



Gene Discovery of Characteristic Metabolic Pathways in the Tea Plant (*Camellia sinensis*) Using ‘Omics’-Based Network Approaches: A Future Perspective

Shihua Zhang¹, Liang Zhang¹, Yuling Tai², Xuwen Wang³, Chi-Tang Ho⁴ and Xiaochun Wan^{1*}

¹ State Key Laboratory of Tea Plant Biology and Utilization, Institute of Applied Mathematics, Anhui Agricultural University, Hefei, China, ² School of Life Sciences, Anhui Agricultural University, Hefei, China, ³ Department of Genetics, University of Georgia, Athens, GA, United States, ⁴ Department of Food Science, Rutgers University, New Brunswick, NJ, United States

OPEN ACCESS

Edited by:

Yi Zhao,
Institute of Computing
Technology (CAS), China

Reviewed by:

Haibo Liu,
University of Massachusetts Medical
School, United States
Wei-Lun Hung,
University of Florida, United States

*Correspondence:

Xiaochun Wan
xcwan@ahau.edu.cn

Specialty section:

This article was submitted to
Bioinformatics and Computational
Biology,
a section of the journal
Frontiers in Plant Science

Received: 15 October 2017

Accepted: 29 March 2018

Published: 04 June 2018

Citation:

Zhang S, Zhang L, Tai Y, Wang X,
Ho C-T and Wan X (2018) Gene
Discovery of Characteristic Metabolic
Pathways in the Tea Plant (*Camellia
sinensis*) Using ‘Omics’-Based
Network Approaches: A Future
Perspective. *Front. Plant Sci.* 9:480.
doi: 10.3389/fpls.2018.00480

Characteristic secondary metabolites, including flavonoids, theanine and caffeine, in the tea plant (*Camellia sinensis*) are the primary sources of the rich flavors, fresh taste, and health benefits of tea. The decoding of genes involved in these characteristic components is still significantly lagging, which lays an obstacle for applied genetic improvement and metabolic engineering. With the popularity of high-throughput transcriptomics and metabolomics, ‘omics’-based network approaches, such as gene co-expression network and gene-to-metabolite network, have emerged as powerful tools for gene discovery of plant-specialized (secondary) metabolism. Thus, it is pivotal to summarize and introduce such system-based strategies in facilitating gene identification of characteristic metabolic pathways in the tea plant (or other plants). In this review, we describe recent advances in transcriptomics and metabolomics for transcript and metabolite profiling, and highlight ‘omics’-based network strategies using successful examples in model and non-model plants. Further, we summarize recent progress in ‘omics’ analysis for gene identification of characteristic metabolites in the tea plant. Limitations of the current strategies are discussed by comparison with ‘omics’-based network approaches. Finally, we demonstrate the potential of introducing such network strategies in the tea plant, with a prospects ending for a promising network discovery of characteristic metabolite genes in the tea plant.

Keywords: the tea plant, characteristic metabolic pathway, plant-specialized metabolite, transcriptomics, metabolomics, gene discovery, network approach

Abbreviations: ADC, arginine decarboxylase; AIDA, alanine decarboxylase; ALT, alanine aminotransferase; ANR, anthocyanidin reductase; ANS, anthocyanidin synthase; CHI, chalcone isomerase; DEG, differentially expressed gene; DFR, dihydroflavonol 4-reductase; EGCG, epigallocatechin-3-gallate; F3H, flavanone 3-hydroxylase; F3’H, flavonoid 3’-hydroxylase; F3’5’H, flavonoid 3’,5’-hydroxylase; GC, gas chromatography; GCH, galloylated catechins hydrolase; GLS, glutaminase; GOGAT, glutamate synthase; GS, glutamine synthetase; GSL, glucosinolate; GST, glutathione S-transferase; LAR, leucocyanidin reductase; LC, liquid chromatography; MS, mass spectrometry; NGS, next generation sequencing; NMR, nuclear magnetic resonance; PAL, phenylalanine ammonia-lyase; PCC, Pearson correlation coefficient; RNA-seq, deep mRNA sequencing; TF, transcription factor; TS, theanine synthetase; UGT, UDP-glycosyltransferase.

INTRODUCTION

The tea plant (*Camellia sinensis*) in the family *Theaceae* is an important commercial crop that is extensively cultivated in Asian, African, Latin American, and Oceanian countries (ITC, 2014). Leaves of this crop serve as the source of a popular non-alcoholic beverage known as “tea” due to its abundant production of many valuable secondary metabolites, such as polyphenols, alkaloids, theanine, vitamins, minerals, and volatile oils (Zhang et al., 2006, 2012; Chen et al., 2009; Sugimoto et al., 2009; Guo et al., 2011; Lu et al., 2011; Ghiringhelli et al., 2012; Weerawatanakorn et al., 2015). Among these small-molecular compounds, flavonoids, theanine, and caffeine represent the three major characteristic secondary metabolites that are main determinants of the rich flavors, fresh taste, and health benefits of tea (Yamamoto et al., 1997). For example, theanine and caffeine are the main taste compounds of green tea and contribute to the umami and bitterness, respectively (Narukawa et al., 2014). There have accumulated plenty of reports about the health benefits of EGCG, such as anti-oxidation, anti-inflammation, and anti-tumor (Tipoe et al., 2007). From the view of their biosynthesis, flavonoids originate from diverse branches of the phenylpropanoid pathway and include flavones, flavonols, isoflavones, flavanones, flavanols, anthocyanidins, and dihydroflavonols (Dixon and Pasinetti, 2010). Theanine biosynthesis starts from glutamine and pyruvate, and depends on the enzymatic processes of TS, GS, GLS, ALT, and ADC (Sasaoka et al., 1965). Caffeine is a purine alkaloid and its biosynthetic pathway comprises purine biosynthesis and purine modification steps (Li et al., 2015). In the tea plant, the disclosure of genes involved in the characteristic components biosynthesis is still lagging far behind the model *Arabidopsis thaliana* and even many non-model plants [e.g., sorghum (Blomstedt et al., 2015) and tomato (Verpoorte and Memelink, 2002)], which inevitably lays an obstacle to the potential applications in genetic improvement and metabolic engineering.

Recent advances in high-throughput phenotyping technologies, such as transcriptomics and metabolomics, have accumulated massive ‘omics’ datasets that quantify the expression/accumulation profile of transcripts/metabolites and facilitate the evaluation of interactions among these cellular components (transcriptional regulatory networks, metabolic pathways) and networks (Barabasi and Oltvai, 2004). In case of the associations of genes and plant-specialized (secondary) metabolites, two elements in cellular networks, namely nodes and edges, denote genes and/or metabolites and gene-to-gene and/or gene-to-metabolite interactions, respectively. In the past few decades, parallel and integrated analysis of the multi-‘omics’ datasets in a network fashion, such as gene co-expression network and gene-to-metabolite network, has become efficient ways to identify genes underlying specialized metabolism in the model *Arabidopsis thaliana* and other non-model plants (Oksman-Caldentey et al., 2004; Hirai, 2009). Such gene discovery strategies are based on a simple assumption that genes involved in a specialized metabolic pathway are coordinately regulated under a shared regulatory system, using the ‘guilt-by-association’ principle (Oliver,

2000; Saito et al., 2008). With the large-scale ‘omics’ datasets generated in the tea plant, it is now a possible active area of gene discovery in characteristic secondary metabolism of this crop by borrowing the above-mentioned powerful ‘omics’-based network approaches.

With the above considerations, this review focuses on the agriculturally important crop, tea plant, in which the key genes of characteristic metabolites remains poorly understood. Firstly, we introduce recent advances in transcriptomics and metabolomics for transcript and metabolite profiling, and highlight different ‘omics’-based network strategies for the gene discovery in plant-specialized metabolism using successful examples that are applied in the model *Arabidopsis thaliana* and non-model plants (e.g., tomato, wheat). Further, we summarize recent progress in the ‘omics’ analysis for gene discovery of characteristic secondary metabolism in the tea plant, and limitations of the current strategies are discussed by comparison with ‘omics’-based network approaches. Finally, the potential of introducing ‘omics’-based network approaches in the tea plant are demonstrated, with a prospects ending for the promising network discovery of characteristic metabolite genes in the tea plant.

RECENT ADVANCES IN TRANSCRIPTOMICS AND METABOLOMICS

Deep mRNA Sequencing (RNA-Seq) for Transcript Profiling

Microarray and deep sequencing are useful technologies in transcript profiling (transcriptomics) due to their high-throughput and coverage (Malone and Oliver, 2011). In the tea plant, microarray has seldom been used because of the lack of biological resources (especially genomic sequences) that can aid in molecular probe design (Shi et al., 2011). In this review for the tea plant, phenotyping technology ‘transcriptomics,’ specifically RNA-seq (regardless of rRNA and non-coding RNA), can be applied in non-model species without reference genomes (Trapnell et al., 2012). There have accumulated many RNA-seq examples for the tea plant in the past 5 years. As a NGS technology, RNA-seq has become an efficient functional genomics tool in generating large-scale, low-cost mRNA expression data in model plants (*Arabidopsis thaliana* and rice) and non-model plants, such as crop (legumes, maize, and wheat), vegetables (cabbage and tomato), and trees (*Populus*) (Agarwal et al., 2014). Currently, several public repositories, such as Sequence Read Archive (SRA) at National Center for Biotechnology Information (Leinonen et al., 2010b) and European Nucleotide Archive (ENA) at European Molecular Biology Laboratory (Leinonen et al., 2010a), have made a vast volume of RNA-seq data available, serving as valuable resources for the data-driven knowledge discovery in plant secondary metabolism and many other fields.

