

SCIENTIFIC REPORTS



OPEN

Characterizing co-expression networks underpinning maize stalk rot virulence in *Fusarium verticillioides* through computational subnetwork module analyses

Man S. Kim^{1,2}, Huan Zhang¹, Huijuan Yan¹, Byung-Jun Yoon² & Won Bo Shim¹

Fusarium verticillioides is recognized as an important stalk rot pathogen of maize worldwide, but our knowledge of genetic mechanisms underpinning this pathosystem is limited. Previously, we identified a striatin-like protein Fsr1 that plays an important role in stalk rot. To further characterize transcriptome networks downstream of Fsr1, we performed next-generation sequencing (NGS) to investigate relative read abundance and also to infer co-expression networks utilizing the preprocessed expression data through partial correlation. We used a probabilistic pathway activity inference strategy to identify functional subnetwork modules likely involved in virulence. Each subnetwork modules consisted of multiple correlated genes with coordinated expression patterns, but the collective activation levels were significantly different in *F. verticillioides* wild type versus *fsr1* mutant. We also identified putative hub genes from predicted subnetworks for functional validation and network robustness studies through mutagenesis, virulence and qPCR assays. Our results suggest that these genes are important virulence genes that regulate the expression of closely correlated genes, demonstrating that these are important hubs of their respective subnetworks. Lastly, we used key *F. verticillioides* virulence genes to computationally predict a subnetwork of maize genes that potentially respond to fungal genes by applying cointegration-correlation-expression strategy.

Maize stalk rot is a complex disease, primarily caused by a series of fungal pathogens. Charcoal rot (by *Macrophomina phaseolina*), Fusarium stalk rot (by *Fusarium verticillioides*), Gibberella stalk rot (by *F. graminearum*) and Anthracnose stalk rot (by *Colletotrichum graminicola*) are the major stalk rots that devastate maize-growing regions in the US^{1,2}. Losses due to stalk rot come in several different forms including stalk breakage, lodging, premature death of the plant, and the interruption of normal grain filling process. Pathogens typically overwinter in crop residue from the previous year and produce spores in the next growing season that will serve as the primary inoculum source. It is generally perceived that when crops experience abiotic stress, particularly at the end of the growing season, pathogens take advantage and colonize vulnerable stalk tissues²⁻⁴. But overall, we still lack a clear understanding of how these stalk rot fungi colonize and progress through pathogenesis.

To better understand the mechanism of Fusarium stalk pathogenesis, we screened for loss-of-virulence *F. verticillioides* mutants and identified a gene, *FSR1*, that is responsible for the deficiency in stalk rot virulence⁵. Microscopic examination of inoculated stalks revealed the wild-type fungus vigorously colonizing vascular bundles and causing rot, whereas the mutant showed limited colonization and rot in stalks. *FSR1* encodes a protein that shares high similarity with striatins, a group of proteins found in eukaryotes that form complexes with

¹Department of Plant Pathology and Microbiology, Texas A&M University, College Station, TX, 77843, United States.

²Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, 77843, United States. Man S. Kim and Huan Zhang contributed equally to this work. Correspondence and requests for materials should be addressed to W.B.S. (email: wbskim@tamu.edu)

kinases and phosphatases to regulate diverse cellular functions^{6–8}. Recent studies have demonstrated important cellular and physiological roles of striatin protein and striatin-interacting phosphatase and kinase (STRIPAK) complex in *Sordaria macrospora*, *Neurospora crassa*, *Aspergillus nidulans*, and *C. graminicola*^{9–12}. In ascomycete fungus *S. macrospora*, Pro11, Pro22, Pro45, SmMob3, and PP2Ac1, key STRIPAK complex components, were shown to play critical roles in fruiting body formation, vegetative growth, and hyphal fusion^{13,14}. In *N. crassa*, STRIPAK components are central for regulating important MAP kinase signaling pathways that regulate cell wall integrity and fruiting body development⁹. But more recently, investigation of striatin function in maize pathogens *C. graminicola*¹² and *F. verticillioides*^{15,16} supports our hypothesis that Fsr1/striatin-mediated signal transduction plays a critical role in regulating stalk rot pathogenesis.

One of the intriguing questions we are aiming to answer is the impact of Fsr1 in cellular signaling associated with *F. verticillioides* virulence. Biological functions, including virulence, are executed through elaborate collaboration of various biomolecules, and there has been increasing interest in the computational identification of functional modules from large-scale experimental data. To unravel the complex web of genetic interactions in *F. verticillioides* and maize, we decided to take advantage of next-generation sequencing (NGS) and explore the transcriptomic subnetwork modules underpinning *FSR1*-mediated fungal virulence by computational network-based analysis. Our goal was to develop probabilistic and systematic models to investigate the interrelationship between genes rather than relying on quantitative comparison of transcript abundance as a measure of significance. Our NGS study was designed to capture dynamic changes in gene expression during maize stalk colonization by *F. verticillioides* wild type and *fsr1* mutant. To capture dynamic changes in transcriptome, samples were harvested from three distinct phases of stalk pathogenesis: i) establishment of fungal infection, ii) colonization and movement into the vascular bundles, and iii) host destruction and collapse¹⁷. A total of six independent biological replications were prepared and analyzed for each sample, since increasing the number of replicates was important for us to implement our computational analysis for identifying subnetwork modules that show strong differential expression.

As described in our previous work¹⁷, our strategy is to first construct the co-expression network of *F. verticillioides* using partial correlation, and search through these networks to detect subnetwork modules that are differentially expressed in the two *F. verticillioides* strains. Subsequently, we use the probabilistic pathway activity inference scheme¹⁸ to predict the activity level of potential subnetworks, followed by applying a computationally efficient branch-out technique to find the subnetworks that display the largest differential expression. Each subnetwork consisted of multiple genes with coordinated expression patterns, but more importantly we targeted subnetworks whose collective activation level is significantly different in the wild type versus the mutant. We then applied a series of mathematical criteria to predict the hub gene in each network and functionally tested their role in *F. verticillioides* virulence and the maintenance of network robustness.

Results

NGS data preparation and relative expression analysis. We performed NGS using Illumina HiSeq 2000 and generated 36 independent libraries (*i.e.*, six libraries per each time point - 3 dpi [infection], 6 dpi [colonization], and 9 dpi [rot] - for wild type and *fsr1* mutant). For analysis and prediction in this study, we used 24 sample libraries from the last two time points (6 dpi and 9 dpi) to focus on gene regulation mechanism in the latter stages of maize-fungal interaction. The general information of our NGS datasets is shown in Fig. 1A. From these genes, we selected 324 most significant differentially and highly expressed genes either in wild type or *fsr1* mutant from our datasets, where all replicates were normalized and analyzed for their individual relative expression levels at three different time points. As shown in a heat map with three distinct time points (Fig. 1B), 155 genes (red) are expressed significantly higher in the wild type and 169 genes (blue) are expressed significantly higher in *fsr1* mutant (Fig. 1B and Table S1). As explained earlier, the relative abundance was acquired by the two-step normalization by each gene length as well as β -tubulin genes, and was selected by *t*-test statistics score measurement ($|t\text{-test score}| > 3.5$). However, this common NGS analysis focuses on relative expression of individual genes but does not allow us to predict gene-gene associations and system-level changes across correlated genes during pathogenesis.

Identification of *F. verticillioides* subnetwork modules. We developed a computational workflow that allows us to build co-expression networks from *F. verticillioides* NGS datasets¹⁷. We first inferred the co-expression networks for the wild type as well as *fsr1* mutant utilizing the preprocessed gene expression data by using partial correlation¹⁹ (Supplementary Information). In this co-expression network, we applied five distinct thresholds (0.965, 0.97, 0.975, 0.98, and 0.985), thereby constructing five different co-expression networks. The number of genes and edges between genes are shown in Fig. 1C. When these co-expression networks are illustrated with all member genes and possible edges, we can generate a complex web of scale-free networks (Fig. 1D). However, the aim of our proposed network-based NGS data analysis is to search through these co-expression networks to identify subnetwork modules that are differentially activated between the *F. verticillioides* wild type and *fsr1* mutant, that can considerably differ in terms of virulence potentials (Fig. 2).

