2021

Integration of the metabolome and transcriptome reveals the metabolites and genes related to nutritional and medicinal value in Coriandrum sativum

WU Tong

FENG Shu-yan

YANG Qi-hang

Preetida J. BHETARIYA

GONG Ke

See next page for additional authors

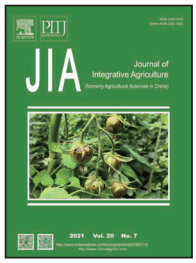
Authors

WU Tong, FENG Shu-yan, YANG Qi-hang, Preetida J. BHETARIYA, GONG Ke, CUI Chun-lin, SONG Jie, PING Xiao-rui, PEI Qiao-ying, YU Tong, and SONG Xiao-ming

Journal of Integrative Agriculture 2021, 20(7): 1807–1818

Available online at www.sciencedirect.com

ScienceDirect



RESEARCH ARTICLE

Integration of the metabolome and transcriptome reveals the metabolites and genes related to nutritional and medicinal value in *Coriandrum sativum*

WU Tong[1], FENG Shu-yan[1], YANG Qi-hang[1], Preetida J BHETARIYA[2], GONG Ke[1], CUI Chun-lin[1], SONG Jie[1], PING Xiao-rui[1], PEI Qiao-ying[1], YU Tong[1], SONG Xiao-ming[1, 3, 4]

[1] *College of Life Sciences, North China University of Science and Technology, Tangshan 063210, P.R.China*
[2] *Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA*
[3] *Food Science and Technology Department, University of Nebraska-Lincoln, Lincoln, NE 68526, USA*
[4] *Center for Informational Biology, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, P.R.China*

Abstract

Coriandrum sativum (Coriander) or Chinese parsley is a culinary herb with multiple medicinal effects, which is widely used in cooking and traditional medicine. It is enriched with essential oils and anti-oxidant compounds with unknown significance. To explore the untapped reservoir of Coriander, we studied the transcriptome and metabolic profiles from three developmental stages. Here, we identified 10 tyrosine metabolic pathway-related genes (*TMPRGs*), six porphyrins and chlorophyll metabolic pathway-related genes (*PCMPRGs*), and five Vitamin E metabolic pathway-related genes (*VEMPRGs*). These genes were associated with the early development of Coriander. Our analysis suggests that these pathways are involved in the production of critical phenolic metabolites. Furthermore, we constructed the interaction network between these pathway-related genes and transcription factors (TFs), which supported the regulatory pathways for phenolic metabolites. Interestingly, we identified several nutritional or medicinally relevant metabolites, including 59 phenols, two polyamines, 12 alkaloids, and one terpenoid. The higher concentrations of metabolites were from caffeic acid, agmatine, and its derivatives. We found higher levels of caffeic acid and agmatine at 30 days compared to 60 or 90 days. This study provides evidence to stimulate further investigation of the role of these metabolites in medicinal and nutritional research.

Keywords: metabolomics, transcriptome, regulatory network, phenolic compounds, Coriander

1. Introduction

Coriandrum sativum (Coriander) is a herbaceous plant with both nutritional and medicinal properties (Yella *et al*. 2019). Originating in the Mediterranean region, Coriander is now a global crop consumed in most parts of the world. Coriander leaves can stimulate appetite and help digestion. It secretes abundant phenolic compounds (Laribi *et al*.

Received 24 March, 2020 Accepted 29 July, 2020
WU Tong, E-mail: WuTongHBLG@163.com; Correspondence SONG Xiao-ming, Tel/Fax: +86-315-8805607, E-mail: songxm@ncst.edu.cn

doi: 10.1016/S2095-3119(20)63358-5

2015) with medicinal and nutritional properties. These phenolic compounds play essential roles in human health by regulating metabolism, reducing chronic diseases, and inhibiting cell proliferation (Cory *et al*. 2018).

The caffeic acid extracted from Coriander has antioxidant and inhibitory effects on leukemic cell growth (Habtemariam 2017). It is known to inhibit the hepatitis C virus (Tanida *et al*. 2015) and *Staphylococcus aureus* bacterium (Matejczyk *et al*. 2017). Most of these compounds are still under investigation for their anti-inflammatory, anti-fungal, and anti-tumor activities. Coriander is also rich in agmatine, which is a polyamine analog transformed by arginine decarboxylase (Laube and Bernstein 2017). Agmatine possesses a wide range of biological effects including stimulating expression of nutritional factors and adult neurogenesis (Neis *et al*. 2017). Agmatine is also a candidate drug for the treatment of severe depression (Freitas *et al*. 2016) and has been found to inhibit nicotine withdrawal-induced cognitive deficits in rats (Wiśniewska *et al*. 2017; Kotagale *et al*. 2018).

As mentioned above, Coriander has high nutritional and medicinal value, including vitamins, anti-tumor activity, and others. However, much less is known about the secretory pathways and genes involved in the production of these compounds, including phenols, caffeic acid, and agmatine. Recently, high-quality next-generation sequencing datasets from Coriander were developed through advanced methods like PacBio and Illumina sequencing technology (Choudhary *et al*. 2019; Song *et al*. 2019; Tulsani *et al*. 2020). To explore the metabolic profile, we further studied the transcriptomic profile of Coriander at three developmental stages. This study provides the guideline for harvest time with regard to the cultivation, consumption, and processing of Coriander without losing most of its nutritional value. Here, using RNA-seq and metabolomics datasets, we report the differentially expressed genes (DEGs) and differential metabolites (DMs) from three developmental stages of Coriander.

2. Materials and methods

2.1. Materials and growth conditions

The Coriander line "SJ01" was selected for this study, as the genome sequence from this species has been published recently (Song *et al*. 2019). Seeds were grown in pots containing soil:vermiculite mixture (3:1) in a controlled-environment growth chamber. The normal growth conditions were programmed for 16/8 h at 22/18°C for day/night, relative humidity of ~60%, and a light intensity of 3 000 lx.

According to the growth phases of Coriander, three developmental stages of young seeding, metaphase, and anaphase at 30, 60, and 90 days, respectively, were examined in this study. At 30 days, Coriander was in the fast-growing stage, and the average height was ~9.6 cm. At 60 days, it was in the stable growth stage, and the average height was ~15.8 cm. At 90 days, Coriander was in the late vegetative growth stage, and the average height was ~19.5 cm. We randomly selected Coriander leaves from 11 samples at each developmental stage, which were used for transcriptomics (three biological replicates) and metabolic (eight replicates) studies.

2.2. Transcriptomics analysis

RNA extraction, library construction, and sequencing All samples were treated with liquid nitrogen for quick freezing and stored at –80°C until further use. The RNA was extracted with an RNA Kit (TIANGEN, Beijing, China) following the manufacturer's instructions and then reverse transcribed into cDNA using PrimeScript RT Reagent Kit (TaKaRa, Dalian, China). The ribosomal RNA was removed with Ribo-zero™ rRNA Removal kit. The mRNA sample, thus enriched, was subjected to fragmentation buffer and then cDNA synthesis using random hexamer primers.

Libraries were prepared per manufacturer's instructions and sequenced on Illumina HiSeq™ 4000 for paired-end sequencing. For the transcriptome, three replicates at 30 days were named as 30 d1–3, replicates at 60 days were named as 60 d1–3, and replicates at 90 days were named as 90 d1–3. The RNA-seq dataset produced is available in the Genome Sequence Archive in BIG Data Center and CGDB database (accession no. CRA001656) (http://bigd.big.ac.cn/gsa) (BIG Data Center Members 2019; Song *et al*. 2020).

