

Resequencing and association mapping of the generalist pathogen *Botrytis cinerea*

Susana Atwell¹, Jason A. Corwin^{1,2,3}, Nicole Soltis^{1,2}, Wei Zhang^{1,4}, Daniel Copeland¹, Julie Feusier¹, Robert Eshbaugh¹, and Daniel J. Kliebenstein^{1,5*}

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

¹Plant Biology Graduate Group, University of California, Davis, One Shields Avenue, Davis, CA, 95616, USA

³Department of Ecology and Evolution Biology, University of Colorado, 1900 Pleaseant Street, 334 UCB, Boulder, CO, 80309-0334, USA

⁴National and Local Joint Engineering Laboratory for Energy Plant Bio-oil Production and Application, Key Laboratory of Bio- resource and Eco-environment, Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, 610064, PR China

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

Abstract

We performed whole genome resequencing of 84 field isolates of *Botrytis cinerea*, largely collected from a local set of plant species. Combined with 13 previously resequenced isolates sampled from diverse locations, this gave a collection of 97 isolates for studies of natural variation. Alignment to the reference sequence T4 and SNP detection provided further data for population genetics analysis including a mapping population for association studies. Although much of the genomic diversity was captured in the original 13 isolates, the additional genomes increased total diversity in the population by a third. Surprisingly, the same additional genomes increase mitochondrial diversity 2 fold. Across the population, LD was limited and decayed rapidly, reflecting frequent outcrossings. Effectively, this sampling strategy increased the level of genetic diversity available, whilst limiting the problem of population stratification and enabling GWAS of several phenotypes on common *Arabidopsis* plants effected in disease pathways. Overlap of results using all GWAS methods revealed numerous candidate genes / pathways that potentially contribute to its broad host range and offer conceivable pathogen decrease targets.

Introduction

Interactions between plants and pathogens are a key factor in crop yield, with a large body of research studying its evolutionary and mechanistic basis. The majority of plant pathogen research has however focused on specialist pathogens that infect a small number of economically significant species and frequently cause epidemic bursts of disease. This specialization of epidemic pathogens to limited hosts has generated a model of pathogen evolution that is dominated by the generation of new large-effect virulent loci in the pathogen that are countered by equally large-effect resistance alleles in the plant. In contrast to epidemic specialist pathogens are the endemic generalist pathogens such as *B. cinerea* and *S. sclerotiorum*. The isolates of these pathogens have the ability to affect nearly any plant host and are widely distributed in the natural environment. However, the evolutionary and genomic pressures that are imparted by this generalist endemic life-style are less well understood.

Botrytis cinerea is a necrotrophic plant-pathogenic fungus, the causal agent of gray mold. As a generalist fungus, it infects most vegetable and fruit crops, and a large variety of shrubs, flowers, trees and weeds, including many crops of economic importance (Dean et al., 2012). It is the most important postharvest decay pathogen (Romanazzi & Feliziani, 2014) and the second most economically significant plant pathogen (Dean et al., 2012). Whilst there is some evidence for host specialization / increased virulence by some isolates of *B. cinerea* on specific host species (Cotoras & Silva, 2005), the

effects of geography, host selection and agronomic practices on diversity are poorly understood. Field collected isolates are almost entirely haploid and the presence of a small sequenced genome makes this pathogen a useful model to investigate how a generalist endemic life cycle may structure the genomic variation in the pathogen.

A key component of the evolutionary pressures on generalist pathogens is the interplay of resistance and virulence mechanisms between the pathogens and the host. It is well known on the plant side that the resistance to generalist pathogens is typically dominated by a highly polygenic architecture involving 100s to 1000s of genes (Corwin & Kliebenstein, 2017; H. C. Rowe & Kliebenstein, 2008). To date, there are only a limited number of highly successful quantitative genetic studies of pathogen traits suggesting that virulence can be affected by multiple genes in specialist pathogens (M. H. Lendenmann, Croll, Palma-Guerrero, Stewart, & McDonald, 2016; Mark H. Lendenmann, Croll, & McDonald, 2015; Mark H. Lendenmann, Croll, Stewart, & McDonald, 2014) . Similarly, there is a wide range of known *B. cinerea* pathogenicity genes / effectors /gene clusters and potential pathways such as Botrydial and Botcinic acid, Velvet, PolyKetide Synthases (PKS) and Non Ribosomal Peptide Synthetases (NRPS) suggesting the potential for polygenic variation in *B. cinerea* virulence (Dalmais et al., 2011; Pinedo et al., 2008; Schumacher et al., 2012). However, we have little understanding if generalist pathogen virulence is equally polygenic to the host plant.

One approach that is useful to uncovering polygenic architectures, especially in species without structured mapping populations is Genome wide association mapping. Fungi with sufficient recombination and diversity should be key candidates for using genome wide association mapping to understand the genetic architecture of virulence. However, there are only a few genome-wide association studies (GWAS) in necrotrophic fungal pathogen. And these have largely been limited in their ability to look at polygenic architecture either by low SNP coverage (Gao et al., 2016; Talas, Kalih, Miedaner, & McDonald, 2016) or small genotype sample (Dalman et al., 2013; Palma-Guerrero et al., 2013; Wang et al., 2018). To utilize GWA to understand the architecture of virulence, we developed a large GWAS population within the generalist endemic pathogen *B. cinerea*. This provides a more detailed image of the genomic architecture and polygenic nature of virulence variation.

We performed whole-genome resequencing of 84 isolates of *B. cinerea* to enhance our understanding of variation in this pathogen. This data, combined with 13 previously resequenced isolates from our pilot study, found high levels of polymorphism, including 1,516,518 unique SNPs when mapped

against the ~ 41 Mb of the reference sequence from the *B. cinerea* isolate T4. Similar levels were found when the isolates were aligned to an additional *B. cinerea* isolate reference BO5.10. We describe the genomic pattern of polymorphism, the population structure and low levels of linkage disequilibrium (LD) decay of the isolates sampled, loss of function mutations and selective sweeps found. The SNP data additionally enabled the creation of a GWAS mapping populations of *B. cinerea* that were phenotyped on 4 genetic backgrounds of *Arabidopsis thaliana*, Col-0, *npr1*, *coi1* and *pad3*. Using multiple GWA mapping approaches with different SNP sets identified a highly consistent picture of virulence variation in the pathogen. Specifically, virulence is highly polygenic within *B. cinerea*. Further, the identified candidate genes were predominantly dependent on the hosts resistance mechanism(s) but there were some general virulence loci. Thus, the genetic architecture is highly dependent on the interaction of the host and pathogens genomes. The results identified potential targets to assist in future control of this pathogen.

Mat&Methods

Polymorphism Detection

DNA was purified from 97 single spore derived *B. cinerea* isolates as described previously (Atwell et al., 2015). Genomic sequencing, sequence clean-up, alignment of isolates against both the T4 and later the BO5.10 reference genomes and variant calling was carried out as described previously in (Atwell et al., 2015). SNPs and indels were called in the 97 haploid isolates with a minimum coverage of 6 and a quality score of 30. Analysis was mostly carried out using the T4 data. Additionally, alignment and SNP calling to the BO5.10 mitochondrial and later the genomic reference sequence, that was published during our analyses (Van et al., 2017), was also conducted using the same methodologies.

Phylogenomic analysis

Consensus sequences were created from all reference based alignments and converted to fasta format using GATK. A distance-based approach was used for phylogenetic inference. Radial phylogenetic trees were reconstructed from pair-wise distances using SEAVIEW (Gouy, Guindon, & Gascuel, 2010) and plotted using Figtree (Rambaut, A. and Drummond, A, 2008). This revealed that 11 genomic aligned isolates sampled from an organic field were highly similar. As such further analysis was conducted using 87 isolates, including a single representative isolate from the 11 highly similar isolates.

Population structure

We tested for the presence of population structure and admixture using STRUCTURE 2.3.4 (Pritchard, Stephens, & Donnelly, 2000) on a filtered SNP subset with a minimum interval of 500bp between SNPs. STRUCTURE was run in the admixture model and 1 to 10 genetic clusters (K) were tested, repeated ten times per K with a burn-in of 1000 iterations followed by 1000 MCMC iterations. To estimate the most likely value of K , the likelihood of the data for a range of K values was calculated by creating posterior probabilities of K . The different Q matrices were merged with CLUMPP (Jakobsson & Rosenberg, 2007) and plotted with strplot (Ramasamy, Ramasamy, Bindroo, & Naik, 2014).

Principle Component Analysis (PCA) of Isolate Variation

PCA was conducted using pca2d (Khang, Soo, Tan, & Lim, 2016) on the 87 isolates aligned to T4 based on geography (origin of sample location) and family (plant family from which the isolate was sampled).

Population Genetic Statistics

Haplotype diversity, Nucleotide diversity and ZnS were calculated using PopGenome (Pfeifer, Wittelsbürger, Ramos-Onsins, & Lercher, 2014) and SNPs from the 87 isolates with a minor allele frequency of at least 0.1 and complete information in nonoverlapping 5-kb windows. **Decay in linkage disequilibrium (LD)** was determined using PopLDdecay (C. Zhang, Dong, Xu, He, & Yang, n.d.) with the collection of 87 isolates. Additional subsets of isolates grouped by host plant family or location of isolate were also utilized to assess for differential decay. **Selective sweeps** were determined from the 87 isolates using SweeD_v3.3.2. (Pavlidis, Živković, Stamatakis, & Alachiotis, 2013), a likelihood based detection method of selective sweeps based on allele frequencies which was used to identify candidate genomic regions that are indicative of strong selection pressure at the population level. The composite likelihood ratio (CLR) was calculated for each contig with a resolution of 5kb bins.

