Running title: Botrytis eQTL of co-transcriptome

**Pathogen genetic control of transcriptome variation in the *Arabidopsis thaliana* – *Botrytis cinerea* pathosystem**

**Nicole Soltis1,2, Wei Zhang3, Jason A. Corwin4, Susana Atwell2, and Daniel J. Kliebenstein2,5\***

1Department of Plant Sciences, University of California, Davis, One Shields Avenue, Davis, CA, 95616, USA

2Plant Biology Graduate Group, University of California, Davis, One Shields Avenue, Davis, CA, 95616, USA

3Department of Plant Pathology, Kansas State University, 1712 Clafflin Road, Throckmorton Hall, Manhattan, KS, 66506, USA

4Department of Ecology and Evolution Biology, University of Colorado, 1900 Pleasant Street, 334 UCB, Boulder, CO, 80309-0334, USA

5DynaMo Center of Excellence, University of Copenhagen, Thorvaldsensvej 40, DK-1871, Frederiksberg C, Denmark

**\*Correspondence:** Daniel J. Kliebenstein, Department of Plant Sciences, University of California, Davis, One Shields Ave, Davis, CA, 95616, USA.

Kliebenstein@ucdavis.edu

**Keywords: Host-Pathogen Interaction, Pathosystem, Arabidopsis, *B. Cinerea*, RNA-Seq, Genome Wide Association, Dual transcriptome**

**ABSTRACT**

**INTRODUCTION**

1. **Disease is an interaction.**

**Dominant focus on large effect loci but majority are quantitative (including symbiotic interactions)**

**Genomics of quantitative interactions via transcriptome**

**Response / counterresponse in host and pathogen interactions**

**Mechanisms of interspecific communication not well known**

1. **Cotranscriptome with At/ Botrytis**

**Not dominated by large effects, so don’t lose the more complex interactions**

**Can measure both host and pathogen**

**Tons of diversity within genomes**

**But causality unknown**

1. **To begin finding causal hypotheses: eGWA on host/ pathogen**
2. **Results summary paragraph**

Infectious disease is an interaction between host and pathogen, determined by the genetics of both and their interactions. Plant-pathogen interactions are often parsed as qualitative, in which few genetic variants interact to determine binary disease outcomes, or quantitative, in which a spectrum of outcomes may occur due to genetic variation between the host and pathogen. The past decade has revealed the molecular basis of numerous large-effect loci on either the host side or the pathogen side that control qualitative interactions. Studies have begun to elucidate the host genetic basis of quantitative plant-pathogen interactions. The pathogen genetic basis is less clear and is being elucidated through the study of phenotypic variation across genetically diverse hosts and pathogens, finding links to genetic variation.

There is a lack of evidence for qualitative virulence/ resistance genes underlying quantitative disease outcomes in plant-pathogen interactions. Rather, the genetic basis of plant resistance in these interactions is highly polygenic (Glazebrook 2005, Nomura, Melotto et al. 2005, Goss and Bergelson 2006, Rowe and Kliebenstein 2008, Barrett, Kniskern et al. 2009, Corwin, Copeland et al. 2016, Fordyce, Soltis et al. 2018).

On the pathogen side, recent studies have accumulated evidence for a polygenic basis of virulence as well (Corwin, Copeland et al. 2016, Bartoli and Roux 2017, Wu, Sakthikumar et al. 2017, Zhang, Corwin et al. 2017, Atwell, Corwin et al. 2018, Soltis, Atwell et al. 2019). These studies provide many candidate loci for pathogen resistance, with diverse functional annotations. Some of these loci appear to modulate virulence across multiple virulence phenotypes, including lesion size across multiple hosts and independent phenotypes of lesion growth (Corwin, Copeland et al. 2016, Fordyce, Soltis et al. 2018, Soltis, Atwell et al. 2019). However, thus far we know little about the molecular mechanism of action by which these genes affect virulence outcomes, or how the host and pathogen genetics interact.

One study has taken a genome-wide association (GWA) approach to the plant and pathogen genomes simultaneously, as well as their genetic interactions (Zhang, Corwin et al. 2018). In this system, most of the variation in disease resistance was determined by a polygenic structure in the pathogen, with small genetic effects from the host genome or the interaction between them. Similarly, viral load in the human x HIV pathosystem is better explained by pathogen than host diversity (Bartha, McLaren et al. 2017).

Many quantifiable phenotypes result from the interaction of plant and pathogen and can be used in GWA studies, including gene expression responses. Each expression profile may be considered a unique indicator of the progression of the interaction between host and pathogen. As such, analysis summarizing information across transcriptomes can elucidate the common and specific genetics underlying virulence phenotypes and hypothesize causal relationships between genetic variation and expression responses. Individual gene expression profiles can be treated as phenotypes for analysis using recombinant inbred lines (RIL) or GWA. Expression quantitative trait loci (eQTL) are the markers correlated with variation in transcripts’ expression profiles and are hypothesized as points of direct or indirect genetic control over expression variation. Locally acting (*cis*) eQTL may indicate regulatory variation within the expressed gene itself, or nearby. Additional markers distant from the responding gene are classified as *trans*-eQTL. *trans*-eQTL may be due to genes present in a common regulatory network, or transcription factors acting upon the expressed gene. *trans*-eQTL hotspots (loci linked to expression variation across many transcripts) may point to master regulators, with extensive pleiotropy across many genes.

eQTL analysis focusing on one side of the plant-pathogen interaction can identify pathogen loci controlling pathogen gene expression in planta, or plant loci controlling plant gene expression over the course of infection. This provides a view of how variation in one of the interacting organisms leads to the shared disease phenotype and can identify virulence loci in the pathogen and resistance loci in the plant (Chen, Hackett et al. 2010, Christie, Myburg et al. 2017). However, few studies have conducted genome-wide association for any disease traits on the pathogen side, let alone expression traits for eQTL studies (Bartoli and Roux 2017). Further, these analyses cannot account for the signaling and decision-making that occur between the two organisms’ genomes over the course of infection.

