Identification of fasciclin-like arabinogalactan proteins in textile hemp (Cannabis sativa L.): in silico analyses and gene expression patterns in different tissues

Gea Guerriero1*, Lauralie Mangeot-Peter1,2, Sylvain Legay1, Marc Behr1,3, Stanley Lutts3, Khawar Sohail Siddiqui4 and Jean-Francois Hausman1

Abstract

Background: The fasciclin-like arabinogalactan proteins (FLAs) belong to the arabinogalactan protein (AGP) superfamily and are known to play different physiological roles in plants. This class of proteins was shown to participate in plant growth, development, defense against abiotic stresses and, notably, cell wall biosynthesis. Although some studies are available on the characterization of FLA genes from different species, both woody and herbaceous, no detailed information is available on the FLA family of textile hemp (Cannabis sativa L.), an economically important fibre crop.

Results: By searching the Cannabis genome and EST databases, 23 CsaFLAs have been here identified which are divided into four phylogenetic groups. A real-time qPCR analysis performed on stem tissues (isolated bast fibres and shivs sampled at three heights), hypocotyls (6-9-12-15-17-20 days-old), whole seedlings, roots, leaves and female/male flowers of the monoecious fibre variety Santhica 27, indicates that the identified FLA genes are differentially expressed. Interestingly, some hemp FLAs are expressed during early phases of fibre growth (elongation), while others are more expressed in the middle and base of the stem and thus potentially involved in secondary cell wall formation (fibre thickening). The bioinformatic analysis of the promoter regions shows that the FLAs upregulated in the younger regions of the stem share a conserved motif related to flowering control and regulation of photoperiod perception. The promoters of the FLA genes expressed at higher levels in the older stem regions, instead, share a motif putatively recognized by MYB3, a transcriptional repressor belonging to the MYB family subgroup S4.

Conclusions: These results point to the existence of a transcriptional network fine-tuning the expression of FLA genes in the older and younger regions of the stem, as well as in the bast fibres/shivs of textile hemp. In summary, our study paves the way for future analyses on the biological functions of FLAs in an industrially relevant fibre crop.

Keywords: Fasciclin-like arabinogalactan proteins, Bast fibres, Cell wall, RT-qPCR, Cannabis sativa

* Correspondence: gea.guerriero@list.lu
1Environmental Research and Innovation (ERIN) Department, Luxembourg Institute of Science and Technology (LIST), 5, Avenue des Hauts-Fourneaux, L-4362 Esch/Alzette, Luxembourg
Full list of author information is available at the end of the article

© The Author(s). 2017 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Background

Arabinoxyproline-rich glycoproteins (AGPs) are cell surface glycoproteins belonging to the hydroxyproline-rich glycoprotein superfamily ([1] and references therein) which are involved in many aspects of plant development, i.e. pattern formation, phytohormone interaction, tissue differentiation, reproduction, response to (a)biotic stresses, cell expansion and secondary cell wall deposition [2, 3]. These heavily glycosylated proteins are subdivided into four main classes: classical AGPs, AG peptides, Lys-rich AGPs, fasciclin-like AGPs (FLAs) [3–6].

FLAs are characterized by the occurrence of one or two AGP domains, as well as one or two fasciclin (FAS) domains [7]. FAS domains were first identified in the fruit fly Drosophila melanogaster and later found in many other organisms, from bacteria to higher plants to animals [5]. Although a consensus sequence for the FAS domains is lacking, two regions are highly conserved, named H1 and H2 (of ca. 10 amino acids) [5]. Additionally, most FLAs show an N-terminal signal peptide and a C-terminal glycosylphosphatidylinositol (GPI) membrane anchor [5, 7], mediating attachment to the cell surface.

FLAs constitute multigene families in plants: for example, 21 FLAs have been identified in thale cress, 24 in rice, 35 in poplar, 34 in wheat, 19 in cotton, 33 in Chinese cabbage and 18 in eucalypt [5, 7–11]. Molecular studies focused on FLAs are important, since they increase the understanding of the molecular functions of this protein family: the available literature on the topic has shown that FLAs in plants are not only related to tissue-specific functions, but also involved in generalized responses to environmental constraints, both biotic and abiotic [3, 7, 11, 12].

Additionally, a strong body of evidence in the literature has highlighted the importance of FLAs in regulating aspects linked to cell wall biosynthesis and, more generally, to stem mechanics in both herbaceous and woody species, as well as fibre growth. For instance, in Arabidopsis, insertion mutants of Atfla11 and Atfla12 and Atfla11/fla12 double mutants show modified stem mechanics, due to a decrease in cellulose, arabinose and galactose in secondary cell walls [12]. Likewise, in Eucalyptus, FLAs belonging to the subgroup A [5, 12] are involved in stem mechanics [11]: in particular EgrFLA2 is linked to cellulose microfibril angle. In poplar, antisense expression of PtFLA6 alters secondary cell wall composition in the xylem, by affecting the biosynthesis of lignin and cellulose [13]. In cotton, GhiFLA1 is involved in fibre initiation and elongation: its overexpression increases fibre length, while its silencing results in shorter fibres with an altered primary cell wall composition [14]. In the fibre crop flax, some FLAs were shown to be upregulated at the snap point, a physical region marking the transition from elongation to cell wall thickening, hence confirming the potential function of these genes in the regulation of fibre development [15, 16].

Textile hemp (Cannabis sativa L.) is an economically important bast fibre-producing crop, with several applications in industry, namely the biocomposite, textile, construction sector [17]. This plant is not only important as a multi-purpose crop, but also useful for fundamental studies centered on cell wall biosynthesis/remodeling [18], because its stem tissues show strong differences in cell types and cell wall composition [19, 20]. The core of hemp stems (a.k.a hurd/shiv) is indeed woody, while the cortex harbors long gelatinous fibres, the bast fibres, with a high content in crystalline cellulose and poor in lignin [21]. The different stem heights correspond to distinctive stages of bast fibre development (from intrusive growth to thickening; Fig. 1a). It is hence possible to study the mechanisms involved in the development of cellulose and woody fibres by separating the stem tissues of the same plant. The cortex can be peeled from the hursds and the bast fibres can be separated from the surrounding parenchymatic cells with the use of 80% ethanol, a mortar and a pestle [20, 22, 23].