Gene expression analysis and qRT-PCR Raw reads were filtered to obtain high quality reads using base-calling software from Illumina. Filtered reads were aligned to the Coriander genome using HISAT Software (Kim *et al*. 2015). Using Python-based HTSeq Software (v0.11.2), we analyzed the gene expression level for each transcript (Anders *et al*. 2015). The expected number of fragments per kilobase of transcript sequence per million of base pairs sequenced (FPKM) was used as the normalization method, which considers the effect of sequencing depth and gene length on fragment counts (Trapnell *et al*. 2010). DESeq2 Software was used to conduct DEGs detection with padj<0.05 (Love *et al*. 2014).

The expression of selected genes was also checked using quantitative real-time PCR (qRT-PCR) with three replicates according to our previously reported methods (Song *et al*. 2014, 2016). The mRNA was translated to cDNA using PrimeScript cDNA Synthesis Kit (TaKaRa, Dalian, China). Then, the cDNA was detected quantitatively by CFX96™ Real-Time System (C1000™ Thermal Cycler, Bio-Rad) using SYBR_Premix *Ex Taq*™ II Kit (TaKaRa, Dalian, China). The primer sequences of examined genes

are list in the Appendix A1.

DEGs detection and cluster analysis The hypothesis test probability was calculated according to a negative binomial distribution model. The false discovery rate was derived using multiple hypothesis test corrections. Finally, DEGs were defined by the criteria of padj<0.05 and $\log_2$ |fold-change|>1 (Anders and Huber 2010).

We used the H-cluster R script to cluster the relative expression levels of differential genes. The clustering algorithm classified the DEGs into several clusters, and the genes in the same cluster had similar expression trends. The common and specific numbers of DEGs groups were plotted with Venn diagrams using R script.

Gene function annotation and enrichment analyses The Gene Ontology (GO) enrichment analysis of DEGs was performed using GOseq Software based on non-central hypergeometric distribution with the corrected *P*-value<0.05 (Young *et al*. 2010). GO annotations were obtained through the HMMscan search. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was conducted using KOBAS Software (Kanehisa *et al*. 2008).

2.3. Metabolomics analysis

Metabolite extraction and detection The Coriander leaves were also simultaneously collected for metabolomics analysis and subjected to liquid chromatography-mass spectrometry (LC-MS) technology (Artati *et al*. 2019). Eight replicates at 30 days were named as xM1–xM8 (30 d), replicates at 60 days were named as yM1–yM8 (60 d), and replicates at 90 days were named as zM1–zM8 (90 d).

The data for retention times and peak shapes were extracted using Compound discover software from ThermoFisher scientific (https://www.thermofisher.com/us/en/home.html). The full-scan mass spectrogram was used to determine the exact molecular weight, mass number deviation, and ion information. The compounds were identified by searching through the mzCloud database (https://www.mzcloud.org/), which consists of the fragmentation mode, energy, actual collected secondary, and multi-stage mass spectrometry data.

Differential metabolites (DMs) detection We conducted a multivariate statistical analysis of metabolites, including principal component analysis (PCA) and partial least squares discrimination analysis (PLS-DA), to reveal the differences of the metabolic compositions among different developmental stages (David and Jacobs 2014; Gromski *et al*. 2015). The variable importance in the projection (VIP) value of the first principal component with the PLS-DA Model and the *P*-value of *T*-test were used to detect the DMs (Cho *et al*. 2008). The thresholds were set as VIP>1, fold-change >2.0 or <0.5, and *P*-value<0.05 according to a previous report (Li *et al*. 2019). The hierarchical clustering and metabolite-metabolite correlation analysis were conducted to reveal the relationship between metabolites and samples (Beshir *et al*. 2019). The *Z*-score was calculated from the mean and standard deviation of the reference dataset (Ben Salah *et al*. 2017).

2.4. Integrative analysis of metabolome and transcriptome

The Pfam database (http://pfam.sanger.ac.uk/) was searched to identify transcription factors (TFs) with the *E*-value<1e-4 setting. Pearson's correlation coefficients (PCCs) between metabolic pathway-related genes and TFs were calculated using in-house Perl scripts from the gene expression values of the three developmental stages. The positive and negative regulatory relationships were defined as PCC>0.99 and PCC<–0.99, respectively. The interaction networks between metabolic pathway-related genes and TFs were constructed using Gephi Software (https://gephi.org). For pathway-specific gene sequences, such as tyrosine and Vitamin E metabolism pathways, we compared Coriander sequences with the related *Arabidopsis* genes in these pathways using Basic Local Alignment Search Tool (BLAST), and the screening criteria were *E*-value<1e-5 with identity value>50%.

3. Results

3.1. Transcriptomic analyses of the three developmental stages

DEGs analyses The high-quality bases from nine samples, including three developmental stages of Coriander with three replicates, were over 6.98 Gb (Appendix B1-a). The base percentages of Q30 in all samples were over 91.07% (Appendix B1-b). The total matching rates of filtered reads to the reference genome were more than 92.1% (Appendix B1-c). A total of 30 026 (73.68%) Coriander genes were detected in all three developmental stages, and 6 801 genes had no expression in any of the three developmental stages. The PCCs among each replicate were all greater than 0.84 (Appendix A2), which indicated that the replicates were repeatable and accurate.

We further identified a list of DEGs from the three developmental stages of Coriander (Appendices A3–A5). Throughout the manuscript, we will refer to 30 d *vs*. 60 d as Group I, 60 d *vs*. 90 d as Group II, and 30 d *vs*. 90 d as Group III for the comparisons. In Group I, we found a total of 168 DEGs, of which 149 were up-regulated, and 19 were down-regulated (Fig. 1-A). Among Group II, we found 409 DEGs, including 224 up-regulated and 185 down-regulated.

In Group III, only 19 DEGs were obtained, including 12 up-regulated and seven down-regulated. Six candidate genes, mainly involved in the chlorophyll metabolic pathway, were further selected for qRT-PCR examination. The quantitative expression with PCR was consistent with the transcriptome results (Appendix B2).

The Venn diagram shows the shared and unique numbers of genes among all three groups (Fig. 1-B). There were 96 unique DEGs for Group I, 324 DEGs for Group II, and five DEGs for Group III. Interestingly, one gene *Cs09G04288* was found to be a common gene expressed among all developmental stages. Based on the H-cluster analysis, six sub-clusters were obtained according to the expression trends or patterns at the three developmental stages (Appendix B3). Within each sub-cluster, the genes had similar expression trends, and so these genes might play a synergistic role in different development stages of Coriander.

Functional enrichment analyses The GO enrichment analysis showed the distribution of genes among various functions (Appendices A6–A8). In total, 31 and eight terms were enriched in up-regulated genes in 30 d compared with 60 and 90 d, respectively (Appendices A6–A7, B4). However, only two enrichment terms were detected in down-regulated genes in 60 d *vs*. 90 d (Appendices A8 and B5). Interestingly, we found most of the up-regulated genes in 30 d compared with the other two stages were enriched in the photosynthesis pathway.

Genes found to be upregulated in Group I (Appendices B6 and A9), Group II, and Group III (Appendices B6, A10–A11) were further explored with KEGG metabolic pathway analysis. The results showed the interaction of pathways between photosynthesis, photosynthesis-antenna protein, Porphyrin, and chlorophyll metabolism, which was consistent with the GO enrichment analyses.

