GENETIC AND TRANSCRIPTOMIC ANALYSES OF THE *PSEUDOPERONOSPORA CUBENSIS*-*CUCUMIS SATIVUS* INTERACTION

By

Elizabeth Ann Savory

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Plant Pathology

2012

ABSTRACT

GENETIC AND TRANSCRIPTOMIC ANALYSES OF THE *PSEUDOPERONOSPORA CUBENSIS*-*CUCUMIS SATIVUS* INTERACTION

By

Elizabeth Ann Savory

Pseudoperonospora cubensis [(Berkeley & M. A. Curtis) Rostovzev] is the causal agent of cucurbit downy mildew, a foliar disease responsible for devastating losses worldwide of cucumber, cantaloupe, pumpkin, watermelon, and squash. In the United States, cucurbit downy mildew is the major threat to cucumber (*Cucumis sativus*) production and has been a significant limiting factor since the 2004 and 2005 growing seasons. Prior to 2004, host resistance had been an effective means of controlling the disease and, as such, limited research had been conducted to study the biology, genetics, or virulence of *Ps. cubensis*. The first step to expand the knowledge base of *Ps. cubensis* was the generation of a 64.4 Mbp genomic assembly of the MSU-1 isolate and prediction of 23,519 loci and 23,522 gene models. Similar to other oomycete plant pathogens, *Ps. cubensis* utilizes RXLR and RXLR-like effector proteins, which can function as either virulence or avirulence determinants during the course of host infection. Using *in silico* analyses, 271 candidate effector proteins were identified with variable RXLR motifs, including 20 different amino acids at position R1. Of these, only 109 (41%) had a putative ortholog in *Phytophthora infestans* and evolutionary rate analysis of these orthologs shows that they are evolving at a significantly faster rate than most other genes. One *Ps. cubensis* effector protein, RXLR protein 1 (*Psc*RXLR1) was characterized in detail. *Psc*RXLR1 was shown to be up-regulated during the early

stages of host infection, and elicits a cell death response in *Nicotiana benthamiana*. *Psc*RXLR1 was also demonstrated to be a product of alternative splicing, marking this as the first example of an alternative splicing event in plant pathogenic oomycetes transforming a non-effector gene into a functional effector protein.

We present the first large-scale global gene expression analysis of *Ps. cubensis* infection of a susceptible *C. sativus* cultivar, 'Vlaspik', and identification of both pathogen and host genes involved in infection and the defense response, respectively. Using mRNA-Seq, we captured differential expression of 2383 *Ps. cubensis* genes in sporangia and at 1, 2, 3, 4, 6, and 8 days post-inoculation (dpi). Co-expression analyses identified distinct modules of *Ps. cubensis* genes that were representative of early, intermediate, and late infection stages. Additionally, the expression of 15,286 cucumber genes was detected, of which 14,476 were expressed throughout the infection process from 1 dpi to 8 dpi. The rapid induction of key defense related genes, including catalases, chitinases, lipoxygenases, peroxidases, and protease inhibitors was detected within 1 dpi, suggesting recognition by *C. sativus* of the initial stages of *Ps. cubensis* infection. Co-expression network analyses revealed transcriptional networks with distinct patterns of expression including down-regulation at 2 dpi of known defense responsive genes suggesting coordinated suppression of host responses by the pathogen. In total, the work described herein presents an in-depth analysis of the interplay between host susceptibility and pathogen virulence in an agriculturally important pathosystem.

I would like to dedicate this dissertation to my family.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Brad Day for his unwavering support, advice, and guidance over the past 5 years.

I would also like to thank my committee members Drs. C. Robin Buell, Raymond Hammerschmidt, and Rebecca Grumet for their support, collaboration, and advice over the course of my dissertation research.

I would like to thank the Michigan State University Project GREEEN, Pickle Packers International, Inc, National Science Foundation East Asia and Pacific Summer Institutes and the United States Department of Agriculture National Institute of Food and Agriculture, Agriculture and Food Research Initiative for funding this reasearch.

Finally, I would like to thank my lab mates, Caleb Knepper, Katie Porter, Alyssa Burkhardt, Patricia Santos, and Masaki Shimono who always made coming to lab worthwhile, who were always there whether the data looked good or bad, and without whom I would not have made it through. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

The cucurbit downy mildew pathogen *Pseudoperonospora cubensis*

This review was previously published in Molecular Plant Pathology.

Savory EA, Granke LL, Quesada-Ocampo LM, Varbanova M, Hausbeck M, and Day B. 2011. The cucurbit downy mildew pathogen *Pseudoperonospora cubensis*. Mol Plant Pathol 12:217-226.

SUMMARY

Pseudoperonospora cubensis [(Berkeley & M. A. Curtis) Rostovzev], the causal agent of cucurbit downy mildew, is responsible for devastating losses worldwide of cucumber, cantaloupe, pumpkin, watermelon, and squash. Although downy mildew has been a major issue in Europe since the mid-1980s, in the United States, downy mildew on cucumber was successfully controlled for many years through host resistance. However, since the 2004 growing season, host resistance is no longer effective, and as a result, the control of downy mildew on cucurbits now requires an intensive fungicide program. Chemical control is not always feasible due to the high costs associated with fungicides and their application. Moreover, the presence of pathogen populations resistant to commonly used fungicides limits the long-term viability of chemical control. This review summarizes the current knowledge of taxonomy, disease development, virulence, pathogenicity and control of *Ps. cubensis*. In addition, topics for future research that aim to develop both short- and long-term control measures of cucurbit downy mildew are discussed.

Taxonomy: Kingdom Straminipila; Phylum Oomycota; Class Oomycetes; Order Peronosporales; Family Peronosporaceae; Genus *Pseudoperonospora*; Species *Pseudoperonospora cubensis*.

Disease Symptoms: Angular chlorotic lesions bound by leaf veins on the foliage of cucumber. Symptoms vary on different cucurbit species and varieties, specifically in

terms of lesion development, shape, and size. Infection of cucurbits by *Ps. cubensis* impacts fruit yield and overall plant health.

Infection process: Sporulation on the underside of leaves results in the production of sporangia that are wind-dispersed. Upon arrival to a susceptible host, sporangia germinate in free water on the leaf surface, producing biflagellate zoospores that swim to and encyst on stomata where they form germ tubes. An appressorium is produced and forms a penetration hypha, which enters the leaf tissue through the stomata. Hyphae grow through the mesophyll and establish haustoria, specialized structures for the transfer of nutrients and signals between host and pathogen.

Control: Management of downy mildew in Europe requires the use of tolerant cucurbit cultivars in conjunction with fungicide applications. In the U.S., an aggressive fungicide program with sprays every five to seven days for cucumber, and every seven to ten days for other cucurbits, has been necessary to control outbreaks and prevent crop loss.

Useful websites: http://www.daylab.plp.msu.edu/pseudoperonospora-cubensis/ (Day Lab website with research advances in downy mildew), http://veggies.msu.edu/ (Hausbeck Lab website with downy mildew news for growers), http://cdm.ipmpipe.org/ (Cucurbit downy mildew forecasting homepage), http://ipm.msu.edu/downymildew.htm (Downy mildew information for Michigan's vegetable growers).

INTRODUCTION

Cucurbit downy mildew (caused by *Pseudoperonospora cubensis*) is one of the most important foliar diseases of cucurbits, causing significant yield losses in the U.S., Europe, China and Israel [1]. The pathogen has a wide geographic distribution and has been reported in over 70 countries, including environments ranging from semi-arid to tropical. In addition, *Ps. cubensis* has a wide host range, infecting approximately 20 different genera of cucurbits [2,3]. The cucurbit crops grown in the U.S. that are susceptible to this aerially dispersed oomycete pathogen are valued at more than $246.2 million [4].

Control of downy mildew relies on application of fungicides and the use of host resistance. Nonetheless, fungicide-resistant *Ps. cubensis* populations have been documented throughout the world [3,5,6,7,8], and host resistance is no longer sufficient to control the disease as it once was in the U.S. [9]. The additional cost of fungicides, coupled with potential yield losses of up to 100% caused by downy mildew, threaten the long-term viability of cucurbit crop production [3,8,10].

Detailed knowledge of *Ps. cubensis* epidemiology, infection processes, and population genetics is currently lacking, but is necessary to guide future efforts in developing new resistant varieties and fungicides, as well as preventing pathogen populations from overcoming host resistance and chemical control. Studies of pathogen epidemiology and global population genetics, evaluation of fungicides for disease control, and

development of a high-coverage draft genome sequence will assist in our understanding of the pathogen, as well as in developing effective diagnostics and control measures for *Ps. cubensis*. The aim of this review is to briefly summarize what is currently known about the cucurbit downy mildew pathogen, *Ps. cubensis*, including taxonomy, disease development, virulence, pathogenicity and management.

TAXONOMY AND MORPHOLOGY

Ps. cubensis is the type species of the genus *Pseudoperonospora*, which includes five accepted species: *Ps. cubensis*, *Ps. humuli*, *Ps. cannabina*, *Ps. celtidis* and *Ps. urticae*, [11]. In addition, there are reports of a sixth species, *Ps. cassiae*, which, while rare, may also be a true species of *Pseudoperonospor*a [12]. Originally named *Peronospora cubensis* when discovered in Cuba by Berkeley and Curtis [13], *Ps. cubensis* was reclassified in 1903 after further observations of sporangia germination [14]. *Pseudoperonospora* species have true sporangia that germinate via cytoplasmic cleavage to produce zoospores (Figure 1.2c); whereas, species of *Peronospora* have sporangia that germinate directly via a germ tube [2,14].

Recent work has shown that there are no significant morphological differences between *Ps. cubensis* and the hop (*Humulus* spp.) pathogen *Ps. humuli;* nonetheless, there is no evidence that *Ps. humuli* can infect cucurbits, and limited support for *Ps. cubensis* pathogenicity on hops [15]. Molecular evidence also shows conflictive results. Internal transcribed spacer (ITS) region sequences of both pathogens are highly similar, which

suggests that *Ps. humuli* could be a taxonomic synonym of *Ps. cubensis* [11]. However, a recent study using single nucleotide polymorphisms (SNPs) indicates that two nuclear and one mitochondrial gene support the separation of *Ps. cubensis* from *Ps. humuli* [15]. In addition, host range studies have demonstrated pathogenic differences between *Ps. cubensis* and *Ps. humuli* that further supports separation of these species [16]. Overall, these genetic, phenotypic and physiological characterizations of *Ps. cubensis* and *Ps. humuli* provide support for the distinction between these species. Further studies including evidence from hundreds of loci would be helpful to fully resolve the phylogeny of these closely related species.

Morphological characters may not provide enough information for characterization of *Ps. cubensis* isolates or even for differentiating species of *Pseudoperonospora* [17]. *Ps. cubensis* sporangiophore morphology can vary with temperature, while sporangia dimensions are influenced by cucurbit host [12,18]. Recent work with a single isolate of *Ps. cubensis* inoculated onto six different cucurbit species has shown that the host cell matrix plays a role in influencing five morphological criteria, including sporangiophore length, length of ultimate branchlets, sporangial length and width, and the ratio between sporangial length and width [17]. The differences among these morphological characteristics were more obvious in phylogenetically unrelated hosts. These results indicate that it is desirable to include information from genetic markers when resolving phylogenetic relationships in species of *Pseudoperonospora*.

SYMPTOMS AND SIGNS

Downy mildew of cucurbits is a foliar disease and is easily recognizable by the development of chlorotic lesions on the adaxial leaf surface, sometimes with necrotic centers. These lesions can be restricted by the leaf vein, as in cucumber, giving them an angular appearance (Figure 1.1). In other cucurbits, the symptoms may vary slightly in terms of shape and color. For example, in both melon and watermelon, foliar lesions are less defined than those on cucumber, and are not always bound by leaf veins [1]. As infection progresses, the chlorotic lesions expand and may become necrotic [19], with necrosis occurring more quickly in hot, dry weather [20]. Leaves colonized by *Ps. cubensis* undergo changes in temperature and transpiration rates, which vary during the course of infection and over the leaf surface [19,21]. Low temperatures can delay symptom development while still promoting colonization of the leaf tissue, while higher temperatures result in faster lesion chlorosis that may inhibit pathogen growth [22]. As the downy mildew disease progresses, entire leaves may die within days following the initial infection, as lesions expand and coalesce [1]. A reduced canopy leads to cessation of fruit development and increased sun exposure of the fruit, allowing for sun scald and secondary rots [23]. Ultimately, crop yield and fruit quality are affected (Figure 1.1).

When temperatures are below those that allow lesion formation and relative humidity is ≥90%, sporulation, the eponymous “downy” appearance on the lower leaf surface, may be the first sign of disease (Figure 1.1) [24]. Hyaline sporangiophores (180-400 μm)

Figure 1.1

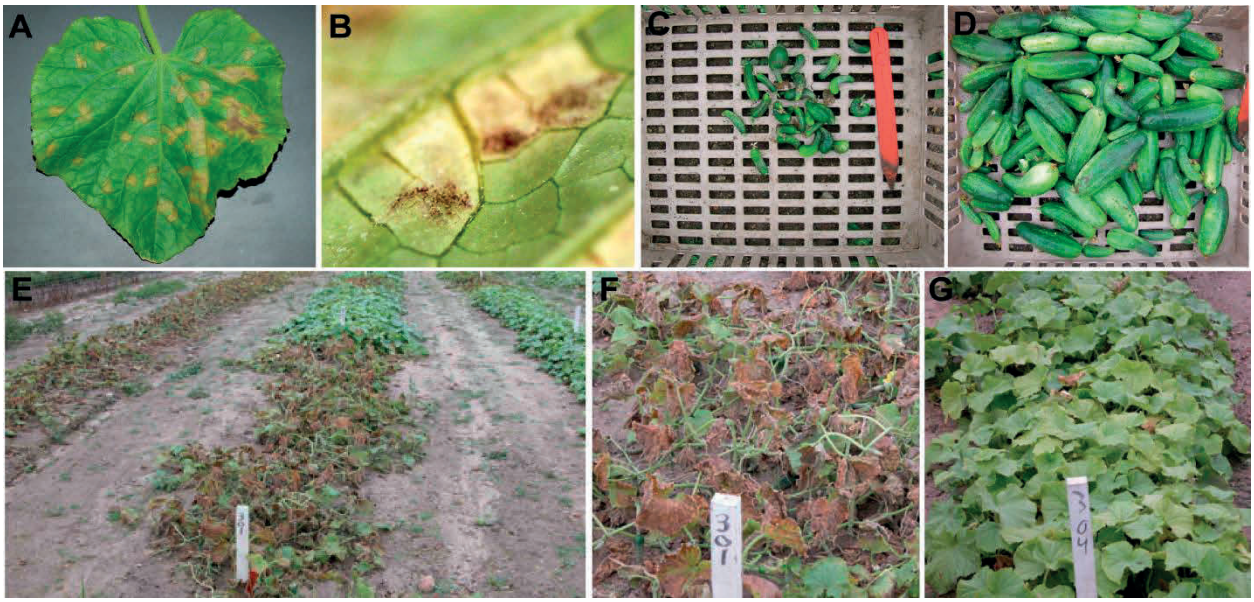

Figure 1.1 Symptoms caused by *Pseudoperonospora cubensis*. (A) Yellow angular lesions on a cucumber leaf. (B) Typical "downy" appearance on abaxial leaf surface caused by sporulation. Yields from untreated (C) and treated (D) cucumbers infected with *Ps. cubensis*. E. Field showing typical disease symptoms. Severe symptoms on untreated (F) versus treated (G) cucumbers. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

bearing papillate, lemon-shaped, gray-purple sporangia (20-40 x 14-25 µm) on sterigmata emerge in groups of 1-6 from stomata on the abaxial surface of infected leaves (Figure 1.2) [11,25]. While leaf wetness is prohibitory for sporangium production, a period of near-saturated relative humidity must occur for ≥6 hours to induce sporulation [26]. Sporulation, as in other downy mildews, is dependent on the diurnal cycle, and is enhanced by longer photoperiods [27]. Differentiation of sporangia requires a minimum 6 hour dark period [28]. The optimum temperature for sporangia production is 15-20°C, but sporangia may form on cucumber at temperatures from 5 to

Figure 1.2

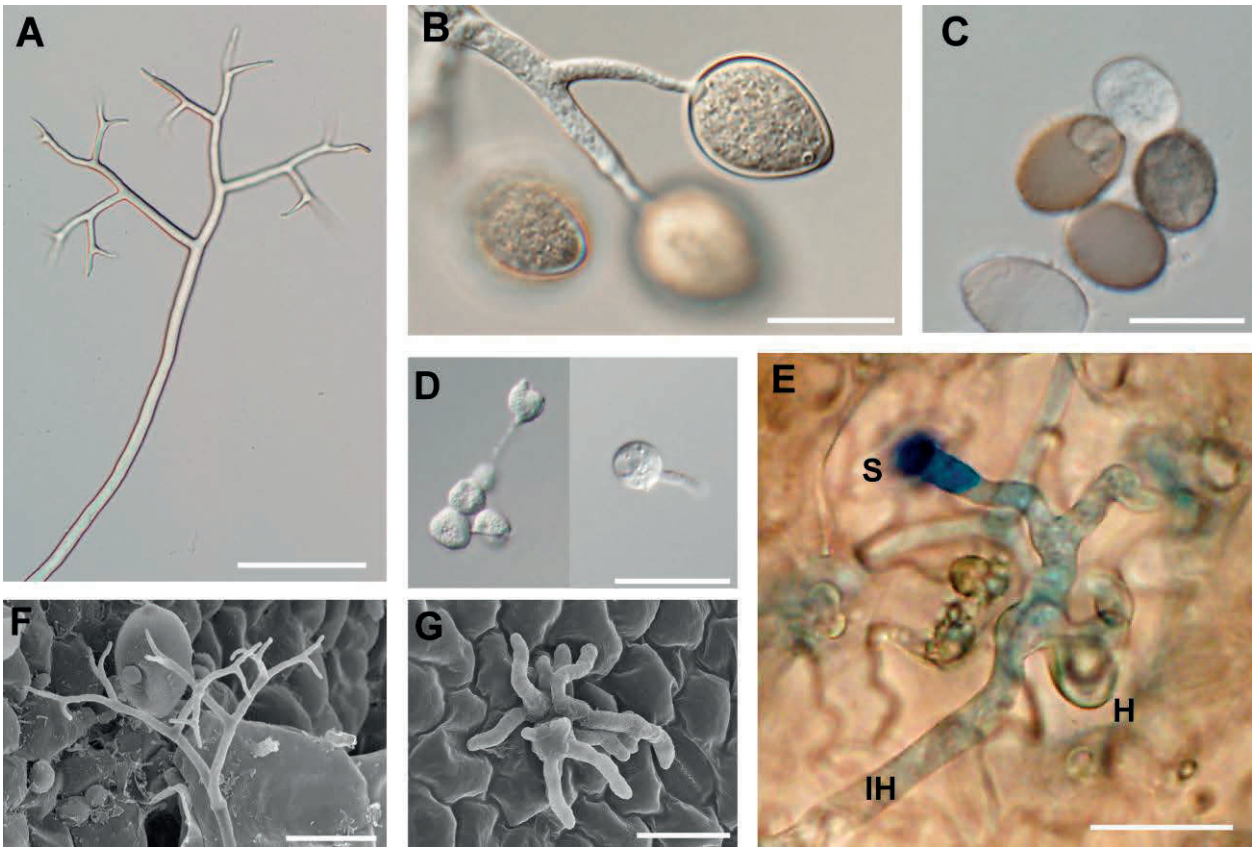

Figure 1.2 Morphology of *Pseudoperonospora cubensis*. (A) Sporangiophore (bar = 50 µm). (B) Sporangia attached to distal end of sporangiophores. (C) Sporangia germinating via cytoplasmic cleavage. (D) Left panel: zoospores. Right panel: encysted zoospore with germ tube. (E) Intercellular growth: S, stomata; H, haustorium; IH, intercellular hyphae. (F) Scanning electron micrograph (SEM) of sporangiophore (bar = 20 µm). (G) SEM of multiple sporangiophores emerging through stomata (bar = 20 µm). Bars = 25 µm, except where noted.

30°C [1]. Other factors such as the host species, cultivar, host nutritional status, and host age may also affect sporulation [26].

DISPERSAL AND SURVIVAL

Ps. cubensis cannot overwinter in geographic locations with killing frosts. Instead, the pathogen overwinters in areas with mild winter temperatures that permit cucurbit hosts to be grown year round [29] or in greenhouses [1]. It was recently demonstrated that *Ps. cubensis* could infect a perennial member of the Cucurbitaceae, *Bryonia dioica*, in the laboratory, and the pathogen could potentially overwinter on this host in Central and Northern Europe [30]. However, this has not been supported with observations in the field [31], and it is unknown if *B. dioica* plays an important role in the life cycle of *Ps. cubensis* [30]. While oospores have been observed in both temperate and tropical regions including India, Japan, Austria, Russia, China, Italy, and Israel [2,32,33,34,35,36], the production of oospores is very rare [1,2]. The rare occurrence of thick-walled resting structures, *ie.*, oospores, limits *Ps. cubensis* survival in the absence of a living host. It is currently unknown if oospores play an important role in the disease cycle (Figure 1.3).

As an obligate biotroph *Ps. cubensis* requires live host tissue for reproduction and dispersal. Copious asexual sporangia are produced on infected foliage, which may be liberated to the air following a reduction in relative humidity when hygroscopic twisting movements of sporangiophores actively release sporangia into air currents [37]. Hence, airborne *Ps. cubensis* sporangia concentrations are greater in the morning and early afternoon when changes in relative humidity and leaf wetness tend to occur [27]. The distance sporangia travel depends upon where in the canopy the sporangia are

Figure 1.3

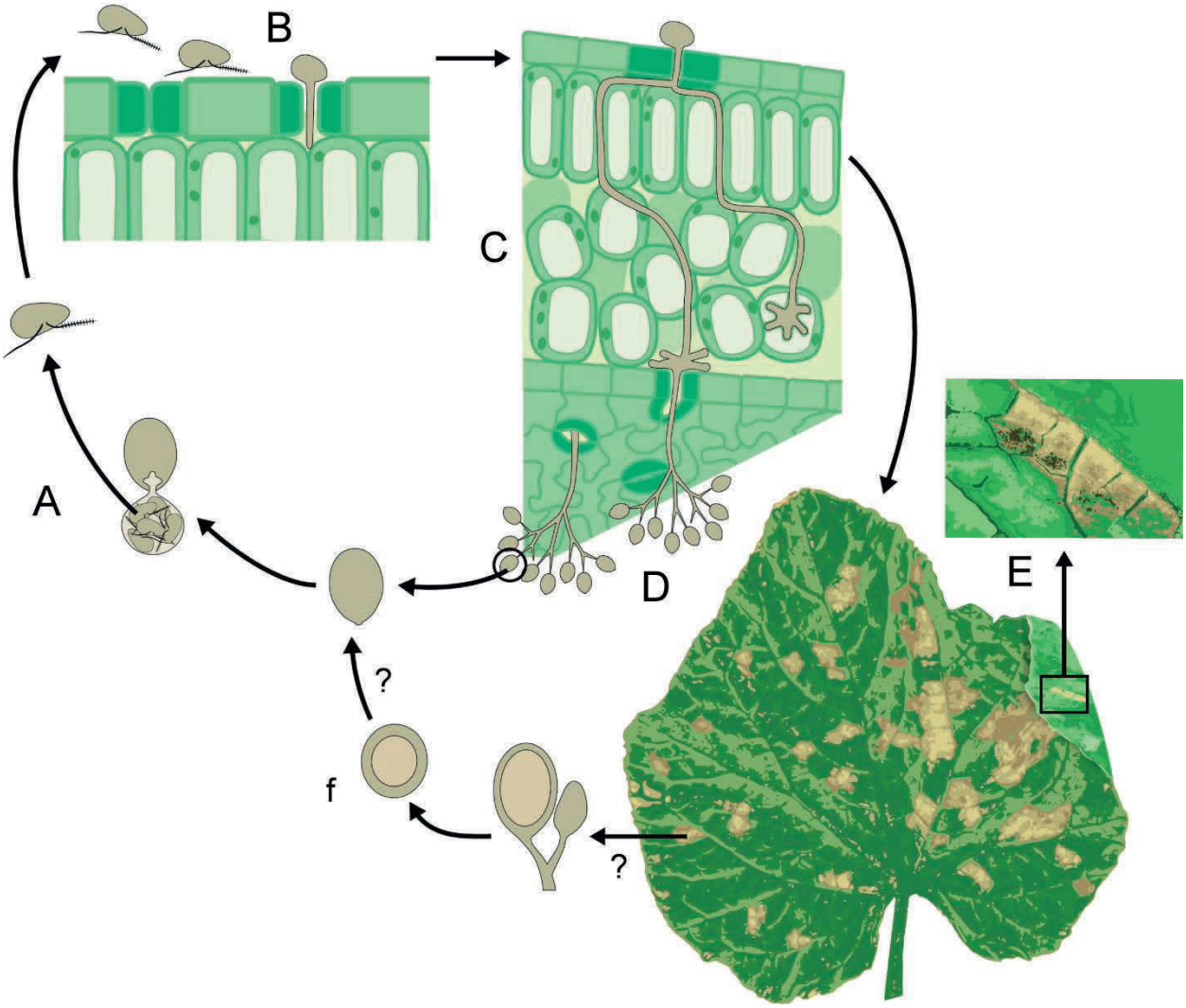

Figure 1.3 Life cycle of *Pseudoperonospora cubensis*. (A) Aerially dispersed lemon-shaped gray-purple sporangia land on the leaf surface and germinate in free moisture to form biflagellate zoospores. (B) Zoospores swim to and encyst in stomata, then penetrate the leaf surface via a germ tube. (C) Hyphae colonizes the mesophyll layer, establishing clavate-branched haustoria within plant cells. (D) The diurnal cycle triggers sporulation and up to 6 sporangiophores emerge through each stomate, bearing sporangia at their tips. Sporangia are dislodged from sporangiophores by changes in hydrostatic pressure and are picked up by wind currents that carry them to their next host. (E) Chlorotic, angular lesions bound by leaf veins are a symptom of *Ps. cubensis* infection visible on the adaxial leaf surface. On the lower leaf surface, sporulation is visible (inset). (F) The role of the sexual stage of *Ps. cubensis* is unknown.

produced, as well as wind conditions as they become airborne [38]. Like other downy mildews [39,40], *Ps. cubensis* sporangia may be wind-dispersed over long distances [1]. As such, it has been proposed that in the U.S., *Ps. cubensis* overwinters in frost-free areas of the southern states (e.g., Florida and Texas) and spreads northward each growing season via wind currents [41,42,43]. Likewise, sporangia that infect cucurbits in Central Europe originate from year-round production areas in Southeast Europe, and are transported via wind currents [44]. Subsequent local transport of secondary inoculum occurs primarily via wind, but sporangia also may be dispersed by rain splash or physical transfer on equipment within a field [1].

Sporangial survival during transport is limited to 1 to 16 days after dispersal [26,45] depending on temperature, relative humidity [1] and solar radiation [46]. Once a sporangium lands on a host plant, that sporangium must survive until environmental conditions are favorable for infection. Sporangial survival is favored by conditions of low relative humidity, lower temperature and cloudy days [45,46].

INFECTION MECHANISMS

While liberation and dispersal of sporangia occurs under conditions of low leaf moisture, leaf wetness is necessary for the pathogen to successfully germinate and infect the host plant. At 15°C, the optimum temperature for infection, at least 2 hours of leaf wetness are required for infection when high levels of inoculum are present. Sporangia may also germinate and cause infection at temperatures ranging from 5-28°C [26], but longer

periods of leaf wetness are required [22,24] under these conditions or when less inoculum is present [26]. The incubation period depends on temperature, photoperiod, inoculum concentration, and leaf wetness duration, and can range from 4 to 12 days [1,22].

Sporangia germinate via cytoplasmic cleavage, resulting in the release of 2-15 motile, biflagellate zoospores [2], which preferentially swim to open stomata where they encyst [47] (Figure 1.2; Figure 1.3). Germ tubes form from encysted zoospores and produce appressoria. A penetration hypha develops from the appressorium and enters through the stomatal aperture into the leaf tissue. Hyaline coenocytic hyphae subsequently form and grow intercellularly through the mesophyll and palisade tissues. Clavate-branched haustoria are established within mesophyll cells where they invaginate the plant cell membrane [48,49] (Figure 1.2; Figure 1.3). These specialized structures are the site of nutrient uptake by the pathogen and delivery of effector proteins that could potentially function to redirect host metabolism and suppress defense responses [50,51].

PATHOGENICITY AND VIRULENCE

Ps. cubensis has a broad host range, infecting over 49 species in 20 genera within the Cucurbitaceae, including 19 species in the genus *Cucumis* [2,52]. Cucumber (*Cucumis sativus* L.), melon (*Cucumis melo* L.), watermelon (*Citrullus lanatus* Matsum. and Nakai), and squash (*Cucurbita* spp.) are the four major food crops that are hosts to *Ps. cubensis*. Other cucurbits infected by *Ps. cubensis* include loofa (*Luffa acutangula* (L.)

Roxb.), bottle gourd (*Lagenaria siceraria* (Molina) Stand.), wax gourd (*Benincasa hispida* (Thunb.) Cogn.) and bitter melon (*Momordica charantia* L.) [2].

Ps. cubensis isolates show differences in virulence and pathogenicity depending on the cucurbit variety. To date, six physiological races have been identified in the U.S., Israel and Japan, and additional evidence suggests that many more exist in Europe [36,53,54]. Using *Citrullus*, *Cucumis*, and *Cucurbita* spp., Thomas *et al.* [55] identified five distinct physiological races of *Ps. cubensis*: 1 and 2 from Japan, 3 from Israel, and 4 and 5 from the U.S.. In 2003, Cohen *et al.* identified a sixth physiological race in Israel based on its pathogenicity to a wider range of susceptible cucurbits compared to race 3. All six physiological races that have been described are pathogenic on cucumber and muskmelon (*C. melo* var. *reticulatus)* but show differences in pathogenicity on watermelon, squash or pumpkin. Subsequently, Lebeda and Widrlechner (2003) developed a set of differential taxa that included 12 representatives from 6 genera, *Benincasa*, *Citrullus*, *Cucumis*, *Cucurbita*, *Lagenaria* and *Luffa*, which represent natural hosts of *Ps. cubensis*. Using this set of hosts, the authors evaluated the differences in pathogenicity of 22 additional isolates from the Czech Republic, Spain, France and the Netherlands, which were classified as 13 physiological races based on their virulence [52,53]. Collectively, these studies were the first to describe differences in virulence and pathogenicity of *Ps. cubensis* in Europe in detail; however, the genetic basis for differences among physiological races has not been established.

Differences in effector content could be a potential explanation for differences among physiological races. Effector proteins have been shown to play roles as both virulence and avirulence determinants in other oomycetes [56,57,58,59]. Oomycete effector proteins contain the RXLR motif that is located downstream of the signal peptide and are under diversifying selection at the C-terminal domain [60]. Preliminary sequence data of the *Ps. cubensis* genome has yielded 61 putative effector proteins [61,62]. Of these, 32 were secreted proteins containing the RXLR motif typical of previously identified oomycete effector proteins, while the remaining 29 had an R to Q substitution at the first amino acid (*ie*., QXLR). A family of QXLR containing effectors, *PcQNE* (*Ps. cubensis* QXLR nuclear localized effectors), was shown to localize to the nucleus and the C-terminal domain was under diversifying selection, as has been observed for RXLR effectors [61]. Understanding the diversity and role of effector proteins is key to understanding pathogenicity and the genetic basis for virulence differences between isolates.

The pathogenic and genetic diversity of *Ps. cubensis* has been shown to vary temporally and geographically [2,55,63]. In a study by Lebeda and Urban [3], *Ps. cubensis* isolates were collected over a three-year period in the Czech Republic, and a general population shift from isolates with low pathogenicity to those with high pathogenicity was observed. This work also demonstrated that the variability of the population decreased from 33 different physiological races in 2001 to only 13 physiological races in 2003 [3]. An increased representation of a highly pathogenic physiological race that was capable of causing infection on all cucurbits studied was

also observed over the course of the experiments, indicating a population shift to high pathogenicity isolates [3]. Recently, using a combination of amplified fragment length polymorphism (AFLP) and ITS sequence analyses, molecular polymorphisms were identified among populations of *Ps. cubensis* from Greece, the Czech Republic, the Netherlands and France [54]. While there was no clear grouping of isolates based on their pathogenicity, AFLP analysis indicated genetic differentiation between the Greek isolates and those from central and western Europe [54]. The work by Sarris and colleagues (2009) suggested that the clustering of *Ps. cubensis* isolates from Greece corresponded to their geographical distribution rather than their pathogenicity or virulence on cucurbit hosts. Nevertheless, this study was supported on SNP evidence from one single loci and AFLP evidence with error rates of 1-4% due to band scoring. A similar study with information from several loci, a reliable polymorphism scoring system, and a more extensive sampling of regions of interest is needed to support or disprove what Sarris and colleagues (2009) proposed in their paper.

DISEASE MANAGEMENT

Ps. cubensis outbreaks over the past several decades have been responsible for annual yield losses of up to 80%, and as a result, cucurbit downy mildew is currently the most destructive disease of cucumbers for both field and greenhouse production in Europe [3,44]. Before 1984, downy mildew was not a major issue for Central [64,65] and Northern Europe [66,67]. However, from ~1985 onward, epidemics of downy mildew have been a challenge for cucurbit production in Europe [65]. In Europe, it has

been suggested that tolerant cultivars should be used in conjunction with fungicide applications when conditions are favorable for downy mildew [68,69]. In the U.S., host resistance introduced in the 1950s was effective in limiting losses caused by downy mildew without the use of fungicides. A resurgence of the disease was observed in many states along the Eastern Seaboard and the upper Midwest in 2004 and 2005, respectively [9]. This loss in resistance has led to the management of downy mildew through a spray program, with recommended fungicides applied every five to seven days on cucumbers [8,70] and every seven to ten days on other cucurbits [71].

An aggressive spray program is essential, as plants must have a protective barrier of fungicide prior to sporangium deposition to avoid yield losses. However, additional fungicide applications to control downy mildew greatly increase the cost of production. In Michigan, for example, major outbreaks of the disease have been observed since 2005, and the cost of additional fungicide sprays is over 6 million dollars annually (Hausbeck, *unpublished data*). Since fungicide applications are expensive, downy mildew sporangia trapping and forecasts can be useful tools to alert growers to when airborne sporangia are present or likely to be present so they can make an informed decision about when to initiate fungicide applications [43]. Delaying the initiation of a fungicide spray program may reduce management costs for growers and the amount of fungicides in the environment.

The efficacy of chemical control measures may be diminished if *Ps. cubensis* populations develop resistance to key fungicides; *Ps. cubensis* was the first oomycete

with documented resistance to metalaxyl and reduced sensitivity to mancozeb [63,72]. In addition, populations of *Ps. cubensis* resistant to strobulurin fungicides have been described [73]. In the Eastern U.S. field and greenhouse fungicide trials, products containing mefenoxam and strobilurins as the active ingredients failed to provide adequate downy mildew disease control, indicating that resistance is widespread in this region [5,8,70,74]. Since *Ps. cubensis* has a high potential of developing resistance to fungicides [75], it is important that populations are carefully monitored for resistance to currently registered products and that new active ingredients are tested.

While fungicide applications are currently necessary for adequate disease control [76], resistant varieties and cultural techniques are important components of a management strategy. The full genetic parameters controlling resistance to downy mildew in cucumber are unknown. The original source of host resistance (i.e., the recessive *dm1* gene) was identified in cucumber accession PI 197087 and first described in India in 1954 [77]. The resistance response governed by *dm1* is characterized by sparse pathogen sporulation, small necrotic lesions, tissue browning, and rapid cell death, indicative of the classical hypersensitive response (HR) type resistance. Since the 1950s, resistance conferred by *dm1* has been widely used in commercial cultivars for cucumber production in the U.S. and was sufficient to prevent losses due to downy mildew until 2004 [9]. Cultivars containing the *dm1* gene still show some level of resistance; unfortunately, the high level of resistance once observed is now lost. Additionally, susceptible cultivars without the *dm1* gene become infected earlier in the season, and exhibit more severe damage than was previously observed [43].

While the *dm1* remains useful in a disease management program, a robust source of resistance is highly desirable. Current breeding research for resistance to downy mildew in cucumber is focused on the identification of resistant germplasm(s) and cultivars via large scale screening trials [78,79]. Tolerant and high yielding germplasm has been identified in these studies, but a source of complete resistance to downy mildew in cucumber has been elusive, likely due to limited genetic diversity for *Ps. cubensis* resistance in cucumber [52,79]. Other *Cucumis* spp., such as melon, may be more relevant for identifying effective sources of resistance [52]. Wild melon line PI 124111F [PI], for example, has been shown to be resistant to the 6 physiological races of *Ps. cubensis* via two constitutively expressed glyoxylate aminotransferase-encoding genes, *At1* and *At2* [80]. These two genes are known as enzymatic resistance (eR) genes which when expressed at high levels in either wild type or transgenic plants, confer complete resistance to infection by *Ps. cubensis.*

FUTURE PROSPECTS

The re-emergence of cucurbit downy mildew in the U.S. and its persistence across much of Europe and Asia, represents a significant threat to cucurbit production worldwide. While the disease was successfully managed for decades within the U.S. using host resistance (i.e., the *dm-1* locus), severe epidemics have occurred since 2004. Whether this is due to a change in pathogen populations or a change in the environment is currently unknown. To this end, we need to investigate the changes in

pathogen populations, environmental factors, and how the pathogen-environment interaction affects the host-pathogen interaction and disease development.

Studies have investigated some aspects of the basic biology of *Ps. cubensis* and its interactions with various cucurbit hosts. However, additional research is needed to further clarify the taxonomy, variations in virulence and pathogenicity among physiological races, and the pathogenicity determinants of the pathogen. A better understanding of each of these components will ultimately facilitate the development of durable host resistance. To this end, the forthcoming genome sequence will provide molecular tools for gene discovery and the development of molecular markers, which may then be used to investigate population and evolutionary biology of the pathogen. Such studies will yield information about possible migration events or evolutionary changes within pre-existing U.S. pathogen populations that resulted in strains with increased virulence.

An integrated research approach that includes all factors affecting disease development (pathogen, host, and environment) is essential to control and predict future cucurbit downy mildew epidemics. First, it would benefit growers if new fungicides that are more economical and provide effective control were identified. Second, screening of breeding lines and wild germplasm will help identify durable sources of genetic resistance that would be preferable to an aggressive spray program. Finally, studies to determine the effects of environment on inocula and disease development will serve as a first step in developing rapid, and highly specific forecasting systems. In summary, research efforts contributing to the development of sustainable management strategies such as durable

host resistance are a priority to ensure the long-term viability of the cucurbit production industry.

ACKNOWLEDGEMENTS

The authors thank members of the Day lab for critical reading of the manuscript. Cucurbit downy mildew research in the Day lab is funded by the Michigan Agricultural Experiment Station (MAES), Project GREEEN (Award numbers GR06-0099D and GF07-077), the Michigan State University Office of the Vice President for Research and Graduate Studies, the National Science Foundation (Award number IOS-0641319), and a joint grant awarded to MKH and BD from the Agricultural Research Fund of Pickle Packers International Inc.. Work in the Hausbeck lab is supported by the Pickle and Pepper Research Committee of MSU, Fresh Vegetable Growers of Ontario, North Central IPM Center (Sub award 2003-51120-02111 S4256), and Project GREEEN (Award Numbers GR07-077 and GR06-0099D).

REFERENCES

REFERENCES

1. Thomas CE (1996) Downy mildew. In: Zitter TA, editor. Compendium of cucurbit diseases. Ithaca, NY: Cornell University Press. pp. 25-27.

2. Palti J, Cohen Y (1980) Downy mildew of cucurbits (*Pseudoperonospora cubensis*): the fungus and its hosts, distribution, epidemiology, and control. Phytoparasitica 8: 109-147.

3. Lebeda A, Urban J (2007) Temporal changes in pathogenicity and fungicide resistance in *Pseudoperonospora cubensis* populations. Acta Hort 731: 327-336.

4. Anonymous (2009) National Online Statistics. US Dep Agric Natl Agric Stat Serv.

5. Colucci SJ, Holmes GJ (2007) Fungicide insensitivity and pathotype determination of *Pseudoperonospora cubensis*, causal agent of cucurbit downy mildew. Phytopathology 97: S24.

6. Zhu S, Liu P, Liu X, Li J, Yuan S, et al. (2008) Assessing the risk of resistance in *Pseudoperonospora cubensis* to the fungicide flumorph in vitro. Pest Management Science 64: 255-261.

7. Mitani S, Araki S, Yamaguchi T, Takii Y, Ohshima T, et al. (2001) Biological properties of the novel fungicide cyazofamid against *Phytophthora infestans* on tomato and *Pseudoperonospora cubensis* on cucumber. Pest Manag Sci 58: 139-145.

8. Hausbeck MK, Cortright BD (2009) Evaluation of fungicides for control of downy mildew of pickling cucumber, 2007. PDMR 3: V112.

9. Holmes GJ, Thomas CE (2006) The history and re-emergence of cucurbit downy mildew. Phytopathology 99: S171.

10. Holmes G, Wehner T, Thornton A (2006) An old enemy re-emerges. American Vegetable Grower Feb.: 14-15.

11. Choi YJ, Hong SB, Shin HD (2005) A re-consideration of *Pseudoperonospora cubensis* and *P. humuli* based on molecular and morphological data. Mycol Res 109: 841-848.

12. Waterhouse GM, Brothers MP (1981) The taxonomy of *Pseudoperonospora*. Mycological Papers 148: 1-18.

13. Berkeley MS, Curtis A (1868) *Peronospora cubensis*. J Linn Soc Bot 10: 363.

14. Rostovzev SI (1903) Beitrage zur Kenntnis der Peronosporeen. Flora 92: 405-430.

15. Mitchell MN, Ocamb C, Gent D (2009) Addressing the relationship between *Pseudoperonospora cubensis* and *Pseudoperonospora humuli* by multigenic characterization and host specificity. Phytopathology 99: S87.

16. Gent DH, Mitchell MN, Holmes GJ (2009) Genetic and pathogenic relatedness of *Pseudoperonospora cubensis* and *P. humuli:* Implications fo detection and management. Phytopathology 99: S171.

17. Runge F, Thines M (In Press) Host matrix has major impact on the morphology of *Pseudoperonospora cubensis*. Eur J Plant Pathol: 1-10.

18. Iwata Y (1942) Specialization in *Pseudoperonospora cubensis* (Berk. et Curt.) Rostov. II. Comparative studies of the morphologies of the fungi from *Cucumis sativus* L. and *Cucurbita moschata* Duchesne. Ann Phytopathol Soc Japan 11: 172-185.

19. Oerke EC, Steiner U, Dehne HW, Lindenthal M (2006) Thermal imaging of cucumber leaves affected by downy mildew and environmental conditions. J Exp Bot 57: 2121-2132.

20. Cohen Y, Rotem J (1971) Rate of lesion development in relation to sporulating potential of *Pseudoperonospora cubensis* in cucumbers. Phytopathology 61: 265-268.

21. Lindenthal M, Steiner U, Dehne HW, Oerke EC (2005) Effect of downy mildew development on transpiration of cucumber leaves visualized by digital infrared thermography. Phytopathology 95: 233-240.

22. Cohen Y (1977) The combined effects of temperature, leaf wetness, and inoculum concentration on infection of cucumbers with *Pseudoperonospora cubensis*. Can J Bot 55: 1478-1487.

23. Keinath AP, Holmes GJ, Everts KL, Egel DS, Langston DB (2007) Evaluation of combinations of chlorothalonil with azoxystrobin, harpin, and disease forecasting for control of downy mildew and gummy stem blight on melon. Crop Protection 26: 83-88.

24. Rotem J, Cohen Y, Bashi E (1978) Host and environmental influences on sporulation in vivo. Ann Rev Phytopathol 16: 83-101.

25. Palti J (1975) *Pseudoperonospora cubensis* (Berk & M. A. Curtis) Rost. C M I Descr Pathog Fungi Bact 457: 1-2.

26. Cohen Y (1981) Downy mildew of cucurbits. In: Spencer DM, editor. The Downy Mildews. London: Academic Press. pp. 341-354.

27. Cohen Y, Rotem J (1971) Field and growth chamber approach to epidemiology of *Pseudoperonospora cubensis* on cucumbers. Phytopathology 61: 736-737.

28. Cohen Y (1977) Growth and differentiation of sporangia and sporangiophores of *Pseudoperonospora cubensis* on cucumber cotyledons under various combinations of light and temperature. Physiol Plant Pathol 10: 93-103.

29. Bains SS, Jhooty JS (1976) Over wintering of *Pseudoperonospora cubensis* causing downy mildew of muskmelon. Indian Phytopathol 29: 213-214.

30. Runge F, Thines M (2009) A potential perennial host for *Pseudoperonospora cubensis* in temperate regions. Eur J Plant Pathol 123: 483-486.

31. Lebeda A, Cohen Y (In Press) Cucurbit downy mildew (*Pseudoperonospora cubensis*)—biology, ecology, epidemiology, host-pathogen interaction and control. Eur J Plant Pathol: 1-36.

32. Bains SS, Sokhi SS, Jhooty JS (1977) *Melothria maderaspatana* - a new host of *Pseudoperonospora cubensis*. Indian J Mycol Plant Pathol 7: 86.

33. Hiura M, Kawada S (1933) On the overwintering of *Peronoplasmopara cubensis* Jap J Bot 6: 507-513.

34. D'Ercole N (1975) La peronoospora del cetriolo in coltura protetta. Inftor Fitopath 25: 11-13.

35. Bedlan G (1989) First detection of oospores of *Pseudoperonospora cubensis* (Berk et Curt.) Rost. on glasshousse cucumbers in Austria. Pflantzenschutzberichte 50: 119-120.

36. Cohen Y, Meron I, Mor N, Zuriel S (2003) A new pathotype of *Pseudoperonospora cubensis* causing downy mildew in cucurbits in Israel. Phytoparasitica 31: 458-466.

37. Lange L, Eden U, Olson LW (1989) Zoosporogenesis in *Pseudoperonospora cubensis*, the causal agent of cucurbit downy mildew. Nord J Bot 8: 497-504.

38. Aylor DE (1990) The role of intermittent wind in the dispersal of fungal pathogens. Ann Rev Phytopathol 28: 73-92.

39. Aylor DE, Taylor GS (1982) Aerial dispersal and drying of *Peronospora tabacina* conidia in tobacco shade tents. Proc Natl Acad Sci USA 79: 697-700.

40. Wu BM, van Bruggen AHC, Subbarao KV, Pennings GGH (2001) Spatial analysis of lettuce downy mildew using geostatistics and geographic information systems. Phytopathology 91: 134-142.

41. Doran WL (1932) Downy mildew of cucumbers. Massachusetts Agr Exp Sta Bull 283.

42. Van Haltern F (1933) Spraying cantaloupes for the control of downy mildew and other diseases. Georgia Exp Sta Bull.

43. Holmes GJ, Main CE, Keever ZT (2006) Cucurbit downy mildew: a unique pathosystem for disease forecasting. In: Spencer-Phillips PTN, Jeger M, editors. Advances in downy mildew research. Dordrecht, The Netherlands: Kluwer academic publishers.

44. Lebeda A, Schwinn FJ (1994) The downy mildews - an overview of recent research progress. J Plant Dis Protection 101: 225-254.

45. Cohen Y, Rotem J (1971) Dispersal and viability of sporangia of *Pseudoperonospora cubensis*. Trans Br Mycol Soc 57: 67-74.

46. Kanetis L, Holmes GJ, Ojiambo PS (2009) Survival of *Pseudoperonospora cubensis* sporangia exposed to solar radiation. Plant Pathology 59: 313-323.

47. Iwata Y (1949) Studies on the invasion of cucumber plants by downy mildew. (In Japanese, with English summary). Ann Phytopathol Soc Japan 13: 60-61.