By following this proposed strategy, we developed two potential subnetwork modules differentially activated during *F. verticillioides* pathogenesis. We performed additional analyses with six different characteristics for selecting hubs and their modules followed by our network-based comparative analysis approach¹⁷ (Fig. 3). In the subnetwork module fine-tuning process, two modules in Fig. 4 showed the minimum discriminative power increase for the entire module adjustment as 22% and 27% while over 90% of modules displayed smaller than 20% increase. Note that our approach probabilistically focuses on generating subnetwork modules whose member genes have high likelihood of showing associated expression patterns to each other across all replicates using the log-likelihood-ratio (LLR) matrix that demonstrates how likely each gene would express in *F. verticillioides* wild

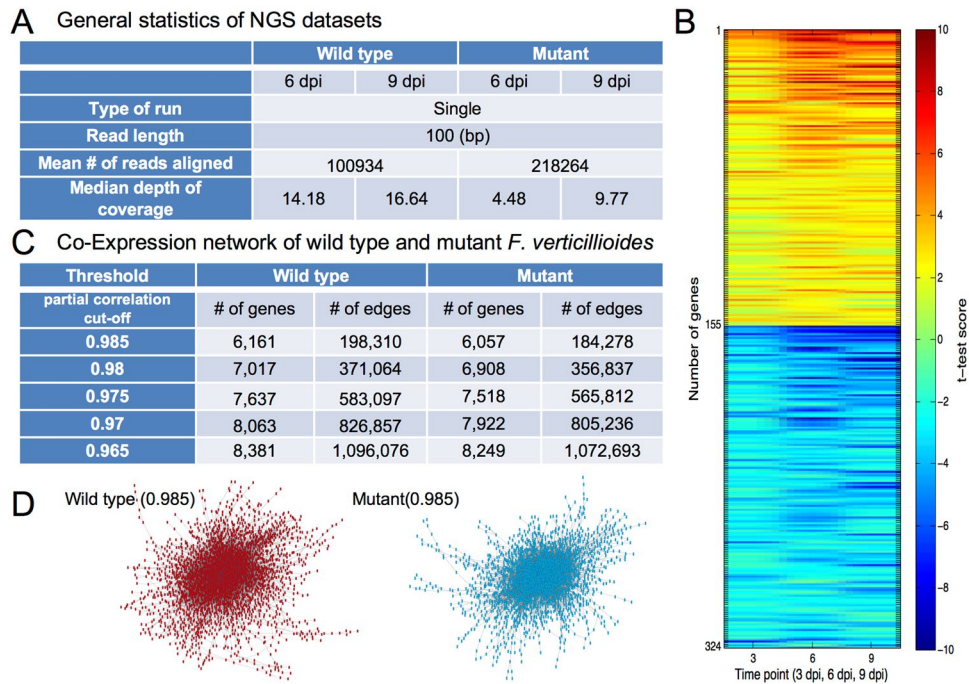


Figure 1. NGS statistics of significant differentially expressed genes. **(A)** General statistics of next-generation sequencing (NGS) datasets. Reads from wild type and *fsr1* mutant (6 dpi and 9dpi) were mapped to *F. verticillioides* reference genome. **(B)** The heat map provides a schematic overview of 324 most significant differentially expressed genes at three distinct time points. A total of 155 genes are expressed significantly higher in the wild type (red) while a total of 169 genes are expressed significantly higher in the mutant (blue). In this selection, genes whose absolute *t*-test statistics score is higher than 3.5 were chosen, where the relative abundance was acquired by the two-step normalization which considers each gene length as well as relative expression against β -tubulin genes over all the time points (3 dpi, 6 dpi, and 9 dpi). Next, *t*-test scores of the selected genes were again measured over relative expression levels normalized at each time point, and displayed in colors. **(C)** Co-expression network of wild type and mutant *F. verticillioides*. We applied five distinct threshold levels to generate five different co-expression networks. The table shows number of genes and all possible edge combinations in each co-expression network. **(D)** Schematic depiction of wild type and mutant co-expression networks at threshold level 0.985.

type or the mutant. As a result, our network-based computational analysis approach found potential subnetwork modules that show harmonious coordination of member genes as well as strong differential activity between the two strains.

Computational characterization of two key *F. verticillioides* subnetwork modules. From our network-based comparative analysis, we identified two potential pathogenicity-associated subnetwork modules differentially activated between the wild-type and *fsr1* mutant strains (Fig. 4). Module A was composed of ten *F. verticillioides* genes, where 80% of these were annotated with a significant GO term cytoplasmic component (GO:0044444) (<http://biit.cs.ut.ee/gprofiler/index.cgi>)²⁰. However, it is important to note that majority of these genes have no known function and these GO functions were chosen solely based on predicted protein motifs (Table S2). Module B was comprised of fifteen genes, where four (FVEG_07930, FVEG_00890, FVEG_11886, and FVEG_00594) were annotated with a significant GO term transport (GO:0006810). The eleven other genes were hypothetical proteins with some knowledge of their functional domains. But this module showed relatively higher percentage of genes with no GO terms and no functional protein domains compared to module A (Table S2). Once we defined these subnetwork modules, we analyzed all member genes *in silico* to predict potential hub genes that may hold a key role in *F. verticillioides* pathogenicity.

In module A, we selected FVEG_11622 as a potential pathogenicity-associated hub gene based on following observations: (i) FVEG_11622 deteriorated the differential probabilistic activity level of its given module from wild type to mutant by 26% (the mean of other member genes was 16%), (ii) correlation coefficients of FVEG_11622 decreased from wild type to mutant by 0.26 and 0.34 for Pearson's and Spearman rank, respectively (the mean of other member genes was 0.14 and 0.19), (iii) FVEG_11622 contained four edges to other member genes (the mean of other member genes was 2.8), (iv) FVEG_11622 demonstrated significant expression decrease from wild type to mutant (*t*-score of 4.4), and (v) orthologous gene of FVEG_11622 in *Botrytis cinerea* is recognized as having a role in fungal virulence^{21,22}. The predicted hub gene FVEG_11622, which was tentatively designated as *FvEBP1*, encodes a putative 238-AA hypothetical protein that harbors Emopamil-binding protein (EBP) domain (pfam05241).

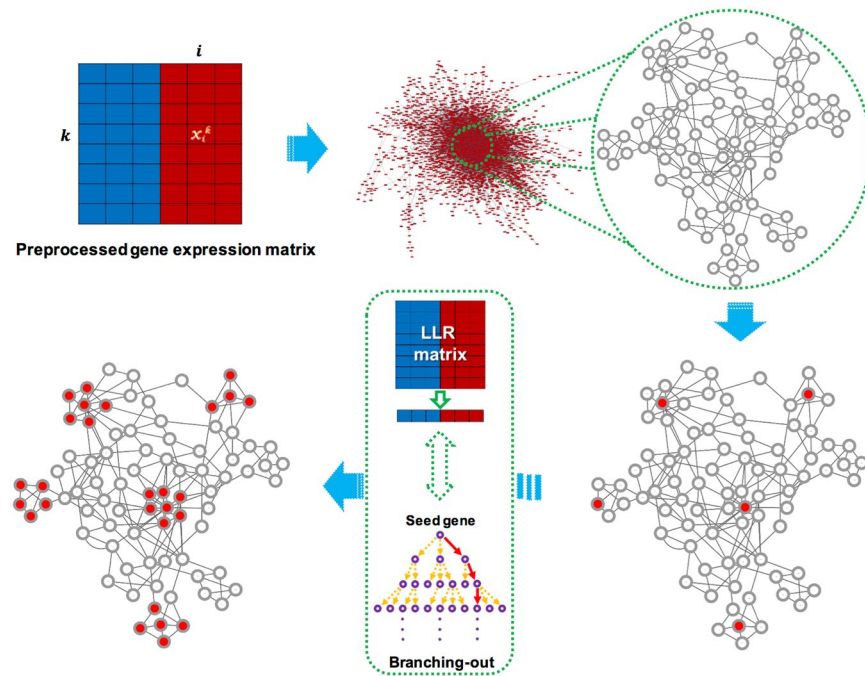


Figure 2. Schematic overview of our network-based NGS data analysis. Our aim is to search through large co-expression networks to identify subnetwork modules that are differentially activated between two different conditions (e.g. wild type versus *fsr1* mutant). Initial selection of seed genes (i.e. top 1% differentially expressed genes) is followed by a series of computational procedures described previously¹⁷. Through this process, we can identify subnetwork modules that show significant difference in virulence potentials.

