Transcriptome and Small RNAome Dynamics during a Resistant and Susceptible Interaction between Cucumber and Downy Mildew

Alyssa Burkhardt and Brad Day*

Abstract
Cucumber (Cucumis sativus L.) downy mildew, caused by the obligate oomycete pathogen Pseudoperonospora cubensis (Berk. and Curt.) Rostov., is the primary factor limiting cucumber production. Although sources of resistance have been identified, such as plant introduction line PI 197088, the genes and processes involved in mediating resistance are still unknown. In the current study, we conducted a comprehensive transcriptome and small RNAome analysis of a resistant (PI 197088) and susceptible ('Vlaspik') cucumber during a time course of P. cubensis infection using Illumina sequencing. We identified significantly differentially expressed (DE) genes within and between resistant and susceptible cucumber leaves over a time course of infection. Weighted gene correlation network analyses (WGCNA) created coexpression modules containing genes with unique expression patterns between Vlaspik and PI 197088. Recurring data trends indicated that resistance to cucumber downy mildew is associated with earlier response to the pathogen, hormone signaling, and regulation of nutrient supply. Candidate resistance genes were identified from multiple transcriptome analyses and literature support. Additionally, parallel sequencing of small RNAs (sRNAs) from cucumber and P. cubensis during the infection time course was used to identify and quantify novel and existing microRNA (miRNA) in both species. Predicted miRNA targets of cucumber transcripts suggest a complex interconnectedness of gene expression regulation in this plant–pathogen system. This work bioinformatically uncovered gene expression patterns involved in the mediation of or response to P. cubensis resistance. Herein, we provide the foundation for future work to validate candidate resistance genes and miRNA-based regulation proposed in this study.

PSEUDOPERONOSPORA CUBENSIS, the causal agent of cucurbit downy mildew, is an obligate oomycete pathogen capable of infecting more than 20 genera within the Cucurbitaceae, including cucumber, melon (Cucumis melo L.), and watermelon [Citrullus lanatus (Thunb.) Matsum. & Nakai] (Savory et al., 2010). Within the first 24 h of the host–pathogen association, pathogen infection occurs, which includes sporangia germination, zoospore encystment, and the initiation of hyphae formation. In parallel, the onset of host symptoms include the development of angular chlorotic lesions on the upper leaf surface that expand and coalesce, with pathogen sporulation occurring on the lower surface. In the following 2 to 6 d, pathogen hyphal development continues, leading to the formation of specialized structures called haustoria, through which metabolites, protein, and nucleic acids are transferred between the pathogen and host, many of which likely function in enhancing pathogen virulence. In the final stage of its life cycle, P. cubensis forms sporangioles, which release sporangia into the air, leading to the establishment of new infection cycles.

A. Burkhardt and B. Day, Graduate Program in Cell and Molecular Biology, Michigan State Univ., East Lansing, MI 48824; B. Day, Graduate Program in Genetics and Dep. of Plant, Soil and Microbial Sciences, Michigan State Univ., East Lansing, MI 48824. Received 3 June 2015. Accepted 28 Sept. 2015. *Corresponding author (bday@msu.edu).

Abbreviations: cDNA, complementary DNA; DE, differentially expressed; dpi, days postinoculation; GO, gene ontology; ITS, internal transcribed space; lsiRNA, long small interfering RNA; miRNA, microRNA; mRNA, messenger RNA; nt, nucleotide; PCR, polymerase chain reaction; qPCR, quantitative real-time PCR; QTL, quantitative trait loci; RTA, Real-Time Analysis; SA, salicylic acid; SE, single-end; siRNA, small interfering RNA; sRNA, small RNA; WGCNA, weighted gene correlation network analysis.
The molecular-genetic basis of host resistance and pathogen virulence in downy mildew species has been best characterized through the study of Arabidopsis thaliana (L.) Heynh. and Hyaloperonospora arabidopsidis, an obligate pathosystem through which the genetic diversity of host and pathogen has revealed a complex interaction of several host resistance (R) genes and race-specific effectors from in H. arabidopsidis (Nemri et al., 2010; Krasileva et al., 2011; Coates and Beynon, 2010; Mohr et al., 2010). In addition, other studies have characterized the contribution of broad spectrum resistance responses, collectively demonstrating the genetic interactions between quantitative and qualitative trait loci within the host (Lapin et al., 2012; Gao et al., 2014).

In some cases, specific transcription factors have been shown to contribute to resistance, which is the case for transcription factors VvWRKY33 and VvWKRY1, which mediate resistance to grapevine (Vitis vinifera L.) downy mildew [Plasmopara viticola] (Berk. & M.A. Curtis) Berl. & De Toni] (Marchive et al., 2013; Merz et al., 2014).

Transcription-based resistance has also been identified in cucurbits, the best example of which is through resistance to P. cubensis in melon, which is conferred through the constitutive overexpression of glyoxylate aminotransferases (At1 and At2) (Taler et al., 2004). In addition to host genes that contribute to resistance signaling, numerous other genes have been associated with susceptibility, including those in Arabidopsis in response to H. arabidopsidis infection, such as 2-oxoglutarate (2OG)-Fe(II) oxygenase DMR6 (van Damme et al., 2008), the malecin-like receptor kinase IOS1 (Hok et al., 2011, 2014), and the negative regulator of plant defense response GSL5/ PMR4 (Wawrzynska et al., 2010). In total, these genes were initially identified based on transcriptional differences or were found to control transcriptional activities in the plant during infection. Collectively, these studies have demonstrated the importance of transcriptional regulation of defense and susceptibility during host–oomycete interactions.

In addition to the identification of host genes associated with resistance signaling during pathogen infection, a growing body of literature has demonstrated that sRNAs play essential roles in plant–pathogen interactions through the regulation of transcriptional, or posttranscriptional, processes. For example, host-derived sRNAs have been demonstrated to promote defense signaling in response to pathogen infection, or reciprocally, by the pathogen to promote infection and the downregulation of host immune responses (Qiao et al., 2013; Weber et al., 2013, 2014). Among the best characterized of these sRNAs are miRNAs, a class of single stranded, 20 to 22 nucleotide (nt) RNA molecules that posttranscriptionally regulate gene expression. As a group, miRNAs are hypothesized to play a key role as regulators of plant defense signaling including regulating processes associated with pathogen-associated-molecular-pattern-triggered and effector-triggered defense responses (Weber et al., 2014; Zhao et al., 2015). Specifically, recent work has shown that sRNAs from Botrytis cinerea Pers. can move from the pathogen to the host, where they function to silence resistance-associated genes (Weiberg et al., 2013). In cucumber, miRNAs have also been identified (Mao et al., 2012; Martinez et al., 2011; Hu et al., 2014; Li et al., 2014; Jin and Wu, 2015). However, none of these studies investigated the expression of sRNAs during a resistant response to biotic stress.

Historically, cucumbers have been bred for resistance to P. cubensis, yet in recent years, downy mildew epidemics in the United States suggest that the widely incorporated cucumber recessive resistance locus, dm-1, is no longer sufficient in providing durable resistance (Call et al., 2012a,b). To define alternate sources of resistance, screening of numerous cucumber cultigens has resulted in the identification of a plant introduction line, PI 197088, which displays robust resistance to a field isolate of P. cubensis (Call et al., 2012b). More recently, work by Caldwell et al. (2014) used quantitative trait loci (QTL) mapping to identify three loci located on chromosome 2, 4, and 5 as having a contribution to resistance to P. cubensis, with the QTL on chromosome 5 being the most robust (~24%) (Caldwell et al., 2014). However, a separate study by Yoshioka et al. (2014) identified a different constellation of QTL for P. cubensis resistance from PI 197088-derived cucumber lines. Taken together, the sum of these two these studies support the hypothesis that downy mildew resistance in cucumber is likely due to the combinatorial effects of several genes.

