

ANALYSES OF *BOESENBERGIA ROTUNDA* CELL SUSPENSION CULTURES WITH OVEREXPRESSED *CHALCONE SYNTHASE* (CHS) MANIFESTING CHANGES ON THE EXPRESSION LEVEL OF FLAVONOID-RELATED GENES

Nurnadiah Roslan^{1a} , Teh Ser Huy^{2a} , Lee Yean Kee^{3b} , Noorsaadah Abd Rahman^{4b} and Zulqarnain Mohamed^{5a*}

^aInstitute of Biological Sciences, Faculty of Science, University of Malaya, 50603, Kuala Lumpur, MALAYSIA. Email: nur_nadiah@ymail.com¹ ; serhuy@gmail.com² ; zulq@um.edu.my⁵

^bDepartment of Chemistry, Faculty of Science, University of Malaya, 50603, Kuala Lumpur, MALAYSIA. Email : yeankee@um.edu.my³ ; noorsaadah@um.edu.my⁴

Corresponding author: zulq@um.edu.my

Received: 17th May 2019

Accepted: 16th Oct 2019

Published: 29th Feb 2020

DOI: <https://doi.org/10.22452/mjs.vol39no1.2>

ABSTRACT Over the years, hundreds of potentially important and useful compounds from *Boesenbergia rotunda* (*B. rotunda*) have been successfully isolated and studied as these compounds have been shown to exhibit pharmacological and traditional medicinal properties in treating illness and diseases. In this study, a high-throughput large-scale RNA sequencing (RNA-Seq) was used to generate massive transcriptome sequence data on cell suspension cultures with overexpressed *Chalcone synthase*. These data were further analysed to understand the influence of *CHS* overexpression in cell suspension cultures of *B. rotunda* and further examine the regulation of genes associated with flavonoids accumulation. Based on the distribution of differentially expressed genes (DEGs), 305 up-regulated genes and 181 down-regulated genes which were consistently shown between wildtypes and transgenics. However, further comparative analysis against the KEGG database highlighted the fact that expression level of flavonoid-related genes were down-regulated, which is deviated from our assumption. Also, due to the overexpression of *BrCHS2* transcript, unigenes that are responsible in the plant-pathogen interaction pathway also increased as a response towards stress. With a sufficient amount of data and transcripts obtained, identification of potential unigenes could be used for the biosynthesis of specific compounds such as flavonoids and plant defence metabolites in *B. rotunda*. Also, from these analyses, there is a potential biosynthesis pathway diverges from main pathways could be strongly supported and suggested.

Keywords: *Boesenbergia rotunda*, flavonoids, RNA-Seq, differentially expressed genes, rhizomes

1. INTRODUCTION

Boesenbergia rotunda (*B. rotunda*) is a monocotyledonous plant belonging to the family Zingiberaceae in the major group of Angiosperms (a group of flowering plants). It is a type of ginger species, which is synonymously known as *Boesenbergia pandurata* or *Kaempferia pandurata*. The locals recognise it as Temukunci, Krachai, Chinese keys, and Fingerroot ginger. The expression of *CHS* gene and its involvement in

flavonoid production has been well studied in a number of plant species, such as *Gerbera hybrida* (Deng et al., 2014), *Triticum aestivum* (Trojan et al., 2014), *Malus domestica* (Dare et al., 2013), *Oncidium Gower Ramsey* (Liu et al., 2012), *Juglans regia* (Cheniany et al., 2012) and *Arabidopsis thaliana* (Dao et al., 2011). Most people in many Asian countries use this *B. rotunda* plant as a condiment in food due to its aromatic flavour, in which its young rhizomes and shoots are cooked as vegetable or eaten raw. The fresh rhizomes are

practically used in post-partum tonic mixtures such as the popular Indonesian tonic, 'jamu' for women after childbirth, as a stomachic to improve digestion and to treat illness such as muscle pain (Tewtrakul et al., 2009). More focus has been put on the role of this *B. rotunda* plant and this has drawn a wide attention for researchers to isolate and elucidate more bioactive compounds that could therefore explain the significance of its traditional usage. These compounds were reported to exhibit an antimutagenic effect (Trakontivakorn et al., 2001), anticancer activity (Kirana et al., 2007), anti-dengue (Tan et al. 2006) and antioxidant activity (Shindo et al., 2006).

Available data on transcriptomic sequences in model organisms, such as *Arabidopsis*, *Oryza sativa* and *Zea mays* have greatly improved the knowledge of the growth and development in many higher plants (Dukowic-Schulze et al., 2014; Wakasa et al., 2014; Wang et al., 2014). However, these model organisms encounter high sequence repeat regions although they have been used to generate very large genomes (Guo, Gao, & Qian, 2014). This doesn't hamper researchers from utilising the available techniques to perform transcriptome sequencing which is a powerful attractive approach to generate genetic resources for species that have economic importance to the world. Examples of such initiatives include garlic (Sun et al., 2015), fern (Der et al., 2011), pea (Franssen et al., 2011), chickpea (Garg et al., 2011) and Japanese knotweed (Hao et al., 2011). To the contrary, for non-model plants such as *B. rotunda*, very little research has been carried out to producing limited information of genes that have been deposited in the GenBank database.

CHS genes are constitutively expressed during plant development which is mainly restricted to floral tissues, for example in *Lilium oriental* (Liu et al., 2011). Meanwhile, in *Physcomitrella patens*, there were eight classes of *CHS* genes expressed during different developmental stages or under different growth conditions (Jiang et al., 2006).

Furthermore, other *CHS* genes were reported to be transcriptionally induced by UV-B and wounding treatments such as in *Ginkgo biloba* (Pang et al., 2005). However, insufficient information on metabolic changes upon *CHS* overexpression is available for *B. rotunda* plant species. Therefore, this study forms a part of multidisciplinary research towards a better understanding of *CHS* and its functions on the compound synthesis and plant resistance.

