In-silico Study of the Role of Differentially Expressed Genes and MicroRNAs in Host-Pathogen Interaction

Nilubon Kurubanjerdjit, Ching–Yi Chen, Ke-Rung Tzeng, Jin-Shuei Ciou, Jeffrey J.P Tsai,
Chien-Hung Huang *, Ka-Lok Ng *

***Abstract*—In this study, the microarray data for *Arabidopsis thaliana* infected with *Xanthomonas campestris pv. campestris* (*Xcc*) is analyzed, where differentially expressed genes (DEGs) are identified, and Gene Set Enrichment Analysis (GSEA) are employed for analysis. As a result, highly relevant pathogen resistant pathways are inferred. Furthermore, the results of DEGs for various different conditions; such as, infection by different strains of *Agrobacterium tumefaciens*, are determined. The results can be accessed at http://ppi.bioinfo.asia.edu.tw/R_At_xcc/index.htm.**

Furthermore, protein-protein interactions (PPIs) play an important role in the host-pathogen interactions. Gene Ontology (GO) annotation for microRNA-regulated PPI, pathogen resistant genes and transcription factors information are implemented, such resources can provide new insights for microRNA-regulated PPI networks in host-pathogen interaction study. The database is freely accessible at http://ppi.bioinfo.asia.edu.tw/At_miRNA/.

***Index Terms* — microRNA, microarray, host-pathogen interaction, *Arabidopsis*, *Xanthomonas campestris* pv. *Campestris (Xcc)*, differentially expressed genes, gene set enrichment analysis, protein interaction**

I. Introduction

GAINING a better understanding of the biotic and abiotic stress responses for plant systems provide a model system for studying human diseases and drug-related research. Understanding how plant systems defense against environment stress is of great significance for the world's food and agricultural production.

It is well known that *Arabidopsis thaliana* (*A. thaliana*), a long day plant, is a good model organism for plant science [1]. *A. thaliana* is chosen as the model system for two reasons: (1) the complete genome sequence has been known since 2000; and (2) there are many molecular tools, such as cDNA, genomic libraries, bacterial artificial chromosomes, microarrays and ESTs, are available for the study of its biological functions [1]. Only a small number of bacteria are pathogenic on *A. thaliana*, where more than 3,000 proteins are directly related to the plant defense response mechanism [2-3]. *A. thaliana* also play a crucial role as a model organism for the study of plant-pathogen interaction, many model systems have been developed to better understand the interactions between plants and bacteria, fungal, viral and nematode pathogens. *A. thaliana* has been successfully implemented in the study of the interaction between plants and disease-causing pathogens.

Xanthomonas campestris pv. campestris (*Xcc*) is one of the pathogenic gram-negative bacteria that cause blights and rots in plants [4-7]. Host infections caused by *Xcc* can occur in any stage of the plant life cycle. Symptoms resulted from this pathogen have been reported in many previous research works [4-8]. In addition, *Xcc* is considered the most important and most destructive disease of crucifers, infecting all cultivated varieties of brassicas worldwide. Host infection by *Xcc* can occur at any stage of the plant life cycle.

Plants are continuously invaded by pathogens including bacteria, fungi, nematodes, viruses and insect pests. Generally, a pathogenic bacterium attacks hosts in many ways including sticking and colonizing host tissues, secreting degradation enzymes and toxins release. Pathogen-associated molecular patterns (PAMPs) trigger plant defenses when perceived by surface-localized immune receptors. PAMP-triggered immunity (PTI) plays an important role in the resistance of plants to pathogens. Many plant pathogens, including bacteria and viruses, can deliver a variety of effector proteins into the host plant cell to inhibit PTI signaling [9-10]. In response, plant resistance proteins sense effectors to activate effector-triggered immunity (ETI), which is a second inducible defense layer [11-12].

Time series microarray experiments were studied to infer pathogen induced genes. Differentially expressed genes (DEGs) were identified by using the *Bioconductor* statistical package, EBAYES.

