Int.J.Curr.Microbiol.App.Sci (2020) *9*(10): 2534-2541

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Volume 9 Number 10 (2020)**
Journal homepage: http://www.ijcmas.com

Original Research Article https://doi.org/10.20546/ijcmas.2020.910.305

Identification of Green Tissue Specific Genes in Cotton Employing Transcriptome Sequencing

D. Anne Kitty Deborah[1*], Ganesan Balasubramani[2] and Vijay N. Waghmare[3]

[1]Women Scientist Scheme-A of Department of Science and Technology, Biotechnology unit, ICAR-Central Institute for cotton Research, Post Bag No-2, Shankar nagar post office, Nagpur-440010, Maharashtra
[2]Biotechnology unit, ICAR-Central Institute for Cotton Research, Post Bag No-2, Shankar nagar post office, Nagpur-440010, Maharashtra
[3]ICAR_Central Institute for cotton Research, Post Bag No-2, Shankar nagar post office, Nagpur-440010, Maharashtra, India

**Corresponding author*

Keywords

Transcriptome sequencing, Differential gene expression, Green tissue

Article Info

Accepted:
20 September 2020
Available Online:
10 October 2020

A B S T R A C T

The present study aims at identification of green tissue specific genes in *Gossypium hirsutum* variety, Suraj for promoter isolation and characterization in the future. For this, transcriptome sequencing of three different tissues viz., seed, leaf and boll rind was done. Differential gene expression analysis between seed (non green) and leaf and bollrind (green) tissues was done to identify genes which are highly expressed in green tissues i.e., leaf and boll rind and very low or no expression in non green tissues i.e., seed. Fifty genes were selected for validation through Real Time PCR studies. For Real Time PCR studies, apart from leaf , boll rind and seed at 25 DPA, boll rind and seed @ 10 DPA, pollen and petals were also studied. Five genes were found to show high expression in green tissues and very low or no expression in non green tissues. As expected, most of them were found to be chloroplastic genes. These are the potential genes for isolation and characterization of green tissue specific promoters.

Introduction

Transgenic crop plants have been a great relief in crops lacking natural source of resistance to various pests and diseases. In any genetic engineering programme of developing transgenic plants, promoter plays an important role to express "transgene" in the host plants. Strong constitutive expression promoters such as the CaMV 35S

(cauliflower mosaic virus 35S) promoter and the maize Ubi-1 (Ubiquitin) gene promoter (Odell *et al.,* 1985; Cornejo *et al.,* 1993) are widely used in the development of economically useful transgenic crops. Unfortunately, the constitutive expression of foreign genes may be harmful to the host plant and reported to cause sterility, retarded development, abnormal morphology, yield penalty, altered grain composition or transgene silencing (Sinha *et al*., 1993; Matzke *et al*., 2000; Kurek *et al.,* 2002; Cheon *et al*., 2004 and Xu *et al*., 2006). It was also reported that considerable amounts of Bt protein can accumulate within cotton seeds. This has caused public concern about using transgenic crops as food. Using green tissue-specific promoters will avoid or reduce the expression of transgene in seed or fruit besides reducing the constitutive expression penalty. However until now, there are few green tissue specific promoters with strong and reliable expression that have been used in biotechnology. Photosynthetic-related gene promoters, such as *rbcS* (ribulose 1,5 bisphosphate carboxylase), *LHCPII* (light-harvesting chlorophyII a/b binding protein of photosystem II), *D540* (photosystem II 10 kDa polypeptide), and *cyFBPase* (cytosolic fructose-1,6-bisphosphatase), have been shown to activate expression in green tissues (Tada *et al*., 1991; Luan and Bogorad 1992; Kyozuka *et al*., 1993; Jang *et al*., 1999; Nomura *et al*., 2000; Si *et al*., 2002; Cai *et al.,* 2007). Other genes like rice rubisco activase, *DX1*, Leaf Panicle 2, Pyruvate, orthophosphate dikinase (PPDK), Sucrose-phosphate synthase, *Psak* and *Pharbitis nil* Leucine zipper (Yang *et al*., 2012; Ye *et al*., 2012; Thilmony *et al*., 2009; Taniguchi *et al*., 2000; Ana *et al*., 2000 ; Lin *et al*., 2017 and Wang *et al*., 2016) also showed green tissue specific expression.

