Plant Omics Journal POJ

POJ 6(6):464-473 (2013) ISSN:1836-3644

Global characterization of Arabidopsis protein interactome

Jian Yang, Jiabin Wang, Yuling Hu, Yiheng Tang, Xugang Hou, Xiao Li *

College of Life Science, Sichuan University, Key Laboratory of Resource Biology and Eco-environment, Ministry of Education, Sichuan Key Laboratory of Molecular Biology and Biotechnology, Chengdu, 610064, China

***Corresponding author: lix@scu.edu.cn.**

Abstract

The network of all protein-protein interactions (PPIs) in an organism, named as the interactome, is a powerful tool for understanding biological processes as an integrated system. Studies in several model species indicate that interactomes exhibit some unique features, such as the scale-free topology and the correlations with genomic information. Here we constructed an Arabidopsis interactome comprised of high-quality protein interactions from several publicly available databases. The interactome contained a total of 3,432 distinct PPIs, among 1,679 proteins identified uniquely by genome locus IDs. The analysis showed that the Arabidopsis interactome shares similar features with those of model organisms at the global level, but also exhibits a local divergence. We found that two interacting proteins show a much higher level of co-localization than randomly-selected protein pairs. Our investigation also indicates that there is a significant correlation between interacting proteins and biological functions. Interactome comparison between Arabidopsis and non-plant species showed that protein complexes involved in the core biological processes may undergo more evolutionary pressure to remain conserved. To facilitate the plant focused research, we also constructed a database server called PlaPID (Plant Protein Interaction Database). It is an integrative information platform for system-level understanding of gene function and biological processes in plant and is free to access at http:www.plapid.net.

Keywords: *Arabidopsis thaliana*; database; functional genomics; interaction; interactome; protein-protein interaction.
Abbreviations: BIND_Biomolecular Interaction Network Database; BT_BTB and TAZ domain protein; bZIP_basic leucine-zipper transcription factor; CaM_calmodulin; CaMK_CaM kinase; CDPK_Ca2+-dependent protein kinase; co-AP_co-affinity purification; DIP_Database of Interacting Proteins; FKBP_FK506-binding protein; GO_Gene Ontology; HSP_heat shock protein; MINT_ Molecular Interactions Database; MIPS_Munich Information Center for Protein Sequences; MS_mass spectrometry; PlaPID_Plant Protein Interaction Database; PPI_protein-protein interaction; SUBA_Arabidopsis Subcellular Database; Y2H_yeast two-hybrid.

Introduction

Protein–protein interactions (PPIs) are fundamental to all biological processes, and play important roles in defining most cellular functions. Traditional study of one protein or one interaction at a time limits our global understanding of biological systems. Given the proteomic makeup of an organism, the complete set of possible protein interactions constitutes its interactome which has quickly become a valuable resource to research protein function and understand the molecular mechanisms underlying diseases (Ideker and Sharan, 2008; Park et al., 2009). With the development of high-throughput protein interaction detection technologies, such as yeast two-hybrid (Y2H) screens and co-affinity purification (co-AP) followed by mass spectrometry (MS), interactome mapping projects have been done in yeast (*Saccharomyces cerevisiae*), fruit fly (*Drosophila melanogaster*), nematode worm (*Caenorhabditis elegans*), and human (*Homo sapiens*) (Morell et al., 2009; Wodak et al., 2009). Many previous studies on these interactomes have revealed that complex protein interaction networks have some unique topological and biological attributes that may reflect biological phenomena. In particular, the interactomes appear to be a scale-free network which has a power law degree distribution, in which most proteins are of low degree, whereas a few proteins act as "highly connected hubs" (Jeong et al., 2000; Jeong et al., 2001). Power law topology might be related to genetic robustness (Albert et al., 2000; Barabasi and Oltvai, 2004). Furthermore, analyses of these large datasets have revealed interesting biological properties within interactomes. For instance, two interacting proteins show a much higher level of co-localization than randomly-selected protein pairs, especially for those involved in the same biochemical reaction and the same protein complex (Grigoriev, 2001; Fraser et al., 2004).

In this respect, progress in plants is still lagging behind (Uhrig, 2006; Morsy et al., 2008; Yuan et al., 2008). The previous works were carried out on the interactomes of several model organisms, but little is known about the properties of the plant interactome, even for *Arabidopsis thaliana*. As the joint efforts of the community, a large number of protein interactions have been obtained from the study of individual proteins during the last few decades for the model plant Arabidopsis, most of which have been published in thousands of literatures. Recently, several public-available databases were devoted to maintain high-quality protein interaction information by manual curation of the literature. Thus, we can present an Arabidopsis interactome by means of collecting interaction data deposited in these databases, which enables us to conduct a preliminary study of network properties in plants. In this study, using the currently-available Arabidopsis interaction data, we carried out

Table 1. Intersects between each two datasets of Arabidopsis interactions.

IntAct (2538)					
BIND (1530)	370 (9.10%)				
TAIR (855)	206 (6.07%)	85 (3.56%)			
MINT (91)	23 (0.87%)	3 (0.19%)	10 (1.06%)		
DIP (13)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	
	IntAct	BIND	TAIR	MINT	DIP

The total number of interactions obtained from each database is presented in italic type beside the database name. The overlapping rate, calculated as the number of interactions in intersects divided by the number of interactions in the smaller dataset of the two, is indicated beside the number of interactions in intersects.

a comprehensive investigation whether the global features of the interactomes observed in model organisms (e.g. yeast) and humans also hold true in plant species: namely, (1) the scale-free topology and the centrality–lethality rule, (2) co-localization of interacting proteins, and (3) interacting proteins being involved in the same biological process. The conservation of protein interaction between Arabidopsis and other model organisms was next analyzed in details.

