POPULATION GENOMICS REVEALS MOLECULAR DETERMINANTS OF SPECIALIZATION TO TOMATO IN THE POLYPHAGOUS FUNGAL PATHOGEN *BOTRYTIS CINEREA*

Alex Mercier[1,2], Adeline Simon[1], Nicolas Lapalu[1], Tatiana Giraud[3], Marc Bardin[4], Anne-Sophie Walker[1†], Muriel Viaud[1†] and Pierre Gladieux[5*†]

[1] Université Paris-Saclay, INRAE, AgroParisTech, UMR BIOGER, 78850 Thiverval-Grignon, France

[2] Université Paris-Saclay, Orsay, France

[3] Ecologie Systématique Evolution, CNRS, Université Paris-Saclay, AgroParisTech, 91400 Orsay, France

[4] UR0407 Pathologie Végétale, INRAE, 84143 Montfavet, France

[5] UMR BGPI, INRAE, CIRAD, Montpellier Supagro, Université de Montpellier, 34398 Montpellier, France

[*] Corresponding author: pierre.gladieux@inrae.fr

[†] contributed equally to this study

Abstract

Many fungal plant pathogens encompass multiple populations specialized on different plant species. Understanding the factors underlying pathogen adaptation to their hosts is a major challenge of evolutionary microbiology, and it should help preventing the emergence of new specialized pathogens on novel hosts. Previous studies have shown that French populations of the grey mould pathogen *Botrytis cinerea* parasitizing tomato and grapevine are differentiated from each other, and have higher aggressiveness on their host-of-origin than on other hosts, indicating some degree of host specialization in this polyphagous pathogen. Here, we aimed at identifying the genomic features underlying the specialization of *B. cinerea* populations to tomato and grapevine. Based on whole genome sequences of 32 isolates, we confirmed the subdivision of *B. cinerea* pathogens into two genetic clusters on grapevine and another, single cluster on tomato. Levels of genetic variation in the different clusters were similar, suggesting that the tomato-specific cluster has not recently emerged following a bottleneck. Using genome scans for selective sweeps and divergent selection, tests of positive selection based on polymorphism and divergence at synonymous and non-synonymous sites and analyses of presence/absence variation, we identified several candidate genes that represent possible determinants of host specialization in the tomato-associated population. This work deepens our understanding of the genomic changes underlying the specialization of fungal pathogen populations.

Keywords: Host specialization; grey mould; gene content variation; selective sweeps; positive selection

Introduction

Many fungal plant pathogens encompass multiple lineages, host races or *formae speciales* specialized on different plant species. Understanding the proximate (*i.e.* molecular) and ultimate (*i.e.* eco-evolutionary) factors underlying adaptation to hosts is a major goal for evolutionary microbiology, because emerging diseases are often caused by the appearance and spread of new pathogen populations specialized onto new hosts (Fisher *et al.* 2012; Stukenbrock & McDonald 2008). Evolutionary theory predicts that pathogen specialization should facilitate the emergence of new populations onto novel hosts, because specialization restricts encounters of potential mates within hosts and reduces the survival of offspring due to maladaptation of immigrants and hybrid offspring, thereby reducing gene flow between ancestral and emerging populations (Giraud *et al.* 2010; Nosil *et al.* 2005). The role of specialization as a barrier to gene flow is expected to be strong for pathogens mating within or onto their hosts because, for individuals evolving the ability to infect novel hosts, mating automatically becomes assortative with respect to host use, and reproductive isolation arises as a direct consequence of adaptive divergence (Gladieux *et al.* 2011; Servedio *et al.* 2011). Evolutionary theory also predicts that specialization, and the associated emergence of new populations, could be facilitated by the molecular basis of plant-pathogen interactions, because compatibility is often determined by a limited number of genes in the host and the pathogen, and selection is more efficient when it acts on a smaller number of genes (Giraud *et al.* 2010; Schulze-Lefert & Panstruga 2011). However, despite the apparent ubiquity of specialized fungal pathogens and the negative impact that specialized populations can have on food security and ecosystem health, the genomic features involved in host specialization

remain largely unknown. Acquiring knowledge about the genomic features underlying pathogen specialization can provide key insights into the mechanisms of specialization.

The ascomycete *Botrytis cinerea* is often presented as a textbook example of a polyphagous plant pathogen, parasitizing more than 1400 host plant species belonging to 580 genera (Elad *et al.* 2016). Yet, previous work revealed population subdivision in *B. cinerea,* with genetic differentiation between populations infecting tomato (*Solanum lycopersicum*) and grapevine (*Vitis vinifera*), respectively, in the field (Walker *et al.* 2015). This population structure in *B. cinerea* was shown to be stable in time and was observed in multiple regions in France. Furthermore, this structure was later associated with differences in performance on the two hosts, with pathogens isolated from tomato being more aggressive on tomato than pathogens isolated from grapevine, and reciprocally (Mercier *et al.* 2019). Altogether, these data were consistent with a certain degree of specialization of *B. cinerea* populations onto these two host plants.

Here, we aimed to identify the molecular basis of host specialization in the *B. cinerea* tomato- and grapevine-associated populations, by addressing the following questions: (1) Can we confirm the genetic subdivision between *B. cinerea* populations from tomato and grapevine using genomic data, and what is the degree of divergence between them? (2) Can we identify genes showing extreme population differentiation, consistent with divergent selection, and what are their predicted functions? (3) Can we identify genes with footprints of positive selection in the genomes of isolates specialized on one host, and what are their predicted functions? (4) Is there variation in gene content between *B. cinerea* populations on tomato and grapevine? To address these questions, we used a set of *B. cinerea* isolates collected on tomato and grapevine in different regions of France. We Illumina-sequenced

their genomes and identified single nucleotide polymorphisms by mapping sequencing reads against a high-quality reference genome (van Kan *et al.* 2017). Because some tests of selection can be biased by population subdivision, while other tests are based on patterns of population differentiation, we first analyzed the population structure of *B. cinerea* collected on tomato and grapevine. To detect genes potentially involved in the specialization of *B. cinerea* to tomato, we searched for signatures of positive selection, by scanning genomes in the tomato population for selective sweeps, and estimating the direction and intensity of selection using McDonald-Kreitman tests on genic sequences. Furthermore, we investigated signatures of divergent selection using an F_{ST} outlier approach, and we characterized variations in the presence/absence of predicted genes between populations collected on tomato and grapevine using *de novo* genome assemblies, gene prediction, and orthology analysis.

Experimental procedures

Sample collection

Botrytis cinerea samples were taken randomly in a collection of isolates collected in three areas of France (Champagne, Occitanie and Provence) and previously characterized using analyses of population structure based on microsatellite markers and pathogenicity tests (Walker *et al.* 2015 ; Mercier *et al.*, 2019). Collection sites were 15 to 133 km apart within regions, and 204 to 722 km apart between regions (Supplementary Figure 1). We chose 32 isolates randomly from the following hosts: (i) tomato (*Solanum lycopersicum*; fruits; 13 isolates), (ii) grapevine (*Vitis vinifera*; berries; 16 isolates), (iii) bramble (*Rubus fruticosus*;

berries; two isolates) and (iv) hydrangea (*Hydrangea macrophylla*; flower buds; one isolate; Table 1). Samples from tomato had been collected either under tunnels (Occitanie region) or in glasshouses (Provence and Champagne regions). Single spores were isolated for each sample and cultured on malt-yeast-agar (MYA; 20 $g.L^{-1}$ malt extract, 5 $g.L^{-1}$ yeast extract, 15 $g.L^{-1}$ agar) at 23°C, under continuous light until sporulation. Single-spored isolates were produced and kept as conidial suspensions in glycerol 20% at -80°C until use.

