



# Genome-wide identification of fasciclin-like arabinogalactan proteins in jute and their expression pattern during fiber formation

Md. Sabbir Hossain<sup>1</sup> · Borhan Ahmed<sup>1,2</sup> · Md. Wali Ullah<sup>1</sup> · Nasima Aktar<sup>1</sup> · Md. Samiul Haque<sup>1,2</sup> · Md. Shahidul Islam<sup>1,2</sup>

Received: 12 June 2020 / Accepted: 23 September 2020 / Published online: 4 October 2020  
© Springer Nature B.V. 2020

## Abstract

Fasciclin-like arabinogalactan proteins (FLAs), a class of arabinogalactan proteins (AGPs) are involved in plant growth and development via cell communication and adhesion. FLAs were also associated with fiber and wood formation in plants but no information is available about the roles of FLA proteins during fibre development of jute. Here, we performed molecular characterization, evolutionary relationship and expression profiling of FLAs proteins in jute (*Corchorus olitorius*). In total, nineteen CoFLA genes have been identified in jute genome, which were divided into four classes like FLAs of other species based on protein structure and similarity. All CoFLAs have N-terminal signal peptide and one or two FAS domain while two FLAs lack well defined AGP region and eight FLAs were devoid of C-terminal glycosylphosphatidylinositol (GPI) anchor. Expression analysis of different regions of jute stem suggested their involvement in different fiber development stages. Four genes CoFLA 11, 12, 20, and 23 were highly or predominately expressed in fiber containing bark tissues while the expression levels of six CoFLA genes 02, 03, 04, 06, 14 and 19 were comparatively higher in stick. Higher transcripts levels of CoFLA 12 and 20 in the middle bark tissues suggest their involvement in fiber elongation. In contrast, the CoFLA 11 and 23 were more expressed in bottom bark tissues suggesting their potential involvement in secondary cell wall synthesis. Our study can serve as solid foundation for further functional exploration of FLAs and in future breeding program of jute aiming fiber improvement.

**Keywords** Fasciclin domain · *Corchorus olitorius* · Fiber elongation · Secondary cell wall · RT-qPCR

## Introduction

Plant cell wall, mainly composed of polysaccharides, cross-linking glycans and proteins, plays an important role during plant development processes and adaptation to various environmental conditions through interaction plasma membrane [1, 2]. Proteins such as enzymes, expansins, cell wall-associated kinases and hydroxyproline (Hyp)-rich glycoproteins comprise a small portion of primary cell wall

(less than 10% of dry weight) with potential role in cell wall and plasma membrane interactions [1, 3]. Arabinogalactan proteins (AGP) belonging to Hyp-rich glycoprotein super-family are abundantly present in cell wall and plasma membrane of different plant species, and involved in various biological functions such as plant growth, reproduction, adaptation [4–7]. According to protein structure, AGPs are divided into six major classes including typical AGPs, Lys-rich AGPs, AG peptides, fasciclin-like AGPs (FLA), non-classical AGPs and chimeric AGPs [6, 8, 9] where the FLAs have one or two fasciclin (FAS) domains along with one or two AGP domain [10]. FAS was first identified in fruit fly *Drosophila melanogaster* and later discovered in many species such as bacteria, algae, plants and animals [3, 11–13]. About 110–150 amino acids long FAS domains are characterized by two highly conserved sequences (H1 and H2) and one short conserved motifs ([Y/F]-H) [1–3]. The proportion of Pro, Ala, Ser and Thr (PAST) is less than 35% in FLAs although it contains AGP like glycosylated sequences

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11033-020-05858-w>) contains supplementary material, which is available to authorized users.

✉ Md. Shahidul Islam  
shahidul@jutegenome.org; nshahidul@gmail.com

<sup>1</sup> Basic and Applied Research on Jute Project, Bangladesh Jute Research Institute, Dhaka, Bangladesh

<sup>2</sup> Bangladesh Jute Research Institute, Dhaka, Bangladesh

which are repeat sequence of [S/T/A]-P and [S/T/A]-P-P residues [9], whereas, the classical AGPs have more than 50% PAST [14]. FLAs may associate with cell–cell interactions and cell–matrix adhesion as FAS and AGP domain of FLA mediate protein–protein interactions and protein–carbohydrate interactions, respectively [2]. Moreover, most FLAs also contain an N-terminal signal peptide and glycosylphosphatidylinositol (GPI) membrane anchor, through which proteins bind to cell surface and participate in cell communications [15, 16].

FLAs are encoded by multigene families in plants and identified in several plant to species; 24 in rice, 34 in wheat [1], 33 in Chinese cabbage [2], 21 in Arabidopsis [3], 35 in poplar [10], 19 in cotton [15], 23 in hemp [16] and 18 in eucalypt [17]. However, only few FLAs have been functionally characterized so far. For example, the AtFLA4 mutant has thinner cell wall, abnormal swollen cells in root tip and sensitive to salt [18]. Moreover, AtFLA1 and AtFLA3 have role in lateral root initiation, shoot development, and microspore development in Arabidopsis, respectively [10, 19]. Additionally, FLAs are involved in the regulation of cell wall biosynthesis in some species. For instance, mutant AtFLA11 and AtFLA12 have ability to modify stem biomechanics by decreasing polysaccharides [20]. Likewise, in eucalypt, EgrFLA2 and EgrFLA3 are associated with cellulose microfibril angle and stem strength, respectively [17]. In poplar, PtrFLA6 and 26 are involved in altering xylem cell wall matrix and stem strength by affecting the biosynthesis of cellulose and lignin [10, 21]. FLAs may also be associated with fiber cell development, for example, overexpression of GhFLA1 in cotton promoted fiber elongation, leading to an increase in fiber length while it's down-regulation produces opposite effects which indicate their linkage in fiber initiation and elongation [22]. Furthermore, previous study in hemp revealed the association of some CsaFLAs including CsaFLA 2, 6 and 24 in bast fiber cell initiation and elongation whereas CsaFL A3, 12, 13, 16 and 19 participates in secondary cell wall biosynthesis [16]. Some FLAs from flax, another fiber crop, have potential functions in the regulation of phloem fiber formation from initiation to cell wall thickening [23–25], suggesting that their involvement in cell wall biosynthesis specially phloem fiber development, although the underlying molecular mechanisms are still unknown.

Jute (*Corchorus* spp.) is economically and eco-friendly important ligno-cellulosic bast fibre crop and in terms of production it is the second largest natural fiber crop next to cotton [26]. It is not only used for manufacturing hessian bags, rope, packaging, carpet, paper, pulp, automotive headlining and geo-textile but also valued for bio-fuel production such as bio-ethanol and yarn [27, 28]. Jute fiber is now getting considerable attention due to its annually renewable and biodegradable nature, lower cost in production and broad-spectrum applications [27]. However, this natural fiber has limitations

due to the short ultimate fiber length (0.8–6.0 mm) which affects spinning processes. Furthermore, the lignin content of jute fiber is higher (11 to 14%) than any other fiber crops [29, 30], causing fiber stiffness and spine problems. Therefore, it is necessity to know the molecular mechanisms underlying fiber cell development for increasing the length and reducing lignin content of fiber. In recent years, some studies have been conducted to investigate regulation of jute fiber development using transcriptome data [31–33]. However, we have still little information about the roles of different genes during fiber formation. Our research group released the genome of *Corchorus olitorius* [30], providing the opportunity to identify and explore the role of genes associated with fiber formation and development.

Jute is not only an economically important crop but it can also be used as model plant for the study of phloem fiber development, where the woody stick tissues (the core of jute stem) are completely different in cell types and composition from the bark tissues containing phloem fiber. Furthermore, different stem heights possess three distinctive stages of fiber cell formation including fiber initiation, cell elongation and secondary cell wall deposition [34]. Jute fiber develops in phloem, arranged in trapezoid wedges (pyramid) of sclerenchymatous fiber cell bundles and alternate with medullary ray of soft tissue which are individually  $\leq 3$  mm in length. Fiber cells are originated from primary and secondary phloem [35] and the cell elongation starts by a combination of coordinated growth with the surrounding cells and intrusive growth between cells in the stem [36]. After the fibers reach their final length, they begin to start secondary cell wall thickening [37–39]. After secondary wall formation is complete, the protoplast becomes progressively dying and an internal lumen develops [40]. During maturation, fibers become lignified [41], it is assumed that lignification of the secondary walls of the secondary phloem fiber bundles is one of the key determinants for bast fiber development in jute [42]. The mature cell composed of 60–65% cellulose, 20–25% hemi-cellulose, 11–14% lignin. In this study we focused on identification and characterization of FLA genes in *C. olitorius* (CoFLA) using bioinformatics approach including gene structure, motifs, physio-chemical properties, domain analysis, subcellular localization and phylogenetic relationship other species. Subsequently, the expression patterns of the FLA genes of stick and bark tissues from different regions were also investigated. This study serves as a foundation for future FLA genes research in jute.

## Materials and methods

### Identification of FLA gene family members in *C. olitorius*

In order to identify the genes encoding FLA in *C. olitorius*, all proteins of *C. olitorius* (Accession ID PRJNA215141) [31] were downloaded from National Center for Biotechnology Information (NCBI) database. The Hidden Markov Model (HMM) profile of the fasciclin (FAS) domain (PF02469) from pfam database (<https://pfam.sanger.ac.uk>) was used to retrieve the candidate FLA genes in *C. olitorius* using HMMER3.0 [43]. In addition, 21 Arabidopsis FLAs from TAIR database (<https://www.arabidopsis.org>) were used as the query against the *C. olitorius* reference proteome to identify putative *C. olitorius* FLA genes using BLASTP. The obtained candidate FLA proteins of *C. olitorius* were further screened to confirm the presence of FAS domain through the Simple Modular Architecture Research Tool (SMART, <https://smart.embl-heidelberg.de>) and pfam database (<https://pfam.xfam.org>) [44, 45]. Sequences with incomplete domain and other complications were discarded. The N-terminal signal peptide was identified by SignalP 5.0 (<https://www.cbs.dtu.dk/services/SignalP/index.php>) [46]. The big-PI Plant Predictor ([https://mendel.imp.ac.at/sat/gpi/gpi\\_server.html](https://mendel.imp.ac.at/sat/gpi/gpi_server.html)) was used to predict C-terminal GPI anchor signals [47]. After excluding the FAS domains, N and C-terminal signals, the remaining sequence were manually scrutinized to identify the AGP region containing two or more continuous (A/S/T)-P motifs. Subsequently the proportion of PAST of the candidate CoFLAs were calculated by an in-house Perl script.

