TRANSCRIPTOME PROFILING OF RESISTANT AND SUSCEPTIBLE THEOBROMA CACAO FOLLOWING ARTIFICIAL INFECTION WITH COCOA POD BORER INSECT (*CONOPOMORPHA CRAMERELLA*) TO STUDY THEIR HOST-PATHOGEN INTERACTION

Roslina Mohd Shah,
Malaysian Cocoa Board,
Commercial Zone 1, Norowot Road,
Kota Kinabalu Industrial Estate,
88460 Kota Kinabalu, Sabah
Email: roslina@koko.gov.my

David Afong,
Malaysian Cocoa Board,
Commercial Zone 1, Norowot Road,
Kota Kinabalu Industrial Estate,
88460 Kota Kinabalu, Sabah

Navies Maisin,
Malaysian Cocoa Board,
Centre of Cocoa Research and Development,
Mile 10, Apas Road,
P.O.Box 60237, 91012 Tawau, Sabah
Email: navies@koko.gov.my

Rosmin Kasran,
Malaysian Cocoa Board,
Commercial Zone 1, Norowot Road,
Kota Kinabalu Industrial Estate,
88460 Kota Kinabalu, Sabah
Email: rosmin@koko.gov.my

Lee Ping Chin,
Faculty of Science and Natural Resources,
Universiti Malaysia Sabah,
Jln U.M.S., 88400 Kota Kinabalu, Sabah, Malaysia
Email : leepc@ums.edu.my

Larry Croft,
Malaysian Genomics Resource Centre Berhad (652790-V),
27-9, Level 9 Boulevard Signature Office, Mid Valley City,
59200 Kuala Lumpur, Malaysia

ABSTRACT

Plants developed different mechanisms to reduce insect attack, including specific responses that activate different metabolic pathways, which considerably alter their chemical and physical aspects. On the other hand, insects developed several strategies to overcome plant defence barriers, allowing them to feed, grow, and reproduce on their host plants. The Cocoa Pod Borer (CPB), also known as Cocoa Moth, is caused by the insect Conopomorpha cramerella, causing considerable losses in the cocoa industry during the 1890s and 1900s. CPB attacks both young and mature cocoa pods. A common symptom of infested pods is unevenness and premature ripening. The transcriptome will be sequenced from Theobroma cacao tissues (pod husk and placenta) of resistant and susceptible clones, before and after infection. The unique characterization of the differences observed, if any, in pod husk, placenta, and other relevant tissues may be used to understand the molecular variations underlying the phenotypic and disease resistance characteristics, and appropriate molecular markers identified can be used in downstream breeding research.

Key words: cocoa, CPB, resistance, susceptible, genes

INTRODUCTION

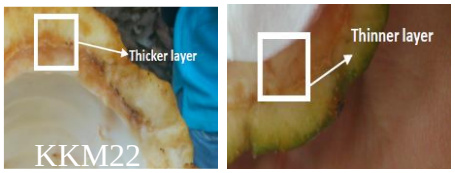
The Cocoa Pod Borer (CPB), also known as Cocoa Moth, is caused by the insect *Conopomorpha cramerella* (Ooi, 2005), causing considerable losses in the cocoa industry during the 1890s and 1900s (Valenzuela et al., 2014). CPB attacks both young and mature *Theobroma cacao* pods (Ooi., 2005). Malaysian Cocoa Board has identified and classified cocoa clones based on their yield, butter contents, tolerance to pest and diseases, and their availability to be planted in specific or in divest area. KKM22 and PBC123 are among the top clones selected and recommended to be grown in cocoa fields. However, both clones shown a distinct resistance toward CPB as KKM22 is more tolerant of CPB attack, whereas PBC123 are more susceptible to the attack by the insects.

From cocoa breeders evaluation, both pods for KKM22 and PBC123 clones are very distinct as one of them is more rounded and fewer ridges, while the others are longer and bigger with more ridges around the pod and different in pod colour too.

Figure 1: Pods of KKM22 and PBC123 (Malaysian Cocoa Clones, 2012)

The other significant differences between these two clones are the thickness of their sclerotic layer. KKM22 has a thicker layer compare to PBC123 with a thinner sclerotic layer.

