Published with Open Access at Journal BiNET
Vol. 26, Issue 02: 2185-2191
Journal of Bioscience and Agriculture Research
Journal Home: www.journalbinet.com/jbar-journal.html

Phenylalanine ammonia-lyase gene family (PAL): Genome wide characterization and transcriptional expression in jute (*Corchorus olitorius*)

Md. Sabbir Hossain[1], Rasel Ahmed[1], Md. Wali Ullah[1], Ummay Honi[1], Md. Zablul Tareq[1], Mohammad Saiful Alam Sarker[1], Borhan Ahmed[1,2] and Md. Shahidul Islam[1,2]

[1]Basic and Applied Research on Jute Project, Bangladesh Jute Research Institute, Dhaka, Bangladesh
[2]Bangladesh Jute Research Institute, Dhaka, Bangladesh

✉ For any information: sabbirbge@gmail.com (Hossain, MS)
Article received: 19.10.2020; Revised: 07.12.2020; First published online: 10 December 2020.

ABSTRACT

Jute is one of the important ligno-cellulose bast fiber crops next to cotton. High lignin content in jute fiber makes hindrances during spinning in the textile industry. Phenylalanine ammonia-lyase (PAL), encoded by multigene family, is the first enzyme in the phenylpropanoid pathway which involved in biosynthesis of different secondary metabolites including lignin. A total of 4 PAL genes were identified in jute (Corchorus. olitorius) genome which was being distributed in two chromosomes and clustered into three subfamilies based on phylogenetic analysis. Like PAL genes in other species, CoPALs had similar molecular properties and structure organizations. Expression analysis revealed that CoPAL1 and CoPAl2 were differentially expressed in various jute tissues. Among them, CoPAL1 was predominately expressed in stem tissues suggesting its involvement in lignin accumulation in fiber and can act as a potential target for reducing lignin in jute. Our study provides useful information for future functional characterization of PAL genes in jute.

Key Words*: Jute, PAL gene, Genome wide analysis, Gene structure and Gene expression.*

Cite Article: Hossain, M. S., Ahmed, R., Ullah, M. W., Honi, U., Tareq, M. Z., Sarker, M. S. A., Ahmed, B. and Islam, M. S. (2020). Phenylalanine ammonia-lyase gene family (PAL): Genome wide characterization and transcriptional expression in jute (*Corchorus olitorius*). Journal of Bioscience and Agriculture Research, 26(02), 2185-2191. **Crossref:** https://doi.org/10.18801/jbar.260220.267

I. Introduction

The phenylpropanoid pathway is one of the most essential pathways which is required for biosynthesis of a wide range of secondary metabolites including flavonoids, isoflavonoids, anthocyanins, plant hormones, phytoalexins, and lignins (Fraser and Chapple, 2011). These metabolites are associated with various plant functions such as growth, development and adaptation (Dong et al., 2016). Phenylalanine ammonia-lyase (PAL) is the first enzyme that is widely present in plants, in some fungi and yeast (Yan et al., 2019). PAL catalyzes phenylalanine to cinnamic acid and synthesis of secondary metabolites depends on PAL activity as it connects primary metabolism with secondary metabolism (Dong et al., 2016). Therefore, many studies have been conducted to identify the PAL genes in different plant species

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

such as Arabidopsis, poplar, pear, eucalyptus, tomato and potato, etc. (Dong et al., 2016; Yan et al., 2019). Moreover, the functional characterization of PAL genes in some plants was also conducted to know their involvement in plant growth and development. For example, *AtPAL1* and 2 were predominately expressed in Arabidopsis stem tissue and their down regulation significantly reduce lignin content. Moreover, silencing of PAL gene in tobacco leads to alter flower pigmentation and diminish plant growth (Yoon et al., 2015). PAL genes were also involved in flower and fruit development in raspberry, *EgrPALs* were associated with lignification and flower development, and higher expression patterns of *PbPALs* were observed in pear flower, stem and fruit (Dong et al., 2016; Carocha et al., 2015; Cao and Li, 2019). Therefore, PAL genes get more attention from researchers for future breeding programs.