### Phylogenetic analysis, multiple sequence alignment and naming of CoFLA genes

To understand the phylogenetic relationships among the FLA genes of Arabidopsis [3], Cannabis [16] and *C. olitorius*, initially the protein sequences were aligned by ClustalW and constructed a phylogenetic tree using the default parameters of neighbor-joining (NJ) method of MEGA X (version 10.0.5) [48] with a bootstrap value of 1000 replications. All the CoFLA genes were assigned according to phylogenetic tree and homology with Arabidopsis FLAs where a blastp program, with an e value of  $10^{-10}$ , was performed to find the homologous CoFLAs with Arabidopsis FLA proteins. The sequences of FAS domain from CoFLAs were aligned using Clustal Omega [49] software to find the highly conserved sequenced (H1 and H2) and Tyr-His motif as these conserve sequences are depicted in SM00554.

### Molecular characterization of jute FLA genes

Gene and protein length of the CoFLA genes were predicted from the *C. olitorius* reference genome database [31]. Protein properties such as isoelectric point (PI), grand average of hydropathicity (GRAVY) and molecular weight (MW) were calculated using ExPASy ([https://web.expasy.org/compute\\_pi](https://web.expasy.org/compute_pi)) [50]. Plant-mLoc (<https://www.csbio.sjtu.edu.cn/bioinf/plant-multi>) [51] was used to predict subcellular localization of CoFLAs. In house perl script was used to get the chromosomal/scaffold positions of CoFLAs. Unavailability of the chromosome based assembly led to use scaffolds to determine the position of CoFLAs. The exon/intron structures of the genes were illustrated through Gene Structure Display Server (GSDS, V.2, <https://gsds.cbi.pku.edu.cn>) [52]. Multiple Expectation Maximization for Motif Elicitation (MEME, <https://meme-suite.org/tools/meme>) [53] was employed to identify the conserved motifs with default settings except number of motifs where the number of motif was set as 15. Gene ontology (GO) annotation of CoFLAs for describing biological processes, cellular components and molecular functions were determined by the BLAST2GO software [54]. The output files of blastp and interproscan tools in Blast2GO were used to annotate GO categories and generate figures. The 1000 bp upstream sequences from transcription site of all CoFLA genes were retrieved from *C. olitorius* genome through in-house perl script using generic file format (GFF). The *cis*-acting regulatory elements were predicted through the web based PlantCAREprogram [55].

### Preparation of plant material

Tossa jute (*C. olitorius* cv. O-4) seeds were grown in greenhouse under controlled conditions with 70–80% humidity at  $33 \pm 2$  °C on a 13 h light/11 h dark cycle using cool white fluorescent light. After 45 days of growth, samples were collected from different stem regions with respect to snap point, a transitional point of bast fiber development where the elongation of the fiber cells is completed and secondary cell wall thickening starts [34]. Snap point was determined through cross section of jute stem every 0.5 cm from 3 to 8 cm (below the stem apex) and counting the number of fiber cells. We defined the snap point where numbers of fiber cells are constant or not increasing. In addition, we also confirmed snap point physically according to Gorshkova et al. [34] and Koziel [56] and it was 6.0 cm to 7.0 cm below the stem apex. Therefore, stem samples were taken from 3 to 9 cm from the stem apex and divided into 3 regions such as top part (3–5 cm) comprising the above portion of the snap point, middle region containing the snap point (5–7 cm) and bottom segment (7–9 cm) is composed of the lower portion of the snap point (S1 Fig). The bark tissues were collected separately by peeling off from the sticks. Then bark and stick

samples from the different regions were quickly frozen into liquid nitrogen and stored at  $-80\text{ }^{\circ}\text{C}$  until use. Three independent biological replications were taken for each sample and 10 plants were pooled for each replication.

### Primer design, RNA extraction and expression profiling

The primers of CoFLA genes (S1 Table) were designed using web based tool from IDT (<https://sg.idtdna.com>) and GenScript ([www.genscript.com](http://www.genscript.com)) followed by verification with Multiple Primer Analyzer (<https://www.thermofisher.com>). Primer specificity was evaluated by 1% agarose gel electrophoresis and melting curve analysis during qRT-PCR.

One gram of sample tissue was disrupted into fine powder with mortar-pestle, and then in-house modified CTAB [57] protocol was used to extract total RNA. Genomic DNA contamination was removed by amplification grade DNase I (Sigma-Aldrich, Germany). NanoDrop 2000 spectrophotometer (NanoDrop, Thermo Scientific) was used to determine the concentration of RNA and subsequently, the integrity of RNA was evaluated by 1% agarose gel electrophoresis. The first strand complementary DNA (cDNA) was synthesized using the RevertAidFirst Strand cDNA Synthesis Kit (Thermo Fisher Scientific, USA) and random primers from 1  $\mu\text{g}$  of total RNA according to the manufacturer's protocol. Then the samples were treated with RNaseH (Thermo Fisher Scientific, USA) to remove residual RNA. The quantitative RT-PCR was performed with a QuantStudio Real-time PCR system (Applied Biosystems, USA) using 96 well-plate with a final reaction volume of 20  $\mu\text{L}$  containing 5 ng cDNA, 10  $\mu\text{l}$  of 2X PowerUp<sup>TM</sup> SYBR<sup>TM</sup> Green Master Mix (Applied Biosystems, USA) and 300 nM of each primer. The qRT-PCR cycling for all genes were carried out according to following conditions: initial activation step of 2 min at  $50\text{ }^{\circ}\text{C}$ , then denaturation step for 10 min at  $95\text{ }^{\circ}\text{C}$  followed by 40 cycles of  $95\text{ }^{\circ}\text{C}$  for 15 s, primer specific annealing temperatures (S1 Table) for 20 s, and  $72\text{ }^{\circ}\text{C}$  for 20 s. To check the specificity of the primers, a melting curve was generated at the end of qRT-PCR cycles by constant increment of temperature from 60 to  $95\text{ }^{\circ}\text{C}$ . The qRT-PCR reactions were performed with three biological replicates and three technical replicates for each biological replicates along with no template control to ensure reliability and reproducibility. Expression level of CoFLA genes were analyzed with the comparative CT method ( $2^{-\Delta\Delta\text{Ct}}$ ) described by Livak et al. [58] and two housekeeping genes such as PP2Ac and EF2 [59] were used as internal control genes for normalization of qRT-PCR analysis.

## Results

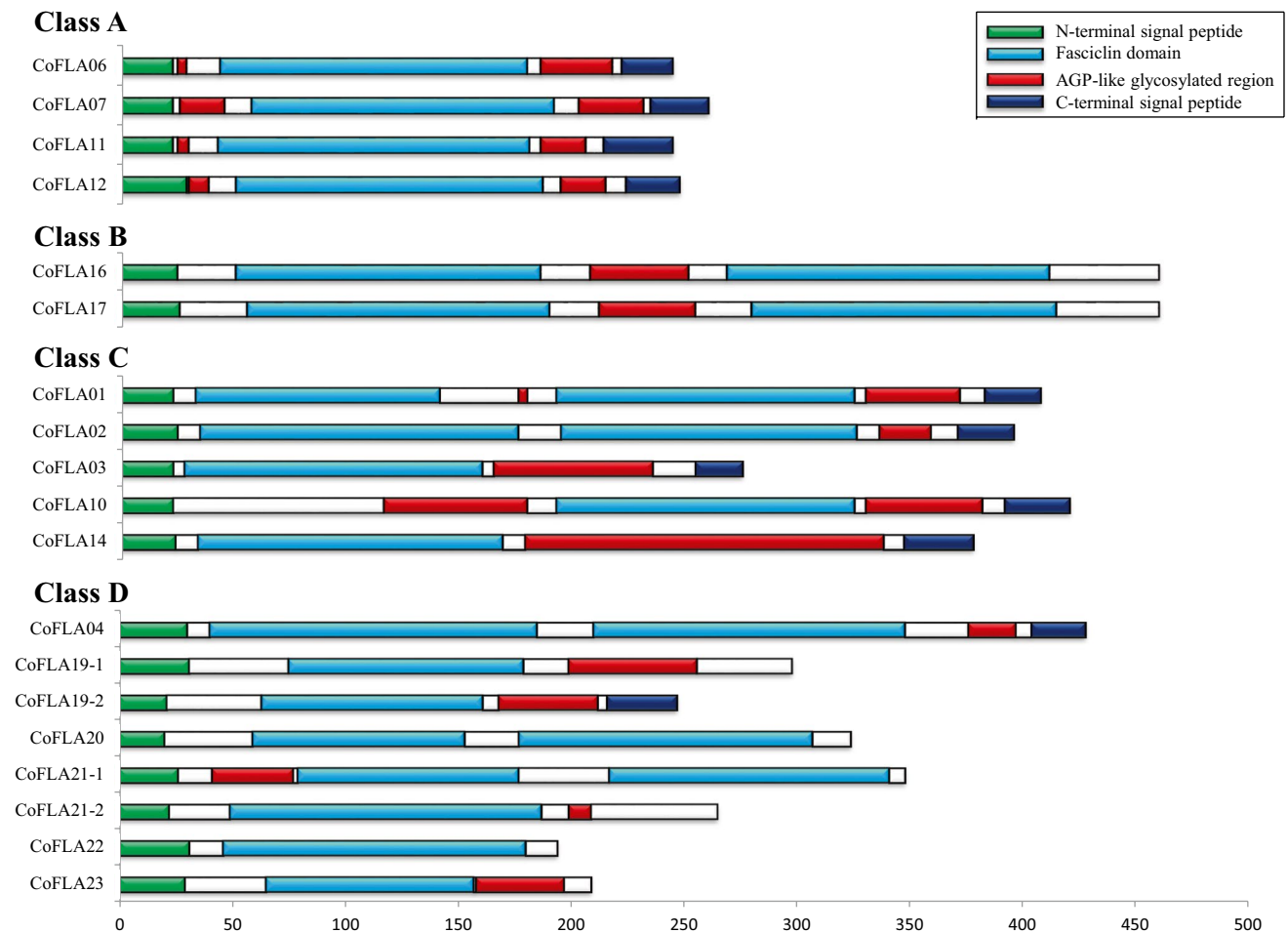
### Genome-wide identification of putative FLA genes in *C. olitorius*

HMM profile of FAS domain (PF02469) and AtFLA protein sequences were blast against *C. olitorius* protein database led to the identification of FLA genes in jute. The retrieved protein sequences were further screened for the presence of FAS domain using SMART and Pfam database followed by N-terminal signal peptide detection. Then the sequences were further screened for the presence of AGP-like glycosylated regions that contains at least two noncontiguous Hyp residues. Finally, 19 FLA genes were identified in *C. olitorius* genome and assigned as CoFLA (S2 Table) containing N-terminal signal peptide sequences and one or two FAS domains. Seven FLA genes CoFLA (01, 02, 04, 16, 17, 20, 21-1) harbored two FAS domain while the remaining had only one FAS domain (Fig. 1). PAST constitutes a certain proportion of CoFLAs and it ranges from 23 to 42% except CoFLA14 (54%) (S3 Table).