48. Voglmayr H, Riethmuller A, Goker M, Weiss M, Oberwinkler F (2004) Phylogenetic relationships of *Plasmopara, Bremia* and other genera of downy mildew pathogens with pyriform haustoria based on bayesian analysis of partial LSU rDNA sequence data. Mycol Res 108: 1011-1024.

49. Fraymouth J (1956) Haustoria of the Peronosporales. Trans Br Mycol Soc 39: 79-107.

50. Hahn M, Mendgen K (2001) Signal and nutrient exchange at biotrophic plant-fungus interfaces. Curr Opin Plant Biol 4: 322-327.

51. Whisson SC, Boevink PC, Moleleki L, Avrova AO, Morales JG, et al. (2007) A translocation signal for delivery of oomycete effector proteins into host plant cells. Nature 450: 115-118

52. Lebeda A, Widrlechner MP (2003) A set of Cucurbitaceae taxa for differentiation of *Pseudoperonospora cubensis* pathotypes. J Plant Dis Protection 110: 337-349.

53. Lebeda A, Gadasova V (2002) Pathogenic variation of *Pseudoperonospora cubensis* in the Czech Republic and some other European countries. Acta Hort 588: 137-141.

54. Sarris P, Abdelhalim M, Kitner M, Skandalis N, Panopoulos N, et al. (2009) Molecular polymorphisms between populations of *Pseudoperonospora cubensis* from Greece and the Czech Republic and the phytopathological and phylogenetic implications. Plant Pathology Doi: 10.111/j.1365-3059.2009.02093.x.

55. Thomas CE, Inaba T, Cohen Y (1987) Physiological specialization in *Pseudoperonospora cubensis*. Phytopathology 77: 1621-1624.

56. Schornack S, Huitema E, Cano LM, Bozkurt TO, Oliva R, et al. (2009) Ten things to know about oomycete effectors. Molecular Plant Pathology 10: 795-803.

57. Hogenhout SA, Van der Hoorn RAL, Terauchi R, Kamoun S (2009) Emerging concepts in effector biology of plant-associated organisms. Mol Plant-Microbe Interact 22: 115-122.

58. Oliva R, Win J, Raffaele S, Boutemy L, Bozkurt TO, et al. Recent developments in effector biology of filamentous plant pathogens. Cellular Microbiology 12: 705-715.

59. Thines M, Kamoun S (2010) Oomycete-plant coevolution: recent advances and future prospects. Curr Opin Plant Biol 13: 427-433.

60. Win J, Morgan W, Bos J, Krasileva K, Cano L, et al. (2007) Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. The Plant Cell 19: 2349-2369.

61. Tian M, Win J, Savory EA, Kamoun S, Day B (2011) 454 genome sequencing of *Pseudoperonospora cubensis* reveals effector proteins with a putative QXLR translocation motif. Mol Plant-Microbe Interact 24.

62. Savory EA, Tian M, Win J, Kamoun S, Day B (2009) Genome characterizarion and discovery of novel QXLR effector motif in the cucurbit downy mildew pathogen *Pseudoperonospora cubensis* 14th International ISM-MPMI Congress. Quebec, Canada.

63. Thomas CE, Jourdain EL (1992) Host effect on selection of virulence factors affecting sporulation by *Pseudoperonospora cubensis*. Plant Dis 76: 905-907.

64. Lebeda A (1986) Epidemic occurrence of *Pseudoperonospora cubensis* in Czechoslovakia. Temperate Downy Mildews Newsletter 4: 15-17.

65. Lebeda A, Schwinn FJ (1994) The downy mildews--an overview of recent research progress. J Plant Dis Protection 101: 225-254.

66. Forsberg AS (1986) Downy mildew-*Pseudoperonospora cubensis* in Swedish cucumber fields. Vaxtskyddsnotiser 50: 17-19.

67. Tahvonen R (1985) Downy mildew of cucurbits found for the first time in Finland. Vaxtskyddsnotiser 49: 42-44.

68. Chaban VS, Okhrimchuk VN, Sergienko VG (2000) Optimization of chemical control of *Pseudoperonospora cubensis* on cucumber in Ukraine. EPPO Bulletin 30: 213-215.

69. Urban J, Lebeda A (2006) Fungicide resistance in cucurbit downy mildew-methodological, biological, and population aspects. Annals of applied biology 149: 63-75.

70. Gevens AJ, Hausbeck MK (2006) Control of downy mildew of cucumbers with fungicides, 2005. F & N Tests: V062.

71. Hausbeck MK. Downy mildew and *Phytophthora* control in vine crops; 2009; Syracuse, NY. pp. 193-195.

72. Reuveni M, Eyal H, Cohen Y (1980) Development of resistance to Metalaxyl in *Pseudoperonospora cubensis* Plant Disease 64: 1108-1108.

73. Heaney SP, Hall AA, Davis SA, Olaya G. Resistance to fungicides in the QoI-STAR cross-resistance group: current perspectives; 2000. pp. 755-762.

74. Keinath AP, DuBose VB, Lassiter AW (2008) Evaluation of fungicides to manage downy mildew on pickling cucumber in Charleston, South Carolina. Plant Disease Management Reports 2: V024.

75. Russell PE (2004) Sensitivity baselines in fungicide resistance research and management. FRAC Monograph 3.

76. Gisi U (2002) Chemical control of downy mildews. In: Spencer-Phillips PTN, Gisi U, Lebeda A, editors. Advances in downy mildew research. Dordrecht, The Netherlands: Kluwer academic publishers. pp. 119-159.

77. Barnes WC, Epps WM (1954) An unreported type of resistance to cucumber downy mildew. Plant Disease Rep 38: 620.

78. Wehner TC, Shetty NV (1997) Downy mildew resistance of the cucumber germplasm collection in North Carolina field tests. Crop Science 37: 1331-1340.

79. Shetty NV, Wehner TC, Thomas CE, Doruchowski RW, Vasanth Shetty KP (2002) Evidence for downy mildew races in cucumber tested in Asia, Europe, and North America. Scientia Horticulturae 94: 231-239.

80. Taler D, Galperin M, Benjamin I, Cohen Y, Kenigsbuch D (2004) Plant eR genes that encode photorespiratory enzymes confer resistance against disease. Plant Cell 16: 172-184.

CHAPTER 2

Alternative splicing of a multi-drug transporter from *Pseudoperonospora cubensis* generates an RXLR effector protein that elicits a rapid cell death

This chapter was originally published in PLoS ONE.

Savory EA, Zou C, Adhikari BN, Hamilton JP, Buell CR, Shiu S-H, and Day B (2012) Alternative splicing of a multi-drug transporter from *Pseudoperonospora cubensis*. PLoS ONE 7(4): e34701. doi:10.1371/journal.pone.0034701.

Author Contributions:

Conceived and designed the experiments: EAS, S-HS, CRB, and BD. Performed the experiments: EAS, CZ, and BNA. Analyzed the data: EAS, CZ, BNA, JPH, S-HS, CRB, and BD. Contributed reagents/materials/analysis tools: EAS, CZ, BNA, and JPH. Wrote the paper: EAS and BD.

ABSTRACT

Pseudoperonospora cubensis, an obligate oomycete pathogen, is the causal agent of cucurbit downy mildew, a foliar disease of global economic importance. Similar to other oomycete plant pathogens, *Ps. cubensis* has a suite of RXLR and RXLR-like effector proteins, which likely function as virulence or avirulence determinants during the course of host infection. Using *in silico* analyses, we identified 271 candidate effector proteins within the *Ps. cubensis* genome with variable RXLR motifs. In extending this analysis, we present the functional characterization of one *Ps. cubensis* effector protein, RXLR protein 1 (*Psc*RXLR1), and its closest *Phytophthora infestans* ortholog, PITG_17484, a member of the Drug/Metabolite Transporter (DMT) superfamily. To assess if such effector-non-effector pairs are common among oomycete plant pathogens, we examined the relationship(s) among putative ortholog pairs in *Ps. cubensis* and *P. infestans*. Of 271 predicted *Ps. cubensis* effector proteins, only 109 (41%) had a putative ortholog in *P. infestans* and evolutionary rate analysis of these orthologs shows that they are evolving significantly faster than most other genes. We found that *Psc*RXLR1 was up-regulated during the early stages of infection of plants, and moreover, that heterologous expression of *Psc*RXLR1 in *Nicotiana benthamiana* elicits a rapid necrosis. More interestingly, we also demonstrate that *Psc*RXLR1 arises as a product of alternative splicing, making this the first example of an alternative splicing event in plant pathogenic oomycetes transforming a non-effector gene to a functional effector protein. Taken together, these data suggest a role for *Psc*RXLR1 in pathogenicity, and in total, our data provide a basis for comparative analysis of

candidate effector proteins and their non-effector orthologs as a means of understanding function and evolutionary history of pathogen effectors.

INTRODUCTION

The identification and characterization of secreted effector proteins from plant pathogens has anchored the recent evolution of molecular plant pathology [1,2,3]. As components of many pathogenic microorganisms' secretomes, effector proteins represent a key component of phytopathogenicity, contributing to both the virulence and avirulence capacity of the invading pathogen. Numerous studies have identified and characterized the activities of secreted effector proteins from a broad range of phytopathogens [reviewed in 1,4,5]. Collectively, these works have revealed two primary functions for pathogen effector molecules. First, as virulence molecules, effector proteins can enhance a pathogen's ability to cause disease, likely through abrogating host processes that would otherwise block pathogen infection, growth, and proliferation within the host [5,6]. Secondly, as avirulence determinants, effector proteins function to modulate the activation of host defense responses by perturbing the activity of host resistance (R) proteins [5,6].

For infection, colonization, and subsequent propagation within their hosts to occur, pathogens must dampen multiple layers of plant defense responses. Often described as basal resistance, the initial perception and elicitation of defenses requires the recognition of pathogen associated molecular patterns (PAMPs; e.g., chitin, flagellin,

LPS) [5,6], highly conserved molecules essential for the lifestyle and survival of the microorganism. The recognition of PAMPs, which are highly specific elicitors, occurs through receptors on the host membrane surface, and following initiation of this receptor-ligand interaction, a rapid first response known as PAMP-triggered immunity (PTI) is elicited [5,6]. Overall, the PTI response provides a basal level of resistance against a wide range of microorganisms, often utilizing conserved signaling pathways such as the up-regulation of the mitogen-activated protein kinase (MAPK) pathway, the generation of reactive oxygen species, and the induction of defense-related genes [5,6]. To overcome PTI, phytopathogens, including bacteria and oomycetes, rely on the delivery and activity of secreted effector proteins to abrogate this initial basal level of defense, as well as to further promote virulence [3,5,6]. In response, pathogen effectors can be recognized by R (resistance) proteins, leading to the activation of effector-triggered immunity (ETI) [5,6] best illustrated as an amplified and sustained layer of defense. ETI is a robust response that is often associated with the activation of a specific type of programmed cell death referred to as the hypersensitive response (HR) [5,6]. Over time, as this cycle of subversion and recognition evolves, host specificity and subsequent interactions between pathogen and host are modulated by the interplay between the activity and recognition of secreted pathogen effector molecules and their host counterparts.

Oomycetes are a phylogenetically distinct eukaryotic lineage within the Stramenopiles, which as a group, are among the best-studied and most economically important plant pathogens. In recent years, the genomes of several agriculturally important oomycete

pathogens have been sequenced, including *Phytophthora infestans*, *Phytophthora ramorum*, and *Phytophthora sojae*, the causal agents of late blight of potato and tomato, sudden oak death, and soybean root rot, respectively [7,8]. The genomes of two other oomycete pathogens, *Pythium ultimum*, which causes damping off and root rot on a wide range of hosts, and *Hyaloperonospora arabidopsidis*, a pathogen of *Arabidopsis thaliana,* have also been sequenced [9,10]. These investigations, through the analysis of genome content and structure, have provided a wealth of information, both towards understanding the nature of the host-pathogen interaction (e.g., host specificity, virulence strategies), as well as insight into the evolution of the interaction itself. Central to the analysis of phytopathogen genomes, the identification and characterization of oomycete effector proteins has moved swiftly into the forefront in the field of plant-pathogen interactions, due in large part to the aforementioned available genomic resources. At a primary level, the identification of a highly conserved N-terminal translocation motif (i.e., RXLR; *Arg-X-Leu-Arg*, where "X" is any amino acid) demonstrated to be necessary for effector delivery into host cells, has been a seminal discovery in the field of plant-oomycete interactions [11,12]. Similar in function to phytopathogenic bacterial effector proteins, oomycete RXLR-containing effectors have been demonstrated to suppress PTI [13], as well as to activate ETI [11,14,15,16,17,18]. Structurally, oomycete effector proteins display a modular organization, consisting of a N-terminal signal peptide, a conserved RXLR translocation motif, followed by a variable C-terminal effector domain [3]. It is the function and activity of the variable C-terminal effector domain that drives the activity of these molecules [3,4].

Alternative splicing (AS) of pre-mRNA drives the generation of multiple protein isoforms through assembly of different combinations of splice sites within a single gene. In total, this process represents a conserved mechanism found in eukaryotes which drives proteome complexity within organisms with a finite number of genes [19]. In oomycetes, there are few reports of intron processing [20,21,22], and to date, these analyses has been strictly *in silico* [20,22], with little functional validation [21]. Costanzo et al. [21] characterized alternative processing in *P. sojae* family 5 endoglucanases revealing the generation of both coding and non-coding RNA isoforms. Additionally, based on their large-scale analysis of intronic structure and alternative splicing in *P. sojae*, Shen and colleagues [22] validated splice variants leading to premature translation termination.

Ps. cubensis is an obligate biotrophic oomycete pathogen of cucurbits (i.e., cucumber, melon, squash, watermelon, etc.), and is the causal agent of cucurbit downy mildew, an economically important foliar disease [23]. Capable of rapid defoliation of fields in short periods of time (i.e., <10 days), *Ps. cubensis* is the primary factor limiting cucurbit production in the United States. Despite obvious economic importance, very little is known about the genetic determinants of virulence and pathogenicity of *Ps. cubensis*, as well as the molecular-genetic basis of resistance in the cucurbits.

Similar to related oomycete pathogens of plants, *Ps. cubensis* possesses a suite of effector proteins that likely function to promote virulence and suppress host defense responses [3,24]. Recent work by Tian et al. [24] identified and characterized a preliminary set of effector proteins from a draft genome sequence of *Ps. cubensis*

obtained using 454 pyrosequencing. In brief, this set of 61 candidate effectors included a large class of variants with sequence similarity to the canonical RXLR motif found in other oomycete plant pathogens [24]. Specifically, this work characterized the function of a QXLR-containing effector, designated *Pc*QNE, which was shown to be a member of a large family of *Ps. cubensis* QXLR nuclear-localized effectors, up-regulated during infection of cucumber. Additionally, internalization of *Pc*QNE was shown to require the QXLR-EER motif, thereby establishing a basic homology to the well-characterized *Phytophthora* spp. effector proteins.

In the current study, we describe the identification and evolutionary potential of the *Ps. cubensis* effector repertoire. First, through characterization of a RXLR effector protein, *Psc*RXLR1, we investigated the localization and *in planta* activity, and similarly to some oomycete effector proteins described to date, *Psc*RXLR1 induces a rapid cell death response when delivered into plant cells. Additionally, using whole transcriptome sequencing analyses, as well as RT-PCR, we show that *Psc*RXLR1 is a product of alternative splicing of the *Psc*_781.4 gene which encodes a putative multi-drug transporter. Coupled with the induction and expression of *Psc*RXLR1 mRNA during *Ps. cubensis* infection of cucumber, as well as a complement of bioinformatic, cell biology and *in vivo* analyses, we provide evidence suggestive of a virulence role for *Psc*RXLR1. Finally, we used *Psc*RXLR1 as template for assessing the conservation and evolutionary potential of oomycete effector proteins from *Ps. cubensis*, identifying and analyzing orthologous pairs of *Ps. cubensis* effector proteins and *P. infestans* non-effector proteins. Using more robust methods, we identified additional candidate

effectors from *Ps. cubensis* for these analyses and showed that, like other oomycete effectors, they tend to be influenced by positive selection. Assessment of evolutionary rate and conservation of secretion signals between orthologous pairs revealed that *Ps. cubensis* effectors are undergoing adaptive evolution and conservation of signal peptides are similar among effector and non-effector proteins in *Ps. cubensis*. Overall, our study provides support for the investigation of relationships among oomycete effectors and their non-effector orthologs, and in total, the analysis presented herein establishes a foundation for understanding the evolution of effector repertoires and host-pathogen specificity.

RESULTS

Genome sequencing of *Ps. cubensis*

Next generation sequencing with the Illumina Genome Analyzer II platform was used to generate an assembly of the *Ps. cubensis* MSU-1 genome. A total of 4.5 Gb of cleaned paired end reads from two libraries were used to generate the assembly using Velvet, a *de novo* short read assembler [25]. The final assembly contains 35,546 contigs with an N50 contig size of 4.0 Kbp representing 64.4 Mbp. Protein coding genes in the draft assembly were annotated using MAKER [26] which incorporated *ab initio* gene predictions, protein evidence, and transcript evidence from other sequenced oomycete genomes. In total, 23,519 loci and 23,522 gene models were predicted.

Identification of the *Ps. cubensis* effector repertoire

Our initial analysis of the effector complement of *Ps. cubensis* in an earlier draft assembly [24] identified 61 sequences containing the conserved RXLR, or novel QXLR, motif found in known oomycete effector proteins. This number is significantly less than the effector count predicted for other plant pathogenic oomycetes (i.e., 563 effectors in *P. infestans*, 396 in *P. sojae*, 374 in *P. ramorum*, and 134 in *H. arabidopsidis*; [7,10,27]), and is likely the result of limited coverage generated from an initial 454 pyrosequencing [24]. Generation of genomic sequences using the Illumina Genome Analyzer platform and their subsequent assembly generated a more comprehensive dataset. Using this resource, 269 additional sequences were identified as putative effector proteins. Interestingly, the putative *Ps. cubensis* effectors showed more variation at the R1 position of the RXLR motif than previously shown [24], with 18 amino acids predicted at the R1 position, in addition to R and Q (Table S2.1; Available at www.plosone.org, e34701). Moreover, we have evidence for expression for at least one predicted effector with any one of 19 amino acids (except Y, *Tyr*) at position R1, during the course of infection on a susceptible cucumber cultivar [28], supporting the hypothesis of an expanded translocation motif repertoire in *Ps. cubensis*. In total, including the previously characterized *Pc*QNE, the current predicted effector complement of *Ps. cubensis* contains 271 members.

Nature of selection on *Ps. cubensis* paralogs

Based on comparative genomic analyses of several oomycete plant pathogens, positive selection has been postulated to act disproportionately on effectors gene[s] compared to other genes in the genome [27,29]. To this end, we examined the strength of selection acting upon the predicted effector complement of *Ps. cubensis* by estimating ω, the ratio of the non-synonymous substitution rate (*Ka*) to the synonymous substitution rate (*Ks*). Among all *Ps. cubensis* effector paralogs, the median ω is 0.54, which is significantly higher than that of *Ps. cubensis* paralogous genes in general (ω = 0.24, Wilcox Rank Sum Test, $p < 2.2e-16$). For comparison, we also examined *P. infestans* effectors and arrived at the same conclusion (Wilcox Rank Sum Test, $p < 4.0e-14$). Because more recent duplicates tend to have elevated ω, we examined if the higher ω values among effector paralogs can be attributed to recent duplication. We found that, using *Ks* as a proxy of time, the ω values for effector pairs are in general significantly higher than other paralogs in *Ps. cubensis* in a *Ks* bin (Figure 2.1A). Thus, the elevated ω values among effectors are not exclusively due to relaxation of selection among recent duplicates. The results for *P. infestans* effectors are similar (Figure 2.1B), although the ω values of *P. infestans* are in general higher than those in *Ps. cubensis*.

Taken together, *Ps. cubensis* effectors either have experienced a significantly lower degree of selective constraints, or tend to be positively selected. Consistent with the latter, 6.3% of *Ps. cubensis* effector paralog pairs have ω > 1, compared to 3.2% of all other paralogous gene pairs. In parallel to our observations in *Ps. cubensis*, 4.6% of *P.*

Figure 2.1

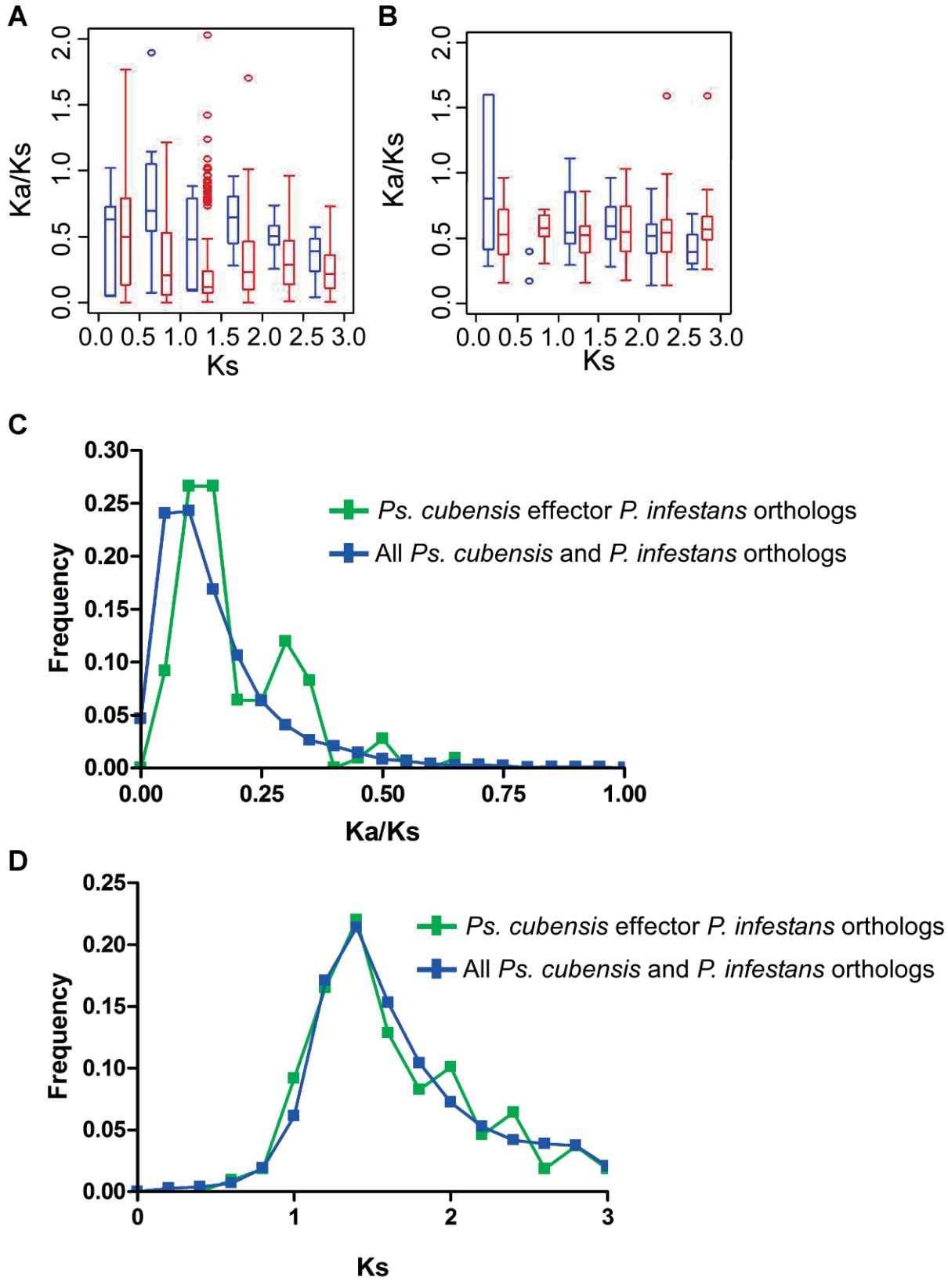
A
Ka/Ks
0.0
0.5
1.0
1.5
2.0
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Ks
B
Ka/Ks
0.0
0.5
1.0
1.5
2.0
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Ks
C
Frequency
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.00 0.25 0.50 0.75 1.00
Ka/Ks
Ps. cubensis effector P. infestans orthologs
All Ps. cubensis and P. infestans orthologs
D
Frequency
0.00
0.05
0.10
0.15
0.20
0.25
0 1 2 3
Ks
Ps. cubensis effector P. infestans orthologs
All Ps. cubensis and P. infestans orthologs

Figure 2.1 (cont'd) Strength of purifying selection on *Pseudoperonospora cubensis* effectors. Comparison of selective constraints on effector paralogs (blue) and all other genes (red) in (A) *Ps. cubensis* and (B) *P. infestans*. Frequency distributions (C) of *ω*, the ratio between the nonsynonymous substitution rate (*Ka*) and the synonymous substitution rate (*Ks*) of *Ps. cubensis* and *P. infestans* sequence pairs. Distributions of *Ks* values (D) of *Ps. cubensis* and *P. infestans* sequence pairs. Green symbols indicate *Ps. cubensis* effector – *P. infestans* non-effector ortholog pairs. Blue symbols represent other orthologous gene pairs between the two species.

infestans effector paralog pairs have ω > 1, compared to 3.2% for all other paralogs. Although there is no clear evidence suggesting that most effectors are subjected to positive selection, it is interesting that even among relatively ancient effector duplicates, the rate of evolution among effectors is significantly higher than most genes. Given that older duplicates that survive for tens to hundreds of millions of years tend to be subjected to substantially stronger selective constraints than young duplicates [30], this would suggest that, perhaps, effectors function in a way that do not require as strong a constraint on their primary sequence. Alternatively, it is also possible that pathogen effectors, even those having undergone ancient duplication events, experience some degree of continuous positive selection.

Relationship between *Ps. cubensis* effectors and their *P. infestans* orthologs

Subsequent *in silico* analysis of candidate *Ps. cubensis* effectors and comparisons to annotated genes in *P. infestans* revealed that there were a number of orthologs between *Ps. cubensis* effector proteins and both effector and non-effector genes in *P. infestans.* We hypothesized that this scenario (i.e., effector with non-effector ortholog)

may provide a foundation for the analysis of the evolution of effectors from non-effector proteins. Therefore, we identified orthologous pairs of predicted *Ps. cubensis* effector proteins and their non-effector counterparts in *P. infestans* considering sequence similarity and synteny (see Materials and Methods). With this approach, 11,601 orthologous gene pairs were identified between *Ps. cubensis* and *P. infestans* for comparison. Of 271 *Ps. cubensis* effector sequences, 109 had a predicted ortholog in *P. infestans* (TABLE S2.2; Available at www.plosone.org, e34701). As shown in Figure 2.1C, the *Ps. cubensis* effector *P. infestans* (*Psc*E-*Pi*) ortholog pairs have significantly higher ω values as compared to the baseline pairs (Kolmogorov-Smirnov test, $p < 7.9e-06$), consistent with what was found with the effector paralogs (Figure 2.1A). Additionally, the distribution of ω for the *Psc*E-*Pi* pairs appears multi-modal. Given that the first effector ortholog peak (at ω ~0.15) is mostly overlapping with that of the other orthologs, these effector paralogs are more highly conserved. The second peak at ω ~0.3 likely indicates the presence of a group of effectors that are more quickly evolving (Figure 2.1C). However, we cannot rule out the possibility that these peaks are present simply due to the small effector ortholog sample size.

To determine if the overall higher ω value among effector orthologs is an artifact due to mis-identification of orthologous genes, we examined if putative effectors, as well as the other orthologs, have similar "age". As shown in Figure 2.1D, the distributions of *Ks* values for the effector and the other orthologs are highly similar and are statistically indistinguishable. Thus, mis-identified orthologous pairs likely do not significantly impact our findings.

Signal peptide conservation among ortholog pairs

Signal peptides are essential components of oomycete effector proteins, as they are required for translocation of the protein from the pathogen haustorium to the extrahaustorial matrix prior to uptake by the host cell membrane [3]. As such, all 109 of the *Ps. cubensis* effector sequences in the *Psc*E-*Pi* dataset are predicted to have signal peptides (Figure S2.1). However, only 71 (65%) of the corresponding *P. infestans* orthologs were predicted to be secreted proteins. For comparison, predictions of signal peptides were made for 10,383 of the 11,601 ortholog pairs. Of these, there were 688 (6.63%; *Psc*-sec/*Pi*-sec) ortholog pairs where both members were predicted to have signal peptides, 428 (4.12%; *Psc*-sec/*Pi*-non) pairs where the *Ps. cubensis* protein was predicted to be secreted and the *P. infestans* ortholog was not, and 622 (5.99%; *Psc*-non/*Pi*-sec) where the *Ps. cubensis* sequence did not have a predicted signal peptide and its corresponding *P. infestans* sequence was predicted to be secreted. Additionally, there were 8,645 (83.3%; *Psc*-non/*Pi*-non) ortholog pairs where neither member was predicted to be secreted. For statistical analysis, the *Psc*-sec/*Pi*-sec and *Psc*-sec/*Pi*-non datasets from the *Ps. cubensis* effector-*P. infestans* orthologs were compared to their respective genome-wide datasets. Using the Fisher's exact test, no significant difference ($p = 0.5354$) was identified between the two datasets, indicating that presence of signal peptide prediction is not a good indicator of potential selection for effector peptide evolution.

Identification of *Pseudoperonospora cubensis* effector *Psc*RXLR1

Using the RXLR effector identification pipeline [29], we previously identified 61 candidate effector protein sequences from a draft genome assembly of *Ps. cubensis* [24]. Initial analysis of these sequences using the Basic Alignment Analysis Search Tool (BLAST) against the proteome of *P. infestans* revealed that only 7 of these sequences aligned with annotated proteins within the *P. infestans* genome database; moreover, only 1 of these was predicted to be a secreted RXLR effector. Of these sequences, one (contig01871_F1) had 75% amino acid identity to *P. infestans* protein PITG_17484, a putative member of the drug/metabolite transporter (DMT) superfamily (CLO184; 2.2). Additional cloning via 3' RACE PCR and subsequent analysis revealed that the *Ps. cubensis* coding sequence, hereafter referred to as *Ps. cubensis* RXLR protein 1 (*Psc*RXLR1), appeared significantly shorter (i.e., 132 amino acids), compared to its *P. infestans* ortholog PITG_1784 (i.e., 332 amino acids). This apparent truncation in *Psc*RXLR1 results in a protein coding sequence lacking the EamA functional domain (PF00892; formerly called DUF6) found in members of the DMT family [31].

SignalP analysis of *Psc*RxLR1 identified a putative 26 amino acid signal peptide at the N-terminus of the protein (Figure 2.2). Based on the conserved features and domain organization of oomycete effectors, the presence of a canonical RXLR motif was identified at amino acid position 70 (Figure 2.2). However, unlike several previously characterized oomycete effector proteins, *Psc*RXLR1 does not contain an EER motif, which has also been implicated in oomycete effector translocation into the host cell

[12,24,27]. The *P. infestans* ortholog, PITG_17848, while not having a predicted signal peptide, does contain an RXLR-like motif (i.e., RFMR; Figure 2.2A). To eliminate the possibility that PITG_17484 was mis-annotated and did in fact contain a signal peptide upstream of the predicted coding sequence, the region 500 bp upstream of the annotated PITG_17484 sequence was examined and a canonical signal peptide sequence was not identified. We therefore concluded that PITG_17484 is not an RXLR effector protein.

The absence of a predicted signal peptide in PITG_17848 suggests that *Psc*RXLR1 may have evolved this function independently. To address this possibility, and to further explore the ancestral function of these proteins, orthologous *Psc*RXLR1 sequences in additional plant oomycete pathogen species were identified, including *P. sojae*, *P. ramorum*, and *Py. ultimum*. Not surprisingly, the sequences from *P. sojae* and *P. ramorum* were more similar to *Psc*RXLR1 than those from *Py. ultimum* (76% and 72%, respectively, compared to 59%; Figure S2.2). Additionally, while none of these orthologs had predicted signal peptides, they did contain EamA functional domains, indicating that they were also members of the DMT superfamily (Figure S2.2). Phylogenetic analysis to infer evolutionary relationships between *Psc*RXLR1 and orthologs from *P. sojae*, *P. ramorum*, and *Py. ultimum* supported these observations (Figure 2.2C).

Figure 2.2

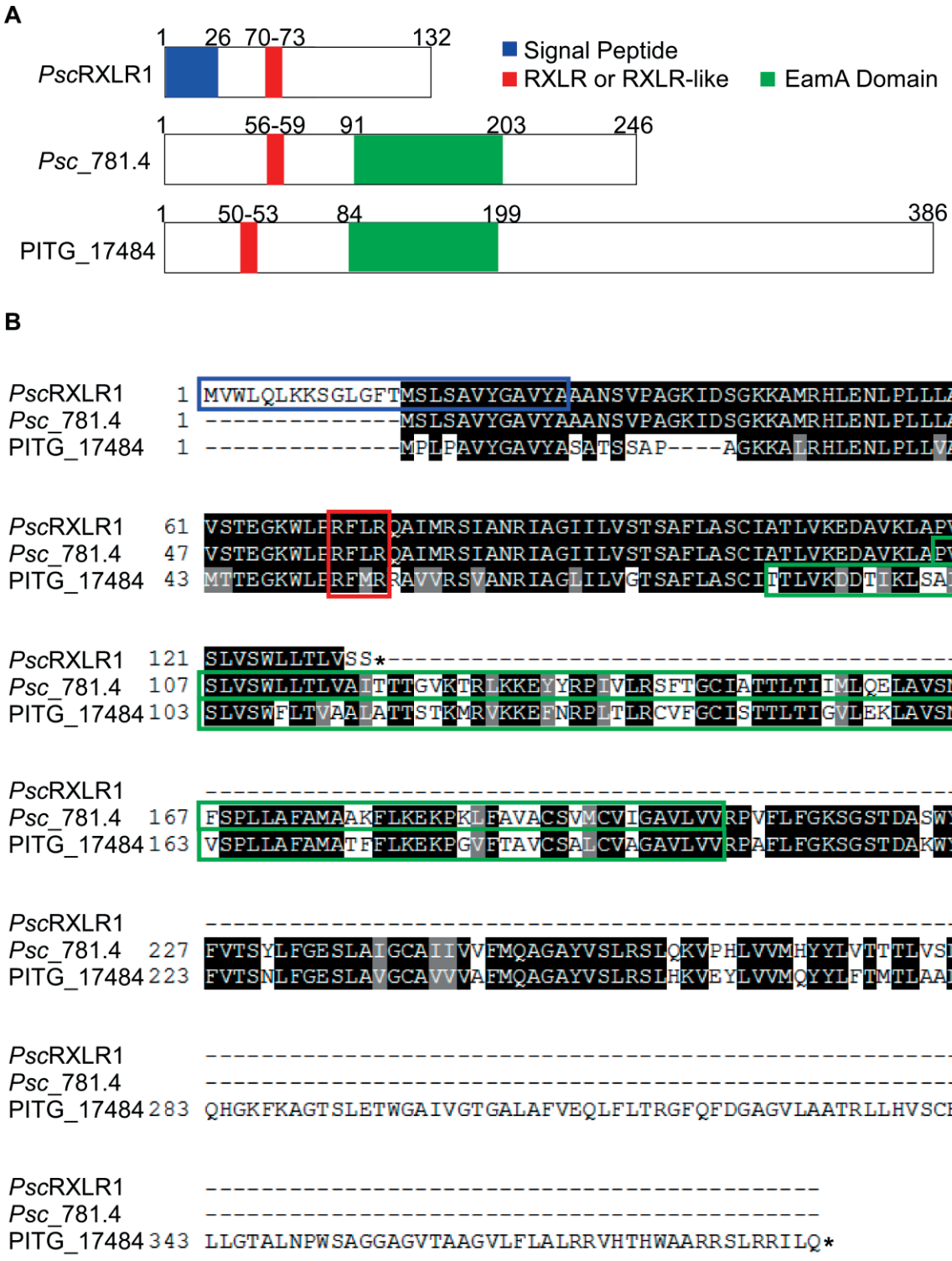
A
PscRXLR1
1
26
70-73
132
Signal Peptide
RXLR or RXLR-like
EamA Domain
Psc_781.4
1
56-59
91
203
246
PITG_17484
1
50-53
84
199
386
B
PscRXLR1 1 MVWLQLKKSGLGFTMSLSAVYGAVYAAANSVPAGKIDSGKKAMRHLENLPLLLASDSLES
Psc_781.4 1 ----------MSLSAVYGAVYAAANSVPAGKIDSGKKAMRHLENLPLLLASDSLES
PITG_17484 1 ----------MPLPAVYGAVYASATSSAP----AGKKALRHLENLPLLVASDSLES
PscRXLR1 61 VSTEGKWLERFLRQAIMRSIANRIAGIILVSTSAFLASCIATLVKEDAVKLAPVEILFWR
Psc_781.4 47 VSTEGKWLERFLRQAIMRSIANRIAGIILVSTSAFLASCIATLVKEDAVKLAPVEILFWR
PITG_17484 43 MTTEGKWLERFMRRAVVRSVANRIAGLILVGTSAFLASCITTLVKDDTIKLSAIEALFWR
PscRXLR1 121 SLVSWLLTLVSS*----------
Psc_781.4 107 SLVSWLLTLVAITTTGVKTRLKKEYYRPIVLRSFTGCIATTLTIIMLQELAVSNAIAITY
PITG_17484 103 SLVSWFLTVAALATTSTKMRVKKEFNRPLTLRCVFGCISTTLTIGVLEKLAVSNATAVTY
PscRXLR1 ----------
Psc_781.4 167 FSPLLAFAMAAKFLKEKPKLFAVACSVMCVIGAVLVVRPVFLFGKSGSTDASWYRRSMTS
PITG_17484 163 VSPLLAFAMATFFLKEKPGVFTAVCSALCVAGAVLVVRPAFLFGKSGSTDAKWYHRSMAS
PscRXLR1 ----------
Psc_781.4 227 FVTSYLFGESLAIGCAIIVVFMQAGAYVSLRSLQKVPHLVVMHYYLVTTTLVSLP*----
PITG_17484 223 FVTSNLFGESLAVGCAVVVAFMQAGAYVSLRSLHKVEYLVVMQYYLFTMTLAALAAMLGI
PscRXLR1 ----------
Psc_781.4 ----------
PITG_17484 283 QHGKFKAGTSLETWGAIVGTGALAFVEQLFLTRGFQFDGAGVLAATRLLHVSCEFAWGVI
PscRXLR1 ----------
Psc_781.4 ----------
PITG_17484 343 LLGTALNPWSAGGAGVTAAGVLFLALRRVHTHWAARRSLRRILQ*

Figure 2.2 (cont'd)

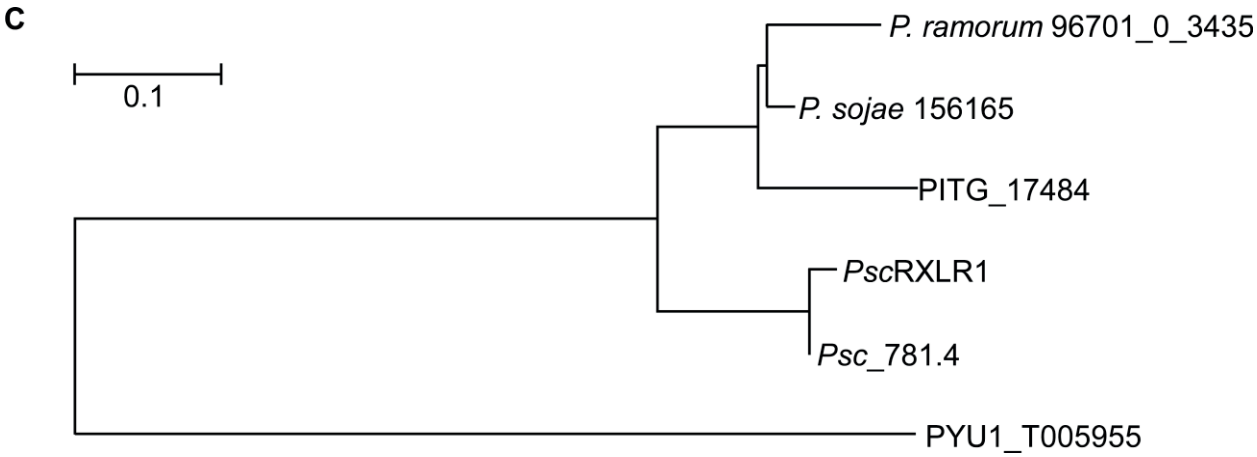

Figure 2.2 *Psc*RXLR1 encodes a RXLR-containing effector protein with homology to a non-effector protein in *Phytophthora infestans*. (A) Schematic representations of *Psc*RXLR1, *Psc*_781.4 and PITG_17484 from *P. infestans*. (B) Multiple sequence alignment of *Psc*RXLR1, *Psc*_781.4 and PITG_174984. Protein sequences were aligned using CLUSTALW and shaded for consensus using BOXSHADE. The asterisk indicates a stop codon. (C) The full length protein sequences of *Psc*RXLR1, *Psc*781.4 and their orthologs from *P. infestans* (PITG_17484), *Phytophthora ramorum* (*P.* ramorum 96701 0 3435), *Phytophthora sojae* (*P. sojae* 156165) and *Pythium ultimum* (PYU_T005955) were aligned using Muscle and evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model [1] using MEGA5[2]. 500 bootstrap runs were performed.

Functional validation of the *Psc*RXLR1 signal peptide

A primary characteristic of oomycete effector proteins is signal peptide-mediated secretion from the haustoria into the extrahaustorial matrix (EHM) prior to translocation into the host cell [3]. *Psc*RXLR1 contains a 26 amino acid signal peptide as predicted by SignalP 3.0 (HMM Probability, 0.966), whereas its closest *P. infestans* ortholog, PITG_17484, does not have a predicted signal peptide. To determine if the predicted

signal peptide from *Psc*RXLR1 is functional, we used a yeast signal trap assay based on the requirement of invertase secretion for yeast growth on media with raffinose as the sole carbon source [32]. This assay has been used previously to confirm predicted signal peptide sequences in candidate effector proteins from both *P. infestans* and *Ps. cubensis* [24,33]. As shown in Figure 2.3A, pSUC2-*Psc*RXLR1 (column 4) is able to grow on medium containing only raffinose (YPRAA), indicating that the signal peptide of *Psc*RXLR1 is sufficient for secretion of invertase. As a second confirmation of signal peptide function, we monitored the reduction of 2,3,5-triphenyltetrazolium chloride (TTC) to the red-colored compound triphenylformazan. Again, pSUC2-*Psc*RXLR1 (column 4) was confirmed as having a functional signal peptide. In contrast, neither the control yeast strains (i.e., YPK12 [column 1] or pSUC2 [column 2]), nor the pSUC2-PITG_17484 (column 3) yeast strain containing a PITG_17484-invertase fusion construct, grew on YPRAA, nor were they able to reduce TTC. Our positive control, PcQNE-SP1 (5) was, as shown previously [24], both able to grow on YPRAA medium and reduce TTC. These data support the annotation of *Psc*RXLR1 as a secreted RXLR effector protein and confirm that its *P. infestans* ortholog is a non-secreted protein.

*Psc*RXLR1 and PITG_1784 localize to the plant plasma membrane

In planta localization of effector proteins has been successfully used to guide functional analysis and to infer the function itself [34,35,36,37]. To identify a possible role for *Psc*RXLR1 in the pathogenicity of *Ps. cubensis*, and to provide clues as to its putative role *in planta*, we investigated the localization of *Psc*RXLR1. To this end, a C-terminal

Figure 2.3

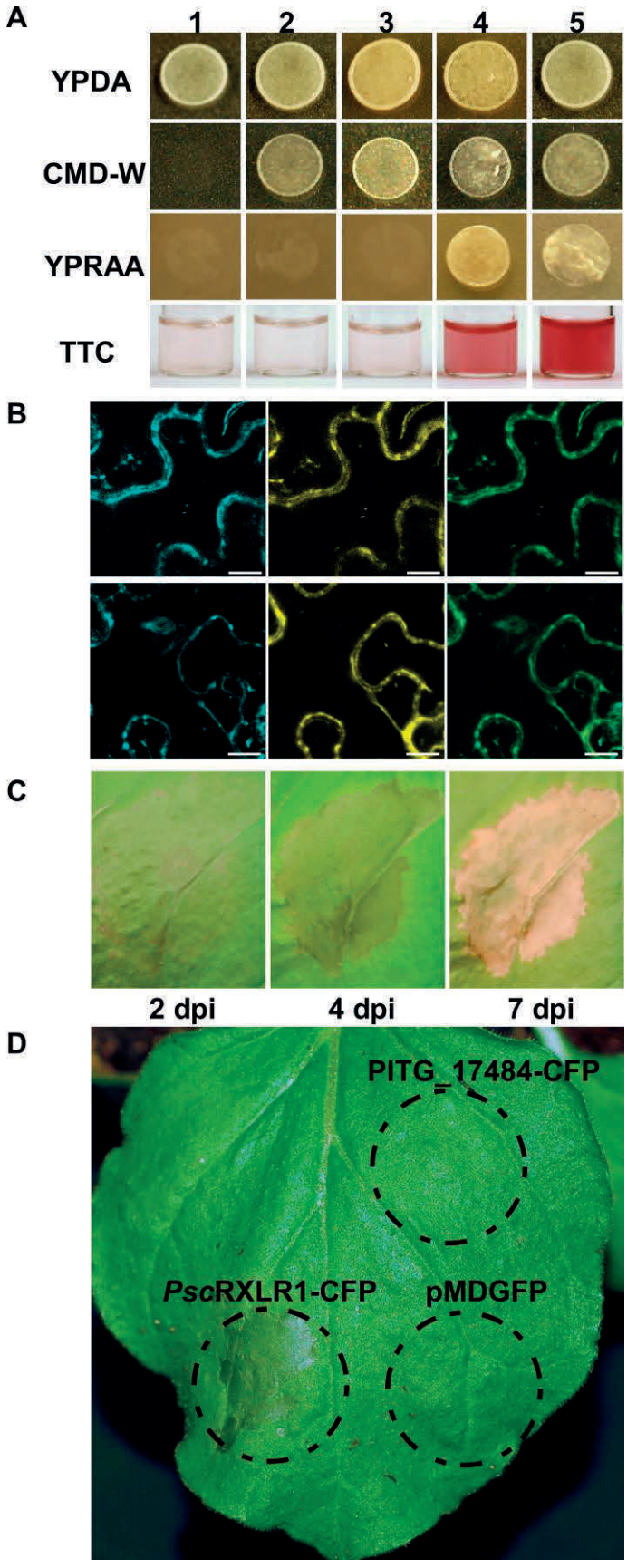
A
1
2
3
4
5
YPDA
CMD-W
YPRAA
TTC
B
C
2 dpi
4 dpi
7 dpi
D
PITG_17484-CFP
PscRXLR1-CFP
pMDGFP

Figure 2.3 (cont'd) Functional characterization of *Psc*RXLR1 and PITG_17484. (A) *Psc*RXLR1 has a functional signal peptide validated by the yeast signal trap assay. Yeast strains were grown on media with raffinose (YPRAA) as the sole carbon source or in the presence of 2,3,5-triphenyltetrazolium chloride (TTC). Yeast strains YTK12 and YTK12:pSUC2 EV both lack a functional invertase gene and cannot grow on YPRAA medium or reduce TTC to red formazan. The functional signal peptide of *Psc*RXLR1, when fused in-frame to the mature yeast invertase (pSUC2-*Psc*RXLR1), allows for secretion of invertase, resulting in growth on YPRAA medium, as well as reduction of TTC to red formazan. PITG_17484, as predicted, does not have a functional signal peptide (pSUC2-PITG_17484). (B) Both *Psc*RXLR1_CFP (top row) and PITG_17484_CFP (bottom row) co-localize with a plasma membrane-specific *At*PIP2A_YFP marker. Left panels: C-CFP fusion protein only. Center panels: *At*PIP2A_YFP. Right panels: Merge of CFP and YFP images. Scale bar = 10 µm. (C, D) Heterologous expression of *Psc*RXLR1 induces chlorosis and necrosis in *Nicotiana benthamiana*. *Agrobacterium tumefaciens* C58-C1 expressing *Psc*RXLR1_CFP, PITG_17484_CFP, or an *At*PIP2A-YFP construct were infiltrated into *N. benthamiana*. The chlorosis and necrosis phenotype of *Psc*RXLR1_CFP infiltrated leaves (C) was photographed at 2, 3, and 7 days post-inoculation (dpi). Leaf areas infiltrated with either PITG_17484_CFP or *At*PIP2A-YFP (D) as designated by the dash line circles did not shown any phenotype at 3 dpi.

