



The Toolbox for Fiber Flax Breeding: A Pipeline From Gene Expression to Fiber Quality

Dmitry Galinousky^{1,2}, Natalia Mokshina¹, Tsimafei Padvitski³, Marina Ageeva⁴, Victor Bogdan⁵, Alexander Kilchevsky² and Tatyana Gorshkova^{6*}

¹ Laboratory of Plant Glycobiology, Kazan Institute of Biochemistry and Biophysics, FRC Kazan Scientific Center of RAS, Kazan, Russia, ² Laboratory of Ecological Genetics and Biotechnology, Institute of Genetics and Cytology, The National Academy of Sciences of Belarus, Minsk, Belarus, ³ Cellular Network and Systems Biology Group, University of Cologne, CECAD, Cologne, Germany, ⁴ Laboratory of Microscopy, Kazan Institute of Biochemistry and Biophysics, FRC Kazan Scientific Center of RAS, Kazan, Russia, ⁵ Laboratory of Fiber Flax Breeding, Institute of Flax, Ustie, Belarus, ⁶ Laboratory of Plant Cell Growth Mechanisms, Kazan Institute of Biochemistry and Biophysics, FRC Kazan Scientific Center of RAS, Kazan, Russia

OPEN ACCESS

Edited by:

Gea Guerriero,
Luxembourg Institute of Science
and Technology, Luxembourg

Reviewed by:

Michael Deyholos,
University of British Columbia,
Canada
Marc Behr,
Université libre de Bruxelles, Belgium

*Correspondence:

Tatyana Gorshkova
gorshkova@kibb.knc.ru

Specialty section:

This article was submitted to
Plant Genomics,
a section of the journal
Frontiers in Genetics

Received: 31 July 2020

Accepted: 22 October 2020

Published: 12 November 2020

Citation:

Galinousky D, Mokshina N,
Padvitski T, Ageeva M, Bogdan V,
Kilchevsky A and Gorshkova T (2020)
The Toolbox for Fiber Flax Breeding:
A Pipeline From Gene Expression
to Fiber Quality.
Front. Genet. 11:589881.
doi: 10.3389/fgene.2020.589881

The goal of any plant breeding program is to improve quality of a target crop. Crop quality is a comprehensive feature largely determined by biological background. To improve the quality parameters of crops grown for the production of fiber, a functional approach was used to search for genes suitable for the effective manipulation of technical fiber quality. A key step was to identify genes with tissue and stage-specific pattern of expression in the developing fibers. In the current study, we investigated the relationship between gene expression evaluated in bast fibers of developing flax plants and the quality parameters of technical fibers measured after plant harvesting. Based on previously published transcriptomic data, two sets of genes that are upregulated in fibers during intrusive growth and tertiary cell wall deposition were selected. The expression level of the selected genes and fiber quality parameters were measured in fiber flax, linseed (oil flax) cultivars, and wild species that differ in type of yield and fiber quality parameters. Based on gene expression data, linear regression models for technical stem length, fiber tensile strength, and fiber flexibility were constructed, resulting in the identification of genes that have high potential for manipulating fiber quality. Chromosomal localization and single nucleotide polymorphism distribution in the selected genes were characterized for the efficacy of their use in conventional breeding and genome editing programs. Transcriptome-based selection is a highly targeted functional approach that could be used during the development of new cultivars of various crops.

Keywords: transcriptome-based selection, flax (*Linum L.*), gene expression, fiber quality, intrusive growth, tertiary cell wall, linear regression models

INTRODUCTION

For thousands of years mankind has made effective use of fiber crops for various purposes. Currently, fiber crops produce the raw materials for the textile industry, and also for eco-friendly building materials, cosmetics, car manufacturing, and medicine, etc. Interest in natural fibers is also growing due to environmental issues and new legislation that is driving the development of a market for biodegradable and recyclable materials (Smole et al., 2013).

There are numerous plant sources of natural fibers, which are often classified according to their plant tissue of origin. Accordingly, there are three main categories: (1) bast fibers of the stem (flax, hemp, jute, ramie, kenaf, etc.), (2) leaf fibers (sisal, banana, manila hemp, pineapple, etc.), and (3) seed fibers (coir, oil palm, etc.) (Jawaid and Khalil, 2011). More than eight million households worldwide are involved in the production of plant-based fibers from the above listed groups (excluding cotton). With a global production of around six million metric tons, the market share of these plant-based fibers was around 5.7% of total global fiber production volume in 2018 (based on FAOstat and Total Global Production Volumes compiled by Textile Exchange¹). The global flax fiber and tow production in 2018 is estimated at around 780 metric tons, with 93% of flax production coming from Europe (data for the last decade²). Flax is grown for its cellulose-rich fibers and seeds, which can be ground into a meal or turned into linseed oil. The long bast fibers are traditionally used in the textile industry for the production of linen or mixed fiber textiles, and, together with shorter xylem fibers they are also used in the automobile and construction industries (Baley et al., 2006; Kymäläinen and Sjöberga, 2008).

Flax bast fibers are primary fibers that originate from the procambium (Esau, 1943) and are similar to a number of primary bast fibers of other plant species (hemp, ramie, and nettle) that form a thick cellulose-rich cell wall that we have named the tertiary cell wall (TCW) (Gorshkova et al., 2018a). Before cell wall thickening is initiated, the fibers undergo a specific type of cell elongation, called intrusive growth, that occurs in the depths of other stem tissues (Ageeva et al., 2005; Snegireva et al., 2010). This stage of development has a notable impact on the yield and quality of fibers, and it has been largely underestimated and poorly characterized from a molecular-genetic point of view (Snegireva et al., 2015; Mokshina et al., 2018). Structure and composition of the bast fiber TCW is similar to the G-fiber cell wall in tension wood (Gorshkova et al., 2015, 2018a), and is quite distinct from the other cell wall types. The high cellulose content (up to 90%), the presence of rhamnogalacturonan I, and absence of xylan and lignin are specific distinguishing features of the TCW. The polymer arrangement also differentiates TCW from primary and secondary cell walls; cellulose microfibrils in the TCW lie close to the axial orientation and are poorly interlaid by matrix polymers (Gorshkova et al., 2018a). Both the chemical composition and spatial organization of polymers in the cell wall supramolecular structure contribute to the natural strength, flexibility, smoothness, and linearity of flax fibers.

Fiber crop producers aim to achieve excellent fiber quality; this is a complex feature that is characterized by a number of physical parameters measured after plant harvest and partial processing. These parameters include tensile strength, density, flexibility, and color grade (Akin, 2005) and they are widely used for the evaluation of technical fibers (here and after, the term “technical fiber” is used for fiber bundles obtained from harvested plants, while the term “fiber” is used in its biological sense, meaning individual plant cell with specific characteristics).

Quality depends both on the peculiarities of fiber development *in planta* and the way in which the stems are processed after crop harvest. The importance of the biological basis of fiber quality is reflected in pronounced differences in technical fiber parameters between various crop cultivars. There are numerous studies describing the anatomical, morphological and biochemical differences between various cultivars that are related to fiber quality (Elhaak et al., 1999; Couture et al., 2002, 2004; Rennebaum et al., 2002). However, to manipulate these quality parameters it is necessary to understand their, as yet, elusive genetic basis. The comprehensive approaches of genomics and transcriptomics may help to fill this gap.

Flax is a bast fiber crop on which the most advanced molecular studies of fiber development have been conducted. The whole genome sequence of cultivated flax was assembled in 2012 (Wang et al., 2012), and was used for SNP (single nucleotide polymorphism) discovery (Kumar et al., 2012), transcriptomic profiling of different tissues (Dmitriev et al., 2016, 2017; Zhang and Deyholos, 2016) and construction of a high resolution consensus map of flax that integrated genetic and physical data (Cloutier et al., 2012). These approaches considerably advanced the earlier important studies to large-scale molecular characterization of the flax genome and transcriptome (Day et al., 2005; Roach and Deyholos, 2007; Fenart et al., 2010). Further RNA-Seq experiments revealed those genes that are highly upregulated during intrusive growth (Gorshkova et al., 2018b) and TCW deposition (Gorshkov et al., 2017), and described genes with fiber-specific expression that are barely transcribed in other tissues (Mokshina et al., 2020). These studies opened the possibility of relating the structure and expression of genes important for fiber development to fiber quality.

In the current study, we concentrated on the establishment of a pipeline that could be used to select a few genes that have high potential for manipulating fiber quality. Our approach included five key steps: (1) selection of candidate genes using previously published RNA-Seq data (Gorshkov et al., 2017, 2019; Mokshina et al., 2020); (2) evaluation of their expression in various developing flax genotypes; (3) determination of technical fiber quality parameters in these flax genotypes after plant harvest; (4) identification of the optimal sets of genes whose expression is closely related to fiber quality features by regularized regression models; and (5) analysis of SNP abundance across the coding regions of genes important for bast fiber development using RNA-Seq data obtained for several different flax genotypes. We suggest that the obtained results may help to advance conventional breeding and genome editing programs for flax and other fiber crops.

MATERIALS AND METHODS

Plant Material and Experimental Design

Ten different flax genotypes were used as the plant objects: six cultivars of fiber flax (Aramis, Eden, Grant, Laska, Drakkar, and Mogilevsky), two cultivars of linseed flax (Lirina and Orpheus), and two wild flax species [*L. usitatissimum* ssp. *bienne* (Mill.)

¹<https://textileexchange.org/>

²<http://www.fao.org/home/en/>

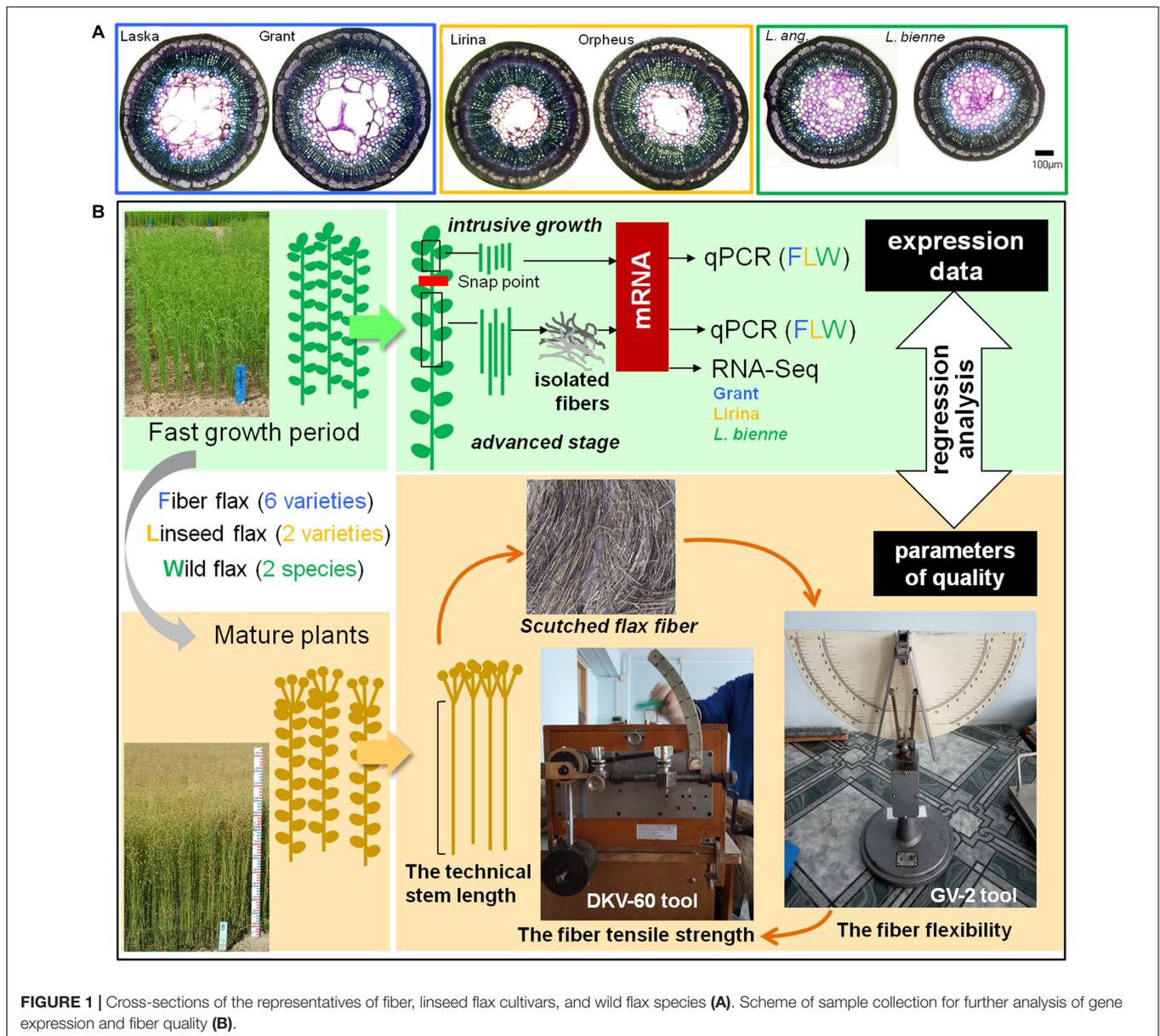


FIGURE 1 | Cross-sections of the representatives of fiber, linseed flax cultivars, and wild flax species (A). Scheme of sample collection for further analysis of gene expression and fiber quality (B).

Stank., and *L. angustifolium* Huds.]. Cultivars of fiber flax and linseed flax belong to *L. usitatissimum*. Altogether, the analyzed genotypes were designated as the FLW group [from fiber (F), linseed (L), and wild (W) flax]. Flax plants were grown on an experimental field of the Institute of Flax of the National Academy of Sciences of Belarus (Ustie, Vitebsk region, Belarus; 54°26′02″ S, 30°21′05″ W in 2018 and 54°26′03″ S, 30°20′50″ W in 2019). The experimental plots were organized according to the type of plant collection nursery, based on guidelines for studying the collection of flax (*Linum usitatissimum* L.) (No author listed, 1988). Plant material for each genotype was collected in 2018 and 2019.

Samples for gene expression analysis were collected during the fast growth period of plant development (around 45 days after sowing) when both intrusive growth and TCW deposition

are well pronounced. The stem segments located above the snap point, a marker of fiber developmental transition from elongation to cell wall thickening (Gorshkova et al., 2003), were used to evaluate the expression of genes whose products were involved in intrusive growth (Figure 1). For this, a 1 cm portion was removed from the top of the stem and the underlying 2 cm stem portion was collected. All leaves were removed from the collected stem fragments.

Gene expression at the advanced stage of fiber specialization was determined in fibers isolated from the stem part below the snap point. To isolate fibers at the stage of TCW deposition, a 10 cm-long stem portion, starting from 1 cm below the snap point was cut downward the stem (Figure 1). The fiber-enriched part was peeled off and the peel was washed in 80% ethanol to isolate bast fibers (Mokshina et al., 2014).

The collected stem parts used to characterize gene expression at fiber intrusive growth and isolated fibers to study TCW deposition were frozen in liquid nitrogen and used for RNA extraction. Three biological replicates at each of the analyzed developmental stages were used for RNA extraction. Each biological replicate included samples from five individual plants. Samples for the determination of technical fiber quality were obtained from mature plants (**Figure 1**).

RNA Extraction and RT-qPCR

RNA was extracted using “ExtractRNA” (Evrogen, Moscow, Russia) and afterward purified using an RNeasy Mini Kit (Qiagen, Hilden, Germany). Obtained RNA samples were treated using a Turbo DNase Kit (Ambion, Austin, TX, United States). RNA concentration and integrity were tested with agarose gel electrophoresis and a NanodropND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, United States). Reverse transcription was performed using RevertAid H Minus Reverse Transcriptase (Thermo Fisher Scientific, Waltham, MA, United States) and Oligo (dT) 12-18 Primer (Invitrogen, Carlsbad, CA, United States) according to the manufacturer’s recommendations. Every cDNA synthesis reaction was accompanied by NRT (no reverse transcriptase control). The obtained cDNA and NRT samples were diluted 10 times and used as templates for qPCR. qPCR was performed using a 2.5× Reaction Mix for qPCR with EVA Green dye (Syntol, Moscow, Russia) and accompanied by NRT and NTC (no template control). Sequences for gene-specific primers for the analyzed genes are presented in **Supplementary Table S1**. *LusGAPDH*, *LusETIF1*, and *LusETIF5A* genes were used as reference genes for qPCR data normalization (Huis et al., 2010). The stability of reference gene expression was tested by the geNorm algorithm implemented in CFX Maestro Software (Bio-Rad, Hercules, CA, United States) (Vandesompele et al., 2002). We computed the ΔCq value for evaluation of gene expression and gene clustering. The ΔCq values were calculated as the difference between the geometric mean of three reference gene Cq values and the Cq values of the genes of interest. The off-scale Cq values were assigned the highest Cq value measured for this gene in other samples plus 2 (Bergkvist et al., 2010). Two technical repetitions were performed for each of three biological replicates. All obtained ΔCq values can be found in **Supplementary Table S2**.

