

RESEARCH

Open Access



Homologous mapping yielded a comprehensive predicted protein–protein interaction network for peanut (*Arachis hypogaea* L.)

Fangping Gong^{1†}, Di Cao^{1†}, Xiaojian Sun¹, Zhuo Li¹, Chengxin Qu¹, Yi Fan¹, Zenghui Cao¹, Kai Zhao¹, Kunkun Zhao¹, Ding Qiu¹, Zhongfeng Li¹, Rui Ren¹, Xingli Ma¹, Xingguo Zhang¹ and Dongmei Yin^{1*}

Abstract

Background Protein–protein interactions are the primary means through which proteins carry out their functions. These interactions thus have crucial roles in life activities. The wide availability of fully sequenced animal and plant genomes has facilitated establishment of relatively complete global protein interaction networks for some model species. The genomes of cultivated and wild peanut (*Arachis hypogaea* L.) have also been sequenced, but the functions of most of the encoded proteins remain unclear.

Results We here used homologous mapping of validated protein interaction data from model species to generate complete peanut protein interaction networks for *A. hypogaea* cv. ‘Tifrunner’ (282,619 pairs), *A. hypogaea* cv. ‘Shitouqi’ (256,441 pairs), *A. monticola* (440,470 pairs), *A. duranensis* (136,363 pairs), and *A. ipaensis* (172,813 pairs). A detailed analysis was conducted for a putative disease-resistance subnetwork in the Tifrunner network to identify candidate genes and validate functional interactions. The network suggested that DX2UEH and its interacting partners may participate in peanut resistance to bacterial wilt; this was preliminarily validated with overexpression experiments in peanut.

Conclusion Our results provide valuable new information for future analyses of gene and protein functions and regulatory networks in peanut.

Keywords Protein interactome, PPIs, Bioinformatic analysis, Disease-resistance proteins, *Ralstonia solanacearum*

Background

Proteins often cooperate with one another to carry out specific functions, and gene expression often depends on interactions between proteins [1]. At the molecular level,

protein–protein interactions (PPIs) play important roles in processes including transcription factor recruitment, enzyme activation or inactivation, cytoskeleton assembly, protein phosphorylation, and transporter activation [2–4]. PPIs are involved in various biological regulatory processes, including stress responses, signal transduction, organ formation, and even dynamic balance control [5, 6]. In plants, PPIs are critical in the regulation of signal transduction. Protein kinases and phosphatases are important components of signal transduction pathways.

[†]Fangping Gong and Di Cao contributed equally to this work.

*Correspondence:

Dongmei Yin
yindm@henau.edu.cn

¹ College of Agronomy, Henan Agricultural University, Zhengzhou 450000, People’s Republic of China



Thus, PPIs are necessary for critical physiological, pathological, and developmental processes [7].

Various methods falling into two general categories can be used to identify PPIs: physical interaction inferences, based on experimental methods; and functional relationship predictions, based on computational analyses. After the yeast two-hybrid (Y2H) technique was developed in the 1980s, many other experimental methods emerged for physical validation of PPIs. These approaches include fluorescence resonance energy transfer, bimolecular fluorescence complementation, luciferase complementation (LUC), and co-immunoprecipitation assays [8, 9]. Each method has unique features and shortcomings. Thus, PPIs generally require the use of multiple methods to verify their authenticity. However, such experimental methods generally require extensive manpower, time, and financial resources, and produce results with low accuracy and efficiency [8–11]. PPIs can therefore be predicted with computational methods such as homologous mapping, protein sequence-based prediction, and predictions with classical machine learning or deep learning techniques [10, 11]. Significant progress has been made in predicting PPIs using these methods [12–16].

A PPI network (PPIN) refers to a conceptual representation of physical interactions between proteins. Analysis of a complete protein interactome provides a valuable framework for understanding the functional organization of protein groups. Based on interactions with proteins of known function, PPINs can increase our understanding of the functions of unannotated proteins and the molecular bases of target traits. The gradual maturity of high-throughput sequencing technologies and experimental protein detection methods have enabled comprehensive PPIN construction for humans and several other model species; researchers can use bioinformatics methods to integrate data from the fields of biology, computer science, physics, chemistry, mathematics, and other disciplines to obtain genome-scale PPINs.

There are several high-profile examples in the literature of PPINs yielding valuable biological insights into plants. For example, to identify the potential functions of known protein kinases, a comprehensive protein kinase network was established for rice to predict the molecular and biological functions of kinases, especially those related to plant defense activities [6]. To understand plant abiotic stress responses, a PPIN was constructed from Y2H data, which included over 200 rice genes related to abiotic stress and seed germination [17]. Integration of a PPIN with gene expression data and quantitative trait locus data can enable systemic analyses of plant responses to abiotic and biotic factors; in rice, such a network facilitated discovery of genes previously unknown to be related to disease resistance [17]. A wheat protein

interactome for abiotic stress and development has also been constructed based on Y2H data, including 73 proteins involved in 97 interactions [18]. In that network, all of the bait proteins and the corresponding interactors were connected, which revealed complex interactions among transcription factors during flower development, abscisic acid signal transduction, and abiotic stress [18]. A genome-wide protein interactome of tea tree has also been constructed to improve understanding of molecular defense mechanisms against biotic and abiotic stresses [19]. These prior studies demonstrate that the huge datasets generated from combinations of sequencing and experimental efforts can uncover protein complexes and establish the biological roles of protein interactions [20].

Full genome sequencing has now been completed for several wild ancestors of modern peanut (*Arachis hypogaea* L.) in addition to multiple domesticated cultivars. Fully annotated genomes are available for the diploid wild species *A. duranensis* and *A. ipaensis* [21–23], the tetraploid wild species *A. monticola* [24, 25], and the tetraploid *A. hypogaea* cultivars ‘Tifrunner’ [26] and ‘Shitouqi’ [27]. However, peanut PPINs are extremely scarce to date. Deeper understanding of biological functionality in peanut thus require urgent supplementation via methods including PPIN analysis.

We here used homologous mapping to predict PPIs in peanut based on data from model species, then generated a full genome-wide peanut PPIN. This allowed us to fill gaps in our understanding of PPIs in peanut using known protein interactions in model species to shed light on the peanut proteome. The results provide a solid foundation for future analyses of PPIs in peanut. Importantly, the peanut PPIN also suggested candidate genes for future targeted breeding efforts to increase yield, disease resistance, and abiotic stress resistance in this economically important crop.

Results

Prediction of peanut PPIs

PPI data from nine model species were mapped to Tifrunner, Shitouqi, *A. monticola*, *A. duranensis*, and *A. ipaensis* (Fig. 1, Table 1, Supplementary Tables S1–S5). Tifrunner and Shitouqi are tetraploid cultivars and showed similar numbers of PPIs. The number of PPIs in *A. duranensis* and *A. ipaensis* combined was comparable to the number in Tifrunner alone. Perhaps due to the relatively high quality of the *A. monticola* assembly [24, 25], this species contained a greater number of identified proteins than either of the two tetraploid cultivars, resulting in a significantly higher number of mapped PPIs. Because Tifrunner is the most commonly used reference genome in peanut research, all subsequent analyses were conducted using data from this cultivar.

