Sains Malaysiana 47(12)(2018): 2993–3002
http://dx.doi.org/10.17576/jsm-2018-4712-08

Reconstruction of the Transcriptional Regulatory Network in *Arabidopsis thaliana* Aliphatic Glucosinolate Biosynthetic Pathway

(Pembinaan Semua Jaringan Pengawal Atur Transkripsi Tapak Jalan Biosintesis Glukosinolat Alifatik dalam *Arabidopsis thaliana*)

KHALIDAH-SYAHIRAH ASHARI, MUHAMMAD-REDHA ABDULLAH-ZAWAWI, SARAHANI HARUN & ZETI-AZURA MOHAMED-HUSSEIN*

ABSTRACT

Aliphatic glucosinolate is an important secondary metabolite responsible in plant defense mechanism and carcinogenic activity. It plays a crucial role in plant adaptation towards changes in the environment such as salinity and drought. However, in many plant genomes, there are thousands of genes encoding proteins still with putative functions and incomplete annotations. Therefore, the genome of Arabidopsis thaliana *was selected to be investigated further to identify any putative genes that are potentially involved in the aliphatic glucosinolate biosynthesis pathway, most of its gene are with incomplete annotation. Known genes for aliphatic glucosinolates were retrieved from KEGG and AraCyc databases. Three co-expression databases i.e., ATTED-II, GeneMANIA and STRING were used to perform the co-expression network analysis. The integrated co-expression network was then being clustered, annotated and visualized using Cytoscape plugin, MCODE and ClueGO. Then, the regulatory network of* A. thaliana *from AtRegNet was mapped onto the co-expression network to build the transcriptional regulatory network. This study showed that a total of 506 genes were co-expressed with the 61 aliphatic glucosinolate biosynthesis genes. Five transcription factors have been predicted to be involved in the biosynthetic pathway of aliphatic glucosinolate, namely SEPALLATA 3 (SEP3), PHYTOCHROME INTERACTING FACTOR 3-like 5 (AtbHLH15/PIL5), ELONGATED HYPOCOTYL 5 (HY5), AGAMOUS-like 15 (AGL15) and GLABRA 3 (GL3). Meanwhile, three other genes with high potential to be involved in the aliphatic glucosinolates biosynthetic pathway were identified, i.e., methylthioalkylmalate-like synthase 4 (MAML-4) and aspartate aminotransferase (ASP1 and ASP4). These findings can be used to complete the aliphatic glucosinolate biosynthetic pathway in* A. thaliana *and to update the information on the glucosinolate-related pathways in public metabolic databases.*

Keywords: Aliphatic glucosinolate biosynthesis; co-expression analysis; regulatory network

ABSTRAK

Glukosinolat alifatik merupakan metabolit sekunder penting di dalam mekanisme pertahanan tumbuhan dan aktiviti karsinogen. Glukosinolat juga penting di dalam penyesuaian terhadap persekitaran seperti kemasinan dan kemarau. Namun begitu dalam kebanyakan genom tumbuhan, masih banyak fungsi gen yang mengekod protein adalah putatif dan tidak lengkap. Oleh itu, genom Arabidopsis thaliana *telah dipilih untuk dikaji dengan lebih mendalam untuk mengenal pasti gen putatif yang berpotensi terlibat di dalam tapak jalan biosintesis glukosinolat alifatik. Gen biosintetik glukosinolat alifatik telah dikumpul daripada pangkalan data KEGG dan AraCyc manakala pangkalan data ATTED-II, GeneMANIA dan STRING digunakan dalam analisis pengekspresan bersama. Integrasi jaringan pengekspresan bersama telah dilakukan dengan menggunakan perisian Cytoscape, MCODE dan ClueGO. Kesemua gen pengekspresan bersama yang terlibat dipetakan menggunakan set data jaringan pengawal atur daripada pangkalan data AtRegNet. Hasil kajian ini berjaya mengenal pasti 506 gen yang telah diekspreskan bersama dengan 61 gen biosintetik glukosinolat alifatik. Lima faktor transkripsi telah berjaya dikenal pasti dan didapati terlibat di dalam mengawal atur biosintetis glukosinolat alifatik iaitu SEPALLATA 3 (SEP3), PHYTOCHROME INTERACTING FACTOR 3-like 5 (AtbHLH15/PIL5), ELONGATED HYPOCOTYL 5 (HY5), AGAMOUS-like 15 (AGL15) dan GLABRA 3 (GL3). Kajian ini mengukuhkan lagi penglibatan gen berpotensi di dalam tapak jalan biosintesis glukosinolat alifatik melalui penemuan gen methylthioalkylmalate-like synthase 4 (MAML-4) dan aspartate aminotransferase (ASP4 dan ASP1) menggunakan kaedah yang telah dijalankan.*

Kata kunci: Analisis pengekspresan bersama; biosintesis glukosinolat alifatik; jaringan pengawal atur

INTRODUCTION

Glucosinolates are secondary metabolites that contain the core structure of β-D-thioglucose group linked to the sulphonated aldoxime moiety and amino acid side chain (Redovniković et al. 2012). They are mostly found in *Brassicaceae* family (Sønderby et al. 2010). These metabolites are divided into three groups, i.e., aromatic, aliphatic and indolic (Redovniković et al. 2012).

Aromatic glucosinolates are derived from phenylalanine and tyrosine, indolic glucosinolates are derived from tryptophan and aliphatic glucosinolates are derived from methionine, alanine, leucine, isoleucine and valine (Halkier & Gershenzon 2006).

The accumulation of aliphatic glucosinolates can be divided into three main phases; i.e., the elongation of amino acid side chain, followed by the formation of core structure of glucosinolate and the secondary modification to synthesize aliphatic glucosinolates with specific function (Nour-Eldin & Halkier 2009). The synthesized aliphatic glucosinolates are important not only in plant growth and development, they also play an important role in the adaptation towards environmental stress such as salt toleration adaptation in salinity (Martínez-Ballesta et al. 2015). Moreover, aliphatic glucosinolates are also nutrient elements that are important in carcinogenic activity (Rameeh 2015) and plant defenses (Redovniković et al. 2008). When the plant tissues are being attacked by the herbivores, it was found that glucosinolates will activate myrosinase enzyme (Rohr et al. 2009). This enzyme hydrolyzes aliphatic glucosinolates and produces several toxic compounds such as isothiocyanate, thiocyanate, nitrile and epithionitrile (Halkier & Gershenzon 2006). Besides, the hydrolysis of aliphatic glucosinolates also produce other toxic compounds such as sulforaphane and sinigrin that inhibit proliferation and catalyze the apoptosis of cancer cells hence reducing the risk of cancer (Atwell et al. 2015).

