The EuroBiotech Journal
SHORT COMMUNICATION
Bioinformatics / System Biology

Analysis of public RNA-seq data in studies of flax fiber biogenesis

Dmitry Galinousky and Tsimafei Padvitski

Abstract

In this work we used publicly available raw RNA-seq data to elucidate mechanisms of flax fiber biogenesis by measuring expression of cell-wall related genes (cellulose synthase, cellulose synthase-like and chitinase-like genes) in stem of flax (*Linum usitatissimum cv*. Bethune). Using public RNA-sequence data we have quantified and characterised the expression of the specific cell-wall genes in the top, middle and bottom parts of the Bethune flax stem. The most prominent findings are:

- Secondary cell-wall cellulose synthase (CesA) genes are expressed differentially in phloem and xylem in all parts of Bethune stem, in contrast with primary cell-wall cellulose synthase genes.
- Total expression level of cellulose synthase-like (Csl) genes is tissue invariant (although, *CslG* and *CslE* are differentially expressed) and smaller than the total expression of cellulose synthase genes. The *CslD2D3* subgroup dominates total expression of *CslD* genes in both xylem and phloem.
- Expression levels of all expressed chitinase-like (Ctl) genes are tissue dependent in all parts of stem. Total expression level of chitinase-like genes is much higher than expression of cellulose synthase and cellulose synthase-like genes in both tissues.

Laboratory of Ecological Genetics and Biotechnology, Institute of Genetics & Cytology, Belarus.

Corresponding author: Tsimafei Padvitski

E-mail: tsimafeipadvitski@gmail.com

Published online: 9 May 2017
doi:10.24190/ISSN2564-615X/2017/02.10

Introduction

The flax fiber classified as a bast fiber formed by cell wall deposition (1). One promising approache in the study of flax fibre biogenesis is transcriptomic research of genes involved in cell-wall biosynthesis. The major group of such genes includes cellulose-synthase, cellulose-synthase like and chitinase-like genes. In this study, we quantified the expression of *CesA*, *Csl* and *Ctl* genes in Bethune flax using publicly available raw RNA-seq data.

Materials and Methods

We used public data from the PRJNA251268 project that was performed under DOE Joint Genome Institute program and contains flax bast fiber and xylem profiling under normal conditions. The project comprises 18 RNA-seq experiments: transcriptomes of phloem and xylem from three different parts of stem (top, middle and bottom), each condition sequenced in triplicate. The flax strain that was used in the experiment is CDC Bethune, sequencing was performed on the Illumina hi-seq 2500 machine, paired-end library layout was used.

The analysis pipeline included 3 steps: data retrieval, pseudo alignment and transcripts quantification, and were carried out as described below:

1. Nucleotide sequences from samples were downloaded from SRA database in FASTA format using the fastq-dump program with the following parameters: --split-3 --fasta --skip-technical -I -W.

2. Pseudo-alignment tool kallisto was used for rapid, near-optimal RNA-Seq quantification of TPM values of the annotated genes (2). The following parameters were used -b 100 -t 4. Indexes were built based on the full set of *L. usitatissimum* coding sequences - Lusitatissimum_200_v1.0.cds.fa - that was downloaded from the JGI genome portal (5). This set

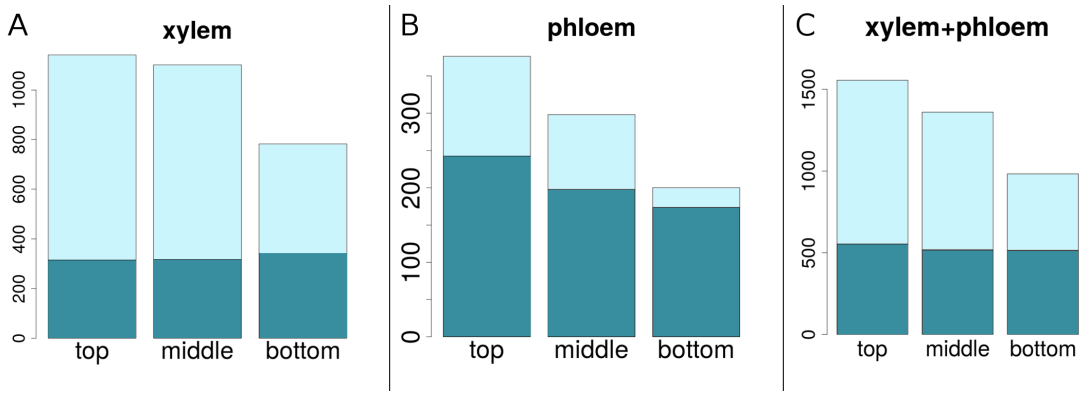

Figure 1. Total expression level (in TPM) of primary cell-wall genes ( ) and secondary cell-wall genes ( ) in the xylem (A), phloem (B) and in combination of the both tissues (C), measured in different parts of the flax stem.

was extended to include manually extracted coding sequences of two putative *CesA7* and one *CesA4* genes.