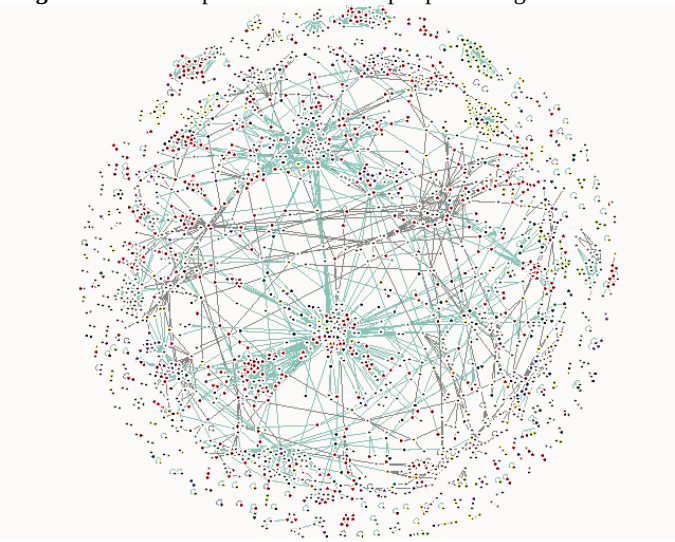
Results and Discussion

The integrated Arabidopsis interactome

We built up an Arabidopsis interactome map based on the public-available PPI databases. These databases have been devoted to maintain high-quality protein interaction information by manually curated literature. Additional 78 interactions were manually collected from literature. To integrate the interaction data, protein identifiers were mapped to genome locus ID by BioMart (Durinck et al., 2005). By removing same interactions, we gained a non-redundant high-quality Arabidopsis PPI dataset which constitutes the Arabidopsis interactome containing a total number of 3,432 distinct PPIs among 1,679 proteins identified uniquely by genome locus IDs. The comparison between datasets from different databases show that only a few interactions overlap among diverse databases (Table 1). More than 80% of interactions were detected based on individual *in vivo* experiments (such as coimmunoprecipitation) or *in vitro* experiments (such as pull-down), and about 20% based on small-scale Y2H assays or protein chip (Supplementary Table 1 for detailed statistics). These small-scale Y2H assays mainly focused on the identification of interactions among members within several gene families of transcription factor, such as MADS box, bHLH and bZIP gene family, many of which were further confirmed by individual biochemical experiments (Colombo et al., 2008; Chandler et al., 2009; Weltmeier et al., 2009).

The Arabidopsis interactome network is shown in Fig. 1, where proteins (nodes) are highlighted in different color according to their subcellular localizations. To visualize the network, each protein was assigned a unique subcellular compartment using a winner-take-all method (see Methods). The analysis of topological features shows that the degree distributions of the interactome network is well fit to a scale-free characteristic with decay exponent γ =1.65. Moreover, the degree distributions for essential (258 genes) and non-essential genes (1,456 genes) are significantly different (Wilcoxon rank sum test; $P < 1.5 \times 10^{-4}$). The average clustering coefficient C_i is 0.15 for the interactome network, which is slightly higher than those of the human (0.10), yeast (0.11) and worm (0.08) networks (Barabasi and Albert, 1999; Barabasi and Oltvai, 2004; Li et al., 2006).

Fig 1. The Arabidopsis interacome map representing an overall



network of Arabidopsis protein interactions. Nodes indicate proteins highlighted in different color according to primary site of subcellular compartment; edges represent interactions.

These global features are consistent with the previous observations in yeast and humans, suggesting that the centrality–lethality rule may also hold true in plants. In the Arabidopsis interactome, several isoforms of Calmodulin (CaM) have the largest number of interactions, including CaM4 (127 interactions), CaM7 (119 interactions), CaM9 (114 interactions), CaM6 (107 interactions) and CaM8 (76 interactions), most of which were identified based on a high-density protein microarray (Popescu et al., 2007).

Subcellular localizations for Arabidopsis PPIs

It has been confirmed that two proteins that interact with each other tend to colocalize in the same subcellular compartment in humans and yeast (Gandhi et al., 2006). Among the total of 1,679 proteins in the Arabidopsis interactome, only 58% (950) proteins are annotated to their subcellular localizations by experiments referring to the Arabidopsis Subcellular Database (SUBA) (Heazlewood et al., 2007); and 220 proteins are assigned to multiple subcellular compartments. Based on the limited annotations, we examined the enrichment or depletion of protein interactions, where both interacting proteins are localized in the same subcellular compartments. Fig. 2a shows the numbers of interactions as a function of the subcellular localization of their interacting proteins and *P* values associated

Table 2. The top 30 of interologs ranked by the number of organisms, among which the interactions are conserved.

Locus A	Protein A	Locus B	Protein B	Biological process	Species
At4g02520	Glutathione S-transferase	At4g02520	Glutathione S-transferase	toxin catabolic process	5
At1g32750	Transcription initiation factor TFIID subunit 1-A	At1g55520	Transcription initiation factor TFIID TBP-2 subunit	transcription	4
At1g32750	Transcription initiation factor TFIID subunit 1-A	At3g13445	Transcription initiation factor TFIID TBP-1 subunit	transcription	4
At5g58590	Ran-binding protein 1 homolog c	At5g55190	GTP-binding nuclear protein Ran-3 (Ras-related nuclear protein 3)	protein transport	4
At5g58590	Ran-binding protein 1 homolog c	At5g20020	GTP-binding nuclear protein Ran-2 (Ras-related nuclear protein 2)	protein transport	4
At5g58590	Ran-binding protein 1 homolog c	At5g20010	GTP-binding nuclear protein Ran-1 (Ras-related nuclear protein 1)	protein transport	4
At1g54140	Transcriptional activation factor TAFII32, putative	At1g04950	Putative TATA binding protein-associated factor	transcription	4
At1g54140	Transcriptional activation factor TAFII32, putative	At1g54360	transcription initiation factor	transcription	4
At5g20850	DNA repair protein RAD51 homolog 1	At5g20850	DNA repair protein RAD51 homolog 1	DNA repair	4
At5g67380	Casein kinase II subunit alpha-1	At5g47080	Casein kinase II subunit beta-1	protein amino acid phosphorylation	4
At5g67380	Casein kinase II subunit alpha-1	At3g60250	Casein kinase II subunit beta-3	protein amino acid phosphorylation	4
At5g61210	Synaptosomal-associated protein SNAP25-like 1	At1g08560	Syntaxin-related protein KNOLLE	vesicle-mediated transport	3
At1g08560	Syntaxin-related protein KNOLLE	At1g12360	SNARE-interacting protein KEULE	vesicle-mediated transport	3
At4g29810	Mitogen-activated protein kinase kinase 2	At2g43790	Mitogen-activated protein kinase 6	signal transduction	3
At2g03710	Agamous-like MADS-box protein AGL3	At5g60910	Agamous-like MADS-box protein AGL8	regulation of transcription	3
At2g45650	Agamous-like MADS-box protein AGL6	At5g15800	Agamous-like MADS-box protein AGL2	regulation of transcription	3
At5g60910	Agamous-like MADS-box protein AGL8	At5g15800	Agamous-like MADS-box protein AGL2	regulation of transcription	3
At1g24260	Agamous-like MADS-box protein AGL9	At1g26310	Agamous-like MADS-box protein AGL10	regulation of transcription	3
At1g09020	putative activator subunit of SNF1-related protein kinase	At3g01090	SNF1-related protein kinase catalytic subunit alpha KIN10	protein amino acid phosphorylation	3
At1g09020	putative activator subunit of SNF1-related protein kinase	At3g29160	SNF1-related protein kinase catalytic subunit alpha KIN11	protein amino acid phosphorylation	3
At3g43810	Calmodulin-7	At1g74740	Calcium-dependent protein kinase 30	response to stimulus	3
At5g21274	Calmodulin-6	At1g74740	Calcium-dependent protein kinase 30	response to stimulus	3
At1g66410	Calmodulin-4	At1g74740	Calcium-dependent protein kinase 30	response to stimulus	3
At4g38130	Histone deacetylase 19	At1g24190	Histone deacetylase complex subunit Sin3	regulation of transcription	3
At5g53530	Vacuolar sorting protein-like	At2g17790	Putative vacuolar sorting protein 35	protein transport	3
At5g20010	GTP-binding nuclear protein Ran-1 (Ras-related nuclear protein 1)	At2g16950	Putative transportin	protein transport	3
At3g43810	Calmodulin-7	At3g19100	Calcium-dependent protein kinase	response to stimulus	3
At3g43810	Calmodulin-7	At4g23650	Calcium-dependent protein kinase	response to stimulus	3
At5g21274	Calmodulin-6	At3g19100	Calcium-dependent protein kinase	response to stimulus	3
At5g21274	Calmodulin-6	At4g23650	Calcium-dependent protein kinase	response to stimulus	3