The molecular steps involved in the regulation of bast fibre initiation, development and intrusive growth comprise many still unexplored aspects [24–26]; hence an increased knowledge in these mechanisms would favor the development of biotechnological tools focused on bast fibre improvement.

In the light of the above-mentioned relationships between FLAs and cell wall-related processes and considering the industrial applications of C. sativa, we here sought to identify and study the expression patterns of hemp FLA genes in the different stem tissues, as well as in other organs. By using bioinformatics coupled to RT-qPCR, we show that some FLA genes are highly expressed in bast fibres. Moreover, we identify groups of FLAs, upregulated either at the top or the bottom of the stem, which share putative conserved elements in their promoters. Our study therefore lays the foundation to further molecular analyses on a unique family of proteins in an important herbaceous crop.

Methods

Plant material and growth conditions

A hemp monoecious fibre variety (C. sativa cv. Santhica 27) was studied in this work. Plants were grown and sampled as described in [20]. Briefly, after six weeks of growth in controlled chambers, samples were taken along three stem regions localized at different heights with respect to the “snap point” (e.g. an empirically-defined reference region marking the transition from elongation to secondary cell wall thickening; [27]). The “TOP” segment internode corresponds to the region
right below the apex (above the snap point), the “MID” (middle) segment is the internode containing the snap point and the “BOT” (bottom) segment is located two internodes below the “MID” sample (for clarity, a cartoon depicting the sampling strategy is shown in Fig. 1b). A segment of 2.5 cm was collected in the middle of each internode to avoid too much variation in gene expression, due to the varying developmental stages of the cell types.

Fibres were separated from the shivs by peeling the cortical tissues and by quickly processing them as described in [23]. The shivs were directly plunged in liquid nitrogen and stored at −80 °C. The number of independent biological replicates is four, with the exception of the BOT core tissues, for which the biological replicates are three. A total of 13 plants were pooled for each replicate.

Two leaves (sampled below the TOP region from 4 biological replicates, each composed of a pool of 16 plants) were frozen in liquid nitrogen after removal of the midrib with a scalpel and subsequently stored at −80 °C. The number of independent biological replicates is four, with the exception of the BOT core tissues, for which the biological replicates are three. A total of 13 plants were pooled for each replicate.

The hypocotyls, aged from 6 to 20 days after sowing, were grown and sampled as described in [18]. Three biological replicates, each consisting of a pool of 20 hypocotyls, were used.

**Identification of CsaFLA genes using bioinformatics**

In order to identify the FLA genes in *C. sativa* (hereafter referred to as CsaFLAs for the genes and CsaFLAs for the corresponding proteins), different databases were searched: the Medicinal Plant Genomics Resource (http://medicinalplantgenomics.msu.edu/mpgr_external_blast.shtml) and the *Cannabis sativa* Genome Browser Gateway (http://genome.ccbio.utoronto.ca/cgi-bin/hgBlat?command=start&org=C.sativa&db=finola1&hgsid=93256). CsaFLAs were identified by using orthologous FLA protein sequences of *Arabidopsis thaliana* [5] and *Populus trichocarpa* [7]. These sequences were used to perform a BLAT analysis against the hemp Finola and Purple Kush.
database (*Cannabis* Genome Browser Gateway; [28]) and a BLASTP in the MPGR database. Several incomplete sequences were retrieved when using the MPGR database; however it was possible to deduce their full length sequences either by querying the *Cannabis* Genome Browser Gateway, or the EST database at NCBI (dbEST; available at http://www.ncbi.nlm.nih.gov/dbEST/).

**In silico** and phylogenetic analyses of *CsaFLA* protein sequences
Putative FAS domains were identified with the Motif Scan algorithm (http://myhits.isb-sib.ch/cgi-bin/motif_scan), N-terminal signal peptides were identified with SignalP (http://www.cbs.dtu.dk/services/SignalP/) and SignalBlast (http://sigpep.services.came.sbg.ac.at/signalblast.html); the subcellular localization was predicted with TargetP (http://www.cbs.dtu.dk/services/TargetP/).

The big-PI Plant Predictor program ([29]; available at http://mendel.imp.ac.at/gpi/plant_server.html) was used to identify the glycosylphosphatidylinositol (GPI) anchor. The 3D homology models of the hemp FLA 10 and FLA 11 were generated with iTASSER Suite ([30] using 4ut1 and 1o70 as targets respectively; available at http://zhanglab.ccmb.med.umich.edu/I-TASSER/) employing LOMETS, SPICKER and TM-align. The models were then refined using REMO by optimizing the backbone hydrogen-bonding networks and FG-MD by removing the steric clashes and improving the torsion angles. The H1 and H2 conserved regions, motifs and residues implicated in adhesion in both proteins were manually annotated according to Johnson et al. [5]. The final structures showing various domains, conserved regions, motifs and residues involved in adhesion were visualized with Swiss PDB Viewer v4.1 [31]. Conserved motifs in the *CsaFLA* promoter sequences (retrieved at the *Cannabis sativa* Genome Browser Gateway) were identified using the MEME Suite 4.11.2 ([32]; available at http://meme-suite.org/doc/cite.html?man_type=web). The identified motifs were subsequently analyzed with Tomtom ([33]; available at http://meme-suite.org/tools/tomtom) for a comparison against the available motifs in the JASPAR CORE plant database 2016 [34]. For the phylogenetic analysis, full-length sequences were aligned with Clustal-Omega (http://www.ebi.ac.uk/Tools/msa/clustalo) and the generated alignment submitted to PHYML (http://www.phylogeny.fr) to obtain a maximum likelihood phylogenetic tree. The Maximum Likelihood tree was constructed using an aLRT (approximate likelihood ratio test) for non-parametric branch support, based on a Shimodaira-Hasegawa-like procedure. The tree was visualized with iTOL-Interactive Tree Of Life (http://itol.embl.de/). Intron-exon junctions were visualized with Gene Structure Display Server 2.0 (GSDS, http://gsds.cbi.pku.edu.cn/) [35].