Virulence Phenotyping

The virulence of the collection of *B. cinerea* isolates on the *Arabidopsis thaliana* (*A. thaliana*) genotypes (Col-0, *pad3*, *npr1*, *coi1*) was previously measured using a detached leaf assay that allows for high-throughput analysis and is typically consistent with whole plant assays (Boydom, Dawit, & W/Ab, 2013; Denby, Kumar, & Kliebenstein, 2004; Govrin & Levine, 2000; Mengiste, Chen, Salmeron, & Dietrich, 2003; Mulema & Denby, 2012). The virulence measurements for Col-0, *coi1-1*

and *npr1-1* were obtained from the previous publication (W. Zhang et al., 2017). In the same experiments as previously reported for these three genotypes, we had also included the *Arabidopsis* genotype *pad3* that is deficient in camalexin synthesis, a major defense metabolite against *Botrytis*. Pathogen inoculation and utilization of digital images to determine lesion area protocols were as described previously (W. Zhang et al., 2017). Lesion data was analyzed using a generalized linear model with a Gaussian link function for lesion area in an individual model for each plant genotype as described previously (W. Zhang et al., 2017). Using this model, we calculated the least-squared means of the phenotypes within each plant genotype as the phenotype input for GWAS.

GWAS mapping population

For GWAS mapping, the vcf files of SNPs from the 87 isolates aligned to the reference T4 and separately to the reference BO5.10 and SNPs were filtered (minimum 6x coverage, minimum quality score of 30 and MAF 0.2) and formatted to tab separated files using vcftools (Danecek et al., 2011). This gave 345,485 loci from isolates aligned to the T4 reference and GWAS analysis was conducted using the methods Wilcoxon (Wilcoxon, 1945) and bigRR (Shen, Alam, Fikse, & Rönnegård, 2013). Because bigRR provides an estimated effect size, but not a p-value, we performed permutation analyses to determine effect significance. We permuted the phenotypes 1000x and re-ran the analysis, to establish 95%, 99%, and 99.9% thresholds for significance. GWAS analysis was also carried out with GEMMA (X. Zhou & Stephens, 2012) using 237,878 SNPs, at MAF 0.20 or greater, aligned to the BO5.10 reference, and less than 10% missing SNP calls as described above. To determine significance of SNPs by GEMMA, we used 1000 permutations to determine p-value significance at the 99%, and 99.9% thresholds (Corwin et al., 2016a; Doerge & Churchill, 1996; Shen et al., 2013).

From the significant associations, candidate gene lists were created by including genes found as having a candidate SNP within the overlap of the top 500 SNPs in both the Wilcoxon and bigRR mapping methods and were annotated to specific genes using SNPdat (Doran & Creevey, 2013) and the T4 gene models for genomic DNA (broadinstitute.org, (Staats & Kan, 2012a). The resulting genes were compared to the results from the BO5.10 mapping population GEMMA using annotation from Inra
<http://dx.doi.org/10.15454/IHYJCX>

Results

Isolate sampling

To investigate the genomic architecture of genetic variation within the broad host-range generalist *B. cinerea*, we obtained a collection of isolates using single spore isolation approaches. This collection of natural isolates was obtained mainly from California with a sampling of available international germplasm (Supplemental Table 1). Previous work on a subset of this collection has shown that this represents a population displaying a wide range of genomic, phenotypic and virulence variation (Atwell et al., 2015; Corwin et al., 2016b; Heather C. Rowe & Kliebenstein, 2007). By collecting and extending the population using a local set of diverse plant species in combination with a sampling of isolates available from around the world, we worked to minimize extensive linkage disequilibrium (LD) and population structure. This strategy should increase the level of genetic diversity that is available, while limiting the problem of population stratification.

Genomic diversity

Genomic sequencing and polymorphism detection in 84 new isolates was carried out identical to a previous pilot study (Atwell et al., 2015). In combination with 13 isolates from this pilot study, we generated genomic sequencing data on a total of 97 isolates with a mean of 164X-fold coverage per isolate (32–214X, Supplemental Table1). Genome wide polymorphisms in the nuclear genome were identified by alignment first to the previously used sequenced reference T4 (Staats & Kan, 2012b). Additionally, alignment and SNP calling to the BO5.10 reference sequence (Van et al., 2017), which was published during our analyses, was also carried out to discern if the reference genome affected the observed level and/or patterns of sequence variation. The BO5.10 reference has the advantage of final chromosome designations as.

This analysis identified 1,516,518 unique SNPs against the T4 reference; an additional 536,189 SNPs than were found using the 13 isolates from the pilot study (Atwell et al., 2015). The SNP polymorphisms determined against the T4 reference yielded an average of at least one polymorphism between the isolates every 27 bp across all pairwise combinations, giving an average density of 37 polymorphic SNPs per kb within the population. Single isolates had between 127,779 and 392,802 SNPs per isolate against T4. Similarly, the complete collection of isolates identified 396,638 indel polymorphisms, an increase of 161,474 above the original collection of 13 isolates. There was an average of 46,433 (29,003-67,508) insertions per isolate and 37,682 (22,879-57,620) deletions in comparison to the reference genome. Combining the polymorphisms (both SNP and indel) from all pairwise combinations of isolates against the T4 reference yielded an average of at least one polymorphism between the isolates every 23bp, giving an average density of 43 polymorphisms per kb within the population. There was no significant correlation between polymorphism per isolate and depth of sequencing per isolate (Supplemental Table1).

To test the influence of the reference genome on these estimates, we also mapped the re-sequencing data against the BO5.10 reference genome (Van et al., 2017). This identified a similar rate of polymorphism with 1,420,317 SNPs and 368,981 indels. Single isolates (not including our BO5.10 isolate) had between 165,566 and 402,227 SNPs per isolate, between 14,028 and 65,465 insertions and 12,105 and 56,410 deletions against the BO5.10 reference (Supplemental Table1). This suggests that the two reference genomes give a similar image of a highly polymorphic species. Resequencing the BO5.10 genome using our sample of this genotype identified 125 SNPs against the reference BO5.10 genomic sequence. It is not presently clear if these are mutations or errors in the reference genome or resequencing efforts. Previous work has suggested that spontaneous mutation rates in *B. cinerea* are too low to account for this level of SNPs suggesting that these are technical discrepancies. Given that this rate is >1000 fold lower than the SNP frequency using wild isolates, this indicates that >99.9% of the SNPs identified are real.

Mitochondrial diversity

Work on quantitative traits in a large number of species has shown the potential for genetic variation in the organelle to influence adaptive traits both alone and in interaction with nuclear polymorphisms (Joseph, Corwin, & Kliebenstein, 2015; Paliwal, Fiumera, & Fiumera, 2014). To measure variation in

the organellar genome, we reconstructed the mitochondrial genome in each of the isolates against the BO5.10 reference mitochondrial sequence. Only the BO5.10 mitochondrial genome was used because the T4 mitochondrial sequence was not published. This found a mean 225X-fold coverage (36–249X) of the mitochondrial genome and identified 388 SNPs, 1,360 insertions and 669 deletions across the isolates. This is 335 more SNPs than were found in the original collection of only 13 diverse isolates. Single isolates varied between 2 and 288 SNPs per isolate against the reference BO5.10 mitochondrial genome (Supplemental Table 1). Insertion / deletion patterns were similar with individual isolates having between 5 to 703 insertions, and 1 to 401 deletions (Supplemental Table 1). There was a single SNP and zero indel variation between our BO5.10 sample and the reference sequence (JA Kan Van et al., 2017).

Combining the SNP and indel polymorphisms against the BO5.10 mitochondrial reference yielded an average density of 29 polymorphisms per kb within the population. This showed that mitochondrial genomic variation was approximately 20% lower than for the nuclear genome with an average unique polymorphism every 34bp from the combined isolates. At first impression, this seemed to be a high rate of mitochondrial variation compared to the pilot study (Atwell et al., 2015). However, further investigation revealed that this increase in variation was due to 3 isolates, 1 with a dramatically higher SNP variation and 2 with greatly increased numbers of indels compared to the other isolates that were sequenced. A closer look at mitochondrial sequence of the two isolates with increased indel number revealed numerous insertions which ranged from 10-32 bp. Blasting (Altschul et al., 1997) the majority of these insertions returned matches to *Botrytis* species but 3 came back as potential matches from other species. However, without further de novo mitochondrial sequencing and sequencing of additional *Botrytis* mitochondria, we cannot determine if these are highly degenerate mitochondria or horizontally transferred mitochondria from other *Botrytis* species. The increase in mitochondrial variation at first impression seemed to be dramatically high (in comparison to that shown in the much smaller pilot study). However, the resequencing of just 18 isolates of *Lachancea kluyveri* revealed similar levels of polymorphism with an average diversity of 28.5 SNPs/kb and 6.6 indels/kb (Jung, Friedrich, Reisser, Hou, & Schacherer, 2012). Thus, the mitochondria may contribute to phenotypic variation in the fungus.