In the present study, transcriptome sequencing of seed, leaf and boll rind of *Gossypium hirsutum* variety, Suraj has been carried out on Illumina NextSeq500 platform. Differential gene expression analysis between seed (non-green) and leaf and boll rind (green) tissues was carried out to identify putative green tissue-specific genes whose promoters could be useful in the development of economically important transgenic cotton that would help in targeted expression of transgenes.

RNA isolation and Library preparation

High-quality RNA was isolated from 4 weeks old leaf and 25 DPA boll rind and seed of *Gossypium hirsutum* variety, Suraj. RNA-Sequence paired end sequencing libraries were prepared from the Quality Control passed RNA samples using Illumina TruSeq Stranded mRNA sample Prep kit. Briefly, mRNA was enriched from the total RNA using poly-T attached magnetic beads, followed by enzymatic fragmentation, 1st strand cDNA conversion using SuperScript II and Act-D mix to facilitate RNA dependent synthesis. The 1st strand cDNA was then synthesized to second strand using second strand mix. The double stranded cDNA was then purified using AMPure XP beads followed by A-tailing, adapter ligation and then enriched by a limited number of PCR cycles.

Quantity and quality check (QC) of library on Agilent 4200 Tape Station

The PCR enriched libraries were analyzed in 4200 Tape Station system (Agilent Technologies) using high sensitivity D1000 Screen tape as per manufacturer instructions. After obtaining the Qubit concentration for the libraries and the mean peak sizes from Agilent Tape Station profile, the PCR Enriched illumina libraries were loaded onto NextSeq500 for cluster generation and sequencing. Bioinformatic tools were used to create high quality reads

The high-quality reads of the 3 samples of cotton were mapped on the reference genome of *Gossypium hirsutum* using TopHat v 2.1.1 with default parameters

Differential Gene Expression (DGE) Analysis

Differential Gene Expression (DGE) Analysis was done using Cufflinks v 2.2.1 program. The analysis was carried out for commonly expressed genes reported between control and treated samples respectively. Seed was considered as control and leaf and boll rind as treatments. There were two combinations for differential gene expression analysis. In the first combination, seed was compared with leaf and in the second combination, seed was compared with bollrind. FPKM (Fragments per Kilobase of transcript per Million Mapped reads) values were used to calculate the log fold change as $\log_2$ (FPKM Experimental/ FPKM Control). Log_2 Fold Change (FC) values greater than zero were considered up-regulated whereas less than zero were down-regulated along with P-value threshold of 0.05 for statistically significant results.

Heat map

An average linkage hierarchical cluster analysis was performed on top 50 differentially expressed genes using multiple experiments viewer (MeV v 4.9.0). The heatmap shows the level of gene abundance. Differentially expressed genes were analyzed by hierarchical clustering.

The summary of differential gene expression Analysis is as follows:

Combination 1: It shows the differentially expressed genes between seed and leaf. In combination 1, there are 1,241 genes exclusively expressed in leaf, 932 genes exclusively expressed in seed, 39,104 genes expressed both in seed and leaf, 800 genes upregulated in leaf and 510 genes downregulated in leaf (Table 1).

Combination 2: It shows the differentially expressed genes between seed and boll rind. In combination 2, there are 354 genes exclusively expressed in boll rind, 2,128 exclusively expressed in seed, 36,487 genes expressed both in seed and boll rind, 663 genes upregulated in boll rind and 582 genes downregulated in boll rind (Table 1).

Identification of putative green tissue specific genes

A total of 42,587 genes were commonly expressed in combination 1 (seed vs leaf) and 40,214 genes in combination 2 (seed vs boll rind). Fifty common genes expressed in leaf and boll rind were shortlisted for validation of relative expression analysis through Real Time PCR study. These 50 genes contains genes which are either upregulated or only expressed in green tissues (leaf and boll rind) when compared to non-green tissue (seed). Gene specific Primers were designed for these 50 genes based on coding sequence from NCBI with the help of IDT software.