**Immunohistochemistry**
Immunohistochemical analyses were performed on resin-embedded tissue sections, as previously described [18]. The LM14 antibody (PlantProbes) was diluted 1:10 in milk protein (MP)/PBS (5% w/v). Sections were incubated for 1.5 h, rinsed three times in PBS and subsequently incubated for 1.5 h with the anti-rat IgG coupled to FITC (Sigma) diluted 100-fold in MP/PBS.

**RNA extraction and RT-qPCR**
Total RNA was extracted using a modified CTAB extraction protocol combined with an RNeasy Plant Mini Kit (Qiagen) according to [23]. The RNA concentration and quality were measured by using a Nanodrop ND-1000 (Thermo Scientific) and a 2100 Bioanalyzer (Agilent), respectively. One microgram of RNA was retrotranscribed into cDNA using the ProtoScript II RTase (NEB) and random primers, according to the manufacturer’s instructions.

The cDNA was diluted to 2 ng/μL and 2 μl used for the RT-qPCR analysis in 384-wells microplates. An automated liquid handling robot (epMotion 5073) was used to prepare the 384-wells microplates (10 μl final volume). A tissue maximization design was used to prepare the microplates [36]. The expression of each *CsaFLA* was normalized using 5 reference genes (tubulin, CDPK, RAN, clathrin and F-box, which geNORMPLUS identified as sufficient for appropriate data normalization) for the stem tissues, as described in [20], and 3 (RAN, TIP41 and F-box) for the other tissues (leaves, seedlings, flowers and roots). For statistical analysis, the normalized relative quantities exported from qBasePLUS were log2 transformed. A one-way ANOVA was carried out using IBM SPSS Statistics v19. A Tukey’s HSD was performed as post-hoc test. The normal distribution of the data was verified with a Kolmogorov–Smirnov test.

**Primer design**
Primers were designed using Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3-plus.cgi/) and verified with the OligoAnalyzer 3.1 tool from Integrated DNA technologies (http://eu.idtdna.com/calc/analyze), Primer efficiencies were checked via qPCR using a serial five-fold dilution of cDNA (25, 5, 1, 0.2, 0.04, 0.008 ng/μL). The primer sequences, amplicon length and Tm, amplification efficiencies and R² are indicated in Additional file 1: Table S1.

**Sequencing of some representative *CsaFLA* promoters**
To determine the homology of the promoter sequences of the variety Santhica 27 with those from the PurpleKush and Finola reference genomes, primers were designed on 3 representative genes (*CsaFLA2–7–16*) using the available sequences at the *Cannabis sativa* Genome Browser to
perform nested PCRs (Additional file 2: Table S2). Genomic DNA was extracted from stem tissues (whole internodes) by using a CTAB-based protocol coupled to the Nucleospin Plant II kit ( Macherey-Nagel). Briefly, 500 μl of extraction buffer (2% CTAB, 2.5% PVP-40, 2 M NaCl, 100 mM Tris-HCl pH 8.0, 25 mM EDTA and 10 μl RNase) were added to 100 mg of finely ground sample and the slurry was vortexed vigorously. After an incubation step at 60 °C for 10 min, 20 μl β-mercaptoethanol buffer were added and the samples were further incubated for 20 min at 60 °C. Subsequently, 500 μl chloroform/isoamyl alcohol 24:1 were added, and the DNA was precipitated for 1 h at 20 °C. After this stage, the Nucleospin II columns were used to bind the DNA and the manufacturer’s instructions were followed to elute genomic DNA.

PCRs were performed using 50 ng DNA and the Q5 Hot Start High-Fidelity 2X Master Mix, following the manufacturer's instructions. The optimal annealing temperatures were computed using the NEB T_m calculator (available at http://tmcalculator.neb.com/#!/).

PCR products were ligated into the pGEM-T Easy vector, following the manufacturer’s instructions and cloned into JM109 chemically competent cells. Three positive clones for each gene promoter were grown o/n at 37 °C in LB medium supplemented with ampicillin 100 μg/ml. Plasmids were extracted using the QIAGEN plasmid miniprep kit and sequenced on an Applied Biosystems 3500 Genetic Analyser using the BigDye Terminator v3.1 Cycle Sequencing and the BigDye XTerminator Purification kits, according to the manufacturer’s instructions.

Results
Identification of putative FLAs in C. sativa: Protein architecture and phylogenetic analysis
BLAST/BLAT analyses of the 21 A. thaliana sequences (AtFLAs) performed against the Medicinal Plant Genomics Resource, the NCBI EST and the Cannabis Genome Browser Gateway databases led to the identification of 23 CsaFLAs (Additional file 3: Table S3). It should be noted that, during the database queries, a contig, i.e. csa_locus_44222_iso_1_len_407_ver_2, which was initially called CsaFLA22 and retrieved at the Medicinal Plant Genomics Resource (MPGR), was also found. However, we believe that this partial gene was erroneously attributed to C. sativa, since we never amplified any product with different primers designed on it and the reported FPKM values at the MPGR are 0 for all the tissues examined. We discarded this gene from our analyses, but kept the original nomenclature given to the hemp FLA genes (i.e. CsaFLA1–24), as at this stage we cannot rule out the existence of this gene in textile hemp.

The intron-exon structure analysis highlighted the presence of 5 genes (CsaFLA2-17-8-12-4) containing one intron: of these genes, CsaFLA12 and 4, both possessing a small intron, group together according to the maximum likelihood (ML) phylogenetic tree, while CsaFLA2-17-8, containing longer introns, are in a different branch (Additional file 4: Figure S1). To check that the putative CsaFLAs belong to the FLA family, the occurrence of the following features was checked: the presence of at least one FAS domain, a signal peptide at the N-terminus, (in some cases) a GPI anchor at the C-terminus and the presence of AGP domains (Additional file 5: Table S4; Additional file 6: Figure S2). The identified FLAs show a PAST% ranging from 19.4 to 44.5% (Table 1), which is in agreement with the values reported for poplar FLAs [7].