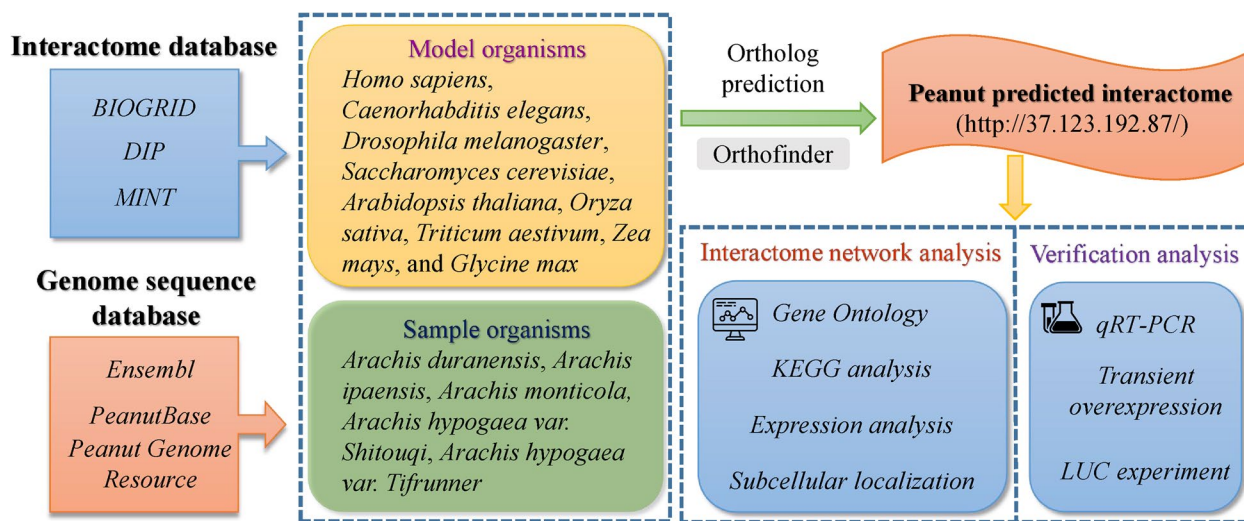


Fig. 1 Workflow showing generation of the predicted peanut protein interactome

Table 1 The summary of proteins and predicted protein interactions among various peanut reference genome

Type	Material	Total protein number	KEGG number	GO number	Predicted protein–protein interaction number	Proteins in PPI
Tetraploid cultivars	Tifrunner	84,714	22,536	38,092	282,619	17,626
	Shitouqi	88,661	14,694	10,199	256,441	17,166
Tetraploid wild species	<i>Arachis monticola</i>	75,226	22,605	35,586	440,470	19,932
Diploid wild species	<i>Arachis duranensis</i>	36,734	6230	4134	136,363	11,998
	<i>Arachis ipaensis</i>	41,840	4503	6692	172,813	13,202

Topological analysis of the peanut PPIN

Homology mapping yielded a total of 282,619 PPIs among 17,626 proteins in Tifrunner. These interactions were used to generate a predicted peanut PPIN (Fig. 2A, Supplementary Figure S1). Analysis of the network topology (Supplementary Table S6) showed that the PPIN formed one large main network with several smaller networks; the main network comprised 17,242 proteins, accounting for 97.82% of the total number of proteins present in the PPIN. Each protein had between one and 1054 degrees of connection (i.e., interactions) (mean=32) (Fig. 2B). A majority of the proteins had between one and 20 degrees of interaction, and there was a negative correlation between the degree value and the number of proteins with that number of interactions. The shortest path length in the predicted peanut PPIN was generally between one and six (Fig. 2C), indicating relatively short path lengths between any two proteins in the network. This suggested strong fault tolerance and stability in this network.

Stress centrality was determined by calculating the number of shortest paths that included a given protein.

The peanut PPIN contained a large number of highly stress-central nodes (Fig. 2D), with 72.32% of nodes having a stress centrality $> 1 \times 10^4$; this indicated high interconnectivity of the network. Nodes through which a large number of shortest paths passed were likely to correspond to key proteins that carry out important functions and have significant impacts on peanut life activities. The neighborhood connectivity showed a decreasing trend as the degree of connection increased (Fig. 2E). Some low-degree proteins interacted with highly connected proteins in their surroundings, whereas the proteins interacting with moderate-degree proteins tended to have similarly high degrees of connection. The neighborhood connectivity values of some proteins with degrees of connection between 200 and 400 were also high.

Functional annotation of the peanut PPIN

Of the 84,714 proteins encoded by the Tifrunner genome, 38,092 had Gene Ontology (GO) annotations, accounting for 44.96% of the total. However, among the 17,626 Tifrunner proteins in the PPIN, 12,558 (71.25%) had GO annotations. Thus, compared to the full protein set,

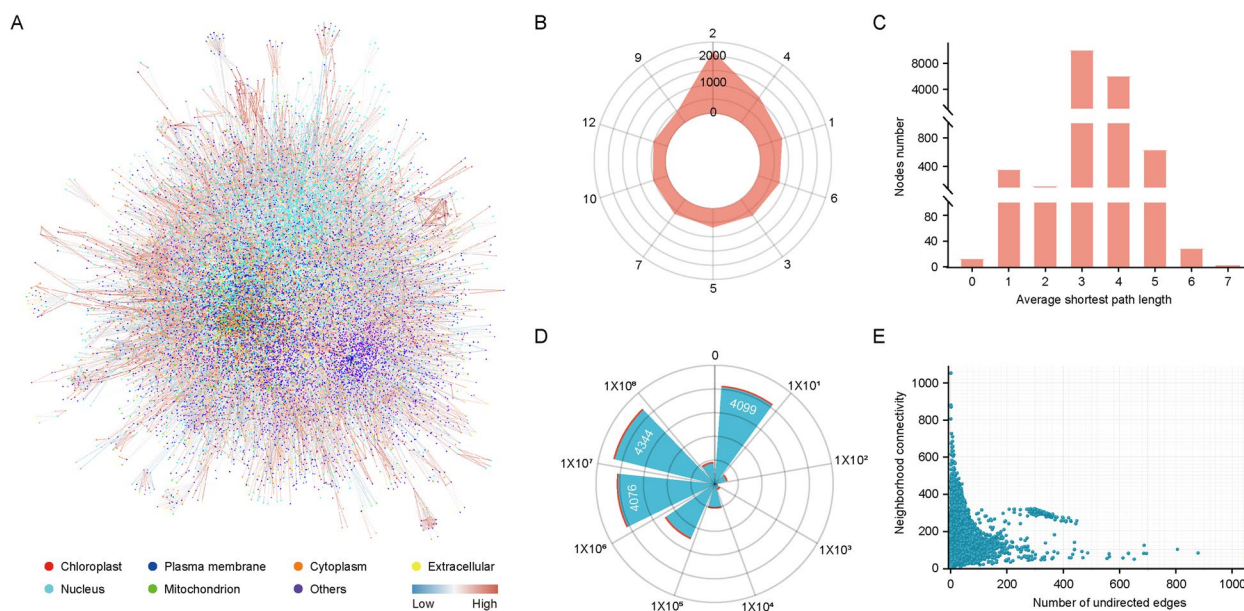


Fig. 2 Topological structure of the predicted peanut protein interaction network for Tifrunner. **A** Predicted Tifrunner master protein interaction network. **B** Distribution of protein interaction degrees. **C** Distribution of average shortest path length between all possible protein pairs. **D** Stress centrality. **E** Neighborhood connectivity

proteins present in the PPIN had higher GO annotation coverage. Statistical and clustering analyses were conducted for the GO terms, including the biological process (BP), molecular function (MF), and cellular component (CC) annotations. The most abundant BP annotation in the PPIN was “oxidation–reduction process”, which accounted for 7.8% of BP annotations. The most abundant MF term was protein binding, which accounted for 10.97% of MF annotations. The CC annotations “membrane”, “intracellular anatomical structure”, and “nucleus” had similarly high proportions, which accounted for 14.48%, 13.82%, and 16.68% of CC annotations, respectively (Fig. 3A).

Clustering was also performed on the entire set of peanut proteins with GO annotations and on the proteins in the PPIN with GO annotations. The most abundant BP, MF, and CC terms were similar between the two datasets (Fig. 3B), demonstrating that the GO annotations of proteins in the PPIN were consistent with those of the full proteome. There were 175,382 PPI pairs for which both members had GO annotations. The G-SESAME algorithm was therefore used to calculate the GO-specific annotation similarity. The Relative specificity similarity (RSS) scores were unevenly distributed, and the proportions of GO-RSS scores were very high for each annotation type: 43.22% for BP-RSS scores, 24.39% for MF-RSS scores, and 49.64% for CC-RSS scores (Fig. 3C). These PPIs therefore had high specificity and similarity in GO annotations. Overall, the results showed that most pairs

of proteins predicted to interact via homologous mapping had annotations for identical or similar functions, supporting the validity of the predicted interactions between members of that pair.

