Journal of Cell and Molecular Research (2018) 10 (1), 18-26 DOI: 10.22067/jcmr.v10i1.67357

Research Article

Global Analysis of Gene Expression and Identification of Modules in *Echinacea purpurea* Using Systems Biology Approach

Ahmad Tahmasebi[1], Farzaneh Aram[2], Hassan Pakniyat[1], Ali Niazi[2], Elahe Tavakol[1], Esmaeil Ebrahimie[2,3,4,5*]

[1] *Department of Crop Production and Plant Breeding, College of Agriculture, Shiraz University, Shiraz, Iran*
[2] *Institute of Biotechnology, Shiraz University, Shiraz, Iran*
[3] *School of Medicine, The University of Adelaide, Adelaide, Australia*
[4] *School of Information Technology and Mathematical Sciences, Division of Information Technology, Engineering and the Environment, University of South Australia, Adelaide, Australia*
[5] *School of Biological Sciences, Faculty of Science and Engineering, Flinders University, Adelaide, Australia*

Received 10 September 2017 Accepted 25 November 2017

Abstract

Co-expression analysis is a useful tool to analysis data and detection of genes that act in the same pathway or biological process. *Echinacea purpurea* is one of the most important medicinal plant of the Asteraceae family that is known as antioxidative and antiviral agent. Despite medicinal importance of *E. purpurea*, very few reports are available for metabolic mechanisms in this plant. With the aim to elucidate the gene expression profiling and identification of modules in *E. purpurea*, we performed a systems biology analysis on publicly available transcriptome data. Gene ontology and KEGG pathway enrichment analysis revealed that the unigenes were highly related to the cellular process, primary metabolic process, carbon metabolism and biosynthesis of antibiotics. The co-expression networks divided genes into multiple modules. Of these, module M2 associated with secondary metabolic process. Moreover, a total of 47 transcription factor families such as bHLH, bZIP, C2H2, MYB and WRKY in modules were identified. These findings can provide an overall picture for better understanding the gene expression patterns and common transcriptional mechanisms in *E. purpurea*.

Keywords**:** *Co-expression analysis, Echinacea purpurea, Transcriptomics data*

Introduction

The emergence of sequencing technology and bioinformatics tools provide an opportunity for study important aspects of metabolic processes and complexity of transcriptome in non-model plant species. Gene co-expression analyses can uncover the functional interaction between genes. A co-expression network represents pairwise interactions among genes which based on these relationships, it is possible to find gene clusters (modules) that involved in common biological pathways (Guo et al., 2014; Lee et al., 2010). Currently, availability of transcriptomics data for various medicinal plants also can be used as tools to explore transcriptome information and molecular mechanisms, specifically related to the secondary metabolites. The biosynthesis of secondary metabolites is a complex process and regulate by transcriptional, translational and post-translational mechanisms. The accumulation and synthesis of these metabolites affect by different factors. Transcriptome analysis can facilitate identification of key genes and regulators associated with such bioactive compounds (Graham et al., 2010; Xiao et al., 2013). Investigation of the network controlling the biosynthesis, transportation, accumulation of secondary metabolites will be helpful for production of these metabolites. Engineering biosynthetic pathways can be considered as an approach but is still limited, partially due to complexity of biosynthesis pathways (Yang et al., 2012). Identification and annotation of transcription factors as regulatory proteins of gene expression can assist greatly in this regard.

Echinacea purpurea is an important medicinal plant which belongs to Asteraceae family. The roots, flowers and aerial parts are used mainly for medicinal purposes. In recent years, *E. purpurea* has been considered for high medicinal value and industrial applications in pharmaceutical and food industries. It has antioxidant, anti-inflammatory and immuno-stimulating properties and is used for treating viral and inflammatory diseases. Nowadays, numerous photochemical constituents include polysaccharides, caffeic acid derivatives and

* Corresponding author E-mail: esmaeil.ebrahimie@adelaide.edu.au

alkylamides; have been identified from *E. purpurea* (Pellati et al., 2004; Tsai et al., 2012). Despite the importance of this plant, mechanism and regulation of biosynthesis of metabolites have been rarely investigated. In this study, by using transcriptome data and gene co-expression network analysis, we provide global expression patterns of genes and detected pathways, modules and transcription factors to achieve insights into particularly molecular mechanisms of secondary metabolism-related genes in *E. purpurea*.

