J Biosci (2020)45:129
DOI: 10.1007/s12038-020-00101-x

A predicted protein functional network aids in novel gene mining for characteristic secondary metabolites in tea plant (*Camellia sinensis*)

Shihua Zhang[1,†], Yong Ma[2,†], Rui Zhang[2], Xiaolong He[3], Ying Chen[3], Jingke Du[3], Chi-tang Ho[4], Youhua Zhang[2], Guomin Han[5]* and Xiaoyi Hu[6]*

[1]*College of Life Science and Health, Wuhan University of Science and Technology, Wuhan, China*

[2]*College of Information and Computer Science, Anhui Agricultural University, Hefei, China*

[3]*State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei, China*

[4]*Department of Food Science, Rutgers University, New Brunswick, NJ 08901, USA*

[5]*College of Life Science, Anhui Agricultural University, Hefei, China*

[6]*School of Forestry and Landscape Architecture, Anhui Agricultural University, Hefei, China*

**Corresponding authors (Emails, guominhan@ahau.edu.cn; hxytong@126.com)*

[†]*These authors have contributed equally to this work.*

MS received 15 November 2019; accepted 22 September 2020

Modeling a protein functional network in concerned species is an efficient approach for identifying novel genes in certain biological pathways. Tea plant (*Camellia sinensis*) is an important commercial crop abundant in numerous characteristic secondary metabolites (e.g., polyphenols, alkaloids, alkaloids) that confer tea quality and health benefits. Decoding novel genes responsible for tea characteristic components is an important basis for applied genetic improvement and metabolic engineering. Herein, a high-quality *p*rotein functio*n*al *n*etwork for *tea* plant (TeaPoN) was predicted using cross-species protein functional associations transferring and integration combined with a stringent biological network criterion control. TeaPoN contained 31,273 non-redundant functional interactions among 6,634 tea proteins (or genes), with general network topological properties such as scale-free and small-world. We revealed the modular organization of genes related to the major three tea characteristic components (theanine, caffeine, catechin) in TeaPoN, which served as strong evidence for the utility of TeaPoN in novel gene mining. Importantly, several case studies regarding gene identification for tea characteristic components were presented. To aid in the use of TeaPoN, a concise web interface for data deposit and novel gene screening was developed (*http://teapon.wchoda.com*). We believe that TeaPoN will serve as a useful platform for functional genomics studies associated with characteristic secondary metabolites in tea plant.

Keywords. Characteristic secondary metabolites; novel gene mining; protein functional network; tea plant

Electronic supplementary material: The online version of this article (https://doi.org/10.1007/s12038-020-00101-x) contains supplementary material, which is available to authorized users.

http://www.ias.ac.in/jbiosci
Published online: 15 October 2020

1. Introduction

Tea plant (*Camellia sinensis*), a member plant of *Theaceae* family, is an important perennial, evergreen, woody commercial crop that is widely cultivated in China, India, Kenya, and many other countries (Zhang *et al.* 2018). It is well known that leaves of tea plant can be used as source materials for a popular non-alcoholic beverage known as "tea" due to its numerous characteristic secondary metabolites, such as polyphenols, alkaloids, theanine, polysaccharides, vitamins, minerals, and volatile oils (Zhang *et al.* 2016). However, tea plant is a non-model plant with big genome and complicated genetic background (Shi 2011); its fundamental biology research is lagging far behind the model crop rice (Mao *et al.* 2017) and even many other non-model crops, such as sorghum (R Cantoro *et al.* 2013), tomato (Thirumalaikumar *et al.* 2017) and wheat (Xin *et al.* 2013). A urgently-to-be-resolved issue is that many of tea characteristic component related genes, including catalyzing enzymes, signal proteins, and regulatory transcription factors (TFs), have not been functionally characterized, which inevitably lays an obstacle for the applied genetic improvement and metabolic engineering in tea plant.

With the development of crop functional genomics, many computational strategies (e.g., genome-wide association study, feature-based classification) have been applied for decoding key genes responsible for important agricultural traits, such as stress resistance, edible quality, and yield (Kryukov and Gladyshev 2002; Milenkovic *et al.* 2010; Visscher *et al.* 2017). Among these, network-aided gene screening has been proved as an efficient approach, particularly in the identification of novel genes associated with plant-specialized (secondary) metabolism, by using a simple assumption that genes (or proteins) in a secondary metabolic pathway usually functionally interact to achieve a metabolite accumulative phenotype (Higashi and Saito 2013). In different forms of molecular networks, *p*rotein functi*o*nal *n*etwork (PoN) of a species is widely used and logically accepted because it considers protein as a basic functional unit in actual cellular context (Shim *et al.* 2017). In a PoN, two network elements, nodes and edges, denote proteins and protein-to-protein functional interactions, respectively. Luckily, a high-quality reference genomic map of tea plant is now present after our collaborative efforts of more than ten years (Wei *et al.* 2018), which provides a good opportunity for the construction of a PoN for tea plant and the possible application in novel gene screening of tea characteristic secondary metabolites.

