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**

Plant-pathogen interactions can be classified in two groups; qualitative interactions, in which few genetic variants interact to determine binary disease outcomes, or quantitative, in which a spectrum of interactions may occur due to genetic variation between the host and pathogen. The genetic basis of quantitative plant-pathogen interactions is less understood and is being explored through the study of phenotypic variation across genetically diverse hosts and pathogens, finding links to genetic variation. Interactions between plants and generalist pathogens, which do not specialize on particular hosts, are more often quantitative due to a lack of reciprocal co-evolution, and generalist pathogens often harbor higher genetic diversity (Williamson, Tudzynski et al. 2007).

Within a generalist pathogen, there are several theories for how one species can adapt to a broad host range. Perhaps the pathogen has evolved a singular resistance gene of broad effect that disables core defense pathways across diverse plant host species; a sort of silver bullet. This mechanism would predict low diversity in the pathogen as the silver bullet gene is fixed in the population. However, we see very high diversity in *B. cinerea* and no evidence of a single gene with high linkage to virulence, so we can eliminate this hypothesis. A second prediction for an extreme generalist species is specialization at the level of the individual. This would predict high population structure and moderate diversity. However, we observe low population in *B. cinerea*, high diversity, and a lack of evidence for individual specialization to hosts. Our final prediction is specialization at the gene or allele level, which would select for very high diversity and low population structure as the different genetic strategies are intermixed within individuals. This is consistent with the high SNP diversity and low population structure observed in previous studies of *B. cinerea*, due to a combination of random mating and frequent recombination {Williamson 2007; Rowe 2010; Atwell 2015; Corwin 2016; Zhang 2016}.

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). X to X loci are implicated with relatively small effect sizes, and these genes have varied functions including X.

On the pathogen side, recent studies have accumulated evidence for a polygenic basis of virulence as well (Corwin, Copeland et al. 2016, 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.

Many measurable phenotypes result from the interaction of plant and pathogen, 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 such as genome-wide association (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 linked (*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.

Studies encompassing transcriptomic variation in both a 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). A small number of previous studies have identified cross-species trans-eQTL, as a way to hypothesize causal relationships between individual genes in the interspecific interaction (Wu, Cai et al. 2015, Guo, Fudali et al. 2017). eQTL identified through this approach include one locus in a plant pathogenic nematode which modulates expression of >60 genes in its host *Medicago truncatula* (Guo, Fudali et al. 2017). Further, in *M. truncatula* a total of 213 genes in the host were linked to one or more loci in the pathogen, and functionally enriched for transcription factors as well as defense-related enzymes and enzymes involved in essential amino acid biosynthesis (Guo, Fudali et al. 2017). However, this study works with the limited natural variation of 98 F2 progeny from recombinant inbred line (RIL) between two geographically distant parents, of which only one can reproduce on common bean (Guo, Fudali et al. 2017).

In mouse, 1054 host genes were linked to one or more loci in the pathogen, across a total of 208 pathogen eQTL (Wu, Cai et al. 2015). These host genes were enriched for ATP response, metabolic functions, and antimicrobial and inflammatory immune responses, and genes from the same host network often shared the same eQTL (Wu, Cai et al. 2015). Many of the host genes were linked to multiple parasite eQTL (Wu, Cai et al. 2015). They were able to validate 14 out of 15 host genes selected based on predicted immune activity (Wu, Cai et al. 2015). However, this study works with the limited natural variation of 24 progeny from a RIL between two parental lines from distinct subspecies of *Plasmodium* (Wu, Cai et al. 2015).

On the host side, similar methods can identify human host genetic polymorphisms affecting bacterial parasite gene expression; three bacterial genes were regulated by these identified host eQTL (Guo, Fudali et al. 2017).

For this work, 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 virulenceThese interactions are well-characterized phenotypically, and we have previous information on some of the potentially relevant genetic factors on both the pathogen and host side. This also gives us the opportunity to connect our findings, particularly in plant genetic targets and affected pathways, to many previous datasets.