To date, studies on the biosynthesis of aliphatic glucosinolates are still poorly understood. This is because *A. thaliana* genome contains thousands of genes encoding protein with unknown functions and incomplete annotation. Incomplete annotation limits the genome-scale metabolic model prediction activity (Bradbury et al. 2013). To address this issue, this study was carried out to identify these genes and to predict if they have the potential to be involved in the aliphatic glucosinolates biosynthetic pathway using genome-context analysis such as co-expression and regulatory analysis. In this study, the model plant *A. thaliana* was used as its leaves have high content of aliphatic glucosinolates (Beekwilder et al. 2008). Transcriptomic data were utilized in the gene expression analysis and the co-expression network of aliphatic glucosinolate biosynthesis was built using three public databases, ATTED-II (Obayashi et al. 2007), GeneMANIA (Warde-Farley et al. 2010) and STRING (Szklarczyk et al. 2017). Further analyses were also conducted that include gene clustering using MCODE (Bader & Hogue 2003) and functional enrichment by ClueGO (Bindea et al. 2009) in Cytoscape. According to Serin et al. (2016), genes that have similar function or involve in the same regulatory pathway tend to share similar expression pattern, forming modules in the network. From those modules, the known genes can be used to predict the function of co-expressed putative genes (van Dam et al. 2017). The transcriptional regulatory network aliphatic glucosinolate biosynthesis was constructed to support other biological information and to identify the functional relationship between the regulatory genes and the candidate genes.

Materials and Methods

Gene Mining

Kyoto Encyclopedia of Genes and Genomes (KEGG) version 80.0 (Kanehisa et al. 2016) was used in finding the aliphatic glucosinolate biosynthetic pathway. AraCyc version 14.0 (Mueller et al. 2003) database was used in searching for the aliphatic glucosinolate genes in *A. thaliana*. List of genes from both databases were compared, combined and integrated to eliminate the redundant genes.

Co-expression Network Analysis

Co-expression network analysis was conducted to identify the genes that are possibly involved in the aliphatic glucosinolate biosynthesis. Li et al. (2015) stated that genes having similar expression profiles are usually connected by the interaction in the network. The co-expressed genes are predicted to have the same roles and involved in the same biological processes (Roy et al. 2016). All genes obtained from KEGG and AraCyc databases were used as an input to generate the co-expression network using three co-expression databases; i.e., GeneMANIA (Warde-Farley et al. 2010), STRING version 10.0 (Szklarczyk et al. 2017) and ATTED-II version 8.0 (Obayashi et al. 2007). Raw data in these databases were obtained from the expression data repository known as Gene Expression Omnibus (GEO). In addition, in GeneMANIA, STRING and ATTED-II, microarray data were used to calculate the gene expression. The constructed co-expression networks and other expression information such as expression level were then being visualized, integrated and analyzed using Cytoscape (Shannon et al. 2003) software.

Gene Clustering

Genes in the integrated co-expression network were clustered based on the guilt-by-association principle where genes that were co-expressed with the aliphatic glucosinolate biosynthesis genes in the same module, do share the same biological function or involve in the same metabolic process (Nützmann et al. 2016). This co-expression network was clustered using Cytoscape plugin, MCODE (Bader & Hogue 2003). MCODE clusters the genes by identifying the region based on the interaction density in the network and this region may represent the molecular complex (Mao et al. 2009). The parameters in the clustering analysis were set as default whereby both degree cutoff and k-value were set to 2.0. The interactions with k-value < 2.0 were eliminated and modules with scores > 2.0 and have at least three nodes were selected for the functional enrichment analysis.

FUNCTIONAL ENRICHMENT

A Cytoscape plugin, ClueGO (Bindea et al. 2009) was applied to perform the functional enrichment of the modules and it also provides interactive visualization. Hypergeometric method was selected for statistical test and the generated p-values were corrected with the Bonferroni method where the cut-off value was <0.05. Cluepedia was also employed in this analysis in order to visualize the involvement of the genes in the biological process.

RECONSTRUCTION OF THE REGULATORY NETWORK

Microarray data were integrated with the regulatory information where co-expressed genes in the microarray analysis tend to share similar *cis*-regulatory elements (Serin et al. 2016). Thus, the target genes that are regulated by the same transcription factor with the aliphatic glucosinolate genes were also predicted to be involved in the same biological process. The transcriptional regulatory data were retrieved from *A. thaliana* Regulatory Network (AtRegNet) database (Yilmaz et al. 2011) that stores *A. thaliana* transcriptional regulatory data. List of the transcription factors and their target genes were downloaded, visualized and mapped onto the co-expression network using Cytoscape and the reconstructed regulatory network was then being analyzed.

RESULTS AND DISCUSSION

GENE MINING

A total of 61 genes involved in the aliphatic glucosinolate biosynthesis pathway were identified where four of them were retrieved from KEGG database and 46 of them were identified from AraCyc database. Meanwhile, eleven of them were found from both metabolic databases. These 61 aliphatic glucosinolate genes consist of 13 genes in the elongation of amino acid side chain phase, 14 genes in the formation of glucosinolate core structure phase and 34 genes in the secondary modification phase (Table 1).

According to Redovniković et al. (2012), aliphatic glucosinolate biosynthesis can be divided into three main phases. The first phase is the elongation of amino acid side chain followed by the formation of core structure glucosinolate and the secondary modification of glucosinolate side chain. Thirteen of them were found involved in the elongation of amino acid side chain. This phase includes the deamination of amino acid to 2-oxo acid catalyzed by the BCAT genes (Lächler et al. 2015). BCAT1 encodes the mitochondrial BCAT (Schuster et al. 2006) while BCAT2, BCAT3, and BCAT5 encode the chloroplast BCAT in valine-, leucine- and isoleucine-derived of aliphatic glucosinolates (Binder 2010). BCAT4 encodes the cytosol BCAT, functioning in the elongation of amino acid side chain methionine-derived of aliphatic glucosinolate (Lächler et al. 2015). The 2-oxo acid then undergoes the transformation cycle including the condensation of acetyl CoA by methylthioalkylmalate synthase (MAM), isomerization of isopropyl isomerase (IPMI2) and oxidative decarboxylation by isopropylmalate dehydrogenase (IMD1, IMD2, IMD3, AT1G80555.1 and IIL1) (Field et al. 2006). MAM1 catalyzes the condensation in elongation of the C1 and C3 of methionine (Sønderby et al. 2010). Meanwhile, MAM3 catalyzes the condensation in elongation from C1 until C6 (Textor et al. 2007).