The focus of this research was to study the effect of overexpression of *chalcone synthase (BrCHS2)* transcript in cell suspension cultures of *B. rotunda* by utilizing Illumina paired-end sequencing technology to characterise the transcriptome and annotate particular potential genes. Previously, *BrCHS2* transcript was found to be expressed predominantly in the rhizome. As such this transcript was introduced into *B. rotunda* cell suspension cultures for downstream analyses. Therefore, to elucidate those genes, triplicates of cell suspension cultures overexpressed with *BrCHS2* were harvested and analysed together with wild-type cultures for further observation on the DEGs. With the findings obtained, these data could serve as a reference for other Zingiberaceae families provided in the plant databases.

2. MATERIAL AND METHODS

2.1 Preparation of transgenic suspension cultures for RNA-Seq analysis

Fine cell suspension culture was established from a single bud that emerged from the rhizome and propagated in Murashige and Skoog (MS) (Murashige and Skoog, 1962) according to Wong et al. (2013)'s instructions. In order to study the effects of overexpression of *CHS* on the flavonoid accumulation, *BrCHS2* transcript which is predominantly transcribed in the rhizome was cloned in pCAMBIA1304 expression vector and expressed in *B. rotunda* cell suspension cultures via *Agrobacterium tumefaciens* strain 4404. After few screenings on the presence of

BrCHS2 transgene in transgenic lines cultured in LB media containing 300mg/L of Cefotaxime and 15mg/L of Hygromycin, detection of reporter gene expression by (β -glucuronidase) GUS assay and PCR analysis, those transgenic line were assayed on its enzymatic activity and subjected to HPLC analysis. In this analysis, six standards (MicroCombiChem, Germany) were applied to check the present of respective compounds (Suppl. 1a) and retention time for each standard was recorded and used as reference. Three transgenic lines showing highest detection amount of pinocembrin chalcone (Suppl. 1b) and three replicates of wildtypes that served as negative control were propagated in MS liquid medium (1962) supplemented with 0.5mg/L of 6-Benzylaminopurine (BAP) and 1mg/L of 2,4-dichlorophenoxyacetic acid (2,4-D) in a 250mL conical flask. The suspension cultures were incubated at $25 \pm 2^\circ\text{C}$ in a shaker at 100rpm in the plant growth room. Subcultures were carried out every 10 days up to two months to attain a sufficient number of suspension cultures.

2.2 Preparation for RNA extraction, cDNA library, and RNA-Seq

All six cell suspension cultures (including three wildtypes) were harvested after two months, dried and rapidly frozen in liquid nitrogen prior to storing at -80°C . The samples were transported to Beijing Genome Institute (BGI), Hong Kong, China for RNA extraction, cDNA library preparation and RNA-Seq analysis, following manufacturer's instructions. RNA quantity was determined with a Qubit Fluorometer and RNA integrity was confirmed using an Agilent 2100 Bioanalyzer (Agilent, Waldbronn, Germany). RNA concentration with more than 150 ng/ μL with an optical density reading OD_{260/280} of 1.8-2.21, rRNA ratio (28S/18S) \geq 1.9-2.4, and RNA Integrity Number (RIN) \geq 7.5 was used for cDNA library preparation. Next, total RNA was converted into a library of template molecules suitable for high-throughput DNA sequencing for subsequent cluster generation. Finally, high-throughput sequencing of six

libraries clones with 200bp short-insert library was carried out on the Illumina HiSeq™ 2000 (Illumina Inc., USA) platform.

2.3 Transcript assembly by Trinity and Unigene functional annotation

De novo transcriptome was carried out by assembling the clean data into contigs by short reads assembly program called Trinity (<http://trinityrnaseq.sourceforge.net/>). Functional annotation of the assembled unigenes was conducted and aligned against public protein databases such as non-redundant (Nr), Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO) and COG by performing a BlastX homology search with a cut-off e-value of $<1.00\text{E}^{-05}$ as well as a nucleotide (Nt) database by BlastN with a cut-off e-value of $<1.00\text{E}^{-05}$. Moreover, pathway annotation can be developed upon obtaining KEGG annotation to predict possible functions of the unigenes. Based on the KEGG database, a biological complex of all genes can be studied.

2.4 Differentially expressed (DE) genes analysis

The calculation of unigenes expression was determined using Fragments Per kb per Million reads (FPKM) method that can be directly used to compare the differences of gene expression between the samples. In order to identify the significantly expressed genes between wildtype and transgenic suspension cultures, analysis of differentially expressed genes was achieved via Poisson distribution calculations (Audic & Claverie, 1997). Using this analysis, the ratio of two-fold increase ($|\log_2\text{Ratio}|\geq 1$) and a lower false discovery rate (FDR) ($\text{FDR} \leq 0.001$) were used in all samples for the whole experiment. DEGs were then carried out into KEGG Pathway analysis.

2.5 KEGG pathway analysis on DEGs

The unigenes expression profiles between three replicate experiments of transgenic and wild-type suspension cultures

resulted in the formation of three main groups; Group 1: WT-1 vs TL-1, Group 2: WT-2 vs TL-2 and Group 3: WT-3 vs TL-3. Since overexpression of *BrCHS2* transcript could affect the gene expression in flavonoid and related plant pathways, DEGs between three groups were filtered according to several parameters against KEGG databases. The parameters were group with all DEGs conforming to be a) present in three groups with b) a fold change $\geq +1, \geq -1$, c) FDR of ≤ 0.001 and have d) significance, p-values of $p \leq 0.01$. These parameters were used as thresholds to obtain the significant values of the up- and down-regulated unigenes between both samples in the three groups. Flavonoid and related plant pathways with significantly enriched DEGs were investigated after comparing with the whole transcriptome background.

2.6 Validation analysis

Transcriptional results from RNA sequencing were validated by selecting 5 unigenes for reverse transcription-quantitative real-time PCR (RT-qPCR) analysis in all three groups. The unigenes included two random up-regulated unigenes (Unigene 28275_All and CL4317.Contig3_All) and three random down-regulated unigenes (CL1414.Contig2_All, Unigene 5234_All and CL12450.Contig1_All). The primers used for the validation are listed (Suppl. 2). A reference gene, elongation factor, *EF1 α* , was used for normalization. RT-qPCR was performed on QuantStudio™ 12K Flex real-time PCR platform (Applied Biosystem, USA) using Power SYBR® Green RNA-to-C_T™ 1-Step (Applied Biosystem, USA) to detect transcript

abundance. The amplification was conducted according to the thermal cycling conditions: reverse transcription step at 48°C for 30 min, enzyme activation at 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 15 sec, finally annealing and extension at 60°C for 1 min. The dissociation curve was established at the end of PCR cycle at 95°C for 15 sec, 60°C for 1 min followed by 95°C for 15 sec. The relative expression levels of the selected unigenes normalised to elongation factor were calculated using the 2^{- $\Delta\Delta$ Ct} method. All reactions were performed in three replicates.