A similar analysis was next conducted for Kyoto Encyclopedia of Genes and Genomes (KEGG) biochemical pathway annotations. In the entire peanut proteome, 22,536 proteins had KEGG annotations, accounting for 26.60% of all proteins. Of the 17,626 proteins in the peanut PPIN, 7615 had KEGG annotations, corresponding to 43.20%. Thus, comparable to the GO annotations, the KEGG annotation coverage was higher in the PPIN than in the whole proteome. The 7615 PPIN proteins with KEGG annotations were distributed in 128 pathways and participated in 210,308 PPIs. The coverage of these 128 pathways ranged from 0–1 (Supplementary Figure S2A), with the autophagy pathway having the highest coverage at 95.65%. Of the 23 proteins in the autophagy pathway, 22 were present in the PPIN. The pathway with the second-highest coverage was the plant hormone signal transduction (PHST) pathway (91.11%); the coverage values of the ribosome, plant mitogen-activated protein kinase (MAPK) signaling, plant circadian rhythm, and spliceosome pathways all exceeded 80%. KEGG enrichment analysis demonstrated that pathways with higher coverage in the PPIN were also more highly enriched (Supplementary Figure S2B). The exception was the PHST pathway, which had high coverage but was not among the 20 most highly enriched pathways. However,

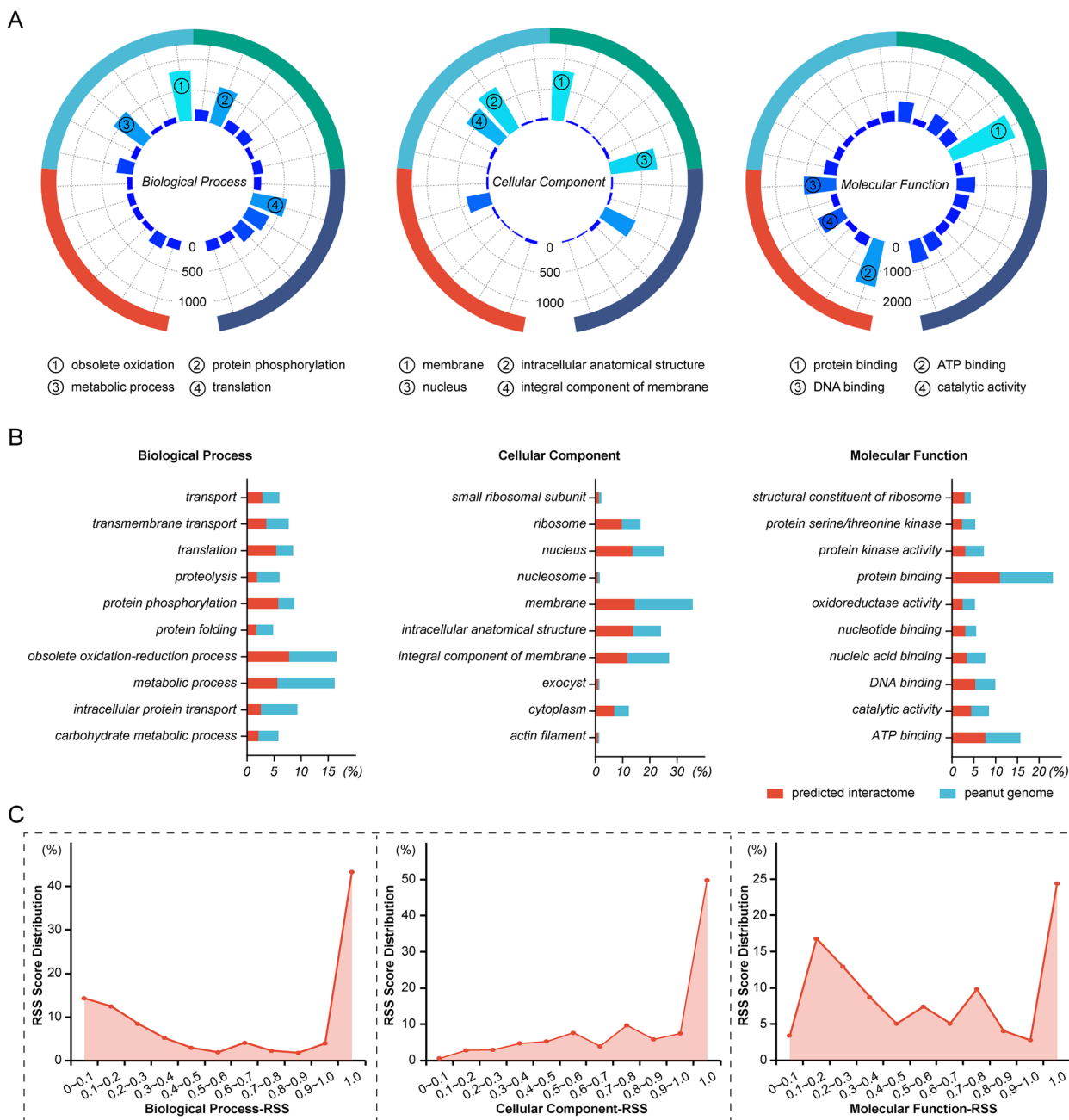


Fig. 3 Gene Ontology (GO) statistics for the predicted peanut protein interaction network for Tifrunner. **A** GO annotation statistics for the interaction network. **B** Comparison of GO clustering for the entire peanut proteome and for proteins included in the interaction network. **C** Relative specificity similarity scores distribution of the predicted peanut protein interaction network

the PHST pathway contains a small number of proteins, which contributed to the relatively high *p*-value and thus the low ranking.

Subcellular colocalization of peanut PPIN members

Predicted subcellular localization data were available for 17,623 proteins in the PPIN; these proteins were

distributed among 14 cellular regions (Supplementary Table S7). The most enriched regions were the nucleus (27.84%), the membrane (15.91%), the cytoplasm (15.54%), and the chloroplast membrane (12.42%). Of the 282,619 total predicted PPIs, members of 59,622 interactions (21.10%) had consistent predicted subcellular localizations. The largest number of predicted co-localized

protein pairs (27,728, 46.51%) were annotated as localized to the nucleus. There were 14,661, 11,288, and 2807 interacting proteins co-localized to the cytoplasm, membrane, and chloroplast membrane, respectively, accounting for 24.59%, 18.93%, and 4.71% of the total, respectively (Fig. 4A).

Interacting proteins co-localized to the nucleus formed an internal core network, with some proteins also participating in several sub-networks (Fig. 4B). KEGG analysis showed that nuclear-localized peanut proteins and their partners belonged to 70 biochemical pathways. The most

highly enriched pathways were the PHST, MAPK signaling, protein processing in the endoplasmic reticulum (PPER), ubiquitin-mediated proteolysis (UBMP), and plant-pathogen interaction (PPAI) pathways (Fig. 4C). Furthermore, 29 nuclear proteins were shared by the PHST and MAPK pathways; 31 were shared by the PPER and UBMP pathways; 22 were shared by the MAPK and PPAI pathways; and five proteins (LL1IKB, RI4L2F, FS6JVA, 3M3KZ3, 7VWF6Q) were shared by the PHST, PPAI, and MAPK pathways. There were direct correlations between the PHST and MAPK, the PPER and

