BIOINFORMATICS ORIGINAL PAPER

Vol. 30 no. 1 2014, pages 94–103
doi:10.1093/bioinformatics/btt605

Systems biology

Advance Access publication October 22, 2013

Inferring the soybean (*Glycine max*) microRNA functional network based on target gene network

Yungang Xu[1,2], Maozu Guo[1,*], Xiaoyan Liu[1], Chunyu Wang[1] and Yang Liu[1]

[1]School of Computer Science and Technology, and [2]School of Life Science and Technology, Harbin Institute of Technology, Harbin 150001, P.R. China

Associate Editor: Ivo Hofacker

ABSTRACT

Motivation: The rapid accumulation of microRNAs (miRNAs) and experimental evidence for miRNA interactions has ushered in a new area of miRNA research that focuses on network more than individual miRNA interaction, which provides a systematic view of the whole microRNome. So it is a challenge to infer miRNA functional interactions on a system-wide level and further draw a miRNA functional network (miRFN). A few studies have focused on the well-studied human species; however, these methods can neither be extended to other non-model organisms nor take fully into account the information embedded in miRNA–target and target–target interactions. Thus, it is important to develop appropriate methods for inferring the miRNA network of non-model species, such as soybean (*Glycine max*), without such extensive miRNA-phenotype associated data as miRNA-disease associations in human.

Results: Here we propose a new method to measure the functional similarity of miRNAs considering both the site accessibility and the interactive context of target genes in functional gene networks. We further construct the miRFNs of soybean, which is the first study on soybean miRNAs on the network level and the core methods can be easily extended to other species. We found that miRFNs of soybean exhibit a scale-free, small world and modular architecture, with their degrees fit best to power-law and exponential distribution. We also showed that miRNA with high degree tends to interact with those of low degree, which reveals the disassortativity and modularity of miRFNs. Our efforts in this study will be useful to further reveal the soybean miRNA–miRNA and miRNA–gene interactive mechanism on a systematic level.

Availability and implementation: A web tool for information retrieval and analysis of soybean miRFNs and the relevant target functional gene networks can be accessed at SoymiRNet: http://nclab.hit.edu.cn/SoymiRNet.

Contact: maozuguo@hit.edu.cn

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on April 29, 2013; revised on September 3, 2013; accepted on October 17, 2013

1 INTRODUCTION

MicroRNAs (miRNAs), ~22 nt in length, are endogenous small non-coding RNAs that repress gene expression by binding 3′-untranslated regions (3′-UTRs) of their target messenger RNAs (mRNAs), leading to direct destructive cleavage or translational repression. In plant, mature miRNAs are generated from stem-loop regions of longer RNA precursors mainly by an endoribonuclease III-like enzyme, dicer like-1 (DCL1). The processed and methylated miRNA/miRNA* duplex is then exported to the cytosol. MiRNAs that are incorporated into an argonaute protein containing RNA-induced silencing complex (RISC) can affect the target gene expression (Jones-Rhoades *et al.*, 2006; Voinnet, 2009). Different from of animals, most plant miRNAs have highly complementary recognition sites on their targets; it is considered to be a prominent feature of cleavage-guided repression. MiRNAs, one of the most important components of the cell, play critical roles in many important biological processes, such as development, nutrient homeostasis, abiotic stress responses and pathogen responses in plant. So far, miRBase (Griffiths-Jones *et al.*, 2006) (Release 19: August 2012) contains 21 264 entries representing hairpin precursor miRNAs, expressing 25 141 mature miRNA products, in 193 species. It is predicted that in eukaryotes, miRNA genes accounted for ~1% of the total genes and their target genes may reach 10–30%.

The limited miRNAs are claimed to be able to control the larger set of genes or processes through interaction, in which multiple miRNAs work interactively to control individual genes. To date, several elegant experiments have been carried out, unraveling intriguing miRNA interactions. For example, lin-4 and let-7, the first two experimentally verified miRNAs, are cooperative in Drosophila (Enright, *et al.*, 2004). Krek *et al.* (2005) pointed out that there are interactive relations between miRNAs when they proposed the algorithm of PicTar and further experimentally validated that miR-375, miR-124 and let-7b coordinately regulate *Mtpn*. Additionally, numerous miRNAs were identified to be involved in the aging signaling pathway, such as miR-1, miR-21, miR-24, miR-34, miR-100, miR-106, miR-132, miR-145, miR-146, miR-199, miR-206, miR-217, miR-320 and miR-449 [reviewed by Chen *et al.* (2010)]. The same interactions also exist in plant; for example, miR160 and miR167 are involved in *Arabidopsis* adventitious rooting program. All these experimental results bring to a new area of miRNA research that focuses on network rather than individual miRNA interaction. With the availability of large data derived from fine-scale experiments and computer algorithms, several results have been achieved. Shalgi *et al.* (2007) first built an miRNA network by analyzing the target genes predicted by PicTar and TargetScan. Integrated analysis of miRNA expression profiling data showed that the higher correlation coefficient of miRNA co-expressions, the higher intensity in

*To whom correspondence should be addressed.

their synergy; miRNAs with negative correlation between their expressions tend to avoid the synergies. Recently, Xu *et al.* (2011) constructed a human miRNA–miRNA functional synergistic network via co-regulating functional modules. They found that, contrasting to non-disease miRNAs, disease miRNAs have more synergism and are the global central cores of the network, indicating their higher complexity of function.

The prerequisite of constructing an miRNA functional network (miRFN) is to define the functions of the miRNAs. However, the functions of the most recently identified miRNAs are unknown due to the lack of experimental strategy for systematic identification of their regulating targets. To better understand miRNAs, it is increasingly necessary to measure their functional similarity (FS) and further construct a functional network, thus in turn to infer novel potential functions for miRNAs. For the FS of protein-coding genes, many successful methods have been proposed either based on sequence or expression similarities (Horvath *et al.*, 2006; Lin *et al.*, 2007) or based on Gene Ontology (GO) annotations (Lee *et al.*, 2004; Lord *et al.*, 2003; Pesquita *et al.*, 2008, 2009; Wang *et al.*, 2007). Nevertheless, we cannot indiscriminately imitate these successful methods for inferring the FS of miRNAs. First, all miRNAs are short in length and miRNA–mRNA duplex allows mismatch. So miRNAs with similar sequence and pre-miRNAs with similar structure may have distinctive functions. Second, although expression similarity can interpret part of the FS of miRNA, like protein-coding genes, the similar expression pattern does not always signify FS. Third, it is impractical to measure miRNAs FS directly based on their functions due to the absence of functional databases such as GO annotations for coding genes to annotate miRNAs.

Therefore, new methods were developed to measure miRNAs' FS. Wang *et al.* (2010) used human miRNA-disease association data to infer miRNAs' FS by measuring the similarity of their associated diseases structured as directed acyclic graphs, which is similar to infer the similarity of protein-coding genes by measuring the semantic similarity of their annotating GO directed acyclic graphs (Wang *et al.*, 2007). However, few species has datasets about the relationships between miRNAs and specific phenotypes, such as miRNA-diseases associations in human. More generally and directly, miRNAs' function as repressive regulator on their target transcripts. The final functionality of miRNAs that is embedded in the miRNA–target interactions mainly depends on the functions of the downstream target genes. Thus, the target genes should play the major roles in full elucidation of the biological meanings as well as the FSs of the miRNAs. Based on target genes, all existing approaches for miRNA's FS calculation can be summarized as shown in Figure 1A. The simplest method reported by Shalgi *et al.* (2007) defined the FS as the proportion of the common target genes, which is represented as a Jaccard similarity between two target gene sets regulated by two miRNAs (Fig. 1A (a)). Yu *et al.* (2010) presented a second method for systematic study of FSs among human miRNAs by using their target genes' FSs presented as GO semantic similarities [Fig. 1A (b)]. Xu *et al.* (2011) used the third method shown as in Figure 1A (c). For each pairwise miRNAs, they test the significance of their co-regulated genes sharing the same GO category (called as co-regulating functional modules) using hypergeometric test.