In order to investigate the evolutionary relationship between known plant FLA proteins, a maximum-likelihood phylogenetic tree was built using the identified 23 CsaFLAs, 21 AtFLAs, 18 Eucalyptus (EgrFLAs) and 35

| FLA   | Proline% | Alanine% | Serine% | Threonine% | PAST% | Class |
|-------|----------|----------|---------|------------|-------|-------|
| CsaFLA1 | 5.7      | 13.8     | 6       | 7.6        | 33.1  | C     |
| CsaFLA2 | 3.6      | 7.2      | 9.3     | 9.3        | 29.4  | C     |
| CsaFLA3 | 8        | 9.9      | 9.9     | 6.9        | 34.7  | A     |
| CsaFLA4 | 8.2      | 9.8      | 12.8    | 8          | 38.8  | C     |
| CsaFLA5 | 14.2     | 4.7      | 13.9    | 6.4        | 39.2  | B     |
| CsaFLA6 | 7.9      | 8.8      | 9.6     | 7.9        | 34.2  | A     |
| CsaFLA7 | 9.6      | 10.3     | 11.9    | 8.4        | 40.2  | A     |
| CsaFLA8 | 8.3      | 6.6      | 7.5     | 5.8        | 28.2  | B     |
| CsaFLA9 | 8.4      | 7.3      | 10.2    | 10.2       | 36.1  | A     |
| CsaFLA10| 8.9      | 11.2     | 11.8    | 9.4        | 41.3  | C     |
| CsaFLA11| 6.4      | 9.2      | 11.2    | 9.6        | 36.4  | A     |
| CsaFLA12| 8.3      | 8        | 16      | 11.5       | 43.8  | A     |
| CsaFLA13| 12.6     | 8.4      | 12.3    | 10.2       | 43.5  | A     |
| CsaFLA14| 8.7      | 15.5     | 12.9    | 7.4        | 44.5  | C     |
| CsaFLA15| 8.6      | 11.7     | 11.7    | 10.5       | 42.5  | A     |
| CsaFLA16| 8.2      | 10.4     | 14.6    | 8.8        | 42    | A     |
| CsaFLA17| 8.2      | 6.9      | 7.8     | 5          | 27.9  | B     |
| CsaFLA18| 8.8      | 8.2      | 9.8     | 8.5        | 35.3  | A     |
| CsaFLA19| 8.2      | 8.2      | 15.2    | 11.7       | 43.3  | A     |
| CsaFLA20| 6.4      | 7.6      | 10.9    | 5.8        | 30.7  | D     |
| CsaFLA21| 6.5      | 7.3      | 10.1    | 7.6        | 31.5  | D     |
| CsaFLA22| 7.7      | 6.5      | 13.6    | 9.3        | 37.1  | D     |
| CsaFLA24| 2.5      | 3        | 10.4    | 3.5        | 19.4  | D     |

The classes for each target are also indicated.
poplar sequences (PtrFLAs) (Fig. 2 and Additional file 6: Figure S2). It should be noted that we chose to perform the phylogenetic analysis using full-length CsaFLA sequences to conform to the previously published tree of poplar FLAs [7]. Different results may be obtained if the mature protein sequences are used.

CsaFLAs cluster into four major classes (A-D). Class A is the largest clade, with 11 hemp members (CsaFLA3/6/7/9/11/12/13/15/16/18/19) containing a single FAS domain flanked by two AGP domain and a GPI anchor at the C-terminus. Class B includes CsaFLA5/8/17 which contains two FAS domains and a single AGP region. Class C comprises 5 members (CsaFLA1/2/4/10/14), characterized by two FAS, two AGP domains and a GPI anchor. The last class includes 4 proteins (CsaFLA20/21/23/24) with no distinctive protein architecture (although CsaFLA20 and CsaFLA23 have 2 FAS domains like group B, however their lengths are smaller). The percentage of CsaFLA identity with the putative orthologs from thale cress is between 32 and 76% (Table 2).

**CsaFLA expression patterns in hemp tissues**

An immunohistochemical analysis carried out with the LM14 antibody (recognizing AGPs) revealed that the epitope is distributed in different tissues of the hemp stem (Additional file 7: Figure S3): this result shows the broad distribution of these proteins in the different hemp stem tissues. In particular, in the bottom internode, AGPs are present in the core tissues (cell walls of fibres and vessels), cortical parenchyma/collenchyma and, notably, in the inner region of the fibres, i.e. the layer (plasma membrane) delimiting the fibre cell lumen.

Of the 23 CsaFLAs identified, 22 were expressed in the stem tissues (Fig. 3; Additional file 8: Figure S4). CsaFLA14 was detected at very low levels in the stem tissues (Ct > 32).

In hemp bast fibres, the heat-map hierarchical clustering shows 5 major expression trends (Fig. 3). These are the following: 1) a group of genes (CsaFLA2–6-24) is upregulated at the middle internode containing the snap point (in the core the expression decreases towards the base of the stem); 2) CsaFLA1-4-7-8-10-20-23 are expressed at higher levels in the top and decreased towards the bottom internode; 3) two FLAs, CsaFLA5 and 21, are downregulated at the snap point; 4) three genes, CsaFLA9-11-17, show a tendency to upregulation at the snap point, although the pattern is less marked with respect to group I (and in the core the expression increased towards the stem base); 5) the last group comprises FLAs upregulated at the bottom (CsaFLA3-12-13-15-16-18-19).

Gene expression analysis was carried out on other hemp tissues, i.e. leaves, roots, male/female flowers and
seedlings, to check whether hemp FLAs expressed at low levels in the stem showed a distinctive expression pattern in other hemp tissues (Fig. 4; Additional File 9: Figure S5). The heat-map hierarchical clustering reveals the presence of 4 main expression patterns (Fig. 4). More specifically, a group of genes, comprising CsaFLA1-4-5-6-8-10-17, is expressed at higher levels in hemp leaves; in group II, two genes specifically upregulated in male flowers are present (CsaFLA14 and 23); group III comprises 9 FLAs expressed at higher levels in hemp roots (CsaFLA3-7-11-12-13-15-16-18-19); the fourth group is represented by CsaFLA2-9-20-21-24, which are overall more expressed in female flowers.

The expression of some FLAs belonging to the aforementioned 5 stem groups has also been investigated in hypocotyls, aged 6-9-12-15-17-20 (H6 to H20). The hemp hypocotyl was proven to be a suitable model to study cell wall-related processes accompanying secondary growth [18], therefore the goal was to verify whether their expression pattern highlighted the same trend observed in adult stems. CsaFLA1-2-8-21 were more...
expressed in young hypocotyls (H6); CsaFLA3-9-11-13 were more expressed in H15, H17 and H20 (Fig. 5).

