Published with Open Access at Journal BiNET
Vol. 26, Issue 02: 2204-2210
Journal of Bioscience and Agriculture Research
Journal Home: www.journalbinet.com/jbar-journal.html

Whole transcriptome sequencing and analysis of jute (*Corchorus olitorius*) fiber cell

Rasel Ahmed[1], Md. Sabbir Hossain[1], Shah Md. Tamim Kabir[1], Borhan Ahmed[1], Rajnee Hasan[1], Mohammad Saiful Alam Sarker[1], Md. Zablul Tareq[2], Emdadul Mannan Emdad[1] and Md. Shahidul Islam[1]

[1]Basic and Applied Research on Jute Project, Bangladesh Jute Research Institute, Dhaka, Bangladesh
[2]Jute Agriculture Experimental Station, Bangladesh Jute Research Institute, Jagir, Manikganj, Bangladesh

✉ For any information: rahmed.bg@gmail.com (Ahmed, R)
Article received: 15.10.2020; Revised: 30.11.2020; First published online: 10 December 2020.

ABSTRACT

The demand for products made by jute fiber is increasing day-by-day for its biodegradable nature regarding environmental concerns. To gain this opportunity correctly, the development of high yielding and improved fiber quality jute variety is essential for ensuring diversified use of jute fiber. The major developed jute varieties, so far, are the outcome of conventional breeding which is a very time consuming process. Improvement of fiber quality and yield through genetic modification approach is highly desired. However, very little is known about the molecular mechanism behind fiber cell formation in jute. Here, we attempted to do the whole transcriptome sequencing of fiber cell RNA to reveal the molecular mechanisms were happening in the premises of jute fiber cells. We performed RNA isolation from jute fiber cells followed by whole transcriptome sequencing. De novo assembly of sequencing reads resulted in 21,294 contigs representing the transcriptome size of 4.07 Mbp. Gene ontology analysis assigned 14144 genes (52.21%) for biological process, 8399 genes (31%) involved in molecular function and 4549 genes (16.79%) for cellular component. Total 66 fiber related genes were found from reference based annotation where 9 genes involved in fiber cell initiation and elongation and the rest 57 for secondary cell wall development. We presented the overall view of the jute fiber cell transcriptome in this study. These findings help for understanding the molecular basis of fiber formation in jute plant.

Key Words*: Jute, Fiber cell, Transcriptome, Sequencing and De novo assembly*

Cite Article: Ahmed, R., Hossain, M. S., Kabir, S. M. T., Ahmed, B., Hasan, R., Sarker, M. S. A., Tareq, M. Z., Emdad, E. M. and Islam, M. S. (2020). Whole transcriptome sequencing and analysis of jute (*Corchorus olitorius*) fiber cell. Journal of Bioscience and Agriculture Research, 26(02), 2204-2210.
Crossref: https://doi.org/10.18801/jbar.260220.269

I. Introduction

Jute has played a historic role as a golden fiber in the socio-economic development of Bangladesh. Jute is the second largest source of natural fiber next to cotton. Unfortunately, due to the aggression of

Published with open access at journalbinet.com
EISSN: 2312-7945, © 2020 The Authors, Research paper

synthetic commodities, it lost its demand and value in the global market. But, now-a-days, with concerns of environmental issues the need for jute-fiber made products is increasing for its biodegradability. To gain this opportunity correctly, development of high yielding and improved fiber quality jute variety is essential for ensuring diversified use of jute fiber.

Jute fiber is composed of sclerenchyma cells which give mechanical support to the plants. Fibers are cambial derivatives originated from primary growth via the procambium and occurring as bundles in secondary phloem (Esau, 1965; Fahn, 1990) which are intermingled with medullary ray cells and other soft tissues as triangular wedges (Kundu et al., 1959). After initiation of fiber cell, it goes to elongation process i.e. combination of coordinated growth with other cells and intrusive growth between the neighboring cells (Schoch-Bodmer and Huber, 1951). Before completion of the elongation process, secondary cell wall thickening is started and maturity comes through lignin deposition which is one of the critical determinants of bast fiber development in jute (Esau 1977; Kundu et al., 2011).

So far, the major developed jute varieties are the outcome of conventional breeding which is a very time consuming process. Improvement of fiber quality and yield through genetic modification approach is highly desired. But, very little is known about the molecular mechanism behind fiber cell formation in jute (Chakraborty et al., 2015). Bangladesh Jute Research Institute has already decoded the jute genome and predicted more than 300 genes related to fiber bio-synthesis pathway (Islam et al., 2017). But we don't know how those genes are interacting with each other or what the fates of those genes in fiber cell environment. In these circumstances, we attempted to do the whole transcriptome sequencing of fiber cell RNA to decipher the code of molecular events happening in jute fiber cells.