Any genes linked to expression variation of many members of the previously described *A. thaliana* and *B. cinerea* virulence co-expression networks both affirms the biological relevance of the pathway and suggests a genetic control factor in pathway-level expression variation. Determining the pathogen genetic control of both host and pathogen gene expression over the course of infection can give us inference into points of genetic control over virulence pathways in the pathogen. Further, it can elucidate the sensitive host pathways, to inspire a search for potential resistance alleles among host variants. If we consider the full transcriptome of host and pathogen, this provides us thousands of phenotypes to test in genome-wide association, and we can deepen our search for loci which control multiple phenotypic measures of the progression of the plant-pathogen interaction. We can build inference on which genes in the pathogen are core factors in the virulence interaction, and which are uniquely controlling specific attributes of the interaction.

Studies encompassing transcriptomic variation in both the host and pathogen, and genomic variation within one of the interacting organisms, can look for signs of interspecific *trans*-eQTL; loci in the pathogen that modulate expression in the infected host, or loci in the host that modulate expression in the infecting pathogen. A few studies have examined variation in host-pathogen interactions in this way, validating the ability of this approach to identify pathogen loci modulating host expression levels, and thus candidate loci for interspecific signals (Wu, Cai et al. 2015, Guo, Fudali et al. 2017). These cross-species *trans*-eQTL can be used to hypothesize causal relationships between individual genes in the interspecific interaction (Wu, Cai et al. 2015, Guo, Fudali et al. 2017).

Previous studies have identified hundreds of plant host genes with one or more eQTL in the pathogen modulating their expression (Wu, Cai et al. 2015, Guo, Fudali et al. 2017), and one small study identified 3 bacterial parasite genes modulated by human host eQTL (Guo, Fudali et al. 2017). Hundreds of pathogen eQTL linked to host expression changes (Wu, Cai et al. 2015), and a single eQTL may modulate expression of over 60 host genes (Guo, Fudali et al. 2017). Thus, the majority of the disease interactions appear to be modulated by very few loci within the pathogen genome.

We have previously successfully conducted GWA in this pathogen population for the phenotype of lesion size {Soltis 2019}.

Previous studies in the *A. thaliana* - *B. cinerea* pathosystem point to transcriptome variation on both sides of the interaction. Detached leaves of wildtype *A. thaliana* and major immune pathway mutants were inoculated with 96 genetically variable isolates of *B. cinerea,* and at 18 hours post inoculation, mRNA was collected. Variation in expression of *A. thaliana* genes was very sensitive to pathogen genetic variation; expression of host genes was under approximately equal regulation from genetic variation across the *B. cinerea* isolates, and host immune-pathway responsive variation across the *B. cinerea* isolates (Zhang, Corwin et al. 2017). This far exceeded the contribution of the major host immune pathway variants to variation in gene expression (Zhang, Corwin et al. 2017). The host-pathogen genetic interactions target four major host response networks; jasmonic acid and salicylic acid signaling and camalexin biosynthesis, defense and cell cycle, and two photosynthesis networks (Zhang, Corwin et al. 2017). The authors analyzed co-expression of genes across the *B. cinerea* isolates and *A. thaliana* immune pathway mutants. Genes were condensed into co-expression networks, which hypothesize causal links between many genes in an interacting web. Within this host-pathogen system, all transcriptome variation is a result of the interaction of the two species, and not independently determined by one of the organisms. However, these analyses do not untangle the directionality of effect from one gene, one pathway, or one genome to another.

Previous analysis showed that the vast majority of transcripts are affected by variation in the *B. cinerea* genome. Of the differentially expressed *B. cinerea* genes, 74% showed significant variation in their expression due to *B. cinerea* genetic variation, with an average broad-sense heritability of 0.152 attributed to *B. cinerea* isolate (Zhang, Corwin et al. 2018). Of the differentially expressed *A. thaliana* genes, 85% showed significant variation in their expression due to *B. cinerea* genetics, with an average broad-sense heritability of 0.108 due to *B. cinerea* isolate (Zhang, Corwin et al. 2017). As such, we conducted genome-wide association (GWA) to look for evidence of loci in *B. cinerea* that may be modulating this heritable expression variation.

In this study, we focus on an extreme generalist pathogen with high genetic diversity, *B. cinerea*, and the model plant host, *A. thaliana*. *B. cinerea* exhibits highly quantitative virulence, facilitated through natural genetic variation (Rowe, Walley et al. 2010, Corwin, Copeland et al. 2016, Zhang, Corwin et al. 2018). These interactions are well-characterized phenotypically, and we have previous information on some of the potentially relevant genetic factors on both the pathogen and host sides of the interaction (Denby, Kumar et al. 2004, Rowe and Kliebenstein 2008, Zhang, Corwin et al. 2017, Atwell, Corwin et al. 2018, Soltis, Atwell et al. 2019). This also gives us the opportunity to connect our findings, particularly in plant genetic targets and affected pathways, to many previous datasets.

In this study, we ask how genetics within the pathogen may modulate expression variation over the course of infection. We work with the gene expression data from Zhang *et al*., performing genome-wide association (GWA) of variation in individual transcript expression profiles with SNP level variation within the *B. cinerea* genome when infecting the wildtype host Col-0 *A. thaliana*. This gives us a hypothesis of directionality; any locus in *B. cinerea* linked to expression variation in the host or pathogen is directly or indirectly modulating expression. With numerous traits in this analysis, we focused on general patterns of eQTL distribution across the genome, and identification of major hotspots of eQTL.