To provide a useful platform for the above-focused issues, we inferred a high-quality *p*rotein functi*o*nal *n*etwork for *tea* plant (TeaPoN) based on a computational pipeline of transferring, integration, and trimming of orthologous protein functional associations from six data-rich model species to tea plant. TeaPoN contained 31,273 functional associations among 6,634 tea proteins, with a giant connected component of 6,143 proteins and an array of small components of a few proteins. The utility of TeaPoN in novel gene mining was confirmed with several case studies. In addition, a web interface was developed for this promising application. In the near future, TeaPoN might act as a central resource in tea biochemistry and molecular biology community.

2. Materials and methods

2.1 *Protein functional association prediction for tea plant using an associalog method*

The total 33,932 deduced protein sequences of *Camellia sinensis* var. *sinensis* (CSS) were retrieved from our international tea plant genome sequencing consortium (*http://pcsb.ahau.edu.cn:8080/CSS*) (Wei *et al.* 2018). With these protein sequences, we used a web server JiffyNet to predict protein functional associations for tea plant (Kim *et al.* 2013). JiffyNet applies an associalog method that can transfer orthologous protein functional associations from data-rich model species (template species) to a species of interest (query species) with its protein sequences available. It is noted that JiffyNet extends protein-to-protein physical interactions to general protein functional interactions, thus the predicted protein functional associations (include protein-to-protein physical and indirect interactions) are more complete than those based on only protein physical interaction (using interolog method). This is logically supported by a fact that two proteins may function together in the same biological pathway with no direct physical interactions (Creighton 1993). In our study, six template species genome-scale protein functional association networks, including EcoliNet, YeastNet, WormNet, HumanNet, AraNet and RiceNet, were chosen for an integration of tea protein functional associations after transferring the associations from *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Homo sapiens*, *Arabidopsis thaliana* and *Oryza sativa*, respectively.

2.2 *Protein functional association redundancy filtering*

Protein functional associations of tea plant can be represented as an undirected graph (i.e., biological network) that is different from directed transcriptional regulatory network or metabolic network (Barabasi and Oltvai 2004). As a typical biological molecular association network, redundant associations (interactions) should be firstly filtered. That is to say, the functional association that protein A is related to protein B (in A-B form) is equivalent to the functional association that protein B is related to protein A (in B-A form) in a undirected network, and in such case, either A-B or B-A form is retained in the filtering process. Based on this principle, we developed in-house Python scripts to achieve the redundancy filtering for tea plant.

2.3 *Construction of TeaPoN using stringent biological network criterions*

For each of the above-filtered protein functional associations in tea plant, adjusted log likelihood (LLS) score was transferred from the template network as an edge weight score that can be used as confidence score for the associations. Thus, these protein functional associations can be represented using a undirected weighted network which was thereafter called as *p*rotein functi*o*nal *n*etwork for *tea* plant (TeaPoN). In the network, node denoted protein, and a link was placed between two tea proteins indicating their functional association. To our knowledge, biological networks are mostly characterized as scale-free, and in a biological network, modular structure with high network density usually reveals actual cellular functional organization (Barabasi and Oltvai 2004). Therefore, these two network topological properties can be used as stringent biological network criterions for the rational selection of a adjusted LLS threshold in the refining of TeaPoN (Romero-Campero *et al.* 2016). To achieve this, a range of adjusted LLS threshold were considered to generate a family of tea protein functional networks. For these member networks, we can estimate how well an individual network satisfies a scale-free property using the model fitting index R^2 of the linear regression for the logarithmic transform of the node degree distribution where the degree of a node represents the number of nodes linked to this node. In addition, we applied average node degree for these individual networks to measure the network density that is indicative of the modular property of a network.

2.4 *Modular characterization of characteristic secondary metabolites related genes in TeaPoN*

It is known that functionally related genes may work together in a functional module, such as protein complex, pathway and cellular organelle, to perform a desired cellular function (Romanov *et al.* 2019). Therefore, the modular organization of functional related gene in a network can be used to identify novel genes involved in a biological pathway using a 'guilt-by-association' approach described in (Wolfe *et al.* 2005). We used the two modular measures, dyadicity (D) and heterophilicity (H), proposed by (Park and Barabási 2007), to quantify the modular properties of secondary secondary metabolites related genes in TeaPoN. Dyadicity evaluates the enrichment of links between nodes with a common property over the number expected if the property were distributed randomly on the network. Heterophilicity is quantization of the tendency of nodes to connect with other nodes sharing a common property. In general, the findings of $D>1$ and $H<1$ can be used to as an evidence that genes involved in a secondary metabolic pathway have a clear clustering tendency in the network, which then can be considered as a prerequisite for the next novel gene mining in concerned pathways of characteristic components. In TeaPoN, genes (proteins) in a certain metabolic pathway, such as theanine, caffeine, and catechins, were regarded to have the common property. For the computation of D and H values, known genes in a metabolic pathway were expert-curated. The interaction tendency among these genes (or with other genes) in the pathway was computed and statistically evaluated using make-up pathway as random control, where the same number of genes as in the real pathway was randomly selected from the whole TeaPoN.

