Structural and Functional Characterization of a *Xanthomonas* Type III Effector

Shuchi Wu

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Horticulture

Bingyu Zhao, Committee Chair
Eric Beers
Boris A. Vinatzer
John M. McDowell

April 20th, 2015
Blacksburg, VA

Keywords: Rice leaf streak disease, *Xanthomonas*, Stomatal defense, Cystatin

Structural and Functional Characterization of a *Xanthomonas* Type III Effector

Shuchi Wu

ABSTRACT

Rice bacterial leaf streak disease caused by *Xanthomonas oryzae* pv. *oryzicola* (*Xoc*) is one of the most important rice bacterial diseases. *Xanthomonas* type III effector gene *avrRxo1* is conserved in diverse *Xoc* strains and its homologues have been identified from several other gram-negative bacteria species such as *Burkholderia* and *Acidovorax*. In this research, we studied the protein structure of AvrRxo1 and illustrated its virulence mechanism.

We determined the three-dimensional structure of the complex of AvrRxo1 and its cognate chaperone Arc1 (AvrRxo1 required chaperone 1). The AvrRxo1: Arc1 complex is structurally similar to the Zeta-epsilon family of toxin: antitoxin systems from the human bacterial pathogen *Streptococcus pyogenes*. AvrRxo1 and Arc1 have toxin: antitoxin-like activity in bacteria and the toxin activity of AvrRxo1 is required for its virulence function *in planta*. These findings suggest that AvrRxo1 evolved from an endogenous bacterial toxin-antitoxin system.

Furthermore, AvrRxo1 was shown to have virulence functions in diverse host plants including *Arabidopsis thaliana*. The ectopic expression of wild type *avrRxo1* in *Arabidopsis* suppresses plant basal defense. *AtVOZ* (*Arabidopsis* **v**ascular **o**ne **z**inc-finger transcription factor), which has two homologues in the *Arabidopsis* genome, *VOZ1* and *VOZ2*, was identified as one of AvrRxo1's candidate interactor. The knockout of *voz1/voz2* renders the plants more susceptible to the virulent pathogen *Pseudomonas syringae* pv. *tomato* (*Pst*) DC3000, but compromises the virulence function of AvrRxo1. The expression profiling of transgenic *Arabidopsis* plants expressing the *avrRxo1* gene allowed us to identify *Arabidopsis* genes regulated by AvrRxo1 and VOZ1/2. AvrRxo1 interacts with and stabilizes VOZ2 *in vivo* and directly binds to the promoter region of *AtCYS2* (*Arabidopsis* phyto**CYS**tatin 2) to induce its expression. The overexpression of *CYS2* in increased stomatal aperture size, and enhanced plant susceptibility to *Pst*. Therefore one of AvrRxo1's virulent functions is to regulate the expression of *CYS2* by manipulating VOZ2, resulting in increased stomatal aperture. Presumably, this renders the host leaf more susceptible to colonization via the stomata.

Another component of my dissertation was based on a genome-wide survey of *Arabidopsis* papin-like cysteine protease genes (PLCPs). The *Arabidopsis* genome has 31 PLCP and 7 cystatin genes, and they often worked in pairs to regulate signaling pathways in response to biotic and abiotic stress. The coordinated transcriptional regulation of all *Arabidopsis* PLCP and cystatin genes has never been systematically investigated. In order to unveil the mechanism of stomata-related plant immunity regulated by CYS2, we analyzed the expression patterns of 28 PLCPs and 7 cystatins in *Arabidopsis* in response to biotic or abiotic stress, by reprocessing and integrating microarray data from the AtGenExpress database. We also performed enzyme assays and evaluated the inhibition specificity of seven cystatins to the five most abundant PLCPs in *Arabidopsis*. Finally, we utilized the SVMs (support vector machines) package in R software to predict a functional network of PLCP-cystatin interplay in *Arabidopsis*. We identified the PLCP protein PAP4 as one of the putative targets of CYS2. The co-expression profiling indicated that the expression patterns of *PAP4* and *CYS2* were strongly correlated during virulent bacterial infection, and weakly correlated under drought stress. Therefore, *PAP4* was determined to be a promising gene in regulating stomatal aperture size. Further research on the interplay of PAP4-CYS2 could be important for understanding AvrRxo1's virulence mechanism and regulation of plant stomatal immunity.

Acknowledgements

I would like to express my sincere appreciation for my advisor, Dr. Bingyu Zhao, for his mentorship and advice during my graduate studies. I feel grateful to have the opportunity to work in his lab. Under his guidance, I systematically learned plant science and molecular biology techniques. And far beyond the knowledge and experimental skills, I was trained to be a capable scientist who is an independent researcher, and at the same time, also a compatible cooperator. In addition to being a research mentor, Dr. Bingyu Zhao has also been a life mentor. I will always remember his suggestions on magnanimity, responsibility and optimism. Dr. Bingyu Zhao has also been a great friend. I will never forget his Christmas chocolate and Spring Festival parties every year, especially the delicious steam breads. And I am grateful to all the members of Dr. Bingyu Zhao's Lab. I would like to thank Dr. Changhe Zhou and Dr. Bin Xu for teaching me many experimental skills. And I would like to thank Taylor Frazier who helped me with English writing and proof reading. Additionally, I would like to thank all my colleagues in the MPS program and other departments in Virginia Tech. I would like to thank Dr. Song Li, who helped me with bioinformatics skills. I would like to thank Hardik Zatakia for helping with EMSA. And I would like to thank Dr. Gunjune Kim and Dr. Alan Dickerman for helping with RNAseq data analysis.

I also would like to thank my committee members: Dr. Eric Beers, Dr. Boris Vinatzer and Dr. John McDowell. They were the best advisory committee and their advice and encouragement was very helpful for my research and education. Especially the Plant Physiology class taught by Dr. Eric Beers and the Plant Pathology class taught by Dr. Boris Vinatzer and Dr. John McDowell, which led me, as an almost outsider in to the plant science world.

I would like to specifically thank my former advisor, Dr. Fabricio Medina-Bolivar, who initiated my scientific training. Many of my good habits for scientific research including the spirit of suspicion, perfection and enterprise, arose from his foresight and guidance.

I would like to thank my family and friends for their love, support and encouragement, especially my wife Wenjun Zheng and my mother Lan Zeng, who have been extremely supportive when I was frustrated. I love you all!

I feel lucky and grateful to have pursued my PhD degree in Horticulture Department of Virginia Tech, where I had been offered a supportive, constructive and very professional research environment. I would like to express my appreciation to the Head of Horticulture Department, Dr. J. Roger Harris, who provided my Teaching Assistance opportunity and other financial supports.

Attribution

List of co-authors

Bingyu Zhao, Department of Horticulture, Virginia Tech
Jianrong Li, Department of Biochemistry, Virginia Tech
Jan E. Leach, Department of Bioagricultural Sciences and Pest Management, Colorado State University
Howard Robinson, Biology Department, Brookhaven National Laboratory
Loïc Deblais, Department of Bioagricultural Sciences and Pest Management, Colorado State University
James Tokuhisa, Department of Horticulture, Virginia Tech
Jiamin Miao, Department of Horticulture, Virginia Tech
Lindsay Triplett, Department of Bioagricultural Sciences and Pest Management, Colorado State University
Yi Liu, Department of Horticulture, Virginia Tech
Shuchi Wu, Department of Horticulture, Virginia Tech
Changhe Zhou, Department of Horticulture, Virginia Tech
Qian Han, Department of Biochemistry, Virginia Tech
Qiang Cheng, Department of Horticulture, Virginia Tech
Bin Xu, Department of Horticulture, Virginia Tech
Zhengxing Shen, Department of Horticulture, Virginia Tech
Zhiyong Yang, Department of Horticulture, Virginia Tech

Contributions of co-authors

Chapter II

Bingyu Zhao and Jianyong Li conceived and designed the experiments and will be corresponding authors.

Qian Han and Changhe Zhou processed the protein crystallization and structural analysis experiments. Shuchi Wu processed the enzyme activity assay and *in planta* experiments. These three authors contributed equality to the project and will be the co-first authors.

Yi Liu and Jiamin Miao processed the protein-protein interaction experiments. James Tokuhisa provided advising on enzyme activity assay. Lindsay Triplett, Loïc Deblais, Howard Robinson and Jan E. Leach processed the *Xanthamonas* related experiments and provided the source of the effector protein. They will be the co-authors.

Chapter III

Bingyu Zhao conceived and designed the experiments and will be the corresponding author.

Shuchi Wu contributed to the execution of most experiments, data analysis and paper writing. Changhe Zhou processed the initial experiments and also contributed to the RNA-seq experiment. They will be the co-first authors.

Yi Liu, Qiang Cheng and Zhiyong Yang processed the protein-protein interaction experiments. Bin Xu and Zhengxing Shen contributed to the T-DNA line establishment and the initial bacterial growth curve assays. They will be co-authors.

Chapter IV

Bingyu Zhao conceived and designed the experiments and will be the corresponding author.

Shuchi Wu contributed to most experiments, data analysis and paper writing. He will be the first author.

Yi Xing processed the initial data acquisition, filtering and organization. Song Li provided advice on bioinformatics. They will be co-authors.

Table of Contents

Chapter I: Literature Review 1
Plant Cysteine Protease: a Double-edged Sword in the Seesaw Battle between Plants and Microbes 1
Summary 1
Part 1: Microbial pathogen-plant interaction: the endless warfare 1
Part 2: Plant cysteine proteases: Important hubs in pathogen-plant interaction ..10
References 18
Chapter II: 29
Crystal structure of the complex between *Xanthomonas* AvrRxo1-ORF1, a type III effector with a polynucleotide kinase domain, and its interactor AvrRxo1-ORF2 .29
Summary 29
Introduction 30
Results and discussions 31
Concluding remarks 42
Material and methods 42
Acknowledgements 46
Supplementary information 47
References 53
Chapter III: 58
Xanthomonas effector AvrRxo1 suppresses plant immunity by regulating the plant stomatal aperture size 58
Abstract 58
Introduction 58
Results 60
Discussions 72
Concluding Remarks 75
Material and methods 77
Acknowledgements 84
Supplementary information 85
References 91
Chapter IV: 98
Co-expression analysis of *Arabidopsis* papin-liked cysteine proteases and cystatins in abiotic and biotic stress conditions 98
Abstract 98
Introduction 98
Results and discussions 102
Conclusions and perspectives 116
Material and methods 116

Acknowledgements 120
Supplementary information 120
References.....123

List of Figures

Figure 1-1| Distribution of protease related publications in NCBI database.....11
Figure 1-2| Distribution of publications of cysteine proteases associated with plant programmed cell death.....12
Figure 2-1| Assembly of AvrRxo1-ORF1:AvrRxo1-ORF2 complex and regions having significant conformational changes revealed by a comparison of four complex structures 33
Figure 2-2| Structural fold and putative active site residues of AvrRxo1-ORF1 35
Figure 2-3| Binding sites of phosphate, sulfate ions and ATP molecule in the AvrRxo1-ORF1 active center 36
Figure 2-4| Polynucleotide substrate binding in T4pnk and AvrRxo1-ORF1.....37
Figure 2-5| Functional assay of wild type and mutant AvrRxo1 39
Figure 2-6| Structural fold and 3D alignment with a similar structure of AvrRxo1-ORF2 41
Figure S2-1| Gel filtration chromatography of AvrRxo1-ORF1: -ORF2 complexes 47
Figure 3-1| AvrRxo1 enhances Pseudomonas bacterial virulence and ectopic expression of AvrRxo1 suppressed plant basal defense response 62
Figure 3-2| AvrRxo1 interacts with AtVOZ1/2 that is required for its virulence function 65
Figure 3-3| AvrRxo1 suppressed VOZ2 degradation *in vivo* 67
Figure 3-4| Luciferase reporter assay indicated wild type AvrRxo1 significantly induces the expression of *Arabidopsis* cystatin gene *CYS2*.....69
Figure 3-5| Electrophoretic mobility shift assay indicated VOZ2 directly binds to the promoter region of *CYS2*.....70
Figure 3-6| Overexpression of CYS2 in *Arabidopsis* increased stomatal aperture size and suppressed plant basal defense against *P. syringae*.....71
Figure 3-7| Model of the virulence mechanism of AvrRxo1.....76
Figure S3-1| AvrRxo1 interacts with OsVOZ2 *in vitro*.....85
Figure S3-2| RT-PCR indicated the compromising of *VOZ1/2* expression in *Arabidopsis* T-DNA lines 86
Figure S3-3| The complementary of *VOZ1* or *VOZ2* expression suppressed the chlorosis phenotype of *voz1/2*, and restored the virulence activity of AvrRxo1.....87
Figure S3-4| Co-expression of AvrRxo1 and VOZ2 in *Arabidopsis* suppressed VOZ2 degradation in *vivo*.....88
Figure S3-5| Pst DC3000-ΔCEL carrying pBZ598-*avrRxo1* induced the expression of *CYS2* in wild type *Arabidopsis* Col-0, but not the *voz1/2* mutant plants 89
Figure S3-6| Increased *Arabidopsis* stomatal aperture size by *Pst* DC3000-ΔCEL carrying pBZ598-*avrRxo1*, or the overexpression of *VOZ2*.....90

Figure 4-1| Expression profiles of seven cystatins and the five most abundant PLCPs of *Arabidopsis* plants infected with *Pseudomonas syringae* ES4326 104
Figure 4-2| Co-expression patterns of *Arabidopsis* PLCPs and cystatins and clustering of PLCPs in biotic stress condition 107
Figure 4-3| Co-expression patterns of *Arabidopsis* PLCPs and cystatins and clustering of PLCPs in wounding or drought stress condition 111
Figure 4-4| Enzyme activities of the five most abundant *Arabidopsis* PLCPs and the enzyme inhibition ability of seven cystatins 113
Figure 4-5| PLCP-cystatin interaction network and the expression profiles of *CYS2* and *PAP4* under drought stress 115
Figure S4-1| Expression levels of 28 PLCP genes in *Arabidopsis* shoot and root tissues 120

List of Tables

Table 1-1 | Relative expression levels of 31 *Arabidopsis* PLCP genes in different plant organs 14
Table S2-1| Lists of hydrogen bonds and salt bridges in the interaction interfaces in the N-AvrRxo1-ORF1:-ORF2 complex 48
Table S2-2| Lists of hydrogen bonds and salt bridges in the interaction interfaces in the ATP-AvrRxo1-ORF1:-ORF2 complex.....50
Table S2-3| HADDOCK score and other parameters of the docking pose used in the comparison study 52
Table S2-4| Sequences of primers used in this study.....53
Table S3-1| Sequences of primers used in this study.....91
Table S4-1| Gene names and TAIR IDs for PLCP, cystatin and housekeeping genes used in this study.....121
Table S4-2| Predicted relationship between seven cystatins and 28 PLCPs 122
Table S4-3| Sequences of primers used in this study.....123

Chapter I: Literature Review

Plant Cysteine Protease: a Double-edged Sword in the Seesaw Battle between Plants and Microbes

Shuchi Wu[1] and Bingyu Zhao[1*]
[1]Department of Horticulture, Virginia Tech, Blacksburg, VA, 24061

Summary

Judging from physical size, the battle between a plant and microbes is comparable to a fight between an elephant and ants, and the microbes have no chance to win. However, in reality there are always microbial pathogens that can break down plant immunity and successfully colonize and cause disease in host plants. To overcome plant immunity, microbial pathogens evolved at least two kinds of survival strategies: first of all, microbes evolve quickly to escape from host recognition; secondly, successful microbial pathogens deploy a collection of virulence effectors into plant cells to suppress host immunity, or even turn host defenses against itself. Plant cysteine proteases play essential roles in programmed cell death, which can efficiently restrict the proliferation of microbial pathogens, but the programmed cell death can also cause severe damage to the host cell itself. To regulate this "suicide" weapon, plants evolved various cysteine protease inhibitors to tightly regulate protease activity. Here, I review the studies of plant cysteine proteases and their cognate inhibitors, and discussed how plantss utilize the protease/ inhibitor system for defense against microbial pathogens, and how microbial pathogens hijack this mechanism.

Part 1: Microbial pathogen-plant interaction: the endless warfare

The First Encounter: Pathogen entry and plant innate immunity

Just like an ant biting an elephant, invading a plant host is never an easy task for microbial pathogens. Over time, plants evolved several levels of defenses. Most healthy plants can effectively prevent most microbes from getting inside of plant tissue, and limit microbial colonization on the leaf surface as epiphytes [1, 2]. On the plant leaf surface, bacteria can often proliferate to 10^6 - 10^7 cells/cm^2[3, 4]. The nutrient poor plant surface is not a “hospitable” environment for most microbes [2]. The epiphytes on leaf surfaces can be easily washed off by heavy rain or killed by intensive UV light [5]. The dramatic fluctuation of temperature and humidity on the leaf surface also casts severe stress on most microbial cells [2]. In order to survive on the leaf surface, some microbes were reported to develop different

strategies to "modify" their living conditions. For example, *Pseudomonas* was found to secrete toxins that increase plant nutrient secretion, and improve the lubricity of the leaf surface [6]. Syringomycin, a phytotoxin secreted by *Pseudomonas syringae* (*P. syringae*), can induce plant ion channels and increase the release of plant metabolites to the leaf surfaces [7-9]. *Pseudomonas tolaasii* was reported to secrete a biosurfactants, tolaasin, allowing bacteria to spread across the rough leaf surface, and reach nutrient rich areas such as wounds or natural openings [10]. Compared to the "barren" surface, the interior plant mesophyll tissue is a much more “fertile” land for microbial pathogens [11]. Therefore, microbial pathogens are always eager to break though the waxy plant surface, to access host nutrients [12, 13].

For microbial pathogens, the first step of invasion is to break through or bypass the waxy and compact surface of host plants [11-13]. Microbial pathogens evolved a number of different strategies to enter host plant tissue, either directly penetrating through the plant surface or entering through natural openings such as wounds, stomata, or hydathode [11]. Fungal and oomycete pathogens can directly penetrate through plant leaf surfaces [12, 13]. During evolution, fungal and oomycete pathogens developed a set of elaborate piercing structures to break through plant surfaces and form infectious structures inside of mesophyll tissues [12]. A typical penetration process of fungal and oomycete pathogens is initiated with germination of germination tube from a spore landing on a plant surface [14]. After a short time of elongation, the end of the germination tube will differentiate into an appressorium cell that can further develop into a special hypha, the penetration peg [14]. The penetration peg will continually grow to break through the plant surface and extend under the plant epidermal cell [14]. Eventually, the penetration peg will differentiate into diverse specialized hyphae like invasive hypha and haustorium, which help the pathogen proliferate in the plant apoplast and absorb nutrients from plant cells [11, 14].

Unlike fungal or oomycete pathogens, bacterial pathogens do not have such elaborate piercing structures to directly penetrate the host plant surface. Therefore, they have to enter through natural openings such as surface wound sites, stomata or hydathodes [15, 16]. Wounding sites are rare on plant surfaces of a healthy plant, and therefore, the more common natural openings, such as stomata, become the major entry sites for most bacterial pathogens [15, 16]. Thus, it is not surprising that both bacterial pathogen and plants evolved diverse strategies to fight for control of stomatal closure and opening [15-17]. For most plant species, stomatal aperture size is controlled by two guard cells [18, 19]. Plants reduce aperture size to inhibit water loss and increase aperture size to promote gas exchange [18, 19]. During bacteria invasion, the host plant immune receptor FLS2 recognizes flg22, a conserved domain of bacterial flagellin to trigger defense responses involved in

salicylic acid (SA) and abscisic acid (ABA) signaling pathways [20]. The defense signal activates guard cells to close the stomatal aperture. While bacterial chemotaxis allows bacterial pathogens to swim toward and accumulate around the guard cells [17, 20]. As previously described, *Pseudomonas* secrete biosurfactants to facilitate swimming toward stomata [10]. Melotto et al. suggested that *P. syringae* also has the ability to move towards open stomata and avoid closed stomata [20]. Besides the ability of moving towards open stomata, some plant bacterial pathogens also evolved the ability to induce the reopening of stomata [17, 20, 21]. Mittal and Davis (1995) and others reported that *P. syringae* can secrete a small compound, called coronatine (COR), which may be essential for bacteria to get inside of host plant mesophyll tissue [21] [22]. Mutant *P. syringae* (*cor*-) lost the ability of secreting coronatine and reduced virulence when spray-inoculated on the leaf surface [21]. However, if the mutant *P. syringae* (*cor*-) was directly infiltrated into plant mesophyll tissue, virulence was fully recovered [21]. Therefore, the authors proposed that COR may be required for bacteria to enter through stomata cells. This hypothesis was proved by Melotto et al. (2006) and other who demonstrated that coronatine is a structural mimic of the jasmonic acid (JA) precursor, that functions through COI1, an E3 ligase subunit, to inhibit stomatal closure [20, 21, 23]. In plant cells, JA and SA signaling pthways are usually antagonistic [1, 17]. SA has key roles of signaling the stomatal closure [17]. Coronatine hijacks JA signaling, and therefore, could indirectly affect the SA signaling to suppress stomata closing [1, 17]. Successfully getting inside of host plant mesophyll tissue and colonize the apoplastic space is only the first step of pathogen invasion. After the loss of the first encounter on the leaf surface, host plants also evolved several other layers of systemic defense strategies to fight against microbial pathogens in the apoplast [1, 11].

Battle in the Apoplastic Space: Pathogen-associated molecular pattern-triggered immunity and effector triggered susceptibility

Unlike animal and human pathogens, plant microbial pathogens usually inhabit in the apoplastic space, but do not invade the inside of cells [11]. Therefore, most interactions between microbial pathogens and host plant cells take place in the apoplastic space [11]. The plant cells infected by pathogens initiate early defense responses by efficiently recognizing invading pathogens, and quickly transduce the alarm signal to the unaffected plant cells [11]. Plants, unlike mammals, do not have specialized immune cells and a somatic adaptive immune system, they mainly rely on innate immunity of each cell [1, 24]. The pattern recognition receptors (PRRs) localized on the surface of plant cells are in involved

in pathogen recognition [25]. PRRs are structurally conserved transmembrane proteins, which include an extracellular leucine-rich repeat (LRR) domain and an intracellular non-arginine–aspartate (non-RD) kinase domain [11, 25]. The LRR domain is in charge of pathogen recognition and the kinase domain serves as trigger of defense signal transduction [25, 26]. Since microbial pathogens evolve fast, plant PRRs recognize the most conserved features or most common components on the surface of pathogens, such as chitin or chitosan (adeacylated derivative of chitin) of fungi and elicitin from oomycetes, or polysaccharides and flagellin of bacterial species [11, 27, 28]. The most intensively studied PRR is FLS2 of *Arabidopsis*, which is a good example of how plants recognize conserved molecular patterns of bacterial species. FLS2 recognizes a 22 amino acid long peptide, flg22, that is conserved in most bacterial flagellin proteins [25]. To trigger downstream defense signaling, FLS2 interacts with at least two kinases: one is the cross-membrane kinase, BAK1 and the other one is the cytoplasmic kinase BIK1 [28]. In the absence of bacteria invasion, the intercellular kinase domains of FLS2 and BAK1 interact with BIK1, respectively [28]. Upon the recognition of flg22 by the LRR domain of FLS2, the intercellular domain of FLS2, BIK1 and BAK1 will form a complex, which triggers phosphorylation of both, FLS2 and BAK1[28]. Subsequently, BAK1 will phosphorylate BIK1, and BIK1 will also promote the phosphorylation of FLS2 and BAK1 as a feedback [28]. The phosphorylated FLS2, BAK1 and BIK1 will trigger the downstream mitogen-activated protein kinase (MAPK)-dependent signal cascade [26, 28]. This plant PRR-mediated innate immune response triggered by the recognition of a pathogen-associated molecular pattern (PAMP) is also called PAMP-triggered immunity (PTI) [1]. Besides recognizing PAMP signals, plant PRRs can also recognize endogenous molecular signals generated by the physical damage of plant components, such as fragmented DNA, proteins, sugars, cell wall components, etc, which are termed damage-associated molecular patterns (DAMPs) [29, 30]. One kind of interesting plant DAMPs are heat shock proteins (HSPs) [31], which normally serve as chaperones that help proteins fold [31]. In immune signaling pathways, several HSPs interact with plant receptors to trigger a plant immune response, which is termed DAMP-triggered immunity (DTI) [30]. DPI can awaken plant immune response under abiotic stress like heat, or mechanical damage like animal biting [30]. On the other hand, DPI serves as supplement to PTI that can strengthen and amplify defense signaling cascades [30]. For example, in *Nicothiana tabacum*, a heat shock protein, Ntshsp17 was reported to be involved in plant defense against *Ralstonia solanacearum* (*R. solanacearum*) [32]. *Ntshsp17* was induced by heat and chemical elicitors such as aminocyclopropane carboxylic acid, hydrogen peroxide, methyl jasmonate, and salicylic acid [32]. Overexpression of *Ntshsp17* enhanced resistance to *R. solanacearum* [32].

Downstream of PTI are at least two major signaling pathways: SA-dependent and JA-dependent pathways in response to different types of pathogens, whereby the two pathways are usually antagonistic to each other [1, 33]. Microbial plant pathogens can be classified as necrotrophs and biotrophs and hemibiotrophs based on their life styles [11]. Necrotrophs kill plant tissue during their invasion and feed on the nutrients released from dead tissue [34]. Biotrophs can only survive in living plant tissues and acquire nutrients from plant metabolism [35]. Hemibiotrophs first colonize living plant tissue but eventually induce plant cell death during late infection [34]. SA-dependent defense signaling is usually triggered by biotrophic and hemibiotropic microbial pathogens [33]. A quick and massive production of reactive oxygen species (ROS) could eventually trigger programmed cell death (PCD), as an effective defense mechanism against the biotrophic and hemibiotropic pathogens [36]. While, the JA-dependent plant defense is usually triggered by necrotrophic pathogens and herbivores [36] [37], which involves the production and secretion of anti-microbial secondary metabolites such as terpenes and phenolic compounds [38, 39]. Besides of SA and JA, other phytohormones are also involved, either positively or negatively, in defense signal transduction pathways [33, 40]. For example, ethylene was reported to enhance SA-dependent defense signaling, resulting in a potentiated expression of the SA-responsive marker gene *PR-1* [39, 41]. ABA can also interact with SA to regulate stomata closure and, therefore, block bacterial envision [42], while auxins may have negative roles in regulating plant defense [43, 44].

The phytohormone-dependent defense signal cascades are essential for triggering and maintaining systemic acquired resistance (SAR), which helps protect the entire plant, including parts distant from the primary infection site[45]. Long distance defense signal transduction helps plants convert the passive defense into active defense against potential microbial pathogens. Although the detailed mechanisms of long distance signal transduction in SAR are not completely understood, the SA-dependent defense signal cascade is believed to play a pivotal role in SAR [1, 45, 46]. *Arabidopsis* mutants abolished either in SA production or SA signaling have reduced SAR upon pathogen infection [45, 47], while transgenic plants that over-produce SA have enhanced plant resistance to various pathogens [47-50]. The SA-dependent defense signaling cascade consists of the upstream ROS production and SA feedback signal-amplification-loop signaling pathways and the downstream outcome is associated with induction of pathogenesis related (PR) genes [51]. Pathogen infection will induce a rapid ROS accumulation, also termed as oxidative burst [1]. Lamb and Dixon (1994) firstly reported this rapid oxidative burst that can be detected a few minutes after PAMP recognition [52]. The primary accumulation of ROS will last only about 60-180 minutes, which is enough to induce SA biosynthesis and the expression of ROS-related genes [52]. A

second phase of ROS production usually begin at 3 to 4 h post pathogen infection, which continuously increases and eventually can trigger the programmed cell death in plant cells [52]. SA participates in the ROS-mediated signal amplification loop, and induce the expression of PR genes through the a key regulator NPR1 [53]. The *Arabidopsis npr1* mutant was identified as loss of the ability to induce PR1 gene expression (Nonexpressor of Pathogenesis Related Proteins 1) after challenge by avirulent pathogens [53]. Pathogen-triggered SA-accumulation is not affected in *npr1* mutant plants, while local overexpression of NPR1 can neither trigger SA accumulation nor PR1 activation in the distant uninfected plant tissue [54]. However, constitutive overexpression of NPR1 in different plant species usually compromises plant development, leads to lesion formation, and plants are significantly more resistant to diverse pathogens [50, 55, 56].

In conclusion, plant PAMP and DAMP-triggered immunity and phytohormones-dependent SAR constitute a tight and systemic innate immune network, which is efficient in resistance to most microbial pathogens.

To survive in the apoplastic space, which is tightly controlled by intertwined plant defense mechanisms, microbial pathogens need resistance to anti-microbial compounds and escape from the recognition by plant PRRs, which allow the microbes to successfully colonize and proliferate in plant mesophyll tissue [11]. Rather than just passive escaping, successful microbial pathogens adapt a more active strategy to suppress host immunity, through delivery of a collection of effectors (virulence factors) into plant cells [1]. Microbial pathogens developed diverse effectors secretion organs or structures, such as the haustorium in fungi and oomycetes [57], and various secretion systems in gram-negative bacteria [58]. Among them, the Type III secretion system (TTSS), which is conserved in many plant bacterial pathogens, has been intensively studied in the past few decades [58-60]. Most effectors are proteins, where one exception is that *Agrobacterium tumefaciens* (*A. tumefaciens*) can deliver T-DNAs (transfer DNA) into plant cells to induce crown gall at the affected tissue [61]. Recently, *Botrytis cinerea* (*B. cinerea*) was also reported to secrete small RNAs as virulence effectors to hijack host plant RNA interfering pathways, and suppress host plant immunity [62]. In this review, we will mainly focus on the virulence mechanism of protein effectors.

Bacterial pathogens can deliver 10-50 effectors into host cells, while *Legionella pneumophila* can secrete over 100 effectors, and the genomes of some oomycetes contain more than 200 putative effector genes [63] [64-67]. For example AvrPtoB, one of Type III effector from *Pseudomonas syringae* pv *tomato* DC3000 (*Pst*DC3000) was identified to degrade *Arabidopsis* FLS2, which is an important initiator of PTI [68]. Gohre et al. (2008) determined that AvrPtoB functions as an E3 ligase, interacting with FLS2 and BAK1 and induces the polyubiquitination of FLS2's kinase domain [68]. The ubiquitinated FLS2 will be

degraded by the plant 26S proteasome degradation pathway [68]. Another example is the *Xanthomonas* effector *avrBsT*, which encodes an acetyltransferase that can modify *Arabidopsis* ACIP1 (for ACETYLATED INTERACTING PROTEIN1) to suppress plant innate immunity [69]. Fungi can also deliver effectors to interfere with PTI [70-74]. Chitin is an essential component of the fungal cell wall that can be recognized by plant PPRs to trigger PTI [75]. In rice (*Oryza sativa*), the receptor protein CEBiP and CERK1 were identified as the PRRs that can recognize chitin through the conserved LysM domains [72, 73]. Bolton et al. (2008), reported that the fungal tomato pathogen *Cladosporium fulvum* can deliver the effector Ecp6 to inhibit chitin-triggered PTI [73]. The fungal effector Ecp6 also contains a LysM domain, and it can compete with CEBiP and CERK1 for chitin-binding. Just like an experienced hunter wiping his footprints, Ecp6 helps pathogens hide their tracea to avoid host recognition [73]. Therefore, pathogen effectors have key roles in suppressing host PTI, which is also termed as effector-triggered susceptibility (ETS). In order to counteract ETS, plants have developed their ultimate weapon: disease resistance (R) proteins [1].

Heavy Resistance: R proteins and programmed cell death

To select disease resistant crop cultivars, plant breeders identified plant germplasm carrying host-specific resistance genes (*R* genes) that are effective against certain crop pathogens [76]. Most *R* genes are dominant and they confer complete resistance to certain pathovars of a given microbial pathogen, although the resistance could be lost due to the dynamics of pathogen populations [76]. To understand the mechanism of *R*-gene mediated disease resistance, Harold Henry Flor proposed the famous “gene-for-gene” hypothesis [77, 78]. When he was studying the flax (*Linum usitatissimum*) rust disease caused by *Melampsora lini*, Flor identified that the plant resistance phenotype determined by a pair of cognate genes, the dominant *R* gene from host plant and dominant avirulence (*avr*) gene from the rust pathogen [77]. It is worth mentioning that the name of "avirulence" does not mean the *avr* genes are the "stupid" products of microbial pathogens in order to suppress their own virulence on host plants. In fact, after intensive research of molecular plant microbe interactions, scientists now confirmed avirulence genes are indeed virulence effectors for overcoming PTI in the susceptible plant genotypes as described previously. The first avirulence gene confirmed to have virulence function on susceptible plant is the *avrBs2* gene cloned from *Xanthomonas campestris* pv. *vesicatoria* [79]. The name “avirulence gene” is now less popular because of its confusion with its virulence function. Instead, these genes are generally named as pathogen effector genes. In response to

ETS, plants evolved *R* genes that specifically recognize the cognate effectors and trigger effector-triggered immunity (ETI) [1]. Most *R* genes encode proteins with conserved nuclear binding (NB) and leucine rich repeat (LRR) domains, which are often named as NB-LRR proteins [1]. The NB domain binds ATP/ADP or GTP/GDP, and the LRR domain may be involved in protein-protein interactions [1]. The NB domain binds ATP/ADP or GTP/GDP, and the LRR domain may involve in protein-protein interactions [1]. NB-LRR proteins can be further classified into TIR-NB-LRR or CC-NB-LRR, based on their unique N-terminal structures: the toll interleukin 1 receptor domain (TIR) of human and *Drosophila* Toll-like receptors and the coiled-coil (CC) [80].

Similarly to PTI, ETI is also involved in specific recognition between host R proteins and pathogen effector proteins [1]. There are two models to explain R-effector protein interaction: (1) the receptor-ligand model to explain the direct interaction between host R and pathogen effector proteins, (2) the guard model to explain the indirect interaction between host R and cognate effector protein, where the pathogen effector can cause modification of host virulence target(s), and the modification can be monitored by R proteins [1]. The ligand-receptor model is supported by several examples, where the pathogen effector directly interacts with the polymorphic LRR domain of the cognate R protein, though recent reports suggest the microbial effector could also interact with the N-terminal CC domains of R proteins [1, 81-88]. For example, the rice R gene *Pi-ta* directly interacts with and recognizes its cognate effector *Avr-Pita* expressed by the rice blast fungus, *Magnaporthe grisea* [81]. The flax R proteins (L5, L6 and L7) also directly interact with rust effectors AvrL5,6,7 [82]. The guard model has been supported by studies of the RPM1/AvrRpm1/RIN4 and RPS2/AvrRpt2/RIN4 systems [1]. RIN4 was originally identified as one of RPM1 interactors [89-91], which was confirmed to be targeted by *Pseudomonas* effector AvrRpm1, AvrB and AvrRpt2 [89-91]. The effectors AvrRpm1 and AvrB interact with RIN4 and promote its phosphorylation and accelerate its degradation [89]. The plant R protein RPM1 monitors the phosphorylation status of RIN4 to initiate ETI [89]. The *Pseudomonas* effector AvrRpt2 encodes a cysteine protease, which can cleave RIN4 *in vivo* [90, 92]. The *Arabidopsis* R protein RPS2 recognizes the cleavage of RIN4, to trigger ETI [90-92]. RIN4 was shown to be not only the target of AvrRpm1 and AvrRpt2, it also functions as negative regulator of Rpm1 and Rps2-mediated ETI [89, 90, 92]. Therefore, RIN4 can be considered as a "bait" to attract multiple effectors. The quick evolution of pathogen effector genes has obvious advantages over the relatively slow evolution of host *R* genes. Thus for the *R* genes, "guarding" the slowly evolving plant target genes is more efficient than directly targeting on the quickly evolving pathogen effector genes.

The co-evolution of effector and *R* genes is like an arm race between pathogen and host that has been described as the "zigzag" model [1]. The "zigzag" model divides the co-evolution of pathogen and host plant into several phases [1]. The first phase is the PTI process, where host PRRs recognize PAMPs or DAMPs of pathogens to trigger innate immunity [1]. The second phase is the ETS process, where successful pathogens develop effectors, and deploy them to suppress host PTI [1]. In the third phase, the host plant develops NB-LRR proteins, which specifically recognize certain effectors to trigger ETI [1]. In the following phase, successful pathogens evolve to escape the recognition of host R proteins, or develop new effectors to suppress ETI [1]. As consequence, host plant will evolve new R proteins to recognize new effectors and trigger ETI again [1]. The relationship between the tomato R protein PTO and the *Pst* effector AvrPtoB provides strong evidence to support the "zigzag" model [95-98]. AvrPtoB is a *P. syringae* effector that suppresses PTI by interacting with BAK1 and BIK1 [96]. It was demonstrated that the N-terminus (1-387aa) of AvrPtoB is sufficient to suppress PTI [96]. This suggests that originally AvrPtoB may only have had the N-terminal functional domain [97]. The tomato R protein FEN recognizes AvrPtoB, and with the help of another NB-LRR protein PRF triggers ETI [98]. Under natural selection, two successful evolutionary events in AvrPtoB allowed it to escape from the recognation by FEN and PRF [98]. One of them is the acquisition of the C-terminal E3 ligase domain. The E3 ligase domain can cause the degradation of Fen [98]. In another pathogen, PsmES4326, AvrPtoB lost the domain that is recognized by Fen resulting in the effector, HopPmaL [98]. In response to the evolution of AvrPtoB, the plant R protein PTO evolved to recognize the E3 ligase domain of AvrPtoB and triggers PTO/PRF-mediated ETI [98].

Compared to PTI, ETI typically culminates in cell death, termed the hypersensitive response (HR) [1]. A typical HR involves the activation of the production of SA, JA, nitride oxide (NO), ROS, and expression of *PR* genes [1]. The HR-like cell death is usually well restricted to the site of infection [1]. It is believed that the well-controlled programmed cell death (PCD) can cut the nutrient supply to pathogens and prevent its further spreading, although the mechanism of stopping bacteria proliferation is still unclear [1]. Many pathogen effectors evolved the ability to suppress PCD [99]. The previously described evolutionary process of AvrPtoB is a good example of how pathogens are under selection to avoid pulling the trigger of ETI and PCD. Many other pathogen effectors can also directly target key operators of PCD such as cysteine proteases [100-103]. In the following part, I will review defense related cysteine proteases, and continue to illustrate the mechanism of plant-pathogen interaction.

Part 2: Plant cysteine proteases: Important hubs in pathogen-plant interaction

Cysteine Proteases and Cystatin: Democles sword and its sheath

Protease, by definition, is an enzyme that degrades proteins. But in plant cells, proteases act as far more than "protein recyclers". Instead, they are key regulators of diverse biological processes such as embryogenesis, flowering, leaf senescence, and plant defense responses [104]. Plant genomes encode numerous proteases. For example, the *Arabidopsis* genome contains over 800 putative protease genes [101, 105-108]. All proteases have a similar biochemical activity, where they all use an oxyanion hole structure to stabilize the oxygen on the substrate peptide, and thus polarize the carbonyl group [104]. Subsequently, their nucleophilic residues attack the exposed carbon atom to initiate the cleavage [104]. However, different proteases may have different catalytic residues. Based on their catalytic residues, plant proteases are classified into five major groups: serine proteases, cysteine proteases, threonine proteases, aspartate proteases, and metalloproteases (using a metal ion, usually zinc) [104, 109, 110]. As shown in Figure 1-1a, surveying publications in the NCBI database suggests that serine proteases and cysteine proteases are the two most intensively studied proteases in the field of plant science. Interestingly, we found 2,286 out of a total of 4,693 publications of cysteine proteases, are related to PCD, a phenomenon closely related to plant immunity as previously described. Therefore, it is not surprising that “cysteine proteases" are one of the hottest topics in the research field of plant immunity (Figure 1-1b).

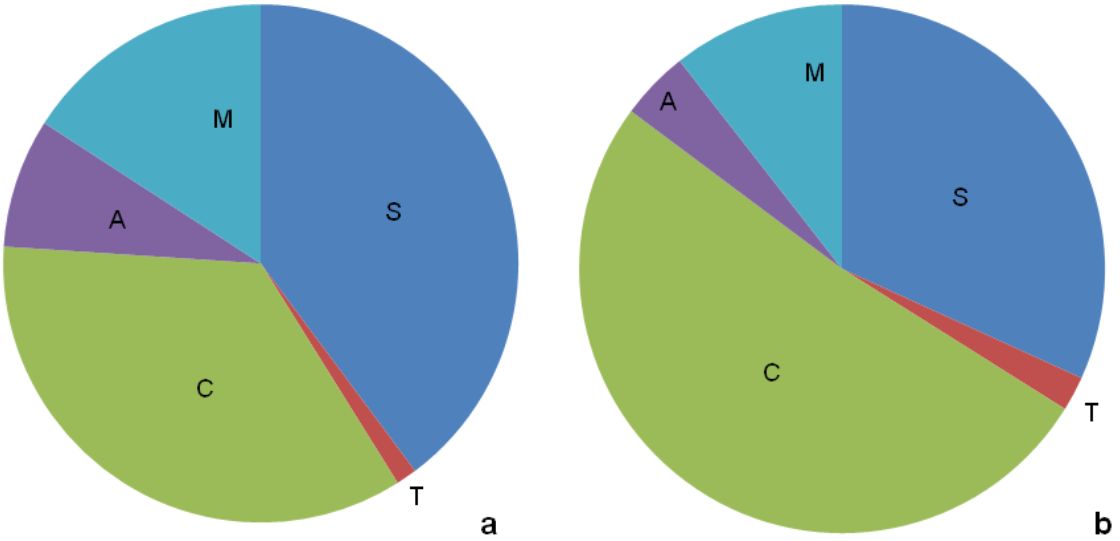


Figure 1-1| Distribution of protease related publications in the NCBI database

a. Publications related with the keyword "plant science". **b.** Publications related with the keyword "plant immunity". Abbreviations: M- metalloproteases, A- aspartate proteases, C- cysteine protease T- threonine proteases, S- serine protease.

Plant genomes encode over 140 cysteine proteases, which can be divided into 15 families of five clans [104, 105]. Many of them are related with PCD and plant immunity, and others are involved in biological process like flowering time, embryo development, epidermal cell aging [101, 108, 111-113]. In this review, I will focus on plant immunity-related cysteine proteases that mainly belong to the C13 family and C1A family (Figure 1-2). Cysteine proteases from C13 and C14 are structurally similar to each other, and both belong to the CD clan [104, 105]. The CD proteases are also known as caspase-like proteases [114], since they share sequence or structural homology to the animal caspases, which have a conserved α/β/α sandwich structure [115]. However, the biological functions of plant and animal CD proteases are different. For example, metacaspases (MCAs) from the C14 family are involved in PCD in somatic embryogenesis [116, 117]. While plant vascular processing enzymes VPEs from the C13 family are involved in defense-related PCD [118]. In *N. benthamiana*, silencing *NbVpe* blocks the hypersensitive cell death triggered by tobacco mosaic virus (TMV) and N-mediated resistance to TMV [118]. In *Arabidopsis*, the *Vpes* gene family includes four family members: *αVpe, βVpe, γVpe, δVpe*. Among them, δVpe regulates PCD during seed coat development, and γVpe was reported to positively regulate plant disease resistance [119].