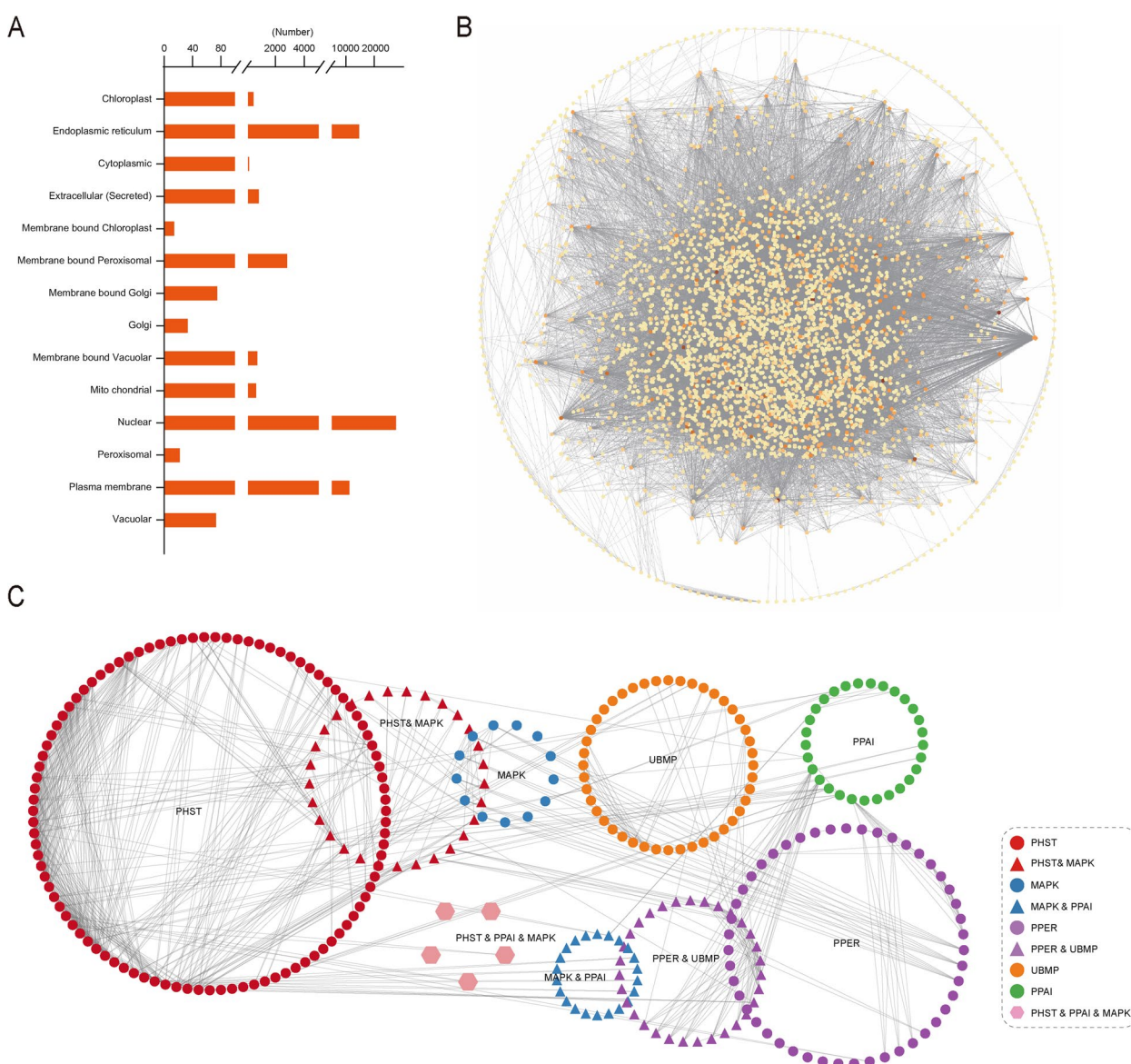


Fig. 4 Subcellular localization analysis of members of the predicted peanut protein interaction network for Tifrunner. **A** Subcellular localization annotations. **B** Visualization of the predicted peanut nuclear interactome. **C** Subnetwork of the five most enriched pathways among nuclear proteins

UBMP, and the MAPK and PPAI pathways through these co-node proteins.

Disease resistance subnetwork construction and expression analysis

Based on the KEGG data of disease resistance, 53 potential disease-resistance genes/proteins and their interacting proteins were screened in the peanut PPIN (Supplementary Table S8). These proteins were involved in 1040 PPIs. The network of disease-resistance genes was relatively dispersed (Supplementary Figure S3A). Three proteins with degrees >50 were identified: DX2UEH, JRZT96, and 458UQU. These three proteins participated in 209 PPIs and formed the center of a star-shaped subnetwork. DX2UEH was predicted to have 82 PPIs, including 13 interactions with proteins annotated as having kinase activity, 11 with proteins having serine/threonine kinase activity, and six with proteins having phosphotransferase activity (Fig. 5A). DX2UEH was connected to F5MKWZ in a branch that also included

AhIDU4K1; in a previous study from our lab, the latter protein was shown to enhance peanut resistance to *Ralstonia solanacearum* [28, 29].

To assess the reliability of the candidate disease-resistance subnetwork, the expressions of potential disease-resistance genes/proteins and their interacting genes/proteins were visualized and analyzed using previously published peanut RNA-seq data for two *A. hypogaea* cultivars after inoculation with *R. solanacearum*: the resistant and susceptible cultivars ‘H108’ and ‘H107’, respectively (Supplementary Figure S4). Further, we analyzed expression levels of *AhDX2UEH*, *AhJRZT96*, and *Ah458UQU* in H108 and H107 by qPCR. All three genes were expressed at significantly higher levels in H108 than in H107 after infection. In H108, the genes were upregulated at 1 d post-inoculation (dpi) and downregulated at 7 dpi. In H107, the genes were gradually downregulated over the entire period (Fig. 5B). In the branch of the network connecting *AhDX2UEH* to *AhF5MKWZ*, expression patterns were more complex. *AhZQ326X*, *AhXP9K23*,