2.5 *Development of a web interface for novel gene mining of tea characteristic secondary metabolites*

Our website for novel gene mining of tea characteristic secondary metabolites was implemented in a free and open source Python Web framework, Flask (*http://flask.pocoo.org*), with a popular relational database management system, MySQL (*https://www.mysql.com*) as the backend database. Its frontend pages are generated via HTML5, CSS3, jQuery (*http://jquery.com*), Bootstrap (*https://getbootstrap.com*), and DataTables (*https://datatables.net*). Protein functional clusters were visualized by vis.js (*http://visjs.org*).

3. Results

3.1 *Overview of protein functional associations in tea plant*

As indicated in Methods, the template species chosen for the prediction of TeaPoN covered species of bacteria, fungi, animals and plants, which allowed a completion of high-confident network modeling. We adopted a parallel orthologous protein functional associations transferring and the follow-up integration to facilitate a high-quality protein functional network for tea plant. As shown in figure 1, we parallelly transferred 31,431 functional associations among 1,935 tea proteins from *Escherichia coli*, 133,837 functional associations among 5,462 tea proteins from *Saccharomyces cerevisiae*, 417,167 functional associations among 5,375 tea proteins from *Caenorhabditis elegans*, 278,203 functional associations among 5,925 tea proteins from *Homo sapiens*, 197,979 functional associations among 8,028 tea proteins from *Arabidopsis thaliana*, and 187,460 functional associations among 9,458 tea proteins from *Oryza sativa*. In this transferring process, the evolutionarily close model plants *Arabidopsis thaliana* and *Oryza sativa* contributed mostly in the number of tea proteins despite of a moderate number of functional associations. It is noted that *Escherichia coli* contributed the least number of both proteins and functional associations. In the next integration process, we pooled all the transferred protein functional associations from the six template species and filtered the redundant associations using the association redundancy filtering criterion described in Methods. Finally, we identified 1,068,875 functional associations among 16,446 proteins for tea plant. In the total functional associations, 655,109 (61.28%) had orthologous protein functional associations in only one template species, and 145,961, 53,998, 23,958, 8,142

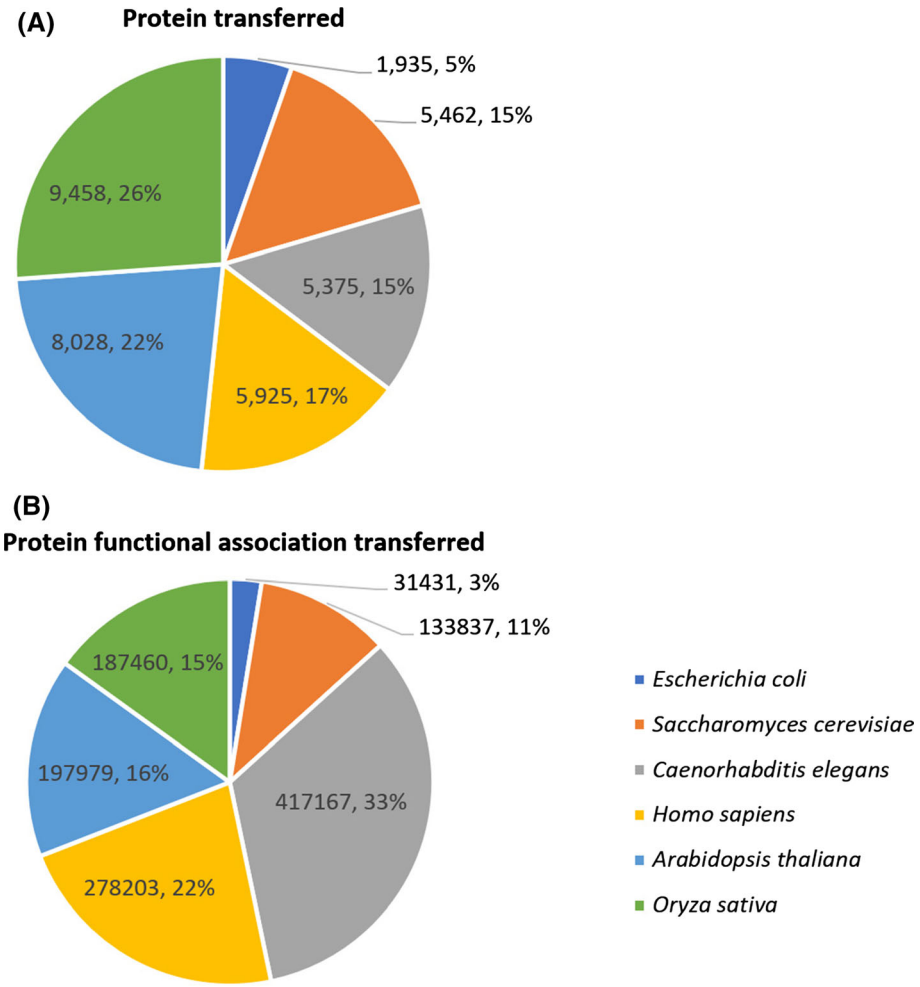

Figure 1. Distribution of cross-species proteins and protein functional associations transferred. Different colors in the two pie charts were used to denote different species from which proteins **(A)** and their corresponding protein functional associations **(B)** were transferred.

