From Wei Zhang 2018: **Genetic Variation in *Botrytis* Controls the Transcriptional Plasticity in *Planta***

METHODS:

whole-genome expression profiles of 96 *B. cinerea* isolates at 16-hours post-infection (HPI) onto the Col-0 accession of Arabidopsis (Zhang et al., 2017). Together with wild-type Col-0, two immune-compromised, single gene knockout mutants *coi1-1* and *npr1-1*, which abolish or diminish the major defense signaling pathways of JA and SA respectively, are also used in the pathosystem (References). The number of reads from each biological replication of the *B. cinerea* infected Arabidopsis leaf tissues across three genotypes were mapped to the *B. cinerea* isolate B05.10 (Van Kan et al., 2017) and Arabidopsis TAIR10.25 cDNA reference genome. The total of 9284 *B. cinerea* gene models were kept with more than 30 counts in one isolate or 300 counts across 96 isolates, which corresponded to 79.36% of the total predicted genes in the *B. cinerea* genome.

We first used a generalized linear model linked with negative binomial (nbGLM) to calculate the least square mean (log2) and standard error for each *B. cinerea* transcript (Supplemental Data Sets 2, 3, and 4). To test how genetic variations in pathogen and hosts influence *B. cinerea* transcriptomic response during infection, we estimated the broad-sense heritability (H2) for each *B. cinerea* transcript that contributed by pathogen, plant hosts, and their interaction (Figure 3 and Supplemental Data Set 4). Compared with host genotypes, both pathogen genotypes and interaction between pathogen and host altered a large amount of *B. cinerea* transcripts expression profiles (**H2Isolate = 0.152**, H2Host = 0.010, **H2Isolate x Host = 0.116**, respectively) (Figure 3A and Supplemental Dataset 5).

The top 100 genes that significantly influenced by pathogen genotypes showed the highest heritability ranging from 0.338 to 0.460

***B. cinerea* Genotyping**

A total of 96 *B. cinerea* isolates were selected for their phenotypic and genotypic differences (Rowe and Kliebenstein, 2007; Atwell et al., 2015; Corwin et al., 2016; Zhang et al., 2016). Isolates of *B. cinerea* were grown on potato dextrose agar (PDA) covered with sterile cellophane for approximately 4-5 days. The young leading mycleial front was collected using a spatula and total DNA was collected using the Qiagen DNeasy Plant Mini kit (Qiagen, Hilden, Germany). Illumina DNA libraries were created by randomly fragmenting 1g of DNA using the Fragmentase enzyme mix (New England Biolabs, Ipswich, MA) incubated at 37⁰C for 30 minutes followed by end-repair, A-tailing, adapter ligation, and fragment enrichment as described for the RNASeq libraries (Reference AtBc Paper). The resulting DNA libraries were pooled in batches of 8 libraries each and submitted for 150 bp paired-end sequencing on four lanes of Illumina HiSeq 2500 (San Diego, CA) at the UC Davis Genome Center - DNA Technologies Core (Davis, CA).

Raw reads of genomic libraries were cleaned using Fastq-mcf from the ea-utils package (http://code.google.com/p/ea-utils) to remove reads with an average Phred quality score less than 30 and trim the first 9 bp of each read. Alignment and SNP calling was preformed as described (Atwell et al., 2015). Briefly, the alignment program STAMPY (Lunter and Goodson, 2010) was used to align the fragments to the T4 Botrytis cinerea reference from Botrytis cinerea Sequencing Project at Broad Institute of Harvard and MIT (http://broadinstitute.org/). Alignment files were prepared using Samtools (Li et al., 2009) and Picard (http://sourceforge.net/projects/picard/) and high quality (MQ > 85) SNPs were identified using the GATK UnifiedGenotyper (DePristo et al., 2011).

**Plants Material and Growth Conditions**

The *Arabidopsis thaliana* (Arabidopsis there after) ecotype Columbia-0 (Col-0) is the genetic background of the wild-type Col-0 and two mutants. We utilized the wild-type Col-0 in conjunction with *coi1-1* and *npr1-1* mutant plant genotypes that abolish the major defense perception pathways of jasmonate and salicylic acid, respectively. These mutations are EMS single point mutant alleles that produce modified version of *COI1* and *NPR1* proteins that confer dominant JA- and SA-insensitive phenotypes. We grew three accessions in two randomized complete blocks across over 6 flats with 15 plants per flat for a total of 90 plants per experiment and two independent experiments were conducted separately. The *Arabidopsis* seeds were vernalized in 0.1% phytoagar at 4 oC for 4 days in the dark. Three seeds of each accession were placed in the center of a cell filled with soil (Sunshine Mix #1, Sun Gro Horticulture, Agawam, MA) and covered the flat with a transparent plastic hood to maintain the humidity during germination. The transparent hood was removed one week after germination and the plants were thinned to one plant per cell. We watered plants twice a week using nutrient-enriched water and kept them grown in a growth chamber at 22 oC under a photoperiod of 10 h light/14 h dark. At five weeks after sowing, the first five true leaves were harvested from each plant for *B. cinerea* isolate infection and placed on the 1% phytoagar in large plastic trays.