Jute (*Corchorus spp*.) is an important fiber crop which demands are continuously increasing globally due to its annually renewable and biodegradable nature, lower cost in production and broad-spectrum applications (Hossain et al., 2019). However, short fiber length and higher lignin content in jute make it difficult to use in the textile industry during spinning process (Hossain et al., 2020). Therefore, it is important to investigate the lignin biosynthesis process for dropping lignin content in jute fiber. Recently, the availability of genome sequence of jute (Islam et al., 2017) allows us to identify the genes associated with lignin biosynthesis. Although genome wide characterization of PAL genes has been conducted in many plant species, no systematic study has been found in bast fiber crops so far. Therefore, in this study, we focused on the identification of PAL genes in jute and their molecular characterization through bioinformatics approaches. Moreover, available transcriptome data allow us to analyze gene expression to know the function of PAL genes in jute plant development, especially in lignin biosynthesis. Our results will serve as a solid foundation for further characterization of PAL genes in jute and as well as in the future breeding program.

II. Materials and Methods

Genomic data mining and PAL gene identification

To identify the PALs protein sequence in jute (*C. olitorius*), PALs protein sequence of *Arabidopsis thaliana* were chosen as query sequences. The Arabidopsis and jute data were downloaded from TAIR (http://www.arabidopsis.org) and National Center for Biotechnology Information (NCBI, Accession ID PRJNA215141) database (Islam et al., 2017), respectively. The BLAST tool was used for the identification of PAL homolog in jute with a cut-off e-value of e^{-10}. The amino acid sequence of candidate genes of *CoPAL* was further verified to confirm the presence of conserved domains Aromatic amino acid lyase (PF00221) using the Pfam database (https://pfam.xfam.org/).

Phylogenetic analysis naming of *CoPAL* genes

To understand the evolutionary relations among the PAL genes, a phylogenetic tree was constructed using MEGA X (Kumar et al., 2018). Initially, the amino acids of PAL genes from *C. olitorius*, Arabidopsis, *Eucalyptus grandis* and *Populus trichocarpa* (Cao and Li, 2019; Carocha et al., 2015) were aligned by ClustalW and then multiple aligned file was applied to MEGA X to build a phylogeny tree using neighbor-joining (NJ) method with bootstrap value of 1000 replications. The name *CoPAL* genes were assigned based on the phylogenetic tree and the homology of *AtPAL*.

Molecular characteristics and structural analysis of *CoPALs*

The web-based tool Prot Param (http://expasy.org/tools/protparam.html) was used to predict the molecular characteristics including protein length, isoelectric point (pI), grand average of hydropathicity (GRAVY) and molecular weight (MW) of *CoPALs*. The exon/intron structures were illustrated by a web application named Gene Structure Display Server using GSDS (http://gsds.cbi.pku.edu.cn/). Multiple expectation maximization for motif elicitation (MEME) (http://meme.sdsc.edu/meme/meme.html) was used to identify the conserved motifs of *CoPALs*.

Expression profiling of *CoPAL* genes

Transcriptome data of different tissues in various growth stages of *C. olitorius* were downloaded from NCBI under accession numbers PRJNA597180 (Yang et al., 2020) Y and PRJNA215141 (Islam et al., 2017) were used for expression analysis. These transcriptome data were generated from 3 days old seedlings (SD), leaf vegetative growth period (LVGP), leaf flowering period (LFP), stem vegetative growth period (SVGP), bast vegetative growth period (BVGP), bast flowering period (BFP), bast

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

technical mature period (BTMP), mature flower (MF), fruit 1–2 cm in length (FT1), fruit 2-4 cm in length (FT2) and greater than 4 cm in length (FT3). All SRA data were converted to FASTQ file using FastQC v.0.11.9 followed by filtered low quality reads through Trimmomatic *v.*0.36. Then high quality read pairs were mapped to *C. olitorius* reference genome with TopHat (version 2.1.1). Cuffdiff of Cufflinks2 v.2.1.1 suite (Trapnell et al., 2012) was applied to quantify gene expression using mapped bam file. The cuffdiff calculated gene expression based on FPKM values (fragment per kilobase per million mapped reads).

