A THESIS FOR THE DEGREE OF MASTER OF SCIENCE

MagNet: the protein-protein interaction network of the rice blast fungus *Magnaporthe oryzae*

벼 도열병 균의 단백질 상호작용 네트워크 데이터베이스 구축

AUGUST, 2017

HYUNBIN KIM

INTERDISCIPLINARY PROGRAM IN AGRICULTURAL GENOMICS

COLLEGE OF AGRICULTURE AND LIFE SCIENCES

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

MagNet: the protein-protein interaction network of the rice blast fungus *Magnaporthe oryzae*

벼 도열병 균의 단백질 상호작용 네트워크 데이터베이스 구축

지도교수 이 용 환

이 논문을 농학석사학위논문으로 제출함
2017년 7월

서울대학교 대학원
협동과정 농생명유전체학전공

김 현 빈

김현빈의 농학석사학위논문을 인준함
2017년 6월

위 원 장 최도일 (인)

부 위 원 장 이용환 (인)

위 원 이진희 (인)

MagNet: the protein-protein interaction network of the rice blast fungus *Magnaporthe oryzae*

UNDER THE DIRECTION OF DR. YONG-HWAN LEE
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF
SEOUL NATIONAL UNIVERSITY

BY
HYUNBIN KIM

MAJOR IN FUNGAL BIOINFORMATICS
INTERDISCIPLINARY PROGRAM IN AGRICULTURAL GENOMICS
THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

JULY, 2017

APPROVED AS A QUALIFIED THESIS OF HYUNBIN KIM
FOR THE DEGREE OF MASTER OF SCIENCE
BY THE COMMITTEE MEMBERS

CHAIRMAN — Doil Choi, Ph.D.

VICE-CHAIRMAN — Yong-Hwan Lee, Ph.D.

MEMBER — Jin-Hoe Huh, Ph.D.

MagNet: The Protein-Protein Interaction Network of the Rice Blast Fungus *Magnaporthe oryzae*

HYUNBIN KIM

INTERDISCIPLINARY PROGRAM IN AGRICULTURAL GENOMICS

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

ABSTRACT

Magnaporthe oryzae, the rice blast fungus, plays a role as a model organism in the area of molecular plant-microbe interaction research, and its pathogenic and signaling pathways in appressorium development were well understood. As multi-omics data being available, several genomic-level researches have been conducted to uncover the biological process underlying the pathogenesis of this fungus. Genome-wide research of protein-protein interaction (PPI) network is one of the useful research methods by which can lead to better understandings of signaling and regulatory pathways, but existing biological network resources of *M. oryzae* were not sufficient for researches on fungal plant pathology. In this study, PPI network analysis platform of *M. oryzae*, MagNet was

constructed with three methods: homology-based ‘Interolog’ search, co-expression network construction, and domain-domain interaction based prediction. Interologs within proteins which have orthologs in model species were predicted with six PPI repositories. Co-expression networks were built with RNA sequencing data from the infection stage and those from the vegetative stage. In addition, we used the information of interacting domain pairs to extract 3,121,109 interactions within 6,226 proteins. With three approaches all together, the pan-network with 5,600,976 interactions was generated including highly confident 215,731 interactions found in >3 subnetworks. Experimental results on *M. oryzae* PPIs demonstrate that our highly confident PPI network can predict PPIs with higher sensitivity (89.65%) and specificity (78.57%) compared with the previously constructed databases. MagNet provides integrated biological network data which can help to understand the pathogenic mechanisms of plant fungal pathogens.

Keywords**:** protein-protein interaction (PPI), *Magnaporthe oryzae*, interolog, domain-domain interaction (DDI), co-expression network

Student number: 2015-23148

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

PPI	Protein-protein interaction
DDI	Domain-domain interaction
MAPK	Mitogen-activated protein kinase
GO	Gene Ontology
PCC	Pearson correlation coefficient

INTRODUCTION

Proteins are social molecules. Since a biological process is not arbitrarily driven by a single protein, understanding the total protein-protein interactions (PPIs) in a cell has become an important challenge in the post-genomic era. Network-based research is needed to find out what is happening in a cell beyond identifying what is present in the cell. However, mapping the total interaction of proteins in an organism that is known as an interactome is a labor-intensive and time-consuming work. Total possible combinations of binary interaction of proteins are hundreds of millions and it is almost impossible to check the entire interactions with experimental methods such as yeast-two-hybrid or bimolecular fluorescence complementation. As an alternative, computational prediction methods are widely adopted in construction of genome-wide PPI networks.

For computational prediction of interactome, various types of genomic data such as evolutionary relationship, 3D protein structure, genomic position, domain, and primary protein structure have been used in PPI prediction (Pitre *et al.,* 2008). Among these methods, searching interaction-ortholog (interolog) is one of the general *in silico* methods to predict unrevealed PPI network. The presumption applied to this method is that if a pair of proteins in one organism interact with each other, the interaction between the homologous pair of the proteins in another organism is likely to be conserved due to the similarity of protein properties. STRING (Search Tool for Recurring Instances of

Neighbouring Genes) (Szklarczyk *et al.,* 2015) is a well-known database of predicted PPI constructed with interologs search methods. The other majorly used method in PPI prediction is the identification of domain-domain interactions (DDIs). The DDIs data can provide useful clues for deciphering PPI and databases including DOMINE (Yellaboina *et al.,* 2011) and IDDI (Integrated Domain-Domain Interaction anaylsis system) (Kim *et al.,* 2012) were constructed for utilization of DDI information.

In plant-microbe interaction, PPIs play a key role in both the process by which plants recognize the surface of pathogens and the process by which pathogens modulate the plant defense response (Staskawicz *et al.,* 2001). Due to the importance of PPIs in plant pathogenesis, PPI databases have already been created with the proteomes of plant pathogens including *Fusarium graminearum* (Zhao *et al.,* 2009), *Magnaporthe oryzae* (He *et al.,* 2008) and *Ustilaginoidea virens* (Zhang *et al.,* 2017). In wheat head blight fungus *F. graminearum*, the prediction of pathogenic genes was conducted by analysis of sub-network of PPI network which built upon pathogenic seed genes (Lysenko *et al.,* 2013).

M. oryzae is a fungal plant pathogen which causes blast disease on monocot crops. Rice is the major host of this fungi, and it causes severe loss of rice harvest. Not only can the fungus damage rice production, but also is able to infect the heads of wheat raising wheat-blast outbreaks globally (Sadat and Choi, 2017). In addition to the severity of the rice blast disease, the availability to culture and carry out genetic experiments makes *M. oryzae* a model phytopathogen for understanding molecular plant-microbe interactions.

M. oryzae penetrates the host cell with a dome-shaped specialized infection

structure called an appressorium. Appressorium development is induced by surface sensing mechanisms and the melanization of appressorium is essential to create enough turgor pressure for penetration of plant cell (Hamer and Talbot, 1998). Genetic researches revealed core components of signaling pathways that connect mechanosensing system to appressorium differentiation with regard to cell wall remodeling and an increase of solute concentration, suggesting that the protein-protein interactions (PPIs) in cAMP-dependent pathway and MAPK cascade signaling are known to play key roles in the appressorium formation (Ebbole, 2007). As the genomic sequence of *M. oryzae* is available (Dean *et al.,* 2005), researches on the pathogenesis of the fungus are underway at multi-omics levels (Franck *et al.*, 2015; Gokce *et al.*, 2012; Kim and Lee, 2012) as well as functional genetics approach through targeted gene deletion and disruption. The whole set of PPIs is also one of the major research target for omics level studies, and two predicted databases, MPID and STRING have been created for *M. oryzae* PPIs. However, both databases were constructed only with the methods for searching interologs, which makes it difficult to find *M. oryzae* species-specific interactions.

Here, I provide improved PPI network of *M. oryzae*. I made an integrative prediction using multiple databases and multiple methods in an integrative way, and constructed a high-confidence network of *M. oryzae* genes, MagNet. I also verified the network with computational simulations and experimental evidences from previous researches. With MagNet, I obtained systemic understanding through clustering the network and overlapping of existing data.

MATERIALS AND METHODS

Sequences and functional annotation data for *Magnaporthe oryzae*

Protein sequences and functional annotation data of *M. oryzae* were downloaded from *Magnaporthe* comparative database (*Magnaporthe* comparative Sequencing Project, Broad Institute of Harvard and MIT (http://www.broadinstitute.org/)). Gene Ontology (GO) annotation is a way of functional annotation with standardized term, and GO annotation of *M. oryzae* proteins were downloaded from Gene Ontology Consortium (Meng *et al.,* 2009).

Construction of the PPI network

Integrated gene network was built through three methods: Homology-based protein interaction prediction, co-expression, and domain-domain interaction (DDI) based prediction (Figure 1). The integrated network was named "MagNet" and visualization of the network was conducted with Cytoscape v3.4.0 (Shannon *et al.,* 2003). Parameters of the constructed network were calculated with NetworkAnalyzer plugin (Assenov *et al.,* 2008) included in Cytoscape.