Flax Samples Used for Comparison of Transcriptome Profiles to Select Candidate Genes

Bioinformatic analysis of previously published RNA-Seq data (Mokshina et al., 2020) was performed after normalization applied to all samples simultaneously to ensure that the expression values were comparable across all samples. The procedure of normalization was described previously (Mokshina et al., 2020). Briefly, Illumina raw reads (as FASTQ files) of each sample were filtered using the BBDuk utility of BBTtools v 37.02³.

³<https://sourceforge.net/projects/bbmap>

The clean reads were mapped onto the flax genome sequence scaffolds using HISAT2 v2.1.0.83 (Kim et al., 2015). The aligned reads were aggregated with StringTie v2.0 (Pertea et al., 2016) using the annotation of flax genes as *Lusitatisimum_200_v1.0.gff3*⁴ (Wang et al., 2012) to provide total gene read (TGR) counts for each gene. The R package DESeq2 v.1.14.1 (Love et al., 2014) was used to perform the differential expression analysis of mRNA from all samples using TGR counts. The DESeq estimateSizeFactors and estimateDispersions functions (with the default options) were used to obtain normalization factors for each sample and to normalize the TGR counts. Most of the transcriptomic data were obtained for the fiber flax cultivar Mogilevsky, with the exception of SAM (shoot apical meristem), LEAF, and ROOT samples (**Table 1**).

For the selection of genes upregulated at fiber intrusive growth, genes that had a four-fold increased level of expression in intrusively growing fibers (iFIB) compared to other tissues, and a total gene reads (TGR) ≥ 50 in iFIB were considered. Out of 28 genes that fitted these criteria, 17 genes were selected (close paralogs were excluded).

For the selection of genes upregulated at TCW deposition, genes that had four-fold increased levels of expression in fibers forming TCW (tFIBa) compared to other tissues, and a TGR ≥ 250 in tFIBa were considered. Out of 38 genes

⁴<https://genome.jgi.doe.gov/portal/pages/dynamicOrganismDownload.jsf?organism=Lusitatisimum>

TABLE 1 | The sources of used transcriptomic data.

Sample name	Description	References	Project accession
SAM	Shoot apical meristem, 0.5 mm	Zhang and Deyholos (2016)	PRJNA229810
cPAR	Cortical parenchyma, 0.3–2.5 cm from the stem apex, the same part of stem as iFIB	Gorshkov et al. (2019)	PRJNA475325
iFIB	Intrusively growing fibers with PCW, 0.3–2.5 cm from the apical part (microdissected), average for iFIBa and iFIBb samples	Gorshkova et al. (2018b, 2019)	PRJNA475325
tFIBa	Fibers depositing tertiary cell wall, 5 cm long stem portion starting 1 cm below snap point, earlier stage	Gorshkov et al. (2019)	PRJNA475325
tFIBb	Fibers depositing tertiary cell wall, 5 cm long stem portion, at the 3 cm above cotyledon leaves, later stage	Mokshina et al. (2020)	PRJNA631357
sXYLa	Xylem part enriched in secondary cell walls (the same segment of the stem as for tFIBa)	Gorshkov et al. (2019)	PRJNA475325
sXYLb	Xylem part enriched in secondary cell walls, collected from the same stem segment as tFIBb	Mokshina et al. (2020)	PRJNA631357
LEAF	Upper leaves from individual plants under normal conditions	Dmitriev et al. (2016)	PRJNA340243
ROOT	Root tips (5 mm) of control plants	Dmitriev et al. (2017)	PRJNA412801

that fitted these criteria, 27 were selected. Additionally, five genes related to cellulose biosynthesis (*LusCESA1*, *LusCESA3-B*, *LusCESA4*, and *LusCESA8-A*) and chitinase-like gene (*LusCTL2*) were included into the list.

SNP Search in Coding Regions of the Selected Genes After RNA-Seq Analysis of Several Genotypes

RNA-Seq analysis was performed for fibers at the TCW-deposition stage of development isolated from three flax genotypes: fiber flax (Grant), linseed flax (Lirina), and a wild species (*L. bienne*) harvested in 2019. Total RNA was extracted and purified from the same samples as used for qPCR. The RNA integrity test using an Agilent 2100 bioanalyzer and pair-end RNA sequencing using an Illumina HiSeq 6000 were performed at Novogene in Europe⁵. Raw data were analyzed using Illumina CASAVA 1.8, and three types of sequence reads, namely, reads with adapter, no exact base information, or low-quality reads, were filtered and removed. The genome of a medium-late-maturing oilseed flax (*Linum usitatissimum* L.) CDC Bethune (Rowland et al., 2002) was used as reference genome⁴. The processed data with sequence reads and alignments were generated based on HISAT2 software. Mapping results were provided in BAM format by Novogene. BAM-files were used for SNP calling. Variant calling (SNP and InDel searching) was performed for scaffolds from the reference genome that contained the selected 32 genes, with all the BAM-files simultaneously. We used VarScan v2.3.9 with default settings except for “Minimum read depth at a position to make a call,” which was increased to 20. Data for the analysis were pre-generated by the Samtools v1.5 mpileup command⁶. The obtained variants were annotated using RegTools to differentiate variants located in genes’ exons from others. We included in the final table only those SNPs that were identified in all three biological replicates for each sample. All identified SNPs (at least in one biological replicate) can be found in **Supplementary Table S3**.

We anchored the selected genes into the physical map of flax chromosomes in order to assess their potential linkage with mapped markers (**Supplementary Table S4**). We used chromosome-scale pseudomolecules (You et al., 2018), genetic maps (Cloutier et al., 2012), as well as gene annotation of the phytozome.gov portal⁷.

RNA-Seq data for 32 genes (TCW-upregulated) were used for validation of qPCR data (**Supplementary Figure S1**).

Technical Fiber Quality Evaluation

Flax straw was subjected to dew-retting and processed on a scutching-and-breaking machine SMT-200M according to the quality standards in Belarus (STB1195, 2008; Dyagilev et al., 2016). We evaluated the technical length of the plant stem, tensile strength, and flexibility of scutched fiber. At least three

independent measurements were performed for each quality feature in each biological replicate.

The technical stem length of flax is defined as the length of the stem without side branches (Diederichsen and Fu, 2006). It was measured in cm as a length from the root collar to the point where the stem splits into floral side branches.

The fiber flexibility of the scutched flax fiber was measured using the GV-2 tool (**Figure 1**). We took a 27-cm-long fiber bundle weighing 420 mg and placed it on the horizontal surface of the GV-2 tool. Then supporters on the horizontal surface of the tool were released and fiber sag (in mm) under natural gravity was determined. The value was measured for both sides of the sample and the mean was calculated.

The tensile strength of the scutched flax fiber was evaluated using a DKV-60 tool (modification of the stelometer that was adapted for flax fiber; **Figure 1**). The same sample that was used for flexibility measurement was used here. The fiber sample was fixed between clamps on the device; the distance between the clamps was 100 mm. The force applied to the clamps was then increasing until the fiber broke. The force (in Newtons) was recorded at the time of fiber rupture. The cross-sectional area of the analyzed technical fiber sample was about 1 mm². Therefore, the force of tensile strength measured in Newtons was numerically equal to the tensile strength in MPa.

Statistical Analyses

The obtained data were analyzed in R (R Core Team, 2013); *ggplot2* and *ggpubr* packages were used for data visualization, *stargazer* package was used to report regression output.

The arithmetic means were calculated for independent values of ΔCq and fiber quality features in biological replicates, and then used for computing the gene-phenotype correlation. The data obtained for technical fiber quality and relative expression were tested for normal distribution using the Shapiro–Wilk test. For data that were normally distributed, we used Pearson’s correlation coefficient (PC) and ANOVA test, and the Spearman’s correlation coefficient (SC) and the Kruskal–Wallis test, if not. When necessary, the results of statistical tests were adjusted for the false discovery rate to account for multiple testing.

Co-expression heatmaps were generated with the heatmaply v1.1.0 R package (Galili et al., 2018) using ΔCq of gene expression and Spearman’s correlation coefficient, respectively. Genes and samples (flax cultivars) were clustered using Euclidean distances and the complete-linkage hierarchical clustering method. Gene co-expression heatmaps were generated with corrplot v0.84 R package (Wei and Simko, 2017) using Pearson correlation of ΔCq expression values.

A linear regression was used to model each fiber quality parameter (dependent variable) with ΔCq of the studied gene(s) [independent variable(s)]. Simple and multiple linear regression models were constructed, and the coefficient of determination (R^2), the adjusted R^2 , and statistical significance of the beta coefficients were used to assess the quality of obtained models. To select a few predictive features for each of the mechanical parameters we used a form of regularized regression called LASSO (Tibshirani, 1996) with stability selection (Meinshausen and Bühlmann, 2010). This machine learning approach attempts

⁵<https://en.novogene.com/>

⁶<http://dkoboldt.github.io/varscan/using-varscan.html>

⁷<https://phytozome.jgi.doe.gov/pz/portal.html>

to optimize the variance-bias trade-off in a high dimensional setting, i.e., when the number of independent variables is greater than the number of observations. Alternatively, the stepwise forward selection regularization (Efroymson, 1960) was used to select larger models with better predictive performance for the price of higher bias.

RESULTS

Selection of Genes to Relate Their Expression to Quality Parameters

Genes whose expression had the potential to influence fiber quality were selected on the basis of data from previously published RNA-Seq experiments. We used transcriptomes of various flax tissues (**Table 1**) to identify the genes important for two major stages of bast fiber development: intrusive growth and TCW deposition.

Genes whose expression was important during fiber intrusive growth were selected among the specifically and significantly upregulated genes (>4-fold) at this stage of fiber development (sample iFIB), as compared to all other analyzed samples from other plant tissues and other stages of fiber development (**Table 2A**). The 17 selected genes (designated as Intrusive-upregulated) encoded products with different functional specialization: transcription factors, enzymes for polysaccharide modification, hormonal response proteins, and one protein with unknown function. For example, genes for putative transcription factor *LusBZIP5* (orthologous to *AT3G49760*) and transcriptional activator *LusASML2* [orthologous to *AT3G12890* encoding activator of spomin::LUC2 that regulates the expression of at least a subset of sugar-inducible genes (Masaki et al., 2005)], and auxin-responsive GH3 family protein *Lus10003598* (orthologous to *AT5G54510* and *DFL1*) were among the selected genes. Also, a number of O-Glycosyl hydrolases belonging to family 17 (four genes) that can be involved in callose metabolism were selected, because of their specific activation during intrusive growth (Gorshkova et al., 2018b).

Among 32 genes encoding the products that could play a significant role during TCW formation, 27 genes designated as TCW-upregulated had higher expression levels in fibers at this stage of development as compared to intrusively growing fibers (tFIBa/iFIB) and the apical part of the stem (tFIBa/SAM, **Table 2B**). Furthermore, they were significantly (>4-fold) upregulated in fibers forming TCW as compared to other samples such as xylem, leaves, and roots (**Table 2B**). Since the TCW is known as a cellulose-rich structure, we added five genes related to cellulose biosynthesis: four cellulose synthase genes (*LusCESA1*, *LusCESA3-B*, *LusCESA4*, and *LusCESA8-A*) and the chitinase-like gene (*LusCTL2*). In ramie, the *CESA3* and *CESA8* were among the cell wall related DEGs revealed in comparative transcriptomics of stem bark samples collected from three regions associated with bast fiber development (Xie et al., 2020); in hemp hypocotyl, the orthologs of *A. thaliana CESA4*, *CESA7* and *CESA8* were among the SCW-related genes upregulated at later stages of development (Behr et al., 2016). Along with cellulose, an important role in TCW formation and functioning is

played by pectic RG-I, despite its relatively low content of around 7% (Gorshkova et al., 2018a). For this reason, we chose a number of genes associated with RG-I biosynthesis and modifications, such as *LusRRT1*, *LusGT106-1* (biosynthesis of backbone; shown for *Arabidopsis*, Takenaka et al., 2018), *LusGALS* (biosynthesis of galactan side chains; shown for *Arabidopsis*, Liwanag et al., 2012), *LusBGAL12* (modification of galactan side chains; Roach et al., 2011), and *LusRGL6-A* (putative modification of RG-I backbone; Mokshina et al., 2019). The list of genes for analysis also included three glycosyltransferases belonging to 47 and 31 families (*LusGT47-1* and *LusGT31*) and the gene for cellulose synthase-like protein (*LusCSLG3*). The functions of these genes, even when putative, remain unknown, but they could also refer to RG-I biosynthesis (Gorshkov et al., 2017; Galinousky et al., 2020) and were specifically upregulated in fibers at the TCW-deposition stage (**Table 2B**). Most of the selected genes had a much higher mRNA abundance at an earlier stage of TCW deposition as evidenced by the comparison of TGR values for tFIBa and tFIBb.

Thus, two gene sets containing 17 and 32 genes important for intrusive growth and TCW formation, respectively, were used for further analysis with qPCR.

Expression of Selected Genes in Samples From Various Flax Genotypes

Expression of the selected genes was evaluated in samples of 10 flax genotypes belonging to three groups of flax that have different applications and technical fiber quality: six cultivars of fiber flax (Aramis, Eden, Grant, Laska, Drakkar, and Mogilevsky), two cultivars of linseed flax (Lirina and Orpheus), and two wild flax species (*L. bienne* and *L. angustifolium*). We named the three listed groups as FLW group [from fiber (F) flax, linseed (L) flax, and wild (W) flax] though intraspecific and interspecific taxonomy of the *Linum* genus is difficult and is presented differently by various authors. For instance, *L. bienne* and *L. angustifolium* are considered synonymous in several databases (The Plant List, Flora Europaea, GRIN taxonomy, and NCBI-Taxonomy). Genome analysis with chromosome and molecular markers show that *L. bienne* must be considered as a subspecies of *L. usitatissimum* rather than a separate species (Muravenko et al., 2003), while morphologically *L. angustifolium* and *L. bienne* are closely related (Kutuzova et al., 2015). Based on breeding experiments, *L. angustifolium* was suggested to be the ancestor of Colchian flax, from which the linseed, intermediate, and finally fiber forms consequently evolved; the ancestor of *L. usitatissimum* ssp. *biene* (hereafter, *L. bienne*) could be *L. angustifolium* or Colchian flax (Kutuzova et al., 2015).

The cross-sections of the various representatives of FLW group are exemplified in **Figure 1**. The number of fibers on the cross-sections of FLW representatives ranged from 444 ± 90 in Orpheus to 732 ± 80 in Drakkar, being lowest in linseed cultivars, intermediate in wild species, and the highest in fiber flax (data not shown). The average length of individual fibers obtained from mature plants ranged from 14.7 ± 2.0 mm in *L. angustifolium* to 21.5 ± 1.2 mm in Drakkar. The proportion of fibers longer

TABLE 2 | RNA-Seq data (TGR) for the selected genes upregulated in fibers in the course of intrusive growth **(A)** and during tertiary cell wall formation **(B)** in different flax tissues.


