Understanding Plant Pathosystems in Wild Relatives of Cultivated Crop Plants

Michael Gerald Fedkenheuer

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

Plant Pathology, Physiology, and Weed Science

John M. McDowell

Boris A. Vinatzer

Mohammad A. Saghai-Maroof

Xiaofeng Wang

06/06/2016

Blacksburg, VA

Keywords: Glycine max, Glycine soja, Wild Species, Transcriptomics, Oomycete(s)

Understanding Plant Pathosystems in Wild Species of Cultivated Crop Plants

Michael Gerald Fedkenheuer

ABSTRACT

As the global population rises, the demand for food increases which underscores a need for improvement in food security. Disease pressures are a major concern surrounding sustainable agriculture. Static crop populations, containing little to no genetic diversity, are vulnerable to diverse pathogen populations. Wild relatives of crop plants are a reservoir for new disease resistance traits that can be introgressed into cultivated crops. The identification of novel disease resistance is of paramount importance because pathogen co-evolution is not only defeating current resistance genes (*R* genes) but chemical controls as well. *Phytophthora sojae* (*P. sojae*), the causal agent of *Phytophthora* root and stem rot disease, reduces soybean harvests worldwide. We developed an approach to screen for new *R* genes that recognize core effectors from *P. sojae*. We expect *R* genes identified by these screens to be durable because *P. sojae* requires core effectors for virulence. We utilized effector-based screening to probe *Glycine soja* germplasm with core RXLR effectors from *P. sojae* to search for novel *R* genes. We developed segregating populations from crosses of *P. sojae* resistant *G. soja* germplasm with susceptible *G. max* cultivar Williams to determine inheritance of potential *R* genes in germplasm that responded to core effectors. We are using marker-assisted breeding to map disease resistance traits in recombinant inbred (RI) lines. To better understand pathosystems, we examined host resistance and susceptibility using bioinformatics. We analyzed the interaction between *Arabidopsis thaliana* ecotype Col-0 and *Hyaloperonospora arabidopsidis* isolate Emwa1 using a publicly available RNA

time-course experiment. We describe a new algorithm to sort genes into time-point specific clusters using activation and repression parameters. Gene ontology annotations were used to identify defense genes with unique expression profiles, and *A. thaliana* null mutants for these genes were significantly more susceptible to Emwa1 than wild-type. We plan to use these tools to rapidly identify and guide introgression of durable disease resistance into crop species.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank the many people who have helped me develop scientifically at Virginia Tech. I would like to acknowledge funding sources which include USDA-NIFA, NSF, and the TPS program at Virginia Tech. I would like to thank the scientific advisory board whom advised our USDA funded work and provided necessary technical advice.

I would like to specifically acknowledge Kevin E. Fedkenheuer, Colin L. Davis, and John M. McDowell for their leadership roles in many projects. The work would not be possible without their contributions. I would like to thank Joel L. Shuman for his help with disease assay, travel, and greenhouse expertise. I would like to remember and thank Bhadra Gunesekera for her tireless help in making the lab run smoothly. I would like to recognize M. A. Saghai Maroof for his breeding expertise and for his collaborative spirit. I would like to thank Brett M. Tyler for collaboration and input given on a weekly basis. Lastly, I would like to thank the committee members and Dr. Song Li who have reviewed my work.

ATTRIUBTIONS

Projects Managed and Maintained by Graduate Students:

Kevin E Fedkenheuer[1]

Michael G Fedkenheuer[1]

Colin L. Davis[2]

Projects Designed and Advised by PhD Advisors:

John M. McDowell[1]

M. A. Saghai Maroof[2]

Brett M. Tyler[3]

[1]Department of Plant Pathology, Physiology, and Weed Science, Virginia Polytechnic Institute, Blacksburg, VA

[2]Department of Crop and Soil Environmental Sciences, Blacksburg, VA

[3]Department of Botany and Plant Pathology, Corvallis, OR

Chapter 1

Review: Introgression of Disease Resistance from Wild Species to Cultivated Crops

Michael G. Fedkenheuer[1] and John M. McDowell[1]

[1]Department of Plant Pathology, Physiology, and Weed Science, Virginia Tech.

Blacksburg, VA

[*]for correspondence, johnmcd@vt.edu

In preparation for submission to Theoretical and Applied Genetics

Author Contributions: MF developed and wrote the manuscript with advice from JM.

Abstract

Disease pressures are a major concern surrounding sustainable agriculture as the demand for food is driven upwards by global population growth. Diverse pathogen populations, capable of rapid evolution, pose a threat to cultivars with little genetic diversity or adaptive variation. Wild relatives of cultivated crop species present an important opportunity to acquire new traits for disease resistance that can be introgressed into cultivated crop species. Identifying novel disease resistance genes can counteract the pathogen co-evolution that is defeating current resistance genes and chemical controls. Many useful resistance traits have been identified in wild species. This review will describe how disease resistance can be mapped and introgressed from wild to cultivated crop species. Advancements in molecular techniques have accelerated this process as genotyping technologies have become faster, more accurate, and less costly. Marker assisted selection can be used to guide breeding for simple or complex disease resistance traits in segregating populations. As a result, wild species are becoming increasingly accessible sources of novel resistance against persistent, adaptable, and destructive plant pathogens.

Disease Pressures in Agriculture

Crop Production is Threatened by Diverse Plant Pathogens

Sustainable agriculture is essential as we move through the 21st Century. Agricultural producers strive to improve yield year after year to provide affordable food for our rapidly growing global population. As we approach the maximum limitations of yield in a conventional sense (fruit size, oil content, etc.), attention is being shifted towards improving other aspects of crop fitness such as disease resistance. It is conservatively estimated that we lose 20-40% of agricultural productivity as a direct consequence of pests and weeds [1]. Thus, reduction of these losses can significantly boost productivity.

Genetic resistance and chemical treatments have been used effectively for decades and remain as cornerstones of integrated disease management strategies. However, diverse pathogens are emerging that can overcome chemical controls and resistance (*R*) genes [2]. Disease management problems will further exacerbated by climate change. Pathogen populations can be influenced by environmental factors. Consistent periods of heavy rainfall or flooding commonly exacerbate disease problems, while periods of dry weather often have an opposite effect [3]. In 1999, Coakley et al. suggested that global climate change could alter rates of pathogen development and host resistances [4]. We can expect a shift in geographic pathogen and crop distribution that will severely test the efficacy of current disease management strategies [4]. These challenges underscore a need for tools to reduce disease losses and mitigate food security concerns. This review

will focus on the improvement of crops though introgression of traits from wild progenitors.

Disease Susceptibility in Crop Plants

Cultivated crops planted today are typically elite cultivars encompassing a relatively low amount of genetic diversity within the species. Until the last century, monoculture was only practiced in a few plant species such as wheat, maize, and rice [5]. Due to its convenience, monoculture practices have been extended to other species [5]. These practices decrease genetic diversity in crop species, leaving cultivars susceptible to severe disease outbreaks. Monoculture also promotes the overuses of pesticides and fertilizers that can destroy natural ecosystems [5, 6].

Resistance in Wild Species

The most primitive form of plant breeding was the domestication of wild species for cultivation [7]. During this process, many alleles providing variability were lost [7]. Breeders are now turning to these species' wild progenitors for a wide variety of agronomic traits. During domestication, many deleterious alleles were selected that were linked to more traditional agronomic traits such as growth and yield [8]. Many traits such as disease resistance against certain pathogens were lost. We will broadly refer to disease resistance genes, alleles, and loci as resistance traits throughout this review.

Important disease resistance traits are available in wild species of cultivated crop plants. In this review, we will discuss introgression of disease resistance traits from wild species to major crop plants with a focus on potential management strategies. Major crop categories as defined by the USDA are Corn, Cotton and Wool, Fruit and Tree Nuts, Rice, Soybean and Oil Crops, Sugars and Sweeteners, Vegetables and Pulses, and Wheat

[9]. Working with wild species presents unique challenges but significant opportunities for identifying new and untapped gene pools.

Oomycete Plant Pathogens

Oomycetes, commonly referred to as water molds, cause numerous diseases of agronomic importance. The class *Oomycota* is in the kingdom *Chromalveolata* and belongs to the phylum *Heterokontophyta* [10] . Pathogenic oomycetes are located in the order *Leptomitales*, more specifically in the families *Aphanomycetes*, *Pythiaceae*, *Albuginaceae*, and *Peronosporaceae* species [11-13]. In 2015, oomycetes and fungi were described as the biggest threats to modern agriculture [14]. This point is illustrated historically by the Irish Potato famine caused by *Phytophthora infestans* (*P. infestans*) [15]. This oomycete plant pathogen is conservatively estimated to cause 5 to 6 billion dollars in global potato losses each year [16]. Sexual recombination is uncommon in *P. infestans*; however when it does occur the resulting lineages can rapidly displace old clonal lineages [17]. This phenomenon was observed in European pathogen populations when compatible mating types apparently migrated from Mexico in the 1980s [17]. Another destructive oomycete disease is soybean root and stem rot, caused by the oomycete pathogen *Phytophthora sojae* (*P. sojae*) [18]. This disease was estimated to annually suppress 1% of global soybean yield [18, 19].

Unlike *P. sojae* and *P. infestans* which cause disease on specific cultivated species, some oomycete pathogens have broad host ranges. For example, *Pseudoperonospora cubensis* (*P. cubensis*) causes disease on all cultivated cucurbit species [20]. *P. cubensis* isolates were recently described that break known resistance in cucumber [20, 21]. Prior to 2003, the *dm-1* resistance gene provided broad spectrum

resistance to *P. cubensis*; however over the last ten years sexual recombination events have continued to create resistance-breaking isolates [20]. When resistance was broken in 2004, Colucci et al. reported yield losses of 40% for US farmers [22]. Due to this disease, future cucurbit production is in jeopardy [20, 21]. Another oomycete pathogen with a broad host range is *Phytophthora cinnamomi* (*P. cinnamomi*) which has been reported to impact over 2,000 plant species [23]. This pathogen causes damage to natural ecosystems in regions that receive significant rainfall [23]. It is clear that oomycete pathogens are a major food security threat due to their dynamic ability to adapt through sexual recombination and cause disease epidemics among crop plants.

Obligate Oomycete Pathogens

Life Cycle

Many oomycete pathogens require a plant host for growth and reproduction. During infection, asexual spores land on the leaf surface and germinate producing a germ tube and appressorium [24]. Spores from obligate oomycete pathogens commonly lack flagella. Penetration hyphae extend from the appressorium to invade the leaf intercellularly. As hyphae extend through the plant tissue, the cells are penetrated by structures known as the haustoria [24]. The interface that separates the plant cell membrane and the haustorial cell wall is referred to as the extrahaustorial matrix (EHM) [25]. Haustoria are thought of as feeding structures because they ultimately function to draw nutrients from the cell for reproduction; however these structures serve another primary purpose. At the EHM, effector proteins are secreted by the pathogen to subvert and suppress plant immunity [24]. After successful colonization, sporangiophore are produced on the leaf's surface [26]. Sporangiophore production is triggered by increased humidity and spores are dispersed by wind or water splash to infect susceptible plants nearby. Sexual spores called oospores are produced later in infection or in response to stress [26]. These thick walled structures are very durable and can survive for many years dormant in plant material.

Peronosporaceae *Pathogens*

The two most destructive *Peronosporaceae* pathogens cause downy mildew disease on grapes and cucurbit species. One is the aforementioned *Pseudoperonospora cubensis* that has been ravaging cucurbit production in the USA since host resistance was

broken in 2004 [27]. *P. cubensis* has a diverse genetic structure that makes management nearly impossible [28]. The second example is grape downy mildew, caused by *Plasmopara viticola*, which causes significant global economic losses [29, 30]. This pathogen thrives in warm, humid climates and infects the *Vitaceae* family [31]. *P. viticola* infects leaves and fruit clusters and is most pathogenic on *Vitis vinifera*, the most commonly used grapevine in the global wine industry [32]. Producers rely heavily on oomycete-targeted fungicide treatments to control seasonal downy mildew infections. However, increasing attention is being focused on identifying natural sources of resistance in wild *Vitis* species [32].

In addition to grape and cucumber, most lettuce cultivars are susceptible to downy mildew. *Bremia lactucae* (*B. lactucae*), the causal agent of lettuce downy mildew, is especially destructive in California which contains the highest pathotype diversity [33]. Pathotypes of *B. lactucae* have also been reported to infect artichoke, cornflower, and strawflower [34]. Downy mildew disease has also been described on a number of specialty crop species. *Peronospora belbarhii*, the causal agent of basil downy mildew, re-emerged in global populations in 2001 and within six years made its way into the USA [35]. Basil downy mildew was first observed in the state of Florida in 2007 and since then has spread through most of the country including Hawaii [35]. *P. belbarhii* has been responsible for over ten million dollars in crop losses since its introduction into the US [35]. Sunflower downy mildew is an emerging pathogen which causes significant crop losses and is caused by the obligate pathogen *Plasmopara halstedii* [36]. Although resistance exists in commonly used cultivars, reports of resistance breaking isolates

indicate that it is just a matter of time before pathogen populations are restructured to overcome these resistance mechanisms [36].

Albuginaceae *Pathogens*

The *Albugo* genus causes white blister rust (WBR) disease on many crop plants. *Albugo candida* has been reported to cause disease on mustard plants like crucifers, cabbage, and vegetables and on radish varieties [37]. WBR disease in sunflower is caused by *Pustula helianthicola*. In 1995, van Wyk et al reported that this disease suppresses yield up to 80% of sunflower production, primarily due to stem fractures [38]. The severe foliar symptoms that accompany this disease have negatively impacted ornamental sunflowers [39]. WBR foliar symptoms have suppressed yield in the growing spinach markets as well [40].

Non-Obligate Oomycete Pathogens

Life Cycle

Many aspects of oomycete biology are similar; however lifecycles can be different depending on infection strategy. Non-obligate oomycete pathogens generally infect roots or leaves with hyphae or with motile zoospores [41]. Oomycete zoospores are unique because they contain an anterior "tinsel" flagellum and a posterior "whiplash" flagellum [10]. This morphological trait distinguishes oomycetes from fungi [10]. Zoospores encyst and penetrate soft root tissue with intercellular hyphae [41]. These pathogens produce haustoria as well. During a successful infection, plant roots and stems are rotted as the pathogen feeds on plant cells to extract nutrients for reproduction [41]. Sporangia often contain motile zoospores, while robust oospores are generated inside plant tissue for overwintering [41]. Oospores are sexual structures and represent a source of new genetic diversity, likely giving rise to diverse isolates in future seasons [41]. Many of these pathogens transition from an early biotrophic phase to necrotrophy late in infection, a characteristic which classifies them as "hemi-biotrophic" pathogens [42]. Pathogens that do not have a descernable biotrophic phase are classified as nectrophs. Additioanlly, it is common for necrotrphoic pathogens to lack haustoria [42].

Pythiaceae *Pathogens*

Phytophthora species can cause significant disease on crop plants. *Phytophthora* pathogens thrive in wet and humid conditions, a characteristic shared by most oomycete pathogens. Among the most destructive pathogens in this family are *P. infestans and P. sojae*. *P. infestans'* genetic diversity is in a constant state of flux which makes

determining effective management difficult [15]. Similarly, *P. sojae* is a difficult pathogen to control due to isolate complexity. Diverse *P. sojae* isolates have been described that overcome all commercially deployed *R* genes in soybean [43]. *P. sojae* optimally infects soybean seedlings between 25°C to 30°C causing pre- and post-emergence damping off [43]. In mature soybean plants this disease causes root and stem rot [43].

An emerging pathogen, *Phytophthora capsici,* causes significant disease on peppers and can infect cucurbits, eggplant, snap peas, and lima beans [44]. The broad host range of this pathogen makes it difficult to manage but offers a unique opportunity for studying host range and virulence in oomycete pathogens [44]. *Phytophthora cinnamomi* also inflicts damage on a wide range of hosts, causing damage to many nursery crops and ornamentals [45]. *Phytophthora ramorum* causes sudden oak death, a major problem for California forests and can also infect over 100 species of forest trees [46]. *Phytophthora* pathogens are primarily treated with the oomycete-targeted fungicides metalaxyl and mefenoxam [43, 47]. Other less common but effective oomycete-targeted fungicides include azoxystrobin, cymoxanil, fluopicolide, mandipropamid, and cyazfamid [47].

Necrotrophic oomycete pathogens are predominantly represented by *Pythium* species. These pathogens cause root rot and seedling disease on a wide variety of crops in the USA [48]. *Pythium ultimum* (*P. ultimum) and P. irregulare* cause damping off in wheat seedlings, while rice seedlings are primarily infected by *P. irregulare*, *P. arrhenomanes*, and *P. graminicola* causing uneven stand height [48]. *P. ultimum* and *P. aphanidermatum* cause *Pythium* leak disease, a common postharvest disease in potato

tubers [49]. Changes in cultural practices such as earlier planting, minimum tillage, and fungicide use have increased seedling establishment problems in soybean and corn [50, 51]. Insensitivity to commonly used fungicides has been reported against mefenoxam, azoxystrobin, trifloxystrobin, and captan [50]. Often times, management of *Pythium* species is not effective when a single chemical treatments is employed. This is largely due to the fact that repeated use of a fungicides can promote the emergence of fungicide insensitivity [51].

Aphanomycetes *Pathogens*

There are only two major pathogens of agronomic importance in the *Aphanomyces* genus. *Aphanomyces euteiches* has been reported to cause severe root rot in cultivated peas [52]. This pathogen thrives in cool and damp conditions, and it can persist in the soil for over a decade [52]. This disease has been reported to suppress pea yield up to 80% of production [53]. *Aphanomyces cochlioides* is the causal agent of blackroot disease on sugarbeet [52]. *A. cohlioides* has been reported to increase in response to continual sugarbeet cropping [52].

The Plant Immune System

In the next section I introduce key concepts about the molecular function of the plant immune system, to set the stage for discussion of mining new *R* genes from wild relatives of crops. I will not discuss physical barriers to infection, but rather focus on two layers of inducible defense that provide resistance to pathogens that are adapted to crop hosts [54].

Pattern-Triggered Immunity

The first layer of plant immunity can be activated by recognition of so-called "pathogen-associated molecular patterns" (PAMPs), which comprise a diverse collection of pathogen-derived molecules. Most commonly, these molecules are conserved over broad taxonomic distances, ranging from genera to kingdoms (see examples below). PAMPS are recognized by plant pattern-recognition receptors (PRRs) that typically bind the PAMP on the exterior of the plant cell and in turn activate a generic host immune response, referred to as pattern-triggered immunity (PTI) [54, 55]. PRR proteins are organized into two groups, receptor-like kinases (RLKs) and receptor-like proteins, both of which contain an extracellular domain for ligand detection and a transmembrane domain [56]. RLKs, but not RLPs, contain an intracellular kinase domain as well [56]. The best-studied examples of PAMPs from plant pathogens are fragment of the bacterial flagellin protein [57, 58]. For example, in *A. thaliana*, the PRR, FLS2, recognizes a conserved 22 amino acid peptide from bacterial flagellum [58]. The first oomycete PAMP was isolated from the soybean pathogen, *P. sojae* [59]. In early infection, β-1,3 glucanase degrades the pathogen cell wall [60]. This releases β- glucans that stimulate the

production of antimicrobial compounds called phytoalexins [59, 60]. While PTI can effectively limit growth of many pathogen species, more specialized or so-called "adapted pathogens" can overcome this defense through deployment of secreted effector proteins (described below).

Effector-Triggered Immunity

The second mechanism by which plants sense pathogen attack is through recognition of pathogen effector proteins that are secreted to the interior of plant cells [54, 61]. This Effector Triggered Immunity (ETI) occurs when a plant resistance protein (*R* protein) recognizes a specific pathogen effector [54, 61]. Most *R* genes belong to a family of genes known as NBS-LRRs which describes the protein structure (nucleotide binding site-leucine rich repeat). Effector recognition by plant *R* genes triggers a strong and specific defense response [54, 61]. ETI triggered by oomycete pathogens causes localized cell death at the site of pathogen infection preventing colonization soon after pathogen attack [54, 61]. When recognized by a plant *R* gene pathogen effectors are referred to as avirulence (avr) proteins. Gene-for-gene resistance occurs when an avr/*R* combination triggers ETI, and often comprises a strong phenotype with a simple (i.e., monogenic) inhertance. Thus, gene-for-gene resistance is commonly selected for in breeding studies aimed at mapping or introgressing disease resistance traits.

Plant Pathogen Effectors

Oomycete Effectors

Secreted effector proteins appear to be essential components of virulence for oomycete pathogens. For example, the RXLR superfamily has been revealed as a significant class of effectors in some oomycete species [62]. Most RXLR effectors contain an N-terminal secretion signal followed by "RXLR" and "dEER" amino acid motifs that are followed by a variable effector domain that provides the virulence function [63]. These characteristics make oomycete effectors identifiable using bioinformatic techniques. Effector proteins are modular in design and contain W, Y, and L motifs [64]. It is thought that by re-assortment of these domains exposed surface residues can be altered in a targeted fashion [61]. Additionally, effector genes are located in dynamic portions of the genome which favor genetic recombination and mutation [61, 62]. RXLR effector proteins are secreted by plant pathogens to suppress PTI. If an *R* protein is not present, effector-triggered susceptibility can occur. If an *R* protein recognizes the effector, ETI occurs, resulting in resistance [65]. Pathogen effectors can sometimes suppress ETI mechanisms and manipulate physical barriers such as stomata adding multiple layers of complexity to the plant/pathosystem [61].

Fungal Effectors

Fungi represent a major class of plant pathogens that cause disease in crop plants. Two major groups of fungi contain biotrophic phases and are encompassed by rust fungi (*Basidiomycota*) and powdery mildew (*Ascomycota*) [66]. Many species such as the rice blast pathogen *Magnaporthe oryzae* (*M. oryzae*) have early biotrophic phases.

Hemibiotrophic or biotrophic fungi secrete effectors to subvert immunity in plant species through interfaces with the plant cell [66]. Most identified fungal effectors are secreted cysteine-rich proteins and those that have been identified inhibit plant chitinases and cysteine proteases [67]. Similar to oomycete pathogens, many fungi are capable of producing haustoria; however this is not observed in all species [66]. The haustorial interface has been well studied in rust pathogens that grow intercellularly [68]. Analysis of effectors from fungal pathogens that are obligate in nature has revealed an N-terminal motifs that are similar to the RXLR motif in oomycetes [69].

Bacterial Effectors

Bacterial plant pathogens are another major group of pathogens that cause disease on crop plants. Some of the most significant bacterial pathogens are *Pseudomonas syringae* (*P. syringae*) pathovars and *Xanthomonas* species [70]. *P. syringe* one of the largest number of effector genes among bacterial pathogens [71]. Researchers are now trying to understand how different combinations of virulence factors influence pathogen host range [71]. Similar to most biotrophic plant pathogens, effectors secreted by bacteria are an integral part of a successful infection. Some bacterial pathogens secrete effector proteins through a molecular syringe known as the type III secretion system [72-74]. At the N terminus of each T3SS effector gene is a leader sequence which guides the proteins through the T3SS [73]. TypeIII effectors are best studied in the Pseudomonas genus, where they are designated Hrp outer protein (HOP) [73]. In 2012, Colmer et al. identified 57 effector families from 17 *Pseudomonas* species [73]. Secreted effectors have also been well-studied in *Xanthomonas* species, which have been shown to contain transcription

activator-like (TAL) effectors [74]. These effectors suppress host immunity by modulating transcription in these plant pathosystem interactions [74].

Nematode and Aphid Effector Proteins

Nematodes and Aphids are major pests on crop plants. Most nematodes and aphids feed at the phloem and secrete effectors through the stylet [75]. Effector proteins contain a stylet secretion signal and some contain identifying sequences similar to oomycete plant pathogens. The green peach aphid, *Myzus persicae* (*M. persicae*) is a well sequenced and studied aphid that feeds on *N. benthamiana*. A total of 48 effectors for identified and tested from this species to identify unique phenotypes, immune suppression, and fecundity in which several candidates were identified [76]. Additionally, cyst nematode infections are common in many crop species and contain diverse effector repertoires that are secreted through the stylet [77].

Viral Effectors

Viral plant pathogens are well studied in crop plants. Some of the most important viral pathogens are mosaic, wilt, and streak viruses [78]. Viral avirulence proteins and non-viral effectors are recognized by plant resistance genes [79]. Like effectors from nearly every biotrophic plant pathogens, viral effectors almost always are virulence factors that suppress host immunity [79]. Thus, recognition of these pathogen proteins provokes ETI and systemic acquired resistance [79].

Disease Resistance

Identifying Resistance in Wild Species of Cultivated Crop Plants

As described above, host recognition of PAMPS and effectors is a conserved mechanism in all plant-pathosystems. Plant defense genes provide robust resistance and have been the focus of many introgression studies in cultivated crop plants. Wild species of cultivated crop plants are often encompassed by diverse germplasm collections. Resistance is commonly determined through pathogen assays or empirical observation. Large scale screens of wild germplasm are becoming more common as sequencing and genome sampling technologies become more available. Identifying accessions with resistance is always the starting points to begin targeted mapping and introgression studies.

Genetic Mapping in Plants

The goal of most early studies probing wild germplasm was to identify qualitative (single-gene) and/or quantitative disease resistance traits (multi-gene) [80]. Genetic mapping in segregating progeny (derived from crosses of resistant and susceptible parents) is used to determine the relative genomic position of an unknown gene(s) between genetic markers [81]. Mapping is frequently applied to disease resistance to compare genetic markers within segregating populations generated from crossing resistant and susceptible plants [81]. By analyzing the genotypes of segregating populations with respect to resistant versus susceptible phenotypes, genes which contribute to these phenotypes can be mapped to their genomic positions in parental accessions. As is the case with other traits amenable to breeding, development of high

quality linkage maps that can be used to localize genes or identify loci related to resistance requires careful consideration and judicious application of new technologies [81]. For example, sample sizes and type of marker(s) used for genotyping should be designed or optimized to maximize genome coverage in mapping populations. In this context, molecular genetic markers such as restriction fragment length polymorphisms (RFLPs), simple sequence repeat (SSRs), and single nucleotide polymorphisms (SNPs) have been exploited to facilitate mapping of both qualitative and quantitative trait loci (QTL) [82, 83]. Linkage maps can be improved by using more advanced mapping populations (e.g., developing recombinant inbred lines), by improving screening techniques, or by increasing the coverage of genetic markers [81, 82]. In sum, the advent of molecular genetics and genomic tools carries the same potential benefits for disease resistance breeding as for other plant traits.

Plant Breeding

Breeding starts from segregating populations in which resistance loci can be mapped. Breeding can be complicated when the wild species from which a trait is desired is not sexually compatible with the cultivated species [84]. If mating types do not match, sexual somatic hybrids can be generated [85]. Somatic cell hybrids can be used to map resistance traits or to provide large amounts of genetic diversity from wild species. Bridge species can sometimes be used if both species are sexually compatible with an intermediate species [85]. Conventional breeding was used in most early studies to map or introgress resistance from wild species using sexually compatible wild relatives, sexual somatic cell hybrids, or through bridge species [86]. Linkage drag is inevitable during introgression and results in the transfer of unwanted genes that can negatively affect crop

performance. Linkage drag can be minimized through the selection of tightly flanking markers [86]. Although linkage drag can be minimized, tightly linked elements can take many backcrossed generations to separate.

Applying New Technologies to Wild Crop Species

Marker-Assisted Breeding

As the availability of markers increases for reference genomes and wild species, identification of resistance loci becomes more precise and accurate. Markers can be used for cultivar identification, assessment of genetic diversity, parental selection, and confirmation of hybrids [86]. Thus marker-assisted selection (MAS) can be used to analyze population structures of wild and cultivated species. Since modern agriculture is predominantly monoculture, wild species and non-commercial cultivars are often the subject of genetic studies. Accessions containing a trait of interest are usually crossed with deeply-characterized cultivars known as a reference accession. MAS can be used to identify genes or loci of interest and track them during development of elite cultivars [87]. The use of markers to guide the breeding processes for specific gene(s) is known as marker assisted breeding (MAB). A genetic locus can be mapped in more detail by increasing the quantity or specificity of genetic markers at the chromosomal region of interest or by utilizing inbred populations to minimize heterozygosity [88]. Since resistance loci often contain tightly packed clusters of NBS-LRR proteins, mapping is often used to identify candidate genes which can then be confirmed in molecular studies [89].

High-Throughput Sequencing and Genotyping

Once a locus is mapped, genetic markers can be generated to facilitate introgression of specific loci from wild germplasm into cultivated species. Improvements in next generation sequencing (NGS) technologies and microarray platforms allow for

fast and cost effective large scale genetic screens [90]. Single nucleotide polymorphisms are the best represented form of sequence variation in plants and animals [91]. SNPs are binary in nature and can be detected with high accuracy through sequencing making them markers of choice for most modern studies [89]. As a result, many chip-based platforms have been developed specifically for SNP identification in addition to NGS technologies [89, 90]. Chip-based genotyping platforms may provide an effective alternative to sequencing when performing large scale genetic screens, a technique referred to as genome-wide association studies (GWAS) [89, 90, 92]. GWAS has been successfully used to identify thousands of associations to complex disease traits. GWAS can be used in fine-mapping studies where genomic constraints are placed on mapping populations to determine linkage disequilibrium [92]. Variants identified in this analysis are then annotated to draw functional linkages with disease resistance [92]. In many studies custom arrays are generated to increase the coverage of markers in a genomic region of interest [89].

Examples of Disease Resistance Introgression

Historically, disease resistance is a well-studied trait. With respect to wild progenitors of cultivated crop plants, early research was limited by practicality. In many early examples, identification and mapping of disease resistance traits signified a massive undertaking due to the limited genetic resources available for wild species. In many cases, studies were further limited by the quantity and quality of genetic markers in cultivated species and reference accessions. The advent of reliable, cost-effective sequencing technologies and microarray platforms has greatly impacted research into disease resistance traits from wild species. This technology has not only allowed markers to be identified in wild species, but has also been used to extensively characterize reference accessions. This makes the mapping process easier and more accurate, thus improving the efficiency of introgression into elite cultivars.

*Soybean (*Glycine*)*

One of the first examples of understanding disease resistance from wild soybean was based on identifying three amplified fragment length polymorphisms (AFLPs) involved in resistance to soybean mosaic virus (SMV) [93]. These were identified using near-isogenic lines derived from crosses with *G. soja* [93]. Recently, the Rsv4 locus which provides resistance to SMV was fine-mapped to a 120 kb interval [94]. Four intergenic SNPs were perfectly linked to resistance and two haplotypes were identified. Haplotype analysis suggested that the *Rsv4* locus was recently introgressed from *G. soja* [94]. The Rsv4 gene has been identified in *G. max* accessions Haman and

Ilpumgeomjeong and is used as a strong source of resistance against diverse SMV strains [94].

Resistance loci *rhg1* and *Rhg4* against soybean cyst nematode (SCN) were first identified by converting AFLP bands into polymorphic sequence-tagged-sites [95]. Candidate genes for *rhg1* and *Rhg4* were compared between *Glycine max* and *G. soja* cultivars [96]. The authors identified 31 population-specific haplotypes. 15 of these were specific to *G. soja* indicating two major haplotypes. This indicates that new haplotypes were generated during domestication while others were lost [96]. QTL resistance against soybean cyst nematode (SCN) by minor genes was described in *G. soja* accession PI464925B. The recessive *rhg1* locus was used to develop a natto variety of soybean known as Suzhime which contains resistance to SCN 1 and 3 [97]. Unfortunately, the *Rhg4* resistance locus is closely linked to agronomic traits such as seed pigmentation, limiting the effectiveness of this loci in practical agriculture settings [98]. Additionally resistance to major pathogens such as SCN and rust has been documented in perennial *Glycine* species, native to Australia [99].

Genetic resources can be developed for cultivated species which improve mapping and introgression of traits from wild progenitors. High resolution linkage maps have been constructed in cultivated *Glycine max* that will improve breeding with wild *Glycine* accessions [100]. A comprehensive screen of *G. soja* germplasm has recently been conducted to identify accessions with novel resistance to *P. sojae* [101]. This germplasm was screened with a combination of *P. sojae* pathotypes that overcome currently utilized Resistance to *Phytopthora sojae* (*Rps*) genes [101]. Large screens such as these lay groundwork for molecular breeding and mapping studies.

*Potato (*Solanum*)*

Resistance against nematodes and aphids is well studied in wild *Solanum* species. One of the first example of targeted introgression of disease resistance from a wild species was against potato nematode caused by *Heterodera rostochiensis* [102]. Hybrids of cultivated potato were generated with wild *Solanum* species that produced resistance responses to diverse European races of the potato nematode [102]. Resistance to potato root knot nematode, *Meloidogyne chitwoodi* races 1 and 2 was identified in *Solanum hougasii* [103]. Hybridization with cultivated potato, produced hybrids with resistance was specific to race 2 [103]. More recently the wild potato species, *Solanum spegazzinii*, was shown to contain three QTLs for resistance against potato cyst nematode [104]. One of these QTLs on chromosome XII appears to be in a cluster of orthologous nematode resistance genes, indicating that novel resistance genes can be found using QTL based approaches [104].

R genes against *P. infestans, Pi-mcd* and *Pi-ber*, were identified in wild species *Solanum microdontum* and *Solanum berthaultii* [105, 106]. In field studies it was shown that Pi-mcd delayed infection by three days and *Pi-ber* delayed infection by three weeks [107]. Pyramiding these genes did however appear to have an additive effect on *P. infestans* resistance because lines containing both resistance alleles were delayed in infection by five weeks in field studies [107]. Resistance against *P. infestans* from chromosome 5 was identified in wild potato [108]. Resistance QTLs were linked to QTLs which negatively impacted maturity, fruit size, and plant architecture [108]. This highlights the difficulties that still remain in trying to eliminate linkage drag when breeding with wild species. In a recent study, interspecific hybrids between *Solanum tuberosum* and a nutritional African variety of potato known as *Solanum villosum* were

shown to have resistance against late blight [109]. Identification of novel and agrinomically valuable *R* genes against *P. infestans* remains a top research concern moving forward.

Hybridzation has been used to introgress resistance from a variety of wild germplasm against diverse viral, bacterial, and fungal pathogens. Hybrids between *Solanum tuberosum* and *Solanum brevidens* showed resistance against potato leaf roll virus and potato virus Y [110]. Since these species are sexually incompatible, hybrids were generated using electro and chemical fusion techniques to create somatic hybrids. The authors measured disease resistance by virus titer and found significantly increased resistance among hybrids [110]. Resistance against bacterial wilt of potato was acquired using protoplast fusions between cultivated *Solanum* and *Solanum chacoense* [111]. Using primarily SSR markers, the authors were able to identify three alleles which were involved in resistance [111].

*Tomato (*Lycopersicon*)*

Tomato is also a solanaceous species known as *Solanum lycopersicum*. In tomato, the l2 locus, which confers resistance to *Fusarium oxysporum f. sp. Lycopersici* race 2, was transferred from *Solanum pimpinelifolium* to *Solanum esculentum*. RFLP analysis was used in the mapping of this resistance locus [112]. Resistance derived from the wild species *Solanum hirsutum* against early blight caused by *Alternaria solani* was later analyzed for QTLs [113]. Wild solanaceous tomato, *Solanum lycopersicoides*, was found to contain strong resistance to fungal pathogens *Botrytis cinerea* and *Alternaria solani* [114]. Resistance and susceptibility was analyzed using RNA sequencing to

identify putative defense genes and defense-related categories involved in immunity [114].

Important aphid and nematode resistance genes have been identified in wild tomato germplasm. Kaloshian et al. identified two genetically linked loci, Mi and Meu1 in tomato which provided resistance to the potato aphid *Macrosiphum euphorbiae* [115]. It is hypothesized that these two tightly linked resistance traits are derived from *Lycopersicon peruvianum* [115]. Further studies showed that the Mi locus contains multiple resistance genes [116]. Specifically, the Mi-3 gene was shown to be derived from *L. peruvianum* for resistance against root-knot nematode [116]. This highlights the utility of resistance loci as sources of resistance to multiple related pathogens.

One of the most well studied resistance genes, *PTO*, was introgressed from wild species into cultivated tomato species for resistance against bacterial speck caused by *Pseudomonas tomato*. Recognition of effector proteins AvrPto or AvrPtoB by Pto kinase activates multiple signal transduction pathways for resistance to this pathogen [117]. Pto has been used to effectively control bacterial speck since its discovery in the late 1970s [118].

Resistance to tomato chlorosis virus was observed in 2 interspecific hybrid crosses between wild *Solanum* and cultivated tomato [119]. Resistance was provided by a single major dominant loci in both *Solanum peruvianum* and *Solanum chmielewskii* [119]. An additive x dominance epistatic interaction with at least one other gene was also observed in hybrids generated with *S. chmielewskii* suggesting that introgression was more quantitative [119]. A goal of many recent studies is to more accurately determine linkage between genes in mapped resistance loci.

*Rice (*Oryza*)*

In 1997, Kush et al. described the need for broadening genetic pools in rice [120]. The authors describe a major hurdle to overcome in rice breeding being the low crossability and recombination with wild rice species. The authors used embryo rescue following hybridization to generate favorable hybrids containing traits such as male sterility, resistance to grassy stunt virus, and bacterial blight [120]. They suggested future research should be aimed at enhancing recombination of chromosomes and introgression of QTLs from wild species [120].

Introgression of resistance to rice blast in the elite cultivar IR64 is achieved using the Pi33 locus which is derived from the wild species, *Oryza rufipogon [121]*. The authors report unusually large introgressions at the Pi33 locus corresponding to nearly half of the short arm of chromosome 8 [121]. It is hypothesized that this introgression was mainted during backcrossing because of selective pressure during the breeding process [121]. In other words this positive trait was produced by linkage drag and unintentionally selected for by environmental disease pressures. In 2012, another novel rice blast was gene, Pi54rh was cloned from wild *Oryza rhizomatis* [122]. The authors used an allele mining approach to identify this orthologue of the Pi45 which provides broad spectrum resistance to *M. oryzae* [122]. Recently, Pi54of was cloned from *Oryza officinalis* and interacts with Avr-Pi54 through a novel non-LRR domain [123]. Introgression of multiple *R* genes or loci is at the foreground of engineering robust disease resistance. A comparison of the *Magnaporthe oryzae* resistance gene *Pi-ta* between invasive and weedy rice species showed that resistance is derived from a single origin in wild red rice known as “strawhull awnless” [124]. They also show that the susceptible allele, pi-ta, has been introgressed from cultivated rice to wild rice by out-

crossing [124]. This underscores the importance of preserving native germplasm and limiting natural outcrossing with cultivated species which may contain engineered traits.

As in potato, resistance in rice to aphids is often most effectivly and economically managed by *R* genes.The *R* genes *BPH6* and *BPH12*, effective against brown planthopper aphids, were identified in wild rice species and introgressed into japonica rice [125]. By pyramiding *BPH6* and *BPH12*, the authors were able to achieve stronger resistance responses [125]. This suggests that further pyramiding of *BPH* genes will be effective at producing varieties with durable resistance.

*Wheat (*Triticea*)*

The National Institute of Agricultural Botany has organized collection, evaluation, and utilization of biodiversity in wheat germplasm. Although this program focuses on transferring stress tolerance from wild to cultivated species, diverse germplasm developed by these organizations can be used to improve resistance as well [126]. Accessions of wild species of *Triticea* have been shown to have resistance against *Puccinia recondita f.sp. tritici* which causes what leaf rust and *Erysiphe graminis* which causes wheat powdery mildew. Sequence tagged site marker analysis was used to examine markers associated with resistance to these two pathogens in wild species and NIL of cv. Thatcher [127]. Later studies showed that the Lr1 resistance gene, conferring resistance to wheat rust was present in bread wheat and the wild species *Aegilops tauschii* [128].

The availability of new technologies for molecular engineering have led to the advancements in DNA hybridization techniques. The molecular staining technique giemsa C-banding genomic in situ hybridization (GISH) was used to show that 3Ns

chromosome of the wild wheat species, *Psathyrostac hyshuashanica*, contains resistance gene(s) against stripe rust [129]. This technique utilizes a chromosomal stain known as Giemsa which stains heterochromatic regions close to centromeres. When paired with probe based in situ hybridization, specific DNA sequences can be localized within the cell.

*Peanuts (*Arachis*)*

A recent study used artificial hybridization to introgress resistance from wild to cultivated peanut [130]. Molecular studies of resistance derived from the wild peanuts species, *Arachis diogoi*, against the fungal disease late leaf spot caused by *Phaeoisariopsis personata* showed 233 differentially expressed genes [131]. The authors used cDNA-AFLP and 2D proteomics to identify these defense gene candidates [131]. A large screen for resistance against leaf spot, rust, and scab was performed by [132]. The authors screened 43 wild accessions for resistance to these fungal pathogens and demonstrated that several varieties contained enhanced resistance to these diseases [132].

*Grapevine (*Vitis*)*

A selective sweep at *Rpv3*, a major resistance locus for grapevine downy mildew showed that selection favored haplotypes introgressed from wild vines [133]. Although resistance loci such as *Rpv3* represent important sources of resistance against *P. viticola*, the authors report that alleles at these loci are hard to manipulate by conventional breeding due to the close genetic distances between them [133]. QTL analysis was used to identify the loci *Rpv10* which contains eight resistance gene analogues [134]. *Rpv10*, like *Rpv3*, was introgressed from the wild grapevine species, *Vitis amurensis*. Pyramided populations containing both loci showed enhanced disease resistance [134]. RNA

sequencing of three Chinese wild grape species discovered new genes and SNPs which may facilitate future studies of disease resistance or other valuable agronomic traits [135]. A screen of 57 wild Chinese grapevine accessions with powdery mildew isolate En.NAFU1 yielded cultivars with enhanced resistance; however none of the tested cultivars exhibited full immunity [136].

*Cotton (*Gossypium*) and Barley (*Hordeum*)*

Authors studying **barley** took advantage of the secondary gene pool of *Hordeum bulbosum* to generate agronomically useful hybrids [137]. A more recent study identified the viral resistance gene, *Ryd4*, from wild germplasm that confers complete resistance to the barley yellow dwarf virus [138]. Introgression of genes for resistance against cotton leaf roll virus has been reported from the wild **cotton** species, *Gossypium stocksii* [139]. In F1 families, the authors report an additional unlinked trait for increased fiber strength was transferred during hybridization [139]. This is another example of positive traits being introgressed as an unintended consequence of linkage drag.

Minor Crops

Resistance to scab from wild **apple** is conferred by HcrVf2 [140]. Wild **cucurbitaceous** species show resistance to Tozuchini yellow mosaic virus [141]. The first resistance gene analogues in **strawberry** were identified from wild and cultivated strawberry [142]. *Helianthus argophyllus,* a wild accession of **sunflower** was used to map the PI(ARG) resistance locus against the emerging oomycete pathogen *P. halstedii*, the causal agent of sunflower downy mildew [143].