In the current study, we advance previous genome- and transcriptomics-centered research in the P. cubensis–cucumber pathosystem (Tian et al., 2011; Burkhardt et al., 2015; Adhikari et al., 2012; Savory et al., 2012a,b) through employing a comprehensive parallel transcriptome and small RNAome sequencing analysis to identify changes in gene and sRNA expression that are associated with a resistant or susceptible interaction. By analyzing temporal changes in gene expression over a time course of infection in susceptible (Vlaspik) and resistant (PI 197088) cucumber varieties, we identified more than 10,000 host genes that were differentially expressed. Using this approach, we compared the expression of these transcripts, including the pathogen-induced responses during resistance and susceptibility, to provide compelling evidence in support of the hypothesis that multiple genes—specifically at the earliest stages of the host–pathogen interaction—are responsible for the initiation of resistance to P. cubensis. Finally, we characterize the coexpression of Dicer-derived sRNAs from both host and pathogen and identify novel miRNA in both cucumber and P. cubensis that are predicted to regulate host gene expression during pathogen infection. Taken together, the work presented herein provides the first comprehensive parallel analysis of the transcriptome and small RNAome changes during the P. cubensis–cucumber interaction, shedding light on processes that control resistance and susceptibility during obligate parasitism.
Materials and Methods

Plant Growth and *Pseudoperonospora cubensis* Inoculation

Cucumber Vlaspik (Osborne seeds, sourced from Semi-nis hybrid Vlaspik M F1 88% Vlaspik, 12% Sire) and plant introduction line PI 197088 (seeds provided by Todd Wehner, NC State) were grown at 22°C in 16:8 h light/dark cycles in Redi-earth potting mix (Sun Gro Horticulture) in a BioChambers Bigfoot Series Model AC-60 growth chamber. *Pseudoperonospora cubensis* isolate MSU-1 was propagated on the susceptible cucumber AC-60 growth chamber.

Infection experiments were performed by pipetting multiple 10-μL droplets of a 1 × 10³ sporangia mL⁻¹ solution of *P. cubensis* onto the abaxial side of fully expanded 4-wk-old leaves. For mock-inoculated samples, water was pipetted onto the leaf surface. Inoculated plants were stored in the dark at nearly 100% humidity before being moved to the growth chamber (Adhikari et al., 2012). Inoculated leaf samples were collected using a #3 cork borer.

Disease Phenotype Analysis

For microscopic analyses of infected samples, samples were cleared in 100% ethanol and stained with trypan blue in a 1:1:1 solution of glyceral, lactic acid, and water. Visualization of destained tissues was performed using an Olympus IX71 inverted light microscope as previously described (Savory et al., 2012a).

RNA Isolation

RNA for real-time polymerase chain reaction (PCR) was extracted from sporangia and from flash-frozen leaf tissue using the RNeasy Plant Mini Kit (Qiagen). Contaminating genomic DNA was removed from the RNA sample by performing the on-column DNase treatment (Qiagen). Total RNA was extracted from sporangia and flash-frozen leaf tissue for messenger RNA (mRNA) and sRNA sequencing using the AllPrep DNA–RNA–miRNA Universal kit (Qiagen) and contaminating DNA was removed using on-column DNase treatment. RNA quality was assessed using the 2100 Bioanalyzer (Agilent) with the Agilent RNA 6000 Pico kit.

Pathogen Quantification Quantitative Real-Time Polymerase Chain Reaction

Complementary DNA (cDNA) was prepared from RNA using the USB first-strand cDNA synthesis kit with random hexamer primers (USB Affymetrix). Samples were prepared using the HotStart SYBR Green qPCR Master Mix (2x; USB). A Mastercycler ep Realplex real-time PCR (Eppendorf AG) was used to perform quantitative real-time PCR (qPCR). *Pseudoperonospora cubensis* internal transcribed space (ITS) primers were used to quantify the level of pathogen in the infected leaf material, as previously described (Tian et al., 2011). Values were normalized based on the relative expression of the pathogen ITS region as compared with cucumber-specific actin primers. For qPCR analysis, the following cycling parameters were used: 1 hold of 95°C for 2 min, 40 to 50 cycles of 95°C (15 s), 56°C (15 s), and 72°C (30 s). Relative expression was calculated where relative expression = 2^-ΔΔCt and where ΔCt = CtITS - Ctactin (Porter et al., 2012). Error bars represent the standard error from four biological replicates. Data were analyzed and processed using Prism (Ver. 6.0b; GraphPad Software, Inc., San Diego, CA). A two-way ANOVA was used to compare time and plant, and the Sidak’s multiple comparisons test was used to correct for multiple comparisons. Alpha ≤ 0.05 was used to determine significance.

Library Preparation and Next-Generation Sequencing

Libraries from two biological replicates each of Vlaspik and PI 197088 cucumber leaves that were mock inoculated or collected 1, 2, 3, 4, and 6 d postinoculation (dpi) with *P. cubensis* were prepared using the Illumina TruSeq Stranded mRNA Library Preparation Kit LT for mRNA or the Illumina TruSeq Library Preparation Kit for sRNA by the Michigan State Research Technology Support Facility. For sRNA samples, products were gel purified using a 6% polyacrylamide gel. Samples were pooled and run on the Illumina HiSeq 2500 Rapid Run flow cell (v1) using Rapid SBS reagents. Each mRNA sample was sequenced with at least 20 million 50-nt single-end (SE) reads and each sRNA sample was sequenced with at least 10 million 50-nt SE reads.

Processing and Alignment of Messenger RNA Sequencing Reads

Illumina Real-Time Analysis (RTA) software v1.17.21.3 was used to call bases. The output of RTA was demultiplexed and converted to FastQ files with Illumina Bcl2Fasta v1.8.4. Reads were deposited in the Sequence Read Archive as a BioProject at the National Center for Biotechnology Information under the accession number PRJNA285071 (to be released on publication). The reads were evaluated using FastQC (Babraham Bioinformatics), and the sequencing adapters were removed using Cutadapt v. 1.4.1. The first 14 bases were trimmed using fastx_trimmer from the FASTX v. 0.0.13 program. The processed reads were uniquely mapped to the annotated cucumber reference genome of Chinese long version 2 (ftp://www.icugi.org/pub/genome/cucumber/Chinese_long/v2/) using Bowtie v1.0.0 (Langmead et al., 2009) and TopHat v1.4 (Trapnell et al., 2009). The minimum and maximum intron sizes allowed were 50 and 50,000 bp, respectively (Adhikari et al., 2012). Accepted BAM hits from TopHat were converted to SAM files using SAMTools v0.1.18 (Li et al., 2009). The number of reads that aligned to each annotated gene of the cucumber genome were counted using HTSeq count v. 0.6.0 using the -s reverse and -t gene options, which is part of the HTSeq Python package (Anders et al., 2014). The percentage of uniquely mapped reads was calculated by dividing the total number of reads.
that mapped to genes using HTSeq count by the total number of reads that were obtained from sequencing. Normalized read counts were calculated by dividing the raw counts from HTSeq count by the library size calculated for each sample by DESeq2 (Love et al., 2014). The data were then log2 transformed and negative values were converted to zero. The correlation of biological replicates was determined and the values for the replicates were averaged for each time point for downstream WGCNA. The spliced version of each gene was used to report the annotation and gene ontology (GO) terms of each gene.

Analysis of Differential Gene Expression
The raw HTSeq count data were used as the input to measure differential gene expression using DESeq v. 1.18.0 implemented in R (R Development Core Team, 2010) (Seyednasrollah et al., 2013; Anders et al., 2013). DESeq analyses were performed using the standard procedures for pairwise comparisons with the $p$-adjusted value < 0.05 as the cutoff for determining significance. DESeq was used to find significantly differentially expressed genes between Vlaspik and PI 197088 at each time point and between all time points, including the mock. DESeq2 v. 1.6.3 was used to identify genes that were differentially expressed across the time course (Love et al., 2014). Count data from HTSeq was used as input, and the data set design included the variables of plant line, time, and the interaction between the plant line and time. The likelihood ratio test with a reduced design allowed for the contributions of each variable to be measured by comparing the full model with the reduced model as a method to explain the variation in the calculated values of gene expression; the $p$-values reported the likelihood that the source of variation within the model could be attributed the removed variable. In this way, the significance of the contributions of the plant line, time, and the interaction of the plant line and the time to the pattern and changes in gene expression observed in all plant lines at all time points was determined.