and 85 (39.72% in total) had functional associations in two, three, four, five, and six template species, respectively (figure 2), indicating a conserved protein functional association profile among tea plant and the six well-established model species.

3.2 *Topological properties of constructed TeaPoN*

As described in Methods, the constructed TeaPoN with the above-obtained protein functional associations in tea plant was highly dense and did not conform to the general topological properties of actual biological network, such as scale-free and modularity (Zhang *et al.* 2009). Thus, using the stringent biological network criterions, we measured the scale-free and modular properties of a family of tea protein functional networks based on a wide range of adjusted LLS thresholds with interval of 0.1. As indicated in figure 3, with the increasing LLS threshold, the network density (i.e. average node connectivity, node connectivity is quantified by the number of edges linked to a node) gently decreased whereas the scale-free model fit (R^2) tardily increased. According to the strategy described by (FJ Romero-Campero *et al.* 2016), we arrived a compromise between the establishment of a scale-free network and a high density network when the adjusted LLS threshold was selected as 2.700. At this LLS threshold, the scale-free model fit exhibited a maximal R^2 value of 0.9332 with a moderate average node connectivity (degree) of 9.428. We called this network generated using this threshold as the final and refined TeaPoN.

TeaPoN contained 31,273 functional associations among 6,634 tea proteins, with a giant connected component of 6,143 proteins (92.6% of the entire network) and a batch of small components with a few proteins (supplementary figure 1). This global network connectivity of TeaPoN is consistent with previous constructed protein networks in model and non-model plants (Ding *et al.* 2014; Gu *et al.* 2011). As shown in figure 4A, the prominent scale-free property of TeaPoN was confirmed by the finding that it had a degree-distribution of power-law (AL BarabÃ si and E Bonabeau 2003); that is to say, most protein nodes in TeaPoN had low degrees, whereas a few protein nodes had very high degrees. Among the total 6,634 tea proteins, only 165 (2.48%, act as "hubs") have more than 100 functional associations with other tea proteins, suggesting that these few highly connected hubs held the network together. Of the top 20 hub protein nodes, we noticed that the majority of them belong to the ribosomal protein family, indicating the important role of ribosomal proteins in the whole tea protein functional interactome (supplementary table 1). Figure 4B showed the shortest path length distribution of TeaPoN, where the majority of node pairs have a shortest length of 1 to 11, a mean shortest-path length of 5.502. Therefore, any two tea proteins in TeaPoN can be

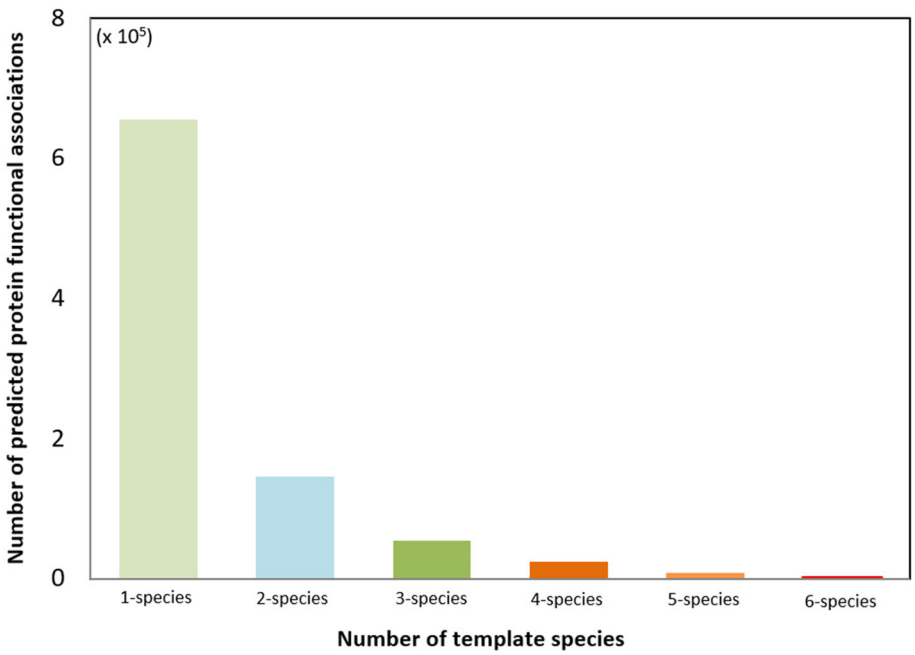

Figure 2. Enrichment of the predicted protein functional associations on data-rich template species. It is noted that the majority (61.28%) of the protein functional associations in tea plant was inferred from only one orthologous model species.