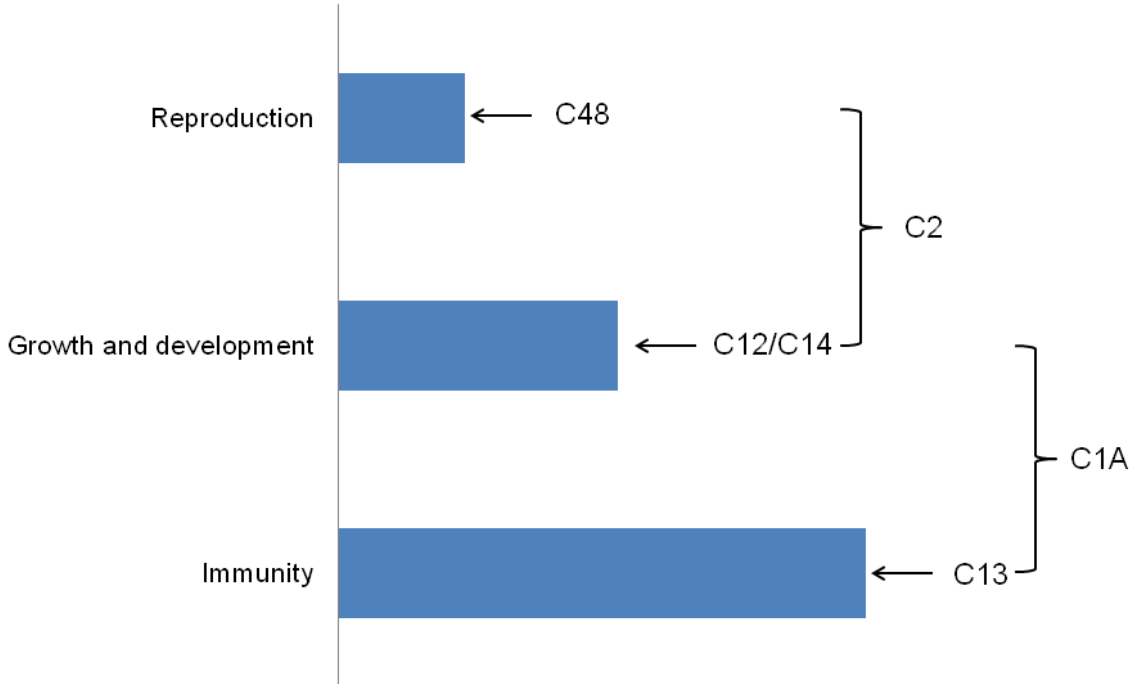


Figure 1-2| Distribution of publications of cysteine proteases associated with plant programmed cell death

The distribution of publications of cysteine proteases associated with plant programmed cell death (C12, C13, C14 and C48).

C1A is a subfamily of C1 cysteine proteases, which are also known as papin-like cysteine proteases (PLCPs) [104, 105]. The other C1 subfamily is C1B, which is not present in plant genomes [104]. Compared to C1B proteases, C1A proteases mainly target proteins with disulfide bridges and they usually accumulate in the vacuole and apoplast [105]. As shown in Figure 1-2, C1As were the most frequently reported proteases in studies of plant immunity. Therefore, in many plant pathology publications, the name of "cysteine protease" usually refers to the proteases belonging to C1A or PLCP. In the remainder of this review, I will focus on the roles of PLCPs in plant-pathogen interactions. PLCPs are involved in plant defense by regulating PCD during ETI [104]. Although the phenotypes of PCD triggered by different R proteins are similar, the cell death signaling pathways could be independent [104]. For example, one of the PLCP genes, *Cathepsin B* (*CathB*) from *N. benthamiana* was found to be induced during the HR process [120]. However, silencing of *NbCathB* could suppress the HR triggered by interaction of Avr3a/R3a [120], but not the interaction of Avr4/Cf-4 [120]. This study indicates NbCathB is specifically required for Avr3a/R3a-mediated ETI [120].

In plant cells, PCD is the strongest defense response that can be considered as a partial "suicide" defense strategy [118]. As the key operator of PCD, proteases, especially PLCPs, can be potentially damaging when overexpressed or miss-activated at the translational level [11]. Therefore, plants developed diverse negative protease regulators to tightly regulate PLCP-related PCD. For the PLCPs, plants evolved at least three layers of regulatory mechanisms to control the PLCP activities. The first one is transcriptional regulation. Arabidopsis has 31 PLCLPs [121]. Analyzing the RNA-seq data of expression profiling of Arabidopsis genes in different plant tissue (NCBI GEO accession ID: GSE38612), I found that the 31 Arabidopsis PLCP genes have differential expression patterns. They are only expressed in certain plant organs (Table 1-1). The second layer of PLCP regulation is at the post-translational level. All PLCPs carried an auto-inhibitory predomain, which prevents the premature activation of the protease [110, 122]. Besides the pre-domain, PLCPs also have multiple signal peptides targeting them to different subcellular localizations [110]. Some PLCPs also carry a granulin-like domain at their C-terminus with unknown function [110, 121, 122]. After being translated into a pre-protein, PLCPs need go through a series of post-translational processing to become mature active proteases [110]. Here, I am going to use Arabidopsis PLCP RD21A as an example to illustrate the maturing process of PLCP [122]. RD21A has been widely reported to be associated with plant defenses and abiotic stress responses [122, 123]. The pre-protein precursor of RD21A has a N-terminal signal peptide, an auto-inhibitory pro-domain, the protease domain, a proline-rich domain and a granulin-like domain [122]. The first step of RD21A maturation consists in entering the secretion pathway due to recognition of the signal peptide, which will be removed after this process [122]. Then RD21A will lose its pro-domain and turn into an intermediate active isoform. The final maturation step of RD21A is removal of its C-terminal granulin domain, and some mature RD21A isoforms can continue with removal of the proline-rich domain [122]. The activation of PLCP often requires certain elicitors [11]. For example, cysteine protease activity can be induced by the SA-associated HR triggered by the interaction of cowpea and cowpea rust [124]. Some PLCPs can be activated by H_2O_2, SDS, or low pH [120, 125].

Table 1-1 | Relative expression levels of 31 *Arabidopsis* PLCP genes in different plant organs

Gene	TAIR ID	Leaf	Root	Flower	Seed
RD21A	AT1G47128	201.51	241.1	310.24	227.66
RD21B	AT5G43060	23.877	189.18	55.857	115.22
RD21C	AT3G19390	0.9993	58.348	27.154	0.6574
RDL1	AT4G36880	0.5273	6.3687	10.399	0.3067
RDL2	AT3G19400	10.639	19.056	11.368	18.401
RDL3	AT3G43960	5.236	11.296	27.104	38.025
RDL4	AT4G11310	0.0039	1.9153	1.402	0.0315
RDL5	AT4G11320	19.293	850.14	62.072	9.3265
RDL6	AT4G23520	0	0.0092	0.342	2.6559
CEP1	AT5G50260	0.0682	0.422	3.4813	24.435
CEP3	AT3G48350	8.7012	6.0746	3.2843	1.7459
CEP2	AT3G48340	0.0644	73.627	15.977	3.5965
XCP1	AT4G35350	10.929	62.702	14.878	13.657
XCP2	AT1G20850	27.86	61.194	37.414	34.458
XBCP3	AT1G09850	9.6307	19.843	17.511	12.992
THI1	AT1G06260	1.0414	0.6018	63.709	0.9433
SAG12	AT5G45890	0.0085	0	100.19	19.222
PAP1	AT2G34080	0.5532	9.6054	0.3399	0.3312
PAP2	AT1G29090	0.1206	0.506	0.4038	0
PAP3	AT1G29080	0	0	0.0285	1.1595
PAP4	AT2G27420	1.7533	0.0132	1.8958	0.7659
PAP5	AT3G49340	0.0146	0	0.0048	0.0833
RD19A	AT4G39090	324.8	639	168.32	443.55
RD19B	AT2G21430	0.8851	1.0212	16.028	60.994
RD19C	AT4G16190	207.03	516.51	287.88	230.8
RD19D	AT3G54940	0.0289	1.0598	0.6456	1.0394
AALP	AT5G60360	249.45	299.37	237.96	221.42
ALP2	AT3G45310	13.315	79.153	26.443	27.002
CTB1	AT1G02300	23.904	11.207	14.042	24.346
CTB2	AT1G02305	52.662	124.06	77.039	122.41

CTB3	AT4G01610	45.247	205.61	91.522	93.897

To guarantee restriction of protease activity, plants evolved specific protease inhibitors as the third layer of regulatory mechanism [126]. Protease inhibitors are not only conserved in plants but are also widely distributed in animals and microorganisms [126-129]. The inhibition mechanism of the protease inhibitor is also conserved, including either direct or indirect blocking of the activation site of the cognate protease [126]. Protease inhibitors were previously classified into different groups based on either sequence homology of the inhibition domain or the catalytic type of protease that is inhibited [126]. However, later studies proved this classification to be less rigorous, because many protease inhibitors were identified to be able to suppress proteases across different families [126]. For example, some protease inhibitors from the serpin family have inhibition activity against both, serine and cysteine proteases [126]. Nevertheless, the historical "less rigorous" classification is still widely used. Names, such as "cysteine-" or "serine-" protease inhibitor, are still more commonly used than the rigorous names, such as "Family 14" protease inhibitor. In this review, I mainly focus on the phytocystatins (plant cystatins), which belong to the cystatin superfamily [126, 130]. Cystatins are competitive protein inhibitors of PLCPs [130]. They have a conserved Gln-Xaa-Val-Xaa-Gly motif in the central region of the polypeptide. Additionally, plant cystatins also have a conserved Gly residue in the N-terminus and a Pro-Trp (or Leu-Trp) motif in the C-terminus [130]. The center motif (Gln-Xaa-Val-Xaa-Gly) of cystatins is the functional part that physically interacts with the activation site of cysteine proteases [130]. The C-terminal Pro-Trp (or Leu-Trp) motif together with the center motif form a hairpin-loop structure that helps the center part to enter the activation site of the cysteine proteases [130]. These highly conserved protein structure and inhibitory motifs suggest cystatins play vital roles in regulation of PLCPs activity in plant cells [130]. The close relationship between PLCPs and cystatins suggests that plant PCD could be regulated by the tight interplay of PLCP-cystatin pairs. Contrary to the function of PLCPs, plant cystatins are usually functioning as negative regulators of PCD or HR [130]. For example, Arabidopsis cystatin CYS1 was found to suppress PCD triggered by pathogens and wounding [131].

However, it needs to be mentioned that protease inhibitors cannot be simply considered as the negative regulators of plant immunity [126]. In addition to suppressing *in planta* proteases, protease inhibitors can also inhibit the *ex planta* proteases such as the digestive enzymes secreted by insects and nematodes [126, 132, 133]. For example, the trypsin inhibitors from soybean were reported to be toxic to the larvae of the flour beetle (*Tribolium confusum*) [132]. And trypsin inhibitors from cowpea show broad inhibition of the growth of the nematodes

Globodera tabacum, *Globodera pallida*, and *Meloidogyne incognita* [133]. Bacteria also secrete protease-like effectors to induce cleavage of plant proteins involved in plant immunity [92, 127]. For example, *P. syringae* effector AvrRpt2 is a cysteine protease that can degrade the Arabidopsis RIN4 protein to negatively regulate plant basal defenses [92]. Another *P. syringae* effectorm AvrPphB, also encodes a cysteine protease that can cleave the plant kinase PBS1 to trigger Rps5-mediated resistance [134]. Although there is no evidence that plant cystatins participate in AvrRpt2-triggered defense responses, plant protease inhibitors could potentially suppress the protease activity of various bacterial protease-like effectors.

Pushing Hands: Fight for the control of cysteine protease

In plant-microbe interactions, microbial pathogens evolve relatively fast, which allows them to evolve new effectors to escape from host recognition or suppress the plant defense system [11]. Successful co-evolved *R* genes often "guard" the virulence target of pathogen effectors [1], which usually do not change in the “arm race” game. That way, during co-evolution of pathogen effectors and plant R proteins, both of them will target some common proteins, which are the converged points of plant immunity. Mukhtar et al (2011) endorsed this co-evolution model through their comprehensive study of effector interactomes [135]. Mukhtar et al employed the yeast two-hybrid method to map the physical interactions between Arabidopsis proteins and effectors of *P. syringae* and *H. arabidopsidis*, and they found although *P. syringae* and *H. arabidopsidis* diverged about two billion years ago, their effectors target an overlapping subset of plant proteins, which are common hubs in plant-microbe interactions [135]. Some PLCPs belong to the immune-related hubs that are often targeted by diverse pathogen effectors [100, 101, 111, 123]. In the following section, I will review several publications to illustrate the mechanism of PLCPs-regulated plant immunity.

Maud Bernoux et al (2008) found that the *R. solanacearum* effector PopP2 interacts with one Arabidopsis PLCP, RD19A (At4g39090) [111]. RD19A is required for *RRS1-R* mediated resistance to *R. solanacearum* [111]. The expression of *RD19A* is up-regulated during disease [111]. Therefore, the authors hypothesized that RD19A could be the virulence target of PopP2 [111]. Transient expression of RD19A in *N. benthamiana* revealed that it can co-localize with another PLCP, AALP (also named as aleurain, At5g60360) in the lytic vacuole [111]. The subcellular localization of RD19A was altered by co-expression of PopP2 in *N. benthamiana* changing to the plant nucleus [111]. PopP2 also interacts

with RRS1-R in the plant nucleus [111]. Interestingly, PopP2 also encodes a cysteine protease. Therefore, RD19A may have a function similar to RIN4, where it can be cleaved by AvpRpt2 [111]. However, there is not enough evidence to support that RD19A is required for virulence function of PopP2. Furthermore, PopP2 cannot clave or modify RD19A [136]. It is possible that the interaction of Pop2/RD19A and Pop2-mediated nuclear localization of RA19A only occurs in *N. benthamiana* plant cells when both, Pop2 and RD19A, are over-expressed. Therefore, with the current evidence, it is still difficult to draw a clear conclusion about the role of RD19A in plant immunity. It is required for the RRS1-R-mediated defense, but what is the specific mechanism? Does RD19A function as a positive regulator of Pop2-mediated ETS or is it an actual target of PopP2? Further research on RD19A is still needed to answer these questions.

Compared to the "ambiguous" relationship between PopP2 and RD19A, the effector Pit2 of the fungal maize pathogen *Ustilago maydis* has been clearly demonstrated to be a cysteine protease inhibitor [100]. Mueller et al (2013) reported that Pit2 interacts with four maize cysteine proteases (CP2, CP1A, CP1B and XCP2) *in vitro*, where all these PCLPs are related to SA-associated defenses [100]. In a protease enzyme assay, the authors demonstrated that Pit2 can inhibit the activity of CP2, CP1A and XCP2 [100]. The conserved protease-inhibitor domain of Pit2 is required for its ability to inhibit plant proteases [100]. Like many other effectors, to avoid the recognition by the plant immune system, Pit2 homologues identified in fungal pathogens such as *U. maydis*, *S. reilanum* and *U. hordei* have diverged [100]. There are only 15 identical residues of all Pit2 homologues, where eight of them are located in the conserved protease-inhibitor domain [100]. This discovery endorses the model that evolution of pathogen effectors will not change their virulence targets and activities. It will be interesting to further investigate if Pit2 has virulence activity and whether its virulence function is linked to protease inhibition. Besides directly secreting a protease inhibitor, the fungal pathogen *U. maydis* can also regulate the expression of maize cystatin CC9, which is required for pathogen virulence [137]. Interestingly, transgenic maize plants that overexpress CC9 suppress the expression of *PR1* and the protease activity of CP2, CP1A, CP1B and XCP2, which is usually regulated by SA, and eventually suppress plant immunity [137].

The studies of *U. maydis* indicate that plant PLCPs play essential roles in maize immunity. Since plant PLCPs are often being targeted by microbial pathogens, host plants also evolved different strategies to "guard" them with R proteins [101]. For example, the fungal effector Avr2 secreted by *Cladosporium fulvum* encodes a cysteine-rich protein [101]. As a substrate mimic of cysteine protease, Avr2 interacts with tomato cysteine protease Rcr3, and inhibits its enzyme activity [101]. the tomato R protein Cf-2 recognizes the interaction of

Avr2 and Rcr3, and triggers ETI [101]. Interestingly, neither inhibiting Rcr3 by artificial protease inhibitor E-64, nor completely knocking out Rcr3 can trigger the Cf-2-mediated resistance response [101]. Therefore, the Avr2-Rcr3 complex, instead of the individual protein, is guarded by Cf-2 [101]. Like other defense related cysteine proteases, the expression of *Rcr3* can be induced by pathogen infection [101]. However, the detailed mechanism of the Rcr3-mediated plant defense response still needs further investigation.

Conclusion: You may know by a handful the whole sack

From reviewing the studies of plant-microbe interaction and cysteine protease-associated plant immunity, we can epitomize the multifarious endless warfare of pathogens and plants in regard to cysteine proteases. Cysteine proteases, especially the PLCPs, are essential components involved in plant PTI or ETI (Figure1-1 and Figure1-2). Thus, they are often being attacked (or modified) by various pathogen effectors [100, 101, 137]. In response to the pathogen attack, host plants evolved R proteins or other strategies to protect their PLCPs as explained in the guard model [1]. On the other hand, many PLCPs are important regulators in PCD signaling pathways [104]. PCD is a powerful defense response to restrict pathogen growth, but it is also a "suicide" bomb, which needs to be tightly controlled to be active only under certain conditions in defined tissues [11, 118]. To this end, plants evolved several layers of strategies to regulate cysteine proteases activity, such as the specialized cysteine protease inhibitors cystatins [126, 130]. Successful pathogens also evolved to utilize the power of cystatins to "modify" plant cysteine protease enzyme activity. Therefore, we could consider that "protease enzyme activity", rather than a given protein, is the actual target of pathogen effectors. The arms race between pathogen effectors and host R proteins through the co-evolutionary process is endless, while cysteine proteases as essential hubs of the plant immune response never change.

In summary, we still need to further investigate the detailed mechanisms of how cysteine proteases contribute to plant immunity. In future studies, we need to consider the cysteine protease and cystatin pairs as a balanced system that may coordinately regulate plant immunity. Studying cysteine protease-associated immunity may eventually help us understand the general mechanism of plant-microbe interaction, and allow us to design novel durable crop disease resistance.

References

1. Jones, J.D. and J.L. Dangl, *The plant immune system.* Nature, 2006. **444**(7117): p. 323-9.

2. Lindow, S.E. and M.T. Brandl, *Microbiology of the phyllosphere.* Applied and environmental microbiology, 2003. **69**(4): p. 1875-1883.
3. Andrews, J.H. and R.F. Harris, *The ecology and biogeography of microorganisms on plant surfaces.* Annual review of phytopathology, 2000. **38**(1): p. 145-180.
4. Beattie, G.A. and S.E. Lindow, *The secret life of foliar bacterial pathogens on leaves.* Annual review of phytopathology, 1995. **33**(1): p. 145-172.
5. Wilson, M., S. Hirano, and S. Lindow, *Location and survival of leaf-associated bacteria in relation to pathogenicity and potential for growth within the leaf.* Applied and Environmental Microbiology, 1999. **65**(4): p. 1435-1443.
6. Bunster, L., N.J. Fokkema, and B. Schippers, *Effect of surface-active Pseudomonas spp. on leaf wettability.* Applied and environmental microbiology, 1989. **55**(6): p. 1340-1345.
7. Quigley, N.B. and D. Gross, *Syringomycin production among strains of Pseudomonas syringae pv. syringae: conservation of the syrB and syrD genes and activation of phytotoxin production by plant signal molecules.* MOLECULAR PLANT MICROBE INTERACTIONS, 1994. **7**: p. 78-78.
8. Mo, Y.-Y. and D.C. Gross, *Expression in vitro and during plant pathogenesis of the syrB gene required for syringomycin production by Pseudomonas syringae pv. syringae.* Mol. Plant-Microbe Interact, 1991. **4**: p. 28-36.
9. Hutchison, M.L., M.A. Tester, and D. Gross, *Role of biosurfactant and ion channel-forming activities of syringomycin in transmembrane ion flux: a model for the mechanism of action in the plant-pathogen interaction.* Mol. Plant-Microbe Interact, 1995. **8**: p. 610-620.
10. Hutchison, M.L. and K. Johnstone, *Evidence for the involvement of the surface active properties of the extracellular toxin tolaasin in the manifestation of brown blotch disease symptoms by Pseudomonas tolaasii on Agaricus bisporus.* Physiological and Molecular Plant Pathology, 1993. **42**(5): p. 373-384.
11. Doehlemann, G. and C. Hemetsberger, *Apoplastic immunity and its suppression by filamentous plant pathogens.* New Phytol, 2013. **198**(4): p. 1001-16.
12. Howard, R.J. and B. Valent, *Breaking and entering: host penetration by the fungal rice blast pathogen Magnaporthe grisea.* Annual Reviews in Microbiology, 1996. **50**(1): p. 491-512.
13. Meng, S., et al., *Common processes in pathogenesis by fungal and oomycete plant pathogens, described with Gene Ontology terms.* BMC Microbiol, 2009. **9 Suppl 1**: p. S7.

14. Bourett, T.M. and R.J. Howard, *In vitro development of penetration structures in the rice blast fungus Magnaporthe grisea.* Canadian Journal of Botany, 1990. **68**(2): p. 329-342.
15. Melotto, M., W. Underwood, and S.Y. He, *Role of stomata in plant innate immunity and foliar bacterial diseases.* Annual review of phytopathology, 2008. **46**: p. 101.
16. Zeng, W., M. Melotto, and S.Y. He, *Plant stomata: a checkpoint of host immunity and pathogen virulence.* Current opinion in biotechnology, 2010. **21**(5): p. 599-603.
17. Schulze-Lefert, P. and S. Robatzek, *Plant pathogens trick guard cells into opening the gates.* Cell, 2006. **126**(5): p. 831-4.
18. Fan, L.M., Z. Zhao, and S.M. Assmann, *Guard cells: a dynamic signaling model.* Curr Opin Plant Biol, 2004. **7**(5): p. 537-46.
19. Schroeder, J.I., et al., *Guard cell signal transduction.* Annual review of plant biology, 2001. **52**(1): p. 627-658.
20. Melotto, M., et al., *Plant stomata function in innate immunity against bacterial invasion.* Cell, 2006. **126**(5): p. 969-980.
21. Mittal, S. and K.R. Davis, *Role of the phytotoxin coronatine in the infection of Arabidopsis thaliana by Pseudomonas syringae pv. tomato.* MPMI-Molecular Plant Microbe Interactions, 1995. **8**(1): p. 165-171.
22. Ma, S.W., V.L. Morris, and D.A. Cuppels, *Characterization of a DNA region required for production of the phytotoxin coronatine by Pseudomonas syringae pv. tomato.* Mol. Plant Microbe Interact., 1991. **4**: p. 69-77.
23. Katsir, L., et al., *COI1 is a critical component of a receptor for jasmonate and the bacterial virulence factor coronatine.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(19): p. 7100-7105.
24. Zipfel, C. and G. Felix, *Plants and animals: a different taste for microbes?* Current opinion in plant biology, 2005. **8**(4): p. 353-360.
25. Gómez-Gómez, L. and T. Boller, *FLS2: An LRR receptor–like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis.* Molecular cell, 2000. **5**(6): p. 1003-1011.
26. Asai, T., et al., *MAP kinase signalling cascade in Arabidopsis innate immunity.* Nature, 2002. **415**(6875): p. 977-983.
27. Denny, T., *Involvement of bacterial polysaccharides in plant pathogenesis.* Annual review of phytopathology, 1995. **33**(1): p. 173-197.
28. Lu, D., et al., *A receptor-like cytoplasmic kinase, BIK1, associates with a flagellin receptor complex to initiate plant innate immunity.* Proceedings of the National Academy of Sciences, 2010. **107**(1): p. 496-501.

29. Zipfel, C., *Early molecular events in PAMP-triggered immunity.* Current opinion in plant biology, 2009. **12**(4): p. 414-420.
30. Boller, T. and G. Felix, *A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors.* Annual review of plant biology, 2009. **60**: p. 379-406.
31. van Wijk, F. and B. Prakken, *Editorial: Heat shock proteins: Darwinistic immune modulation on dangerous grounds.* Journal of leukocyte biology, 2010. **88**(3): p. 431-434.
32. Maimbo, M., et al., *Induction of a small heat shock protein and its functional roles in Nicotiana plants in the defense response against Ralstonia solanacearum.* Plant physiology, 2007. **145**(4): p. 1588-1599.
33. Dong, X., *SA, JA, ethylene, and disease resistance in plants.* Current opinion in plant biology, 1998. **1**(4): p. 316-323.
34. Horbach, R., et al., *When and how to kill a plant cell: infection strategies of plant pathogenic fungi.* Journal of plant physiology, 2011. **168**(1): p. 51-62.
35. Glazebrook, J., *Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens.* Annu. Rev. Phytopathol., 2005. **43**: p. 205-227.
36. Morel, J.-B. and J.L. Dangl, *The hypersensitive response and the induction of cell death in plants.* Cell death and differentiation, 1997. **4**(8): p. 671-683.
37. Thaler, J.S., et al., *Jasmonate- mediated induced plant resistance affects a community of herbivores.* Ecological Entomology, 2001. **26**(3): p. 312-324.
38. Gundlach, H., et al., *Jasmonic acid is a signal transducer in elicitor-induced plant cell cultures.* Proceedings of the National Academy of Sciences, 1992. **89**(6): p. 2389-2393.
39. Memelink, J., R. Verpoorte, and J.W. Kijne, *ORCAnization of jasmonate-responsive gene expression in alkaloid metabolism.* Trends in plant science, 2001. **6**(5): p. 212-219.
40. Yang, D.L., Y. Yang, and Z. He, *Roles of plant hormones and their interplay in rice immunity.* Mol Plant, 2013. **6**(3): p. 675-85.
41. van Loon, L.C., B.P. Geraats, and H.J. Linthorst, *Ethylene as a modulator of disease resistance in plants.* Trends Plant Sci, 2006. **11**(4): p. 184-91.
42. Cao, F.Y., K. Yoshioka, and D. Desveaux, *The roles of ABA in plant-pathogen interactions.* J Plant Res, 2011. **124**(4): p. 489-99.
43. Cui, F., et al., *The Pseudomonas syringae type III effector AvrRpt2 promotes pathogen virulence via stimulating Arabidopsis auxin/indole acetic acid protein turnover.* Plant Physiol, 2013. **162**(2): p. 1018-29.
44. Chen, Z., et al., *Pseudomonas syringae type III effector AvrRpt2 alters Arabidopsis thaliana auxin physiology.* Proc Natl Acad Sci U S A, 2007. **104**(50): p. 20131-6.

45. Durrant, W. and X. Dong, *Systemic acquired resistance.* Annu. Rev. Phytopathol., 2004. **42**: p. 185-209.
46. Gaffney, T., et al., *Requirement of salicylic acid for the induction of systemic acquired resistance.* Science, 1993. **261**(5122): p. 754-756.
47. Malamy, J., et al., *Salicylic acid: a likely endogenous signal in the resistance response of tobacco to viral infection.* Science, 1990. **250**(4983): p. 1002-1004.
48. Métraux, J., et al., *Increase in salicylic acid at the onset of systemic acquired resistance in cucumber.* Science, 1990. **250**(4983): p. 1004-1006.
49. CAMERON, R.K., et al., *Accumulation of salicylic acid and PR-1 gene transcripts in relation to the systemic acquired resistance (SAR) response induced by Pseudomonas syringae pv. tomato in Arabidopsis.* Physiological and Molecular Plant Pathology, 1999. **55**(2): p. 121-130.
50. Chern, M., et al., *Overexpression of a Rice NPR1 Homolog Leads to Constitutive Activation of Defense Response and Hypersensitivity to Light.* Molecular Plant-Microbe Interactions, 2005. **18**(6): p. 511-520.
51. Vlot, A.C., D.M.A. Dempsey, and D.F. Klessig, *Salicylic acid, a multifaceted hormone to combat disease.* Annual review of phytopathology, 2009. **47**: p. 177-206.
52. Levine, A., et al., *H 2 O 2 from the oxidative burst orchestrates the plant hypersensitive disease resistance response.* Cell, 1994. **79**(4): p. 583-593.
53. Cao, H., et al., *The Arabidopsis NPR1 gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats.* Cell, 1997. **88**(1): p. 57-63.
54. Spoel, S.H., et al., *NPR1 modulates cross-talk between salicylate-and jasmonate-dependent defense pathways through a novel function in the cytosol.* The Plant Cell Online, 2003. **15**(3): p. 760-770.
55. Lin, W.-C., et al., *Transgenic tomato plants expressing the Arabidopsis NPR1 gene display enhanced resistance to a spectrum of fungal and bacterial diseases.* Transgenic Research, 2004. **13**(6): p. 567-581.
56. Zhang, X., et al., *Over-expression of the Arabidopsis NPR1 gene in citrus increases resistance to citrus canker.* European Journal of Plant Pathology, 2010. **128**(1): p. 91-100.
57. Catanzariti, A.-M., P.N. Dodds, and J.G. Ellis, *Avirulence proteins from haustoria-forming pathogens.* FEMS microbiology letters, 2007. **269**(2): p. 181-188.
58. Hueck, C.J., *Type III protein secretion systems in bacterial pathogens of animals and plants.* Microbiology and molecular biology reviews, 1998. **62**(2): p. 379-433.

59. Alfano, J.R. and A. Collmer, *Type III secretion system effector proteins: double agents in bacterial disease and plant defense.* Annu. Rev. Phytopathol., 2004. **42**: p. 385-414.
60. Galán, J.E. and A. Collmer, *Type III secretion machines: bacterial devices for protein delivery into host cells.* Science, 1999. **284**(5418): p. 1322-1328.
61. Klee, H.J., et al., *Agrobacterium tumefaciens T-DNA.* Genes & Development, 1987. **1**: p. 86-96.
62. Weiberg, A., et al., *Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways.* Science, 2013. **342**(6154): p. 118-123.
63. Ninio, S. and C.R. Roy, *Effector proteins translocated by Legionella pneumophila: strength in numbers.* Trends in microbiology, 2007. **15**(8): p. 372-380.
64. Petnicki-Ocwieja, T., et al., *Genomewide identification of proteins secreted by the Hrp type III protein secretion system of Pseudomonas syringae pv. tomato DC3000.* Proceedings of the National Academy of Sciences, 2002. **99**(11): p. 7652-7657.
65. Deslandes, L. and S. Rivas, *Catch me if you can: bacterial effectors and plant targets.* Trends in plant science, 2012. **17**(11): p. 644-655.
66. Morgan, W. and S. Kamoun, *RXLR effectors of plant pathogenic oomycetes.* Current opinion in microbiology, 2007. **10**(4): p. 332-338.
67. Ochiai, H., et al., *Genome sequence of Xanthomonas oryzae pv. oryzae suggests contribution of large numbers of effector genes and insertion sequences to its race diversity.* Japan Agricultural Research Quarterly: JARQ, 2005. **39**(4): p. 275-287.
68. Göhre, V., et al., *Plant pattern-recognition receptor FLS2 is directed for degradation by the bacterial ubiquitin ligase AvrPtoB.* Current Biology, 2008. **18**(23): p. 1824-1832.
69. Cheong, M.S., et al., *AvrBsT acetylates Arabidopsis ACIP1, a protein that associates with microtubules and is required for immunity.* PLoS Pathog, 2014. **10**(2): p. e1003952.
70. Kemen, E., et al., *Identification of a protein from rust fungi transferred from haustoria into infected plant cells.* Molecular Plant-Microbe Interactions, 2005. **18**(11): p. 1130-1139.
71. Kloppholz, S., H. Kuhn, and N. Requena, *A secreted fungal effector of Glomus intraradices promotes symbiotic biotrophy.* Current Biology, 2011. **21**(14): p. 1204-1209.
72. de Jonge, R., et al., *Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants.* Science, 2010. **329**(5994): p. 953-955.

73. Bolton, M.D., et al., *The novel Cladosporium fulvum lysin motif effector Ecp6 is a virulence factor with orthologues in other fungal species.* Molecular microbiology, 2008. **69**(1): p. 119-136.
74. Stergiopoulos, I. and P.J. de Wit, *Fungal effector proteins.* Annual review of phytopathology, 2009. **47**: p. 233-263.
75. Mengiste, T., *Plant immunity to necrotrophs.* Annual review of phytopathology, 2012. **50**: p. 267-294.
76. Hammond-Kosack, K.E. and J. Jones, *Resistance gene-dependent plant defense responses.* The Plant Cell, 1996. **8**(10): p. 1773.
77. Flor, H.H., *Current status of the gene-for-gene concept.* Annual review of phytopathology, 1971. **9**(1): p. 275-296.
78. Loegering, W.Q. and A. Ellingboe, *HH Flor: Poineer in Phytopathology.* Annual review of phytopathology, 1987. **25**(1): p. 59-66.
79. Gassmann, W., et al., *Molecular evolution of virulence in natural field strains of Xanthomonas campestris pv. vesicatoria.* J Bacteriol, 2000. **182**(24): p. 7053-9.
80. Eitas, T.K. and J.L. Dangl, *NB-LRR proteins: pairs, pieces, perception, partners, and pathways.* Current opinion in plant biology, 2010. **13**(4): p. 472-477.
81. Jia, Y., et al., *Rice Pi-ta gene confers resistance to the major pathotypes of the rice blast fungus in the United States.* Phytopathology, 2004. **94**(3): p. 296-301.
82. Dodds, P.N., et al., *Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes.* Proceedings of the National Academy of Sciences, 2006. **103**(23): p. 8888-8893.
83. Jia, Y., et al., *Direct interaction of resistance gene and avirulence gene products confers rice blast resistance.* EMBO J, 2000. **19**(15): p. 4004-14.
84. Dodds, P.N., et al., *Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes.* Proc Natl Acad Sci U S A, 2006. **103**(23): p. 8888-93.
85. Wang, C.I., et al., *Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity.* Plant Cell, 2007. **19**(9): p. 2898-912.
86. Krasileva, K.V., D. Dahlbeck, and B.J. Staskawicz, *Activation of an Arabidopsis resistance protein is specified by the in planta association of its leucine-rich repeat domain with the cognate oomycete effector.* Plant Cell, 2010. **22**(7): p. 2444-58.

87. Kanzaki, H., et al., *Arms race co-evolution of Magnaporthe oryzae AVR-Pik and rice Pik genes driven by their physical interactions.* The Plant Journal, 2012. **72**: p. 894-907.
88. Chen, Y., Z. Liu, and D.A. Halterman, *Molecular determinants of resistance activation and suppression by Phytophthora infestans effector IPI-O.* PLoS Pathog, 2012. **8**(3): p. e1002595.
89. Mackey, D., et al., *RIN4 interacts with Pseudomonas syringae type III effector molecules and is required for RPM1-mediated resistance in Arabidopsis.* Cell, 2002. **108**(6): p. 743-754.
90. Mackey, D., et al., *Arabidopsis RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance.* Cell, 2003. **112**(3): p. 379-389.
91. Axtell, M.J. and B.J. Staskawicz, *Initiation of RPS2-specified disease resistance in Arabidopsis is coupled to the AvrRpt2-directed elimination of RIN4.* Cell, 2003. **112**(3): p. 369-377.
92. Kim, H.-S., et al., *The Pseudomonas syringae effector AvrRpt2 cleaves its C-terminally acylated target, RIN4, from Arabidopsis membranes to block RPM1 activation.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(18): p. 6496-6501.
93. Meyers, B.C., et al., *Genome-wide analysis of NBS-LRR–encoding genes in Arabidopsis.* The Plant Cell Online, 2003. **15**(4): p. 809-834.
94. Buell, C.R., et al., *The complete genome sequence of the Arabidopsis and tomato pathogen Pseudomonas syringae pv. tomato DC3000.* Proceedings of the National Academy of Sciences, 2003. **100**(18): p. 10181-10186.
95. Mucyn, T.S., et al., *The tomato NBARC-LRR protein Prf interacts with Pto kinase in vivo to regulate specific plant immunity.* The Plant Cell Online, 2006. **18**(10): p. 2792-2806.
96. Abramovitch, R.B., et al., *Pseudomonas type III effector AvrPtoB induces plant disease susceptibility by inhibition of host programmed cell death.* The EMBO Journal, 2003. **22**(1): p. 60-69.
97. Xiao, F., et al., *The N- terminal region of Pseudomonas type III effector AvrPtoB elicits Pto- dependent immunity and has two distinct virulence determinants.* The Plant Journal, 2007. **52**(4): p. 595-614.
98. Rosebrock, T.R., et al., *A bacterial E3 ubiquitin ligase targets a host protein kinase to disrupt plant immunity.* Nature, 2007. **448**(7151): p. 370-374.
99. Abramovitch, R.B. and G.B. Martin, *Strategies used by bacterial pathogens to suppress plant defenses.* Current opinion in plant biology, 2004. **7**(4): p. 356-364.

100. Mueller, A.N., et al., *Compatibility in the Ustilago maydis-maize interaction requires inhibition of host cysteine proteases by the fungal effector Pit2.* PLoS Pathog, 2013. **9**(2): p. e1003177.
101. Rooney, H.C., et al., *Cladosporium Avr2 inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance.* Science, 2005. **308**(5729): p. 1783-1786.
102. Krüger, J., et al., *A tomato cysteine protease required for Cf-2-dependent disease resistance and suppression of autonecrosis.* Science, 2002. **296**(5568): p. 744-747.
103. Gilroy, E.M., et al., *Involvement of cathepsin B in the plant disease resistance hypersensitive response.* The Plant Journal, 2007. **52**(1): p. 1-13.
104. Van der Hoorn, R.A., *Plant proteases: from phenotypes to molecular mechanisms.* Annu. Rev. Plant Biol., 2008. **59**: p. 191-223.
105. Rawlings, N.D., D.P. Tolle, and A.J. Barrett, *MEROPS: the peptidase database.* Nucleic Acids Research, 2004. **32**(suppl 1): p. D160-D164.
106. Bozhkov, P.V., et al., *Cysteine protease mcII-Pa executes programmed cell death during plant embryogenesis.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(40): p. 14463-14468.
107. Reeves, P.H., et al., *early in short days 4, a mutation in Arabidopsis that causes early flowering and reduces the mRNA abundance of the floral repressor FLC.* Development, 2002. **129**(23): p. 5349-5361.
108. Diaz-Mendoza, M., et al., *C1A cysteine protease–cystatin interactions in leaf senescence.* Journal of experimental botany, 2014: p. eru043.
109. Beers, E.P., B.J. Woffenden, and C. Zhao, *Plant proteolytic enzymes: possible roles during programmed cell death*, in *Programmed Cell Death in Higher Plants*. 2000, Springer. p. 155-171.
110. van der Hoorn, R.A., et al., *Activity profiling of papain-like cysteine proteases in plants.* Plant Physiol, 2004. **135**(3): p. 1170-8.
111. Bernoux, M., et al., *RD19, an Arabidopsis cysteine protease required for RRS1-R–mediated resistance, is relocalized to the nucleus by the Ralstonia solanacearum PopP2 effector.* The Plant Cell Online, 2008. **20**(8): p. 2252-2264.
112. Lid, S.E., et al., *The defective kernel 1 (dek1) gene required for aleurone cell development in the endosperm of maize grains encodes a membrane protein of the calpain gene superfamily.* Proceedings of the National Academy of Sciences, 2002. **99**(8): p. 5460-5465.
113. Feller, U., I. Anders, and K. Demirevska, *Degradation of rubisco and other chloroplast proteins under abiotic stress.* Gen Appl Plant Physiol, 2008. **34**(1-2): p. 5-18.

114. van der Hoorn, R.A. and J.D. Jones, *The plant proteolytic machinery and its role in defence.* Current opinion in plant biology, 2004. **7**(4): p. 400-407.
115. Chen, J.-M., et al., *Identification of the active site of legumain links it to caspases, clostripain and gingipains in a new clan of cysteine endopeptidases.* FEBS letters, 1998. **441**(3): p. 361-365.
116. Bozhkov, P., et al., *VEIDase is a principal caspase-like activity involved in plant programmed cell death and essential for embryonic pattern formation.* Cell Death & Differentiation, 2004. **11**(2): p. 175-182.
117. Suarez, M.F., et al., *Metacaspase-dependent programmed cell death is essential for plant embryogenesis.* Current Biology, 2004. **14**(9): p. R339-R340.
118. Hatsugai, N., et al., *A cellular suicide strategy of plants: vacuole-mediated cell death.* Apoptosis, 2006. **11**(6): p. 905-911.
119. Nakaune, S., et al., *A vacuolar processing enzyme, δVPE, is involved in seed coat formation at the early stage of seed development.* The Plant Cell Online, 2005. **17**(3): p. 876-887.
120. Solomon, M., et al., *The involvement of cysteine proteases and protease inhibitor genes in the regulation of programmed cell death in plants.* The Plant Cell Online, 1999. **11**(3): p. 431-443.
121. Richau, K.H., et al., *Subclassification and biochemical analysis of plant papain-like cysteine proteases displays subfamily-specific characteristics.* Plant Physiol, 2012. **158**(4): p. 1583-99.
122. Yamada, K., et al., *A Slow Maturation of a Cysteine Protease with a Granulin Domain in the Vacuoles of Senescing Arabidopsis Leaves.* Plant Physiology, 2001. **127**(4): p. 1626-1634.
123. Shindo, T., et al., *A role in immunity for Arabidopsis cysteine protease RD21, the ortholog of the tomato immune protease C14.* PloS one, 2012. **7**(1): p. e29317.
124. D'Silva, I., G.G. Poirier, and M.C. Heath, *Activation of cysteine proteases in cowpea plants during the hypersensitive response—a form of programmed cell death.* Experimental cell research, 1998. **245**(2): p. 389-399.
125. Yamada, T., et al., *Purification of a novel type of SDS-dependent protease in maize using a monoclonal antibody.* Plant and cell physiology, 1998. **39**(1): p. 106-114.
126. Habib, H. and K.M. Fazili, *Plant protease inhibitors: a defense strategy in plants.* Biotechnology and Molecular Biology Review, 2007. **2**(3): p. 68-85.
127. Supuran, C.T., A. Scozzafava, and B.W. Clare, *Bacterial protease inhibitors.* Med Res Rev, 2002. **22**(4): p. 329-72.
128. Haq, S.K., S.M. Atif, and R.H. Khan, *Protein proteinase inhibitor genes in combat against insects, pests, and pathogens: natural and engineered*

phytoprotection. Archives of Biochemistry and Biophysics, 2004. **431**(1): p. 145-159.
129. Valueva, T. and V. Mosolov, *Role of inhibitors of proteolytic enzymes in plant defense against phytopathogenic microorganisms.* Biochemistry (Moscow), 2004. **69**(11): p. 1305-1309.
130. Benchabane, M., et al., *Plant cystatins.* Biochimie, 2010. **92**(11): p. 1657-66.
131. Belenghi, B., et al., *AtCYS1, a cystatin from Arabidopsis thaliana, suppresses hypersensitive cell death.* European Journal of Biochemistry, 2003. **270**(12): p. 2593-2604.
132. Gatehouse, A.M. and D. Boulter, *Assessment of the antimetabolic effects of trypsin inhibitors from cowpea (Vigna unguiculata) and other legumes on development of the bruchid beetle Callosobruchus maculatus.* Journal of the Science of Food and Agriculture, 1983. **34**(4): p. 345-350.
133. Urwin, P.E., et al., *Transgenic resistance to the nematode Rotylenchulus reniformis conferred by Arabidopsis thaliana plants expressing proteinase inhibitors.* Molecular Breeding, 2000. **6**(3): p. 257-264.
134. Shao, F., et al., *Cleavage of Arabidopsis PBS1 by a bacterial type III effector.* Science, 2003. **301**(5637): p. 1230-3.
135. Mukhtar, M.S., et al., *Independently evolved virulence effectors converge onto hubs in a plant immune system network.* science, 2011. **333**(6042): p. 596-601.
136. Tasset, C., et al., *Autoacetylation of the Ralstonia solanacearum effector PopP2 targets a lysine residue essential for RRS1-R-mediated immunity in Arabidopsis.* PLoS pathogens, 2010. **6**(11): p. e1001202.
137. van der Linde, K., et al., *A maize cystatin suppresses host immunity by inhibiting apoplastic cysteine proteases.* The Plant Cell Online, 2012. **24**(3): p. 1285-1300.

Chapter II:

Crystal structure of the complex between *Xanthomonas* AvrRxo1-ORF1, a type III effector with a polynucleotide kinase domain, and its interactor AvrRxo1-ORF2

Qian Han[a,b,1], Changhe Zhou [c,1], Shuchi Wu[c,1], Yi Liu[c], Lindsay Triplett[d,e], Jiamin Miao[c], James Tokuhisa[c], Loïc Deblais[d], Howard Robinson[f], Jan E. Leach[d], Jianyong Li[b,2] and Bingyu Zhao[c,2]

[a]Laboratory of Tropical Veterinary Medicine and Vector Biology, and Hainan Key Laboratory of Sustainable Utilization of Tropical Bioresources, College of Agriculture, Hainan University, Haikou 570228, Hainan, China

[b]Department of Biochemistry, Virginia Tech, Blacksburg, Virginia 24061, USA

[c]Department of Horticulture, Virginia Tech, Blacksburg, Virginia 24061, USA

[d]Department of Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, CO 80523-1177

[e]Present address: Department of Plant Pathology and Ecology, The Connecticut Agricultural Experiment Station, New Haven, CT 06511

[f]Biology Department, Brookhaven National Laboratory, Upton, NY 11973

[1]These authors contributed equally to this work.