Weighted Gene Correlation Network Analysis
Genes with correlated expression patterns were identified using the WGCNA implemented in R (R Development Core Team, 2010) (Langfelder and Horvath, 2008). Genes included in this analysis were such that a gene was required to have a DESeq2 $p$-adjusted value of less than $1 \times 10^{-5}$ to be included in the analysis. Such a stringent cutoff was used to reduce the number of genes classified as significant. The input expression values for each gene were the processed HTSeq counts as described above. Modules were generated from these 12,532 genes using data for Vlaspik, PI 197088, or both. When genes did not exhibit sufficient variance or had zero value, they were excluded from the modules. All WGCNA parameters were used in their default settings with the following modifications: the soft threshold ($\beta$) parameter of 16 was used to create the Vlaspik modules, and a $\beta$ of 18 was used to create the PI 197088 and Vlaspik–PI 197088 combined modules. For all modules, the tree-cut parameter was 0.9, and the resulting modules were merged using a distance threshold of 0.2. Modules were visualized by plotting the Z-scores for all genes within a given module (Childs et al., 2011).

Gene Ontology Enrichment Analysis
Gene Ontology terms for each predicted cucumber gene were obtained by running the predicted cucumber protein sequences through InterPro Scan (Zdobnov and Apweiler, 2001). A custom script was written to parse the file to create a list of all of the unique GO terms for each protein translated from the spliced version of each gene. The GO terms for each gene in an enrichment list (i.e., a list of genes within a specific module) were selected for GO term enrichment analysis, which was performed using the singular enrichment analysis online tool on the AgriGo website (http://bioinfo.cau.edu.cn/agriGO/index.php). This analysis was run using the standard parameters, which were the Fisher statistical method with the Yekutieli multi-test adjustment method at a significance level of 0.05.

Processing and Alignment of Small RNA Reads
Illumina RTA software (v1.17.21.3) was used for base calling. The output of RTA was demultiplexed and converted to FastQ files with Illumina Bcl2Fast a 1.8.4. Reads were evaluated using FastQC, and sequencing adapters were removed using Cutadapt (v1.4.1). Sickle (v. 1.33) was used to trim reads with a quality score below 30 and reads longer than 15 nt were used for further analysis. The high-quality, trimmed reads were aligned to both the P. cubensis genome and the cucumber genome using Bowtie v. 1.0.0 using the options -v 1, -best, -strata, -k 50 as was used to identify sRNA in A. thaliana (Axtell, 2013b). Reading mappings to only one of the genomes were used for sRNA analyses in either P. cubensis or cucumber. Small RNAs were identified and mature miRNAs were quantified using the default settings of the plant-optimized sRNA bioinformatics program ShortStack v. 2.1.0, which used other sRNA programs including RNAfold and RNAplot to identify hairpins and miRNA (Axtell, 2013b; Shahid and Axtell, 2013). The number of reads mapping to each identified sRNA cluster were counted with HTSeq using the alignment files generated by ShortStack. The miRNA read counts were divided by the DESeq2 library size, and the resulting data was log2 transformed and negative values were converted to zero. Biological replicates were combined.

MicroRNA Family Identification and Target Prediction
The families of each predicted miRNA locus were identified from the 2 to 3 most abundant mature miRNA sequences predicted by ShortStack. These mature miRNA were then compared with the miRNAs present in miRBase through a BLASTn search using an E-value cutoff of 1.0 to be considered as part of the known miRNA family. The miRBase database was also used to identify conserved miRNA in all species present in the database. Cucumber targets of the mature miRNA sequences were
identified using psRNATarget with a maximum expectation value of 3.0 and with default program setting (Dai and Zhao, 2011). TAPIR, the target prediction for plant miRNAs (Bonnet et al., 2010), was also used to predict targets from among the C. sativus v. 2 transcripts using the score cutoff of 4 and the free energy ratio cutoff of 0.7. A gene was considered to be a target of a predicted sRNA if it was identified using both prediction methods.

DESeq2 for MicroRNA
Differentially expressed miRNA were identified using DESeq2 (Love et al., 2014) following the procedures described for miRNA. The cutoff for determining significance was a $p$-adjusted value of less than 0.05. The expression patterns of significantly expressed miRNA were correlated with the expression patterns of their predicted targets by calculating and plotting the $Z$-scores for the normalized count values.

Quantitative Real-Time Polymerase Chain Reaction Validation of RNA Sequencing Data
RNA was extracted with the Qiagen All-Prep kit and was synthesized into cDNA as previously described. Samples were prepared for qPCR as previously described but using the HotStart SYBR Green qPCR Master Mix with ROX. A 7500 Fast real-time PCR machine (Applied Biosystems) was used to perform qPCR using the following parameters: 1 hold of 95°C for 10 min and 40 cycles of 95°C (15 s) and 72°C (10 s). A melt curve analysis was run to ensure that only one amplicon was generated per reaction. Primers used to amplify the genes of interest are in Supplemental Table S1. Relative expression was calculated, whereby relative expression = $2^{\Delta \Delta Ct}$ and where $\Delta Ct = Ct_{\text{gene of interest}} - Ct_{\text{actin}}$ (Porter et al., 2012). Error bars represent the standard error from three biological replicates. A Grubbs’ test was performed to remove outliers ($\alpha < 0.05$).

Results and Discussion

Pseudoperonospora cubensis Growth and Development is Abrogated in the Resistant Cucumber Accession PI 197088
A time course of P. cubensis infection of resistant (PI 197088) and susceptible (Vlaspik) cucumber varieties was conducted to investigate changes in phenotypes and gene expression between the two plant lines during the infection cycle. As shown in Fig. 1, pathogen infection of PI 197088 and Vlaspik showed increasingly divergent phenotypes, as demonstrated through microscopy and leaf lesions as well as supported by qPCR to measure pathogen proliferation. In the susceptible cultivar Vlaspik, the area of the leaf inoculated with P. cubensis showed a water-soaking phenotype and developed chlorotic lesions (Fig. 1A). In contrast, the resistant variety PI 197088 showed smaller, less chlorotic lesions. Microscopic examination of infected lesions supported the observed whole-leaf phenotypes, revealing that sporangiophore development occurred as early as 6 dpi in Vlaspik but was not observed in infected PI 197088 (Fig. 1B). The lack of sporulation on PI 197088 does not appear to be the result of a block in pathogen entry as indicated by encystment and hyphal growth (Fig. 1B; arrows at 1 and 4 dpi). However, it is unclear from microscopy observations if the pathogen is living. Furthermore, we did not observe pathogen elicitation of the hypersensitive response, which is consistent with breeding data (Caldwell et al., 2014; Yoshioka et al., 2014).

To quantitatively assess pathogen growth as a means to correlate pathogen load with precise stages of host infection, qPCR was used (Fig. 1C). We observed a steady increase in the abundance of P. cubensis in Vlaspik over time with an ~12-fold increase at 6 dpi, as compared with samples analyzed from 1 dpi. In contrast, the level of P. cubensis in PI 197088 remained relatively low, and the level of pathogen growth never exceeded threefold of that observed at 1 dpi (Fig. 1C). The greatest contrast in pathogen growth on either Vlaspik or PI 197088 was observed at 6 dpi. A two-way ANOVA indicated that the differences attributed to time, plant line, and the interaction were all statistically significant.

As noted above, a key feature of resistance observed in PI 197088 is that the pathogen was not able to sporulate; however, phenotypic and qPCR data indicate the early stages of infection are occurring in the resistant line. We posit that once the initial entry of the pathogen occurs, the pathogen is capable of growing, if only for a brief period of time, in the resistant plant line. Given that P. cubensis is an obligate biotroph and requires host resources for all of its nutrition, we hypothesize that one possible contributing mechanism that functions in the abrogation of pathogen growth is host restriction of P. cubensis nutrient acquisition. Alternatively, or in addition, a second hypothesis is that changes in plant metabolism, such as changes in hormones or nutrients, may play a role in resistance.

