BioTechnologia vol. 100 (3) • pp. 227–240 • 2019
Journal of Biotechnology, Computational Biology and Bionanotechnology RESEARCH PAPERS
http://doi.org/10.5114/bta.2019.87582

Identification of candidate genes related to aroma in rice by analyzing the microarray data of highly aromatic and nonaromatic recombinant inbred line bulks

ZAHRA ZINATI *, AZAR DELAVARI

Department of Agroecology, College of Agriculture and Natural Resources of Darab, Shiraz University, Shiraz, Iran

Abstract

Aroma is an important factor affecting the rice quality. However, the breeding strategies for improving aroma, the molecular basis of the aroma formation, and aroma-related genes are yet to be investigated. In this study, microarray analysis was performed followed by gene ontology (GO) enrichment analysis and protein–protein interaction (PPI) network analysis for discovering the differentially expressed genes in the bulks of an aromatic and a nonaromatic recombinant inbred line (RIL) derived from a cross between a basmati rice variety, Pusa 1121, and a nonaromatic rice variety, Pusa 1342. GO enrichment analysis categorized a total of 2577 differentially expressed genes into 24 functional groups. The metabolic process was the most highly overrepresented category enriched by the differentially expressed genes. One of the most paramount minor subcategories in the metabolic process category was "cellular aromatic compound metabolic process (GO: 0006725)" with 21 genes that may serve as novel candidates for genes involved in aroma formation. According to the network analysis, six downregulated genes, as well as two upregulated genes, were identified as critical in the PPI network. These results will provide a perspective for unveiling the key components of the mechanisms behind the aroma formation in rice and give the required information on client-oriented breeding.

Key words: rice, protein-protein interaction network, aroma-related genes, gene ontology enrichment analysis

Introduction

Aromatic rice (*Oryza sativa*) is gaining immense popularity among consumers throughout the world (Bhattacharjee et al., 2002); hence, its market price is considerably higher than that of the nonaromatic variant (Qiu and Zhang, 2003). Since the rice aroma, which is a polygenic quantitative trait characterized by an intricate inheritance pattern, is significantly influenced by the environment, analyzing the contributing genes is quite laborious (Pachauri et al., 2010). The availability of whole-genome microarray chips for diverse organisms has paved the way for deciphering the flow of biological information underlying the complex traits by a simultaneous analysis of expression of almost all the genes (Kathiresan et al., 2006). In addition, microarrays have been widely used for detecting the genes that are accountable for a specific phenotype (Rapaport et al., 2007).

The discovery of candidates for genes responsible for the aroma of rice grains has been gaining interest since the studies reporting the identification of a single dominant gene (3 : 1 segregation) (Kadam and Patankar, 1938), a single recessive gene (*fgr*) (Jodon, 1944), and three Quantitative Trait Loci, namely *aro3.1*, *aro4.1*, and *aro8.1*, (Amarawathi et al., 2008) were published. However, the molecular basis of the aroma formation, as well as the aroma-related genes, is poorly understood. Over the last few years, microarray-based gene expression profiling, also called transcriptome profiling, has been considered as a practical approach for gene identification (Pachauri et al., 2014). Integrating the gene expression analysis, gene ontology (GO) analysis, and network analysis provides exciting opportunities to investigate the presumptive molecular processes underlying the aroma formation and to identify novel target genes for improving aroma.

* Corresponding author: Department of Agroecology, College of Agriculture and Natural Resources of Darab, Shiraz University, Shiraz, Iran; e-mail: zahrazinati@shirazu.ac.ir

Thus, the aim of this study was to investigate the molecular mechanisms involved in the aroma formation and to identify the aroma-related genes for improving the aroma of rice. To achieve this, the microarray data of the bulks of highly aromatic and nonaromatic recombinant inbred lines (RILs) derived from a cross between Pusa 1121, a basmati rice variety, and Pusa 1342, a nonaromatic rice variety (Pachauri et al., 2014), were employed. The use of transcriptome profiles of the bulk of inbred lines has led to the understanding of the background along with a number of individual samples to be assayed (Pachauri et al., 2010). Accordingly, the microarray data of the bulks of aromatic and nonaromatic RILs may serve as a suitable material for the analysis of the transcriptome.

In this study, different genes that are significantly up- and downregulated were screened, and a variety of bioinformatic tools including GO enrichment and protein–protein interaction (PPI) network were utilized to shed light on the biological mechanisms behind the aroma formation and identify the aroma-related genes. Thus, the study unveils a number of genes involved in the aroma formation in rice, which are proposed as promising candidates for breeding rice cultivars and enhancing the rice aroma.

Materials and methods

Microarray analysis

To unravel the mechanisms associated with the formation of aroma in aromatic rice, the microarray data of the bulks of highly aromatic and nonaromatic RILs derived from a cross between a basmati rice variety, Pusa 1121, and a nonaromatic rice variety, Pusa 1342 (Pachauri et al., 2014) (available online at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE47812) were used.

The differential expression analysis was accomplished using Genevestigator, a database containing data on thousands of experiments on public microarrays and RNAseq. All the data are manually curated and well described in the database. Genevestigator allows visualizing the gene expression profiles in various biological contexts such as nutrients, genotypes, diseases, chemicals, tissues, or biotic and abiotic stress conditions (https://genevestigator.com/gv/doc/intro_plant.jsp). The *Diff-Expression* tool embedded in the database enables easy and quick finding of the genes that are significantly differentially expressed between the two conditions. In the Data Selection dialogue box, "GSE47812: Gene expression data for rice grain (*Oryza sativa* L.) aroma was performed using bulk pools of aromatic and nonaromatic RILs derived from a cross between Pusa 1121 and Pusa 1342" was searched in the experiments part and the data were selected for analysis in the study. In the Define Comparison dialogue box, the samples for groups X and Y were defined as "aromatic RILs/non-aromatic RILs."

For the GO enrichment analysis, construction of the PPI network, and identification of the candidate genes in the aroma PPI network, the genes with a log2-fold change greater than 0.6 and a false discovery rate (FDR) lesser than 0.01 as upregulated candidates and those with a log2-fold change lesser than 0.6 and an FDR lesser than 0.01 as downregulated candidates were screened.