Figure 2: Sclerotic layer of KKM22 and PBC123



Cocoa Pod Borer is the most critical devastating insect-pest for cocoa throughout South-East Asia.

Figure 3: Adult CPB, CPB larvae and damaged cocoa pod from CPB infestations

This species is found in the Philippines, Malaysia, Indonesia, and Papua New Guinea. Today CPB is one of the most devastating pests of cocoa in Southeast Asia and Pacific archipelagos. This pest is currently responsible for an average of 40-60% loss of cocoa production in the entire Southeast Asian region, and up to 80% losses in an unmanaged farm.
Infested pods will produce clumped beans and become unusable and must be discarded (McMahon et al., 2015). The use of pesticides as standard practice in the cocoa field, although useful, in the long run, may also lead to insecticide resistance and leaves harmful effects on humans and the environment. Growers have no choice either to adapt strategies or to suffer from a dramatic loss of income.

Performing analysis of the transcriptome and using high-throughput technologies are necessary to elucidate the molecular mechanisms and to identify biomarkers with the potential to be developed as markers that can be monitored in different agronomic traits. To gain insights into the genetic basis of the natural resistance of cocoa clones to CPB, a comparison between transcriptional response to the controlled CPB attacks of susceptible and resistant clones by RNAseq analysis fifteen days after infection were performed (Chen, et al., 2013; Gupta et al., 2015). The unique characterization of the differences observed; if any; in pod husk, placenta and other relevant tissues may be used to understand the molecular reasons behind the phenotypic and disease resistance characteristics, and appropriate molecular markers may be identified that can be used in downstream breeding research (Wang et al., 2009; Salgotra et al., 2014).

MATERIALS AND METHODS:

Rearing of CPB moths and selection of suitable pods for caging and controlled infection were performed before any laboratory works to be done. These works were done inside of the laboratory with adequate containment to limit the release of wild CPB in unaffected cocoa fields. Resistant and susceptible clones; KKM22 and PBC123 (Malaysian Cocoa Clones, 2012; Haya et al., 2007) were selected, and the controlled infection was done in one cocoa farmers plot in Kota Marudu. After fifteen days of infection with CPB moths, pods were collected and divided into tissues of pod husk and placenta of resistant and susceptible clones, before and after infection (Tan *et al.*, 2009). Transcriptomes were performed for samples listed as below;

a) Resistant, pod husk, before the infection
b) Resistant, pod placenta, before the infection
c) Resistant, pod husk, after infection
d) Resistant, pod placenta, after infection
e) Resistant, larvae + placenta, after infection

f) Susceptible, pod husk, before the infection
g) Susceptible, pod placenta, before the infection
h) Susceptible, pod husk, after infection
i) Susceptible, pod placenta, after infection
j) Susceptible, larvae + placenta, after infection
k) Larvae from an artificial diet
l) Larvae treated with caffeine from the artificial diet

RESULTS AND DISCUSSION:

RNA extraction was performed for eight plant tissue samples, and four RNA larvae were also extracted (Goh et al., 2014; Tan et al., 2009). All extraction samples were subjected to quality assessment process measured using QiaXpert (Qiagen) to verify the quality and quantity of the RNA

It is crucial to maintain a high RNA integrity number (RIN) for each sample before any deep sequencing to be performed. The selected samples of quality assessment results are shown in **Table 1.** Next, the sample is processed through messenger RNA isolation and cDNA synthesis using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® according to the manufacturer's protocol. Quantification of the synthesized cDNA was made using Qubit 2.0 DNA Broad Range Assay (Invitrogen, U.S.A.).