III. Results and Discussion

Identification and phylogenetic analysis *CoPALs*

We have identified 4 *CoPALs* genes in *C. olitorius* genome using BLASTP and all *CoPALs* contained PAL domain namely Aromatic amino acid lyase (PF00221). The number of *CoPLAs* was comparable to the number of PAL genes in other species such as 4 in Arabidopsis and *Medicago truncatula,* six in *Populus trichocarpa* and eight in *Glycin max* (Carocha et al., 2015; Dong et al., 2016*).* However, some studies have reported that some species contain many members of PAL gene family; for example, 16 in *Vitis vinifera*, 20 in *Zea mays* (Rawal et al., 2013). The difference in number of PAL genes among various species might be the result of genome size, species type, number of protein coding genes and evolution. The phylogenetic analysis revealed that PAL genes of different species were closely related and divided into three clades were closely related, however, *CoPAL5* was out grouped along with *EgrPAL2* (Figure 01).

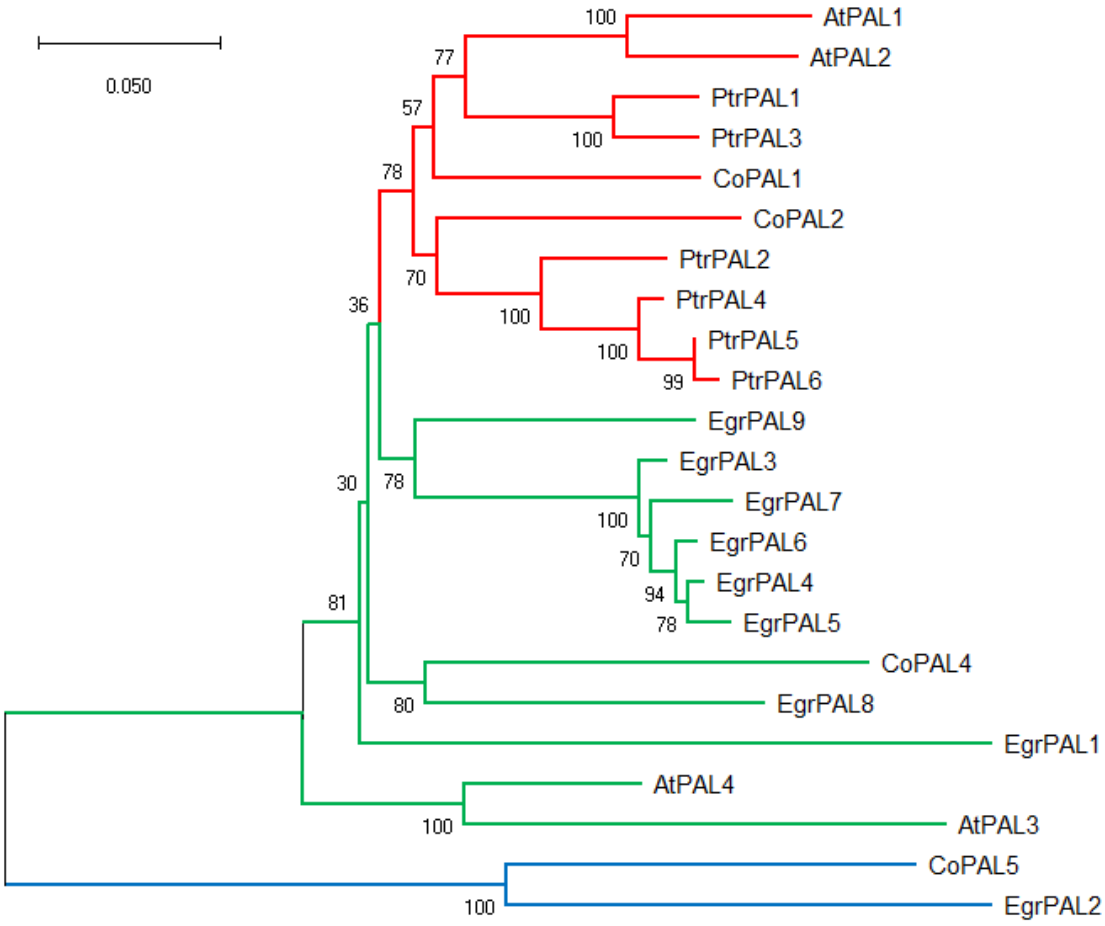

Figure 01. The phylogenetic relationships of PAL proteins of in *C. olitorius*, Arabidopsis, Eucalyptus, and Poplar. The Neighbor-Joining tree was constructed with 4 *AtPAL*, 9 *EgrPAL*, 6 *PtrPAL*, 5 *CoPAL* using MEGAX. The bootstrap value was 1000 replicates.