Gene ontology enrichment analysis of the differentially expressed genes

The GO enrichment analysis was carried out to extract meaningful information from the differentially expressed genes (Du et al., 2010) (available from http://bioinfo.cau.edu.cn/agriGO/index.php) under the three main categories namely biological process, molecular function, and cellular component using the AgriGO web-based tool. The Singular Enrichment Analysis (SEA) was done setting "Rice Gramene Locus" as a reference, "Yekutieli (FDR under dependency)" as a multitest adjustment method, "Fisher" as a statistical test method, 0.01 as a *P*-value cutoff, and "10" as a minimum number of mapping entries. In the SEA, the GO term enrichment was calculated one by one in a linear model by comparing a given set of genes to a standard or customized annotated reference list. To detect the possible biological processes and molecular functions that are enriched by the identified genes, an enriched *P*-value was computed. This would allow identifying from the whole set of differentially expressed genes those that match a given biological category as compared to pure random chance (Du et al., 2010).

Construction of the protein–protein interaction network

To construct a PPI network among the genes that are significantly differentially expressed, the Predicted Rice Interactome Network (PRIN; http://bis.zju.edu.cn/prin/) was used. PRIN is the first database generated for the rice PPI network, and was built on the basis of the PPI

data of six model organisms including fruit fly (*Drosophila melanogaster*), worm (*Caenorhabditis elegans*), *Escherichia coli* K12, human (*Homo sapiens*), yeast (*Saccharomyces cerevisiae*), and *Arabidopsis thaliana* (Gu et al., 2011). PRIN contains 76,585 nonredundant protein interaction pairs among 5049 rice proteins (Gu et al., 2011). This database is often updated, and thus an extremely useful, up-to-date tool for assessing the PPI data and performing a prospective study on the functional biology and systems biology of rice (Gu et al., 2011; Fukao, 2012). To avoid bias toward the upregulated genes in the network, both upregulated and downregulated genes were utilized as query genes in PRIN to construct the PPI network. A tree layout was used for visualizing the network. This layout arranges graphs as a tree with the nodes organized hierarchically so that the display is clear and easy to understand (Agapito et al., 2013).

Identification of the candidate genes in the aroma PPI network

To identify the candidate genes in the aroma PPI network, all relations in the network were transferred to the Cytoscape software, version 2.5.0 (Shannon et al., 2003). Then, Cyto-Hubba, a hub finder/analyzer software, was utilized to find the critical proteins/hubs using several topological algorithms, including Betweenness Centrality, Bottleneck, Closeness Centrality, Clustering Coefficient, Degree, Density of Maximum Neighborhood Component, EcCentricity, Edge Percolated Component, Maximal Clique Centrality, Maximum Neighborhood Component (MNC), Radiality Centrality, and Stress Centrality (Chin et al., 2014). The top 10 nodes ranked, or given a default value, by each mentioned algorithm, were identified. Finally, on the basis of the overall results obtained by applying different algorithms, some genes were determined as critical.

Results

Microarray analysis

A comparison was made between the bulks of aromatic and nonaromatic RILs derived from a cross between Pusa 1121 and Pusa 1342. It was found that a total of 3170 genes in six highly aromatic RIL bulks were differentially expressed compared to six nonaromatic RIL bulks. Of these genes, 1154 and 1616 were found to be significantly up- and downregulated, respectively.

Gene ontology enrichment analysis of the differentially expressed genes

Out of the 3170 genes, 2577 were annotated with GO terms. The GO enrichment analysis categorized these 2577 differentially expressed genes into 24 functional groups. The number of major subcategories was as follows: 14 in "biological process" (Fig. 1), seven in "molecular function" (Fig. 2), and three in "cellular component" (Fig. 3). According to the enrichment analysis of biological process, the metabolic process was the most highly overrepresented category with a large number of differentially expressed genes (1063 genes). A further investigation into the metabolic process revealed the presence of a terpenoid metabolic process (GO:0006721) (13 genes), a cellular carbohydrate metabolic process (GO : 0044262)(46 genes), a cellular aromatic compound metabolic process (GO : 0006725) (21 genes), a cellular lipid metabolic process (GO : 0044255) (46 genes), a cellular ketone metabolic process (GO : 0042180) (83 genes), and a cellular amino acid and a derivative metabolic process (GO : 0006519) (60 genes) – Figure 4.

Construction of the PPI network

The PPI network contained 174 nodes, 491 edges, and 43 connected components. The average number of neighbors was 4.414. In the tree layout of the network, LOC_Os03g17700.1 (Os03g0285800) (MAP kinase) constituted the root of the tree (Fig. 5).

Identification of the candidate genes in the aroma PPI network

On the basis of the overall intersection of algorithms (Table 1), LOC_Os07g30980.1 (Os07g0492100) (uvrD/REP helicase family protein, putative, expressed), LOC_Os05g38530.1 (Os05g0460000) (DnaK family protein, putative, expressed), LOC_Os10g38278.1 (chloroplast 30S ribosomal protein S2, putative), LOC_Os02g24612.1 (ribosomal protein S2, putative), LOC_Os04g48410.1 (Os04g0573200) (copper chaperone for superoxide dismutase, putative, expressed), LOC_Os05g49250.1 (Os05g0567500) (hhH-GPD superfamily base excision DNA repair protein, putative, expressed), and LOC_Os10g21234.1 (chloroplast 30S ribosomal protein S2, putative) were determined as ideal candidate genes for aroma enhancement in rice.

As mentioned above, in the tree layout of the network, LOC_Os03g17700.1 (Os03g0285800) (MAP ki-

Fig. 1. Gene ontology enrichment analysis of significantly differentially expressed genes in aromatic RIL bulks compared to non-aromatic RIL bulks. The diagram shows GO terms of the significantly enriched biological processes determined by SEA analysis. To highlight the significance, the squares are color coded. In terms of importance degree and ranking, yellow and red are calibrated as 1 (least significance level) and 9 (most significant) respectively. GO categories are determined by squares proceeded by bracketed significance level, GO category name and gene ratio in both the list and on the microarray associated to that classification

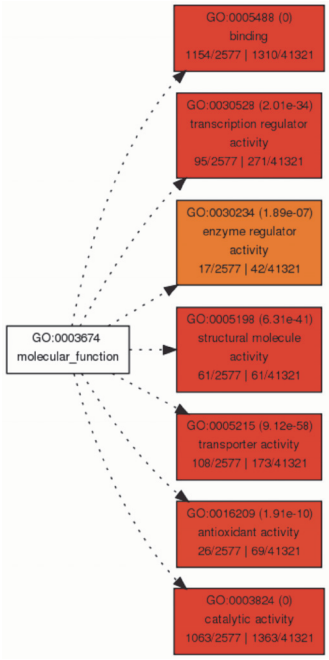

Fig. 2. Gene ontology enrichment analysis of significantly differentially expressed genes in aromatic RIL bulks compared to non-aromatic RIL bulks. The diagram shows GO terms of the significantly enriched molecular functions determined by SEA analysis. To highlight the significance, the squares are color coded. In terms of importance degree and ranking, yellow and red are calibrated as 1 (least significance level) and 9 (most significant) respectively. GO categories are determined by squares proceeded by bracketed significance level, GO category name and gene ratio in both the list and on the microarray associated to that classification

nase) constituted the root. This gene was also identified when the Betweenness and Stress Centrality algorithms were applied. It should be pointed out that almost all the critical genes, except LOC_Os07g30980.1 and LOC_Os05g49250.1, were downregulated in an aromatic rice variety.

