



# A rice protein interaction network reveals high centrality nodes and candidate pathogen effector targets



Bharat Mishra<sup>a,1</sup>, Nilesh Kumar<sup>a,1</sup>, M. Shahid Mukhtar<sup>a,b,c,\*</sup>

<sup>a</sup> Department of Biology, University of Alabama at Birmingham, 1300 University Blvd., Birmingham, AL 35294, USA

<sup>b</sup> Nutrition Obesity Research Center, University of Alabama at Birmingham, 1300 University Blvd., Birmingham, AL 35294, USA

<sup>c</sup> General Surgery Gastrointestinal Section, Department of Surgery, UAB School of Medicine, Birmingham, AL 35294, USA

## ARTICLE INFO

### Article history:

Received 13 January 2021

Received in revised form 10 April 2022

Accepted 17 April 2022

Available online 21 April 2022

### Keywords:

Biotic stress

Infection

Interactome

Oryzae

Plant-pathogen interactions

Prioritization

## ABSTRACT

Network science identifies key players in diverse biological systems including host-pathogen interactions. We demonstrated a scale-free network property for a comprehensive rice protein-protein interactome (RicePPInets) that exhibits nodes with increased centrality indices. While weighted  $k$ -shell decomposition was shown efficacious to predict pathogen effector targets in Arabidopsis, we improved its computational code for a broader implementation on large-scale networks including RicePPInets. We determined that nodes residing within the internal layers of RicePPInets are poised to be the most influential, central, and effective information spreaders. To identify central players and modules through network topology analyses, we integrated RicePPInets and co-expression networks representing susceptible and resistant responses to strains of the bacterial pathogens *Xanthomonas oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola* (*Xoc*) and generated a Rice-*Xanthomonas* Interactome (RIXIN). This revealed that previously identified candidate targets of pathogen transcription activator-like (TAL) effectors are enriched in nodes with enhanced connectivity, bottlenecks, and information spreaders that are located in the inner layers of the network, and these nodes are involved in several important biological processes. Overall, our integrative multi-omics network-based platform provides a potentially useful approach to prioritizing candidate pathogen effector targets for functional validation, suggesting that this computational framework can be broadly translatable to other complex pathosystems.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Complex systems including physical, biological, and social systems can be represented and analyzed as networks to reduce

**Abbreviations:** AI-1<sub>MAIN</sub>, Arabidopsis interactome-1 MAIN; BB, bacterial blight; BC, betweenness centrality; BLS, bacterial leaf streak; EAI, effector-activated immunity; EAS, effector-activated susceptibility; EC, eigenvector centrality; ETI, effector-triggered immunity; ETS, effector-triggered susceptibility; FDR, False discovery rate; GEO, gene expression omnibus; GRN, gene-regulatory network; IC, information centrality; PPI, protein-protein interaction; PTI, pattern-triggered immunity; RDP, degree-preserving random network; RicePPInets, Rice protein-protein interaction network; RIXIN, rice-*Xanthomonas* interactome; RXICoNet, rice-*Xanthomonas* interaction co-expression network; T3S, type III secretion system; TAL, transcription activator-like; WGCNA, weighted co-expression network analysis; *Xoc*, *Xanthomonas oryzae* pv. *oryzicola*; *Xoo*, *Xanthomonas oryzae* pv. *oryzae*.

\* Corresponding author at: Department of Biology, University of Alabama at Birmingham, 1300 University Blvd., Birmingham, AL 35294, USA.

E-mail address: [smukhtar@uab.edu](mailto:smukhtar@uab.edu) (M. Shahid Mukhtar).

<sup>1</sup> These authors contributed equally.

<https://doi.org/10.1016/j.csbj.2022.04.027>

2001-0370/© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

complexity [1,2]. Similar to other networks, biological networks are represented as correlation-based- gene co-expression networks (GCN) and, protein-protein interaction networks (PPI) or interactomes [3,4]. These associations maintain the global circuitry of diverse biological processes and signaling pathways induced during any stress [5]. Exploration of the structural and functional properties of these interactomes holds high potential to reveal a wide range of information on specific proteins and interactions, the formation of diverse gene modules, and distinctive nodes within signaling cascades [6–9]. Some of the well-studied network structural topological properties (centralities) employed in diverse biological networks are degree, betweenness centrality, eigenvector centrality, and information centrality [7,10,11]. For instance, the degree is the total number of connections of a node, betweenness centrality is the bottleneck property among two subnetworks, eigenvector centrality is the most influential nodes based on the mutual effect of node neighbors and their influence in a network, and information centrality is the ability of a node to spread

information throughout the network based on the shortest path [7,11–16]. These centralities define the network's robustness and as well as resilience and identify the significant components in the network [17]. Therefore, decoding the network structural architecture and exploiting these topological properties are crucial to revealing the novel nodes/modules in a multifaceted system, which may lead to the identification of key players in biological processes that can be validated by experimental procedures [18–20].

Pathogens including bacteria, viruses, fungi, oomycetes, insects, and others employ a compendium of pathogenic proteins termed “effectors” that interact with host proteins termed as “targets” to manipulate the host machinery to establish disease [21–24]. Since most of such pathogen target proteins exhibit high centrality nodes within a given network for efficient pathogenesis or host survival [11,22,25], investigating network structural analyses will expand our understanding of the nature of plant-pathogen interactions. Towards this, several intra- and inter-species host-pathogen interactomes have been generated and investigated in different model systems including worms, plants, mice, and humans [4,11,24,26]. By exploiting the network-based approaches, such systems-level host-pathogen interactomes resulted in the identification of several novel molecular players, structural modules, functional components, and overall cellular regulatory perturbations [23,27,28]. For instance, several reports revealed that diverse interactomes exhibit similar network architecture and display scale-free network topology [24]. To discover pathogens' contact points, neighborhood-based degree, and path-based betweenness centrality analyses have been applied [3,4,11]. Indeed, nodes with high connections (hubs) and high betweenness (bottlenecks) were previously shown to be the preferred contact points in diverse host-pathogenesis including diverse plant-pathogen interactomes, human-viral, and human-bacterial interactomes, as well as these topological centralities associated with different cancer immunopathogenesis and other diseases [21–25,29–38]. Additionally, pathogen target proteins were found closer (based on the shortest path) to the pathogen-induced differentially expressed genes [14]. Specific to the bacterial plant pathosystem, the network analysis of several interactomes has been explored and revealed that the aforementioned network centrality features are significantly associated with the most venerable pathogen targets [3,12,14,21,25,39–42].

Bacterial blight (BB) and bacterial leaf streak (BLS) are major diseases in many rice (*Oryza sativa*) growing areas worldwide, leading to a reduced yield and quality of this staple crop and degrading plant quality [43]. BB is caused by *Xanthomonas oryzae* pv. *oryzae* (*Xoo*), and BLS is caused by *X. oryzae* pv. *oryzicola* (*Xoc*). These pathogens attack plants in part by injecting proteins (effectors) secreted through the type III secretion system (T3S) [44,45], which act in the plant cell to create conditions conducive to pathogen multiplication and spread, which has been referred to as effector-triggered susceptibility (ETS) [45]. The plant can detect pathogen-associated molecular patterns (PAMPs) and effector molecules to activate patterns- and effector-triggered immunity (PTI and ETI), respectively [45–47]. Largely, *Xanthomonas* pathogens possess two types of pathogen effector proteins, transcription activator-like (TAL) effectors and non-TAL effectors (*Xanthomonas* outer proteins, Xops) [48,49]. The TAL effectors act as transcription factors and directly transcriptionally activate host target genes. Some of these targets' function as disease susceptibility (S) genes and others as resistance (R) genes, respectively mediating what we refer to as effector-activated-susceptibility (EAS) and immunity (EAI). EAS and EAI may be considered subclasses of ETS and ETI, with the important distinction that the outcome is mediated not by a protein-protein interaction between the pathogen and host, but by direct transcriptional activation of a host

gene by a TAL effector. A TAL effector finds its target via contiguous interactions between polymorphic repeat sequences in the protein and specific bases in the DNA [50,51]. This modular DNA-recognition mechanism enables the experimental and computational identification of candidate TAL effector targets based on 1) TAL effector-dependent upregulation and 2) the presence of a matching DNA sequence in the promoter of the candidate gene [52,53]. Additional experimentation is needed to validate candidates and to determine whether their activation plays a role in disease. Because there is degeneracy in the base-specificity of some repeat types, and because TAL effector-dependent gene expression changes may include indirect results of TAL effector activity, candidates may be numerous. A rational means of prioritizing candidates for such validation is thus desirable.

Cross-species transcriptome and interactome studies have identified the most significant players in host-pathogen interactions through integrative network systems biology techniques [54]. Transcriptome sequence data generated from different rice genotypes inoculated with different virulent or avirulent strains of *Xoo* or *Xoc* have been published, which we integrated with the collection of rice computational interactomes to understand the rice-*Xanthomonas* interplay [53]. Here, we performed network-based analyses including standard centralities and a newly improved version of weighted *k*-shell decomposition analysis on a comprehensive rice protein-protein interaction network (RicePPInets) that encompasses 17,421 nodes and 759,851 interactions. Additionally, we generated multiple gene co-expression networks representing five interaction outcome categories for rice-*Xanthomonas* interactions (PTI, ETI, ETS, EAS, and EAI). Subsequently, we integrated the co-expression networks (34,320 genes and 71,110,592 interactions) with RicePPInets and reported a Rice-*Xanthomonas* Interactome (RIXIN) with 9,603 proteins and 110,281 interactions. Focusing on BLS and five pathogenic TAL effectors conserved across multiple *Xoc* strains (called Tal2g, Tal3a, Tal3b, Tal9a, and Tal11b in strain BLS256 [Wilkins, 2015]), we discovered that high centrality nodes as well as proteins located in the inner layers of RIXIN are enriched in previously identified candidate TAL effector targets. Moreover, we also discovered that RIXIN exhibits significant properties of biological networks and can serve as a computational resource representing the most influential proteins during *Xoc*-rice interaction. Thus, we propose network analysis as one means of prioritizing candidate pathogen effector candidates for validation experiments.