Sequencing

A minimum of 10ng cDNA was fragmented using a Covaris S220 (Covaris Inc, U.S.A.) to a targeted size of 200 – 300 bp. The fragmented cDNA was then end-repaired, ligated to NEBNext adapters, and was further PCR-enriched using NEBNext® Ultra™ RNA Library Prep Kit according to manufacturer's protocol. Quantification of the final sequencing libraries was performed using K.A.P.A. kit (K.A.P.A. Biosystem, U.S.A.) on Agilent Stratagene Mx-3005p quantitative PCR (Agilent, U.S.A.). Sizes were confirmed using Agilent Bioanalyzer High Sensitivity D.N.A. Chip (Agilent, U.S.A.). Finally, the libraries were sequenced using an Illumina flow cell with 209 cycles on the Illumina HiSeq 2000 platform (Illumina, U.S.A.). For this project, the sequencing run has generated a total number of 96.63 Gb raw data.

Cocoa Transcriptome Assembly Statistics

After the pre-processing stage, the clean reads from the data sets were assembled using the *M.G.R.C. Transcriptome Pipeline*. All reference transcripts were generated by combining all clean reads of the

Illumina sequencing data sets (**Table 2**). A more straightforward representation of transcripts length distribution identified in Cocoa transcriptome data sets were shown in **Figure 4**.

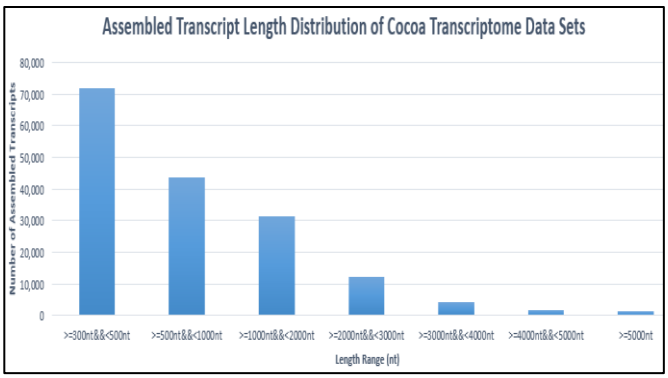
Table 1: Sample Quality Assessment Results

Tube No.	Sample Category	Sample Description	QiaXpert					RIN
			Conc (ng/ul)	A260/ 280	A260/ 230	Vol.	Total (ug)	
1	Cocoa RNA Larvae	Larvae from KKM22 in pulp infected with CPB	680.700	2.030	1.530	50	34.040	9.6
2	Cocoa RNA Larvae	Artificial Diet 1 RNA Larvae	708.100	1.790	2.130	50	39.910	9.1
3	Cocoa RNA Larvae	Caffeeine 1 RNA Larvae	228.800	1.830	1.880	50	11.440	8.6
4	Cocoa RNA Larvae	Larvae from pbc123 in pulp infected with CPB	304.700	2.090	1.400	30	9.141	9.5
5	Plant RNA	KKM22 Control POD HUSK 1	84.300	1.440	0.840	30	1.260	8.5
6	Plant RNA	KKM22 CPB Infected POD HUSK 1	24.000	1.610	0.470	40	0.960	8.7
7	Plant RNA	KKM22 Control POD PULP 1	24.500	1.830	1.340	30	0.740	9.3
8	Plant RNA	KKM22 CPB Infection POD PULP	152.400	0.880	2.250	40	6.096	7.9
9	Plant RNA	PBC 123 Control POD HUSK 1	33.100	1.990	1.910	30	0.993	9.5
10	Plant RNA	PBC 123 CPB Infected POD HUSK 1	75.900	1.990	1.830	30	2.277	7.0
11	Plant RNA	PBC 123 Control PULP 1	115.000	1.740	1.080	30	3.450	8.7
12	Plant RNA	PBC 123 CPB Infected PULP 1	36.600	1.820	1.790	30	1.098	7.0

Table 2: Statistics of assembled transcripts identified in Cocoa transcriptome data sets.