For the photosynthesis pathway, we detected 12 DEGs in Group I (Appendix A9). Among them, seven and five genes were found to be part of the photosynthetic system I and II related genes, respectively. All of these genes were found to be up-regulated during the early developmental stage (Appendices A3 and A9). Similarly, eight DEGs in Group II also belonged to photosynthesis pathway, including five from the photosynthetic system I, and three from system II. The expression analysis showed that all of them were up-regulated in 30 d compared with 90 d (Appendices A4 and A10).

We identified a total of 20 and 13 upregulated genes in Group I and Group II, respectively. These genes were part of the photosynthesis-antenna protein pathway (Appendices A9–A10). Moreover, these DEGs were all up-regulated in 30 d compared with 60 d and 90 d, and they were involved in the chlorophyll protein complex synthesis (Appendices A3–A4, A9–A10). In the porphyrin and chlorophyll metabolism pathway, we identified seven DEGs, and all of them were up-regulated in 30 d compared with 60 d (Appendix A9). Specific enzymes encoded by these seven DEGs were involved in the synthesis of chlorophyll (Appendices A3 and A9).

In conclusion, the up-regulated genes in photosynthesis and chlorophyll metabolism pathways were enriched in the early developmental stages compared to late stages. This indicates these genes play a crucial role in the early development of Coriander.

3.2. Metabolic analyses of the three developmental stages

Quantitative assessment of metabolites We present the metabolic analyses of the three developmental stages (30,

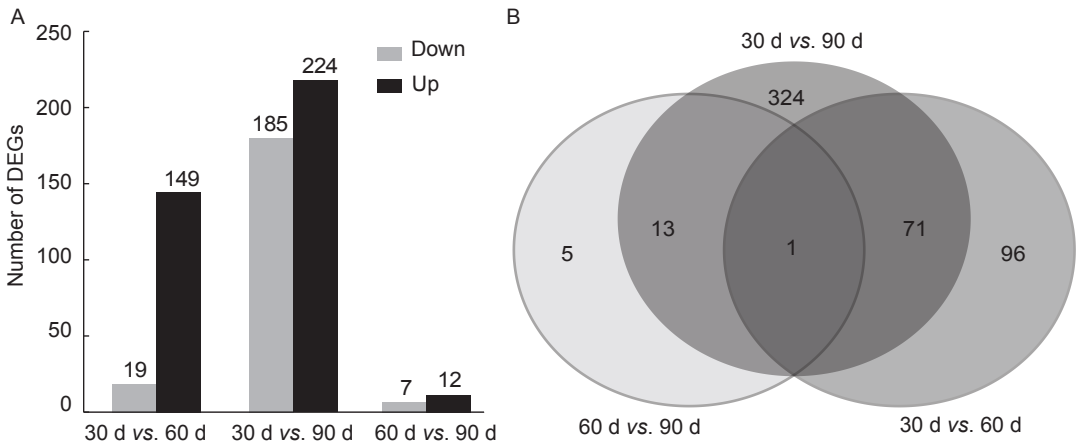


Fig. 1 Transcriptomics analysis of the three developmental stages in Coriander. A, numbers of differentially expressed genes (DEGs) among the three developmental stages. B, the Venn diagrams of DEGs among the three developmental stages.

60, and 90 d) of Coriander with eight replicates (Appendix A12). The correlation analysis yielded PCC close to one which assured the high-quality of samples and stability of the detection method (Appendix A13).

The metabolites with positive (pos) metabolites and negative (neg) ion modes were quantitatively analyzed and compared among different stages of the samples. The comparison groups were 60 d *vs*. 30 d (neg/pos), 90 d *vs*. 30 d (neg/pos), 90 d *vs*. 60 d (neg/pos) (Appendices A14–A19). The PCA diagram of all samples showed the independent grouping of the metabolites among samples from 60 d *vs*. 30 d, and 90 d *vs*. 30 d (Appendix B7-a–d). However, the grouping among metabolites in 90 d *vs*. 60 d was not very clear (Appendix B7-e–f). This phenomenon showed that there was a significant difference in the types of metabolites produced in 60 d *vs*. 30 d or 90 d *vs*. 30 d.

Furthermore, the PLS-DA models for each comparison group were established. The model evaluation parameter R^2 was obtained by 7-fold cross-validation. We found R^2 values of all samples were larger than 0.93, indicating the model used in this study was stable and reliable. Using this model, we identified the trend of metabolites between groups (Appendix B8). We later sorted and verified the PLS-DA model for all comparisons. After verification, the Q2 of the model did not exceed the corresponding lines, indicating that the model did not appear to be overfitting (Appendix B9). Not "overfitting" indicated that the model could better describe the sample and could be used as a prerequisite for the model biomarker search. The mutual verification of these two models showed that it was feasible to group and compare metabolites in different developmental stages.

Differential metabolites (DMs) analyses The nutrients in Coriander leaves have not been fully studied until now. Based on our metabolic study, we detected higher concentrations of phenolic (59), two polyamines, 12 alkaloids, and one diterpene compounds. The secretion of these compounds differed in the three developmental stages (Appendix A20). The subsequent analysis showed that 16 phenolic compounds belong to caffeic acid and its derivatives. The concentrations of these compounds were higher at 30 d compared to 60 and 90 d (Fig. 2-A). We found the concentrations of two polyamines, i.e., agmatine and its derivatives, were significantly higher at 30 days than the later stages (Appendix B8-a). Agmatine has been discussed as a putative neurotransmitter (Stickle *et al*. 1996). Caffeic acid is a known antioxidant compound and has been reported by several groups.

Overall, there were more significant DMs in 90 d *vs*. 30 d than the other two comparisons, whether using positive ion (1 025) or negative ion (442) modes (Table 1). In contrast, the least significant DMs were found in 90 d *vs*. 60 d. Interestingly, we found that all the DMs which were down-regulated with both ion modes were more than those up-regulated in both 60 d *vs*. 30 d and 90 d *vs*. 30 d. This phenomenon indicated that the content of similar types of DMs in 30 d was higher than in 60 and 90 d. However, there were no differences for up- or down-regulated DMs in 90 d *vs*. 60 d. In addition, the results showed that in the comparison of 60 d *vs*. 30 d and 90 d *vs*. 30 d, the DMs with positive ion mode numbered more than with the negative ion mode (Fig. 2-B; Table 1).

The Venn diagram showed the numbers of common and specific DMs between the comparison groups (Fig. 2-C). The results showed that several specific DMs with negative/positive ion modes were detected in 60 d *vs*. 30 d (83/202), 90 d *vs*. 30 d (90/288), and 90 d *vs*. 60 d (68/74). A total of 16 DMs with negative and 21 with positive ion modes were common among all three comparisons. In addition, the *Z*-score analysis showed a significant difference in the relative contents of DMs with negative ion mode at the same level of all three comparisons (Appendices B10 and B11).

KEGG pathway enrichment analysis for metabolites KEGG path enrichment analysis of DMs showed the changes in metabolic pathways during Coriander development (Appendices A21 and A22). Interestingly, both 60 d *vs*. 30 d and 90 d *vs*. 30 d DMs with negative ion mode were significantly enriched in the tyrosine metabolism pathway (Fig. 2-D; Appendices B12–B13, A21–A22). In this pathway, 17 DMs were detected, which were down-regulated in 60 d compared with 30 d (Appendix A21). A total of 16 DMs was found in 90 d *vs*. 30 d, and all of them were down-regulated in 90 d compared with 30 d (Appendix A22).

It is worth noting that most DMs of tyrosine metabolism pathways were the same in 60 d *vs*. 30 d and 90 d *vs*. 30 d, and only DMs Com-41-6neg (caffeic acid and its derivatives) was different between the two comparisons (Appendices A21 and A22). Compared with the middle and late stages of Coriander development, the metabolic pathways in the early stage were significantly different, which was consistent with the results obtained by the transcriptome enrichment analysis.

