A predicted protein interactome identifies conserved global networks and disease resistance subnetworks in maize

Bryan Musungu1, Deepak Bhatnagar2, Robert L. Brown2, Ahmad M. Fakhoury3 and Matt Geisler1*

1 Department of Plant Biology, Southern Illinois University, Carbondale, IL, USA, 2 Food and Feed Safety Research, Southern Regional Research Center, United States Department of Agriculture, Agricultural Research Service, New Orleans, LA, USA, 3 Department of Plant Soil and Agriculture Systems, Southern Illinois University, Carbondale, IL, USA

Interactomes are genome-wide roadmaps of protein-protein interactions. They have been produced for humans, yeast, the fruit fly, and Arabidopsis thaliana and have become invaluable tools for generating and testing hypotheses. A predicted interactome for Zea mays (PiZeaM) is presented here as an aid to the research community for this valuable crop species. PiZeaM was built using a proven method of interologs (interacting orthologs) that were identified using both one-to-one and many-to-many orthology between genomes of maize and reference species. Where both maize orthologs occurred for an experimentally determined interaction in the reference species, we predicted a likely interaction in maize. A total of 49,026 unique interactions for 6004 maize proteins were predicted. These interactions are enriched for processes that are evolutionarily conserved, but include many otherwise poorly annotated proteins in maize. The predicted maize interactions were further analyzed by comparing annotation of interacting proteins, including different layers of ontology. A map of pairwise gene co-expression was also generated and compared to predicted interactions. Two global subnetworks were constructed for highly conserved interactions. These subnetworks showed clear clustering of proteins by function. Another subnetwork was created for disease response using a bait and prey strategy to capture interacting partners for proteins that respond to other organisms. Closer examination of this subnetwork revealed the connectivity between biotic and abiotic hormone stress pathways. We believe PiZeaM will provide a useful tool for the prediction of protein function and analysis of pathways for Z. mays researchers and is presented in this paper as a reference tool for the exploration of protein interactions in maize.

Keywords: Zea mays, interactome, network, protein-protein interactions, disease resistance, tool
Introduction

Understanding the biological interactions within an organism is vital for the comprehension of its functions. The interactome provides a large scale mapping of protein-protein interactions (PPIs). Interactomes of model organisms such as Arabidopsis thaliana and Saccharomyces cerevisiae were built using high throughput experimental methodologies (Consortium, 2011). However, predicted interactomes in species of agronomic importance, like Citrus sinensis, Oryza sativa, and Glycine max, have provided insight into disease resistance. There was no plant interactome until the Arabidopsis predicted interactome was released in 2007 (Geisler-Lee et al., 2007). It was based on orthologs (genes separated by speciation) of S. cerevisiae (Yu et al., 2008), Drosophila melanogaster (Giot et al., 2003), Caenorhabditis elegans (Li et al., 2004), and Homo sapiens (Rual et al., 2005). This predicted plant interactome successfully provided hypotheses for testing interactions, including those involving membrane proteins, which are otherwise difficult to elucidate using forward and reverse genetic approaches (Lalonde et al., 2010; Nejad et al., 2012). Although experiment-based interactomes for A. thaliana are now being made (Consortium, 2011; Chen et al., 2012), the predicted interactome still makes many useful predictions for interactions not yet found in the growing experimental dataset. For instance, studies by Guo et al. (2009) which tackled the complexity of germination and the involvement of plant hormone pathways, found interacting partners of Rack1 (receptor for activated kinases1) from a candidate list of 88 partners using a predicted interactome. Plant predicted interactomes have also aided in determining proteins involved in resistance to the destructive bacterial pathogen Huanglongbing in citrus (Martinelli et al., 2012, 2013), as well as to the soybean cyst nematode (SCN) in soybean (Lightfoot, 2014). Moreover, the human interactome was used to link the differential expression of genes with protein interactions in the analysis of cancer tissues, allowing researchers to analyze the connectivity between known and novel targets (Wachi et al., 2005). Thus, interactomes allow for hypotheses to be generated with a posteriori and a priori knowledge of a biological system.

The underlying principle for a predicted interactome is that evolutionarily conserved proteins tend to have conserved interactions when the proteins retain orthologous functions. Software programs such as Inparanoid (Ostlund et al., 2010), OrthoMCL (Li et al., 2003), and MSOAR (Geer et al., 2006), along with many others, have been developed in order to discover all orthologs and outparalogs (duplications prior to divergence of species) between two or more genomes, and to separate these from inparalogs (duplication within a lineage). PPIs can thus be predicted across an entire genome by high throughput computational methods using whole genome ortholog prediction (Geiser-Lee et al., 2007; Schuette et al., 2015). These methods have been successfully used to predict interactomes for A. thaliana, G. max, Coffea robusta, H. sapiens, C. sinensis, D. melanogaster, O. sativa (rice), and P. patens (a moss) (Giot et al., 2003; Li et al., 2004; Brown and Jurisica, 2005; Rual et al., 2005; Guan et al., 2008; Consortium, 2011; Geisler and Fitzek, 2011; Gu et al., 2011; Ho et al., 2012; Ding et al., 2014; Lightfoot, 2014). Moreover, physically interacting proteins tend to be encoded by genes co-expressed in response to different stimuli in many species (Giot et al., 2003; Bhardwaj and Lu, 2005; Rual et al., 2005). Expression data, such as microarray and RNA-Seq, can thus be used as an additional layer of support for PPIs predicted through orthology.

An interactome can be visualized as a field of circles (nodes) that represent proteins and connections (edges) between nodes representing PPIs. Each node can be rated based on the number of connections, referred to as the connectivity or degree of that node. Protein interactomes typically contain a few highly connected hubs (proteins with >10 partners) and numerous smaller hubs (proteins with 3–10 partners), pipes (2 partners) and free ends (1 partner only). This distribution is an inverse power relationship between node frequency and connectivity, and is similar to that of other small-world network structures such as social networks and electrical power grids (Watts and Strogatz, 1998; De Silva et al., 2006; Gu et al., 2011). The small-world topology is a compromise between efficiency and robustness. Having fewer interacting partners involved in a pathway results in increased efficiency in terms of how fast a product or outcome can be processed. The highly connected hubs represent proteins that are conserved through different organisms and are under less selection pressure for mutations (Batada et al., 2006; Zotenko et al., 2008; Ning et al., 2010). For example, heat shock proteins (i.e. Hsp70) are some of the highly connected hubs in most interactome networks (Gopinath et al., 2014). These proteins are vital for helping other proteins to fold properly, assemble, and translocate. They have also been implicated in reactions to abiotic and biotic stress (Wang et al., 2004). A robust Hsp70 system includes redundant pathways, auto-regulation, and feedback for increased stability.