The Future of Disease Resistance in Wild Species

A Need for Durable Resistance

As mentioned above, the evolution of plant pathogens is being driven by planting in monoculture [5]. Monoculture is attractive because single varieties are easier to plant, harvest, and market [5]. However, these varieties will be uniformly susceptible to a compatible pathogen isolate [5]. In 2000, Zhu et al. reported that crop heterogeneity is a potential solution to this disease problem [6]. They worked with Chinese agricultural extension agents in the Yunnan Province to plant diverse rice varieties, consisting of mixtures of resistant and susceptible cultivars, in five townships [6]. When planted with resistant varieties, the yield of susceptible varieties improved by 89% and the severity of rice blast decreased by 94% when compared to susceptible varieties grown in monoculture [6]. Another option for improving disease resistance in crop plants is to use molecular techniques and breeding to stack multiple *R* genes in elite cultivars, a technique referred to as pyramiding resistance. If we mobilized an array of *R* genes into yield optimized (elite) cultivars, we could provide robust disease resistance to susceptible varieties. These approaches have proven difficult due to the limited genetic resistance available in many cultivated species. Obviously, *R* genes from wild relatives could fit well into this approach.

Applying Molecular Techniques

Recently, there have been several significant examples of successful mapping and/or introgression of disease resistance from wild species and related species of cultivated crop plants. Jones et al. applied marker assisted breeding to map and introgress

resistance from the wild leguminous species *Cajanus cajan* (*C. cajan*) against the causal agent of Asian soybean rust (ASR), *Phakopsora. pachyrhizi* (*P. pachyrhizi*) into cultivated soybean *G. max* [144]. Jones et al. used traditional pathogen based screening to identify durable resistance against diverse isolates of *P. pachyrhizi* that was mapped to a 200kb region. The authors cloned possible NB-LRR resistance genes and used heterologous expression in *G. max* to identify a single gene which confers resistance to ASR. In chapter three, *P. sojae* resistance *G. soja* accessions were probed with core pathogen effectors to identify durable resistance. Germplasm containing resistance was crossed with the *G. max* cultivar Williams to create inbred populations for mapping and introgression in elite cultivars. Recent advancement against potato light blight have focused on Swedish varieties which contain resistance to the causal agent *P. infestans* [145]. The authors screened with over 50 *P. infestans* effectors to identify a specific resistance response against Avr2 [145]. Eight R2 homologues were introgressed to identify a single gene responsible for Avr2 based resistance [145]. It is clear from these examples that disease resistance in wild species can be mapped and introgressed into cultivated crop plants to defend against emerging pathogens.

Conclusions

Over the past two to three decades, understanding of plant pathosystems has increased tremendously. We have discovered many new pathogens and witnessed the re-emergence of others. Groups have begun to study plant-pathogen interactions at the country and global levels, and comprehensive disease assessments are made on a semi-annual basis for most staple crop plants. Advancements in molecular techniques give researchers the ability to track the emergence of new pathotypes and inform growers of potential threats.

By understanding pathogen population structures, it may be possible to predict future outbreaks and take pre-emptive measures. Recent examples focus on the use of new sequencing technologies and large screens to understand resistance derived from wild species. Understanding *R* genes and resistance loci has been the topic of recent studies in most recent studies. By studying alleles of resistance genes from wild populations, new resistance genes can be identified. In many wild species, resistance to various pathogens is being identified through large screens. Pyramiding *R* genes and/or loci have been shown to have a cumulative effect on immunity; however, difficulties separating overlapping loci or tightly clustered genes within *R* loci continues to be problematic. Recent studies have focused on GWSS techniques to improve mapping in segregating populations. Additionally, it is becoming common practice for groups to screen with diverse isolates and pathotypes to ensure durability of the resistance trait. This allows durable disease resistance genes or loci to be rapidly introgressed from wild species into cultivated crop plants for deployment in field settings. A significant shift toward translational plant science applications is underway as these techniques become increasingly more refined. These advancements have allowed small laboratories an opportunity to produce translational products for global markets which has begun ushering in a new age in plant research and development.

Chapter 2

Effector-Directed Screens for New Disease Resistance Genes against the Soybean Root and Stem Rot Pathogen Phytophthora sojae

Kevin E Fedkenheuer[1], Michael G Fedkenheuer[1], Colin L. Davis[2], Joel L. Shuman[1], Nick A. Dietz[1], Brett M. Tyler[3], M. A. Saghai Maroof[2], and John M. McDowell[1*]

[1]Department of Plant Pathology, Physiology, and Weed Science, Latham Hall, Virginia Tech. Blacksburg, VA 24060-0390

[2]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR

[3]Department of Crop, Soil, and Environmental Sciences, Latham Hall, Virginia Tech, Blacksburg, VA 24060-0390

[*]for correspondence, johnmcd@vt.edu

In preparation for submission to *New Phytologist*

Author Contributions: KF wrote and developed this manuscript with advice from JM, BT, MF, and CD. JM, BT, and SM advised this project. JS helped with plant maintenance and *P. sojae* culture advice. KF developed the *Pseudomonas* effector delivery system in soybean. KF trained MF and CD in the procedure. MF and CD helped optimize this procedure for screening segregating $F_{2:3}$ populations. KF, MF, CD, and ND screened the $F_{2:3}$ populations used in this study. MF developed a scoring system to analyze segregating populations.

Abstract

Phytophthora root and stem rot disease suppresses global soybean production. We developed a system to screen soybean germplasm for *Resistance to Phytophthora sojae* (*Rps*) genes that recognize core *Phytophthora sojae* RXLR effectors. New *Rps* genes that target core effectors will likely be effective and durable against diverse *P. sojae* isolates. We developed a system using bacterial *Pseudomonas* to deliver individual *P. sojae* effectors into soybean by Type III secretion. If the delivered effector is recognized by a resistance (*R*) gene, a visual hypersensitive response (HR) is produced. We used *P. sojae* avirulence (*Avr*) effectors to validate that this approach can recapitulate known *Avr/R* interactions. We delivered core *P. sojae* effectors *Avh*16, *Avh*180, and *Avh*240 into disease-resistant *Glycine max* germplasm. *G. max* accessions that produced a HR in response to core effector(s) were selected for crosses with *P. sojae*-susceptible cultivar Williams. We assayed genetic segregation of the response to three effectors in $F_{2:3}$ progeny derived from crosses of resistant X susceptible soybean. In all cases, we observed a simple, genetically dominant inheritance pattern that is consistent with gene-for-gene resistance and suggestive of amenability to breeding. A screen with additional core *P. sojae* effectors suggested that the resistant germplasm likely contains many potential new *Rps* genes. We will characterize and breed these potential *Rps* loci for commercial deployment against a destructive disease of soybean.

Introduction

Phytophthora sojae, an oomycete pathogen, is the causal agent of *Phytophthora* root and stem rot disease on cultivated soybean, *Glycine max* [43, 146]. The United States of America (USA) is the world leader in soybean production, devoting 29% and 31% of its cropping area to soybean in 2013 and 2014 respectively [147]. Soybeans are crushed to extract valuable oil. The crushed remainder is processed into soybean meal [148]. Soybean meal is marketed globally as a food source for animals and fish [148]. In 2006, soybean crop losses due to *Phytophthora* root and stem rot were estimated at 1.46 million metric tons in the USA and over 2.32 million metric tons worldwide [19]. This was a loss of 1.75% of all soybean produced during that year in the USA [19]. Robust management options are needed to mitigate these losses.

P. sojae overwinters in dried plant tissue and soil, typically as oospores [43, 149]. Germinating oospores can infect seedlings directly, or can produce sporangia, containing asexual zoospores [43, 149]. Once released, zoospores will swim towards root exudates, encyst, and invade root tissue with filamentous hyphae [150]. *P. sojae* hyphae grow intercellularly and produce feeding structures, called haustoria, which penetrate the cell wall and interface with the plant cell membrane [151]. *P. sojae* secretes effector proteins into soybean cells at this interface to suppress plant immunity and to promote virulence [43, 149, 151]. A compatible infection is propagated asexually during the soybean growing season [43, 149].

Pests and pathogens secrete effectors to subvert plant immunity [54]. The genome sequence of *P. sojae* revealed hundreds of genes encoding candidate effectors [152-155]. One large effector family, RXLR effectors, can enter plant cells [54]. These effectors

contain an N-terminal signal peptide, RXLR-dEER motif, and an effector domain [63]. The *P. sojae* genome contains nearly 400 genes encoding candidate RXLR effectors [156]. Many effector genes have variable sequence and expression in diverse *P. sojae* isolates [152, 156]. If a *P. sojae* effector is recognized by a soybean *Rps* gene (effector-triggered immunity or ETI), the plant produces localized cell death (i.e., a Hypersensitive Response or HR) [149]. These gene-for-gene interactions trigger a robust immune response which impedes pathogen growth [54].

Gene-for-gene resistance has been effective at controlling *Phytophthora* root and stem rot under field conditions [43, 149, 157]. A total of 24 genes has been identified. Of these, seven (*Rps*1a, *Rps*1b, *Rps*1c, *Rps*1k, *Rps*3a, *Rps*6, and *Rps*8) have been commercially deployed to manage this disease [43, 157]. Because *R* gene function is dependent on effector(s) recognition, *Rps* genes can be defeated when *P. sojae* genetic diversity includes isolates for which *Avr* effectors are mutated or silenced [43, 157]. *P. sojae* isolates that defeat all commercial *Rps* genes have been described [43]. The average effectiveness of a soybean *R* gene in the field was estimated at 8-15 years, underscoring the need to identify new sources of genetic resistance against *P. sojae* [158].

Effectors have been identified that interact with all of these *Rps* genes except for *Rps*8 [43]. In every case (*Avr*1a, *Avr*1b, *Avr*1c, *Avr*1k, *Avr*3a/5, *Avr*3b, *Avr*3c, and *Avr*4/6), all recognized effectors are from the RXLR family [149]. With the exception of *Avr*3b, these effectors make minor contributions to *P. sojae* virulence on soybean [149]. Non-essential *P. sojae* effector genes can be silenced, deleted, or mutated in response to *Rps*-mediated selective pressure with little or no fitness cost to the pathogen. These

mechanisms for evading host immunity can subvert conventional *R* genes, rendering the *R* gene useless for disease control.

Effector-directed breeding strategies (termed "effectoromics") can accelerate disease resistance breeding against plant pathogens [159]. For example, effectoromics has been successfully applied to improve potato resistance against potato late blight [160]. Vleeshouwers et al assayed 54 effectors from *Phytophthora infestans* for *Avr* activity on wild *Solanum* [160]. The RXLR family lpiO induced a HR on three species which revealed the *Rpi*-blb1 *R* gene. They identified the cognate effector *Avr*-blb1 by transiently expressing IpiO and *Rpi*-blb1 in a heterologous *Nicotiana benthamiana* system [160].

In soybean, Staskawicz et al. used effectoromics to explore race compatibility of *P. s. glycinea* [161, 162]. They constructed a genomic DNA library for *P. s. glycinea* race 6 and mobilized it into *P. s. glycinea* race 5 [161]. One transformed clone changed *P. s glycinea* race 5 from virulent to avirulent on the appropriate soybean cultivars [161]. Staskawicz et al. applied a similar approach to characterize two avirulence genes from *P. s. glycinea* race 0 and race 1, *Avr*B0 and *Avr*C [162]. The identification of *Avr*B1 provided evidence of an *Avr/R* interaction between *Avr*B1 and the *Rpg*1 locus [162].

We chose to identify novel *Rps* genes which recognize core *P. sojae* effectors. We define core effectors as highly expressed in early infection, conserved amongst diverse isolates, and necessary for virulence based on reverse genetic criteria. *Rps* genes against core *P. sojae* effectors will likely be more durable and more effective against genetically diverse isolates when compared to traditional *Rps* genes, because *P. sojae* cannot easily discard a core effector without a major fitness penalty. In this study, we designed a bacterial *Pseudomonas* system for delivering oomycete effectors into soybean. This

system was used to screen soybean germplasm for potential new *Rps* genes against numerous core *P. sojae* effectors. *G. max* accessions with potential new *Rps* genes were crossed with susceptible cultivar Williams. Inheritance of three candidate *Rps* genes in $F_{2:3}$ populations suggested single, dominant loci, suggesting that the effectors are recognized by R proteins that can be easily bred into elite soybean cultivars.

Results

Overview of the Screen

Our approach tests *G. max* for a localized cell death (hypersensitive response or HR) against specific pathogen effectors, when the effector is delivered to the inside of plant cells by a bacterial surrogate. A visual HR suggests the presence of a *R* gene in a tested population. To identify novel and durable *R* genes, we assayed *P. sojae* resistant soybean germplasm for HRs against core *P. sojae* effectors. *G. max* accessions which contained predicted *R* genes were crossed with susceptible Williams. *Avr* activity was assessed in $F_{2:3}$ populations to determine gene inheritance. An overview of the strategy is provided in Figure 2.1.

Identification of Resistant Soybean Germplasm

In a recently published study, Matthiesen et al. screened 1,019 Glycine accessions with a mixture of three *P. sojae* isolates which defeat 13 commercial *Rps* genes [101]. They predicted that 17% of the *G. max* accessions contained novel *Rps* loci [101]. We used these accessions in our screen with core effectors. We listed the *G. max* accession IDs and its corresponding plant introduction (PI) number for tested *G. max* germplasm in Table 2.1.

Identification of Core Effectors from P. sojae

This work will be described in a forthcoming manuscript [163]. In this study, Wise et al. identified 30 RXLR effectors (from ~400 predicted RXLR genes [164]) which are highly expressed at early time points in the infection and which are genetically monomorphic amongst four *P. sojae* reference [163]. These isolates collectively

encompass much of the *P. sojae* genetic diversity in the USA [156]. RNA silencing of effector genes was used to test whether seventeen effectors are important for virulence [163]. The results indicated that three effectors have major effects on virulence, four have no effect, and the remaining ten have intermediate effects [163]. The three essential effectors (*Avh*16, *Avh*180 and *Avh*240) were subsequently silenced in an additional three reference isolates and were shown to be essential in all three isolates [163]. *Avh*16, *Avh*180 and *Avh*240 were shown to be highly expressed in all four *P. sojae* reference isolates [163]. These results suggest that *Avh*16, *Avh*180 and *Avh*240 are broadly important across the species, and we therefore selected them as the first effectors for our screen [163]. Germplasm was tested with nine additional effectors: *Avh*53, *Avh*137, *Avh*261, *Avh*488, Snel1, *Avh*94, *Avh*241, *Avh*23, or *Avh*110 [163]. This second set of effectors was chosen for screening based on gene expression and sequence conservation [163]. RNA interference experiments indicate that *Avh94* and *Avh241* make major contributions to virulence and that *Avh110* and *Avh23* only make moderate contributions to virulence. The remaining genes have not yet been tested with RNAi.

Development of a Bacterial System for Transient Effector Delivery

Our primary objective was to develop a system for delivering *P. sojae* effectors, one at a time, to the interior of plant cells so that recognition of the effector by an *R* protein would trigger a visible cell death response. The screens with *P. infestans* effectors were accomplished using *Agrobacterium*-mediated transient transformation [160]. This approach works well in tomato and related species, but soybean is recalcitrant to *Agrobacterium*-mediated transient transformation (K. Fedkenheuer, data not shown). Therefore, we decided to focus on developing a system through which the effectors could

be delivered by the *Pseudomonas* Type III secretion system. Effector delivery was accomplished using the effector-detector vector (pEDV6) that provides a Type III secretion system (T3SS) signal [165]. Cloning into pEDV6 produces a *AvrRps*4-RXLR fusion product [165]. The bacterial *AvrRps*4 leader sequence enables secretion of the construct through the T3SS into plant cells [165]. This approach has proven useful when screening for virulence or avirulence activity of *Hyaloperonospora arabidopsidis* RXLR effectors in *Arabidopsis* [165].

We chose several *Pseudomonas* strains to evaluate for the ability to deliver *P. sojae* effectors into soybean. *Pseudomonas syringae pathovar tomato DC3000* is a widely used reference strain. D28E is a mutant version of the reference strain *Pst DC3000* in which 28 Type III effectors have been removed by targeted deletions [166]. D28E (*Avr*PtoB) is D28E with the *Pst DC3000* effector *Avr*PtoB, which can potentially suppress PTI [166]. *Pseudomonas syringae pathovar glycinea* (*Psg*) race 4 is a virulent pathogen on soybean [162]. *Pseudomonas fluorescens (Pfo)* is a soil bacterium that is non-pathogenic and does not contain Type III effectors nor a T3SS [167]. *Pfo* strain EtHAN contains a transgenic, stably integrated operon of genes encoding the *Pst DC3000* Type III secretion components [167].

First, we tested whether these strains provoke a HR in soybean. *Pst DC3000* produced a visual HR on all tested soybean cultivars. No visual HR was observed on leaves in response to infiltration with *Pst DC3000* D28E, *Pst DC3000* D28E (*Avr*PtoB), *P. s. glycinea* (*Psg*) race 4, or EtHAN. Next, we tested whether these strains could deliver *P. sojae* effectors into soybean. We cloned *P. sojae Avr* effectors *Avr*1k, *Avr*1b, and *Avr*4/6 into pEDV6 and transformed the five *Pseudomonas* strains with these constructs.

For each strain, we challenged cultivars containing *Rps*1k, *Rps*1b, and *Rps*4 with their cognate effectors. Each *Pseudomonas* strain was evaluated for its ability to produce a visual HR on *Rps* cultivars in known *Avr/R* interactions.

We performed *in vivo* bacterial growth assays following infiltrations of *Psg-Avr*1k and EtHAN-*Avr*1k. We did not recover any *Psg race 4* carrying the pEDV6-*Avr*1k plasmid at 5 DPI (Figure 2.2). EtHAN carrying *Avr*1k grew only 1 log $(cfu/cm)^2$ at 5 DPI in the presence *Rps*1k (Figure 2.2). The unexpected absence of *P. s. glycinea-Avr*1k bacteria at 5 DPI led us to test the stability of the pEDV6-effector constructs in our five *Pseudomonas* strains. We performed an experiment in which the strains were assayed for a loss of the plasmid prior to inoculation into soybean. The strains were grown in King's Broth (KB) media with antibiotic selection and were re-suspended in 10 mM $MgCl_2$. The cells were incubated at 28ºC and were assayed for presence of the pEDV6-effector construct by plating the bacteria on selective and non-selective media. The results indicated that, in all strains except EtHAN, pEDV6-effector constructs were rapidly lost during a 15 minute incubation in 10 mM $MgCl_2$ (Table 2.2). This suggests that the pEDV6-effector construct is lost from all *Pseudomonas* strains except EtHAN when selective pressure is removed (Table 2.2).

We hypothesized that *P. sojae* effectors could be toxic to virulent *Pseudomonas* strains and potentially to non-pathogenic EtHAN. To test this possibility, we measured the growth rate of strains in minimal media with antibiotic selection to force plasmid retention. All three *P. sojae* effectors were mildly toxic to *P. s. glycinea* and to EtHAN, reducing WT growth rates (Figure 2.3). Despite mild sensitivity to *P. sojae* effectors,

EtHAN was selected as the only candidate because it did not eject pEDV6-effector constructs (Table 2.2 and Figure 2.3).

Optimization of the Growth of EtHAN in planta

To improve HR consistency and intensity, we explored options to enhance effector delivery in EtHAN. We hypothesized that increasing the *in planta* growth of EtHAN might improve visual signal in response to *P. sojae* effectors. This hypothesis was based on prior work by Keen et al. that demonstrated that the *in planta* growth of *P. fluorescens* could be improved by co-infiltration with a virulent strain of *Psg* [168].

We confirmed these results by assaying the growth of EtHAN in soybean leaves with and without *Psg* race 4. The addition of *Psg* race 4 by co-infiltration improved the *in planta* growth of EtHAN two-fold in a compatible interaction (Figure 2.4). In an incompatible interaction, the addition of *Psg* race 4 augments the growth of EtHAN tenfold (Figure 2.4). This was seen in incompatible interactions of EtHAN delivering *Avr*1k, *Avr*1b, or *Avr*4/6 on soybean cultivars *Rps*1k, *Rps*1b, and *Rps*4 respectively (Figure 2.4). Based on these results, we chose to use *P. s. glycinea* race 4 to augment the growth of EtHAN for screening soybean germplasm. There are no known *R* genes against *P. s. glycinea* race 4 [162]. We confirmed that our starting germplasm was susceptible to this race by pre-screening accessions with this *P. s. glycinea* race 4.

Optimization of Plant Growth

We optimized several aspects of plant growth to achieve consistent results with our system. We found that a light intensity of 250-300 μM to be key for the production of consistent macroscopic HRs during screening. It is critical that plants are consistently watered. Drought stress will cause early senescence of the unifoliate leaves. Flats were

watered with 1.5 L of H_2O per flat every 36-48 hours or more frequently as needed. Humidity was held at 80% RH.

Approximately 12-14 days post planting, *G. max* cultivars will begin producing immature trifoliate leaves. This is described as the V2 growth stage for soybean [169]. Because results are scored seven days post infiltration, senescence of the unifoliate leaves can interfere with data collection. We discovered that pruning the first set of trifoliate leaves causes a reprogramming of the unifoliate leaves (Figure 2.5): After two days the unifoliate leaves become thicker, darker, and resilient (Figure 2.5). By removing trifoliate leaves at the stalk apex, unifoliate leaves are exposed to more direct light as they are unshaded. In *Avr/R* interactions, effector-responses were stronger and more consistent on unifoliate leaves from pruned plants compared to unifoliate leaves from unpruned plants.

A Facile Assay for P. sojae *Virulence*

We also developed an assay to determine *P. sojae* virulence on detached trifoliate leaves. Mycelial plugs from *P. sojae* race 2 produced large lesions on trifoliate leaves from susceptible cultivars which were wounded at the plug site (leaf center): Williams and Harosoy (Figure 2.6). This isolate did not produce an expanded lesion on the resistant control cultivar Haro 15 (Figure 2.6). Using this technique, we confirmed *P. sojae* resistance in *G. max* accessions, accession 32 (gm32) and gm326 (Figure 2.6).

Validation of the Pseudomonas*/EDV Delivery System*

After optimizing plant growth, the next step was to test whether *Avr*1k, *Avr*1b, and *Avr*4/6 can elicit a strong and consistent HR when delivered by EtHAN to soybean cultivars *Rps*1k, *Rps*1b, and *Rps*4 respectively (Figure 2.7). We quantified the consistency of the macroscopic HR as a percentage of infiltration sites. *Avr1k* had the

highest percent HR on *Rps*1k and *Avr4/6* had the lowest percent HR on *Rps*4 (Figure 2.7b). We produced similar results using *Rps* cultivars in the Williams background (Table 2.3). EtHAN alone, or with *P. s. glycinea race 4*, did not produce a HR on any tested soybean cultivar (Table 2.4).

To quantify cell death, we measured ion leakage from the soybean leaves into water using a conductivity meter. Infiltration of strains carrying *Avr*1k, *Avr*1b, and *Avr*4/6 into cultivars containing a cognate *R* gene produced higher levels of ion leakage (20-30% higher) when compared with the Harosoy control (Figure 2.8). Co-infiltration of EtHAN and *Psg* race 4 produced some background cell death (Figure 2.8). These results indicate that the *Pseudomonas* delivery system is suitable for screening *G. max* germplasm to find novel gene-for-gene interactions.

Screening of Resistant G. max *Germplasm with* P. sojae *Core Effectors*

To identify potential new *R* genes, we screened 32 *G. max* accessions which are predicted to contain novel *Rps* loci [101]. Our first set of screens utilized EtHAN strains carrying the three most promising core effectors, *Avh16*, *Avh180*, and *Avh240*. From the 32 *P. sojae*-resistant *G. max* accessions, six accessions responded to *Avh*16 with a visible HR, six accessions responded to *Avh*180, and six accessions responded to *Avh*240 (Table 2.5). *Avh*7a, a homolog of *Avh*16, produced a HR on the same accessions that responded to *Avh*16 (Table 2.6). Three accessions responded to two effectors, and one accession responded to all three effectors (Table 2.5). Seventeen accessions did not respond to these three effectors. Additionally, six *G. max* accessions with partial resistance against *P. sojae* were included as blind negative controls and did not respond to effector treatment.

Screening Resistant G. max *Germplasm with Additional Effectors*

The positive results from screens with *Avh*16, *Avh*180, and *Avh*240, encouraged us to screen the same 32 *G. max* accessions with a second set of core *P. sojae* effectors, namely *Avh*53, *Avh*137, *Avh*261, *Avh*488, Snel1, *Avh*94, *Avh*241, *Avh*23, and *Avh*110 (Table 2.7). These effectors were selected by the same criteria as *Avh*16, *Avh*180 and *Avh*240, except that their function in P7076, P7074 and P7063 was not tested by silencing. We found that many *G. max* accessions respond to one or more of these effectors (Table 2.7). *Avh*23 and *Avh*110 did not provoke a HR on any tested accessions.

Genetic Segregation of Effector Response in $F_{2:3}$ Populations

Seven effector-responsive accessions were selected for crossing with susceptible Williams (Table 2.8). We chose to breed resistant accessions with Williams because it has a sequenced genome [170] and no *Rps* genes [171]. A sequenced genome enables us to design single nucleotide polymorphism (SNP) chips which are essential for soybean locus mapping. A SNP is a nucleotide variation at a genomic position within a population and these chips can assay multiple SNPs concurrently [172]. Williams was chosen for breeding based on maturity group as well. The maturity group refers to the cultivar development period which is characterized by growth stages [169]. Soybean producers use maturity group information to accurately time chemical treatments. Williams' 3.9 maturity group is ideal for growth in the Midwest.

We analyzed $F_{2:3}$ seed for two crosses, accession 32 (gm32) x Williams and gm326 x Williams. Because most known *R* genes segregate as single, genetically dominant loci, we predicted that the HR response to the effectors would segregate similarly, producing a 1:2:1 phenotypic ratio in segregating $F_{2:3}$ populations.

We tested segregating $F_{2:3}$ families for *Avr* activity using the effector delivery system. For gm32, we observed a phenotypic ratio not significantly different ($x > 0.05$) than 1:2:1 in response to *Avh*16 and in response to *Avh*240 (Table 2.9). For gm326, we observed a phenotypic ratio not significantly different ($x > 0.05$) than 1:2:1 in response to *Avh*180 (Table 2.9). We can conclude that gm32 likely contains two dominant *R* genes and gm326 likely contains one dominant *R* gene. Was Fig 2.9 discussed at all?

Discussion:

Phytophthora root and stem rot disease on soybean was first described in 1954 by soybean farmers in southwestern Ontario and in the Midwest [173]. *P. sojae* owes its "success" as a crop pathogen to a plastic genome which supports rapid evolution. *P. sojae* isolates have been described which defeat all commercial *Rps* genes. To mitigate this disease in the future, we need new and robust disease management options. Genetic resistance is the most economic tool, but most *R* genes against *Phytophthora* pathogens are not durable because they recognize effectors that do not play essential roles in virulence.

Our strategy is based on screening for novel *R* genes against core *P. sojae* effectors and to breed these *R* genes for commercial deployment. To probe soybean germplasm recognition of core *P. sojae* effectors, we first needed a flexible system for transient expression in soybean. We exploited the *Pseudomonas* T3SS for effector delivery into soybean, drawing from early work by Staskawicz et al., who produced visual cell death symptoms by transforming *P. s. glycinea* race 1, race 4, and race 5 with molecular components from *P. s glycinea* race 6 [161]. AvrB1 was shown to produce an HR on Rpg1 [162]. While analyzing the *Rpg*2, *Rpg*3, and *Rpg*4 loci in 1991, Keen and Buzzell transformed *P. s. glycinea* race 4 with *Avr*D from *Pst* and observed visual incompatibility [174]. They suggested that *Rpg*4 may be responsible for *G. max* resistance against *Pseudomonas* carrying *Avr*D [174]. These experiments demonstrate that *P. s glycinea* is capable of delivering exogenous, plasmid borne effectors to soybean leaves in sufficient quantity to produce *Avr* activity in *Avr/R* interactions. Using this information, we constructed a *Pseudomonas* system for transient expression in soybean.

Our system uses Type III secretion by *Pseudomonas fluorescens* strain EtHAN to deliver *P. sojae* effectors. A critical component of effector delivery to *G. max* by EtHAN is the addition of *P.s. glycinea* race 4 to the inoculum. We hypothesize that *P. s. glycinea* race 4 improves the growth of EtHAN *in planta* by suppressing plant immunity. To validate this system, we exploited known *Avr/R* interactions between *P. sojae Avr* effectors and *G. max R* proteins. When *Avr*1k, *Avr*1b, and *Avr*4/6 were delivered to *G. max* cultivars which contained a cognate *R* gene, we observed visual HRs. We confirmed cell death symptoms using an ion leakage assay. This new system gave us the unique ability to study oomycete effectors in soybean. Importantly, the bacterial system recapitulated the same gene-for-gene specificity as is seen in experiments with *P. sojae* containing the same *Avr* genes. In fact, the previously documented strength of resistance was recapitulated in our system: The response to *Avr*1K was strongest and the response to *Avr*4/6 was weakest, mirroring the strength of the responses in either *P. sojae*-based assays or when the effectors are delivered by biolistics [175, 176]. Thus, this system was validated for studying oomycete effectors in soybean. The broad applicability of EtHAN as an effector-screening system is further underscored by recent reports of its successful use in monocot wheat leaves [177, 178].

We tested 12 core effectors, and observed that ten produced a visual HR on at least one of the 32 tested *G. max* accessions. Six accessions responded to one effector, twelve accessions responded to two effectors, seven accessions responded to three effectors and one accession responded to four effectors. By selecting for *P. sojae* resistant germplasm and by then selecting for accessions with no known resistance, we likely enriched our germplasm with accessions containing new *Rps* genes.

We selected seven accessions which produced a HR in response to a core effector(s) for breeding with susceptible Williams. We observed *Avr* activity in $F_{2:3}$ progeny from crosses, gm32 x Williams and gm326 x Williams. In $F_{2:3}$ populations, a 1:2:1 phenotypic ratio was expected and observed suggesting dominant inheritance. We concluded that gm32 likely contains *Rps* genes which recognize core *P. sojae* effectors *Avh*16 and *Avh*240. Gm326 likely contains one dominant *Rps* gene which targets *Avh*180.

In addition to resistance loci *Rps*1 to *Rps*6 for which *Avr* effectors have been identified, twelve additional *Rps* loci have been described [157]. These uncharacterized resistance loci may explain some of our *Avr* activity. In 2013, Lin et al. independently mapped two *R* genes in PI567139B, *Rps*Un1 and *Rps*Un2 [179]. We independently identified effector-responses in PI567139B against *Avh*16 and *Avh*240. This effector-response may correlate to *Rps*Un1 and *Rps*Un2. It is unlikely that effector-responses were due to effector-toxicity in soybean leaves because effector-responses were accession specific and were similar to effector-responses by *Rps* cultivars in known *Avr/R* interactions.

Effector-directed breeding provides many advantages over classical pathogen-based breeding. Effector-directed breeding allows for the identification of potential *R* genes which recognize core pathogen effectors. This approach can be applied to find potential *R* genes against any pathogen with a sequenced genome and identifiable effectors. Non-host plants can be screened for effector recognition as well. Effector-directed breeding can facilitate stacking *R* genes.

We suggest using both effector-directed breeding and pathogen-based screening to identify novel, durable R genes. This may help to eliminate accidental duplication of *R*

loci during resistance breeding. Mapping the same population using both *Avr* activity and pathogen screens will create a multidimensional map to more accurately determine resistance mechanisms. For this reason, we have developed a technique to detect *Avr* activity in simultaneous effector and pathogen assays. This system will yield effector data and pathogen data for a single individual.

In principle, our effector delivery system could be used to screen any host plant for *R* genes against a pathogen for which effectors are characterized. EtHAN was used to successfully deliver effectors to wheat leaves [177, 178]. This demonstrates that EtHAN may be useful for effector delivery in a broad range of systems. Our immediate goal is to use effector data and pathogen data to map potential *Rps* genes in soybean RI lines. In the future, we hope to introduce new and durable *Rps* genes to soybean growers for protection against *Phytophthora* root and stem rot.

Materials and Methods

Effector Plasmid Clones

Clones encoding RXLR effectors were cloned into pENTR using the pENTR™/D-TOPO® Cloning Kit – ThermoFischer Scientific or they were provided by the Tyler lab at Oregon State University. The coding region of the ORF began with a methionine codon, followed by the first codon following the signal peptide cleavage site as predicted by SignalP. An LR reaction was used to clone the effector genes into the destination vector pEDV6 [167]. Constructs were transformed into10 *E. coli*. Single colonies were selected on KB agar containing Gentamycin (10ug/ul) and were screened for the desired plasmid by PCR, using a pEDV6 forward (F) primer and a gene specific reverse (R) primer. Positive transformants were maintained at 37 °C under Gentamycin selection.

Transformation of EtHAN by Tri-Parental Mating

P. fluorescens strain EtHAN was provided by the Chang lab at Oregon State University [167]. EtHAN was transformed using tri-parental mating as follows: Bacterial strains RK600, EtHAN, and positive transformants in top10 *E. coli* were streaked onto KB agar plates with the proper selection. Onto one King's Broth (KB) agar plates, we streaked our positive transformants in top10 then the helper strain RK600 followed by EtHAN. The plates were incubated for three days at 28 °C. Following the incubation, colonies were streaked onto KB agar containing Chloramphenicol (10ug/ul), Nitrofurantoin (20ug/ul), and Gentamycin (10ug/ul). Single colonies were screened by PCR using the pEDV6 F primer and a gene specific R primer. All bacterial strains are described in Table 2.10. A list of primers used for cloning can be found in Table 2.11.

Culture Preparation

Bacterial cultures were prepared from single colonies and inoculated in 5 mL aliquots of KB media containing the appropriate antibiotic selection. The cultures were incubated at 28 °C for 18-22 h. The bacteria was pelleted and re-suspended in 10 mM $MgSO_4$. EtHAN and EtHAN-effector constructs were prepared at an OD = 0.6. For co-infiltration, EtHAN and *Psg* were mixed for at ODs = 0.6 and ODs = 0.3, respectively for a final OD = 0.9.

Plant Growth Conditions

G. max was planted in Fafner P2 potting media. About 1.5 kg Fafard P2 potting media was hydrated with 2 L of water with Cal-Mag fertilizer. 4 inch pots were packed firmly with media. 5 seeds were sown per pot and covered with about 2-3 cm of loosely packed potting media. The top soil was treated with Marathon insecticide, put into flats (8 plants per flat), and covered with clear plastic lids to maintain humidity during germination. After 24-36 h, the lids were removed from the flats. Germination rates were >95% for all *G. max* tested. *G. max* were germinated and maintained under diurnal light (16 h light at 22 ^{0}C; 6 h dark at 20 ^{0}C). Light intensity was maintained between 250-300 µM for all experiments.

Effector Infiltrations

Plants were infiltrated after 15-17 days (3 days post pruning). Bacterial solutions were prepared 1 hour prior to infiltration. *G. max* unifoliates from pruned plants infiltrate best 1 h after the lights turn on. Needless syringes were used to pressure infiltrate unifoliate leaves. Each unifoliate received 4 infiltrations of 100-200 uL of bacterial solution. The identical process is performed on the adjacent unifoliate. Co-infiltration of

EtHAN and *Psg* did not produce a visible HR on any tested *G. max* cultivar. The best negative control is to co-infiltrate EtHAN-effector and *Psg* into *G. max* cultivars which contain no *R* gene against the cloned effector. A list of soybean cultivars which were used for these experiments is detailed in Table 2.12.

Scoring

Seven days post infiltration (DPI), the unifoliate leaves were detached, imaged, and scored. A plant scored positive if ≥ 5/8 infiltration spots produced a visual HR on both unifoliate leaves (Figure 2.9b). Plants were scored as negative if ≤2/8 infiltration spots produced a visible HR (Figure 2.9b). Plants that fell between 3-4/8 were scored as ambiguous (Figure 2.9b). Any data, scored as ambiguous, was discarded, and the experiment was repeated until the appropriate sample size was reached. To calculate % HR, we used the equation, $\%\ HR = (\#\ of\ responding\ plants) \div (\#\ of\ infiltrated\ plants)$.

If too much pressure is applied during infiltration, the damage may induce a cell death response around edges of infiltration sites (Figure 2.9a). This can make scoring weak HRs more difficult. If the cell death response is weak, the infiltration site will turn brown with very small HRs throughout the site (Figure 2.9a). Cell death can occur on negative control plants if disease or insects are present (Figure 2.9c). The Compendium of Soybean Diseases was used as a reference to identify disease [169].

Determination of R *Gene Inheritance in* $F_{2:3}$ *Populations*

Plants from segregating $F_{2:3}$ families were screened with a single effector at a time. Prior to screening, we determined the background cell death each effector produced

on Williams. $F_{2:3}$ families were tested with an average sample size c. 15 individuals per family or line. Seed was limited.

Determining $F_{2:3}$ Phenotypic Ratios

After compiling the results, we analyzed the data as % HR with variation represented as standard deviation (σ). We determined our range for homozygous dominant using the equation, $RR = 100\%\ to\ (100\% - 1\ parental\ \sigma)$. We determined our range for homozygous recessive using the equation, $rr = 0\%\ to\ (Williams\ \%\ HR + 1\ Williams\ \sigma)$. Families that fell within the two ranges are considered to be segregating.

In planta *Bacterial Growth Curves*

All bacterial strains were grown at 28 °C on KB agar containing the appropriate antibiotic selection. Single colonies were inoculated in 5 mL liquid KB media containing antibiotic selection. The cultures were incubated at 28 °C for 18-22 h. Bacterial cultures were then pelleted by centrifugation at 2,000 x g for 15 minutes and resuspended in 3-5 mL 10 mM $MgSO_4$. Optical densities (OD) were calculated by measuring absorbance at 600nm (OD_{600}). Bacterial solutions were then diluted to appropriate ODs with 10 mM $MgSO_4$. In bacterial growth assays containing only one strain of bacteria, solutions were pressure infiltrated into soybean unifoliates using a needless syringe at an OD_{600} = 0.01. In the case of co-infiltration experiments, the target strain (EtHAN) for the assay was infiltrated at an OD =0.01 while the non-target strain (*Psg*) was infiltrated at an OD = 0.3. Leaf disks were collected at 0 days post inoculation (0 DPI) and 5 days post inoculation (5 DPI). Leaf disks were homogenized in 200 µL 10 mM $MgSO_4$ using 1 mm glass beads

in a bead beater. Samples were transferred to a sterile 96 well plate and diluted 1:10 serially. The dilution series was plated on KB agar containing the appropriate antibiotic selection. Plates were placed at 28 °C for 2 days at which point colony forming units (CFU) were counted (n = 5 plants / 10 leaves per replicate). Three biological replicates were performed.

Ion Leakage Assay

Soybean leaves were inoculated according to the methods for the visual hypersensitive response assay. After seven days, six circular leaf punches (1 cm in diameter) were taken from inoculated areas. Leaf disks were then floated in 45 mL sterile ddH_20 to reduce background ion leakage from cutting. After one hour, leaf disks were transferred to new tubes with 5 mL of water. After a three hour incubation, the water was removed and transferred to a 10 mL beaker. A handheld conductivity meter was used to measure the ions in solution. After conductivity readings were taken, the water was added back to the sample tubes. Samples were then boiled for 8 minutes to release all of the ions from the leaf disks. New conductivity readings were then taken to measure total ion leakage. To determine percent ion leakage, the initial conductivity reading was divided by the total ion leakage. Three biological replicates were performed (n = 3 plants / 6 leaf punches per replicate).

In vitro *and* in planta *Plasmid Ejection Assay*

Plasmid ejection of pEDV6 containing *P. sojae* avirulence genes *Avr*1K, *Avr*1B, or *Avr*4/6 was measured in all *Pseudomonas* strains used in this study. Bacterial colonies were grown in a 5 mL liquid KB cultures containing the appropriate antibiotic selection. Bacteria were pelleted and resuspended in 10 mM $MgSO_4$. For the *in vitro* assay, the

bacterial solution was left to incubate at 28 °C and 10 µL aliquots were plated every 15 minutes onto KB agar containing an antibiotic to select for the bacteria and Gentamycin to select for the pEDV6-effector plasmid. Three independent biological replicates were performed (n = 3 cultures per replicate). The plates were then incubated for two days at 28 °C and CFU were counted. The *in planta* bacterial growth curve described above was used to observe plasmid ejection after pressure infiltration of the culture into soybean leaves. For the *in planta* assay, time points were at 0 DPI and 3 DPI.

Effector Toxicity Assay

In order to test the toxicity of effector proteins *Avr*1K, *Avr*1B, and *Avr*4/6 on EtHAN and on *Psg*, bacterial cultures were grown in 5 mL liquid KB cultures containing the appropriate antibiotic selection. Cells were pelleted by centrifugation and diluted to an OD_{600}= 0.1 in 5 mL minimal media (50 mM KH_2PO_4, 5 mM K_2HPO_4, 1.7 mM $MgCl_2$*$6H_2O$, 7.6 mM $(NH_4)2SO_4$, 1.7 mM NaCl, 10 mM mannitol, and 10 mM fructose, filter sterilized) maintaining proper antibiotic selection. We measured OD_{600} at 5 hour intervals to assess effector impact on bacterial growth. Effector impact on growth was measured for each strain separately and compared to an untransformed control. Three independent biological replicates were performed (n = 5 cultures per replicate).

P. sojae *Mycelium Plug Infection Assay on Detached Trifoliate Leaves*

P. sojae was maintained on a PARP V8 plate. Three weeks prior to infection, a mycelium plug is transferred from a PARP V8 plate to a V8 plate. After three weeks, the culture should colonize most of the V8 plate. Contaminated plates should be discarded. Fresh mycelium plugs should be taken from freshly colonized V8 plates for the best results.

At 14 days post planting, trifoliate leaves should appear expanded. Per plant, remove each trifoliate and place each leaf in a wetted petri dish with a piece of filter paper. Experimental design will depend on leaf and culture sample size. On average, each petri dish will contain three leaves from three different individual plants. This allows for testing individuals with multiple pathogen isolates during a single replicate.

When ready to infect, scar each leaf in the center and place a 1cm mycelium plug on top of the scarred area. Parafilm the plates to maintain high humidity. Disease symptoms will begin at about 48 hours post infection. The symptoms conclude after seven days. Each culture should be passage every 1.5 months to maintain virulence.