Previous studies in the *A. thaliana* - *B. cinerea* pathosystem point to control of expression variation on the host side 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.

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.

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.

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.

**RESULTS**

**eQTL indicate polygenic transcriptome modulation**

We performed genome-wide association (GWA) for eQTL detection across all genes expressed in the *B. cinerea* - *A. thaliana* pathosystem, including transcripts from the host and from the pathogen. To perform genome-wide association (GWA) and identify expression quantitative trait loci (eQTL) for both host and pathogen, we associated the expression profiles of these 9,267 *B. cinerea* genes and 23,947 *A. thaliana* genes to genome-wide SNP variation in the *B. cinerea* genome. We used a previous genome-wide SNP dataset that has 237,878 SNPs with a minimum minor allele frequency of 0.20 {Atwell 2018}. We performed Genome-wide Efficient Mixed Model Association (GEMMA) (Zhou and Stephens 2012) to rapidly estimate the significance of all markers for each expression profile. We first controlled for the effects of population structure within our *B. cinerea* isolates by calculating and including a relatedness matrix in the downstream analysis. GEMMA estimates the significance of effects of each SNP on the focal trait as a p-value. To determine significance of SNP effects across tens of thousands of traits (individual expression profiles in *B. cinerea* and *A. thaliana*), we permuted each trait across the 96 B. cinerea isolates five times, and repeated GEMMA analysis for each. We calculated the minimum p-value per SNP per transcript across these five permutations and used this to threshold our data. However, we found this approach overly conservative due to the magnitude of variation from one trait to another. Instead, we considered p-values below the 5% permutation threshold as significant. For individual gene expression traits, we find 0 to XX loci with significant p-values for *B. cinerea*, and 0 to XX loci for *A. thaliana* transcripts (Figure N1). In total, genetic variation in *B. cinerea* appeared to significantly affect the expression of 5,213 *A. thaliana* genes and 1,616 *B. cinerea* genes.

In looking at only the top 1 SNP per trait, we find that in B. cinerea 69% of the genes show a lower p-value of SNP-trait association from the observed data than the maximum across all 5 permutations.

Overall, we find a highly polygenic basis of loci modulating transcriptome variation.

**Absence of transcriptome cis-effect dominance**

Previous eQTL mapping studies using both GWA or structured mapping populations in a wide range of species show a dominance of loci that map to the gene itself, i.e. *cis-*eQTL {CITE}. To test if the *B. cinerea* transcriptome shows a similar cis-eQTL predominance, we looked for a *cis*-diagonal signature of associations between transcript center and GWA SNP hits. 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, we find that few genes have a top SNP hit within the same chromosome, and even fewer within 1Mb (Figure N2). We conclude that most of this genetic variation is *trans*-acting, as we do not see a strong *cis*-diagonal signal when comparing transcript center to top SNP hit. This pattern holds whether we examine the top 1 SNP per transcript (Figure N2a) or the top 10 SNPs per transcript (FigureN2b). Further, we do visualize vertical stripes of SNPs, indicative of *trans*-eQTL hotspots; loci which modulate expression variation across many of the pathogen genes (Figure N2).

In these genome-wide transcript-to-SNP associations, we find that *cis*- control of gene expression is largely drowned out by patterns of *trans*-acting variation. To further search for a signature of dominant cis-effects, 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 those from the association of random transcript profiles to top SNPs (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 control of gene expression variation in *B. cinerea* on *A. thaliana* appears to be *trans*-acting.

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

The absence of a dominant signal for *cis*-eQTL could arise from *B. cinerea* having a different pattern of causal eQTL variation from our expectation. Alternatively, this could arise from *B. cinerea* having a sufficiently high genetic diversity that leads to false-negatives due to complex haplotype structures at causal loci {CITE}. To test between these possibilities, we narrowed our focus to a set of three biosynthetic pathways that exist as gene clusters and have known presence-absence polymorphisms. to identify the role of *cis*-eQTL. These biosynthetic pathways contribute to *B. cinerea* virulence and the botcynic 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) (Zhang, Corwin et al. 2018). Critically, the transcripts within each of these pathways are highly correlated across the isolates suggesting that their genetic variation is controlled by pathway-specific variation {Zhang 2018}. None of the genes showed a *cis*-eQTL within the GWA suggesting that the identified causal variation is solely in *trans* to the pathways.