3. RESULTS

RNA profiles of all six cell suspension cultures on their effects of *BrCHS2* overexpression were examined through RNA-Seq analysis. A total of 31.5 million sequence reads equivalent to 1.2 Gb of raw sequence reads were successfully generated. After stringent data cleaning and quality check, the sequences were estimated to have an average GC content of 49.19% with an average Q20 bases quality read of 97.95% (base quality was more than 20) as shown in Table 1. Upon clustering using TGI Clustering Tools (TGICL), a total of 102,195 unigenes with a total length of 103,043,426 nucleotides (nt) were generated. All-unigenes sequences from the six samples yielded a mean size of assembled unigenes of 1008 basepairs (bp). The dominance of this longer unigenes indicated the high quality of assembly in this study which was also represented by the N50 value of 1646 bp.

Table 1. Summary of reads assembly generated by Trinity from wildtype and transgenic *B. rotunda* cell suspension cultures

Samples	Total Raw Reads	Total Clean Reads	Total Clean Nucleotides (nt)	Q20 percentage	GC percentage
WT-1	71,160,282	64,400,870	5,796,078,300	97.94%	49.58%
WT-2	66,868,696	61,111,682	5,500,051,380	97.93%	50.23%
WT-3	54,766,806	49,816,986	4,483,528,740	97.80%	49.50%
TL-1	56,634,676	52,543,920	4,728,952,800	97.96%	48.83%
TL-2	69,256,744	64,557,498	5,810,174,820	97.99%	48.29%
TL-3	63,821,100	57,834,176	5,205,075,840	98.06%	48.68%

On the other hand, there were more than 300 bp in size in all assembled unigenes and the maximum length of unigenes was 101,630 bp. Among these, 61,838 unigenes (60.85%) were in the length range of 300 to 1000, 25,301 unigenes (24.89%) with length ranging from 1000 to 2000, while 14,536 unigenes (14.30%) were longer than 2000 bp (Suppl. 3). Compared with previous *B. rotunda* transcriptome databases (Md-Mustafa et al., 2014), the average length and the N50 of unigenes in this study were much longer with nearly half of total unigenes were more than

1000bp in size. Therefore, all the data generated in this study indicated a relatively good overall integrity of the unigenes. Most of the unigenes were successfully annotated using the Nr database (63.9%), followed by Nt (49.6%), Swiss-Prot (44.1%), KEGG (40.8%), GO (46.1%) and COG 26.7%) (Table 2). However, out of the total 102,195 of All-unigenes, only 67,251 unigenes (65.8%) were successfully annotated showing similarity to known proteins in all databases. The remainder which accounted for 34.2 % were left unannotated.

Table 2. Functional annotation of the *B. rotunda* cell suspension cultures in public databases

Annotated database	No. of unigene hits	Percentage
Nr	65296	63.9 %
Nt	50704	49.6 %
Swiss-Prot	45050	44.1 %
KEGG	41749	40.8 %
GO	47111	46.1 %
COG	27330	26.7 %
Total annotated unigenes	67251	65.85 %

Three output data were obtained after DEG analysis: Group 1 (WT-1 vs TL-1), Group 2 (WT-2 vs TL-2) and Group 3 (WT-3 vs TL-3). Overall, a total of 3148 unigenes were significantly differential-expressed (1905 genes were up-regulated and 1043 genes were down-regulated). We scaled down the data

group by filtering the DEGs based on the other three parameters (fold change $\geq +1$, ≥ -1 , FDR ≤ 0.001 and $p \leq 0.01$). The inter-sectioned among three comparison groups comprises 305 up-regulated genes and 181 down-regulated genes (Figure 1), and further evaluation was done on this data group.

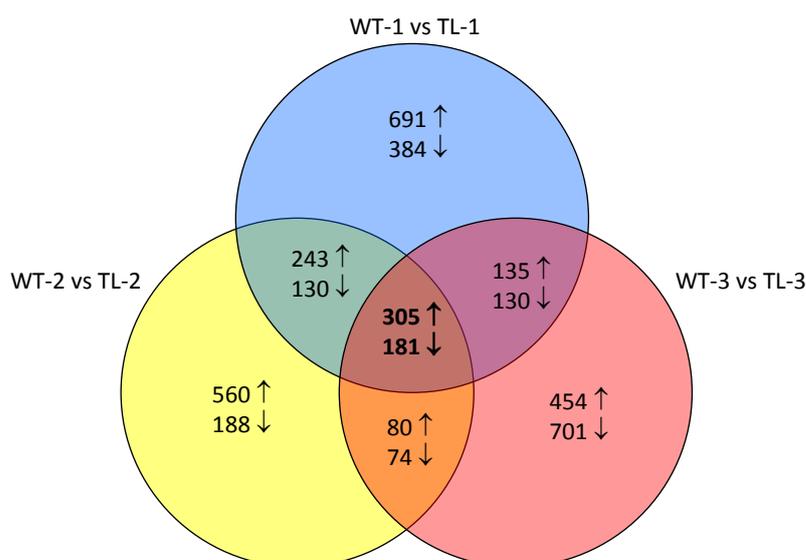


Figure 1. Venn diagram analysis on DEGs in three comparison groups (WT-1 vs TL-1, WT-2 vs TL-2 and WT-3 vs TL-3). The number of up-regulated unigenes are represented by symbol ↑ while down-regulated unigenes are represented by symbol ↓.