Transcriptome Profiling of the Pseudoperonospora cubensis–Cucumber Interaction
RNA sequencing was used to undertake comprehensive transcriptome profiling for five infection time points—1, 2, 3, 4, and 6 dpi—and mock-inoculated leaf samples from both Vlaspik and PI 197088. Two independent and parallel time course inoculations were conducted on both Vlaspik and PI 197088 to yield two biological replicates. On average, following trimming and removing lower quality reads, 74% of the sequenced reads from each sample uniquely mapped to the cucumber genome (Supplemental Table S2). The correlation of biological replicates for the normalized read counts of all the genes between each sample was calculated to have an $R^2$ value of at least 0.95 (Supplemental Fig. S1). The normalized read counts for each gene at each time point can be found in the supplemental files in the author-recommended online resources. In total, 19,581 genes were expressed during at least one time point in at least one cultigen.
Fig. 1. Phenotypic and quantitative analysis of *Pseudoperonospora cubensis* infection of ‘Vlaspik’ (susceptible) and PI 197088 (resistant) cucumber. (A) Leaf phenotypes of the lower (abaxial) and upper (adaxial) leaf surface inoculated with 10 μL of $1 \times 10^{-5}$ *P. cubensis* sporangia mL$^{-1}$ during a time course. (B) Light microscopy of inoculated Vlaspik and PI 197088 leaf punches were cleared and stained with trypan blue. Scale bars = 50 μm for Vlaspik at 2 to 4 d postinoculation (dpi), 100 μm for Vlaspik at 1 dpi, and PI 197088 at 1 to 4 dpi, 200 μm for PI 197088 at 6 dpi, and 500 μm for Vlaspik at 6 dpi. Arrows indicate pathogen structures. (C) Pathogen quantification using quantitative real time polymerase chain reaction analysis. DNA primers specific for *P. cubensis* internal transcribed space normalized to *Cucumis sativus* actin. Statistical analysis was performed using a two-way ANOVA using the Sidak’s multiple comparisons test. The differences attributed to time, plant line, and the interaction were all statistically significant ($p < 0.05$). Error bars represent the standard error of four biological replicates.
Differentially Expressed Genes in PI 197088 and Vlaspik Reveal Early Time Points are Important in *Pseudoperonospora cubensis* Infection Response

To identify differentially expressed host genes during a resistant and susceptible interaction, we next used DESeq analyses to perform a pairwise comparison between Vlaspik and PI 197088. Using this approach, we found that the greatest number of differentially expressed genes existed between the mock-inoculated plant and each inoculation time point (Table 1). The number of DE genes compared with mock was lowest at 1 dpi for both plants, yet compared with Vlaspik, PI 197088 still had several thousand more DE genes than mock-inoculated samples harvested at 1 dpi. For 2 to 6 dpi, the number of DE genes was similar to mock, with the greatest number of DE genes occurring at 4 dpi in Vlaspik and 2 dpi in PI 197088. Together, these data suggest that PI 197088 responds earlier to the pathogen than Vlaspik. Furthermore, the number of genes that were differentially expressed when comparing the later time points in either cultigen was very low, with the number of DE genes in the hundreds when comparing 3 or 4 dpi to 6 dpi and below 100 when comparing the nearest time points. This could suggest that more gradual changes in gene expression exist once the pathogen has established itself within the host (in the case of Vlaspik) or has been resisted by the plant (in the case of PI 197088).

In addition, trends were observed in the number of significantly DE genes that were up- or downregulated in either Vlaspik or PI 197088 at some time points. For example in Vlaspik, roughly 18% of the genes were downregulated at 1 dpi compared with the mock, while in PI 197088 over 58% of the genes were downregulated at 1 dpi compared with the mock (Table 1). Similarly, in Vlaspik only about 5% of the genes were downregulated at 2 dpi compared with 1 dpi, while in PI 197088, the same comparison resulted in about 39% of the genes being downregulated (Table 1). Among the downregulated genes of PI 197088 at 1 dpi, several significant GO terms were associated with photosynthesis processes, transcription factor activity, or signaling cascades. The GO term of response to hormone stimuli was attributed to several auxin response factor genes; this is significant because auxin has been shown to promote pathogen growth and is often negatively regulated in initial responses to pathogens (Huot et al., 2014).

### Metabolism and Hormone Signaling Might Play a Role in Resistance Signaling

Pairwise comparisons between each time point and the mock samples were further analyzed by comparing lists of genes, which were commonly or oppositely up- or downregulated at each time point (Supplemental Fig. S2). Among the expressed genes identified as uniquely upregulated in PI 197088, common GO terms included photosynthesis and transcription factor activity at 2 to 6 dpi. In contrast, glycolysis related GO terms were common among the gene groups only upregulated in Vlaspik during infection. This might indicate that while PI 197088 is able to continue with photosynthesis during infection, Vlaspik must break down more of its carbohydrates. Genes associated with exocytosis were also enriched in upregulated genes in only Vlaspik at 4 and 6 dpi. Among the genes that were uniquely downregulated in PI 197088, were intracellular transport associated GO terms including those related to the Golgi and endoplasmic reticulum (ER). In addition, Zn ion binding genes and genes associated with translation and signal transduction, including protein phosphatases, were also included only in PI 197088 downregulated genes. Beyond the uniquely induced or downregulated genes from each plant line, the genes that were oppositely regulated between Vlaspik and PI 197088 were of particular interest, and generally increased in number across the time course (Supplemental Fig. S2). For example, two separate ethylene-responsive transcription factors, Csa3G017320 and Csa2G191300, were downregulated in PI 197088 between 2 and 6 dpi and upregulated in Vlaspik; this pattern in differential regulation supports the role of hormone signaling in defense regulation (Supplemental Fig. S3).

Next, DESeq was used to compare genes expressed at each time point between plant lines. From this analysis, we identified 655 DE genes between the Vlaspik and PI 197088 mock-inoculated plants. Among the genes identified, the only enriched GO terms were those related to photosynthesis. As a result, additional analyses, below, included the mock in DESeq2 analysis without using it as a baseline for comparison. Still, in the pairwise comparisons, it was evident that an important change in gene transcription patterns was occurring between 1 and 2 dpi, as the number of DE genes between Vlaspik and PI 197088 at 1 dpi was relatively low (~472), while the number of DE genes at 2 dpi was considerably high.

### Table 1. Differentially expressed genes as determined by DESeq v. 1.18.0 from cucumber ‘Vlaspik’ (susceptible) and PI 197088 (resistant) inoculated with *Pseudoperonospora cubensis*. The numbers in parentheses indicate the percentage of significantly (\(p_{adj} < 0.05\); Benjamini and Hochberg, 1995) downregulated genes.

<table>
<thead>
<tr>
<th>Time (dpi)</th>
<th>Vlaspik</th>
<th>PI 197088</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 dpi</td>
<td>1969 (18)</td>
<td>4864 (58)</td>
</tr>
<tr>
<td>2 dpi</td>
<td>3699 (5)</td>
<td>5990 (39)</td>
</tr>
<tr>
<td>3 dpi</td>
<td>23 (30)</td>
<td>51 (55)</td>
</tr>
<tr>
<td>4 dpi</td>
<td>28 (79)</td>
<td>2 (0)</td>
</tr>
<tr>
<td>6 dpi</td>
<td>5733 (40)</td>
<td>7059 (45)</td>
</tr>
</tbody>
</table>

† dpi, days postinoculation.
Interestingly, a GO term that was significantly enriched at 1 dpi among the downregulated genes in PI 197088 was related to secondary active transmembrane transporter activity and included genes Csa2G175170 and Csa3G135080, which encode phosphate transporters, and Csa3G094510 and Csa6G409380, which encode general solute carriers (Supplemental Fig. S3). From this, we posit that if any of these transporters are functioning in the delivery of nutrients to *P. cubensis*, it supports the hypothesis that a host-driven restriction in nutritional compounds could be a significant contributor to host resistance (Stassen and Van den Ackerveken, 2011). As noted above, in contrast to the low number of DE genes at 1 dpi, there were 3538 genes differentially expressed between Vlaspik and PI 197088 at 2 dpi, with 1624 being downregulated in PI 197088 (66 enriched GO terms) and 1914 being upregulated (45 enriched GO terms). Among the GO terms that were significantly enriched among the downregulated genes at 2 dpi were GO terms associated with molecular processes such as protein phosphorylation and ubiquitination processes. Both of these processes have been extensively characterized as key processes associated with immune and defense signaling pathways in plants (Antolín-Llovera et al., 2014; Coca and San Segundo 2010; Furniss and Spoel 2015; Belkhadir et al., 2014).