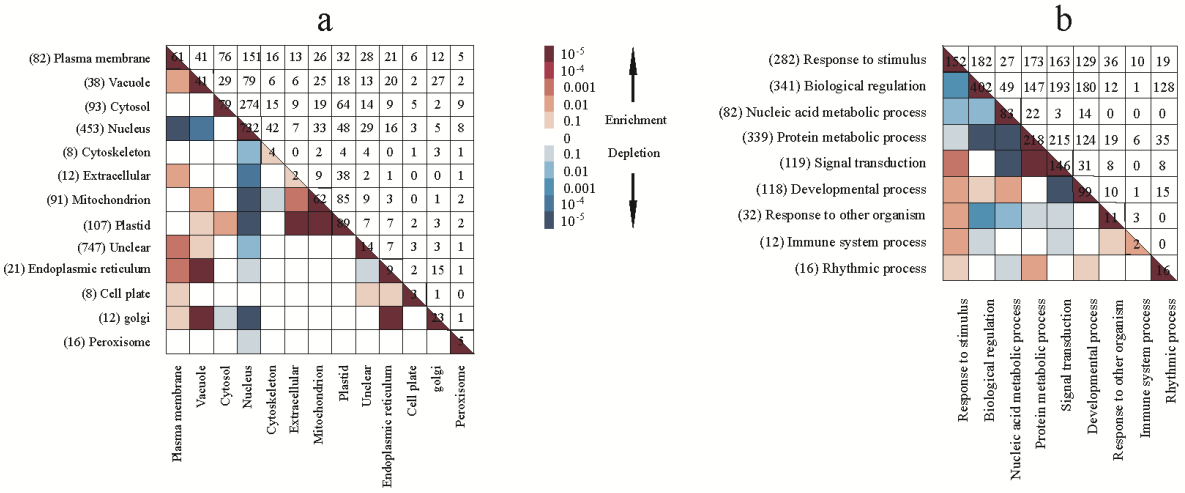


Fig 2. Correlation of protein interactions with subcellular localization and biological process. (a) Correlation of protein interactions with subcellular localization of proteins. The numbers of individual protein numbers is shown beside compartment names. The distribution of protein interactions is represented as a matrix, where each axis indicates the primary subcellular compartment of one of proteins in an interaction pair. The higher diagonal of the matrix shows the numbers of interacting protein pairs, in which both proteins are assigned to a subcellular compartment, while the lower diagonal of the matrix indicates fold enrichment or depletion represented in colors. The statistic enrichment or depletion is calculated as the ratio of the number of edges in the observed network to the expected number of edges in an ensemble of random networks (see Methods). (b) Correlation of biological process categories with protein interactions. Each axis represents the category of biological process according to the annotations of Gene Ontology (GO).

with the statistical significance against an ensemble of random network (see Materials and Methods). We observed that there is a statistically significant enrichment of interactions for most of subcellular locations, showing that two proteins interact with each other generally when they reside in the same subcellular compartment. There are two exceptions where the interactions cannot occur between two proteins from within the cytoskeleton or the extracellular space, which is due to the little number of protein interactions in the two subcellular compartments. Some protein interactions are enriched across subcellular compartments. For example, the proteins in the vacuole interact with those not only in the vacuole but also in the endoplasmic reticulum and Golgi apparatus, which can be partly explained by that some protein complexes involved in the endomembrane trafficking pathway are composed of members from the three subcellular compartments. Moreover, current evidence shows that it is difficult to experimentally distinguish proteins in these compartments, which also can explain partially the enrichments in these cases (Dunkley et al., 2006). The other pairs of subcellular compartments with an enrichment of interactions are the plastid/mitochondrion and plastid/extracellular space, which may be due to that some proteins in the plastid are also localized in other compartments (e.g. the mitochondrion) or the experimental error in raw data. As both the mitochondrion and the plastid hold organelle genomes, two organelles share the proteins that are required to translate the genetic information encoded in organelle genomes (e.g. transcription and translation of protein-encoding genes), which can explain the enrichment of interactions between the two subcelluar compartments. The enrichment of interactions across different organelles was also observed in the human interactome (Gandhi et al., 2006). In contrast to the enrichment, we found that there is no significant depletion of interactions for two proteins from different compartments, except those for which one protein is from nuclear and the other from the plasma membrane, mitochondrion or Golgi apparatus. Although the pair of location (nuclear-plasma membrane) is significant depletion (*P*-value: 9.68×10^{-5}) of interactions, we still observed a quantity of interactions (115) across the two subcellular compartments, which could be caused by the experimental inaccuracy of original subcellular localization data.

Biological process for Arabidopsis PPIs

We examined the association between protein interactions and biological functions involving them. Interaction partners are expected to have the same biological function, as they commonly are involved in a same pathway or are members of a protein complex. In humans, there is a significant correlation between interacting proteins and biological functions or disease categories (Gandhi et al., 2006). Our investigation indicates that the significant correlation also holds true in Arabidopsis.

The Gene Ontology (GO) annotations were used as a controlled vocabulary to describe the functions of genes or gene products in a hierarchical manner at three levels (cellular component, biological process and molecular function) (Harris et al., 2004). Thus, two proteins with functional relationship will share common GO terms. For the Arabidopsis interactome, only 58% of interacting proteins shared at least one GO, in term of biological process annotations. We classified the proteins within the Arabidopsis interactome into 9 non–mutually exclusive biological process categories that have the largest number of interactions for both two proteins sharing the same biological process. The enrichment analysis showed that there is a significant enrichment of interactions within almost all the biological process categories (Fig. 2b), indicating that interactions generally occur between proteins involved in the same biological process.