Pathogenicity tests on tomato plants

Isolates of *B. cinerea* collected from tomato or grape were cultivated on MYA medium in a growth chamber (21°C, 14 hours light) for 14 days. Spores were then washed with sterile distilled water. The cell suspension was filtered through a 30 µm mesh sterile filter to remove mycelium fragments. The conidial concentration was determined with a haemacytometer and adjusted to 10^6 spores/mL. Seeds of tomato var. Clodano (Syngenta) were sown in compost and transplanted after one week in individual pots. Plants were grown in a glasshouse for 6 to 8 weeks where they received a standard commercial nutrient solution once or twice a day, depending on needs. Plants had at least eight fully expanded leaves when inoculated with a single isolate spore solution. For each isolate, five plants were inoculated, from each of which two leaves were removed, leaving 5-10 mm petiole stubs on the stems. Each pruning wound was inoculated with 10 µl of a spore suspension of a given *B. cinerea* isolate at 10^6 spores/mL. Plants were then incubated in a growth chamber in conditions conducive to disease development (21°C, 16h-photoperiod, 162 $\mu mol.s^{-1}.m^{-2}$, relative humidity > 80%). Due to a growth chamber area that did not allow all the isolates to be tested at the same time, nine series of tests were conducted with 10-12 isolates each plus

the BC1 reference isolate, collected in a tomato glasshouse in Brittany (France) in 1989 (Decognet *et al.* 2009). In the end, two to four independent repetitions of the test were performed for each isolate. Lesion sizes (in mm) were assessed daily between the 4th and the 7th day post-infection and the Area Under the Disease Progress Curve (AUDPC; Simko & Piepho 2012) was computed to take into account the kinetics of disease development for each isolate. To compare the aggressiveness of the different isolates tested, an aggressiveness index (AI), relative to the reference isolate BC1, was computed as follows: $AI_{isolate} = 100 \times (AUDPC_{isolate}/AUDPC_{BC1})$, where $AUDPC_{isolate}$ was the average AUDPC for a given isolate and $AUDPC_{BC1}$ is the average AUDPC for the reference isolate BC1. Using the AI index calibrating the AUDPC of a given isolate with that of the reference isolate BC1 in the same test allows comparing isolates aggressiveness while taking into account the variability occurring among assays (e.g. plant physiological state; Leyronas *et al.* 2018). Two statistical analyses were performed with STATISTICA. In the first analysis, the relative aggressiveness indexes of all isolates were compared in a variance analysis, considering the average values for each of the independent tests as replications. In the second analysis, variance analysis was used to compare the aggressiveness of isolates contrasted for their host of origin (tomato vs. grape), considering the independent repetitions of the test as blocks and the 13 isolates from tomato and 16 isolates from grapevine as replicates. For each analysis, when a significant effect was observed, the multiple comparison test of Student-Newman-Keuls was used to compare the means.

DNA preparation and sequencing

Isolates randomly selected for sequencing were cultivated for 48 h on MYA + cellophane medium, at 23 °C and in obscurity to obtain young mycelium. Cultures were then scraped off

and grounded using a mortar and pestle in liquid nitrogen. DNA was extracted using a standard sarkosyl procedure (Dellaporta *et al.* 1983). Paired-end libraries were prepared and sequenced (2 x 100 nucleotides) on a HiSeq4000 Illumina platform at Integragen (Evry, France). Genomic data were deposited at SRA under accession number PRJNA624742. Read quality was checked using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Depth of sequence coverage ranged from 58 to 305 genome equivalents.

SNP calling and filtering

SNPs were detected with the same workflow as described in Zhong *et al.* (2017). Sequencing reads were preprocessed with Trimmomatic v0.36 (Bolger *et al.* 2014). Preprocessed reads were mapped onto the *B. cinerea* B05.10 reference genome (van Kan *et al.* 2017) using BWA v0.7.15 (Li & Durbin 2009). Aligned reads were filtered based on quality using SAMTools v1.3 (Li *et al.* 2009) and Picard Tools (http://broadinstitute.github.io/picard/) to remove secondary alignments, reads with a mapping quality <30 and paired reads not at the expected distance. SNP calling was performed with FreeBayes v1.1 (Garrison & Marth 2012). We kept only biallelic SNPs supported by more than 90% of aligned reads, detected outside low-complexity regions or transposable elements (as identified in the reference isolate B05.10: https://doi.org/10.15454/TFYH9N; Porquier *et al.* 2016) and with coverage lower than twice the standard deviation from the mean depth coverage. Furthermore, the genes previously identified in the small accessory chromosomes BCIN17 and BCIN18 of the isolate B05.10 (23 and 19 genes respectively; van Kan *et al.* 2017) were removed from datasets as these chromosomes are absent in several of the analyzed isolates.

174

175 *Population subdivision*

176 We used the sNMF program to infer individual ancestry coefficients in *K* ancestral
177 populations. This program is optimized for the analysis of large datasets and it estimates
178 individual admixture coefficients based on sparse non-negative matrix factorization, without
179 assuming Hardy-Weinberg equilibrium (Frichot et al., 2014). We used SPLITSTREE version 4 to
180 visualize relationships between genotypes in a phylogenetic network, with reticulations
181 representing the conflicting phylogenetic signals caused by recombination or incomplete
182 lineage sorting. The position of the root was determined using a *B. fabae* isolate as the
183 outgroup. *Botrytis fabae* is one of the closest known relatives of *B. cinerea* (Amselem *et al.*
184 2011; Walker 2016).

185

186 *Genome assembly and gene prediction*

187 Illumina paired-reads were assembled using a combination of VELVET (Zerbino & Birney
188 2008), SOAPDENOVO and SOAPGAPCLOSER (Luo *et al.* 2012). Genomic regions mapping to
189 transposable elements previously identified in *B. cinerea* (https://doi.org/10.15454/TFYH9N;
190 Porquier *et al.* 2016) were masked with REPEATMASKER (http://www.repeatmasker.org) prior
191 to gene prediction. Genes were predicted using the FGENESH *ab initio* gene-finder (Solovyev
192 *et al.* 2006; http://www.softberry.com/berry.phtml) with the provided *Botrytis* specific
193 gene-finding parameters.

194 We used ORTHOFINDER (Emms & Kelly 2015) in order to define groups of orthologous
195 sequences (hereafter "groups of orthologs") based on sequences of predicted genes

translated into protein sequences. To reduce the impact of incomplete gene prediction (*e.g.* truncated genes in small contigs), groups of orthologs were then manually checked for the presence/absence of the protein-encoding genes in the genomes using TBLASTN (Johnson *et al.* 2008). Proteins were functionally annotated using INTERPROSCAN (Jones *et al.* 2014), and SIGNALP (Almagro Armenteros *et al.* 2019) was used to predict secretion signal peptides. As we detected traces of bacterial contamination in the genome of the Sl13 isolate, this genome was not included in orthology analysis. However, we have kept this genome for analyzes based on polymorphism, because bacterial reads cannot map to the reference genome.

Tests of positive selection based on polymorphism and divergence at synonymous and non-synonymous sites

We estimated the intensity and direction of selective pressures exerted on genes in populations using the McDonald-Kreitman test based on polymorphism and divergence at synonymous and non-synonymous sites (McDonald & Kreitman 1991). This test is based on the number of nucleotide polymorphisms and substitutions in gene sequences, and assumes that synonymous mutations are neutral. Pseudo-sequences for all genes in the reference genome were generated using the VCF file and the reference sequence in Fasta format. Numbers of substitutions were computed using *B. fabae* as the outgroup. Computations were carried out using a python script available on Github (https://github.com/alexrjcmercier/mcdonald_kreitman_test). The neutrality index, defined as $(P_N/D_N)/(P_S/D_S)$ (Stoletzki & Eyre-Walker 2011), was computed for every gene, with P_N and D_N the numbers of non-synonymous polymorphisms and substitutions, respectively, and P_S

and D_S the numbers of synonymous polymorphisms and substitutions, respectively. Pseudocounts of one were added to each cell of the McDonald-Kreitman tables to ensure the NI was always defined (*i.e.* no division by zero).

Tests of positive selection based on linkage disequilibrium and the site frequency spectrum

We used POPLDDECAY (Zhang *et al.* 2019) to measure linkage disequilibrium (r^2) as a function of the distance between pairs of SNPs. POPLDDECAY was configured with a maximum distance between SNPs of 300 kbp, a minor allele frequency of 0.005 and a maximum ratio of heterozygous allele of 0.88. The dataset used for this analysis did not have missing data.