The elongated side chain of amino acid is then subjected to the formation of glucosinolate core structure involving 14 identified aliphatic glucosinolate genes. This phase starts with the conversion of amino acid to aldoxime by cytochrome P450 (CYP79) (Mostafa et al. 2017). CYP79F1 metabolizes all the elongated methione-derived side chain and CYP79F2 only metabolizes penta-homomethione and hexa-homo-methionine (Redovniković et al. 2008). Aldoximes are then being oxidized by CYP83A1 into the active form and transformed into thiohydroximates through conjugation of glutathione and C-S lyase (SUR1) reaction (Grubb & Abel 2006). The thiohydroximates are then converted into the glucosinolate structure catalyzed by S-glucosyltransferase (UGT74B1 and UGT74F1) and sulfotransferase (SOT16,

TABLE 1. List of genes involved in different aliphatic glucosinolate biosynthesis phases

Biosynthesis phase	Gene involved
Elongation of amino acid side chain	*AT1G80555.1, BCAT1, BCAT2, BCAT3, BCAT4, BCAT5, IIL1, IMD1, IMD2, IMD3, IPMI2, MAM1, MAM3*
Formation of core structure glucosinolate	*AT4G28410.1, AT4G28420.2, CYP79C1, CYP79C2, CYP79F1, CYP79F2, CYP83A1, SOT16, SOT17, SOT18, SUR1, TAT3, UGT74B1, UGT74F1*
Secondary modification	*2A6, AOP2, AOP3, AT1G03400.1, AT1G04350.1, AT1G04380.1, AT1G06620.1, AT1G06640.1, AT1G06645.1, AT1G06650.2, AT1G12130.1, AT1G12160.1, AT1G62600.1, AT1G62620.1, AT1G63340.1, AT1G63370.1, AT1G63390.1, AT2G25450.1, AT2G30830.1, AT2G30840.1, AT3G61400.1, AT5G07800.1, AT5G43440.1, AT5G43450.1, AT5G59530.1, AT5G59540.1, AT5G61290.1, FMO, FMO_GS-OX1, FMO_GS-OX2, FMO_GS-OX3, FMO_GS-OX4, FMO_GS-OX5, NOGC1*

SOT17 and SOT18) (Ishida et al. 2014). In *A. thaliana*, aliphatic glucosinolate can be modified into six different glucosinolates (Nour-Eldin & Halkier 2009). This phase includes the oxidation of methylthio sulphur to methylsulfinyl by flavin monooxygenase (FMOGS-OX1, FMOGS-OX2, FMOGS-OX3, FMOGS-OX4 and FMOGS-OX5) in methylsulfinyl glucosinolate biosynthesis (Sønderby et al. 2010). Meanwhile, S-oxygenated glucosinolate can be converted into hydroxyalkyl or into alkenyl glucosinolate by AOP2 and AOP3 genes (Neal et al. 2010).

CO-EXPRESSION NETWORK AND CLUSTER ANALYSIS

The aliphatic glucosinolate biosynthesis genes were then used as a query in co-expression analysis. The co-expression networks from ATTED-II, GeneMANIA and STRING were visualized and integrated using Cytoscape 3.4.0 software, yielding a network consisting of 506 genes that co-expressed with 61 aliphatic glucosinolate genes (Figure 1). Clustering analysis employed by MCODE showed a total of 37 modules in the network. However, the main focus was paid to the modules containing aliphatic glucosinolate genes that were mined before, with the assumption that genes that co-expressed with those aliphatic glucosinolate genes in the same module, potentially share similar function in the same biological pathway. Four modules which are Module 1, Module 2, Module 3 and Module 4 score of 30.0, 19.103, 17.034 and 14.8, respectively (Figure 2).

The co-expression network was then being clustered using MCODE plugin yielding four modules containing the known genes (Figure 2 & Table 2). The results showed that Module 1 had highly dense interconnected region compared to the other modules. Genes from Module 1 were found to co-express with four genes from the elongation of side chain amino acid phase, suggesting that all genes in this module might participate in the first phase of biosynthesis. IMS1 and MAML-4 genes were predicted to have the same function with IMS2 and MAM1 that catalyze the condensation of methylthio-2-oxoalkanoic with acetyl CoA to synthesize 2-methylthioalkylmalate (Textor et al. 2007). This reaction occurs in the elongation of methionine-derived side chain phase in aliphatic glucosinolate biosynthesis (Sønderby et al. 2010). Besides, IPMI1 gene was predicted to have the same functional role with IPMI2 and IIL1 that convert the 2-isopropylmalate to 3-isopropylmalate in leucine biosynthesis pathway (Binder 2010). Similar to Module 1, 22 genes from Module 2 were also predicted to be involved in the elongation of side chain amino acid phase based on their co-expression relationship with the seven known genes; i.e., IMD1, IMD2, IMD3, BCAT1, BCAT2, BCAT3, BCAT4 and AT1G800555 in this module.