Isolate clustering and admixture

To visualize the genetic relationship amongst these isolates, we plotted their phylogenetic relationships. This utilized the nuclear genome wide SNP diversity obtained from all 97 isolates in combination with a neighbor-joining phylogenetic tree using pairwise SNP differences to T4 to generate the alignments (Figure 1A). The un-rooted radial phylogeny of the 98 isolates (including the reference T4) showed slight clustering and 11 isolates sampled closely from one organic grape field appeared nearly genetically identical. As such all further aligned analysis utilized only 87 isolates, including just one representative isolate from the collection of nearly identical isolates. This phylogeny was similar if the SNPs were called using the T4 or BO5.10 reference genome (Figures 1A and 1B). In contrast to the nuclear phylogeny, the mitochondrial phylogeny (aligned to BO5.10) was dominated by the 3 isolates showing hypervariable mitochondria. We removed these isolates to allow better visualization of the remaining isolates mitochondrial phylogeny (Figure 1C). Together, these trees suggest that there is admixture amongst the isolates as supported by previous evidence of recombination in the genome (Atwell et al., 2015).

Previous studies with microsatellites and other genetic markers have not clearly resolved the genetic structure of genomic variation in this species. The pilot-study mitochondrial phylogeny suggested the presence of four clusters/populations but the nuclear genomic sequencing of the 13 isolates was less clear. The mitochondrial phylogeny of the larger population (minus the 3 hypervariable isolates) again suggests the presence of four clusters (Figure 1C). Using the nuclear genomic variation with principal component analysis (PCA) while indicating partitioning also indicated some level of admixture with four significant axes that explain ~40% of the genetic variation (Figure 2A). However, none of the vectors show a clear division of isolates by geographical location (Figure 2B) or family of plant from which the isolate was sampled were revealed (Figure 2C). Further, these vectors were dominated by a few isolates and did not affect the main body of the population.

To directly assess the presence of population structure and admixture, Bayesian clustering analyses were performed assuming an admixture model using the software STRUCTURE (Pritchard et al., 2000) yielding results consistent with the neighbor-joining phylogeny analyses. From the STRUCTURE and CLUMPP (Jakobsson & Rosenberg, 2007; Pritchard et al., 2000) results, $K=4$ (the smallest stable K value represents the optimum value) was the best clustering solution of the 87 isolates (Figure 2D). 41 isolates were placed in distinct clusters (made up of 19, 9, 5 and 8 isolates) with the remaining 46 isolates being admixtures of 2-4 of these clusters. This suggests that within our

collection of isolates, there is weak population structure that does not appear to map to geography or host plant. Further, half of these isolates represent admixtures of these four theorized populations which increases the potential for genome wide association mapping to be useful.

Linkage Disequilibrium

Structure analysis suggest the presence of extensive admixture among the isolates. This result is supported by previous work with genomic sequencing and AFLPs (Atwell et al., 2015; Walker et al., 2015) that suggested the potential for extensive recombination within this fungal species. To test how this may influence genomic variation within this larger collection of isolates, we measured the decay of linkage disequilibrium across the genome. Across most of the contigs, LD decayed rapidly and mean r^2 across the genome fell to approximately one-half of the initial value within 1 kb (Figure 3A). Further supporting the presence of relatively frequent outcrossing and recombination within this species. As has been found in a number of other species, mean LD was variable (0.0128 – 0.827, ZnS (Kelly, 1997), with some localized regions of high LD (Figure 3B and 3C). A closer look at LD decay by grouping the isolates by the plant family from which they were isolated or the international isolates by themselves showed similar rates of LD decay as within the entire population (Supplemental Figures 1A-D), additionally supporting that host source or geography is not driving genomic architecture. Thus, as noted in the pilot study via recombination breakpoint analysis (Atwell et al., 2015), there is extensive recombination within this species on a genome wide level, limiting mean LD decay to relatively small regions (< 1,000 bp).

Diversity and Selective Sweeps

The localized regions of higher LD suggested the potential presence of localized variation in selective pressures/sequence diversity. To investigate the potential for local variation in sequence diversity, we shifted from a whole genome analysis to discrete local analysis. Across the entire genome, the levels of nucleotide diversity estimated in non-overlapping 5-kb windows ranged between 0.0001 and 0.044 per site amongst the 84 isolates (Figure 3B and 3C), with a mean nucleotide diversity of 0.005 across all combined contigs. Measurements of haplotype diversity in non-overlapping 5kb blocks gave elevated but not unusual levels ranging between 0.02 and 0.99, with an average of 0.85 (Figure 3B and 3C). Thus, there is extensive local variation around the mean genomic average sequence diversity.

One potential process that can alter local sequence variation are selective sweeps that are indicative of

strong directional selection pressure at the population level. The presence of low levels of genome wide haplotype diversity warranted investigation of selective sweeps. Analyzing the genomic variation with SweeD (Pavlidis et al., 2013), identified 56 peaks genome-wide with a CLR under the 0.95 criterion calculated with a resolution of 5kbp bins across the whole genome (Supplemental Figure 2). Similar to the polymorphisms that were predicted to cause major-loss of function in the pilot study, these 56 regions identified genes associated with mating and virulence. Particularly heterokaryon incompatibility genes and Nacht domain protein encoding genes that are associated with vegetative incompatibility in fungi (Glass, Jacobson, & Shiu, 2000; Glass & Kaneko, 2003). Known virulence genes including Velvet where the selective sweep is against a minor allele linked to altered virulence and growth (Schumacher et al., 2012) were also found. In addition, there are sweeps affecting Chitin Synthase, Tannase (a lignin degrading enzyme), Patatin and unknown Cytochromes p450. Thus, there appear to be potential selective sweeps in *Botrytis* that are linked to known mating and virulence loci.

Genomic and mitochondrial loss of function mutations

To investigate the potential functional consequences of this high polymorphism rate, we extracted polymorphisms predicted to have major effects on gene function using SnpEff (Cingolani et al., 2012) and the annotation from T4 (<http://www.broadinstitute.org>; *B. cinerea* T4). These could either be via premature protein truncation, triggering of nonsense mediated decay or other major effect polymorphisms. This analysis identified 14,498 polymorphisms likely to have major loss-of-function effects on 2,697 genes (out of a possible 10,401 genes annotated in the T4 reference). This represented an increase of only about 2 fold over the original 13 isolates when adding 73 additional isolates. This suggests that we are beginning to resample the major functional polymorphisms present within the species but the extensive recombination is allowing us to sample new combinations of alleles.

An analysis of these polymorphisms showed that the vast majority were not singleton polymorphisms as might be expected if they were highly deleterious to the organism. However, half of the loss of function polymorphisms had a frequency lower than 20% (Supplemental Figure 3). Importantly, we were able to find the previously reported loss of function polymorphism for the Velvet gene (Schumacher et al., 2012) segregating within the population. Additionally, like the selective sweep analysis in this study and the loss-of-function polymorphisms from the previous pilot-study, there was

a tendency for loss-of-function polymorphisms to occur in heterokaryon incompatibility genes and Nacht domain protein encoding genes, Tannase, Patatin and Cellulase genes. These control the ability of the isolates to have vegetative mating or to degrade plant cell walls. Thus, there is a large pool of large-effect polymorphisms segregating within *Botrytis* potentially contributing to its broad host range.

534 mitochondrial polymorphisms were predicted to have large functional consequences in 29 of the 38 genes annotated in the BO5.10 mitochondria, with 9 genes having no SNP or Indel polymorphism major or otherwise. These conserved regions are evenly distributed across the whole mitochondria (Supplemental Figure 4). However, annotation is not informative as to why they are non-polymorphic. Future work is required to test how frequently divergent mitochondria appear within a vastly larger collection of this and other *Botrytis* species in addition to mitochondrial de novo sequencing.

GWAS and Polygenic Virulence

The linkage disequilibrium and polymorphism data suggests that it should be possible to conduct a GWAS study in this population to identify naturally variable *B. cinerea* genes that influence differential virulence on specific hosts. To test this possibility, we individually inoculated all of the resequenced isolates on the common lab *A. thaliana* accession Col-0 and on three *A. thaliana* plants compromised in specific resistance pathways *coi1*, *pad3* and *npr1*. We obtained average virulence for each pathogen isolate by host genotype combination using eight-fold independent randomized replication across two experiments. The traits followed a normal distribution (Figures 4A-D) except that lesion scoring on Col-0 was slightly right-skewed due to the isolate A517 (Figure 4A). Heritability of virulence across the *B. cinerea* isolates, among the *Arabidopsis* genotypes and interaction between isolate and host genotype have been previously published ($H^2_{\text{Isolate}} = 0.164$, $H^2_{\text{HostGenotype}} = 0.151$, $H^2_{\text{Isolate} \times \text{HostGenotype}} = 0.047$ respectively; (W. Zhang et al., 2017)).

For the GWAS, we utilized up to 345,485 high quality SNPs that were evenly distributed across the *B. cinerea* contigs from the T4 reference. We utilized three methods for GWAS, the first was a Wilcoxon non-parametric SNP by SNP analysis (Wilcoxon, 1945). The second was a ridge-regression analysis, bigRR (Shen et al., 2013) that tests all of the SNPs in a single model. This approach has previously been shown to have a high validation rate within *A. thaliana* and the rank order of the SNPs is highly similar to that obtained using the EMMA methodology (Kang et al., 2008). The benefit of the ridge

regression is that it can be conducted quickly enough to do direct permutation analysis to find empirical thresholds based on the phenotypic distribution (Shen et al., 2013). The final method was to use a mixed model to control for population structure within the GEMMA (X. Zhou & Stephens, 2012) pipeline and a SNP mapping population created against the BO5.10 reference. Using the Wilcoxon and bigRR methods, we identified the genes linked to the top 500 SNPs for virulence on each host genotype (Figure 5, Table 1). We chose the top 500 SNPs, as permutation with bigRR showed that all of these were significant for all traits and we wanted to compare similar sized gene lists across the approaches. SNPs of interest were distributed over all contigs and between 40 and 60 genes were found per phenotype that overlapped between the three methods (Figure 5). Including the GEMMA results showed that the three methods and two mapping populations largely found the same genes or when there was a discrepancy, they identified neighboring genes (Table 1). Using the combination identified between 5 and 9 genes per phenotype as being of immediate interest but these were entirely uncharacterized genes (Table 1).