	Ref_Cocoa
Number of assembled transcripts	166,735
Total size of assembled transcripts	163,117,937
Longest assembled transcripts	28,452
Shortest assembled transcripts	300
Number of assembled transcripts > 1K nt	50,984
Number of assembled transcripts > 10K nt	55
Mean assembled transcripts size	978
N50 assembled transcripts length	1,516
Assembled transcripts %A	29.57
Assembled transcripts %C	20.38
Assembled transcripts %G	20.54
Assembled transcripts %T	29.51
Assembled transcripts %N	0
Assembled transcripts %non-ACGTN	0
Number of assembled transcripts non-ACGTN nt	0

Figure 4: Graphical length distribution summary of transcripts identified in Cocoa transcriptome data sets



Cocoa Transcriptome Alignment Statistics

Alignment statistics were reported from the S.A.M. format alignment files using custom Perl scripts (**Table 3**). A large number of the reads (96.16%) aligned back to the transcripts as expected. The reads that did not map back to the assembled transcripts corresponded to either low-quality reads or lowly-expressed transcripts that could not be assembled due to the minimum length requirement for transcripts (≥ 300nt). In this case, only a small portion of the reads (less than 3% of the reads) did not align back to the reference Cocoa transcriptome.

Table 3: Reads alignment statistics of transcripts identified in Cocoa transcriptome data sets

Alignment Statistics of Cocoa Transcriptome Data Sets	
Category	**Ref_Cocoa**
Total Reads	841,436,646
Reads Aligned	809,088,224
% Reads Aligned	96.16
Assembled Transcripts Length (in bases)	163,117,937
Total Assembled Transcripts Covered (in bases)	160,099,755
% Total Assembled Transcripts Covered	98.15
Average Read Depth	179.02

Downstream Analysis

In the final step, annotation and gene expression analysis were performed to characterize and identify potential resistance genes in *T.cacao* towards Cocoa Pod Borer insect.

Gene Ontology (G.O.)

Annotated transcripts were further categorized according to the three main G.O. domains: biological processes, cellular components, and molecular function, as shown in **Table 4.**

Differential Expression Analysis

Paired-end reads were aligned back to the assembled transcripts (length ≥ 300bp) through the *M.G.R.C. Transcriptome Pipeline*. For expression analysis, the M.G.R.C. Transcriptome Pipeline was used to estimate the abundance of transcripts. The results showed that around 86% of the paired-end reads from the Cocoa transcriptome datasets were mapped back to the reference transcripts. The lower mapping libraries around 73% are larvae samples. This is likely due to the lower depth of transcriptome sequencing coupled with the many orders of magnitude increased complexity of the insect transcriptome.

Figure 5: Differential expression analysis of kkm_husk_cpb vs pbc_HuSK_CPB at the gene level

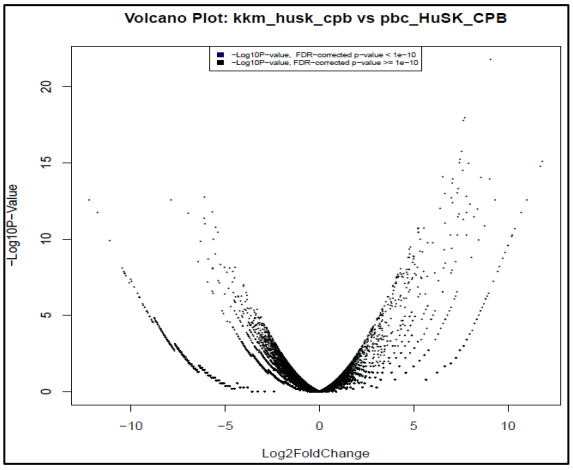
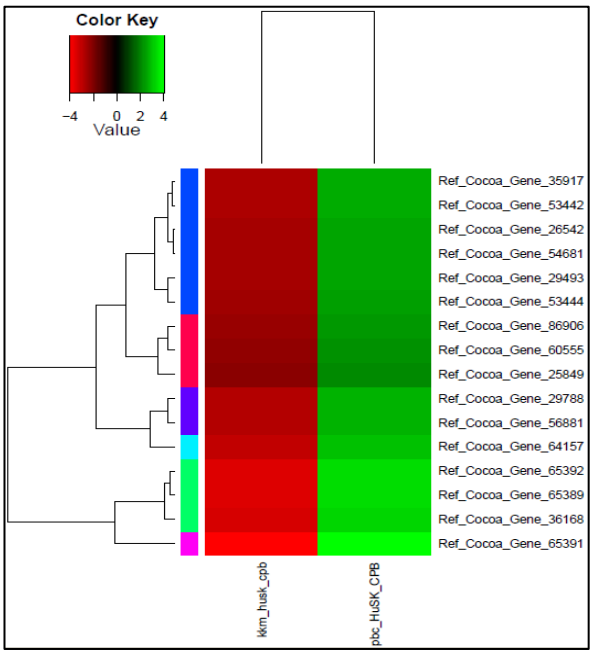


