Supplementary Note

Life-style transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses

Richard J O'Connell, Michael R Thon, Stéphane Hacquard, Stefan G Amyotte, Jochen Kleemann, Maria F Torres, Ulrike Damm, Ester A Buiate, Lynn Epstein, Noam Alkan, Janine Altmüller, Lucia Alvarado-Balderrama, Christopher A Bauser, Christian Becker, Bruce W Birren, Zehua Chen, Jaeyoung Choi, Jo Anne Crouch, Jonathan P Duvick, Mark L Farman, Pamela Gan, David Heiman, Bernard Henrissat, Richard J Howard, Mehdi Kabbage, Christian Koch, Barbara Kracher, Yasuyuki Kubo, Audrey D Law, Marc-Henri Lebrun, Yong-Hwan Lee, Itay Miyara, Neil Moore, Ulla Neumann, Karl Nordström, Daniel G Panaccione, Ralph Panstruga, Michael Place, Robert H Proctor, Dov Prusky, Gabriel Rech, Richard Reinhardt, Jeffrey A Rollins, Steve Rounsley, Christopher L Schardl, David C Schwartz, Narmada Shenoy, Ken Shirasu, Usha R Sikhakolli, Kurt Stüber, Serenella A Sukno, James A Sweigard, Yoshitaka Takano, Hiroyuki Takahara, Frances Trail, H Charlotte van der Does, Lars M Voll, Isa Will, Sarah Young, Qiandong Zeng, Jingze Zhang, Shiguo Zhou, Martin B Dickman, Paul Schulze-Lefert, Emiel Ver Loren van Themaat, Li-Jun Ma, Lisa J Vaillancourt

Nature Genetics: doi:10.1038/ng.2372

Supplementary Note

1. Background information

Colletotrichum is the anamorphic stage of the genus *Glomerella* and belongs to the Ascomycota (Pezizomycotina, Sordariomycetes, Hypocreomycetidae, Glomerellales, Glomerellaceae)[1,2]. *Colletotrichum* is one of the most widespread and economically damaging genera of plant pathogenic fungi and represents a major threat to global food security. Members of the genus cause anthracnose leaf spot diseases, blights and post-harvest rots on an enormous range of agronomic and horticultural crops throughout the world[3,4]. Economic and social impacts are especially severe in tropical and sub-tropical regions, where *Colletotrichum* limits the production of maize, sorghum, grain legumes, cassava, yam and many fruit crops.

C. higginsianum causes disease on many cultivated forms of *Brassica* and *Raphanus*, especially turnip, radish, pak-choi and Chinese cabbage. It also infects wild species of Brassicaceae, including the model plant *Arabidopsis thaliana*. This provides an attractive pathosystem for studying the molecular basis of fungal pathogenicity and plant resistance/susceptibility, in which both partners can be genetically manipulated[5,6]. *C. higginsianum* strains mutated in the *Ku70* DNA repair gene provide enhanced rates of homologous recombination, which facilitates targeted mutagenesis[7]. *C. graminicola* causes anthracnose stalk rot and leaf blight of maize and, in contrast to *C. higginsianum*, is believed to have a narrow host range confined to *Zea* and close relatives[4,8,9]. Maize is one of the most important commodities worldwide and is the dominant crop in the United States, with a value of more than 76 billion dollars in 2011 (USDA/NASS). *C. graminicola* is estimated to cause annual losses of more than 1 billion dollars in the U.S.A. alone[10]. *C. graminicola* is also among the most intensively studied and genetically tractable species of *Colletotrichum*, and one of the few in which sexual crosses can be performed[8,11].

1.1. Infection process

Colletotrichum pathology characteristically involves the sequential formation of a series of specialized cell types associated with a multi-stage 'hemibiotrophic' infection process[12]. Asexual spores (conidia) germinate to form dome-shaped appressoria, which are darkly pigmented with melanin and mediate initial host penetration (**Supplementary Figs 1a,b** and **2a,b**). A penetration hypha emerging from the base of the appressorium breaks through the plant cuticle and cell wall (**Supplementary Fig. 1b,c,d**) by means of turgor-generated mechanical force and lytic enzymes[13,14]. Inside host epidermal cells, the fungus then differentiates bulbous primary hyphae (**Supplementary Figs 1c,e** and **2c-f**), which expand and invaginate the host plasma membrane (**Supplementary Figs 1b,c** and **2f**). This symptomless stage of infection is termed 'biotrophic' because the penetrated host cells remain alive[6]. In the case of *C. higginsianum*, both the primary hyphae and the biotrophic phase are entirely confined to the first infected epidermal cell. Approximately 24 hours after initial penetration, infected cells die and *C. higginsianum* switches to destructive 'necrotrophic' growth, associated with a change in hyphal morphology: thin, filamentous secondary hyphae emerging from the tips of primary hyphae invade adjacent cells and rapidly colonize the host tissue, killing cells and degrading cell walls ahead of infection (**Supplementary Fig. 1f,g**). In

contrast to *C. higginsianum*, the primary hyphae of *C. graminicola* colonize many epidermal, as well as mesophyll cells, becoming highly constricted at cell-to-cell penetration sites as they pass through host cell walls (**Supplementary Fig. 2c,d**). Host cells die 12-24 hours after penetration by primary hyphae, while at the same time more recently invaded cells appear alive. Eventually, thin secondary hyphae develop within the initially invaded cells, associated with the appearance of disease symptoms (tissue maceration and browning). Invasion of living host cells by primary hyphae continues at the expanding margins of the lesion, while host cells at the center are dead and macerated[15]. Thus, biotrophy and necrotrophy occur simultaneously at the later stages of *C. graminicola* infection, whereas the switch from biotrophy to necrotrophy is complete in the case of *C. higginsianum*. In both species, the asexual cycle concludes when sporulating structures (acervuli) erupt from the surface of the dead plant tissue (**Supplementary Figs 1h** and **2a**).

1.2. Mating type genes

Sexual reproduction is rare in most *Colletotrichum* species. The *C. higginsianum* strain used in this study does not appear to be fertile, either on its own or when paired with other potential mating partners (O'Connell, unpublished data). *G. graminicola* was reported to be homothallic[16], but the majority of strains, including the one in this study, are self-sterile and cross-fertile[11]. Mating in *Colletotrichum* is inconsistent with the typical one locus, two allele (idiomorph) mating system utilized by most filamentous ascomycetes to regulate mating compatability[17]. *C. graminicola* and *C. higginsianum* each have a single *Mat1-2* gene encoding a characteristic HMG box DNA binding domain (GLRG_04643, CH063_09403, 67.9% sequence identity). However, there is no strong evidence for a *Mat1-1* idiomorph in either species, consistent with many previous studies that failed to detect the conserved *Mat1-1* alpha-box motif by hybridization in *C. graminicola* and other *Colletotrichum* species[17-21]. Analysis of synteny in the region of the *Mat1-2* locus in *C. graminicola* and other Sordariomycetes revealed a conserved group of 14 genes: *Cia30*/*Apc5*/*Cox13*/*Apn2*/*Mat1*/*Sla2*/L21e/S4-9/*Slu7*/*Rev3*/*Tex2*/*Ami1*/Cwc24/*Atg3*. A similar analysis for *C. higginsianum* could not be completed due to the fragmented nature of the genome assembly, with no more than 3 genes on a single contig and several genes incomplete. Nonetheless, the region appears to be syntentic between the two species.

1.3. Phylogeny

To position the two sequenced genomes reported in this study within the context of the genus *Colletotrichum*, we generated a phylogeny based on sequencing five genes. Genomic DNA of the selected isolates was extracted using the method of Damm *et al.*[22]. The 5.8S nuclear ribosomal gene with the two flanking internal transcribed spacers (ITS), and partial sequences of the actin (ACT), chitin synthase 1 (CHS-1), beta-tubulin (TUB2) and histone3 (HIS3) genes were amplified and sequenced using the primer pairs ITS-1F[23] + ITS-4[24], ACT-512F + ACT-783R[25], CHS-354R + CHS-79F[25], BT2Fd + BT4R[26] and CYLH3F + CYLH3R[27], respectively. The PCRs were performed as described by Damm *et al.*[28] and Woudenberg *et al.*[26]. The DNA sequences obtained from forward and reverse primers were used to obtain consensus sequences using Bionumerics v. 4.60 (Applied Maths, St-Marthens-Lathem, Belgium), which were added to the outgroup (*Monilochaetes infuscans*, strain CBS 869.96).

The alignment was assembled and manually adjusted using Sequence Alignment Editor v. 2.0a11[29]. To determine whether the seven sequence datasets were congruent and combinable, tree topologies of 70 % reciprocal Neighbour-Joining bootstrap with Maximum Likelihood distances (10,000 replicates) were compared visually[30]. Substitution models were determined separately for each partition using MrModeltest v2.3[31]. A maximum parsimony analysis was performed on the multilocus alignment (ITS, CHS-1, HIS3, ACT, TUB2) with PAUP (Phylogenetic Analysis Using Parsimony) v.4.0b10[32] using the heuristic search option with 100 random sequence additions and tree bisection and reconstruction (TBR) as the branch-swapping algorithm. Alignment gaps were treated as missing and all characters were unordered and of equal weight. The robustness of the trees was evaluated by 500 bootstrap replications with 2 random sequence additions[33]. Tree length, consistency index (CI), retention index (RI), rescaled consistency index (RC) and homoplasy index (HI) were calculated for the resulting tree. Sequences derived in this study were lodged at GenBank and the alignment was submitted to TreeBASE (www.treebase.org/treebase-web/home.html).

The five sequence datasets did not show any conflicts in tree topology for the 70 % reciprocal bootstrap trees, which allowed us to combine them. In the multigene analyses of 28 isolates including the outgroup, 2150 characters including the alignment gaps were processed, of which 594 characters were parsimony-informative, 229 parsimony-uninformative and 1327 constant. After a heuristic search using PAUP, the most parsimonious tree was retained (length=2331 steps, CI=0.542, RI=0.641, RC=0.347, HI=0.458) and is shown in **Supplementary Fig. 5a**.

These analyses resulted in detection of 10 major clades representing a selection of different *Colletotrichum* species and species complexes. We selected *Colletotrichum* strains that have been (or are in the process of being) genome-sequenced, or which have been extensively used to study host-pathogen interactions, as well as ex-type strains if available (**Supplementary Fig. 5b**). The two *C. higginsianum* strains, IMI 349063 and MAFF 305635, do not show any sequence variation and group in the destructivum clade together with the closely related *C. fuscum* and four other taxa, namely *C. tabacum*, *C. destructivum*, *C. linicola* and *Glomerella truncata*, that form only short branches. Latunde-Dada & Lucas[34] showed that "*C. truncatum*" from lentil is closely related to *C. destructivum*. Armstrong-Cho & Banniza[35] named the teleomorph formed by mating two of these strains (one of which is included in this phylogeny) *G. truncata*, assuming this to be the sexual stage of *C. truncatum.* However, this phylogeny shows that *G. truncata* belongs to the destructivum clade and is not closely related to the ex-epitype strain of *C. truncatum*[28].

The graminicola clade contains seven taxa on comparatively long branches that are all pathogens restricted to the plant family Poaceae, with some species probably specialised to single host species or genera, and all characterized by falcate conidia and dark setae[9]. The graminicola clade is a sister clade to the destructivum clade and a single-strain clade representing *C. coccodes*. The two *C. graminicola* strains do not show any sequence variation and both group together with *C. navitas* next to *C. nicholsonii*. These three taxa form a sister clade to that formed by *C. sublineolum*, *C. falcatum* and *C. caudatum*, while *C. cereale* has a basal position within the

graminicola clade. The orbiculare clade, represented by *C. orbiculare* and *C. lindemuthianum*, is basal to all other *Colletotrichum* species and species complexes included in this phylogeny.

2. Genome sequencing, assembly and annotation

2.1. Selection of isolates

C. graminicola strain M1.001 (also known as M2) was collected in Missouri from infected maize[36]. This strain was selected for sequencing because it is the most commonly used laboratory strain and is very aggressive on both maize leaves and stalks. It also undergoes sexual crosses as male or female with most other field strains[11]. The strain is available from the Fungal Genetics Stock Center (culture #10212).

The *C. higginsianum* strain IMI 349063 was originally obtained from leaves of *Brassica campestris* in Trinidad and Tobago (CABI culture collection, Wallingford, UK). This isolate was chosen because many genomic resources are available for this genotype, including large EST collections and random insertional mutants[37,38]. This isolate appears to be asexual[6].

2.2. Sequencing and assembly

Genome assemblies of *C. graminicola* strain M1.001 were generated at the Broad Institute employing a Whole Genome Shotgun (WGS) sequencing approach. The hybrid assembly (NCBI accession number ACOD01000000), combining sequence reads obtained using Sanger and 454 pyrosequencing technologies, was produced using an improved Newbler hybrid approach (**Supplementary Table 2**). In particular, paired-end reads from 468,734 plasmids and 67,151 fosmids were used to improve continuity of the assembly, resulting in a high quality draft assembly that consists of 653 sequence scaffolds with a total length of 50.87 Mb and an N_{50} scaffold length of 579 Kb (that is, 50% of all bases are contained in scaffolds of at least 579 Kb). The 454 sequencing reads provided deeper coverage (>9X) and enhanced the confidence level at each base. In the assembled genome, more than 98.5% of the sequence bases have quality scores >40, i.e. one error per 10^4 bases.

The genome assembly of *C. higginsianum* was generated by GATC Biotech AG (Konstanz, Germany) using a hybrid sequencing approach, combining 454 GS-FLX shotgun reads and Illumina GAII mate-pair reads. In addition, a fosmid library was end-sequenced using Sanger technology (1,728 reads from 864 clones). After removing dinucleotide repeats, the 454 shotgun reads were co-assembled together with fosmid end sequences using the SeqMan NGen assembly tool (DNAStar Inc., Madison, WI 53705, USA). The contigs were then sorted into scaffolds using paired-end information derived from an Illumina 3 kb insert mate-pair library (2x36 bp reads), again using the NGen assembler. The scaffolds were manually edited to correct falsely joined contigs and falsely arranged scaffolds. After editing, the assembly comprised 10,269 contigs (total contig length = 48.2 Mb, contig N50 length = 14 kb). Of these, 9,648 contigs (97%) were grouped into 366 scaffolds (scaffold N50 length = 264 kb). Only 621 contigs (~1 Mb) were not included in scaffolds. To correct homopolymer sequencing errors in the 454 data, Illumina GA data (76-fold coverage) were mapped to the scaffolded contigs and the depth of coverage

was used to create a final corrected consensus sequence (**Supplementary Table 2**). This assembly has NCBI accession number CACQ02000000.

The *C. graminicola* and *C. higginsianum* genome assemblies are available from the Broad Institute Colletotrichum Genome Database (http://www.broadinstitute.org/annotation/genome/colletotrichum_group).

After genome annotation had been performed (see below), the *C. higginsianum* sequence data were re-assembled using a combination of Newbler and Velvet assemblers[39]. After initial assembly of the 454 data using Newbler (version 2.3), Velvet was used to coassemble the Illumina reads together with contigs obtained with Newbler. This hybrid assembly has fewer contigs (6,566) than the annotated NGen assembly but aligns less well to the optical map. This assembly is available from the Max Planck Institute for Plant Breeding Research website (http://www.mpipz.mpg.de/14157/fungal_genomes).

2.3. Gene annotation

To provide a training set for the gene-calling pipeline and for validating gene models, we generated ESTs as follows. For *C. graminicola*, Sanger ESTs were obtained from a normalized cDNA library produced from saprophytic mycelium grown on rich medium (18,812 reads) and a non-normalised library from nitrogen-starved mycelium (9,612 reads). For *C. higginsianum*, Sanger ESTs were obtained from appressoria *in vitro* (1,447 reads[38]) and biotrophic hyphae purified from infected tissue using flow cytometry (4,027 reads[40]). In addition, 454-pyrosequencing was used[41] to obtain ESTs from appressoria *in planta* (38,024 reads), purified biotrophic hyphae (318,604 reads), necrotrophic mycelium (143,637 reads) and appressoria *in vitro* (322,853 reads)[42].

Protein coding genes were annotated in *C. graminicola* using multiple lines of evidence from BLAST, PFAM searches and EST alignments. First, we aligned ESTs to the genome using BLAT and then collapsed them into distinct transcript clusters using a program called CallReferenceGenes developed at the Broad Institute (described below). EST alignments with 90% identity over 50% of the EST length and with canonical splice junctions were considered valid alignments suitable for building gene models. In addition, EST-based gene models were built manually. Gene structures were predicted using a combination of gene models from the computational gene prediction programs FGENESH (Softberry Inc.), GENEID[43], GeneMark[44] and EST-based automated and manual gene models. Both GENEID and FGENESH were trained using a set of high confidence EST-based gene models generated by clustering blat-aligned species-specific ESTs. In addition, the gene-finding programs SNAP[45] and Augustus[46] were used. After training the gene prediction programs, their performance was evaluated by comparing the gene models with EST and BLAST evidence. Those that performed adequately were used in the automated gene calling pipeline developed at the Broad Institute[47]. By combining BLAST, EST and *ab initio* predictions, manual annotators built additional gene models that were otherwise missed by the automated annotation. *C. graminicola* was predicted to have 12,006 gene models, of which 39% were verified by the alignment of 13,600 Sanger EST reads. The *C. higginsianum* gene set was created similarly, filtered using TBLASTN alignments from 10,661 of the *C. graminicola* gene models (<1e-10); a further 1,564 gene models were

based on evidence from *C. higginsianum* ESTs and 600 were based on EVidenceModeler (EVM) models having BLAST hits to proteins in the UniRef90 database. In total, *C. higginsianum* was predicted to have 16,172 protein-coding genes, of which 89% were validated by alignment of 135,923 ESTs from 454 sequencing.

2.4. Evaluation of gene space coverage and annotation

We used the Core Eukaryotic Genes Mapping Approach pipeline (CEGMA v. 2.0)[48] to identify orthologs of 248 single-copy, core eukaryotic genes (CEGs) in both *Colletotrichum* genomes. In the *C. higginsianum* genome assembly, CEGMA annotated 228 (91.9%) of the CEGs as complete and eight (3.2%) as partial genes, while in the *C. graminicola* genome assembly 245 (98.8%) CEGs were complete and only one (0.4%) was a partial gene. It is important to note that *C. graminicola* was sequenced largely using Sanger technology, which provides less fragmented assemblies than the short-read data obtained from next-generation sequencing (NGS) with 454 or Illumina platforms. Consequently, the CEGMA annotation pipeline typically detects more complete CEGs in Sanger-sequenced genomes. The percentages of complete and partial CEGs found in *C. higginsianum* are comparable to results obtained from other NGS-sequenced genomes[49]. The CEGMA core genes identified as missing in *C. higginsianum* were manually inspected by aligning the protein sequence to the genome with TBLASTN. Six of the 12 genes identified as missing by CEGMA could be aligned, showing that the true coverage of the gene space in *C. higginsianum* is approximately 97.6%. Two of the missing genes (KOG1466 and KOG3232) could not be aligned to either *Colletotrichum* genome, but also could not be aligned to the genomes of *Magnaporthe oryzae* or *Neurospora crassa*. If these two genes are excluded from the analysis, then the estimated gene space coverage for *C. graminicola* and *C. higginsianum* are 100% and 98.4%, respectively.

To obtain an independent estimate of the number of transcripts missing from the *C. higginsianum* genome assembly, we used data obtained from RNA-Sequencing (see **Section 6**). Sequence reads from the necrotrophic phase were excluded because these included significant contamination with bacterial sequences. The reads were filtered and trimmed using fastq_quality_trimmer (-t 25, -l 50), a component of FASTX toolkit (version 0.0.13). Next, reads that aligned to the *Arabidopsis thaliana* genome using tophat (a=10, g=5, version 1.2.0)[50] were removed. The reads not mapping to the *A. thaliana* genome were assembled using Trinity (version 21-05-2012,)[51] and the previously filtered reads were mapped back to the assembly to obtain coverage information. Assembled transcripts less than 300 bases in length or with less than 10X coverage were removed, as were transcripts that still mapped to the *A. thaliana* genome using megablast (alignment length > 100, identity > 98). The remaining 6,710 transcripts were aligned to the *C. higginsianum* genome using megablast. Of these, 6,491 transcripts (96.7%) could be aligned (alignment length > 100, identity > 98), at least partially, to the *C. higginsianum* genome assembly and 6,598 transcripts (98.3%) mapped to the more complete Velvet assembly. This indicates that few transcripts are completely missing from the assembly, consistent with the results obtained using CEGMA. As a further measure of the completeness of the genome assembly, we searched for 25 genes that were previously identified from *C. higginsianum*[7,37,38] or available from the NCBI nucleotide database. Using BLASTN, we found that all of these genes are present in the genome assembly.

Due to the fragmented nature of the *C. higginsianum* genome assembly, resulting from the sequencing and assembly strategy use for this species, we suspected that some of the predicted gene models were truncated or split with respect to the true gene structure. To determine the extent of these artifacts, BLASTN was used to align the transcript sequences of one hundred CEGMA genes from the *C. graminicola* genome to the *C. higginsianum* transcripts predicted by the annotation pipeline. The BLAST results were then examined manually to identify truncated gene models (truncated at 5' or 3' ends with respect to the *C. graminicola* model), and those that are split (more than one *C. higginsianum* transcript aligning with high sequence identity over distinct regions of the *C. graminicola* model). Nine of the 100 (9%) models examined showed evidence of truncation only, while a further seven (7%) showed evidence of being split into two truncated models. We also manually examined 225 additional gene models during comparative analyses described in this manuscript and found four gene models were truncated only, while a further 10 were both split and truncated. Combining the results from both analyses, we estimate that 5.2% of the *C. higginsianum* gene models represent genes that were split into two or more models, while 4.0% are truncated representations of the true gene structure.

2.5. Optical mapping and genome organization

In the absence of genetic mapping information, optical maps were constructed for both *Colletotrichum* genomes. Protoplasts of *C. graminicola* and *C. higginsianum* were generated according to Thon *et al.*[52], washed three times with PBS buffer to remove storage buffer and glycerol, then lysed in TE pH 8.0 with 5 mM EGTA and 1 mg/ml proteinase K by heating at 70° C for 15 min. The extract was then incubated at 37° C overnight to ensure full digestion of proteins from the lysed protoplasts and the autodigestion of excess proteinase K. T7 DNA (final concentration ~30 pg/μl; Yorkshire Bioscience Ltd, York, UK) was added as a sizing standard. DNA solutions were loaded into a silastic microchannel device[53,54] for presenting DNA molecules on optical mapping surfaces. Mounted DNA molecules were digested by restriction endonuclease *Mlu*I (A^CGCGT, average fragment size for both genomes = 9.2 kb) in NEB buffer 3 (100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl2, 1 mM dithiothreitol, pH 7.9; New England Biolabs) with omission of BSA, and addition of 0.02% Triton X-100. Digested DNA molecules were then stained with 12 μL of 0.2 μM YOYO-1 solution (Molecular Probes, Eugene, OR) in TE containing 20% β-mercaptoethanol. Fully automated imaging workstations[53,55] were used for generating single molecule optical data sets for the genomes of *C. graminicola* and *C. higginsianum* and high resolution whole genome physical maps for each genome were constructed using a shotgun optical mapping approach[55-60]. We used the optical mapping system to acquire more than 300X coverage of raw single DNA molecule maps (Rmaps) for each genome that were assembled into genome-wide optical maps, spanning each chromosome. These contigs were assembled by the map assembler[61] using divide-and-conquer and iterative assembly strategies[55,57,60,62] for distributing the computational load[54,55,60]. For the *C. graminicola* genome, we assembled 13 chromosome-wise optical map contigs ranging from 505 kb to 7,429 kb with a total size of 57.44 Mb (**Supplementary Table 1a**). Thus, approximately 6.5 Mb is missing from the *C. graminicola* sequence assembly (total length 50.9 Mb). For the *C. higginsianum* genome, 12 chromosome-wide optical map contigs were assembled, and these contigs range

from 610 kb to 6,287 kb with a total size of 53.35 Mb (**Supplementary Table 1b**). These findings are consistent with results obtained for these two species by cytological karyotyping (Masatoki Taga, personal communication).

For *C. higginsianum*, we also calculated genome size by dividing the total number of bases generated by Illumina sequencing by the genome coverage, which was estimated as a weighted mean from the average coverage in all contigs >1 kb, weighted according to contig length. The coverage per contig was calculated by mapping 20.2 million Illumina reads (878.4Mb) to the genome using the BWA program[63]. In total, 18.9 million reads mapped to the genome, resulting in a weighted mean coverage of 16.7 (± 1.9). Contigs with >50x coverage were discarded from the coverage estimate because they likely represent mitochondrial DNA or repetitive regions. Based on these results, we estimate the *C. higginsianum* genome size to be 52.6 Mb, comparable to the estimate obtained from optical mapping (53.35Mb). This indicates that 3.3 to 4.1 Mb is missing from the current sequence assembly (total length of 49.3 Mb).

Map alignment software[64] was used to investigate whether homologous chromosomes or chromosomal segments could be detected between the genomes at the optical map level. The software allows for both local and global map pairwise alignments. However, global map alignment approaches will fail to detect local significant instances of variation, such as insertions, deletions or partial genome duplication, which are expected when comparing optical maps derived from different strains or different chromosome homologs. Our approach for extracting multiple high-scoring alignments was based on the efficient linear scaling approach of Huang and Miller[65]. We generated confidence scores using an approach similar to that used by Waterman and Vingron[66] for sequence alignments. However, this analysis yielded only two map segments that could be aligned between the two *Colletotrichum* genomes: (a) the ribosomal repeat regions, which span a region of >1Mb on *C. graminicola* chromosome (Chr) 8 and *C. higginsianum* Chr 9, and (b) a segment of ~47 kb located on *C. graminicola* Chr 9 (475,161-522,931) and *C. higginsianum* Chr 8 (1,526,687-1,573,653), which could be inversely aligned. Overall, these results indicate that the two *Colletotrichum* species are relatively divergent at sequence level so that very few homologies can be identified at optical map level.

Using map aligner software developed at the Broad Institute, we also aligned the *in silico* maps of genome sequence contigs or scaffolds, based on lengths of the predicted *Mlu*I fragments, to the finished optical maps of *C. graminicola* and *C. higginsianum*. This facilitated ordering and orientating contigs or scaffolds, estimating gaps between contigs or scaffolds, and revealed assembly errors. The results were incorporated into the final genome sequence assembly. Overall, 70 % of the *C. graminicola* optical maps were covered by assembled sequence scaffolds (**Supplementary Table 1c**). The optical linkage group maps can be accessed at: http://www.broadinstitute.org/annotation/genome/colletotrichum_group/maps/ViewMap.html?sp=0.