Lus_transcript	At description	Lus symbol	SAM	cPAR	iFIB	tFIBa	tFIBb	sXYLa	sXYLb	LEAF	ROOT	iFIB/iFIBa	iFIB/SAM	iFIBa/cPAR	iFIB/sXYLa	iFIB/LEAF	iFIB/ROOT	At_ID
Lus10024535	O-Glycosyl hydrolases 17	LusGH17-1	107	323	2967	620	505	348	330	242	97	5	28	9	9	12	31	AT4G31140
Lus10019898	WRKY DNA-binding protein 2	LusWRKY2	619	126	2609	51	36	86	17	171	2	51	4	21	30	15	1152	AT5G56270
Lus10012986	Serine/threonine-protein kinase WNK	LusWNK-1	155	199	1920	134	94	49	56	90	9	14	12	10	39	21	219	AT1G60060
Lus10017740	O-Glycosyl hydrolases 17	LusGH17-2	15	87	1353	11	28	4	27	136	10	121	87	15	345	10	138	AT3G24330
Lus10003598	Aux-responsive, dwarf in light 1	LusDFL1	143	92	978	30	47	6	5	125	116	33	7	11	157	8	8	AT5G54510
Lus10023309	Glycosyl hydrolase 17	LusGH17-3A	17	19	736	22	13	3	1	3	134	34	44	39	272	218	5	AT4G18340
Lus10038501	Glycosyl hydrolase 17	LusGH17-3B	3	31	653	1	1	1	1	7	21	653	215	21	653	89	31	AT4G18340
Lus10010167	GRAM domain-containing protein	LusGRAM	68	21	552	29	44	3	1	24	47	19	8	26	204	23	12	AT5G08350
Lus10013561	Pathogenesis-related thaumatin	LusTLP	20	17	547	20	71	2	8	24	31	27	27	32	362	22	18	AT4G38660
Lus10021272	Al-activated malate transporter	LusAMLT	4	8	383	1	3	5	2	12	12	383	98	49	77	31	32	AT3G11680
Lus10014905	Receptor like protein 7	LusRLP7	75	14	330	8	16	1	1	15	1	43	4	24	251	21	259	AT1G47890
Lus10011722	Basic leucine-zipper 5	LusBZIP5	16	28	311	39	21	15	23	8	41	8	19	11	20	40	8	AT3G49760
Lus10033510	Nodulin MtN21/EamA-like transporter	LusWAT1-R	5	56	283	35	1	1	1	8	1	8	60	5	283	33	283	AT5G64700
Lus10028910	Pectin methylesterase inhibitor	LusPMEI10	6	14	224	40	51	17	34	38	6	6	40	16	13	6	39	AT1G62760
Lus10031192	Activator of spomin::LUC2	LusASML2	1	1	130	23	3	5	5	3	1	6	130	130	25	47	130	AT3G12890
Lus10005815	RING/U-box protein	LusATL78	1	3	101	2	3	1	1	4	12	47	91	35	101	24	8	AT1G49230
Lus10017628	Unknown protein	Lus10017628	1	1	77	2	1	1	1	1	1	46	77	77	77	77	77	AT5G37730

(Continued)

TABLE 2 | Continued

Lus_transcript	At description	Lus symbol	SAM	cPAR	iFIB	tFIBa	tFIBb	sXYLa	sXYLb	LEAF	ROOT	tFIBa/ iFIB	tFIBa/ tFIBb	tFIBa/ sXYLa	tFIBa/ LEAF	tFIBa/ ROOT	At_ID
Cellulose																	
Lus10018902	Cellulose synthase 1	LusCESA1	2938	4701	4428	4959	785	10289	3465	1569	8769	1	6	0	3	1	AT4G32410
Lus10007538	Cellulose synthase 3	LusCESA3-B	2308	8222	8607	9706	1508	9476	3433	2491	12148	1	6	1	4	1	AT5G05170
Lus10008225	Cellulose synthase A4	LusCESA4	13	22	47	2464	136	16913	9847	11	97	53	18	0	224	25	AT5G44030
Lus10007296	Cellulose synthase 8	LusCESA8-A	3	1	43	12124	594	41514	18429	26	87	282	20	0	475	140	AT4G18780
Lus10037737	Chitinase-like protein 2	LusCTL2	61	24	26	4317	2049	18230	34486	24	97	164	2	0	176	44	AT3G16920
Lus10010864	Homolog of carrot EP3-3 chitinase	LusCTL19	1	1	1	210	775	3	2	122	1	210	0	64	2	210	AT3G54420
Lus10010866	Homolog of carrot EP3-3 chitinase	LusCTL20	1	1	1	507	551	45	40	39	35	507	1	11	13	14	AT3G54420
Lus10036114	FASCICLIN-like arabinogalactan-protein 11	LusFLA11-2	5	1	2	3570	2061	197	438	8	2	1586	2	18	451	1774	AT5G03170
Lus10036112	FASCICLIN-like arabinogalactan-protein 11	LusFLA11-1	2	3	3	18687	4248	2567	1840	18	1	7302	4	7	1023	15908	AT5G03170
Lus10009917	Expansin A8	LusEXPA8	1	1	1	298	42	13	3	1	1	298	7	23	298	244	AT2G40610
Lus10026418	Lipid transfer protein 3	LusLTP3	1	10	2	1541	307	25	4	11	3	764	5	62	134	452	AT5G59320
Lus10041197	Glycosylphosphatidylinositol-anchored LTP	LusLTPG5	13	1	1	1239	244	54	15	5	2	1239	5	23	261	497	AT3G22600
Rhamnogalacturonan I																	
Lus10028848	Beta-galactosidase 12	LusBGAL12-A	84	21	7	29212	3998	988	687	62	72	3915	7	30	469	408	AT4G26140
Lus10038387	Galactan synthase 2	LusGALS2	339	335	401	11094	1615	1106	857	491	507	28	7	10	23	22	AT5G44670
Lus10007975	RG-I Rhamnosyltransferase	LusRRT1-A	229	348	447	5919	807	1318	475	220	355	13	7	4	27	17	AT5G15740
Lus10004281	Rhamnogalacturonate lyase	LusRGL6-A	1	1	1	38928	3654	2707	817	1	5	38928	11	14	38928	7786	AT2G22620
Lus10013503	RG-I Rhamnosyltransferase	LusRRT1-B	371	439	537	3192	858	1904	897	294	419	6	4	2	11	8	AT5G15740
Lus10027753	O-Fucosyltransferase	LusGT106-1	6	5	2	4702	816	179	25	3	1	2729	6	26	1345	3395	AT3G54100
Lus10023057	Cellulose synthase like G3	LusCSLG3.2	1	1	1	645	215	8	3	122	7	645	3	78	5	98	AT4G23990
Lus10010869	Glycosyltransferase (DUF604)	LusGT31	3	42	8	641	88	326	42	1	1	76	7	2	641	538	AT2G37730
Lus10013790	Exostosin family protein	LusGT47-1	1	5	1	242	171	5	4	146	37	172	1	52	2	7	AT5G03795
Others																	
Lus10022485	BEL1-like homeodomain 2	LusBLH2-A	22	41	68	1749	137	427	187	114	39	26	13	4	15	45	AT4G36870
Lus10020682	P-loop nucleoside triphosphate hydrolase	LusKIN14H	2	1	4	1568	177	58	73	25	10	394	9	27	63	151	AT1G09170
Lus10025322	Putative Na ⁺ /H ⁺ antiporter	LusCHX20	1	1	1	1233	425	42	64	5	30	1233	3	30	254	42	AT3G53720
Lus10011335	SWEET sucrose efflux transporter	LusSWEET15	1	1	1	1049	2386	24	31	20	26	1049	0	43	52	40	AT5G13170
Lus10009582	Lectin protein kinase family	LusCES101	1	1	1	810	660	9	1	46	1	810	1	87	17	810	AT3G16030
Lus10016790	BEL1-like homeodomain 2	LusBLH2-B	8	4	10	808	143	129	75	159	17	81	6	6	5	48	AT4G36870
Lus10037225	Thioredoxin H-type 8	LusTH8	1	1	6	737	1377	42	66	11	1	124	1	17	65	508	AT1G69880
Lus10010766	Metacaspase 1	LusMC1	2	1	28	735	279	19	34	15	4	27	3	39	50	203	AT1G02170
Lus10014387	Protein kinase family protein	LusWAKL20	1	1	1	621	502	50	38	8	1	621	1	12	75	621	AT5G02070
Lus10028015	Isopentenyltransferase 3	LusIPT	9	1	8	264	535	12	94	8	6	34	0	22	33	43	AT3G63110
Lus10014347	Alpha/beta-hydrolases	LusABH	1	1	1	255	1	3	1	1	15	255	255	81	255	17	AT1G52700

SAM, shoot apical meristem (Zhang and Deyholos, 2016); cPAR, cortical parenchyma; iFIB, intrusively growing fibers (Gorshkova et al., 2018b); tFIBa, fibers depositing tertiary cell wall, earlier stage (Gorshkov et al., 2019); tFIBb, fibers depositing tertiary cell wall, later stage (Gorshkov et al., 2017); sXYLa, xylem part, the same segment of the stem as for tFIBa (Gorshkov et al., 2019); sXYLb, xylem part, the same segment of the stem as for tFIBb (Mokshina et al., 2020); LEAF, upper leaves from individual plants under normal conditions (Dmitriev et al., 2016); ROOT, root tips, average for control samples (Dmitriev et al., 2017). When calculating the FC, to avoid division by zero, +1 was added to all the TGR values. The blue-red color scale reflects the level of expression, the green-yellow one – the ratio between the designated samples.

than 30 mm constituted over 20% in fiber flax, while in L and W genotypes it was 8% and 12%, respectively. The proportion of short fibers (below 15 mm) was especially high in *L. angustifolium* (Supplementary Figure S2).

Transcript abundance of the selected genes was measured by RT-qPCR using the reverse transcription method in stem portions containing fibers at the intrusive growth stage and in isolated fibers at the advanced stage of fiber specialization. We did not isolate fiber cells at the intrusive growth stage and extracted RNA from whole stem parts because previously we have analyzed the transcriptome of fibers isolated by laser microdissection at this stage of development and revealed genes especially abundant in fiber cell at the intrusive growth stage as compared to other tissues (Gorshkova et al., 2018b); only such genes were analyzed in the current study to characterize expression at fiber elongation (Supplementary Table S1).

Most of the selected genes had increased expression in the fiber flax cultivars compared with linseed cultivars (*LusBGAL12-A*, *LusRGL6-A*, *LusRRTs*, *LusEXPA8A*, *LusGT31*, *LusFLA11s*, *LusGT47-1*, *LusLTP3*, *LusLTPG5*, *LusGT106-1*, *LusGALS2*, *LusBLHs*, and *LusSWEET15*), while in wild species gene expression was, as a rule, higher than in linseed ones, but lower than in fiber flax. The expression of some genes did not differ significantly in the FLW group (*LusCESA1*, *LusCESA3-B*, *LusCTL2*, *LusCTL20*, *LusMC1*, and *LusWAKL20*) (Supplementary Figures S3, S4).

The expression of *LusGH17-1* and *LusWNK-1* genes, as with *LusASML2*, *LusWNK-1*, and *LusWRKY2*, was relatively stable and did not differ in different samples during the 2-year experimental period. Variability of *Lus10017628* and *LusAMLT* expression was especially pronounced in fiber flax cultivars analyzed during the 2 years, while several other genes (*LusBZIP5*, *LusGH17-3A*, and *LusGRAM*) showed similar tendency (differences in the expression depended on the year) in all analyzed samples. *LusDFL1* was downregulated in fiber flax cultivars compared to linseed cultivars and wild species, this was independent of the year, while *LusWAT1-R* showed the opposite tendency (Figure 2 and Supplementary Figure S3).

The expression level of some selected genes differed according to the year of fiber collecting. For example, the expression of 10 from 49 analyzed genes (*LusABH*, *LusCES101*, *LusCESA4*, *LusCXH20*, *LusCTL20*, *LusCTL2*, *LusCTL20*, *LusFLA11-1*, *LusKIN14H*, and *LusLTP3*) was different when comparing the results of 2018 and 2019. The remaining genes, on the contrary, showed a more stable level of expression (*LusCESA1* and *LusCESA3-B*; Figure 2 and Supplementary Figures S3, S4). The expression of *CESA* genes that encode various isoforms of cellulose synthases is an interesting example here. The expression of the *LusCESA4* gene that encodes glycosyltransferase involved in the synthesis of the secondary cell wall (Taylor et al., 2004), varied strongly in all studied flax samples (Figure 2), while *LusCESA1* and *LusCESA3-B* (involved in primary cell wall synthesis, Desprez et al., 2007) and *LusCESA8-A* (secondary cell wall synthesis, Taylor et al., 2004) genes did not show such differences (Figure 2 and Supplementary Figures S3, S4).

Some of the selected genes were co-expressed (Figure 3). Among the genes upregulated during intrusive growth,

LusBZIP5, *LusGH17s*, *LusGRAM*, and *LusPMEI10* were positively co-expressed with each other, while all of them were negatively co-expressed with *LusALMT12* (Figure 3A). *LusBGAL12-A*, *LusCESA8-A*, *LusFLA11-2*, *LusGT31*, and *LusLTPG5*, all genes that were upregulated during TCW formation, exhibited positive co-expression between each other and with other genes; all of them were negatively co-expressed with *LusIPT* (Figure 3B). Upon clustering of the samples on the base of gene expression, they were mainly combined in accordance with the FLW group (fiber, linseed, wild; Figure 4). Two major clusters were revealed independently from the year of sampling; almost all fiber flax cultivars were grouped together, while another cluster was formed by linseed cultivars with wild species and three fiber flax cultivars (Drakkar, Mogilevsky, and Laska) sampled in 2019. Notably, most of the samples collected through the 2 years were grouped together.

Analysis of Parameters of Technical Fiber Quality in Members of the FLW Group

We evaluated the quality of FLW technical fibers harvested in 2018 and 2019. The fiber flax group differed significantly from other groups; the pronounced differences were also often detected between samples of the same cultivar collected in different years (p -value < 0.05; Figure 5).

Technical fibers scutched from fiber flax cultivars were notably stronger than those from linseed and wild flax (Figure 5). The strongest technical fibers were obtained from Laska and Grant in 2019 – 315 and 298 N, respectively; while in linseed and wild species it was below 100 N. In 2019, the fiber tensile strength was higher than in 2018 in all members of the FLW group, except Eden and *L. angustifolium* which showed no difference.

Flexibility was less variable among FLW representatives than the tensile strength. It had a multidirectional trend change over the two experimental years in the different sample groups. Fiber flax cultivars produced more flexible fiber in 2018, and the remainder of the flax samples did so in 2019. The changes in Orpheus's flexibility were statistically not significant (p -value = 0.081; Figure 5). The most flexible technical fiber was isolated from the stems of Laska cultivar in 2018 (56.2 mm).

The technical stem length was longer in 2018 than in 2019 for most fiber flax cultivars and *L. bienne* (Figure 5). Laska and *L. angustifolium* had a similar stem technical length in both experimental years. The technical stem length difference over the two experimental years can largely be accounted for different soil moisture levels during the period of fast growth. Soil moisture was assessed as sufficiently humid during the period of the fast growth of flax plants in 2018, and dry in 2019: Selyaninov Hydrothermal Coefficient (Taparauskienė and Miseckaitė, 2017) was, respectively, 1.03 and 0.47 during the fast growth period as was determined by the Institute of Flax.

Altogether, a considerable range of fiber quality parameters was demonstrated in various representatives of the FLW group and in plants grown under different weather conditions. This range was further related to selected gene expression values.