Materials and Methods

Data collection and functional annotation analysis

The raw expression data of different tissues of *E. purpurea* were obtained from the Medicinal Plant Genomics Resource database (MPGR, http://medicinalplantgenomics.msu.edu/) (Góngora-Castillo et al., 2012). All expressed transcripts were searched against the Arabidopsis Information Resource (TAIR) (http://www. Arabidopsis.org) using the BLASTX.

GO enrichment analysis of the transcripts was performed using Gene Ontology (GO, http://wego.genomics.org.cn/cgi-bin/wego/index.pl) database for categories of biological process (BP), molecular function (MF) and cellular components (CC). DAVID database (Dennis et al., 2003) was carried out to find important pathways according to the KEGG pathway database. Significantly enriched pathways were determined using a Benjamini test (*P* value <0.05).

Co-expressed gene network analysis

A Co-expression analysis was conducted with the Weighted Gene Co-expression Network Analysis (WGCNA) (Langfelder and Horvath, 2008). First, transcripts with low expression were filtered out, then expression matrix was used to cluster groups of highly co-expressed genes. The pairwise gene correlation matrix was transformed into a weighted matrix with a soft thresholding power of 10 to calculate a topological overlap matrix (TOM) and creating a hierarchical cluster tree. Finally, the modules were identified with minimum module size of 100, the power of 8 and TOMType of signed.

Enrichment analysis of modules

To interpret of functional characterization of modules that were identified by the WGCNA analysis, a separate enrichment analysis was performed to transcript list of modules using agriGO (Du et al., 2010). Moreover, to identify and classify transcription factors (TFs) corresponding to each module, transcripts were aligned to AtTFDB database (http://arabidopsis.med.ohio-state.edu/AtTFDB/) using BLASTX with a cut-off E-value ≤ 10−5.

Results and Discussion

Gene ontology (GO) and KEGG pathway analysis

To obtain insight into functions and characterize the *E. purpurea* transcriptome, all unigenes from different tissues were annotated using BLASTX searches against the TAIR database. A total of 73,643 unigenes (69.74%) were matched to TAIR database. GO analysis was used to predict the function of unigenes by classifying them into three major functional categories (biological process, cellular component and molecular function) (Figure 1). In the category of molecular function, catalytic activity (GO:0003824) with 4,168, binding (GO:0005488) with 2,421 and transporter activity (GO:0005215) with 1,071 transcripts were predominant. In the biological process group, unique sequences related to cellular process (GO:0009987) with 4056, primary metabolic process (GO:0044238) with 3872, localization (GO:0051179) with 1408, and cellular component organization (GO:0071840) with 999 transcripts were found. Moreover, secondary metabolic process (GO:0019748) was one of most common categories. In the cellular component domain, cell part (GO:0044464), organelle (GO:0043226), and macromolecular complex (GO:0032991) were shown to be the top 3 clusters.

KEGG pathway enrichment analysis of unigenes was also carried out in web-based DAVID. Of the 15,085 unigenes, 2,129 genes were categorized into 41 pathways which 7 pathways include carbon metabolism, biosynthesis of antibiotics, glycolysis/gluconeogenesis, peroxisome, glycerolipid metabolism, endocytosis and fatty acid degradation, were significantly enriched (Table 1).

Identification of the genes associated with phenylpropanoid biosynthesis

In plants, phenylpropanoid pathway is a source of metabolites and starting point for the generation of secondary metabolites such as phenolic volatiles, flavonoids and lignans (Fraser and Chapple, 2011; Vogt, 2010). The phenylpropanoids and their derivatives play also key roles in plant responses to biotic and abiotic stresses. In this study, a total of 491 unigenes were obtained that related to