3.3. Integrative analysis of transcriptome and metabolome

Based on the analysis of transcriptomes and metabolomes in different development stages of Coriander, we found that the expression trend of several genes was consistent with the trend of downstream metabolite content in multiple metabolic pathways, including *TMPRGs*, *PCMPRGs*, and *VEMPRGs*. Furthermore, we constructed an interaction network between these pathway-related genes and TFs according to PCCs using Gephi Software (Fig. 3).

Tyrosine metabolic pathway The tyrosine metabolic

pathway was significantly enriched in 60 d *vs*. 30 d and 90 d *vs*. 30 d (Appendices A21–A22; Fig. 4-A). The contents of fumarylacetoacetic acid, caffeic acid, and their derivatives of this pathway in 60 and 90 d were significantly lower than in 30 d (Fig. 4-B; Appendix A23).

Caffeic acid is a phenolic compound, which plays an important role in regulating plant growth (Li *et al*. 2018; Liu *et al*. 2018). However, the scope of our study is limited to this finding, and we did not explore the role of these compounds in plant growth. We found that 16 compounds belonging to the caffeic acid category were enriched after 60 d, and the corresponding tyrosine metabolic pathway genes were upregulated. This finding warrants the exploration of these phenolic compounds in detail.

Meanwhile, we retrieved five protein sequences that encode enzymes from the tyrosine metabolism pathway in *Arabidopsis* and one protein sequence from carrot. We obtained Coriander homologous genes by comparing these protein sequences using BLAST (Appendix A24). By integrating RNA-seq and metabolomics analysis, we

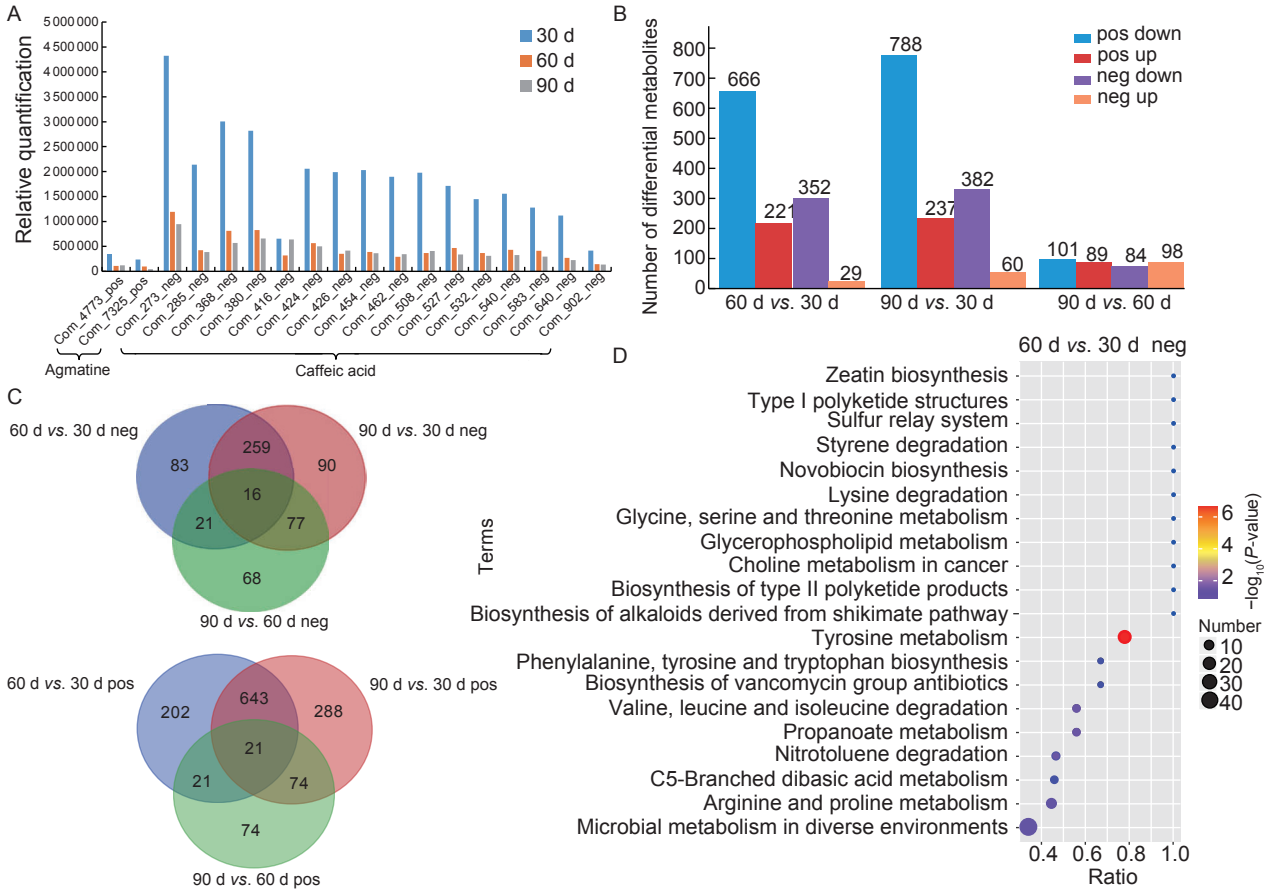


Fig. 2 Differential metabolite (DM) analysis of the three developmental stages in Coriander. A, the relative quantification of caffeic acid along with its derivatives, and agmatine. B, the number of DMs among the three developmental stages. C, the Venn diagrams of metabolites with positive (pos) and negative (neg) ion models. D, scatter plot of KEGG enriched pathways of the DMs in 60 d *vs*. 30 d.

Table 1 Summary of metabolites for comparisons of the three developmental stages of Coriander

Compared sample[1)]	Number of metabolites	Number of significant metabolites	Number of up-rich metabolites	Number of down-rich metabolites
60 d *vs*. 30 d pos	4 998	887	221	666
90 d *vs*. 30 d pos	4 998	1 025	237	788
90 d *vs*. 60 d pos	4 998	190	89	101
60 d *vs*. 30 d neg	1 569	379	27	352
90 d *vs*. 30 d neg	1 569	442	60	382
90 d *vs*. 60 d neg	1 569	182	98	84

[1)] The pos indicated the positive ion, and the neg indicated the negative ion.