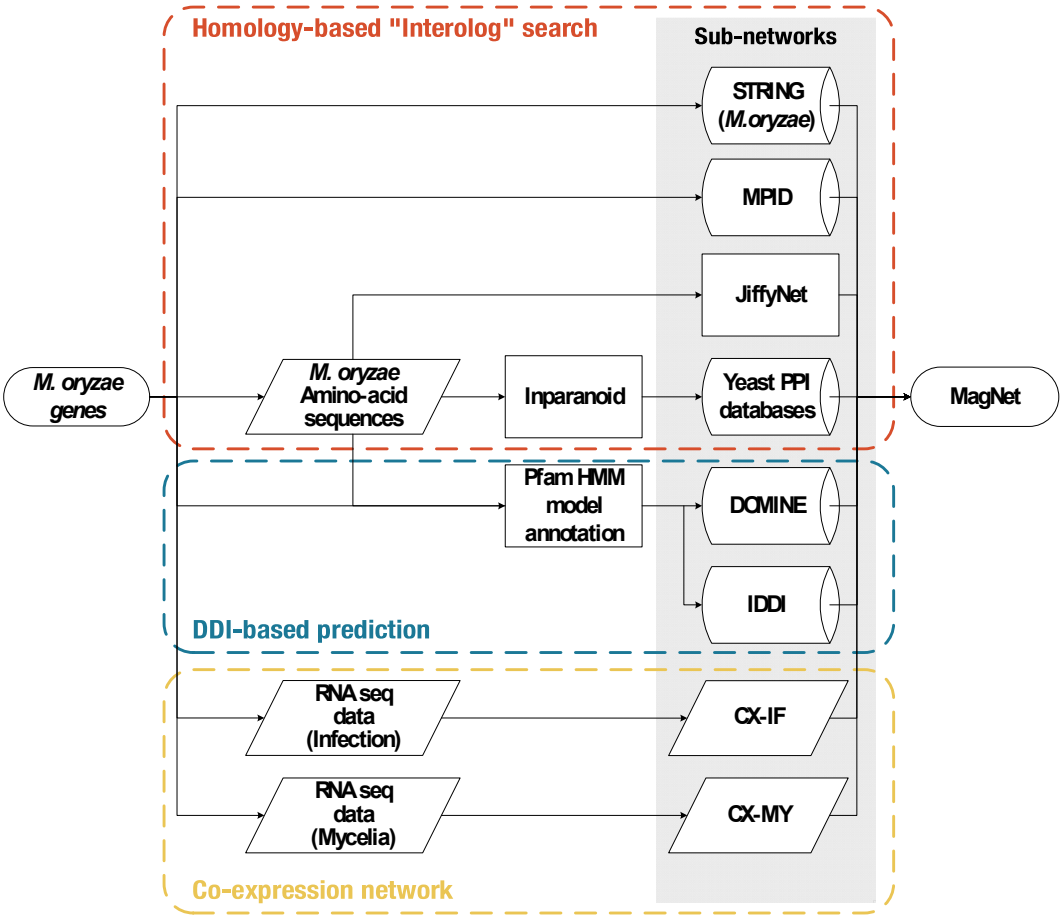

Figure 1. Schematic overview of construction of multiplex network of *M. oryzae*. Three main methods were used to build MagNet database; Homology-based "Interolog" search, DDI-based prediction, and co-expression network. In "Interolog" finding, *M. oryzae* homologs of interacting protein pairs from fungal PPI databases were found with Inparanoid 4.0. BioGRID, DIP, MINT, PINA, and INTACT were used as PPI repositories and the interacting proteins are mainly from *Saccharomyces cerevisiae*. *M.oryzae* proteins were uploaded to JiffyNet and predicted interaction network was generated based on

template networks of model organisms; *Escherichia coli, S. cerevisiae, Caenorhabditis elegans, Homo sapiens, Arabidopsis thaliana and Oryza sativa.* Pfam IDs of *M. oryzae* proteome were downloaded from UniProt, and domain interaction data of DOMINE and IDDI were applied to predict PPIs by DDI-based prediction. I used in-house and public RNAseq expression data of *M. oryzae* to build co-expression network, and two network was generated based on time-series expression data during the infection stage and expression data during the vegetative growth. Integrated network was named as MagNet, and validated with literature evidences and computational simulations.

Homology-based prediction

"Interolog" is a putative interaction between two proteins whose homologs are proved to interact each other in other species. Databases containing fungal PPIs which are experimentally validated were used to find interologs in *M. oryzae*; DIP (Xenarios *et al.,* 2002), MINT (Licata *et al.,* 2012), PINA (Cowley *et al.,* 2012), INTACT (Kerrien *et al.,* 2007), and BIOGRID (Chatr-Aryamontri *et al.,* 2015). While fungal PPI data of BioGRID derives from five fungal species; *Aspergillus nidulans*, *Candida albicans*, *Neurospora crassa, S. cerevisiae*, *Schizosaccharomyces pombe*, and *Ustilago maydis*, PPIs in the other databases comes from *S. cerevisiae*. Fungal proteins of six species were imported to Inparanoid4.1 (Remm *et al.,* 2001), and their orthologs of *M. oryzae* were found. Among the potential orthologs predicted by InParanoid, only the protein of which score is 1.000 were applied to find interologs.

JiffyNet (Kim *et al.,* 2013) is a web-based tool for construction of homology-based PPI network which founds interologs of uploaded protein sequences. JiffyNet uses orthologs of uploaded proteins in model organisms with Inparanoid algorithm and identifies interactions of query proteins based on six template networks; EcoliNet (Kim *et al.,* 2015), YeastNet (Kim *et al.,* 2014), WormNet (Cho *et al.,* 2014), HumanNet (Hsu *et al.,* 2011), AraNet (Lee *et al.*, 2015a) and RiceNet (Lee *et al.*, 2015b). Proteome of *M. oryzae* genome version 8 was uploaded to JiffyNet web server and interologs were found.

Domain-domain interaction

Using the information of *M. oryzae* proteome stored in UniProt, I assigned Pfam domain to *M. oryzae* proteins and verified it with the Interproscan version 58 (Quevillon *et al.,* 2005). DDI information was downloaded from two databases: DOMINE and IDDI. DOMINE contains DDIs observed in PDB entry and predicted DDIs, and was used in PPI prediction of *N. crassa* (Wang *et al.,* 2011). DOMINE is composed of two DDI databases and 13 predicted DDI databases, and IDDI added 10 databases including databases used in DOMINE. DOMINE contains 26,219 DDIs in 5,140 Pfam domains, and IDDI contains 204,705 DDIs in 7,351 Pfam domains.

Co-expression network

If the expression patterns of the two proteins are similar, the two proteins are likely to interact with each other because they tend to co-exist at the same time as they are controlled by the same regulation mechanism. The average degree of co-expression between protein pairs has been used as a measure of PPI network accuracy. However, in this research, co-expression was used as a prediction method for finding putative interacting protein pairs. Based on the disease cycle of the fungi, two co-expression networks were generated; one network based on the expression profile of the infection stage *in planta* RNA-Seq and the other with the expression profile during the vegetative growth stage. For the co-expression network of the infection stage, expression profiles of KJ201 (in-house data) and 98-06 (Dong *et al.,* 2015) were used and each profile consists of five

time points. RNA-Seq data during mycelial growth, which was used as control during the experiment, was downloaded from the Gene Expression Omnibus (GSM752000, GSM1072034, GSM1375971-3 and GSE51597) and another co-expression network was constructed with the data downloaded.

Co-expression network of each stage was constructed with ExpressionCorrelation plugin in Cytoscape. Pearson correlation coefficient (PCC) was used as a measure for co-expression, and PCC cutoff was set to 0.903 for the infection stage co-expression network and 0.846 for the vegetative stage co-expression network corresponding to p-value 0.001. Pairs with negative PCC were discarded.

Validation of Predicted PPIs

To assess the confidence of the predicted PPI network, 97 PPI experimental evidences of *M. oryzae* were collected from 17 references (Chen *et al.,* 2008; Choi *et al.,* 2009; Cui *et al.,* 2015; Ding *et al.,* 2010; Jacob *et al.,* 2015; Kulkarni and Dean, 2004; Li *et al.,* 2010; Liu *et al.,* 2010; Mehrabi *et al.,* 2008; Park *et al.,* 2006; Qi *et al.,* 2012; Yin *et al.,* 2016; Zhang *et al.,* 2011; Zhang *et al.,* 2014; Zhao *et al.,* 2005; Zhou *et al.,* 2012; Zhou *et al.,* 2011). Results of yeast two-hybrid, affinity purification, bimolecular florescence complementation, and co-immunoprecipitation were compared with predicted PPI databases constructed in previous researches; MPID and STRING, and MagNet from this research.

Addition to literature-based validation, assessment of MagNet was conducted via

two computational experiments. Firstly, confidence of each method used to construct MagNet was tested with the intersection of MPID and STRING. Secondly, shared GO terms of proteins were identified. As proteins of similar biological functions are likely to interact, interacting protein pairs tend to share more specific biological process involved or molecular function than non-interacting pairs. I compared shared GO term depths of protein pairs in MagNet with those of pairs in random network.

Information of interaction partners

I used gProfilerR (Reimand *et al.,* 2007) to determine which GO terms were enriched in the putative interaction partners of each gene. As an enrichment test parameter, the p-value cut was set to 0.05 and corrected through the FDR correction (Meng *et al.,* 2009). In addition, WolfPSort (Horton *et al.,* 2007) was used to predict the subcellular localization information of the interaction partners.

Identification of hubs and clusters

I found densely connected regions in the high-confidence network with ClusterONE algorithm (Nepusz *et al.,* 2012). Only the clusters identified with p-value smaller than 0.05 were considered as clusters. Each cluster was named based on the WordCloud (Oesper *et al.,* 2011) summary of *M. oryzae* gene description and GO annotations, and GO enrichment analysis of the cluster was conducted with ClueGO (Bindea *et al.,* 2009) plugin working in Cytoscape. I also identified hub genes in the high-

confidence network with cytoHubba plugin in Cytoscape (Chin *et al.,* 2014). CytoHubba calculates 11 parameters of nodes in given network, and I identified hub lists based on the parameters.

Sub-network analysis

Known pathogenic gene list of *M. oryzae* was downloaded from PHI-base (Urban *et al.,* 2015). The sub-network of pathogenic genes was found, and two co-expression networks were compared in the pathogenic gene sub-network. Additionally, a sub-network was extracted with the list of up-regulated genes during appressorium development (Soanes *et al.,* 2012).

MoHox2, one of the transcription factors which possesses a homeobox domain and is indispensable for conidiation, was chosen as a target for additional sub-network analysis. I extracted a sub-network of genes predicted to interact with MoHox2 while differentially expressed in *ΔMohox2* (Kim and Lee, 2012). I also compared the interaction partners of MoHox2, MoHox7, and MoHox8, which are known to regulate distinct developmental processes (Kim *et al.,* 2009).