A complete pipeline of RNA-seq consists of biological sample preparation, library construction, deep sequencing on a sequencing platform, and the downstream bioinformatics analysis (Wang et al., 2009). Routine analysis of RNA-seq data includes sequence alignment and/or *de novo* assembly, gene pathway and function annotation, gene expression qualification, and statistically identification of DEGs, which usually depends on whether a reference genome is sequenced for a species of interest (Conesa et al., 2016). Advanced analysis of RNA-seq data may be related to gene regulatory network reconstruction, gene module and motif analysis, and others (Iancu et al., 2012; Tierney et al., 2012). For experimental biologists without bioinformatics skills, there are several standalone softwares and online web-servers, such as RobiNA (Lohse et al., 2012) and TRUFA (Kornobis et al., 2015), for the easily-operated RNA-seq data analysis.

Metabolomics Technology for Metabolite Profiling

Plants are a rich source of diverse specialized metabolites that have been used for a very long period as fragrances, flavors, colorants, insecticides, and pharmaceuticals (Facchini et al., 2012). A metabolome, known as a complete set of small-molecule metabolites in an organism, represents the resulting phenotype of cells deduced by the perturbation of gene expression, which are usually governed by external environmental changes (Saito and Matsuda, 2010). Therefore, metabolomics is of great importance in understanding cellular systems and decoding gene functions. A main concern in this field is metabolite profiling, i.e., targeted or non-targeted measurement of hundreds or potentially thousands of metabolites, such as amino acids, alkaloids, polyphenols, minerals, phenolics, and vitamins (Joyce and Palsson, 2006). To achieve this, a combination of sample extraction protocols, separation techniques such as GC and LC, and spectroscopic techniques such as MS and NMR spectroscopy are required for the quantitative and qualitative analysis of metabolites extracted from isolated plant cells or tissues (Kopka et al., 2004).

One of the disadvantages of current metabolomics is the limitation of public databases of metabolite accumulation in plants (include the tea plant) under different conditions; this is in striking contrast to transcriptomics for which many useful databases are readily available for the potential functional genomics research (Saito et al., 2008). To compensate this, several metabolome resources and tools have been recently developed. Among these, PMR (Bais et al., 2010), MPMR (Wurtele et al., 2012), and MeKO (Fukushima et al., 2014) are 'metabolite profiling'-oriented databases that facilitate the sharing of comprehensive metabolome datasets in plants. In addition, organism-specific databases, such as MoTo DB (Moco et al., 2006) and SoyMetDB (Joshi et al., 2010), are now emerging in certain plant species of interest. Accompanying with these available metabolome resources, useful statistical and visualized tools, such as MeltDB (Neuwegger et al., 2008), MBRole (Chagoyen and Pazos, 2011) and MetaMapp (Barupal et al., 2012), have also been developed to reinforce this promising field.

A SURVEY OF 'OMICS'-BASED NETWORK STRATEGIES FOR THE IDENTIFICATION OF SPECIALIZED METABOLITE GENES IN PLANTS

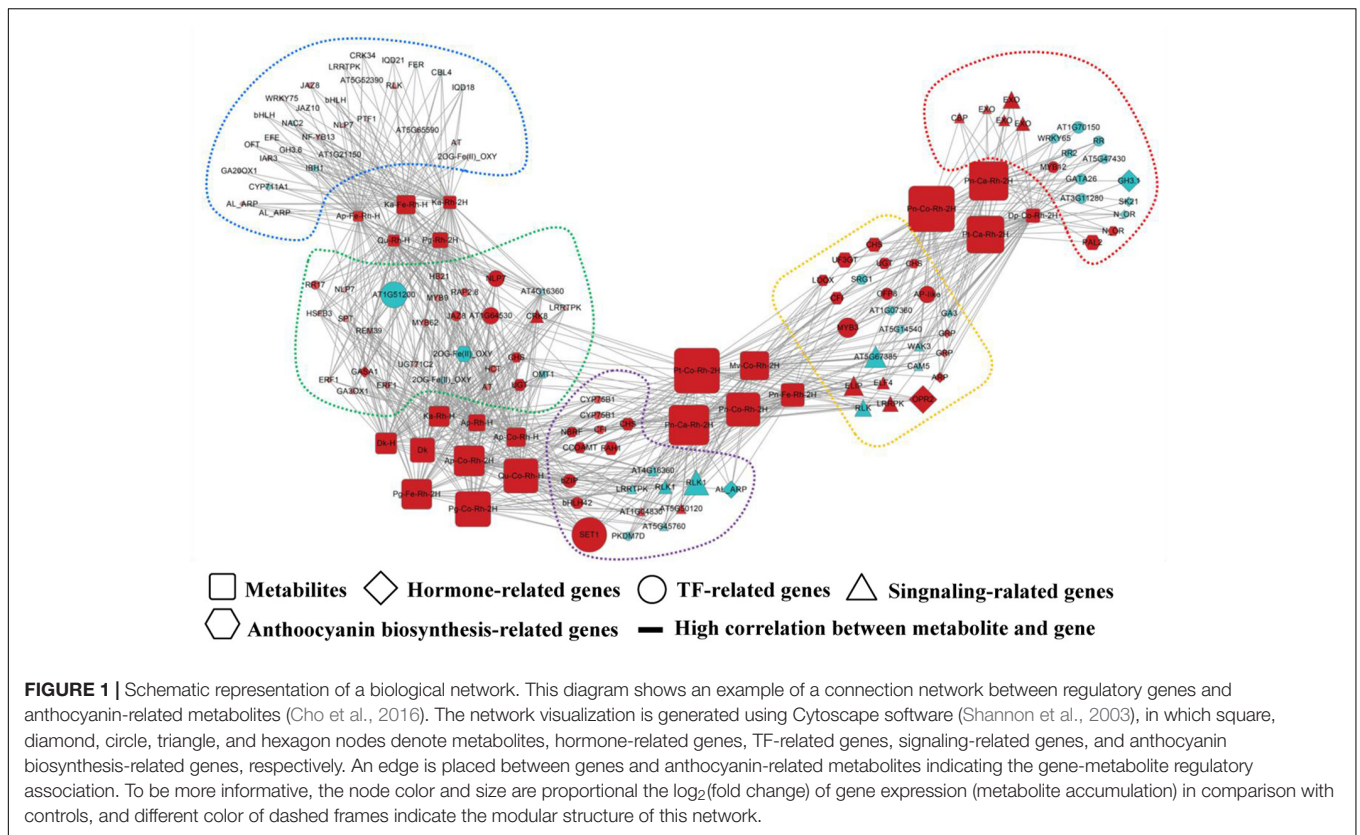
Introduction to Biological Network

Biological networks are graph representations of molecular interactions in a biological cell system. A network can be defined as a set of nodes (or vertices), denoting metabolites, genes or gene products, and a set of directed or undirected edges (**Figure 1**), denoting the interactions between them (e.g., regulatory relationships, direct physical interactions, functional associations). The cell can be viewed as an overlay of at least three types of networks, which describes transcriptional regulations (directed), protein-protein interactions (undirected), and metabolic reactions (directed) (Alon, 2003). Similar to other naturally occurring networks such as those seen in computer science, power grids, social communication and the World Wide Web, biological networks have the characteristic topological organization, such as small-world and scale-free properties (Dorogovtsev and Mendes, 2013). In many fields of plant sciences including secondary metabolism focused here, it has been demonstrated that networking modeling using single- or multi-'omics' data has the possibility in capturing many of the essential characteristics of complicated biological cell systems (Hecker et al., 2009). More detailed information on biological networks, such as network reconstruction, visualization, and topological analysis, can be seen in several comprehensive review articles (Ravasz and Barabási, 2003; Barabasi and Oltvai, 2004).

Gene Co-expression Network

After a glance of the basic concept of biological network, several types of 'omics'-based network approaches for the identification of plant-specialized metabolite genes will be further introduced. The commonly used in this field is the application of transcriptome-based gene co-expression network (Higashi and Saito, 2013). In diverse types of biological networks, gene co-expression network can be attributed to a form of gene functional association network as it is inferred from the similarity of gene expression patterns across a wide array of experimental conditions. Publicly available datasets and in-house datasets can be used to compute the gene expression similarity, resulting in condition-independent and condition-dependent gene co-expression network, respectively [**Table 1** (Hirai, 2009)]. Different selection strategies of experimental conditions have been summarized by Usadel et al. (2009) for a comprehensive assessment. The measure commonly used for gene expression similarity is PCC because most secondary metabolism in plants proceeds along linear pathways and consists of irreversible chemical reactions governed by a single enzyme or a TF. Different measures used for gene expression similarity were summarized in **Table 2** and their advantages/disadvantages have been discussed by Schaefer et al. (2017).

In systems biology, a logically conceivable assumption is that a set of genes involved in a particular biological process (more practically in a secondary metabolic pathway) are



co-regulated and thus co-expressed under the control of a shared regulatory system. This is a typical example of the so-called ‘guilt-by-association’ principle. Based on this, gene co-expression network is widely utilized to identify genes in particular secondary metabolism in plants using guide-gene (i.e., bait gene) or non-targeted protocols (Aoki et al., 2007). In the model *Arabidopsis thaliana*, many enzyme genes and regulatory TF genes have been disclosed in specialized metabolic pathways, such as GSL [see **Figure 2** as a pioneering example (Hirai et al., 2007)], phenylpropanoid (Tokimatsu et al., 2005), cellulose (Brown et al., 2005), brassinosteroid (Lisso et al., 2005), hemicelluloses (Cocuron et al., 2007), flavonoid (Wei et al., 2006), and isoprenoid (Wille et al., 2004).