Using the same approach, we identified FVEG_00594 as the potential pathogenicity-associated hub gene in module B based on following observations: (i) FVEG_00594 reduced the differential probabilistic activity level of its detected module from wild type to mutant by 24% (the mean of other member genes was 13%); (ii) correlation coefficient difference of FVEG_00594 between wild type and mutant was 0.34 and 0.4 for Pearson's and Spearman rank, respectively (the mean of other member genes was 0.24 and 0.2); (iii) FVEG_00594 included four edges to other member genes (the mean of other member genes was 3.7); (iv) the difference of expression level of FVEG_00594 was higher in wild type although it did not show high significance (t-score as 0.8), and v) the ortholog of FVEG_00594 in *M. oryzae* is recognized as having a role in intracellular trafficking and pathogenicity²³. FVEG_00594, designated *FvSYN1*, encodes a putative 377 amino-acid protein that harbors two well-recognized domains syntaxin N-terminal domain (cd00179) and SNARE domain (cd15849).

Functional characterization of predicted hub genes associated with virulence. To test our hypothesis that *FvEBP1* (FVEG_11622) and *FvSYN1* (FVEG_00594) are putative hub genes of subnetwork modules A and B, respectively, and that they are important for *F. verticillioides* virulence, we generated gene knockout mutants $\Delta fvebp1$ and $\Delta fvsyn1$ through homologous recombination following our standard split marker protocol²⁴. Hygromycin B phosphotransferase (*HPH*) was used as the selective marker, and homologous recombination outcomes were confirmed by PCR (data not shown) and Southern blots (Fig. S1). We first compared vegetative growth of these mutants on synthetic media (PDA, V8 agar and defined medium agar). While $\Delta fvsyn1$ strain showed reduced colony growth, $\Delta fvebp1$ strain exhibited no growth defect (Fig. 5A). The mutant $\Delta fvsyn1$ showed restricted radial vegetative grow while exhibiting more dense and fluffier mycelial growth on solid media when compared to the wild type (Fig. 5B). When cultures were harvested from YEPD broth, we did not observe a significant difference in fungal mass production (Fig. 5C). For spore production on V8 plates, $\Delta fvsyn1$ produced significantly reduced spores when compared to other strains (Fig. 5D).

To test virulence, we inoculated B73 maize seedling mesocotyls with spore suspension of wild-type, $\Delta fvebp1$, and $\Delta fvsyn1$ strains (along with water as a negative control) following the previously described procedure²⁵. When symptoms were observed after a 2-week incubation period, $\Delta fvebp1$ and $\Delta fvsyn1$ mutants showed significantly decreased levels of rot when compared with the wild-type progenitor (Fig. 5E). Mutants $\Delta fvebp1$ and $\Delta fvsyn1$ showed approximately 70% and 60% reduction in virulence when analyzed by average mesocotyl rot area (Fig. 5E). In order to test whether the mutant phenotype is due to a targeted gene replacement, we generated complementation strains of $\Delta fvebp1$ and $\Delta fvsyn1$ by co-transforming each mutant protoplasts with the respective wild-type gene (*FvEBP1* and *FvSYN1* with their native promoter and terminator) along with the geneticin-resistance gene. PCR was performed to confirm reintroduction of wild-type genes in complemented strains. *FvSYN1*C strain showed complete restoration of virulence on maize seedlings whereas *FvEBP1*C showed partial (~75%) recovery (Fig. 5E). These results suggested that *FvEBP1* and *FvSYN1* play an important role in

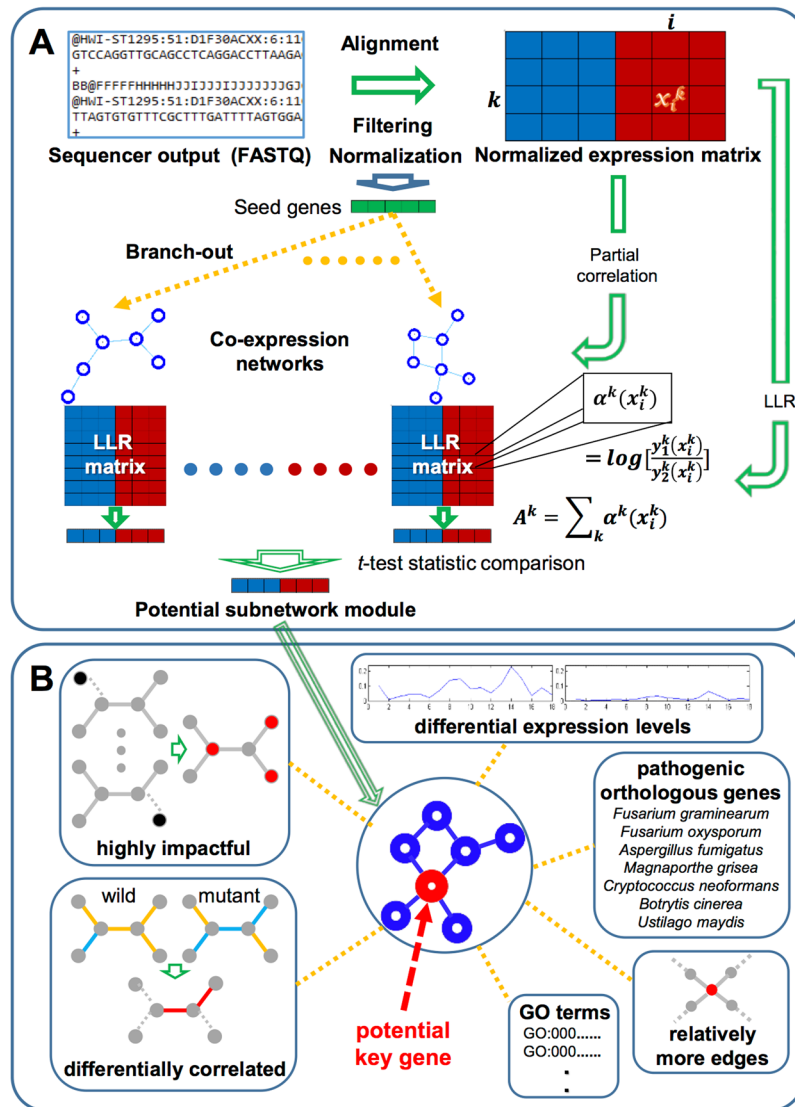


Figure 3. Computational prediction procedure for identifying key potential pathogenicity genes. (A) Raw NGS datasets are preprocessed, *i.e.* alignment, filtering, and normalization, before they are applied for inferring *F. verticillioides* co-expression networks by means of partial correlation. In addition, these datasets were also converted into a log-likelihood ratio (LLR) matrix for downstream analysis. Next, subnetwork modules are extended from seed genes, significantly differentially expressed between the two different conditions (*F. verticillioides* wild type vs. mutant), as long as they keep sufficient strength of differential activity between the two strains [1–2]. (B) Each potential hub (virulence-associated) gene is predicted in its detected subnetwork module based on several criteria: (i) highly impactful in a probabilistic manner, (ii) relatively differentially correlated between two strains (wild type vs. *fsr1* mutant), (iii) relatively more connected in the given module, (iv) relatively significantly differentially expressed, (v) orthologous to known pathogenicity-associated genes of other fungal species, (vi) annotated to significant GO terms with other member genes. Through this proposed analysis approach, we identified potential functional genes showing significant differential activity between the two conditions as well as strong association with virulence.

virulence on maize seedling rot, and further convinced us that these two genes serve as the predicted hub gene of their respective subnetwork module.

Testing network robustness in gene deletion mutants. A very important feature of these subnetwork modules is having robustness, *i.e.* the ability to respond to and withstand external as well as internal stimuli while maintaining its normal behavior²⁶. However, it is reasonable to predict that when we eliminate or disable a critical node (*i.e.* a hub gene), the network could be disrupted and shattered into isolated nodes. If a hub gene is eliminated from the subnetwork, we can hypothesize that other member genes, particularly those sharing direct edges, will exhibit disparate expression patterns.