Identification of conserved motifs in the promoters of some CsaFLAs
A bioinformatic analysis was carried out on the available promoter regions (between 116 and 1064 bp retrieved at the Cannabis Genome Browser Gateway) of the CsaFLA genes showing a distinctive expression pattern in the bast fibres. The promoters of the genes within group I, II and V were selected (Additional File 10), in the light of their upregulation at the snap point, at the younger and older stem regions respectively (Fig. 3). While no conserved motifs could be found with the MEME suite tool for the genes in group I, 1 motif was found for the genes within group II and V (Table 2). The search carried out with the conserved sequence of group II and V FLAs in the JASPAR CORE 2016 plants database identified SOC1 (lowest p-value among the matches retrieved) and MYB3 as candidates recognizing similar motifs (Table 3). In addition to that, the promoters of three representative CsaFLAs showing upregulation at the snap point, at the top, or at the bottom (CsaFLA2, CsaFLA7 and CsaFLA16) were amplified and cloned. The purpose was to verify the sequence conservation between the fibre variety Santhica 27, the Purple Kush strain and Finola variety [28]. The motifs identified by MEME were cross-verified with PlantPAN 2.0 (http://plantpan2.itps.ncku.edu.tw/) and are highlighted in yellow in the promoters of CsaFLA7 and CsaFLA16 from Santhica 27 (Additional File 11).

Domains, conserved regions, motifs and residues mediating adhesion in CsaFLAs from class A and C.
Homology models of one representative each from Class C (FLA 10) and Class A (FLA 11) showing FAS, AGP and GPI anchor domains, H1 and H2 conserved regions, putative amino acids involved in adhesion and [YFL]H motif are given in Fig. 6 and Table 4. The conserved sequences for FLA10 for the first FAS domain are LTVLVLVLSNGA and LSVFEVDQVL for H1 and H2 regions respectively and the [YF]H motif is AL-X-LH-VV (Fig. 6a; Table 4). The conserved sequences for FLA10 for the second FAS domain are LTLFAPNDEA and LVIFTVDNVL for H1 and H2 regions respectively and the [YF]H motif is VL-X-YH-SL (Fig. 6a; Table 4). The conserved sequences for FLA11 for a single FAS domain are ITVFAPTDSA and LSIFAPEVL for H1 and H2 regions respectively and the [YF]H motif is LV-X-YH-VL (Fig. 6b; Table 4).

The two hydrophobic residues preceding and after [YF]H motif are thought to be implicated in mediating adhesion of these proteins. It is interesting to note that these residues are generally aliphatic amino acids such as I, L and V [5], however in FLA10 an Ala and polar Ser are found in the first and second FAS domains respectively (Fig. 6a).

It is noteworthy that [YF]H residues are either completely or partially buried, however both residues that

---

**Fig. 5** Expression analysis of 8 CsaFLAs in the hypocotyls of C. sativa. Error bars indicate the standard error of the mean (n = 3). Different letters indicate statistically significant values at the one-way ANOVA test (p < 0.05) with a Tukey post-hoc test. * indicates a non-normal distribution (Kolmogorov-Smirnov test). H6-H9-H12-H15-H17-H20 indicate hypocotyls aged 6–9–12–15–17–20 days.
flank [YF]H motif on the N-terminal side (AL, VL, LV) are completely solvent-exposed in both FLA10 and FLA11 (Table 4). In contrast, aliphatic residues (VV) on the C-terminal side of [YF]H motif in H1 domain of FLA10 are either completely or partially buried, whereas polar Ser is partially buried, but Leu is exposed in the H2 domain (Fig. 6a, space-filled). For FLA11 both VL residues on the C-terminal side of [YF]H motif are solvent exposed (Fig. 6b, space-filled). In general, residues belonging to FLA11 and those located towards the N-terminal side are exposed more favorably to mediate adhesion than FLA10 and those located on the C-terminal side. This suggests that adhesion for both these proteins may be mediated via hydrophobic interactions.

Discussion
The FLAs identified in C. sativa group into the previously described four phylogenetic classes (Fig. 2) [5]. A nomenclature of CsaFLAs is hereby also proposed which follows the Arabidopsis classification (i.e. when the phylogenetic tree highlighted clustering of a CsaFLA proteins with a specific AtFLA, the same number was assigned to C. sativa).

Within class A, the largest, it is possible to observe a separate clade represented by CsaFLA3-12-13-15-16-18-19 which is highly expressed at the snap point and in the older stem regions, both in the bast fibres and the shivs (Fig. 3). A subset of class A genes (CsaFLA3-9-11-13) was more expressed in the old hypocotyls (peaking at H17 with high values at H15 and H20). As previously shown [18], the hypocotyl undergoes secondary growth in H9 and later time-points. The phylogenetic position of this cluster of FLAs, together with their common expression pattern, might indicate a specific role in secondary growth. This group of genes may indeed represent hemp-specific single FAS domain FLAs specialized in secondary growth, in a manner analogous to what was previously shown in eucalypt and thale cress [11, 12]. Hemp is unfortunately recalcitrant to transformation, therefore homologous testing, as previously performed on e.g. eucalypt FLAs [11], is cumbersome. However heterologous testing in a more amenable system, e.g. Nicotiana tabacum, can confirm or refute the hypothesis.

It is here worth discussing also the phylogenetic position of CsaFLA11 in a clade grouping AtFLA11, EgrFLA2b and EgrFLA3b (Fig. 2). These genes were shown to affect stem mechanics, as well as cell wall architecture [11, 12]. The AtFLA11 transcript was detected in the xylem and interfascicular fibres in inflorescence stem, preceding the lignification of those two tissues [37]; CsaFLA11 also shows a gradual increase in expression towards the older regions of the stem and it is slightly more expressed in the older hypocotyl too (Fig. 5). This FLA represents another interesting candidate putatively involved in cell wall-related processes in textile hemp.

Within class C, CsaFLA4 and CsaFLA1 group together with the characterized orthologs from thale cress. AtFLA4 (SOS5) is involved in cell expansion [38] and AtFLA1 was shown to regulate root and shoot development in tissue culture [39]. CsaFLA8 was more expressed in the TOP region of the stem, as well as in H6, suggesting a role in elongating tissue. However, it remains to be shown whether the hemp genes are involved in the same regulatory networks as in Arabidopsis.

