ICBAA2017-10

ANALYSIS OF BANANA TRANSCRIPTOME IN BANANA ROOTS IN RESPONSE TO INFECTION BY TROPICAL RACE 4 OF *Fusarium oxysporum* f. sp. *cubense*

Umaiyal Munusamy[1*], Yusmin Mohd-Yusuf[2,3], Nadiya Akmal Baharum[4,5], Kamilatulhusna Zaidi[3,5], Siti Nor Akmar Abdullah[1], Rofina Yasmin Othman[3,5*]

[1]*Institute of Plantation Studies, Universiti Putra Malaysia, 43400 UPM Serdang Selangor, Malaysia*
[2]*Centre for Foundation Studies in Science, University of Malaya, 50603 Kuala Lumpur, Malaysia*
[3]*Centre for Research in Biotechnology for Agriculture (CEBAR), University of Malaya, 50603 Kuala Lumpur, Malaysia*
[4]*Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*
[5]*Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia*

Abstract: Pisang Berangan, is the most popular dessert banana cultivar in Malaysia. This AA type is grouped within highly nutritious type of fruits. Therefore, its production to cater for human needs is crucial. However, fusarium wilt disease is threatening world banana production. Availability of banana transcriptome will be highly useful for improving banana gene annotation and for biological research. The knowledge of global gene expression patterns that are influenced by the Fusarium races, the varieties of infected banana cultivar and the environmental conditions will help to provide better understanding on the host responses towards the infection. RNA samples from laboratory infected Berangan were pooled for transcriptomics using the Illumina Technology. Analysis of the banana transcriptome led to the identification of overexpressed genes. Differential gene expression (DGE) profiling analysis reveals that the infection by Fusarium caused changes in the global gene expression profiles in the banana roots 2 hrs post infection. The infection led to the induction of many well known defense related genes. This transcriptome profiling analysis indicates that early detection is possible and through this any infected plant can be identified and further essential precaution can be carried out.

Keywords**:** *Musa acuminata*, Berangan*, Fusarium*, transcriptomics, early detection

INTRODUCTION

Musa acuminata is a popular fruit globally produced in a large scale in Asia and Africa. In this region, India is the largest producer of *Musa acuminata* with 29.8 million tons production yearly followed by Uganda, China and the Philippines (Swarupa et al., 2014). Although Cavendish is the variety of *Musa acuminata* found in the global market, cultivation of banana varieties in different regions differs widely based on local consumption. *Musa acuminata* cv *Berangan* is native to the Asian tropics including Malaysia, Indonesia, the Philippines, Australia and also East Africa. In Malaysia, this is the popular variety consumed as dessert (Ee Shian et al., 2012).

Basically, all varieties are under attack by wilting diseases, caused by the soil-borne fungus *Fusarium oxysporum* f. sp. *cubense* (Foc) (Heslop-Harrison and Trude Schwarzacher, 2007). Foc is a complex pathogen manifesting as subtypes or races that have resulted in outbreaks in Latin America, the

Caribbean Islands, Taiwan, the Philippines, Malaysia, Indonesia, the Northern Territory of Australia and China (Thangavelu et al., 2012; IPPC, 2013; García-Bastidas et al., 2014). Additionally, the very virulent Foc Tropical Race 4 (FocR4) is now not limited to tropical regions in Asia but has alarmingly been diagnosed in Mozambique and Jordan (Butler, 2013) and recently diagnosed in Northern Queensland, indicating that the disease is emerging in new regions (Biosecurity Queensland, 2015).