Inspiringly, a large proportion of these candidate genes have been confirmed using reverse-genetic and biochemical experiments. Following these good examples, specialized metabolite genes have been continuously identified in non-model plants, such as *Cannabis sativa*, *Papaver somniferum*, *Solanum lycopersicum*, and *Catharanthus roseus* (Higashi and Saito, 2013). In public, there have appeared several co-expression databases and tools, such as ATTED-II (Obayashi et al., 2008), CSB.DB (Steinhauser et al., 2004), and GeneCAT (Mutwil et al., 2008) for model plant, primarily *Arabidopsis thaliana*, but also rice, poplar, barley, and others, which can be easily accessed for experimental biologists who have no any programming skills.

TABLE 1 | Gene co-expression network classification based on experimental condition selection.

Network type	Transcriptome data source	Experimental condition	Gene co-expression correlation
Condition-dependent	In-house dataset	Specific condition of interest	Condition-biased
Condition-independent	Public dataset	A wide range of conditions	No bias

TABLE 2 | Similarity measures used to calculate gene co-expression relationship.

Similarity measure	Statistic method	Computational efficiency	Gene co-expression relationship
Pearson correlation coefficient (PCC)	Parametric	Inexpensive	Linear
Spearman correlation coefficient (SCC)	Non-parametric	Moderate	Non-linear
Kendall correlation coefficient (KCC)	Non-parametric	Moderate	Non-linear
Mutual information (MI)	Non-parametric	Expensive	Linear and non-linear

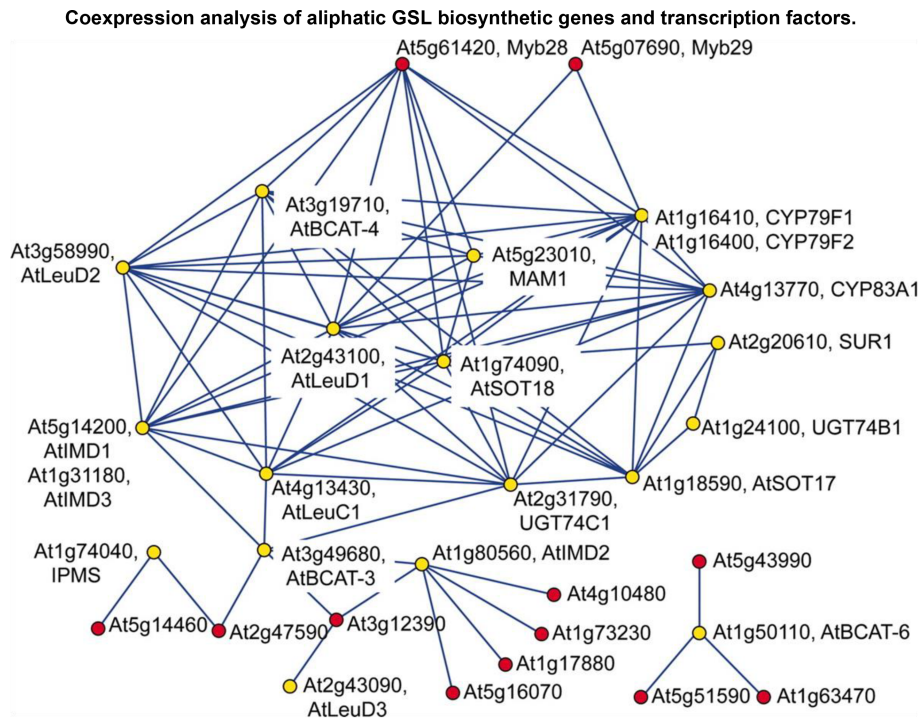


FIGURE 2 | A gene co-expression network application in the gene discovery of plant-specialized metabolism. A gene co-expression module consisting of aliphatic GSL biosynthetic genes (yellow) and TF genes (red) is illustrated in this figure (Hirai et al., 2007), which was derived from the exhaustive analysis of co-expression between GSL genes and TF genes using a dataset of publicly condition-independent transcriptome profiles. This network visualization revealed that the known aliphatic GSL biosynthetic genes (as bait genes) were clustered in the same module together with two uncharacterized TF genes, *Myb28* and *Myb29*, suggesting that these two TF genes may be positive regulators of aliphatic GSL biosynthesis (experimentally validated in the following reverse-genetic and molecular experiments).

Comparative Co-expression Analysis: From the View of Cross-Species Conversation

One possible caveat in gene discovery using gene co-expression relationships in single-species analysis is the risk of ‘false positives,’ i.e., co-expressed genes might be accidentally co-expressed rather than being functionally related. Generally, gene co-expression relationships in certain biological processes, such as cell cycle and protein synthesis and degradation, are conserved across diverse species (Stuart et al., 2003). Thus, two co-expressed genes from one species usually have their orthologs in another species that in turn are also co-expressed. According to this paradigm, the examination of conserved gene co-expression relationships across different species can be used to minimize the risk of ‘false positives’ via the knowledge transfer from model organisms to non-model plants (Movahedi et al., 2012). There are several tools, such as ATTED-II (Obayashi et al., 2011), CoP (Ogata et al., 2009) and StarNet (Jupiter et al., 2009), which allows between-species comparison of gene co-expression networks. In addition, the NetworkComparer pipeline deployed in PlaNet (Mutwil et al., 2011) can aid in multi-species comparison (Mutwil et al., 2011). This tool bins genes into gene families according to their Pfam annotation (Finn et al., 2016) and compares gene vicinity networks from the query genes for re-occurring Pfams. Using *AtPAL1* (responsible for the initial

step for monolignol synthesis) as bait gene in NetworkComparer, Ruprecht and Persson (2012) inferred a consensus network of *AtPAL1*-orthologs in barley, *Medicago*, polar, rice, soybean, and wheat, from which gene families corresponding to the subsequent steps in lignin biosynthetic pathway, such as *4CL* (‘AMP_binding’), *C4H* (‘p450’), *HCT* (‘Transferase’), *CCoAOMT* (‘Methyltransf_3’), *CCR* (‘Epimerase’), and *CAD* (‘ADH_zinc_N’), were identified and experimentally confirmed.

Gene Co-functional Network: From Single- to Multi-Network

As described above, gene co-expression network is a form of gene functional association network, which is only based on the similarity of gene expression patterns across a variety of experimental conditions. There exist other types of molecular relationships, such as physical interaction of proteins, subcellular co-localization, domain co-occurrence, co-citation in literature, mutant phenotypes, gene neighbors, genetic interactions, conserved motif sequences, and enzymatic reaction. These different levels of relationships are necessarily to be integrated in the original gene co-expression network to expand the network concept from single- to multi-network that can be called as gene co-functional network (Fukushima and Kusano, 2014). To increase the predictive power for gene identification, publicly available resources, such as AraNet (Lee et al., 2015),

ATTED-II (Obayashi et al., 2011), and PlaNet (Mutwil et al., 2011), have integrated one or several of those valuable molecular relationships into the gene co-expression network, facilitating a more comprehensive gene co-functional network resource for users. It is noted that these efforts are limited to *Arabidopsis thaliana* because it is a data-rich model species (Lee et al., 2015). For other plants species, high coverage and confidence gene co-functional relationships can also be predicted via efficient computational pipelines. For instance, genome-wide gene co-functional networks have recently been integratively achieved in *Glycine max* (Kim E. et al., 2017), *Solanum lycopersicum* (Kim H. et al., 2017), and *Triticum aestivum* (Lee et al., 2017) by translating molecular interaction knowledge from data-rich model species (e.g., *Arabidopsis thaliana* and *Saccharomyces cerevisiae*) in them and using their in-species gene expression data.

Gene-to-Metabolite Network: Co-occurrence Analysis of Genes and Metabolites

The metabolome in plant cell is the final product of a series of gene actions. Hence, metabolomics, when integrated with transcriptomics, provides a potential for the study of gene-to-metabolite networks that control specialized metabolism in plants, both at the catalytic and regulatory levels. Currently, this integrative analysis (especially the co-occurrence analysis in a network fashion) of gene expression and metabolite accumulation has emerged as an alternative strategy for the identification of novel gene functions involved in plant-specialized metabolism. Hirai et al. (2005) presented one of the very first articles successfully analyzing the gene/metabolite associations in a gene-to-metabolite network via the integration of time-series transcriptome and metabolome datasets. On the basis of known genes in GSL biosynthesis, the authors identified the following genes as candidates involved in GSL pathways: three putative sulfotransferase genes (At1g74100, At1g18590, and At1g74090), an S-glucosyltransferase gene (At1g24100), a putative Tyr aminotransferase gene (At5g36160), and two putative GST genes (At3g03190 and At1g78370). To date, some of these candidate genes have been experimentally characterized in concurrent studies using biochemical approaches (Grubb et al., 2004; Piotrowski et al., 2004).

RECENT PROGRESS IN 'OMICS' ANALYSIS FOR THE IDENTIFICATION OF CHARACTERISTIC METABOLITE GENES IN THE TEA PLANT

Transcriptome Analysis

We conducted the first RNA-seq based specialized metabolite gene analysis for the tea plant in the early 2011 through deep sequencing, *de novo* assembling and functional annotation of the transcriptome of a pooled sample of seven tissues including tender shoots, young leaves, mature leaves, stems, young roots, flower buds, and immature seeds (Shi et al., 2011).