Compared with the early time points, the number of DE genes between Vlaspik and PI 197088 was highest at 4 and 6 dpi, where the number of DE genes was 5297 and 6867, respectively. This pattern of expression suggests a strong divergence in the biology of the two genotypes—perhaps because one had successfully resisted infection and the other was succumbing to the pathogen. A unique GO term in the downregulated genes at 4 dpi included the molecular process of extracellular-glutamate gated ion channel activity, for which several genes contributing to this GO term belong to the plant defense-related clade 3 of this protein family (Forde and Roberts 2014). Common GO terms associated with cellular processes in downregulated genes in PI 197088 compared with Vlaspik at 4 and 6 dpi included those associated with the proteasome complex, the endoplasmic reticulum, and the exocyst. Some of these genes could be involved in the formation of the extrahaustorial membrane surrounding haustoria and therefore might be allowing less nutrient flow to the pathogen when they are downregulated (Lu et al., 2012). Within the groups for genes upregulated at 4 and 6 dpi in PI 197088, is the heat-shock protein GO term, which has prior evidence for being involved in downy mildew resistance in cucumber (Li et al., 2012).

Significantly Differentially Expressed Genes Group into Distinct Modules of Coexpression

To reduce the number of biologically relevant genes for WGCNA, DESeq2 was used to filter the number of genes that were significantly differentially expressed not only in pairwise comparisons but also in multiple comparisons across both time and between plant lines. Using this method, we found that the gene with the most significance for the interaction between plant line and time of infection was Csa1G051760, which encodes an inducer of CBF expression and was highly induced only in PI 197088 (Supplemental Fig. S3). As a function of plant defense signaling, previous work showed that a homolog of this gene was induced by jasmonic acid (Hu et al., 2013) and was also regulated by ABA, a key hormone associated with downy mildew defense signaling in *Arabidopsis* (Knight 2004; Hok et al., 2011, 2014). The second most significantly differentially regulated gene for the interaction factor of plant and time was Csa4G641000, an N-hydroxycinnamoyl/benzyoltransferase, a homolog that catalyzes a committed step in phytoalexin synthesis (Yang et al., 1997), which was induced and highly expressed only in PI 197088 (Supplemental Fig. S3). This is of particular significance, as phytoalexin is an important secondary metabolite in plant defense signaling (Hammerschmidt, 1999; Zhao et al., 2005).

Following DESeq2 filtering, ~12,000 genes were further segregated based on patterns of expression. Using this approach, when either Vlaspik or PI 197088 expression values were analyzed, 12 separate modules were constructed, while nine modules were constructed when Vlaspik and PI 197088 expression values were analyzed together. For example, as shown in Fig. 2 and 3, modules were found to illustrate distinct patterns of gene expression, whereby the plotted values (i.e., Z-scores) describe how many standard deviations each data point is from the mean of the normalized expression values for each gene across all time points. Genes with the same or exact opposite expression patterns were grouped in the same module. For example, in PI 197088, Module K contained the fewest genes (i.e., 34), while Module G contained the most genes (4236). The average module size was ~1000 genes (Fig. 2A; Supplemental Fig. S4). For the grouping of genes from Vlaspik, Module H contained the fewest genes (i.e., 48), while Module E contained the most genes (5156) with the average module size of ~1000 genes (Fig. 2B; Supplemental Fig. S4). Overall, modules constructed for Vlaspik contained genes that were regulated in the same directional pattern, with few oppositely regulated genes, while in PI 197088, several of the modules included genes with opposite regulation patterns. PI 197088 also had more genes in modules that changed in expression pattern at an early time point (i.e., 1 dpi) compared with Vlaspik, whose modules showed delayed transcriptional changes. For the combined Vlaspik and PI 197088 data, the average module size was approximately 1200 genes, with the fewest genes (84) in Module B and the most genes in Module D (2853) (Fig. 3; Supplemental Fig. S4).

Coexpression Modules in Vlaspik and PI 197088 Reveal Differing Patterns of Gene Regulation among Significantly Differentially Expressed Genes

In PI 197088, several modules consisted of genes with divergent patterns of expression at the mock time point and at early time points during infection (Fig. 2A; Supplemental Fig. S4). In Module E, enriched GO terms in the
genes that were down in the mock time point included DNA replication, phosphorylation, ubiquitination, and peroxidase activity, some of which could be involved in a signaling response to the pathogen (Fig. 2A). In Module C, enriched GO terms among genes there were highly expressed at 1 dpi included translation and glycolysis (Fig. 2A). The genes in Module C that were down at 1 dpi were enriched in terms including several transporters like cation transporters and ion transmembrane transporters. In Module G, genes that were upregulated at 1 dpi have enriched GO terms that included intracellular transport of proteins, Golgi vesicle transport, and phospholipid transport (Fig. 2A). Finally, genes that were down at 1 dpi in Module G included GO terms enriched for photosynthesis and glycogen biosynthesis; these genes were later expressed at average or slightly above average levels, suggesting that the plant’s photosynthetic processes are not disrupted by the pathogen at later time points. Furthermore, two genes of interest in Module G involved in thiamine synthesis, Csa3G078800 (thiamine thiazole synthase) and Csa1G528590 (ThiC), are of particular interest as thiamine has been implicated in initiating a resistance response against fungi in rice (*Oryza sativa* L.) and cucumber (Ahn et al., 2005) and against grapevine downy mildew (Boubakri et al., 2012). Mechanistically, thiamine has been shown to induce resistance through priming via a systemic acquired resistance pathway in *Arabidopsis* (Ahn et al., 2007) and is able to manipulate the phenylpropanoid pathway in grapevine in resistance to downy mildew (Boubakri et al., 2013).

In Vlaspik, several modules consisted of genes that peaked in expression around 2 or 3 dpi and were then downregulated, or at the baseline level of expression, before or after that time point (Fig. 2B; Supplemental Fig. S4). For example, Module G showed patterns of a peak expression at 1 dpi and a tapering off of transcript level throughout the rest of the time course (Fig. 2B). Within Module G, multiple genes encoding transcription factors, including WRKYs and Zn finger proteins were classified. Similarly, in Module D, expression patterns peaked at 2 dpi. Among unique significant GO terms in this module were genes annotated to encode heat-shock proteins, which are a group of genes that have previously be associated with resistance (Li et al., 2012). Finally, Module K
displayed a relative baseline level of transcription, with a decrease at approximately 4 dpi and an increase at 6 dpi (Fig. 2B). In this module, were several kinases, some of which are defense-associated leucine-rich receptor kinases (Antolín-Llovera et al., 2014; Belkhadir et al., 2014).

**Coexpression of Significant Genes across Genotypes Shows Distinct Patterns of Regulation during *Pseudoperonospora cubensis* Infection**

The modules that combined the expression values of Vlaspik and PI 197088 were especially informative, as genes with opposite patterns of expression were clearly identified. For example, as shown in Fig. 3, Modules A and D revealed opposite expression trends, in that in Module A, genes from PI 197088 were upregulated, while genes from Vlaspik were downregulated. The observed trend in Module A was enriched in GO terms including photosynthesis, RNA methyltransferase activity, and amino acid biosynthesis. In contrast, the main trend Module D was enriched in several GO terms including glycolysis, protein catabolism and ubiquitination, serine–threonine kinase, and the endomembrane system. These apparent opposing expression trends support the hypothesis that photosynthesis in PI 197088 is relatively unperturbed, while Vlaspik must metabolize energy stores in response to pathogen growth. Among the genes in Module A were the aforementioned inducers of CBF expression and the N-hydroxycinnamoyl/benzoyltransferase. Additional genes of interest classified within Module A included isochorismate synthase (Csa1G008580) and thioredoxin genes (Supplemental Fig. S3) (Wildermuth et al., 2001). Similarly, Module D also contained several genes of interest, including a suite of transcription factors as well as a group of genes encoding proteins of unknown function that displayed strong patterns of differential expression between PI 197088 and Vlaspik. Significant genes in this module with lesser-known
functions in resistance were a WAT1-related drug and metabolite transporter protein (Csa2G151580), a cupredoxin (Csa7G432520), and a lipase (Csa6G490910) that were expressed lowly in PI 197088 but was continually upregulated in Vlaspik (Supplemental Fig. S3). Additionally, Module B was interesting because it revealed gene expression patterns with extreme opposing patterns of expression (Fig. 3). For example, and of particular note, were two genes that encode protein kinases, Csa6G176420 and Csa6G176430, as well as a gene encoding a receptor-like protein kinase (Csa5G495960). In this module, each of these genes were identified as being highly expressed in PI 197088 and lowly or not expressed in Vlaspik. In contrast, two glutamate receptors, Csa2G363540 and Csa2G363530, were expressed only in Vlaspik; this observation was surprising, as glutamate receptors have been proposed to have a role in pathogen recognition and subsequent resistance signaling (Forde and Roberts, 2014). Module F revealed a difference in the timing of response to pathogen infection, as the genes in this module from PI 197088 showed a declining trend of expression at an earlier time point than the same genes in Vlaspik (Fig. 3). For the observed trend of Module F, several GO terms were enriched including intracellular protein transport, glycolipid metabolism, transcription, and protein phosphatase 2C, which is relevant because it serves as a negative regulator of ABA responses, and a resistance-associated hypersensitive response to ABA via the removal of a negative ABA regulator has been shown to confer resistance to downy mildew (Hok et al., 2014).