## 2. Materials and methods

### 2.1. Dataset acquisition

The large-scale high-confidence computational interactome (RicePPInets) was compiled by merging the interactions of four rice protein-protein interaction networks (RicePPInet [55], PRIN [56], STRING [57], and ORFome [58]) for network biology analysis. The transcriptomics datasets were downloaded from the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) database. We included seven transcriptome datasets (GEO series ID: GSE16793, GSE19239, GSE33411, GSE34192, GSE36272, GSE63047, and GSE67588) in this study representing different classes (outcomes) of host-pathogen interaction (EAI, ETS, PTI, EAS, and ETI). A total of 478 pathogen effector targets were included in this study comprising 378 Candidate, 19 validated *Xoc* TAL effector targets, which interact with well-studied five TALs (Tal2g, Tal3a, Tal3b, Tal9a, and Tal11b) extracted from Wilkins et al. 2015 [53] and Cernadas et al. 2014 [52], and 81 rice orthologs of Arabidopsis host-pathogen targets obtained from Ahmed et al. 2019 [12].

## 2.2. Network analyses

The network centrality measures for the rice interactome (RicePPInets) were analyzed using the NetworkX [59] library in Python 2.7.13. Comprehensively, we computed neighborhood-based, path-based, and iterative refinement centralities – degree, betweenness centrality, eigenvector centrality, and information centrality of the network with  $n$  number of nodes and edges.

Further, the weighted  $k$ -shell decomposition network measure was performed as described in Wei et al [60] and Ahmed et al [12]. This method generates the shells by weighing the node's degree and the connected edges. The calculation is done by the equation.

$$k_i^w = \alpha_i^k + (1 - \alpha) \sum_{j \in \Gamma_i} w_{ij}$$

where  $\Gamma_i$  are sets of neighboring nodes of  $i$  and  $w_{ij}$  is the edge weight defined as  $w_{ij} = K_i + K_j$ . The value of  $\alpha$  was set to 0.5.

Additionally, all network analysis was performed on two random networks generated from RicePPInets for a comparative study. One of the random networks preserved the degree distribution while another random network did not preserve the pre-existing degree distribution of RicePPInets. All networks were visualized using Cytoscape 3.7.1 [61].

**Weighted  $k$ -shell decomposition implementation.** We implemented the weighted  $k$ -shell algorithm in Python (3.7). The scripts can be accessed on (GitHub).

## 2.3. Co-expression network construction

We constructed a *Xoc*-rice interaction co-expression network from seven transcriptome datasets. First, seven GEO datasets (GSE16793 [52], GSE19239 [62], GSE33411, GSE34192, GSE36272, GSE63047 [63], and GSE67588 [53]) were downloaded from the NCBI GEO database. Second, the datasets were partitioned into five rice categories, based on the nature of the interaction based on outcome (EAI, ETS, PTI, EAS, and ETI; see Supplemental Information). To determine respective similar gene expressions, we implemented a correlation-based R package, weighted gene co-expression network analysis (WGCNA) [64]. A soft-threshold power of 18–12 along with a scale-free model fit was utilized for maximum scale-free topology, preserving high mean connectivity, and rejecting lesser correlations based on the number of samples per dataset. We created individual elementary WGCNA networks using flashClust() and cutreeDynamic() algorithms, incorporating all cleaned expression values [65]. The resulting networks were merged into five different co-expression networks based on the five interaction categories (EAI co-expressed pairs, EAS co-expressed pairs, ETI co-expressed pairs, ETS co-expressed pairs, and PTI co-expressed pairs). Finally, a massive rice-*Xanthomonas* interaction co-expression network (RXICoNet) was constructed. Further, RXICoNet was integrated with RicePPInets to get the Rice-*Xanthomonas* Interaction Network (RIXIN) using in-house python scripts.

## 2.4. Functional analyses

Gene Ontology analysis and enrichment of inner layers' proteins and candidate TAL effector targets in RicePPInets and RIXIN were achieved through the panther gene ontology tool [66].

## 2.5. Statistical analyses

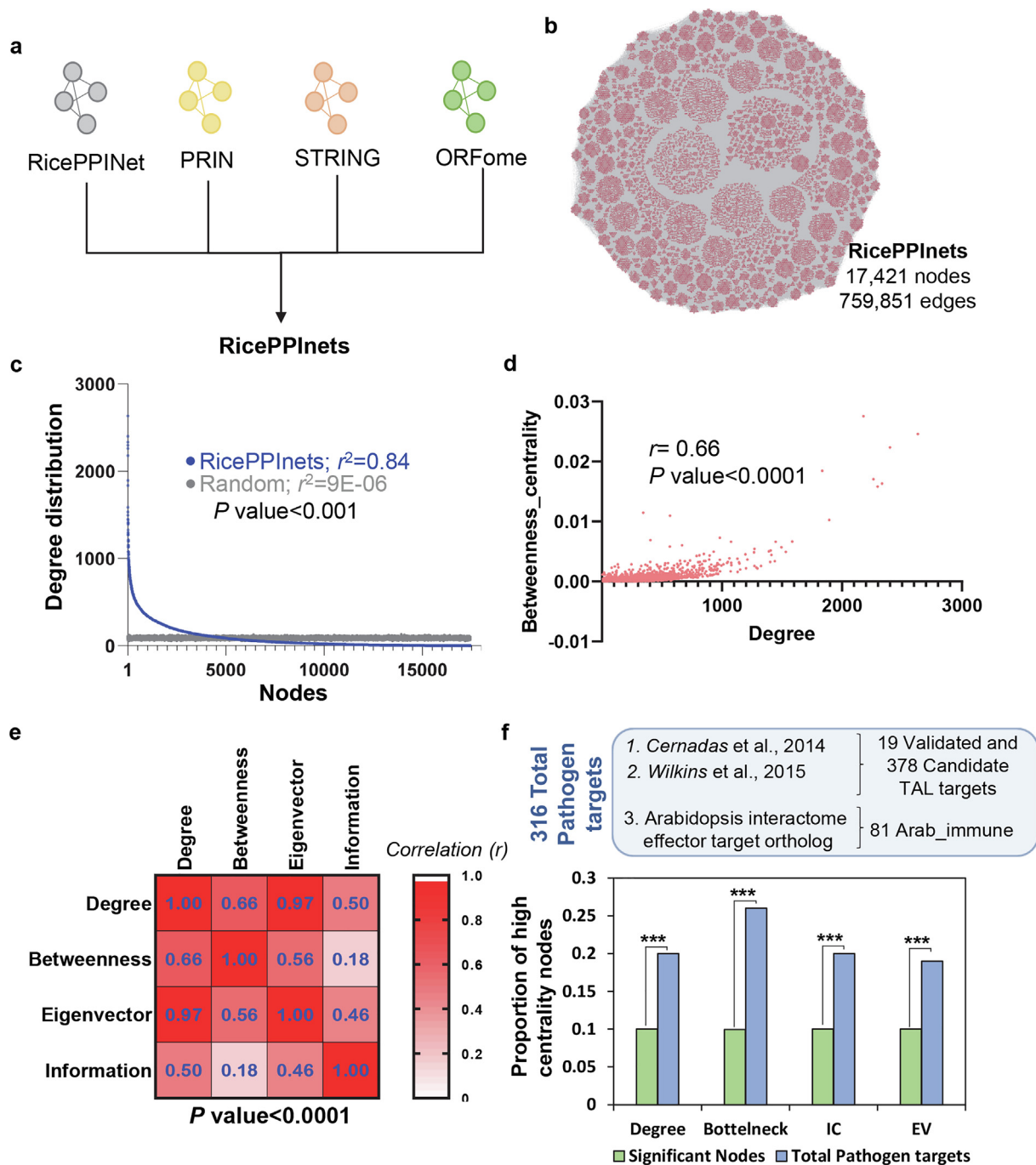
We carried out power-law correlation ( $r^2$ ) analysis, Pearson correlation analysis, the Mann–Whitney–Wilcoxon test, and a test for hypergeometric enrichment using R version 3.3.1 as well as the

online Stat Trek tool. The top 10 % nodes with high centrality values for the following centrality measures: degree, betweenness, information centrality, and eigenvector were classified as the significant property of RicePPInets and RIXIN proteins. Similarly, four shells cutoffs (10 %, 15 %, 20 %, and 33 %) for the weighted  $k$ -shell decomposition method were considered to classify internal layers' proteins, and others were peripheral layers' proteins.