Therefore, developing new strategies to overcome these emerging diseases is crucial. Transcriptomic studies are being widely utilized to conduct investigation on gene expression in response to infection in banana. With this technique, in-depth knowledge and information on the interactive responses of the fungal pathogen and the susceptible host can be identified and further treatment, or necessary precaution can be taken to save the plant from deterioration. Huang et al. (2012) have reported new information on insect herbivory-induced biochemical process in banana plants through a transcriptomic approach. Li et al. (2014) had highlighted differentially expressed genes in elucidating bacterial pathogenesis also in banana where enough data was successfully obtained using transcriptomic approach. Furthermore, the breakthroughs of *Musa acuminata* and *Musa balbisiana* genome sequencing have greatly facilitated the designing, collecting and analyzing the gene expression data. With this unlimited information, the complexity of the disease infection network can be more accurately understood and explored for disease management task (Davey et al., 2013). The objective of this study was to generate RNA sequence data that would constitute a strong basis for transcriptomic studies in *Musa accuminata* cv Berangan by identifying the differentially expressed genes within the early stages of infection by FOC Race 4.

METHODS

Bioinformatic Analysis

Computing resources, programming, file formats, quality control

CLC Genomic Workbench software version 9.5 was used. NCBI database was utilized (http://www.ncbi.nlm.nih.gov/genome/10976) to download genome sequences and gene annotation files. The reads were functionally annotated using Blast2Go software. Softwares such as Aspera, FileZilla and Blast2Go were downloaded in LINUX system. Cross-platform file formats including fasta, fastq, sam, bam, gtf and gff files were used. Raw data with adjunct sequences were processed before being mapped.

Differential gene expression analysis (DGE)

The frequency of each DGE library was statistically analyzed to identify the differentially expressed genes. The false discovery rate (FDR) was used to determine the threshold of *P* value in multiple tests by manipulating the FDR value. Gene expressions were considered significantly different with a cut-off value FDR < 0.05 and |log2 ratio| > 1.

Venn diagram and heat map

Venn diagram was generated using Venny based on pairwise comparison (Oliveros et al. 2007-2015). The venn diagram was drawn based on the read counts. Heatmap was generated using Rscript for heatmap.

Gene ontology and KEGG prediction

The transcript sequence was extracted from banana gff files using per script. The extracted sequences were search against the SwissProt database using BLASTX. Transdecoder were used to predict the

peptides from the transcript sequences to serve as an input into Trinotate SQLite DB. The BLASTX results were loaded into Trinotate SQLite DB which will automatically link to Gene Ontology (GO). Trinotate results were generated with e-value threshold of 0.000001. SwissProt ID was extracted from Trinotate table. The SwissProt ID was mapped to KEGG Orthologous and KEGG pathway. All the results were combined into the CLC GWB format using custom scripts.

RESULTS AND DISCUSSION

Differential gene expression (DGE)

We examined the number of DGE from two different points (2 hrs compare with 48 hrs and 2 hrs compare with 96 hrs). We found that our RNA Seq results provide a wide difference in the number of DGE based on tagwise dispersion fold change being >2 and the FDR corrected p-value being smaller than 0.05. From the DGE analysis, three unigenes were significantly expressed after 48 hrs post infection relative to 2 hrs post infection. However, only one unigene was assigned to cellular component. The other two which were LOC103969247 and LOC103974217 are considered as the unique unigenes. In sample infected for up to 96 hrs, 38 unigenes were differentially expressed relative to 2 hrs of infection and only 10 were assigned with gene ontology. The remaining 28 unigenes may be considered as being exclusive to this particular plant pathogen interaction. According to Raherison et al. (2012), non-annotated sequences could be either unique to Fusarium and banana, are highly diverged to permit annotation or due to differential loss/acquisition of genes among taxa. In addition, these results also indicate that a broader range of unigenes may be actively expressed in 96 hrs infected sample than in 48 hrs infected samples relative to 2 hrs (Jayakodi et al., 2014).

Venn diagram

A total of 32 709 transcripts were detected. Figure 1 shows the total number of unigene from each of the two assemblies 2 hrs:48 hrs and 2 hrs:96 hrs and the number of unigene shared between two assemblies. Between 2n48 and 2n96 assemblies 30 977 and 30 613 unigenes were shared, respectively out of a total of 32 709 unigene sequences from each pairwise comparison. Among all two assemblies, only less than ~2000 unigenes were not being shared. Distinct unigene expressed in 2 hrs and reduction in the number of unigenes were noticed in unigene derived from 48 hrs and 96 hrs. This shows that more unigenes can be detected during the early detection stage, as at this time point, the plant is actively preparing itself for defense against fusarium infection. These comparisons indicate that patterns of gene expression are quite different in various time frame examined, yielding a diverse array of transcript sequences (Rowland et al., 2012).