A total of 3 common genes which were exclusively expressed in leaf and boll rind when compared to seed were selected whose FPKM values were more than ten in leaf and more than two in boll rind.

The genes which were upregulated in green tissues are of utmost importance as they are the predicted green tissue specific genes. So, a total of twenty eight genes were selected with FPKM values more than 50 and 10 in leaf and boll rind respectively. Of them, 5 are chloroplastic genes.

Chloroplasts are the organelles which give green colour due to the presence of

chlorophyll. The genes which are involved in the organisation of chloroplast are obviously a source of green tissue specific genes. Hence, eleven genes are selected which are chloroplastic and have FPKM values more than eighty five and twenty in leaf and boll rind respectively

Top ten genes with highest FPKM values in leaf and boll rind were also selected for relative expression analysis. The tabular form of the shortlisted genes is given below in table 2.

Validation of putative Green tissue specific genes through Real Time PCR studies

For Real Time PCR studies, apart from leaf, 25 DPA boll rind and 25DPA seed, boll rind and seed @ 10 DPA, petals and pollen were selected. Leaf tissue was taken as control and other tissues were compared with leaf for differential expression. Real Time PCR studies were done with Agilent technologies Stratagene MX3005p system.

Based on Real Time PCR studies, 5 genes were shortlisted for putative green tissue specific gene promoter isolation. The genes which were showing very low or no expression in non green tissues like seed, petal and pollen were selected as putative green tissue specific genes. They are chlorophyll a, b binding protein-4. Photosystem-II repair protein PSB 27-H1, Tetrapyrrole binding protein and Salicylate carboxymethyl transferase like protein and Cytosolic sulfotransferase like protein (Table 3). As we expected most of them were found to be expressed in chloroplast.

Table.1 Differential gene expression analysis summary

S.No	Description	Combination 1 (seed vs leaf)	Combination 2 (seed vs boll rind)
1	Exclusive leaf or boll rind	1,241	354
2	Exclusive seed	932	2,128
3	Expressed both	39,104	36,487
4	Upregulated	800	663
5	Downregulated	510	582
Total		42,587	40,214

Table.2 Criteria for shortlisting common genes in leaf and boll rind

S.No.	Category	FPKM Leaf	FPKM Boll rind	FPKM Seed	TOTAL GENES
1	Exclusively expressed in leaf and boll rind	>10	>2	-	3
2	Upregulated genes in leaf and Boll rind	>50	>10	-	23
	Chloroplastic	>50	>10	-	5
3	Chloroplastic, expressed in leaf & Boll rind	>85	>20	<1	11
	Top 10	741.8-13603.3	75.7-4752.3	1.98 - 159.7	8
Total					50

Table.3 Real time PCR results

S.No	Gene name	Gene Description	2^-ddct						
			Leaf	Boll rind @10 DPA	Boll rind @25DPA	Seed @10DPA	Seed @25DPA	petal	pollen
1	LOC107953509	Chlorophyll a,b binding protein-4, chloroplastic	1.00	0.113	0.336	0.001	0.018	0.000	0.002
2	LOC107946154	Photosystem II repair protein PSB 27-H1, chloroplastic	1.00	0.165	0.406	0.000	0.014	0.001	0.000
3	LOC107957175	Tetrapyrrole binding protein, chloroplastic like	1.00	0.082	0.286	0.001	0.008	0.002	0.003
4	LOC107954540	Salicylate carboxymethyl transferase like	1.00	0.145	0.380	0.000	0.000	0.006	0.036
5	LOC107951743	Cytosolic sulfotransferase -16 - like	1.00	0.053	0.230	0	0	0.000	0.000

Various fold change expression levels of the above 5 genes is as under (data not shown).

Chlorophyll a,b binding protein gene has shown various folds of lower expression in different tissues when compared to leaf tissue. It showed 8, 3, 1000, 55, very low and 500 folds lower expression in 10 DPA boll rind, 25 DPA boll rind, 10 DPA seed, 25 DPA seed, petals and pollen respectively when compared to leaf.