## 3. Results

### 3.1. Network centralities identify the most influential proteins in RicePPInets

To investigate the protein relationships in rice interactomes and subsequently identify potential important nodes in rice, we developed a computational pipeline that encompasses a series of network analyses (Fig. S1). Towards this, we compiled a comprehensive and non-redundant computational rice interactome; RicePPInets from four diverse sources encompassing 17,421 nodes (proteins) and 759,851 edges (interactions) (see Methods, Fig. 1A–B, Table S1). To examine the structural properties of RicePPInets, we generated two random networks *i.e.*, a degree preserving random network and a non-degree preserving random network. To evaluate network robustness, we calculated the scale-free property of RicePPInets and compared it with non-degree preserving random network. As expected, the network degree centrality revealed that RicePPInets follows a scale-free property based on the power-law distribution ( $r^2 = 0.84$ ); few proteins have a higher number of connections than other proteins. In contrast, the non-degree preserving random network doesn't follow the scale-free property, with >75 % of proteins having the same degree distribution (Fig. 1C, Table S1). Next, we performed the standard network centrality analysis (degree, betweenness, information, and eigenvector) on RicePPInets to explore the topological features of nodes. It is well documented that the highly connected nodes are also central to the network, thus we performed the correlation analysis on all centralities. The quantitative relationship between proteins corresponding to degree, betweenness, information, and eigenvector centrality distribution revealed that highly connected nodes in RicePPInets are also bottlenecks (Pearson correlation  $r = 0.66$ ,  $P < 0.0001$ ) and significant information spreaders (Pearson correlation  $r = 0.5$ ,  $P < 0.0001$ ) (Fig. 1D and E, Table S1). At the same time, eigenvector centrality correlation with a degree was also high (Pearson correlation  $r = 0.97$ ) (Fig. 1E) RicePPInets. Further, we classified the significant nodes based on the top 10 % of centrality nodes for each topological centrality. In RicePPInets, the top 10 % of proteins with the maximum number of connections are found to be “hub<sup>248</sup>(hub)” with 248 or more connections. Subsequently, betweenness centrality distribution revealed that a subset of proteins is significantly central in RicePPInets, termed “bottlenecks.” with selected “0.000268” as the cutoff for significant bottlenecks representing the top 10 % of proteins in RicePPInets (Fig. 1F, Table S1). Additionally, information centrality distribution on the largest subnetwork of RicePPInets revealed that most of the proteins in the subnetwork have high information spreader (top 10 %, IC cutoff 0.000329) properties (Fig. 1F), and eigenvector centrality distribution also identified the top 10 % of nodes with high eigenvector cutoff 0.011198984 in RicePPInets (Fig. 1F). The functional analysis for significant proteins revealed that some of the significantly enriched ontologies are a response to stimulus, response to stress, signal transduction, organic compound transport, defense response, translation elongation, cellular response to nitrogen starvation, and immune system processes ( $P < 0.01$ , Table S1). To explore the topological locations of 478 pathogenic effector targets (see methods; Fig. 1F), we mapped and identified



**Fig. 1.** Structural centrality measures identifies RicePPInets most influential proteins in the network. (A) RicePPInets encompassing four rice protein–protein interaction networks (RicePPInet, PRIN, STRING, and ORFome). (B) RicePPInets enlists 17,421 nodes (proteins) and 759,851 edges (interactions). (C) The graph illustrates the degree distribution of RicePPInets and a random network. RicePPInets follows scale free network properties ( $r^2 = 0.84$ ) where few proteins have higher degree than others. The random network does not, with >75 % of proteins having the same degree distribution. (D) The quantitative relationship between node degree and betweenness distribution to identify the highly central and lethal nodes (Pearson correlation coefficient of  $r = 0.66$ ,  $P < 0.0001$ ). (E) The quantitative relationship among four standard centralities distribution illustrates most of highly connected nodes have high information centrality, eigenvector and betweenness properties (Pearson correlation coefficient of  $r = 0.5–0.97$ ,  $P < 0.0001$ ). (F) We collected a compendium of previously reported pathogen effector targets in rice and effector targets orthologs of Arabidopsis (Arab Immune) as potential total pathogen effector targets (on top). Network analysis identified that total pathogen effector targets are significantly enriched in the proteins with high properties including hub, bottleneck, information centrality (IC) and eigenvector centrality (EV) (hypergeometric test  $p < 0.001$ ) (bottom).

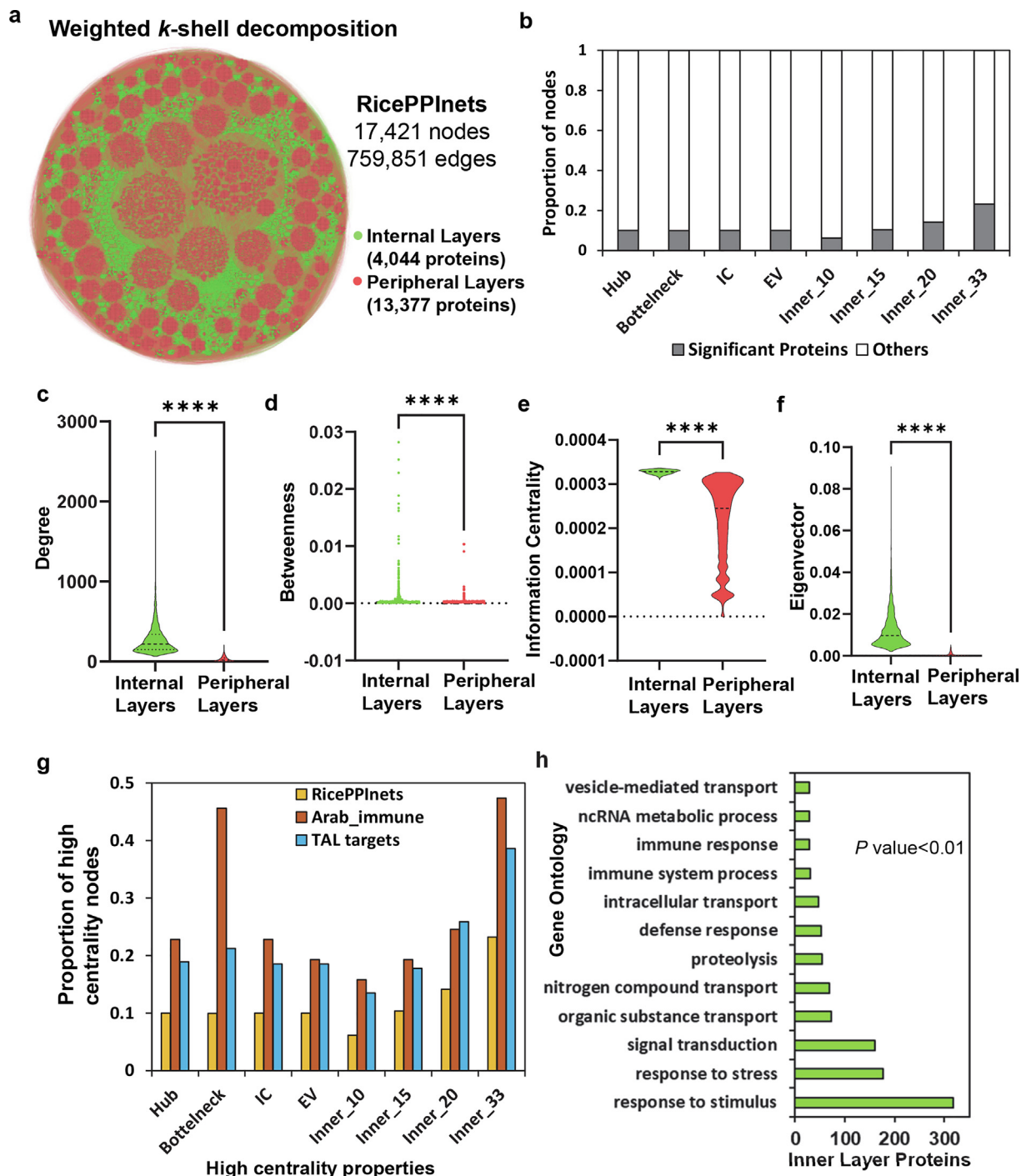
316 total pathogen effector targets in RicePPInets (Table S1). This includes 8 experimentally validated TAL targets, 259 candidate TAL targets [52,53] and 57 orthologs of effector targets in Arabidopsis interactome (Arab Immune; Table S1) [12]. Interestingly, we identified that these 316 pathogen effector targets are significantly enriched as hubs, bottleneck, information, and eigenvector centrality (hypergeometric test,  $P < 0.001$ ). This analysis also iden-

tified that standard centrality measures identified at least 20 % of total pathogen effector targets with high parameters. Taken together, these analyses supported earlier reports and strengthens the established notion that a small section of nodes in the interactome possess significantly increased standard network centrality indices, thus controlling a large proportion of network functionality [12,14,21,40].