On this basis, many putative candidate genes involved in the three major secondary metabolic pathways (flavonoids, theanine, and caffeine) that tightly related to tea quality and taste were target-disclosed. Among these, several genes associated with theanine and flavonoid biosynthesis were experimentally validated using low throughput RT-PCR and qRT-PCR analysis. Generally, gene expression changes underlying specialized metabolite accumulation in plants at specific experimental conditions, developmental time points or different tissues is necessary to be measured for the disclosure of complicated regulatory mechanisms of specialized metabolites biosynthesis. In these cases, knowledge-gain aimed experimental designs (e.g., control/treat coupling and sample repetition) are preferentially considered to subject to the downstream data analysis namely DEG identification in routine RNA-seq analytical pipeline. Evidently, our effort is based on a multi-tissue composed mixture sample without the above biological concerns. With the popularity of RNA-seq analysis and increasing decrease of sequencing cost, many interesting hypotheses have been developed in the later 5 years by focusing specialized metabolism mechanisms of the tea plant in certain abiotic/biotic stress conditions [e.g., drought (Wang W. et al., 2016) and pathogen attack (Wang Y.N. et al., 2016)], different tissues [e.g., bud, stem, flower, and seed (Li et al., 2015)], and different developmental stages [e.g., leaf tissues at color-changing (Li et al., 2017) and form-shaping stages (Lin et al., 2017)].

Comparative Transcriptome Analysis

Beyond the routine transcriptome analysis, comparative transcriptome analysis has become a valuable strategy in dissection of significant differences in genes and their expressions among different biological samples of a certain species, similar biological samples of different cultivars in a certain species, or even similar biological samples in different species. It is known that oil tea (*Camellia oleifera*) from the same genus *Camellia* lacks the three characteristic metabolites (flavonoids, theanine, and caffeine) that the tea plant (*Camellia sinensis*) predominately possesses. To uncover the genetic components underlying the biosynthesis of characteristic components in tea, we applied a cross-species transcriptome comparison by choosing bud and leaf tissues from the two *Camellia* plants (Tai et al., 2015). Based on the RNA-seq analysis, we experimentally confirmed that several enzyme genes associated with flavonoid, theanine and caffeine pathways, such as PAL, CHI, DFR, and F3H, exhibited considerably different expressions in tea compared to oil tea using qRT-PCR analysis. Thus, it can be speculated that the differential expressions in certain genes behave as the contributing genetic basis for the divergence of metabolite contents in the two plants of the same *Camellia* genus. Another representative example is that conducted by Wu et al. (2014) where leaf transcriptomes of the four tea plant cultivars, 'Yunnan Shilixiang,' 'Chawan Sanhao,' 'Rucheng Maoyecha,' and 'Anji Baicha,' with different percentages of various catechins, were subjected to a deep comparative analysis. In this effort, three catechin closely-related enzyme genes, ANS, ANR and LAR, were unraveled to be as key factors involved in the changed catechin percentages in different tea plant cultivars.

Integrated Transcriptome and Metabolome Analysis

Single metabolome analysis applied in the tea plant is mainly focused on the accumulation patterns of chemical components in certain tea products [e.g., green (Lee et al., 2011) and black (Pan et al., 2017)] or plant parts [mainly leave (Pauli et al., 2016; Ryu et al., 2017)]. In these studies, the elucidation of metabolite-related genes has not been concerned. It is promising that the connection between gene expression and metabolite accumulation should be considered to study the ‘cause-to-effect’ relationships in the biosynthesis of specialized metabolites in plants because a metabolome in cell system represents the phenotype effect of gene actions. In the past few years, there has appeared several attempts that focused on integrated transcriptomics and metabolomics analysis in the tea plant, which is a popular fashion used from the model plant *Arabidopsis thaliana* to vegetables (Alba et al., 2005), fruits (Savoi et al., 2016), crops (Kovinich et al., 2011), and trees (Hamanishi et al., 2015). In a recent effort, Li et al. (2016) presented a combined transcriptomics and metabolomics analysis of ‘Anji Baicha’ (*Camellia sinensis*) leaves at yellow-green, albescent, and re-greening stages. In theanine biosynthetic pathway, one of the three characteristic metabolic pathways, the authors found that the expressions of four genes, GOGAT, AIDA, GS, and TS, were significantly correlated with the concentrations of ethylamine (GOGAT), glutamine (GOGAT, AIDA, GS, and TS), and theanine (AIDA) in this pathway, which are likely the causes of the leaf metabolite variability among the three color and developmental stages.

LIMITATIONS OF THE CURRENT GENE DISCOVERY STRATEGIES OF CHARACTERISTIC METABOLITES IN THE TEA PLANT IN CONTRAST WITH ‘OMICS’-BASED NETWORK STRATEGIES

The exploration of genes responsible for characteristic components biosynthesis is an important branch of tea biochemistry research. Previously, characteristic metabolite genes were mostly discovered through Sanger sequencing (Park et al., 2004; Singh et al., 2009). With the advent of NGS technology (specifically RNA-seq), the gene decoding has made great achievements in the determination of gene structures and expression profiles relying on its high throughput and coverage superiority (Ozsolak and Milos, 2011). However, current RNA-seq based ‘omics’ analysis has several limitations in metabolite gene discovery by comparison with ‘omics’-based network approaches: (1) genes are always identified from assembled unigenes through homology-based function annotation based on well-characterized gene references in data-rich model species. This commonly used strategy in non-model species may shield the possibility in discovery of novel enzyme genes and the underlying regulatory TFs related to certain characteristic metabolic pathways in the

tea plant. As emphatically discussed in this review, genes in a specific specialized pathway are usually co-regulated at the transcriptional level. Thus, it is logical that the gene-to-gene associations (e.g., gene co-expression network) at a genome-wide scale should be established to predict novel genes of a characteristic metabolic pathway in the tea plant based on the ‘guilt-by-association’ principle, using well-characterized bait genes. (2) Although integrated transcriptome and metabolome analysis can help disclose genes that contribute to certain metabolite accumulation pattern in certain biological conditions in the tea plant (see illustrated example above), the associations between genes and metabolites has still not been quantitatively measured into a gene-to-metabolite network (can be seen in other species) based on a multi-sample statistical model. That is to say, integrated transcriptome and metabolome analysis currently applied in the tea plant is intrinsically parallel-isolated at the two levels of gene expression and metabolite accumulation. (3) Among the three major characteristic components in the tea plant, theanine has been found in some *Camellia* species and in a mushroom, *Xerocomus badius* (Casimir et al., 1960). Therefore, the metabolic pathway associated with theanine biosynthesis has no reference pathway from other model plants to identify metabolite genes using the homology-based knowledge translation. (4) *Camellia sinensis* (the tea plant) is evolutionarily far-distant from *Arabidopsis thaliana* and other secondary metabolism well-established model plants. Thus, cross-species gene knowledge translation (homology-based) may hinder the elaborate disclosure of the specific biosynthesis of characteristic components in the tea plant.

POTENTIAL OF INTRODUCING ‘OMICS’-BASED NETWORK STRATEGIES IN THE IDENTIFICATION OF CHARACTERISTIC METABOLITE GENES IN THE TEA PLANT

As summarized in section “A survey of ‘omics’-based network strategies for the identification of specialized metabolite genes in plants,” several prerequisites are required for the ‘omics’-based network strategies applied in the tea plant, such as sufficient sample size and well-characterized bait gene set in a certain characteristic metabolic pathway. We searched NCBI SRA (Leinonen et al., 2010b), a representative NGS sequence database, using the keyword “*Camellia sinensis*,” and manually checked biological samples that documented relevant RNA-seq applications in the tea plant. As of January 2018, more than 200 biological samples in the tea plant in different biological conditions are publicly available. In addition, dozens of in-house RNA-seq examples concerned with different biological questions have accumulated in our own lab in recent years. Thus, the RNA-seq sample size in the tea plant is now sufficient to allow for the statistical computation of co-expression relationships of pairwise genes [see sample requirement for gene co-expression decision in review (Aoki et al., 2007)]. As to bait genes, researchers in our tea lab at Anhui Agricultural University

and colleagues around the world have contributed considerable efforts in biological molecular experiments that well-characterize genes related to the three characteristic metabolic pathways in the tea plant, such as UGT (Cui et al., 2016), *F3'H* (Zhou et al., 2016), GCH (Jiang et al., 2013), TS (Okada et al., 2006), and *F3'5'H* (Wang et al., 2014). Most recently, the genome of tea plant has been sequenced and released in public (Xia et al., 2017). Using this resource as a reference, we can accurately track genes and compute their expressions in diverse sequenced samples. In addition, researchers can also use the reference to computationally predict several molecular functional associations, such as protein–protein physical interaction and proteins subcellular co-localization, to enhance single-network of gene co-expression into multi-network of gene co-function. With respect to the gene-metabolite associations, parallel experimental design and integrated analysis related to gene expression and metabolite accumulation in concerned biological conditions can be now readily achieved for the possible gene-to-metabolite network inference in the gene identification of characteristic secondary metabolism of the tea plant.

