RESULTS

We focused our analysis on *B. cinerea* genes with variable expression in *B. cinerea* on *A. thaliana*, and *A. thaliana* genes with variable expression in this system. Of these 9,284 *B. cinerea* genes, 74% (5,244) varied significantly in their expression due to *B. cinerea* genetic variation, with an average broad-sense heritability of 0.152 due to *B. cinerea* isolate (Zhang, Corwin et al. 2018). Of the 23,898 *A. thaliana* genes, 85% (20,328) varied significantly in their expression due to *B. cinerea* genetic variation, with an average broad-sense heritability of 0.108 due to *B. cinerea* isolate (Zhang, Corwin et al. 2017).

We associated these 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. To associate this expression variation to genomic variation, we performed Genome-wide Efficient Mixed Model Association (GEMMA). 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 phenotype as a p-value.

To search for hotspots of quantitative trait loci (QTL) potentially controlling expression variation, for the SNP with the strongest evidence (lowest p-value) of association by GEMMA per transcript, we plotted the p-value and location of each top SNP. Hotspots are visualized as peaks composed of many high p-value SNPs.

To validate SNP peaks in our hotspot analysis, 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 phenotypes, one from each measured *B. cinerea* 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 X1). We then defined hotspots in our data as any peaks formed by SNPs with p-values below a 100% threshold defined by a 5x permutation of randomized transcript profiles.

These randomized analyses do not produce any eQTL hotspots, thus we are confident the hotspots detected in our original analysis cannot be explained by random associations between transcriptional variation and genomic variation.

We identified regulatory hotspots on most chromosomes, except chromosome 11, 17, and 18 (Figure X1a). These hotspots of controlling variation may be due to *cis*- or *trans*-acting loci. Genome-wide, we identified XX hotspots, from X to X per chromosome. Hotspots were defined by X to X SNPs, with an average length of X kb.

To look for *B. cinerea* loci that are *trans*-acting to control expression variation in the affected host, we also examined patterns of association between *A. thaliana* expression variation and *B. cinerea* genomic variation. We identified hotspots putatively controlling *A. thaliana* expression variation for all *B. cinerea* chromosomes except 17 and 18 (Figure X2b). These *B. cinerea* loci may contribute to regulation of gene expression in the host. We looked for evidence of eQTL hotspots that are common across both the *B. cinerea* and *A. thaliana* analysis by… XXXX.

To search for a pattern of transcriptome-wide *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 X3a). These distances are similar to those from the association of random transcript profiles to top SNPs (Figure X3b). 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.

In these transcript-to-SNP associations, we find that *cis*- control of loci is largely drowned out by patterns of *trans*-acting variation. 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 to *cis*-acting 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 X4). We again conclude that most of this genetic variation is *trans*-acting.

As we did not detect a transcriptome-wide pattern of *cis*-effect variation controlling expression patterns of *B. cinerea* on *A. thaliana*, we narrowed our focus to a subset of localized gene networks to identify the role of *cis*-eQTL in controlling this expression variation. These focal networks were among those most highly conserved across *B. cinerea* isolates (Zhang, Corwin et al. 2018). These include the botcynic acid biosynthetic pathway (13 genes, 55.8 kb), botrydial biosynthesis (7 genes, 26 kb), and Network 5 (Zhang, Corwin et al. 2018), which contains XX (10 genes, 46.5 kb).

[Primary analysis on all 3 of these networks]

We focused further analysis on the botcynic acid biosynthesis network. Upon visualizing SNP-level variation surrounding genes in this pathway, we found evidence of a deletion common to 12 of the 96 isolates. We selected a focal region encompassing the deletion endpoints (53.5 kb) and an additional 2 genes beyond the deletion boundaries (Bcin01g00170, Bcin01g00190)(Figure X5). 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. Hierarchical clustering based on these loci assigned the 96 isolates into three major clusters, and one small two-isolate cluster (B05.10, Fd1) (Figure X2). The largest cluster (cluster 4) includes a sub-cluster containing all of the isolates in the deletion group (cluster 3). We removed isolates with the whole-region deletion, then performed ANOVA across the 3 major clusters (1, 4, 5) to see whether SNP variation within the botcynic acid biosynthetic network predicts expression level of genes within the region. We found no significant effect of cluster membership on expression profile (F(1,74)=0.36, p=0.55). This suggests that the major *cis*-effect loci controlling expression variation in this network are not captured by SNP-level variation.