Zea mays (maize) is one of the three major global crops (FAOstat, 2009; Ranum et al., 2014). Although its genome was sequenced (Schnable et al., 2009) a few years ago, it still contains many poorly annotated genes with provisional functional annotation based on sequence or domain homology. In addition, proteins with a known biological role in a species are often found to have, in another species, even more biological roles that were previously unknown or overlooked. A systematic approach to the analysis of the connectivity between known and novel proteins will help identify these respective biological roles, and will add to existing Z. mays gene annotation. PiZeaM is thus a useful tool to identify networks of connections among proteins in many key Z. mays processes. The recent increase in the amount of genome and interactome data for model organisms has made it possible to systematically predict PPIs in Z. mays from experimentally determined reference PPIs maps.

In addition to gene discovery and functional annotation, the ultimate goal in systems biology is to gain an understanding of the networks of molecular processes in an organism. These systems biology efforts at network analysis have begun in model organisms such as A. thaliana (Consortium, 2011), and Synechocystis sp. PCC6803 (Kim et al., 2008). However, where these efforts are urgently needed is in the major cereal crops. Z. mays is one of the three major food crops. Moreover, over the...
last decade, there has been an increase in the use of Z. mays to produce the biofuel ethanol as a substitute to gasoline. In fact, it has been estimated by the Renewable Fuels Association that 26% of the Z. mays grown in the US in 2012 was used for the production of ethanol (RFA, 2012). Increasing the yield in Z. mays through better management practices and improved hybrid genetics is a goal actively sought by the researchers around the world.

In this work we present PiZeaM—a predicted Z. mays interactome—to help functional annotation of gene encoding proteins, to understand their connectivity and to aid systems biology approaches. This interactome can expand knowledge of protein functions by analysis of interacting proteins’ connectivity and by providing maps of interconnected pathways. We demonstrated the accuracy of the published prediction method by comparing the predicted A. thaliana interactome in 2007 with the experimentally determined interactomes published in 2008–2014. We believe PiZeaM will provide a useful tool for the prediction of protein function and analysis of pathways for Z. mays researchers.

Materials and Methods

Choice of Prediction Method
The interolog method was chosen to predict protein interactions in Z. mays due to its success in other plant species (Gu et al., 2011; Ho et al., 2012), including O. sativa, another member of the grass family. Bioinformatic algorithms and programs, and corresponding parameters and weights used to produce this interactome were the same as those used in O. sativa (Ho et al., 2012).

Sources of Protein Sequences
Amino acid sequences from Z. mays (ZmB73 4a.53 sequence) were retrieved from the GRAMENE database (http://www.gramene.org; Chandler and Brendel, 2002; Youens-Clark et al., 2011). The accession numbers were later mapped onto a more recent version ZmB73 5a to gain access to updated annotation tools in Phytozome (phytozome.jgi.doe.gov). Those of the 13 reference species—nine eukaryotes Orzya sativa, Arabidopsis thaliana, Homo sapiens, Mus musculus, Rattus norvegicus, Drosophila melanogaster, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and four prokaryotes Escherichia coli, Bacillus subtilis, Helicobacter pylori, Campylobacter jejuni, and Synechococcus—were retrieved from ENSEMBL www.ensembl.org/index.html (access date November 2011), (Flicke et al., 2012) and NCBI http://www.ncbi.nlm.nih.gov, (Geer et al., 2010).

Prediction of Orthologs
The amino acid sequences were then processed with the Linux based program Inparanoid 3.0 (Ostlund et al., 2010), BLOSUM (BLOcks SUbstitution Matrix) 80 was used for three plant species, A. thaliana, O. sativa, and Z. mays; BLOSUM 62 for five animal species C. elegans, D. melanogaster, H. sapiens, M. musculus, R. norvegicus, and two fungal species S. pombe, and S. cerevisiae; and BLOSUM 45 for the four prokaryotes B. subtilis, C. jejuni, E. coli, and Synechococcus. One ortholog pair with a bootstrap score of 100% was chosen from each ortholog cluster for the one-to-one interactome. A second interactome was built with all pairwise combination of proteins in each ortholog cluster referred to as the many-to-many interactome. The latter allowed all potential paralogs in both Z. mays and reference species to substitute for one another, while outparalogs are in separate gene clusters. Outparalogs will retain interactions within their cluster but not to other outparalogs. Gene orthology, unlike homology by blast score or “% similarity,” is considered to be a better predictor of the conservation of function, and thus conservation of protein interactions (Koonin, 2005; Ostlund et al., 2010). The number of orthologs found between Z. mays and the reference organisms was dependent on the evolutionary distance, and often more distantly related organisms (like bacteria and animals) had fewer orthologs with Z. mays.

Predicting Maize Interactions from Conserved Orthologs
PiZeaM was built from a comprehensive analysis of physical interactions between proteins of Z. mays that were predicted based on experimentally determined interactions from five major interactome databases [BioGRID (version 3.1.84; www.thebiogrid.org) (Stark et al., 2006); DIP (November 2011 release) (Salwinski et al., 2004); IntAct (downloaded November 5, 2011; http://www.ebi.ac.uk/intact) (Aranda et al., 2010); and MINT (downloaded November 5, 2011; http://mint.bio.uniroma2.it/mint) (Chatr-Aryamontri et al., 2007)]. The interactome thus consisted of a large set of data, listing pairs of interacting proteins (see Supplemental Table 1). The entire interactome can be visualized graphically using the Cytoscape software packages (Shannon et al., 2003; Cline et al., 2007) in which proteins are indicated by 2 dimensional shapes (circles) and interactions are indicated by connecting lines. Heterologous interactions were ordered (larger ID as protein A: smaller ID as protein B) as all interactions were considered bidirectional. Duplicates were then removed from the master database to produce a unique interaction table where each interaction is represented once.

Calculation of the Confidence Value (CV) of Experimental Support
To determine the confidence value (CV), the following formula was used (Geisler-Lee et al., 2007):

\[ CV = N \times S \times E \]

N: the number of times that an interaction appeared in PiZeaM.
E: the number of times an experiment appeared with an interaction in the mined interactomes.
S: the number of times an interaction appeared in the reference species.

A scale was then used to rank the CV of the interactions that comprise the interactome. This last factor (E) was an important consideration as different methods for determining protein interactions will not always give converging results rates.
(for example yeast-2-hybrid false positives are not likely to be the same as those that occur using in vitro co-precipitation or peptide arrays)(Yu et al., 2008). Protein pairs that had a CV = 1 were considered to have low confidence interactions, a CV = 2–10 was considered to indicate a medium confidence interaction, and a CV > 10 was considered to represent a high confidence interaction. Interactions with the highest CVs were more likely to be conserved across species.