Manual checking of the Hyp sequence motifs revealed the presence of at least one [A/S/T]-P-X(0–10)-[A/S/T]-P except for CoFLA 21-2 and 22 (S4 Table). Three genes such as CoFLA 03, 14 and 21-2 had [A/S/T]-P-P-P motifs while 8 CoFLA genes 04, 07, 10, 14, 16, 17, 19-2 and 23 possessed [A/S/T]-P-P motif. However, it is noted that CoFLA 20 and 22 were seemed to lack two or more non-contiguous [A/S/T]-P motifs when excluding FAS domain, N- and C-terminal signal sites. Glycosylphosphatidylinositol (GPI) membrane anchor in C-terminal that helps in protein anchorage to the cell surface of eukaryotes [15] was another distinguished features were found in FLA. Among 11 CoFLA genes out of 19, (57.89%) were predicted to have a GPI anchor sites (Fig. 1).

### Phylogeny, naming, multiple sequence alignment and classification of CoFLAs

To explore the evolutionary relationships among the FLA genes of different species, a neighbor-joining (NJ) phylogeny tree was constructed using 21 AtFLA genes, 23 CsaFLAs and identified 19 CoFLA genes with 1000 bootstrap replicates. All the CoFLA genes were assigned according to their positions in the phylogenetic tree and homology with the Arabidopsis FLA genes (S5 Table). Like FLA genes in other species [3, 10, 15, 16], CoFLA genes were also classified into four phylogenetic classes (class I–IV) as shown in Fig. 2. The proteins, namely, CoFLA 06, 07, 11, 12 and CoFLA 01, 02, 03, 10, 14 belonged to class I and III, respectively. Class II had the lowest number (2)



**Fig. 1** Schematic representation of putative CoFLA genes. The CoFLAs are classified into four classes (a–d). The color regions indicate the N-terminal signal peptide (green), fasciclin domain (blue), AGP

glycosylated region (red), C-terminal signal (dark blue) and remaining protein regions (white). (Color figure online)

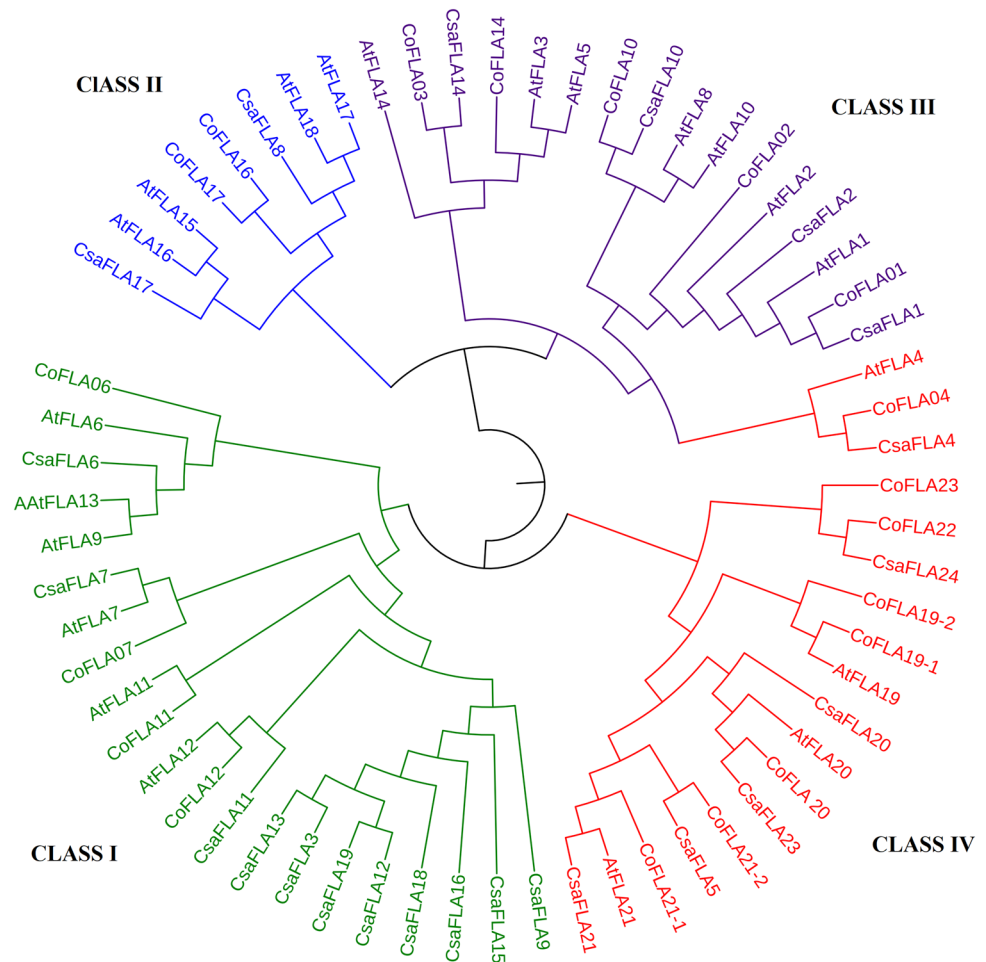
of CoFLA genes such as CoFLA 16 and 17 while class IV contained the highest number (8) of CoFLA genes namely, CoFLA 04, 19-1, 19-2, 20, 21-1, 21-2, 22, and 23. Three pairs of orthologous genes including AtFLA11/CoFLA11, AtFLA12/CoFLA12 and AtFLA19/CoFLA19-1 were identified in *Arabidopsis* and *C. olerarius*.

A multiple sequence alignment of the FAS domain of CoFLA genes identified the H1 and H2 conserved regions (Fig. 3) which are common for all FAS domain. The H1 and H2 region is characterized by [S/T]-[V/L/I]-F-A-P-X-[D/E/N]-X-A and [V/L/I]-[F/Y/H/Q]-X-[V/L/I]-X-X-[V/L/I]-[V/L/I]-[V/L/I]-P sequence, respectively. In addition, another conserved region characterized by [YF]H was located between the H1 and H2 regions (Fig. 3). The Thr residue in H1 regions of all CoFLA was conserved just like in cabbage [2], *Arabidopsis* [3], poplar [10] and cotton [15]. Moreover, Ile or Leu was close to the H1 regions in most CoFLAs that might be involved in maintaining FAS domain structure and/or cell adhesion [60]. An Asn/Asp/Glu residue was also

present at the sixth position after the Thr residues. The C terminal in H2 regions also possessed conserved sequence with small hydrophobic amino acids such as Val, Leu and/or Ile except for the CoFLA21-1 as shown in Fig. 3. Only CoFLA 01 and 22 contained Cys and Ser residue instead of His in [YF]H motif, respectively. Additionally, Tyr/Phe residue was substituted by Asn, Val, Leu, Arg and Ser in some FAS domain of *C. olerarius*. Pro residue was conserved in the [YF]H region for the majority of CoFLA genes. Moreover, the [YF]H motif was flanked by [L/V/I]-[L/V/I] in most of the CoFLA.

Like FLA genes in other species, CoFLA genes were also classified into four classes (Fig. 1) on the basis of FAS domains, AGP-like regions, GPI anchor and amino acid similarity as described earlier by Johnson et al. in *Arabidopsis* [3]. The proteins CoFLA 06, 07, 11 and 12 belonged to class A with a single FAS domain flanked by two AGP regions and a C terminal GPI anchor sites. Class B contained only two CoFLA proteins such as CoFLA 16 and 17 that were

**Fig. 2** The phylogenetic relationships of FLA proteins of in *C. olierius*, Arabidopsis and hemp. The Neighbor-Joining tree was constructed with 21 genes of Arabidopsis (AtFLA), 23 genes of hemp (CsaFLA) and 19 genes of *C. olierius* (CoFLA) using MEGAX. The bootstrap value was 1000 replicates. The FLA proteins were clustered into four distinct classes (I–IV), represented by different colors. (Color figure online)



characterized by two FAS domain and a single AGP region between two FAS domain without any AGP site. Class C included five CoFLAs such as CoFLA 01, 02, 03, 10 and 14 where they had one or two FAS domain, one or two AGP regions and a GPI anchor site. Although CoFLA 03 and 14 had similar protein architecture to class A but they were placed in class C because their sequences had more similarity with other genes of class C. Class D was the largest class that comprised CoFLA 04, 19-1, 19-2, 19-3, 20, 21-1, 21-2 and 22 with no distinct characteristics and less similar among the classes or to any other CoFLA proteins.