We used window sizes of 10, 50 and 100 kbp for genome scans. We detected signatures of selective sweeps along genomes using three different softwares, each implementing a different approach. The SWEED software (Pavlidis *et al.* 2013) implements a composite likelihood ratio (*CLR*) test based on the SWEEPFINDER algorithm (Nielsen *et al.* 2005), which uses the site frequency spectrum (SFS) of a locus to compute the ratio of the likelihood of a selective sweep at a given position to the likelihood of a null model hypothesis without selective sweep. The nS_L method implemented in the NSL software (Ferrer-Admetlla *et al.* 2014) is based on the decay of the haplotype homozygosity as a function of recombination distance along chromosomes. Both *CLR* and nS_L are population-specific statistics and analyses were thus run independently for each population. Lastly, the *hapFLK* method (Fariello *et al.* 2013) implemented in the eponym HAPFLK software is based on the original *FLK* method by Bonhomme *et al.* (2010). This metric can be used to test the null hypothesis of neutrality by contrasting allele frequencies at a given locus in different populations.

241 *hapFLK* extends the FLK method to account for the haplotype structure in the sample, and
242 the method is robust to the effects of bottlenecks and migration.
243

244 ## Results

245 *Whole-genome sequencing and analysis of population structure*

246 Previous work using microsatellite data revealed differentiation between populations of *B.*
247 *cinerea* occurring on tomato and grapevine (Mercier *et al.* 2019; Walker *et al.* 2015). To
248 investigate the genetic basis of specialization of the tomato- and grapevine-infecting
249 populations, we randomly selected for sequencing 32 isolates collected on tomato (13
250 isolates) and grapevine (16 isolates), *Rubus* (two isolates) and *Hydrangea* (one isolate).
251 Isolates from *Rubus* and *Hydrangea* were previously shown to belong to generalist
252 populations (*i.e.* assigned to clusters found on all sampled hosts; Mercier *et al.* 2019). One
253 isolate of the sister species *B. fabae* was also sequenced and used as the outgroup.
254 Information about the sequenced isolates is summarized in Table 1. Alignment of sequencing
255 reads to the B05.10 reference genome (van Kan *et al.* 2017) followed by SNP calling
256 identified 249,084 high-quality SNPs.

257 In clustering analyses based on sparse nonnegative matrix factorization algorithms, as
258 implemented in the sNMF program (Frichot et al., 2014), the model with *K*=4 clusters was
259 identified as the best model on the basis of the cross-entropy criterion (Supplementary
260 Figure 2) and models with K>4 mostly further subdivided the clusters identified at *K*=4. One
261 cluster was associated with tomato (hereafter referred to as "Sl" cluster after the Latin name
262 of the host), two clusters were associated with grapevine (hereafter referred to, after the

Latin name of the host, as "Vv1" for the largest, and "Vv2" for the smallest), and one cluster was formed by the isolates Rf1 and Hm1 from *R. fruticosus* and *H. macrophylla*, respectively (Figure 1). The isolate Rf2 collected on wild blackberry displayed ancestry in multiple clusters, and the reference isolate B05.10 had ancestry in the Vv2 and Sl clusters. No pattern of geographical subdivision was observed, consistent with previous findings (Walker *et al.* 2015; Mercier *et al.,* 2019). The neighbor-net network inferred with SPLITSTREE revealed three main groups, corresponding to the three clusters identified with sNMF, with cluster Vv2 branching more closely from Sl than from Vv1 (Figure 2). Together, analyses of population subdivision revealed three clearly defined populations (two on grapevine, and one on tomato) and we therefore focused on these populations to identify the genes underlying differences in host specialization.

On average across core chromosomes BCIN01 to BCIN16, nucleotide diversity π and Tajima's D were comparable in the three clusters (π : from 0.044/bp in Vv1 to 0.056/bp in Sl; D: from 1.00 in Sl to 1.25 in Vv2; Supplementary Table 2). The distance at which linkage disequilibrium (LD) decayed to 50% of its maximum was an order of magnitude longer in Vv2 (LDdecay50: 16,800bp), than in Sl (LDdecay50: 8600bp) and Vv1 (LDdecay50: 3100bp) (Supplementary Figure 3).

Isolates from the Sl population are more aggressive on tomato plants.

To test whether isolates collected on tomato are more aggressive on their host of origin, compared to isolates collected on grapevine, independent pathogenicity assays were achieved on whole tomato plants in controlled conditions (Figure 3). Isolate effect was found significant (ANOVA, $p < 0.001$), as confirmed by the wide range of aggressiveness levels

observed for the 29 isolates. We also observed a significant effect of the host of origin (ANOVA, $p < 0.001$) and Sl isolates collected on tomato were on average 2.7 times more aggressive on tomato plants than Vv1 or Vv2 isolates collected on grapevine.

Gene content of tomato- and grapevine-associated populations

As variation in gene content can be involved in adaptation to novel hosts (Cummings *et al.* 2004; Inoue *et al.* 2017; Langridge *et al.* 2015), we sought to identify genes specific to the tomato (Sl) and grapevine (Vv1 and Vv2) populations. We built *de novo* assemblies of the genomes of Sl, Vv1 and Vv2 isolates. The genome assembly size of the 28 isolates ranged from 41 Mb to 42.5 Mb (Supplementary Table 1), which was slightly smaller than the genome assembly size of the B05.10 reference isolate (42.6 Mb). We predicted genes *ab initio* using FGENESH. The number of predicted genes ranged from 11,109 to 11,311. In order to compare the sets of proteins in the Sl, Vv1 and Vv2 populations, we used ORTHOFINDER to identify 12,319 groups of homologous sequences. The number of groups of homologs shared by pairs of isolates within populations was higher that between populations (Supplementary Table 3). By looking for orthologous groups that were present in at least 75% of the genomes of a focal population and missing in other populations, we identified 21 Vv1-specific genes, a single Vv2-specific gene, five genes specific to the Vv populations, two genes missing specifically in the Vv1 population, and a single gene specific to Sl (Supplementary Table 4). Functional annotations and signal peptides were predicted using INTERPROSCAN (Jones *et al.* 2014) and SIGNALP (Almagro Armenteros *et al.*, 2019), respectively.

The only gene specific to Sl was a family 71 glycoside hydrolase (OG0011490, an α-1,3-glucanase; Lombard *et al.* 2014) acting on fungal cell wall. Among the genes specific to Vv1

and Vv2, we identified four adjacent genes (in OG0010989, OG0010990, OG0010991 and OG0010997) that could correspond to a secondary metabolism gene cluster. Among these four genes, one is coding for a NRPS-like enzyme similar to the protein MelA of *Aspergillus terreus* involved in the biosynthesis of an α-keto acid dimer (Geib *et al.* 2016), two other genes encode putative biosynthetic enzymes (FAD-binding and enoyl reductase domains) while the fourth gene had no functional prediction. Among the genes specific to Vv1, we found a family 10 glycoside hydrolase (in OG0011469; Lombard *et al.* 2014) acting on plant cell wall (*i.e.* hemicellulose). The proteins encoded by the other Vv1-specific genes had no functional prediction though four of them shared a domain typical of metalloenzymes (IPR11249) with putative peptidase activities and three other ones showed a versatile protein-protein interaction motif involved in many functions (IPR011333). Two proteins with secretion signal peptides were also found specific to Vv1 (OG0011305 and OG0011366), with only OG0011366 having a predicted function of interferon alpha-inducible protein-like (Rosebeck & Leaman 2008). Together, these analyses revealed a limited number of gene gains or losses between *B. cinerea* populations, which emphasizes the need to investigate differences in allelic content at shared genes for elucidating the genomic basis of host specialization.

