Co-expression network analysis and identification of core genes in the interaction between wheat and Puccinia Striiformis f. sp. tritici

Yibo Wang
Yunnan Academy of Tobacco Agricultural Sciences, National Tobacco Genetic Engineering Research Centre
Ke Zhang
Yunnan Tobacco Quality Inspection & Supervision Station
Dan Chen
Yunnan Tobacco Quality Inspection & Supervision Station
Kai Liu
Yunnan Tobacco Quality Inspection & Supervision Station
Wei Chen
Yunnan Tobacco Quality Inspection & Supervision Station
Fei He
Chinese Academy of Sciences
Zhijun Tong
Yunnan Academy of Tobacco Agricultural Sciences, National Tobacco Genetic Engineering Research Centre
Qiaoling Luo
lql2014@genetics.ac.cn

Chinese Academy of Sciences

Research Article

Keywords: Wheat stripe rust, Transcriptome, Co-expression, WGCNA, Interaction gene

Posted Date: December 8th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3695555/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Archives of Microbiology on May 2nd, 2024. See the published version at https://doi.org/10.1007/s00203-024-03925-5.

Abstract

The epidemic of stripe rust, caused by the pathogen *Puccinia Striiformis* f. sp. *tritici* (*Pst*), would reduce wheat (*Triticum aestivum*) yields seriously. Traditional experimental methods are difficult to discover the interaction between wheat and *Pst*. Multi-omics data analysis provides a new idea for efficiently mining the interactions between host and pathogen. We used 140 wheat-*Pst* RNA-Seq data to screen for differentially expressed genes (DEGs) between disease-resistant and disease-susceptible samples, and carried out Gene Ontology (GO) enrichment analysis. Based on this, we constructed a gene co-expression network, identified the core genes and interacted gene pairs from the conservative modules. Finally, we checked the distribution of Nucleotide-binding and leucine-rich repeat (NLR) genes in the co-expression network and drew the wheat NLR gene co-expression network. In order to provide accessible information for related researchers, we built a web-based visualization platform to display the data. Based on the analysis, we found that various heat shock proteins (HSPs), protein kinases, and glycosidases frequently appeared in the network. They were likely to be involved in the biological processes of *Pst* infecting wheat. We also found that HSPs was significantly co-expressed in wheat and *Pst*, suggesting that there might be direct or indirect interactions between them. This study can assist scholars in conducting studies on the pathogenesis and help to advance the investigation of wheat-*Pst* interaction patterns.

Introduction

Wheat (*Triticum aestivum*) is one of the three major food crops in the world. As a main food source in Asia, Europe, and the Americas, it provides a large amount of carbohydrate for humans. In the face of global climate change, maintaining stable wheat yields is key to reducing world food risk [1]. Wheat stripe rust, also known as yellow rust (Yr), is an important, world-wide, disease dispersed by the wind. *Puccinia striiformis* f. sp. *tritici* (*Pst*), as a pathogenic fungus causing wheat stripe rust, mainly infects leaves, and even stalks, leaf sheaths and glumes in severe case. In China, wheat stripe rust is found throughout the main wheat-producing areas like northwest, southwest, north China and the middle and lower reaches of Yangtze River. It can generally cause yield losses of 5–10% in years of occurrence, and up to 30% during epidemic years [2].Historically, several epidemic of *Pst* have caused severe yield losses in wheat, and it is urgent to understand the mechanisms of wheat resistance to *Pst*.

Wheat stripe rust is fueled by low temperatures, heavy rain and exists mainly in spring, and autumn. The fastest sporulation rate of the *Pst* occurs when the temperature is between 15 and 25°C and it survives only for a short time when the temperature reaches 35°C or higher [3]. Air humidity and wind also affect the germination and spread of *Pst*. When humidity is high, dew will form on the plant surface and promote *Pst* to infest the plant; in windy environment, *Pst* can rely on wind for medium and long distance spread. Annual cycles of stripe rust includes four stages: oversummering, infection of seedlings in autumn, overwintering and spring epidemic. When the environment become disease-favorable, *Pst* keeps expanding and spreading during the cycle, leading to a large-scale epidemic. Consensus among

scientists on the control of wheat stripe rust is breeding disease-resistant wheat varieties, which is the most economical, environmentally and effective method.

On November 25, 2020, Nature published a research paper entitled "Multiple wheat genomes reveal global variation in modern breeding". The study selected 16 representative wheat varieties for whole genome sequencing, providing a comprehensive wheat genome map, allowing scientists and breeders to quickly identify genes related to wheat disease-resistance, thus providing great help in breeding disease-resistant wheat varieties [4]. The completion of wheat genome sequencing and the development of multi-omics technologies provide new ideas for wheat disease-resistance breeding. Furthermore, by constructing a framework for plant immunity, scholars can efficiently locate and clone wheat resistance genes. Subsequently, they can use multi-omics technologies to analyze the complete triggering process of plant immunity, integrate resistance genes, and build broad-spectrum resistance in wheat.

The basic framework of modelling plant immunity was described as early as 2006. The framework of the plant immune system, consisting of one layer of non-specific and one layer of specific immune response called PTI (PAMPs triggered immunity) and ETI (Effector triggered immunity), respectively, provides resistance to various pathogens. Plants trigger a non-specific defense response PTI after infected by pathogenic bacteria. Pathogen-associated molecular patterns (PAMPS) are recognized by pattern recognition receptors (PRRs) on the cell membrane surface and thus plant immunity is activated. Plant PRRS are generally receptor-like kinases, while pathogen PAMPs are mainly flagellin and titin shells, etc. The immune responses of PTI are the accumulation of plant callose, and burst of reactive oxygen species [5]. PTI is an important component of plant immune system, and such immune-initiated resistance is generally broad-spectrum, but weak and cannot efficiently suppress pathogen infestation. During the long period competing of plant and pathogen, pathogens evolves a class of virulence effector proteins inhibiting the plant PTI response effectively, and manage to infest the plant eventually. In turn, plants evolves a receptor protein, NLR (Nucleotide-binding domain and Leucine-rich Repeat), which specifically recognizes pathogenic effectors and triggers the ETI immune response. ETIs are generally race-specific and can be more effective against pathogenic bacteria, making them ideal targets for disease-resistance breeding in various crops [5]. In recent years, scholars found that PTI and ETI are not independent of each other. ETI depends to some extent on the PTI pathway, and the triggering of ETI can also enhance PTI, suggesting that plants can synergistically activate both immune responses to resist the invasion by pathogenic bacteria.