Plant microRNAs (miRNAs) are usually perfectly complementary to their targets and cause the cleavage of their targets by a RNA-induced silencing complex (RISC). The translational inhibition by miRNAs has been thought of

Manuscript received December 8, 2013. The work of Chien-Hung Huang is supported by the National Science Council of Taiwan under grant NSC 101-2221-E-150-088-MY2. The work of Jeffrey J.P Tsai is supported by the grant NSC 102-2632-E-468-001-MY3, the work of Ka-Lok Ng is supported by NSC 102-2221-E-468-024 and NSC 102-2632-E-468-001-MY3.

Nilubon Kurubanjerdjit, Ching–Yi Chen, Ke-Rung Tzeng, Jin-Shuei Ciou, and Jeffrey J.P Tsai are with the Department of Biomedical Informatics, Asia University, Taiwan 41354.

* Corresponding author, Chien-Hung Huang is with the Department of Computer Science and Information Engineering, National Formosa University, Taiwan 632 (Phone: +886-56315588; fax: +886-56330456; e-mail: chhuang@nfu.edu.tw).

* Corresponding author, Ka-Lok Ng is with the Department of Biomedical Informatics, Asia University, Taiwan 41354 (Phone: +886-423394541; fax: +886-423320718; e-mail: ppiddi@gmail.com).

as a major mechanism in animal systems while mRNA degradation or post-transcriptional regulation has been considered as a major regulatory mechanism in plants [13].

MiRNAs play crucial roles in *A. thaliana* biological processes, such as leaf sidedness, flower development, hormone signaling, metabolism, and stress response. Due to the limitations of the current techniques, high-throughput target validation via biological experiments is not practical. Given these circumstances, a lot of computational target prediction methods have been developed, and online open resources are developed to fulfill the need of scientists performing miRNA research.

Host-pathogen interaction (HPI) mechanism involves host-pathogen protein-protein interaction (PPI). PPI is an essential process of living cells [14]. It also plays a crucial role in some critical interspecies interactions such as HPIs and pathogenicity [15]. Recently high throughput proteomic technology has uncovered a large number of PPI, particularly in interspecies protein interactions of plants and bacteria [16]. Therefore, comprehensive knowledge of host-pathogen PPI and interactome analysis can help accelerating protein annotations and elucidate a plant's immune system against bacteria. It is known that aberrant miRNA expression or defects in PPI can possibly induce diseases.

In this study, results are deployed as a web-based platform. This platform provides the following information; miRNA-targeted mRNA data, *A. thaliana* PPI records obtained from BioGrid [17], pathogen resistant genes (PRGs) [18] and transcription factors (TFs) [19] data.

II. Materials and Methods

A. Datasets

Microarray data for the *A. thaliana* (Columbia wild type, col-0) infected with the *Xcc*147 strain was downloaded from PLEXdb [20] with an experiment ID, AT-87 or from ArrayExpress [21] with ID, E-GEOD-9674. AT-87, an Affymetrix microarray platform, compared gene expression levels between samples at 0 minute, and 90-105 minutes, 2-4 hours and 6 hours after inoculation dissected from the leave tissue. Since there is only one sample available for the 6 hours time point, therefore, it is not used in the present study.

B. Differentially expressed gene identification

To identify *Xcc* induced DEGs, the following analyses was preformed, (i) the zero minute samples are compared with the 90-105 minutes samples, and (ii) the zero minute samples are compared with the 2-4 hrs samples.

The EBAYES algorithm computes moderated t-statistics, moderated F-statistic, and log-odds of differential expression by empirical Bayes shrinkage of the standard errors towards a common value.

SAM is a statistical method for identifying DEGs by comparing two or more groups of samples. It uses repeated permutations of the data to estimate False Discovery Rate (FDR) based on observed versus expected score, which is obtained from randomized data. A gene which has an observed score that deviates significantly from the expected score is consider as a DEG. EBAM performs one and two class analyses using either a modified t-statistic or standardized Wilcoxon rank statistic, and a multiclass analysis using a modified F-statistic. Moreover, this function provides a EBAM procedure for categorical data such as SNP data and the possibility of employing a user-written score function. Our previous study [22] suggested that, EBAYES, SAM, and EBAM, achieve a similar level of cancer gene prediction accuracy, i.e. around 20%, therefore, EBAYES is adopted in the present analysis.