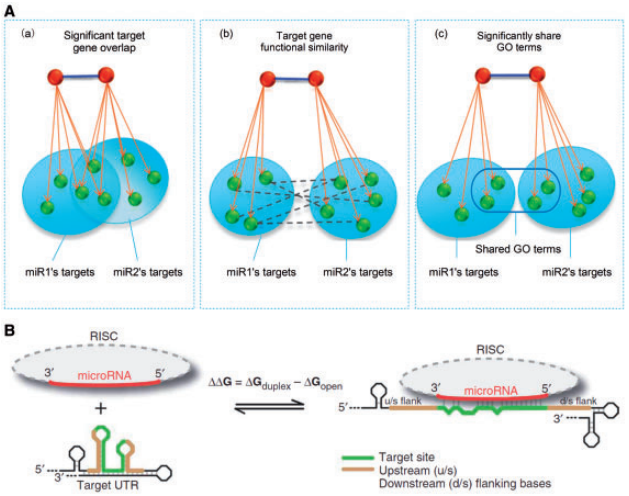

Fig. 1. (**A**) Three approaches for measuring FS of miRNAs. (a) FS was directly calculated as the proportion of the common target genes between two miRNAs, which is represented as a Jaccard similarity between two target gene sets. (b) FS was calculated by computing the semantic similarity between their target gene sets based on GO. (c) FS was represented as the significance of their co-regulated genes sharing the same GO category, which is identified by hypergeometric test. (**B**) Illustration of accessibility, represented as for mRNA–target interaction. The left part shows the unbound RISC and its partial-paired target UTR; the right shows the RISC-target interactions. $\Delta\Delta G$ is computed as the free energy gained by transitioning from the state in which miRNA and target are unbound (left) to the state in which the miRNA binds its target (right). The region of the target site that needs to be unpaired for a miRNA–target interaction includes the miRNA bound region and likely additional flanks. This figure was completely cited from Kertesz *et al.* (2007)

However, all aforementioned methods ignored two critical aspects of information embedded in the miRNA-target regulatory process, which would greatly improve their performances. One is the site accessibility between miRNA and its target mRNA. The other is the interactive context between target genes.

Target accessibility, first proposed by Kertesz *et al.* (2007), has been identified as one of important factors that are involved in target recognition because the secondary structure (stem etc.) around target site will prevent miRNA and target mRNA from contacting. As shown in Figure 1B, the target accessibility is represented by $\Delta\Delta G$, the energy computed as the free energy gained by transitioning from the state in which miRNA and its target are unbound (left) to the state in which the miRNA binds its target (right). The less energy means the more possibility that miRNA is able to contact (and cleave) target mRNA (i.e. high accessibility). From this point of view, an miRNA pair that has less $\Delta\Delta G$ with their common target mRNAs should be more similar in functionality (more detailed illustrations are provided in Supplementary Materials S1 and Supplementary Fig. S1). Thus, we should take into account target accessibility when measuring the FS between miRNAs to improve the sensitivity of functional comparison.

Interactive contexts between miRNAs' target genes have also attracted the attentions of several researchers. Satoh *et al.* (2011) suggested that a set of miRNA targets regulated by a single miRNA generally constitute the biological network of

functionally associated molecules in human cells. Hsu *et al.* (2008) have performed topological analysis to elucidate the global correlation between miRNA regulation and protein–protein interaction network in humans. They showed that target genes of individual miRNA tend to be hubs and bottlenecks in the network. Although proteins directly regulated by miRNA might not form a network module themselves, the miRNA targets and their interacting neighbors jointly showed significantly higher modularity. These studies indicate that target genes constitute not just a simple gene set, but also a subnetwork of a larger scale functional gene network they involved in. The interactive contexts in such target gene network should also be used for inferring the FS of miRNAs.

We propose a novel approach to measuring the FS of miRNAs, considering both their target site accessibility and the topology of target functional gene networks (FGN), and further construct four soybean (*Glycine max*) miRFNs. We by topology analysis indicated that miRFNs of soybean exhibit a scale-free, small world and modular architecture; with their degrees fit best to power-law and exponential distribution. Degree correlation showed that miRNAs of high degrees in miRFNs tend to interact with miRNAs of low degrees, revealing the disassortativity and modularity of miRFNs. More researches on soybean genome and microRNome based on miRFNs and FGNs are underway.

2 METHODS

2.1 Data source

2.1.1 MiRNAs of soybean Soybean miRNA sequences were downloaded from miRBase (Griffiths-Jones *et al.*, 2006) (Release 19, August 2012). It contains 555 mature miRNAs generated from 506 hairpin precursors, wherein 367 miRNAs belong to 81 miRNA families. All miRNAs are denoted by their mature miRNA names, such as 'gma-miR159a-5p'.

2.1.2 3′-UTRs mRNAs' 3′-UTRs were downloaded from Soybase (http://www.soybase.org/dlpages/flank/utrindex.php). It contains 14 136 3′-UTRs transcripted from 12 082 distinct gene ensembls of Wm82 (annotation version: Glyma 1.1).

2.1.3 Target gene networks MiRNAs' target gene networks were extracted from three soybean FGNs (SoyFGNs), which were constructed based on GO annotation of three orthogonal aspects: biological process (BP), cellular component (CC) and molecular function (MF). The work about SoyFGNs will be published in another article synchronously. Finally, we use gene ensembl IDs, such as Glyma10g28890, to represent corresponding genes.

2.2 MiRNA target prediction

Owing to the lack of experimentally validated targets of soybean miRNAs, *in silico* methods for miRNA target prediction are used in this study. To reduce the biases introduced by individual predicting method, we incorporate all miRNA–target interactions that occur in three accepted predicting methods using their default settings, respectively. The three methods are PITA (http://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html), psRNATarget (http://plantgrn.noble.org/psRNATarget/) and TAPIR (http://bioinformatics.psb.ugent.be/webtools/tapir/; using precise search). The $\Delta\Delta Gs$ of all miRNA-target duplexes predicted by PITA are given by PITA itself, whereas the $\Delta\Delta Gs$ of those predicted by psRNATarget and TAPIR are calculated using Vienna RNA Package (Mückstein *et al.*, 2006). Because the highly complementary sequence feature of miRNA-target duplexes in plants, to reduce the false positive, we set a relatively strict criteria of >7 bases in seed length, no G:U wobble or loops, no mismatch and $\Delta\Delta G \leq -10$ Kcal/mol to obtain a target set of high confidence.

2.3 Methods for measuring miRNA FS

As demonstrated in Section 1, miRNAs perform their functions by down-regulating their target genes, and the intensity of which is closely associated with the accessibility of binding the target site. Thus, a more accurate measurement should take fully into account both the FS between their targets and the regulatory strengths they exert on their targets. In this study, we propose an approach to measure the regulatory strength based on target site accessibility and an approach to compute the FS between the target genes of two miRNAs based on FGNs and further to infer the FS of these two miRNAs by integrating these two critical factors. Figure 2 shows the complete workflow of the approaches. Given two miRNAs, miR_i and miR_j, we firstly predict their target gene sets, $Tar\_gene_i = \{g_1, g_2, \ldots, g_m\}$ and $Tar\_gene_j = \{g_1, g_2, \ldots, g_n\}$, respectively (Fig. 2(a)). After that, the union of their target gene sets, $Tar\_gene_{ij} = Tar\_gene_i \cup Tar\_gene_j = \{g_1, g_2, \ldots, g_N\}$ was subjected to two subsequent steps. One is to extract the network topology from the FGNs for further measuring the FS between these two target gene sets (Fig. 2 (c), (d), (e), (f), (g); inside light gray shade); the other is to calculate the regulatory strength for further getting the co-regulation coefficient (Fig. 2(h), (k); inside dark gray shade). Finally, these two values were integrated into the final FS of the given two miRNAs (Fig. 2(l)). The next three subsections will give the details of these three steps.