It is essential to map and clone stripe rust resistance genes in wheat which are involved in PTI and ETI. By transgenesis and gene editing, we can achieve significant disease control at the genetic level. Recently, scholars have cloned stripe rust resistance genes in wheat such as *Yr5/YrSp, Yr7, Yr10, Yr15, Yr27, Yr36, Yr18, Yr47, YrAS2388R, YrU1*, and *YrNAM* [6–15]. However, limited by the slow rate of gene cloning, some newly discovered resistance genes have long been used in breeding but cannot study their important role in disease resistance. Traditional experimental methods are difficult to tap the wheat-*Pst* interactions and inefficient for mining resistance genes among different varieties, and these problems have been plaguing breeders.

Transcriptome focuses on RNA modifications that affect gene expression at the transcriptional level. Transcriptome sequencing technology can comprehensively provide transcript information of a species-specific tissue or organ, which can be used for various gene expression level studies. In recent years, using transcriptome for gene spatiotemporal expression analysis, scholars identified and detected several significant functional genes in biosynthetic pathways, confirming the feasibility of transcriptome analysis for mining functional genes. Transcriptome analysis has been also successfully applied to plant disease resistance, for examples: downy mildew, powdery mildew, gray mold of grapes; blight, bacterial wilt, bacterial streak of rice; citrus yellow dragon disease; mango deformation and anthracnose [16–24], which makes it become an effective tool for studying the expression patterns and molecular mechanisms of plant disease-resistance genes.

With the development of transcriptome and sequencing technologies, gene co-expression network analysis becomes a common research tool. Based on it, scholars can correlate unknown functional genes with biological processes and to explore the core genes that play significant functions. Such network analysis was first successfully used in animal and medical fields and gradually extended to plant fields. Several genes were identified and reported by network analysis in various plants such as tomato, Arabidopsis, lily, and rice [25–28]. Weighted gene co-expression network analysis (WGCNA) is a common tool used by scholars to detect co-expressed genes. It can divide the network into modules and associate them with target traits to screen the core genes associated with the target traits, which could explain phenotypic variation at gene expression level [29]. Scholars can quickly call functions for customized co-expression network construction due to its availability in R. This study used the WGCNA tool to construct the network for subsequent analysis.

In this study, we constructed co-expression networks using published wheat-*Pst* transcriptome data. We classified gene modules and correlated them with infestation phenotypes, screened conserved modules related to phenotypes and extracted core genes from them for annotation. Finally we predicted multiple gene pairs with interaction relationships based on the connectivity of genes in the network.

Materials and methods

Acquisition of transcriptome raw data

The Rust expression browser is an open, expVIP-based constructed transcriptome database (www.rust-expression.com). It contains 1024 RNA-Seq datasets of *Pst* and wheat, which were infested one-to-one. All can be downloaded and used by scholars for subsequent analysis [30]. We downloaded this data and selected 140 high-quality transcriptomic samples containing the infestation phenotype as raw data for subsequent analysis. The data of 140 transcriptome samples containing infestation leaf phenotypes were downloaded from the Rust expression browser database. A total of 140 samples were collected from different wheat varieties. Each sample contained a wheat gene expression matrix and a *Pst* gene expression matrix with one-to-one infestation relationships, sampled from 2015 to 2017. Among the 140

samples, 43 were disease-susceptible and 97 were disease-resistant. Most of the samples were originated from Europe (Fig. 1A).

Differential expression analysis

Differential expression analysis refers to grouping samples from different treatments or phenotypes and statistically screening for genes with significantly different expression. These genes are considered to be most likely involved in biological processes regulating phenotype construction. In this study, samples were divided into disease-resistant and disease-susceptible groups based on infestation percentage data. We classified samples as susceptible if their infestation percentage was higher than 50%, and as resistant if it was lower. After grouping, differential expression analysis was performed using the R package Deseq2 to screen for genes that were differentially expressed in disease-resistant and disease-susceptible samples. The thresholds were set as p-value < 0.05 and |log2Foldchange|>0.5.

Gene Ontology enrichment analysis

Gene Ontology (GO) is an internationally standardized gene function classification system that provides a set of dynamically updated standard vocabularies to comprehensively describe the properties of genes in an organism. There are a total of three ontologies in GO, which describe molecular functions, cellular components, and biological processes. Evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG) is an online gene function annotation web tool (http://eggnog-mapper.embl.de/) that provides a free and fast gene function annotation service [31]. We annotated the differentially expressed genes using eggNOG-mapper with the default threshold settings. We used TBtools software [32] for GO enrichment of the annotated differentially expressed genes with p-value < 0.05.

Co-expression network construction

Differentially expressed genes (DEGs) were used as the raw data for constructing the co-expression network. The wheat-*Pst* co-expression network was constructed using the R package WGCNA 1.72 [33]. We checked the missing values in the filtered gene expression matrix for sample clustering and removed outlier samples. The soft threshold value was calculated using the "pickSoftThreshold" function, the optimal weight value was estimated using the "powerEstimate" function. The soft threshold value was selected according to the plotted change in the scale-free topology fitting index, and the mean connectivity of genes in the network under different soft thresholds was counted. Finally, the soft threshold value power beta = 7 was chosen, and other parameters are listed in Table 1. The network partitioning module was constructed by the "blockwiseModules" function. The modules with a similarity greater than 0.75 were merged.

Network module identification

After the network was constructed, we randomly splited the gene expression data into two groups and used the "modulePreservation" function to calculate their conservativeness: whether the two groups of data could be equally divided into the same gene modules as before. Z values of these gene modules were calculated, and we categorized them as follows: Z < 2 indicated no conservation, 2 < Z < 10

indicated mild to moderate conservation, and Z > 10 indicated high conservation. Afterwards, we performed a correlation analysis of these modules with the infection percentage data and calculated the correlation coefficients between each module eigenvalues and the phenotype. Those with higher absolute values of correlation coefficients were considered as modules that might contain host-pathogen gene interactions.

Annotation

All genes in the network were annotated using seven metrics: Log2foldchange (L2FC), Module, Z-score, Module-trait-relationship, GeneSignificant (GS), ModuleMembership (MM), and Protein anno. Among them, L2FC refers to the differential expression ploidy of the gene; Module refers to the module to which the gene belongs; Z-score refers to the conserved Z value of the module to which the gene belongs; Module-trait-relationship refers to the correlation between the module to which the gene belongs and the phenotype; GS refers to the correlation between the gene and the phenotype; MM refers to the correlation between the gene and the module; and Protein anno refers to the protein annotation of the gene.

Interacting gene pairs mining

The weight value of the gene pairs in the module was calculated, which can reflect the correlation between two genes: the closer to 1, the stronger the correlation. If the weight value was more than 0.1, the wheat-*Pst* gene pairs was considered had interacting relationships in the highly conserved module [34]. After screening the interacting gene pairs, we annotated their protein functions. The co-expression network was mapped by Gephi 0.10.1 [35].