Table 4: Top ten G.O. terms in each G.O. domain for Cocoa/Pod Borer transcriptome data sets

GO ID	Biological Process	Number of Ref_Cocoa Transcripts
GO:0006351	transcription,_DNA-templated	1,426
GO:0006355	regulation_of_transcription,_DNA-templated	1,154
GO:0005975	carbohydrate_metabolic_process	817
GO:0015074	DNA_integration	700
GO:0055085	transmembrane_transport	652
GO:0006412	translation	499
GO:0007165	signal_transduction	421
GO:0006886	intracellular_protein_transport	377
GO:0035556	intracellular_signal_transduction	376
GO:0007018	microtubule-based_movement	374
GO ID	Cellular Component	Number of Ref_Cocoa Transcripts
GO:0016021	integral_component_of_membrane	6,235
GO:0005634	nucleus	3,839
GO:0016020	membrane	1,210
GO:0005737	cytoplasm	1,146
GO:0005886	plasma_membrane	1,018
GO:0005739	mitochondrion	843
GO:0009507	chloroplast	730
GO:0005576	extracellular_region	661
GO:0005829	cytosol	632
GO:0005622	intracellular	591
GO ID	Molecular Function	Number of Ref_Cocoa Transcripts
GO:0005524	ATP_binding	5,923
GO:0008270	zinc_ion_binding	3,879
GO:0003676	nucleic_acid_binding	3,460
GO:0003677	DNA_binding	3,052
GO:0046872	metal_ion_binding	2,786
GO:0004674	protein_serine/threonine_kinase_activity	1,838
GO:0003700	sequence-specific_DNA_binding_transcription_factor_activity	1,215
GO:0016787	hydrolase_activity	1,121
GO:0003723	RNA_binding	945
GO:0000166	nucleotide_binding	922

A Volcano plot of –Log10 (p-value) versus Log2FoldChange was plotted for each analyzed pair of transcriptomes. The Volcano plot assists in the identification of genes or transcripts with a definite statistically significant expression difference.

Figure 6: Heatmap of gene expression values show clustering (LHS dendrogram) of genes (RHS) between samples (kkm_husk_cpb vs pbc_HuSK_CPB) based on expression of mRNAs (color differences) for a set of significant genes (padj < 1e-10).



The heatmap of **Figure 6** is a scale colour image for representing the observed values of two different samples (kkm_husk_cpb vs. pbc_HuSK_CPB). The dendrogram at the side shows hierarchical clustering of the significant genes (padj < 1e-10). In the main panel, the sample names are represented in columns, and the significant genes are expressed in rows. The clustering algorithm groups related rows together by expression similarity. Low expression values of a sample tend towards red, while higher expression values tend to green in color.

Differential Expression Analysis

Pod Borer component of the transcriptome

A P450 cytochrome was identified as differentially expressed between regularly fed and caffeine fed (caffeine is a close analog of the drug theobromine found in cocoa). This is likely to be the P450 responsible for detoxification of these purinergic xenobiotics from the insect (Magnanou et al., 2013). No differential expression of this gene was found between the insects grazing on tolerant and susceptible cultivars. This suggests there is no chemical defence difference between tolerant and susceptible cultivars. An endogenous chemical insecticide is less likely to be the cause of cocoa cultivar resistance.

Other differentially expressed genes between borers fed caffeine and are not likely to be due to retarded life stage (fibroin), irritation of the insect digestive system (intestinal mucin, peritrophin). The LigB-like gene may also be involved in caffeine metabolism, but it is paradoxically down expressed in the caffeine fed borers.

Cocoa component of the transcriptome

Aluminum detoxification genes (Ryan et al., 2011) are active in infected plants, but not in uninfected plants. Low soil pH (under 4) mobilizes aluminum. The low pH or the uptake of aluminum due to low pH may stress the cocoa plant and make both resistant and susceptible plants more open to attack by borers.