### 3.2. Weighted *k*-shell decomposition uncovers the core proteins

Recently, *k*-shell decomposition and its derivative “weighted *k*-shell decomposition” analyses presented an extraordinary platform for the classification of the significant proteins in interactomes, outperforming standard network centrality indices

comprising degree, betweenness, and PageRank [12,67]. However, the current version of weighted *k*-shell decomposition is limited to small-scale networks. Towards making it applicable to larger-scale networks, we first improved its computational code (see Methods). This implementation reduced the execution time from weeks to hours. Moreover, the current implementation is easy to execute



**Fig. 2.** Weighted *k*-shell decomposition identifies internal layer proteins have increased pathogen effector targets than standard centralities in RicePPInets. **(A)** Weighted *k*-shell decomposed with  $\alpha = 0.5$  computes two layers, internal and peripheral in green and red color, respectively. The internal layers have 4,044 nodes while peripheral layers have the remaining 13,377 nodes in RicePPInets. **(B)** The proportion of RicePPInets nodes with significant centrality cutoff. Grey are significant proteins and white are other protein of RicePPInets. **(C-F)** The centrality distribution of both internal and peripheral layers of RicePPInets illustrates that internal layers' proteins have significantly high degree, betweenness, information centrality and eigenvector centrality. (Mann-Whitney test  $P < 0.0001$ ). **(G)** The proportion of proteins in high centralities corresponding to known effector targets in Arabidopsis (Arab-immune) and pathogen effector target proteins (Pathogen targets). **(H)** The gene ontology analysis of inner layers' protein from Fig. 2A identifies stimulus, stress, signal transduction, transport, defense and immune system proteins are significantly enriched ( $P < 0.01$ ). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

on a wide variety of operating systems, is user-friendly, and does not depend upon bash and Java run time in the environment. To identify the most influential node in RicePPInets, we employed this improved version of weighted  $k$ -shell decomposition to RicePPInets and both random networks with  $\alpha = 0.5$  as a default parameter [12]. The analysis parsed the network into different shells within the network to compute the core of the network. As a result, we discovered 2,835 shells (1/3rd of total shells) from the core of the network as the internal layers. The remaining 5,756 shells of RicePPInets are designated as a peripheral (outer) layers (Fig. 2A, Table S1). The internal and peripheral layers consist of 4,044 and 13,377 nodes, respectively. Furthermore, to evaluate the shells with other standard network centralities, we selected three more cutoffs for internal layers proteins for 10%, 15%, and 20% of shells, which resulted in 1,097, 1,812, and 2,467 proteins, respectively. The comparison against standard centralities and internal layers for the identification of significant proteins revealed that the top 10% nodes with centralities have  $\sim 1,742$  proteins, whereas shells with 10% inner layers represent only 6% of all nodes in RicePPInets, whereas 33% of inner layers represents 23% of all nodes in RicePPInets (Fig. 2B). Hereafter, we mapped standard network centralities against internal and peripheral layers' proteins to establish the structural influence of proteins. Strikingly, we found that the degree, betweenness, information centrality, and eigenvector centrality of internal layers' proteins are significantly higher than those of the peripheral layers' proteins (Fig. 2C–E, Mann-Whitney test,  $P < 0.0001$ ). Further, to explore the distribution of total pathogen targets in inner layers of RicePPInets, we mapped the above-described 316 pathogen targets to a weighted  $k$ -shell decomposed network. Interestingly, the analysis identified that the inner layers of 33% of shells encompasses 40% of TAL effector targets and  $\sim 45\%$  of Arab\_immune (Fig. 2G). This is a major improvement in the pathogen effector target discovery through interactome analysis. Furthermore, the high enrichment of Arab\_immune as bottleneck nodes also aligns with the finding that generally conserved proteins possess high betweenness centrality. Additionally, to test the biological significance of inner layers' proteins, we performed gene ontology analysis by panther database. The analysis revealed that some of the significantly enriched ontologies for internal layers proteins are a response to stimulus, response to stress, signal transduction, organic compound transport, defense response, and immune system processes ( $P < 0.01$ , Table S1). Taken together these observations imply that the proteins residing within the internal layers of RicePPInets are poised to be the most influential, central, and poised to be a preferential target of pathogen effectors.

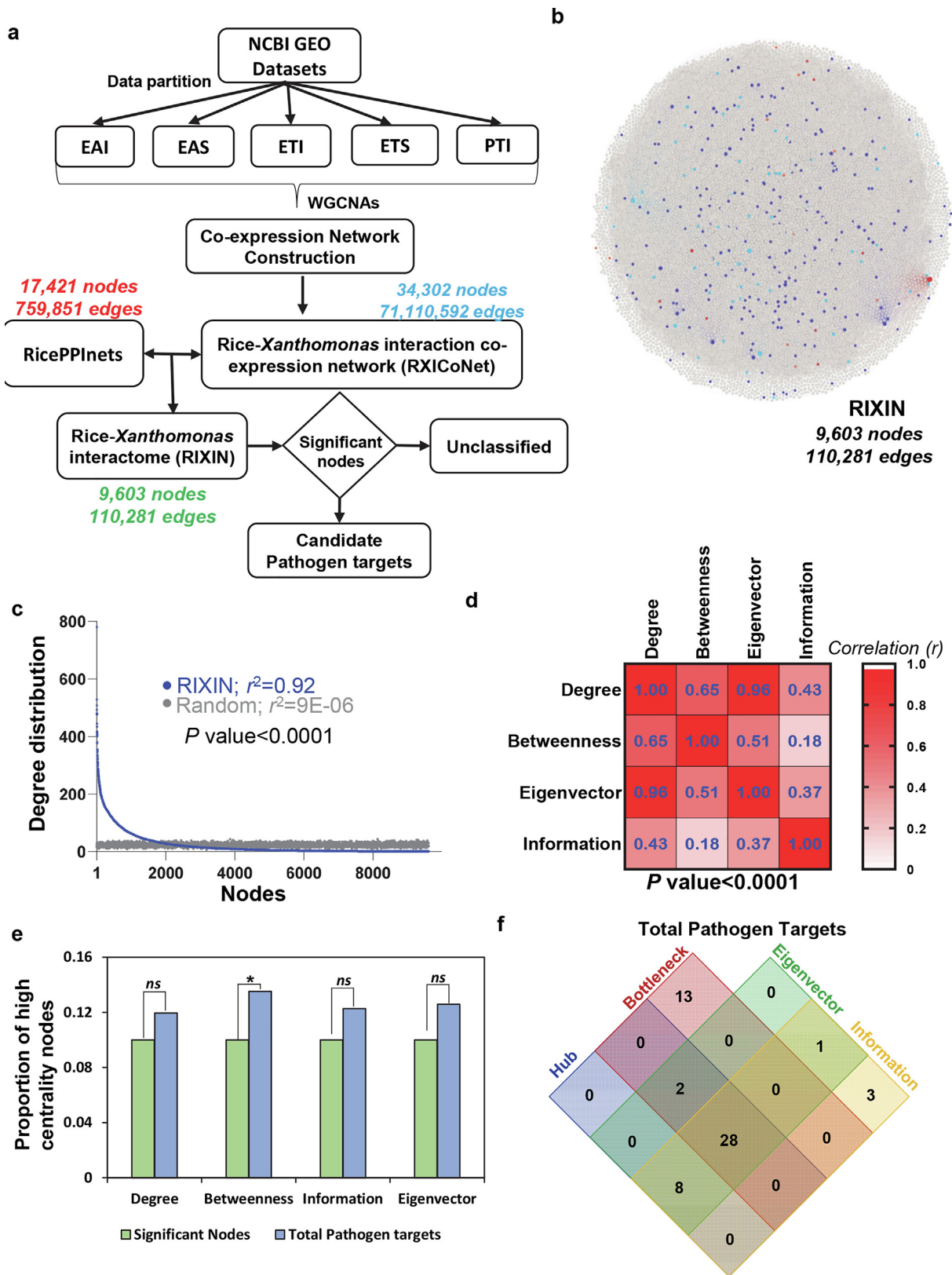
### 3.3. The Rice-Xanthomonas interactome (RIXIN) as a potential source to discover candidate TAL effector targets

Having established the robust, scale-free nature of RicePPInets, we sought to investigate its utility for examining putative molecular contact points in a specific host-pathogen interaction. For this, we took advantage of extensive existing gene expression data for different rice genotypes interacting with different strains of *Xoo* and *Xoc* representing different plant-pathogen interaction outcome categories (EAI, ETS, PTI, EAS, and ETI) to filter the interactions (Fig. 3A; NCBI GEO database accessions GSE16793, GSE19239, GSE33411, GSE34192, GSE36272, GSE63047, and GSE67588). First, we classified each dataset into one of the five categories based on the interaction represented (Fig. 3A, Supplemental Information). Subsequently, we performed a weighted co-expression network analysis (WGCNA) for each of the transcriptome datasets. This analysis yielded five individual co-expression networks partitioned into five interaction categories (EAI: 184,371 pairs, EAS: 86,064 pairs, ETI: 44,341,444 pairs, ETS: 20,127,875 pairs, and PTI:

2,196,789 pairs, Supplemental Information). To understand the comprehensive *Xanthomonas*-infected rice gene co-expression network, we merged five individual co-expression networks. Consequently, we recovered a massive rice-*Xanthomonas* interaction co-expression network (RXICoNet) with 34,302 nodes and 71,110,592 non-redundant edges. Given the massive size of RXICoNet, we integrated the RicePPInets (17,421 nodes and 759,851 edges) with RXICoNet to get the *Xanthomonas*-infected rice specific PPI network, which we termed the rice-*Xanthomonas* interactome (RIXIN). The RIXIN network encompasses 9,603 proteins with 110,281 interactions (Fig. 3B, Table S2). To verify the robustness of RIXIN, we performed standard network centralities analyses on it (Figs. S4A and B, Fig. S2A). Interestingly, RIXIN follows a scale-free property-based power-law distribution ( $r^2 = 0.92$ ) as most real-world networks [20]. Also, few nodes of RIXIN have high betweenness centrality, similar to most previously studied interactomes [9,12,14,40,68]. Furthermore, the quantitative relationship between proteins corresponding to a degree, betweenness, information, and eigenvector centrality distribution revealed that highly connected nodes in RIXIN have high eigenvector values (Pearson correlation  $r^2 = 0.96$ ,  $P < 0.0001$ ) and significant bottleneck and information spreaders (Pearson correlation  $r^2 = 0.43$ – $0.65$ ,  $P < 0.0001$ ) (Fig. 3D). Afterward, we computed the significant property proteins with a high degree, betweenness, information, and eigenvector centrality with a standard cutoff of the top 10% nodes for each centrality. As a result, we have 960 proteins with high values for each centrality. Consequently, to test the hypothesis that pathogen effector targets are central, we mapped the total pathogen effector proteins in RIXIN. Intriguingly, we observed that most of the pathogen effector targets are significantly enriched only as bottlenecks but not as a hub, information, or eigenvector centrality in RIXIN (Fig. 3E, Table S2). To verify the enrichment of pathogen effector proteins in the hub and bottleneck of the interactome, we constructed two additional random networks, i.e., a random RIXIN (rRIXIN; Table S2, Fig. S2A) and degree-preserving random RIXIN (rdp\_RIXIN; Table S2, Fig. S2C) network. The network centrality analysis identified 34 and 25 total pathogen targets are part of hub and bottlenecks, which are neither enriched in rRIXIN nor in rdp\_RIXIN networks (Table S2, Fig. S2B, and D; hypergeometric test  $p$ -value  $> 0.05$ ). Furthermore, we also tested the enrichment of pathogen targets in inner layers' proteins of the rdp\_RIXIN network. Again, we reported that 87 pathogen targets are present in the inner layers' proteins, which are not significantly enriched in inner layers' proteins of rdp\_RIXIN (Table S2, Fig. S2D; hypergeometric test  $p$ -value  $> 0.05$ ). Next, we compared the total pathogen effector targets with high centrality parameters among four network topologies. The analysis found that most of the pathogen targets (28) share all four high parameters including degree, betweenness, information, and eigenvector centrality. Whereas 13 pathogen effector targets are uniquely enriched as bottlenecks relating to the significance of betweenness to the validated TAL targets and Arab\_immune. These analyses strongly suggest that RIXIN reflects significant properties of biological networks and can serve as a computational resource for studying the most influential and central proteins with additional topological measures in the *Xanthomonas*-infected rice interactome.

