



Article

Genetic Determinants of Fiber-Associated Traits in Flax Identified by Omics Data Integration

Alexander Kanapin ¹, Tatyana Rozhmina ², Mikhail Bankin ³ , Svetlana Surkova ³ , Maria Duk ^{3,4}, Ekaterina Osyagina ³ and Maria Samsonova ^{3,*}

¹ Centre for Computational Biology, Peter the Great St. Petersburg Polytechnic University, 195251 St. Petersburg, Russia

² Laboratory of Breeding Technologies, Federal Research Center for Bast Fiber Crops, 172002 Torzhok, Russia

³ Mathematical Biology & Bioinformatics Laboratory, Peter the Great St. Petersburg Polytechnic University, 195251 St. Petersburg, Russia

⁴ Theoretical Department, Ioffe Institute, 194021 St. Petersburg, Russia

* Correspondence: m.g.samsonova@gmail.com; Tel.: +7-812-290-9645

Abstract: In this paper, we explore potential genetic factors in control of flax phenotypes associated with fiber by mining a collection of 306 flax accessions from the Federal Research Centre of the Bast Fiber Crops, Torzhok, Russia. In total, 11 traits were assessed in the course of 3 successive years. A genome-wide association study was performed for each phenotype independently using six different single-locus models implemented in the GAPIT3 R package. Moreover, we applied a multivariate linear mixed model implemented in the GEMMA package to account for trait correlations and potential pleiotropic effects of polymorphisms. The analyses revealed a number of genomic variants associated with different fiber traits, implying the complex and polygenic control. All stable variants demonstrate a statistically significant allelic effect across all 3 years of the experiment. We tested the validity of the predicted variants using gene expression data available for the flax fiber studies. The results shed new light on the processes and pathways associated with the complex fiber traits, while the pinpointed candidate genes may be further used for marker-assisted selection.

Keywords: flax; GWAS; fiber; GEMMA; gene expression



Citation: Kanapin, A.; Rozhmina, T.; Bankin, M.; Surkova, S.; Duk, M.; Osyagina, E.; Samsonova, M. Genetic Determinants of Fiber-Associated Traits in Flax Identified by Omics Data Integration. *Int. J. Mol. Sci.* **2022**, *23*, 14536. <https://doi.org/10.3390/ijms232314536>

Academic Editor: Bartolome Sabater

Received: 27 September 2022

Accepted: 15 November 2022

Published: 22 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Flax fibers are used as a textile raw material for production of cords, weaving yarn, fashionable garments and high-quality fabric upholstery [1,2]. Since the 1930s, flax fibers have become an important reinforcement component in composite materials due to its unique mechanical properties [3,4]. Currently, composites made of flax fibers are widely used in the automotive industry [5]. The average mechanical performance of flax fibers is very close to that of glass, however, contrary to glass-based composites, flax fibers have reduced environmental impact when incorporated into composite materials. [5].

An elementary fiber is an individual sclerenchyma cell with an exceptionally thick cell wall and an extreme length to diameter ratio (above 1000) [4]. The cells originate from the apical meristem and develop in two stages: elongation and wall thickening. Initially, the elongation of flax fibers proceeds synchronously with neighboring cells; however, after several hours an intrusive growth begins, leading to fibers that reach several centimeters in length. The cell wall thickening includes the deposition of several new layers from the inside of the primary cell wall, and it begins when elongation stops. The first additional layer (so-called S layer) is very thin, while the further deposited layers are characterized by high cellulose content, axial orientation of cellulose microfibrils, almost complete absence of xylan and lignin and the presence of rhamnogalacturonan-I [6–8]. Flax fibers are gathered in bundles located at the periphery of the stalk, providing for its reinforcement [1,9].

Each bundle is composed of 15–50 elementary fibers embedded in a gel-like matrix and joined together.

A technical quality of the fiber is a complex trait defined by plant genotype, environment and after-harvest processing. The fiber traits manifest a significant degree of variation between cultivars [3,4,10]. The most important morphological traits associated with the fiber quality are the following: (a) technical stem length (from the bottom of the stem to the first floral ramification), (b) number of internodes and (c) stem diameter. Flax cultivars with a longer technical stem have compact, dense bast bundles constructed of long elementary fibers. Plants with long internodes and a small number of leaves have a higher fiber quality, as the fiber bundles are torn at the attachment point of the leaf blades. Cultivars with thicker stems have sparse and enlarged fiber bundles with well-developed woody cores, thus producing coarse and less flexible fiber with low spinning quality. The cylindrical shape of the stalk indicates a uniform distribution of the elementary fibers along its length, which ensures homogeneous fiber maturation and even color distribution [10].

There are two common approaches to harness phenotypic variability for dissection of the genetic architecture of agronomic traits. Linkage mapping is based on the analysis of progeny from biparental crosses, and its power predominantly relies on recombination events taking place in the F1 generation. This limits the practical application of linkage mapping for breeding, as the estimated marker effects are specific to the same or genetically related populations and are often not transferable to other genetic backgrounds [11]. GWAS, on the contrary, exploits natural variation and the long history of recombination events captured in large germplasm collections. Inbred plant species such as flax have the practical effect of immortalizing the genotype, providing an unlimited supply of identical seeds that can be grown and assayed in replications for a variety of traits and under various conditions.

Nevertheless, in spite of the increase in size of GWAS populations, most GWAS analyses still have insignificant statistical power to detect associations with alleles of small effect and low frequency that collectively explain a large fraction of the trait heritability. One of the approaches to improve the power of GWAS is to combine the multiple related traits in a multivariate framework [12]. Multitrait GWAS increases statistical power by accounting for trait correlations and pleiotropic effects of polymorphisms, as well as by aggregating weak genetic effects across traits [13]. Additional advantages of multitrait GWAS in comparison with a standard univariate approach lie in its ability to complement missing information in one of phenotypes with the other phenotypes and to capture indirect genetic effects when an SNP affects one phenotype through its effect on the other functionally related phenotype [14,15].

Recently, numerous studies on dissecting the genetic architecture of agronomic traits in flax have been reported. Most of these analyses interrogated traits related to yield, phenology and fatty acid content [16–24]. Regions associated with plant height were also identified in several analyses [17,20,25,26]. Other fiber-related traits received less attention. The current information on genomic regions associated with fiber quality and content is not only limited but is also quite inaccurate, as the studies employ reduced representation sequencing methods and, in most cases, use the biparental crosses. Several studies analyzing RIL populations from crosses between linseed and fiber flax parents report QTLs associated with the following phenotypes: fiber yield and fiber content [25], fiber technical length [26], straw weight and cell wall [27]. Another study of 224 flax samples and 34,932 SFAF markers used GLM and MLM models and identified four SNP loci, of which two were associated with fiber percentage and two with fiber technical length [28].

In addition to GWAS, several high-throughput transcriptomics studies suggested candidate genes involved in fiber-intrusive elongation and tertiary cell wall deposition [8,29,30]. However, these studies were confined to a limited number of cultivars that preclude generalization due to varietal variation. Therefore, to gain an ultimate understanding of mechanisms defining fiber quality, and to prioritize genes for further in-depth experiments,

it is important to make the most of the available genomic and genetic resources, as well as the analytical approaches such as GWAS.

Here, we explore the genetic control of fiber traits by mining 306 flax accessions maintained at the Federal Research Centre of the Bast Fiber Crops (Tables S12 and S13). The dataset was whole-genome sequenced at 10× depth coverage to characterize genetic diversity and population structure, as well as thoroughly phenotyped [10,31]. An in-depth characterization of genetic regions controlling fiber traits could provide a foundation for further genomic selection and breeding efforts aimed at the development of cultivars with superior fiber quality.

2. Results and Discussion

2.1. Phenotypic Variation

All traits showed abundant variation among 306 accessions (Table S1, Appendix A), substantiating GWAS application for detection of genomic regions involved in fiber control. The most variable traits are weight-related characteristics: under three different plantings (in 2019 and two times in 2020), the maximum coefficient of variation for fiber weight (FW) was 54.7% and for technical part length (TW) 43.1%. The variations in stem slenderness (SSI) and tapering (STI) indices were also among the highest, reaching 35.3% and 36.5% in 2019Y, respectively. Compared to variation in indices values, the variation coefficient for elementary fiber length (EFL) was about two times smaller, ranging between years from 12.2% to 15.8%.

Within planting comparisons, we observed statistically significant superiority of most fiber traits when fiber genotypes were compared with linseed genotypes and fiber cultivars were compared with landraces. Non-significant differences were also observed in juxtaposition of fiber cultivars and landraces for stem diameter (SD) in 2020Y-1 and 2020Y-2, distance between internodes (DI) in 2020Y-1 and stem slenderness index (SSI) in 2019Y (Table S2).

Most fiber-related traits are correlated (Table S3, Figure 1). A notable negative correlation was observed between stem diameter (SD) and fiber content (FC). The highest positive correlation was detected in pairwise comparisons between elementary fiber length (EFL), technical stem length (TL), plant height (PH) and the number of internodes (NI). When estimating trait correlation between different plantings, no correlation was revealed for stem tapering index (STI) and distance between internodes (DI), with small positive correlation for stem diameter (SD) and moderate positive correlation for technical part weight (TW) (Table S4). All other traits had between-planting correlation coefficient values higher than 0.7.

2.2. Association Mapping of Fiber Traits

After SNP calling and filtering with thresholds for an MAF of 0.05 and a call rate of 0.85, a total of 72,526 variants remained. Principal component analysis did not produce a clear separation of linseed and fiber genotypes (not shown).

A univariate GWAS analysis was performed using the GLM, MLM, CMLM, FarmCPU, SUPER and Blink models implemented in the GAPIT3 R package [32]. Proceeding from the GAPIT estimation of Bayesian information content, we retained five principal components to be used in GWAS. The six aforementioned methods yielded 1059, 2102 and 3170 QTNs in 2019, 2020Y-1 and 2020Y-2 plantings (FDR adjusted *p*-value 0.05, see Table S5). All models performed well in controlling population structure and family relatedness, as *p*-values deviated from the expected values at the end of distribution in all Q-Q-plots. All QTNs were discovered by at least two models. A total of 873 QTNs were detected in two years (Table S6), while 4 QTNs were found in all plantings. A total of 717 QTN detected in at least two plantings were pleiotropic. Among pleiotropic QNTs, the most often found association was with plant height, being found in 346 of 717 cases. The majority of the identified QTNs demonstrated significant allelic effect across all three plantings, as is shown in Figure 2,

Figures S1 and S2 for plant height (PH). Negative alleles decrease the plant height, while positive alleles increase it.

Multitrait analysis was conducted with GEMMA [33] package for five traits, namely, plant height (PH), technical stem length (TL), number of internodes (NI), fiber content (FC) and fiber weight (FW). The phenotype values have been normalized with quantile normalization for each phenotype, and SNPs were filtered with the following thresholds: MAF 0.05 and call rate 0.85. We used linear mixed model and score test for GEMMA calculation, and the threshold for false discovery rate was set at 0.05. Finally, 162 QTNs were found, which occur in all three plantings and satisfy the FDR threshold (Table S7).

Corrected to chromosome-specific LD, several QTNs detected by GEMMA and GAPIT models coincided (Table S8a). With whole-genome sequencing information and advanced statistical methods at hand, we were able to achieve higher precision in QTL detection in comparison with previous studies based on the reduced representation sequencing methods, identification of SNPs without a reference sequence and biparental crosses. Nevertheless, several QTNs detected by us were in close proximity to QTLs predicted from biparental crosses [25–27] (Table S8b), providing for general agreement between the studies.

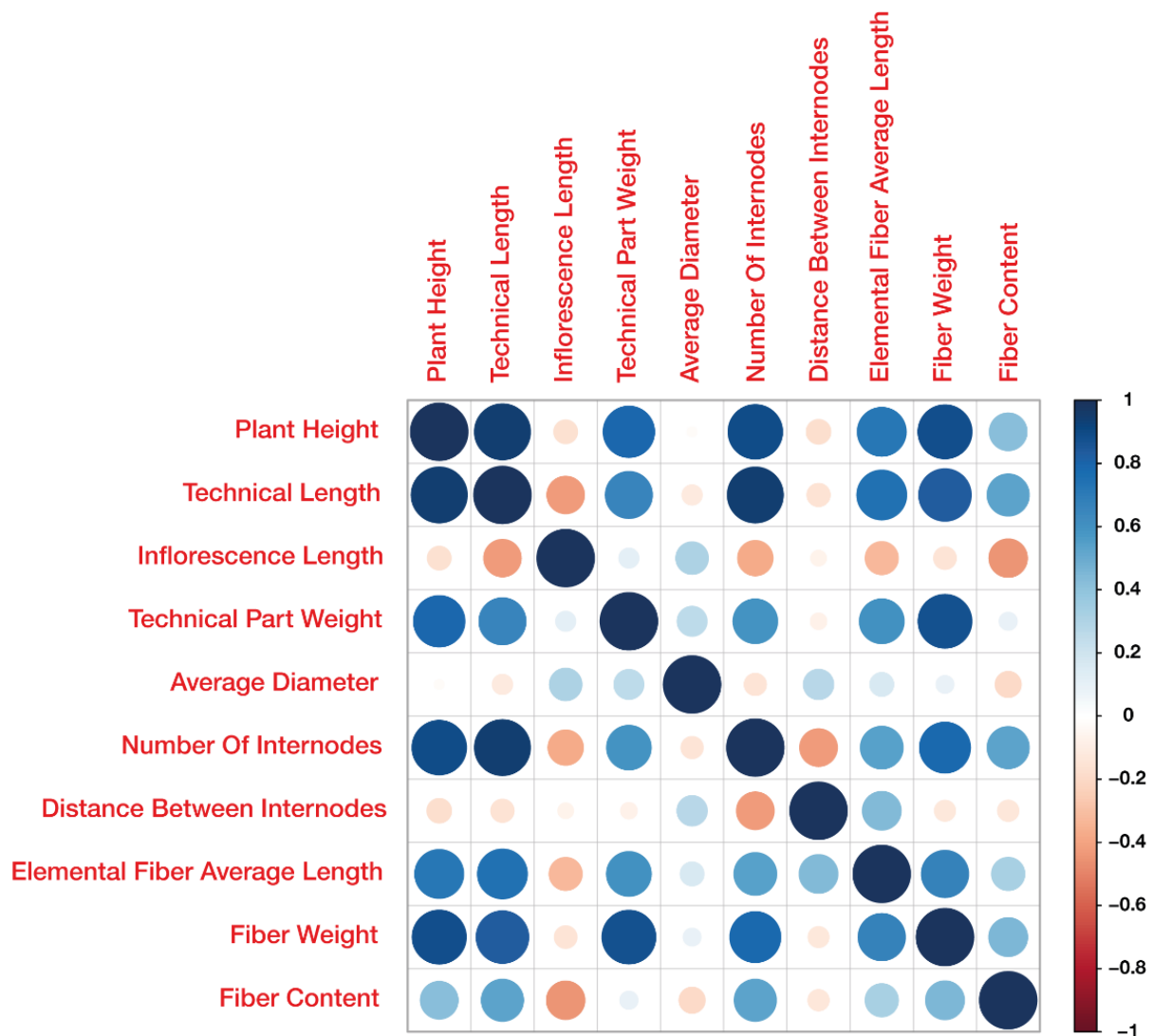


Figure 1. Pearson correlation between traits in 2019Y.