There are clear plant defence responses to borers in husk and pod in both resistant and susceptible cultivars. The susceptible cultivar has a more significant response than the resistant cultivar. The most likely explanation is that the quantity of larvae in the pod determines the strength of the response. The resistant cultivar has fewer borers, and so the defence response is smaller.

This is suggestive of the resistant plants being physically more protected from borer attacks than the susceptible plants. There is a more significant response in the susceptible plants as more borers are getting through. This fits current models that the husk is thicker and more challenging for the borers to penetrate in the resistant cultivars (Haya et.al, 2007). It may also be resistant plants are less attractive to borers, or outer husk biochemical defence is more robust, which would both limit borer attack and hence plant defence activation.

General transcriptome differences between the susceptible and tolerant cultivars

Expression differences between the two cultivars are small in the samples uninfected by borers. Significant differences appear when the plant defence system becomes active after infection by borers. In uninfected pulp, the resistant cultivar seems to have higher oil production genes, possibly cultivar differences in oil content unrelated to insect tolerance.

The susceptible cultivar has more terpene biosynthesis activity in the husk than the resistant cultivar. Terpenes are often used as plant defences against insects and bacteria. Increased terpene production in the susceptible cultivar does not discourage the borers, so if the terpene biosynthesis is a defence against borers, it is ineffectual.

Looking at a comparison of the borer infected but tolerant cultivar and the borer infected susceptible cultivar there are many expression differences. Quite a number are insect defence-related. In the cocoa pulp, there are a set of chitinases genes (for digesting insect cuticle or fungal hyphae), which are strongly upregulated in the resistant cultivar. These are upregulated along with osmotin, which is a plant defence gene for bacteria, suggesting a plant defence response that is much more active in the resistant cultivar. This defence appears not to distinguish between bacteria and insects, or insect ingress brings in bacteria which the plant reacts to. In husk tissue, similar significant differences are observable between the cultivars.

Potential pigment genes are upregulated in the resistant cultivar along with again, a set of chitinases for defence against insect/fungal predation. These chitinases appear to be effective insect inhibitors as an insect fat storage protein is strongly upregulated in the resistant sample, suggesting the borers are struggling in the resistant cultivar, and fat and healthy in the susceptible cultivar.

The resistant cultivar also synthesizes more enzymes for phytoalexins (known to inhibit insect feeding P.M.I.D.: 24302045). These compounds are mostly antimicrobial, and the lack of P450 response in the larvae fed with cocoa pulp suggests these phytoalexin-like compounds are not effective poisons against borers. Still, the phytoalexin response is are of the general response of the resistant cultivar to attack (whether insect, fungi, or bacterial).

Components of the salicylic acid biosynthesis pathway are also upregulated in the resistant cultivar. This plant defence signaling hormone and the enhanced phytoalexin biosynthesis suggests a more active defence system is present in the resistant cultivar. The cocoa defence is comprised of a set of chitinases for digesting the borer cuticle and possibly phytoalexins, which may deter the borers with unpleasant flavours. A possible defensin locus is also upregulated. The whole defence system for any threat appears to be more active in the resistant cultivar.

CONCLUSION

This transcriptomic study of two cultivars of cocoa, one tolerant of cocoa pod borer attack and the other susceptible, has thrown a great deal of light on the plant defence mechanisms involved in borer tolerance. The resistant cultivar appears to have a heightened response to the attack, but also, there is less activity of defensive genes, possibly due to the lower burden of borers. Two loci identified by previous genetic methods show differential expression. The first locus has a small expression difference, which may or may not be mechanistically significant. The second locus has a massive expression difference between the two cultivars, but the function of the genes in this region is unknown. The transcriptome of the cocoa pod borer has also been assembled from this dataset, and a high confidence gene for xenobiotic resistance identified.

REFERENCES;

Chen, T., Lv, Y., Zhao, T., Li, N., Yang, Y., (2013) Comparative Transcriptome Profiling of a Resistant vs. Susceptible Tomato (*Solanum lycopersicum*) Cultivar in Response to Infection by Tomato Yellow Leaf Curl Virus. PLoS ONE 8(11): e80816.