Differential expression pattern associated with biosynthesis of secondary metabolites were studied and further focused on elucidating potential unigenes being regulated in Flavonoid and its related plant pathways (Flavone and flavonol, Isoflavonoid and Anthocyanin pathway) as highlighted in the Table 3. Comparative analysis showed that in these pathways, there were 16 differentially expressed genes with most of them coded for down-regulated downstream enzymes. CL3398.Contig3_All was the recombinant *BrCHS2* transcript that was introduced into transgenic lines. As expected, it was highly expressed in TL samples (5.2 fold). The nucleotide sequence alignment between CL3398.Contig3_All and *BrCHS2* transcript is listed in Suppl. 4. The other four down-regulated unigenes (CL3398.Contig1_All (-5.6 fold), CL3398.Contig2_All (-4.1 fold), CL7278.Contig3_All (-4.3 fold) and Unigene8648_All (-7.5 fold) that were annotated as CHS were probably from endogenous *CHS* genes. Apart from flavonoid

and its related plant pathways, the expression level of numerous unigenes were also found in plant-pathogen interaction pathway (Table 3). Most of unigenes were belongs to plant resistance (R) genes and were significantly differential-expressed coded for different domains such as Leucine Rich Repeats (LRR), Amino acid domain (WRKY) and Kinase.

The reliability of the transcriptome profiling data was validated by RT-qPCR of five randomly selected unigenes (two up-regulated and three down-regulated unigenes). These unigenes were found to exhibit some discrepancies in fold change for up- and down-regulation in all comparison groups, however, similar expression patterns were also observed in all the selected unigenes compared to RNA-Seq results (Suppl. 5). The discrepancies in fold-change values were generated due to the use of different platforms (RT-qPCR and RNA-Seq) as different platforms use different formulas for fold change calculation.

Table 3. Numbers of DEGs in biosynthesis of flavonoid pathways and plant-pathogen interaction pathways

Pathway	Enzyme name	Unigene numbers	
		Up	Down
Flavonoid	Shikimate O-hydroxycinnamoyltransferase (HCT)	1	1
	6'-deoxychalcone synthase (DOCS)	1	-
	Chalcone synthase (CHS)	1	4
	Chalcone isomerase (CHI)	1	1
	Isoflavone reductase (IFR)	-	1
	Flavonol synthase (FLS)	-	1
	Naringenin 3-dioxygenase (F3H)	-	1
Flavone and Flavonol	Flavonol 3-O-methyltransferase	-	2
Isoflavonoid	Flavonoid 6-hydroxylase	-	1
Anthocyanin	None	-	-
Plant-pathogen interaction	Disease resistance protein RPM1	6	5
	Disease resistance protein RPS2	4	-
	DNA-directed RNA polymerases I, II, and III subunit RPABC1	7	-
	Interleukin-1 receptor-associated kinase 4	7	7
	Calcium-binding protein CML	5	3
	WRKY transcription factor 33	5	2
	WRKY transcription factor 22	2	-
	WRKY transcription factor 25	1	-
	Phytochrome-interacting factor 3	2	-
	Mitogen-activated protein kinase	2	1

brassinosteroid insensitive 1-associated receptor kinase 1	4	-
Serine/threonine-protein kinase PBS1	1	1
Calcium-dependent protein kinase	1	1
Hypothetical protein kinase	2	-
Aprataxin	-	1
Guanine nucleotide-exchange factor	1	-
Jasmonate ZIM domain-containing protein	3	1
aarF domain-containing kinase	-	1
ubiquitin-conjugating enzyme E2 H	1	-
Extracellular signal-regulated kinase 1/2	2	-
Calmodulin	2	1
somatic embryogenesis receptor kinase 1	-	1
Leucine-rich repeat (LRR) protein	3	-
LRR receptor-like serine/threonine-protein kinase FLS2	7	3
Transcription factor MYC2	2	1

4. DISCUSSIONS

With the help of the transcriptomic analysis, the potential genes can be elucidated and specialized pathways of secondary metabolite biosynthesis in non-model medicinal plants can be explored. In this study, RNA-Seq platform was utilised to analyse gene regulatory network in cell suspension cultures overexpressed with *BrCHS2* transcript. About 3 Gbps of clean reads and a total of 102,195 unigenes were generated, of which 67,251 (65.8%) were successfully annotated against public databases, suggesting the presence of their relatively conserved functions. However, the remaining unigenes (34.2% of the assembled unigenes) were absent or failed to generate significant homology to existing genes due to limited homologous sequences in the public databases. Also, there was a large proportion (34.45%) of unigenes that were shorter than 500bp, that were too short to allow statistically meaningful matches. This elucidated that there was more potential unigenes which remained unannotated and might fill in the gaps in the regulation networks among differentially regulated genes generated in this study.

Further analysis of the transcriptomic datagroups of *B. rotunda* cell suspension cultures harbouring *BrCHS2* transgene with significantly higher amounts of pinocembrin chalcone was carried out. In this study, DEGs were analysed and potential candidate genes that were responsible for the metabolite production were identified. As expected, the *BrCHS2* transgene (CL3398.Contig3_All) was found to be up-regulated in the flavonoid biosynthesis pathway as it yielded a significantly higher amount of pinocembrin chalcone in transgenic than that of the wild-type (Suppl. 1). However, endogenous *BrCHS* were found to be down-regulated and this finding was supported by the previous prediction on the introduction of a transgene that caused unintended modifications in transcription and translation of native genes in host plants (Kanobe et al., 2013).

The overall picture on distribution of differentially expressed genes manifested that there were relatively higher numbers of upregulated genes in the TL suspension cultures compared to WT as a result of *BrCHS2* overexpression.

However, further comparative analyses against the KEGG database highlighted the fact that for flavonoid biosynthesis and its related pathways (Flavone and Flavonol, Isoflavonoid and Anthocyanin pathway), there were actually more down-regulated genes. Overexpression of *BrCHS2* transcripts might directly switch on the expression of other genes in the other branching pathways as proposed by previous researchers (Tan et al., 2015; Yadnya-Putra et al., 2014). Reports on the transcriptomic analysis in phenylalanine-

treated suspension cultures of *B. rotunda* (Md-Mustafa et al., 2014; Tan et al., 2012), also showed possible branching pathways that could actually emerge after the production of pinocembrin chalcone. The existence of the yet to be uncharacterized plant biosynthesis pathways could be speculated to occur and may produce other novel bioactive compounds. This enables future effort to produce new and valuable flavonoids derived from pinocembrin chalcone compound.