Molecular characteristics and structural analysis of *CoPALs*

The length of the identified *CoPAL* genes was ranged from 630 to 730 amino acids with molecular weight from 75.08 to 79.30 Kda as shown in Table 01. All *CoPALs* were slightly acidic as the values of pI were varied from 5.97 to 6.59. In addition, the GRAVY score for all *CoPALs* were negative suggesting that they were hydrophilic in nature (Table 01). These molecular properties were comparable to PAL genes of other species with some variation (Dong et al., 2016). Like other species, the putative *CoPALs* were

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

localized in the cytoplasm (Yan et al., 2019). To understand the genomic distribution, PAL genes were mapped to the reference genome. *CoPALs* were unevenly distributed because three *CoPALs* were located on chromosome 2 and the rest of one (*CoPAL4*) was in chromosome 4 (Table 01). The gene structure analysis is crucial to know the diversification of species, therefore intron-exon organization and motifs were also investigated in the study. Three *CoPALs* had two exons separated by a single intron while *CoPAL5* had only one exon (Figure 02). This result was also coordinated with the phylogenetic tree as the *CoPAL5* was placed in different clade in the tree (Figure 01). The length of exons of other *CoPALs* was similar, however, the length of intron was varied. Consistent with our results, similar intron-exon structure was observed in walnut and watermelon (Dong et al., 2016). Moreover, 15 conserved motifs were identified in *CoPALs* and most of motifs were found in the *CoPALs* except 14 and 15 (Figure 03). These motifs were absent in the *CoPAL5*. The predicted jute PAL gene motifs ranged from 14 to 50 amino acids which were similar to other studies (Yan et al., 2019).

Table 01. Molecular Properties of PAL genes in *C. olitorius*

Gene Name	AA	MW	pI	Gravy	Subcellular localization	Chromosomal location
CoPAL1	730	79.30	5.97	-0.176	Cytoplasm	Chromosome 2
CoPAL2	716	78.10	6.13	-0.171	Cytoplasm	Chromosome 2
CoPAL4	697	75.90	6.59	-0.185	Cytoplasm	Chromosome 4
CoPAL5	683	75.08	6.19	-0.085	Cytoplasm	Chromosome 2

Legends: AA- protein length (number of amino acid); pI- theoretical isoelectric point; MW- molecular weight (kDa); GRAVY- grand average of hydropathicity

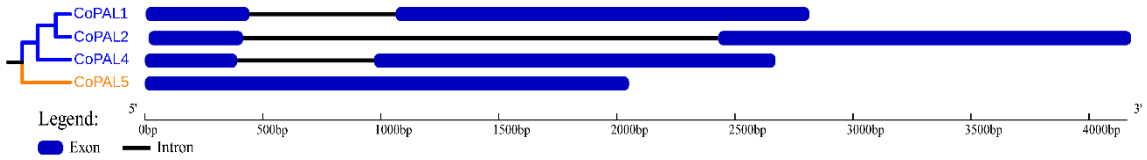

Figure 02. Graphical representation of exon/intron structure of *CoPALs* genes using online tool GSDS.