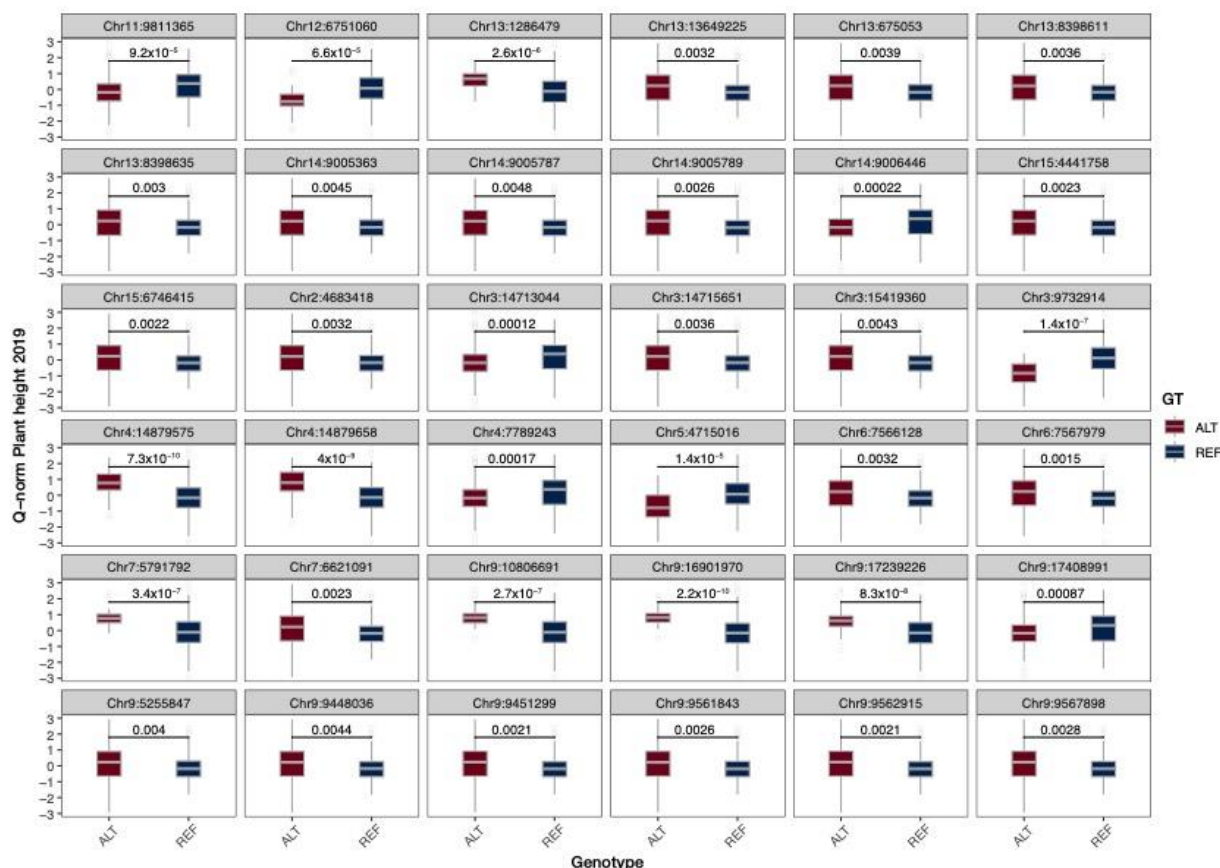


Figure 2. Box plots of allelic effects for representative QTNs associated with plant height in 2019. Numbers are p -values according to the Mann–Whitney–Wilcoxon non-parametric test.

2.3. Fiber-Related Candidate Genes

A two-stage procedure was introduced to pinpoint fiber-related candidate genes. First, we defined intervals centered on stable QTNs detected in at least two plantings and set according to LD decay estimated for each chromosome (Table S9). We called such intervals QTN loci. In total, we found 1645 genes and 1105 genes in QTN loci detected with GAPIT3 and GEMMA, respectively.

Next, we checked the expression of these genes using transcriptome libraries from four flax tissues: bast fibers isolated at the stage of intrusive elongation (iFIB) or at the stage of tertiary cell wall deposition (tFIB), cortical parenchyma (cPAR) and the xylem part of the stem (sXYL) [8,29]. The genes specifically expressed in fibers were determined by contrasting data at the stage of intrusive elongation versus cortical parenchyma and at the stage of tertiary cell wall deposition versus the xylem part of the stem. The dynamics of gene expression during fiber development were interrogated by contrasting data at the stages of intrusive elongation and tertiary cell wall deposition. A total of 206 QTNs detected with GAPIT3 and 97 variants detected with GEMMA were associated with at least one gene differentially expressed in these comparisons (Tables S10 and S11). Figure 3 presents the number of differentially expressed genes in QTN regions. Genes with $|\log_2FC| \geq 2$, $padj0.01 \leq 0.01$ were selected for further annotation based on the functions of orthologous genes in *Arabidopsis* and other plant species. Proceeding from annotation, the candidate genes fell into five groups: (1) genes related to the synthesis and modification of cell wall components; (2) genes controlling cell fate determination, growth and elongation; (3) genes associated with vesicular transport, intracellular trafficking of the cell wall components and solute transport across membranes; (4) genes involved in hormonal regulation; (5) genes associated with resistance to plant diseases (Tables 1 and 2).

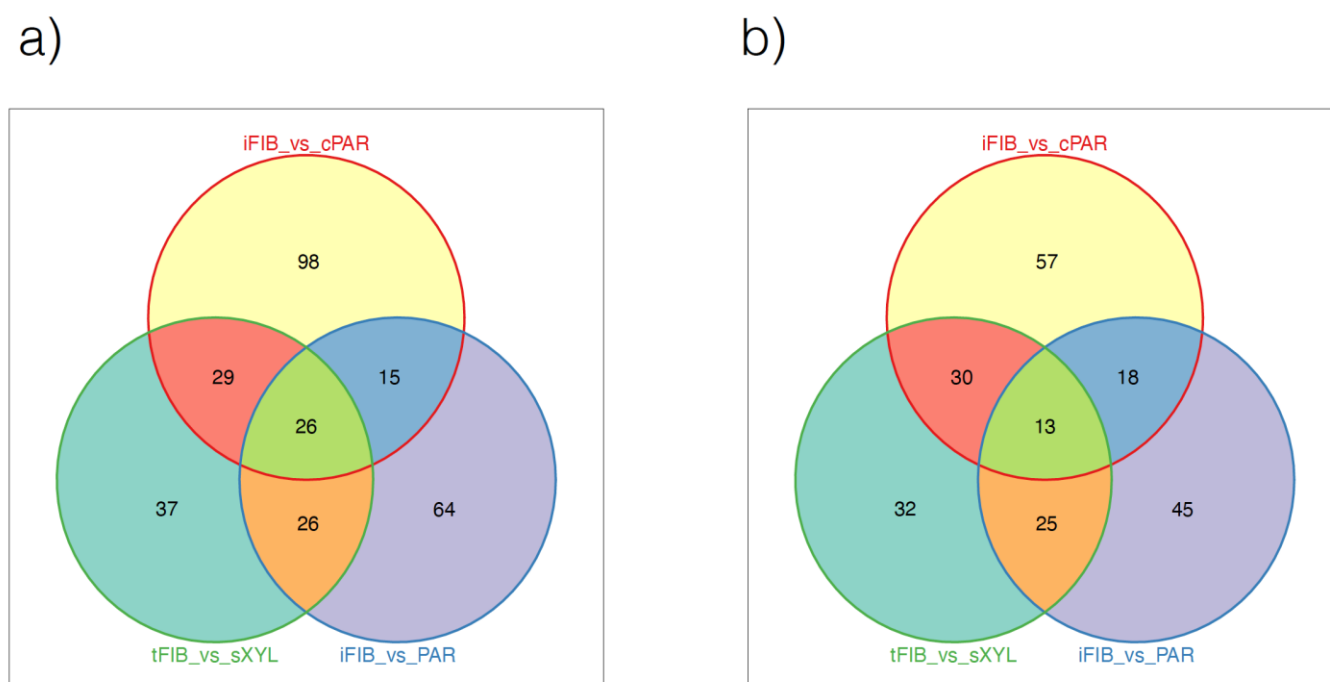


Figure 3. Venn diagram representing the number of differentially expressed genes in QTN loci detected with GAPIT (a) and GEMMA (b).

Table 1. Candidate genes predicted from the intersection of GAPIT3 R hits with data on gene expression in fibers.

QTN	Gene	Functional Annotation	<i>Arabidopsis</i> Gene Name	Reference	Predicted Function
Chr1:27748943	<i>Lus10010138</i>	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	<i>AT4G01240</i>	[34]	methylation and demethylation of cell wall proteins
Chr2:261056	<i>Lus10005814</i>	RING/U-box superfamily protein	<i>AT1G49230</i>	[35]	signaling proteins for receptor kinases, regulation of hormone signaling and abiotic/biotic stress response
	<i>Lus10005815</i>	RING/U-box superfamily protein	<i>AT1G49230</i>		
	<i>Lus10005818</i>	RING/U-box superfamily protein	<i>AT1G49230</i>		
Chr2:2513654	<i>Lus10016317</i>	cell wall/vacuolar inhibitor of fructosidase 1	<i>AT1G47960, ATC/VIF1, C/VIF1</i>	[36,37]	regulation of ABA response and salt tolerance
Chr2:4360628	<i>Lus10020193</i>	auxin efflux carrier family protein	<i>AT5G16530, PIN5</i>	[38,39]	regulates intracellular auxin homeostasis and metabolism
Chr2:4683418	<i>Lus10013618</i>	RING/U-box superfamily protein	<i>AT5G17600</i>	[35]	signaling proteins for receptor kinases, regulation of hormone signaling and abiotic/biotic stress response
	<i>Lus10013631</i>	peroxidase superfamily protein, class III	<i>AT5G1782, AtperoxP57</i>	[40,41]	cell wall localized proteins tightly associated to its loosening and stiffening
Chr2:4872459	<i>Lus10013655</i>	HSP20-like chaperones superfamily protein	<i>AT1G54400</i>	[42]	enhances membrane stability and detoxifies ROS by regulating the antioxidant enzymes system
Chr2:6885020	<i>Lus10038682</i>	major facilitator superfamily protein	<i>AT2G39210</i>	[43]	facilitates movement of small solutes across cell membranes

Table 1. Cont.