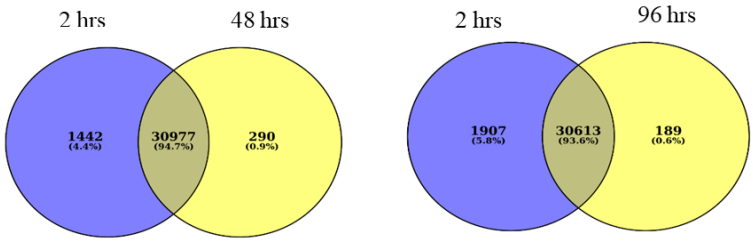

Figure 1: Venn diagram showed the number of distinct transcripts among the tested time points.

Heatmap

This two-component (2 hrs and 48 hrs and 2 hrs and 96 hrs) seemed to play significant roles in recognition and adaptation of the duration of infection. This heat map demonstrated that fusarium infection could modulate the expression of genes in the plant. When compared between 2 and 48 hrs, only 3 similar unigenes were detected and were upregulated at 48 hrs. While the comparison between 2 and 96 hrs showed more transcripts. However, the upregulated of similar transcript at 2 hrs was down-regulated at 96 hrs, while the down-regulated transcript at 2 hrs was identified to be up -regulated at 96 hrs. These upregulated gene can be used as markers to detect early fusarium infection. The general low level of transcripts expression may suggest that they were expressed in fewer tissues or were tightly regulated in some manner, perhaps playing a more specialized role in metabolism or development upon infection (Zhao et al., 2014). In addition, defence related genes in plants often have low expression in conditions without pathogen challenge, but are rapidly displayed upon pathogen attack (Cheng et al., 2015).

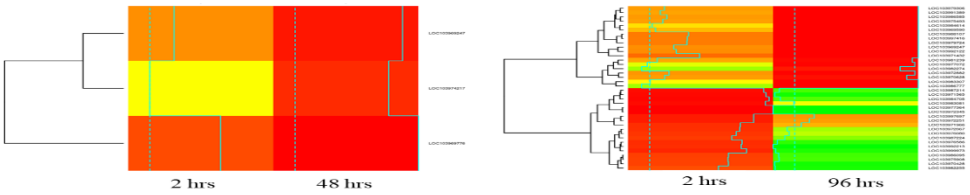

Figure 2: Analysis of differential gene expression (DGEs) in 2 stages-48 hrs and 96 hrs in comparison with 2 hrs of infection. Overall cluster analysis of DGEs for these two stages. FPKM (fragments per kilo base of transcript per million base pairs sequenced) was used to estimate the level of gene expression. The color change from red (highly expressed) to green (low expression) represents expression level value log 2 (ratios).

Gene ontology (GO) annotation and KEGG prediction

Only 11 unigenes were successfully assigned for GO terms (Table 1). All these unigenes were assigned into cellular components, molecular function and biological processes. In the 48 hrs infected sample, 3 unigenes were upregulated while in 96 hrs, 19 unigenes were upregulated and another 19 were down regulated. To gain insight into the effect of time on infection, we illustrated expression patterns with a heat map obtained via a hierarchical cluster analysis. This clustering revealed the relatedness of the various unigenes. Transcription factors are important for controlling many other genes. Seven unigenes were identified with transcription factor characteristics (LOC1039-70428, 71565,76080,82255,86095,92213,99973). Among them, four were upregulated (LOC1039-70428, 82255,86095,92213,99973) while the rest were down regulated. As for the GO terms, 16 classes were related to biological processes, nine for cellular component and only four for molecular function indicating a large biological process of genes in the transcriptomic data. Within the biological processes, seven unigenes were identified to be involved in transcription process followed by five unigenes participating in the ethylene biosynthesis pathway and four unigenes in the defense response particularly in defacing fungus. Plants can respond to pathogen infection by inducing a broad spectrum resistance, a phenomenon called systematic acquired resistance (SAR). Induction of SAR includes induction of pathogens responsive genes, salicyclic acid and ethylene release compounds (Guo and Ecker, 2004). Five KEGG pathways were

