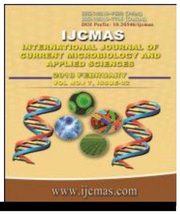
Int.J.Curr.Microbiol.App.Sci (2018) *7*(2): 2362-2366

EXCELLENT PUBLISHERS

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Volume 7 Number 02 (2018)**
Journal homepage: http://www.ijcmas.com



Original Research Article https://doi.org/10.20546/ijcmas.2018.702.288

Genome-Wide Identification of Genes, Transcription Factors and Transposable Elements in Sesame (*Sesamum indicum* L.)

P. Supriya[1*] and K.V. Bhat[2]

[1]*Division of Bioinformatics, ICAR-Indian Agricultural Research Institute, New Delhi, India*
[2]*Division of Genomic Resources, not Bioinformatics, India*

**Corresponding author*

ABSTRACT

Sesame is an ancient oil seed crop belonging to family Pedaliaceae. It is valued for high quality oil and seeds are rich source of unsaturated fatty acids, minor nutrients such as vitamins and minerals which consists of essential health and medicinal values. In the present study, genome-wide identification of genes, transcription factors and transposable elements was performed using draft genome assembly of Swetha variety of sesame. A total of 24579genes were identified among which 8244 genes had significant matches in Plant TFDB corresponding to 58 transcription factor families. In addition, a total of 25069 transposable elements were identified from scaffolds of draft assembly of sesame. Our study enhances the genomic resources of sesame and provides ample amount of information about genes and transposable elements which could serve as a valuable basis for future studies.

Keywords

Assembly, Annotation, Gene, Transcription factor, Transposon

Article Info

Accepted:
20 January 2018
Available Online:
10 February 2018

Introduction

Sesame (*Sesamum indicum* L.) belonging to the family pedaliaceae, is an ancient oilseed crop cultivated for its edible oil, seed and flavorsome. It is commonly known as Gingelly and widely as “Queen of Oilseeds” due to its high degree of resistance to oxidative rancidity (Bedigian and Harlan 1986). It is an excellent source of highly nutritive edible oil (50-60%) is rich in polyunsaturated fatty acids (PUFA) besides, medicinally important antioxidants, sesamin, sesamolin, and tocopherol (Fukuda *et al.*, 1985). Sesame seeds are rich source of nutritional components that offer numerous nutritional and physiological benefits. Sesame seeds contain high amounts of protein, dietary fiber, and vitamins, in addition to phosphorous, iron, magnesium calcium, manganese, copper and zinc. Despite having high nutritional value, research on this valuable oilseed is still at infancy. Completion of the whole genome sequencing of *Arabidopsis thaliana* was a major breakthrough, being the first plant genome to be sequenced. Understanding how the genome comprehensively encodes development and other environmental responses is the foremost challenge for all plant genome projects.

Repetitive DNA, particularly mobile genetic elements, represent substantial portions of many plant genomes (Bureau *et al.*, 1996). Though the evolutionary roles of these mobile genetic elements yet to be defined clearly (Lonnig and Saedler, 1997), the earlier studies reported that they could serve as a major source of mutation (Wessler, 1996; Britten, 1997). In flowering plants, transposable elements (TEs) are the most abundant genomic components. For instance, approximately 80% of the maize genome (Meyers *et al.*, 2001) and 40% of the rice genome is occupied by TEs (Sasaki, 2005). Based on transposition mechanisms, mobile elements are classified into two basic groups i.e. DNA transposons which are mobilized via DNA intermediates (Federoff, 1989) and retro transposons that are mobilized through RNA intermediates (Weiner *et al.*, 1986). Although whole genome sequence information is available for oil seed crops such as soy bean (Schmutz *et al.*, 2010), castor (Chan *et al.*, 2010), *Brassica rapa* (Wang *et al.*, 2011) and jatropha (Sato *et al.*, 2011), the ample amount of genomic information is lacking for sesame crop. Hence, the present study was undertaken to identify genes, transcription factors and transposable elements which can greatly aid in further understanding of their functions and mechanisms.