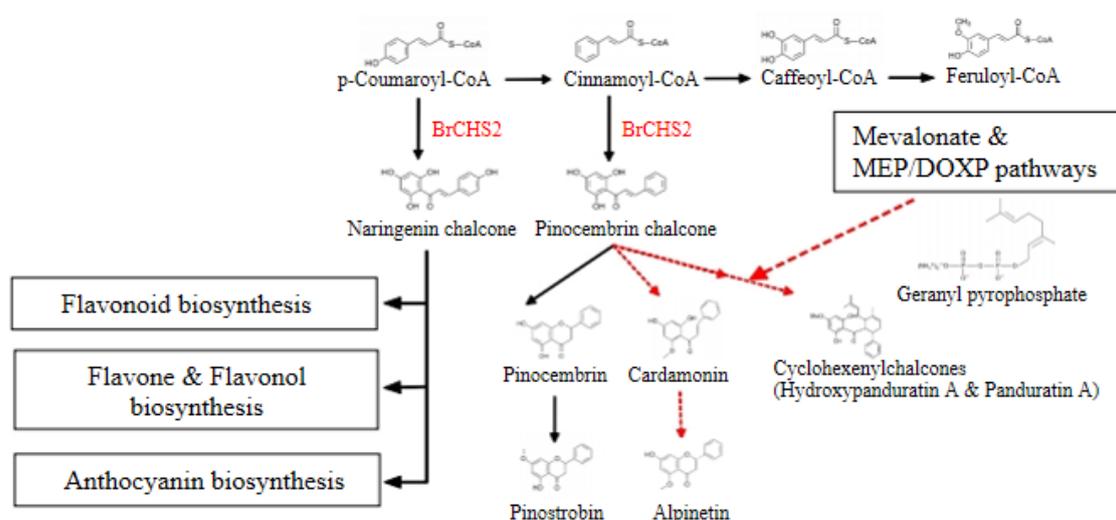


Figure 2. Proposed biosynthesis pathway after overexpression of *BrCHS2* transcript in *B. rotunda* cell suspension culture. The red dashed arrows show the potential biosynthesis pathways of secondary metabolites as proposed by Tan et al. (2015) and Yadnya-Putra et al. (2014).

Unigenes that were responsible for plant pathogen interaction pathway were also increased and thus play an important function on plant resistance. In the present study, transcriptomic analysis exhibited that *CHS*-accumulating lines induced a group of plant defence-related genes alongside transcription factors in the plant pathogen interaction pathway. Disease resistance proteins RPM1 and RPS2 have been abundantly expressed in this particular pathway with an up-regulation of DE genes for each of the protein in response to microbial/pathogen attack. In this condition, antimicrobial compounds such

as phytoalexins are produced where they are derived from different classes of metabolites such as flavonoids, isoflavonoids, stilbenoids, sesquiterpenoids, steroids and alkaloids (Angelova et al., 2010; Ibraheem et al., 2010; Malacarne et al., 2011; Sobolev et al., 2011; Cho & Lee, 2015; Li et al., 2015).

In addition, there is still interest in finding plant defence metabolites that could serve as a potential drug candidate for the inhibition of pathogenic activities from *L. monocytogenes*, *B. cereus*, *S. aureus*, *E. coli* and *S. Typhimurium* particularly in the

other Zingiberaceae species such as *Boesenbergia pandurata* Schltr. that exhibited strong inhibitory activity against these pathogens (Pattaratanawadee et al., 2006). Therefore, with the advancement of RNA-Seq technology, more comprehensive research on plant infections towards different types of pathogens could be conducted in the near future to reveal potential candidates of plant defence genes involved in various mechanisms to act against pathogens in Zingiberaceae species. In conclusion, potential genes responsible for flavonoid biosynthesis coupled with identification of plant defence-related genes in the genus *Boesenbergia* mined from the transcriptomic data could give valuable addition of genomic data to the publicly available *B. rotunda* especially for non-model organisms that lack genomic sequences.

5. ACKNOWLEDGEMENTS

We hereby acknowledge the financial supports from University of Malaya through these research grants FL 002-2010 and PV085-2012.

6. REFERENCES

- Angelova, S., Buchheim, M., Frowitter, D., Schierhorn, A., & Roos, W. (2010). Overproduction of alkaloid phytoalexins in california poppy cells is associated with the co-expression of biosynthetic and stress-protective enzymes. *Molecular Plant*, 3(5), 927–939.
- Audic, S., & Claverie, J. M. (1997). The significance of digital gene expression profiles. *Genome Research*, 7(10), 986–995.
- Cheniary, M., Ebrahimzadeh, H., & Masoudi-nejad, A. (2012). Expression of chalcone synthase influences flavonoid content and frequency of rhizogenesis in microshoots of *Juglans regia* L. *Plant Cell, Tissue and Organ Culture*, 109, 51–59.
- Cho, M. H., & Lee, S. W. (2015). Phenolic phytoalexins in rice: Biological functions and Biosynthesis. *International Journal of Molecular Sciences*, 16(12), 29120–29133.
- Dao, T. T. H., Linthorst, H. J. M., & Verpoorte, R. (2011). Chalcone synthase and its functions in plant resistance. *Phytochemistry Reviews*, 10(3), 397–412.
- Dare, A. P., Tomes, S., Jones, M., McGhie, T. K., Stevenson, D. E., Johnson, R. a., ... Hellens, R. P. (2013). Phenotypic changes associated with RNA interference silencing of *chalcone synthase* in apple (*Malus domestica*). *Plant Journal*, 74, 398–410.
- Deng, X., Bashandy, H., Ainasoja, M., Kontturi, J., Pietiäinen, M., Laitinen, R. A. E., ... Teeri, T. H. (2014). Functional diversification of duplicated chalcone synthase genes in anthocyanin biosynthesis of *Gerbera hybrida*. *New Phytologist*, 201, 1469–1483.
- Der, J. P., Barker, M. S., Wickett, N. J., DePamphilis, C. W., & Wolf, P. G. (2011). De novo characterization of the gametophyte transcriptome in bracken fern, *Pteridium aquilinum*. *BMC Genomics*, 12(1), 99.
- Dukowic-Schulze, S., Harris, A., Li, J., Sundararajan, A., Mudge, J., Retzel, E. F., ... Chen, C. (2014).