2.3.1 Regulatory strength of miRNA Kertesz *et al.* (2007) had indicated that target accessibility was a critical factor in miRNA function, which was presented as the free energy ($\Delta\Delta G$) gained from the formation of the miRNA-target duplex and the energetic cost of unpairing the target to make it accessible to the miRNA. They experimentally showed that diminishment of target accessibility substantially reduced the miRNA-mediated translational repression. Based on such accessibility, we define the regulatory strength (denoted by R) to quantify the effect of translational repression. First, we calculate the $\Delta\Delta G$ of each miRNA-target duplex in the targets prediction. Because an miRNA may be predicted to bind more than one site on the same gene transcript, then we use Equation (1) to integrate these multiple $\Delta\Delta Gs$ ($\Delta\Delta G_1, \cdots, \Delta\Delta G_i, \cdots, \Delta\Delta G_n$) into a unique one for each miRNA-target duplex.

$$\Delta\Delta G = \log \sum_{i=1}^{n} e^{\Delta\Delta G_i} \tag{1}$$

Second, we assign a regulatory strength value (R) for a given miRNA–target interaction ($miR_i - g_k$) as defined in Equation (2), for the reason that all $\Delta\Delta Gs$ are less than zero and the smaller its value the greater the regulatory strength will be.

$$R_{i,g_k} = \frac{(\Delta\Delta G_{\max} - \Delta\Delta G)}{(\Delta\Delta G_{\max} - \Delta\Delta G_{\min})} \tag{2}$$

where $\Delta\Delta G$ is the integrated unique accessibility of $miR_i - g_k$ interaction, $\Delta\Delta G_{max}$ and $\Delta\Delta G_{min}$ are the maximum and minimum integrated unique $\Delta\Delta G$ of all miRNA–target interactions.

Given two miRNAs, miR_i and miR_j, we can get the regulatory strengths that they exerted on each gene of their union target set, $Tar\_gene_{ij} = Tar\_gene_i \cup Tar\_gene_j = \{g_1, g_2, \ldots, g_N\}$, respectively, as defined in Equation (2). If there is no predicted interaction for $miR_i - g_k$, we define $R_{i,g_k} = 0$. Putting these two sets of regulatory strengths in an identical fixed order, we form two vectors, $\vec{V}_i$ and $\vec{V}_j$ (Fig. 2 (h)):

$$\begin{aligned} \vec{V}_i &= (R_i, g_1, \cdots, R_i, g_k, \cdots, R_i, g_N) \\ \vec{V}_j &= (R_j, g_1, \cdots, R_j, g_k, \cdots, R_j, g_N) \end{aligned} \quad N = |Tar\_gene_{ij}|$$

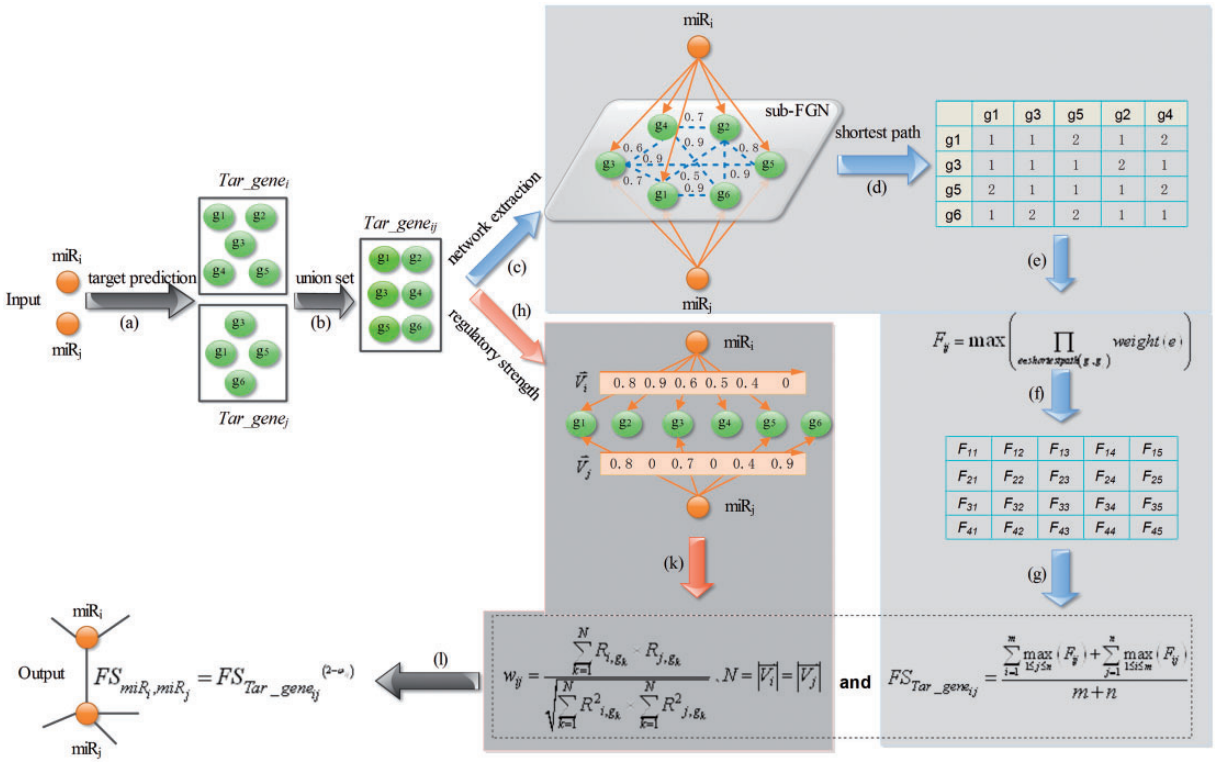

Fig. 2. A schematic view of the workflow for measuring the FS of two miRNAs. (**a**) Target prediction of two given miRNAs by PITA, paRNATarget and TAPIR. (**b**) Union for these two target gene sets. (**c**) The extraction of the network topology (sub-FGN) from the global gene functional networks for further measuring the FS between these two target gene sets (inside light gray shade). (**d**) Computing the shortest paths. (**e**) Computing the FS of genes based on shortest paths. (**f**) Generating a new FS matrix. (**g**) Computing the FS of two target gene sets. (**h**) The calculation of the regulatory strength for further getting the co-regulation coefficient (inside dark gray shade). (**k**) Computing the co-regulation coefficient. (**l**) The integration of target FS and co-regulation coefficient to the final miRNA FS

Finally, we use co-regulation coefficient (w_{ij}) to represent the relationship between the regulatory strengths of these two miRNAs, miR_i and miR_j.w_{ij} are calculated as the cosine function of the angle between the two representative vectors in the N-dimensional space of composition vectors as defined in Equation (3) (Fig. 2 (k)). w_{ij} ranges between 0 and 1. Note that, w_{ij} was only calculated for pairwise miRNAs that have at least one target gene in common. Thus the w_{ij} of miRNAs having no common targets will be 0.

$$w_{ij} = \frac{\vec{V}_i \cdot \vec{V}_j}{\|\vec{V}_i\| \cdot \|\vec{V}_j\|} = \frac{\sum_{k=1}^{N} R_{i,g_k} \cdot R_{j,g_k}}{\sqrt{\sum_{k=1}^{N} R_{i,g_k}^2 \cdot \sum_{k=1}^{N} R_{j,g_k}^2}} \quad N = \left|\vec{V}_i\right| = \left|\vec{V}_j\right| = \left|Tar\_gene_{ij}\right| \tag{3}$$

2.3.2 FS between target gene sets MiRNAs implement their functions through regulating their target genes. Thus the FS between miRNAs should be obviously represented by the FS between their targets to a large extend, though it also suffers fine-tuning of regulatory strength. Therefore, to fully reveal the FS between targets is the key for measuring the FS between miRNAs. Previous studies have shown that miRNA-regulated genes function in a modular and hierarchical fashion and the closer the two genes in a network, the more similar their functions are likely to be. In this study, we propose a network-based method to calculate the FS between two target gene sets.