CFP-fusion protein (i.e., *Psc*RXLR1-CFP) was transiently expressed in *Nicotiana benthamiana,* and protein localization visualized using laser scanning confocal microscopy. Based on the similarity of *Psc*RXLR1 with members of the DMT superfamily in *Phytophthora* spp. and *Py. ultimum*, *Psc*RXLR1 was predicted to localize to the plant plasma membrane, despite the absence of a predicted transmembrane domain when analyzed using InterProScan. To confirm this, *Psc*RXLR1-CFP was co-expressed with a YFP-tagged construct encoding the aquaporin gene *At*PIP2A, a marker for plasma membrane localization [38]. As predicted, *Psc*RXLR1-CFP co-localized with *At*PIP2A-YFP (Figure 2.3B), confirming that *Psc*RXLR1 localizes to the plasma membrane *in planta*. Additionally, a C-terminal CFP fusion was made to the *P.*

infestans ortholog PITG_17484, which was also confirmed to be plasma membrane localized (Figure 2.3B).

*Psc*RXLR1, but not its *P. infestans* ortholog, elicits a rapid cell death response when expressed in *Nicotiana benthamiana*

The obligate nature of a plant pathogen often presents challenges towards functional characterization of virulence and susceptibility within their respective host(s). To circumvent this limitation, transient heterologous systems have been developed and have proved valuable in their functional characterization [34,39,40,41,42]. To investigate the activity of *Psc*RXLR1 *in planta*, we utilized heterologous expression in *N. benthamiana* as means to characterize and determine the function of *Psc*RXLR1. As shown in Figure 2.3C, expression of *Psc*RxLR1 resulted in leaf chlorosis throughout the entire infiltration zone by 2 dpi, followed by browning and initiation of necrosis at 4 dpi. By 7 dpi, the zone of infiltration was completely dehydrated. In comparison, neither infiltration with PITG_17484, nor pMDGFP, resulted in any detectable cell death-type phenotype in *N. benthamiana* leaves at 4 dpi (Figure 2.3D).

*Psc*RXLR1 expression is induced during *Ps. cubensis* infection of cucumber

The function of pathogen effector molecules is to enhance the virulence of the pathogen during its lifecycle, as well as to dampen host defense responses activated during infection. In this regard, the temporal expression of effector molecules during infection

and pathogen development often signals critical stages in the host-pathogen interaction. Expression of *Psc*RXLR1 mRNA was measured using quantitative reverse transcription (qRT)-PCR following infection of *Ps. cubensis* on the susceptible cucumber cultivar 'Vlaspik'. As shown in Figure 2.4, expression of *Psc*RXLR1 was significantly ($p < 0.001$) induced during infection, beginning at 1 dpi and continuing through 4 dpi, as compared to the basal expression level in sporangia. Induction of gene expression at 1 dpi corresponds with zoospore encystment in the stomata, the first stage of pathogen entry into the host (Figure 2.4B, left panel). Subsequent expression observed through 4 dpi corresponds with hyphal penetration through the stomata, growth throughout the mesophyll, and formation of haustoria (Figure 2.4B, center and right panels). This pattern of expression supports a potential role for *Psc*RXLR1 in initial establishment of the infection possibly through dampening host defense responses. Additionally, this pattern is consistent with the expression patterns observed in other oomycete plant pathogen effectors, further supporting the prediction of *Psc*RXLR1 as an effector protein with a role in infection and disease development.

*Psc*RXLR1 is a splice variant of *Psc*_781.4

Automated annotation of the Illumina-generated *Ps. cubensis* assembly described in this study resulted in *Psc*_781.4, a gene model at the *Psc*RXLR1 locus that more closely mirrored PITG_17484 than our prediction for *Psc*RXLR1 and what was obtained through molecular cloning (Figure 2.5), with the primary difference between the two predictions being that intron 1 is either spliced in *Psc*_781.4, or retained in *Psc*RXLR1

Figure 2.4

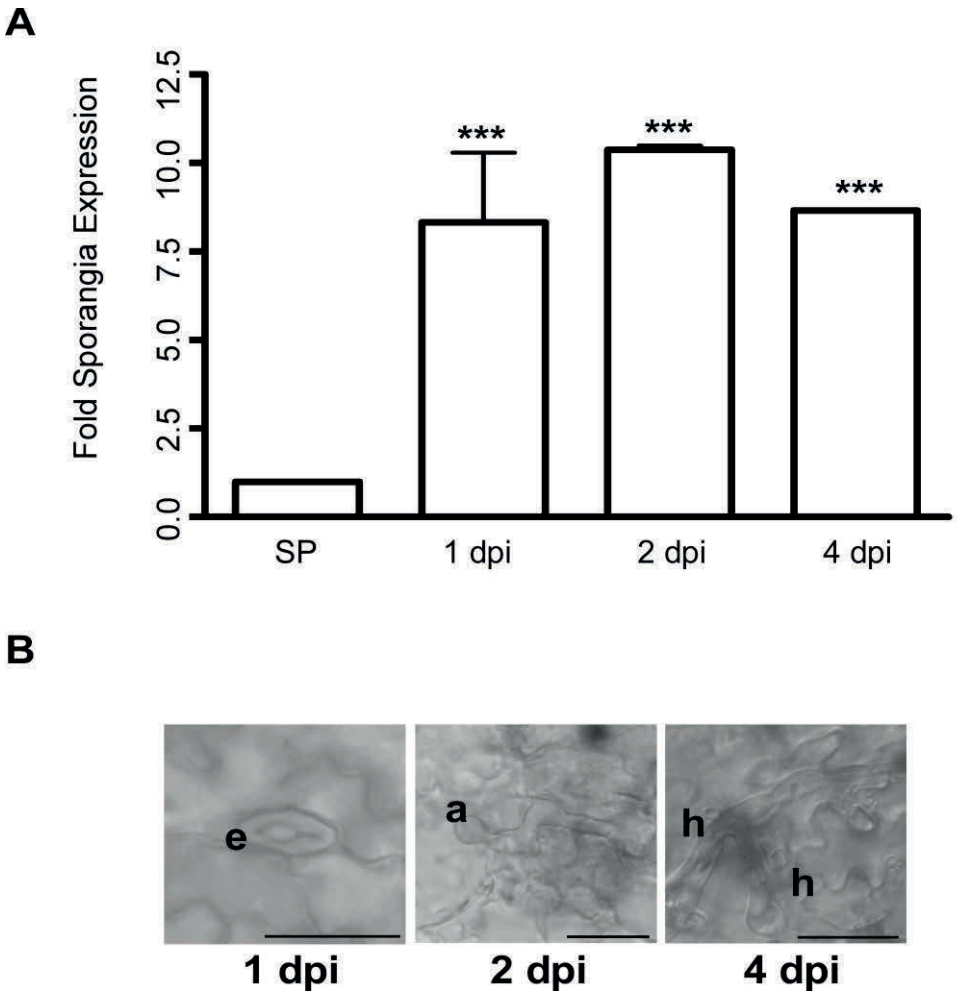

Figure 2.4 *Psc*RXLR1 mRNA expression is up-regulated during *Pseudoperonospora cubensis* infection of cucumber. (A) Expression levels of *Psc*RXLR1 in sporangia (SP) and at 1, 2, and 4 days post-inoculation (dpi). Expression is displayed as fold sporangia expression and all time points are significantly different compared to SP control (*** indicates $p < 0.001$) using Two-way ANOVA. Error bars represent the standard error of the mean of 2 technical replicates from each of 2 biological replicates. (B) Differential interference (DIC) microscopy images of stages of *Ps. cubensis* infection on cucumber where e = encysted zoospore; a = appressorium; and h = haustorium. Scale bars = 25 µm.

(Figure 2.5A). Empirical whole transcriptome sequence data (RNA-seq) from *Ps. cubensis* sporangia (unpublished results) provides support for both isoforms at this locus. When the first intron is retained, a stop codon is also brought into frame, yielding a truncated transcript (i.e., putative effector *Psc*RXLR1), and subsequently, a smaller protein, which, as described in Figure 2.2, lacks the EamA functional domain (Figure 2.5A).

Based on our *in silico* predictions we confirmed which gene model, or both, was represented *in vivo*. Using an RT-PCR-based approach, we were able to amplify both splice variants from purified sporangia (SP), as well as from infected leaf material harvested at 4 and 8 days post-inoculation (dpi) (Figure 2.5B), suggesting that both isoforms are present throughout the infection process. As an added control, the resultant PCR products were cloned and sequenced to confirm that they corresponded to the appropriate splice variant (Figure S2.4). Additional functional analysis of *Psc*_781.4 confirmed transient expression in *N. benthamiana* does not elicit a cell death response *in planta*, indicating that it likely has no virulence function in *Ps. cubensis* (Figure S2.5). In total, these independent methods confirm our conclusion that *Psc*_781.4 is alternatively spliced leading to generation of a functional RXLR effector protein.

Figure 2.5

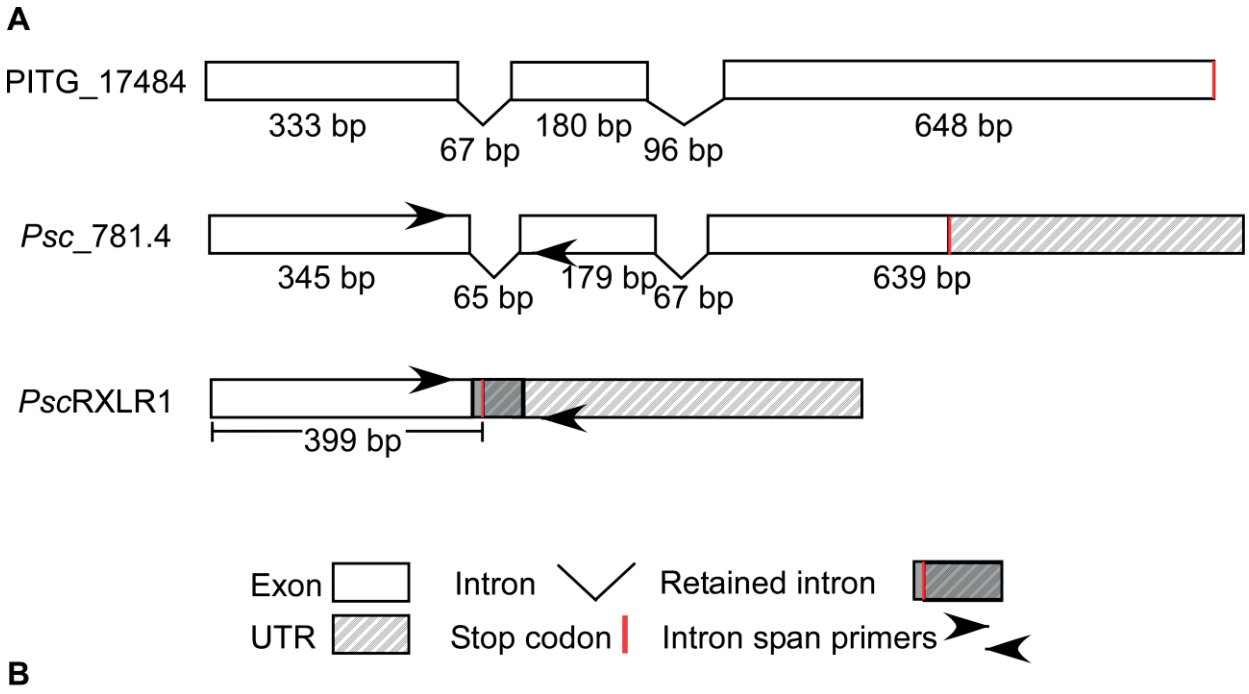

Figure 2.5 *Psc*RXLR1 is a splice variant of *Psc*_781.4. (A) Schematic representation of intron and exon structures of PITG_17484, *Psc*_781.4, and *Psc*RXLR1. (B) RT-PCR analysis of alternative splicing in *Psc*_781.4 in sporangia (SP) and at 4 and 8 days post-inoculation (dpi). RT-PCR products were subcloned into the pGEM vector and DNA was amplified by PCR using intron spanning primers as shown in (A). S, transcript with spliced intron 1. NS, non-spliced transcript.

DISCUSSION

In this study, we describe a candidate RXLR-type effector from *Ps. cubensis* that results from a splice variant of a putative multi-drug transporter protein, and additionally expands the scope of our understanding of the function and evolutionary history of the *Ps. cubensis* effector repertoire. While *Ps. cubensis* is an oomycete pathogen of worldwide economic importance, insight into the mechanism(s) underlying its virulence and pathogenicity remain limited [43]. A recent study has provided a foundation for investigating the genetic basis for virulence and pathogenicity in *Ps. cubensis* through generation of a large scale genomic dataset [24]. We build upon this previous work using a combination of *in silico* analyses, gene expression studies, and cell biology to functionally characterize *Psc*RXLR1 and establish a potential role in promoting *Ps. cubensis* infection and proliferation.

Alternative splicing has been previously described in oomycetes pathogens of plants; specifically related to the family 5 endoglucanases (EGL5) from *P. sojae* [21], as well as in gene families such as Crinklers (CRNs), protein kinases, and transcription factors [22]. In *P. sojae*, EGL5 proteins have a role in infection of soybean and are highly up-regulated during the early stages of infection. As part of these analyses, four different mechanisms of alternative splicing were described: intron skip, exon skip, alternative donor site, and alternative acceptor site, with intron skip, where the intron is retained, being the most commonly observed mechanism [22]. In agreement with this previous observation, we propose that the *Psc*_RXLR1 transcript is generated *via* a retained

intron from *Psc*_781.4, yielding an RXLR-type effector. From an evolutionary standpoint, alternative-splicing functions to expand the capacity of an organism's proteome, thus enabling the generation of multiple functional isoforms from a single coding sequence. Over time, new isoforms may be maintained if they have a beneficial function [44], or lost, if their function is not critical to the lifecycle of the organism. In the case of plant pathogens, this process could potentially serve an adaptive role to allow for generation of isoforms of "housekeeping" type genes that gain new function(s), potentially to promote virulence or infection. Alternatively, the *Psc*RXLR1 splice variant could represent a step in evolutionary time, as it is generated from the same coding sequence as *Psc*_781.4 and maintained in the coding repertoire, but has not been duplicated or retained as a separate sequence.

Like other oomycete effectors characterized to date, *Psc*RXLR1 has a functional 26 amino acid signal peptide necessary for secretion from the haustorium into the extrahaustorial matrix prior to translocation into the host cytoplasm. Interestingly, *Psc*RXLR1 is also localized to the host plasma membrane, despite the lack of a predicted transmembrane domain or localization signal (Figure 2.3). To further examine this, we surveyed the genomes of additional oomycete plant pathogens for orthologous sequences. We hypothesize that if functional characterization data for orthologs in any of these better characterized species (e.g., *P. infestans*, *P. ramorum* and *P. sojae*) existed, it might provide insight into both the function and conservation of *Psc*RXLR1. Through BLAST analysis of the *P. infestans*, *P. sojae*, *P. ramorum* and *Py. ultimum* genomes, we identified orthologous sequences in each of the pathogens, all of which

were annotated as members of the Drug/Metabolite Transporter (DMT) superfamily [31]. The DMT superfamily encompasses 19 families; the orthologs described here are members of the EamA family, named for the O-acetylserine/cysteine export gene in *Escherichia coli* [45]. While *Psc*RXLR1 is lacking the EamA transmembrane domains that are characteristic of these transporter proteins, our data clearly demonstrated plasma membrane localization.

Monitoring the expression of both pathogen and host genes during infection can provide insight into the interplay between resistance and susceptibility [46]. Using qRT-PCR, we demonstrated that expression of *Psc*RXLR1 mRNA is up-regulated during the early stages of infection of cucumber. While we were unable to distinguish between isoforms, expression was induced nearly 10-fold at 1, 2 and 4 days post-inoculation (dpi), corresponding with zoospore encystment, appressorium formation and penetration, and proliferation and haustoria establishment, respectively (Figure 2.5). Several effector proteins from *P. infestans* have also been demonstrated to have distinct temporal patterns of expression, and are often expressed during the pre- and early stages of infection, representative of the biotrophic phase of the *P. infestans* life cycle [12]. Based on the robust induction of *Psc*RXLR1 mRNA during the early stages of infection, as well as the aggressive nature of the necrosis-inducing activity observed in *N. benthamiana*, we hypothesize that the expression pattern of *Psc*RXLR1 could support a role in the initial infection process, possibly through dampening of host defense responses. Indeed, effectors from other oomycete plant pathogens, including *Pc*QNE from *Ps. cubensis* and members of the CRN family from *P. infestans*, have also been shown to

elicit similar phenotypes when transiently expressed in *N. benthamiana* [3,24,47], supporting the classification of *Psc*RXLR1 as an effector protein with a putative role in virulence.

In the current study, ongoing analysis of the *Ps. cubensis* genome has expanded the candidate effector complement of *Ps*. *cubensis* to 271 sequences, revealing significant variation in the conserved translocation motif. While previous analyses revealed a near equal distribution of RXLR:QXLR motifs in *Ps. cubensis*, our current work, based on a higher coverage draft genome sequence and predicted protein sequences rather than open reading frames (ORFs), predicts sequences with 20 different amino acid possibilities at the R1 position of the XXLR motif. Of these 20 predicted R1 substitutions, 19 have expression support (Table S2.1). While all 20 R1 substitutions have yet to be functionally validated, it is not surprising that *Ps. cubensis* effectors may in fact utilize a more diverse set of translocation motifs compared to the *Phytophthora* spp., given its obligate lifestyle and relatively narrow host range. Among *Phytophthora* spp., the conservation of the RXLR motif is well-established, yet there are additional classes of oomycete effectors, such as the CRN family, that appear to utilize disparate translocation motifs [34,47]. Moreover, analysis of the *Py. ultimum* genome has identified an additional predicted translocation motif, YxSL[RK] [9]. Indeed, divergence of transport signal sequences is even more pronounced between oomycetes and the true fungi, which have no obvious conserved motifs that could function in transport and show high degrees of variation even within the same species [48]. For example, the effectors AvrM and AvrL567 from *Melampsora lini*, an obligate rust fungi with a similar

lifestyle to *Ps. cubensis*, rely on unique N-terminal sequences for uptake [48]. These sequences, while different in regards to sequence similarity from the RXLR motif observed in *Phytophthora* spp., are similar in that they feature positively charged residues, implying that secondary protein structure may be a factor contributing to uptake of these proteins. Both *M. lini* and *Ps. cubensis* are obligate biotrophs with specific host ranges, which may have influenced the evolution of their effector repertoires to select for unique translocation motifs compared to those found in *Phytophthora* spp.

Preliminary analysis of the *Ps. cubensis* effector repertoire reveals minimal orthology with annotated effector proteins from *P. infestans,* similar to what has been observed when comparing the effector complement from *P. infestans* with *P. ramorum*, *P. sojae* or *H. arabidopsidis* [8,27,29]. Through extensive analysis using both evolutionary and comparative genomics, *Phytophthora* RXLR effector genes have been shown to be undergoing accelerated rates of birth and death evolution as well as both widespread gene duplication and loss events [7,8]. As such, among the predicted RXLR effector genes from *P. infestans, P. sojae*, and *P. ramorum*, there are few genes with high degrees of sequence similarity or 1:1:1 orthology [7]. Similarly, the same phenomenon was observed in comparing candidate effector sequences from *Ps. cubensis* to those of *P. infestans.* Of 271 predicted *Ps. cubensis* sequences, less than half (41%) of these had significant similarity (e-value < $1e^{-5}$) to predicted *P. infestans* proteins, and only 3 of these sequences had similarity to annotated *P. infestans* RXLR effector proteins. These results indicate that the effector repertoire *Ps. cubensis* utilizes to promote its virulence

and pathogenicity on its hosts is, as could be predicted, very different than that utilized by *P. infestans*, and likely the other *Phytophthora* spp. as well. This is likely due to differing selective pressures on *Ps. cubensis* resulting from host specificity as well as differences in lifestyle between the two pathogens (i.e., obligate vs. non-obligate).

In this study, we identified minimal conservation between the predicted *Ps. cubensis* effector complement and effectors from *P. infestans*. We hypothesize that the identification and analysis of effector to non-effector relationships among oomycete plant pathogens is a valid measure to assess conservation and rates of evolution. Additionally, with the identification of *Psc*RXLR1, a splice variant of a non-effector gene, we posit that these types of analyses as well as a more thorough analysis of alternative splicing may provide a preliminary baseline to not only investigate evolutionary differences among oomycete plant pathogens, but to also infer the relationship between effector repertoire and the host-pathogen specificity and lifestyle. We have used several criteria (i.e., prediction of selection pressure, secretion, etc.) to identify and analyze the relationship between predicted *Ps. cubensis* effectors and their orthologs in *P. infestans*. As observed for other effector proteins, some *Ps. cubensis* effectors may have experienced stronger positive selection than most other proteins within the genome. Interestingly, in addition to varying significantly from the genome average, the distribution of ω for the *Psc*E-*Pi* pairs has two distinct peaks, representing groups of effectors under different levels of selection pressure. Thus, it appears that aside from acting as effectors during infection, some of these slower evolving genes may have additional, "housekeeping" roles that are yet to be uncovered. Despite computational

evidence indicating that these slower evolving genes are likely effectors, their role(s) in pathogenesis remain to be established.

MATERIALS AND METHODS

Ps. cubensis culture and growth conditions

Ps. cubensis was maintained on *Cucumis sativus* cv. 'Vlaspik' as previously described [24]. Cucumber plants were grown at 22 °C with a 12 h light/dark photoperiod. For *Ps. cubensis* inoculation, sporangia were collected from heavily sporulating leaves by washing with cold sterile distilled water and collecting sporangia in a centrifuge tube. Sporangia were enumerated with a hemocytometer and suspended to a concentration of 1 x 10^5 sporangia/ml in sterile distilled water. The underside of fully expanded 2^{nd} or 3^{rd} true leaves of 4-week-old cucumber plants were spray-inoculated, until run-off, with the suspension, and incubated for 24 h at 100% humidity in the dark. After 24 h, inoculated plants were moved to a growth chamber (22 °C with a 12 h light/dark photoperiod).

DNA and RNA extraction

Genomic DNA of *Ps. cubensis* was isolated from sporangia of isolate MSU-1 using the DNeasy Plant Mini kit (Qiagen, Germantown, MD) with slight modifications. Sporangia were washed from heavily sporulating leaves with sterile distilled water and filtered

through a 40 μm nylon cell strainer to remove residual plant debris. The resultant sporangia suspension was centrifuged, and the supernatant decanted. Sporangia were suspended in buffer AP1 containing RNase and 5 μl of Proteinase K and incubated at 37 °C for 20 min. 50 μl of 425-600 μm acid washed beads were added to the sporangia suspension and sporangia disrupted for 3 min using a vortex. Subsequent DNA extraction steps were performed according to manufacturer's instructions.

Ps. cubensis total RNA was isolated as follows: sporangia were collected as described above for DNA isolation, yet re-suspended in buffer RLT from the RNeasy Plant Mini Kit (Qiagen, Germantown, MD) and disrupted as above. RNA isolation was performed according to the manufacturer's instructions. RNA samples were treated with DNase (Promega, Madison, WI) prior to use.

Sequence, assembly, and annotation of the *Ps. cubensis* genome

Genomic DNA was isolated from *Ps. cubensis* MSU-1 as described above and libraries constructed using the Illumina Genomic DNA Sample kit (Illumina, San Diego, CA). Two separate paired-end libraries were end sequenced using an Illumina Genome Analyzer II (Illumina, San Diego, CA) at the UC-Davis Genome Center. The first library was sequenced with 84bp reads and an insert size of 180bp yielding 7.8 Gbp of sequence. The second library was sequenced with 100bp reads and an insert size of 480bp yielding 5.5 Gbp of sequence. Illumina reads were trimmed to 51 bp to remove low quality regions at the 3' end of the reads. Reads with more than one N base or a base

with a quality score less than 20 were removed. The reads were then searched against the Cucumber genome assembly [49] with Bowtie v0.12.7 [50] and matching reads were removed; 4.5 Gbp of sequence was retained following trimming and cleaning the reads. The trimmed and cleaned reads were assembled using Velvet v1.0.14 [25]. Three Velvet runs were performed with hash lengths of 31, 41, and 51 and coverage cutoffs of 7, 3.6, and 2, respectively. A minimum contig size cutoff of 200bp was used for all the assembly runs. The contigs from each Velvet run were then merged into one assembly using the Minimus2 pipeline (http://sourceforge.net/apps/mediawiki/amos/index.php?title=Minimus2). Contaminant-containing and mitochondrial contigs were removed; the final assembly contains 35,546 contigs with an N50 contig size of 4.0 Kbp; the total assembly is 64.4 Mbp. Reads were deposited in the Sequence Read Archive at the National Center for Biotechnology Information under study number SRP011018. The assembly is available at NCBI under accession AHJF00000000 and via http://www.daylab.plp.msu.edu/wp-content/uploads/psc_merged_contigs.fasta.zip. The annotation can be downloaded at http://www.daylab.plp.msu.edu/wp-content/uploads/psc_merged_contigs.gff3.zip.

The assembly was annotated using the MAKER [26] annotation pipeline. The FGENESH gene finder [51] was used with the *Phytophthora* matrix to produce the initial gene calls for the pipeline. All transcript and protein sequences from sequenced oomycete genomes were provided to MAKER to improve the quality of the annotation. In total, 23,519 loci and 23,522 gene models were predicted. Putative functional

annotation was assigned by searching the gene models against UniRef100 using BLASTX (cutoff: 1E-5) and transferring the first hit with informative annotation.

Identification and cloning of *Psc*RXLR1, *Psc*_781.4 and PITG_17484

Amplification of the coding sequence of *Psc*RXLR1 was performed using DNA primers that correspond to the open reading frame of *Psc*RxLR1 (Figure 2.5). Subsequent isolation and cloning of *Psc*RxLR1 was performed by PCR using gene-specific primers and genomic DNA from *Ps. cubensis* sporangia. Resultant amplicons were cloned into the TA cloning vector pGEM-T-Easy (Promega), generating pGEM_*Psc*RXLR1. To ensure identification of a complete coding sequence, as well as to verify the absence of introns in the sequence, 3′ RACE (Rapid Amplification of cDNA Ends) was performed using the SMARTer RACE cDNA Amplification Kit (Clontech, Mountain View, CA). Amplification of *Psc*_781.4 was performed using 3' RACE as described above and the final coding sequence was amplified using gene specific primers (Figure 2.5). Fidelity of all sequences was confirmed by DNA sequencing using the ABI 3730 Genetic Analyzer (Applied Biosystems, Foster City, CA).

P. infestans clone PITG_17484 was amplified from cDNA of *P. infestans* by PCR using gene-specific primers. Amplicons were subcloned into pGEM-T-Easy and sequences confirmed by sequencing.

DNA cloning and construct preparation

To validate the predicted signal peptide (SP) of *Psc*RXLR1, the yeast signal trap assay was used [24,32,33]. A DNA fragment corresponding to the predicted 26 amino acid signal peptide (including start codon) was amplified by PCR using gene specific primers (Table S3) modified to include 5′ *Eco*R1 and 3′ *Xho*I restriction sites. Resultant amplicons were cloned into TA cloning vector pGEM-T-Easy (Promega), to yield pGEM-*Psc*RXLR1-SP. The plasmid pGEM-*Psc*RXLR1-SP was digested with *Eco*R1 and *Xho*I, the 84 bp SP fragment was gel purified, and ligated into the *Eco*R1/*Xho*I sites of the yeast signal trap vector pSUC2T7M13ROI [32], generating pSUC2-*Psc*RXLR1.

The SP sequence of PITG_17484 was similarly amplified and cloned, generating pSUC2-PITG_17484. Plasmids pSUC2T7M13ROI, pSUC2-*Psc*RXLR1 and pSUC2-PITG_17484 were transformed into the yeast *SUC2* minus strain YTK12 using the Frozen-EZ Yeast Transformation IITM kit (Zymo Research, Orange, CA) following the manufacturer's instructions. Transformants were selected on CMD-W plates (0.67% yeast nitrogen base without amino acids, 0.075% tryptophan dropout supplement, 2% sucrose, 0.1% glucose, and 2% agar) for 3 days at 28 °C. To confirm signal peptide function, YTK12 containing the pSUC2 constructs were grown on raffinose-containing YPRAA plates (1% yeast extract, 2% peptone, 2% raffinose, 2 μg/L Antimycin A, and 2% agar). Yeast strains were replicated onto YPDA plates (1% yeast, 2% peptone, 2% glucose, 0.003% adenine hemisulfate, and 2% agar) and CMD-W plates as equal viability controls. TYK12 without pSUC2 was used as a negative control. The detection

of the secreted invertase activity with 2,3,5-triphenyltetrazolium chloride (TTC) was performed as described by Tian et al. [24].

For construction of plasmids used for localization and phenotype studies, the open reading frames of *Psc*RXLR1, *Psc*_781.4 and PITG_17484 (minus stop codons) were amplified and cloned into the Gateway entry vector pENTR/D-TOPO (Invitrogen, Carlsbad, CA), yielding pENTR-*Psc*RXLR1-GW, pENTR-*Psc*_781.4-GW, and pENTR-PITG_17484-GW, respectively. The destination vector pVKH18En6gw-cCFP [24] was used to create the C-terminal CFP fusions, using heterologous recombination *via* LR Clonase, as per the manufacturer's instructions (Invitrogen).

Transient expression and localization in *N. benthamiana*

Infiltration and transient expression in *N. benthamiana* using *A. tumefaciens* was performed on 4-6 week old plants as described in Tian et al. [24]. *A. tumefaciens* strains were grown overnight at 28 °C on Luria-Bertani (LB) plates containing 50 µg/mL rifampicin and 25 µg/mL kanamycin. *A. tumefaciens* clones were re-suspended in induction buffer (10 mM MES, pH 5.6, 10 mM $MgCl_2$, 150 µm acetosyringone) and incubated at room temperature, shaking in the dark, for 2 hours prior to infiltration. *A. tumefaciens* suspensions were infiltrated at a final concentration of OD_{600} = 0.8.

A. tumefaciens-mediated transient expression in *N. benthamiana* for localization of *Psc*RXLR1-CFP, *Psc*_781.4-CFP, and PITG_17484-CFP with *At*PIP2A-YFP [38] was

performed as described above. Visualization of fluorescently tagged proteins was observed using an Olympus Fluoview 1000 laser scanning confocal microscope. Images were adjusted for contrast in Canvas X (ACD Systems).

Quantitative real time PCR

First strand cDNA was synthesized from 1 μg total RNA using the first-strand cDNA synthesis kit (USB, Cleveland, OH). Quantitative RT-PCR was performed using a Mastercycler ep Realplex real-time PCR (Eppendorf, Hamburg, Germany) using HotStart SYBR Green qPCR Master Mix (2x) (USB), as previously described [52]. For amplification of *Psc*RXLR1 transcripts, gene specific primers were designed to amplify a 50 bp fragment (Forward: 5'-TGCGTAGCATCGCCAACCGA-3 and Reverse: 5'-TCTTGCCAGCTGCATCGCGA-3'). Primers specific for the *Ps. cubensis* internal transcribed spacer (ITS) region were used as an endogenous control (Table S3). Cycle parameters were as follows: 95 °C for 2 min, followed by 40 cycles of: 95 °C (15 sec), 60 °C (15 sec) and 72 °C (45 sec). Fold expression was calculated based on expression in sporangia. Data were analyzed by two-way ANOVA using Prism 4 (GraphPad Software).

Splice variant analysis

Primers spanning the region of intron 1 were used (Figure S2.4) to amplify RT-PCR products from SP, 4, and 8 dpi cDNA samples and resultant products were cloned into

the TA cloning vector pGEM-T-Easy. Fidelity of all sequences was confirmed by DNA sequencing as described above.

Ortholog identification and sequence analysis

Candidate effector proteins were identified from the predicted proteome of *Ps. cubensis* generated from the draft genome assembly using a modified RXLR effector identification pipeline [29]. Orthologs of *Psc*RXLR1 were identified in *P. infestans* (http://www.broadinstitute.org/), *P. sojae* (http://genome.jgi-psf.org), *P. ramorum* (http://genome.jgi-psf.org) and *Py. ultimum* (http://pythium.plantbiology.msu.edu) with BLAST [53]. Signal peptides were predicted using SignalP 3.0 ([54], http://www.cbs.dtu.dk/services/SignalP/) and protein motifs identified using InterProScan ([55], http://www.ebi.ac.uk/Tools/pfa/iprscan/). Amino acid alignments were generated using ClustalW2 ([56], http://www.clustal.org), and resultant figures generated using BoxShade v3.21 (http://www.ch.embnet.org/software/BOX_form.html). *P. infestans* effector sequences from Haas *et al*. [7] were used for analyses and *P. infestans* protein models were obtained from the *Phytophthora infestans* Sequencing Project (http://www.broadinstitute.org/).

Defining paralogs and orthologs and evolutionary rate estimates

Synonymous and non-synonymous substitution rates (*Ks* and *Ka*, respectively) were determined using the yn00 program in PAML [57]. Protein sequences were aligned first

and “back-translated” to coding sequence alignments. Very few pairs had run errors (e.g., NAN in PAML output), and those with run errors were excluded. Sequence pairs that were too similar (*Ks* ≤ 0.005) or too divergent (*Ks* > 3) were also excluded from further analysis. For each *Ps. cubensis* or *P. infestans* effector protein, the closest paralogous genes were identified using within-species BLAST searches and used for rate calculation. Rates between putative orthologs were calculated as well. Putative orthologs were identified globally between *Ps. cubensis, P. infestans* or *P. ultimum* by first determining pairwise sequence similarities between all genes in these species. For each *Ps. cubensis* protein, X, a protein in a second species, Y, is considered an ortholog if the following two conditions are met: 1) X is the reciprocal best match of Y and 2) X is located in a syntenic block where Y is found. Syntenic regions were established using Multiple Collinearity Scan [58], with $1e^{-5}$ as an alignment significance threshold, match size ≥5, and average intergenic distance.

Molecular phylogenetic analysis

The full-length protein sequences of *Psc*RXLR1 and its orthologs were aligned using default parameters with MUSCLE [59]. The multiple sequence alignment was used to infer phylogenetic relationships between *Psc*RXLR1 and its orthologs using the Maximum Likelihood method, based on the JTT matrix-based model [60] with MEGA5 [61]. Bootstrap values (based on 500 replicates) for each node are given if >25%.

ACKNOWLEDGEMENTS

We thank members of the Day lab for critical reading of the manuscript. Dr. Joe Win (The Sainsbury Laboratory) is gratefully acknowledged for his contributions in the preliminary analysis of genomic sequences from *Ps. cubensis*.

APPENDIX

Figure S2.1

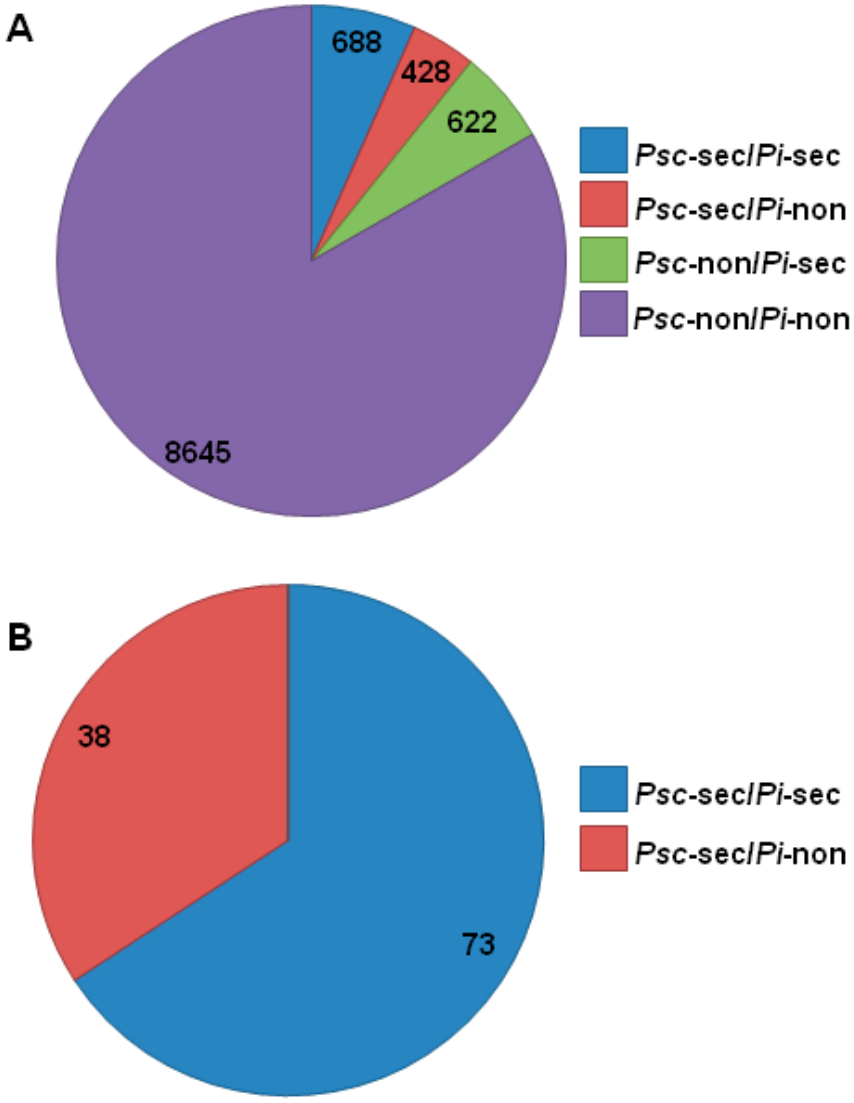

Figure S2.1 Signal peptide distribution among ortholog pairs. (A) Distribution of secreted or non-secreted proteins in the *Pseudoperonospora cubensis* – *Phytophthora infestans* ortholog baseline dataset. (B) Distribution of *P. infestans* orthologs of *Ps. cubensis* effectors that are predicted to be secreted. *Psc*-sec = *Ps. cubensis* secreted protein. *Psc*-non = *Ps. cubensis* non-secreted protein. *Pi*-sec = *P. infestans* secreted protein. *Pi*-non = *P. infestans* non-secreted protein.

Figure S2.2

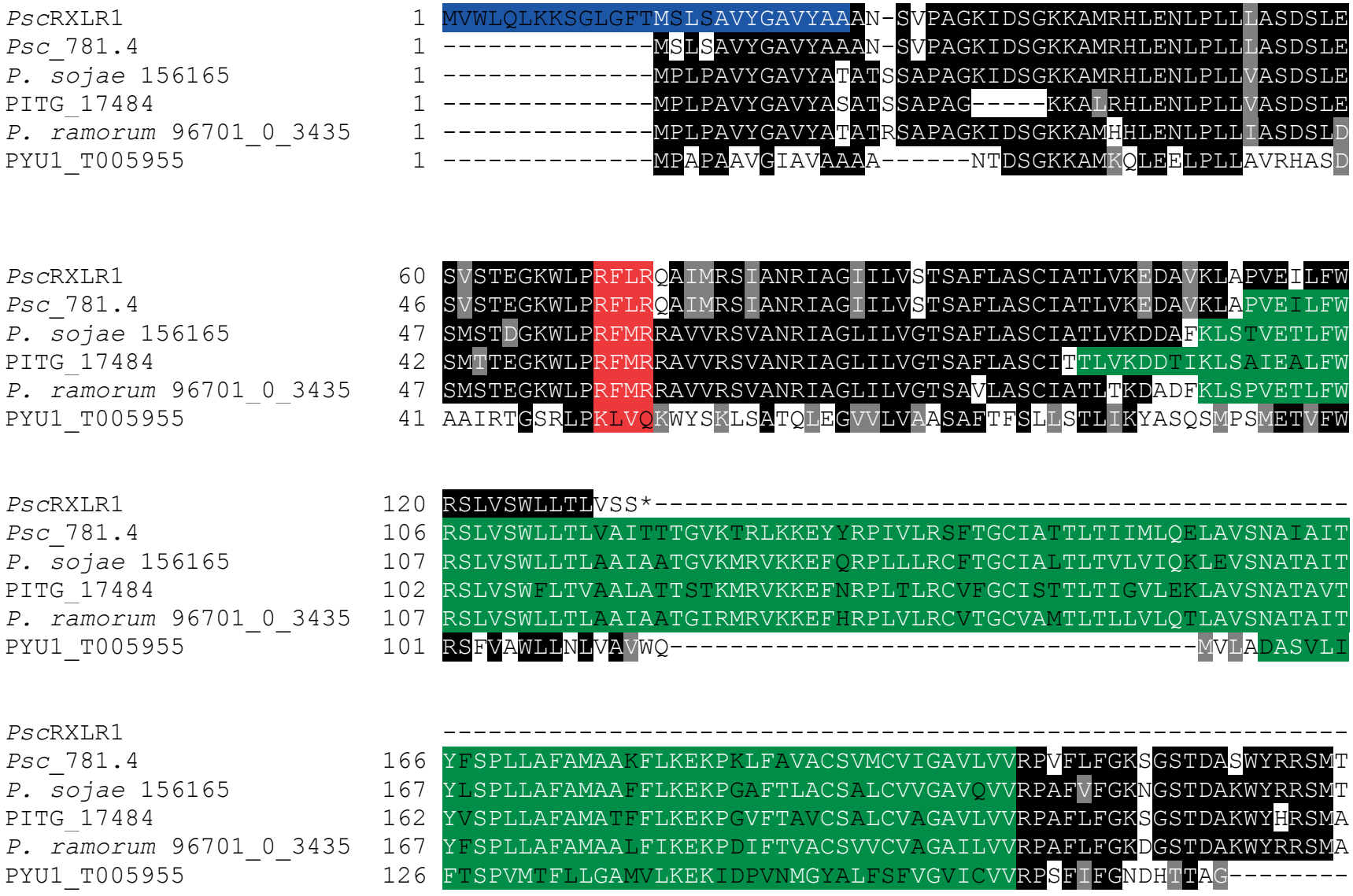
PscRXLR1 1 MVWLQLKKSGLGFTMSLSAVYGAVYAAAN-SVPAGKIDSGKKAMRHLENLPLLLASDSLE
Psc_781.4 1 ----------MSLSAVYGAVYAAAN-SVPAGKIDSGKKAMRHLENLPLLLASDSLE
P. sojae 156165 1 ----------MPLPAVYGAVYATATSSAPAGKIDSGKKAMRHLENLPLLVASDSLE
PITG_17484 1 ----------MPLPAVYGAVYASATSSAPAG-----KKALRHLENLPLLVASDSLE
P. ramorum 96701_0_3435 1 ----------MPLPAVYGAVYATATRSAPAGKIDSGKKAMHHLENLPLLIASDSLD
PYU1_T005955 1 ----------MPAPAAVGIAVAAAA------NTDSGKKAMKQLEELPLLAVRHASD
PscRXLR1 60 SVSTEGKWLPRFLRQAIMRSIANRIAGIILVSTSAFLASCIATLVKEDAVKLAPVEILFW
Psc_781.4 46 SVSTEGKWLPRFLRQAIMRSIANRIAGIILVSTSAFLASCIATLVKEDAVKLAPVEILFW
P. sojae 156165 47 SMSTDGKWLPRFMRRAVVRSVANRIAGLILVGTSAFLASCIATLVKDDAFKLSTVETLFW
PITG_17484 42 SMTTEGKWLPRFMRRAVVRSVANRIAGLILVGTSAFLASCITTLVKDDTIKLSAIEALFW
P. ramorum 96701_0_3435 47 SMSTEGKWLPRFMRRAVVRSVANRIAGLILVGTSAVLASCIATLTKDADFKLSPVETLFW
PYU1_T005955 41 AAIRTGSRLPKLVQKWYSKLSATQLEGVVLVAASAFTFSLLSTLIKYASQSMPSMETVFW
PscRXLR1 120 RSLVSWLLTLVSS*----------
Psc_781.4 106 RSLVSWLLTLVAITTTGVKTRLKKEYYRPIVLRSFTGCIATTLTIIMLQELAVSNAIAIT
P. sojae 156165 107 RSLVSWLLTLAAIAATGVKMRVKKEFQRPLLLRCFTGCIALTLTVLVIQKLEVSNATAIT
PITG_17484 102 RSLVSWFLTVAALATTSTKMRVKKEFNRPLTLRCVFGCISTTLTIGVLEKLAVSNATAVT
P. ramorum 96701_0_3435 107 RSLVSWLLTLAAIAATGIRMRVKKEFHRPLVLRCVTGCVAMTLTLLVLQTLAVSNATAIT
PYU1_T005955 101 RSFVAWLLNLVAVWQ----------
PscRXLR1 ----------
Psc_781.4 166 YFSPLLAFAMAAKFLKEKPKLFAVACSVMCVIGAVLVVRPVFLFGKSGSTDASWYRRSMT
P. sojae 156165 167 YLSPLLAFAMAAFFLKEKPGAFTLACSALCVVGAVQVVRPAFVFGKNGSTDAKWYRRSMT
PITG_17484 162 YVSPLLAFAMATFFLKEKPGVFTAVCSALCVAGAVLVVRPAFLFGKSGSTDAKWYHRSMA
P. ramorum 96701_0_3435 167 YFSPLLAFAMAALFIKEKPDIFTVACSVVCVAGAILVVRPAFLFGKDGSTDAKWYRRSMA
PYU1_T005955 126 FTSPVMTFLLGAMVLKEKIDPVNMGYALFSFVGVICVVRPSFIFGNDHTTAG--------

Figure S2.2 (cont'd)