II. Materials and Methods

Plant materials, total RNA isolation and sequencing

Seeds of O-4 (*Corchorus olitorius*) cultivar of jute were sown at the field of jute agricultural experiment station in Manikganj. After 45 days of growth, stems of about 15cm in length spanning from 5 cm to 7 cm below the top of the plant were cut off. Bark was separated from the middle stick of collected stem by hand. Green epidermal layer other than fiber strip was removed from the bark by gentle rubbing with a spatula. Immediately after rubbing the fiber strip was fixed in liquid nitrogen. Fiber strips collected from 50 plants of both species are pooled separately. Fixed fiber strips were lyophilized (freeze drying) for 2 hours. After completion of freeze drying, fiber strip was washed in 80% ethanol with hard pressing by pestle in the mortar at least 3-4 times until individual fiber thread was separated from strips. The mass of individual fiber threads was again washed in 80% ethanol with vigorous jerking in glass bottle. Finally washed fiber threads were examined under a light microscope to check the purity of separated fiber cells. After confirmation of purity, RNA was isolated from the fiber cells following the guanidine thiocyanate (GNTC) method (Chomczynski and Sacchi, 1987) with modification. RNA-seq library was prepared using Ion Total RNA-seq kit and then library was sequenced in next generation Ion Proton sequencing platform according to manufacturer's instructions (Life Technologies, USA).

Pre-processing of raw Reads and mapping

Pre-processing of reads involved quality trimming, removal of adapter, primer sequence and duplicate reads using in-house perl script, cutadapt program and Fastx toolkit. The Bowtie mapper was used to map reads against genome data of jute plant to valid the reads.

De Novo transcriptome assembly

To assemble the RNA-seq reads into contigs the Trinity software package (version trinity rnaseq_r20140413) was applied with default parameters and minimum contig length of ≥ 50 bp.

Functional annotation of raw reads and assembled contigs

To identify a putative function, raw reads and the contigs obtained from trinity assembly were blasted against annotated protein sequence of *C. olitorius* using the Basic Local Alignment Search Tool (BLAST) (Altschul, 1990) for nucleotide (BLASTN). Each of reads and contigs showing significant similarity (E- value ≤ 1e-30, non-overlap and align coverage more than 30%) was assigned the same

Published with open access at journalbinet.com
EISSN: 2312-7945, ©2020 The Authors, Research paper

putative function as that of corresponding *C. olitorius* protein. The protein sequences were compared to NCBI non-redundant (nr) database using protein BLAST (BLASTP), the output of BLASTP in xml format was used in Blast2GO program (Conesa et al., 2005) to assign gene ontology (GO) terms for biological process, molecular function and cellular components. The predicted genes were analyzed for functional domains by using InterProScan (Zdobnov and Apweiler, 2001).

III. Results

RNA isolation, sequencing and mapping

Total RNA was isolated from separated fiber cells and confirmed in gel electrophoresis (Figure 01). In the present study, the transcriptome of fiber RNA from *C. olitorius*, a cultivated jute variety was sequenced using Ion Proton Platform. It generated more than 63 million reads with an average length of 101 bp covering a total of 6.4 Gb (Giga base pairs) data. These reads were filtered by in-house script, cut adapt and fastx toolkits to remove adapter, duplicate sequence and low quality read to provide good quality reads for assembly. More than 16 million reads were obtained after quality trimming of the raw reads with an average length of 110 bp. These high quality reads were mapped with Bowtie2 to the reference jute (*C. olitorius*) genome sequences (Islam et al., 2017) and obtained 94.32% aligned reads. The summary of sequencing data generated from fiber cells and their quality trimming is given in Table 01.