**RESULTS**

**eQTL indicate polygenic transcriptome modulation**

To better understand how natural genetic variation in the pathogen influences both the host and pathogen transcriptomes, we performed genome-wide association (GWA) for eQTL detection across all genes expressed in both species within the *B. cinerea* - *A. thaliana* pathosystem. This incorporated the expression profiles of 9,267 *B. cinerea* genes and 23,947 *A. thaliana* genes,each as individual traits. For each trait, we used a used Genome-wide Efficient Mixed Model Association (GEMMA) mode with a previous genome-wide SNP dataset of 237,878 SNPs with a minimum minor allele frequency of 0.20 {Zhou 2012; Atwell 2018}. GEMMA estimates the significance of effects of each SNP on the focal trait as a p-value after accounting for potential effects of population structure within the *B. cinerea* isolates. In total, GEMMA was able to identify *B. cinerea* SNPs linked to transcriptional variation in 5,213 *A. thaliana* genes and 1,616 *B. cinerea* genes. For these genes with significant SNPs, there was a median of XX SNPs per transcript (Range 1 to 16,818 SNPs) for *B. cinerea*, and a median of XX SNPs per transcript (Range 1 to 24,623 SNPs) for *A. thaliana* transcripts (Figure N1). Further, the distribution of p-values for significant SNP found little evidence for large effect polymorphisms suggesting a highly polygenic basis of loci modulating transcriptome variation.

Given the scale of this dataset, it was not viable to estimate empirical significance thresholds for each and every transcript using 1,000 or more permutations. However, we permuted the whole dataset across all of the tens of thousands of traits five times and repeated the GEMMA to get a feel for the potential for dominant patterns that may exist randomly (individual expression profiles in *B. cinerea* and *A. thaliana*). We then compared the permuted minimum p-value per transcript across all SNPs to the data obtained from real traits. This showed that the top SNP per trait for most genes show a stronger association in our observed data than across any of the 5 permutations. In *B. cinerea*, the observed p-value is lower for 69% of genes, and in *A. thaliana* the observed p-value is lower for 58% of genes. Thus, to develop genomic images of the results, we focused on the top SNP per transcript for the remaining analysis.

**Absence of transcriptome cis-effect dominance**

A hallmark of eQTL mapping studies using both GWA or structured mapping populations in a wide range of species is the occurrence of large-effect loci that map to the gene itself, i.e. *cis-*eQTL (Brem, Yvert et al. 2002, Schadt, Monks et al. 2003, Monks, Leonardson et al. 2004, Keurentjes, Fu et al. 2007, West, Kim et al. 2007). However, we did not identify a large number of outlier p-values as would be expected if there were numerous large effect *cis*-eQTL. To test if the *B. cinerea* transcriptome shows a similar cis-eQTL pattern, we plotted the position of the transcript’s genomic position against the top GWA SNP for all the *B. cinerea* transcripts. We first focused on the single top SNP hit per transcript, with the highest probability (lowest p-value) of significant effect on expression in the gene of interest. If control of gene expression is localized to the gene itself or toproximate loci, we would expect a strong linear (*cis*-diagonal) association between the center of each gene and the genomic location of its top SNP hit. However, there was no evidence of any cis-diagonal (Figure N2). This pattern holds whether we examine the top 1 SNP per transcript (Figure N2a) or the top 10 SNPs per transcript (FigureN2b). In contrast, there was evidence for *trans*-eQTL hotspots; loci which modulate expression variation across many of the pathogen genes (Figure N2).

To test if there might be a bias towards *cis*-effects that may function at a close distance, we calculated the distance between the center of each transcript and the top associated SNP. If *cis*- acting loci contribute the bulk of genetic control of expression variation, we would expect to see a high frequency of short-distance associations, and a rapid decline to a plateau moving away from the gene of interest. However, we observe that distances between transcript center and top SNP as far as 2 Mb are common (Figure N3). These distances are similar to what would happen if the causal SNPs had no *cis* association and were instead scattered across the genome (Figure SX1). As such, we do not see evidence for *cis*-effect loci overrepresented in the top candidates for control of expression variation. Rather, most of the loci that we can associate with potentially influencing gene expression variation in *B. cinerea* on *A. thaliana* is *trans*-acting.

**Search for *cis*-effects through focus on gene networks with presence-absence polymorphism**

The absence of a dominant *cis*- pattern in the genome-wide transcript-to-SNP associations could be caused by a relative absence of *cis-* variation. Alternatively haplotype heterogeneity or allele frequency may complicate the ability to accurately identify *cis*-polymorphisms {}. To test between these possibilities, we conducted a more focused analysis on three biosynthetic pathways that exist as gene clusters. These biosynthetic pathways contribute to *B. cinerea* virulence and the botcinic acid biosynthetic pathway (13 genes, 55.8 kb), botrydial biosynthetic pathway (7 genes, 26 kb), and a putative cyclic peptide pathway (10 genes, 46.5 kb) (Deighton, Muckenschnabel et al. 2001, Colmenares, Aleu et al. 2002, Porquier, Morgant et al. 2016, Zhang, Corwin et al. 2018). These pathways have known presence-absence polymorphisms and should have *cis*-eQTL but none were detected (Siewers, Viaud et al. 2005, Pinedo, Wang et al. 2008, Zhang, Corwin et al. 2018). Critically, the transcripts within each of these pathways are highly correlated across the isolates, suggesting that their expression variation is controlled by pathway-specific variation (Zhang, Corwin et al. 2018). Thus, these loci may have false-negative issues that prevented the detection of real *cis*-eQTL.

To test if these pathways have undetected cis-eQTL we used all of the SNPs for each biosynthetic cluster to align the *B. cinerea* isolates and investigate haplotype diversity. We first investigated the botcinic acid cluster which identified a number of distinct haplotypes with a few individual outlier isolates (e.g. B05.10, Fd1) (Figure N4a). We then utilized the haplotypes to test for specific effects on transcript expression for the biosynthetic pathway. This identified a single clade of isolates with a distinctly lower level of expression than the other clusters (Figure N4b). Investigating the short-reads and SNP calls showed that these 12 isolates share a 53.5 kb deletion that removes the entire biosynthetic cluster (Figure N4c). After removing the major deletion, we found no remaining significant effect of cluster membership in the remaining 3 major clusters on expression profile (F(1,74)=0.36, p=0.55). However, within each of these clusters there are independent isolates with very low pathway expression, suggesting loss-of-expression polymorphisms (Noble Rot, 01.04.03, Apple 517, 02.04.09) (Figure N4b). While these isolates each contain smaller deletions that are independent of each other, it is not clear what is functionally leading to the loss of botcinic acid biosynthetic pathway expression (Figure N4c). This suggests that for this clustered pathway, there are undetected *cis*-effect polymorphisms, a large common deletion and rarer additional events.