Figures and Tables

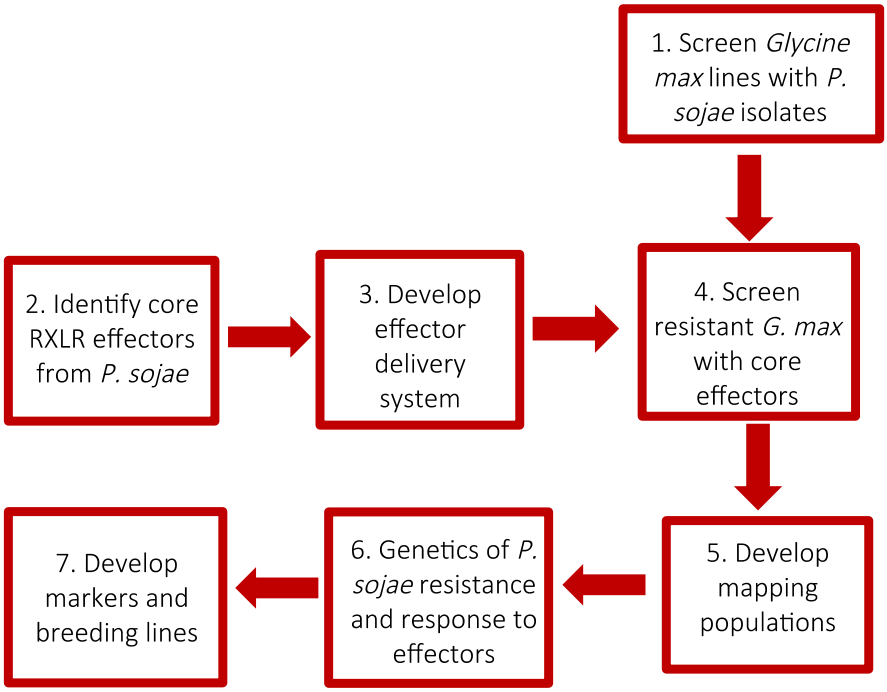

Figure 2.1 An overview of effector-directed breeding in soybean. This summarizes the development of soybean lines with *R* genes that recognize core *P. sojae* effectors. *Glycine max* germplasm is screened with a collection of isolates that collectively break all commercially available *Rps* genes. 2. Core effectors are identified with functional genomics tools. 3. A bacterial system to deliver those effectors individually to plants is developed. 4. The *P. sojae*-resistant germplasm from Step 1 is screened with the core effectors. 5. Lines with putative *R* genes against the effectors are crossed to a susceptible line, and progeny from the cross are used to develop mapping populations. 6. Genetic analysis is used to validate the phenotypes and identify lines in which effector recognition and pathogen resistance are conferred by the same locus. 7. Such lines will be used to develop breeding lines.

Table 2.1 The PI numbers for all referenced accessions are listed.

G. max **Parental Accession**	**PI Number**
gm13	PI407985
gm20	PI408319C
gm31	PI398440
gm32	PI567139B
gm41	PI200553
gm48	PI274508
gm78	PI398666
gm92	PI398775
gm94	PI398791
gm96	PI398946
gm98	PI398996
gm99	PI399004
gm106	PI399036
gm121	PI408287
gm123	PI423741
gm144	PI424477
gm291	PI399079
gm314	PI408015
gm315	PI408020A
gm317	PI408029
gm320	PI408097
gm321	PI408111
gm326	PI408132

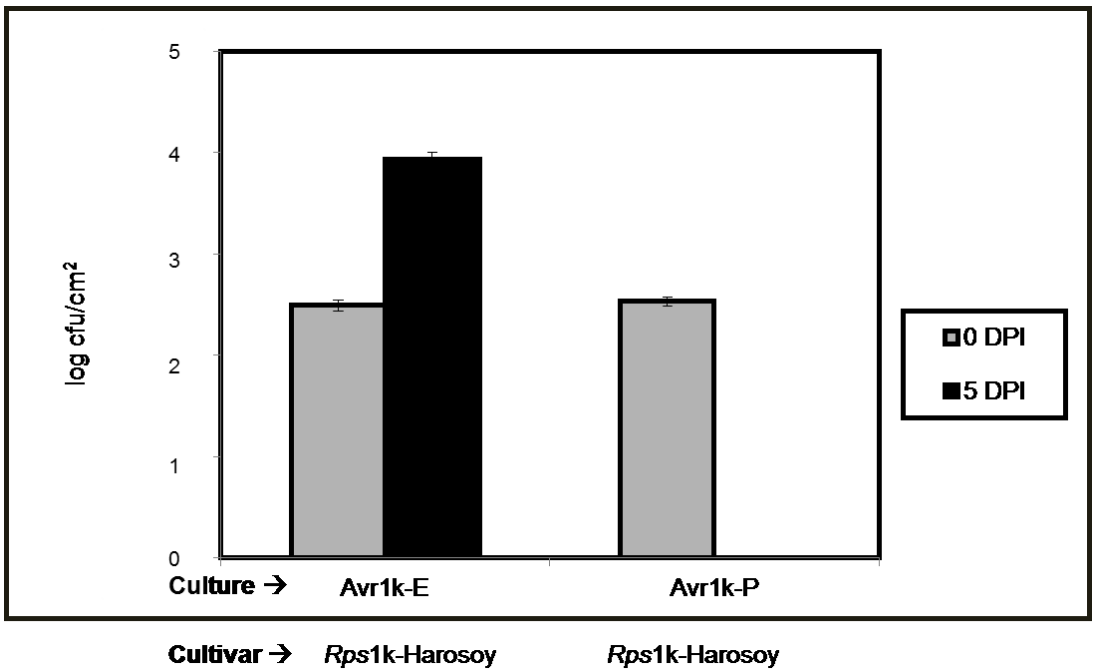

Figure 2.2 *Psg* likely ejects the *Avr*1k-pEDV6 plasmid prior to or during effector infiltration. *In planta* growth assay with *Psg* race 4 carrying *Avr*1k (*Avr*1k-P) and EtHAN carrying *Avr*1k (*Avr*1k-E) on HARO15. We recovered no *Avr*1k-P at 5 DPI. This data was produced from one biological replicate (n = 5 plants / 10 leaves per culture). The standard deviation represents leaf-to-leaf variation. Two other biological replicates produced similar results.

Table 2.2 EtHAN is the only strain that does not eject pEDV6 containing *Avr*1k, *Avr*1b, or *Avr*4/6. After a 15 minute incubation in 10mM $MgSO_4$, strains were plated onto selective media. Only *P. f.* EtHAN retained the plasmids. Grey shading indicates plasmid ejection. Black indicates plasmid retention. This data was produced from three independent biological replicates (n = 3 cultures per replicate).

Pseudomonas **Strains**	**Plasmid Ejection**		
	***Avr*1k**	***Avr*1b**	***Avr*4/6**
P. f. EtHAN			
P. syringe DC3000			
P. syringe DC3000 D28E			
P. syringe DC3000 D28E + AvrPtoB			
P. s. glycinea race 4			

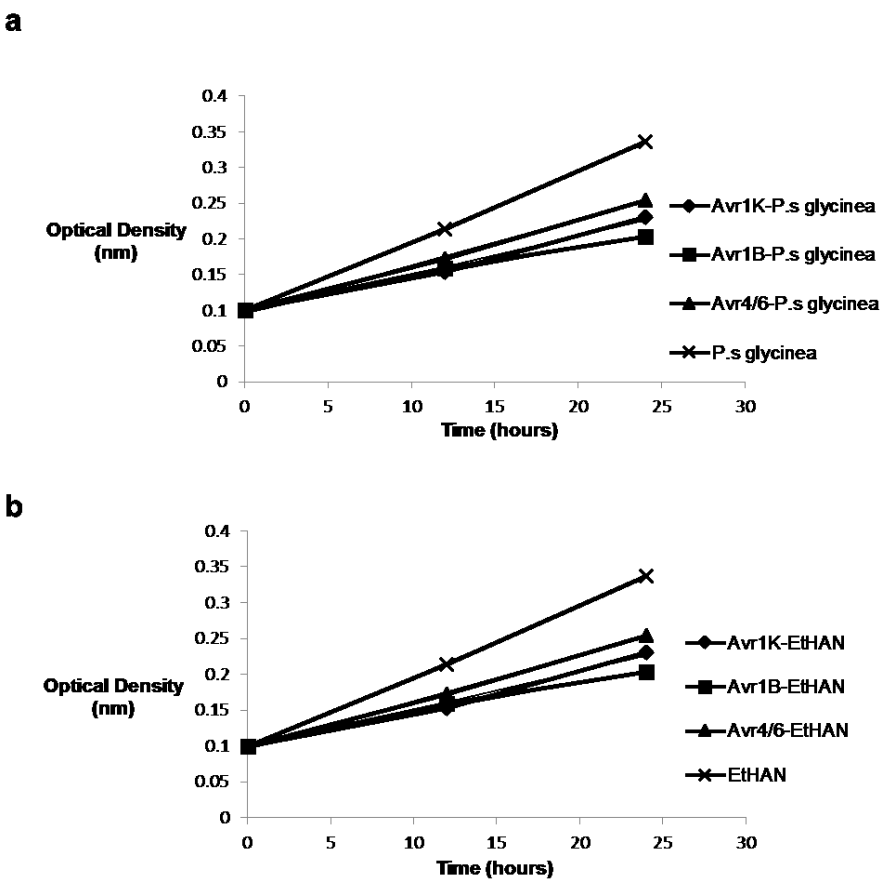

Figure 2.3 *P. sojae* effector toxicity on *Pseudomonas*. Bacterial strains carrying *Avr*1k, *Avr*1b, or *Avr*4/6 were cloned into pEDV6 and were grown in minimal media with antibiotic selection for 24 hours. Minimal media stress activates effector production and antibiotic selection forces plasmid retention. *P. sojae* effector expression inhibited the growth of a) *Psg* and b) *P. f.* EtHAN. This data was produced from one replicate (n = 5 cultures per replicate). Two other biological replicates produced similar results.

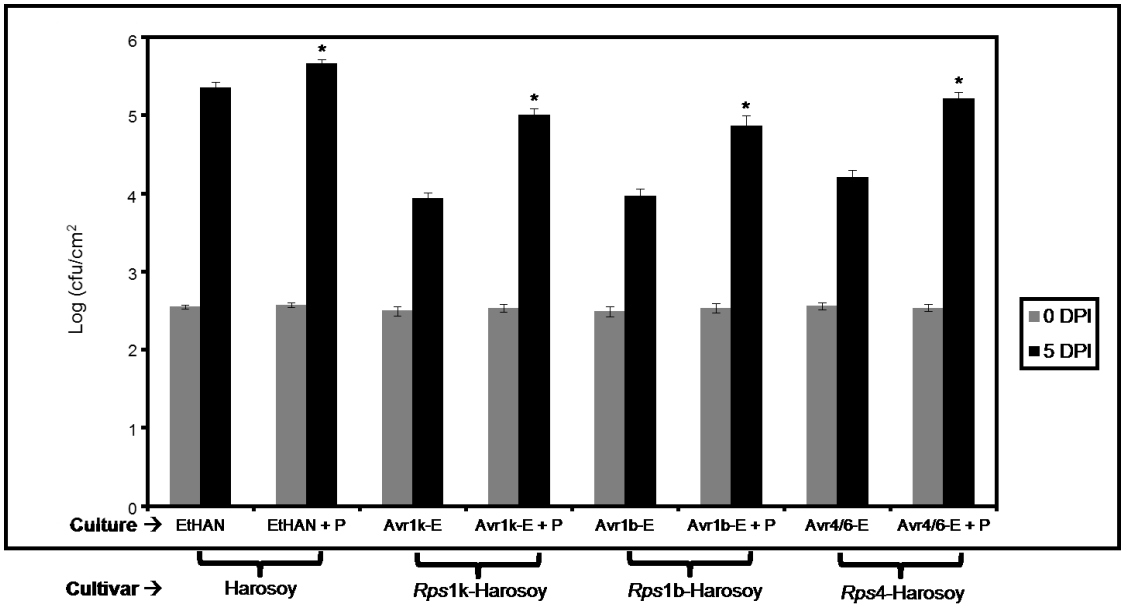

Figure 2.4 Co-infiltration with virulent *Psg* race 4 improves the growth of EtHAN *in planta*. EtHAN or EtHAN with the indicated *Avr* genes was infiltrated into the indicated plant cultivars that carry *R* genes against the cognate *Avr* effector proteins. The addition of *Psg* race 4 to a culture is indicated by a "+ P". These data were produced from one biological replicate (n = 5 plants / 10 leaves per culture). Standard deviation represents leaf-to-leaf variation. Statistical significance of *Avr-Psg* race 4 CFU was determined by a pairwise t-test against the cognate *Avr-P. f.* EtHAN CFU ($p < 0.05$) and is denoted by an *. Two additional biological replicates produced similar results.

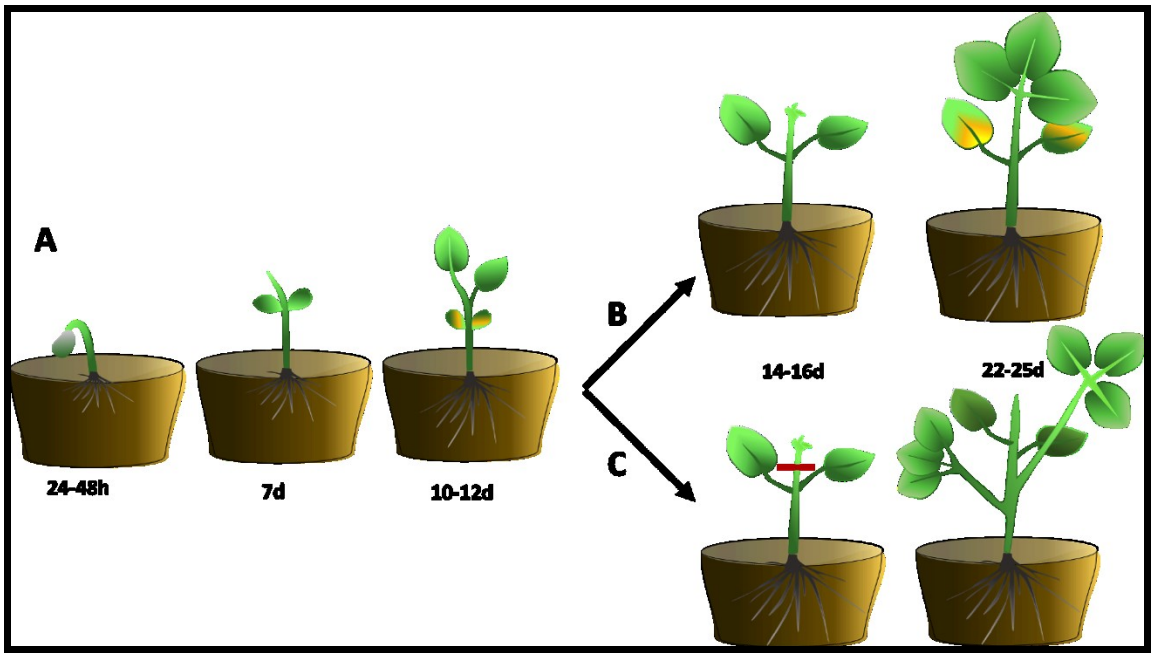

Figure 2.5 Soybean growth and pruning techniques for optimizing effector-response. *G. max* plants germinate, first producing cotyledons. **a)** Unifoliate leaves emerge after 10-12 days at which point the cotyledons senesce. **b)** The plant will then begin producing trifoliate leaves which will mature, leading to senescence of unifoliate leaves. **c)** If the immature trifoliate leaves are pruned between days 14-16 (Red Hash Mark) the unifoliate leaves will become thicker and darker. New inflorescence will be generated below the point of pruning.

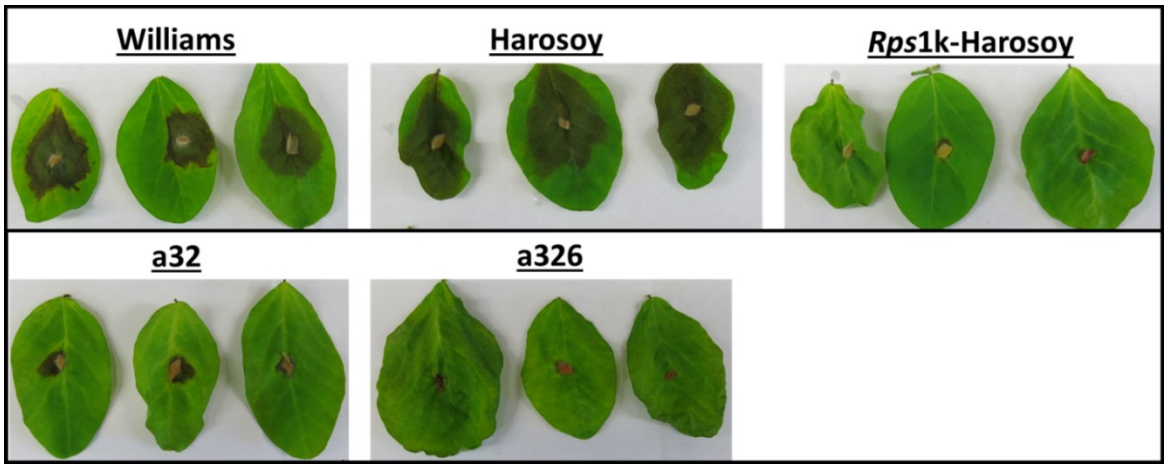

Figure 2.6 A simple *P. sojae* mycelial plug assay for testing detached soybean leaves. We detach trifoliate leaves after the V2 growth stage. These trifoliate leaves are inoculated with *P. sojae* plug. The infection is scored at 7 DPI. *P. sojae* produced large brown/black legions on susceptible cultivars Williams and Harosoy in all cases. *Rps*1k-Harosoy is included as a positive control for resistance against *P. sojae* race 2 and showed no expanded disease legion. *P. sojae* race 2 did not produce expanded disease legions on gm32 or gm326 suggesting that they are resistant to this isolate. This figure was produced from one biological replicate (n = 3 plants / 3 leaves per replicate). Two additional biological replicates produced similar results.

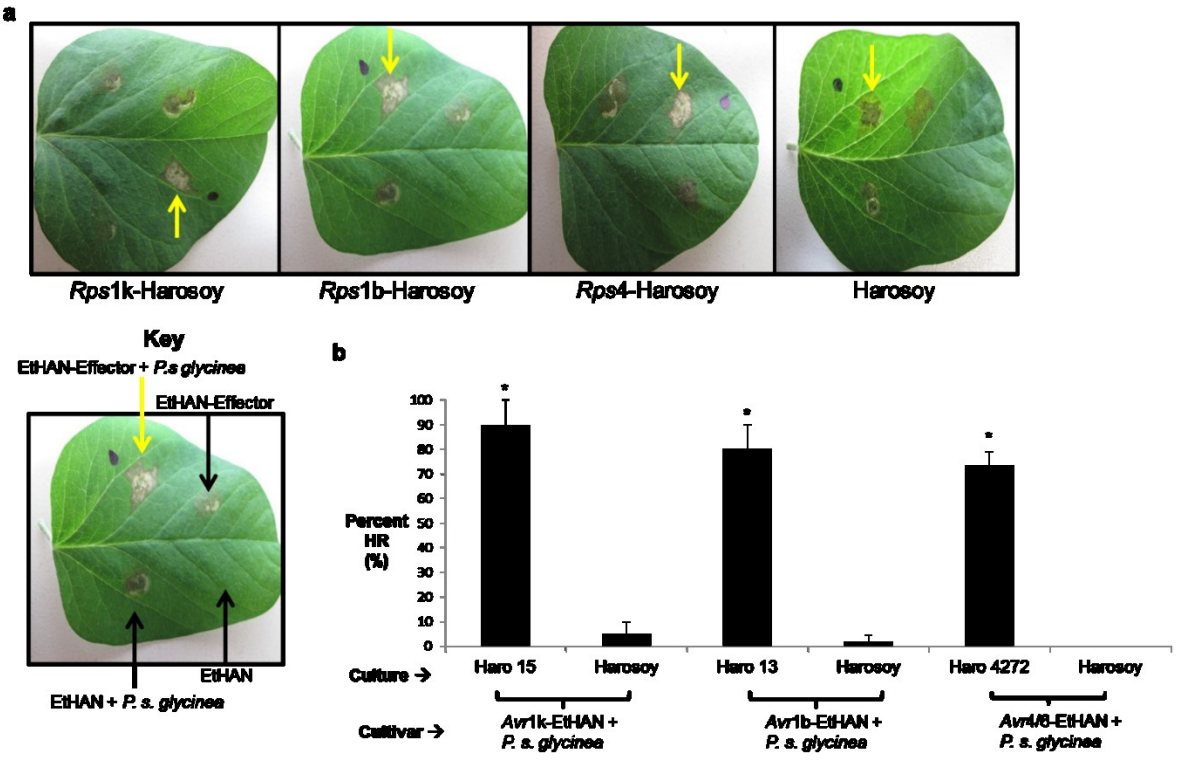

Figure 2.7 Validating the *Pseudomonas* effector delivery system using known *Avr/R* interactions. **a)** *P. sojae* avirulence proteins trigger gene-for-gene resistance when delivered from *P. fluorescens* by Type III secretion. The key depicts the infiltration strategy. The yellow arrow indicates the sites at which EtHAN expressing the avirulence effector was co-infiltrated with *Psg* race 4. The adjacent infiltration is EtHAN expressing the avirulence effector without *P. s. glycinea* race 4. The negative controls for each treatment were infiltrated on the opposite side of the leaf. **b)** The graph quantifies the amount of times a visual HR was produced on a cultivar, represented as % HR. Harosoy was used as a negative control. The error bars depict standard deviation between the three biological replicates (n = 10 plants/ 20 leaves per replicate). Statistical significance was determined by a pairwise t-test against Harosoy ($p < 0.05$) and is represented by an asterisk.

Table 2.3 Validating effector delivery using *Avr*1k, *Avr*1b, and *Avr*4/6 on *Rps* cultivars in the Williams background. These effectors produce about 30% background cell death on cultivar Williams. Three independent biological replicates were performed (n = 10 plants / 20 leaves per replicate). Standard deviation was calculated between the three replicates.

G. max Cultivar	**% HR *Avr*1k**	**% HR *Avr*1b**	**% HR *Avr*4/6**
Williams82	84 ± 11	28 ± 14	30 ±6
L77-1863	46 ± 7	72 ± 16	28 ± 11
L85-2352	22 ± 8	27 ± 6	68 ± 18
Williams	34 ± 11	25 ± 6	32 ± 6

Table 2.4 A EtHAN and *Psg* race 4 mixture of bacteria do not produce a HR on tested cultivars. Three independent biological replicates were performed (n = 10 plants / 20 leaves per replicate). Standard deviation was calculated between the three replicates.

Isoline	**Resistance Gene**	**EtHAN + *Psg* race 4**
Williams 82	Rps1k	8 ± 2
Haro 15	Rps1k	4 ± 1
L77-1863	Rps1b	12 ± 5
Haro 13	Rps1b	2 ± 1
L85-2352	Rps4	13 ± 3
Haro 4272	Rps4	0

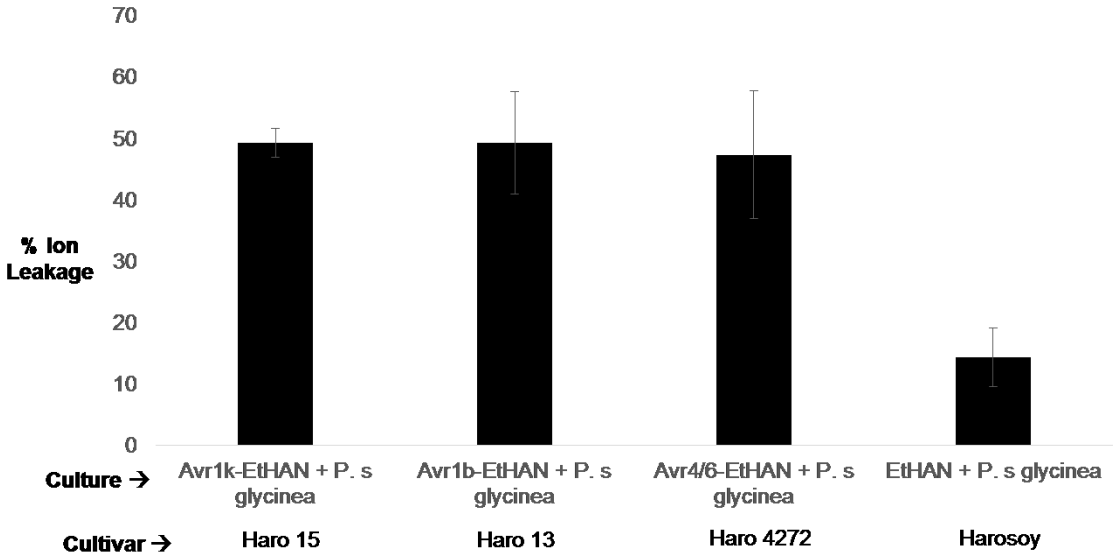

Figure 2.8 Ion leakage assay to quantify cell death symptoms in visual assays. *Avr*1k, *Avr*1b, and *Avr*4/6 were inoculated onto cultivars which contained a cognate *R* gene. EtHAN + *Psg* race 4 on Harosoy was used to measure background cell death. Three independent biological replicates were performed (n = 3 plants / 6 leaf punches per replicate). Standard deviation was calculated between the three replicates.

Table 2.5 Soybean germplasm respond to core *P. sojae* effectors, *Avh*16, *Avh*180, and *Avh*240. Responses of 30 *G. max* accessions to core effectors *Avh*16, *Avh*180, and *Avh*240. Data are presented as percentage of infiltration sites responding with a macroscopic HR, with standard deviation between the three biological replicates (n = 10 plants / 20 leaves per replicate).. Statistical significance was determined by pairwise t-test against Harosoy ($p < 0.05$). No background cell death was recorded for the *P. sojae* resistant *G. max* germplasm.

***G. max* Parental Accession**	***Avh*16 % HR**	***Avh*180 % HR**	***Avh*240 % HR**	***Avh*16 T-Test (p)**	***Avh*180 T-Test (p)**	***Avh*240 T-Test (p)**
99	54 ± 7	0	0	P<0.001	0	0
321	51 ± 28	0	0	P<0.04	0	0
48	67 ± 31	0	0	P<0.03	0	0
144	70 ± 24	0	0	P<0.02	0	0
31	0	86 ± 4	0	0	P<0.0003	0
326	0	86 ± 16	0	0	P<0.005	0
314	0	70 ± 10	0	0	P<0.003	0
20	0	0	60 ± 10	0	0	P<0.004
13	0	0	73 ± 21	0	0	P<0.01
32	88 ± 11	0	93 ± 7	P<0.001	0	P<0.0006
291	0	73 ± 15	77 ± 15	0	P<0.005	P<0.006
78	0	90 ± 10	91 ± 9	0	P<0.002	P<0.001
320	68 ± 16	63 ± 6	97 ± 6	P<0.008	P<0.001	P<0.0005
Harosoy	3	0	0	0	0	0

Table 2.6 Homologous effectors, *Avh*16 and *Avh*7a, provoke similar response patterns in resistant germplasm. We identified *Avh*7a as a homolog of *Avh*16. Resistant *G. max* accessions which responded to *Avh16* with a HR, were tested with *Avh7a* as well. We obtained similar results in responding cultivars. This data was produced from three biological replicates (n = 10 plants / 20 leaves per replicate). Standard deviation was calculated between the three replicates.

***G. max* Parental Accession**	***Avh*16 % HR**	***Avh*7a %HR**
99	54 ± 7	87 ± 6
321	51 ± 28	100 ± 0
48	67 ± 31	67 ± 6
144	70 ± 24	63 ± 12
32	88 ± 11	93 ± 6
320	68 ± 16	100 ± 0

Table 2.7 Response of *G. max* germplasm to twelve core *P. sojae* effectors. Blank cells indicate that the accession did not respond to the corresponding effector with an HR % >30. The delivery of *Avh*23 and *Avh*110 to the resistant germplasm did not produce a HR on any accession. Three independent biological replicates were performed (n = 5 plants / 10 leaves per replicate). Standard deviation was calculated between the three replicates.

***G. max* Parental Accession**	**Response to Core *P. sojae* Effectors**									
	***Avh*16**	***Avh*180**	***Avh*240**	***Avh*53**	***Avh*137**	***Avh*261**	***Avh*488**	**Snel1**	***Avh*94**	***Avh*241**
13			73 ± 21						55 ± 16	84 ± 10
20			60 ± 10							
31		86 ± 4		82 ± 8					71 ± 16	
32	88 ± 11		93 ± 7							
40				83 ± 10			85 ± 8			
41				85 ± 8	81 ± 14					
48	67 ± 31								86 ± 4	
78		90 ± 10	91 ± 9							
92					77 ± 7					
94				55 ± 6	62 ± 10	81 ± 6				
96					81 ± 6	84 ± 15				
98									72 ± 10	
99	54 ± 7									
106				86 ± 8	60 ± 24					
121				78 ± 6		62 ± 14				80 ± 10
123				54 ± 26		76 ± 14				
144	70 ± 24									
291		73 ± 15	77 ± 15		56 ± 10		66 ± 14			
314		70 ± 10			83 ± 6					
315									77 ± 12	
317								84 ± 8		60 ± 24
320	68 ± 16	63 ± 6	97 ± 6							
321	51 ± 28				62 ±14					
326		86 ± 16					66 ± 20			

Table 2.8 *G. max* accession which responded to core effectors were bred with susceptible Williams. We tested multiple individuals from segregating $F_{2:3}$ populations for a response to the effector with a HR. The tested $F_{2:3}$ populations have a * next the cross name.

Cross	ID	*Avh*16	*Avh*180	*Avh*240	Maturity Group
32(R) x Williams(S)*	cr32				V
326(R) x Williams(S)*	cr326				IV
144(R) x Williams(S)	cr144				IV
284(R) x Williams(S)	cr284				IV
291(R) x Williams(S)	c291				IV
13(R) x Williams (S)	cr13				IV
320(R) x Williams(S)	cr320				IV

Table 2.9 Effector-response phenotypic ratios in $F_{2:3}$ populations. We screened segregating $F_{2:3}$ families for the phenotype of response to an effector with a HR. We observed a 1:2:1 phenotypic ratio in all crosses. Four independent biological replicates were performed (n = 5 plants / 10 leaves per replicate). Statistical significance of our observed ratio was determined using a chi-squared test against a 1:2:1 expected ratio ($x > 0.05$). The observed ratios are not statistically significant suggesting that the response to the effector is provided by a single, dominant locus.

Effector	Cross	Number of F3 Families			Expected Ratio	x^2 value (2 df)
		HmR	Seg	HS		
*Avh*180	*G. max* 326(R) x Williams(S)	14	27	13	1:2:1	0.071
*Avh*16	*G. max* 32(R) x Williams(S)	21	45	23	1:2:1	0.101
*Avh*240	*G. max* 32(R) x Williams(S)	19	47	17	1:2:1	1.56

Table 2.10 Bacterial strains and their corresponding selectable markers are listed. All *Psuedomonas* strains were grown at 28°C and all *E. coli* strains were grown at 37°C.

Bacterial Strain	**Selectable Marker**	**Temperature (°C)**
Top 10 *E. coli*	None	37
P. fluorescens, EtHAN	Chloramphenicol	28
E. coli, RK600	Chloramphenicol	37
P. s pv glycinea race 4	Riffampicin	28
P. syringae DC3000	Tetracyclin	28
P. s DC3000, D28E	Riffampicin / Spectinomicin	28
P. s DC3000, D28E + *Avr*PtoB	Riffampicin / Spectinomicin	28

Table 2.11 The diagnostic primers used screen for positive transformants into pEDV6 are listed. In all cases, the pEDV6 F primer was used with a gene specific *R* primer to confirm the insert and proper directionality.

Primer	**Sequence (5'-3')**
pEDV6 F	GGCACCCCAGGCTTTACACTTTATG
*Avr*1K R	TCAGATAATCATGATGCTGT
*Avr*1B R	CCGGTGAAAGGTGTATCCGTTGTAG
*Avr*4/6 R	CGTTAGGTGGTGTAGTCCGACGGAC
*Avh*16 R	CCATCTTCTTTGCTTCCTTAGC
*Avh*7a R	CTACAAGCCGTCCGTGTTCATGCCCA
*Avh*180 R	CTAAGCGATGTTCGTCTG
*Avh*240 R	CTAGTTTGCGGGTTGG
*Avh*23 R	TCATGCATTGTCGGAAAGTTTGAAGT
*Avh*110 R	TTATCCTACGCGGACACTCGCTACGC
*Avh*53 R	GTTATTTTCGTTAGCCCCATA
*Avh*137 R	CGGGTAAATGTATCCCTCAAG
*Avh*261 R	AATGTGATTTTGCGGGTTGTC
*Avh*488 R	CGGCGACTGGCCCATGCGAGC
Snel1 R	CGATCGCCGGTGCTGACGACT
*Avh*94 R	CTAAGCGGTGTCCTCCTCCTCC
*Avh*241 R	GTTCAGCTCGTGCCACTTGAA
*Avh*7a R	CTACAAGCCGTCCGTGTTCATGCCCA

Table 2.12 Williams and Harosoy isolines are listed.

Resistance Gene	Cultivar	Isoline
Rps1k	Williams	Williams 82
Rps1k	Harosoy	Haro 15
Rps1b	Williams	L77-1863
Rps1b	Harosoy	Haro 13
Rps4	Williams	L85-2352
Rps4	Harosoy	Haro 4272

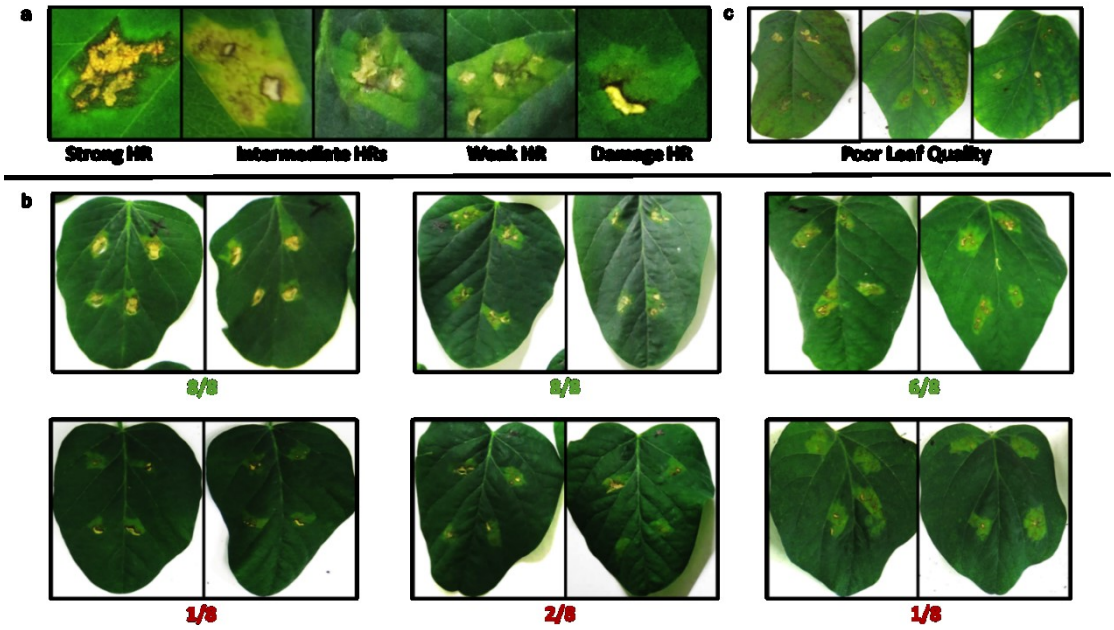

Figure 2.9 Effector-response was scored by presence or absence of a HR. **a)** The strength of the effector response (aka Strong/Medium/Weak) was not recorded for individual infiltration site. **b)** During all screens, each plant received 8 effector inoculations, 4 per unifoliate leaf. Each of the 8 infiltration spots were scored as either 1 or 0 for the presence or absence of a HR. If ≥5 individual sites produced a HR, the plant was considered a positive responder. If ≤2 spots produced a HR the plants were negative. If between 3-4 infiltration spots responded, the data was considered too ambiguous to score.

Chapter 3

Development of an Effector-Based Screening System to Identify Novel R *Genes in a Wild Relative of Cultivated Soybean*

Michael G. Fedkenheuer[1], Kevin E. Fedkenheuer[1], Colin L. Davis[1], Brett M. Tyler[2], M. A. Saghai-Maroof[3], and John M. McDowell[1*]

[1]Department of Plant Pathology, Physiology, and Weed Science, Latham Hall, Virginia Tech. Blacksburg, VA 24060-0390

[2]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR

[3]Department of Crop, Soil, and Environmental Sciences, Latham Hall, Virginia Tech, Blacksburg, VA 24060-0390

[*]for correspondence, johnmcd@vt.edu

In preparation for submission to *MPMI*

Author Contributions: MF wrote and developed this manuscript with advice from JM and KF. JM, BT, and SM advised this project. The majority of effector-based screening data was generated by MF. KF helped with development of effector-based screening in *G. soja* by providing technical expertise throughout experiments for troubleshooting and optimization. CD was involved with screening the segregating population and helped develop the batch style scarification technique.

Abstract

*T*he oomycete *Phytophthora sojae* causes a destructive root and stem rot disease of cultivated soybean (*Glycine max*). Over time, commercially available resistance (*R*) genes are losing efficacy against *P. sojae*. We hypothesized that *Glycine soja*, a sexually compatible wild relative of *Glycine max*, could be exploited as a useful source of new *R* genes against soybean pathogens. We developed a system to screen *G. soja* with core RXLR effectors from *P. sojae* to identify novel *R* genes. We focused our screening on conserved RXLR effectors that are essential for virulence. *R* genes that recognize core effectors are expected to be durable, because mutations in the effector genes will result in a fitness penalty for the pathogen. We screened 24 *P. sojae*-resistant *G. soja* accessions with individual RXLR effectors delivered by *Pseudomonas fluorescens* Type III secretion. We identified several *G. soja* accessions for which delivery of one or more core effectors triggered a localized cell death, suggestive of effector-triggered immunity. We developed a segregating population from a cross of one *G. soja* line to a susceptible *G. max* line, to track inheritance of the *R* gene. Response to the effector *Avh*240 segregated in a 1:2:1 phenotypic ratio in the $F_{2:3}$ generation suggesting that this line contains one dominant *R* gene against the core *P. sojae* effector *Avh*240. These results indicate that an effector-based screening approach can be utilized to identify useful new *R* genes from a wild relative of soybean.

Introduction

Soybean is an important crop which accounts for 60% of world oilseed production [180]. The United States is a world leader in the production of various soybean products such as oil, crush, and meal [181]. Soybean production and trade continues to increase globally and in the US [181]. *Phytophthora* root and stem rot is one of the most destructive disease affecting soybean yield [182]. In 2006, losses associated with *Phytophthora* root and stem rot in the United States were 1,464 metric tons, equivalent to nearly $500 million USD [182].

Phytophthora sojae (*P. sojae*), the causal agent of this disease, is responsible for pre- and post-emergence damping off of susceptible seedlings [149]. On mature soybean plants, *P. sojae* causes root and stem browning and leaf chlorosis [169]. The disease symptoms generally appear 1 to 2 weeks after heavy rainfall [149]. *P. sojae* produces sporangia and oospores during infection. Sporangia release motile zoospores that swim to root exudates through wet soil. Zoospores then encyst and penetrate roots, producing hyphae that grow in the spaces between cells [183]. The filamentous hyphae create specialized feeding structures, known as haustoria, that draw nutrients out of the cell [151]. Additionally, these structures function as an interfaces where virulence factors, called effectors, are secreted to subvert immunity and reprogram the cell for feeding [151]. Oospores are thick-walled sexual structures produced in plant tissue that can survive for many years dominant in the soil and endure winter frosts. Oospores germinate in the spring and infect emerging seedlings. Heavy rains will induce oospore germination making this disease very volatile in seasons with extended periods of flooding, which typically initiates disease epidemics [183].

P. sojae secretes several different types of effector proteins to subvert plant immunity and alter host cell structure and function [62, 64]. The best-studied effector protein superfamily, called RXLR effectors, are defined by an N-terminal secretion signal, followed by the consensus amino acid motif RXLR [184]. The RXLR motif is often, but not always, followed a short motif enriched for acidic amino acids ("DEER") and by so-called "W" and "Y" domains that form helical "WY folds" which are unique to RXLR effectors in the *Phytophthora* and downy mildew families [62, 185]. These domains can be arranged to form complex folds, which make the WY fold a highly evolvable platform that supports structured surface polymorphisms upon rearrangement [64]. Virulence targets of RXLR proteins have been validated as key regulators of immunity, and it is becoming increasingly clear that RXLR proteins play a major role in promoting virulence of oomycetes in the *Phytophthora* genus [186].

Despite their importance for virulence, RXLR proteins can also be detrimental to the pathogen if recognition of an RXLR effector by a plant *R* protein leads to gene-for-gene resistance (more recently termed "effector triggered immunity" or ETI [187]). Indeed, every case of ETI against *Phytophthora* species, characterized at the molecular level, involves recognition of an RXLR protein by a nucleotide-binding, leucine-rich repeat immune receptor protein [64]. ETI is often characterized by localized cell death (termed the hypersensitive response or HR) which impedes pathogen growth [187]. This type of resistance is generally regarded as one of the most cost-effective tools to manage diseases caused by *Phytophthora* pathogens.

In soybean, there are 24 known Resistance to *Phytophthora sojae* (*Rps*) genes that provide gene-for-gene resistance against *P. sojae* infection. Seven of these genes have

been commercially deployed with varying degrees of effectiveness against the most common *P. sojae* pathotypes [43, 188]. Unfortunately, even the most effective *Rps* genes are now being overcome by pathogen co-evolution. This co-evolution produces diverse *P. sojae* strains ("pathotypes") characterized by different patterns of pathogenicity on *Rps* soybean cultivars. In other words, a soybean cultivar containing a *R* gene against *P. sojae* may be resistant to one pathotype and susceptible to a different pathotype [43]. It would be desirable to identify *R* genes with resistance to most or all of the *P. sojae* pathotypes encountered in field situations.

Genome analyses of oomycete phytopathogens is revealing the mechanisms behind the rapid evolution of pathotype diversity. *Phytophthora* genomes contain large, variable collections of RXLR-encoding genes which in turn has given rise to multiple *P. sojae* pathotypes [188]. Pathogen genomes are often organized by a bipartite distribution of fast and slow evolving genetic elements. It is hypothesized that pathogens create a two 'speed-speed genome' to modulate the evolutionary speeds of genes [189]. Effector genes are located in a part of the genome where rates of polymorphisms, copy-number variation, and recombination are extremely high [190]. Effector genes can be deleted or altered to avoid *R* gene detection. Due to genomic instability, expression of effectors is often variable. Transcriptional changes can cause effectors expression to fall below detection thresholds [64]. Polymorphisms that affect the coding sequence can cause effectors to have variability or off target catalytic activity. By this mechanism effector function can be changed or modulated to achieve new virulence mechanisms [64]. The proximity of RXLR effectors to transposable elements suggest a potential epigenetic adaptation to globally silence a group of RXLR effectors under environmentally

favorable conditions [191]. These silencing events could potentially be reversed if the environmental pressures for the silencing are reduced, thereby providing the pathogen with a powerful mechanism for rapid co-evolution. Altogether, these mechanisms illustrate why host resistance to *Phytophthora* species is short-lived, and underscore the challenges of breeding soybean with durable resistance to *P. sojae* field isolates with diverse pathotypes, as are frequently encountered in the field [192].