***B. cinerea* Growing Conditions and Inoculation**

All *B. cinerea* isolates were cultured and inoculated on three *Arabidopsis* genotypes. Small canned peach slices (~3 cm3) in petri plates were inoculated from frozen glycerol stocks of isolate spores and were incubated on the bench top for one week. Spores were collected by submerging the sporulating peach slices in 5 mL sterile ultra-pure water and agitated with a flame sterilized glass rod. The solution was filtered through a syringe containing a small plug of sterilized glass wool followed by a mild centrifugation in a bucket rotor at 1,000 rpm for 15 min. The supernatant was discarded and the spore pellet was re-suspended in sterilized 1/2x organic grape juice (Santa Cruz Organics, Pescadero, CA). Spore concentration was determined using a hemacytometer and spore solutions were diluted to 10 spores/μL. Five-week old detached leaves of Col-0, *coi1-1* and *npr1-1* were inoculated with 4 μL of the diluted spore solution and were incubated at room temperature on flat containing ~2 cm of 1% phytoagar covered with a humidity hood. *B. cinerea* isolates were inoculated in a randomized complete block design across the six planting blocks. Two of the six blocks were harvested at 16 hours for transcriptome analysis by transferring the infected leaf into a 2 mL Eppendorf tube containing one 4 mm and four 2.4 mm stainless steel ball bearings and immediately submerging the tube in liquid nitrogen. Frozen tubes were stored at -80 oC until processing. The infected leaf tissues were first taken pictures for lesion analysis and then placed in 400 μL of 90% methanol for camalexin extraction at 72 hours-post inoculation (HPI).

**RNASeq library Construction, Sequencing, and Alignment**

RNASeq library construction was mainly based on previous method (Kumar et al., 2012) and we made a minor modification in a high throughput way. Frozen infected leaves were homogenized by rapid agitation in a bead beater followed by direct mRNA isolation and purification from tissue lysate using the DynaBeads® Oligo (dT)25 Kit (Invitrogen, Carlsbad, CA). The first strand synthesis of mRNA was created using the SuperScript III kit (Invitrogen, Carlsbad, CA) with random primers and RNaseOut. The second strand synthesis was accomplished using 50 units of DNA Pol I (Fermentas, Vilnius, Lithuania) with an equal mix of 2.5 mM dNTPs and 1.6 units of RNase H. The resulting cDNA was purified using 0.8x of AMPure beads and fragmented using the Fragmentase enzyme (New England Biolabs, Ipswich, MA) for 20 min at 37oC. Fragmented cDNA was repaired using the End Repair kit (New England Biolabs, Ipswich, MA) and an A-base was added using the exo- Klenow fragment (New England Biolabs, Ipswich, MA) with 0.8x AMPure bead clean up steps between both reactions. A barcoded sequencing adapter (NEXTflex DNA Indexed Adapters, Bioo Scientific, Austin, TX) was ligated to sample fragments using the NEB Quick Ligase ligation kit (New England Biolabs, Ipswich, MA). Adapter-ligated fragments were size selected using 0.8x Ampure beads and PCR-enriched using the Phusion Master Mix (New England Biolabs, Ipswich, MA), 10 μM PE sequencing primers. PCR conditions ran for 98oC for 30 minutes followed by 14 cycles of 98oC for 30 sec; 62oC for 30 sec; 72oC for 30 sec and final extension at 72oC for 5 min. A total of 2 μL of the PCR-enriched samples were run on a 1.5% Agarose gel to test for successful amplification. Successfully amplified samples were size selected using 0.7x AMPure beads. Amplified and size selected libraries were pooled in group of 96 libraries according to the unique indexed barcodes within the adapter and submitted for single read, 50 bp sequencing on a single lane of Illumina HiSeq 2500 (San Diego, CA) at the U.C. Davis Genome Center - DNA Technologies Core (Davis, CA).

Fastq files from individual HiSeq lanes were separated by adapter index into individual RNASeq library samples. Individual libraries were qualitatively assessed for overall read quality and over-represented sequences using FastQC software (Version 0.11.3, [www.bioinformatics.babraham.ac.uk/projects/](http://www.bioinformatics.babraham.ac.uk/projects/)). Bowtie 1 V.1.1.2 (<http://sourceforge.net/projects/bowtie-bio/files/bowtie/1.1.2/>) was used to align the processed reads against the *Arabidopsis* *thaliana* TAIR10.25 cDNA reference genome (<http://ftp.gramene.org/archives/PAST_RELEASES/release44/data/fasta/arabidopsis_thaliana/cdna/)> using phred33 quality scores and trimming the first 10 bp due to low quality at the beginning of reads by a customer R script (Langmead et al., 2009; Li et al., 2009). Gene counts were pulled from the resulting sam file generated from Bowtie 1 and then summed to reduce overrepresentation of genes with multiple splice variants.

**Statistical analysis methods**

All the analyses were conducted using R V3.2.1 statistical environment (R Development Core Team, 2011).