[2]To whom correspondence should be addressed: E-mail: bzhao07@vt.edu or lij@vt.edu

Summary

Xanthomonas oryzae pv. *oryzicola* (*Xoc*) causes bacterial leaf streak (BLS) disease on rice plants. *Xoc* delivers the type III effector AvrRxo1 into rice plant cells that can be recognized by the disease resistance (R) protein Rxo1 and confers resistance to BLS. However, the mechanism and virulence role of AvrRxo1 are not known. In the genome of *Xoc*, *AvrRxo1-ORF1* is adjacent to another gene, *AvrRxo1-ORF2*, which was predicted to encode a molecular chaperone of AvrRxo1-ORF1. We report the co-purification and crystallization of the AvrRxo1-ORF1:AvrRxo1-ORF2 tetramer complex at 1.64 Å resolutions. AvrRxo1-ORF1 has a T4 polynucleotide kinase domain, and expression of AvrRxo1-ORF1 suppresses bacterial growth in a manner dependent on the kinase motif. Although AvrRxo1-ORF2 binds AvrRxo1-ORF1, it is structurally different from typical effector-binding chaperones, in that it has a distinct fold containing a novel kinase-binding domain. We show that AvrRxo1-ORF2 functions to suppress the bacteriostatic activity of AvrRxo1-ORF1 in bacterial cells.

Introduction

A common feature of gram-negative bacterial pathogens of plants or animals is the use of evolutionarily conserved type III secretion systems (T3SS) to deliver effectors (T3Es) into host cells in order to suppress host immunity [1-4]. Efficient secretion of some T3Es requires T3S chaperones, which bind to the N-terminus of the cognate effector [5, 6]. While in many plant pathogens chaperones are encoded downstream of the effector in the same operon, *Xanthomonas* sp. typically encode a single chaperone, HpaB, that binds to multiple effectors. Most T3E-binding chaperones share a high degree of structural homology despite having divergent amino acid sequences [6].

In some plant pathosystems, plants evolved disease resistance genes (*R*) to specifically recognize cognate T3Es and trigger a disease resistance response that frequently culminates in a hypersensitive response, which consists in programmed cell death and other concomitant plant responses [1]. This interaction is also referred to as effector-triggered immunity [7]. Most cloned *R* genes encode structurally conserved nucleotide binding domain and leucine-rich repeat (NB-LRR) proteins [8-11]. However, in contrast to the conserved structural similarity of plant R proteins, T3Es share little structural or sequence homology [12-15], and it is difficult to predict their biochemical functions based on the primary protein sequence. Some of the bacterial T3Es may suppress host immunity through a variety of strategies, including functional mimicry of proteins in the host cell. However, the majority of effectors do not have predicted biochemical functions. Solved three dimensional (3D) structures of pathogen effectors have been key in identification of the putative functions of several effectors [5, 16-21]. For example, the recolved structures of AvrPtoB and XopL led to the identification of unique E3 ligase domains that were not identifiable from their amino acid sequences [16, 22].

Rice bacterial leaf streak (BLS) disease, caused by *Xanthomonas oryzae* pv. *oryzicola* (*Xoc*), is one of the most important bacterial rice diseases [23-27]. Almost all rice cultivars, especially hybrid rice cultivars, are highly susceptible to BLS [23, 26]. We previously isolated a T3E gene, *avrRxo1-ORF1,* from the genome of *Xoc* [26]. AvrRxo1-ORF1 is conserved among all Asian *Xoc* strains tested [26]. Although no major BLS disease resistance genes has been found in rice germplasm, AvrRxo1-ORF1 is recognized by the maize NB-LRR type disease resistance protein Rxo1, which triggers disease resistance in both, maize and transgenic rice plants [27]. In all *Xoc* strains tested, *avrRxo1-ORF1* is encoded directly upstream of *avrRxo1-ORF2*, predicted to encode a small protein that was hypothesized to act as a chaperone of AvrRxo1-ORF1 [26]. However, ORF2 is not

required for HR induction by AvrRxo1-ORF1, demonstrating AvrRxo1-ORF2 is not required for secretion or translocation of AvrRxo1-ORF1 [26]. The virulence role and molecular host targets of AvrRxo1-ORF1 are unknown, but AvrRxo1 has been reported to be toxic to both yeast and plant cells [26, 28, 29]. Both AvrRxo1-ORF1 and AvrRxo1-ORF2 have no significant sequence homology to known proteins in protein databases. The molecular interactions between AvrRxo1-ORF1 and -ORF2 are also unknown. Therefore, a structural understanding of AvrRxo1-ORF1 and AvrRxo1-ORF2 would provide insight to the molecular mechanism of AvrRxo1's growth-inhibitory capacity and of AvrRxo1/Rxo1-mediated disease resistance. Therefore, we determined the crystal structures of a heterotetramer complex formed by AvrRxo1-ORF1 and AvrRxo1-ORF2. In addition to its toxicity to plant cells, we found that AvrRxo1-ORF1 contributes to *Xanthomonas oryzae* proliferation on rice plants, and strongly suppresses bacterial growth in the absence of AvrRxo1-ORF2. AvrRxo1-ORF1 is a T4-Polynucleotide kinase domain-containing protein with conserved catalytic sites required for virulence and toxicity. AvrRxo1-ORF2, which has no structural similarity to characterized T3S effector-binding chaperones [6], contains a novel kinase-domain binding fold.

Results and discussions

Determination of the crystal structure of AvrRxo1-ORF1:AvrRxo1-ORF2 complexes

Purification of AvrRxo1-ORF1 after expression in *Escherichia coli* cells was problematic because the protein was not soluble. However, co-expression of AvrRxo1-ORF1 with AvrRxo1-ORF2 resulted in a stable and soluble protein complex, allowing purification of the AvrRxo1-ORF1:-ORF2 complex of both native and selenomethionine-substituted proteins. The first crystal of selenomethionine-substituted AvrRxo1-ORF1:-ORF2 complex (Se-Met-AvrRxo1-ORF1:AvrRxo1-ORF2) was diffracted to 1.88 Å. The structure was solved using the single-wavelength anomalous dispersion (SAD) method. Five selenium sites were found in the structure with hkl2map (http://webapps.embl-hamburg.de/hkl2map/), giving an overall figure of Merit of 0.5 calculated by PHENIX.Autosol (https://www.phenix-online.org/documentation/reference/autosol.html). The further PHENIX run using AutoBuild resulted in an initial model with more than 75% of the residues built. This model was subjected to iterative cycles of refinement and model rebuilding until the final model reached acceptable parameters in geometry and fitting. N-AvrRxo1-ORF1:AvrRxo1-ORF2 (2.30 Å, native crystal of AvrRxo1-

ORF1:AvrRxo1-ORF2 complex), SO_4-AvrRxo1-ORF1:AvrRxo1-ORF2 (1.64 Å, crystallized in ammonium sulfate salt) and ATP-AvrRxo1-ORF1:AvrRxo1-ORF2 (1.64 Å, co-crystallized with ATP and magnesium acetate) crystal structures were solved by molecular replacement using the solved Se-Met: AvrRxo1-ORF1:AvrRxo1-ORF2 complex structure as a search model.

The complex proteins have a molecular weight of around 100 kDa compared with protein standards in gel-filtration chromatography (Figure S2-1 in Supplementary Information). Based on the molecular weight, the complex consists of two AvrRxo1-ORF1 and two AvrRxo1-ORF2 molecules, therefore it appears to form a heterotetramer. The structure of N-AvrRxo1-ORF1:AvrRxo1-ORF2 complex is depicted in Figure 2-1a and 2-1b. The complex is a heterotetramer analyzed by PISA [31], which is consistent with the gel filtration results. AvrRxo1-ORF1 and AvrRxo1-ORF2 form a compact heterodimer (Figure 2-1c), and two heterodimers form a tetramer with 2-fold rotational symmetry (Figure 2-1a). In the N-AvrRxo1-ORF1:AvrRxo1-ORF2 structure, the interface area between AvrRxo1-ORF1 and AvrRxo1-ORF2 is 1608 $Å^2$, the interface area between AvrRxo1-ORF2 and AvrRxo1-ORF2 is 988 $Å^2$, and the interface area between AvrRxo1-ORF1 and AvrRxo1-ORF1 is 279 $Å^2$. The lists of hydrogen bonds and salt bridges in the interfaces were generated using PISA. There are 18 hydrogen bonds and nine salt bridges between AvrRxo1-ORF1 and AvrRxo1-ORF2 molecules, four hydrogen bonds and four salt bridges between two AvrRxo1-ORF2 molecules and no hydrogen bond and salt bridge formed between two AvrRxo1-ORF1 molecules in the N-AvrRxo1-ORF1:AvrRxo1-ORF2 complex (Table S2-1 in Supplementary Information).

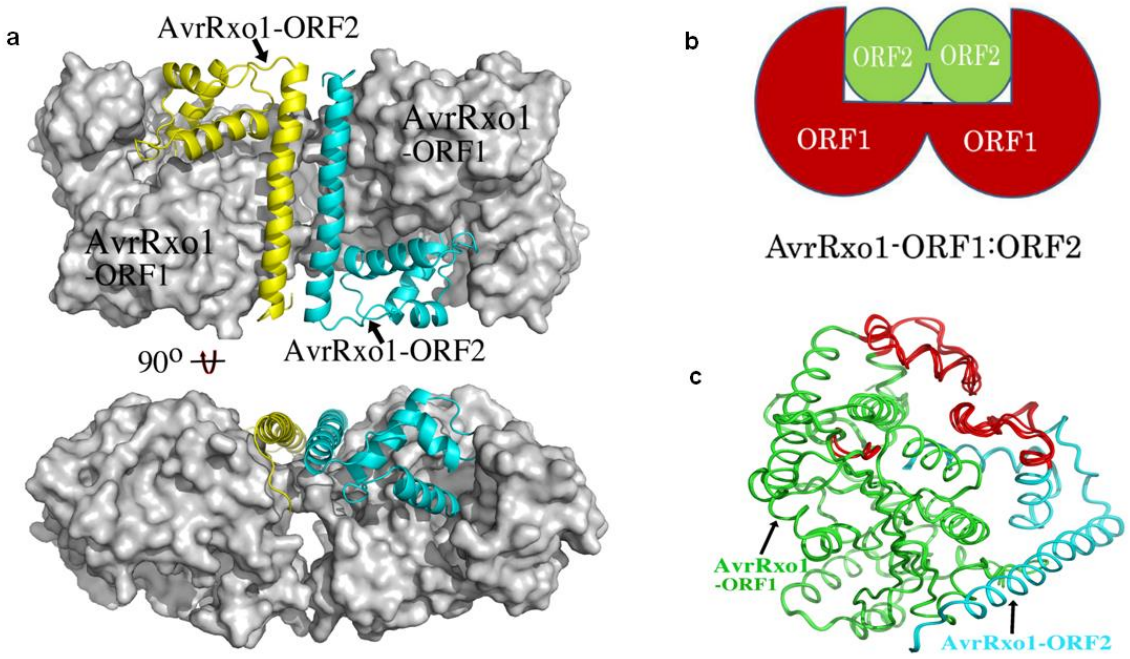


Figure 2-1| Assembly of AvrRxo1-ORF1:AvrRxo1-ORF2 complex and regions having significant conformational changes revealed by a comparison of four complex structures

a. Two molecules of AvrRxo1-ORF1 are presented in the surface presentation with light grey color. Two molecules of AvrRxo1-ORF2 are shown in cartoon presentation using yellow and cyan color. The lower image is a view of a 90° horizontal rotation of the top view. **b.** Tetramer organization of AvrRxo1-ORF1: -ORF2 complex. In the complex both AvrRxo1-ORF1 and -ORF2 are involved in tetramer formation. **c.** A superposition of all four complexes, with the backbones shown in ribbon presentation. AvrRxo1-ORF1 is shown in green, AvrRxo1-ORF2 in cyan and regions having conformational changes are in red.

Complex formation and the interaction interfaces

By carefully comparing the interface areas and number of hydrogen bonds and salt bridges involved in the complex formation of the following four complex structures, we did not observe major differences: structures crystallized in the different conditions (N-AvrRxo1-ORF1:AvrRxo1-ORF2 and SO_4-AvrRxo1-ORF1:AvrRxo1-ORF2), and between native and selenomethionine-substituted proteins (N-AvrRxo1-ORF1:AvrRxo1-ORF2 and Se-Met-AvrRxo1-ORF1:AvrRxo1-ORF2). However, when co-crystallized with ATP and Mg^{++} (ATP-AvrRxo1-ORF1:AvrRxo1-ORF2) the complex is very different from the previous complexes regarding interface areas, number of hydrogen bonds and salt

bridges involved in complex formation. The molecules bind more tightly to one another in the ATP-AvrRxo1-ORF1:AvrRxo1-ORF2 complex than in other complexes. In the ATP-AvrRxo1-ORF1:AvrRxo1-ORF2 complex, the interface area between AvrRxo1-ORF1 and AvrRxo1-ORF2 is 1787 $Å^2$, the interface area between AvrRxo1-ORF2 and AvrRxo1-ORF2 is 944 $Å^2$, and the interface area between AvrRxo1-ORF1 and AvrRxo1-ORF1 is 502 $Å^2$. There are 18 hydrogen bonds and eight salt bridges between AvrRxo1-ORF1 and AvrRxo1-ORF2 molecules, ten hydrogen bonds and six salt bridges between two AvrRxo1-ORF2 molecules and 12 hydrogen bonds between two AvrRxo1-ORF1 molecules in the ATP-AvrRxo1-ORF1:AvrRxo1-ORF2 complex (Table S2-2). From the structural data, ATP and Mg^{++} may affect the complex formation or the equilibration between the different stages of complex polymerization.

To determine if ATP and Mg^{++} could affect conformational change inside each individual molecule, we did a comparison of four complexes, which may provide a valuable insight into the dynamics of the molecular interaction in the complex. A superposition of four independent AvrRxo1-ORF1:AvrRxo1-ORF2 complexes can be seen in Fig. 1B, and the regions that are less conformationally constrained are illustrated in the figure. These significant conformational changes include two parts of the AvrRxo1-ORF1 molecule and one of the AvrRxo1-ORF2 molecules (Figure 2-1c). However, these conformational changes are not associated with ligand binding or crystallization condition.

Structural features of AvrRxo1-ORF1 and its function

A Dali search [32] with the AvrRxo1-ORF1 structure revealed that AvrRxo1-ORF1 is most similar to Zeta toxin protein, a toxin from *Streptococcus pyogenes* [33] (pdb code: 1gvn, Z = 16.6, rmsd = 3.4, sequence identity = 12%). The Dali search result listed eight unique proteins (40 different chains) with Z score over 9.2. All eight proteins have a P-loop (Fig. S2B), a common motif in ATP- and GTP-binding proteins and have been assigned as kinases or phosphotransferases [34-40]. 3-D alignments of eight kinase structures and AvrRxo1-ORF1 revealed that they all have common fold core structures of the N-terminal kinase domain of the T4 polynucleotide kinase (T4pnk) from an enterobacteria phage [34] (pdb code 1ly1, Z = 10.0, rmsd = 2.8, sequence identity =13%), which are conserved in spite of structure-based sequence alignments that reveal very low pair-wise sequence identity. AvrRxo1-ORF1 is composed of a conserved T4pnk domain flanked by additional N-terminal and C-terminal domains. The N-terminal domain of AvrRxo1-ORF1 consists of one long α-helix and two short α-helices, the latter of which are involved in complex formation. The C-terminal domain is also involved in complex formation and is comprised of three

β-strands and three major helices (Fig. 2-2a). All AvrRxo1-ORF1 structures from the four complexes reported here bind either two phosphate or two sulfate ions in the putative ATP binding site revealed by T4pnk structures [34, 35] (Figure 2-3). By superimposing all four complex structures, we noticed that phosphate (Figure 2-3a and 2-3c) and sulfate ions (Figure 2-3b and 2-3c) occupied slightly different positions in AvrRxo1-ORF1. Based on the phosphate and sulfate binding information, we defined an ATP binding site by using molecular docking. Interestingly, one of the phosphate ions occupied the γ-phosphate position of ATP and one of the sulfate ions occupied the α-phosphate position of ATP (Figure 2-3d). The phosphate or sulfate ions occupied the ATP binding site of AvrRxo1-ORF1, likely resulting in steric hindrance that prevented ATP interaction and co-crystallization with AvrRxo1-ORF1.

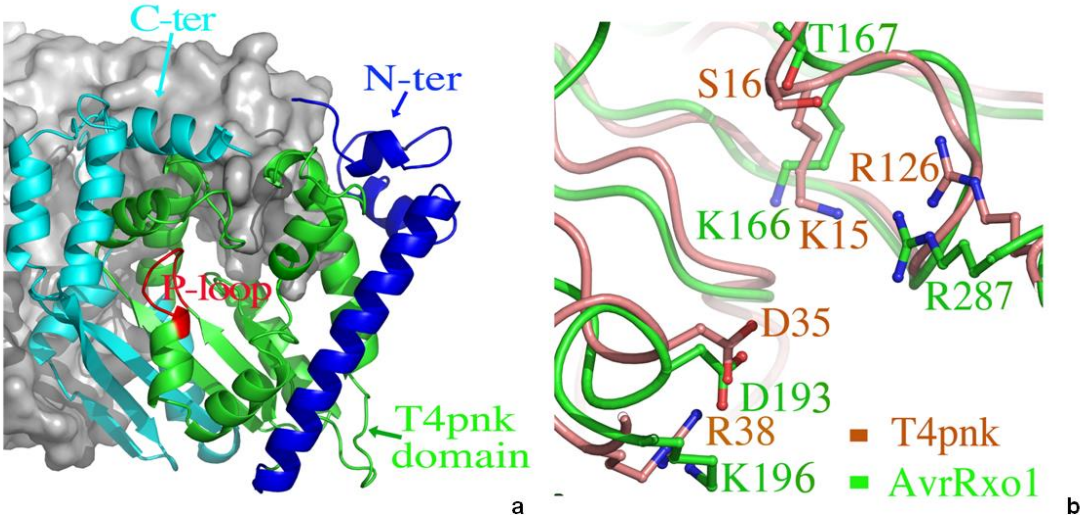


Figure 2-2| Structural fold and putative active site residues of AvrRxo1-ORF1

One AvrRxo1 molecule in the complex is shown in cartoon presentation while other parts of the complex are in surface presentation in light green. **a.** AvrRxo1 consists of three major domains: N-terminal domain (blue), T4pnk kinase domain (green), and C-terminal domain (cyan). P-loop is red. **b.** A superposition of AvrRxo1 (green) and the T4pnk kinase domain (pdb, 1ly1) (brown). Key residues for T4pnk kinase catalysis identified previously are shown in sticks, with brown carbon, red oxygen, and blue nitrogen atoms. The equivalent residues from AvrRxo1 are shown in sticks, with green carbon, red oxygen, and blue nitrogen atoms.

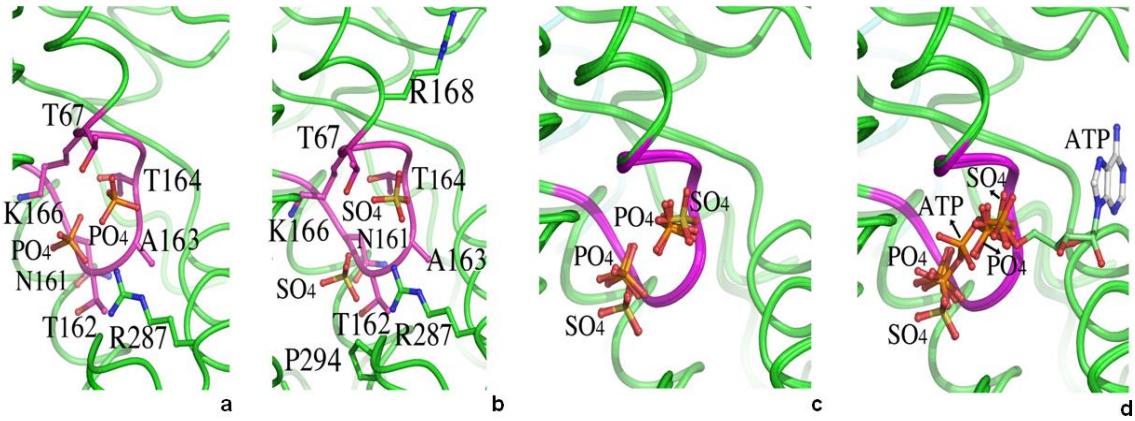


Figure 2-3| Binding sites of phosphate, sulfate ions, and ATP in the AvrRxo1-ORF1 active center

Parts of the AvrRxo1-ORF1 structures are shown in ribbon presentation with green color and P-loops shown in magenta. **a.** Phosphate ion binding sites in the active center of the AvrRxo1 molecule from the N-AvrRxo1-ORF1:-ORF2 structure. Residues around 4 Å of two phosphate ions are shown in sticks and residue numbers are labeled. **b.** Sulfate ion binding sites in the active center of the AvrRxo1 molecule from the SO_4-AvrRxo1-ORF1:-ORF2 structure. Residues around 4 Å of two sulfate ions are shown in sticks and residue numbers are labeled. **c.** Superposition of all four complexes. The phosphate ion (orange) binding sites, seen in N-AvrRxo1-ORF1:-ORF2 and Se-Met-AvrRxo1-ORF1:-ORF2 structures, are similar, and the sulfate ion (yellow) binding sites, revealed in SO_4-AvrRxo1-ORF1:-ORF2 and ATP-AvrRxo1-ORF1:-ORF2 structures, are the same. **d.** A superposition of the AvrRxo1-ORF1: ATP complex obtained by molecular docking onto the four superimposed structures shown in panel **C**. Note that one of the phosphate ions occupied the γ-phosphate position of ATP and one of the sulfate ions occupied the α-phosphate position of ATP.

The T4pnk kinase domain catalyzes the transfer of the γ-phosphate from ATP or other nucleoside triphosphates to the 5′-OH terminus of DNA, RNA, and nucleoside 3′-monophosphates [41]. Residues K15, S16, D35, R38, R126 in T4pnk are key residues responsible for its kinase activity [41, 42]. Interestingly, all of the other eight structures share conserved T4pnk active site residues. AvrRxo1-ORF1 has equivalent residues K166 (K15), T167 (S16), D193 (D35), K196 (R38), R287 (R126), which fold and occupy similar positions as the key active site residues in T4pnk (Figure 2-2b). The substrate binding sites of the T4pnk kinase domain have been identified[43]. A superposition of AvrRxo1-ORF1 onto the T4pnk kinase domain revealed that AvrRxo1-ORF1 has a similar site for nucleotide binding as

the T4pnk kinase domain (Figure 2-4b). We used HADDOCK Protein-DNA docking [44] to see if the same T4pnk binding substrate, a single strand tri-nucleotide (GTC), can be docked to the similar position of AvrRxo1-ORF1. The docking result showed that the binding site (Figure 2-4b and 2-4c) with the best HADDOCK score (Table S2-3 in Supplementary Information) is similar to the binding site of GTC in the T4pnk kinase domain (Figure 2-4). Interestingly, this nucleotide binding cavity is occupied by the AvrRxo1-ORF2 in the complex, and renders AvrRxo1-ORF1 in the complex in an inactive form. Based on the structure of AvrRxo1-ORF1, it is highly likely that AvrRxo1-ORF1 is a kinase of an unknown substrate, and that AvrRxo1-ORF2 blocks this kinase activity.

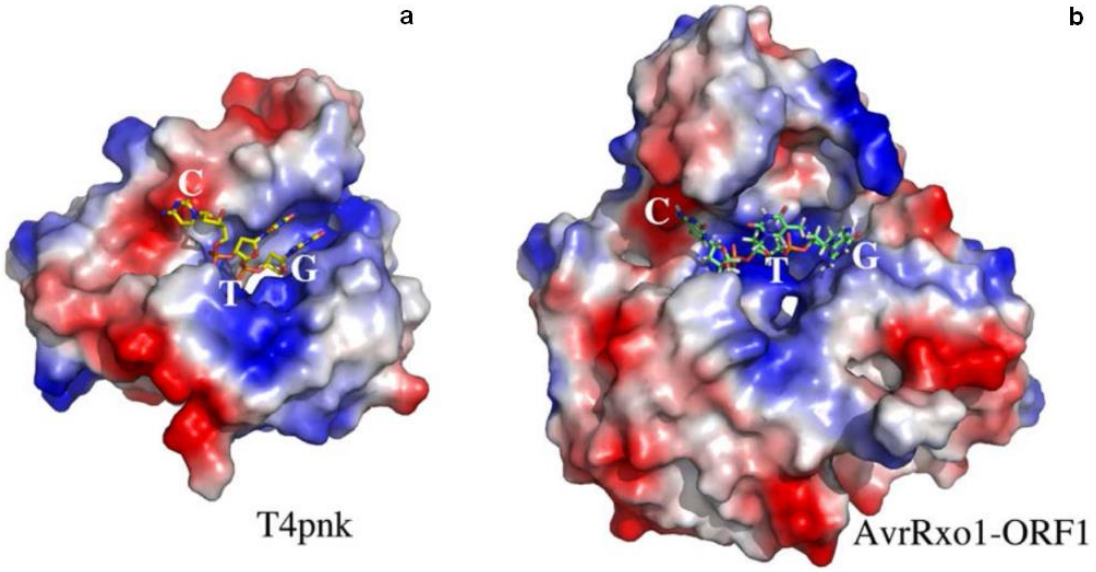


Figure 2-4| Polynucleotide substrate binding in T4pnk and AvrRxo1-ORF1

a. T4pnk is shown in surface presentation, and colored with electrostatic property. Positive potentials are drawn in blue, negative potentials in red. The single strand tri-nucleotide, GTC binding site of T4pnk was generated from the T4pnk: GTC complex structure (pdb code: 1rc8). GTC is shown in sticks, with yellow carbon, red oxygen, blue nitrogen and orange phosphorus atoms. **b.** GTC binding site in AvrRxo1-ORF1 was obtained by protein-DNA docking. The AvrRxo1-ORF1 is shown in surface presentation and colored to show its electrostatic property. GTC is shown in sticks, with green carbon, red oxygen, blue nitrogen, grey hydrogen and orange phosphorus atoms.

To determine whether key catalytic sites important for kinase activity of T4pnk correspond to sites important to AvrRxo1-ORF1 activity, AvrRxo1-ORF1 was inactivated at the sites of the predicted substrate-binding residue D193 or the

ATP-binding residue T167. Since AvrRxo1-ORF1 has structural similarity to Zeta toxin, a microbial toxin that inhibits the proliferation of *E. coli* [39], we tested the effect of AvrRxo1-ORF1 expression on the growth of *E. coli*. As shown in Figure 2-5a, expression of AvrRxo1-ORF1 inhibits bacterial growth and this activity are abolished by mutations of the predicted substrate-binding residue D193 or the ATP-binding residue T167. Therefore, wild type AvrRxo1 indeed possesses a growth suppressive function that depends on its putative kinase activity. AvrRxo1-ORF1 bacteriostatic activity is inhibited by co-expression with AvrRxo1-ORF2 (Figure 2-5a), consistent with the structural placement of AvrRxo1-ORF2 at the predicted substrate-binding site. The conserved kinase domain suggests that AvrRxo1-ORF1 phosphorylates substrates that are critical for bacterial growth. In addition to inhibiting bacterial growth, AvrRxo1-ORF1 also triggers T167-dependent plant cell death after transient expression in diverse plant species (Figure 2-5b). Interesting, the cell death phenotype triggered by AvrRxo1-ORF1 can be completely inhibited by co-expression of AvrRxo1-ORF2 (Figure 2-5b). This may suggest that AvrRxo1-ORF1 has a phosphorylation target that is conserved in prokaryotes and eukaryotes, and AvrRxo1-ORF2 may function as antitoxin to inhibit the putative kinase active of AvrRxo1-ORF1 in both prokaryotes and eukaryotes.

To test if AvrRxo1 enhances *Xanthomonas* bacterial virulence and if the virulence function depends on the kinase domain of AvrRxo1, we introduced a plasmid that contains either the native AvrRxo1-ORF1:ORF2 gene [26] or mutant AvrRxo1-ORF1-T167N:ORF2 gene into the *Xanthomonas oryzae* (*Xo*) strain X11-5A. X11-5A is a less virulent *Xo* strain isolated in the USA [45]. As shown in Figure 2-5c, when inoculated on rice plants (cv. Kitaake), populations of X11-5A carrying wild type AvrRxo1-ORF1:ORF2 increase to higher numbers after two days than *Xo* strain X11-5A carrying AvrRxo1-ORF1-T167N:ORF2. Surprisingly, the ATP binding site mutant reached higher population numbers than the vector control, possibly because the mutant protein retained ability to bind and sequester the hypothetical phosphorylation target. Thus, AvrRxo1 functions to enhance bacterial multiplication in plants, and full virulence function is dependent on its putative kinase activity.

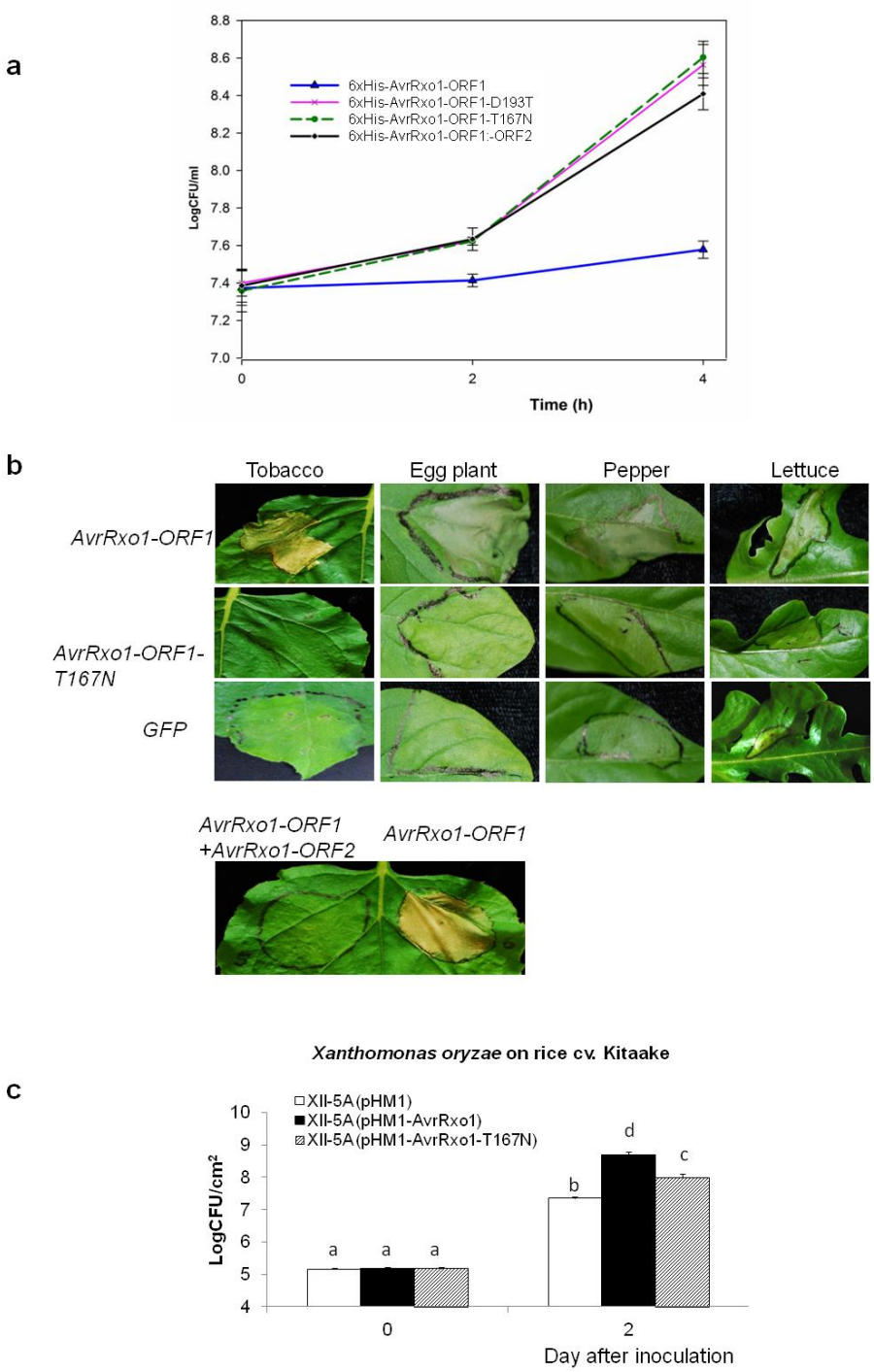


Figure 2-5| Functional assay of wild type and mutant AvrRxo1

a. Wild type AvrRxo1, but not AvrRxo1-ORF1:AvrRxo1-ORF2, or mutant AvRxo1-ORF1, inhibits bacterial growth. *E. coli* strains carrying pDEST527-*AvrRxo1-ORF1*, *AvrRxo1-ORF1-T167N*, *AvrRxo1-ORF1-D193T*, or *AvrRxo1-*

ORF1: -ORF2 were grown at 37°C in LB liquid medium. Protein expression was induced by adding 0.5 mM IPTG. Bacterial growth was monitored by dilution plating. The mean data of three replicates are presented (The means ± s.d.; n=3). **b.** Transient expression of AvrRxo1, but not AvrRxo1-T167N, triggered a cell death phenotype in diverse plant species. *Agrobacterium* strain GV2260 carrying construct pEarleyGate101-AvrRxo1, pEarleyGate101-AvrRxo1-T167N, or pEarleyGate101-GFP (A600=0.4), was infiltrated into the leaf tissue of wild type tobacco (*Nicotiana benthamiana*), eggplant (*Solanum melongena* cv. Black Beauty), pepper (*Capsicum annuum* cv. ECW), and lettuce (*Lactuca sativa* cv. Giant Caesar). Photographs were taken at 3 days post inoculation. **c.** Co-expression of AvrRxo1-ORF2 with AvrRxo1-ORF1 suppressed the cell death phenotype. *Agrobacterium* strain GV2260 carrying construct pEarleyGate101-AvrRxo1-ORF1 alone (A600=0.4) or mixed equally with a suspension of *Agrobacterium* carrying pEarleyGate202-AvrRxo1-ORF2, were infiltrated into the leaf tissue of *Nicotiana benthamiana*, and the responses were photographed at 3 days post inoculation. **d.** Expression of wild type AvrRxo1 in *Xanthomonas* enhanced bacterial proliferation on rice plants. Leaves of rice variety Kitaake were inoculated with derivatives of *Xanthomonas oryzae* strain X11-5A carrying pHM1-AvrRxo1, pHM1-AvrRxo1-T167N, or the empty vector pHM1. Bacterial multiplication in leaves was determined in triplicate at 0 and 2 days post inoculation. The means ± s.e.; (n=9; $p<0.01$, Student's t-test, different letters indicate significant difference) are presented.

Structural features of AvrRxo1-ORF2

A Dali search [32] with the AvrRxo1-ORF2 structure did not reveal any homolog structure with a Z score higher than 3.9. The most similar structure is the C-terminal domain of DUF199/WhiA from *Thermatoga maritime*[46] (pdb code: 3hyi, Z = 3.9, rmsd = 4.1). 3-D structural alignment showed that the sequences from both structures only partially aligned, including the first α-helices from both structures and the third α-helix from AvrRxo1-ORF2 and the fourth α-helix from the C-terminal domain of DUF199/WhiA (Figure 2-6a and 2-6c). The AvrRxo1-ORF2 structure represents a new protein fold comprised of four α-helices. The first and longest α-helix is key to tetramer formation by forming hydrogen bonds, salt bridges and hydrophobic interactions between two AvrRxo1-ORF2 molecules in the complex. α-helices two to four and the loop/turn links between them bind to the predicted substrate binding cavity of the AvrRxo1-ORF1 molecule (Figure 2-6b) and this structural feature represents a probable new kinase-binding domain. Although *AvrRxo1-ORF2* is encoded downstream of a T3S effector and produces an effector-binding protein, our previous work showed that it is not essential for

translocation into host cells[26], and this study revealed that AvrRxo1-ORF2 has no structural similarity to characterized T3S effector chaperones. Therefore, we conclude that AvrRxo1-ORF2 is not a typical T3S effector-binding chaperone, but that it may instead function as an antitoxin to protect the bacteria from the bacteriostatic activity of AvrRxo1-ORF1. Further studies are needed to determine whether -ORF2 has any activities of typical T3S effectors.

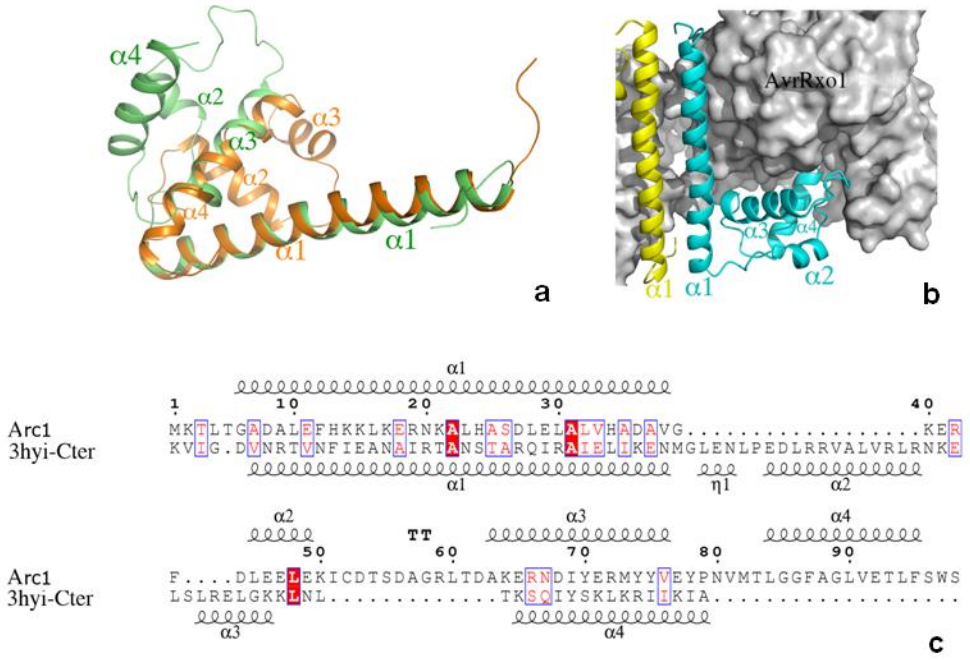


Figure 2-6| Structural fold and 3D alignment with a similar structure of AvrRxo1-ORF2

The most similar structure is the C-terminal domain of DUF199/WhiA from *Thermatoga maritime* (Kaiser et al., 2009) (pdb code: 3hyi, Z = 3.9, rmsd = 4.1). 3D structural alignment showed that the sequences from both structures were only partially aligned, the first α-helices from both structures and the third α-helix from AvrRxo1-ORF2 and the fourth α-helix from the C-terminal domain of DUF199/WhiA. **a.** Cartoon backbone representation of the 3D alignment of the 3hyi C-terminal domain (orange) and AvrRxo1-ORF2 (green). Labels describing secondary structural information are provided for both proteins. **b.** Part of the AvrRxo1-ORF1:-ORF2 complex is shown. Surface presentation of two AvrRxo1 molecules (light grey) and cartoon presentation of two AvrRxo1 molecules (cyan and yellow) are shown. The first α-helix (α1) from one AvrRxo1-ORF2 molecule (cyan) contacts α1 from the other AvrRxo1-ORF2 molecule (yellow) and α-helices 2 to 4 (α2 to α4) and the loop/turn links between them bind to a big cavity of the AvrRxo1 molecule. **c.** The secondary structural information of both proteins

is provided next to their protein sequences in the alignment. AvrRxo1-ORF2 residue numbers were labeled on top of the alignment.

Concluding remarks

We characterized the *Xanthomonas* T3S effector AvrRxo1-ORF1 using structural biology and plant transient assay methods. AvrRxo1-ORF1 folds into three domains, the N-terminal region (residue 87-151), a central domain that contains the ATP binding site (residue 152-322), and a C-terminal domain (residue 323-421). AvrRxo1-ORF1 forms a heterotetramer with AvrRxo1-ORF2, in which -ORF2 binds to the central kinase domain of AvrRxo1-ORF1. AvrRxo1-ORF1 inhibits bacterial growth in a manner dependent on the catalytic and ATP binding sites predicted by similarity to T4pnk domains. An *X. oryzae* strain expressing wild type AvrRxo1-ORF1-ORF2 demonstrated enhanced bacterial proliferation in rice plants, and the bacterial strain expressing a AvrRxo1-T167N mutant was diminished in that ability, further suggesting that the predicted kinase activity of AvrRxo1-ORF1 is important to its function. Based on its functional activity and structural similarity to polynucleotide kinases, it is possible that AvrRxo1-ORF1 targets a universally important oligonucleotide or nucleotide sugar as a substrate. To date, plant pathogen T3Es are known to act as proteases, phosphatases, acetyltransferases, ubiquitin ligases, nucleotide transferases, phosphothreonine lyases, and transcriptional activators [21, 47]. Modifying phosphorylation states of signaling proteins is a common mechanism of suppressing plant effector triggered immunity, and several effectors act to block phosphorylation or remove phosphate groups from host proteins[48]. AvrB and AvrRpm1 increase phosphorylation of RIN4 by the host kinase RIPK [49, 50], but are not thought to act as kinases themselves. The finding that AvrRxo1 has a T4pnk domain provides intriguing evidence that an effector might directly phosphorylate a host target, although demonstration of AvrRxo1 kinase activity will be needed to demonstrate this conclusively.

Material and methods

Cloning, expression, and purification of AvrRxo1-ORF1: -ORF2 complex

DNA fragments containing AvrRxo1-ORF1 (65-421aa) and -ORF2 genes[26] were amplified from the genomic DNA of *Xanthomonas oryzae* pv. *oryzicola* strain BLS256 and cloned into pENTR-D-TOPO vector (Invitrogen, Carlsbad, CA) (Primers listed in Table S2-4). The N-terminal 65 amino acids, presumed to encode a secretion signal, were omitted in purification experiments.

The *AvrRxo1-ORF1*(65-421aa) and -*ORF2* genes were subcloned into a modified pGEX4T-1 destination vector through LR Gateway cloning (Invitrogen), where a ccdB cassette was inserted into the *Bam*HI and *Sal*I sites in pGEX4T-1 (GE Healthcare Bio-Sciences, Pittsburgh, PA). To facilitate the removal of GST tag, we also introduced a TEV protease cleavage site between GST and the targeted protein. The resulting plasmid was transformed into *E. coli* C41 cells (Lucigen, Middleton, WI) and grown overnight in 50 ml of LB medium containing 100 mg/L ampicillin at 37°C. The culture was transferred into 2 liter of LB medium containing 100 mg/L ampicillin to reach an A600 of 0.8 units. Isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added to a final concentration 1 mM with subsequent incubation at 250 rpm at 28°C for 8 h. The bacterial cells were harvested and the cell pellets were broken by incubating with 1 μg/ml lysozyme on ice for 1 h, followed by sonication on ice. The lysate was centrifuged at 12,000 g for 20 min at 4^{o}C and the supernatant was gently collected and used for subsequent purification. Glutathione Sepharose 4B (GenScript, Piscataway, NJ) affinity resin, Q-sepharose ion exchange column were used for to purify the complex. The purity of the purified proteins was evaluated both by 12% SDS–PAGE and 8% native-PAGE. Protein concentration was determined by a protein assay kit from Bio-Rad (Hercules, CA) using bovine serum albumin as a standard. Selenomethionine-substituted proteins were produced in the same strain of *E. coli*. The incorporation of selenomethionine into the recombinant proteins was accomplished by using a special medium, Selenomethionine Expression Media (Medicilon Inc, Shanghai, China). The purification of selenomethionine-containing AvrRxo1-ORF1:-ORF2 complex followed the same protocols as those of the unmodified complexes.