### 3.4. Candidate TAL effector target predictability improved by weighted $k$ -shell decomposition indices in the rice-pathogen interaction interactome

The evolution of specialized pathogen effectors to target the most influential, predominantly core nodes of host intracellular networks to the pathogen's advantage has been widely reported [4,12,14,69]. To investigate whether the candidate and validated TAL effector targets might reside in the inner layers of RIXIN, we



employed the weighted  $k$ -shell decomposition analysis on RIXIN and the random versions of RIXIN. We restricted the analysis with three inner shell cutoffs including 10 %, 20 %, and 33 %, which identified 959, 1,920, and 2,653 nodes in the inner shells, respectively (Fig. 4A, Table S2). Here, we report that inner shells 33 % is the best cutoff to determine the internal layers' proteins and peripheral layers' proteins. Next, to explore the topological residence of pathogen effector targets, we mapped the compendium of 478 total pathogen effector targets (see methods) (Fig. 4B, Table S1). Furthermore, we explored the enrichment of 318 out of 478 pathogen effector target proteins (including 57 Arabidopsis immune orthologues, 252 candidate, and 9 validated TAL targets) among significant proteins corresponding to hubs, bottlenecks, influential information spreaders, eigenvector, and different inner shells in RIXIN (Table S1). Interestingly, we found that hub, information centrality, and eigenvector centrality only discover a total of 38, 39, and 40 pathogen effector targets, respectively including two validated TAL effectors targets (OsTFX1, SWEET14) in degree and information centrality and one in eigenvector centrality (OsTFX1) with high values (Fig. 4C, hypergeometric  $P > 0.05$ , n.s). Similarly, the bottleneck discovers a total of 43 pathogen effector targets including five validated TAL effector targets (OsTFX1, SWEET11, SWEET13, SWEET14, and OsHEN1), XA1; an R protein that recognizes TALEs independent of their transcriptional activity, and nine Arabimmune (Fig. 4C, hypergeometric  $P < 0.05$ ). Whereas the inner layers proteins of RIXIN encompass a total of 60, 80, and 102 pathogen effector targets in Inner shells\_10, Inner shells\_20, and Inner shells\_33, respectively (Fig. 4C, hypergeometric  $P < 0.05$ ). Interestingly, we found that two validated TAL effector targets (OsTFX1, SWEET14) are part of each inner shell cutoff. However, weighted  $k$ -shell decomposition identified 46, 60, and 81 candidate TAL effector targets in Inner shells\_10, Inner shells\_20, and Inner shells\_33, respectively (Fig. 4C). Furthermore, we found that 12, 18, and 19 Arabimmune is present in Inner shells\_10, Inner shells\_20, and Inner shells\_33, respectively (Fig. 4C). Additionally, the analysis uncovered that there are twice, or more total candidate TAL effector targets discovered by the weighted  $k$ -shell decomposition method than the standard centrality measures. Taken together, these observations substantiate previous reports suggesting that weighted  $k$ -shell decomposition analysis (32%) significantly increases the predictability of effector targets compared to standard network centralities (11.9%, 13.5%, 12.2%, and 12.5% for hubs, bottlenecks, information centrality, and eigenvector centrality, respectively).

To test the biological significance of internal layers' proteins at different inner shells cutoff, we used gene ontology (GO) enrichment analysis by panther [66]. Interestingly, the GO analysis revealed that most of the inner shell 10 proteins are significantly enriched in biological regulation, response to stimulus, response to stress, signal transduction, organic substance transport, intracellular transport, immune system, and MAPK cascade (Fig. 4D,  $P < 0.01$ , Table S2). Interestingly we also observe that more proteins from inner shells 20 and 30 are significantly enriched among

biological regulation, response to stimuli, response to stress, and signal transduction process supporting the significance of internal layers' proteins in RIXIN during pathogen infection. Furthermore, most of the internal layers' protein biological roles are not well characterized during pathogen infection, which suggests the need for extensive future functional determination.

Given the aforementioned observations, we extracted the validated TAL targets (OsTFX1, AAO1, OsHEN1, SULTR3:6, SWEET11, SWEET13, SWEET14, and LOC\_Os12g42970) and XA1; R protein subnetwork from RIXIN to explore the integral contribution of integrative network biology to pinpoint potential TAL effector target proteins. Interestingly, the subnetwork highlighted most of the previous candidate TAL effector targets (TFIIA-1, PRX92, LOC\_Os06g03080, DRE1J, DRIP1, FKP1, HDAC3, MLO12, PBF1, PP7, PICAKM1A, TOL6, SAPK2, SAR1, LOC\_Os12g30180, LOC\_Os02g30100, LOC\_Os02g38220, and VHA-F), and Arabimmune (CLH1, PP2A, ROC5) as the closest partners of validated TAL effector targets (Fig. 4E, Table S3). These highlighted proteins along with others should be investigated experimentally for their contribution to rice-xanthomas interaction. Given the unknown intricacy of rice-xanthomas interaction, the integrative multi-omics product; RIXIN will further expand our understanding of pivotal crosstalk and the intersection between immunity and signal transduction during pathogen infection. Moreover, these system-wide techniques can be implemented for the accelerated investigation of significantly enriched unclassified proteins from the inner vicinity of the multi-omics protein-protein interaction network and their contribution to host-pathogen interplay.