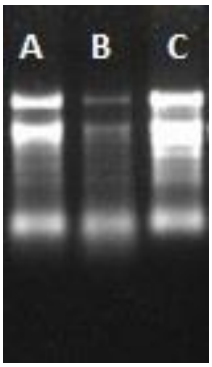

Figure 01. Electrophoresis of isolated RNA in agarose gel. Lane A and B for RNA isolated from *C. olitorius* fiber cell. Lane C for control RNA isolated from 3 days old *C. olitorius* seedling

Table 01. Overview of the sequencing information

Parameters	Raw Data	Data after Trimming
No. of reads	63,321,934	16,813,592
Total Bases (bp)	6,402,674,226	1,852,652,917
GC content (%)	~44	~42
Minimum read (bp)	8	16
Maximum read (bp)	364	364
Average read length (bp)	101.11	110.19

De Novo assembly

The assembly is one of the essential steps of next generation sequencing (NGS) data analysis for downstream analysis. The filtered data were assembled by using the Trinity software package (Grabherr et al., 2011) and the resulting fiber cell transcriptome, with 21,294 contigs, N50 of 189 bp and average length of 191 bp, was used for further downstream analysis (Table 02).

Table 02. Summary of assembly results

Parameters	Trinity assembly metrics
No. of contigs	21,294
Total Bases (bp)	4,073,868
GC content (%)	~44%
Minimum contig length (bp)	101
Maximum contig length (bp)	2,906
Average contig length (bp)	191.32
N50 (bp)	189

Published with open access at journalbinet.com
EISSN: 2312-7945, © 2020 The Authors, Research paper

Functional annotation of raw reads and assembled contigs

We performed BLAST for nucleotide searches of reads and contigs against reference jute mRNA database using E-value 1e−30, non-overlap and align coverage at least 30% for annotation. Of all the reads, a total 11,561 have found hits against the reference database but only 4,920 (42.56%) reads that followed the above criteria in nucleotide BLAST (BLASTN) process, were taken to further analysis. Similarly, among 21,294 contigs, 3,226 (15.15%) showed significant homology to known proteins in the reference database. We also compared blast hits from reads and conigs and found 3,017 genes in common. In addition, InterProScan determined motifs and domains of genes against protein databases. InterPro (IPR) domains are presented in Table 03. To identify the major gene groups Gene Ontology (GO) assignments systems were used and assigned genes into three main categories resulted in 14144 genes (52.21%) for biological process, 8399 genes (31%) involve in molecular function and 4549 genes (16.79%) for cellular component (Figure 02). Total 66 fibers related genes were found from reference based annotation where 9 genes involved in fiber cell initiation and elongation and the rest 57 for secondary cell wall development (Table 04)

Table 03. InterPro domains within the *C. olitorius* fiber cell transcriptome. Only the most abundant domains (≥30 hits) are listed

IPR_ID	IPR_Name	Gene Count
IPR000719	Protein kinase, catalytic domain	278
IPR000504	RNA recognition motif domain	127
IPR001245	Serine-threonine/tyrosine-protein kinase catalytic domain	125
IPR001841	Zinc finger, RING-type	120
IPR001680	WD40 repeat	107
IPR001810	F-box domain, cyclin-like	88
IPR002885	Pentatricopeptide repeat	81
IPR001005	SANT/Myb domain	73
IPR001650	Helicase, C-terminal	53
IPR001128	Cytochrome P450	50
IPR001611	Leucine-rich repeat	49
IPR001932	Protein phosphatase 2C (PP2C)-like	47
IPR000225	Armadillo	46
IPR000008	C2 calcium-dependent membrane targeting	45
IPR002048	Calcium-binding EF-hand	44
IPR001461	Peptidase A1	42
IPR001806	Small GTPase superfamily	38
IPR011598	Myc-type, basic helix-loop-helix (bHLH) domain	37
IPR011990	Tetratricopeptide-like helical	37
IPR003593	AAA+ ATPase domain	35
IPR000767	Disease resistance protein	35
IPR001623	DnaJ domain	35
IPR000620	Drug/metabolite transporter	35
IPR000608	Ubiquitin-conjugating enzyme, E2	35
IPR003690	Mitochodrial transcription termination factor-related	34
IPR005123	Oxoglutarate/iron-dependent dioxygenase	33
IPR002198	Short-chain dehydrogenase/reductase SDR	33
IPR011989	Armadillo-like helical	31
IPR000109	Proton-dependent oligopeptide transporter family	31

Table 04. Comparison of fiber related genes between fiber cell and reference seedling stranscriptome

Sample	Initiation and Elongation	Secondary Cell wall development	Total
Fiber cell	9	57	66
Seedlings	114	191	305

Published with open access at journalbinet.com
EISSN: 2312-7945, ©2020 The Authors, Research paper