McDonald-Kreitman tests of positive selection

We investigated differences in the direction and intensity of natural selection driving the evolution of gene sequences in the two populations with the greatest difference in terms of quantitative pathogenicity on grape and tomato (Vv1 and Sl). In particular, we searched for genes with signatures of positive selection in one population, but not in the other

population. The direction and intensity of selection was estimated using neutrality indexes computed for each individual gene in each population based on McDonald-Kreitman tables of polymorphisms and substitutions at synonymous and non-synonymous sites, using *B. fabae* as the outgroup. The neutrality index is expected to be below one under positive selection (due to an excess of non-synonymous substitutions) and above one under negative selection (due to a deficit of non-synonymous polymorphisms). To identify genes potentially involved in host specialization, we selected genes having both low neutrality index values in one population (log [neutrality index] ≤-0.5), and high values in the other (log [neutrality index]>=0). This analysis identified 136 genes in Vv1 and 168 genes in Sl (Supplementary Table 5 and Supplementary Table 6). Candidate genes appeared to be evenly distributed along chromosomes, with highest densities in the BCIN01 contig for Vv1 and in the BCIN05 contig for Sl. Functional enrichment analyses revealed contrasting results in the Vv1 and Sl populations (Supplementary Table 7). In the Vv1 population, we identified a two-fold significant enrichment in transporter-encoding genes among the 136 genes with signatures of positive selection. These transporters corresponded to putative amino-acid transporters (two genes), mitochondrial transporters (two genes) as well as proteins of the major facilitator superfamily (MFS; four genes) that could have roles in various processes, including obtaining nutrients from the host, efflux of fungi-toxic compounds or the export of fungal phytotoxins (Hartmann *et al.* 2018; Maruthachalam *et al.* 2011; Perlin *et al.* 2014). In addition, the McDonald-Kreitman test indicated that the gene coding for the key enzyme in abscisic acid production (*Bcaba3;* Otto *et al.* 2019) was under positive selection in the Vv1 population. In the Sl population, the list of 168 genes with signatures of positive selection showed a two-fold enrichment in genes encoding for proteins that are secreted at the early

355 stage of germination on media that contain plant extracts (seven genes; (Espino *et al.* 2010);
356 Supplementary Table 7). More generally, the list included nine Carbohydrate-Active Enzymes
357 (CAZymes) with at least four of them acting on the plant cell wall by acting either on
358 cellulose (glycosides hydrolases GH7 and GH61) or on pectin (GH28 and GH88). The *Bcnep1*
359 gene that encodes a phytotoxic protein (Schouten *et al.* 2008) was also found under positive
360 selection in the Sl population. It is worth noting that this gene was previously shown to be
361 under positive selection across species in the *Botrytis* genus (Staats et al., 2007).
362
363 *Genome scans for selective sweeps*
364 We scanned the genomes of Vv1 and Sl populations for signatures of selective sweeps, by
365 computing metrics of selective sweeps in non-overlapping 10, 50 and 100 kb windows.
366 Because results were consistent across window sizes, and distance at which LD decayed to
367 half of its maximum value in 3 to 17kb in the three populations, only the results for 10 kb
368 windows are presented. To identify genomic regions with signatures of selective sweeps, we
369 conducted three different genome scans, using different features of the data: hapFLK, based
370 on patterns of differentiation between populations; nS_L, based on patterns of linkage
371 disequilibrium within populations; and SweeD's CLR, based on the site frequency spectrum.
372 The *CLR* and nS_L metrics are population-specific and were computed for each population
373 independently, while the *hapFLK* metric is F_{ST}-based and was thus computed for populations
374 Vv1 and Sl. For each metric, we identified outliers as being in the top 5% quantile. Out of the
375 569 and 577 windows in the top 5% quantile of at least one of the three metrics for Vv1 and
376 Sl, respectively, 59 and 52 were in the top 5% quantile of at least two metrics, and five and
377 four in the top 5% quantile of all three metrics. Results are presented in Figure 5 and a

summary of the distribution of regions detected as being under positive selection can be found in Supplementary Table 8. For regions detected in the top 5% quantile with at least two selective sweep metrics, 6.8% and 1.9% were found within 10 kb of centromeric regions, and 22.0% and 21.2% within 10 kb of a transposable element. Among windows in the top 5% for the three metrics, for Vv1 and Sl populations respectively, 16.9% and 15.4% were in the bottom 5% quantile of Tajima's *D*, 20.3% and 11.5% were in the bottom 5% quantile of Fay & Wu's *H*, and 50.8% and 63.5% were in the top 5% quantile of F_{ST} (Supplementary Figure 4).

Windows in the top 5% quantile of the three metrics encompassed 21 and 15 genes in Vv1 and Sl populations, respectively, with 11 and 9 having a predicted function. Windows in the top 5% quantile of at least two metrics encompassed 188 and 155 annotated genes in Vv1 and Sl populations, respectively. No significant functional enrichment was detected (Supplementary Table 7) and the regions did not include any of the genes with signatures of positive selection according to McDonald-Kreitman tests. Candidate genes in the selective sweep regions of each population included genes coding for proteins with functions consistent with a role in infections, such as transporters, CAZymes and other secreted proteins. It is unlikely that all genes in selective sweep regions have been direct targets of positive selection, most of them being possible hitch-hikers; however, we note that one selective sweep region identified in the Sl population (chromosome BCIN14) exclusively contains the two genes of an uncharacterized secondary metabolism gene cluster *i.e.* the polyketide synthase gene *Bcpks11* and the non-ribosomal peptidase synthase *Bcnrps9* (Amselem et al., 2011). This suggests that the synthesis of the corresponding unknown metabolite may have been under positive selection in the Sl population.

Discussion

Genetic differentiation in B. cinerea *between generalist populations found on grapevine and a specialist population found on tomato*

Previous research concluded that *B. cinerea* populations were subdivided according to some of their hosts, including tomato, grapevine and to a lesser extent, bramble (Fournier & Giraud 2008; Walker *et al.* 2015). More recently, we have shown that *B. cinerea* populations parasitizing tomato are specialized to this host (Mercier *et al.* 2019). Here we show that populations parasitizing tomato and grapevine are subdivided into three populations, two being associated with grapevine (Vv1 and Vv2), and one with tomato (Sl). This pattern of population genetic structure differs from previous findings (Walker et al., 2015), as only a single population parasitizing grapevine had previously been detected, and but the difference with our study in terms of population structure can be explained by the higher resolving power of our 249,084 SNP dataset compared to the set of 8 SSR markers used previously.

Multiple factors can contribute to reduce gene flow between generalist populations parasitizing grapevine and the specialist population parasitizing tomato. A first possible factor limiting gene flow is adaptation to host. Mating in *B. cinerea* occurs on the host after infection, between individuals that were thus sufficiently adapted to infect the same host, which induces assortative mating with respect to host use and reduce opportunities for inter-population crosses (Giraud et al. 2006; Giraud 2008; Giraud et al. 2010). Another factor possibly limiting gene flow between populations infecting tomato and grape is habitat isolation (*i.e.* reduced encounters caused by mating in different habitats). Tomatoes are grown in nurseries before being dispatched to the fields or greenhouses, and this may

424 generate habitat isolation if sexual reproduction in the pathogen occurs in nurseries for the
425 tomato-infecting population of *B. cinerea*. Such habitat isolation may contribute to promote
426 adaptation to tomato, by preventing the immigration of alleles which are favorable for
427 infection of non-tomato hosts but not favorable for infection of tomato.
428
429 *Widespread signatures of selection along genomes*
430 We identified little variation in the gene content of Sl, Vv1 and Vv2 populations, with one
431 gene specific to Sl, 22 genes specific to to Vv1 and five genes shared between Vv1 and Vv2
432 but not Sl, suggesting that gene gain or loss is not the main driver of adaptation to tomato,
433 or that variation in gene content was not captured by our sequencing approach based on
434 short reads. In parallel to our analysis of presence/absence variation, our genome scans for
435 positive selection allowed us to pinpoint several genomic regions which may harbour
436 determinants of ecological differentiation between the population specialized to tomato and
437 the population parasitizing grapevine. In order to cover multiple time scales and different
438 positive selection signatures, we used a variety of analytical approaches. The McDonald-
439 Kreitman test focuses on genes and detects repeated episodes of selective sweeps fixing
440 non-synonymous substitutions, thus generating a higher ratio of amino acid divergence to
441 polymorphism (Ka / Pa), relative to the ratio of silent divergence to polymorphism (Ks / Ps),
442 than expected under neutrality. The values of nucleotide diversity and Tajima's D measured
443 in the Sl population specialized in tomatoes were very close to the values measured for the
444 two other populations, which is not consistent with a very recent origin of this population
445 and justifies the use of the McDonald-Kreitman test. Genome scans for selective sweeps and
446 differentiation outliers detect more recent events, and by construction these methods can

also detect genes that are not directly the target of selection, but may have hitch-hiked due to physical linkage with sites under positive selection. However, the LD decay values measured for the Sl population remain moderate (9kb), and the combination of different genome scan approaches substantially shortened the list of candidate genes (15 genes in the top 5% of three metrics and 162 in the top 5% of two metrics), which suggests that the impact of genetic hitch-hiking on our list of genes under recent positive selection remains limited. The majority of 10kb windows in the top 5% of the three selective sweep metrics were also in the top 5% of F_{ST} values, which indicates that our approach effectively targets regions under divergent selection.