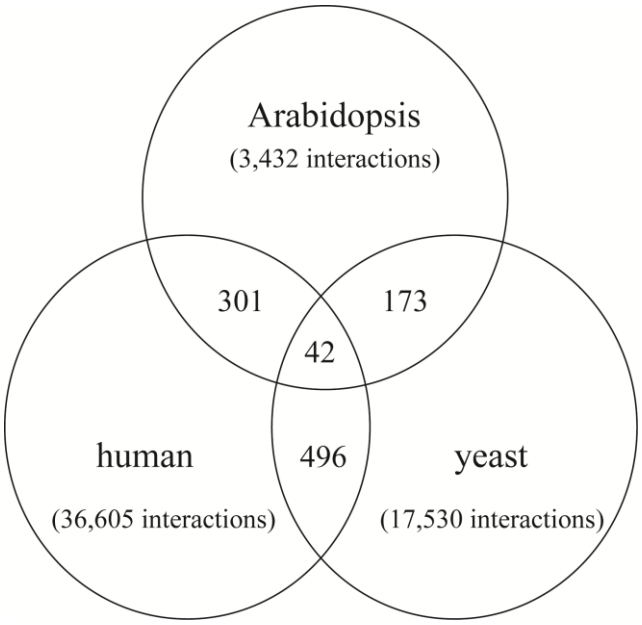


Fig 3. Venn diagram showing overlaps of protein interactions between Arabidopsis, human and yeast datasets. The yeast (*Sacchromyces cerevisia*) interaction dataset was downloaded from the Database of Interacting Proteins (Salwinski et al., 2004), and the human interaction dataset was obtained from the Human Protein Reference Database (Popescu et al., 2008). The interologs between two organisms were identified by protein orthologs defined in the Inparanoid database (O'Brien et al., 2005).

The only exception is the case, where proteins functioning in signal transduction interact with those in protein metabolic process.

Interactome comparison between Arabidopsis and non-plant species

Physical protein-protein interactions conserved across species are called "interologs" (Matthews et al., 2001). The interspecies comparisons of PPI data from yeast, worm, fly and human were carried out to identify conserved PPIs or sub-networks (Gandhi et al., 2006; Mika and Rost, 2006; Saeed and Deane, 2008). We investigated the extent to which plant protein interactions overlap with those reported in the yeast, worm, fly and human datasets. The comparison of PPI datasets among plant, animal and fungi allowed us to identify the common protein interactions and protein complexes in eukaryotic kingdoms.

Among the total number of 1,679 proteins in the Arabidopsis interactome, most of them (1,346 proteins) can be found to have their respective orthologs in animals and fungi. However, there is a small fraction of interactions conserved in at least one species, only roughly 15% (490/3,687) of the Arabidopsis interactome. Particularly, we observed 10% and 6.4% of Arabidopsis interactions conserved in humans and yeast, respectively. We focused on the comparison of the Arabidopsis interactome with the yeast and human interaction datasets, which are considered as the most comprehensive and reliable protein interactomes thus far. As indicated by a Venn diagram, the overlaps between the Arabidopsis, yeast and human interaction datasets (Fig. 3), 42 interactions are common to the Arabidopsis, yeast and human datasets and only one interaction is common to all the five organisms.

Table 2 lists the top 30 interologs ranked by the number of organisms, among which the interactions are conserved (the comprehensive list of the Arabidopsis interologs is provided in Supplementary Table 2). We noticed that most of the highly conserved interactions are dimmers, including homodimers (a complex of two identical protein molecules) and heterodimers (an interaction consisting of two distinct protein molecules). We further found that among these conserved heterodimers, most of them (except for those between CaMs and their target proteins) are composed of two distinct members encoded by a same gene family, which suggests they might be evolved from homodimers through gene duplications. The further investigation showed that dimmers process 37% of 490 conversed interactions, but only 5% of all interactions, suggesting that interactions enriched in complexes are conserved preferentially, compared to transient interactions in the Arabidopsis interactome. These highly conserved protein complexes are mainly involved in the core biological processes, including transcription, transcription regulation, protein phosphorylation, transport and DNA repair, which suggests that they are fundamental and/or vital to all eukaryotes. Further examination showed over 25% of 490 conserved interactions is involved in transcription or transcription regulation. One prominent exception is the homodimer of glutathione S-transferase that is the only interaction common to all the 5 species. It is known that the dimeric enzyme originates from prokaryotes and plays a vital role in the detoxification of both endogenous and xenobiotic compounds. In summary, our analysis indicates that protein complexes involved in the core biological processes may undergo more evolutionary pressure to remain conserved.

The comparison of interactomes among animals, fungi and plants enabled us to identify not only the core protein interactions that may be common and fundamental to eukaryotic kingdoms, but also the species-specific interactions that may provide insights into proteins' evolution and their functional diversity. Our analysis showed that a number of the most highly conserved interactions are between CaMs and their interactors (Table 2). CaM is the most predominant Ca^{2+} sensor present in all eukaryotic cells and plays a crucial role in the regulation of a wide variety of cellular processes by modulating the activities of numerous target proteins. In Arabidopsis, there has been a major effort directed toward the detection of interacting partners of CaMs, and more than 550 targets of CaMs have been identified by either traditional biochemical or high-throughput approaches. The interactions between CaMs and their targets constitute the largest sub-network in the Arabidopsis interactome, which therefore, allows us to carry out a comprehensive comparison of interaction networks of CaM/target between plants and animals. We grouped the CaM targets into several classes according to their gene families and conducted the comparison by representing the CaM target interactions as a bipartite network. As observed from Fig. 4, there is a high divergence of CaM target interactions between Arabidopsis and humans. Among the total number of >700 interactions, only very few (31) are common to the two organisms. These conserved CaM interactors include heat shock protein (HSP) encoded by the HSP90 family, FK506-binding protein (FKBP) belonging to the large peptidyl prolyl *cis/trans* isomerase superfamily, casein kinase and several members of the Ca^{2+}-dependent protein kinase (CDPK or CPK) family. In mammals, these conserved interactions are well-studied and known to be involved in a wide range of

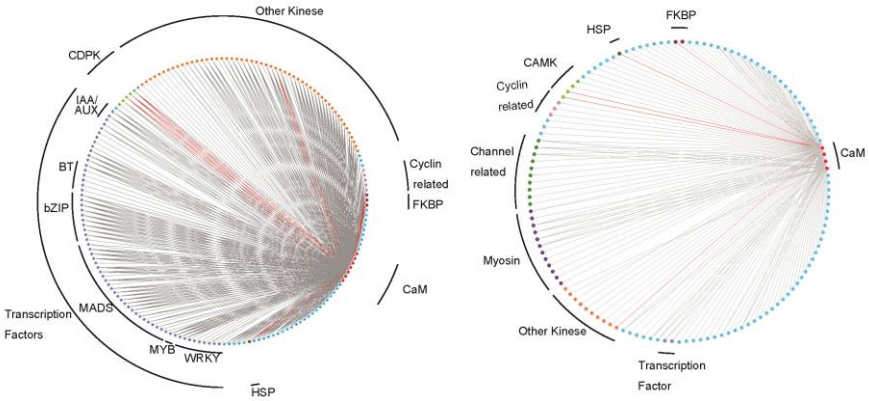


Fig 4. The interaction network of CaMs and their targets. (a) The interaction network of CaMs and their targets in Arabidopsis. The interaction subnetwork extracted from the Arabidopsis interactome and is represented as a bipartite graphic, where proteins (nodes) are classified and highlighted in different colors according to their gene families or superfamilies, and the interactions (edges) conserved between Arabidopsis and humans are highlighted in red. (b) The interaction network of CaMs and targets in humans. The human CaM/target interaction network is extracted from a combined human interaction dataset (see Methods). CaM, calmodulin; CDPK, Ca^{2+}-dependent protein kinase; FKBP, FK506-binding protein; HSP, heat shock protein; CaMK, CaM kinase; BT, BTB and TAZ domain protein; bZIP, basic leucine-zipper transcription factor; MADS, MADS-box transcription factor; MYB, MYB transcription factor; WRKY, WRKY transcription factor.