To test if this result may be complicated by the architecture of the pathway loci, we initially focused on the botcynic acid biosynthesis network, which is known to have presence-absence polymorphisms in the species. We obtained all of the SNPs for this biosynthetic cluster and conducted an alignment of the *B. cinerea* isolates. This showed evidence of a number of distinct haplotypes with some relatively isolated isolates (e.g. B05.10, Fd1) (Figure N4a). Looking at the expression of the transcripts for the biosynthetic pathway showed a single clade that had a distinctly lower level of expression than the other clusters. Using SNP-level variation surrounding genes in this pathway showed that these 12 isolates share a 53.5 kb deletion that removes the entire biosynthetic cluster(Figure N4c). We then tested if the remaining clusters had any link to transcript expression by performing ANOVA across the 3 major clusters (2, 5, 6). After removing the major deletion, we found no remaining significant effect of cluster membership on expression profile (F(1,74)=0.36, p=0.55). Although within each of these clusters there are independent isolates with low pathway expression polymorphisms (Noble Rot, 01.04.03, Apple 517, 02.04.09) (Figure N4b), these isolates each contain smaller independent deletions that likely abolish expression of the botcynic acid biosynthetic pathway (Figure N4c). This suggests that for this locus, there is a major *cis*-effect deletion and a number of rarer additional deletions that control expression variation. However, this *cis*-effect is not captured by the SNP-level data utilized for the GWA. Rather, our GWA analysis misses the major *cis*-eQTL signal of the network deletion. If insertion and deletion events account for the majority of localized control of expression variation, our GWA analysis will not detect these *cis*-effect loci. Thus, we investigated the other two biosynthetic pathways for additional evidence of *cis*-acting genetic variation.

The botrydial biosynthetic network, and the cyclic peptide pathway, exhibit similar cis-effect patterns to the botcynic acid biosynthetic network. Hierarchical clustering within each of these networks by genic SNP variation divides the isolate population into two groups, and mean pathway expression across all isolates is not differentiated by this clustering for either network (Figure S1, Figure S2). We find that SNP state does not detect the major cis-effects polymorphisms… We conclude that large cis effect loci exist for transcriptional regulation in this pathosystem, but that GWA across SNPs misses these patterns. This missing cis-effect likely amplifies the apparent magnitude of trans-acting loci.

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

Our 11 significant *trans­-*eQTL hotspots for *B. cinerea* range from 22 to 129 linked genes, with minimal overlap to expression modulation in *A. thaliana* (a maximum of 56 genes). These hotspots are dispersed across the genome, at least 0.1 Mb apart and across 9 chromosomes (Figure N7). The 11 significant cross-species *trans*-eQTL hotspots for *A. thaliana* are also dispersed across the genome of *B. cinerea*, covering 8 chromosomes with at least 0.1 Mb between hotspots (Figure N7). These range from 114 to 634 linked *A. thaliana* transcripts, with very low overlap with *B. cinerea* transcripts (a maximum of 3 genes).

The other typical predominant pattern in eQTL studies is the presence of hotspots, whereby variation in numerous transcripts links to specific loci. 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*. To conduct a conservative search and to simplify the analysis, we focused on solely the single most-significant SNP (i.e. strongest evidence) that is linked to a given transcript. Taking these top SNPs for all *the B. cinerea* and *A. thaliana* transcripts, we queried for hotspots per transcriptome (Figure N6). By permuting the SNP positions, we identified maximum permuted hotspot sizes for *B. cinerea* as 11 genes, and 80 genes for *A. thaliana*. 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 hotSNPs influencing *the A. thaliana* transcriptome (Figure N6, Figure N7). The *trans*-eQTL hotSNPs are spread throughout the genome, present on all chromosomes except 7, 11, 14, 15 (Figure N6, Table N1).