Genes under divergent natural selection

We identified a number of genes showing signatures of natural selection using the approaches discussed above and highlighting potential candidates for their role in host specialization. Functional annotation of the *B. cinerea* genome and previous experimental studies provided lists of genes involved in host-pathogen interaction and in other developmental processes (Espino *et al.*, 2010; Amselem *et al.* 2011; Nakajima & Akutsu 2014; Rodriguez-Moreno *et al.* 2018). We used these published lists of genes to investigate whether some specific biological processes are subjected to positive selection in the different populations. Our data revealed that the 163 genes showing signatures of selection in the Sl population with McDonald-Kreitman tests were enriched in genes coding for secreted proteins that were detected at the early stage of germination on media that contain plant extracts suggesting that they could be involved in the infection process (Espino *et al.* 2010). Notably, several secreted CAZymes targeting compounds of the plant cell wall,

i.e. cellulose and pectin (Amselem et al., 2011; Lombard *et al.* 2014), could be identified among these 163 genes, and additional Plant Cell Wall Degrading Enzymes (PCWDEs) were found in genomic regions identified as selective sweeps. Necrotrophic species have important repertoires of CAZymes especially those corresponding to PCWDEs which are known to act as major virulence factors in fungi (Zhao *et al.* 2013; Rodriguez-Moreno *et al.* 2018). The genome of the reference isolate of *B. cinerea* (B05.10) revealed 118 PCWDEs (Amselem et al., 2011) and our data suggest that, within this repertoire, some cellulases and pectinases may be of particular importance for the degradation of tomato cell wall.

In addition to PCWDEs, the unique gene that we found as present in the Sl population but not in the Vv1 and Vv2 populations is a CAZyme acting on the fungal cell wall, an α-1,3-glucanase classified as a member of the GH71 family. In the fungal cell wall, α-1,3-glucan is a major component that encloses the α-(1,3)-glucan-chitin fibrillar core. Because of its external localization and specific composition, α-1,3-glucan of pathogenic fungi plays a major role in infection-related morphology and host recognition (Beauvais *et al.* 2013; King *et al.* 2017). A dozen of genes of *B. cinerea* encode for enzymes of the GH71 family (Amselem et al., 2011). The Sl-specific GH71 CAZyme might therefore have been retained in the Sl population as a mean to specifically facilitate infection of tomato by modification of the fungal cell wall resulting in adaptive morphological changes or impairment of host recognition.

Our genome scans also pointed out some small molecules possibly acting in the interaction with the tomato host. The first evidence was the identification of the gene that encodes the small phytotoxic protein NEP1 (Schouten et al., 2008). This gene is not only under positive selection in the Sl population but also across *Botrytis* species parasitizing

493 different host plants (Staats et al., 2007). Our data therefore suggest the importance of the
494 toxic protein NEP1 in the interaction of *B. cinerea* with tomato. In addition to small secreted
495 proteins, fungal secondary metabolites could also act as effectors in the infection process
496 (Rodriguez-Moreno *et al.* 2018; Scharf *et al.* 2014). In *B. cinerea*, the sesquiterpene botrydial
497 and the polyketide botcinic acid were shown to be virulence factors (Dalmais et al., 2011). In
498 our study, a genomic region involved the synthesis of another, unknown, secondary
499 metabolite was identified in a selective sweep region in the SI population by all three
500 genome scan methods. The gene cluster includes two key genes *i.e.* a polyketide synthase
501 (*Bcpks11)* and a non-ribosomal peptide synthetase (*Bcnrps9*). These two genes are co-
502 localized with other genes that could contribute to further steps in the synthesis of
503 metabolites, *i.e.* genes encoding two P450-monooxygeneases and one amino-transferase. It
504 is not yet known whether this cluster of genes is responsible for the synthesis of both one
505 polyketide and one peptide or for a single hybrid compound. Based on our results, it could
506 be hypothesized that the resulting metabolite(s) may be particularly important in the
507 specialization of the SI population to its tomato host. In a recent comparative genomic
508 analysis, it was suggested that some PKS-NRPS hybrids may be involved in the capacity of the
509 wheat pathogens *Eutiarosporella darliae* and *E. pseudodarliae* to infect woody plants as host
510 reservoirs, thus shaping the host range and evolution pattern of these emerging wheat
511 pathogens (Thynne *et al.* 2019).
512
513 *Concluding remarks*
514 We identified a population of *B. cinerea* specialized to tomato, which diverged from a more
515 generalist, grapevine-associated population. Genome scans for selective sweeps and

McDonald-Kreitman tests revealed widespread signatures of positive selection that identify genes that may contribute to the pathogen's adaptation to its host tomato. Candidate genes for specialization to tomato included PCWDEs, key enzymes in secondary metabolite clusters, and transporters of ABC or MFS families. Our work sets the stage for future studies aiming to elucidate the phenotypic and fitness effects of the candidate genes for specialization of *B. cinerea* to tomato, for instance by knocking-out or replacing candidate genes for host specialization.

Acknowledgements

AM was supported by a grant from the Doctoral School « Sciences du Végétal », Université Paris-Saclay. We are grateful to INRAE-LIPM and Genotoul Bioinformatics Platforms Toulouse Occitanie for providing computational support and resources. This work was supported by a grant overseen by the French National Research Agency (ANR) as part of the "Investments d'Avenir" Programme (LabEx BASC; ANR-11-LABX-0034) and by the INRAE department SPE. The BIOGER unit is also grateful to the "Saclay Plant Science" EUR.

References

Almagro Armenteros JJ, Tsirigos KD, Sonderby CK, *et al.* (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature biotechnology* **37**, 420-423.

Amselem J, Cuomo CA, van Kan JA, *et al.* (2011) Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS genetics* **7**, e1002230.

Beauvais A, Bozza S, Kniemeyer O, *et al.* (2013) Deletion of the α-(1,3)-glucan synthase genes induces a restructuring of the conidial cell wall responsible for the avirulence of *Aspergillus fumigatus*. *PLoS pathogens* **9**, e1003716.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120.

Bonhomme M, Chevalet C, Servin B, *et al.* (2010) Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics* **186**, 241-262.

Cummings CA, Brinig MM, Lepp PW, van de Pas S, Relman DA (2004) Bordetella species are distinguished by patterns of substantial gene loss and host adaptation. *Journal of bacteriology* **186**, 1484-1492.

Dalmais B, Schumacher J, Moraga J, *et al.* (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**, 564-579.

Decognet V, Bardin M, Trottin-Caudal Y, Nicot PC (2009) Rapid change in the genetic diversity of *Botrytis cinerea* populations after the introduction of strains in a tomato glasshouse. *Phytopathology* **99**, 185-193.

Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA minipreparation: Version II. *Plant Molecular Biology Reporter* **1**, 19-21.

Elad Y, Pertot I, Prado AMC, Stewart A (2016) Plant hosts of Botrytis spp. In: *Botrytis–the fungus, the pathogen and its management in agricultural systems*, pp. 413-486. Springer.

Emms DM, Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome biology* **16**, 157.