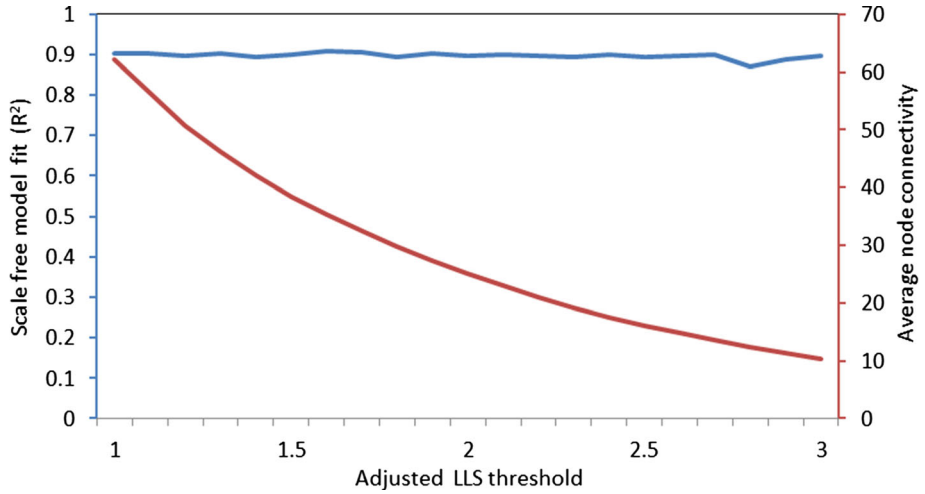

Figure 3. Network density and scale-free model fit (R^2) of TeaPoN based on adjusted LLS threshold. In this plot, the abscissa denoted adjusted LLS threshold from 1 to 3. The left-blue ordinate denoted scale-free model fit (R^2), and the right-red ordinate denoted network density (average node connectivity) of a resulted network generating from a certain adjusted LLS threshold.

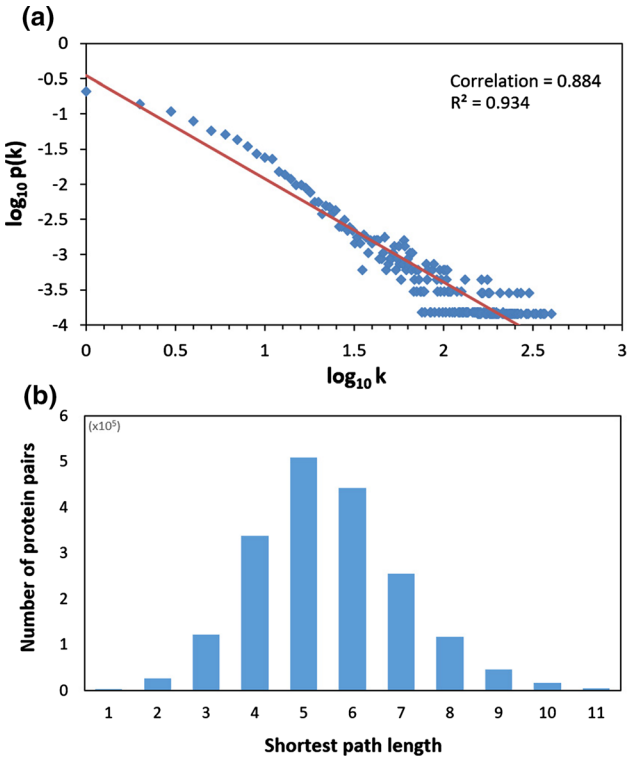

Figure 4. Topological properties of TeaPoN. Double logarithmic coordinates were used for quantization of node degree and the corresponding occurring probability **(A)** and the distribution of the number of node pairs with the shortest path length ranging from 1 to 11 **(B)**.

reached to each other with a path of only a few links. All these findings showed that TeaPoN has a small-world property (MEJ Newman and DJ Watts 1999).

3.3 *Modular nature of characteristic secondary metabolite related genes in TeaPoN*

In tea plant, theanine, caffeine and catechins are actively studied and considered as the three major characteristic secondary metabolites. Here, the modular property (clustering tendency in a network) of their related genes was evaluated using both the dyadicity (*D*) and heterophilicity (*H*) measures (see Methods). By computation, we found that theanine, caffeine and catechins have clear modular organization in TeaPoN because all the three characteristic components associated genes were dyadic (*D*-value>1, table 1) and the two characteristic components (theanine, caffeine) associated genes were heterophobic (*H*-value<1). Therefore, it could be speculated that the three characteristic components have distinct separation among them in tea metabolic flux origin. In addition, we noted that only catechins showed heterophilic (*H*-value>), indicating that catechins related genes have interactions with other metabolite related genes. It is logically reasonable that the catechins biosynthetic pathway is non-linear (with multiple catalyzing branches) and have cross-talks with others tea characteristic components (e.g., theanine), which are demonstrated in the previous studies (Li *et al.* 2015; Tai *et al.* 2018).