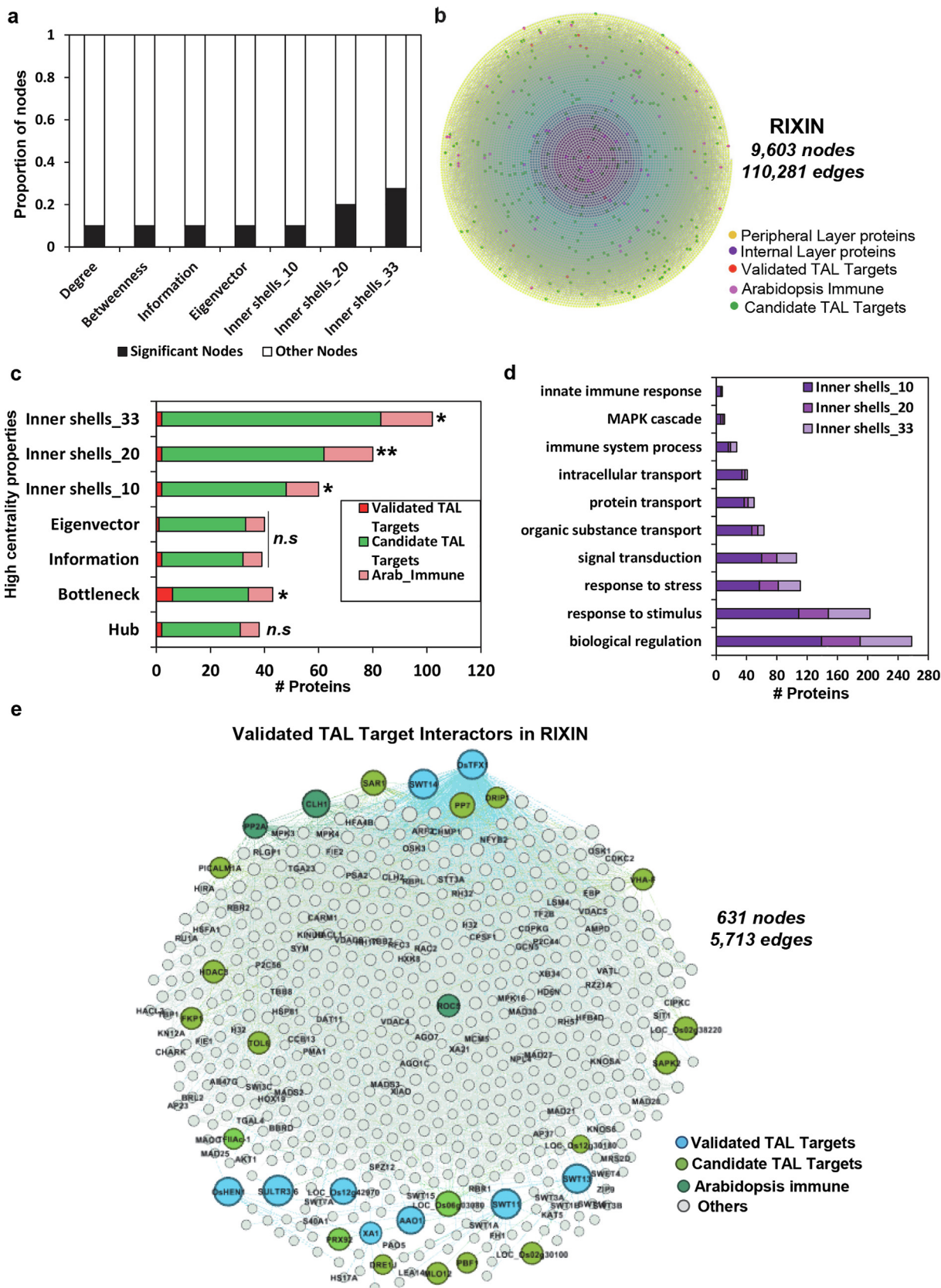
#### 4. Discussion

In the last two decades, network science has provided an excellent opportunity to evaluate diverse interactomes across animal and plant models [4,26,70]. Standard network centrality measures have been applied to interactomes to extract significant and emerging molecular components participating in different biological processes [9,12–14]. In this study, network centralities and weighted  $k$ -shell decomposition methods were used to reveal the most influential players in a comprehensive rice interactome that was built from four major sources and was termed RicePPInets. We established that RicePPInets follows the universal scale-free network property and that some highly connected nodes are central as well as the most influential information spreaders. This real-world network-specifying property has been discussed in the context of network robustness for several years. Nevertheless, some groundbreaking studies have delineated its fundamental importance [17,54,71]. Regardless of the model system under study, several prokaryotic and eukaryotic system-wide interactomes have demonstrated scale-free node distribution [13,17,23–26,72]. We also showed that a few nodes with high connectivity serve as bottlenecks, eigenvectors, and significant information spreaders (Pearson  $r = 0.66$ , 0.50, and 0.97, respectively). After



**Fig. 3.** Pathogen effector targets are enriched in Rice-Xanthomonas Interaction Network (RIXIN). (A) For comprehensive rice-Xanthomonas rice gene co-expression network construction, we followed the pipeline (see methods), which generated a massive network, RXICoNet with 34,302 nodes and 71,110,592 edges. RXICoNet was integrated with RicePPInets to get the Rice-Xanthomonas Interaction Network (RIXIN). (B) RIXIN, with 9,603 nodes and 110,281 edges. The candidate TAL effector targets by royal blue, orthologs of effector targets in Arabidopsis (Arabimmune) in and validated effector targets in red. (C) The graph illustrates the degree distribution of RIXIN and its random network. RIXIN follows scale free network properties ( $r^2 = 0.92$ ) where few proteins have higher degree than others. The degree distribution among both networks is significantly different ( $P < 0.0001$ ). (D) The quantitative relationship among four standard centralities distribution illustrates most of highly connected nodes have high information centrality, eigenvector and betweenness properties (Pearson correlation coefficient of  $r = 0.43$ – $0.96$ ,  $P < 0.0001$ ). (E) Network analysis of RIXIN identified that total pathogen targets are significantly enriched in the proteins with high betweenness but not in hub, information centrality (IC) and eigenvector centrality (EV) (hypergeometric test  $p < 0.05$ ). (F) Most of candidate pathogen targets (28) with high centrality are shared by four network parameters. Only bottleneck have 13 candidate pathogen targets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)





highlighting hubs as critical nodes in interactomes, the expansion of network theory elucidated bottlenecks in different model systems [13,73,74]. Several proteome-scale interactome studies have established the notion that hub nodes often are not essential and are generally associated with functionally redundant gene duplicates [73,75–78]. These studies paved the future directions in network science by introducing the concept of network rewiring and transcriptional reprogramming by minimizing the severity of “so-called” highly vulnerable nodes [68,79–81]. Irrespective of these debatable divergences, hub and bottleneck indices remained pivotal in host-pathogen interactome studies in plants and animals [9,12,14,21,23,40,82]. Our observations further emphasize the significance of highly connected and central nodes involved in defense and immune systems in RicePPInets. Moreover, by comparing the nodes that are associated with standard centrality features and located within the internal layers of the network, we observed that although most of the biological processes are the same, translation elongation and nitrogen starvation processes are unique to significant proteins, whereas processes including developmental, brassinosteroid-mediated signaling pathway, mRNA splicing, and translational initiation process are uniquely enriched in internal layers’ proteins. We also found that internal layers of RicePPInets have several MAP kinases (M2K1, MPK1, MPK12, MPK5, and MPK3), RLKs (RK176, RK185, CRK10, XA21), LRR receptor kinase (SERK2), and several TFs (LG2, TGA21, TGAL1, TGAL5, TGAL6, and TGAL7), which are major immune system players in host-pathogen interplay. Furthermore, we revealed that some nodes with enhanced connectivity, bottlenecks, eigenvector, and information spreaders that reside in the inner layers of the interactome are the most vulnerable proteins, which can either be the candidate or validated targets of candidate TAL effectors (Fig. 2C-F). Indeed overall, we detected the enrichment of TAL effector targets and Arab\_immune in the internal layers by two-folds as compared to standard network centralities. Interestingly, the high presence of Arab\_immune in bottlenecks suggests the evolutionary conserved immune-related proteins are central in rice interactome.

RicePPInets corresponds to system-wide PPI, so, many of the proteins are likely to be involved in several biological functions [55]. Prior to the work presented here, there was no genome-wide PPI network representing the *Xanthomonas*-infected rice interactome. Since it has been shown that hubs and bottlenecks are enriched in conditional phenotypes and immune-related nodes [12], we focused on understanding the relationship between the interactome and transcriptome specifically by examining RicePPInets in the context of rice-*Xanthomonas* interactions. It has been shown that genes expressed similarly cluster together in networks and this aids in the discovery of novel players and modules in plant-pathogen interactions [83]. Therefore, we modeled a rice-*Xanthomonas* protein-protein interaction network (RIXIN) by integrating several co-expression networks representing several categories of rice-*Xanthomonas* interaction based on the outcome of the interaction. The significantly high enrichment of TAL effector targets in the internal layers’ nodes of RIXIN demonstrates the util-

ity of integrative network biology to identify emergent proteins and modules in biotic/abiotic stress [12]. RIXIN will serve as a high-confidence computational interactome resource to study different emergent molecular players in rice-*Xanthomonas* interactions, and potential interactions of rice with other pathogens.

Consistent with previous studies, we demonstrated here that candidate and validated targets of Xoc TAL effectors and Arab\_immune nodes possess significantly enhanced centrality numbers in RicePPInets and RIXIN. While standard network centralities were predictors of a small fraction (~6.5 %) of pathogen effector targets in the Arabidopsis interactome (AI-1<sub>MAIN</sub>) [12], they predict approximately 19–26 % of the total candidate pathogen effector targets (including candidates, validated, and Arab\_immune) in Rice interactome (RicePPInets) and 11–13% of the total pathogen effector targets (including candidates, validated, and Arab\_immune) in rice-*Xanthomonas* interactome (RIXIN). Using the weighted *k*-shell network decomposition method, we achieved a tremendous improvement of 25 % (to 40 %) in RicePPInets and an improvement of 19 % (to 32 %) in RIXIN. These results are consistent with previous studies to identify pathogen target proteins in plants and humans [9,12]. Our observations further exemplify the power of network decomposition analysis, in conjunction with standard network matrices to identify hubs, bottlenecks, eigenvector, and information spreaders that allow researchers to identify the most instrumental and vulnerable points in interactomes of complex biological systems. Further, our work adds support to the general conclusion that most pathogen targets are in or in close proximity to the core, with significantly high average centralities, rather than in the periphery of the interactome [6,24,25]. Moreover, the functional analysis of inner layers’ proteins of RIXIN highlighted some of the core pathways hijacked by the pathogen to manipulate the host to its benefit, including biological regulation, response to stimulus, stress, signal transduction, transport, defense, and immune system proteins, which have been garnering increased attention recently in studies of plant-pathogen interactions [84–86]. Additionally, nine experimentally verified TAL targets (OsTFX1, XA1, AAO1, OsHEN1, SULTR3:6, SWEET11, SWEET13, SWEET14, and LOC\_Os12g42970) and their neighbors in RIXIN should be investigated for the expansion of putative pathogen targets and their contribution in plant-pathogen interaction and defense mechanisms.