NLR gene co-expression network construction

ANNA: an Angiosperm NLR Atlas. ANNA is an open database that collects annotated protein sequences of the largest class plant disease resistance gene family from angiosperm genomes. ANNA now contains over 90,000 NLR genes from 304 angiosperm genomes, including 18,707 TNL (TIR-NBS-LRR) genes, 70,737 CNL (CC-NBS-LRR) genes, and 1,847 RNL (RPW8-NBS-LRR) genes [36]. To verify the usability of the wheat-*Pst* co-expression network, we observed the distribution of 2298 wheat NLR genes identified by the ANNO database in this network. We screened the co-expressed genes in the network that were linked to these NLR genes, and setting the weight value > 0.1 to consider them to have an interaction with NLR genes. Finally, co-expression network of NLR genes in wheat were built and mapped by Gephi 0.10.1 [35].

Webpage building

Dynamic webpages were written to exhibit the results based on the shiny package, a web application framework in R that allows users to package the data analysis results into web apps. Considering that most of the algorithms in this study rely on R for programming and that Shiny has a complete data interface with R, we chose Shiny to visualize the webpages.

Results

Differential expression analysis of disease-resistant and disease-susceptible samples

We screened 8322 DEGs between the disease-resistant and disease-susceptible samples, including 7959 wheat genes and 363 *Pst* genes. The expression of 4361 wheat genes was significantly up-regulated, and 3598 wheat genes were significantly down-regulated in the disease-resistant samples, while 144 *Pst* genes were significantly up-regulated and 219 *Pst* genes were significantly down-regulated in the disease-resistant samples (Fig. 1B). Observing the distribution of wheat genes on chromosomes, we found that they were somewhat enriched at the terminal regions of chromosomes, such as chromosomes 3A and 3D (Fig. 1C). We speculated that the differential expression of these genes could explain the differences in the infestation phenotypes between the two sample groups and could be applied to the subsequent construction of co-expression networks. GO enrichment analysis showed that there was none significant GO term enriched in *Pst* due to the limited number of DEGs. So here just described the GO enrichment results of DEGs in wheat. DEGs were enriched in 401 pathways, including 107 molecular functional pathways, 10 cellular component pathways, and 284 biological process pathways. Sorting the number of genes enriched in these pathways revealed that more genes were enriched in the pathways of biological process, such as response to stimulus, response to stress, and even response to fungus and defense response to fungus, which were relevant to this study. In terms of cellular components, genes were mainly enriched in plastid stroma, cell-cell junction, obsolete cell wall part, etc. In terms of molecular functions, more genes were enriched in transcriptional regulatory activity, DNA-binding transcription factor activity, and other transcription-related molecular functional pathways. The above results showed that some of the DEGs might participate hormone metabolism, transcriptional regulation, and signaling processes related to plant immunity. The result further validated their feasibility for performing co-expression network construction. Some of the pathways related to plant immune regulation were listed in Fig. 1D.

Wheat-Pst co-expression network construction and module analysis

We clustered the samples and obtained 100 groups of available samples after eliminating the outliers. The DEGs of wheat and *Pst* of these 100 groups of samples were used as input data to construct the wheat-*Pst* co-expression network. The scale-free topology fitting index and the network mean connectivity were calculated to determine the optimal soft threshold value of the network. It was found that the scale-free topology fit index tended to stabilize after $\beta = 7$ and $R2 > 0.80$, while the network connectivity started to be smooth at the same time. It can be considered that the network has good connectivity and conforms to the scale-free network distribution (Fig. 2A). We set $\beta = 7$ as the soft threshold to construct the network. The genes were clustered into modules based on gene co-expression relationship. As shown in Fig. 2B, each color represented a module and the gray module consisted of genes that couldn't be classified into any module. After constructing the network, we obtained 28 gene

co-expression modules. The number of genes in each module was counted, and it was found that the turquoise module contained the largest number of genes with 3182, the white module contained the least number of genes with 32. A number of 588 genes were classified into the gray module, and the number of module genes is shown in Fig. 2C. To determine whether these modules are conserved and biologically significant, we calculated the Z-values of each module and the correlation coefficient between modules and phenotypes. It was found that a total of 16 modules with Z-values > 10 showed high conservation, 11 modules with Z-values > 2 and < 10 had moderate conservation and one module with Z-value < 2 did not have conservation (Fig. 2D). The association analysis showed that the gene expression in most modules was negatively correlated with the phenotype, and only 7 modules and the grey module exhibited a positive correlation (Fig. 2E). These 7 modules had a higher percentage of *Pst* genes compared with other modules, and the absolute values of the correlation coefficient were higher, which meant the modules might contain more wheat-*Pst* interactions. Further, it was shown that most of these positively correlated modules had Z-values greater than 10, and all of them were greater than 2. We believed that the genes within these modules have high confidence of co-expression. Interestingly, the correlation coefficient between the grey module and the phonotypes came to a maximum value of 0.46. So we should pay attention to the individual genes in the grey module though they could not be classified into any module and had no co-expression.

Annotation of core genes

To statistically describe the core genes and interactions in 28 conserved modules, we annotated all genes in the network with seven indicators: L2FG, module, Z-score, module-trait-relationship, GS, MM, protein anno. The genes with MM > 0.8 and |GS| > 0.2 were selected as core genes, and 390 core genes were obtained, including 351 wheat genes (mainly from the turquoise module) and 39 *Pst* genes (mainly from the cyan module). In addition, some of the non-core genes in the module encoded protein domains associated with plant immunity. It suggested that the selection of threshold values in the core gene screening needs to be improved, and we should not only focus on core genes in the subsequent analysis. Some of the core and non-core genes encoded plant immune-related structural domains were listed in Table 2.