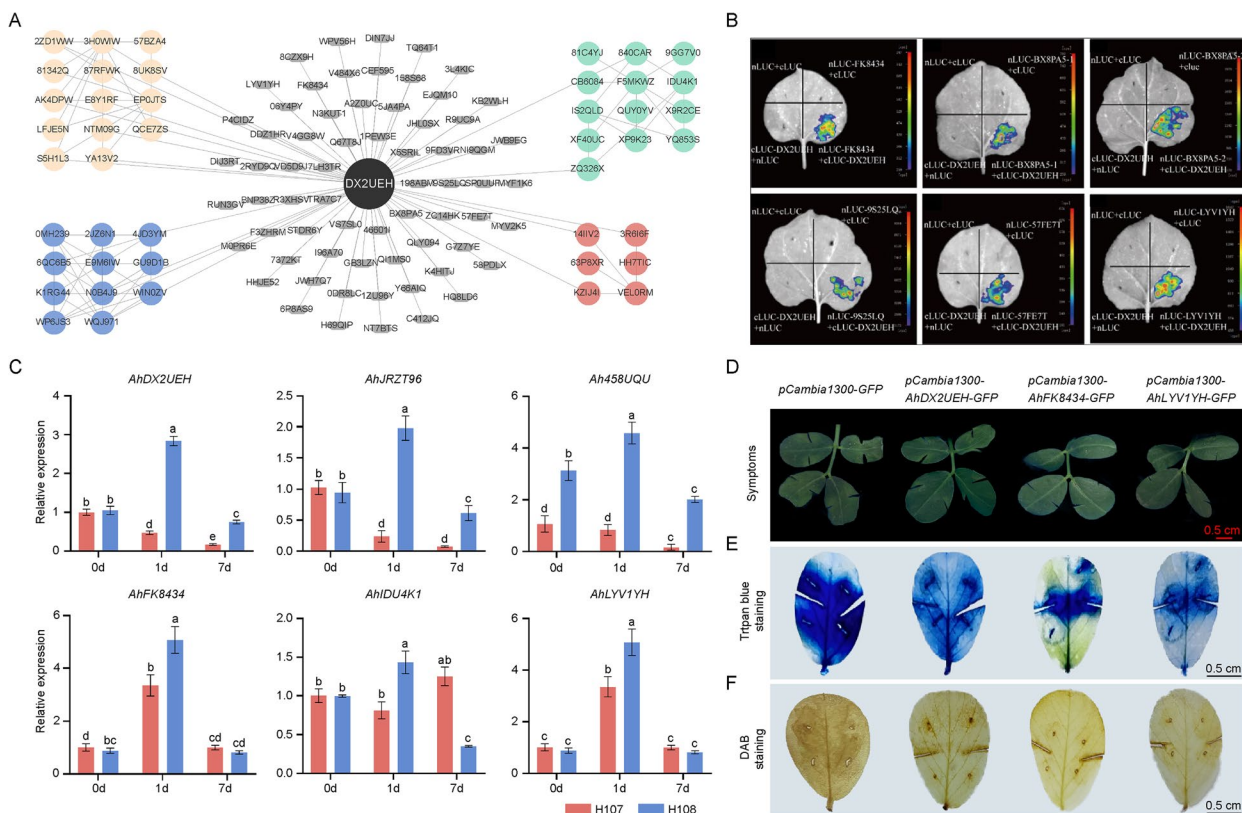


Fig. 5 Analysis of putative disease-resistance proteins based on the predicted peanut protein interaction network for Tifrunner. **A** Putative peanut disease-resistance subnetwork (DX2UEH interactors). **B** Validation of predicted peanut disease-resistance protein–protein interactions via luciferase complementation assays. **C** Expression levels of selected genes in the disease-resistance subnetwork after inoculation with *R. solanacearum*. Error bars represent the mean ± SD of three biological replicates, and different letters indicate statistically significant differences at *p* < 0.05 based on the Tukey–Kramer test. **D–F** Symptoms (**D**), trypan blue staining (**E**), and 3, 3'-diaminobenzidine (DAB) staining (**F**) in peanut leaves transiently overexpressing *AhDX2UEH*, *AhFK8434*, or *AhLYV1YH* after inoculation with *R. solanacearum*

AhYQ853S, and *AhF5MKWZ* each showed an initial increase followed by a decrease in expression in both cultivars, but *AhZQ326X*, *AhXP9K23*, and *AhF5MKWZ* were expressed at significantly higher levels in H108 than H107. *AhF5MKWZ* was upregulated by nearly six-fold at 1 dpi in H108, compared to an increase of just two-fold in H107. In H107, *AhIS2QLD* and *AhIDU4K1* were down-regulated at 1 dpi, but expression levels later recovered; in H108, the same genes were first upregulated, then down-regulated (Fig. 5B, Supplementary Figure S3B).

Six proteins predicted to interact with DX2UEH were randomly selected from the disease-resistance subnetwork for validation with LUC assays: LYV1YH, FK8434, BX8PA5-1, BX8PA5-2, 57FE7T and 9S25LQ. All of these proteins except 2RYD9Q exhibited significant fluorescence signals, indicating interactions with DX2UEH *in vivo*. Interestingly, amplification of *AhBX8PA5* for the LUC assay yielded a previously unidentified transcript, *AhBX8PA5-2*. This transcript contained an additional 156-bp fragment in the middle of the previously identified coding sequence (CDS), and the corresponding protein was found to interact with AhDX2UEH (Fig. 5C).

To further validate the functions of these proteins, peanut leaves transiently overexpressing *AhDX2UEH*, *AhFK8434*, or *AhLYV1YH* were inoculated with *R. solanacearum*, then the level of resistance was estimated. Control leaves were susceptible, showing wilting at 3 dpi. In contrast, leaves transiently overexpressing *AhDX2UEH-GFP*, *AhFK8434-GFP*, or *AhLYV1YH-GFP* showed a normal (uninfected) phenotype (Fig. 5D). After trypan blue staining, the empty-vector control leaves were a much deeper shade of blue than leaves overexpressing any of the three putative resistance-related genes (Fig. 5E). Finally, H₂O₂ accumulation in the leaves was estimated with 3, 3'-diaminobenzidine (DAB) staining. H₂O₂ accumulation was comparable between the leaves overexpressing any of the three genes and uninoculated leaves (Fig. 5F). Overall, peanut leaves transiently overexpressing *AhFK8434* or *AhLYV1YH* showed strong *R. solanacearum* resistance, whereas those overexpressing *AhDX2UEH* showed significant but weak resistance.

Discussion

Predicted peanut PPIs

We here conducted a comprehensive analysis of PPIs in several peanut species via homologous mapping. For comparison, the STRING database (<https://cn.string-db.org/>) [30] was used to predict PPIs among several peanut cultivars using the homologous mapping method, which yielded 71,122 PPIs. This was much smaller than the number of PPIs obtained with the method used in the present study, validating our approach.

Numerous studies have used homologous mapping to identify putative PPIs in model species. For example, PPI data collected from yeast, nematode, fruit fly, and human were used to predict PPIs in *Arabidopsis thaliana* via homologous mapping; this generated 19,979 predicted PPIs among 3,617 *Arabidopsis* proteins [31]. Zhu et al. used homologous mapping to translate PPI data from *Arabidopsis*, yeast, human, fruit fly, nematode, and *Escherichia coli* to rice, resulting in 76,585 predicted PPIs among 5049 proteins [32]. Finally, PPI data obtained from many species, including nine eukaryotes (rice, *Arabidopsis*, human, mouse, rat, fruit fly, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, and *Schizosaccharomyces pombe*), four prokaryotes (*E. coli*, *Bacillus subtilis*, *Helicobacter pylori*, and *Campylobacter jejuni*), and *Chlamydomonas reinhardtii* were mapped to maize, resulting in 49,026 predicted PPIs among 6004 proteins [33]. These earlier studies were conducted with fewer available PPI data from model species, partially explaining the reduced number of predicted PPIs compared to the results of the present study in peanut. The larger number of PPIs here may also have been due to the relatively large size of the peanut genome. We therefore expect that future additions to proteomic and PPI datasets in various species will allow predictions of still more PPIs in peanut.

In addition to homologous mapping, several bioinformatics methods have been used to predict PPIs from combinations of features, including amino acid sequences, gene co-expression data, functional associations, and phylogenetic relationships. For example, a machine learning model trained on these features was used to predict 50,220 PPIs in *Arabidopsis* [34]. A program called DeepPPI was constructed using deep neural networks to effectively learn PPIs from commonly used protein descriptors [35]. A deep ensemble learning method, EnAmDNN, has also been designed to predict PPIs [36]. In all of these cases, various protein features are extracted, then classical machine learning or deep learning methods are applied to predict PPIs.

Most known PPIs occur in animals and microorganisms; relatively few have been experimentally identified in plants. At present, there are too few published PPIs for peanut to enable the use of machine learning to predict additional PPIs. However, machine learning methods could be used to predict PPIs in peanut after training on datasets from other plant species. Combining these predictions with existing PPI data obtained through experimentation or homologous mapping could yield more reliable and larger-scale PPI data for peanut. This would combine the distinct advantages of several complementary approaches.

Network analysis of predicted peanut PPINs

Of the 84,714 proteins produced by Tifrunner, 17,626 were predicted to be involved in 282,619 PPINs; these predicted interactions were used to generate the peanut PPIN. There was a very large main network consisting of 17,242 proteins, consistent with the phenomenon of most proteins having multiple interaction partners. Topological analysis [37] of the PPIN revealed 107 connected components, comparable to prior results in *Oryza sativa* and *C. elegans* [38]. The average degree of each node in the network was 32, which was also similar to results in *O. sativa* and human [39] and indicated relatively tight internal connections. Most pairs of proteins in the network had path lengths between one and six, and the network overall had a relatively small average shortest path length, indicating that it had the small-world property [40] and was thus relatively stable. For example, these characteristics indicated that peanut should be able to respond quickly to external stressors and to compensate for the loss of a given protein through other pathways with relative ease.

Proteins with high centrality in a PPIN are likely to have key functions, which is important in exploring protein functionality both in the PPIN and biologically. Furthermore, proteins with low connectivity overall but high neighborhood connectivity are likely to be important links between pathways. Proteins with high overall connectivity generally occupy a central position in a PPIN, whereas proteins with intermediate connectivity but high neighborhood connectivity should occupy position just outside the core. Proteins with high centrality and high neighborhood connectivity are likely to be critical in various biological activities and should therefore be prioritized for future in-depth studies.