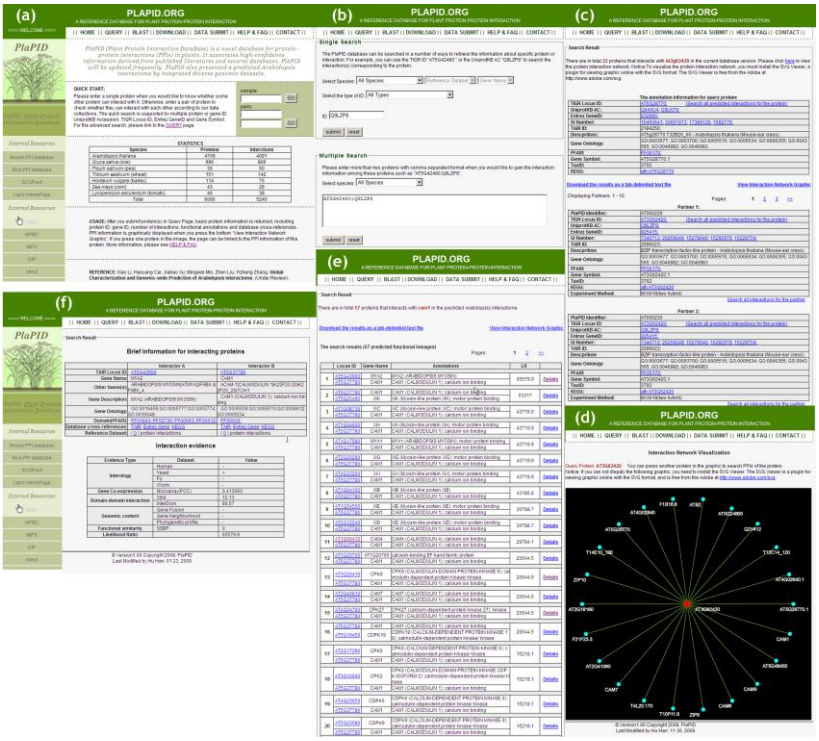


Fig 5. Database and web services of PlaPID. (a) The homepage of PlaPID. (b) The query interface of PlaPID database. The query interface allows users to search interactions among proteins of interest by submitting a variety of gene or protein identifiers (e.g. TAIR LocusID, Swiss-Prot ID and Gene Symbol). Also, users can query interaction information using sequence-based search against a BLAST server. (c) The query results on the reference dataset in PlaPID. The HTML page displays the query results for known interactions among the protein AtBZIP53 (LocusID:At3g62420). The resulted page lists the annotation information of the query protein and interacting proteins, and the detailed information (e.g. experimental method) of each interacting pair as well. (d) Visualization of the interaction network. The interaction network of AtBZIP53 and its' interacting proteins is visualized online as a scalable vector graphics (SVG) file, where the hyperlink on a node can lead to more detailed information of the corresponding interacting protein when clicked. (e) The query results on the predicted Arabidopsis dataset in PlaPID. The HTML page displays the query results for predicted interactions among the protein AtCAM1 (LocusID:At3g62420). (f) Annotation information for the interacting proteins and supporting evidences for the predicted interactions. For each predicted interacting pair, PlaPID providers the basic information (e.g. Gene name, GO annotation and KEGG annotation) for two interacting proteins, and also can show the supporting evidences for interaction between them.

biological processes, but their significances with regard to biochemical or physiological functions in plants are still poorly understood. Involvement in heat shock signal transduction may be the most universal and ancestral role of Ca^{2+}/CaM signal system, where the interaction of CaM with heat stress transcription factor (HSF) is required for induction of the HSP expression (Liu et al., 2008). CaM interaction with both HSP90 and FKBP forms a ternary complex that is conserved in animals and plants, as well as in fungi and certain groups of protists (e.g. *Plasmodium falciparum*). But the physiological relevance of the complex displays differences in distinct eukaryotic kingdoms (Geisler and Bailly, 2007). Reports show that the mammalian FKBP is activated by binding to Ca^{2+}/CaM, which is required for the regulation of Bcl-2 function and; thereby, participates in the promotion of apoptosis in neuronal tissues (Edlich et al., 2005). By contrast, the plant Ca^{2+}/CaM/FKBP complex is thought to be involved in phytohormone function and heat shock response (Geisler and Bailly, 2007).

Interestingly, we found that CaM interactions with several plant CDPKs are interologous to those with animal CaM kinase (CaMKI and CaMK Ⅱ). The plant CDPKs are encoded by a large multigene family and involved in regulation of metabolic enzymes and abiotic and biotic stress signal transduction pathways. The previous studies suggest that the CDPK gene family is widely distributed in the plant kingdom and certain groups of protists (e.g. *Plasmodium*), but not found in other eukaryotic kingdoms, including animals and fungi (Hrabak et al., 2003). However, in the definition of InParanoid database, some members of Arabidopsis CDPKs (e.g. CPK6 and CPK30) are clustered into an orthologous group with the animal CaM kinases, suggesting they are probably homologous.

We performed a multiple alignment of the plant CDPKs with the mammalian CaMKs. Result showed that they have high sequence identity at amino acid level and similar domain architecture (Supplementary Fig. 1). The mammalian CaM kinases as well as their homologs are well-characterized. They are phosphorylated and activated by interacting with Ca^{2+} and one or more CaMs, and constitute a Ca^{2+}/CaM/CaMK signaling cascade that is important for many normal physiological processes and can lead to a variety of disease states when irregulated (Colomer and Means, 2007). However, the biological role played by CaM binding to CDPKs remains to be elucidated because the Ca^{2+}/CaM/CaMK signaling cascade seems lack in Arabidopsis (Colomer and Means, 2007). Several members (e.g. CPK4, CPK11, CPK6 and CPK30) of the Arabidopsis CDPK family are found to be activated in a CaM-independent manner and play an important role in the regulation of abscisic acid signal transduction pathway (Mori et al., 2006; Zhu et al., 2007). Taken together, these data indicates that some CaM/target interactions conserved in both animals and plants are involved in the different biological processes and thereby have significantly functional divergence.