POSSIBLE ISSUES REGARDING 'OMICS'-BASED NETWORK APPROACHES

When using 'omics'-based network approaches for the identification of specialized metabolite genes in plants, several pitfalls and limitations should be concerned. As surveyed above, gene association network models (e.g., gene co-expression and co-functional networks) utilize the "guilt-by-association" principle to prioritize candidate genes that might be involved in a particular secondary pathway. However, genes in the same pathway are not necessarily co-expressed or have no co-functional relationships. For example, genes in a certain secondary pathway with post-transcriptional regulation may have no significant correlations among them in a gene model, and therefore they can't be recognized using the above "guilt-by-association" method. More information about the limitation of the "guilt-by-association" principle in gene association network analysis can be seen in a critical article reported by Gillis' group (Gillis and Pavlidis, 2012). When researchers integrate a large number of heterogeneous transcriptomic and/or metabolomic 'omics' data from public databases or their own collections in a network modeling pipeline, it should be noted that several technical issues such as batch effects and missing values intrinsically exist and may lead to misguided conclusions (Scherer, 2009). Inspiringly, there have appeared several computational techniques such as generalized R^2 statistic model (Leek et al., 2010) and missing value imputation algorithm (Liew et al., 2010) that are developed to deal with batch effects and missing values. Another notification for users is the downstream experimental validation of selected candidate genes. False positive genes are possibly subjected to time-consuming biochemical test. Among the candidate genes, relevant ones can be manually screened out via the KEGG/GO (Gene Ontology Consortium, 2014;

Kanehisa et al., 2015) functional annotation or other function database annotation and/or extensive literature reviewing of interest genes.

CONCLUSION AND FUTURE PROSPECTS

Considerable research examples demonstrate that 'omics'-based network approaches are powerful tools for the gene discovery in plant-specialized metabolism (Hirai et al., 2004; Mercke et al., 2004). However, different 'omics'-based network strategies for the identification of specialized metabolite genes in plants are not summarized in a single review article, which conceal the possible connections and differences of such systems biology approaches. Moreover, the advantages of such 'omics'-based network strategies have not been discussed in a single paper by comparison with the traditional 'omics' analysis in the identification of specialized metabolite genes in plants. As to the tea plant focused in this review, apart from the high complexity in its genome, characteristic metabolic pathways in this crop have their intrinsic features different from other plant species. For example, as an ammonium-tolerant and perennial plant species, the tea plant is different from other plants in nitrogen metabolism, which systematically governs the three characteristic metabolic pathways and makes them particular (Britto and Kronzucker, 2002). Therefore, network-fashion systems biology approaches are necessary to be introduced to facilitate the *de novo* identification of key genes involved in characteristic metabolic pathways in the tea plant. With these considerations, this review focuses on the agriculturally important crop tea plant (can also extend to others plant species) in which characteristic secondary metabolites are primary determinants of tea quality and taste and key characteristic metabolite genes are still not fully understood. We highlight different 'omics'-based network approaches for the gene discovery in plant-specialized metabolism using successful examples that are employed in model and non-model plants. Limitations of the traditional 'omics' strategies in discovery of specialized metabolite genes are discussed by comparison with 'omics'-based network approaches using the tea plant as an instance. Particularly, the potential of introducing these strategies in the relevant field of the tea plant are particularly demonstrated. We believe this will provide novel directions in the exploration of functional genes associated with characteristic components in the tea plant, which is a critical basis for applied genetic improvement and metabolic engineering.

As useful large-scale computational pipelines, 'omics'-based network approaches can provide clues about the potential candidate genes involved in certain characteristic metabolic pathways in the tea plant. Hence, a main concern is that tea biochemists should perform wet lab experiments to validate the predicted gene functions. Due to the complicated genetic background, the transformation system in the tea plant has still not been established. Thus, reverse genetics approaches, such as gene knockout and over-expressing, are not feasible for the experimental confirmation of identified gene functions

in the tea plant. In such case, we can consider the correlation analysis of gene expression and metabolite accumulation in multi-samples statistical model to achieve it, following the regular molecular experiments such as clone, protein recombination, and enzymatic activity. With the reference genome of the tea plant currently available, an active research field of metabolite-based genome-wide association study [or mGWAS (Luo, 2015)] should be focused, as it is a good way to identify genomic loci associated to key metabolic pathways by re-sequencing a collection of tea plant cultivars or recombination lines around the world and profiling their metabolomes of tissues and conditions of interest. The loci information from mGWAS analysis can be considered together with the gene information inferred from 'omics'-based network analysis to provide subtle clues for a specialized component biosynthesis in the tea plant. To date, detailed information regarding the bait genes in the three characteristic metabolic pathways in the tea plant has been scattered in published studies. As such, several tasks should be carried out to gain a comprehensive list of such bait genes, using manual curation from publications (Zhang et al., 2013) or more effective literature-mining tools (Jensen et al., 2006). In addition, exhaustive experimental characterization of characteristic metabolites genes as a more comprehensive bait gene list should be an ongoing program for gaining an optimized prediction. In the past 5 years, we have seen massive accumulation of transcriptome and metabolome datasets from different labs worldwide, which are valuable resources in tea secondary metabolism community. Thus, it is an urgent and promising task to develop the corresponding deposit and analysis platforms, which can aid in the network identification of characteristic component genes in the tea plant for experimental investigator without any

bioinformatics skills. On this base, novel algorithms in biological network analysis, such as motif and module mining, should be borrowed to help find gene- and/or metabolite-mediated regulatory sub-structures [e.g., frequently-appeared feed-forward loop in plants, FFL (Sakuraba et al., 2015)] that might control a specific specialized metabolite pathway in the tea plant. In addition to the three characteristic components in the tea plant, there are also several other specialized metabolites with important nutritional values and health benefits, such as saponins (Sur et al., 2001) and volatile terpenes (Yang et al., 2013). Now it still remains as a virgin field that should call for the analogous studies in gene discovery to advance a comprehensive understanding of secondary metabolic profile in the tea plant.

AUTHOR CONTRIBUTIONS

SZ and XW conceived the project and wrote the manuscript. LZ and YT wrote the contents regarding metabolomics and transcriptomics. XW and C-TH provided scientific criticisms and manuscript proofreading.

FUNDING

This work was funded by grants from the National Natural Science Foundation of China (31301248 and 31170283), the Young Elite Scientist Sponsorship Program by CAST (2016QNR001), the Fund of State Key Laboratory of Tea Plant Biology and Utilization (SKLTOF20160113), and the China Agriculture Research System (CARS-19).

REFERENCES

- Agarwal, P., Parida, S. K., Mahto, A., Das, S., Mathew, I. E., Malik, N., et al. (2014). Expanding frontiers in plant transcriptomics in aid of functional genomics and molecular breeding. *Biotechnol. J.* 9, 1480–1492. doi: 10.1002/biot.201400063
- Alba, R., Payton, P., Fei, Z., McQuinn, R., Debbie, P., Martin, G. B., et al. (2005). Transcriptome and selected metabolite analyses reveal multiple points of ethylene control during tomato fruit development. *Plant Cell* 17, 2954–2965. doi: 10.1105/tpc.105.036053
- Alon, U. (2003). Biological networks: the tinkerer as an engineer. *Science* 301, 1866–1867. doi: 10.1126/science.1089072
- Aoki, K., Ogata, Y., and Shibata, D. (2007). Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant Cell Physiol.* 48, 381–390. doi: 10.1093/pcp/pcm013
- Bais, P., Moon, S. M., He, K., Leitao, R., Dreher, K., Walk, T., et al. (2010). PlantMetabolomics.org: a web portal for plant metabolomics experiments. *Plant Physiol.* 152, 1807–1816. doi: 10.1104/pp.109.151027
- Barabasi, A. L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113. doi: 10.1038/nrg1272
- Barupal, D. K., Haldiya, P. K., Wohlgemuth, G., Kind, T., Kothari, S. L., Pinkerton, K. E., et al. (2012). MetaMapp: mapping and visualizing metabolomic data by integrating information from biochemical pathways and chemical and mass spectral similarity. *BMC Bioinformatics* 13:99. doi: 10.1186/1471-2105-13-99
- Blomstedt, C. K., O'Donnell, N. H., Bjarnholt, N., Neale, A. D., Hamill, J. D., Möller, B. L., et al. (2015). Metabolic consequences of knocking out UGT85B1, the gene encoding the glucosyltransferase required for synthesis of dhurrin in *Sorghum bicolor* (L. Moench). *Plant Cell Physiol.* 57, 373–386. doi: 10.1093/pcp/pcv153
- Britto, D. T., and Kronzucker, H. J. (2002). NH₄⁺ toxicity in higher plants: a critical review. *J. Plant Physiol.* 159, 567–584. doi: 10.1078/0176-1617-0774
- Brown, D. M., Zeef, L. A., Ellis, J., Goodacre, R., and Turner, S. R. (2005). Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics. *Plant Cell* 17, 2281–2295. doi: 10.1105/tpc.105.031542
- Casimir, J., Jadot, J., and Renard, M. (1960). Separation and characterization of N-ethyl-γ-glutamine from *Xerocomus badius*. *Biochim. Biophys. Acta* 39, 462–468. doi: 10.1016/0006-3002(60)90199-2
- Chagoyen, M., and Pazos, F. (2011). MBRole: enrichment analysis of metabolomic data. *Bioinformatics* 27, 730–731. doi: 10.1093/bioinformatics/btr001
- Chen, L., Chen, Q., Zhang, Z., and Wan, X. (2009). A novel colorimetric determination of free amino acids content in tea infusions with 2,4-dinitrofluorobenzene. *J. Food Compos. Anal.* 22, 137–141. doi: 10.1016/j.jfca.2008.08.007
- Cho, K., Cho, K.-S., Sohn, H.-B., Ha, I. J., Hong, S.-Y., Lee, H., et al. (2016). Network analysis of the metabolome and transcriptome reveals novel regulation of potato pigmentation. *J. Exp. Bot.* 67, 1519–1533. doi: 10.1093/jxb/erv549
- Cocuron, J.-C., Lerouxel, O., Drakakaki, G., Alonso, A. P., Liepman, A. H., Keegstra, K., et al. (2007). A gene from the cellulose synthase-like C family encodes a β-1, 4 glucan synthase. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8550–8555. doi: 10.1073/pnas.0703133104
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17:13. doi: 10.1186/s13059-016-0881-8

