Nebraska Soybean Board
Final Research Findings Report

Soybean Phytophthora root rot disease control by computationally identifying virulence effectors in the causal agent, Phytophthora

PI: Chi Zhang
Funding period: Oct. 1st, 2014 - Sept. 30th, 2016.

Abstract

Phytophthora root rot is a major disease of soybean and causes the soybean industry millions of dollars lost each year. *Phytophthora Sojae* (*P. Sojae*), a notorious oomycete pathogen, is the causal agent, which secretes virulence effectors as major 'weapons' to attack the target host organisms. Understanding the biological functions of virulence effector is the key step to reveal the mechanism of oomycete pathogenicity to hosts and control the disease. Interestingly, despite the extreme importance, the knowledge about *P. Sojae* virulence effectors and their interactions with soybean proteins is very limited. Under the support from NE Soybean Board from 2014 to 2016, we conducted research on genome-wide identification of all virulence effectors in *P. Sojae*, and explore the so-far blank research area of protein interactions between *P. Sojae* virulence effectors and soybean proteins. We analyzed all proteins in *P. Sojae* genome, compared them with the known virulence effectors in oomycetes and fungi, which were collected by us. We designed a novel statistical model to identify the proteins that are highly similar to the known virulence effectors, and consider them as putative virulence factors in *P. Sojae*, which are potential targets for the disease control. There are 26,584 proteins in *P. Sojae*, and we found 596 putative virulence effectors and more than 18K soybean proteins as their potential targets. We conducted validation on the identified virulence effectors based on the transcriptome studies and genome comparisons. As a result of this research, the mechanism under which *P. Sojae* interacts with and infects soybean can be elucidated, which may directly lead to feasible and efficient methods to prevent soybean from the Phytophthora root rot disease, and hence help Nebraska soybean farmers prevent the yield-loss due to this disease. In the future, we will model the protein structures to all discovered virulence effector, and predicted their active sites/functions.

Introduction

Oomycetes, a kind of fungus-like eukaryotic microorganisms, are economically important because they are aggressive plant pathogens and *Phytophthora Sojae* (*P. Sojae*) is a notorious oomycete pathogen to soybeans, which causes Phytophthora root rot, a major soybean disease. Yield losses caused by this disease are substantial; entire fields can be destroyed. It costs the soybean industry millions of dollars each year. On the other hand, this disease is difficult to effectively control, *e.g.* under conditions that are favorable to the pathogen, even resistant cultivars can become severely infected (Xiao et al., 2002). Therefore, understanding the intricate interactions between this pathogen and soybean is important because it is a key step towards the disease control.

The pathogenicity of *P. Sojae* is inducted by virulence factors, the molecules, such as proteins, produced by pathogens as major 'weapons' to attack the target host organisms. The development of plant diseases is a complicated process, and virulence factors play various roles in this process, including involvement in the production of infection structures, in the penetration of the cuticle and cell walls, and in fungal nutrition *etc.* Some of them, called effectors, are secreted into the apoplast or xylem of their host plants and subsequently translocated into host cells. The first known *P. Sojae* effector *Avr1a* was cloned in 2009 (Dong et al., 2009), and is has been discovered that *P. Sojae* effector proteins target different host plant tissues (Kamoun, 2006; Morgan and Kamoun, 2007), such as inhibitors of plant hydrolases in the extracellular space(Tyler et al., 2006). These *P. Sojae* effector proteins result in the induction of hypersensitive cell death and immunity. It is believed that

virulence effectors facilitate virulence by suppressing pathogen-associated molecular pattern-triggered immunity and inducing effector-triggered immunity in plants. Because they directly enable the pathogen to infect its host plant and cause disease, understanding the biological functions of virulence effector can reveal the mechanism of pathogenicity to plants, including soybeans, and lead to efficient methods for disease control. Interestingly, despite the extreme importance, the knowledge about oomycete virulence effectors and their interactions with soybean proteins is very limited. On the other hand, effector genes undergo rapid sequence diversification and their products typically show high rates of amino acid polymorphisms, particularly nonsynonymous substitutions with signatures of positive selection (Win et al., 2007). These properties of oomycete effector proteins enable them to be identified with data mining tools.