The *B. cinerea* genes identified as linked to variation in virulence on *Arabidopsis* were highly dependent on the specific host genotype with Col-0, *npr1-1* and *coi1-1* illuminating different pathogen loci. In contrast, *pad3* and *coi1-1* identified largely overlapping sets of virulence associated genes in *B. cinerea*. This similarity between *coi1-1* and *pad3* is possibly because both mutants lead to a loss of production of the major defense metabolite, camalexin, upon infection by these isolates (Schuhegger et al., 2006; N. Zhou, Tootle, & Glazebrook, 1999 (Schuhegger et al., 2006; N. Zhou, Tootle, & Glazebrook, 1999). This suggests that the collection of virulence loci found when using the Col-0 genotype but not the *coi1-1* and *pad3* hosts may be associated with how *B. cinerea* counteracts this specific metabolic defense. Additionally, 3 genes were found for linked to altered virulence on 3 or more host genotypes; NMT1/THI5 like- Thiamin biosynthesis -vit B1 (NMT1), for fungi, thiamine is essential for vegetative growth, development, sporulation, host invasion, and adaptation to environmental stress (Hohmann & A Meacock, 1998) Berberine bridge enzyme like (BBE), which is part of the PolyKetide Synthase19 (PKS19) cluster; and a Nacht domain gene. Together, this shows that it is possible to use GWA within the pathogen to find loci that control both general virulence and host specific virulence mechanisms.

Discussion

Given the economic cost of crop damage caused by *B. cinerea*, this plant pathogen has undergone extensive investigation, with great advancements in predicted gene functionality made recently via sequencing projects (Amselem et al., 2011; Blanco-Ulate, Allen, Powell, & Cantu, 2013; JA Kan Van et al., 2017; Staats & Kan, 2012b). One limitation of these studies is that *B. cinerea* is a broad host pathogen and it is unclear as to how much of the species variation, genetic and phenotypic, can be captured solely by two reference genomes and one predominant lab isolate. To address these questions, we conducted whole genome sequencing on a collection of *B. cinerea* isolates, to begin understanding the genomic architecture of this broad host pathogen and how this variation may contribute to this host range. In this study we provide the first comprehensive resequencing data of a collection of *B. cinerea* isolates which enabled the creation of the first *B. cinerea* GWAS mapping population. The availability of this data will be a valuable resource for further *B. cinerea* – plant interactions, identification of genes responsible for traits and potential targets for *B. cinerea* control.

The genomic resequencing identified a high level of genetic variation amongst these isolates including major effect polymorphisms that create natural knockouts in nearly 25% of the annotated genes. This genomic resequencing showed that there was no clear signal of host source or geography as has been found in other studies using AFLPs or microsatellites (Ma & Michailides, 2005). Comparing polymorphism identified with the 84 new isolates vs an original set of 13 isolates showed that 2/3rds of the polymorphisms were found in the original 13 resequenced isolates. This suggests that there is a large body of moderate frequency polymorphisms within the species. In combination with the lack of population structure and the rapid LD decay, the new isolates are sampling new allelic combinations more than new genetic diversity as would be essential to maximize the power of any GWAS.

LD, Life-History and GWA

Fungal pathogens were considered to have a largely clonal life-style similar to Oomycete pathogens (Taylor, Hann-Soden, Branco, Sylvain, & Ellison, 2015). This would limit the utility of GWA to identify causal variants influencing variation in virulence across a species. However, recent genomic studies are beginning to show that this is not an accurate view. Recent work has shown that in *S. cerevisiae*, global LD decays within 500bp (Peter et al., 2018), similar to the 1kb found in our study of

B. cinerea. Smaller, more focused studies in plant pathogens have shown that LD is similar in *Fusarium graminearum* (mean LD ~1 kb, (Talas & McDonald, 2015) but higher in *Parastagonospora nodorum* (mean LD ~5–10 kb, (Gao et al., 2016). This could signify that similar pathogens may be more suitable for GWAS than the previous ideas of clonality signified. Our data shows that *B. cinerea* has the combination of high frequency SNPs and low LD required to enable effective GWAS. Further, *B. cinerea* has a lack of substantial sub-structuring suggesting that structure will not be a dominant driver in any GWA study. The fact that the LD decay remained rapid upon dissection of the isolates based on geography or host plant, bodes well for addition of further samples to increase power and allows for fine mapping of genes which can be targeted to suppress the destruction of this pathogen.

Using this LD and diversity, we tested the potential to use GWA by measuring virulence of the isolates on four defined *A. thaliana* genotypes that differ in their defense capacities. This showed the potential to identify both general and host specific loci that could influence virulence within this *B. cinerea* population. Using three different GWA approaches within a specific host plant identified largely overlapping sets of potential causal genes regardless of if the method controlled for population structure, GEMMA, tested SNP by SNP, Wilcoxon, or conducted a genome prediction approach, bigRR. Whilst there was overlap in the genes found between the three approaches, the SNPs found within the putative causal genes were not always identical (Table 1). Thus, *B. cinerea* GWA using this population will be an excellent tool to generate gene level hypothesis to follow up.

GWA across the host genotypes showed some consistency and some host specific genes. Intriguingly, whilst the results of our GWAS highlighted several genes and pathways of interest we did not find many of the known classical virulence genes. Selective sweep analysis and the identification of major effect polymorphisms showed this is at least partly due to allele in these classical virulence genes are below what was included in the SNP mapping populations (<20%). GWAS is usually performed with SNPs, because they have the highest frequency and SNPs of minor frequency can lead to spurious associations (Turner et al., 2011). As such, this suggests that the best way to utilize this population to understand the architecture of virulence and pathogen adaptation is to combine the GWA with a separate analysis that investigates regions like selective sweeps. Together, this combined list of polymorphisms can be used to better develop hypothesis on how a generalist pathogen evolves to a broad array of hosts.

References

- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389–3402.
- Amselem, J., Cuomo, C. A., Kan, J. A. L. van, Viaud, M., Benito, E. P., Couloux, A., ... Dickman, M. (2011). Genomic Analysis of the Necrotrophic Fungal Pathogens Sclerotinia sclerotiorum and Botrytis cinerea. *PLOS Genetics*, 7(8), e1002230. <https://doi.org/10.1371/journal.pgen.1002230>
- Atwell, S., Corwin, J. A., Soltis, N. E., Subedy, A., Denby, K. J., & Kliebenstein, D. J. (2015). Whole genome resequencing of Botrytis cinerea isolates identifies high levels of standing diversity. *Frontiers in Microbiology*, 6. <https://doi.org/10.3389/fmicb.2015.00996>
- Blanco-Ulate, B., Allen, G., Powell, A. L., & Cantu, D. (2013). Draft Genome Sequence of Botrytis cinerea BcDW1, Inoculum for Noble Rot of Grape Berries., Draft Genome Sequence of Botrytis cinerea BcDW1, Inoculum for Noble Rot of Grape Berries. *Genome Announcements*, *Genome Announcements*, 1, 1(3, 3), 311–324. <https://doi.org/10.1128/genomeA.00252-13>, 10.1128/genomeA.00252-13
- Boydom, A., Dawit, W., & Getaneh. (2013). Evaluation of Detached Leaf Assay for Assessing Leaf Rust (*Puccinia triticina* Eriks.) Resistance in Wheat. *Journal of Plant Pathology & Microbiology*. Retrieved from <http://agris.fao.org/agris-search/search.do?recordID=US2016B01151>
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, 6(2), 80–92. <https://doi.org/10.4161/fly.19695>
- Corwin, J. A., Copeland, D., Feusier, J., Subedy, A., Eshbaugh, R., Palmer, C., ... Kliebenstein, D. J. (2016a). The Quantitative Basis of the Arabidopsis Innate Immune System to Endemic Pathogens Depends on Pathogen Genetics. *PLOS Genetics*, 12(2), e1005789. <https://doi.org/10.1371/journal.pgen.1005789>
- Corwin, J. A., Copeland, D., Feusier, J., Subedy, A., Eshbaugh, R., Palmer, C., ... Kliebenstein, D. J. (2016b). The Quantitative Basis of the Arabidopsis Innate Immune System to Endemic Pathogens Depends on Pathogen Genetics. *PLOS Genetics*, 12(2), e1005789. <https://doi.org/10.1371/journal.pgen.1005789>
- Corwin, J. A., & Kliebenstein, D. J. (2017). Quantitative Resistance: More than just perception of a pathogen. *The Plant Cell*, tpc.00915.2016. <https://doi.org/10.1105/tpc.16.00915>
- Cotoras, M., & Silva, E. (2005). Differences in the initial events of infection of Botrytis cinerea strains isolated from tomato and grape. *Mycologia*, 97(2), 485–492. <https://doi.org/10.1080/15572536.2006.11832824>
- Dalmais, B., Schumacher, J., Moraga, J., Pêcheur, P. L., Tudzynski, B., Collado, I. G., & Viaud, M. (2011). The Botrytis cinerea phytotoxin botcinic acid requires two polyketide synthases for production and has a redundant role in virulence with botrydial. *Molecular Plant Pathology*, 12(6), 564–579. <https://doi.org/10.1111/j.1364-3703.2010.00692.x>
- Dalman, K., Himmelstrand, K., Olson, Å., Lind, M., Brandström-Durling, M., & Stenlid, J. (2013). A Genome-Wide Association Study Identifies Genomic Regions for Virulence in the Non-Model Organism Heterobasidion annosum s.s. *PLOS ONE*, 8(1), e53525. <https://doi.org/10.1371/journal.pone.0053525>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Dean, R., Kan, J. a. L. V., Pretorius, Z. A., Hammond-Kosack, K. E., Pietro, A. D., Spanu, P. D., ... Foster, G. D. (2012). The Top 10 fungal pathogens in molecular plant pathology. *Molecular Plant Pathology*, 13(4), 414–430. <https://doi.org/10.1111/j.1364-3703.2011.00783.x>