Discussion

As an economic trait influencing the rice quality, aroma has received special attention in recent years, and hence, enhancing aroma is of high priority in rice breeding programs. However, this is a serious challenge and identifying the related genes and biological processes is certainly important.

Microarray analysis

A comparative expression analysis of the bulks of aromatic and nonaromatic RILs revealed a total of 3170

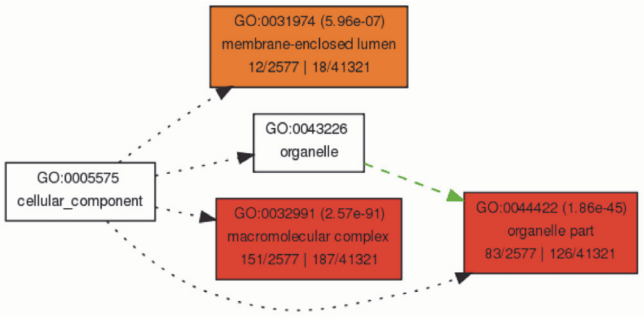

Fig. 3. Gene ontology enrichment analysis of significantly differentially expressed genes in aromatic RIL bulks compared to non-aromatic RIL bulks. The diagram shows GO terms of the significantly enriched cellular components determined by SEA analysis. To highlight the significance, the squares are color coded. In terms of importance degree and ranking, yellow and red are calibrated as 1 (least significance level) and 9 (most significant) respectively. GO categories are determined by squares proceeded by bracketed significance level, GO category name and gene ratio in both the list and on the microarray associated to that classification

genes that were differentially expressed between the tested rice cultivars. This finding indicates that aroma is a very complicated trait in rice and is developed as a result of an organized expression of a large number of genes.

Gene ontology enrichment analysis of the differentially expressed genes

A popular approach to deciphering the microarray data is performing a gene enrichment analysis on the basis of functional annotation of the differentially expressed genes. This is an appropriate option to determine whether the differentially expressed genes are attributed to a definite biological process or a molecular function.

According to the enrichment analysis of biological process, 14 categories were enriched by differentially expressed genes (Fig. 1). Consistent with these results of GO enrichment analysis, all biological categories except the biological process "death (GO: 0016265)" have been previously reported in a study conducted by Wei et al. (2016) on aroma metabolites and emission of pear. Our results indicated that the biological process "death" might be related to aroma formation in rice in addition to other reported biological processes. According to the enrichment analysis of biological process, the metabolic process was the most highly overrepresented category with a large number of differentially expressed genes.

One of the minor subcategories in the metabolic process category was a "terpenoid metabolic process (GO:0006721)" with 13 genes. According to El Hadi et al. (2013), terpenoids constitute the largest class of secondary plant metabolites with many representative volatile compounds. Besides, other significant minor subcategories in the metabolic process category were the cellular carbohydrate metabolic process (GO:0044262), cellular lipid metabolic process (GO:0044255), cellular amino acid and derivative metabolic process (GO:0006519), and cellular ketone metabolic process (GO:0042180). Consistent with the results of the enrichment analysis of biological process, the most important aroma compounds were found to be lipid-derived compounds, amino acid-derived compounds, phenolic derivatives, and mono- and sesquiterpenes (Schwab et al., 2008). Volatiles that are important for aroma and flavor have been mentioned to be derived from membrane lipids, amino acids, and carbohydrates in a study by Sanz et al. (1997).

One of the most paramount minor subcategories in the metabolic process category was the "cellular aromatic compound metabolic process (GO:0006725)" with 21 genes. Distinguishing aroma genes to generate new improved varieties of aromatic rice, as well as expediting studies on the cellular and molecular mechanisms underlying aroma formation, is most intriguing for rice breeders. The 21 genes enriching the "cellular aromatic compound metabolic process (GO:0006725)" subcategory are listed in Table 2. In this subcategory, 11 and 10 genes were up- and downregulated, respectively. Among the downregulated genes, Os02g0626100, Os02g 0627100, Os02g0626400 (phenylalanine ammonia-lyase, PAL), Os08g0448000 (probable 4-coumarate-CoA ligase 2, 4CL2), and Os02g0177600 (probable 4-coumarate-CoA ligase 1, 4CL1) are involved in the phenylpropanoids biosynthesis pathway, along with three isoforms of laccase, namely Os11g0641500 (Laccase-19), Os10g0346300 (Laccase-15), and Os12g0258700 (Laccase-24). PAL is the entry-point enzyme in this pathway that converts L-phenylalanine (L-Phe) to trans-cinnamic acid, and can be inhibited by its product, trans-cinnamic acid (Zhang and Liu, 2015). 4CL catalyzes the synthesis of p-coumaroyl-CoA from p-coumarate, and 4CL1-2 participates in the lignin biosynthesis (Ehlting et al., 1999; Lindermayr et al., 2003). Laccase also takes part in the lignin synthe-sis by producing monolignol radicals that can bind together