Given a miRNA pair, miR_i–miR_j, the union set of their target genes was subjected to the network-extraction from three global FGNs (SoyFGNs), which have been constructed by us previously (Fig. 2 (c)). Such a FGN is a weighted undirected graph in which each edge is weighted by the FS of its two end-node genes. Given two target genes, g_i and g_j, we firstly find out the shortest path(s) between them based on the topology of the FGNs (Fig. 2 (d)). Then we define the FS of these two genes as (Fig. 2 (e, f)):

$$F_{ij} = \max\left(\prod_{e \in shortestpath(g_i, g_j)} weight(e)\right), \tag{4}$$

where $shortestpath(g_i, g_j)$ contains all the edges on a shortest path between g_i and g_j in FGN; $weight(e)$ denotes the FS of two genes linked by the edge e. The function max(x) means that F_{ij} is the product of all weights on the shortest path, which can produce the maximum one, if there are more than one shortest path between g_i and g_j. Note that not all target genes are included in SoyFGNs. These absent genes are regarded as orphan nodes in SoyFGNs and have zero similarity with other genes when being used in Equation (4). Finally, we combine all these pairwise similarities into a unique one by Equation (5) (Fig. 2 (g)), which is same with the best-match average approach used for integrating pairwise GO semantic similarity into pairwise FS of genes (Xu, *et al.*, 2013).

$$FS_{Tar\_gene_{ij}} = \frac{\sum_{i=1}^{m} \max_{1 \le j \le n}(F_{ij}) + \sum_{j=1}^{n} \max_{1 \le i \le m}(F_{ij})}{m+n}, \tag{5}$$

where, $FS_{Tar\_gene_{ij}}$ ranges between 0 and 1; m, n are the number of target genes of miR_i and miR_j.

2.3.3 FS between miRNAs Then, we have obtained two quantitative indicators, the co-regulation coefficient (w_{ij}) between two miRNAs and

the FS ($FS_{Tar\_gene_{ij}}$) between their target gene sets. Based on the aforementioned point of view, the final similarity between miRNAs on functionality should be the FS of their target gene sets fine-tuned by their co-regulation coefficient. Thus we give the final calculating formula based on these two indicators as follows:

$$FS_{miR}(miR_i, miR_j) = FS_{Tar\_gene_{ij}}^{(2-\omega_{ij})}, \tag{6}$$

where $FS_{miR}(miR_i, miR_j)$ also ranges between 0 and 1.

Because we have constructed three SoyFGNs based on three GO aspects (as hereinbefore abbreviated as BP, MF and CC), each miRNA pair will get three FSs. We may also need a single integrated FS for each miRNA pair. Here we calculate the weighted average of the three FSs as their integration (hereinafter denoted by Integration), which can be formulated as

$$FS_{\text{wAvg}} = \frac{w_1 \cdot FS_{bp} + w_2 \cdot FS_{mf} + w_3 \cdot FS_{cc}}{w_1 + w_2 + w_3} \tag{7}$$

where, FS_{bp}, FS_{mf} and FS_{cc} are three FSs for each miRNA pair; w_1, w_2 and w_3 are the corresponding weights. Though the absence of a criterion to quantify the weights of the different aspects of GO on miRNAs' function, we think that the aspect giving a larger similarity should have a greater weight. Thus we let the weight be equal to the corresponding FS. The final formula for integration is

$$FS_{\text{wAvg}} = \frac{FS_{bp} \cdot FS_{bp} + FS_{mf} \cdot FS_{mf} + FS_{cc} \cdot FS_{cc}}{FS_{bp} + FS_{mf} + FS_{cc}} \tag{8}$$

where, FS_{wAvg} also ranges between 0 and 1.

2.4 Construction of miRFN

An miRFN here is a weighted undirected graph, in which miRNAs represent the nodes and their functional interactions represent the edges, which are weighted by the pairwise FSs of miRNAs they linked. As described earlier in the text, our method can give a more reliable miRNA FS, a critical problem of miRFN construction; in doing so, we can calculate any pairwise FSs for a list of interesting miRNAs and further construct the functional network. However, there is still a serious problem needed to be addressed that two miRNAs of how similar in function can be connected in the network. We should set an appropriate threshold to ensure that miRNA pairs with FSs greater than or equal to the threshold will be connected by edges; otherwise, they are not connected directly.

In this study, we adopt the clustering–coefficient-based threshold selection. The clustering coefficient (C_i) of a node (i) in a network is defined as $C_i = 2n_i/k_i(k_i - 1)$, where n_i represents the number of edges between $k_i(>1)$ first neighbors of the miRNA i; if $k_i = 1$, we define $C_i = 0$. While the clustering coefficient of a network is defined as the average clustering coefficient of all of its nodes,

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i, \tag{9}$$

where N is the number of nodes in the network. If $N = 0$, we define $C = 0$.

The construction of a miRNA network can be viewed as a process that links are removed from the original complete graph by gradually increasing the FS threshold. We set a series of incremental thresholds τ from 0 to 1 with an increment of 0.01. For each threshold τ, we construct a network comprises the miRNA pairs with FSs greater than or equal to τ. In system biology, a genuine biological network should be scale-free and highly modular; its clustering coefficient, denoted by $c(\tau)$, should be significantly higher than that of the corresponding random network, denoted by $c_r(\tau)$. Here, we denote the difference between $c(\tau)$ and $c_r(\tau)$ by $\Delta c(\tau)$, i.e. $\Delta c(\tau) = c(\tau) - c_r(\tau)$. We conjecture that the most appropriate threshold should be the maximum τ, which can produce a monotonically increasing $\Delta c(\tau)$ when removing the links gradually as the threshold increasing from 0 to τ. More specifically, we formulate this as a discrete optimization problem, where the critical cutoff threshold was determined by finding the first τ, which lets $\Delta c(\tau + 0.01) - \Delta c(\tau) < 0$ over a set of τ gradually increasing from 0 to 1, i.e. $\tau = \arg\ \max_{0\le\tau\le1}(\Delta c(\tau))$. Note that calculating $c_r(\tau)$ of the random networks was non-trivial via Equation (9) because it is not clear which and how many random network models should be used for this purpose. Hence we adopt a statistical method proposed by Elo *et al.* (2007) for its solution. If we denote by N the total number of nodes and by k_i the degree of a node i for the generated network using threshold τ, then $c_r(\tau)$ was the expected value of the clustering coefficient, which can be calculated as:

$$c_r(\tau) = \frac{\left(\bar{k^2} - \bar{k}\right)^2}{\bar{k}^3 N}, \tag{10}$$

where $\bar{k} = 1/N\sum_i^N k_i$ and $\bar{k^2} = 1/N\sum_i^N k_i^2$.

Finally an miRFN can be constructed and represented as $G(V, E, W, T)$, where $V = \{miR_1 miR_2 \cdots, miR_N\}$ represents the miRNAs involved in the network, $E = \{e_{ij} = \prec miR_i, miR_j \succ \mid FS_{miR}(miR_i, miR_j) \ge T\}$ represents the edges between miRNA pairs with FSs greater than or equal to the threshold T and $W = FS_{miR}(miR_i, miR_j)$ represents the weights of the edges that are the FSs between each two connected miRNAs.