### Candidate-Based Approach to Identify Resistance-Associated Genes

In addition to taking a nondirected approach to identify resistance-associated genes, we also employed a literature-informed approach to identify candidates. From the two QTL studies mapping downy mildew resistance, an overlapping QTL, dm5.3, was identified in both studies and was associated with a relatively high level to resistance (Yoshioka et al., 2014; Caldwell et al., 2014). As a result, DE genes located within dm5.3 were evaluated as potential candidates for resistance. The two most differentially expressed genes in this region were two genes of unknown function that were highly expressed in the resistant line. Other common annotations significantly differentially expressed genes in this region included proteases and methyltransferases. Of particular interest within this group, and among genes with previous annotations previously associated with resistance, were several transcription factors including Csa5G604920 that encodes a homeobox leucine zipper protein that is homologous to ATHB13, a transcription factor in Arabidopsis that confers broad-spectrum resistance and is upregulated at some time points in PI 197088 (Supplemental Fig. S3)(Gao et al., 2014).

A candidate-based approach was also taken through the identification of cucumber homologs of genes previously demonstrated to function in downy mildew resistance signaling in other plant systems (Table 2). Interestingly, and in support of the work presented herein, many of these directed candidates were also identified through our bioinformatics analyses, above. For example, the homologs of two aminotransferase genes previously shown to confer resistance to P. cubensis

### Table 2. Candidate cucumber resistance-associated genes.

<table>
<thead>
<tr>
<th>Cucumber gene</th>
<th>Gene ID</th>
<th>Plant gene</th>
<th>Function and resistance association</th>
<th>Cucumber expression pattern</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Csa5G604920</td>
<td>AT1G69780</td>
<td>ATHB13</td>
<td>HD-zip transcription factor conferring broad-spectrum resistance</td>
<td>Significantly induced in PI 197088 compared with mock at 2 to 6 dpi, Significantly higher in PI 197088 compared with ‘Vlaspik’, Module 6 for PI 197088 and Vlaspik located on dm5.3</td>
<td>Gao et al., 2014</td>
</tr>
<tr>
<td>Csa6G128000</td>
<td>AT4G03550</td>
<td>PMR4/GSL5</td>
<td>Negative regulator of plant defense</td>
<td>Significantly induced in Vlaspik at 4 to 6 dpi compared with mock, significantly higher in Vlaspik compared with PI 197088 at 4 to 6 dpi</td>
<td>Wawrzynska et al., 2010</td>
</tr>
<tr>
<td>Csa6G433300</td>
<td>AAL62332.1</td>
<td>AT2</td>
<td>Aminotransferase 2, overexpression in resistant melon</td>
<td>Significantly induced in Vlaspik and PI 197088 compared with mock at 2 to 6 dpi, Significantly higher in PI 197088 compared with Vlaspik at 2 to 6 dpi, Module 6 for PI 197088 and Vlaspik</td>
<td>Taler et al., 2004</td>
</tr>
<tr>
<td>Csa6G176370</td>
<td>AAL47679.1</td>
<td>AT1</td>
<td>Aminotransferase 1, overexpression in resistant melon</td>
<td>Significantly higher in PI 197088 compared with Vlaspik at 2 to 6 dpi, Module A for PI 197088 and Vlaspik</td>
<td>Taler et al., 2004</td>
</tr>
<tr>
<td>Csa2G028490</td>
<td>Csa2G028490</td>
<td>LOX</td>
<td>Lipoygenase</td>
<td>Significantly up in PI 197088 at 6 dpi compared with mock, Significantly up in PI 197088 compared with Vlaspik at 2, 4, and 6 dpi, Module A in PI 197088 and Vlaspik</td>
<td>Huang et al., 2009</td>
</tr>
<tr>
<td>Csa3G078800</td>
<td>OsDR8</td>
<td>TH11</td>
<td>Thiamine thiazole synthase</td>
<td>Significantly up in PI 197088 from 2 to 6 dpi compared with Vlaspik, PI 197088 Module 6</td>
<td>Wang et al., 2006</td>
</tr>
<tr>
<td>Csa1G008580</td>
<td>AT1G74710</td>
<td>ICS1</td>
<td>Isoschisomate synthase (salicylic acid synthesis)</td>
<td>Significantly up in PI 197088 from 2 to 6 dpi compared with Vlaspik, Module A for PI 197088 and Vlaspik</td>
<td>Wildermuth et al., 2001</td>
</tr>
<tr>
<td>Csa4G641000</td>
<td>AT2G19070</td>
<td>HCBT</td>
<td>Involved in phytoalexin synthesis</td>
<td>Significantly up in PI 197088 from 2 to 6 dpi compared with Vlaspik, Module A for PI 197088 and Vlaspik</td>
<td>Yang et al., 1997</td>
</tr>
</tbody>
</table>

1 dpi, days postinoculation.
in melon (Taler et al., 2004) were identified as having significantly higher levels of expression in the resistant cucumber line (Supplemental Fig. S3). Additionally, a lipoxygenase gene, Csa2G028490, had significantly higher expression in PI 197088 than Vlaspik, and previous work suggests that lipoxygenases may play a role in disease resistance in cucumber (Supplemental Fig. S3) (Huang et al., 2009). Finally, the thiamine thiazole synthase gene, Csa3G078800, which showed over a 25-fold increase in PI 197088 compared with Vlaspik at 6 dpi, represents a potentially interesting candidate, as its rice homolog, OsDR8, is an upstream regulator of the rice resistance pathway to both bacterial and fungal pathogens. Briefly, it was shown that reduced expression of this gene resulted in increased susceptibility and that resistance could be rescued through the application of exogenous thiamine to plants (Wang et al., 2006). Furthermore, increased thiamine levels have been shown to influence salicylic acid (SA)-induced defense response and upregulate the phenylpropanoid pathway (Boubakri et al., 2013). This finding is consistent with our data (Supplemental Fig. S3) showing that isochorismate synthase, which is involved in phytoalexin synthesis, is both upregulated in the resistant line, PI 197088 (Fig. 3).

**Identification of Small RNAs during the Cucumber–Pseudoperonospora cubensis Interaction**

In parallel to the mRNA analyses described above, sRNAs were sequenced (from the same samples from which the mRNA was collected) to establish a foundation for defining the role of sRNAs in the regulation of the cucumber transcriptome during *P. cubensis* infection. The reads from both cucumber and *P. cubensis* were first mapped to the *P. cubensis* and cucumber genomes to remove any reads that map to both species, as previously described (Weiberg et al., 2013). The percentage of reads that uniquely mapped to either *P. cubensis* or cucumber are shown in Supplemental Table S3. The sample derived from *P. cubensis* sporangia consists of 73% of reads mapping to *P. cubensis*. The remaining 27% of reads mapping to cucumber is due to the contamination of the sporangia with leaf tissue, which is an unavoidable consequence of the collection process. Reads mapping to *P. cubensis* in the mock-inoculated cucumber sample (~2–3%) aligned with chloroplast or plant-derived sequences, as well as some bacterial species, suggesting a low level of plant or bacterial sequence contamination in the previously published *P. cubensis* genome assembly, respectively. In the resistant interaction, the percentage of reads derived from *P. cubensis* is 24% at 1 dpi but decreases to less than 10% at all other time points, which is in agreement with our microscopy and qPCR analyses, indicating that the amount of pathogen decreases in PI 197088 over time. In contrast, the percentage of *P. cubensis* reads from infected Vlaspik samples comprises about 20% of all of the inoculated samples, an observation that is consistent with successful pathogen infection and in agreement with our microscopy and qPCR analyses.