- Comparative transcriptomics of early meiosis in arabidopsis and maize. *Journal of Genetics and Genomics*, 41(3), 139–152.
- Franssen, S., Shrestha, R., Brautigam, A., Bauer, E., & Weber, A. (2011). Comprehensive transcriptome analysis of the highly complex *Pisum sativum* genome using next generation sequencing. *BMC Genomics*, 12(1), 227.
- Garg, R. O., Patel, R. A. V. I. K., Tyagi, A. K. K., & Jain, M. U. (2011). *De novo* assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Research*, 18, 53–63.
- Guo, L., Gao, Z., & Qian, Q. (2014). Application of resequencing to rice genomics, functional genomics and evolutionary analysis. *Rice*, 7(4), 1–10.
- Hao, D. C., Ge, G., Xiao, P., Zhang, Y., & Yang, L. (2011). The first insight into the tissue specific taxus transcriptome via Illumina second generation sequencing. *PloS One*, 6(6), e21220.
- Ibraheem, F., Gaffoor, I., & Chopra, S. (2010). Flavonoid phytoalexin-dependent resistance to anthracnose leaf blight requires a functional yellow seed1 in *Sorghum bicolor*. *Genetics*, 184(4), 915–926.
- Jiang, C., Schommer, C. K., Kim, S. Y., & Suh, D.-Y. (2006). Cloning and characterization of chalcone synthase from the moss, *Physcomitrella patens*. *Phytochemistry*, 67(23), 2531–2540.
- Kanobe, M. N., Rodermel, S. R., Bailey, T., & Scott, M. P. (2013). Changes in endogenous gene transcript and protein levels in maize plants expressing the soybean ferritin transgene. *Frontiers in Plant Science*, 4, 196.
- Kirana, C., Jones, G. P., Record, I. R., & McIntosh, G. H. (2007). Anticancer properties of panduratin A isolated from *Boesenbergia pandurata* (Zingiberaceae). *Journal of Natural Medicines*, 61, 131–137.
- Li, R., Tee, C.-S., Jiang, Y.-L., Jiang, X.-Y., Venkatesh, P. N., Sarojam, R., & Ye, J. (2015). A terpenoid phytoalexin plays a role in basal defense of *Nicotiana benthamiana* against potato virus X. *Scientific Reports*, 5, 9682.
- Liu, X. J., Chuang, Y. N., Chiou, C. Y., Chin, D. C., Shen, F. Q., & Yeh, K. W. (2012). Methylation effect on chalcone synthase gene expression determines anthocyanin pigmentation in floral tissues of two *Oncidium* orchid cultivars. *Planta*, 236, 401–409.
- Liu, Y., Lou, Q., Xu, W., Xin, Y., Bassett, C., & Wang, Y. (2011). Characterization of a chalcone synthase (CHS) flower-specific promoter from *Lilium oriental* “Sorbonne.” *Plant Cell Reports*, 30(12), 2187–2194.
- Md-Mustafa, N. D., Khalid, N., Gao, H., Peng, Z., Alimin, M. F., Bujang, N., Othman, R. Y. (2014). Transcriptome profiling shows gene regulation patterns in a flavonoid pathway in response to exogenous phenylalanine in *Boesenbergia rotunda* cell culture. *BMC Genomics*, 15(1), 984.

- Malacarne, G., Vrhovsek, U., Zulini, L., Cestaro, A., Stefanini, M., Mattivi, F., ... Moser, C. (2011). Resistance to *Plasmopara viticola* in a grapevine segregating population is associated with stilbenoid accumulation and with specific host transcriptional responses. *BMC Plant Biology*, 11(1), 114.
- Murashige, T. and F. S. (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiologia Plantarum*, 15, 473–497.
- Pang, Y., Shen, G., Wu, W., Liu, X., Lin, J., Tan, F., ... Tang, K. (2005). Characterization and expression of chalcone synthase gene from *Ginkgo biloba*. *Plant Science*, 168(6), 1525–1531.
- Pattaratanawadee, E., Thongson, C., Mahakarnchanakul, W., & Wanchaitanawong, P. (2006). Antimicrobial activity of spice extracts against pathogenic and spoilage microorganisms. *The Proceedings of the 44th Kasetsart University Annual Conference*, Kasetsart, Subject: Agro-Industry, Economics, Business Administration, 165, 297–307.
- Shindo, K., Kato, M., Kinoshita, A., Kobayashi, A., & Koike, Y. (2006). Analysis of antioxidant activities contained in the *Boesenbergia pandurata* Schult. Rhizome. *Bioscience, Biotechnology, and Biochemistry*, 70(9), 2281–2284.
- Sobolev, V. S., Khan, S. I., Tabanca, N., Wedge, D. E., Manly, S. P., Coy, S. J. C., ... Gloer, B. S. A. N. J. B. (2011). Biological activity of peanut (*Arachis hypogaea*) phytoalexins and selected natural and synthetic stilbenoids. *Journal of Agricultural and Food Chemistry*, 59(5), 1673–1682.
- Sun, W., Meng, X., Liang, L., Jiang, W., Huang, Y., He, J., ... Wang, L. (2015). Molecular and biochemical analysis of chalcone synthase from *Freesia hybrida* in flavonoid biosynthetic pathway. *PLoS ONE*, 10(3), 1–18.
- Tan, B. C., Tan, S. K., Wong, S. M., Ata, N., Rahman, N. A., & Khalid, N. (2015). Distribution of flavonoids and cyclohexenyl chalcone derivatives in conventional propagated and in vitro-derived field-grown *Boesenbergia rotunda* (L.) Mansf. *Evidence-Based Complementary and Alternative Medicine*, 1–7.
- Tan, E. C., Karsani, S. A., Foo, G. T., Wong, S. M., Abdul Rahman, N., Khalid, N., ... Yusof, R. (2012). Proteomic analysis of cell suspension cultures of *Boesenbergia rotunda* induced by phenylalanine: Identification of proteins involved in flavonoid and phenylpropanoid biosynthesis pathways. *Plant Cell, Tissue and Organ Culture*, 111(2), 219–229.
- Tan, S. K., Pippen, R., Yusof, R., Ibrahim, H., Khalid, N., & Rahman, N. A. (2006). Inhibitory activity of cyclohexenyl chalcone derivatives and flavonoids of fingerroot, *Boesenbergia rotunda* (L.), towards dengue-2 virus NS3 protease. *Bioorganic and Medicinal Chemistry Letters*, 16(12), 3337–3340.
- Tewtrakul, S., Subhadhirasakul, S., Karalai, C., Ponglimanont, C., & Cheenpracha, S. (2009). Anti-