2.5 Topological characterization of miRFN

One way to characterize biological networks is to study their topological properties. We by using Cytoscape 2.8.2 (Smoot *et al.*, 2011) investigate the global properties of the four miRNA networks firstly. In addition, we conduct an in-depth analysis of the degree distribution and degree correlation in next two subsections. In absence of special description, degree refers to the nodal degree.

2.5.1 Degree distribution Many previous studies have observed that biological networks are generally scale-free and their degree distributions follow the power law (Arita, 2005; Khanin and Wit, 2006). Some studies have also argued that there are other distributions, such as the lognormal distribution, which explains the degree distribution better than power law does (Pržulj *et al.*, 2004; Stumpf and Ingram, 2007). Here we use four models to investigate the distributions of the four resulting functional miRNA networks. These models are Lognormal, Power law, Exponential and Poisson. All model fittings and visualizations are completed by using Origin 9 (http://www.originlab.com).

2.5.2 Degree correlation Degree correlation is a basic structural metric that calculates the likelihood that nodes link to nodes of similar or dissimilar degree. The former case is called positive degree correlation, whereas the latter is called negative degree correlation. In the social sciences, a network with positive degree correlation is called assortative, whereas one with negative degree correlation is called disassortative (Newman, 2002). Three ways of characterizing the amount of degree correlation are used, each containing less detail and expressing the result in more compact terms. They are the Joint Degree Distribution (JDD), the k-nearest neighbors (*knn*) and the Pearson Degree Correlation (PDC).

The JDD is defined as the distribution in which each entry D_{ij} is the number of edges that the nodes at its endpoints have degrees i and j, respectively. It is actually a 2D distribution of the number of edges with respect to the degree of their connected nodes.

Instead of recording every pair of nodes, as JDD does, *knn* simply averages the degrees of the neighbors of each node of a given degree and plots the results as linear, semi-log and log-log plots. If a degree is missing, it is skipped in the graph. If *knn* rises as nodal degree rises, this indicates that nodes of similar degree tend to be linked, whereas if *knn* falls as degree rises, this indicates the opposite.

The PDC is the most condensed way to characterize the degree-link structure of a network. It consists of the conventional Pearson correlation

calculation applied to each pair of linked nodes. The result always lies in the range [−1, 1] with a negative result indicating that nodes of dissimilar degree tend to be linked and a positive result indicating that nodes of similar degree tend to be linked.

In this section, we will analyze the JDD, *knn* and PDC for miRNA networks to investigate their assortativity.

3 RESULTS

3.1 MiRNA target prediction

We used three methods (PITA, psRNATarget and TAPIR) to predict soybean miRNA targets followed by a strict criterion to control the false discovery rate. As a result, 554 of 555 soybean miRNAs (except for gma-miR1530) target 34 718 sites of 8731 transcripts (mRNAs) derived from 7598 distinct soybean genes. A brief statistic is shown in Figure 3A and the detailed information of predicted targets is provided in Supplementary File S1 (XLS).

3.2 FS of miRNAs

We applied our method to calculate the pairwise FS of 554 miRNAs based on three target FGNs and the accessibility of each miRNA-target duplex. As results, 153 182(99.64%), 150 980(98.20%), 152 630(99.28%) and 153 730(99.99%) miRNA pairs of 153 735(554 × 555/2) gained non-zero similarities in BP, MF, CC and Integration, respectively. The zero similarities appear to be the case of zero similarities between their target genes due to the absence of these targets in SoyFGNs. The pairwise FSs and their dendrogram of Integration, as an example, were shown as a heat map in Figure 3B. The full size of all heat maps in BP, MF, CC and Integration were provided as Supplementary File S2 (PDF).

3.3 MiRNAs in the same family or cluster show high FS

A family of miRNAs includes similar mature miRNAs and complete identical seed regions, which are widely used as the key regions for miRNA target recognition (Bartel, 2009). Therefore, miRNAs of the same family tend to show high FS. A cluster of miRNAs incorporates those are located closely in genome and thus are usually transcribed and expressed synchronously and functions coordinately. Therefore, miRNAs in the same cluster are expected to have higher FS. To evaluate the reliability of the FSs computed by our method, we compared the FS of miRNAs in the same family and the same cluster with that of the random-selected miRNAs, which are neither from the same family nor the same cluster. We firstly downloaded soybean miRNA family data and the genome coordinate data from miRBase and then identified miRNA clusters by setting the distance between consecutive miRNAs <50 kb on the genome coordinate, as previously suggested by Baskerville and Bartel (2005). We next calculated the FS of miRNAs in the same family and miRNAs in the same cluster, and then compared these with that of random-selected miRNAs. As a result, the average FSs of miRNAs in 81 families (denoted by 'miR-Family'), miRNAs in 59 clusters (denoted by 'Cluster') and 81 random-selected miRNAs (denoted by 'Random') are shown as Figure 3C. The analysis of variance indicates that the FSs of miRNAs in the same family and miRNAs in the same cluster

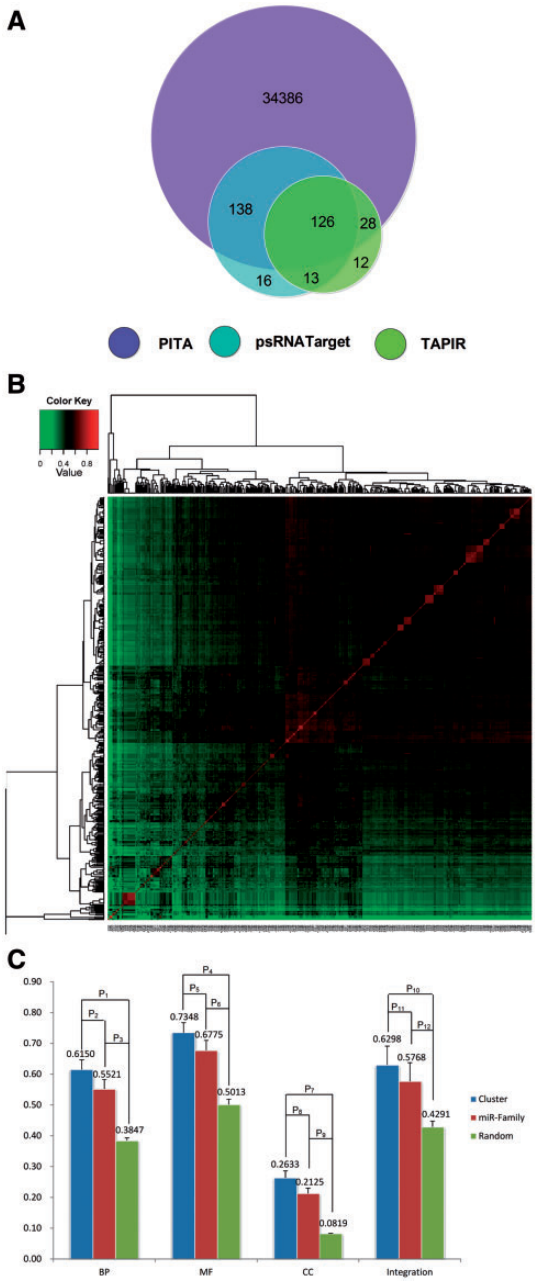

Fig. 3. **A**) A brief statistic (Venn diagram) of the number of targets predicted by PITA, psRNATarget and TAPIR. Numbers in the figure denote the number of predicted target genes in each part. (**B**) The heat map of the pairwise FSs and their dendrogram of Integration. The full size of all heat maps in BP, MF, CC and Integration was provided as Supplementary File S2. (**C**) A comparison of the mean FS of miRNAs in the same family, miRNAs in the same cluster and miRNAs of random pairs. The error bars represent variance of the means. $P_1 \sim P_{12}$ are the *P*-values by analysis of variance ($\alpha = 0.01$): $P_1 = 9.4E\text{-}111$, $P_2 = 5E\text{-}08$, $P_3 = 8.5E\text{-}111$, $P_4 = 9.67E\text{-}86$, $P_5 = 1.12E\text{-}06$, $P_6 = 7.4E\text{-}109$, $P_7 = 2.3E\text{-}130$, $P_8 = 9E\text{-}09$, $P_9 = 9.3E\text{-}124$, $P_{10} = 2.6E\text{-}128$, $P_{11} = 2.8E\text{-}09$, $P_{12} = 1.3E\text{-}138$