RESULTS

Statistics of the integrated network

In order to increase the coverage of the PPI network compared to existing networks and extract the high-confidence network, I merged the networks generated by different methods. A total of 826,128 interactions were predicted as interologs, 3,121,109 interactions as DDIs, and 1,967,109 interactions as co-expression. Interologs and DDIs covered 7,138 and 6,226 proteins respectively, and co-expression network covered 10,759 proteins. The integrated network with a total of 5,600,976 interactions was generated with three main methods described above, and covered total 11,734 proteins. The property of each subnetwork in MagNet was calculated with Cytoscape plugin NetworkAnalyzer. (Table 1)

Table 1. Network statistics of MagNet sub-networks

Method	Network	Proteins	Interactions	Avg. num of neighbors	Clustering coefficient	Network centralization	Connected components	Network diameter	Network density
Homology based network	STRING	6,498	1,097,966	168.97	0.309	0.445	71	9	0.026
	MPID	3,017	11,674	7.234	0.105	0.041	108	9	0.002
	BioGRID	3,503	166,492	61.961	0.207	0.375	4	5	0.018
	DIP	2,344	10,110	8.438	0.130	0.054	28	9	0.004
	MINT	2,542	23,868	8.094	0.127	0.045	30	10	0.003
	PINA	2,849	38,336	26.688	0.251	0.493	2	6	0.009
	INTACT	2,825	31,756	21.848	0.239	0.543	5	6	0.008
	JiffyNet	4,690	266,409	113.607	0.303	0.184	2	7	0.024
Co-expression network	Vegetative stage	10,356	1,173,767	226.683	0.546	0.075	6	15	0.022
	Infection stage	8,131	819,480	201.569	0.539	0.071	5	11	0.025
Domain interaction network	DOMINE	5,727	1,450,218	372.438	0.663	0.412	182	10	0.065
	IDDI	6,231	4,247,425	991.802	0.662	0.514	51	6	0.159
High-confidence	MagNet	6,005	217,531	71.936	0.324	0.223	57	11	0.012
Total	MagNet	11,734	5,600,976						

The property of each subnetwork was calculated with NetworkAnalyzer of Cytoscape.

High-confidence network extracted from the integrated network

Since it is important to identify the putative interaction partners subjected to the experiment, I have extracted high-confidence interactions among all the predicted interactions (Figure 2). The high-confidence interaction was determined with the cutoff of three or more interaction evidences because the average number of interaction evidences were 3.88 in true positives and 2.54 in false positives in validation with available experimental evidences. The high-confidence network covered 217,531 interactions between 6,005 *M. oryzae* genes and the interacting proteins in the high-confidence network had 72 interacting partners on average.

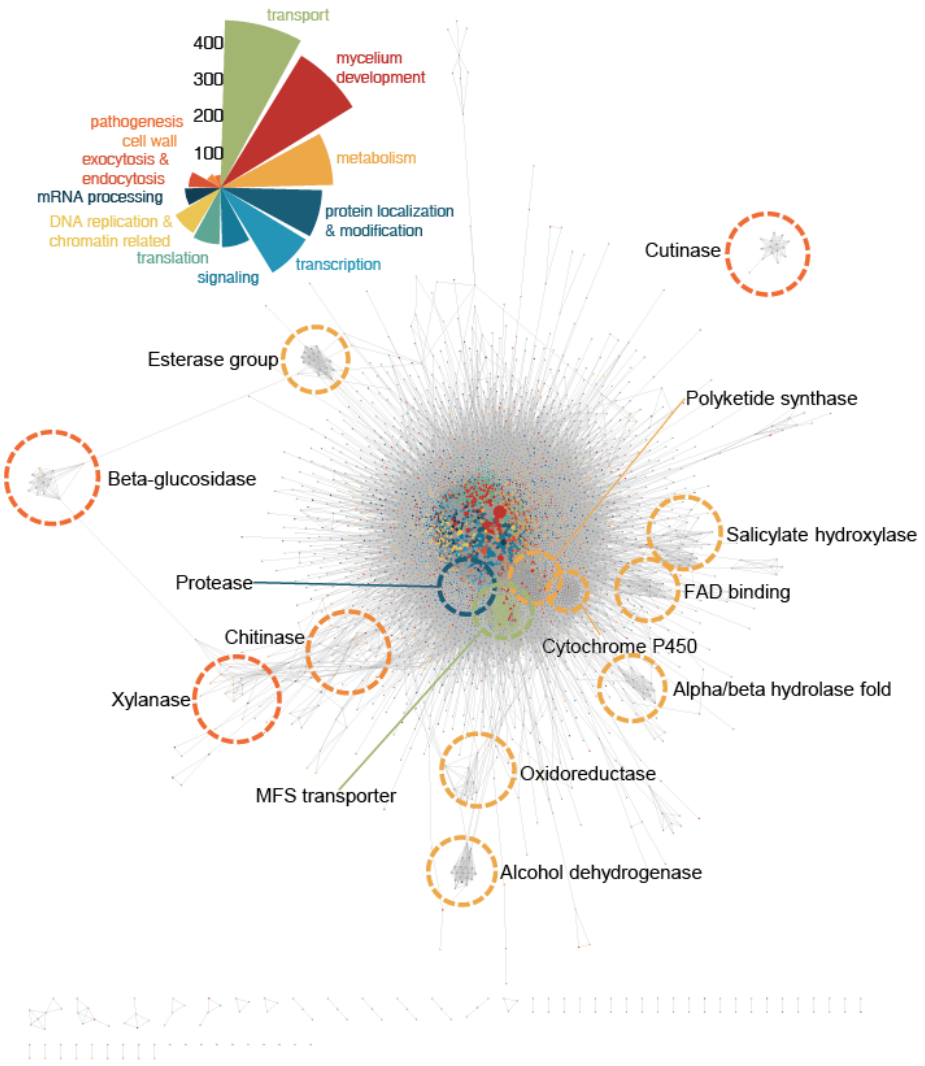

Figure 2. Global view of high-confidence network of *M. oryzae* proteins. Interactions of MagNet predicted in three or more resources were grouped as a high-confidence network. High-confidence network is visualized with Cytoscape 3.4.0. Each circle represents a protein and the lines connecting circles mean interactions between proteins. Proteins were colored based on the biological processes predicted by assigned GO terms. Sizes of circles

were determined according to the number of connected edges. ClusterONE algorithm was adopted to find edge-dense regions in the network. Each edge-dense clusters were named with Wordcloud result of assigned GO terms and descriptions of clustered genes. MFS transporter, Cytochrom P450, and other enzymes were found to be clustered within the network.

Topologically, MagNet is similar to known biological networks with the power-law degree distribution (Figure 3). It is consistent with the results of yeast network in which a small number of high-degree nodes connect many other low-degree nodes (Albert, 2005). Higher average clustering coefficient was observed in MagNet (0.324) than random network (0.012) which is consistent to previously constructed PPI networks (Zhang *et al.,* 2017). Based on the degree of the proteins in the network, twenty most highly connected *M. oryzae* proteins were extracted (Table 2). Hubs consists of proteins that act in post-translational modification like polyubiquitin, heat shock proteins or kinases. The average expression level of hub proteins (6350.8, FPKM) was higher than the average expression level of total proteins in the high confidence network (481, FPKM), which shows that hub proteins are likely to be essential for the organism.

Total 5,600,976 protein interactions are saved in the database table format as a pair of interacting proteins with the presence/absence vector from sub-networks. The subcellular localizations of *M. oryzae* proteins were predicted with WolfPsort (Horton *et al.*, 2007), and the localization profile of the interacting partners of query gene CPKA is shown with the pie graph. (Figure 4)

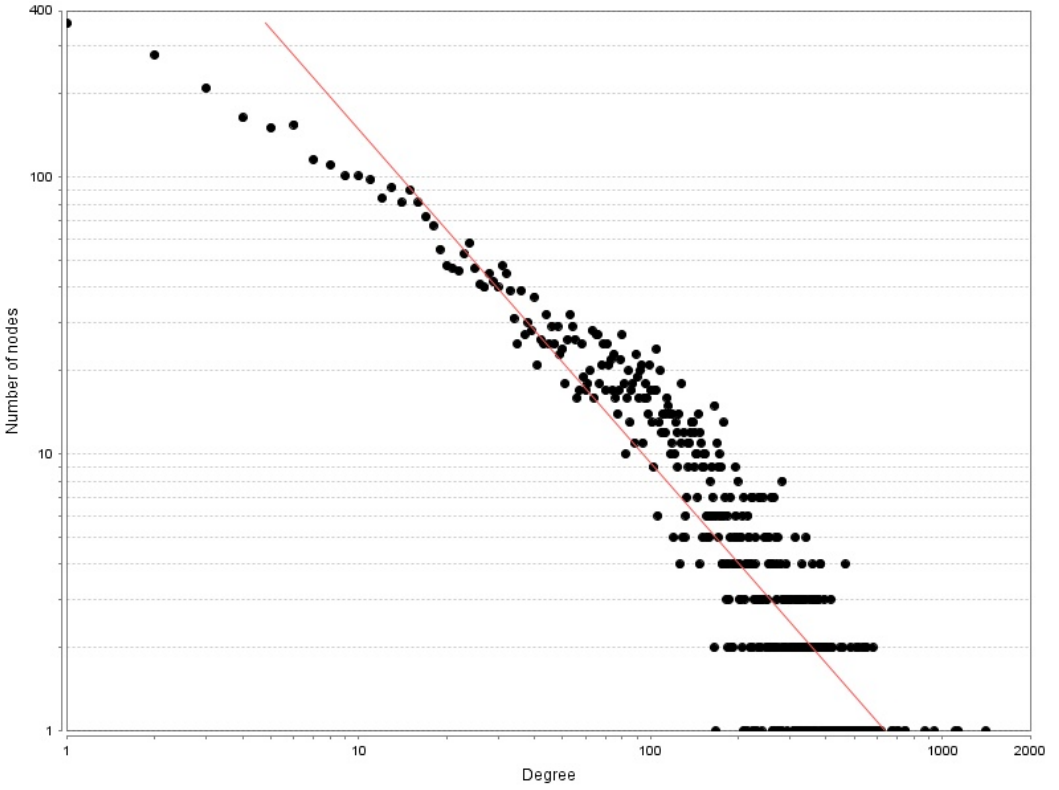

Figure 3. Degree distribution of high-confidence MagNet. The degree distribution of MagNet follows a power law ($y = 2377.1x^{-1.204}$, $r^2 = 0.853$) which is a well-known characteristic of scale-free networks. In the plot, the x-axis represents the degree of the nodes in the high-confidence network on logarithmic scale while the y-axis stands for the number of the nodes.