We examined mean botcynic acid pathway expression across all isolates, grouped by cluster membership. Expression levels across the network are reduced in the deletion group (cluster 3) but we also see independent low-expression isolates outside of this deletion group (Noble Rot, 01.04.03, Apple 517, 02.04.09) (Figure X7). These isolates also contain deletions within the botcynic acid region (Figure X5), likely of independent origin. This deletion extends X kb and includes SNP X at the X’ end of the chromosome, indicating a teleomeric loss on chromosome X. As such, the major *cis­*-effect variation controlling expression in the botcynic acid region appears to be explained by deletion events, which are not captured by our SNP analysis. 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.

METHODS

Experimental design

We used a previously described collection of B. cinerea isolates 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). 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).

Expression analysis

RNASeq libraries were prepared as previously described (Kumar, Ichihashi et al. 2012, Zhang, Corwin et al. 2017). 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/](http://www.bioinformatics.babraham.ac.uk/projects/)), and trimmed (fastx, http://hannonlab.cshl.edu/fastx\_toolkit/commandline.html). Reads were aligned to the Arabidopsis TAIR10.25 cDNA reference genome (Bowtie 1 V.1.1.2, <http://sourceforge.net/projects/bowtie-bio/files/bowtie/1.1.2/>) {Langmead, 2009}, and we pulled gene counts (SAMtools, (Li, Handsaker et al. 2009). We summed counts across gene models, and normalized gene counts as previously described (Zhang, Corwin et al. 2017).

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). 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 data

For GEMMA mapping, we used 95 isolates with a total of 237,878 SNPs against the *B. cinerea* B05.10 genome with MAF 0.20 or greater and less than 10% missing calls.

GWAS

We used haploid binary SNP calls with MAF > 0.20 and <20% missingness. We matched these phenotypes (9,267 *B cinerea* gene expression profiles and 23,947 *A. thaliana* gene expression profiles) to the SNP data with custom R scripts, for a total of 95 isolates. We ran GEMMA once per phenotype, across 9,267 transcripts.

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%.

FIGURE LEGENDS

Figure X1. Manhattan-type plot of GEMMA results of transcriptome-wide *B. cinerea* expression phenotypes.

Manhattan-type plot of top 1 SNP per *B. cinerea* transcript on Col-0 *A. thaliana*. Panel a is from measured expression profiles, panel b is from a randomized expression phenotype. 9,267 *B. cinerea* expression profiles were randomized across the 96 isolates prior to GWA analysis. We repeated this permutation five times and report the mean p-value across all permutations. Random max from 1 permutation: 6.81522 (p = 1.530313e-07).

Figure X2. Interspecific hotspot comparison of eQTL on *B. cinerea* Chromosome 1.

a. Top SNP (lowest p-value) per each of 9,267 *B. cinerea* gene expression profiles on Col-0 A. thaliana detached leaves.

b. Top SNP (lowest p-value) per each of 23,898 Col-0 *A. thaliana* gene expression profiles under *B. cinerea* infection.

Figure X3. 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. Panel a is from measured expression profiles; panel b is from a single randomization of the *B. cinerea* expression profiles.

Figure X4. Cis-diagonal plot correlating *B. cinerea* gene center to position of top associated SNP. We retained only the SNPs with highest probability (lowest p-value) of significant effect on expression of the transcript of interest. Panel a is single top SNP per transcript, panel b is top 10 SNPs per transcript. Each point represents a single transcript from our *B. cinerea* expression profile, with y axis of transcript center and x axis of top SNP location. Chromosome locations are indicated as red bars along the x axis.

Figure X5. SNP state of *B. cinerea* isolates within botcynic acid biosynthesis network. Binary states are color coded as 2 (yellow) for B05.10 reference, 1 (green) for non-reference. SNPs missing data are coded as 0 (white). All SNP calls within the deletion region have been coded as 0. Isolates are arbitrarily ordered to visualize grouping by genotype.

Figure X6. Hierarchical clustering of *B. cinerea* isolates from SNPs in 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%.

Figure X7. 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).

REFERENCES

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**.

Lenth, R. (2018). "emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.3.0. <https://CRAN.R-project.org/package=emmeans>."

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.

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."