Espino JJ, Gutierrez-Sanchez G, Brito N, *et al.* (2010) The *Botrytis cinerea* early secretome. *Proteomics* **10**, 3020-3034.

Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B (2013) Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics* **193**, 929-941.

Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R (2014) On detecting incomplete soft or hard selective sweeps using haplotype structure. *Molecular biology and evolution* **31**, 1275-1291.

Fisher MC, Henk DlA, Briggs CJ, *et al.* (2012) Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484**, 186.

Fournier E, Giraud T (2008) Sympatric genetic differentiation of a generalist pathogenic fungus, *Botrytis cinerea*, on two different host plants, grapevine and bramble. *Journal of Evolutionary Biology* **21**, 122-132.

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. *arXiv:1207.3907 [q-bio]*.

Geib E, Gressler M, Viediernikova I, *et al.* (2016) A non-canonical melanin biosynthesis pathway protects *Aspergillus terreus* ionidia from environmental stress. *Cell chemical biology* **23**, 587-597.

Giraud T, Gladieux P, Gavrilets S (2010) Linking the emergence of fungal plant diseases with ecological speciation. *Trends in ecology & evolution* **25**, 387-395.

Gladieux P, Guerin F, Giraud T, *et al.* (2011) Emergence of novel fungal pathogens by ecological speciation: importance of the reduced viability of immigrants. *Molecular ecology* **20**, 4521-4532.

576 Hartmann FE, McDonald BA, Croll D (2018) Genome-wide evidence for divergent selection between
577 populations of a major agricultural pathogen. *Molecular ecology* **27**, 2725-2741.
578 Inoue Y, Vy TTP, Yoshida K, *et al.* (2017) Evolution of the wheat blast fungus through functional losses in a host
579 specificity determinant. *Science* **357**, 80-83.
580 Johnson M, Zaretskaya I, Raytselis Y, *et al.* (2008) NCBI BLAST: a better web interface. *Nucleic acids research* **36**,
581 W5-W9.
582 Jones P, Binns D, Chang HY, *et al.* (2014) InterProScan 5: genome-scale protein function classification.
583 *Bioinformatics* **30**, 1236-1240.
584 King R, Urban M, Lauder RP, *et al.* (2017) A conserved fungal glycosyltransferase facilitates pathogenesis of
585 plants by enabling hyphal growth on solid surfaces. *PLoS pathogens* **13**, e1006672.
586 Langridge GC, Fookes M, Connor TR, *et al.* (2015) Patterns of genome evolution that have accompanied host
587 adaptation in *Salmonella*. *Proceedings of the National Academy of Sciences of the United States of*
588 *America* **112**, 863-868.
589 Leyronas C, Bardin M, Berthier K, *et al.* (2018) Assessing the phenotypic and genotypic diversity of *Sclerotinia*
590 *sclerotiorum* in France. *European journal of plant pathology* **152**, 933-944.
591 Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*
592 **25**, 1754-1760.
593 Li H, Handsaker B, Wysoker A, *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*
594 **25**, 2078-2079.
595 Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The Carbohydrate-Active enZymes
596 database (CAZy) in 2013. *Nucleic acids research* **42**, D490-495.
597 Luo R, Liu B, Xie Y, *et al.* (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo
598 assembler. *GigaScience* **1**, 18.
599 Maruthachalam K, Klosterman SJ, Kang S, Hayes RJ, Subbarao KV (2011) Identification of pathogenicity-related
600 genes in the vascular wilt fungus *Verticillium dahliae* by *Agrobacterium tumefaciens*-mediated T-DNA
601 insertional mutagenesis. *Molecular Biotechnology* **49**, 209-221.
602 McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* **351**, 652-
603 654.
604 Mercier A, Carpentier F, Duplaix C, *et al.* (2019) The polyphagous plant pathogenic fungus *Botrytis cinerea*
605 encompasses host-specialized and generalist populations. *Environmental microbiology*.
606 Nakajima M, Akutsu K (2014) Virulence factors of *Botrytis cinerea*. *Journal of General Plant Pathology* **80**, 15-23.
607 Nielsen R, Williamson S, Kim Y, *et al.* (2005) Genomic scans for selective sweeps using SNP data. *Genome*
608 *research* **15**, 1566-1575.
609 Nosil P, Vines TH, Funk DJ (2005) Perspective: reproductive isolation caused by natural selection against
610 immigrants from divergent habitats. *Evolution; international journal of organic evolution* **59**, 705-719.
611 Otto M, Teixeira PG, Vizcaino MI, David F, Siewers V (2019) Integration of a multi-step heterologous pathway in
612 *Saccharomyces cerevisiae* for the production of abscisic acid. *Microbial Cell Factories* **18**, 205.
613 Pavlidis P, Zivkovic D, Stamatakis A, Alachiotis N (2013) SweeD: likelihood-based detection of selective sweeps
614 in thousands of genomes. *Molecular biology and evolution* **30**, 2224-2234.
615 Perlin MH, Andrews J, Toh SS (2014) Essential letters in the fungal alphabet: ABC and MFS transporters and
616 their roles in survival and pathogenicity. *Advances in genetics* **85**, 201-253.
617 Porquier A, Morgant G, Moraga J, *et al.* (2016) The botrydial biosynthetic gene cluster of *Botrytis cinerea*
618 displays a bipartite genomic structure and is positively regulated by the putative Zn(II)2Cys6
619 transcription factor BcBot6. *Fungal genetics and biology : FG & B* **96**, 33-46.
620 Rodriguez-Moreno L, Ebert MK, Bolton MD, Thomma BPHJ (2018) Tools of the crook- infection strategies of
621 fungal plant pathogens. *The Plant journal : for cell and molecular biology* **93**, 664-674.
622 Rosebeck S, Leaman DW (2008) Mitochondrial localization and pro-apoptotic effects of the interferon-inducible
623 protein ISG12a. *Apoptosis : an international journal on programmed cell death* **13**, 562-572.

624 Scharf DH, Heinekamp T, Brakhage AA (2014) Human and plant fungal pathogens: the role of secondary
625 metabolites. *PLoS pathogens* **10**, e1003859-e1003859.
626 Schouten A, van Baarlen P, van Kan JA (2008) Phytotoxic Nep1-like proteins from the necrotrophic fungus
627 *Botrytis cinerea* associate with membranes and the nucleus of plant cells. *The New phytologist* **177**,
628 493-505.
629 Schulze-Lefert P, Panstruga R (2011) A molecular evolutionary concept connecting nonhost resistance,
630 pathogen host range, and pathogen speciation. *Trends in plant science* **16**, 117-125.
631 Servedio MR, Van Doorn GS, Kopp M, Frame AM, Nosil P (2011) Magic traits in speciation: 'magic' but not rare?
632 *Trends in ecology & evolution* **26**, 389-397.
633 Simko I, Piepho H-P (2012) The area under the disease progress stairs: calculation, advantage, and application.
634 *Phytopathology* **102**, 381-389.
635 Solovyev V, Kosarev P, Seledsov I, Vorobyev D (2006) Automatic annotation of eukaryotic genes, pseudogenes
636 and promoters. *Genome biology* **7**, S10.11-12.
637 Stoletzki N, Eyre-Walker A (2011) Estimation of the neutrality index. *Molecular biology and evolution* **28**, 63-70.
638 Stukenbrock EH, McDonald BA (2008) The origins of plant pathogens in agro-ecosystems. *Annual review of*
639 *phytopathology* **46**, 75-100.
640 Thynne E, Mead OL, Chooi Y, McDonald MC, Solomon PS (2019) Acquisition and loss of secondary metabolites
641 shaped the evolutionary path of three emerging phytopathogens of wheat. *Genome biology and*
642 *evolution* **11**, 890-905.
643 van Kan JA, Stassen JH, Mosbach A, *et al.* (2017) A gapless genome sequence of the fungus *Botrytis cinerea*.
644 *Molecular plant pathology* **18**, 75-89.
645 Walker A (2016) Diversity within and between species of *Botrytis*. In: *Botrytis - The Fungus, the Pathogen and*
646 *its Management in Agricultural Systems* (eds. Fillinger S, Elad Y), pp. 91-125.
647 Walker A, Gladieux P, Decognet V, *et al.* (2015) Population structure and temporal maintenance of the
648 multihost fungal pathogen *Botrytis cinerea*: causes and implications for disease management.
649 *Environmental microbiology* **17**, 1261-1274.
650 Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome*
651 *research* **18**, 821-829.
652 Zhang C, Dong SS, Xu JY, He WM, Yang TL (2019) PopLDdecay: a fast and effective tool for linkage
653 disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786-1788.
654 Zhao Z, Liu H, Wang C, Xu JR (2013) Comparative analysis of fungal genomes reveals different plant cell wall
655 degrading capacity in fungi. *BMC genomics* **14**, 274.
656 Zhong Z, Marcel TC, Hartmann FE, *et al.* (2017) A small secreted protein in *Zymoseptoria tritici* is responsible for
657 avirulence on wheat cultivars carrying the Stb6 resistance gene. *The New phytologist* **214**, 619-631.
658