- Denby, K. J., Kumar, P., & Kliebenstein, D. J. (2004). Identification of *Botrytis cinerea* susceptibility loci in *Arabidopsis thaliana*. *The Plant Journal*, 38(3), 473–486. <https://doi.org/10.1111/j.0960-7412.2004.02059.x>
- Doerge, R. W., & Churchill, G. A. (1996). Permutation Tests for Multiple Loci Affecting a Quantitative Character. *Genetics*, 142(1), 285–294.
- Doran, A. G., & Creevey, C. J. (2013). Snpdat: Easy and rapid annotation of results from de novo snp discovery projects for model and non-model organisms. *BMC Bioinformatics*, 14(1), 45. <https://doi.org/10.1186/1471-2105-14-45>
- Gao, Y., Liu, Z., Faris, J. D., Richards, J., Brueggeman, R. S., Li, X., ... Friesen, T. L. (2016). Validation of Genome-Wide Association Studies as a Tool to Identify Virulence Factors in *Parastagonospora nodorum*. *Phytopathology*, 106(10), 1177–1185. <https://doi.org/10.1094/PHYTO-02-16-0113-FI>
- Glass, N. L., Jacobson, D. J., & Shiu, P. K. T. (2000). The Genetics of Hyphal Fusion and Vegetative Incompatibility in Filamentous Ascomycete Fungi. *Annual Review of Genetics*, 34(1), 165–186. <https://doi.org/10.1146/annurev.genet.34.1.165>
- Glass, N. L., & Kaneko, I. (2003). Fatal Attraction: Nonself Recognition and Heterokaryon Incompatibility in Filamentous Fungi. *Eukaryotic Cell*, 2(1), 1–8. <https://doi.org/10.1128/EC.2.1.1-8.2003>
- Gouy, M., Guindon, S., & Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, 27(2), 221–224. <https://doi.org/10.1093/molbev/msp259>
- Govrin, E. M., & Levine, A. (2000). The hypersensitive response facilitates plant infection by the necrotrophic pathogen *Botrytis cinerea*. *Current Biology*, 10(13), 751–757. [https://doi.org/10.1016/S0960-9822\(00\)00560-1](https://doi.org/10.1016/S0960-9822(00)00560-1)
- Hohmann, S., & A Meacock, P. (1998). *Thiamin metabolism and thiamin diphosphate-dependent enzymes in the yeast *Saccharomyces cerevisiae*: Genetic regulation* (Vol. 1385). [https://doi.org/10.1016/S0167-4838\(98\)00069-7](https://doi.org/10.1016/S0167-4838(98)00069-7)
- JA Kan Van, J. H. Stassen, Mosbach, A., Van, T. D. L., Faino, L., Farmer, A. D., ... Scalliet, G. (2017). A gapless genome sequence of the fungus *Botrytis cinerea*. *Molecular Plant Pathology*, 18(1), 75–89. <https://doi.org/10.1111/mpp.12384>
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14), 1801–1806. <https://doi.org/10.1093/bioinformatics/btm233>
- Joseph, B., Corwin, J. A., & Kliebenstein, D. J. (2015). Genetic Variation in the Nuclear and Organellar Genomes Modulates Stochastic Variation in the Metabolome, Growth, and Defense. *PLOS Genetics*, 11(1), e1004779. <https://doi.org/10.1371/journal.pgen.1004779>
- Jung, P. P., Friedrich, A., Reisser, C., Hou, J., & Schacherer, J. (2012). Mitochondrial Genome Evolution in a Single Protoploid Yeast Species. *G3: Genes, Genomes, Genetics*, 2(9), 1103–1111. <https://doi.org/10.1534/g3.112.003152>
- Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., & Eskin, E. (2008). Efficient control of population structure in model organism association mapping. *Genetics*, 178(3), 1709–1723. <https://doi.org/10.1534/genetics.107.080101>
- Kelly, J. K. (1997). A Test of Neutrality Based on Interlocus Associations. *Genetics*, 146(3), 1197–1206.
- Khang, T. F., Soo, O. Y. M., Tan, W. B., & Lim, L. H. S. (2016). Monogenean anchor morphometry: systematic value, phylogenetic signal, and evolution. *PeerJ*, 4, e1668. <https://doi.org/10.7717/peerj.1668>
- Krzywinski, M. I., Schein, J. E., Birol, I., Connors, J., Gascoyne, R., Horsman, D., ... Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*.

- <https://doi.org/10.1101/gr.092759.109>
- Lendenmann, M. H., Croll, D., Palma-Guerrero, J., Stewart, E. L., & McDonald, B. A. (2016). QTL mapping of temperature sensitivity reveals candidate genes for thermal adaptation and growth morphology in the plant pathogenic fungus *Zymoseptoria tritici*. *Heredity*, 116(4), 384–394.
<https://doi.org/10.1038/hdy.2015.111>
- Lendenmann, Mark H., Croll, D., & McDonald, B. A. (2015). QTL mapping of fungicide sensitivity reveals novel genes and pleiotropy with melanization in the pathogen *Zymoseptoria tritici*. *Fungal Genetics and Biology*, 80, 53–67. <https://doi.org/10.1016/j.fgb.2015.05.001>
- Lendenmann, Mark H., Croll, D., Stewart, E. L., & McDonald, B. A. (2014). Quantitative Trait Locus Mapping of Melanization in the Plant Pathogenic Fungus *Zymoseptoria tritici*. *G3: Genes, Genomes, Genetics*, g3.114.015289. <https://doi.org/10.1534/g3.114.015289>
- Ma, Z., & Michailides, T. J. (2005). Genetic Structure of *Botrytis cinerea* Populations from Different Host Plants in California. *Plant Disease*, 89(10), 1083–1089. <https://doi.org/10.1094/PD-89-1083>
- Mengiste, T., Chen, X., Salmeron, J., & Dietrich, R. (2003). The BOTRYTIS SUSCEPTIBLE1 Gene Encodes an R2R3MYB Transcription Factor Protein That Is Required for Biotic and Abiotic Stress Responses in *Arabidopsis*. *The Plant Cell*, 15(11), 2551–2565. <https://doi.org/10.1105/tpc.014167>
- Mulema, J. M. K., & Denby, K. J. (2012). Spatial and temporal transcriptomic analysis of the *<Emphasis Type="Italic">Arabidopsis thaliana</Emphasis>–<Emphasis Type="Italic">Botrytis cinerea</Emphasis>* interaction. *Molecular Biology Reports*, 39(4), 4039–4049.
<https://doi.org/10.1007/s11033-011-1185-4>
- Paliwal, S., Fiumera, A. C., & Fiumera, H. L. (2014). Mitochondrial-Nuclear Epistasis Contributes to Phenotypic Variation and Coadaptation in Natural Isolates of *Saccharomyces cerevisiae*. *Genetics, genetics*.114.168575. <https://doi.org/10.1534/genetics.114.168575>
- Palma-Guerrero, J., Hall, C. R., Kowbel, D., Welch, J., Taylor, J. W., Brem, R. B., & Glass, N. L. (2013). Genome Wide Association Identifies Novel Loci Involved in Fungal Communication. *PLOS Genetics*, 9(8), e1003669. <https://doi.org/10.1371/journal.pgen.1003669>
- Pavlidis, P., Živković, D., Stamatakis, A., & Alachiotis, N. (2013). SweeD: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. *Molecular Biology and Evolution*, 30(9), 2224–2234. <https://doi.org/10.1093/molbev/mst112>
- Peter, J., Chiara, M. D., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., ... Schacherer, J. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*, 556(7701), 339–344.
<https://doi.org/10.1038/s41586-018-0030-5>
- Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses in R. *Molecular Biology and Evolution*, 31(7), 1929–1936. <https://doi.org/10.1093/molbev/msu136>
- Pinedo, C., Wang, C.-M., Pradier, J.-M., Dalmais, B., Choquer, M., Le Pêcheur, P., ... Viaud, M. (2008). Sesquiterpene Synthase from the Botrydial Biosynthetic Gene Cluster of the Phytopathogen *Botrytis cinerea*. *ACS Chemical Biology*, 3(12), 791–801. <https://doi.org/10.1021/cb800225v>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, 155(2), 945–959.
- Ramasamy, R. K., Ramasamy, S., Bindroo, B. B., & Naik, V. G. (2014). STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. *SpringerPlus*, 3(1), 431. <https://doi.org/10.1186/2193-1801-3-431>
- Rambaut, A. and Drummond, A. (2008). FigTree: Tree figure drawing tool, version 1.2. 2. *Institute of Evolutionary Biology, University of Edinburgh*.
- Romanazzi, G., & Feliziani, E. (2014). Chapter 4 - *Botrytis cinerea* (Gray Mold). In S. Bautista-Baños (Ed.), *Postharvest Decay* (pp. 131–146). San Diego: Academic Press. <https://doi.org/10.1016/B978-0-12-411552-1.00004-1>
- Rowe, H. C., & Kliebenstein, D. J. (2008). Complex Genetics Control Natural Variation in *Arabidopsis*