We first tested correlated gene expression patterns in the wild type versus $\Delta fvebp1$ mutant by qPCR. We learned that gene expression levels of FVEG_03416, FVEG_04142, and FVEG_08818 were drastically lowered in

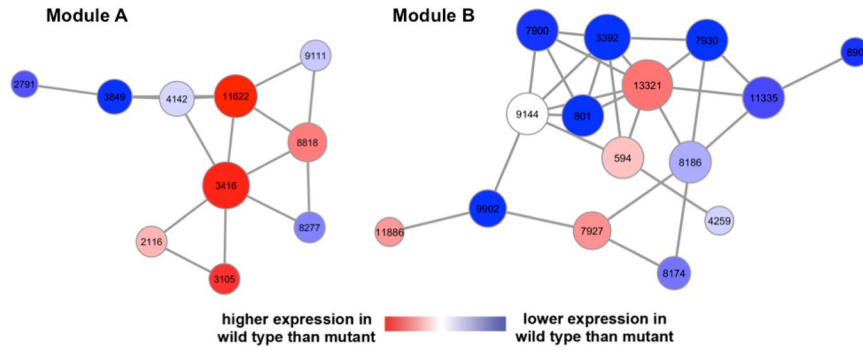


Figure 4. Potential subnetwork modules associated with the *F. verticillioides* pathogenicity. Our network-based comparative analysis identified two potential virulence-associated subnetwork modules differentially activated between the two *F. verticillioides* strains (wild type vs. mutant). Module A is composed of ten *F. verticillioides* genes, where 80% of these were annotated to a significant GO term GO:0044444 “cytoplasmic part”. Module B is comprised of fifteen genes, where four of these (FVEG_07930, FVEG_00890, FVEG_11886, and FVEG_00594) were annotated to a significant GO term GO:0006810 “transport”. The node sizes are directly proportional to their number of edges and the node colors vary according to their discriminative power measured by *t*-test scores.

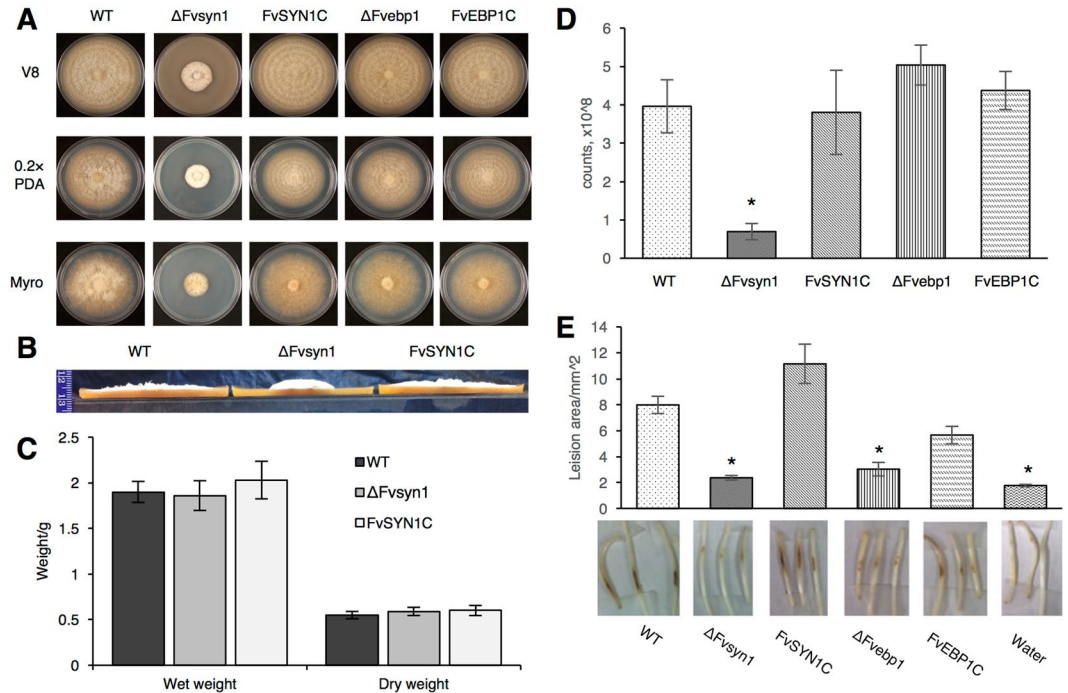


Figure 5. Functional characterization of *F. verticillioides* FvSYN1 and FvEBP1. (A) Vegetative growth of wild-type (WT), $\Delta Fvsyn1$, $\Delta Fvebp1$ and their complementation strains (FvSYN1C and FvEBP1C) were examined on V8, 0.2XPDA, and Myro agar plates. Strains were point inoculated with an agar block (0.5 cm in diameter) and incubated for 6 days at 25 °C under 14 h light/10 h dark cycle. (B) Spores (2×10^7) of WT, $\Delta Fvsyn1$ and FvSYN1C were inoculated in the center of V8 plates for 6 days at 25 °C under 14 h light/10 h dark cycle. Vegetative growth of WT, $\Delta Fvsyn1$ and FvSYN1C on V8 agar plates, strain growth condition was the same as described above. Agar plates were cut into half and pictures were taken from a side view. (C) Fungal mass production of WT, $\Delta Fvsyn1$ and FvSYN1C strains was tested in YEPD broth. Spore suspension ($5 \mu\text{l}$, $10^8/\text{ml}$ concentration) was inoculated and incubated for 4 days at 25 °C and shaking at 150 rpm. Fungal mass production was quantified by weighing wet and dry fungal mass. The data presents the average and standard deviation of three independent experiments. (D) Spore production of WT, $\Delta Fvsyn1$, FvSYN1C, $\Delta Fvebp1$ and FvEBP1C on V8 agar plates, strain growth condition was the same as described above. Spores were collected from agar plates and counted. (E) One-week-old B73 seedlings were inoculated with spore suspension ($5 \mu\text{l}$, $10^8/\text{ml}$ concentration) of fungal strains on mesocotyls. Lesion areas were quantified by Image J software after 2-week incubation. Asterisk above the column indicates statistically significant difference ($P < 0.05$) analyzed by *t*-Test.