It is important to note that pathogen effector targets including validated and candidate TAL effector targets as well as Arab\_immune are not enriched among hubs or bottlenecks as well as in the inner layers of degree-preserving and non-degree preserving random networks, which further highlights the specificity of our network analyses. Enrichment of candidate or validated TAL effector targets in PPInets or RIXIN does not necessarily mean that all the inner layers proteins are potential TAL effector targets. As of today, knock-out mutants of a handful of TAL effector targets have shown a measurable pathological phenotype. How the other validated or candidate TAL effector targets contribute to the pathogenesis of this deadly disease is an area of future research. It is plausible that knocking out mutants corresponding to a set of these



**Fig. 4.** Weighted *k*-shell decomposition discovered more pathogen effector target proteins with high network properties than other standard centralities. **(A)** The proportion of RIXIN nodes with significant centrality cutoff. Black are significant proteins and white are other protein of RIXIN. **(B)** The weighted *k*-shell decomposition discovered internal layers’ proteins (purple), peripheral layers’ proteins (yellow) in RIXIN. We also mapped the 318 proteins as total pathogen effector targets including 241 candidates (green), 9 validated (red) Xoc TAL effector targets, 11 predicted TAL targets (pink), and 57 proteins are orthologs of effector targets in Arabidopsis interactome (AI-1<sub>MAIN</sub>; Arab\_immune) (light purple). **(C)** The total number of pathogen effector target proteins including Arabidopsis (Arab-immune), validate and candidate pathogen target proteins reside in RIXIN with high centralities. **(D)** The gene ontology analysis of inner layers’ protein for three cutoffs of inner shells in RIXIN identifies biological regulation, response to stimulus, stress, signal transduction, transport, defense and immune system proteins are significantly enriched ( $P < 0.01$ ). **(E)** The subnetwork representing the association of validated pathogen effector targets and their first neighbors in RIXIN. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

TAL targets would not give rise to a measurable phenotype. Alternatively, the other such TAL targets may act as decoy molecules in this evolutionary tug-of-war between rice and its bacterial blast-inducing pathogens. Irrespective of this important scientific puzzle, our data established the fact that both candidate and validated TAL effector targets are enriched in the inner layers of these rice interactomes and that we increased the prediction power by implicating a weighted *k*-shell centrality measure. Therefore, it provides an important resource for future studies to prioritize the candidate TAL effector targets and rice effectors' targets in general.

In conclusion, we analyzed a genome-wide rice interactome to identify significant players, integrated *Xanthomonas*-infected rice co-expression networks to model a rice pathogen interactome representing different possible outcomes of infection and discovered an enrichment of candidate pathogen effector targets at central nodes, using several network-centric approaches. A similar pipeline can be applied to various PPI networks including human and plant interactomes to investigate the intricacy of host-pathogen interactions toward therapeutic interventions.

### Data availability

All supporting data from this study are available from the article and Supplementary Information files, or from the corresponding author upon reasonable request. Moreover, the improved weighted *k*-shell algorithm implemented in python can be accessed at <https://github.com/nilesh-iita/WkShell.2.2.1>.

### Funding

This work was supported by the National Science Foundation (IOS-1557796 and IOS-2038872) to M.S.M.

### Author contributions

M.S.M., and B.M. conceived and designed the study. B.M. and N. K. performed network and computational analyses. N.K improved the weighted *k*-shell decomposition from java to python. B.M. drafted of the manuscript. All authors discussed the results, critically reviewed the manuscript, and approved the final version.

### Competing interests

The authors declare no competing interests.

Correspondence and requests for materials should be addressed to M.S.M.

### Acknowledgments

The authors wish to acknowledge Drs. Morgan Carter and Adam Bogdanove for the conceptualization of the study. Dr. Karolina Mukhtar for critical reading of the manuscript.

### Appendix A. Supplementary data

Supplementary data (Supplemental Figs and Tables) to this article can be found online at <https://doi.org/10.1016/j.csbj.2022.04.027>.

### References

- [1] Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU. Complex networks: structure and dynamics. *Phys Rep* 2006;424(4):175–308.
- [2] Brandes U, Robins G, McCranie ANN, Wasserman S. What is network science? *Network Sci* 2013;1(1):1–15.
- [3] Garbutt CC, Bangalore PV, Kannar P, Mukhtar MS. Getting to the edge: protein dynamical networks as a new frontier in plant-microbe interactions. *Front Plant Sci* 2014;5:312.
- [4] McCormack ME, Lopez JA, Crocker TH, Mukhtar MS. Making the right connections: network biology and plant immune system dynamics. *Curr Plant Biol* 2016;5:2–12.
- [5] Das J, Fragoza R, Lee HR, Cordero NA, Guo Y, Meyer MJ, et al. Exploring mechanisms of human disease through structurally resolved protein interactome networks. *Mol BioSyst* 2014;10(1):9–17.
- [6] Borgatti SP. Centrality and network flow. *Social Networks* 2005;27(1):55–71.
- [7] Lü L, Chen D, Ren X-L, Zhang Q-M, Zhang Y-C, Zhou T. Vital nodes identification in complex networks. *Phys Rep* 2016;650:1–63.
- [8] Raman K, Damaraju N, Joshi GK. The organisational structure of protein networks: revisiting the centrality-lethality hypothesis. *Syst Synth Biol* 2014;8(1):73–81.
- [9] Kumar N, Mishra B, Mehmood A, Mohammad A, Mukhtar MS. Integrative network biology framework elucidates molecular mechanisms of SARS-CoV-2 pathogenesis. *iScience* 2020;23(9):101526.
- [10] Grassler J, Koschutski D, Schreiber F. CentiLib: comprehensive analysis and exploration of network centralities. *Bioinformatics* 2012;28(8):1178–9.
- [11] Mishra B, Kumar N, Mukhtar MS. Systems biology and machine learning in plant-pathogen interactions. *Mol Plant Microbe Interact* 2019;32(1):45–55.
- [12] Ahmed H, Howton TC, Sun Y, Weinberger N, Belkhadir Y, Mukhtar MS. Network biology discovers pathogen contact points in host protein-protein interactomes. *Nat Commun* 2018;9(1):2312.
- [13] Hu JX, Thomas CE, Brunak S. Network biology concepts in complex disease comorbidities. *Nat Rev Genet* 2016;17(10):615–29.
- [14] Mishra B, Sun Y, Ahmed H, Liu X, Mukhtar MS. Global temporal dynamic landscape of pathogen-mediated subversion of Arabidopsis innate immunity. *Sci Rep* 2017;7(1):7849.
- [15] Mishra B, Sun Y, Howton TC, Kumar N, Mukhtar MS. Dynamic modeling of transcriptional gene regulatory network uncovers distinct pathways during the onset of Arabidopsis leaf senescence. *npj Syst Biol Appl* 2018;4:35.
- [16] Pei S, Muchnik L, Andrade Jr JS, Zheng Z, Makse HA. Searching for superspreaders of information in real-world social media. *Sci Rep* 2014;4:5547.
- [17] Gao J, Barzel B, Barabasi AL. Universal resilience patterns in complex networks. *Nature* 2016;536(7615):238.
- [18] Cafarelli TM, Desbuleux A, Wang Y, Choi SG, De Ridder D, Vidal M. Mapping, modeling, and characterization of protein-protein interactions on a proteomic scale. *Curr Opin Struct Biol* 2017;44:201–10.
- [19] Lopez J, Mukhtar MS. Mapping protein-protein interaction using high-throughput yeast 2-hybrid. *Methods Mol Biol* 2017;1610:217–30.
- [20] Mishra B, Kumar N, Mukhtar MS. Network biology to uncover functional and structural properties of the plant immune system. *Curr Opin Plant Biol* 2021;62:102057.
- [21] Mukhtar MS, Carvunis AR, Dreze M, Epple P, Steinbrenner J, Moore J, et al. Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science* 2011;333(6042):596–601.
- [22] Wessling R, Epple P, Altmann S, He Y, Yang L, Henz SR, et al. Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. *Cell Host Microbe* 2014;16(3):364–75.
- [23] Pan A, Lahiri C, Rajendiran A, Shanmugham B. Computational analysis of protein interaction networks for infectious diseases. *Brief Bioinform* 2016;17(3):517–26.
- [24] Vidal M, Cusick ME, Barabasi AL. Interactome networks and human disease. *Cell* 2011;144(6):986–98.
- [25] Arabidopsis Interactome Mapping, C., *Evidence for network evolution in an Arabidopsis interactome map*. *Science*, 2011. 333(6042): p. 601–7.
- [26] Luck K, Sheynkman GM, Zhang I, Vidal M. Proteome-scale human interactomics. *Trends Biochem Sci* 2017;42(5):342–54.
- [27] Bosl K, lanevski A, Than TT, Andersen PI, Kuivanen S, Teppor M, et al. Common nodes of virus-host interaction revealed through an integrated network analysis. *Front Immunol* 2019;10:2186.
- [28] Ackerman EE, Kawakami E, Katoh M, Watanabe T, Watanabe S, Tomita Y, et al. Network-guided discovery of influenza virus replication host factors. *mBio* 2018;9(6).
- [29] Pfefferle S, Schopf J, Kogl M, Friedel CC, Muller MA, Carbajo-Lozoya J, et al. The SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors. *PLoS Pathog* 2011;7(10):e1002331.
- [30] Rozenblatt-Rosen O, Deo RC, Padi M, Adelmant G, Calderwood MA, Rolland T, et al. Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. *Nature* 2012;487(7408):491–5.
- [31] Gulbahce N, Yan H, Dricot A, Padi M, Byrdson D, Franchi R, et al. Viral perturbations of host networks reflect disease etiology. *PLoS Comput Biol* 2012;8(6):e1002531.
- [32] Abreu AL, Souza RP, Gimenes F, Consolaro ME. A review of methods for detect human Papillomavirus infection. *Virology* 2012;9:262.
- [33] Calderwood MA, Venkatesan K, Xing L, Chase MR, Vazquez A, Holthaus AM, et al. Epstein-Barr virus and virus human protein interaction maps. *Proc Natl Acad Sci U S A* 2007;104(18):7606–11.
- [34] de Chasse B, Navratil V, Tafforeau L, Hiet MS, Aublin-Gex A, Agaogou S, et al. Hepatitis C virus infection protein network. *Mol Syst Biol* 2008;4:230.
- [35] Roohvand F, Maillard P, Lavergne JP, Boulant S, Walic M, Andreo U, et al. Initiation of hepatitis C virus infection requires the dynamic microtubule