X-ray crystallography

Diffraction quality crystals of N-AvrRxo1-ORF1:AvrRxo1-ORF2 (native protein) and Se-Met-AvrRxo1-ORF1:AvrRxo1-ORF2 (selenomethionine-labeled) were obtained using the hanging drop method with initial droplets created by mixing 1 μL of protein (10.0 mg/mL) with 1 μL of a crystallization solution composed of 0.8 M NaH_2PO_4, 0.8 M KH_2PO_4, 15% Glycerol, 8 mM DTT, 0.1 M HEPES, pH7.5. Crystals of SO_4-AvrRxo1-ORF1:AvrRxo1-ORF2 (sulfate complex) were obtained using the same method in a different crystallization buffer (1.6 M $(NH_4)_2SO_4$, 18% Glycerol, 8 mM DTT, 100 mM Na-Citrate, pH 5.0). ATP-AvrRxo1-ORF1:AvrRxo1-ORF2 (co-crystallized with ATP and magnesium acetate) was co-crystallized with 30 mM ATP and 60 mM Magnesium Acetate in the same buffer for SO_4-AvrRxo1-ORF1:AvrRxo1-ORF2 crystals. All crystals appeared after 2 days at 22 °C.

Data collection and processing

Diffraction data of crystals were collected at the Brookhaven National Synchrotron Light Source beam line X29A (λ = 0.9790 Å or λ =1.075 Å). Data were collected using an ADSC Q315 CCD detector. All data were indexed and integrated using HKL-2000 software. Scaling and merging of diffraction data were performed using the program SCALEPACK[51].

Structure determination

The structure of Se-Met-AvrRxo1-ORF1:AvrRxo1-ORF2 complex was determined by the SAD phasing technique with selenium anomalous signals. Heavy atom sites were located using hkl2map[52] and the initial SAD phases were calculated by PHENIX.Autosol[53]. Initial models were built using PHENIX.AutoBuild[53]. The model was refined using SAD refinement with optimization of the selenium occupancy in Refmac 5.2[54]. Final models were produced after numerous iterative rounds of manual re-building in Coot[55] and refinement in Refmac 5.2. The structures of native AvrRxo1-ORF1:AvrRxo1-ORF2 complexes were determined by the molecular replacement method using the solved structure as a starting model. The programs MOLREP[56] and Phaser[57] were employed to calculate both cross-rotation and translation of the model. The initial model was subjected to iterative cycles of crystallographic refinement with the Refmac 5.2 and graphic sessions for model building using the program Coot[55]. Solvent molecules were automatically added and refined with ARP/wARP[58] together with Refmac 5.2.

Molecular Docking

To obtain insight into adenosine triphosphate (ATP) binding of AvrRxo1-ORF1, AutoDock Vina was used for molecular docking[59]. ATP molecules for AutoDock Vina were prepared by using Marvin 5.3.7, 2010, ChemAxon (http://www.chemaxon.com). AutoDock Tools 1.5.4 (http://mgltools.scripps.edu/documentation/how-to/citing-pmv-adt-and-vision) were used to prepare the ligand and the receptor, after which docking was performed using AutoDock Vina. All protein side chains were immobilized and the rotatable bonds were left free to rotate for the ATP molecule. The X, Y, and Z dimensions of the grid were set based on phosphate and sulfate binding sites in the putative ATP binding sites revealed by similar ATP binding protein (T4pnk) in the AvrRxo1-ORF1:AvrRxo1-ORF2 complex structure. The grid box covered the whole cavity. The lowest energy configuration was considered the best docking

pose for ATP. Protein-DNA docking was performed with default parameters using the web server version of HADDOCK[44]. The HADDOCK score was used to rank the generated poses. It is a weighted sum of intermolecular electrostatic, van der Waals, desolvation and ambiguous interaction restraint energies. The final structures were then clustered using the pairwise backbone rmsd at the interface using a 7.5 Å cutoff. The top ranked cluster based on the HADDOCK scores is the most reliable, and the first complex from Cluster 1 was used in the comparison of the binding sites of a single stand nucleotide, GTC in AvrRxo1-ORF1 and T4pnk[43].

Structure analysis

Superposition of structures was done using Lsqkab[60] in the CCP4 suite. Figures were generated using PyMOL[61]. Protein surface properties were analyzed using PyMO[61]. 3D structural alignment was made using STRAP[62].

Plasmid Construction and Site-Directed Mutagenesis in Function Studies

Wild type *AvrRxo1-ORF1* (corresponding to residues 65-421) and *AvrRxo1-ORF2* was amplified from p1-9[26] with primers listed in *SI* text and cloned into to TopoEntrD (Invitrogen). This TopoEntrD-AvrRxo1-ORF1 construct was used to develop *AvrRxo1-ORF1* variants using primers listed in Table S2-4. For *Agrobacterium*-mediated transient assays, wild type or mutant AvrRxo1-ORF1-T167N were cloned to the Gateway compatible pEarleyGate101, and AvrRxo1-ORF2 was cloned pEarleyGate202 to through LR cloning[63], where the targeted genes were driven by the CaMV35S promoter.

For the plant infection assay, *avrRxo1-ORF1-avrRxo1-ORF2* variants were subcloned into a broad host spec vector pHM1[26]. pHM1-AvrRxo1-ORF1-T167N, carrying AvrRxo1-ORF1 and ORF2 with a T167N ATPase site substitution in AvrRxo1-ORF1, was generated by cloning the *Sph*I fragment from pHM1-AvrRxo1-ORF1[26] into the vector pGEM T-Easy (Promega, Madison, WI). The site-directed mutation was introduced using primers AvrRxo1-ORF1_T167N for and AvrRxo1-ORF1_T167N rev. The *Sph*I fragment was cloned into pHM1 for *X. oryzae* expression. Electrotransformation was used to transfer plasmid DNA (empty vector pHM1, pHM1-AvrRxo1-ORF1 (T167N), and pHM1-AvrRxo1-ORF1) into *X. oryzae* strain X11-5A as described previously [26].

Monitoring the growth rates of E.coli strains expressing AvrRxo1

The AvrRxo1-ORF1, AvrRxo1-ORF1: -ORF2, AvrRxo1-ORF1-D193T, and AvrRxo1-ORF1-T167 were cloned into the Gateway-compatible pDEST527 (Addgene.org) and expressed in *E. coli* strain BL21 (DE3). Strains were grown at 37 °C in LB liquid medium supplemented with 100 μg/ml ampicillin. The bacterial culture was diluted to A600=0.01, and expression was induced by adding 0.5 mM Isopropyl-1-thio-β-D-galactopyranoside (IPTG). The controls were the bacterial cultures without induction. The bacterial growth was monitored by dilution plating. The mean data of three replicates presented. The experiments were repeated three times with similar results.

Plant Disease Assay and Agrobacterium-mediated transient assay

Xanthomonas oryzae strain X11-5A is a weakly virulent strain isolated in the USA, and was used to express wild type and mutant AvrRxo1-ORF1-T167N for inoculation on rice cv. Kitaake plants. For bacterial population studies, four-week old rice plants (cv. Kitaake) grown in growth chamber were inoculated with derivatives of *X. oryzae* strain X11-5A carrying pHM1, pHM1-AvrRxo1-ORF1:-ORF2, or pHM1-AvrRxo1-ORF1-T167N:-ORF2 as described previously[27]. Bacterial multiplication in the leaves was determined in triplicate at 0 and 2 days post inoculation. Leaf samples were surface-sterilized, ground in 1 ml of sterile ddH_2O, serially diluted, and plated on nutrient agar [26]. The experiments were repeated three times with similar results.

Agrobacterium tumefaciens GV2260 was used for transient expression in *Nicotiana benthamiana*. *Agrobacterium* strains carrying different constructs were adjusted to A600=0.4, and used to infiltrate to plant mesophyll tissue using the blunt end of syringes without needles as described previously [64]. The plant cell death phenotype was photographed at 3 days post inoculation.

Acknowledgements

This work was carried out in part at the National Synchrotron Light Source, Brookhaven National Laboratory. The project was supported by an NSF grant (IOS-0845283) to B. Z. and USDA NIFA (#2011-67012-30570 and 2014-67013-21564) to L.T. and J.E.L., an integrated, internal competitive grant from VAES, VCE, and the College of Agriculture and Life Sciences at Virginia Tech, and a Hainan University Start-up Grant for New Faculty.

Supplementary information

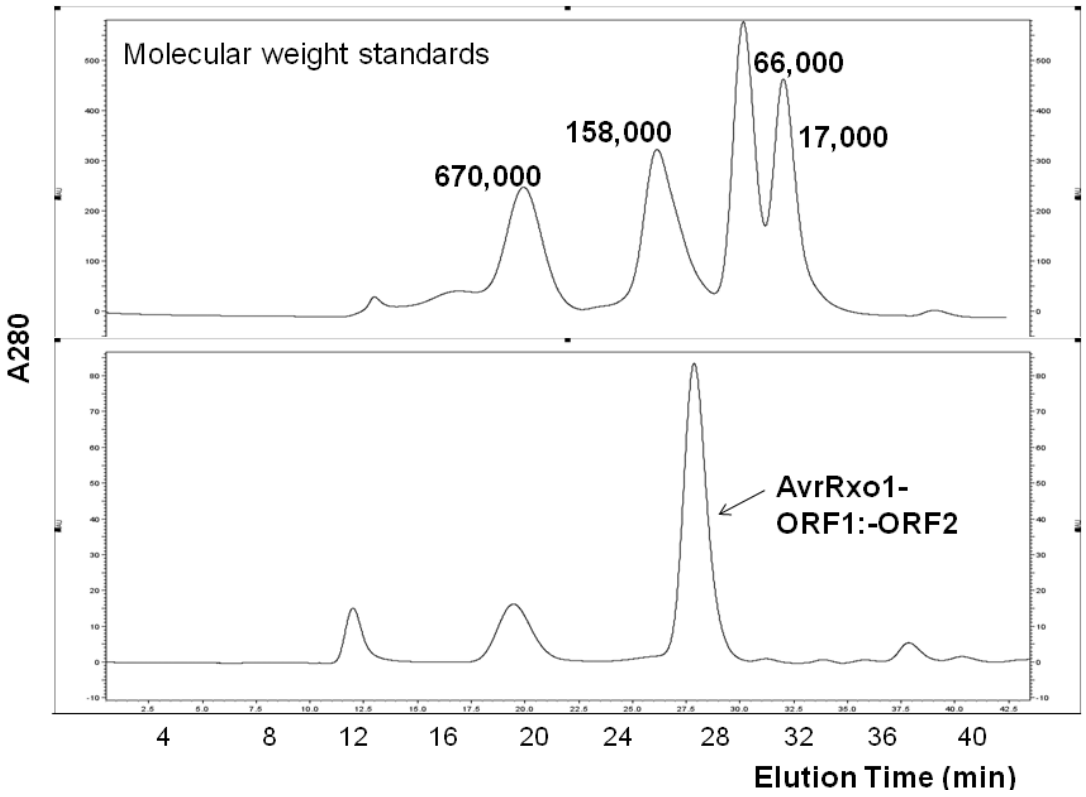


Figure S2-1| Gel filtration chromatography of AvrRxo1-ORF1: -ORF2 complexes

Purified AvrRxo1-ORF1: -ORF2 complex fraction was subjected to gel filtration chromatography using a Pharmacia Superose-6 column (1 x 30 cm). A mobile phase consisted of 50 mM HEPES buffer (pH 6.5) containing 120 mM NaCl was used for protein separation at a flow rate of 0.5 mL min^{-1}. At the applied condition of chromatography, the retention time of the AvrRxo1-ORF1: -ORF2 complex was about 28 min. Based on comparison with molecular weight standards (thyroglobulin, 670,000, γ-globulin 158,000, ovalbumin 66,000 and myoglobin, 17,000), the AvrRxo1-ORF1:-ORF2 complex behaved as a protein with a molecular weight of 120,000 D that is likely to be a tetramer.

Table S2-1| Lists of hydrogen bonds and salt bridges in the interaction interfaces in the N-AvrRxo1-ORF1:-ORF2 complex

Hydrogen bonds

##	AvrRxo1-ORF2	Distance (Å)	AvrRxo1-ORF1
1	B:LEU 4[N]	2.96	A:VAL 351[O]
2	B:HIS 13[ND1]	2.77	A:TYR 357[OH]
3	B:HIS 13[NE2]	2.73	A:ASP 316[OD1]
4	B:LYS 17[NZ]	2.90	A:ASP 316[OD2]
5	B:ASN 20[ND2]	3.49	A:MET 312[SD]
6	B:ASN 20[ND2]	2.86	A:THR 361[OG1]
7	B:ASN 20[ND2]	3.57	A:THR 361[O]
8	B:SER 55[OG]	3.64	A:GLY 106[O]
9	B:ARG 59[NH2]	3.45	A:SER 255[OG]
10	B:TYR 78[OH]	3.78	A:PRO 298[O]
11	B:TYR 78[OH]	2.99	A:SER 302[OG]
12	B:SER 98[OG]	2.44	A:ASP 192[OD1]
13	B:LEU 4[O]	2.97	A:VAL 351[N]
14	B:ASP 56[OD2]	3.34	A:TYR 251[OH]
15	B:GLU 65[OE1]	3.74	A:TYR 251[N]
16	B:GLU 77[OE2]	2.78	A:SER 302[OG]
17	B:SER 98[OG]	3.85	A:LYS 195[NZ]
18	B:SER 98[O]	3.22	A:ARG 246[NH1]

Salt bridges

##	AvrRxo1-ORF2	Distance (Å)	AvrRxo1-ORF1
1	B:HIS 13[NE2]	2.73	A:ASP 316[OD1]
2	B:LYS 17[NZ]	3.91	A:ASP 316[OD1]
3	B:LYS 17[NZ]	2.90	A:ASP 316[OD2]
4	B:ARG 19[NE]	3.34	A:GLU 358[OE2]
5	B:ARG 19[NH1]	3.48	A:GLU 358[OE2]
6	B:ARG 19[NH2]	3.47	A:GLU 358[OE2]
7	B:ARG 59[NH1]	3.93	A:ASP 222[OD1]
8	B:SER 98[O]	3.27	A:ARG 246[NH2]
9	B:SER 98[O]	3.22	A:ARG 246[NH1]

Hydrogen bonds

##	AvrRxo1-ORF2	Distance (Å)	AvrRxo1-ORF2
1	B:ARG 19[NH1]	2.90	B:SER 26[OG]
2	B:ARG 19[NH1]	2.90	B:ASP 27[OD1]
3	B:SER 26[OG]	2.90	B:ARG 19[NH1]
4	B:ASP 27[OD1]	2.90	B:ARG 19[NH1]

Salt bridges			
##	AvrRxo1-ORF2	Distance (Å)	AvrRxo1-ORF2
1	B:ARG 19[NH1]	2.90	B:ASP 27[OD1]
2	B:ARG 19[NH2]	3.61	B:ASP 27[OD1]
3	B:ASP 27[OD1]	2.90	B:ARG 19[NH1]
4	B:ASP 27[OD1]	3.61	B:ARG 19[NH2]

Table S2-2| Lists of hydrogen bonds and salt bridges in the interaction interfaces in the ATP-AvrRxo1-ORF1:-ORF2 complex

Hydrogen bonds

##	AvrRxo1-ORF2	Distance (Å)	AvrRxo1-ORF1
1	B:LEU 4[N]	2.95	A:VAL 351[O]
2	B:HIS 13[ND1]	2.60	A:TYR 357[OH]
3	B:HIS 13[NE2]	2.78	A:ASP 316[OD1]
4	B:LYS 17[NZ]	2.73	A:ASP 316[OD2]
5	B:ASN 20[ND2]	3.21	A:THR 361[O]
6	B:ASN 20[ND2]	2.96	A:THR 361[OG1]
7	B:ARG 59[NH1]	3.39	A:SER 255[OG]
8	B:ARG 59[NH2]	3.66	A:TYR 256[OH]
9	B:TYR 78[OH]	3.66	A:PRO 298[O]
10	B:TYR 78[OH]	2.86	A:SER 302[OG]
11	B:ASN 80[ND2]	3.54	A:ASP 370[OD2]
12	B:SER 98[OG]	2.60	A:ASP 192[OD2]
13	B:LEU 4[O]	2.91	A:VAL 351[N]
14	B:ASP 53[OD2]	3.38	A:TRP 110[NE1]
15	B:ASP 56[OD2]	2.38	A:TYR 251[OH]
16	B:GLU 65[OE1]	3.08	A:TYR 251[N]
17	B:GLU 77[OE2]	2.61	A:SER 302[OG]
18	B:SER 98[O]	3.04	A:ARG 246[NH2]

Salt bridges

##	AvrRxo1-ORF2	Distance (Å)	AvrRxo1-ORF1
1	B:HIS 13[NE2]	2.78	A:ASP 316[OD1]
2	B:LYS 17[NZ]	3.68	A:ASP 316[OD1]
3	B:LYS 17[NZ]	2.73	A:ASP 316[OD2]
4	B:ARG 19[NE]	3.31	A:GLU 358[OE2]
5	B:ARG 19[NH1]	3.66	A:GLU 358[OE2]
6	B:ARG 19[NH2]	3.39	A:GLU 358[OE2]
7	B:SER 98[O]	3.43	A:ARG 246[NH1]
8	B:SER 98[O]	3.04	A:ARG 246[NH2]

Hydrogen bonds

##	AvrRxo1-ORF2	Distance (Å)	AvrRxo1-ORF2
1	B:LYS 15[NZ]	3.44	B:GLU 29[OE2]
2	B:ARG 19[NH1]	3.81	B:LEU 23[O]
3	B:ARG 19[NH1]	2.94	B:SER 26[OG]
4	B:ARG 19[NH1]	2.90	B:ASP 27[OD1]
5	B:ARG 19[NH2]	3.21	B:ASP 27[OD1]

6	B:GLU 29[OE2]	3.44	B:LYS 15[NZ]
7	B:SER 26[OG]	2.94	B:ARG 19[NH1]
8	B:ASP 27[OD1]	2.90	B:ARG 19[NH1]
9	B:LEU 23[O]	3.81	B:ARG 19[NH1]
10	B:ASP 27[OD1]	3.21	B:ARG 19[NH2]

Salt bridges

##	AvrRxo1-ORF2	Distance (Å)	AvrRxo1-ORF2
1	B:LYS 15[NZ]	3.44	B:GLU 29[OE2]
2	B:ARG 19[NH1]	2.90	B:ASP 27[OD1]
3	B:ARG 19[NH2]	3.21	B:ASP 27[OD1]
4	B:GLU 29[OE2]	3.44	B:LYS 15[NZ]
5	B:ASP 27[OD1]	2.90	B:ARG 19[NH1]
6	B:ASP 27[OD1]	3.21	B:ARG 19[NH2]

Hydrogen bonds

##	AvrRxo1-ORF1	Distance (Å)	AvrRxo1-ORF1
1	A:SER 366[OG]	2.60	A:GLU 355[OE1]
2	A:VAL 367[N]	2.96	A:GLU 358[OE1]
3	A:VAL 367[N]	3.33	A:GLU 355[OE2]
4	A:THR 368[N]	2.70	A:GLU 355[OE2]
5	A:ASN 386[ND2]	3.61	A:GLU 379[OE1]
6	A:ASN 386[ND2]	3.61	A:ASP 377[OD2]
7	A:GLU 355[OE1]	2.60	A:SER 366[OG]
8	A:GLU 358[OE1]	2.96	A:VAL 367[N]
9	A:GLU 355[OE2]	3.33	A:VAL 367[N]
10	A:GLU 355[OE2]	2.70	A:THR 368[N]
11	A:ASP 377[OD2]	3.61	A:ASN 386[ND2]
12	A:GLU 379[OE1]	3.61	A:ASN 386[ND2]

Table S2-3| HADDOCK score and other parameters of the docking pose used in the comparison study

HADDOCK score	-85.0 ± 1.8
Cluster size	116
RMSD from the overall lowest-energy structure (Å)	1.3 ± 0.1
Van der Waals energy (kcal/mol)	-36.6 ± 1.9
Electrostatic energy (kcal/mol)	-274.5 ± 15.0
Desolvation energy (kcal/mol)	5.5 ± 1.7
Restraints violation energy (kcal/mol)	9.2 ± 9.6
Buried Surface Area (Å2)	1054.3 ± 62.6

Table S2-4| Sequences of primers used in this study

Name	Sequences
AvrRxo1-ORF1_196For,	CACCGGATCCATGCATGAATCACTGCCGCATTCGGTT
AvrRxo1-ORF1_1263Rev,	GAGCTCAATTAGCTCGCTGTGAGCAGCTA
AvrRxo1-ORF1_T167N for,	GAAATACCGCAACGGGTAAAAACCGGATTGCGACAAAA
AvrRxo1-ORF1_T167N rev,	TTTTGTCGCAATCCGGTTTTTACCCGTTGCGGTATTTC
AvrRxo1-ORF1_D193T For,	GGTTGCGTGAATCCAACAGTATTCAAGAGTTCGCT
AvrRxo1-ORF1_D193T Rev,	AGCGAACTCTTGAATACTGTTGGATTCACGCAACC
AvrRxo1- ORF2 BamSacFor,	CACCGGATCCGAGCTCATGAAAACTTTGACAGGCGCA
AvrRxo1- ORF2 BamSacRev,	GGATCCGAGCTCTCATGACCACGAGAAAAGTGTCT

References

1. Chisholm, S.T., et al., *Host-microbe interactions: shaping the evolution of the plant immune response.* Cell, 2006. **124**(4): p. 803-14.
2. Ausubel, F.M., *Are innate immune signaling pathways in plants and animals conserved?* Nat Immunol, 2005. **6**(10): p. 973-9.
3. Jones, J.D.G. and J.L. Dangl, *The plant immune system.* Nature, 2006. **444**(7117): p. 323-329.
4. Abramovitch, R.B., J.C. Anderson, and G.B. Martin, *Bacterial elicitation and evasion of plant innate immunity.* Nat Rev Mol Cell Biol, 2006. **7**(8): p. 601-11.
5. Singer, A.U., et al., *Crystal structures of the type III Effector protein AvrPphF and its chaperone reveal residues required for plant pathogenesis.* Structure, 2004. **12**(9): p. 1669-1681.
6. Parsot, C., C. Hamiaux, and A.L. Page, *The various and varying roles of specific chaperones in type III secretion systems.* Curr Opin Microbiol, 2003. **6**(1): p. 7-14.

7. Boller, T. and S.Y. He, *Innate immunity in plants: An arms race between pattern recognition receptors in plants and effectors in microbial pathogens.* Science, 2009. **324**(5928): p. 742-744.
8. Ellis, J., P. Dodds, and T. Pryor, *Structure, function and evolution of plant disease resistance genes.* Curr Opin Plant Biol, 2000. **3**(4): p. 278-284.
9. Meyers, B.C., S. Kaushik, and R.S. Nandety, *Evolving disease resistance genes.* Curr Opin Plant Biol, 2005. **8**(2): p. 129-134.
10. Bai, J., et al., *Diversity in nucleotide binding site-leucine-rich repeat genes in cereals.* Genome Res, 2002. **12**(12): p. 1871-1884.
11. Meyers, B.C., et al., *Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis.* Plant Cell, 2003. **15**(4): p. 809-834.
12. Chang, J.H., et al., *A high-throughput, near-saturating screen for type III effector genes from Pseudomonas syringae.* Proc Natl Acad Sci, 2005. **102**(7): p. 2549-2554.
13. Thieme, F., et al., *Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium Xanthomonas campestris pv.* vesicatoria revealed by the complete genome sequence. J Bacteriol, 2005. **187**(21): p. 7254 - 7266.
14. Potnis, N., et al., *Comparative genomics reveals diversity among xanthomonads infecting tomato and pepper.* BMC Genomics, 2011. **12**(1): p. 146.
15. Stavrinides, J., W. Ma, and D.S. Guttman, *Terminal reassortment drives the quantum evolution of type III effectors in bacterial pathogens.* PLoS Pathog, 2006. **2**(10): p. e104.
16. Janjusevic, R., et al., *A bacterial inhibitor of host programmed cell death defenses is an E3 ubiquitin ligase.* Science, 2006. **311**(5758): p. 222-226.
17. Lee, C.C., et al., *Crystal structure of the type III effector AvrB from Pseudomonas syringae.* Structure, 2004. **12**(3): p. 487-494.
18. Wang, C.I., et al., *Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity.* Plant Cell, 2007. **19**(9): p. 2898-912.
19. Boutemy, L.S., et al., *Structures of Phytophthora RXLR effector proteins: a conserved but adaptable fold underpins functional diversity.* J Biol Chem. **286**(41): p. 35834-42.
20. Chou, S., et al., *Hyaloperonospora arabidopsidis ATR1 effector is a repeat protein with distributed recognition surfaces.* Proc Natl Acad Sci. **108**(32): p. 13323-8.
21. Wirthmueller, L., A. Maqbool, and M.J. Banfield, *On the front line: structural insights into plant-pathogen interactions.* Nat Rev Micro, 2013. **11**(11): p. 761-776.

22. Singer, A.U., et al., *A pathogen type III effector with a novel E3 ubiquitin ligase architecture.* PLoS Pathog, 2013. **9**(1).
23. Nino-Liu, D.O., P.C. Ronald, and A.J. Bogdanove, *Xanthomonas oryzae pathovars: model pathogens of a model crop.* Mol Plant Pathol, 2006. **7**(5): p. 303-24.
24. Salzberg, S., et al., *Genome sequence and rapid evolution of the rice pathogen Xanthomonas oryzae pv. oryzae PXO99A.* BMC Genomics, 2008. **9**(1): p. 204.
25. Zhao, B., et al., *The Rxo1/Rba1 locus of maize controls resistance reactions to pathogenic and non-host bacteria.* Theor Appl Genet, 2004. **109**(1): p. 71-79.
26. Zhao, B., et al., *The avrRxo1 gene from the rice pathogen Xanthomonas oryzae pv. oryzicola confers a nonhost defense reaction on maize with resistance gene Rxo1.* Mol Plant Microbe Interact, 2004. **17**(7): p. 771-9.
27. Zhao, B., et al., *A maize resistance gene functions against bacterial streak disease in rice.* Proc Natl Acad Sci 2005. **102**(43): p. 15383-15388.
28. Liu, H., et al., *Domain dissection of AvrRxo1 for suppressor, avirulence and cytotoxicity functions.* PLoS One, 2014. **9**(12): p. e113875.
29. Salomon, D., et al., *Expression of Xanthomonas campestris pv. vesicatoria type III effectors in yeast affects cell growth and viability.* Mol Plant Microbe Interact, 2011. **24**(3): p. 305-14.
30. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.
31. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* J Mol Biol, 2007. **372**(3): p. 774-97.
32. Holm, L., et al., *Using Dali for structural comparison of proteins.* Curr Protoc Bioinformatics, 2006. **Chapter 5**: p. Unit 5 5.
33. Meinhart, A., et al., *Crystal structure of the plasmid maintenance system epsilon/zeta: functional mechanism of toxin zeta and inactivation by epsilon 2 zeta 2 complex formation.* Proc Natl Acad Sci U S A, 2003. **100**(4): p. 1661-6.
34. Wang, L.K., C.D. Lima, and S. Shuman, *Structure and mechanism of T4 polynucleotide kinase: an RNA repair enzyme.* EMBO J, 2002. **21**(14): p. 3873-80.
35. Galburt, E.A., et al., *Structure of a tRNA repair enzyme and molecular biology workhorse: T4 polynucleotide kinase.* Structure, 2002. **10**(9): p. 1249-60.
36. Bernstein, N.K., et al., *The molecular architecture of the mammalian DNA repair enzyme, polynucleotide kinase.* Mol Cell, 2005. **17**(5): p. 657-70.

37. Yuen, M.H., et al., *Crystal structure of the H256A mutant of rat testis fructose-6-phosphate,2-kinase/fructose-2,6-bisphosphatase. Fructose 6-phosphate in the active site leads to mechanisms for both mutant and wild type bisphosphatase activities.* J Biol Chem, 1999. **274**(4): p. 2176-84.
38. Khoo, S.K., et al., *Molecular and structural characterization of the PezAT chromosomal toxin-antitoxin system of the human pathogen Streptococcus pneumoniae.* J Biol Chem, 2007. **282**(27): p. 19606-18.
39. Mutschler, H., et al., *A novel mechanism of programmed cell death in bacteria by toxin-antitoxin systems corrupts peptidoglycan synthesis.* PLoS Biol, 2011. **9**(3): p. e1001033.
40. Sherrer, R.L., et al., *C-terminal domain of archaeal O-phosphoseryl-tRNA kinase displays large-scale motion to bind the 7-bp D-stem of archaeal tRNA(Sec).* Nucleic Acids Res, 2011. **39**(3): p. 1034-41.
41. Wang, L.K. and S. Shuman, *Mutational analysis defines the 5'-kinase and 3'-phosphatase active sites of T4 polynucleotide kinase.* Nucleic Acids Res, 2002. **30**(4): p. 1073-80.
42. Wang, L.K. and S. Shuman, *Domain structure and mutational analysis of T4 polynucleotide kinase.* J Biol Chem, 2001. **276**(29): p. 26868-74.
43. Eastberg, J.H., J. Pelletier, and B.L. Stoddard, *Recognition of DNA substrates by T4 bacteriophage polynucleotide kinase.* Nucleic Acids Res, 2004. **32**(2): p. 653-60.
44. de Vries, S.J., M. van Dijk, and A.M. Bonvin, *The HADDOCK web server for data-driven biomolecular docking.* Nat Protoc, 2010. **5**(5): p. 883-97.
45. Triplett, L.R., et al., *Genomic analysis of Xanthomonas oryzae isolates from rice grown in the United States reveals substantial divergence from known X. oryzae pathovars.* Appl Environ Microbiol. **77**(12): p. 3930-7.
46. Kaiser, B.K., et al., *The structure of a bacterial DUF199/WhiA protein: domestication of an invasive endonuclease.* Structure, 2009. **17**(10): p. 1368-76.
47. Deslandes, L. and S. Rivas, *Catch me if you can: bacterial effectors and plant targets.* Trends Plant Sci, 2012. **17**(11): p. 644-55.
48. Block, A., et al., *Phytopathogen type III effector weaponry and their plant targets.* Curr Opin Plant Biol, 2008. **11**(4): p. 396-403.
49. Liu, J., et al., *A receptor-like cytoplasmic kinase phosphorylates the host target RIN4, leading to the activation of a plant innate immune receptor.* Cell Host Microbe, 2011. **9**(2): p. 137-46.
50. Chung, E.H., et al., *Specific threonine phosphorylation of a host target by two unrelated type III effectors activates a host innate immune receptor in plants.* Cell Host Microbe, 2011. **9**(2): p. 125-36.

51. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Methods in Enzymology, 1997. **276**: p. 307-326.
52. Pape, T. and T.R. Schneider, *HKL2MAP: a graphical user interface for phasing with SHELX programs.* J Appl Cryst, 2004. **37**(843-844).
53. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
54. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method.* Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 3): p. 240-55.
55. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr, 2004. **D60**: p. 2126-2132.
56. Vagin, A. and A. Teplyakov, *MOLREP: an automated program for molecular replacement.* J Appl Cryst, 1997. **30**: p. 1022-1025.
57. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Cryst, 2007. **40**: p. 658-674.
58. Perrakis, A., et al., *wARP: improvement and extension of crystallographic phases by weighted averaging of multiple refined dummy atomic models.* Acta Cryst D, 1997. **53**: p. 448-455.
59. Trott, O. and A.J. Olson, *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading.* Journal of Computational Chemistry 2010. **31**: p. 455-461.
60. Kabsch, W., *Crystal physics, diffraction, theoretical and general crystallography.* Acta Cryst. , 1976. **A32**: p. 922-923.
61. DeLano, W.L., *The PyMOL molecular graphics system* in *The PyMOL molecular graphics system*. 2002, Delano Scientifc: San Carlos, CA, USA. .
62. Gille, C. and C. Frommel, *STRAP: editor for structural alignments of proteins.* Bioinformatics, 2001. **17**(4): p. 377-8.
63. Earley, K.W., et al., *Gateway-compatible vectors for plant functional genomics and proteomics.* Plant J, 2006. **45**(4): p. 616-29.
64. Zhao, B., et al., *Computational and biochemical analysis of the Xanthomonas effector AvrBs2 and its role in the modulation of Xanthomonas type three effector delivery.* PLoS Pathog, 2011. **7**(12): p. e1002408.

Chapter III:

Xanthomonas effector AvrRxo1 suppresses plant immunity by regulating the plant stomatal aperture size

Shuchi Wu[1,2], Changhe Zhou[1,2], Yi Liu[1], Qiang Cheng[1], Bin Xu[1], Zhengxing Shen[1], Zhiyong Yang[1], Bingyu Zhao[1*]
[1]Department of Horticulture, Virginia Tech, Blacksburg, VA, 24061
[2]These authors contributed equally to this work.

Abstract

Rice bacterial leaf streak disease caused by *Xanthomonas oryzae* pv. *oryzicola* (*Xoc*) is one of the most important rice diseases. The *Xanthomonas* type III effector gene *avrRxo1* is conserved in diverse *Xoc* strains and its homologues have been identified from several other gram-negative bacteria species. AvrRxo1 makes a significant contribution to bacterial virulence on diverse hosts when expressed in *Xanthomonas* and *Pseudomonas syringae* strains. This virulence function requires a conserved ATP-biding domain. Ectopic expression of *avrRxo1* in *Arabidopsis* suppresses plant basal defense responses. *Arabidopsis VOZ* (**V**ascular **O**ne **Z**inc-finger transcription factor), which has two homologues in the *Arabidopsis* genome, *VOZ1* and *VOZ2*, was identified as one of AvrRxo1 interactors. The double knockout of *VOZ1*, and its homolog *VOZ2*, renders the plants more susceptible to the virulent pathogen *Pseudomonas syringae* pv. *tomato* (*Pst*) DC3000. Additionally, the virulence function of AvrRxo1 is negated in a *voz1/voz2* mutant, genetically validating that VOZ proteins are a major target of AvrRxo1. Expression profiling of transgenic *Arabidopsis* plant expressing *avrRxo1* genes allowed us to identify *Arabidopsis* genes regulated by AvrRxo1 and VOZ1/2. AvrRxo1 interacts with and stabilizes VOZ2 *in vivo* and binds to the promoter region of *AtCYS2* (*Arabidopsis* phytocystatin 2) to induce its expression. Overexpression of *CYS2* in *Arabidopsis* increased stomata aperture size, and enhanced plant susceptibility to *Pst*. Therefore, one of AvrRxo1's virulence functions is to regulate the expression of *CYS2* through the function of VOZ2, resulting in increased stomatal aperture size that facilitates bacterial colonization of the interior of the leaf.

Introduction

Plant-pathogen interactions are comprised of an endless battle between the invading pathogens and defensive host plants. Both plants and microbes are continuously evolving diverse strategies to overcome invasion or defense

mechanisms. The first layer of plant defense is physical barriers such as epidermal waxy cuticles and the fairly rigid cell walls etc., which are effective against most microbes [1]. Some microbial pathogens can invade internal tissue either by directly breaking down plant cell walls or via natural openings such as stomata and wounds [1, 2]. In response to pathogen invasion, plants evolved basal defense responses that are usually initiated by the recognition of highly conserved pathogen-associated molecular patterns (PAMPs) of microbes, which activate PAMP-triggered immunity (PTI) to confine pathogens at the infection sites [3]. Successful microbial pathogens, such as gram-negative bacteria, evolved type III secretion systems (T3SS) to deliver a collection of effectors into plant cells to suppress PTI [3]. Plants evolved a second layer of defense based on host-specific disease resistance (*R*) genes that directly or indirectly recognize pathogen effectors to initiate effector-triggered immunity (ETI) [3]. Thus far, the exact mechanisms and signaling pathways involved in PTI and ETI are not well defined.

Most pathogen effectors have unknown biochemical functions, though a few have been identified to possess eukaryotic enzyme activities that can modify host targeting proteins [5]. Pathogen effectors can target conserved host proteins, or hubs, which play key roles in regulating plant development or stress response [6]. For example, several pathogen effectors target plant transcription factors and may reprogram the plant transcriptome and regulate plant growth to benefit microbial proliferation [6-10]. TAL-effectors (transcription activator-like effectors) can directly bind to plant promoters and activate plant gene expression [11]. Therefore, plant transcription factors are key targets in both pathogen invasion and plant defense pathways.

Arabidopsis vascular one zinc factor (VOZ) is a plant specific NAC-type transcription factor family with two members (VOZ1/VOZ2) that are conserved in diverse vascular plant species [12]. Recently, it has been demonstrated that VOZ2 is a master regulator of plant biotic and abiotic stress, and also has a key role in regulating *Arabidopsis* flowering time [13-15]. However, it has not been shown if pathogen effectors can target and modify this key transcription factor to suppress plant immunity.

Plant stomata control waster transpiration and gas exchange, and therefore, are critical for plant photosynthesis [16]. Plants have evolved a multi-layer regulatory system to control stomatal opening/closure in response to both abiotic and biotic stress [16-18]. Since natural openings on plant surface, such as surface wound sites, stomata or hydathodes, are major entrance for bacterial pathogen invasion [17], some bacterial pathogens have evolved different mechanisms of interfering with stomatal opening/closure to facilitate pathogen invasion [17]. However, the exact regulation mechanisms of stomatal opening are still unclear, especially with regards to pathogen infection.

Rice bacterial leaf streak (BLS) disease caused by the gram-negative bacterium *Xanthomonas oryzae* pv. *oryzicola* (*Xoc*) is one of the most important rice diseases in Asia and Africa [5]. *Xoc* has been selected as one of regulated agents (http://www.selectagents.gov/SelectAgentsandToxinsList.html), which hindered the *Xoc* research in both laboratory and field conditions. We previously identified a maize disease resistance (*R*) gene, *Rxo1,* that provides resistance to *Xoc* strains carrying the T3E gene *avrRxo1* in maize and transgenic rice plants [6, 7]. In the *Xoc* genome, *avrRxo1* is adjacent to a small open reading frame that was originally named as *avrRxo1-ORF2*, while the *avrRxo1* gene was originally designated as *avrRxo1-ORF1* [8]. We recently co-crystallized the AvrRxo1 ORF1:ORF2 complex, and demonstrated that AvrRxo1 ORF1:ORF2 is structurally similar to the zeta-epsilon family of toxin:antitoxin systems. AvrRxo1 ORF1:ORF2 has toxin:antitoxin-like activity in bacteria, and the toxin activity of AvrRxo1-ORF1 is required for its virulence function in rice plants. Here, we use *avrRxo1* to specifically refer to *avrRxo1-ORF1*. Both *avrRxo1* and *avrRxo1-ORF2* are conserved in all Asian *Xoc* strains, which suggests that AvrRxo1 plays an essential role in suppressing plant immunity. However, the virulence mechanism and the host targets of AvrRxo1 haven't been characterized.

In this study, we employed the *Arabidopsis thaliana* and *Pseudomonas syringae* pv. *tomato* (*Pst*) system to study the AvrRxo1 virulence mechanism and identified a conserved plant transcription factor, VOZ2, as one of the AvrRxo1 virulence targets. We demonstrated that wild type AvrRxo1, but not the mutant AvrRxo1-T167N, can interact with and suppress the degradation of VOZ2 *in vivo*. Transcriptomics revealed that AvrRxo1 can induce the expression of a cysteine protease inhibitor, *CYS2*. We demonstrated that VOZ2 specifically binds to the promoter region of *CYS2* to induce its expression. Transgenic Arabidopsis plants overexpressing *CYS2* display increased stomatal aperture and enhanced susceptibility to *Pst* DC3000. Therefore, we conclude that one of AvrRxo1's virulence mechanisms is to regulate stomatal opening, and suppress Arabidopsis stomatal-mediated immunity.

Results

AvrRxo1 Enhances Pseudomonas Bacterial Virulence, and this Function Depends on its Putative ATP Binding Site

To evaluate the virulence function of AvRxo1, both the wild type and ATP-binding mutant *avrRxo1* (T167N) genes, along with the putative AvrRxo1 specific chaperone (ORF2) were cloned into the effector delivery vector pBZ598, which expresses cloned genes from *avrRpm1* promoter and provides a *P. syringae* T3S

signal [11]. Then *avrRxo1* and *avrRxo1* (T167N) genes were expressed in *Pseudomonas syringae pv. tomato* (*Pst*) DC3000-ΔCEL (an attenuated strain deprived of the effector genes *hopM1* and *avrE1* [17]). The isogenic bacterial strains were infiltrated into leaves of *Nicotiana benthamiana* and *Arabidopsis* ecotype Col-0, and the bacterial populations were assessed at 3 and 4 days post inoculation (dpi). As shown in Figure 3-1, *Pst* DC3000-ΔCEL expressing the wild type *avrRxo1* increases in population size about half logCFU/cm^2 units more than the isogenic strain expressing the *avrRxo1-T167N* gene or *Pst* DC3000-ΔCEL with an empty vector when infiltrated in both *N. benthamiana* (Figure 3-1a) and *Arabidopsis* plants (Figure 3-1b).

The *avrRxo1* gene was also expressed in the wild type *Pst* DC3000 strain and spray-inoculated on *Arabidopsis* Col-0 plants. The bacterial populations were assessed at three dpi. The *Pst* DC3000 strain expressing wild type *avrRxo1* proliferates to population size about two log CFU/cm^2 units higher than the strains expressing *avrRxo1*-T167N or the empty vector control (Figure 3-1c). Therefore, we conclude AvrRxo1 has significant virulence functions when expressed either in the attenuated *Pst* strain DC3000-ΔCEL [29] when evaluated by infiltration, or when expressed in highly virulent strain *Pst* DC3000 when evaluated by spray inoculation on two host plants.