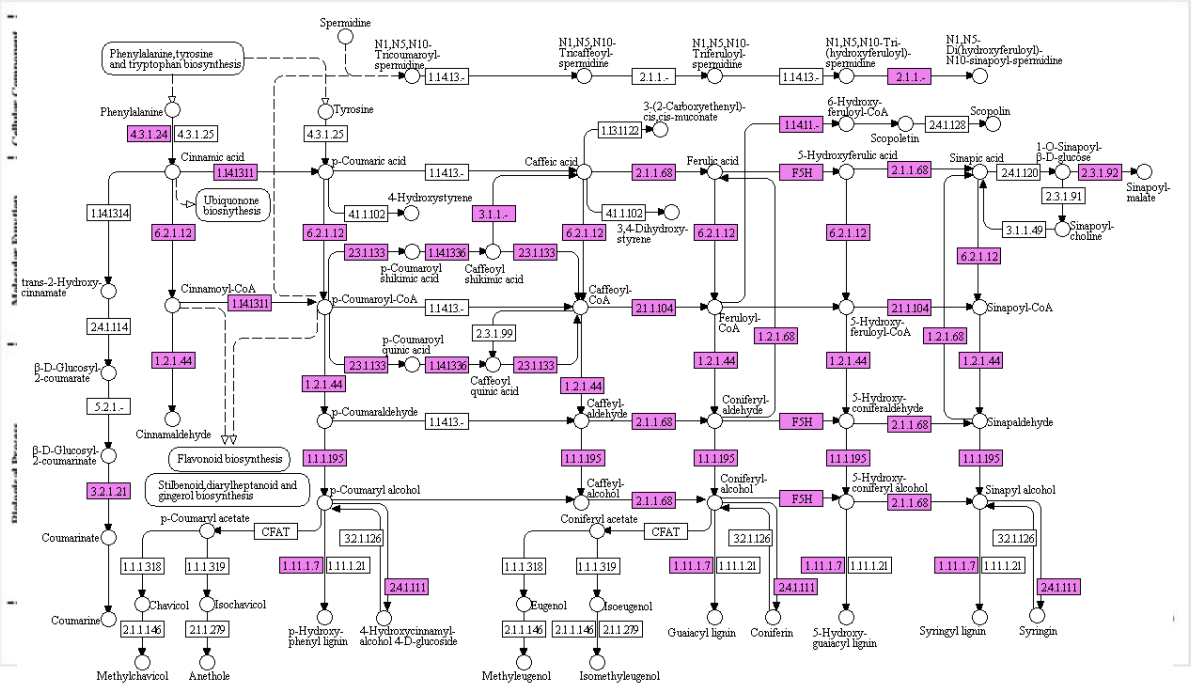

Figure 1. Pathway classification of the unigenes into five major categories. Colored boxes indicate the molecular function identified genes.

Table 1. The KEGG pathway enrichment analysis of the total unigenes.

Term	Gene count*	Percentage	p-value
Carbon metabolism	217	1.5	0.000
Biosynthesis of antibiotics	353	2.4	0.000
Glycolysis / Gluconeogenesis	97	0.6	0.009
Peroxisome	76	0.5	0.013
Glycerolipid metabolism	48	0.3	0.013
Endocytosis	117	0.8	0.026
Fatty acid degradation	38	0.3	0.040

* The number and percentage of genes in each pathway.

phenylpropanoid biosynthesis pathway (Figure 2). Several key enzymes such as phenylalanine ammonia-lyase [EC:4.3.1.24], peroxidase [EC:1.11.1.7] and caffeoyl-CoA O-methyltransferase [EC:2.1.1.104] that involved in phenylpropanoid biosynthesis were expressed. Phenolic compounds are derived from phenylalanine through the phenylpropanoid pathway. The pathway is initiated by ammonia-lyase (PAL) and cinnamate-4-hydroxylase (C4H), which their activities are positively correlated to phenylpropanoid product accumulation (Ali and McNear, 2014; Lee and Scagel, 2010). The enhancement of PAL and C4H activities was accompanied with an increase in the cichoric acid content in *E.purpurea.* In addition, previous studies have reported that enzyme activity of PAL changes by application of GA3 that has resulted in accumulation of flavonoids, lignin and starch (Abbasi et al., 2012).