Table 2. Twenty most highly connected *M. oryzae* protein interaction hubs in the high-confidence network.

Locus	Protein description	Degree
MGG_06958	hsp70-like protein	1408
MGG_01282	polyubiquitin	1137
MGG_11513	heat shock protein SSB1	1108
MGG_01362	CMGC/CDK/CDC2 protein kinase	939
MGG_06759	heat shock protein 90	872
MGG_04660	CMGC/CDK/CDK5 protein kinase	749
MGG_06044	ubiquitin-60S ribosomal protein L40	741
MGG_07928	ubiquitin-40S ribosomal protein S27a	708
MGG_01656	UV excision repair protein Rad23	697
MGG_04719	guanine nucleotide-binding protein subunit beta-like protein	679
MGG_01318	deubiquitination-protection protein dph1	669
MGG_09960	PLK protein kinase	616
MGG_04905	CMGC/CDK protein kinase	608
MGG_06320	STE/STE20/PAKA protein kinase	600
MGG_04943	CMGC/MAPK protein kinase	599
MGG_14620	CAMK/CAMKL protein kinase	591
MGG_14773	AGC/AKT protein kinase	582
MGG_00803	CAMK/CAMKL/AMPK protein kinase	580
MGG_02503	glucose-regulated protein	576
MGG_09887	NEDD8	576

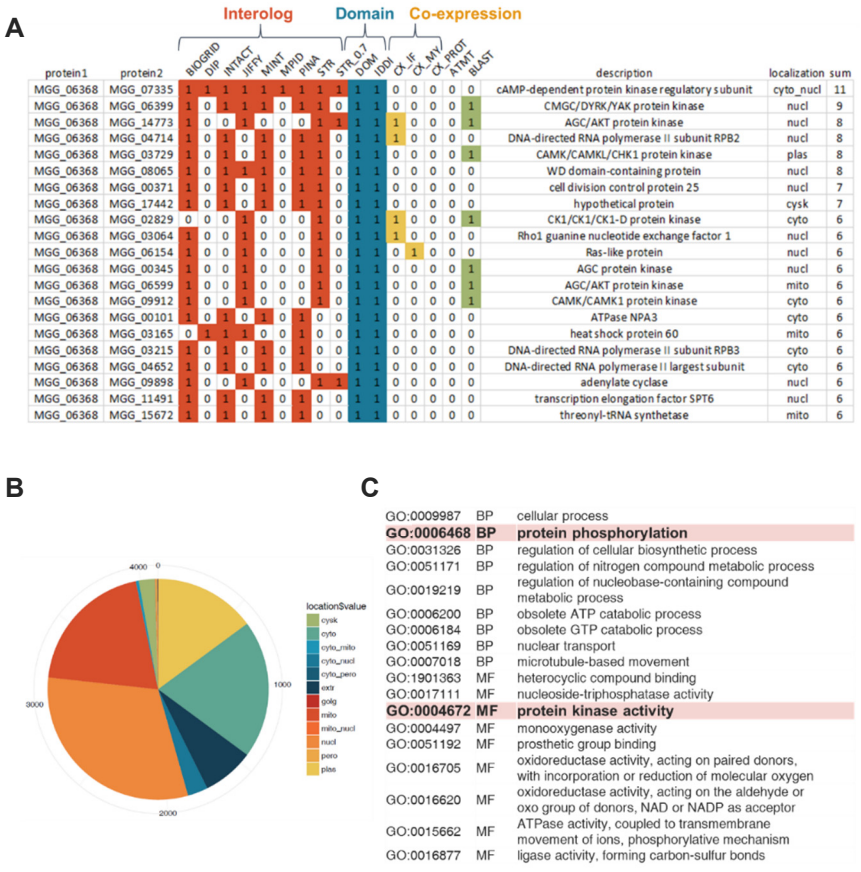

Figure 4. Interaction partners of CPKA (MGG_06368) predicted in MagNet. (A) CPKA gene encodes the catalytic subunit of cAMP-dependent protein kinase A, and the interaction with MGG_07335, the regulatory subunit of PKA, was predicted in 11 databases out of 13 resources, which is the most. **(B)** The pie chart at the left bottom shows the subcellular localization information of CPKA interaction partners. **(C)** The table at the right bottom is the most enriched GO terms in interaction partners of CPKA, which includes “protein phosphorylation” in the biological process and “protein kinase activity” in the molecular function.

Validation with experimental evidences from literatures

The experimental evidences showed that the high-confidence network of MagNet is more precise than the predictions of MPID and STRING (Table 3). The integrated MagNet had the most number of true positives in the tested networks. In the 43 true positives of MagNet, 8 interactions were predicted only in MagNet; 6 interactions from IDDI and 2 interactions from the infection stage co-expression network. MagNet specific false positives were also identified and five predictions were from IDDI and one prediction from JiffyNet.

Validation with computational methods

The accuracy of each prediction method used in MagNet was compared using the intersection of MPID and STRING. (Figure 5a) As a result, BioGRID and IDDI showed good classification results with the intersection of previously created databases, but co-expression, JiffyNet and DOMINE did not fit to the intersection. However, the combination of several methods showed more accurate classification results. As interacting proteins are likely to participate in similar biological process, they tend to share more specific biological terms. The depth of shared GO terms in protein pairs predicted with MagNet was compared to those of randomly generated pairs. The result showed that pairs in MagNet tend to share more specific GO terms than randomly generated pairs (Figure 5b).

Table 3. Validation with literature evidences

Parameter	MPID	STRING	MagNet	MagNet (HC)*
True positve (No.)	7	34	43	26
True negative (No.)	14	9	3	11
False positive (No.)	0	5	11	3
False negative (No.)	76	49	40	57
Precision	1.000	0.872	0.796	0.897
Accuracy	0.216	0.443	0.474	0.381
F1 Score	0.156	0.557	0.628	0.464

*MagNet (HC): High-confidence subnetwork of MagNet which has been predicted with three or more evidences

No false positives were found in MPID, which appears to be due to the small interaction volume of MPID. Integrated MagNet had the most number of true positives and false positives, which led to the highest accuracy and F1 score.

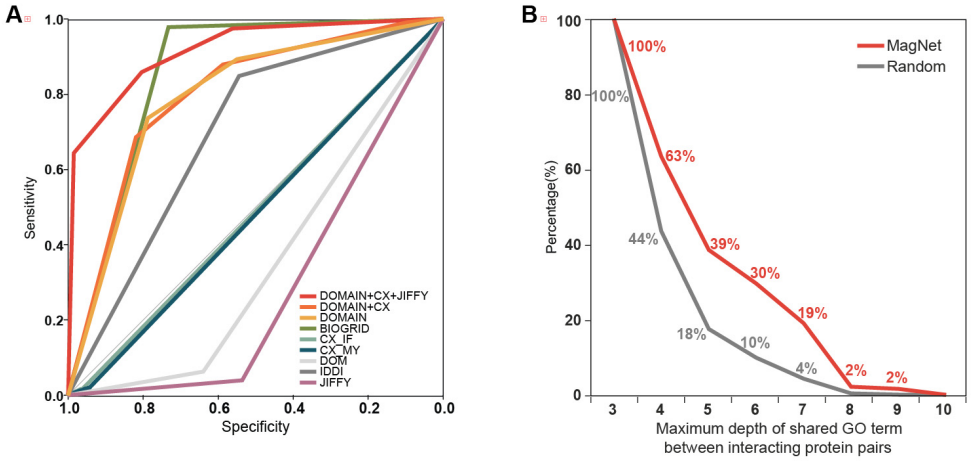

Figure 5. Computational validation result of MagNet. (A) Receiver operating curve of MagNet resources tested with intersection set of MPID and STRING shows the performance of each predicteion method. **(B)** Validation of the predicted PPI network based on shared GO term depth. Pairs in MagNet tend to share more specific terms compared to randomly generated pairs.

I compared the edge coverage of MPID, STRING and MagNet in the known pathways. The number of interactions are 11,674 in MPID, 1,097,966 in STRING, 5,600,976 in MagNet, and 217,531 in MagNet with three or more evidences (MagNet (HC)). Although MagNet (HC) contains fewer interactions than STRING, and the numbers of interactions found in MAPK pathway are similar between STRING (55 interactions) and MagNet (HC) (56 interactions). (Figure 6) Similar results were also confirmed within the interactions in septin complex.