C. Gene Set Enrichment Analysis (GSEA)

Functional annotation of the DEGs is given by implementing the Database for Annotation, Visualization and Integrated Discovery, i.e. DAVID [23]. DAVID provides functional annotation tools, which mainly supply gene GO term enrichment analysis to highlight the most relevant GO terms associated with a given gene list. The list of *Arabidopsis* DEGs was submitted to DAVID for clustering; hence, enriched pathways were obtained.

D. MiRNA Target Gene Prediction Algorithms

There are many miRNA target gene prediction algorithms are available. In the present study, we chose RNAHybrid, miRanda and PITA algorithms for the prediction. The motivation for choosing these tools is that they take into account of various prediction features, such as sequence complementary, thermodynamics properties, target site accessibility and multiple binding. RNAHybrid was developed by Kruger and Rehmsmeier [24], which predict miRNA targets by calculating the minimum free energy (MFE) of hybridization between target genes and miRNA sequences. MiRanda is another algorithm for identifying miRNA targets in Drosophila and humans. For each miRNA, miRanda selects target genes on the basis of three properties; sequence complementarity using a position-weighted local alignment, free energies of RNA-RNA duplexes, and conservation of target sites in related genomes. PITA is a target prediction tool, which calculates the thermodynamic free energy of the binding event and assigns scores to both single binding sites and multiple binding sites. This algorithm incorporates target accessibility into miRNA target site prediction to take into account the secondary structure of the miRNA-target hybrid.

There may be concern that those tools were developed specifically for animals or human, which may not be applicable in plants. We note that RNAHybrid has been adopted to predict microRNA targets in *A. thaliana*. Furthermore, both miRanda and PITA also calculate the free energies of RNA–RNA duplexes, therefore; these two tools are adopted in our analysis.

To prepare the training set, a set of experimentally confirmed miRNA-target pairs was downloaded from ASRP. These pairs were derived from a set of 118 miRNAs (BLAST e-values are somewhere between $2*10^{-10}$ and 0.62) and a set of 205 mRNAs. This experimentally confirmed set was processed by the three machine learning classifiers; SVM, random forest (RF) and neural network (NN). Then, target pairs predicted by each algorithm were merged. The positive training set (406 pairs) are experimentally confirmed pairs that satisfied the three algorithms' parameter settings. The negative set, a total of 9938, comprised pairs that satisfied the three algorithms' settings with the positive set subtracted. The test set was generated

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

by computing the three prediction scores for a set of 243 *A. thaliana* miRNA and a genome wide set of UTR.

For each classifier, parameter setting was identified by observing the accuracy (ACC), specificity (SPC), sensitivity (SEN), and F-score (F1) performance. Firstly, each classifier was trained by using the training set to conduct a 10-fold cross-validation test. Next, a set of parameters for each algorithm was systematically changed to observe the prediction performance. For instance, the RF model requires six parameter settings; starting from the default settings, then, the first parameter value was changed systematically from its default value while keeping other parameters fixed. The parameter value will be frozen if it returns the highest F1 value. Next, the above steps will be repeated for other parameter to find the highest F1. Their values will be frozen before moving on the next parameter. Finally, the optimal parameter setting of each model with the highest F1 was obtained.

E. MiRNA-regulated Protein-Protein Interaction Pathways

It is known that defects in PPI can possibly induce diseases. In this study, miRNA-targeted mRNA data, PPI records from BioGrid, PRGs and TFs data are integrated. To quantify the relationship among miRNAs, target genes, and their PPIs, the importance of miRNA-PPI coupled networks are ranked by performing enrichment analysis. There is a tendency for two directly interacting proteins to participate in the same biological process or share the same molecular function. Enrichment analysis was performed by computing the *Jaccard* coefficient (*JC*) to rank the significance of such relations.

III. Results

A. The results of GSEA

Using EBAYES with an adjusted *p*-value less than 0.05, DEGs due to *Xcc* inoculation were identified. When adopting GSEA, sets of data obtained at different time points were studied. Both 90-105 minutes and 2-4 hrs samples are used. These two measurements allow us to infer how fast is the host response to the infection. After the clustering step, enriched KEGG pathways were obtained. Table 1 summarized the enriched KEGG pathways for the 90-105 minutes data. The last column represented the number of genes, N_g, identified among the pathway genes, N_{total}.