QTN	Gene	Functional Annotation	Arabidopsis Gene Name	Reference	Predicted Function
Chr2:25126362	<i>Lus10006462</i>	cellulase (glycosyl hydrolase family 5) protein	<i>AT1G13130</i>	[44]	hydrolyses the glycosidic bond
Chr3:4704531	<i>Lus10019171</i>	microtubule-associated proteins 70-2	<i>AT1G24764, ATMAP70-2</i>	[45]	outlining tracks of cellulose production during SCW formation and delivery of cell wall components
Chr3:5077352	<i>Lus10007972</i>	UDP-glucosyl transferase 89B1	<i>AT1G73880, UGT89B1</i>	[46]	
Chr3:5883916	<i>Lus10040598</i>	high-affinity K ⁺ transporter 1	<i>AT4G10310, ATHKT1</i>	[47]	reducing Na ⁺ toxicity through K ⁺ uptake
Chr3:17134662	<i>Lus10033485</i>	plant glycogenin-like starch initiation protein 1	<i>AT3G18660, GUX1, PGSIP1</i>	[48]	glucuronyl transferase responsible for the addition of GlcA residues onto xylan and for secondary wall deposition
Chr3:17878746	<i>Lus10007051</i>	peroxidase superfamily protein, class III	<i>AT2G4148, AtperoxP25</i>	[40,41]	cell wall localized proteins tightly associated to its loosening and stiffening
Chr3:25379963	<i>Lus10037568</i>	homeobox-leucine zipper family protein/lipid-binding START domain-containing protein	<i>AT1G52150, ATHB15, CNA, ICU4</i>	[49]	confers positional information which is required to establish the number and pattern of vascular bundles in the stem, and is involved in repression of secondary cell wall development in <i>Arabidopsis</i>
Chr3:26449997	<i>Lus10002912</i>	integrase-type DNA-binding superfamily protein	<i>AT5G25390, SHN2</i>	[50]	TF, regulates floral organs' epidermal cell elongation by affecting pectin metabolism
Chr4:184242	<i>Lus10030156</i>	glucose-1-phosphate adenylyltransferase family protein	<i>AT2G39770, CYT1, EMB101, GMP1, SOZ1, VTC1</i>	[51]	provides GDP-mannose, which is used for cell wall carbohydrate biosynthesis, including lignin and protein glycosylation
Chr4:466800	<i>Lus10030216</i>	cytochrome P450, family 78, subfamily A, polypeptide 6	<i>CYP78A6, EOD3</i>	[52–54]	Cell expansion, enhancer of DA-1 3 and seed and fruit development
Chr4:647773	<i>Lus10030255</i>	DYNAMIN-like 1B	<i>AT3G61760, ADL1B</i>	[55]	membrane trafficking, essential for plant cytokinesis and polarized cell growth
Chr4:1229010	<i>Lus10030366</i>	alpha/beta-hydrolases superfamily protein	<i>AT3G62860</i>	[56]	breaking of carbon–carbon bonds
Chr4:2763834	<i>Lus10029506</i>	cation/hydrogen exchanger 28	<i>AT3G52080, chx28</i>	[57]	cation/hydrogen exchange
Chr4:7789243	<i>Lus10039564</i>	F-box/RNI-like superfamily protein	<i>AT5G67140</i>	[58]	SCF ubiquitin ligase complex
Chr4:14879575 Chr4:14879658	<i>Lus10041581</i>	lysine histidine transporter 1	<i>AT5G40780, LHT1</i>	[30]	solute transporter
Chr4:15210519	<i>Lus10041644</i>	fucosyltransferase 1	<i>AT2G03220, ATFT1, ATFUT1, FT1, MUR2</i>	[59]	xyloglucan synthesis
	<i>Lus10041651</i>	cinnamoyl CoA reductase 1	<i>AT1G15950, ATCCR1, IRX4</i>	[60]	lignin biosynthesis
Chr7:14140267	<i>Lus10025431</i>	RAC-like 2	<i>AT5G45970, ATRAC2, ATROP7</i>	[61]	secondary cell wall of xylem vessels
Chr8:1099671	<i>Lus10013589</i>	syntaxin of plants 121	<i>AT3G11820, ATSYR1, PEN1</i>	[62]	secretory traffic to the plasma membrane

Table 1. Cont.

QTN	Gene	Functional Annotation	Arabidopsis Gene Name	Reference	Predicted Function
Chr9:693002 Chr9:695025	<i>Lus10010153</i>	NOD26-like intrinsic protein 5;1	<i>AT4G1038, NIP5;1, NLM6, NLM8</i>	[63]	aquaporin
	<i>Lus10010190</i>	serine carboxypeptidase-like 34	<i>AT5G23210, SCPL34</i>	[64]	present in cell wall of elongating cells
Chr9:1929944	<i>Lus10008974</i>	beta-galactosidase 12	<i>AT4G26140, BGAL12</i>	[65]	plant cell wall remodeling
Chr9:5255847	<i>Lus10004364</i>	alpha/beta-hydrolases superfamily protein	<i>AT4G16820, PLA-I[beta]2</i>	[56]	breaking of carbon-carbon bonds
	<i>Lus10004377</i>	homeobox protein 2	<i>AT4G1678, ATHB2, HAT4, HB-2</i>	[66]	cell expansion and cell proliferation in the response to auxin
Chr9:8233979	<i>Lus10007455</i>	disease resistance protein (TIR-NBS-LRR class) family	<i>AT3G4448, cog1, RPP1</i>	[67]	plant immunity
Chr9:11359847	<i>Lus10019631</i>	alpha/beta-hydrolases superfamily protein	<i>AT1G72620</i>	[56]	breaking of carbon-carbon bonds
Chr9:17239226	<i>Lus10042571</i>	aluminum-activated malate transporter family protein	<i>AT5G46600</i>	[68]	transporter
Chr9:17408991	<i>Lus10042531</i>	NAC domain-containing protein 47	<i>AT3G0407, NAC047, SHYG</i>	[69]	NAC TF, localized longitudinal cell expansion
	<i>Lus10042537</i>	malectin/receptor-like protein kinase family protein	<i>AT3G04690, ANX1</i>	[70]	perception and relay processes at cell membranes
Chr9:18936691	<i>Lus10024808</i>	WUSCHEL-related homeobox 4	<i>AT1G46480, WOX4</i>	[71]	regulates the dynamic balance of division and differentiation of plant stem cells
	<i>Lus10024818</i>	cytochrome P450, family 81, subfamily D, polypeptide 5	<i>AT4G3732, CYP81D5</i>	[72]	defense response
Chr9:19858075	<i>Lus10000494</i>	flavin-binding monooxygenase family protein	<i>AT4G2872, YUC8</i>	[73]	indole-3-pyruvate monooxygenase YUCCA8, responsible for auxin biosynthesis in leaves
Chr9:20032070- Chr9:20032149	<i>Lus10011868</i>	CLAVATA3/ESR-RELATED 41	<i>AT3G24770, CLE41</i>	[74]	regulates vascular cell division, vascular organization and xylem differentiation in vascular tissue
	<i>Lus10011858</i>	cytochrome B5 isoform B	<i>AT2G32720, CB5-B</i>	[75]	lignin biosynthesis
Chr12:16047776	<i>Lus10043075</i>	O-glycosyl hydrolase family 17 protein	<i>AT5G20870</i>	[76]	hydrolytic activity towards different types of hemicelluloses or callose
Chr13:1286479	<i>Lus10010665</i>	UDP-glycosyltransferase superfamily protein	<i>AT1G2240, ATUGT85A1, UGT85A1</i>	[77,78]	cell cycle regulation, influences trans-zeatin (cytokinin) homeostasis likely through O-glycosylation
	<i>Lus10010666</i>	annexin 8	<i>AT5G12380, ANNAT8</i>	[79]	cell wall protein
Chr13:3759775	<i>Lus10026123</i>	hydroxycinnamoyl-CoA shikimate/quinic acid hydroxycinnamoyl transferase	<i>AT5G48930, HCT</i>	[80]	lignin biosynthesis
Chr15:1632466	<i>Lus10013935</i>	alpha/beta-hydrolases superfamily protein	<i>AT4G18550</i>	[56]	breaking of carbon-carbon bonds
	<i>Lus10013936</i>	alpha/beta-hydrolases superfamily protein			
	<i>Lus10013945</i>	O-methyltransferase family protein	<i>AT4G35160</i>		transfer of methyl groups to various biomolecules

Table 2. Candidate genes predicted from the intersection of GEMMA hits with data on gene expression in fibers.

QTN	Gene	Functional Annotation	<i>Arabidopsis</i> Gene Name	Reference	Predicted Function
Chr2:734912 Chr2:748984 Chr2:752999	<i>Lus10019240</i>	glycosyl hydrolase family protein	<i>AT5G10560</i>	[76]	hydrolysis of glycosidic bonds
	<i>Lus10019273</i>	basic leucine-zipper 6	<i>AT2G22850, AtbZIP6, bZIP6</i>		transcription factor
Chr2:3633762 Chr2:3633790	<i>Lus10003436</i>	glutamate receptor 2.7	<i>AT2G29120, GLR2.7</i>	[81]	non-selective cation channel
Chr2:6831421	<i>Lus10038682</i>	major facilitator superfamily protein	<i>AT2G39210</i>	[43]	facilitates movement of small solutes across cell membranes
Chr2:17384359 Chr2:17384362	<i>Lus10033212</i>	RAD-like 6	<i>AT1G75250, RL6, RSM3</i>		MYB family TF
Chr3:12697091	<i>Lus10005850</i>	aldolase superfamily protein	<i>AT4G26530</i>	[82]	modulates V-ATPase-dependent vesicular trafficking events and actin cytoskeleton remodeling
Chr4:460524	<i>Lus10030216</i>	cytochrome P450, family 78, subfamily A, polypeptide 6	<i>AT2G46660, CYP78A6, EOD3</i>	[52–54]	cell expansion, enhancer of DA-13, seed and fruit development
	<i>Lus10029458</i>	OBF-binding protein 3	<i>AT3G55370, OBP3</i>	[83]	TF, pays a role in plant growth and development, induced by SA and auxin
Chr4:2376445	<i>Lus10029464</i>	nodulin MtN21/EamA-like transporter family protein	<i>AT2G39510</i>	[84–86]	amino acid and auxin transporters
	<i>Lus10029065</i>	peroxidase superfamily protein, class III	<i>AT5G05340, AtperoxP52</i>	[40,41]	cell wall localized proteins tightly associated to its loosening and stiffening
Chr4:9042565	<i>Lus10029063</i>	major facilitator superfamily protein	<i>AT2G40460</i>	[43]	facilitates movement of small solutes across cell membranes
Chr4:17106278	<i>Lus10004043</i>	myb domain protein 20	<i>AT1G66230</i>	[87]	activates lignin and phenylalanine biosynthesis genes during secondary wall formation
Chr4:18493683	<i>Lus10030007</i> <i>Lus10030008</i>	glycosyl hydrolase 9B8	<i>AT2G32990, GH9B8</i>	[88]	membrane endo (1→4)-β-D-glucanase (cellulase), wall assembly and cell elongation, fruit ripening and floral abscission
Chr8:4371945 Chr8:4833196	<i>Lus10034491</i>	major facilitator superfamily protein	<i>AT4G34950</i>	[43]	facilitates movement of small solutes across cell membranes
	<i>Lus10023931</i>	plant regulator RWP-RK family protein	<i>AT2G17150</i>	[89]	control of cell differentiation
Chr8:18364995	<i>Lus10007808</i> , <i>Lus10007809</i> , <i>Lus10007810</i> , <i>Lus10007811</i> , <i>Lus10007813</i> , <i>Lus10007814</i>	disease resistance protein (TIR-NBS-LRR class) family	<i>AT5G36930</i>	[90]	plant immunity
Chr8:19022659	<i>Lus10002243</i>	subtilase family protein	<i>AT1G04110, SDD1</i>	[91]	cell wall enzyme
	<i>Lus10007528</i> , <i>Lus10007529</i>	TRICHOME BIREFRINGENCE-LIKE 33	<i>AT2G40320, TBL33</i>	[92]	contribute to the synthesis and deposition of secondary wall cellulose, presumably by influencing the esterification state of pectic polymers
Chr9:971861	<i>Lus10007532</i>	laccase 5	<i>AT2G40370, LAC5</i>	[93]	lignin degradation and detoxification of lignin-derived product

Table 2. Cont.

QTN	Gene	Functional Annotation	Arabidopsis Gene Name	Reference	Predicted Function
Chr9:7152630	<i>Lus10000957</i>	expansin-like B1	<i>AT4G17030, ATEXLB1, ATEXPR1, ATHEXPBETA 3.1, EXLB1, EXPR</i>	[79]	cell wall loosening
Chr9:7487990 Chr9:7488554	<i>Lus10024499</i>	IQ-domain 22	<i>AT4G23060, IQD22</i>	[94,95]	Ca ²⁺ /CaM-regulated scaffold for cellular transport of specific cargo along microtubules
	<i>Lus10024485</i>	KNOTTED-like homeobox of <i>Arabidopsis thaliana</i> 7	<i>AT1G62990, IXR11, KNAT7</i>	[96]	Secondary-wall-associated transcription factor
Chr9:8240615 Chr9:8240618	<i>Lus10007455</i>	disease resistance protein (TIR-NBS-LRR class) family	<i>AT3G44480, cog1, RPP1</i>	[90]	plant immunity
Chr9:9241973	<i>Lus10006152</i>	cytochrome P450, family 90, subfamily D, polypeptide 1	<i>AT3G13730, CYP90D1</i>	[72]	modification of cyclic terpenes and sterols in the brassinosteroid, abscisic acid and gibberellin pathways
Chr9:17172377 Chr9:17190919	<i>Lus10042571</i>	aluminum-activated malate transporter family protein	<i>AT5G46600</i>	[68]	transporter
Chr9:17511839 Chr9:17513033	<i>Lus10042531</i>	NAC domain-containing protein 47	<i>AT3G04070, anac047, NAC047</i>	[69]	transcription factor
	<i>Lus10042516</i>	sugar transporter 1	<i>AT1G11260, STP1</i>	[97]	H ⁺ /monosaccharide cotransporter
Chr9:18121668	<i>Lus10029638</i> <i>Lus10029639</i>	GDSL-like lipase/acylhydrolase superfamily protein	<i>AT1G54790</i>	[98–100]	growth, biotic stress response, in cotton the GDSL (GhGDSL) lipase/hydrolase gene (CotAD_74480) is expressed during SCW biosynthesis
Chr11:11843614	<i>Lus10021553</i>	auxin-responsive family protein	<i>AT3G25290</i>	[101]	TF, bind to AuxRE in the promoters of auxin-regulated genes
Chr12:16474392	<i>Lus10027867</i>	reticulon family protein	<i>AT3G19460</i>	[102]	endoplasmic reticulum-Golgi trafficking, vesicle formation and membrane morphogenesis
Chr13:12832925	<i>Lus10010638</i>	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	<i>AT1G27660, PEA5</i>	[103]	governs the competence of pericycle cells to initiate lateral root primordium formation
Chr13:18496268	<i>Lus10030901</i>	cysteine-rich RLK (RECEPTOR-like protein kinase) 2	<i>AT1G70520, CRK2</i>	[104]	plant immunity
Chr14:3502060	<i>Lus10020525</i>	disease resistance protein (TIR-NBS-LRR class) family	<i>AT4G16950, RPP5</i>	[105]	plant immunity
Chr14:13328874	<i>Lus10032902</i>	HSP20-like chaperones superfamily protein	<i>AT1G54400</i>	[42]	enhances membrane stability and detoxifies ROS by regulating the antioxidant enzymes system
Chr15:124776	<i>Lus10007632</i> <i>Lus10007633</i>	cysteine-rich RLK (RECEPTOR-like protein kinase) 25	<i>AT4G05200, CRK25</i>	[106]	plant immunity