- thaliana Resistance to Botrytis cinerea. *Genetics*, 180(4), 2237–2250.
<https://doi.org/10.1534/genetics.108.091439>
- Rowe, Heather C., & Kliebenstein, D. J. (2007). Elevated Genetic Variation Within Virulence-Associated Botrytis cinerea Polygalacturonase Loci. *Molecular Plant-Microbe Interactions*, 20(9), 1126–1137. <https://doi.org/10.1094/MPMI-20-9-1126>
- Schuhegger, R., Nafisi, M., Mansourova, M., Petersen, B. L., Olsen, C. E., Svatoš, A., ... Glawischnig, E. (2006). CYP71B15 (PAD3) Catalyzes the Final Step in Camalexin Biosynthesis. *Plant Physiology*, 141(4), 1248–1254. <https://doi.org/10.1104/pp.106.082024>
- Schumacher, J., Pradier, J.-M., Simon, A., Traeger, S., Moraga, J., Collado, I. G., ... Tudzynski, B. (2012). Natural Variation in the VELVET Gene bcvel1 Affects Virulence and Light-Dependent Differentiation in Botrytis cinerea. *PLOS ONE*, 7(10), e47840.
<https://doi.org/10.1371/journal.pone.0047840>
- Shen, X., Alam, M., Fikse, F., & Rönnegård, L. (2013). A Novel Generalized Ridge Regression Method for Quantitative Genetics. *Genetics*, genetics.112.146720.
<https://doi.org/10.1534/genetics.112.146720>
- Staats, M., & Kan, J. A. L. van. (2012a). Genome Update of Botrytis cinerea Strains B05.10 and T4. *Eukaryotic Cell*, 11(11), 1413–1414. <https://doi.org/10.1128/EC.00164-12>
- Staats, M., & Kan, J. A. L. van. (2012b). Genome Update of Botrytis cinerea Strains B05.10 and T4. *Eukaryotic Cell*, 11(11), 1413–1414. <https://doi.org/10.1128/EC.00164-12>
- Talas, F., Kalih, R., Miedaner, T., & McDonald, B. A. (2016). Genome-Wide Association Study Identifies Novel Candidate Genes for Aggressiveness, Deoxynivalenol Production, and Azole Sensitivity in Natural Field Populations of Fusarium graminearum. *Molecular Plant-Microbe Interactions*, 29(5), 417–430. <https://doi.org/10.1094/MPMI-09-15-0218-R>
- Talas, F., & McDonald, B. A. (2015). Genome-wide analysis of Fusarium graminearum field populations reveals hotspots of recombination. *BMC Genomics*, 16(1), 996.
<https://doi.org/10.1186/s12864-015-2166-0>
- Taylor, J. W., Hann-Soden, C., Branco, S., Sylvain, I., & Ellison, C. E. (2015). Clonal reproduction in fungi. *Proceedings of the National Academy of Sciences of the United States of America*, 112(29), 8901–8908. <https://doi.org/10.1073/pnas.1503159112>
- Turner, S., Armstrong, L. L., Bradford, Y., Carlson, C. S., Crawford, D. C., Crenshaw, A. T., ... Ritchie, M. D. (2011). Quality Control Procedures for Genome Wide Association Studies. *Current Protocols in Human Genetics / Editorial Board, Jonathan L. Haines ... [et Al.]*, CHAPTER, Unit1.19.
<https://doi.org/10.1002/0471142905.hg0119s68>
- Walker, A.-S., Gladieux, P., Decognet, V., Fermaud, M., Confais, J., Roudet, J., ... Fournier, E. (2015). Population structure and temporal maintenance of the multihost fungal pathogen Botrytis cinerea: causes and implications for disease management. *Environmental Microbiology*, 17(4), 1261–1274.
<https://doi.org/10.1111/1462-2920.12563>
- Wang, M., Roux, F., Bartoli, C., Huard-Chauveau, C., Meyer, C., Lee, H., ... Bergelson, J. (2018). Two-way mixed-effects methods for joint association analysis using both host and pathogen genomes. *Proceedings of the National Academy of Sciences*, 201710980.
<https://doi.org/10.1073/pnas.1710980115>
- Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, 1(6), 80–83.
<https://doi.org/10.2307/3001968>
- Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M., & Yang, T.-L. (n.d.). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*.
<https://doi.org/10.1093/bioinformatics/bty875>
- Zhang, W., Corwin, J. A., Copeland, D., Feusier, J., Eshbaugh, R., Chen, F., ... Kliebenstein, D. J. (2017). Plastic Transcriptomes Stabilize Immunity to Pathogen Diversity: The Jasmonic Acid and Salicylic Acid Networks Within the Arabidopsis/Botrytis Pathosystem. *The Plant Cell*, tpc.00348.2017.

<https://doi.org/10.1105/tpc.17.00348>

Zhou, N., Tootle, T. L., & Glazebrook, J. (1999). *Arabidopsis PAD3*, a Gene Required for Camalexin Biosynthesis, Encodes a Putative Cytochrome P450 Monooxygenase. *The Plant Cell*, 11(12), 2419–2428. <https://doi.org/10.1105/tpc.11.12.2419>

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, 44(7), 821–824. <https://doi.org/10.1038/ng.2310>

Figure 1. Genomic diversity within 97 isolates of *B. cinerea*. Unrooted genomic phylogeny determined based on pairwise SNP differences in the alignments of (A) 97 *B. cinerea* strains to and including the reference T4 and (B) 97 *B. cinerea* strains to and including the reference BO5.10. Branch lengths are proportional to the number of segregating sites that differentiate each pair of isolates. (C) Unrooted mitochondrial phylogeny determined based on pairwise SNP differences to B05.10 in the alignments of 97 *B. cinerea* strains (including BO5.10 reference). Isolates Peachy, Rasp and B24 were removed to enable visualization.

Figure 2. Admixture in *B. cinerea*. (A) **Scree plot** of principal components extracted from all SNP genotypes simultaneously, showing variance contribution based on their eigenvalues. **PCA of genotypes** does not show any (B) geographic or (C) family of plant sampled from clustering, suggesting admixture. (D) **strplot** (Ramasamy et al., 2014) showing admixture between 4 ancestral populations from STRUCTURE (Pritchard et al., 2000) results and using CLUMPP (Jakobsson & Rosenberg, 2007) to align the ten replicates for $K = 4$ (all runs were performed with 1000 burnin period and 1000 MCMC repeats after burnin).

Figure 3. Genome-Wide polymorphism in 87 isolates of *Botrytis cinerea* (A) LD decay estimated from the mean linkage disequilibrium was plotted against pairwise distance in kb of the 87 isolates aligned to T4. (B) Circos (Krzywinski et al., 2009) diagram of 56 contigs detailing (from outer to inner track) haplotype diversity, nucleotide diversity (Nei, 1987) and mean linkage disequilibrium (Kelly, 1997). Coloring and numbers on the outside of the circle show annotated chromosomes (C) Close-up of Chromosome 2.

Figure 4. Virulence diversity of the *B. cinerea* isolates on *Arabidopsis thaliana*. Kernel density estimate plots of phenotype distribution. Phenotypes are least-squared means of the lesion area determined following inoculation of each of the resequenced isolates scored on the (A) Col-0, (B) *npr1*, (C) *coi1* and (D) *pad3* plant genotypes.

Figure 5. Polygenic architecture of virulence. GWAS Manhattan plot of the Wilcoxon SNP *p*-values from visible lesion phenotypes scored on the Col-0 *Arabidopsis* plant genotype. Negative log₁₀ *p*-values are plotted against genomic position, the blue highlighted SNPs are genes (determined from available annotation) found in the overlap of the top 500 SNPs using all GWAS methods (Wilcoxon and bigRR with SNPs from the T4 alignment and GEMMA with SNPs from the BO5.10 alignment), the red horizontal line denotes genome-wide significance.

Table 1. Candidate virulence genes for *B. cinerea* on *A. thaliana*.

Genes of interest (determined from available annotation) found in the overlap of the top 500 SNPs using all GWAS methods, Wilcoxon and bigRR with SNPs from the T4 alignment and GEMMA with SNPs from the BO5.10 alignment. Genes are listed by plant genotype on which they were found to be significant.

Supplemental Table 1. Genetic polymorphism summary.

A summary of genomic sequencing on each isolate. The mean sequencing depth and number of SNPs, insertions and deletions in respect to the T4 and B05.10 genomes for the nuclear genome are listed. The polymorphisms in comparison the B05.10 mitochondrial genome is also provided. The host and location of collection for each isolate is listed.

Supplemental Figure 1. Host collection does not influence LD decay. The LD decay against the mean linkage disequilibrium was plotted against distance in kb using just isolates sampled from (**A**) Vitaceae, (**B**) Rosaceae, (**C**) Solanaceae and (**D**) international collection sites.