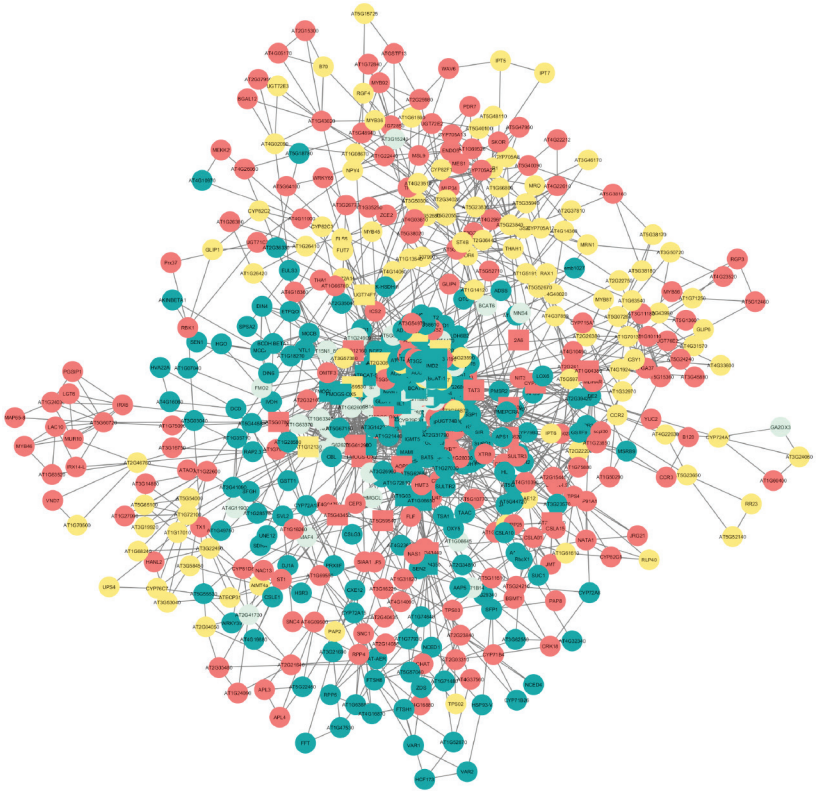

FIGURE 1. The integrated co-expression network. The genes were being categorized by the type of genes and their expression level. The square nodes represent the known genes while the round nodes represent for the potential genes. Meanwhile, different color of nodes shows different range of expression level. Yellow nodes represent genes having expression level below than 5.0, red nodes represent genes having expression level between 5.0 and 8.0, dark blue nodes represent genes having expression level higher than 8.0, and the soft blue nodes represent for the genes that did not have any expression level records from the database

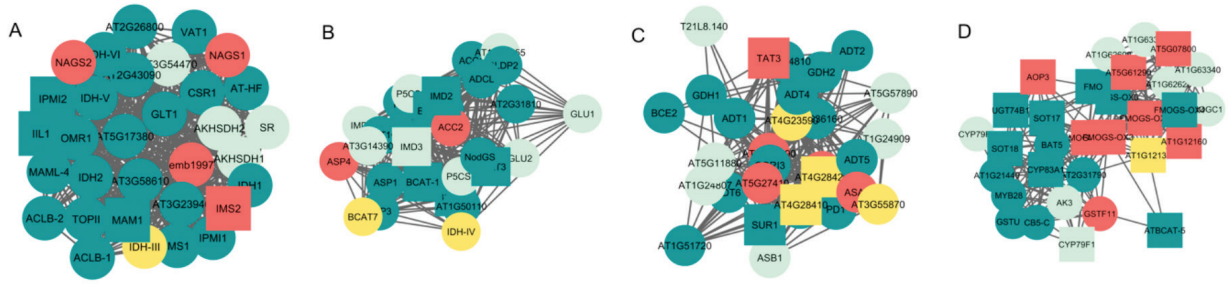

FIGURE 2. Modules containing known genes and the potential genes. (A) Module 1, (B) Module 2, (C) Module 3, (D) Module 4

Meanwhile, genes from Module 3 such as cystine lyase (CORI3) and tyrosine aminotransferase (TAT7) were shown to be potentially involved in the formation of core structure glucosinolate phase based on their interactions with the known genes such as SUR1, TAT3, AT4G28420, and AT4G28410 that cleave the conjugated S-alkylthiohydroximate into thiohydroximate, pyruvate and ammonia (Grubb & Abel 2006).

FUNCTIONAL ENRICHMENT

Functional enrichment analysis was performed in order to identify the biological processes of the genes in each module. As a result, genes from Modules 1, 2, 3 and 4 were enriched with the significant GO terms of biological process (Table 3). Genes in Module 1 were found to have 22 significant GO terms. However, we are focusing to the biological processes that contribute to the biosynthesis of

TABLE 2. List of modules with genes in the aliphatic glucosinolate biosynthesis co-expression network

Module	Score	Known genes	Co-expressed genes
1	30.0	*IIL1, IPMI2, MAM1, MAM3*	*OMR1, ACLB-2, CSR1, IMS1, MAML-4, IPMI1*
2	19.103	*BCAT-2, BCAT3, BCAT-1, IMD3, IMD2, BCAT4, IMD1*	*ASA1, ASP3, PAT1, ACC2, GLDP2, ACC1, ASP5, GLU2, GLU1, ASP1, ASP4*
3	17.034	*TAT3, SUR1, AT4G28420, AT4G28410*	*TAT7, T21L8.140, PD1, GDH3, GDH2, GDH1, CORI3, BCE2, ASB1, ASA2, ADT6, ADT5, ADT4, ADT2*
4	14.8	*CYP83A1, CYP79F1, AOP3, FMOGS-OX1, SOT17, UGT74B1, FMOGS-OX2, CYP79F2*	*AT1G63370, CB5-C, GSTU20, GSTF11, NOGC1, MYB28, AK3, BAT5, AT1G62600*

TABLE 3. GO biological process of aliphatic glucosinolate biosynthesis genes in modules

Module	GO ID	GO term	Genes
1	16143	S-glycoside metabolic process	*IIL1, IMS2, IPMI1, IPMI2, MAM1*
	44272	Sulphur compound biosynthesis process	*ACLB-1, ACLB-2, IIL1, IMS2, IPMI1, IPMI2, MAM1*
	9097	Isoleucine biosynthesis process	*CSR1, OMR1, VAT1*
	9082	Branched-chain amino acid biosynthesis process	*CSR1, IIL1, IMS1, IMS2, IPMI1, IPMI2, MAML-4, OMR1, VAT1*
	19761	Glucosinolate biosynthesis process	*IIL1, IMS2, IPMI1, IPMI2, MAM1*
2	9081	Branched-chain amino acid metabolic process	*BCAT-1, BCAT-2, BCAT3, BCAT4, IMD1*
	6103	2-oxoglutarate metabolic process	*ASP1, ASP3, ASP4, ASP5*
	6531	Aspartate metabolic process	*ASP1, ASP3, ASP4, ASP5*
4	9625	Response to insects	*CYP79F1, CYP83A1, MYB28*
	44272	Sulphur compound biosynthesis process	*BAT5, CYP79F1, CYP79F2, CYP83A1, MYB28, SOT17, SOT18, UGT74B1*
	19761	Glucosinolate biosynthesis process	*BAT5, CYP79F1, CYP79F2, CYP83A1, MYB28, SOT17, SOT18, UGT74B1*

aliphatic glucosinolate. Therefore, 11 biological processes related to the aliphatic glucosinolate biosynthesis were identified including isoleucine biosynthesis process, sulphur compound biosynthesis and branched-chain amino acid biosynthesis. While, genes from Module 2 were found significantly enriched in three related biological processes such as branched-chain amino acid metabolic process, aspartate metabolic process and 2-oxoglutarate metabolic process. This enrichment analysis also showed that genes from Module 4 are involved in several biological processes related to the aliphatic glucosinolate biosynthesis such as response to insects and S-glycoside biosynthesis. However, there was no related biological process enriched by the genes from Module 3.