Gene co-expression network analysis

To detect networks of co-expressed genes in *E. purpurea*, systems biology approach was performed by weighted gene correlation network analysis (WGCNA). WGCNA clusters genes based on gene expression similarity into modules. After low expression filtering, a FPKM (fragments per kilobase of transcript per million fragments mapped) matrix of different tissues include 21,363

transcripts were obtained for construct gene co-

expression networks. The genes were clustered based on similar patterns. A dendrogram and correlation heat map was generated to visualize topological overlap values between genes in the modules (Figure 3). In total, 25 modules were identified which ranged in size from 1,783 members to 245 members and tagged with different colors. Turquoise (containing 1,783 genes), blue (1,495 genes), brown (1,472 genes), yellow (1,326 genes) and green (1,310 genes) were five major modules. Based on the eigengenes and a minimum cut height (0.5), modules grouped in seven distinct clades (Figure 4). Subsequently, the MEtur and MEsalmon modules had very similar expression profiles, thus these modules were merged and renamed as M1. Similarly, MEdarkturquoise and MEpink were merged and renamed as M2, whereas MEcyan, MEgreen, MEbrown and MEmidnightblue were merged and renamed as M3. MEdarkgreen, MEred, MElightgreen and MEyellow were renamed as M4. MEmagenta, MEturquoise, MEroyalblue and MEtan were renamed as M5. MEgrey60, MEorange, MEdarkred, MElightyellow, MEblack and MEgreenyellow were merged and renamed as M6. MElightcyan, MEdarkgrey and MEpurple were renamed as M7.

Functional annotation of modules

To find the biological functions associated with modules, the modules were submitted to enrichment analysis. A FDR<0.01 was used to define significantly enriched terms. All of the modules were significantly enriched for at least one GO term. The biological process GO terms of each module are presented in Additional file. Module M1 was associated with aspects of the cellular process according to gene ontology, and the module was highly enriched for cellular component organization (GO:0016043), reproductive process (GO:0022414), reproduction (GO:0000003), metabolic process (GO:0008152) and response to stimulus (GO:0050896) terms that are related to the immune system (Table S1). Also, 583 genes were identified from this module that were involved in primary metabolic process. Module M2 enriched for functions related to metabolic process. The module also was associated with several GO terms that involved in developmental process (GO:0032502) and multicellular organismal process (GO:0032501). Interestingly, 37 and 20 genes were associated with secondary metabolic process (GO:0019748) and pigment metabolic process (GO:0042440), respectively (Table S2). Module M3 linked to localization (GO:0051179), immune system process (GO:0002376), response to stimulus (GO:0050896) and growth (GO:0040007).