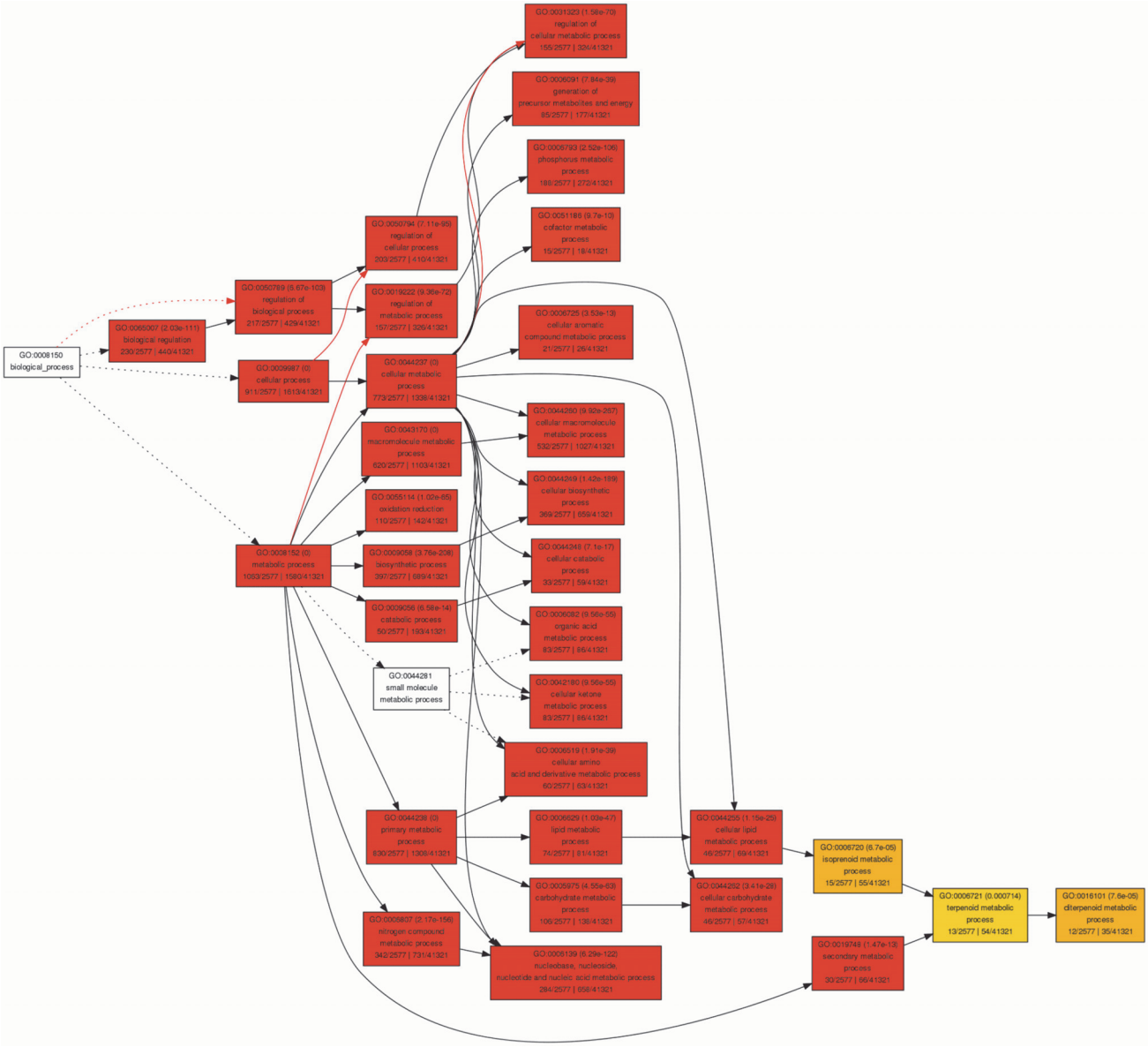

Fig. 4. The miner subcategories in metabolic process category. The diagram shows GO terms of the significantly enriched biological processes determined by SEA analysis. To highlight the significance, the squares are color coded. In terms of importance degree and ranking, yellow and red are calibrated as 1 (least significance level) and 9 (most significant) respectively. GO categories are determined by squares proceeded by bracketed significance level, GO category name and gene ratio in both the list and on the microarray associated to that classification

randomly and polymerize to form lignin (Cai et al., 2006; Castrillón-Arbeláez and Délano Frier, 2016).

Phenylpropanoids biosynthesis pathway produces various aromatic compounds that have been classified into different subclasses such as flavonoids, lignins, coumarins, tannins, and quinones (Fraser and Chapple, 2011). These compounds play essential roles in the growth, development, and protection against UV irradiation in plants (Dixon and Paiva, 1995; Vanholme et al., 2010). In this study, we have demonstrated that the enzymes that participate in the phenylpropanoids biosynthesis are downregulated in highly aromatic RIL bulks, and consequently, their protein products are reduced in number. Downregulation of these enzymes indicates that the aroma of Pusa 1121 cultivar is not produced through the phenylpropanoids biosynthesis pathway. However, further investigation is required to understand the reason behind the reduction in the levels of the related enzymes in this pathway.