Photosystem II repair protein PSB 27-H1 has shown similar kind of expression pattern. The gene expression was 6, 4, very low, 71, 1000 folds lower and very low in 10 DPA boll rind, 25 DPA boll rind, 10 DPA seed, 25 DPA seed, petals and pollen respectively when compared to leaf.

The expression of Tetrapyrrole binding protein gene was 12, 3, 1000, 125, 500 and 333 folds lower in 10 DPA boll rind, 25 DPA boll rind, 10 DPA seed, 25 DPA seed, petals and pollen respectively when compared to leaf.

Salicylate carboxymethyl transferase like gene has shown 7, 3, very low, very low, 166 and 28 folds lower expression in 10 DPA boll rind, 25 DPA boll rind, 10 DPA seed, 25 DPA seed, petals and pollen respectively when compared to leaf.

Cytosolic sulfotransferase -16 like protein gene has shown almost very low expression in non green tissues like seed, petals and pollen whereas in 10 DPA and 25 DPA bollrinds, it was eighteen and four folds lower when compared to leaf.

All of these genes showed very low expression in non green tissues like seed, pollen and petals when compared to leaf tissue, whereas boll rind has shown very high expression when compared to seed, pollen and petals.

It is concluded, in the present study, five genes which were highly expressed in green tissues i.e., leaf and boll rind and very low or no expression in non green tissues i.e., seed, pollen and petals were identified. Most of them were found to be chloroplastic genes as expected. These genes can be used to isolate and characterize promoter regions for green tissue specific expression of foreign genes.

Acknowledgements

We acknowledge Department of Science and Technology, New Delhi for funding the project under DST Women Scientist Scheme-A and also thank the director of ICAR-CICR, Nagpur for all his support.

References

Ana T. C., Valdez-Alarcon, J. J., Martınez-Trujillo, M, Chen, L., Xoconostle-Cazares, B., Lucas, W. J. and Herrera-Estrella L. (2000) Tissue-Specific and Developmental Pattern of Expression of the Rice sps1 Gene. Plant Physiology. 124, 641–653.

Cai, M., Wei, J., Li, X. H., Xu, C. G. and Wang, S. P. (2007) A rice promoter containing both novel positive and negative cis-elements for regulation of green tissue-specific gene expression in transgenic plants. Plant Biotechnology. 5, 664–674.

Cheon, B.Y., Kim, H. J., Oh, K. H., Bahn, S. C., Ahn, J. H., Choi, J. W., Ok, S. H., Bae, J. M. and Shin, J. S. (2004) Overexpression of human erythropoietin (EPO) affects plant morphologies: retarded vegetative growth in tobacco and male sterility in tobacco and Arabidopsis. Transgenic Research. 13, 541–549.

Cornejo, M.J., Luth, D., Blankenship, K.M., Anderson, O.D. and Blechl, A.E. (1993) Activity of a maize ubiquitin promoter in transgenic rice. Plant