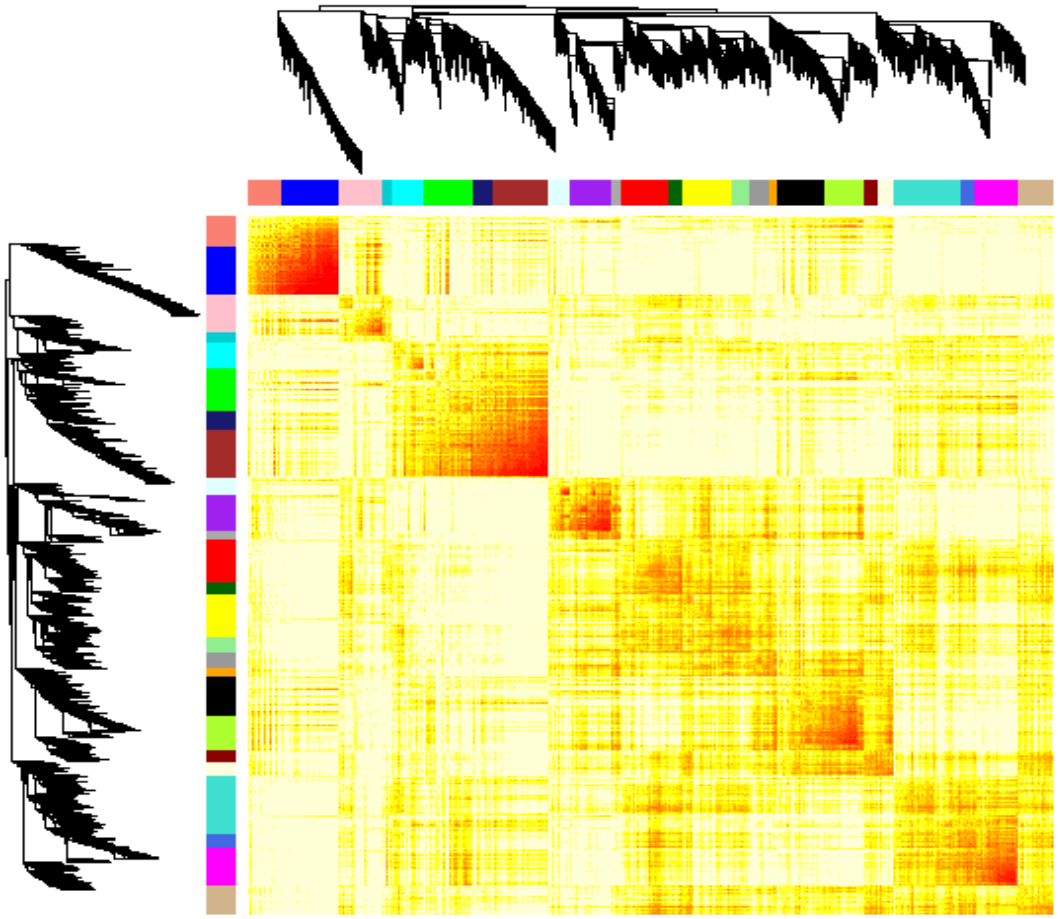

Figure 3. Topological overlap matrix plot (TOMplot) for all WGCNA modules. Modules are illustrated by different colors. Red color in the heat map represents the genes that showed high correlation.

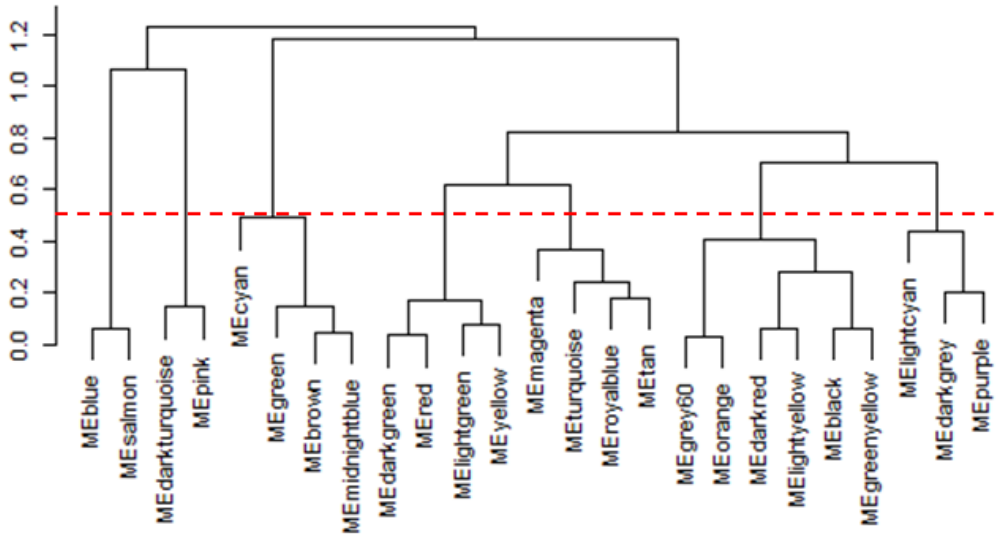

Figure 4. Clustering dendrogram of modules. The horizontal line showing merge cut height of 0.5.

Under biological regulation, 205 genes also were associated with regulation of primary metabolic process (GO:0080090). For the response to stimulus, 51 genes were associated with the response to auxin stimulus (GO:0009733) (Table S3). Module M4 was related to reproductive process (GO:0022414), primary metabolic process (GO:0044238) and macromolecule metabolic process (GO:0043170). Moreover, genes that are involved in organ development (GO:0048513) were in this module (Table S4). Likewise, in module M5, M6 and M7, enriched for functions related to chromatin organization multicellular organismal process (GO:0032501), developmental process (GO:0032502) and reproduction (GO:0000003) (Table S5, S6 and S7). The results showed that these modules involved in various biological processes.