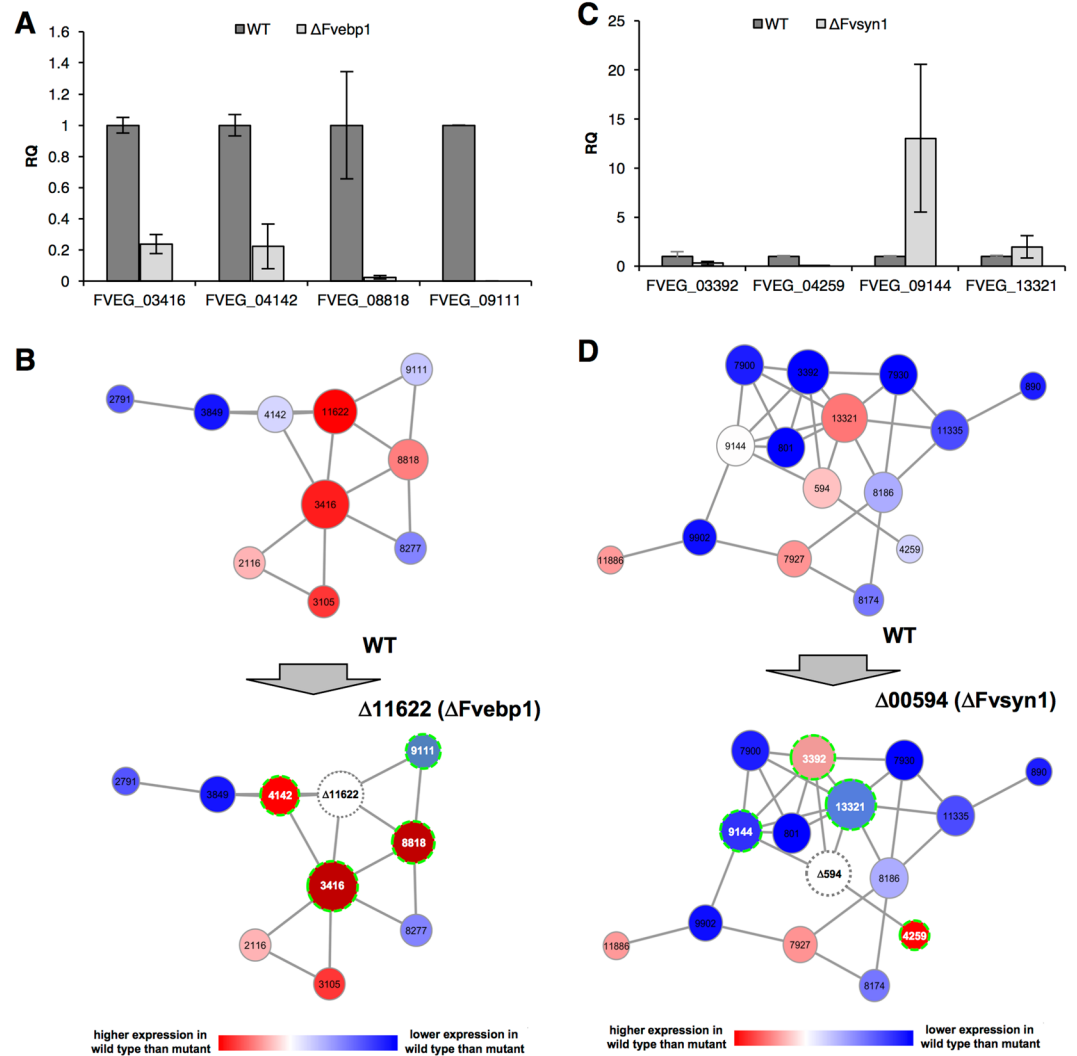


Figure 6. Altered expression of select neighboring genes as detected by qPCR. **(A)** Relative quantification (RQ) of four neighboring genes (FVEG_04142, FVEG_08818, FVEG_03416, FVEG_09111) to predicted hub *FvEBP1* (FVEG_11622) in wild type (WT) versus $\Delta Fvebp1$. **(B)** Schematic overview of transcriptional changes of four neighboring genes (highlighted) of *FvEBP1* (FVEG_11622) observed in WT versus $\Delta Fvebp1$. **(C)** RQ of four neighboring genes (FVEG_03392, FVEG_09144, FVEG_13321, FVEG_04259) to predicted hub *FvSYN1* (FVEG_00594) in wild type (WT) versus $\Delta Fvsyn1$. RQ levels of four genes in WT were normalized to 1. **(D)** Schematic overview of transcriptional changes of four neighboring genes (highlighted) of *FvSYN1* (FVEG_00594) observed in WT versus $\Delta Fvsyn1$.

the $\Delta fvebp1$ mutant than those observed in the wild-type progenitor (Fig. 6A). Furthermore, the FVEG_09111 gene expression level was not detectable in the mutant. Particularly, FVEG_04142 and FVEG_09111 showed higher levels of expression in *fsr1* mutant when compared to the wild type, and in $\Delta fvebp1$ that expression pattern is now reversed (Fig. 6B). These results show that when *FvEBP1* is no longer present in the subnetwork, expression levels of these genes, FVEG_03416, FVEG_04142, and FVEG_08818, and FVEG_09111, are drastically suppressed (Fig. 6A,B), suggesting *FvEBP1* is critical for proper regulation of these neighboring genes.

In the $\Delta fvsyn1$ strain, we comparatively studied the expression pattern of four genes that directly share edges with *FvSYN1*. Three of the four genes tested, FVEG_03392, FVEG_04259 and FVEG_09144 showed a significant difference in expression levels between the wild type and $\Delta fvsyn1$ mutant. Significantly, FVEG_03392 and FVEG_04259, which showed lower expression level in the wild type when compared with *fsr1* mutant, reversed its course and showed higher expression in the wild type when compared with $\Delta fvsyn1$ (Fig. 6C). FVEG_09144, which showed no difference in expression between wild type and *fsr1* mutant, showed significantly higher expression in $\Delta fvsyn1$. FVEG_13321, which showed higher expression in the wild type compared to *fsr1* mutant, now exhibits statistically similar expression in wild-type and $\Delta fvsyn1$ (Fig. 6C,D). Collectively these data showed that *FvSYN1* and *FvEBP1* are important for regulating the expression of closely correlated genes, further providing evidence that these are important hub genes of their respective subnetworks.

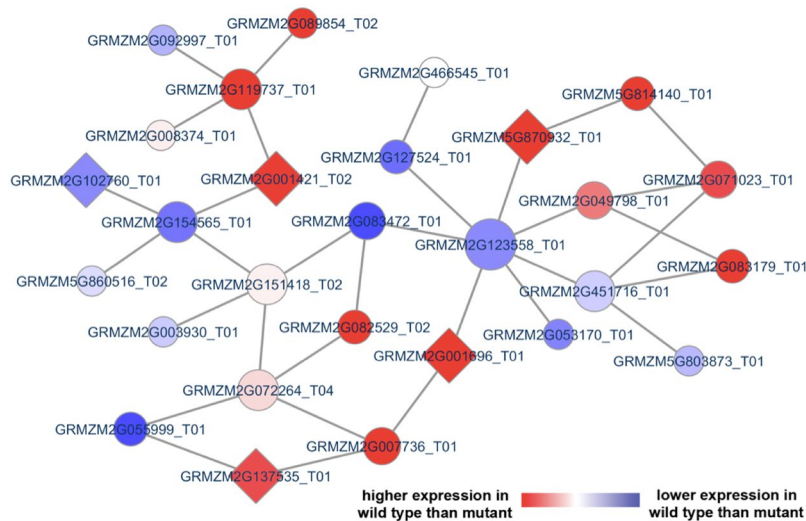


Figure 7. Potential subnetwork module associated with maize stalk defense response. Our network-based comparative analysis of maize genes applied cointegration-correlation-expression strategy to identify a potential maize subnetwork module differentially activated between the two strains (*F. verticillioides* wild type vs. *fsr1* mutant). This predicted maize module is comprised of 28 member genes, and five were annotated to a significant GO term GO:0009814 “defense response, incompatible interaction”. These five genes were indicated by diamond-shaped illustration.

Predicting maize defense-associated co-expression subnetwork modules. We used the cointegration-correlation-expression approach to build potential defense-associated modules on maize co-expression networks²⁷. An illustration of the potential subnetwork module associated with maize defense response against the *F. verticillioides* pathogenicity genes is shown in Fig. 7. We selected three *F. verticillioides* pathogenicity genes identified from our current work (FVEG_11622, FVEG_00594, and FVEG_09767 [*FSR1*]) to predict maize stalk response subnetworks. The subnetwork module associated with maize defense system was composed of 28 maize genes, where genes significantly expressed in wild type-infected maize are indicated in red and genes significantly expressed in mutant-infected maize are indicated in blue (Table S3). In this potential maize defense gene subnetwork module, we noticed that five maize genes (GRMZM2G102760_T01, GRMZM5G870932_T01, GRMZM2G001421_T02, GRMZM2G001696_T01, and GRMZM2G137535_T01) were annotated with a significant GO term defense response/incompatible interaction (GO:0009814), defined as “a response of a plant to a pathogenic agent that prevents the occurrence or spread of disease”. However, as seen in *F. verticillioides* subnetwork modules, we recognize that a large percentage of member genes encode hypothetical proteins and that transcriptional coordination does not always result in functional correlation. In addition, unlike *F. verticillioides* genes we characterized in this study, generating null mutants and performing network robustness assays are more strenuous for maize.

Discussion

In this study, we assembled a streamlined computational network analysis pipeline to investigate the system-level coordinated changes across differentially activated genes rather than simply focusing on differential transcript abundance of individual genes, and to detect subtle processes that are not likely to be revealed by examining a small list of highly significant genes in this host-pathogen interaction. To generate meaningful prediction from limited datasets, a comprehensive and rigorous investigation was needed. Thus, we mainly searched for comparable expression patterns probabilistically using a log-likelihood ratio matrix over replicates instead of just considering differential expression for identifying potential subnetwork modules. Also, we analytically investigated the given subnetwork modules with multidirectional analysis considering factors such as probabilistic impact, and differential correlation. Significantly, this comprehensive approach can help identify novel virulence-associated subnetwork modules as well as the key functional “hub” genes in fungal pathogens, such as *F. verticillioides*. This assembly of tools will be instrumental as we continue our effort to harness new and meaningful information from NGS datasets as we try to better understand complex pathosystems.

Our study mainly focused on analyzing the underlying transcriptional regulation in host-pathogen interactions. However, we do recognize that complex intercellular web of interactions in a living cell, not to mention between a host and its pathogen, are not limited to gene-gene association. Numerous constituents of the cell, e.g. DNA, RNA, protein, and metabolites, contribute to the structure and the dynamics of cellular network and ultimately behavior. However, in contrast to DNA and RNA, the resources available for us to generate systems-level proteome and metabolome datasets for network analyses are currently limited. In addition to these challenges, majority of host-pathogen systems have very limited genetic information available. For instance, as one can see from our two predicted modules (Fig. 4 and Table S2) many member genes encode hypothetical proteins with unknown, and vaguely predicted, functions. We primarily focused on developing this computational approach

with the intent of investigating not-well defined biological systems with minimal bias toward existing genetic information, *i.e.* allocating higher scores toward known virulence genes in given species.