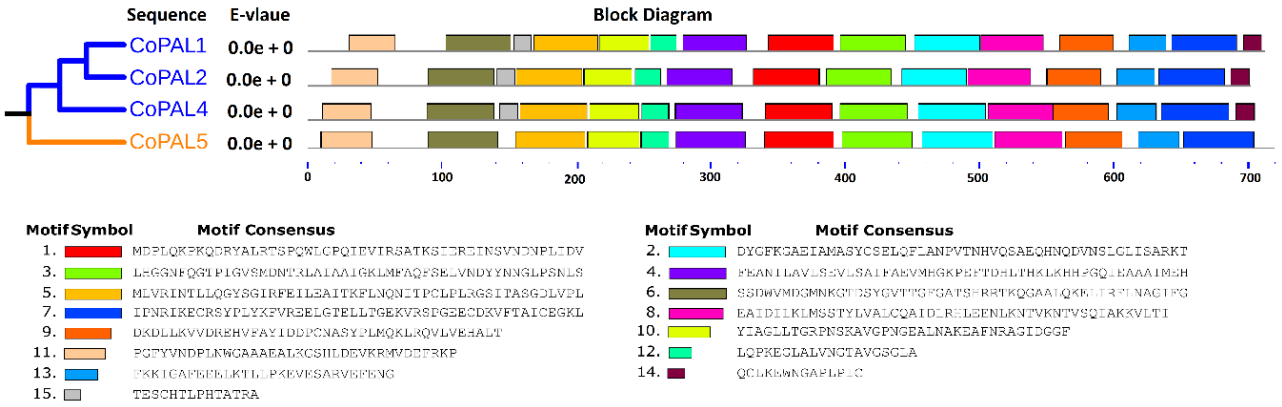

Figure 03. Schematic illustration of *CoPALs* motif analysis. Each motif is represented by different colored boxes

Expression profiling of *CoPALs* genes

To get the function of the *CoPALs* a comparative gene expression analysis was carried out using transcriptome data among the different tissues in various growth stages such as seedlings, Leaf, stem, bast, flower and fruit. RNA-seq analysis revealed that *CoPAL1* and 2 were higher than other two *CoPALs* genes; especially transcripts level of *CoPAL1* gene was higher in stem than in any other tissues (Table 02 and Figure 04) because stem tissues comprise both xylem and phloem. These xylem and phloem tissues are higher lignified tissue than other tissues as they provide mechanical strength and rigidity to the plants. Similar results were also observed in other studies, for example, *PbPAL1* and *EgrPAL3* were highly expressed in pear and eucalyptus (Cao and Li, 2019; Carocha et al., 2015), respectively. However, *CoPAL1* exhibited moderate expression in all jute organs or tissue. Because PAL genes are not only

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

involved in lignin biosynthesis but also associated with other metabolites production such as flavonoids, isoflavonoids, anthocyanins, plant hormones and phytoalexins. On the other hand, *CoPAL2* had higher expression in seedlings and stem while moderate expression was observed in other tissues suggesting that *CoPAL2* was associated with the development of different tissues. The results of expression analysis suggesting that these *CoPAL1* and *2* were mainly involved in phenylpropanoid/monolignol-pathway in jute. Since a higher amount of lignin in jute fiber is the main hindrance for use in the textile industry, therefore, more functional studies of these two PAL genes are further necessary to know the specific functions of the genes in lignin synthesis to reduce lignin content in jute.