Mapping of Maize Orthologs to Gene Ontology
Gene ontology (GO) information from Z. mays was downloaded from ENSEMBL (www.ensembl.org/index.html) (Plicek et al., 2012), Phytozome Version 10.1 (http://phytozome.jgi.doe.gov/pz/portal.html), and GRAMENE Release 45 (http://www.gramene.org). These GO terms described the biological function (process), subcellular localization (compartment) and molecular function of each Z. mays protein, based on experimental evidence or on prediction based on protein domain composition, signal sequences and global homology (Carbon et al., 2009). Then, a custom ontology file and a reference annotation were developed for updated annotation. GO-term enrichment analysis of the interactions within PiZeaM was conducted by comparing proteins within the interactome (6004 proteins) with the whole genome of Z. mays. This comparison established which biological functions were captured by the aforementioned interolog method and determined which types of proteins were conserved as interacting ortholog pairs. A second GO comparison was made of 1st neighbors (prey) of proteins with the GO term “response to other organism” (GO: 0051707) (bait). The Cytoscape Bingo plugin was then used to look at enrichment in the predicted interactome contrasted with the whole annotation (Maere et al., 2005). The GO terms were then sorted by their enrichment or depletion P-value, and for terms reflecting molecular function, biological process, and cellular component (Supplemental Table 3).

Co-Expression Analysis
For expression data analysis and determination of co-expression between various genes, microarray data was downloaded from the Gene Expression Omnibus (Edgar et al., 2002) from 68 biotic and abiotic stress data sets generated using the same Affymetrix Z. mays Gene chip (see Supplemental Table 2). To generate a co-expression matrix, the change in levels of mRNA was normalized using global intensity normalization and computed as an M-value (Log base 2 of the ratio of stimulus over control).

\[ M - value = \log_2 \left(\frac{\text{treatment}}{\text{control}}\right) \]

Pearson correlation was calculated for each of the data sets using the R statistical language (Edgar et al., 2002; Stuart et al., 2003; Bhardwaj and Lu, 2005). The correlation function was used to determine the Pearson correlation between random proteins and true interologs. Pearson correlation coefficient (r):

\[ r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]} \]

N: the number of expression samples.
X: the expression level of gene X in ith sample.
Y: the expression level of gene Y in the ith sample.

The value of \( r \) is between \(-1 \) and 1 (i.e. \(-1 < r < 1\)). A Pearson correlation coefficient (r) was generated for each pair of interacting proteins in the interactome. Interacting proteins that had a high \( r \)-value were considered to be more likely to be coordinately co-expressed (Narayanan et al., 2010).

Visualization of PiZeaM
One objective of this study was to make PiZeaM readily available for other researchers to use. To visualize PiZeaM, Cytoscape 3 (Clines et al., 2007) can be used by importing the provided Excel files (Supplemental Table 1), or by opening the compiled Cytoscape file (Supplemental File 1). The information can be easily stored and manipulated in MySQL (a database language) and Microsoft Excel to address different research needs. To visualize a protein of choice from a pathway, only requires retrieving protein IDs for the Z. mays proteins of choice from Ensembl. One limitation of the interactome is a possible under-representation of interactions exclusive to Z. mays, since a substantial amount of the data used in this study to build the interactome was derived from interactions reported in non-grasses. This limitation can be alleviated in the future by incorporating more data from future research involving grasses and cereals, once these experimental datasets are built and shared with the public.

Results

Overview of PiZeaM
Using one-to-one interolog as a method of predicting interactions, a total of 34,107 unique interactions were found for 4843 Z. mays proteins from the 110,185 Z. mays proteins tested (Supplemental Table 1). A summary of the relative contribution of each reference species to the predicted interactions is shown in (Table 1). PiZeaM represented less than 5% of the Z. mays proteome due to the exclusion of paralogous and duplicated genes, which constitute a relatively large proportion of the Z. mays genome. When duplicated genes are included in the prediction of the interactome, using a many-to-many ortholog matching method that allows the inclusion of paralogs, 1161 Z. mays proteins that were only in the many-to-many set, as well as 14,919 unique interologs were added to the unique interactome. This resulted in a combined 49,026 unique interactions comprising 6004 Z. mays proteins (Table 2). A premade cytoscape formatted graphical visualization of PiZeaM for this combined set of proteins is included in Supplemental File 1).

Confidence Value, Conservation and Connectivity in Predicted Z. mays Interactions
The initial analysis of PiZeaM was performed to determine accuracy and to determine if the network of predicted protein interactions in a major plant crop species such as Z. mays had the same structure as those described for model organisms such as A.
**TABLE 1** | Reference organism data used for maize interolog prediction.

| Organism                        | Total Proteins | 1 to 1 interologs | M to M interologs | M to M fold change | Total interactions |
|---------------------------------|----------------|-------------------|-------------------|-------------------|--------------------|
| Arabidopsis thaliana            | 33602          | 4289              | 7331              | 1.70926E+03       | 11620              |
| Saccharomyces cerevisiae         | 7126           | 42250             | 51873             | 1.22776E+03       | 94123              |
| Drosophila melanogaster         | 15246          | 1508              | 2222              | 1.48010E+03       | 3740               |
| Rattus norvegicus               | 29516          | 22                | 78                | 3.54545E+04       | 100                |
| Mus musculus                    | 37991          | 105               | 269               | 2.56190E+04       | 374                |
| Schizosaccharomyces pombe       | 7031           | 2766              | 3121              | 1.12834E+04       | 5887               |
| Homo sapiens                    | 57945          | 2735              | 5005              | 1.82998E+04       | 7740               |
| Bacillus subtilis               | 4371           | 32                | 32                | 1                | 64                 |
| Caenorhabditis elegans          | 20447          | 1                 | 1                 | 1                 | 2                  |
| Campylobacter jejuni            | 3673           | 60                | 923               | 15.38333E+03      | 983                |
| Synechocystis                   | 3264           | 161               | 501               | 3.11180E+03       | 662                |
| Oryza Sativa                    | 58058          | 220               | 248               | 1.12727E+03       | 468                |
| Escherichia coli K12            | 4618           | 2496              | 2444              | 0.97916E+03       | 4940               |

Total 130703

| Only orthologs used for interologs are listed. |

**TABLE 2** | Predicted maize interactions.

| Orthology                          | Proteins | Interactions |
|------------------------------------|----------|--------------|
| Combined Total (One to One/Many to Many) | 6004     | 130703       |
| One to One Total                   | 4843     | 56645        |
| Many to Many Total                 | 5673     | 74058        |
| Combined Unique (One to One/Many to Many) | 6004     | 49028        |
| One to One Unique                  | 331      | 4737         |
| Many to Many Unique                | 1161     | 14919        |
| Overlap between Unique (One to One/Many to Many) | 4512     | 29370        |

**thaliana** and **S. cerevisiae**. This analysis helped in establishing and differentiating the value of each predicted interaction and protein in the network, and allowed for determining the weights within the sub-networks.

Confidence values (CVs) for each interaction in PiZeaM are listed in [Supplemental Table 1](#) and added to the network visualization in [Supplemental File 1](#) as an edge feature. PiZeaM had 1079, 38,851, and 9096 low, medium and high confidence interactions, respectively. Thus most interactions were at least medium confidence with more than one supporting line of evidence.