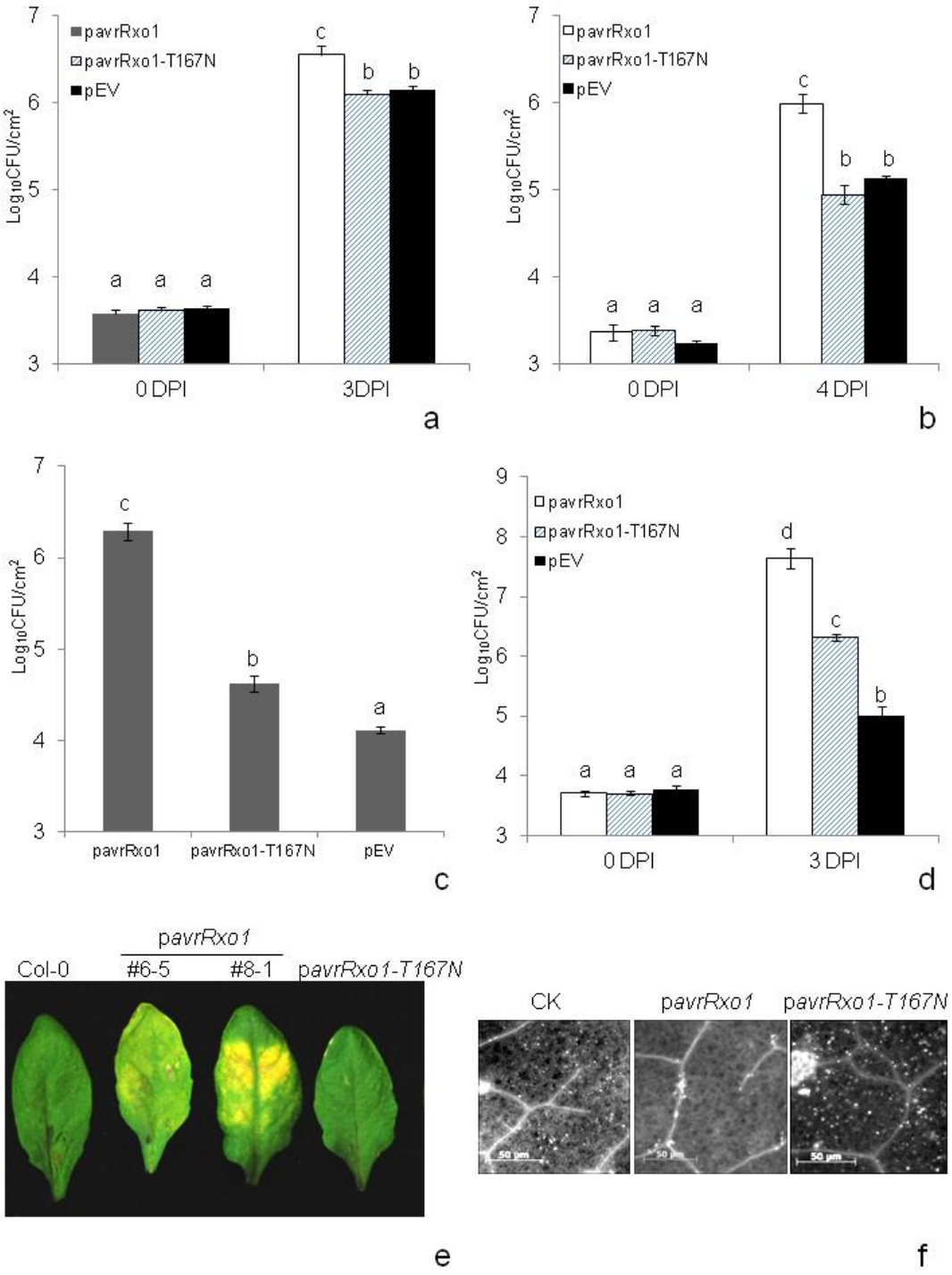


Figure 3-1| AvrRxo1 enhances Pseudomonas bacterial virulence and ectopic expression of AvrRxo1 suppressed plant basal defense response

a. *N. benthamiana* leaves were inoculated by infiltration with *Pst* DC3000-ΔCEL carrying pBZ598-*avrRxo1*, pBZ598-*avrRxo1*-T167N or pBZ598-EV. Bacterial multiplication in the leaves was determined in triplicate at 0 and 3 days

post inoculation. **b.** *Arabidopsis* (Col-0) leaves were inoculated by infiltration with *Pst* DC3000-ΔCEL carrying pBZ598-*avrRxo1*, pBZ598-*avrRxo1*-T167N or pBZ598-EV. Bacterial multiplication in the leaves was determined in triplicate at 0 and 4 days post inoculation. **c.** *Arabidopsis* leaves were inoculated by spraying with *Pst* DC3000-ΔCEL carrying pBZ598-*avrRxo1*, pBZ598-*avrRxo1*-T167N, or pBZ598-EV. Bacterial multiplication in the leaves was determined in triplicate at 3 days post inoculation. **d.** Leaves of *Arabidopsis* plants ectopically expressing pTA7001-*avrRxo1*, pTA7001-*avrRxo1*-T167N or pTA7001-EV were inoculated by infiltration with *Pst* DC3000-ΔCEL after turning on the ectopic genes expression by dexamethasone (DEX) induction for 12 hours. Bacterial multiplication in the leaves was determined in triplicate at 0 and 3 days post inoculation. **e.** Representative leaves demonstrating the chlorosis phenotype caused by ectopic expression of *avrRxo1* in *Arabidopsis*. **f.** Ectopic expression of wild type but not mutant *avrRxo1* in *Arabidopsis* could suppress the callose deposition triggered by *Pst*-Δ*hrc*.

All the experiments were repeated three times with similar results. Bacterial growth was monitored by dilution plating, and the mean data of three replicates presented (means ± s.d.; n=3, $p<0.05$, Student's t-test, different letter indicate significant statistically difference).

Ectopic expression of AvrRxo1 suppresses plant basal defense responses

To further characterize the virulence functions of AvrRxo1, we generated transgenic *Arabidopsis* plants (Col-0) expressing the wild type *avrRxo1*, mutant *avrRxo1*-T167N, and the empty vector by using DEX-inducible expression vector pTA7001 [30]. These transgenic plants were inoculated with the T3S-deficient mutant strain *Pst*DC3000-*hrc*, which does not cause disease on wild type *Arabidopsis* plants [31]. As shown in Figure 3-1d, transgenic plants that express wild type *avrRxo1*, but not the *avrRxo1*-T167N mutant, have compromised plant immunity, thus allowing more bacterial proliferation. The ectopic expression of wild type *avrRxo1* also caused a chlorosis phenotype at the bacterial inoculation site (Figure 3-1e). The callose deposition triggered by *Pst*DC3000-*hrc* was suppressed by the wild type but not the mutant AvrRxo1-T167N (Figure 3-1f). Therefore, we conclude AvrRxo1 can suppress plant immunity in *Arabidopsis* plant cells.

AvrRxo1 interacts with AtVOZ1/2 which is required for its virulence function

To identify the virulence targets of AvrRxo1 in *Arabidopsis* plant cells, we performed yeast two-hybrid (Y2H) screening by using AvrRxo1-T167N as the bait. Screening of two *Arabidopsis* cDNA libraries led us to identify more than fifty putative AvrRxo1 interactors. One of the putative AvrRxo1 interactor genes, *VOZ2*, encodes a vascular one-zinc-finger transcription factor that has been reported to be a key regulator of *Arabidopsis* flowering time and response to abiotic and biotic stress [32-34]; this gene was chosen for further characterization.

VOZ2 belongs to a small gene family with two members, *VOZ1* (At1g28520) and *VOZ2* (At2g42400). Both, VOZ1 and VOZ2, interact with AvrRxo1-T167N in Y2H assays under high stringency conditions (Figure 3-2a). The interaction between VOZ2 and AvrRxo1 was also confirmed by co-immunoprecipitation (co-IP) assay *in vivo* (Figure 3-2b). VOZ is conserved in different plant species including rice. We also confirmed that AvrRxo1 interacts with the rice VOZ2 in a Y2H assay (Figure S3-1 in the Supplementary Information).

Arabidopsis VOZ1 and VOZ2 share 53% amino acid sequence similarity [32]. Since *Arabidopsis* VOZ1 easily becomes degraded *in vivo* [34], we mainly characterized the functions of VOZ2. To characterize the biological functions of *VOZ1/2* and their roles in AvrRxo1 signaling pathways, we identified T-DNA knockout mutants of *voz1* and *voz2* (WISCDSLOX489-492O10 and SALK_115813.40.55), generated a *voz1/2* double homozygous mutant (designated as line 3-9) (Figure S3-2), and developed *VOZ2* overexpression (*OE-VOZ2*) lines. The knockout mutants were infiltrated with *Pst*DC3000-ΔCEL. As shown in Figure 3-2c, the *voz1/2* double mutants are significantly more susceptible to *Pst*DC3000-deltaCEL than the wild type control, and *voz1* and *voz2* single mutants are also significantly more susceptible to the same pathogen, but not as much as the double mutant. The *OE-VOZ2* transgenic lines were spray-inoculated with *Pst*-DC3000, and they are also slightly more susceptible to *Pst*-DC3000 than the wild type controls (Figure 3-2d). While *OE-VOZ2* plants don't have any obvious phenotype, the *voz1/2* double mutant plants developed chlorotic lesions along the vascular tissue (Figure S3-3b).

*Pst*DC3000-ΔCEL strains carrying *AvrRxo1* or an empty vector (EV) grew to a similar population level at 4 dpi on the *voz1/2* mutant plants (Figure 3-2f). Complementing the *voz1/2 Arabidopsis*mutant lines by expressing either *VOZ1* or *VOZ2* completely suppressed the "chlorosis" phenotype and also restored AvrRxo1-enhanced bacterial growth (Figure S3-3a). Therefore, we conclude VOZ1/2 is required for the virulence function of AvrRxo1.

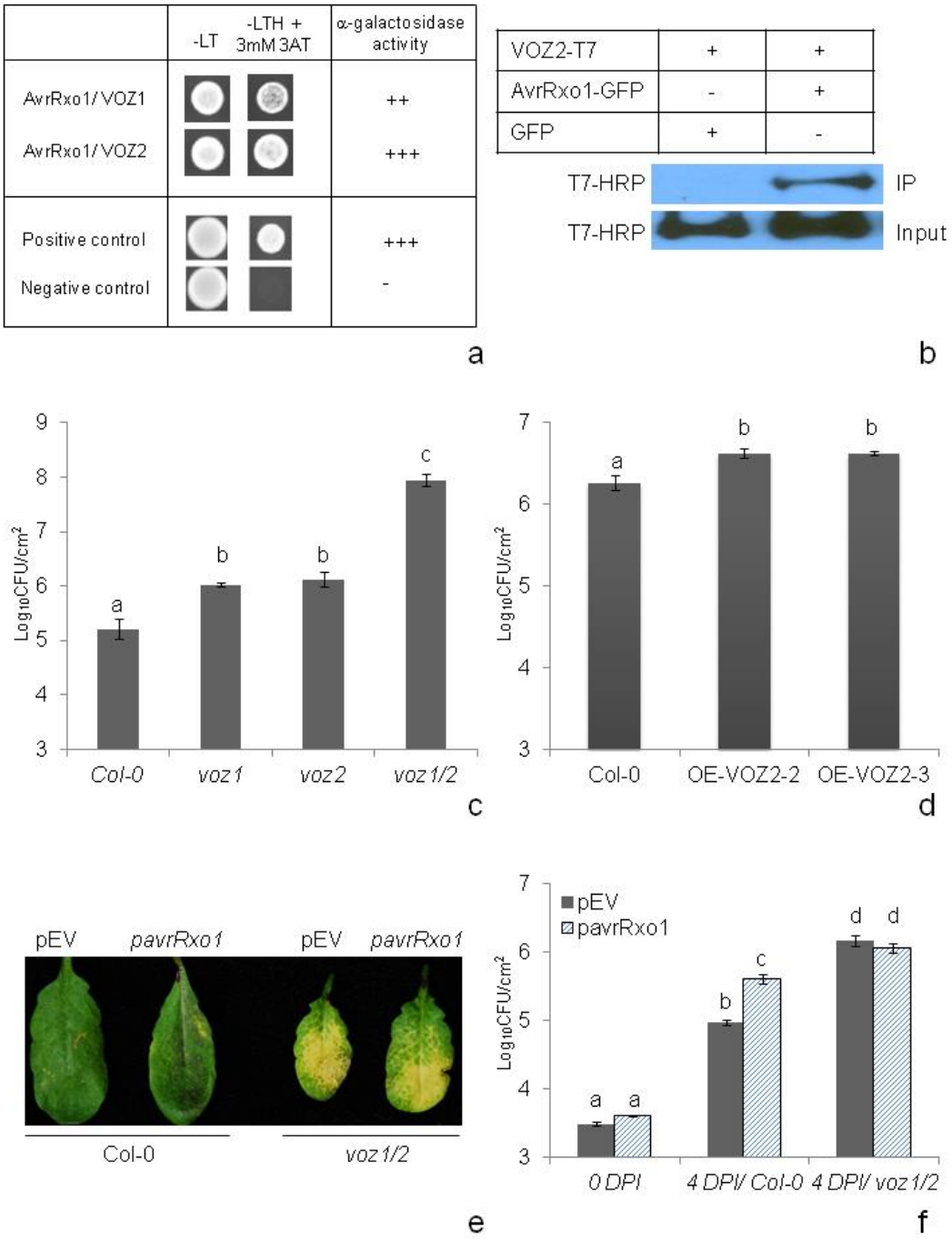


Figure 3-2| AvrRxo1 interacts with AtVOZ1/2 that is required for its virulence function

a. AvrRxo1 interacts with *Arabidopsis* VOZ1 and VOZ2 in yeast two-hybrid (Y2H) assays. The transformed yeast clones carrying pGADT7-*VOZ1* or pGADT7-*VOZ2*, along with pGBKT7-*avrRxo1*-T167N were grown on yeast SD medium lacking leucine, tryptophan, and histidine (-LTH) supplemented with 3

mM 3-Amino-1,2,4-triazole (3-AT). Both the alpha-galactosidase activity and growth on –LTH medium as marked with "+" suggested AvrRxo1-T167N interacted with the *Arabidopsis* VOZ1 and VOZ2. **b.** AvrRxo1-T7 co-immunoprecipitates (co-IP) with the *Arabidopsis* VOZ2 fused with YFP *in vivo*. The co-IP and input protein was detected with anti-T7-HRP or anti-GFP-HRP. **c.** Leaves of *Arabidopsis* T-DNA lines of *voz1*, *voz2* and *voz1/2* were inoculated by infiltration with *Pst* DC3000-ΔCEL. Bacterial multiplication in the leaves was determined in triplicate at 3 days post inoculation. **d.** Leaves of *Arabidopsis* overexpressing *VOZ2* (Lines OE-VOZ2-2 and OE-VOZ2-3) were inoculated by spraying with *Pst* DC3000. Bacterial multiplication in the leaves was determined in triplicate at 3 days post inoculation. **e.** *voz1/2* double knockout line developed chlorosis along the leave vascular system. **f.** Leaves of wild type *Arabidopsis* (Col-0) and *voz1/2* were inoculated by infiltration with *Pst* DC3000-ΔCEL carrying pBZ598-*avrRxo1* or pBZ598-EV. Bacterial multiplication in the leaves was determined in triplicate at 0 and 4 days post inoculation.

All growth curve experiments were repeated three times with similar results. Bacterial growth was monitored by dilution plating, and the mean data of three replicates presented (means ± s.d.; n=3, $p<0.05$, Student's t-test, different letter indicate significant statistically difference).

AvrRxo1 interacts with VOZ2 and suppresses its degradation in vivo

VOZ2 can be degraded through the 26S proteasome pathway [34]. To test if AvrRxo1 interacts with VOZ2 and inhibits its degradation, we infected transgenic *Arabidopsis* plants overexpressing *VOZ2* with *Pst* DC3000-ΔCEL carrying the wild type or mutant *avrRxo1*-T167N, or an empty vector (pBZ598) [11]. The crude protein extract of infected *Arabidopsis* leaves were used for an *in vitro* 26S proteasomal protein degradation assay [35]. As shown in Figure 3-3, VOZ2, isolated from *Arabidopsis* leaves infected by *Pst* DC3000-ΔCEL carrying an empty vector or *avrRxo1*-T167N, was completely degraded after one hour incubation at room temperature. The VOZ2 degradation could be completely inhibited by the 26S proteasome degradation inhibitor MG115 [36]. Interestingly, VOZ2 isolated from *Arabidopsis* leaves infected by *Pst* DC3000-ΔCEL carrying wild type *avrRxo1* is also resistant to the degradation even without the addition of M115. Therefore, we conclude AvrRxo1 interacts with VOZ2 and suppresses its degradation *in vivo*.

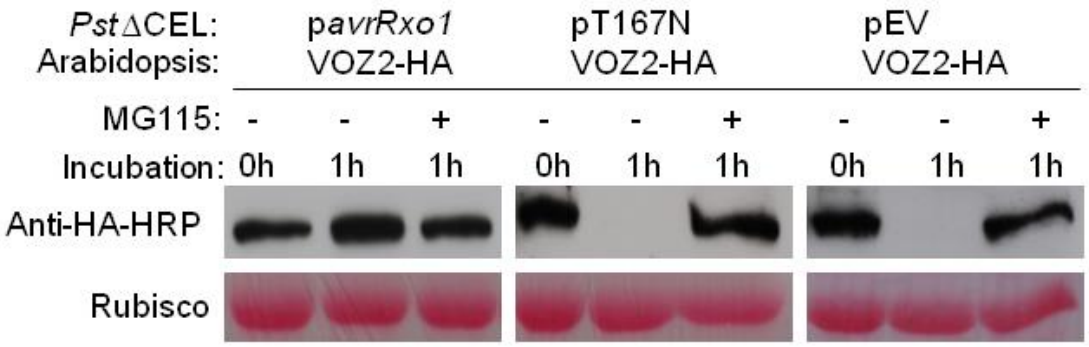


Figure 3-3| AvrRxo1 suppressed VOZ2 degradation *in vivo*

After infected with *Pst* DC3000-ΔCEL carrying pBZ598-*avrRxo1* (T167N) or pBZ598-EV, the crude extract of VOZ2 from the leaves of *Arabidopsis* overexpressing pEarleygate101-*VOZ2*-HA was degraded completely after a one-hour incubation at room temperature. The degradation of VOZ2 could be inhibited by proteasomal degradation inhibitor MG115 (40 mM). After infected with *Pst* DC3000-ΔCEL carrying pBZ598-*avrRxo1*, the crude extract of VOZ2 was not degraded after a one-hour incubation at room temperature.

As an alternative method to confirm that AvrRxo1 inhibits VOZ2 degradation, we generated transgenic *Arabidopsis* plants co-expressing *VOZ2* (tagged with HA) with either the wild type *avrRxo1* gene, or an empty vector (pTA7001) [30]. The crude extract of VOZ2 protein was used for a similar protein degradation assay as described above [35]. As shown in Figure S3-4, VOZ2 isolated from *Arabidopsis* plants expressing an empty vector was completely degraded after one hour incubation at room temperature. Contrastingly, VOZ2 isolated from *Arabidopsis* plants expressing wild type AvrRxo1 is resistant to the degradation even without the addition of M115, similar to what was observed from bacterial delivery of AvrRxo1 described above.

AvrRxo1 significantly induces the expression of a cysteine protease inhibitor, CYS2 in Arabidopsis

To identify *Arabidopsis* genes regulated by AvrRxo1, the transgenic *Arabidopsis* plants expressing wild type *avrRxo1*, or the empty vector pTA7001 were used for gene expression profiling via RNA-seq. The expression of *avrRxo1* gene was turned on by DEX induction for 3 hours, 6 hours, 9 hours and 12 hours. A collection of *Arabidopsis* genes that are specifically induced by the expression of *avrRxo1* were identified (data not shown). The *Arabidopsis CYS2* gene that encodes a cysteine protease inhibitor [37], which was dramatically induced at three hours after the induction of *avrRxo1*, was chosen for further characterization.

To further confirm the induction of *CYS2* by AvrRxo1, we spray inoculated the wild type *Arabidopsis* plants (Col-0) with *Pst*-DC3000ΔCEL expressing the wild type or mutant T167N of AvrRxo1, or an empty vector. The RT-PCR result suggests *CYS2* can be strongly induced by *Pst*-DC3000ΔCEL (*pAvrRxo1*) at 6 h post inoculation. Interesting, *Pst*-DC3000ΔCEL (*pAvrRxo1-T167N*) can also induce the expression of *CYS2* at a lower level (Figure S3-5).

To test if the induction of *CYS2* requires the function of VOZ2, we fused the *CYS2* promoter to a luciferase gene (Pro-*CYS2*-Luc) and transiently co-expressed with the wild type or mutant T167N of AvrRxo1 into protoplast prepared from either the wild type Col-0 plants or the *voz1/2* mutant plants. As shown in Figure 3-4b and 3-4c, the luciferase gene expression was highly induced by the co-expression of AvrRxo1 but not the mutant T167N or GFP control in the wild type *Arabidopsis* protoplast cells. The AvrRxo1-mediated luciferase gene induction was completely abolished when the same set of constructs were expressed in the *voz1/2* protoplast cells. The expression of AvrRxo1 and T167N in *voz1/2* protoplast cells was confirmed by western blot (Figure 3-4d). Therefore, we conclude that AvrRxo1 induces *CYS2* expression through the function of VOZ1/2.

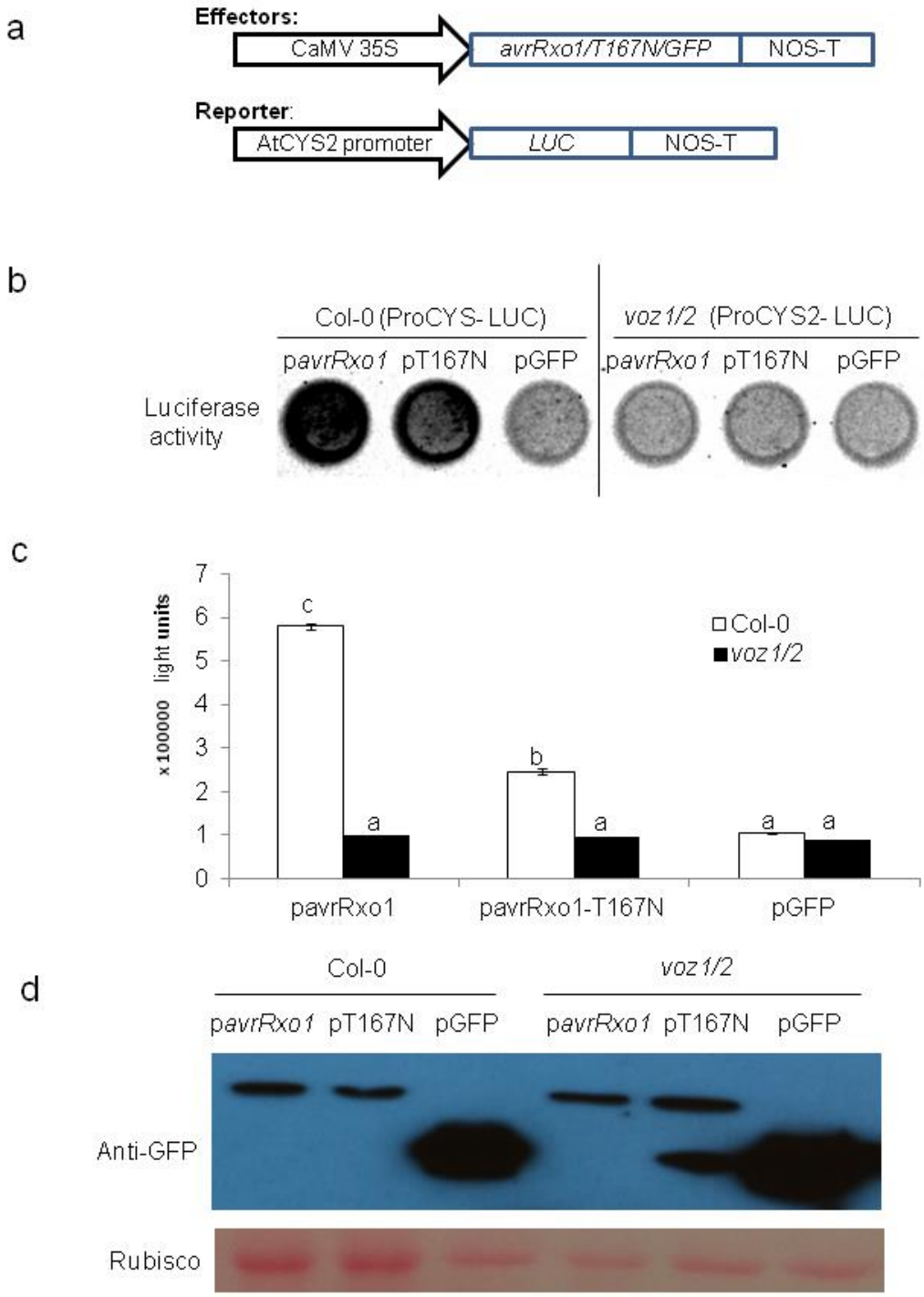


Figure 3-4| Luciferase reporter assay indicated wild type AvrRxo1 significantly induces the expression of *Arabidopsis* cystatin gene *CYS2*

a. The promoter of *CYS2* was fused with a luciferase gene (Pro-*CYS2*-Luc) and transiently co-expressed with *avrRxo1*, *avrRxo1*-T167N or GFP in protoplasts prepared from either wild type Col-0 plants or the *voz1/2* mutant plants. **b.** The

luciferase signal was captured by a CCD camera. **c.** The mean values of luciferase signal were presented (mean ± s.d, n=5, $p<0.05$, Student's t-test, different letter indicate significant statistically difference). **d.** Western blot to assess the expression of AvrRxo1, AvrRxo1-T167N and GFP in Arabidopsis protoplast cells.

VOZ2 specifically binds to the promoter region of CYS2 in an EMSA assay

To test if VOZ2 binds to the promoter region of *CYS2*, we used the recombinant 6xHIS-VOZ2 fusion proteins to incubate with a DNA fragment containing the *CYS2* promoter for an electrophoretic mobility shift assay (EMSA) [24]. As shown in Figure 3-5, the DNA fragment containing *CYS2* promoter region shifted on the DNA electrophoresis gel when it was mixed with 6xHIS-VOZ2 recombinant proteins. The shifted DNA band could be inhibited by supplementing large amounts of unlabeled competitor DNA of *CYS2* promoter. Interestingly, we also confirmed that *CYS2* is up regulated in the OE-VOZ2 transgenic plants. Therefore, we conclude that VOZ2 directly binds to the *CYS2* promoter and regulates its expression.

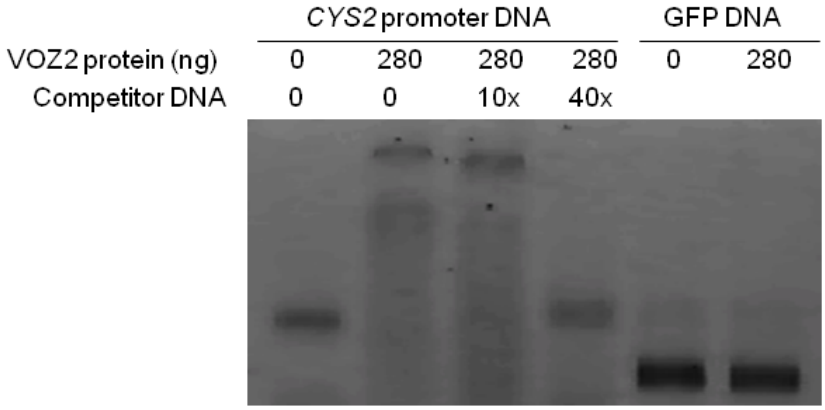


Figure 3-5| Electrophoretic mobility shift assay indicated VOZ2 directly binds to the promoter region of *CYS2*

The DNA fragment contains *CYS2* promoter region was shifted on the DNA electrophoresis gel when it was mixed with 6xHIS-VOZ2 recombinant proteins. The shifted DNA band could be inhibited by supplemental of 40 folds of unlabeled competitor DNA of *CYS2* promoter.

Overexpression of CYS2 increases Arabidopsis stomatal aperture size and enhances plant susceptibility to Pseudomonas syringae

A previous report suggested *CYS2* is predominately expressed in *Arabidopsis* guard cells [38]. To test the biological functions of CYS2, we generated transgenic *Arabidopsis* plants with overexpression of *CYS2* (*OE-CYS2*).

As shown in Figure 3-6a and 3-6b, the stomatal aperture size in *OE-CYS2* transgenic plants is significantly larger than that of wild type controls. Therefore, overexpression of *CYS2* could increase stomatal aperture size.

The *OE-CYS2* transgenic plants were inoculated with *Pst*-DC3000 by infiltration and spray methods. As shown in Figure 3-6c, *OE-CYS2* transgenic plants were significantly more susceptible to *Pst*-DC3000 than the control plants when they were spray-inoculated. However, there was no difference between the *OE-CYS2* plant and wild type controls when they were inoculated by the infiltration method (Figure 3-6d). Therefore, we conclude the overexpression of *CYS2* could increase *Arabidopsis* stomatal aperture size and increase plant susceptibility to *Pst*-DC3000.

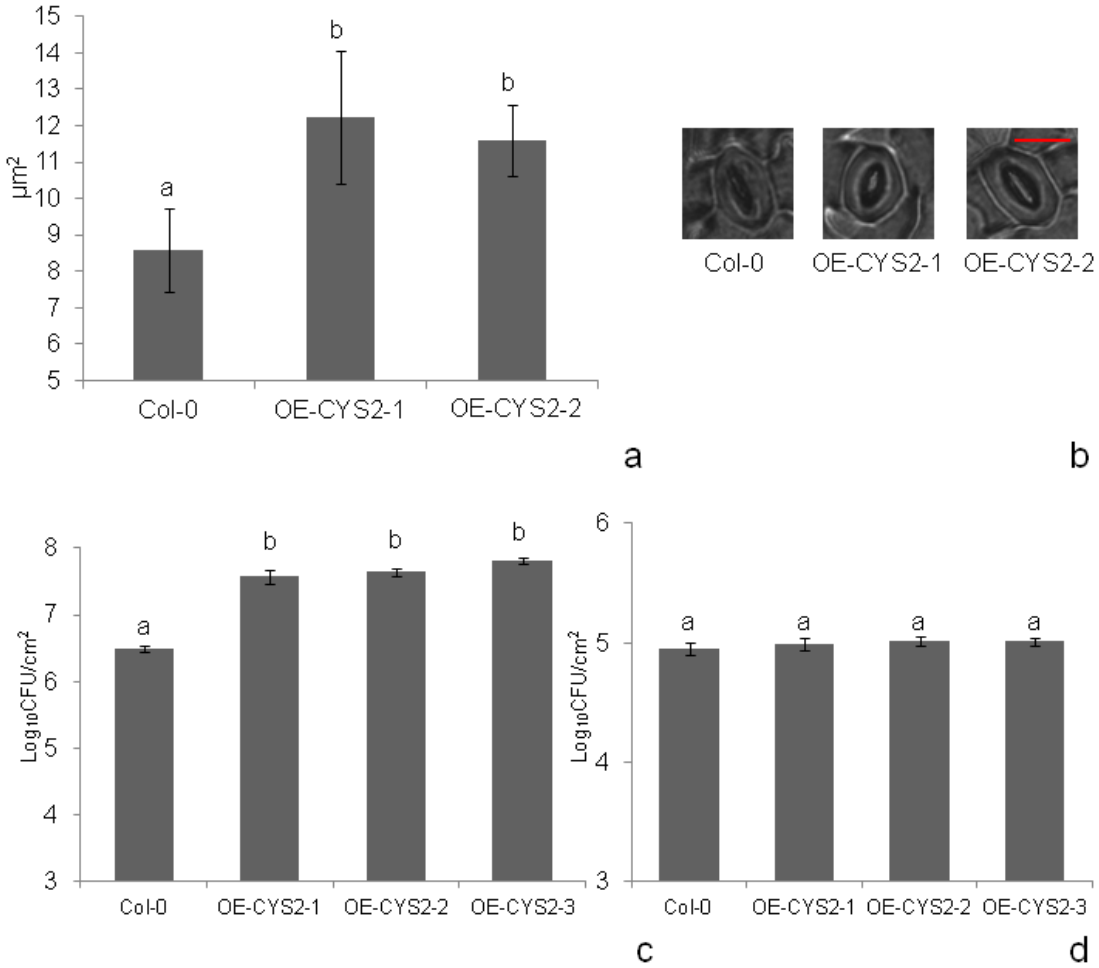


Figure 3-6| Overexpression of CYS2 in *Arabidopsis* increased stomatal aperture size and suppressed plant basal defense against *P. syringae*

a. The *Arabidopsis* stomatal aperture size was measured with 400X magnification of a light microscope, and was estimated by the product of stomatal

opening length and width. The mean value of stomatal aperture size was presented (mean ± s.d, n=10, $p<0.05$, Student's t-test, different letter indicate significant statistically difference). Similar results in two independent lines of *Arabidopsis* overexpressing *CYS2* (OE-CYS2) indicated that overexpression of *CYS2* in *Arabidopsis* promoted stomatal aperture size. **b.** The image of stomatal aperture of Col-0 and OE-CYS2 plants was captured under 400X magnification of a light microscope. The red bar represented 50 μm. **c.** Leaves of Col-0 and OE-CYS2 plants were inoculated by spraying with *Pst* DC3000. Bacterial multiplication in the leaves was determined in triplicate at 3 days post inoculation. Similar results of bacterial growth curve in three independent lines of *Arabidopsis* overexpressing *CYS2* indicated that overexpression of *CYS2* suppressed plant defense against *P. syringae* that was inoculated by spraying to leaf surface. **d.** Leaves of Col-0 and OE-CYS2 plants were inoculated by infiltration with *P. syringae* strain DC3000-ΔCEL. Bacterial multiplication in the leaves was determined in triplicate at 3 days post inoculation. Similar results of bacterial growth curve in three independent lines of *Arabidopsis* overexpressing *CYS2* indicated that overexpression of *CYS2* did not affect plant defense against *P. syringae* that was inoculated by infiltrating to leaf mesophyll.

All the experiments were repeated three times with similar results. Bacterial growth was monitored by dilution plating, and the mean data of three replicates presented (means ± s.d.; n=3, $p<0.05$, Student's t-test, different letter indicate significant statistically difference).

To test if AvrRxo1 expressed by *Pst* can also increase stomata opening, we inoculated wild type *Arabidopsis* plants with the coronatine mutant *Pst*-DC3000-ΔCEL-COR$^-$ [39] carrying p*avrRxo1*, p*avrRxo1*-T167N or an empty vector, and monitored the stomatal aperture size at 1 dpi. As shown in Figure S3-6a, *Arabidopsis* plants inoculated with *Pst*-DC3000-ΔCEL-COR$^-$ (p*avrRxo1*) indeed have significantly larger stomatal aperture sizes.

We also confirmed that the transgenic *Arabidopsis* overexpressing *VOZ2* had increased stomatal aperture sizes (Figure S3-6b). Therefore, as summarized in Figure 8, we conclude that at least one of AvrRxo1's virulence functions is to regulate *Arabidopsis* stomata opening to enhance bacterial infection in the early stage of the disease process.

Discussions

Plant bacterial pathogens usually invade host tissue through the natural openings on the leaf surface, such as stomata or wounds [1]. Plant stomatal guard cells can perceive conserved bacterial molecular patterns and control stomatal closure to restrict pathogen invasion [18, 40]. In response to stomatal closure, some

microbial pathogens evolved mechanisms to subvert stomatal immunity and enhance microbial entry through the stomata [18, 40-42]. Here, we provided evidences showing bacterial pathogens can utilize T3Es to regulate stomatal aperture size. We demonstrated that the *Xanthomonas* T3E AvrRxo1 interacts with and stabilizes the plant transcription factor VOZ2 to specifically induce the expression of *CYS2* (Figure 3-2, 3-3 and 3-4). VOZ2 specifically binds to the promoter of *CYS2* (Figure 3-5). Overexpression of either *VOZ2* or *CYS2* can increase stomatal aperture size and enhance plant susceptibility to *Pseudomonas syringae* pv. *tomato* DC3000 (Figure 3-6).

Regulation of Stomatal Opening by Phytotoxins and Microbial Effectors

It is well established that the phytotoxin coronatine, produced by many *P. syringae* strains, can reopen stomata closed during PTI [18]. Recently, *Xanthomonas campestris* pv *campestris* (*Xcc*) was shown to secrete a small unidentified molecule to enforce plant stomatal re-opening [43]. The fungal toxin fusicoccin, produced by *Fusicoccum amygdali*, has the ability to regulate stomatal aperture size as well [44]. Oxalate produced by the fungal pathogen *Sclerotinia sclerotiorum* can also enhance stomatal opening [45]. Gram-negative bacterial pathogens can deliver up to 40 T3Es and the majority of them have been found to suppress PTI [46]. At least three *P. syringae* effectors (HopF2, HopM1, and HopX1) have been demonstrated to enhance stomatal opening [47-49]. However, thus far, no *Xanthomonas* effector has been reported with the ability to provoke stomatal re-opening. Our results showed that the *Xanthomonas* effector AvrRxo1 enhances stomatal opening, which reinforces that stomatal regulation is a universal virulence mechanism of diverse phytopathogens.

Plant bacterial pathogens can express and deliver more than 40 effectors into host cells. One open question is whether the expression of T3Es is chronologically regulated, and if the effectors are also sequentially delivered inside of host cells as described in *Colletotrichum higginsianum* [50, 51]? Considering the fact that some effectors, including AvrRxo1, have the ability to enforce stomatal opening to help bacteria enter host tissue, we speculate that these effectors may be important in the early stages of pathogenesis. However, some effectors, like HopF2, function to regulate both stomatal aperture size and suppress apoplastic immunity. Thus, they may have virulence functions in both early and late stages of pathogenesis. In this study, we demonstrated that a *Pst* strain expressing AvrRxo1 has higher virulence than the control, even when it was injected into mesophyll tissue to bypass the stomatal defense. Therefore, AvrRxo1 also has dual roles in both early and late stages of bacterial infection. It will be interesting to identify effectors that only

have virulence function in the early stages, where they may only interact with host proteins expressed in guard cells.

VOZ, a Putative Master Regulator of Abiotic and Biotic Stress

In this study, we found that AvrRxo1 specifically interacts with the plant transcription factor VOZ2 (Figure 3-2 and S3-1). Previous studies suggested VOZ2 to be a master regulator in response to both abiotic and biotic stress, and of flowering time [52-54]. VOZ2 belongs to a small gene family with two members that are conserved in diverse plant species [32]. Therefore, it is not surprising that microbial pathogens would try to regulate this important and conserved transcription factor with the aim of reprograming the host transcriptome to benefit microbial infection and proliferation in host tissues. A more recent report suggested *Xanthomonas oryzae* pv. *oryzae* effector XopN may also interact with rice VOZ2 [55]. Therefore, it will be interesting to identify other microbial effectors including fungal and oomycete effectors that also target VOZ2, which will confirm if VOZ2 represents a conserved immune-related hub [56].

Arabidopsis VOZ2 was reported to be mainly expressed in the vascular system [34]. Interestingly, our *voz1/2* mutant *Arabidopsis* plant showed chlorosis phenotype along the vascular system of old leaves, which haven't been reported [54]. It is possible that the growth condition, including light, temperature, humidity and soil nutrient, etc., could contribute to the development of chlorosis phenotype. Another interesting note is that rice leaf streak disease symptoms are also including chlorosis along leaf veins [57-59]. As shown in Figure S3-1, AvrRxo1 interacts with rice VOZ2 *in vitro*, therefore, rice VOZ could be also regulated by AvrRxo1 when it is expressed by *Xoc*, using the same mechanism as described in this study, which we will study further.

A previous study suggested that Voz2 can be degraded by the 26S proteasome degradation pathway [54]. In this study, we also confirmed that VOZ2 is unstable *in vivo* (Fig 3-3 and S3-4). Interestingly, wild type AvrRxo1, but not T167N interacts presumably modifies VOZ2, which prevents its degradation *in vivo*. We previously reported the crystal structure of AvrRxo1-ORF1:ORF2, and demonstrated that AvrRxo1 possess a conserved kinase domain that required for its virulence and toxicity (Han et al, 2015). However, the kinase substrate of AvRxo1 is still waiting to be identified. Therefore, we speculate AvrRxo1 may function as a kinase that can phosphorylate VOZ2 *in vivo*, whereby phosphorylation prevents the degradation caused by an unknown E3 ligase. Therefore, further investigation of the AvrRxo1-mediated modification on VOZ2 and identification of the E3 ligase responsible for VOZ2 degradation is warranted.

Controlling Stomata by Cysteine Protease Inhibitors

We identified an *Arabidopsis* cystatin gene, *AtCYS2*, which encodes a cysteine protease inhibitor that can be specifically induced by AvRxo1 within 6-hour post inoculation (Figure S3-5). Previous report suggests *CYS2* is predominately expressed in stomatal guard cells, and can be induced by wounding [38]. In this study, we demonstrated that overexpression of *CYS2* increases stomatal aperture size and suppress plant immunity (Figure 3-6). Although cystatins have been reported to have roles in regulating plant defense, especially programmed cell death [60], our data is the first time linking *Arabidopsis* CYS2 to plant stomatal immunity. Recent studies suggest a maize cystatin gene, *CC9* [61], can be induced by the virulent fungal pathogen *Ustilago maydis*, while silencing *CC9* compromised pathogen virulence [62]. More importantly, maize CC9 suppresses plant immunity by inhibiting a group of cysteine proteases that normally can be induced by SA [61]. However, it is unknown if the induction of *CC9* can also increase maize stomatal aperture size and suppress immunity to bacterial pathogens, which is worthy to be investigated in the future. It is known that both fungal and oomycete pathogens can secrete cystatin-like proteins to suppress enzyme activities of host-expressed cysteine proteases [63-66]. For example, the maize fungal pathogen *Ustilago maydis* effector Pit2 encodes a cystatin-like protein that can inhibit the enzyme activity of several maize cysteine proteases and suppress host immunity [64]. Therefore, cystatin-related immunity may be important for both fungal and oomycete and bacterial pathogens. Considering most bacterial genomes are much smaller than those of fungi or oomycetes, and thus far, no cystatin-like proteins have been identified in any bacterial pathogens. Our result demonstrated that bacterial pathogens, instead of directly secreting cystatin-like effectors into host cells, may utilize their T3Es to regulate host cystatin gene expression, where the mis-regulation of host cystatin genes may suppress immunity.

Since the major biochemical function of cystatin is to suppress the enzyme activity to plant cysteine protease [67], we speculate that CYS2 also specifically suppresses immune-related cysteine proteases, where those cysteine proteases may have key roles in regulating plant stomatal immunity.

Concluding Remarks

Our study identified the first *Xanthomonas* T3E, AvrRxo1 that regulates stomatal immunity. AvrRxo1 targets a conserved plant transcription factor VOZ to regulate the expression of *Arabidopsis* cystatin gene *AtCYS2* and increases stomatal aperture size (Figure 3-7). We show for the first time that a cysteine protease inhibitor, and possibility a cysteine protease, may be involved in

regulation of stomatal immunity. The discovery further confirmed that plant stomata are a key battle field of host-bacterium interaction, where both plants and pathogens have evolved diverse strategies to control the closure/opening of stomata. The link of a cysteine protease inhibitor to stomata opening represents a significant advance in our understanding of the regulation of plant stomatal regulation.

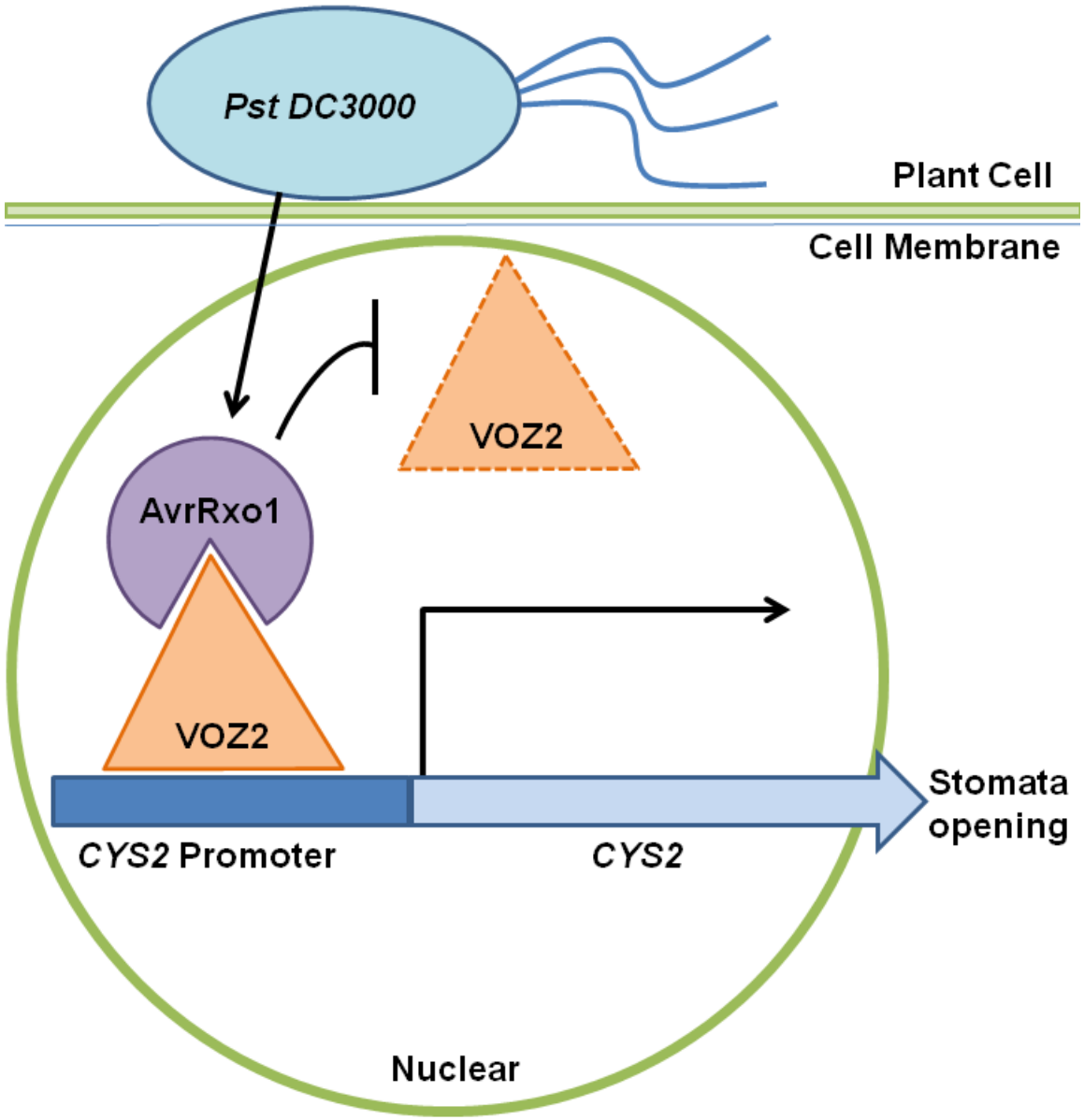


Figure 3-7| Model of the virulence mechanism of AvrRxo1

Bacterial pathogen delivery AvrRxo1 into plant cell, which interacts with and stabilizes plant transcription factor VOZ2, and induced the expression of *CYS2*. The induction of *CYS2* increases plant stomatal aperture size and suppresses plant stomatal immunity.