To test for the ability of natural variation within *B. cinerea* to impact disease related phenotypes, we ran the following Negative Binomial Generalized linear model for camalexin, lesion area, and all transcripts:

Camalexin/Lesion Area: Yegai = Ee + Ee(Gfg) + Ee(Gfg(Afa)) + Ii

Transcripts in infected samples in host genotypes:

Yegai = Ee + Ee(Gfg) + Ee(Gfg(Afa)) + Ii + Hh + Hh\* Ii

where the main effects E, I, and H are denoted as experiment, isolate genotype, and plant host genotype, respectively. Nested effects of the growing flat (Gf) within the experimental replicates and agar flat (Af) nested within growing flat are also accounted. Gene count data obtained from RNASeq experiments were subjected to multiple statistical tools. Normalization on gene counts was first conducted using the TMM method in function calcNormFactors() from “edge R” package (Robinson and Smyth, 2008; Robinson et al., 2010; Robinson and Oshlack, 2010; Nikolayeva and Robinson, 2014). The linear model was conducted on normalized gene counts using function glm.nb () from “MASS” package (Venables and Ripley, 2002). Model corrected means and standard errors for each isolate was determined using the “lsmeans” V2.19 package (Lenth, 2016). p-Values for F- and ChiSq-test were determined by Type II sums of squares using function anova()from the “car” package (Fox and Weisberg, 2011). False discovery rate (FDR) (p-Value < 0.05) was used for correction p-Value for multiple tests of significance (Benjamini et al., 2001; Strimmer, 2008). Broad-sense heritabilities of each phenotype and transcripts as the proportion of variance attributed to *B. cinerea* genotype, *Arabidopsis* genotype, or their interaction effect to the total variance within the model were estimated. We then compared broad-sense heritabilities of transcripts involved in biosynthesis pathway of camalexin, tryptophan, JA, and SA in each Arabidopsis genotype. The violin plots were generated to depict the distribution of heritability using violin package.

Gene co-expression networks for all Arabidopsis transcripts were generated using the model corrected means for each isolate. Spearman’s rank correlation coefficient of these model corrected means of transcript in each infected *Arabidopsis* genotype plant tissues were calculated for all gene pairs using function cor () in R. For gene pairs with a positive Spearman’s correlation greater than 0.95 in each dataset were selected as co-expressed genes and overlapped genes from three selected gene datasets were used to construct gene co-expression networks. Gene co-expression network were visualized using Cytoscape V3.2.1 (Java version:1.8.0\_60) (Shannon et al., 2003).Genes identified from each *Arabidopsis* genotype represented the nodes in the networks and were named based on AGI Locus IDs as published by TAIR (The Arabidopsis Information Resource, [https://www.Arabidopsis.org/](https://www.arabidopsis.org/)). Genes involved in biosynthesis of camalexin, tryptophan, JA and SA were named using their general name and marked in red, yellow, orange and blue, respectively. *Arabidopsis* chloroplast-encoded genes were marked in purple in networks. The *B. cinerea* induced transcript profiles of genes involved in camalexin, tryptophan, SA and JA pathways, as well as in plant immune systems were illustrated by violin plots (Adler, 2005). Tukey’s multiple comparison was performed using the HSD.test function from “agricolae” package (Mendiburu, 2016).

To investigate the associations among Arabidopsis wild-type Col-0, *coi1-1*and *npr1-1* challenging by 96 diverse *B. cinerea* isolates, we conducted a mantel test on correlation matrices generated by model corrected transcript means in each *Arabidopsis* genotype using a customer R script based on Function getPermuteMatrix() from **“**vegan” V2.3.0 package (Mantel, 1967; Oksanen et al., 2016). Mantel test was inferred following 999 permutations using Spearman’s Rank coefficient method.

We selected top 2000 genes that expression was significantly regulated by *B. cinerea* infection. Principal component analysis (PCA) was performed on model corrected means of each transcript in Arabidopsis wild-type Col-0 using princomp function in R to capture the unobserved data structure. The first two principal component score values for two mutant datasets as well as transcripts in the mock control samples were predicted using wild-type Col-0 as training model. The PCA was then conducted on transcripts of each *Arabidopsis* mutant to obtain the data structure. The contributions of the first five principal components were illustrated by bar chart. The relationships among the first five components were displayed by scatter plot. We employed an ANOVA model to identify the statistical significance of influence of host genotype, each principal component, or a specific interaction between host genotype and a given principal component. The ANOVA model was Lesion area/ Camalexin = Host genotype + Each principal component + Host genotype x Each principal component + error.

Heat map visualization used for clustering analysis of lesion and camalexin was based on heatmap.2 function from “gplots” R package (Warnes et al., 2016). Gene annotation was obtained using <http://www.uniprot.org/>. Gene Ontology enrichment analysis was performed with a BiNGO plugin in Cytoscape environment by using Fisher’s Extract Test with Multiple Testing Correction of FDR at 0.05 (Benjamini and Hochberg, 1995; Maere et al., 2005). KEGG pathway analysis were conducted using online tool DAVID (<https://david.ncifcrf.gov/)> (Huang da et al., 2009b, a).