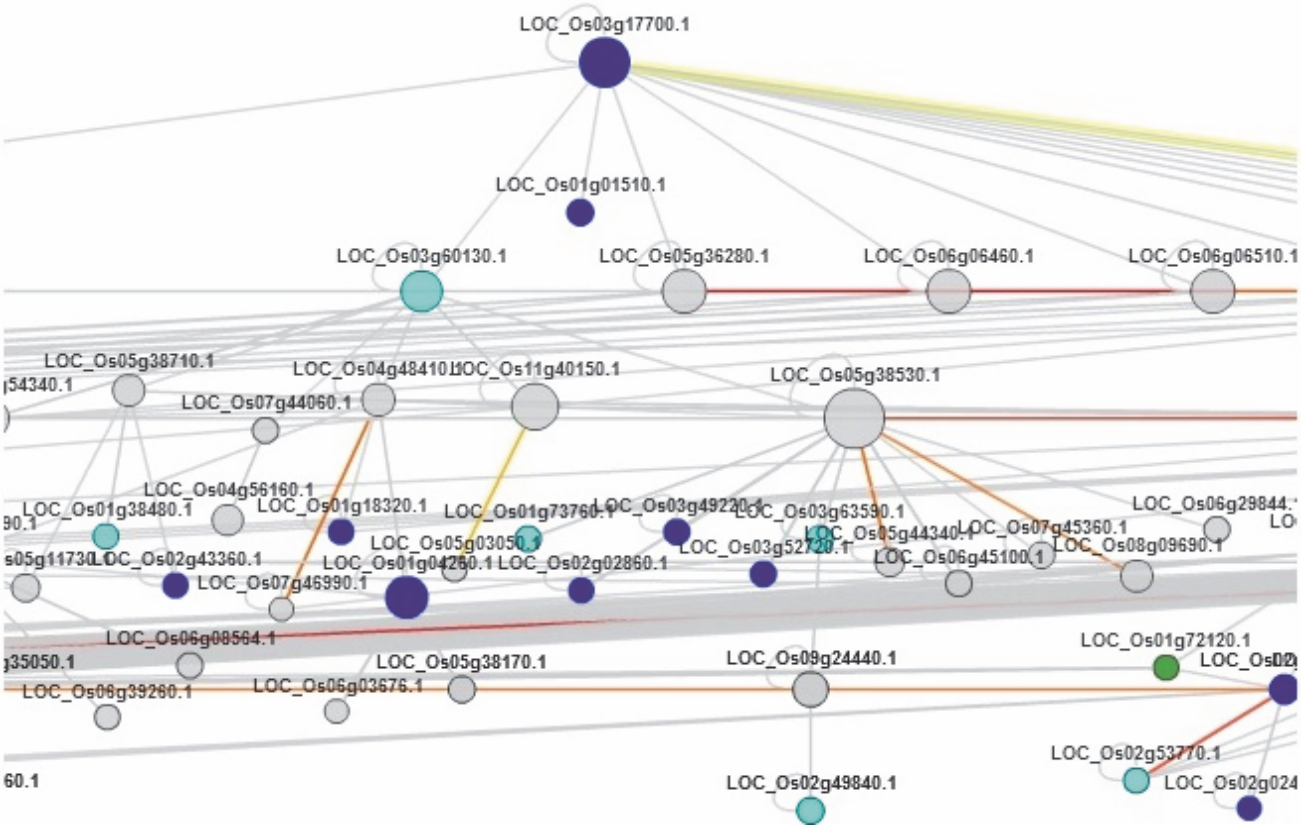

Fig. 5. Part of protein-protein interaction (PPI) network of significantly differentially expressed genes in aromatic RIL bulks compared to non-aromatic RIL bulks. The web based tool PRIN (a Predicted Rice Interactome Network; http://bis.zju.edu.cn/prin/) was used to predict the interactions

Os10g0523700 is one of the upregulated genes identified in this study. It encodes arogenate dehydratase (ADT)/prephenate dehydratase 6 (PDT) and plays a role in Phe biosynthesis. The ADT/PDT activity in rice is subjected to a feedback regulation and inhibited by Phe (Wakasa and Ishihara, 2009). In plant species, Phe catabolism occurs through three pathways: production of phenylpropanoids (Knaggs, 2003), production of 2-phenylethanol (Jezussek et al., 2002; Tieman et al., 2006), and production of phenyl-glucosinolates (Mikkelsen et al., 2004). Since the enzymes involved in phenylpropanoids production, namely PAL, 4CL, and LAC, were found to be downregulated in this study, it seems that Phe catabolism occurs via the other two pathways (i.e., production of 2-phenylethanol and/or production of phenyl-glucosinolates).

Another gene found upregulated in highly aromatic RIL bulks was Os03g0797400, named indole-3-glycerol phosphate lyase. It has been proven that this gene plays a role in the biosynthesis of volatile indole in maize (Frey et al., 2000). Hinge et al. (2016) indicated that indole and 2-acetyl-1-pyrrole were two major compounds that served to distinguish the scented from the nonscented rice cultivars. In our study, Os03g0111100, which encodes a folylpolyglutamate synthase (FPGS), was found upregulated in a scented rice cultivar. It has been demonstrated that this enzyme affects the lignin biosynthesis in maize (Li et al., 2015b). Two isoforms of FPGS in *O. sativa* were shown to be involved in the development of seeds (Anukul et al., 2010). Os10g0483500 (cytokinin dehydrogenase 3) and Os05g0374200 (cytokinin dehydrogenase 9) were both found upregulated in highly aromatic RIL bulks. Deborah et al. (2017) pointed out that Os05g0374200 (similar to cytokinin dehydrogenase 1 precursor) was involved in the accumulation of cytokinins and influenced the grain size. The expression of Os02g0613900 (adenine phosphoribosyltransferase 4, APRT) gene was also upregulated in highly aromatic RIL bulks as shown by our analysis. APRT is involved in the caffeine biosynthesis pathway in tea leaves (Li et al.,

Table 1. The list of critical genes identified based on overall results of applying different algorithms embedded in Cyto-Hubba including Maximal Clique Centrality (MCC), Density of Maximum Neighborhood Component (DMNC), Maximum Neighborhood Component (MNC), Degree, Edge Percolated Component (EPC), Bottleneck (BN), EcCentricity, Closeness centrality, Radiality centrality, Betweenness centrality, Stress centrality, and Clustering coefficient

	Name	MCC	MNC	Degree	EPC	BottleNeck	EC centricity	Closeness	Radiality	Betwenness	Stress	Clustering coefficient	The number of algorithms that identified critical genes
LOC_Os07g30980.1	12	0.23452	19	20	51.512	71	0.12165	58.93333	5.46835	4220.9039	46426	0.18421	7
LOC_Os05g38530.1	11	0.14572	17	18	30.154	22	0.14598	51.11667	5.07444	2992.17201	13636	0.11765	6
LOC_Os10g38278.1	0	0.39566	21	22	55.197	1	0.12165	58.46667	5.36408	802.86462	11458	0.30303	6
LOC_Os02g24612.1	0	0.39566	21	22	55.473	26	0.12165	58.46667	5.36408	802.86462	11458	0.30303	6
LOC_Os04g48410.1	3	0.3733	11	12	35.75	26	0.14598	52.68333	5.3467	2468.76287	20098	0.33333	5
LOC_Os05g49250.1	1	0.32048	15	16	50.074	24	0.12165	54.08333	5.2656	1809.39382	25026	0.26667	5
LOC_Os10g21234.1	0	0.39566	21	22	55.093	1	0.12165	58.46667	5.36408	802.86462	11458	0.30303	5

Table 2. The list of the genes enriched "cellular aromatic compound metabolic process (GO : 0006725)"