2.4. Candidate Genes Related to the Synthesis and Modification of Cell Wall Components

The flax primary cell wall is composed of cellulose microfibrils embedded in a hydrated matrix of hemicelluloses and pectin [107]. Hemicelluloses include xylans, mannans, mixed linkage β -glucans and xyloglucans. During fiber cell elongation and intrusive growth, the size and morphology of the cell changes driven by a coordinated series of

biochemical reactions result in the biosynthesis and degradation of cell wall components [8]. Numerous enzymes participate in these processes to provide the remodeling of cell wall structure via selective synthesis and degradation of some polysaccharides. A number of candidate genes were predicted to be involved in the synthesis of cell wall carbohydrates. The *Lus10041644* gene within Chr4:15210519 QTN (Table 1) likely encodes fucosyltransferase 1, involved in xyloglucan synthesis [59]. In dicot plants, xylan consists of a linear backbone of $\beta(1,4)$ -linked D-Xyl residues, some of which are substituted with $\alpha(1,2)$ -linked glucuronic acid (GlcA). The *Lus10033485* gene at Chr3:17134662 QTN was predicted to encode plant glycogenin-like starch initiation protein 1, which is glucuronyl transferase, responsible for the addition of GlcA residues onto xylan [48].

Several candidate genes were predicted to participate in the enzymatic modification of the cell wall polysaccharides. For example, *Lus10006462* and *Lus10030007* genes at QTNs Chr2:25126362 (Table 1) and Chr4:18493683 (Table 2), respectively, were predicted to encode endoglucanases capable of hydrolyzing glucans and callose [44,88]. Callose is a plant polysaccharide, which is localized to plasmodesmata and also produced as a temporary cell wall constituent in response to stress [108]. Callose degradation may accompany the intercalation of elongating fibers, as they are simplistically isolated and have no plasmodesmata [109].

Several genes, *Lus10030366*, *Lus10019631*, *Lus10013935* and *Lus10013936* at QTNs Chr4:1229010, Chr9:11359847 and Chr15:1632466 (Table 1), belonged to the alpha/beta-hydrolase protein superfamily, known to break carbon-carbon bounds [56]. The *Lus10043075* at Chr12:16047776 QTN was annotated to encode the O-glycosyl hydrolase family 17 protein with hydrolytic activity towards different types of hemicelluloses or callose [76]. *Lus10008974* at Chr9:1929944 was predicted to encode beta-galactosidase 12 [65]. In the flax tertiary cell wall, long galactan chains of the pectin rhamnogalacturonan-I (RG-I) are located between the cellulose microfibrils, thus preventing their lateral interaction. Beta-galactosidase shortens RG-I, which facilitates the lateral interaction of cellulose microfibrils and leads to maturation of the cell wall [7].

The flax primary cell wall also contains a variety of glycoproteins [110]. The QTN Chr13:1286479 (Table 1) harbored the *Lus10010666* gene, predicted to encode annexin, a soluble protein capable of Ca²⁺-dependent or -independent association with membrane phospholipids [79]. The *Lus10000957* gene (Table 2) within the Chr9:7152630 QTN encodes expansin-like B1, a cell wall loosening protein [79]. Class III peroxidases are cell wall localized proteins that can oxidize cell wall aromatic compounds within proteins and phenolics that are either free or linked to polysaccharides. Thus, they are also tightly associated to cell wall loosening and stiffening [40,41]. *Lus10013631*, *Lus10007051* and *Lus10029065* at QTNs Chr2:4683418 (Table 1), Chr3:17878746 (Table 1) and Chr4:9042565 (Table 2) were predicted to encode class III peroxidase. The Chr8:19022659 (Table 2) QTN was associated with the *Lus10002243* gene, which was annotated as encoding subtilase. Most subtilases are cell wall enzymes involved in the control of growth and development by regulating the properties of the cell wall and the activity of extracellular signaling molecules [91].

In flax fibers, the middle lamella is enriched in pectins, and rhamnogalacturonan-I is one of the constituents of the tertiary wall [8,107]. Pectin is synthesized in the Golgi apparatus, with most of the carboxyl groups methyl esterified, which results in pectin with low charge density. This likely increases the pectin flexibility and mobility in the wall. The transfer of methyl groups to pectin is catalyzed by methyltransferases [107]. *Lus10010138* at Chr1:27748943 QTN and *Lus10013945* at Chr15:1632466 QTN (Table 1) were annotated to encode methyltransferases, of which the former belonged to the S-adenosyl-L-methionine-dependent methyltransferase superfamily [34].

A unique feature that provides strength to flax fibers is the presence of additional cell wall layers; of particular interest are candidate genes, the predicted function of which may be associated with secondary and tertiary wall deposition. The regulatory pathways implicated in secondary wall synthesis are well studied in *Arabidopsis* [111]. In this plant,

the secondary cell wall develops in the xylary elements and fibers of the vascular system to provide mechanical strength to all plant organs. Among candidate genes, the *Lus10037568* gene at Chr3:25379963 QTN (Table 1) was annotated to encode class III homeodomain leucine zipper transcription factor (HD-ZIP III TFs) CORONA (CAN/AtHB150), which regulates the biosynthesis of secondary cell walls in *Arabidopsis* [49,112]. The *Lus10007528* and *Lus10007529* genes within QTN Chr9:971861 (Table 2) are orthologs to the *ATTBL33* gene encoding the TRICHOME BIREFRINGENCE-LIKE 33 protein, contributing to the deposition of secondary wall cellulose in *Arabidopsis* and, therefore, may be implicated in secondary cell wall synthesis in flax [92]. *Lus10024485* at Chr9:7487990 QTN (Table 2) was annotated as an ortholog of *A. thaliana* genes encoding the KNOTTED-like homeobox protein, which is a transcription factor involved in the secondary cell wall synthesis [96]. *Lus10025431* at Chr7:14140267 QTN (Table 1) probably encodes a small GTPase from the RAC/ROP family that is involved in the development of secondary cell walls of xylem vessels in *Arabidopsis* [61].

The flax secondary cell wall is composed of 30% lignin, a polymer made by cross-linking phenolic precursors, while the tertiary cell wall lacks this compound. A number of candidate genes may be involved in the synthesis and degradation of lignin. In particular, *Lus10041651* at QTN Chr4:1521051 and *Lus10026123* at Chr13:3759775 were predicted to encode enzymes of monolignol biosynthesis [80]. The *Lus10004043* gene at Chr4:17106278 (Table 2) encodes myb domain protein 20 that activates the lignin and phenylalanin biosynthesis genes during secondary wall formation [87]. The *Lus10007532* gene within QTN Chr9:971861 (Table 2) encodes laccase 5, which degrades lignin [93].

Lus10029639 at Chr9:18121668 (Table 2) was predicted to encode the GDSL-like lipase/acylhydrolase superfamily protein. GDSL-like lipase/acylhydrolases are enzymes with broad substrate specificity and can hydrolyze many kinds of substrates such as thioesters, aryl esters, phospholipids and amino acids. In cotton, the GDSL lipase/hydrolase gene *CotAD_74480* is expressed during secondary cell wall biosynthesis [98–100].

The orientation of the cellulose fibrils defines mechanical properties of flax secondary and tertiary cell walls. The direction of microfibrils in the wall is determined by movements of the cellulose synthase complex along the plasma membrane, partly guided by elements of the underlying cytoskeleton [107]. The *Lus10019171* gene at Chr3:4704531 QTN (Table 1) annotated as encoding microtubule-associated protein 70-2 may be involved in the control of the cellulose synthase movement [45].

2.5. Candidate Genes Controlling Cell Fate Determination, Growth and Elongation

In the plant embryo, the establishment of the apical–basal axis is followed by formation of the radial axis and finally of bilateral symmetry. The SAM arose between the two out-growing cotyledons [113]. *Lus10004377* at Chr9:5255847 QTN was predicted to encode the HD-Zip II transcription factor [66], which in *Arabidopsis* controls the embryonic apical patterning and the SAM function partly by interacting with HD-Zip III proteins, including ATHB15/CORONA implicated also in the control of the secondary cell wall biosynthesis [49,112]. The HD-Zip III genes are also involved in patterning along the radial and adaxial–abaxial axes of the embryo [113].

In *Arabidopsis* and other seed plants, the lateral roots are mainly derived from the pericycle cells located at the outermost cell layer of the stele along the roots. Plants can form lateral roots flexibly depending on environments due to the ability of the pericycle cells to maintain their cell division potential for a long time. *Lus10010638* at Chr13:12832925 (Table 2) was annotated as encoding a basic helix-loop-helix (bHLH) transcription factor. Its ortholog in *Arabidopsis* *PFB5* governs the competence of the pericycle cells to initiate the formation of the lateral root primordium [103].

Candidate genes *Lus10030216*, *Lus1002912*, *Lus10042531* and *Lus10010190* are probably involved in cell elongation. The *Lus10030216* gene associated with the Chr4:460524 QTN (Table 2) is an ortholog of the *ENOD3* gene, which in *Arabidopsis* encodes cytochrome P450 78A6 and controls seed size [52–54]. *Lus10002912* at the Chr3:26449997 QTN (Table 1)

was predicted to encode the integrase-type DNA-binding superfamily protein, which is a SHINE transcription factor. SHINE transcription factors control the elongation of floral organs' epidermal cells by affecting pectin metabolism [50]. Among their targets are genes from the CYP86A cytochrome P450 protein family. The *Lus10042531* gene at Chr9:17511839 QTN was annotated as the NAC047 transcription factor (Table 2). In *Arabidopsis*, *NAC047* controls the longitudinal cell expansion at the lower (abaxial) side of the leaf petiole [69]. The *Lus10010190* gene within the Chr9:693002 QTN (Table 1) is an ortholog of the *A. thaliana* *AT5G23210* gene for serine carboxypeptidase-like 34. In *Nicotiana tabacum*, the extracellular type III carboxypeptidase controls cell elongation [64]. *Lus10023931* at Chr8:4371945 OTN (Table 2) may be involved in the control of cell differentiation. It was annotated to encode an RWP-RK domain family protein, which in *Arabidopsis* regulates the differentiation of the female gametophyte cells [89].

2.6. Candidate Genes Associated with Vesicular Transport, Intracellular Trafficking of the Cell Wall Components and Solute Transport across Membranes

Fundamental to cell elongation is the supplying of carbohydrates, proteins and other molecules to the new plasma membrane and cell wall. *Lus10013589* at QTN Chr8:1099671 (Table 1) was predicted to encode a SNARE protein critical for the movement of membrane and cargo proteins within the cell [62,114]. *Lus10030255* associated with Chr4:647773 (Table 1) is an ortholog of *AT3G61760* for DYNAMIN-like 1b required for the de novo assembly and maintenance of the plasma membrane during cell plate formation and cell expansion [55]. The *Lus10005850* gene within the Chr3:12697091 QTN (Table 2) likely encodes aldolase, which modulates V-ATPase-dependent vesicular trafficking events and actin cytoskeleton remodeling [82]. The *Lus10024499* gene associated with Chr9:7487990 QTN (Table 2) was annotated as calmodilin-binding protein IQD22, involved in cellular transport of specific cargo along microtubules [94,95]. The *Lus10027867* gene at Chr12:16474392 QTN (Table 2) was an ortholog of *AT3G19460*, encoding reticulon involved in endoplasmic reticulum–Golgi trafficking, vesicle formation and membrane morphogenesis [102].

Numerous QTNs were associated with genes encoding various transporters (Tables 1 and 2). For cell wall expansion, sugar, potassium and malate transporters are of particular importance [115]. The *Lus10029506* gene at Chr4: 2,763,834 QTN (Table 1) was annotated to encode the K^+/H^+ transporter [57]. *Lus10042571* within the Chr9:17172377 QTN likely encodes aluminum malate transporter [68], and the *Lus10042516* gene contained within Chr9:17511839 QTN (Table 2) was predicted to encode sugar transporter [98]. A high concentration of osmotically active solutes results in an influx of water mediated by aquaporins. *Lus10010153* at Chr9:693002 QTN (Table 1) was predicted to encode NOD26-like intrinsic protein 5;1, an aquaporin [63]. Vacuolar invertase has long been considered as a major player in cell expansion. *Lus10016317*, harbored by QTN Chr2:2513654 (Table 1), encodes a cell wall/vacuolar inhibitor of fructosidase 1, an invertase that cleaves sucrose [36,37].