In functional annotation observed that some genes of module M2 involved in secondary metabolic process. To evaluate details of metabolites related to module M2, metabolome data of *E. purpurea* available at PMR database (PMR; http://metnetdb.org/PMR/) (Hur et al., 2013) was used. The genes present in module M2 were searched against correlation data of metabolomes to determine metabolites synthesized by module M2. The results indicated that module M2 was significantly correlated with biosynthesis of monoterpenoids, fatty acid derivatives, isoflavonoids and anthocyanidins.

Identification of transcription factors association with modules

Transcription factors (TFs) are key set of proteins that regulate gene expression and function as central regulators in many important biological processes such as growth, development and secondary metabolism (Mitsuda and Ohme-Takagi, 2009; Patra et al., 2013). To assess the regulatory mechanisms of detected modules, unigene sequences of first five WGCNA modules were searched against AtTFDB database. Based on BLASTX, TFs were assigned to different modules and classified into different families. Generally, 47 TF families such as AP2/ERF-ERF, bHLH, bZIP, C2H2, MYB and WRKY were identified in all five modules (Table 2). The turquoise module was enriched in bHLH, HB-HD-ZIP, bZIP, TUB and WRKY TF families. The blue module also included C2H2, AP2/ERF-ERF and bHLH TF families which associated with gibberellic acid mediated signaling pathway. The brown module displayed the largest number of GRAS, AP2/ERF-ERF and WRKY TF families. The yellow and green modules were particularly enriched for AP2/ERF-ERF, WRKY and bHLH.

The different TF families were also reported to regulate anthocyanin and flavonoid biosynthesis. The anthocyanin biosynthesis starts from amino acid phenylalanine. Most of the structural genes in the pathway are coordinately regulated by the MBW complex, which is composed of the MYB, bHLH and WDR proteins. TFs belonging to these three groups are functionally conserved among plant species. In Arabidopsis, early genes of anthocyanin pathway are positively regulated by MYB TF family, MYB11, MYB12 and MYB111, however late pathway genes are regulated by MBW complex. However, the MYB (phMYB27) negatively regulates anthocyanin biosynthesis in Petunia. The phMYB27 is up-regulated in shadow-grown leaves and is repressed under high light. Light-induced expression pattern of TFs as activator or repressor in Arabidopsis and Petunia is similar that suggest the conserved controlling mechanism of anthocyanin synthesis in vegetative tissues (Patra et al., 2013). In Arabidopsis, over-expression of the MYB TF (AtPAp1) Led to up-regulation of PAL, CHS and DER genes that subsequently enhanced production of lignin (Deluc et al., 2006). Additionally, NAC TF family has transcriptional activation effect on anthocyanin pathway genes in high light condition (Patra et al., 2013).This TF family especially present in brown and green modules. AP2/ERF-ERF and WRKY TF families present in all modules. These TF families act as regulatory proteins of major indole alkaloids pathway (Patra et al., 2013). C2H2 TF, as a member of Zinc-finger proteins (ZFPs) transcription factor family, is involved in controlling various abiotic stress and phytohormone responses, floral development (Wu et al., 2008), secondary metabolism and cell wall structure (Liu et al., 2015). In *Aspergillus nidulans*, C2H2 TF (MtfA) have been identified that regulate the secondary metabolism (Ramamoorthy et al., 2013). GRAS proteins are an important TF family that play key regulatory role in the development, abiotic stress and phytochrome signaling (Hirsch and Oldroyd, 2009). According to our data, GRAS TF family was strongly enriched in brown, yellow and green modules, this result suggests that this TF might associate with secondary metabolite production. Moreover, GRAS proteins function as regulator in GA3 signaling and biosynthesis. In *E. purpurea*, GA3 treatment resulted in higher production of caftaric acid, cichoric acid and anthocyanins. We propose GRAS TF family has the potential to secondary metabolic engineering in *E. purpurea*.