Analyses of two subnetworks led us to discover two potential pathogenicity-associated hub genes *FvEBP1* and *FvSYN1*. *FvEBP1* encodes a protein that harbors Emopamil-binding protein (EBP) domain (pfam05241), and the EBP protein family in mammalian systems is known to be associated with endoplasmic reticulum and plays a critical role in sterol isomerization and lipoprotein internalization²⁸. An emopamil binding protein BcPIE3 in *Botrytis cinerea* which shares significant structural similarities to mammalian EBPs was shown to be important for virulence²⁹. Predicted FvSyn1 protein contains syntaxin N-terminal domain (cd00179) and SNARE domain (cd15849). In budding yeast, the SNARE protein complex is involved in membrane fusion and protein trafficking for new synthesis and recycling of organelles³⁰. SNAREs were originally classified into v-SNAREs and t-SNAREs according to their vesicle or target membrane localization³¹. Syntaxins belong to t-SNARE proteins and are shown to play an important role in membrane fusion in eukaryotic cells^{32,33}. Syntaxins are known as a family of membrane-associated receptors for intracellular transport vesicles. Syntaxin and SNAREs are also known to anchor these newly synthesized and recycled proteins to the cytoplasmic surface³⁴. SNARE proteins play critical and conserved roles in intracellular membrane fusion in eukaryotic cells³⁵. They were known to mediate membrane fusion during all trafficking steps of the intracellular communication process, including the secretory and endocytic pathways³⁶.

Subnetwork module analyses further allowed us to identify genes that show strong coordinated changes with these hub genes. *FvEBP1* has four direct edges to FVEG_03416, FVEG_04142, FVEG_08818 and FVEG_09111. FVEG_03416 is an alginate lyase gene, and contains an alginate lyase domain which is important for fructose and mannose metabolism. FVEG_04142 is a V-type proton ATPase subunit F. V-type ATPases have hydrogen-exporting ATPase activity and are involved in ATP hydrolysis coupled proton transport. FVEG_09111 is a hypothetical protein, containing a PX-associated domain. The function of this protein is unknown, but its N-terminus is always found to be associated to a PX domain which is involved in targeting of proteins to cell membranes. FVEG_08818 is a hypothetical protein with a methyltransferase domain. *FvSYN1* has four directly associated genes in the subnetwork module: FVEG_03392, FVEG_04259, FVEG_09144 and FVEG_13321. Three of these genes (FVEG_03392, FVEG_04259, FVEG_09144) encode hypothetical proteins with no known functional motif thus making it difficult to predict their role. While FVEG_13321 is a hypothetical protein, it does contain a fungal Zn₂Cys₂ binuclear cluster domain, which is typically found in the family of fungal zinc cluster transcription factors^{37,38}.

Furthermore, there is a greater challenge in refining subnetwork module development for host organisms that typically has larger and more complex genomes. Over 95% of our NGS data generated in this study came from maize, suggesting that maize stalk is actively responding to the pathogen invasion and colonization at transcriptional level²⁷. In our earlier study, we developed a computational analysis pipeline, including the cointegration-correlation-expression approach, to predict potential maize defense-associated genes that show strong differential activation and coordination with known *F. verticillioides* virulence genes²⁷. Here, we used three *F. verticillioides* pathogenicity genes to build maize response subnetworks. While we acknowledge the limitation associated with this approach, *e.g.* solely relying on transcriptional coordination, these predicted host subnetworks can provide us an understanding of how host genes are responding collectively against the pathogen aggression. Comparing host co-expression subnetworks at different time points, *e.g.* early invasion versus colonization, can also help us understand the dynamics of host-pathogen association at the transcriptome level.

Difficulties mentioned above remain as obstacles. However our effort demonstrates that the proposed network-based analysis pipeline can improve our understanding of the biological mechanisms that underlie host-pathogen interactions. This approach also has the potential to unveil novel genetic subnetwork modules and hub genes critical for virulence in fungal pathogens. We are currently in the process of improving our experimental pipeline with computational network querying that can estimate node correspondence probabilities to find novel functional pathways in biological networks^{39–42}. This gene subnetwork approach can lead to the discovery of new quantifiable cellular subnetworks that can bridge the knowledge gaps in the maize-*F. verticillioides* system and be further applied to other plant-microbe pathosystems.

Methods

Fungal strains, maize line, and RNA sample preparation. *F. verticillioides* strain 7600⁴³ and *fsr1* mutant⁵ were cultured at 25 °C on V8 juice agar (200 ml of V8 juice, 3 g of CaCO₃ and 20 g of agar powder per liter). Maize inbred B73, a progenitor of numerous commercial hybrids with no inherent resistance to stalk rot, was inoculated with *F. verticillioides* wild type and *fsr1* mutant spore suspension as described previously^{5,25}. Maize stalk samples were collected 3, 6, and 9 dpi using manual sectioning, and microscopically inspected to identify host tissue damage and/or fungal colonization, particularly in the vascular bundles. For each sample, sectioning was performed on at least three independent stalk samples from each stage of infection, and isolated tissues were pooled for RNA extraction with TRIzol reagent (Invitrogen). For each time point, we collected six pooled samples, thus thirty-six RNA samples in total. Standard QA/QC procedure for RNA samples was implemented at the Texas A&M AgriLife Research Genomics and Bioinformatics Service (College Station, TX) prior to sequencing.

RNA Sequencing and data preprocessing. RNA sequencing was processed at the Texas A&M AgriLife Research Genomics and Bioinformatics Service using Illumina HiSeq 2000 as described previously²⁷. We sequenced a total of 36 sample libraries, *i.e.* six libraries per each time point (3 dpi, 6 dpi, and 9 dpi) for wild type and *fsr1* mutant inoculated maize stalks, but it is worth noting that we only used sequencing data for the last two time points in this study (6 dpi and 9 dpi, hence 24 sample libraries in total) to focus on gene regulation mechanism in the latter stages of maize-*F. verticillioides* interaction. Percentages of the two β -tubulin read abundance were traced over 24 replicates in 6 dpi and 9 dpi to examine their expression consistency. Mean (μ) and standard

deviation (σ) of the percentages for β -tubulins were $\mu = 0.058\%$, $\sigma = 0.0056$ for FVEG_04081 and $\mu = 0.035\%$, $\sigma = 0.0044$ for FVEG_05512.

Prediction of *F. verticillioides* subnetwork modules associated with stalk rot. A procedure of identifying candidate functional subnetwork modules follows computational analysis pipeline described earlier¹⁷ with some modifications, and additional detail is provided in Supplementary Information. From the preprocessed gene expression data, we performed conversion into log likelihood ratio (LLR) matrix, construction of co-expression networks based on partial correlation, and selection of the most significant differentially expressed genes (*i.e.*, top 1%) between the two strains (wild type vs *fsr1* mutant) as seed genes. Based on this preparation, we applied computationally efficient branching out searching from a seed gene until it does not meet minimum discriminative power increase for each subnetwork module. The entire search process was reiterated for every seed gene and for the five co-expression networks (Supplementary Information). Note that our approach probabilistically searches for subnetwork modules whose member genes have highly coordinated expression patterns to each other over all the replicates using the LLR matrix that demonstrates how likely each gene would be regulated in wild type or *fsr1* mutant. As a result, our network-based computational analysis approach found candidate subnetwork modules that show harmonious coordination of member genes as well as strong differential activity between the wild type and *fsr1* mutant.