Wild relatives of domesticated crop species often contain useful agronomic traits such as tolerance to abiotic stresses, yield enhancement, and disease resistance [99]. These traits can be bred into commercial cultivars by conventional breeding if the relative is sexually compatible with the crop or if somatic cell hybrids and bridge species can be used to circumvent compatibility issues. Working with sexually compatible species is advantageous because it simplifies the breeding process. There have been successful reports of breeding resistance from wild relatives into commercial varieties. One of the first examples of introgression of resistance from a wild species was in potato [102]. Since then, disease resistance traits have been introgressed from wild species in most major crop plants.

Soybean was domesticated 6,000-9,000 years ago in Eastern Asia [193]. Although many aspects of domestication remain unclear, *Glycine soja* (*G. soja*) is accepted as the closest known wild relative to *G. max* [193]. *G. max* seed contains about 17% more oil than *G. soja* seed. *G. max* germinates consistently to produce uniform plants with large seed and thin seed coats. *G. soja* produces small, vine-like plants with small seed and thick seed coats. Despite these phenotypic differences, *G. max* and *G. soja* are sexually compatible [169], such that useful traits from *G. soja* can, in principle, be bred into

soybean cultivars. Many studies of soybean genetic diversity suggest that most landraces of *G. max* are monophyletic, and *G. soja* germplasm is reported to contain more genetic diversity than *G. max* germplasm. Wen et al. and Joshi et al. both hypothesized six *G. soja* genetic subgroups [194, 195]. The subgroups are separated geographically and correspond to Japan, Korea, Russia, northeast China, southern China, and the Huanghuai Valleys in China. Long-term seed dispersal is rare and does not appear to impact population structure of native populations.

Resistance against soybean cyst nematode [196], soybean mosaic virus [197], and *Phytophthora* root and stem rot [101] have been reported in *G. soja*. There are no reports of using directed breeding to intentionally introgress disease resistance traits from *G. soja* to *G. max*. Using molecular diagnostics, *G. soja* can be exploited as a source of resistance genes against *P. sojae*. Our goal is to identify novel and durable *R* genes against *P. sojae* in *G. soja* accessions, using an approach that combines traditional, pathogen-based screens with effector-based screen in which plant germplasm is probed for immunity triggered by delivery of individual effectors. This effector-based approach was pioneered against *Phytophthora infestans* [160]. 54 RXLR effectors from *P. infestans* were assayed for avirulence activity, by delivering effectors one-at-a-time to *Solanum* germplasm via *Agrobacterium*-mediated transient expression. In this system, *Agrobacterium* suspensions were infiltrated into plant leaves, which were then screened for a macroscopic cell death response that could result from hypersensitive cell death triggered by recognition of the effector. Vleeshouwers et al. used wild *Solanum* species to identify the *Rpi*-blb1 *R* gene using this approach. The avirulence response was triggered by the IpiO RXLR family [160]. More recently, Oh et al. bred novel *R* genes *Rpi*-blb1 and *Rpi*-blb2 against potato

late blight from a wild *Solanum* species into potato [198]. Efforts are now underway to commercialize these genes in potato. This example illustrates how useful *R* genes were discovered by probing wild *Solanaceae* with RXLR effectors.

Following the *P. infestans* example, our approach is based on screening with RXLR effectors; however we focused our approach on "core effectors" that are highly expressed early in infection, conserved and monomorphic among the four most common *P. sojae* pathotypes, and required for virulence. By focusing on effectors that are conserved and contribute most to virulence, we hope to find resistance genes that target the pathogen's most static or most vulnerable points. This would make it more difficult for the pathogen to evolve quickly by mutating or silencing the gene and activating genes with redundant functions. In addition, *R* genes that recognize conserved effectors are more likely to provide recognition and resistance to a broad range of field isolates.

This study builds on a recent effort in which comparative genomics and transcriptomics were used to identify RXLR effector genes in *P. sojae* which were highly expressed early in infection and conserved among the four most common *P. sojae* pathotypes [199, 200]. Then, transient gene silencing was used to suppress expression of candidate core *P. sojae* effectors. The silencing of core *P. sojae* effectors *Avh*16, *Avh*180, and *Avh*240 greatly reduced pathogen growth on the susceptible cultivars, while the silencing of 7 other effectors *Avh*53, *Avh*137, *Avh*23, *Avh*110, *Avh*488, SNEL1 and *Avh*241showed moderately reduced pathogen growth. These are the effectors that we used in the current study.

Our screening strategy was based on an effector-based screening approach that was developed in *G .max* (Chapter 2). Importantly, this study developed an alternative

system for effector delivery (see below). We were able to effectively optimize techniques developed in this study for use in wild soybean. After normalizing the variables of germination and growth stage in wild soybean we found that the effector-based screening was suitable for use with minor technical modifications. Using effector-based screening we were able to identify putative resistance genes in accessions containing novel resistance to *P. sojae*. Additionally, we were able to confirm simple inheritance of one putative *R* gene in segregating populations. This suggests that this technique can be used to identify *R* genes in *G. soja* and is sensitive enough to map resistance in crosses between *G. soja* and *G. max*, thereby validating its utility for disease resistance breeding.

Results

Overview of the Strategy

Our approach for delivering effectors was based on a screening system described in Fedkenheuer et al. 2016 (Chapter 2). Because *Agrobacterium*-mediated transient transformation is not efficient in soybean, we developed a system for delivering effectors via Type III secretion from *Pseudomonas*. Trials with different *Pseudomonas* species revealed that the best strain for delivery in soybean is the soil bacterium *P. fluorescens*, EtHAN. EtHAN contains an artificially engineered type III secretion system which can be utilized to deliver individual effectors. *P. sojae* effectors were cloned into the Effector Detector Vector (pEDV6) that contains the bacterial *AvrRps*4 promoter and leader sequence which guides proteins through the type III secretion system [167]. We co-infiltrated EtHAN with *P. syringae pathovar glycinea* (*Psg*) for effector-based screens to improve signal intensity. *Psg* has been shown to improve the growth of EtHAN *in planta* presumably by suppressing host immunity (Chapter 2). Consequential to improved growth, more effector protein is likely delivered. If a host *R* gene recognizes the delivered effector, a macroscopic HR is triggered that can be scored visually.

In this study, we validated that the EtHAN-effector system efficiently delivers core *P. sojae* effectors into *G. soja* unifoliate leaves. Recently, Matthiesen et al. screened ~1,000 Glycine accessions with a mixture of three *P. sojae* isolates which defeat 13 commercial *Rps* genes [101]. From these accessions, we chose 24 *G. soja* accessions for effector-based screening. These resistant lines are likely to contain novel *R* gene loci or

new alleles at known loci. The effector-based approach described below provides a mechanism through which to sort and prioritize these gene for breeding.

Optimizing the Screening System

The first objective of this study was to optimize growth conditions for *G. soja*, because our experience with this screen in *G. max* demonstrated that robust plant health is of key importance to obtaining reliable results from bacterial inoculations. This required several modifications of the protocols developed for the *G. max* screen .*G. soja* seeds are small and are coated with a thick outer wall. Germination of *G. soja* seed requires removal of the seed coat (scarification). We found that *G. soja* seeds germinated best when scarified in 100% H_2SO_4. We determined that 5 minutes in these conditions was sufficient to remove all seed coats. Immediately following chemical treatment, seeds were removed and rinsed with diH_2O to remove all traces of acid.

Another complication is that vegetative growth and development is more variable in *G. soja* than in cultivated soybean. Unlike *G. max*, *G. soja* must be germinated at 100% humidity to achieve consistent germination. Another important step was to prune of the first set of trifoliates leaves at three days after germination. This allowed us to standardize growth stage prior to effector infiltration. This is critical when screening diverse accessions of *G. soja* because timing of unifoliate leaf maturity is often variable among accessions. Pruning of trifoliate leaves also improves unifoliate leaf quality. After pruning, unifoliate leaves became thicker and acquired a darker green color. In plants treated in this manner, resistance responses are stronger and the infiltration zone is more clearly visible. When scoring for visual HRs, it is important to judge the spread of cell death outside of the zone where the syringe interfaced with the leaf surface. This is done

to avoid false positives caused from damage induced by the infiltration process. The negative control and the effector treatment must be done on different plants because a strong effector response can trigger cell death in the negative control (a priming effect). Individuals were subjected to only one treatment for this reason. When screening individual effectors, the experimental treatment can be treated as a negative control on non-responding accessions. When recognition does not occur, levels of cell death are often below those of the untransformed control (Figure 3.1).

Screening G. soja *accessions with 3* P. sojae *pathotypes*

As mentioned above, our screen focused on *G. soja* accessions identified in a previous study to be resistant to a combination of *P. sojae* pathotypes which collectively overcome all known *Rps* genes. Table 3.1 lists the relevant accession information including PI numbers and geographic origin. To confirm disease resistance responses we performed disease screening on *G. soja* germplasm using individual isolates rather than the combination of three used in the previous study. We believed it was important to look at infections on an individual level because it is impossible to predict the complexity of interactions when using isolates in combination. For example, a strong immune response to one pathotype may override the virulence mechanisms of another pathotype. We optimized infectivity of our cultures at monthly intervals by passaging isolates through infected soybean (cv. Williams). The process of refreshing *P. sojae* cultures is essential to maintain a healthy strength for infection. A loss in virulence over time was observed in all isolates during stem and leaf inoculation assays.

We developed a disease assay in which trifoliate leaves were inoculated with agar plugs from each of three virulent accessions: Pt1005, Pt2004, and Race 7 (Figure 3.2).

This assay is reproducible, convenient, and allows the same plants to be scored simultaneously for pathogen resistance and effector response. Wounding of the leaf prior to inoculation was essential for reproducibility. After 5-7 days, resistance versus susceptibility was very clear, and we scored the plants as resistant or susceptible without taking measurements of disease lesions. The *G. max* accession Williams was used as a control for susceptibility. We confirmed that many of the *G. soja* accessions were resistant to all three isolates; however, several accessions were partially susceptible to one or two pathotypes. No accessions were susceptible to all three isolates. Hypocotyl inoculations were done to confirm instances of susceptibility in foliar assays. In all cases, we observed perfect agreement between hypocotyl and leaf inoculation. The major technical challenge of these assays is providing a consistently moist infection site for pathogen infection following inoculation. This refined pathotype information will be important for prioritizing accessions for further characterization and for accurately mapping pathogen resistance loci in segregating populations. Hypocotyl inoculations were laborious and had limited throughput when compared to the trifoliate assay.

Three G. soja *Accessions Contain Novel Resistance to* Psg *race 4, a Bacterial Pathogen of Soybean*

As mentioned above, our effector delivery system is based on co-inoculation with EtHAN (containing the effector transgene) and *Psg* race 4 (to boost growth of the EtHAN strain). Thus, it was necessary to pre-screen *G. soja* accessions for immune responses to *Psg* race 4 that would interfere with effector delivery during the co-inoculation with EtHAN + *Psg*. This was important to eliminate false positives from our screens that could occur from ETI triggered by the native effectors secreted by *Psg* race 4. We found three

G. soja accessions (gs2763, gs730, and gs727) that produced a HR when pressure infiltrated with *Psg* race 4 (Figure 3.3 and Table 3.2). Visual HRs were apparent one to two days post inoculation (DPI). Gs2763 and gs730 produced strong HRs. Accession gs727 produced a weak and inconsistent visual HR against *Psg*. Thus, we were unable to screen for effector specific HRs in these accessions due to the immune interaction with the helper strain *Psg*. However, these accessions likely contain potentially useful resistance genes against *Psg*.

Effector-based screening

Our first screen of 20 *P. sojae* resistant *G. soja* accessions was conducted with three of the most promising core *P. sojae* effectors: *Avh*16*, Avh*180, and *Avh*240. These effectors are strongly expressed and make major contributions to virulence in all four of the reference *P. sojae* accessions, so *R* genes against any of these effectors could be very effective. Effectors were delivered individually for each of the 20 accessions via co-infiltration with EtHAN and *Psg* race 4, as described in Figure 3.1. We scored responses to these effectors with a HR as a percentage (%HR). We observed 3 accessions (gs2777, gs2514, and gs2292) that produced HRs against *Avh*16, *Avh*180, and *Avh*240 respectively (Figure 3.4). Effector responses in these accessions were strong and extremely specific (Table 3.3). Gs2292 produced a HR against *Avh*240 in 95% of individuals tested. Accessions gs2777 and gs2514 produced HRs against *Avh*16 and *Avh*180 respectively in 90% of individuals tested. These accessions produce strong and consistent HRs against their corresponding effectors and display no background cell death from our internal control of untransformed EtHAN co-infiltrated with *Psg*.

A second round of screens was performed with core effectors *Avh53*, *Avh*137, *Avh*23, *Avh*110, *Avh*488, SNEL1 and *Avh*241. *Avh*53 and *Avh*137 elicited resistance responses in many accessions from multiple subgroups (Figure 3.5). Additionally, we observed paired responses to effectors *Avh*53 and *Avh*137 only in accessions collected in Primorsky, Russia. This observed geographic distribution suggests that these accessions share common resistance loci. *Avh*23 and *Avh*110 did not produce a visual HR on any tested *P. sojae* resistant *G. soja* accession (Table 3.4 A-G). We compiled effector screening data and accession information in Table 3.5

Determining the Inheritance of Potential R *Genes in a Segregating $F_{2:3}$ Population*

P. sojae-resistant *G. soja* accessions which responded with a HR to a core *P. sojae* effector were crossed with the susceptible *G. max* cultivar Williams as part of our breeding program. In order to test the viability of using effector-based screening to map a resistance gene, we looked at the inheritance of a potential *R* gene in a segregating $F_{2:3}$ population (Figure 3.6). We analyzed the cross, gs2292 (R) x Williams (S), for the phenotype of response to core *P. sojae* effector *Avh*240 with a HR. Figure 3.7 B describes the screening and scoring individuals within segregating $F_{2:3}$ populations. We collected data from 89 $F_{2:3}$ families with a sample size n>10. Because most *R* genes segregate as major-effect, genetically dominant loci, we expected a 1:2:1 phenotypic ratio. HR strength was variable throughout biological replicates, however the presence or absence of visual HRs was extremely consistent. As expected, we observed a 1:2:1 phenotypic ratio in the $F_{2:3}$ progeny of a cross between gs2292 (R) x Williams (S) (Figure 3.7 and Table 3.6). By analyzing this population, we determined that effector based screening is both sensitive and accurate enough to identify resistance responses on an individual plant

basis. Additionally, the response to *Avh*240 in this accession appears to segregate as a single locus, easily amenable to introgression into *G. max* cultivars.

Discussion

Soybean is harvested for oil and protein, and ranked second for area planted in the US in 2014 [201]. In 2015, 55% of vegetable oil consumed in the United States of America (USA) came from soybean, underscoring the importance of this crop. *Phytophthora* root and stem rot substantially affects the profitability of soybean by reducing soybean harvests in pathogen suited environments across the world [182]. In the USA, *Phytophthora* root and stem rot is most severe in the Midwest, and resistance to *P. sojae* can be broken by many factors [192]. In the past, soybean *Rps* genes were reported to confer full resistance against most *P. sojae* pathotypes. Unfortunately, *P. sojae* is evolving to evade recognition by soybean *Rps* genes, and the average effectiveness of an *Rps* gene in the field is 8-15 years [158]. Thus, it is important to identify novel *R* genes against *P. sojae* which are durable in the field and which recognize the potentially broad diversity of isolates encountered in field situations.

We began by assessing *G. soja* germplasm for resistance to *P. sojae* to confirm and extend the results from Matthiesen et al. [101]. We used the same three pathotypes, which collectively overcome all known *Rps* resistance genes except Rps8, that were used by [101]. We developed a trifoliate infection assay to complement our new effector based screening system. Agar plugs were used to infect a wounded trifoliate leaf. Our trifoliate agar plug infection assay can be used to measure quantitative resistance by measuring lesion diameter, or it can be used to measure qualitative resistance by assaying for the presence or absence of disease lesions in pronounced examples. We used specialized tools to quickly generate uniform agar chunks, greatly increasing throughput and consistency. Most importantly, this assay can be used in combination with effector-based

screening since trifoliates are removed during the pruning process. Using this disease screening technique, we evaluated *G. soja* germplasm for novel resistance to *P. sojae*. Screening with pathotypes individually versus in combination generated uniquely complimentary data. For example, the individual isolate approach generates a more accurate pathotype profile; while screening in combination allows us to infer the strength of the resistance response. Resistance observed in screening with a combination of pathotypes indicate that resistance responses will not be easily overcome through diversification of effector profiles.

We chose to screen *P. sojae* resistant *G. soja* germplasm with core *P. sojae* effectors to identify new sources of *R* genes against *Phytophthora* root and stem rot disease. A core *P. sojae* effector is highly expressed in early *P. sojae* infection, conserved among the 4 most common *P. sojae* pathotypes, and essential for *P. sojae* virulence. To assess effector impact on virulence, Tyler et al. used transient gene silencing to suppress expression of candidate core *P. sojae* effectors. The silencing of core *P. sojae* effectors *Avh*16, *Avh*180, and *Avh*240 greatly reduced pathogen growth on the susceptible *G. max* cultivar Williams. We selected these 3 effectors as our most promising *P. sojae* core effectors. The silencing of core *P. sojae* effectors *Avh*53, *Avh*137, *Avh*23, *Avh*110, *Avh*488, and SNEL1 moderately decreased pathogen growth. We used these 9 core *P. sojae* effectors to probe for potential *R* genes in *P. sojae* resistant *G. soja* germplasm against core *P. sojae* effectors. Because these *P. sojae* effectors are monomorphic and important for virulence, we hypothesize that *P. sojae* will be unable to easily discard core effectors to evade *R* gene detection. For this reason, we predict that a *R* gene against a core *P. sojae* effector will provide durable resistance against *P. sojae* in the field. We

hypothesize that *R* genes against core effector targets will provide broad spectrum resistance because these core effectors contain conserved sequences present among the most common field isolates of *P. sojae*.

Prior to effector-based screening of *G. soja* germplasm with our library of *P. sojae* core effector candidates we established the background of *Psg* in all accessions. This screen identified three *G. soja* accessions: gs2763, gs730, and gs727 which contain resistance to *Psg* race 4. *Psg*, a bacterial pathogen, is the causal agent of bacterial blight on soybean [202]. *Psg* is most virulent in the northern most states and into parts of Canada [202]. By testing *Psg* resistant accessions with other pathotypes (races) of *Psg* whether this resistance is durable against all isolates of this pathogen. We can use segregating populations available in our germplasm collection to track resistance against both *Psg* and *P. sojae* using the assays developed in this study. It will be interesting to determine if there is a linkage between bacterial and oomycete resistance mechanisms.

Resistance responses against *Psg* in these three accessions excluded these accessions from effector-based screening. We compiled effector based screening data from the remaining 20 *G. soja* accessions with 10 effectors. We found putative *R* gene candidates against 8/10 core effector candidates. Effector responses for 6 core effectors were very specific (only one or two responding accessions per effector). We observed strong and clear resistance responses against our top candidate effectors: *Avh*16, *Avh*180, and *Avh*240. In some cases, we found a correlation between the effector response of a *P. sojae* resistant *G. soja* accession and geographic collection location. *Avh*53 provoked responses in 6 accessions, while *Avh*137 provoked resistance responses in 9 accessions. We confirmed that these HRs were effector specific by extensively screening *Avh*53 and

*Avh*137 on non-responding *G. soja* accessions. Four *P. sojae* resistant Russian *G. soja* accessions responded with a HR to both core *P. sojae* effectors *Avh*53 and *Avh*137. From the 2 remaining *P. sojae* resistant Russian *G. soja* accessions, one accession responded to *Avh*53 and one accessions responded to *Avh*137. No other effector responses were observed in *P. sojae* resistant Russian *G. soja* accessions. This could indicate that *R* genes against *Avh*53 and *Avh*137 are under positive selective pressure in Russia. We found multiple *P. sojae* resistant South Korean *G. soja* accessions which responded to *Avh*137 and 1 South Korean accession which responded to *Avh*53. One Japanese *G. soja* accession responded to *Avh*137 as well. Since, subgroups are genetically isolated from each other we hypothesize that evolution has either converged on these effector targets or that these accessions share *R* genes from ancestral origins.

We used the *Pseudomonas* effector screening system to analyze a segregating $F_{2:3}$ population, gs2292 (R) x Williams (S). Similar to the *G. soja* parental accession, $F_{2:3}$ seed required scarification prior to planting. We found that scarifying *G. soja* seeds and the $F_{2:3}$ seeds in 100% H_2SO_4 optimized seed germination and experimental throughput. Individuals in $F_{2:3}$ families shared morphological characteristics more similar to *G. max* cultivar Williams. We observed a 1:2:1 phenotypic ratio for the phenotype of response to the core *P. sojae* effector *Avh*240 with a HR. This suggests a classical dominance inheritance. All *P. sojae* resistant *G. soja* accessions which produced a HR against a core *P. sojae* effector were crossed with *G. max* cultivar Williams as part of our breeding program. We will use recombinant inbred lines to further evaluate and fine-map potential *R* genes. A limitation of our approach is applying this system in the *G. max* Williams background. Effectors which produce a strong HR will often produce background cell

death in Williams. This makes scoring individuals within segregating populations more challenging. We hope to mitigate these limitations by further optimization of our system or by exploring other elite accessions. Since effector-based screening was not paired with pathogen infection assays in $F_{2:3}$ populations it is impossible to say for certain whether our identified effector response genes are truly responsible for pathogen resistance. The parental accession gs2292 is resistant to three accessions which overcome all known *Rps* resistance. Additionally, this accession responds to *Avh*240, one of our most promising core effectors, in effector-based screens. These data taken together suggest that these techniques will complement each other in more advanced populations. We will analyze individuals within recombinant inbred populations using our trifoliate infection assay in combination with our effector-based screening assay to understand this interplay in future studies.

Effector-based screening is a new but powerful technique. We believe that this technique is an excellent complement to classic pathogen infection assays. These techniques can be used together to map resistance loci more accurately. Additionally populations containing multiple resistance loci can be more clearly separated. Some of the most successful examples of breeding for resistance have come from wild species. For example, the *R* gene *Pto* which confers resistance to *Pseudomonas syringae pathovar tomato* has been used to control speck disease for over 70 years [117]. Until recently, the only resistance against root-knot nematode in tomato was the *R* gene Mi-1[116]. Introgressed from the wild tomato species, this gene from *Lycopersicon peruvianum* is an excellent example of an effective resistance gene derived from a wild species. In potato, the *R* gene *Pi-ber* was derived from *Solanum berthaultii* and confers complete resistance

against the *P. infestans* isolate US8 [106]. We built on work in *G. max* to develop an effector-based screening system for *G. sojae* (Chapter 2). We were able to utilize bacterial type III secretion to deliver individual effectors to the interior of *G. soja* unifoliates. After growth conditions were optimized, this system was seamlessly transitioned into *G. soja*, underscoring the potential of this system to be deployed across broad spectrum of crop plants. We used resources (germplasm and cultivars with specific *Rps* genes) available in *G. max* to develop a disease system that would be effective in both *G. max* and *G. soja* for any pathotype.

By targeting our effector screens to identify *R* genes against core *P. sojae* effectors, we can look for durable resistance in a unique way. Additionally, we designed a disease assay which can be performed concurrently with effector response screens on the same individual plants, to facilitate linking effector responses with pathogen resistance. The introduction of these genes into cultivated soybean by traditional breeding allows for product introgression into all soy-based markets. We hope to map these genes to specific loci by examining phenotypes in segregating progeny. Mapping resistance genes using genetic markers is an effective and proven method [93]. We hope to use the unique information acquired from an intelligently designed effector-based screens and disease assays to predict *R* gene specificity and durability in a field setting [203].

Materials and Methods

Culture Preparation and Maintenance

Phytophthora sojae cultures of pathotypes pt1005, pt2004, and Race 7 were kindly provided by the Robertson Lab at Iowa State University [101]. Agar plugs were

used to inoculate trifoliate leaves of susceptible Williams. Seven days post inoculation, visibly infected areas were cut and washed in 70% EtOH. Sterilized plant material was washed with diH_2O, dried, and plated on to PARP-V8 plates. Cultures were incubated in the dark at room temperature. Agar plugs containing *P. sojae* mycelium were taken from the edges of the growing colonies and used to start new V8 plates. *P. sojae* grows quickly on V8 media and fully colonized plates within two weeks. Pathogen material prepared in this fashion was used for infection assays. On average, *P. sojae* will colonize a PARP-V8 plate in 6-8 weeks. After 8 weeks, *P. sojae* pathotypes were re-isolated using the process outlined above, to maintain virulence.

Trifoliate Leaf Agar Plug Infection Assay

We used agar plugs from pathotypes pt1005, pt2004, and Race 7 to infect mature trifoliate leaves from our *G. soja* germplasm collection. Trifoliate leaves were detached and placed into Petri dishes containing filter paper saturated with water. A scalpel was used to wound trifoliate leaves. Wound size, shape, and depth are not important within reason; however the position and angle of the wound will sometime effect the direction of the disease lesion. Agar plugs were then placed over the wound and a drop of H_2O was added to the agar plug to hydrate the site of infection. We used a specialized tool, specifically a Large/Jumbo Double End Steel Amalgam Carrier ®Miltex Instrument Co. to rapidly generate consistent sized agar plugs. Plates were sealed with parafilm to prevent moisture loss and maintain humidity during infection. Sealed plates were stored under normal growing conditions until imaging. After 7-10 days, disease lesions on susceptible plants covered 80-100% of the leaf surface. While disease lesions are often present on resistant accessions, these lesions do not expand past the wound site.

Hypocotyl Inoculation Assay

We confirmed our results using standard hypocotyl inoculation assays [183]. We applied agar plugs to wounded hypocotyls of soil grown plants. After inoculation, lesions developed in resistant and susceptible accessions. Lesions did not expand outside of the site of inoculation in resistant germplasm. Lesions expanded rapidly in susceptible individuals making differences in susceptibility clear. We confirmed resistance and susceptibility phenotypes using this technique to validate our trifoliate screen.

Scarification

Seeds were scarified in 100% H_2SO_4, for 5 minutes. Seeds were then rinsed vigorously in H_2SO_4 and directly planted or dried for storage. As a high-throughput modification, accessions were scarified using a 98 well 1.5 mL centrifuge tube holder with holes bored through the wells (Figure 3.8). This allowed seed to be scarified in batch, while keeping accessions separate. Manual scarification using coarse sandpaper is effective; however germination rates and emergence times were improved with acid scarification.

Planting and Maintenance

For all screens, *G. max* was planted in Fafner P2 potting media. Fafner P2 potting media was hydrated with 2 L of water and Cal-Mag Fertilizer®. 2.5 inch pots were packed firmly with media. 5 seeds were sown per pot and covered with about 2-3 cm of loosely packed potting media. The top soil was treated with Marathon insecticide, put into flats and covered with clear plastic lids to maintain humidity during germination. After 24-36 h, lids were removed from the flats. Germination rates were between 80-95%

for all *G. soja* accessions tested. *G. soja* were germinated and maintained under diurnal light (16 h light at 22 ^{0}C; 6 h dark at 20 ^{0}C). Light intensity was maintained between 250-300 µM for all experiments. Light intensity is critical for the production of consistent macroscopic HRs during screening. Flats were watered with 1.0 L of H_2O per flat every 36-48 hours or as needed. Before the end of the V2 stage of development, trifoliate leaves were pruned as described in the Results section.

Culture Preparation

Gateway clones of 10 RXLR core effectors (*Avh*16, *Avh*180, *Avh*240, *Avh*53, *Avh*137, *Avh*241, *Avh*488, SNEL1, *Avh*23, and *Avh*110) were generously provided by the Tyler Lab at Oregon State University. The unstructured N terminus and secretion peptides of core RXLR effectors were removed during entry vector cloning in order to complement the design of the pEDV6 system. Bacterial cultures were prepared from single colonies and inoculated in 5 mL aliquots of KB media containing the appropriate antibiotic selection. The bacteria was pelleted as mentioned above and re-suspended in 10 mM $MgSO_4$. Single inoculums were prepared at ODs = 0.3-0.6 and pressure infiltrated into soybean unifoliates. Co-infiltration of EtHAN and *Psg* were done by mixing the two at ODs 0.6 and 0.3 respectively for a final OD = 0.9.

Infiltrations

Plants were infiltrated after 15-17 days (3 days after pruning the trifoliate leaves). Bacterial solutions were prepared 1 hour prior to infiltration. Needless syringes were used to pressure infiltrate attached unifoliate leaves. Each unifoliate received 4 infiltrations of 50-100 µL of bacterial solution. The identical process is performed on the adjacent

unifoliate. Negative controls were performed on separate plants to avoid priming induced by effector recognition.

Scoring

Seven days post infiltration (DPI), unifoliate leaves were detached, imaged and, scored. A plant was scored positive if ≥6/8 infiltration spots produced a visual HR on both unifoliate leaves. Plants were scored negative if ≤2/8 infiltration spots produced a visible HR. Plants that fell between 3-5/8 were scored as ambiguous. Any data, scored as ambiguous, was discarded, and the experiment was repeated until the appropriate sample size was reached (Chapter 2). To calculate % HR, we used the equation, $\%\ HR = (\#\ of\ responding\ plants) \div (\#\ of\ infiltrated\ plants)$.

Delivery of P. sojae *Core Effectors to Resistant* G. soja *Germplasm*

In experiments with *Avh16*, *Avh180*, and *Avh240*, cultivars were screened with effectors individually in three biological replicates (n = 20 plants per replicate). Standard deviation was calculated between the three replicates. Statistical significance ($p<0.05$) was determined by a pairwise t-test against Harosoy.

Determination of R *Gene Inheritance in* $F_{2:3}$ *Populations*

Individuals from segregating $F_{2:3}$ families were screened with Avh240 as described above. Before effector screening, we determined the background that *Avh*240 produces in Williams. $F_{2:3}$ families were tested with an average sample size of n≥10 individuals per family. We determined our range for homozygous dominant as ≥85 % HR. We determined our range for homozygous recessive as the background response of

Avh240 on Williams ≤30 % HR. Families that fell within the two ranges are considered to be segregating. We use the Chi squared (χ^2) value to determine the probability at which the deviation is expected.

Figures and Tables

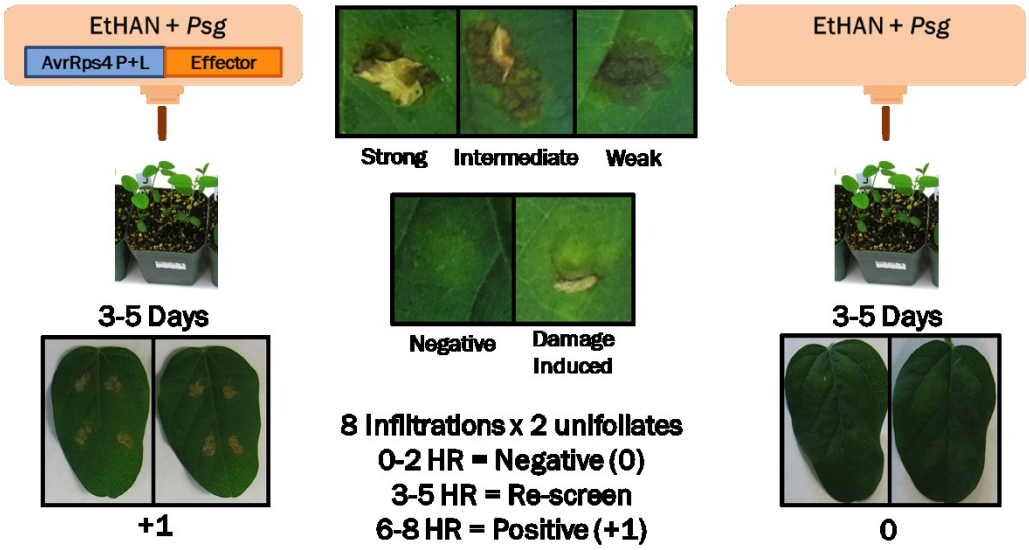

Figure 3.1 Overview of the procedure for screening *G. soja* accessions with core *P. sojae* RXLR effectors. *P. fluorescens* EtHAN was used to deliver *P. sojae* effector proteins to the interior of plant cell. The pEDV6 vector contains the Avr*Rps*4 promoter and leader sequence which were used to guide effector proteins through the Type III secretion system. We co-infiltrated EtHAN with *Psg* race 4 in a 2:1 ration to suppress PTI. This suppression creates a favorable environment for *P. fluorescens* to achieve high levels of growth, such that it delivers higher quantities of our desired effector and produces a more robust phenotype. Pressure infiltration was used to inoculate four areas per unifoliate leaf and we infiltrated both unifoliate leaves per plant. Plants were scored 3-5 DPI. If at least 6 out of 8 spots produced a visual HR, the plant was scored +1. If 0 out of 2 spots produced a visual HR, the plant was scored as 0, and if 3-5 spots produced a visual HR, the data point was discarded and the population was retested. An internal control of untransformed *P. fluor*escens co-infiltrated with *Psg* was used during screening.

Table 3.1 Accession Information for all tested *G. soja* germplasm is listed.

PI #	Accession ID	Origin
PI464935	gs2410	China: Jiangsu
PI549046	gs2761	China: Ningxia
PI549048	gs2763	China: Changping
PI464889A	gs727	China: Jilin
PI479746B	gs56	China Jilin
PI562538	gs2771	South Korea: Gyeon Gi Do
PI562539	gs2772	South Korea: Gyeon Gi Do
PI562542	gs725	South Korea: Gyeon Gi Do
PI407162	gs730	South Korea: Gyeon Gi Do
PI562544	gs2777	South Korea: Chung Ch'ong Nam Do
PI562547	gs2780	South Korea: Chung Ch'ong Nam Do
PI562549	gs2782	South Korea: Chung Ch'ong Nam Do
PI507582	gs2474	Japan: Aomori
PI507587	gs2479	Japan: Aomori
PI507597	gs2490	Japan: Nakajo
PI507621	gs2514	Japan: Nakajo
PI407076	gs2292	Japan: Unknown
PI507767	gs2601	Russia: Primorskaya Province
PI507788	gs2618	Russia: Primorskaya Province
PI507794	gs2623	Russia: Primorskaya Province
PI507814	gs714	Russia: Primorskaya Province
PI522215	gs719	Russia: Primorskaya Province
PI522216	gs720	Russia: Primorskaya Province

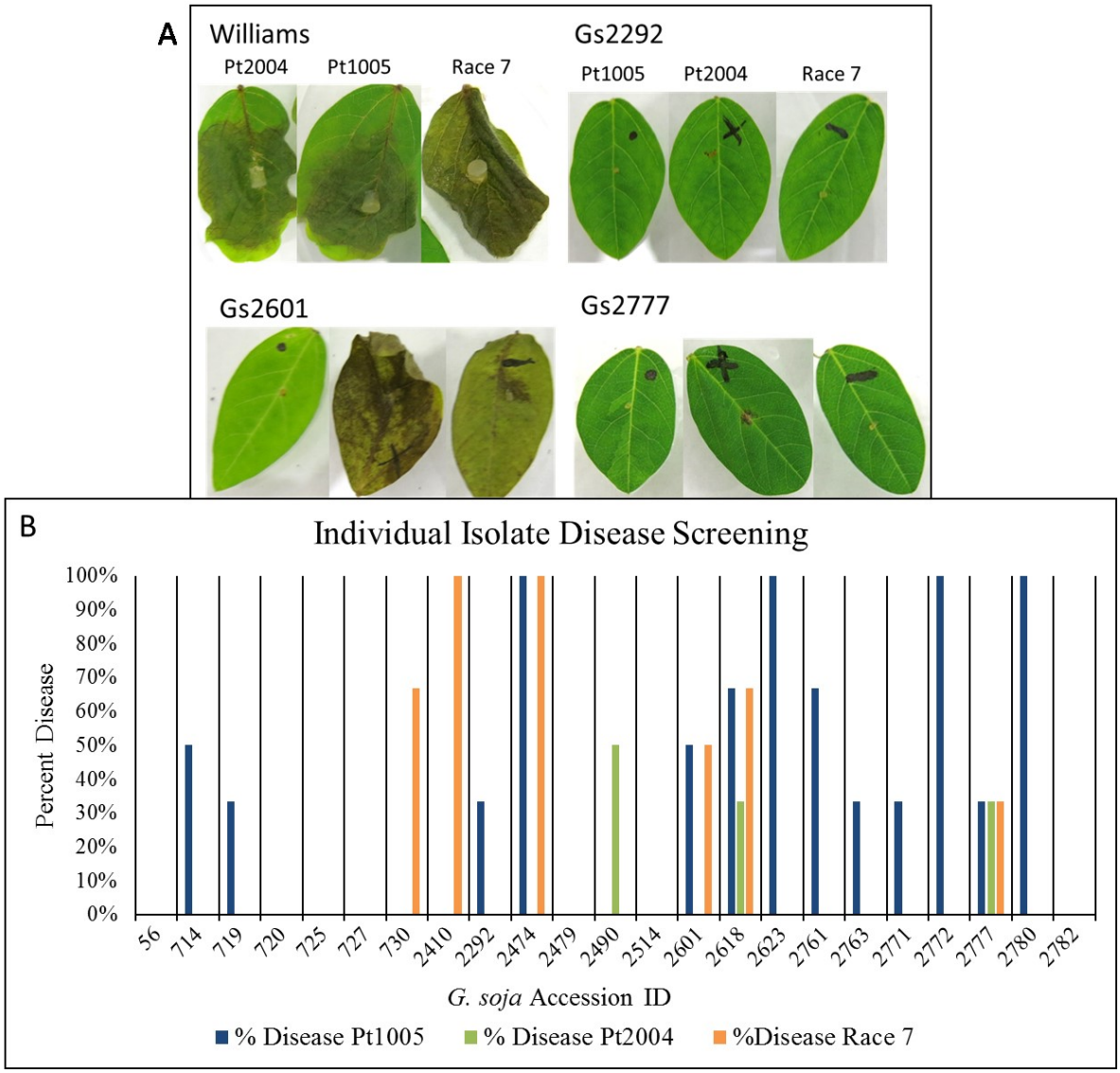

Figure 3.2 Disease screening of *G. soja* accessions with three common field isolates of *P. sojae*. Agar plugs containing mycelia were used to infect trifoliate leaves with isolates Pt2004 (X) Pt1005 (.) and PtRace7 (-). We found susceptible interactions among *G. soja* accessions; however most cultivars showed nearly complete resistance. On the X axis, *G. soja* germplasm is listed and on the Y axis, resistance or susceptibility was measured for n = 5 individuals in two biological replicates.

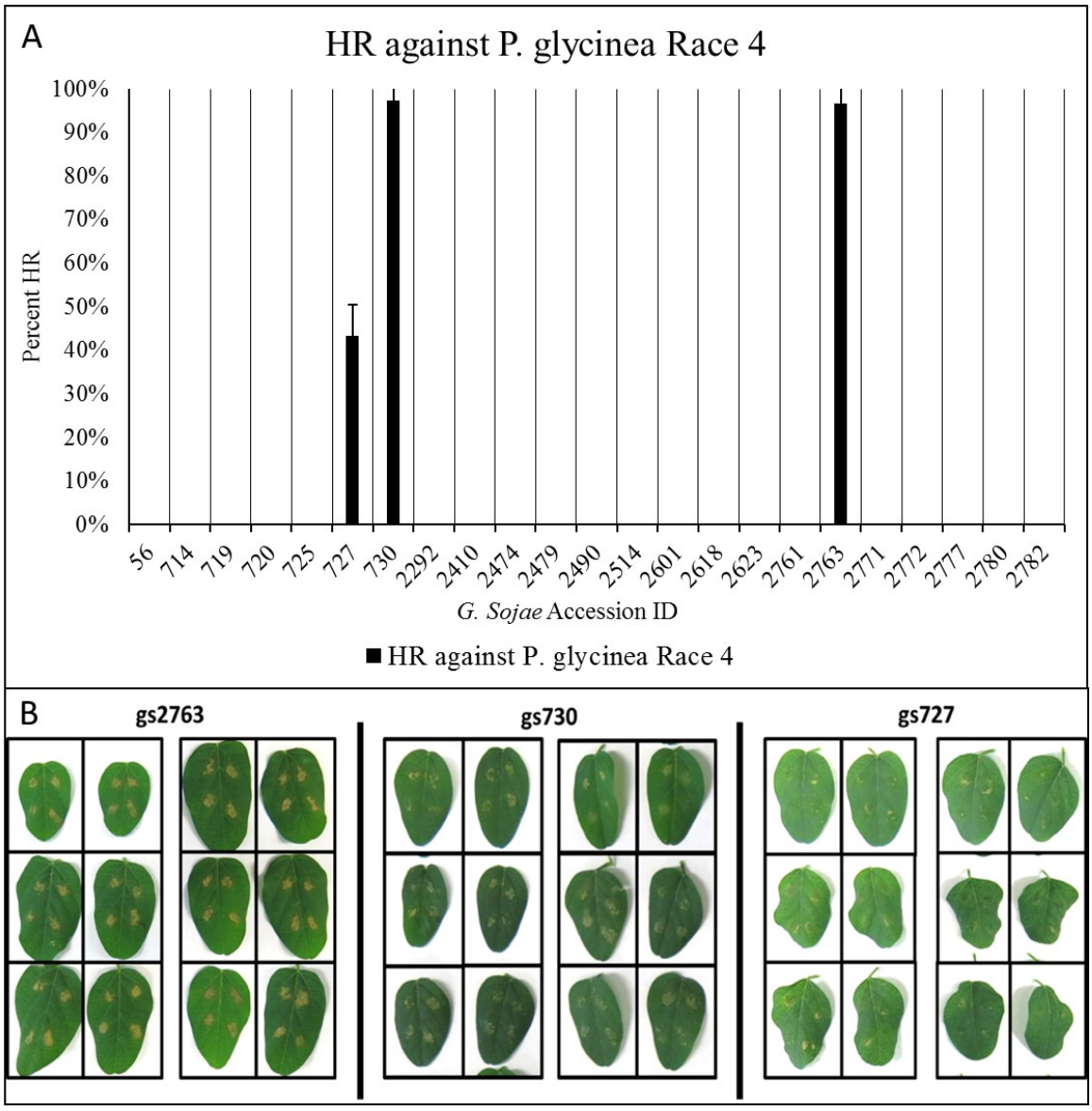

Figure 3.3 Gs2763, gs730, and gs727 are incompatible with *Psg* race 4. These *P. sojae* resistant *G. soja* accessions produced a faster, stronger HR when inoculated with *Psg* as opposed to when inoculated with a core *P. sojae* effector. A) The %HR in response to *Psg* was recorded over 3 biological replicates. B) The images were taken from 1 biological replicate where these accessions were infiltrated with *Psg*.