This project conducted genome-wide identification to all virulence effectors in *P. Sojae*, which causes the Phytophthora root rot disease. Understanding the molecular level interactions between this pathogen and soybean is important because it is a key step towards the disease control. First, we predict putative virulence effectors of *P. Sojae*. For this step, we developed a novel accurate statistical model to identify putative virulence effectors in the pathogen. In the second step, we validated the candidates of virulence effectors with bioinformatics and experimental approaches. Finally, we explored interactions between *P. Sojae* virulence factors and soybean proteins, which may directly lead to feasible and efficient methods to prevent soybean Phytophthora root rot disease.

Results

1. The novel bioinformatics method for identification of putative virulence effectors

As knowing a pathogenic protein *p* with a domain d_p interacts with a host protein *h* with a domain d_h, based on the Bayes' theorem, the two domains interaction probability can be calculated with Equation (1):

$$P(d_p,d_h \mid p,h) = \frac{P(p,h \mid d_p,d_h)P(d_p,d_h)}{\sum P(p,h)} \quad (1)$$

From the BioGRID dataset, the pathogen and host protein-protein interaction pairs are abstracted based on their taxonomy ids, and all the proteins parse the Pfam domains to calculate the domain-domain interaction probability (DDI) with Equation (1). When calculate the probability between the domains, only host-pathogen domains are considered. With the DDI datasets, a pair of pathogen and host protein can be predicted their interaction probability using Equation (2):

$$P(p,h) = 1 - \prod_p \prod_h (1 - P(d_p,d_h \mid p,h)) \quad (2)$$

We found that previous work did not consider the fact that a host protein and pathogen protein sharing the same domain may have different functions in their interactions. For example, some tip structures on bacteria surface can pierce their host membrane, while a tip like structure on the host surface does not have this function. Based on these observations, we separate the proteins from the pathogen and host into two groups, and only considered the directed domain-domain interactions - the same interaction pairs have the same direction either in host-pathogen or pathogen-host. As most genes in pathogen are only work for the cell basic function, these genes are filtered to reduce the test dataset base on their notations. When a pathogen factor interacts with its hosts, it should be out of the cell, on the cell membrane or in the matrix. Therefore, these genes that have secretion signal sequence or located on cell membrane are kept, no matter what their notation are. The algorithm was implemented with Python.

2. The predicted protein-protein interactions in *P. Sojae*

Three different methods used to predict the *P. sojae* and its host soybean protein-protein interactions, and the results summarized in Table 1. It is clear that the method based on homology only fond too few interaction pairs, whereas previous DDI method fond too many interaction pairs. The modified DDI method, our new method, can reduce the interaction pairs to limit the false

positive. False positive PPIs are common in experiments and predictions. As the proteins in a cell are separated in different locations and expressed in different times, a PPI prediction only based on their sequences or structures cannot avoid predict false positive interactions – predicted interacted-proteins are in different locations and/or expressed in different conditions. In order to reduce these false positive, a predict model must include other information about the protein profile, such as the cell organ location, expression stage, etc. For pathogen-host PPI, one of important cues to predict the PPI is the components of the proteins, such as the structure domain. In this modified method, the domain specificities are considered, and it significantly reduced more than 90% false positive.