Materials and Methods

Gene prediction and functional annotation

The draft assembly of Swetha variety of sesame (Bioproject: PRJNA219369, Biosample number SAMN02357081) has been collected from NCBI website (http://www.ncbi.nlm.nih.gov/) and used for further analysis. The assembled scaffolds were used for GENSCAN gene prediction (Burge and Karlin, 1997) based on *Arabidopsis* model matrix. The functional annotation was performed through BLASTx against the NCBI non-redundant (Nr) protein database. Gene Ontology (GO) terms were assigned using Blast2GO software (Conesa *et al.*, 2005) and functional classification was achieved through WEGO software (Ye *et al.*, 2006). GO terms were used to classify the functions of the predicted genes. The GO mapping provides ontology of defined terms suggesting gene product properties in three categories viz., biological process, cellular component and molecular function. GO mapping was performed to retrieve GO terms for functionally annotated genes using BLASTx. The resulted accession IDs were used for the retrieval of gene names. The identified gene names were searched against the species specific entries of the gene product tables of GO database.

Identification of transcription factors and transposable elements

The genes identified in the present study were searched against Plant TFDB 4.0 (Jin *et al.*, 2013) using BLASTx with E-value threshold of 0.1 to identify the putative transcription factors. The identified transcription factors were further classified into different transcription factor families. A combination of homology-based and structure-based approaches was followed to identify TEs from the draft genome assembly (Wicker *et al.*, 2007).

Results and Discussion

A sequence similarity search was conducted against the NCBI non-redundant protein (Nr) database, the Swiss-Prot protein database, using BLASTx algorithm with an E-value threshold of 10^{-5}. In this study, a total of 24579 genes were predicted from whole genome assembly of Swetha variety of sesame. Blast2Go software was used to assign Gene ontology terms to classify the functions of the predicted genes.

Fig.1 Distribution of GO terms of genes

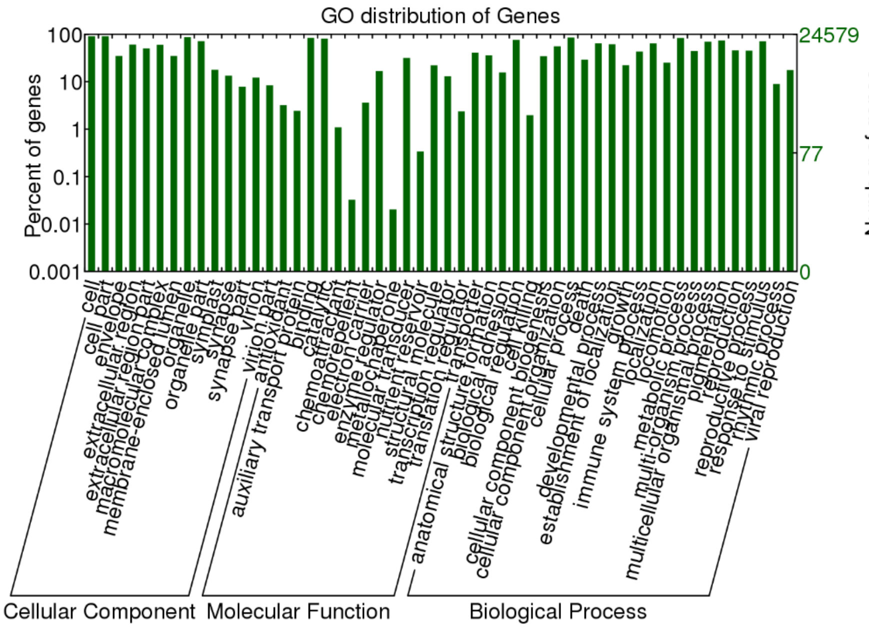
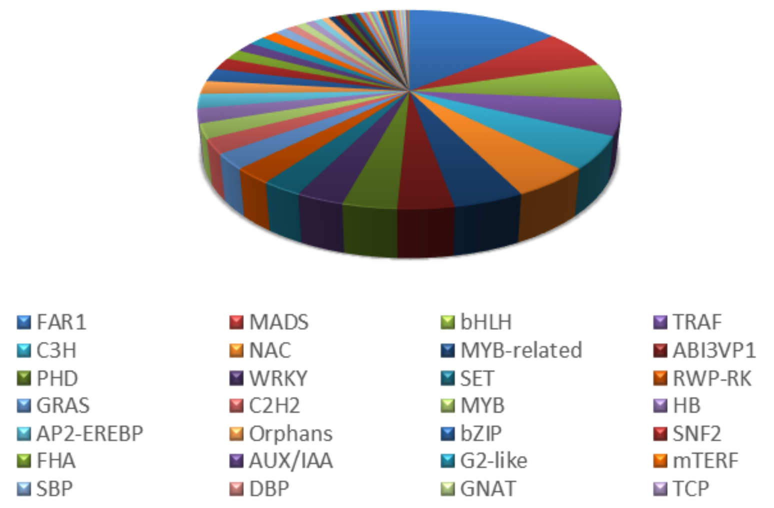