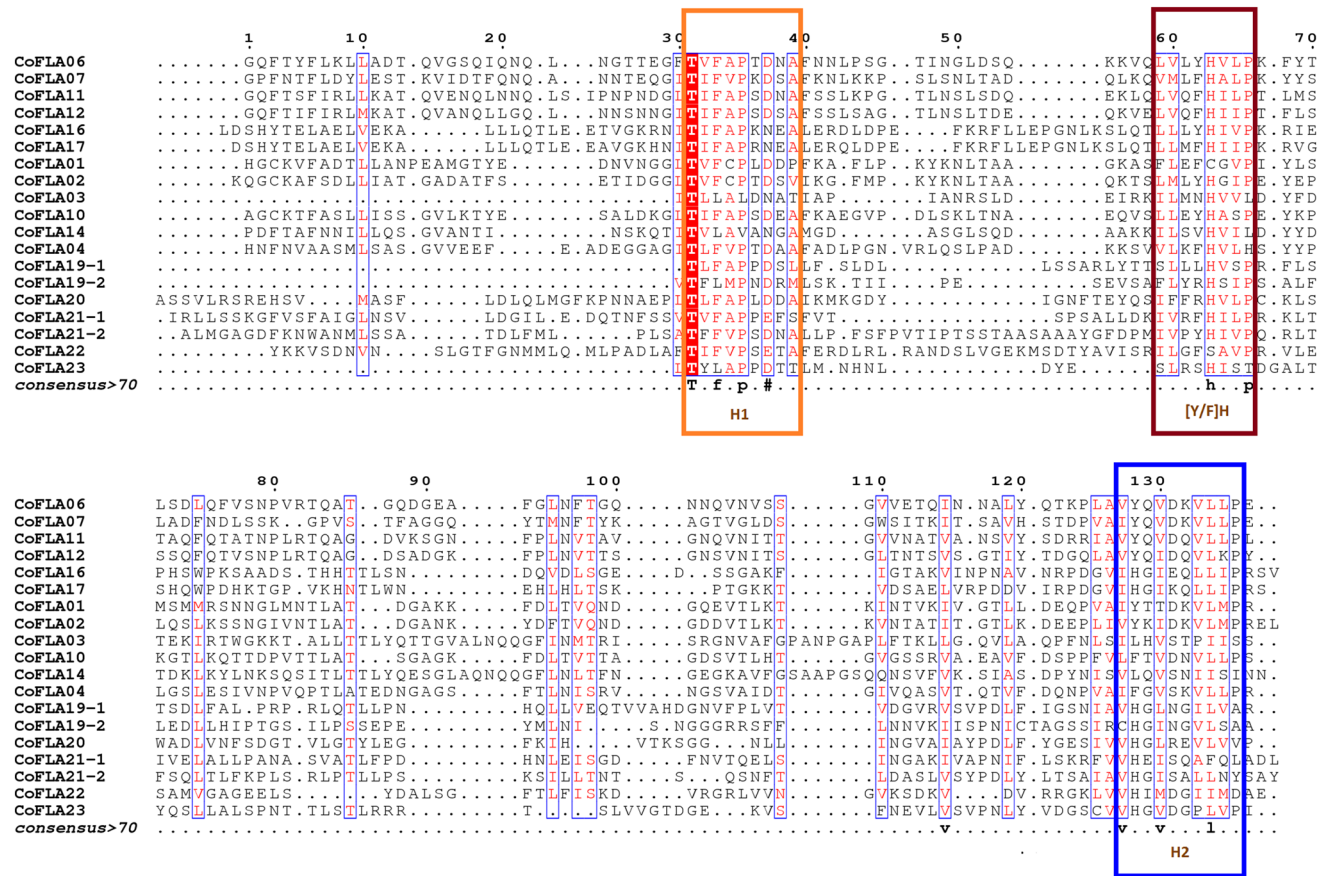
### Molecular characteristics of CoFLA genes

The characteristics of the 19 CoFLA genes such as protein size, theoretical isoelectric point (pI), molecular weight (MW), grand average of hydropathicity (GRAVY) and subcellular localization are summarized in Table 1. The length of CoFLAs ranged from 194 to 463 aa where class B CoFLA genes were longer than others. Although the theoretical isoelectric point (pI) ranged from 4.94 to 9.28, most of the CoFLAs displayed acidic pI (4.94–6.39) except for CoFLA

11, 14 and 23. The molecular weights (MW) of CoFLA proteins varied from 21.35 to 51.01 kDa. The GRAVY values of five CoFLAs were negative suggesting their hydrophilic behavior while the rest of them were hydrophobic as their GRAVY values were positive. The subcellular localization predictions revealed that all CoFLA proteins were located in the cell membrane except CoFLA 17, 22 and 23 as they were also located in nucleus and chloroplasts, respectively. Moreover, some genes were located in mitochondria and endoplasmic reticulum but the reliability was very low.

Among 10 CoFLAs out of 19 were unevenly distributed across 4 chromosomes and rest was located in unanchored scaffolds as listed in Table 1. Four genes were distributed on chromosome 2, three genes were located in chromosome 4, two genes were placed in chromosome 3 and chromosome 1 had only one gene whereas around half of the genes were located on unanchored scaffolds due to unavailability of chromosome based assembly.

The exon–intron arrangements may play an important role in the process of diversification for a gene family. Therefore, we analyzed exon–intron organizations using GSDS software together with a phylogenetic tree among CoFLA



**Fig. 3** Multiple sequences alignment of the CoFLA fasciclin domains. The alignment was generated by clustalW. The conserved regions such as H1, H2 and [Y/F]H of fasciclin domain were indicated by orange, magenta and red box, respectively. (Color figure online)

genes. Only two CoFLA 01 and 02 had one intron flanked by two exons and they were paralogous, however, the length of intron significantly varied. Besides this, all other CoFLA genes harbored only one exon (Fig. 4).

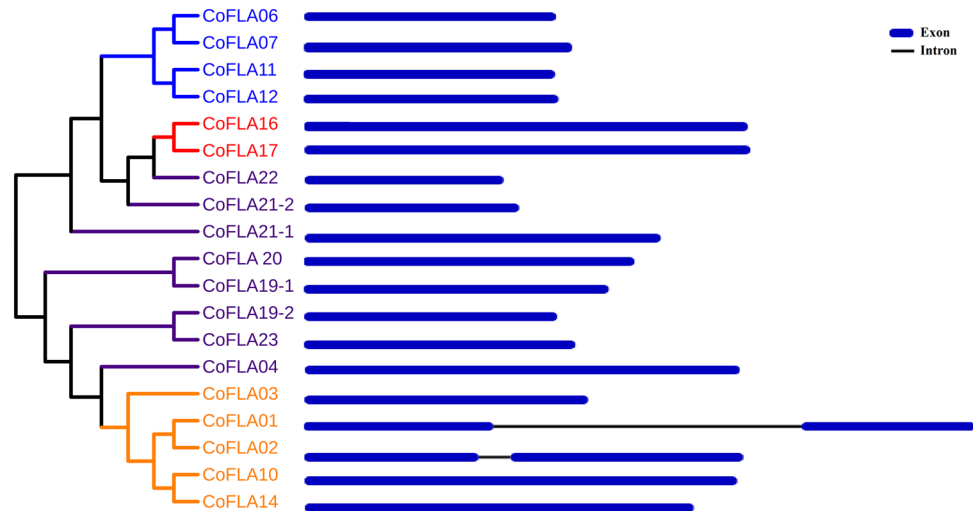
The classifications of CoFLAs were also supported by conserved motif analysis. In the study, we have identified 15 distinct motifs ranged from 7 to 50 aa (S6 Table) and number of motifs varied from 1 to 11 for CoFLAs as shown in Fig. 5. As expected, each subfamily shared common motifs suggesting their functional similarity excluding the dissimilar class D. For example, motif 1, 2, 3, 4 and 13 were found in all genes of class A. Motif 3, 4, 6–9 and 11–13 were specific to class B. Seven motifs (1–5 and 12, 13) were shared by the CoFLAs of class C. Few motifs were found in class D and shared a common motif 4. However, CoFLA4 of class D, had conserved motifs similar to that of class C. It was observed that motifs 3 and 4 commonly shared by almost all of the CoFLA proteins. The presence or absence of conserved motifs might play vital role in protein functions and needs further research for characterization. The sequence and logo of the motifs from CoFLA proteins are presented in S6 Table and S2 Fig.

The results of Blast2GO indicated that the CoFLAs were enriched in cell (GO:0005623), integral component of membrane (GO:0016021), anchored component of plasma membrane (GO:0046658) and plant-type cell wall (GO:0009505) of the cellular component categories (Fig. 6a, S7 Table). Transferase activity (GO:0103068), catalytic activity (GO:0102953, GO:0004672) and binding of molecules such as ATP, polysaccharide (GO:0030246, GO:0005524) were the main activities in molecular function categories as shown in Fig. 6b. In the biological processes categories, cell adhesion (GO:0007155) was highly enriched followed by response to hormone including auxin, gibberellin, abscisic acid and cytokinin (GO:0009733, GO:0009739, GO:0009737, GO:0009735). Response to salt stress (GO:0009651), cell wall biogenesis (GO:0009834, GO:0009833) and seed development (GO:0090378, GO:0010262) were the other functions in this categories (Fig. 6c, S7 Table).

**Table 1** Characteristics of FLA genes of *C. olitorius*

Gene ID	Class	Protein length (aa)	Localization	Mol. Wt (kDa)	pI	Gravy	Chromosome /scaffold no
CoFLA 01	C	410	Cell membrane	43.81	6.25	0.019	scf7180001685597
CoFLA 02	C	398	Cell membrane	42.68	5.8	-0.098	scf7180001684603
CoFLA 03	C	278	Cell membrane	28.88	5.25	0.031	3
CoFLA 04	D	429	Cell membrane	45.58	5.79	0.256	3
CoFLA 06	A	247	Cell membrane	25.60	5.75	-0.033	scf7180001684361
CoFLA 07	A	263	Cell membrane	27.87	5.75	0.058	4
CoFLA 10	C	424	Cell membrane	43.47	5.35	0.156	scf7180001685705
CoFLA 11	A	247	Cell membrane	26.11	8.71	0.204	2
CoFLA 12	A	250	Cell membrane	26.49	6.27	0.051	2
CoFLA 14	C	381	Cell membrane	38.10	9.28	-0.134	4
CoFLA 16	B	457	Cell membrane	48.09	6.14	-0.3	scf7180001684849
CoFLA 17	B	463	Cell membrane/chloroplast	51.01	6.16	-0.258	2
CoFLA 19-1	D	299	Cell membrane	32.16	6.39	0.13	scf7180001685832
CoFLA 19-2	D	248	Cell membrane	43.81	6.25	0.263	scf7180001684888
CoFLA 20	D	324	Cell membrane	35.00	4.98	0.281	4
CoFLA 21-1	D	348	Cell membrane	38.59	5.69	0.287	2
CoFLA 21-2	D	265	Cell membrane	27.97	5.66	0.493	scf7180001685709
CoFLA 22	D	194	Cell membrane/nucleus	21.35	4.94	0.213	1
CoFLA 23	D	210	Cell membrane/nucleus	22.53	8.84	0.094	scf7180001685638