The distribution of all uniquely mapped reads for *P. cubensis* and cucumber shows strong enrichment for reads that are between 20 and 25 nt, which is consistent with the expected lengths of sRNA (Axtell, 2013a) (Fig. 4). In plants, miRNAs are typically 20 to 24 nt with the mode being 21 nt; this is consistent with the read distribution shown in Fig. 5. Other sRNAs within the distribution shown in Fig. 5 include small interfering RNAs (siRNAs; 20–24 nt), which are derived from double-stranded RNA. Additionally, long siRNA (lsiRNAs; 30–40 nt), which are siRNA-like sRNAs, have also been identified as being induced during bacterial infection (Katiyar-Agarwal et al., 2007). It is noteworthy that the enrichment of reads with a length of approximately 35 nt in the read distribution graph (Fig. 5) may reflect an enrichment in lsiRNA as a result of a response to pathogen infection in cucumber. Small RNAs have been identified in other oomycetes and have been found to be enriched in populations of reads that were predominantly 21 or 25 nt and derived from either inverted repeats and miRNA or transposable elements, respectively (Fahlgren et al., 2013). *Phytophthora infestans* sRNAs were also found to be enriched around 32 nt (Vetukuri et al., 2012). The read length distribution of *P. cubensis* is consistent with that of other oomycete plant pathogens (Fig. 4).

Of the reads that uniquely mapped to *P. cubensis* or cucumber, the number of Dicer-derived sRNAs, and specifically miRNAs, were predicted and counted using...
Fig. 5. Z-score plots of normalized expression values of predicted cucumber microRNA (miRNA) and corresponding predicted cucumber target genes. (A) Novel cucumber miRNA from csa-novel-1 (Cluster 11080) targeting CsaG000160. (B) csa-miR-156r (Cluster 11464) targeting Csa3G567830, Csa3G809420, and Csa4G075190. (C) csa-miR-396 (Cluster 7011) targeting Csa3G651860 and Csa2G354030. (D) csa-miR-167l (Cluster 6579) targeting Csa1G536820. P1, PI 197088 at 1 d postinoculation (dpi); P2, PI 197088 at 2 dpi; P3, PI 197088 at 3 dpi; P4, PI 197088 at 4 dpi; P6, PI 197088 at 6 dpi; V1, 'Vlaspik' at 1 dpi; V2, Vlaspik at 2 dpi; V3, Vlaspik at 3 dpi; V4, Vlaspik at 4 dpi; V6, Vlaspik at 6 dpi. The miRNA precursor structure is shown with the red and pink regions indicating the location of the mature miRNA.
ShortStack and HTSeq, respectively. Of the reads that uniquely mapped to each species, only a small percentage (less than 10%) mapped to predicted sRNA loci, and even fewer mapped to a predicted miRNA locus (Table 3). This reduction from the number of sequenced reads to the number of reads mapping to sRNA loci has been observed in other studies and (Visser et al., 2014; Martínez et al., 2011; Mao et al., 2012) and the number of reads used to predict and quantify sRNAs in this study was much higher than that used in previous studies identifying sRNAs in cucumber (Martínez et al., 2011; Mao et al., 2012, Li et al., 2014, Jin and Wu, 2015). In addition, the greater number of P. cubensis-derived sRNA and miRNA loci in the inoculated Vlaspik samples compared with the inoculated PI 197088 samples further supports that P. cubensis growth is greatly reduced in the resistant plant.

MicroRNA Analysis in Cucumber Reveals Conserved and Novel Loci

By analyzing pooled reads from both plant lines, novel miRNAs and miRNA families previously identified in cucumber and other species were found. As shown in Table 4, 28 conserved miRNA families were identified in this study including 23 families that were previously identified in cucumber (Mao et al., 2012; Martínez et al., 2011; Li et al., 2014, Jin and Wu, 2015). Of the remainder, four families were previously identified in other species recorded in miRBase yet not in cucumber. Many of these conserved families, notably miR156, miR166, miR167, and miR171, were all represented by several loci and mature miRNA from the cucumber samples analyzed in this study. In addition, many of the conserved miRNA target transcription factors. For example, miR156, miR157, miR169, and miR171 all target squamosa promoter-binding transcription factors; miR159 and miR319 target MYB transcription factors; and miR166 targets Class III homeobox leucine zipper proteins. We posit that miRNA targeting of transcription factors is relevant, as these conserved miRNAs could play a role in signaling pathways involved in resistance during infection. Other notable targets of the identified conserved miRNAs included auxin-related genes, targeted by miR160, the auxin responsive factor, targeted by miR167, and additional auxin-signaling genes, targeted by miR393. We hypothesize that this additional layer of regulation is relevant to the host–pathogen interaction, as a number of studies have demonstrated that auxin and SA function in apparent opposing ways to regulate the tradeoff between plant growth and defense (Huot et al., 2014). Finally, and of particular noteworthiness, was the identification of miR482, which was first discovered in cucumber in this study and was predicted to target a CC-NBS-LRR R-gene.

### Table 3. Read processing from total sequenced reads to total reads mapped to microRNA loci in *Pseudoperonospora cubensis*, cucumber ‘Vlaspik’, and cucumber PI 197088.

<table>
<thead>
<tr>
<th>Category of reads</th>
<th>Sporangia</th>
<th>Vlaspik mock</th>
<th>Inoculated Vlaspik</th>
<th>PI 197088 mock</th>
<th>Inoculated PI 197088</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total sequenced reads</td>
<td>28,775,108</td>
<td>22,199,524</td>
<td>118,004,094</td>
<td>15,511,102</td>
<td>102,910,883</td>
</tr>
<tr>
<td>Quality-filtered Reads</td>
<td>20,022,634</td>
<td>16,236,220</td>
<td>77,997,535</td>
<td>10,158,836</td>
<td>67,422,425</td>
</tr>
<tr>
<td>Reads unique to cucumber</td>
<td>7,622,556</td>
<td>30,989,042</td>
<td>4,855,957</td>
<td>2,771,988</td>
<td>12,516,322</td>
</tr>
<tr>
<td>Cucumber dicer-derived small RNA</td>
<td>3,651,841</td>
<td>11,075,300</td>
<td>2,771,988</td>
<td>12,516,322</td>
<td>12,516,322</td>
</tr>
<tr>
<td>Cucumber miRNA</td>
<td>1,265,624</td>
<td>2,898,729</td>
<td>820,587</td>
<td>3,082,744</td>
<td></td>
</tr>
<tr>
<td>Reads Unique to <em>P. cubensis</em></td>
<td>8,742,311</td>
<td>7,571,862</td>
<td>3,248,927</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P. cubensis</em> dicer-derived small RNA</td>
<td>803,862</td>
<td>390,125</td>
<td>54,608</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>P. cubensis</em> miRNA</td>
<td>199,165</td>
<td>93,174</td>
<td>7801</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 4. Conserved microRNA (miRNA) families from cucumber or in other plant species in miRBase. The number of loci and mature miRNA were identified in this study using ShortStack and comparing the predicted mature miRNA to miRBase.

<table>
<thead>
<tr>
<th>Conserved miRNA family</th>
<th>Loci in cucumber</th>
<th>Mature miRNA</th>
<th>Cucumis sativus</th>
<th>Arabidopsis thaliana</th>
<th>Vitis vinifera</th>
<th>Cucumis melo</th>
<th>Glycine max</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR156</td>
<td>6</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR157</td>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR159</td>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR160</td>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR162</td>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR164</td>
<td>3</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR166</td>
<td>5</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR167</td>
<td>6</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR168</td>
<td>2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR169</td>
<td>5</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR171</td>
<td>6</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR172</td>
<td>2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR191</td>
<td>3</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR390</td>
<td>2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR393</td>
<td>2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR396</td>
<td>2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR398</td>
<td>2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR399</td>
<td>4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR408</td>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR477</td>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR482</td>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR827</td>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR1175</td>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR2111</td>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR2950</td>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR3627</td>
<td>2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR7741</td>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>miR8210</td>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