Functional enrichment analysis demonstrated that genes from Module 1 were significantly enriched in different biological process associated with aliphatic glucosinolate biosynthesis including isoleucine biosynthesis, sulphur compound biosynthesis and many more. Acetolactate synthase 1 (CSR1), L-O-methylthreonine resistant (OMR1) and valine-tolerant 1 (VAT1) were collectively enriched in isoleucine biosynthesis. These processes are important in the accumulation of aliphatic glucosinolate. Besides, sulphur compound biosynthesis, S-glycoside biosynthesis, branched-chain amino acid biosynthesis and glucosinolate biosynthesis processes were also enriched by several genes from this module including 2-isopropylmalate synthase 2 (IMS2) and IPMI1. All these processes become part of aliphatic glucosinolate biosynthesis. This is because aliphatic glucosinolate is a plant secondary metabolite containing nitrogen and sulphur compounds (Redovniković et al. 2008). Besides, the core structure of glucosinolate contains β-D-thioglucoside group linked to the sulphonated aldoxime moiety and amino acid side chain (Lee et al. 2017).

Some of the genes from Module 2 such as aspartate aminotransferase (ASP), BCAT and IMD1 were annotated in three associated biological processes such branched-chain amino acid metabolic process, 2-oxoglutarate metabolic process and aspartate metabolic processes. 2-oxoglutarate metabolic process was predicted to be associated with aliphatic glucosinolate biosynthesis. According to Farrow and Facchini (2014), 2-oxoglutarate is an organic acid in TCA cycle and it is also a substrate for 2-oxoglutarate-dependent dioxygenase (2-ODD) found in *A. thaliana* glucosinolate biosynthetic pathway.

GO enrichment analysis showed that bihomomethionine N-hydroxylase (CYP79F1), cytochrome P450 83A1 (CYP83A1) and MYB28 genes from Module 4 are involved in response to insects' attack by catalyzing the hydrolyzation of glucosinolate and producing several toxic compounds such as including thiocyanate, isothiocyanate, nitrile and epithionitrile to deter the insects (Hopkins et al. 2009). Moreover, a total of eight genes from this module such as bile acid transporter (BAT5) and MYB28 genes were associated in the glucosinolate biosynthetic and sulphur containing compound biosynthetic pathways.

REGULATORY NETWORK

The transcriptional regulatory network of aliphatic glucosinolate biosynthesis was built by mapping the regulatory network from AtRegNet onto the co-expression network. The constructed regulatory network consisted of 185 nodes that represent genes and transcription factors connected by 228 regulatory pairs. The analysis, however, was only focused on the transcription factors that regulates the aliphatic glucosinolate genes. Thus, genes that share the same transcription factor with the aliphatic glucosinolate genes were predicted to have a tendency to be involved in the same biological pathway. As a result, five regulatory genes including sepallata 3 (SEP3), phytochrome interacting factor 3-like 5 (PIL5), long hypocotyl 5 (HY5), agami-like 15 (AGL15) and glabra 3 (GL3) were known to contribute in the regulation of aliphatic glucosinolate genes (Figure 3).

Transcriptional regulation is determined through the identification of *cis*-element by transcription factor binding sites domain (Suryamohan & Halfon 2015). The binding of the transcription factor at this *cis*-element region of the promoter site and their interactions in the regulatory network influence the expression of the target genes (Smita et al. 2015). Therefore, the transcriptional regulatory network was constructed to explore the involvement of transcription factors in the aliphatic glucosinolate biosynthesis pathway.

Five transcriptional factors (Table 4) were identified to play a crucial role in a plant defense system including abiotic and/or biotic stress response. Based on Kaufmann et al. (2009), SEP3 is a MADS-domain protein expressed during flower development by modulating auxin response to facilitate morphogenesis and the growth of the floral organ. While PIL5 is a negative regulator that binds to the phytochrome (Moon et al. 2008). Phytochrome is a photoreceptor that stimulates the seed proliferation in plants and the inhibition of this phytochrome by PIL5 drops the level of gibberellic acid (Oh et al. 2009). Low level of gibberellic acid activates DELLA protein that stimulates the biosynthesis of abscisic acid (Kim et al. 2008). The abscisic acid produced are found to participate in response to pathogen attacks (Cao et al. 2011) and response to a variety of environmental stresses such as drought and salinity (Seo & Koshiba 2002).

Another transcription factor found was HY5 and it regulates the photomorphogenesis positively (Cluis et al. 2004) by activating the genes that stimulate photosynthesis, photopigment biosynthesis as well as chloroplast and seed cotyledon development (Toledo-Ortiz et al. 2014). It also stimulates the accumulation of anthocyanin and reactive oxygen species (ROS) (Zhang et al. 2011). Based on the reconstructed regulatory network, HY5 was found to activate methylthioalkylmalate synthase 1 (MAM1) gene and inhibit late embryogenesis abundant-containing gene (AT1G72100). AT1G72100 was potentially involved in the aliphatic glucosinolate genes due to its ability to tolerate abiotic stresses such as cold, drought and salinity. This gene was abundantly found during the dehydration of plant seeds in the late embryogenesis phase (Hundertmark & Hincha