These levels of confidence allow users to select specific levels of false discovery when the data is used to build networks or to develop hypotheses. The most confident interactions were self-interactions for AAA ATPases, Topoisomerases and DNA-Repair ([Supplemental Table 1](#)). On the other hand, the most confident hetero-interactions were with proteins involved in core molecular processes, such as elements of the DNA repair machinery, the basal promoter complex, the proteosome, and proteins associated with the regulation of the cell cycle ([Table 3](#)). DNA repair machinery is conserved throughout eukaryotes ([Liu et al., 1999; Ohbayashi et al., 1999](#)) and was recovered in PiZeaM ([Supplemental File 1](#)). High CV correlates with conserved strong interactions that are detected by multiple methodologies in different species.

“Connectivity” of PPIs determines the number of interacting partners for a given protein. Proteins with many interacting partners, referred to as having “high connectivity,” are of biological interest as they may represent the “circuit hubs” central to signaling and information processing in the organism. The distribution of the connectivity of elements in most information systems has implications on features such as robustness and efficiency of the system ([Alon, 2007](#)). In the PiZeaM, the highly connected proteins were ubiquitous partners and co-factors such as chaperones, scaffolding proteins, and protein involved in degradation pathways (see [Table 4](#)).

Highly connected hubs are reported to represent the most evolutionarily conserved proteins and to form the backbone of core processes ([Evlampiev and Isambert, 2008](#)). Proteins in **Z. mays** with a large number of interacting partners were also found to be highly conserved (Table 4). The protein with the highest connectivity was Ubiquitin 2 (GRMZM2G118637_P01), with 797 different predicted protein partners. This protein is orthologous to ubiquitin in **S. cerevisiae**, involved in targeted protein degradation. In **A. thaliana**, ubiquitination has been shown to play a role in abiotic stresses, biotic stresses, and in other cellular processes, including auxin based growth stimulation ([Dong et al., 2006](#)). Some of the other highly interconnected conserved proteins are chaperonins, heat shock proteins, and members of large protein complexes such as the ribosome and the proteosome. Specific conserved interactions were similar to those found in **A. thaliana**, including interactions between histones, proteosome components, MutS type DNA repair proteins, and tubulin binding to the spindle pole body Spc97/Spc98. Thus, in **Z. mays**, despite the large changes in genome structure and increased number of genes compared to the reference species used, the conserved interactions remain intact, and the hypothesis that highly connected hubs have deeply conserved interactions across species, genera and even kingdoms ([Bork et al., 2004; Consortium, 2011](#)) is also evident in domesticated crop species like **Z. mays**.

Phosphorylations of Serine-threonine/tyrosine-protein kinases and their connected transcription factors are...
instrumental in signaling pathways involved in modulating responses to abiotic and biotic stresses, cell growth, and development (Rudrabhatla et al., 2006). These transcription factors and signaling proteins are underrepresented in the Z. mays interactome, although family members are found in all used reference organisms. For example the protein serine/threonine kinase activity (GO:0004674) is found 1135 times in the whole maize genome while only 181 times in the interactome (it was expected 299.3 times based on size of population sample), thus is 1.65-fold depleted. This indicates a lack of conservation of their interactions despite their important function, and likely represents lineage specific rewiring of regulatory networks. Changes in differential gene regulation are thought to be the earliest step in evolutionary divergence, followed only later by changes in gene function (Evlampiev and Isambert, 2008). Some regulatory processes (ubiquitination, methylation, chromatin remodeling) are highly conserved, and more importantly, maintain their conserved interacting partners and complexes throughout eukaryotic evolution (Hershko and Ciechanover, 1998; Kaiser and Huang, 2005; Soltes et al., 2011). These may represent only the core mechanisms for regulation, while the specificity of gene and protein targets in other pathways such as map kinases may diverge or rewire over the course of evolution (Mosca et al., 2012).

Topology of PiZeaM

Topology looks at patterns in networks and compares network properties. The frequency distribution for node connectivity was calculated for PiZeaM (Figure 1B). The majority of proteins were intermediate sized hubs with 10–100 interacting protein partners. The unique interactome had an average connectivity of 15.86 neighbors per node, similar to D. melanogaster and O. sativa (Giot et al., 2003; Stark et al., 2006; Gu et al., 2011). The distribution of nodes by connectivity (Supplemental Table 1) follows an inverse power relationship between node frequency and connectivity, which is typical of “small world” topology networks, and is frequently seen in biological and social networks, and in transportation hubs (Watts and Strogatz, 1998; Alon, 2007).

The shortest path length within a network is the path with the fewest number of intervening nodes between two given nodes. When all protein pairs in PiZeaM were evaluated there was an enriched frequency of path lengths between 2 and 4 with a characteristic path length of 3.942. This suggests that a majority of interactions can be characterized as short path interactions (shown in Figure 1). This is similar to the 2.6 node characteristic path length of S. cerevisiae (Gursoy et al., 2008). The A. thaliana interactome has a characteristic path length of 3.4 (Chen et al., 2012) and the human interactome has a characteristic path length that varies among databases between 1 and 3 (Taylor et al., 2009). The overall topology of the PiZeaM resembles that of experimentally determined networks. Although the resemblance is not surprising (as it was built from experimentally determined reference interactomes), this confirms that Z. mays has the protein orthlogs to generate a predicted interactome of normal topology. This lends confidence that the prediction process has not introduced systematic errors that altered the overall structure of the interactome.

Co-Expression of Interacting Proteins

The distribution of gene co-expression for interacting proteins in PiZeaM was different from what was reported in other predicted interactomes, including those of A. thaliana and O. sativa. When the entire distribution of correlation coefficients for every interacting protein pair was analyzed, it was found to be

### TABLE 3 | The most confident hetero-interactions in the predicted maize interactome*

| Protein A** | Protein B** | Protein A description | Protein B description | CV |
|-------------|-------------|-----------------------|-----------------------|----|
| 008327_P01  | 061287_P01  | Cell division control protein 2 homolog | Cyclin superfamily protein, putative | 12960 |
| 056075_P01  | 110212_P01  | DNA mismatch repair protein MSH2 | MUS2 protein | 10752 |
| AC149818.2_FGP004 | 008327_P01 | Cyclin-dependent kinases regulatory subunit | Cell division control protein 2 homolog | 9450 |
| 045314_P01  | 064732_P01  | Guanine nucleotide-binding protein subunit beta | G protein alpha subunit 1 | 6720 |
| 008327_P01  | 017081_P01  | Cell division control protein 2 homolog | Uncharacterized protein | 6120 |
| 033626_P02  | 061745_P01  | 26S proteasome non-ATPase regulatory subunit 14 | Uncharacterized protein | 5880 |
| 080305_P01  | 174757_P01  | Eukaryotic translation initiation factor 3 subunit A | Uncharacterized protein | 5772 |
| 040152_P02  | 088162_P01  | Unknown | Putative uncharacterized protein | 5780 |
| 047774_P01  | 104609_P02  | Unknown | Unknown | 5616 |
| 028313_P03  | 452026_P01  | Putative translation elongation/initiation factor | Eukaryotic peptide chain release factor subunit 1-1 | 5124 |
| 071518_P01  | 085970_P01  | Spc97/Spc98 family of spindle pole body (SBP) component | Tubulin gamma-2 chain | 5040 |
| 046021_P01  | 177974_P02  | Histone acetyltransferase of the GNAT family 1 | Uncharacterized protein | 4320 |
| 043484_P01  | 148924_P02  | Histone-lysine N-methyltransferase EZ3 | Polycytoplasm group protein RIE2 | 3969 |
| 161418_P02  | 168096_P01  | TATA-box-binding protein 2 | DNA binding | 3640 |
| 021069_P01  | 066101_P02  | Minichromosome maintenance protein | Unknown | 3584 |
| 139352_P01  | 356935_P01  | MUTL-homolog 1 | Uncharacterized protein | 3432 |
| AC234199.1_FGP003 | 048067_P01 | Unknown | Uncharacterized protein | 3360 |
| 152328_P01  | 556845_P01  | Actin-1 | Unknown | 3315 |
| 054225_P01  | 402295_P01  | DNA-directed RNA polymerase | Unknown | 3312 |
| 042371_P01  | 100872_P01  | Elongator protein 2 | Uncharacterized protein | 3024 |