Tables

Table 1. *Botrytis* isolates included in the study. Isolates from grapevine berries and tomatoes were collected in three regions of France between September 2005 and June 2007 for Champagne and Provence isolates (Walker *et al.* (2015), and between May and June 2009 for Occitanie isolates. B05.10 is used as the reference isolate for genomic analysis (van Kan et al. 2016)

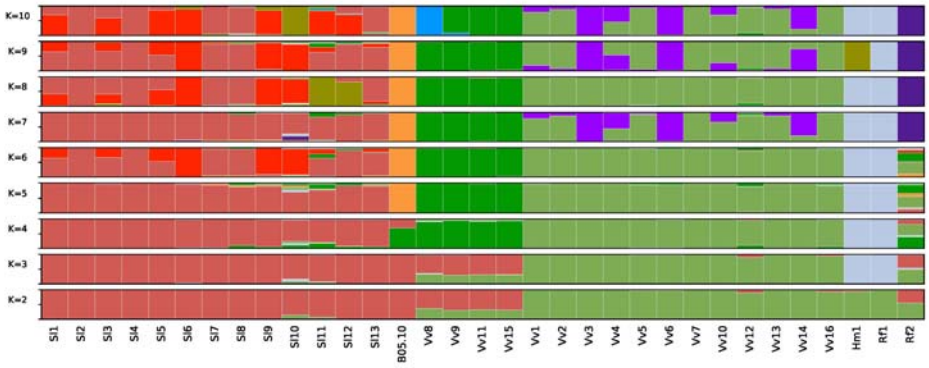
Species	Isolate	Other reference	Host, cultivar	Region, city
B. cinerea	Vv1	VC636	*Vitis vinifera,* Pinot noir	Champagne, Hautvilliers
	Vv2	VC095	*Vitis vinifera,* Pinot noir	Champagne, Vandières
	Vv3	VC621	*Vitis vinifera,* Pinot noir	Champagne, Hautvilliers
	Vv4	VC671	*Vitis vinifera,* Pinot noir	Champagne, Hautvilliers
	Vv5	VC224	*Vitis vinifera,* Pinot meunier	Champagne, Courteron
	Vv6	VC624	*Vitis vinifera,* Pinot noir	Champagne, Hautvilliers
	Vv7	ACBER342	*Vitis vinifera,* Grenache	Provence, Berre
	Vv8	ACBER356	*Vitis vinifera,* Grenache	Provence, Berre
	Vv9	ACBER358	*Vitis vinifera,* Grenache	Provence, Berre
	Vv10	ACSAR333	*Vitis vinifera,* Grenache	Provence, Sarrians
	Vv11	ACSAR334	*Vitis vinifera,* Grenache	Provence, Sarrians
	Vv12	ACSAR335	*Vitis vinifera,* Grenache	Provence, Sarrians
	Vv13	ACSAR342	*Vitis vinifera,* Grenache	Provence, Sarrians
	Vv14	ACSAR354	*Vitis vinifera,* Grenache	Provence, Sarrians
	Vv15	ACSAR357	*Vitis vinifera,* Grenache	Provence, Sarrians
	Vv16	VC610	*Vitis vinifera,* Pinot noir	Champagne, Hautvilliers
	Sl1	VA714	*Solanum lycopersicum,* Moneymaker	Champagne, Foissy-sur-Vanne
	Sl2	VC800	*Solanum lycopersicum,* Moneymaker	Champagne, Courceroy
	Sl3	VC806	*Solanum lycopersicum,* Moneymaker	Champagne, Courceroy
	Sl4	AABER19	*Solanum lycopersicum,* Alison	Provence, Berre
	Sl5	ACBER304	*Solanum lycopersicum,* Alison	Provence, Berre
	Sl6	ACPIE306	*Solanum lycopersicum,* Hipop	Provence, Pierrelatte
	Sl7	ADPIE463	*Solanum lycopersicum,* Hipop	Provence, Pierrelatte
	Sl8	ADPIE475	*Solanum lycopersicum,* Hipop	Provence, Pierrelatte

	Sl9	65_TT8	*Solanum lycopersicum*, Brenda	Occitanie, Alenya
	Sl10	5_TT8	*Solanum lycopersicum*, Brenda	Occitanie, Alenya
	Sl11	13_TT8	*Solanum lycopersicum*, Brenda	Occitanie, Alenya
	Sl12	9_TT8	*Solanum lycopersicum*, Brenda	Occitanie, Alenya
	Sl13	66_TT8	*Solanum lycopersicum*, Brenda	Occitanie, Alenya
	Rf1	VC902	*Rubus fruticosus*, wild	Champagne, Foissy-sur-Vanne
	Rf2	VC399	*Rubus fruticosus*, wild	Champagne, Courteron
	Hm1	MSN-Bot 2556	*Hydrangea macrophylla*, Leuchtfeuer	Anjou, Angers
	B05.10	-	-	-
B. fabae	Bfab	MSN-Bot 2220	*Vicia faba*	Region of Tunis (Tunisia)

665

666

667 ## Figures



668

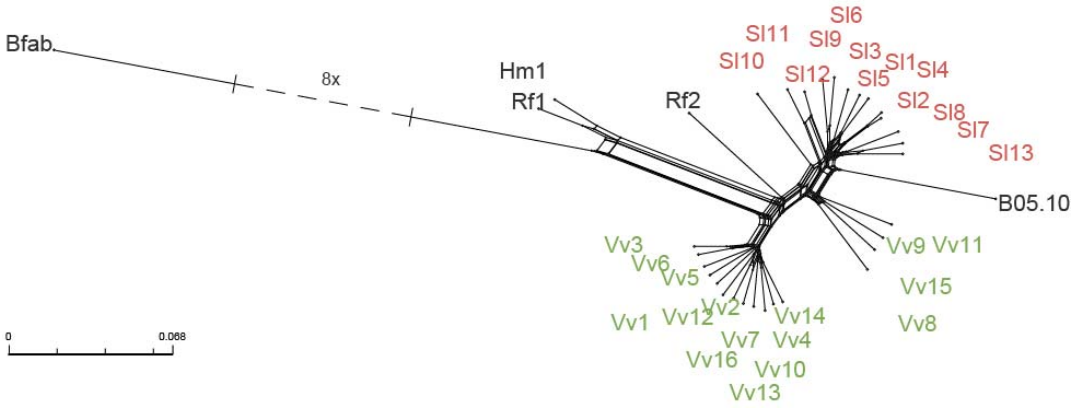
669 **Figure 1.** Ancestry proportions in K clusters based on SNPs identified in 27 isolates of *Botrytis cinerea* collected

670 on tomato (Sl prefix, *Solanum lycopersicum*), grape (Vv prefix, *Vitis vinifera*), bramble (Rf prefix, *Rubus*

671 *fruticosus*) and hydrangea (Hm prefix, *Hydrangea macrophylla*), as estimated with the sNMF program. Each

672 multilocus genotype is represented by a vertical bar divided into K segments, indicating membership in K

673 clusters. B05.10 is the reference genome for *B. cinerea* (van Kan et al. 2016).