upregulated in 2 hrs of infection while four in 96 hrs of infection. For example, the pentose and glucuronate interconversion pathway is the series of metabolic reactions involves in the interconversion of the monosaccharide pentose and glucuronate (Yew et al., 2005). According to Kwasiborski et al. (2014) pentose phosphate pathway seems to be enhanced and would provide the needed nucleic acids and energy for the wound colonization during infection. The glucuronate pathway is essential for plant cell wall formation during pathogen infection (Usadel et al., 2004).

Table 1: Gene Ontology Classification of assembled unigenes.

Ontology	Class	Number of Unigene
Biological Process	Defense response/fungus	4
	Abscisic acid	1
	Ethylene	5
	Jasmonic acid	1
	Salt stress	2
	Transcription	7
	Regulation Transcription	1
	Nucleosome assembly	2
	Anoxia	1
	Cold Acclimation	1
	Photosynthetic	1
	Chitin	3
	Heat	1
	Oxidative stress	1
	UV-B	1
	Wounding	1
Cellular Component	Nucleus	8
	Nucleolus	2
	Nucleosome	3
	Plasma membrane	2
	Plasmodesma	2
	Tylakoid	2
	Vaculor membrane	2
	Chloroplast	2
	Cytosol	2
Molecular Function	Chromatin binding	1
	DNA binding	8
	Sequence specific DNA binding transcription factor activity TF	6
	Metal ion binding	1

CONCLUSIONS

Through the analysis of the transcriptome data obtained through RNA-Seq, we identified 38 putative banana genes that were previously annotated and nine KEGG pathways were successfully predicted. Obvious difference was found in the early infection process (2 hrs). Fusarium infection induced expression

of many genes commonly responsive to infection. Pathogen related genes were successfully identified at early stage of infection even though the total number of genes that was differentially expressed are lesser. This information can be used to save the particular infected plant from succumbing to the disease or being a source for infecting other plants in the same field.

REFERENCES

Biosecurity Queensland (2015). Panama disease tropical race 4: Standards and guidelines.

Butler, D. (2013). Fungus threatens top banana. Nature. Issue 504, volume 7479: 195-196. Retrieved June, 2nd, 2014. http://www.nature.com/news/fungus-threatens-top-banana-1.14336.

Cheng, W., Li, H.P., Zhang, J.B., Du, H.J., Wei, Q.Y., Huang, T., Yang, P., Kong, X.W., Liao, Y.C. (2015). Tissue-specific and pathogen-inducible expression of a fusion protein containing a *Fusarium*-specific antibody and a fungal chitinase protects wheat against *Fusarium* pathogens and mycotoxins. Plant Biotechnol J. 13: 664–674.

Davey, M.L., Heimdal, R., Ohlson, M., Kauserud, H. (2013). Host- and tissue-specificity of moss-associated *Galerina* and *Mycena* determined from amplicon pyrosequencing data. Fungal Eco. 6:179–186.

Ee Shian, T., Abdullah, A., Hamid Musa, K., Yusof Maskat, M., ABD Ghani, M. (2012). Antioxidant Properties of Three Banana Cultivars (*Musa acuminata* 'Berangan', 'Mas' and 'Raja') Extracts. Sains Malaysiana. 41: 319–324.

García-Bastidas, F., Konkol, J., Al-Qasim, M., Naser, Z., Abdelwali, M., Salem, N., Waalwijk, C., Ploetz, RC., Kema, G.H.J. (2014). First Report of *Fusarium oxysporum* f. sp.*cubense* Tropical Race 4 Associated with Panama Disease of Banana outside Southeast Asia. Disease Notes. 98: 5-694.