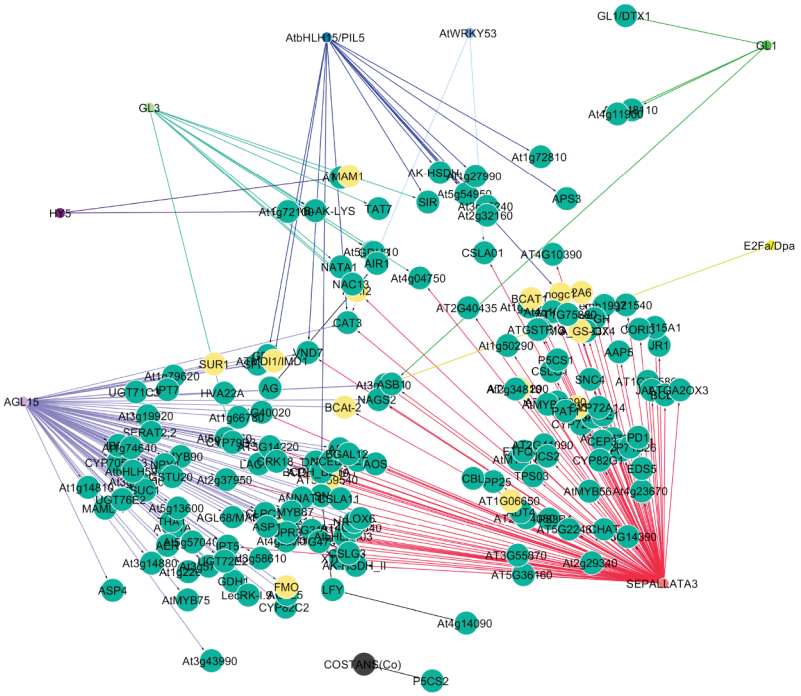

FIGURE 3. Aliphatic glucosinolate biosynthesis regulatory network. Yellow nodes represent the known genes and green nodes represent the potential genes involved in the aliphatic glucosinolate biosynthesis. These genes were connected by different colors of nodes and edges that represent as transcription factors

TABLE 4. List of transcription factors with regulated genes

Transcription factor	Target genes	
	Known genes	Potential genes
SEP 3	*2A6, BCAT1, BCAT-2, FMO_GS-OX4, IMS2, IPMI2, NOGC1*	*AAP5, ADT4, AK-HSDH_II, ANNAT1, AOS, APK, APL3, ASB1, ASP1, AtMYB28, AtMYB37, AtMYB56, AtMYB76, AtMYB87*
PIL5	*ATMDI1/IMD1, NOGC1*	*AK-HSDH_I, APS3, ASN1, CAT3, GDH2, SIR, VND7*
HY5	*MAM1*	*AT1G72100*
AGL15	*IMD1, BCAT-2, FMO, IMS2, SUR1*	*AER, AOS, AOX1A, APK, ASB1, ASN1, ASP1, ASP4*
GL3	*IPMI2*	*AT-HF, CYP79B3, GDH3, NAC13, SIR, TAT7*

2008). Similar to PIL5, Zheng et al. (2013) stated that AGL15 transcription factor also inhibits the expression of genes encoding auxin receptor in somatic embryogenesis, leading to the biosynthesis of abscisic acid which is essential in the biotic and abiotic response of the plants.

On the other hand, previous studies showed that GL3 participate in the regulation of trichome development (Wada et al. 2014). Trichome is a hairy epidermis cell that appeared on the leaves of most plants and high level of trichome can protect the plants from herbivore attacks (Yoshida et al. 2009). This is due to the high level of aliphatic glucosinolate in the trichomes of *A. thaliana*'s leaves that activate the glucosinolate-myrosinase system, producing toxic compounds to herbivores (Frerigmann et al. 2012).

Among 506 identified genes, methylthioalkylmalate-like synthase 4 (MAML-4) has the highest potential to be involved in the aliphatic glucosinolate biosynthetic pathway, followed by aspartate aminotransferase (ASP1 and ASP4) genes. In this study, we found that MAML-4 was the highest potential gene involved in the aliphatic glucosinolate biosynthetic pathway. MAML-4 was co-expressed with IPMI2 and IIL1 in Module 1 and annotated with several biological processes related to the accumulation of aliphatic glucosinolate such as carboxylic acid biosynthetic process, leucine metabolic process and branched-chain amino acid biosynthetic process. In addition, it was also regulated by the same transcription factor with IMD1 and IMS2 genes, proving that it might carry similar function with these genes. This

finding was supported by the previous study done by Field et al. (2006). Their findings stated that MAML-4 encodes an enzyme that catalyzes the condensation of acetyl-CoA of elongated 2-oxo acid in the elongation of amino acid side chain leucine- or methionine-derived glucosinolate phase.

Meanwhile, ASP1 and ASP4 were the second highest predicted genes participated in the biosynthesis pathway. Based on the clustering analysis, ASP1 and ASP4 genes showed high tendency to have a similar role with branched-chain amino acid transaminase genes (BCAT). Furthermore, these two genes were also annotated with the aliphatic glucosinolate biosynthesis-related biological processes such as aspartate metabolic process and share similar transcription factor with BCAT1 and BCAT2 in the regulatory analysis. The involvement of ASP1 and ASP4 in the aliphatic glucosinolate biosynthesis was in agreement with the findings Miesak and Coruzzi (2002), demonstrating that ASP1 encodes for an aspartate amino acid enzyme in mitochondria, while, ASP4 encodes for aspartate aminotransferase in cytosol.

CONCLUSION

The identification of potential genes will enhance understanding of aliphatic glucosinolate biosynthesis mechanism in plant defense system. The information discovered in this study can be used to complete and update the aliphatic glucosinolate biosynthetic pathway in the KEGG and Aracyc databases, providing a clearer understanding of the biosynthesis of glucosinolate specifically in *A. thaliana*.

ACKNOWLEDGEMENTS

This research was carried out at the Centre for Bioinformatics Research, Institute of Systems Biology (INBIOSIS), and funded by ERGS/1/2013/STG07/UKM/02/3 awarded to ZAMH by the Ministry of Higher Education, (Malaysia).

REFERENCES

Atwell, L.L., Beaver, L.M., Shannon, J., Williams, D.E., Dashwood, R.H. & Ho, E. 2015. Epigenetic regulation by sulforaphane: Opportunities for breast and prostate cancer chemoprevention. *Current Pharmacology Reports* 1(2): 102-111.

Bader, G.D. & Hogue, C.W. 2003. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4(1): 2.