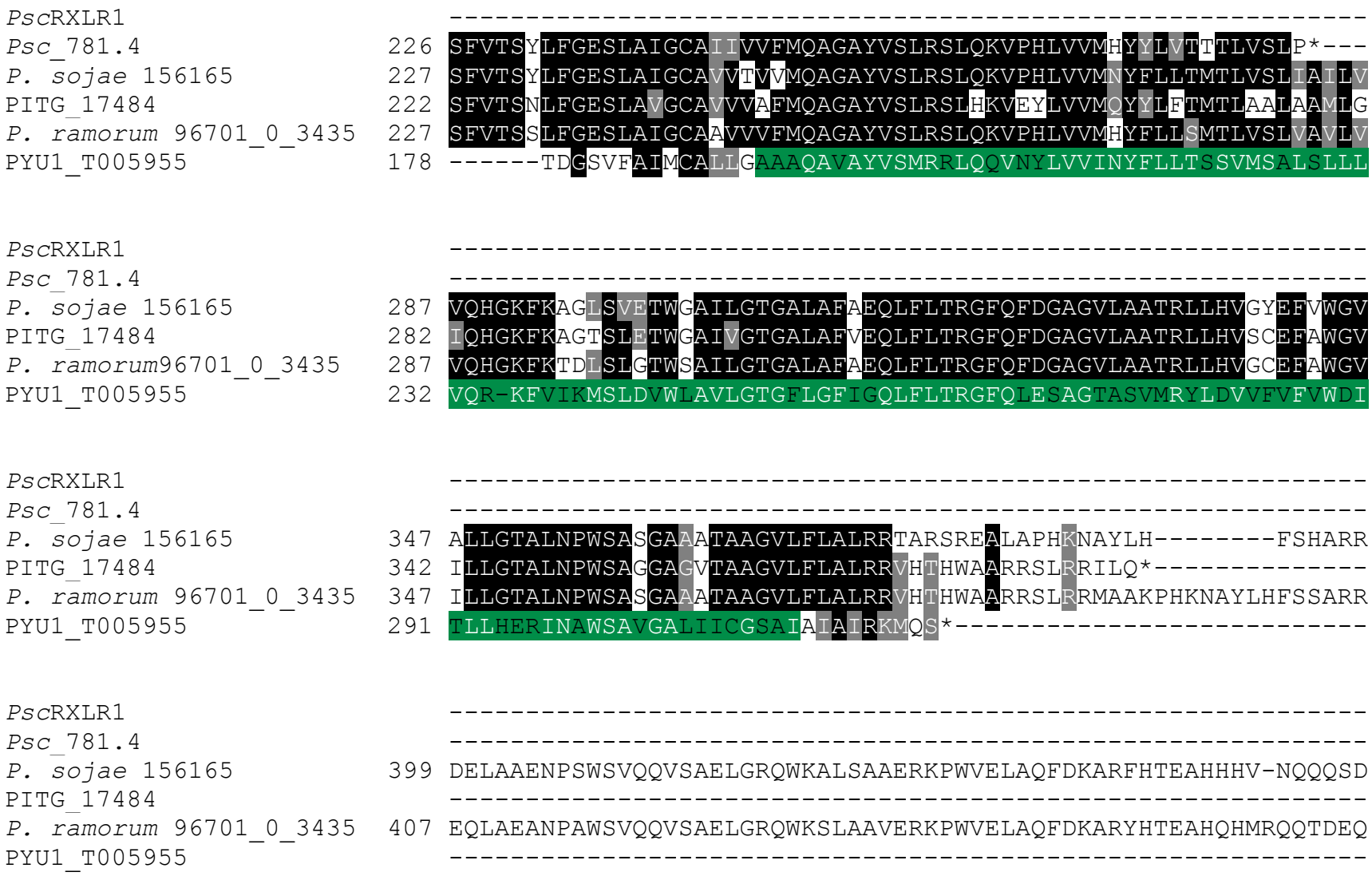

Figure S2.2 (cont'd)

```
PscRXLR1                        ------------------------------------------------------------
Psc_781.4                       ------------------------------------------------------------
P. sojae 156165            458  EQPEQAPPKRKKQSNEPRQPDTAYICFWKSQRPEVVAANPFLAAPLVSKEVGRQWRALSD
PITG_17484                      ------------------------------------------------------------
P. ramorum 96701_0_3435    467  PERPHLPTKRKKRPNEPRQPDTAYICFWKSRRPEVIAENPLLAAPSVSREEIARFQP---
PYU1_T005955                    ------------------------------------------------------------


PscRXLR1                        ------------------------------------------------------------
Psc_781.4                       ------------------------------------------------------------
P. sojae 156165            518  DERQPTLAATTPDMLSALKTPLKDPFAPKPAKTAFQLFMSHNRESFMLLNMTINEFRAEM
PITG_17484                      ------------------------------------------------------------
P. ramorum 96701_0_3435    524  ------ALVATSEIPPALKAPLKDPFAPKPAKTGFQLFMSHNRESFTLLNMTINEFRTEM
PYU1_T005955                    ------------------------------------------------------------


PscRXLR1                        ------------------------------------------------------------
Psc_781.4                       ------------------------------------------------------------
P. sojae 156165            578  SQLWKRLSDADKAEWHELAKEDQRRYDTEMNAYKPPAYMDLVVQRSHKRMEELRRLARED
PITG_17484                      ------------------------------------------------------------
P. ramorum 96701_0_3435    578  SQLWKRLSDADKNEWYELAKLDERRYETEMNAYKPPAYMESAVQRAHKRLDELRRLARRD
PYU1_T005955                    ------------------------------------------------------------


PscRXLR1                        ------------------------------------------------------------
Psc_781.4                       ------------------------------------------------------------
P. sojae 156165            638  SAAPRLPMNAYNCYLSAKRQELVDRRPGRKNPEIMREIGVTWKALSDDERAVYQRKADED
PITG_17484                      ------------------------------------------------------------
P. ramorum 96701_0_3435    638  AAAPRLPMNAYNCYLSKERQELAVQRPDLKNPEIMREIGVTWKALSEDERASFQRKAEDD
PYU1_T005955                    ------------------------------------------------------------
```

Figure S2.2 (cont'd)

```
PscRXLR1                        ------------------------------------------------------------
Psc_781.4                       ------------------------------------------------------------
P. sojae 156165            698  VERFRAEMEAHIAKKNEEEAANPLTKRRPRKRKEFDDEEELVKTPTVPRKKRKSGPPRRP
PITG_17484                      ------------------------------------------------------------
P. ramorum 96701_0_3435    698  VERFRADMEAYLTQQEEQRAVQEEQDVEPEVVREVVHEVVKEP--VVGRKKRKSVSPRRP
PYU1_T005955                    ------------------------------------------------------------

PscRXLR1                        ------------------------------------------------------------
Psc_781.4                       ------------------------------------------------------------
P. sojae 156165            758  KTAYNLMYMSKRTELLSTYQMSHNECSALCGKLWRQMSEAEREPYKRMAAEDKHRYEAEL
PITG_17484                      ------------------------------------------------------------
P. ramorum 96701_0_3435    756  KTAYNLMYMSKRAELLSTYQMSHNECSALCGRLWRQMSEEEREPYKRMAAEDKRRYETEM
PYU1_T005955                    ------------------------------------------------------------

PscRXLR1                        ------------------------------------------------------------
Psc_781.4                       ------------------------------------------------------------
P. sojae 156165            818  QVYNAQQEEANNKTLRDSAGFRHFLEAKRRENEAISSDEAAAIWQEMTEPHQLLWTELAR
PITG_17484                      ------------------------------------------------------------
P. ramorum 96701_0_3435    816  EIYNAEVDAANKKTLRESAGFSYFLEAKRRENEQISEGEAAAIWRDMLAPHQMLWTELAS
PYU1_T005955                    ------------------------------------------------------------

PscRXLR1                        ------------------------------------------------------------
Psc_781.4                       ------------------------------------------------------------
P. sojae 156165            878  DNKHKTSVERTAVDVLDTLL*------------------------------------
PITG_17484                      ------------------------------------------------------------
P. ramorum 96701_0_3435    876  DTPAPASSNATHQKGQTGMNGGRWTEQEHQSFLAGLRLYGREWKKVAAKIKTRTSAQIRS
PYU1_T005955                    ------------------------------------------------------------
```

Figure S2.2 (cont'd)

```
PscRXLR1                          ------------------------------------------------------------
Psc_781.4                         ------------------------------------------------------------
P. sojae 156165                   ------------------------------------------------------------
PITG_17484                        ------------------------------------------------------------
P. ramorum 96701_0_3435  936 HAQKYFAKLARDDEMRKHSGLSMIMAGSIGYFSDGGSSVAQNSGDDDAEASDASRQMARA
PYU1_T005955                      ------------------------------------------------------------

PscRXLR1                          ------------------------------------------------------------
Psc_781.4                         ------------------------------------------------------------
P. sojae 156165                   ------------------------------------------------------------
PITG_17484                        ------------------------------------------------------------
P. ramorum 96701_0_3435  996 RSAGQSKGTAAILIAPMGSAVSGLYKQTTGATKKRARAAVTGFDGQLEMGAATSSFPYKL
PYU1_T005955                      ------------------------------------------------------------

PscRXLR1                          ------------------------------------------------------------
Psc_781.4                         ------------------------------------------------------------
Psojae_156165                     ------------------------------------------------------------
PITG_17484                        ------------------------------------------------------------
P. ramorum 96701_0_3435      QKRQNDATRVEYLPSQEELLAKASPNLRHRLSSLIEAELCALQVLSCYAMLQQQEQISAP
PYU1_T005955                      ------------------------------------------------------------

PscRXLR1                          --------------------------------
Psc_781.4                         --------------------------------
Psojae_156165                     --------------------------------
PITG_17484                        --------------------------------
P. ramorum 96701_0_3435 1116 RQKTKRQGSAKASTLGLPMLSTEQMPPTSSIY*
PYU1_T005955                      --------------------------------
```

Figure S2.2 (cont'd) Relationship between *Psc*RXLR1 and oomycete orthologs. Alignment of *Psc*RXLR1, *Psc*_781.4, PITG_17484 (*P. infestans*), PYU_T005955 (*Py. ultimum*), *P. ramorum* 96701_0_3435, and *P. sojae* 156165 amino acid sequences were generated using ClustalW and represented with BoxShade. *Psc*RXLR1 signal peptide is boxed in blue. The RXLR or RXLR-like domains are boxed in red. The green boxes represent the EamA domains found in each protein sequence. Stop codons are represented by asterisks.

Figure S2.3

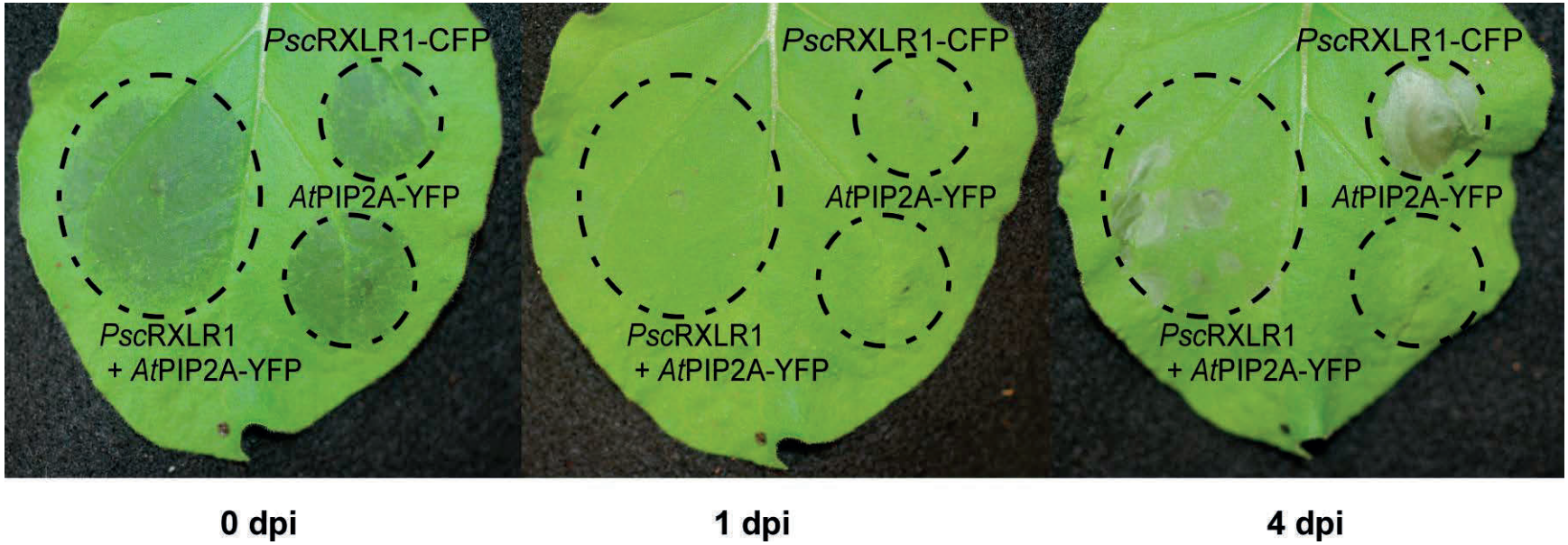

Figure S2.3 Heterologous expression of *Psc*RXLR1 specifically results in cell death in *Nicotiana benthamiana*. Infiltration of *Psc*RXLR1_CFP with or without the plasma membrane marker construct *At*PIP2A-YFP results in chlorosis and necrosis 4 days post-inoculation (dpi). Circles mark the infiltration zones, visible at 0 dpi. Infiltration with *At*PIP2A-YFP alone does not result in any observable phenotype in *N. benthamiana* leaves.

Figure S2.4

A

```
781_4 exon 1(last 100 bp)      TGCCAGCTGC ATCGCGACGT TGGTGAAAGA GGACGCCGTT AAATTAGCAC
781_4 exon1 F primer           ---------- ---------- ---------- ---------- ----------
4 dpi #6-8                     ---------- ---------- ---------- ---------- ----------
8 dpi #4-12                    ---------- ---------- ---------- ---------- ----------
SP #5-9                        ---------- ---------- ---------- ---------- ----------
INTRON 1                       ---------- ---------- ---------- ---------- ----------
781_4 intron 1 span R primer   ---------- ---------- ---------- ---------- ----------
781_4 exon 2 (first 100 bp)    ---------- ---------- ---------- ---------- ----------

781_4 exon 1(last 100 bp)      CCGTGGAGAT TTTGTTTTGG CGCTCACTCG TGTCTTGGCT GCTAACGCTT
781_4 exon1 F primer           ---------- ------TTGG CGCTCACTCG TGTCTTG--------------
4 dpi #6-8                     ---------- ------TTGG CGCTCACTCG TGTCTTGGCT GCTAACGCTT
8 dpi #4-12                    ---------- ------TTGG CGCTCACTCG TGTCTTGGCT GCTAACGCTT
SP #5-9                        ---------- ------TTGG CGCTCACTCG TGTCTTGGCT GCTAACGCTT
INTRON 1                       ---------- ---------- ---------- ---------- ----------
781_4 intron 1 span R primer   ---------- ---------- ---------- ---------- ----------
781_4 exon 2 (first 100 bp)    ---------- ---------- ---------- ---------- ----------

781_4 exon 1(last 100 bp)      ---------- ---------- ---------- ---------- ----------
781_4 exon1 F primer           ---------- ---------- ---------- ---------- ----------
4 dpi #6-8                     GTAAGCTCTT GACCCACTGA TATTGTACGA TGATTGCCTA ACAAATTCTT
8 dpi #4-12                    GTAAGCTCTT GACCCACTGA TATTGTACGA TGATTGCCTA ACAAATTCTT
SP #5-9                        GTAAGCTCTT GACCCACTGA TATTGTACGA TGATTGCCTA ACAAATTCTT
INTRON 1                       GTAAGCTCTT GACCCACTGA TATTGTACGA TGATTGCCTA ACAAATTCTT
781_4 intron 1 span R primer   ---------- ---------- ---------- ---------- ----------
781_4 exon 2 (first 100 bp)    ---------- ---------- ---------- ---------- ----------
```

Figure S2.4 (cont'd)

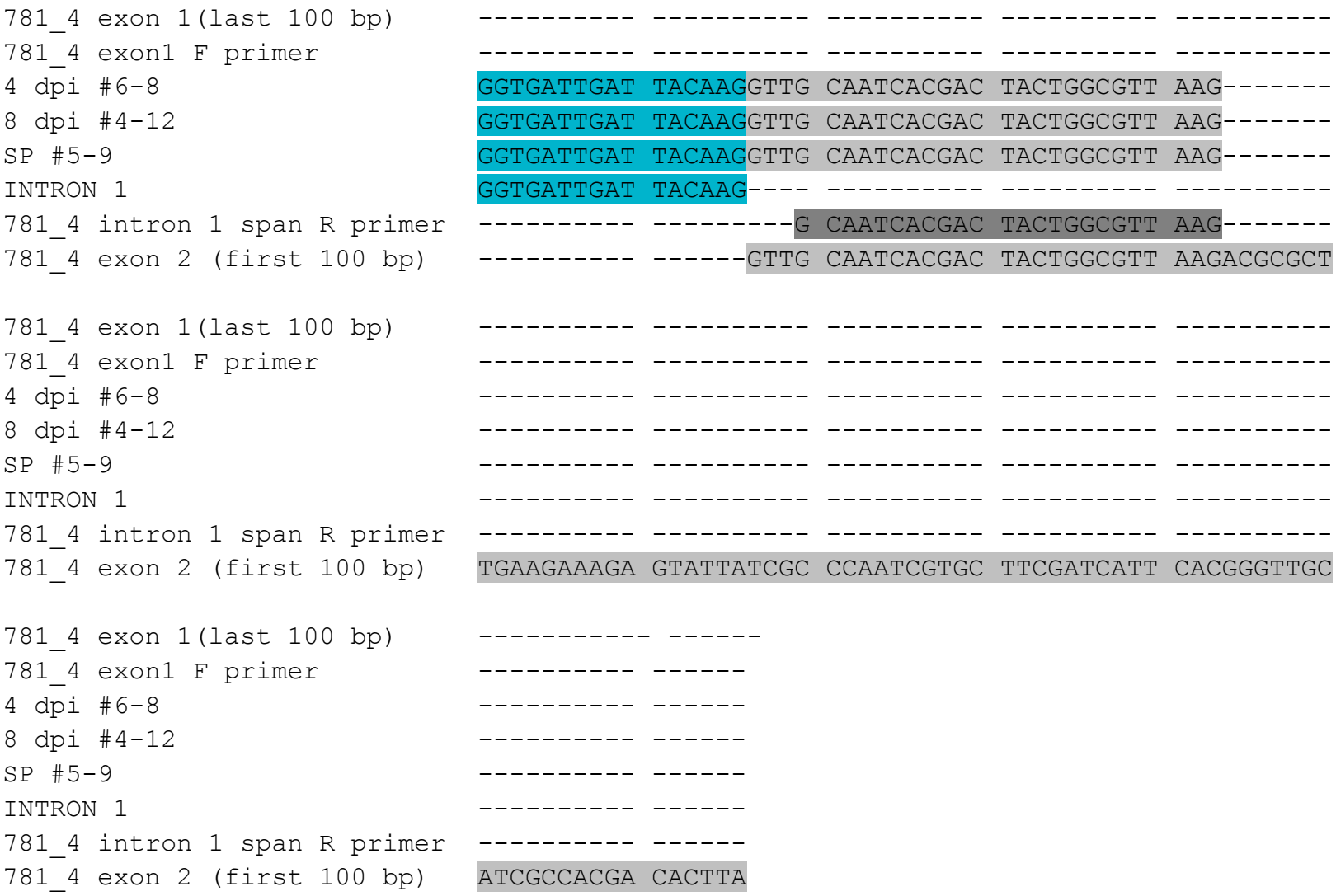

Figure S2.4 (cont'd)

B

```
781_4 exon1 (last 100 bp)      TGCCAGCTGC ATCGCGACGT TGGTGAAAGA GGACGCCGTT AAATTAGCAC
781_4 exon1 F primer           ---------- ---------- ---------- ---------- ----------
4 dpi #3-1                     ---------- ---------- ---------- ---------- ----------
8 dpi #4-7                     ---------- ---------- ---------- ---------- ----------
SP #2-3                        ---------- ---------- ---------- ---------- ----------
781_4 intron 1 span R primer   ---------- ---------- ---------- ---------- ----------
781_4 exon 2 (first 100 bp)    ---------- ---------- ---------- ---------- ----------

781_4 exon1 (last 100 bp)      CCGTGGAGAT TTTGTTTTGG CGCTCACTCG TGTCTTGGCT GCTAACGCTT
781_4 exon1 F primer           ---------- ------TTGG CGCTCACTCG TGTCTTG--- ----------
4 dpi #3-1                     ---------- ------TTGG CGCTCACTCG TGTCTTGGCT GCTAACGCTT
8 dpi #4-7                     ---------- ------TTGG CGCTCACTCG TGTCTTGGCT GCTAACGCTT
SP #2-3                        ---------- ------TTGG CGCTCACTCG TGTCTTGGCT GCTAACGCTT
781_4 intron 1 span R primer   ---------- ---------- ---------- ---------- ----------
781_4 exon 2 (first 100 bp)    ---------- ---------- ---------- ---------- ----------

781_4 exon1 (last 100 bp)      ---------- ---------- ---------- ---------- ----------
781_4 exon1 F primer           ---------- ---------- ---------- ---------- ----------
4 dpi #3-1                     GTTGCAATCA CGACTACTGG CGTTAAG--- ---------- ----------
8 dpi #4-7                     GTTGCAATCA CGACTACTGG CGTTAAG--- ---------- ----------
SP #2-3                        GTTGCAATCA CGACTACTGG CGTTAAG--- ---------- ----------
781_4 intron 1 span R primer   ---GCAATCA CGACTACTGG CGTTAAG--- ---------- ----------
781_4 exon 2 (first 100 bp)    GTTGCAATCA CGACTACTGG CGTTAAGACG CGCTTGAAGA AAGAGTATTA
```

Figure S2.4 (cont'd)

```
781_4 exon1 (last 100 bp)     ---------- ---------- ---------- ---------- ----------
781_4 exon1 F primer          ---------- ---------- ---------- ---------- ----------
4 dpi #3-1                    ---------- ---------- ---------- ---------- ----------
8 dpi #4-7                    ---------- ---------- ---------- ---------- ----------
SP #2-3                       ---------- ---------- ---------- ---------- ----------
781_4 intron 1 span R primer  ---------- ---------- ---------- ---------- ----------
781_4 exon 2 (first 100 bp)   TCGCCCAATC GTGCTTCGAT
```

Figure S2.4. Multiple sequence alignments of splice variant isoforms. (A) Alignment representing non-spliced, *Psc*RXLR1 isoform. (B) Alignment representing spliced isoform indicative of *Psc*_781.4.

Figure S2.5

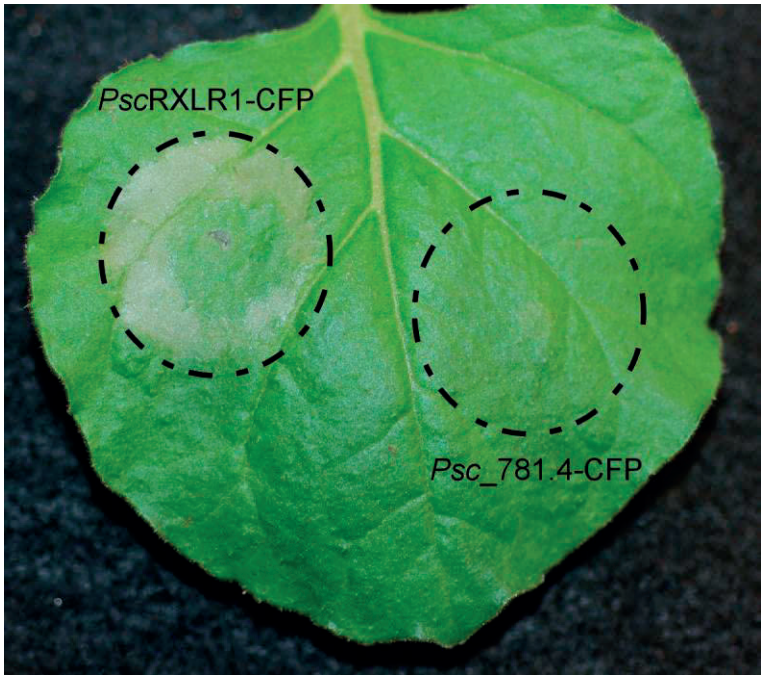

Figure S2.5 Heterologous expression of *Psc*_781.4 in *Nicotiana benthamiana*. Infiltration and expression of *Psc*_781.4 does not result in any observable phenotype in *N. benthamiana* leaves at 4 days post-inoculation (dpi). Circles mark the infiltration zones, visible at 9 dpi.

REFERENCES

REFERENCES

1. De Wit PJGM, Mehrabi R, Van Den B, Harrold A, Stergiopoulos I (2009) Fungal effector proteins: past, present and future. Mol Plant Pathol 10: 735-747

2. Mansfield JW (2009) From bacterial avirulence genes to effector functions via the hrp delivery system: an overview of 25 years of progress in our understanding of plant innate immunity. Mol Plant Pathol 10: 721-734

3. Schornack S, Huitema E, Cano LM, Bozkurt TO, Oliva R, et al. (2009) Ten things to know about oomycete effectors. Mol Plant Pathol 10: 795-803

4. Phytopathology TPCRi (2006) A catalogue of the effector secretome of plant pathogenic oomycetes. Ann Rev Phytopathol 44: 41-60.

5. Chisholm ST, Coaker G, Day B, Staskawicz BJ (2006) Host-microbe interactions: shaping the evolution of the plant immune response. Cell 124: 803-814.

6. Jones JD, Dangl JL (2006) The plant immune system. Nature 444: 323-329.

7. Haas BJ, Kamoun S, Zody MC, Jiang RH, Handsaker RE, et al. (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature 461: 393-398.

8. Tyler BM (2006) *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313: 1261-1266

9. Levesque CA, Brouwer H, Cano L, Hamilton JP, Holt C, et al. (2010) Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. Genome Biol 11: R73.

10. Baxter L, Tripathy S, Ishaque N, Boot N, Cabral A, et al. (2010) Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. Science 330: 1549-1551.

11. Rehmany AP, Gordon A, Rose LE, Allen RL, Armstrong MR, et al. (2005) Differential recognition of highly divergent downy mildew avirulence gene alleles by RPP1 resistance genes from two Arabidopsis lines. Plant Cell 17: 1839-1850.

12. Whisson SC, Boevink PC, Moleleki L, Avrova AO, Morales JG, et al. (2007) A translocation signal for delivery of oomycete effector proteins into host plant cells. Nature 450: 115-118

13. Bos JI, Kanneganti TD, Young C, Cakir C, Huitema E, et al. (2006) The C-terminal half of *Phytophthora infestans* RXLR effector AVR3a is sufficient to trigger R3a-mediated hypersensitivity and suppress INF1-induced cell death in *Nicotiana benthamiana*. Plant J 48: 165-176.

14. Allen RL (2004) Host-parasite coevolutionary conflict between Arabidopsis and downy mildew. Science 306: 1957-1960

15. Armstrong MR, Whisson SC, Pritchard L, Bos JI, Venter E, et al. (2005) An ancestral oomycete locus contains late blight avirulence gene Avr3a, encoding a protein that is recognized in the host cytoplasm. Proc Natl Acad Sci U S A 102: 7766-7771.

16. Shan W, Cao M, Leung D, Tyler BM (2004) The Avr1b locus of *Phytophthora sojae* encodes an elicitor and a regulator required for avirulence on soybean plants carrying resistance gene Rps 1b. Mol Plant-Microbe Interact 17: 394-403.

17. Qutob D, Tedman-Jones J, Dong S, Kuflu K, Pham H, et al. (2009) Copy number variation and transcriptional polymorphisms of *Phytophthora sojae* RXLR effector genes Avr1a and Avr3a. PLoS ONE 4: e5066

18. Dong S, Qutob D, Tedman-Jones J, Kuflu K, Wang Y, et al. (2009) The *Phytophthora sojae* avirulence locus Avr3c encodes a multi-copy RXLR effector with sequence polymorphisms among pathogen strains. PLoS ONE 4: e5556

19. Matlin AJ, Clark F, Smith CWJ (2005) Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol 6: 386-398.

20. Win J, Kanneganti TD, Torto-Alalibo T, Kamoun S (2006) Computational and comparative analyses of 150 full-length cDNA sequences from the oomycete plant pathogen *Phytophthora infestans*. Fungal Genet Biol 43: 20-33.

21. Costanzo S, Ospina-Giraldo M, Deahl K, Baker C, Jones R (2007) Alternate intron processing of family 5 endoglucanase transcripts from the genus *Phytophthora*. Current Genetics 52: 115-123.

22. Shen D, Ye W, Dong S, Wang Y, Dou D (2011) Characterization of intronic structures and alternative splicing in *Phytophthora sojae* by comparative analysis of expressed sequence tags and genomic sequences. Can J Microbiol 57: 84-90.

23. Savory EA, Granke LL, Quesada-Ocampo LM, Varbanova M, Hausbeck MK, et al. (2011) The cucurbit downy mildew pathogen *Pseudoperonospora cubensis*. Mol Plant Pathol 12: 217-226.

24. Tian M, Win J, Savory E, Burkhardt A, Held M, et al. (2011) 454 Genome sequencing of *Pseudoperonospora cubensis* reveals effector proteins with a QXLR translocation motif. Mol Plant-Microbe Interact 24: 543-553.

25. Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18: 821-829.

26. Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, et al. (2008) MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18: 188-196.

27. Jiang RHY, Tripathy S, Govers F, Tyler BM (2008) 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 105: 4874-4879.

28. Savory EA, Adhikari BN, Hamilton JP, Vaillancourt B, Buell CR, et al. (2012) mRNA-Seq Analysis of the *Pseudoperonospora cubensis* Transcriptome During Cucumber (*Cucumis sativus* L.) Infection. PLoS ONE 7: e35796.

29. Win J, Morgan W, Bos J, Krasileva KV, Cano LM, et al. (2007) Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. Plant Cell 19: 2349-2369.

30. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290: 1151-1155.

31. Jack DL, Yang NM, Saier MH, Jr. (2001) The drug/metabolite transporter superfamily. Eur J Biochem 268: 3620-3639.

32. Jacobs KA, Collins-Racie LA, Colbert M, Duckett M, Golden-Fleet M, et al. (1997) A genetic selection for isolating cDNAs encoding secreted proteins. Gene 198: 289-296.

33. Oh SK, Young C, Lee M, Oliva R, Bozkurt TO, et al. (2009) In planta expression screens of *Phytophthora infestans* RXLR effectors reveal diverse phenotypes, including activation of the *Solanum bulbocastanum* disease resistance protein Rpi-blb2. Plant Cell 21: 2928.

34. Schornack S, van Damme M, Bozkurt TO, Cano LM, Smoker M, et al. (2010) Ancient class of translocated oomycete effectors targets the host nucleus. Proc Natl Acad Sci U S A 107: 17421-17426.

35. Gurlebeck D, Jahn S, Gurlebeck N, Szczesny R, Szurek B, et al. (2009) Visualization of novel virulence activities of the *Xanthomonas* type III effectors AvrBs1, AvrBs3 and AvrBs4. Mol Plant Pathol 10: 175-188.

36. Jin P, Wood MD, Wu Y, Xie Z, Katagiri F (2003) Cleavage of the *Pseudomonas syringae* type III effector AvrRpt2 requires a host factor(s) common among eukaryotes and is important for AvrRpt2 localization in the host cell. Plant Physiol 133: 1072-1082.

37. Shan L, Thara VK, Martin GB, Zhou JM, Tang X (2000) The *Pseudomonas* AvrPto protein is differentially recognized by tomato and tobacco and is localized to the plant plasma membrane. Plant Cell 12: 2323-2338.

38. Nelson BK, Cai X, Nebenführ A (2007) A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. Plant J 51: 1126-1136.

39. Chaparro-Garcia A, Wilkinson RC, Gimenez-Ibanez S, Findlay K, Coffey MD, et al. (2011) The receptor-like kinase SERK3/BAK1 is required for basal resistance against the late blight pathogen *Phytophthora infestans* in *Nicotiana benthamiana*. PLoS ONE 6: e16608.

40. Win J, Kamoun S, Jones AM (2011) Purification of effector-target protein complexes via transient expression in *Nicotiana benthamiana*. Methods Mol Biol 712: 181-194.

41. Liu T, Ye W, Ru Y, Yang X, Gu B, et al. (2011) Two host cytoplasmic effectors are required for pathogenesis of *Phytophthora sojae* by suppression of host defenses. Plant Physiol 155: 490-501.

42. Halterman DA, Chen Y, Sopee J, Berduo-Sandoval J, Sanchez-Perez A (2010) Competition between *Phytophthora infestans* effectors leads to increased aggressiveness on plants containing broad-spectrum late blight resistance. PLoS ONE 5: e10536.

43. Savory EA, Granke LL, Quesada-Ocampo LM, Varbanova M, Hausbeck MK, et al. (2011) The cucurbit downy mildew pathogen *Pseudoperonospora cubensis*. Molecular Plant Pathology 12: 217-226.

44. Keren H, Lev-Maor G, Ast G (2010) Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet 11: 345-355.

45. Franke I, Resch A, Dassler T, Maier T, Bock A (2003) YfiK from *Escherichia coli* promotes export of O-acetylserine and cysteine. J Bacteriol 185: 1161-1166.

46. Chandran D, Tai YC, Hather G, Dewdney J, Denoux C, et al. (2009) Temporal global expression data reveal known and novel salicylate-impacted processes and regulators mediating powdery mildew growth and reproduction on Arabidopsis. Plant Physiol 149: 1435-1451.

47. Torto TA (2003) EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora*. Genome Res 13: 1675-1685.

48. Rafiqi M, Gan PH, Ravensdale M, Lawrence GJ, Ellis JG, et al. (2010) Internalization of flax rust avirulence proteins into flax and tobacco cells can occur in the absence of the pathogen. Plant Cell 22: 2017-2032.

49. Huang S, Li R, Zhang Z, Li L, Gu X, et al. (2009) The genome of the cucumber, *Cucumis sativus* L. Nat Genet 41: 1275-1281.

50. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25.

51. Salamov AA, Solovyev VV (2000) Ab initio Gene Finding in Drosophila Genomic DNA. Genome Res 10: 516-522.

52. Knepper C, Savory EA, Day B (2011) Arabidopsis NDR1 is an integrin-like protein with a role in fluid loss and plasma membrane-cell wall adhesion. Plant Physiol 156: 286-300.

53. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.

54. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nature Protocols 2: 953-971.

55. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, et al. (2005) InterProScan: protein domains identifier. Nucleic Acids Res 33: W116-120.

56. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948.

57. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comp App Biosci 13: 555-556.

58. Tang H, Wang X, Bowers JE, Ming R, Alam M, et al. (2008) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome Res 18: 1944-1954.

59. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792-1797.

60. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comp App Biosci 8: 275-282.

61. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol doi: 10.1093/molbev/msr121.

CHAPTER 3

mRNA-Seq Analysis of the *Pseudoperonospora cubensis* transcriptome during cucumber (*Cucumis sativus* L.) infection

This chapter was originally published in PLoS ONE.

Savory EA*, Adhikari, BN*, Hamilton JP, Vaillancourt, B, Buell CR, and Day B (2012) mRNA-Seq Analysis of the *Pseudoperonospora cubensis* transcriptome during cucumber (*Cucumis sativus* L.) infection. PLoS ONE 7(4): e35796. doi:10.1371/journal.pone.0035796.

* These authors contributed equally to this work

Author Contributions:

Conceived and designed the experiments: EAS, BNA, CRB, and BD. Performed the experiments: EAS, BNA, JPH, and BV. Analyzed the data: EAS, BNA, JPH, CRB, and BD. Contributed reagents/materials/analysis tools: EAS, BNA, JPH, CRB, and BD. Wrote the paper: EAS, BNA, CRB, and BD.

ABSTRACT

Pseudoperonospora cubensis, an oomycete, is the causal agent of cucurbit downy mildew, and is responsible for significant losses on cucurbit crops worldwide. While other oomycete plant pathogens have been extensively studied at the molecular level, *Ps. cubensis* and the molecular basis of its interaction with cucurbit hosts has not been well examined. Here, we present the first large-scale global gene expression analysis of *Ps. cubensis* infection of a susceptible *Cucumis sativus* cultivar, 'Vlaspik', and identification of genes with putative roles in infection, growth, and pathogenicity. Using high throughput whole transcriptome sequencing, we captured differential expression of 2383 *Ps. cubensis* genes in sporangia and at 1, 2, 3, 4, 6, and 8 days post-inoculation (dpi). Additionally, comparison of *Ps. cubensis* expression profiles with expression profiles from an infection time course of the oomycete pathogen *Phytophthora infestans* on *Solanum tuberosum* revealed similarities in expression patterns of 1,576-6,806 orthologous genes suggesting a substantial degree of overlap in molecular events in virulence between the biotrophic *Ps. cubensis* and the hemi-biotrophic *P. infestans*. Co-expression analyses identified distinct modules of *Ps. cubensis* genes that were representative of early, intermediate, and late infection stages. Collectively, these expression data have advanced our understanding of key molecular and genetic events in the virulence of *Ps. cubensis* and thus, provides a foundation for identifying mechanism(s) by which to engineer or effect resistance in the host.

INTRODUCTION

The phytopathogenic oomycete *Pseudoperonospora cubensis,* the causative agent of cucurbit downy mildew [1,2], infects a wide range of cucurbits, including cucumber (*Cucumis sativus* L.), squash (*Cucurbita* spp.), and melon (*Cucumis melo* L.). As an obligate biotroph, *Ps. cubensis* is dependent on its host for both reproduction and dispersal, and as such, has evolved a highly specialized host range limited to members of the *Cucurbitaceae*. At present, downy mildew is the most important foliar disease of cucurbits, affecting cucurbit production throughout the world [1,2]. Under favorable conditions, *Ps. cubensis* is capable of infecting and defoliating a field in less than two weeks, and as a result, is responsible for devastating economic losses. For more than 50 years, control of downy mildew on cucumber in the U.S. was maintained through genetic resistance; however, since 2004, the likely introduction of a new pathotype into U.S. pathogen populations has resulted in a loss of this resistance [1]. While minimal knowledge of the genetic variation within *Ps. cubensis* exists - specifically related to virulence, pathogenicity, and host specificity among physiological races - the genetic basis of these processes, and the underlying mechanism(s) associated with infection have not been elucidated [1,2,3,4].

To date, analyses of the *Ps. cubensis*-*C. sativus* interaction have been limited to the identification of the aforementioned physiological races, and have largely focused on the utilization of variation in host specificity for the identification and classification of pathotypes [3,5]. To this end, six physiological pathotypes, or races, have been identified within populations in the U.S., Israel, and Japan, as well as additional races

throughout Europe [1,2,4]. In the U.S., increased disease pressure on cucumber production since 2004 is hypothesized to be the result of the introduction of a new, more virulent pathotype, capable of overcoming the downy mildew resistance gene *dm-1*, that has been widely incorporated into commercial cucumber varieties since the 1940's [6]. While genetic analyses such as Amplified Fragment Length Polymorphism have been used to differentiate these physiological races [4] and some effort has been made to refine the species within *Pseudoperonospora* [6,7], there is limited information available about pathogenicity or virulence genes in *Ps. cubensis* or the molecular-genetic basis of resistance to this pathogen in the cucurbits.

Recent work generated the first sequence assembly of the *Ps. cubensis* genome and subsequent *in silico* analysis has identified candidate effector proteins that may have either virulence or avirulence roles in *Ps. cubensis* infection [8,9]. Structurally, oomycete effector proteins display a modular organization, consisting of a N-terminal signal peptide, a conserved RXLR (*Arg-X-Leu-Arg*, where "X" is any amino acid) translocation motif, followed by a variable C-terminal effector domain [10]. In short, it is the function and activity of the variable C-terminal effector domain that drives the activity of these molecules [10,11]. A set of 61 candidate effectors were identified in the first draft of the *Ps. cubensis* genome [8] and included a large class of variants with sequence similarity to the canonical RXLR motif. Specifically, the function of a QXLR-containing effector, designated *Pc*QNE, was characterized and shown to be a member of a large family of *Ps. cubensis* QXLR nuclear-localized effectors, which was up-regulated during infection of cucumber [8]. Additionally, internalization of *Pc*QNE into the host cell was shown to

require the QXLR-EER motif, thereby establishing a basic functional homology with the well-characterized *Phytophthora* spp. effector proteins [8]. While this work serves as a substantial development in understanding the genetic basis for pathogenicity in *Ps. cubensis*, additional work is needed to identify and characterize additional effectors and other proteins involved in establishment of infection and pathogen proliferation.

The accessibility of oomycete pathogen genome sequences, combined with gene expression data from both pathogen and host throughout the course of infection, can serve as a basis for identification and curation of genes that may have important roles in both virulence and avirulence [12,13,14,15]. To date, oomycete RXLR effectors have been demonstrated to suppress basal host resistance [16,17], as well as to activate effector-triggered immunity (ETI) [18,19,20,21]. In addition to the RXLR class, other cytoplasmically-localized effectors have been identified in *Phytophthora* spp. [11]. The Crinkler (CRN) family, for example, has a conserved LXLFLAK motif necessary for translocation into the host cytoplasm and subsequent import into plant nuclei where they elicit a rapid cell death response [22,23]. Finally, oomycete effectors have also been shown to function within the host apoplast, including functions as enzyme inhibitors [24,25,26,27,28], small cysteine-rich proteins [22,29,30], the Nep1-like family of proteins [31,32], and CBEL (Cellulose Binding, Elicitor, and Lectin-like) proteins [33,34].

The initial stages of pathogen infection of a plant host involve adhesion, penetration, and invasive growth within the host cell tissue. As such, cell wall degrading enzymes

(CWDE), such as endoglucanases, cutinases, cellulases, and β-glucanases have evolved as essential components of an oomycete's repertoire for cell wall penetration [20]. Numerous CWDE have been identified computationally from the genomic sequences of several plant pathogenic oomycetes, including *Phytophthora sojae, Phytophthora ramorum, Hyaloperonospora arabidopsidis,* and *Pythium ultimum* [30,35,36]. In *P. sojae*, members of the family 5 and family 12 endoglucanases have been shown to be up-regulated during early stages of infection [37,38]. However, in *H. arabidopsidis*, which causes downy mildew of *Arabidopsis thaliana*, CWDE-encoding mRNAs are reduced [36]. This could indicate an adaptation in downy mildew pathogens for evasion of recognition by their host, as break-down products from plant cell wall components can function as elicitors of defense responses [39].

Recent advancements in sequencing technologies have led to an explosive growth in the analysis of *in planta*-expressed genes of biotrophic plant pathogens [12,40,41,42,43,44]. In the current study, we present the first global gene expression analysis of the infection stages of cucumber by the obligate oomycete pathogen *Ps. cubensis*, the causal agent of cucurbit downy mildew. Through the analysis of a susceptible cucumber cultivar interaction, we describe the identification of genes with putative roles in infection, growth and pathogenicity. Using next-generation sequencing technology, we assessed gene expression in *Ps. cubensis* in sporangia and at six time points of infection. By combining visual assessment of symptoms with light microscopy to monitor infection stages as well as minimizing collection of non-inoculated tissues, we were able to capture expression of 7,821 *Ps. cubensis* genes ranging from 159 genes at

1 days post inoculation (dpi) to 7,698 at 8 dpi. In total, this work represents a comprehensive examination of the key infection stages of *Ps. cubensis* growth and development. In total, the work described herein provides a foundation for further dissection of genes relevant to virulence in this obligate phytopathogen.

RESULTS AND DISCUSSION

Characterization and sampling of *Ps. cubensis* infection stages

While *Ps. cubensis* is a major pathogen of cucumber and other cucurbits, limited resources describing the infection process and/or virulence determinants of this obligate oomycete are available. In the current study, we sought to identify *Ps. cubensis* gene expression from both purified sporangia, as well as from a time course of infected cucumber tissues, representing a wide range of infection stages from 1 to 8 dpi. In total, our goal was to gain a broad perspective of *in planta* gene expression during infection of a susceptible cucumber host and to correlate this expression with the development of outwardly visible symptoms, as well as the development of microscopic pathogen infection structures. Like other phytopathogenic downy mildews and biotrophic fungi, *Ps. cubensis* is non-culturable, and proliferates and reproduces only on a susceptible cucurbit host. As with previously published reports on analyzing gene expression in biotrophic phytopathogens, optimization of sampling techniques is key to maximize pathogen tissue compared to host, particularly at early stages of infection (Figure 3.1) [12,43,44,45].

Plants were inoculated on the abaxial leaf surface with purified *Ps. cubensis* sporangia, and samples were collected using a cork borer, minimizing the amount of non-infected tissue in each sample (Figure 3.1). Initial symptoms of downy mildew infection can be observed on the abaxial leaf surface at 1-3 dpi as water soaking at the site of inoculation, while no visual symptoms are apparent on the upper leaf surface (Figure 3.2). At 1 dpi, zoospores were encysted upon stomata on the lower leaf surface, and by 2 dpi, appressoria and initial penetration hyphae were visible beneath stomata. The yellow angular lesions typical of cucurbit downy mildew were apparent on the upper leaf surface by 4 dpi, and over time, became more chlorotic and necrotic as the infection progressed. By 3 to 4 dpi, multiple haustoria formed within the mesophyll layer.

mRNA-Seq data analyses

Expression profiling of *Ps. cubensis* sporangia, as well as infection stages at six time points of cucumber infection, were performed using mRNA-Seq. For each time point, two biological replicates were sequenced. The total number of reads produced for each time point ranged from 55 to 59 million reads, with a median of 57 million reads. Reads were mapped to the *Ps. cubensis* genome which was generated by assembly of Illumina next generation reads; in total the *Ps. cubensis* genome encompasses 67.9 Mb, with 23,519 protein coding genes and 23,522 gene models (AHJF00000000;

Figure 3.1

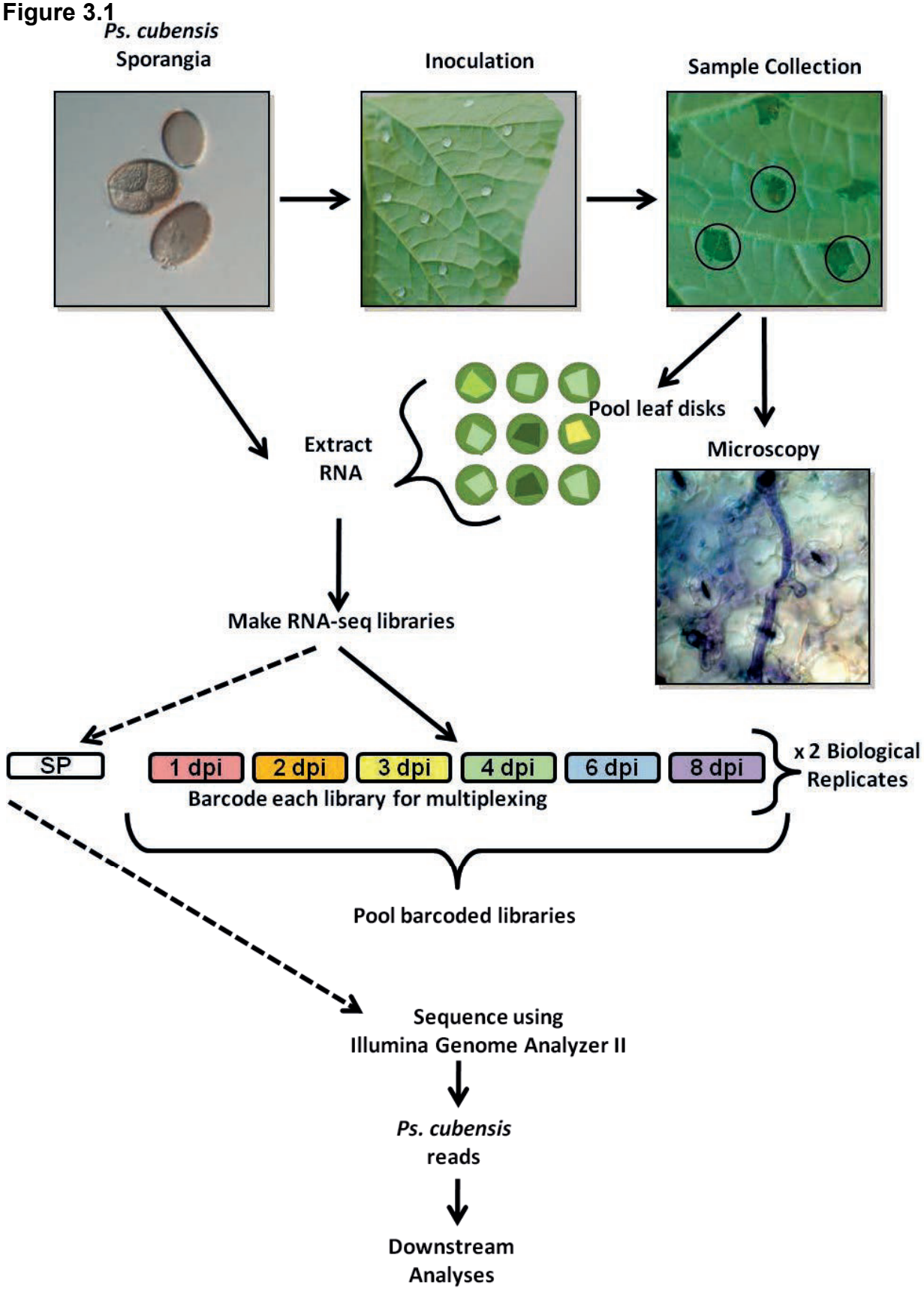
Ps. cubensis
Sporangia
Inoculation
Sample Collection
Pool leaf disks
Extract
RNA
Microscopy
Make RNA-seq libraries
SP
1 dpi
2 dpi
3 dpi
4 dpi
6 dpi
8 dpi
x 2 Biological
Replicates
Barcode each library for multiplexing
Pool barcoded libraries
Sequence using
Illumina Genome Analyzer II
Ps. cubensis
reads
Downstream
Analyses

Figure 3.1 (cont'd) Experimental design and sample collection. A 1 x 10e5 sporangia/ml solution of *Pseudoperonospora cubensis* was used to inoculate the abaxial leaf surface of cucumber cultivar 'Vlaspik'. Samples were collected using a #3 cork borer to minimize uninfected tissue (black circles) at 1, 2, 3, 4, 6, and 8 days post-inoculation (dpi). Leaf disks were used for microscopic analysis of infection stages or pooled for RNA extraction. RNA-Seq libraries were made for each time point from 2 biological replicates. Within a biological replicate, libraries were barcoded and sequenced in multiple lanes. The sporangia-only library (SP) was not barcoded and was sequenced on its own.