In conclusion, this study, using systems biology approach, provides a transcriptional overview of *E. purpurea*. The analysis also highlights several

Table 2. List of TF families in the first five modules

Turquoise		Blue		Brown		Yellow		Green	
TF family	N*	TF family	N	TF family	N	TF family	N	TF family	N
bHLH	5	C2H2	5	GRAS	15	GRAS	10	AP2/ERF-ERF	8
HB-HD-ZIP	5	AP2/ERF-ERF	3	AP2/ERF-ERF	14	AP2/ERF-ERF	7	WRKY	7
bZIP	4	bHLH	3	WRKY	14	MADS-MIKC	6	bHLH	6
TUB	4	MYB-related	2	C2H2	13	C2H2	6	HSF	5
WRKY	4	Trihelix	2	bHLH	9	MYB-related	6	MYB-related	5
AP2/ERF-ERF	3	C2C2-CO-like	2	NAC	8	bZIP	4	bZIP	5
B3	3	C3H	1	Tify	6	Trihelix	4	NAC	5
GARP-G2-like	3	MYB	1	MYB	5	C3H	4	GARP-G2-like	4
HB-other	2	TCP	1	HSF	4	HB-HD-ZIP	3	GRAS	4
MYB-related	2	bZIP	1	B3	3	bHLH	3	HB-HD-ZIP	3
NAC	2	FAR1	1	FAR1	3	TCP	3	HB-other	2
FAR1	2	NAC	1	bZIP	3	LOB	2	B3	2
PLATZ	2	WRKY	1	C3H	3	NAC	2	MYB	2
C2H2	2	CSD	1	DBP	2	C2C2-Dof	2	HB-BELL	2
C3H	2	GARP-G2-like	1	MYB-related	2	HB-other	2	zf-HD	2
GRAS	2	OFP	1	C2C2-GATA	2	SBP	2	C2C2-Dof	1
AP2/ERF-AP2	1	HB-BELL	1	GARP-G2-like	2	FAR1	2	RWP-RK	1
E2F-DP	1	SBP	1	C2C2-Dof	1	MYB	2	B3-ARF	1
TCP	1			HB-HD-ZIP	1	SRS	2	C2H2	1
EIL	1			TUB	1	GARP-G2-like	2	C2C2-GATA	1
Tify	1			MADS-MIKC	1	CSD	1	Trihelix	1
C2C2-GATA	1			NF-YB	1	B3	1	FAR1	1
HB-BELL	1			BES1	1	DBB	1	NF-YA	1
B3-ARF	1			C2C2-CO-like	1	NF-YB	1	C3H	1
HSF	1			LOB	1	WRKY	1		
RWP-RK	1			NF-YC	1	B3-ARF	1		
ULT	1			zf-HD	1	C2C2-GATA	1		
SRS	1			HB-other	1	E2F-DP	1		
BES1	1					HB-KNOX	1		
LOB	1					MADS-M-type	1		
						BBR-BPC	1		
						HB-BELL	1		

important TFs which their function as regulatory elements involved in secondary metabolism and stress responses. Although, some pathways have not yet been identified for *E. purpurea*, these TFs such as GRAS, AP2/ERF-ERF, bHLH, bZIP, C2H2, MYB and WRKY can be considered as promising candidates for pathway studies and manipulation toward accumulation of valuable products.

References

1. Abbasi BH., Stiles AR., Saxena PK. and Liu C-Z. (2012) Gibberellic acid increases secondary metabolite production in Echinacea purpurea hairy roots. Applied biochemistry and biotechnology 168: 2057-2066.
2. Ali MB. and McNear DH. (2014) Induced transcriptional profiling of phenylpropanoid pathway genes increased flavonoid and lignin content in Arabidopsis leaves in response to microbial products. BMC plant biology 14: 84.
3. Deluc L., Barrieu F., Marchive C.,