- Cui, L., Yao, S., Dai, X., Yin, Q., Liu, Y., Jiang, X., et al. (2016). Identification of UDP-glycosyltransferases involved in the biosynthesis of astringent taste compounds in tea (*Camellia sinensis*). *J. Exp. Bot.* 67, 2285–2297. doi: 10.1093/jxb/erw053
- Dixon, R. A., and Pasinetti, G. M. (2010). Flavonoids and isoflavonoids: from plant biology to agriculture and neuroscience. *Plant Physiol.* 154, 453–457. doi: 10.1104/pp.110.161430
- Dorogovtsev, S. N., and Mendes, J. F. (2013). *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford: Oxford University Press.
- Grubb, C. D., Zipp, B. J., Ludwig-Müller, J., Masuno, M. N., Molinski, T. F., and Abel, S. (2004). Arabidopsis glucosyltransferase UGT74B1 functions in glucosinolate biosynthesis and auxin homeostasis. *Plant J.* 40, 893–908. doi: 10.1111/j.1365-313X.2004.02261.x
- Facchini, P. J., Bohlmann, J., Covello, P. S., De Luca, V., Mahadevan, R., Page, J. E., et al. (2012). Synthetic biosystems for the production of high-value plant metabolites. *Trends Biotechnol.* 30, 127–131. doi: 10.1016/j.tibtech.2011.10.001
- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344
- Fukushima, A., and Kusano, M. (2014). A network perspective on nitrogen metabolism from model to crop plants using integrated 'omics' approaches. *J. Exp. Bot.* 65, 5619–5630. doi: 10.1093/jxb/eru322
- Fukushima, A., Kusano, M., Mejia, R. F., Iwasa, M., Kobayashi, M., Hayashi, N., et al. (2014). Metabolomic characterization of knockout mutants in Arabidopsis: development of a metabolite profiling database for knockout mutants in Arabidopsis. *Plant Physiol.* 165, 948–961. doi: 10.1104/pp.114.240986
- Gene Ontology Consortium (2014). Gene ontology consortium: going forward. *Nucleic Acids Res.* 43, D1049–D1056. doi: 10.1093/nar/gku1179
- Ghiringhelli, F., Rebe, C., Hichami, A., and Delmas, D. (2012). Immunomodulation and anti-inflammatory roles of polyphenols as anticancer agents. *Anti Cancer Agents Med. Chem.* 12, 852–873. doi: 10.2174/187152012802650048
- Gillis, J., and Pavlidis, P. (2012). "Guilt by association" is the exception rather than the rule in gene networks. *PLoS Comput. Biol.* 8:e1002444. doi: 10.1371/journal.pcbi.1002444
- Guo, L., Liang, Q., and Du, X. (2011). Effects of molecular characteristics of tea polysaccharide in green tea on glass transitions of potato amylose, amylopectin and their mixtures. *Food Hydrocoll.* 25, 486–494. doi: 10.1002/jfsa.4247
- Hamanishi, E. T., Barchet, G. L., Dauwe, R., Mansfield, S. D., and Campbell, M. M. (2015). Poplar trees reconfigure the transcriptome and metabolome in response to drought in a genotype- and time-of-day-dependent manner. *BMC Genomics* 16:329. doi: 10.1186/s12864-015-1535-z
- Hecker, M., Lambeck, S., Toepfer, S., Van Someren, E., and Guthke, R. (2009). Gene regulatory network inference: data integration in dynamic models—a review. *Biosystems* 96, 86–103. doi: 10.1016/j.biosystems.2008.12.004
- Higashi, Y., and Saito, K. (2013). Network analysis for gene discovery in plant-specialized metabolism. *Plant Cell Environ.* 36, 1597–1606. doi: 10.1111/pce.12069
- Hirai, M. Y. (2009). A robust omics-based approach for the identification of glucosinolate biosynthetic genes. *Phytochem. Rev.* 8, 15–23. doi: 10.1007/s11101-008-9114-4
- Hirai, M. Y., Klein, M., Fujikawa, Y., Yano, M., Goodenowe, D. B., Yamazaki, Y., et al. (2005). Elucidation of gene-to-gene and metabolite-to-gene networks in Arabidopsis by integration of metabolomics and transcriptomics. *J. Biol. Chem.* 280, 25590–25595. doi: 10.1074/jbc.M502332200
- Hirai, M. Y., Sugiyama, K., Sawada, Y., Tohge, T., Obayashi, T., Suzuki, A., et al. (2007). Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6478–6483. doi: 10.1073/pnas.0611629104
- Hirai, M. Y., Yano, M., Goodenowe, D. B., Kanaya, S., Kimura, T., Awazuhara, M., et al. (2004). Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U.S.A.* 101, 10205–10210. doi: 10.1073/pnas.0403218101
- Iancu, O. D., Kawane, S., Bottomly, D., Searles, R., Hitzemann, R., and McWeeney, S. (2012). Utilizing RNA-Seq data for de novo coexpression network inference. *Bioinformatics* 28, 1592–1597. doi: 10.1093/bioinformatics/bts245
- ITC (2014). *Annual Bulletin of Statistics*. London: International Tea Committee, 71–81.
- Jensen, L. J., Saric, J., and Bork, P. (2006). Literature mining for the biologist: from information retrieval to biological discovery. *Nat. Rev. Genet.* 7, 119–129. doi: 10.1038/nrg1768
- Jiang, X., Liu, Y., Li, W., Zhao, L., Meng, F., Wang, Y., et al. (2013). Tissue-specific, development-dependent phenolic compounds accumulation profile and gene expression pattern in tea plant [*Camellia sinensis*]. *PLoS One* 8:e62315. doi: 10.1371/journal.pone.0062315
- Joshi, T., Yao, Q., Levi, D. F., Brechenmacher, L., Valliyodan, B., Stacey, G., et al. (2010). "SoyMetDB: the soybean metabolome database," in *Proceedings of the 2010 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Hong Kong, 203–208. doi: 10.1109/BIBM.2010.5706563
- Joyce, A. R., and Palsson, B. Ø. (2006). The model organism as a system: integrating 'omics' data sets. *Nat. Rev. Mol. Cell Biol.* 7, 198–210. doi: 10.1038/nrm1857
- Jupiter, D., Chen, H., and VanBuren, V. (2009). S TAR N ET 2: a web-based tool for accelerating discovery of gene regulatory networks using microarray co-expression data. *BMC Bioinformatics* 10:332. doi: 10.1186/1471-2105-10-332
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2015). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070
- Kim, E., Hwang, S., and Lee, I. (2017). SoyNet: a database of co-functional networks for soybean *Glycine max*. *Nucleic Acids Res.* 45, D1082–D1089. doi: 10.1093/nar/gkw704
- Kim, H., Kim, B. S., Shim, J. E., Hwang, S., Yang, S., Kim, E., et al. (2017). TomatoNet: a genome-wide co-functional network for unveiling complex traits of tomato, a model crop for fleshy fruits. *Mol. Plant* 10, 652–655. doi: 10.1016/j.molp.2016.11.010
- Kopka, J., Fernie, A., Weckwerth, W., Gibon, Y., and Stitt, M. (2004). Metabolite profiling in plant biology: platforms and destinations. *Genome Biol.* 5:109. doi: 10.1186/gb-2004-5-6-109
- Kornobis, E., Cabellos, L., Aguilar, F., Frias-Lopez, C., Rozas, J., Marco, J., et al. (2015). TRUFA: a user-friendly web server for de novo RNA-seq analysis using cluster computing. *Evol. Bioinform. Online* 11, 97–104. doi: 10.4137/EBO.S23873
- Kovinch, N., Saleem, A., Arnason, J. T., and Miki, B. (2011). Combined analysis of transcriptome and metabolite data reveals extensive differences between black and brown nearly-isogenic soybean (*Glycine max*) seed coats enabling the identification of pigment isogenes. *BMC Genomics* 12:381. doi: 10.1186/1471-2164-12-381
- Lee, J.-E., Lee, B.-J., Chung, J.-O., Shin, H.-J., Lee, S.-J., Lee, C.-H., et al. (2011). 1 H NMR-based metabolomic characterization during green tea (*Camellia sinensis*) fermentation. *Food Res. Int.* 44, 597–604. doi: 10.1016/j.foodres.2010.12.004
- Lee, T., Hwang, S., Kim, C. Y., Shim, H., Kim, H., Ronald, P. C., et al. (2017). WheatNet: a genome-scale functional network for hexaploid bread wheat, *Triticum aestivum*. *Mol. Plant* 10, 1133–1136. doi: 10.1016/j.molp.2017.04.006
- Lee, T., Yang, S., Kim, E., Ko, Y., Hwang, S., Shin, J., et al. (2015). AraNet v2: an improved database of co-functional gene networks for the study of Arabidopsis thaliana and 27 other nonmodel plant species. *Nucleic Acids Res.* 43, D996–D1002. doi: 10.1093/nar/gku1053
- Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., et al. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* 11, 733–739. doi: 10.1038/nrg2825
- Leinonen, R., Akhtar, R., Birney, E., Bower, L., Cerdeno-Tárraga, A., Cheng, Y., et al. (2010a). The European nucleotide archive. *Nucleic Acids Res.* 39, D28–D31. doi: 10.1093/nar/gkq967
- Leinonen, R., Sugawara, H., Shumway, M., and International Nucleotide Sequence Database Collaboration (2010b). The sequence read archive. *Nucleic Acids Res.* 39, D19–D21. doi: 10.1093/nar/gkq1019
- Li, C.-F., Xu, Y.-X., Ma, J.-Q., Jin, J.-Q., Huang, D.-J., Yao, M.-Z., et al. (2016). Biochemical and transcriptomic analyses reveal different metabolite biosynthesis profiles among three color and developmental stages in 'Anji