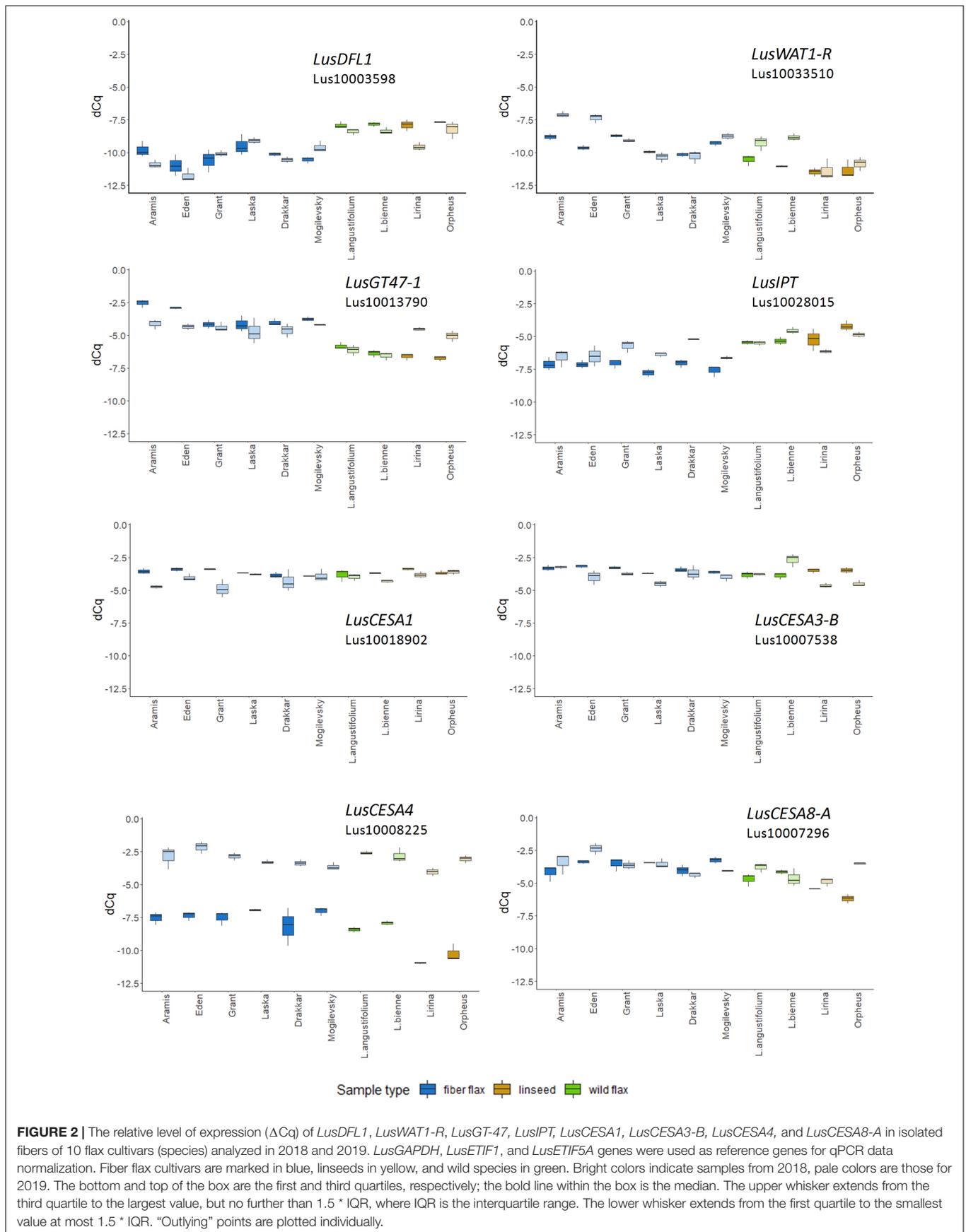
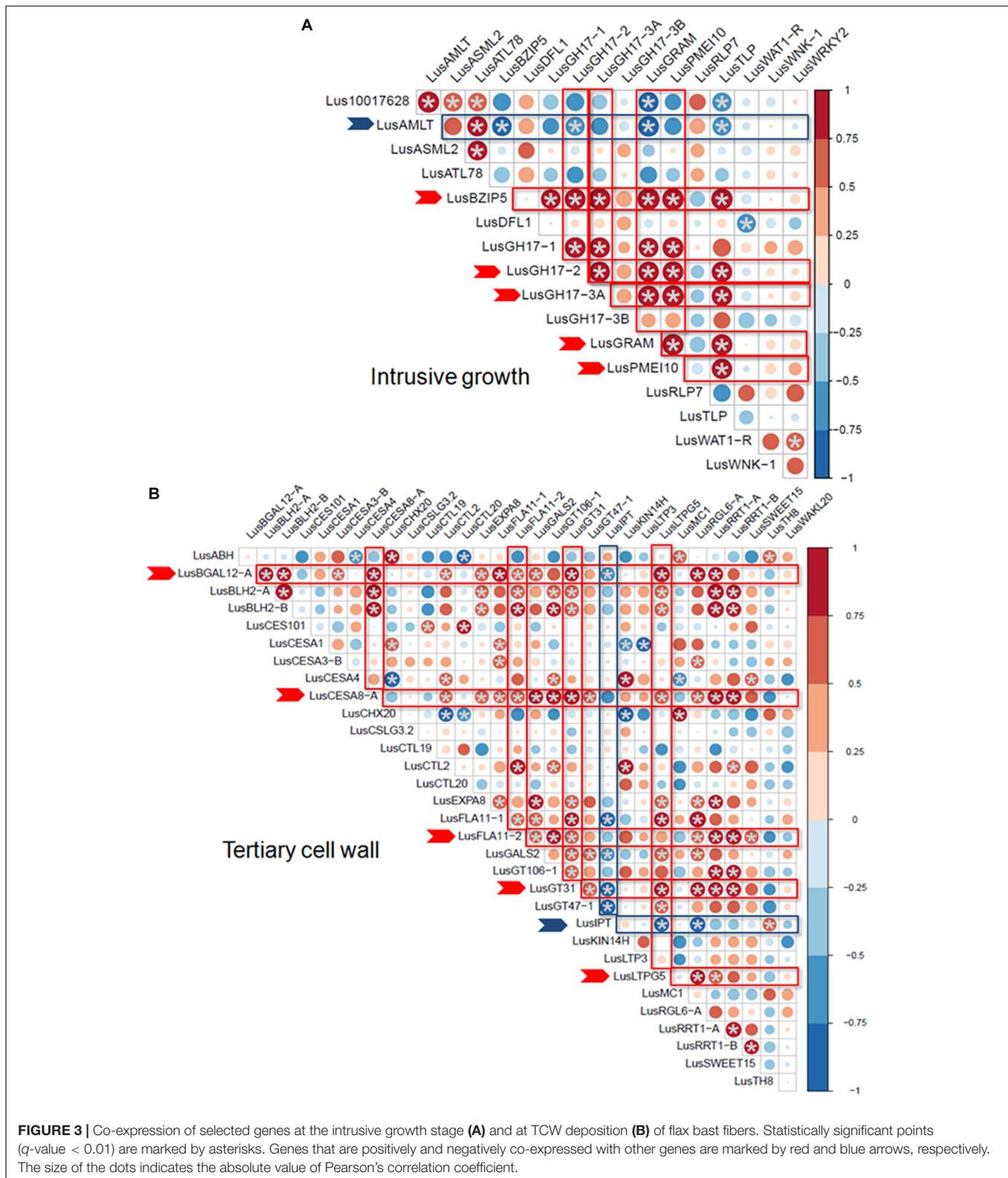


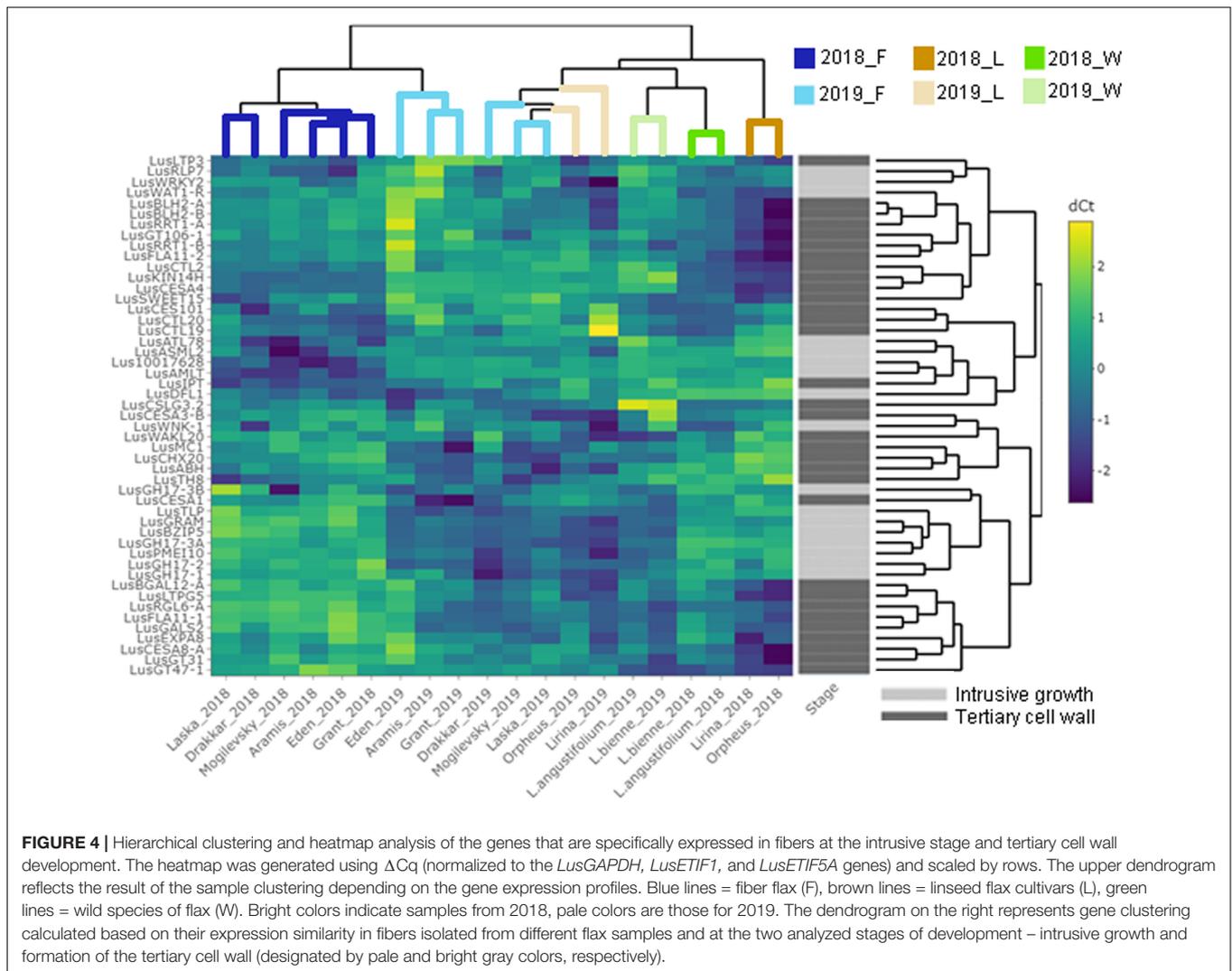
FIGURE 2 | The relative level of expression (ΔCq) of *LusDFL1*, *LusWAT1-R*, *LusGT-47*, *LusIPT*, *LusCESA1*, *LusCESA3-B*, *LusCESA4*, and *LusCESA8-A* in isolated fibers of 10 flax cultivars (species) analyzed in 2018 and 2019. *LusGAPDH*, *LusETIF1*, and *LusETIF5A* genes were used as reference genes for qPCR data normalization. Fiber flax cultivars are marked in blue, linseeds in yellow, and wild species in green. Bright colors indicate samples from 2018, pale colors are those for 2019. The bottom and top of the box are the first and third quartiles, respectively; the bold line within the box is the median. The upper whisker extends from the third quartile to the largest value, but no further than $1.5 \times IQR$, where IQR is the interquartile range. The lower whisker extends from the first quartile to the smallest value at most $1.5 \times IQR$. "Outlying" points are plotted individually.



Regression Analysis

A linear regression was used to model each of the fiber quality parameters (dependent variable) with ΔCq of the studied genes

(independent variables). The R-squared of the single-gene models ranged from 0.00 to 0.75 (Figure 6). The highest scores were obtained for *LusDFL1*, *LusGT47-1*, and *LusTH8* in relation to



fiber tensile strength; for *LusIPT* in relation to fiber flexibility; and for eight genes, *LusGT47-1*, *LusDFL1*, *LusIPT*, *LusLTPG5*, *LusGT3*, *LusRGL6-A*, *LusBGAL12-A*, and *LusBLH2-A* in relation to technical stem length (Supplementary Figure S5). In general, the studied genes were able to explain the technical length of stem and fiber tensile strength better than the flexibility (Figure 6).

Two alternative regularization methods were used to produce multiple regression models of the fiber quality features. The first one is the LASSO method that performs selection of a small set of highly predictive features. The second one is the stepwise forward-selection, which uses more features and fits a more predictive model on the expense of higher bias (Hastie et al., 2017). LASSO models of “tensile strength” and “technical stem length” included two independent variables – *LusGT47-1* and *LusDFL1*, together accounting for 58 and 80% of the variability of the respective parameters (Table 3). The expression level of the *LusTH8* gene also showed relatively high value of the determination coefficient in single-gene regression models (Supplementary Figures S5H,K), but did not meet the specified statistical parameters of the LASSO-analysis (Supplementary

Figure S6). The model of the tensile strength obtained by stepwise forward selection contained twelve independent variables (Table 3 and Supplementary Figure S7) and could explain almost 97% of variation in the tensile strength. The stepwise forward selection model of “technical stem length” included five independent variances and could explain 94% of the technical length variation (Table 3 and Supplementary Figure S7). *LusIPT* was the only gene included in the LASSO model of “flexibility” (Supplementary Figure S6). The stepwise forward-selection model of this parameter included four explanatory genes (Table 3), which together explained 60% of the flexibility dispersion.

Chromosome Localization of the Studied Genes and Polymorphism in Their Coding Sequences

Polymorphic variants of genes can be valuable source material for flax breeding. To check for the presence of polymorphism in coding regions of genes expressed in fibers, RNA-Seq analysis

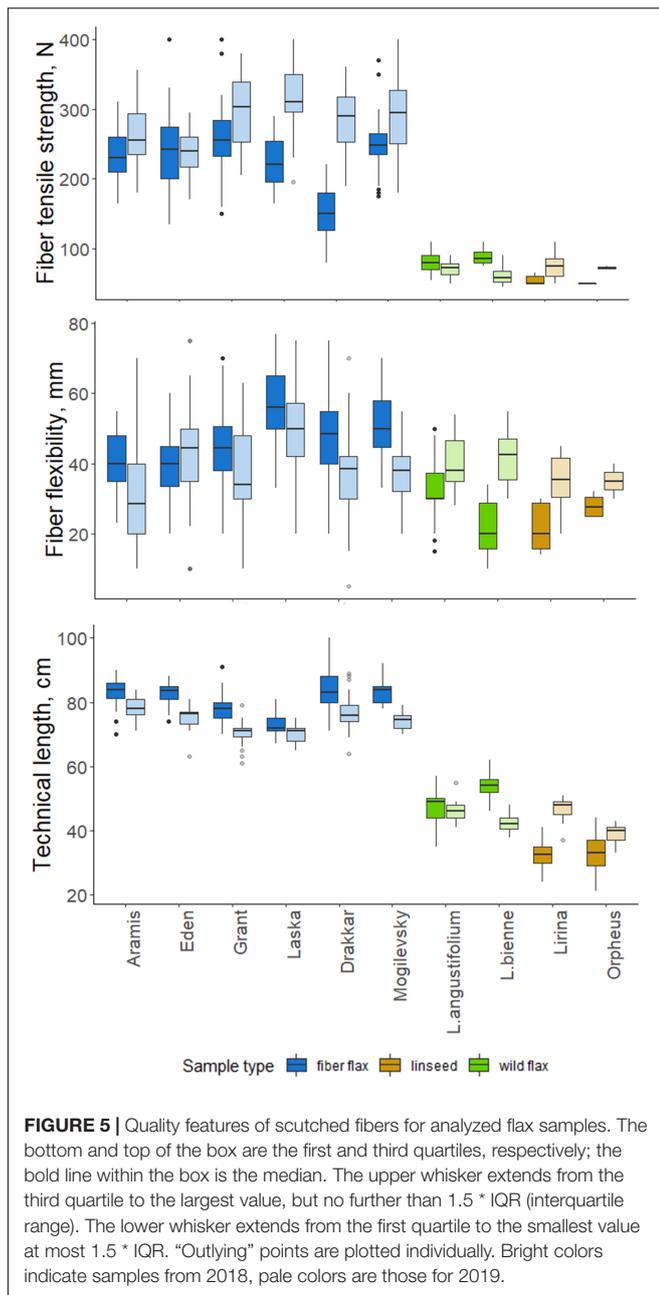


FIGURE 5 | Quality features of scutched fibers for analyzed flax samples. The bottom and top of the box are the first and third quartiles, respectively; the bold line within the box is the median. The upper whisker extends from the third quartile to the largest value, but no further than $1.5 \times$ IQR (interquartile range). The lower whisker extends from the first quartile to the smallest value at most $1.5 \times$ IQR. “Outlying” points are plotted individually. Bright colors indicate samples from 2018, pale colors are those for 2019.

of transcriptomes was performed for fibers isolated from the members of each FLW group, namely fiber flax Grant, linseed Lirina, and wild flax *L. bienne*. We analyzed transcriptomes of isolated fibers collected at the advanced stage of fiber specialization and therefore compared coding sequences (CDS) of the selected 32 TCW-upregulated genes with the reference flax genome. We did not consider genes that were upregulated during fiber intrusive growth, because they are poorly expressed at this stage of fiber development.

We found 176 SNPs for which at least one of the replicates showed variation from the reference (**Supplementary Table S3**). The CDS of eight genes (*LusMCL1*, *LusCTL19*,

LusCTL20, *LusGT31*, *LusWAKL20*, *LusFLA11-1*, *LusFLA11-2*, and *LusLTPG5*) were identical to the reference genome. At least one SNP in at least one biological replicate was present in the sequences of the remaining genes. These variations could arise both from the SNPs of the individual plants and from the inherent errors of the reverse transcriptase reaction (Ji and Loeb, 1992). Only genes that met the SNP-calling criteria in all three biological replicates of each sample were considered as polymorphic (**Table 4**). Ten such genes were revealed having 42 SNPs in total. The distribution of SNPs between the genes was not uniform. The largest number of SNPs was identified in *LusCESA3-B* (11 SNPs, CDS length 3279 bp) and *LusCSLG3.2* (9 SNPs, CDS length 2232 bp). One SNP per gene was found in *LusCES101* (2250 bp), *LusGT47-1* (1716 bp), *LusBLH2-A* (2229 bp), and *LusCTL2* (993 bp). Thus, the level of CDS polymorphisms in analyzed genes varied from 3.4 SNPs per 1 Kbp in *LusCESA3-B* to zero in the eight genes listed above.