We then investigated the other two biosynthetic pathways for additional evidence of missed *cis*-acting genetic variation. The botrydial biosynthetic network, and the cyclic peptide pathway, exhibit a similar lack of dominant *cis*-effect SNP patterns to the botcinic acid biosynthetic network. Hierarchical clustering within each of these networks by genic SNP variation divides the isolate population into two groups that are not associated with mean pathway expression (Figure S1, Figure S2). However, within the cyclic peptide pathway, minor deletions within the intergenic regions correlate with low expression, and two isolates with partial deletions within the genes early in the pathway exhibit very low pathway expression (1.05.16, 1.05.22) (Figure S4). In contrast, there was no evidence for *cis-*effects although this pathway did not have any obvious loss-of-expression events (Figure S3). As such, we are able to detect cis-acting variation in the form of deletions for two of the biosynthetic pathways. This suggests that there are missing *cis*-effects within the *B. cinerea* GWA that is missed due to SNP data not incorporating structural variation, as well as this structural variation often being below the minor allele cutoffs. Testing whether insertion and deletion events account for the majority of localized control of expression variation would require long-read sequencing to accurately identify these structural variants and computational approaches that can blend SNP and indel information {Wang 2018}.

**Detection and annotation of *trans*-eQTL hotspots**

While *cis*-effects are difficult to identify, there was a strong signature of SNPs that appeared to affect more transcripts than expected by chance (Figure N2). These are considered positions where there is a causal polymorphism that influences the regulation of numerous genes in *trans*, i.e. a *trans*-eQTL hotspot. In this dataset, we can extend this analysis to look for *trans*-eQTL hotspots that extend beyond *B. cinerea* and influence the expression of genes in the host, *A. thaliana*. We queried for hotspots in both the *B. cinerea* and *A. thaliana* transcriptome by using the top SNP per transcript (Figure N6). By permuting the SNP positions, we identified maximum permuted hotspots as SNP associated with 11 *B. cinerea* transcripts or 80 *A. thaliana* transcripts. For further analysis of hotspots, we utilized a conservative threshold of 20 linked transcripts for *B. cinerea* and 150 transcripts for *A. thaliana*. This analysis identified 13 SNPs as potential *trans*-eQTL hotspots for the *B. cinerea* transcriptome and 12 SNPs as potential cross-species *trans*-eQTL influencing *the A. thaliana* transcriptome (Figure N6, Figure N7). The *trans*-eQTL hotspots are spread throughout the genome (Figure N6, Table N1).

The benefit of a co-transcriptome is that it should be possible to map how polymorphisms cause effects in the pathogen and how these effects transmit to an altered transcriptome in the host. This would suggest that a *trans*-eQTL hotspot for *B. cinerea* transcripts may control virulence pathways and thus cause an associated *trans*-eQTL hotspot in the *A. thaliana* response. However, we found no overlap in eQTL hotspots across the two transcriptomes; hotspots targeting *B. cinerea* gene expression linked to 0 to 56 transcripts in *A. thaliana*, and hotspots targeting *A. thaliana* gene expression linked to 0 to 3 *B. cinerea* transcripts. All of these are values that are below the permutation threshold. To test if this is caused by using solely the top SNP per transcript, we repeated the full analysis by selecting the top 10 SNPs per transcript. This again identified a limited number of *trans*-eQTL hotspots with little overlap between the two species’ transcriptomes (Table SX1; Figure SX3). This suggests that the pathogen’s influence on the host’s transcriptome is not solely limited to major interactions between *tran*s-eQTL hotspots but can involve narrower changes in the pathogen that are magnified in the host’s response.

**Annotation of eQTL hotspots and targets**

We annotated these hotspots with gene and functional information, to understand possible mechanisms of expression modulation in the host and pathogen transcriptomes. While we find annotation information suggestive of metabolic interactions between host and pathogen, and known virulence mechanisms, 32% of our hotspot genes and 21% of the downstream targets do not yet have gene ontology (GO) information (Table N1, Table N2). Thus, this study identifies a large number of loci potentially involved in novel virulence mechanisms of *B. cinerea*.

We examined annotations of the genes linked to our eQTL hotspots, to hypothesize mechanisms of regulation by these hotspots. We annotated the genes targeted by these hotspots with functional information, including links to co-expression networks and lesion size variation from previous RNAseq analysis of these transcripts (Zhang, Corwin et al. 2017, Zhang, Corwin et al. 2018) (Table X1). eQTL hotspots linked to these co-expression networks could indicate regulatory points for these modules of expression variation.

The 12 *A. thaliana* hotspots, annotated to 11 genes, included 4 enzymes and 2 genes associated with isolate compatibility (Table N1). From *B. cinerea* expression profiles, the 13 hotspots annotated to 11 genes included 4 enzymes (Table N1). Nine of the *A. thaliana* eQTL hotspots were also linked to genes from two major genotype-dependent *A. thaliana* co-expression networks when infected with *B. cinerea* (Figure N8). These networks contain genes pointing to network function in jasmonate and salicylic acid signaling processes and camalexin biosynthesis (Network I), or photosynthesis in the host (Network IV).