*Unique one-to-one interactome interactions with the highest confidence values. **GRMZM2G prefix for identifiers removed for space.
### Table 4 | The most connected maize proteins.

| Protein ID | Annotation                                           | Connectivity* |
|------------|------------------------------------------------------|---------------|
| GRMZM2G118637_P01 | Putative ubiquitin family protein                     | 797           |
| GRMZM2G012280_P01 | Uncharacterized protein                               | 523           |
| GRMZM2G003626_P02 | 26S proteasome non-ATPase regulatory subunit 14      | 496           |
| GRMZM2G012631_P01 | HSP protein                                          | 432           |
| GRMZM2G327635_P01 | Ubiquitin carboxyl-terminal hydrolase                 | 399           |
| GRMZM2G147671_P01 | 26S proteasome non-ATPase regulatory subunit 4       | 348           |
| GRMZM2G1416120_P01 | Chaperonin CPN80-2I                                  | 334           |
| GRMZM2G155384_P02 | Uncharacterized protein                               | 324           |
| GRMZM2G075719_P02 | Uncharacterized protein                               | 308           |
| GRMZM2G046021_P01 | Histone acetyltransferase of the GNAT family 1       | 281           |
| GRMZM2G152328_P01 | Actin-1                                               | 274           |
| GRMZM2G008410_P01 | Transcribed sequence 1087 protein                    | 250           |
| GRMZM2G038964_P02 | Uncharacterized protein                               | 248           |
| GRMZM2G153815_P01 | Heat shock 70 kDa protein                            | 248           |
| GRMZM2G171080_P01 | Uncharacterized protein                               | 232           |
| GRMZM2G008327_P01 | Cell division control protein 2 homolog              | 230           |
| GRMZM2G027282_P01 | Uncharacterized protein                               | 230           |
| GRMZM2G303752_P01 | ATP-dependent rRNA helicase spb4                     | 215           |
| GRMZM2G14876_P01  | Prefoldin 5                                           | 199           |
| GRMZM2G107540_P01 | Core histone H2A/H2B/H3/H4 domain                    | 195           |

*Number of predicted interacting protein partners for the unique interactome.

uniform and only slightly skewed to the right with the predicted Z. mays interologs, indicating a slightly higher likelihood of co-expression for interacting proteins (Figure 1C). This was not like the tight correlation shown in other plant interactomes (Geisler-Lee et al., 2007; Ho et al., 2012). Randomized interologs (random pairs of proteins from the interactome) displayed a left shifted higher peak in protein interactions distribution than the predicted interologs. A paired T-test was performed to compare correlations and a significant difference ($P$-value of $9.5 \times 10^{-39}$) was found between random and predicted interologs. This showed that, though the difference was slight, there was a significant difference between the randomized network and PiZeaM.

**Go-Enrichment for PiZeaM**

When PiZeam was evaluated for enriched and depleted GO terms there was an overall enrichment of interactions of proteins localized to the chloroplast, mitochondrion, plasma membrane, nucleolus, and cytoplasm (Supplemental Table 3). Proteins localized to the chloroplast (GO: 9507), for example, occurred in 538 interactions in the network ($p$-value of $1.0 \times 10^{-128}$ vs. occurrence by chance). Several biological processes were enriched in the entire interactome, including cell division and vitamin biosynthetic pathways (GO: 9110). There was also enrichment of processes thought to be plant specific, such as

![Figure 1](#)
as 14 proteins that have been predicted to be involved in defense against fungal pathogens (GO: 50382). PiZeaM is thus biased (enriched or depleted) for specific cellular and biological processes.

**Conserved Subnetworks**

The most conserved interactions were identified in an initial analysis of the predicted interactome and its underlying biology. Highly conserved interactions represent ancient pathways that formed in early common ancestors and have remained intact as eukaryotes diverged into their extant forms. There was a large number of unique interactions (6650) with orthologs in more than 4 reference species (Figure 1A). These conserved interactions also had high confidence values in PiZeaM (1644 had CV greater than or equal to 100). This was not surprising as one of the factors used to determine the confidence level of a predicted interaction was the number of species (S) where such an interaction has been previously reported. Subnetworks were created for interactions with interologs in greater than 4 and greater than 5 species (Figures 2A,B). Overall, these included fewer proteins and a significantly smaller ratio of interactions per protein (3.0, 1.2, respectively, in Figures 2A,B) in comparison with the complete interactome (8.2). In the most stringently conserved subnet (5 species, Figure 2B), many well-known complexes such as the COP9 signalosome, the U2 splicosome, the mitotic spindle, histone interactions, and 26S proteosome are represented as interaction clusters (example in Figure 2C). The conserved subnetworks also showed connectivity between some clusters; i.e., the connections between the 26S proteosome and the translation initiation complex (Figure 2B upper left).

**Hypothesis Generation and Data Mining in a Disease Resistance Subnetwork**

PiZeaM is like a detailed road and town map of an entire continent when visualized graphically. It is visually cluttered due to the number of nodes and edges (for example Figure 2A). However, it is useful when a small portion of interest is focused on (i.e., Figure 2C) for hypothesis generation and data mining. Identification of protein-protein interacting subnetworks in a given biological process in *Z. mays* is vital to achieve a better understanding of that process and how it connects to other processes.