The PROmer program of the Mummer package[67], revealed that the chromosomal structure of the two genome assemblies was similar (**Fig. 2a** and **Supplementary Table 4**). The homologous Chr 8 in *C. graminicola* and Chr 9 in *C. higginsianum* contain the rDNA repeats. Overall about 35% of the genome space in *C. graminicola* has homologous matches in *C. higginsianum*. Although most chromosomes have about 40% coverage, the three

minichromosomes in the *C. graminicola* assembly[68] lack significant homologous hits in the *C. higginsianum* assembly (**Fig. 2b** and **Supplementary Table 4**). Similarly, one of the smallest chromosomes (Chr 12) in *C. higginsianum* was too short to be mapped to the chromosome map, and the other minichromosome (Chr 11) also lacked homologous hits in the *C. graminicola* assembly. Extensive chromosomal rearrangements were observed between these two genomes (**Fig. 2a**). However, with the exception of Chr 9 in the *C. graminicola* assembly, which has homologous sequences dispersed among several different *C. higginsianum* chromosomes, as well as many unmapped *C. higginsianum* scaffolds, homologous relationships were detectable for most chromosomes (**Supplementary Table 5**). From the global comparison, there are more homologous sequences in the *C. higginsianum* assembly that cannot be mapped to chromosomes, reflecting the fragmented nature of the genome assembly. For the *C. graminicola* assembly, most unmapped sequences are caused by repetitive sequences.

3. Transposable element analysis

Repetitive DNA elements were identified as described in the **Online Methods**. The *C. graminicola* genome was found to contain 13 species of LTR retrotransposons, 6 species of DNA transposons and at least 3 additional repetitive elements of unknown classification (**Supplementary Table 3**). In *C. graminicola*, repetitive DNA comprises 5.5 % of sequence scaffolds located on the 10 larger chromosomes, 22.7 % of sequence scaffolds located on the three minichromosomes and 49.8 % of the unanchored scaffolds, totalling 12.2 % of the genome overall. The percentage and diversity of the repetitive DNA is similar to that found in the genomes of other Sordariomycetes such as *Magnaporthe oryzae*[69] and *Fusarium oxysporum*[70]. A similar analysis of the *C. higginsianum* genome identified less 1.22 % repetitive DNA. However, the sequencing and assembly strategy used for this species excluded most repetitive sequences from the final assembled contigs. For this reason, we did not attempt to further characterize the transposable elements (TEs) of *C. higginsianum*.

The distribution of TEs on *C. graminicola* Chr 1 is shown in **Supplementary Figure 3**. The TEs are organized into clusters and the clusters can be found throughout the genome. In other fungi, TEs are associated with increased rates of recombination and gene evolution[71,72]. Therefore we determined whether there is a correlation between the presence of TE clusters and specific features in the *C. graminicola* genome. We examined the 10 largest chromosomes in *C. graminicola*; the three remaining chromosomes were too small to test using this method. For the purpose of this analysis, we define paralogous genes as those that are found in MCL clusters containing more than one protein from the same species. Using Pearson's 'r' correlation-coefficient test, we found a statistically significant correlation ($p < 0.001$) between the presence of TEs and paralogous gene copies on 7 of the 10 largest chromosomes. We identified orthologous genes between the two *Colletotrichum* species by computing reciprocal best blast hits, prepared a global alignment of each pair of orthologs, and determined the pair-wise sequence identity. We found a significant negative correlation between the presence of TEs and transcript sequence identity to their *C. higginsianum* ortholog on 6 out of 10 chromosomes. These two observations suggest that the *C. graminicola* genome contains regions enriched with genes undergoing rapid evolution. When we examined the distribution of genes encoding secreted proteins, 5 out of 10 chromosomes

had a significant positive correlation with the presence of TEs. Our current understanding of plant-pathogen interactions indicates an important role for the secretome in pathogenicity, and since secreted proteins potentially interact with host components, the 'arms race' hypothesis predicts that genes encoding such proteins would be under strong selective pressure[73]. Thus, the correlation between the location of TEs and genes encoding secreted proteins is consistent with the secretome's role in adaptation of the fungus to its host. We also performed this test with genes encoding CSEPs and secondary metabolism enzymes but failed to find a correlation between TEs and either of these gene categories. It is likely that the number of genes in these categories was not high enough to yield a statistically significant result with the test used.

4. Orthology and multigene families

To identify differences in gene family size between *C. graminicola* and *C. higginsianum*, we clustered their proteomes using the clustering application MCL[74]. An all vs. all BLASTP search was performed using the program's default parameters, followed by clustering with MCL using an inflation value of 2.0. We also included the proteomes of 11 additional species of filamentous fungi that represented various lifestyles, including four necrotrophs (*Alternaria brassisicola*, *Botrytis cinerea*, *Stagonospora nodorum*, *Sclerotinia sclerotiorum*), two obligate biotrophs (*Melampsora laricis-populina*, *Blumeria graminis* f.sp. *hordei*) two saprotrophs (*Aspergillus nidulans*, *Neurospora crassa*) and three additional hemibiotrophs (*Leptosphaeria maculans*, *Magnaporthe oryzae*, *Mycosphaerella graminicola*). The proteomes of *Fusarium graminearum* and *Verticillium albo-atrum* were also included as representatives of genera closely related to *Colletotrichum*. Sequences were aligned using MAFFT[75] and phylogenetic trees were reconstructed using the neighbor-joining method, followed by a bootstrap test with 100 replications. Sequence editing, alignment, and phylogenetic analyses were performed using Geneious Pro (version 5.5)[76].

We identified 1,326 orphan proteins (proteins with no identifiable homologs) in *C. higginsianum* and only 582 in *C. graminicola* (**Supplementary Table 6**). The *C. higginsianum* genome also contains nearly twice as many multi-copy genes as *C. graminicola*. When we compared the *Colletotrichum* genomes to those of three other closely related members of the Sordariomycetes we found that *C. higginsianum* has the largest number of both orphan and multi-copy genes. Assuming most of the orphan genes originated by gene duplication but are now too distantly related to their originating family to be correctly classified by the clustering algorithm, these observations are consistent with a higher rate of gene duplication and faster evolution in the *C. higginsianum* genome.

The most expanded protein cluster in *C. higginsianum*, relative to *C. graminicola*, was a cluster of polyketide synthases (PKS) containing 55 more copies than *C. graminicola* (**Supplementary Fig. 6a**). Manual curation of the PKS genes confirmed a smaller expansion of this family in *C. higginsianum*, with 19 additional members compared to *C. graminicola* (**Supplementary Fig. 10a** and **Supplementary Table 8**). We also found two expanded clusters of cytochrome P450 proteins (**Supplementary Fig. 6a**). Together with the PKS enzymes, the presence of P450 proteins among the most expanded clusters suggests a requirement for greater secondary metabolite diversity in this species. Additionally, three clusters of membrane transporters are present among the

most expanded MCL clusters. Two of these are likely to mediate the export of secondary metabolites, and thus, the size and diversity of these families may have evolved concomitantly with secondary metabolism genes.

Among the protein clusters most expanded in *C. graminicola* relative to *C. higginsianum* was a family of small secreted proteins of unknown function that is expanded by five copies (**Supplementary Fig. 6a**). Another expanded cluster comprises a family of secreted histidine acid phosphatases (putative phytases), which could contribute to fungal nutrition by mobilizing phytic acid, the main form of stored phosphorus in plant tissues. One member of this cluster (GLRG_11131; *BOUP2*) was previously identified among the most strongly up-regulated genes in *C. graminicola* biotrophic hyphae isolated by laser capture microdissection[77]. Also more expanded in *C. graminicola* was a family of atypical cellulases (GH61 glycosyl hydrolases), which are specifically induced during the necrotrophic phase of infection (see **Section 5.3** and **Supplementary Figure 8b**).

Considering the top ten most expanded clusters in each species, the *C. higginsianum* clusters have 17-56 more copies than *C. graminicola*, while the top ten expanded clusters in *C. graminicola* only have 4-5 more copies than *C. higginsianum*. Thus, both the larger numbers of orphan genes, as well as the presence of larger gene families, contribute to the larger proteome of *C. higginsianum*. Phylogenetic trees of the *C. higginsianum* expanded families (**Supplementary Fig. 6b**) contain both ancient and species-specific gene duplications, indicating that gene duplication and gene loss in these families is an ongoing process, and is not restricted to the *C. higginsianum* lineage.

5. Gene category descriptions

5.1. Secretome annotation

To identify proteins that are transported out of the cell and into the extracellular space, we first evaluated the accuracy of several programs that are commonly used to identify protein localization, namely SignalP[78], TMHMM[79], WoLF-PSORT[80] and the Glycosylphosphatidylinositol-modification site (GPI anchor) described by de Groot *et al.*[81]. The accuracy of the programs was tested using 4,351 fungal proteins with experimentally validated subcellular localizations, including 273 extracellular proteins, obtained from the UniProt database. Various combinations of the programs were evaluated for their ability to correctly classify the test protein data set. The program WoLF-PSORT correctly classified the most extracellular secreted proteins (92.7%) and incorrectly classified only 4.1% of the intracellular proteins as extracellular. WoLF-PSORT was therefore selected to determine the secretomes of the *Colletotrichum* genomes. We identified 1,650 secreted proteins in the *C. graminicola* genome and 2,142 in the *C. higginsianum* genome, accounting for 13.7% and 13.2% of the proteome, respectively (**Supplementary Fig. 9a**).

5.2. Candidate Secreted Effector Proteins (CSEPs)

Many plant pathogens secrete effector proteins that facilitate infection by reprogramming host cells and modulating plant immune responses[82]. In this study, we identified candidate secreted effector proteins (CSEPs)

as predicted extracellular proteins with no significant BLAST homology to sequences in the UniProt database (both the SwissProt and TrEMBL components) using an e-value threshold of 1e-3 and excluding any BLAST hits from the genus *Colletotrichum*. A similar BLAST search was performed using a database of proteins sequences from *C. graminicola* and *C. higginsianum*. From these BLAST searches, we defined two sets of CSEPs, those that lack homology to any other protein and those that have homology to proteins only within the genus *Colletotrichum*. Among the predicted secreted proteins, we identified 177 CSEPs in *C. graminicola* and 365 in *C. higginsianum* (**Supplementary Fig. 9a**). In *C. higginsianum*, more than 72% of the CSEPs are species-specific, also lacking homologs in *C. graminicola*, while in *C. graminicola*, 48% are species-specific.

The *Colletotrichum* CSEPs have several properties that are consistent with our current understanding of fungal proteinaceous effectors. They are small proteins, with an average of 110 amino acids in *C. higginsianum* and 175 amino acids in *C. graminicola*, whereas the average sizes of proteins in the proteomes of each species are 365 and 466 amino acids, respectively. They are also cysteine-rich, with CSEPs less than 100 amino acids in length having approximately twice the cysteine content of similar sized proteins in the whole proteomes (**Supplementary Fig. 9b**).

To compare sequence diversification among the *Colletotrichum* CSEPs and those of the hemibiotroph *Magnaporthe oryzae* and the obligate biotroph *Blumeria graminis* f.sp. *hordei*, we predicted the CSEP repertoires of these pathogens using the same computational pipeline and made multiple sequence alignments (**Supplementary Fig. 9c**). The pattern of CSEP diversification observed in *C. higginsianum* and *C. graminicola* is similar to that found in *M. oryzae* but strikingly different to the *B. graminis* f. sp. *hordei* CSEP repertoire, which is characterized by the presence of large gene families[41]. Of the CSEPs identified in *C. higginsianum* and *C. graminicola*, only 14% are organized in gene families (see MCL analysis) and the largest family encompasses only 5 members in *C. higginsianum* and 3 in *C. graminicola*. As reported for *M. oryzae* effectors[83], this absence of expanded gene families might indicate that *C. higginsianum* and *C. graminicola* effectors evolved mainly by gain/loss processes rather than by nucleotide substitution.

5.3. Carbohydrate active enzymes (CAZymes)

We searched the genomes of *C. higginsianum* and *C. graminicola* for genes encoding putative carbohydrate-active enzymes using the CAZy annotation pipeline[84] (http://www.cazy.org), and compared these inventories to those of 13 other sequenced fungi representing a range of different life-styles (**Fig. 3a**; see **Section 4**). For each CAZy class, the number of enzyme modules and the families they belong to are shown in **Supplementary Table 9**. Data for enzymes acting on pectin, hemicellulose and chitin, as well as proteins with CBM18 and CBM50 carbohydrate-binding motifs, were manually curated.

All the fungi examined encode similar numbers of glycosyltransferases (GTs), i.e. enzymes that build complex carbohydrates from activated sugar donors. This trend is consistent with the GTs being involved in basal activities of the fungal cell (e.g. fungal cell wall synthesis, glycogen cycle, trehalose cycle), whereas glycoside hydrolases

(GHs), polysaccharide lyases (PLs) and carbohydrate esterases (CEs) are mostly secreted enzymes adapted to exploiting the great diversity of carbohydrates present in the exterior milieu as nutrient sources.

The genomes of *C. higginsianum* and *C. graminicola* encode a total of 361 and 307 GHs, respectively, more than all other fungi examined in this study. This large arsenal of sugar-cleaving enzymes is further extended by polysaccharide lyases belonging to families PL1, PL3, PL4 and PL9, all of which target pectin. The total number of PLs in *C. higginsianum* (39) surpasses *C. graminicola* (15) and all other fungi sequenced to date. In total, *C. higginsianum* encodes 86 enzyme modules attacking pectin, compared to only 42 in *C. graminicola* (**Fig. 3b**). The two *Colletotrichum* species possess similar numbers of enzymes degrading cellulose (GH5, GH6, GH7, GH12, GH45, GH61) and hemicellulose (CE1, GH10, GH11, GH26, GH29, GH43, GH51, GH53, GH54, GH62, GH67, GH74, GH93). Overall, the CAZyme arsenals of *C. higginsianum* and *C. graminicola* more closely resemble those of hemibiotrophic and necrotrophic plant pathogens such as *Stagonospora*, *Magnaporthe*, *Fusarium* and *Alternaria* than those of obligate biotrophs such as *Melampsora* and *Blumeria* (**Fig. 3a** and **Supplementary Table 9**).

Genome-wide expression profiling by RNA-Seq (see **Section 6**) revealed that CAZyme genes were expressed at all stages of fungal development *in planta*, but there were marked differences in their patterns of expression at the early appressorial stage, associated with penetration of the host cuticle and epidermal cell wall, and during the later necrotrophic phase, when host tissues are extensively degraded and rotted. CAZyme expression levels were overall higher during necrotrophy: thus, in *C. higginsianum*, 23 out of the top 100 most highly up-regulated genes during necrotrophy vs. appressoria were CAZymes, compared to only one among the 100 most up-regulated genes in appressoria vs. necrotrophy (**Supplementary Table 13**). Similarly in *C. graminicola*, 25 of the top 100 genes induced during necrotrophy were CAZymes, compared to only three of the top 100 genes induced in appressoria (**Supplementary Table 14**).

To compare the spectra of CAZymes expressed at early and late infection stages, we identified genes that were significantly up-regulated (log2 fold change >2, *p*-value <0.05) in pair-wise comparisons of appressoria vs. necrotrophic phase and *vice versa* (**Supplementary Tables 13** and **14**). To consider only CAZymes likely to degrade plant or fungal components, we excluded genes encoding GTs and CBM proteins without enzymatic domains from the analysis. In both species, more CAZymes were up-regulated during necrotrophy (122 in *C. graminicola* and 146 in *C. higginsianum*) than in appressoria (30 in *C. graminicola* and 87 in *C. higginsianum*, see **Supplementary Fig. 8a**). During the necrotrophic phase, each species appears to employ a different strategy to deconstruct host cell walls (**Supplementary Fig. 8a,b**). Thus, a larger number of pectin-degrading enzymes (51, including 31 PLs) were induced during necrotrophy in *C. higginsianum* compared to *C. graminicola* (16, including 9 PLs). Conversely, a larger set of enzymes degrading cellulose and hemicellulose were induced during necrotrophy in *C. graminicola* (52) than in *C. higginsianum* (27). These findings are consistent with the higher pectin content of dicot cell walls (~35%, compared to 10% in maize), and the higher hemicellulose content of maize cell walls (up to 61%, compared to ~30% in dicots)[85,86]

Remarkably, *C. graminicola* activates a much larger set of 22 GH61 genes during necrotrophy (**Supplementary Fig. 8b**), including six with appended CBM1 cellulose-binding (CBM1) domains, and nine GH61s are among the 100 most highly induced genes during necrotrophy in this species (**Supplementary Table 17b**). In contrast, *C. higginsianum* activates only six GH61s at this stage, none of which feature among the 100 most highly induced genes. The GH61 proteins that have been experimentally characterized are copper-dependent monooxygenases which act in concert with endoglucanases and cellobiohydrolases to enhance cellulose hydrolysis[87,88], although in principle GH61 enzymes could also cleave other cell wall components. The contrasting numbers of these enzymes deployed during necrotrophy by each *Colletotrichum* species may reflect differences in cell wall digestibility between maize and dicots. Thus, grass cell walls are enriched with phenolic molecules such as hydroxycinnamates and lignins[86], which can act as cofactors for GH61s, but also make cellulose more resistant to classical hydrolases[87,89].

Although fewer CAZyme genes were up-regulated during the appressorial phase of *C. higginsianum*, 35 different CAZyme families were represented. These included enzymes with the potential to degrade most major plant cell wall polymers, e.g. cutin (4), cellulose (8), hemicelluloses (8) and pectin (9). Such enzymes may contribute to the initial penetration of host epidermal cell walls, acting in concert with mechanical forces[14]. However, the largest set of genes induced at this early infection stage encode enzymes for modifying fungal cell walls (**Supplementary Fig. 8a**), including 12 enzymes involved in glycoprotein synthesis/processing (GH47, GH76, GH92), seven chitin deacetylases (CE4), five chitinases (GH18) and 15 enzymes cleaving β-1,3-glucans (GH16, GH17, GH55, GH64), although the latter could potentially also degrade plant-derived glucans deposited at fungal entry sites (callose papillae). Remodelling of the fungal cell surface during appressorium-mediated penetration may be required for evading detection by plant immune receptors or attack by plant hydrolases[90]. A smaller and less diverse set of CAZyme genes belonging to 18 different families were induced in *C. graminicola* appressoria, among which GH16 β-1,3-glucanases (8) were the most abundant. This may indicate a greater role for mechanical penetration in the appressoria of this species.

5.3.1. Proteins containing CBM50 modules

CBM50 carbohydrate-binding modules, also known as LysM or lysin motifs, comprise 40-60 amino acid residues and bind to chitin and peptidoglycans[91]. Based on the CAZy annotation, both *Colletotrichum* genomes encode more proteins with CBM50 modules than all other fungi examined (**Supplementary Table 9**). In addition to these predictions, we identified further proteins containing putative CBM50 modules based on the Broad Institute annotation and NCBI Conserved Domain Database. In total, we found 20 proteins containing CBM50 modules but no enzymatic domains in *C. higginsianum* and 14 in *C. graminicola*, of which 9 and 5, respectively, were predicted to be extracellular secreted proteins by both SignalP and WolfPSort (**Supplementary Table 10**). These CBM50 proteins may be chitin-binding lectins. The CAZy annotation indicates an additional three proteins with CBM50 domains (CH063_04609, CH063_06344 and GLRG_11925) that also contain GH18 chitinase catalytic

domains. The non-chitinase CBM50 genes appear to be evolving rapidly because ortholog assignment between the two species was not possible.

Secreted CBM50 proteins could function as effectors, contributing to a fungal stealth strategy by concealing surface-exposed chitin or sequestering chitin fragments to evade recognition by plant immune receptors[91,92]. In *C. lindemuthianum*, a secreted protein with tandem CBM50 modules called CIH1 was previously localised specifically on the surface of intracellular biotrophic hyphae using a monoclonal antibody[93,94], which also labels the biotrophic hyphae of *C. higginsianum*[40]. Two *C. higginsianum* genes, CH063_13023 and CH063_04445, showing the greatest homology to *CIH1* (41% and 50% amino acid identity, respectively) encode small secreted proteins of 170 and 176 amino acids, each containing two CBM50 modules. Transcriptome profiling showed that both genes are strongly upregulated *in planta* during the biotrophic and early necrotrophic stages (**Supplementary Table 13**). There was no ortholog of CH063_13023 in *C. graminicola*, but the ortholog of CH063_04445 (GLRG_02947) is also significantly induced during early infection (**Supplementary Table 14**). Four additional *C. higginsianum* LysM proteins are induced in the necrotrophic stage: (CH063_05398, CH063_06689, CH063_04323 and CH063_08917). Most of these lack orthologs in *C. graminicola*, while the ortholog of CH063_05398 (GLRG_06565) is poorly expressed at all stages.

5.3.2. Proteins containing CE4 modules

CE4 polysaccharide deacetylase modules are likely to function in the deacetylation of chitin. Both *C. graminicola* and *C. higginsianum* have 15 gene models encoding proteins with predicted CE4 modules, a total surpassed only by the obligate biotroph *Melampsora* among the fungi examined in our comparative analysis (**Supplementary Table 9** and **Supplementary Table 11**). Orthologs between the two *Colletotrichum* species were found for 14 of the 15 chitin deacetylases, with amino acid identity of 59 to 95% (median 83%). In both species, nine of these enzymes are predicted to be secreted, including one containing a potential C-terminal glycosyl-phosphatidylinositol (GPI) anchor mediating linkage to the extracellular face of the plasma membrane or to β-1,6-glucans in the fungal cell wall[95]. By converting chitin to chitosan, deacetylase enzymes may contribute to pathogen counter-defence by avoiding lysis of the fungal cell wall by plant chitinases[90]. Given that both chitin and chitosan can be recognised by plants as microbe-associated molecular patterns[96], such cell wall remodelling is less likely to contribute to pathogen concealment. Two orthologous genes, GLRG_02208 and CH063_14486, share 47% and 48% amino acid identity with *Magnaporthe oryzae* CBP1, a chitin deacetylase with two chitin-binding CBM18 modules implicated in surface contact-induced development of appressoria[97]. Transcriptome profiling in *C. higginsianum* showed that three genes encoding secreted chitin deacetylases are differentially regulated *in planta*, being induced in pre-penetration appressoria (CH063_15385), the biotrophic phase (CH063_02920) or the necrotrophic phase (CH063_03329) (**Supplementary Table 13**). In *C. graminicola*, orthologs of all three genes (GLRG_11238, GLRG_07915, and GLRG_04084) are upregulated later during infection (**Supplementary Table 14**).

5.4. Secreted proteases

To search for *C. higginsianum* and *C. graminicola* genes encoding secreted peptidase enzymes, sequences of proteins predicted to have an extracellular location by WolfPSort (see Supplementary section 5.1) were subjected to MEROPS Batch Blast analysis[98]. For comparison, we also analysed the predicted secretomes of 13 other sequenced fungi: *Alternaria brassisicola*, *Botrytis cinerea*, *Stagonospora nodorum*, *Sclerotinia sclerotiorum*, *Melampsora laricis-populina*, *Blumeria graminis* f.sp. *hordei*, *Aspergillus nidulans*, *Neurospora crassa*, *Leptosphaeria maculans*, *Magnaporthe oryzae*, *Mycosphaerella graminicola*, *Fusarium graminearum* and *Verticillium albo-atrum*. Peptidase family S9 homologs were excluded from the final inventories because they are likely to be alpha/beta hydrolases other than peptidases (e.g. carboxylesterases and lipases). Numbers of serine proteases were manually curated for *C. higginsianum*.

We found that *C. higginsianum* encodes a larger number of putative secreted proteases (151) than *C. graminicola* (110), also surpassing all other fungi included in the comparison (**Supplementary Fig. 7**). In particular, *C. higginsianum* shows a lineage-specific expansion of the S08A family of subtilisin serine proteases with 36 members, compared to only 12 in *C. graminicola*. Both *Colletotrichum* species also encode larger numbers of metalloproteases than all other fungi examined. In *C. higginsianum*, 14 serine protease and 22 metalloproteases were significantly induced *in planta* at the transition to necrotrophy (**Supplementary Table 13** and **Supplementary Fig. 11**). Serine proteases and metalloproteases may contribute to plant cell wall degradation at this stage by targeting structural proteins such as hydroxyproline-rich glycoproteins[99,100]. In contrast to *C. higginsianum*, more secreted proteases were upregulated in *C. graminicola* during the biotrophic phase (**Supplementary Tables 14** and **17**). This suggests that proteolysis is more important for fungal nutrition or degrading plant defense-related proteins, or involved in localised degradation of the plant cell wall during cell-to-cell penetration by biotrophic hyphae.

5.5. Membrane transporters

Based on BLAST searches against the Transporter Collection Database (www.tcdb.org), the *Saccharomyces* Genome Database (www.yeastgenome.org) and the NCBI Reference Sequences database (www.ncbi.nlm.nih.gov), we identified a total of 647 and 754 membrane transporter genes in *C. graminicola* and *C. higginsianum*, respectively (**Supplementary Table 7a**). The Major Facilitator Superfamily (MFS), Oligopeptide (OPT) and ATP-binding cassette (ABC) transporter families are substantially expanded in both *Colletotrichum* species compared with biotrophic fungi, including *Puccinia graminis* f.sp. *tritici*, *Melampsora larici-populina*, *Blumeria graminis* f.sp. *hordei* and *Ustilago maydis* (**Supplementary Table 7b**). Respectively, *C. graminicola* and *C. higginsianum* have 299 and 363 MFS transporters, 16 and 18 OPT transporters and 52 and 68 ABC transporters. The overall numbers of MFS and ABC transporters are 2 to 4 times higher in *Colletotrichum* than in the biotrophs, with similar numbers present in other hemibiotrophic and necrotrophic phytopathogens, including *Magnaporthe oryzae*, *Botrytis cinerea*, *Verticillium albo-atrum* and *Fusarium oxysporum* (**Supplementary Table 7b**). Among the MFS transporters, the Sugar Porter and Anion:Cation Symporter subfamilies are four and six

times larger in *Colletotrichum* than in the biotrophic fungi, respectively. The Drug:H+-Antiporter subfamilies DHA1 and DHA2 are also three times larger in *Colletotrichum*. The expansion of transporter families required for the efflux of host or fungal metabolites and nutrient acquisition suggests that evasion of host defences, delivery of small molecule toxins and/or effectors and assimilation of a broad range of host nutrients are important aspects of *Colletotrichum* pathogenesis. One MFS inorganic phosphate transporter (CH063_07455) was previously found to be essential for pathogenicity of *C. higginsianum*[37].