Lauvergeat V., Decendit A., Richard T., Carde J-P., Mérillon J-M. and Hamdi S. (2006) Characterization of a grapevine R2R3-MYB transcription factor that regulates the phenylpropanoid pathway. Plant physiology 140: 499-511.
4. Dennis G., Sherman BT., Hosack DA., Yang J., Gao W., Lane HC. and Lempicki RA. (2003) DAVID: database for annotation, visualization, and integrated discovery. Genome biology 4: R60.
5. Du Z., Zhou X., Ling Y., Zhang Z. and Su Z. (2010) agriGO: a GO analysis toolkit for the agricultural community. Nucleic acids research 38: W64-W70.
6. Fraser CM. and Chapple C. (2011) The phenylpropanoid pathway in Arabidopsis. The Arabidopsis Book, e0152.
7. Góngora-Castillo E., Childs KL., Fedewa G., Hamilton JP., Liscombe DK., Magallanes-Lundback M., Mandadi KK., Nims E., Runguphan W. and Vaillancourt B. (2012) Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. PloS one 7: e52506.
8. Graham IA., Besser K., Blumer S., Branigan CA., Czechowski T., Elias L., Guterman I., Harvey D., Isaac PG. and Khan AM. (2010) The genetic map of Artemisia annua L. identifies loci affecting yield of the antimalarial drug artemisinin. science 327: 328-331.
9. Guo K., Zou W., Feng Y., Zhang M., Zhang J., Tu F., Xie G., Wang L., Wang Y. and Klie S. (2014) An integrated genomic and metabolomic framework for cell wall biology in rice. BMC genomics 15: 596.
10. Hirsch S. and Oldroyd GE. (2009) GRAS-domain transcription factors that regulate plant development. Plant signaling & behavior 4: 698-700.
11. Hur M., Campbell AA., Almeida-de-Macedo M., Li L., Ransom N., Jose A., Crispin M., Nikolau BJ. and Wurtele ES. (2013) A global approach to analysis and interpretation of metabolic data for plant natural product discovery. Natural product reports 30: 565-583.
12. Langfelder P. and Horvath S. (2008) WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics 9: 559.
13. Lee I., Ambaru B., Thakkar P., Marcotte EM. and Rhee SY. (2010) Rational association of genes with traits using a genome-scale gene network for Arabidopsis thaliana. Nature biotechnology 28: 149-156.
14. Lee J. and Scagel CF. (2010) Chicoric acid levels in commercial basil (Ocimum basilicum) and Echinacea purpurea products. Journal of functional foods 2: 77-84.
15. Liu Q., Wang Z., Xu X., Zhang H. and Li C. (2015) Genome-wide analysis of C2H2 zinc-finger family transcription factors and their responses to abiotic stresses in poplar (Populus trichocarpa). PloS one 10: e0134753.
16. Mitsuda N. and Ohme-Takagi M. (2009) Functional analysis of transcription factors in Arabidopsis. Plant and Cell Physiology 50: 1232-1248.
17. Patra B., Schluttenhofer C., Wu Y., Pattanaik S. and Yuan L. (2013) Transcriptional regulation of secondary metabolite biosynthesis in plants. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms 1829: 1236-1247.
18. Pellati F., Benvenuti S., Magro L., Melegari M. and Soragni F. (2004) Analysis of phenolic compounds and radical scavenging activity of Echinacea spp. Journal of Pharmaceutical and Biomedical Analysis 35: 289-301.
19. Ramamoorthy V., Dhingra S., Kincaid A., Shantappa S., Feng X. and Calvo AM. (2013) The putative C2H2 transcription factor MtfA is a novel regulator of secondary metabolism and morphogenesis in Aspergillus nidulans. PLoS One 8: e74122.
20. Tsai Y-L., Chiu C-C., Chen JY-F., Chan K-C. and Lin S-D. (2012) Cytotoxic effects of Echinacea purpurea flower extracts and cichoric acid on human colon cancer cells through induction of apoptosis. Journal of ethnopharmacology 143: 914-919.
21. Vogt T. (2010) Phenylpropanoid biosynthesis. Molecular plant 3: 2-20.
22. Wu C., You C., Li C., Long T., Chen G., Byrne ME. and Zhang Q. (2008) RID1, encoding a Cys2/His2-type zinc finger transcription factor, acts as a master switch from vegetative to floral development in rice. Proceedings of the National Academy of Sciences 105: 12915-12920.
23. Xiao M., Zhang Y., Chen X., Lee E-J., Barber CJ., Chakrabarty R., Desgagné-

Penix I., Haslam TM., Kim Y-B. and Liu E. (2013) Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. Journal of biotechnology 166: 122-134.

24. Yang CQ., Fang X., Wu XM., Mao YB., Wang LJ. and Chen XY. (2012) Transcriptional regulation of plant secondary metabolism. Journal of integrative plant biology 54: 703-712.