---

**FIGURE 2 | Conserved subnetworks.** *(A)* A sub-network consisting of interactions identified by matching interologs in at least 4 (out of 13) reference species. The majority of proteins were found in a single connected network, with several smaller unconnected subnetworks. *(B)* A more stringent subnetwork of interactions with at least 5 interolog reference species. Clear clusters were concentrated in an organic layout by having more inter-group connections than between-group connections. These clusters consisted primarily of proteins with the same biological role, indicated by dotted circles. One such cluster *(C)* was an unconnected subnetwork consisting entirely of members of the COP9 signalosome. EIF3, elongation initiation factor complex 3; U2, U2 splicosome; MCM, mini-chromosome maintenance proteins complex.
Data Mining in Biotic Stress Signaling

A map of the interconnections within the biotic stress signaling and response pathways was developed using a bait-and-prey approach. This approach was taken specifically to examine defense pathways against plant pathogens. All 154 proteins with the GO annotation “Response to other organism” were used as “bait” (see Disease subnet 1 in Supplemental File 1). These proteins include many known pathogen response proteins, such as NON-EXPRESSION OF PATHOGEN RESISTANCE GENE 1 (NPR1), GRMZM2G076450 in Z. mays (Chern et al., 2001; Mou et al., 2003; Spoel et al., 2003), which is the key signaling protein for plant systemic acquired resistance (SAR) (Chern et al., 2001; Ferrari et al., 2003; Mou et al., 2003; Dong et al., 2006; Griebel and Zeier, 2008). The entire interactome was searched for proteins directly interacting (i.e., first neighbors) with the bait. This recovered 1424 “prey” proteins. Subsequent GO term enrichment analysis of the prey indicated enrichment for cytoplasm localized proteins, pyrophosphatases and acid anhydride hydrolases and proteins involved in ketone, small molecule, and primary metabolism (full list in Supplemental Table 4).

The prey dataset was then limited to key regulators and metabolic proteins for reactive oxygen species (ROS), and the hormones jasmonic acid (JA), salicylic acid (SA), ethylene (ET), and abscisic acid (ABA) pathways (Ferrari et al., 2003; Baxter et al., 2014). This list of 81 proteins included 163 interactions, and formed a large connected subnetwork (Figure 3A). There were, however, no clear clusters of specific roles, unlike the highly conserved subnetworks in Figure 2B. The degree of cross connection is somewhat expected given that proteins such as enhanced disease susceptibility 1 protein (EDS1) have been shown to be involved in multiple signaling pathways, including the salicylic acid pathway and the regulation of the jasmonate pathway, to allow specific responses to pathogens (Heidrich et al., 2011). Regulators such as the EDS1-PAD4 protein (phytoalexin deficient 4) are involved in the response to the hemibiotrophic pathogen Blumeria graminis, the causal agent of powdery mildew on grasses (Parker et al., 1996; Yun et al., 2003; Wiermer et al., 2005). The subnetwork demonstrates that the three hormone pathways implicated in interactions with necrotrophic plant pathogens (Jasmonic acid, Abscisic acid, and Salicylic acid) (El-Zahaby et al., 1995; De Gara et al., 2003; Baxter et al., 2014) are physically interacting with proteins involved in ROS signaling. Interestingly, the influence of light flux rate and red/far-red shifts is connected to the network through GRMZM2G013478_P01, a predicted ROS response protein interacting with GRMZM2G092174_P01, a phytochrome protein (Griebel and Zeier, 2008; Moreno et al., 2009). Interactions such as these are presumed to be key to plant-microbe defense

FIGURE 3 | Disease response and hormone subnetworks. (A) Interaction between all proteins in PiZeaM with the gene ontology annotation “response to other organism,” which includes many disease response proteins, and their first neighbors from ROS and stress hormone signaling and metabolic pathways. (B) A nested subnetwork taken from (A) of MAPK signaling proteins and first neighbors. (C) A nested subnetwork taken from (A) of S-adenosyl methionine synthase (ethylene biosynthesis) and first neighbors.
responses due to the interplay between light and susceptibility to pathogens (Mühlenbock et al., 2008; Roden and Ingle, 2009).

A nested subnetwork was also built using S-adenosyl methionine synthase (SAM) and its first neighbors (Figure 3B). SAM is a key step in the biosynthesis of the hormone ethylene, involved in responses to both biotic and abiotic stresses (Yang and Hoffman, 1984). This subnetwork included connections to a superoxide dismutase (which breaks down the ROS superoxide into another ROS hydrogen peroxide), and a ROS responsive GDP mannose pyrophosphorylase, an enzyme that generates an activated mannose important in n-linked glycosylation. SAM synthase is also predicted to interact with an ABC transporter and an RNA polymerase linked to biotic responses, and a MAPK signaling protein. Several additional mitogen activated protein kinases (MAPKs) were also included in the “response to other organism subnetwork” (Figure 3A), when these were isolated along with their interacting partners (Figure 3C). This illustrates connections between different layers of the signaling cascade. A number of other proteins were connected to this cascade, including a protein kinase involved in cell division, a WRKY transcription factor, an S-adenosyl methionine (SAM) synthase, and an uncharacterized protein with a domain of unknown function (DUF26).

Discussion

How Accurate are Interolog Predictions in Plant Species?
The interolog method is useful in identifying interactions that are conserved between different organisms. This has been demonstrated in A. thaliana and other key organisms. It allows for a cost effective high-throughput analysis of potential protein-protein interactions, building on previous experimental evidence acquired from other species (De Bodt et al., 2009). It makes use of the Inparanoid method which has been shown to be able to compete with tree based methods for detection of orthology (Altenhoff and Dessimon, 2009). Moreover, relying solely on experimental approaches to identify PPIs is technically challenging, resource intensive and time consuming, especially since relying on just one experimental method to identify PPIs can lead to false positives. Due to these limitations, the interolog method allows for novel hypotheses to be developed, using existing hard to obtain experimental data, including PPIs data gathered in other organisms (De Bodt et al., 2009).

After the predicted interactome for A. thaliana was released in 2007, a high throughput experimental interactome was created, making the comparison between the predicted interactome and the experimental interactome possible. The two interactomes were compared to confirm that plant protein interactions can indeed be predicted using non-plant reference species. We compared the 72,266 interolog interactions in A. thaliana predicted by Geisler-Lee to the dataset of 37,645 experimentally determined PPIs collected at the BioArrayResource (http://bar.utoronto.ca/welcome.htm) (Toufighi et al., 2005). The observed overlap between the predicted and experimental datasets was 1450 interactions, compared to an expected overlap by chance of about 91 interactions (as two random subsets of (V² + V)/2, where V is the number of proteins common to both sets). The observed overlap was calculated to be enriched 15-fold over chance, with a p-value of less than 10⁻¹⁰⁰ in a simple 1° of freedom chi-square test. There is also considerable bias in experimentally determined interactions toward plant specific interactions, while the A. thaliana interolog predictions are focused on evolutionarily conserved interactions among eukaryotes identified by the ortholog methodology. Neither dataset is likely to be a complete map of all protein interactions in A. thaliana. The interactome of S. cerevisiae has over 230,000 known unique interactions for a very small genome (6600 genes). The number of interactions in a multicellular organism with over 30,000 proteins can be expected to be much higher with well over one million interactions, by simple extrapolation of genome size. Finding 1450 PPIs in the A. thaliana predicted interactome that were also experimentally validated is an indicator that the interolog method is reliable for the discovery of protein interactions in plants, even when using non-plant organisms as reference. All confidence levels of interactions performed much better than chance, with high confidence interactions (CV > 10) performing slightly better than low and medium confidence interactions. There was, however, no significant improvement in overlap with experimentally validated interactions (data not shown).