Table 3.2 Screening *G. soja* germplasm for a HR in response to *Psg* race 4. HRs against *Psg* race 4 were observed in *G. soja* accessions gs730, gs2763, and gs727.

G. soja **Line ID**	**% HR against *Psg* race 4**	**% Standard Deviation**
gs56	0%	0%
gs714	0%	0%
gs719	0%	0%
gs720	0%	0%
gs725	0%	0%
gs727	43%	7%
gs730	97%	4%
gs2292	0%	0%
gs2410	0%	0%
gs2474	0%	0%
gs2479	0%	0%
gs2490	0%	0%
gs2514	0%	0%
gs2601	0%	0%
gs2618	0%	0%
gs2623	0%	0%
gs2761	0%	0%
gs2763	97%	4%
gs2771	0%	0%
gs2772	0%	0%
gs2777	0%	0%
gs2780	0%	0%
gs2782	0%	0%

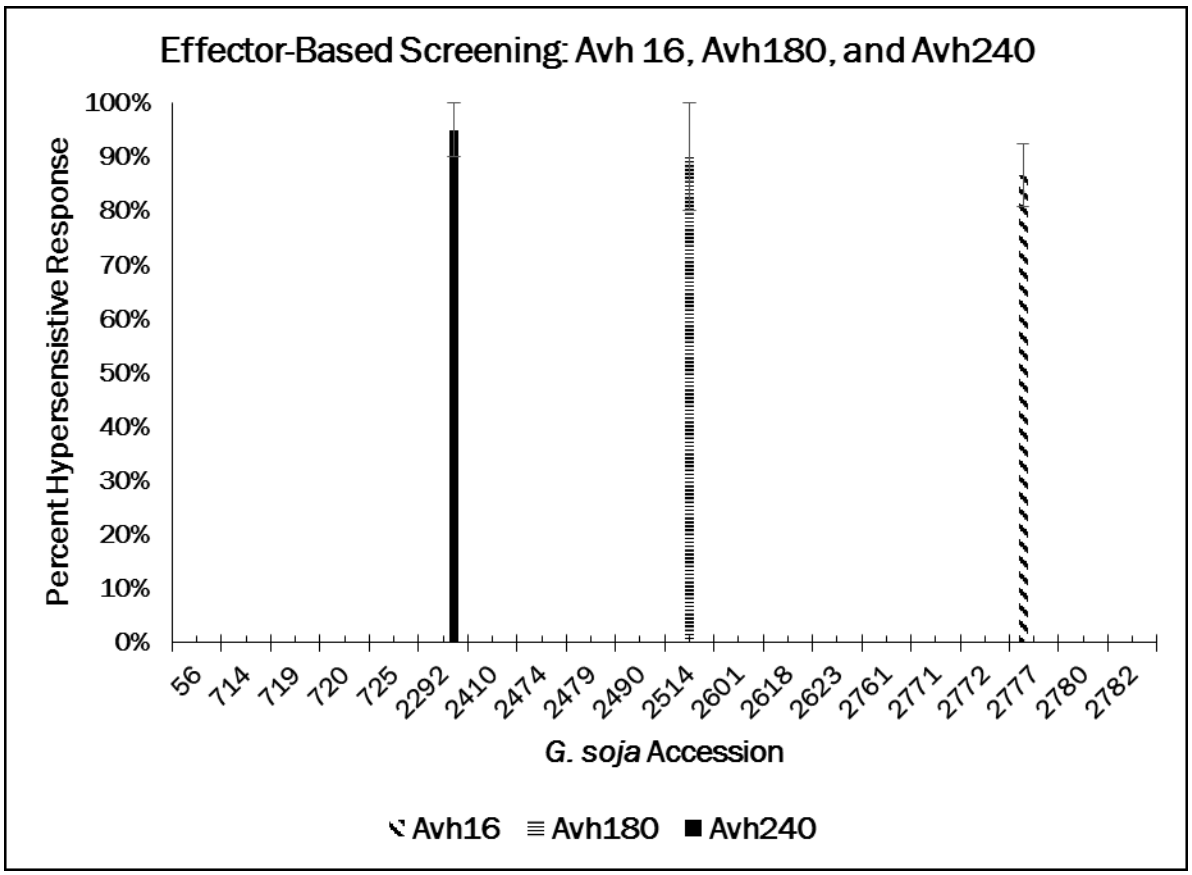

Figure 3.4 Gs2777, gs2514, and gs2292 produced HRs in response to *Avh*16, *Avh*180, and *Avh*240 respectively. On the X axis, *G. soja* accessions subjected to effector-based screening are listed. On the Y axis, percent hypersensitive response among individuals within accessions recorded. We confirmed the observed effector responses over 3 biological replicates to determine standard deviation (n≥60). The error bars represents the percent standard deviation between 3 biological replicates. Neither *P. fluorescens* nor *Psg* produced HRs individually or in combination in these accessions.

Table 3.3 Total *Avh*16, *Avh*180, and *Avh*240 effector-response in *G. soja* germplasm. Percent hypersensitive response and percent standard deviation is reported for the effectors *Avh*16, *Avh*180, and *Avh*240.

***G. soja* Line ID**	***Avh*16 % HR**	***Avh*16 % StDev**	***Avh*180 % HR**	***Avh*180 % StDev**	***Avh*240 % HR**	***Avh*240 % StDev**
gs56	0%	0%	0%	0%	0%	0%
gs714	0%	0%	0%	0%	0%	0%
gs719	0%	0%	0%	0%	0%	0%
gs720	0%	0%	0%	0%	0%	0%
gs725	0%	0%	0%	0%	0%	0%
gs2292	0%	0%	0%	0%	95%	5%
gs2410	0%	0%	0%	0%	0%	0%
gs2474	0%	0%	0%	0%	0%	0%
gs2479	0%	0%	0%	0%	0%	0%
gs2490	0%	0%	0%	0%	0%	0%
gs2514	0%	0%	90%	10%	0%	0%
gs2601	0%	0%	0%	0%	0%	0%
gs2618	0%	0%	0%	0%	0%	0%
gs2623	0%	0%	0%	0%	0%	0%
gs2761	0%	0%	0%	0%	0%	0%
gs2771	0%	0%	0%	0%	0%	0%
gs2772	0%	0%	0%	0%	0%	0%
gs2777	87%	6%	0%	0%	0%	0%
gs2780	0%	0%	0%	0%	0%	0%
gs2782	0%	0%	0%	0%	0%	0%

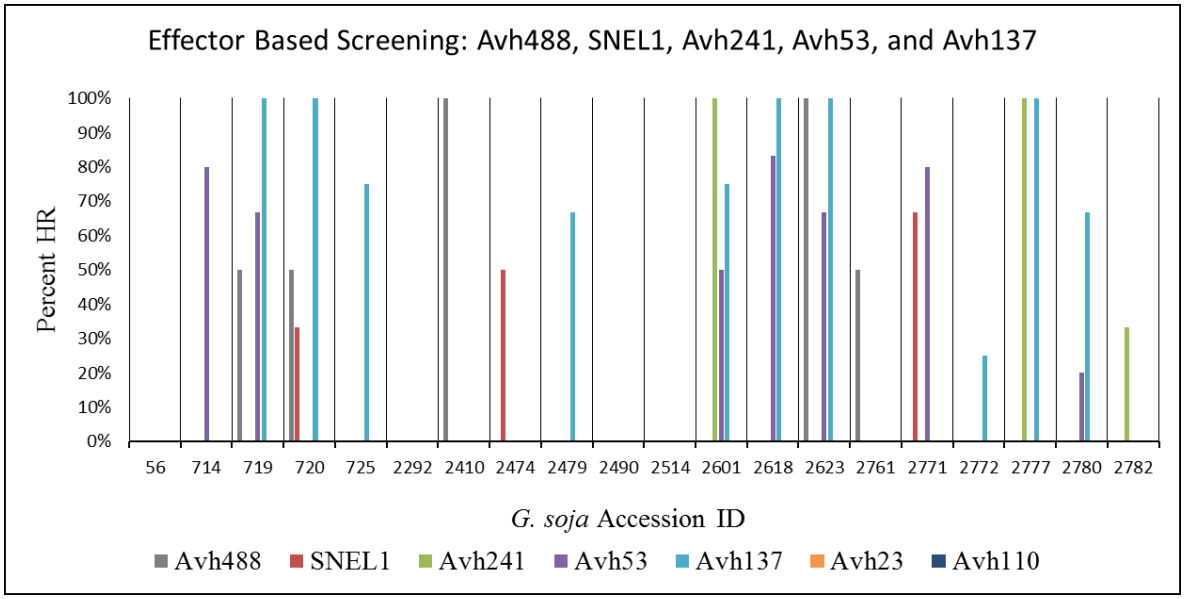

Figure 3.5 Cumulative effector based screening with *Avh*488, SNEL1, *Avh*241, *Avh*53, and *Avh*137. On the Y axis, percent hypersensitive response is reported for effectors screened individually on *G. soja* germplasm.

Table 3.4 A-G Effector-based screening data for *Avh*53*,* *Avh*137, *Avh*488, SNEL1, *Avh*241, *Avh*23, and *Avh*110. Response to a given effector (*Avh* HR) divided by total plants screened (Total *Avh*) gives percent hypersensitive response (*Avh* % HR).

A

G. soja **Line ID**	***Avh*53 HR**	**Total *Avh*53**	***Avh*53 % HR**
gs56	-	-	NT
gs714	4	5	80%
gs719	4	6	67%
gs720	0	3	0%
gs725	0	5	0%
gs2292	0	4	0%
gs2410	0	4	0%
gs2474	0	6	0%
gs2479	0	7	0%
gs2490	0	3	0%
gs2514	0	2	0%
gs2601	3	6	50%
gs2618	5	6	83%
gs2623	2	3	67%
gs2761	0	2	0%
gs2771	4	5	80%
gs2772	0	5	0%
gs2777	0	6	0%
gs2780	1	5	20%
gs2782	0	9	0%

B

G. soja **Line ID**	***Avh*137 HR**	**Total *Avh*137**	***Avh*137 % HR**
gs56	-	-	NT
gs714	0	2	0%
gs719	4	4	100%
gs720	2	2	100%
gs725	3	4	75%
gs2292	0	2	0%
gs2410	0	3	0%
gs2474	0	2	0%
gs2479	2	3	67%
gs2490	0	2	0%
gs2514	0	1	0%
gs2601	3	4	75%
gs2618	5	5	100%
gs2623	4	4	100%
gs2761	0	5	0%
gs2771	0	3	0%
gs2772	1	4	25%
gs2777	5	5	100%
gs2780	2	3	67%
gs2782	0	4	0%

C

G. soja Line ID	*Avh*488 HR	Total *Avh*488	*Avh*488 % HR
gs56	0	3	0%
gs714	0	2	0%
gs719	2	7	29%
gs720	1	5	20%
gs725	0	0	NT
gs2292	0	2	0%
gs2410	3	3	100%
gs2474	0	3	0%
gs2479	0	0	NT
gs2490	0	3	0%
gs2514	0	2	0%
gs2601	0	3	0%
gs2618	0	5	0%
gs2623	3	3	100%
gs2761	3	4	75%
gs2771	0	4	0%
gs2772	0	3	0%
gs2777	0	5	0%
gs2780	0	1	0%
gs2782	0	2	0%

D

G. soja Line ID	SNEL1 HR	Total SNEL1	SNEL1 % HR
gs56	0	3	0%
gs714	0	3	0%
gs719	0	6	0%
gs720	1	3	33%
gs725	0	0	NT
gs2292	0	1	0%
gs2410	1	5	20%
gs2474	0	3	0%
gs2479	0	0	NT
gs2490	0	1	0%
gs2514	0	1	0%
gs2601	0	0	NT
gs2618	0	5	0%
gs2623	0	0	NT
gs2761	0	5	0%
gs2771	0	2	0%
gs2772	0	2	0%
gs2777	0	4	0%
gs2780	0	1	0%
gs2782	3	3	100%

E

G. soja Line ID	*Avh*241 HR	Total *Avh*241	*Avh*241 % HR
gs56	0	6	0%
gs714	4	4	100%
gs719	1	8	13%
gs720	0	2	0%
gs725	0	3	0%
gs2292	0	6	0%
gs2410	0	1	0%
gs2474	0	11	0%
gs2479	0	3	0%
gs2490	0	4	0%
gs2514	0	2	0%
gs2601	2	7	29%
gs2618	2	8	25%
gs2623	0	6	0%
gs2761	0	7	0%
gs2771	3	4	75%
gs2772	0	11	0%
gs2777	5	7	71%
gs2780	0	2	0%
gs2782	2	6	33%

F

G. soja Line ID	*Avh*23 HR	Total *Avh*23	*Avh*23 % HR
gs56	0	4	0%
gs714	0	5	0%
gs719	0	2	0%
gs720	0	3	0%
gs725	0	3	0%
gs2292	0	2	0%
gs2410	0	3	0%
gs2474	0	2	0%
gs2479	0	4	0%
gs2490	0	5	0%
gs2514	0	3	0%
gs2601	0	3	0%
gs2618	0	3	0%
gs2623	0	2	0%
gs2761	0	5	0%
gs2771	0	2	0%
gs2772	0	2	0%
gs2777	0	3	0%
gs2780	0	2	0%
gs2782	0	3	0%

G

G. soja Line ID	*Avh*110 HR	Total *Avh*110	*Avh*110 % HR
gs56	0	3	0%
gs714	0	3	0%
gs719	0	3	0%
gs720	0	4	0%
gs725	0	3	0%
gs2292	0	2	0%
gs2410	0	2	0%
gs2474	0	5	0%
gs2479	0	2	0%
gs2490	0	3	0%
gs2514	0	3	0%
gs2601	0	2	0%
gs2618	0	2	0%
gs2623	0	4	0%
gs2761	0	4	0%
gs2771	0	2	0%
gs2772	0	2	0%
gs2777	0	2	0%
gs2780	0	2	0%
gs2782	0	3	0%

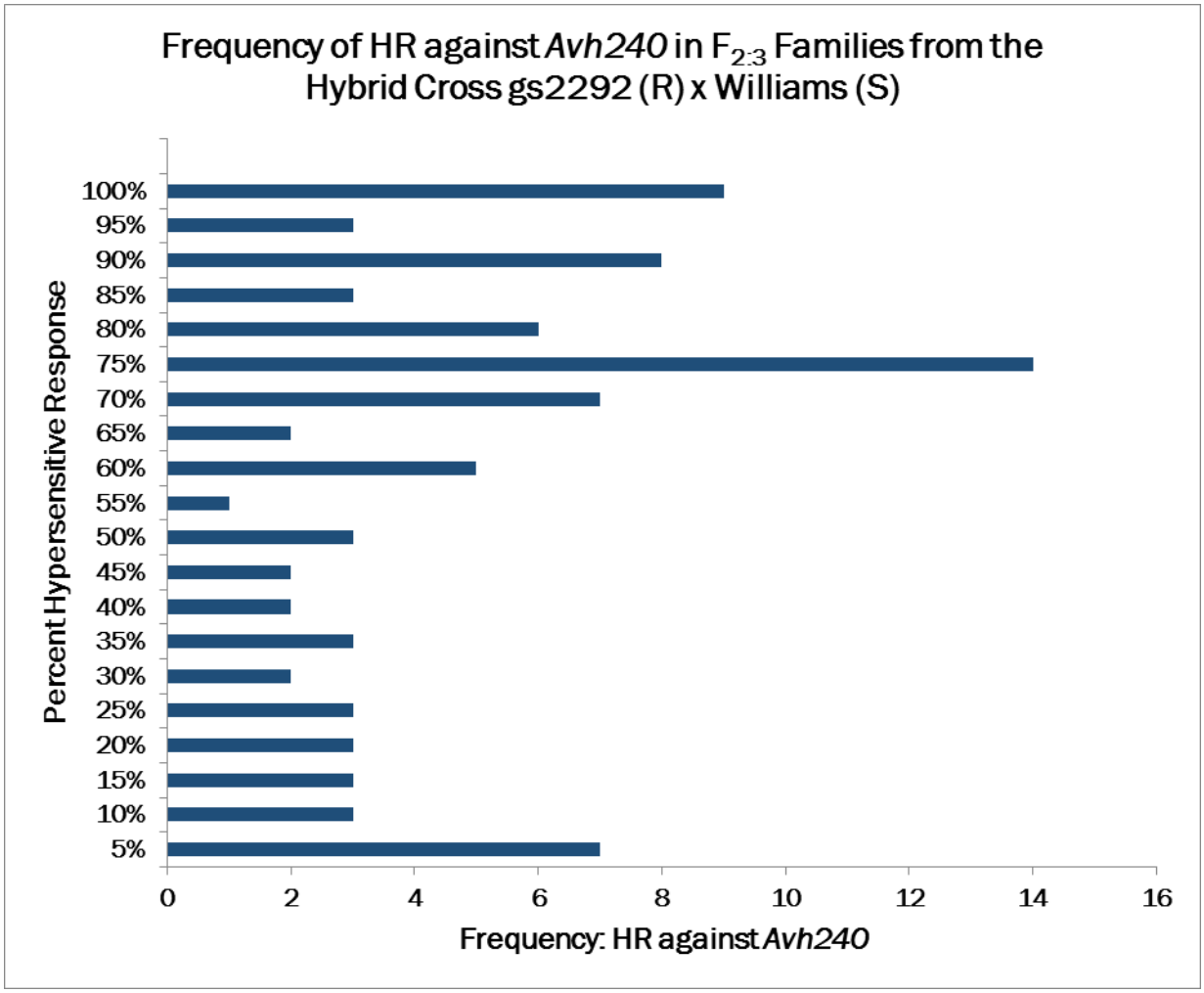

Figure 3.6 Histogram of effector-based screening on 88 $F_{2:3}$ hybrid families. The distribution of effector response in segregating populations suggests the presence of a single dominant *R* gene against *Avh*240.

A

Cross	# of $F_{2:3}$ Families			Expected Ratio	χ^2 value	Probability
	HgD	SEG	HgR			
gs2292 (R) X Williams (S)	21	48	20	1:2:1	0.230	>98%

B

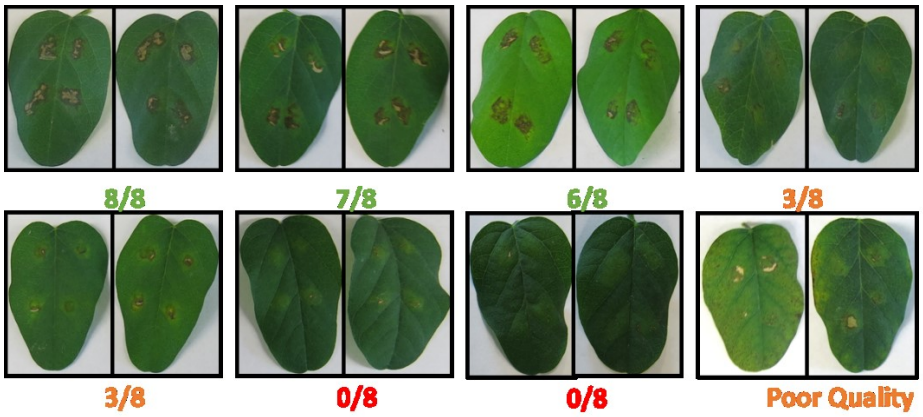

Figure 3.7 Segregating $F_{2:3}$ families were scored for response to *Avh*240. **a)** This table shows the distribution of families in response to the trait HR against *Avh*240. Families were separated into 3 groups homozygous dominant (HgD), homozygous recessive (HgR), and segregating (SEG). **b)** HR strength was variable in $F_{2:3}$ families. Individuals that produced HRs against *Avh*240 in ≥6/8 infiltrations were scored positive, whereas individuals which produced a HR in ≤2 infiltrations were scored negative. Individuals that produced HRs in 3-5/8 infiltrations were not included in scored data.

Table 3.5 Cumulative effector-based screening of wild soybean accessions. Shaded boxes indicate which *G. soja* accession responds to which effector. Percentages are included with standard deviation when possible.

Accession ID	*Avh*488	SNEL1	*Avh*241	*Avh*53	*Avh*137	*Avh*16	*Avh*180	*Avh*240	Subgroup
gs2410	100%								1
gs2761	75%								2
gs727									3
gs728									3
gs2771			75%	80%					4
gs2772									4
gs725					75%				4
gs2777			71%		100%	87% ± 6			4
gs2780					67%				4
gs2782		100%							4
gs2474									5
gs2479					67%				5
gs2490									5
gs2514							90% ± 10		5
gs2292								95% ± 5	5
gs2601				50%	75%				6
gs2618				83%	100%				6
gs2623				67%	100%				6
gs714			100%	80%					6
gs719				67%	100%				6
gs720					100%				6

Table 3.6 *Avh*240 effector-based screening data. - Cross gs2292 (R) x Williams (S). Replicate 1 *Avh*240 responding plants are labeled by the column header (R1P). Replicate 1 total plants are labeled by the column header (R1T). The remaining replicates were labeled by the above scheme. Responders (P) divided by total plants screened (T) gives percent hypersensitive response (%HR) for all biological replicates.

Cross	R1P	R2P	R3P	R4P	R5P	R1T	R2T	R3T	R4T	R5T	P	T	%HR
1	0	3	0	2	0	0	3	3	3	0	5	9	56%
2	0	1	0	0	0	0	3	1	1	2	1	7	14%
3	0	4	2	0	0	0	4	5	0	0	6	9	67%
4	0	0	0	0	0	0	0	5	2	3	0	10	0%
5	0	1	1	1	3	0	1	1	2	4	6	8	75%
6	0	3	2	2	3	0	4	3	4	4	10	15	67%
7	0	1	0	0	0	0	3	1	3	0	1	7	14%
8	0	0	1	1	0	0	0	1	1	2	2	4	50%
9	0	3	0	0	3	0	4	2	0	4	6	10	60%
11	0	1	2	0	0	0	3	5	1	0	3	9	33%
12	0	0	1	1	3	0	4	3	2	5	5	14	36%
13	0	0	0	0	0	0	2	2	0	5	0	9	0%
14	0	0	1	0	1	0	2	3	0	4	2	9	22%
15	0	4	3	3	3	0	4	5	5	4	13	18	72%
17	0	0	0	0	0	0	3	2	0	0	0	5	0%
18	0	2	3	3	0	0	4	3	4	0	8	11	73%
19	0	0	2	1	3	0	2	4	2	3	6	11	55%
20	0	0	1	4	3	0	2	1	4	4	8	11	73%
21	0	0	2	1	0	0	0	2	1	0	3	3	100%
22	0	3	1	4	0	0	3	2	5	0	8	10	80%
23	0	3	2	3	0	0	3	2	3	1	8	9	89%
24	0	0	0	2	3	0	0	1	4	3	5	8	63%
25	0	4	2	2	5	0	4	5	4	5	13	18	72%
26	0	5	3	4	2	0	5	3	4	3	14	15	93%
27	0	3	2	1	0	0	3	2	1	0	6	6	100%
28	0	0	5	4	5	0	0	5	4	5	14	14	100%
29	0	2	5	5	3	0	3	5	5	4	15	17	88%
30	0	3	1	0	2	0	3	2	0	3	6	8	75%
31	0	5	2	0	4	0	5	5	0	4	11	14	79%
32	0	4	2	0	0	0	4	4	0	2	6	10	60%
33	0	3	3	0	2	0	4	4	0	3	8	11	73%

34	0	2	1	0	0	0	2	5	0	0	3	7	43%
35	0	1	1	4	0	0	1	1	5	1	6	8	75%
36	0	4	2	1	3	0	4	4	1	3	10	12	83%
37	0	4	2	1	0	0	5	4	1	0	7	10	70%
39	0	4	1	0	3	0	4	2	2	3	8	11	73%
40	0	0	0	0	1	0	0	3	1	5	1	9	11%
41	0	0	2	1	0	0	4	4	1	2	3	11	27%
42	0	0	0	0	1	0	4	2	0	4	1	10	10%
44	0	0	0	0	0	0	2	3	0	3	0	8	0%
45	0	0	0	1	0	0	2	2	2	0	1	6	17%
46	0	2	3	4	0	0	4	4	4	1	9	13	69%
49	0	0	1	2	0	0	2	1	2	0	3	5	60%
51	2	0	2	0	0	4	2	4	0	3	4	13	31%
52	3	3	0	0	0	4	3	0	0	0	6	7	86%
53	2	3	3	0	3	3	3	3	1	3	11	13	85%
54	4	0	3	0	2	4	0	3	0	2	9	9	100%
55	4	4	5	0	4	4	4	5	0	4	17	17	100%
56	3	0	5	4	5	5	0	5	4	5	17	19	89%
57	0	2	0	0	0	1	2	1	1	0	2	5	40%
59	0	1	1	1	1	1	2	2	0	1	4	6	67%
60	3	1	2	0	2	4	2	3	0	4	8	13	62%
61	0	0	1	0	0	6	2	4	0	2	1	14	7%
63	2	0	0	0	2	4	1	1	0	2	4	8	50%
64	0	0	0	0	0	3	2	0	0	0	0	5	0%
65	0	0	1	0	2	3	4	4	1	5	3	17	18%
66	2	1	0	0	1	5	4	0	0	3	4	12	33%
67	0	1	2	0	1	3	1	2	0	1	4	7	57%
68	3	0	2	1	0	4	0	2	1	0	6	7	86%
70	4	1	1	0	0	5	1	1	0	1	6	8	75%
71	5	2	0	0	0	5	2	0	0	1	7	8	88%
72	2	0	0	0	0	4	2	4	0	3	2	13	15%
73	2	2	2	1	2	2	2	4	1	2	9	11	82%
74	4	5	3	1	2	4	5	5	3	3	15	20	75%
75	3	2	2	1	3	4	2	3	1	4	11	14	79%
76	2	0	1	0	2	3	4	2	0	2	5	11	45%
77	1	0	1	0	0	1	0	1	0	0	2	2	100%
78	1	0	0	0	1	5	6	5	3	5	2	24	8%
79	3	1	3	3	1	4	2	4	4	2	11	16	69%
80	1	0	0	0	0	2	1	0	0	1	1	4	25%
81	3	0	0	2	0	3	0	0	2	0	5	5	100%
82	4	3	4	1	1	5	4	4	3	3	13	19	68%

83	3	3	2	2	1	5	3	2	3	1	11	14	79%
85	3	2	3	1	1	3	2	3	1	1	10	10	100%
86	5	0	3	4	0	5	0	5	4	0	12	14	86%
87	5	4	4	4	2	5	4	4	5	2	19	20	95%
88	5	0	3	1	0	5	0	3	1	0	9	9	100%
89	2	1	0	0	1	4	2	0	0	3	4	9	44%
90	3	4	2	0	3	4	4	2	0	3	12	13	92%
91	5	1	0	1	0	5	2	0	1	0	7	8	88%
92	2	0	0	0	0	4	2	1	0	0	2	7	29%
93	4	1	2	0	3	5	3	3	0	3	10	14	71%
94	2	1	1	0	0	2	2	1	0	0	4	5	80%
95	2	2	1	0	0	4	2	1	0	0	5	7	71%
96	0	0	0	0	0	1	0	1	0	0	0	2	0%
97	0	1	0	1	0	2	2	1	3	0	2	8	25%
98	0	0	0	0	0	1	0	1	0	0	0	2	0%
99	4	2	0	3	3	5	3	1	3	4	12	16	75%
100	5	1	3	0	2	5	2	3	2	2	11	14	79%

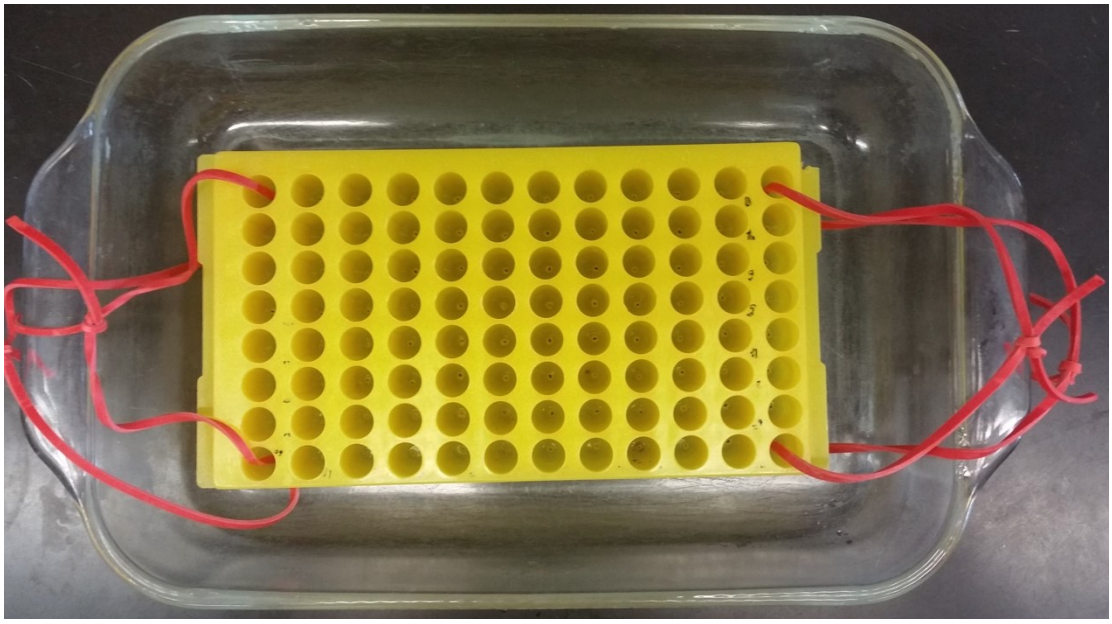

Figure 3.8 An apparatus for seed coat removal is shown in a photograph. Batch style acid scarification was used to remove the thick seed coats of *G. soja* seeds in a high throughput fashion. Small holes were bored through the bottom of a 1.5 mL micro-centrifuge tube holder for scarification in 100% H_2SO_4. After 5 minutes, the apparatus is carefully picked up and residual acid is drained back into the bath. Washes were performed in diH_2O to remove residual acid. The appropriate safety precautions are essential due the volatility of this strong acid. All material must be acid resistant and care must be taken to avoid splashing or dipping.

Chapter 4

Development of a Differential Clustering Algorithm to Identify Resistance and Susceptibility Mechanisms in the *Hpa*/Arabidopsis Plant Pathosystem

Michael G. Fedkenheuer[1*], Kevin E. Fedkenheuer[1], John M. McDowell[1]

[1]Department of Plant Pathology, Physiology, and Weed Science, Virginia Polytechnic Institute, Blacksburg, VA

[*]for correspondence fedkenmg@vt.edu

Author Contributions: Michael G. Fedkenheuer wrote this manuscript, prepared all figures, and generated all gene clusters, and performed percent representation analyses. Kevin E. Fedkenheuer developed the percent annotation representation analysis and provided expertise in Bayesian Biclustering and helped with developing the differential time point approach. Dr. Song Li reviewed the methodology in this manuscript.

Abstract

In this study we utilize complementary data sets to analyze the plant/pathogen interaction between *Arabidopsis thaliana* ecotype Col-0 and *Hyaloperonospora arabidopsidis (Hpa)* isolate Emwa1. We analyzed microarray data to understand how nearly 6,000 genes are differently modulated during a time course following inoculation of *Hpa* onto resistant and susceptible plant accessions. We optimized large gene sets and used Bayesian Biclustering (BBC) to identify a novel susceptibility cluster that shared expression patterns with a published pathogen susceptibility target, CCA1. BBC was limited in that it was not able to place 1/3 of genes into expression clusters and many expression clusters were redundant (many "unique" clusters had extremely similar expression profiles) or contained overlapping expression profiles. To overcome the aforementioned limitations, we developed a new algorithm to sort genes into time point-specific clusters using a gene specific activation/repression parameter. This clustering schematic is amenable to a differential analysis because there is a limited number of biologically relevant gene patterns. *RPP4*, responsible for *R* gene-mediated immunity, was part of a major resistance cluster (MRC) containing genes with activation or repression specific to 2 days post inoculation. We determined percent annotation representation within clusters for cellular compartment, biological processes, and molecular function to compare percent annotation representation between clusters. We used these relationships to determine mechanisms of plant resistance and pathogen susceptibility underlying *RPP4*-mediated immunity. Additionally, we applied a targeted approach to identify defense gene candidates that are involved in immunity. T-DNA

insertion mutants were screened with *Hpa* Emwa1 to verify defense gene candidates selected by our analysis. We observed strongly enhanced susceptibility in knockouts of the *Zar1* and *RLK7* genes, which had not been previously associated with resistance to *Hpa*.

Introduction

Plant disease is not always apparent but outbreaks can be severe. Oomycete plant pathogens have been reported as a major agronomic risk factor. We began our study with the intention of understanding resistance and susceptibility in a plant-pathogen interaction. We chose to study the model system *Arabidopsis thaliana* (*A. thaliana)* and the oomycete pathogen *Hyaloperonospora arabidopsidis* (*Hpa*) [24]. *Hpa* belongs to a group of obligate biotrophs known as downy mildew pathogens. Downy mildews are closely related to *Phytophthora* species [204]. *Hpa* has reduced gene families for necrosis and ethylene-inducing (NEP1)-like proteins and RXLR effectors, consistent with evolution toward an obligate relationship with the host [205]. Sequencing of the *Hpa* isolate Emoy2 revealed reductions or losses in metabolic pathways as well as a significantly reduced effectorome when compared to species in the *Phytophthora* genus. During evolution, *Hpa* shed genes related to motility and adhesion to evade PAMP triggered immunity [205]. Downy mildews cause disease on over 800 plant species including important crop plants [206]. By studying the interaction between *Hpa* and *A. thaliana* we hope to gain insight into combatting devastating downy mildew pathogens of crop plants.

Due to its obligate nature, one of the best ways to study *Hpa* is to study the host during a time course after inoculation. In this study, we analyzed a time course microarray experiment published in 2011 by Wang et al. [199]. In this study, *A. thaliana* seedlings from the resistant accession Columbia (Col-0) were inoculated with Emwa1 and analyzed over a six-day time course experiment. In Col-0, resistance is conferred by

the resistance *(R*) gene RPP4 [199]. The RPP4 resistance gene provides protection not only against the isolate Emwa1 but also against Noco2 and Emoy2 [207]. This *R* gene directly or indirectly impacts at least 12 defense components [207]. However, much remains to be learned about the mechanisms underpinning resistance in this interaction.

After analyzing data from microarray experiments on resistant and susceptible plants, Wang et al. selected 106 differentially expressed genes from the 2 DPI time point [199]. SALK T-DNA insertion mutants were screened using the *Hpa* isolates Emwa1 and Noco2. This analysis revealed 22 genes that showed enhanced susceptibility when challenged with *Hpa* [199]. This analysis demonstrated that RPP4 regulates at least two separate responses. They speculated that RPP4-mediated resistance not only drives programmed cell death but activates genes involved in general defenses [199]. Additionally, the authors of this study demonstrated that the gene CCA1, a circadian regulator, is important for RPP4-mediated resistance. To increase information from early time points in infection, Wang et al. used RNA Annealing Selection Ligation-Sequencing (RASL-seq) to examine the expression pattern of 22 defense genes that showed enrichment for an "evening element" in promoter analysis [199]. This analysis showed that RPP4-mediated defense does not disrupt the circadian clock; rather, it activates CCA1. Although the authors uncovered a novel aspect of the *Hpa-Arabidopsis* interaction, they did not provide a comprehensive analysis of the data set beyond the aspects described above. Thus, an opportunity exists to mine these data for additional insights into the molecular determinants of the interaction between *Arabidopsis* and *Hpa*.

As has been documented for other plant-pathogen interactions, a cursory examination of these *Hpa* infection data sets revealed huge differences in gene regulation

between resistant and susceptible plants. We initially attempted to analyze these complex data using established clustering techniques to organize genes into a biologically relevant context. We used a comparative approach to eliminate genes that were not differentially modulated during pathogen infection. Gene expression is often analyzed using biclustering, a technique which was first introduced by Cheng et al. [208]. We used a well-accepted technique for analyzing gene expression data known as Bayesian Biclustering (BBC) [209]. Determining the total number of clusters for a particular data set is difficult due to the inherent lopsided dimensionality of transcriptome data. The Bayesian information criterion is often used to provide a metric to establish comparisons (BIC) [210].

We used Bayesian biclustering to examine genes that are specifically induced or repressed in either the resistant or susceptible interaction, as well as those that were induced/repressed similarly in both data sets compared to a mock inoculated control (shared). We had limited success using the BBC model. We were able to identify a unique susceptibility cluster which shared infection specific expression with CCA1 [209]. Unfortunately, BBC generated redundant clusters with overlapping expression patterns and could not identify appropriate clusters for many genes. In response to these limitations, we developed a new approach. We developed a differential clustering algorithm to group genes based on time point specific expression. Using this strategy we were able to cluster all significantly expressed genes. Among these clustered genes were a set of 2,000 genes which shared expression in resistant (R) and susceptible (S) gene sets. This group of genes could not be clustered by established protocols.

We validated our methodology by screening SALK T-DNA insertion mutants for genes identified by this analysis. We challenged Col-0 mutant accessions with *Hpa* Emwa1 to validate our technique as well as to assist in understanding both pathogen infection strategies and host defense response. From this analysis we identified *Zar1* and *RLK7* mutants as having strong enhanced susceptibility phenotypes.

Results

Bioinformatics Approaches and Transcriptomics Data

Transcript profiling can be used to analyze interactions from many different perspectives. Analyses can be tailored to focus on small gene sets related to specific processes or gene families (micro-analyses), while other analyses can focus on large gene sets to identify broad changes (macro-analyses). To generate a complete picture, data should be analyzed at multiple levels to achieve complete understanding. Fundamentally, transcriptomics data are organized in array format. A simple array can be designed by ordering X and Y parameters such that they are linked by position to at least one dimension. In both microarray and RNA sequencing experiments, the X parameter is the gene name and the Y parameter is the expression level of that gene. Because *A. thaliana* contains over 26,000 genes, arrays generated from transcriptomics experiments are not evenly distributed. In an evenly distributed array there is an equal number of X and Y parameters. Although gene expression level (Y) changes over time gene ID (X) does not. This allows the complexity of the array to be increased in the Y direction by adding time points in parallel. Thus, in time course experiments the arrays are more balanced because expression values are stacked together (added) while the X parameters are merged. As a result this array now encompasses a new variable time (Z). Time course experiments should evaluate a complete period of infection in susceptible plants encompassing tissue entry, colonization, and reproduction. Using the period of infection in susceptible plants for comparison, it is possible to evaluate when defense responses begin and culminate in resistant plants.

Designing a Bioinformatics Approach to Compare Resistant and Susceptible Plants

The array parameters in our study were gene ID (At#) x Relative Fluorescence Units (RFUs). When analyzing microarray data, RFUs can be described as light units emitted during data collection. Since these light units are relative for each gene depending on the composition of the microarray chip, units are standardized by Affymetrix. After RFUs are normalized the array is redefined as At# × Relative Gene Expression (RGE). The relationship between At# and RGE was kept constant during all analyses. We focused on a transcript profiling experiment conducted by Wang et al. [199]. The experiment compared transcriptomes from *Hpa* infection of *At*Col-0 (resistant, "R") and the Col-*rpp4* mutant (susceptible, "S") using material collected at the 0.5, 2, 4, and 6 days post infection (DPI). Importantly, resistance in Col-0 is activated by the immune receptor protein *RPP4*, which perceives a *Hpa* effector protein Atr4. In *rpp4*Col-0, this gene is mutated and non-functional, thereby disabling effector-triggered immunity and allowing pathogen colonization. Thus, transcriptomic comparisons of these two genotypes provides an opportunity to identify components of the defense response that function in, or downstream of, RPP4-dependent pathogen resistance.

Structuring our Hpa Infection Array to Analyze Resistance versus Susceptibility

We began by generating separate time course arrays for R and S by ordering time points sequentially. We merged these arrays by linking gene expression parameter with their respective designation, R or S. When the designation R or S is added, a new dimension is added to the array. Since the R or S designation does not affect the arrangement of At#s, this second dimension should intersect At#. If both R and S data sets are normalized to a control array these planes should be parallel. If these planes are

parallel and are bisected by a common plane the difference between resistance and susceptibility should be evenly spaced by this geometric relationship. We used the water mock-inoculated sample collected by Wang et al. at 0 DPI as a control array. We normalized RGE at each time point to the water mock to create a ratio for each time point. This ratio was transformed using a Log2 ratio such that gene expression at the 0 DPI for every gene in the array was 0.

Clustering Genes for Downstream Analyses

By generating expression clusters for resistant (R) and susceptible (S) plants, the differential relationships between R and S clusters can be explored to determine correlation with other variables. The amount of identical annotations can be catalogued for genes within resistant or susceptible expression clusters. By comparing the amount of like annotations among genes with the total number of genes within a cluster we established the metric, “percent annotation representation”, for each cluster [211]. Thus percent representation is not only a factor of X but is now a factor of the clusters, XY. We can derive this relationship further by comparing percent representation across resistant and susceptible clusters to establish a link between gene function and expression with respects to pathogen infection. Since annotations can range from broad to specific and do not contain exclusively one or the other, broad annotations were used. An example would be adding together genes from all cellular compartment annotations for chloroplast regardless of whether they are annotated for stroma, thylakoid, etc.

Identifying Significant Genes in Resistant and Susceptible Plants

We assessed normalized distribution of gene expression using volcano plots. Using these plots we observed a significant fraction of genes above and below a 1.5 Fold

Change (FC) expression in response to *Hpa* Emwa1 [212]. We grouped genes into a broad category known as the *Hpa* specific gene cluster (*Hpa*C) defined by gene expression exceeding a 1.5 FC cutoff for upregulation or downregulation at any time point in resistant or susceptible plants. If a gene met the criteria in one or multiple time points in R but did not meet the criteria in S, the gene was considered to be part of the Resistant Cluster (RC). If the inverse statement was true, the genes were considered to be part of the Susceptible Cluster (SC). If a gene was significantly activated or repressed at time points in both resistant and susceptible plants, it was considered part of the Shared Cluster (RSC).

Understanding Gene Flux in Response to Hpa Emwa1

We analyzed gene sets at each stage of the clustering process to gain a unique perspective on the data after each transformation. After removing non-significant expression using FC cutoffs, nearly 6,000 genes were implicated during infection (Table 4.1). The amount of genes with significant changes in gene expression in an example of single gene mediated resistance versus susceptibility was striking. 2237 genes were specific to the resistant interaction and 1317 genes were specific to the susceptible interaction (Figure 4.2 A). Additionally, 2054 genes showed expression in both R and S data sets. To identify more specific host gene targets which might be manipulated during compatible and incompatible interactions with *Hpa* Emwa1, we used Bayesian Biclustering (BBC) software to cluster RCs, SCs, and RSCs further by co-expression in individual data sets.