We examined the reliability of the constructed PPIN from various perspectives through analyses of similarity in GO and KEGG annotations, subcellular localization, and gene expression between members of a PPI pair. For example, a GO-specific similarity score was calculated for each predicted interaction pair [40]; a score > 0.5 was classified as significant. For the BP, MF, and CC annotations, 51,480, 71,622, and 52,263 pairs of proteins with a score > 0.5 were found to interact with each other, accounting for 56.97% of BP-RSS scores, 53.35% of MF-RSS scores, and 83.97% of CC-RSS scores, respectively (Fig. 2). These pairs were considered highly likely to interact. High scores in the subcellular co-localization and gene co-expression analyses further increased the credibility of specific predicted PPINs.

Websites such as IntAct and BioGRID have scoring criteria for PPINs [30, 41, 42] that incorporate factors such as gene co-expression, protein co-localization, the number of articles that validate the interaction, and validation

methods. Some researchers have also used existing PPI data for a given species to functionally evaluate a predicted PPIN. The results of the present study were based on predicted PPINs; we therefore used a method combining co-expression, co-localization, and GO similarity to estimate predicted PPI validity. Future experimental data can be used to further validate the predicted peanut PPIN.

Disease-resistance subnetwork analysis

We here identified putative disease-resistant proteins and their interacting proteins in the putative disease-resistance subnetwork for Tifrunner. DX2UEH, JRZT96, and 458UQU are predicted disease-resistance proteins in the CC-NBS-LRR family; all three had interaction degrees > 50, indicating centrality in some biological process. DX2UEH had the highest degree, and most of its predicted protein interactors were annotated as having kinase activity. To validate the roles of these proteins in plant defense, expression levels of the genes encoding them were analyzed after plant infection with *R. solanacearum*. *AhDX2UEH*, *AhJRZT96*, and *Ah458UQU* were upregulated in the resistant peanut cultivar H108 after inoculation, and there were significant differences in the expression levels of the three genes between H108 and the susceptible cultivar H107 at both 1 and 7 dpi. We therefore hypothesized that *AhDX2UEH*, *AhJRZT96*, and *Ah458UQU* had either a direct positive response or a positive regulatory effect in peanut that promoted disease resistance.

We also examined expression levels of the genes encoding the proteins present in the branch network connected to DX2UEH and F5MK2Z after infection. These responses were markedly more complex, exhibiting both positive and negative responses in H107 and H108. *AhIDU4K1* has previously been shown to enhance peanut resistance to bacterial wilt disease [28, 29], supporting the validity of the PPIN in revealing proteins associated with disease resistance. Further studies are needed to investigate the mechanism by which the branch network regulates and affects peanut disease resistance phenotypes.

Six proteins in the disease-resistance subnetwork predicted to interact with DX2UEH were randomly selected for experimental validation with LUC assays. Five of the proteins were found to interact with DX2UEH *in vitro*, for a validation rate of 83.3%. Although the sample size was small, it did verify the accuracy of the predicted peanut PPIN obtained with homology mapping. Of the six genes in the five validated interactions, five (all except *BX8PA5*) were upregulated after inoculation with *R. solanacearum*. Furthermore, the five interactions were all related to regulation of peanut resistance to bacterial

wilt disease. Importantly, the validation experiment incidentally uncovered a novel transcript, *BX8PA5-2*, and the encoded protein also interacted with DX2UEH. The protein encoded by the novel transcript and its interactions with DX2UEH are promising candidates for further experimental investigation. Finally, transient overexpression experiments in peanut leaves indicated that DX2UEH and its protein interactors induced varying degrees of *R. solanacearum* resistance. Further study will be required to establish the mechanisms by which members of the subnetwork collaborate to participate in disease resistance.

Conclusion

We here used homology mapping to predict proteome-wide PPIs in several peanut species and cultivars. The PPIN of Tifrunner formed a large main network with tight internal connections and overall stability. Topological analysis revealed some key proteins with high degrees of interaction and high centrality. Proteins contained in the PPIN included most of the GO terms and KEGG pathways annotated in peanut, including many important biological processes. Five out of six randomly selected predicted PPIs were experimentally confirmed through LUC assays. Both analysis of a putative disease-resistant subnetwork and experimental validation indicated that proteins in the subnetwork were indeed involved in enhancing peanut disease resistance. The results of this study provide valuable new avenues for basic research into peanut proteins associated with agronomically important traits such as high yield, high oil content, stress resistance, and high nutritional value. Future studies should focus on experimental protein interactome validation; those data can then be used to train machine learning models for genome-wide peanut PPI predictions.

Materials and methods

Source of protein sequence data and experimental protein interaction data

The protein sequence data of nine model organisms, including *Homo sapiens*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Oryza sativa*, *Triticum aestivum*, *Zea mays*, and *Glycine max* was retrieved from Ensembl database (<http://asia.ensembl.org/index.html>) (Supplementary Table S9). For peanut, the protein sequence data of Tifrunner1.0, *A. duranensis* and *A. ipaensis* was retrieved from PeanutBase (<https://www.peanutbase.org/>), Shitouqi was from Peanut Genome Resource (<http://peanutgr.fafu.edu.cn/>), and *A. monticola* was from our own laboratory (Table 1) [24, 25]. Protein–protein interaction data from the above model organisms were downloaded

from the public protein–protein interaction databases, including BioGrid (<https://www.ebi.ac.uk/intact/>), IntAct (<https://www.ebi.ac.uk/intact/>), DIP (<https://comp-sys-bio.org/dipos/>), and MINT (<http://cbm.bio.uniroma2.it/mint/>) (Supplement Table S1). No protein–protein interaction data for peanut was found in those database for now.

Protein interaction mapping

The OrthoFinder software was used to identify orthologous proteins and orthologous protein groups between each species and the peanut genome. Interolog mapping, a method based on evolutionary conservation of protein–protein interactions across species, was used to map protein interactions from each species onto the peanut proteome. Interolog mapping is a well-established method for predicting protein–protein interactions, based on the fundamental premise that these interactions co-evolve with the conservative evolution of proteins. If Protein A and Protein B interact in one species, it is highly likely that their homologs in another species, Protein C (homologous to Protein A) and Protein D (homologous to Protein B), also interact with each other.

The acquisition of peanut protein Gene Ontology (GO) data and KEGG data

Tifrunner1.0 protein sequence was annotated using EggNOG-mapper [43] with a one-to-one orthologous annotation method to obtain the KEGG annotation. All KEGG pathway K numbers were downloaded from <https://www.kegg.jp/kegg/> along with 136 pathways for peanuts (KEGG organisms ID: ahf), and organized to list the K numbers included in each KEGG pathway for peanuts. GO annotation information provided by Bertoli et al. was used [26]. The genome annotation data was used for Shitouqi and *A. monticola*, while EggNOG-mapper was used for *A. duranensis* and *A. ipaensis*. The KEGG and GO enrichment analysis and visualization were performed using TBtools software [44] and the OmicStudio platform (<https://www.omicstudio.cn/index>).

Co-expression analysis

We downloaded RNA-seq data from 22 different tissues of Tifrunner1.0 at different growth and developmental stages [45]. The raw sequencing data were mapped to the Tifrunner1.0 genome using Hisat2 software [46]. FeatureCounts software [47] was used for expression quantification to obtain the Fragments Per Kilobase per Million mapped reads (FPKM) expression data for each gene. The Pearson correlation coefficient was calculated for each pair of genes using the R language. Each pair of

interacting proteins has a gene coexpression correlation coefficient that ranges from -1 to 1, with highly correlated values indicating a greater likelihood of being co-expressed and functionally related [48].

Subcellular localization prediction of peanut proteins

ProtComp (<http://www.softberry.com/berry.phtml>) integrates several protein localization prediction methods: neural network-based predictions; direct comparison with homologous proteins of known localization; analysis of pentamer distribution to compute queries and database sequences; and prediction of certain functional peptide sequences, such as mitochondrial and chloroplast signal peptides, transport peptides, and transmembrane segments. By combining these methods, the predicted locations are scored on a scale from 0 to 1, with higher scores indicating a higher probability of the protein being localized to that region. Here, ProtComp was used for subcellular localization prediction of peanut proteins, with a winner-takes-all approach that considers the highest-scoring subcellular region as the protein's localization region. If two interacting proteins are predicted to localize to the same region, they are considered to be co-localized in the same subcellular compartment.