Prediction strategy for hub genes in each subnetwork module. Our proposed computational approach in this study simultaneously inferred potential *F. verticillioides* pathogenicity-associated hub or key functional genes in each subnetwork module while branching out the modules by performing multidirectional analysis on the detected candidate subnetwork modules with six different criteria, as depicted in Fig. 3B. First, we investigated how each gene is probabilistically impactful in its subnetwork module utilizing the probabilistic inference strategy applied in our previous work¹⁷. We estimated probabilistic differential activity of each gene by comparing discriminative power on the two phenotypes (wild type vs *fsr1* mutant) between its given module with and without the gene. As described in our previous study¹⁷, we computed discriminative power difference (estimated by *t*-test statistics) between both activity levels (one with the gene and the other without the gene) by supposing $\zeta = \{g_1, g_2, \dots, g_n\}$, member genes in a subnetwork module, and $e = \{e^1, e^2, \dots, e^n\}$, expression levels of the given genes. The discriminative power difference $D(d)$ was calculated as follows.

$$D(d) = \left[\sum_{k=1}^n \log \left(\frac{f_1^k(e^k)}{f_2^k(e^k)} \right) \right]_{\zeta}^{t\text{-test score}} - \left[\sum_{k=1}^n \log \left(\frac{f_1^k(e^k)}{f_2^k(e^k)} \right) \right]_{\zeta - \{g_d\}}^{t\text{-test score}} \quad (1)$$

where $\log \left(\frac{f_1^k(e^k)}{f_2^k(e^k)} \right)$ is the log-likelihood ratio (LLR), and $f^k(e)$ is each conditional probability density functions (*i.e.*, either wild or mutant). We subsequently considered genes whose discriminative power deterioration $D(d)$ relatively larger as candidate key genes. Second, we examined differential correlation of each gene with its connected gene in each module between the wild type versus the mutant using two correlation methods, *i.e.* Pearson's correlation and Spearman rank correlation. We selected genes whose correlation coefficients with its neighboring genes were not only relatively significantly different between the two different networks, but also higher in the wild-type network as candidate hub genes. Third, we calculated number of edges of each gene to other member genes in each module since it is reasonable to predict that genes with more edges will exhibit more meaningful influence on the module. Fourth, we investigated expression difference of each gene in a given module between the two strains since we predict that altered expression of genes downstream of *FSR1* can follow that of *FSR1*. We selected genes that were significantly and differentially expressed between the two conditions and relatively high expression in the wild type. Fifth, we listed orthologs of known pathogenicity genes in other well-studied fungal species such as *F. graminearum* (FG), *F. oxysporum* (FO), *Aspergillus fumigatus* (AF), *Botrytis cinerea* (BC1G), *Magnaporthe grisea* (MGG), *Ustilago maydis* (UM), and *Cryptococcus neoformans* (CNAG). We noted genes in a given module shown in the list of pathogenic genes as potential key genes. Finally, we considered whether each gene in a given module was associated with a significant GO term. We applied *p*-values of Benjamini-Hochberg false discovery rate (FDR) method⁴⁴ to find the most relevant GO term to each gene and its given module based on g:Profiler (<http://biit.cs.ut.ee/gprofiler/index.cgi>)²⁰. We chose genes annotated with a GO term that is the most significantly associated term with the given module as candidate key genes.

After identifying a candidate hub (or key functional) gene in its given subnetwork module, we performed additional fine-tuning procedure for more robust and reliable module prediction. Briefly, once we identified a candidate gene satisfying the above-mentioned six criteria through subnetwork module extension process with 10% minimum discriminative power enhance, we performed module adjustment process by implementing the whole module extension for the given module as well as identifying the same potential gene by escalating the minimum discriminative power enhance by 1%. We repeatedly applied this fine-tuning process until the predicted hub (or key functional) gene did not meet all conditions. This process also stopped when the given subnetwork module grew smaller down to the arbitrary set minimum size of seven genes.

Nucleic acid manipulation, polymerase chain reaction (PCR), and fungal transformation. Standard molecular manipulations, including PCR and Southern hybridization, were performed as described previously²⁴. Fungal genomic DNA was extracted using the OminiPrep genomic DNA extraction kit (G Biosciences, Maryland heights, MO, USA). The constructs for transforming *F. verticillioides* were generated with a split-marker approach described earlier²⁴. Briefly, DNA fragments of 5' and 3' flanking regions of each gene were PCR amplified from wild-type genomic DNA. Partial Hygromycin B phosphotransferase (*HPH*) gene (*HP-* and *-PH*)

fragments were amplified from pBS15 plasmid. 5' and 3' flanking region fragments were then fused with PH- and -HP fragments by single-joint PCR, respectively. The single-joint PCR products were transformed into wild-type fungal protoplast. For complementation, respective wild-type genes driven by its native promoter was co-transformed with a geneticin-resistant gene (*GEN*) into mutant protoplasts. All primers used in this study were listed in Table S4. *F. verticillioide*s protoplast were generated and transformed following standard protocol²⁴ with minor modifications. Murinase (2 mg/ml) was replaced with Driselase (5 mg/ml) (Sigma, St Louis, MO, USA) in the protoplast digestion solution. Transformants were regenerated and selected on regeneration medium containing 100 µg/ml of hygromycin B (Calbiochem, La Jolla, CA, USA) and/or 150 µg/ml G418 sulfate (Cellgro, Manassas, VA, USA) as needed. Respective drug-resistant colonies were screened by PCR and further verified by Southern analysis.

Maize infection assays. Maize seedling rot pathogenicity assay was performed on 2-week old maize inbred lines B73 seedlings as previously described²⁵ with minor modifications. Briefly, spore suspensions in YEPD broth along with YEPD control were inoculated on maize B73 mesocotyls. Plant mesocotyls were first slightly wounded by a syringe needle about 3 cm above the soil. A 5-µl spore suspension (1×10^8 /ml) was applied to the wound site. The seedlings were immediately covered with a plastic cover to create a high moisture environment suitable for infection and colonization. The seedlings were collected and analyzed after a 2-week growth period in the dark room. At least three biological and three technical replicates were performed for each fungal strain.

Expression analysis of subnetwork member genes linked to predicted hub genes. Total RNA extractions were conducted by using RNeasy plant mini kit (Qiagen) according to manufacturer's specifications and was quantified by Nanodrop. RNA was converted into cDNA using the Verso cDNA synthesis kit (Thermo Fisher Scientific, Waltham, MA) following the manufacturer's protocol. qRT-PCR analyses were performed using the SYBR Green Dynamo Color Flash qPCR kit (Thermo Fisher Scientific) on an Applied Biosystems 7500 Real-Time PCR system. The *F. verticillioide*s β -tubulin gene (*TUB-2*) was used as the endogenous calibrator. The amplification data analysis was done according to the manufacturer's protocol.

Identification of potential maize defense subnetwork module. We followed our previous analysis strategy²⁷ to identify potential subnetwork modules associated with maize stalk defense response against *F. verticillioide*s virulence genes. We began searching for maize modules possibly responsible for its defense mechanism through cointegration-correlation-expression analysis: i) Cointegration was applied to track an interrelationship of expression levels between maize and *F. verticillioide*s over all replicates. We applied the Engle-Granger correlation method to measure single cointegrating relations and p -value ≤ 0.05 was used to identify candidate maize genes that appear to have significant association with the identified *F. verticillioide*s pathogenicity-associated genes, ii) correlation was utilized to trace patterns of expression levels between maize and *F. verticillioide*s over all replicates. We used Pearson's correlation coefficients to estimate their linear relationship and condensed maize genes into candidates whose expression patterns are highly correlated with that of *F. verticillioide*s pathogenicity-associated genes (*i.e.*, p -value ≤ 0.005), iii) We considered expression levels of maize genes over replicates and filtered out insignificantly expressed maize genes (*i.e.*, maize genes whose mean expression levels were in the bottom 20% or not expressed in at least one replicate). We adjusted p -values of the cointegration-correlation-expression approach to have 50% of candidate maize genes predicted from a given *F. verticillioide*s pathogenicity-associated gene were also among the candidates inferred from other given *F. verticillioide*s gene. Based on the maize genes narrowed down through the cointegration-correlation-expression analysis, we identified subnetwork modules associated with maize defense response using our network-based comparative analysis approach.

Availability of Materials and Data. The datasets generated during and/or analyzed during the current study are publicly available at NCBI Sequence Read Archive (SRA) database (record# GSE108365).