are both significantly greater than that of random-selected miRNAs are in BP, MF, CC and Integration, respectively (Fig. 3C). This result suggests that our method can reflect the FS of miRNAs correctly and is consistent with expectations. Additionally, miRNAs in a cluster are also significantly more similar in function than miRNAs in the same family, which is contrary with the conclusions in the study of human miRNA similarity by Wang, *et al*. (2010). This contradiction infers that the clustered distribution of miRNAs in the genome may have different importance between animals and plants. To confirm that using target network can improve the accuracy of computing miRNA FS directly using GO similarity, an additional comparison was carried out between the FSs based on these two kinds of information in Supplementary Materials SM2 and Figure S2. The results indicated that, target genes' interactive context enhanced the ability of our method on reflecting the fact that miRNAs in the same family or in the same cluster are more similar in function than those that are randomly selected.

3.4 Construction of miRFN

By setting a series of incremental thresholds τ (from 0 to 1) with an increment of 0.01, we used each threshold to filter the original complete networks (including all pairwise similarity of 554 miRNAs) in BP, MF, CC and Integration, respectively. As a result, we obtained 99, 98, 67 and 98 networks in BP, MF, CC and Integration, respectively. Using our own JAVA script, we calculated the cluster coefficient of each network [$c(\tau)$] and that of its random model [$c_r(\tau)$]. As shown in Figure 4A, the first stop of monotonically increasing of the $\Delta c(\tau)$ occurs at τ =0.53, 0.68, 0.13 and 0.85 in BP, MF, CC and Integration, respectively, which indicates that these thresholds are the most appropriate ones to construct the miRFNs in BP, MF, CC and Integration, respectively.

Using above thresholds, we constructed four miRFNs in BP, MF, CC and Integration, respectively. A graphic view of the network in Integration is shown as Figure 4B. The full size of the graphic views of these four networks was provided in Supplementary File S3 (PDF).

3.5 Topology of soybean miRFN

Analyzed by Cytoscape 2.8.2, the global properties of the miRFNs in BP, MF, CC and Integration are shown as Table 1. These four networks cover 462 (83.2%), 454 (81.8%), 512 (92.3%) and 474 (85.4%) miRNAs of soybean, respectively. All networks manifest the typical common characteristics of biological networks: high clustering coefficient, low diameter and density and high centralization (shown in Table 1).

3.5.1 Degree distribution Figure 5A shows the graphic views of the degree distribution and four fitted models for each miRNA network. The detailed parameters of these models and their performances (represented by R-squares, R^2) are listed in Table 2. Our results demonstrate that the Power law and Exponential models fit best to the degree distribution on the whole, whereas the Poisson fits worst. These degree distributions indicate that miRFNs have the typical characteristics of biological networks, e.g. scale-free small world, rather than that of random network, of which the degree distribution fits Poisson best.

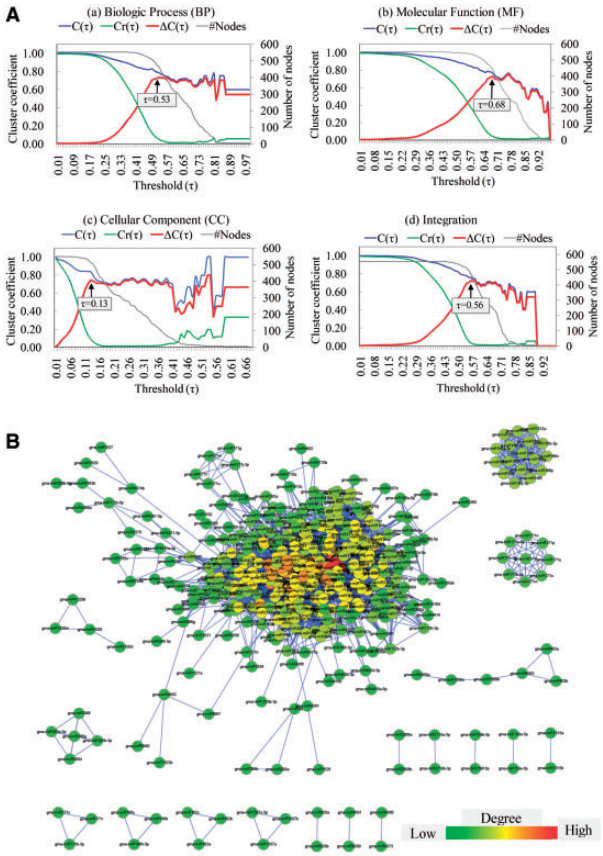

Fig. 4. (**A**) Cluster coefficient and nodes number of the network under each threshold in BP, MF, CC and Integration, respectively. Black arrows point to the first peaks of the red curve and rectangular boxes show the corresponding threshold values. $c(\tau)$ (blue curve)represents the cluster coefficient of the created network, $c_r(\tau)$ (green curve) the cluster coefficient of the corresponding random network and $\Delta c(\tau)$ (red curve) the difference between $c(\tau)$ and $c_r(\tau)$. '#Nodes' represents the number of nodes at each threshold (gray curve). (**B**) A graphic view of the integrated soybean miRFN. The color of nodes from green to red corresponds to node degree from low to high. (Visualized by Cytoscape 2.8.2)

Table 1. The summary properties of soybean miRFNs in BP, MF, CC and Integration

Property	BP	MF	CC	Integration
Number of nodes	462	454	512	472
Number of edges	7858	8271	16 813	7038
Cluster coefficient	0.782	0.779	0.832	0.762
Connected components	13	15	5	11
Diameter	6	5	6	7
Radius	1	1	1	1
Centralization	0.503	0.467	0.511	0.487
Shortest paths	185 394	162 316	253 520	193 276
Characteristic path length	2.372	2.287	2.104	2.573
Average number of neighbors	34.017	36.436	65.676	29.822
Density	0.074	0.080	0.129	0.063
Heterogeneity	1.281	1.292	1.135	1.323

Note: All these properties are calculated using Cytoscape 2.8. BP, biological progress; MF, molecular function; CC, cellular component; Integration, the integrated network based on the integrated FS.