2.7. Candidate Genes Involved in Hormonal Regulation

Plant hormones regulate cell wall expansion and cell growth. Among them, auxin plays a vital role in controlling plant growth and development via promotion of cell proliferation, cell expansion and elongation, as well as cell differentiation [38]. Several candidate genes were either involved in auxin biosynthesis and transport or relayed regulatory signals to their targets via the auxin pathway. *Lus10000494* at QTN Chr9:1985807 (Table 1) likely encodes indole-3-pyruvate monooxygenase YUCCA8, catalyzing the last step in auxin biosynthesis, the oxidation of indole-3-pyruvic acid (IPA) to indole-3-acetic acid (IAA) [73]. *Lus10020193* at QTN Chr2:4360628 (Table 1) encodes the auxin efflux carrier family protein [38,39], while *Lus10029464* encodes the nodulin MtN21/EamA-like transporter family protein involved, in addition to auxin, in amino acid transport [84–86]. *Lus10029458*, *Lus10021553* (Table 2) and *Lus10004377* (Table 1) genes were predicted to encode transcription factors inducible by auxin [83,101]. *Lus10029458* was annotated as encoding a Dof transcription factor OBP3, which plays important roles in plant growth and

development. In *Arabidopsis*, the expression level of the OBP3 protein increased following treatment with auxin, and transgenic plants overexpressing OBP3 showed a severe growth defect with altered root development and yellowish leaves [83]. The *Lus10004377* gene associated with Chr9:5255847 (Table 1) was predicted to encode HD-ZIP II transcription factor ATHB2, an early auxin-inducible gene in both *Arabidopsis* and wheat [66,113].

Cytokinins are a class of plant hormones that promote cell division in plant roots and shoots. The *Lus10010665* gene at the Chr13:1286479 QTN (Table 1) was predicted to encode UDP-glycosyltransferase 85A1, involved in cell cycle regulation by influencing a homeostasis of trans-zeatin, a cytokinin [77,78].

2.8. Candidate Genes Associated with Resistance to Plant Diseases

Plant cell wall alteration directly influences growth and stress response pathways [116]. This is especially relevant during plant responses to external stimuli. It is not surprising, therefore, that we found several abiotic-stress- and pathogenesis-related genes associated with QTNs for fiber traits. For example, a cluster of six genes, *Lus10007808–Lus10007814* within Chr8:18364995 QTN (Table 2), the *Lus10007455* gene at Chr9:8240615 QTN and the *Lus10020525* gene at Chr14:3502060 QTN (Table 2) were predicted to encode disease resistance TIR-NBS-LRR proteins [90].

Protein regulation is an important system to accomplish adaptation to different abiotic and biotic stresses. In plants, the ubiquitin-mediated system regulates various proteins in response to developmental signals and different environmental stresses [35]. Numerous candidate genes, namely *Lus10005814*, *Lus10005815*, *Lus10005818*, *Lus10013618*, *Lus10034490* and *Lus10039564* (Table 1), were predicted to encode RING/U-box superfamily proteins.

2.9. Network Analysis

To provide further evidence for the involvement of the selected QTNs in trait variation, we constructed a co-expression network using candidate genes. Next, the network of most highly connected genes was obtained by filtering weighed network adjacencies with a threshold set to 0.9 (Figure 4). The annotation of the hub genes in this network highlighted the subnetwork of co-expressed genes, which contained genes involved in the secondary cell wall formation and modification (*Lus10033485*, *Lus10007529*, *Lus10025431*, *Lus10029638* and *Lus10037568*), as well as two genes (*Lus10041651* and *Lus10011858*) (Tables 1 and 2) which were predicted to encode lignin-synthesizing enzymes: cinnamoyl CoA reductase [60] and cytochrome B5 [75]. This subnetwork also included two genes for transporter proteins, *Lus10042571* and *Lus10029063*, which also may be involved in flax secondary cell wall formation. Interestingly, the Chr3:25379963 and Chr4:15210519 QTNs and their candidate genes *Lus10037568* and *Lus10041651* were about 85 kb and 160 kb away from the *uq.C3-1* and the *QPLH-Lu4.3* QTLs, respectively, both associated with plant height (Table S8b) [26,117]. Taking into account the low resolution of QTL mapping, we may speculate that genomic regions detected in GWAS and QTL analyses are in close proximity or even coincide.

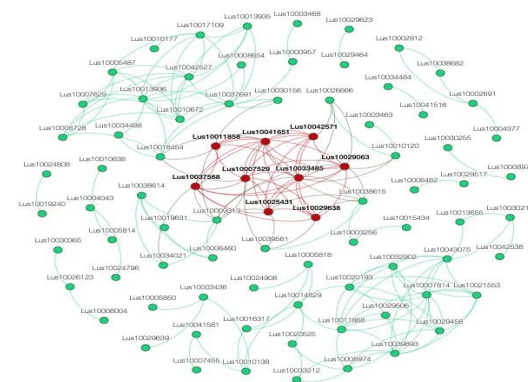


Figure 4. The network of most highly connected candidate genes. A node size is proportional to its adjacency value. Red node color defines gene membership in a subnetwork.

3. Materials and Methods

3.1. Plant Material and Phenotyping

A total of 306 flax specimens containing 182 fiber flax accessions, 120 linseed flax accessions and 4 accessions of unknown morphotype were selected from the collection of the Federal Research Center for Bast Fiber Crops (Table S12). Linseed group was further subdivided into the following subgroups: 99 intermediate accessions, 5 large-seeded accessions and 16 crown accessions. The dataset included landraces, elite cultivars and breeding lines (Table S13). The geographic origin/release of the accessions encompasses 30 countries and all continents.

All plants were grown in the experimental field of Federal Research Center for Bast Fiber Crops in Torzhok, Tver region, Russia (57°02' N, 34°58' E; altitude: 165 m). A detailed description of plant growing and phenotyping was published in [10]. In brief, the accessions were planted using a completely randomized design in the spring of 2019 and two times in the spring of 2020 with a shift of two weeks. For each specimen, 20 seeds were sown in one hole. Ten rows of holes (200 sown seeds and 10 accession plots) alternated with a single row of standard flax varieties (controls), namely, Caesar and Severny cultivars, representing fiber and oil flax, correspondingly. Harvesting was carried out in the yellow ripening stage.

Phenotyping was performed in each planting. For replicate plots of each accession, we first measured 6 morphological traits informative of fiber content and quality: plant height, length of technical part of the stem, technical part weight, the diameter of the stem (at the top, at its middle part and at the butt), the number of internodes and distance between internodes. Next, the technical fiber was extracted manually from each stalk after retting. The isolated fiber was then weighed, and its content was assessed as a percentage of total straw weight, resulting in two additional phenotypic traits.

A challenging issue faced by flax growers and textile workers is the high labor intensity of methods for estimation of fiber content and predicting its quality. Consequently, three indirect indices of the prospective fiber quality assessment were introduced, which for each specimen enable to predict the proportion of fibrous matter in the stem and the feasibility of its isolation in the form of long fibers during mechanical processing. The stem slenderness index (SSI) is a ratio of the length of the technical part of the stem to its diameter. Plants with high SSI values are tall and contain high-quality fine fiber. The stem tapering index (STI) is the difference between the diameters in the lower and upper parts of the stem. The third index, the elementary fiber length (EFL), is calculated as the following:

$$D_f = 0.69D_{int} + 0.096D_{tlength} + 1.04D_{stem} - 0.34 \quad (1)$$

where D_{int} —average internode length of the stem (mm). The internode length is calculated as the ratio of the length of the technical part of the stem to the number of leaves:

$D_{tlength}$ —technical length of the stem (mm);

D_{stem} —the diameter of the stem (the average between diameter values at the top, at its middle part and at the butt, mm).

We calculated all three indices in addition to morphology and fiber traits, bringing the total number of traits measured for each accession to eleven (Table S1).

3.2. DNA Sequencing and Variant Calling

The DNeasy Plant Mini Kit (Qiagen, Germantown, MD, USA) was used to extract DNA from collected leaves. The Illumina protocol generating paired-end reads of 150 bp was used to sequence DNA at the BGI (China). A total of 1143.850625 Gbytes of raw data comprising 7.626 billion reads with an average of $9.3\times$ coverage or 3.7 Gbp per sample were generated. Reads were processed and aligned to NCBI flax reference genome assembly ASM22429v2 with bwa-mem using default parameters [118]. NGSEP [119] version 4.0. was used to call variants identified in the 3,416,829 biallelic SNPs. VCF tools [120] was applied to retain SNPs with minor allele frequency (MAF) > 5% and genotype call-rate > 85%. A total of 72,526 SNPs passed all filters and remained for further analysis.

3.3. Genetic Data Analysis

Principal component analysis (PCA) was performed using the PCA tools R package (Blighe internet). The linkage disequilibrium (LD) decay was evaluated using squared Pearson's correlation coefficient (r^2). The PopLDdecay [121] version 3.4.1 was run to calculate r^2 in a 500 kb window. The LD decay was calculated based on r^2 and the distance for each pair of SNPs using an R script in accordance with Hill–Weir approximation [122]. We applied Mann–Whitney–Wilcoxon test [123] to make group comparisons. Flax genome annotation [117] was kindly provided by Cloutier group (Ottawa Research and Development Centre).

3.4. GWAS

The univariate GWAS analysis was performed with GAPIT3 R package [32]. We selected the following models according to the workflow proposed by the package authors for multiple model testing: GLM, MLM, CMLM, FarmCPU, SUPER and Blink. The phenotype values were quantile normalized to a standard normal distribution. The FDR adjusted p -value threshold for candidate QTNs' selection was set at 0.05.

Because of the high correlation of fiber traits, in addition to univariate GWAS models, a multivariate linear mixed model implemented in the GEMMA package [33] was applied. GEMMA tests marker associations with multiple phenotypes and controls both for population structure and for estimation of genetic correlation among complex phenotypes. A preliminary correlation analysis of 11 phenotypic traits using GEMMA selected five of them for further analysis: number of internodes (NI), technical length (TL), fiber weight (FW), plant height (PH) and fiber content (FC). The total number of single-nucleotide variants after filtering (call_rate 0.15, MAF 0.05) was 24,281 for 5 phenotypic traits. The threshold for determining statistically significant variants was 10^{-6} .

3.5. Candidate Genes

A function of the candidate genes containing the identified QTNs was inferred from the function of their *Arabidopsis* orthologs in the TAIR database, from flax genome annotations kindly provided by Cloutier group (Ottawa Research and Development Centre, Canada) [117], as well as from the function of the homologous genes in other plant species, as described in the literature. We also searched for candidate genes within a window surrounding detected QTNs, where width was defined as proceeding from the LD decay estimated for each chromosome (Table S8).

3.6. Differential Gene Expression in Fiber Cells

The transcriptome libraries for bast fibers isolated at the stage of intrusive elongation (iFIB) and tertiary cell wall deposition (tFIB), cortical parenchyma (cPAR) and xylem part of the stem (sXYL) were obtained from SRA (BioProject PRJNA475325) as raw data in FASTQ format. The alignment to the flax reference genome version 2.0 was performed with STAR aligner [124] version 2.6.0c. The differential expression of genes was estimated with the DESeq2 [125] BioConductor package and the adjusted p -value threshold of 0.01.

3.7. Co-Expression Network Construction

To gain further insight into possible biological functions of the candidate genes, we used the transcriptome libraries described above (Section 3.6) to construct a co-expression network with the BioNERO package [126]. The package implements the Weighted Gene Co-expression Network Analysis (WGCNA) algorithm [127] to infer weighted networks from expression data. To estimate gene distances in the network, correlation values were raised to a power $\beta = 6$ that provided best fit to scale-free topology. The threshold 0.9 was used to filter weak weighed network adjacencies and highlight hub genes together with their connections. The network visualization was conducted in Gephi [128].

4. Conclusions

Genome-wide association studies use whole-genome sequencing information and are able to achieve higher precision in QTL detection in comparison with more conventional methods of QTL mapping. At the same time, dissecting GWAS loci to decipher the underlying biology is a labor-intensive multi-step process. High linkage disequilibrium between many variants often necessitates utilizing statistical fine-mapping approaches, which still have insufficient power to detect associations with alleles of small effect and low frequency and require the application of the FDR-controlling procedures for false discoveries and the use of functional genomic annotations for prioritization of variants before experimental testing. Recently, transcriptomics studies emerged as an approach to prioritize genomic regions for experimental validation of variants. In this study we integrated the results of GWAS and transcriptomics analyses to pinpoint genomic regions controlling fiber-related traits in flax. This approach revealed a number of genomic variants associated with different fiber traits, implying their complex and polygenic control. The variants mark potential candidate genes known to have a role in the synthesis and modification of cell wall components, in the control of cell fate determination, growth and elongation, in vesicular and intracellular trafficking of the cell wall components and solute transport across membranes, and in hormonal regulation and stress responses. Our results suggest tight genetic control of the fiber cell fate determination, as well as a leading role of auxin and cytokinins in promotion of fiber cell wall expansion and growth. The comprehensive set of QTNs identified in this study adds to better dissection of complex fiber traits, while the pinpointed candidate genes will be further used for in-depth investigations of their functional role and in marker-assisted selection.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms232314536/s1>.