The expression of the 23 CsaFLAs was first investigated in the different tissues of the stem, because we wanted to identify those genes specifically associated

| FLAs | Conserved motif (MEME) | Similar motifs (Tomtom/JASPAR CORE 2016 plants) | Function | p-value |
|------|------------------------|-----------------------------------------------|----------|----------|
| Group II | MA0554.1 (SOC1) | | Transcription activator controlling flowering time. Probably also involved in photoperiod perception. | 5.39e-05 |
| Group V | MA1038.1 (MYB3) | | Repression of the phenylpropanoid biosynthesis-related genes. Response to salt stress, wounding, ABA, SA. | 1.59e-02 |

The first motif in each table cell is the one found in the database, the second motif is the one found in the promoters of hemp FLAs. The function and computed p-value are indicated.
with a tissue-type and a stem region. Among them, we would like to draw the reader’s attention on the first group of genes, represented by CsaFLA2-6-24, because they show a different expression profile in the bast fibres and the shivs. The expression in the shivs shows a decrease from the top to the bottom of the stem, while in the bast fibres their expression peaks at the snap point. This is quite interesting if we consider that the snap point is the region marking a shift in the stem mechanical properties, as it determines the transition from cell elongation to thickening [27]. It was shown that the young stem regions of hemp at the vegetative stage of growth are characterized by the presence of ca. 66% glucose, while older regions have about 82%: this result confirms that during their transition from elongation to thickening, bast fibres require great amounts of glucose for the synthesis of cellulose [40]. The 3 FLAs may therefore be involved in cell wall-related processes occurring during this transition. Additionally, this is in agreement with the flax microarray data showing upregulation of certain FLAs around the snap point [15] and with the increased expression of poplar FLAs in tension wood, which, like bast fibres, is composed of a cellulosic G-layer [41, 42]. As previously discussed for

Table 4  Solvent accessibility of residues within conserved motifs

| FLA  | Conserved motif | Solvent accessibility |
|------|-----------------|-----------------------|
| CsaFLA10 | (AL)-X-LH-(VL) | (9,8)-2,4-(4,0) |
| CsaFLA11 | (LV)-X-YH-(VL) | (9,9)-1,6-(6,8) |

Residues within brackets are thought to be involved in adhesion. The solvent accessibility of X residue is not given. Solvent accessibility is determined by iTASSER suite and the range is from 0 (completely buried) to 9 (completely solvent exposed)
involvement of the G-layer formation in the stems, specific chitinases are highly expressed in bast fibres and may regulate G-layer formation in these cell types [22]. Therefore, it is reasonable to assume that the concerted action of specific FLAs and chitinases may be involved in the transition from elongation to G-layer formation in hemp.

In group II and V are FLAs which, in the bast fibres, show a gradual decrease from the apical to the basal part of the stem and an increase in expression, respectively. A similar trend was observed in the hypocotyls: CsaFLA8 (belonging to the stem group II) was more expressed in H6; CsaFLA13 (belonging to the stem group V) was more expressed in H15, H17 and H20 (Fig. 5). In addition, the hypocotyl expression pattern of CsaFLA3 was similar to the one of CsaFLA13. Our study therefore identified specific FLAs likely involved in bast fibre elongation during intrusive growth (CsaFLA1-4-7-8-10-20-23) and others involved in secondary cell wall deposition during the thickening stage (CsaFLA3-12-13-15-16-18-19).

The expression of hemp FLAs was also investigated in other tissues, notably leaves, roots, male/female flowers and in seedlings (Fig. 4).

The genes belonging to group III in stem tissues (Fig. 3) are highly expressed in roots: within this cluster of FLAs are the orthologs of AtFLA7 and AtFLA11 (Fig. 2) for which a higher number of ESTs was retrieved in the roots of thale cress [5].

In reproductive organs, the RT-qPCR results show that some genes are highly expressed in male and female flowers. This suggests that some FLAs are involved in hemp inflorescence formation.

In seedlings, two genes showed higher expression, i.e. CsaFLA2 and 24 (Fig. 4). Notably, CsaFLA2 is the ortholog of AtFLA2 (Fig. 1), whose EST number is higher in Arabidopsis flower buds [5].

In order to investigate whether specific regulatory elements occurred in the promoters of the genes showing specific expression patterns in the stem tissues, we analyzed the genes from group I-II and V (Fig. 3). While for group I no conserved motifs could be obtained, 2 conserved sequences were found for group II and V (Table 2). A conserved motif recognized by the MADS box transcription factor SOC1 could be identified in the promoters of the genes upregulated in the apical stem regions: this finding suggests that they may be involved in a developmental program regulating the transition from vegetative to reproductive growth and/or the response to hormonal regulation (e.g. via gibberellin). In this respect it is noteworthy that in A. thaliana SOC1 was shown to control the annual growth habit [43]: soc1 ful mutants show indeed woody growth reminiscent of the perennial lifestyle. Hence the FLAs upregulated at the top of the stems might belong to a regulatory circuit controlling elongation and suppressing secondary growth.

The genes in group V show the presence of a conserved motif putatively recognized by MYB3, which is an R2R3 MYB transcriptional repressor belonging to subgroup S4 together with the characterized AtMYB4 [44]. MYB4 negatively regulates phenylpropanoid biosynthesis (more specifically, in thale cress it is a negative regulator of hydroxycinnamic acid metabolism and it exerts its silencing function by displacing the activators binding to the MYB motifs present in many promoters of genes involved in the phenylpropanoid metabolism; [44]). It is therefore possible that the identified element is involved in the coordination of phenylpropanoid biosynthesis in bast fibres and might regulate the hypolignification observed in these cells [45, 46]. In our recently-published transcriptomic dataset [47], we observed an upregulation of the SOC1 gene at the top (4-fold induction with respect to the bottom and 1.3-fold induction with respect to the middle) and MYB4 at the bottom (1.7-fold induction with respect to the top and 4.6-fold induction with respect to the middle). This result therefore strengthens the existence of a putative regulatory circuit (controlling, among other genes, the expression of CsaFLAs) at the top and bottom of adult hemp plants.