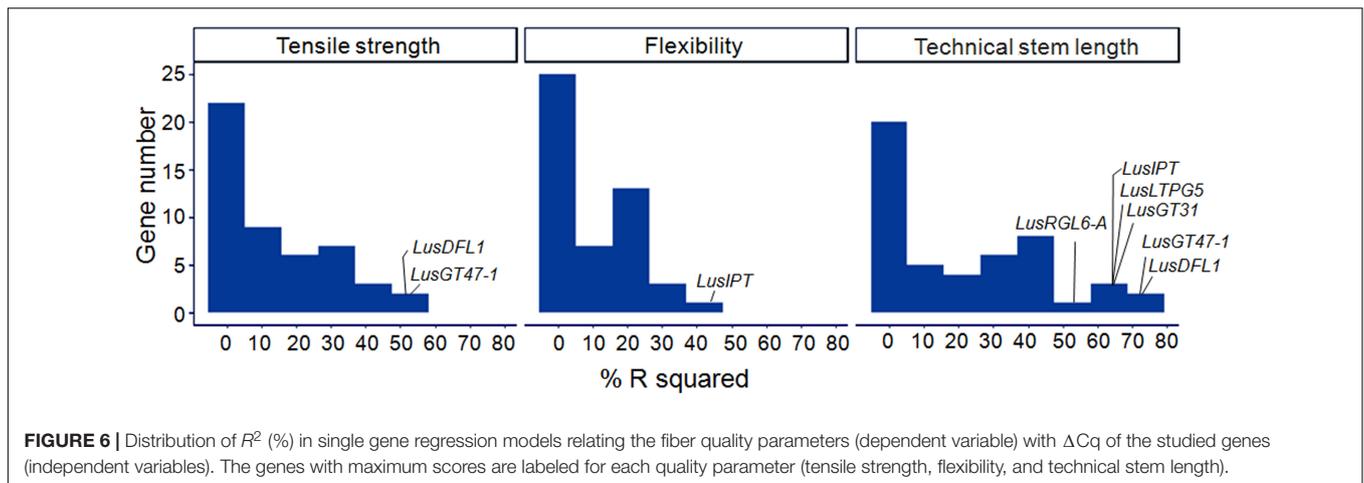
Members of the FLW group had various degrees of SNP specificity (**Table 4**). Three specific SNPs (2 SNPs in *LusKIN14H* and 1 SNP in *LusCESA3-B*) were found in Lirina, and one specific SNP (in *LusGALS2*) in *L. bienne*. Specific SNPs were not found for Grant: all polymorphisms found in the CDS of Grant cultivar genes were also found in *L. bienne* (5 SNPs in *LusCESA8-A*, 2 SNPs in *LusCSLG3.2*, and 1 SNP in *LusBLH2-A*), Lirina (3 SNPs in *LusCSLG3.2*), or *L. bienne* and Lirina combined (25 SNPs) (**Table 4**).

Most discovered point mutations did not cause amino acid substitutions in the primary peptide structure due to redundancy in the genetic code. For instance, all SNPs identified in the *LusCESA3-B* gene were silent mutations. The largest number of missense mutations (four SNPs) was found in the *LusCSLG3.2* gene (**Table 4**). The *in silico* prediction of functional effect of sequence variants with SNAP2 tool⁸ indicated that non-synonymous SNPs (P592T, K17T, P12S; data not shown) found in the protein encoded by *LusCSLG3.2* are most likely not deleterious. According to another *in silico* prediction tool⁹, these three amino acids are likely to be buried rather than exposed on the protein surface, meaning that these residues are more likely to affect the protein conformation but not the binding with other proteins or ligands. One missense mutation was identified in each of *LusCESA8-A* and *LusGT47-1* (**Table 4**). This fact may be of practical interest since the expression levels of these genes are included as independent variables in regression models for fiber quality feature.

Next, we anchored the selected genes into the physical map of flax chromosomes in order to assess their potential linkage with mapped markers (**Supplementary Table S4**). We used chromosome-scale pseudomolecules (You et al., 2018), as well as gene annotation from the phytozome.gov portal. The empirically established genome wide ratio of 239 Kb/cM was used to translate physical distance (in bp) into genetic distance (in cM) (Cloutier et al., 2012). We took into account the gene sequence orientation inside the scaffold, and also scaffold orientation inside the chromosome pseudomolecules when physical distance

⁸<https://roslab.org/services/snap2web/>

⁹<https://predictprotein.org/>



was calculated. We also corrected physical distance for genes lying on scaffolds split into several parts (You et al., 2018; **Supplementary Table S4**).

The largest number of studied genes (seven genes) were anchored onto the third chromosome, no genes were mapped onto chromosome 5 and 10 (**Supplementary Table S4**). The distance between the markers and the mapped genes varied from 0.007 cM (*LusWAKL20*) to 7.437 cM (*LusSWEET15*) owing to the uneven marker saturation of the genetic map. The *LusSWEET15* and *LusASML* genes were located in chromosomes sites least covered by markers. The distances to the nearest mapped marker were 7.437 and 4.609 cM, respectively. The closest to each other genes, *LusGT31*, *LusCTL20*, and *LusCTL19*, were located on the eighth chromosome (**Supplementary Table S4**).

DISCUSSION

Approaches to Find Candidate Genes for Manipulation of Fiber Quality Parameters

Breeders are urgently tasked to clarify molecular links between genetic background and economically valuable crop traits. Bast fiber quality is a complex feature that includes a number of mechanical and physical parameters of fiber bundles that are measured after crop harvest. These parameters are influenced by the post-harvest processing of stems, but are largely determined

during the course of plant development by the reached length of individual fibers, number of fibers and their bundles on a stem cross-section, and the width and composition of the cell wall (Mokshina et al., 2018). These biological parameters are determined by a complex combination of environmental biotic and abiotic factors as well as the genetic status of the plant. Differences in final fiber properties between flax cultivars and species are largely determined by genetic background; this creates a need to identify important genetic elements that can be used in crop quality improvement programs.

Genetic data have been used during different approaches toward improving industrial crops. One of these is the identification of quantitative trait loci (QTL) (Collard et al., 2005). QTLs associated with lint percentage have been identified and mapped for cotton, resulting in the development of simple sequence repeat (SSR) markers for the control of their inheritance (Abdurakhmonov et al., 2007). A promising example of QTL utilization for flax breeding is the model of pasmo (a fungal disease) resistance prediction (He et al., 2019). The developed model includes 500 QTLs as independent variables. Although the determination coefficient of individual QTLs does not exceed 0.25, the model itself has a very high accuracy of pasmo resistance prediction (He et al., 2019). For linseed, SSR markers related to parameters of oil production have been revealed (Soto-Cerda et al., 2014). However, QTL mapping has not yet been successfully used in the identification of flax fiber quality and

TABLE 3 | Linear regressions of scutched fiber quality features on studied gene relative expression levels.

Quality feature	Independent variable	Adjusted R^2	p -Value	Method of regression regularization
Tensile strength	<i>LusDFL1</i> , <i>LusGT47-1</i>	0.58	2.3E-04	LASSO
	<i>LusDFL1</i> , <i>LusTH8</i> , <i>LusWNK-1</i> , <i>LusSWEET15</i> , <i>LusCESA4</i> , <i>LusLTP3</i> , <i>LusCTL2</i> , <i>LusABH</i> , <i>LusCTL19</i> , <i>LusCESA8-A</i> , <i>LusFLA11-2</i> , <i>LusCESA1</i>	0.97	8.8E-06	Stepwise
Flexibility	<i>LusIPT</i>	0.42	2.1E-03	LASSO
	<i>LusIPT</i> , <i>LusKIN14H</i> , <i>LusAMLT</i> , <i>LusDFL1</i>	0.58	1.6E-03	Stepwise
Technical stem length	<i>LusDFL1</i> , <i>LusGT47-1</i>	0.80	4.2E-07	LASSO
	<i>LusGT47-1</i> , <i>LusWRKY2</i> , <i>LusLTP3</i> , <i>LusWAKL20</i> , <i>LusLTPG5</i>	0.94	5.9E-09	Stepwise

yield traits. A significant association has been established with the “plant height” trait for two molecular markers in flax: *LuFAD3B* (encodes desaturases capable of desaturating linoleic acid) ($R^2 = 0.117$, $p < 0.01$) and *CV8824* (encodes leucyl aminopeptidase) ($R^2 = 0.095$, $p < 0.01$) (Nandy, 2012). Low coefficients of the determination indicate that these markers explain a small part of the variation in the phenotypic “plant height” trait. Genome-Wide Association Study (GWAS) has also not been very effective in fiber flax agronomic studies since few genes could be associated with important agronomic traits of

fiber flax (Xie et al., 2018). *Lus10016354* that encodes a xanthoxin dehydrogenase potentially involved in the synthesis of abscisic acid has been suggested as a causal gene for fiber percentage in flax species and *Lus10016125* encoding ABC transporter (ATP-binding cassette transporters) is speculated to be the candidate gene for flax plant height. The candidate genes from the detected two loci related to technical length could not be determined under the applied GWAS approach (Xie et al., 2018).

The current paper presents a more targeted functional approach based on the analysis of gene expression during

TABLE 4 | SNPs detected in coding sequences of TCW-upregulated genes.

Gene name	Gene ID	Scaffold	Scaffold position	SNP	Samples with SNP	CDS position	Triplet substitution	AA substitution
<i>LusCESA8-A</i>	Lus10007296.g	scaffold859	119591	t>c	<i>L. bienne</i> , Grant	c.A2920G	auu>guu	Ile>Val
			120290	g>a	<i>L. bienne</i> , Grant	c.C2322T	auc>auu	Silent
			121671	a>g	<i>L. bienne</i> , Grant	c.T1308C	ggg>ggc	Silent
			121731	g>a	<i>L. bienne</i> , Grant	c.C1248T	cgc>cgu	Silent
			122013	c>t	<i>L. bienne</i> , Grant	c.G966A	ccg>cca	Silent
			122508	a>c	<i>L. bienne</i> , Grant	c.T546G	guu>gug	Silent
			123246	g>a	<i>L. bienne</i> , Grant	c.C165T	agc>agu	Silent
			<i>LusCESA3-B</i>	Lus10007538.g	scaffold259	26161	c>t	<i>L. bienne</i> , Grant, Lirina
26194	t>a	<i>L. bienne</i> , Grant, Lirina				c.T195A	ccu>cca	Silent
28415	g>a	<i>L. bienne</i> , Grant, Lirina				c.G1467A	gag>gaa	Silent
28850	g>a	<i>L. bienne</i> , Grant, Lirina				c.G1699A	ccg>cca	Silent
28964	c>t	<i>L. bienne</i> , Grant, Lirina				c.C1812T	uuc>uuu	Silent
29573	t>c	<i>L. bienne</i> , Grant, Lirina				c.T2151C	uuu>uuc	Silent
29938	c>t	<i>L. bienne</i> , Grant, Lirina				c.C2361T	ggc>ggg	Silent
29983	c>t	<i>L. bienne</i> , Grant, Lirina				c.C2406T	cgc>cgu	Silent
30043	t>c	<i>L. bienne</i> , Grant, Lirina				c.T2466C	auu>auc	Silent
30163	a>g	Lirina				c.A2586G	aga>agg	Silent
30166	t>c	<i>L. bienne</i> , Grant, Lirina				c.T2589C	uuu>uuc	Silent
<i>LusCES101</i>	Lus10009582.g	scaffold1331				314229	t>c	<i>L. bienne</i> , Grant, Lirina
<i>LusGT47-1</i>	Lus10013790.g	scaffold1168	492105	g>t	<i>L. bienne</i> , Grant, Lirina	c.G1108T	guc>uuc	Val>Phe
<i>LusBLH2-B</i>	Lus10016790.g	scaffold903	514884	g>a	<i>L. bienne</i> , Grant, Lirina	c.C948T	gac>gau	Silent
			514893	g>c	<i>L. bienne</i> , Grant, Lirina	c.C939G	ggc>ggg	Silent
			515143	t>g	<i>L. bienne</i> , Grant, Lirina	c.A689C	gac>gcc	Asp>Ala
			515193	t>a	<i>L. bienne</i> , Grant, Lirina	c.A639T	gca>gcu	Silent
<i>LusKIN14H</i>	Lus10020682.g	scaffold303	171179	c>a	Lirina	c.G2446T	gcu>ucu	Ala>Ser
			171294	c>g	<i>L. bienne</i> , Grant, Lirina	c.G2331C	ugg>ugc	Trp>Cys
			171309	t>c	<i>L. bienne</i> , Grant, Lirina	c.A2316G	aca>acg	Silent
			171607	a>g	<i>L. bienne</i> , Grant, Lirina	c.T2130C	cau>cac	Silent
			172853	a>g	Lirina	c.T1432C	uug>cug	Silent
<i>LusBLH2-A</i>	Lus10022485.g	scaffold38	778141	a>g	<i>L. bienne</i> , Grant	c.T1726C	ucg>ccg	Ser>Pro
<i>LusCSLG3.2</i>	Lus10023057.g	scaffold325	175435	g>t	<i>L. bienne</i> , Grant	c.C1774A	ccc>acc	Pro>Thr
			176191	t>a	<i>L. bienne</i> , Grant	c.A1473T	aau>auu	Silent
			177347	a>t	<i>L. bienne</i> , Grant, Lirina	c.T963A	gcu>gca	Silent
			177784	t>c	<i>L. bienne</i> , Grant, Lirina	c.A738G	caa>cag	Silent
			177823	g>a	<i>L. bienne</i> , Grant, Lirina	c.C699T	aac>aau	Silent
			177865	t>g	<i>L. bienne</i> , Grant, Lirina	c.A657C	aga>agc	Arg>Ser
			179212	t>g	Grant, Lirina	c.A50C	aag>acg	Lys>Thr
			179226	g>a	Grant, Lirina	c.C36T	ccc>ucu	Pro>Ser
			179228	g>a	Grant, Lirina	c.C34T		
<i>LusCTL2</i>	Lus10037737.g	scaffold196	1324196	c>t	<i>L. bienne</i> , Grant, Lirina	c.G136A	gug>aug	Val>Met
<i>LusGALS2</i>	Lus10038387.g	scaffold28	1041978	t>c	<i>L. bienne</i> , Grant, Lirina	c.835G	acc>gcc	Thr>Ala
			1042130	t>g	<i>L. bienne</i>	c.A683C	caa>cca	Gln>Pro

specific stages of bast fiber development. The whole pipeline includes the identification of temporal and spatial localization of fiber intrusive elongation and TCW deposition in the stems of developing flax plants (Gorshkova et al., 2003, 2004), transcriptomic analysis of fibers at certain developmental stage (Gorshkov et al., 2017; Gorshkova et al., 2018b), identification of genes with fiber-specific expression by determination of tau score (Mokshina et al., 2020), and the relation of fiber quality parameters to the expression of genes important at key stages of bast fiber development (current paper). The identification of genes contributing to fiber intrusive elongation and TCW deposition is an important step in this pipeline because it reduces information noise in the analysis and determines the effectiveness of the subsequent steps.

Fully understanding that gene expression may not have a direct relation to the activity of the encoded protein owing to various post-translational modifications and the other means of enzyme activity modulation, we checked for the presence of genes with expression levels in developing plants that were highly related to final crop traits. As an argument for such an approach, comparison of various flax genotypes indicated that some of the selected genes kept a relative level of expression (normalized to the reference genes) in accordance to the FLW group despite the different conditions throughout the 2-year experimental period (Figure 2 and Supplementary Figure S3). This was confirmed by clusterization performed for the expression data of all analyzed genes obtained for the 2 years of observation (Figure 4).

Studies that attempt to find a correlation between gene expression and plant traits have been stimulated by the development of transcriptome technology. For example, correlation analysis of growing leaf transcriptome with mature leaf parameters has been performed in a maize recombinant inbred line population by Baute et al. (2015). Hundreds of transcripts putatively linked with 17 traits of potato, representing both well-known and novel associations were described based on gene expression level (Alexandersson et al., 2020). In a study devoted to cotton fiber (trichome) quality, the expression of some genes was evaluated by qPCR. Inbred lines of *Gossypium* species of higher and lower quality showed a correlation between the expression of a number of genes and fiber (trichome) quality: the expression of the *cinnamate-4-hydroxylase* gene was correlated with length uniformity index ($R^2 = 0.73$), the expression of the *Lim domain* gene was correlated with mean length ($R^2 = 0.68$), and the expression of the *pectin methylesterase* gene was correlated with micronaire (resistance to airflow per unit mass of fiber) ($R^2 = 0.55$) (Al-Ghazi et al., 2009). Our approach helped to effectively identify flax genes whose expression level highly correlated to bast fiber strength, flexibility, and technical length.

Genes Whose Expression Levels Are Highly Related to the Fiber Quality Parameters: Possible Functions of the Encoded Products

The set of genes whose expression is highly related to quality parameters includes those genes upregulated both at fiber intrusive elongation and TCW deposition stages, with both

positive and negative correlation coefficients (Supplementary Figure S5). Genes included in the multiple regression models (Table 3) belonged to different co-expression groups (Figure 3). This may indicate the participation of these gene products in various biochemical processes that are important for the formation of high-quality fiber.