Beekwilder, J., Van Leeuwen, W., Van Dam, N.M., Bertossi, M., Grandi, V., Mizzi, L., Soloviev, M., Szabados, L., Molthoff, J.W., Schipper, B. & Verbocht, H. 2008. The impact of the absence of aliphatic glucosinolates on insect herbivory in *Arabidopsis*. *PLoS ONE* 3(4): 1-12.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W., Pagès, F., Trajanoski, Z., Galon, J., Team, A., Immunology, I.C. & Descartes, U.P. 2009. ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25(8): 1091-1093.

Binder, S. 2010. Branched-chain amino acid metabolism in *Arabidopsis thaliana*. *The Arabidopsis Book* e0137: 1-14.

Bradbury, L.M.T., Niehaus, T.D. & Hanson, A.D. 2013. Comparative genomics approaches to understanding and manipulating plant metabolism. *Current Opinion in Biotechnology* 24: 278-284.

Cao, F.Y., Yoshioka, K. & Desveaux, D. 2011. The roles of ABA in plant-pathogen interactions. *Journal of Plant Research* 124(4): 489-499.

Cluis, C.P., Mouchel, C.F. & Hardtke, C.S. 2004. The *Arabidopsis* transcription factor HY5 integrates light and hormone signaling pathways. *The Plant Journal* 38(2): 332-347.

Farrow, S.C. & Facchini, P.J. 2014. Functional diversity of 2-oxoglutarate/Fe(II)-dependent dioxygenases in plant metabolism. *Frontiers in Plant Science* 5(October): 1-15.

Field, B., Furniss, C., Wilkinson, A. & Mithen, R. 2006. Expression of a Brassica isopropylmalate synthase gene in *Arabidopsis* perturbs both glucosinolate and amino acid metabolism. *Plant Molecular Biology* 60(5): 717-727.

Frerigmann, H., Bottcher, C., Baatout, D. & Gigolashvili, T. 2012. Glucosinolates are produced in trichomes of *Arabidopsis thaliana*. *Frontiers in Plant Science* 3: 242.

Grubb, C.D. & Abel, S. 2006. Glucosinolate metabolism and its control. *Trends in Molecular Medicine* 11(2): 89-100.

Halkier, B.A. & Gershenzon, J. 2006. Biology and biochemistry of glucosinolates. *Annual Review of Plant Biology* 57: 303-333.

Hopkins, R.J., van Dam, N.M. & van Loon, J.J.A. 2009. Role of glucosinolates in insect-plant relationships and multitrophic interactions. *Annual Review of Entomology* 54: 57-83.

Hundertmark, M. & Hincha, D.K. 2008. LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* 9: 118.

Ishida, M., Hara, M., Fukino, N., Kakizaki, T. & Morimitsu, Y. 2014. Glucosinolate metabolism, functionality and breeding for the improvement of Brassicaceae vegetables. *Breeding Science* 64: 48-59.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research* 44: D457-D462.

Kaufmann, K., Muiño, J.M., Jauregui, R., Airoldi, C.A., Smaczniak, C., Krajewski, P. & Angenent, G.C. 2009. Target genes of the MADS transcription factor sepallata3: Integration of developmental and hormonal pathways in the *Arabidopsis* flower. *PLoS Biology* 7(4): 0854-0875.

Kim, D.H., Yamaguchi, S., Lim, S., Oh, E., Park, J., Hanada, A., Kamiya, Y. & Choi, G. 2008. SOMNUS, a CCCH-Type zinc finger protein in *Arabidopsis*, negatively regulates light-dependent seed germination downstream of PIL5. *The Plant Cell* 20(5): 1260-1277.

Lächler, K., Imhof, J., Reichelt, M., Gershenzon, J. & Binder, S. 2015. The cytosolic branched-chain aminotransferases of *Arabidopsis thaliana* influence methionine supply, salvage and glucosinolate metabolism. *Plant Molecular Biology* 88(1-2): 119-131.

Lee, J.G., Lim, S., Kim, J. & Lee, E.J. 2017. The mechanism of deterioration of the glucosinolate-myrosynase system in radish roots during cold storage after harvest. *Food Chemistry* 233: 60-68.

Li, Y., Pearl, S.A. & Jackson, S.A. 2015. Gene networks in plant biology: Approaches in reconstruction and analysis. *Trends in Plant Science* 20(10): 664-675.

Mao, L., Van Hemert, J.L., Dash, S. & Dickerson, J.A. 2009. *Arabidopsis* gene co-expression network and its functional modules. *BMC Bioinformatics* 10: 1-24.

Martínez-Ballesta, M., Moreno-Fernández, D.A., Castejón, D., Ochando, C., Morandini, P.A. & Carvajal, M. 2015. The impact of the absence of aliphatic glucosinolates on water transport under salt stress in *Arabidopsis thaliana*. *Frontiers in Plant Science* 6(July): 1-12.

Miesak, B.H. & Coruzzi, G.M. 2002. Molecular and physiological analysis of *Arabidopsis* mutants defective in cytosolic or chloroplastic aspartate aminotransferase. *Plant Physiology* 129(2): 650-660.

Moon, J., Zhu, L., Shen, H. & Huq, E. 2008. PIF1 directly and indirectly regulates chlorophyll biosynthesis to optimize the greening process in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* 105(27): 9433-9438.

Mostafa, I., Yoo, M.J., Zhu, N., Geng, S., Dufresne, C., Abou-Hashem, M., El-Domiaty, M. & Chen, S. 2017. Membrane proteomics of *Arabidopsis* glucosinolate mutants cyp79b2/b3 and myb28/29. *Frontiers in Plant Science* 8(April).

Mueller, L.A., Zhang, P. & Rhee, S.Y. 2003. AraCyc: A biochemical pathway database for *Arabidopsis*. *Plant Physiology* 132: 453-460.

Neal, C.S., Fredericks, D.P., Griffiths, C.A. & Neale, A.D. 2010. The characterisation of AOP2: A gene associated with the biosynthesis of aliphatic alkenyl glucosinolates in *Arabidopsis thaliana*. *BMC Plant Biology* 10: 1-16.

Nour-Eldin, H.H. & Halkier, B.A. 2009. Piecing together the transport pathway of aliphatic glucosinolates. *Phytochemistry Reviews* 8(1): 53-67.

Nützmann, H.W., Huang, A. & Osbourn, A. 2016. Plant metabolic gene clusters - from genetics to genomics. *New Phytologist* 211(3): 771-789.