3. R package Sleuth (3) was used to normalize results of transcript abundance quantifications (performed by kallisto) across samples and for analysis of differential expression. Some exploratory analysis of transcript estimates and plotting was made in R.

Pseudo-alignment tool kallisto and R package Sleuth were used for the rapid quantification of TPM values (4) of all annotated CesA, Csl and Ctl genes, normalization of results across samples and for DE analysis.

Results

CesA-genes. Expression of at least two transcripts of each *CesA* gene has been detected in Betune's stem. The xylem transcription pool is dominated by the secondary cell wall (SCW) genes, among which *CesA4* gene transcripts are the most abundant. Conversely, expression of primary cell wall (PCW) genes dominate over SCW in the phloem, comprising 40-60% of the total pool. That being said, *CesA4* is relatively highly expressed in the phloem as well. In general, an expression level of the primary cell wall *CesA* genes is similar in the xylem and phloem, but the expression level of the secondary cell-wall *CesA*s is significantly higher (7-8 times) in the xylem (**Fig. 1 A,B**).

Total expression pool of *CesA* genes in the stem is dominated by the SCW genes (figure 1 A, B) and slightly decreases in direction top-middle-bottom mostly due to the decrease in SCW genes (**Fig. 1 C**).

Csl-genes. Many *Csl* genes are not expressed or expressed weakly in Betune's stem. For instance, expression of the only transcript of *CslB* family and of *CslD* subgroups such as *CslD1*, *CslD4*, *CslD6* was negligible. Total expression level of *Csl* is approximately equal in the phloem and xylem and smaller than the total expression of *CesA* genes (10x lower in the xylem and 2-5x lower in the phloem).

The major part of total *Csl* expression in xylem and phloem (~80% and ~60% respectively) constitutes expression of *CslD* genes (see figure 2 A, B), among which subgroup *CslD2D3* is an expression leader. Expression of *CslD* genes is largely tissue-invariant, while *CslB* and *CslG* genes show different expression patterns in phloem (figure B) and in xylem (figure A), where these genes are virtually not expressed. Interestingly, in contrast to *CesA* genes, total expression pool of *Csl* genes doesn't diminish but rather increases in stem in direction top-middle-bottom according to RNAseq data.

Ctl-genes. The majority of *Ctl* genes are not expressed or expressed weakly in Betune's stem in both tissues. In contrast,

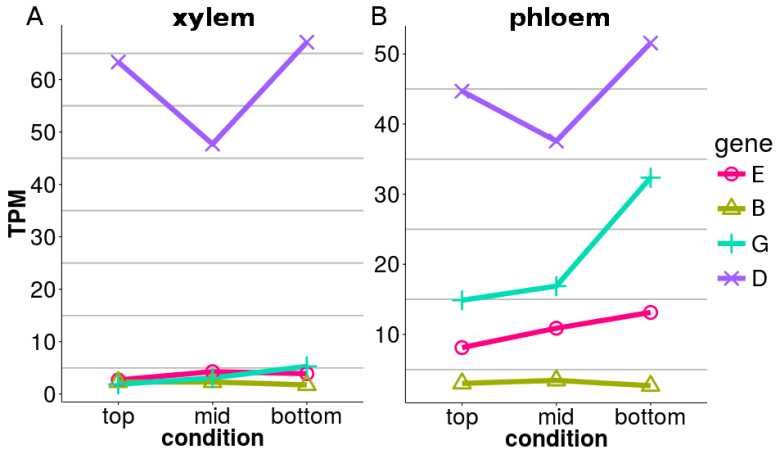

Figure 2. Summary expression of *CslB,E,G,D* groups of genes in xylem (A) and phloem (B) from different part of flax stem quantified from RNA-seq data.