GO specific similarity calculation method

GO is an important semantic description system in the field of life science research. It integrates information from multiple databases, annotates and classifies gene function using structured terms, and represents hierarchical relationships between terms using a directed acyclic graph. Although GO provides GO terms for each gene, a challenge remains in accurately measuring the semantic similarity between two GO terms to determine the functional similarity between genes. In this experiment, we adopt an algorithm called G-SESAME [49, 50] to calculate the similarity between GO terms of two genes and provide a score. The algorithm was according to the previous report [50].

Using algorithms provided by previous studies, we developed a Python script that can run in batches to calculate the GO-specific similarity of protein interactions in a protein interaction prediction network that have GO annotations. The semantic similarity between the GO terms of the two genes obtained from the formula ranges between 0 and 1, with a higher similarity indicating a stronger correlation. The reliability of the peanut protein interactions at the GO annotation level was verified by calculating the RSS scores [40, 51, 52], which also demonstrated the degree of correlation between protein interactions at the GO annotation level.

Plant materials and inoculation by *Ralstonia solanacearum*

A. hypogaea var. H108 (resistant to *R. solanacearum*) and H107 (susceptible to *R. solanacearum*) [28, 29] were used as materials. The method of plant inoculation by *R. solanacearum* was according to our previous report [28, 29]. Leaves after inoculation for 0, 1, and 7 days were used for RNA extraction and further qRT-PCR. The 3–5 week-old tobacco (*Nicotiana benthamiana*) plants were used for LUC experiment.

qRT-PCR

Primers were designed using Primer 6.0 (Supplementary Table S10) and synthesized by Genaray Biotech (Shanghai) Co., Ltd. Total RNA was extracted using the TransZol Plant RNA extraction kit (TransGen Biotech), and cDNA was synthesized using the EasyScript One-Step gDNA Removal and cDNA Synthesis SuperMix kit (TransGen Biotech). qRT-PCR was performed using the PerfectStart Green qPCR SuperMix kit (Quanta Bio) following the instructions provided for specific operation steps and system configuration. Quantitative data was analyzed using the $2^{-\Delta\Delta Ct}$ method, with variance analysis performed using SPSS software and graphs created using Prism.

Luciferase complementary assay

Gene primers for LUC experiment were shown in Supplementary Table S11. PCR products were recovered by gel extraction using Gel Extraction Kit (OMEGA). The target gene and linearized vector (pCAMBIA 1300-nLUC and pCAMBIA1300-cLUC) were subjected to homologous recombination using Seamless Assembly Cloning Kit (Clone Smarter). For DH5 α with successful bacterial transformation, their plasmid was extracted using Plasmid Mini Kit (OMEGA). The extracted plasmid was transformed into *Agrobacterium* EHA105. After suspension culture, the back of tobacco leaves was injected with bacterial liquid. The leaves were photographed and observed using a plant live imaging system. The experiment was set up with three independent biological replicates.

Transient overexpression and trypan blue and diaminobenzidine staining in peanut leaves

Disease resistance related genes were selected and were transiently overexpressed in peanut leaves through the *Agrobacterium*-mediated method according to our previous study [28, 29]. The inoculated peanut leaves were further stained by diaminobenzidine (DAB) and trypan blue also according to Zhao et al. [28, 29]. The experiment was set up with three independent biological replicates.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-024-05580-w>.

Supplementary Material 1. Table S1. Predicted protein–protein interaction data for the tetraploid *Arachis hypogaea* cv. 'Tifrunner'. Table S2. Predicted protein–protein interaction data for the tetraploid *Arachis hypogaea* cv. 'Shitouqi'. Table S3. Predicted protein–protein interaction data for the tetraploid wild species *Arachis monticola*. Table S4. Predicted protein–protein interaction data for the diploid wild species *Arachis duranensis*. Table S5. Predicted protein–protein interaction data for the diploid wild species *Arachis ipaensis*. Table S6. Proteins predicted to participate in protein–protein interactions in Tifrunner. Table S7. Subcellular co-localization data for the predicted interacting protein pairs in Tifrunner. Table S8. Information about interacting proteins in the putative disease-resistance subnetwork for Tifrunner. Table S9. Information about model species collected from multiple databases. Table S10. Primers for quantitative reverse transcription (qRT)-PCR. Table S11. Primers for luciferase complementation assays.

Supplementary Material 2. Figure S1. Detailed information for the predicted peanut protein interaction network. A: Distribution of protein interaction degrees. B: Subnetwork for chloroplast-localized proteins. C: Subnetwork for plasma membrane-localized proteins. D: Subnetwork for cytoplasm-localized proteins. E: Subnetwork for extracellular-localized proteins. F: Subnetwork for nuclear-localized proteins. G: Subnetwork for mitochondrial-localized proteins. H: Subnetwork for proteins located in other cellular compartments

Supplementary Material 3. Figure S2. Kyoto Encyclopedia of Genes and Genomes (KEGG) coverage and enrichment analyses in the predicted peanut protein interaction network. A: KEGG pathway coverage. B: KEGG enrichment analysis

Supplementary Material 4. Figure S3. Disease-resistance subnetwork and selected putative disease-resistance gene expression analysis. A: Peanut disease-resistance subnetwork. B: Expression levels of selected genes encoding putative disease-resistance proteins. Error bars represent the mean \pm SD of three biological replicates, and different letters indicate statistically significant differences at $p < 0.05$ based on the Tukey-Kramer test

Supplementary Material 5. Figure S4. The expression of RNA-seq data for genes in the putative disease-resistance subnetwork in the peanut cultivars H108 and H107 after inoculation with *Ralstonia solanacearum*.

Acknowledgements

Not applicable.

Authors' contributions

Dongmei Yin and Fangping Gong conceived the project and designed research; Chengxin Qu and Di Cao performed research; Chengxin Qu, Xiaojian Sun, Zhan Li, and Di Cao, contributed new reagents/analytic tools; Fangping Gong, Dongmei Yin and Chengxin Qu analyzed data and wrote the paper; Kuopeng Wang, Yi Fan, Zhuo Li, Zenghui Cao, Kai Zhao, Kunkun Zhao, Ding Qiu, Zhongfeng Li, Rui Ren, Xingli Ma, Xingguo Zhang reviewed and approved the manuscript for publication.

Funding

This work was supported by grants from the Key Program of National Natural Science Foundation of China (NSFC)-Henan United Fund (No. U1704232), Key Scientific and Technological Project of Henan Province (No. 221111110500; 161100111000; HARS-22-05-G1), the Key Scientific Research Project of Henan Higher Education Institutions (24A210007), National Science Foundation of Henan Province (No. 222300420178).

Availability of data and materials

The full PPI data for five peanut materials (the tetraploid *A. hypogaea* cultivars Tifrunner and Shitouqi, the tetraploid wild species *A. monticola*, and the diploid wild species *A. duranensis* and *A. ipaensis*) have been uploaded to github (<https://github.com/PeanutLaboratory/PeanutPredictedPPI>). The data are also publicly available at <http://37.123.192.87/>.

Data availability

The full PPI data for five peanut materials (the tetraploid *A. hypogaea* cultivars Tifrunner and Shitouqi, the tetraploid wild species *A. monticola*, and the diploid wild species *A. duranensis* and *A. ipaensis*) is provided within the manuscript or supplementary information files and also have been uploaded to github (<https://github.com/PeanutLaboratory/PeanutPredictedPPI>). The data are also publicly available at <http://37.123.192.87/>.