Table 1. The output discovered domain interactions with different methods

	Homology	Non-Dirc-DDI (cutoff >=0.003)	Dirc-DDI (cutoff >=0.003)
G. max-P. sojae PPI	56513	2384395	145596
P. sojae protein (Total 26584)	946	8826	596
G. max protein (Total 73320)	6406	34678	18142

3. Functions of discovered candidate effectors

Protein domains, usually very conserved, are the basic building block of proteins, and each domain has a simple function. We used domain function annotation to understand the functions of candidate effectors. For effectors, the protein domains have functions including cell adhesion, protein Kinase domain, ATPase domain of histidine kinase, membrane transport, while for soybean proteins, most domain functions include Zinc-finger domain, WD40-repeat domain for signal transduction, cell membrane proteins. In all candidates of effectors in *P. sojae,* the most commonly used domain is PF12796, the ankyrin repeat domain, whose function is for protein–protein interactions in eukaryotes. The second most used domain in *P. sojae* effectors is PF00400, which has a variety of functions ranging from signal transduction and transcription regulation to cell cycle control, autophagy and apoptosis. Another ranked in top 10 domain in *P. sojae* candidates but used less in *soybean* is PF00027, Cyclic nucleotide-binding domain, which binds cyclic nucleotides (cAMP or cGMP). The difference of protein domain components between pathogen and its host indicates that, in the pathogen and host interactions, pathogens have favorite weapons to attack special host targets.

We also adopted Gene Ontology (GO) database to annotate protein functions. Enriched GO terms by predicted *P. sojae* effectors are shown in Figure 1. These GO terms have functions relevant to Binding functions (GO:0005515, GO:0046906, GO:0020037), Kinase (GO:0019200, GO:0004396, GO:0008865), Ligase activity (GO:0016879, GO:0016881). The most enriched GO terms by *P. sojae* pathogens are structural constituent of ribosomes (GO:0003735) and structural molecule activity (GO:0005198). The results indicate that *P. sojae* pathogens interfere the genetic information flowing by using protein domains with functions of cell signals and DNA binding.

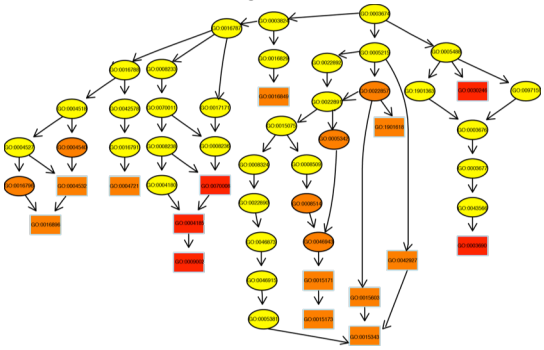

Figure 1. GO terms, shown in orange and red, are enriched by discovered effectors.

4. Validations to discovered effectors

To validate the candidate effectors that we discovered, we analyzed the gene expression profiles

for both soybean and *P. sojae* genes from both microarray assays and RNA-seq experiments. From the analysis of the Affymetrix dataset GSE7124 from GEO database, 4355 expressed *P. sojae* genes were identified, and 183 of these genes were identified by our method as candidate effectors. We identified a *P. sojae* protein (ID PS258303) as a putative effector, which has been studied as Avh295 (Wang et al., 2011), and we predicted this protein has interaction with soybean protein Glyma16g22880. After analyzed microarray data for 3-day infection, we discovered that soybean genes, Glyma16g22880, Glyma03g29230 and Glyma03g14740, were significantly up-regulated.

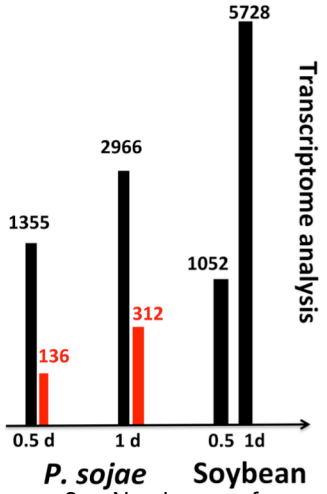

Figure 2. Numbers of genes expressed in *P. sojae* and soybean at different stages (black bars). The numbers of expressed candidate effector genes are shown as red bars.