The transporter families that were most expanded in *Colletotrichum* compared to the biotrophic fungi also showed the greatest differential gene gain/loss between *C. graminicola* and *C. higginsianum*, with 25 to 60% having no identifiable ortholog (**Supplementary Table 7a**), suggesting that transporter genes are evolutionarily dynamic in *Colletotrichum*. Overall, *C. higginsianum* has more MFS transporters than *C. graminicola*, with the Anion:Cation Symporter, Pleiotropic Drug Resistance, Drug:H+-Antiporter DHA2, Monocarboxylate Porter, Sugar Porter, and Siderophore Iron Transporter subfamilies being particularly expanded (**Supplementary Table 7a** and **Supplementary Figure 6**). Expansion of Siderophore Iron Transporters in *C. higginsianum* (6 additional copies) suggests that during host infection iron may be less available to this species, consistent with the greater alkalinization of *Arabidopsis* tissue observed during *C. higginsianum* infection (**Supplementary Figure 13**), which is likely to reduce the bioavailability of Fe ions.

Transcript profiling of *C. higginsianum* infecting *Arabidopsis* revealed that the expression of many membrane transporter genes is highly stage-specific. Comparing RNA-seq data from *in vitro* and *in planta* appressoria, no fewer than 131 transporter genes were significantly induced upon host contact (log2 fold change>2, p<0.05) (**Supplementary Table 13**). Among the most highly induced transcripts (log2 fold change>3) were 14 Sugar Porters, 18 Drug:H^+ Antiporters (9 DHA1 and 9 DHA2), 6 Pleiotropic Drug Resistance (PDR) and 5 Drug Conjugate Transporters (DCT). While small molecule influx/efflux could represent an important function of appressoria formed *in planta*, it is also possible that transcripts coding for sugar transporters already accumulate in appressoria in order to provide the corresponding transport proteins at the onset of host cell invasion and establishment of the biotrophic phase.

In contrast to the large number of transporter genes induced in appressoria *in planta*, only 38 were significantly induced in the biotrophic phase compared to appressoria *in planta* (log2 fold change>2, p<0.05) (**Supplementary Table 13**). Among those, only three sugar transporters were identified (CH063_06868, CH063_02254, CH063_06485) including one hexose transporter. However, none of them were preferentially expressed during the biotrophic phase since their transcripts showed higher accumulation in the necrotrophic phase. Analysis of gene expression level of the 94 *C. higginsianum* sugar porters across the four fungal developmental stages also revealed that only five are preferentially expressed in the biotrophic phase, including an ortholog of the characterized *C. graminicola* hexose transporter *CgHXT1* (CH063_06203[101]). In addition, some sugar porter transcripts were detected in more than one infection stage, such as the *C. higginsianum* ortholog of the *C. graminicola* hexose transporter *CgHXT3* (CH063_11110[101]) and the melibiose transporter *CgMBT1* (CH063_00539[102]). Although no specific reprogramming of nutrient transporter transcription was observed during

biotrophy, the basal expression of a relatively small number of sugar porters at this stage might be sufficient for significant uptake of hexoses and disaccharides. Many more genes encoding transporters of small molecules were induced during biotrophy, including 22 MFS transporters (5 DHA1, 4 DHA2) and 4 ABC transporters (PDR family). These findings suggest that plasma membrane efflux and vacuolar sequestration of fungal secondary metabolites and/or host antimicrobial compounds are major activities during intracellular colonization of living host cells.

A large set of 75 transporter genes were significantly induced in the necrotrophic phase compared to the biotrophic phase (log2 fold change>2, p<0.05), including 18 sugar transporters (**Supplementary Table 13**). Among those, 17 showed higher expression during necrotrophy compared to the three other fungal developmental stages, suggesting a specific role during necrotrophy. This suggests that specialized suites of transporters are required for uptake of the diverse sugars released by lytic enzymes during the necrotrophic phase, and that *C. higginsianum* is well-adapted to this dynamic spectrum of carbon sources. In addition, four Na^+ P-type ATPases (CH063_09484, CH063_11329, CH063_03065 and CH063_09618) were strongly induced. These are homologs of yeast ENA proteins, which are involved in sodium and pH homeostasis and are known to be induced under conditions of high ambient pH[103]. Expression of these sodium efflux pumps might reflect the strong alkalinisation of host tissue at this stage of *C. higginsianum* infection (**Supplementary Figure 13**).

5.6. Secondary metabolism-related genes

Fungi produce a broad range of secondary metabolites (SM) with important functions, such as antibiosis, protection from stresses and pathogenicity to animals and plants[104]. Fungi produce four major groups of SM: polyketides produced by polyketide synthases (PKS); peptides produced by nonribosomal peptide synthases (NRPS); alkaloids produced by dimethylallyl tryptophan synthases (DMATS); and terpenes produced by terpene synthases (TS)[104]. PKS, NRPS, and PKS-NRPS hybrids are the most abundant SM biosynthetic genes in fungi[104-106]. *Colletotrichum* species produce diverse SM, including flavones, peptides and terpenes[107,108], as well as DHN (1,8-dihydroxynaphthalene) melanin, essential for appressorium-mediated host penetration[109]. Although SM have not been described in *C. higginsianum*, several were characterized in *C. graminicola*, including the antifungal compounds monorden and monicillins I, II, and III[110] and mycosporine-alanine, a spore germination inhibitor[111]. Deletion of *PPT1*, a gene that encodes a cofactor essential for the enzymatic function of all PKS and NRPS, resulted in decreased pathogenicity in *C. graminicola*[112]. Necrotrophs and hemibiotrophs generally produce large numbers of SM, while biotrophs produce very few[41]. We evaluated the capacity of the hemibiotrophic *C. higginsianum* and *C. graminicola* to produce SM by analyzing their repertoires of SM genes and the expression of these during infection.

5.6.1. Annotation of secondary metabolism genes

Candidate SM genes were initially identified by MCL analysis and gene family searches using the Broad Institute *Colletotrichum* database, BLAST searches against the NCBI databases, and InterproScan analysis[74]. The Fungal Cytochrome P450 Database (FCPD) pipeline (http://p450.riceblast.snu.ac.kr/) was also used to annotate

cytochrome P450 enzymes. The Secondary Metabolite Unknown Region Finder (SMURF) web-based program[106] was applied to both *Colletotrichum* genomes. In the case of *C. higginsianum*, SMURF was applied to the less fragmented Velvet assembly (see **Section 2.2**). SMURF was also applied to the most current lists of protein sequences from several other sequenced fungi.

Candidate PKS, NRPS, PKS-NRPS hybrid, DMATS, and TS genes identified using automated searches were subjected to manual annotation, including visual inspection of nucleotide/protein sequences, alignments with protein sequences of known enzymes and identification of conserved domains using CDD (http://www.ncbi.nlm.nih.gov/Entrez). Phylogenetic analysis of PKS, NRPS, PKS-NRPS, TS, and DMATS proteins was performed using phylogeny.fr (http://www.phylogeny.fr/) with Muscle (V3.7) and PhyML (V3.0). Statistical branch support was provided by a standard Likelihood Ratio Test, aLRT[112]. Trees were constructed using the AMP binding adenylation domains from NRPS genes, and the ketoacyl synthase N-terminal and C-terminal domains from PKS genes. PKS-NRPS, DMATS and TS were analyzed by aligning whole protein sequences, and by clustering with OrthoMCL and COCO-CL (http://www.orthomcl.org/cgi-bin/OrthoMclWeb.cgi).

5.6.2. Terpene synthases

Terpene synthases produce terpenes (e.g. carotenoids, gibberellins and trichothecenes) from dimethylallyl diphosphate *via* the isoprenoid pathway. There are 17 putative TS genes in *C. higginsianum* and 14 in *C. graminicola* (**Supplementary Table 8**). These TS include CH063_03786 of *C. higginsianum*, and GLRG_02475 of *C. graminicola*, which are orthologous to fungal squalene/phytoene synthases involved in carotenoid biosynthesis[113]. Phylogenetic and Ortho-MCL/COCO-CL analyses revealed that 10 TS are shared between the two species, while 4 are unique to *C. graminicola* and 7 are unique to *C. higginsianum*.

5.6.3. Dimethylallyl tryptophan synthases

DMATS produce alkaloids (e.g. ergotamine) from tryptophan *via* the indole pathway. There are 10 putative DMATS in *C. higginsianum* and 7 in *C. graminicola* (**Supplementary Table 8**). Phylogenetic and Ortho-MCL/COCO-CL analyses showed 6 DMATS are shared between these two species, while 1 is found only in *C. graminicola* and 4 only in *C. higginsianum*.

5.6.4. Polyketide synthases

PKS genes contain three essential modules encoding the ketoacyl CoA synthase (KS), acyl transferase (AT), and acyl carrier (ACP) domains, and up to three optional modules encoding ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) domains. Both *C. graminicola* and *C. higginsianum* have a very large number of predicted PKS genes. A phylogenetic tree based on alignments of the highly conserved KS domains from the Broad Institute gene models[114], showed evidence of gene expansion in this family. For example *C. higginsianum* genes CH063_00511 and CH063_01345 appear to be paralogs, as do GLRG_03511 and GLRG_05714. However, some of the putative *C. higginsianum* paralogous PKS genes matched to a single predicted ORF in the

Velvet assembly (e.g. CH063_10861 and CH063_02506). After resolving these overlapping gene models as far as possible, we conclude that *C. higginsianum* has 58 PKS genes while *C. graminicola* has 39 (**Supplementary Table 8**). However, these numbers may still be inaccurate, particularly for *C. higginsianum*. Based on phylogenetic analysis of KS domains from the Broad Institute gene models (45 from *C. higginsianum* and 34 from *C. graminicola*), at least 15 are shared between the two species. Further analysis is required before final conclusions can be reached about this group of genes.

5.6.5. Non-ribosomal peptide synthases

Non-ribosomal peptides are produced by multimodular enzymes encoded by NRPS genes. Each NRPS contains several conserved domains including an AMP binding domain (A), a peptidyl carrier domain (T), and a condensation domain (C). The highly conserved A domain, required for amino acid recognition and activation, has been used previously to construct NRPS phylogenetic trees[115]. Due to their large size and modular nature, NRPS genes are often fragmented in the Broad Institute annotation. We identified many cases in which two or more *C. higginsianum* gene models are actually fragments of a larger NRPS gene. For example, CH063_15533, CH063_15485, CH063_10172, CH063_01270, CH063_01271, CH063_10216, CH063_012138 and CH063_05450 are all fragments of a single 36 Kb NRPS gene, predicted to contain 10 ATC modules (*ChNRPS2*). Adenylation domains from the re-annotated NRPS genes were used to construct a phylogenetic tree, on the basis of which we conclude there are 7 NRPS genes in *C. graminicola*, 6 being shared with *C. higginsianum*, while *C. higginsianum* has 12 NRPS, including 6 unique to that species (**Supplementary Table 8**).

5.6.6. PKS-NRPS hybrids

PKS-NRPS hybrids consist of a PKS fused to a single NRPS module and are involved in production of several important SM, e.g. lovastatin and fusarin C[105]. Using a combination of manual annotation and phylogenetic analysis, we concluded that *C. higginsianum* has 6 PKS-NRPS hybrids, while *C. graminicola* has 7 (**Supplementary Table 8**). Only 2 PKS-NRPS are shared between the two species, while 4 are unique to *C. higginsianum* and 5 are unique to *C. graminicola*.

Overall, we conclude that *C. higginsianum* has more SM genes than *C. graminicola* and that this results from expansion of specific SM gene families. Both *Colletotrichum* species have an unusually large number of SM genes when compared to other sequenced fungi with different lifestyles (**Supplementary Fig. 10a**), consistent with hemibiotrophs generally having more SM genes[41,105,116,117].

5.6.7. Cytochrome P450 enzymes

Cytochrome P450 monooxygenases can play important roles in modifying secondary metabolites and are often included in SM gene clusters[118]. Using the FCPD pipeline, a total of 230 and 148 putative P450 proteins were identified from *C. higginsianum* and *C. graminicola*, respectively. The CYP65 and CYP68 classes, which are typically involved in secondary metabolism[118], are well-represented, comprising 12% and 5% of the P450s in

each species, respectively. Lists of genes encoding P450 proteins in each species are given in **Supplementary Tables 13** and **14**, and complete results can be found at: http://p450.riceblast.snu.ac.kr/species.php.

5.6.8. Secondary metabolism gene clusters

SM genes are typically clustered in the genome. We used the web-based prediction tool SMURF to predict SM gene clusters. For *C. graminicola*, 70% of SMURF-predicted "backbone" genes were placed into clusters by the program, compared with only 42% in *C. higginsianum*. Underestimation of the number of clusters in *C. higginsianum* by SMURF probably results from the more fragmented nature of this genome assembly. SMURF predicted 39 SM gene clusters for *C. higginsianum*, ranging in size from 2 to 16 genes (average 8; **Supplementary Table 12**). For *C. graminicola*, 42 clusters were predicted, ranging from 2 to 24 genes (average 7; **Supplementary Table 12**). SMURF identified a larger number of SM clusters in both *Colletotrichum* species compared to most other sequenced fungi with various life-styles (**Supplementary Fig. 10b**). Nevertheless, the list of clusters predicted by SMURF for both fungi is incomplete, since it included neither the melanin nor the carotenoid biosynthesis gene clusters. SMURF also failed to predict several clusters containing PKS-NRPS and terpene synthase genes, and frequently predicted the borders of SM clusters incorrectly, as evidenced by the differential regulation of genes within some predicted clusters (**Supplementary Tables 13** and **14**).

Although not predicted by SMURF, the melanin biosynthesis gene cluster comprises orthologs of *CMR1, T_4HR1, CMR2, PKS1* and *FET3*[119-122]. This cluster is perfectly conserved in gene content, order and orientation between *C. higginsianum* and *C. graminicola* as well as *C. orbiculare* and the distantly related Sordariomycete *M. oryzae*, although in the latter species the orientation of *CMR2* is reversed compared to *Colletotrichum*. Two other structural genes, *SCD1* and *THR1*[123,124], are not clustered in either *Colletotrichum* or *M. oryzae*, in contrast to melanin gene clusters in *Aspergillus* and *Penicillium* species which include *SCD1* orthologs[125,126]. Nine other clusters predicted by SMURF, together with the carotenoid biosynthetic cluster and one terpene synthase cluster, are conserved between the two *Colletotrichum* species, although gene order differs (**Supplementary Table 12**). In one of the most highly conserved clusters (*C. higginsianum* cluster 10, *C. graminicola* cluster 18), all genes are shared, but gene inversion has affected their order (**Fig. 4b**). This cluster is related to the Alternapyrone biosynthesis cluster of *Alternaria*[127]. Despite this high level of conservation, the expression patterns are strikingly different between the two *Colletotrichum* species (**Fig. 4b** and **Supplementary Fig. 12**). In contrast, the less well-conserved cluster 35 in *C. graminicola* and cluster 15 in *C. higginsianum*, which share only half of their genes, are both highly induced in appressoria (**Supplementary Tables 13** and **14**). Overall, 60% of *C. higginsianum* clusters contain at least one syntenic block comprising two or more *C. graminicola* genes, suggesting a common ancestry.

The majority of SM gene clusters appear to be lineage-specific. One *C. higginsianum*-specific cluster contains a hybrid PKS-NRPS (CH063_03067) orthologous to *ACE1* of *M. oryzae*, encoding a secondary metabolite effector that is recognized by the rice Pi33 resistance gene[105,128]. The *ACE1* cluster, not predicted by SMURF, was reconstructed by manual annotation. It includes 8 genes with *M. oryzae* orthologs, three of which were not predicted in the Broad Institute annotation of *C. higginsianum*. The PKS-NRPS CH063_03253 and CH063_03254

matched to a single FGENESH open reading frame orthologous to *M. oryzae SYN2*[105]. Previously, *Chaetomium* was the only other fungus apart from *M. oryzae* to possess both *ACE1* and *SYN2* clusters[129]. In *M. oryzae*, both *ACE1* and *SYN2* clusters are up-regulated in appressoria[105]. However, in *C. higginsianum*, only the *ACE1* cluster was similarly induced at early infection stages, whereas the *SYN2* cluster was silent under the conditions examined. While *C. graminicola* lacks both of these gene clusters, it has one (Cluster 38) not present in *C. higginsianum* that is identical in gene content and gene order to one producing monorden in *Pochonia chlamydosporia*[130,131], and is therefore likely to produce the same compound in *C. graminicola*.

The 42 SMURF-predicted SM gene clusters of *C. graminicola* were manually inspected for the presence of repetitive elements: 71 % of the clusters were associated (within 10 kb) with one or more repetitive sequences including LTR transposons, DNA transposons, and MITE repeats. All of the SM gene clusters located on small, unanchored scaffolds, where the only genes on the scaffold belong to an SM gene cluster, have repetitive DNA at both ends, e.g. *C. graminicola* clusters 35 and 38. The frequent association of repetitive sequences with SM gene clusters could provide an explanation for the cluster diversity between the two species, since portions of the clusters could recombine within these repetitive elements. A similar analysis of *C. higginsianum* was not informative because most repetitive DNA had been excluded from the assemblies.

5.6.9. Phytohormone biosynthesis genes

Plant pathogenic microorganisms can affect their hosts by manipulation of host hormone levels[132,133]. The genomes of *C. higginsianum* and *C. graminicola* were surveyed to identify genes that could play roles in manipulation of plant hormone metabolism (**Supplementary Table 8**). Using homology-based analyses, potential orthologs of an auxin efflux transporter as well as enzymes for auxin biosynthesis *via* indole-3-acetamide and indole-3-pyruvate pathways were identified in both genomes, consistent with experiments demonstrating the presence of intermediates for both pathways in *C. gloeosporioides*[134]. Further, in support of a study on *C. gloeosporioides* suggesting that levels of tryptophan, rather than enzymes, are the limiting factor for fungal auxin synthesis, expression of the genes involved were relatively constant between the developmental stages analysed[135]. Several *C. higginsianum* genes were also found to share homology with enzymes involved in ethylene, jasmonic acid and gibberellin synthesis. Experiments have indicated that *C. musae*, which infects bananas, is able to synthesize ethylene, although this ability does not appear to be important for infection[136]. On the other hand, the genes potentially involved in gibberellin synthesis were not conserved in *C. graminicola* and experimental evidence for the synthesis of jasmonic acid and gibberellin by *Colletotrichum* sp. is lacking.

Apart from genes involved in plant hormone synthesis, potential salicylic acid (SA) and ethylene degradation enzymes were also identified. SA-signaling is known to be important for resistance against biotrophic pathogens and interestingly we found one of the three SA hydroxylases encoded by *C. higginsianum* (CH063_02138) is highly expressed during biotrophy. Ethylene-insensitive plants have been shown to be more susceptible to *Colletotrichum* infection[137], indicating that pathogen degradation of ethylene *in planta* could contribute to pathogenicity. Potential ethylene receptors (GLRG_10611; CH063_11183) are encoded by both species, and a

potential ABA receptor was detected in *C. higginsianum* (CH063_09383). These could be required for pathogen perception of plant signals during infection. The induction of germination and appressorium formation in *C. gloeosporioides* by exogenous ethylene is consistent with an ability to sense this plant hormone[138].

5.7. Transcription factors

Genome sequences from two *Colletotrichum* species were uploaded to the Fungal Transcription Factor Database (FTFD)[139] and subjected to an embedded analysis pipeline for systematic comparison between and among species. FTFD identified 542 putative transcription factors from *C. higginsianum* and 551 from *C. graminicola*, comprising 3.36% and 4.58% of the total proteome, respectively. These putative transcription factors belong to 39 families, none of which were notably expanded in either species. Lists of genes encoding transcription factors in each species are given in **Supplementary Tables 13** and **14**, and are also available from http://ftfd.snu.ac.kr/.

6. Whole-genome transcriptome profiling

6.1. Sample preparation and RNA extraction

C. higginsianum: *Arabidopsis* leaves with high-density *C. higginsianum* infections were obtained as described previously[40]. To sample the pre-penetration stage (PA, 22 hours post-inoculation, hpi) and the early biotrophic stage (BP, 40 hpi), pre-penetration appressoria, and post-penetration appressoria with young biotrophic hyphae, respectively, were collected by epidermal peeling. For this, the adaxial leaf surface was adhered with double-sided tape and the epidermis was quickly stripped off using tweezers. Pieces of remaining mesophyll were quickly cut off and the epidermal peels were flash-frozen in liquid nitrogen. Microscopic spot-checks confirmed the identity of samples, i.e. absence of biotrophic hyphae in PA, absence of necrotrophic hyphae in BP. To obtain RNA from the switch between biotrophy and necrotrophy (NP), sections of the leaf lamina including the very first appearing pin-point water-soaked lesions (approx. 60 hpi) were harvested. RNA was extracted using the RNeasy Mini Kit (Qiagen), including DNase treatment. Each experimental repetition was based on RNA extracted from ~300 leaves (typically 15 mm^2 per leaf). RNA from *in vitro* appressoria was prepared as described previously[38]. Total RNA was checked on an Agilent 2100 Bioanalyzer to confirm RNA integrity.

C. graminicola: Maize leaf sheaths were inoculated essentially as described by Kankanala *et al.*[140] with minor modifications. Maize inbred Mo940 plants at the V3 stage of development were used. Leaf sheaths were removed from the oldest true leaf on each plant and cut into segments approximately 5cm long. The segments were placed in Petri dishes containing wet filter paper, supported in a horizontal position with the midrib at the bottom. Spores of *C. graminicola* strain M1.001 were collected from 2-week-old potato dextrose agar (PDA) cultures, rinsed three times in sterile water, and the concentration was adjusted to $5x10^5$ spores/mL. Two 10 µl drops of inoculum were applied to each leaf sheath, approximately 1 cm from either end, and the inoculated sheaths were incubated at 23°C under continuous light. Sheaths containing mature pre-penetration appressoria (PA, approximately 24 hpi), intracellular biotrophic hyphae prior to symptom development or emergence of

secondary hyphae (BP, approximately 36 hpi), and necrotrophic stages in which water soaked lesions were evident and secondary hyphae were visible (NP, approximately 60 hpi) were collected. Each leaf sheath was inspected with a light microscope to confirm the identity of samples, i.e. absence of primary hyphae in PA, and absence of secondary hyphae in BP. For the BP and NP samples, each sheath segment was gently cleaned with a sterile cotton swab to remove unadhered spores and superficial mycelia. For the PA and BP samples, each sheath was shaved with a sharp razor blade to remove uninfected cell layers below the epidermal layer as described by Kankanala *et al.*[140]. Each sheath was trimmed to include only the inoculated area, and the pieces were transferred into microfuge tubes and flash-frozen in liquid nitrogen. Sheath pieces were maintained at -80 °C until RNA extraction. Total RNA was extracted as described by Metz *et al.*[141]. First the frozen, ground tissue was extracted with Trizol reagent (Invitrogen), followed by further RNA purification by using the RNeasy Plant Mini Kit (Qiagen). DNase A digestion was performed according to the manufacturer's protocol. Approximately 15 sheaths were pooled for each experimental repetition. RNA integrity and quantity were measured with an Agilent 2100 Bioanalyzer before sequencing.

6.2. RNA sequencing, data analysis and heatmaps

Twelve *C. higginsianum* libraries (4 developmental stages, 3 biological replicates) and nine *C. graminicola* libraries (3 developmental stages, 3 biological replicates) were prepared according to the Illumina TruSeq™ RNA Sample Preparation Kit. Two µg of PolyA-enriched RNA per sample was processed, including chemical fragmentation, first-strand cDNA synthesis, second-strand cDNA synthesis, end-repair, 3'end adenylation, adapter ligation (including different barcoding indices) and PCR amplification. After quality control analysis, libraries were pooled (3 libraries per pool), and for *C. higginsianum* this was followed by gel-based size selection and single-read sequencing (1-2 lanes/pool) with a read length of 100 bp (including 7 bp for the barcode adaptor) using the Illumina Genome Analyzer IIx, according to standard Illumina protocols. For *C. graminicola*, 1-4 lanes per pool were sequenced, with a read length of 76 bp (including 7 bp for the barcode adaptor). The clusters were identified, signal purity assessed, and the bases called using Illumina Realtime Analysis (RTA 1.8) and Sequence Control Software (RTA 2.8). The samples were demultiplexed and sequences were converted to fastq format using BclConverter-1.7.1 and CASAVA-1.7.0 (demultiplex.pl and GERALD.pl). For *C. higginsianum*, 11.5M to 13.3M filtered reads (1.23 Gb) were obtained for each of the VA samples, and 24.3M to 27.8M filtered reads (2.58 Gb) were obtained for the BP, PA, and NP samples. For *C. graminicola*, the totals were 119.7M (8.4 Gb), 33.5M (2.3 Gb), and 33.7M (2.4 Gb) filtered reads for the PA, BP, and NP samples, respectively.

The RNAseq reads were mapped to the annotated genomes with TopHat (a=10, g=5)[50] and transformed into counts per annotated gene per sample with the 'coverageBed' function from the BEDtools suite[142] and custom R scripts. Differentially expressed genes between two developmental stages were detected using the 'exactTest' function from the R package EdgeR[143] with a common dispersion estimate and library size as offset. All *P* values were corrected for false discoveries due to multiple hypthosis testing using the Benjamin-Hochberg procedure. To calculate fold changes, the number of reads for each gene in each library was normalized by the total number of mapped reads for the library. Transcripts with a significant *P* value (<0.05) and more than a two-fold change

(log2) in transcript level were considered as differentially expressed. All codes for the RNA-seq pipeline are available upon request. The *C. higginsianum* RNA seq data were also mapped onto the un-annotated Velvet genome assembly using bowtie[144] and visualised with samtools (http://samtools.sourceforge.net/) and the IGV-browser (http://www.broadinstitute.org/igv/). Heatmaps of gene expression profiles were generated using the Genesis expression analysis package[145]. To derive expression patterns of genes, relative expression indexes (REI) were calculated as the ratio between the normalised number of reads detected for a given gene at a given fungal developmental stage and the geometrical mean number of reads (normalised) calculated across the four fungal developmental stages[146]. To generate lists of differentially expressed genes, the 100 most highly expressed genes that were significantly regulated ($P < 0.05$, log2 fold-change > 2) in each fungal developmental stage were selected.

6.3. Quantitative RT-PCR

C. higginsianum: First-strand cDNA was synthesized from 1 μg DNase-treated total RNA using the iScript cDNA synthesis kit (Bio-Rad) in a total volume of 20 μL. For amplification, 10 ng of RT products were amplified in 1X iQ SYBR Green Supermix (Bio-Rad) with 1.6 μM of specific 5'- and 3'-primers using the iQ5 Real-time PCR detection system (Bio-Rad). Specific primers amplifying fragments ranging from 87 to 329 nucleotides were designed with the Primer 3 and Amplify 3X programs and primer efficiency, ranging between 80 and 119%, was assessed for each target sequence in preliminary experiments. BLASTN searches against the *C. higginsianum* and *A. thaliana* genomes were performed to rule out cross-annealing artefacts. GeNorm[147] (http://medgen.ugent.be/~jvdesomp/genorm/) was used to assess the expression stability of five commonly used reference genes encoding alpha-tubulin, beta-tubulin, elongation factor 1-alpha, actin and glucose-6-phosphate 1-dehydrogenase[148], of which alpha-tubulin and actin were the most stable (stability value 0.047 and 0.051, respectively) and used to normalize gene expression employing Pfaffl calculation[149].