TABLE 1
ENRICHED KEGG PATHWAYS FOR THE 90-105 MINUTES SAMPLES

Pathway	*Ng/Ntotal*
"Protein processing in endoplasmic reticulum"	115/138
"Plant-pathogen interaction"	138/148
"Protein export"	38/46
"Indole alkaloid biosynthesis"	6/7
"Phenylpropanoid biosynthesis"	96/109
"Amino sugar and nucleotide sugar metabolism"	93/102
"Ribosome"	184/311
"Phenylalanine, tyrosine and tryptophan biosynthesis"	27/52
"Vitamin B6 metabolism"	8/9
"Endocytosis"	60/72
"Protein processing in endoplasmic reticulum"	115/138

Enrichment analysis was conducted using two different inputs, one of the analysis was performed using a complete set of data as input (cases *a* and *b* in Table 2); the same analysis was performed using DEGs only (cases *c* and *d* in Table 2). Table 2 lists the top five pathways given by DAVID: (i) plant-pathogen interaction, (ii) protein processing in endoplasmic reticulum, (iii) phenylalanine, tyrosine and tryptophan biosynthesis, (iv) phenylpropanoid biosynthesis and (v) protein export pathways.

TABLE 2
THE TOP FIVE ENRICHED KEGG PATHWAYS FOR CASES a, b, c and d.

Pathway	*Nrep*	Case
Plant-pathogen interaction	4	*a, b, c, d*
Protein processing in endoplasmic reticulum	4	*a, b, c, d*
Phenylalanine, tyrosine and tryptophan biosynthesis	3	*a, b, c*
Phenylpropanoid biosynthesis	3	*a, b, c,*
Protein export	3	*a, c, d*

a denotes analysis without DEG identification for the 90-105 minutes samples, *b* denotes analysis without DEG identification for the 2-4 hrs samples.；*c* denotes DEGs identification for the 90-105 minutes samples, *d* denotes DEGs identification for the 2-4 hrs. samples.
Nrep denotes the number of times KEGG pathways are found in the four cases.

Table 3 depicted the results of the genes involving in the top five enriched KEGG pathways.

TABLE 3
GENES INVOLVE IN THE TOP FIVE ENRICHED KEGG PATHWAYS

Pathway	Gene symbol
Plant-pathogen interaction	For PTI - CDPK1, FRK
	For ETI - RPM1, SGT1A, HSP90
Protein processing in endoplasmic reticulum	CNX1, CRT1b, SAR1, DER1, p97, RAD23, Png1, HSP70, sHSF
Phenylalanine, tyrosine and tryptophan biosynthesis	ASK1, EMB1144
Phenylpropanoid biosynthesis	PAL1, PAL2, HCT, CAD5, CCR2
Protein export	SEC61(α, β, γ), SEC11, SRPRP

PTI denotes PAMP-triggered immunity, ETI denotes effector-triggered immunity

Plant-pathogen interaction pathway

Plant systems have developed the PTI and ETI mechanisms to defense against bacterial infection. When the host is first attacked by bacterial, it releases pattern-recognition receptors (PRR) protein to recognize pathogen-associated molecular patterns (PAMP), which are associated with pathogens. In response to PTI, pathogens may release the effector proteins (such as AvrPphB, AvrRpm1, AvrRpt2), to suppress PTI; thus trigger ETI, the host's second line of defense. At this stage, host resistance genes (R genes) are activated to recognize effector proteins. Most of the bacteria employ the Type III secretion system (T3SS) to inject effector proteins into the host cell. ETI associates with a program cell death called hypersensitive response (HR). It is known that HR is a typical response for *A. thaliana* after infection with *Xcc* [25]. Furthermore, it is suggested that the SGT1 and HSP90 protein complexes utilize the SKp1 protein in the ubiquitin-proteasome system

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

to regulate the hypersensitivity resistance mechanism, which is mediated by the resistance protein RPM1.

Protein processing in endoplasmic reticulum

Protein folding is taken place in the endoplasmic reticulum (ER) with the help of the chaperones. Properly folded proteins are translocate to Golgi, while incorrect folded proteins are refolded with the help of chaperones at ER. Misfolded proteins are digested through the endoplasmic reticulum-associated protein degradation process. Furthermore, ER stress can activate the unfolded protein response (UPR) signaling transduction pathway. In a more serious incident, it can lead to cell apoptosis [26].