Co-expression network comparison

Stage-specific co-expressed proteins were identified by comparing two co-expression networks; 2,502 proteins for vegetative stage, and 696 proteins for infection stage. Enriched GO terms of vegetative stage-specific co-expressed proteins are “Cellular macromolecule metabolic process”, “Cellular protein metabolic process”, “Cellular component organization at cellular level”, “Mycelium development” and “Translation”. On the other hand, two terms, “Interaction with host via substance released outside of symbiont” and “Response to cAMP”, were enriched in the infection stage co-expressed proteins. Clustering of co-expression network showed that distinct sets of proteins are co-expressed during the infection stage and the vegetative stage of *M. oryzae*. As the result of clustering co-expression network with ClusterONE algorithm, I identified 16 clusters of significant p-value in infection stage co-expression network, and 37 clusters in vegetative stage co-

expression network. GO terms which are enriched in the clusters were assigned to 10 out of 16 infection stage co-expressed clusters and 11 out of 37 vegetative stage co-expressed clusters with gProfileR. Co-expression clusters during vegetative growth contain clusters involved in rRNA processing, oxidation-reduction process, multicellular organism/ mycelium development, transmembrane transport and DNA replication. On the other hand, co-expression clusters during infection process include clusters on translation, protein glycosylation, protein phosphorylation and single-organism developmental process.

Pathogenic gene network

Genes in PHI-base were grouped based on the phenotype pattern, and genes related to the phenotype of “loss of pathogenicity” were likely to have higher degrees in the network compared to genes related to phenotype of “unaffected pathogenicity” (t-test, p-value 0.03). However, there was no statistically significant difference between the genes with "loss of pathogenicity" and genes with "reduced pathogenicity" and between "reduced pathogenicity" and "unaffected pathogenicity".

To confirm that the infection co-expression network is related to the pathogenesis, two co-expression networks were compared within the pathogenic gene network (Figure 7). The result indicated that the pathogenic genes are co-expressed during the infection process rather than during the vegetative growth.

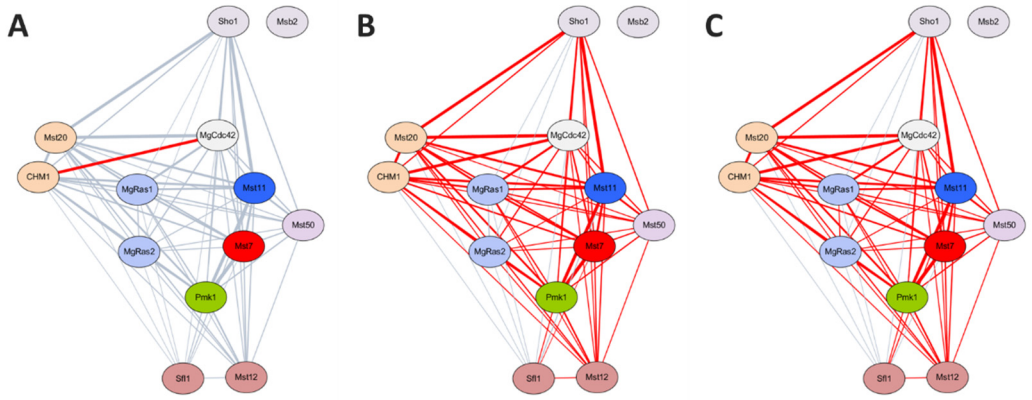

Figure 6. PPIs in MAPK pathway. Red lines indicate the interactions predicted in **(A)** MPID, **(B)** STRING and **(C)** MagNet with three or more evidences (MagNet (HC)). Total edges are more in STRING than MagNet (HC) (STRING: 1097966, MagNet (HC): 217531). However, the edge coverages of two networks are similar, which means MagNet (HC) may have less false positives than STRING.

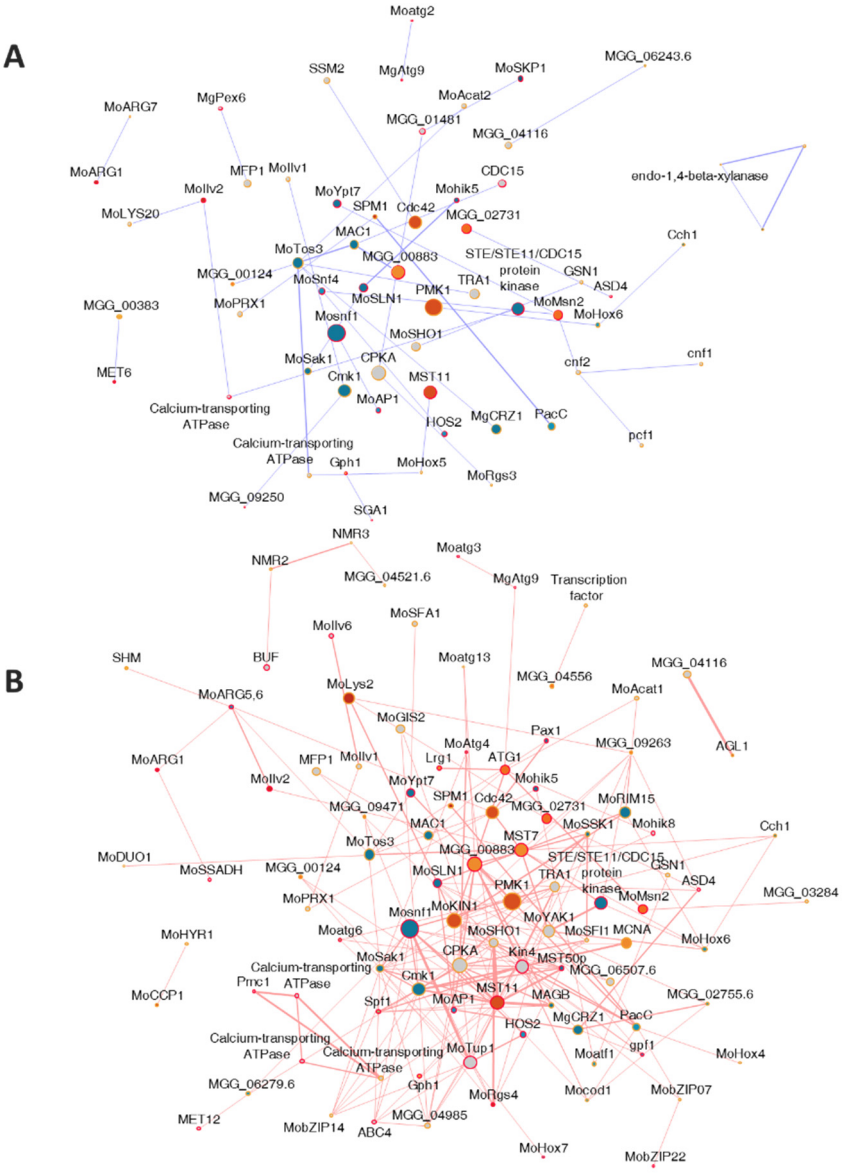

Figure 7. Co-expression networks in the pathogenic gene network. (A) Co-expression during vegetative growth was colored blue in the network figure and **(B)** the co-expression infection process was colored red. The red lines connect more pathogenic genes and generate denser connection than blue lines do.

Differentially expressed network

To identify differentially expressed gene interaction complex during the pathogenesis by this fungi, the sub-network of differentially expressed genes (DEGs) during appressorium development was extracted (Figure 8). Compared to the clusters identified in the high-confidence MagNet, clusters acting on signal transduction, melanin biosynthesis, MFS transporter and cytochrome P450 were up-regulated during appressorium development.

As *ΔMohox2* cannot generate conidia, the DEGs in *ΔMohox2* may be regulated by MoHOX2 and be associated with conidial development. To find protein components which regulate conidiation, the intersection of DEGs in *ΔMohox2* and interaction partners of MoHOX2 was found (Figure 9). Of the partner DEGs, genes with kinase activity were repressed by MoHOX2.

I also compared the putative interaction partners of MoHOX2, MoHOX7 and MoHOX8 because the deletion mutants of these HOX genes showed phenotypes without only a specific developmental process; *ΔMohox2* without conidiation, *ΔMohox7* without appressorium development and *ΔMohox8* without invasive growth in infection process (Figure 10). The intersection of putative interaction partners of three HOX genes includes genes that regulate cellular processes, while many carbohydrate transporters belong to MoHOX7 specific partners which are presumed to be participate in appressorium development.

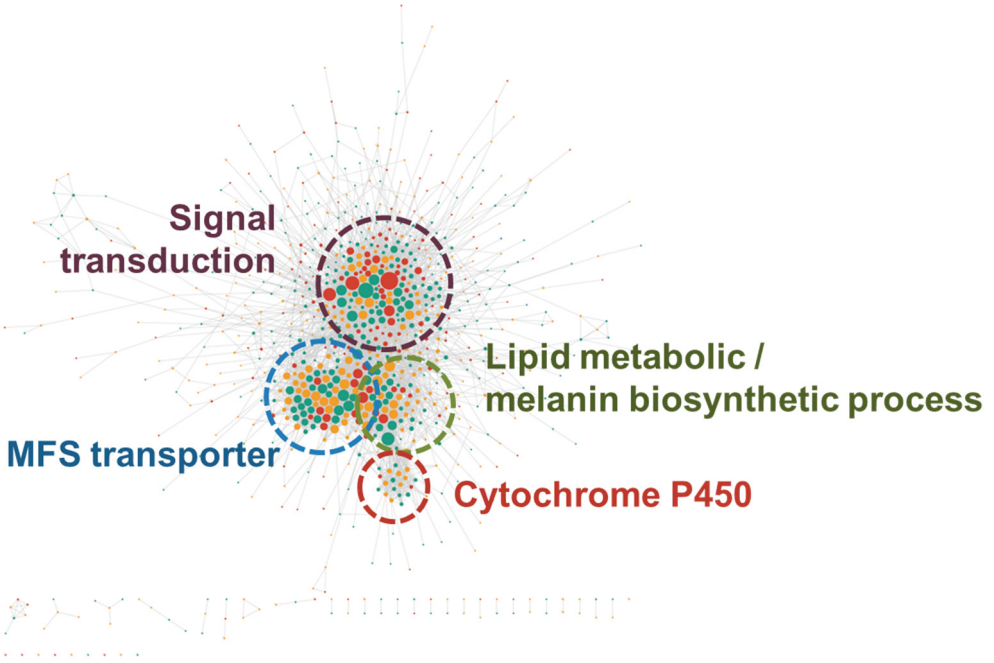

Figure 8. Differentially up-regulated gene network during appressorium development. The red circles stands for the DEGs at 4-8 hpi, green for the DEGs at 14-16 hpi and yellow for the DEGs at both time points. Clusters were found in the sub-network of DEGs and clusters working in signal transduction, melanin biosynthetic process, cytochrome P450 and MFS transporter remained.