The optimal model of the tensile strength obtained by the stepwise method contained 12 genes (Table 3 and Supplementary Figure S7): *LusDFL1*, *LusTH8*, *LusWNK-1*, *LusSWEET15*, *LusCESA4*, *LusLTP3*, *LusCTL2*, *LusABH*, *LusCTL19*, *LusCESA8-A*, *LusFLA11-2*, *LusCESA1*. Three of them are genes for cellulose synthases, including *CESA1* that is considered as specific for primary cell wall and those for secondary cell wall (*CESA8*, *CESA4*). All these genes are activated in fibers during TCW deposition (Mokshina et al., 2017). Also, several potential co-factors (or closely related proteins) for cellulose synthase complex activity are in the list, such as *LusLTP3*, *LusCTL2*, *LusFLA11-2*; *LusCTL19*. The products of these genes can add specific details to the process of cellulose deposition during formation of TCW in phloem fibers. The precise role of FLA proteins and CTL proteins in cellulose biosynthesis is not established yet, but several studies confirmed that some isoforms of these proteins are involved in cell wall biosynthesis (MacMillan et al., 2010; Sánchez-Rodríguez et al., 2012; Huang et al., 2013; Guerriero et al., 2017). Flax *LusLTP3* is predicted to be a 119 amino acid protein; it shows a high homology with the *AtLTP3* protein encoded by *AT5G59320*. A common feature of LTP is their ability to bind and transfer lipids (Finkina et al., 2016), and individual isoforms likely play specific or multiple roles in important biological processes (Yeats and Rose, 2008). LTPs have been found to impact cell wall loosening and promotion of cell extension under tension in tobacco and wheat plants (Nieuwland et al., 2005). *LTP3* of cotton (GenBank accession #AF228333) is reported to interact with the zinc-binding domain of cellulose-synthase *GhCESA1* (Doblin et al., 2002). Cell wall localization (presence of signal peptide) was predicted for *LusLTP3* by TargetP¹⁰. In hemp, *LTP3* gene was also activated in the stem region, containing fibers forming TCW as compared to the upper part of the stem with fibers at an earlier stage of development (Guerriero et al., 2017). The role of *LusWNK-1* [serine/threonine-protein kinase WNK (with no lysine)-like protein, ortholog to *AT1G60060*] from the group of genes upregulated at fiber intrusive growth is difficult to predict, since such protein kinases modulate a wide spectrum of cellular processes through phosphorylating downstream protein substrates (Cao-Pham et al., 2018).

LusDFL1 (*Lus10003598*), expressed at the stage of fiber intrusive growth, was included as the independent variable in regression models that explained each of three studied fiber quality parameters. *LusDFL1* is the ortholog of the *AtDFL1* gene (*DFL1*, *DWARF IN LIGHT 1*; *GH3.6*, *Gretchen Hagen 3.6*, *AT5G54510*) that encodes an IAA-amido synthetase, which conjugates Ala, Asp, Phe, and Trp to auxin (Staswick et al., 2005). *AtDFL1* is involved in auxin signal transduction and inhibits shoot and hypocotyl cell elongation and lateral root cell

¹⁰<http://www.cbs.dtu.dk/services/TargetP/>

differentiation under light (Nakazawa et al., 2001). *Arabidopsis* lines overexpressing this gene accumulate IAA-Asp and are hypersensitive to several auxins. Over-expression of antisense *DFL1* results in larger shoots and an increase in the number of lateral roots. *LusDFL1* has increased expression in oilseed flax cultivars, as well as in wild flax species; this may cause a higher activity of IAA-amido synthetase and lead to faster inactivation of phytohormones in the bast fibers of these genotypes. Lower auxin conjugation may result in longer fibers in fiber flax cultivars, as compared to the remainder of the FLW group (**Supplementary Figure S2**), leading to higher scutched fiber strength. Of note, putative IAA-amido synthetases GH3.9 (four isoforms) were activated in the fiber-containing bark peeled from elongating internode of ramie, as compared to fully elongated internode (Xie et al., 2020). Enhanced expression of *GH3* genes in *Arabidopsis* leads to dwarf plants and de-etiolation in darkness (Hagen and Guilfoyle, 2002). According to study devoted to expression of *GH3* genes in cotton, expression of some *GhGH3* genes may play a role in cotton seed hair initiation and development by regulating the IAA content of ovules (Yu et al., 2018). One more potentially hormone-related gene in the list was *LusABH* (ortholog to *AT1G52700*) encoding alpha/beta hydrolase that belongs to the one of the largest groups of enzymes with diverse catalytic function (Holmquist, 2000); In various species, some ABH domain proteins have been characterized and shown to play roles in lipid homeostasis (Lord et al., 2013). Among other functions, it was suggested that ABH fold serves as the core structure for hormone receptors in the gibberellin, strigolactone, and karrikin signaling pathways in plants (Mindrebo et al., 2016).

Co-expression analysis showed the significant negative correlation of *LusDFL1* with *LusWAT1-R* (walls are thin1-related proteins) (**Figure 3A**). *AtWAT1* encodes a tonoplast-localized auxin transporter, which facilitates auxin export from vacuoles to the cytoplasm (Ranocha et al., 2013). In *Arabidopsis*, it is required for secondary wall formation in fibers (Ranocha et al., 2010). Gene for *WAT1* was also up-regulated in hemp tissues, at the advanced stage of fiber development (Guerriero et al., 2017). In ramie stem, several genes for *WAT1*-related proteins were up-regulated in the bark of elongating internode as compared to non-elongating one (Xie et al., 2020). Thus, two genes encoding proteins involved in keeping auxin sub-cellular homeostasis and specifically upregulated at flax fiber intrusive elongation are in the list of genes whose expression is highly related to the quality of technical fibers.

The *LusSWEET15* gene was related with the tensile strength, it is the orthologous gene of *ATSWEET15* (*AT5G13170*, known as *SAG29*). The SWEET family (Sugar Will Eventually be Exported Transporters) includes plant transmembrane proteins that bidirectionally transport mono- and disaccharides across the membrane along the substrate concentration gradient (Jeena et al., 2019). These transporters are known as key proteins for improving crop stress tolerance and yield through their involvement in carbohydrate partitioning (Wei et al., 2019). Upregulation of *SWEET15* has been demonstrated for seeds and flowers in several plant species (Chen et al., 2015; Wang et al., 2019), as well as in *Arabidopsis* shoots under osmotic stress (Durand et al., 2018). Flax phloem fibers are living cells

with quite active photosynthesis, as indicated by chloroplast functional ultrastructure (Salnikov et al., 2008) and enrichment with transcripts associated with photosynthesis (Gorshkov et al., 2017). At the same time, they are powerful sinks of assimilates due to the active deposition of the cell wall. The activation of sugar transporters may be critical for carbohydrate turnover, balancing of assimilates, and maintenance of osmotic status, all of which accompany TCW deposition in bast fibers. The analysis of cell wall composition and transcriptomics in stem tissues of stinging nettle (*Urtica dioica* L.) revealed up-regulation of *SWEET15* gene in the cortical tissues sampled at the bottom; it was proposed that product of this gene could be involved in the transfer of sucrose to the thickening bast fibers (Xu et al., 2019).

LusTH8 (ortholog to *AT1G69880*) could be considered as potential marker of fiber quality; its expression has a high negative correlation with fiber quality parameters (**Supplementary Figures S5H,K**). In accordance, fiber flax cultivars had a lower expression level of *LusTH8* compared to linseed cultivars and wild species (**Supplementary Figure S4**). Thioredoxins are small and powerful disulfide reductases (Holmgren, 1985), they act as hydrogen donors for redox enzymes (Laurent et al., 1964) through regulating a growing number of biological processes. Through the reduction of intramolecular disulfide bonds, thioredoxin H has been shown to promote the degradation of reserve substances during seed development (Yano et al., 2001). Suppression of *Trx h9* expression in wheat promotes retardation of germination and delayed or decreased expression of the associated enzymes, leading to suppression of preharvest sprouting (Li et al., 2009). It may be surmised that *LusTH8* upregulation in fibers of linseed cultivars and wild species can promote maturation of these cells, while in fiber flax cultivars their prolonged development is necessary for thicker cell wall formation. Another possible role of thioredoxins may be associated with signal transduction and induction of TCW formation. In poplar stems, 36% of identified proteins that were differentially expressed after gravistimulation (i.e., induction of TCW deposition in xylem fibers) have been reported as potential thioredoxin targets; the involvement of thioredoxin H in the first events of signal transduction in inclined poplar stems has been suggested (Azri et al., 2013). Though the expression level of *LusTH8* is demonstrated to be highly related to technical fiber quality, the exact function of *LusTH8* in bast fibers is still unknown.

A list of genes that showed the high value of the adjusted coefficient of determination in the model using technical stem length as the dependent variable included *LusGT47-1*, *LusWRKY2*, *LusLTP3*, *LusWAKL20*, *LusLTPG5*. The *LusGT47-1* gene encodes a glycosyltransferase, which is an enzyme that could potentially be associated with RG-I synthesis (Gorshkov et al., 2017), and it is included in a co-expression network with genes that are likely also related to RG-I biosynthesis and modification (*LusGT31*, *LusGALS2*, *LusBGAL12-A*, *LusRRT1-A*, and *LusRRT1-B*; **Figure 3B**). Probable glycosyltransferase homologous to *At5g03795* (*GT47*) was activated in ramie bark from fully elongated internode region comparing to bark from elongating internode, as well as a number of galactosidases *BGALs* (Xie et al., 2020). *LusWRKY2* is ortholog to *At5g56270* that encodes

TABLE 5 | The most promising genes regarding the manipulation of technical fiber quality as revealed in the current study.

Gene name	Gene ID	Preferential expression	Biological function	Impact on technical fibers properties	Product localization*
<i>LusDFL1</i> IAA-amido synthetase	Lus10003598	Intrusively growing fiber	Phytohormone biosynthesis and signal transduction	Tensile strength; technical stem length	Cytoplasm, expected accuracy 84%
<i>LusIPT</i> Isopentenyl transferase	Lus10028015	TCW depositing fiber	Phytohormone biosynthesis and signal transduction	Flexibility	Chloroplast, expected accuracy 89%
<i>LusGT47-1</i> Glycosyl transferase GT47	Lus10013790	TCW depositing fiber	Rhamnogalacturonan I biosynthesis	Tensile strength; technical stem length	Golgi apparatus membrane, expected accuracy 88%

*According to *LocTree3*¹¹ (Goldberg et al., 2014).

WRKY transcription factor 2, a zinc-finger protein (Chen et al., 2017). Interestingly, the ortholog of WRKY2 in poplar was activated in tension wood samples that formed TCW in xylem fibers (Andersson-Gunnerås et al., 2006). Using transgenic tobacco plants constitutively over-expressing the *Vitis vinifera* L. transcription factor VvWRKY2 it was demonstrated that WRKY2 can modulate expression of genes involved in lignin biosynthesis pathway and cell wall formation (Guillaumie et al., 2010). Thus, this transcription factor could be prominent candidate for regulator involved in thickening of fiber cell wall. *LusWAKL20*, cell wall-associated receptor kinase-like 20, serine/threonine-protein kinase that may function as a signaling receptor of extracellular matrix component (Verica and He, 2002), but exact function in TCW needs to be elucidated in further.

The expression level of the *LusIPT* gene (*Lus10028015*) was related to the fiber flexibility and also had a high score in the stepwise regression model for technical stem length (Table 3 and Supplementary Figures S5C,O). This gene (orthologous to *AT3G63110*, *IPT3*) encodes the adenylate isopentenyltransferase (cytokinin synthase), which is involved in cytokinin biosynthesis (Takei et al., 2001). Among *Arabidopsis IPTs*, *AtIPT3* is expressed in the vasculature throughout the plant (Miyawaki et al., 2004). The promoter activity of *AtIPT3* is high in the phloem but not in the cambium during secondary growth in *Arabidopsis* plants (Matsumoto-Kitano et al., 2008). For hemp fibers forming TCW, it was speculated that cytokinins regulate the initiation of lignin deposition in secondary tissues, while later lignification is tuned by other factors (Behr et al., 2016). According to our data, *IPT3* was activated in flax fibers during deposition of TCW that lacks lignin; the role of cytokinin in fiber development needs to be better characterized.

Several described genes (*LusTH8*, *LusLTP3*, and *LusIPT*) are specifically expressed in developing bast fibers as indicated by a high tau-score (Mokshina et al., 2020). This means that manipulation of their activity through conventional breeding or genome editing would have limited effect on cells other than fibers.

Based on the regression analysis and the biological function investigation we may provide a short list of genes with high potential to influence the technical fiber quality (Table 5).

Analysis of SNPs Using RNA-Seq Reveals a Highly Variable Level of Polymorphism in Coding Sequences of Genes Important for Fiber Development

To incorporate those genes in which expression is highly related to fiber quality into breeding programs, it is important to understand their polymorphism in various flax genotypes. Genome-wide SNP analysis in flax through next generation sequencing was performed by Kumar et al. (2012). In the latter study, SNP discovery was performed through *in silico* analysis of the sequencing data for eight flax genotypes against the whole genome shotgun sequence assembly of flax genotype CDC Bethune. It was shown that 25% of the found SNPs were present in genic regions, of which only 8% were present in coding regions (Kumar et al., 2012).

According to the current data, 25% of the studied genes were totally devoid of polymorphism when compared to the reference genome. In 31% of the analyzed genes, nucleotide sequence variations were reliably identified as present in all three replicates. The identified SNPs were irregularly distributed through the CDS of the studied genes. The highest SNP range was revealed in *LusCESA3-B* CDS (3.4 SNPs per 1 Kbp). Such SNP frequency was significantly higher than in the genome-wide average; the latter varies, according to different investigations, from 0.17 SNP/Kbp (Kumar et al., 2012) to 1.2 SNP/Kbp (Xie et al., 2018). The average frequency of SNPs detected in CDS of all analyzed genes and present in all three replicates was 0.86 SNP/Kbp. Of note was the absence of polymorphism sites specific for the fiber flax cultivar (Table 4). Higher polymorphism in linseed supports the idea that flax was first domesticated for oil, rather than fiber (Allaby et al., 2005).

Many of the revealed CDS variations were not reflected in protein sequences due to redundancy in the genetic code. Those that do result in amino acid substitutions are most likely not deleterious since they were found in phenotypically normal plants. Unfortunately, detailed analysis of the effects of the reported SNPs on the proteins is not possible due to a lack of resolved protein structures and, in some cases, even full amino acid sequences. It can also be hypothesized that the differences in fiber quality related to gene expression in the FLW group are determined by regulation of gene expression via specific transcription factors and other regulatory mechanisms (such as miRNA or methylation sites) rather than by the structure of the

¹¹<https://roslab.org/services/loctree3/results.php?id=29faa499-389e-4436-91e7-c938ee161b60>

encoded proteins. A search for the regulatory elements will be the focus of future studies.

CONCLUSION

Modern plant breeding has embraced many aspects of biological engineering, and it is ready to adopt modern genetic knowledge and approaches. New super-domesticated cultivars that could not be selected under natural environmental conditions or developed by common breeding techniques are now gaining a footing in the agricultural industry (Pérez-de-Castro et al., 2012). This paper proposed an approach that establishes a link between genetic background and valuable crop traits, the key step of which was the identification of genes with tissue and stage characters of expression. Identification was performed using transcriptome data analysis, which allowed the identification of flax genes that were highly correlated to bast fiber strength, flexibility, and technical length.

Based on linear regression modeling of quality traits with gene expression, genes that could be considered as potential markers for flax fiber quality were identified; among them, genes for products, involved in phytohormone biosynthesis and signal transduction (*LusDFL1*, *LusABH*, *LusIPT*, and *LusTH8*), lipid binding and transport (*LusLTP3* and *LusLTPG5*), rhamnolacturonan I biosynthesis and modification (*LusGT47-1*), cellulose deposition (*LusCESA1*, *LusCESA4*, *LusCESA8-A*, *LusFLA11-2*, *LusCTL2*, and *LusCTL19*), carbohydrate (*LusSWEET15*) and malat (*LusAMLT*) transport, protein phosphorylation and cell signaling (*LusWRKY2*, *LusWAKL20*, and *LusWNK-1*) were revealed. Taking into account that these genes are specifically expressed in fibers, their active manipulation through conventional breeding or genome editing could impact bast fiber quality without affecting other tissues. Two missense point mutations were detected in genes included in the regression models (c.A2920G of *LusCESA8-A* and c.G1108T of *LusGT47-1*); these can be useful for flax breeding purposes. Thus, the algorithm to reveal and characterize those genes whose expression is important for certain fiber quality parameters was established, and this may be used in future breeding and genome editing programs.

DATA AVAILABILITY STATEMENT

RNA seq data were deposited in the NCBI Sequence Read Archive (SRA) under accession number: PRJNA658125.