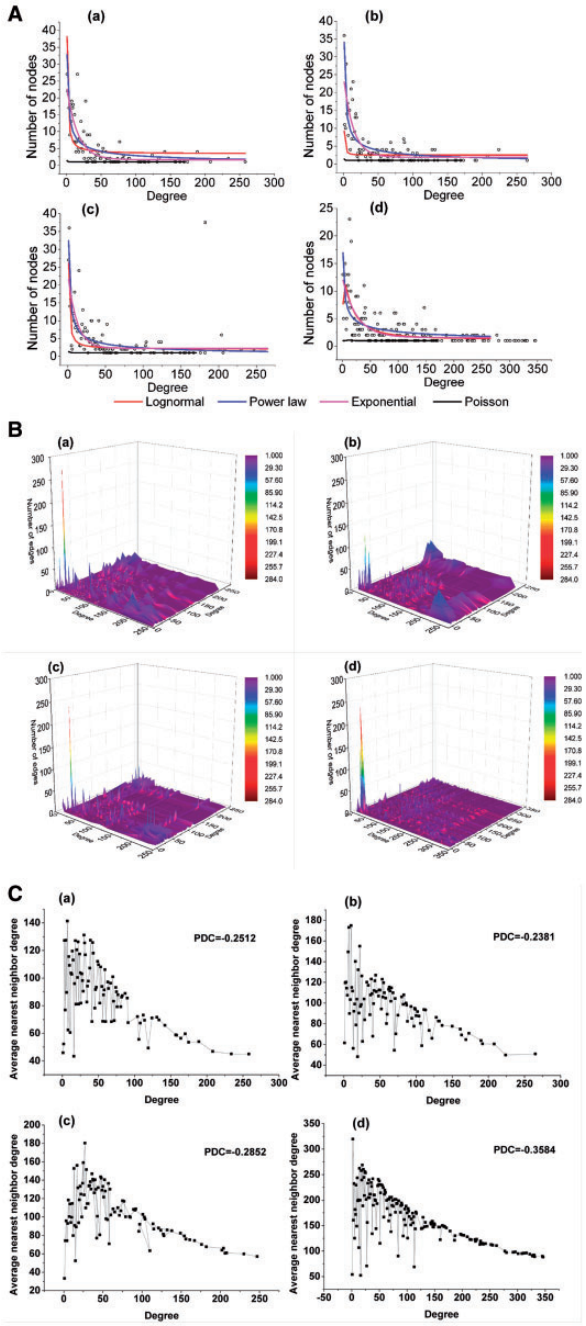

Fig. 5. (**A**) The graphic view of the degree distributions and fitted models for each miRFN. (**B**) The JDD of miRNA networks. The *x*- and *y*-axes represent the nodal degrees and *z*-axis represents the number of edges of pairwise degrees. (**C**) The k-nearest neighbors (*knn*) and Pearson degree correlation (PDC) of miRNA network. Note that, PDCs in this figure are values calculated according to the degrees of two endpoints of all edges, rather than the parameters of this graph. (a) Integration; (b) biological progress (BP); (c) molecular function (MF) and (c) cellular component (CC)

3.5.2 Degree correlation The JDD of miRNA networks in BP, MF, CC and Integration are visualized as 3D surface graph in Figure 5B. The results suggest several important points. First, in all miRFNs, most of degree pairs have small number of edges. The average numbers of edges are 3.5428, 3.6300, 2.6260 and 3.5473 for network in BP, MF, CC and Integration, respectively, which indicate the low network densities, as the same as shown in Table 1. Second, the extremely sharp protrusions, especially in MF, CC and Integration, show that little nodes share a large number of edges, indicating the existence of local dense functional modules in miRFN. Third, the majority of apparent peaks (the number of edges $\geq$30) appear in the degree of low-high and high-low node pairs, followed by the low-low node pairs. This result suggests that the nodes of high degree tend to connect to that of low degree in miRFNs, indicating their disassortative feature.

Similar results were obtained in the analysis of the knn and the PDC (Fig. 5C). The descending knn and negative PDC indicate that nodes of dissimilar degree tend to be linked in all four miRFNs.

4 DISCUSSION

In this study, we first proposed a novel approach to measuring the FS of miRNAs, both considering their site accessibility and the topology of target gene functional network. Based on pairwise FS, we further constructed four soybean (*G.max*) miRFNs via a clustering–coefficient-based threshold selection. Our method in this article made up for the defects of existing methods either that some of them failed to infer the miRNA network of non-model species without such extensive miRNA-phenotype associated data as miRNA-disease association in human, or that some of them did not consider the target site accessibility and the interactive context between target genes. The successful application in soybean indicates that our methods can be easily extended to other species. In next five subsections, we will discuss the highlights of this article and the research prospects.

4.1 Indispensable target site accessibility

Accessibility proposed by Kertesz *et al.* (2007) has been identified as one of critical factors that are involved in target recognition. In this study, we further demonstrated its important role for inferring miRNA FS, and hence used it to infer the FS of miRNAs based on such an argument that a miRNA pair has higher accessibility with their common targets should be more similar in function. Taking accessibility into account improved the sensitivity of functionally comparing miRNAs based on their targets. Especially for those miRNAs having the same targets can also be distinguished regarding their functionality due to the different accessibility. Kertesz *et al.* (2007) defined the accessibility as the free energy gained by transitioning from the state in which miRNA and target are unbound to the state in which the miRNA binds its target (Fig. 1B). After a series of reasonable transformation, we converted the pairwise accessibility between miRNAs and their targets to regulatory strengths, and further to a fine-tuning factor, referred to as co-regulation coefficient. Such defined co-regulation coefficient reveals the intensity that two miRNAs co-regulate the same group of target genes, which is conducive to reflect the similarity of miRNAs in functionality more accurately and sensitively (Supplementary Materials S1 and Fig. S1).

Table 2. Four types of fitted models of the degree distribution for each network

Model	Parameter	BP	MF	CC	Integration
Lognormal $y = y_0 + \frac{A}{\sqrt{2\pi}\sigma x} e^{-\frac{\left[\ln\frac{x}{\mu}\right]}{2\sigma^2}}$	y_0	2.50 ± 0.836	2.00 ± 1.458	1.25 ± 0.4153	3.40 ± 0.537
	μ	1.00 ± 12.752	2.33 ± 121.428	28.63 ± 11.806	6.9852
	σ	1.79 ± 14.907	14.07 ± 845.528	1.41 ± 0.198	9.431
	A	30.32 ± 28.450	885.86 ± 190.54	379.10 ± 67.686	8.22
	R^2	0.02802	0.41554	**0.50255**	0.468
Power law $y = a \cdot x^b$	a	34.13 ± 3.008	32.41 ± 2.774	16.93 ± 1.801	32.91 ± 2.953
	b	-0.56 ± 0.040	-0.57 ± 0.039	-0.41 ± 0.038	-0.53 ± 0.039
	R^2	0.62591	**0.62385**	0.41513	**0.63514**
Exponential $y = y_0 + Ae^{bx}$	y_0	1.86 ± 0.525	2.24 ± 0.515	1.58 ± 0.291	1.64 ± 0.566
	A	22.71 ± 2.205	21.081 ± 2.4179	10.752 ± 1.0470	22.31 ± 1.913
	b	-0.07 ± 0.002	-0.85 ± 0.016	-0.39 ± 0.007	-0.06 ± 0.010
	R^2	**0.65389**	0.57212	0.50189	0.70027
Poisson $y = y_0 + \frac{\lambda^x}{x!} e^{-\lambda}$	y_0	1	1	1	5.607 ± 0.720
	λ	1	2	13	2.042 ± 33.56
	R^2	−0.01149	−0.01136	−0.00847	−0.00403

Note: R-squares (R2) in bold font and gray background represent the best-fitted model for each network.

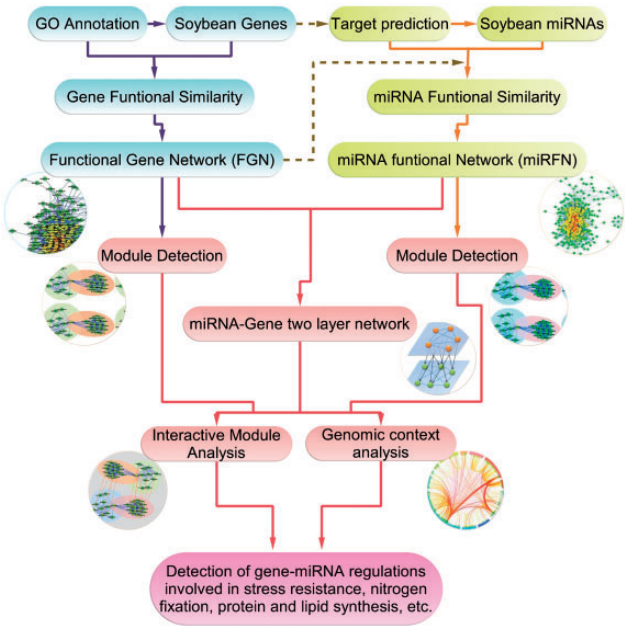

Fig. 6. A schematic view of our work in this article as part of the whole research project for soybean functional network. Rounded rectangles with blue backgrounds represent the construction of FGN of soybean, the result of which is a part of the source data of this article and will be published in another article; yellow backgrounds represent our works in this article; and red backgrounds represent our subsequent research prospects