Guo, H., Ecker, J.R. (2004). The ethylene signaling pathway: new insights. Curr Opin Plant Biol. 7: 40–49.

Heslop-Harrison, J.S., Schwarzacher, T. (2007). Domestication, Genomics and the Future for Banana. Annals of Botany. 100:1073–1084.

Huang, Y.H., Wang, R.C., Li, C.H., Zuo, C.W., Wei, Y.R., Zhang, L., Yi, G.J. (2012). Control of fusarium wilt in banana with Chinese leek. NIH Public Access. 134: 87-95.

IPPC- International Plant Protection Convention (2013). New banana disease found in Mozambique (*Fusarium oxysporum* f.sp. *cubense* Tropical Race 4). Retrieved June 2, 2014.

Jayakodi, M., Lee, S.C., Lee, Y. S., Park, H.S., Kim, N.H., Jang, W., … Yang, T.-J. (2014). Comprehensive analysis of *Panax ginseng* root transcriptomes. BMC Plant Biology. 15: 138.

Kwasiborski, A., Bajji, M., Renaut, J., Delaplace, P., Jijakli, M.H. (2014). Identification of Metabolic Pathways Expressed by *Pichia anomala* Kh6 in the Presence of the Pathogen *Botrytis cinerea* on Apple: New Possible Targets for Biocontrol Improvement. PLoS ONE9(3). e91434.

Li, H., Chen, S., Song, A., Wang, H., Fang, W., Guan, Z., Jiang, J., Chen, F. (2014). RNA-Seq derived identification of differential transcription in the chrysanthemum leaf following inoculation with *Alternaria tenuissima*. BMC Genomics. 15: 9.

Oliveros, J.C. (2007-2015). Venny. An interactive tool for comparing lists with Venn's diagrams.

Raherison, E., Rigault, P., Caron, S., Poulin, P.L., Boyle, B., Verta, J.-P., … MacKay, J. (2012). Transcriptome profiling in conifers and the Picea GenExpress database show patterns of diversification within gene families and interspecific conservation in vascular gene expression. BMC Genomics. 13: 434.

Rowland, L.J., Alkharouf, N., Darwish, O., Ogden, E.L., Polashock, J.J., Bassil, N.V., Main, D. (2012). Generation and analysis of blueberry transcriptome sequences from leaves, developing fruit, and flower buds from cold acclimation through de-acclimation. BMC Plant Biology. 12: 46

Swarupa, V., Ravishankar, K.V., Rekha, A. (2014). Plant defense response against *Fusarium oxysporum* and strategies to develop tolerant genotypes in banana. Planta. 239: 735-751.

Thangavelu, R., Perumal, G.D., Mustaffa, M., Sreeramanan, S., Rathinam, X. (2012). Genomics of *Fusarium oxysporum* f. sp. *cubense* Causing Wilt Disease in Banana (*Musa* spp.) in Genetics, Genomics and Breeding of Bananas. Taylor & Francis Group.

Usadel, B., Schluter, U., Molhoj, M., Gipmans, M., Verma, R., Kossmann, J., Reiter, W.D., Pauly, M. (2004). Identification and characterization of a UDP-D-glucuronate 4-epimerase in *Arabidopsis.* FEBS Letters. 569: 327-331.

Yew, W.S., Akana, J., Wise, E.L., Rayment, I., Gerlt, J.A., (2005). Evolution of enzymatic activities in the orotidine 5'-monophosphate decarboxylase superfamily: enhancing the promiscuous D-arabino-hex-3-ulose 6-phosphate synthase reaction catalyzed by 3-keto-L-gulonate 6-phosphate decarboxylase. Biochem. 44: 1807-15.

Zhao, W., Xia, W., Li, J., Sheng, S., Lei, L., Zhao, S. (2014). Transcriptome profiling and digital gene expression analysis of *Fallopia multiflora* to discover putative genes involved in the biosynthesis of 2,3,5,4'-tetrahydroxy stilbene-2-O-β-D-glucoside. Gene. 547: 126-135.