Author Contributions: Conceptualization, M.S.; methodology, A.K.; software, M.D. and A.K.; validation, M.B., E.O. and M.S.; investigation, T.R. and S.S.; resources, T.R.; data curation, T.R. and M.B.; writing—original draft preparation, M.S.; writing—review and editing, M.S. and A.K.; funding acquisition, M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Science and Higher Education of the Russian Federation under the strategic academic leadership program “Priority 2030” (Agreement 075-15-2021-1333 dated 30 September 2021).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are openly available at Figshare <https://figshare.com/s/86a68ecfac6872ef239> (accessed on 26 September 2022) (phenotypes) and at European Variation Archive (EVA) Project: PRJEB46073, Analyses: ERZ2775743 (genetic variants).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The following abbreviations were used for phenotyping traits:

Plant height	(PH)
Technical stem length	(TL)
Technical part weight	(TW)
Stem diameter	(SD)
Number of internodes	(NI)
Distance between internodes	(DI)
Fiber content	(FC)
Fiber weight	(FW)
Elementary fiber length	(ELF)
Stem slenderness index	(SSI)
Stem tapering index	(STI)

References

1. Goudenhooff, C.; Bourmaud, A.; Baley, C. Flax (*Linum usitatissimum* L.) fibers for composite reinforcement: Exploring the link between plant growth, cell walls development, and fiber properties. *Front. Plant Sci.* **2019**, *10*, 411. [[CrossRef](#)] [[PubMed](#)]
2. Allaby, R.G.; Peterson, G.W.; Merriwether, D.A.; Fu, Y.-B. Evidence of the domestication history of flax (*Linum usitatissimum* L.) from genetic diversity of the Sad2 locus. *Theor. Appl. Genet.* **2005**, *112*, 58–65. [[CrossRef](#)] [[PubMed](#)]
3. Goudenhooff, C.; Alm eras, T.; Bourmaud, A.; Baley, C. The remarkable slenderness of flax plant and pertinent factors affecting its mechanical stability. *Biosyst. Eng.* **2019**, *178*, 1–8. [[CrossRef](#)]
4. Goudenhooff, C.; Bourmaud, A.; Baley, C. Varietal selection of flax over time: Evolution of plant architecture related to influence on the mechanical properties of fibers. *Ind. Crop. Prod.* **2017**, *97*, 56–64. [[CrossRef](#)]
5. Fogorasi, M.; Barbu, I. The potential of natural fibres for automotive sector—Review. *Iop. Conf. Ser. Mater. Sci. Eng.* **2017**, *252*, 012044. [[CrossRef](#)]
6. Mellerowicz, E.J.; Gorshkova, T.A. Tensional stress generation in gelatinous fibres: A review and possible mechanism based on cell-wall structure and composition. *J. Exp. Bot.* **2012**, *63*, 551–565. [[CrossRef](#)] [[PubMed](#)]
7. Roach, M.J.; Mokshina, N.Y.; Badhan, A.; Snegireva, A.V.; Hobson, N.; Deyholos, M.K.; Gorshkova, T.A. Development of cellulosic secondary walls in flax fibers requires β -galactosidase. *Plant Physiol.* **2011**, *156*, 1351–1363. [[CrossRef](#)] [[PubMed](#)]
8. Gorshkova, T.; Chernova, T.; Mokshina, N.; Gorshkov, V.; Kozlova, L.; Gorshkov, O. Transcriptome analysis of intrusively growing flax fibers isolated by laser microdissection. *Sci. Rep.* **2018**, *8*, 14570. [[CrossRef](#)] [[PubMed](#)]
9. Baley, C.; Goudenhooff, C.; Gibaud, M.; Bourmaud, A. Flax stems: From a specific architecture to an instructive model for bioinspired composite structures. *Bioinspir. Biomim.* **2018**, *13*, 026007. [[CrossRef](#)] [[PubMed](#)]
10. Rozhmina, T.; Bankin, M.; Samsonova, A.; Kanapin, A.; Samsonova, M. A comprehensive dataset of flax (*Linum usitatissimum* L.) Phenotypes. *Data Brief* **2021**, *37*, 107224. [[CrossRef](#)] [[PubMed](#)]
11. Barabaschi, D.; Tondelli, A.; Desiderio, F.; Volante, A.; Vaccino, P.; Val e, G.; Cattivelli, L. Next generation breeding. *Plant Sci.* **2016**, *242*, 3–13. [[CrossRef](#)]
12. Porter, H.F.; O'Reilly, P.F. Multivariate simulation framework reveals performance of multi-trait GWAS methods. *Sci. Rep.* **2017**, *7*, 38837. [[CrossRef](#)]
13. Casale, F.P.; Rakitsch, B.; Lippert, C.; Stegle, O. Efficient set tests for the genetic analysis of correlated traits. *Nat. Methods* **2015**, *12*, 755–758. [[CrossRef](#)] [[PubMed](#)]
14. Stephens, M. A unified framework for association analysis with multiple related phenotypes. *PLoS ONE* **2013**, *8*, e65245. [[CrossRef](#)] [[PubMed](#)]
15. Ritchie, M.D.; Holzinger, E.R.; Li, R.; Pendergrass, S.A.; Kim, D. Methods of integrating data to uncover genotype–phenotype interactions. *Nat. Rev. Genet.* **2015**, *16*, 85–97. [[CrossRef](#)] [[PubMed](#)]
16. Soto-Cerda, B.J.; Duguid, S.; Booker, H.; Rowland, G.; Diederichsen, A.; Cloutier, S. Genomic regions underlying agronomic traits in linseed (*Linum usitatissimum* L.) as revealed by association mapping. *J. Integr. Plant Biol.* **2014**, *56*, 75–87. [[CrossRef](#)] [[PubMed](#)]
17. Chandrawati; Singh, N.; Kumar, R.; Kumar, S.; Singh, P.K.; Yadav, V.K.; Ranade, S.A.; Yadav, H.K. Genetic diversity, population structure and association analysis in linseed (*Linum usitatissimum* L.). *Physiol. Mol. Biol. Plants* **2017**, *23*, 207–219. [[CrossRef](#)] [[PubMed](#)]
18. Soto-Cerda, B.J.; Cloutier, S.; Quian, R.; Gajardo, H.A.; Olivos, M.; You, F.M. Genome-wide association analysis of mucilage and hull content in flax (*Linum usitatissimum* L.) seeds. *Int. J. Mol. Sci.* **2018**, *19*, 2870. [[CrossRef](#)] [[PubMed](#)]
19. You, F.M.; Xiao, J.; Li, P.; Yao, Z.; Jia, G.; He, L.; Kumar, S.; Soto-Cerda, B.; Duguid, S.D.; Booker, H.M.; et al. Genome-wide association study and selection signatures detect genomic regions associated with seed yield and oil quality in flax. *Int. J. Mol. Sci.* **2018**, *19*, 2303. [[CrossRef](#)] [[PubMed](#)]
20. Xie, D.; Dai, Z.; Yang, Z.; Tang, Q.; Deng, C.; Xu, Y.; Wang, J.; Chen, J.; Zhao, D.; Zhang, S.; et al. Combined genome-wide association analysis and transcriptome sequencing to identify candidate genes for flax seed fatty acid metabolism. *Plant Sci.* **2019**, *286*, 98–107. [[CrossRef](#)] [[PubMed](#)]
21. Guo, D.; Jiang, H.; Yan, W.; Yang, L.; Ye, J.; Wang, Y.; Yan, Q.; Chen, J.; Gao, Y.; Duan, L.; et al. Resequencing 200 flax cultivated accessions identifies candidate genes related to seed size and weight and reveals signatures of artificial selection. *Front. Plant Sci.* **2020**, *10*, 1682. [[CrossRef](#)]
22. Soto-Cerda, B.J.; Aravena, G.; Cloutier, S. Genetic dissection of flowering time in flax (*Linum usitatissimum* L.) through single- and multi-locus genome-wide association studies. *Mol. Genet. Genom.* **2021**, *296*, 877–891. [[CrossRef](#)] [[PubMed](#)]
23. Sertse, D.; You, F.M.; Ravichandran, S.; Soto-Cerda, B.J.; Duguid, S.; Cloutier, S. Loci harboring genes with important role in drought and related abiotic stress responses in flax revealed by multiple GWAS models. *Theor. Appl. Genet.* **2021**, *134*, 191–212. [[CrossRef](#)] [[PubMed](#)]
24. Xie, D.; Dai, Z.; Yang, Z.; Sun, J.; Zhao, D.; Yang, X.; Zhang, L.; Tang, Q.; Su, J. Genome-wide association study identifying candidate genes influencing important agronomic traits of flax (*Linum usitatissimum* L.) using SLAF-seq. *Front. Plant Sci.* **2018**, *8*, 2232. [[CrossRef](#)] [[PubMed](#)]
25. Wu, J.; Zhao, Q.; Zhang, L.; Li, S.; Ma, Y.; Pan, L.; Lin, H.; Wu, G.; Yuan, H.; Yu, Y.; et al. QTL mapping of fiber-related traits based on a high-density genetic map in flax (*Linum usitatissimum* L.). *Front. Plant Sci.* **2018**, *9*, 885. [[CrossRef](#)] [[PubMed](#)]
26. Zhang, J.; Long, Y.; Wang, L.; Dang, Z.; Zhang, T.; Song, X.; Dang, Z.; Pei, X. Consensus genetic linkage map construction and QTL mapping for plant height-related traits in linseed flax (*Linum usitatissimum* L.). *BMC Plant Biol.* **2018**, *18*, 160. [[CrossRef](#)]