674

675 **Figure 2**. Neighbor-net phylogenetic network estimated with SPLITSTREE based on SNPs identified in 27 isolates

676 of *Botrytis cinerea* and one isolate of *B. fabae* (Bfab) used as the outgroup. Reticulations indicate phylogenetic

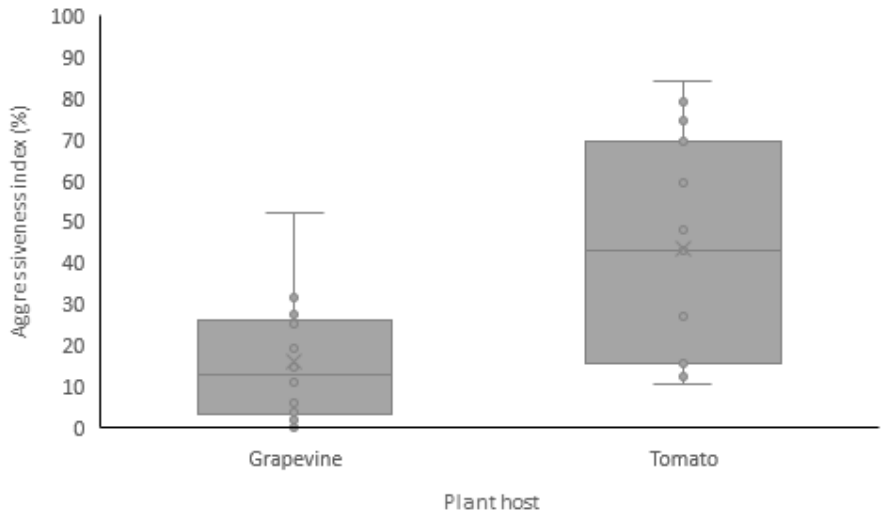
677 conflicts caused by recombination or incomplete lineage sorting. Isolates were sampled from *Vitis vinifera* (Vv

678 prefix, green) *Solanum lycopersicum* (Sl prefix, red), bramble (Rf prefix, *Rubus fruticosus,* black) and hydrangea

679 (Hm prefix, *Hydrangea macrophylla,* black). B05.10 is the reference genome for *B. cinerea* (van Kan et al. 2016).

680

681



682

683 **Figure 3.** Boxplots representing the extent of aggressiveness (% relative to the reference isolate BC1) on

684 tomato of the *Botrytis cinerea* isolates collected on grapevine (n=16) and tomato (n=13). For each boxplot,

685 mean (crosses) (median (horizontal lines), values of the aggressiveness index (circles), 25-75% quartiles, and

686 maximum and minimum values are represented.

687

Figure 4. Log_{10} of the neutrality index estimated based on polymorphism and divergence at 11,917 genes in, respectively, thirteen and twelve isolates of the SI and Vv1 populations of *Botrytis cinerea*. A positive Log_{10} neutrality index indicates negative selection, and a negative Log_{10} neutrality index indicates positive selection. Genes potentially involved in host specialization of the SI population onto tomato (red dots) and to the Vv1 population onto grape (green dots) were identified as genes having both low neutrality index values in the focal population (log[neutrality index]<=-0.5), and high values in the alternate population (log[neutrality index]>=0).

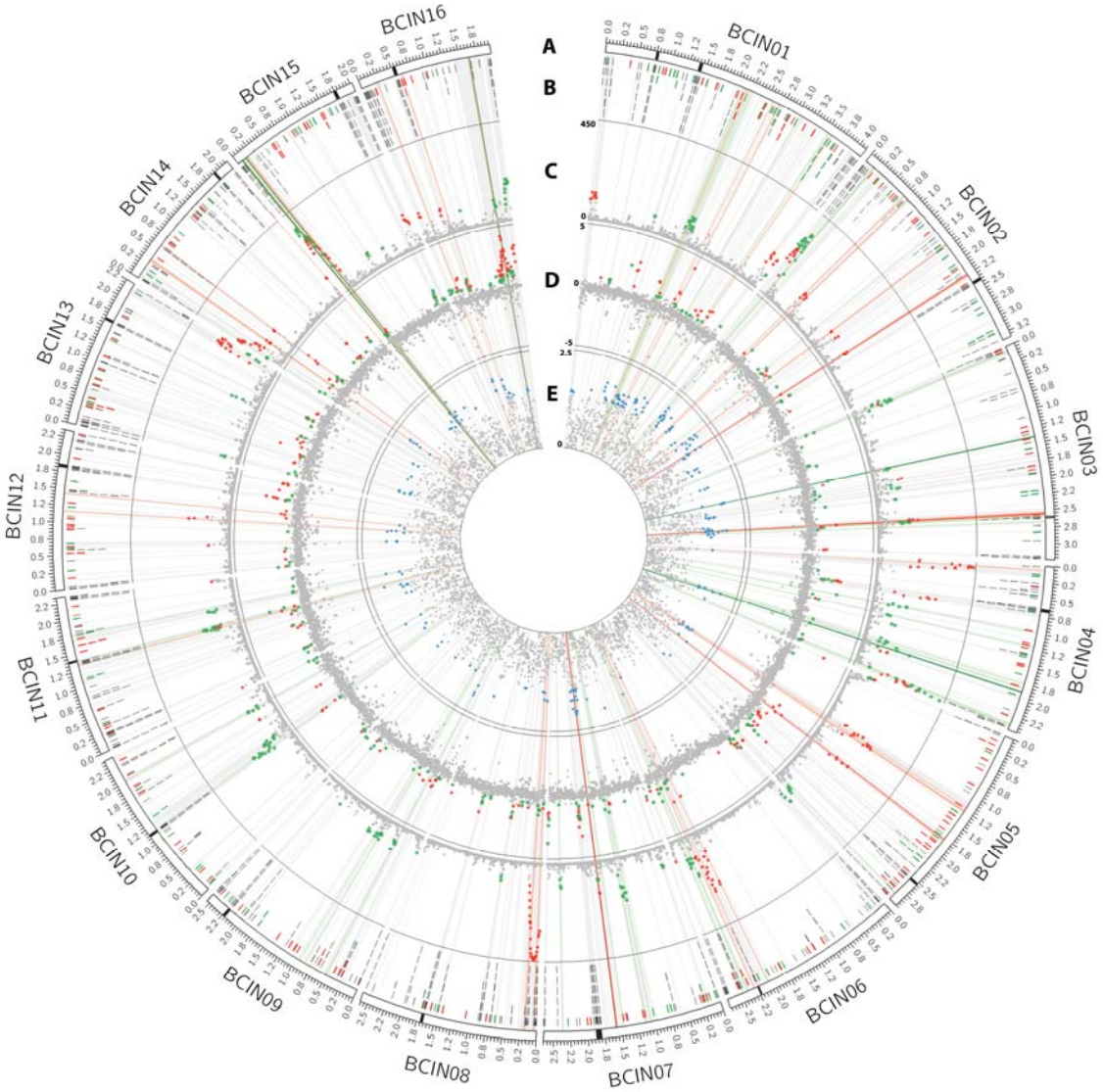


Figure 5. Genome scans for positive selection in 10kb genomic windows of the Vv1 and SI populations of *Botrytis cinereal*, parasitizing grapevine and tomato, respectively. Values in the top 5% of positive selection metrics are represented as green dots for the Vv1 population, as red dots for the SI population, and in blue for the whole dataset. (A) Core chromosomes represented using white bars with putative centromeric regions in black. Dispensable chromosomes (BCIN17 and BCIN18) are not represented. (B) Position of transposable elements (grey) and candidate genes for specialization identified using the McDonald-Kreitman tests for Vv1 and SI populations (see Figure 4). (C) Composite likelihood ratio estimated using the SweeD software (Nielsen *et al.* 2005; Iidis *et al.*, 2013). (D) Number of segregating sites by length, nS_L, estimated using the nSL software

(Ferrer-Admetlla *et al.* 2014). (E) hapFLK statistic estimated using the HAPFLK software (Bonhomme *et al.* 2010; Fariello *et al.* 2013). Radial lines show regions in the top 5% quantile for all three methods (darker red/green), for two out of three methods (lighter red/green), or by one method only (light grey).