Exploration of wheat-Pst interaction relationships

A total of 919,606 pairs of interaction relationships were obtained from 28 conserved modules. These included 3493 pairs of *Pst-Pst* interacting gene pairs, 902,034 pairs of wheat-wheat interacting gene pairs, and 14,078 pairs of wheat-*Pst* interacting gene pairs. To explore the interactions between wheat and *Pst*s at the gene expression level, we screened 4429 wheat-*Pst* interactions with a weight value greater than 0.2, and of which a total of 932 gene pairs could be fully annotated. Protein annotation of these pairs showed that HSP17.A, HSP70, HSP82, CLPB1, IP5P11, and B120 proteins were more frequent in wheat, and HSP16, HSP78, HSP104, CEX, GSK3, and HSS1 proteins more frequent in *Pst*

(Fig. 3), and most of them play important functions in plant immunity. We briefly introduced their biological functions below:

1. Heat shock proteins, such as HSP, CLPB, and DNAJ proteins, appeared many times in the wheat-*Pst* interactions. HSP proteins are protective proteins that are commonly found in various organisms. It plays an important role in plant growth and the regulation of stress by binding to other heat shock proteins to maintain proteostasis and repair degenerated proteins. HSP proteins are divided into five families: HSP100, HSP90, HSP70, HSP60, and small molecule HSP, and each of them contains a variety of proteins [37];
2. Protein kinases can receive a wide variety of ligands to help plants sense their cellular environment. It regulates plant growth, development, response to adversity and trigger immunity [38]. They repeatedly appeared in the wheat-*Pst* interaction gene pairs. In this study, B120, At2g42960, HK6, MIK2, CRK6, CRK7, RGA3, XA21, and other protein kinases were annotated among the interaction genes. Most of these proteins were verified to be involved in responding to plant stress and have the classical domains of plant immune receptor-like protein kinases. Among them, CRK6 protein conferred broad-spectrum resistance to wheat leaf blight [39]; RGA3 conferred resistance to rice blast [40]; and rice XA21 conferred resistance to bacterial infestation [41];
3. Various glycosidases, such as CEX, XGEA, XLNA, and other xyloglucanases, repeatedly appeared in wheat-*Pst* interaction gene pairs. Such proteins are thought to be key enzymes in the catabolic remodeling of plant cell wall. It plays a role in *Pst* infestation of wheat: inducing disease resistance defenses in plants [42].

The frequency of HSP-HSP interaction gene pairs was high. It was found that the heat stress proteins from wheat and *Pst* had a significant co-expression relationship, with several interaction gene pairs having a weight value of 0.5. Previously, it was found that virus infestation could up-regulate the expression level of HSP protein in plant hosts [43]; While virus infestation also promoted the expression of the virus's own HSP, with a clear co-expression phenomenon between them [43, 44]. Our study indicated that HSPs co-expression also exists in fungal-plant interactions, suggesting that there may be direct or indirect interactions between different HSPs from wheat and *Pst*. These pairs of interactions also included effector and receptor protein kinases (who may trigger downstream immune regulation), which had some value in validating protein interactions. Finally, we mapped the wheat-*Pst* interaction network composed of these 4429 interacting gene pairs (Fig. 4A).

Interaction of NLR genes in the network

NLR genes represent one of the largest plant disease resistance gene families. NLR immune receptors can rapidly recognize pathogens and trigger ETI responses to ensure normal plant growth. Here we collected 2298 wheat NLR genes identified from the ANNO database. We found that 107 of those NLR genes were differentially expressed in wheat stripe rust resistant and susceptible samples, with most of them (77/107) distributed in the turquoise module, and the remaining in other modules in small

numbers. The threshold value of weight > 0.1 was set to screen NLR interaction gene pairs. A total of 42,227 interaction gene pairs were obtained, including 42,101 wheat-wheat pairs and 126 wheat-*Pst* pairs. Oxidase-NLR gene pair and glycosidase-NLR gene pairs were repeatedly found in the interaction gene pairs. Table 3 showed some of the NLR interaction gene pairs, and Fig. 4B showed the NLR gene co-expression network.

A database of wheat- Pst gene co-expression network

To enhance the significance of this study, we proposed to use the network as a database of interactions between wheat and *Pst*. Using the web we can infer the potential interactions of various cloned or putative resistance genes by observing the distribution of their connections in the network. We built a visualization tool which could combine all annotated gene results and network interaction data from our analysis. The link to the webpage is: http://8.130.121.203:3838/WRCA/. Here we briefly described several functions of the webpage:

(1) Searching for gene annotation information on the web

The webpage provides annotation information of the genes involved in the study, including genes Gene id, L2FG, Module, Z-score, Module-trait-relationship, GS, MM, and Protein anno. Users can search for relevant genes by entering a gene list or module, and the search results will be displayed in a table and allow users to download and save.

(2) Searching for interactions in the network

The webpage provides all the co-expressed gene pairs in the network, users can search for the pairs by entering the gene list or annotation information. The search results will be displayed in a table and allow users to download and save.

(3) Selecting genes or gene pairs to generate network diagrams

After searching and selecting genes or gene pairs, the webpage allows the search results to be used as raw input data to generate network diagrams. The generated network diagrams can be downloaded and saved by users.

(4) Downloading the results of this study

The webpage provides a link to download the complete data involved in this study, which users can use for their research.

Discussion

Feasibility and limitations of WGCNA analysis

The massive transcriptome data samples provide the data base for various types of network analysis. Gene co-expression networks, which can link genes of unknown function to biological processes of interest, have become an inherent model for the study of gene interactions. WGCNA, as the first tool for co-expression analysis, has effectively identified the relationship between multiple phenotypes and gene modules. For example, co-expression analysis of RNA-Seq data from lean and obese pigs identified obesity-associated modules and identified CCR1, MSR1, and SPI1 as regulators of this process [45]; co-expression analysis identified gene modules at different developmental stages in humans and mice, etc. [46]. WGCNA was well known by the great success in animal and medical research fields, and it have been extended to the plant fields. Phenotype-associated gene modules have been identified and reported in tomato, arabidopsis, lily, and rice [25–28]. In this study, we innovatively considered wheat-*Pst* as an interaction system and combined their RNA-Seq data as input data to construct a co-expression network. Consequently, we identified several gene co-expression modules related to the infestation phenotype and screened several core genes and wheat-*Pst* interaction gene pairs, which could provide a data base for analyzing the immunity mechanism of wheat against stripe rust. Although WGCNA has shown superiority in co-expression analysis, it is worth noting that the behavior of attaching biological significance to modules sometimes leads to erroneous conclusions. When the modules as a whole has correlation with the phenotype, it does not mean that every gene in the module is necessarily associated with the phenotype. Actually, the proportion of genes associated with the phenotype in the module is usually below 20%. This reminds us that even if we identify phenotype-associated modules with high confidence, we still need to perform deeper screening of genes within the modules based on functional annotation, connectivity, and other indicators.

Exploration of wheat-Pst interactions

Researches on wheat-*Pst* interactions and the pathogenesis have been established. Tang et al. (2022) identified *Pst*_A23, an effector located in the host nucleus, which promote *Pst* pathogenesis by regulating alternative splicing in wheat [47]; Wei et al. (2023) identified Hasp98 as a *Pst*-specific effector, which promoted *Pst* infection by interfering with the MAPK signaling pathway in wheat [48]. In this study, we constructed a co-expression network using wheat-*Pst* RNA-Seq data and mined 4429 pairs of wheat-*Pst* interaction gene pairs. We innovatively found that HSP respectively from wheat and *Pst* have significant co-expression relationships, and they may have direct or indirect interactions. This study can assist scholars in conducting studies on the pathogenesis and help to advance the investigation of wheat-*Pst* interaction patterns.