Conclusions

In conclusion, our work has identified (at least) 23 genes coding for FLAs in textile hemp, some of which specific to distinct stages of bast fibre development. Bioinformatics has highlighted the occurrence of conserved motifs in the promoters of genes upregulated either at the top or at the bottom of the stem. This finding points to the existence of a fine regulatory network controlling bast fibre elongation and cell wall composition. Future functional analyses carried out on heterologous systems will shed more light on the functions of the identified genes.

Additional files

Additional file 1: Table S1. List of primers used to amplify CsaFLAs in the study. The details concerning the primer sequences, amplicon length and Tm, PCR efficiency and regression coefficient are given. (DOCX 16 kb)
Acknowledgements

The authors wish to thank Aude Corvisy and Laurent Solinich for their technical support.

Funding

The Fonds National de la Recherche, Luxembourg, (Project CANCNC C13/SR/5774202), is gratefully acknowledged for financial support. The funding agency had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, and in the decision to publish the results.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Authors’ contributions

GG conceived and designed the experiments; LM-P performed the experiments; KSS performed the RT-qPCR and bioinformatics analyses; MB contributed to the RT-qPCR and bioinformatics analyses; all authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

In March 2017, GG has filed a patent (“Genetically engineering of plant fibres and plant thereof”) describing the promoters of the FLA genes in hemp, which might potentially pose a competing interest. The patent is owned by the Luxembourg Institute of Science and Technology that has no affiliation to any commercial entity. The protection procedures of the associated intellectual property do not alter the adherence to BMC Genomics policies on sharing data and materials.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

1. Environmental Research and Innovation (ERIN) Department, Luxembourg Institute of Science and Technology (LIST), 5, Avenue des Hauts-Fourneaux, L-4362 Esch/Alzette, Luxembourg. 2. Present address: Institut National de la Recherche Agronomique, Université de Lorraine, UMR 1136, Interactions Arbres-Microorganismes, Champenoux, France. 3. Groupe de Recherche en Physiologie Végétale, Earth and Life Institute-Agronomy, Université catholique de Louvain, Louvain-la-Neuve, Belgium. 4. Life Science Department, King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia.

References

1. Tan L, Shouvalter A, Egelund J, Hernandez-Sanchez A, Dobbin M, Bacic A. Arabinogalactan proteins and the research challenges for these enigmatic plant cell surface proteoglycans. Front Plant Sci. 2012;3:140.
2. Seifert GJ, Roberts K. The Biology of Arabinogalactan Proteins. Ann Rev. Plant Biol. 2007;58:137–61.
3. Pereira AM, Pereira LG, Coimbra S. Arabinogalactan proteins: rising attention and novel insights. Genet Genomics. 2007;278:149–58.
4. Schultz CJ, Rumsewicz MP, Johnson KL, Jones BJ, Gaspar YM, Bacic A. Using Genomic Resources to Guide Research Directions. The Arabinogalactan Protein Gene Family as a Test Case. Plant Physiol. 2002;129:1448–63.
5. Johnson KL, Jones BJ, Bacic A, Schultz CJ. The Fasciclin-Like Arabinogalactan Proteins of Arabidopsis. A Multigene Family of Putative Cell Adhesion Molecules. Plant Physiol. 2003;133:1911–25.
6. Ellis M, Egelund J, Schultz CJ, Bacic A. Arabinogalactan-Proteins: Key Regulators at the Cell Surface? Plant Physiol. 2010;153:609–19.
7. Zang L, Zheng T, Chu Y, Ding C, Zhang W, Huang Q, et al. Genome-Wide Analysis of the Fasciclin-Like Arabinogalactan Protein Gene Family Reveals Differential Expression Patterns, Localization, and Salt Stress Response in Populus. Front Plant Sci. 2015;6:1140.
8. Faik A, Abouzouhair J, Sarhan F. Putative fasciclin-like arabinogalactan-proteins (FLA) in wheat (Triticum aestivum) and rice (Oryza sativa): identification and bioinformatic analyses. Mol Genet Genomics. 2006;276:478–94.
9. Huang QX, Xu WL, Gong SY, Li B, Wang XL, Xu D, et al. Characterization of 19 novel cotton FLA genes and their expression profiling in fiber development and in response to phytohormones and salt stress. Physiol Plantarum. 2008;134:348–59.
10. Jun L, Xiaoming W. Genome-wide identification, classification and expression analysis of genes encoding putative fasciclin-like arabinogalactan proteins in Chinese cabbage (Brassica rapa L.). Mol Biol Rep. 2012;39(10):541–55.
11. MacMillan CP, Taylor L, Bi Y, Southerton SG, Evans R, Spokevicius A. The fasciclin-like arabinogalactan protein family of Eucalyptus grandis contains members that impact wood biology and biomechanics. New Phytol. 2015;206:1314–27.
12. MacMillan CP, Mansfield SD, Stachunki ZH, Evans R, Southerton SG. Fasciclin-like arabinogalactan proteins: specialization for stem biomechanics and cell wall architecture in Arabidopsis and Eucalyptus. Plant J. 2010;62:689–703.
13. Wang H, Jiang C, Wang C, Yang Y, Yang L, Gao X, et al. Antisense expression of the fasciclin-like arabinogalactan protein FLA6 gene in Populus inhibits expression of its homologous genes and alters stem biomechanics and cell wall composition in transgenic trees. J Exp Bot. 2015;66:3291–302.
14. Huang Q, Gong SY, Xu WL, Li W, Li P, Zhang C, et al. A Fasciclin-Like Arabinogalactan Protein, GhFLA1, is Involved in Fiber Initiation and Elongation of Cotton. Plant Physiol. 2013;161:1278–90.
15. Roach MJ, Deyholos MK. Microarray analysis of flax (Linum usitatissimum L.) stems identifies transcripts enriched in fibre-bearing phloem tissues. Mol Genet Genomics. 2007;278:149–65.
16. Roach MJ, Deyholos MK. Microarray Analysis of Developing Flax Hypocotyls Identifies Novel Transcripts Correlated with Specific Stages of Phloem Fibre Differentiation. Ann Bot. 2008;102:317–30.
17. Guerriero G, Hausman JF, Strauss J, Ertan H, Siddiqui KS. Lignocellulosic biomass: Biosynthesis, degradation, and industrial utilization. Eng Life Sci. 2016;16:1–16.
18. Behr M, Legay S, Ziković E, Motyka V, Dobreův P, Hausman JF, et al. Studying Secondary Growth and Bast Fiber Development: The Hemp Hypocotyl Peeks behind the Wall. Front. Plant Sci. 2016;7:1733.