27. Kumar, S.; You, F.M.; Duguid, S.; Booker, H.; Rowland, G.; Cloutier, S. QTL for fatty acid composition and yield in linseed (*Linum usitatissimum* L.). *Theor. Appl. Genet.* **2015**, *128*, 965–984. [[CrossRef](#)] [[PubMed](#)]
28. Xie, D.; Dai, Z.; Yang, Z.; Tang, Q.; Sun, J.; Yang, X.; Song, X.; Lu, Y.; Zhao, D.; Zhang, L.; et al. Genomic variations and association study of agronomic traits in flax. *BMC Genom.* **2018**, *19*, 512. [[CrossRef](#)] [[PubMed](#)]
29. Gorshkov, O.; Chernova, T.; Mokshina, N.; Gogoleva, N.; Suslov, D.; Tkachenko, A.; Gorshkova, T. Intrusive growth of phloem fibers in flax stem: Integrated analysis of MiRNA and mRNA expression profiles. *Plants* **2019**, *8*, 47. [[CrossRef](#)] [[PubMed](#)]
30. Mokshina, N.; Gorshkov, O.; Galinousky, D.; Gorshkova, T. Genes with bast fiber-specific expression in flax plants—Molecular keys for targeted fiber crop improvement. *Ind. Crop. Prod.* **2020**, *152*, 112549. [[CrossRef](#)]
31. Duk, M.; Kanapin, A.; Rozhmina, T.; Bankin, M.; Surkova, S.; Samsonova, A.; Samsonova, M. The genetic landscape of fiber flax. *Front. Plant Sci.* **2021**, *12*, 764612. [[CrossRef](#)] [[PubMed](#)]
32. Wang, J.; Zhang, Z. GAPIT version 3: Boosting power and accuracy for genomic association and prediction. *bioRxiv* **2020**. bioRxiv: 2020.11.29.403170. [[CrossRef](#)] [[PubMed](#)]
33. Zhou, X.; Stephens, M. Efficient algorithms for multivariate linear mixed models in genome-wide association studies. *Nat. Methods* **2014**, *11*, 407–409. [[CrossRef](#)]
34. Struck, A.W.; Thompson, M.L.; Wong, L.S.; Micklefield, J. S-adenosyl-methionine-dependent methyltransferases: Highly versatile enzymes in biocatalysis, biosynthesis and other biotechnological applications. *Chembiochem.* **2012**, *13*, 2642–2655. [[CrossRef](#)] [[PubMed](#)]
35. Azevedo, C.; Santos-Rosa, M.J.; Shirasu, K. The U-box protein family in plants. *Trends Plant Sci.* **2001**, *6*, 354–358. [[CrossRef](#)]
36. Yang, W.; Chen, S.; Cheng, Y.; Zhang, N.; Ma, Y.; Wang, W.; Tian, H.; Li, Y.; Hussain, S.; Wang, S. Cell wall/vacuolar inhibitor of fructosidase 1 regulates ABA response and salt tolerance in *Arabidopsis*. *Plant Signal. Behav.* **2020**, *15*, 1744293. [[CrossRef](#)] [[PubMed](#)]
37. Wang, L.; Li, X.-R.; Lian, H.; Ni, D.-A.; He, Y.; Chen, X.-Y.; Ruan, Y.-L. Evidence that high activity of vacuolar invertase is required for cotton fiber and *Arabidopsis* root elongation through osmotic dependent and independent pathways, respectively. *Plant Physiol.* **2010**, *154*, 744–756. [[CrossRef](#)] [[PubMed](#)]
38. Mravec, J.; Skůpa, P.; Bailly, A.; Hoyerová, K.; Křeček, P.; Bielach, A.; Petrášek, J.; Zhang, J.; Gaykova, V.; Stierhof, Y.-D.; et al. Subcellular homeostasis of phytohormone auxin is mediated by the ER-localized PIN5 transporter. *Nature* **2009**, *459*, 1136–1140. [[CrossRef](#)] [[PubMed](#)]
39. Majda, M.; Robert, S. The role of auxin in cell wall expansion. *Int. J. Mol. Sci.* **2018**, *19*, 951. [[CrossRef](#)] [[PubMed](#)]
40. Fawal, N.; Li, Q.; Savelli, B.; Brette, M.; Passaia, G.; Fabre, M.; Mathé, C.; Dunand, C. PeroxiBase: A database for large-scale evolutionary analysis of peroxidases. *Nucleic Acids Res.* **2013**, *41*, D441–D444. [[CrossRef](#)] [[PubMed](#)]
41. Francoz, E.; Ranocha, P.; Nguyen-Kim, H.; Jamet, E.; Burlat, V.; Dunand, C. Roles of cell wall peroxidases in plant development. *Phytochemistry* **2015**, *112*, 15–21. [[CrossRef](#)] [[PubMed](#)]
42. Ul Haq, S.; Khan, A.; Ali, M.; Khattak, A.M.; Gai, W.-X.; Zhang, H.-X.; Wei, A.-M.; Gong, Z.-H. Heat shock proteins: Dynamic biomolecules to counter plant biotic and abiotic stresses. *Int. J. Mol. Sci.* **2019**, *20*, 5321. [[CrossRef](#)] [[PubMed](#)]
43. Drew, D.; North, R.A.; Nagarathinam, K.; Tanabe, M. Structures and general transport mechanisms by the major facilitator superfamily (MFS). *Chem. Rev.* **2021**, *121*, 5289–5335. [[CrossRef](#)]
44. Liberato, M.V.; Silveira, R.L.; Prates, É.T.; de Araujo, E.A.; Pellegrini, V.O.A.; Camilo, C.M.; Kadowaki, M.A.; Neto, M.d.O.; Popov, A.; Skaf, M.S.; et al. Molecular characterization of a family 5 glycoside hydrolase suggests an induced-fit enzymatic mechanism. *Sci. Rep.* **2016**, *6*, 23473. [[CrossRef](#)]
45. Korolev, A.V.; Chan, J.; Naldrett, M.J.; Doonan, J.H.; Lloyd, C.W. Identification of a novel family of 70° KDa microtubule-associated proteins in *Arabidopsis* cells. *Plant J.* **2005**, *42*, 547–555. [[CrossRef](#)]
46. Lim, E.-K.; Doucet, C.J.; Li, Y.; Elias, L.; Worrall, D.; Spencer, S.P.; Ross, J.; Bowles, D.J. The activity of *Arabidopsis* glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates. *J. Biol. Chem.* **2002**, *277*, 586–592. [[CrossRef](#)] [[PubMed](#)]
47. Kaundal, A.; Sandhu, D.; Duenas, M.; Ferreira, J.F.S. Expression of the high-affinity K⁺ transporter 1 (PpHKT1) gene from almond rootstock ‘Nemaguard’ improved salt tolerance of transgenic *Arabidopsis*. *PLoS ONE* **2019**, *14*, e0214473. [[CrossRef](#)] [[PubMed](#)]
48. Rennie, E.A.; Hansen, S.F.; Baidoo, E.E.K.; Hadi, M.Z.; Keasling, J.D.; Scheller, H.V. Three members of the *Arabidopsis* glycosyltransferase family 8 are xylan glucuronosyltransferases. *Plant Physiol.* **2012**, *159*, 1408–1417. [[CrossRef](#)]
49. Ochando, I.; González-Reig, S.; Ripoll, J.-J.; Vera, A.; Martínez-Laborda, A. Alteration of the shoot radial pattern in *Arabidopsis thaliana* by a gain-of-function allele of the class III HD-Zip gene INCURVATA. *Int. J. Dev. Biol.* **2004**, *52*, 953–961. [[CrossRef](#)]
50. Shi, J.X.; Malitsky, S.; Oliveira, S.D.; Branigan, C.; Franke, R.B.; Schreiber, L.; Aharoni, A. SHINE transcription factors act redundantly to pattern the archetypal surface of *Arabidopsis* flower organs. *PLoS Genet.* **2011**, *7*, e1001388. [[CrossRef](#)] [[PubMed](#)]
51. Nishigaki, N.; Yoshimi, Y.; Kuki, H.; Kunieda, T.; Hara-Nishimura, I.; Tsumuraya, Y.; Takahashi, D.; Dupree, P.; Kotake, T. Galactoglucomannan structure of *Arabidopsis* seed-coat mucilage in GDP-mannose synthesis impaired mutants. *Physiol. Plant.* **2021**, *173*, 1244–1252. [[CrossRef](#)] [[PubMed](#)]
52. Vercruyse, J.; Baekelandt, A.; Gonzalez, N.; Inzé, D. Molecular networks regulating cell division during *Arabidopsis* leaf growth. *J. Exp. Bot.* **2019**, *71*, 2365–2378. [[CrossRef](#)]
53. Ito, T.; Meyerowitz, E.M. Overexpression of a gene encoding a cytochrome P450, CYP78A9, induces large and seedless fruit in *Arabidopsis*. *Plant Cell* **2000**, *12*, 1541–1550. [[CrossRef](#)] [[PubMed](#)]

54. Fang, W.; Wang, Z.; Cui, R.; Li, J.; Li, Y. Maternal control of seed size by EOD3/CYP78A6 in *Arabidopsis thaliana*. *Plant J.* **2012**, *70*, 929–939. [[CrossRef](#)]
55. Kang, B.-H.; Busse, J.S.; Bednarek, S.Y. Members of the *Arabidopsis* dynamin-like gene family, ADL1, are essential for plant cytokinesis and polarized cell growth. *Plant Cell* **2003**, *15*, 899–913. [[CrossRef](#)] [[PubMed](#)]
56. Mindrebo, J.T.; Nartey, C.M.; Seto, Y.; Burkart, M.D.; Noel, J.P. Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom. *Curr. Opin. Struc. Biol.* **2016**, *41*, 233–246. [[CrossRef](#)]
57. Pires, I.S.; Negrão, S.; Pentony, M.M.; Abreu, I.A.; Oliveira, M.M.; Purugganan, M.D. Different evolutionary histories of two cation/proton exchanger gene families in plants. *BMC Plant Biol.* **2013**, *13*, 97. [[CrossRef](#)] [[PubMed](#)]
58. Kuroda, H.; Yanagawa, Y.; Takahashi, N.; Horii, Y.; Matsui, M. A comprehensive analysis of interaction and localization of *Arabidopsis* SKP1-LIKE (ASK) and F-Box (FBX) proteins. *PLoS ONE* **2012**, *7*, e50009. [[CrossRef](#)]
59. Rocha, J.; Cicéron, F.; de Sanctis, D.; Lelimosin, M.; Chazalet, V.; Lerouxel, O.; Breton, C. Structure of *Arabidopsis thaliana* FUT1 reveals a variant of the GT-B class fold and provides insight into *Xyloglucan fucosylation*. *Plant Cell* **2016**, *28*, 2352–2364. [[CrossRef](#)] [[PubMed](#)]
60. Tamasloukht, B.; Lam, M.S.-J.W.Q.; Martinez, Y.; Tozo, K.; Barbier, O.; Jourda, C.; Jauneau, A.; Borderies, G.; Balzergue, S.; Renou, J.-P.; et al. Characterization of a cinnamoyl-CoA reductase 1 (CCR1) mutant in maize: Effects on lignification, fibre development, and global gene expression. *J. Exp. Bot.* **2011**, *62*, 3837–3848. [[CrossRef](#)]
61. Brembu, T.; Winge, P.; Bones, A.M. The small GTPase AtRAC2/ROP7 is specifically expressed during late stages of xylem differentiation in *Arabidopsis*. *J. Exp. Bot.* **2005**, *56*, 2465–2476. [[CrossRef](#)] [[PubMed](#)]
62. Waghmare, S.; Lileikyte, E.; Karnik, R.; Goodman, J.K.; Blatt, M.R.; Jones, A.M.E. SNAREs SYP121 and SYP122 mediate the secretion of distinct cargo subsets. *Plant Physiol.* **2018**, *178*, 1679–1688. [[CrossRef](#)] [[PubMed](#)]
63. Liu, Q.; Wang, H.; Zhang, Z.; Wu, J.; Feng, Y.; Zhu, Z. Divergence in function and expression of the NOD26-like intrinsic proteins in plants. *BMC Genomics* **2009**, *10*, 313. [[CrossRef](#)] [[PubMed](#)]
64. Irshad, M.; Canut, H.; Borderies, G.; Pont-Lezica, R.; Jamet, E. A new picture of cell wall protein dynamics in elongating cells of *Arabidopsis thaliana*: Confirmed actors and newcomers. *BMC Plant Biol.* **2008**, *8*, 94. [[CrossRef](#)] [[PubMed](#)]
65. Moneo-Sánchez, M.; Izquierdo, L.; Martín, I.; Labrador, E.; Dopico, B. Subcellular location of *Arabidopsis thaliana* subfamily A1 β -galactosidases and developmental regulation of transcript levels of their coding genes. *Plant Physiol. Biochem.* **2016**, *109*, 137–145. [[CrossRef](#)] [[PubMed](#)]
66. He, G.; Liu, P.; Zhao, H.; Sun, J. The HD-ZIP II transcription factors regulate plant architecture through the auxin pathway. *Int. J. Mol. Sci.* **2020**, *21*, 3250. [[CrossRef](#)] [[PubMed](#)]
67. Schreiber, K.J.; Bentham, A.; Williams, S.J.; Kobe, B.; Staskawicz, B.J. Multiple domain associations within the *Arabidopsis* immune receptor RPP1 regulate the activation of programmed cell death. *PLoS Pathog.* **2016**, *12*, e1005769. [[CrossRef](#)]
68. Liu, J.; Zhou, M. The ALMT gene family performs multiple functions in plants. *Agronomy* **2018**, *8*, 20. [[CrossRef](#)]
69. Rauf, M.; Arif, M.; Fisahn, J.; Xue, G.-P.; Balazadeh, S.; Mueller-Roeber, B. NAC transcription factor SPEEDY HYPONASTIC GROWTH regulates flooding-induced leaf movement in *Arabidopsis*. *Plant Cell* **2013**, *25*, 4941–4955. [[CrossRef](#)]
70. Rajaraman, J.; Douchkov, D.; Hensel, G.; Stefanato, F.L.; Gordon, A.; Ereful, N.; Caldararu, O.F.; Petrescu, A.-J.; Kumlehn, J.; Boyd, L.A.; et al. An LRR/malectin receptor-like kinase mediates resistance to non-adapted and adapted powdery mildew fungi in barley and wheat. *Front. Plant Sci.* **2016**, *7*, 1836. [[CrossRef](#)]
71. Li, Z.; Liu, D.; Xia, Y.; Li, Z.; Jing, D.; Du, J.; Niu, N.; Ma, S.; Wang, J.; Song, Y.; et al. Identification of the WUSCHEL-related homeobox (WOX) gene family, and interaction and functional analysis of TaWOX9 and TaWUS in wheat. *Int. J. Mol. Sci.* **2020**, *21*, 1581. [[CrossRef](#)]
72. Xu, J.; Wang, X.; Guo, W. The cytochrome P450 superfamily: Key players in plant development and defense. *J. Integr. Agr.* **2015**, *14*, 1673–1686. [[CrossRef](#)]
73. Eggers, R.; Jammer, A.; Jha, S.; Kerschbaumer, B.; Lahham, M.; Strandback, E.; Toplak, M.; Wallner, S.; Winkler, A.; Macheroux, P. The scope of flavin-dependent reactions and processes in the model plant *Arabidopsis thaliana*. *Phytochemistry* **2021**, *189*, 112822. [[CrossRef](#)]
74. Etchells, J.P.; Provost, C.M.; Mishra, L.; Turner, S.R. WOX4 and WOX14 act downstream of the PXY receptor kinase to regulate plant vascular proliferation independently of any role in vascular organisation. *Development* **2013**, *140*, 2224–2234. [[CrossRef](#)] [[PubMed](#)]
75. Gou, M.; Yang, X.; Zhao, Y.; Ran, X.; Song, Y.; Liu, C.-J. Cytochrome b 5 is an obligate electron shuttle protein for syringyl lignin biosynthesis in *Arabidopsis*. *Plant Cell* **2019**, *31*, 1344–1366. [[CrossRef](#)]
76. Davies, G.; Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure* **1995**, *3*, 853–859. [[CrossRef](#)]
77. Woo, H.-H.; Jeong, B.R.; Hirsch, A.M.; Hawes, M.C. Characterization of *Arabidopsis* AtUGT85A and AtGUS gene families and their expression in rapidly dividing tissues. *Genomics* **2007**, *90*, 143–153. [[CrossRef](#)]
78. Jin, S.-H.; Ma, X.-M.; Kojima, M.; Sakakibara, H.; Wang, Y.-W.; Hou, B.-K. Overexpression of glucosyltransferase UGT85A1 influences trans-zeatin homeostasis and trans-zeatin responses likely through O-glucosylation. *Planta* **2013**, *237*, 991–999. [[CrossRef](#)] [[PubMed](#)]
79. Baucher, M.; Pérez-Morga, D.; Jaziri, M.E. Insight into plant annexin function: From shoot to root signaling. *Plant Signal. Behav.* **2012**, *7*, 524–528. [[CrossRef](#)] [[PubMed](#)]