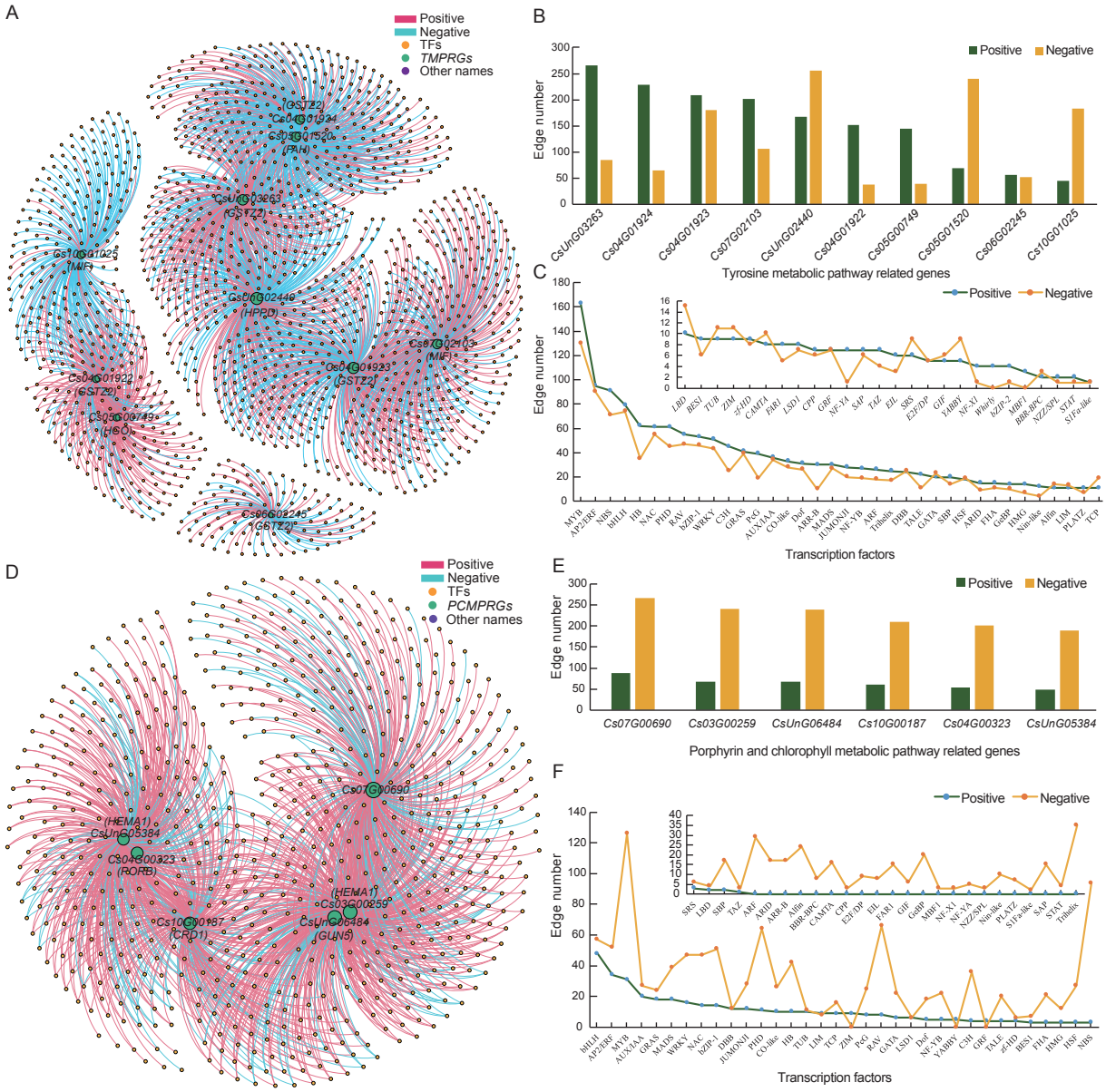


Fig. 3 The interaction network analyses between transcription factors (TFs) and tyrosine metabolic pathway-related genes (*TMPRGs*) or porphyrin and chlorophyll metabolic pathway-related genes (*PCMPRGs*) in Coriander. A, the interaction network constructed according to the Pearson's correlation coefficient (PCC) between *TMPRGs* and TFs. B, edge number of positive (PCC>0.99) and negative (PCC<–0.99) regulatory interactions for each examined *TMPRG*. C, the edge number of interactions in the *TMPRGs* network for each TF. D, the interaction network constructed according to the PCC between *PCMPRGs* and TFs. E, edge number of positive and negative regulatory interactions for each examined *PCMPRG*. F, edge number of interactions in the *PCMPRG* network for each TF.

found the expression of several genes in tyrosine metabolic pathway had similar trends with the higher contents of these metabolites among the three stages, such as *Cs05G01520*, *Cs10G01025*, and *CsUnG02440* (Fig. 4-B; Appendices B14–B15, A25). The Arabidopsis gene *AT2G02380* was matched with five genes in Coriander (Appendix A24). Among these genes, three genes (*Cs04G01922*, *Cs04G01923*, and *Cs04G01924*) were on chromosome 4, and they belonged to tandem repeat genes. The homology score of gene *Cs04G01923* was the lowest in the BLAST analysis, which indicated that this gene may have a different arrangement of tandem repeats and is unique to Coriander. The annotation showed that *Cs04G01922* encoded an unnamed protein product; *Cs04G01923* and *Cs04G01924* encoded glutathione *S*-transferase zeta class family protein. We also found that the expression of *Cs04G01923* and *Cs04G01924* was lower than that of *Cs04G01922* at the three stages (Fig. 4-D). The role of unknown gene