MSU-ID	RAP-ID	Aromatic RIL bulks compared to non-aromatic RIL bulks	GO term	Description
LOC_Os08g34790.1	Os08g0448000	down regulated	GO : 0016207 4-coumarate-CoA ligase activity GO : 0009698 phenylpropanoid metabolic process GO : 0005524 ATP binding	probable 4-coumarate–CoA ligase 2
LOC_Os05g48360.1	Os05g0557400	down regulated	GO : 0009651 response to salt stress GO : 0052542 callose deposition during defense response GO : 0010337 regulation of salicylic acid metabolic process GO : 0009626 plant-type hypersensitive response	–
LOC_Os11g42200.1	Os11g0641500	down regulated	GO : 0046274 lignin catabolic process GO : 0005615 extracellular space GO : 0016023 cytoplasmic membrane-bounded vesicle GO : 0048046 apoplast GO : 0055114 oxidation reduction GO : 0005507 copper ion binding GO : 0008471 laccase activity	laccase-19
LOC_Os02g41630.2	Os02g0626100	down regulated	GO : 0005737 cytoplasm GO : 0006559 L-phenylalanine catabolic process GO : 0016211 ammonia ligase activity GO : 0009058 biosynthetic process GO : 0045548 phenylalanine ammonia-lyase activity GO : 0009698 phenylpropanoid metabolic process	phenylalanine ammonialyase
LOC_Os03g02030.1	Os03g0111100	up regulated	GO : 0009396 folic acid and derivative biosynthetic process GO : 0004326 tetrahydrofolylpolyglutamate synthase activity GO : 0005739 mitochondrion GO : 0005524 ATP binding	Os03g0111100 protein
LOC_Os09g36030.1	Os09g0529900	up regulated	GO : 0016830 carbon-carbon lyase activity GO : 0009536 plastid GO : 0006725 cellular aromatic compound metabolic process	Os09g0529900 protein
LOC_Os05g31040.1	Os05g0374200	up regulated	GO : 0016023 cytoplasmic membrane-bounded vesicle GO : 0050660 FAD binding GO : 0009690 cytokinin metabolic process GO : 0019139 cytokinin dehydrogenase activity GO : 0009055 electron carrier activity	putative uncharacterized protein
LOC_Os02g41680.1	Os02g0627100	down regulated	GO : 0005737 cytoplasm GO : 0006559 L-phenylalanine catabolic process GO : 0016211 ammonia ligase activity GO : 0009058 biosynthetic process GO : 0045548 phenylalanine ammonia-lyase activity GO : 0009698 phenylpropanoid metabolic process	phenylalanine ammonia-lyase

LOC_Os10g20610.1	Os10g0346300	down regulated	GO : 0046274 lignin catabolic process GO : 0005615 extracellular space GO : 0016023 cytoplasmic membrane-bounded vesicle GO : 0048046 apoplast GO : 0055114 oxidation reduction GO : 0005507 copper ion binding GO : 0008471 laccase activity	laccase-15
LOC_Os03g47610.1	Os03g0679700	up regulated	GO : 0009536 plastid GO : 0009228 thiamin biosynthetic process	Os03g0679700 protein
LOC_Os06g35050.1	Os06g0542200	down regulated	GO : 0006571 tyrosine biosynthetic process GO : 0004665 prephenate dehydrogenase (NADP+) activity GO : 0005488 binding GO : 0005739 mitochondrion	Os06g0542200 protein
LOC_Os10g37980.1	Os10g0523700	up regulated	GO : 0009094 L-phenylalanine biosynthetic process GO : 0009536 plastid GO : 0004664 prephenate dehydratase activity	Os04g0406600 protein
LOC_Os01g47420.1	Os01g0663500	up regulated	GO : 0006729 tetrahydrobiopterin biosynthetic process GO : 0008124 4-alpha-hydroxytetrahydrobiopterin dehydratase activity GO : 0005739 mitochondrion	putative uncharacterized protein
LOC_Os05g38170.1	Os05g0455700	up regulated	GO : 0006223 uracil salvage GO : 0009536 plastid GO : 0004845 uracil phosphoribosyltransferase activity GO : 0009116 nucleoside metabolic process	Os05g0455700 protein
LOC_Os10g34230.1	Os10g0483500	up regulated	GO : 0022900 electron transport chain GO : 0050660 FAD binding GO : 0009690 cytokinin metabolic process GO : 0019139 cytokinin dehydrogenase activity	Os10g0483500 protein
LOC_Os03g58300.1	Os03g0797400	up regulated	GO : 0009536 plastid GO : 0006568 tryptophan metabolic process GO : 0004834 tryptophan synthase activity	Os03g0797400 protein
LOC_Os02g53810.1	Os02g0778500	up regulated	GO : 0045430 chalcone isomerase activity GO : 0009813 flavonoid biosynthetic process GO : 0005739 mitochondrion	Os07g0571600 protein
LOC_Os02g41650.1	Os02g0626400	down regulated	GO : 0005737 cytoplasm GO : 0006559 L-phenylalanine catabolic process GO : 0016211 ammonia ligase activity GO : 0009058 biosynthetic process GO : 0045548 phenylalanine ammonia-lyase activity GO : 0009698 phenylpropanoid metabolic process	phenylalanine ammonia-lyase

LOC_Os12g15680.1	Os12g0258700	down regulated	GO : 0046274 lignin catabolic process GO : 0005615 extracellular space GO : 0016023 cytoplasmic membrane-bounded vesicle GO : 0048046 apoplast GO : 0055114 oxidation reduction GO : 0005507 copper ion binding GO : 0008471 laccase activity	laccase-24
LOC_Os02g08100.1	Os02g0177600	down regulated	GO : 0016207 4-coumarate-CoA ligase activity GO : 0009698 phenylpropanoid metabolic process GO : 0005524 ATP binding	probable 4-coumarate–CoA ligase 1
LOC_Os02g40010.1	Os02g0613900	up regulated	GO : 0009536 plastid GO : 0003999 adenine phosphoribosyltransferase activity GO : 0009116 nucleoside metabolic process GO : 0006168 adenine salvage	Os12g0589100 protein

2015a). Besides, APRT converts N^6-(Δ^2-isopentenyl) adenine (iP) to t-zeatin that is catabolized by cytokinin dehydrogenase (Sakakibara, 2006). Therefore, upregulation of APRT may cause upregulation of the cytokinin dehydrogenase, the role of which was explained in a previous study. The present study revealed the genes that were upregulated and downregulated in the scented rice cultivar, but experimental studies are essential to demonstrate the contribution of these genes to the mechanism of aroma formation in rice cultivars.

According to the enrichment analysis of molecular function, seven categories were enriched by differentially expressed genes. Consistent with these results of GO enrichment analysis, six out of seven molecular function categories including binding (GO : 0005488), enzyme regulator activity (GO : 0030234), structural molecule activity (GO : 0005198), transporter activity (GO : 0005215), antioxidant activity (GO : 0016209), and catalytic activity (GO : 0003824) have been previously reported in a study conducted by Wei et al. (2016) on aroma metabolites and emission of pear. The molecular function "transcription regulator activity (GO : 0030528)" has not been previously reported as associated with aroma. Therefore, the genes in this molecular function category might be related to the aroma formation in rice.