Among the 22 hotspot genes, three of the *B. cinerea* hotspot genes and two of the *A. thaliana* hotspot genes are strongly positively correlated to lesion size variation across all tested *A. thaliana* genotypes, including immune pathway mutants (Table N1) {Zhang 2018}. A fourth *B. cinerea* hotspot gene is consistently identified as a top GWA hit controlling lesion size across host genotypes and association methods (Table N1){Atwell 2018}. Nine of the 11 *B. cinerea* eQTL hotspot genes were linked to genes in one or more of four major *B. cinerea* co-expression networks on *A. thaliana* (Figure N8). In particular, two of these were host-specific networks functionally associated with virulence, with 7 of the 11 *B. cinerea* hotspot genes associated with one of these virulence co-expression networks. As such, we hypothesize that these major points of *B. cinerea* gene expression modulation may also exhibit regulation of virulence strategies on *A. thaliana*.

Many of the *B. cinerea* genes targeted by the hotspots are enzymatic (34% of our list of annotated hotspot target genes) (Table N2). These genes also include 11 transcription factors and two genes associated with virulence.

We observe particularly strong links between 4 of the 11 eQTL hotspots of *B. cinerea* modulation of *A. thaliana* gene expression and immune networks in *A. thaliana* (Figure N8). These frequent links suggest that our eQTL hotspots may exhibit regulatory control over co-expressed modules of genes active in virulence interactions between *B. cinerea* and its host. If these eQTL hotspots are modulating expression of many genes, and affecting lesion size, they may be major *B. cinerea* control points in the plant-pathogen interaction.

We looked for functional overrepresentation among the *A. thaliana* genes targeted by hotspots (Table N1, Table N3). Many of the hotspots target genes active in metabolism, particularly primary metabolism including carbohydrate metabolism, nucleic acid metabolism, and amino acid biosynthesis. Three of the hotspots, including hotspots annotated as two enzymes and a heterokaryon incompatibility locus, show an overrepresentation of photosynthesis-related functions among their targeted genes. Two of the hotspots show many target genes with abiotic stress response functions. Only two of the hotspots are linked to expected plant defense loci, including chitin response and microbe defenses.

**DISCUSSION**

**Dispersed interactions across host and pathogen genomes**

**Haplotype diversity and polygenic genetic modulation of expression**

**Detection of pathogenicity genes and novel loci**

**Connecting from genome to transcriptome to phenotype (future directions)**

**METHODS**

**Experimental design**

We used a previously described collection of *B. cinerea* genotypes that were isolated as single spores from natural infections of fruit and vegetable tissues collected in California and internationally (Atwell, Corwin et al. 2015, Zhang, Corwin et al. 2017, Fordyce, Soltis et al. 2018, Zhang, Corwin et al. 2018). We focused analysis on the *A. thaliana* accession Columbia-0 (Col-0), and all plants were grown as described in a previous study, with 4-fold replication of the full randomized complete block experimental design across two independent experiments (Zhang, Corwin et al. 2017, Zhang, Corwin et al. 2018). The original study included wildtype Col-0 A. thaliana hosts, as well as knockouts to the salicylic acid pathway (*npr1-1*) and to jasmonic acid sensitivity (*coi1-1*). Leaves were harvested 5 weeks after sowing, and inoculated in a detached leaf assay with spores of each of 96 *B. cinerea* isolates (Zhang, Corwin et al. 2017, Zhang, Corwin et al. 2018).

**Expression analysis**

RNASeq libraries were prepared as previously described (Kumar, Ichihashi et al. 2012, Zhang, Corwin et al. 2017, Zhang, Corwin et al. 2018). Briefly, we prepared mRNA from leaves frozen at 16 hours post inoculation, and pooled amplified, size-selected libraries into four replicate groups of 96 barcoded libraries. Sequencing was completed on a single Illumina HiSeq 2500 (San Diego, CA) lane as single 50bp reads at the U.C. Davis Genome Center- DNA Technologies Core (Davis, CA). Individual libraries were then separated by adapter index from fastq files, evaluated for read quality and overrepresentation (FastQC Version 0.11.3, www.bioinformatics.babraham.ac.uk/projects/), and trimmed (fastx, http://hannonlab.cshl.edu/fastx\_toolkit/commandline.html). Reads were aligned to the *A. thaliana* TAIR10.25 cDNA reference genome, followed by the *B. cinerea* B05.10 cDNA reference genome, and we pulled gene counts (Langmead, Trapnell et al. 2009, Li, Handsaker et al. 2009, Van Kan, Stassen et al. 2017, Zhang, Corwin et al. 2017, Zhang, Corwin et al. 2018). We summed counts across gene models, and normalized gene counts as previously described (Zhang, Corwin et al. 2017, Zhang, Corwin et al. 2018).

We used as input the model-adjusted means per transcript from negative binomial linked generalized linear models in previously published studies in the *A. thaliana* transcriptome and *B. cinerea* transcriptome (Zhang, Corwin et al. 2017, Zhang, Corwin et al. 2018). *A. thaliana* and *B. cinerea* transcript phenotypes were from least square means of normalized gene counts in a negative binomial generalized linear model (nbGLM) (Zhang, Corwin et al. 2017, Zhang, Corwin et al. 2018). We calculated linear models from the transcript data including the effects of isolate and host genotype. We z-scaled all transcript profiles prior to GWA.

**Genome wide association**

For GEMMA mapping, we used 95 isolates with a total of 237,878 SNPs against the *B. cinerea* B05.10 genome (Atwell, Corwin et al. 2018). We used haploid binary SNP calls with MAF > 0.20 and <20% missingness. We ran GEMMA once per phenotype, across 9,267 *B cinerea* gene expression profiles and 23,947 *A. thaliana* gene expression profiles.

**Genome wide association of permuted phenotypes**

To validate SNPs as significantly associated with transcript variation, we performed a comparative analysis of randomized phenotypes. Taking each transcriptional profile, we randomized the assignment of phenotypes across the 96-isolate collection. This analysis includes 9,267 randomized *B. cinerea* phenotypes and 23,947 randomized *A. thaliana* phenotypes, one from each measured expression profile. We repeated this randomization in a 5x permutation. We ran GEMMA on each of these permutations, and plotted SNP p-value vs. position (Figure N5). To threshold our individual expression profile GEMMA outputs, we considered p-values below the average 5% permutation threshold as significant; p < 1.96e-05 for *B. cinerea* and p < 2.90e-05 for *A. thaliana*. Permutation approaches are often more effective than p-value thresholding for determining significance across GWA studies with many phenotypes (Evans and Cardon 2006).