In contrast to the conserved interactions, the majority of CaM/target interactions appear to be organism-specific. In Arabidopsis, the most unique set of CaM targets is composed of proteins encoded by several transcription factor gene families, including MYB, bZIP, MADS and WRKY gene family, whereas the animal CaM is found uncommonly to bind to these types of transcription factors. More and more lines of evidence now indicate that the plant CaM plays key roles in the mediation of cellular responses to developmental and environmental stimuli via regulation of gene expression (Kim et al., 2009). More surprisingly, a study finds that one of the Arabidopsis isoforms (CaM7) can function as transcription factor that directly binds to Z-/G-box elements located in the promoter region of light-responsive genes and; therefore, induces the expression of these genes (Kushwaha et al., 2008). By contrast, many members of the myosin superfamily have been identified as CaM targets, and compose a major target set of the animal CaM, in comparison with the plant CaM. The myosin family of actin-based molecular motors contains one or more IQ motifs that are responsible for binding to CaM (Bahler and Rhoads, 2002). It is well-characterized that the various myosin classes are modulated by Ca^{2+}/CaM and carry out a diversity of cell and physiological functions, such as smooth muscle contraction (Ding et al., 2009). As Ca^{2+} is the most prominent second messengers and CaM is the most major decoder of Ca^{2+} signals in both plants and animals, the binding specificity of CaM may reflect their discrepancy in responses to environmental stimuli modulated by Ca^{2+} and CaM signal pathway. In contrast to animals that can be adapted to environment through their movements, plants, as sessile organisms, are more likely to employ differential gene expression in response to changes from environment stimuli, which partially is supported by a great number of transcription factors identified as CaM target proteins in Arabidopsis.

Database and web server

To facilitate the research in plant, we also constructed a database server called PlaPID (Plant Protein Interaction Database). PlaPID was implemented to deposit the putative interactions and known interactions. In addition to Arabidopsis, protein interactions detected experimentally in other 6 plant species, including rice (*Oryza sativa*), pea (*Pisum sativum*), wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), corn (*Zea mays*) and tomato (*Lycopersicon esculentum*), were also collected and deposited into our database, and thus, PlaPID could be an integrative reference database of plant protein interactions. To facilitate plant researchers, we had also developed a user-friendly web server to search and download PPI data in the database (Fig. 5). The query interface allows users to search for interactions among proteins of interest by using a variety of gene identifiers or by sequence-based search against a BLAST server. The result interactions and their annotation information will be displayed via HTML pages. Moreover, the interaction network can be visualized online as a SVG (scalable vector graphics) file, which provides the means for a fast, visual evaluation of the protein's interaction environment and allows for a navigation of the protein interaction network. For putative interactions, PlaPID also provides predictive information on supporting evidences and likelihood ratios, which enables users to assess the reliability of interactions and choose more reliable interactions by setting higher cutoff values of likelihood ratios. All interaction data presented in PlaPID can be downloaded freely with tab-delimited or PSI-MI standard format files through our web interface.

Materials and Methods

Datasets

We extracted the Arabidopsis PPI data stored in several public-available databases such as Biomolecular Interaction Network Database (BIND; 1530 interactions) (Bader et al., 2003), Database of Interacting Proteins (DIP; 13 interactions) (Salwinski et al., 2004), Molecular Interactions Database (MINT; 91 interactions) (Chatr-aryamontri et al., 2007), The Arabidopsis Information Resource (TAIR; 885 interactions) (Swarbreck et al., 2008) and IntAct (2,538 interactions) (Kerrien et al., 2007). All PPI data were downloaded in March 2008. The protein structure and functional annotation information were obtained from the TAIR website (Swarbreck

et al., 2008). The Arabidopsis protein subcellular localizations were defined in the SUBA database (Heazlewood et al., 2007). The Arabidopsis lethal gene dataset was obtained from Kocabek et al. (2006), where the total of 780 genes were defined as lethal genes, 245 of which were contained in the the Arabidopsis interactome.

The human interaction data contained one literature-curated dataset from Human Protein Reference Database (HPRD; 32,345 interactions) (Prasad et al., 2008), two Y2H datasets generated by Rual et al. (3,182 interactions) (Rual et al., 2005) and Stelzl et al. (2,754 interactions) (Stelzl et al., 2005), respectively, and one MS dataset by Ewing et al. (4,248 interactions) (Ewing et al., 2007). We incorporated these human interaction datasets into a combined human PPI dataset that was composed of 41,898 PPIs among 10,041 proteins identified uniquely by Entrez GeneIDs. The yeast, worm and fly interaction data were downloaded from the Database of Interacting Proteins (DIP) (Salwinski et al., 2004). The protein ortholog dataset was downloaded from the InParaniod database (version 7.0) (O'Brien et al., 2005).

Network visualization and statistical analysis

The Arabidopsis protein interaction networks were visualized by the Cytoscape software, and the topological features were analyzed by using the NetAnalysis, a java plug-in of Cytoscape (Shannon et al., 2003; Barsky et al., 2007). The topological features of a protein (node) in a network are commonly characterized by two parameters, degree and clustering coefficient. In a network, the degree of a protein is the number of interactions (edges) linking it; and the clustering coefficient (C_i) of a protein represents the ratio of the number of links between a node's neighbors to the number of edges of the node. The average clustering coefficient of a network is calculated as follows:

$$C_i = \sum_{i}^{n} \frac{2e_i}{k_i(k_i - 1)} / n$$

Where, i is the ^{i}th node among the total number (n) of nodes, k_i is the degree of the node i, and e_i is the number of edges existing between the k_i nodes connected to the node i.

To visualize the network, we used a winner-takes-all method of Gandhi et al. (2006) to assign a single subcellular compartment to every protein, because some proteins have more than one subcellular localization. In the SUBA database, 6,345 proteins are annotated to the 12 distinct subcellular compartments (plasma membrane, vacuole, cytosol, nucleus, cytoskeleton, extracellular space, mitochondrion, plastid, endoplasmic reticulum, cell plate, Golgi and peroxisome) by direct or indirect experimental evidence derived from 5 sources of information (GFP fusion experiments, mass spectrometry studies, AmiGO annotation, Swiss-Prot annotation, and TAIR gene descriptions). If a protein is annotated to multiple subcellular localizations, the protein is designated in a unique compartment that is supported by the most sources of evidence. The statistical difference of degree distributions between lethal and non-lethal genes was analyzed using the Wilcoxon rank sum test (also called as "Mann-Whitney test"), which is a non-parametric statistical hypothesis test for the case of two related samples (Hollander and Wolfe, 1973).