Material and methods

Bacterial strains and growth

Escherichia coli (*E. coli*) DH5α and Agrobacterium tumefaciens (*A. tumefaciens*) GV2260 [9] strains were grown on Luria agar medium at 37 and 28 ºC, respectively. *Pst* DC3000, *Pst* DC3000-ΔCEL, *Pst* DC3000-ΔHrc, and *Pst* DC3000-*cor*- strains were grown on nutrient glycerol agar (NYGA) [10] at 28 ºC. *E. coli* antibiotic selections were 50 µg/ml kanamycin, 100 µg/ml carbenicillin, 100 µg/ml Spectinomycin, Chloramphenicol 34 µg/ml, and 50 µg/ml gentamycin. *Pst* and *A. tumefaciens* antibiotic selection were rifampicin, and/or 50 µg/ml kanamycin.

Gene cloning, site-directed mutagenesis, and plasmid construction

A 1.5kb DNA fragment contains *avrRxo1*-ORF1 and ORF2 was amplified from the genomic DNA of *Xoc* strain BLS256 using primers "AvrRxo1-196For" and "AvrRxo1-ORF2 Rev" (Table S3-1 in Supplementary Information), and cloned into the pENTR-D-TOPO vector (Invitrogen, Carlsbad, CA), The sequences of all primer used this study are listed in Table S1. The pENTR-*avrRxo1* construct was used for generating mutant *AvrRxo1*-T167N (ORF1-ORF2) by site-directed mutagenesis using primers "AvrRxo1-T167NFor" and "AvrRxo1-T167NRev". pENTR-*avrRxo1*-(ORF1-ORF2) and pENTR-*avrRxo1*-T167N-(ORF1-ORF2) were subcloned into a effector delivery vector pBZ598 vector [11] by LR Gateway cloning (Invitrogen) to generate pBZ598-*avrRxo1* and pBZ598-*avrRxo1*-T167N. An empty vector control of pBZ598 (pEV) was generated by removing the ccdB cassette. These three constructs were transformed into bacterial strains *Pst* DC3000, *Pst* DC3000-ΔCEL, *Pst* DC3000-Δ*Hrc*, and *Pst* DC3000-*cor*-, and used for bacteria growth curve on *Arabidopsis*.

Both wild type *avrRxo1*, and mutant *avrRxo1*-T167N (196-1263) without the type III secretion signal of AvrRxo1 and AvrRxo1-ORF2, were also amplified from pENTR-AvrRxo1-T167N-(ORF1-ORF2) using primers "AvrRxo1_196For" and "AvrRxo1 _1263Rev" and cloned into pENTR-D-TOPO (Invitrogen) to generate pENTR-*avrRxo1*-(196-1263) and pENTR-*avrRxo1*-T167N-(196-1263), respectively. A truncated AvrRxo1 gene was also amplified from pENTR-avrRxo1-T167N-(ORF1-ORF2) using primers "AvrRxo1_196For" and "AvrRxo1 _1134Rev" and generated the pENTR-*avrRxo1*-(196-1134). To generate constructs for transient expression of *avrRxo1* genes in *Nicotiana benthamiana* (*N. benthamiana*), pENTR-*avrRxo1*-(196-1263), pENTR-*avrRxo1*-T167N-(196-1263), and pENTR-*avrRxo1*-(196-1134) was LR cloned into pEarleygate101 vector [12],

pEarleygate102-Des-T7 (modified from pEarleygate 102 [12] by replacing the CFP-HA DNA fragment with a T7 epitope tag sequence). The derived constructs were named as pEarleygate101-*avrRxo1*-(196-1263), pEarleygate101-*avrRxo1*-(196-1263)-T167N, and pEarleygate102- *avrRxo1*-(196-1134)-T7.

The opening reading frame (ORF) *VOZ2* was amplified from the cDNAs of wild type *Arabidopsis* accession Columbia (Col-0) using primers of "AtVOZ2For" and "AtVOZ2Rev" and cloned to generate pENTR-*VOZ2*. The putative promoter region of *VOZ1* (about 3.3kb) and *VOZ2* (about 1.2kb) were PCR amplified from the genomic DNA of Col-0 using primers ProAtVOZ1For, ProAtVOZ1Rev, ProAtVOZ2For and ProAtVOZ2Rev, and subcloned into pDONR P4-P1R (Invitrogen) to generate pDONR-VOZ1 promoter and pDONR-VOZ2 promoter.

The ORF of *CYS2* was amplified from cDNAs of Col-0 using primers of "AtCYS2For" and "AtCYS2Rev", and cloned to generate pENTR-*CYS2*. The putative promoter region (1,580 bp DNA fragment upstream of the start codon) of *AtCYS2* was PCR amplified from genomic DNA of Col-0 using primers "AtCYS2_promoterFor" and "AtCYS2_promoterRev" as a *Bam*HI-*Nco*I fragment, and cloned to generate pENTR-Pro*CYS2*.

A GFP gene was amplified from pHBT35S-Egfp-NosT [7] using primers “GFP_Xho For” and “GFP_Xho rev” and cloned to generate pENTR-*CYS2*.

RNA isolation and RT-PCR

For RT-PCR, total RNAs were isolated from *Arabidopsis* leave tissue using TRIzol Reagent (Invitrogen, Carlsbad, CA) according to the manufacturer’s instructions. DNA contamination was eliminated by treating total RNA with UltraPure DNase I (Invitrogen). The integrity and quantity of total RNA were checked by running through a 0.8% agarose gel and through a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE). cDNA synthesis was performed using the SuperScript III First-Strand System for RT-PCR Kit (Invitrogen) with an oligo-dT primer. The cDNAs were amplified by PCR using the Iproof DNA polymerase (Bio-Rad, Hercules, CA).

Arabidopsis T-DNA knockout and transgenic lines

In this study, all transgenic lines or T-DNA knockout lines are in *Arabidopsis* accession Columbia (Col-0) background only less it was specified. All transgenic *Arabidopsis* plants were developed using flower-dipping method [13]. To generate transgenic *Arabidopsis* plant with condition expression of *avrRxo1* genes, pENTR-*avrRxo1*-(196-1263) and pENTR-*avrRxo1*-T167N-(196-1263) were LR cloned into a Gateway-compatible, DEX (dexamethasone) inducible

expression vector pTA7001-Des [14], to generate pTA7001-*avrRxo1*-(196-1263) and pTA7001- *avrRxo1*-(196-1263)-T167N, respectively. The two *avrRxo1* gene expression constructs, along with the pTA7001 empty vector (EV) were transformed in to *Arabidopsis* (Col-0).

Arabidopsis T-DNA lines *voz1* (WISCDSLOX489-492O10) and *voz2* (SALK_115813.40.55) were purchased from *Arabidopsis* Biological Resource Center (ABRC) (Columbus, OH). The two T-DNA lines were crossed and selected for the *Arabidopsis voz1* and *voz2* (*voz1/2*) double homozygous knock-out T-DNA lines that were re-designated as line 3-9. RT-PCR was used to confirm the compromising of expression of both *voz1* and *voz2* genes in T-DNA lines. The primers used for RT-PCR were: "AtVOZ1For" and "AtVOZ1Rev" for *VOZ1*, "AtVOZ2For" and "AtVOZ2Rev" for VOZ2, respectively. To generate transgenic *Arabidopsis* plants with overexpression of VOZ2, the pENTR-*VOZ2* was LR cloned into a gateway compatible vector pEarleygate101 [12] to generate pEarleygate 101-*VOZ2*, which has been transformed into either wild type Col-0 or the transgenic plants expressing *avrRxo1* genes or controls as describe previously.

To develop constructs for complementary of the *voz1/2* mutant, the pDONR-VOZ1 promoter and pEntr-Native-VOZ1, pDONR-VOZ2 promoter and pEntr-Native-VOZ1 were LR clone into the gateway compatible vector pK7m24GW,3[15], by using the LR Clonase II plus mixture (Invitrogen) following the manufacture's instruction. The derived construct were named as pNative-*VOZ1* and pNative-*VOZ2*, which was transformed into line 3-9 (*voz1/2*).

To develop transgenic *Arabidopsis* plant with overexpression of *CYS2*, the pENTR-*AtCYS2* plasmid was LR cloned into pEarleygate101 [12] to generate pEarleygate101-*AtCYS2*, and transformed into wild type Col-0.

Protein-protein interaction in yeast

The yeast two-hybrid (Y2H) assays was performed according to the user's manual of the Matchmaker Gold yeast 2-hybrid system (Clontech, Palo Alto, CA). The prey and bait vectors pGBKT7 and pGADT7 were modified as Gateway compatible vectors (pGBKT7-Des and pGADT7-Des) by inserting the Gateway recombination cassette (Invitrogen) at the NdeI site. Different genes were then subcloned from TopoEntrD vectors into either pGBKT7-Des or pGADT7-Des vectors.

The yeast strain Y2HGold (Clontech) transformed with wild type pGBKT7-avrRxo1-(196-1263) doesn't grow well, but the yeast strains expressing mutant avrRxo1-T167N (196-1263) grew well. Therefore, we used the mutant pGADT7-avrRxo1-T167N (196-01263) for Y2H assay. Three *Arabidopsis* cDNA libraries obtained from ARBC (CD4-10, CD4-22, and CD4-30) were used for Y2H

screening using AvrRxo1-T167N as bait. The indicated prey and bait constructs were co-transformed into Y2HGold Yeast Strain (Clontech), and selected on yeast SD medium that is lacking leucine, tryptophan, and Histidine (-LTH) supplemented with 3mM 3-Amino-1,2,4-triazole (3-AT). Yeast clones that were grown on selection medium were selected for plasmid DNAs isolation, and sequenced for further characterization according to the user's manual of the Matchmaker Gold yeast 2-hybrid system (Clontech).

To validate the interaction between VOZ2 and AvrRxo1, the pENTR-VOZ1 and pENTR-VOZ2 were LR cloned into the bait vector pADT7-Des to generate pGADT7-VOZ1 and pGADT7-VOZ2, and co-transformed with pGBKT7-avrRxo1-T167N (196-01263). The transformed yeast cells were selected on –LTH medium supplemented with 3mM 3-AT, and the alpaha-galactosidase activity was measured according to the user's manual of the Matchmaker Gold yeast 2-hybrid system (Clontech).

Agrobacterium-mediated transient assay, western blot, and co-immunoprecipation

Agrobacterium strains carrying different constructs were adjusted to A600=0.4, and infiltrate to the mesophyll tissue using blunt end syringes without needles. Lead disks (1 cm^2) were collected at 3 dpi and grinded in 100µl 1X Laemmli SDS-PAGE buffer. 25 µl protein extract samples were loaded on 8-10% SDS-PAGE gel. The proteins samples were blotted to PVDF membrane and hybridized with appropriate antibodies [anti-GFP-HRP (Miltenyi biotec, 120-002-105, 1:5,000), anti-HA-HRP (sigma, H6533, 1:500), anti-T7-HRP (Novagen, 69048-3, 1:5,000)]. The western blot signal was detected by using the ECL kit (Thermo Scientific, USA). After the western blot detection, the PVDF membrane with stained with 0.5 % Ponceau S solution to detect the Rubisco protein as the equal loading control.

The Agroinfiltration, plant protein isolation and co-immunopericipation (co-IP) was performed as described by Krasileva KV et al (2010)[16]. The YFP or GFP tagged fusion protein was immunoprecipitated with anti-GFP (Abcam, ab290) and the co-IP samples were detected with anti-T7-HRP. Before the immunoprecipitation, 25µl of samples were saved as the input control. The immunoprecipitated proteins and input controls were loaded on 8% SDS-PGAGE gel, blotted to nitrocellulose membrane, probed with either anti-GFP-HRP or anti-T7-HRP.

Bacteria growth curve assay

Bacterial proliferation on inoculated *Arabidopsis* plants was measured by standard growth curve assay. To inoculation methods (infiltration and spray) were used in this study [17, 18]. In brief, for the infiltration method, The *Pst* bacteria was cultivated on NYGA media supplemented with appropriate antibiotics at 28 ˚C for two days, the bacterial cells were resuspended in 10 mM $MgCl_2$ and diluted to $5X10^5$ cfu/ml. The bacterial inoculum was infiltrated into the back side of plant leaf using a blunt ended syringe. The inoculated plants were kept at 100% moisture for 24 h (10 h light / 22 ºC, 14 h dark/20 ºC), and then cultured at 50% moisture for two-four days (10 h light / 22 ºC, 14 h dark/20 ºC). *Arabidopsis* leaf discs (588 mm^2) were randomly sampled from inoculated leaves at 0 day and the third or fourth day. For the spray inoculation, the *Pst* bacterial strains were cultivated on NYGA media plate at 28˚C for two days, and the bacterial cells were collected and resuspended in 10mM $MgCl_2$ with 0.02% Silwet L77 and diluted to OD_{600} =0.2. The *Arabidopsis* plants were sprayed with ddH_2O, covered and kept in in dark one day before the inoculation. The bacterial inoculum was sprayed on the leaf of *Arabidopsis* plants using a spray bottle. The inoculated plants were covered and sealed to be kept in 100% moisture for two days, and then cultivated in a growth chamber that was set to 50% moisture for one more day (10 h light / 22 ºC, 14 h dark/20 ºC). Leaf discs (588 mm^2) were randomly sampled from inoculated leaves at the third day for growth curve assay [17, 18].

The sampled leaf discs were grounded in 588 µL of 10mM $MgCl_2$, and vortexed for one minute before the dilution and plating on NYGA media supplemented with proper antibiotics. The plates were cultivated at 28˚C until the bacteria colonies can be counted. The bacteria colony numbers were used to calculate the bacteria proliferation ratio (Log10 CFU/cm^2) [17, 18]. All growth curves have three biological repeats with replicates.

Luciferase reporter assay

The luciferase reporter assay using the *Arabidopsis* protoplast was performed as described previously [19]. A *Bam*HI-*Nco*I DNA fragment contains the *AtCYS2* promoter was subcloned from pENTR-Pro*CYS2* into the *Bam*HI and *Nco*I sites of pUC19-ProFRK1-LUC [20] to generate pCYS2pro-LUC. The pENTR-*avrRxo1*-(196-1263), pENTR-*avrRxo1*-T167N (196-1263), pENTR-GFP, were subcloned into p2GWF7 [21] by LR cloning (Invitrogen). In the p2GWR7.0 constructs, the *avrRxo1* genes were fused with GFP gene and driven by the CaMV35S promoter. The p2GWF7-GFP construct was used as negative control.

Wild type Col-0 and *voz1/2* T-DNA line were used for protoplast isolation. The pEarleygate-mediated protoplast transformation were performed as described previously [22]. The *Arabidopsis* protoplasts were diluted to a density of 2 x 10^5

cells/ml. Twenty microgram plasmid DNAs (the ratio of reporter and effector plasmid DNAs is 1:1) were transformed to 600 µL protoplasts and incubate overnight at 30 ºC. 1mM luciferase substrate, Firefly D-luciferin (Sigma-Aldrich, St. Louis, MO) was added to the protoplast and chemical fluorescence signals were detected using the CCD camera of Gel Doc™ XR+ System (Bio-Rad), and quantified with a luminometer (Promega, Madison, WI). After the LUC assay, the protoplast was harvested for Western blot assay. The expressed proteins were detected by using anti-GFP-HRP.

In vitro proteasomal degradation assay

Stable transgenic *Arabidopsis* plants co-expressing pTA7001-*avrRxo1* orpTA7001-empty vector with pEarleygate101-*VOZ2* was used for VOZ2 cell-free degradation assays were carried out following the protocol described previously [23]. In brief, four weeks old *Arabidopsis* plants were sprayed with 1 µM Dex to induce the expression of *avrRxo1*. Twenty-four hours post induction, 0.2 g leaf tissue for each treatment group was collected and grinded in 400 µL reaction buffer (25 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 10 mM NaCl, 10 mM ATP, and with or without 5 mM DTT) with protease inhibitors (Protease Inhibitor Cocktail Tablet-EDTA free, Roche Applied Science, Indianapolis, IN). The leaf extract was centrifuged (14,000g) for 10 min at 4°C to obtain crude protein extract. The crude protein extract was divided equally into two sets and one set was added 40 mM Proteasome Inhibitor MG-115 (Sigma). All samples were then incubated at room temperature for 0, 0.5, and 1 hrs, before adding 3X Laemmli buffer and boiling for five minutes. The protein samples were loaded on 8% SDS-PAGE gel, blotted to PVDF membrane and hybridized with anti-HA-HRP. The western blot signal was detected by using an ECL kit (Thermo Scientific).

As an alternative method, the *Pst* DC3000 ΔCEL strains (OD_{600}=0.1) carrying either pBZ598-*avrRxo1*-(ORF1-ORF2), pBZ598-*avrRxo1*-T167N-(ORF1-ORF2), or pBZ598-EV were infiltrated with a blunt end syringe into the leaf mesophyll tissue of four weeks-old transgenic *Arabidopsis* plants expressing pEarleygate101-*VOZ2*. Leaf samples were collected at twenty-four hours post infiltration, and processed with the same protocol as described above.

Recombinant protein expression and purification from E. coli

The pENTR-*VOZ2* was subcloned into a modified gateway-compatible vector pET28a-Des, where a *ccdB* cassette was inserted into the *Nde*I site of pET28a+ (GE Healthcare Bio-Sciences, Pittsburgh, PA) through LR cloning (Invitrogen). The resulting plasmid was transformed into *E. coli* C41 cells

(Lucigen, Middleton, WI) and grown overnight in 50 ml of LB medium containing 50 mg/L Kanamycin at 37°C. The culture was transferred into 1 liter of LB medium containing 50 mg/L Kanamycin to reach an A600 of 0.8 units. Isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added to a final concentration 1 mM with subsequent incubation at 250 rpm at 28°C for 8 h. The bacterial cells were harvested and the cell pellets were broken by incubating with 1 µg/ml lysozyme on ice for 1 h, followed by sonication on ice. The lysate was centrifuged at 12,000 g for 20 min at 4^{o}C and the supernatant was gently collected and used for subsequent purification. Nickel affinity resin (GenScript, Piscataway, NJ) were used for to purify the 6xHis-VOZ2 recombinant protein. The purity of the purified proteins was evaluated both by 10% SDS–PAGE. Protein concentration was determined by a protein assay kit from Bio-Rad (Hercules, CA) using bovine serum albumin as a standard.

Electrophoretic mobility shift assay (EMSA)

The EMSA was performed as described previously [24, 25]. The *AtCYS2* promoter DNA (1580 bp) was PCR amplified from pENTR-Pro*CYS2* using FAM labeled "M13For" and "M13Rev" primers [26]. The unlabeled competitor DNA of *CYS2* promoter was amplified with regular M13 For/Rev primers. A DNA fragment of the *GFP* gene was used as negative control, which was PCR amplified from pENTR-*GFP* using the FAM labeled M13For/Rev primers. In brief, 500ng FAM labeled DNA fragment of the *CYS2* promoter (or *GFP*) was mixed with 280 ng VOZ2 protein, along with different amount of unlabeled competitor DNA fragment in a reaction buffer (10 mM Tris, 100 mM KCl, 1 mM EDTA, 0.1 mM DTT, 5 %v/v glycerol, 0.01 mg/mL BSA, pH 7.5) at room temperature for 20 min. The reactions were separated on 1% agarose gel in TAE buffer. The fluorescence signal was captured by a gel scanner, Typhoon FLA7000 equipped with a 635nm laser filter (GE Healthcare, Piscataway, NJ).

Arabidopsis stomatal aperture size measurement

The *Arabidopsis* stomatal aperture sizes were measured following a protocol as described previously [27, 28]. In brief, a thin layer of clear nail polish was applied on the lower epidermis of *Arabidopsis* leaf. After air dry, one piece of clear tape was peeled away from the leaf, which had the impression of the leaf surface. The tape was viewed and pictured with 400 X magnification of a light microscope (Zeiss Axio Observer.A1, Carl Zeiss MicroImaging, Inc., Thornwood, NY). The stomatal aperture size was estimated by the measured length times the width of the stomata aperture. At least 100 guard cells were measured in each sample.

Expression profiling by RNA-seq

The transgenic *Arabidopsis* plants carrying pTA7001-*avrRxo1*-(196-1263), or pTA7001-*avrRxo1*-T167N-(196-1263), or pTA7001 empty vector control was used for the RNA-seq experiment. Four weeks-old *Arabidopsis* plants was sprayed with 10μM Dex to induce the gene expression of *avrRxo1* or *avrRxo1*-T167N. One hundred milligram leaf tissue of each lines was collected at 0, 3, 9, and 12 hrs after spraying DEX. The total RNAs were isolated using an RNeasy Plant Mini Kit (Qiagen, Germantown, MD) following the instruction of the user manual. Total RNAs of two biological replicates were processed and sequenced at BGI Americas (Cambridge, MA). The raw RNA-seq data was analyzed using the CLC genome workbench (6.0) software.

Chemicals

All chemicals used in this study were purchased from Sigma-Aldrich (St Louis, MI), only less it was specified. All oligo primers including the FAM-labelled M13 primes were synthesized from Integrated DNA Technologies, Inc. (Coralville, Iowa).

Statistical Analysis

All the mentioned experiments with three independent biology replicates were used to collected data for a *t*-test to test significant differences. The *p*-value<0.05 was considered as significant differences.

Acknowledgements

We thank Dr. Jake Tu (Virginia Tech) for allowing us to use the luminometer in his laboratory. Financial support was provided by grants from US National Science Foundation (NSF) (IOS-0845283 to BZ), the US-Israel Binational Agricultural Research and Development Fund (BARD) (US-4216-09 to BZ), Jeffery Trust Foundation (xx to BZ), and the Virginia Agricultural Experiment Station (VA135872). This project was also funded, in part, with an integrated, internal competitive grant from VAES, VCE, and the College of Agriculture and Life Sciences at Virginia Tech.

Supplementary information

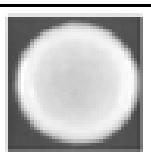
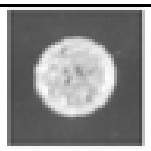
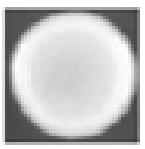
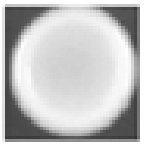
	-LT	-LTH + 3mM 3AT	α-galactosidase activity
AvrRxo1/ OsVOZ2			++
Positive control			+++
Negative control			-

Figure S3-1| AvrRxo1 interacts with OsVOZ2 *in vitro*

AvrRxo1 interacted with rice VOZ2 in yeast two-hybrid (Y2H) assays. The transformed yeast cells carrying pGADT7-Os*VOZ2* and pGBKT7-*avrRxo1*-T167N were grown on yeast SD medium lacking leucine, tryptophan, and histidine (-LTH) supplemented with 3 mM 3-Amino-1,2,4-triazole (3-AT). Both the alpha-galactosidase activity and growth on –LTH medium as marked with "+" suggested AvrRxo1-T167N interacted with the rice VOZ2.

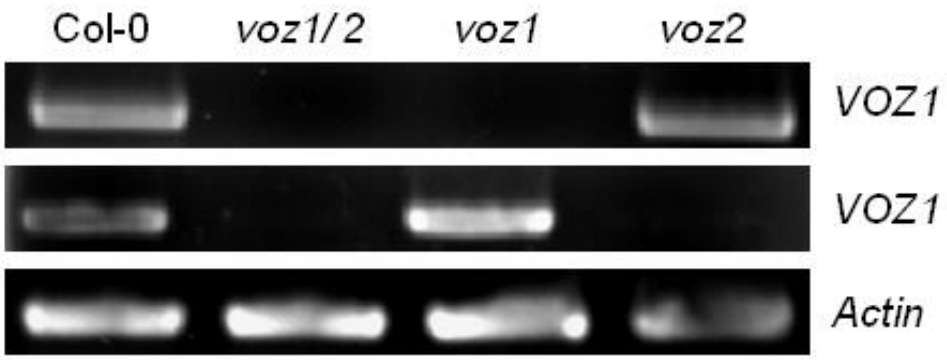


Figure S3-2| RT-PCR indicated the compromising of *VOZ1/2* expression in *Arabidopsis* T-DNA lines

The cDNA of *Arabidopsis* T-DNA lines of *voz1*, *voz2* and voz1/2 was used as PCR template to amplify *VOZ1* or *VOZ2* genes, for 26 cycles. The PCR product of *Actin* gene was used as a positive control that reflected equal loading. The RT-PCR results indicated the expressing of *VOZ1* or *VOZ2* was compromised in the *Arabidopsis* T-DNA lines.

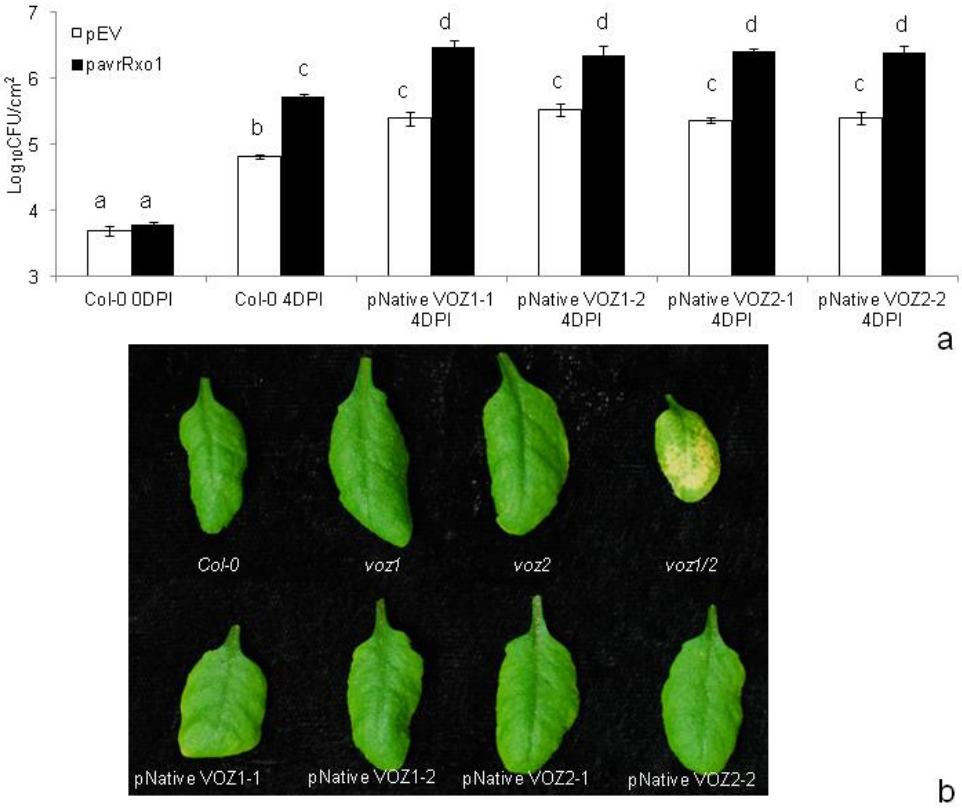


Figure S3-3| *VOZ1* or *VOZ2* expression suppressed the chlorosis phenotype of *voz1/2*, and restored the virulence activity of AvrRxo1

a. Leaves of Col-0 and *voz1/2* complementary lines of expressing either pNative-*VOZ1* or pNative-*VOZ2* were inoculated by infiltration with *Pst* DC3000-ΔCEL carrying pBZ598-*avrRxo1* or pBZ598-EV. Bacterial multiplication in the leaves was determined in triplicate at 0 and 4 days post inoculation. All the growth curve experiments were repeated three times with similar results. Bacterial growth was monitored by dilution plating, and the mean data of three replicates presented (means ± s.d.; n=3, $p<0.05$, Student's t-test, different letter indicate significant statistically difference). **b.** The chlorosis phenotype in *voz1/2* mutant was suppressed in the *voz1/2* complementary lines of expressing pNative-*VOZ1* or pNative-*VOZ2*.

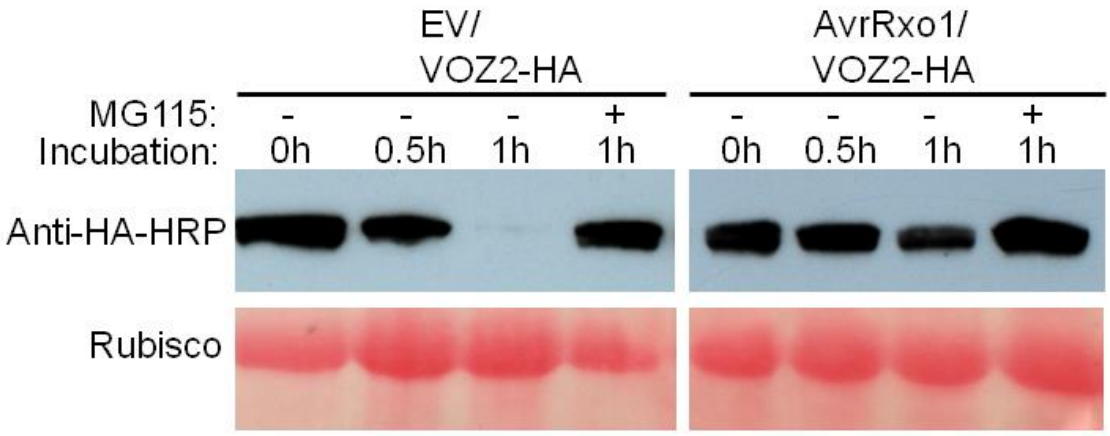


Figure S3-4| Co-expression of AvrRxo1 and VOZ2 in *Arabidopsis* suppressed VOZ2 degradation in *vivo*

Stable transgenic *Arabidopsis* plants co-expressing pTA7001-*avrRxo1* or pTA7001-EV with pEarleygate101-*VOZ2*-HA was used for VOZ2 degradation assay. The crude extract of VOZ2 from *Arabidopsis* plants co-expressing pTA7001-EV and *VOZ2* was completely degraded after one hour incubation at room temperature. The degradation of VOZ2 could be inhibited by proteasomal degradation inhibitor MG115 (40 mM). Crude extract of VOZ2 from *Arabidopsis* plants co-expressing pTA7001-*avrRxo1* with pEarleygate101-*VOZ2-HA* was stabled after one hour incubation at room temperature.

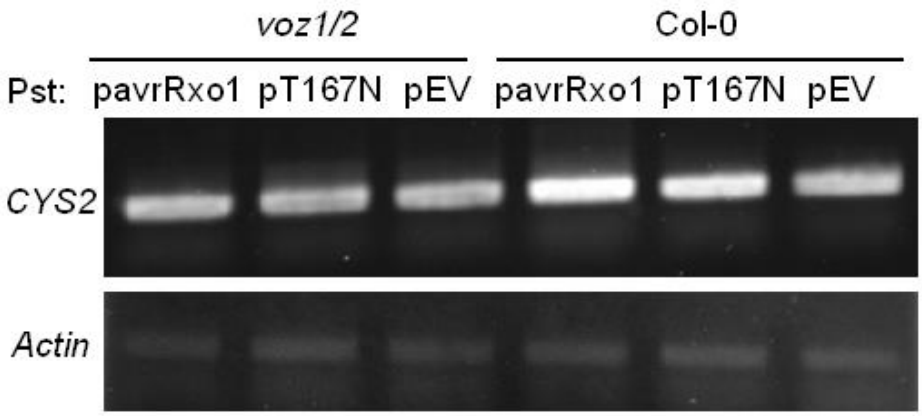


Figure S3-5| Pst DC3000-ΔCEL carrying pBZ598-*avrRxo1* induced the expression of *CYS2* in wild type *Arabidopsis* Col-0, but not the *voz1/2* mutant plants

Arabidopsis Col-0 or mutant voz1/2 were infiltrated with *Pst* DC3000-ΔCEL carrying pBZ598-*avrRxo1*, pBZ598-*avrRxo1*-T167N or pBZ598-EV (10^8 CFU/ml). RT-PCR of *CYS2* gene indicated the expression of *CYS2* in *Arabidopsis* Col-0 can be significantly induced by the infection of *Pst* DC3000-ΔCEL carrying pBZ598-*avrRxo1*, and slightly induced by the infection of *Pst* DC3000-ΔCEL carrying pBZ598-*avrRxo1*-T167N, at 6-hour post inoculation. The expression of *CYS2* in *Arabidopsis* mutant *voz1/2* cannot be induced by infection with the same bacteria.

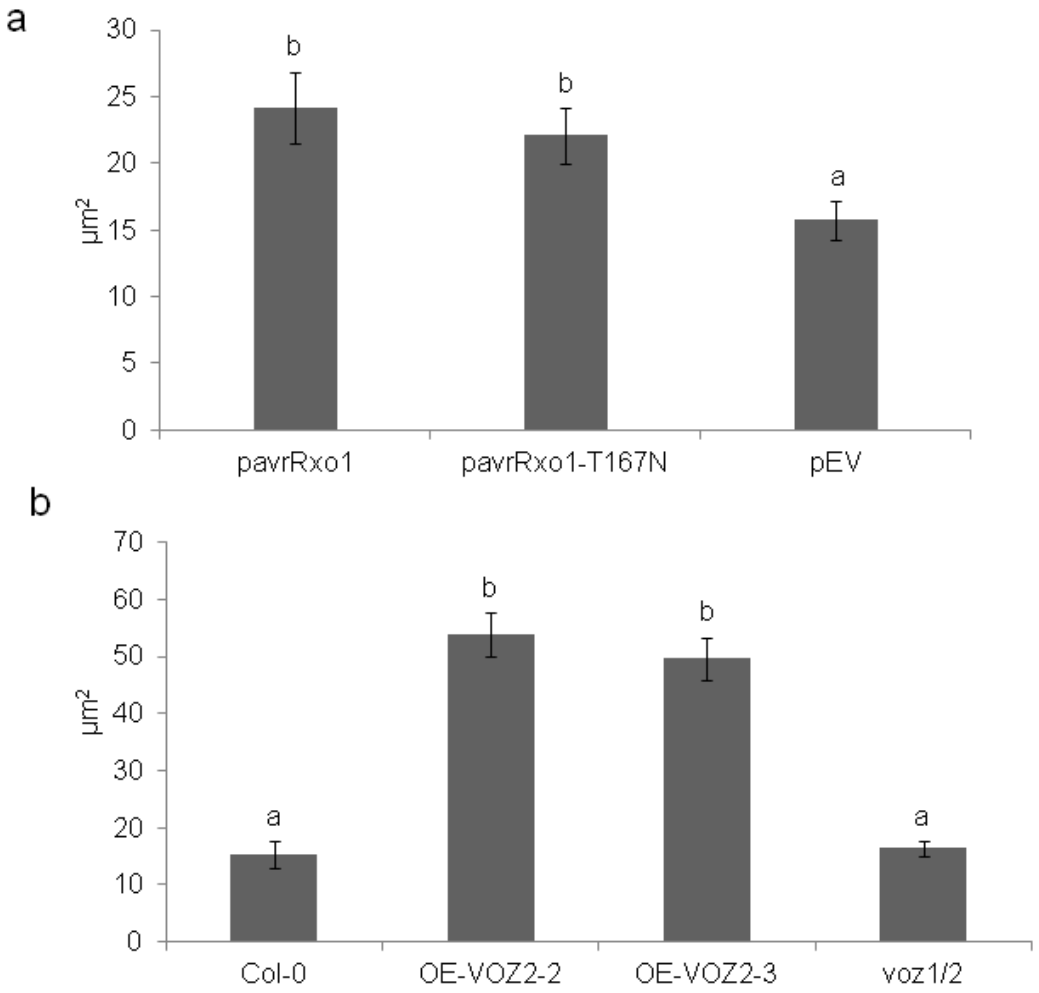


Figure S3-6| Increased *Arabidopsis* stomatal aperture size by *Pst* DC3000-ΔCEL carrying pBZ598-*avrRxo1*, or the overexpression of *VOZ2*

The *Arabidopsis* stomatal aperture size was measured under 400X magnification of a light microscope, and was estimated by the product of stomatal opening length and width. The mean value of stomatal aperture size was presented (mean ± s.d, n=10, $p<0.05$, Student's t-test, different letter indicate significant statistically difference). **a.** The *Arabidopsis* stomatal aperture size was increased after the infection of *Pst* DC3000-ΔCEL carrying pBZ598-*avrRxo1* or pBZ598-*avrRxo1*-T167N, but not pBZ598-EV. **b.** The stomatal size of *Arabidopsis* plant with overexpressing of *VOZ2* (OE-VOZ2) was induced in comparison to wild type Col-0. The size of *Arabidopsis* mutant *voz1/2* was not significantly different from that of the wild type Col-0.

Table S3-1| Sequences of primers used in this study

AvrRxo1 T167NFor	GAAATACCGCAACGGGTAAAAACCGGATTGCGACAAAA
AvrRxo1 T167NRev	TTTTGTCGCAATCCGGTTTTTACCCGTTGCGGTATTTC
AvrRxo1_196For	CACCGGATCCATGCATGAATCACTGCCGCATTCGGTT
AvrRxo1 _1263Rev	AGAAAGCTGGGTAGAGCTCAATTAGCTCGCTGTG
AvrRxo1 _1134Rev	GAGCTCGTCGATTACCTTATCAGCCAAAT
AtVOZ2For	CACCTCTAGAATGTCAAACCACCCGAAGATCACATC
AtVOZ2Rev	CTCCTTACGACCTTTGGTTGGAG
AtCYS2For	cacctctagaATGGCTACCATGTTGAAGG
AtCYS2Rev	gtcgacGTAGACAGGACTGACAACAGGAG
AtCYS2_promoterFor	CACCcctaggGAATTCGAGACTCTTACGCTTAGGG
AtCYS2_promoterRev	Tctagatgcttctcacttttgccttcttc
AtVOZ1For:	ATGACGGGGAAGCGATCAAAGAC
AtVOZ2Rev:	GGGGATATAATAGTCGCTTAG
ProAtVoz1For:	AGAAAAGTTGtgccagttaaaaccacagcctggt
ProAtVoz1Rev:	TTTGTACAAACTTGtAAGAAATCTCACCTCCCCACCA
ProAtVoz2For:	AGAAAAGTTGtgccatgggatcagcccagcttc
ProAtVoz2Rev:	TTTGTACAAACTTGtCTCCTCTAAACCCTGACCAAA

References

1. Lindow, S.E. and M.T. Brandl, *Microbiology of the phyllosphere.* Applied and environmental microbiology, 2003. **69**(4): p. 1875-1883.

2. Doehlemann, G. and C. Hemetsberger, *Apoplastic immunity and its suppression by filamentous plant pathogens.* New Phytol, 2013. **198**(4): p. 1001-16.

3. Jones, J.D. and J.L. Dangl, *The plant immune system.* Nature, 2006. **444**(7117): p. 323-9.

4. Göhre, V., et al., *Plant pattern-recognition receptor FLS2 is directed for degradation by the bacterial ubiquitin ligase AvrPtoB.* Current Biology, 2008. **18**(23): p. 1824-1832.

5. Zhao, B., et al., *A maize resistance gene functions against bacterial streak disease in rice.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(43): p. 15383-15388.

6. Zhao, B., et al., *A maize resistance gene functions against bacterial streak disease in rice.* Proc Natl Acad Sci 2005. **102**(43): p. 15383-15388.

7. Zhao, B.Y., et al., *The Rxo1/Rba1 locus of maize controls resistance reactions to pathogenic and non-host bacteria.* Theoretical and Applied Genetics, 2004. **109**(1): p. 71-79.

8. Zhao, B.Y., et al., *The avrRxo1 gene from the rice pathogen Xanthomonas oryzae pv. oryzicola confers a nonhost Defense reaction on maize with resistance gene Rxo1.* Molecular Plant-Microbe Interactions, 2004. **17**(7): p. 771-779.

9. Traore, S.M. and B. Zhao, *A novel Gateway(R)-compatible binary vector allows direct selection of recombinant clones in Agrobacterium tumefaciens.* Plant Methods, 2011. **7**(1): p. 42.

10. Turner, P., C. Barber, and M. Daniels, *Behaviour of the transposons Tn5 and Tn7 in Xanthomonas campestris pv. campestris.* Molecular and General Genetics MGG, 1984. **195**(1-2): p. 101-107.

11. Rentel, M.C., et al., *Recognition of the Hyaloperonospora parasitica effector ATR13 triggers resistance against oomycete, bacterial, and viral pathogens.* Proc Natl Acad Sci U S A, 2008. **105**(3): p. 1091-6.

12. Earley, K., et al., *Gateway-compatible vectors for plant functional genomics and proteomics.* Plant J, 2006. **45**(4): p. 616 - 629.

13. Clough, S.J. and A.F. Bent, *Floral dip: a simplified method forAgrobacterium-mediated transformation ofArabidopsis thaliana.* The plant journal, 1998. **16**(6): p. 735-743.

14. Zuo, J. and N.-H. Chua, *Chemical-inducible systems for regulated expression of plant genes.* Current Opinion in Biotechnology, 2000. **11**(2): p. 146-151.

15. Karimi, M., B. De Meyer, and P. Hilson, *Modular cloning in plant cells.* Trends in Plant Science, 2005. **10**(3): p. 103-105.

16. Krasileva, K.V., D. Dahlbeck, and B.J. Staskawicz, *Activation of an Arabidopsis resistance protein is specified by the in planta association of its leucine-rich repeat domain with the cognate oomycete effector.* Plant Cell, 2010. **22**(7): p. 2444-58.

17. Katagiri, F., R. Thilmony, and S.Y. He, *The Arabidopsis thaliana-Pseudomonas syringae interaction.* The *Arabidopsis* book/American Society of Plant Biologists, 2002. **1**.

18. Melotto, M., et al., *Plant stomata function in innate immunity against bacterial invasion.* Cell, 2006. **126**(5): p. 969-980.

19. He, P., L. Shan, and J. Sheen, *The use of protoplasts to study innate immune responses.* Methods Mol Biol, 2007. **354**: p. 1-9.

20. Lu, D., et al., *A receptor-like cytoplasmic kinase, BIK1, associates with a flagellin receptor complex to initiate plant innate immunity.* Proceedings of the National Academy of Sciences, 2010. **107**(1): p. 496-501.

21. Karimi, M., D. Inze, and A. Depicker, *GATEWAY vectors for Agrobacterium-mediated plant transformation.* Trends Plant Sci, 2002. **7**: p. 193 - 195.

22. He, P., L. Shan, and J. Sheen, *The use of protoplasts to study innate immune responses*, in *Plant-Pathogen Interactions*. 2007, Springer. p. 1-9.

23. Spoel, S.H., et al., *Proteasome-mediated turnover of the transcription coactivator NPR1 plays dual roles in regulating plant immunity.* Cell, 2009. **137**(5): p. 860-72.

24. Van Eck, L., et al., *The transcriptional network of WRKY53 in cereals links oxidative responses to biotic and abiotic stress inputs.* Functional & integrative genomics, 2014. **14**(2): p. 351-362.

25. Zatakia, H.M., et al., *ExpR coordinates the expression of symbiotically important, bundle-forming Flp pili with quorum sensing in Sinorhizobium meliloti.* Applied and environmental microbiology, 2014. **80**(8): p. 2429-2439.

26. Schuelke, M., *An economic method for the fluorescent labeling of PCR fragments.* Nature biotechnology, 2000. **18**(2): p. 233-234.