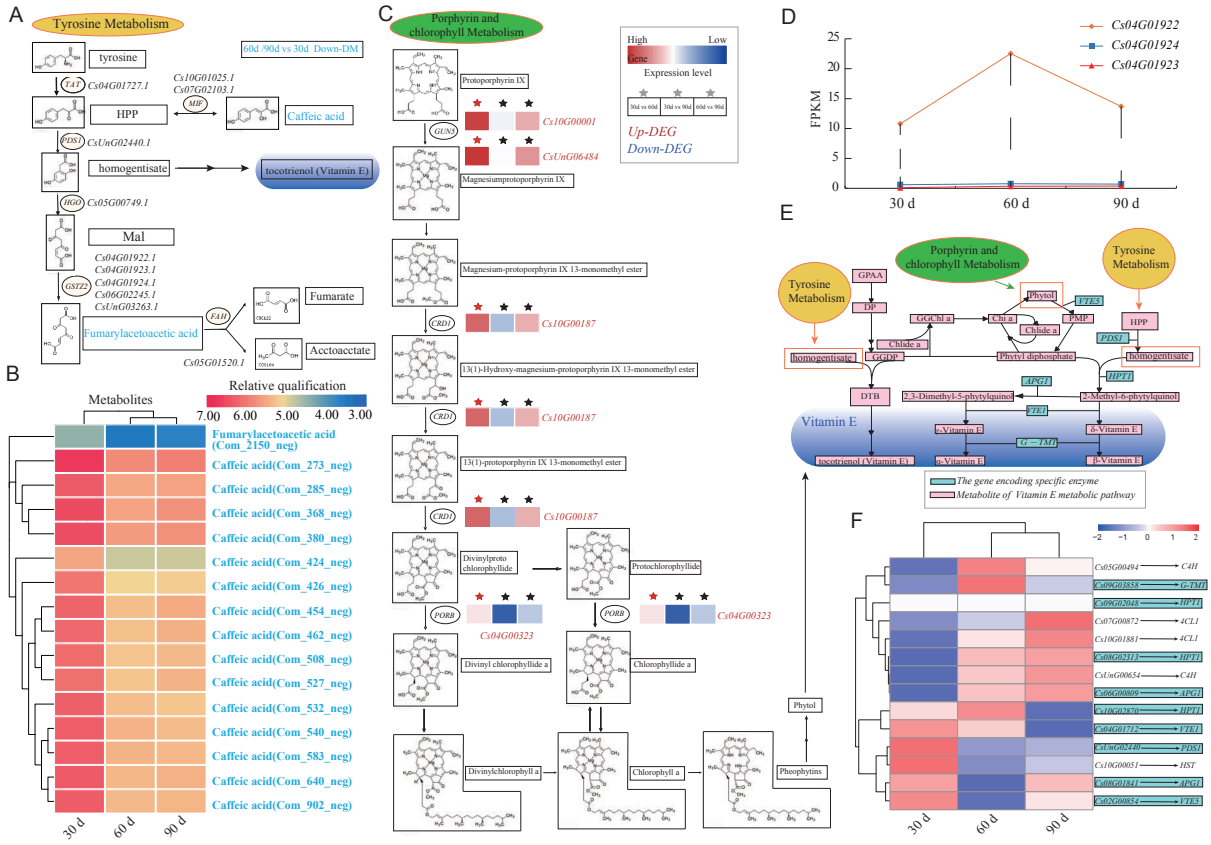


Fig. 4 Integration analysis of transcriptome and metabolome in Coriander. A, tyrosine metabolism pathway. Down-regulated differential metabolites (DMs) at 60 and 90 d compared with 30 d are marked by blue color. B, heat map of DMs in tyrosine metabolism at the three developmental stages. C, porphyrin and chlorophyll metabolism pathway. Up-regulated differentially expressed genes (DEGs) at 30 d compared with 60 and 90 d are marked by red color. D, a line diagram of the gene expression for three genes. E, vitamin E metabolic pathway. F, the heatmap of *DEGs* in important genes in vitamin E metabolic pathways. The expression values were log-transformed for the cluster analyses. Red and blue colors represent high and low expression genes, respectively. The descriptions of abbreviations for genes and enzymes are list in Appendices A24, A28, A31, and A32.

Cs04G01922 may be of interest considering its potential role in development.

There are many studies indicating the role of transcription factors in the development of plants (Zhang *et al*. 2019; Zhao *et al*. 2019). Based on the expression of *TMPRGs* and TFs, we constructed the interaction network between them according to the PCC using Gephi Software (Fig. 3). In the network, we detected 1 534 and 1 241 positive and negative regulatory interactions among *TMPRGs* and TFs, respectively (Fig. 3-A; Appendix A26). There were more interactive edges for three *TMPRGs* (*CsUnG02440*, *Cs04G01923*, and *CsUnG03263*), and the total edge number was over 350 (Fig. 3-B; Appendix A26). Seven *TMPRGs* had more edges with positive regulatory with TFs than positive regulatory, while we did not find a similar effect for the other three genes. In total, 61 kinds of TFs had a regulatory relationship with *TMPRGs* (Fig. 3-C; Appendix A27). The *MYB* TFs (293) had the most edges in the network, followed by *AP2*/*ERF* (185), *NBS* (162), and *bHLH* (153). In addition, 45 kinds of TFs (72.58%) had higher edges number with the positive regulatory than negative with *TMPRGs* (Fig. 3-C; Appendix A27).

Porphyrin and chlorophyll metabolism pathway In the transcriptome analysis of 30 d *vs*. 60 d, we identified seven significantly up-regulated genes enriched in porphyrin and chlorophyll metabolism using the KEGG (Appendices A9, A28 and B16). Four of these seven genes were involved in the formation of pheophytin (Fig. 4-C), which can produce phytol under the action of chlorophyll-degrading enzymes (Krautler 2016). In the integrative analysis of RNA-seq and metabolomics, we found the expression of these seven genes had similar trends with the content of metabolites of aminolevulinate at the three stages (Fig. 4-C; Appendices B17 and A28).

Based on the expression, we constructed the interaction network between *PCMPRGs* and TFs according to PCCs (Fig. 3). We detected 387 positive and 1 348 negative regulatory interactions between six *PCMPRGs* and TFs

(Fig. 3-D; Appendix A29). In this network analysis, there were more interactive edges for only one *PCMPRG* (*Cs07G00690*), and the total edge number was over 350 (Fig. 3-E; Appendix A29). All *PCMPRGs* had more edges with negative regulatory with TFs, which was significantly different from the *TMPRGs* network. In total, 60 kinds of TFs had a regulatory relationship with *PCMPRGs* (Fig. 3-F; Appendix A30). However, most kinds of TFs (54, 90.00%) had higher edges number with the negative regulatory than positive regulatory with *PCMPRGs*. Only four TFs showed the higher edges number with positive regulatory with *PCMPRGs*, which was just the opposite from the *TMPRGs* network (Fig. 3-F; Appendix A30).

Vitamin E metabolic pathway Based on the results above, we sorted out the pathway of synthesizing vitamin E from homogentisate and phytol (Fig. 4-E; Appendix A31) (Soll *et al*. 1985; DellaPenna 2005). We compared the related protein sequences of the Vitamin E metabolic pathway in *A. thaliana* with the protein sequences of Coriander by BLAST (Appendix A32). Combining it with the RNA-seq, we found that the expression of *VTE1*, *PDS1*, *HST*, and *VTE5* was higher at 30 d than that at 60 and 90 d (Fig. 4-F; Appendix A33).

Based on the expression, we constructed the interaction network between *VEMPRGs* and TFs according to PCCs (Appendix B18). In the network, 503 positive and 505 negative regulatory relationships between five VEMPRGs and TFs were detected (Appendices B18-a and A34). We found more than 313 interactive edges for gene *Cs08G02313* (Appendices B18-b and A34). We found a total of 60 kinds of TFs had a regulatory relationship with *VEMPRGs* (Appendices B18-c and A35).

4. Discussion

In this study, a large number of metabolites were significantly enriched during the growth stage of 30 days compared with 60 and 90 d in Coriander. Our results showed no significant differences in transcriptomics and metabolomics changes between 60 and 90 d of growth. We believe that the best harvest period for Coriander with enhanced nutrient compounds and yield is 30–60 d. Several key genes and metabolites were detected, which were involved in the tyrosine metabolic pathway, porphyrin and chlorophyll metabolism pathway, and vitamin E metabolic pathway.

Based on previous research, it has been established that tyrosine is catalyzed by tyrosine aminotransferase (TAT) to form 4-hydroxyphenypyruvate (HPP) in the tyrosine metabolic pathway (Munir *et al*. 2019). Caffeic acid and its derivatives can also produce HPP under the action of macrophage migration inhibitory factor (MIF). Then, p-hydroxyphenylpyruvate dioxidase (PDS1) catalyzes HPP to reduce homogentisate, which is oxidized by homogentisate 1,2-dioxygenase (HGO) to produce maleylacetoacetate. This forms fumarylacetoacetic acid under the catalysis of glutathione transferase (GSTZ2) (Wang *et al*. 2019). Finally, fumarylacetoacetic acid forms fumaric acid and acetoacetic acid (Schenck and Maeda 2018). In this study, we found the contents of fumarylacetoacetic acid, caffeic acid, and their derivatives in 30 d were significantly higher than in 60 or 90 d. These results indicated that the tyrosine pathway plays important roles in the coriander development at the early stage.

In plants, homogentisate can produce vitamin E under the action of specific enzymes (Grusak and DellaPenna 1999). Vitamin E is synthesized on the chloroplast membrane to prevent photooxidative stress, thus reducing the production of toxic free radicals and protecting photo-cooperative devices (Ivanov 2014). Vitamin E is one of the key substances related to photosynthesis (Smirnoff and Wheeler 2000). Previous studies have shown that vitamin E can affect plant development by regulating the content of jasmonic acid in leaves (Munne-Bosch 2005; Munne-Bosch *et al*. 2007). Therefore, we speculated that the decrease of caffeic acid induces the reduction of homogentisate, leading to a low amount of vitamin E production. This result partially explained the decrease of photosynthesis found at 30 d *vs*. 60 d according to the transcriptomics analyses.

Similar studies have shown that the free phytol produced by pheophytin degradation can be used as the precursor of vitamin E synthesis through the compensation pathway (Karunanandaa *et al*. 2005). The content of vitamin E might be reduced with the decrease in phytol degradation. We believe the slow growth of the Coriander at 60 days could be linked with decreased vitamin E.