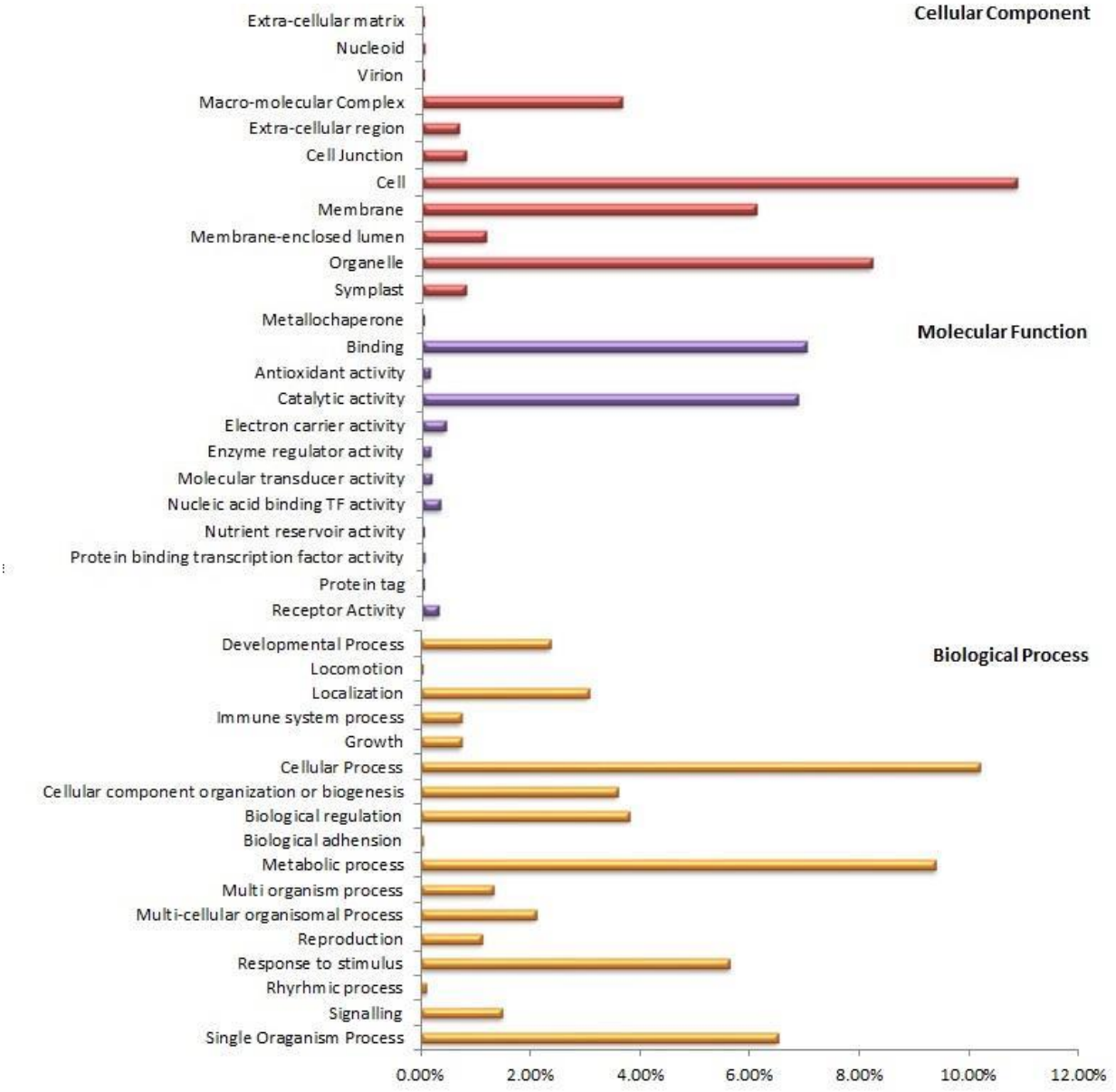

Figure 02. Histogram presentation of GO classification of fiber cell transcriptome (The results are summarized in three main groups: Cellular component, Molecular function and Biological process. The Y-axis indicates the percent of genes in a category)

IV. Discussion

In this study, we presented a comprehensive view of the de novo assembled jute fiber cell transcriptome and identified gene categories. For transcriptome sequencing, RNA isolation is a crucial part and it is necessary to isolate good quality RNA with sufficient quantity for NGS. In this case, our isolated RNA from fiber cell was less in amount per gram of fiber cell. One of the reasons behind that may be the mechanical procedure of isolating fiber cells where cells undergo huge scorching and washing that leads to RNA degradation. Besides that, this isolation process takes too long time resulting in low quality RNA. As during the maturation process of fiber cell, the protoplast becomes progressively dying (Esau, 1977), we hardly expect the whole intact mRNA from fiber cell. Moreover, it is quite impossible to isolate fiber cells at the immature stage (Gorshkova et al., 2012).