We also analyzed RNA-seq data for different infection stages based on a 3'-tag digital gene expression protocol. The GEO accession ID is GSE29651. The numbers of expressed genes in *P. sojae* and soybean at different stages are shown in Figure 2. For both Soybean and *P. sojae,* more genes are expressed at 1d after infection than those at 0.5 day. Especially, more number of the predicted candidate effector genes is expressed in 1 day after infection than 0.5 day. This indicates the discovered candidate effector genes mainly are real effectors, because they are highly expressed by the pathogens after 1 day infection. To conduct the analysis of this transcriptomic dataset, we developed a new bioinformatics tool, which was published in 2015 (Dou et al., 2015).

We also compared our discovered effectors and the interactions between them and soybean proteins with the experimental results. *Rps* genes are soybean genes that respond to *P. sojae* infection. Lin et al. identified the *Rps* genes by sequencing, analyzing, and comparing the transcriptomes of 10 near soybean isogenic lines (Lin et al., 2014). They identified 40 *Rps* interactions with *P. sojae* proteins. We predicted that 1156 *P. sojae* genes have interactions with 35 *Rps* genes, including ETR2, MPK4, BAK1, NPR1-like, JAZ1, JAZ6, JAZ8, JAZ12 and EBF1. In all *Rps* genes, FLS2 family has the most interactions with our discovered effectors, including 589 proteins in *P. sojae*. The FLS2 gene Glyma08g08810 has two domains PF00069.20 and PF13504.1 in Pfam IDs. The domain PF00069.20 is a protein kinase domain (Nye et al., 2005), and the domain PF13504.1 is a LRR (leucine-rich repeat) domain. Both domains are found to be involved in plant pathogen reaction genes (Christopoulou et al., 2015).

5. Genome comparison

We applied our new method to other four Phytophthora strains*: P. capsici*, *P. cinnamoni*, *P. infestans,* and *P. ramorum.* Based on the directional domain-domain interactions in all five Phtophthora strains and their hosts, there are 746 protein-protein interaction pairs predicted in which most pathogen proteins have one or more ankyrin (ANK) domains, which mediate protein-protein interactions (Mosavi et al., 2004). Among these 746 pathogen proteins, proteins with IDs of PS502668, PS262839, PS258303 in *P. sojae*, PR86674 in *P. ramorum*, PCA93010 in *P. capsici*, PCI245732 in *P. cinnamomi*, and PI01937 in *P. infestans* are homologies of our discovered effectors and other Phytophthora effectors obtained from the Uniprot database.

One of the largest conserved Phytophthora–host protein-protein interaction network in all the five strains has pathogen proteins with cell surface domains and host proteins with integral membrane domains. All the pathogen proteins have the domain with Pfam ID PB003914, which was found in protein DTFA (Uniprot ID Q54HK7) in *Dictyostelium discoideum*, and had been reported as a cell surface protein important for both cell adhesion and cell sorting (Ginger et al., 1998). All host proteins in the cluster have the domain PF04893, which is an integral membrane domain containing four transmembrane alpha helices, and involving vesicular transport and

GTPases interaction (Yang et al., 1998). Another PPI network found in all the five species composes NmrA family proteins (PF05368) in Phytophthora, which is a negative transcriptional regulator modifies the transcription factor AreA and controlling nitrogen metabolite repression in fungi (Stammers et al., 2001).

There is a specific PPI network in *P. sojae,* shown in Figure 3. This network includes 14 proteins from *P. sojae, which* are predicted to have interactions with soybean kinases, including stress-induced receptor kinases Glyma13g09794 and Glyma17g32782. These 14 proteins have a Pfam domain PB003732 and a Pfam domain PF00589.17, which covalently linked to the DNA through a catalytic tyrosine residue at the carboxyl end of the alignment and cleave DNA substrates by a series of staggered cuts. These 14 proteins are homologous to *P. sojae* effector Avr4/6 (Iliades et al., 2004), but they do not have the RXLR-dEER domain (Dou et al., 2010). Interestingly, they all located near the known Avr4/6 effectors on the *P. sojae* genome sequence.