C. graminicola: One μg of DNase and RNase-treated total RNA was used to synthesize the first strand of cDNA, using the SuperScript II reverse transcriptase kit (Invitrogen) in a volume of 20 μl. Primers were designed to amplify fragments between 100-200 bp, using PrimerQuest (Integrated DNA Technologies) software. The reaction mix for real-time PCR contained 0.4 mM of each primer, 10μl of SYBR green PCR Master Mix (Applied Biosystems), 5 μl of a 1:5 dilution of the cDNA product, and DEPC water to a final volume of 20 μl. Cycling conditions were as follows: 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 seconds and 60 °C for 1 min. The reactions were carried out in fast 96-well reaction plates on the ABI 7900HT fast RT-PCR system (Applied Biosystems). Transcript levels were normalized using the actin gene as an internal standard and relative expression was calculated using the Pfaffl method[149].

6.4. Most abundant fungal transcripts at each developmental stage

C. higginsianum: In total, 34.84 million reads could be mapped to predicted *C. higginsianum* genes, with an average of 1.7, 1.3, 4.9 and 3.7 million reads for BP, PA, NP and VA samples. Out of 16,172 genes, only 1,200

had less than 5 mapped reads. As expected, several known appressoria-associated genes were detected among the most highly expressed genes in VA samples (**Supplementary Table 13**), including orthologs of *Magnaporthe oryzae ORP1, MAS1* and *ASG1* (also known as *MAS3* and *GAS1*), *Colletotrichum gloeosporioides CAS1* and *CAP22* and *Blumeria graminis GEgh16*, as well as the *C. higginsianum* melanin-related polyketide synthase (CH063_03518). Remarkably, there was little overlap between VA and PA samples, and most of the highly accumulated transcripts in PA and BP samples were not detectable, or only weakly detected, in VA samples, suggesting plant-dependent induction. In the PA and BP samples, many of the most highly expressed genes were putative effectors and secondary metabolite genes. Among the 100 most highly expressed genes in NP samples, half encoded ribosomal proteins, indicating that intense translational activity occurs at this late stage of infection (**Supplementary Table 13**).

C. graminicola: In total 47.11 million reads were mapped to predicted genes, including a total of 3.7, 10, and 33.3 million reads for PA, BP, and NP, respectively. Of the 12,006 annotated genes, only 1,194 had less than 5 mapped reads. Orthologs of *ORP1*, *MAS1*, *ASG1*, *CAS, CAP22*, *GEgh16*, and the melanin-related polyketide synthase (GLRG_04203) were all expressed in *C. graminicola* appressoria formed *in planta* (**Supplementary Table 14**). However, none of these were among the top 100 most highly expressed genes. One of the most highly expressed genes in *C. graminicola* appressoria was an ortholog of biotrophy-associated secreted protein 2 (*BAS2*), an effector of unknown function specific to intracellular biotrophic hyphae of *Magnaporthe oryzae*[150]. An ortholog of *BAS3* was among the most highly expressed genes in all samples (PA, BP, and NP) of *C. graminicola*. Primary metabolic genes dominated in the BP samples. In the *C. graminicola* necrotrophic phase, the most highly expressed genes encoded proteins involved in primary metabolism and membrane transporters. An ortholog of a *M. oryzae* cyclophilin was highly expressed in NP samples of both *C. graminicola* and *C. higginsianum*: in *M. oryzae* this gene was shown to contribute to fungal virulence[151]. As with *C. higginsianum*, transcripts encoding ribosomal proteins were over-represented in the NP samples of *C. graminicola* (**Supplementary Table 14**).

6.5. Significantly regulated genes at each developmental stage

To unravel the genetic programs associated with each infection stage of *C. higginsianum* on *Arabidopsis* leaves, and *C. graminicola* on maize leaf sheaths, pair-wise expression ratios were calculated as described below.

C. higginsianum: **Supplementary Table 15a** summarizes the number of significantly regulated genes (log2 fold-change > 2, *p*-value < 0.05) identified between the different stages analysed. The highest number of significantly regulated genes (log2 fold-change > 2, *p*-value < 0.05) was observed between the stages representing the start and end points of pathogenesis (VA vs. NP or PA vs. NP), suggesting that their highly contrasted functions are reflected by their transcriptomes. However, a large proportion of transcripts were also significantly regulated between successive developmental stages. This result indicates that fungal gene expression is extremely dynamic during plant colonization. We focused our analysis on *C. higginsianum* genes induced between the PA vs VA, BP vs PA and NP vs BP stages. Among the 100 most highly induced transcripts in PA vs. VA

(**Supplementary Table 16a**), many encode secondary metabolism-related proteins, including 10 polyketide synthases, 10 cytochrome P450s and 3 non-ribosomal peptide synthases. These transcripts are likely activated in response to an unknown plant signal and could be critical for *C. higginisianum* pathogenicity. However, the nature of these metabolites and the role they play in the infection process remains to be determined. Remarkably, the transcript showing the most dramatic increase in expression level (>17 log2 fold-change) encodes a CSEP of low molecular weight (9.5 kDa) with no homology to proteins in public databases (ChEC6[42]). The strong plant-dependent induction observed at the transcript level suggests that this protein plays a crucial role at early stages of leaf infection and might be required to establish and maintain biotrophic infection. Only three CSEPs were highly induced in PA vs. VA. In contrast, 13 CSEPs were among the 100 most highly induced transcripts comparing BP vs. PA (**Supplementary Table 16b**). This suggests that most candidate effectors are induced during and after fungal penetration in *C. higginsianum*. The transcript showing the highest induction in BP vs. PA (> 7 fold-change, log2) encodes an alcohol dehydrogenase which is involved in metabolism of sugar alcohols, suggesting that the biotrophic hyphae might still be utilizing glycerol derived from stored lipids. Unexpectedly, no amino acid, oligopeptide or sugar transporters were identified among the most highly induced genes specific to the biotrophic stage. This provides further support for the hypothesis that biotrophic hyphae utilize stored nutrient sources. In contrast, many of these transporters were specifically and highly induced in the necrotrophic stage (**Supplementary Table 16c**). In addition, numerous carbohydrate-active enzymes and secreted proteases were strongly induced at this late stage of fungal infection, including 18 transcripts encoding pectate lyases (PL1, PL3, PL9), pectin esterases (CE8) and glycosyl hydrolases (GH28, GH105). These might work in concert to catalyze the depolymerisation of pectin, an important structural cell wall component of *Arabidopsis* and other dicot plants[85].

C. graminicola: **Supplementary Table 15b** summarizes the number of significantly regulated genes (log2 fold-change > 2, *p*-value < 0.05) identified between the different stages analysed. The highest number of significantly regulated genes occurred between the extreme time-points (PA vs. NP or NP vs. PA), and changes in gene expression also occurred between successive infection stages (PA vs. BP and NP vs BP), as observed in *C. higginsianum*. However, these changes were not as pronounced in *C. graminicola* as in *C. higginsianum*, indicated by the smaller number of genes that were specific to each stage. This may reflect the relative lack of synchronization in *C. graminicola* infections when compared to *C. higginsianum* on Arabidopsis (**Fig. 1b,c**). Thus, the *C. graminicola* BP samples comprise a mixture of living host cells containing intracellular biotrophic hyphal tips, and dead or dying plant cells containing intercalary hyphae that may be undergoing different processes than those at the advancing infection front. Furthermore, the NP samples contain narrow secondary hyphae that are correlated with tissue maceration, in addition to the two types of fungal-plant interaction present in BP samples.

In *C. graminicola*, analysis of the composition of the top 100 differentially expressed genes in the BP vs PA pair-wise comparison revealed that more than half encoded secreted proteins, including 13 secreted proteases (**Supplementary Table 17a**). Membrane transporters comprised another 11% of the differentially expressed genes, in marked contrast to *C. higginsianum* where transporters were absent from the top 100 differentially expressed genes. Also in contrast to *C. higginsianum*, there were few transcripts encoding CSEPs (1) or secondary metabolism enzymes (3). In the *C. graminicola* NP vs BP comparison (**Supplementary Table 17b**), as

in *C. higginsianum*, genes encoding cell-wall degrading enzymes were well-represented, including 21 glycosyl hydrolases. Transcripts encoding genes potentially involved in secondary metabolism were particularly abundant, including four of the genes comprising secondary metabolite cluster 22. In *C. graminicola* the majority of gene clusters (**Supplementary Table 14**) appeared to be induced during necrotrophic development, suggesting that secondary metabolism is more important during that phase in *C. graminicola* than in *C. higginsianum*.

Comparisons of the top 100 differentially expressed genes in BP vs PA and NP vs BP in *C. graminicola* and *C. higginsianum* were made. In the BP vs PA comparison, four up-regulated genes were orthologous between the two species. One of them is predicted to encode a secreted serine protease of the subtilase class (CH063_01323, GLRG_05319). Another conserved gene encodes a putative beta-glucosidase C (CH063_11281, GLRG_00795) which could be involved in cellulose degradation. In the NP vs BP comparison, three up-regulated genes were orthologous between the two species. Two were predicted to encode amino acid transporters. The third encodes a putative glucose oxidase (CH063_02718, GLRG_06868). By producing H_2O_2 from glucose, such an enzyme could induce host cell death during necrotrophic growth.

6.5. Gene expression profiling during C. higginsianum *infection of* Arabidopsis *leaves*

Since the developmental synchronization within the *C. higginsianum* samples was high, and the distinction between phases of infection was exceptionally clear, we used these samples to gain further insight into the dynamics of *Colletotrichum* gene expression during plant infection. We used hierarchical clustering to generate heatmaps of gene expression profiles during different phases of disease development. REI were calculated to derive gene expression patterns (see **Section 6.2**). We analysed seven functional categories likely to be associated with fungal pathogenesis, namely genes encoding membrane transporters, carbohydrate-active enzymes (CAZymes), secreted proteases, candidate secreted effector proteins (CSEPs), cytochrome P450 enzymes, secondary metabolism-related genes and transcription factors. Genes in these different functional categories showed highly contrasted expression patterns, suggesting particular biological activities are deployed by the fungus in a stage-specific manner (**Fig. 5b**, **Supplementary Fig. 11**).

Numerous secondary metabolism-related genes share a similar expression pattern, marked by a strong induction in PA and BP (**Fig. 5b**), suggesting they might play a key role in establishment of the pathogenic interaction. Remarkably, eight secondary metabolite gene clusters showed this expression pattern, representing a total of 89 genes. Furthermore, almost half of these genes were identified among the 100 most highly induced transcripts in PA vs. VA (**Supplementary Table 16a**). The secondary metabolite gene cluster showing the most dramatic induction at the transcriptional level (cluster 10) is composed of 16 genes encoding two O-methyltransferases, three short chain dehydrogenases, two MFS transporters, a polyketide synthase, a transcription factor, a multicopper oxidase, a FAD binding domain-containing protein, a GA4 desaturase and four hypothetical proteins (**Fig. 4a**). While genomic rearrangements have affected the overall structure of the cluster, most of the genes are conserved between *C. higginsianum* cluster 10 and *C. graminicola* cluster 18. A gene encoding a ferric reductase in the *C. graminicola* cluster is associated with a transposable element in *C. higginsianum* and is located in a different genomic region. An additional gene encoding an MFS transporter (CH063_00314) was present only in

the *C. higginsianum* cluster and might complement the function of a second MFS transporter (CH063_00308) for which no transcript induction was detected in penetrating appressoria. Strikingly, this secondary metabolite gene cluster is also well-conserved in the phylogenetically distant fungus *M. oryzae*. The dramatic transcript induction detected specifically in *C. higginsianum* appressoria formed *in planta* suggests activation by an unknown plant signal. Surprisingly, expression of this cluster was not seen at any stage of *C. graminicola* infection *in planta* (**Supplementary Fig. 12**), suggesting that *Colletotrichum* species also use transcriptional regulation to produce different spectra of small molecules.

Analysis of the gene expression profiles of CSEPs revealed distinct waves of expression associated with each fungal developmental stage, and suggests that effector secretion is highly dynamic during plant colonization (**Fig. 5b**). A large proportion of CSEP transcripts preferentially accumulated during the biotrophic phase, suggesting that primary hyphae are a key interface for effector delivery to host cells. While secondary metabolites and effector proteins appear to be deployed primarily during the early stages of leaf infection (i.e. penetration and biotrophic growth), transcripts encoding secreted proteases, CAZymes and transporters share a common expression pattern, characterized by a massive accumulation at the switch to necrotrophy. This suggests that high levels of degradative activity, together with concomitant uptake of the products of this activity, occur during the necrotrophic phase (**Fig. 5b** and **Supplementary Fig. 11**). A similar up-regulation of CAZymes was evident during the necrotrophic phase of *C. graminicola* (**Supplementary Table 17c**). In contrast to the other gene categories examined, transcription factor genes showed more subtle changes in expression at each infection stage (**Fig. 5b**). It is likely that slight variations in the transcript abundance of such genes is sufficient to trigger dramatic changes in the expression of their target genes.

6.6. Validation of the RNA-Seq data

To validate the RNA-Seq approach, the expression levels of 31 transcripts encoding *C. higginsianum* transcription factors and CSEPs, and 21 transcripts encoding *C. graminicola* secondary and primary metabolic proteins and house-keeping proteins, were assessed using Reverse Transcription Polymerase Chain Reaction (RT-qPCR). Log2 transcript fold-changes (PA vs. VA, BP vs. PA and NP vs. BP) measured by RNAseq and RT-qPCR were then plotted to assess the correlation between gene expression profiles obtained using the two approaches. High linear regressions were observed in both cases: for *C. higginsianum*, $R^2 = 0.883$; and for *C. graminicola*, $R^2 = 0.867$. Moreover, in both cases the slope of the regression line was close to unity: for *C. higginsianum*, y = 0.97; and for *C. graminicola*, y = 1.08. These values indicate minimal variation in the expression data and high consistency between the two techniques.

7. Host tissue alkalinization

Local alkalinization of host tissue, resulting from the fungal secretion of ammonia at low ambient pH, has been demonstrated for several fruit-infecting *Colletotrichum* species[152-155]. Alkalinization of host tissue ensures that secreted hydrolases such as pectin-degrading enzymes are expressed under the pH conditions at which they function most effectively[152,156,157]. However, it is unclear whether leaf-inhabiting species such as *C. higginsianum*

and *C. graminicola* also induce local alkalinization of host tissues. The pH and ammonia concentrations in infected maize and *Arabidopsis* leaf tissues were measured as described in the **Online Methods**. In *Arabidopsis* leaf tissue at 5 days post inoculation with *C. higginsianum* (corresponding to the necrotrophic phase), the pH in host cells close to hyphal tips was estimated as pH 8.0, almost 3 pH units higher than in uninfected leaves (**Supplementary Fig. 13a**). The host cell cytoplamic pH decreased with increasing distance from the hyphal tips. Ammonia concentrations were nearly six-fold higher in infected *Arabidopsis* leaves at the same time-point compared to uninfected leaves (**Supplementary Fig. 13b**). However, in maize leaf tissue at an equivalent stage of *C. graminicola* infection, host cell pH adjacent to fungal hyphae was only 1.6 pH units higher than in uninfected tissue (**Supplementary Fig. 13c**). This weaker alkalization of host tissue was correlated with lower production of ammonia *in planta* by *C. graminicola* (**Supplementary Fig. 13d**). These findings suggest that the production of ammonia by both species locally increases host tissue pH to high levels around fungal hyphae at the necrotrophic stage of infection, this effect being most pronounced in the case of *C. higginsianum*.

8. Acknowledgments

This work was primarily supported by USDA-CSREES Grant 2007-35600-17829 (L.J.V., M.B.D, S.R., M.R.T.), the Max Planck Society (R.J.O.) and Deutsche Forschungsgemeinschaft (SPP1212) grant OC 104/1-3 (R.J.O.). Other funding sources included USDA-CSREES Grant 2009-34457-20125 (L.J.V.), University of Kentucky College of Agriculture Research Office (L.J.V.), Ministerio de Ciencia e Innovación (MICINN) of Spain Grants AGL2008-03177/AGR and AGL2011-29446/AGR (M.R.T.) and the Programme for Promotion of Basic and Applied Research for Innovations in Bio-oriented Industry (P.G. and K.S.). We are grateful to Britta Maegde and Bernard Dumas for critical reading of the manuscript, E. I. du Pont de Nemours and Company for releasing their *Colletotrichum* genome data to the public, Masatoki Taga for sharing unpublished cytological karyotype data and James Polashock for providing unpublished sequence data for *Colletotrichum fioriniae*. We thank Dennis Duross for artwork in Figure 1, Gary Munkvold (Iowa State University) and Laura Ruttman (Pioneer HiBred International, Johnston, IA USA) for providing images used in Figure 1c. We thank Gillian Turgeon for her helpful advice on PKS and NRPS phylogenetics. We thank Carl Simmons and Saurabh Kulshrethsa for providing unpublished data and materials that were not ultimately used in the paper. We thank Etta Nuckles, Doug Brown and Wolfgang Schmalenbach for technical support, staff of the Advanced Genetics Technology Center at the University of Kentucky including Jennifer Webb and Abbe Kesterson, staff of the Texas Agrilife Institute for Plant Genomics and Biotechnology including Charles Johnson, Richard Metz, Scott Schwartz, and Noushin Ghaffari and the Bioinformatics team at University of Kentucky including Jerzy Jaromczyk, Tommy Bullock, and Jola Jaromczyk.

9. References

1. Réblová, M., Gams, W. & Seifert, K.A. Monilochaetes and allied genera of the Glomerellales, and a reconsideration of families in the Microascales. *Stud. Mycol.* **68**, 163-191 (2011).
2. Zhang, N. *et al*. An overview of the systematics of the Sordariomycetes based on a four-gene phylogeny. *Mycologia* **98**, 1076-1087 (2006).
3. Bailey, J.A. & Jeger, M.J. (eds). *Colletotrichum - Biology, Pathology and Control* (CAB International, Wallingford, UK, 1992).
4. Hyde, K.D. *et al*. *Colletotrichum* - names in current use. *Fungal Divers.* **39**, 147-183 (2009).
5. Narusaka, Y. *et al.* RCH1, a locus in *Arabidopsis* that confers resistance to the hemibiotrophic fungal pathogen *Colletotrichum higginsianum*. *Mol. Plant Microbe Interact*. **17**, 749-762 (2004).
6. O'Connell, R. *et al*. A novel *Arabidopsis-Colletotrichum* pathosystem for the molecular dissection of plant-fungal interactions. *Mol. Plant Microbe Interact.* **17**, 272-282 (2004).
7. Ushimaru, T. *et al*. Development of an efficient gene targeting system in *Colletotrichum higginsianum* using a non-homologous end-joining mutant and *Agrobacterium tumefaciens*-mediated gene transfer. *Mol. Genet. Genomics* **284**, 357-371 (2010).
8. Bergstrom, G.C. & Nicholson, R.L. The biology of corn anthracnose: knowledge to exploit for improved management. *Plant Dis.* **83**, 596-608 (1999).
9. Crouch, J.A. & Beirn, L.A. Anthracnose of cereals and grasses. *Fungal Divers.* **39**, 19-44 (2009).
10. Frey, T.J., Weldekidan, T., Colbert, T., Wolters, P.J.C.C. & Hawk, J.A. Fitness evaluation of *Rcg1*, a locus that confers resistance to *Colletotrichum graminicola* (Ces.) G.W. Wils. using near-isogenic maize hybrids. *Crop Sci.* **51**, 1551-1563 (2011).
11. Vaillancourt, L.J. & Hanau, R.M. A method for genetic analysis of *Glomerella graminicola* (*Colletotrichum graminicola*) from maize. *Phytopathology* **81**, 530-534 (1991).
12. Perfect, S.E., Hughes, H.B., O'Connell, R.J. & Green, J.R. *Colletotrichum* - A model genus for studies on pathology and fungal-plant interactions. *Fungal Genet. Biol*. **27**, 186-198 (1999).
13. Bechinger, C. *et al.* Optical measurements of invasive forces exerted by appressoria of a plant pathogenic fungus. *Science* **285**, 1896-1899 (1999).
14. Deising, H.B., Werner, S. & Wernitz, M. The role of fungal appressoria in plant infection. *Microbes Infect.* **2**, 1631-1641 (2000).
15. Mims, C.W. & Vaillancourt, L.J. Ultrastructural characterization of infection and colonization of Maize leaves by *Colletotrichum graminicola*, and by a *C. graminicola* pathogenicity mutant. *Phytopathology* **92**,803-812 (2002).
16. Politis, DJ. The identity and perfect state of *Colletotrichum graminicola*. *Mycologia* **67**, 56-62 (1975).
17. Vaillancourt, L., Du, M., Wang, J., Rollins, J. & Hanau, R. Genetic analysis of cross fertility between two self-sterile strains of *Glomerella graminicola*. *Mycologia* **92**, 430-435 (2000).
18. Crouch, J., Clarke, B.B. & Hillman, B.I. Unraveling evolutionary relationships among the divergent lineages of *Colletotrichum* causing anthracnose disease in turfgrass and corn. *Phytopathology* **96**, 46-60 (2006).

19. Du, M., Schardl, C.L., Nuckles, E.M. & Vaillancourt, L.J. Using the mating-type sequences for improved phylogenetic resolution of *Colletotrichum* species complexes. *Mycologia* **97**, 641-658 (2005).
20. Rodríguez-Guerra, R. *et al*. Heterothallic mating observed between Mexican isolates of *Glomerella lindemuthiana*. *Mycologia* **97**, 793-803 (2005).
21. García-Serrano, M., Alfaro Laguna, E., Rodriguez-Guerra, R. & Simpson, J. Analysis of the *MAT1-2-1* gene of *Colletotrichum lindemuthianum*. *Mycoscience* **49**, 312-317 (2008).
22. Damm, U., Mostert, L., Crous, P.W. & Fourie, P.H.. Novel *Phaeoacremonium* species associated with necrotic wood of Prunus trees. *Persoonia* **20**, 87-102 (2008).
23. Gardes, M. & Bruns, T.D. ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizae and rusts. *Mol. Ecol.* **2**, 113-118 (1993).
24. White, T.J., Bruns, T., Lee. S. & Taylor, J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR Protocols: a guide to methods and applications* (eds Innis, M.A., Gelfand, D.H., Sninsky, J.J. & White, T.J.) 315-322 (Academic Press, San Diego, 1990).
25. Carbone, I. & Kohn, L.M. A method for designing primer sets for speciation studies in filamentous ascomycetes. *Mycologia* **91**, 553-556 (1999).
26. Woudenberg, J.H.C., Aveskamp, M.M., de Gruyter, J., Spiers, A.G. & Crous, P.W. Multiple *Didymella* teleomorphs are linked to the *Phoma clematidina* morphotype. *Persoonia* **22**, 56-62 (2009).
27. Crous, P.W., Groenewald, J.Z., Risede, J.M. & Hywel-Jones, N.L. *Calonectria* species and their *Cylindrocladium* anamorphs: species with sphaeropedunculate vesicles. *Stud. Mycol.* **50**, 415-430 (2004).
28. Damm, U., Woudenberg, J.H.C., Cannon, P.F. & Crous, P.W. *Colletotrichum* species with curved conidia from herbaceous hosts. *Fungal Divers.* **39**, 45-87 (2009).
29. Rambaut A (2002). Sequence Alignment Editor. Version 2.0. Department of Zoology. University of Oxford, Oxford.
30. Mason-Gamer, R.J. & Kellogg, E.A. Chloroplast DNA analysis of the monogenomic *Triticeae*: phylogenetic implications and genome-specific markers. In *Methods of Genome Analysis in Plants* (ed. Jauhar, P.P.) 301–325 (CRC Press, Boca Raton, 1996).
31. Nylander, J.A.A. *MrModeltest v.2. Program distributed by the author* (Evolutionary Biology Centre, Uppsala University, 2004).
32. Swofford, D.L. (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts, USA.
33. Hillis, D.M. & Bull, J.J. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* **42**, 182-192 (1993).
34. Latunde-Dada, A.O. & Lucas, J.A. Localized hemibiotrophy in *Colletotrichum*: cytological and molecular taxonomic similarities among *C. destructivum*, *C. linicola* and *C. truncatum*. *Plant Pathol.* **56**, 437-447 (2007).
35. Armstrong-Cho, C.L. & Banniza, S. *Glomerella truncata* sp. nov., the teleomorph of *Colletotrichum truncatum*. *Mycol. Res.* **110**, 951-956 (2006).

36. Forgey, W.M., Blanco, M.H. & Loegering, W.Q. Differences in pathological capabilities and host specificity of *Colletotrichum graminicola* on *Zea mays* (maize). *Plant Dis. Rep.* **62**, 573-576 (1978).
37. Huser, A., Takahara, H., Schmalenbach, W. & O'Connell, R. Discovery of pathogenicity genes in the crucifer anthracnose fungus *Colletotrichum higginsianum*, using random insertional mutagenesis. *Mol. Plant Microbe Interact.* **22**, 143-156 (2009).
38. Kleemann, J., Takahara, H., Stüber, K. & O'Connell, R. Identification of soluble secreted proteins from appressoria of *Colletotrichum higginsianum* by analysis of expressed sequence tags. *Microbiology* **154**, 1204-1217 (2008).
39. Zerbino, D.R. & Birney, E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821-829 (2008).
40. Takahara, H., Dolf, A., Endl, E. & O'Connell, R. Flow cytometric purification of *Colletotrichum higginsianum* biotrophic hyphae from *Arabidopsis* leaves for stage-specific transcriptome analysis. *Plant J.* **59**, 672-683 (2009).
41. Spanu, P.D. *et al*. Genome expansion and gene loss in powdery mildew fungi reveal functional trade-offs in extreme parasitism. *Science* **330**, 1543-1546 (2010).
42. Kleemann, J. *et al*. Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. *PLoS Pathogens* **8**, e1002643 (2012).
43. Parra, G., Blanco, E. & Guigó, R. GeneID in *Drosophila*. *Genome Res.* **10**, 511-515 (2000).
44. Borodovsky, M. & McIninch, J. GeneMark: parallel gene recognition for both DNA strands. *Comput. Chem.* **17**, 123-133 (1993).
45. Korf I. Gene finding in novel Genomes. *BMC Bioinformatics* **5**, 59 (2004).
46. Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32**, W309-W312 (2004).
47. Haas, B.J., Zeng, Q., Pearson, M.D., Cuomo, C.A. & Wortman, J.R. Approaches to fungal genome annotation, *Mycology* **2**, 118-141 (2011).
48. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061-1067 (2007).
49. Kemen, E. & Jones, J.D.G. Obligate biotroph parasitism: can we link genomes to lifestyles? *Trends in Plant Science*, DOI: 10.1016/j.tplants.2012.04.005 (2012).
50. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111 (2009).
51. Grabherr, M.G. et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol*. **29**, 644-652 (2011).
52. Thon, M.R., Nuckles, E.M. & Vaillancourt, L.J. Restriction enzyme-mediated integration used to produce pathogenicity mutants of *Colletotrichum graminicola*. *Mol. Plant Microbe Interact.* **13**, 1356-1365 (2000).
53. Dimalanta, E.T. *et al*. A microfluidic system for large DNA molecule arrays. *Anal. Chem.* **76**, 5293-5301 (2004).