**Fig. 4** Schematic representation of exon/intron structure of CoFLA genes. A phylogenetic tree was constructed with CoFLA genes with MEGAX and the different groups marked with different colors. Blue box represents exons and black line indicates intron. (Color figure online)



### Cis-elements analysis of CoFLA genes

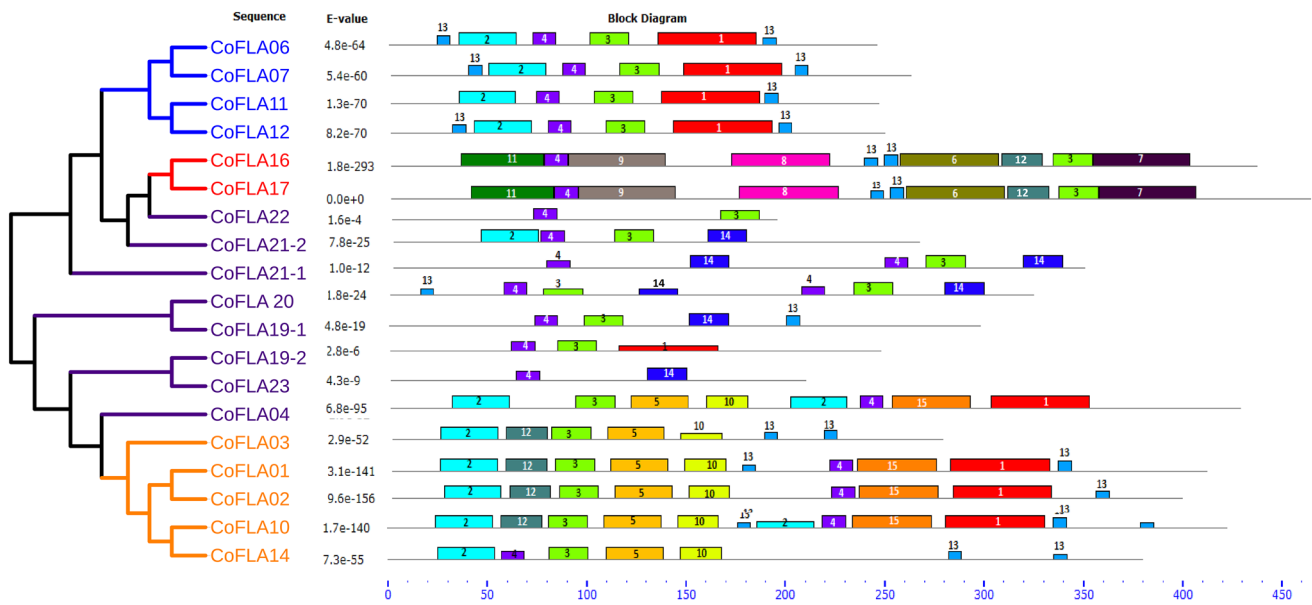
*Cis*-regularity elements are present in the upstream of transcriptional start sites regulating the expression of the genes. A total of 78 *cis*-elements found in the promoter regions that were divided into four major groups such as light-responsive, hormonal/environment responsive, sites binding-related element and promoter core/function element (S8 Table and S3 Fig). The number of promoter function elements such as TATA and CAAT box was higher than any other *cis*-elements. Furthermore, light-responsive and

hormonal/environment responsive elements were also abundantly found in promoter regions which indicated that COFLAs might be involved in plant growth and development. Additionally, the upstream of all CoFLAs contain MYB that belongs to sites binding-related element.

### Expression analysis

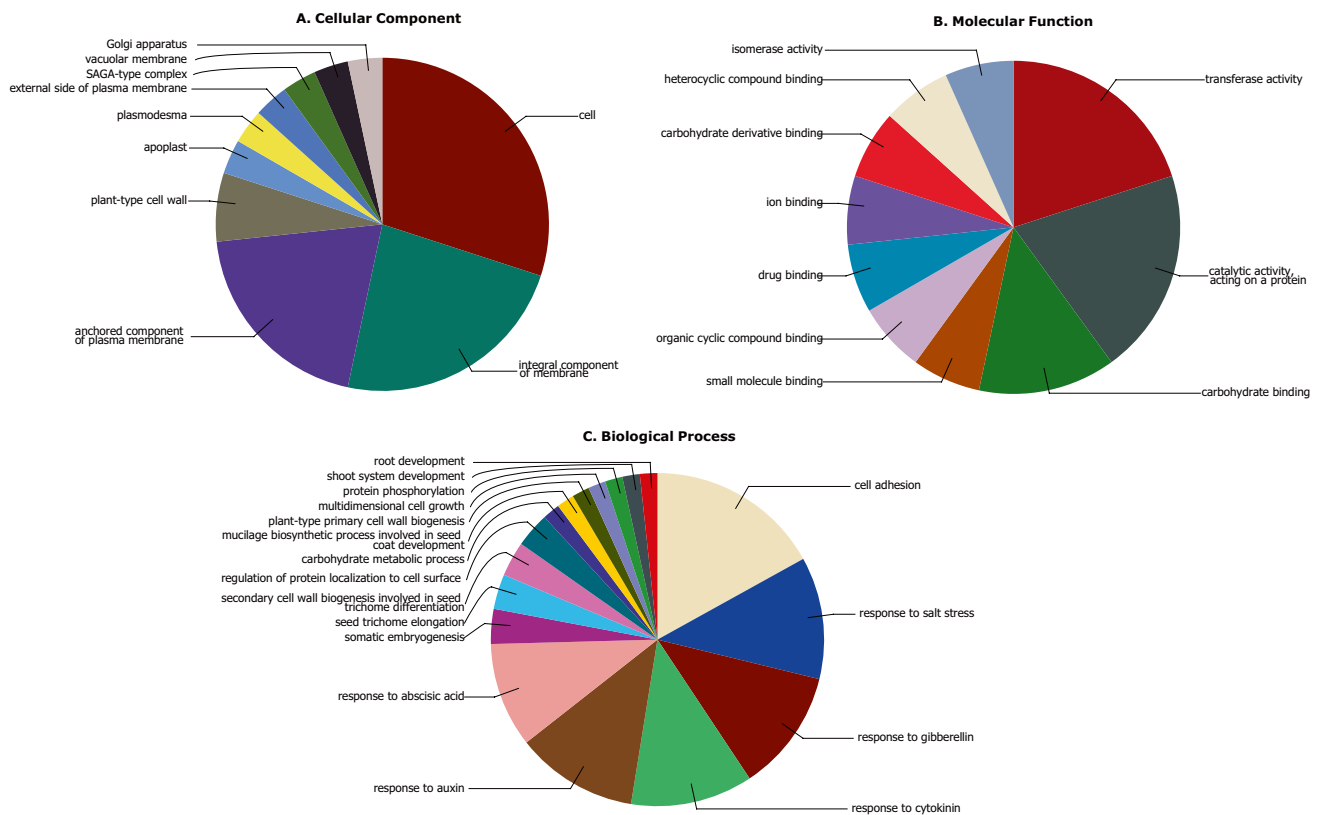
Jute is important crop for unraveling the mechanisms of phloem fiber development and FLAs are associated with cell interaction, adhesion and cell wall biosynthesis [10, 16].





**Fig. 5** Schematic representation of CoFLA motif analysis. A phylogenetic tree was constructed with CoFLA genes with MEGAX and different groups are marked with different colors. Distinct motifs

were identified by MEME suite and each motif are represented by different colored boxes with number 1–15. (Color figure online)

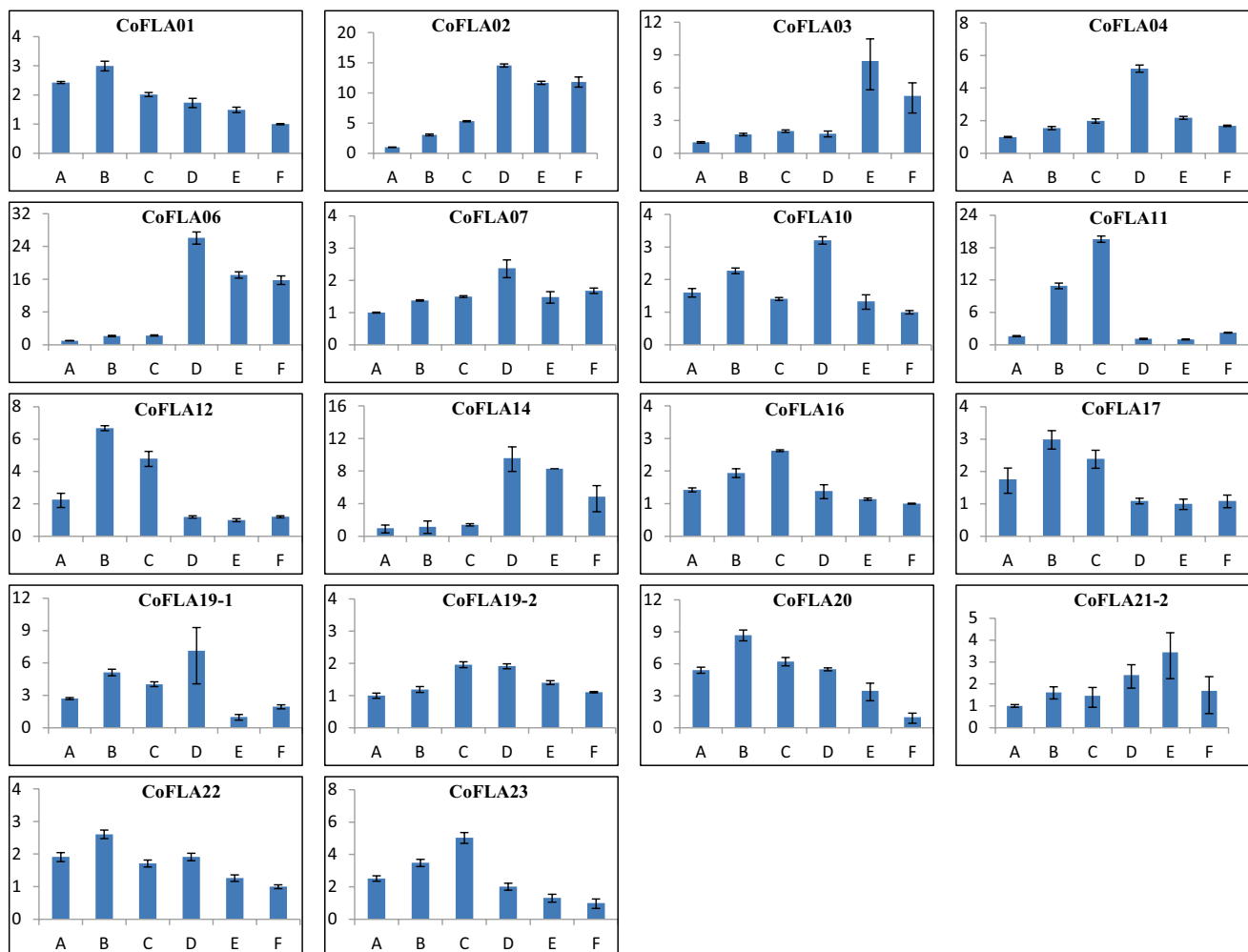


**Fig. 6** Gene ontology (GO) distributions for the CoFLAs in jute. The Blast2GO program were used to define the gene ontology under three categories. **a** Cellular component, **b** molecular function and **c** biological processes