27. Hilu, K.W. and J.L. Randall, *Convenient method for studying grass leaf epidermis.* Taxon, 1984: p. 413-415.

28. Yang, Z., et al., *Production of Autopolyploid Lowland Switchgrass Lines Through In Vitro Chromosome Doubling.* BioEnergy Research, 2014. **7**(1): p. 232-242.

29. DebRoy, S., et al., *A family of conserved bacterial effectors inhibits salicylic acid-mediated basal immunity and promotes disease necrosis in plants.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(26): p. 9927-9932.

30. McNellis, T.W., et al., *Glucocorticoid-inducible expression of a bacterial avirulence gene in transgenic Arabidopsis induces hypersensitive cell death.* The Plant Journal, 1998. **14**(2): p. 247-257.

31. Wei, W., et al., *The gene coding for the Hrp pilus structural protein is required for type III secretion of Hrp and Avr proteins in Pseudomonas syringae pv. tomato.* Proceedings of the National Academy of Sciences, 2000. **97**(5): p. 2247-2252.

32. Mitsuda, N., et al., *VOZ; Isolation and Characterization of Novel Vascular Plant Transcription Factors with a One-Zinc Finger from Arabidopsis thaliana.* Plant and Cell Physiology, 2004. **45**(7): p. 845-854.

33. Nakai, Y., et al., *Vascular plant one-zinc-finger protein 1/2 transcription factors regulate abiotic and biotic stress responses in Arabidopsis.* The Plant Journal, 2013. **73**(5): p. 761-775.

34. Yasui, Y., et al., *The phytochrome-interacting VASCULAR PLANT ONE–ZINC FINGER1 and VOZ2 redundantly regulate flowering in Arabidopsis.* The Plant Cell Online, 2012. **24**(8): p. 3248-3263.

35. Mou, Z., W. Fan, and X. Dong, *Inducers of plant systemic acquired resistance regulate NPR1 function through redox changes.* Cell, 2003. **113**(7): p. 935-944.

36. Lopes, U.G., et al., *p53-dependent induction of apoptosis by proteasome inhibitors.* Journal of Biological Chemistry, 1997. **272**(20): p. 12893-12896.

37. Hwang, J.E., et al., *Distinct expression patterns of two Arabidopsis phytocystatin genes, AtCYS1 and AtCYS2, during development and abiotic stresses.* Plant Cell Rep, 2010. **29**(8): p. 905-15.

38. Hwang, J.E., et al., *Distinct expression patterns of two Arabidopsis phytocystatin genes, AtCYS1 and AtCYS2, during development and abiotic stresses.* Plant cell reports, 2010. **29**(8): p. 905-915.

39. Mittal, S. and K.R. Davis, *Role of the phytotoxin coronatine in the infection of Arabidopsis thaliana by Pseudomonas syringae pv. tomato.* MPMI-Molecular Plant Microbe Interactions, 1995. **8**(1): p. 165-171.

40. Schulze-Lefert, P. and S. Robatzek, *Plant pathogens trick guard cells into opening the gates.* Cell, 2006. **126**(5): p. 831-4.

41. Zeng, W., M. Melotto, and S.Y. He, *Plant stomata: a checkpoint of host immunity and pathogen virulence.* Current opinion in biotechnology, 2010. **21**(5): p. 599-603.

42. Melotto, M., W. Underwood, and S.Y. He, *Role of stomata in plant innate immunity and foliar bacterial diseases.* Annual review of phytopathology, 2008. **46**: p. 101.

43. Gudesblat, G.E., P.S. Torres, and A.A. Vojnov, *Xanthomonas campestris overcomes Arabidopsis stomatal innate immunity through a DSF cell-to-cell signal-regulated virulence factor.* Plant Physiol, 2009. **149**(2): p. 1017-27.

44. Turner, N.C. and A. Graniti, *Fusicoccin: a Fungal Toxin that opens Stomata.* Nature, 1969. **223**(5210): p. 1070-1071.

45. Godoy, G., et al., *Use of mutants to demonstrate the role of oxalic acid in pathogenicity of Sclerotinia sclerotiorum on Phaseolus vulgaris.* Physiological and Molecular Plant Pathology, 1990. **37**(3): p. 179-191.

46. Chisholm, S.T., et al., *Host-microbe interactions: shaping the evolution of the plant immune response.* Cell, 2006. **124**(4): p. 803-14.

47. Hurley, B., et al., *The Pseudomonas syringae type III effector HopF2 suppresses Arabidopsis stomatal immunity.* PLoS One, 2014. **9**(12): p. e114921.

48. Lozano-Duran, R., et al., *The bacterial effector HopM1 suppresses PAMP-triggered oxidative burst and stomatal immunity.* New Phytol, 2014. **202**(1): p. 259-69.

49. Gimenez-Ibanez, S., et al., *The bacterial effector HopX1 targets JAZ transcriptional repressors to activate jasmonate signaling and promote infection in Arabidopsis.* PLoS Biol, 2014. **12**(2): p. e1001792.

50. Kleemann, J., et al., *Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen Colletotrichum higginsianum.* PLoS Pathog, 2012. **8**(4): p. e1002643.

51. Carruthers, V.B. and L.D. Sibley, *Sequential protein secretion from three distinct organelles of Toxoplasma gondii accompanies invasion of human fibroblasts.* Eur J Cell Biol, 1997. **73**(2): p. 114-23.

52. Nakai, Y., et al., *Overexpression of VOZ2 confers biotic stress tolerance but decreases abiotic stress resistance in Arabidopsis.* Plant Signal Behav, 2013. **8**(3): p. e23358.

53. Yasui, Y. and T. Kohchi, *VASCULAR PLANT ONE-ZINC FINGER1 and VOZ2 repress the FLOWERING LOCUS C clade members to control flowering time in Arabidopsis.* Biosci Biotechnol Biochem, 2014. **78**(11): p. 1850-5.

54. Yasui, Y., et al., *The phytochrome-interacting vascular plant one-zinc finger1 and VOZ2 redundantly regulate flowering in Arabidopsis.* Plant Cell, 2012. **24**(8): p. 3248-63.

55. Cheong, H., et al., *Xanthomonas oryzae pv. oryzae type III effector XopN targets OsVOZ2 and a putative thiamine synthase as a virulence factor in rice.* PLoS One, 2013. **8**(9): p. e73346.

56. Mukhtar, M.S., et al., *Independently evolved virulence effectors converge onto hubs in a plant immune system network.* Science, 2011. **333**(6042): p. 596-601.

57. Rose, D., *Epidemiology of maize streak disease.* Annual Review of Entomology, 1978. **23**(1): p. 259-282.

58. Nino-Liu, D.O., P.C. Ronald, and A.J. Bogdanove, *Xanthomonas oryzae pathovars: model pathogens of a model crop.* Mol Plant Pathol, 2006. **7**(5): p. 303-24.

59. Wang, L., et al., *Novel candidate virulence factors in rice pathogen Xanthomonas oryzae pv. oryzicola as revealed by mutational amalysis.* Applied and Environmental Microbiology, 2007. **73**(24): p. 8023-8027.

60. Belenghi, B., et al., *AtCYS1, a cystatin from Arabidopsis thaliana, suppresses hypersensitive cell death.* European Journal of Biochemistry, 2003. **270**(12): p. 2593-2604.

61. Massonneau, A., et al., *Maize cystatins respond to developmental cues, cold stress and drought.* Biochim Biophys Acta, 2005. **1729**(3): p. 186-99.

62. van der Linde, K., et al., *A maize cystatin suppresses host immunity by inhibiting apoplastic cysteine proteases.* Plant Cell, 2012. **24**(3): p. 1285-300.

63. Rooney, H.C., et al., *Cladosporium Avr2 inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance.* Science, 2005. **308**(5729): p. 1783-6.

64. Mueller, A.N., et al., *Compatibility in the Ustilago maydis-maize interaction requires inhibition of host cysteine proteases by the fungal effector Pit2.* PLoS Pathog, 2013. **9**(2): p. e1003177.

65. Tian, M., B. Benedetti, and S. Kamoun, *A Second Kazal-like protease inhibitor from Phytophthora infestans inhibits and interacts with the apoplastic pathogenesis-related protease P69B of tomato.* Plant Physiol, 2005. **138**(3): p. 1785-93.

66. Tian, M., et al., *A Phytophthora infestans cystatin-like protein targets a novel tomato papain-like apoplastic protease.* Plant Physiol, 2007. **143**(1): p. 364-77.

67. Turk, V. and W. Bode, *The cystatins: protein inhibitors of cysteine proteinases.* FEBS letters, 1991. **285**(2): p. 213-219.

Chapter IV:

Co-expression analysis of *Arabidopsis* papin-liked cysteine proteases and cystatins in abiotic and biotic stress conditions

Shuchi Wu[1], Yi Xing[2], Song Li[3], Bingyu Zhao[1*]
[1]Department of Horticulture, Virginia Tech, Blacksburg, VA, 24061
[2]College of Computer and Information Science, Northeastern University, Boston, MA, 02115
[3]Department of Crop and Environmental Sciences, Virginia Tech, Blacksburg, VA, 24061

Abstract

Plant papin-liked cysteine proteases (PLCPs) and plant cystatins are the major protease/protease inhibitor pairs involved in signaling pathways in response to biotic and abiotic stress. The *Arabidopsis* genome has 31 PLCPs and 7 cystatins. The co-expressional regulation of all *Arabidopsis* PLCPs and cystatins has never been systematically investigated. In this study, we analyzed the expression patterns of 28 PLCPs and seven cystatins in *Arabidopsis* in response to biotic or abiotic stress, by reprocessing and integrating microarray data from the AtGenExpress database. We also performed enzyme assays and evaluated the inhibition specificity of seven *Arabidopsis* cystatins to five most abundant PLCPs in *Arabidopsis*. Finally, we utilized the SVMS (support vector machines) package in R to predict gene-gene interaction network of PLCP-cystatin genes in *Arabidopsis*. Our results suggest that *Arabidopsis* has different sets of PLCP and cystatin genes in response to virulent or avirulent bacterial pathogen infection. The PLCP and cystatin genes involved in response to bacterial pathogen infection and wounding stress were completely distinguished from each other. Most PLCP and cystatin genes involved in drought resistance also participated in plant basal defense to bacterial pathogen.

Introduction

Proteases are widely distributed in all organisms and have key roles in regulation of diverse biological processes [1-5]. In animal and plant genomes, about two percent of genes code for proteases [6]. Though the basic biochemical function of proteases is acting as proteolytic enzymes to degrade nonfunctional proteins, many proteases act as far more than "protein recyclers" *in vivo* [1, 2]. Indeed, the catalytic events mediated by proteases can serve as signals in regulating diverse biological processes in plants, such as embryogenesis,

flowering, leaf senescence, and plant defense response [7-10]. Plant proteases can be divided into five major groups based on their catalytic residues: serine proteases, cysteine proteases, threonine proteases, aspartate proteases and metalloproteases (using a metal ion, like zinc) [1, 11, 12]. Among the different kinds of plant proteases, serine and cysteine proteases are the groups that have been intensively studied in the past 50 years with many cysteine proteases having been reported to be involved in plant immunity [1, 12]. Plant genome encode over 140 cysteine proteases, which can be divided into 15 families of 5 clans [1, 6]. Many of them are involved with programmed cell death (PCD) and plant immunity, and others are involved in biological process such as flowering time, embryo development, epidermal cell aging, and responses to abiotic stress etc. [8-10, 13-15].

Cysteine proteases belonging to family C1A, which are also known as papin-like cysteine proteases (PLCPs), are the major cysteine proteases associated with PCD and plant immunity [1, 6]. As one of the key operators of PCD, proteases, especially PLCPs, can be potentially damaging to plant cells when they are overexpressed or miss-activated at the translational level [16]. Therefore, plants evolved diverse mechanisms to restrict PLCPs' activity. All plant PLCPs contain auto-inhibitory pro-domains in their precursors, which can be automatically processed to generate mature and active proteases [12, 17]. And the PLCP activation process usually depends on low pH or is elicited by H_2O_2 or SDS [18, 19]. In addition to self-auto-inhibition of PLCPs, plant genomes also have a collection of specialized protease inhibitors to inhibit PLCPs activities. These PLCP inhibitors are also widely conserved in both animals and microorganisms [2-5]. The protease inhibitors can either directly or indirectly block the activation site of the cognate PLCPs, through conserved mechanisms [2]. Among the PLCP inhibitors, Cystatins are a kind of competitive protein inhibitors of PLCPs [2, 20]. All cystatins contain the conserved Gln-Xaa-Val-Xaa-Gly motif in the central region of their polypeptide [20, 21]. Plant cystatins also have a conserved Gly residue in the N-terminus, and a Pro-Trp (or Leu-Trp) motif in the C-terminus [20]. The center motif (Gln-Xaa-Val-Xaa-Gly) of the cystatin is the functional part, which physically interacts with the activation site of the PLCP [20]. Besides controlling PLCPs enzyme activity at the protein level, plants also tightly regulate their gene expression at the transcriptional level [18, 22]. PLCPs expression is usually maintained at a base level and defense-related PLCP genes will be highly induced by certain pathogens or plant defense responses [13, 23, 24]. Wounding, for example damage caused by insects, could potentially positively or negatively regulate the expression of PLCP and cystatin genes [25, 26].

Land plants have evolved at least two layers of immunity: (1) the basal defense response against the majority of microbes that is usually triggered by plant

recognition of pathogen-associated molecular patterns (PAMPs), which is designated as PAMP triggered immunity (PTI); (2) the second layer of plant immunity is designated as effector triggered immunity (ETI), which is a more specific and amplified defense response and involves plant resistance genes (*R* genes) that directly or indirectly recognize cognate pathogen effectors [27]. ETI usually results in hypersensitive response or PCD at the local pathogen infection sites, which may function as "suicide" mechanism to restrict pathogen proliferation within the infected tissue [27]. Several PLCPs have been reported to be key signaling components of plant immunity. For example, tomato *RCR3* encodes a cysteine protease that is specifically required by *Cf2*-mediated resistance to *Cladosporium fulvum* [28]. The *Arabidopsis* PLCP gene *RD19A* is up-regulated during bacterial wilt disease caused by *Ralstonia solanacearum,* and its protein product RD19A physically interacts with *R. solanacearum* type III effector PopP2 *in vivo* [13]. *Arabidopsis* PLCP RD21A was also reported to be a positive regulator of *Arabidopsis* defense to the necrotrophic pathogen *Botrytis cinerea* [29]. A group of maize *PLCPs* (*CP2, CP1A, CP1B* and *XCP2*) are induced by Salicylic Acid (SA), and also have integrated roles in SA-associated defense signaling pathways [30, 31].

Plant cystatins function instead as negative regulators of plant immunity. For example, maize cystatin *CC9* is strongly induced by the fungal pathogen *Ustilago maydis (U. maydis) during compatible interactions [31]*. Transgenic maize plants that overexpress *CC9* suppress the activity of several *PLCPs (CP2, CP1A, CP1B and XCP2)*, and also reduce the expression of PR genes induced by SA treatment, and enhance maize susceptibility to *U. maydis [31]*. *Arabidopsis* cystatin AtCYS1 was also found to suppress PCD triggered by pathogens and wounding [25]. Our previous result (Wu et al, 2015) suggests that *AtCYS2* can be regulated by the type III effector AvrRxo1 to suppress stomatal-related immunity. It is also known that both fungal and oomycete pathogens can secrete cystatin-like proteins to suppress plant immunity [10, 30]. Therefore, the interplay of PLCP and cystatin plays key roles in plant immunity to phytopathogens and in the response to physical wounding, where the molecular mechanism still needs further investigation.

Plant PLCPs and cystatins may also have roles in response to drought stress. For example, among the 31 *Arabidopsis* PLCP genes, 13 of them were named as Responsive to Desiccation (RD) or RD-liked genes [32]. And *RD21A* was originally identified as the most abundant cysteine protease gene induced by dehydration, whose expression level can be unregulated by more than five times in response to drought stress [33, 34]. RD21A might serve a signal transducer of regulating plant drought resistance since it was reported to be the substrate of the E3 Ub ligase AtAIRP3/LOG2, which is a positive regulator of the ABA-mediated drought stress tolerance mechanism [35].

*Arabidopsis*Other examples are the tomato cysteine protease gene TDI-65 and a group of four to five cysteine protease genes from wheat, which were all shown to be induced under drought stress [36, 37]. Cystatins were also reported to be regulating plant drought tolerance. For example, ectopic overexpression of rice cystatin (oryzacystatin-I) leads to enhanced drought tolerance in soybean and *Arabidopsis* [38]. A group of 5 Corn Cystatin genes (*CCII*, *CC3*, *CC4*, *CC5* and *CC9*) were reported to be down-regulated in response to water starvation [39]. Under drought conditions, plants regulate the stomatal opening to prevent water loss [40, 41]. Importantly, plants also regulate stomata opening in response to bacterial infection and certain bacterial pathogens can enforce the stomatal re-opening as a virulence mechanism [42]. Thus, plants may employ shared signaling pathways to regulate stomata opening in response to drought stress and bacterial pathogen infection [43]. The interplay of PLCP and cystatins that may co-regulate abiotic and biotic stress has not been intensively studied.

Co-expression network analysis has been used for effectively associating genes with unknown function to a targeted signaling pathway [44]. Co-expression analysis assumes that genes in the same pathway, regulated by the same genes or being a part of the functional complex, should present similar expression patterns under the same conditions. Leal and colleagues employed co-expression network analysis to identify a large collection of *Arabidopsis* immune related genes, where many of them were previously uncharacterized [45]. The public micro array database, AtGenExpress (by TAIR, www.*Arabidopsis*.org) provides a good source for performing co-expression network analysis on *Arabidopsis* PLCP and cystatin genes, and predicting the associated PLCP-cystatin gene pairs involved in the targeted signaling pathway.

Support Vector Machines (SVMs) are supervised computer learning models that were developed by Cortes & Vapnik (1995) for binary classification [46]. SVMs include associated learning algorithms that can learn from the classification of a given set of training examples, and build a model to classify new examples [46]. For Biologists, SVMs provide big advantages in analyzing large sets of gene expression data, such as micro array data, and predicting gene clustering [47]. In the specific case of analyzing PLCP-cystatin interplay, a subset of putative interacting PLCP-cystatin gene pairs can serve as the training set for SVMs. Then SVMs can build a classification model to predict the remaining PLCP-cystatin gene interaction.

Thus far, studies on PLCP-associated plant immunity and abiotic stress are isolated and fragmentary. Comprehensive expression profiling of all PLCP and cystatin genes, characterization of the regulation of protease activities, and study of the interplay between PLCPs and cystatins in plant immunity haven’t been reported.

In this study, we analyzed co-expression patterns of 28 PLCP and 7 cystatin genes in *Arabidopsis* in response to biotic or abiotic stress, by reprocessing and integrating microarray data from the AtGenExpress database. There are 31 candidate PLCP genes and 7 candidate cystatin genes present in the *Arabidopsis* genome (Table S4-1, list of the PLCP and Cystatin genes, in Supplementary Information), 28 PLCP and 7 cystatin genes are present in our microarray dataset, while other three PLCP genes were not presented by probes or were excluded from this study because of conflicting problems. For studying PLCPs associated with biotic stress, we analyzed the microarray data of wild type *Arabidopsis* plants inoculated with either virulence and avirulent bacterial pathogens (TAIR Accession ID: 1008031517) [48]. For studying PLCPs associated with abiotic stress, we analyzed the micro array data of *Arabidopsis* in response to wounding and drought stress (TAIR Accession ID: 1007966439; 1007966668) [49]. In addition, we also performed protease enzyme activity and inhibition assays of the five most abundant PLCPs and seven cystatins in *Arabidopsis*. Finally, we predicted the relationship between the 23 PLCPs and 7 cystatins by using the SVMs package "e1071" of R (Version 3.2.0) [50].

Results and discussions

Expression Profiles of Seven Cystatins and Five PLCPs of Arabidopsis Infected with Pseudomonas syringae ES4326

Several *Arabidopsis* PLCP and cystatin genes were reported to be involved in plant immunity [1, 2, 6]. In this study, we characterized the expression patterns of seven *Arabidopsis* cystatins and five PLCPs under the condition with pathogen inoculation. The five PLCPs were selected because of their relatively higher expression levels either in leave or root tissue (Figure S4-1 in Supplementary Infomation). The gene expression patterns were generated from the AtGenExpress microarray data of *Arabidopsis thaliana* (Col-0) plants challenged with *Pseudomonas syringae* strain (*P. syringae* ES4326), with or without the avirulence effector AvrRpt2 (TAIR Accession ID: 1008031517). Figure 4-1a and 4-1b summarizes the gene expression patterns of the selected genes in the condition with virulent or avirulent pathogen inoculation.

Although the expression levels of the five selected PLCP genes did not change significantly in response to either virulent or avirulent pathogens, *RD21A* was slightly downregulated and *CTB3* was slightly upregulated in response to both virulent and avirulent pathogens (Figure 4-1a). The relative stable expression of the five most abundant PLCP genes suggested their activities might be regulated at translational level. In contrast to the stable expression of five PLCP genes, the

expression levels of seven *Arabidopsis* cystatin genes have relatively more significant fluctuations (Figure 4-1b). *CYS1*, *CYS3*, *CYS6* and *CYS7* did not change significantly in response to either virulent or avirulent pathogens. Therefore, these four cystatin genes may not closely related to plant immunity to *P. syringae*. The expression of *CYS2* is suppressed at 16 h and 48 h in response to the virulent *P. syringae* ES4326, while the suppression of *CYS2* was compromised in response to avirulent *P. syringae* ES4326 carrying AvrRpt2. Therefore, *CYS2* and its regulated PLCPs are more likely to associate with plant basal defense. *CYS4* can be specifically induced at 16 h after the inoculation of *P. syringae* ES4326 carrying AvrRpt2, which suggesting *CYS4* and its related PLCPs might be associated with AvrRpt2 triggered-ETI. It is also known AvrRpt2 encodes a functional cysteine protease [51], therefore, it will be interesting to test if CYS4 can actually inhibit the protease activity of AvrRpt2 *in vivo*. *CYS5* was reduced two folds in response to *P. syringae* ES4326 with or without *avrRpt2*. Therefore, *CYS5* and its related PLCPs might also participate in plant defense response.

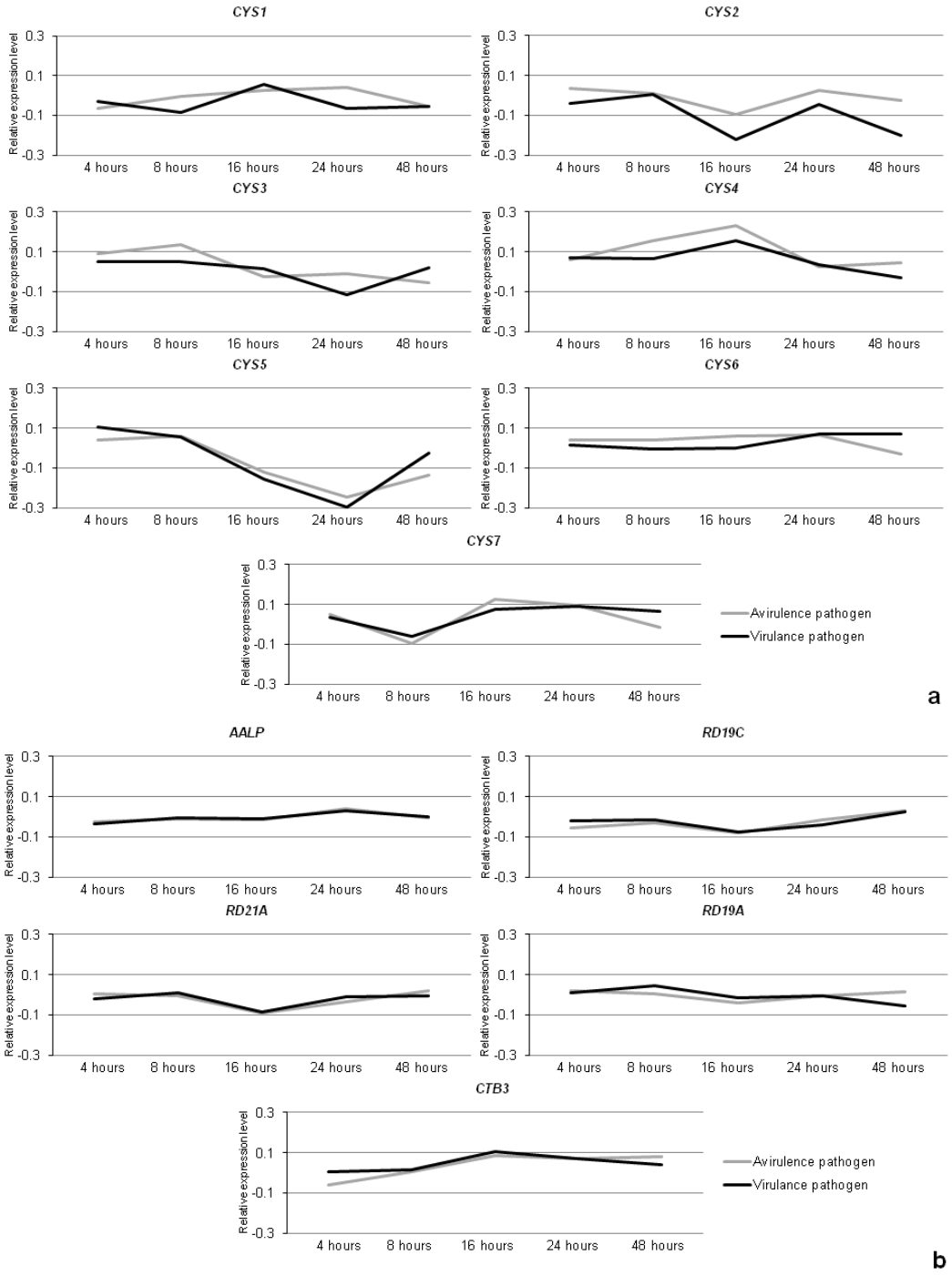


Figure 4-1| Expression profiles of seven cystatins and the five most abundant PLCPs of *Arabidopsis* plants infected with *Pseudomonas syringae* ES4326

Gene expression patterns were generated from the AtGenExpress microarray data of *Arabidopsis thaliana* (Col-0) plants challenged with *Pseudomonas syringae* strain ES4326, with or without the avirulence effector AvrRpt2 (TAIR Accession ID: 1008031517). The relative expression value represented the difference of the

expression value between bacteria infected treatment and mock treatment. The gray line represented the condition of bacterial infection of *Pst* ES4326 with AvrRpt2, and the black line represented the condition of bacterial infection of *Pst* ES4326. **a.** The line charts reflected the relative expression values of different cystatin genes at different time points post bacterial infection. **b.** The line charts reflected the relative expression values of different cystatin genes at different time points post bacterial infection.

Co-expression Analysis of Arabidopsis PLCPs and Cystatins that May Involve in Plant Immunity

To identify if any PLCPs are co-regulated with certain cystatins in response to the infection of bacterial pathogens, we acquired the same microarray data set as described above, and filtered it with 31 *Arabidopsis* PLCPs and seven cystatins related probes. Three PLCP genes (*RDL4*, *RDL5* and *CEP2*) that have probe missing or conflicting problems (Table S4-1). Two housekeeping genes (*UCP022280* and *PTB1*) were also selected in the dataset as controls. To select two housekeeping genes as controls was to avoid the accident expression of one control. The raw microarray data was processed into relative gene expression value as described in the materials and methods. The relative gene expression values of 28 PLCPs, the two control genes, and seven cystatins were used to calculate the correlation value between PLCPs and cystatins, which was presented as a heat map (Figure 4-2a and 4-2b). The darkness of the brick was inversely proportional to the correlation value. And the description of correlation strength was followed by the definition in the Statistic book by Dancey et al (2004) [52]. In brief, the correlation value less than 0.4 was considered as weak correlation and the correlation value greater than 0.6 was considered as strong (or significant) correlation. And the correlation value between 0.4 and 0.6 was considered as modulate correlation. At the selected signification level, none of the cystatins have significant correlation with the two housekeeping genes at the same time.

When inoculated with virulent *P. syringae* ES4326, *Arabidopsis* will mount a quick but weak basal defense response that will be suppressed by the virulence pathogen [27]. Therefore, the PLCPs that have positively roles in immunity could be induced at the early stage but suppressed at the late stage of the infection with virulent bacterial pathogens. The cystatins that are co-regulated with the virulent bacterial as those PLCPs could be considered as the specific inhibitors to regulate PCLP enzyme activities at protein level. When the wild type *Arabidopsis* plants are inoculated with *P. syringae* ES4326 expressing AvrRpt2, the *Arabidopsis* R protein Rps2 recognizes AvrRpt2 and triggers ETI [53]. The PLCPs that have

positively roles in AvrRpt2-triggered ETI will be induced and their corresponding cystatins should be co-regulated to keep the balance of cysteine protease activity.

As shown in Figure 4-2a, when inoculated with *P. syringae* ES4326, the *Arabidopsis* PLCPs could be divided into five groups based on their co-expression pattern in relation to different cystatins. The top 11 PLCPs showed weak correlation with most cystatins. And four of them (*RD19B*, *SAG12*, *RDL1* and *PAP5*) did not have significant correlation with any cystatin, that suggested they might not be related with *Arabidopsis* basal defense to *P. syringae*. The other 20 PLCPs showed significant correlation with at least one cystatin. And five of them (*RD21B*, *XCP2*, *AALP*, *ALP2* and *RDL3*) have significant correlation value with two or three cystatins. It is possible that these five PLCPs have a positive roles in *Arabidopsis* immunity, whose activities are suppressed by cystatins under disease susceptible conditions. In the condition of AvrRpt2 triggered ETI (Figure 4-2b), the correlation pattern of PLCP-cystatin pairs did not show clear clustering. ETI is an amplified defense response, where the HR associated PCD might regulate the expression of PLCPs and cystatins.

By comparing the correlation heat map in Figure 4-2a and 4-2b, *XCP2* and *PAP3* did not show significant correlation with any cystatin under the AvrRpt2 triggered ETI condition, but showed significant correlation with two or three cystatins under susceptible condition triggered by the virulence bacteria. Thus *XCP2* and *PAP3* could be classified as candidate genes that are specifically related with AvrRpt2 triggered ETI, whose activity was tightly controlled by cystatins in plant basal defense, but not under the ETI condition. The expression of *CTB2*, *RD19C*, and *THI1* are strongly correlated with the expression *CYS4* when *Arabidopsis* plants were inoculated with virulent *P. syringae* ES4326, but only has weak correlation with *CYS4* under the condition of AvrRpt2 triggered ETI. Thus, we may predict CTB2, RD19C and THI1 are essential for the AvrRpt2 triggered-ETI. However, the effector AvrRpt2 itself is a cysteine protease that might affect the function of cystatins and affect the accuracy of our prediction.

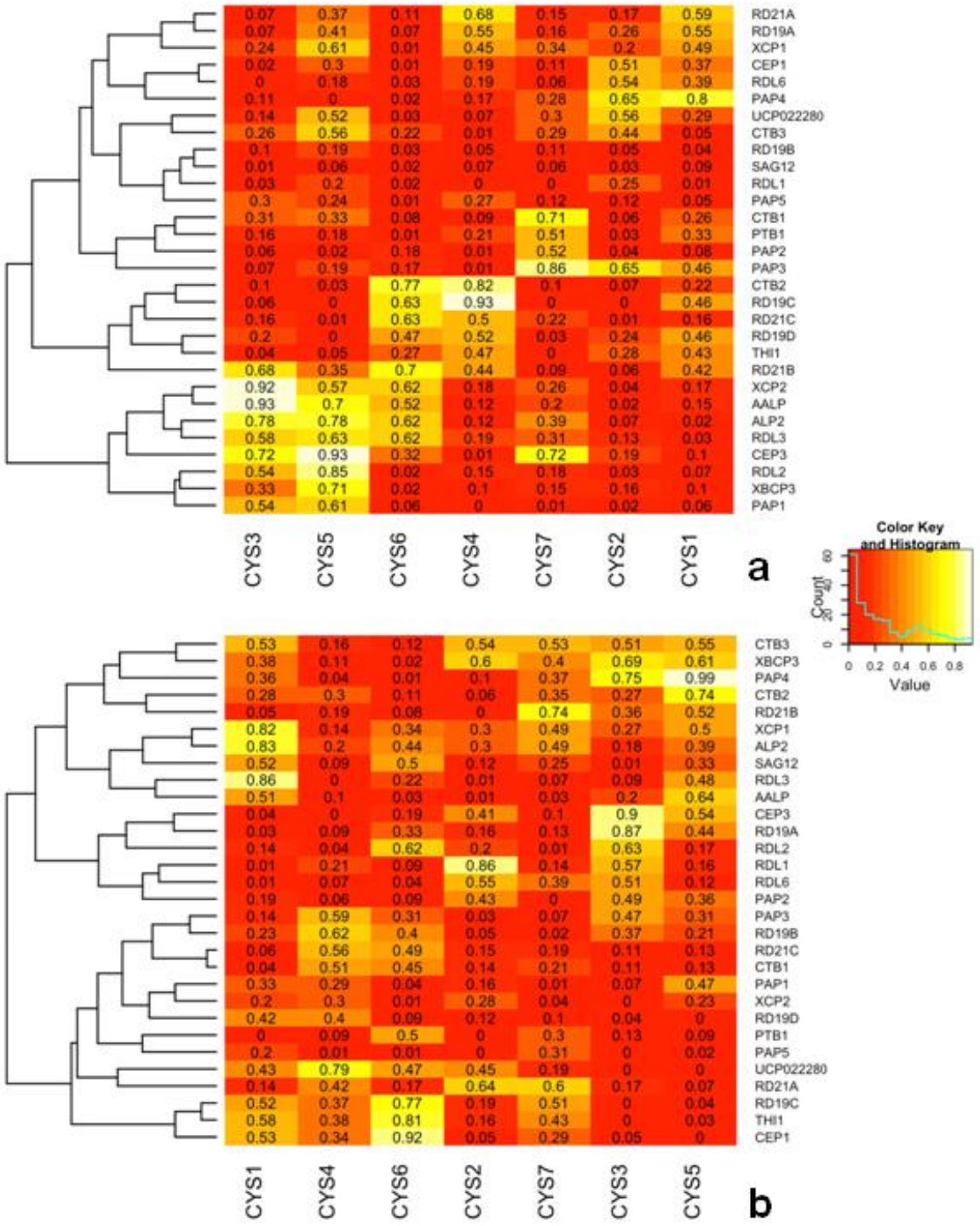


Figure 4-2| Co-expression patterns of *Arabidopsis* PLCPs and cystatins and clustering of PLCPs in biotic stress condition

Gene co-expression patterns were generated from the AtGenExpress microarray data of *Arabidopsis thaliana* (Col-0) plants challenged with

Pseudomonas syringae strain ES4326, with or without the avirulence effector AvrRpt2 (TAIR Accession ID: 1008031517). The relative expression value was the difference of the expression value between bacteria infected treatment and mock treatment. The correlation value of PLCP genes (or housekeeping genes: *PTB1* and *UCP022280*) and cystatin genes were calculated based on the relative expression values of every gene at different time point post the bacterial infection. The correlation values were presented as heatmaps of co-expression pattern. The clustering of PLCP genes were based on the correlation values. **a.** The co-expression pattern of 28 PLCP genes (and two housekeeping genes) and seven cystatin genes in the condition of *Pseudomonas syringae* strain ES4326 infection were presented as the heatmap. **b.** The co-expression pattern of 28 PLCP genes (and two housekeeping genes) and seven cystatin genes in the condition of *Pseudomonas syringae* strain ES4326 with the avirulence effector AvrRpt2 infection were presented as the heatmap.

Co-expression Analysis of Arabidopsis PLCPs and Cystatins under Wounding and Drought Stress

We generated PLCP-cystatin correlation heat maps by analyzing the wounding and drought stress microarray data set (TAIR Accession ID: 1007966439 and 1007966668).

The microarray data collected from leaf tissue of *Arabidopsis* under the wounding stress was used to generate heat maps (Figure 4-3a). The PLCPs could be roughly divided into four groups based on the correlation with cystatin genes. The top 15 PLCP genes show weak or modulate correlation with all cystatins. *RD21C*, *XCP2*, *RDL1* and *CEP3* are strongly correlated with *CYS1* or *CYS3*. *CTB3*, *RDL2*, *CTB2* and *RD19C* are strongly correlated with *CYS4* or *CYS5*. The rest PLCP genes (*RD21B*, *THI1*, *PAP1*, *PAP5* and *SAG12*) are strongly correlated with *CYS2* or *CYS7*.

The co-expression patterns of PLCP-cystatin under wounding stress (Figure 4-3a) are much different from the patterns under the disease condition (Figure 4-2a). For example, under disease conditions, the expression of *XCP2*, *AALP*, *ALP2* and *RDL3* are significantly correlated with at least three cystatin genes (*CYS3*, *CYS5* and *CYS6*) (Figure 4-2a). While in response to wounding stress, these four PLCP genes are only weakly correlated with most cystatin (Figure 4-3a). As another example, the expression of *SAG12* and *PAP5* is correlated with most cystatin genes except *CYS4* and *CYS6* under the wounding stress, while these two PLCP genes only showed weak correlations with all seven cystatin genes in response to the inoculation *of P. syringae* ES4326.

Mechanic wounding is a mimic of damage caused by insects or animals [2, 27]. Under wounding stress, protease inhibitors, including cystatins, might positively regulate of plant defense by suppressing *ex planta* proteases such as the digesting enzymes secreted by insects and nematodes [2]. While PLCPs, that have role of enhancing PCD and positively regulate defense to bacterial pathogens, could be negatively regulating plant defense to wounding or damage caused by inserts [1, 6]. The different co-expression patterns of PLCP-cystatins shown in Figure 4-2a and 4-3a might be consistent with this speculation.

We also analyzed the expression patterns of PLCP and cystatin genes using the microarray data collected from root tissue of *Arabidopsis* under the wounding stress. As shown in Figure 4-3b, the PLCPs could be roughly divided into three groups based on their correlation values. The top seven PLCPs showed strong correlation with *CYS2*, *CYS5* or *CYS6*. The last ten PLCPs showed strong correlation with *CYS1*, *CYS3*, *CYS4*, *CYS6* or *CYS7*. And the rest of PLCP genes only showed weak correlation with all cystatin genes. The PLCP-cystatin co-expression patterns in leaf tissue (Figure 4-3a) are much different from the patterns in root tissue (Figure 4-3b). For example, six PLCP genes in *Arabidopsis* shoot tissue (*RDL6, RD19B*, *XCP1*, *XBCP3*, *RDL3* and *CEP1*) are weakly correlated with all cystatin genes, while five different PLCPs (*RD21C*, *RD19C*, *CTB1*, *CEP3* and *PAP2*) in root tissue show weak correlation with all cystatins. Therefore, *Arabidopsis* might have two sets of PLCPs response to wounding stress in leaf and root tissue, respectively.

Many *Arabidopsis* PLCPs were reported to be associated with drought tolerance, and nearly half of them were named as Responsive to Desiccation (RD) or RD-liked genes [32]. In this study, we also analyzed the co-expression patterns of PLCP-cystatin of *Arabidopsis* under drought conditions (Figure 4-3c and 4-3d). We firstly analyzed the co-expression patterns of PLCPs and cystatins in *Arabidopsis* shoot tissue under drought stress. As shown in Figure 4-3c, the PLCPs could be roughly divided into two groups based on their correlation values. The top twelve PLCP genes were significantly correlated with *CYS5*, *CYS6* or *CYS7*. And the rest of PLCP genes only showed weak correlation with all cystatins. As shown in Figure 4-2a, when *Arabidopsis* plants infected with virulent bacteria, there are nine PLCP genes (*RD19A*, *XCP1*, *CPE1*, *RDL6*, *CTB3*, *RD19B*, *SAG12*, *RDL1* and *PAP5*) showed weak correlation with all cystatins. And six of them (*XCP1*, *CPE1*, *RD19B*, *SAG12*, *RDL1* and *PAP5*) show a similar co-expression pattern under the drought stress condition, where they weakly correlated with all cystatin genes. Therefore, *Arabidopsis* leaves response to drought stress and bacterial infection by using similar sets of PLCP and cystatin genes.

Five cystatin genes (*CYS3*, *CYS4*, *CYS5*, *CYS6*, and *CYS7*) showed strong correlation with most PLCP genes under the virulent bacterial infection condition,

while only *CYS5*, *CYS6* and *CYS7* showed strong correlation with most PLCP genes under drought condition. Therefore, the three cystatins (*CYS5*, *CYS6* and *CYS7*) might involve in regulating response to both drought stress and bacterial infection. Since both drought stress and bacterial pathogen can regulate stomatal closure, we speculate if these drought/disease shared PLCP-cystatins may have roles in regulating stomatal closure, which is worthy for further investigation in the future.

The expression of *CYS2* is weakly correlated with most PLCP genes (except *CEP1* and *PAP5*) (Figure 4-3c). Our previous report suggests the overexpressing of *CYS2* can increase stomatal opening (Wu et al, 2015). Interestingly, the expression of *CYS2* was suppressed under drought condition, which is usually triggering stomatal closing. Therefore, the *CYS2* expression pattern is consistent with its role in regulating stomatal closing. It will be interesting to test if the CYS2 overexpression plants are also more sensitive to drought stress.

We also analyzed the co-expression patterns of PLCP and cystatin genes in *Arabidopsis* root tissue under drought stress, and the result was presented in Figure 4-3d. Comparing to the co-expression patterns of PLCP and cystatin genes in shoot tissue under drought stress (Figure 4-3c), the overall co-expression values of all the PLCP-cystatin were reduced. In root tissue, the expression patterns of most PLCPs, are weakly or moderately correlated with all cystatins (Figure 4-3d), where most PLCPs are only correlated with a few cystatins in leave tissue (Figure 4-3c). Therefore, *Arabidopsis* might have different regulating strategies in root and shoot tissue under drought stress. Nevertheless, in root tissue, *CYS5*, *CYS6* and *CYS7* still showed strong or modulate correlation with most PLCPs (Figure 4-3d), which is consistent with the result in shoot tissue (Figure 4-3c). Therefore, these three cystatin genes (*CYS5*, *CYS6* and *CYS7*) might be involved a common signaling pathway in response to drought stress. While the expression of *CYS2* was slightly upregulated in root tissue, and it is significantly correlated with *XBCP3,* and moderately correlated with *CEP*, *AALP,* and *RD21B*. It is possible that *CYS2* is specifically involved in regulating stomatal aperture, which is more sensitive in shoot than root tissue under the drought stress condition.