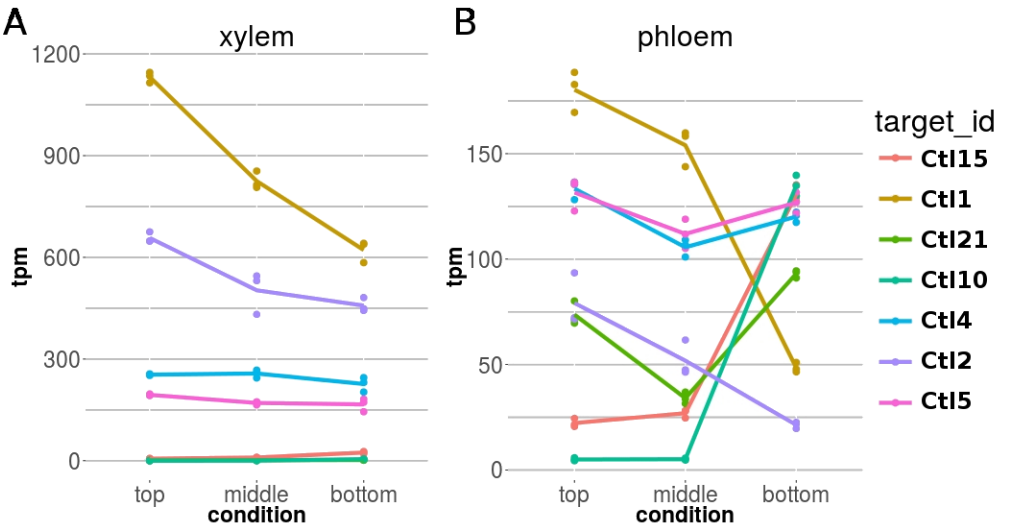

Figure 3. Expression of *Ctl* genes in xylem (A) and phloem (B) from different part of flax stem quantified from RNA-seq data.

several genes have very high level of expression: *Ctl1, Ctl2, Ctl4, Ctl5, Ctl10, Ctl15* and *Ctl21* (**Fig. 3**).

Expression levels of all expressed *Ctl* genes are significantly different in xylem (**Fig. 3A**) and phloem (**Fig. 3B**) and thus tissue-dependent. For instance, expression of genes *Ctl1, Ctl2, Ctl4, Ctl5* is an order of magnitude higher in the xylem than in phloem. Regarding the expression pattern on top-middle-bottom axis, it is tissue-invariant for some *Ctl* genes (e.g. *Ctl1,Ctl2*) and tissue-dependent for others (e.g. *Ctl10, Ctl15*).

Total expression level of *Ctls* is higher than expression of CesAs and of Csls genes in both phloem (1,7-4,2x and 7,5-9x, respectively) and xylem (1,7-2x and 20-30, respectively).

Discussion

We observed a tissue specific pattern of the secondary cell-wall *CesA*s gene expression with great overexpression of these genes in xylem, compared with phloem. This is a quite unexpected result, since the phloem tissue of flax is reported to be enriched in cellulose in comparison with the xylem (6). Thus, the observed phenomena should be studied in future researches.

Stable and relatively high level of *CslD2D3* subgroup expression in both tissues was shown in our analysis. The same results were obtained in our previous qPCR experiments on two different flax varieties that allows us to draw conclusion about robust expression of these genes across flax varieties. Another interesting observation that worth further investigation is tissue specificity of *CslG* and *CslE* genes, which are expressed only in xylem.

Expression levels of all expressed *Ctl* genes are very dependent on the tissue and in many cases also changes on the top-middle-bottom axis of stem; there are few Ctl genes that exhibit very high level of expression in some conditions. Interestingly, we also observed high expression levels and striking differences in expression of some *Ctl* genes in our earlier qPCR experiments where two varieties of flax were compared. This may point to the involvement of these gene in the dynamic regulation of cell-wall biosynthesis and organisation in flax fiber.

All candidate genes should be further studied in other varieties of flax, including fiber-producing ones, using a quantitative approach to advance our understanding of the flax fiber formation process.

Acknowledgment

The work was supported by the National Academy of Sciences of Belarus, Agreement B15UK/A-041 and by the Belarusian Republican Foundation for Fundamental Research (BRFFR), Agreement B15M-101 and B15-147.

Conflict of Interest Statement

The authors declare the absence of any conflict of interest.

References

1. Gorshkova T, Brutch N, Chabbert B, Deyholos M, Hayashi T, Lev-Yadun S, Mellerowicz EJ, Morvan C, Neutelings G, Pilate G. Plant Fiber Formation: State of the Art, Recent and Expected Progress, and Open Questions. Critical Reviews In Plant Sciences 2012; 31 (3): 201-228.
2. Bray NL, Pimentel H, Melsted P, Pachter l. Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology 2016; 34: 525–527.
3. Pimentel HJ, Bray N, Puente S, Melsted P, Pachter L. Differential analysis of RNA-Seq incorporating quantification uncertainty, bioRxiv 2016; 058164.
4. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci 2012; 131: 281.
5. Wang ZW, Hobson N, Galindo L, Zhu SL, Shi DH, McDill J, et al. The genome of flax (*Linum usitatissimum*) assembled *de novo* from short shotgun sequence reads. Plant J 2012; 72: 461–473.
6. Mikshina P, Chernova T, Chemikosova S, Ibragimova N, Mokshina N, Gorshkova T. Cellulosic Fibers: Role of Matrix Polysaccharides in Structure and Function, Cellulose - Fundamental Aspects, Dr. Theo G.M. Van De Ven (Ed.), 2013; InTech.