Optimizing Bayesian Biclustering and Understanding the Limitations of this Technique for Analyzing Complex Interactions

We analyzed resistant and susceptible specific gene sets (RC, SC) using BBC. This technique was used to identify genes which shared unique expression profiles in resistant or susceptible plants. BBC identified over 50 RCs and 50 SCs many of which had very similar expression profiles. Many of these clusters were not significant with respect to metrics such as standard deviation and skewedness. Further inspection revealed that several of these clusters contained multiple overlapping gene expression patterns. We manually adjusted Bayesian clusters by merging similar clusters and removing others which did not show congruent expression. An analysis of finalized clusters revealed a large repressed cluster which shared a consistent expression profile with CCA1 in susceptible plants, characterized by downregulation at the 0.5 day time point. Genes with shared expression in resistant and susceptible plants (RSC) could not be accurately clustered using BBC, because the Bayesian algorithm is not effective with small gene sets. This is because BBC compares the expression of every gene against that of every other gene in the array at each time point to determine unique clusters. Thus, the composition of expression clusters varies depending on the number of genes available for comparison. As the X parameter of the array is reduced, the relationships described by the BBC can change making optimization difficult. Changing the X parameter also influences the total number of clusters that BBC will generate. Since the total number of clusters is determined using maximum likelihood estimations, a metric that depends on the distribution of X and Y in the array, changing the total number of genes will also affect the total number of clusters produced. Since the amount of clusters and cluster composition changes with array complexity it was difficult to further optimize BBC.

Developing a New Clustering Technique

We began to develop a technique that could better describe expression patterns over a time course of infection. Expression level can be biased when comparing any two genes. This is due to the intrinsic variability between the maximum and minimum amount of transcript produced for any given gene within the *A. thaliana* genome. For example, NBS-LRR genes are expressed at very low levels in uninfected tissues because activation of transcription of these genes above a threshold point triggers auto-immunity. Thus maximum and minimum levels of transcription will vary based on functional thresholds, which can vary for these genes. This is an example of how expression level bias can be a challenge when assessing what constitutes significant changes of expression in diverse gene families. In order to accurately profile gene expression we calculated gene specific parameters to describe significant activation versus repression. These parameters were used to determine a binary relationship which described whether a gene was significantly up or downregulated at particular time points during infection. We sequentially sorted this binary output with respect to time point to generate activated time course profiles for all genes. Thus gene sets were organized based on the time point at which genes were first activated in response to the pathogen. Gene sets were further grouped by the relationship between time points in resistant and susceptible interactions. Our clustering approach is outlined in Figure 4.1.

Activated versus Repressed Gene Clusters

Although NBS-LRR-dependent immunity is traditionally described as "gene-for-gene" resistance, in fact thousands of genes related to immunity are in flux during an *Hpa* infection of *A. thaliana*. We will refer to increases in gene expression by the term

activation, while we will refer to decreases in gene expression as repression. It is important to keep in mind that if a gene is expressed at low levels, increased repression may not be evident during transcript profiling. We used the log2 transformed values 0.3/-0.3 as a cutoff to establish a baseline for significant activation or repression in this stage of our analysis. Of the 6,000 genes we identified as being significantly expressed during infection, the majority were only either activated or repressed during time course experiments. We began with gene sets from RC, SC, and RSC which produced optimal clusters in BBC. We separated these categories by up and downregulation which yielded six clusters; resistance activated cluster (RaC), resistance repressed cluster (RrC), susceptibility activated cluster (SaC), susceptibility repressed cluster (SrC), shared activated cluster (RSaC), and shared repressed cluster (RSrC) (Table 4.2). Using this analysis we identified 500 genes which showed profiles of significant upregulation and downregulation during infection. These genes were clustered into the resistant, susceptible, and shared activated and repressed gene cluster (RarC, SarC, and RSarC) (Table 4.2). Shared clusters were then sorted into R and S specific activated and repressed clusters using the differential approach described below (Table 4.3).

Defining Gene Specific Activation/Repression Parameters to Normalize Differences in Rates of Transcription

We used minimum and maximum expression values determined from resistant and susceptible data sets to describe the limits of gene expression during pathogen attack. Maximum and minimum expression levels were multiplied by ½ to determine at what time point(s) a given gene was reaching exponential levels of activation or repression with respect to basal levels of expression at the 0 DPI. Since genes were normalized

using the log2 ratio derived from the water mock, this parameter is gene specific and describes the middle point on a logarithmic scale during a period of infection. We define this term as the activation/repression parameters. We consider this maximum or minimum value to be the pseudo Michaelis-Menten (M-M) Vmax of transcription. By using the ½ of maximum and/or minimum gene expression, we generate a pseudo M-M km to determine when catalytic efficiency of transcription is approaching maximum levels in response to pathogen attack. The intrinsic limitation of defining the activation parameter in this way is the uncertainty of whether the maximum and minimum expression values truly encompass maximum levels of activation or repression for a given gene. To minimize this effect and avoid skewing, we relied on FC cutoffs to minimize the contribution of genes that do not reach significant levels of expression during infection. We verified the activation/repression parameter through multiple clustering experiments.

Time Point Clustering of Resistance and Susceptibility Clusters (RCs, SCs)

We applied sorting to organize genes into smaller gene clusters which more accurately describe gene expression patterns in resistant or susceptible plants. We sorted upregulated genes using the binary output from the equation (if x>activation factor output 1; if x<activation factor output 0). We processed downregulated genes separately using the equation (if x<repression factor output 1; if x>activation factor output 0). The binary output from these equations was maintained in array format so sorting could be used to group genes by their activation parameter. We used sequential sorting to create time point specific clusters. By sorting for initial transcriptional activation/repression, we generated clusters for 0.5, 2, 4, and 6 DPI time points for up and downregulated genes (Table 4.4,

Figure 4.2 B). Since a gene could be implicated in more than one time point, sequential sorting allows us to expand time point clustering to account for the 32 possible combinations of binary gene movement over the time course. In this case the 0 DPI is included because all genes were normalized to the water mock. This allows us to use zero as a starting point for gene expression. Using this approach we identified 18 clusters which ranged in size from 1-555 genes per cluster (Table 4.5). We saw a clearly uneven distribution of genes within our time point clusters indicating a non-random distribution (Figure 2D). This approach is similar to algorithms used for clustering short gene expression data [213]. The time point clustering algorithm designed by Ernst et al. yields similar results to our algorithm in number of clusters and cluster composition when used to analyze short time course experiments [213]. We did not perform a direct comparison because this method was not amenable to the differential analysis performed on RS clusters.

To analyze these clusters we plotted them using normalized expression values. An inspection of the clusters revealed that expression profiles for clusters larger than 25 genes were tight, consistent, and most fell within standard error between 0.01 and 0.03. A cursory examination of time point clusters revealed that resistance clusters were activated or repressed at 0.5 and 2 DPI. This is in stark contrast to the susceptible interaction. In the susceptible interaction, clusters are split between 0.5, 4, and 6-day time points, with no specific gene activation/repression occurring at 2 DPI (Figure 4.2 B).

Differential Time Point Clustering of Shared (RS) Genes

Genes in the RSC have highly variable expression with respects to time point. The expression profiles of these genes are too variable to analyze with respect to either

individual time course experiment. Since gene expression in RSCs is shared between both arrays, we developed a differential clustering approach to derive significant differences with respects to either variable R or S. We use the same equations from the previous section to describe the relationship between shared genes with respects to R and S by using the difference (=R-S) of binary activation/repression parameters. R and S specific activated and repressed clusters are listed in Table 4.3. The differential array was then sorted by time point to generate initial time point clusters specific to R, S or a combination of R and S time points (Figure 2 C). If we take into account transcriptional up and downregulation this analysis can generate 32 possible R or S specific activation/repression clusters. We used sequential sorting to generate unique clusters ranging in size from 1-248 genes per cluster (Table 4.6, Figure 4.2 E). Clusters that met the activation/repression parameter specific to time points in R and S are listed in the final section in Table 4.6. Additionally, we listed activated and repressed genes which did not show a difference with activation/repression sorting in the categories NdaC and NdrC.

Major Resistance and Susceptibility Clusters were identified by Analyzing the Duality of Gene Expression in Finalized Clusters

Since expression patterns were very clear in resistant and susceptible specific unique time point clusters, we explored the relationship between activated and repressed clusters with matching expression proflies (Figure 4.3 A, B). If standard deviation between absolute values of activated and repressed clusters with matching pattern (e.g. RaC0.5a and RrC0.5a) was less than 0.01, we merged these clusters to generate major resistance or susceptibility clusters. These clusters contained activated and repressed

genes with matcing patterns and similar intesnsities specific to resistance or susceptibiltiy to *Hpa* Emwa1 (Figure 4.4 A, B). We referred to these paired clusters as major resistance or susceptibility clusters (Table 4.7).

Analyzing Resistance and Susceptibility Specific Clusters

We identified the major player in this interaction, *RPP4*, as a part of the Major Resistance Cluster 2 (MRC2a). In a resistant interaction, we would expect these stimuli to be indicatative of pathogen recognition. The presence of major resistance clusters MRC0.5a0, MRC0.5a1, and MRC0.5e suggests that multiple recognition events are subverted in the absence of the resistance gene *RPP4* (Table 4.7). Genes in the MRC0.5a0 were of particular interest because activation or repression of these transcripts were maintained throughout infection. MRC0.5a0 peak in expression at 0.5 DPI; however culmination of the immune response occurs at the 2 DPI during peak upregulation or downreuglation of genes in MRC2a, MRC2b, and MRC2d (Figure 4.4 A). Since early (0.5 DPI) MRCs clusters are not activated in susceptible plants and transcriptional activation of RPP4 occurs after these gene clusters, we can assume that *RPP4* protects early defense signaling even at basal levels of expression.

In susceptible plants, expression data suggests that immunity is subverted by modulation of transcription from before the 2 DPI. Of the 1317 genes with significant expression specific to susceptible plants, 258 genes showed repression specific to 0.5 DPI (Figure 4.2 E). Genes in these categories belonged to the major susceptibility cluster identified in BBC clustering of gene expression specific to susceptible plants. These repressed genes were part of major susceptibility clusters MSC0.5e, MSC0.5h (Figure 4.4 B). In a susceptible interaction, genes in the MSC0.5e and MSC0.5h are activated early in

infection but after this initial impetus return to basal at 2 DPI. It is posisble that these clusters are modulated by circadian timing due to the diurnal pattern. Interestingly, no genes were activated or repressed specific to susceptibility at 2 DPI. For these reasons we believe that in suceptible plants, immunity is subverted prior to the 2 DPI. Additionally, the unique drop in activation of all gene clusters at 2 DPI giving further evidence to suggest that the pathogen has achieved "stealth" at this point in infection. Genes in the MSC4a cluster were differentially expressed late in infection (Figure 4.4 B). This unique activation of susceptibility specific factors late in infection suggests that Emwa1 is resuming subversion of transcription late in infection to prepare for sporulation.

Utilizing Cellular Location, Biological Process, and Molecular Function to Gain Perspective on the Hpa/Arabidopsis Pathosystem

A. thaliana, the model plant system, is utilized for its extensive genomic resources. Gene localization and/or function information is available for most *A. thaliana* genes. The coverage of annotations amongst genes in the most commonly described categories; biological process (BP), molecular function (MF), and cellular component (CC) was between 25-50%. The cellular component category was the most well represented category and contained localization information for nearly half of the *A. thaliana* genes significantly expressed in our data set. Although the remaining categories were somewhat incomplete, genes involved in pathogen related processes have been extensively described and annotated. The function of LRR proteins such as RPP4 exemplify important immune mechanisms such as gene-for-gene resistance observed in numerous cultivar/pathotype interactions. Gene expression and function are intrinsic factors of each other because the amount of available transcript can directly impact the

speed and efficiency of translation. Since protein function is generally constant during pathogen attack while expression is variable, expression and function should intersect each other in accurate R or S expression clusters. Due to the relationship between expression and function we expect percent representation to increase in gene groupings as we apply annotations to smaller more accurate clusters. We observed this increased representation in many R, S, and RS gene clusters. Table 4.8 lists all annotation represented by ≥ 5 genes and also encompassing ≥ 10% of the total genes that cluster for the GO category cellular component. GO categories for biological process and molecular function that contained ≥ 3 genes and encompassed ≥ 5% of the total genes in that cluster are shown. The top three to four annotations are shown for each group when possible.

Annotations were applied to Finalized Clusters Hpa/Arabidopsis to Establish an Infection Framework

We organized all unique clusters and annotated the most significant processes and localizations to produce functional resistant versus susceptible *Hpa* interaction frameworks (Figure 4.5 A, B). In a resistant interaction, the plant responds by heavily regulating transcription to quickly respond to pathogen attack after recognition. Resistant clusters are enriched for membrane and nuclear cellular components in most unique clusters (Figure 4.5A). Resistant expression clusters show ubiquitous enrichment in the categories signal transduction, oxidation-reduction, ribosomal composition, protein binding, and metal ion binding throughout infection (Table 4.8). We compared annotations between activated and repressed clusters to determine that unique 0.5DPI activated clusters were enriched specifically for activation of signal transduction. Unique 2 DPI activated clusters were enriched for regulation of transcription, defense genes, and

kinase activity when compared to unique 0.5 and 2 DPI repressed clusters that were enriched for specific processes related to ribosomal processing. The shared cluster uRS-RaC2d showed the highest representation among any cluster for genes related to defense response (18%) and signal transduction (13%) (Figure 4.5A).

We observed very significant enrichment in shared expression clusters with repression specific to resistance. The cluster uRS-RrC2b was repressed specific to a resistant interaction at 2 DPI contained genes whose repression is delayed in a susceptible interaction. 30% of annotated genes in the uRS-RrC2b were localized to the chloroplast. These genes are involved in a variety of stress and hormone responses (Table 4.8). Additionally, this cluster was heavily enriched for ribosomal components, translation, and ribosome biogenesis. This cluster contained the highest number of genes annotated as structural constituents of ribosome (31%) (Figure 4.5A). The composition of uRS-RrC2b was similar to another RS cluster with resistant specific repression at 2 and 4 DPI, uRS-RrC2d, which was also enriched for ribosomal proteins.

In a susceptible interaction several processes are ubiquitously enriched during infection. Oxidation-reduction, metal ion binding, protein binding were enriched in most unique susceptibility clusters (Figure 4.5 B). These clusters were primarily enriched for chloroplast, membrane, and nuclear components. At the 0.5 DPI, defense genes, kinases, signaling pathways, and transcription factors are specifically repressed when compared to activated clusters. All but one of these genes return to basal levels at the 2DPI and there are no significant gene clusters activated or repressed specific to 2 DPI in a susceptible interaction. Late in a susceptible interaction starting at 4 DPI transcription factors are activated. At the 6 DPI signal transduction, hydrolase activity, and nutrient reservoirs are

enriched in activated susceptibility clusters. Processes related to ribosomal processing were heavily enriched in 4 and 6 DPI repressed clusters (Figure 4.5B).

Using our differential analysis, we clustered genes with variable expression profiles during pathogen attack. Genes from late differential shared activated clusters were best represented by the groups of kinases, defense response genes, and transcriptional regulators (Table 4.8). Genes in RS clusters with activation specific to 6 DPI were enriched for defense genes and kinases. This is likely due to the activation of immune responses that are no longer suppressed or able to be suppressed by *Hpa*. Since these genes were activated in the resistant interaction at earlier time points but the resistance response has concluded at the 4 and 6 day time points, we can assume that these genes would likely have remained upregulated with continual pathogen stimuli.

Using Infection Frameworks to Perform Targeted Analyses

RPP4 is one of the most highly expressed NBS-LRRs during infection with Emwa1. Additionally, it belongs to a major resistance cluster identified in our framework (Figure 4.5). We developed a targeted analysis to identify genes similar to *RPP4*. One drawback to this approach is that we cannot compare the true differential expression of *RPP4* since the gene is knocked out in this compatible interaction. Annotations were used to identify defense genes. We examined genes annotated as serine/threonine kinases (Ser/Thr kinaes), NBS-LRRs, receptor like kinase (RLKs), and receptor like proteins (RLPs). We utilized differential clustering to group genes in the method described above. We selected defense genes with the highest expression specific to initial time point cluster. We selected defense genes from many different gene clusters to understand the contribution of defense genes across multiple infection specific clusters. This type of

analysis can be applied to specific biological processes or gene families to elucidate subtle manipulations of host processes that are less evident in percent annotation analysis.

Analyzing Defense Genes using T-DNA Insertion Mutants

We acquired T-DNA insertion mutants from the SALK institute. Mutants were selected by tailoring our approach to find defense genes which were significantly regulated during infection (Table 4.9). Since *RPP4* is specific to the MRC2 we believe that NBS-LRRs and defense genes from other infection specific clusters might show contributions to resistance. Mutants which produced an enhanced susceptibility phenotype were used to confirm the importance of many defense genes in infection specific immune clusters. We believe that some of these genes provide basal resistance against *Hpa* while others are involved *RPP4* immunity. We identified receptor kinases *Zar1* and *RLK7* as being strong contributors to immunity because they produced the strongest enhanced disease susceptibility (EDS) phenotypes (Table 4.9). Although these genes appear to be involved in basal immunity during an *Hpa* infection, their resistant and susceptible profiles are different in the presence or absence of *RPP4*. This suggests that regulation of these genes is indirectly related to *RPP4* resistance mechanisms.

Discussion

We developed a new clustering approach in order to better understand changes in gene expression during compatible and incompatible *Hpa* infection. We began by studying immunity mediated by *RPP4*, which activated resistance of *A. thaliana* to *Hpa* Emwa1. We acquired microarray data from Wang et al. in hopes of identifying genes involved in innate immunity or additional components of *RPP4* triggered immunity. We investigated susceptibility to understand general susceptibility mechanisms. We approached this analysis from a broad to a specific perspective by generating increasingly more focused gene clusters (Figure 4.1). We used clusters to correlate gene function and expression during compatible and incompatible pathogen infections. We used an annotated framework to analyze the *Hpa* pathosystem by comparing infection specific time point clusters against one another. We identified highly significant correlations between function and expression profile through these comparisons.

We began by removing non-significant data from both arrays. We used fold change cutoffs to remove roughly 18,000 genes which were not significantly up or downregulated in response to pathogen attack (Table 4.1). We were unable to achieve complete and accurate clustering of the entire data set using BBC. Optimization was difficult because results vary with sizes of gene sets. To get a more accurate profile of resistance versus susceptibility we developed a clustering algorithm to derive significant changes in expression pattern. We clustered sequentially by time point specific activation or repression to analyze time course data. We determined maximum and minimum expression levels of genes to define the activation/repression parameter. We consider this

maximum or minimum value to be the pseudo-Vmax of transcription. We determined our activation/repression parameter by taking ½ of maximum and/or minimum gene expression. We use this value as a pseudo km to determine when catalytic efficiency of transcription is approaching maximum levels in response to pathogen attack. This technique is advantageous because the number of clusters is defined by the complexity of the time course experiment, groupings are not biased by the size of the gene set, and it is amenable to differential analysis.

We began by grouping genes by resistant and susceptible specific expression clusters (RCs and SCs). Genes which exhibited significant expression in at least one time point in R and S were considered shared (RSC) (Table 4.2, Figure 4.2 A). The activation parameter was used to sort genes for significant activation at specific time points (Figure 4.2 B). By sorting expression values against the activation parameter we were able to identify which time point a gene was first activated in response to *Hpa*. We applied a similar approach to genes which has significant shared (RS) expression among compatible and incompatible *Hpa* interactions (Figure 4.2 C). We were unable separate genes in shared clusters in individual (R or S) using any bioinformatic technique including our activation based sorting. Instead we clustered genes by the difference in activation between R and S with respects to the time course of infection. Genes were sorted into initial time point clusters for shared genes by applying this method (Table 4.4). We used the first activated time point to define a genes overall grouping. We sorted genes sequentially for time points 0.5-6 DPI to develop unique expression patterns. Since we used a binary relationship to describe gene activation, a maximum of 32 possible gene clusters were available between upregulated and downregulated clusters (Table 4.5).

Since genes with shared expression were sorted using our differential algorithm a maximum of 32 possible clusters were also possible for this set of genes (Table 4.6). We refer to these more specific clusters as unique time point clusters unique time point patterns. Unique clusters were analyzed with respect to cellular component, biological process, and molecular function (Table 4.8). We used these annotated clusters as a framework to describe the *Hpa* interaction and identify mechanisms involved in resistance or susceptibility to *Hpa* Emwa1 (Figure 4.5 A, B).

We began validating unique resistance and susceptibility clusters using statistics and graphical approaches (Figure 4.3 A, B). We observed that certain expression profiles for upregulated genes exactly matched clusters of downregulated genes in both pattern and intensity. We observed this phenomenon in 6 resistance clusters and 4 susceptibility clusters (Figure 4.4 A, B). The presence of matching pairs indicate that both up and downregulated genes are being regulated by the same stimuli. Activated and repressed genes that shared this unique relationship, were grouped together if standard deviation and standard error improved (Table 4.7). We referred to these highly significant groupings as major resistance or susceptibility clusters.

We observed significant gene flux in both compatible and incompatible interactions. Although gene flux of 20-30 percent is common during pathogen attack we did not expect to see that these percentages when comparing resistance versus susceptibility caused by a single gene knockout. We identified 4,291 genes which were implicated in the resistant data set and 3,371 genes which were significantly up or downregulated in the susceptible data set. Of these significantly regulated genes, we identified 2,054 genes with shared expression in both data sets (Table 4.1). A total gene

analysis of differentially expressed genes shows the clearest distribution in gene expression at 2 DPI (Figure 4.2 B). There are nearly 1800 genes from 10 unique expression clusters which show high levels of activation at 2 DPI during incompatibility. Conversely, in a susceptible interaction, there were no significant clusters at this time point. In fact only 9 genes were significantly modulated at this time point. Very few genes are expressed during an incompatible interaction following 2 DPI (Figure 4.2 B). This indicates that the immune response culminates at the 2 DPI. RPP4 reaches peak expression at 2 DPI and belongs to an activated initial time point cluster containing 525 genes with significant expression specific to the 2 DPI (Table 4.4). An analysis of major resistance clusters revealed 0.5 DPI resistance clusters reached peak expression before significant activation of RPP4 and MRC2a (Figure 4.4 A). Since these early clusters are not implicated in a susceptible interaction, we can assume that effector recognition by RPP4 preserves the integrity of multiple defense gene clusters. This indicates that recognition of the pathogen takes place in the absence of RPP4 suggesting regulation by non-RPP4 machinery. By the same reasoning it is also likely that the effector atr4 functions to silence pathogen recognition and defense response. Despite being transcriptionally activated later in infection, basal levels of RPP4 appear to be sufficient for protecting the integrity of multiple defense signaling pathways.

Our analysis of major resistance and susceptibility clusters indicates that RPP4 recognition leads to the activation of PCD and a subset of defense genes as previously reported [199]. We were able identify an RPP4 dependent cluster which peaked in expression during at 2 DPI. A global analysis of the *Hpa*-Arabidopsis interactions was recently performed by Krasileva et al. [214]. This group showed that recognition of

ATR1 and ATR13 is evolutionarily dynamic. The observed variability in gene-for-gene resistance led them to propose a broad genome-for-genome perspective instead [214]. Our data indicates that *RPP4* mediated resistance cannot be explained by a simple gene-for-gene model in a *Hpa* Emwa1 infection of *A. thaliana*.

We analyzed enrichment in gene function and localization using annotated resistance versus susceptible *Hpa* interaction frameworks (Figure 4.5 A, B). In a resistant interaction, the plant responds by heavily regulating transcription to quickly respond to pathogen attack after recognition. Resistant clusters are enriched for membrane and nuclear cellular components in most unique clusters (Figure 4.5A). We compared annotations between activated and repressed clusters to determine that unique 0.5DPI activated clusters were enriched specifically for activation of signal transduction. Unique 2 DPI resistance activated clusters were enriched for regulation of transcription, defense genes, and kinase activity when compared to unique 0.5 and 2 DPI. The shared cluster uRS-RaC2d showed the highest representation among any cluster for genes related to defense response (18%) and signal transduction (13%) (Figure 4.5A). Resistant repressed clusters specific to the 2 DPI were enriched for specific processes related to ribosomal processing and translation.

Shared cluster, uRS-RrC2b, was repressed specific to a resistant interaction at the 2 DPI and showed significant enrichment for localization to the chloroplast. These genes were involved in a variety of stress and hormone responses (Table 4.8). Additionally, this cluster was a heavily enriched for ribosomal components, translation, and ribosome biogenesis. This cluster contained the highest amount of genes annotated as structural constituents of ribosome (31%) (Figure 4.5A). The composition of uRS-RrC2b was

similar to another RS cluster with resistant specific repression at 2 and 4 DPI, uRS-RrC2d, which was also enriched for ribosomal proteins. This indicates that repression of translation is part of the culmination of the defense response.

In a susceptible interaction several processes are ubiquitously enriched during infection. Oxidation-reduction, metal ion binding, protein binding were enriched in most unique susceptibility clusters (Figure 4.5 B). These clusters were primarily enriched for chloroplast, membrane, and nuclear components. At 0.5 DPI, defense genes, kinases, signaling pathways, and transcription factors are repressed. All but one of these gene clusters return to basal levels at 2DPI. There are no significant gene clusters activated or repressed specific to the 2 DPI in a susceptible interaction, indicating stealth. Late in a susceptible interaction starting at 4 DPI transcription factors are activated. At 6 DPI signal transduction, hydrolase activity, and nutrient reservoirs are enriched in activated susceptibility clusters. This suggests that *Hpa* is acquiring nutrients for sporulation. Processes related to ribosomal processing were repressed in susceptible plants at 6 DPI which may indicate a delayed immune response at 6 DPI (Figure 4.5A, B).

Genes from late differential shared activated clusters were best represented by the groups of kinases, defense response genes, and transcriptional regulators (Table 4.8). Genes in RS clusters with activation specific to 6 DPI were enriched for defense genes and kinases. This is likely due to the activation of immune responses that are no longer suppressed or able to be suppressed by *Hpa*. Since these genes were activated in the resistant interaction at earlier time points but the resistance response has concluded at the 4-6 day time points, we can assume that these genes would likely have remained upregulated with continual pathogen stimuli. This gives further evidence to a delayed

immune response which begins to occur prior to sporulation in the absence of the resistance protein RPP4. The genes in these clusters are likely related to a basal resistance to *Hpa*.

A major battleground during an Emwa1 infection of *A. thaliana* is transcription. It is clear from both data set that large gene clusters are activated specific to compatibility and incompatibility with *Hpa* Emwa1. In a resistant interaction, there are a large proportion resistant specific genes involved in transcriptional regulation which are not activated during susceptibility (Figure 4.5 A). Additionally the most well represented cellular compartment across resistant specific clusters is the nucleus. We observed a smaller subset of genes related to transcriptional processes in early susceptibility clusters. Late downregulated susceptibility clusters are enriched for fatty acid biosynthesis and translation (Figure 4.5 B). This provides further evidence that Emwa1 is reprogramming host processes to make the cellular environment more suitable for sporulation. It is clear from this data that the plant regulates transcription to coordinate a defense response during incompatibility, while the pathogen regulates transcription to subvert recognition during compatibility.

There was a significant enrichment in chloroplast localization in both resistant and susceptibility gene clusters. An enrichment of 20-35% was seen ubiquitously amongst repressed susceptibility clusters. Resistant repressed clusters were also enriched for chloroplast localization; however percent representation was only 15-25%. A closer inspection of biological process and molecular function categories indicate that a wide variety of processes are being repressed in these clusters encompassed by response to hormones, response to abiotic stress, and metabolism. These processes were similar

between resistance and susceptibility clusters indicating that these processes are targets of pathogen manipulation.

A targeted annotation analysis was used to look at the regulation of defense genes during infection. *A. thaliana* T-DNA gene disruption mutants were acquired from the SALK institute for the most significantly expressed Ser/Thr kinases, RLPs, RLKs, and NBS-LRRs from R, S, and RS clusters (Table 4.9). This analysis was used to identify strong enhanced susceptibility mutants *Zar1* and *RLK7*. We optimized differential activation sorting to minimize elimination of potentially significant data. The *RLK7* gene was identified as a border line candidate in our analysis; however visual inspection revealed that its expression pattern in R was offset by half a day from its expression in S. We selected *RLK7* because of this unique relationship. From our defense gene analysis, we identified *RLM2* which belonged to the earliest resistant specific gene cluster and was highly expressed (Table 4.9). Based on expression and pattern we expected a knockout in this gene to produce a strong phenotype. We observed weak susceptibility phenotype when *RLM2* was disrupted. When analyzing this result in our framework we see that in the presence of *RPP4* causes rapid signal transduction leading to the transcription of defense genes. Even if RLM2 were responsible for driving a large defense gene cluster, the cumulative defense responses is coordinated by at least 9 other unique gene clusters during a resistant interaction. Since receptor kinases *Zar1* and *RLK7* produced strong EDS phenotypes and belonged to shared expression clusters, we implicated these genes in basal immunity to *Hpa*. The *RPP4* dependent regulation of these genes early infection suggests that regulation of these genes is indirectly related *RPP4* resistance mechanisms.

We performed a cursory examination of the recent RNA-seq time course experiment between Col-0 and *Hpa* isolates Waco9 (S) and Emoy2 (R) [215]. The Emwa1 susceptibility mutants, Zar1 and RLK7, were differentially modulate for resistance and susceptibility in these data sets as well. Zar1 has been shown to be a major player in resistance against *Pseudomonas syringae [216]*, and RLK7 is involved in germination speed and tolerance to oxidative stress [217]. Additionally, we have determined that although, RLM2 is not upregulated in either interaction, a related LRR, RLM3 is heavily activated early in infection with Emoy2. From the literature we know that RLM proteins and Zar1 appear linked in the *A. thaliana* evolutionary history against the causal agent of blackleg disease, *Leptosphaeria maculans* [218]. Zar1 has been implicated in immunity against bacteria, fungi, and now in immunity against an oomycete pathogen.

It can be difficult to compare data when data is generated by a different technique and collected at different time points. We hope to utilize our complete *Hpa* infection frameworks to compare transcriptome experiments with different experimental variables. Clustering can be used to organize genes into gene sets to begin building an infection framework. Clusters can be compared over the time course of infection to create an arbitrary plane upon which new variables be normalized and integrated. We used this strategy to define the infection period and normalize uneven sampling between time points. We classified clusters by early, middle, or late infection instead of a strict time point scaling (Figure 4.5). Levels of sensitivity between data acquisition methods are comparable because significant regulation is measured by the activation parameter, a metric that normalizes maximum and minimum expression.

Building a comparable framework to describe the interaction between *A. thaliana* and *Hpa* isolates Waco9 and Emoy2 is the next step in this process. Our framework is designed to be flexible enough to normalize experimental variables. The flexibility offered by our system allows the user to increase or decrease the stringency of the framework to fit the desired analysis. The addition of biologically relevant resistant and susceptible *Hpa* data sets will allow us to understand conserved immune mechanisms and to clarify isolate/accession specific differences. Unfortunately no analysis can make up for incomplete sampling. For example, when Jones et al. sampled resistant and susceptible interactions using *Hpa* Waco9 and *Hpa* Emoy2 at 1, 3 and 5 DPI they did not account for the resistance response of *A. thaliana* to Emoy2 being nearly cooled down by 3 DPI. This makes a comparative time point analysis difficult since gene movement at 3 and 5 DPI in R is minimal. It is important the sampling be done to capture early time points in pathogen infection.

From our analysis of compatible and incompatible *Hpa* Emwa1 interactions it is clear that resistance culminates at 2 DPI. Although defense responses culminate at 2 DPI, clear difference between resistance and susceptibility are clearly observed at 0.5 DPI. Strong defense responses are observed at the earliest point sampled leading us to conclude that the determination of resistance versus susceptibility took place prior to sampling at the 0.5 DPI. A similar trend is observed when analyzing early time points in the compatible interaction. Defense responses are suppressed and susceptibility clusters are activated early in infection. This rapid transcriptional reprogramming induced by *Hpa* must be further analyzed to increase our understanding of early infection. We hope to utilize increasingly available transcriptome and annotation data to draw stronger

functional relationships between the interaction between *Hpa* and *A. thaliana*. Applying this analysis to other complementary data sets will help gauge the effectiveness of our differential algorithm in describing relationships in diverse systems.

Materials and Methods

Curating the Data

The raw data is available online at NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE22274. This data is backed up on a lab external hard drive. Q values, obtained from a mixed model ANOVA, were sent electronically by Dr. Wang via email and are saved on a lab external hard drive. The Affymetrix Chip contains 20,922 unique genes. The data was analyzed statistically using Q values cutoffs. The data was transformed by Log2 Expression. Data were normalized to the water mock at 0 DPI. All transformation of the data from this point were done using expression data in the form of Log2 normalized values.

Merging Data Sets

We began by merging curated time course data from data sets Emwa1 x Col-0 and Emwa1 x *rpp4*Col-0. Each data set contained expression data from time points 0.5 DPI, 2 DPI, 4 DPI, and 6 DPI. Thus, merging the resistant and susceptible time course experiments yielded 8 total data points per gene (167,376 total data points). Each time point contained a normalized expression value for each gene in resistant and susceptible.

Emwa1 Disease Factors

To identify genes that had gene expression patterns unique to a resistant or susceptible interaction, we measured the difference between gene expression values of resistant and susceptible data sets. We did this at each time point to generate 4 unique difference values. If expression values for any gene fell below 1.5 FC for all data points, that gene was removed from the analysis. From this analysis we removed 9,936 genes.

Resistant Specific, Susceptible Specific, and Shared Clusters

We then looked within the Emwa1 disease factors for genes with significant expression or repression during infection. After this analysis we were left with 2,237 genes that met this cutoff in resistant only; 1,317 genes that met this cutoff in susceptible only; and 2,054 genes that met this criteria in both interactions. We named these clusters Resistant Specific Clusters (RCs), Susceptible Specific Cluster (SCs), and Shared Cluster (RSCs).

Activation and Repression

In the next step we clustered genes in the RC, SC, and RSC by transcriptional activation or repression, specific to their respective data sets. Genes that fell above 0.3 normalized expression FC at all time point were considered activated and those that fell below -0.3 normalized expression were considered repressed. The clusters were referred to as follows: Resistance activated Cluster (RaC), genes that met the above criteria specifically for activation in the resistant data set; Resistance repressed Cluster (RrC), genes that met the above criteria for repression in the resistant data set. Susceptible activated Cluster (SaC), genes that met the above criteria for activation in the susceptible data set; Susceptible repressed Cluster (SrC), genes that met the above criteria for repression the susceptible data set. Shared Upregulated Clusters (RSUC), genes that met the above criteria for activation both the resistant and the susceptible data sets. Shared repressed Cluster (RSrC), genes that met the above criteria for repression in both the resistant and the susceptible data sets.

Initial Time-Point Clustering of RaC, RrC, SaC, SrC

We defined whether a gene was activated or repressed specific to a time-point using ½MAX for activated clusters or ½Min for repressed clusters. RCs were analyzed using values from the resistant data set and SC were analyzed using values from the susceptible data set. Each time point was compared against its activation or repression parameter. We sorted genes by time points where meaningful activation or repression was occurring. From this analysis we clustered genes in the RaC, RrC, SaC, and SrC into initial time point clusters (Table 4.4). These were referred to as iRaC0.5, iRaC2, iRaC4, iRaC6, iRrC0.5, iRrC2, iRrC4, iRrC6, iSaC0.5, iSaC2, iSaC4, iSaC6, iSrC0.5, iSrC2, iSrC4, and iSrC6.

Unique Time-Point Clusters (fTPC) of RaC, RrC, SaC, SrC

Finalized clustering was produced by taking the analysis done in initial activation and applying sorting to include activation at all time points. Within each initial time point cluster, we clustered genes based on activation or a lack of activation at the remaining time points. This analysis produces a maximum of 16 unique expression clusters for each of the following clusters RaC, RrC, SaC, and SrC. 1,1,1,1 = 0.5a // 1,1,1,0 = 0.5b //… // 1,0,0,0 =0.5h//0,1,1,1 = 2a// etc

Comparing Up and Downregulation in Finalized R and S specific Clusters to Determine Major Resistance and Susceptibility Clusters

Comparable activated and repressed final time point clusters (based on sorting pattern) were analyzed against each other. We used the absolute value of repressed clusters to compare how similar activated and repressed cluster moved with regards to expression in regards to both activation pattern and amplitude. We grouped activated and

repressed clusters moving with near identical patterns. We refer to these paired clusters as Major Resistance or Susceptibility Clusters (MRCs, MSCs).

Resistant and Susceptible Specific Differential Activation Clustering of RSCs

Genes that fell into the shared expression clusters (RSCs) showed significant expression in both the resistant and susceptible data sets. Genes in Shared Clusters were unable to be sorted effectively by co-expression or co-activation. We chose to analyze differential activation in order to cluster these genes into biologically meaningful groups. We used the activation or repression parameters ½MAX and ½MIN from all data points for a particular gene to determine time point specific activation in both data sets. We analyzed each time point for genes that were not activated at either time point, activated in R versus S, activated in S versus R, or activated in both. This analysis produced 6 clusters: Shared activated Cluster specific to R (RS-RaC); Shared activated Clusters specific to S (RS-SaC); Shared Upregulated Cluster with no differential activation (RS-NdaC); Shared repressed Cluster specific to R (RS-RrC); Shared repressed Clusters specific to S (RS-SrC); Shared Repressed Cluster with no differential repression (RS-NdrC). Genes which were not activated in both data sets were removed from the analysis.

Differential iTPC and uTPC of Shared Clusters (RSCs)

We looked used the differential between the binary result from comparison with the activation/repression parameter. Sorting was done at a time point specific level to generate differential iTPC (iRSTPC). We named these clusters RS-RaC0.5, RS-RaC2, RS-RaC4, RS-RaC6, RS-SaC0.5, RS-SaC2, RS-SaC4, RS-SaC6, RS-RrrC0.5, RS-RrC2, RS-RrC4, RS-RrC6, RS-SrC0.5, RS-SrC2, RS-SrC4, and RS-SrC6 (Table 4.5). We further divide these clusters to determine unique expression profiles between the resistant

and susceptible data set. We utilized sequential sorting to sort R versus S specific activation with respects to all time points.

Analyzing Clusters using an Annotation Based Approach

We wanted to determine whether unique gene clusters shared similarities with regards to biological process, molecular function, or cellular compartmentalization. We used panther to annotate gene clusters from multiple stages during the analysis to highlight both broad and specific factors involved in resistance or susceptibility. For global analysis we counted similar annotations. We utilized annotations that were highly represented across our clusters to describe the *Hpa* infection from a functional standpoint. We applied a more specific method to annotate specific categories of interest. In this analysis, we generated all known sub-annotations of genes specific to the process or family. We chose to investigate defense that may be involved in resistance or susceptibility based on their specific activation or repression profiles. We combined like annotations to generate a complete list of *A. thaliana* LRR genes. We identified top defense gene candidates from several expression clusters. We selected LRR knockouts based on maximum expression among clusters of interest. Our predominant focus was on genes which fell into shared expression clusters to confirm the importance of these clusters in resistance or susceptibility of *A. thaliana* to *Hpa* Emwa1.

Defense Gene Analysis: Emwa1 Innoculation

We obtained LRR KO mutants from the SALK Institute. Seed from these knockout mutants was planted and bulked prior to inoculations. *Hpa* Emwa1 was inoculated on 11 day old seedlings. Seedlings were infected with a spore suspension of 50,000 spores per mL. 6 days post inoculation, Seedlings were placed under 100%

humidity for 24 hours. The number of conidiophore were counted on mutant and control plants Col-0, rpp4Col-0, EDS1-1, and WS. We determined susceptibility or resistance of KO mutants in comparison to Col-0 CS6000 (The background for all mutant lines). We used rpp4Col-0 as a reference for susceptibility to judge the strength of our phenotypes. We used 3 biological replicates to validate selected mutants. Accessions which produced similar levels of susceptibility to *rpp4*Col-0 were considered heavily susceptible. Although we consider these heavy susceptibility phenotypes it is important to note that much more extreme susceptibility phenotypes were observed in *Hpa* infections of *A. thaliana* cv. WS and EDS1-1.

Figures and Tables

Table 4.1 General Information for the resistance dataset Emwa1 x Col-0, and susceptibility dataset Emwa1 x *rpp4*Col-0. A 1.5 fold change cutoff was implemented to remove genes that were not significantly activated or repressed in response to pathogen infection. Nearly 6,000 *A. thaliana* genes were significantly regulated during a compatible and/or incompatible interaction with *Hpa* Emwa1.