Conclusion

In this study, 140 RNA-Seq data of wheat infected by *Pst* near the Mediterranean Sea in northern Europe were used for differential gene expression analysis. A total of 8322 DEGs were obtained to construct a co-expression network. And we got 28 gene modules from the co-expression network, which included 16 high conserved modules and 11 moderate conserved modules. Importantly, seven modules were associated with the infestation phenotype. We also identified 919,606 pairs of interactions from the

network. Among them, we found obvious co-expression in wheat-*Pst* interactions and drew a wheat-*Pst* interactions network. Further, we revealed that there was a significant co-expression of heat stress proteins respectively from wheat and *Pst*, which is similar with the previous findings in plant-virus interactions. Moreover, we identified the distribution of NLR genes in the network, screened the interaction gene pairs with co-expression relationships, and mapped the NLR gene co-expression network. Finally, we built a webpage by pooling the data for the exhibition of our results.

Abbreviations

Pst	*Puccinia Striiformis* f. sp. *tritici*
DEGs	Differentially expressed genes
GO	Gene ontology
NLR	Nucleotide-binding and leucine-rich repeat
HSPs	Heat shock proteins
Yr	Yellow rust
PAMPS	Pathogen-associated molecular patterns
PTI	PAMPs triggered immunity
ETI	Effector triggered immunity
PRRs	pattern recognition receptors
WGCNA	Weighted gene co-expression network analysis
L2FC	Log2foldchange
GS	GeneSignificant
MM	ModuleMembership

Declarations

Author Contributions: Y.W., Q.L. and K.Z. conceived and designed the experiments. Y.W., D.C., K.L. analyzed the RNA-Seq data. Y.W., W.C. built the webpage. Y.W. wrote the manuscript. Q.L., F.H. and Z.T. supervised the research and edited the manuscript. All authors have read and agreed to the submitted version of the manuscript.

Data availability: All data generated or analyzed during this study are included in the published article.

Funding: This project was supported by the Research Program for Network Security and Information of the Chinese Academy of Sciences (CAS-WX2021SF-0109) and Research on Key Technologies for Identification and Inspection of Tobacco Raw Materials and Cigarette-like Products Based on DNA Molecular Markers of Yunnan Branch of China National Tobacco Corporation (2023530000241027).

Conflicts of Interest: The authors declare that they have no conflict of interest.

Ethical approval: The current study does not involve any animal or human data.

Informed consent: Not applicable for current study.

References

1. Si Z, Qin A, Liang Y et al (2023) A review on regulation of irrigation management on wheat physiology, grain yield, and quality. Plants 12:692. https://doi.org/10.3390/plants12040692
2. Chen WQ, Wu LR, Liu TG et al (2009) Race dynamics, diversity, and virulence evolution in *Puccinia Striiformis* f. sp. *tritici*, the causal agent of wheat stripe rust in china from 2003 to 2007. Plant Dis 93:1093–1101. https://doi.org/10.1094/PDIS-93-11-1093
3. Chen XM (2005) Epidemiology and control of stripe rust [*Puccinia Striiformis* f. sp. *tritici*] on wheat. Can J Plant Pathol 27:314–337. https://doi.org/10.1080/07060660509507230
4. Walkowiak S, Gao L, Monat C et al (2020) Multiple wheat genomes reveal global variation in modern breeding. Nature 588:277–283. https://doi.org/10.1038/s41586-020-2961-x
5. Deng Y, Liu Y, Wang J et al (2021) Strategic thinking and research on crop disease and pestresistance in China. Sci Sin Vitae 51:1435–1446. https://doi.org/10.1360/SSV-2021-0199
6. Marchal C, Zhang J, Zhang P et al (2018) BED-domain-containing immune receptors confer diverse resistance spectra to yellow rust. Nat Plants 4:662–668. https://doi.org/10.1038/s41477-018-0236-4
7. Liu W, Frick M, Huel R et al (2014) The stripe rust resistance gene *Yr10* encodes an evolutionary-conserved and unique cc–nbs–lrr sequence in wheat. Mol Plant 7:1740–1755. https://doi.org/10.1093/mp/ssu112
8. Klymiuk V, Yaniv E, Huang L et al (2018) Cloning of the wheat *Yr15* resistance gene sheds light on the plant tandem kinase-pseudokinase family. Nat Commun 9:3735. https://doi.org/10.1038/s41467-018-06138-9
9. Athiyannan N, Abrouk M, Boshoff WHP et al (2022) Long-read genome sequencing of bread wheat facilitates disease resistance gene cloning. Nat Genet 54:227–231. https://doi.org/10.1038/s41588-022-01022-1
10. Fu D, Uauy C, Distelfeld A et al (2009) A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. Science 323:1357–1360. https://doi.org/10.1126/science.1166289
11. Krattinger SG, Lagudah ES, Spielmeyer W et al (2009) A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. Science 323:1360–1363.