Fig.2 Distribution of transcription factors in sesame



The GO terms were classified into 59 functional groups (Figure 1) including biological processes (25), cellular components (19), and molecular functions (15). In the biological process category, metabolic process was the most prevalent (30.4% of sequences), followed by cellular process (28.4%) and response to stimulus (11.2%). In the molecular function category, binding was the most prevalent (46.4%), followed by catalytic activity (39.5%), structural molecule activity (4.95%) and transporter activity (4.95%). In the cellular component category, cell was the most dominant term (33.7%), followed by organelle (26.45%) and macromolecular complex (6.1%). The predicted genes were found to involve in pathways such as fatty acid biosynthesis, purine metabolism, starch and sucrose metabolism etc. Furthermore, a total of 8244 genes had significant matches in Plant TFDB corresponding to 58 families (Figure 2). Most abundant TF family was FAR1 (1132) followed by MADS (517), bHLH (499), NAC (420) whereas HRT (7) and SAP (2) were least abundant TF families. We also identified 20530 retrotransposons (Class I) and 3006 DNA transposons (Class II) from the scaffolds of draft assembly of sesame.

In the present study, a total of 24579 genes were predicted and functionally annotated. The identified genes in this study will not only facilitate the understanding of genetic basis of fatty acid biosynthesis, but also accelerate genetic improvement through marker-assisted selection in sesame. Transcription factors and transposons identified in this study will serve as valuable source for the sesame research community to understand the structural and functional characterization of TEs and their interaction with genes in sesame and related oilseed crops.

References

Bedigian, D. and Harlan, J. R. (1986). Evidence for cultivation of sesame in the ancient world. *Economic Botany*, *40*(2), 137-154.

Britten, R. J. (1997). Mobile elements inserted in the distant past have taken on important functions. *Gene*, *205*(1), 177-182.

Bureau, T. E., Ronald, P. C. and Wessler, S. R. (1996). A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. *Proceedings of the National Academy of Sciences*, *93*(16), 8524-8529.

Burge, C. and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *Journal of molecular biology*, *268*(1), 78-94.

Chan, A. P., Crabtree, J., Zhao, Q., Lorenzi, H., Orvis, J., Puiu, D. and Cahoon, E. B. (2010). Draft genome sequence of the oilseed species *Ricinus communis*. *Nature Biotechnology*, *28*(9), 951-956.

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M. and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*(18), 3674-3676.

Federoff, N. V. (1989). About maize transposable elements and development. *Cell*, *56*, 181-191.

Fukuda, Y., Osawa, T., Namiki, M. and Ozaki, T. (1985). Studies on antioxidative substances in sesame seed. *Agricultural and Biological Chemistry*, *49*(2), 301-306.

Jin, J., Zhang, H., Kong, L., Gao, G. and Luo, J. (2013). Plant TFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Research*, *42*(D1), D1182-D1187.

Lonnig, W. E. and Saedler, H. (1997). Plant transposons: contributors to evolution? *Gene*, *205*(1), 245-253.

Meyers, B. C., Tingey, S. V., and Morgante, M. (2001). Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Research*, *11*(10), 1660-1676.

Sasaki, T. (2005). The map-based sequence of the rice genome. *Nature*, *436*(7052), 793.

Sato, S., Hirakawa, H., Isobe, S., Fukai, E., Watanabe, A., Kato, M. and Takahashi, C. (2010). Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. *DNA Research*, *18*(1), 65-76.

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W. and Xu, D. (2010). Genome sequence of the palaeopolyploid soybean. *Nature*, *463*(7278), 178-183.

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S. and Huang, S. (2011). The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics*, *43*(10), 1035-1039.

Weiner, J. (1986). How competition for light and nutrients affects size variability in Ipomoea tricolor populations. *Ecology*, 1425-1427.

Wessler, S. R. (1996). Plant retrotransposons: turned on by stress. *Current Biology*, *6*(8), 959-961.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B. and Paux, E. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, *8*(12), 973-982.

Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z. and Wang, J. (2006). WEGO: a web tool for plotting GO annotations. *Nucleic Acids Research*, *34*(suppl_2), W293-W297.

How to cite this article:

Supriya, P. and Bhat, K.V. 2018. Genome-Wide Identification of Genes, Transcription Factors and Transposable Elements in Sesame (*Sesamum indicum* L.). *Int.J.Curr.Microbiol.App.Sci.* 7(02): 2362-2366. doi: https://doi.org/10.20546/ijcmas.2018.702.288