About 6.4 Gbp raw sequence was generated, which is lower in comparison to reference seedling transcriptome (approximately 10 Gb) from ion proton platform. In the data filtering process, a considerable number of duplicates (more than 73%) were found in the raw reads. De novo assembly from filtered reads was performed by Trinity and resulted in 21,294 contigs representing transcriptome size of 4.07 Mb. The assembly produced matrix such as N50, average contig length, consensus size etc. were presented in Table 02 which suggested the challenges we faced during

Published with open access at journalbinet.com
EISSN: 2312-7945, © 2020 The Authors, Research paper

downstream analysis. As the prediction of full length gene is difficult from such kind of fragmented data, we went for gene prediction from raw reads against reference genome. In addition, we also predicted the genes from trinity generated contig file. From raw reads, we got 4,920 predicted genes whereas 3,226 genes were found in trinity assembly file. It is important to be noted that 5.21% of raw reads were unassembled in trinity. Therefore, we got less number of genes in the assembled file compared to raw reads. The predicted genes from raw reads of fiber cellsare only 15% of the total gene (32,594) predicted from reference seedling transcriptome whereas it is only 13.29% compared to the genes (37,031) predicted from the reference genome. As mentioned earlier that, during maturation process of fiber cell, the protoplast becomes progressively dying and an internal lumen develops (Esau, 1977), RNA is being fragmented and degraded. It could be the core reason to get less number of genes in fiber RNA.

After identifying genes in fiber cell transcriptome, we classified the raw read based annotated genes under three categories of gene ontology term (GO) and compared this GO classification with that of reference seedlings transcriptome. Among three main GO categories, the 'cell', 'organelle', 'binding', 'catalytic activity', 'cellular process' and 'metabolic process' were dominant (Figure 02). We have found 54.5% genes in biological process, 19.4% in molecular process and 25.9% in cellular component in case of genes predicted in reference seedlings transcriptome. In fiber cells, genes under biological process (52.21%) and molecular function (16.79%) were found less compared to seeding transcriptome. In contrast, fiber cell contains more than 5% transcripts than seedlings in cellular component category. Let's look at the physiology of fiber cell which is undergoing the process of maturity. We can see that in this stage, secondary cell wall deposition is happening followed by creation of lumen. Those processes are likely under the cellular component category and this fraction contained 31% genes in fiber cell RNA.

In reference transcriptome, 305 genes were predicted to be associated with fiber biosynthesis pathway (Islam et al., 2017). Among them, 114 genes are involved in fiber cell initiation and elongation and the rest 191 genes for secondary cell wall development. In contrast, total 66 genes are found as fiber genes in fiber cell RNA; 9 for initiation and elongation, and 57 for secondary cell wall development. In fiber cells, about 16% of fiber genes are involved in initiation and elongation and in case of seedlings, it is more than 37%. On the other hand, most of the fiber genes are expressed in favor of secondary cell wall development (84.21%) which is higher compared to seedlings RNA (62.6%). As we isolated RNA from the fiber cell during maturation stage, it was expected that most of the genes would be responsible for secondary cell wall development rather than initiation and elongation. The above analysis also supports this prediction.

V. Conclusion

In this study, we presented the overall view of the jute fiber cell transcriptome in terms of the genes expressed in fiber cells. Total 66 fiber related genes were found where 9 genes are involved in fiber cell initiation & elongation process and 57 genes are responsible for secondary cell wall development. This work helps our understanding of the molecular basis of fiber formation in jute plant. However, to get a better experience and reveal concrete molecular mechanism in fiber cell biosynthesis, it is necessary to do comparative transcriptome analysis between fiber and surrounding non-fiber cells. Also, gene expression analysis from fiber and non-fiber cells and even between different stages in fiber cell development is required to get the complete scenario behind fiber biosynthesis process.

VI. References

[1]. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

[2]. Chakraborty, A., Sarkar, D., Satya, P., Karmakar, P., Singh, N. (2015) Pathways associated with lignin biosynthesis in lignomaniac jute fibres. Mol Genet Genomics, 290, 1–20.