https://doi.org/10.1126/science.1166453

12. Wang H, Zou S, Li Y et al (2020) An ankyrin-repeat and WRKY-domain-containing immune receptor confers stripe rust resistance in wheat. Nat Commun 11:1353. https://doi.org/10.1038/s41467-020-15139-6
13. Moore JW, Herrera-Foessel S, Lan C et al (2015) A recently evolved hexose transporter variant confers resistance to multiple pathogens in wheat. Nat Genet 47:1494–1498. https://doi.org/10.1038/ng.3439
14. Zhang C, Huang L, Zhang H et al (2019) An ancestral NB-LRR with duplicated 3'UTRs confers stripe rust resistance in wheat and barley. Nat Commun 10:4023. https://doi.org/10.1038/s41467-019-11872-9
15. Ni F, Zheng Y, Liu X et al (2023) Sequencing trait-associated mutations to clone wheat rust-resistance gene *YrNAM*. Nat Commun 14:4353. https://doi.org/10.1038/s41467-023-39993-2
16. Kang J, Gong P, Ge M et al (2021) The PLCP gene family of grapevine (*Vitis vinifera* L.): characterization and differential expression in response to *Plasmopara Viticola*. BMC Plant Biol 21:499. https://doi.org/10.1186/s12870-021-03279-w
17. Jiang C, Wang D, Zhang J et al (2021) *VqMYB154* promotes polygene expression and enhances resistance to pathogens in Chinese wild grapevine. Hortic Res 8:151. https://doi.org/10.1038/s41438-021-00585-0
18. Gao P, Zhang H, Yan H et al (2021) *RcTGA1* and glucosinolate biosynthesis pathway involvement in the defence of rose against the necrotrophic fungus *Botrytis cinerea*. BMC Plant Biol 21:223. https://doi.org/10.1186/s12870-021-02973-z
19. Xia Y, Fei B, He J et al (2017) Transcriptome analysis reveals the host selection fitness mechanisms of the *Rhizoctonia solani* AG1IA pathogen. Sci Rep 7:10120. https://doi.org/10.1038/s41598-017-10804-1
20. Wang C, Tariq R, Ji Z et al (2019) Transcriptome analysis of a rice cultivar reveals the differentially expressed genes in response to wild and mutant strains of *Xanthomonas oryzae* pv. *oryzae*. Sci Rep 9:3757. https://doi.org/10.1038/s41598-019-39928-2
21. Xie X, Chen Z, Zhang B et al (2020) Transcriptome analysis of *xa5*-mediated resistance to bacterial leaf streak in rice (*Oryza sativa* L.). Sci Rep 10:19439. https://doi.org/10.1038/s41598-020-74515-w
22. Curtolo M, de Souza Pacheco I, Boava LP et al (2020) Wide-ranging transcriptomic analysis of *Poncirus trifoliata*, *Citrus sunki*, *Citrus sinensis* and contrasting hybrids reveals HLB tolerance mechanisms. Sci Rep 10:20865. https://doi.org/10.1038/s41598-020-77840-2
23. Liu F, Wu J, Zhan R, Ou X (2016) Transcription profiling analysis of mango–*Fusarium mangiferae* interaction. Front Microbiol 7:. https://doi.org/10.3389/fmicb.2016.01443
24. Hong K, Gong D, Zhang L et al (2016) Transcriptome characterization and expression profiles of the related defense genes in postharvest mango fruit against *Colletotrichum gloeosporioides*. Gene 576:275–283. https://doi.org/10.1016/j.gene.2015.10.041

25. DiLeo MV, Strahan GD, den Bakker M, Hoekenga OA (2011) Weighted correlation network analysis (WGCNA) applied to the tomato fruit metabolome. PLoS ONE 6:e26683. https://doi.org/10.1371/journal.pone.0026683
26. Li F, Hu Q, Chen F, Jiang JF (2021) Transcriptome analysis reveals Vernalization is independent of cold acclimation in *Arabidopsis*. BMC Genom 22:462. https://doi.org/10.1186/s12864-021-07763-3
27. Shi S, Duan G, Li D et al (2018) Two-dimensional analysis provides molecular insight into flower scent of *Lilium* "Siberia." Sci Rep 8:5352. https://doi.org/10.1038/s41598-018-23588-9
28. Zeng Z, Zhang S, Li W et al (2022) Gene-coexpression network analysis identifies specific modules and hub genes related to cold stress in rice. BMC Genom 23:251. https://doi.org/10.1186/s12864-022-08438-3
29. van Dam S, Võsa U, van der Graaf A et al (2017) Gene co-expression analysis for functional classification and gene–disease predictions. Brief Bioinform bbw139. https://doi.org/10.1093/bib/bbw139
30. Adams TM, Olsson TSG, Ramírez-González RH et al (2021) Rust expression browser: an open source database for simultaneous analysis of host and pathogen gene expression profiles with expVIP. BMC Genom 22:166. https://doi.org/10.1186/s12864-021-07488-3
31. Cantalapiedra CP, Hernández-Plaza A, Letunic I et al (2021) EggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol 38:5825–5829. https://doi.org/10.1093/molbev/msab293
32. Chen C, Chen H, Zhang Y et al (2020) TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant 13:1194–1202. https://doi.org/10.1016/j.molp.2020.06.009
33. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinform 9:559. https://doi.org/10.1186/1471-2105-9-559
34. Zhang H, Fu Y, Guo H et al (2019) Transcriptome and proteome-based network analysis reveals a model of gene activation in wheat resistance to stripe rust. Int J Mol Sci 20:1106. https://doi.org/10.3390/ijms20051106
35. Bastian M, Heymann S, Jacomy M (2009) Gephi: an open source software for exploring and manipulating networks. ICWSM 3:361–362. https://doi.org/10.1609/icwsm.v3i1.13937
36. Liu Y, Zeng Z, Zhang Y-M et al (2021) An angiosperm NLR Atlas reveals that NLR gene reduction is associated with ecological specialization and signal transduction component deletion. Mol Plant 14:2015–2031. https://doi.org/10.1016/j.molp.2021.08.001
37. Timperio AM, Egidi MG, Zolla L (2008) Proteomics applied on plant abiotic stresses: role of heat shock proteins (HSP). J Proteomics 71:391–411. https://doi.org/10.1016/j.jprot.2008.07.005
38. Dievart A, Gottin C, Périn C et al (2020) Origin and diversity of plant receptor-like kinases. Annu Rev Plant Biol 71:131–156. https://doi.org/10.1146/annurev-arplant-073019-025927
39. Saintenac C, Cambon F, Aouini L et al (2021) A wheat cysteine-rich receptor-like kinase confers broad-spectrum resistance against *Septoria tritici* blotch. Nat Commun 12:433.

https://doi.org/10.1038/s41467-020-20685-0

40. Xu X, Hayashi N, Wang C-T et al (2014) Rice blast resistance gene Pikahei-1(t), a member of a resistance gene cluster on chromosome 4, encodes a nucleotide-binding site and leucine-rich repeat protein. Mol Breeding 34:691–700. https://doi.org/10.1007/s11032-014-0067-6
41. Ercoli MF, Luu DD, Rim EY et al (2022) Plant immunity: Rice XA21-mediated resistance to bacterial infection. Proc Natl Acad Sci USA 119:e2121568119. https://doi.org/10.1073/pnas.2121568119
42. Beliën T, Van Campenhout S, Van Acker M, Volckaert G (2005) Cloning and characterization of two endoxylanases from the cereal phytopathogen *Fusarium graminearum* and their inhibition profile against endoxylanase inhibitors from wheat. Biochem Biophys Res Commun 327:407–414. https://doi.org/10.1016/j.bbrc.2004.12.036
43. Alam SB, Rochon D (2016) Cucumber necrosis virus recruits cellular heat shock protein 70 homologs at several stages of infection. J Virol 90:3302–3317. https://doi.org/10.1128/JVI.02833-15
44. Medina V, Peremyslov VV, Hagiwara Y, Dolja VV (1999) Subcellular localization of the HSP70-homolog encoded by beet yellows closterovirus. Virology 260:173–181. https://doi.org/10.1006/viro.1999.9807
45. Kogelman LJA, Cirera S, Zhernakova DV et al (2014) Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA Sequencing in a porcine model. BMC Medical Genom 7:57. https://doi.org/10.1186/1755-8794-7-57
46. Xue Z, Huang K, Cai C et al (2013) Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. Nature 500:593–597. https://doi.org/10.1038/nature12364
47. Tang C, Xu Q, Zhao J et al (2022) A rust fungus effector directly binds plant pre-mRNA splice site to reprogram alternative splicing and suppress host immunity. Plant Biotechnol J 20:1167–1181. https://doi.org/10.1111/pbi.13800
48. Wei J, Wang X, Hu Z et al (2023) The *Puccinia Striiformis* effector Hasp98 facilitates pathogenicity by blocking the kinase activity of wheat TaMAPK4. J Integr Plant Biol 65:249–264. https://doi.org/10.1111/jipb.13374

Tables

Tables 1 to 3 are available in the Supplementary Files section.