References

- Stack, J. Common stalk rot diseases of corn. *Plant Dis.* **6**, 1–8 (1999).
- White, D. G. *Compendium of Corn Disease*. 3rd edn, (APS Press, 1999).
- Dodd, J. L. The role of plant stresses in development of corn stalk rots. *Plant Dis.* **64**, 533–537 (1980).
- Michaelson, M. E. Factors affecting development of stalk rots of corn caused by *Diplodia zeae* and *Gibberella zeae*. *Phytopathology* **47**, 499–503 (1957).
- Shim, W. B. *et al.* *FSR1* is essential for virulence and female fertility in *Fusarium verticillioide*s and *F. graminearum*. *Mol. Plant Microbe Int.* **19**, 725–733 (2006).
- Bartoli, M. *et al.* Down-regulation of striatin, a neuronal calmodulin-binding protein, impairs rat locomotor activity. *J. Neurobiol.* **40**, 234–243 (1999).
- Castets, F. *et al.* Zinedin, SG2NA, and striatin are calmodulin-binding, WD repeat proteins principally expressed in the brain. *J. Biol. Chem.* **275**, 19970–19977 (2000).
- Moreno, C. S. *et al.* WD40 repeat proteins striatin and S/G(2) nuclear autoantigen are members of a novel family of calmodulin-binding proteins that associate with protein phosphatase 2A. *J. Biol. Chem.* **275**, 5257–5263 (2000).
- Dettmann, A. *et al.* HAM-2 and HAM-3 are central for the assembly of the *Neurospora* STRIPAK complex at the nuclear envelope and regulate nuclear accumulation of the MAP kinase MAK-1 in a MAK-2-dependent manner. *Mol. Microbiol.* **90**, 796–812 (2013).
- Wang, C. L., Shim, W. B. & Shaw, B. D. *Aspergillus nidulans* striatin (StrA) mediates sexual development and localizes to the endoplasmic reticulum. *Fungal Genet. Biol.* **47**, 789–799 (2010).
- Pöggeler, S. & Kück, U. A WD40 repeat protein regulates fungal cell differentiation and can be replaced functionally by the mammalian homologue striatin. *Eukaryot. Cell* **3**, 232–240 (2004).
- Wang, C. L., Shim, W. B. & Shaw, B. D. The *Colletotrichum graminicola* striatin orthologue Str1 is necessary for anastomosis and is a virulence factor. *Mol. Plant Pathol.* **17**, 931–942 (2016).

13. Bloemendal, S. *et al.* A homologue of the human STRIPAK complex controls sexual development in fungi. *Mol. Microbiol.* **84**, 310–323 (2012).
14. Beier, A., Teichert, I., Krisp, C., Wolters, D. A. & Kück, U. Catalytic subunit 1 of protein phosphatase 2A is a subunit of the STRIPAK complex and covers fungal sexual development. *MBio* **7** (2016).
15. Yamamura, Y. & Shim, W. B. The coiled-coil protein-binding motif in *Fusarium verticillioides* Fsr1 is essential for maize stalk rot virulence. *Microbiol-SGM.* **154**, 1637–1645 (2008).
16. Zhang, H., Mukherjee, M., Kim, J.-E., Yu, W. & Shim, W. B. Fsr1, a striatin homolog, forms an endomembrane-bound signaling complex that regulates virulence in maize pathogen *Fusarium verticillioides*. *Mol. Plant Pathol.* **19**, 812–826 (2018).
17. Kim, M., Zhang, H., Woloshuk, C. P., Shim, W. B. & Yoon, B.-J. Computational prediction of pathogenic network modules in *Fusarium verticillioides*. *IEEE/ACM T. Comp. Bioinformatics* **15**, 506–515 (2018).
18. Su, J. J., Yoon, B. J. & Dougherty, E. R. Accurate and reliable cancer classification based on probabilistic inference of pathway activity. *PLoS One* **4**, e8161 (2009).
19. Hero, A. & Rajaratnam, B. Hub discovery in partial correlation graphs. *IEEE T. Inform. Theory* **58**, 6064–6078 (2012).
20. Reimand, J., Arak, T. & Vilo, J. G. Profiler-a web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Res.* **39**, W307–W315 (2011).
21. Amselem, J. *et al.* Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genetics* **7** (2011).
22. Blanco-Ulate, B. *et al.* Genome-wide transcriptional profiling of *Botrytis cinerea* genes targeting plant cell walls during infections of different hosts. *Front. Plant Sci.* **5** (2014).
23. Qi, Z. Q. *et al.* The syntaxin protein (MoSyn8) mediates intracellular trafficking to regulate conidiogenesis and pathogenicity of rice blast fungus. *New Phytol.* **209**, 1655–1667 (2016).
24. Sagaram, U. S., Shaw, B. D. & Shim, W. B. *Fusarium verticillioides* GAPI, a gene encoding a putative glycolipid-anchored surface protein, participates in conidiation and cell wall structure but not virulence. *Microbiol-SGM.* **153**, 2850–2861 (2007).
25. Christensen, S. A. *et al.* The novel monocot-specific 9-Lipoxygenase ZmLOX12 is required to mount an effective jasmonate-mediated defense against *Fusarium verticillioides* in maize. *Mol. Plant Microbe Int.* **27**, 1263–1276 (2014).
26. Barabasi, A. L. & Oltvai, Z. N. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–115 (2004).
27. Kim, M., Zhang, H., Woloshuk, C., Shim, W. B. & Yoon, B. J. Computational identification of genetic subnetwork modules associated with maize defense response to *Fusarium verticillioides*. *BMC Bioinformatics* **16**(Suppl 13), S12 (2015).
28. Seo, K. W., Kelley, R. L., Okano, S. & Watanabe, T. Mouse Td(ho) abnormality results from double point mutations of the emopamil binding protein gene (Ebp). *Mamm. Genome* **12**, 602–605 (2001).
29. Gioti, A. *et al.* A *Botrytis cinerea* emopamil binding domain protein, required for full virulence, belongs to a eukaryotic superfamily which has expanded in eucoscomycetes. *Eukaryot. Cell* **7**, 368–378 (2008).
30. Nicholson, K. L. *et al.* Regulation of SNARE complex assembly by an N-terminal domain of the t-SNARE Sso1p. *Nat. Struct. Biol.* **5**, 793–802 (1998).
31. Sollner, T., Bennett, M. K., Whiteheart, S. W., Scheller, R. H. & Rothman, J. E. A protein assembly-disassembly pathway *in vitro* that may correspond to sequential steps of synaptic vesicle docking, activation, and fusion. *Cell* **75**, 409–418 (1993).
32. Hong, W. J. SNAREs and traffic. *BBA-Mol. Cell Res.* **1744**, 120–144 (2005).
33. Gupta, G. D. & Heath, I. B. Predicting the distribution, conservation, and functions of SNAREs and related proteins in fungi. *Fungal Genet. Biol.* **36**, 1–21 (2002).
34. Yuan, Q. A. & Jantti, J. Functional analysis of phosphorylation on *Saccharomyces cerevisiae* syntaxin 1 homologues Sso1p and Sso2p. *PLoS One* **5** (2010).
35. Chen, Y. A. & Scheller, R. H. SNARE-mediated membrane fusion. *Nat. Rev. Mol. Cell Biol.* **2**, 98–106 (2001).
36. Hong, W. J. & Lev, S. Tethering the assembly of SNARE complexes. *Trends Cell Biol.* **24**, 35–43 (2014).
37. MacPherson, S., Larochele, M. & Turcotte, B. A fungal family of transcriptional regulators: The zinc cluster proteins. *Microbiol. Mol. Biol. Rev.* **70**, 583–604 (2006).
38. Schillig, R. & Morschhauser, J. Analysis of a fungus-specific transcription factor family, the *Candida albicans* zinc cluster proteins, by artificial activation. *Mol. Microbiol.* **89**, 1003–1017 (2013).
39. Qian, X., Sze, S. H. & Yoon, B. J. Querying pathways in protein interaction networks based on hidden Markov models. *J. Comput. Biol.* **16**, 145–157 (2009).
40. Qian, X. & Yoon, B. J. Effective identification of conserved pathways in biological networks using hidden Markov models. *PLoS One* **4**, e8070 (2009).
41. Sahraeian, S. M. E. & Yoon, B. J. Fast network querying algorithm for searching large-scale biological networks. *IEEE Int. Conf. Acoust. Speech Sig. Process.*, 6008–6011 (2011).
42. Su, J., Yoon, B. J. & Dougherty, E. R. Identification of diagnostic subnetwork markers for cancer in human protein-protein interaction network. *BMC Bioinformatics* **11**(Suppl 6), S8 (2010).
43. Ma, L. J. *et al.* Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**, 367–373 (2010).
44. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B Met.* **57**, 289–300 (1995).

Acknowledgements

This work was supported in part by the National Research Initiative Competitive Grants Program (No. 2007-35319-18334) and the Agriculture and Food Research Initiative (No. 2013-68004-20359) from the US Department of Agriculture. We would like to thank Mr. Trent LeMaster (Bioenvironmental Science, Texas A&M) and Mr. Jae-Young Jun (Electrical & Computer Engineering, Texas A&M) for excellent technical assistance.

Author Contributions

Research concepts: M.K., H.Z., W.B.S., B.-J.Y. experiments performed: M.K., H.Z. and H.Y. manuscript editing: M.K., H.Z., W.B.S. All authors reviewed and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-26505-2>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018