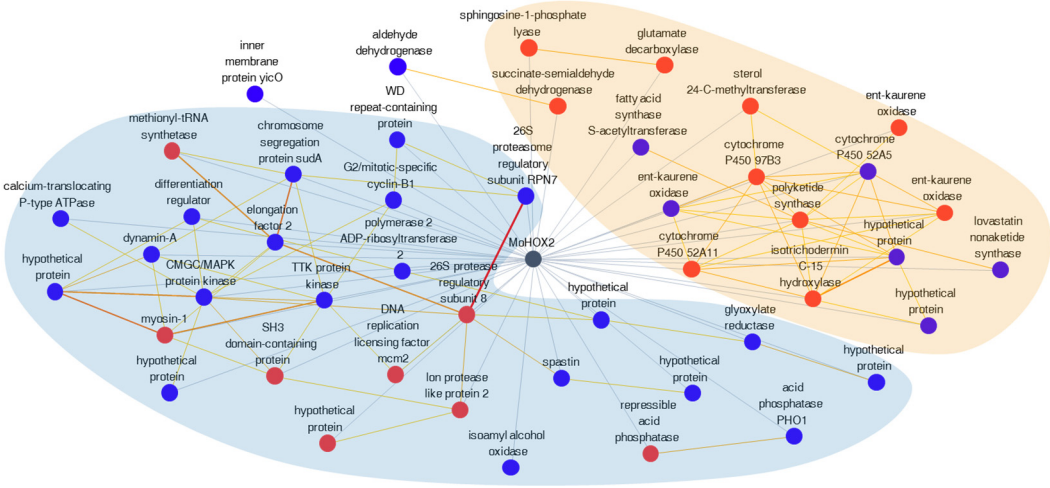

Figure 9. Intersection of differentially expressed genes in *ΔMohox2* and MoHox2 interaction partners. Genes repressed in wild type conidiation and induced in *ΔMohox2* are colored blue and the genes with opposite expression pattern are colored red. The interactions between proteins of same expression pattern are more than those between proteins of different expression patterns, which reveal protein complex regulated by MoHOX2.

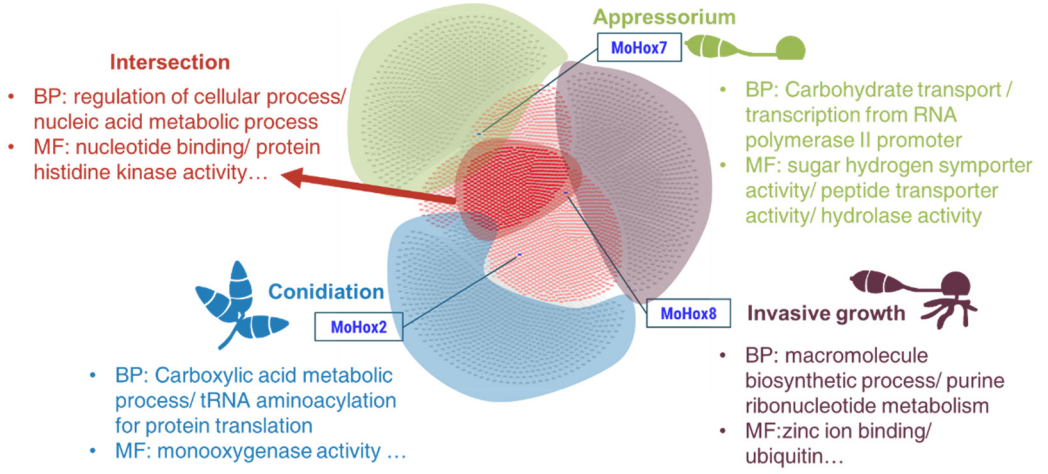

Figure 10. Interaction partners of *M. oryzae* HOX proteins. The interaction partners of MoHOX2, MoHOX7 and MoHOX8 were compared. The proteins which interact with all 3 genes were shown to be involved in the regulation of cellular processes by binding to nucleotides or by kinase activity. The specific interaction partners of each HOX gene may function in the developmental process which the HOX gene regulates.

DISCUSSION

High-confidence PPI prediction in *M. oryzae* with multiple methods

The rice blast fungus is a valuable model for studying the pathogenesis mechanism of plant pathogenic fungi as well as a pathogen that seriously damage crop yields. Protein-protein interaction plays a key role in the developmental differentiation control for the infection of this fungi, so profiling PPIs is essential for establishing the theoretical basis of the disease management. I generated a high-confidence predicted PPI network of *M. oryzae,* MagNet, containing over 200 thousand PPIs. Validation of MagNet was conducted with computational methods and experimental evidences from previous researches, and the result showed high-confidence MagNet has better precision than the other networks. Compared to STRING, MagNet could cover most interactions in known pathways with fewer total edges.

In addition, MagNet has improved the accuracy of prediction by diversifying the prediction method and using the updated databases. Previous studies on fungal plant pathogens have used the degree of co-expression for validating predicted interactions (Zhang *et al.*, 2017; Zhao *et al.*, 2009). This implies that co-expression can be used to predict PPIs. In the validation results with 97 experimental evidences, two interactions were not predicted in any other way but only predicted with co-expression network. Since many fungal gene products are not assigned to domains, co-expression can offer clues of

interactions to proteins without assigned domains.

Among the databases which are not used in MPID and STRING, IDDI used as a DDI database could find more true-positives compared to any other methods with 43 true-positives out of 97 experimental evidences. Sensitivity seems to be improved because of more data amount than DOMINE. BioGRID, which has the highest amount of data in the PPI repositories, found the most number of true-positives among the databases used in the interolog search.

Analysis to extract biological meanings from the network

There is no example of applying the multiple method integrally to the PPI prediction of *M. oryzae*, and I found functionally associated sub-network modules by applying the clustering algorithm to the network constructed in an integrative manner (Ames *et al.*, 2013). With identifying interaction-dense clusters in the network, I could isolate groups of interacting proteins that participate in the same biological processes or that perform together for specific biological functions. This functionally well-clustered network can be utilized for the annotation of unannotated genes.

I also analyzed high-degree genes that act as hubs in the network. As high-degree genes link several pathways, it is known that genes with high degrees tend to be essential genes with lethal phenotypes from yeast network studies (Krogan *et al.*, 2006). In this study, I showed that genes of which the deletion mutants lose pathogenicity have higher degree

in the network than the genes that do not cause phenotypic changes, which is consistent with previous findings that high-degree genes are essential. Previous studies conducted on genomic scale have led to extraction of candidate genes related to specific biological processes. However, if the interaction relationship of each protein is available, it is possible to add interaction information to the simple gene list obtained from genomic research. With this approach, it becomes possible to find an interaction-active hub in a specific condition or to find a pathway linked to the condition.

Clustering analysis was performed on the infection-stage co-expression network. Sixteen clusters were detected, including clusters of biologic processes such as amide biosynthesis, protein phosphorylation, and protein glycosylation, and clusters of oxidoreductase and hydrolase. Infection-stage co-expressed cluster associated with glycosylation includes MoPmt2 (MGG_07190). *ΔMopmt2* exhibits a phenotype with reduced cell wall integrity and conidia germination, leading to reduced pathogenicity (Guo *et al.*, 2016). MoPmt2 was clustered with MgKIN1 (MGG_01279), and MgKIN1 was found to be a kinase that regulates conidia germination (Luo *et al.*, 2014). In addition, MoPmt2 was predicted to interact with MCK1 (MGG_00883), and MCK1 is a kinase that regulates cell wall integrity (Yin *et al.*, 2016). This confirms that the predicted interaction relationship in MagNet is linked to the phenotype association. To confirm whether the predicted interaction in MagNet is actually associated to the phenotype, I found gene pairs that showed same phenotypes among the T-DNA inserted mutants from ATMT database (Jeon *et al.*, 2007) and compared pairs with the predicted interactions generated in MagNet.

In the high-confidence network, 19 gene pairs shared same phenotypes in ATMT database.

Limitation and conclusion

One of the limitation of this research is that there is no prediction scoring based on the confidence of each prediction method. This was because there was no gold standard set to measure the confidence of each method. Therefore, as an alternative, I reflected the confidence of method by increasing the number of databases used in the method. Another limitation is that I focused on PPIs in *M. oryzae*, not on the interactions between *Oryza sativa* proteins and *M. oryzae* proteins. More accurate secretome information of *M. oryzae* and the integration of two high-confidence PPI networks are needed to predict inter-species PPI network.

In this study, PPI network of *M. oryzae*, MagNet, was constructed with three approaches: Homology-based "Interolog" search, Domain-Domain interaction, Co-expressed gene network. Subnetworks of pathogenic genes and DEGs were found from high-confidence PPI network, and can help to understand pathogenic pathways related to appressorium formation and conidiation. The information of MagNet can be used in identification of putative gene functions, construction of hypothesis, and integration of biological information.

References

Albert, R. (2005) Scale-free networks in cell biology. *Journal of Cell Science* **118**: 4947-4957.

Ames, R.M., Macpherson, J.I., Pinney, J.W., Lovell, S.C., Robertson, D.L. (2013) Modular biological function is most effectively captured by combining molecular interaction data types. *PloS One* **8**: e62670.

Assenov, Y., Ramirez, F., Schelhorn, S.E., Lengauer, T., Albrecht, M. (2008) Computing topological parameters of biological networks. *Bioinformatics* **24**: 282-284.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pages, F., Trajanoski, Z., Galon, J. (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**: 1091-1093.