Figures

Fig. 1

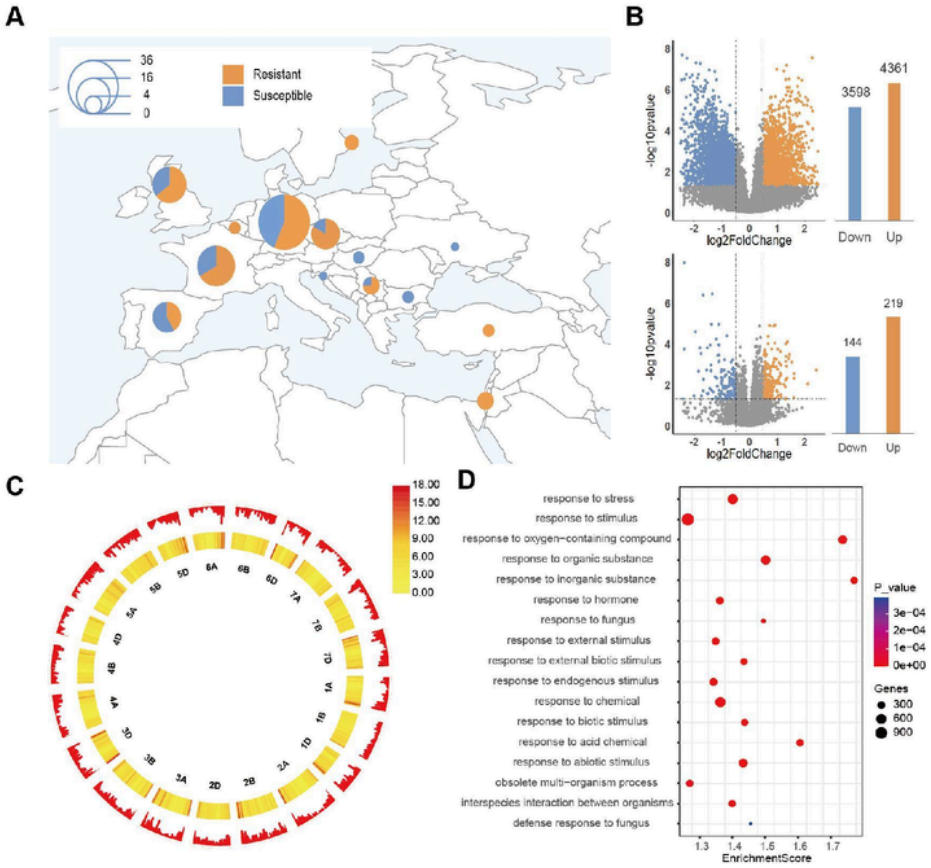

Figure 1

A Sample source, samples mainly originated from Europe. B Volcano map of differentially expressed genes in wheat in the upper panel and differentially expressed genes in *Pst* in the lower panel. C Distribution of differentially expressed genes in wheat on the chromosomes. D Some of the GO terms related to plant immunity

Fig. 2

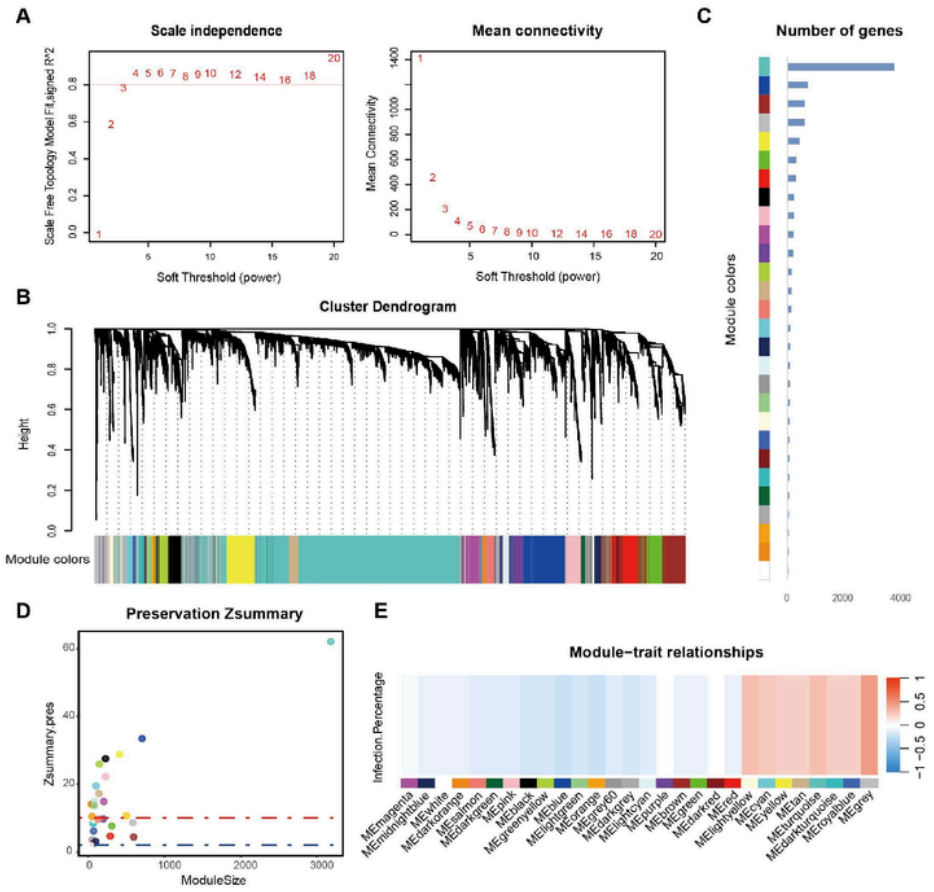

Figure 2

A Scale-free topology fit index and network connectivity. B gene module cluster dendrogram, which included 28 gene co-expression modules with different color. C the number of genes contained in each module. D Z-value of each module. E correlation of modules with infestation percentage