54. Zhou, S., Herschleb, J. & Schwartz, D.C. A single molecule system for whole genome analysis. In *New Methods for DNA Sequencing* (ed. Mitchelson, K.R.) 269-304 (Elsevier, Amsterdam, 2007).
55. Zhou, S. *et al*. Validation of rice genome sequence by optical mapping. *BMC Genomics* **8**, 278 (2007).
56. Zhou, S. *et al*. Whole-genome shotgun optical mapping of *Rhodobacter sphaeroides* strain 2.4.1 and its use for whole-genome shotgun sequence assembly. *Genome Res.* **13**, 2142-2151 (2003).
57. Zhou, S. *et al*. Shotgun optical mapping of the entire *Leishmania major* Friedlin genome. *Mol. Biochem. Parasitol.* **138**, 97-106 (2004).
58. Lim, A. *et al*. Shotgun optical maps of the whole *Escherichia coli* O157:H7 genome. *Genome Res.* **11**, 1584-1593 (2001).
59. Lin, J. *et al*. Whole-genome shotgun optical mapping of *Deinococcus radiodurans*. *Science* **285**, 558-562 (1999).
60. Zhou, S. *et al*. A single molecule scaffold for the maize genome. *PLoS Genet.* **5**, e1000711 (2009).
61. Valouev, A., Schwartz, D.C., Zhou, S.G. & Waterman, M.S. An algorithm for assembly of ordered restriction maps from single DNA molecules. *Proc. Natl. Acad. Sci. USA* **103**, 15770-15775 (2006).
62. Teague, B. *et al*. High-resolution genome structure by single molecule analysis. *Proc. Natl. Acad. Sci. USA* **107**, 10848-10853 (2010).
63. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* **25**,1754-1760 (2009).
64. Valouev, A. *et al*. Alignment of optical maps. *J. Comput. Biol.* **13**, 442-462 (2006).
65. Huang, X. & Miller, W. A time-efficient, linear space local similarity algorithm. *Adv. Appl. Math.* **12**, 337-357 (1991).
66. Waterman, M. & Vingron, M. Sequence comparison significance and poisson approximation. *Stat. Sci.* **9**, 401-418 (1994).
67. Kurtz, S. *et al*. Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
68. Rollins, J.A. *The characterization and inheritance of chromosomal variation in Glomerella graminicola.* Ph.D. thesis, Purdue University, West Lafayette, Indiana, USA (1996).
69. Dean, R.A. *et al*. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**, 980-986 (2005).
70. Ma, L.-J. *et al*. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**, 367-373 (2010).
71. Rouxel, T. *et al*. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nat Commun.* **2**, 202 (2011).
72. Thon, M.R. *et al*. The role of transposable element clusters in genome evolution and loss of synteny in the rice blast fungus *Magnaporthe oryzae*. *Genome Biol.* **7**, R16 (2006).
73. Aguileta, G., Refrégier, G., Yockteng, R., Fournier, E. & Giraud, T. Rapidly evolving genes in pathogens: Methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. *Infect. Genet. Evol.* **9**, 656-670 (2009).

74. Enright, A.J., Van Dongen, S. & Ouzounis, C.A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575-1584 (2002).
75. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059-3066 (2002).
76. Drummond, A.J. *et al*. Geneious v5.5. (2011). At http://www.geneious.com
77. Tang, W.H., Coughlan, S., Crane, E., Beatty, M., & Duvick, J. The application of laser microdissection to in planta gene expression profiling of the maize anthracnose stalk rot fungus *Colletotrichum graminicola. Mol. Plant Microbe Interact.* **19**,1240-1250 (2006).
78. Bendtsen, J.D., Nielsen, H., von Heijne, G. & Brunak, S. Improved Prediction of Signal Peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783-795 (2004).
79. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E.L.. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567-580 (2001).
80. Horton, P. *et al*. WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* **35**, W585-W587 (2007).
81. de Groot, P.W.J., Hellingwerf, K.J. & Klis, F.M. Genome-wide identification of fungal GPI proteins. *Yeast* **20**, 781-796 (2003).
82. Jones, J.D. & Dangl, J.L. The plant immune system. *Nature* **444**, 323-329 (2006).
83. Yoshida, K. *et al*. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *Plant Cell* **21**, 1573-1591 (2009).
84. Cantarel, B.L. *et al*. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37**, D233-D238 (2009).
85. Cosgrove, D.J. Assembly and enlargement of the primary cell wall in plants. *Ann. Rev. Cell Dev. Biol.* **13**, 171-201 (1997).
86. Vogel, J. Unique aspects of the grass cell wall. *Curr. Opin. Plant Biol.* **11**, 301-307 (2008).
87. Quinlan, R.J. *et al*. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc. Natl Acad. Sci. U S A.* **108**, 15079-15084 (2011).
88. Beeson, W.T., Phillips, C.M., Cate, J.H. & Marletta, M.A. Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases. *J. Am. Chem. Soc.* **134**, 890-892 (2012).
89. Dimarogona, M., Topakas, E., Olsson, L. & Christakopoulos, P. Lignin boosts the cellulase performance of a GH-61 enzyme from *Sporotrichum thermophile*. Bioresource Technol. **110**, 480-487 (2012).
90. El Gueddari, N.E., Rauchhaus, U., Moerschbacher, B.M. & Deising, H.B. Developmentally regulated conversion of surface-exposed chitin to chitosan in cell walls of plant pathogenic fungi. *New Phytol.* **156**, 103-112 (2002).
91. de Jonge, R., & Thomma, B.P.H.J. Fungal LysM effectors: extinguishers of host immunity? *Trends Microbiol.* **17**, 151-157 (2009).
92. de Jonge, R. *et al*. Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. *Science* **329**, 953-955 (2010).

93. Pain, N.A., O'Connell, R.J., Mendgen, K. & Green, J.R. Identification of glycoproteins specific to biotrophic intracellular hyphae formed in the *Colletotrichum lindemuthianum*-bean interaction. *New Phytol.* **127**, 233-242 (1994).
94. Perfect, S.E., O'Connell, R.J., Green, E.F., Doering-Saad, C., & Green, J.R. Expression cloning of a fungal proline-rich glycoprotein specific to the biotrophic interface formed in the *Colletotrichum*-bean interaction. *Plant J.* **15**, 273-279 (1998).
95. de Groot, P.W.J., Ram, A.F. & Klis, F.M. Features and functions of covalently linked proteins in fungal cell walls. *Fungal Genet. Biol.* **42**, 657-675 (2005).
96. Iriti, M. & Faoro, F. Chitosan as a MAMP, searching for a PRR. *Plant Signal Behav.* **4**, 66-68 (2009).
97. Kamakura, T., Yamaguchi, S., Saitoh, K., Teraoka, T. & Yamaguchi, I. A novel gene, CBP1, encoding a putative extracellular chitin-binding protein, may play an important role in the hydrophobic surface sensing of *Magnaporthe grisea* during appressorium differentiation. *Mol. Plant Microbe Interact.* **15**, 437-444 (2002).
98. Rawlings, N.D., Barrett, A.J. & Bateman, A. MEROPS: the peptidase database. *Nucleic Acids Res.* **38**, D227-D233 (2010).
99. Dunasevsky, Y.E., Matveeva, A.R., Beliakova, G.A., Domash, V.I. & Belozersky, M.A. Extracellular alkaline proteinase of *Colletotrichum gloeosporioides*. *Biochemistry (Mosc).* **72**, 345-350 (2007).
100. Dow, J.M., Davies, H.A. & Daniels, M.J. A metalloprotease from *Xanthomonas campestris* that specifically degrades proline/hydroxyproline-rich glycoproteins of the plant extracellular matrix. *Mol. Plant Microbe Interact.* **11**, 1085-1093 (1998).
101. Lingner, U., Münch, S., Deising, H.B. & Sauer, N. Hexose transporters of a hemibiotrophic plant pathogen: functional variations and regulatory differences at different stages of infection. *J. Biol. Chem.* **286**, 20913-20922 (2011).
102. Lingner, U., Münch, S., Sode, B., Deising, H.B. & Sauer, N. Functional characterization of a eukaryotic melibiose transporter. *Plant Physiol.* **156**, 1565-1576 (2011).
103. Ruiz, A. & Ariño, J. Function and regulation of the *Saccharomyces cerevisiae* ENA sodium ATPase system. *Eukaryot. Cell* **6**, 2175-2183 (2007).
104. Keller, N.P., Turner, G. & Bennett, J.W. Fungal secondary metabolism – from biochemistry to genomics. *Nat. Rev. Microbiol.* **3**, 937-947 (2005).
105. Collemare, J., Billard, A., Bohnert, H.U., & Lebrun, M.H. Biosynthesis of secondary metabolites in the rice blast fungus *Magnaporthe grisea*: the role of hybrid PKS-NRPS in pathogenicity. *Mycol. Res.* **112**, 207-215 (2008).
106. Khaldi, N. *et al*. SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genet. Biol.* **47**, 736-741 (2010).
107. Garcia-Pajón, C.M. & Collado, I.G. Secondary metabolites isolated from *Colletotrichum* species. *Nat. Prod. Rep.* **20**, 426-431 (2003).
108. Singh, J., Quereshi, S., Banerjee, N. & Pandey, A.K. Production and extraction of phytotoxins from *Colletotrichum dematium* FGCC# 20 effective against *Parthenium hysterophorus* L. *Braz. Arch. Biol. Technol.* **53**, 669-678 (2010).

109. Kubo, Y. & Furusawa, I. Melanin biosynthesis: Prerequisite for successful invasion of the plant host by appressoria of *Colletotrichum* and *Pyricularia*. In *The Fungal Spore and Disease Initiation in Plants and Animals* (eds Cole, G.T. & Hoch, H. C.) 205-217 (New York, Plenum Publishing, 1991).

110. Wicklow, D.T., Jordan, A.M. & Gloer, J.B. Antifungal metabolites (monorden, monocillins I, II, III) from *Colletotrichum graminicola*, a systemic vascular pathogen of maize. *Mycol. Res.* **113**, 1433-1442 (2009).

111. Leite, B. & Nicholson, R.L. Mycosporine-alanine: a self-inhibitor of germination from the conidial mucilage of *Colletotrichum graminicola*. *Exp. Mycol.* **16**, 76-86 (1992).

112. Horbach, R. *et al*. Sfp-type 4'-phosphopantetheinyl transferase is indispensable for fungal pathogenicity. *Plant Cell* **21**, 3379-3396 (2009).

112. Dereeper, A. *et al*. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465-W469 (2008).

113. Linnemannstöns, P., Prado, M., Fernandez-Martin, R., Tudzynski, B. & Avalos, J. A carotenoid biosynthesis gene cluster in *Fusarium fujikuroi*: the genes *carB* and *carRA*. *Mol. Genet. Genomics* **267**, 593-602 (2002).

114. Kroken, S., Glass, N.L., Taylor, J.W., Yoder, O.C., & Turgeon, B.G. Phylogenomic analysis of type 1 polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proc. Natl. Acad. Sci. USA* **100**, 15670-15675 (2003).

115. Bushley, K.E. & Turgeon, B.G. Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. *BMC Evol. Biol.* **10**, 26 (2010).

116. Amselem, J. *et al*. Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. PLoS Genet. **7**, e1002230 (2011).

117. Schmidt, S.M. & Panstruga, R. Pathogenomics of fungal plant parasites: what have we learnt about pathogenesis? *Curr. Opin. Plant Biol.* **14**, 392-399 (2011).

118. Cresnar, B. & Petric, S. Cytochrome P450 enzymes in the fungal kingdom. *Biochim. Biophys. Acta* **1814**, 29-35 (2011).

119. Tsuji, G. *et al*. Novel fungal transcriptional activators, Cmr1p of *Colletotrichum lagenarium* and Pig1p of *Magnaporthe grisea*, contain Cys2His2 zinc finger and Zn(II)2Cys6 binuclear cluster DNA binding motifs, and regulate transcription of melanin biosynthesis genes in a development specific manner. *Mol. Microbiol.* **38**, 940-954 (2000).

120. Tsuji, G. *et al*. Evidence for involvement of two naphtholreductases in the first reduction step of melanin biosynthesis pathway of *Colletotrichum lagenarium*. *Mycol. Res.* **107**, 854-860 (2003).

121. Fujii, I. *et al*. Enzymatic synthesis of 1,3,6,8-tetrahydroxynaphthalene solely from malonyl coenzyme A by a fungal interactive typeI polyketide synthase PKS1. *Biochemistry* **39**, 8853-8858 (2000).

122. Edens, W. A., Goins, T. Q., Dooley, D. & Henson, J. M. Purification and characterization of a secreted laccase of *Gaeumannomyces graminis* var. *tritici*. *Appl. Environ. Microbiol.* **65**, 3071-3074 (1999).

123. Kubo, Y. *et al*. Cloning and structural analysis of the melanin biosynthesis gene *SCD1* encoding scytalone dehydratase in *Colletotrichum lagenarium*. *Appl. Environment. Microbiol.* **62**, 4340-4344 (1996).

124. Perpetua, N. S., Kubo, Y., Yasuda, N., Takano, Y. & Furusawa, I. Cloning and characterization of a melanin biosynthetic THR1 reductase gene essential for appressorial penetration of *Colletotrichum lagenarium*. *Mol. Plant Microbe Interact*. **9**, 323-329 (1996).
125. Tsai, H.F., Wheeler, M.H., Chang, Y.C. & Kwon-Chung, K.J. A developmentally regulated gene cluster involved in conidial pigment biosynthesis in *Aspergillus fumigatus*. *J. Bacteriol.* **181**, 6469-6477 (1999).
126. Woo, P.C. *et al*. High diversity of polyketide synthase genes and the melanin biosynthesis gene cluster in *Penicillium marneffei*. *FEBS J.* **277**, 3750-3758 (2010).
127. Fujii, I., Yoshida, N., Shimomaki, S., Oikawa, H. & Ebizuka, Y. An iterative type I polyketide synthase PKSN catalyzes synthesis of the decaketide Alternapyrone with regio-specific octa-methylation. *Chem. Biol.* **12**, 1301-1309 (2005).
128. Böhnert, H.U. *et al*. A putative polyketide synthase/peptide synthetase from *Magnaporthe grisea* signals pathogen attack to resistant rice. *Plant Cell* **16**, 2499-2513 (2004).
129. Khaldi, N., Collemare, J., Lebrun, M.-H. & Wolfe, K.H. Evidence for horizontal transfer of a secondary metabolite gene cluster between fungi. *Genome Biol.* **9**, R18 (2008).
130. Reeves, C.D., Hu, Z., Reid, R. & Kealey, J.T. Genes for biosynthesis of the fungal polyketides hypothemycin from *Hypomyces subiculosus* and radicicol from *Pochonia chlamydosporia*. *Appl. Environ. Microbiol.* **74**, 5121-5129 (2008).
131. Wang, S. *et al*. Functional characterization of the biosynthesis of radicicol, an Hsp90 inhibitor resorcyclic acid lactone from *Chaetomium chiversii*. *Chem. Biol.* **15**, 1328-1338 (2008).
132. Pieterse, C.M.J., Leon-Reyes, A., Van der Ent, S. & Van Wees, S.C.M. Networking by small-molecule hormones in plant immunity. *Nat. Chem. Biol.* **5**, 308-316 (2009).
133. Maor, R. & Shirasu, K. The arms race continues: battle strategies between plants and fungal pathogens. *Curr. Opin. Microbiol.* **8**, 399-404 (2005).
134. Robinson, M., Riov, J. & Sharon, A. Indole-3-acetic acid biosynthesis in *Colletotrichum gloeosporioides* f. sp. *aeschynomene*. *Appl. Environ. Microbiol.* **64**, 5030-5032 (1998).
135. Maor, R., Haskin, S., Levi-Kedmi, H. & Sharon, A. *In planta* production of indole-3-acetic acid by *Colletotrichum gloeosporioides* f. sp. *aeschynomene*. *Appl. Environ. Microbiol.* **70**, 1852-1854 (2004).
136. Daundasekera, W.A.M., Joyce, D.C., Adikaram, N.K.B. & Terry, L.A. Pathogen-produced ethylene and the *Colletotrichum musae–banana* fruit pathosystem. *Austral. Plant Pathol.* **37**, 448-453 (2008).
137. Chen, N., Goodwin, P.H. & Hsiang, T. The role of ethylene during the infection of *Nicotiana tabacum* by *Colletotrichum destructivum*. *J.Exp. Bot.* **54**, 2449 -2456 (2003).
138. Flaishman, M.A. & Kolattukudy, P.E. Timing of fungal invasion using host's ripening hormone as a signal. *Proc. Natl. Acad. Sci. USA* **91**, 6579-6583 (1994).
139. Park, J. *et al*. FTFD: an informatics pipeline supporting phylogenomic analysis of fungal transcription factors. *Bioinformatics* **24**, 1024-1025 (2008).
140. Kankanala, P., Czymmek, K. & Valent, B. Roles for rice membrane dynamics and plasmodesmata during biotrophic invasion by the blast fungus. *Plant Cell* **19**, 706–724 (2007).

141. Metz, R.P., Kwak, H.-I., Gustafson, T., Laffin, B. & Porter, W.W. Differential transcriptional regulation by mouse single-minded 2s. *J. Biol. Chem.* **281**, 10839-10848 (2006).
142. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
143. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. EdgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
144. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
145. Sturn, A., Quackenbush, J. & Trajanoski, Z. Genesis: cluster analysis of microarray data. *Bioinformatics* **18**, 207-208 (2002).
146. Duplessis, S. *et al*. *Melampsora larici-populina* transcript profiling during germination and timecourse infection of poplar leaves reveals dynamic expression patterns associated with virulence and biotrophy. *Mol. Plant Microbe Interact.* **24**, 808-818 (2011).
147. Vandesompele, J. *et al*. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, 0034.1 (2002).
148. Hacquard, S. *et al*. Validation of *Melampsora larici-populina* reference genes for in planta RT-quantitative PCR expression profiling during time-course infection of poplar leaves. *Physiol. Mol. Plant Pathol.* **75**, 106-112 (2011).
149. Pfaffl, M.W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45 (2001).
150. Mosquera, G, Giraldo, M.C., Khang, C.H., Coughlan, S. & Valent, B. Interaction transcriptome analysis identifies *Magnaporthe oryzae* BAS1-4 as biotrophy-associated secreted proteins in rice blast disease. *Plant Cell* **21**, 1273-1290 (2009).
151. Viaud, M.C., Balhadère, P.V. & Talbot, N.J. A *Magnaporthe grisea* cyclophilin acts as a virulence determinant during plant infection. *Plant Cell* **14**, 917-930 (2002).
152. Drori, N. *et al*. A combination of external pH and nitrogen assimilation affects secretion of the virulence factor pectate lyase by *Colletotrichum gloeosporioides. Appl. Environ. Microbiol.* **69**, 3258-3262 (2003).
153. Prusky, D. & Yakoby, N. Pathogenic fungi: leading or led by ambient pH? *Mol. Plant Pathol.* **4**, 509-516 (2003).
154. Alkan, N., Fluhr, R., Sherman, A. & Prusky, D. Role of ammonia secretion and pH modulation on pathogenicity of *Colletotrichum coccodes* on tomato fruit. *Mol. Plant Microbe Interact.* **21**, 1058-1066 (2008).
155. Diéguez-Uribeondo, J., Förster, H. & Adaskaveg, J.E. Visualization of localized pathogen-induced pH modulation in almond tissues infected by *Colletotrichum acutatum* using confocal scanning laser microscopy. *Phytopathology* **98**, 1171-1178 (2008).
156. Yakoby, N., Kobiler, I., Dinoor, A. & Prusky, D. pH regulation of pectate lyase secretion modulates the attack of *Colletotrichum gloeosporioides* on avocado fruits. *Appl. Environ. Microbiol.* **66**, 1026-1030 (2000).
157. Kramer-Haimovich, H. *et al*. Effect of ammonia production by *Colletotrichum gloeosporioides* on *pelB* activation, pectate lyase secretion, and fruit pathogenicity. *Appl. Environ. Microbiol.* **72**, 1034-1039 (2006).

158. Czymmek, K.J., Bourett, T.M., Howard, R.J. Fluorescent probes in fungi. In: *Microbial Imaging: Methods in Microbiology*. (eds. Savidge, T., Pothoulakis, C.) **34,** 27-62 (Microbial Imaging Elsevier, Amsterdam, the Netherlands, 2005).

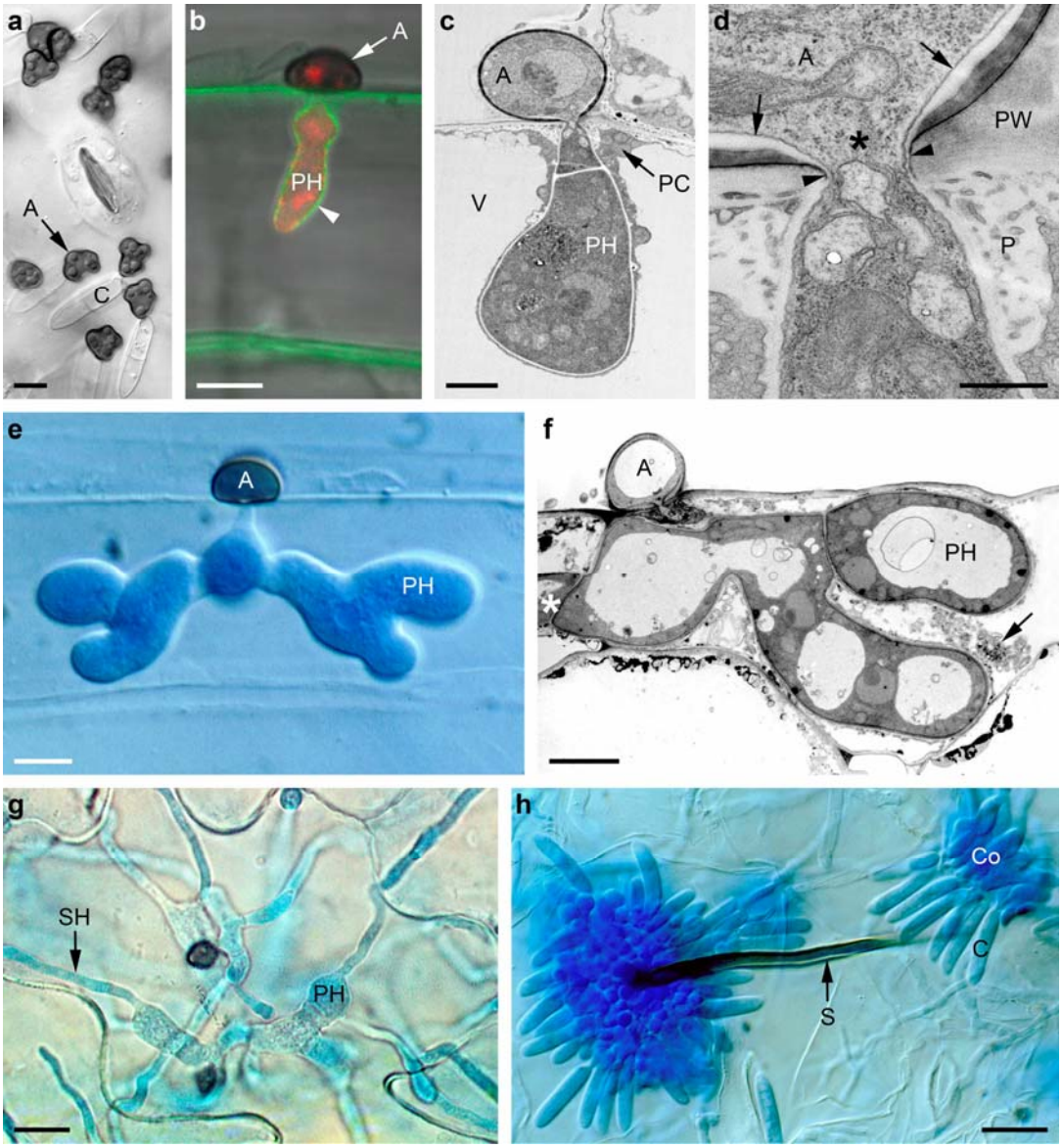
159. Lücking, R., Huhndorf, S., Pfister, D.H., Plata, E.R. & Lumbsch, H.T. Fungi evolved right on track. *Mycologia* **101**, 810-822 (2009).

160. Alkan, N., Davydov, O., Sagi, M., Fluhr, R. & Prusky, D. Ammonium secretion by *Colletotrichum coccodes* activates host NADPH oxidase activity enhancing host cell death and fungal virulence in tomato fruits. *Mol. Plant Microbe Interact.* **22**, 1484–1491 (2009).