19. Andre CM, Hausman JF, Guerriero G. Cannabis sativa: The Plant of the Thousand and One Molecules. Front Plant Sci. 2016;7:179.

20. Mangeot-Peter L, Legay S, Hausman JF, Esposito S, Guerriero G. Identification of Reference Genes for RT-qPCR Data Normalization in Cannabis sativa Stem Tissues. Int J Mol Sci. 2016;17:1556.

21. Guerriero G, Sergeant K, Hausman JF. Integrated -Omics: A Powerful Approach to Understanding the Heterogeneous Lignification of Fibre Crops. Int J Mol Sci. 2013;14:10,958–78.

22. Mokshtina N, Goshkova T, Deyholos MK. Chitinase-Like and Cellulose Synthase Gene Expression in Gelatinous-Type Cellulosic Walls of Flax (Linum usitatissimum L.) Bast Fibers. PLOS ONE. 2014;9:e97949.

23. Guerriero G, Mangeot-Peter L, Hausman JF, Legay S. Extraction of High Quality RNA from Cannabis sativa Bast Fibres: A Vademecum for Molecular Biologists. Fibers. 2016;4:23.

24. Lev-Yadun S. Plant fibers: Initiation, growth, model plants, and open questions. Russ. J. Plant Physiol. 2010;57:205–15.

25. Snegireva A, Chernova T, Ageeva M, Lev-Yadun S, Goshkova T. Intrusive growth of primary and secondary phloem fibres in hemp stem determines fibre-bundle formation and structure. AoB Plants. 2015;7:pxv061.

26. Guerriero G, Hausman JF, Cai G. No Stress! Relax! Mechanisms Governing Growth and Shape in Plant Cells. Int J Mol Sci. 2014;15:5094-5114.

27. Goshkova TA, Salnikov W, Chemikosova SB, Ageeva MV, Pavlencheva NV, van Dam JEG. The snap point: a transition point in Linum usitatissimum bast fiber development. Ind Crops Prod. 2003;18:213–21.

28. van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, et al. The Arabidopsis SOS5 Locus – Time PCR Experiments. Clin Chem. 2009;55:611-616.

29. Eisenhaber B, Wildpaner M, Schultz CJ, Borner GH, Dupree P, Eisenhaber F. Glycosylphosphatidylinositol Lipid Anchoring of Plant Proteins: Sensitive Prediction from Sequence- and Genome-Wide Studies for Arabidopsis and Rice. Plant Physiol. 2003;133:1691–701.

30. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The T-ASSER Suite: Protein structure and function prediction. Nat Methods. 2015;12:7–8.

31. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-Pdb Viewer. An environment for comparative protein modeling. Electrophor. 1997;18:2714–23.

32. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Edited by Edited by AAAI press 1994:28-36.

33. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying similarity between motifs. Genome Biol. 2007;8:R24.

34. Mathelier A, Nesvizhskii AI, Arieninas DJ, Chen CY, Denay G, Lee J, et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2016;44:D110–5.

35. Hu S, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2015;31:1296–7.

36. Bustin SA, Benes V, Garson JA, Helttemans J, Hoggett J, Kubista M, et al. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Clin Chem. 2009;55:611–22.

37. Ito S, Suzuki Y, Miyamoto K, Ueda J, Yamaguchi I. AtFLA11, a Fasciclin-Like Arabinogalactan-Protein, Specifically Localized in Sclerenchyma Cells. Biosci. Biotech. Bioch. 2005;69:1963–9.

38. Shi H, Kim Y, Guo Y, Stevenson B, Zhu JK. The Arabidopsis SOS5 Locus Encodes a Putative Cell Surface Adhesion Protein and Is Required for Normal Cell Expansion. Plant Cell. 2003;15:19–32.

39. Johnson KL, Kibble NA1, Bacic A, Schultz CJ, A Fasciclin-Like Arabinogalactan-Protein (FLA) Mutant of Arabidopsis thaliana, fla1, Shows Defects in Shoot Regeneration. PLOS ONE. 2011;6:e25154.

40. Cròtner D, Monties B, Chabbert B. Structure and chemical composition of bast fibers isolated from developing hemp stem. J Agric Food Chem. 2005;53(21):8279–8289.

41. Lafarguette F, Leplé JC, Déjardin A, Laurans F, Costa G, Legeaye-Caumes MC, et al. Polylar genes encoding fasciclin-like arabinogalactan proteins are highly expressed in tension wood. New Physiol. 2004;164:107–21.

42. Grichtch C, Wan Y, Mitchell RAC, Sheevey PR, Harley SJ, Karp A. G-fibre cell wall development in willow stems during tension wood induction. J Exp Bot. 2015;66:447–59.

43. Melzer S, Lenz F, Genni J, Vanneste S, Rohde A, Beeckman T. Flowering-time genes modulate meristem determinacy and growth form in Arabidopsis thaliana. Nat Genet. 2008;40:1489–92.

44. Jin H, Corninelli E, Bailey P, Parr A, Mehtens F, Jones J, et al. Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. EMBO J. 2000;19:6150–61.

45. Day A, Ruel K, Neutelings G, Gronier D, David H, Hawkins S, et al. Lignification in the flax stem: evidence for an unusual lignin in bast fibers. Planta. 2005;222:334–45.

46. Huś R, Moreel K, Fliniaux O, Lucau-Dabila A, Fénart S, Grec S, et al. Natural Hypolignification Is Associated with Extensive Oligolignol Accumulation in Flax Stems. Plant Physiol. 2012;158:1893–915.

47. Guerriero G, Behr M, Legay S, Mangeot-Peter L, Zorzan S, Ghoniem M, et al. Transcriptomic profiling of hemp bast fibres at different developmental stages. Sci Rep. 2017;7:4961.