Fig. 3

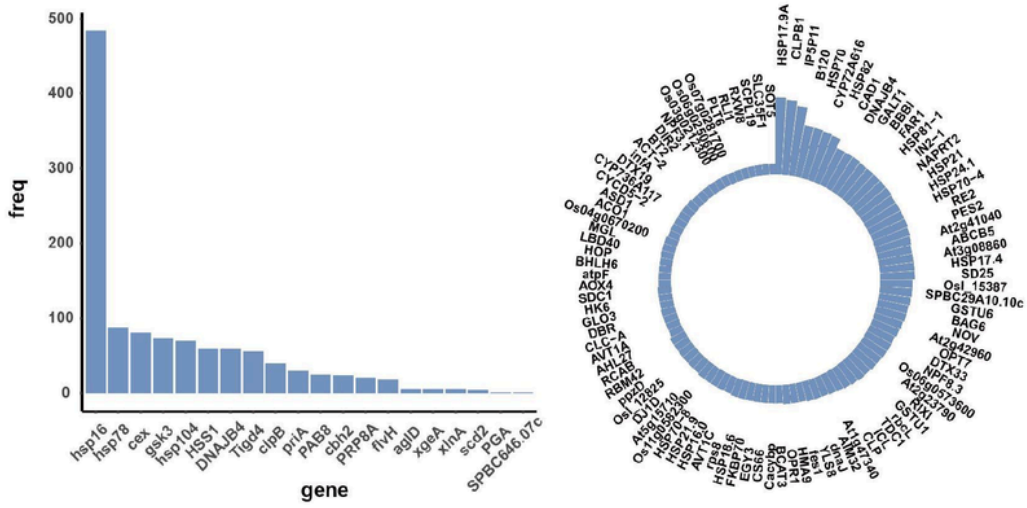

Figure 3

The frequency of the wheat-striped rust interaction gene pair, the left panel is the strip rust genes, and the right panel is the wheat genes (based on protein annotation)

Fig. 4

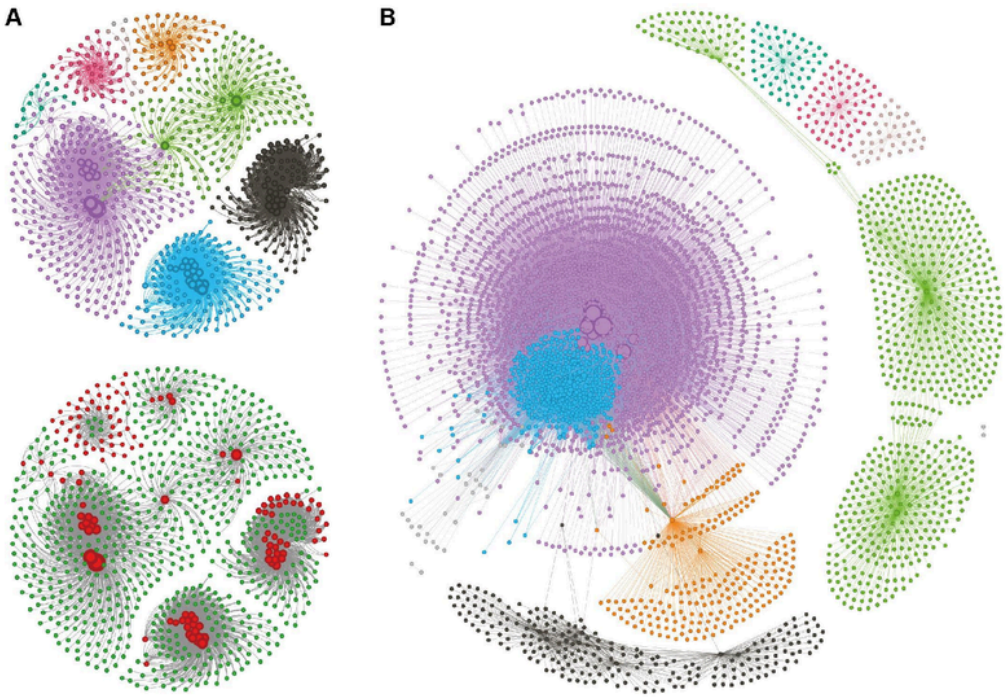

Figure 4

A Wheat-*Pst* interaction network diagram, clustered by module, with red nodes for *Pst*genes and green nodes for wheat genes in the panel below. B NLR gene co-expression network

Fig. 5

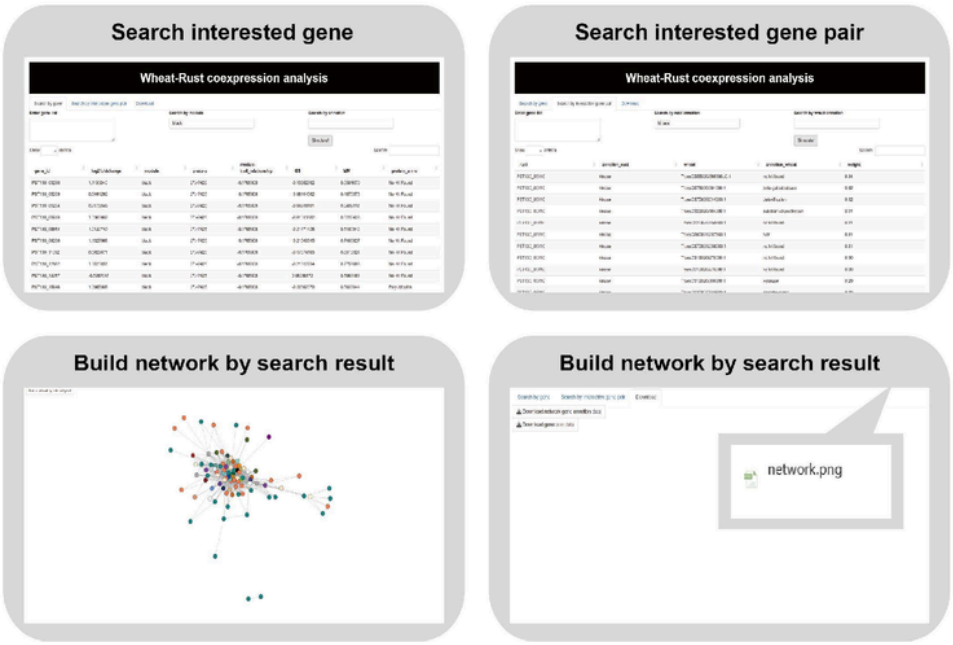

Figure 5

Web Features & Screenshots

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Tables.pdf
- SupplementaryFiles.zip