80. Hoffmann, L.; Besseau, S.; Geoffroy, P.; Ritzenthaler, C.; Meyer, D.; Lapierre, C.; Pollet, B.; Legrand, M. Silencing of hydroxycinnamoyl-coenzyme a shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. *Plant Cell* **2004**, *16*, 1446–1465. [[CrossRef](#)] [[PubMed](#)]
81. Hebda, A.; Liszka, A.; Lewandowska, A.; Lyczakowski, J.J.; Gabryś, H.; Krzeszowiec, W. Upregulation of GLRs expression by light in *Arabidopsis* leaves. *BMC Plant Biol.* **2022**, *22*, 197. [[CrossRef](#)]
82. Merkulova, M.; Hurtado-Lorenzo, A.; Hosokawa, H.; Zhuang, Z.; Brown, D.; Ausiello, D.A.; Marshansky, V. Aldolase directly interacts with ARNO and modulates cell morphology and acidic vesicle distribution. *Am. J. Physiol.-Cell Physiol.* **2011**, *300*, C1442–C1455. [[CrossRef](#)] [[PubMed](#)]
83. Kang, H.; Singh, K.B. Characterization of salicylic acid-responsive, *Arabidopsis* Dof domain proteins: Overexpression of OBP3 leads to growth defects. *Plant J.* **2000**, *21*, 329–339. [[CrossRef](#)]
84. Ranocha, P.; Dima, O.; Nagy, R.; Felten, J.; Corratgé-Faillie, C.; Novák, O.; Morreel, K.; Lacombe, B.; Martinez, Y.; Pfrunder, S.; et al. *Arabidopsis* WAT1 is a vacuolar auxin transport facilitator required for auxin homeostasis. *Nat. Commun.* **2013**, *4*, 2625. [[CrossRef](#)]
85. Ranocha, P.; Denancé, N.; Vanholme, R.; Freydier, A.; Martinez, Y.; Hoffmann, L.; Köhler, L.; Pouzet, C.; Renou, J.; Sundberg, B.; et al. Walls are thin 1 (WAT1), an *Arabidopsis* homolog of medicago truncatula NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers. *Plant J.* **2010**, *63*, 469–483. [[CrossRef](#)] [[PubMed](#)]
86. Denancé, N.; Szurek, B.; Noël, L.D. Emerging functions of nodulin-like proteins in non-nodulating plant species. *Plant Cell Physiol.* **2014**, *55*, 469–474. [[CrossRef](#)] [[PubMed](#)]
87. Geng, P.; Zhang, S.; Liu, J.; Zhao, C.; Wu, J.; Cao, Y.; Fu, C.; Han, X.; He, H.; Zhao, Q. MYB20, MYB42, MYB43, and MYB85 regulate phenylalanine and lignin biosynthesis during secondary cell wall formation. *Plant Physiol.* **2019**, *182*, 1272–1283. [[CrossRef](#)] [[PubMed](#)]
88. Minic, Z.; Jouanin, L. Plant glycoside hydrolases involved in cell wall polysaccharide degradation. *Plant Physiol. Biochem.* **2006**, *44*, 435–449. [[CrossRef](#)]
89. Tedeschi, F.; Rizzo, P.; Rutten, T.; Altschmied, L.; Bäumlein, H. RWP-RK domain-containing transcription factors control cell differentiation during female gametophyte development in *Arabidopsis*. *New Phytol.* **2017**, *213*, 1909–1924. [[CrossRef](#)]
90. De Araújo, A.C.; Fonseca, F.C.D.A.; Cotta, M.G.; Alves, G.S.C.; Miller, R.N.G. Plant NLR receptor proteins and their potential in the development of durable genetic resistance to biotic stresses. *Biotechnol. Res. Innov.* **2019**, *3*, 80–94. [[CrossRef](#)]
91. Schaller, A.; Stintzi, A.; Rivas, S.; Serrano, I.; Chichkova, N.V.; Vartapetian, A.B.; Martínez, D.; Guiamét, J.J.; Sueldo, D.J.; Hoorn, R.A.L.; et al. From structure to function—A family portrait of plant subtilases. *New Phytol.* **2018**, *218*, 901–915. [[CrossRef](#)] [[PubMed](#)]
92. Bischoff, V.; Nita, S.; Neumetzler, L.; Schindelasch, D.; Urbain, A.; Eshed, R.; Persson, S.; Delmer, D.; Scheible, W.-R. TRICHOME BIREFRINGENCE and its homolog AT5G01360 encode plant-specific DUF231 proteins required for cellulose biosynthesis in *Arabidopsis*. *Plant Physiol.* **2010**, *153*, 590–602. [[CrossRef](#)] [[PubMed](#)]
93. Bris, P.L.; Wang, Y.; Barbereau, C.; Antelme, S.; Cézard, L.; Legée, F.; D’Orlando, A.; Dalmais, M.; Bendahmane, A.; Schuetz, M.; et al. Inactivation of LACCASE8 and LACCASE5 genes in brachypodium distachyon leads to severe decrease in lignin content and high increase in saccharification yield without impacting plant integrity. *Biotechnol. Biofuels* **2019**, *12*, 181. [[CrossRef](#)] [[PubMed](#)]
94. Bürstenbinder, K.; Savchenko, T.; Müller, J.; Adamson, A.W.; Stamm, G.; Kwong, R.; Zipp, B.J.; Dinesh, D.C.; Abel, S. *Arabidopsis* Calmodulin-binding protein IQ67-domain 1 localizes to microtubules and interacts with kinesin light chain-related protein-1. *J. Biol. Chem.* **2013**, *288*, 1871–1882. [[CrossRef](#)]
95. Bürstenbinder, K.; Möller, B.; Plötner, R.; Stamm, G.; Hause, G.; Mitra, D.; Abel, S. The IQD family of calmodulin-binding proteins links calcium signaling to microtubules, membrane subdomains, and the nucleus. *Plant Physiol.* **2017**, *173*, 1692–1708. [[CrossRef](#)] [[PubMed](#)]
96. Zhong, R.; Richardson, E.A.; Ye, Z.-H. The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in *Arabidopsis*. *Plant Cell* **2007**, *19*, 2776–2792. [[CrossRef](#)]
97. Cordoba, E.; Aceves-Zamudio, D.L.; Hernández-Bernal, A.F.; Ramos-Vega, M.; León, P. Sugar regulation of SUGAR TRANSPORTER PROTEIN 1 (STP1) expression in *Arabidopsis thaliana*. *J. Exp. Bot.* **2015**, *66*, 147–159. [[CrossRef](#)]
98. Ma, R.; Yuan, H.; An, J.; Hao, X.; Li, H. A *Gossypium hirsutum* GDSL lipase/hydrolase gene (GhGLIP) appears to be involved in promoting seed growth in *Arabidopsis*. *PLoS ONE* **2018**, *13*, e0195556. [[CrossRef](#)]
99. Lai, C.-P.; Huang, L.-M.; Chen, L.-F.O.; Chan, M.-T.; Shaw, J.-F. Genome-wide analysis of GDSL-type esterases/lipases in *Arabidopsis*. *Plant Mol. Biol.* **2017**, *95*, 181–197. [[CrossRef](#)] [[PubMed](#)]
100. Yadav, V.K.; Yadav, V.K.; Pant, P.; Singh, S.P.; Maurya, R.; Sable, A.; Sawant, S.V. GhMYB1 regulates SCW stage-specific expression of the GhGDSL promoter in the fibres of *Gossypium hirsutum* L. *Plant Biotechnol. J.* **2017**, *15*, 1163–1174. [[CrossRef](#)] [[PubMed](#)]
101. Li, S.-B.; Xie, Z.-Z.; Hu, C.-G.; Zhang, J.-Z. A review of auxin response factors (ARFs) in plants. *Front. Plant Sci.* **2016**, *7*, 47. [[CrossRef](#)]
102. Yang, Y.S.; Strittmatter, S.M. The reticulons: A family of proteins with diverse functions. *Genome Biol.* **2007**, *8*, 234. [[CrossRef](#)]
103. Zhang, Y.; Mitsuda, N.; Yoshizumi, T.; Horii, Y.; Oshima, Y.; Ohme-Takagi, M.; Matsui, M.; Kakimoto, T. Two types of BHLH transcription factor determine the competence of the pericycle for lateral root initiation. *Nat. Plants* **2021**, *7*, 633–643. [[CrossRef](#)] [[PubMed](#)]

104. Gu, J.; Sun, J.; Liu, N.; Sun, X.; Liu, C.; Wu, L.; Liu, G.; Zeng, F.; Hou, C.; Han, S.; et al. A novel cysteine-rich receptor-like kinase gene, TaCRK2, contributes to leaf rust resistance in wheat. *Mol. Plant Pathol.* **2020**, *21*, 732–746. [[CrossRef](#)]
105. Parker, J.E.; Coleman, M.J.; Szabò, V.; Frost, L.N.; Schmidt, R.; van der Biezen, E.A.; Moores, T.; Dean, C.; Daniels, M.J.; Jones, J.D. The *Arabidopsis* downy mildew resistance gene RPP5 shares similarity to the toll and interleukin-1 receptors with N and L6. *Plant Cell* **1997**, *9*, 879–894. [[CrossRef](#)]
106. Piovesana, M.; Wood, A.K.M.; Smith, D.P.; Deery, M.J.; Bayliss, R.; Carrera, E.; Napier, J.A.; Kurup, S.; Matthes, M.C. A point mutation in the kinase domain of CRK10 leads to xylem vessel collapse and activates defence responses. *bioRxiv* **2021**. bioRxiv: 2021.08.16.456532. [[CrossRef](#)]
107. Cosgrove, D.J. Plant cell growth and elongation. In *eLS*; John Wiley & Sons, Ltd.: New York, NY, USA, 2014. [[CrossRef](#)]
108. Wu, S.-W.; Kumar, R.; Iswanto, A.B.B.; Kim, J.-Y. Callose balancing at plasmodesmata. *J. Exp. Bot.* **2018**, *69*, 5325–5339. [[CrossRef](#)]
109. Ageeva, M.V.; Petrovská, B.; Kieft, H.; Sal'nikov, V.V.; Snegireva, A.V.; van Dam, J.E.G.; van Veenendaal, W.L.H.; Emons, A.M.C.; Gorshkova, T.A.; van Lammeren, A.A.M. Intrusive growth of flax phloem fibers is of intercalary type. *Planta* **2005**, *222*, 565–574. [[CrossRef](#)]
110. Cassab, G.I. Plant cell wall proteins. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1998**, *49*, 281–309. [[CrossRef](#)] [[PubMed](#)]
111. Hellmann, E.; Ko, D.; Ruonala, R.; Helariutta, Y. Plant vascular tissues—Connecting tissue comes in all shapes. *Plants* **2018**, *7*, 109. [[CrossRef](#)]
112. Du, Q.; Wang, H. The role of HD-ZIP III transcription factors and MiR165/166 in vascular development and secondary cell wall formation. *Plant Signal. Behav.* **2015**, *10*, e1078955. [[CrossRef](#)] [[PubMed](#)]
113. Turchi, L.; Carabelli, M.; Ruzza, V.; Possenti, M.; Sassi, M.; Peñalosa, A.; Sessa, G.; Salvi, S.; Forte, V.; Morelli, G.; et al. *Arabidopsis* HD-Zip II transcription factors control apical embryo development and meristem function. *Development* **2013**, *140*, 2118–2129. [[CrossRef](#)]
114. Sanderfoot, A.A.; Kovaleva, V.; Bassham, D.C.; Raikhel, N.V. Interactions between syntaxins identify at least five SNARE complexes within the golgi/prevacuolar system of the *Arabidopsis* cell. *Mol. Biol. Cell* **2001**, *12*, 3733–3743. [[CrossRef](#)] [[PubMed](#)]
115. Wang, L.; Ruan, Y.-L. Unraveling mechanisms of cell expansion linking solute transport, metabolism, plasmodesmal gating and cell wall dynamics. *Plant Signal. Behav.* **2010**, *5*, 1561–1564. [[CrossRef](#)]
116. Malinovsky, F.G.; Fangel, J.U.; Willats, W.G.T. The role of the cell wall in plant immunity. *Front. Plant Sci.* **2014**, *5*, 178. [[CrossRef](#)] [[PubMed](#)]
117. You, F.M.; Cloutier, S. Mapping quantitative trait loci onto chromosome-scale pseudomolecules in flax. *Methods Protoc.* **2020**, *3*, 28. [[CrossRef](#)]
118. Li, H.; Durbin, R. Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)] [[PubMed](#)]
119. Tello, D.; Gil, J.; Loaiza, C.D.; Riascos, J.J.; Cardozo, N.; Duitama, J. NGSEP3: Accurate variant calling across species and sequencing protocols. *Bioinformatics* **2019**, *35*, 4716–4723. [[CrossRef](#)]
120. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [[CrossRef](#)]
121. Zhang, C.; Dong, S.-S.; Xu, J.-Y.; He, W.-M.; Yang, T.-L. PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **2018**, *35*, 1786–1788. [[CrossRef](#)] [[PubMed](#)]
122. Hill, W.G.; Weir, B.S. Variances and covariances of squared linkage disequilibria in finite populations. *Theor. Popul. Biol.* **1988**, *33*, 54–78. [[CrossRef](#)]
123. Mann, H.B.; Whitney, D.R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **1947**, *18*, 50–60. [[CrossRef](#)]
124. Dobin, A.; Gingeras, T.R. Mapping RNA-seq reads with STAR. *Curr. Protoc. Bioinform.* **2015**, *51*, 11–14. [[CrossRef](#)]
125. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)]
126. Almeida-Silva, F.; Venancio, T.M. BioNERO: An all-in-one R/bioconductor package for comprehensive and easy biological network reconstruction. *bioRxiv* **2021**. bioRxiv: 2021.04.10.439287. [[CrossRef](#)]
127. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [[CrossRef](#)]
128. Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An open source software for exploring and manipulating networks. *Proc. Int. AAAI Conf. Web Soc. Media* **2009**, *3*, 361–362. [[CrossRef](#)]