Therefore, the expression patterns of CoFLAs were analyzed among bark tissues containing fiber cells and stick of different region based on snap-point as described in ‘Materials and Methods’ to investigate the role of the CoFLA genes during fiber cell formation. All the CoFLA genes were expressed in all the tissues except CoFLA21-1 which was expressed at very low level ( $Ct > 34$ ). Therefore, we did not perform the expression analysis of the CoFLA21-1 among the tissues. The results revealed that the expression of CoFLA 01, 11, 12, 16, 17, 20, 22 and 23 was higher in bark compared to stick tissues (Fig. 7). Among them, CoFLA 01, 12, 17, 20 and 22 was expressed at high levels in the middle portion of bark containing the snap point. In detail, the expression of CoFLA 12 and 20 increased about 7 and eightfold, respectively while CoFLA 01, 17 and 22 were expressed approximately threefold higher in middle bark

tissue. By contrast, CoFLA 11, 16 and 23 showed higher expression in bottom bark tissues (below the snap point). CoFLA11 was significantly expressed (about 20 fold) in bottom region of bark while the transcripts levels of CoFLA 16 and 23 were accumulated by three and sixfold higher respectively, in lower part of bark tissues.

On the other hand, CoFLA 02, 03, 06, 14 and 19-1 were predominately expressed in stick tissues while CoFLA 04, 07, 10, 21-2 had moderately higher expression than bark as shown in Fig. 6. CoFLA 02, 06 and 14 were expressed at higher levels in the top stick tissues and decreased towards the bottom stick tissues where the expression levels of CoFLA 02 and 06 were up to 15 and 28 fold higher, respectively in top stick tissues. Similarly, CoFLA 14 and 19-1 were also highly expressed in upper part of stick tissues. The transcript level of CoFLA03 showed a distinctive expression



**Fig. 7** Expression analysis of CoFLA genes using qRT-PCR. The  $2^{-\Delta\Delta Ct}$  method was used to calculate the transcript level of each CoFLA where expression level was calculated to highest average Ct value of the sample. CoPP2Ac and CoEF2 were used as internal control genes. Data represents average of three technical replicates from

three biological replicates. The error bars indicate standard error from the replication (total  $3 \times 3 = 9$  replicates). **a** top (3–5 cm) bark, **b** middle (5–7 cm) bark, **c** bottom (7–9 cm) bark, **d** top (3–5 cm) stick, **e** middle (5–7 cm) stick, **f** bottom (7–9 cm) stick

in the middle part of stick tissues containing snap-point. In addition, the CoFLA 04, 07, 10 and 21-2 were preferentially expressed (about 3-fivefold) in top stick tissues.

## Discussion

FLAs are one of the classes of arabinogalactan proteins that have significant impact on plant growth and development especially on secondary cell wall biosynthesis and responses to adverse conditions [10]. The FLAs has been characterized in some plant, however, very little information is known about jute FLA genes but recently characterizes the bast fiber-specific FLA genes in flax [61]. Jute is the second most important fiber crop which is becoming popular due to its environmental friendly nature and some evidence has indicated that FLAs may have vital role in fiber development [16, 22]. Therefore, based on draft genome of jute (*C. olitorius*) [31], we identified 19 FLA genes in jute which were named as CoFLA. The number of CoFLAs was comparable to the number of FLA genes in other species such as 21, 19, 23 and 18 genes in Arabidopsis [3], cotton [15], hemp [16] and eucalyptus [17], respectively, however, a total of 35 FLA proteins were identified in poplar [10]. A nomenclature of CoFLA was applied according to phylogenetic tree and homology to AtFLA because it is easy for the readers to distinguish the Arabidopsis homolog. Like other species, CoFLAs were classified into four classes based on protein structure and sequence similarity (Fig. 1). In addition, the genes of each class were placed in same clade of the phylogenetic tree except CoFLA04 as shown in Fig. 2. All the identified CoFLAs had an N terminal signal peptide, one or two FAS domain where majority of CoFLAs contained AGP regions and half of the CoFLAs presented C terminal GPI anchor site. Class A included four CoFLAs (06, 07, 11 and 12) containing single FAS domain flanked by two AGP regions, C terminal GPI anchor. The protein sequence similarity among the members of class A was from 31 to 72%. This class was located into class I in the phylogenetic tree where the FLAs in class A of Arabidopsis and hemp also belonged to class I [3, 16]. Class B, the smallest class, containing CoFLA 16 and 17, had over 73% similarity which is in accordance with the values reported in Arabidopsis [3] and poplar [10]. The class B CoFLAs contained two FAS domains and one AGP regions between FAS domains without any C-terminal GPI anchor region. This class B belongs to class II in the phylogenetic tree. Five CoFLAs formed class C that had one or two fasciclin domains, one or two AGP motifs and a C-terminal GPI anchor. Although CoFLA 03 and 14 had similar protein structure to class A, these two genes were placed in class C instead of class A. Because these two genes showed more amino acids similarity with the members of class C than class A. The similarities among

the members of class C were from 34 to 61% whereas the sequence similarities between class A and C were 26% to 36%. Class D was the largest class with 8 CoFLAs that had no relationship with each other or to any other CoFLAs and showed remarkably low similarity. Most of the CoFLAs in these class had no C-terminal GPI anchor regions except for CoFLA 04 and 19-2. CoFLA 04, 20 and 21-1 contained two fasciclin domains while other members of class D had one FAS domain. Except for CoFLA 20 and 22, the CoFLAs of class D possess one AGP like glycosylated region. CoFLA04 had similar structural arrangement with CoFLA 01 and 02 of class A but it was placed in class D because of low sequence similarity among them (below 23%). However, the phylogenetic position of CoFLA04 was more close to class C than class D. Consistent with our result, AtFLA4 which is homolog with CoFLA04 was also placed in class D, although it showed similar structure arrangement with class C [3]. Class C and D were positioned in class III and IV, respectively in the phylogenetic tree.

All CoFLAs possessed (A/S/T)-P/PP/PPP glycomodules, however, CoFLA 20 and 22 lacked well-defined AGP like glycosylated region as AGP region defined as [A/S/T]-P-X(0–10)-[A/S/T]-P or at least one [A/S/T]-P<sub>(2–4)</sub> excluding FAS domain, N and C terminal signal regions [1, 13]. Absence of AGP regions in FLA protein was also reported in wheat, rice, cabbage and poplar [1, 2, 10] however, all the AtFLAs had well characterized O-glycosylation regions [3]. The CoFLA 20 and 22 may not undergo O-glycosylation as they did not contain AGP glycomodules, however non-AGP protein such as sporamin undergoes arabinogalactosylation [1]. So it would be interesting to investigate the pattern of O-glycosylation in CoFLAs 20 and 22 and as well as FLAs from other species which had no well-characterized AGP regions. The length of AGP regions was varied among CoFLAs where longest glycosylated region was found in CoFLA14 (Fig. 1) consisted of 160 amino acids with 30 O-glycosylation sites. Interestingly, this AGP region is significantly longer than the AGP regions of FLA from Arabidopsis and poplar [3, 10]. On the other hand, the smallest AGP region was found in CoFLA 01, 02 and 21-1 with four amino acids containing two potential O-glycosylation sites.

GO annotations is a useful tool for describing gene localizations and functions. In this study, GO annotation analysis revealed that the majority of the CoFLAs were located in cell, membrane and cell wall. Additionally, the most of these genes were involved in cell adhesion, hormonal regulation, abiotic stress, cell wall biogenesis, catalytic and binding activity. These features support the previous findings as FLAs are associated with cell adhesion, cell wall biogenesis and anchoring molecules. Similarly, the *cis*-regulatory elements in promoter region of CoFLAs suggested that FLAs were associated with plant development especially in cellular development and stress tolerance/regulation. Because,

the elements of promoter core function, site binding, light, hormone and environment responses were more abundant than other elements in the upstream region of CoFLAs.

The snap-point is a mechanically defined region where fibers transit from elongation to thickening [34]. This point allows us to elucidate the molecular events and identify the key genes associated with fiber development that can help in future breeding program for the development of new varieties with agro-economic traits fiber quality, yield, etc. Different fiber developmental stages such as fiber initiation, elongation and secondary cell wall formation elucidated the number, length and composition of the fibers [62, 63] have led to the identification a region of stem at 6–7 cm from the shoot apex as the snap point of jute which was corroborated by the results of other studies in flax and hemp [34, 63]. In order to identify the roles of FLA proteins in fiber cell initiation, elongation and secondary cell wall synthesis we extracted RNA from different regions of bark containing fiber tissues. Moreover, stick of different development regions were also considered to compare the expression pattern of CoFLAs during the development of xylem and phloem in jute. The results of qRT-PCR analysis suggested that some CoFLAs might have function in jute fiber development (Fig. 6). Within class A, CoFLA 11 and 12 were highly expressed in bark tissues while CoFLA06 was exclusively detected in stick tissues. Among them, CoFLA11 showed a gradual increase of expression from the top to bottom part of bark tissue. With consistent to our result, AtFLA11 (homolog to CoFLA11) was significantly expressed in inter fascicular fiber and lignifications in this tissues [64]. As below the snap-point, the fiber cells start formation of secondary cell wall related material such as lignin, CoFLA11 might play an important role in secondary cell wall development of jute fibers. A similar expression trend was observed in hemp fiber tissues: CsaFLA11 was more expressed in bottom bast tissues [16]. On the other hand, CoFLA12 displayed highest mRNA level in middle part of bark suggesting the probability of FLA gene having a significant role in fiber cell expansion/elongation. Consistent with our results, AtFLA12, (homologous to CoFLA12) showed stem specific expression in Arabidopsis [20]. Moreover, fiber length of cotton was increased due to over-expression of GhFLA1 which was orthologous to AtFLA12 whereas silencing of GhFLA1 by RNAi had adverse effect [22]. However, CsaFLA12 of hemp was thought to be involved in cell wall development process as it had more expression in the middle and bottom part of the stem [16]. Like CsaFLA06, CoFLA06, another member of class A, was specifically expressed in top stick tissues suggesting their role in stick development [16].