Protein export pathway

Protein secretion is a basic cellular process found in every species. The *Sec*-dependent pathway is the general protein transport system that moves proteins in or across the plasma membrane. The SEC pathway, also known as type II secretion system (T2SS) is highly conserved in prokaryotes, and the Sec translocation channel is make of a highly conserved membrane protein complex [27], which is compose of several subunits, i.e. Sec61-alpha, Sec61-beta and Sec61-gamma [28].Other protein export systems have been identified in gram-negative bacteria as well, such as the type I, II, III, IV and V secretion system [29].

Plants produce a lot of secondary metabolites, which are found to play an important role in defending plants against pathogens [30]. Secondary metabolites can be divided into three major groups: phenolics, terpenes, and nitrogen-containing compounds.

Phenylalanine, tyrosine and tryptophan biosynthesis

Phenylalanine, tyrosine, and tryptophan are the end-products of the shikimate pathway. A lot of secondary metabolites are also synthesis through this pathway [31].

Tyrosine/tryptophan is the amino acid precursor for isoquinoline/indole biosynthesis, there is evidence for a role of indole-derived compounds, in pathogen resistant [32]. Phenylalanine is the amino acid precursor for phenylpropanoid biosynthesis.

Phenylpropanoid biosynthesis pathway

The formation of many plant phenolics, including phenylpropanoid, lignin and anthocyanins, starts from phenylalanine. Lignin mechanically strengthens cell walls that play a role in pathogen infection [31].

DEGs due to *Xcc* inoculation under other conditions were determined; for instance, (i) MYB30 over-expression (90-105 minutes and 2-4 hours), (ii) wildtype and MYB30 over-expression (90-105 minutes and 2-4 hours), and (iii) infection by the bacterium, *Agrobacterium tumefaciens* subtype C58 and GV3101 (at two different time points, 3 hours and 6 days). These results can be accessed at http://ppi.bioinfo.asia.edu.tw/R_At_xcc/index.htm.

B. Predicted miRNA-target Interaction

Genome-wide miRNA target prediction was performed using the three target predictors, where feature vectors of the 158,750 miRNA-target interactions were input into the NN, SVM and RF classifiers. Table 4 depicted the results of the performance of the three classifiers as well as their combination, i.e. NN+SVM+RF, where it achieves the best classification accuracy, and rank second in SPE. RF classifier also achieves comparable performance.

TABLE 4
PERFORMANCE OF THE MACHINE LEARNING CLASSIFIERS

Classifier	ACC	SPE	SEN	F1
NN	0.9817	0.9821	0.970	0.9765
SVM	0.9832	0.9881	0.8618	0.9207
RF	0.9841	0.9845	0.9734	0.9789
NN+SVM+RF	0.9862	0.9871	0.9650	0.9759

As a result of classification, 3-vote (NN+SVM+RF), 2-vote (NN+SVM, NN+RF, or SVM+RF) and 1-vote groups had 6808, 24457 and 20784 interactions respectively; in other words, the number of positive miRNA target pairs is 6808 predicted by the 3-vote study.

The 3-vote group was selected to identify the interacting module among miRNA and their targets. It was found that one can classify the interactions in three types of modules, i.e. one-to-one, one-to-many, and many-to-many.

We have established a web-based platform, the platform provides the following information: (i) for a given miRNA ID, output the targeted mRNAs and PPI partners up to four levels (more protein levels can be done if necessary), (ii) search for PRG or TF specific target gene PPI partners; (iii) depending on the user interest, either experimentally verified or predicted miRNA-regulated PPI pathways can be selected; and (iv) based on GO annotation, *JC* is given to characterize the GO similarity of a PPI.

The database is freely accessible at http://ppi.bioinfo.asia.edu.tw/At_miRNA. Figure 1 shows the miRNA-regulated PPI web page.