---

14 OF 19  THE PLANT GENOME  MARCH 2016  VOL. 9, NO. 1
In addition to identifying previously described miRNA families, 57 novel mature miRNAs were identified. This represents a significant expansion in the number of cucumber miRNAs identified to date, and we posit that this increase in the number of miRNAs is likely attributable to the deeper miRNA sequencing undertaken in the current study. Indeed, previous analyses generated 209,331 (Martínez et al., 2011), 4,012,509 (Mao et al., 2012), or 27,765,704 (Jin and Wu, 2015) high-quality reads, whereas we generated over 70 million high-quality cucumber reads (Table 3). An additional, or alternative, factor influencing this increased number may also be a reflection of biotic stress induction with both a resistant and susceptible plant (Inal et al., 2014). Among the novel cucumber miRNA found in this study, four were previously identified as novel miRNA or miRNA* sequences in other studies (Martínez et al., 2011; Li et al., 2014; Jin and Wu, 2015). Among the predicted targets of novel cucumber miRNAs was a mildew resistance locus O (MLO)-like protein, which has a role in mediating powdery mildew defense (Fig. 6) (Acevedo-Garcia et al., 2014). This predicted target suggests a role for newly discovered miRNAs in plant defense. In addition, other miRNAs unique to this study are predicted to target various kinases, some of which may have a role in biotic stress signaling. In addition, many miRNAs have unidentified targets; this could indicate that some of the miRNAs are lineage-specific miRNAs that are typically lower abundance and have poorly defined targets (Ma et al., 2010; Fahlgren et al., 2010; Axtell 2013a).
MicroRNA Regulation of Vlaspik and PI 197088 Targets during Pseudoperonospora cubensis Infection

After identifying the suite of conserved and novel miRNAs described above, we analyzed the miRNA count data for each sample to identify significantly differentially expressed miRNAs. Twenty-seven miRNA loci were predicted to significantly change in expression over time and 10 miRNA loci were differentially expressed between the plant lines. For the miRNAs for which targets were predicted, the Z-scores of the normalized expression of the miRNA were compared with the Z-scores of the normalized expression of the target cucumber transcript. Using this information, opposing patterns in gene regulation were used in support of the hypothesis that the miRNA targets were correctly predicted. For example, and as shown in Fig. 5, miRNAs, as well as their predicted target gene expression patterns, suggest that the transcripts of the targets may be regulated by miRNA such that when the miRNA level is high, a parallel reduction in the target transcript level is observed. Conversely, when the miRNA level is reduced, an increased transcript level is observed. For example, as shown in Fig. 5A, the novel mature miRNA predicted to target Csa3G000160, an MLO-like protein, showed an expression pattern that was oppositely correlated with the expression of the MLO-like transcript. This MLO-like gene has membership in Vlaspik Module K and PI 197088 Module C (Fig. 2). Additionally, csa-miR-156r had an expression pattern that oppositely correlated with expression pattern of the predicted targets encoding squamosa promoter binding proteins (Csa3G567830 and Csa3G809420) (Fig. 5B). These genes were significantly DE genes and were in the Vlaspik–PI 197088 Module G (Supplemental Fig. S3). Likewise, csa-miR-396 had evidence of targeting Csa3G651860 and Csa2G354030, which are both growth promoting factors and are also members of Module G from Vlaspik–PI 197088 (Fig. 5C; Supplemental Fig. S3). Finally, csa-miR-167l targeted an auxin response factor encoded by Csa1G536820, which was in the Vlaspik–PI 197088 Module G (Fig. 5D; Supplemental Fig. S3).

Discovery of MicroRNA Loci in Pseudoperonospora cubensis

In the Arabidopsis–Botrytis interaction, sRNAs from the pathogen have been shown to directly target host defense-related genes (Weiberg et al., 2013). To determine if P. cubensis-derived sRNA exist, we sequenced the small RNAome of P. cubensis and predicted cucumber target genes. Of the sRNA loci identified in P. cubensis, four new miRNA loci and 11 mature miRNA were identified that were unique to P. cubensis. Similarly, six miRNA loci were observed in another oomycete, P. infestans, which is consistent with the number of miRNA loci in P. cubensis (Vetukuri et al., 2012). The miRNA loci identified in P. cubensis all displayed characteristic hairpin loop precursors (Fig. 6). For each predicted miRNA loci, 1 to 3 mature miRNA were identified, and the 5’ nt of each miRNA was either a uracil or a cytosine.

In a paradigm-shifting study involving the fungus B. cinerea and Arabidopsis, it was shown that fungal miRNA could be transported from the pathogen into the plant to target and silence genes involved in immunity (Weiberg et al., 2013). Based on this observation, potential cucumber targets of P. cubensis miRNA were identified (Table 5). Two predicted P. cubensis miRNA, PcmiRNA2 and PcmiRNA4, had predicted cucumber genes targets, while the remaining two P. cubensis miRNA did not. Interestingly, PcmiRNA2 was predicted to target the TIR-NBS-LRR cucumber gene, Csa5G647510. Future work will be needed to confirm the mode of action and specificity of the predicted miRNA and targets.

Conclusions

Based on the sum of the work presented herein, we posit that the timing of the plant response is key to resistance in response to P. cubensis infection, as our data show that the resistant line PI 197088 displayed a faster, more robust, response to pathogen infection. For example, the number of genes differentially expressed at early time points and genes grouped in modules indicated that a large-scale change in transcription occurred more rapidly.
in PI 197088 in response to *P. cubensis*. In addition, our observed regulation of several host gene groups and pathways suggested that resistance in PI 197088 might be attributed to a combination of hormones and regulation of nutrient supply. For example, genes associated with SA and vitamin B1 (thiamine) synthesis were strongly differentially regulated between the resistant and susceptible plant over the time course of infection. However, the work from this study focused only on transcript data and was unable to identify which genes may be a cause or effect of the resistance response in PI 197088. Future studies are necessary to investigate the in planta protein and metabolite levels of genes of interest to further test the hypotheses developed in this study. In addition to sequencing transcripts, this work also sequenced sRNA and predicted miRNAs in cucumber and *P. cubensis* that contribute to the complex interactions between the host and pathogen. Many miRNAs had predicted cucumber targets that could mediate signal transduction pathways within cucumber. In total, this work provides a foundation for the core set of candidate resistance-associated genes and miRNAs uncovered through this study.

### Supplementary Materials

**Supplemental Table S1.** List of primer sequences used for qPCR to validate expression patterns of mRNA.

**Supplemental Table S2.** The percent of trimmed, high quality reads uniquely mapping to the Chinese long *Cucumis sativus* genome v. 2 using Bowtie 1.0.0 for each biological replicate of Vlaspik and PI 197088.

**Supplemental Table S3.** Percentage of trimmed and quality filtered reads mapping to either *P. cubensis* genome or to the cucumber Chinese long genome v. 2 using Bowtie 1.0.0. *Sp.* = sporangia, dpi = days post inoculation.

**Supplemental Fig. S1.** Correlation of expression values in two biological replicates of PI 197088 and Vlaspik inoculated with *Pseudoperonospora cubensis* or mock-inoculated with water. Reads were mapped to the *Cucumis sativus* genome v. 2 using Bowtie 1.0.0 and TopHat v. 1.4. The number of reads mapping to each gene were counted using HTSeq v. 0.6 and were normalized divided by the library size for each sample calculated from DESeq2. The data was then log2 transformed and negative values were converted to zero. $R^2$ is the correlation coefficient. dpi = days post inoculation.

**Supplemental Fig. S2.** Venn diagrams of significantly differentially expressed genes from DESeq at each time point compared to mock between Vlaspik and PI 197088. (A) 1 dpi (B) 2 dpi (C) 3 dpi (D) 4 dpi (E) 6 dpi. dpi = days post inoculation.

**Supplemental Fig. S3.** Validation of RNA-Seq data from select genes with qPCR. RNA-Seq plots are log2 transformed normalized RNA-Seq counts with error bars representing the standard error from 2 biological replicates. qPCR plots represent the relative expression of each gene of interest normalized to actin with the error bars representing the standard error of 3 biological replicates.

### Supplemental Fig. S4.** Trend plots from the weighted gene correlation network analysis of normalized gene expression values from Vlaspik, PI 197088, and combined Vlaspik and PI 197088. dpi = days post inoculation.

### Author-recommended Supplemental Files

http://www.daylab.plp.msu.edu/aburkardt-supporting-files

Username: data
Password: ABManuscriptFiles

### Acknowledgments

AB is supported by a Michigan State University Graduate Student Distinguished Fellowship. Research in the laboratory of BD is supported by funding from the Michigan State University Rackham Foundation, Michigan State University Project GREEEN (GR13-007), and a grant from the USDA Specialty Crops Research Initiative (2011-5181-30661). We would like to thank Kevin Childs and Robin Buell for valuable discussions at the onset and throughout the research presented herein.

### References