**Defining significant hotspots**

We plotted the number of transcripts linked to each SNP, summed across all 5 permutations, to calculate permuted hotspot size. For any SNPs that linked to permuted hotspots of over 5 transcripts in *B. cinerea* or 10 transcripts in *A. thaliana*, we removed these SNPs from downstream analysis as likely false positives. The maximum hotspot size across any of the 5 permutations was 11 genes in *B. cinerea* and 80 genes in *A. thaliana*. We then conservatively defined significant hotspots as SNP peaks exceeding 20 transcripts in *B. cinerea* and 150 transcripts in *A. thaliana*. We further annotated hotspot SNPs to the nearest gene within a 2kb window. The average LD decay in the B. cinerea genome is < 1kb, so we can be relatively confident of SNPs tagging particular genes at the hotspot peaks (Atwell, Corwin et al. 2018). Three genes are annotated to pairs of neighboring hotSNPs, the rest are unique genes. Two genes on chromosome 12 denoting hotSNPs from *A. thaliana* gene expression appear closely linked; in fact, they are separated by ~80kb on the *B. cinerea* genome.

**Annotation of gene ontology and network membership**

*A. thaliana* co-expression analysis identified 131 genes across four major networks (Zhang, Corwin et al. 2017). Network architecture varied by plant host, but a constitutive core was conserved across *A. thaliana* genotypes. We compared our eQTL hotspots (both the gene at eQTL hotspot SNP and all associated transcript profiles) to the largest *A. thaliana* network lists (*npr1-1* background) to estimate all possible regulatory ties. We identified gene overlap with two of the major networks; Network I, camalexin biosynthesis; Network IV, chloroplast function.

*B. cinerea* co-expression analysis identified ten major co-expression networks containing 5 to 242 genes (Zhang, Corwin et al. 2018). We identified gene overlap with four of these networks, including one likely involved in fungal vesicle virulence processes including growth and toxin secretion (vesicle/ virulence), one involved in translation and protein synthesis (translation/ growth). These networks maintained a consistent core across the 3 *A. thaliana* host genotypes, but linkages varied; as such we compared our gene lists with the networks across all 12 hosts, and included both host-dependent and host-independent annotations of our hotspots.

**Pathway focus**

We focused further *cis*-effects analysis on three networks which were highly conserved across *B. cinerea* isolates (Zhang, Corwin et al. 2018). We clustered isolates by SNP data within focal networks. Hierarchical clustering was computed using the R package pvclust based on mean linkage (UPGMA), with correlation distance and 1000 bootstrap replications (Suzuki and Shimodaira 2015). AU p-values are reported in red, BP values in green. Edges with high AU values are considered strongly supported by the data, and clustering is drawn according to these edges with AU > 95%.

For botcinic acid biosynthesis, the major deletion extends 53.5 kb and includes SNP 4kb from the 5’ end of the chromosome, indicating a teleomeric loss on chromosome 1. We selected a focal region encompassing the deletion endpoints (1.4029, 1.82614) and an additional 2 genes beyond the deletion boundaries (Bcin01g00170, Bcin01g00190) (Figure N4c). We removed 10 SNPs that were likely miscalled (SNP state ~ inverse compared to surrounding region) and called all SNPs within the deletion region as missing.

**TABLE LEGENDS**

**Table N1. Annotation of the hotSNPs identified from *B. cinerea* and *A. thaliana* eQTL.**

**Table N2. Annotation of the *B. cinerea* genetic targets of *B. cinerea* hotSNPs.**

**Table N3. Annotation of the *A. thaliana* genetic targets of *B. cinerea* hotSNPs.**

**FIGURE LEGENDS**

**Figure N1. Manhattan plot examples for 1 transcript per species.** Panel a is an example plot of p-values for all *B. cinerea* SNP associations to a single *B. cinerea* transcript, from Bcin01g00170. Panel b is an example plot of p-values for all *B. cinerea* SNP associations to a single *A. thaliana* transcript.

**Figure N2. *cis*-diagonal plot comparing *B. cinerea* gene center to position of top associated SNP, for all 9,284 transcripts.** We retained only the SNPs with highest probability (lowest p-value) of significant effect on expression for each transcript. Panel a depicts the single top SNP per transcript. Panel b depicts the top 10 SNPs per transcript. Chromosomes are delimited by red bars along the x-axis. Vertical striping of SNP positions indicate genomic locations of putative *trans*-eQTL hotspots.

**Figure N3. Distance between transcript center and top SNP location for all *B. cinerea* expression profiles on Col-0 *A. thaliana*.** Data include the top 1 SNP identified by GEMMA association with each transcript expression profile (lowest p-value for association). Distances are in Mb, including only top SNPs on the same chromosome as the focal gene.

**Figure N4. *cis*-effect analysis of the botcinic acid biosynthetic gene network.** Panel a is hierarchical clustering of *B. cinerea* isolates from SNPs within the botcinic acid biosynthetic gene network. Clustering was based on mean linkage (UPGMA), with correlation distance and 1000 bootstrap replications. AU p-values are reported in red, BP values in green. Edges with high AU values are considered strongly supported by the data, and clustering is drawn according to these edges with AU > 95%. Panel b is Violin plots of botcinic acid network-level expression within *B. cinerea* clusters. Isolates are clustered based membership in groups defined by hierarchical clustering of the SNPs within the botcinic acid biosynthesis network (Figure X5). Panel c is the gene models of the biosynthetic gene network, with the cluster 3 deletion indicated as a triangle.