- inflammatory effects of compounds from *Kaempferia parviflora* and *Boesenbergia pandurata*. *Food Chemistry*, 115(2), 534–538.
- Trakoontivakorn, G., Nakahara, K., Shinmoto, H., Takenaka, M., Onishi-Kameyama, M., Ono, H., Tsushida, T. (2001). Structural analysis of a novel antimutagenic compound, 4-Hydroxypanduratin A, and the antimutagenic activity of flavonoids in a Thai spice, fingerroot (*Boesenbergia pandurata* Schult.) against mutagenic heterocyclic amines. *Journal of Agricultural and Food Chemistry*, 49(6), 3046–3050.
- Trojan, V., Musilová, M., Vyhnánek, T., Klejdus, B., Hanáček, P., & Havel, L. (2014). *Chalcone synthase* expression and pigments deposition in wheat with purple and blue colored caryopsis. *Journal of Cereal Science*, 59(1), 48–55.
- Wakasa, Y., Oono, Y., Yazawa, T., Hayashi, S., Ozawa, K., Handa, H., Takaiwa, F. (2014). RNA sequencing-mediated transcriptome analysis of rice plants in endoplasmic reticulum stress conditions. *BMC Plant Biology*, 14(1), 101.
- Wang, H., Chung, P. J., Liu, J., Jang, I. C., Kean, M. J., Xu, J., & Chua, N. H. (2014). Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. *Genome Research*, 24, 444–453.
- Wong, S. M., Salim, N., Harikrishna, J. A., & Khalid, N. (2013). Highly efficient plant regeneration via somatic embryogenesis from cell suspension cultures of *Boesenbergia rotunda*. *In Vitro Cellular and Developmental Biology - Plant*, 49(6), 665–673.
- Yadnya-Putra, A. A. G. R., Chahyadi, A., & Elfahmi. (2014). Production of panduratin A, cardamomin and sitosterol using cell cultures of fingerroot (*Boesenbergia pandurata* (Roxb.) schlechter)). *Biosciences Biotechnology Research Asia*, 11(1), 43–52.

Supplementary 1. Retention time of six reference compounds were listed in Suppl. 1a. Selection of transgenic *B. rotunda* suspension cultures based on enzymatic activity. Three transgenic lines that showing highest detection amount of pinocembrin chalcone (TL3, TL10 and TL9) during HPLC analysis were chosen for further subcultures prior to transcriptome profiling (Suppl. 1b).

Table of Suppl 1a.

Reference compounds	Retention time (min)
Naringenin	6.531±0.10
Naringenin chalcone	6.571±0.10
Pinocembrin	10.304±0.10
Pinocembrin chalcone	8.913±0.10
Pinostrobin	12.107±0.10
Pinostrobin chalcone	10.727±0.10

Table of Suppl 1b.

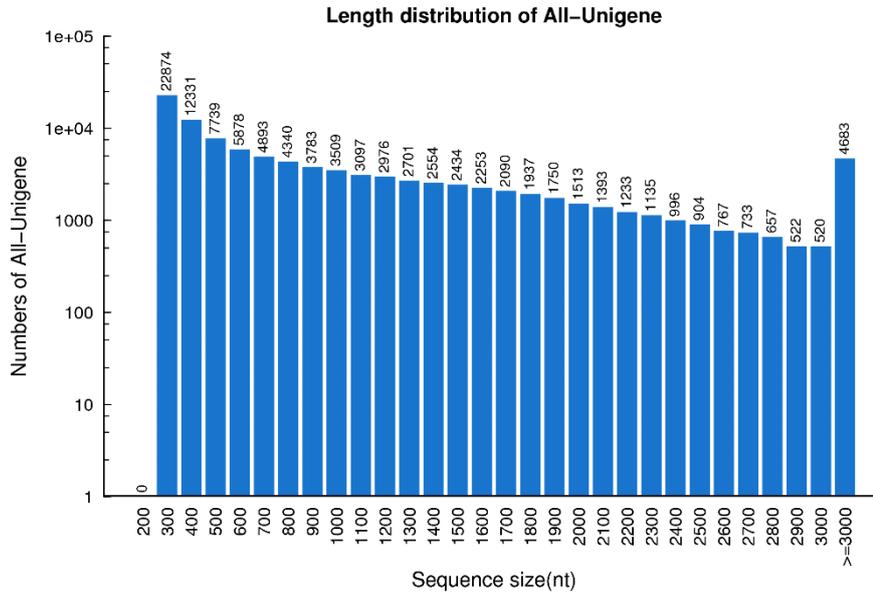
Compound	Pinocembrin Chalcone	
Retention time (min)	8.90±0.10	
Suspension cultures	Amount (mg/g)	S.E.
WT	64.67	6.67E-02
TL1	154.43*	7.42E-01
TL2	112.20	1.00E-01
TL3	275.63*	7.69E+00
TL5	131.83*	2.50E+00
TL9	207.30*	5.47E+00
TL10	231.03*	1.83E+01