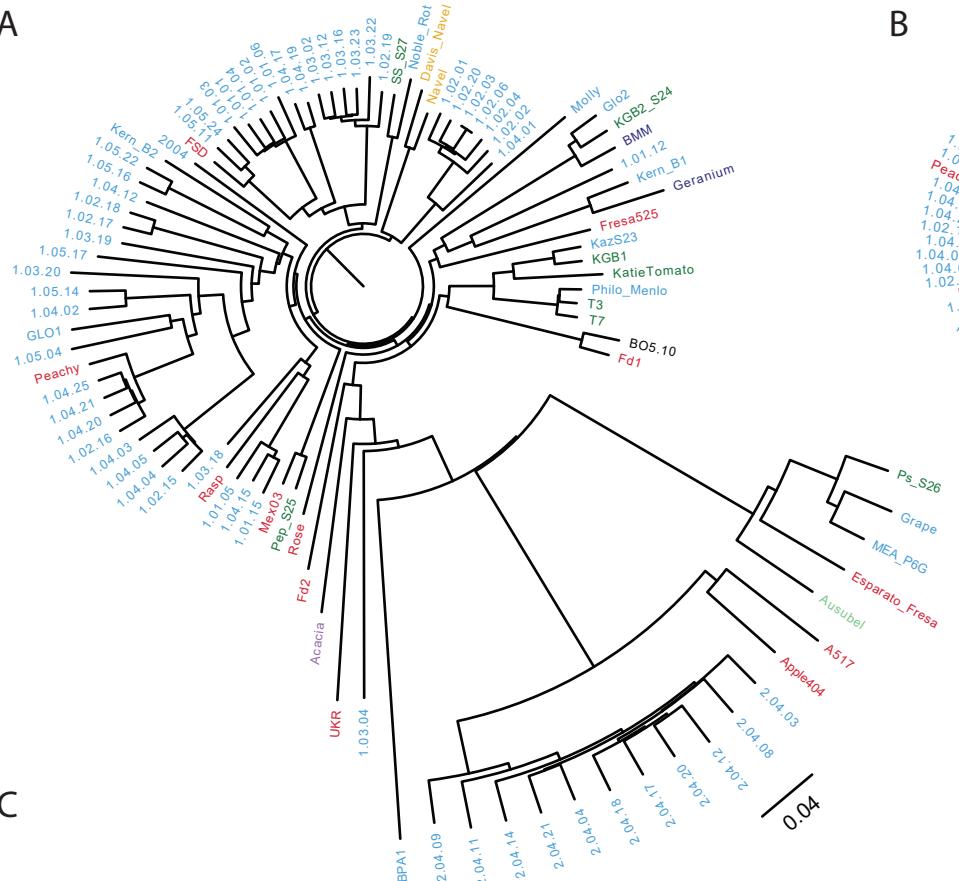
Supplemental Figure 2. Potential selective sweeps. Selective sweeps were determined from the 87 isolates aligned to reference T4 using SweeD_v3.3.2.(Pavlidis et al., 2013) which implements a composite likelihood ratio (*CLR*) test. Blue colored SNPs represent SNPs that cross the 0.95 criterion with d=5kbp.

Supplemental Figure 3. Frequency of major loss of function mutations from the T4 aligned data.

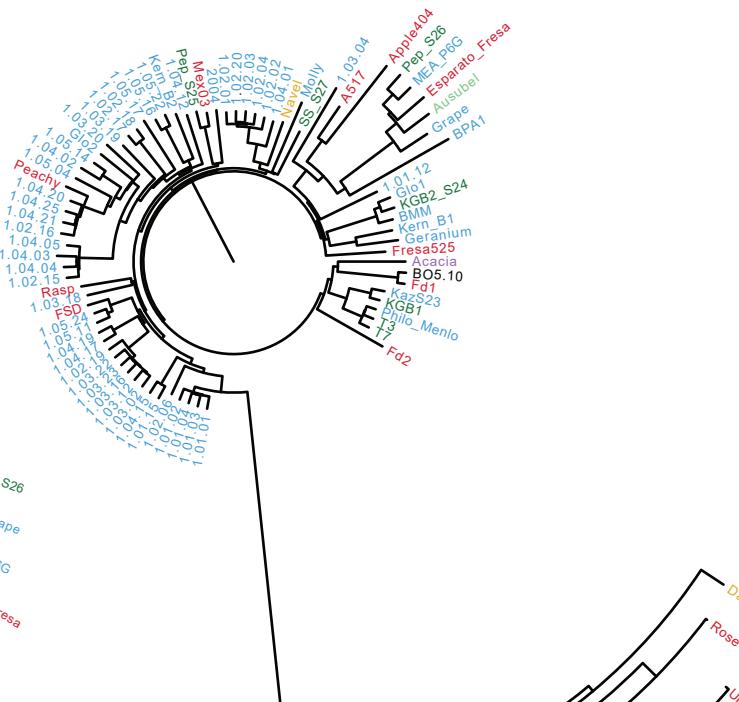
Supplemental Figure 4. Mitochondrial diversity. SNP and Indel Polymorphisms found from the BO5.10 mitochondrial alignment are plotted with the frequency .

Supplemental Figure 5. Polygenic architecture of virulence. GWAS Manhattan plot of the Wilcoxon SNP *p*-values from visible lesion phenotypes scored on the (A) *npr1*, (B) *pad3* and (C) *coi1* plant genotypes. Negative log₁₀ *p*-values are plotted against genomic position, the red and blue lines are suggestive and genome-wide significance lines respectively.