Molecular Biology. 23, 567–581.
https://ccb.jhu.edu/software/tophat/index.shtml
https://www.mybiosoftware.com/mev-4-6-2-multiple-experiment-viewer.html
Jang, I. C., Nahm, B. H. and Kim, J. K. (1999) Subcellular targeting of green fluorescent protein to plastids in transgenic rice plants provides a high-level expression system. Molecular Breeding. 5, 453–461.
Kurek, I., Stoger, E., Dulberger, R., Christou, P. and Breiman, A. (2002) Overexpression of the wheat FK506-binding protein 73 (FKBP73) and the heat-induced wheat FKBP77 in transgenic wheat reveals different functions of the two isoforms. Transgenic Research. 4, 373-9.
Lin, Z., Yan, J., Yan, H. and Wang, F. (2017) Characterization of a strong green tissue-specific motif in rice photosystem I gene promoter *Ppsak*. Plant Biotechnology Rep. 11, 87–95.
Luan, S. and Bogorad, L. (1992) A rice cab gene promoter contains separate cis-acting elements that regulate expression in dicot and monocot plants. Plant Cell. 4, 971–981.
Matzke, M. A., Mette, M. F. and Matzke, A. J. M. (2000) Transgene silencing by the host genome defense: implications for the evolution of epigenetic control mechanisms in plants and vertebrates. Plant Molecular Biology. 43, 401–415.
Nomura, M., Katayama, K., Nishimura, A., Ishida, Y., Ohta, S. and Komari, T. (2000) The promoter of rbcS in a C3 plant (rice) directs organ-specific, light dependent expression in a C4 plant (maize), but does not confer bundle sheath cell-specific expression. Plant Molecular Biology. 44. 99–106.
Odell, J. T., Nagy, F. and Chua, N. H. (1985) Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature. 313, 810–812.
Si, L. Z., Wang, L., Cao, S. Y., Chu, C. C. (2002) Deletion of 93 bp far upstream fragment of rice cytosolic fructose-1,6-bisphosphatase promoter completely alter its expresssion pattern. Acta Botanica Sinica. 44 (11),1339–1345.
Sinha, N. R., Williams, R. E. and Hake, S. (1993) Overexpression of the maize homeo box gene, KNOTTED-1, causes a switch from determinate to indeterminate cell fates. Genes Dev. 7, 787–795.
Kyozuka, J., McElroy, D., Hayakawa, T., Xie, Y., Wo, R., Shimamoto, K. (1993) Light-regulated and cell-specific expression of tomato rbcS-gusA and rice rbcS-gusA fusion genes in transgenic rice. Plant Physiology. 102, 991–1000.
Tada, Y. M., Sakamoto, M., Matsuoka and Fujimuram, T. (1991) Expression of a monocot lhcp promoter in transgenic rice. European Molecular Biology Organization. 10, 1803–1808.
Taniguchi, M., Izawa, K., Ku, MSB., Lin, J. H., Saito, H., Ishida, Y., Ohta, S., Komari, T., Matsuoka, M. and Sugiyama, T. (2000) The promoter for the maize C4 pyruvate, orthophosphate dikinase gene directs cell- and tissue-specific transcription in transgenic maize plants. Plant Cell Physiology. 41, 42–48.
Thilmony, R., Guttman, M., Thomson, J. G. and Blech, A. E. (2009) The LP2 leucine-rich repeat receptor kinase gene promoter directs organspecific, light-responsive expression in transgenic rice. Plant Biotechnology. 7, 867–882.
Wang, Q., Zhu, Y., Sun, L., Li, L., Jin, S. and Zhang, X. (2016) Transgenic Bt cotton driven by the green tissue-specific

promoter shows strong toxicity to lepidopteran pests and lower Bt toxin accumulation in seeds. Science China Life Sciences. 59 (2), 172–182.

Xu, R., Zhao, H., Dinkins, R.D., Cheng, X., Carberry, G. and Li, Q. Q. (2006) The 73 kD subunit of the cleavage and polyadenylation specificity factor (CPSF) complex affects reproductive development in Arabidopsis. Plant Mol. Biol. 61, 799–815.

Yang, Y., Yu, Y., Yang, G., Zhang, J., and Zheng, C. (2009) Tissue-specific expression of the PNZIP promoter is mediated by combinatorial interaction of different cis-elements and a novel transcriptional factor. Nucleic Acids Research 37, 2630–2644.

Ye, R., Zhou, F. and Lin, Y. (2012) Two novel positive cis-regulatory elements involved in green tissue-specific promoter activity in rice (Oryza sativa L. ssp.). Plant Cell Rep. 31, 1159–1172.

How to cite this article:

Anne Kitty Deborah, D., Ganesan Balasubramani and Vijay N. Waghmare. 2020. Identification of Green Tissue Specific Genes in Cotton Employing Transcriptome Sequencing. *Int.J.Curr.Microbiol.App.Sci.* 9(10): 2534-2541. doi: https://doi.org/10.20546/ijcmas.2020.910.305