In this study, we found a greater number of up-regulated genes enriched in the photosynthesis and chlorophyll metabolism pathways in 30 days. Our analysis showed increased amounts of metabolites in 30 d relative to 60 and 90 d. This phenomenon might be due to the fact that Coriander was in a fast-growing stage at 30 d, and needed to synthesize a large number of substances for growth. Therefore, the expression of genes related to photosynthesis and chlorophyll metabolism was significantly up-regulated, and the metabolite content was significantly increased at 30 d compared to the other two stages.

Interestingly, caffeic acid is a substance with great benefits to human health. The antioxidant, anti-inflammatory, anti-fungal, and anti-tumor activities of caffeic acid indicate its important role in human health (Bhat *et al*. 2020; Ning *et al*. 2020; Zhang *et al*. 2020). Coriander contains an agmatine compound, which is a known neuromodulator and has a wide range of biological effects, such as the treatment of severe depression, anti-atherosclerotic action, and anti-oxidant

activity in neuronal cell injury (Ferlazzo *et al*. 2020; Li *et al*. 2020). In the metabolic analysis, the content of agmatine in Coriander growing at 30 days was significantly higher than that in 60 days. Furthermore, the significantly higher expression of caffeic acid and agmatine related genes indicated that they play important roles at the early-stage development. This study provides the basis for further research on these phenolic compounds extracted from Coriander for their effects on human health and medicine.

5. Conclusion

Our analyses indicated that the tyrosine metabolic, porphyrin, chlorophyll, and vitamin E metabolic pathways are involved in the production and regulation of important phenolic metabolites in Coriander. These pathways and genes are also crucial for early development of the plant. The interaction network construction of these pathway-related genes and TFs provides strong support for the activation of genes related to the nutritional and medicinal value of Coriander.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (31801856), the Hebei Province Higher Education Youth Talents Program, China (BJ2018016), and the Hebei Province Postgraduate Demonstration Course (Genomics) Construction Project in 2018, China (KCJSX2018053).

Declaration of competing interest

The authors declare that they have no conflict of interest.

Appendices associated with this paper are available on http://www.ChinaAgriSci.com/V2/En/appendix.htm

References

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biology*, **11**, R106.

Anders S, Pyl P T, Huber W. 2015. HTSeq — A Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.

Artati A, Prehn C, Adamski J. 2019. LC-MS/MS-based metabolomics for cell cultures. *Methods in Molecular Biology*, **1994**, 119–130.

Ben Salah N, Bejar D, Snène H, Ouahchi Y, Mehiri N, Louzir B. 2017. The Z-score: A new tool in the interpretation of spirometric data. *La Tunisie Medicale*, **95**, 767–771.

Beshir W F, Tohge T, Watanabe M, Hertog M, Hoefgen R, Fernie A R, Nicolaï B M. 2019. Non-aqueous fractionation revealed changing subcellular metabolite distribution during apple fruit development. *Horticulture Research*, **6**, 98.

Bhat W F, Ahmed A, Abbass S, Afsar M, Bano B, Masood A. 2020. Deciphering the nature of caffeic acid to inhibit the HSA aggregation induced by glyoxal. *Protein and Peptide Letters*, **27**, 725–735.

BIG Data Center Members. 2019. Database resources of the BIG data center in 2019. *Nucleic Acids Research*, **47**, D8–D14.

Cho H W, Kim S B, Jeong M K, Park Y, Miller N G, Ziegler T R, Jones D P. 2008. Discovery of metabolite features for the modelling and analysis of high-resolution NMR spectra. *International Journal of Data Mining and Bioinformatics*, **2**, 176–192.

Choudhary S, Naika M B N, Sharma R, Meena R D, Singh R, Lal G. 2019. Transcriptome profiling of Coriander: A dual purpose crop unravels stem gall resistance genes. *Journal of Genetics*, **98**, 19.

Cory H, Passarelli S, Szeto J, Tamez M, Mattei J. 2018. The role of polyphenols in human health and food systems: A mini-review. *Frontiers in Nutrition*, **5**, 87.

David C C, Jacobs D J. 2014. Principal component analysis: A method for determining the essential dynamics of proteins. *Methods in Molecular Biology*, **1084**, 193–226.

DellaPenna D. 2005. A decade of progress in understanding vitamin E synthesis in plants. *Journal of Plant Physiology*, **162**, 729–737.

Ferlazzo N, Curro M, Giunta M L, Longo D, Rizzo V, Caccamo D, Ientile R. 2020. Up-regulation of HIF-1α is associated with neuroprotective effects of agmatine against rotenone-induced toxicity in differentiated SH-SY5Y cells. *Amino Acids*, **52**, 171–179.

Freitas A E, Neis V B, Rodrigues A L S. 2016. Agmatine, a potential novel therapeutic strategy for depression. *European Neuropsychopharmacology*, **26**, 1885–1899.

Gromski P S, Muhamadali H, Ellis D I, Xu Y, Correa E, Turner M L, Goodacre R. 2015. A tutorial review: Metabolomics and partial least squares-discriminant analysis — a marriage of convenience or a shotgun wedding. *Analytica Chimica Acta*, **879**, 10–23.

Grusak M A, DellaPenna D. 1999. Improving the nutrient composition of plants to enhance human nutrition and health. *Annual Review of Plant Physiology and Plant Molecular Biology*, **50**, 133–161.

Habtemariam S. 2017. Protective effects of caffeic acid and the Alzheimer's brain: An update. *Mini Reviews in Medicinal Chemistry*, **17**, 667–674.

Ivanov B N. 2014. Role of ascorbic acid in photosynthesis. *Biochemistry Biokhimiia*, **79**, 282–289.

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Research*, **36**, D480–D484.

Karunanandaa B, Qi Q, Hao M, Baszis S R, Jensen P K, Wong Y H, Jiang J, Venkatramesh M, Gruys K J, Moshiri F, Post-Beittenmiller D, Weiss J D, Valentin H E. 2005. Metabolically

engineered oilseed crops with enhanced seed tocopherol. *Metabolic Engineering*, **5–6**, 384–400.

Kim D, Langmead B, Salzberg S L. 2015. HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, **12**, 357–360.

Kotagale N R, Ali M T, Chopde C T, Umekar M J, Taksande B G. 2018. Agmatine inhibits nicotine withdrawal induced cognitive deficits in inhibitory avoidance task in rats: Contribution of α(2)-adrenoceptors. *Pharmacology Biochemistry and Behavior*, **167**, 42–49.

Krautler B. 2016. Breakdown of chlorophyll in higher plants — phyllobilins as abundant, yet hardly visible signs of ripening, senescence, and cell death. *Angewandte Chemie International Edition*, **55**, 4882–4907.

Laribi B, Kouki K, M'Hamdi M, Bettaieb T. 2015. Coriander (*Coriandrum sativum* L.) and its bioactive constituents. *Fitoterapia*, **103**, 9–26.

Laube G, Bernstein H G. 2017. Agmatine: Multifunctional arginine metabolite and magic bullet in clinical neuroscience? *The Biochemical Journal*, **474**, 2619–2640.

Li X F, Zhu J Y, Tian L X, Ma X Y, Fan X, Luo L, Yu J, Sun Y, Yang X, Tang W Q, Ma W, Yan J, Xu X, Liang H P. 2020. Agmatine protects against the progression of sepsis through the imidazoline I2 receptor-ribosomal S6 kinase 2-nuclear factor-κB signaling pathway. *Critical Care Medicine*, **48**, e40–e47.