The expression patterns of CoFLA 16 and 17 were comparatively high in bark than in stick specifically CoFLA16 was expressed more in bottom while CoFLA17 in middle bark tissues. However, these results were not completely

concordant with the expression of the flax FLA 16, 17 and 18 (orthologous to CoFLA 16 and 17) that were up-regulated below the snap point [25]. Therefore, class B FLA genes might be associated with cell wall related processes and fiber elongation. All CoFLAs in class C were up-regulated in top or middle part of stick except CoFLA01 indicating that this group might be involved in stick development. Specifically, CoFLA 02, 03 and 14 had quite different FLA structure such as two FAS domain or long AGP like region along with GPI anchor (Fig. 1) which helps in cell adhesion during stick development. This result was consistent with previous studies in industrial hemp, where CsaFLA 01 and 08 of group C had more expression in top part of stick [16]. Similarly, AtFLA4 (homologous to CoFLA4) was involved in cell expansion and cell wall synthesis [18]. The higher expression of CoFLA04 of class D in apical part of stick suggests their role in cell expansion or adhesion in stick development. The CoFLA20 was up-regulated in middle part of bark tissues while the expression of CoFLA23 was higher in bottom part of the bark suggesting that these two genes might be involved in fiber elongation and secondary cell wall synthesis process, respectively. To sum-up, CoFLAs may have function in different stages of fiber cell development in jute, especially CoFLA 12 and 20 and CoFLA 11 and 23 may have function in elongation and secondary cell wall deposition, respectively. Since increasing of fiber length and reducing lignin content in jute fiber are essential for its suitability in textile industry, therefore alteration of FLAs expression might provide the way for developing long fiber cell or low lignin content jute variety.

## Conclusion

FLAs are a subclass of AGP proteins which play an important role in plant growth and development, notably in cell adhesion and cell wall biosynthesis. Based on jute genome, a total of 19 FLA genes were identified and characterized through a systematic approach including gene structure analysis, phylogeny, subcellular localization, motif and promoter analysis. Furthermore, expression profiling of the CoFLAs in different stages of phloem fiber development revealed their involvement in fiber elongation and cell wall deposition. This comprehensive study of FLAs in jute provides foundation for understanding functional roles in bast fiber development and may help to identify candidate genes in breeding program for jute variety development with specific agronomic traits.

**Acknowledgements** The authors are thankful to Emdadul Mannan Emdad for his technical support during bioinformatics analysis.

**Author contributions** MSI and MS Hossain designed the research. MS Hossain carried out the experiments, analyzed the data and wrote the manuscript. BA and MS Hossain contributed in bioinformatics analysis. MWU and NA revised the manuscript. MS Haque and MSI directed and reviewed the manuscript. All the authors read, discussed and approved the final manuscript.

**Funding** The author(s) received no specific funding for this work.

**Data availability** The datasets supporting the conclusions and description of a complete protocol can be found within the manuscript and its additional files.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no competing interests.

## References

1. Faik A, Abouzouhair J, Sarhan F (2006) Putative fasciclin-like arabinogalactan-proteins (FLA) in wheat (*Triticum aestivum*) and rice (*Oryza sativa*): identification and bioinformatic analyses. *Mol Genet Genom* 276(5):478–494. <https://doi.org/10.1007/s00438-006-0159-z>
2. Jun L, Xiaoming W (2012) Genome-wide identification, classification and expression analysis of genes encoding putative fasciclin-like arabinogalactan proteins in Chinese cabbage (*Brassica rapa* L.). *Mol Biol Rep* 39(12):10541–10555. <https://doi.org/10.1007/s11033-012-1940-1>
3. Johnson KL, Jones BJ, Bacic A, Schultz CJ (2003) The fasciclin-like arabinogalactan proteins of Arabidopsis. A multigene family of putative cell adhesion molecules. *Plant Physiol* 133(4):1911–1925. <https://doi.org/10.1104/pp.103.031237>
4. Showalter AM (2001) Arabinogalactan-proteins: structure, expression and function. *Cell Mol Life Sci* 58(10):1399–1417. <https://doi.org/10.1007/pl00000784>
5. Gaspar Y, Johnson KL, McKenna JA, Bacic A, Schultz CJ (2001) The complex structures of arabinogalactan-proteins and the journey towards understanding function. *Plant Mol Biol* 47(1–2):161–176
6. Seifert GJ, Roberts K (2007) The biology of arabinogalactan proteins. *Annu Rev Plant Biol* 58(1):137–161. <https://doi.org/10.1146/annurev.arplant.58.032806.103801>
7. Pereira AM, Pereira LG, Coimbra S (2015) Arabinogalactan proteins: rising attention from plant biologists. *Plant Reprod* 28(1):1–15. <https://doi.org/10.1007/s00497-015-0254-6>
8. Schultz CJ, Rumsewicz MP, Johnson KL, Jones BJ, Gaspar YM, Bacic A (2002) Using genomic resources to guide research directions. The arabinogalactan protein gene family as a test case. *Plant Physiol* 129(4):1448–1463. <https://doi.org/10.1104/pp.003459>
9. Showalter AM, Keppler B, Lichtenberg J, Gu D, Welch LR (2010) A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. *Plant Physiol* 153(2):485–513. <https://doi.org/10.1104/pp.110.156554>
10. Zang L, Zheng T, Chu Y, Ding C, Zhang W, Huang Q, Su X (2015) 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* 6:1140. <https://doi.org/10.3389/fpls.2015.01140>
11. Elkins T, Zinn K, McAllister L, Hoffmann FM, Goodman CS (1990) Genetic analysis of a drosophila neural cell adhesion molecule: interaction of fasciclin I and Abelson tyrosine kinase mutations. *Cell* 60(4):565–575. [https://doi.org/10.1016/0092-8674\(90\)90660-7](https://doi.org/10.1016/0092-8674(90)90660-7)
12. Huber O, Sumper M (1994) Algal-CAMs: isoforms of a cell adhesion molecule in embryos of the alga *Volvox* with homology to drosophila fasciclin I. *EMBO J* 13(18):4212–4222
13. Kawamoto T, Noshiro M, Shen M, Nakamasu K, Hashimoto K, Kawashima-Ohya Y, Gotoh O, Kato Y (1998) Structural and phylogenetic analyses of RGD-CAP/beta ig-h3, a fasciclin-like adhesion protein expressed in chick chondrocytes. *Biochim Biophys Acta* 1395(3):288–292. [https://doi.org/10.1016/s0167-4781\(97\)00172-3](https://doi.org/10.1016/s0167-4781(97)00172-3)
14. Tan L, Leykam JF, Kieliszewski MJ (2003) Glycosylation motifs that direct arabinogalactan addition to arabinogalactan-proteins. *Plant Physiol* 132(3):1362–1369. <https://doi.org/10.1104/pp.103.021766>
15. Huang GQ, Xu WL, Gong SY, Li B, Wang XL, Xu D, Li XB (2008) Characterization of 19 novel cotton FLA genes and their expression profiling in fiber development and in response to phytohormones and salt stress. *Physiol Plant* 134(2):348–359. <https://doi.org/10.1111/j.1399-3054.2008.01139.x>
16. Guerriero G, Mangeot-Peter L, Legay S, Behr M, Lutts S, Siddiqui KS, Hausman JF (2017) Identification of fasciclin-like arabinogalactan proteins in textile hemp (*Cannabis sativa* L.): in silico analyses and gene expression patterns in different tissues. *BMC Genom* 18(1):741. <https://doi.org/10.1186/s12864-017-3970-5>
17. MacMillan CP, Taylor L, Bi Y, Southerton SG, Evans R, Spokovic A (2015) The fasciclin-like arabinogalactan protein family of Eucalyptus grandis contains members that impact wood biology and biomechanics. *New Phytol* 206(4):1314–1327. <https://doi.org/10.1111/nph.13320>
18. Shi H, Kim Y, Guo Y, Stevenson B, Zhu JK (2003) The Arabidopsis SOS<sub>3</sub> locus encodes a putative cell surface adhesion protein and is required for normal cell expansion. *Plant Cell* 15(1):19–32. <https://doi.org/10.1105/tpc.007872>
19. Johnson KL, Kibble NA, Bacic A, Schultz CJ (2011) A fasciclin-like arabinogalactan-protein (FLA) mutant of Arabidopsis thaliana, flal, shows defects in shoot regeneration. *PLoS ONE* 6(9):e25154. <https://doi.org/10.1371/journal.pone.0025154>
20. MacMillan CP, Mansfield SD, Stachurski ZH, Evans R, Southerton SG (2010) Fasciclin-like arabinogalactan proteins: specialization for stem biomechanics and cell wall architecture in Arabidopsis and Eucalyptus. *Plant J* 62(4):689–703. <https://doi.org/10.1111/j.1365-3113.2010.04181.x>
21. Wang H, Jiang C, Wang C, Yang Y, Yang L, Gao X, Zhang H (2015) 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* 66(5):1291–1302. <https://doi.org/10.1093/jxb/eru479>
22. Huang GQ, Gong SY, Xu WL, Li W, Li P, Zhang CJ, Li DD, Zheng Y, Li FG, Li XB (2013) A fasciclin-like arabinogalactan protein, GhFLA1, is involved in fiber initiation and elongation of cotton. *Plant Physiol* 161(3):1278–1290. <https://doi.org/10.1104/pp.112.203760>
23. Roach MJ, Deyholos MK (2007) Microarray analysis of flax (*Linum usitatissimum* L.) stems identifies transcripts enriched in fibre-bearing phloem tissues. *Mol Genet Genom* 278(2):149–165. <https://doi.org/10.1007/s00438-007-0241-1>
24. Roach MJ, Deyholos MK (2008) Microarray analysis of developing flax hypocotyls identifies novel transcripts correlated with specific stages of phloem fibre differentiation. *Ann Bot* 102(3):317–330. <https://doi.org/10.1093/aob/mcn110>
25. Gorshkova T, Chernova T, Mokshina N, Gorshkov V, Kozlova L, Gorshkov O (2018) Transcriptome analysis of intrusively growing