Chatr-Aryamontri, A., Breitkreutz, B.J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., Stark, C., Breitkreutz, A., Kolas, N., O'Donnell, L., Reguly, T., Nixon, J., Ramage, L., Winter, A., Sellam, A., Chang, C., Hirschman, J., Theesfeld, C., Rust, J., Livstone, M.S., Dolinski, K., Tyers, M. (2015) The BioGRID interaction database: 2015 update. *Nucleic Acids Research* **43**: D470-478.

Chen, J., Zheng, W., Zheng, S., Zhang, D., Sang, W., Chen, X., Li, G., Lu, G., Wang, Z. (2008) Rac1 is required for pathogenicity and Chm1-dependent conidiogenesis in rice fungal pathogen *Magnaporthe grisea*. *PLoS Pathogens* **4**: e1000202.

Chin, C.H., Chen, S.H., Wu, H.H., Ho, C.W., Ko, M.T., Lin, C.Y. (2014) cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology* **8 Suppl 4**: S11.

Cho, A., Shin, J., Hwang, S., Kim, C., Shim, H., Kim, H., Lee, I. (2014) WormNet v3: a network-assisted hypothesis-generating server for *Caenorhabditis elegans*. *Nucleic Acids Research* **42**: W76-82.

Choi, J., Kim, Y., Kim, S., Park, J., Lee, Y.H. (2009) *MoCRZ1*, a gene encoding a calcineurin-responsive transcription factor, regulates fungal growth and pathogenicity of *Magnaporthe oryzae*. *Fungal Genetics and Biology : FG & B* **46**: 243-254.

Cowley, M.J., Pinese, M., Kassahn, K.S., Waddell, N., Pearson, J.V., Grimmond, S.M., Biankin, A.V., Hautaniemi, S., Wu, J. (2012) PINA v2.0: mining interactome modules. *Nucleic Acids Research* **40**: D862-865.

Cui, X., Wei, Y., Wang, Y.H., Li, J., Wong, F.L., Zheng, Y.J., Yan, H., Liu, S.S., Liu, J.L., Jia, B.L., Zhang, S.H. (2015) Proteins interacting with mitochondrial ATP-dependent Lon protease (MAP1) in *Magnaporthe oryzae* are involved in rice blast disease. *Molecular Plant Pathology* **16**: 847-859.

Dean, R.A., Talbot, N.J., Ebbole, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., Thon, M., Kulkarni, R., Xu, J.R., Pan, H., Read, N.D., Lee, Y.H., Carbone, I., Brown, D., Oh, Y.Y., Donofrio, N., Jeong, J.S., Soanes, D.M., Djonovic, S., Kolomiets, E., Rehmeyer, C., Li, W., Harding, M., Kim, S., Lebrun, M.H., Bohnert, H., Coughlan, S., Butler, J., Calvo, S., Ma, L.J., Nicol, R., Purcell, S., Nusbaum, C., Galagan, J.E., Birren, B.W. (2005) The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**: 980-986.

Ding, S.L., Liu, W., Iliuk, A., Ribot, C., Vallet, J., Tao, A., Wang, Y., Lebrun, M.H., Xu, J.R. (2010) The tig1 histone deacetylase complex regulates infectious growth in the rice blast fungus *Magnaporthe oryzae*. *The Plant Cell* **22**: 2495-2508.

Dong, Y., Li, Y., Zhao, M., Jing, M., Liu, X., Liu, M., Guo, X., Zhang, X., Chen, Y., Liu, Y., Ye, W., Zhang, H., Wang, Y., Zheng, X., Wang, P., Zhang, Z. (2015) Global genome and transcriptome analyses of *Magnaporthe oryzae* epidemic isolate 98-06 uncover novel effectors and pathogenicity-related genes, revealing gene gain and lose dynamics in genome evolution. *PLoS Pathogens* **11**: e1004801.

Ebbole, D.J. (2007) *Magnaporthe* as a model for understanding host-pathogen interactions. *Annual Review of Phytopathology* **45**: 437-456.

Franck, W.L., Gokce, E., Randall, S.M., Oh, Y., Eyre, A., Muddiman, D.C., Dean, R.A. (2015) Phosphoproteome analysis links protein phosphorylation to cellular remodeling and metabolic adaptation during *Magnaporthe oryzae* appressorium development. *Journal of Proteome Research* **14**: 2408-2424.

Gokce, E., Franck, W.L., Oh, Y., Dean, R.A., Muddiman, D.C. (2012) In-depth analysis of the *Magnaporthe oryzae* conidial proteome. *Journal of Proteome Research* **11**: 5827-5835.

Guo, M., Tan, L., Nie, X., Zhu, X., Pan, Y., Gao, Z. (2016) The Pmt2p-mediated protein O-mannosylation is required for morphogenesis, adhesive properties, cell wall integrity and full virulence of *Magnaporthe oryzae*. *Frontiers in Microbiology* **7**: 630.

Hamer, J.E., Talbot, N.J. (1998) Infection-related development in the rice blast fungus *Magnaporthe grisea*. *Current Opinion in Microbiology* **1**: 693-697.

He, F., Zhang, Y., Chen, H., Zhang, Z., Peng, Y.L. (2008) The prediction of protein-protein

interaction networks in rice blast fungus. *BMC Genomics* **9**: 519.

Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., Nakai, K. (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Research* **35**: W585-587.

Hsu, C.L., Huang, Y.H., Hsu, C.T., Yang, U.C. (2011) Prioritizing disease candidate genes by a gene interconnectedness-based approach. *BMC Genomics* **12 Suppl 3**: S25.

Jacob, S., Foster, A.J., Yemelin, A., Thines, E. (2015) High osmolarity glycerol (HOG) signalling in *Magnaporthe oryzae*: Identification of *MoYPD1* and its role in osmoregulation, fungicide action, and pathogenicity. *Fungal Biology* **119**: 580-594.

Jeon, J., Park, S.Y., Chi, M.H., Choi, J., Park, J., Rho, H.S., Kim, S., Goh, J., Yoo, S., Park, J.Y., Yi, M., Yang, S., Kwon, M.J., Han, S.S., Kim, B.R., Khang, C.H., Park, B., Lim, S.E., Jung, K., Kong, S., Karunakaran, M., Oh, H.S., Kim, H., Kang, S., Choi, W.B., Lee, Y.H. (2007) Genome-wide functional analysis of pathogenicity genes in the rice blast fungus. *Nature Genetics* **39**: 561-565.

Kerrien, S., Alam-Faruque, Y., Aranda, B., Bancarz, I., Bridge, A., Derow, C., Dimmer, E., Feuermann, M., Friedrichsen, A., Huntley, R., Kohler, C., Khadake, J., Leroy, C., Liban, A., Lieftink, C., Montecchi-Palazzi, L., Orchard, S., Risse, J., Robbe, K., Roechert, B., Thorneycroft, D., Zhang, Y., Apweiler, R., Hermjakob, H. (2007) IntAct--open source resource for molecular interaction data. *Nucleic Acids Research* **35**: D561-565.

Kim, E., Kim, H., Lee, I. (2013) JiffyNet: a web-based instant protein network modeler for newly sequenced species. *Nucleic Acids Research* **41**: W192-197.

Kim, H., Shim, J.E., Shin, J., Lee, I. (2015) EcoliNet: a database of cofunctional gene network for *Escherichia coli*. *Database (Oxford)* **2015**: bav001.

Kim, H., Shin, J., Kim, E., Hwang, S., Shim, J.E., Lee, I. (2014) YeastNet v3: a public database of data-specific and integrated functional gene networks for *Saccharomyces cerevisiae*. *Nucleic Acids Research* **42**: D731-736.

Kim, K.S., Lee, Y.H. (2012) Gene expression profiling during conidiation in the rice blast pathogen *Magnaporthe oryzae*. *PloS One* **7**: e43202.

Kim, S., Park, S.Y., Kim, K.S., Rho, H.S., Chi, M.H., Choi, J., Park, J., Kong, S., Goh, J., Lee, Y.H. (2009) Homeobox transcription factors are required for conidiation and appressorium development in the rice blast fungus *Magnaporthe oryzae*. *PLoS Genetics* **5**: e1000757.

Kim, Y., Min, B., Yi, G.S. (2012) IDDI: integrated domain-domain interaction and protein interaction analysis system. *Proteome Science* **10 Suppl 1**: S9.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., Punna, T., Peregrin-Alvarez, J.M., Shales, M., Zhang, X., Davey, M., Robinson, M.D., Paccanaro, A., Bray, J.E., Sheung, A., Beattie, B., Richards, D.P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M.M., Vlasblom, J., Wu, S., Orsi, C., Collins, S.R., Chandran, S., Haw, R., Rilstone, J.J., Gandi, K., Thompson, N.J., Musso, G., St Onge, P., Ghanny, S., Lam, M.H., Butland, G., Altaf-Ul, A.M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J.S., Ingles, C.J., Hughes, T.R., Parkinson, J., Gerstein, M., Wodak, S.J., Emili, A., Greenblatt, J.F. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**: 637-643.

Kulkarni, R.D., Dean, R.A. (2004) Identification of proteins that interact with two regulators of appressorium development, adenylate cyclase and cAMP-dependent protein kinase A, in the rice blast fungus *Magnaporthe grisea*. *Molecular Genetics and Genomics : MGG* **270**: 497-508.

Lee, T., Yang, S., Kim, E., Ko, Y., Hwang, S., Shin, J., Shim, J.E., Shim, H., Kim, H., Kim, C., Lee, I. (2015a) AraNet v2: an improved database of co-functional gene networks for the study of *Arabidopsis thaliana* and 27 other nonmodel plant species. *Nucleic Acids Research* **43**: D996-1002.

Lee, T., Oh, T., Yang, S., Shin, J., Hwang, S., Kim, C.Y., Kim, H., Shim, H., Shim, J.E., Ronald, P.C., Lee, I. (2015b) RiceNet v2: an improved network prioritization server for rice genes. *Nucleic Acids Research* **43**: W122-127.