AUTHOR CONTRIBUTIONS

DG, NM, and TG planned and designed the experiments. DG, TP, and MA implemented the study and acquired the data. DG, TP, and NM prepared figures and tables. DG, NM, TP, and TG drafted the manuscript. VB and AK helped to revise the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was partially supported by the Russian Science Foundation (grant #17-76-20049, evaluation of flax fiber quality for different flax cultivars and species, qPCR; grant #19-14-00361, RNA-Seq analysis). Part of the study (data interpretation) was performed by NM and TG with financial support from the government assignment for FRC Kazan Scientific Center of RAS. qPCR was performed on equipment belonging to the CSF-SAC FRC KSC RAS and the Shareable core facilities laboratory “Genome” of the Institute of Genetics and Cytology of NAS of Belarus.

ACKNOWLEDGMENTS

We thank Alena Andronik, the Head of the Linseed Breeding Laboratory (Institute of Flax, Belarus), for growing linseed samples in field experimental plots. We also thank Danat Yermakovich, researcher of bioinformatic group (Institute of Genetics and Cytology, Belarus), for assistance with gene variant calling.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.589881/full#supplementary-material>

Supplementary Figure 1 | Validation of qPCR data for 32 genes (TCW-upregulated) using RNA-Seq data.

Supplementary Figure 2 | The distribution of elementary fiber length (**A**) and average length of individual fibers (**B**) in representatives of FLW groups.

Supplementary Figure 3 | The relative expression levels (qPCR) of each studied gene (in alphabetical order) in all analyzed flax genotypes.

Supplementary Figure 4 | The relative expression level of each studied gene (qPCR), averaged for three groups of flax genotypes (fiber cultivars, linseed cultivars, wild species).

Supplementary Figure 5 | The relation between the technical stem length, the technical fiber strength, and flexibility and the expression level (ΔCq -value) of the selected genes in members of FLW groups.

Supplementary Figure 6 | Dependence modeling between tensile strength, flexibility, and technical length (used as dependent variable) and relative expression level (ΔCq -value) of Intrusive-upregulated and TCW-upregulated genes (used as independent variables) using LASSO method.

Supplementary Figure 7 | Linear regression of flax phenotypic traits on relative expression level (ΔCq -value).

Supplementary Table 1 | The primers used for qPCR in this study and gene sequences according to Phytozome DB.

Supplementary Table 2 | ΔCq values of each studied gene in all analyzed flax samples.

Supplementary Table 3 | Single-nucleotide polymorphisms (SNPs) found in the studied genes (TCW-upregulated).

Supplementary Table 4 | Mapping of analyzed genes (scaffolds, chromosomes, and position on the physical and genetic maps).

REFERENCES

- Abdurakhmonov, I. Y., Buriev, Z. T., Saha, S., Pepper, A. E., Musaev, J. A., Almatov, A., et al. (2007). Microsatellite markers associated with lint percentage trait in cotton, *Gossypium hirsutum*. *Euphytica* 156, 141–156. doi: 10.1007/s10681-007-9361-2
- Ageeva, M. V., Petrovská, B., Kieft, H., Sal'nikov, V. V., Snegireva, A. V., Van Dam, J. E. G., et al. (2005). Intrusive growth of flax phloem fibers is of intercalary type. *Planta* 222, 565–574. doi: 10.1007/s00425-005-1536-2
- Akin, D. E. (2005). Standards for flax fiber. *Standard. News* 33, 22–25.
- Alexandersson, E., Kushwaha, S., Subedi, A., Weighill, D., Climer, S., Jacobson, D., et al. (2020). Linking crop traits to transcriptome differences in a progeny population of tetraploid potato. *BMC Plant Biol.* 20:120. doi: 10.1186/s12870-020-2305-x
- Al-Ghazi, Y., Bourot, S., Arioli, T., Dennis, E. T., and Llewellyn, D. J. (2009). Transcript profiling during fiber development identifies pathways in secondary metabolism and cell wall structure that may contribute to cotton fiber quality. *Plant Cell Physiol.* 50, 1364–1381. doi: 10.1093/pcp/pcp084
- Allaby, R. G., Peterson, G. W., Merriwether, D. A., and Fu, Y. B. (2005). Evidence of the domestication history of flax (*Linum usitatissimum* L.) from genetic diversity of the sad2 locus. *Theor. Appl. Genet.* 112, 58–65. doi: 10.1007/s00122-005-0103-3
- Andersson-Gunnerås, S., Mellerowicz, E. J., Love, J., Segerman, B., Ohmiya, Y., Coutinho, P. M., et al. (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. doi: 10.1111/j.1365-313X.2005.02584.x
- Azri, W., Brunel, N., Franchel, J., Ben Rejeb, I., Jacquot, J. P., Julien, J. L., et al. (2013). Putative involvement of Thioredoxin h in early response to gravitropic stimulation of poplar stems. *J. Plant Physiol.* 170, 707–711. doi: 10.1016/j.jplph.2012.12.017ff
- Baley, C., Perrot, Y., Busnel, F., Guezenoc, H., and Davies, P. (2006). Transverse tensile behaviour of unidirectional plies reinforced with flax fibres. *Mater. Lett.* 60, 2984–2987. doi: 10.1016/j.matlet.2006.02.028
- Baute, J., Herman, D., Coppens, F., De Block, J., Slabbinck, B., Dell'Acqua, M., et al. (2015). Correlation analysis of the transcriptome of growing leaves with mature leaf parameters in a maize RIL population. *Genome Biol.* 16:168. doi: 10.1186/s13059-015-0735-9
- Behr, M., Legay, S., Zizkova, E., Motyka, V., Dobrev, P. I., Hausman, J. F., et al. (2016). Studying secondary growth and bast fiber development: the hemp hypocotyl peeks behind the wall. *Front. Plant Sci.* 7:1733. doi: 10.3389/fpls.2016.01733
- Bergkvist, A., Rusnakova, V., Sindelka, R., Garda, J. M. A., Sjögren, B., Lindh, D., et al. (2010). Gene expression profiling - clusters of possibilities. *Methods* 50, 323–335. doi: 10.1016/j.ymeth.2010.01.009
- Cao-Pham, A. H., Urano, D., Ross-Elliott, T. J., and Jones, A. M. (2018). Nudge-nudge, WNK-WNK (kinases), say no more? *New Phytol.* 220, 35–48. doi: 10.1111/nph.15276
- Chen, F., Hu, Y., Vannozzi, A., Wu, K., Cai, H., Qin, Y., et al. (2017). The WRKY transcription factor family in model plants and crops. *Critic. Rev. Plant Sci.* 36, 311–335. doi: 10.1080/07352689.2018.1441103
- Chen, L. Q., Lin, I. W., Qu, X. Q., Soso, D., McFarlane, H. E., Londono, A., et al. (2015). A cascade of sequentially expressed sucrose transporters in the seed coat and endosperm provides nutrition for the *Arabidopsis* embryo. *Plant Cell* 27, 607–619. doi: 10.1105/tpc.114.134585
- Cloutier, S., Ragupathy, R., Miranda, E., Radovanovic, N., Reimer, E., Walichowski, A., et al. (2012). Integrated consensus genetic and physical maps of flax (*Linum usitatissimum* L.). *Theor. Appl. Genet.* 125, 1783–1795. doi: 10.1007/s00122-012-1953-0
- Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., and Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica* 142, 169–195. doi: 10.1007/s10681-005-1681-5
- Couture, S. J., Asbil, W. L., DiTommaso, A., and Watson, A. K. (2002). Comparison of European fibre flax (*Linum usitatissimum*) cultivars under Eastern Canadian growing conditions. *J. Agric. Crop Sci.* 188, 350–356. doi: 10.1046/j.1439-037X.2002.00582.x
- Couture, S. J., DiTommaso, A., Asbil, W. L., and Watson, A. K. (2004). Evaluation of fibre flax (*Linum usitatissimum* L.) performance under minimal and zero tillage in Eastern Canada. *J. Agric. Crop Sci.* 190, 191–196. doi: 10.1111/j.1439-037X.2004.00092.x
- Day, A., Addi, M., Kim, W., David, H., Bert, F., Mesnage, P., et al. (2005). ESTs from the fibre-bearing stem tissues of flax (*Linum usitatissimum* L.): expression analysis of sequences related to cell wall development. *Plant Biol.* 7, 23–32. doi: 10.1055/s-2004-830462
- Desprez, T., Juraniec, M., Crowell, E. F., Jouy, H., Pochylova, Z., Parcy, F., et al. (2007). Organization of cellulose synthase complexes involved in primary cell wall synthesis in *Arabidopsis thaliana*. *Proc. Nat. Acad. Sci. U.S.A.* 104, 15572–15577. doi: 10.1073/pnas.0706569104
- Diederichsen, A., and Fu, Y. (2006). Phenotypic and molecular (RAPD) differentiation of four infraspecific groups of cultivated flax (*Linum usitatissimum* L. subsp. *usitatissimum*). *Genet. Resour. Crop Evol.* 53, 77–90. doi: 10.1007/s10722-004-0579-8
- Dmitriev, A. A., Krasnov, G. S., Rozhmina, T. A., Novakovskiy, R. O., Snezhkina, A. V., Fedorova, M. S., et al. (2017). Differential gene expression in response to *Fusarium oxysporum* infection in resistant and susceptible genotypes of flax (*Linum usitatissimum* L.). *BMC Plant Biol.* 17(Suppl. 2):253. doi: 10.1186/s12870-017-1192-2
- Dmitriev, A. A., Kudryavtseva, A. V., Krasnov, G. S., Koroban, N. V., Speranskaya, A. S., Krinitsina, A. A., et al. (2016). Gene expression profiling of flax (*Linum usitatissimum* L.) under edaphic stress. *BMC Plant Biol.* 16:237. doi: 10.1186/s12870-016-0927-9
- Doblin, M. S., Kurek, I., Jacob-Wilk, D., and Delmer, D. P. (2002). Cellulose biosynthesis in plants: from genes to rosettes. *Plant Cell Physiol.* 43, 1407–1420. doi: 10.1093/pcp/pcf164
- Durand, M., Mainson, D., Porcheron, B., Maurousset, L., Lemoine, R., and Pourtau, N. (2018). Carbon source-sink relationship in *Arabidopsis thaliana*: the role of sucrose transporters. *Planta* 247, 587–611. doi: 10.1007/s00425-017-2807-4
- Dyagilev, A., Biziuk, A., Kogan, A. (2016). “Estimation and prediction of long scutched flax spinning ability,” in *Proceedings of The 90th Textile Institute World Conference*, Poznań, 66–72.
- Efroymson, M. A. (1960). “Multiple regression analysis,” in *Mathematical Methods for Digital Computers*, eds A. Ralston and H. S. Wilf (New York, NY: Wiley), 191–203.
- Elhaak, M. A., El-Shourbagy, M. N., and El-Nagar, R. (1999). Effect of edaphic factors on technical properties of flax fibre. *J. Agric. Crop Sci.* 182, 113–120. doi: 10.1046/j.1439-037x.1999.00273.x
- Esau, K. (1943). Vascular differentiation in the vegetative shoot of *Linum*. III. The origin of the bast fibers. *Am. J. Bot.* 30, 579–586. doi: 10.1002/j.1537-2197.1943.tb10302.x
- Fenart, S., Ndong, Y.-P., Duarte, J., Rivière, N., Wilmer, J., van Wuytswinkel, O., et al. (2010). Development and validation of a flax (*Linum usitatissimum* L.) gene expression oligo microarray. *BMC Genomics* 11:592. doi: 10.1186/1471-2164-11-592
- Finkina, E. I., Melnikova, D. N., Bogdanov, I. V., and Ovchinnikova, T. V. (2016). Lipid transfer proteins as components of the plant innate immune system: structure, functions, and applications. *Acta Nat.* 8, 47–61. doi: 10.32607/20758251-2016-8-2-47-61
- Galili, T., O'Callaghan, A., Sidi, J., and Sievert, C. (2018). Heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics* 34, 1600–1602. doi: 10.1093/bioinformatics/btx657
- Galinosky, D., Padvitski, T., Mokshina, N., Gorshkov, O., Khotyleva, L., Gorshkova, T., et al. (2020). Expression of cellulose synthase-like genes in two phenotypically distinct flax (*Linum usitatissimum* L.) subspecies. *Genet. Resour. Crop Evol.* 67, 1821–1837. doi: 10.1007/s10722-020-00943-2
- Goldberg, T., Hecht, M., Hamp, T., Karl, T., Yachdav, G., Nielsen, H., et al. (2014). LocTree3 prediction of localization. *Nucleic Acids Res.* 42, W350–W355.
- Gorshkov, O., Chernova, T., Mokshina, N., Gogoleva, N., Suslov, D., Tkachenko, A., et al. (2019). Intrusive growth of phloem fibers in flax stem: integrated