The enrichment analysis of cellular component revealed that three categories were enriched by differentially expressed genes. In accordance with these results of GO enrichment analysis, all cellular component categories including membrane-enclosed lumen (GO : 0031974), macromolecular complex (GO : 0032991), and organelle part (GO : 00044422) have been previously reported in a study by Wei et al. (2016).

Construction of the PPI network

The complex changes in gene expression can be addressed by organizing the genes in networks (Pedicini et al., 2010). The PPI study helps in uncovering the unknown molecular functions of proteins and provides insight into the sophisticated cellular networks and molecular mechanisms of diverse physiological events in plants (Fukao, 2012). However, there is no report on the PPI networks that underlie the aroma biosynthesis in rice. According to the PPI network created in this study, LOC_Os03g17700.1 (Os03g0285800) (MAP kinase), the network root, comes into interaction with regulatory genes such as protein kinases, namely LOC_Os03g17700.1 (MAP kinase BIMK1) and LOC_Os04g58910.1 (receptor-like protein kinase-like protein), and transcription factor LOC_Os09g35790.1 (heat stress transcription factor 24) (Supplementary Table 1). This finding suggests the involvement of a protein kinase signal transduction in the expression of genes associated with the aroma formation in rice. Here, it should be noted that the mitogen-activated protein kinase signaling pathway is one of the important pathways of the secondary metabolite biosynthesis (Liu et al., 2014).

Identification of candidate genes in the aroma PPI network

The genes that are critical in the aroma PPI network (Table 1) may serve as novel candidates for genes implicated in the aroma formation. The subnetworks of critical nodes open up the opportunity for a more comprehensive investigation of the biological functions associated with the aroma formation. Interestingly, the genes recognized as critical have not been reported previously to be involved in aroma formation and may thus be used as a source of novel genes for further experiments on improving the aroma of rice cultivars. Further investigation into the PPI network revealed that the critical gene LOC_Os04g48410.1 (Os04g0573200) (copper chaperone for superoxide dismutase, putative, expressed) hadindirect interaction with LOC_Os05g38170.1 (Os05g 0455700) (putative uracil phosphoribosyltransferase) via LOC_Os01g04260.1 (putative sterol-C5(6)-desaturase) – Supplementary Table 1. As shown in Table 2, LOC_Os05g38170.1 was one of the genes that enriched the "cellular aromatic compound metabolic process (GO : 0006725)." This finding may corroborate the significance of the critical gene LOC_Os04g48410.1 (copper chaperone for superoxide dismutase, putative, expressed) in the aroma formation in rice.

Conclusion

Rice is one of the most economically important crops in the world. The formation of aromatic compounds in rice is known to be a sophisticated process. However, information on the manipulation of genes related to the aroma formation in rice is very limited. Therefore, we carried out GO enrichment analysis of the genes that are differentially expressed between the aromatic and non-aromatic rice varieties to shed light on the mechanisms underlying aroma formation. According to the analysis,

the presence of aroma in Pusa 1121, a basmati rice variety, was probably associated with the terpenoid metabolic process, the cellular carbohydrate metabolic process, the cellular aromatic compound metabolic process, the cellular lipid metabolic process, the cellular ketone metabolic process, and the cellular amino acid and derivative metabolic processes. Our study is the first to predict the PPI network of the aroma formation in rice and provides useful information and perspective for future studies. Furthermore, the number of differentially expressed genes in rice was reduced by the GO enrichment analysis and the PPI network to allow screening for potential candidate genes for aroma enhancement in rice. It should be emphasized that searching of genes involved in aroma formation using molecular biology methods is both time-consuming and labor-intensive, while an integrative bioinformatic strategy produces a list of candidate genes that might help in deciphering the regulatory networks of aroma formation, and consequently, engineering metabolic pathways for improving aroma in rice in a more economic way. However, further experimental work on the functions of genes predicted to be involved in aroma formation in rice is needed.

Supplementary material

Supplementary 1 (http://s9.picofile.com/file/8349508584/Supplementary_Table_1.xlsx.html).

References

Agapito G., Guzzi P.H., Cannataro M. (2013) *Visualization of protein interaction networks: problems and solutions*. BMC Bioinformatics 14 Supplementary 1: S1.

Amarawathi Y., Singh R., Singh A.K., Singh V.P., Mohapatra T., Sharma T.R., Singh N.K. (2008) *Mapping of quantitative trait loci for basmati quality traits in rice*. Mol. Breed 21: 49–65.

Anukul N., Ramos R.A., Mehrshahi P., Castelazo A.S., Parker H., Dievart A., Bennett M.J. (2010) *Folate polyglutamylation is required for rice seed development*. Rice 3: 181–193.

Bhattacharjee P., Singhal R.S., Kulkarni P.K. (2002) *Basmati rice: a review*. Int. J. Food Sci. Technol. 37: 1–12.

Cai X., Davis E.J., Ballif J., Liang M., Bushman E., Haroldsen V., Torabinejad J., Wu Y. (2006) *Mutant identification and characterization of the laccase gene family in Arabidopsis*. J. Exp. Bot. 57: 2563–2569.

Castrillón-Arbeláez P.A., Délano Frier J.P., (2016) [in:] *Abiotic and biotic stress in plants – recent advances and future perspectives*. INTECH.

Chin C.H., Chen S.H., Wu H.H., Ho C.W., Ko M.T., Lin C.Y. (2014) *CytoHubba: identifying hub objects and subnetworks from complex interactome*. BMC Syst. Biol. 8 (Suppl 4): S11.

Deborah D.A., Vemireddy L.R., Roja V. et al. (2017) *Molecular dissection of QTL governing grain size traits employing association and linkage mapping in Basmati rice*. Mol. Breed. 37: 77.

Dixon R., Paiva N. (1995) *Stress-induced phenylpropanoid metabolism*. Plant Cell 7(7): 1085–1097.

Du Z., Zhou X., Ling Y., Zhang Z., Su Z. (2010) *AgriGO: a GO analysis toolkit for the agricultural community*. Nucl. Acids Res. 38: W64–W70.