Nature Genetics: doi:10.1038/ng.2372

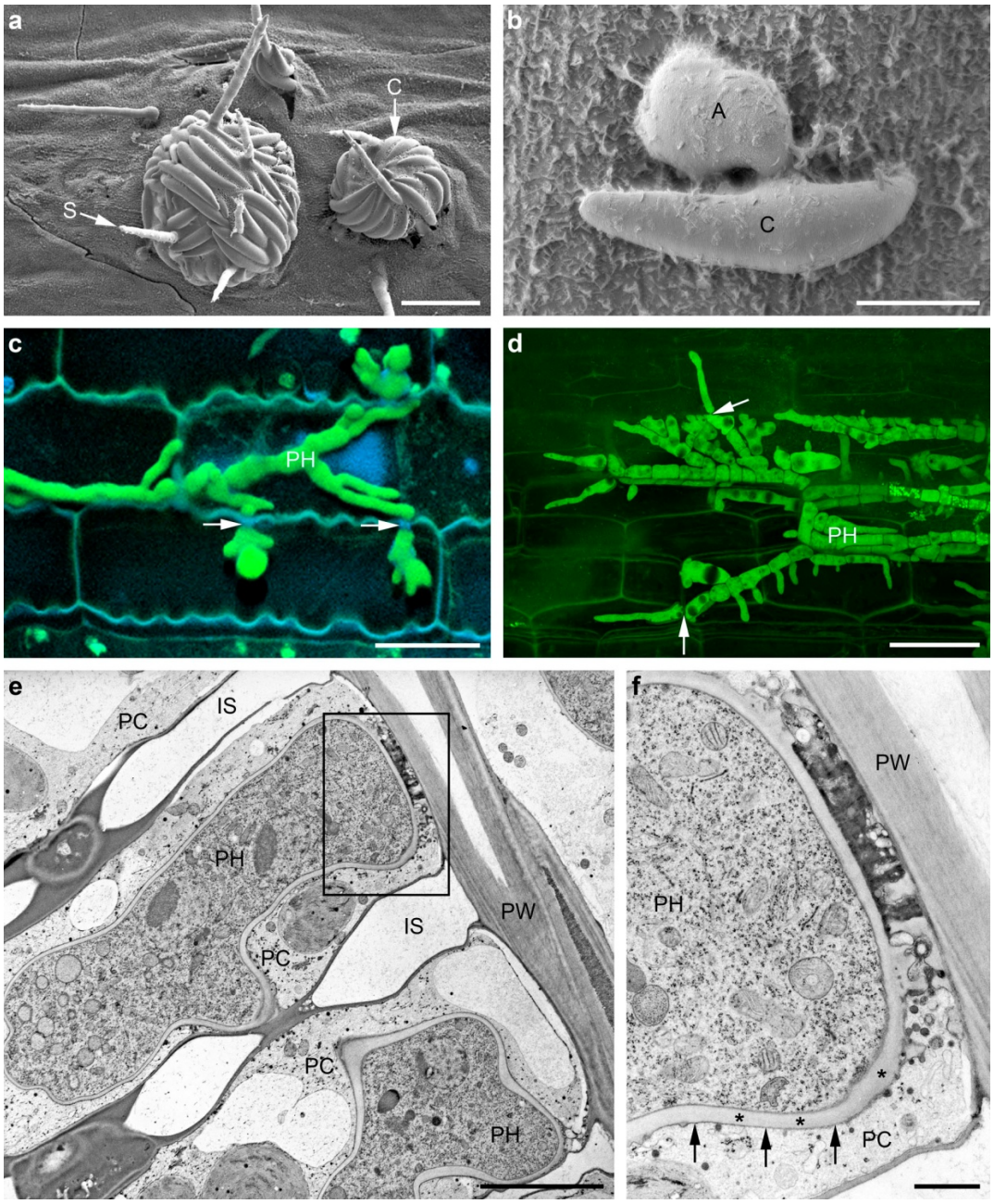
11. Supplementary Figures and Tables

Supplementary Figure 1. Infection process of *C. higginsianum* on *Arabidopsis*.



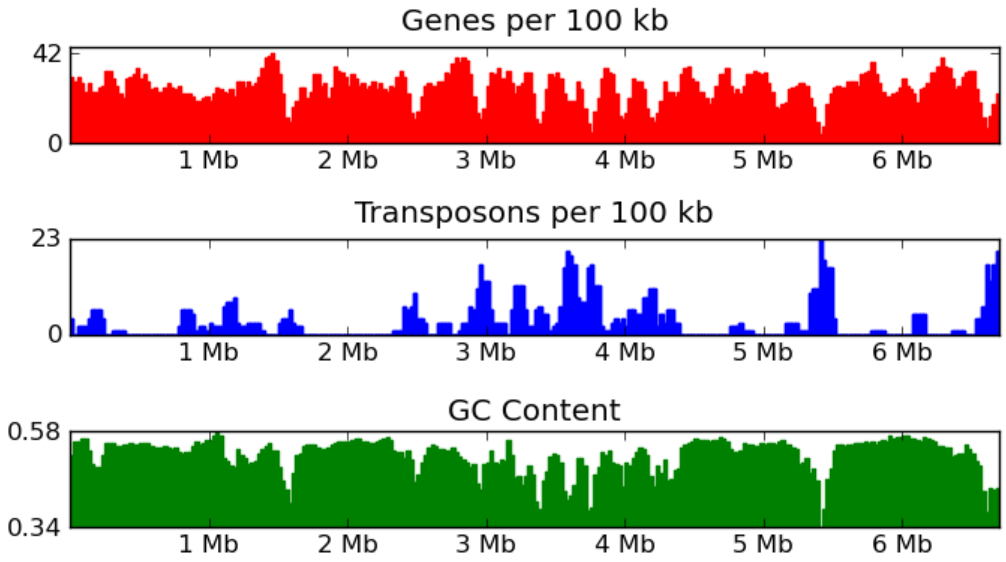
Germination: (**a**) Conidia (C) germinating on the leaf surface to form melanised appressoria (A). Bar = 5 µm. **Biotrophic phase:** (**b**) Confocal micrograph showing epidermal cell penetrated by young primary hypha (PH) surrounded by an intact host plasma membrane (arrow head) labelled with PEN1-GFP. Bar = 5 µm. (**c**) TEM cross-section through young primary hypha surrounded by intact plant cytoplasm (PC). V, vacuole. Bar = 2 µm. (**d**) Enlargement from (**c**) showing appressorial penetration pore (asterisk). The inner wall layer of the appressorium (arrows) is continuous with the pentration hypha cell wall. Note downward deformation of plant cuticle (arrowheads) and papilla (P) deposited by host cell at fungal entry site. PW = plant cell wall. Bar = 500 nm. (**e**) Mature, fully expanded primary hypha inside epidermal cell. Bar = 5 µm. **Necrotrophic phase:** (**f**) TEM section through vacuolated primary hyphae producing necrotrophic secondary hypha (asterisk). Cytoplasm of the infected cell is disrupted (arrow). Bar = 5 µm. (**g**) Thin, filamentous secondary hyphae (SH) emerging from bulbous primary hyphae. Bar = 10 µm. **Sporulation:** (**h**) Acervuli erupting from surface of dead tissue comprise melanised setae (S), conidiophores (Co) and asexual spores (conidia, C). Bar = 10 µm. (**e**) and (**h**) are reprinted from O'Connell et al. (2004)[6] *Mol. Plant Microbe Interact.* **17**, 272-282 with permission of the American Phytopathological Society.

Supplementary Figure 2. Infection process of *C. graminicola* on maize.

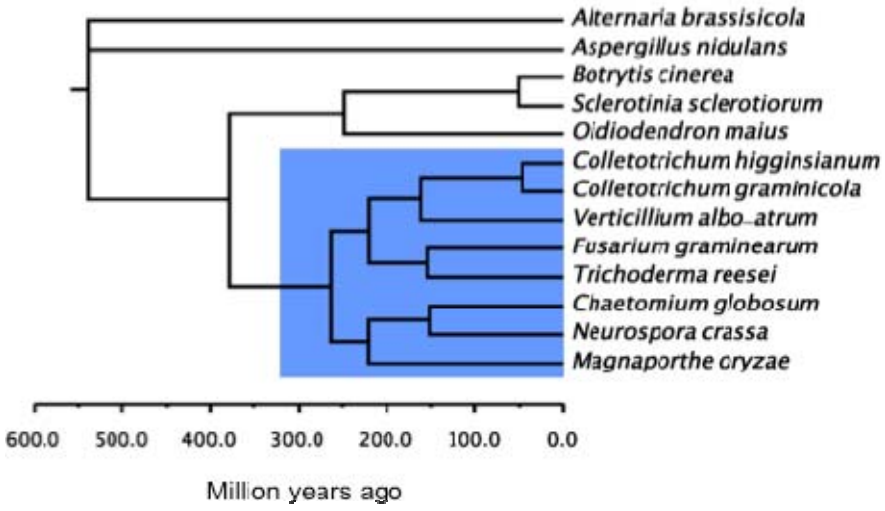


Sporulation: (**a**) Cryo-SEM image showing hair-like setae (S) and conidial mass (C) comprising acervuli erupting through the leaf surface. Bar = 25 µm. **Germination:** (**b**) Cryo-SEM image showing conidium (C) germinating to form an appressorium (A) on the leaf surface. Epicuticular wax crystals adhere to the fungal cells. Bar = 10 µm. **Biotrophic phase:** (**c**) Confocal image showing primary hyphae (PH) colonizing leaf epidermal cells. Arrows indicate hyphal constrictions at sites of cell-to-cell penetration. Hyphae expressing AmCyan were imaged by single photon excitation at 458 nm and cell wall autofluorescence by multiphoton excitation at 730 nm. Bar = 50 µm. (**d**) Confocal micrograph showing primary hyphae at the margin of an expanding lesion within leaf sheath. Hyphae expressing AmCyan were imaged as in (**c**) Bar = 50 µm. Arrows indicate sites of cell-to-cell penetration. (**e**,**f**) TEM images showing sheath cells infected by intracellular primary hyphae, 48 h after inoculation. (**f**) Enlargement showing hypha enveloped by host plasma membrane (arrows), making close contact with the fungal cell wall (asterisks). Prepared by high pressure freezing and freeze substitution. Bars = 5 µm (**e**), 1 µm (**f**). IS, intercellular space; PW, plant wall; PC, plant cytoplasm. Images provided by Keith Duncan (**a**,**b**), Timothy Bourett (**d**) and Nancy Rizzo (**e**,**f**), courtesy of Pioneer Hi-Bred Intl, DuPont Agricultural Biotechnology, Wilmington, DE USA. (**c**) Reprinted from Czymmek *et al*. (2005)[158] *Methods Microbiol.* **34**, 27-62 with permission from Elsevier.

Supplementary Figure 3. Distribution of genes, transposable elements and GC content on *C. graminicola* chromosome 1.



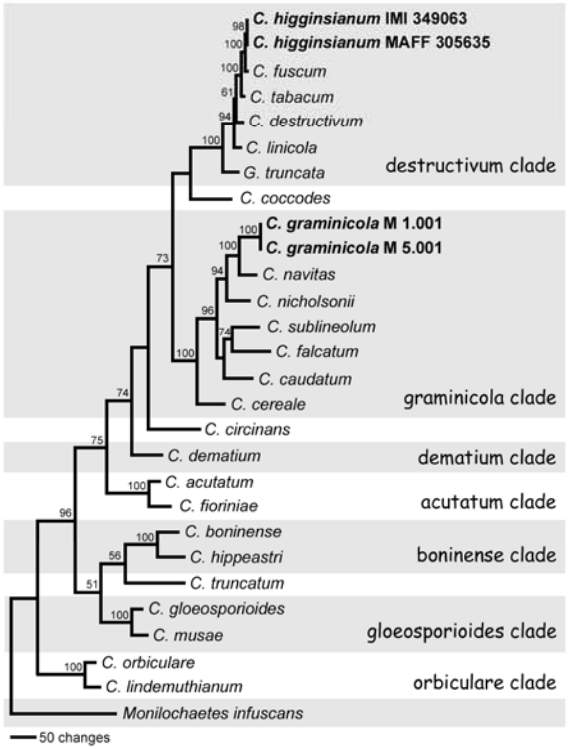
Supplementary Figure 4. *Colletotrichum* evolutionary divergence date estimation.



The phylogenetic tree was constructed from protein sequences of 64 MCL (Markov clustering) clusters containing only one gene from each species. To calibrate the tree, we used the date estimate of 290-380 million years ago (mya) for divergence of the Sordariomycetes (blue square)[159]. On this basis we obtained an age estimate of 46.8 mya for the most recent common ancestor of the two *Colletotrichum* species.

Supplementary Figure 5. Phylogeny of *Colletotrichum* species.

a

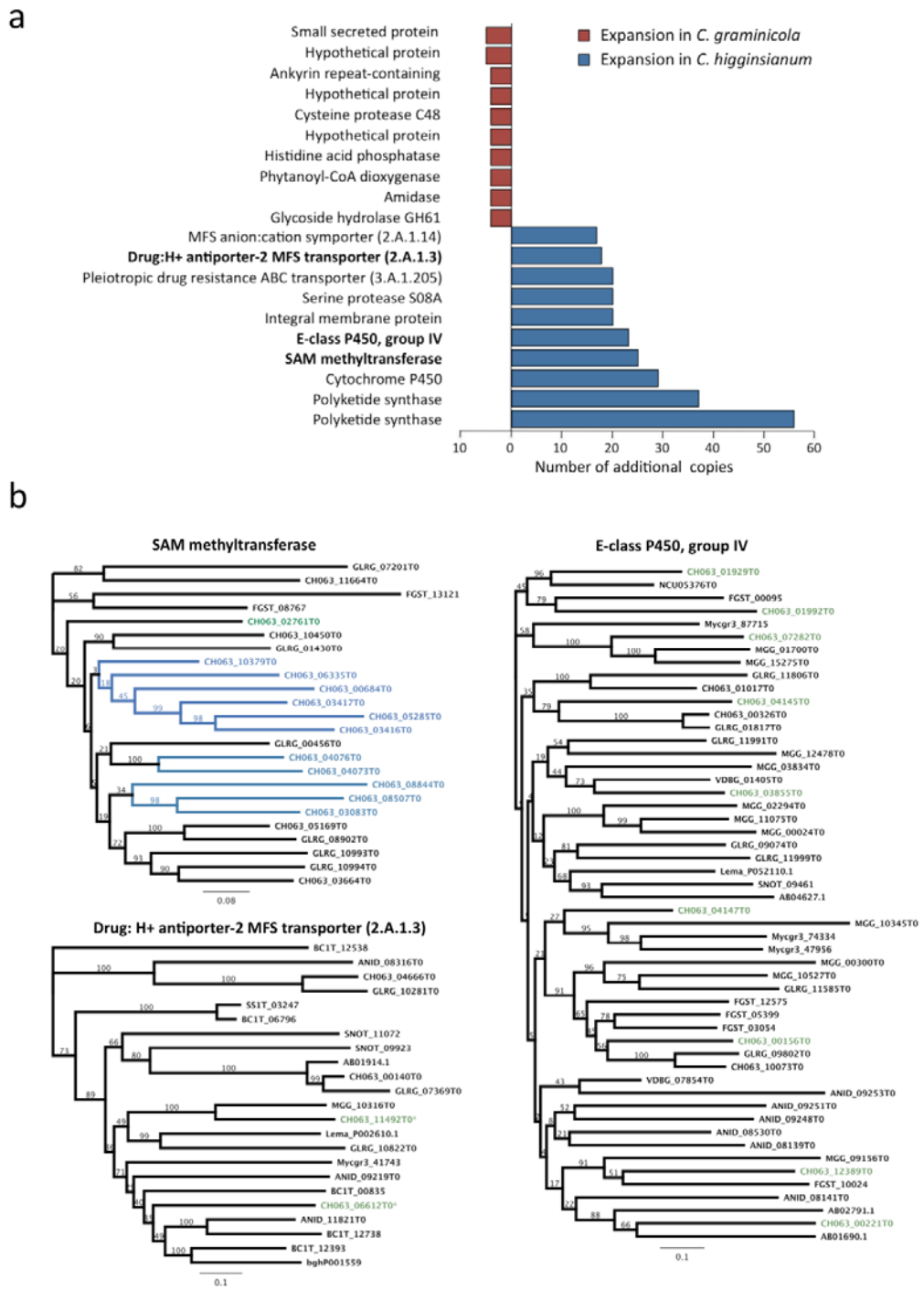


b

Species	Accession number#	Host plant	Country	GenBank accession numbers: ITS/CHS-1/HIS3/ACT/TUB2
C. higginsianum	**IMI 349063**	*Brassica campestris*	Trinidad and Tobago	JQ005760/JQ005781/JQ005802/JQ005823/JQ005844
C. higginsianum	**MAFF 305635**	*Brassica rapa* var. *komatsuna*	Japan	JQ005761/JQ005782/JQ005803/JQ005824/JQ005845
C. fuscum	CBS 130.57	*Digitalis lanata*	unknown	JQ005762/JQ005783/JQ005804/JQ005825/JQ005846
C. tabacum	CBS 161.53	*Nicotiana tabacum*	Zambia	JQ005763/JQ005784/JQ005805/JQ005826/JQ005847
C. destructivum	CBS 149.34	*Trifolium* sp.	Netherlands	JQ005764/JQ005785/JQ005806/JQ005827/JQ005848
C. linicola	CBS 172.51	*Linum usitatissimum*	Netherlands	JQ005765/JQ005786/JQ005807/JQ005828/JQ005849
G. truncata	CBS 127604, DAOM 235316, CT21	*Lens culinaris*	Canada	JQ005766/JQ005787/JQ005808/JQ005829/JQ005850
C. coccodes	CBS 164.49	*Solanum tuberosum*	Netherlands	JQ005775/JQ005796/JQ005817/JQ005838/JQ005859
C. graminicola	**CBS 130836, M1.001***	*Zea mays*	USA	JQ005767/JQ005788/JQ005809/JQ005830/JQ005851
C. graminicola	**CBS 130839, M5.001**	*Zea mays*	Brazil	JQ005768/JQ005789/JQ005810/JQ005831/JQ005852
C. navitas	CBS 125086*	*Panicum virgatum*	USA	JQ005769/JQ005790/JQ005811/JQ005832/JQ005853
C. nicholsonii	MAFF 511115*	*Paspalum dilatatum*	Japan	JQ005770/JQ005791/JQ005812/JQ005833/JQ005854
C. sublineola	CBS 131301, S3.001*	*Sorghum* sp.	Burkina Faso	JQ005771/JQ005792/JQ005813/JQ005834/JQ005855
C. falcatum	CBS 147945, CGMCC 3.14187*	*Saccharum officinarum*	Indonesia	JQ005772/JQ005793/JQ005814/JQ005835/JQ005856
C. caudatum	MAFF 305700	*Imperata cylindrica*	Japan	JQ005773/JQ005794/JQ005815/JQ005836/JQ005857
C. cereale	CBS 129663, KS20BIG*[1]	*Bromus inermis*	USA	JQ005774/JQ005795/JQ005816/JQ005837/JQ005858
C. circinans	CBS 221.81*	*Allium cepa*	Serbia	GU227855/GU228345/GU228051/GU227953/GU228149
C. dematium	CBS 125.25*	*Eryngium campestre*	France	GU227819/GU228309/GU228015/GU227917/GU228113
C. acutatum	**CBS 112996, ATCC 56816*[2]**	*Carica papaya*	Australia	JQ005776/JQ005797/JQ005818/JQ005839/JQ005860
C. fioriniae	**BB#10**	*Vaccinium* sp.	USA	JQ040070/JQ040060/JQ040065/JQ040050/JQ040075
C. boninense	CBS 123755, MAFF 305972*	*Crinum asiaticum* var. *sinicum*	Japan	JQ005153/JQ005327/JQ005414/JQ005501/JQ005588
C. hippeastri	CBS 125376, CSSG1*	*Hippeastrum vittatum*	China	JQ005231/JQ005405/JQ005492/JQ005579/JQ005665
C. truncatum	CBS 151.35*	*Phaseolus lunatus*	USA	GU227862/GU228352/GU228058/GU227960/GU228156
C. gloeosporioides	**CBS 112999, CBS 953.97, IMI 356878***	*Citrus sinensis*	Italy	JQ005152/JQ005326/JQ005413/JQ005500/JQ005587
C. musae	CBS 116870*	*Musa* sp.	USA	JQ005777/JQ005798/JQ005819/JQ005840/JQ005861
C. orbiculare	**CBS 514.97, LARS 414, 104-T**	*Cucumis sativus*	Japan	JQ005778/JQ005799/JQ005820/JQ005841/JQ005862
C. lindemuthianum	CBS 144.31	*Phaseolus vulgaris*	Germany	JQ005779/JQ005800/JQ005821/JQ005842/JQ005863
Monilochaetes infuscans	CBS 869.96	*Ipomoea batatas*	South Africa	JQ005780/JQ005801/JQ005822/JQ005843/JQ005864

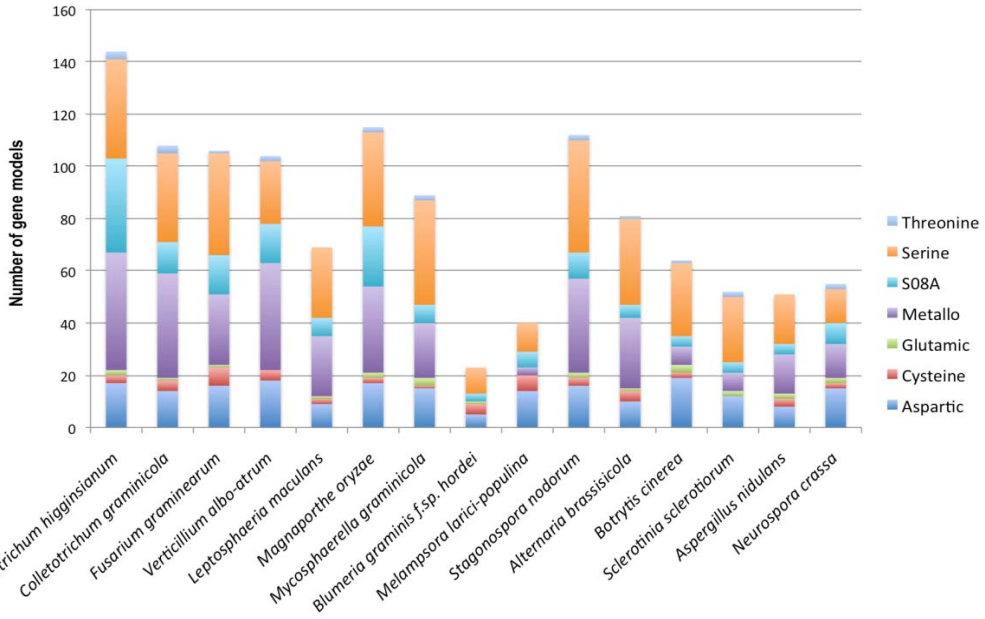
(**a**) Phylogeny of 27 *Colletotrichum* species and species complexes obtained from a maximum parsimony analysis based on a multilocus alignment (ITS, CHS-1, HIS3, ACT, TUB2). *C. higginsianum* and *C. graminicola* strains are highlighted in bold. Bootstrap support values (500 replicates) above 70 % are shown at the nodes. *Monilochaetes infuscans* was used as an outgroup. (**b**) List of *Colletotrichum* strains used for the phylogeny shown in (**a**), indicating their culture collection accession numbers, host plants, country of origin and Genbank accession numbers of the sequences generated. # 'ATCC': American Type Culture Collection, Manassas, Virginia, USA; 'CBS': Culture Collection of the Centraalbureau voor Schimmelcultures, Fungal Biodiversity Centre, Utrecht, The Netherlands; 'CGMCC': China General Microbiological Culture Collection Center, Beijing, China; 'DAOM': National Mycological Herbarium, Ottawa, Canada; 'IMI': Culture Collection of CABI Europe UK Centre, Egham, UK; 'MAFF': MAFF Genebank Project, Ministry of Agriculture, Forestry and Fisheries, Tsukuba, Japan; * ex-holotype, ex-epitype or ex-neotype cultures or forthcoming typifications by [1]J. A. Crouch and [2]U. Damm; accession numbers of strains included in genome sequence projects are bold.

Supplementary Figure 6. Expanded protein families in *C. higginsianum* and *C. graminicola* identified using MCL (Markov clustering).



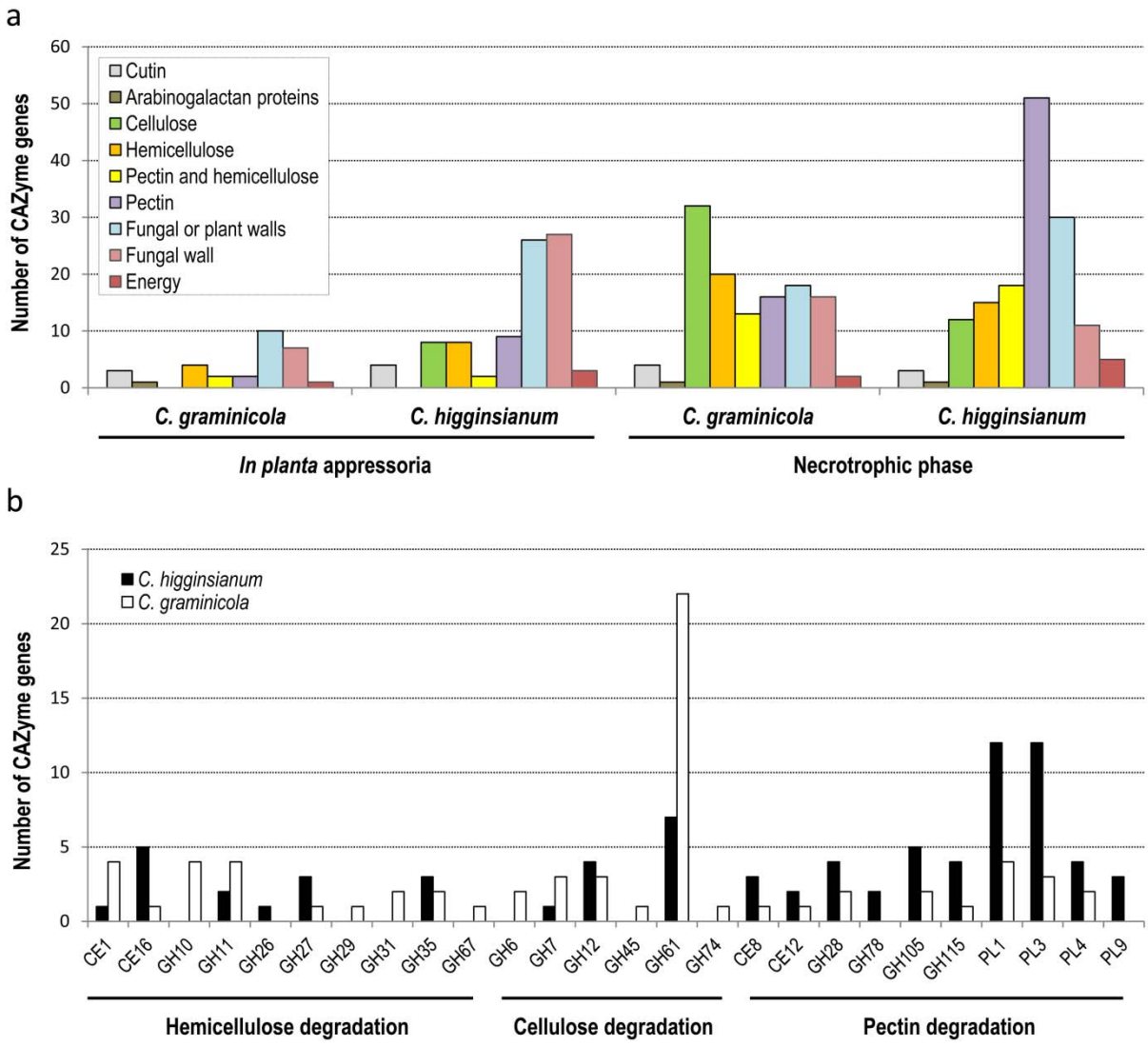
(**a**) Selected MCL protein clusters expanded by the indicated copy numbers in *C. higginsianum* and *C. graminicola* relative to the other species. Families highlighted in bold are detailed in (**b**). (**b**) Phylogenetic analysis of three MCL cluster that are expanded in *C. higginsianum* relative to *C. graminicola*. Representative subtrees of the full tree are shown. Examples of lineage specific expansion in *C. higginsianum* are evident (blue branches and nodes). Examples of ancient duplications that have been retained in *C. higginsianum* but lost in *C. graminicola* and other species. are also evident (green nodes). Protein ID prefixes indicate the species: CH063: *C. higginsianum*, GLRG: *C. graminicola*, VDBG: *Verticillium albo-atrum*, ANID: *Aspergillus nidulans*, FGST: *Fusarium graminearum*, AB: *Alternaria brassisicola*, SNOT: *Stagonospora nodorum*, SS1T: *Sclerotinia sclerotiorum*, BC1T: *Botrytis cinerea*, bgh: *Blumeria graminis*, Mycgr3: *Mycosphaerella graminicola*, Lema: *Leptosphaeria maculans*, MGG: *Magnaporthe oryzae*.