- analysis of miRNA and mRNA expression profiles. *Plants* 8:47. doi: 10.3390/plants8020047
- Gorshkov, O., Mokshina, N., Gorshkov, V., Chemikosova, S., Gogolev, Y., and Gorshkova, T. (2017). Transcriptome portrait of cellulose-enriched flax fibers at advanced stage of specialization. *Plant Mol. Biol.* 93, 431–449. doi: 10.1007/s11103-016-0571-7
- Gorshkova, T., Chernova, T., Mokshina, N., Ageeva, M., and Mikshina, P. (2018a). Plant “muscles”: fibers with a tertiary cell wall. *New Phytol.* 218, 66–72. doi: 10.1111/nph.14997
- Gorshkova, T., Chernova, T., Mokshina, N., Gorshkov, V., Kozlova, L., and Gorshkov, O. (2018b). Transcriptome analysis of intrusively growing flax fibers isolated by laser microdissection. *Sci. Rep.* 8:14570. doi: 10.1038/s41598-018-32869-2
- Gorshkova, T., Mokshina, N., Chernova, T., Ibragimova, N., Salnikov, V., Mikshina, P., et al. (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. doi: 10.1104/pp.15.00690
- Gorshkova, T. A., Chemikosova, S. B., Salnikov, V. V., Pavlencheva, N. V., Gurjanov, O. P., Stoll-Smits, T., et al. (2004). Occurrence of cell-specific galactan is coinciding with bast fibre developmental transition in flax. *Ind. Crops Prod.* 19, 217–224. doi: 10.1016/j.indcrop.2003.10.002
- Gorshkova, T. A., Salnikov, V. V., Chemikosova, S. B., Ageeva, M. V., Pavlencheva, N. V., and van Dam, J. E. G. (2003). The snap point: a transition point in *Linum usitatissimum* bast fiber development. *Ind. Crops Prod.* 18, 213–221. doi: 10.1016/S0926-6690(03)00043-8
- Guerriero, G., Behr, M., Legay, S., Mangeot-Peter, L., Zorzan, S., Ghoniem, M., et al. (2017). Transcriptomic profiling of hemp bast fibres at different developmental stages. *Sci. Rep.* 7:4961. doi: 10.1038/s41598-017-05200-8
- Guillaumie, S., Mzid, R., Léon, C., Hichri, I., Destrac-Irvine, A., et al. (2010). The grapevine transcription factor WRKY2 influences the lignin pathway and xylem development in tobacco. *Plant Mol. Biol.* 72, 215–234. doi: 10.1007/s11103-009-9563-1
- Hagen, G., and Guilfoyle, T. (2002). Auxin-responsive gene expression: genes, promoters and regulatory factors. *Plant Mol. Biol.* 49, 373–385. doi: 10.1023/A:1015207114117
- Hastie, T., Tibshirani, R., and Tibshirani, R. J. (2017). Extended comparisons of best subset selection, forward stepwise selection, and the lasso. *arXiv* Available online at: <https://arxiv.org/abs/1707.08692> (accessed October 18, 2019).
- He, L., Xiao, J., Rashid, K. Y., Jia, G., Li, P., Yao, Z., et al. (2019). Evaluation of genomic prediction for psmo resistance in flax. *Int. J. Mol. Sci.* 20:E359. doi: 10.3390/ijms20020359
- Holmgren, A. (1985). Thioredoxin. *Annu. Rev. Biochem.* 54, 237–271. doi: 10.1146/annurev.bi.54.070185.001321
- Holmquist, M. (2000). Alpha/beta-hydrolase fold enzymes: structures, functions and mechanisms. *Curr. Protein Pept. Sci.* 1, 209–235. doi: 10.2174/1389203003381405
- Huang, G. Q., Gong, S. Y., Xu, W. L., Li, W., Li, P., Zhang, C. J., et al. (2013). A fasciclin-like arabinogalactan protein, GhFLA1, is involved in fiber initiation and elongation of cotton. *Plant Physiol.* 161, 1278–1290. doi: 10.1104/pp.112.203760
- Huis, R., Hawkins, S., and Neutelings, G. (2010). Selection of reference genes for quantitative gene expression normalization in flax (*Linum usitatissimum* L.). *BMC Plant Biol.* 19:71. doi: 10.1186/1471-2229-10-71
- Jawaid, M., and Khalil, A. (2011). Cellulosic/synthetic fibre reinforced polymer hybrid composites: a review. *Carbohydr. Polym.* 86, 1–18. doi: 10.1016/j.carbpol.2011.04.043
- Jeena, G. S., Kumar, S., and Shukla, R. K. (2019). Structure, evolution and diverse physiological roles of SWEET sugar transporters in plants. *Plant Mol. Biol.* 100, 351–365. doi: 10.1007/s11103-019-00872-4
- Ji, J., and Loeb, L. A. (1992). Fidelity of HIV-1 reverse transcriptase copying RNA *in vitro*. *Biochemistry* 31, 954–958. doi: 10.1021/bi00119a002
- Kim, D., Landmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–U121. doi: 10.1038/Nmeth.3317
- Kumar, S., You, F. M., and Cloutier, S. (2012). Genome wide SNP discovery in flax through next generation sequencing of reduced representation libraries. *BMC Genomics* 13:684. doi: 10.1186/1471-2164-13-684
- Kutuzova, S. N., Porokhvinova, E. A., and Pendinen, G. I. (2015). Origin and evolution of *Linum usitatissimum* L. *Proc. Appl. Bot. Genet. Breed.* 176, 436–455. doi: 10.30901/2227-8834-2015-4-436-455
- Kymäläinen, H. R., and Sjöberga, A. M. (2008). Flax and hemp as raw materials for thermal insulations. *Build. Environ.* 43, 1261–1269. doi: 10.1016/j.buildenv.2007.03.006
- Laurent, T. C., Moore, E. C., and Reichard, P. (1964). Enzymatic synthesis of deoxyribo-nucleotides: isolation and characterization of thioredoxin, the hydrogen donor from *Escherichia coli* B. *J. Biol. Chem.* 239, 3436–3444.
- Li, Y.-C., Ren, J.-P., Cho, M.-J., Zhou, S.-M., Kim, Y.-B., Guo, H.-X., et al. (2009). The level of expression of thioredoxin is linked to fundamental properties and applications of wheat seeds. *Mol. Plant* 2, 430–441. doi: 10.1093/mp/ssp025
- Liwanag, A. J., Ebert, B., Verhertbruggen, Y., Rennie, E. A., Rautengarten, C., Oikawa, A., et al. (2012). Pectin biosynthesis: GALSI in *Arabidopsis thaliana* is a β -1,4-galactan β -1,4-galactosyltransferase. *Plant Cell* 24, 5024–5036. doi: 10.1105/tpc.112.106625
- Lord, C. C., Thomas, G., and Brown, J. M. (2013). Mammalian alpha beta hydrolase domain (ABHD) proteins: lipid metabolizing enzymes at the interface of cell signaling and energy metabolism. *Biochim. Biophys. Acta* 1831, 792–802. doi: 10.1016/j.bbali.2013.01.002
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8
- MacMillan, C. P., Mansfield, S. D., Stachurski, Z. H., Evans, R., and Southerton, S. G. (2010). Fasciclin-like arabinogalactan proteins: specialization for stem biomechanics and cell wall architecture in *Arabidopsis* and *Eucalyptus*. *Plant J.* 62, 689–703. doi: 10.1111/j.1365-313X.2010.04181.x
- Masaki, T., Mitsui, N., Tsukagoshi, H., Nishii, T., Hattori, T., Morikami, A., et al. (2005). Activation tagging of a gene for a protein with novel class of CCT-domain activates expression of a subset of sugar-inducible genes in *Arabidopsis thaliana*. *Plant J.* 43, 142–152. doi: 10.1111/j.1365-313X.2005.02439.x
- Matsumoto-Kitano, M., Kusumoto, T., Tarkowski, P., Kinoshita-Tsujimura, K., Václavíková, K., Miyawaki, K., et al. (2008). Cytokinins are central regulators of cambial activity. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20027–20031. doi: 10.1073/pnas.0805619105
- Meinshausen, N., and Buhlmann, P. (2010). Stability selection (with discussion). *J.R. Stat. Soc.* 72, 417–473. doi: 10.1111/j.1467-9868.2010.00740.x
- Mindrebo, J. T., Nartey, C. M., Seto, Y., Burkart, M. D., and Noel, J. P. (2016). Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom. *Curr. Opin. Struct. Biol.* 41, 233–246. doi: 10.1016/j.sbi.2016.08.005
- Miyawaki, K., Matsumoto-Kitano, M., and Kakimoto, T. (2004). Expression of cytokinin biosynthetic isopentenyltransferase genes in *Arabidopsis*: tissue specificity and regulation by auxin, cytokinin, and nitrate. *Plant J.* 37, 128–138. doi: 10.1046/j.1365-313x.2003.01945.x
- Mokshina, N., Chernova, T., Galinousky, D., Gorshkov, O., and Gorshkova, T. (2018). Key stages of fiber development as determinants of bast fiber yield and quality. *Fibers* 6:20. doi: 10.3390/fib6020020
- Mokshina, N., Gorshkov, O., Galinousky, D., and Gorshkova, T. (2020). Genes with bast fiber-specific expression in flax plants - Molecular keys for targeted fiber crop improvement. *Ind. Crops Prod.* 152:112549. doi: 10.1016/j.indcrop.2020.112549
- Mokshina, N., Gorshkov, O., Ibragimova, N., Chernova, T., and Gorshkova, T. (2017). Cellulosic fibres of flax recruit both primary and secondary cell wall cellulose synthases during deposition of thick tertiary cell walls and in the course of graviresponse. *Funct. Plant Biol.* 44, 820–831. doi: 10.1071/FP17105
- Mokshina, N., Gorshkova, T., and Deyholos, M. K. (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. doi: 10.1371/journal.pone.0097949
- Mokshina, N., Makshakova, O., Nazipova, A., Gorshkov, O., and Gorshkova, T. (2019). Flax rhamnogalacturonan lyases: phylogeny, differential expression and modeling of protein structure. *Physiol. Plant* 167, 173–187. doi: 10.1111/pl.12880
- Muravenko, O. V., Lemesh, V. A., Samatadze, T. E., Amosova, A. V., Grushetskaya, Z. E., Popov, K. V., et al. (2003). Genome comparisons with chromosomal and molecular markers for three closely related flax species and their hybrids. *Russ. J. Genet.* 39, 414–421. doi: 10.1023/A:1023309831454
- Nakazawa, M., Yabe, N., Ichikawa, T., Yamamoto, Y. Y., Yoshizumi, T., Hasunuma, K., et al. (2001). DFL1, an auxin-responsive GH3 gene homologue, negatively regulates shoot cell elongation and lateral root formation, and positively regulates the light response of hypocotyl length. *Plant J.* 25, 213–221. doi: 10.1046/j.1365-313x.2001.00957.x
- Nandy, S. (2012). *Elucidating the Genetic Basis of Bast Fibre Production in Flax (Linum usitatissimum L.)*. PhD thesis. Saskatoon, SK: University of Saskatchewan.
- Nieuwland, J., Feron, R., Huisman, B. A., Fasolino, A., Hilbers, C. W., Derksen, J., et al. (2005). Lipid transfer proteins enhance cell wall extension in tobacco. *Plant Cell* 17, 2009–2019. doi: 10.1105/tpc.105.032094
- No author listed. (1988). *Methodical Guidelines for Studying the Collection of Flax (Linum usitatissimum L.)*. Leningrad: 28.
- Pérez-de-Castro, A. M., Vilanova, S., Cañizares, J., Pascual, L., Blanca, J. M., Díez, M. J., et al. (2012). Application of genomic tools in plant breeding. *Curr. Genomics* 13, 179–195. doi: 10.2174/138920212800543084
- Perte, M., Kim, D., Perte, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT StringTie and Ballgown. *Nat. Protoc.* 11, 1650–1667. doi: 10.1038/nprot.2016.095
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Ranocha, P., Denancé, N., Vanholme, R., Freyrier, A., Martinez, Y., Hoffmann, L., et al. (2010). 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.* 63, 469–483. doi: 10.1111/j.1365-313X.2010.04256.x
- Ranocha, P., Dima, O., Nagy, R., Felten, J., Corratgé-Faillie, C., Novák, O., et al. (2013). *Arabidopsis* WAT1 is a vacuolar auxin transport facilitator required for auxin homeostasis. *Nat. Commun.* 4:2625. doi: 10.1038/ncomms3625
- Rennebaum, H., Grimm, E., Warnstorff, K., and Diepenbrock, W. (2002). Fibre quality of linseed (*Linum usitatissimum* L.) and the assessment of genotypes for use of fibres as a by-product. *Ind. Crops Prod.* 16, 201–215. doi: 10.1016/S0926-6690(02)00048-1
- Roach, M. J., and Deyholos, M. K. (2007). Microarray analysis of flax (*Linum usitatissimum* L.) stems identifies transcripts enriched in fiber-bearing phloem tissues. *Mol. Genet. Genomics* 278, 149–165. doi: 10.1007/s00438-007-0241-1
- Roach, M. J., Mokshina, N. Y., Badhan, A., Snegireva, A. V., Hobson, N., Deyholos, M. K., et al. (2011). Development of cellulosic secondary walls in flax fibers requires beta-galactosidase. *Plant Physiol.* 156, 1351–1363. doi: 10.1104/pp.111.172676
- Rowland, G. G., Hormis, Y. A., and Rashid, K. Y. (2002). CDC Bethune flax. *Can. J. Plant Sci.* 82, 101–102. doi: 10.4141/P01-066
- Salnikov, V. V., Ageeva, M. V., and Gorshkova, T. A. (2008). Homofusion of Golgi secretory vesicles in flax phloem fibers during formation of the gelatinous secondary cell wall. *Protoplasma* 233, 269–273. doi: 10.1007/s00709-008-0011-x
- Sánchez-Rodríguez, C., Bauer, S., Hématy, K., Saxe, F., Ibáñez, A. B., Vodermaier, V., et al. (2012). Chitinase-like1/pom-pom1 and its homolog CTL2 are glucan-interacting proteins important for cellulose biosynthesis in *Arabidopsis*. *Plant Cell* 24, 589–607. doi: 10.1105/tpc.111.094672
- Smole, M. S., Hribernik, S., Stana-Kleinschek, K., and Kreze, T. (2013). “Plant fibers for textile and technical applications,” in *Advances in Agrophysical Research*, ed. S. Grundas (Rijeka: IntechOpen), 369–387.
- Snegireva, A., Chernova, T., Ageeva, M., Lev-Yadun, S., and Gorshkova, T. (2015). Intrusive growth of primary and secondary phloem fibres in hemp stem determines fibre-bundle formation and structure. *AoB Plants* 7:lv061. doi: 10.1093/aobpla/plv061
- Snegireva, A. V., Ageeva, M. V., Amenitskii, S. I., Chernova, T. E., Ebskamp, M., and Gorshkova, T. A. (2010). Intrusive growth of sclerenchyma fibers. *Russ. J. Plant Physiol.* 57, 342–355. doi: 10.1134/S1021443710030052
- Soto-Cerda, B. J., Duguid, S., Booker, H., Rowland, G., Diederichsen, A., and Cloutier, S. (2014). Genomic regions underlying agronomic traits in linseed (*Linum usitatissimum* L.) as revealed by association mapping. *J. Integr. Plant Biol.* 56, 75–87. doi: 10.1111/jipb.12118
- Staswick, P. E., Serban, B., Rowe, M., Tiryaki, I., Maldonado, M. T., Maldonado, M. C., et al. (2005). Characterization of an *Arabidopsis* enzyme family that conjugates amino acids to indole-3-acetic acid. *Plant Cell* 17, 616–627. doi: 10.1105/tpc.104.026690
- STB1195. (2008). *Flax Fiber Scutched Long. Introduced 2008-04-30*. Minsk: State Standard of the Republic of Belarus.
- Takei, K., Sakakibara, H., and Sugiyama, T. (2001). Identification of genes encoding adenylate isopentenyltransferase, a cytokinin biosynthesis enzyme, in *Arabidopsis thaliana*. *J. Biol. Chem.* 276, 26405–26410. doi: 10.1074/jbc.M102130200
- Takenaka, Y., Kato, K., Ogawa-Ohnishi, M., Tsuruhama, K., Kajiuira, H., Yagyuet, K., et al. (2018). Pectin RG-I rhamnosyltransferases represent a novel plant-specific glycosyltransferase family. *Nat. Plants* 4, 669–676. doi: 10.1038/s41477-018-0217-7
- Taparuskienė, L., and Miseckaitė, O. (2017). Comparison of watermark soil moisture content with Selyaninov hydrothermal coefficient. *Agrofor. Int. J.* 2, 106–115. doi: 10.7251/AGRENG1702106T
- Taylor, N. G., Gardiner, J. C., Whiteman, R., and Turner, S. R. (2004). Cellulose synthesis in the *Arabidopsis* secondary cell wall. *Cellulose* 11, 329–338. doi: 10.1023/B:CELL.0000046405.11326.a8
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B* 58, 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paep, A., et al. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3:research0034.1. doi: 10.1186/gb-2002-3-7-research0034
- Verica, J. A., and He, Z. H. (2002). The cell wall-associated kinase (WAK) and WAK-like kinase gene family. *Plant Physiol.* 129, 455–459. doi: 10.1104/pp.011028
- Wang, S. D., Yokosho, K., Guo, R., Whelan, J., Ruan, Y. L., Ma, J. F., et al. (2019). The soybean sugar transporter GmSWEET15 mediates sucrose export from endosperm to early embryo. *Plant Physiol.* 180, 2133–2141. doi: 10.1104/pp.19.00641
- Wang, Z. W., Hobson, N., Galindo, L., Zhu, S., Shi, D., McDill, J., et al. (2012). The genome of flax (*Linum usitatissimum*) assembled de novo from short shotgun sequence reads. *Plant J.* 72, 461–473. doi: 10.1111/j.1365-313x.2012.05093.x
- Wei, T., and Simko, V. (2017). *R package “corrplot”: Visualization of a Correlation Matrix (Version 0.84)*. Available online at: <https://github.com/taiyun/corrplot> (accessed October 18, 2019).
- Wei, Y., Xiao, D., Zhang, C., and Hou, X. (2019). The expanded SWEET gene family following whole genome triplication in *Brassica rapa*. *Genes* 10:722. doi: 10.3390/genes10090722
- Xie, D., Dai, Z., Yang, Z., Sun, J., Zhao, D., Yang, X., et al. (2018). Genome-wide association study identifying candidate genes influencing important agronomic traits of flax (*Linum usitatissimum* L.) using SLAF-seq. *Front. Plant Sci.* 8:2232. doi: 10.3389/fpls.2017.02232
- Xie, J., Li, J., Jie, Y., Xie, D., Yang, D., Shi, H., et al. (2020). Comparative transcriptomics of stem bark reveals genes associated with bast fiber development in *Boehmeria nivea* L. *gaud* (ramie). *BMC Genomics* 21:40. doi: 10.1186/s12864-020-6457-8
- Xu, X., Backes, A., Legay, S., Berni, R., Faleri, C., Gatti, E., et al. (2019). Cell wall composition and transcriptomics in stem tissues of stinging nettle (*Urtica dioica* L.): spotlight on a neglected fibre crop. *Plant Direct.* 3:e00151. doi: 10.1002/pld3.151
- Yano, H., Wong, J. H., Lee, Y. M., Cho, M. J., and Buchanan, B. B. (2001). Redox changes accompanying the degradation of seed storage proteins in germinating rice. *Plant Cell Physiol.* 42, 879–883. doi: 10.1093/pcp/pce119
- Yeats, T. H., and Rose, J. K. (2008). The biochemistry and biology of extracellular plant lipid-transfer proteins (LTPs). *Protein Sci.* 17, 191–198. doi: 10.1110/ps.073300108

- You, F. M., Xiao, J., Li, P., Yao, Z., Jia, G., He, L., et al. (2018). Chromosome-scale pseudomolecules refined by optical, physical and genetic maps in flax. *Plant J.* 95, 371–384. doi: 10.1111/tpj.13944
- Yu, D., Qanmber, G., Lu, L., Wang, L., Li, J., Yang, Z., et al. (2018). Genome-wide analysis of cotton GH3 subfamily II reveals functional divergence in fiber development, hormone response and plant architecture. *BMC Plant Biol.* 18:350. doi: 10.1186/s12870-018-1545-5
- Zhang, N., and Deyholos, M. K. (2016). RNA-Seq analysis of the shoot apex of flax (*Linum usitatissimum*) to identify phloem fiber specification genes. *Front. Plant Sci.* 7:950. doi: 10.3389/fpls.2016.00950

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

Copyright © 2020 Galinousky, Mokshina, Padvitski, Ageeva, Bogdan, Kilchevsky and Gorshkova. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.