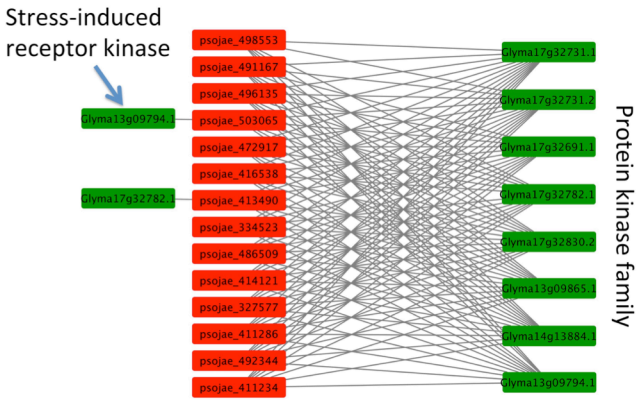

Figure 3. The *P. sojae* specific protein interaction network. Red notes are *P. sojae* and green notes for soybean proteins.

Outcomes

- PI Chi Zhang gave a talk during NE Soybean Board meeting on July 13th, 2016, Broken Bow, NE.
- Based on our research, we edited three manuscripts: one was published, one has been accepted, and the other is under preparation.
  1. Y. Dou, X. Guo, L. Yuan, D.R. Holding, C. Zhang. Differential Expression Analysis in RNA-Seq by a Naive Bayes Classifier with Local Normalization. BioMed Research International (2015) Article ID 789516, doi:10.1155/2015/789516.
  2. Y. Dou, S. Li, W. Yang, K. Liu, Q. Du, G. Ren, B. Yu, C. Zhang. Discovery of Circular RNAs in Arabidopsis Thaliana. *Current Genomics.* (Accepted)
  3. T. Lu, J. Alfano, C. Zhang. Pathogen-host protein-protein interaction of *Phytophthora* based on directional domain-domain interaction. (Preparation)
- We also designed a webpage to show relevant information there: http://sysbio.unl.edu/resources
- As a PI, Chi Zhang submitted several proposals to NSF this year, including one proposal to the ABI program and one to EDEGE track of IOS NSF. All of them are pending.

References:

Christopoulou, M., Wo, S.R.-C., Kozik, A., McHale, L.K., Truco, M.-J., Wroblewski, T., and Michelmore, R.W. (2015). Genome-Wide Architecture of Disease Resistance Genes in Lettuce. G3: Genes|Genomes|Genetics *5*, 2655-2669.

Dong, S., Qutob, D., Tedman-Jones, J., Kuflu, K., Wang, Y., Tyler, B.M., and Gijzen, M. (2009). The Phytophthora sojae avirulence locus Avr3c encodes a multi-copy RXLR effector with sequence polymorphisms among pathogen strains. PLoS One *4*, e5556.

Dou, D., Kale, S.D., Liu, T., Tang, Q., Wang, X., Arredondo, F.D., Basnayake, S., Whisson, S., Drenth, A., Maclean, D., and Tyler, B.M. (2010). Different domains of Phytophthora sojae effector Avr4/6 are recognized by soybean resistance genes Rps4 and Rps6. Mol Plant Microbe Interact *23*, 425-435.