	Emwa1xCol-0 R	**Emwa1xrpp4Col-0 S**
Type of Data Collection	Microarray (Affymetrix)	Microarray (Affymetrix)
# of Genes from MA Chip	20,922	20,922
# of Genes Removed (Q value)	315	345
# of Genes Significantly Changed by Emwa1 Infection	4291	3371
# of Genes Specific to R or S	2237	1317
# of Genes Shared	2054	2054

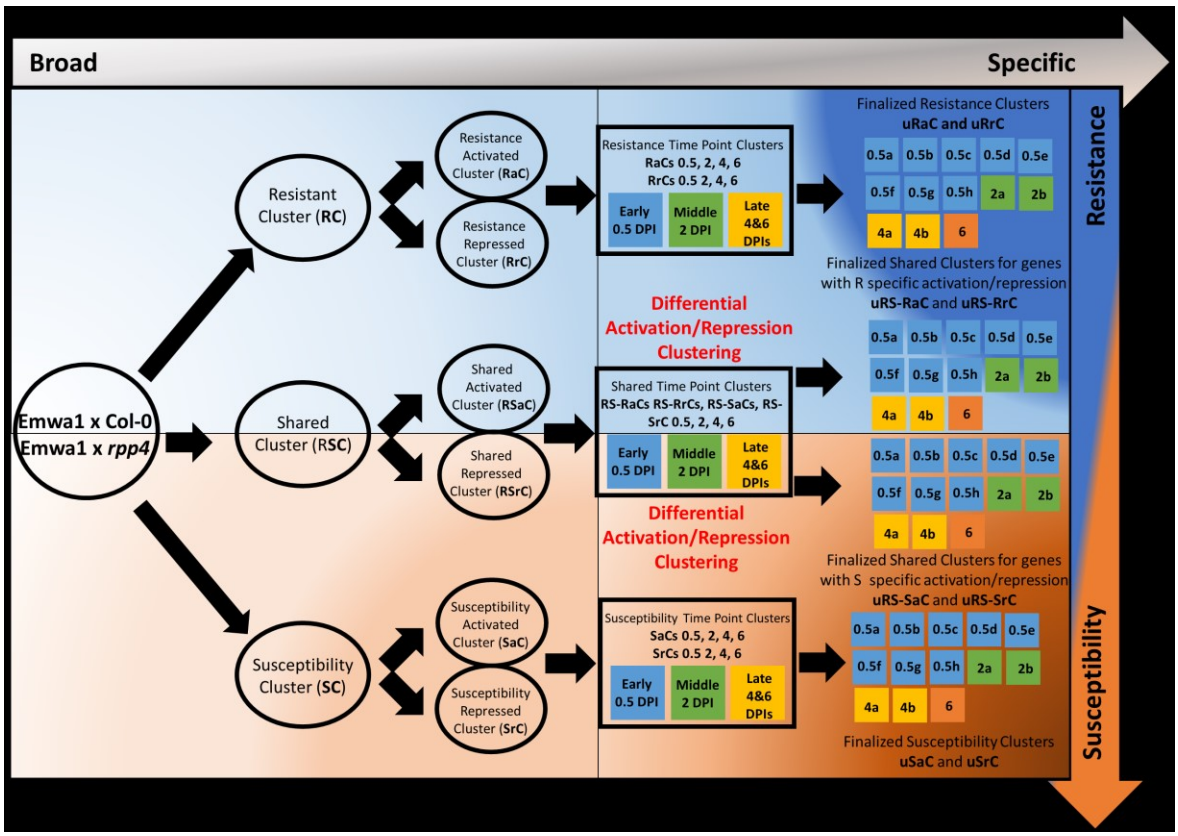

Figure 4.1 Schematic describing differential clustering of complementary *Hpa* infection time course experiments. We analyzed clusters after each transformation to describe the infection from both broad and specific perspectives. We used expression data from both compatible and incompatible interactions to determine whether genes were involved in resistance (R) and susceptibility (S) mechanisms. We determined that a large group of genes with shared expression between R and S, denoted RS. Additionally genes were separated based on transcriptional upregulation versus downregulation expression profiles. We used a 0.3 normalized expression as a cutoff to determine whether a gene was significantly up or downregulated in response to infection. We utilized a term we define as the activation parameter to describe significant expression on an individual gene basis. This term was used to sort for significant gene expression at each time point

generating initial R and S time point clusters by determining at which time point a gene was first activated in response to pathogen attack. Genes with shared expression had variable expression profiles that could not be clustered using R or S data sets individually. Differential gene activation between R and S was used to determine the expression genes with shared expression profiles. Using this approach initial RS time point clusters (iTPC) specific to R or S were generated. Following initial time point clustering R, S, and RS were sequentially sorted using activation to describe gene expression patterns. From this we generated unique R, S, and RS clusters that contained between 25-500 genes.

Table 4.2 Resistant Specific, Susceptible Specific, and Shared Clusters were furthered grouped by Activation or Repression. Genes were grouped into the both (RarC, SarC, RS-arC) category if transcript levels were significantly activated and repressed during the time course.

Activation or Repression	# of Genes
Resistance Activated Cluster (RaC)	901
Resistance Repressed Cluster (RrC)	1317
Resistance Activated and Repressed Cluster (RarC)	19
Susceptibility Activated Cluster (SaC)	538
Susceptibility Repressed Cluster (SrC)	715
Susceptibility Activated and Repressed Cluster (RS-arC)	64
Shared Activated Cluster (RS-aC)	753
Shared Repressed Cluster (RS-rC)	1170
Shared Activated and Repressed Cluster (RS-arC)	448

Table 4.3 *A. thaliana* Genes from shared clusters had significant expression during compatible and incompatible interactions with *Hpa*. We determined whether genes had resistant or susceptible specific time point activation using a differential time point analysis. If a gene was significantly up or downregulated at the same time point in R and S, it was placed in the RS-NDaC or the RS-NDrC. Although genes could fall into multiple clusters during this step, the majority of genes exhibited time point specific activation for either R or S when using differential activation.

Differential Clustering of Shared (RS) Genes	**Total # of Genes**
Shared-Resistance Activated Cluster (RS-RaC)	236
Shared-Resistance Repressed Cluster (RS-RrC)	604
Shared-Susceptibility Activated Cluster (RS-SaC)	214
Shared-Susceptibility Repressed Cluster (RS-SrC)	131

Table 4.4 The activation/repression parameter was used to sort genes from the RaC, RrC, SaC, and SrC into clusters that reflect the timing of gene expression in response to *Hpa*. After binary sorting, normalized expression values specific to R or S were used to determine average expression and deviation among clusters. If a gene was implicated during multiple time points, it was placed into the time point cluster in which it was first activated to generate initial time point clusters.

R and S Time Point Clusters	Total # of Genes	RS Time Point Clusters	Total # of Genes
iRaC0.5	355	iRS-RaC0.5	29
iRaC2	525	iRS-RaC2	135
iRaC4	18	iRS-RaC4	21
iRaC6	3	iRS-RaC6	51
iRrC0.5	695	iRS-RrC0.5	91
iRrC2	617	iRS-RrC2	512
iRrC4	4	iRS-RrC4	0
iRrC6	1	iRS-RrC6	0
iSaC0.5	258	iRS-SaC0.5	57
iSaC2	25	iRS-SaC2	44
iSaC4	197	iRS-SaC4	88
iSaC6	58	iRS-SaC6	25
iSrC0.5	343	iRS-SrC0.5	91
iSrC2	9	iRS-SrC2	3
iSrC4	191	iRS-SrC4	17
iSrC6	172	iRS-SrC6	20

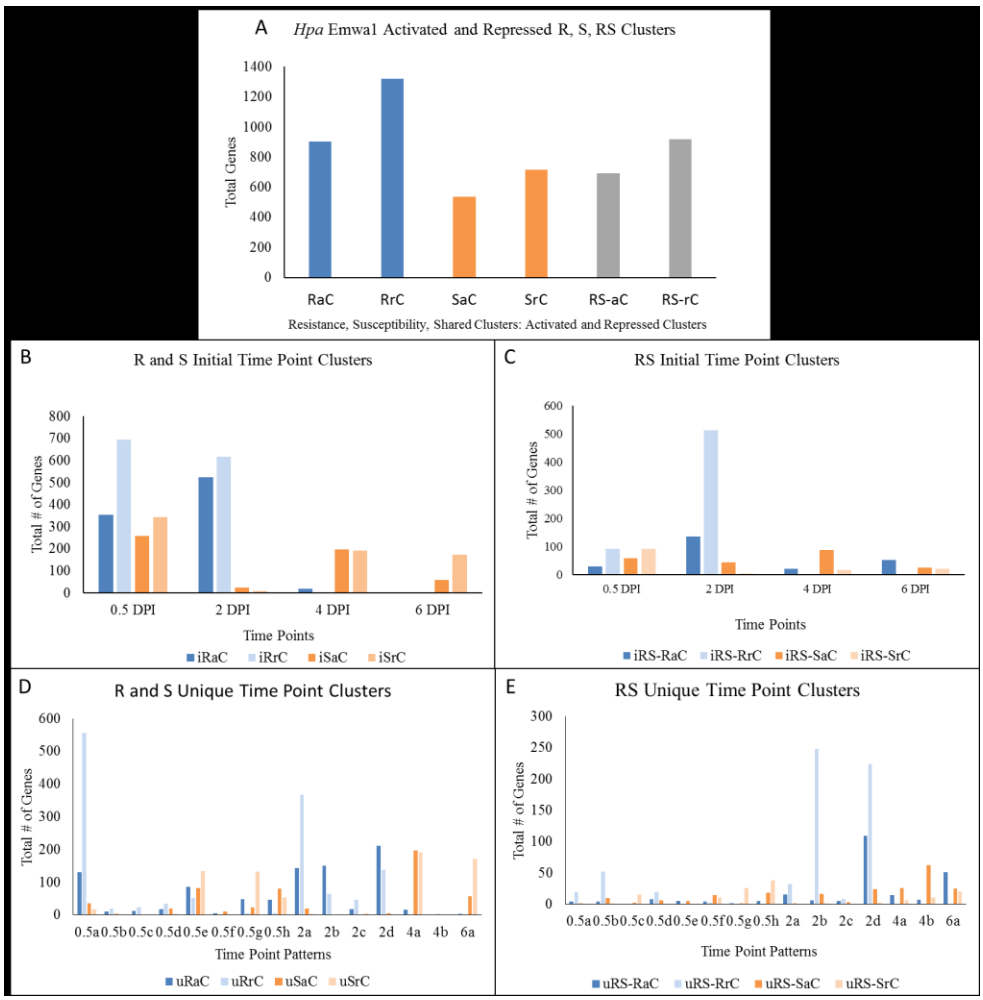

Figure 4.2 A-E Total gene analyses were performed during multiple stages of this analysis. Panel **A** shows the distribution of genes clusters with R specific, S specific, or RS expression profiles. Panel **B** shows the distribution of *Hpa* Emwa1 specific gene clusters (R or S), while panel **C** shows the distribution of shared (RS) gene clusters in initial time point clusters. The activation parameter was used to further distribute genes from initial time point clusters into unique resistance or susceptibility clusters by determining significant activation/repression patterns during infection **(D)**. Differential sorting was used to cluster shared RS clusters into unique R or S specific time point patterns **(E)**.

Table 4.5 Unique resistant and susceptibility clusters were identified using sequential sorting on activated resistant or susceptible specific genes. All unique expression patterns are shown. 32 possible clusters could be achieved for both R and S. Average standard error is reported among genes in finalized R and S clusters with >25 genes.

Unique Time Point Clusters	Total # of Genes	Average Standard Error
uRaC0.5a	130	0.025
uRaC0.5b	10	-
uRaC0.5c	12	-
uRaC0.5d	18	-
uRaC0.5e	86	0.016
uRaC0.5f	5	-
uRaC0.5g	48	0.017
uRaC0.5h	46	0.025
uRaC2a	144	0.012
uRaC2b	151	0.005
uRaC2c	18	-
uRaC2d	212	0.005
uRaC4a	16	-
uRaC4b	2	-
uRaC6a	3	-
uRrC0.5a	555	0.014
uRrC0.5b	20	-
uRrC0.5c	24	-
uRrC0.5d	33	0.021
uRrC0.5e	51	0.017
uRrC0.5f	0	-
uRrC0.5g	6	-
uRrC0.5h	6	-
uRrC2a	368	0.009
uRrC2b	65	0.017
uRrC2c	47	0.021
uRrC2d	137	0.011
uRrC4a	0	-
uRrC4b	4	-
uRrC6a	1	-
uSaC0.5a	36	0.030
uSaC0.5b	4	-
uSaC0.5c	2	-
uSaC0.5d	19	-
uSaC0.5e	82	0.024

uSaC0.5f	11	-
uSac0.5g	24	-
uSac0.5h	80	0.015
uSac2a	19	-
uSaC2b	1	-
uSaC2c	0	-
uSaC2d	5	-
uSaC4a	197	0.015
uSaC4b	0	-
uSaC6a	58	0.021
uSrC0.5a	17	-
uSrC0.5b	1	-
uSrC0.5c	1	-
uSrC0.5d	3	-
uSrC0.5e	134	0.018
uSrC0.5f	1	-
uSrC0.5g	132	0.019
uSrC0.5h	54	0.028
uSrC2a	4	-
uSrC2b	0	-
uSrC2c	5	-
uSrC2d	0	-
uSrC4a	191	0.013
uSrC4b	0	-
uSrC6a	172	0.013

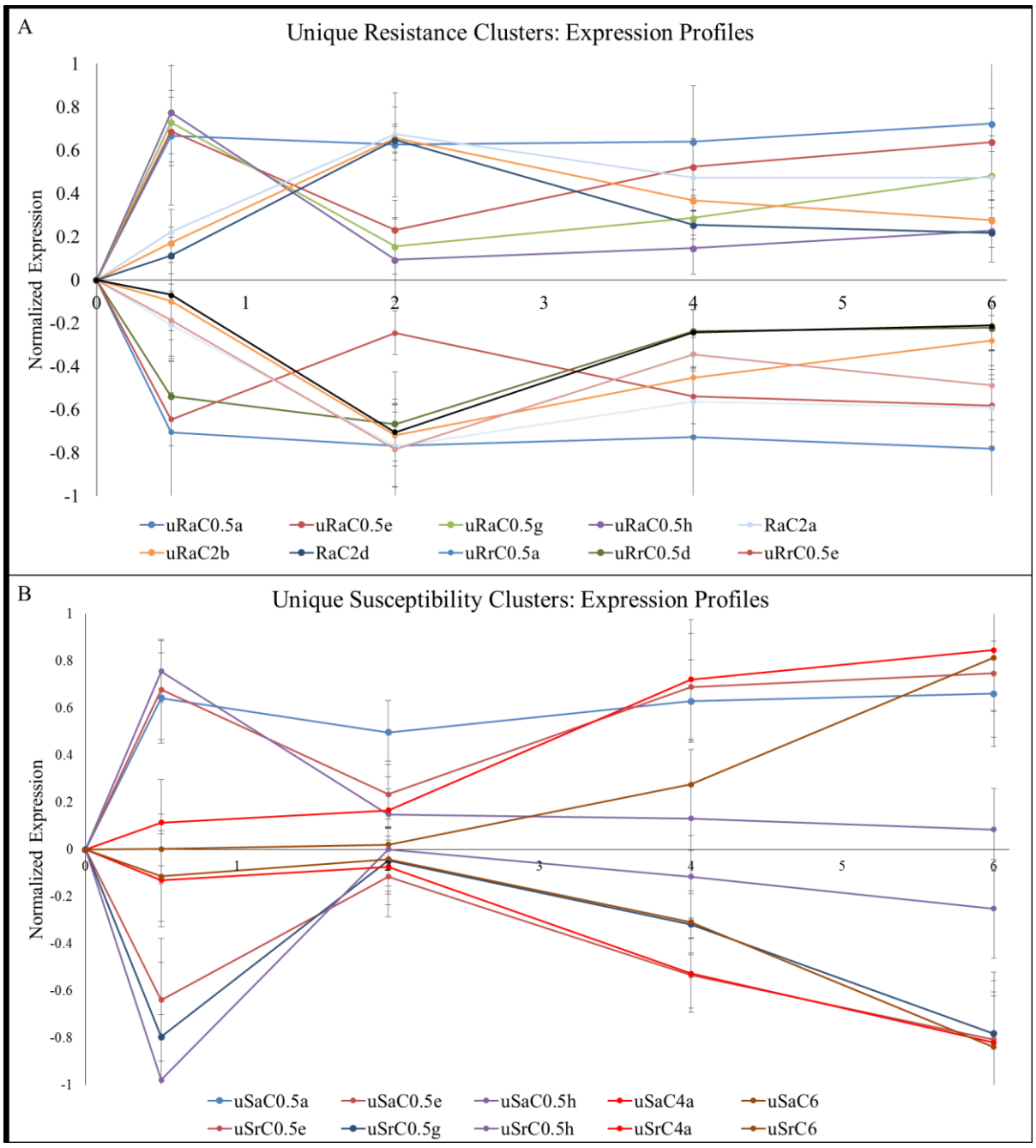

Figure 4.3 A, B Average normalized expression was plotted for finalized resistance **(A)** and susceptibility **(B)** specific gene clusters. Error bars show standard deviation across unique expression clusters. Clusters identified using activation sorting showed expression at one or multiple time points during infection.

Table 4.6 Genes with shared expression profiles were sequentially sorted to determine differential time point activation. Significant gene expression at every time point in R was compared to every time point in S to generate finalized expression clusters.

Unique Shared Time Point Clusters	Total # of Genes	Average Standard Error
uRS-RaC0.5a	4	-
uRS-RaC0.5b	4	-
uRS-RaC0.5c	0	-
uRS-RaC0.5d	8	-
uRS-RaC0.5e	5	-
uRS-RaC0.5f	4	-
uRS-RaC0.5g	1	-
uRS-RaC0.5h	5	-
uRS-RaC2a	15	-
uRS-RaC2b	6	-
uRS-RaC2c	5	-
uRS-RaC2d	109	0.024
uRS-RaC4a	14	-
uRS-RaC4b	7	-
uRS-RaC6a	51	0.036
uRS-RrC0.5a	19	-
uRS-RrC0.5b	52	0.045
uRS-RrC0.5c	0	-
uRS-RrC0.5d	19	-
uRS-RrC0.5e	0	-
uRS-RrC0.5f	2	-
uRS-RrC0.5g	0	-
uRS-RrC0.5h	0	-
uRS-RrC2a	32	0.046
uRS-RrC2b	248	0.015
uRS-RrC2c	8	-
uRS-RrC2d	224	0.015
uRS-RrC4a	0	-
uRS-RrC4b	0	-
uRS-RrC6a	0	-
uRS-SaC0.5a	1	-
uRS-SaC0.5b	10	-
uRS-SaC0.5c	2	-
uRS-SaC0.5d	6	-
uRS-SaC0.5e	5	-
uRS-SaC0.5f	14	-
uRS-SaC0.5g	1	-

uRS-SaC0.5h	18	-
uRS-SaC2a	1	-
uRS-SaC2b	16	-
uRS-SaC2c	3	-
uRS-SaC2d	24	-
uRS-SaC4a	26	0.055
uRS-SaC4b	62	0.039
uRS-SaC6a	25	0.063
uRS-SrC0.5a	0	-
uRS-SrC0.5b	1	-
uRS-SrC0.5c	15	-
uRS-SrC0.5d	0	-
uRS-SrC0.5e	0	-
uRS-SrC0.5f	11	-
uRS-SrC0.5g	26	0.072
uRS-SrC0.5h	38	0.056
uRS-SrC2a	0	-
uRS-SrC2b	1	-
uRS-SrC2c	0	-
uRS-SrC2d	2	-
uRS-SrC4a	6	-
uRS-SrC4b	11	-
uRS-SrC6a	20	-
RS-NdaC	178	0.030
uRS-SaC[2] RaC[6]	32	0.035
RS-NdrC	30	0.056
RS-RrC[2] SrC[4,6]	39	0.022
RS-RrC[2] SrC[4]	28	0.032
RS-RrC[2] SrC[6]	89	0.019
RS-SrC[0.5,4,6] RrC[2]	33	0.033
RS-SrC[0.5,6] RrC[2]	39	0.029
RS-SrC[0.5] rC[2]	49	0.032

Table 4.7 Major resistance and susceptibility clusters were determined by comparing clusters with identical activation/repression patterns. The absolute value of gene expression to determine that levels of transcription were equivalent between up and downregulated genes within certain pairs. In all cases, standard error and standard deviation improved upon combination.

Major Resistance and Susceptibility Clusters	Total # of Genes	Average Standard Error Major Resistance and Susceptibility Clusters
MRC0.5a0	378	0.014
MRC0.5a1	307	0.021
MRC0.5e	137	0.012
MRC2a	512	0.007
MRC2b	216	0.006
MRC2d	349	0.005
MSC0.5e	216	0.014
MSC0.5h	134	0.013
MSC4a	388	0.009
MSC6a	230	0.009

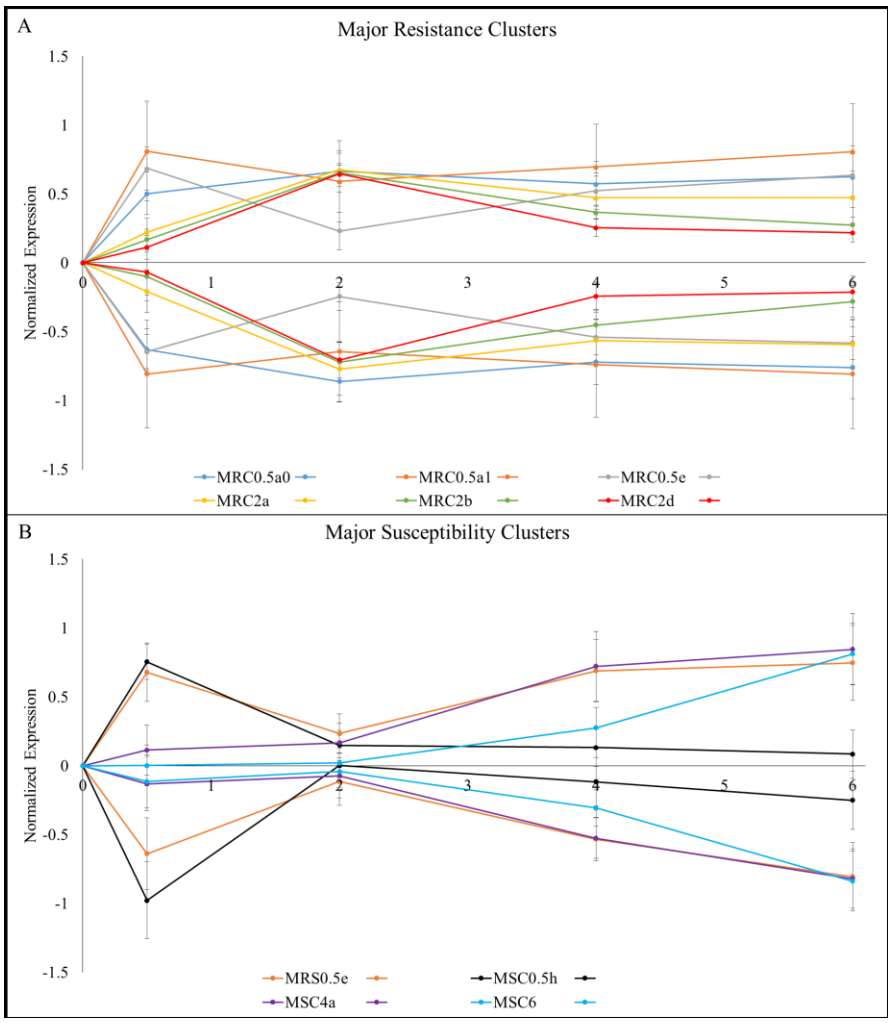

Figure 4.4 (A, B) Major resistance and susceptibility clusters were determined by pairwise comparison of uniquely upregulated and downregulated gene clusters with nearly identical expression. The expression profile of the 6 observed major resistant clusters are shown in **A**, while the 4 observed major susceptibility clusters are shown in **B**. The error bars represent standard deviation among genes in major clusters.

Table 4.8 Unique clusters were functionally annotated for cellular component (CC), biological process (BP), and molecular function (MF). Annotations from these categories were counted for each gene cluster. The total number of annotated genes in CC, BP, and MF are indicated in column 2 as X, X, X respectively. The amount of annotated genes were compared to the total number of annotated genes in the cluster to generate percent representation. The top 3-4 most enriched categories are shown for each expression cluster.

Unique Clusters	# of Annotated Genes (CC, BP, MF)	Cellular Component	Biological Process	Molecular Function
uRaC0.5a	81, 57, 58	Membrane (28%) Nucleus (22%)	Signal Transduction (9%) Protein Phosphorylation (7%)	Protein Binding (14%) Hydrolase (7%)
uRaC0.5e	46, 33, 35	Nucleus (22%) Membrane (13%)	Oxidation-Reduction (9%)	Protein Binding (9%) Metal Ion Binding (9%)
uRaC0.5g	31, 20, 22	Cytoplasm (16%) Nucleus (16%) Chloroplast (13%)	-	-
uRaC0.5h	30, 31, 19	Cytoplasm (20%) Nucleus (17%) Chloroplast (13%)	-	-
uRaC2a	85, 52, 56	Nucleus (26%) Membrane	Protein Phosphorylation (8%)	Heme Binding (9%) Polysaccharide

		(25%) Cytoplasm (13%)	Oxidation- Reduction (8%) Embryo Development -Seed Dormancy (6%)	Binding (5%) ADP Binding (5%)
uRaC2b	82, 44, 42	Nucleus (38%) Cytoplasm (15%) Membrane (13%) Mitochondrion (10%)	Oxidation- Reduction (11%) Regulation of Transcription (7%)	Kinase (10%) Transcription Factor (10%) Zinc Ion Binding (7%)
uRaC2d	118, 80, 81	Nucleus (31%) Cytoplasm (15%) Mitochondrion (13%) Membrane (13%) Extracellular Region (11%)	Oxidation- Reduction (14%) Regulation of Transcription (11%) Cell Wall Organization (5%)	Metal Ion Binding (10%) Transcription Factor (9%)
uRrC0.5a	351, 247, 249	Membrane (24%) Nucleus (16%) Cytoplasm (13%)	Oxidation- Reduction (7%) Regulation of Transcription (4%) Response to Karrakin (3%)	Metal Ion Binding (4%) Zinc Ion Binding (4%) Protein Binding (4%)
uRrC0.5d	23, 18, 18	Membrane (26%) Chloroplast (22%) Plasmodesmata (17%)	-	-
uRrC0.5e	26, 26, 27	Nucleus (27%)	-	-
uRrC2a	223, 161, 160	Membrane (16%)	Translation (9%)	Structural Constituent of

		Cytoplasm (17%) Chloroplast (17%)	Ribosome Biogenesis (6%) Proteolysis (4%) Methylation (4%)	Ribosome (11%) Protein Binding (9%) Metal Ion Binding (9%) Methyltransferase (5%)
uRrC2b	50, 37, 36	Chloroplast (18%) Membrane (18%) Nucleus (16%)	-	-
uRrC2c	29, 25, 25	Chloroplast (31%) Cytoplasm (24%)	-	Protein Binding (12%)
uRrC2d	88, 71, 72	Cytoplasm (19%) Chloroplast (15%) Vacuole (12%) Mitochondrial Respiratory Chain Complex I (6%)	Oxidation-Reduction (7%) Response to Cadmium (7%) ATP Hydrolysis coupled to Proton Transport (6%)	ATP Binding/ATPase (10%) Protein Binding (8%) Zinc Ion Binding (8%) Metal Ion Binding (7%)
uSaC0.5a	20, 11, 12	Membrane (40%)	-	-
uSaC0.5e	52, 38, 40	Membrane (15%) Cytoplasm (11%) Nucleus (10%) Golgi Apparatus (8%)	Oxidation-Reduction (21%)	-
uSac0.5h	60, 41, 39	Membrane (23%) Nucleus (15%) Cytoplasm (12%)	-	Protein Binding (10%) Metal Ion Binding (8%)
uSaC4a	112, 91, 92	Membrane (31%) Nucleus	Oxidation-Reduction (5%)	Metal Ion Binding (9%) Protein Binding

		(15%) Cytoplasm (13%)	Protein Phosphorylation (4%)	(8%) Kinase Activity (5%) Zinc Ion Binding (4%)
uSaC6a	38, 34, 34	Nucleus (29%) Membrane (27%)	Signal Transduction (12%) Oxidation- Reduction (9%)	Nutrient Reservoir (9%) Hydrolase (9%)
uSrC0.5e	93, 84, 82	Membrane (25%) Chloroplast (19%) Nucleus (14%)	Metabolic Process (6%) Regulation of Transcription (6%) Transmembrane Receptor Protein Tyrosine Kinase Signaling Pathway (6%)	Metal Ion Binding (9%) Protein Binding (6%) Transcription Factor (5%) ATP Binding (5%)
uSrC0.5g	76, 56, 53	Chloroplast (30%) Membrane (12%) Nucleus (11%)	Oxidation- Reduction (7%) Defense Response (5%)	Protein Binding (8%) Transcription Factor (6%) Hydrolase (6%)
uSrC0.5h	31, 16, 15	Chloroplast (32%) Nucleus (29%) Mitochondrion (13%)	-	-
uSrC4a	124, 90, 99	Chloroplast (23%) Membrane (19%) Cytoplasm (13%) Nucleus (10%)	Oxidation- Reduction (6%) Regulation of Transcription (4%)	Kinase Activity (8%) Protein Binding (6%) Metal Ion Binding (6%) O-acetyl Transferase (4%)
uSrC6	113, 83, 85	Chloroplast (35%) Cytosolic Large Ribosomal Subunit (8)	Ribosome Biogenesis (17%) Translation (10%) Peptidyl-proline	Structural Constituent of Ribosome (21%) Protein Binding (7%)

			Modification (4%)	Metal Ion Binding (5%)
uRS-RaC2d	86, 73, 69	Membrane (35%) Cytoplasm (20%) Nucleus (14%)	Phosphorylation (16%) Signal Transduction (7%) Defense Response to Fungus (5%)	Kinase (19%) Protein Binding (10%) ATP Binding (9%) Heme Binding (6%) Metal Ion Binding (6%)
uRS-RaC6a	46, 28, 30	Nucleus (24%) Membrane (22%) Chloroplast (9%)	Defense Response to Fungus (11%)	Protein Binding (13%)
uRS-RaC2d	69, 56, 54	Membrane (42%) Cytoplasm (19%) Nucleus (13%)	Phosphorylation (20%) Defense Response (18%) Signal Transduction (13%)	Kinase (20%) Protein Binding (11%) ATP Binding (9%)
uRaC6	36, 21, 24	Nucleus (28%) Membrane (19%)	-	Protein Binding (9%)
uRS-RrC0.5b	17, 12, 13	-	-	-
uRS-RrC2a	15, 12, 13	-	-	-
uRS-RrC2b	135, 104, 98	Chloroplast (30%) Cytosolic Large and Small Ribosome (14%)	Translation (18%) Ribosome Biogenesis (11%)	Structural Constituent of Ribosome (31%) Protein Binding (8%)
uRS-RrC2d	105, 93, 89	Chloroplast (24%) Membrane (22%)	Oxidation-Reduction (6%) Translation (4%)	Structural Constituent of Ribosome (11%) Metal Ion Binding (7%) Oxidoreducatase (7%)
uRS-SaC4a	20, 15, 16	-	-	-

uRS-SaC4b	46, 36, 35	Membrane (30%)	-	-
uRS-SaC6	16, 11, 10	-	-	-
uRS-SrC0.5g	13, 8, 8	Chloroplast (38%)	-	-
uRS-SrC0.5h	14, 13, 11	Chloroplast (36%) Nucleus (36%)	-	-
RS-NdaC	110, 88, 84	Membrane (30%) Nucleus (15%)	Defense/Response to Bacterium (11%) Oxidation-Reduction (7%) Response to Karrakin (5%)	Protein Binding (11%) Kinase Activity (10%) Calcium Ion Binding (6%) ATP Binding (5%)
uRS-SaC[2] RaC[6]	9, 5, 4	-	-	-
RS-NdrC	9, 6, 7	-	-	-
RS-RrC[2] SrC[4,6]	17, 11, 13	Chloroplast (24%)	-	-
RS-RrC[2] SrC[4]	18, 14, 16	Chloroplast (61%)		
RS-RrC[2] SrC[6]	40, 31, 35	Chloroplast (63%)	-	rRNA/RNA Binding (18%) Protein Binding (14%)
RS-SrC[0.5,4,6] RrC[2]	14, 12, 13	-	-	-
RS-SrC[0.5,6] RrC[2]	20, 12, 11	Chloroplast (65%)	-	-
RS-SrC[0.5] rC[2]	11, 7, 9	-	-	-

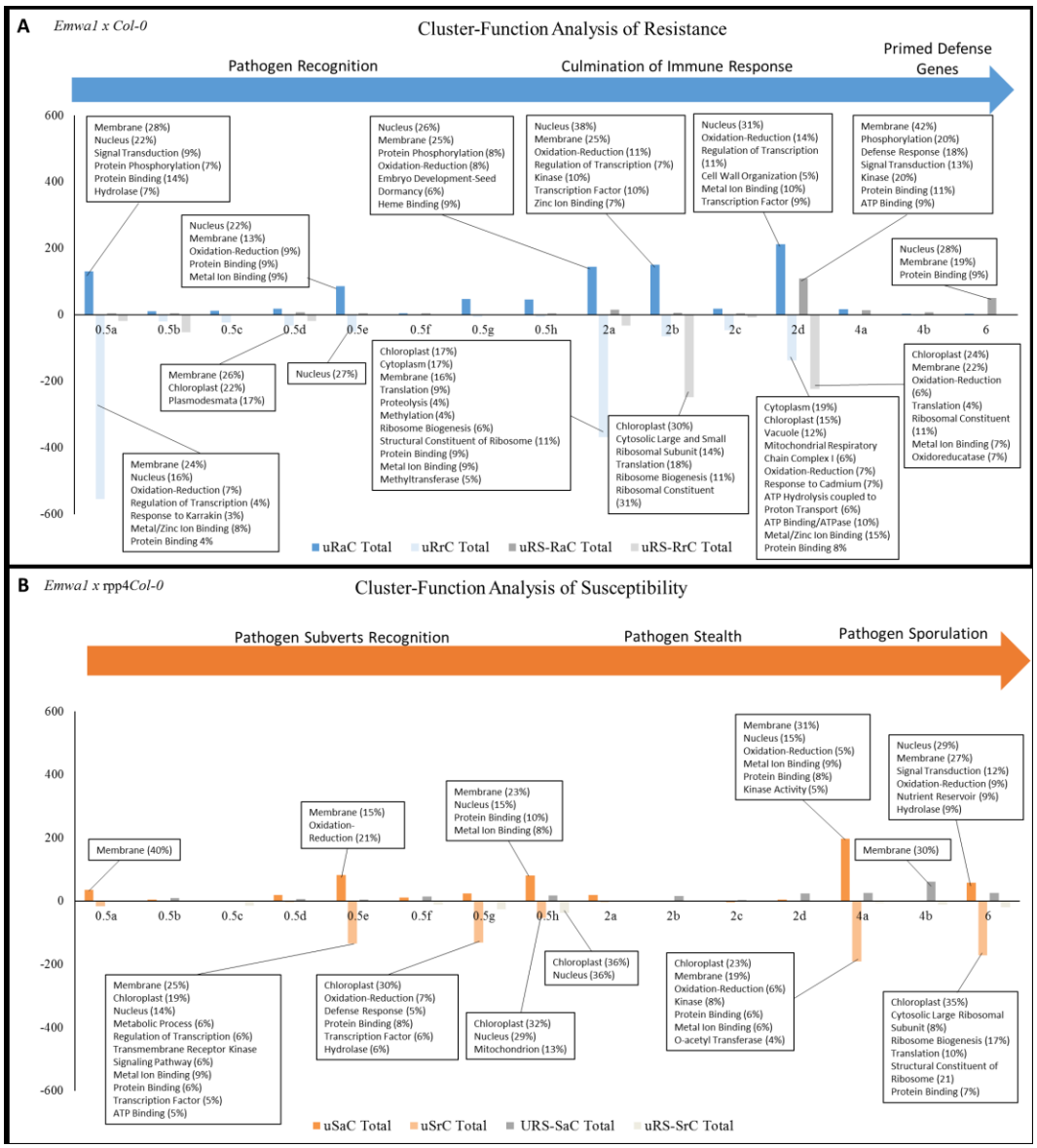

Figure 4.5 (A, B) Resistance **(A)** and susceptibility **(B)** Hpa Infection frameworks. The Y-axis represents the total amount of genes per cluster. We displayed activated clusters as positive values and repressed clusters as negative values. On the X axis, unique clusters are listed sequentially with respects to the time point in which they were first activated. The most significant annotations per clusters are listed above each gene for the following categories; biological process, molecular function, and cellular compartment are listed above the clusters

Table 4.9 Defense Gene Analysis. Accessions containing knockouts in genes implicated in our analysis were obtained from the SALK Institute. When possible, we obtained multiple gene knockouts for a single gene for comparison. Phenotyping was performed by inoculating Emwa1 on all KO mutants in the Col-0 background. Conidiophore counts that were statistically significant to were recorded. Level of susceptibility was determined by comparing conidiophore counts with Col-0. Additionally, we utilized *A. thaliana* cv. WS and *rpp4*Col-0 accessions as benchmarks for susceptibility.

At#	Annotation and Literature Name (when possible)	SALK KO1/KO2	Finalized Cluster(s)	Phenotype Emwa1 x *KOCol-0*
At3g50950	Zar1	CS65962 CS65961	uRS-RaC2d	Strong Susceptible
At1g09970	RLK7	CS351219 CS860338	uRS-SaC4a	Medium Susceptible Strong Susceptible
AT1G63880	RLM2	CS379197 SALK_060981	uRS-RaC0.5a	Weak Susceptible Weak Susceptible
At1g07650	Serine Threonine Kinase	CS847653 SALK_002591	uRS-RaC0.5b	Medium Susceptibility Strong Susceptibility
AT4G16860	RPP4	CS821120 CS349439	RaC2a	Medium Susceptibility
At5g16000	NIK1	SALK_022341 SALK_060808C	RS-arC	Weak Susceptible, Null
At3g14840	LIK1	CS27062 SALK_030855C	uRS-RaC0.5d	No Change
At1g74360	LRR-RLKs LRR@IR protein	CS87681	uRS-RaC2d	No Change
At4g19530	TIR induce cell death	CS25706	uSrC4a	No Change
At1g51800	An Arabidopsis (malectin-like) leucine-rich repeat receptor-like kinase contributes to downy mildew disease.	SALK_137388C	RS-NDaC	No Change

At4g28490	HAE	CS2103477	RS-NDaC	No Change
At1g72300	Tyrosine-sulfated glycopeptide involved in cellular proliferation and expansion in Arabidopsis.	CS876006	uSrC0.5a	No Change
At4g03390	SRF3 incompatible with RPP1 locus EARLY IMMUNITY	CS2103289	uRrC0.5a	Medium Susceptible
At1g53730	SRF6	SALK_054337C	uSrC0.5a	No Change

Chapter 5

Conclusions and Future Direction

Michael Gerald Fedkenheuer[1]

[1]Department of Plant Pathology, Physiology, and Weed Science, Virginia Polytechnic Institute, Blacksburg, VA

Author Contributions: MF wrote this chapter.

Conclusions and Future Direction

Disease pressures are a major concern surrounding sustainable agriculture in monoculture agronomic systems. Static crop populations containing little to no genetic diversity are vulnerable to dynamic pathogen populations. Pathogen populations can adapt at the molecular level to circumvent resistance genes and chemical controls. There is a need to acquire new resistance traits to replace ephemeral control mechanisms and to increase genetic diversity in crop plants.

In Chapter 1 we describe the utility of wild species of cultivated crops as a means of acquiring and introgressing new disease resistance traits. These processes have been accelerated by new technologies such as SNP arrays and NGS sequencing. Molecular markers, generated by these techniques, can be used to assist breeding. This process is known as marker-assisted selection (MAS). MAS can be used to determine inheritance patterns of genetic elements in segregating populations. Resistance traits are commonly mapped using progeny from crosses between resistant and susceptible cultivars. The advent of new sequencing technologies has allowed for the development of numerous genetic markers. This allows for more efficient and more accurate genetic mapping of resistance loci.

Phytophthora sojae (*P. sojae*), the causal agent of *Phytophthora* root and stem rot disease, reduces soybean harvests worldwide. In Chapter 2, we discuss the development of a system to screen soybean germplasm for *Resistance to Phytophthora sojae* (*Rps*) genes that recognize core *P. sojae* RXLR effectors. This screen leverages cell death induced by ETI as an indicator of *Avr/R* interactions. We used the bacterium

Pseudomonas to deliver individual *P. sojae* effectors into soybean by Type III secretion. *P. sojae* core effectors *Avh*16, *Avh*180, and *Avh*240 were screened on disease-resistant *Glycine max* (*G. max*) germplasm. *G. max* accessions that produced a HR in response to core effector(s) were selected for crosses with *P. sojae*-susceptible cultivar Williams. We observed a simple, genetically dominant inheritance pattern that is consistent with gene-for-gene resistance in segregating $F_{2:3}$ progeny. We expect *R* genes identified by these screens to be durable because *P. sojae* requires core effectors for virulence.

A high throughput *P. sojae* infection assay was optimized using *P. sojae* race 2 on *Rps*1k as a resistant control and on susceptible Williams and Harosoy as susceptibility controls. This assay will be used in combination with effector-based screening in recombinant inbred (RI) lines to map resistance loci. We expect that the mapping will be in agreement between effector-based screening and *P. sojae* infection assays. We expect mapped *R* gene candidates to recognize the core effectors identified from screening. Resistance interactions can be tested using heterologous expression in *Nicotiana benthamiana*. Interactions can also be tested through gene knockouts in either pathogen core effectors or plant *R* genes. We have begun working with a modified strain of *P. sojae* race 2 containing a cassette for silencing of *Avh*180 to test whether resistance or susceptibility is impacted by loss of the *Avr/R* interaction. Conversely, we are preparing to use soybean cotyledon silencing assays to knock down soybean *R* gene candidates.

In Chapter 3, we utilized effector-based screening to probe *Glycine soja* (*G. soja*) germplasm with core RXLR effectors from *P. sojae* to identify novel *R* genes. We screened *P. sojae* resistant *G. soja* accessions, identified in a screen with a combination of isolates that overcome all known *Rps* genes in soybean. We challenged these

accessions with 10 effectors and observed responses to 8 core effectors. We developed and utilized the aforementioned high-throughput *P. sojae* infection assay to screen *P. sojae* resistant germplasm with isolates individually rather than in combination. We utilized the three resistance-breaking isolates and found that some germplasm, previously identified as resistant in combinatorial screens, were susceptible to individual isolates. This indicates that complex or overlapping resistance mechanisms can give false positives when screening with multiple pathogen isolates.

We developed segregating populations from crosses of fully resistant germplasm with susceptible *G. max*, to determine inheritance of potential *R* genes in germplasm that responded to core effectors. We observed a clear geographic distribution of effector responses to *Avh*53 and *Avh*137 that was consistent with published subgroups. We analyzed segregating progeny derived from gs2292 which responded to the core effector *Avh*240. We observed a simple, genetically dominant inheritance pattern. We are using MAS to map disease resistance traits in this segregating population. We hope to utilize similar strategies, outlined above, to provide a correlation between disease resistance and *R* gene candidates that were identified in effector-based screens.

Identifying genes involved in resistance and susceptibility can also be achieved through bioinformatics approaches. In Chapter 4, we utilized the plant/pathogen interaction between *Arabidopsis thaliana* and the *Hyaloperonospora arabidopsidis* isolate Emwa1. We developed a framework to analyze more complex or less well studied oomycete/plant pathosystems. We used a publicly available RNA time-course experiment with 0, 0.5, 2, 4, and 6 DPI time-points from both a resistant and a susceptible infection with *Hpa* Emwa1. Susceptibility is conferred by null mutation of a single plant *R* gene,

RPP4, making this an ideal starting point to better understanding gene-for-gene resistance.

We began by clustering genes implicated in resistance and in susceptibility using Bayesian algorithms. However, there are limitations to this method. This led us to develop a novel algorithm by which genes are sorted into time-point specific clusters using a gene specific activation/repression parameter. We used this technique to perform a comprehensive differential analysis to identify resistance and susceptibility mechanisms. Gene ontology annotations were used to identify defense genes with unique expression profiles. *A. thaliana* null mutants were used to identify several defense genes that produced significant susceptibility phenotypes to *Hpa* isolate Emwa1. Among the most significant defense gene mutants were receptor kinases *Zar1* and *RLK7*. In the future we hope to employ this clustering algorithm to identify defense gene against pathogens of crop plants. We are currently working to modify this algorithm for use in diverse scientific disciplines.