Supplementary Figure 7. Classification of genes encoding putative secreted proteases of *C. higginsianum*, *C. graminicola* and 13 other sequenced fungi.



Extracellular secreted proteins were first predicted using WolfPSort and the protease homologs were classified according to the MEROPS database (http://merops.sanger.ac.uk). The serine protease subfamily S08A (subtilisin) is expanded in *C. higginsianum*.

Supplementary Figure 8. Numbers of carbohydrate active enzyme-encoding genes significantly up-regulated by *Colletotrichum* during appressorial penetration or the necrotrophic phase.

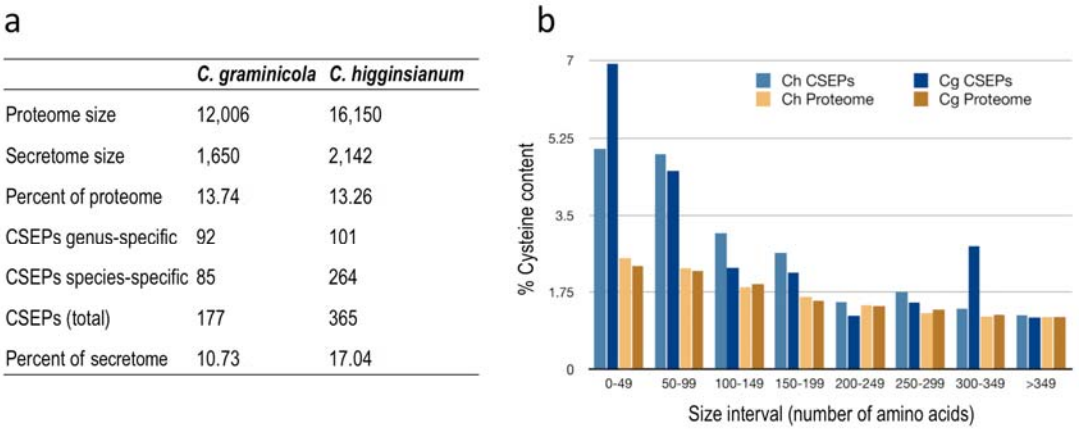


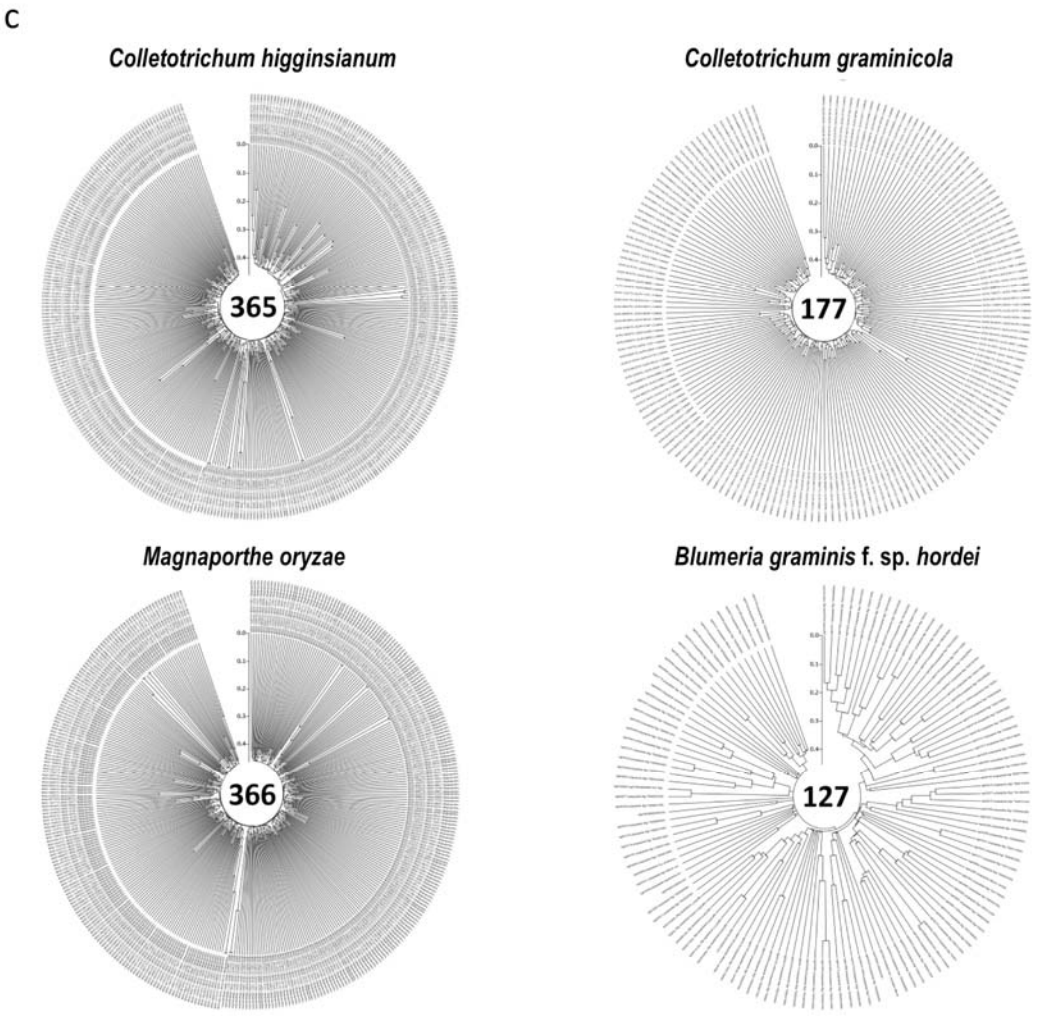
(**a**) Number of CAZyme genes significantly up-regulated (log2 fold-change >2, p-value < 0.05) by *C. graminicol*a and *C. higginsianum* in appressoria *in planta* compared to the necrotrophic phase (left) or in the necrotrophic phase compared to appressoria i*n planta* (right). Genes were categorized according to their substrates/activities (http://www.cazy.org/). (**b**) Number of CAZyme genes involved in hemicellulose, cellulose and pectin dissolution significantly up-regulated (log2 fold-change >2, p-value < 0.05) by *C. graminicol*a and *C. higginsianum* during the necrotrophic phase compared to appressoria *in planta*. CAZyme families were determined according to the CAZy database (http://www.cazy.org/).

Supplementary Figure 9. Secreted proteins and candidate secreted effector proteins (CSEPs) of *C. higginsinum* and *C. graminicola.*

a

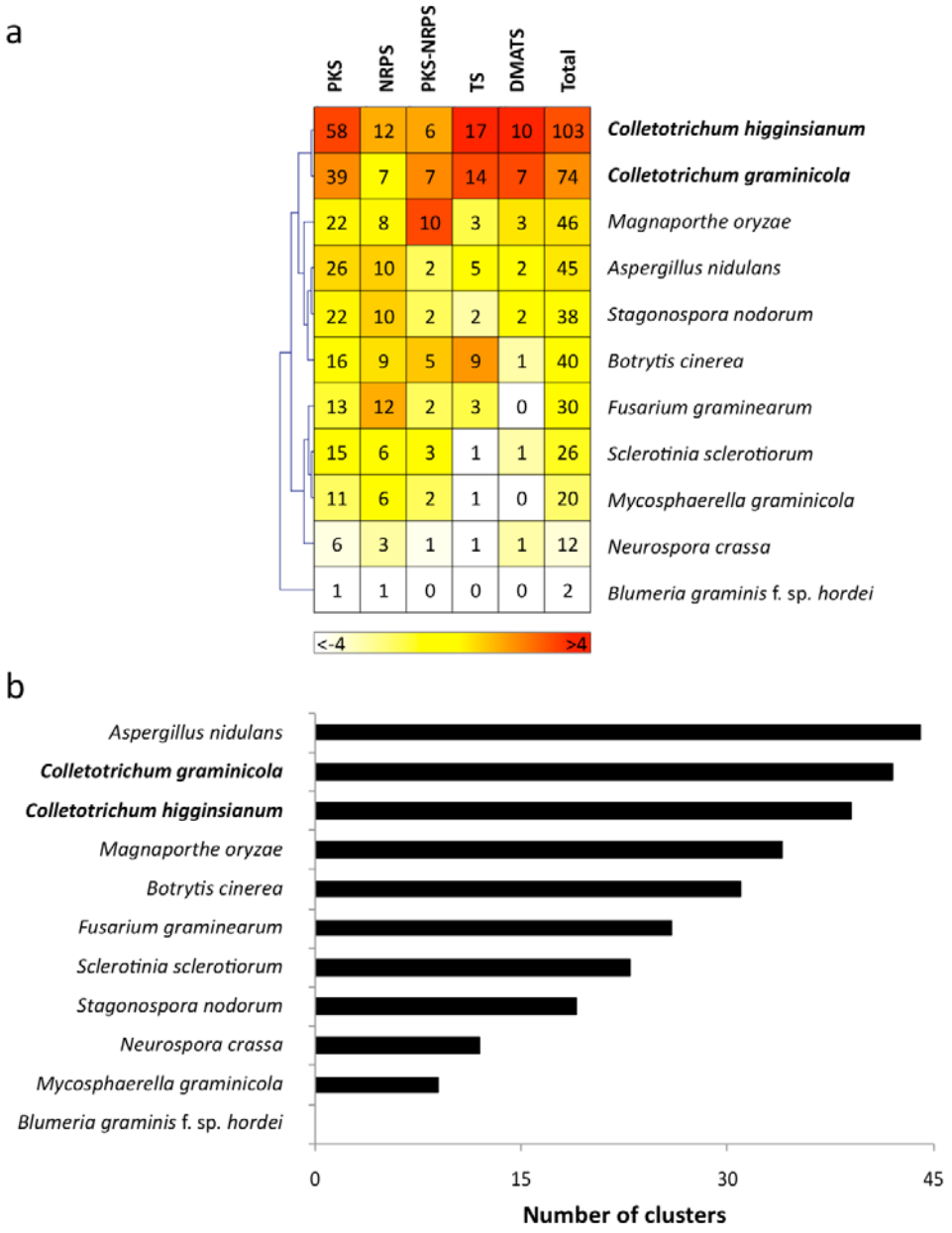
	C. graminicola	*C. higginsianum*
Proteome size	12,006	16,150
Secretome size	1,650	2,142
Percent of proteome	13.74	13.26
CSEPs genus-specific	92	101
CSEPs species-specific	85	264
CSEPs (total)	177	365
Percent of secretome	10.73	17.04





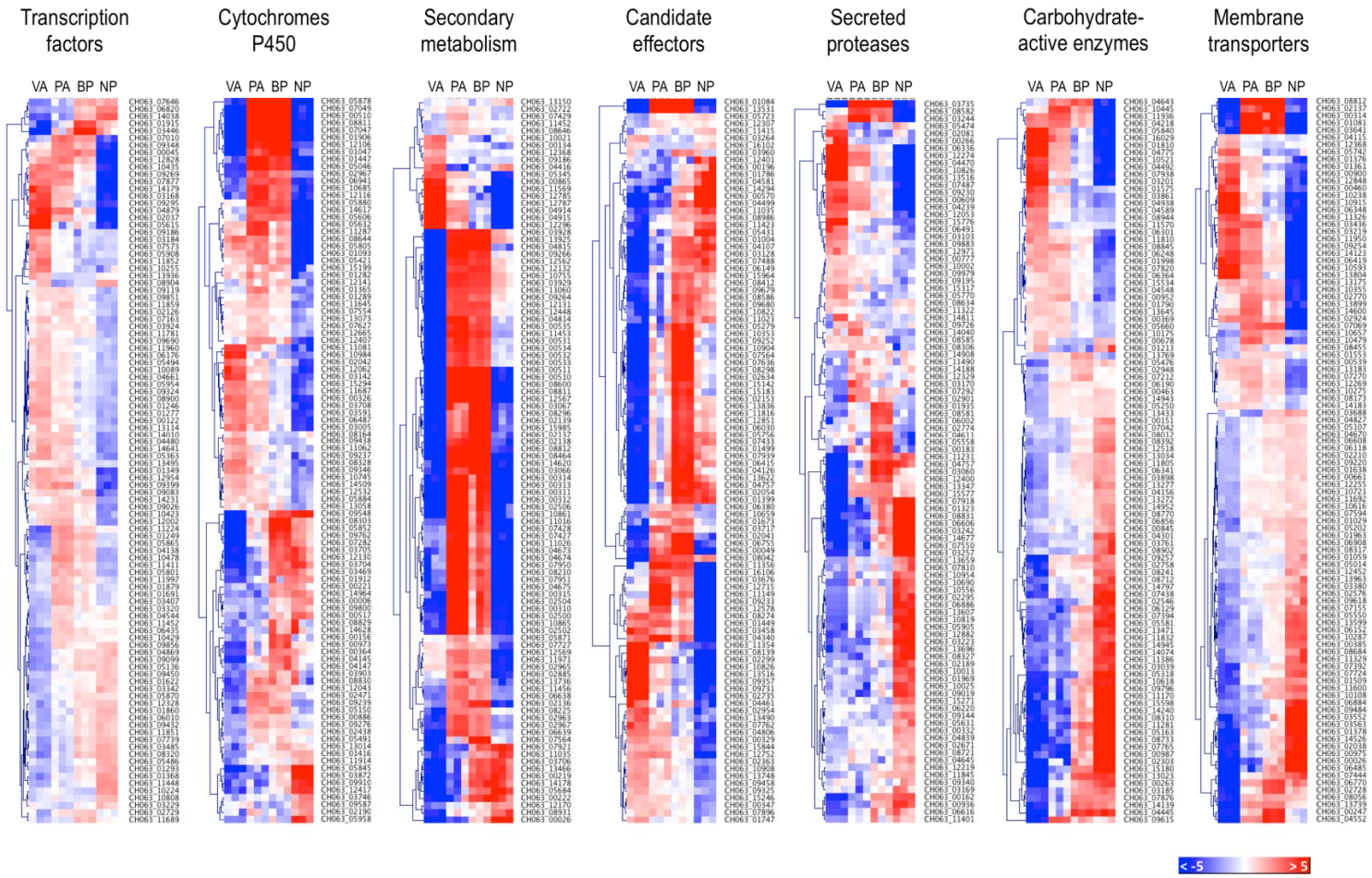
(**a**) Number of secreted proteins and Candidate Secreted Effector Proteins (CSEPs) encoded in the *C. higginsianum* and *C. graminicola* genomes. (**b**) Cysteine content of CSEPs and total proteomes of *C. graminicola* (*Cg*) and *C. higginsianum* (*Ch*). Size intervals refer to the number of amino acids after signal peptide cleavage. (**c**) Diversity of *C. higginsianum* and *C. graminicola* CSEPs and comparison with *Magnaporthe oryzae* and *Blumeria graminis* f. sp. *hordei* CSEP repertoires. Each effector phylogram was constructed based on a multiple alignment of the CSEP amino acid sequences using MEGA4.1 software (http://www.megasoftware.net/). The resulting alignment was then used to generate a neighbor-joining tree using the following parameters (pairwise deletion, amino: *p*-distance). Scale, amino acid substitutions per site. The total number of CSEPs is indicated in the center of each tree.

Supplementary Figure 10. Numbers of genes encoding key secondary metabolism (SM) enzymes and numbers of SM gene clusters in *Colletotrichum* and 9 other fungal genomes.



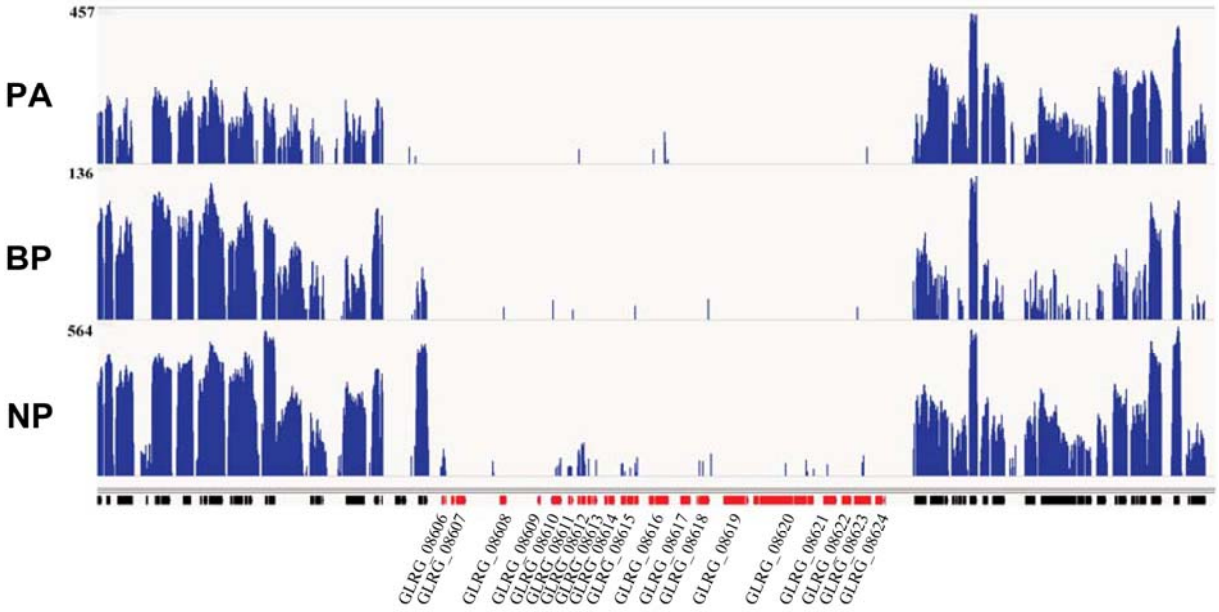
(**a**) Hierarchical clustering of SM genes from *Colletotrichum* and 9 other fungal genomes. PKS (Polyketide synthase), NRPS (*nonribosomal* peptide-synthetase), PKS-NRPS (hybrid *PKS/NRPS*), TS (*terpene* synthase), DMATS (dimethylallyltryptophan synthase). Numbers of enzymes in each genome are presented. Over-represented (orange-red) or under-represented enzymes (pale yellow-white) are depicted as fold-changes relative to the class mean. Data from Collemare et al. 2008[105]; Spanu et al. 2010[41]; Amselem et al. 2011[116] and this study. (**b**) Number of SM clusters predicted by SMURF[106] (http://jcvi.org/smurf/index.php) in *Colletotrichum* and 9 other fungi with different lifestyles.

Supplementary Figure 11. Heatmaps of *C. higginsianum* gene expression showing the 100 most highly expressed and significantly regulated genes in 7 functional categories



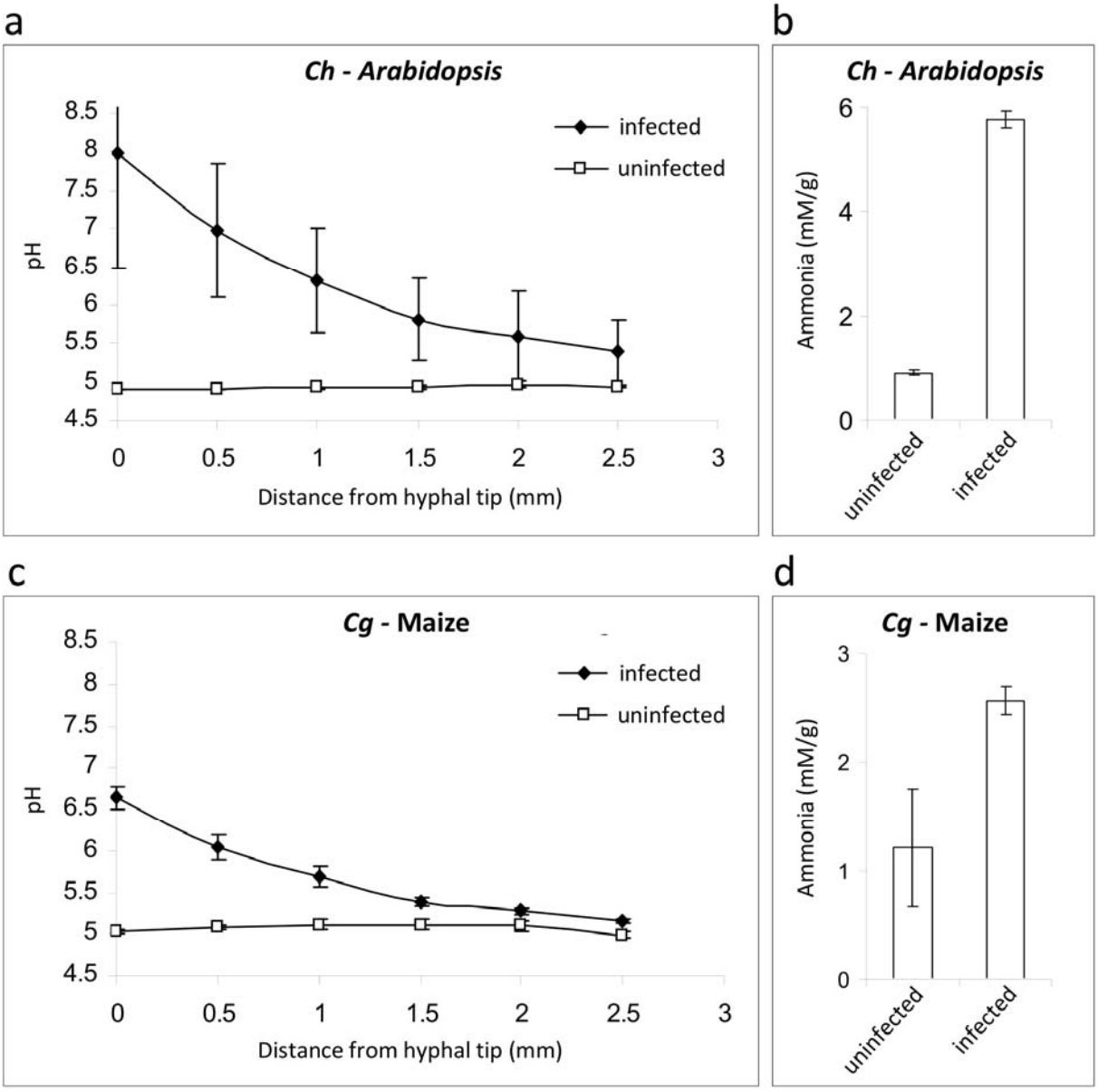
The 100 most highly expressed and significantly regulated genes (log2 fold-change >2, *p*-value <0.05) are presented together with their Broad Institute identification number. Over-represented (pale red-dark red) or under-represented transcripts (pale blue-dark blue) are depicted as log2 fold-changes relative to the mean expression measured across all four stages selected for RNA sequencing (VA: *in vitro* appressoria, PA: *in planta* appressoria, BP: biotrophic phase, NP: necrotrophic phase).

Supplementary Figure 12. Visualization of RNA-seq coverage across the *C. graminicola* polyketide biosynthesis cluster 18.



C. graminicola polyketide biosynthesis cluster 18 (gene models highlighted in red) and flanking genomic regions (gene models highlighted in black) are presented. The blue curves indicate Illumina read coverage for the three samples (PA: *in planta* appressoria, BP: biotrophic phase, NP: necrotrophic phase). The gene cluster is silent whereas flanking genes are transcriptionally active. On the vertical axis, read counts are shown on a log scale.

Supplementary Figure 13. *Colletotrichum* species produce ammonia, resulting in alkalinization of the hyphal cytoplasm and host tissues.



(**a**) and (**b**) show data for *C. higginsianum* (*Ch*) infecting *A. thaliana* leaves, 5 days post inoculation. (**c**) and (**d**) show data for *C. graminicola* (*Cg*) infecting *Zea mays* leaves, 5 days post inoculation. (**a**) and (**c**) Graphs showing measurements of host cell cytoplasmic pH with increasing distance (0.5 mm intervals) from the tips of necrotrophic secondary hyphae of *C. higginsianum* (**a**) or *C. graminicola* (**c**). Mean values from 50 cells (± 1 standard deviation) are presented. Data points for uninfected *A. thaliana* and *Zea mays* leaf cells represent measurements at randomly selected positions in six independent leaves. Cytoplasmic pH was calculated from fluorescence microscope images of tissues stained with the pH sensitive dye BCECF[160]. (**b**) and (**d**) Ammonia concentrations within necrotic lesions on infected leaves and in control leaves measured with the MERCK Ammonium assay kit. Means of 3 independent measurements (± 1 standard deviation) are presented.

Supplementary Table 1. *C. graminicola* and *C. higginsianum* optical map statistics

a. *C. graminicola*

Chromosomes	Size (Mb)	Coverage (fold)#
Chr 1	7.43	50.56
Chr 2	7.41	46.3
Chr 3	6.58	50.93
Chr 4	6.48	49.19
Chr 5	6.21	48.41
Chr 6	4.84	52.05
Chr 7	4.64	61.5
Chr 8*	4.37	64.52
Chr 9	4.01	60.91
Chr 10	3.49	50.61
Chr 11†	0.76	59.24
Chr 12†	0.71	53.82
Chr 13†	0.51	26.84
Total	**57.44**	**51.91**

* Contains rDNA repeats

Average number of single molecule optical maps used to compute the size of each fragment in the chromosome optical map

† Mini chromosome (<1Mb)

b. *C. higginsianum*

Chromosomes	Size (Mb)	Coverage (fold) #
Chr 1	6.29	85.36
Chr 2	6.29	94.39
Chr 3	6.26	82.19
Chr 4	5.99	93.42
Chr 5	5.53	80.39
Chr 6	5.41	92.44
Chr 7	4,70	87.64
Chr 8	4.55	82.82
Chr 9*	3.87	124.47
Chr 10	3.20	103.46
Chr 11†	0.65	80.33
Chr 12†	0.61	97.26
Total	**53.35**	**92.01**

* Contains rDNA repeats

Average number of single molecule optical maps used to compute the size of each fragment in the chromosome optical map.