Dou, Y., Guo, X., Yuan, L., Holding, D.R., and Zhang, C. (2015). Differential Expression Analysis in RNA-Seq by a Naive Bayes Classifier with Local Normalization. BioMed Research International *Article ID 789516*.
Ginger, R.S., Drury, L., Baader, C., Zhukovskaya, N.V., and Williams, J.G. (1998). A novel Dictyostelium cell surface protein important for both cell adhesion and cell sorting. Development *125*, 3343-3352.
Iliades, P., Walker, D.J., Castelli, L., Satchell, J., Meshnick, S.R., and Macreadie, I.G. (2004). Cloning of the Pneumocystis jirovecii trifunctional FAS gene and complementation of its DHPS activity in Escherichia coli. Fungal Genet Biol *41*, 1053-1062.
Kamoun, S. (2006). A catalogue of the effector secretome of plant pathogenic oomycetes. Annu Rev Phytopathol *44*, 41-60.
Lin, F., Zhao, M., Baumann, D.D., Ping, J., Sun, L., Liu, Y., Zhang, B., Tang, Z., Hughes, E., Doerge, R.W., Hughes, T.J., and Ma, J. (2014). Molecular response to the pathogen Phytophthora sojae among ten soybean near isogenic lines revealed by comparative transcriptomics. BMC Genomics *15*, 1.
Morgan, W., and Kamoun, S. (2007). RXLR effectors of plant pathogenic oomycetes. Curr Opin Microbiol *10*, 332-338.
Mosavi, L.K., Cammett, T.J., Desrosiers, D.C., and Peng, Z.Y. (2004). The ankyrin repeat as molecular architecture for protein recognition. Protein Sci *13*, 1435-1448.
Nye, T.M.W., Berzuini, C., Gilks, W.R., Babu, M.M., and Teichmann, S.A. (2005). Statistical analysis of domains in interacting protein pairs. Bioinformatics (Oxford, England) *21*, 993-1001.
Stammers, D.K., Ren, J., Leslie, K., Nichols, C.E., Lamb, H.K., Cocklin, S., Dodds, A., and Hawkins, A.R. (2001). The structure of the negative transcriptional regulator NmrA reveals a structural superfamily which includes the short-chain dehydrogenase/reductases. Embo J *20*, 6619-6626.
Tyler, B.M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R.H.Y., Aerts, A., Arredondo, F.D., Baxter, L., Bensasson, D., Beynon, J.L., Chapman, J., Damasceno, C.M.B., Dorrance, A.E., Dou, D., Dickerman, A.W., Dubchak, I.L., Garbelotto, M., Gijzen, M., Gordon, S.G., Govers, F., Grünwald, N.J., Huang, W., Ivors, K.L., Jones, R.W., Kamoun, S., Krampis, K., Lamour, K.H., Lee, M.-K., McDonald, W.H., Medina, M., Meijer, H.J.G., Nordberg, E.K., Maclean, D.J., Ospina-Giraldo, M.D., Morris, P.F., Phuntumart, V., Putnam, N.H., Rash, S., Rose, J.K.C., Sakihama, Y., Salamov, A.A., Savidor, A., Scheuring, C.F., Smith, B.M., Sobral, B.W.S., Terry, A., Torto-Alalibo, T.A., Win, J., Xu, Z., Zhang, H., Grigoriev, I.V., Rokhsar, D.S., and Boore, J.L. (2006). Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science *313*, 1261-1266.
Wang, Q., Han, C., Ferreira, A.O., Yu, X., Ye, W., Tripathy, S., Kale, S.D., Gu, B., Sheng, Y., Sui, Y., Wang, X., Zhang, Z., Cheng, B., Dong, S., Shan, W., Zheng, X., Dou, D., Tyler, B.M., and Wang, Y. (2011). Transcriptional programming and functional interactions within the Phytophthora sojae RXLR effector repertoire. Plant Cell *23*, 2064-2086.
Win, J., Morgan, W., Bos, J., Krasileva, K.V., Cano, L.M., Chaparro-Garcia, A., Ammar, R., Staskawicz, B.J., and Kamoun, S. (2007). Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. Plant Cell *19*, 2349-2369.
Xiao, K., Kinkel, L.L., and Samac, D.A. (2002). Biological control of Phytophthora root rots on alfalfa and soybean with Streptomyces. Biological Control *23*, 285-295.
Yang, X., Matern, H.T., and Gallwitz, D. (1998). Specific binding to a novel and essential Golgi membrane protein (Yip1p) functionally links the transport GTPases Ypt1p and Ypt31p. Embo J *17*, 4954-4963.