[9]). Of the total reads generated, for each time point, approximately 1.6 to 6.4 million (3-12% of the total; Figure 3.3A) mapped to the *Ps. cubensis* genome. In turn, a majority of reads in each sample were of host origin, and mapped to the cucumber genome (see accompanying paper, [46]) (Figure S3.1). Through this analysis, we found that there was no significant difference in the total number of combined reads from different time points ($p > 0.80$); however, the number of *Ps. cubensis* genes expressed at each time point was significantly different for all time point comparisons ($p < 0.05$; Figure 3.3A). To assess the experimental variation attributable to biological variation, we compared the gene expression pattern of the genes expressed in both of our biological replicates. In total, our experiments showed very high levels of correlation for biological replicates (in all cases examined, Pearson's Correlation Coefficient (PCC) > 0.94; Figure S3.2), indicating that our sampling, assay, and analysis methods are robust.

To evaluate the effect of sampling depth on gene expression detection and to assess whether we have adequately sampled the mixed mRNA-Seq read pool for *Ps. cubensis* transcripts, subsets of five to 30 million reads were randomly selected from the total

Figure 3.2

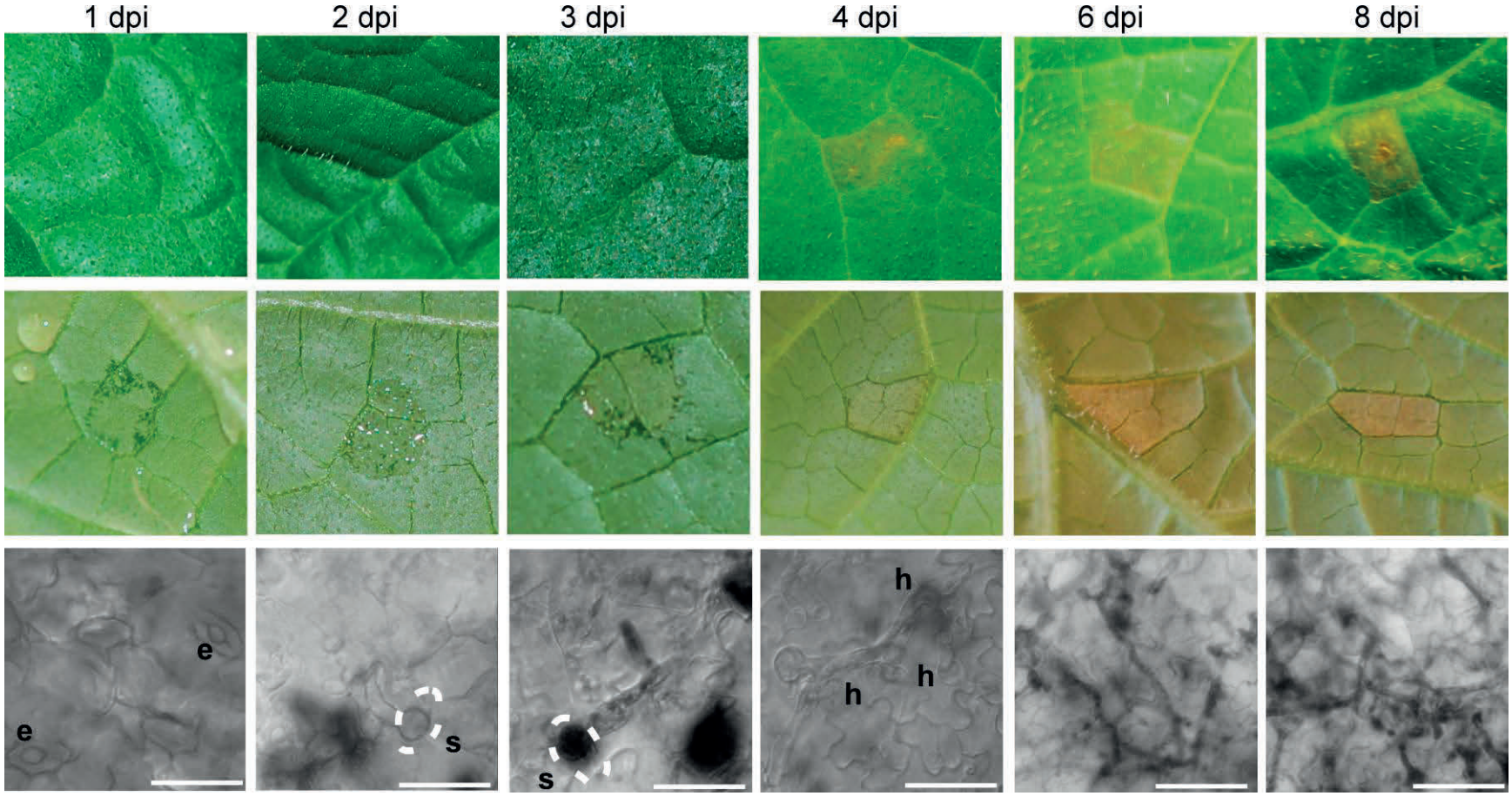

Figure 3.2 Symptoms and microscopy images of *Ps. cubensis* infected *Cucumis sativus* cultivar 'Vlaspik' of time points used for transcriptome analysis. Symptom images were collected of the adaxial (top row) and abaxial (middle row) at 1, 2, 3, 4, 6, and 8 days post-inoculation (dpi). Microscopy (bottom row) to assess stages of *Ps. cubensis* invasion were collected from the same time points using ethanol-cleared, trypan blue stained samples. Scale bars at 1-4 dpi are 25 um. Scale bars at 6 and 8 dpi are 50 um. Dotted lines represent position of stomata relative to the pathogen structure. e = encysted zoospore. s = stomate. h = haustorium.

read pool from each time point and mapped to the *Ps. cubensis* genome. The simulation experiment showed a clear positive relationship between sampling depth and number of expressed genes at the lower to medium sequencing depth (5 to 20 million reads; Figure 3.3B). With the exception of 1 dpi in which few genes are expressed, after 20-25 million reads, the number of expressed genes begin to plateau, corresponding to the minimum sampling depth of all libraries sequenced in this study.

mRNA-Seq transcriptome profiles

In concordance with our visual assessment (Figure 3.2) of pathogen growth throughout the time course, our analyses showed a diversity of transcriptional changes in *Ps. cubensis*, as well as a correlation between gene expression levels and similar stages of pathogen growth. In support of this, we identified 7,821 genes expressed at different time points of infection (Table S3.1, available at www.plosone.org, e35796) and 129 of those genes (Table S3.2, available at www.plosone.org, e35796), mostly housekeeping, were expressed throughout all time points. Analyses of the top 20 highly expressed genes showed that genes expressed at earlier time points have substantially higher FPKM (fragments per kilobase pair of exon model per million fragments mapped) values than the genes expressed at later time points, consistent with the fewer numbers of genes expressed in the early stages of expression and saturation of detection of *Ps. cubensis* expression with our sampling depth (Table S3.3, available at www.plosone.org, e35796). For all time points analyzed, the minimum FPKM value was zero but the maximum FPKM values ranged from 8,528 at 8 dpi to 270,121 at 1 dpi

Figure 3.3

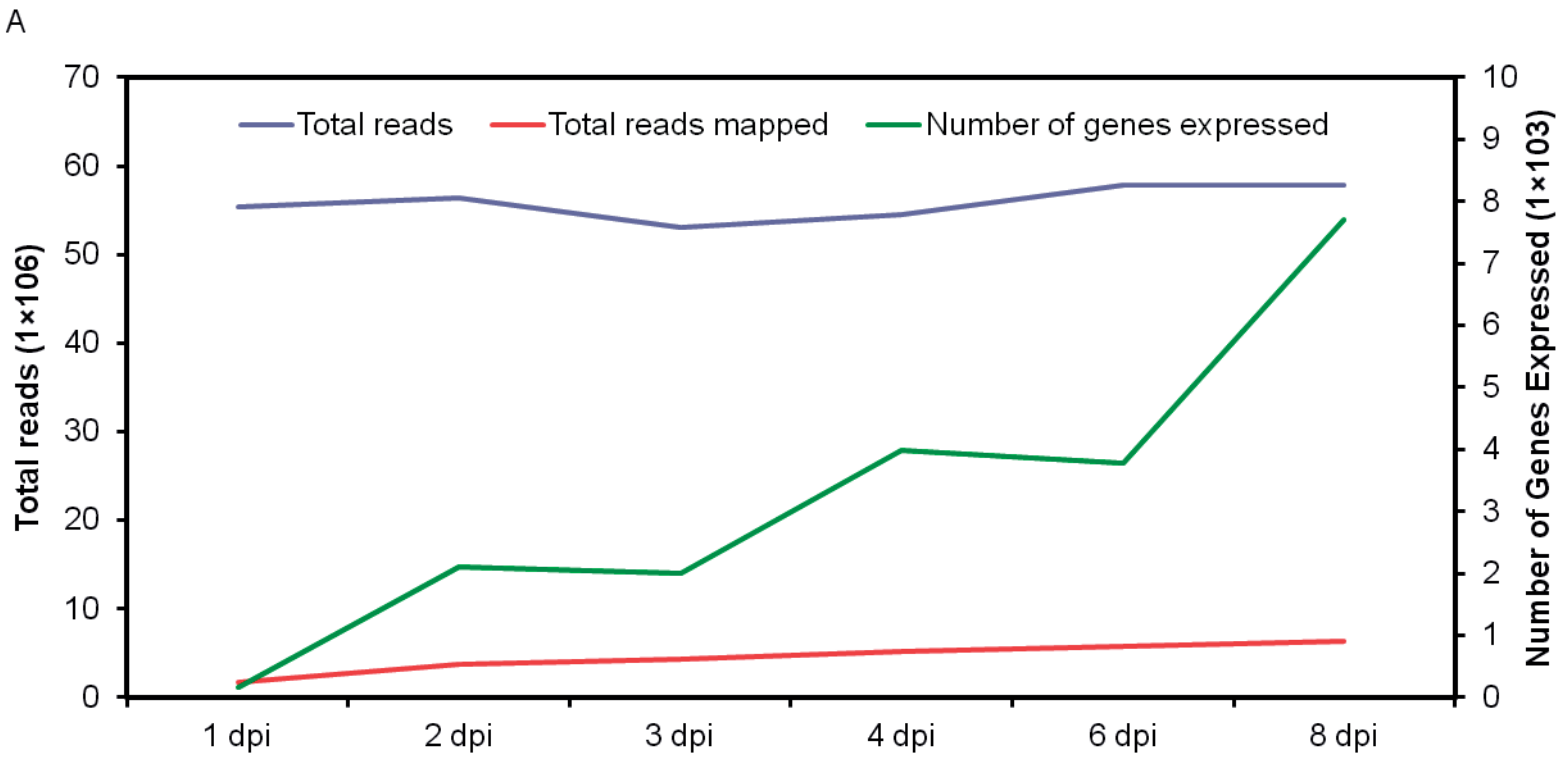
A
Total reads
Total reads mapped
Number of genes expressed
Total reads (1×106)
Number of Genes Expressed (1×103)
0
10
20
30
40
50
60
70
0
1
2
3
4
5
6
7
8
9
10
1 dpi
2 dpi
3 dpi
4 dpi
6 dpi
8 dpi

Figure 3.3 (cont'd)

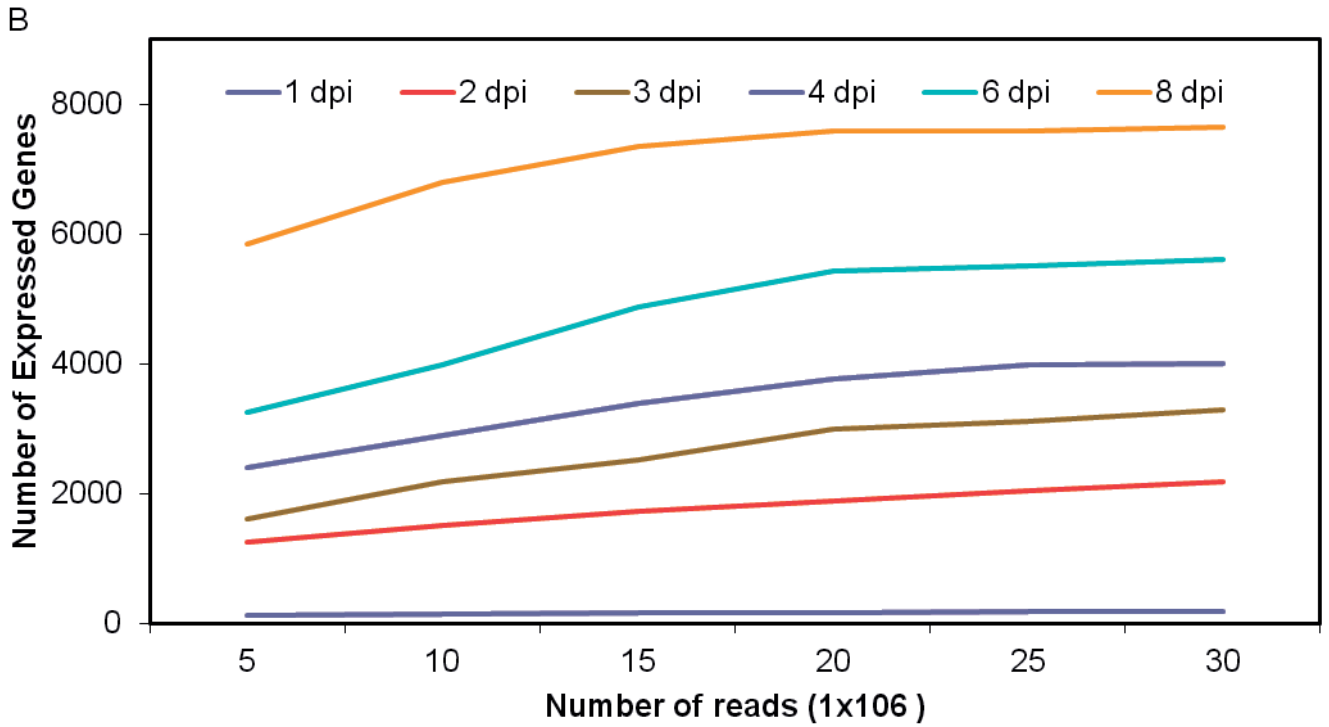

Figure 3.3 Number of total RNA-seq reads, reads mapped, and number of genes expressed. (A). Total number of reads, number of reads mapped to the *Ps. cubensis* genome, and number of genes expressed by *Ps. cubensis* at different time points are shown. Reads were mapped using Bowtie version 0.12.5 and TopHat version 1.2.0. Fragments per kilobase pair of exon model per million fragments mapped (FPKM) values were calculated using Cufflinks version 0.9.3. Genes were considered expressed if the FPKM and 95% confidence interval lower boundary FPKM value was greater than 0.001 and zero, respectively. dpi = days post-inoculation. (B) Comparison between number of expressed genes detected and sampling depth. For all time points 5, 10, 15, 20, 25, and 30 million reads were randomly selected from the total pool of reads from different time points. Read mapping and expression abundances are as described in panel 3A.

(Table S3.3, available at www.plosone.org, e35796). The differences in transcriptome profiles are clearly visible in correlation and cluster analyses between the sampled time points. The Pearson Correlation Coefficient (PCC) values for comparisons of different time points ranged from 0.26 (1 dpi vs. 8 dpi) to 0.79 (3 dpi vs. 4 dpi) (Figure 3.4). Corresponding with our visual assessment of pathogen infection stages showing similar growth at 2 and 3 dpi (i.e., penetration and initial hyphal growth, Figure 3.2), gene expression patterns at 2 dpi are strongly associated with that of 3 dpi. Similarly, genes expressed at 6 dpi showed high correlation with the genes from 8 dpi, which represent comparable stages of pathogen growth and proliferation in the mesophyll (Figure 3.2). Additionally, at 1 dpi when encystment of zoospores is occurring, we observed a poor correlation (PCC ranged from 0.26 to 0.45) between expression at that time point with any other time point, likely due to the unique set of genes that would be involved in this process. Interestingly, gene expression at 3 dpi was highly correlated with other time points (PCC ranged from 0.45 to 0.79), suggesting that events occurring at 3 dpi may represent a transition phase between early and late infection.

Differential gene expression

To identify genes specifically involved in distinct stages of pathogen infection and development, and to assess gene expression pattern changes over the course of infection, we next evaluated differentially expressed genes between all time points. To provide context for the differential expression, we included expression data from mRNA-Seq analysis of sporangia in which 8,254 *Ps. cubensis* genes were expressed (Table

Figure 3.4

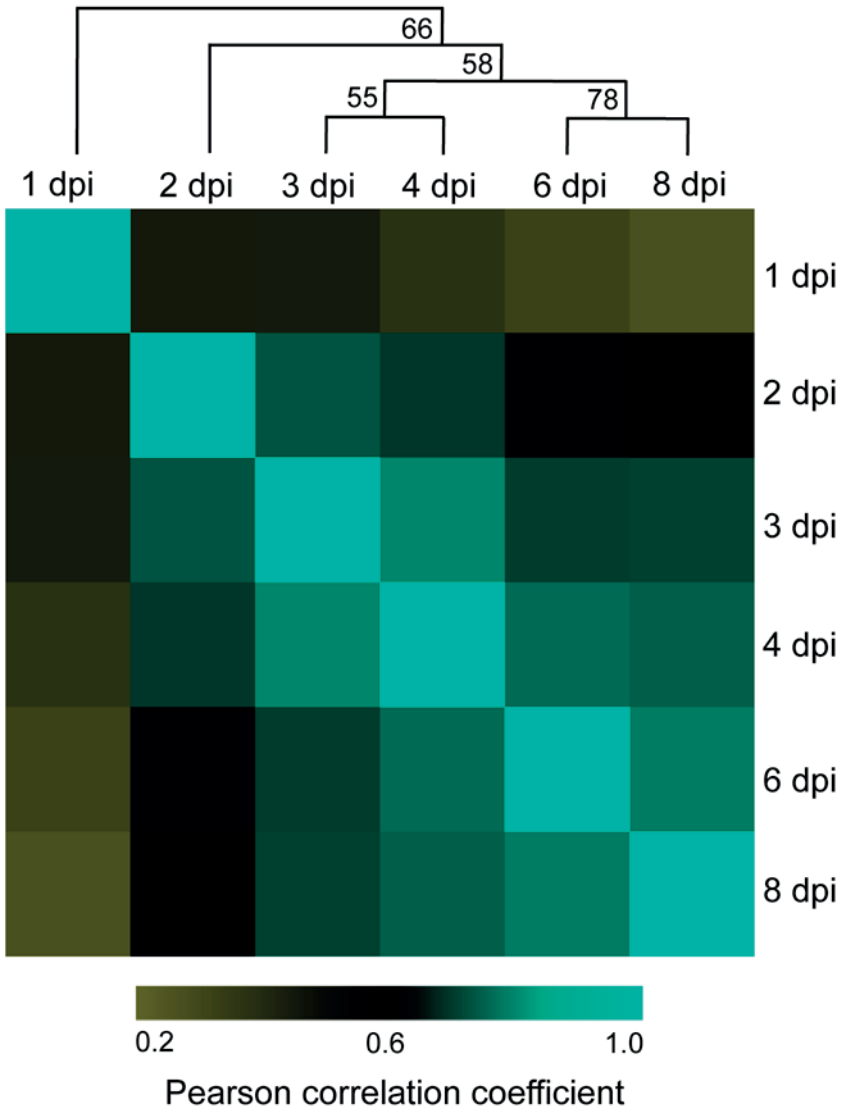

Figure 3.4 Correlation matrix of *Pseudoperonospora cubensis* expression profiles throughout a time course of *Cucumis sativus* infection. Normalized transcript abundances for 7,821 genes were calculated in units of fragments per kilobase pair of exon model per million fragments mapped (FPKM) with Cufflinks version 0.9.3. Pearson product-moment correlations (PCC) of log2 FPKM values were calculated for all pair-wise combinations using R. PCCs were clustered using hierarchical clustering with a Pearson correlation distance metric and average linkage using Multiple Experiment Viewer Software version 4.5. The bootstrap support values shown on tree nodes were obtained from 1000 bootstrap replicates. dpi = days post-inoculation.

S3.4, available at www.plosone.org, e35796). In concordance with the similarities observed during our visualization of pathogen growth, comparisons with the least number of differentially expressed genes are those between early time points, with only 147 (2%) genes differentially expressed between 1 and 2 dpi (Table 3.1). Additionally, 1 and 2 dpi had fewer differentially expressed genes compared to sporangia than the later time points. Of all the combinations tested, 1 dpi had the highest percentage of differentially expressed genes across all pair-wise comparisons, despite having the lowest number of genes expressed, indicating that the events occurring at this stage of infection are unique among the time course. This corresponds both with our cluster analysis above (Figure 3.4) and our microscopic analysis of infection (Figure 3.2). Interestingly, despite the high correlation between expression patterns at 2 and 3 dpi seen using cluster analysis, there were a large number of genes differentially expressed between 2 and 3 dpi. This is additionally supportive of our hypothesis that 3 dpi is a transition phase between early and late infection. Not surprisingly, the highest number of differentially expressed genes were observed in comparison of all other time points and sporangia with 8 dpi, suggestive of an advanced stage of the infection process and a likely transition to processes involved in sporulation.

Expression of genes involved in virulence and pathogenicity

As an obligate biotroph, *Ps. cubensis* must evade and/or overcome basal plant defense responses, as well as effector-triggered immunity, in order to establish growth, proliferate and reproduce within its host. This, as in other phytopathogens, is likely

Table 3.1 Number of differentially expressed genes between each combination of time points and sporangia. Differential expression analysis was conducted using the CuffDiff program in Cufflinks version 0.9.3 using the *Pseudoperonospora cubensis* annotation with a false discovery rate of 0.01.

	2 dpi[†]	3 dpi	4 dpi	6 dpi	8 dpi	Sporangia
1 dpi	147 (50%)[§]	193 (58%)	189 (57%)	175 (60%)	192 (59%)	177 (50%)
2 dpi		848 (28%)	329 (10%)	306 (7%)	898 (19%)	246 (16%)
3 dpi			560 (14%)	301 (7%)	891 (17%)	391 (13%)
4 dpi				342 (7%)	820 (16%)	425 (14%)
6 dpi					644 (10%)	559 (15%)
8 dpi						1,301 (32%)

[†]dpi = days post-inoculation.
[§]Numbers in parenthesis indicate the percent of significantly different tests out of the total number of tests that could be performed for each pairwise comparison.

achieved through the secretion of specific effector proteins that function within the host apoplast to interfere with extracellular plant defense responses, such as the activity of glucanases and proteases or cytoplasmically to suppress defense responses. Thus, the identification and characterization of the temporal expression of pathogen-associated genes throughout the course of infection can assist in the identification of secreted effectors that allow for both the promotion of disease, as well as the avoidance of host recognition. In support of this, we identified a suite of 271 candidate RXLR-type effectors within the *Ps. cubensis* genome [9] with 20 possible amino acid substitutions at position R1, including R and Q. In the current study, we analyzed the expression distribution of all 271-candidate effectors, as well as predicted Crinkler (CRN) effectors,

over our time course of infection. As shown in Figure 3.5, the greatest number of expressed candidate effectors have an ASLR (Ala-Ser-Leu-Arg) motif, and are expressed at 2-8 dpi; candidate effectors with RXLR or QXLR motifs were expressed at every time point. Finally, as noted above, the expanded repertoire at the conserved RXLR motif in *Ps. cubensis,* with a total of 20 amino acids represented at the R1 position, represents a diversity in RXLR-type effectors previously undescribed. The simplest explanation for this expansion at R1 is supported by the hypothesis that RXLR-type effectors may play a role in host range, and that an expanded effector repertoire may impart plasticity. Moving forward, an extensive functional characterization of these RXLR-type effectors will provide insight into both pathogen virulence and host range specificity. Nonetheless, our data suggest that *Ps. cubensis* possesses a potentially highly expanded virulence capacity, of which, we have determined the expression of 271 RXLR-type effectors over an extensive time-course of susceptibility and disease elicitation in cucumber.

Gene families encoding host-targeted hydrolytic enzymes acting on plant proteinases, lipases, and several sugar-cleaving enzymes (carbohydrate active enzymes; CAZymes) were highly expressed in *Ps. cubensis* at 4 to 8 dpi, suggesting a possible role during infection and proliferation (Figure 3.6). Comparison of glycoside hydrolase (GH), glycosyltransferases (GT), polysaccharide lyase (PL), pectin esterase (PE), and carbohydrate esterase (CE) encoding genes revealed significant differences in number that were expressed as well as diversity across different time points. In total, 178 GH, 135 GT, 2 CE, and 15 PE were expressed throughout all the time points sampled

Figure 3.5

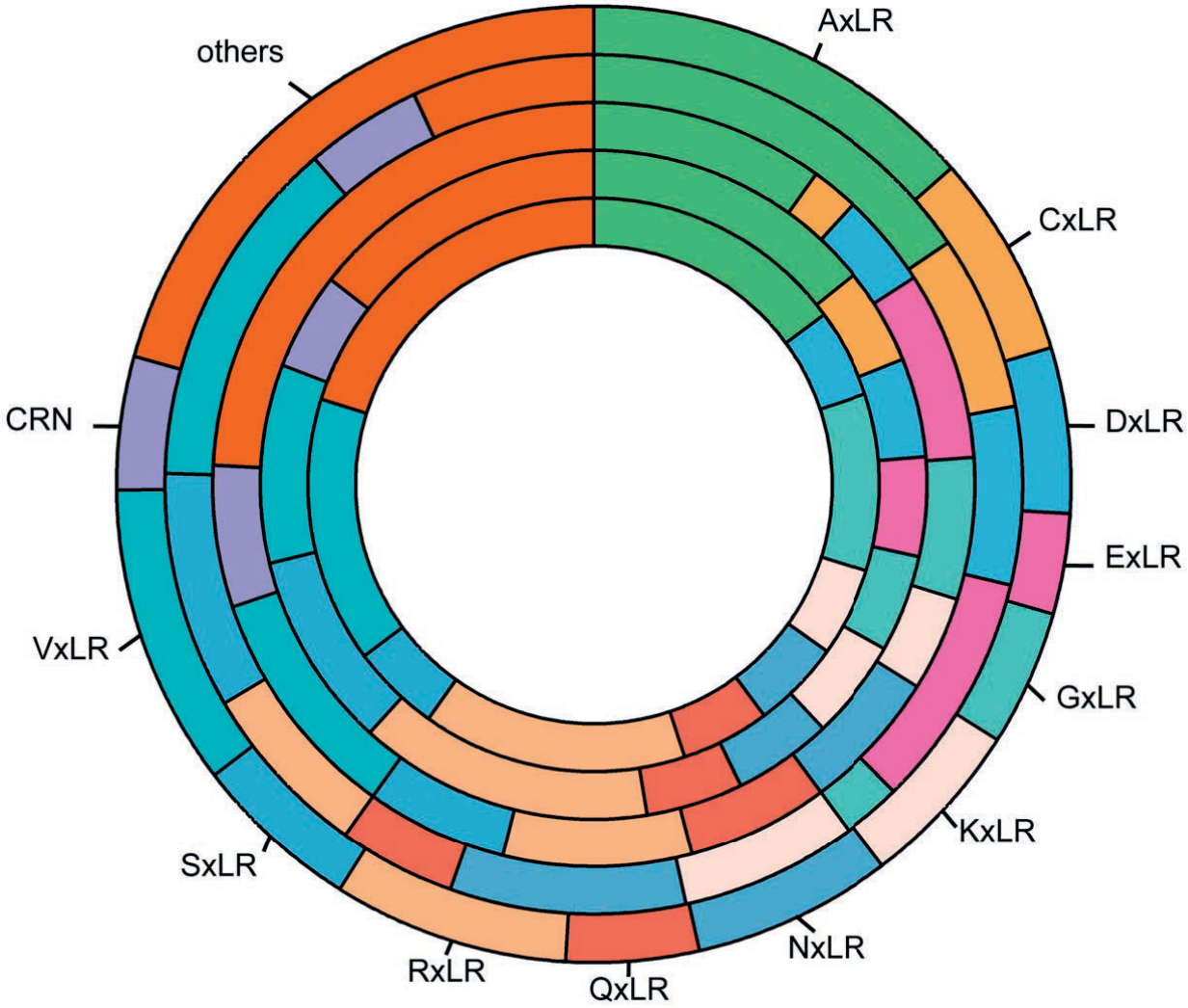

Figure 3.5 Candidate effectors expressed at different timepoints. From inner to outermost circles; 2 dpi, 3 dpi, 4 dpi, 6 dpi and 8 dpi. CRN = Crinkler effectors. dpi = days post-inoculation.

(Figure 3.6). GH was the most represented family, with expression of 30-78 members followed by GT (17-27 members expressed). The most represented GH families identified were GH3 and GH5, while GT20 and GT48 were the most represented among all GTs. Additionally, substantial differences were observed in the types of CAZymes expressed across different time points. For example, several members of GH (GH family 7, 12, and 31), GT (GT family 1), CE (CE family 5), and PL family were absent in

Figure 3.6

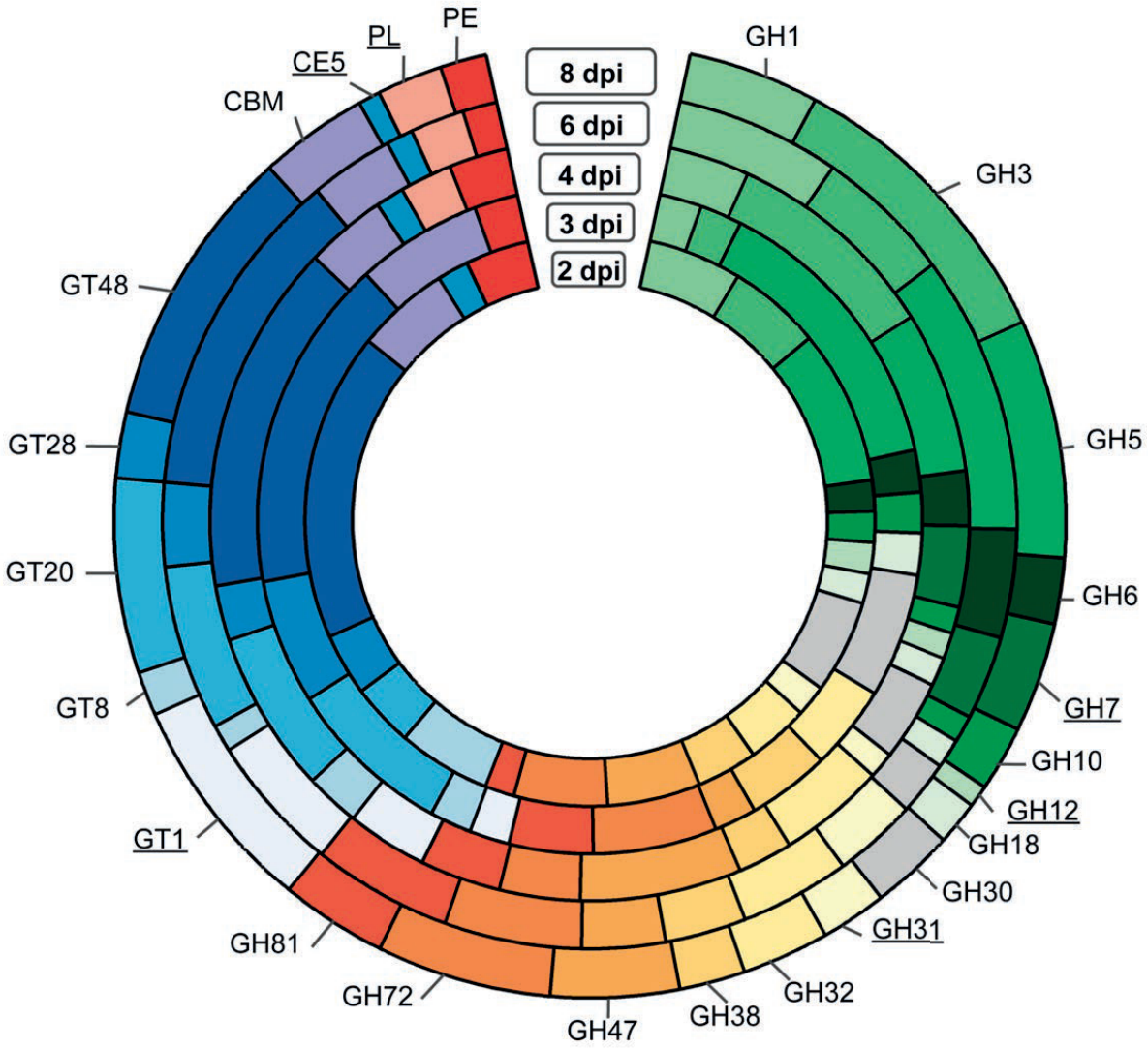

Figure 3.6 CAZymes in *Pseudoperonospora cubensis* expressed during infection on *Cucumis sativus*. The CAZymes coding genes in the *Ps. cubensis* genome were annotated using CAZymes Analysis Toolkit- CAT according to the CAZy database in combination with protein family domain analyses. Gene families absent in at least one time point are underlined. CBM = carbohydrate binding module. CE = carbohydrate esterase. GH = glycoside hydrolase. GT = glycosyl transferase. PE = pectin esterase. PL = polysaccharide lyase. dpi = days post-inoculation.

early time points (i.e., 2 and 3 dpi), yet were expressed at 4 to 8 dpi, suggesting a possible role during the later stage of infection. GH family 12 endoglucanases as well as CE family 5 cutinases have been previously implicated as having a role in infection by *Phytophthora* spp. [38,47,48].

Comparison to genes induced during *P. infestans* infection of potato

The comparison of gene expression patterns between pathogens during infection of their susceptible hosts can allow for identification of common genes that are specifically involved in pathogenesis, as well as enable the discovery of genes unique to either species. To this end, we chose to compare the gene expression pattern of *Ps. cubensis* during infection to that of another economically important oomycete pathogen, *P. infestans*, during the infection of potato, *Solanum tuberosum.* Using clustering analysis of protein coding genes from both pathogens, we identified 7,374 single copy orthologous genes between these two oomycetes. We then compared the gene expression values obtained from our study (PCU) with those from microarray-based expression profiling [10] experiments with *P. infestans-S. tuberosum* (PITG). Spearman rank correlation coefficients (SCCs) of log2 expression values were calculated between the single copy orthologs at all time points in the two datasets; between 1,576 and 5,581 genes were included in the pair-wise comparisons (Figure 3.7). The SCC values among all comparisons ranged from 0.12 to 0.76 (Table S3.5, available at www.plosone.org, e35796). Comparisons between time points reflecting similar stages of pathogen infection showed higher overall correlations (0.29 to 0.76) as compared to

Figure 3.7

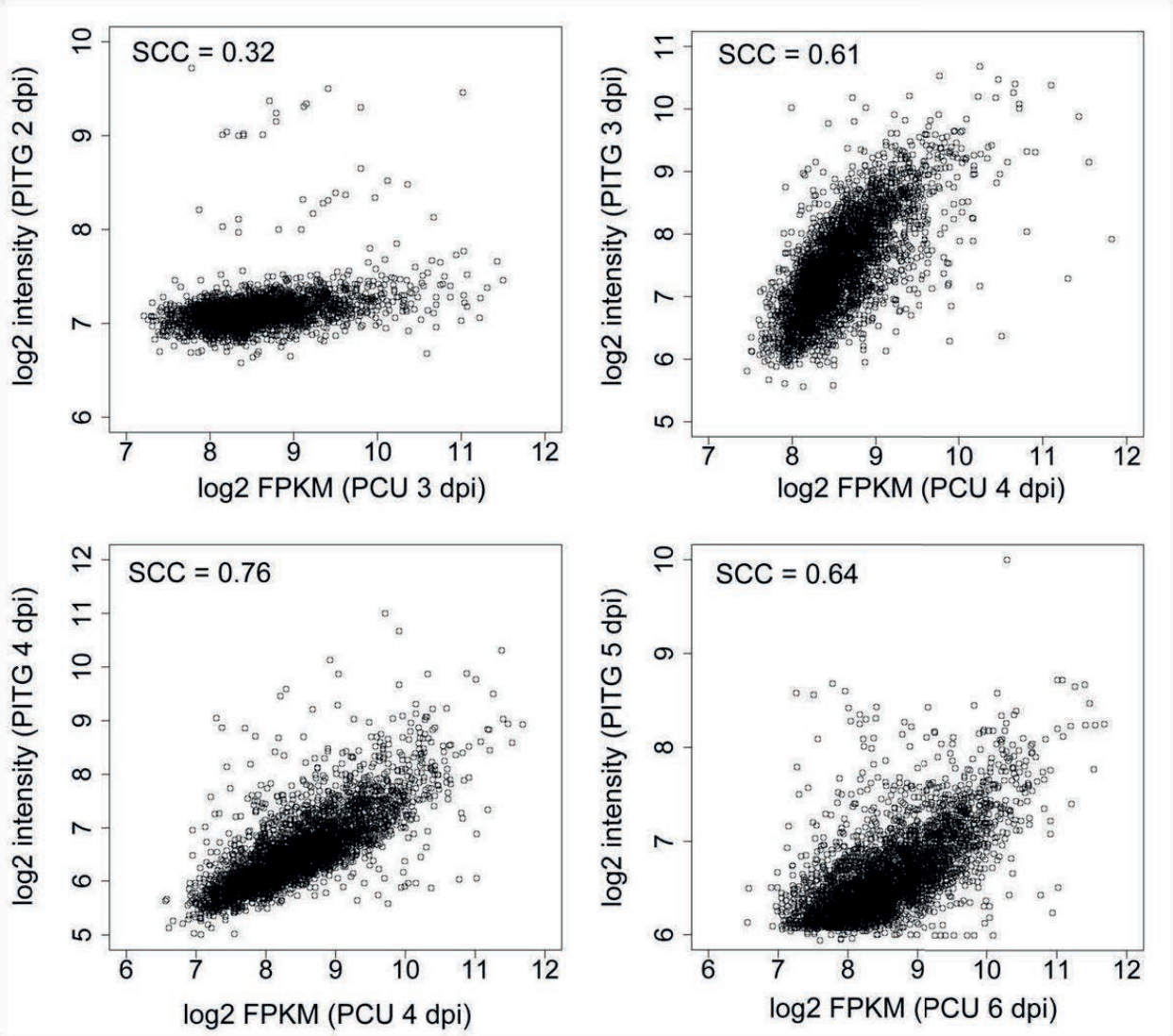

Figure 3.7 Comparison of ribonucleic acid sequencing (RNA-seq) and microarray expression patterns. Microarray expression profiles were obtained from time course analyses of genes expressed in *P. infestans* (PITG) during infection on *S. tuberosum*. Single copy orthologous genes between *P. infestans* and *Ps. cubensis* (PCU) were identified using OrthoMCL. Log2 transformed expression values of single copy orthologous genes present in both the *Ps. cubensis* (log2 of fragments per kilobase pair of exon model per million fragments mapped [FPKM]) and *P. infestans* (log2 intensity) datasets are shown as scatter plots. SCC Spearman correlation coefficient, dpi = days post-inoculation

comparisons between dissimilar time points (0.12 to 0.47). The most highly correlated comparisons were those between genes expressed in *Ps. cubensis* at 4 dpi and *P. infestans* at 4 dpi (SCC = 0.76). In *P. infestans* infection on potato, days 2-4 correspond to haustoria formation [10]; likewise, extensive formation of haustoria by *Ps. cubensis* was observed at 4 dpi (Figure 3.2). Correspondingly, genes expressed at 4 dpi include a haustorium-specific membrane protein, secreted RXLR proteins, as well as an amino acid transporter, which could possibly be involved in nutrient uptake *via* the haustorium to *Ps. cubensis*. Gene expression was also highly correlated (SCC = 0.64) between *Ps. cubensis* 6 dpi samples and *P. infestans* 5 dpi samples. At 6 dpi, symptoms of *Ps. cubensis* infection on cucumber manifest as chlorotic yellow lesions (Figure 3.2). Similarly, at 5 dpi, *Ph. infestans* has entered the mycelial necrotrophic growth stage, showing a similar chlorotic phenotype on its host [10].

Gene co-expression network analyses

Correlation analyses in which associations between gene expression patterns are identified are valuable for inferring common function and/or regulatory relationships [49]. In this study, we were primarily interested in identifying genes that are involved in both establishment and maintenance of *Ps. cubensis* infection, as well as those specifically involved in virulence. To this end, we constructed gene modules to identify highly co-expressed genes, where all members of a module are more highly correlated with each other than to genes outside the module. Using a Coefficient of Variation (CV) cutoff of 1.0, 4,195 genes from an initial total of 7,821 expressed genes were retained for

downstream analyses. Using Weighted Gene Correlation Network Analysis (WGCNA) [50], 3,146 genes were assigned to six different gene modules (Modules 1 to 6) containing 107 to 1,312 genes; 1,049 genes were not assigned to any module (Table S3.6, available at www.plosone.org, e35796; Figure S3.3). Eigengenes [51] were calculated for each module and displayed in a heat map (Figure 3.8) revealing discrete gene expression patterns across different time points.

As described above, the 1 dpi sample represents both an important initial stage in the infection process, as well as a unique gene expression profile among the infection time points analyzed. This is additionally reflected in Module 1, which contains 146 genes that are highly expressed at 1 dpi, including genes involved in pathogenesis and transport (Figure 3.8, Figure S3.3, Table S3.6, available at www.plosone.org, e35796). The genes in this module are also expressed at 3 dpi, indicating that there may be some similarities between processes involved in zoospore adhesion and encystment, and initiation of haustoria formation. Genes expressed in Module 2 could represent those processes involved in the transition from early to late stages of infection and that are involved in the initial suppression of host defenses and establishment. This module, which contains 508 genes, expressed at 2, 3, and 4 dpi, represents gene expression occurring during initial penetration through the stomata into the host tissue, hyphal growth, and initiation of haustoria formation. It includes genes such as candidate RXLR-type effectors, glucanase inhibitors, CRNs, and a haustorium-specific membrane protein (Table S3.6, available at www.plosone.org, e35796), similar to what has been observed

Figure 3.8

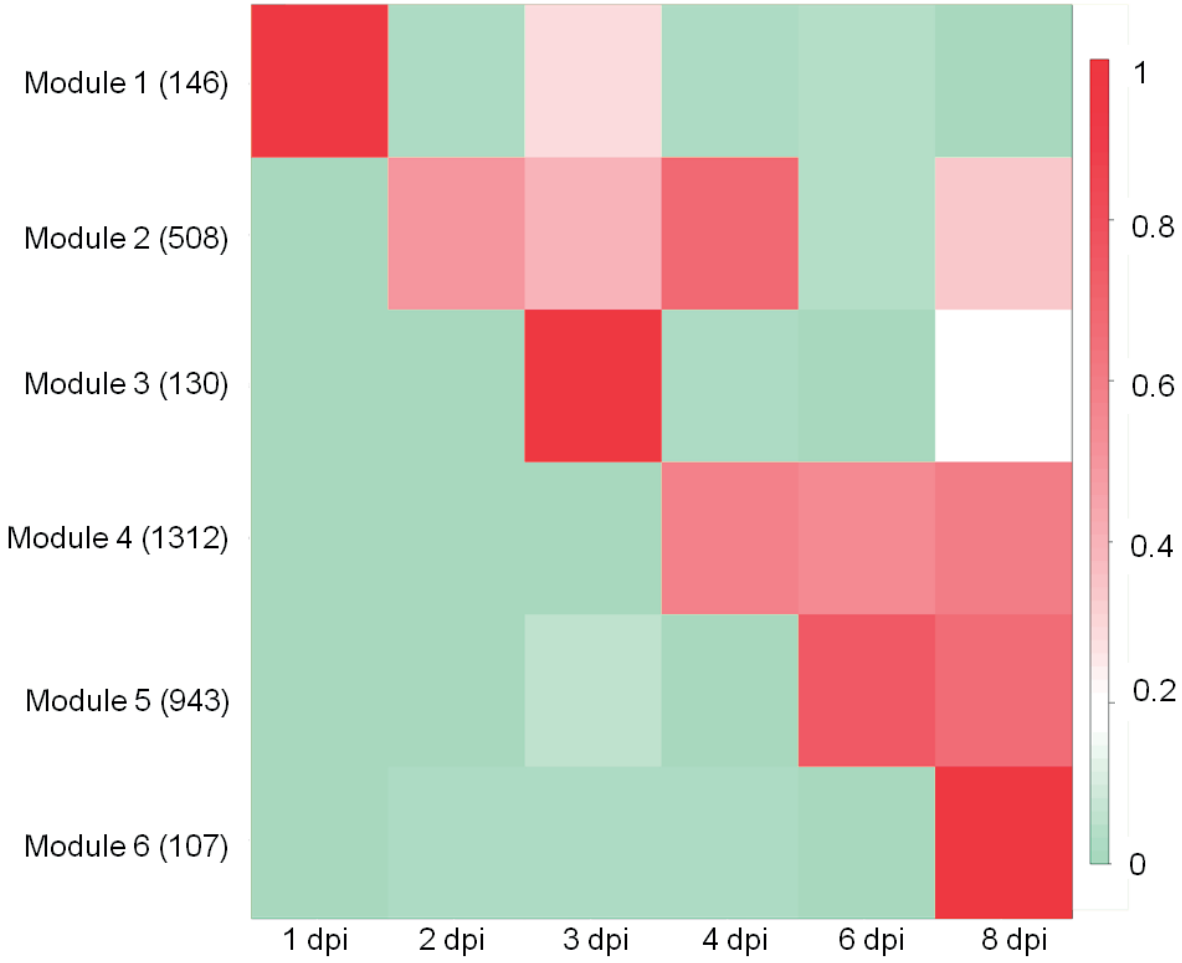

Figure 3.8 Heat map of the eigengenes representing each gene module. The columns in the heat map represent time points, and the rows represent eigengenes for each of the six identified co-expression modules. The numbers of genes in each module are given in parentheses. The cells in the heat map show eigengene values between 0 and 1, indicators of relative expression levels of all genes in the module (see Materials and Methods). dpi = days post-inoculation.

to be up-regulated during *P. infestans* infection on potato [10]. Our WGCNA analyses additionally identified genes that are co-expressed during the later stages of infection, specifically in Modules 4, 5, and 6 (Figure 3.3, Figure S3.2). Module 6, in particular,

represents genes most highly expressed at 8 dpi possibly indicative of those involved in a shift to the reproductive phase and sporulation.

Transcription factors (TFs) are reported to play a key role in the regulation of many biological processes including roles in oomycete pathogenesis [52,53] and within the predicted *Ps. cubensis* proteome 27 transcription factor-related domains in 440 genes were identified (Table S3.7, S3.8, available at www.plosone.org, e35796). A total of 247 of these were expressed throughout the infection process (Table S3.7, available at www.plosone.org, e35796). We also identified genes encoding transcription factor-related Pfam domains in all six co-expression modules. Two modules, 2 and 4, with genes co-expressed across different time contained the majority of the transcription factors. The transcription factor-related genes within those modules could play important role in regulation of genes involved in pathogenesis. The bZIP and Myb, DNA-binding transcription factor, which play an important role in oomycete pathogenesis [54,55], were the most abundant transcription factor-related domains expressed during infection.