Goh, L.P.W., Chia, Y.C., Tan, C.L., (2014), Isolation of D.N.A. and RNA from Cocoa Pod Borer, Conopomorpha cramerella (Snellen) and cloning its genes, Malaysian Cocoa Journal 8: 31-39

Gupta, V., Estrada, A.D., Blakley, I., Reid, R., Patel, K., Meyer, M.D., Andersen, S.U., Brown, A.F., Lila, M.A., Loraine, A.E., (2015), RNA-Seq analysis and annotation of a draft blueberry genome assembly identifies candidate genes involved in fruit ripening, biosynthesis of bioactive compounds, and stage-specific alternative splicing, GigaScience 4:5

Haya, R., Navies, M., Kelvin, L., Albert, L., Alias, A., and Nuraziawati, M.Y., (2007), Screening for cocoa pod borer resistance based morphological descriptors., Presentation at the 2007 Conference on Plantation Commodities, Putra World Trade Centre, Kuala Lumpur, Malaysia, 3-4 July 2007

Magnanou, E., Malenke, J.R., Dearing, M.D., (2013), Hepatic gene expression in herbivores on diets with natural and novel plant secondary compounds, Physiol Genomics 45: 774–785

Ooi, P.A.C., (2005), Report of the A.P.O. Seminar on Nonpesticide Methods for Controlling Diseases and Insect Pests, A.P.O. 2005, ISBN: 92-833-7037-6

Poczai1, P., Varga, I., Laos, M., Cseh, A., Bell, N., Valkonen, J.P.T., Hyvönen, J., (2013), Advances in plant gene-targeted and functional markers: a review, Plant Methods, 9:6

Roslina, M.S., Rosmin, K., Lea, J., Navies, M., Nor Aisyah., Sumayyah, A.K., Zafirah, Z., Fahmie, W., and Croft, L., (2014), Identification of Genetic Markers for Pod Borer Resistance by Sequencing 14 Cocoa Genomes, Malaysian Cocoa Journal 8: 40-48

Ryan, P.R., Tyerman, S.D., Sasaki, T., Furuichi, T., Yamamoto, Y., Zhang, W.H and Delhaize, E., (2011), The identification of aluminum-resistance genes provides opportunities for enhancing crop production on acid soils, Journal of Experimental Botany, Vol. 62, No. 1, pp. 9–20

Salgotra, R.K., Gupta, B.B, Stewart, C.N., (2014), From genomics to functional markers in the era of next-generation sequencing, Biotechnol Lett (36):417–426

Tan, C.L., Verica, J.A., Young, A., Pishak, S., Maximova, S.N., Guiltinan, M.J., (2009), Isolation of pod-specific genes in *Theobroma cacao*, Malaysian Cocoa Journal 5: 37-42

Teixeira, P.J.L., Lima, P.J.P., Thomazella, D.P.T., Prado, O.R.P.F., Rio, M.C.S., Mondego, J.M.C., Mieczkowski, P., Pereiraa, G.A.G., (2014), High-Resolution Transcript Profiling of the Atypical Biotrophic Interaction between *Theobroma cacao* and the Fungal Pathogen *Moniliophthora perniciosa*; The Plant Cell, Vol. 26: 4245–4269

Valenzuela, I., Purung, H., Roush, R.T., Hamilton, A.J.,(2014), Practical yield loss models for infestation of cocoa with cocoa pod borer moth, *Conopomorpha cramerella* (Snellen), Crop Protection 66 (2014) 19-28

Wang, Z., Gerstein, M., Snyder,.M., (2009), RNA-Seq: a revolutionary tool for transcriptomics, *Nat Rev Genet*. 2009 January; 10(1): 57–63

Xu, Q., Liu, C., Biswas, M.K, Pan, Z., Deng, X., (2014), Recent advances in fruit crop genomics, Front. Agric. Sci. Eng. DOI 10.15302/J-FASE-2014002

Malaysian Cocoa Clones (2012)