- Baicha' (*Camellia sinensis*). *BMC Plant Biol.* 16:195. doi: 10.1186/s12870-016-0885-2
- Li, C. F., Zhu, Y., Yu, Y., Zhao, Q. Y., Wang, S. J., Wang, X. C., et al. (2015). Global transcriptome and gene regulation network for secondary metabolite biosynthesis of tea plant (*Camellia sinensis*). *BMC Genomics* 16:560. doi: 10.1186/s12864-015-1773-0
- Li, J., Lv, X., Wang, L., Qiu, Z., Song, X., Lin, J., et al. (2017). Transcriptome analysis reveals the accumulation mechanism of anthocyanins in 'Zijuan' tea (*Camellia sinensis* var. *assamica* (Masters) kitamura) leaves. *Plant Growth Regul.* 81, 51–61. doi: 10.1007/s10725-016-0183-x
- Liew, A. W.-C., Law, N.-F., and Yan, H. (2010). Missing value imputation for gene expression data: computational techniques to recover missing data from available information. *Brief. Bioinform.* 12, 498–513. doi: 10.1093/bib/bbq080
- Lin, J., Wilson, I. W., Ge, G., Sun, G., Xie, F., Yang, Y., et al. (2017). Whole transcriptome analysis of three leaf stages in two cultivars and one of their F₁ hybrid of *Camellia sinensis* L. with differing EGCG content. *Tree Genet. Genomes* 13:13. doi: 10.1007/s11295-016-1089-5
- Lisso, J., Steinhauser, D., Altmann, T., Kopka, J., and Müssig, C. (2005). Identification of brassinosteroid-related genes by means of transcript co-response analyses. *Nucleic Acids Res.* 33, 2685–2696. doi: 10.1093/nar/gki566
- Lohse, M., Bolger, A. M., Nagel, A., Fernie, A. R., Lunn, J. E., Stitt, M., et al. (2012). RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* 40, W622–W627. doi: 10.1093/nar/gks540
- Lu, Y., Zhang, J., Wan, X., Long, M., Li, D., Lei, P., et al. (2011). Intestinal transport of pure theanine and theanine in green tea extract: green tea components inhibit theanine absorption and promote theanine excretion. *Food Chem.* 125, 277–281. doi: 10.1016/j.foodchem.2010.09.027
- Luo, J. (2015). Metabolite-based genome-wide association studies in plants. *Curr. Opin. Plant Biol.* 24, 31–38. doi: 10.1016/j.pbi.2015.01.006
- Malone, J. H., and Oliver, B. (2011). Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol.* 9:34. doi: 10.1186/1741-7007-9-34
- Mercke, P., Kappers, I. F., Verstappen, F. W., Vorst, O., Dicke, M., and Bouwmeester, H. J. (2004). Combined transcript and metabolite analysis reveals genes involved in spider mite induced volatile formation in cucumber plants. *Plant Physiol.* 135, 2012–2024. doi: 10.1104/pp.104.048116
- Moco, S., Bino, R. J., Vorst, O., Verhoeven, H. A., de Groot, J., van Beek, T. A., et al. (2006). A liquid chromatography-mass spectrometry-based metabolome database for tomato. *Plant Physiol.* 141, 1205–1218. doi: 10.1104/pp.106.078428
- Movahedi, S., Van Bel, M., Heyndrickx, K. S., and Vandepoele, K. (2012). Comparative co-expression analysis in plant biology. *Plant Cell Environ.* 35, 1787–1798. doi: 10.1111/j.1365-3040.2012.02517.x
- Mutwil, M., Klie, S., Tohge, T., Giorgi, F. M., Wilkins, O., Campbell, M. M., et al. (2011). PlaNet: combined sequence and expression comparisons across plant networks derived from seven species. *Plant Cell* 23, 895–910. doi: 10.1105/tpc.111.083667
- Mutwil, M., Øbro, J., Willats, W. G., and Persson, S. (2008). GeneCAT—novel webtools that combine BLAST and co-expression analyses. *Nucleic Acids Res.* 36, W320–W326. doi: 10.1093/nar/gkn292
- Narukawa, M., Toda, Y., Nakagita, T., Hayashi, Y., and Misaka, T. (2014). L-Theanine elicits umami taste via the T1R1 + T1R3 umami taste receptor. *Amino Acids* 46, 1583–1587. doi: 10.1007/s00726-014-1713-3
- Neuweger, H., Albaum, S. P., Dondrup, M., Persicke, M., Watt, T., Niehaus, K., et al. (2008). MeltDB: a software platform for the analysis and integration of metabolomics experiment data. *Bioinformatics* 24, 2726–2732. doi: 10.1093/bioinformatics/btn452
- Obayashi, T., Hayashi, S., Saeki, M., Ohta, H., and Kinoshita, K. (2008). ATTED-II provides coexpressed gene networks for Arabidopsis. *Nucleic Acids Res.* 37, D987–D991. doi: 10.1093/nar/gkn807
- Obayashi, T., Nishida, K., Kasahara, K., and Kinoshita, K. (2011). ATTED-II updates: condition-specific gene coexpression to extend coexpression analyses and applications to a broad range of flowering plants. *Plant Cell Physiol.* 52, 213–219. doi: 10.1093/pcp/pcq203
- Ogata, Y., Sakurai, N., Suzuki, H., Aoki, K., Saito, K., and Shibata, D. (2009). The prediction of local modular structures in a co-expression network based on gene expression datasets. *Genome Inform.* 23, 117–127. doi: 10.1142/9781848165632_0011
- Okada, Y., Kosiki, M., and Chu, M. (2006). Protein and cDNA sequences of two theanine synthetases from *Camellia sinensis*. Japan Patent No. 2006254780. Tokyo: Japan Patent Office.
- Oksman-Caldentey, K. M., Inze, D., and Oresic, M. (2004). Connecting genes to metabolites by a systems biology approach. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9949–9950. doi: 10.1073/pnas.0403636101
- Oliver, S. (2000). Guilt-by-association goes global. *Nature* 403, 601–603. doi: 10.1038/35001165
- Ozsolak, F., and Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* 12, 87–98. doi: 10.1038/nrg2934
- Pan, H.-B., Zhang, D., Li, B., Wu, Y.-Y., and Tu, Y.-Y. (2017). A rapid UPLC method for simultaneous analysis of caffeine and 13 index polyphenols in black tea. *J. Chromatogr. Sci.* 55, 495–496. doi: 10.1093/chromsci/bmw197
- Park, J.-S., Kim, J.-B., Hahn, B.-S., Kim, K.-H., Ha, S.-H., Kim, J.-B., et al. (2004). EST analysis of genes involved in secondary metabolism in *Camellia sinensis* (tea), using suppression subtractive hybridization. *Plant Sci.* 166, 953–961. doi: 10.1016/j.plantsci.2003.12.010
- Pauli, E. D., Scarminio, I. S., and Tauler, R. (2016). Analytical investigation of secondary metabolites extracted from *Camellia sinensis* L. leaves using a HPLC-DAD-ESI/MS data fusion strategy and chemometric methods. *J. Chemometr.* 30, 75–85. doi: 10.1002/cem.2772
- Piotrowski, M., Schemenewitz, A., Lopukhina, A., Müller, A., Janowitz, T., Weiler, E. W., et al. (2004). Desulfoglucosinolate sulfotransferases from *Arabidopsis thaliana* catalyze the final step in the biosynthesis of the glucosinolate core structure. *J. Biol. Chem.* 279, 50717–50725. doi: 10.1074/jbc.M407681200
- Ravasz, E., and Barabási, A.-L. (2003). Hierarchical organization in complex networks. *Phys. Rev. E* 67:026112. doi: 10.1103/PhysRevE.67.026112
- Ruprecht, C., and Persson, S. (2012). Co-expression of cell-wall related genes: new tools and insights. *Front. Plant Sci.* 3:83. doi: 10.3389/fpls.2012.00083
- Ryu, H. W., Yuk, H. J., An, J. H., Kim, D. Y., Song, H. H., and Oh, S. R. (2017). Comparison of secondary metabolite changes in *Camellia sinensis* leaves depending on the growth stage. *Food Control* 73, 916–921. doi: 10.1016/j.foodcont.2016.10.017
- Saito, K., Hirai, M. Y., and Yonekura-Sakakibara, K. (2008). Decoding genes with coexpression networks and metabolomics - 'majority report by precogs'. *Trends Plant Sci.* 13, 36–43. doi: 10.1016/j.tplants.2007.10.006
- Saito, K., and Matsuda, F. (2010). Metabolomics for functional genomics, systems biology, and biotechnology. *Annu. Rev. Plant Biol.* 61, 463–489. doi: 10.1146/annurev.arplant.043008.092035
- Sakuraba, Y., Kim, Y.-S., Han, S.-H., Lee, B.-D., and Paek, N.-C. (2015). The Arabidopsis transcription factor NAC016 promotes drought stress responses by repressing *AREB1* transcription through a trifurcate feed-forward regulatory loop involving NAP. *Plant Cell* 27, 1771–1787. doi: 10.1105/tpc.15.00222
- Sasaoka, K., Kito, M., and Onishi, Y. (1965). Some properties of the theanine synthesizing enzyme in tea seedlings. *Agric. Biol. Chem.* 29, 984–988. doi: 10.1080/00021369.1965.10858501
- Savoi, S., Wong, D. C., Arapitsas, P., Miculan, M., Buchetti, B., Peterlunger, E., et al. (2016). Transcriptome and metabolite profiling reveals that prolonged drought modulates the phenylpropanoid and terpenoid pathway in white grapes (*Vitis vinifera* L.). *BMC Plant Biol.* 16:67. doi: 10.1186/s12870-016-0760-1
- Schaefer, R. J., Michno, J.-M., and Myers, C. L. (2017). Unraveling gene function in agricultural species using gene co-expression networks. *Biochim. Biophys. Acta* 1860, 53–63. doi: 10.1016/j.bbagr.2016.07.016
- Scherer, A. (2009). *Batch Effects and Noise in Microarray Experiments: Sources and Solutions*. Hoboken, NJ: John Wiley & Sons. doi: 10.1002/9780470685983
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Shi, C.-Y., Yang, H., Wei, C.-L., Yu, O., Zhang, Z.-Z., Jiang, C.-J., et al. (2011). Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. *BMC Genomics* 12:131. doi: 10.1186/1471-2164-12-131
- Singh, K., Kumar, S., Yadav, S. K., and Ahuja, P. S. (2009). Characterization of dihydroflavonol 4-reductase cDNA in tea [*Camellia sinensis* (L.) O. Kuntze]. *Plant Biotechnol. Rep.* 3, 95–101. doi: 10.1007/s11816-008-0079-y