**Figure N5. Manhattan-type plot of GEMMA results of transcriptome-wide *B. cinerea* expression phenotypes.** Panel a is a Manhattan-type plot of the top 1 SNP hit per *B. cinerea* transcript on Col-0 *A. thaliana*. Panel b is a Manhattan-type plot of the top 1 SNP hit per *A. thaliana* transcript when infected by *B. cinerea*.

**Figure N6. All eQTL hotspots across the *B. cinerea* and *A. thaliana* transcriptomes.** We counted the number of genes (transcripts) associated with each SNP. Panel a is for all *B. cinerea* transcripts, panel b is all *A. thaliana* transcripts.

**Figure N7. Interspecific hotspot comparison on the *B. cinerea* genome.** For each SNP that is a top hit for one or more transcripts, the number of associated transcripts is counted, across both the *B. cinerea* transcriptome and the *A. thaliana* transcriptome.

**Figure N8. Genes linked to eQTL hotspots are in virulence and defense co-expression networks.** Circles along the *B. cinerea* genome map are eQTL hotspots, centered at the gene containing the eQTL and with radius proportional to the number of transcripts linked to this hotspot. The gene center is marked with a white dot. Hotspots for *B. cinerea* transcripts are drawn in blue, hotspots for *A. thaliana* transcripts are drawn in green. The *A. thaliana* networks depicted are the most inclusive of the host-dependent networks, from *npr1-1*. Links between hotspots and co-expression networks are drawn according to the number of genes shared between them. Variable line weight represents the percent of hotspot target genes shared with the co-expression network; 1-25% is dashed, 25-50% is dotted, 50-75% is solid, 75-100% is heavy solid.

REFERENCES

Atwell, S., J. Corwin, N. Soltis and D. Kliebenstein (2018). "Resequencing and association mapping of the generalist pathogen Botrytis cinerea." bioRxiv.

Atwell, S., J. Corwin, N. Soltis, A. Subedy, K. Denby and D. J. Kliebenstein (2015). "Whole genome resequencing of Botrytis cinerea isolates identifies high levels of standing diversity." Frontiers in microbiology **6**: 996.

Barrett, L. G., J. M. Kniskern, N. Bodenhausen, W. Zhang and J. Bergelson (2009). "Continua of specificity and virulence in plant host–pathogen interactions: causes and consequences." New Phytologist **183**(3): 513-529.

Bartha, I., P. J. McLaren, C. Brumme, R. Harrigan, A. Telenti and J. Fellay (2017). "Estimating the respective contributions of human and viral genetic variation to HIV control." PLoS computational biology **13**(2): e1005339.

Bartoli, C. and F. Roux (2017). "Genome-Wide Association Studies In Plant Pathosystems: Toward an Ecological Genomics Approach." Frontiers in plant science **8**.

Brem, R. B., G. Yvert, R. Clinton and L. Kruglyak (2002). "Genetic dissection of transcriptional regulation in budding yeast." Science **296**(5568): 752-755.

Chen, X., C. A. Hackett, R. E. Niks, P. E. Hedley, C. Booth, A. Druka, T. C. Marcel, A. Vels, M. Bayer and I. Milne (2010). "An eQTL analysis of partial resistance to Puccinia hordei in barley." PLoS One **5**(1): e8598.

Christie, N., A. A. Myburg, F. Joubert, S. L. Murray, M. Carstens, Y. C. Lin, J. Meyer, B. G. Crampton, S. A. Christensen and J. F. Ntuli (2017). "Systems genetics reveals a transcriptional network associated with susceptibility in the maize–grey leaf spot pathosystem." The Plant Journal **89**(4): 746-763.

Colmenares, A. J., J. Aleu, R. Duran-Patron, I. G. Collado and R. Hernandez-Galan (2002). "The putative role of botrydial and related metabolites in the infection mechanism of Botrytis cinerea." Journal of chemical ecology **28**(5): 997-1005.

Corwin, J. A., D. Copeland, J. Feusier, A. Subedy, R. Eshbaugh, C. Palmer, J. Maloof and D. J. Kliebenstein (2016). "The quantitative basis of the Arabidopsis innate immune system to endemic pathogens depends on pathogen genetics." PLoS Genet **12**(2): e1005789.

Deighton, N., I. Muckenschnabel, A. J. Colmenares, I. G. Collado and B. Williamson (2001). "Botrydial is produced in plant tissues infected by Botrytis cinerea." Phytochemistry **57**(5): 689-692.

Denby, K. J., P. Kumar and D. J. Kliebenstein (2004). "Identification of Botrytis cinerea susceptibility loci in Arabidopsis thaliana." The Plant Journal **38**(3): 473-486.

Evans, D. M. and L. R. Cardon (2006). "Genome-wide association: a promising start to a long race." Trends in Genetics **22**(7): 350-354.

Fordyce, R., N. Soltis, C. Caseys, G. Gwinner, J. Corwin, S. Atwell, D. Copeland, J. Feusier, A. Subedy, R. Eshbaugh and D. Kliebenstein (2018). "Combining Digital Imaging and GWA Mapping to Dissect Visual Traits in Plant/Pathogen Interactions." Plant Physiology.

Fordyce, R. F., N. E. Soltis, C. Caseys, R. Gwinner, J. A. Corwin, S. Atwell, D. Copeland, J. Feusier, A. Subedy and R. Eshbaugh (2018). "Digital Imaging Combined with Genome-Wide Association Mapping Links Loci to Plant-Pathogen Interaction Traits." Plant physiology **178**(3): 1406-1422.

Glazebrook, J. (2005). "Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens." Annu. Rev. Phytopathol. **43**: 205-227.

Goss, E. M. and J. Bergelson (2006). "Variation in resistance and virulence in the interaction between Arabidopsis thaliana and a bacterial pathogen." Evolution **60**(8): 1562-1573.

Guo, Y., S. Fudali, J. Gimeno, P. DiGennaro, S. Chang, V. M. Williamson, D. M. Bird and D. M. Nielsen (2017). "Networks underpinning symbiosis revealed through cross-species eQTL mapping." Genetics: genetics. 117.202531.