Li, Y., Yan, X., Wang, H., Liang, S., Ma, W.B., Fang, M.Y., Talbot, N.J., Wang, Z.Y. (2010) MoRic8 is a novel component of G-protein signaling during plant infection by the rice blast fungus *Magnaporthe oryzae*. *Molecular Plant-Microbe Interactions* **23**: 317-331.

Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E., Sacco, F., Palma, A., Nardozza, A.P., Santonico, E., Castagnoli, L., Cesareni, G. (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Research* **40**: D857-861.

Liu, T.B., Liu, X.H., Lu, J.P., Zhang, L., Min, H., Lin, F.C. (2010) The cysteine protease MoAtg4 interacts with MoAtg8 and is required for differentiation and pathogenesis in *Magnaporthe oryzae*. *Autophagy* **6**: 74-85.

Luo, Y., Zhang, H., Qi, L., Zhang, S., Zhou, X., Zhang, Y., Xu, J.R. (2014) FgKin1 kinase localizes to the septal pore and plays a role in hyphal growth, ascospore germination, pathogenesis, and localization of Tub1 beta-tubulins in *Fusarium graminearum*. *New Phytologist* **204**:

943-954.

Lysenko, A., Urban, M., Bennett, L., Tsoka, S., Janowska-Sejda, E., Rawlings, C.J., Hammond-Kosack, K.E., Saqi, M. (2013) Network-based data integration for selecting candidate virulence associated proteins in the cereal infecting fungus *Fusarium graminearum*. *PloS One* **8**: e67926.

Mehrabi, R., Ding, S., Xu, J.R. (2008) MADS-box transcription factor mig1 is required for infectious growth in *Magnaporthe grisea*. *Eukaryotic Cell* **7**: 791-799.

Meng, S., Brown, D.E., Ebbole, D.J., Torto-Alalibo, T., Oh, Y.Y., Deng, J., Mitchell, T.K., Dean, R.A. (2009) Gene Ontology annotation of the rice blast fungus, *Magnaporthe oryzae*. *BMC Microbiology* **9 Suppl 1**: S8.

Nepusz, T., Yu, H., Paccanaro, A. (2012) Detecting overlapping protein complexes in protein-protein interaction networks. *Nature Methods* **9**: 471-472.

Oesper, L., Merico, D., Isserlin, R., Bader, G.D. (2011) WordCloud: a Cytoscape plugin to create a visual semantic summary of networks. *Source Code for Biology and Medicine* **6**: 7.

Park, G., Xue, C., Zhao, X., Kim, Y., Orbach, M., Xu, J.R. (2006) Multiple upstream signals converge on the adaptor protein Mst50 in *Magnaporthe grisea*. *The Plant Cell* **18**: 2822-2835.

Pitre, S., Alamgir, M., Green, J.R., Dumontier, M., Dehne, F., Golshani, A. (2008) Computational methods for predicting protein-protein interactions. *Advances in Biochemical Engineering/Biotechnology* **110**: 247-267.

Qi, Z., Wang, Q., Dou, X., Wang, W., Zhao, Q., Lv, R., Zhang, H., Zheng, X., Wang, P., Zhang, Z. (2012) MoSwi6, an APSES family transcription factor, interacts with MoMps1 and is required for hyphal and conidial morphogenesis, appressorial function and pathogenicity of *Magnaporthe oryzae*. *Molecular Plant Pathology* **13**: 677-689.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., Lopez, R. (2005) InterProScan: protein domains identifier. *Nucleic Acids Research* **33**: W116-120.

Reimand, J., Kull, M., Peterson, H., Hansen, J., Vilo, J. (2007) g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Research* **35**: W193-200.

Remm, M., Storm, C.E., Sonnhammer, E.L. (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *Journal of Molecular Biology* **314**: 1041-1052.

Sadat, M.A., Choi, J. (2017) Wheat Blast: A New Fungal Inhabitant to Bangladesh Threatening

World Wheat Production. *The Plant Pathology Journal* **33**: 103-108.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**: 2498-2504.

Soanes, D.M., Chakrabarti, A., Paszkiewicz, K.H., Dawe, A.L., Talbot, N.J. (2012) Genome-wide transcriptional profiling of appressorium development by the rice blast fungus *Magnaporthe oryzae*. *PLoS Pathogens* **8**: e1002514.

Staskawicz, B.J., Mudgett, M.B., Dangl, J.L., Galan, J.E. (2001) Common and contrasting themes of plant and animal diseases. *Science* **292**: 2285-2289.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., Kuhn, M., Bork, P., Jensen, L.J., von Mering, C. (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research* **43**: D447-452.

Urban, M., Irvine, A.G., Cuzick, A., Hammond-Kosack, K.E. (2015) Using the pathogen-host interactions database (PHI-base) to investigate plant pathogen genomes and genes implicated in virulence. *Frontiers in Plant Science* **6**: 605.

Wang, T.Y., He, F., Hu, Q.W., Zhang, Z. (2011) A predicted protein-protein interaction network of the filamentous fungus *Neurospora crassa*. *Molecular BioSystems* **7**: 2278-2285.

Xenarios, I., Salwinski, L., Duan, X.J., Higney, P., Kim, S.M., Eisenberg, D. (2002) DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Research* **30**: 303-305.

Yellaboina, S., Tasneem, A., Zaykin, D.V., Raghavachari, B., Jothi, R. (2011) DOMINE: a comprehensive collection of known and predicted domain-domain interactions. *Nucleic Acids Research* **39**: D730-735.

Yin, Z., Tang, W., Wang, J., Liu, X., Yang, L., Gao, C., Zhang, J., Zhang, H., Zheng, X., Wang, P., Zhang, Z. (2016) Phosphodiesterase MoPdeH targets MoMck1 of the conserved mitogen-activated protein (MAP) kinase signalling pathway to regulate cell wall integrity in rice blast fungus *Magnaporthe oryzae*. *Molecular Plant Pathology* **17**: 654-668.

Zhang, H., Xue, C., Kong, L., Li, G., Xu, J.R. (2011) A Pmk1-interacting gene is involved in appressorium differentiation and plant infection in *Magnaporthe oryzae*. *Eukaryotic Cell* **10**: 1062-1070.

Zhang, H., Zhao, Q., Guo, X., Guo, M., Qi, Z., Tang, W., Dong, Y., Ye, W., Zheng, X., Wang, P.,

Zhang, Z. (2014) Pleiotropic function of the putative zinc-finger protein MoMsn2 in *Magnaporthe oryzae*. *Molecular Plant-Microbe Interactions* **27**: 446-460.

Zhang, K., Li, Y., Li, T., Li, Z.G., Hsiang, T., Zhang, Z., Sun, W. (2017) Pathogenicity genes in *Ustilaginoidea virens* revealed by a predicted protein-protein interaction network. *Journal of Proteome Research* **16**: 1193-1206.

Zhao, X., Kim, Y., Park, G., Xu, J.R. (2005) A mitogen-activated protein kinase cascade regulating infection-related morphogenesis in *Magnaporthe grisea*. *The Plant Cell* **17**: 1317-1329.

Zhao, X.M., Zhang, X.W., Tang, W.H., Chen, L. (2009) FPPI: *Fusarium graminearum* protein-protein interaction database. *Journal of Proteome Research* **8**: 4714-4721.

Zhou, X., Zhang, H., Li, G., Shaw, B., Xu, J.R. (2012) The Cyclase-associated protein Cap1 is important for proper regulation of infection-related morphogenesis in *Magnaporthe oryzae*. *PLoS Pathogens* **8**: e1002911.

Zhou, X., Liu, W., Wang, C., Xu, Q., Wang, Y., Ding, S., Xu, J.R. (2011) A MADS-box transcription factor MoMcm1 is required for male fertility, microconidium production and virulence in *Magnaporthe oryzae*. *Molecular Microbiology* **80**: 33-53.

초 록

벼 도열병균은 식물-미생물 상호작용 연구 분야에 있어 중요한 모델 생물 중 하나이며, 병 발생 시기의 부착기 형성과정에 관여하는 신호전달 경로가 선행 연구에서 많은 부분 밝혀진 바 있다. 유전체 해독 기술의 발달로 멀티오믹스 데이터가 이용가능해지면서, 이 병원균의 병 발생과정을 밝히기 위해 여러 차례의 유전체 단위의 연구들이 진행되어 왔다. 유전체 단위의 단백질 상호작용 네트워크는 신호전달 및 발현조절 경로를 이해하기 위한 연구 분야 중 하나이나, 현재 존재하는 벼 도열병균의 생물학적 네트워크 자원은 병리학적 연구에 직접 이용되기에 불충분하다. 이 연구에서는 벼 도열병균의 단백질 상호작용 네트워크 플랫폼인 MagNet을 상동성 기반의 상동상호작용검색, 도메인 상호작용 정복 기반의 예측 그리고 공발현 네트워크의 세 가지 방법으로 구축했다. 여섯 개의 단백질 상호작용 정보 데이터베이스를 이용해서 모델 생물에 상동성 단백질을 갖고 있는 단백질들 사이의 상동상호작용을 찾았다. 공발현 네트워크는 전사체 해독 데이터를 기반으로 감염시기와 영양생장시기의 두 개의 네트워크를 생성했다. 도메인 상호작용 정보를 이용해서는 6,231개의 벼 도열병균 단백질 사이의 3,121,109개의 상호작용을 예측했고 세 가지 방식을 종합한 결과 5,600,976개의 상호작용과 높은 신뢰도의 215,731개의 상호작용을 추려낼 수 있었다. 단백질 상호작용 실험결과를 통해 이번 연구에서 구축된 높은 신뢰도의 단백질

상호작용 네트워크가 기존의 데이터베이스보다 더 높은 정확도로 단백질 상호작용을 예측했음을 확인했다. 이 연구로 구축된 MagNet은 식물병원균의 병 발생 메커니즘을 이해하는데 있어 통합된 단백질 상호작용 네트워크를 제시할 수 있을 것이다.