CONCLUSIONS

In this study, we present an extensive characterization of the gene expression analysis of the obligate oomycete cucurbit pathogen *Ps. cubensis* during a compatible interaction. This data set represents the first global gene expression profile of a cucurbit pathogen. Using mRNA-Seq, we analyzed the differential expression of pathogen genes

across a time course of infection of cucumber, correlating expression with pathogen infection structures, development, and the onset of disease symptoms. Our study provides a comprehensive examination of the key infection stages of *Ps. cubensis* growth and development and through clustering and co-expression network analyses, describes genes that are specifically expressed during these stages. In addition, our work has identified an expanded effector repertoire, represented by a unique diversity at the canonical RXLR motif. Overall, the work described herein will significantly enhance our understanding of the regulation of infection of oomycete phytopathogens, as well as a baseline for identifying important virulence determinants in *Ps. cubensis.*

MATERIALS AND METHODS

Ps. cubensis inoculation and sample collection

Ps. cubensis MSU-1 was maintained on *Cucumis sativus* cultivar 'Vlaspik' as described previously [8]. Four-week-old cucumber plants were inoculated on the abaxial surface of the first fully-expanded leaf with a 1 x 10^5 sporangia/ml solution with 20-30 10 µl droplets. Inoculated plants were maintained at 100% relative humidity in the dark for 24 hours and then transferred to growth chambers maintained at 22 °C with a 12 h light/dark photoperiod. Samples were collected at 1, 2, 3, 4, 6, and 8 dpi with a #3 cork borer to collect tissue at the site of inoculation. Samples for RNA extraction were frozen in liquid nitrogen and stored at - 80 °C until use. Samples collected for microscopy were cleared in 95 % ethanol and stored at room temperature.

Histological assessment of *Ps. cubensis* growth

Cleared infected leaf discs were stained in a solution of 250 µg/ml trypan blue in equal parts lactic acid, water, and glycerol to visualize infection structures. Microscopy was performed using an Olympus IX71 inverted light microscope. Images were captured using an Olympus DC71 camera and were processed for contrast using Canvas X (ACD Systems International, Inc., Seattle, WA).

Library preparation and sequencing

Sporangia were washed from the abaxial surface of heavily sporulating leaves, filtered through a 40 µm nylon cell strainer, and pelleted *via* centrifugation. For RNA extraction (RNeasy Mini Kit, Qiagen, Valencia, CA), sporangia were resuspended in 450 µl RLT buffer with ~50 µl 425-600 µm acid-washed beads and vortexed for 3 minutes to break cells. Additional extraction steps were followed according to the manufacturer's instructions. RNA concentration and quality was determined using the Bioanalyzer 2100 (Agilent Technologies, San Diego CA). The sporangia library was sequenced in two lanes at the UC DNA Sequencing Facility at University of California, Davis (Davis, CA). RNA samples from the infection time course were processed as described in Adhikari et al. (accompanying paper [46]). In brief, RNA was isolated using the RNeasy Mini Kit (Qiagen, Germantown, MD), treated with DNase (Promega, Madison, WI) and barcoded libraries constructed with the Illumina mRNA-seq kit (Illumina, San Diego CA). Libraries

were sequenced with the Illumina Genome Analyzer II platform generating 35-42 bp single-end reads. Reads from biological replicates were pooled prior to expression abundance measurements. Reads were deposited in the National Center for Biotechnology Information Sequence Read Archive under accession number SRP009350.

mRNA-Seq read mapping and transcript abundance estimation

The assembled and annotated *Ps. cubensis* MSU-1 genome sequence (AHJF00000000; [9]) was used to estimate transcript abundances. mRNA-Seq reads for each time point and control (sporangia) were mapped to the 67.9 Mb *Ps. cubensis* reference genome using the quality aware alignment algorithms, Bowtie version 0.12.7 [56] and TopHat version 1.2.0 [57]. The single-end reads from different time points were aligned in single-end mode while the paired-end reads from the control were aligned in paired-end mode. The minimum and maximum intron length was set to 5 and 50,000 bp, respectively and the insert size for paired-end mode was set to 140 bp.

The aligned read files produced by TopHat were processed by Cufflinks v0.9.3 [58]. A reference annotation of the *Ps. cubensis* genome (23,519 loci and 23,522 gene models) was provided and the maximum intron length was set to 50,000 bp. Normalized gene expression levels were calculated and reported as FPKM. The quartile normalization option was used to improve differential expression calculations of lowly expressed genes [58]; all other parameters were used at the default settings. A gene was

considered expressed in a specific sample if the FPKM value and FPKM 95% confidence interval lower boundary was greater than 0.001 and zero, respectively.

Pearson product-moment correlation analyses of log2 FPKM values among mRNA-Seq libraries were performed using R (http://cran.r-project.org/), with all log2 FPKM values less than zero set to zero. Only tests significant at $p = 0.05$ are shown. Correlation values depicted as a heat map were clustered with hierarchical clustering using a Pearson correlation distance metric and average linkage. The bootstrap support values were calculated from 1000 replicates using Multiple Experiment Viewer Software (MeV) v4.5 [59]. To understand variability among biological replicates, Pearson correlation coefficients were calculated for the log2 transformed FPKM values of the genes expressed in both replicates at a particular time point.

Identification of differentially expressed genes

Once transcript abundance estimation was calculated, differential expression analysis was conducted using the Cuffdiff program within Cufflinks version 0.9.3 [58] utilizing the read alignment files described above. The expression testing was done at the level of genes. Quartile normalization [60] and a false discovery rate of 0.01 after Benjamini-Hochberg correction for multiple testing were used. The *Ps. cubensis* genome and the annotation files were provided as input parameters. All other parameters were used at the default levels. Cuffdiff was used to perform pairwise comparisons of six time points and sporangia.

mRNA-Seq and microarray comparative analyses

Gene expression data from a *P. infestans-S. tuberosum* time course experiment [10] was used to assess gene expression pattern similarities/differences in two ooymcete pathogens. The data set included *P. infestans* gene expression over a five-day (2-5 days) time course of a potato infection. Raw data was downloaded from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) (GSE14480) [61]. The probe intensities were normalized using Robust Multichip Analyses method [62]. For the mRNA-Seq to microarray comparative analysis, single copy orthologous genes were identified using OrthoMCL [63,64] with default parameters. Clustering of 23,522 and 18,140 protein-coding genes from *Ps. cubensis* and *P. infestans*, respectively, yielded 7,374 clusters with single copy genes from both species. Only single copy orthologous genes were used for the analyses. FPKM and probe intensity values were log2 transformed, and Spearman correlation coefficients were calculated using R (http://cran.r-project.org/).

Functional analysis

Functional annotation for all *Ps. cubensis* genes were generated from searches of the UniProt databases (Uuniref100) [65] with BLAST and combined with Pfam [66] protein families assignment performed using HMMER3 [67]. Functional annotations for *Ps. cubensis* sequences were taken from the best possible UniRef sequence match, but if there was no UniRef sequence match, functional annotations were made based on the best Pfam domain alignment. Transcription factors were identified based on PFAM

domains.

Gene co-expression network analysis

Gene co-expression network analysis was done according to the methods described by Childs et al. [68] with some modifications. First, the FPKM gene expression values were log2 transformed and FPKM values less than 1 were transformed to zero. Second, genes showing no variation across time points were filtered out using a coefficient of variance (CV) cutoff (1.0). Third, the β and treecut parameters were 7 and 0.6, respectively. Eigengenes were calculated using the WGCNA package [51]. The heat map of eigengenes for each gene module was constructed using R (http://cran.r-project.org/). Genes assigned to co-expression modules were annotated based on the *Ps. cubensis* functional annotation.

APPENDIX

Figure S3.1

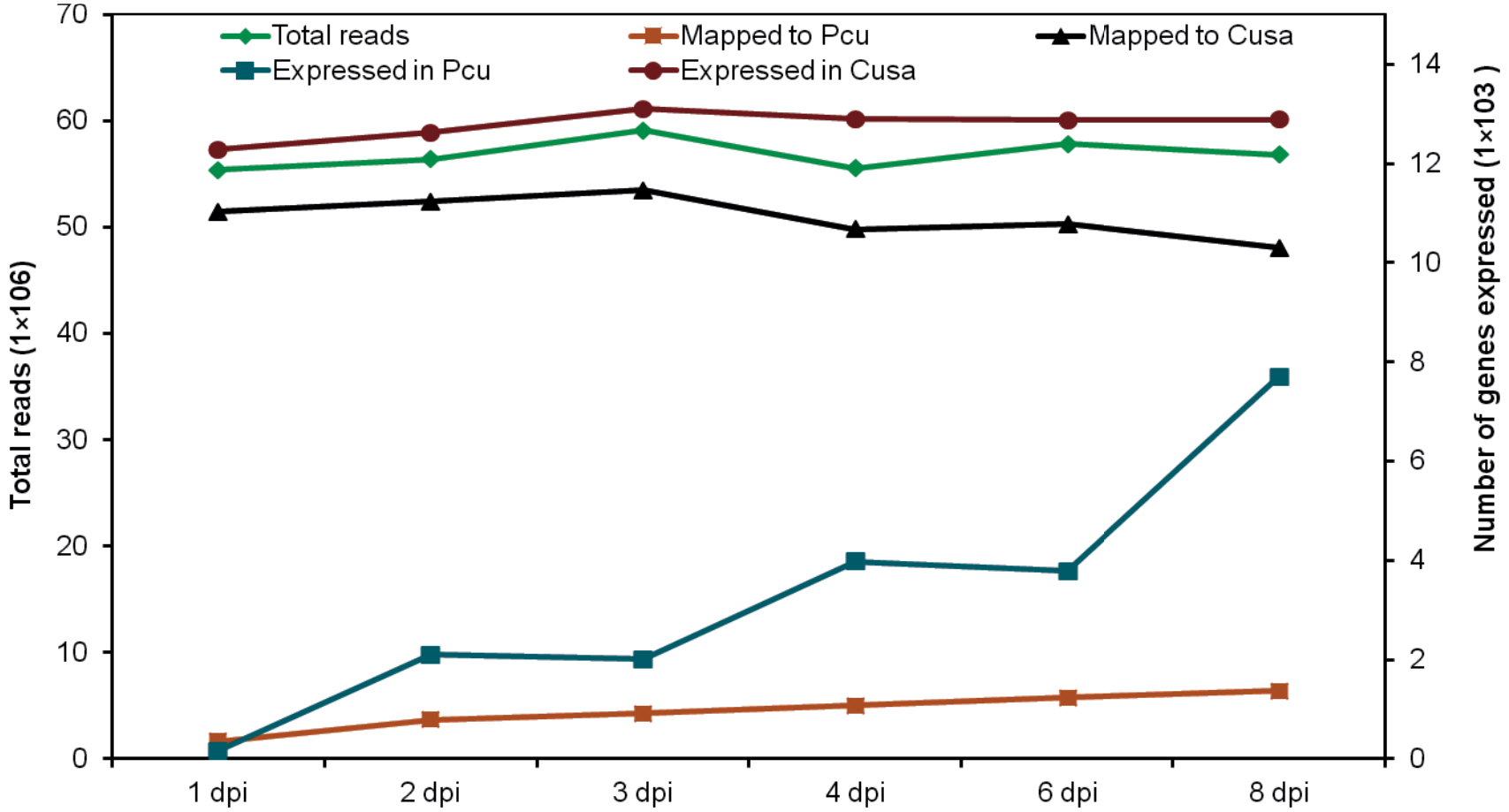

Figure S3.1 Number of total RNA-seq reads, reads mapped, and number of genes expressed at different timepoints. Total number of reads, number of reads mapped, and number of genes expressed in *Cucumis sativus* (Cusa) and *Pseudoperonospora cubensis* (Pcu) at different time-points are shown. Reads were mapped using Bowtie version 0.12.5 and TopHat version 1.2.0. Fragments per kilobase pair of exon model per million fragments mapped (FPKM) values were calculated using Cufflinks version 0.9.3. Genes were considered expressed if the 95% confidence interval lower boundary FPKM value was greater than zero. dpi = days post-inoculation.

Figure S3.2

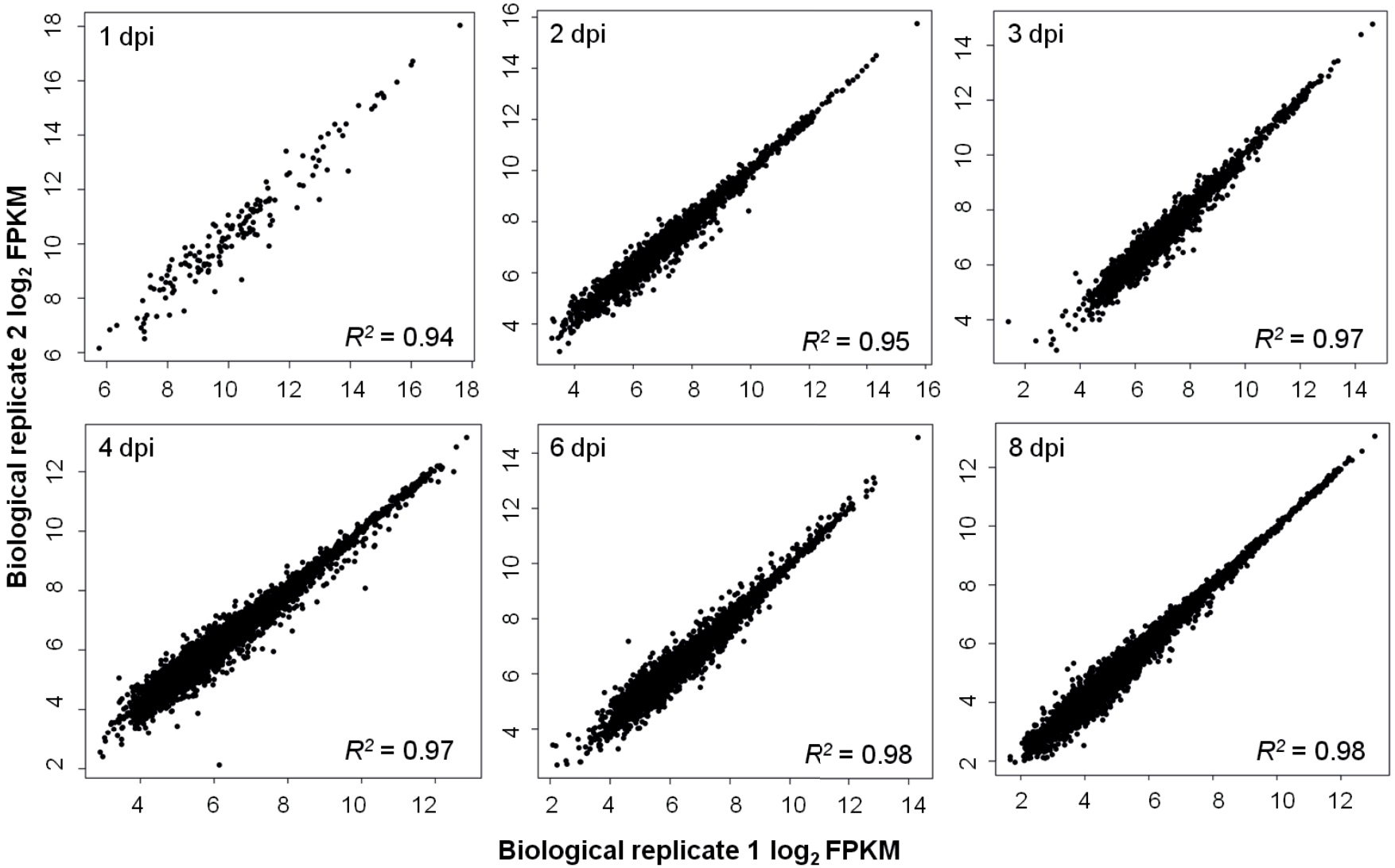
1 dpi
R^2 = 0.94
2 dpi
R^2 = 0.95
3 dpi
R^2 = 0.97
4 dpi
R^2 = 0.97
6 dpi
R^2 = 0.98
8 dpi
R^2 = 0.98
Biological replicate 2 log$_2$ FPKM
Biological replicate 1 log$_2$ FPKM

Figure S3.2 (cont'd) Concordance of FPKM values of the genes expressed in two biological replicates of the *Pseudoperonospora cubensis* transcriptome. Reads from different timepoints were mapped to *Ps. cubensis* genome using Bowtie version 0.12.5 and TopHat version 1.2.0. Fragments per kilobase pair of exon model per million fragments mapped (FPKM) values were calculated using Cufflinks version 0.9.3 and *Ps. cubensis* genome annotations. For each time point Log2 transformed FPKM values of equal number of genes from both replicates are plotted. Pearson Correlation Coefficient (PCC) was calculated using R. dpi, days post-inoculation.

Figure S3.3

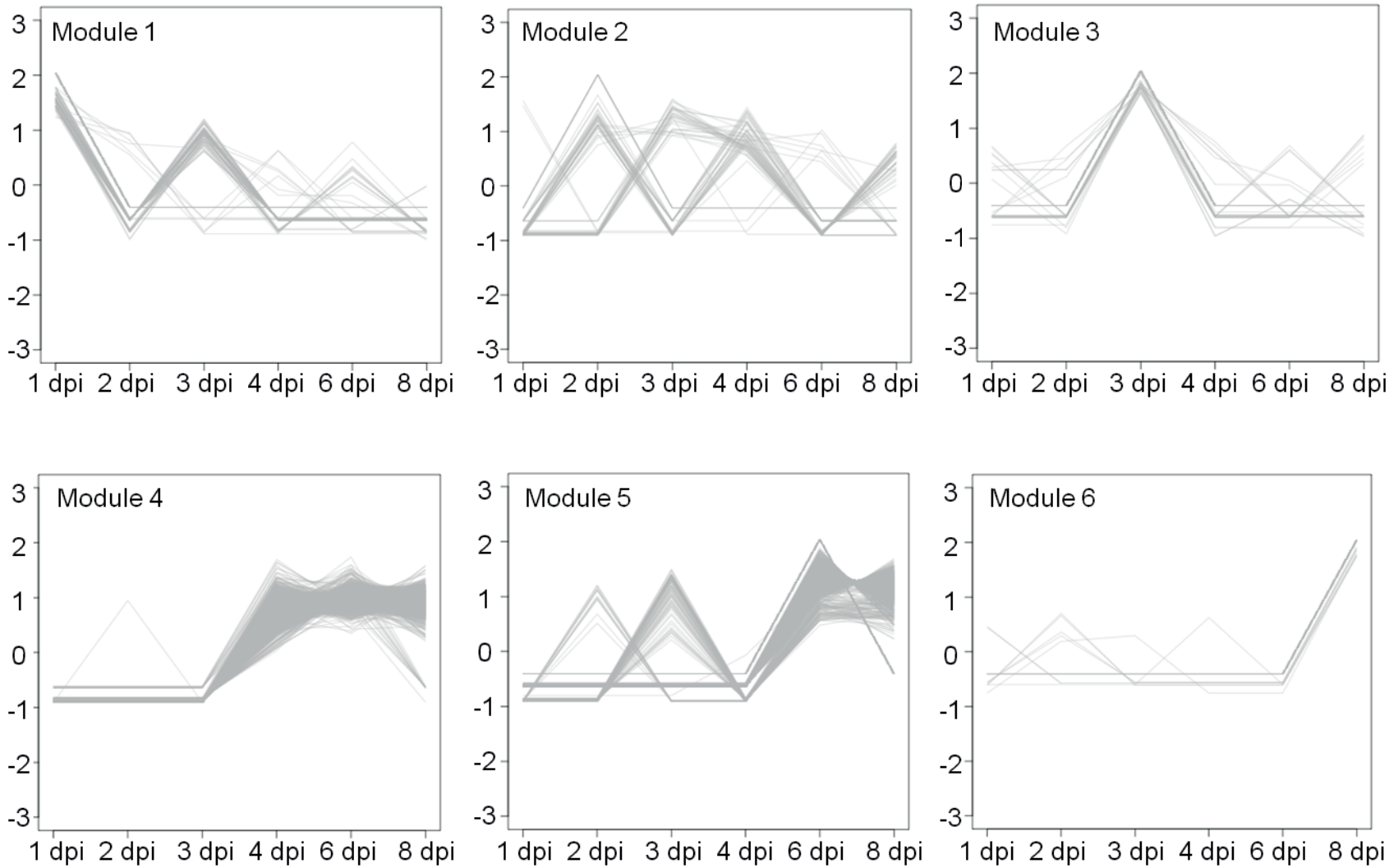

Figure S3.3 Trend plots of the normalized gene expression values for each gene from six identified gene co-expression modules. Modules consisting of genes expressed modules 1, 2, 3, 4, 5, and 6 are shown.

REFERENCES

REFERENCES

1. Savory EA, Granke LL, Quesada-Ocampo LM, Varbanova M, Hausbeck MK, et al. (2011) The cucurbit downy mildew pathogen *Pseudoperonospora cubensis*. Mol Plant Pathol 12: 217-226.

2. Lebeda A, Cohen Y (2011) Cucurbit downy mildew (*Pseudoperonospora cubensis*)—biology, ecology, epidemiology, host-pathogen interaction and control. Eur J Plant Pathol 129: 157-192.

3. Thomas C, Inaba T, Cohen Y (1987) Physiological specialization in *Pseudoperonospora cubensis*. Phytopathol 77: 1621-1624.

4. Sarris P, Abdelhalim M, Kitner M, Skandalis N, Panopoulos N, et al. (2009) Molecular polymorphisms between populations of *Pseudoperonospora cubensis* from Greece and the Czech Republic and the phytopathological and phylogenetic implications. Plant Pathol 58: 933-943.

5. Lebeda A, Widrlechner MP (2003) A set of Cucurbitaceae taxa for differentiation of *Pseudoperonespora cubensis* pathotypes. J Plant Dis Prot 110: 337–349.

6. Runge F, Choi Y-J, Thines M (2011) Phylogenetic investigations in the genus *Pseudoperonospora* reveal overlooked species and cryptic diversity in the *P. cubensis* species cluster. Eur J Plant Pathol 129: 135-146.

7. Choi Y (2005) A re-consideration of *Pseudoperonospora cubensis* and *P. humuli* based on molecular and morphological data. Mycological Res 109: 841-848.

8. Tian M, Win J, Savory E, Burkhardt A, Held M, et al. (2011) 454 Genome sequencing of *Pseudoperonospora cubensis* reveals effector proteins with a QXLR translocation motif. Mol Plant-Microbe Interact 24: 543-553.

9. Savory EA, Zou C, Adhikari BN, Hamilton JP, Buell CR, et al. (2012) Alternative Splicing of a Multi-Drug Transporter from *Pseudoperonospora cubensis* Generates an RXLR Effector Protein That Elicits a Rapid Cell Death. PLoS ONE 7: e34701.

10. Haas BJ, Kamoun S, Zody MC, Jiang RH, Handsaker RE, et al. (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature 461: 393-398.

11. Kamoun S (2006) A catalogue of the effector secretome of plant pathogenic oomycetes. Ann Rev Phytopathol 44: 41-60.

12. Cabral A, Stassen JH, Seidl MF, Bautor J, Parker JE, et al. (2011) Identification of *Hyaloperonospora arabidopsidis* transcript sequences expressed during Infection reveals isolate-specific effectors. PLoS ONE 6: e19328.

13. Sierra R, Rodriguez RL, Chaves D, Pinzon A, Grajales A, et al. (2010) Discovery of *Phytophthora infestans* genes expressed in planta through mining of cDNA libraries. PLoS ONE 5: e9847.

14. Torto-Alalibo TA, Tripathy S, Smith BM, Arredondo FD, Zhou L, et al. (2007) Expressed sequence tags from *Phytophthora sojae* reveal genes specific to development and infection. Mol Plant-Microbe Interact 20: 781-793.

15. Randall TA, Dwyer RA, Huitema E, Beyer K, Cvitanich C, et al. (2005) Large-scale gene discovery in the oomycete *Phytophthora infestans* reveals likely components of phytopathogenicity shared with true fungi. Mol Plant-Microbe Interact 18: 229–243.

16. Bos JI, Kanneganti TD, Young C, Cakir C, Huitema E, et al. (2006) The C-terminal half of *Phytophthora infestans* RXLR effector AVR3a is sufficient to trigger R3a-mediated hypersensitivity and suppress INF1-induced cell death in *Nicotiana benthamiana*. Plant J 48: 165-176.

17. Fabro G, Steinbrenner J, Coates M, Ishaque N, Baxter L, et al. (2011) Multiple Candidate Effectors from the Oomycete Pathogen *Hyaloperonospora arabidopsidis* Suppress Host Plant Immunity. PLoS Pathog 7: e1002348.

18. Armstrong MR, Whisson SC, Pritchard L, Bos JI, Venter E, et al. (2005) An ancestral oomycete locus contains late blight avirulence gene Avr3a, encoding a protein that is recognized in the host cytoplasm. Proc Natl Acad Sci U S A 102: 7766-7771.

19. Allen RL (2004) Host-Parasite Coevolutionary Conflict Between Arabidopsis and Downy Mildew. Science 306: 1957-1960

20. Money NP, Davis CM, Ravishankar JP (2004) Biomechanical evidence for convergent evolution of the invasive growth process among fungi and oomycete water molds. Fung Gen Biol 41: 872-876.

21. Dong S, Qutob D, Tedman-Jones J, Kuflu K, Wang Y, et al. (2009) The *Phytophthora sojae* avirulence locus Avr3c encodes a multi-copy RXLR effector with sequence polymorphisms among pathogen strains. PLoS ONE 4: e5556

22. Torto TA (2003) EST mining and functional expression fssays identify extracellular effector proteins from the plant pathogen *Phytophthora*. Genome Res 13: 1675-1685.

23. Schornack S, van Damme M, Bozkurt TO, Cano LM, Smoker M, et al. (2010) Ancient class of translocated oomycete effectors targets the host nucleus. Proc Natl Acad Sci U S A 107: 17421-17426.

24. Damasceno CM, Bishop JG, Ripoll DR, Win J, Kamoun S, et al. (2008) Structure of the glucanase inhibitor protein (GIP) family from Phytophthora species suggests coevolution with plant endo-beta-1,3-glucanases. Mol Plant-Microbe Interact 21: 820-830.

25. Rose JK, Ham KS, Darvill AG, Albersheim P (2002) Molecular cloning and characterization of glucanase inhibitor proteins: coevolution of a counterdefense mechanism by plant pathogens. Plant Cell 14: 1329-1345.

26. Tian M, Benedetti B, Kamoun S (2005) A Second Kazal-like protease inhibitor from *Phytophthora infestans* inhibits and interacts with the apoplastic pathogenesis-related protease P69B of tomato. Plant Physiol 138: 1785-1793.

27. Tian M, Huitema E, Da Cunha L, Torto-Alalibo T, Kamoun S (2004) A Kazal-like extracellular serine protease inhibitor from *Phytophthora infestans* targets the tomato pathogenesis-related protease P69B. J Biol Chem 279: 26370-26377.

28. Tian M, Win J, Song J, van der Hoorn R, van der Knaap E, et al. (2007) A *Phytophthora infestans* cystatin-like protein targets a novel tomato papain-like apoplastic protease. Plant Physiol 143: 364-377.

29. Liu Z, Bos JI, Armstrong M, Whisson SC, da Cunha L, et al. (2005) Patterns of diversifying selection in the phytotoxin-like scr74 gene family of *Phytophthora infestans*. Mol Biol Evol 22: 659-672.

30. Levesque CA, Brouwer H, Cano L, Hamilton JP, Holt C, et al. (2010) Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. Genome Biol 11: R73.

31. Fellbrich G, Romanski A, Varet A, Blume B, Brunner F, et al. (2002) NPP1, a Phytophthora-associated trigger of plant defense in parsley and Arabidopsis. Plant J 32: 375-390.

32. Cabral A, Oome S, Sander N, Kuefner I, Nürnberger T, et al. (2012) Non-toxic Nep1-like proteins of the downy mildew pathogen *Hyaloperonospora arabidopsidis*; repression of necrosis-inducing activity by a surface-exposed region. Mol Plant-Microbe Interact.

33. Gaulin E, Drame N, Lafitte C, Torto-Alalibo T, Martinez Y, et al. (2006) Cellulose binding domains of a *Phytophthora* cell wall protein are novel pathogen-associated molecular patterns. Plant Cell 18: 1766-1777.

34. Gaulin E, Jauneau A, Villalba F, Rickauer M, Esquerré-Tugayé MT, et al. (2002) The CBEL glycoprotein of *Phytophthora parasitica* var-nicotianae is involved in cell wall deposition and adhesion to cellulosic substrates. J Cell Sci 115: 4565.

35. Tyler BM (2006) *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313: 1261-1266

36. Baxter L, Tripathy S, Ishaque N, Boot N, Cabral A, et al. (2010) Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. Science 330: 1549-1551.

37. Moy P, Qutob D, Chapman BP, Atkinson I, Gijzen M (2004) Patterns of gene expression upon infection of soybean plants by *Phytophthora sojae*. Mol Plant-Microbe Interact 17: 1051-1062.

38. Costanzo S, Ospina-Giraldo MD, Deahl KL, Baker CJ, Jones RW (2006) Gene duplication event in family 12 glycosyl hydrolase from *Phytophthora* spp. Fung Gen Biol 43: 707-714.

39. Yamaguchi Y, Huffaker A (2011) Endogenous peptide elicitors in higher plants. Curr Opinion Plant Biol 14: 351-357.

40. Fernandez D, Tisserant E, Talhinhas P, Azinheira H, Vieira ANA, et al. (2011) 454-pyrosequencing of *Coffea arabica* leaves infected by the rust fungus *Hemileia vastatrix* reveals in planta-expressed pathogen-secreted proteins and plant functions in a late compatible plant–rust interaction. Mol Plant Pathol 13: 17-37.

41. Joly DL, Feau N, Tanguay P, Hamelin RC (2010) Comparative analysis of secreted protein evolution using expressed sequence tags from four poplar leaf rusts (*Melampsora* spp.). BMC Gen 11: 422.

42. Miranda M, Ralph SG, Mellway R, White R, Heath MC, et al. (2007) The transcriptional response of hybrid poplar (*Populus trichocarpa* x *P. deltoides*) to infection by *Melampsora medusae* leaf rust involves induction of flavonoid pathway genes leading to the accumulation of proanthocyanidins. Mol Plant-Microbe Interact 20: 816-931.

43. Duplessis S, Hacquard S, Delaruelle C, Tisserant E, Frey P, et al. (2011) *Melampsora larici*-populina tanscript profiling during germination and timecourse infection of poplar leaves reveals dynamic expression patterns associated with virulence and biotrophy. Mol Plant-Microbe Interact 24: 808-818.

44. Mosquera G, Giraldo M, Khang CH, Coughlan S, Valent B (2009) Interaction transcriptome analysis identifies *Magnaporthe oryzae* BAS1-4 as biotrophy-associated secreted proteins in rice blast disease. Plant Cell 21: 1273-1290.

45. Polesani M, Desario F, Ferrarini A, Zamboni A, Pezzotti M, et al. (2008) cDNA-AFLP analysis of plant and pathogen genes expressed in grapevine infected with *Plasmopara viticola*. BMC Gen 9: 142 - 156.

46. Adhikari BN, Savory EA, Vaillancourt B, Childs KL, Hamilton JP, et al. (2012) Expression Profiling of *Cucumis sativus* in Response to Infection by *Pseudoperonospora cubensis*. PLoS ONE 7: e34954.

47. Ospina-Giraldo M, Griffith J, Laird E, Mingora C (2010) The CAZyome of *Phytophthora* spp.: A comprehensive analysis of the gene complement coding for carbohydrate-active enzymes in species of the genus *Phytophthora*. BMC Gen 11: 1-16.

48. Ospina-Giraldo M, McWalters J, Seyer L (2010) Structural and functional profile of the carbohydrate esterase gene complement in *Phytophthora infestans*. Curr Gen 56: 495-506.

49. Ihmels J, Bergmann S, Berman J, Barkai N (2005) Comparative gene expression analysis by a differential clustering approach: Application to the *Candida albicans* transcription program. PLoS Genet 1: e39.

50. Langfelder P, Zhang B, Horvath S (2008) Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 24: 719 - 720.

51. Langfelder P, Horvath S (2007) Eigengene networks for studying the relationships between co-expression modules. BMC Sys Biol 1: 54.

52. Iyer LM, Anantharaman V, Wolf MY, Aravind L (2008) Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. In J Parasitol 38: 1-31.

53. Wang Y, Dou D, Wang X, Li A, Sheng Y, et al. (2009) The PsCZF1 gene encoding a C2H2 zinc finger protein is required for growth, development and pathogenesis in *Phytophthora sojae*. Micro Path 47: 78-86.

54. Blanco FA, Judelson HS (2005) A bZIP transcription factor from *Phytophthora* interacts with a protein kinase and is required for zoospore motility and plant infection. Mol Micro 56: 638-648.

55. Judelson H, Ah-Fong A (2010) The kinome of *Phytophthora infestans* reveals oomycete-specific innovations and links to other taxonomic groups. BMC Gen 11: 700.

56. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25.

57. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105-1111.

58. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28: 511-515.

59. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, et al. (2006) TM4 microarray software suite. Methods Enz 411: 134-193.

60. Bullard J, Purdom E, Hansen K, Dudoit S (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics 11: 94.

61. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30: 207-210.

62. Irizarry R, Hobbs B, Collin F, Beazer-Barclay Y, Antonellis K, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4: 249-264.

63. Li L, Stoeckert CJ, Jr., Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13: 2178-2189.

64. Chen F, Mackey AJ, Vermunt JK, Roos DS (2007) Assessing performance of orthology detection strategies applied to eukaryotic genomes. PLoS ONE 2: e383.

65. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH (2007) UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics 23: 1282-1288.

66. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, et al. (2000) The Pfam protein families database. Nucleic Acids Res 28: 263-266.

67. Eddy SR (2009) A new generation of homology search tools based on probabilistic inference. Genome Inform 23: 205-211.

68. Childs KL, Davidson RM, Buell CR (2011) Gene coexpression network analysis as a source of functional annotation for rice genes. PLoS ONE 6: e22196.

CHAPTER 4

Expression profiling of *Cucumis sativus* in response to infection by *Pseudoperonospora cubensis*

This chapter was originally published in PLoS ONE.

Adhikari, BN*, Savory EA*, Vaillancourt, B, Childs KL, Hamilton JP, Day B and Buell CR (2012) Expression profiling of *Cucumis sativus* in response to infection by *Pseudoperonospora cubensis.* PLoS ONE 7(4): e34954. doi:10.1371/journal.pone.0034954

* These authors contributed equally to this work.

Author Contributions:

Conceived and designed the experiments: EAS, BNA, KLC, BD and CRB. Performed the experiments: BNA, EAS, KLC, and JPH. Analyzed the data: BNA, EAS, KLC, and JPH. Contributed reagents/materials/analysis tools: BNA, EAS, BV, KLC, JPH, BD and CRB. Wrote the paper: BNA, EAS, BD and CRB.

ABSTRACT

The oomycete pathogen, *Pseudoperonospora cubensis*, is the causal agent of downy mildew on cucurbits, and at present, no effective resistance to this pathogen is available in cultivated cucumber (*Cucumis sativus*). To better understand the host response to a virulent pathogen, we performed expression profiling throughout a time course of a compatible interaction using whole transcriptome sequencing. As described herein, we were able to detect the expression of 15,286 cucumber genes, of which 14,476 were expressed throughout the infection process from 1 day post-inoculation (1 dpi) to 8 dpi. A large number of genes, 1,612 to 3,286, were differentially expressed in pair-wise comparisons between time points. We observed the rapid induction of key defense related genes, including catalases, chitinases, lipoxygenases, peroxidases, and protease inhibitors within 1 dpi, suggesting detection of the pathogen by the host. Co-expression network analyses revealed transcriptional networks with distinct patterns of expression including down-regulation at 2 dpi of known defense response genes suggesting coordinated suppression of host responses by the pathogen. Comparative analyses of cucumber gene expression patterns with that of orthologous *Arabidopsis thaliana* genes following challenge with *Hyaloperonospora arabidopsidis* revealed correlated expression patterns of single copy orthologs suggesting that these two dicot hosts have similar transcriptional responses in response to related pathogens. In total, the work described herein presents an in-depth analysis of the interplay between host susceptibility and pathogen virulence in an agriculturally important pathosystem.

INTRODUCTION

Cucumber (*Cucumis sativus* L.) is an economically important vegetable crop cultivated in over 80 countries, with more than 66 million tons produced annually for both fresh use and processing (http://faostats.fao.org). Cucumber has been utilized extensively as a model system to study sex determination [1], vascular biology [2], and induced defense responses [3,4]. Despite its extensive use as a model system in research, as well as its obvious economic importance, genetic and genomic resources remain limited. In recent years, the publication of both a genetic map [5] and genome sequences [6,7] of cucumber, as well as generation of large-scale expression data sets [8,9], have provided the first comprehensive resources for genetic and genomic based inquiries into cucumber biology. Cucumber has limited genetic diversity, few wild relatives, and only 7 pairs of chromosomes (2n = 14), whereas other *Cucumis* spp., such as melon (*Cucumis melo*), have 12 chromosomes, making interspecific breeding difficult, if not impossible. As such, advances in breeding for important agronomic traits such as increased yield, fruit quality, and disease resistance are slow.

Cucumber production is hindered by diseases caused by bacterial (e.g., *Pseudomonas syringae* pv. *lachrymans*), viral (e.g., *Cucumber mosaic virus*), fungal (e.g., *Sphaerotheca fulginea* and *Erysiphe cichoracearum*), and oomycete (e.g., *Phytophthora capsici* and *Pseudoperonospora cubensis*) pathogens [10,11]. Of these, the most destructive is *Ps. cubensis*, the causal agent of cucurbit downy mildew, which threatens cucumber production in nearly 80 countries and causes severe economic losses

[11,12,13]. *Ps. cubensis* is an obligate, biotrophic oomycete pathogen with a host range limited to the *Cucurbitaceae* [11]. In recent years, a foundation has been established to support advances in this area, including studies on epidemiology [14], host specificity [15,16,17,18], pathogenic variation [19,20,21], and more recently, generation of a draft genome sequence of *Ps. cubensis* [22,23].

The molecular and biochemical mechanisms associated with host resistance have been investigated to a limited extent in cucumber and other cucurbits. In large part, signaling of resistance is primarily associated with systemic acquired resistance (SAR) [24], for which cucumber has historically been a model system [25,26]. In addition to SAR, structural modifications (e.g., callose deposition), as well as the induction of defense-related genes (e.g., peroxidases chitinases, and glucanases) are often associated with the onset of resistance signaling following pathogen infection. Moreover, like other well-characterized plant-pathogen systems, the presence of nucleotide-binding site (NBS)-containing genes encoding protein products that recognize cognate pathogen effector proteins or their perturbations [27] have been postulated to play a role in disease resistance in cucumber. To this end, analysis of the cucumber genome sequence has identified 61 NBS-Resistance genes, considerably less than have been identified in other plant genomes, such as Arabidopsis (200) or rice (600) [6]. However, of the 15 genes known to control disease resistance in cucumber, none have been cloned, nor have they been associated through linkage maps with the 61 predicted NBS-Resistance genes [28]. It is hypothesized that cucumber has an expanded lipoxygenase (LOX)

pathway which may provide an additional mechanism(s) to facilitate responses to biotic stress [6].

While genetically conferred host resistance is the ideal means of disease control in crop species, it has become ineffective in controlling cucurbit downy mildew, particularly in the U.S., where introduction of a new, more virulent pathotype of *Ps. cubensis* is responsible for economic losses in recent years [12,29]. To this end, control methods for cucurbit downy mildew in both Europe and the U.S. require the use of fungicides, coupled with a single host resistance locus, the recessive *dm1* gene, which has been incorporated into most commercial cucumber germplasm [11]. However, the identification of the *dm1* locus, as well as its functional role in resistance signaling remains undefined. In addition to widespread incorporation of *dm1*, breeding of *Ps. cubensis* resistance has focused mainly on genes from melon [30], as limited diversity for resistance is available in cucumber or its wild relative, *Cucumis hardwickii*. Large-scale screening trials to identify tolerant germplasm are in progress, but have yet to identify a source of complete resistance to *Ps. cubensis* [16,31]. As such, new resources must be explored to support development of improved cultivars and identify new sources of resistance for breeding programs.

Next generation sequencing of the transcriptome (mRNA-Seq) permits deep, robust assessments of transcript abundances and transcript structure [32]. When gene expression profiling is applied to host-pathogen interactions of economically important crops, insights into the mechanisms these pathogens use to suppress and subvert the

host defense response can be made [33,34,35]. In the current study, we performed expression profiling, using the susceptible cucumber cultivar 'Vlaspik', over a time course of infection with the downy mildew pathogen *Ps. cubensis* to identify genes, pathways, and systems that are altered during a compatible interaction. Using this technology, deep profiling of both the host and pathogen transcriptome (see accompanying manuscript, [36]) was achieved, providing the first in-depth analysis of this important plant-pathogen interaction. In this study, we cataloged the expression of 14,476 genes from cucumber through an 8-day time course of infection with a virulent *Ps. cubensis* isolate. In total, this work identified major changes in gene expression in cucumber at 1 days post-inoculation (dpi) continuing through 8 dpi, with up to 3,286 genes differentially expressed between time points. Comparative analyses revealed correlated gene expression patterns in cucumber and *Arabidopsis thaliana* leaves infected with downy mildew suggesting orthologous host responses in these two dicotyledonous hosts. Through co-expression network analyses, modules of temporal-specific transcriptional networks were constructed that provide a framework to connect transcription factors with defense response genes.

RESULTS AND DISCUSSION

Response of *C. sativus* leaves to pathogen infection

To correlate gene expression and host responses with observable disease symptoms and pathogen invasion, the progression of infection in susceptible *C. sativus* cv.

'Vlaspik' was monitored at 1, 2, 3, 4, 6, and 8 dpi. As shown in Figure 4.1, the first visible symptoms of *Ps. cubensis* infection were apparent at 1 dpi, in the form of water soaking on the abaxial leaf surface at the inoculation site (Figure 4.1A). These symptoms correspond to zoospore encystment and initial penetration through the stomata into the host (see accompanying paper, [36]). In similar pathosystems systems such as *Hyaloperonospora arabidopsidis*, the causal agent of downy mildew on *A. thaliana*, analogous processes occur in the early stages of infection, with the exception that *H. arabidopsidis* penetrates between anticlinal walls of epidermal cells rather than utilizing natural openings like stomata [37]. While no symptoms are visible on the upper leaf surface in cucumber, water soaking on the lower leaf surface can be seen as early as 1 dpi, and remains present through 4 dpi, during which time hyphal growth through the mesophyll of the host tissue occurs and haustoria formation begins (see accompanying paper, [36]). Yellow angular lesions bound by leaf veins characteristic of cucurbit downy mildew were first visible on the upper leaf surface at 4 dpi (Figure 4.1D), and became more chlorotic and slightly necrotic at the centers by 8 dpi. These symptoms are associated with extensive growth of hyphae through the plant mesophyll (see accompanying paper, [36]).

Gene expression profiling in *C. sativus*

We performed mRNA-Seq profiling of *C. sativus* leaves following infection with *Ps.*

Figure 4.1

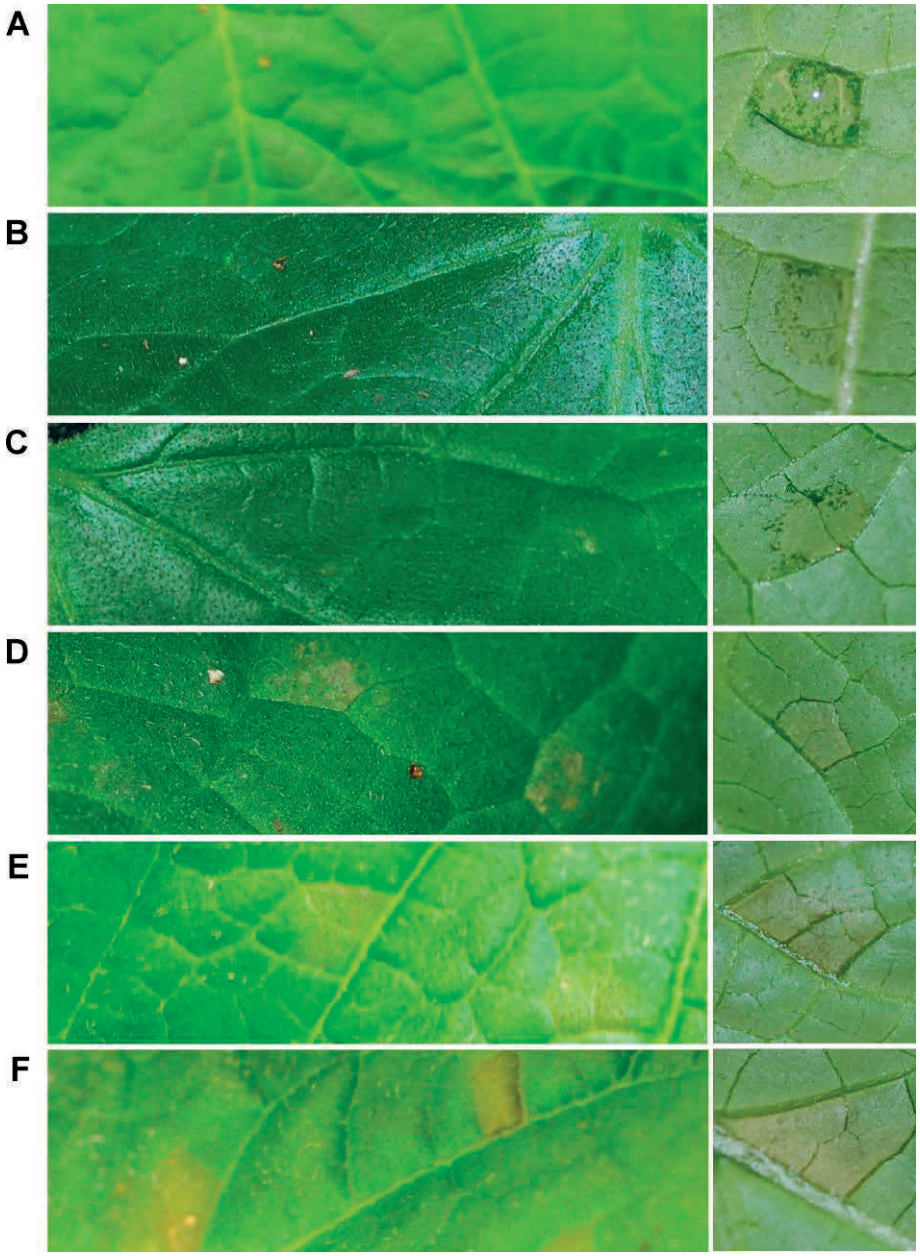

Figure 4.1 Symptoms of *Pseudoperonospora cubensis* infection on susceptible *Cucumis sativus* cv. 'Vlaspik'. Images were collected of the adaxial (left column) and abaxial (right column) leaf surfaces at 1, 2, 3, 4, 6, and 8 days post-inoculation (dpi). **(**A) 1 dpi, (B) 2 dpi, (C) 3 dpi, (D) 4 dpi, (E) 6 dpi, (F) 8 dpi.

maximize the amount of infected tissue in each sample (Figure 4.2, black circles), pooled within a given time point, and RNA was isolated. Two biological replicates of each time point were sequenced, yielding 55 to 59 million reads from both replicates at each time point. Additionally, a mock-inoculated *C. sativus* sample was collected and sequenced, yielding 5.8 million reads. The number of reads that mapped to the *C. sativus* genome ranged from 48 to 53 million (Figure 4.3A, 84-93% of the total reads) per time point while the number of genes expressed at different time points ranged from 12,257 to 13,048 (Figure 4.3A).