Keurentjes, J. J., J. Fu, I. R. Terpstra, J. M. Garcia, G. van den Ackerveken, L. B. Snoek, A. J. Peeters, D. Vreugdenhil, M. Koornneef and R. C. Jansen (2007). "Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci." Proceedings of the National Academy of Sciences **104**(5): 1708-1713.

Kumar, R., Y. Ichihashi, S. Kimura, D. H. Chitwood, L. R. Headland, J. Peng, J. N. Maloof and N. R. Sinha (2012). "A high-throughput method for Illumina RNA-Seq library preparation." Frontiers in plant science **3**.

Langmead, B., C. Trapnell, M. Pop and S. L. Salzberg (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." Genome biology **10**(3): R25.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis and R. Durbin (2009). "The sequence alignment/map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Monks, S., A. Leonardson, H. Zhu, P. Cundiff, P. Pietrusiak, S. Edwards, J. Phillips, A. Sachs and E. Schadt (2004). "Genetic inheritance of gene expression in human cell lines." The American Journal of Human Genetics **75**(6): 1094-1105.

Nomura, K., M. Melotto and S.-Y. He (2005). "Suppression of host defense in compatible plant–Pseudomonas syringae interactions." Current opinion in plant biology **8**(4): 361-368.

Pinedo, C., C.-M. Wang, J.-M. Pradier, B. Dalmais, M. Choquer, P. Le Pêcheur, G. Morgant, I. G. Collado, D. E. Cane and M. Viaud (2008). "Sesquiterpene synthase from the botrydial biosynthetic gene cluster of the phytopathogen Botrytis cinerea." ACS chemical biology **3**(12): 791-801.

Porquier, A., G. Morgant, J. Moraga, B. Dalmais, I. Luyten, A. Simon, J.-M. Pradier, J. Amselem, I. G. Collado and M. Viaud (2016). "The botrydial biosynthetic gene cluster of Botrytis cinerea displays a bipartite genomic structure and is positively regulated by the putative Zn (II) 2Cys6 transcription factor BcBot6." Fungal genetics and biology **96**: 33-46.

Rowe, H. C. and D. J. Kliebenstein (2008). "Complex genetics control natural variation in Arabidopsis thaliana resistance to Botrytis cinerea." Genetics **180**(4): 2237-2250.

Rowe, H. C., J. W. Walley, J. Corwin, E. K.-F. Chan, K. Dehesh and D. J. Kliebenstein (2010). "Deficiencies in jasmonate-mediated plant defense reveal quantitative variation in Botrytis cinerea pathogenesis." PLoS Pathog **6**(4): e1000861.

Schadt, E. E., S. A. Monks, T. A. Drake, A. J. Lusis, N. Che, V. Colinayo, T. G. Ruff, S. B. Milligan, J. R. Lamb and G. Cavet (2003). "Genetics of gene expression surveyed in maize, mouse and man." Nature **422**(6929): 297.

Siewers, V., M. Viaud, D. Jimenez-Teja, I. G. Collado, C. S. Gronover, J.-M. Pradier, B. Tudzynsk and P. Tudzynski (2005). "Functional analysis of the cytochrome P450 monooxygenase gene bcbot1 of Botrytis cinerea indicates that botrydial is a strain-specific virulence factor." Molecular plant-microbe interactions **18**(6): 602-612.

Soltis, N. E., S. Atwell, G. Shi, R. F. Fordyce, R. Gwinner, D. Gao, A. Shafi and D. J. Kliebenstein (2019). "Interactions of tomato and Botrytis genetic diversity: Parsing the contributions of host differentiation, domestication and pathogen variation." The Plant Cell: tpc. 00857.02018.

Suzuki, R. and H. Shimodaira (2015). "pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling. ." R package version 2.0-0.

Van Kan, J. A., J. H. Stassen, A. Mosbach, T. A. Van Der Lee, L. Faino, A. D. Farmer, D. G. Papasotiriou, S. Zhou, M. F. Seidl and E. Cottam (2017). "A gapless genome sequence of the fungus Botrytis cinerea." Molecular plant pathology **18**(1): 75-89.

West, M. A., K. Kim, D. J. Kliebenstein, H. Van Leeuwen, R. W. Michelmore, R. Doerge and D. A. S. Clair (2007). "Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis." Genetics **175**(3): 1441-1450.

Williamson, B., B. Tudzynski, P. Tudzynski and J. A. L. van Kan (2007). "Botrytis cinerea: the cause of grey mould disease." Molecular Plant Pathology **8**(5): 561-580.

Wu, J., B. Cai, W. Sun, R. Huang, X. Liu, M. Lin, S. Pattaradilokrat, S. Martin, Y. Qi and S. C. Nair (2015). "Genome-wide analysis of host-Plasmodium yoelii interactions reveals regulators of the type I interferon response." Cell reports **12**(4): 661-672.

Wu, J. Q., S. Sakthikumar, C. Dong, P. Zhang, C. A. Cuomo and R. F. Park (2017). "Comparative genomics integrated with association analysis identifies candidate effector genes corresponding to Lr20 in phenotype-paired Puccinia triticina isolates from Australia." Frontiers in plant science **8**.

Zhang, W., J. A. Corwin, D. Copeland, J. Feusier, R. Eshbaugh, F. Chen, S. Atwell and D. J. Kliebenstein (2017). "Plastic transcriptomes stabilize immunity to pathogen diversity: the jasmonic acid and salicylic acid networks within the Arabidopsis/Botrytis pathosystem." The Plant Cell: tpc. 00348.02017.

Zhang, W., J. A. Corwin, D. Copeland, J. Feusier, R. Eshbaugh, D. E. Cook, S. Atwell and D. J. Kliebenstein (2018). "Network connections across kingdoms illuminate a potential metabolic battlefield." bioRxiv.

Zhou, X. and M. Stephens (2012). "Genome-wide efficient mixed-model analysis for association studies." Nature genetics **44**(7): 821.