Enrichment analysis

Following the methods of Gandhi et al. (Gandhi et al., 2006), we performed statistical tests for evaluating the enrichment of the Arabidopsis protein interactions in the observed protein network with respect to subcellular localization and biological function against those in an ensemble of random networks. The ensemble of random protein network has the same properties with the observed network as following: (1) the total number of edges, (2) the degree (k) of each node, and (3) the annotation of every protein in their subcellular localizations and biological functions. The edges were grouped into non–mutually exclusive categories with regard to protein annotation types (subcellular localization or biological process). The statistical significance (P-value) of the observed number of edges for a pair of categories ($n_{\alpha\beta}$), where a protein belongs to the category α and other to the category β, was calculated by a Poisson distribution:

$$P(n_{\alpha\beta}) = \begin{cases} \sum_{j=0}^{n} \bar{n}_{\alpha\beta}{}^{j} \exp(-\bar{n}_{\alpha\beta}) / j!, n_{\alpha\beta} < \bar{n}_{\alpha\beta} \\ \sum_{j=n}^{\infty} \bar{n}_{\alpha\beta}{}^{j} \exp(-\bar{n}_{\alpha\beta}) / j!, n_{\alpha\beta} \geq \bar{n}_{\alpha\beta} \end{cases}$$

Where, $\bar{n}_{\alpha\beta}$ is the expected number of edges (a node belongs to the category α and other to the category β) in the ensemble of random network. Finally, we applied a multiple-testing correction as $P = 1-(1-P)^m$, where P is the single-test P value and m is the number of tests. The tests of enrichment (m = number of $\alpha\beta$ pairs with at least one edge in the observed network) and depletion (m = number of $\alpha\beta$ pairs possible in the randomized networks) were tabulated separately. For more details on the analysis, please refer to Methods in Gandhi et al. (2006).

Interactome comparison

We queried the interaction data of three model organisms (yeast, worm and fly) from the DIP database (Salwinski et al., 2004) and downloaded the literature-curated human interaction dataset from the HPRD (Prasad et al., 2008).

To compare the interractome between Arabidopsis and non-plant species, a tough task was to identify orthologs accurately. We used the InParanoid algorithm, which is based on an all-versus-all BLAST search following by clustering into orthologous groups, to identify orthologs (O'Brien et al., 2005). These orthologs were then used for further analysis in this research.

Conclusions

As more and more genomes have been sequenced completely and large-scale 'omic' data are available, systems biology seems to come of age in plant science. The study for interactome is quickly becoming a key area of systems biology. By treating the presently available interaction data as a proxy, we have carried out a comprehensive investigation on the global topological and biological properties of the Arabidopsis interactome. Our investigation indicates that the Arabidopsis interactome has similar global features with as observed previously in animals (humans, fly and worm) and fungi (yeast), but a high evolutionary divergence that only a small fraction of interactions are conserved in non-plant model organisms.

Acknowledgments

We are grateful to Dr. Han Hu for his assistance on the database and web service construction. We also thank Prof. Yongsheng Liu, Dr. Xiangli Niu and Dr. Haiyan Wang for reading the

manuscript and their useful suggestions. This work was supported partially by National Natural Science Foundation of China (Grant No. 61001149) and the National Science and Technology Major Project of the Ministry of Science and Technology of China (Grant No. 2012ZX10005001-010).

References

Albert R, Jeong H, Barabasi AL (2000) Error and attack tolerance of complex networks. Nature. 406(6794): 378-382.

Bader GD, Betel D, Hogue CW (2003) BIND: the Biomolecular Interaction Network Database. Nucleic Acids Res. 31(1): 248-250.

Bahler M, Rhoads A (2002) Calmodulin signaling via the IQ motif. FEBS Lett. 513(1): 107-113.

Barabasi AL, Albert R (1999) Emergence of scaling in random networks. Science. 286(5439): 509-512.

Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet. 5(2): 101-113.

Barsky A, Gardy JL, Hancock RE, Munzner T (2007) Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation. Bioinformatics. 23(8): 1040-1042.

Chandler JW, Cole M, Flier A, Werr W (2009) BIM1, a bHLH protein involved in brassinosteroid signalling, controls Arabidopsis embryonic patterning via interaction with DORNROSCHEN and DORNROSCHEN-LIKE. Plant Mol Biol. 69(1-2): 57-68.

Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, Cesareni G (2007) MINT: the Molecular INTeraction database. Nucleic Acids Res. 35(Database issue): D572-574.

Colombo M, Masiero S, Vanzulli S, Lardelli P, Kater MM, Colombo L (2008) AGL23, a type I MADS-box gene that controls female gametophyte and embryo development in Arabidopsis. Plant J. 54:1037-1048

Colomer J, Means AR (2007) Physiological roles of the Ca2+/CaM-dependent protein kinase cascade in health and disease. Subcell Biochem. 45: 169-214.

Ding HL, Ryder JW, Stull JT, Kamm KE (2009) Signaling processes for initiating smooth muscle contraction upon neural stimulation. J Biol Chem. 284(23): 15541-15548.

Dunkley TPJ, Hester S, Shadforth IP, Runions J, Weimar T, Hanton SL, Griffin JL, Bessant C, Brandizzi F, Hawes C, Watson RB, Dupree P, Lilley KS (2006) Mapping the Arabidopsis organelle proteome. Proc Natl Acad Sci USA. 103(17): 6518-6523.

Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W (2005) BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics. 21(16): 3439-3440.

Edlich F, Weiwad M, Erdmann F, Fanghanel J, Jarczowski F, Rahfeld JU, Fischer G (2005) Bcl-2 regulator FKBP38 is activated by Ca2+/calmodulin. Embo J. 24(14): 2688-2699.

Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, McBroom-Cerajewski L, Robinson MD, O'Connor L, Li M, Taylor R, Dharsee M, Ho Y, Heilbut A, Moore L, Zhang S, Ornatsky O, Bukhman YV, Ethier M, Sheng Y, Vasilescu J, Abu-Farha M, Lambert JP, Duewel HS, Stewart II, Kuehl B, Hogue K, Colwill K, Gladwish K, Muskat B, Kinach R, Adams SL, Moran MF, Morin GB, Topaloglou T, Figeys D (2007) Large-scale mapping of human protein-protein interactions by mass spectrometry. Mol Syst Biol. 3: 89.

Fraser HB, Hirsh AE, Wall DP, Eisen MB (2004) Coevolution of gene expression among interacting proteins. Proc Natl Acad Sci USA. 101(24): 9033-9038.

Gandhi TK, Zhong J, Mathivanan S, Karthick L, Chandrika KN, Mohan SS, Sharma S, Pinkert S, Nagaraju S, Periaswamy B, Mishra G, Nandakumar K, Shen B, Deshpande N, Nayak R, Sarker M, Boeke JD, Parmigiani G, Schultz J, Bader JS, Pandey A (2006) Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. Nat Genet. 38(3): 285-293.

Geisler M, Bailly A (2007) Tête-à-tête: the function of FKBPs in plant development. Trends Plant Sci. 12(10): 465-473.

Grigoriev A (2001) A relationship between gene expression and protein interactions on the proteome scale: analysis of the bacteriophage T7 and the yeast Saccharomyces cerevisiae. Nucleic Acids Res. 29(17): 3513-3519.

Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M, Wood V, de la Cruz N, Tonellato P, Jaiswal P, Seigfried T, White R (2004) The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 32(Database issue): D258-261.

Heazlewood JL, Verboom RE, Tonti-Filippini J, Small I, Millar AH (2007) SUBA: the Arabidopsis Subcellular Database. Nucleic Acids Res. 35(Database issue): D213-218.

Hollander M, Wolfe DA (1973). Nonparametric Statistical Inference. New York, John Wiley & Sons.

Hrabak EM, Chan CW, Gribskov M, Harper JF, Choi JH, Halford N, Kudla J, Luan S, Nimmo HG, Sussman MR, Thomas M, Walker-Simmons K, Zhu JK, Harmon AC (2003) The Arabidopsis CDPK-SnRK superfamily of protein kinases. Plant Physiol. 132(2): 666-680.

Ideker T, Sharan R (2008) Protein networks in disease. Genome Res 18(4): 644-652.

Jeong H, Mason SP, Barabasi AL, Oltvai ZN (2001) Lethality and centrality in protein networks. Nature. 411(6833): 41-42.

Jeong B, Tombor B, Albert R, Oltvai ZN, Barabasi AL (2000) The large-scale organization of metabolic networks. Nature. 407(6804): 651-654.

Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A, Derow C, Dimmer E, Feuermann M, Friedrichsen A, Huntley R, Kohler C, Khadake J, Leroy C, Liban A, Lieftink C, Montecchi-Palazzi L, Orchard S, Risse J, Robbe K, Roechert B, Thorneycroft D, Zhang Y, Apweiler R, Hermjakob H (2007) IntAct--open source resource for molecular interaction data. Nucleic Acids Res. 35(Database issue): D561-565.

Kim MC, Chung WS, Yun DJ, Cho MJ (2009) Calcium and Calmodulin-mediated regulation of gene expression in plants. Mol Plant. 2(1): 13-21.

Kocabek T, Repkova J, Dudova M, Hoyerova K, Vrba L (2006) Isolation and characterization of a novel semi-lethal Arabidopsis thaliana mutant of gene for pentatricopeptide (PPR) repeat-containing protein. Genetica. 128(1-3): 395-407.

Kushwaha R, Singh A, Chattopadhyay S (2008) Calmodulin7 plays an important role as transcriptional regulator in Arabidopsis seedling development. Plant Cell. 20(7): 1747-1759.

Li D, Li J, Ouyang S, Wang J, Wu S, Wan P, Zhu Y, Xu X, He F (2006) Protein interaction networks of Saccharomyces cerevisiae, Caenorhabditis elegans and Drosophila melanogaster: large-scale organization and robustness.

Proteomics. 6(2): 456-461.
Liu HT, Gao F, Li GL, Han JL, Liu DL, Sun DY, Zhou RG (2008) The calmodulin-binding protein kinase 3 is part of heat-shock signal transduction in Arabidopsis thaliana. Plant J. 55(5): 760-773.
Matthews LR, Vaglio P, Reboul J, Ge H, Davis BP, Garrels J, Vincent S, Vidal M (2001) Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs". Genome Res. 11(12): 2120-2126.
Mika S, Rost B (2006) Protein-protein interactions more conserved within species than across species. PLoS Comput Biol. 2(7): e79.
Morell M, Aviles FX, Ventura S (2009) Detecting and interfering protein interactions: towards the control of biochemical pathways. Curr Med Chem. 16(3): 362-379.
Mori IC, Murata Y, Yang Y, Munemasa S, Wang YF, Andreoli S, Tiriac H, Alonso JM, Harper JF, Ecker JR, Kwak JM, Schroeder JI (2006) CDPKs CPK6 and CPK3 function in ABA regulation of guard cell S-type anion- and Ca(2+)-permeable channels and stomatal closure. PLoS Biol. 4(10): e327.
Morsy M, Gouthu S, Orchard S, Thorneycroft D, Harper JF, Mittler R, Cushman JC (2008) Charting plant interactomes: possibilities and challenges. Trends Plant Sci. 13(4): 183-191.
O'Brien KP, Remm M, Sonnhammer EL (2005) Inparanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Res. 33(Database issue): D476-480.
Park J, Lee DS, Christakis NA, Barabasi AL (2009) The impact of cellular networks on disease comorbidity. Mol Syst Biol. 5: 262.
Popescu SC, Popescu GV, Bachan B, Zhang Z, Seay M, Gerstein M, Snyder M, Dinesh-Kumar SP (2007) Differential binding of calmodulin-related proteins to their targets revealed through high-density Arabidopsis protein microarrays. Proc Natl Acad Sci U S A. 104(11): 4730-4735.
Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal, A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, Ray S, Kishore CJ, Kanth S, Ahmed M, Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S, Ranganathan P, Ramabadran S, Chaerkady R, Pandey A (2008) Human Protein Reference Database--2009 update. Nucleic Acids Res. 37(Database issue): D767–772.
Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamosas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE, Roth FP, Vidal M (2005) Towards a proteome-scale map of the human protein-protein interaction network. Nature. 437(7062): 1173-1178.
Saeed R, Deane C (2008) An assessment of the uses of homologous interactions. Bioinformatics. 24(5): 689-695.
Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D (2004) The Database of Interacting Proteins: 2004 update. Nucleic Acids Res. 32(Database issue): D449-451.
Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13(11): 2498-2504.
Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksoz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker EE (2005) A human protein-protein interaction network: a resource for annotating the proteome. Cell. 122(6): 957-968.
Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res. 36(Database issue): D1009-1014.
Uhrig JF (2006) Protein interaction networks in plants. Planta. 224(4): 771-781.
Weltmeier F, Rahmani F, Ehlert A, Dietrich K, Schutze K, Wang X, Chaban C, Hanson J, Teige M, Harter K, Vicente-Carbajosa J, Smeekens S, Droge-Laser W (2009) Expression patterns within the Arabidopsis C/S1 bZIP transcription factor network: availability of heterodimerization partners controls gene expression during stress response and development. Plant Mol Biol. 69(1-2): 107-119.
Wodak SJ, Pu S, Vlasblom J, Seraphin B (2009) Challenges and rewards of interaction proteomics. Mol Cell Proteomics. 8(1): 3-18.
Yuan JS, Galbraith DW, Dai SY, Griffin P, Stewart CN (2008) Plant systems biology comes of age. Trends Plant Sci. 13(4): 165-171.
Zhu SY, Yu XC, Wang XJ, Zhao R, Li Y, Fan RC, Shang Y, Du SY, Wang XF, Wu FQ, Xu YH, Zhang XY, Zhang DP (2007) Two calcium-dependent protein kinases, CPK4 and CPK11, regulate abscisic acid signal transduction in Arabidopsis. Plant Cell. 19(10): 3019-3036.