Genomics Data 6 (2015) 212–213

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

RNA-Seq analysis for indigo biosynthesis pathway genes in *Indigofera tinctoria* and *Polygonum tinctorium*

Bijaya K. Sarangi [a], Yoshiko Minami [b], Sanjog T. Thul [a,*]

[a] Environmental Biotechnology Division, CSIR, National Environmental Engineering Research Institute, Nehru Marg, Nagpur 440020, India
[b] Department of Biochemistry, Okayama University of Science, Japan

ARTICLE INFO

Article history:
Received 22 September 2015
Accepted 27 September 2015
Available online xxxx

Keywords:
Unigenes
Indican
Transcriptome dataset

ABSTRACT

Natural indigo is the most important blue dye for textile dyeing and valuable secondary metabolite biosynthesized in *Indigofera tinctoria* and *Polygonum tinctorium* plants. Present investigation is made to generation of gene resource for pathway enrichment and to understand possible gene expression involved in indigo biosynthesis. The data about raw reads and the transcriptome assembly project has been deposited at GenBank under the accessions SRA180766 and SRX692542 for *I. tinctoria* and *P. tinctorium*, respectively.

Specifications	
Organism/cell line/tissue	*Indigofera tinctoria* and *Polygonum tinctorium*
Sequencer or array type	Illumina HiSeq™ 2000-RNA sequencing
Data format	Raw and processed
Experimental factors	Laboratory grown different species yielding indigo dye
Experimental features	RNA-seq dataset for gene expression profiling in young leaves
Sample source location	India (*Indigofera tinctoria*) and Japan (*Polygonum tinctorium*)

1. Direct link to deposited data

Transcriptome assembly project and raw data have been deposited at GenBank under the accessions SRA180766 (http://www.ncbi.nlm.nih.gov/sra/?term=SRA180766) and SRX692542 (http://www.ncbi.nlm.nih.gov/sra/?term=SRX692542) for *Indigofera tinctoria* and *Polygonum tinctorium*, respectively.

2. Experimental design, materials and methods

2.1. Plant materials

In the mature plants of *P. tinctorium*, younger leaves (especially first and second leaves) contain larger amounts of indican compared to older ones [1], similar observation was made in the case of *I. tinctoria* (data not shown). The seeds of *I. tinctoria* and *P. tinctorium* were germinated in pots containing soil and sand (1:1). Plants were grown in a plant growth chamber under 28 ± 2 °C and ~70% humidity.

2.2. Total RNA extraction and quality control

To maximize the number of indigo biosynthetic pathway genes, fresh leaves from two month old plant were sampled for total RNA extraction. Quantity and purity of extracted total RNA were determined using NanoDrop (Thermo Fisher Scientific Inc., USA) and Agilent 2100 bioanalyzer (Agilent Technologies, USA), respectively.

2.3. Transcriptome sequencing, de novo assembly, annotation and classification

These RNA samples were used for mRNA enrichment, fragmentation and cDNA library construction. The quality check was carried out by Agilent 2100 Bioanaylzer and ABI StepOnePlus Real-Time PCR System. The cDNA library construction and sequencing with an Illumina sequencing platform were performed at the Beijing Genomics Institute (BGI), China. Transcriptome de novo assembly is carried out with short reads assembling program — Trinity (release-20130225) [2] (Table 1). Further, BLASTX search was performed with obtained unigene sequences against several databases, including databases as NCBI Nr (http://www.ncbi.nlm.nih.gov/), Swissprot (http://www.expasy.ch/sprot/), KEGG (http://www.genome.jp/kegg/) and COG (http://www.ncbi.nlm.nih.gov/cog/), using a cut-off E-value of 10^{-5} (Table 2). To classify the unigenes, the Blast2GO program was used to get GO annotation based on molecular function, biological process and cellular component [3].

* Corresponding author.
E-mail addresses: s_thul@neeri.res.in, sanjay.thul@gmail.com (S.T. Thul).

http://dx.doi.org/10.1016/j.gdata.2015.09.021

Table 1
Statistics of sequencing and assembly of the *I. tinctoria* and *P. tinctorium* transcriptome.

Attributes	Value	
	I. tinctoria	*P. tinctorium*
Total raw reads	56,681,030	58,664,978
Total clean reads	53,337,778	54,784,512
Number of contigs	102,459	155,283
Number of unigenes	60,395	96,913
Q20 percentage	98.27%	98.10%
GC percentage	45.28%	49.73%

Table 2
Statistics of functional annotation of the *I. tinctoria* and *P. tinctorium* unigenes.

Databases	NR	NT	Swiss-Prot	KEGG	COG	GO	ALL
Number of unigenes	89,333	74,698	57,053	51,348	31,460	68,956	92,731

3. Conclusion

RNA-seq analysis of *I. tinctoria* and *P. tinctorium* from leaf tissues using Illumina HiSeq™ 2000 platform has led to the identification of transcripts of interest and may be useful in defining the role towards indigo biosynthesis, also several transcription factors. Further, the gene prediction and dataset development in non-model plants by Illumina platform are a much faster and cost-effective approach to enrich the gene resource for molecular and functional genomics of the blue dye producing species.

Conflict of interest

All the authors have approved submission and there are no conflicts of interest.

Acknowledgments

This work was funded by DST-JSPS grant number (Grant ID DST/INT/JSPS/P-163/2013) from DST, Govt. of India — JSPS, Japan. Authors wish to gratefully acknowledge Director, NEERI, Nagpur, and Council of Scientific and Industrial Research (CSIR), New Delhi, India and Okayama University of Science, Japan for providing facilities and encouragement.

References

[1] Y. Minami, O. Nishimura, I. Hara-Nishimura, M. Nishimura, H. Matsubara, Tissue and intracellular localization of indican and the purification and characterization of indican synthase from indigo plants. Plant Cell Physiol. 41 (2) (2000) 218–225.
[2] M.G. Grabherr, B.J. Haas, et al., Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29 (7) (2011) 644–652.
[3] A. Conesa, S. Gotz, M. Garrcia-Gomez, J. Terol, M. Talon, et al., Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21 (2005) 3674–3676.