3.4 *Case studies and implementation of a user interface for novel gene mining*

TeaPoN was decomposed using the Cytoscape plugin MCODE (Bader and Hogue 2003), which is designed

Table 1. Dyadicity (*H*) and heterophilicity (*D*) values of the major three characteristic metabolites related genes in TeaPoN

Metabolic pathway	Number of genes*	In-pathway links	Out-pathway links	*D*-Value	*H*-Value
Theanine	24	141	251	24.8534	0.5253
Caffeine	46	495	826	19.583	1.5624
Catechins	269	1593	4653	10.5276	0.7527

*The number of genes involved in a certain characteristic metabolic pathway.

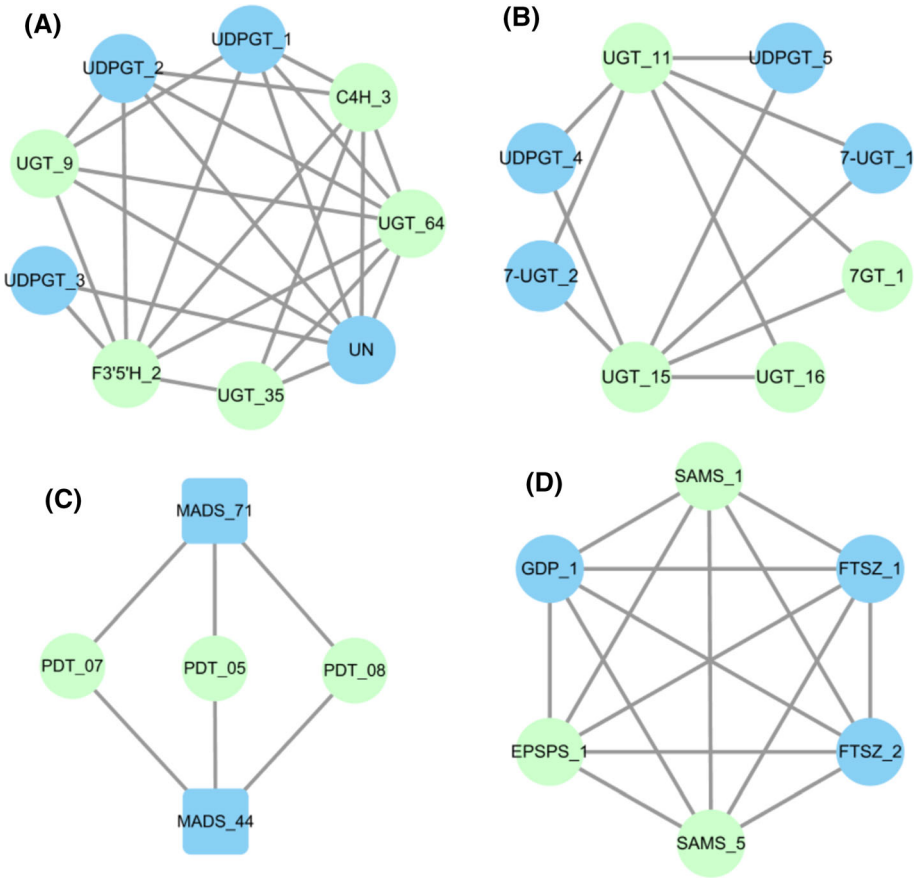

Figure 5. Case studies illustrated in protein functional modules. In these modules, theanine, caffeine, and catechins catalyzing enzyme genes and the possibly involved novel genes were colored in light green, and light blue, respectively. C4H, Cinnamate 4-hydroxylase; EPSPS, 5-enolpyruvylshikimate-3-phosphate synthase; F3′5′H, Flavonoid 3′,5′-hydroxylase; FTSZ, Tubulin/FtsZ family, C-terminal domain; GDP, Probable mannose-1-phosphate guanylyltransferase; 7GT, Flavonol 7-O-glucosyltransferase; MADS, MADS-box transcription factor; PDT, Prephenate dehydratase; SAMS, S-adenosyl-L-methionine synthase; UDPGT, UDP-glucoronosyl and UDP-glucosyl transferase; UGT, UDP-glycosyltransferase; 7-UGT, 7-deoxyloganetin glucosyltransferase; UN, unknown.