Obayashi, T., Kinoshita, K., Nakai, K., Shibaoka, M., Hayashi, S., Saeki, M., Shibata, D., Saito, K. & Ohta, H. 2007. ATTED-II: A database of co-expressed genes and cis elements for identifying co-regulated gene groups in *Arabidopsis*. *Nucleic Acids Research* 35: D863-D869.

Oh, E., Kang, H., Yamaguchi, S., Park, J., Lee, D., Kamiya, Y. & Choi, G. 2009. Genome-wide analysis of genes targeted by PHYTOCHROME INTERACTING FACTOR 3-LIKE5 during seed germination in *Arabidopsis*. *The Plant Cell* 21(2): 403-419.

Rameeh, V. 2015. Glucosinolates and their important biological and anti cancer effects: A review. *Jordan Journal of Agricultural Sciences* 11(1): 1-13.

Redovniković, I.R., Glivetic, T., Delonga, K. & Jasna Vorkapic-Furac. 2008. Glucosinolates and their potential role in plant. *Periodicum Biologorum* 110(4): 297-309.

Redovniković, I.R., Textor, S., Lisni, B. & Gershenzon, J. 2012. Expression pattern of the glucosinolate side chain biosynthetic genes MAM1 and MAM3 of *Arabidopsis thaliana* in different organs and developmental stages. *Plant Physiology and Biochemistry* 53: 77-83.

Rohr, F., Ulrichs, C. & Mewis, I. 2009. Variability of aliphatic glucosinolates in *Arabidopsis thaliana* (L.) - Impact on glucosinolate profile and insect resistance. *Journal of Applied Botany and Food Quality* 82(2): 131-135.

Roy, S., Bhattacharyya, D.K. & Kalita, J.K. 2016. Analysis of gene expression patterns using biclustering. *Methods in Molecular Biology (Clifton, N.J.)* 1375: 91-103.

Schuster, J., Knill, T., Reichelt, M., Gershenzon, J. & Binder, S. 2006. BRANCHED-CHAIN AMINOTRANSFERASE4 is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in *Arabidopsis*. *The Plant Cell* 18: 2664-2679.

Seo, M. & Koshiba, T. 2002. Complex regulation of ABA biosynthesis in plants. *Trends in Plant Science* 7(1): 41-48.

Serin, E.A.R., Nijveen, H., Hilhorst, H.W.M. & Ligterink, W. 2016. Learning from co-expression networks: Possibilities and challenges. *Frontiers in Plant Science* 7(444): 1-18.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. & Ideker, T. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* 13(11): 2498-2504.

Smita, S., Katiyar, A., Chinnusamy, V., Pandey, D.M. & Bansal, K.C. 2015. Transcriptional regulatory network analysis of myb transcription factor family genes in rice. *Frontiers in Plant Science* 6(December): 1-19.

Sønderby, I.E., Geu-flores, F. & Halkier, B.A. 2010. Biosynthesis of glucosinolates - gene discovery and beyond. *Trends in Plant Science* 15(5): 283-290.

Suryamohan, K. & Halfon, M. 2015. Identifying transcriptional cis-regulatory modules in animal genomes. *Wiley Interdisciplinary Reviews. Developmental Biology* 4(2): 59-84.

Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., Jensen, L.J. & von Mering, C. 2017. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Research* 45: D362-D368.

Textor, S., de Kraker, J.W., Hause, B., Gershenzon, J. & Tokuhisa, J.G. 2007. MAM3 catalyzes the formation of all aliphatic glucosinolate chain lengths in *Arabidopsis*. *Plant Physiology* 144(1): 60-71.

Toledo-Ortiz, G., Johansson, H., Lee, K.P., Bou-Torrent, J., Stewart, K., Steel, G., Rodríguez-Concepción, M. & Halliday, K.J. 2014. The HY5-PIF regulatory module coordinates light and temperature control of photosynthetic gene transcription. *PLoS Genetics* 10(6): e1004416.

van Dam, S., Võsa, U., van der Graaf, A., Franke, L. & de Magalhães, J.P. 2017. Gene co-expression analysis for functional classification and gene-disease predictions. *Briefings in Bioinformatics* (January): 1-18.

Wada, T., Kunihiro, A. & Tominaga-Wada, R. 2014. Arabidopsis CAPRICE (MYB) and GLABRA3 (bHLH) control tomato (*Solanum lycopersicum*) anthocyanin biosynthesis. *PLoS ONE* 9(9): e109093.

Warde-Farley, D., Donaldson, S.L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C.T., Maitland, A., Mostafavi, S., Montojo, J., Shao, Q., Wright, G., Bader, G.D. & Morris, Q. 2010. The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research* 38: W214-W220.

Yilmaz, A., Mejia-Guerra, M.K., Kurz, K., Liang, X., Welch, L. & Grotewold, E. 2011. AGRIS: The *Arabidopsis* gene regulatory information server, an update. *Nucleic Acids Research* 39(SUPPL. 1): 1118-1122.

Yoshida, Y., Sano, R., Wada, T., Takabayashi, J. & Okada, K. 2009. Jasmonic acid control of GLABRA3 links inducible defense and trichome patterning in *Arabidopsis*. *Development* 136(6): 1039-1048.

Zhang, Y., Liu, Z., Liu, R., Hao, H. & Bi, Y. 2011. Gibberellins negatively regulate low temperature-induced anthocyanin accumulation in a HY5/HYH-dependent manner. *Plant Signaling and Behavior* 6(5): 632-634.

Zheng, Q., Zheng, Y. & Perry, S.E. 2013. AGAMOUS-Like15 promotes somatic embryogenesis in *Arabidopsis* and soybean in part by the control of ethylene biosynthesis and response. *Plant Physiology* 161(4): 2113-2127.

Khalidah-Syahirah Ashari & Zeti-Azura Mohamed-Hussein*
Centre for Frontier Sciences
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor Darul Ehsan
Malaysia

Muhammad-Redha Abdullah-Zawawi, Sarahani Harun & Zeti-Azura Mohamed-Hussein*
Centre for Bioinformatics Research
Institute of Systems Biology (INBIOSIS)
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor Darul Ehsan
Malaysia

*Corresponding author; email: zeti.hussein@ukm.edu.my

Received: 30 May 2018
Accepted: 19 September 2018