* Significant at $p < 0.001$

Supplementary 2. Primers' sequences that used for RNA-Seq validation

Primers	Sequence (5'-3')
Unigene 28275_All_F	5' - ACTGTGGAGGCAGTGCTC - 3'
Unigene 28275_All_R	5' - TCCATTGATGCACAAGATTCCA - 3'
CL4317.Contig3_All_F	5' - CTACTTTATGGTGCAAATCTCAC - 3'
CL4317.Contig3_All_R	5' - CGGAAGCACCAAACTOCTCCTTCAT - 3'
CL1414.Contig2_All_F	5' - TTCACCATCAAGACCGCCATCG - 3'
CL1414.Contig2_All_R	5' - CATATCTGCAGCCTCCGGGTTGTC - 3'
Unigene 5234_All_F	5' - ATGGCTCCGGGTGCGACGGTT - 3'
Unigene 5234_All_R	5' - CCTCTTCCCTCCGTCACCACTGG - 3'
CL12450.Contig1_All_F	5' - CTCCCTTTCACCATCTTTGACATAG - 3'
CL12450.Contig1_All_R	5' - AGTGGGTAAAAGTCCCGGAGAGTGT - 3'

Supplementary 3. Histogram of the unigenes length distribution



Supplementary 4. Multiple sequence alignment of CL3398.Contig3_ALL and BrCHS2 with * denoted both sequences having the same amino acids. The Multiple Sequence Alignment (MSA) was performed using CLUSTAL O (1.2.4).

```

CL3398.Contig3_All      FFFSFLSINCSLRLFTSHKE*IPAIKRRGNRGRIHTYITMAKVQEIQRQRRAEGPAAIL
BrCHS2                  -----MAKVQEIQRQRRAEGPAAIL
                          *****

CL3398.Contig3_All      AIGTATPTNVVYQADYADYYFRITKSEHLTELKEKFKRMC DKSMIRKRYMHVTEEILKEN
BrCHS2                  AIGTATPTNVVYQADYADYYFRITKSEHLTELKEKFKRMC DKSMIRKRYMHVTEEILKEN
                          *****

CL3398.Contig3_All      PNMCAYPEPSLDERQDIVVVEVPKLGKEAAAKAIKEWGQPKSKITHLIFCTTSGVDMPGA
BrCHS2                  PNMCAYPEPSLDERQDIVVVEVPKLGKEAAAKAIKEWGQPKSKITHLIFCTTSGVDMPGA
                          *****

CL3398.Contig3_All      DYQITKLLGLRPSVNRFFMYQQGCFAGGTVLRLAKDLAENNRGARVLVVCSEITAVTFRG
BrCHS2                  DYQITKLLGLRPSVNRFFMYQQGCFAGGTVLRLAKDLAENNRGARVLVVCSEITAVTFRG
                          *****

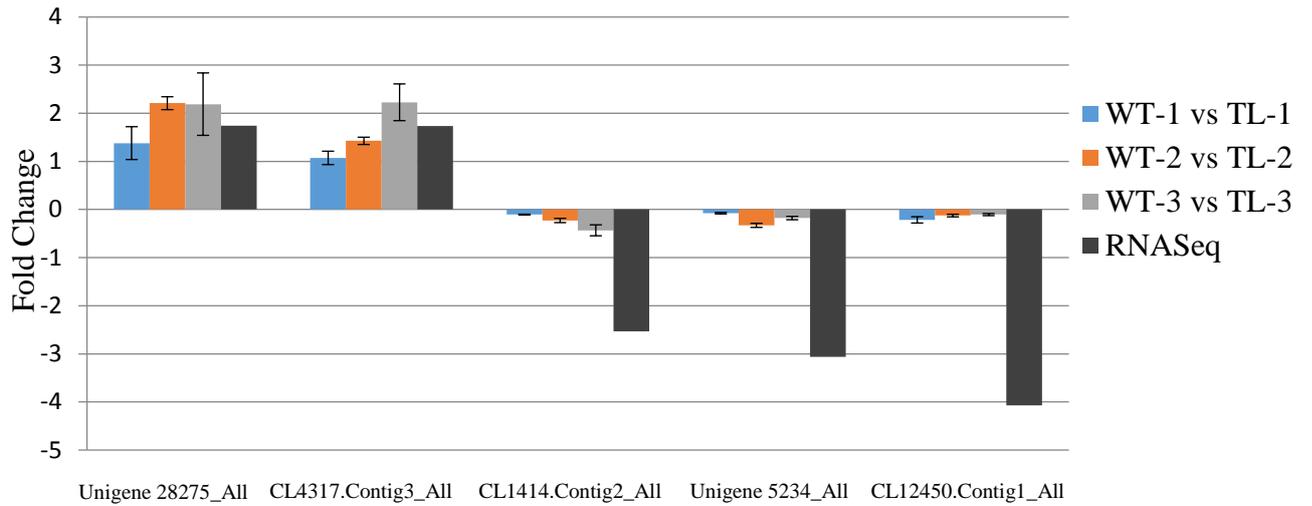
CL3398.Contig3_All      PSESHLDSL VGQALFGDGAGAIIVGADPDLETERPLFELVSASQTILPDSEGAIDGHLRE
BrCHS2                  PSESHLDSL VGQALFGDGAGAIIVGADPDLETERPLFELVSASQTILPDSEGAIDGHLRE
                          *****

CL3398.Contig3_All      VGLTFHLLKDVPLGISKNIEKSLVEAFAPL GIDDWNSIFWIAHPGGPAILDQVEAKLALE
BrCHS2                  VGLTFHLLKDVPLGISKNIEKSLVEAFAPL GIDDWNSIFWIAHPGGPAILDQVEAKLALE
                          *****

CL3398.Contig3_All      KEKMAATRQVLSEYGNMSSACVIFILDEMRRKSAQEGKATTGEGLNWGVLF GFGPGLTVE
BrCHS2                  KEKMAATRQVLSEYGNMSSACVIFILDEMRRKSAQEGKATTGEGLNWGVLF GFGPGLTVE
                          *****

CL3398.Contig3_All      TVVLHSKPINHTSKGEELFTGVVPIILVELDGDVNGHKFSVSGEGEGDATYKGLTLKFICT
BrCHS2                  TVVLHSKPINHXX-----
                          *****

CL3398.Contig3_All      TGKLPVPWPTLVTTFSYGVCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTR
BrCHS2                  -----
    
```



Supplementary 5. Expression pattern validation of selected unigenes by RT-qPCR in three groups. Fold-change in mRNA transcript levels of five randomly selected unigenes represented by blue, orange and grey bars. Black bars indicated the average fold-change value for each unigene that was calculated according to RNA-Seq analysis.