One possibility is that a *trans*-eQTL hotSNP for *B. cinerea* transcripts may control virulence pathways and thus cause an associated *trans*-eQTL hotSNP in the *A. thaliana* response. However, we found no significant overlap in eQTL hotspots across the two genomes; hotSNPs targeting *B. cinerea* gene expression linked to 0 to 56 transcripts in *A. thaliana*, and hotSNPs targeting *A. thaliana* gene expression linked to 0 to 3 *B. cinerea* transcripts. To test the dependency of these results on using solely the top SNP, we repeated the full analysis by selecting the top 10 SNPs per transcript. This again identified a limited number of *trans*-eQTL hotSNPs 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 hotSNPs but can involve more limited changes in the pathogen that are magnified in the host’s response.

**Annotation of hotSNPs and eQTL hotspot targets**

We annotated these hotSNPs with gene and functional information, to understand possible mechanisms of expression modulation in the host and pathogen transcriptomes. From *A. thaliana* expression profiles, the 11 hotSNP gene annotations included 4 enzymes and 2 genes associated with isolate compatibility (Table N1). From *B. cinerea* expression profiles, the 11 hotSNP gene annotations included 4 enzymes (Table N1).

We also annotated the genes targeted by these hotSNPs with functional information, including links to co-expression networks 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.

Among the 11 *B. cinerea* hotspot *trans-*eQTL, three of these hotspot genes are linked to lesion size variation across all tested *A. thaliana* genotypes. A fourth gene is linked to lesion size variation on Col-0 *A. thaliana*. This gene is also linked to the major vesicle/ virulence network of *B. cinerea* co-expression on Col-0 *A. thaliana* (Zhang, Corwin et al. 2018).

Nine of the 11 *B. cinerea* eQTL hotspots were also 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* eQTL hotspots 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*.

Among the 11 *A. thaliana* hotspot trans-eQTL, two of these genes are correlated to *B. cinerea* lesion size across all tested *A. thaliana* genotypes, including immune pathway mutants. An additional gene is correlated with lesion size variation on Col-0 *A. thaliana*, and one on *coi1-1* *A. thaliana.*

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). 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 are modulating expression of many genes, and affecting lesion size, they may be major *B. cinerea* control points in the plant-pathogen interaction.

We further examined functional annotation of the genes linked to our eQTL hotspots, to hypothesize mechanisms of regulation by these hotSNPs.

Many of the *B. cinerea* genes targeted by the hotSNPs are enzymatic (34% of our list of annotated hotSNP target genes, 140/412) (Table N2). These genes also include 11 transcription factors and two genes associated with virulence.

**DISCUSSION**

**Dominance of *trans*-eQTL over *cis*-eQTL patterns**

**Cross-species *trans*-eQTL patterns**

**Potential mechanisms of eQTL hotSNPs**

The 12 *A. thaliana* hotSNPs were annotated to 11 gene functions (Table N1). Among these, 4 were enzymatic, including a glucose/ ribitol dehydrogenase and a glycoside hydrolase. These *B. cinerea* enzymes may alter pathogen metabolism to elicit host responses, detected here as transcriptional regulation. Alternately, a more direct effect is possible if any of these enzymes are secreted, and function in the digestion of host polysaccharides or other metabolites. In fact, one A. thaliana hotSNP is annotated to a secreted glycoside hydrolase, which may directly interact with the host metabolism. Either mechanism would likely stimulate major host responses and thus an expression response (Bcin16g01950, glycoside hydrolase, family 63). Among the 13 *B. cinerea* hotSNPs, 4 were annotated to *B. cinerea* enzymes (Table N1). Further, the targets of these hotSNPs are often enzymes (Table N2). These hotSNP enzymes may alter major branches of the *B. cinerea* metabolic pathways active during the infection of *A. thaliana*.