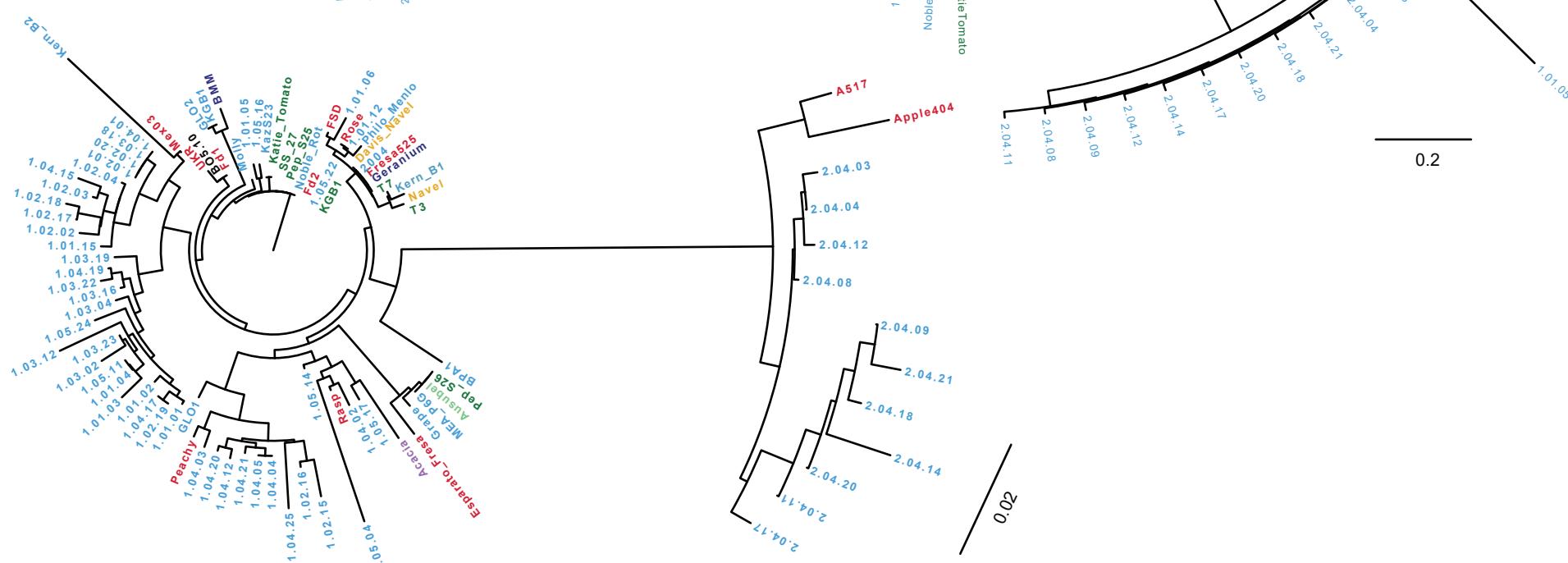
A



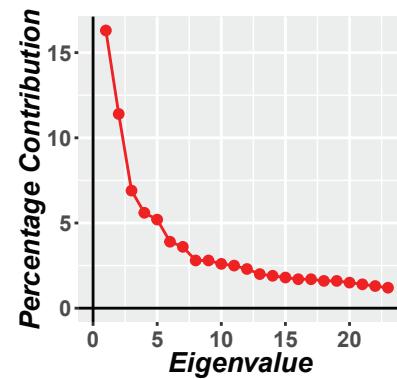
B



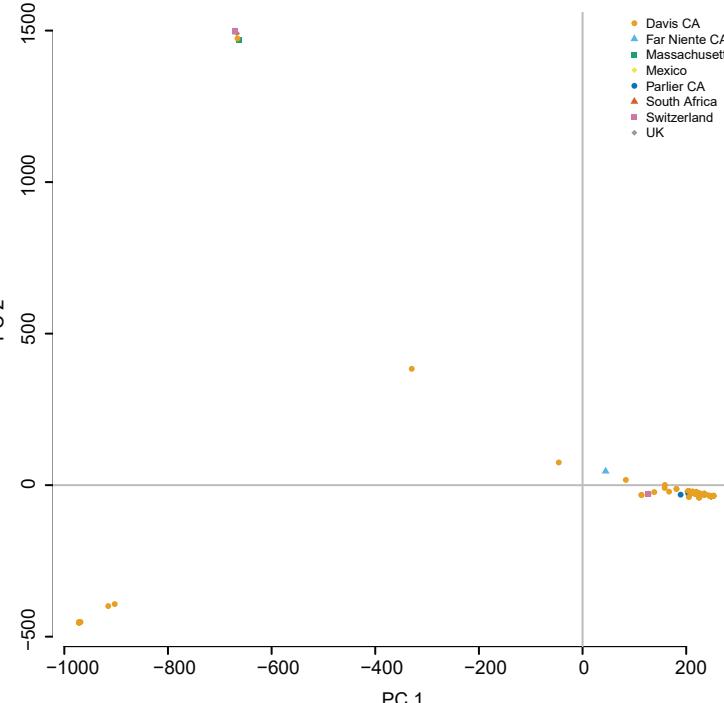
C



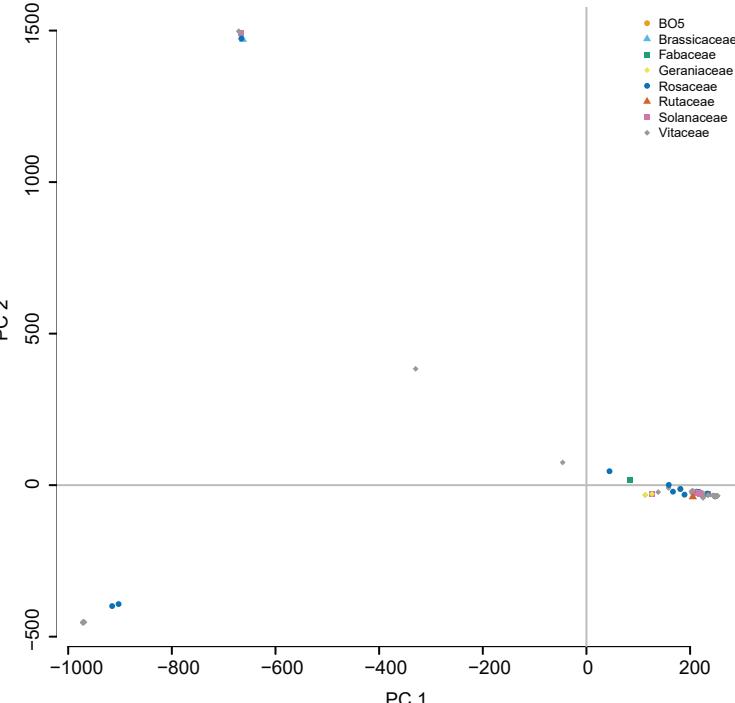
A



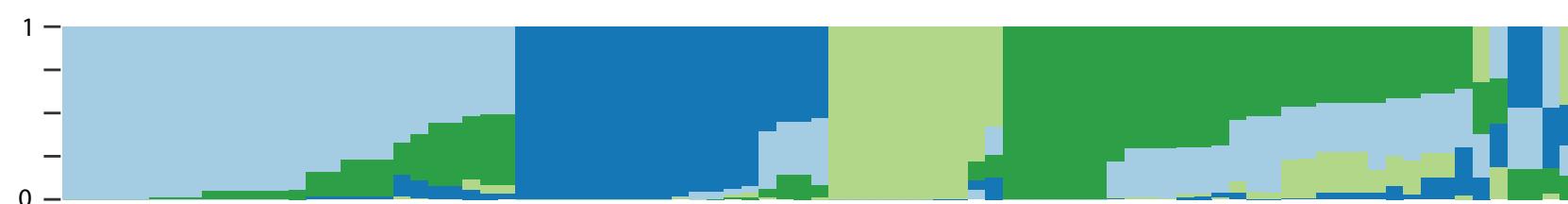
B

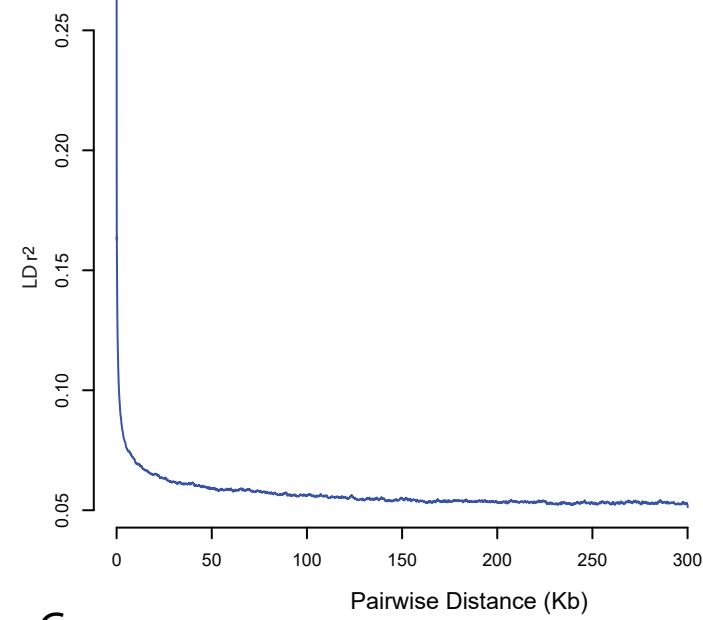
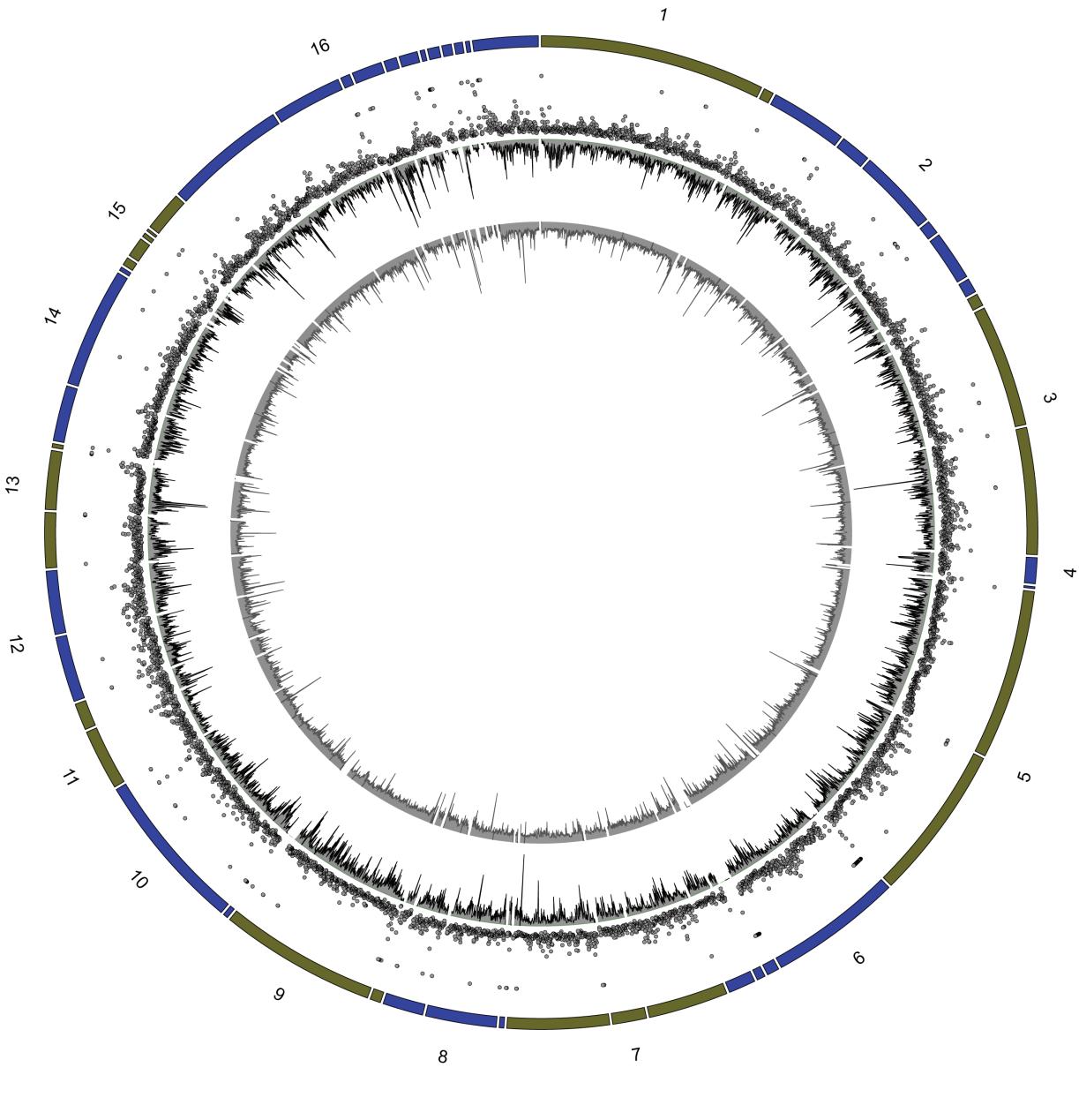


C



D



A**B****C**