The libraries were constructed from inoculated leaves and therefore, the reads represent transcripts from the host (*C. sativus*) and the pathogen (*Ps. cubensis*). At the early time points, nearly all of the reads obtained were of host origin, which is consistent with our microscopy analysis revealing limited pathogen biomass. However, as we are surveying a susceptible interaction, pathogen biomass increases throughout the time course and consequentially, pathogen transcripts increase in the total read pool in the later time points (see accompanying paper, [36]). However, even with the increased percentage of pathogen reads in the later time points, we have saturated sampling of the *C. sativus* transcriptome with our deep read coverage of the libraries. Randomly selected subsets of reads, 5 to 30 million, from the total read pool were used to evaluate the effect of sampling depth on gene expression detection. The simulation demonstrates a positive relationship between sampling depth and numbers of expressed genes at lower sequencing depths (5 to 20 million reads)

Figure 4.2

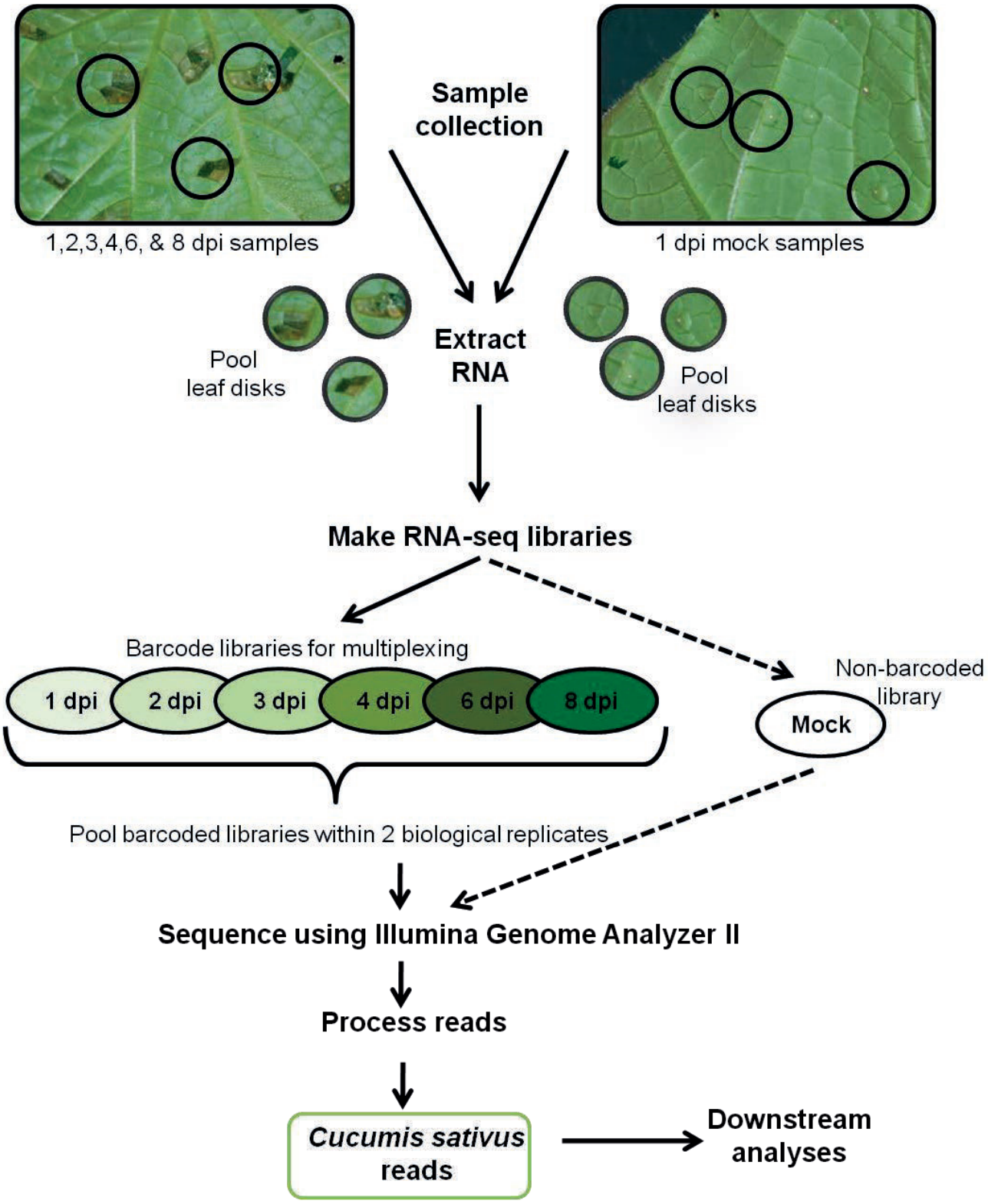
Sample collection
1,2,3,4,6, & 8 dpi samples
1 dpi mock samples
Pool leaf disks
Extract RNA
Pool leaf disks
Make RNA-seq libraries
Barcode libraries for multiplexing
1 dpi
2 dpi
3 dpi
4 dpi
6 dpi
8 dpi
Non-barcoded library
Mock
Pool barcoded libraries within 2 biological replicates
Sequence using Illumina Genome Analyzer II
Process reads
Cucumis sativus reads
Downstream analyses

Figure 4.2 (cont'd) Experimental design and sample collection. Cucumber cv. 'Vlaspik' leaves were inoculated on the abaxial leaf surface with 10-30 10 μl droplets of a 1 x 10^5 sporangia/ml solution. Samples were collected at 1, 2, 3, 4, 6, and 8 days post-inoculation (dpi) using a #3 cork borer to isolate tissue immediately around the infection point (black circles). Samples from cucumber leaves mock-inoculated with 10 μl droplets of dH_2O were collected in the same manner at 1 dpi. Leaf disks from each time point were pooled for RNA extraction. mRNA-Seq libraries were made for each time point from two separate biological replicates. Within a biological replicate, libraries were barcoded and sequenced in multiple lanes with the exception of the mock-inoculated library, which was not barcoded.

(Figure 4.3B). The number of expressed genes, however, begins to plateau at approximately 20 million reads, corresponding to the minimum sampling depth of all libraries in this study. To study the repeatability of two biological replicates, pair-wise scatter plots of gene expression values were generated. For biological replicates of each time point, nearly all genes fell along the diagonal of plots, indicating no major variation with correlation coefficients (R^2) ranging from 0.97 to 0.98, thus providing evidence for high reproducibility of biological replicates (Figure S4.1).

Host transcriptional changes in response to infection

Over the infection period, a total of 14,476 *C. sativus* genes were expressed (Table S4.1, available at www.plosone.org, e34954), with 10,350 genes common to all time points. For all data points analyzed, the minimum fragments per kilobase exon model per million mapped reads (FPKM) value was zero, yet the maximum FPKM values ranged from 8,869 at 1 dpi to 34,017 at 8 dpi (Table S4.1, available at www.plosone.org, e34954). Interestingly, the highest up-regulated gene at 1 dpi was a putative galactinol synthase (Csa6M000080.1), which has

Figure 4.3

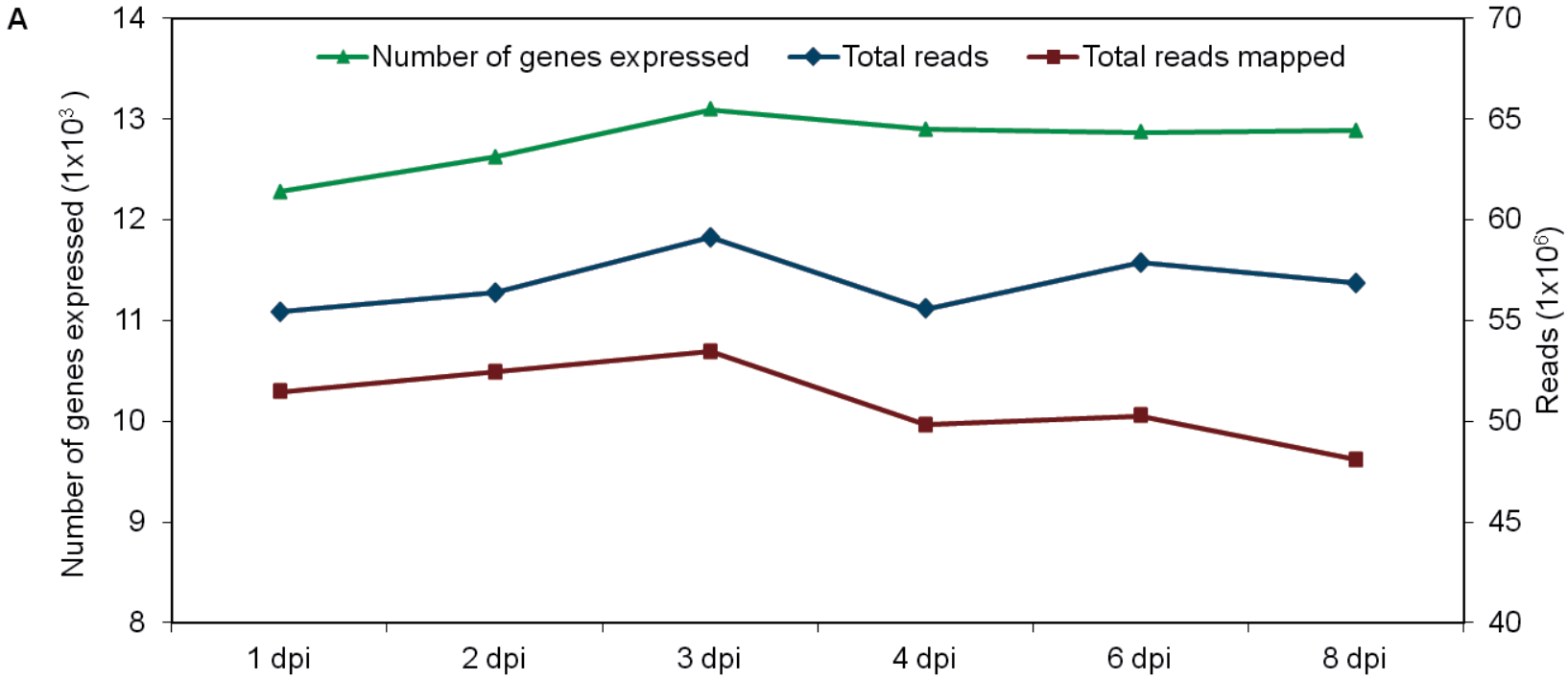
A
Number of genes expressed
Total reads
Total reads mapped
Number of genes expressed (1x10^3)
Reads (1x10^6)
14
13
12
11
10
9
8
70
65
60
55
50
45
40
1 dpi
2 dpi
3 dpi
4 dpi
6 dpi
8 dpi

Figure 4.3 (cont'd)

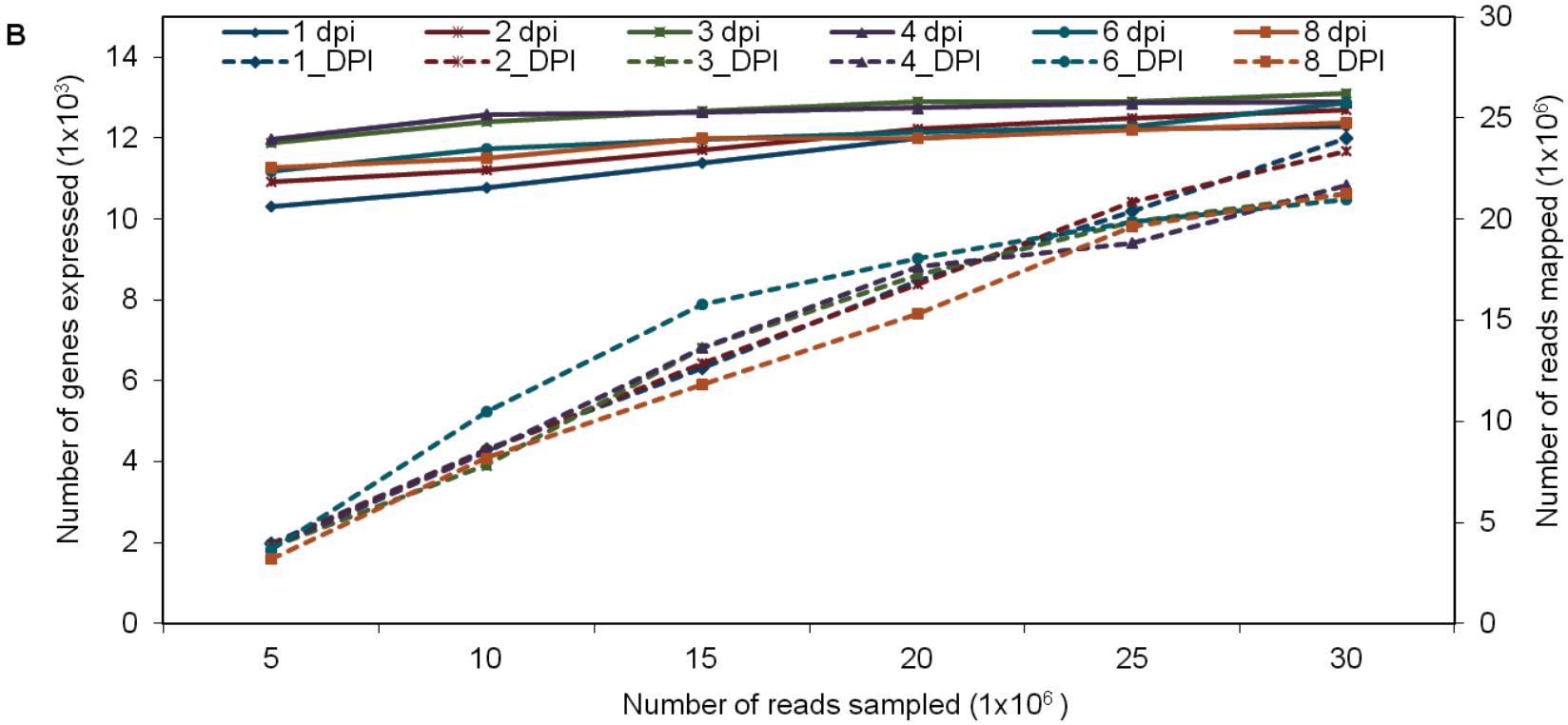

Figure 4.3 Comparison of total mRNA-Seq reads, reads mapped and number of genes expressed. (A) Total number of reads, total reads mapped, and number of genes expressed as determined from pooling of both biological replicates are shown. Reads were mapped to the *C. sativus* genome [6] using Bowtie version 0.12.5 [55] and TopHat version 1.2.0 [53]. Fragments per kilobase pair of exon model per million fragments mapped (FPKM) values for the annotated *C. sativus* genes were calculated using Cufflinks version 0.9.3 [55]. Genes were considered expressed if the FPKM value and 95% confidence interval lower boundary FPKM value was greater than 0.001 and zero, respectively. (B) Effect of sampling depth on detection of expressed genes. For all time points 5, 10, 15, 20, 25, and 30 million reads were randomly selected from the total pool of reads. Read mapping and expression abundance estimates were performed as describe above. Solid lines indicate number of genes expressed and the dashed lines indicate number of reads mapped at different time points. dpi, days post-inoculation.

been shown to be up-regulated in *Cucumis melo* (melon) in response to abiotic stress [38] as well as in an inbred *C. sativus* line 'IL57' with a high level of resistance to downy mildew [39]. The expression patterns of the top 20 highly expressed genes showed expression of genes involved in defense responses including catalases, chitinases, lipoxygenases, peroxidases, and protease inhibitors, beginning at 1 dpi and extending through 8 dpi (Table S4.2, available at www.plosone.org, e34954). The detection of defense-related genes within 1 dpi suggests that there is an active response by *C. sativus* to early infection stages of *Ps. cubensis*, including zoospore encystment, appressorium formation, and penetration via stomata (see accompanying paper, [36]). Additionally, no pathogen defense-related genes are present within the top 20 highly expressed genes in the mock-inoculated samples, which mainly consists of housekeeping genes (Table S4.2, available at www.plosone.org, e34954).

Correlation and cluster analyses were used to identify similarities in transcriptome profiles among the sampled time points. Pearson Correlation Coefficient (PCC) values for time point comparisons ranged from 0.78 to 0.93, with tight clustering readily apparent, revealing patterns that highlight the extent of transcriptional diversity underlying early (1 dpi), intermediate (2, 3, and 4 dpi), and advanced (6 and 8 dpi) stages of disease progression and the corresponding responses in host gene expression (Figure 4.1). As described above, *C. sativus* defense-related genes are expressed within 1 dpi of *Ps. cubensis* inoculation, and based on our correlation analysis (Figure 4.4), these likely represent a distinct cluster of genes responding specifically to initial recognition of sporangia, germination of zoospores, and zoospore

Figure 4.4

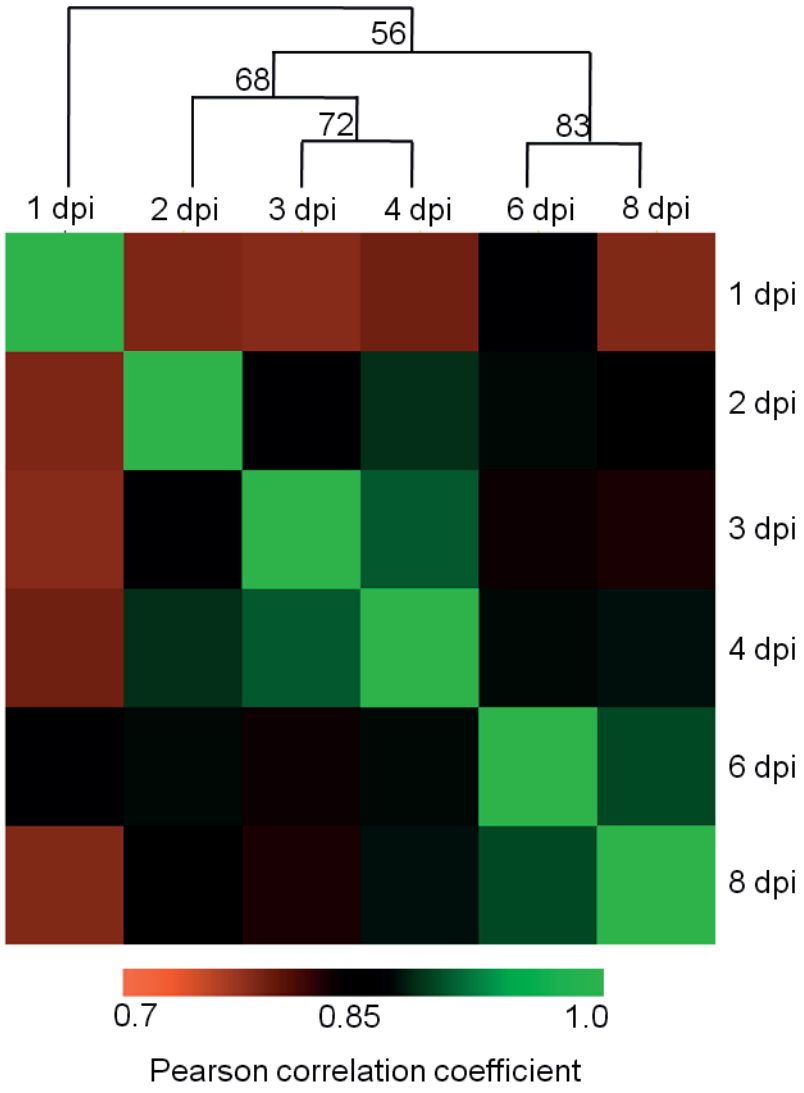

Figure 4.4 Correlation matrix of *Cucumis sativus* expression profiles during infection by *Pseudoperonospora cubensis*. Tissue samples were collected from *C. sativus* at different time points of infection with *Ps. cubensis*. Normalized transcript abundances for 14,476 genes were calculated as fragments per kilobase pair of exon model per million fragments mapped (FPKM) with Cufflinks version 0.9.3 [55]. Pearson correlation coefficient of log2 FPKM values were calculated for all pair-wise comparisons using R. Hierarchical clustering was performed using Pearson correlation distance metric and average linkage with the Multiple Experiment Viewer Software version 4.5 [57]. The bootstrap support values shown on tree nodes were obtained from 1000 bootstrap replicates. dpi, days post-inoculation.

encystment in stomata. Genes expressed at 2-4 dpi also cluster more with each other than to 1 dpi or to later time points, which also reflects the similar stages of *Ps. cubensis* infection apparent at days 2-4 dpi, indicating that these genes may be involved in host responses to hyphal penetration, growth and haustorium formation. The clustering of gene expression at later time points (6 and 8 dpi) likely corresponds to similar symptoms observed (Figure 4.1) as well as plant responses to extensive *Ps. cubensis* hyphal growth that is occurring at those time points (see accompanying paper, [36]).

Conserved host responses in *C. sativus* and *A. thaliana* in compatible interactions as measured through expression profiling

Host responses to pathogen challenge have been well documented in the model species *A. thaliana* [40], including those to the downy mildew pathogen *H. arabidopsidis* [34,37]. To identify genes induced in a compatible interaction with a downy mildew pathogen common to both *C. sativus* and *A. thaliana*, we identified single copy orthologous genes in both plant genomes and analyzed their expression patterns. A total of 7,396 clusters of single copy genes from both species were identified by clustering 23,248 and 27,416 representative protein coding genes from *C. sativus* and *A. thaliana*, respectively. Data from a previous microarray-based expression profiling [41] experiment of a compatible *A. thaliana-H. arabidopsidis* interaction was compared with our mRNA-Seq-based expression data. The *H. arabidopsidis* infection time points were 0, 0.5, 2, 4, and 6 dpi, similar to the 1-8 dpi time points assayed in this study. Spearman rank correlation coefficients (SCCs) of log2 expression values were calculated between the single copy orthologs at all time points in the two datasets;

between 2,136 and 3,446 gene pairs were included in the pair-wise comparisons. Among the six comparisons between similar time points, the SCC values ranged from 0.10 to 0.72 (Table S4.3, available at www.plosone.org, e34954). The correlations between early time points were the lowest between the two interactions, possibly due to the differences in penetration strategies between the two pathogens. The highest correlations were observed between 6 dpi (SCC = 0.72) followed by 2 dpi (SCC = 0.66) for both host-pathogen interactions (Figure 4.5). Overall, correlation coefficients between analogous time points (0.65 to 0.72) were greater than comparisons between dissimilar time points (0.10 to 0.33) indicating similar expression patterns for single copy orthologous genes in *C. sativus* and *A. thaliana*.

Differential gene expression throughout the infection process

Differences in gene expression patterns between time points can provide insight into the host response following pathogen perception and subsequent infection; thus, the mechanism(s) through which pathogenicity occurs may be inferred [34,42,43,44]. For example, a recent publication by Gaulin et al. [45] used a comparative approach analyzing the transcriptomes of *Aphanomyces euteiches* and *Phytophthora* spp. to identify novel pathogenicity factors and expanded repertoires for virulence. Through comparison of gene expression patterns across time points, we identified genes differentially expressed between all time points and the control sample, 1 dpi mock inoculation; between 1,170 and 3,286 genes were differentially expressed in pair-wise

Figure 4.5

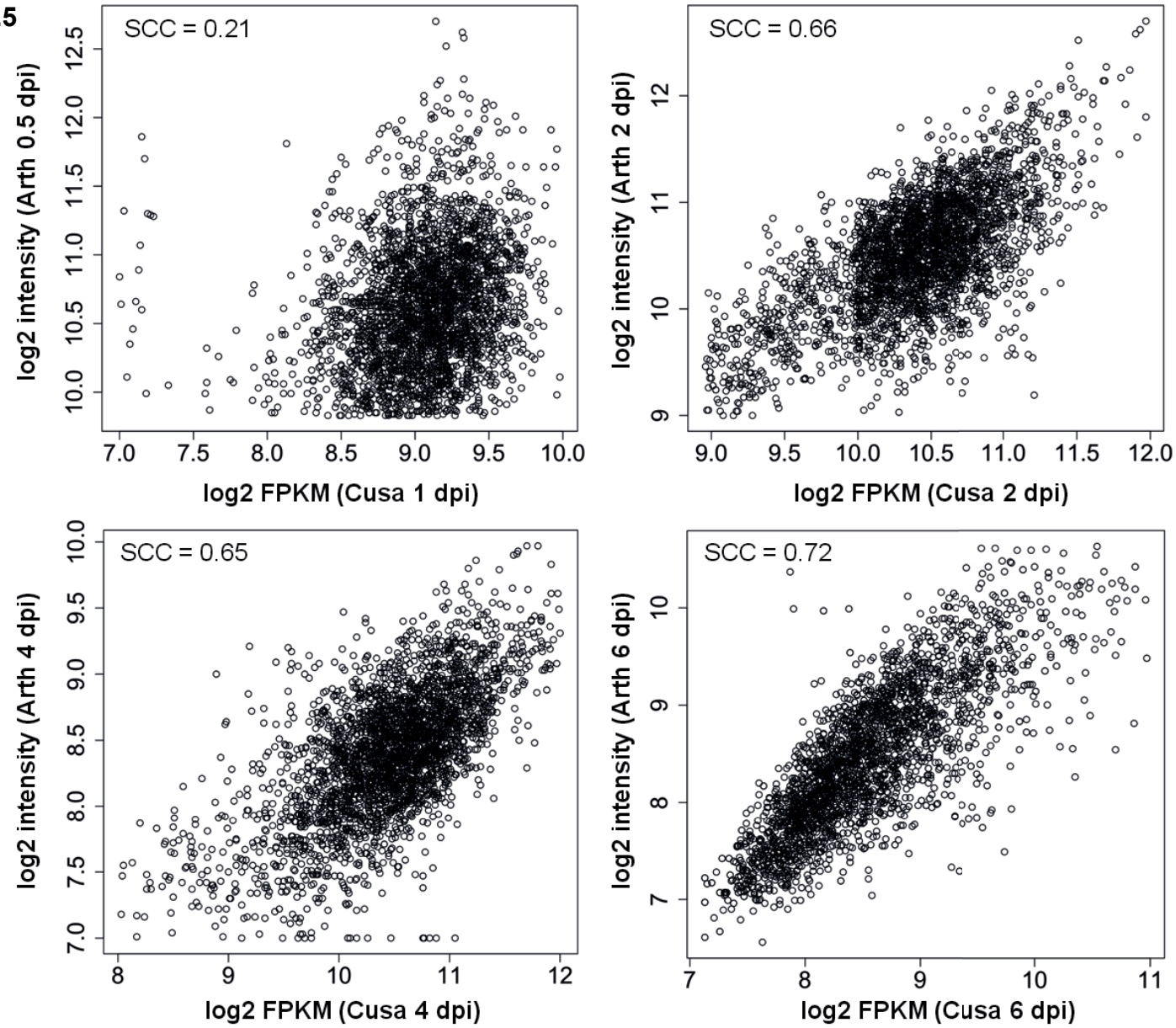
SCC = 0.21
log2 intensity (Arth 0.5 dpi)
log2 FPKM (Cusa 1 dpi)
SCC = 0.66
log2 intensity (Arth 2 dpi)
log2 FPKM (Cusa 2 dpi)
SCC = 0.65
log2 intensity (Arth 4 dpi)
log2 FPKM (Cusa 4 dpi)
SCC = 0.72
log2 intensity (Arth 6 dpi)
log2 FPKM (Cusa 6 dpi)

Figure 4.5 (cont'd) Comparison of orthologous gene expression in *Cucumis sativus* and *Arabidopsis thaliana* in a compatible interaction. Microarray expression profiles were obtained from time-course analyses of genes expressed in *A. thaliana* (Arth) during infection by *Hyaloperonospora arabidopsidis* [39]. Single copy orthologous genes between *C. sativus* (Cusa) and *A. thaliana* were identified using OrthoMCL [60]. Log2 transformed expression values of single copy orthologous genes expressed in the *C. sativus* mRNA-Seq dataset (log2 fragments per kilobase pair of exon model per million fragments mapped [FPKM]) and *A. thaliana* microarray dataset (log2 intensity) are shown as scatter plots. SCC, Spearman correlation coefficient; dpi, days post-inoculation.

comparison between the mock inoculated and/or the inoculated time points (Table 4.1, Table S4.4, available at www.plosone.org, e34954). In general, 12 to 31% of the genes tested under different conditions were differentially expressed in the pair-wise comparisons. In infected samples, comparisons of the three early to intermediate time points (i.e., 1, 2, and 3 dpi) showed a higher number of differentially expressed genes (2,214 to 3,286) than those between three intermediate to later (4, 6 and, 8 dpi) time points (1,612 to 2,074), suggesting more correlated gene expression in later stages of infection.

Gene co-expression pattern analyses

Using Weighted Gene Correlation Network Analysis (WGCNA) [46], we identified sets of highly correlated genes and constructed modules where all members are more highly correlated with each other than they are to genes outside the module [47]. Out of 15,286 genes expressed in the mock control or throughout the infection time course, 4,410 genes passed the Coefficient of Variance (CV) filter (0.4) and were retained for downstream analyses. Of the 4,410 genes, a total of 2,169 were assigned to 11 gene

Table 4.1 Number of genes differentially expressed between different time points.

	1 dpi¥	2 dpi	3 dpi	4 dpi	6 dpi	8 dpi
Mock-Control	1,170 (12%)§	1,906 (19%)	2,596 (26%)	2,509 (25%)	1,713 (17%)	2,556 (26%)
1 dpi		3,286 (31%)	2,948 (30%)	2,849 (27%)	2,010 (20%)	3,014 (30%)
2 dpi			2,214 (23%)	1,736 (18%)	2,718 (27%)	3,006 (29%)
3 dpi				1,590 (16%)	1,840 (18%)	2,218 (23%)
4 dpi					1,842 (18%)	2,074 (21%)
6 dpi						1,612 (16%)

Differential expression analysis was conducted using the cuffdiff program in Cufflinks version 0.9.3 [52], *Cucumis sativus* v2 annotation, and false discovery rate of 0.01.
¥dpi, days post-inoculation.
§Numbers in parenthesis indicate the percent of significantly different tests out of the total number of tests that could be performed for each pair-wise comparison.

modules that contained between 50 and 428 genes (Table S4.5); 2,241 genes were not assigned to any module. To visualize the relationship of the modules to each other with respect to the progression of infection, eigengenes for each module were calculated and displayed in a heat map [48]. As shown in Figure 4.6, some modules are representative of genes with correlated co-expression at primarily a single or two time points (Modules F, G, H, I, J, and K) whereas other modules represent genes that share broader co-expression patterns across multiple time points (Modules A, B, C, D, E). Examination of trend plots for the modules (Figure 4.7, Figure S4.2) reveals the direction and magnitude of gene expression patterns. Genes within Module B are expressed in the mock control and at 1 dpi, but are coordinately down-regulated at 2

dpi, remaining less abundant through 8 dpi (Figure 4.7). This set of 272 genes includes a large suite of genes implicated in resistance including six lipoxygenase genes, four cationic peroxidases, two cinnamate 4-hydroxylases, an anthocyanidin 3-glucosyltransferase, an anthocyanin 5-aromatic acyltransferase, a cysteine protease, and the elicitor-inducible protein EIG-J7. All of these genes have been implicated in defense responses in other plants [49,50,51,52,53,54], and in total, their coordinate down-regulation at 2 dpi is suggestive of an active mechanism by the pathogen to alter host defense responses. For example, the lipoxygenase pathway has been hypothesized to be involved in defense responses in cucumber, due to its expansion within the genome [6], and our data presented herein showing a rapid down-regulation of these transcripts during infection supports this hypothesis. In addition, cinnamate 4-hydroxylase, one of the primary enzymes in the phenylpropanoid pathway responsible for the conversion of cinnamic acid to *p*-coumaric acid, was shown to be rapidly induced in *C. sativus* in response to abiotic stress [51]. In this context, it is reasonable to

hypothesize that the down-regulation of its expression (observed in the present study), as well as that of other genes within the phenylpropanoid pathway, are suggestive of an active virulence mechanism to abrogate host responses associated with stress-induced signaling. In addition to the direct correlation among defense gene expression noted above, this module also includes the coordinated expression of 32 transcription factors that may also have critical roles in regulating genes responsible for the induction of defense signaling in the host. A total of 36 genes of no known function are also in this module, and further examination of their roles in defense responses may provide new

Figure 4.6

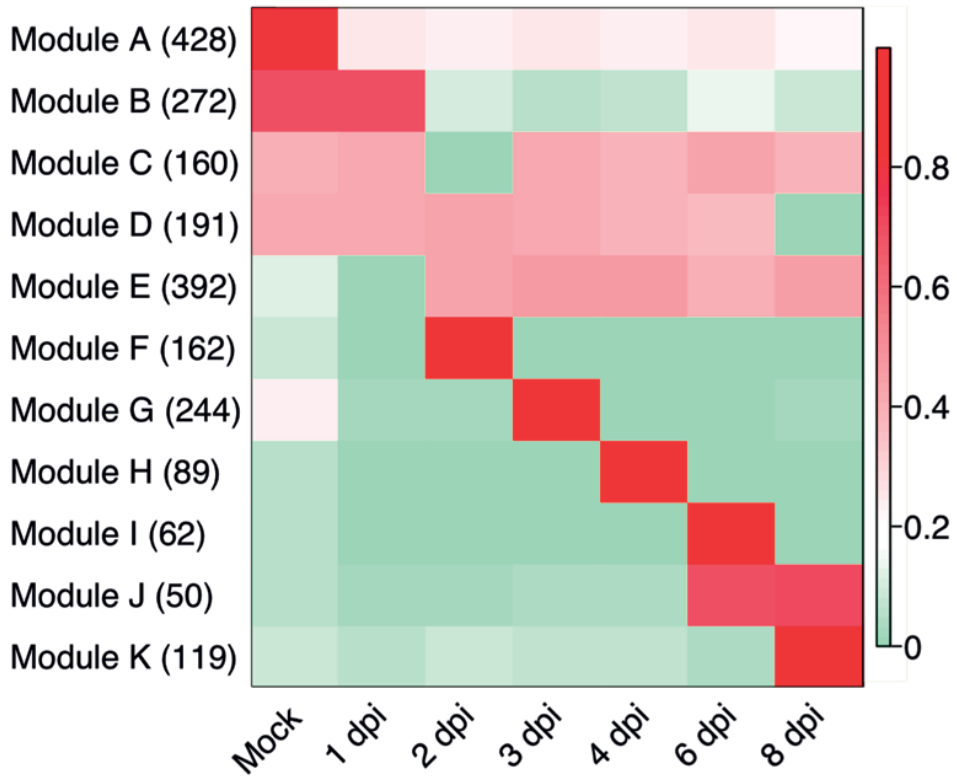

Figure 4.6 Heat map of eigengenes representing each gene module. The mock control and post-infection time points are represented in columns and the eigengenes for each of the 11 identified co-expression modules [46] are presented in rows. The numbers of genes in each module are given in parentheses. The eigengene values, which range from 0 to 1, are a measure of centrality and indicate the relative expression levels of all genes in the module (see Materials and Methods). dpi, days post-inoculation.

insight into critical host genes modulated by virulent pathogens.

Modules F, G, H, I, J, and K all have discrete time points where genes are up-regulated compared to the other sampled time points. As shown in Figure S4.2, genes in Module F (162 genes total) have a peak of expression solely at 2 dpi. This module contains 20 transcription factors which could be key to regulating genes from Module G that have peak expression at 3 dpi (Figure 4.7). Likewise, Module G (244 genes total) has 21

transcription factors that may regulate genes within Module H (89 genes total) that are coordinately up-regulated at 4 dpi (Figure 4.7). Within all of these modules (F, G, H, I, J, and K, 726 genes total) there are 199 genes with no known function and placement of these genes in transcriptional networks provides a new functional annotation and contextual information on their function.

CONCLUSIONS

The work described herein represents the first genome-scale analysis of the cucumber-downy mildew interaction in which we catalog gene expression throughout an 8 day infection period and identify differentially expressed genes that could be correlated with pathogen growth and development *in planta*. With expression profiles for nearly 15,000 genes during a compatible interaction, we have new insight into molecular events at the host-pathogen interface including a suite of defense response-related genes that are down-regulated early upon infection and transcriptional networks that respond in a temporal manner throughout the infection cycle. Most intriguingly, these networks include transcription factors and genes of no known function, which may have a role in the host-pathogen interaction.

Figure 4.7

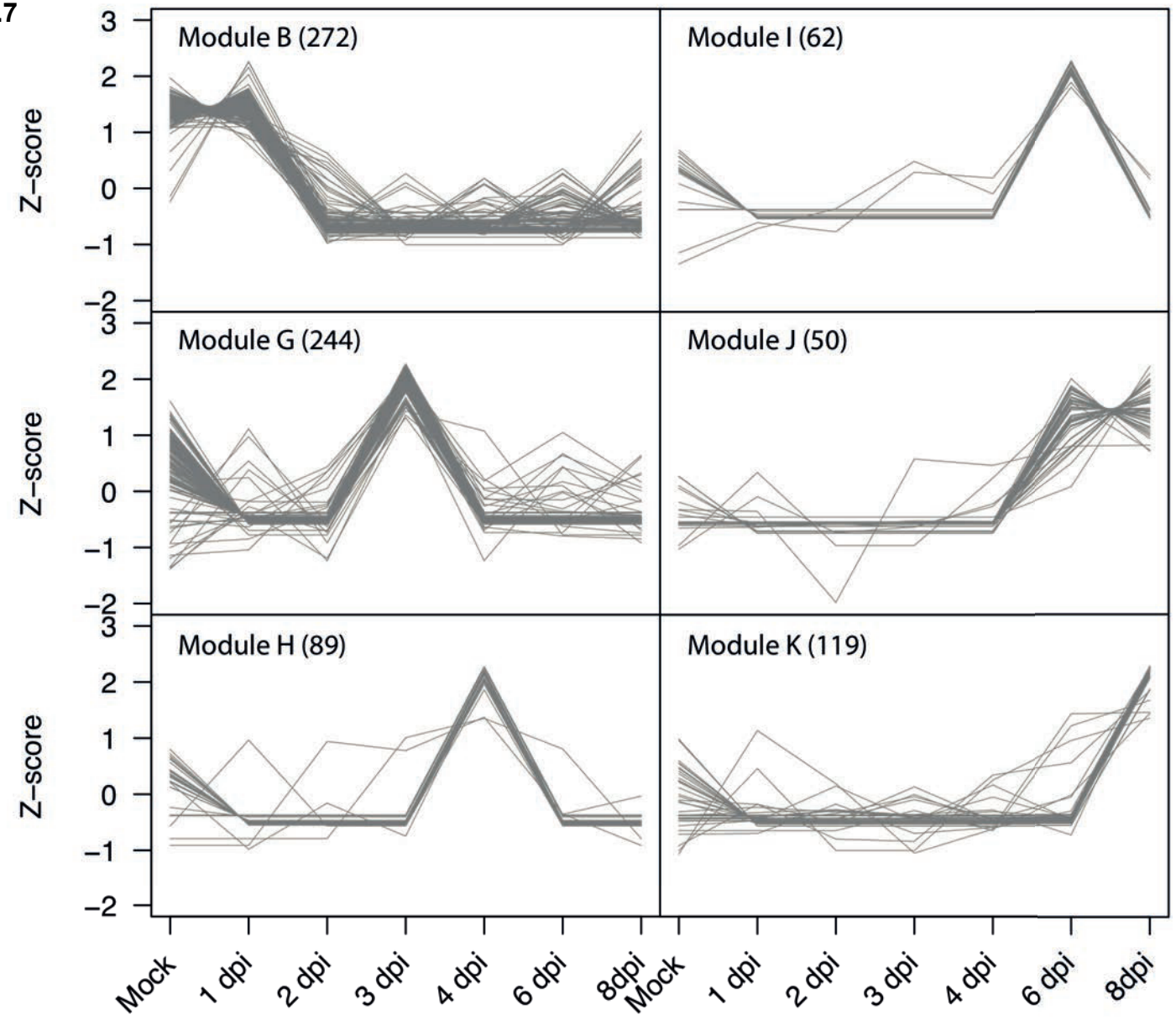
Module B (272)
Module I (62)
Module G (244)
Module J (50)
Module H (89)
Module K (119)
Z−score
3
2
1
0
−1
−2
Mock
1 dpi
2 dpi
3 dpi
4 dpi
6 dpi
8dpi

Figure 4.7 (cont'd) Trend plots of the normalized gene expression values from six identified gene co-expression modules. Modules consisting of genes down-regulated at 2 dpi (Module B), genes up-regulated at a single time point (Modules G, H, I, K) and genes up-regulated at two time-points (Module J) are shown. The number of genes in each module is shown in parentheses.

MATERIALS AND METHODS

Plant materials, inoculum, and pathogen infection

C. sativus cv. 'Vlaspik' was grown in growth chambers maintained at 22 °C with a 12 h light/dark photoperiod. *Ps. cubensis* MSU-1 was maintained as previously described [22]. The first fully expanded leaf of four-week-old cucumber plants was inoculated on the abaxial surface with 20-30 10 µl droplets of a 1 x 10^5 sporangia/ml solution. After inoculation, plants were kept at 100% relative humidity for 24 hours in the dark. Plants were returned to growth chambers for disease progression.

Sampling and experimental design

Samples from two biological replicates were collected at 1, 2, 3, 4, 6, and 8 dpi at the site of inoculation using a #3 cork borer. Additionally, a mock-inoculated sample (i.e.,10 µl dH_2O), which was inoculated as described above and kept at 100% relative humidity for 24 hours in the dark was collected at 1 dpi. Samples were frozen immediately in liquid nitrogen and stored at -80 °C until use.

Library preparation and sequencing

Total RNA was isolated from infected leaf discs using the RNeasy Mini Kit (Qiagen, Germantown, MD) and treated with DNase (Promega, Madison, WI) per the manufacturer's instructions. RNA concentration and quality was determined using the Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). The mRNA-Seq sample preparation was done using the Illumina mRNA-Seq kit (Illumina, San Diego CA) according to the manufacturer's protocol. Parallel sequencing was performed using an Illumina Genome Analyzer II (Illumina, Inc., San Diego, CA) at the Research Technology Support Facility (RTSF) at Michigan State University. Each library within a biological experiment was barcoded, six different barcodes for 6 time-points, pooled and run on multiple lanes. Two biological replicates of each time-point were sequenced multiple times and single-end reads between 35 and 42 bp were generated. Reads from both biological replicates were pooled for determining expression abundances. The mock-inoculated sample library was made as described above, but not barcoded and run in a single lane.

Processing of mRNA-Seq data

mRNA-Seq reads obtained from Illumina Pipeline version 1.3 were quality evaluated on the Illumina purity filter, percent low quality reads, and distribution of phred-like scores at each cycle. Reads were deposited in the National Center for Biotechnology Information Sequence Read Archive under accession number SRP009350. Reads in

FASTQ formats were aligned to the *C. sativus* [6] reference genome using TopHat v1.2.0/Bowtie v0.12.5 [55,56]. A reference annotation of *C. sativus* (version 2) from the Cucurbit Genomics Database (http://www.icugi.org/cgi-bin/ICuGI/misc/download.cgi) was provided in which a representative isoform, the gene model with the longest CDS, was used to estimate expression of the gene; a total of 23,248 gene models from a total of 25,600 gene models were used. All other isoforms were discarded from the annotation set. The minimum and maximum intron length was set to 5 and 50,000 bp, respectively; all other parameters were set to the default values.

Normalized gene expression levels were calculated using Cufflinks v0.9.3 [57] using the quartile normalization option [58] to improve differential expression calculations of lowly expressed genes. The maximum intron length parameter was set to 50,000 bp. All other parameters were used at the default settings. Sampling depth was evaluated on expression estimates by randomly selecting 5, 10, 15, 25 and 30 million reads from the total pool of reads at all time points. Pearson correlation coefficient analyses of log2 FPKM values were performed using R (http://cran.r-project.org/), in which all log2 FPKM values less than zero were set to zero. Only tests significant at $p = 0.05$ are shown. The correlation values were clustered with hierarchical clustering using a Pearson correlation distance metric with average linkage and depicted as a heat map. For each node, bootstrap support values were calculated from 1000 replicates using Multiple Experiment Viewer Software (MeV) v4.5 [59]. To examine biological variation, PCC was calculated for the log2 transformed FPKM values of the genes expressed in both replicates at a particular time point.

Microarray data acquisition and processing

Comparative analyses of host gene expression responses during a compatible interaction with the model species *A. thaliana* utilized microarry-based gene expression data from an *H. arabidopsidis-A. thaliana* time course experiment [41]. The dataset was comprised of *A. thaliana* genes expressed in response to *H. arabidopsidis* infection at 0, 0.5, 2, 4, and 6 dpi using the ATH1 Affymetrix platform. The probe intensity values were downloaded from Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/; GSE22274) [60], normalized using Robust Multichip Analysis method [61]. OrthoMCL [62] was used to identify clusters of orthologs and close paralogs in cucumber and *A. thaliana.* For the mRNA-Seq to microarray comparative analysis only single copy orthologous genes were considered for further analyses. The FPKM and intensity values were log2 transformed, and Spearman rank correlations of the single copy orthologs in both hosts were calculated in R (http://cran.r-project.org/). Only tests significant at $p = 0.05$ are shown.

Identification of differentially expressed genes

The Cuffdiff program within Cufflinks version 0.9.3 [57] was used to identify differentially expressed genes using pair-wise comparisons of the six time points and the control. The minimum number of alignments at a gene required to test was set to 100. Quartile normalization and a false discovery rate of 0.01 after Benjamini-Hochberg correction for multiple testing were used. Significance in the numbers of differentially expressed

genes between time points was tested using the two-sample t-test.

Functional analyses of differentially expressed genes

Functional annotation for all *C. sativus* genes were generated from BLAST searches of the UniProt databases (Uniref100) [63] and combined with Pfam (version 24.0, [64]) protein families assignment performed using HMMER3 [65]. *C. sativus* sequences were functionally annotated based on the best possible UniRef sequence match using a minimum E value cutoff of 1E-5. If there was no UniRef sequence match, functional annotations were assigned using Pfam domains. Transcription factors were annotated based on PFAM domains assignment.

Gene co-expression network analysis

Gene modules of highly correlated genes were identified using the WGCNA method [47] implemented in R. All gene FPKM expression values were log2 transformed and any transformed FPKM value less than 1 was converted to zero. Genes without variation across the mock sample and 6 time points were filtered out using a Coefficient of Variation (CV = σ/μ) cutoff of 0.4. The β and treecut parameters for WGCNA were 13 and 0.9, respectively. All other parameters were used with their default values. Eigengenes for each gene module [48] were calculated and presented as a heat map.

ACKNOWLEDGEMENTS

We thank members of the Day lab for critical reading of the manuscript.