based on a graph-based clustering algorithm to extract densely connected subgraphs (i.e., protein functional modules). We re-decomposed all the functional modules with more than 100 proteins into smaller submodules by considering the hierarchical structures of general biological networks (Newman 2006). A total of 264 protein functional modules were finally identified. Figure 5 showed four representative module examples for illustrating novel gene mining for the three major characteristic metabolic pathways. It is noted that genes in a metabolic pathway tend to group together in a module where catalyzing enzyme genes were colored

in light green and other related genes in light blue. In figure 5A and 5B, two enzyme gene family, *UDPGT* and *7-UGT*, were found to be functionally interacted with catechins enzyme genes (*F3′5′H*, *UGT*, *C4H* and *7GT*), suggesting that they are novel enzymes in catechins pathway. As shown in figure 5C, the enzyme genes *PDT*s in catechins pathway were regulated by two *MADS* members (novel TF regulators), which gave clues for the potential transcriptional regulatory mechanisms in this pathway. We also found that the enzyme gene *SAMS* in caffeine pathway and *EPSPS* in catechins pathway group together in a module (figure 5D). It is reasonable because these two pathways holds the cross-talking property demonstrated in the previous studies (Li *et al.* 2015; Tai *et al.* 2018). From another respect, the related genes *GDP* and *FTSZ* in this module may be novel genes involved in both the two pathways as cooperative contributors for metabolite accumulation. Taken together, the predicted TeaPoN could serve as a useful platform for novel gene mining for characteristic components in tea plant. Inspired by these observations, we implemented an easy-to-use web interface for tea researchers for this valuable application. Upon a query by inputting gene of interest, the corresponding module information as well as functional related genes, interacting intensity, and detailed gene descriptions are present for users. We also provided downloading of the network data and tutorial of the use of this resource.

4. Discussion

Decoding genes responsible for important agronomical traits of a crop species is of great importance for the possible molecular breeding and genetic improvement (Brbaklic *et al.* 2013; Leonova and Budashkina 2017). Tea plant (*Camellia sinensis*) is an important commercial crop featuring its abundant characteristic secondary metabolites, such as theanine, polyphenols, alkaloids, polysaccharides, vitamins, volatile oils, and minerals, in tea leaves that serve as materials of a popular healthy drink "tea". However, only a small fraction of tea metabolites related genes have been functionally dissected. To address this gap, we computationally predicted a *p*rotein functi*o*nal *n*etwork for *tea* plant (TeaPoN) and provided a user-friendly web interface that aids in novel gene mining for tea characteristic secondary metabolites.

TeaPoN represented a high-confident protein-to-protein functional associations in tea plant at a genome-scale because we used six template network in data-rich model species covering bacteria, fungi, animals and plants. In addition, network trimming was used based on stringent biological network criterions that made TeaPoN have general biological network characteristics, such as scale-free and small-world. For the three major characteristic components (theanine, caffeine, and catechins), the topological analysis of TeaPoN showed that their related biosynthetic genes had significant modular property and clear separation among them, which provided an evidence as a prerequisite for the identification of genes related to tea characteristic components. We then decomposed TeaPoN into protein functional modules and illustrated several case studies showing the utility of TeaPoN in novel gene mining for our widely-concerned three major characteristic metabolic pathways.

It should be note that TeaPoN only included conserved protein-to-protein functional interactions between model species and tea plant, and thus tea plant specific protein functional interactions were not considered. It is well-known that tea plant is a newly genome-sequenced non-model plant and there did not accumulate data related to its protein interactions with high confidence. Therefore, the completeness of TeaPoN cannot be guaranteed in the current state. To compensate such a disadvantage of TeaPoN and achieve a more comprehensive gene-to-gene interactions (not limit to protein-to-protein interactions), we can consider the use of tea plant specific genomic data (e.g., gene co-expression derived from large samples of mRNA sequencing data, RNA-Seq) and different levels of functional genomics data in model species (e.g., genetic interaction, subcellular co-localization) into an integrated computational frame that is successfully used in model and non-model species (E Kim *et al.* 2017a; H Kim *et al.* 2017b). In addition to the improved prediction strategy, we will consider a development of a more comprehensive database platform instead of a simple web interface for data deposit, distribution, and application in future. We believe these strategies are promising to increase the data integrity and functional availability of TeaPoN, and promote broader interest from researchers concerned on tea functional genomics studies.

Acknowledgments

Funding was provided by National Natural Science Foundation of China (Grant Nos. 31270714 and 31301248).

References

Bader GD and Hogue CW 2003 An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinform.* **4** 2

Barab AL and Bonabeau E 2003 Scale-free networks. *Sci. Am.* **288** 60–69

Barabasi AL and Oltvai ZN 2004 Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5** 101–U115

Brbaklic L, Trkulja D, Kondicspika A and Kobiljski STB 2013 Detection of QTLs for important agronomical traitsin hexaploid wheat using association analysis. *Czech J. Genet. Plant Breed.* **49** 1–8

Cantoro R, Crocco CD, Benecharnold RL and Rodríguez MVJJOEB 2013 In vitro binding of Sorghum bicolor transcription factors ABI4 and ABI5 to a conserved region of a GA 2-OXIDASE promoter: possible role of this interaction in the expression of seed dormancy. *J. Exp. Bot.* **64** 5721–5735