4.2 Interactive context between target genes

The FS of miRNAs should be obviously represented by that of their targets to a large extend, though it also suffers fine-tuning of regulation strength. As reviewed in the Section 1, a growing number of studies have shown that miRNA-regulated genes are not isolated with each other, but tend to constitute a FGN, which provides an additional global view of the interactive context of each gene. With the realization that interactions encode functional dependencies, the notion of functional comparison has been expanded to multi-activities performed by multiple gene products in conjunction. Hence the opportunity of computational function comparison is the use of a gene's interaction context within the network to compare its function. In this study, we propose a network-based method to calculate the FS between two target gene sets, considering such interaction context. Then the final similarity of miRNAs is subsequently quantified as the similarity of the target genes fine-tuned by co-regulation coefficient. By comparing the similarity of miRNAs in the same family and the same cluster to the similarity of random-selected miRNAs, we confirmed the accuracy of our method.

4.3 The first global view of soybean miRNA functional interactions

Current achievements in miRNA research area are inspiring. Molecular cloning and functional elucidation have greatly advanced our understanding of this small RNA species. As one of the ultimate goals, many efforts have devoted to draw a global view of miRNA interactions. However, such efforts mainly focused on model organisms, especially the human species, while rarely focused on plant. Additionally, they are either difficult to be extended to non-model organisms or fail to take fully into account the information embedded in miRNA–target and target–target interactions. Our work introduced in this article is the first study on soybean miRNAs on the network level. We draw four miRFNs, containing 554 soybean miRNAs, based on three types of their target networks. The topological analysis shows that, similar to other biological networks, miRFN is a scale-free network, and its degree distribution fits best to power-law and exponential distribution. Its degree correlation indicates that the miRNAs of high degrees tend to interact

with those of low degrees, which is referred to as disassortativity. The achievements here are the fundamentals for further studies on interactome of soybean on genome and microRNome level.

4.4 Availability

Based on the work of this research, we have also developed a web platform for information retrieval and analysis of soybean miRNAs and their target gene network. The Web site can be accessed at SoymiRNet: http://nclab.hit.edu.cn/SoymiRNet.

4.5 Prospects

Our functional networks provide a systematic view of the whole microRNome (miRNome) of soybean, and hence such construction of the first miRNome-wide maps has been followed by attempts to discover and predict function within the system as a whole. Therefore, the efforts of this article are just the beginning of the soybean functional genome and microRNome research, and part of a comprehensive study on the soybean functional network. As shown in Figure 6, our whole research project consists of three main focuses: (i) the construction of FGN of soybean (shapes in blue background), the result of which is part of the source data of this article and will be published in another article; (ii) inferring miRNA functional network of soybean as described in this article (shapes in yellow background); and (iii) module detection, miRNA-gene two layer network analysis and further interactive module analysis coupled with genomic context analysis to discover the gene-miRNA regulatory mechanism involved in stress resistance, nitrogen fixation, protein and lipid synthesis and so forth in soybean (shape in red background). All in all, based on the efforts in this article and previous work, we will carry out more comprehensive studies on soybean functional interactome on genome and microRNome level.

ACKNOWLEDGEMENT

The authors thank the editors and reviewers for their valuable comments and suggestions.

Funding: The Natural Science Foundation of China (grant no. 60932008, 61172098, 91335112 and 61271346); and the Specialized Research Fund for the Doctoral Program of Higher Education of China (grant no. 20112302110040).

Conflict of Interest: none declared.

REFERENCES

Arita,M. (2005) Scale-freeness and biological networks. *J. Biochem.*, **138**, 1–4.

Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.

Baskerville,S. and Bartel,D.P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*, **11**, 241–247.

Chen,L.-H. *et al.* (2010) MicroRNA and aging: a novel modulator in regulating the aging network. *Ageing Res. Rev.*, **9**, S59–S66.

Elo,L.L. *et al.* (2007) Systematic construction of gene coexpression networks with applications to human T helper cell differentiation process. *Bioinformatics*, **23**, 2096–2103.

Enright,A.J. *et al.* (2004) MicroRNA targets in *Drosophila*. *Genome Biol.*, **5**, 1.

Griffiths-Jones,S. *et al.* (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.

Horvath,S. *et al.* (2006) Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc. Natl Acad. Sci. USA*, **103**, 17402–17407.

Hsu,C.W. *et al.* (2008) Characterization of microRNA-regulated protein-protein interaction network. *Proteomics*, **8**, 1975–1979.

Jones-Rhoades,M.W. *et al.* (2006) MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.*, **57**, 19–53.

Kertesz,M. *et al.* (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.

Khanin,R. and Wit,E. (2006) How scale-free are biological networks. *J. Comput. Biol.*, **13**, 810–818.

Krek,A. *et al.* (2005) Combinatorial microRNA target predictions. *Nat. Genet.*, **37**, 495–500.

Lee,S.G. *et al.* (2004) A graph-theoretic modeling on GO space for biological interpretation of gene clusters. *Bioinformatics*, **20**, 381–388.

Lin,J. *et al.* (2007) A multidimensional analysis of genes mutated in breast and colorectal cancers. *Genome Res.*, **17**, 1304–1318.

Lord,P.W. *et al.* (2003) Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation. *Bioinformatics*, **19**, 1275–1283.

Mückstein,U. *et al.* (2006) Thermodynamics of RNA–RNA binding. *Bioinformatics*, **22**, 1177–1182.

Newman,M.E. (2002) Assortative mixing in networks. *Phys. Rev. Lett.*, **89**, 208701.

Pesquita,C. *et al.* (2008) Metrics for GO based protein semantic similarity: a systematic evaluation. *BMC Bioinformatics*, **9**, S4.

Pesquita,C. *et al.* (2009) Semantic similarity in biomedical ontologies. *PLoS Comput. Biol.*, **5**, e1000443.

Pržulj,N. *et al.* (2004) Modeling interactome: scale-free or geometric? *Bioinformatics*, **20**, 3508–3515.

Satoh,J.-I. and Tabunoki,H. (2011) Comprehensive analysis of human microRNA target networks. *BioData Min.*, **4**, 1–13.

Shalgi,R. *et al.* (2007) Global and local architecture of the mammalian microRNA–transcription factor regulatory network. *PLoS Comput. Biol.*, **3**, e131.

Smoot,M.E. *et al.* (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431–432.

Stumpf,M.P. and Ingram,P.J. (2007) Probability models for degree distributions of protein interaction networks. *Europhys. Lett.*, **71**, 152.

Voinnet,O. (2009) Origin, biogenesis, and activity of plant microRNAs. *Cell*, **136**, 669–687.

Wang,D. *et al.* (2010) Inferring the human microRNA FS and functional network based on microRNA-associated diseases. *Bioinformatics*, **26**, 1644–1650.

Wang,J.Z. *et al.* (2007) A new method to measure the semantic similarity of GO terms. *Bioinformatics*, **23**, 1274–1281.

Xu,J. *et al.* (2011) MiRNA–miRNA synergistic network: construction via co-regulating functional modules and disease miRNA topological features. *Nucleic Acids Res.*, **39**, 825–836.

Xu,Y. *et al.* (2013) A novel insight into Gene Ontology semantic similarity. *Genomics*, **101**, 368–375.

Yu,G. *et al.* (2010) A new method for measuring FS of microRNAs. *J. Integr. OMICS*, **1**, 49–54.