylogenetic tree construction.

Table S2. List of genes chosen for validation using qPCR, and sequences of primers.