- flax fibers isolated by laser microdissection. *Sci Rep* 8(1):14570. <https://doi.org/10.1038/s41598-018-32869-2>
26. Townsend T, Sette J (2016) Natural Fibres and the World Economy. In: Fangueiro R, Rana S (eds) Natural fibres: advances in science and technology towards industrial applications. RILEM bookseries, vol 12. Springer, Dordrecht. [https://doi.org/10.1007/978-94-017-7515-1\\_30](https://doi.org/10.1007/978-94-017-7515-1_30)
  27. Shafrin F, Ferdous AS, Sarkar SK, Ahmed R, Amin A, Hossain K, Sarker M, Rencoret J, Gutierrez A, Del Rio JC, Sanan-Mishra N, Khan H (2017) Modification of monolignol biosynthetic pathway in jute: different gene, different consequence. *Sci Rep* 7:39984. <https://doi.org/10.1038/srep39984>
  28. Choudhary SB, Kumar M, Chowdhury I, Singh RK, Pandey SP, Sharma HK, Karmakar PG (2016) An efficient and cost effective method of RNA extraction from mucilage, phenol and secondary metabolite rich bark tissue of tossa jute (*C. olitorius* L.) actively developing phloem fiber. *3 Biotech* 6(1):100. <https://doi.org/10.1007/s13205-016-0415-9>
  29. Tanmoy AM, Alum AM, Islam MS, Farzana T, Khan H (2015) Jute (*Corchorus olitorius* var. O-72) stem lignin: variation in content with age. *Bangladesh J Bot* 43(3):309–314. <https://doi.org/10.3329/bjb.v43i3.21603>
  30. del Rio JC, Rencoret J, Marques G, Li J, Gellerstedt G, Jimenez-Barbero J, Martinez AT, Gutierrez A (2009) Structural characterization of the lignin from jute (*Corchorus capsularis*) fibers. *J Agric Food Chem* 57(21):10271–10281. <https://doi.org/10.1021/jf900815x>
  31. Islam MS, Saito JA, Emdad EM, Ahmed B, Islam MM, Halim A, Hossen QM, Hossain MZ, Ahmed R, Hossain MS, Kabir SM, Khan MS, Khan MM, Hasan R, Aktar N, Honi U, Islam R, Rashid MM, Wan X, Hou S, Haque T, Azam MS, Moosa MM, Elias SM, Hasan AM, Mahmood N, Shafiuddin M, Shahid S, Shommu NS, Jahan S, Roy S, Chowdhury A, Akhand AI, Nisho GM, Uddin KS, Rabeya T, Hoque SM, Snigdha AR, Mortoza S, Matin SA, Islam MK, Lashkar MZ, Zaman M, Yuryev A, Uddin MK, Rahman MS, Haque MS, Alam MM, Khan H, Alam M (2017) Comparative genomics of two jute species and insight into fibre biogenesis. *Nat Plants* 3:16223. <https://doi.org/10.1038/nplants.2016.223>
  32. Chakraborty A, Sarkar D, Satya P, Karmakar PG, Singh NK (2015) Pathways associated with lignin biosynthesis in lignomaniac jute fibres. *Mol Genet Genom* 290(4):1523–1542. <https://doi.org/10.1007/s00438-015-1013-y>
  33. Zhang L, Ming R, Zhang J, Tao A, Fang P, Qi J (2015) De novo transcriptome sequence and identification of major bast-related genes involved in cellulose biosynthesis in jute (*Corchorus capsularis* L.). *BMC Genom* 16:1062. <https://doi.org/10.1186/s12864-015-2256-z>
  34. Gorshkova T, Sal'nikov V, Chemiksova S, Ageeva M, Pavlencheva N, Dam J (2003) The snap point: a transition point in *Linum usitatissimum* bast fiber development. *Ind Crops Prod* 18:213–221. [https://doi.org/10.1016/S0926-6690\(03\)00043-8](https://doi.org/10.1016/S0926-6690(03)00043-8)
  35. Esau K (1943) Vascular differentiation in the vegetative shoots of *Linum* III. The origin of the bast fibres. *Am J Bot* 30:579–586
  36. Ageeva MV, Petrovska B, Kieft H, Salnikov VV, Snegireva AV, van Dam JEG, van Veenendaal WLH, Emons AMC, Gorshkova TA, van Lammeren AAM (2005) Intrusive growth of flax phloem fibers is of intercalary type. *Planta* 222:565–574
  37. Gorshkova TA, Wyatt SE, Salnikov VV, Gibeau DM, Ibragimov MR, Lozovaya VV, Carpita NC (1996) Cell-wall polysaccharides of developing flax plants. *Plant Physiol* 110:721–729
  38. Gorshkova TA, Salnikov VV, Pogodina NM, Chemiksova SB, Yablokova EV, Ulanov AV, Ageeva MV, van Dam JEG, Lozovaya VV (2000) Composition and distribution of cell wall phenolic compounds in flax (*Linum usitatissimum* L.) stem tissues. *Ann Bot* 85:477–486
  39. Day A, Ruel K, Neutelings G, Cronier D, David H, Hawkins S, Chabbert B (2005) Lignification in the flax stem: evidence for an unusual lignin in bast fibers. *Planta* 222:234–245
  40. Esau K (1977) Anatomy of seed plants, 2nd edn. Wiley, London
  41. McDougall GJ (1991) Cell wall-associated peroxidases and lignification during growth of flax fibres. *J Plant Physiol* 39:182–186
  42. Kundu A, Sarkar D, Mandal NA, Sinha MK, Bs M (2011) A secondary phloic (bast) fibre-shy (bfs) mutant of dark jute (*Corchorus olitorius* L.) develops lignified fibre cells but is defective in cambial activity. *Plant Growth Regul* 67:45–55
  43. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics (Oxford, England)* 14(9):755–763. <https://doi.org/10.1093/bioinformatics/14.9.755>
  44. Letunic I, Doerks T, Bork P (2012) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res* 40:D302–305. <https://doi.org/10.1093/nar/gkr931>
  45. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A (2006) Pfam: clans, web tools and services. *Nucleic Acids Res* 34:D247–251. <https://doi.org/10.1093/nar/gkj149>
  46. Almagro Armenteros JJ, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 37(4):420–423. <https://doi.org/10.1038/s41587-019-0036-z>
  47. Eisenhaber B, Wildpaner M, Schultz CJ, Borner GH, Dupree P, Eisenhaber F (2003) Glycosylphosphatidylinositol lipid anchoring of plant proteins. Sensitive prediction from sequence- and genome-wide studies for Arabidopsis and rice. *Plant Physiol* 133(4):1691–1701. <https://doi.org/10.1104/pp.103.023580>
  48. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35(6):1547–1549. <https://doi.org/10.1093/molbev/msy096>
  49. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol Syst Biol* 7:539. <https://doi.org/10.1038/msb.2011.75>
  50. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic acids Res* 40:W597–603. <https://doi.org/10.1093/nar/gks400>
  51. Chou KC, Shen HB (2010) Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PLoS ONE* 5(6):e11335. <https://doi.org/10.1371/journal.pone.0011335>
  52. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G (2015) GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics (Oxford, England)* 31(8):1296–1297. <https://doi.org/10.1093/bioinformatics/btu817>
  53. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2:28–36
  54. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 36(10):3420–3435. <https://doi.org/10.1093/nar/gkn176>
  55. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico

- analysis of promoter sequences. *Nucleic Acids Res* 30(1):325–327. <https://doi.org/10.1093/nar/30.1.325>
56. Koziel SP (2010). Genetic analysis of lignification and secondary wall development in bast fibers of industrial Hemp (*Cannabis sativa*). Master's Thesis, University of Alberta, Edmonton, AB, Canada
57. Ahmed R, Hossain MS, Haque M, Alam M, Islam MS (2019) Modified protocol for RNA isolation from different parts of field-grown jute plant suitable for NGS data generation and quantitative real-time RT-PCR. *Afr J Biotechnol* 18:647–658. <https://doi.org/10.5897/AJB2019.16819>
58. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. *Methods (San Diego, Calif)* 25(4):402–408. <https://doi.org/10.1006/meth.2001.1262>
59. Hossain MS, Ahmed R, Haque MS, Alam MM, Islam MS (2019) Identification and validation of reference genes for real-time quantitative RT-PCR analysis in jute. *BMC Mol Biol* 20(1):13. <https://doi.org/10.1186/s12867-019-0130-2>
60. Kim JE, Kim SJ, Lee BH, Park RW, Kim KS, Kim IS (2000) Identification of motifs for cell adhesion within the repeated domains of transforming growth factor-beta-induced gene, betaig-h3. *J Biol Chem* 275(40):30907–30915. <https://doi.org/10.1074/jbc.M002752200>
61. Hobson N, Deyholos MK (2013) LuFLA1<sub>PRO</sub> and LuBGAL1<sub>PRO</sub> promote gene expression in the phloem fibres of flax (*Linum usitatissimum*). *Plant Cell Rep* 32:517–528. <https://doi.org/10.1007/s00299-013-1383-8>
62. Mokshina N, Chernova T, Galinousky D, Gorshkov O, Gorshkova T (2018) Key stages of fiber development as determinants of bast fiber yield and quality. *Fibers* 6:20. <https://doi.org/10.3390/fib6020020>
63. Snegireva A, Chernova T, Ageeva M, Lev-Yadun S, Gorshkova T (2015) Intrusive growth of primary and secondary phloem fibres in hemp stem determines fibre-bundle formation and structure. *AoB Plants*. <https://doi.org/10.1093/aobpla/plv061>
64. Ito S, Suzuki Y, Miyamoto K, Ueda J, Yamaguchi I (2005) AtFLA11, a fasciclin-like arabinogalactan-protein, specifically localized in sclerenchyma cells. *Biosci Biotechnol Biochem* 69(10):1963–1969. <https://doi.org/10.1271/bbb.69.1963>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.