Ehlting J., Büttner D., Wang Q., Douglas C.J., Somssich I.E., Kombrink E. (1999) *Three 4-coumarate: coenzyme A ligases in Arabidopsis thaliana represent two evolutionarily divergent classes in angiosperms*. Plant J. 19: 9–20.

El Hadi M.A.M., Zhang F.J., Wu F.F., Zhou C.H., Tao J. (2013) *Advances in fruit aroma volatile research*. Molecules 18: 8200–8229.

Fraser C.M., Chapple C. (2011) *The phenylpropanoid pathway in Arabidopsis*. Arabidopsis Book 9: e0152.

Frey M., Stettner C., Pare P.W., Schmelz E.A., Tumlinson J.H., Gierl A. (2000) *An herbivore elicitor activates the gene for indole emission in maize*. Proc. Natl Acad. Sci. USA 97: 14801–14806.

Fukao Y. (2012) *Protein–protein interactions in plants*. Plant Cell Physiol. 53(4): 617–625.

Gu H., Zhu P., Jiao Y., Meng Y., Chen M. (2011) *PRIN: a predicted rice interactome network*. BMC Bioinformatics 12: 161.

Hinge V.R., Patil H.B., Nadaf A.B. (2016) *Aroma volatile analyses and 2AP characterization at various developmental stages in Basmati and Non-Basmati scented rice (Oryza sativa L.) cultivars*. Rice 9: 38–60.

Jezussek M., Juliano B.O., Schieberle P. (2002) *Comparison of key aroma compounds in cooked brown rice varieties based on aroma extract dilution analyses*. J. Agric. Food Chem. 50: 1101–1105.

Jodon N.E. (1944) *Inheritance of flower fragrance and other characters in rice*. J. Amer. Soc. Agron. 36: 844–848.

Kadam B.S., Patankar V.K. (1938) *Inheritance of aroma in rice*. Chron. Botan. 4: 32.

Kathiresan A., Lafitte H.R., Chen J., Mansueto L., Bruskiewich R., Bennett J. (2006) *Gene expression microarray and their application in drought stress research*. Field Crops Res. 97: 101–110.

Knaggs A.R. (2003) *The biosynthesis of shikimate metabolites*. Nat. Prod. Rep. 20: 119–136.

Li C.F., Zhu Y., Yu Y., Zhao Q.Y., Wang S.J., Wang X.C., Yao M.Z., Luo D., Li X., Chen L., Yang Y.J. (2015a) *Global transcriptome and gene regulation network for secondary metabolite biosynthesis of tea plant (Camellia sinensis)*. BMC Genom. 16: 560–581.

Li L., Hill-Skinner S., Liu S., Beuchle D., Tang H.M., Yeh C.T., Nettleton D., Schnable P.S. (2015b) *The maize brown*

midrib4 (bm4) gene encodes a functional folly polyglutamate synthase. Plant J. 81: 493–504.

Lindermayr C., Fliegmann J., Ebel J. (2003) *Deletion of a single amino acid residue from different 4-coumarate: CoA ligases from soybean results in the generation of new substrate specificities*. J. Biol. Chem. 278: 2781–2786.

Liu D., Sui S., Ma J., Li Z., Guo Y., Luo D., Yang J., Li M. (2014) *Transcriptomic analysis of flower development in wintersweet (Chimonanthus praecox)*. PLoS One 29; 9(1): e86976.

Mikkelsen M.D., Naur P., Halkier B.A. (2004) *Arabidopsis mutants in the C-S lyase of glucosinolate biosynthesis establish a critical role for indole-3-acetaldoxime in auxin homeostasis*. Plant J. 37: 770–777.

Pachauri V., Mishra V., Mishra P., Singh A.K., Singh S., Singh R., Singh N.K. (2014) *Identification of candidate genes for rice grain aroma by combining QTL mapping and transcriptome profiling approaches*. Cereal Res. Commun. 42(3): 376–388.

Pachauri V., Singh M.K., Singh A.K., Singh S., Shakeel N.A., Singh V.P., Singh N.K. (2010) *Origin and genetic diversity of aromatic rice varieties, molecular breeding and chemical and genetic basis of rice aroma*. J. Plant Biochem. Biotechnol. 19: 127–143.

Pedicini M., Barrena F., Clancy T., Castiglione F., Hovig E., Kanduri K., Santoni D., Benson M. (2010) *Combining network modeling and gene expression microarray analysis to explore the dynamics of Th1 and Th2 cell regulation*. PLoS Comput. Biol. 6(12): e1001032.

Qiu Z.J., Zhang Y.S. (2003) *Why fragrance rice produced in Thailand can be sold worldwide?* World Agric. (China) 2: 33–36.

Rapaport F., Zinovyev A., Dutreix M., Barillot E., Vert J.P. (2007) *Classification of microarray data using gene network*. BMC Bioinformatics 8: 35.

Sakakibara H. (2006) *Cytokinins: activity, biosynthesis, and tanslocation*. Annu. Rev. Plant Biol. 57: 431–49.

Sanz C., Olias J.M., Perez A.G. (1997) *Aroma biochemistry of fruits and vegetables*. [in:] *Phytochemistry of fruit and vegetables*. Oxford University Press Inc.: New York, NY, USA: 125–155.

Schwab W., Davidovich-Rikanati R., Lewinsohn E. (2008) *Biosynthesis of plant-derived flavor compounds*. Plant J. 54: 712–732.

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 Res. 13: 2498–2504.

Tieman D., Taylor M., Schauer N., Fernie A.R., Hanson A.D., Klee H.J. (2006) *Tomato aromatic amino acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde*. Proc. Natl. Acad. Sci. USA 103: 8287–8292.

Vanholme R., Demedts B., Morreel K., Ralph J., Boerjan W. (2010) *Lignin biosynthesis and structure*. Plant Physiol. 153: 895–905.

Wakasa K., Ishihara A. (2009) *Metabolic engineering of the tryptophan and phenylalanine biosynthetic pathways in rice*. Plant Biotechnol 26: 523–533.

Wei S., Tao S., Qin G., Wang S., Tao J., Wu J., Wu J., Zhang S. (2016) *Transcriptome pro ling reveals the candidate genes associated with aroma metabolites and emission of pear (Pyrus ussuriensis cv.).* Sci. Hortic. 206: 33–42.

Zhang X., Liu C.J. (2015) *Multifaceted regulations of gateway enzyme phenylalanine ammonia-lyase in the biosynthesis of phenylpropanoids*. Mol. Plant 8: 17–27.