- Steinhauser, D., Usadel, B., Luedemann, A., Thimm, O., and Kopka, J. (2004). CSB: DB: a comprehensive systems-biology database. *Bioinformatics* 20, 3647–3651. doi: 10.1093/bioinformatics/bth398
- Stuart, J. M., Segal, E., Koller, D., and Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302, 249–255. doi: 10.1126/science.1087447
- Sugimoto, S., Chi, G., Kato, Y., Nakamura, S., Matsuda, H., and Yoshikawa, M. (2009). Medicinal flowers. XXVI. Structures of acylated oleanane-type triterpene oligoglycosides, yuchasaponins A, B, C, and D, from the flower buds of *Camellia oleifera*—gastroprotective, aldose reductase inhibitory, and radical scavenging effects—. *Chem. Pharm. Bull.* 57, 269–275. doi: 10.1248/cpb.57.269
- Sur, P., Chaudhuri, T., Vedasiromoni, J., Gomes, A., and Ganguly, D. (2001). Antiinflammatory and antioxidant property of saponins of tea [*Camellia sinensis* (L) O. Kuntze] root extract. *Phytother. Res.* 15, 174–176. doi: 10.1002/ptr.696
- Tai, Y., Wei, C., Yang, H., Zhang, L., Chen, Q., Deng, W., et al. (2015). Transcriptomic and phytochemical analysis of the biosynthesis of characteristic constituents in tea (*Camellia sinensis*) compared with oil tea (*Camellia oleifera*). *BMC Plant Biol.* 15:190. doi: 10.1186/s12870-015-0574-6
- Tierney, L., Linde, J., Müller, S., Brunke, S., Molina, J. C., Hube, B., et al. (2012). An interspecies regulatory network inferred from simultaneous RNA-seq of *Candida albicans* invading innate immune cells. *Front. Microbiol.* 3:85. doi: 10.3389/fmicb.2012.00085
- Tipoe, G. L., Leung, T.-M., Hung, M.-W., and Fung, M.-L. (2007). Green tea polyphenols as an anti-oxidant and anti-inflammatory agent for cardiovascular protection. *Cardiovasc. Hematol. Disord. Drug Targets* 7, 135–144. doi: 10.2174/187152907780830905
- Tokimatsu, T., Sakurai, N., Suzuki, H., Ohta, H., Nishitani, K., Koyama, T., et al. (2005). KaPPA-view. A web-based analysis tool for integration of transcript and metabolite data on plant metabolic pathway maps. *Plant Physiol.* 138, 1289–1300. doi: 10.1104/pp.105.060525
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016
- Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F. M., Bassel, G. W., Tanimoto, M., et al. (2009). Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant Cell Environ.* 32, 1633–1651. doi: 10.1111/j.1365-3040.2009.02040.x
- Verpoorte, R., and Memelink, J. (2002). Engineering secondary metabolite production in plants. *Curr. Opin. Biotechnol.* 13, 181–187. doi: 10.1016/S0958-1669(02)00308-7
- Wang, W., Xin, H., Mingle Wang, Q. M., Wang, L., Kaleri, N. A., Wang, Y., et al. (2016). Transcriptomic analysis reveals the molecular mechanisms of drought-stress-induced decreases in *Camellia sinensis* leaf quality. *Front. Plant Sci.* 7:385. doi: 10.3389/fpls.2016.00385
- Wang, Y. N., Tang, L., Hou, Y., Wang, P., Yang, H., and Wei, C. L. (2016). Differential transcriptome analysis of leaves of tea plant (*Camellia sinensis*) provides comprehensive insights into the defense responses to *Ectropis obliqua* attack using RNA-Seq. *Funct. Integr. Genomics* 16, 383–398. doi: 10.1007/s10142-016-0491-2
- Wang, Y.-S., Xu, Y.-J., Gao, L.-P., Yu, O., Wang, X.-Z., He, X.-J., et al. (2014). Functional analysis of Flavonoid 3', 5'-hydroxylase from Tea plant (*Camellia sinensis*): critical role in the accumulation of catechins. *BMC Plant Biol.* 14:347. doi: 10.1186/s12870-014-0347-7
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484
- Weerawatanakorn, M., Lee, Y.-L., Tsai, C.-Y., Lai, C.-S., Wan, X., Ho, C.-T., et al. (2015). Protective effect of theaflavin-enriched black tea extracts against dimethylnitrosamine-induced liver fibrosis in rats. *Food Funct.* 6, 1832–1840. doi: 10.1039/c5fo00126a
- Wei, H., Persson, S., Mehta, T., Srinivasainagendra, V., Chen, L., Page, G. P., et al. (2006). Transcriptional coordination of the metabolic network in *Arabidopsis*. *Plant Physiol.* 142, 762–774. doi: 10.1104/pp.106.080358
- Wille, A., Zimmermann, P., Vranová, E., Fűrholz, A., Laule, O., Bleuler, S., et al. (2004). Sparse graphical Gaussian modeling of the isoprenoid gene network in *Arabidopsis thaliana*. *Genome Biol.* 5:R92. doi: 10.1186/gb-2004-5-11-r92
- Wu, Z.-J., Li, X.-H., Liu, Z.-W., Xu, Z.-S., and Zhuang, J. (2014). De novo assembly and transcriptome characterization: novel insights into catechins biosynthesis in *Camellia sinensis*. *BMC Plant Biol.* 14:277. doi: 10.1186/s12870-014-0277-4
- Wurtele, E. S., Chappell, J., Jones, A. D., Celiz, M. D., Ransom, N., Hur, M., et al. (2012). Medicinal plants: a public resource for metabolomics and hypothesis development. *Metabolites* 2, 1031–1059. doi: 10.3390/metabo2041031
- Xia, E.-H., Zhang, H.-B., Sheng, J., Li, K., Zhang, Q.-J., Kim, C., et al. (2017). The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. *Mol. Plant* 10, 866–877. doi: 10.1016/j.molp.2017.04.002
- Yamamoto, T., Juneja, L. R., and Kim, M. (1997). *Chemistry and Applications of Green Tea*. Boca Raton, FL: CRC press.
- Yang, Z., Baldermann, S., and Watanabe, N. (2013). Recent studies of the volatile compounds in tea. *Food Res. Int.* 53, 585–599. doi: 10.1016/j.foodres.2013.02.011
- Zhang, S., Yue, Y., Sheng, L., Wu, Y., Fan, G., Li, A., et al. (2013). PASmiR: a literature-curated database for miRNA molecular regulation in plant response to abiotic stress. *BMC Plant Biol.* 13:33. doi: 10.1186/1471-2229-13-33
- Zhang, X., Liu, Y., Gao, K., Zhao, L., Liu, L., Wang, Y., et al. (2012). Characterisation of anthocyanidin reductase from Shuchazao green tea. *J. Sci. Food Agric.* 92, 1533–1539. doi: 10.1002/jsfa.4739
- Zhang, Z.-Z., Li, Y.-B., Qi, L., and Wan, X.-C. (2006). Antifungal activities of major tea leaf volatile constituents toward *Colletorichum camelliae* Masea. *J. Agric. Food Chem.* 54, 3936–3940. doi: 10.1021/jf060017m
- Zhou, T.-S., Zhou, R., Yu, Y.-B., Xiao, Y., Li, D.-H., Xiao, B., et al. (2016). Cloning and characterization of a flavonoid 3'-hydroxylase gene from tea plant (*Camellia sinensis*). *Int. J. Mol. Sci.* 17:261. doi: 10.3390/ijms17020261

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Zhang, Zhang, Tai, Wang, Ho and Wan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.