An important feature of PiZeaM is the inclusion of other “green” species such as A. thaliana, Synechocystis, and O. sativa as experimentally determined reference interactomes. This helps capture the unique aspects of plant pathways such as photosynthesis, cell wall formation, plant development, and disease resistance pathways. Earlier predicted interactomes, such as those of A. thaliana and O. sativa (Gü et al., 2011; Ho et al., 2012), had a significant number of interactions, but no plant references. The GO-term enrichment analysis in PiZeaM shows significant enrichment of chloroplast localized proteins and plant specific pathways, such as phytochrome light sensing, among the interactors. The successful execution of the methodology was in part measured by the examination of recaptured known conserved networks, including the 26s proteosome, ribosome subunits, DNA repair, splicosome, and COP9 signalosome interactions. The major differences in applying this method to Z. mays were the much larger genome and proteome of this species, which had undergone more extensive gene duplication than A. thaliana or O. sativa. Unlike model organisms, a high number of recent paralogs (inparalogs) in Z. mays are due to its polyploid genome, thus making most of these homologs, or genes duplicated due to polyploidy (Adams and Wendel, 2005).

Although co-expression of gene transcripts is not necessarily required for a physical interaction between proteins, it is an observed trend reported in experimental interactomes that proteins that interact tend to be co-expressed. This is likely the effect of natural selection pressure on improving biological efficiency, i.e., not making proteins without all their necessary interacting partners present (Bhardwaj and Lu, 2005). The low correlation of interaction to co-expression in Z. mays seems to depart from these general observations and could possibly be due to the large number of recent paralogs that Z. mays has in its genome. Duplicated genes may have a faster divergence
of RNA expression due to the relaxation of selection pressure, allowing mutation of promoter elements as well as coding sequences. Further analysis of other species with a similar genetic structure, such as wheat, might confirm or refute this hypothesis.

**Usefulness and Novelty of PiZeaM**

The interactome can be visualized as a field of circles (nodes) that represent proteins and connections (edges) between nodes that represent protein-protein interactions. Each node can be rated based on the number of connections, or the connectivity of that node. As stated previously, protein interactomes typically contain a few highly connected hubs, numerous smaller hubs, pipes, and free ends. This distribution is similar to that of other small-world network structures (Watts and Strogatz, 1998; De Silva et al., 2006; Gu et al., 2011). The small-world topology is a compromise between efficiency and robustness. Having fewer interacting partners involved in a pathway results in increased efficiency in terms of how fast a product or outcome can be processed. The highly connected hubs represent proteins that are conserved through different organisms and are under less selection pressure for mutations (Batada et al., 2006; Zotenko et al., 2008; Ning et al., 2010). For example, heat shock proteins (i.e., Hsp70, Hsp90) are vital for proper folding, assembly, and translocation and have been implicated in abiotic and biotic stress studies (Wang et al., 2004). As previously stated, these proteins are some of the highly connected hubs in the PiZeaM. A robust system includes redundant pathways, autoregulation, and feedback for increased stability. Understanding the interplay between efficiency and robustness is an emerging topic of interactomics that sheds light on the organization of various interactions that take place in organisms. This has led to recent advances in the mathematical field of graph theory to analyze and solve these real world problems. Moreover, there has been an increase in studies to determine the relationship between simple sub-networks and complete publicly available networks (Barabasi and Oltvai, 2004). An example would be studies aimed at comparing networks in maps and Internet social sites and drawing parallels from those studies to analyze biological networks. Another feature of small-world graphs is the enrichment of short path lengths reflecting the number of steps in a path between any two nodes (Supplemental Figure 1). This too was noted in predicted and experimentally determined interactomes such as the PRIN O. sativa interactome (Gu et al., 2011).

By mapping Z. mays proteins to orthologs in other species, prediction of function is also improved, which allowed superior annotation for Z. mays proteins in comparison to previously used methods (Geisler-Lee et al., 2007). Not just relying on orthology, the Geisler-Lee method used to build PiZeaM also utilizes interologs (interacting orthologs), co-expression, graph theory and gene ontology as additional layers of annotation of the network (Bhardwaj and Lu, 2005; Brown and Jurisica, 2005; Peterson et al., 2009). This is vital because orthologs across species are not always phylogenetically closer than paralogs (Koski and Golding, 2001). However, the interolog method makes the assumption that if an interaction occurs in the last common ancestor, and both proteins are retained after divergence, the interaction of the proteins is also retained after divergence.

In conclusion, PiZeaM represents a step forward in developing tools to utilize and integrate publically available genomic and proteomic data to improve our understanding of networks underlying plant-microbe interactions, breeding and development in Z. mays. Future work will analyze dynamic relationships in networks to determine causal relationships underlying Z. mays protein interactions.

**Acknowledgments**

We would like to thank Amanda Warner, Tiffany Blanchette and Jane Geisler-Lee for proofreading the article. This work was funded through the USDA cooperative agreement # 58-6435-5-0-461.

**Supplementary Material**

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fgene.2015.00201/abstract

**Supplemental File 1 | Cytoscape visualized interactome.** This file contains the maize predicted interactome as a Cytoscape ver. 3.11 format file. Interactions shown are indicated and filterable by orthology (many-to-many and one-to-one), confidence value (numerical), and connectivity. The bait proteins used in the construction of subnetworks are indicated by pathway and gene name columns. Subnetworks shown in Figures 2, 3 are included, and colors of bait proteins correspond to that portrayed in the figures.

**Supplemental Figure 1 | Shortest path length distribution.** This shows the frequency distribution shortest path length between every pairwise combination of proteins of the maize interactome using the NetworkAnalyzer tool provided by Cytoscape (ver. 3.11) plugins. A majority of the interactions in the maize interactome have a shortest path length between two and six.

**Supplemental Table S1 | The predicted maize interactome.** The Master Interactome sheet contains all interolog predictions for the maize interactome using either many to many or one to one orthology. This table also lists the reference interactome including organism, orthologous proteins, type of experiment, publication identifier, and type of orthology used. Sheets containing excel versions of tables included in the manuscript are also provided for easy use in calculations. Other sheets contain unique interacting pairs for either one-to-one orthology or all unique interaction datasets. Analyses include a calculated confidence value (CV sheet) for each unique interaction, the number of times each interaction occurred in the reference data, and whether the interaction was a self-interaction or heterologous. A Node Property sheet provides annotation and shows the number of interactions each proteins has in either the unique or combined datasets. The Unique Edges sheet is a unique (each interaction represented once) combination of one to one and many to many interactions used in preparing the cytoscape visualization in Supplemental File 1.

**Supplemental Table S2 | Microarray Accessions.** A listing of GEO accession numbers for all the Affymetrix microarrays downloaded to construct the co-expression matrix used in Figure 1C.