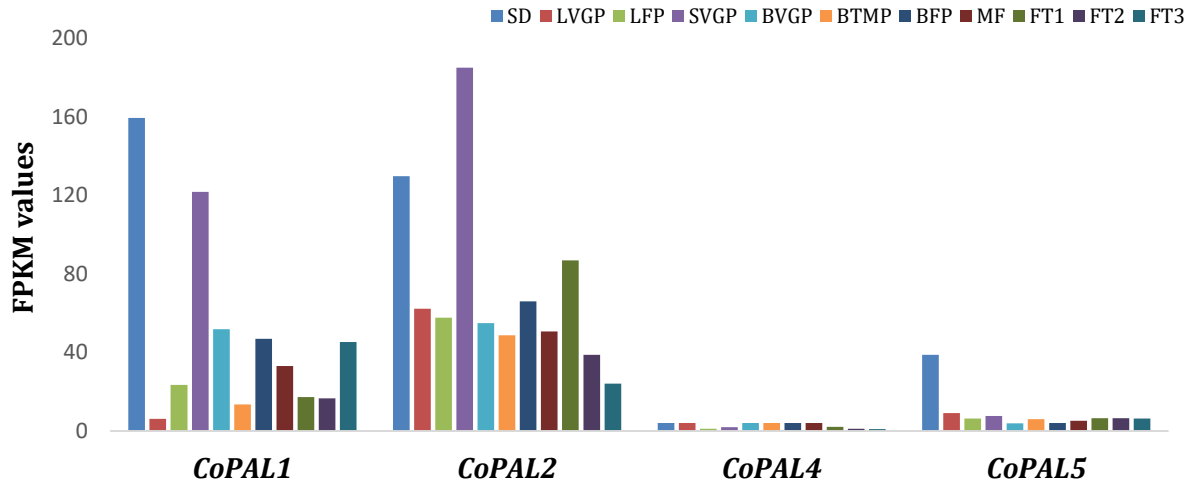

Figure 04. Gene expression analysis using FPKM values obtained from transcriptome data of different jute tissues in various growth stages.

SD- seedlings; LVGP- leaf of vegetative growth period; LFP- leaf of flowering period; SVGP- Stem of vegetative growth period; BVGP- bast of vegetative growth period; BTMP- bast of technical mature period; BFP- bast of flowering period; MF- mature fruit; FT1- fruit1−2 cm in length; FT2- fruit 2−4 cm in length; FT3- fruit greater than 4 cm in length

Table 02. FPKM values (fragments per kilobase per million mapped reads) of different jute tissue in various growth stages obtained from transcriptome data

	SD	LVGP	LFP	SVGP	BVGP	BTMP	BFP	MF	FT1	FT2	FT3
CoPAL1	159.5	6.1	23.6	121.9	51.9	13.6	47.0	33.1	17.4	16.6	45.4
CoPAL2	129.9	62.3	57.7	185.2	55.0	48.7	66.1	50.8	87.0	38.9	24.1
CoPAL4	4.1	4.1	1.2	2.0	4.0	4.0	4.1	4.1	2.1	1.2	1.0
CoPAL5	38.8	9.1	6.3	7.6	4.0	6.1	4.0	5.2	6.5	6.6	6.3

Legends: SD- seedlings; LVGP- leaf of vegetative growth period; LFP- leaf of flowering period; SVGP- Stem of vegetative growth period; BVGP- bast of vegetative growth period; BTMP- bast of technical mature period; BFP- bast of flowering period; MF- mature fruit; FT1- fruit1−2 cm in length; FT2- fruit 2−4 cm in length; FT3- fruit greater than 4 cm in length.

IV. Conclusion

In the present study, a total of 4 PAL genes was identified and distributed in two chromosomes. Molecular properties and structural organizations were comparable to PAL genes in other species. Higher expression of *CoPAL1* and *CoPAl2* suggested that they were associated with plant development especially in lignin biosynthesis. These findings provide data for regulating lignin biosynthesis in jute.

V. References

[1]. Cao, Y. and Li X. (2019). Integrative Analysis of the Core Fruit Lignification Toolbox in Pear Reveals Targets for Fruit Quality Bioengineering. Biomolecules, 9 (9), 504. https://doi.org/10.3390/biom9090504