References

1. Serge Savary, A.F., Jean-Noel Aubertot, and Clayton Hollier, *Crop Losses Due to Disease and their Implications for Global Food Production Losses and Food Security.* Food Security, 2012. **4**(4): p. 519 - 537.
2. Lashermes, D.F.a.P., *Molecular Techniques in Crop Improvement*. Molecular Tools for Improving Coffee Resistance to Parasites, ed. S.M.J.a.D.S. Brar. 2013: Springer Science and Business Media. 328-331.
3. Gilet, T. and L. Bourouiba, *Rain-induced Ejection of Pathogens from Leaves: Revisiting the Hypothesis of Splash-on-Film using High-speed Visualization.* Integrative and Comparative Biology, 2014. **54**(6): p. 974-984.
4. Coakley, S.M., H. Scherm, and S. Chakraborty, *Climate change and plant disease management.* Annu Rev Phytopathol, 1999. **37**: p. 399-426.
5. Wolfe, M.S., *Crop strength through diversity.* Nature, 2000. **406**(6797): p. 681-682.
6. Zhu, Y., et al., *Genetic diversity and disease control in rice.* Nature, 2000. **406**(6797): p. 718-722.
7. Breseghello, F. and A.S. Coelho, *Traditional and modern plant breeding methods with examples in rice (Oryza sativa L.).* J Agric Food Chem, 2013. **61**(35): p. 8277-86.
8. Olsen, K.M. and J.F. Wendel, *A bountiful harvest: genomic insights into crop domestication phenotypes.* Annual Review of Plant Biology, 2013. **64**: p. 47-70.
9. United States Department of Agriculture (USDA), E.R.S.E. *Crops*. 2016; Available from: http://www.ers.usda.gov/topics/crops/.aspx.
10. Virginia Heffer Link1, M.L.P., and Kenneth B. Johnson, *Oomycetes.* The Plant Health Instructor, 2002 (Updated 2012).
11. M. C. Lerclerc, J.G., and M. Deville, *Taxonomic and phylogenetic analysis of Saprolegniaceae (Oomycetes) inferred from LSU rDNA and ITS sequence comparisons.* Antonie van Leeuwenhoek, 2000. **77**(4): p. 369 - 377.
12. Cooke, D.E., et al., *A molecular phylogeny of Phytophthora and related oomycetes.* Fungal Genet Biol, 2000. **30**(1): p. 17-32.
13. Tyler, B.M., *Entering and breaking: virulence effector proteins of oomycete plant pathogens.* Cell Microbiol, 2009. **11**(1): p. 13-20.
14. Bebber, D.P. and S.J. Gurr, *Crop-destroying fungal and oomycete pathogens challenge food security.* Fungal Genet Biol, 2015. **74**: p. 62-4.
15. Fry, W.E., et al., *Five Reasons to Consider Phytophthora infestans a Reemerging Pathogen.* Phytopathology, 2015. **105**(7): p. 966-81.
16. Haverkort, A.J., et al., *Societal Costs of Late Blight in Potato and Prospects of Durable Resistance Through Cisgenic Modification.* Potato Research, 2008. **51**(1): p. 47-57.
17. Runno-Paurson, E., et al., *High genotypic diversity found among population of Phytophthora infestans collected in Estonia.* Fungal Biol, 2016. **120**(3): p. 385-92.

18. Fedkenheuer, K., E., McDowell, J., M., *Recent Advancements in Integrated Management of Phytophthora Root and Stem Rot Disease of Soybean.* Phytopathology, 2016.
19. Wrather, A., Shannon, G., Balardin, R., Carregal, L., Escobar, R., Gupta, G. K., Ma, Z., Morel, W., Ploper, D., and Tenuta, A., *Effects of Disease on Soybean Yield in the Top Eight Producing Countries in 2006.* Plant Management Network, 2010.
20. Cohen, Y., et al., *Resurgence of Pseudoperonospora cubensis: The Causal Agent of Cucurbit Downy Mildew.* Phytopathology, 2015. **105**(7): p. 998-1012.
21. Savory, E.A., et al., *The cucurbit downy mildew pathogen Pseudoperonospora cubensis.* Molecular Plant Pathology, 2011. **12**(3): p. 217-226.
22. Colucci, S., T. Wehner, and G. Holmes, *The downy mildew epidemic of 2004 and 2005 in the Eastern United States.* Cucurbitaceae 2006, Asheville, North Carolina, USA, 17-21 September 2006, 2006: p. 403-411.
23. Hardy, G.E.S.J., S. Barrett, and B.L. Shearer, *The future of phosphite as a fungicide to control the soilborne plant pathogen Phytophthora cinnamomi in natural ecosystems.* Australasian Plant Pathology. **30**(2): p. 133-139.
24. Coates, M.E. and J.L. Beynon, *Hyaloperonospora Arabidopsidis as a pathogen model.* Annu Rev Phytopathol, 2010. **48**: p. 329-45.
25. Soylu, E.M. and S. Soylu, *Light and Electron Microscopy of the Compatible Interaction Between Arabidopsis and the Downy Mildew Pathogen Peronospora parasitica.* Journal of Phytopathology, 2003. **151**(6): p. 300-306.
26. Koch, E. and A. Slusarenko, *Arabidopsis is susceptible to infection by a downy mildew fungus.* Plant Cell, 1990. **2**(5): p. 437-45.
27. Savory, E.A., et al., *The cucurbit downy mildew pathogen Pseudoperonospora cubensis.* Mol Plant Pathol, 2011. **12**(3): p. 217-26.
28. Naegele, R.P., et al., *Regional and Temporal Population Structure of Pseudoperonospora cubensis in Michigan and Ontario.* Phytopathology, 2016. **106**(4): p. 372-9.
29. Fan, J.J., et al., *Characterization of a TIR-NBS-LRR gene associated with downy mildew resistance in grape.* Genet Mol Res, 2015. **14**(3): p. 7964-75.
30. Francesca, S., et al., *Downy mildew (Plasmopara viticola) epidemics on grapevine under climate change.* Global Change Biology, 2006. **12**(7): p. 1299-1307.
31. Liu, R., et al., *Histological responses to downy mildew in resistant and susceptible grapevines.* Protoplasma, 2015. **252**(1): p. 259-70.
32. This, P., T. Lacombe, and M.R. Thomas, *Historical origins and genetic diversity of wine grapes.* Trends Genet, 2006. **22**(9): p. 511-9.
33. Simko, I., et al., *Resistance to Downy Mildew in Lettuce 'La Brillante' is Conferred by Dm50 Gene and Multiple QTL.* Phytopathology, 2015. **105**(9): p. 1220-8.
34. Davis, S.T.K.a.R.M., *UC IPM Pest Management Guidelines: Lettuce.* UC ANR Publication 3450 Diseases, 2012.
35. Wyenandt, C.A., et al., *Basil Downy Mildew (Peronospora belbahrii): Discoveries and Challenges Relative to Its Control.* Phytopathology, 2015. **105**(7): p. 885-94.

36. Gascuel, Q., et al., *The sunflower downy mildew pathogen Plasmopara halstedii.* Mol Plant Pathol, 2015. **16**(2): p. 109-22.
37. Karolina Vrandečić, D.J., Jasenka Ćosić, Jelena Poštić, Zorana Bijelić, *White Blister Species (Abluginaceae) on Weeds.* Poljoprivreda, 2011. **1**: p. 47-51.
38. Van Wyk, P.S., et al., *Early lodging, a novel manifestation of Albugo tragopogonis infection on sunflower in South Africa.* Helia, 1995. **18**: p. 83-90.
39. Lava, S.S., A. Heller, and O. Spring, *Oospores of Pustula helianthicola in sunflower seeds and their role in the epidemiology of white blister rust.* IMA Fungus, 2013. **4**(2): p. 251-258.
40. Correll, J.C., et al., *Spinach: better management of downy mildew and white rust through genomics.* European Journal of Plant Pathology, 2010. **129**(2): p. 193-205.
41. van West, P., A.A. Appiah, and N.A.R. Gow, *Advances in research on oomycete root pathogens.* Physiological and Molecular Plant Pathology, 2003. **62**(2): p. 99-113.
42. Fawke, S., M. Doumane, and S. Schornack, *Oomycete interactions with plants: infection strategies and resistance principles.* Microbiol Mol Biol Rev, 2015. **79**(3): p. 263-80.
43. Dorrance, A.E., D. Mills, A.E. Robertson, M.A. Draper, L. Giesler, and A.Tenuta, *Phytophthora root and stem rot of soybean.* The Plant Health Instructor, 2007.
44. Lamour, K.H., et al., *The oomycete broad-host-range pathogen Phytophthora capsici.* Mol Plant Pathol, 2012. **13**(4): p. 329-37.
45. Ferguson, A.J. and S.N. Jeffers, *Detecting Multiple Species of Phytophthora in Container Mixes from Ornamental Crop Nurseries.* Plant Disease, 1999. **83**(12): p. 1129-1136.
46. Lucas, J.L.P.a.S., *Sudden oak death and ramorum blight.* The Plant Health Instructor, 2008.
47. Saville, A., et al., *Fungicide Sensitivity of U.S. Genotypes of Phytophthora infestans to Six Oomycete-Targeted Compounds.* Plant Disease, 2014. **99**(5): p. 659-666.
48. Okubara, P.A., M.B. Dickman, and A.E. Blechl, *Molecular and genetic aspects of controlling the soilborne necrotrophic pathogens Rhizoctonia and Pythium.* Plant Sci, 2014. **228**: p. 61-70.
49. Peters, R.D., H.W.B. Platt, and C.A. Lévesque, *First report ofPythium sylvaticum causing potato tuber rot.* American Journal of Potato Research. **82**(2): p. 173-177.
50. Broders, K.D., et al., *Characterization of Pythium spp. Associated with Corn and Soybean Seed and Seedling Disease in Ohio.* Plant Disease, 2007. **91**(6): p. 727-735.
51. A. E. Dorrance, S.A.B., P. Bowen, and P. E. Lipps, *Characterization of Pythium spp. from Three Ohio Fields for Pathogenicity on Corn and Soybean and Metalaxyl Sensitivity.* Plant Health Progress, 2004.
52. Papavizas, G.C. and W.A. Ayers, *Aphanomyces species and their root diseases in pea and sugarbeet: a review*. Vol. 1480. 1974: Agricultural Research Service, US Department of Agriculture.
53. Gaulin, E., et al., *Root rot disease of legumes caused by Aphanomyces euteiches.* Molecular Plant Pathology, 2007. **8**(5): p. 539-548.

54. Pieterse, C.M., et al., *Networking by small-molecule hormones in plant immunity.* Nat Chem Biol, 2009. **5**(5): p. 308-16.
55. Tsuda, K. and F. Katagiri, *Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity.* Curr Opin Plant Biol, 2010. **13**(4): p. 459-65.
56. Koller, T. and A.F. Bent, *FLS2-BAK1 extracellular domain interaction sites required for defense signaling activation.* PLoS One, 2014. **9**(10): p. e111185.
57. Gómez-Gómez, L., G. Felix, and T. Boller, *A single locus determines sensitivity to bacterial flagellin in Arabidopsis thaliana.* The Plant Journal, 1999. **18**(3): p. 277-284.
58. 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.
59. Sharp, J.K., B. Valent, and P. Albersheim, *Purification and partial characterization of a beta-glucan fragment that elicits phytoalexin accumulation in soybean.* J Biol Chem, 1984. **259**(18): p. 11312-20.
60. Umemoto, N., et al., *The structure and function of a soybean beta-glucan-elicitor-binding protein.* Proc Natl Acad Sci U S A, 1997. **94**(3): p. 1029-34.
61. Jones, J.D. and J.L. Dangl, *The plant immune system.* Nature, 2006. **444**(7117): p. 323-9.
62. Jiang, R.H., et al., *RXLR effector reservoir in two Phytophthora species is dominated by a single rapidly evolving superfamily with more than 700 members.* Proc Natl Acad Sci U S A, 2008. **105**(12): p. 4874-9.
63. Govers, F. and K. Bouwmeester, *Effector trafficking: RXLR-dEER as extra gear for delivery into plant cells.* Plant Cell, 2008. **20**(7): p. 1728-30.
64. Anderson, R.G., et al., *Recent Progress in RXLR Effector Research.* Mol Plant Microbe Interact, 2015. **28**(10): p. 1063-72.
65. Vleeshouwers, V.G. and R.P. Oliver, *Effectors as tools in disease resistance breeding against biotrophic, hemibiotrophic, and necrotrophic plant pathogens.* Mol Plant Microbe Interact, 2014. **27**(3): p. 196-206.
66. Kankanala, P., K. Czymmek, and B. Valent, *Roles for rice membrane dynamics and plasmodesmata during biotrophic invasion by the blast fungus.* Plant Cell, 2007. **19**(2): p. 706-24.
67. Stergiopoulos, I. and P.J. de Wit, *Fungal effector proteins.* Annu Rev Phytopathol, 2009. **47**: p. 233-63.
68. O'Connell, R.J. and R. Panstruga, *Tete a tete inside a plant cell: establishing compatibility between plants and biotrophic fungi and oomycetes.* New Phytol, 2006. **171**(4): p. 699-718.
69. Godfrey, D., et al., *Powdery mildew fungal effector candidates share N-terminal Y/F/WxC-motif.* BMC Genomics, 2010. **11**: p. 317.
70. Mansfield, J., et al., *Top 10 plant pathogenic bacteria in molecular plant pathology.* Mol Plant Pathol, 2012. **13**(6): p. 614-29.
71. Joardar, V., et al., *Whole-genome sequence analysis of Pseudomonas syringae pv. phaseolicola 1448A reveals divergence among pathovars in genes involved in virulence and transposition.* J Bacteriol, 2005. **187**(18): p. 6488-98.

72. Buttner, D. and S.Y. He, *Type III protein secretion in plant pathogenic bacteria.* Plant Physiol, 2009. **150**(4): p. 1656-64.
73. Lindeberg, M., S. Cunnac, and A. Collmer, *The evolution of Pseudomonas syringae host specificity and type III effector repertoires.* Mol Plant Pathol, 2009. **10**(6): p. 767-75.
74. Zhang, J., Z. Yin, and F. White, *TAL effectors and the executor R genes.* Front Plant Sci, 2015. **6**: p. 641.
75. Rodriguez, P.A. and J.I. Bos, *Toward understanding the role of aphid effectors in plant infestation.* Mol Plant Microbe Interact, 2013. **26**(1): p. 25-30.
76. Bos, J.I., et al., *A functional genomics approach identifies candidate effectors from the aphid species Myzus persicae (green peach aphid).* PLoS Genet, 2010. **6**(11): p. e1001216.
77. Mitchum, M.G., et al., *Nematode effector proteins: an emerging paradigm of parasitism.* New Phytol, 2013. **199**(4): p. 879-94.
78. Rybicki, E.P., *A Top Ten list for economically important plant viruses.* Arch Virol, 2015. **160**(1): p. 17-20.
79. Zvereva, A.S. and M.M. Pooggin, *Silencing and innate immunity in plant defense against viral and non-viral pathogens.* Viruses, 2012. **4**(11): p. 2578-97.
80. St Clair, D.A., *Quantitative disease resistance and quantitative resistance Loci in breeding.* Annu Rev Phytopathol, 2010. **48**: p. 247-68.
81. Semagn, K., Å. Bjørnstad, and M. Ndjiondjop, *Principles, requirements and prospects of genetic mapping in plants.* African Journal of Biotechnology, 2006. **5**(25).
82. Mauricio, R., *Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology.* Nat Rev Genet, 2001. **2**(5): p. 370-81.
83. Jones, N., et al., *Markers and mapping revisited: finding your gene.* New Phytol, 2009. **183**(4): p. 935-66.
84. Acquaah, G., *Principles of plant genetics and breeding*. 2009: John Wiley & Sons.
85. Sattler, M.C., C.R. Carvalho, and W.R. Clarindo, *The polyploidy and its key role in plant breeding.* Planta, 2015. **243**(2): p. 281-296.
86. Collard, B.C. and D.J. Mackill, *Marker-assisted selection: an approach for precision plant breeding in the twenty-first century.* Philosophical Transactions of the Royal Society of London B: Biological Sciences, 2008. **363**(1491): p. 557-572.
87. Tanksley, S.D. and J.C. Nelson, *Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines.* Theor Appl Genet, 1996. **92**(2): p. 191-203.
88. van Berloo, R., et al., *Resistance QTL confirmed through development of QTL-NILs for barley leaf rust resistance.* Molecular Breeding, 2001. **8**(3): p. 187-195.
89. Edenberg, H.J. and Y. Liu, *Laboratory methods for high-throughput genotyping.* Cold Spring Harb Protoc, 2009. **2009**(11): p. pdb top62.
90. Jiang, Z., et al., *Genome Wide Sampling Sequencing for SNP Genotyping: Methods, Challenges and Future Development.* Int J Biol Sci, 2016. **12**(1): p. 100-8.
91. Brookes, A.J., *The essence of SNPs.* Gene, 1999. **234**(2): p. 177-86.

92. Deschamps, S., V. Llaca, and G.D. May, *Genotyping-by-sequencing in plants.* Biology, 2012. **1**(3): p. 460-483.
93. Maughan, P.J., et al., *Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance, and near-isogenic line analysis.* Theor Appl Genet, 1996. **93**(3): p. 392-401.
94. Ilut, D.C., et al., *Identification of haplotypes at the Rsv4 genomic region in soybean associated with durable resistance to soybean mosaic virus.* Theor Appl Genet, 2016. **129**(3): p. 453-68.
95. Meksem, K., et al., *Conversion of AFLP bands into high-throughput DNA markers.* Mol Genet Genomics, 2001. **265**(2): p. 207-14.
96. Li, Y.H., et al., *[Genetic variation of SNP loci based on candidate gene for resistance to soybean cyst nematode].* Yi Chuan, 2009. **31**(12): p. 1259-64.
97. Sunada, K., et al., *A new soybean variety "Suzu-hime".* Bull. Hokkaido Prefect. Agric. Exp. Stn, 1981. **45**: p. 89-100.
98. Suzuki, C., et al., *Genetic relationships of soybean cyst nematode resistance originated in Gedenshirazu and PI84751 on Rhg1 and Rhg4 loci.* Breed Sci, 2012. **61**(5): p. 602-7.
99. Sherman-Broyles, S., et al., *The wild side of a major crop: soybean's perennial cousins from Down Under.* Am J Bot, 2014. **101**(10): p. 1651-65.
100. Song, Q., et al., *Construction of high resolution genetic linkage maps to improve the soybean genome sequence assembly Glyma1.01.* BMC Genomics, 2016. **17**(1): p. 33.
101. Matthiesen, R.L., et al., *A Method for Combining Isolates of Phytophthora sojae to Screen for Novel Sources of Resistance to Phytophthora Stem and Root Rot in Soybean.* Plant Disease, 2016: p. PDIS-08-15-0916-RE.
102. Ross, H. and C.A. Huijsman, *[On the resistance of species of Solanum (Tuberarium) against the European Races of the potato nematode (Heterodera rostochiensis Woll.)].* Theor Appl Genet, 1969. **39**(3): p. 113-22.
103. Brown, C.R., H. Mojtahedi, and G.S. Santo, *Genetic Analysis of Resistance to Meloidogyne chitwoodi Introgressed from Solanum hougasii into Cultivated Potato.* J Nematol, 1999. **31**(3): p. 264-71.
104. Caromel, B., et al., *Mapping QTLs for resistance against Globodera pallida (Stone) Pa2/3 in a diploid potato progeny originating from Solanum spegazzinii.* Theor Appl Genet, 2003. **106**(8): p. 1517-23.
105. Tan, M.Y., et al., *The R(Pi-mcd1) locus from Solanum microdontum involved in resistance to Phytophthora infestans, causing a delay in infection, maps on potato chromosome 4 in a cluster of NBS-LRR genes.* Mol Plant Microbe Interact, 2008. **21**(7): p. 909-18.
106. Rauscher, G., et al., *Quantitative resistance to late blight from Solanum berthaultii cosegregates with R(Pi-ber): insights in stability through isolates and environment.* Theor Appl Genet, 2010. **121**(8): p. 1553-67.
107. Tan, M.Y., et al., *The effect of pyramiding Phytophthora infestans resistance genes R Pi-mcd1 and R Pi-ber in potato.* Theor Appl Genet, 2010. **121**(1): p. 117-25.
108. Haggard, J.E., E.B. Johnson, and D.A. St Clair, *Linkage relationships among multiple QTL for horticultural traits and late blight (P. infestans) resistance on*

chromosome 5 introgressed from wild tomato Solanum habrochaites. G3 (Bethesda), 2013. **3**(12): p. 2131-46.
109. Tarwacka, J., et al., *Interspecific somatic hybrids Solanum villosum (+) S. tuberosum, resistant to Phytophthora infestans.* J Plant Physiol, 2013. **170**(17): p. 1541-8.
110. Gibson, R.W., M.G. Jones, and N. Fish, *Resistance to potato leaf roll virus and potato virus Y in somatic hybrids between dihaploid Solanum tuberosum and S. brevidens.* Theor Appl Genet, 1988. **76**(1): p. 113-7.
111. Chen, L., et al., *Nuclear and cytoplasmic genome components of Solanum tuberosum + S. chacoense somatic hybrids and three SSR alleles related to bacterial wilt resistance.* Theor Appl Genet, 2013. **126**(7): p. 1861-72.
112. Sarfatti, M., et al., *An RFLP marker in tomato linked to the Fusarium oxysporum resistance gene I2.* Theor Appl Genet, 1989. **78**(5): p. 755-9.
113. Foolad, R., et al., *Identification of QTLs for early blight (Alternaria solani) resistance in tomato using backcross populations of a Lycopersicon esculentum x L. hirsutum cross.* Theor Appl Genet, 2002. **104**(6-7): p. 945-958.
114. Smith, J.E., et al., *Resistance to Botrytis cinerea in Solanum lycopersicoides involves widespread transcriptional reprogramming.* BMC Genomics, 2014. **15**: p. 334.
115. Kaloshian, I., W.H. Lange, and V.M. Williamson, *An aphid-resistance locus is tightly linked to the nematode-resistance gene, Mi, in tomato.* Proc Natl Acad Sci U S A, 1995. **92**(2): p. 622-5.
116. Yaghoobi, J., J.L. Yates, and V.M. Williamson, *Fine mapping of the nematode resistance gene Mi-3 in Solanum peruvianum and construction of a S. lycopersicum DNA contig spanning the locus.* Mol Genet Genomics, 2005. **274**(1): p. 60-9.
117. Pedley, K.F. and G.B. Martin, *Molecular basis of Pto-mediated resistance to bacterial speck disease in tomato.* Annu Rev Phytopathol, 2003. **41**: p. 215-43.
118. Pitblado, R. and E. Kerr, *A source of resistance to bacterial speck-Pseudomonas tomato.* Tomato Genet Coop Rep, 1979. **29**: p. 30.
119. Garcia-Cano, E., et al., *Resistance to Tomato chlorosis virus in wild tomato species that impair virus accumulation and disease symptom expression.* Phytopathology, 2010. **100**(6): p. 582-92.
120. Brar, D.S. and G.S. Khush, *Alien introgression in rice.* Plant Mol Biol, 1997. **35**(1-2): p. 35-47.
121. Ballini, E., et al., *Modern elite rice varieties of the 'Green Revolution' have retained a large introgression from wild rice around the Pi33 rice blast resistance locus.* New Phytol, 2007. **175**(2): p. 340-50.
122. Das, A., et al., *A novel blast resistance gene, Pi54rh cloned from wild species of rice, Oryza rhizomatis confers broad spectrum resistance to Magnaporthe oryzae.* Funct Integr Genomics, 2012. **12**(2): p. 215-28.
123. Devanna, N.B., J. Vijayan, and T.R. Sharma, *The blast resistance gene Pi54of cloned from Oryza officinalis interacts with Avr-Pi54 through its novel non-LRR domains.* PLoS One, 2014. **9**(8): p. e104840.
124. Lee, S., et al., *Molecular evolution of the rice blast resistance gene Pi-ta in invasive weedy rice in the USA.* PLoS One, 2011. **6**(10): p. e26260.

125. Qiu, Y., et al., *Development and characterization of japonica rice lines carrying the brown planthopper-resistance genes BPH12 and BPH6.* Theor Appl Genet, 2012. **124**(3): p. 485-94.
126. Farooq, S. and F. Azam, *Production of low input and stress tolerant wheat germplasm through the use of biodiversity residing in the wild relatives.* Hereditas, 2001. **135**(2-3): p. 211-5.
127. Stepien, L., V. Holubec, and J. Chelkowski, *Resistance genes in wild accessions of Triticeae--inoculation test and STS marker analyses.* J Appl Genet, 2002. **43**(4): p. 423-35.
128. Qiu, J.W., et al., *Physical mapping and identification of a candidate for the leaf rust resistance gene Lr1 of wheat.* Theor Appl Genet, 2007. **115**(2): p. 159-68.
129. Wang, Y., et al., *The 3Ns chromosome of Psathyrostachys huashanica carries the gene(s) underlying wheat stripe rust resistance.* Cytogenet Genome Res, 2011. **134**(2): p. 136-43.
130. Favero, A.P., et al., *Successful crosses between fungal-resistant wild species of Arachis (section Arachis) and Arachis hypogaea.* Genet Mol Biol, 2015. **38**(3): p. 353-65.
131. Kumar, D. and P.B. Kirti, *Transcriptomic and proteomic analyses of resistant host responses in Arachis diogoi challenged with late leaf spot pathogen, Phaeoisariopsis personata.* PLoS One, 2015. **10**(2): p. e0117559.
132. Michelotto, M.D., et al., *Identification of Fungus Resistant Wild Accessions and Interspecific Hybrids of the Genus Arachis.* PLoS One, 2015. **10**(6): p. e0128811.
133. Di Gaspero, G., et al., *Selective sweep at the Rpv3 locus during grapevine breeding for downy mildew resistance.* Theor Appl Genet, 2012. **124**(2): p. 277-86.
134. Schwander, F., et al., *Rpv10: a new locus from the Asian Vitis gene pool for pyramiding downy mildew resistance loci in grapevine.* Theor Appl Genet, 2012. **124**(1): p. 163-76.
135. Jiao, C., et al., *Transcriptome characterization of three wild Chinese Vitis uncovers a large number of distinct disease related genes.* BMC Genomics, 2015. **16**: p. 223.
136. Gao, Y.R., et al., *Identification and utilization of a new Erysiphe necator isolate NAFU1 to quickly evaluate powdery mildew resistance in wild Chinese grapevine species using detached leaves.* Plant Physiol Biochem, 2016. **98**: p. 12-24.
137. Pickering, R. and P.A. Johnston, *Recent progress in barley improvement using wild species of Hordeum.* Cytogenet Genome Res, 2005. **109**(1-3): p. 344-9.
138. Scholz, M., et al., *Ryd4 (Hb): a novel resistance gene introgressed from Hordeum bulbosum into barley and conferring complete and dominant resistance to the barley yellow dwarf virus.* Theor Appl Genet, 2009. **119**(5): p. 837-49.
139. Nazeer, W., et al., *Introgression of genes for cotton leaf curl virus resistance and increased fiber strength from Gossypium stocksii into upland cotton (G. hirsutum).* Genet Mol Res, 2014. **13**(1): p. 1133-43.
140. Belfanti, E., et al., *The HcrVf2 gene from a wild apple confers scab resistance to a transgenic cultivated variety.* Proc Natl Acad Sci U S A, 2004. **101**(3): p. 886-90.

141. Csorba, R., E.F. Kiss, and L. Molnar, *Reactions of some cucurbitaceous species Tozucchini yellow mosaic virus (ZYMV).* Commun Agric Appl Biol Sci, 2004. **69**(4): p. 499-506.
142. Martinez Zamora, M.G., A.P. Castagnaro, and J.C. Diaz Ricci, *Isolation and diversity analysis of resistance gene analogues (RGAs) from cultivated and wild strawberries.* Mol Genet Genomics, 2004. **272**(4): p. 480-7.
143. Wieckhorst, S., et al., *Fine mapping of the sunflower resistance locus Pl(ARG) introduced from the wild species Helianthus argophyllus.* Theor Appl Genet, 2010. **121**(8): p. 1633-44.
144. Kawashima, C.G., et al., *A pigeonpea gene confers resistance to Asian soybean rust in soybean.* Nat Biotechnol, 2016.
145. Lenman, M., et al., *Effector-driven marker development and cloning of resistance genes against Phytophthora infestans in potato breeding clone SW93-1015.* Theor Appl Genet, 2016. **129**(1): p. 105-15.
146. Wrather, J.A. and S.R. Koenning, *Estimates of disease effects on soybean yields in the United States 2003 to 2005.* J Nematol, 2006. **38**(2): p. 173-80.
147. www.soystats.com, *U. S. Crop Area Planted*. 2014, American Soybean Association.
148. Dalgaarg, R., Schmidt, J., Halberg, N., Christensen, P., Thrane, M., Pengue, W. , *LCA of Soybean Meal.* Int J LCA, 2008. **13**(3): p. 240 - 254.
149. Tyler, B.M., *Phytophthora sojae: root rot pathogen of soybean and model oomycete.* Mol Plant Pathol, 2007. **8**(1): p. 1-8.
150. Morris, P.F., E. Bone, and B.M. Tyler, *Chemotropic and contact responses of phytophthora sojae hyphae to soybean isoflavonoids and artificial substrates.* Plant Physiol, 1998. **117**(4): p. 1171-8.
151. Catanzariti, A.M., P.N. Dodds, and J.G. Ellis, *Avirulence proteins from haustoria-forming pathogens.* FEMS Microbiol Lett, 2007. **269**(2): p. 181-8.
152. Mao, Y. and B.M. Tyler, *The Phytophthora sojae genome contains tandem repeat sequences which vary from strain to strain.* Fungal Genet Biol, 1996. **20**(1): p. 43-51.
153. Qutob, D., et al., *Copy number variation and transcriptional polymorphisms of Phytophthora sojae RXLR effector genes Avr1a and Avr3a.* PLoS One, 2009. **4**(4): p. e5066.
154. Dong, S., et al., *Sequence variants of the Phytophthora sojae RXLR effector Avr3a/5 are differentially recognized by Rps3a and Rps5 in soybean.* PLoS One, 2011. **6**(7): p. e20172.
155. Anderson, R.G., et al., *Homologous RXLR effectors from Hyaloperonospora arabidopsidis and Phytophthora sojae suppress immunity in distantly related plants.* Plant J, 2012. **72**(6): p. 882-93.
156. Wang, Q., et al., *Transcriptional programming and functional interactions within the Phytophthora sojae RXLR effector repertoire.* Plant Cell, 2011. **23**(6): p. 2064-86.
157. Tyler, B.M., and Gizen, M.J., *Genomics of Plant-Associated Fungi and Oomycetes: Dicot Pathogens*

The Phytophthora sojae genome sequence: foundation for a revolution

ed. R.A. Dean, Lichens-Park, A. and Kole, C. Vol. 7. 2014: Springer.
158. Schmitthenner, A.F., *Problems and progress in control of Phytophthora root rot of soybean.* Plant Disease, 1985. **69**(4): p. 362-368.
159. Vleeshouwers, V., Oliver, R, *Effectors as Tools in Disease Resistance Breeding Against Biotrophic, Hemibiotrophic, and Necrotrophic Plant Pathogens.* Molecular Plant-Microbe Interactions, 2014. **27**(3): p. 196-206.
160. Vleeshouwers, V.G., et al., *Effector genomics accelerates discovery and functional profiling of potato disease resistance and phytophthora infestans avirulence genes.* PLoS One, 2008. **3**(8): p. e2875.
161. Staskawicz, B.J., D. Dahlbeck, and N.T. Keen, *Cloned avirulence gene of Pseudomonas syringae pv. glycinea determines race-specific incompatibility on Glycine max (L.) Merr.* Proc Natl Acad Sci U S A, 1984. **81**(19): p. 6024-8.
162. Staskawicz, B., et al., *Molecular characterization of cloned avirulence genes from race 0 and race 1 of Pseudomonas syringae pv. glycinea.* J Bacteriol, 1987. **169**(12): p. 5789-94.
163. Wise, H., Anderson, R., McDowell, M. J., Wang Y., Tyler, M. B., *Identifying core P. sojae effectors.* Unpublished Communication, 2016.
164. Dou, D., et al., *RXLR-mediated entry of Phytophthora sojae effector Avr1b into soybean cells does not require pathogen-encoded machinery.* Plant Cell, 2008. **20**(7): p. 1930-47.
165. Fabro, G., et al., *Multiple candidate effectors from the oomycete pathogen Hyaloperonospora arabidopsidis suppress host plant immunity.* PLoS Pathog, 2011. **7**(11): p. e1002348.
166. Lindeberg, M., S. Cunnac, and A. Collmer, *Pseudomonas syringae type III effector repertoires: last words in endless arguments.* Trends Microbiol, 2012. **20**(4): p. 199-208.
167. Thomas, W.J., et al., *Recombineering and stable integration of the Pseudomonas syringae pv. syringae 61 hrp/hrc cluster into the genome of the soil bacterium Pseudomonas fluorescens Pf0-1.* Plant J, 2009. **60**(5): p. 919-28.
168. Keen, N.T., Ersek, T., Long, M., Bruegger, B., Hoi, M., *Inhibition of the hypersensitive reaction of soybean to incompatible Pseudomonas spp. by blasticidin streptomycin or elevated temperature.* Physiological Plant Pathology, 1981. **18**: p. 325-337.
169. Hartman, G., Sinclair, J., Rupe, J., *Compedium of Soybean Diseses*. Fourth Edition, ed. G. Hartman, Sinclair, J., Rupe, J. 1999, St. Paul, Minnesota: APS Press.
170. Schmutz, J., et al., *Genome sequence of the palaeopolyploid soybean.* Nature, 2010. **463**(7278): p. 178-183.
171. Ortega, M.A. and A.E. Dorrance, *Microsporogenesis of Rps8/rps8 heterozygous soybean lines.* Euphytica, 2011. **181**(1): p. 77-88.
172. Zhu, Y.L., et al., *Single-Nucleotide Polymorphisms in Soybean.* Genetics, 2003. **163**(3): p. 1123-1134.
173. Hildebrand, A.A., *A Root and Stalk Rot of Soybeans cause by Phytophthora megasperma Drechsler var. sojae var. nov.* Canadian Journal of Botany, 1959. **37**(5): p. 927-957.

174. Keen, N.T. and R.I. Buzzell, *New disease resistance genes in soybean against Pseudomonas syringae pv glycinea: evidence that one of them interacts with a bacterial elicitor.* Theor Appl Genet, 1991. **81**(1): p. 133-8.
175. Kale, S.D. and B.M. Tyler, *Assaying Effector Function in Planta Using Double-Barreled Particle Bombardment*, in *Plant Immunity: Methods and Protocols*, M.J. McDowell, Editor. 2011, Humana Press: Totowa, NJ. p. 153-172.
176. Dou, D., et al., *Different domains of Phytophthora sojae effector Avr4/6 are recognized by soybean resistance genes Rps4 and Rps6.* Mol Plant Microbe Interact, 2010. **23**(4): p. 425-35.
177. Hulbert, C.Y.a.S., *Prospects for functional analysis of effectors from cereal rust fungi.* Euphytica, 2011(179): p. 57-67.
178. Upadhyaya, N.M., et al., *A bacterial type III secretion assay for delivery of fungal effector proteins into wheat.* Mol Plant Microbe Interact, 2014. **27**(3): p. 255-64.
179. Lin, F., et al., *Molecular mapping of two genes conferring resistance to Phytophthora sojae in a soybean landrace PI 567139B.* Theor Appl Genet, 2013. **126**(8): p. 2177-85.
180. USDA/FAS, *Oilseeds: World Markets and Trade*, U.S.D.o.A.a.F.A. Service, Editor. 2016.
181. USDA, *World Agricultural Supply and Demand Estimates.* 2016.
182. Shannon, A.W.a.G., *Effect of Diseases on Soybean Yield in the Top Eight Producing Countries in 2006.* Plant Management Network (PMN), 2010.
183. Dorrance, A.E., D. Mills, A.E. Robertson, M.A. Draper, L. Giesler, and A.Tenuta, *Phytopthora root and stem rot of soybean.* The Plant Health Instructor, 2012.
184. Kale, S.D. and B.M. Tyler, *Entry of oomycete and fungal effectors into plant and animal host cells.* Cell Microbiol, 2011. **13**(12): p. 1839-48.
185. Win, J., et al., *Sequence divergent RXLR effectors share a structural fold conserved across plant pathogenic oomycete species.* PLoS Pathog, 2012. **8**(1): p. e1002400.
186. Bozkurt, T.O., et al., *Oomycetes, effectors, and all that jazz.* Curr Opin Plant Biol, 2012. **15**(4): p. 483-92.
187. Zhang, J., et al., *Effector-triggered and pathogen-associated molecular pattern-triggered immunity differentially contribute to basal resistance to Pseudomonas syringae.* Mol Plant Microbe Interact, 2010. **23**(7): p. 940-8.
188. Tyler, B.M., and Gizen, M.J., *Genomics of Plant-Associated Fungi and Oomycetes: Dicot Pathogens*. The Phytophthora sojae genome sequence: foundation for a revolution ed. R.A. Dean, Lichens-Park, A. and Kole, C. Vol. 7. 2014: Springer.
189. Raffaele, S. and S. Kamoun, *Genome evolution in filamentous plant pathogens: why bigger can be better.* Nat Rev Microbiol, 2012. **10**(6): p. 417-30.
190. Li, X., et al., *Physiological and proteomics analyses reveal the mechanism of Eichhornia crassipes tolerance to high-concentration cadmium stress compared with Pistia stratiotes.* PLoS One, 2015. **10**(4): p. e0124304.
191. Gijzen, M., C. Ishmael, and S.D. Shrestha, *Epigenetic control of effectors in plant pathogens.* Front Plant Sci, 2014. **5**: p. 638.
192. Dorrance, A.E., et al., *Pathotype Diversity of Phytophthora sojae in Eleven States in the United States.* Plant Disease, 2016: p. PDIS-08-15-0879-RE.

193. Kim, M.Y., et al., *Whole-genome sequencing and intensive analysis of the undomesticated soybean (Glycine soja Sieb. and Zucc.) genome.* Proc Natl Acad Sci U S A, 2010. **107**(51): p. 22032-7.
194. Wen, Z., et al., *Genetic diversity and peculiarity of annual wild soybean (G. soja Sieb. et Zucc.) from various eco-regions in China.* Theor Appl Genet, 2009. **119**(2): p. 371-81.
195. Joshi, T., et al., *Genomic differences between cultivated soybean, G. max and its wild relative G. soja.* BMC Genomics, 2013. **14 Suppl 1**: p. S5.
196. Winter, S.M., et al., *QTL associated with horizontal resistance to soybean cyst nematode in Glycine soja PI464925B.* Theor Appl Genet, 2007. **114**(3): p. 461-72.
197. Ilut, D.C., et al., *Identification of haplotypes at the Rsv4 genomic region in soybean associated with durable resistance to soybean mosaic virus.* Theor Appl Genet, 2015.
198. Oh, S.K., H. Kim, and D. Choi, *Rpi-blb2-mediated late blight resistance in Nicotiana benthamiana requires SGT1 and salicylic acid-mediated signaling but not RAR1 or HSP90.* FEBS Lett, 2014. **588**(7): p. 1109-15.
199. Wang, W., et al., *Timing of plant immune responses by a central circadian regulator.* Nature, 2011. **470**(7332): p. 110-4.
200. Tyler and McDowell, *Identification of Core Effectors in P. sojae.* Unpublished_Manuscript.
201. American Soybean Association, S. *U. S. Crop Area Planted*. 2015.
202. Gnanamanickam, S.S. and E.W.B. Ward, *Bacterial blight of soybeans: a new race of Pseudomonas syringae pv. glycinea and variations in systemic symptoms.* Canadian Journal of Plant Pathology, 1982. **4**(1): p. 73-78.
203. Liu, B., et al., *Candidate defense genes as predictors of quantitative blast resistance in rice.* Mol Plant Microbe Interact, 2004. **17**(10): p. 1146-52.
204. Goker, M., et al., *How do obligate parasites evolve? A multi-gene phylogenetic analysis of downy mildews.* Fungal Genet Biol, 2007. **44**(2): p. 105-22.
205. Baxter, L., et al., *Signatures of adaptation to obligate biotrophy in the Hyaloperonospora arabidopsidis genome.* Science, 2010. **330**(6010): p. 1549-51.
206. P.T.N Spencer-Phillips, U.G., A. Lebeda, *Advances in Downy Mildew Research*. 2002, New York, Boston, Dordrecht, London, Moscow: Kluwer Academic Publishers.
207. van der Biezen, E.A., et al., *Arabidopsis RPP4 is a member of the RPP5 multigene family of TIR-NB-LRR genes and confers downy mildew resistance through multiple signalling components.* Plant J, 2002. **29**(4): p. 439-51.
208. Cheng, Y. and G.M. Church, *Biclustering of expression data.* Proc Int Conf Intell Syst Mol Biol, 2000. **8**: p. 93-103.
209. Gu, J. and J.S. Liu, *Bayesian biclustering of gene expression data.* BMC Genomics, 2008. **9 Suppl 1**: p. S4.
210. Schwarz, G., *Estimating the Dimension of a Model.* The Annals of Statistics, 1978.
211. Swarbreck, D., et al., *The Arabidopsis Information Resource (TAIR): gene structure and function annotation.* Nucleic Acids Res, 2008. **36**(Database issue): p. D1009-14.

212. Fedkenheuer, K., *Analysis of Publically Available Transcriptome Data Reveals Arabidopsis thaliana genes which Impact Susceptiblity to Hyaloperonospora arabidopsidis* IN PRESS.
213. Ernst, J., G.J. Nau, and Z. Bar-Joseph, *Clustering short time series gene expression data.* Bioinformatics, 2005. **21 Suppl 1**: p. i159-68.
214. Krasileva, K.V., et al., *Global analysis of Arabidopsis/downy mildew interactions reveals prevalence of incomplete resistance and rapid evolution of pathogen recognition.* PLoS One, 2011. **6**(12): p. e28765.
215. Asai, S., et al., *Expression profiling during arabidopsis/downy mildew interaction reveals a highly-expressed effector that attenuates responses to salicylic acid.* PLoS Pathog, 2014. **10**(10): p. e1004443.
216. Lewis, J.D., et al., *The Arabidopsis ZED1 pseudokinase is required for ZAR1-mediated immunity induced by the Pseudomonas syringae type III effector HopZ1a.* Proc Natl Acad Sci U S A, 2013. **110**(46): p. 18722-7.
217. Pitorre, D., et al., *RLK7, a leucine-rich repeat receptor-like kinase, is required for proper germination speed and tolerance to oxidative stress in Arabidopsis thaliana.* Planta, 2010. **232**(6): p. 1339-53.
218. Staal, J., et al., *Transgressive segregation reveals two Arabidopsis TIR-NB-LRR resistance genes effective against Leptosphaeria maculans, causal agent of blackleg disease.* Plant J, 2006. **46**(2): p. 218-30.