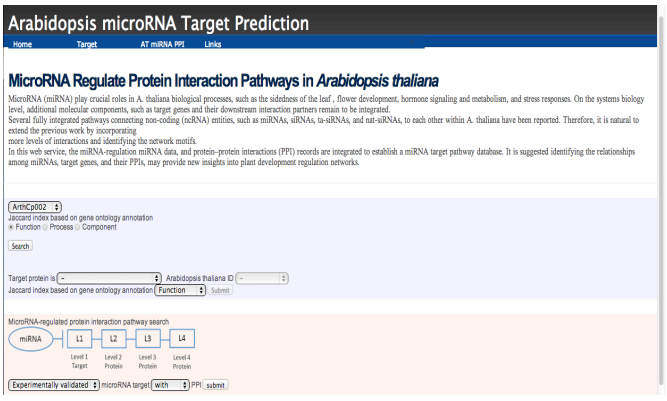

Fig. 1. The miRNA target genes and PPI records are integrated to establish a miRNA-regulated PPI pathway database.

For example, it was found that miR-472a targets a PRG, AT1G12220, where this gene interacts with AT5G13160. AT1G12220, also known as RPS5, mediates resistance against the bacterial pathogen Pseudomonas syringae. AT5G13160 (PBS1) is also a PRG. The *JC* for these two genes' GO biological process annotation is 50%, which indicated that both genes participate in highly similar processes. Identifying the relationships among miRNAs,

ISBN: 978-988-19252-5-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

target genes, and their PPIs, may provide new insights into plant development regulation networks.

IV. Conclusion

We consider the *A. thaliana* and *Xcc* as a model system to investigate the HPI issue by conducting GSEA. Highly relevant pathogen resistant pathways are inferred. The analyses suggested that certain proteins, i.e., SGT1, HSP and SEC, and secondary metabolites are actively involved in plant defense mechanism.

MiRNA and PPI play an important role in the infection process, some critical inter-species interactions such as HPI and pathogenicity occur through PPI [33]. GO annotation of miRNA-regulated PPI with PRG and TF information were implemented. Such resources can provide new insights into miRNA-regulated PPI networks in HPI study. The approach developed in this study should be of value for future studies in understanding the molecular mechanisms enabling *A. thaliana* to respond to pathogen attack.

References

[1] D.F. Mandoli, R. Olmstead, "The importance of emerging model systems in plant biology," *J Plant Growth Regul* 19(3): pp. 249-252, 2000.

[2] J.G. Bishop, A. M. Dean, T. Mitchell-Olds, "Rapid evolution in plant chitinases: molecular targets of section in plant-pathogen coevolution," *P Natl Acad Sci* USA 97(10): pp. 5322-5327, 2000.

[3] O.A. Postnikova, N. Y. Minakova, A.M. Boutanaev, L.G. Nemchinov, "Clustering of pathogen-response genes in the genome of arabidopsis thaliana," *J Integr Plant Biol* 53: pp. 824-834, 2011.

[4] J. Tsuji, S.C. Somerville, "Xanthomonas campestris pv. campestris induced chlorosis in Arabidopsis thaliana," *Arabidopsis Information Service* 26: pp. 1-8, 1988.

[5] J. Tsuji, S.C. Somerville, "First report of the natural infection of arabidopsis thaliana by xanthomonas campestris pv. campestris," *Plant Dis* 76: 539, 1992.

[6] J. Tsuji, S.C. Somerville, R. Hammerschmidt, "Identification of a gene in arabidopsis thaliana that controls resistance to xanthomonas campestris pv. campestris," *Physiol Mol Plant Pathlogy* 38: pp. 57-65, 1991.

[7] C.R. Buell, "Interactions between xanthomonas species and arabidopsis thaliana," *Arabidopsis Book* 1: e0031, 2002.

[8] D. Meyer, E. Lauber, D. Roby, M. Arlat, T. Kroj, "Optimization of pathogenicity assays to study the arabidopsis thaliana–xanthomonas campestris pv. campestris pathosystem," *Mol Plant Pathol* 6(3): pp. 327-333, 2005.

[9] V. Gohre, S. Robatzek, "Breaking the barriers: microbial effector molecules subvert plant immunity," *Annu Rev Phytopathol* 46: pp. 189-215, 2008.

[10] J.M. Zhou, J. Chai, "Plant pathogenic bacterial type III effectors subdue host responses," *Curr Opin Microbiol* 11(2): pp. 179-185, 2008.

[11] S.T. Chisholm, G. Coaker, B. Day, B.J. Staskawicz, "Host-microbe interactions: shaping the evolution of the plant immune response," *Cell* 124(4): pp. 803-814, 2006.