APPENDIX

Figure S4.1

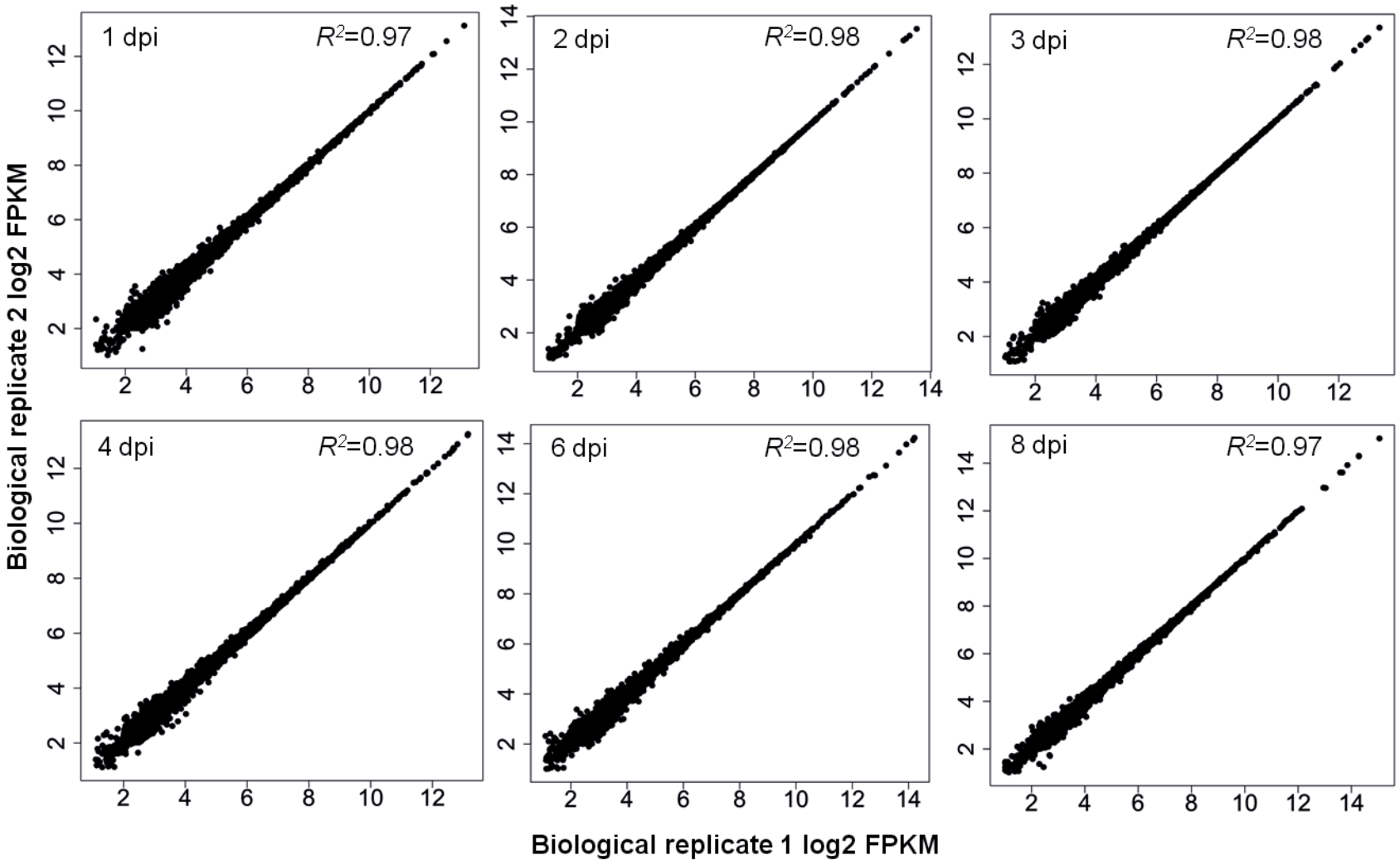
1 dpi
R^2=0.97
2 dpi
R^2=0.98
3 dpi
R^2=0.98
4 dpi
R^2=0.98
6 dpi
R^2=0.98
8 dpi
R^2=0.97
Biological replicate 2 log2 FPKM
Biological replicate 1 log2 FPKM

Figure S4.1 (cont'd) Concordance of expression values in two biological replicates of *Cucumis sativus* during infection by *Pseudoperonospora cubensis*. Reads from different time points were mapped to the *C. sativus* genome using Bowtie version 0.12.5 and TopHat version 1.2.0. Fragments per kilobase pair of exon model per million fragments mapped (FPKM) values were calculated using Cufflinks version 0.9.3 and the *C. sativus* genome annotations. For each time point, log2 transformed FPKM values of equal number of genes from both replicates are plotted. R^2, correlation coefficient; dpi, day post-inoculation.

Figure S4.2

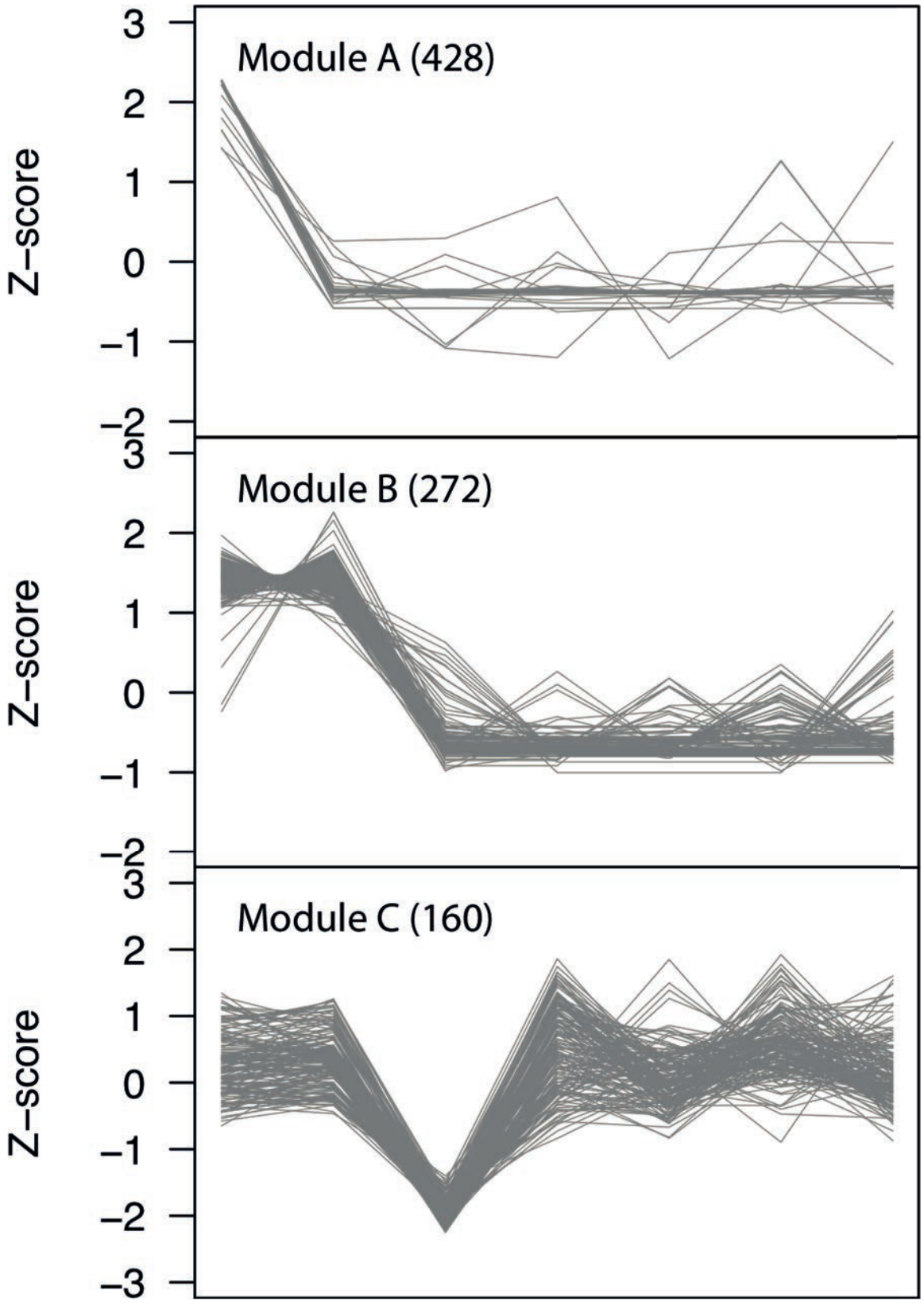

Figure S4.2 (cont'd)

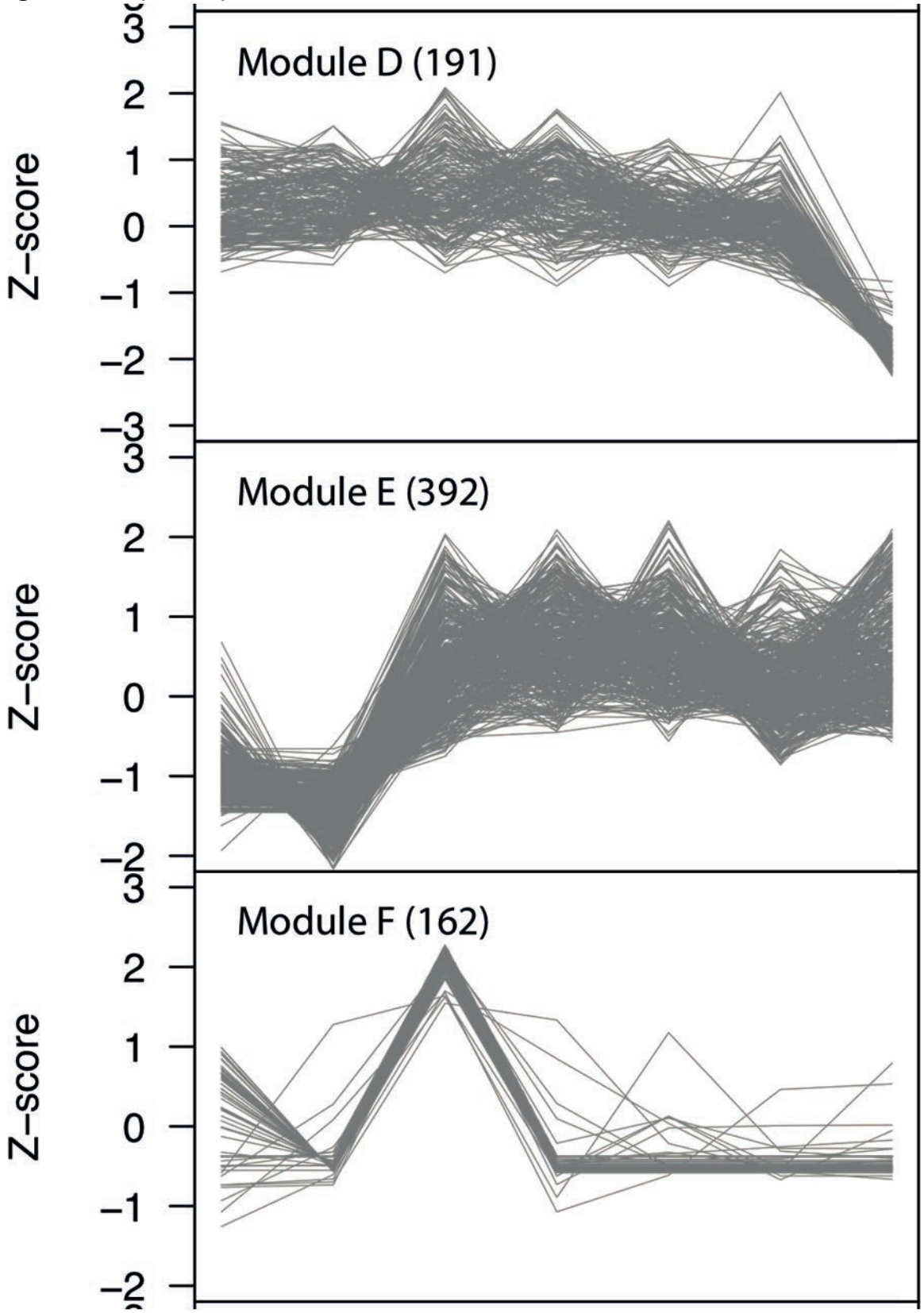

Figure S4.2 (cont'd)

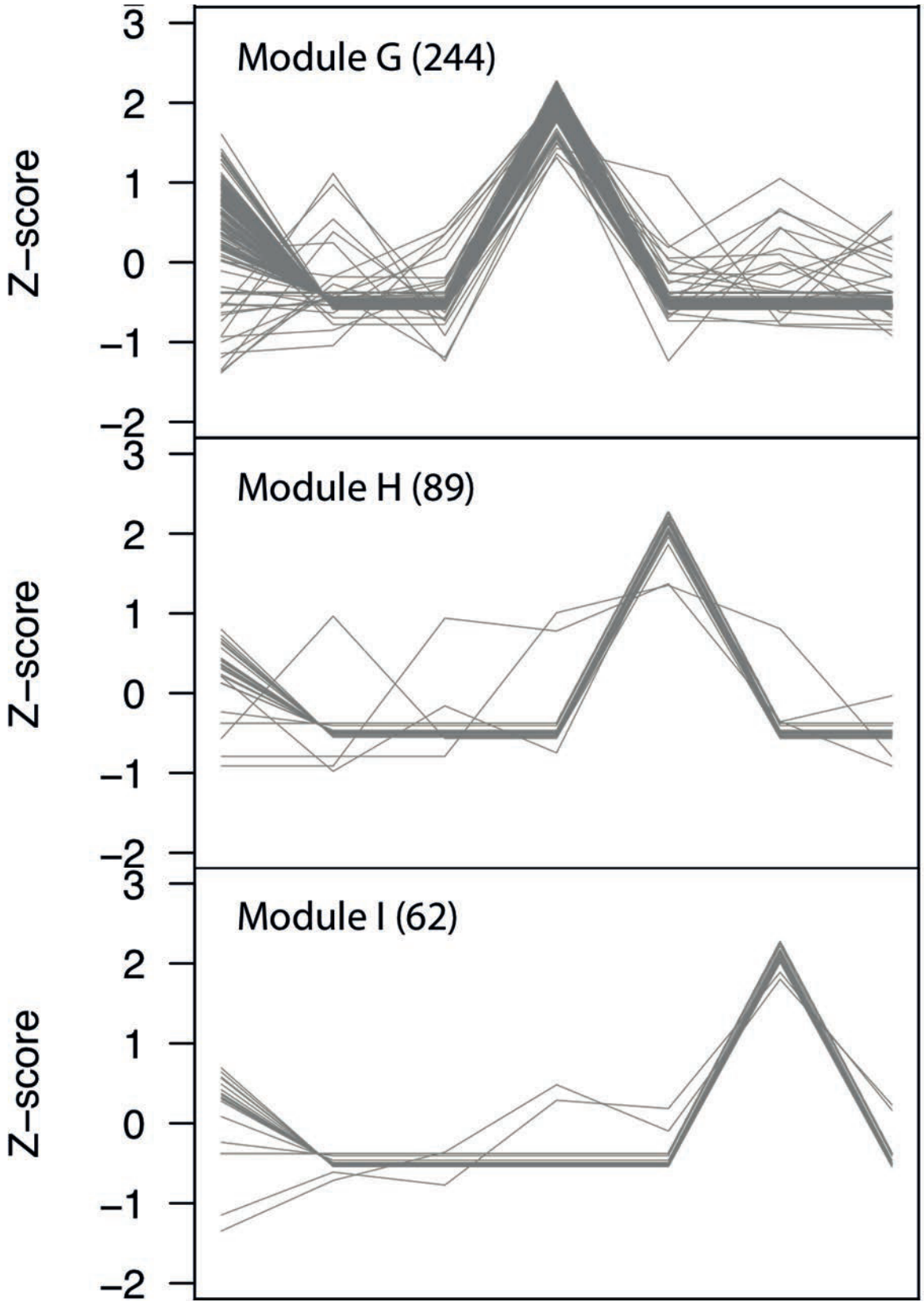

Figure S4.2 (cont'd)

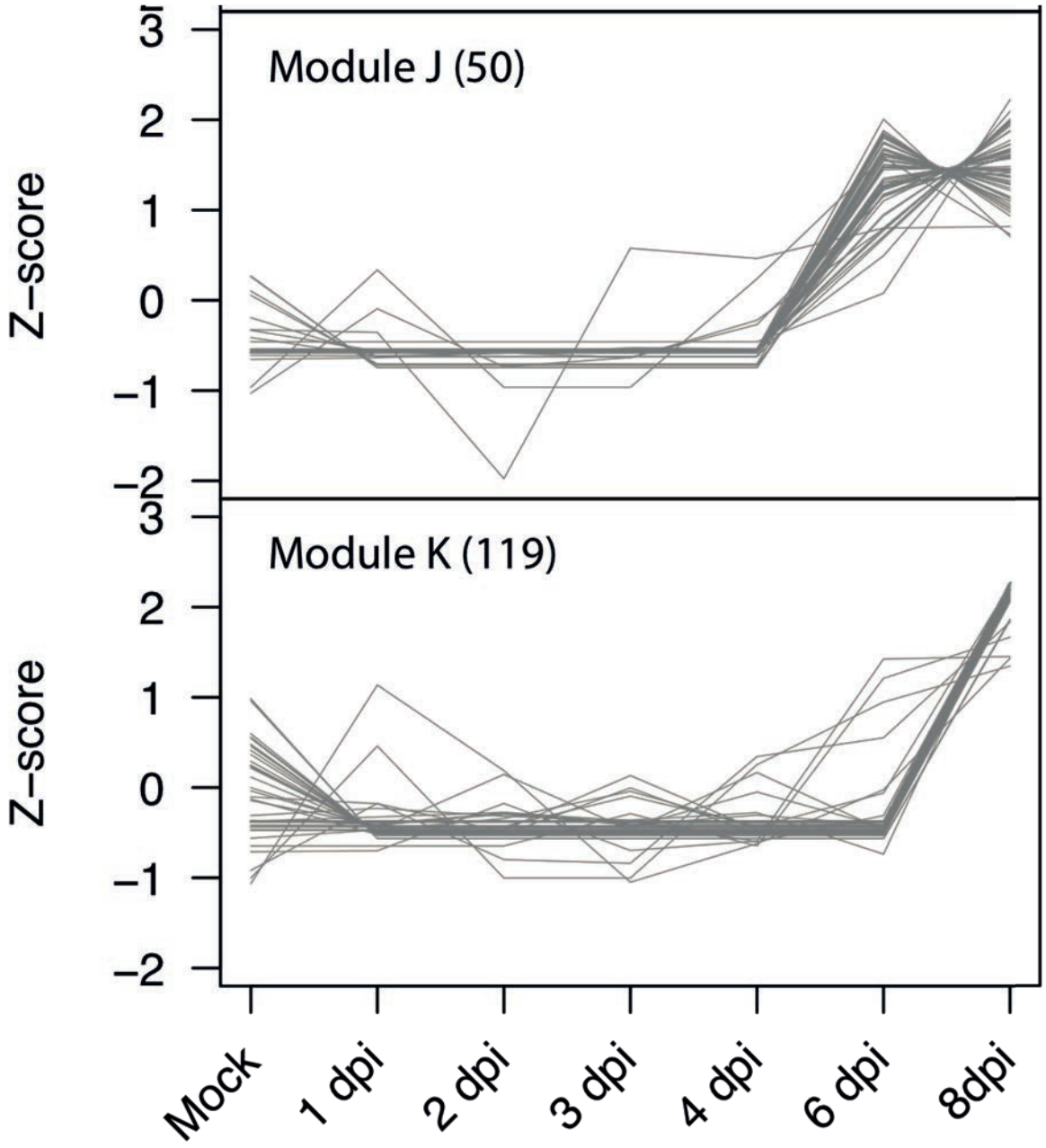

Figure S4.2 Trend plots for all 11 modules. All 11 modules generated using WGCNA are shown (Modules A through K). The number of genes in each module is shown in parentheses.

REFERENCES

REFERENCES

1. Tanurdzic M, Banks JA (2004) Sex-determining mechanisms in land plants. Plant Cell 16 Suppl: S61-71.

2. Lough TJ, Lucas WJ (2006) Integrative plant biology: role of phloem long-distance macromolecular trafficking. Ann Rev Plant Biol 57: 203-232.

3. Hammerschmidt R (1999) PHYTOALEXINS: What have we learned after 60 years? Ann Rev Phytopathol 37: 285-306.

4. Hammerschmidt R (1999) Induced disease resistance: how do induced plants stop pathogens? Physiol Mol Plant Pathol 55: 77-84.

5. Ren Y, Zhang Z, Liu J, Staub JE, Han Y, et al. (2009) An integrated genetic and cytogenetic map of the cucumber genome. PLoS ONE 4: e5795.

6. Huang S, Li R, Zhang Z, Li L, Gu X, et al. (2009) The genome of the cucumber, *Cucumis sativus* L. Nat Genet 41: 1275-1281.

7. Wóycicki R, Witkowicz J, Gawroński P, Dąbrowska J, Lomsadze A, et al. (2011) The genome sequence of the North-European cucumber (*Cucumis sativus* L.) unravels evolutionary adaptation mechanisms in plants. PLoS ONE 6: e22728.

8. Guo S, Zheng Y, Joung JG, Liu S, Zhang Z, et al. (2010) Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types. BMC Gen 11: 384.

9. Ando K, Grumet R (2010) Transcriptional profiling of rapidly growing cucumber fruit by 454-pyrosequencing analysis. J Am Soc Hort Sci 135: 291-302.

10. Olczak-Woltman H, Schollenberger M, Niemirowicz-Szczytt K (2009) Genetic background of host-pathogen interaction between *Cucumis sativus* L. and *Pseudomonas syringae* pv. *lachrymans*. J Appl Genet 50: 1-7.

11. Savory EA, Granke LL, Quesada-Ocampo LM, Varbanova M, Hausbeck MK, et al. (2011) The cucurbit downy mildew pathogen *Pseudoperonospora cubensis*. Mol Plant Pathol 12: 217-226.

12. Colucci SJ, Wehner TC, Holmes GJ (2006) The downy mildew epidemic of 2004 and 2005 in the Eastern United States. In: Holmes G, editor. Proc *Cucurbitaceae* 2006. Raleigh, NC: Universal Press.

13. Lebeda A, Pavelková J, Urban J, Sedláková B (2011) Distribution, host range and disease severity of *Pseudoperonospora cubensis* on cucurbits in the Czech Republic. J Phytopathol 159: 589-596.

14. Arauz LF, Neufeld KN, Lloyd AL, Ojiambo PS (2010) Quantitative models for germination and infection of *Pseudoperonospora cubensis* in response to temperature and duration of leaf wetness. Phytopathol 100: 959-367.

15. Thomas C, Inaba T, Cohen Y (1987) Physiological Specialization in *Pseudoperonospora cubensis*. Phytopathol 77: 1621-1624.

16. Shetty NV, Wehner TC, Thomas CE, Doruchowski RW, Vasanth Shetty KP (2002) Evidence for downy mildew races in cucumber tested in Asia, Europe, and North America. Scientia Hort 94: 231–239.

17. Lebeda A, Widrlechner MP (2003) A set of Cucurbitaceae taxa for differentiation of *Pseudoperonespora cubensis* pathotypes. J Plant Dis Prot 110: 337–349.

18. Cohen Y, Meron I, Mor N, Zuriel S (2003) A new pathotype of *Pseudoperonospora cubensis* causing downy mildew in cucurbits in Israel. Phytoparasitica 31: 458–466.

19. Lebeda A (2002) Pathogenic variation of *Pseudoperonospora cubensis* in the Czech Republic and some other European countries. II International Symposium on Cucurbits 588. pp. 137–141

20. Lebeda A, Urban J (2004) Distribution, harmfulness and pathogenic variability of cucurbit downy mildew in the Czech Republic. Acta Fytotech Zootech 7: 170-173.

21. Sarris P, Abdelhalim M, Kitner M, Skandalis N, Panopoulos N, et al. (2009) Molecular polymorphisms between populations of *Pseudoperonospora cubensis* from Greece and the Czech Republic and the phytopathological and phylogenetic implications. Plant Pathol 58: 933–943.

22. Tian M, Win J, Savory E, Burkhardt A, Held M, et al. (2011) 454 Genome sequencing of *Pseudoperonospora cubensis* reveals effector proteins with a QXLR translocation motif. Mol Plant-Microbe Interact 24: 543-553.

23. Savory EA, Zou C, Adhikari BN, Hamilton JP, Buell CR, et al. (2012) Alternative Splicing of a Multi-Drug Transporter from *Pseudoperonospora cubensis* Generates an RXLR Effector Protein That Elicits a Rapid Cell Death. PLoS One 7: e34701.

24. Durrant WE, Dong X (2004) Systemic acquired resistance. Ann Rev Phytopathol 42: 185-209.

25. Phuntumart V, Marro P, Métraux J-P, Sticher L (2006) A novel cucumber gene associated with systemic acquired resistance. Plant Sci 171: 555-564.

26. Sticher L, Mauch-Mani B, Metraux JP (1997) Systemic acquired resistance. Ann Rev Phytopathol 35: 235-270.

27. Bent AF, Mackey D (2007) Elicitors, effectors, and R genes: The new paradigm and a lifetime supply of questions. Ann Rev Phytopathol 45: 399-436.

28. Call AD, Wehner T (2010) Gene list 2010 for cucumber. Cucurbi Genet Coop Rep 28-29: 105-141.

29. Runge F, Choi Y-J, Thines M (2011) Phylogenetic investigations in the genus *Pseudoperonospora* reveal overlooked species and cryptic diversity in the *P. cubensis* species cluster. Eur J Plant Pathol 129: 135-146.

30. Taler D, Galperin M, Benjamin I, Cohen Y, Kenigsbuch D (2004) Plant eR genes that encode photorespiratory enzymes confer resistance against disease. Plant Cell 16: 172-184.

31. Criswell A (2008) Screening cucumber (*Cucumis sativus)* for resistance to downy mildew (*Pseudoperonospora cubensis*). NCSU Thesis.

32. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, et al. (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456: 470 - 476.

33. Boddu J, Cho S, Muehlbauer GJ (2007) Transcriptome analysis of trichothecene-induced gene expression in barley. Mol Plant-Microbe Interact 20: 1364-1375.

34. Huibers RP, de Jong M, Dekter RW, Van den Ackerveken G (2009) Disease-specific expression of host genes during downy mildew infection of Arabidopsis. Mol Plant-Microbe Interact 22: 1104-1115.

35. Gupta S, Chakraborti D, Sengupta A, Basu D, Das S (2010) Primary metabolism of chickpea is the initial target of wound inducing early sensed *Fusarium oxysporum* f. sp. ciceri race I. PLoS ONE 5: e9030.

36. Savory EA, Adhikari BN, Hamilton JP, Vaillancourt B, Buell CR, et al. (2012) mRNA-Seq Analysis of the *Pseudoperonospora cubensis* Transcriptome During Cucumber (*Cucumis sativus* L.) Infection. PLoS One 7: e35796.

37. Coates ME, Beynon JL (2010) *Hyaloperonospora arabidopsidis* as a pathogen model. Ann Rev Phytopathol 48: 329-345.

38. Volk GM, Haritatos EE, Turgeon R (2003) Galactinol synthase gene expression in melon. J Am Soc Hort Sci 128: 8-15.

39. Li JW, Liu J, Zhang H, Xie CH (2011) Identification and transcriptional profiling of differentially expressed genes associated with resistance to *Pseudoperonospora cubensis* in cucumber. Plant Cell Rep 30: 345-357.

40. Knepper C, Day B (2010) From perception to activation: The molecular-genetic and biochemical landscape of disease resistance signaling in plants. The Arabidopsis Book: e012.

41. Wang W, Barnaby JY, Tada Y, Li H, Tor M, et al. (2011) Timing of plant immune responses by a central circadian regulator. Nature 470: 110-114.

42. Miranda M, Ralph SG, Mellway R, White R, Heath MC, et al. (2007) The transcriptional response of hybrid poplar (*Populus trichocarpa* x *P. deltoides*) to infection by Melampsora medusae leaf rust involves induction of flavonoid pathway genes leading to the accumulation of proanthocyanidins. Mol Plant-Microbe Interact 20: 816-831.

43. Polesani M, Bortesi L, Ferrarini A, Zamboni A, Fasoli M, et al. (2010) General and species-specific transcriptional responses to downy mildew infection in a susceptible (*Vitis vinifera*) and a resistant (*V. riparia*) grapevine species. BMC Gen 11: 117.

44. Polesani M, Desario F, Ferrarini A, Zamboni A, Pezzotti M, et al. (2008) cDNA-AFLP analysis of plant and pathogen genes expressed in grapevine infected with *Plasmopara viticola*. BMC Gen 9: 142.

45. Gaulin E, Madoui MA, Bottin A, Jacquet C, Mathe C, et al. (2008) Transcriptome of *Aphanomyces euteiches*: new oomycete putative pathogenicity factors and metabolic pathways. PLoS ONE 3: e1723.

46. Langfelder P, Zhang B, Horvath S (2008) Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 24: 719 - 720.

47. Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. Stat App Gen Mol Biol 4: Article 17.

48. Langfelder P, Horvath S (2007) Eigengene networks for studying the relationships between co-expression modules. BMC Sys Biol 1: 54.

49. Takemoto D, Demirdöven A (2007) Lipoxygenase in fruits and vegetables: A review. Enz Micro Tech 40: 491-496.

50. Hilaire E, Young SA, Willard LH, McGee JD, Sweat T, et al. (2001) Vascular defense responses in rice: peroxidase accumulation in xylem parenchyma cells and xylem wall thickening. Mol Plant-Microbe Interact 14: 1411-1419.

51. Varbanova M, Porter K, Lu F, Ralph J, Hammerschmidt R, et al. (2011) Molecular and biochemical basis for stress-induced accumulation of free and bound p-coumaraldehyde in cucumber. Plant Physiol 157: 1056-1066.

52. Brader G, Djamei A, Teige M, Palva ET, Hirt H (2007) The MAP kinase kinase MKK2 affects disease resistance in Arabidopsis. Mol Plant-Microbe Interact 20: 589-596.

53. Bernoux M, Timmers T, Jauneau A, Brière C, de Wit PJGM, et al. (2008) RD19, an Arabidopsis cysteine protease required for RRS1-R-mediated resistance, Is

relocalized to the nucleus by the *Ralstonia solanacearum* PopP2 effector. Plant Cell 20: 2252-2264.

54. Takemoto D, Doke N, Kawakita K (2001) Characterization of elicitor-inducible tobacco genes isolated by differential hybridization. J Gen Plant Pathol 67: 89-96.

55. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105-1111.

56. Langmead B, Trapnell C, Pop M, Salzberg S (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25.

57. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28: 511-515.

58. Bullard J, Purdom E, Hansen K, Dudoit S (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics 11: 94.

59. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, et al. (2006) TM4 microarray software suite. Methods Enzymol 411: 134-193.

60. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30: 207-210.

61. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostat 4: 249 - 264.

62. Li L, Stoeckert CJ, Jr., Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13: 2178-2189.

63. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH (2007) UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics 23: 1282-1288.

64. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, et al. (2000) The Pfam Protein Families Database. Nucleic Acids Res 28: 263-266.

65. Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. Nucleic Acids Res 39: W29-37.

CHAPTER 5

Conclusions and Future Directions

CONCLUSIONS

Cucurbit downy mildew, caused by the oomycete pathogen *Pseudoperonospora cubensis*, emerged as a significant economic threat to cucumber production in the United States during the 2004 and 2005 growing seasons and is currently the major limiting factor of cucumber production [1]. Although downy mildew had been a major issue in Europe since the mid-1980's, in the U.S., the disease on cucumber had been successfully controlled since the 1940's through host resistance. As such, only limited research was conducted to understand the biology, genetics, and virulence of the pathogen. The focus of this dissertation was to expand knowledge of this economically important oomycete pathogen using cell biology, genetic, genomic, and transcriptomic approaches to better understand the molecular-genetic basis for the interaction between cucumber (*Cucumis sativus*) and *Ps. cubensis*.

The establishment of genetic resources in the form of a draft genome assembly laid the groundwork for additional analyses, including identification and characterization of candidate effector proteins (Chapter 2) as well as providing a framework for mapping mRNA-Seq reads from a large-scale gene expression study of infection (Chapter 3). One RXLR-type effector protein, *Psc*RXLR1 was characterized in detail and shown to be the result of alternative splicing of a non-effector protein (Chapter 2). Additionally, due to the obligate nature of *Ps. cubensis* and our sampling methods for the mRNA-Seq study, we were able to measure host gene expression changes during *C. sativus* infection with *Ps. cubensis*, providing the first in-depth analysis of this important plant-pathogen interaction (Chapter 4). Overall, the research presented in this dissertation

greatly expands our knowledge of the *C. sativus-Ps. cubensis* interaction and provides much-needed insight into the virulence of *Ps. cubensis* and the defense response of *C. sativus*.

The establishment of genomic resources for major oomycete pathogens such as *Phytophthora infestans* [2], *Phytophthora sojae*, *Phytophthora ramorum* [3], *Pythium ultimum* [4], as well as for the *Arabidopsis thaliana* pathogen, *Hyaloperonospora arabidopsidis* [5], and subsequent identification of RXLR and RXLR-like effector proteins has anchored the recent evolution of molecular plant pathology [6]. The study of oomycete effector proteins has been central to the study of oomycete pathogenicity and virulence and the understanding of the interactions between these pathogens and their hosts [7,8,9,10,11,12,13,14,15,16]. Therefore, the first step in identifying these important pathogenicity and virulence determinants in *Ps. cubensis* was the generation of a 64.4 Mbp genomic assembly of the MSU-1 isolate and prediction of 23,519 loci and 23,522 gene models (Chapter 2). Due to their modular structure, including an N-terminal signal peptide, an RXLR translocation motif, and a C-terminal functional domain, RXLR effectors are readily identified via bioinformatics [17,18]. We identified 271 candidate effector proteins within the *Ps. cubensis* genome with variable RXLR motifs, including the previously identified R and Q, predicting 20 possible amino acids at position R1. Additionally, mRNA-Seq analysis of infection provided expression support for 19 of the 20 possible R1 substitutions (except Y, *Tyr*; Table S2.1, Figure 3.5). While candidate effectors with these R1 amino acid substitutions, with the exception of QXLR [19], have yet to be functionally validated, their expression during a compatible

interaction supports their putative virulence function and role as *bona fide* effector proteins. This variation may provide clues to the diversity among *Ps. cubensis* pathotypes in terms of their virulence and host specificity.

In-depth functional analysis of one *Ps. cubensis* effector protein, RXLR protein 1 (*Psc*RXLR1) led to a surprising and interesting result - it was demonstrated that *Psc*RXLR1 arises as a product of alternative splicing (Figure 2.5), making this the first example of an alternative splicing event in plant pathogenic oomycetes transforming a non-effector gene into a functional effector protein. Using a set of experiments designed to validate effector protein function, it was shown that *Psc*RXLR1 was up-regulated during the early stages of infection and that heterologous expression of *Psc*RXLR1 in *Nicotiana benthamiana* elicits a rapid cell death phenotype (Figure 2.3, Figure 2.4). Additionally, characterization of the closest *P. infestans* ortholog, PITG_17484, a member of the Drug/Metabolite Transporter (DMT) superfamily to *Psc*_RXLR1 confirmed that it did not elicit the same cell death response (Figure 2.3). The orthologous relationship between *Psc*RXLR1 and PITG_17484 led us to question if such effector-non-effector ortholog pairs are common among oomycete plant pathogens. We subsequently examined the relationship(s) among putative ortholog pairs in *Ps. cubensis* and *P. infestans*. Of 271 predicted *Ps. cubensis* effector proteins, only 109 (41%) had a putative ortholog in *P. infestans* and evolutionary rate analysis of these orthologs showed an evolutionary rate significantly faster than most other genes (Figure 2.1). In total, these data provide a basis for comparative analysis of candidate effector proteins and their non-effector orthologs as a means of understanding function

and evolutionary history of pathogen effectors. In addition, the discovery of an effector protein arising from an alternative splicing event may indicate an adaptive mechanism utilized by *Ps. cubensis* and potentially other pathogens to generate proteome complexity during infection.

Next generation sequencing of the transcriptome (mRNA-Seq) permits deep, robust assessments of transcript abundance and structure, and when applied to host-pathogen interactions, enables insight into pathogen mechanisms to suppress and subvert host defense responses [20,21,22,23]. In this dissertation, the first large-scale global transcriptome analysis of *Ps. cubensis* infection of a susceptible *C. sativus* cultivar, 'Vlaspik' is presented, yielding new information about both pathogen and host gene expression during a compatible interaction (Chapter 3, 4). Sampling *Ps. cubensis*-infected *C. sativus* tissue during an 8-day time course and then separating species-specific reads *in silico* enabled simultaneously collection and analysis of infection transcriptome dynamics from both pathogen and host. As described herein, we were able to detect the expression of 7,821 *Ps. cubensis* genes (Chapter 3) and 14,476 cucumber genes (Chapter 4) throughout the infection process from 1 day post-inoculation (1 dpi) to 8 dpi, providing the first genome-scale analysis of the cucurbit downy mildew interaction.

In *Ps. cubensis*, genes involved in virulence, including RXLR and RXLR-like effectors, Crinkler (CRN) effectors, and host-targeted hydrolytic enzymes acting on plant proteinases, lipases, and several sugar-cleaving enzymes were all differentially

expressed throughout the infection time course. Corresponding with our visual assessment of symptoms and infection structures, clustering and co-expression analyses identified distinct modules of *Ps. cubensis* genes that were representative of early, intermediate, and late infection stages. The 1 dpi time point, when encystment of zoospores is occurring, represents a unique pattern of gene expression supported by multiple modes of analysis. Gene expression at this time point poorly correlated with all other time points (Figure 3.4), had the highest percentage of differentially expressed genes across all pair-wise comparisons (Table 3.1), and was represented by a unique module, Module 1 (Figure 3.8), when analyzed using Weighted Gene Correlation Network Analysis (WGCNA). The genes represented here are likely those involved in zoospore encystment, appressorium production, and the initial penetration of stomata. Our data also supports an intermediate, transition stage of infection, represented mainly by overlapping gene expression between the 2, 3, and 4 dpi time points. Module 2, containing 508 genes including candidate RXLR-type effectors, CRNs, and haustorium-specific proteins, represents expression during hyphal penetration, growth, and initiation of haustoria formation, which occurs between 2-4 dpi. The late stage of infection, at 6 and 8 dpi, is characterized by extensive hyphal growth within the mesophyll (Figure 3.1) and a transition to the expression of genes likely involved in sporulation (Module 6, Figure 3.8). Overall, these expression data and analyses and their correlation with pathogen growth have advanced our understanding of molecular and genetic events in the infection of *Ps. cubensis*.

During *Ps. cubensis* infection, expression of defense-related genes, including catalases, chitinases, lipoxygenases, peroxidases, and protease inhibitors, was highly up-regulated in *C. sativus* within 1 dpi (Table S4.2), suggesting an active host defense response to early infection by *Ps. cubensis*. These genes are coordinately down-regulated at 2 dpi (Module B, Figure 4.7) and remain lowly expressed for the continuation of the time course, suggesting that they are suppressed by the pathogen.

With expression profiles for nearly 15,000 genes during a compatible interaction, we have new insight into molecular events at the host-pathogen interface including a suite of defense response-related genes that are down-regulated early upon infection and transcriptional networks that respond in a temporal manner throughout the infection cycle. Most intriguingly, these networks include transcription factors and genes of no known function, which may have a role in the host-pathogen interaction.

Overall, the work presented in this dissertation represents a substantial advance in the understanding of the *Ps. cubensis*-*C. sativus* interaction. Using a variety of methods ranging from cell biology to genetics to large-scale transcriptomic analysis, it serves to advance knowledge of *Ps. cubensis* biology, genomics, virulence determinants, and gene expression during infection, as well as corresponding host responses.

FUTURE DIRECTIONS

While substantial progress has been made over the course of this dissertation project to establish genetic, genomic, and transcriptomic resources for the study of the *Ps. cubensis*-*C. sativus* interaction, it only represents a start to the study of this important and fascinating pathosystem. To build on the work herein, further functional characterization of additional candidate effector proteins, exploration of alternative splicing events over the course of infection, and a more detailed analysis of the *Ps. cubensis* transcriptome data set is needed to provide additional insight into both the lifestyle and virulence of *Ps. cubensis*. For *C. sativus*, additional analysis of the mRNA-Seq data would increase our knowledge of host defense responses in cucumber, and more transcriptome analyses of lines with resistance to *Ps. cubensis* could potentially lead to discoveries of new sources of resistance which could be utilized in breeding of resistant varieties for growers.

This dissertation and recent work [19] have identified 271 candidate effector proteins with variable RXLR motifs and established a pipeline for their characterization using both bioinformatics and molecular biology tools. The prediction of 20 possible amino acid substitutions at the R1 position has expanded our previous knowledge of the distribution of R1 domains in putative *Ps. cubensis* effector proteins, previously thought to be equally distributed between RXLR and QXLR motifs [19]. While expression support exists for 19 of these 20 substitutions, additional functional validation is necessary and would add to both general knowledge of oomycete and fungal effector

proteins and provide insight into the possible mechanisms utilized by *Ps. cubensis* to cause infection.

Functional characterization of the effector protein *Psc*RXLR1 demonstrated that it arises as a product of alternative splicing from *Psc*_781.4, making it the first example of an alternative splicing event in plant pathogenic oomycetes transforming a non-effector gene into a functional effector protein. This was substantiated using a combination of RT-PCR and transcriptome analysis of *Ps. cubensis* gene expression during infection. To build upon this work, analysis to quantify changes in expression levels of *Psc*RXLR1 and *Psc*_781.4 over the course of infection should be completed. Using qRT-PCR, an up-regulation of mRNA was identified at 1 dpi, extending through 4 dpi. However, due to limitations with this technique, we were unable to distinguish between *Psc*RXLR1 and *Psc*_781.4. More in depth analysis of alternative splicing events in *Ps. cubensis* during infection will allow us to distinguish between these isoforms, as well as enable us to identify additional changes in splicing over the course of infection. In addition, such analyses would enable the identification of global splicing events in *Ps. cubensis*, which would expand general knowledge about alternative splicing changes during infection, an as-yet understudied phenomenon.

The large-scale mRNA-Seq data sets generated in this dissertation provide a wealth of knowledge about gene expression during *Ps. cubensis* infection on *C. sativus*, and while extensive, the analyses and data presented herein only represent a small fraction of what could be gleaned from them. The analyses for this project, focused on genes

involved in pathogenicity and infection, such as effector proteins, pathogenicity-related genes, and CAZymes. *Ps. cubensis* is an obligate biotroph and is dependent on *C. sativus* for completion of its life cycle. As such, this mRNA-Seq dataset would be ideal for analyses focused on identifying genes involved in obligate biotrophy and the establishment of this intricate relationship between pathogen and host. This would not only yield insight into the nature of obligate biotrophic relationships which could be extrapolated to other systems, but could also be useful in understanding the pathogenicity of *Ps. cubensis*, potentially leading to improved control strategies. Monitoring the expression of both pathogen and host genes during infection can provide insight into the interplay between resistance and susceptibility. Initial analysis of the mRNA-Seq data focused on studying genes and expression patterns that were either *Ps. cubensis* or *C. sativus* specific. This data set is unique in that expression data from both host and pathogen were collected from the same samples and as such, ideal for correlative analyses that could look at the co-regulation of gene expression between pathogen and host. This could potentially aide in the identification of pathogen effectors that are down-regulating host defense responses or specific host genes that are up-regulated in response to pathogen virulence determinants.

Looking specifically at the *C. sativus* mRNA-Seq data in more detail could increase our knowledge of host defense responses in cucumber. We identified a suite of genes up-regulated at 1 dpi and subsequently down-regulated at 2 dpi (Figure 4.7) - functional characterization of a set of these genes could provide insight into the host defense response of *C. sativus*, as one could hypothesize that they could potentially play a role in defense if pathogen infection results in their down-regulation. Finally, the mRNA-Seq

data presented herein represents a compatible interaction between a virulent pathotype of *Ps. cubensis* and a susceptible *C. sativus* cv., 'Vlaspik'. While ideal for studying pathogen virulence determinants, to better study the host defense response, additional transcriptome analyses of *C. sativus* lines with resistance to *Ps. cubensis* should be done. This could potentially lead to discoveries of new sources of resistance and breeding of resistant varieties for growers.

The *Ps. cubensis*-*C. sativus* interaction, while in its infancy in terms of our understanding of the genetic and biochemical processes that regulate pathogenicity and resistance, presents an opportunity to develop a crop-based model system in which to study oomycete-plant interactions, ultimately focused on identifying mechanisms of host resistance that can be translated into field application. The work presented in this dissertation lays a substantial groundwork to achieve that goal by establishing genomic, genetic and transcriptomic resources for both *Ps. cubensis* and *C. sativus* that can be built upon to better understand the interplay between pathogen and host.

REFERENCES

REFERENCES

1. Savory EA, Granke LL, Quesada-Ocampo LM, Varbanova M, Hausbeck MK, et al. (2011) The cucurbit downy mildew pathogen *Pseudoperonospora cubensis*. Mol Plant Pathol 12: 217-226.

2. Haas BJ, Kamoun S, Zody MC, Jiang RH, Handsaker RE, et al. (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature 461: 393-398.

3. Tyler BM (2006) Phytophthora Genome Sequences Uncover Evolutionary Origins and Mechanisms of Pathogenesis. Science 313: 1261-1266.

4. Levesque CA, Brouwer H, Cano L, Hamilton JP, Holt C, et al. (2010) Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. Genome Biol 11: R73.

5. Baxter L, Tripathy S, Ishaque N, Boot N, Cabral A, et al. (2010) Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. Science 330: 1549-1551.

6. Birch PR, Armstrong M, Bos J, Boevink P, Gilroy EM, et al. (2009) Towards understanding the virulence functions of RXLR effectors of the oomycete plant pathogen Phytophthora infestans. J Exp Bot 60: 1133-1140.

7. Allen RL (2004) Host-Parasite Coevolutionary Conflict Between Arabidopsis and Downy Mildew. Science 306: 1957-1960.

8. Shan W, Cao M, Leung D, Tyler BM (2004) The Avr1b locus of *Phytophthora sojae* encodes an elicitor and a regulator required for avirulence on soybean plants carrying resistance gene Rps 1b. Mol Plant-Microbe Interact 17: 394–403.

9. Rehmany AP, Gordon A, Rose LE, Allen RL, Armstrong MR, et al. (2005) Differential recognition of highly divergent downy mildew avirulence gene alleles by RPP1 resistance genes from two Arabidopsis lines. Plant Cell 17: 1839-1850.

10. Armstrong MR, Whisson SC, Pritchard L, Bos JI, Venter E, et al. (2005) An ancestral oomycete locus contains late blight avirulence gene Avr3a, encoding a protein that is recognized in the host cytoplasm. Proc Natl Acad Sci U S A 102: 7766-7771.

11. Bos JI, Kanneganti TD, Young C, Cakir C, Huitema E, et al. (2006) The C-terminal half of *Phytophthora infestans* RXLR effector AVR3a is sufficient to trigger R3a-mediated hypersensitivity and suppress INF1-induced cell death in Nicotiana benthamiana. Plant J 48: 165-176.

12. Whisson SC, Boevink PC, Moleleki L, Avrova AO, Morales JG, et al. (2007) A translocation signal for delivery of oomycete effector proteins into host plant cells. Nature 450: 115-118.

13. Qutob D, Tedman-Jones J, Dong S, Kuflu K, Pham H, et al. (2009) Copy Number Variation and Transcriptional Polymorphisms of *Phytophthora sojae* RXLR Effector Genes Avr1a and Avr3a. PLoS ONE 4: e5066.

14. Dong S, Qutob D, Tedman-Jones J, Kuflu K, Wang Y, et al. (2009) The *Phytophthora sojae* Avirulence Locus Avr3c Encodes a Multi-Copy RXLR Effector with Sequence Polymorphisms among Pathogen Strains. PLoS ONE 4: e5556.

15. Bos JI, Armstrong MR, Gilroy EM, Boevink PC, Hein I, et al. (2010) *Phytophthora infestans* effector AVR3a is essential for virulence and manipulates plant immunity by stabilizing host E3 ligase CMPG1. Proc Natl Acad Sci U S A 107: 9909-9914.

16. Cabral A, Stassen JH, Seidl MF, Bautor J, Parker JE, et al. (2011) Identification of Hyaloperonospora arabidopsidis Transcript Sequences Expressed during Infection Reveals Isolate-Specific Effectors. PLoS ONE 6: e19328.

17. Schornack S, Huitema E, Cano LM, Bozkurt TO, Oliva R, et al. (2009) Ten things to know about oomycete effectors. Mol Plant Pathol 10: 795–803.

18. Kamoun S (2006) A catalogue of the effector secretome of plant pathogenic oomycetes. Ann Rev Phytopathol 44: 41-60.

19. Tian M, Win J, Savory E, Burkhardt A, Held M, et al. (2011) 454 Genome sequencing of *Pseudoperonospora cubensis* reveals effector proteins with a QXLR translocation motif. Mol Plant-Microbe Interact 24: 543-553.

20. Wang E, Sandberg R, Luo S, Khrebtukova I, Zhang L, et al. (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456: 470 - 476.

21. Boddu J, Cho S, Muehlbauer GJ (2007) Transcriptome analysis of trichothecene-induced gene expression in barley. Mol Plant-Microbe Interact 20: 1364-1375.

22. Huibers RP, de Jong M, Dekter RW, Van den Ackerveken G (2009) Disease-specific expression of host genes during downy mildew infection of Arabidopsis. Mol Plant-Microbe Interact 22: 1104-1115.

23. Gupta S, Chakraborti D, Sengupta A, Basu D, Das S (2010) Primary metabolism of chickpea is the initial target of wound inducing early sensed Fusarium oxysporum f. sp. ciceri race I. PLoS ONE 5: e9030.