Two of the B. cinerea hotSNPs may have direct effects on the transcription machinery (Bcin12g00330, Topoisomerase II-associated protein PAT1; Bcin09g06590, Helicase)(Table N1). Alternately, these genes may affect the number of nuclei per *B. cinerea* mycelial cell, potentially altering the virulence of the pathogen.

Two of the *A. thaliana* hotSNPs are annotated to genes that have been previously shown to predict isolate crossing compatibility and mating type {Atwell 2016}. These loci may structure some of the pathogen diversity, and as such may be major vectors of pathogen variation affecting host response.

From the current analysis and limited annotation information in *B. cinerea*, we cannot distinguish between direct and indirect effects of *B. cinerea* hotSNPs on the *A. thaliana* transcriptome.

**Annotation of genes targeted by eQTL hotSNPs**

The frequency of enzymes in the *B. cinerea* eQTL hotspot target genes suggests a role of *B. cinerea* metabolic shifts as the fungal infection progresses *in planta*.

The presence of many transcription factors among the hotspot targets indicates that the eQTL hotspots may control major networks of *B. cinerea* gene expression variation. The hotSNPs may affect variation in a hierarchical manner, with the targets of the eQTL also regulatory genes modulating the expression of genes downstream. The eQTL hotspots may be near the top of a hierarchical gene regulatory structure.

Two of the *B. cinerea* genes downstream of eQTL hotspots peak in their expression during the outset of colonization (BcPIO5, BcPIO10) {Gioti 2006}. An additional gene is a homeobox transcription factor induced early in infection (BcPIE1, BcHOX8). We detected expression of these genes at a later intermediate infection stage in our experiment, so these genes may play an ongoing role in the progression of infection or remain active at the newly colonized cellular edge of a growing necrotic lesion.

Two of the *B. cinerea* genes are mitogen-activated protein kinases, active in signaling cascades within the fungus to regulate major cellular pathways. These may point to pathways of signal transduction and decision-making as infection of the host progresses.

Looking at the genes targeted by a single hotSNP can reveal potential mechanistic patterns as well. One hotSNP from the *A. thaliana* transcriptome (Bcin16g01950, Bccwh41, glycoside hydrolase, family 63) targets many photosynthesis-related genes (Table N3), as well as members of multiple pathogen defense pathways. These patterns are further reinforced as this hotspot contains several genes from the major photosynthetic co-expression network and the JA/ SA/ Camalexin defense co-expression network (Figure N8).

The two hotSNPs identified on chromosome 12 from the *B. cinerea* transcriptome each target many enzymes, and one targets three genes expressed in planta during infection (Table N2). Both hotSNPs are annotated to the same gene (Bcin12g00330, Bcpat1, Topoisomerase II-associated protein PAT1), whose expression is highly correlated with *B. cinerea* lesion size; this locus may be a control point for major *B. cinerea* metabolic and expression shifts, potentially feeding into virulence strategies.

**Triangle of interaction: genotype to expression to phenotype**

Previous work in the B. cinerea – A. thaliana pathosystem established connections between host polymorphisms and lesion growth, between gene expression and lesion size, and between transcriptomes of the host and pathogen {Corwin 2016; Fordyce 2017; Zhang 2017; Zhang 2018}. To begin establishing causal inference from genome to phenotype, the results of this work fill the gap of connecting genetic variation in the pathogen to expression changes in the interacting transcriptomes.

**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 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 {CITE}.

**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 further analysis as likely false positives. We then conservatively defined actual hotspots as SNP peaks exceeding 20 transcripts in *B. cinerea* and 150 transcripts in *A. thaliana*. We then collapsed hotspots into genes, such that all SNPs were annotated 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 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 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.

**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 botcynic 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 botcynic acid biosynthetic gene network.** Panel a is hierarchical clustering of *B. cinerea* isolates from SNPs within the botcynic 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 botcynic acid network-level expression within *B. cinerea* clusters. Isolates are clustered based membership in groups defined by hierarchical clustering of the SNPs within the botcynic 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.

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.

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.

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.

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.