Li X L, Zhang X P, Ye L, Kang Z J, Jia D H, Yang L F, Zhang B. 2019. LC-MS-based metabolomic approach revealed the significantly different metabolic profiles of five commercial truffle species. *Frontiers in Microbiology*, **10**, 2227.

Li X Y, Wang L H, Wang S M, Yang Q, Zhou Q, Huang X H. 2018. A preliminary analysis of the effects of bisphenol A on the plant root growth *via* changes in endogenous plant hormones. *Ecotoxicology and Environmental Safety*, **150**, 152–158.

Liu Q Q, Luo L, Zheng L Q. 2018. Lignins: Biosynthesis and biological functions in plants. *International Journal of Molecular Sciences*, **19**, 335.

Love M I, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 550.

Matejczyk M, Swislocka R, Kalinowska M, Swidersk G, Lewandowsk W, Jablonska-Trypuo A. 2017. Monitoring of synergistic enhancement of caffeic acid on *Escherichia coli* K-12 reca::GFP strain treated with dacarbazine. *Acta Poloniae Pharmaceutica*, **74**, 809–816.

Munir N, Cheng C Z, Xia C S, Xu X M, Nawaz M A, Iftikhar J, Chen Y K, Lin Y L, Lai Z X. 2019. RNA-seq analysis reveals an essential role of tyrosine metabolism pathway in response to root-rot infection in Gerbera hybrida. *PLoS ONE*, **14**, e0223519.

Munne-Bosch S. 2005. Linking tocopherols with cellular signaling in plants. *The New Phytologist*, **166**, 363–366.

Munne-Bosch S, Weiler E W, Alegre L, Muller M, Duchting P, Falk J. 2007. Alpha-tocopherol may influence cellular signaling by modulating jasmonic acid levels in plants. *Planta*, **225**, 681–691.

Neis V B, Rosa P B, Olescowicz G, Rodrigues A L S. 2017. Therapeutic potential of agmatine for CNS disorders. *Neurochemistry International*, **108**, 318–331.

Ning X H, Ren X B, Xie X F, Yan P, Wang D H, Huang X S. 2020. A caffeic acid phenethyl ester analog inhibits the proliferation of nasopharyngeal carcinoma cells via targeting epidermal growth factor receptor. *Journal of Biochemical and Molecular Toxicology*, **34**, e22491.

Schenck C A, Maeda H A. 2018. Tyrosine biosynthesis, metabolism, and catabolism in plants. *Phytochemistry*, **149**, 82–102.

Smirnoff N, Wheeler G L. 2000. Ascorbic acid in plants: Biosynthesis and function. *Critical Reviews in Biochemistry and Molecular Biology*, **35**, 291–314.

Soll J, Schultz G, Joyard J, Douce R, Block M A. 1985. Localization and synthesis of prenylquinones in isolated outer and inner envelope membranes from spinach chloroplasts. *Archives of Biochemistry and Biophysics*, **238**, 290–299.

Song X M, Liu G F, Duan W K, Liu T K, Huang Z N, Ren J, Li Y, Hou X L. 2014. Genome-wide identification, classification and expression analysis of the heat shock transcription factor family in Chinese cabbage. *Molecular Genetics and Genomics*, **289**, 541–551.

Song X M, Liu G F, Huang Z N, Duan W K, Tan H W, Li Y, Hou X L. 2016. Temperature expression patterns of genes and their coexpression with LncRNAs revealed by RNA-Seq in non-heading Chinese cabbage. *BMC Genomics*, **17**, 297.

Song X M, Nie F L, Chen W, Ma X, Gong K, Yang Q H, Wang J P, Li N, Sun P C, Pei Q Y, Yu T, Hu J J, Li X Y, Wu T, Feng S Y, Li X Q, Wang X Y. 2020. Coriander Genomics Database: A genomic, transcriptomic, and metabolic database for Coriander. *Horticulture Research*, **7**, 55.

Song X M, Wang J P, Li N, Yu J G, Meng F B, Wei C D, Liu C, Chen W, Nie F L, Zhang Z K, Gong K, Li X Y, Hu J J, Yang Q H, Li Y X, Li C J, Feng S Y, Guo H, Yuan J Q, Pei Q Y, *et al*. 2019. Deciphering the high-quality genome sequence of Coriander that causes controversial feelings. *Plant Biotechnology Journal*, **18**, 1444–1456.

Stickle D, Bohrer A, Berger R, Morrissey J, Klahr S, Turk J. 1996. Quantitation of the putative neurotransmitter agmatine as the hexafluoroacetylacetonate derivative by stable Isotope dilution gas chromatography and negative-ion chemical ionization mass spectrometry. *Analytical Biochemistry*, **238**, 129–136.

Tanida I, Shirasago Y, Suzuki R, Abe R, Wakita T, Hanada K, Fukasawa M. 2015. Inhibitory effects of caffeic acid, a coffee-related organic acid, on the propagation of hepatitis C virus. *Japanese Journal of Infectious Diseases*, **68**, 268–275.

Trapnell C, Williams B A, Pertea G, Mortazavi A, Kwan G, van Baren M J, Salzberg S L, Wold B J, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, **28**, 511–515.

Tulsani N J, Hamid R, Jacob F, Umretiya N G, Nandha A K, Tomar R S, Golakiya B A. 2020. Transcriptome landscaping for gene mining and SSR marker development in Coriander (*Coriandrum sativum* L.). *Genomics*, **112**, 1545–1553.

Wang M, Toda K, Block A, Maeda H A. 2019. TAT1 and TAT2 tyrosine aminotransferases have both distinct and shared functions in tyrosine metabolism and degradation in *Arabidopsis thaliana*. *The Journal of Biological Chemistry*, **294**, 3563–3576.

Wiśniewska A, Olszanecki R, Totoń-Żurańska J, Kuś K, Stachowicz A, Suski M, Gębska A, Gajda M, Jawień J, Korbut R. 2017. Anti-atherosclerotic action of agmatine in ApoE-knockout mice. *International Journal of Molecular Sciences*, **18**, 1706.

Yella S S T, Kumar R N, Ayyanna C, Varghese A M, Amaravathi P, Vangoori Y. 2019. The combined effect of *Trigonella foenum* seeds and *Coriandrum sativum* leaf extracts in alloxan-induced diabetes mellitus wistar albino rats. *Bioinformation*, **15**, 716–722.

Young M D, Wakefield M J, Smyth G K, Oshlack A. 2010. Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biology*, **11**, R14.

Zhang N, Chen Y, Zhao Y, Fan D, Li L, Yan B, Tao G, Zhao J, Zhang H, Wang M. 2020. Caffeic acid assists microwave heating to inhibit the formation of mutagenic and carcinogenic PhIP. *Food Chemistry*, **317**, 126447.

Zhang Y B, Tang W, Wang L H, Hu Y W, Liu X W, Liu Y S. 2019. Kiwifruit (*Actinidia chinensis*) R1R2R3-MYB transcription factor AcMYB3R enhances drought and salinity tolerance in *Arabidopsis thaliana*. *Journal of Integrative Agriculture*, **18**, 417–427.

Zhao T T, Wang Z Y, Bao Y F, Zhang X C, Yang H H, Zhang D Y, Jiang J B, Zhang H, Li J F, Chen Q S, Xu X Y. 2019. Downregulation of the *SL-ZH13* transcription factor gene expression decreases drought tolerance of tomato. *Journal of Integrative Agriculture*, **18**, 1579–1586.

Executive Editor-in-Chief HUANG San-wen
Managing Editor WENG Ling-yun