Declarations

Ethics approval and consent to participate

This study was not carried out on animal or endangered species. We declare that all experimental materials were purchased from local authorities of agricultural farm.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 23 January 2024 Accepted: 9 September 2024

Published online: 20 September 2024

References

1. Cho DY, Kim YA, Przytycka TM, et al. Chapter 5: network biology approach to complex diseases. *PLoS Comput Biol*. 2012;8:e1002820.
2. Cui J, Li P, Li G, et al. AtPID: Arabidopsis thaliana protein interactome database—an integrative platform for plant systems biology. *Nucleic Acids Res*. 2008;36:D999-1008.
3. Tai YS. Interactome of signaling networks in wheat: the protein-protein interaction between TaRAR1 and TaSGT1. *Mol Biol Rep*. 2008;35:337–43.
4. Altmann M, Altmann S, Rodriguez PA, et al. Extensive signal integration by the phytohormone protein network. *Nature*. 2020;583:271–6.
5. Morsy M, Gouthu S, Orchard S, et al. Charting plant interactomes: possibilities and challenges. *Trends Plant Sci*. 2008;13:183–91.
6. Ding X, Richter T, Chen M, et al. A rice kinase-protein interaction map. *Plant Physiol*. 2009;149:1478–92.
7. von Mering C, Krause R, Snel B, et al. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*. 2002;417:399–403.
8. Xing S, Wallmeroth N, Berendzen KW, Grefen C. Techniques for the analysis of protein-protein interactions in vivo. *Plant Physiol*. 2016;171(2):727–58.
9. Snider J, Kotlyar M, Saraon P, Yao Z, Jurisica I, Stagljar I. Fundamentals of protein interaction network mapping. *Mol Syst Biol*. 2015;11(12):848.
10. Chang JW, Zhou YQ, Ul Qamar MT, Chen LL, Ding YD. Prediction of protein-protein interactions by evidence combining methods. *Int J Mol Sci*. 2016;17(11):1946.
11. Hu L, Wang X, Huang YA, Hu P, You ZH. A survey on computational models for predicting protein-protein interactions. *Brief Bioinform*. 2021;22(5):bbab036.
12. Wang XW, Madeddu L, Spirohn K, et al. Assessment of community efforts to advance network-based prediction of protein-protein interactions. *Nat Commun*. 2023;14(1):1582.
13. Gao Z, Jiang C, Zhang J, et al. Hierarchical graph learning for protein-protein interaction. *Nat Commun*. 2023;14(1):1093.
14. Roslan R, Othman RM, Shah ZA, et al. Utilizing shared interacting domain patterns and gene ontology information to improve protein-protein interaction prediction. *Comput Biol Med*. 2010;40(6):555–64.
15. Zhang F, Liu S, Li L, Zuo K, Zhao L, Zhang L. Genome-wide inference of protein-protein interaction networks identifies crosstalk in abscisic acid signaling. *Plant Physiol*. 2016;171(2):1511–22.

16. Dong S, Lau V, Song R, et al. Proteome-wide, structure-based prediction of protein-protein interactions/new molecular interactions viewer. *Plant Physiol.* 2019;179(4):1893–907.
17. Cooper B, Clarke JD, Budworth P, et al. A network of rice genes associated with stress response and seed development. *Proc Natl Acad Sci USA.* 2003;100:4945–50.
18. Tardif G, Kane NA, Adam H, et al. Interaction network of proteins associated with abiotic stress response and development in wheat. *Plant Mol Biol.* 2007;63:703–18.
19. Singh G. Genome-wide interologous interactome map (TeaGPIN) of *Camellia sinensis*. *Genomics.* 2021;113:553–64.
20. Petrakis S, Andrade-Navarro MA. Editorial: protein interaction networks in health and disease. *Front Genet.* 2016;7:111.
21. Bertioli DJ, Cannon SB, Froenicke L, et al. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat Genet.* 2016;48:438–46.
22. Chen X, Li H, Pandey MK, et al. Draft genome of the peanut a-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens. *Proc Natl Acad Sci USA.* 2016;113:6785–90.
23. Lu Q, Li H, Hong Y, et al. Genome sequencing and analysis of the peanut B-Genome progenitor (*Arachis ipaensis*). *Front Plant Sci.* 2018;9:604.
24. Yin DM, Ji CM, Ma XL, et al. Genome of an allotetraploid wild peanut *Arachis monticola*: a de novo assembly. *Gigascience.* 2018;7:1.
25. Yin DM, Ji CM, Song QX, et al. Comparison of *Arachis monticola* with diploid and cultivated tetraploid genomes reveals asymmetric subgenome evolution and improvement of peanut. *Adv Sci.* 2020;7:1901672.
26. Bertioli DJ, Jenkins J, Clevenger J, et al. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea* L. *Nat Genet.* 2019;51:877–84.
27. Zhuang WJ, Chen H, Yang M, et al. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat Genet.* 2019;51:865–76.
28. Zhao K, Ren R, Ma XL, et al. Genome-wide investigation of defensin genes in peanut (*Arachis hypogaea* L.) reveals AhDef2.2 conferring resistance to bacterial wilt. *Crop J.* 2022;10:809–19.
29. Zhao K, Li Z, Ke Y, Ren R, et al. Dynamic N6-methyladenosine RNA modification regulates peanut resistance to bacterial wilt. *New Phytol.* 2024;242(1):231–46.
30. Szklarczyk D, Kirsch R, Koutrouli M, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* 2023;51:D638–46.
31. Geisler-Lee J, O'Toole N, Ammar R, et al. A predicted interactome for *Arabidopsis*. *Plant Physiol.* 2007;145:317–29.
32. Zhu P, Gu H, Jiao Y, et al. Computational identification of protein-protein interactions in rice based on the predicted rice interactome network. *Genomics Proteomics Bioinformatics.* 2011;9:128–37.
33. Musungu B, Bhatnagar D, Brown RL, et al. A predicted protein interactome identifies conserved global networks and disease resistance subnetworks in maize. *Front Genet.* 2015;6:201.
34. Ding Z, Kihara D. Computational identification of protein-protein interactions in model plant proteomes. *Sci Rep.* 2019;9:8740.
35. Du X, Sun S, Hu C, et al. DeepPPI: boosting prediction of protein-protein interactions with deep neural networks. *J Chem Inf Model.* 2017;57:1499–510.
36. Li F, Zhu F, Ling X, et al. Protein interaction network reconstruction through ensemble deep learning with attention mechanism. *Front Bioeng Biotechnol.* 2020;8:390.
37. Assenov Y, Ramirez F, Schelhorn SE, Lengauer T, Albrecht M. Computing topological parameters of biological networks. *Bioinformatics.* 2008;24:282–4.
38. Gu H, Zhu P, Jiao Y, et al. PRIN: a predicted rice interactome network. *BMC Bioinformatics.* 2011;12:1–13.
39. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature.* 1998;393:440–2.
40. Wu H, Su Z, Mao F, et al. Prediction of functional modules based on comparative genome analysis and Gene Ontology application. *Nucleic Acids Res.* 2005;33:2822–37.
41. Stark C, Breitkreutz BJ, Reguly T, et al. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 2006;34:D535–9.
42. Orchard S, Kerrien S, Abbani S, et al. Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat Methods.* 2012;9:345–50.
43. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, et al. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol.* 2021;38:5825–9.
44. Chen C, Chen H, Zhang Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant.* 2020;13:1194–202.
45. Clevenger J, Chu Y, Scheffler B, et al. A developmental transcriptome map for allotetraploid *Arachis hypogaea*. *Front Plant Sci.* 2016;7:1446.
46. Kim D, Paggi JM, Park C, et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37:907–15.
47. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014;30:923–30.
48. Narayanan M, Vetta A, Schadt EE, et al. Simultaneous clustering of multiple gene expression and physical interaction datasets. *PLoS Comput Biol.* 2010;6:e1000742.
49. Du Z, Li L, Chen CF, et al. G-SESAME: web tools for GO-term-based gene similarity analysis and knowledge discovery. *Nucleic Acids Res.* 2009;37:W345–9.
50. Wang JZ, Du Z, Payattakool R, et al. A new method to measure the semantic similarity of GO terms. *Bioinformatics.* 2007;23:1274–81.
51. Wu X, Zhu L, Guo J, Zhang DY, Lin K. Prediction of yeast protein-protein interaction network: insights from the gene ontology and annotations. *Nucleic Acids Res.* 2006;34:2137–50.
52. Wu X, Zhu L, Guo J, Fu C, Zhou H, Dong D, Li Z, Zhang DY, Lin K. SPIDER: Saccharomyces protein-protein interaction database. *BMC Bioinformatics.* 2006;7(Suppl 5):S16.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.