- network: role of the viral nucleocapsid protein. *J Biol Chem* 2009;284(20):13778–91.
- [36] Shapira SD, Gat-Viks I, Shum BO, Dricot A, de Grace MM, Wu L, et al. A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell* 2009;139(7):1255–67.
- [37] Simonis N, Rual JF, Lemmens I, Boxus M, Hirozane-Kishikawa T, Gatot JS, et al. Host-pathogen interactome mapping for HTLV-1 and -2 retroviruses. *Retrovirology* 2012;9:26.
- [38] Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, et al. Architecture of the human interactome defines protein communities and disease networks. *Nature* 2017;545(7655):505–9.
- [39] Lee K, Thorneycroft D, Achuthan P, Hermjakob H, Ideker T. Mapping plant interactomes using literature curated and predicted protein-protein interaction data sets. *Plant Cell* 2010;22(4):997–1005.
- [40] Smakowska-Luzan E, Mott GA, Parys K, Stegmann M, Howton TC, Layeghifard M, et al. An extracellular network of Arabidopsis leucine-rich repeat receptor kinases. *Nature* 2018;553(7688):342–6.
- [41] Klopffleisch K, Phan N, Augustin K, Bayne RS, Booker KS, Botella JR, et al. Arabidopsis G-protein interactome reveals connections to cell wall carbohydrates and morphogenesis. *Mol Syst Biol* 2011;7:532.
- [42] Gonzalez-Fuente M, Carrere S, Monachello D, Marsella BG, Cazale AC, Zischek C, et al. EffectorK, a comprehensive resource to mine for Ralstonia, Xanthomonas, and other published effector interactors in the Arabidopsis proteome. *Mol Plant Pathol* 2020;21(10):1257–70.
- [43] Nino-Liu DO, Ronald PC, Bogdanov AJ. Xanthomonas oryzae pathogens: model pathogens of a model crop. *Mol Plant Pathol* 2006;7(5):303–24.
- [44] Buttner D, He SY. Type III protein secretion in plant pathogenic bacteria. *Plant Physiol* 2009;150(4):1656–64.
- [45] Jones JD, Dangl JL. The plant immune system. *Nature* 2006;444(7117):323–9.
- [46] Tsuda K, Katagiri F. Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Curr Opin Plant Biol* 2010;13(4):459–65.
- [47] Chisholm ST, Coaker G, Day B, Staskawicz BJ. Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* 2006;124(4):803–14.
- [48] Buttner D, Bonas U. Regulation and secretion of Xanthomonas virulence factors. *FEMS Microbiol Rev* 2010;34(2):107–33.
- [49] Schulze S, Kay S, Buttner D, Egler M, Eschen-Lippold L, Hause G, et al. Analysis of new type III effectors from Xanthomonas uncovers XopB and XopS as suppressors of plant immunity. *New Phytol* 2012;195(4):894–911.
- [50] Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, et al. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 2009;326(5959):1509–12.
- [51] Moscou MJ, Bogdanov AJ. A simple cipher governs DNA recognition by TAL effectors. *Science* 2009;326(5959):1501.
- [52] Cernadas RA, Doyle EL, Nino-Liu DO, Wilkins KE, Bancroft T, Wang L, et al. Code-assisted discovery of TAL effector targets in bacterial leaf streak of rice reveals contrast with bacterial blight and a novel susceptibility gene. *PLoS Pathog* 2014;10(2):e1003972.
- [53] Wilkins KE, Booher NJ, Wang L, Bogdanov AJ. TAL effectors and activation of predicted host targets distinguish Asian from African strains of the rice pathogen Xanthomonas oryzae pv. oryzicola while strict conservation suggests universal importance of five TAL effectors. *Front Plant Sci* 2015;6:536.
- [54] Yan KK, Wang D, Sethi A, Muir P, Kitchen R, Cheng C, et al. Cross-disciplinary network comparison: matchmaking between hairballs. *Cell Syst* 2016;2(3):147–57.
- [55] Liu S, Liu Y, Zhao J, Cai S, Qian H, Zuo K, et al. A computational interactome for prioritizing genes associated with complex agronomic traits in rice (*Oryza sativa*). *Plant J* 2017;90(1):177–88.
- [56] Gu H, Zhu P, Jiao Y, Meng Y, Chen M. PRIN: a predicted rice interactome network. *BMC Bioinf* 2011;12:161.
- [57] Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* 2021;49(D1):D605–12.
- [58] Wierbowski SD, Vo TV, Falter-Braun P, Jobe TO, Kruse LH, Wei X, et al. A massively parallel barcoded sequencing pipeline enables generation of the first ORFeome and interactome map for rice. *Proc Natl Acad Sci U S A* 2020;117(21):11836–42.
- [59] Los Alamos National Laboratory, United States. Department of Energy, and United States. Department of Energy. Office of Scientific and Technical Information, Exploring network structure, dynamics, and function using networkx.
- [60] Wei B, Liu J, Wei D, Gao C, Deng Y. Weighted k-shell decomposition for complex networks based on potential edge weights. *Phys A* 2015;420:277–83.
- [61] Baryshnikova A. Exploratory analysis of biological networks through visualization, clustering, and functional annotation in cytoscape. *Cold Spring Harb Protoc* 2016;2016(6).
- [62] Zhou YL, Xu MR, Zhao MF, Xie XW, Zhu LH, Fu BY, et al. Genome-wide gene responses in a transgenic rice line carrying the maize resistance gene Rxo1 to the rice bacterial streak pathogen, *Xanthomonas oryzae pv. oryzicola*. *BMC Genomics* 2010;11:78.
- [63] Tanabe S. Spatial regulation of defense-related genes revealed by expression analysis using dissected tissues of rice leaves inoculated with Magnaporthe oryzae. *J Plant Physiol Pathol* 2014;02(04).
- [64] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf* 2008;9:559.
- [65] Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*, 2005. 4: p. Article17.
- [66] Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* 2003;13(9):2129–41.
- [67] Pei S, Muchnik L, Andrade JJS, Zheng Z, Makse HA. Searching for superspreaders of information in real-world social media. *Sci Rep* 2014;4(1):5547.
- [68] Will T, Helms V. Rewiring of the inferred protein interactome during blood development studied with the tool PPICompare. *BMC Syst Biol* 2017;11(1):44.
- [69] Mukhtar MS, McCormack ME, Argueso CT, Pajerowska-Mukhtar KM. Pathogen tactics to manipulate plant cell death. *Curr Biol* 2016;26(13):R608–19.
- [70] Cho DY, Kim YA, Przytycka TM. Chapter 5: Network biology approach to complex diseases. *PLoS Comput Biol*, 2012. 8(12): p. e1002820.
- [71] Barzel B, Barabasi AL. Universality in network dynamics. *Nat Phys* 2013;9.
- [72] He BJ. Scale-free brain activity: past, present, and future. *Trends Cogn Sci* 2014;18(9):480–7.
- [73] Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput Biol* 2007;3(4):e59.
- [74] Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature* 2001;411(6833):41–2.
- [75] Zotenko E, Mestre J, O'Leary DP, Przytycka TM. 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* 2008;4(8):e1000140.
- [76] Yu H, Braum P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, et al. High-quality binary protein interaction map of the yeast interactome network. *Science* 2008;322(5898):104–10.
- [77] Kafri R, Dahan O, Levy J, Pilpel Y. Preferential protection of protein interaction network hubs in yeast: evolved functionality of genetic redundancy. *Proc Natl Acad Sci* 2008;105(4):1243.
- [78] Jadhav A, Shanmugham B, Rajendiran A, Pan A. Unraveling novel broad-spectrum antibacterial targets in food and waterborne pathogens using comparative genomics and protein interaction network analysis. *Infect Genet Evol* 2014;27:300–8.
- [79] Liu Z, Miller D, Li F, Liu X, Levy SF. A large accessory protein interactome is regained across environments. *eLife* 2020;9:e62365.
- [80] Gagnon-Arsenault I, Marois Blanchet F-C, Rochette S, Diss G, Dubé AK, Landry CR. Transcriptional divergence plays a role in the rewiring of protein interaction networks after gene duplication. *J Proteomics* 2013;81:112–25.
- [81] Kim J, Kim I, Han SK, Bowie JU, Kim S. Network rewiring is an important mechanism of gene essentiality change. *Sci Rep* 2012;2(1):900.
- [82] Zaidi SS, Mukhtar MS, Mansoor S. Genome editing: targeting susceptibility genes for plant disease resistance. *Trends Biotechnol* 2018;36(9):898–906.
- [83] Tully JP, Hill AE, Ahmed HM, Whitley R, Skjellum A, Mukhtar MS. Expression-based network biology identifies immune-related functional modules involved in plant defense. *BMC Genomics* 2014;15:421.
- [84] Rich-Griffin C, Stechemesser A, Finch J, Lucas E, Ott S, Schafer P. Single-cell transcriptomics: a high-resolution avenue for plant functional genomics. *Trends Plant Sci* 2020;25(2):186–97.
- [85] Thordal-Christensen H. A holistic view on plant effector-triggered immunity presented as an iceberg model. *Cell Mol Life Sci* 2020;77(20):3963–76.
- [86] Zaidi SS, Naqvi RZ, Asif M, Strickler S, Shakir S, Shafiq M, et al. Molecular insight into cotton leaf curl geminivirus disease resistance in cultivated cotton (*Gossypium hirsutum*). *Plant Biotechnol J* 2020;18(3):691–706.