[12] D.J. Jones, J. L Dangi, "The plant immune system," *Nature* 444: pp. 323-329, 2006.

[13] X. Dai, Z. Zhuang, P.X. Zhao, "Computational analysis of miRNA targets in plants: current status and challenges," *Briefings in Bioinformatics* 12(2): pp. 115-121, 2010.

[14] N. Lin, W. Baolin., R. Jansen, M. Gerstein, H. Zhao, "Information assessment on predicting protein-protein interactions," *BMC Bioinformatics* 5: 154, 2004.

[15] Casadevall A., Pirofski. L. A, "Host-pathogen interactions: basic concepts of microbial commensalism, colonization, infection, and disease," *Infection and Immunity* 68(12): pp. 6511-6518, 2000.

[16] T. H. Tsao, C. H. Chan, Huang C.Y., Lee S.A, "Systems and computational biology - molecular and cellular experimental systems," Ning-Sun Yang (ed.) The prediction and Analysis of Inter- and Intra-Species Protein-Protein Interaction. China, InTech, 2011.

[17] B.J. Breitkreutz et al., "The BioGRID interaction 64 database: 2008 update," *Nucleic Acids Research* 36: D637–D640, 2008.

[18] W. Sanseverino et al., "PRGdb: a bioinformatics platform for plant resistance gene analysis," *Nucleic Acids Research* 38: pp. 814-821, 2010.

[19] V. Matys et al., "TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes," *Nucleic Acids Research* 34: pp. 108-110, 2006.

[20] S. Dash, J. Van Hemert, et al, "PLEXdb: gene expression resources for plants and plant pathogens," *Nucleic acids research*, vol 40, Issue Database issue. pp. 1194-1201, 2012.

[21] A. Brazma et al., "Array Express: a public for microarray gen expression data at the EBI," *Nucleic Acids Res* 31, pp. 68-71, 2003.

[22] S. T. Chen, H.F. Wu, K.L. Ng, "A platform for querying breast and prostate cancer-related microNA genes," *International Conference on Bioinformatics and Biomedical Engineering (ICBBE 2012)*, 1(1), pp. 271-274, Shanghai, May 17-20, 2012.

[23] W. Huang da et al, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature protocols*, vol 4, Issue 1. pp. 44-57, 2009.

[24] J. Kruger, M. Rehmsmeier, "RNAhybrid: miRNA target prediction easy, fast and flexible," *Nucleic Acids Research* 34: pp. 451-454, 2006.

[25] C.R. Buell, "Interactions Between Xanthomonas Species and Arabidopsis thaliana," *The Arabidopsis Book* 1, e0031, 2002.

[26] Kanehisa Laboratories. Protein processing in endoplasmic reticulum - Arabidopsis thaliana (thale cress) [website message]. Retrieved from http://www.genome.jp/kegg-bin/show_pathway?ath04141

[27] Kanehisa Laboratories. Protein export - Arabidopsis thaliana（thale cress）[website message]. Retrieved from http://www.genome.jp/dbget-bin/www_bget?ath03060

[28] E. Park, T.A. Rapoport, "Mechanisms of Sec61/SecY-Mediated Protein Translocation Across Membranes," *Annu Rev Biophys* 41: pp. 21-40, 2012.

[29] Lisette A. de Vries, "Role of Arabidopsis resistance proteins and Pseudomonas syringae effector proteins in hostmicrobe interactions," Master Thesis, UMC Utrecht, 2009.

[30] L. Taiz, E. Zeiger, *Plant Physiology*. MA: Sinauer Associates, Inc. 2006, ch. 13.

[31] M. F. Christopher, C. Clint, "The Phenylpropanoid Pathway in Arabidopsis," *The Arabidopsis Book*, 9, e0152, 2011.

[32] W. M. Truman et al., "Arabidopsis auxin mutants are compromised in systemic acquired resistance and exhibit aberrant accumulation of various indolic compounds," *Plant Physiology*, 152(3): pp. 1562-1573, 2010.

[33] N. Kurubanjerdjit et al., "Prediction of microRNA-regulated protein interaction pathway in Arabidopsis using machine learning algorithms", *Comp Biol Med*, 43(11), pp. 1645-1652, 2013.