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

[2]. Carocha, V., Soler, M., Hefer, C., Cassan-Wang, H., Fevereiro, P., Myburg, A. A., Paiva, J. A. and Grima-Pettenati, J. (2015). Genome-wide analysis of the lignin toolbox of *Eucalyptus grandis*. The New phytologist, 206(4), 1297-1313. https://doi.org/10.1111/nph.13313
[3]. Dong, C-j., Cao, N., Zhang, Z-g and Shang, Q-m. (2016). Phenylalanine ammonia-lyase gene families in cucurbit species: Structure, evolution, and expression. Journal of Integrative Agriculture, 15(6), 1239-1255. https://doi.org/10.1016/S2095-3119(16)61329-1
[4]. Fraser, C. M. and Chapple, C. (2011). The phenylpropanoid pathway in Arabidopsis. The arabidopsis book, 9, e0152. https://doi.org/10.1199/tab.0152
[5]. Hossain, M. S., Ahmed, B., Ullah, M. W., Aktar, N., Haque. M. S. and Islam, M. S. (2020). Genome-wide identification of fasciclin-like arabinogalactan proteins in jute and their expression pattern during fiber formation. Molecular biology reports, https://doi.org/10.1007/s11033-020-05858-w
[6]. Hossain, M. S., Ahmed, R., Haque, M. S., Alam, M. M. and Islam, M. S. (2019). Identification and validation of reference genes for real-time quantitative RT-PCR analysis in jute. BMC Molecular Biology, 20(1), 13. https://doi.org/10.1186/s12867-019-0130-2
[7]. 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. and Wan, X. (2017). Comparative genomics of two jute species and insight into fibre biogenesis. Nature Plants, 3, 16223. https://doi.org/10.1038/nplants.2016.223
[8]. Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Molecular biology and evolution, 35(6), 1547-1549. https://doi.org/10.1093/molbev/msy096
[9]. Rawal, H. C., Singh, N. K. and Sharma, T. R. (2013). Conservation, Divergence, and Genome-Wide Distribution of PAL and POX A Gene Families in Plants. International journal of genomics, 2013, 1-10. https://doi.org/10.1155/2013/678969
[10]. Trapnell, C., Roberts. A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols, 7 (3), 562-578. https://doi.org/10.1038/nprot.2012.016
[11]. Yan, F., Li, H. and Zhao, P. (2019). Genome-Wide Identification and Transcriptional Expression of the PAL Gene Family in Common Walnut (*Juglans Regia* L.). Genes, 10(1), 46. https://doi.org/10.3390/genes10010046
[12]. Yang, Z., Wu, Y., Dai, Z., Chen, X., Wangb, H., Yang, S., Xie, D., Tang, Q., Cheng, C., Xu, Y., Deng, C., Liu, C., Chen, J. and Su, J. (2020). Comprehensive transcriptome analysis and tissue-specific profiling of gene expression in jute (*Corchorus olitorius* L.). Industrial Crops and Products, 146, 112101. https://doi.org/10.1016/j.indcrop.2020.112101.
[13]. Yoon, J., Choi, H. and An, G. (2015). Roles of lignin biosynthesis and regulatory genes in plant development. Journal of integrative plant biology, 57(11), 902-912. https://doi.org/10.1111/jipb.12422.

HOW TO CITE THIS ARTICLE?

Crossref: https://doi.org/10.18801/jbar.260220.267

MLA
Hossain, M. S. et al. "Phenylalanine ammonia-lyase gene family (PAL): Genome wide characterization and transcriptional expression in jute (*Corchorus olitorius*)". Journal of Bioscience and Agriculture Research, 26(02), (2020): 2185-2191.

APA
Hosain, M. S., Ahmed, R., Ullah, M. W., Honi, U., Tareq, M. Z., Sarker, M. S. A., Ahmed, B. and Islam, M. S. (2020). Phenylalanine ammonia-lyase gene family (PAL): Genome wide characterization and transcriptional expression in jute (*Corchorus olitorius*). *Journal of Bioscience and Agriculture Research*, 26(02), 2185-2191.

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

Chicago
Hosain, M. S., Ahmed, R., Ullah, M. W., Honi, U., Tareq, M. Z., Sarker, M. S. A., Ahmed, B. and Islam, M. S. "Phenylalanine ammonia-lyase gene family (PAL): Genome wide characterization and transcriptional expression in jute (*Corchorus olitorius*)". Journal of Bioscience and Agriculture Research, 26(02), (2020): 2185-2191.

Harvard
Hosain, M. S., Ahmed, R., Ullah, M. W., Honi, U., Tareq, M. Z., Sarker, M. S. A., Ahmed, B. and Islam, M. S. 2020. Phenylalanine ammonia-lyase gene family (PAL): Genome wide characterization and transcriptional expression in jute (*Corchorus olitorius*). Journal of Bioscience and Agriculture Research, 26(02), pp. 2185-2191.

Vancouver
Hosain MS, Ahmed R, Ullah MW, Honi U, Tareq MZ, Sarker MSA, Ahmed B and Islam MS. Phenylalanine ammonia-lyase gene family (PAL): Genome wide characterization and transcriptional expression in jute (*Corchorus olitorius*). Journal of Bioscience and Agriculture Research, 2020 December 26(02): 2185-2191.

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