Creighton TE 1993 *Proteins: Structures and Molecular Properties* (W. H. Freeman)

Ding YD, Chang JW, Guo J, Chen DJ, Li S *et al.* 2014 Prediction and functional analysis of the sweet orange protein-protein interaction network. *BMC Plant Biol.* **14** 213

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

Higashi Y and Saito K 2013 Network analysis for gene discovery in plant-specialized metabolism. *Plant Cell Environ.* **36** 1597–1606

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

Kim E, Kim H and Lee I 2013 JiffyNet: a web-based instant protein network modeler for newly sequenced species. *Nucleic Acids Res.* **41** W192–W197

Kim H, Kim BS, Shim JE, Hwang S, Yang S *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

Kryukov GV and Gladyshev VN 2002 Mammalian selenoprotein gene signature: identification and functional analysis of selenoprotein genes using bioinformatics methods. *Methods Enzymol.* **347** 84–100

Leonova IN and Budashkina EB 2017 The study of agronomical traits determining the productivity of the Triticum aestivum/Triticum timopheevii introgression lines with resistance to fungal diseases. *Russ. J. Genet. Appl. Res.* **7** 299–307

Li CF, Zhu Y, Yu Y, Zhao QY, Wang SJ *et al.* 2015 Global transcriptome and gene regulation network for secondary metabolite biosynthesis of tea plant (*Camellia sinensis*). *BMC Genomics* **16** 560

Mao C, Lu S, Lv B, Zhang B, Shen J *et al.* 2017 A rice NAC transcription factor promotes leaf senescence via ABA biosynthesis. *Plant Physiol.* **174** 1747–1763

Milenkovic T, Memisevic V and Ganesan AK 2010 Systems-level cancer gene identification from protein interaction network topology applied to melanogenesis-related functional genomics data. *J. R. Soc. Interface* **7** 423–437

Newman ME 2006 Modularity and community structure in networks. *Proc. Natl. Acad. Sci.* **103** 8577–8582

Newman MEJ and Watts DJ 1999 Renormalization group analysis of the small-world network model. *Phys. Lett. A* **263** 341–346

Park J and Barabási AL 2007 Distribution of node characteristics in complex networks. *Proc. Natl. Acad. Sci. USA* **104** 17916–17920

Romanov N, Kuhn M, Aebersold R, Ori A, Beck M *et al.* 2019 Disentangling genetic and environmental effects on the proteotypes of individuals. *Cell* **177** 1308–1318

Romero-Campero FJ, Ignacio PH, Eva LR, Romero JM and Federico V 2016 ChlamyNET: aChlamydomonasgene co-expression network reveals global properties of the transcriptome and the early setup of key co-expression patterns in the green lineage. *BMC Genomics* **17** 227

Shi CYJBG 2011 Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. *BMC Genomics* **12** 131

Shim JE, Lee T and Lee I 2017 From sequencing data to gene functions: co-functional network approaches. *Anim. Cells Syst.* **21** 77–83

Tai Y, Liu C, Yu S, Hua Y, Sun J *et al.* 2018 Gene co-expression network analysis reveals coordinated regulation of three characteristic secondary biosynthetic pathways in tea plant (*Camellia sinensis*). *BMC Genomics* **19** 616

Thirumalaikumar VP, Devkar V, Mehterov N, Ali S, Ozgur R *et al.* 2017 NAC transcription factor JUNGBRUNNEN1 enhances drought tolerance in tomato. *Plant. Biotechnol. J.* **16** 354–366

Visscher PM, Wray NR, Zhang Q, Sklar P, Mccarthy MI *et al.* 2017 10 Years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* **101** 5–22

Wei C, Yang H, Wang S, Zhao J, Liu C *et al.* 2018 Draft genome sequence of Camellia sinensis var. sinensis provides insights into the evolution of the tea genome and tea quality. *Proc. Natl. Acad. Sci. USA* **115** 201719622

Wolfe CJ, Kohane IS and Butte AJ 2005 Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. *BMC Bioinform.* **6** 227

Xin L, Lihua Y, Xianyao Z, Miaoping Z, Yan L *et al.* 2013 Transgenic wheat expressing *Thinopyrum intermedium* MYB transcription factor TiMYB2R-1 shows enhanced resistance to the take-all disease. *J. Exp. Bot.* **64** 2243–2253

Zhang S, Xuan H, Zhang L, Fu S, Wang Y *et al.* 2016 TBC2health: a database of experimentally validated health-beneficial effects of tea bioactive compounds. *Brief. Bioinform.* **18** 830–836

Zhang S, Zhang L, Tai Y, Wang X, Ho C-T *et al.* 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** https://doi.org/10.3389/fpls.2018.00480

Zhang Z, Lin Y, Gao S, Zhou S, Guan J *et al.* 2009 Trapping in scale-free networks with hierarchical organization of modularity. *Phys. Rev. E* **80** 051120

Corresponding editor: Sreenivas Chavali