**Supplemental Table S3 | GO enrichment analysis for PiZeaM.** The gene ontology terms associated with all 6004 proteins of PiZeaM were compared to the whole maize genome and ranked by multiple-hypothesis adjusted P-value (FDR). The analysis was done using the agriGO tool (http://bioinfo.cau.edu.cn/agriGO/) and Bingo Cytoscape Plugin.

**Supplemental Table S4 | GO enrichment analysis for first neighbors of disease subnet 1.** PiZeaM had 1424 predicted interacting proteins for the 154 proteins with the gene ontology annotation “responds to other organism” (GO:0051707). These were compared to all maize proteins using the agriGO tool (http://bioinfo.cau.edu.cn/agriGO/).
Lalonde, S., Sero, A., Pratelli, R., Pilot, G., Chen, J., Sardi, M. I., et al. (2010). A membrane protein/signaling protein interaction network for Arabidopsis version AMPV2. Front. Physiol. 1:24. doi: 10.3389/fphys.2010.00024

Li, L., Stoheckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–2189. doi: 10.1101/gr.1224503

Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., et al. (2004). A map of the interactome network of the metazoan C. elegans. Science 303, 540–543. doi: 10.1126/science.1091140

Lightfoot, D. A. (2014). “The Soybean Genome Database (SoyGD): a Cytoscape plugin to Parker, J. E., Holub, E. B., Frost, L. N., Falk, A., Gunn, N. D., and Daniels, M. Liu, L., White, M. J., and Macrae, T. H. (1999). Transcription factors and their Mou, Z., Fan, W., and Dong, X. (2003). Inducers of plant systemic acquired resistance to genes in higher plants. Eur. J. Biochem. 262, 247–257. doi: 10.1046/j.1327.1999.00349.x

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plug-in to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21 3448–3449. doi: 10.1093/bioinformatics/bti551

Martinelli, F., Reagan, R. L., Uratsu, S. L., Phu, M. L., Albrecht, U., Zhao, W., et al. (2004). The database of interacting proteins: 2004 update. Nature 428, 449–451. doi: 10.1038/nature02349

Maen, J., Shabalov, P., Koonin, E. V., and Wolf, Y. I. (2005). Evolutionary constraints on structural similarity in orthologs and paralogs. Genome Res. 15, 760–770. doi: 10.1105/tpc.009159

Marcu, M., Yadvinder, S., Chauhan, S., and Bogov, M. (2008). Chloroplast signaling and LESION SIMULATING DISEASE1 regulate crosstalk between salicylate and jasmonate-dependent defense pathways through a novel function in the cytosol. Plant Cell 15, 760–770. doi: 10.1105/tpc.009159

Marti-Renom, M. A., Diekhans, M., and Haussler, D. (2006). Reconstructing gene families with ortholog and paralog information using a Markov model. J. Comput. Biol. 13, 1035–1055. doi: 10.1089/cmb.2006.13.1035

Matteo, A., Mullineaux, P. M., et al. (2008). Chloroplast signaling and jasmonate-dependent defense pathways through a novel function in the TBP-like protein (TLP) distantly related to the Drosophila TBP-related factor. Plant Cell 21, 2546–2552. doi: 10.1105/tpc.10909922

Rual, J.-F., Venkatasekaran, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., et al. (2005). Towards a proteome-scale map of the human protein–protein interaction network. Nature 437, 1173–1178. doi: 10.1038/nature04209

Rudrabhatla, P., Reddy, M., and Rajasekharan, R. (2006). Genome-wide analysis and experimentation of plant serine/threonine-specific protein kinases. Plant Mol. Biol. 60, 293–319. doi: 10.1007/s11105-004-1107-9

Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U., and Eisenberg, D. (2004). The database of interacting proteins: 2004 update. Nucleic Acids Res. 32, D449–D451. doi: 10.1093/nar/gkh086

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Panstrnak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. Science 326, 1122–1125. doi: 10.1126/science.1178534

Schuette, S., Piatkowski, B., Corley, A., Lang, D., and Geisler, M. (2015). Predicted protein–protein interactions in the moss Physcomitrella patens: a new bioinformatic resource. BMC Bioinformatics 16:89. doi: 10.1186/s12859-015-0524-1

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504. doi: 10.1101/gr.1239303

Soltes, G., Muller, E., and D’Aversa, T. G. (2011). Ubiquitin, ubiquitination, and proteasomal degradation in the eukaryotic cell: a review. Bioch 82, 64–71. doi: 10.1893/011.082.0303

Spoel, S. H., Koornneef, A., Claessens, S. M., Korzelius, J. P., Van Pelt, J. A., Mueller, M. J., et al. (2003). NPR1 modulates cross-talk between salicylate- and jasmonate-dependent defense pathways through a novel function in the cytosol. Plant Cell 15, 760–770. doi: 10.1105/tpc.009159

Stark, C., Breitkreutz, B.-J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006). BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 34, D353–D359. doi: 10.1093/nar/gkj109

Stuart, J. M., Segal, E., Koller, D., and Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. Science 302, 249–253. doi: 10.1126/science.1087447

Taylor, I. W., Linding, R., Warde-Farley, D., Liu, Y., Pesquita, C., Faria, D., et al. (2009). Dynamic modularity in protein interaction networks predicts breast cancer outcome network. Nat. Biotechnol. 27, 199–204. doi: 10.1038/nbt.1522

Toufighi, K., Brady, S. M., Austin, R., Ly, E., and Provart, N. J. (2005). The Arabidopsis genome: complexity, diversity, and dynamics. Curr. Opin. Plant Biol. 8, 244–252. doi: 10.1016/j.tplants.2004.03.006

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. Nature 393, 440–442. doi: 10.1038/30918

Wiemer, M., Feys, B. J., and Parker, J. E. (2005). Plant immunity: the EDS1 regulatory node. Curr. Opin. Plant Biol. 8, 383–389. doi: 10.1016/j.pbi.2005.05.010

Wienkoop, S., and Hoffmann, N. E. (2014). Ethylene biosynthesis and its regulation in higher plants. Ann. Rev. Plant Physiol. 35, 155–189. doi: 10.1146/annurev-arplant-031913-110547

Wu, Y., Hu, Z., Lu, S., and Liu, Z. (2009). High-quality binary protein interaction map of the yeast interactome network. Science 322, 104–110. doi: 10.1126/science.1158684

Yin, X., Atkinson, H. A., Gaborit, C., Greenland, A., Read, N. D., Pallas, J. A., et al. (2003). Loss of actin cytoskeletal function and EDS1 activity, in combination, severely compromises non-host resistance in Arabidopsis against wheat powdery mildew. Plant J. 34, 768–777. doi: 10.1046/j.1365-313X.2003.01773.x

Zotenko, E., Mestre, J., O’Leary, D., and Przytycka, T. (2008). Why do hubs in the yeast protein interaction network tend to be essential:
reexamining the connection between the network topology and essentiality. *PLoS Comput. Biol.* 4:e1000140. doi: 10.1371/journal.pcbi.1000140

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