[3]. Chomczynski, P. and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidiniumthiocyanate-phenol-chloroform extraction. Analytical Biochemistry, 162(1), 156–159. https://doi.org/10.1006/abio.1987.9999

Published with open access at journalbinet.com
EISSN: 2312-7945, ©2020 The Authors, Research paper

[4]. Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO, a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics, 21(18), 3674–3676. https://doi.org/10.1093/bioinformatics/bti610
[5]. Esau, K. (1965). Plant Anatomy (2nd Ed.). John Wiley and Sons Inc., New York.
[6]. Esau, K. (1977). Anatomy of Seed Plants (2nded.). John Wiley, London, UK. https://doi.org/10.2307/2418500
[7]. Fahn, A. (1990). Plant Anatomy (4th ed.). Pergamon Press, Oxford, UK.
[8]. Gorshkova, T. A., Brutch, N., Chabbert, B., Deyholos, M., Hayashi, T., Lev-Yadun, S., Mellerowicz, E. J., Morvan, C., Neutelings, G. and Pilate, G. (2012). Plant Fiber Formation: State of the Art, Recent and Expected Progress, and Open Questions. Critical Reviews in Plant Sciences, 31, 201–228. https://doi.org/10.1080/07352689.2011.616096
[9]. Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. and Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology, 29(7), 644–652. https://doi.org/10.1038/nbt.1883
[10]. Islam, M. S., Saito, J. A., Emdad, E. M., Ahmed, B., Islam, M. M., Halim, A., Hossen, Q. M., Hossain, M. Z., Ahmed, R., Hossain, M. S., Kabir, S. M., Khan, M. S., Khan, M. M., Hasan, R., Aktar, N., Honi, U., Islam, R., Rashid, M. M., Wan, X., Hou, S. and Alam, M. (2017). Comparative genomics of two jute species and insight into fibre biogenesis. Nature plants, 3, 16223. https://doi.org/10.1038/nplants.2016.223
[11]. Kundu, A., Sarkar, D., Mandal, N. A., Sinha, M. K. and Mahapatra, B. S. (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. https://doi.org/10.1007/s10725-012-9660-z
[12]. Kundu, B. C., Basak, K. C. and Sarkar, P. B. (1959). Jute in India. The Indian Central Jute Committee, Calcutta. P. 55.
[13]. Schoch-Bodmer, H. and Huber, P. (1951). Das Spitzewachstum der Bastfasernbei Linumusitatissimum und Linumperenne. BerSchweiz Bot Ges, 61, 377-404.
[14]. Zdobnov, E. M. and Apweiler, R. (2001). InterProScan - an integration platform for the signature-recognition methods in InterPro. Bioinformatics, 17, 847-848. https://doi.org/10.1093/bioinformatics/17.9.847

HOW TO CITE THIS ARTICLE?

Crossref: https://doi.org/10.18801/jbar.260220.269

MLA
Ahmed, R. et al. "Whole transcriptome sequencing and analysis of jute (*Corchorus olitorius*) fiber cell". Journal of Bioscience and Agriculture Research, 26(02), (2020):2204-2210.

APA
Ahmed, R., Hossain, M. S., Kabir, S. M. T., Ahmed, B., Hasan, R., Sarker, M. S. A., Tareq, M. Z., Emdad, E. M. and Islam, M. S. (2020). Whole transcriptome sequencing and analysis of jute (*Corchorus olitorius*) fiber cell.*Journal of Bioscience and Agriculture Research*, 26(02), 2204-2210.

Chicago
Ahmed, R., Hossain, M. S., Kabir, S. M. T., Ahmed, B., Hasan, R., Sarker, M. S. A., Tareq, M. Z., Emdad, E. M. and Islam, M. S. "Whole transcriptome sequencing and analysis of jute (*Corchorus olitorius*) fiber cell". Journal of Bioscience and Agriculture Research, 26(02), (2020): 2204-2210.

Harvard
Ahmed, R., Hossain, M. S., Kabir, S. M. T., Ahmed, B., Hasan, R., Sarker, M. S. A., Tareq, M. Z., Emdad, E. M. and Islam, M. S. 2020. Whole transcriptome sequencing and analysis of jute (*Corchorus olitorius*) fiber cell. Journal of Bioscience and Agriculture Research, 26(02), pp. 2204-2210.

Vancouver
Ahmed R, Hossain MS, Kabir SMT, Ahmed B, Hasan R, Sarker MSA, Tareq MZ, Emdad EM and Islam MS. Whole transcriptome sequencing and analysis of jute (*Corchorus olitorius*) fiber cell. Journal of Bioscience and Agriculture Research, 2020 December 26(02): 2204-2210.

Published with open access at journalbinet.com
EISSN: 2312-7945,