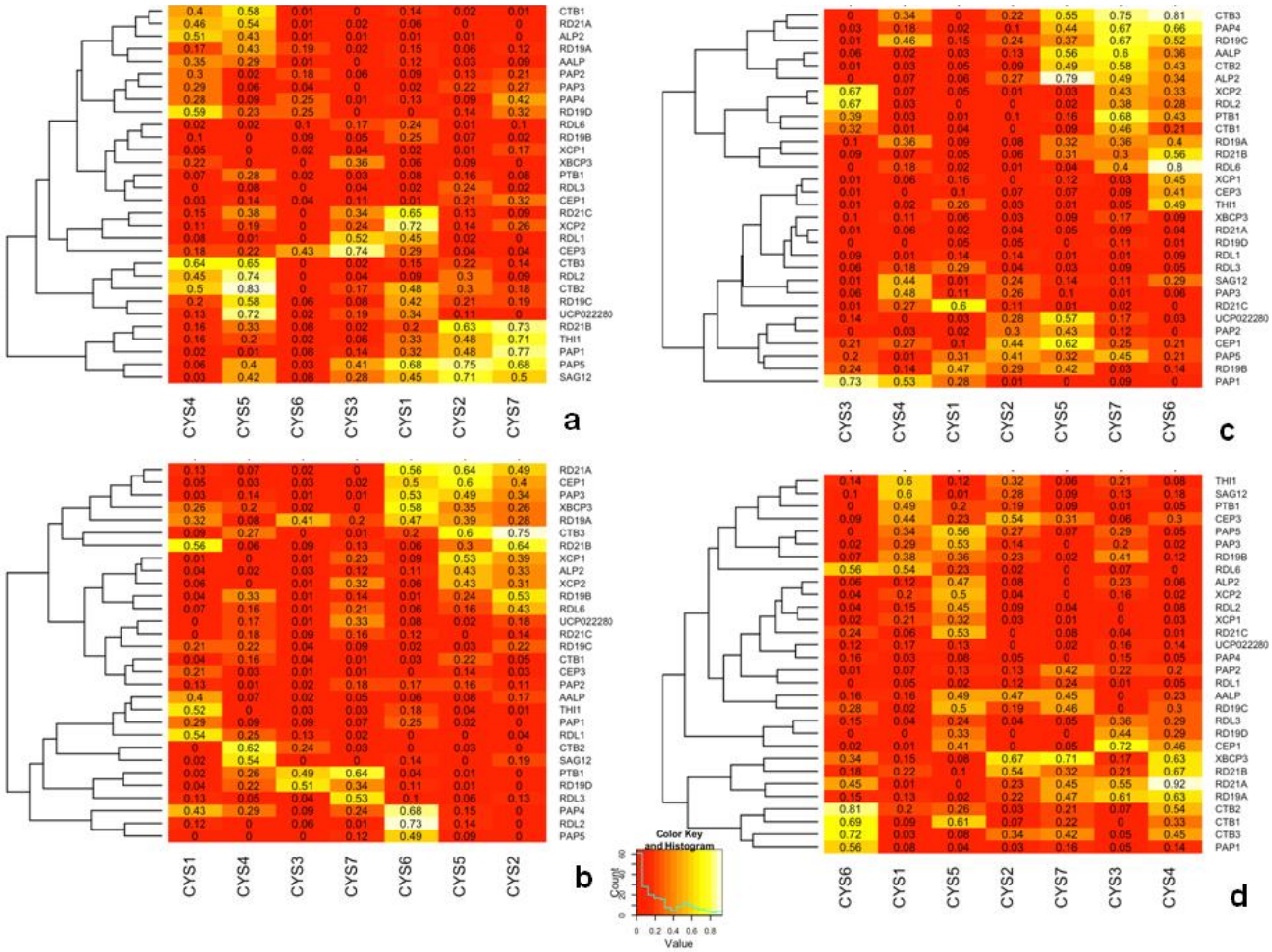


Figure 4-3| Co-expression patterns of *Arabidopsis* PLCPs and cystatins and clustering of PLCPs in wounding or drought stress condition

Gene co-expression patterns were generated from the AtGenExpress microarray data of *Arabidopsis thaliana* (Col-0) plants in response to wounding and drought stress (TAIR Accession ID: 1007966439 and 1007966668). The relative expression value was the difference of the expression value between bacteria infected treatment and mock treatment. The correlation value of PLCP genes (or housekeeping genes: *PTB1* and *UCP022280*) and cystatin genes were calculated based on the relative expression values of every gene at different time point post the wounding or drought stress treatment. The correlation values were presented as heatmaps of co-expression pattern. The clustering of PLCP genes were based on the correlation values. **a.** The co-expression pattern of 28 PLCP genes (and two housekeeping genes) and seven cystatin genes in *Arabidopsis* shoot tissue under wounding stress. **b.** The co-expression pattern of 28 PLCP genes (and two housekeeping genes) and seven cystatin genes in *Arabidopsis* root tissue under wounding stress. **c.** The co-expression pattern of 28 PLCP genes (and two

housekeeping genes) and seven cystatin genes in *Arabidopsis* shoot tissue under drought stress.**d.** The co-expression pattern of 28 PLCP genes (and two housekeeping genes) and seven cystatin genes in *Arabidopsis* root tissue under drought stress.

Evaluate the Enzyme Activities of Five Arabidopsis PLCPs and the Enzyme Inhibition Ability of Seven Cystatins

In addition to the expression analysis of PLCP and cystatin genes at transcription level, we also performed protease enzyme activity and cystatin-inhibition assays to identify the interaction relationship of *Arabidopsis* PLCPs and cystatins at protein level.

Five most abundant PLCPs (*AALP*, *RD19A*, *RD19C*, *RD21A* and *CTB3*) (Figure S4-1) were transiently expressed in *Nicotiana benthamiana* (*N. benthamiana*) and the crude protein extract was used for protease enzyme assays following a protocol as described previously [30]. As summarized in Figure 4-4, different cystatins have different inhibition specificity to a given PLCP. Therefore, though cystatins share conserved inhibition motif, they do evolved specificity against certain cysteine proteases. CYS1 significantly inhibit CTB3, RD19A and RD21A. CYS2 significantly inhibit AALP. CYS3 significantly inhibit RD19A and RD21A. CYS4 significantly inhibit CTB3, RD19A and RD21A. CYS5 significantly inhibit RD19A. CYS6 significantly inhibit RD19C, CTB3 and RD19A. CYS7 significantly inhibit RD19C and RD19A. AALP is the most abundant PLCP in *Arabidopsis* (Figure S4-1). In our inhibition assays, AALP enzyme activity cannot be significantly inhibited by any cystatins except CYS2, where it only have weakly inhibit activity against AALP. We speculate that AALP might serve as one of dominant PLCPs in *Arabidopsis* that involved in more general biological processes. While, RD19A can be significantly inhibit by almost all cystatins, except CYS2, which has less than 50% inhibition activity. Therefore, RD19A might be involved in more specialized signaling pathways in response to certain stress condition such as PCD, and its activity needs to be strictly regulated by different cystatins.

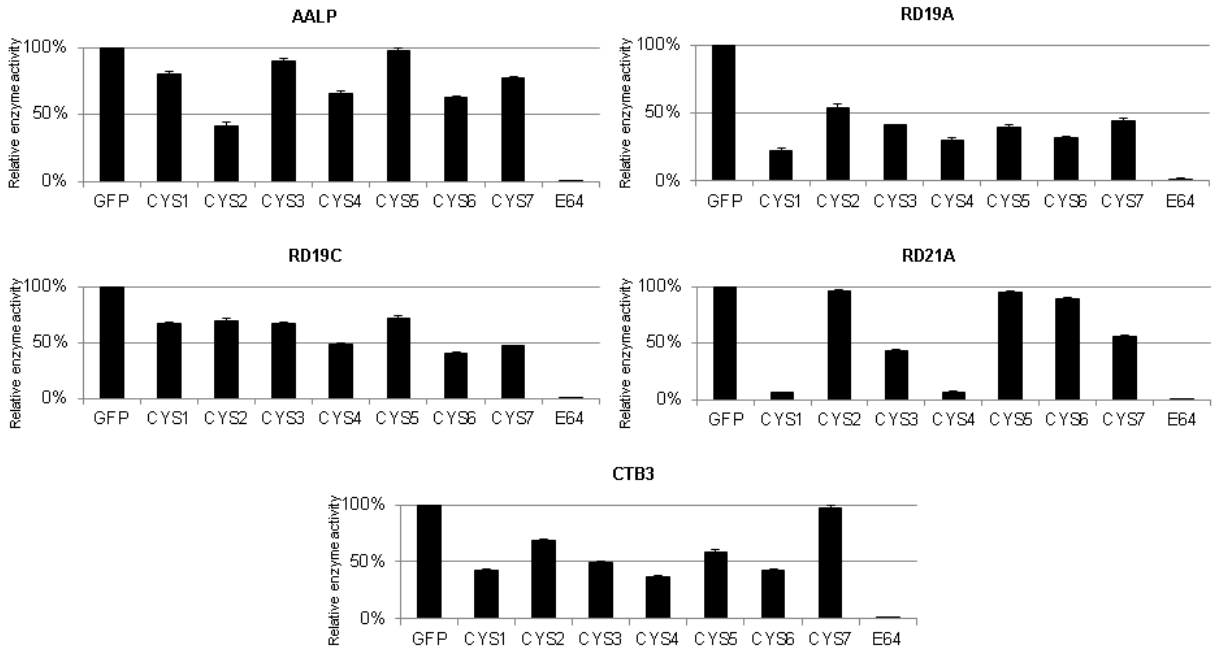


Figure 4-4| Enzyme activities of the five most abundant *Arabidopsis* PLCPs and the enzyme inhibition ability of seven cystatins

Five most abundant *Arabidopsis* PLCPs (Figure S1) were transiently expressed in *Nicotiana benthamiana* and the crude protein extract were used for enzyme activity test. The seven *Arabidopsis* cystatins (and the GFP control) were expressed and purified from *E. coli* based protein expression system. The artificial cysteine protease inhibitor E64 (10μM) inhibited all five PLCPs. The means ± s.e. of enzyme activities of different PLCP and cystatin combinations were presented as the histograms. The experiment was carried out in three independent replicates.

Prediction of Arabidopsis PLCP-Cystatin Interaction Network

We integrated the PLCP-cystatin gene co-expression and protease enzyme inhibition activity data, and predicted the PLCP-cystatin interplay network using the SVMs (Supporting Vector Machines)-based binary classification approach in the R package (e1701) [50]. The parameters of SVMs models were selected based on standard *k*-fold cross-validation [54]. The tuned SVMs models with selected parameters were tested to have 100% accuracy on classifying the test data set (35 pairs of PLCP-cystatin interactions used in the section of *in vitro* enzyme activity assay). The raw output of the classification on the 210 pairs of gene-gene interactions (28 PLCPs by seven cystatins, and 2 housekeeping genes by seven cystatins) was shown in the (Table S4-2 in Supplementary Information). The two housekeeping genes showed no interaction with any cystatins. And the output was translated into visual friendly network in the Cytoscape software (Figure 4-5a).

As shown in Figure 4-5a, CYS3 interacts with at least 12 most PLCPs (RD21A, RD21B, RDL2, RDL3, RDL6, CEP1, XCP1, PAP2, PAP3, PAP4, RD19A and CTB3), while CYS7 interacts with only five PLCPs (CEP3, PAP2, PAP4, RD19A and RD19C). There are ten PLCPs interacts with at least three cystatins (RD21A, RD21B, RDL3, CEP1, XCP1, PAP4, RD19A, RD19B, RD19C, and CTB3). Therefore, these ten PLCPs may have important functions in *Arabidopsis*, where their activities are tightly regulated by several different cystatins. Interestingly, in the RD19 family (Table S4-1), all three family members were belong to this group, which suggested the RD19 family was conserved and involved in regulating some essential biological activities in *Arabidopsis*.

The PLCP-cystatin interaction network provides a general view of the PLCP-cystatin interplay. When characterize biological function of a particular PLCP or cystatin gene, we may consider analyzing the co-related genes as a group in considering of gene redundancy problem. For example, in our previous study, the effector AvrRxo1 from *Xanthomonas oryzae pv. oryzicola* (*Xoc*) was identified to be able specifically induce the expression of *Arabidopsis CYS2*. Overexpression of *CYS2* increases stomata aperture size and suppresses plant immunity. As shown in Figure 4-2a and 4-3c, the expression of *CYS2* had strong correlation with *PAP4* when plants were infected with bacterial pathogen. However, the correlation is not significant when plant were treated with drought stress. As shown in Figure 4-5b, expression of *CYS2* was suppressed at the early stage of drought treatment, while the expression *PAP4* was significantly induced. Considering the overexpression of *CYS2* can increase stomata aperture size, we predict the overexpression or induced expression of *PAP4* may reduce the stomatal aperture size, which can prevent water loss as early response to drought stress.

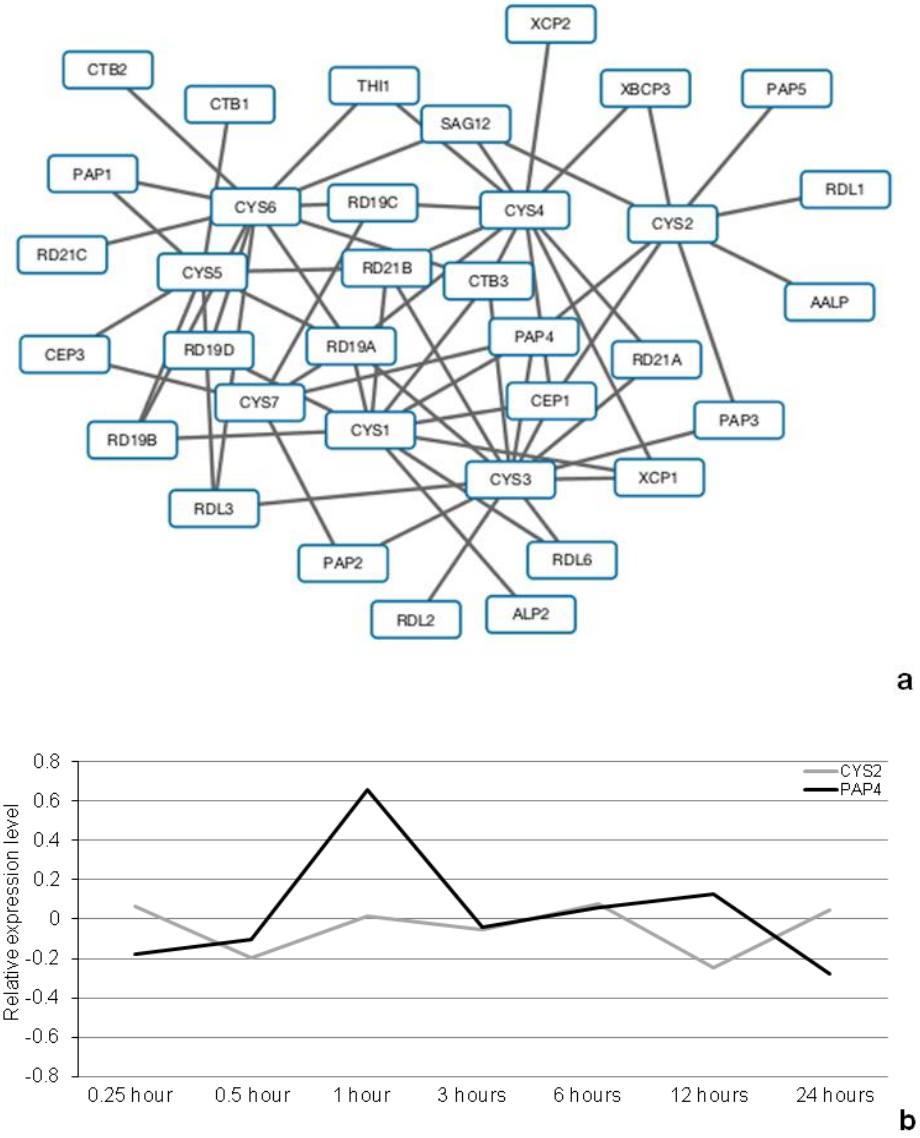


Figure 4-5| PLCP-cystatin interaction network and the expression profiles of *CYS2* and *PAP4* under drought stress

a. The interaction network of 28 PLCP genes and seven cystatin genes were generated based on the prediction result in Table S3, using the Cytospace program. **b.** Gene expression patterns of *CYS2* and *PAP4* were generated from the AtGenExpress microarray data of *Arabidopsis thaliana* (Col-0) plants in response to drought stress (TAIR Accession ID: 1007966668). The line chart represented the relative expression value of *CYS2* and *PAP4* at different time points post the drought stress treatment.

Conclusions and perspectives

In this study, we analyzed the microarray data of *Arabidopsis* plant in response to bacterial pathogen *P. syringae* ES4326, wounding, and drought stress, and predicted the relationship between 28 PLCP and seven cystatin genes in *Arabidopsis*. The heat map of gene co-expression summarizes the correlation between different PLCP and cystatin genes, which provides an overview for the interplay of PLCP-cystatin. In general, the expression of some cystatin genes is regulated by *P. syringae ES4326*, which have roles to regulate PLCP enzyme activity. Both PLCPs and cystatins have variable expression in different tissues (root vs shoot). By comparing the correlation patterns of PLCPs and cystatins under different conditions, we predicted several candidate PLCPs and cystatins, which may specifically involve in a particular signal transduction pathways in response abiotic or biotic stress.

In the future, to improve the interaction network, we need analyze more microarray (or RNA-seq) data, including other effector triggered immunity or other types of pathogen infection. It is known, many PLCPs are only expressed in seed or reproduction organs. Therefore, expression data from more tissue types are needed to be analyzed in the future. In addition to analysis of *Arabidopsis* PLCP and cystatin genes at the transcriptional level, it will be valuable to perform more comprehensive enzyme activity and inhibition assays between seven *Arabidopsis* cystatins and all thirty-one PLCPs at the protein levels. A comprehensive yeast two-hybrid assay or co-immunoprecipitation to validate the physical interactions between the PLCPs and cystatins can also provide insights about the interplay of PLCP-Cystatin. All these information will help us build more comprehensive interaction network of the PLCP-cystatin that will be useful for studying plant immunity and resistance to abiotic stress.

Material and methods

Acquiring the raw microarray data

The microarray data was downloaded from Atgenexpress database on TAIR website (https://www.*Arabidopsis*.org). Three data sets used in this study were: (1) "*Pseudomonas* half leaf injection" (TAIR Accession ID: 1008031517) [48]; (2) "Wounding stress time course" (TAIR Accession ID: 1007966439) [49]; (3) "Drought stress time course" (TAIR Accession ID: 1007966668) [49].

Microarray data processing

The raw microarray data was separated by treatments and time points, and organized into six data sets: (1) virulence bacteria challenge, (2) avirulence bemata challenge, (3) wounding stress in shoot tissue, (4) wounding stress in root tissue, (5) drought stress in shoot tissue, and (6) drought stress in root tissue. We then acquired the information for 28 PLCP genes, seven cystatin genes and two housekeeping genes (Table S4-1). Although 31 putative PLCP genes were predicted in *Arabidopsis* genome, three of them (*RDL4*, *RDL5*, and *CEP2*) have problems of missing or conflicting probes in the microarray data sets, therefore, they were excluded in this study. The averaged microarray signal values of three replicates were transformed into relative expression values by using the formula: Relative expression value = $\log_{10}$ (averaged microarray signal value of the experimental treatment) - $\log_{10}$ (averaged microarray signal value of the mock treatment). The relative expression values were used to calculate the correlation value between each pair of PLCP and cystatin, or the two housekeeping genes. The R software (Version 3.2) with the package of "heatmap.plus" was used for producing the heat maps [56].

Cloning of the PLCPs and cystatin genes

Five *Arabidopsis* PLCP genes (*AALP*, *RD19A*, *RD19C*, *RD21A* and *CTB3*) and seven *Arabidopsis* cystatin genes (*CYS1*, *CYS2*, *CYS4*, *CYS5*, *CYS6*, *and CYS7*) were PCR amplified from the cDNAs of *Arabidopsis* Col-0 using primers list in Table S4-3. The Cys3 was failed to be amplified from cDNAs, and it was synthesized by Genscript (Piscataway, NJ). All full length PLCP and CYS genes were cloned into pDonr207 vector (Invitrogen, Carlsbad, CA) and the correct sequence was confirmed by sequencing as the core facility of Virginia Bioinformatics Institute (Blacksburg, VA). The GFP gene was amplified from p519*gfp* [57]. All cloned genes were subcloned into the Gateway compatible binary vector pEarleygate102 [58] through LR® cloning following the manufacture's instruction (Invitrogen). Seven cystatin and the GFP genes were LR® cloned into a Gateway compatible expression vector pGEX4T-Des vector (Han et al, 2015).

Bacteria based protein expression and purification of cystatins

The expression plasmids were transformed into *E. coli* C41 cells (Lucigen, Middleton, WI) and grown overnight at 37 ˚C in 50 ml of LB medium supplemented with 100 mg/L carbenicillin. The bacteria culture was subcultured

into 1 liter of LB medium containing 100 mg/L carbenicillin till to the A600 is about 0.8 units. Isopropyl-1-thio-β-D30 galactopyranoside (IPTG) was added to a final concentration 0.5mM with subsequent incubation at 28˚C/220 rpm for 8 hours. The bacterial cells were harvested by centrifugation at 5,000 rpm for 10 min at 4˚C and the bacteria pellet was resuspended in 50 ml of binding buffer (50 mM Tris buffer, pH 7.5, 500 mM NaCl, 2 mM 2-mercaptoethanol, 1mM PMSF). The bacterial cells were then broken by incubating with 1 µg/ml lysozyme on ice for 1 hour, followed by three-times of sonication on ice at 50 % duty for 8 seconds/each time. The lysate was centrifuged at 12,000g for 20 min at 4˚C and the supernatant was gently collected and used for subsequent purification. Glutathione resin (GenScript) was equilibrated with binding buffer and added to the cleared supernatants. After incubating at 4˚C for 60 min, the mixture was loaded onto a 5 ml column (1cm x 20cm) (Bio-Rad, Hercules, CA). The column was washed three times with 50 ml of washing buffer (50 mM Tris-HCL buffer, pH 8.0, 300 mM NaCl, 1 mM DTT) to remove the nonspecifically bound proteins. The bound proteins were eluted with 10 x 1 ml of washing buffer containing 10 mM glutathione. The protein concentration was measured using NanoDrop® (Thermo Scientific, Hudson, NH), and protein purity was determined running SDS-PAGE gel and visualized by comas blue staining.

Agrobacterium-mediated transient assay and in vitro protease activity assay

To achieve *in planta* post translational activation of the PLCPs, the five PLCP genes were transiently expressed in *N. benthamiana* following the protocol described previously [59]. In brief, the binary constructs were transformed into *Agrobacterium* strain GV2260 by electroporation [60]. The *Agrobacterium* cells were resuspended in 10 mM $MgCl_2$ and adjusted to A600 = 0.4 units, and infiltrated into the leaf of *N. benthamiana* using a blunt end syringes without needles. The infiltrated plant were incubated at room temperature under continuous light. Leaf disks were collected at four days post infiltration using a cork borer (size 3) [30, 61]. Three leaf disks of each sample were pooled for protein extraction. The crude protein extract was used for *in vitro* protease activity assay as described by Mueller et al.[30]. In brief, the leaf discs were grounded in 100µL reaction buffer (50 mM sodium phosphate, pH 6.0, 600 mM NaCl, 4 mM EDTA, 2 mM DTT) on ice, and centrifuged at 12,000 rpm for 5 minutes at 4˚C. Forty µL of supernatant as the PLCP extracts was aliquot into two micro-tubes as the two replicates. The PLCP extracts were either added with 10 µM GFP or cystatin proteins. As control, 5 mM of E-64 (the cysteine protease inhibitor) (Sigma, St Louis, MI) was added to one set of PLCP samples. The reaction mixture was incubated at room temperature for 10 minutes, followed by adding 10 µL of 10

mM of Z-Phe-Arg- AMC (Sigma, St Louis, MI) as the protease substrate. The mixture was further incubated at room temperature for another 10 minutes. The reduction of fluorescence signal as the reflection of PLCP enzyme activity was measured using a Microplate Reader (BioTeck®, Synergy HT Multi-Mode) (BioTek, Winooski, VT). The values of fluorescence signals were used to compare the difference between PLCP samples incubated with either GFP or Cystatins, or E64.

Prediction of the PLCP-cystatin interaction network

The SVMS (support vector machines) package ('e1071') in R software (Version 3.2) were used for modelling the PLCP-cystatin interaction network [50]. Firstly, we separated the PLCP-cystatin co-expression correlation patterns into either biotic stress or abiotic stress conditions, and grouped them into a six-dimension data set. The dimensions were (1) "virulence, (2) avirulence", (3) "wound_shoot", (4) "wound_root", (5) "drought_shoot", and (6) "drought_root". Each pair of PLCP-cystatin could be considered as one data point in the six-dimension data matrix, and the SVMs can provide a proper binary classification model to divide all the points into two groups (either "related" or "nonrelated"). Secondly, we predicted the PLCP-cystatin interaction pairs based on *in vitro* enzyme activity assay as described previously. The cystatins that suppressed more than 50% reduction of the fluorescence signals were considered as the significant inhibitors to the corresponding PLCPs, and these pairs of cystatin and PLCPs were considered to be co-related to each other. We simplified the training input by considering the cystatin that reduced more than 50% PLCP enzyme activity as the inhibitor of this PLCP and marked this pair of PLCP-cystatin interaction as "1". And the unqualified pairs were marked as "0", to represent no co-relation. The 35 pairs (five PLCPs and seven cystatins) of PLCP-cystatin interaction values ("1" or "0") and their co-expression correlation values were used as the training data to select parameters for the SVMs models. The parameter was selected following the standard protocol of *k*-fold cross-validation [54]. We chose the parameter setting as: cost = 5, gamma = 1, which had the highest score ranking in the cross-validation, and it provided 100% accuracy on classifying the training data set. Finally, we processed the binary classification on the total data set of PLCP-cystatin co-expression correlations, which integrated all the six sets of PLCP-cystatin co-expression data we used in the former analysis. And the raw output of the SVMs modeling was used to generate the interaction network using the Cytoscape software (http://www.cytoscape.org/index.html).

Acknowledgements

Financial support was provided by grants from US National Science Foundation (NSF) (IOS-0845283 to BZ), Jeffery Trust Foundation (xx to BZ), and the Virginia Agricultural Experiment Station (VA135872). This project was also funded, in part, with an integrated, internal competitive grant from VAES, VCE, and the College of Agriculture and Life Sciences at Virginia Tech.

Supplementary information

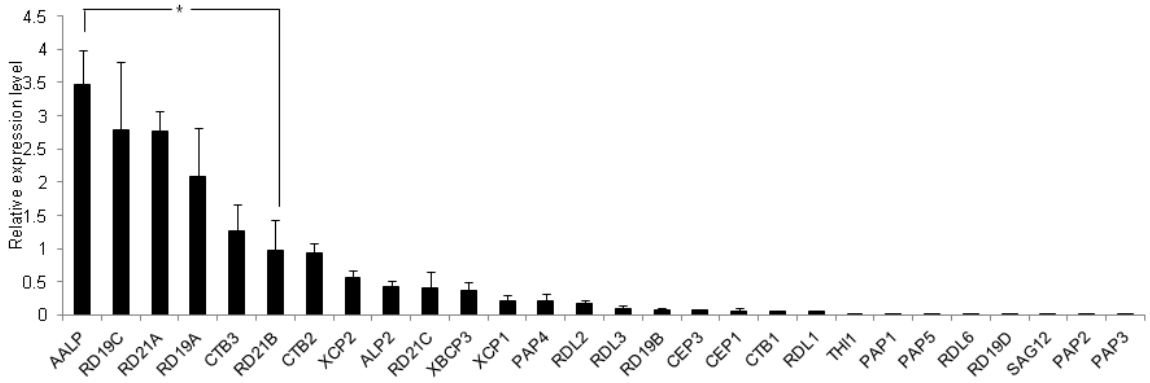


Figure S4-1| Expression levels of 28 PLCP genes in *Arabidopsis* shoot and root tissues

The expression values of 28 PLCP genes (Table S1) in *Arabidopsis* were generated from the AtGenExpress microarray data of *Arabidopsis thaliana* (Col-0) plants challenged with *Pseudomonas syringae* strain ES4326 (TAIR Accession ID: 1008031517), and the wounding and drought stress microarray data set (TAIR Accession ID: 1007966439 and 1007966668). The means ± s.e. of expression values of 28 PLCP genes in "non-treatment" conditions of every micro array data sets were presented. The asterisk represented significant difference (p-value< 0.05, Student's t-test).

Table S4-1| Gene names and TAIR IDs for PLCP, cystatin and housekeeping genes used in this study

The three PLCP genes have probe missing or conflicting problem are labeled in red.

31 PLCP genes		7 Cystatin genes		2 Housekeeping genes	
Gene name	Gene ID	Gene name	Gene ID	Gene name	Gene ID
RD21A	AT1G47128	CYS1	AT5G12140	UCP022280	AT2G45060
RD21B	AT5G43060	CYS2	AT2G31980	PTB1	AT3G01150
RD21C	AT3G19390	CYS3	AT2G40880		
RDL1	AT4G36880	CYS4	AT4G16500		
RDL2	AT3G19400	CYS5	AT5G47550		
RDL3	AT3G43960	CYS6	AT3G12490		
RDL6	AT4G23520	CYS7	AT5G05110		
CEP1	AT5G50260				
CEP3	AT3G48350				
XCP1	AT4G35350				
XCP2	AT1G20850				
XBCP3	AT1G09850				
THI1	AT1G06260				
SAG12	AT5G45890				
PAP1	AT2G34080				
PAP2	AT1G29090				
PAP3	AT1G29080				
PAP4	AT2G27420				
PAP5	AT3G49340				
RD19A	AT4G39090				
RD19B	AT2G21430				
RD19C	AT4G16190				
RD19D	AT3G54940				
AALP	AT5G60360				
ALP2	AT3G45310				
CTB1	AT1G02300				
CTB2	AT1G02305				
CTB3	AT4G01610				
RDL4	AT4G11310				
RDL5	AT4G11320				
CEP2	AT3G48340				

Table S4-2| Predicted relationship between seven cystatins and 28 PLCPs

The predicted relationship between seven cystatin genes and 28 PLCP genes (and two housekeeping control genes) were generated by SVMs modeling with the R package "e1071". As presented in the table, "0" represented "non-related" and "1" represented "related".

PLCP/ control genes	Cystatin genes						
	CYS1	CYS2	CYS3	CYS4	CYS5	CYS6	CYS7
PTB1	0	0	0	0	0	0	0
UCP022280	0	0	0	0	0	0	0
RD21A	1	0	1	1	0	0	0
RD21B	1	0	1	1	1	0	0
RD21C	0	0	0	0	0	1	0
RDL1	0	1	0	0	0	0	0
RDL2	0	0	1	0	0	0	0
RDL3	0	0	1	0	1	1	0
RDL6	1	0	1	0	0	0	0
CEP1	1	1	1	1	0	0	0
CEP3	0	0	0	0	1	0	1
XCP1	1	0	1	1	0	0	0
XCP2	0	0	0	1	0	0	0
XBCP3	0	1	0	1	0	0	0
THI1	0	0	0	1	0	1	0
SAG12	0	1	0	1	0	1	0
PAP1	0	0	0	0	1	1	0
PAP2	0	0	1	0	0	0	1
PAP3	0	1	1	0	0	0	0
PAP4	1	1	1	0	0	0	1
PAP5	0	1	0	0	0	0	0
RD19A	1	0	1	1	1	1	1
RD19B	1	0	0	0	1	1	0
RD19C	0	0	0	1	0	1	1
RD19D	1	0	0	0	0	1	0
AALP	0	1	0	0	0	0	0
ALP2	1	0	0	0	0	0	0
CTB1	0	0	0	0	1	0	0
CTB2	0	0	0	0	0	1	0
CTB3	1	0	1	1	0	1	0

Table S4-3| Sequences of primers used in this study

Restriction sites are labeled in red, and the sequences from cDNA are underlined.

PLCP	**Primers**	
AALP	AAAAAGCAGGCTtggatccATGTCTGCGAAAACAATCCTATC	Forward
	AGAAAGCTGGGTagtcgacTCTCAAGCCACAACGG	Reverse
RD19A	AAAAAGCAGGCTtggatccATGGATCGTCTTAAGCT	Forward
	AGAAAGCTGGGTagtcgacATGGGCGGTGGTTGAGAC	Reverse
RD19C	AAAAAGCAGGCTtggatccATGGATCGTGTGGTC	Forward
	AGAAAGCTGGGTagtcgacCTTGGGTGAGGTATG	Reverse
RD21A	AAAAAGCAGGCTtggatccATGGGGTTCCTTAAGCCAAC	Forward
	AGAAAGCTGGGTagtcgacGGCAATGTTCTTTCTGCCTTG	Reverse
CTB3	AAAAAGCAGGCTtggatccATGGCTGTTTACAAT	Forward
	AGAAAGCTGGGTagtcgacAACCGATGCAACCGG	Reverse
Cystatin	**Primers**	
CYS1	AAAAAGCAGGCTtggatccATGGCGGATCAACAAG	Forward
	AGAAAGCTGGGTagtcgacAACATCGTGAAGGTGG	Reverse
CYS2	CACCtctagaATGGCTACCATGTTGAAGG	Forward
	gtcgacGTAGACAGGACTGACAACAGGAG	Reverse
CYS3	AAAAAGCAGGCTtggatccATGGAATCAAAGACG	Forward
	AGAAAGCTGGGTagtcgacAGAAGATGATTCCTTG	Reverse
CYS4	AAAAAGCAGGCTtggatccATGATGATGAAGTCGTTG	Forward
	AGAAAGCTGGGTagtcgacTAACGCCTTGAAAGACTC	Reverse
CYS5	AAAAAGCAGGCTtggatccATGACTAGTAAGGTCGTC	Forward
	AGAAAGCTGGGTagtcgacAAGGAAGCGACCATTATTG	Reverse
CYS6	AAAAAGCAGGCTtggatccATGATGAGAAGCCG	Forward
	AGAAAGCTGGGTagtcgacGTCATGGTGTTGCTCC	Reverse
CYS7	AAAAAGCAGGCTtggatccATGGATATGAGACGTGC	Forward
	AGAAAGCTGGGTagtcgacCTCATACTTGCCGGTTTG	Reverse
Control gene	**Primers**	
GFP	CACCggatccATGAGTAAAGGAGAAGAACTTTTC	Forward
	gtcgacTCCGGACTTGTATAGTTCAT	Reverse

References

1. Van der Hoorn, R.A., *Plant proteases: from phenotypes to molecular mechanisms.* Annu. Rev. Plant Biol., 2008. **59**: p. 191-223.
2. Habib, H. and K.M. Fazili, *Plant protease inhibitors: a defense strategy in plants.* Biotechnology and Molecular Biology Review, 2007. **2**(3): p. 68-85.

3. Haq, S.K., S.M. Atif, and R.H. Khan, *Protein proteinase inhibitor genes in combat against insects, pests, and pathogens: natural and engineered phytoprotection.* Archives of Biochemistry and Biophysics, 2004. **431**(1): p. 145-159.
4. Supuran, C.T., A. Scozzafava, and B.W. Clare, *Bacterial protease inhibitors.* Med Res Rev, 2002. **22**(4): p. 329-72.
5. Valueva, T. and V. Mosolov, *Role of inhibitors of proteolytic enzymes in plant defense against phytopathogenic microorganisms.* Biochemistry (Moscow), 2004. **69**(11): p. 1305-1309.
6. Rawlings, N.D., D.P. Tolle, and A.J. Barrett, *MEROPS: the peptidase database.* Nucleic Acids Research, 2004. **32**(suppl 1): p. D160-D164.
7. Bozhkov, P.V., et al., *Cysteine protease mcII-Pa executes programmed cell death during plant embryogenesis.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(40): p. 14463-14468.
8. Reeves, P.H., et al., *early in short days 4, a mutation in Arabidopsis that causes early flowering and reduces the mRNA abundance of the floral repressor FLC.* Development, 2002. **129**(23): p. 5349-5361.
9. Diaz-Mendoza, M., et al., *C1A cysteine protease–cystatin interactions in leaf senescence.* Journal of experimental botany, 2014: p. eru043.
10. Rooney, H.C., et al., *Cladosporium Avr2 inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance.* Science, 2005. **308**(5729): p. 1783-1786.
11. Beers, E.P., B.J. Woffenden, and C. Zhao, *Plant proteolytic enzymes: possible roles during programmed cell death*, in *Programmed Cell Death in Higher Plants*. 2000, Springer. p. 155-171.
12. van der Hoorn, R.A., et al., *Activity profiling of papain-like cysteine proteases in plants.* Plant Physiol, 2004. **135**(3): p. 1170-8.
13. Bernoux, M., et al., *RD19, an Arabidopsis cysteine protease required for RRS1-R–mediated resistance, is relocalized to the nucleus by the Ralstonia solanacearum PopP2 effector.* The Plant Cell Online, 2008. **20**(8): p. 2252-2264.
14. Lid, S.E., et al., *The defective kernel 1 (dek1) gene required for aleurone cell development in the endosperm of maize grains encodes a membrane protein of the calpain gene superfamily.* Proceedings of the National Academy of Sciences, 2002. **99**(8): p. 5460-5465.
15. Feller, U., I. Anders, and K. Demirevska, *Degradation of rubisco and other chloroplast proteins under abiotic stress.* Gen Appl Plant Physiol, 2008. **34**(1-2): p. 5-18.

16. Doehlemann, G. and C. Hemetsberger, *Apoplastic immunity and its suppression by filamentous plant pathogens.* New Phytol, 2013. **198**(4): p. 1001-16.
17. Yamada, K., et al., *A Slow Maturation of a Cysteine Protease with a Granulin Domain in the Vacuoles of Senescing Arabidopsis Leaves.* Plant Physiology, 2001. **127**(4): p. 1626-1634.
18. Solomon, M., et al., *The involvement of cysteine proteases and protease inhibitor genes in the regulation of programmed cell death in plants.* The Plant Cell Online, 1999. **11**(3): p. 431-443.
19. Yamada, T., et al., *Purification of a novel type of SDS-dependent protease in maize using a monoclonal antibody.* Plant and cell physiology, 1998. **39**(1): p. 106-114.
20. Benchabane, M., et al., *Plant cystatins.* Biochimie, 2010. **92**(11): p. 1657-66.
21. Turk, V. and W. Bode, *The cystatins: protein inhibitors of cysteine proteinases.* FEBS letters, 1991. **285**(2): p. 213-219.
22. Coupe, S.A., et al., *Identification of dehydration-responsive cysteine proteases during post-harvest senescence of broccoli florets.* Journal of Experimental Botany, 2003. **54**(384): p. 1045-1056.
23. Krüger, J., et al., *A tomato cysteine protease required for Cf-2-dependent disease resistance and suppression of autonecrosis.* Science, 2002. **296**(5568): p. 744-747.
24. Gilroy, E.M., et al., *Involvement of cathepsin B in the plant disease resistance hypersensitive response.* The Plant Journal, 2007. **52**(1): p. 1-13.
25. Belenghi, B., et al., *AtCYS1, a cystatin from Arabidopsis thaliana, suppresses hypersensitive cell death.* European Journal of Biochemistry, 2003. **270**(12): p. 2593-2604.
26. Botella, M.A., et al., *Differential expression of soybean cysteine proteinase inhibitor genes during development and in response to wounding and methyl jasmonate.* Plant Physiology, 1996. **112**(3): p. 1201-1210.
27. Jones, J.D. and J.L. Dangl, *The plant immune system.* Nature, 2006. **444**(7117): p. 323-9.
28. Rooney, H.C., et al., *Cladosporium Avr2 inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance.* Science, 2005. **308**(5729): p. 1783-6.
29. Shindo, T., et al., *A role in immunity for Arabidopsis cysteine protease RD21, the ortholog of the tomato immune protease C14.* PloS one, 2012. **7**(1): p. e29317.
30. Mueller, A.N., et al., *Compatibility in the Ustilago maydis-maize interaction requires inhibition of host cysteine proteases by the fungal effector Pit2.* PLoS Pathog, 2013. **9**(2): p. e1003177.

31. van der Linde, K., et al., *A maize cystatin suppresses host immunity by inhibiting apoplastic cysteine proteases.* The Plant Cell Online, 2012. **24**(3): p. 1285-1300.
32. Yamaguchi-Shinozaki, K., et al., *Molecular Cloning and Characterization of 9 cDNAs for Genes That Are Responsive to Desiccation in Arabidopsis thaliana: SequenceAnalysis of One cDNA Clone That Encodes a Putative Transmembrane Channel Protein.* Plant and Cell Physiology, 1992. **33**(3): p. 217-224.
33. Seki, M., et al., *Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray.* The Plant Journal, 2002. **31**(3): p. 279-292.
34. Koizumi, M., et al., *Structure and expression of two genes that encode distinct drought-inducible cysteine proteinases in Arabidopsis thaliana.* Gene, 1993. **129**(2): p. 175-182.
35. Kim, J.H. and W.T. Kim, *The Arabidopsis RING E3 ubiquitin ligase AtAIRP3/LOG2 participates in positive regulation of high-salt and drought stress responses.* Plant Physiol, 2013. **162**(3): p. 1733-49.
36. Simova-Stoilova, L., et al., *Proteolytic activity and cysteine protease expression in wheat leaves under severe soil drought and recovery.* Plant Physiol Biochem, 2010. **48**(2-3): p. 200-6.
37. Harrak, H., et al., *Isolation and characterization of a gene encoding a drought-induced cysteine protease in tomato (Lycopersicon esculentum).* Genome, 2001. **44**(3): p. 368-374.
38. Quain, M.D., et al., *Ectopic phytocystatin expression leads to enhanced drought stress tolerance in soybean (Glycine max) and Arabidopsis thaliana through effects on strigolactone pathways and can also result in improved seed traits.* Plant Biotechnology Journal, 2014. **12**(7): p. 903-913.
39. Massonneau, A., et al., *Maize cystatins respond to developmental cues, cold stress and drought.* Biochim Biophys Acta, 2005. **1729**(3): p. 186-99.
40. Fan, L.M., Z. Zhao, and S.M. Assmann, *Guard cells: a dynamic signaling model.* Curr Opin Plant Biol, 2004. **7**(5): p. 537-46.
41. Schroeder, J.I., et al., *Guard cell signal transduction.* Annual review of plant biology, 2001. **52**(1): p. 627-658.
42. Melotto, M., et al., *Plant stomata function in innate immunity against bacterial invasion.* Cell, 2006. **126**(5): p. 969-80.
43. Atkinson, N.J. and P.E. Urwin, *The interaction of plant biotic and abiotic stresses: from genes to the field.* Journal of Experimental Botany, 2012.
44. Atias, O., B. Chor, and D.A. Chamovitz, *Large-scale analysis of Arabidopsis transcription reveals a basal co-regulation network.* Bmc Systems Biology, 2009. **3**.

45. Leal, L.G., et al., *Identification of immunity-related genes in Arabidopsis and cassava using genomic data.* Genomics Proteomics Bioinformatics, 2013. **11**(6): p. 345-53.
46. Cortes, C. and V. Vapnik, *Support-vector networks.* Machine learning, 1995. **20**(3): p. 273-297.
47. Brown, M.P., et al., *Knowledge-based analysis of microarray gene expression data by using support vector machines.* Proceedings of the National Academy of Sciences, 2000. **97**(1): p. 262-267.
48. Wang, D., et al., *Induction of protein secretory pathway is required for systemic acquired resistance.* Science, 2005. **308**(5724): p. 1036-1040.
49. Kilian, J., et al., *The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses.* The Plant Journal, 2007. **50**(2): p. 347-363.
50. Meyer, D. and F.T. Wien, *Support vector machines.* The Interface to libsvm in package e1071, 2014.
51. Coaker, G., A. Falick, and B. Staskawicz, *Activation of a phytopathogenic bacterial effector protein by a eukaryotic cyclophilin.* Science, 2005. **308**(5721): p. 548-50.
52. Dancey, C.P. and J. Reidy, *Statistics without maths for psychology*. 2007: Pearson Education.
53. Kunkel, B.N., et al., *RPS2, an Arabidopsis disease resistance locus specifying recognition of Pseudomonas syringae strains expressing the avirulence gene avrRpt2.* The Plant Cell Online, 1993. **5**(8): p. 865-875.
54. Hsu, C.-W., C.-C. Chang, and C.-J. Lin, *A practical guide to support vector classification*. 2003.
55. McLellan, H., et al., *Functional redundancy in the Arabidopsis Cathepsin B gene family contributes to basal defence, the hypersensitive response and senescence.* New Phytol, 2009. **183**(2): p. 408-18.
56. Day, A., *heatmap. plus: Heatmap with more sensible behavior.* R package version, 2012. **1**.
57. Matthysse, A.G., et al., *Construction of GFP vectors for use in Gram-negative bacteria other than Escherichia coli.* FEMS Microbiology Letters, 1996. **145**(1): p. 87-94.
58. Earley, K.W., et al., *Gateway-compatible vectors for plant functional genomics and proteomics.* Plant J, 2006. **45**(4): p. 616-29.
59. Zhao, B., et al., *Computational and biochemical analysis of the Xanthomonas effector AvrBs2 and its role in the modulation of Xanthomonas type three effector delivery.* PLoS Pathog, 2011. **7**(12): p. e1002408.

60. Traore, S.M. and B. Zhao, *A novel Gateway(R)-compatible binary vector allows direct selection of recombinant clones in Agrobacterium tumefaciens.* Plant Methods, 2011. **7**(1): p. 42.
61. Starkey, M. and L.G. Rahme, *Modeling Pseudomonas aeruginosa pathogenesis in plant hosts.* Nat Protoc, 2009. **4**(2): p. 117-24.