† Mini chromosome (<1Mb)

c. *C. graminicola* strain M1.001 sequence assembly anchored to the optical maps

Optical linkage group (Chromosomes)	Size (Mb)*	IDs of mapped Scaffolds#	Mapped Size (bp)	Coverage (%)
Chr 1	7.429	55,66,70,48,16,9,8,2,42	6,168,515	83
Chr 2	7.408	23,54,69,26,59,17,40,30,36,18,50	5,277,658	71
Chr 3	6.58	7,27,14,4,60,32,77	5,023,335	76
Chr 4	6.481	35,6,11,39,1,62	5,234,216	80
Chr 5	6.203	33,56,34,5,38,22,76,29,53,84	4,706,884	76
Chr 6	4.834	47,19,12,24,44,65	3,351,253	69
Chr 7	4.64	25,13,57,10,46	3,339,033	72
Chr 8	4.366	3,15	2,285,310	53
Chr 9	4.004	43,45,20,21,28	2,722,476	68
Chr 10	3.489	51,52,63,41	1,322,627	38
Chr 11†	0.76	101, 115,96	186,627	25
Chr 12†	0.71	105, 104	117,848	17
Chr 13†	0.51	78	138,786	27
Total	**57.414**		**40,156,429**	**70**

* as determined by optical mapping

Scaffolds can be retrieved from the Broad Institute *Colletotrichum* database (http://www.broadinstitute.org/annotation/genome/colletotrichum_group/MultiHome.html)

† Mini chromosome (<1Mb)

Supplementary Table 2. Genome characteristics of *C. graminicola* and *C. higginsianum.*

Genome features	*C. graminicola*	*C. higginsianum*
Chromosome number	13	12
Genome physical size (Mb)	57.44	53.35
NCBI accession no. of assembly	ACOD01000001	CACQ02000000
Total contig length (Mb)	50.87*	49.08
Total aligned sequence (Mb)	48.84	38.6
Contig number	1,151	10,269
Scaffold number	653	367
Average base coverage (Fold)	9.1	101
Sanger	7.9	0.2
Roche 454	1.2	25
Illumina	-	76
N_{50} contig (kb)	228.96	6.15
N_{50} scaffold (Kb)	579.19	265.5
GC-content (%)		
Scaffolds	49.12	55.10
Exons	58.36	59.33
Introns	50.99	51.63
Intergenic regions	44.22	52.78
Protein-coding genes	12,006	16,172
Mean transcript length (bp)	1,397	1,095
Number of exons	32,967	39,537
Mean number of introns/gene	2.7	2.4
Intergenic distance (bp)	2,610	1,112
Percentage coding	32.51	36.08
Genes with Sanger/454 EST evidence	4,682	14,393
Genes with Illumina RNA-Seq evidence†	10,812	14,502
tRNA genes	295	352
Repetitive DNA in genome assembly (%)	12.23	1.22#
Gene space coverage (%)	99.2	95.1

* 98.5% of assembled bases have quality score ≥ than Q40.
† 5 or more mapped Illumina reads
# Underestimated for *C. higginsianum* because assembly programs excluded repetitive elements

Supplementary Table 3. Repetitive elements in the *C. graminicola* genome.

ID	Family	Overall length (bp)	Percent of genome	Number of elements	Length of reference sequence (bp)	Number of full length elements*
repeat_00001	LTR	550,521	1.07	345	9,740	11
repeat_00002	LTR	360,874	0.70	254	7,351	7
repeat_00003	LTR	739,367	1.43	404	7,010	15
repeat_00004	LTR	331,902	0.64	221	7,026	13
repeat_00005	LTR	732,849	1.42	419	11,941	3
repeat_00006	LTR	259,609	0.50	196	8,147	4
repeat_00007	LTR	270,498	0.52	181	6,425	8
repeat_00008	LTR	924,978	1.79	802	6,147	15
repeat_00009	LTR	278,792	0.54	153	5,855	25
repeat_00010	LTR	111,372	0.22	85	5,385	7
repeat_00011	LTR	480,636	0.93	418	8,497	4
repeat_00012	LTR	145,554	0.28	91	6,718	4
repeat_00013	LTR	171,317	0.33	160	7,582	2
repeat_00014	DNA	372,372	0.72	443	3,146	36
repeat_00015	DNA	122,214	0.24	159	2,923	5
repeat_00016	DNA	78,220	0.15	93	1,890	24
repeat_00017	DNA	214,327	0.42	289	1,864	40
repeat_00018	DNA	141,453	0.27	178	1,862	39
repeat_00019	DNA	177,187	0.34	174	1,867	41
repeat_00020	Unknown	51,246	0.10	143	644	23
repeat_00021	Unknown	26,558	0.05	113	597	13
repeat_00022	Unknown	7,518	0.01	29	363	15

* A sequence was considered full-length if within 95% of the length of the reference sequence

Supplementary Table 4. Coverage of *C. graminicola* chromosomes determined from PROmer alignments of *C. higginsianum* sequences.

C. graminicola **chromosomes**	**Length (bp)#**	**Coverage (bp)**	**Coverage (%)**
Chr 1	6,168,515	2,791,603	45.26
Chr 2	5,277,658	2,479,887	46.99
Chr 3	5,023,335	2,325,359	46.29
Chr 4	5,234,216	2,305,646	44.05
Chr 5	4,706,884	1,885,495	40.06
Chr 6	3,351,253	1,474,342	43.99
Chr 7	3,339,033	1,550,402	46.43
Chr 8	2,285,310	885,754	38.76
Chr 9	2,722,476	1,121,752	41.20
Chr 10	1,322,627	958,216	72.45
Chr 11	186,627	1,794	0.96
Chr 12	117,848	0	0.00
Chr 13	138,786	111	0.08
Chr unknown	10,992,380	177,720	1.62
Total	**50,866,948**	**17,976,401**	**35.34**

Total length of anchored scaffold

Supplementary Table 5. Relationship of homologous chromosomes between *C. graminicola* and *C. higginsianum* determined using PROmer alignments.

C. graminicola	*C. higginsianum*
Chr 1-a	Chr 3
Chr 1-b	Chr 5-b
Chr 2	Chr 4
Chr 3	Chr 2
Chr 4	Chr 1
Chr 5-a	Chr 6-a
Chr 5-b	Chr 10-a
Chr 6	Chr 8
Chr 7	Chr 7
Chr 8	Chr 9
Chr 9	dispersed
Chr 10	Chr 5-a, Chr 10-b

Nature Genetics: doi:10.1038/ng.2372

Supplementary Table 6. Summary of gene homology within the MCL-clustered proteome of *C. graminicola*, *C. higginsianum* and four other Sordariomycetes.

Species	Orphans[a]	Conserved single-copy[b]	Multi-copy[c]
Colletotrichum graminicola	582	6,080	5,342
Colletotrichum higginsianum	1,326	4,732	10,077
Magnaporthe oryzae	726	5,299	5,027
Fusarium graminearum	990	5,216	7,108
Verticillium albo-atrum	501	5,059	4,655
Neurospora crassa	898	5,105	3,898

[a] Orphans: proteins having no homology to other proteins.
[b] Conserved single-copy: proteins that have no paralogs but do have orthologs.
[c] Multi-copy: proteins that have one or more paralogs.

Supplementary Table 9. Number of Carbohydrate-Active Enzyme modules of *C. graminicola*, *C. higginsianum* and thirteen other fungal species determined according to the CAZy database (http://www.cazy.org/).

	Glycosyl Hydrolase family (GHs)																																				
Species	1	2	3	5	6	7	9	10	11	12	13	15	16	17	18	20	23	24	25	26	27	28	29	30	31	32	33	35	36	37	38	39	43	45	47	49
Melampsora larici-populina	0	4	3	29	0	8	1	6	0	10	7	4	11	2	15	3	1	0	0	5	7	3	0	0	3	2	0	1	0	3	1	0	8	0	13	0
Blumeria graminis f. sp. *hordei*	0	0	1	3	0	0	0	0	0	0	4	0	9	7	8	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	4	0
Botrytis cinerea	3	2	16	15	1	2	0	2	3	4	10	4	21	6	10	1	0	0	1	2	4	18	0	0	4	1	0	4	2	1	1	1	4	2	8	0
Sclerotinia sclerotiorum	3	2	13	14	1	3	0	2	3	5	10	4	19	6	13	1	0	0	1	1	3	17	0	0	6	1	0	4	2	1	1	0	4	2	8	0
Fusarium graminearum	3	10	22	15	1	2	0	5	3	4	8	3	21	6	19	2	0	0	0	0	2	6	1	0	8	5	1	3	3	2	1	2	17	1	11	0
Verticillium albo-atrum	3	4	15	10	3	2	0	3	4	4	3	3	5	4	7	1	0	1	0	0	2	8	0	1	7	2	1	2	1	1	1	1	14	1	3	1
Colletotrichum graminicola	**2**	**7**	**18**	**15**	**4**	**6**	**0**	**10**	**6**	**6**	**13**	**3**	**22**	**7**	**22**	**3**	**0**	**0**	**0**	**1**	**4**	**9**	**2**	**2**	**6**	**3**	**0**	**5**	**2**	**2**	**2**	**0**	**18**	**2**	**11**	**0**
Colletotrichum higginsianum	**1**	**7**	**30**	**22**	**4**	**8**	**0**	**7**	**5**	**10**	**16**	**5**	**25**	**8**	**24**	**3**	**0**	**0**	**0**	**1**	**3**	**15**	**1**	**3**	**7**	**0**	**0**	**5**	**2**	**3**	**1**	**0**	**24**	**1**	**14**	**0**
Neurospora crassa	1	5	9	7	3	5	0	4	2	1	10	2	14	4	12	0	0	0	1	1	0	2	0	2	5	1	0	2	1	2	1	0	7	1	8	0
Magnaporthe oryzae	2	8	18	13	3	6	0	6	5	3	10	2	18	7	15	3	0	1	0	0	4	3	4	1	6	4	1	0	2	2	2	1	20	1	9	0
Aspergillus nidulans	3	10	20	15	2	3	0	3	2	1	13	2	13	5	19	2	0	1	3	3	3	9	0	0	10	2	0	4	4	1	1	1	15	1	7	0
Mycosphaerella graminicola	2	8	16	9	0	1	0	2	1	1	14	1	13	7	10	0	0	0	0	0	1	2	1	1	7	4	0	2	2	2	1	1	10	1	9	0
Alternaria brassicicola	3	7	13	15	3	5	0	5	4	2	7	2	17	9	8	3	0	0	0	1	6	7	1	2	8	2	0	5	2	2	1	0	19	3	9	0
Stagonospora nodorum	2	10	16	18	4	5	0	7	7	4	7	3	18	6	17	3	0	0	0	0	3	4	0	2	11	4	0	4	2	2	1	1	15	3	9	0
Leptosphaeria maculans	3	7	13	15	3	3	0	3	2	3	7	3	18	8	10	2	0	0	0	1	4	6	1	2	8	2	0	4	2	1	1	1	11	2	8	0

Species	51	53	54	55	61	62	63	64	65	67	71	72	74	75	76	78	79	81	85	88	89	92	93	94	95	105	114	115	125	**Total**
Melampsora larici-populina	3	0	0	0	4	0	1	0	0	0	5	1	0	0	3	0	2	2	2	0	0	2	0	0	0	0	0	0	0	**175**
Blumeria graminis f. sp. *hordei*	0	0	0	1	1	0	1	0	0	0	0	3	0	0	7	1	0	1	0	0	0	2	1	0	0	0	1	0	2	**61**
Botrytis cinerea	3	2	1	4	9	1	1	2	2	0	9	6	0	0	11	8	2	1	0	1	0	5	1	0	2	1	0	1	3	**229**
Sclerotinia sclerotiorum	2	2	1	4	9	0	1	2	2	0	9	6	0	0	11	4	2	1	0	0	1	5	1	0	1	1	1	1	3	**220**
Fusarium graminearum	2	1	1	3	15	1	1	2	0	1	0	3	1	1	8	7	1	1	0	1	0	0	2	0	2	3	2	2	3	**251**
Verticillium albo-atrum	2	1	1	1	14	0	0	2	0	1	0	4	1	1	7	7	1	1	0	3	0	2	0	0	1	3	1	3	1	**176**
Colletotrichum graminicola	**2**	**2**	**1**	**5**	**28**	**2**	**1**	**1**	**1**	**1**	**5**	**6**	**1**	**1**	**9**	**4**	**4**	**2**	**0**	**1**	**0**	**4**	**3**	**0**	**0**	**4**	**1**	**2**	**3**	**307**
Colletotrichum higginsianum	**3**	**2**	**0**	**9**	**25**	**1**	**1**	**2**	**0**	**1**	**8**	**4**	**4**	**1**	**11**	**6**	**4**	**2**	**0**	**1**	**0**	**6**	**1**	**0**	**1**	**7**	**0**	**5**	**3**	**361**
Neurospora crassa	1	1	1	6	14	0	1	2	0	1	6	5	1	1	10	0	2	1	0	0	0	2	2	1	0	1	1	1	3	**177**
Magnaporthe oryzae	3	1	1	4	22	4	2	2	0	1	1	5	1	1	8	3	2	2	0	1	0	6	1	1	1	3	1	3	4	**264**
Aspergillus nidulans	2	1	1	6	9	2	1	0	1	1	5	5	2	2	7	8	1	1	0	2	0	5	2	0	3	3	2	1	1	**252**
Mycosphaerella graminicola	3	2	1	3	2	1	1	4	1	0	1	7	0	1	7	2	4	1	1	0	0	7	1	0	0	2	1	1	3	**186**
Alternaria brassicicola	2	1	1	4	22	2	2	1	1	1	0	5	0	0	9	1	2	2	1	1	0	4	1	1	1	3	1	2	2	**244**
Stagonospora nodorum	2	1	1	3	30	3	2	1	1	1	1	7	0	1	8	4	2	2	0	1	0	7	3	1	2	3	4	2	3	**284**
Leptosphaeria maculans	3	1	0	3	20	1	3	1	1	1	1	6	0	0	7	1	1	2	1	1	0	6	2	1	1	3	2	2	3	**228**

Species	Glycosyltransferase family (GTs)																																											
	1	2	3	4	5	8	15	17	18	20	21	22	24	25	31	32	33	34	35	39	41	48	49	50	55	57	58	59	61	62	64	66	69	71	76	90	**Total**
Melampsora larici-populina	6	14	1	5	2	3	2	0	2	2	1	3	1	0	5	4	1	0	0	3	1	8	1	1	0	2	1	1	0	0	0	1	1	2	1	9	**84**
Blumeria graminis f. sp. *hordei*	4	8	1	4	0	2	2	0	0	2	1	4	1	0	3	3	1	2	1	3	1	1	0	1	0	2	1	0	0	3	0	1	1	0	1	2	**56**
Botrytis cinerea	11	13	1	6	1	7	4	2	0	3	1	4	1	2	6	6	0	1	1	3	1	1	0	1	0	2	1	0	0	3	0	1	3	5	1	3	**95**
Sclerotinia sclerotiorum	10	15	1	4	1	6	4	2	0	3	1	4	1	3	4	4	1	2	1	3	1	1	0	1	0	2	1	1	0	3	0	1	2	2	1	3	**89**
Fusarium graminearum	16	19	1	5	0	6	5	1	0	3	1	4	1	0	5	5	1	2	1	3	1	1	0	1	0	2	1	1	0	2	2	1	1	4	1	6	**103**
Verticillium albo-atrum	8	15	1	5	0	3	2	0	0	2	1	3	1	0	4	1	0	1	1	3	1	0	0	0	0	0	0	0	0	1	0	1	2	0	1	1	**58**
Colletotrichum graminicola	**15**	**16**	**1**	**5**	**3**	**5**	**4**	**2**	**0**	**3**	**1**	**4**	**1**	**2**	**5**	**4**	**1**	**2**	**1**	**3**	**1**	**1**	**0**	**1**	**0**	**2**	**1**	**1**	**0**	**3**	**0**	**1**	**4**	**4**	**1**	**5**	**103**
Colletotrichum higginsianum	**22**	**18**	**1**	**5**	**3**	**6**	**4**	**2**	**0**	**4**	**1**	**6**	**1**	**2**	**6**	**4**	**1**	**2**	**1**	**5**	**1**	**1**	**0**	**1**	**0**	**2**	**1**	**3**	**0**	**3**	**0**	**1**	**5**	**4**	**1**	**6**	**123**
Neurospora crassa	6	11	1	6	2	2	5	1	0	3	1	3	1	1	3	4	1	2	1	3	1	1	0	1	1	2	1	1	0	3	0	1	2	1	1	3	**76**
Magnaporthe oryzae	12	11	1	4	1	2	4	1	0	3	1	4	1	8	4	8	1	2	1	3	1	1	0	1	1	2	1	1	3	3	0	1	4	4	1	6	**102**
Aspergillus nidulans	9	12	1	7	2	5	3	0	0	3	1	4	1	4	5	7	1	3	1	3	1	1	0	1	0	2	1	1	0	3	0	1	3	2	2	2	**92**
Mycosphaerella graminicola	5	14	1	4	3	10	3	2	0	3	1	4	1	3	7	7	1	5	1	3	1	1	0	1	0	2	1	1	0	3	1	1	1	2	1	5	**99**
Alternaria brassicicola	7	14	2	5	0	5	2	0	0	3	1	4	1	2	4	6	1	5	1	3	1	0	0	1	1	1	1	1	0	3	0	1	3	1	0	3	**83**
Stagonospora nodorum	7	16	1	5	0	8	3	0	0	3	1	4	1	3	6	6	1	6	1	3	1	1	0	1	1	2	1	1	0	4	0	1	3	1	1	3	**96**
Leptosphaeria maculans	6	16	1	5	0	9	3	0	0	3	1	4	1	2	5	5	1	7	1	3	1	1	0	1	1	2	1	1	0	3	0	1	2	2	1	4	**94**

Species	Polysaccharide lyase family (PLs)									Carbohydrate esterase family (CEs)											Carbohydrate-binding module family (CBMs)																									
	1	3	4	7	9	11	14	20	**Total**	1	2	3	4	5	8	9	12	15	16	**Total**	1	6	12	13	14	18	20	21	24	32	35	38	42	43	48	50	52	63	**Total**
Melampsora larici-populina	4	0	0	0	0	0	2	0	**6**	1	0	0	18	12	4	0	0	0	1	**36**	0	0	1	0	0	0	0	1	0	0	0	0	0	1	2	0	0	2	**7**
Blumeria graminis f. sp. *hordei*	0	0	0	0	0	0	0	0	**0**	2	0	1	4	2	0	0	0	0	1	**10**	0	0	0	0	0	10	0	1	0	0	0	0	0	1	2	0	0	0	**14**
Botrytis cinerea	6	2	0	1	0	0	0	0	**9**	3	0	1	5	10	5	2	3	0	5	**34**	18	0	0	1	0	18	3	1	15	2	3	0	1	1	3	1	0	1	**68**
Sclerotinia sclerotiorum	4	0	0	1	0	0	0	0	**5**	3	0	1	4	8	5	1	3	1	6	**32**	19	0	0	1	0	18	3	1	15	1	1	0	1	1	3	0	0	1	**65**
Fusarium graminearum	9	7	3	0	1	0	0	1	**21**	5	1	5	7	12	6	1	3	0	2	**42**	12	1	0	2	0	35	2	2	1	4	2	0	1	1	2	2	0	3	**70**
Verticillium albo-atrum	8	6	2	0	1	1	0	0	**18**	5	0	5	7	11	0	1	2	1	3	**35**	20	0	0	0	0	17	3	1	0	0	2	1	2	1	0	3	0	1	**51**
Colletotrichum graminicola	**7**	**4**	**3**	**0**	**1**	**0**	**0**	**0**	**15**	**8**	**0**	**6**	**14**	**12**	**3**	**1**	**4**	**1**	**3**	**52**	**21**	**1**	**0**	**0**	**0**	**29**	**3**	**1**	**5**	**0**	**5**	**0**	**1**	**2**	**3**	**23**	**0**	**1**	**95**
Colletotrichum higginsianum	**15**	**14**	**6**	**0**	**4**	**0**	**0**	**0**	**39**	**7**	**0**	**4**	**15**	**14**	**10**	**1**	**3**	**3**	**5**	**62**	**17**	**0**	**0**	**0**	**0**	**26**	**3**	**1**	**16**	**0**	**5**	**0**	**0**	**1**	**5**	**29**	**0**	**1**	**104**
Neurospora crassa	1	1	1	0	0	0	0	1	**4**	7	0	3	4	3	1	1	1	1	1	**22**	19	0	0	2	0	3	2	1	8	1	0	1	0	1	1	2	1	0	**42**
Magnaporthe oryzae	2	1	1	0	0	0	0	1	**5**	10	1	6	11	18	1	1	3	1	2	**54**	22	2	0	0	0	57	3	1	0	0	3	0	2	2	3	16	1	1	**113**
Aspergillus nidulans	8	5	4	0	1	1	0	2	**21**	3	0	6	8	4	3	1	2	0	3	**30**	7	0	0	0	1	17	4	1	2	0	2	0	1	3	1	2	0	1	**42**
Mycosphaerella graminicola	2	1	0	0	0	0	0	0	**3**	3	0	2	5	6	1	0	0	0	2	**19**	0	0	0	0	0	6	3	1	0	1	1	0	0	1	2	4	1	2	**22**
Alternaria brassicicola	8	11	3	0	1	0	0	0	**23**	4	1	5	6	7	7	2	4	2	3	**41**	9	0	0	0	0	26	2	1	2	1	3	0	1	2	3	5	0	1	**56**
Stagonospora nodorum	4	2	4	0	0	0	0	0	**10**	11	1	4	12	11	6	2	3	1	2	**53**	13	1	0	0	0	46	3	1	0	1	3	0	1	2	2	1	0	1	**75**
Leptosphaeria maculans	9	6	4	0	0	0	0	0	**19**	4	1	2	7	8	3	2	3	2	2	**34**	5	0	0	0	0	29	3	1	0	1	3	0	0	2	2	5	0	1	**52**

Supplementary Table 10. *C. higginsianum* and *C. graminicola* gene models with CBM50 (LysM domains).

gene ID[a]	Signal Peptide[b]	No. LysM domains[c]	Predicted localization[d]	Number of amino acids
C. higginsianum				
CH063_13023	Yes	2	extra	170
CH063_04445	Yes	2	extra	176
CH063_06689	Yes	1	extra	82
CH063_15796	Yes	2	extra	272
CH063_00259	Yes	3	extra	314
CH063_04323	Yes	2	extra	301
CH063_11930	Yes	2	extra	225
no gene called[e]	Yes	1	extra	114
CH063_05398	Yes	2	nucl	270
CH063_05984	Missing N terminus	1	extra	387
CH063_04710	No	1	extra	121
CH063_16015	Missing N terminus	3	extra	377
CH063_11929	Missing N terminus	3	extra	226
CH063_08917	No	2	extra	131
CH063_02233	No	1	plas	186
CH063_05461	No	2	nucl	354
CH063_02695	No	1	nucl	586
CH063_08603	No	1	cyto	99
CH063_04951	No	1	cyto	106
CH063_16043	No	2	cyto	144
C. graminicola				
GLRG_11587	Yes	1 or 3	extra	779
GLRG_07767	Yes	2	extra	153
GLRG_07795	Yes	1	extra	813
GLRG_02709	Yes	1	extra	650
GLRG_06967	Yes	1 or 2	extra	258
GLRG_11081	No	1	extra	347
GLRG_02947	No	2	nucl	154
GLRG_08515	No	1	cyto	358
GLRG_00922	No	1	cyto_nucl	252
GLRG_07095	No	1	plas	186
GLRG_11477	No	1	nucl	141
GLRG_06565	No	1	mito	294
GLRG_11181	No	1	nucl	351
GLRG_06968	No	1 or 2	cyto_nucl	120

[a]Broad Institute identification number (http://www.broadinstitute.org/annotation/genome/colletotrichum_group/MultiHome.html)

[b] Signal peptides were predicted with SignalP

[c]The number of LysM domains were determined by searching the NCBI CDD and using Broad Institute annotation.

[d] Subcellular localization was determined with WoLF PSORT. Extra, extracellular; plas, plasmamembrane; cyto, cytoplasm; mito, mitochondria; nucl, nucleus.

[e]MRSSIIYLALAGLTSFAQANPLNKPPPQTDPRGLITRGLSNVFSPTKRQAAAPPAIPQAGGGDVNNASRNVALTVIAGDTLGQIARLLNSGICDIAKLNNVANPDFIEVGQVLQ (EST evidence)

Supplementary Table 11. *C. higginsianum* and *C. graminicola* gene models with CE4, GH18 or CBM18 domains.

	C. higginsianum		*C. graminicola*		
Gene ID[a]	**SignalP, localization, TMHMM, GPI[b]**	**Gene ID[a]**	**SignalP, localization, TMHMM, GPI[b]**	**Domain(s)**	**No. identical aa/length of alignment**
Chitin deacetylase					
CH063_05487	SP, extra, no TMD, no GPI	GLRG_11241	SP, extra, no TMD, no GPI	CE4 (*Ch*), CE4-CBM18 (*Cg*)	336/406
CH063_13276	SP, extra, no TMD, no GPI	GLRG_00386	SP, extra, no TMD, no GPI	CE4	171/227
CH063_07481	SP, extra, no TMD, no GPI	GLRG_04776	SP, extra, no TMD, no GPI	CE4	629/761
CH063_15412	SP, extra, no TMD, no GPI	GLRG_05587	SP, extra, no TMD, no GPI	CE4	223/254
CH063_02920	SP, extra, no TMD, no GPI	GLRG_07915	SP, extra, no TMD, no GPI	CE4	137/231
CH063_03329	SP, extra, no TMD, GPI	GLRG_04084	SP, extra, no TMD, GPI	CE4	311/409
CH063_14379	SP, extra, no TMD, no GPI	GLRG_11148	NS, nucl, no TMD	CE4-CBM18 (*Ch*); CE4-CBM18-CBM18 (*Cg*)	323/455
CH063_14486	SP, extra, no TMD, no GPI	GLRG_02208	NS, extra, no TMD	CE4-CBM18	348/470
CH063_01681	SP, extra, no TMD, no GPI	GLRG_03854	NS, cyto, noTMD	CE4	178/246
CH063_07732	SA, mito, TMD	GLRG_03976	SP, plas, TMD, no GPI	CE4	154/168
CH063_15385	U, mito, no TMD	GLRG_11238	SP, extra, no TMD, no GPI	CE4 (*Ch*); CE4-CBM18 (*Cg*)	289/311
CH063_06940	NS, cyto, no TMD	GLRG_02688	SP, extra, no TMD, no GPI	CE4	480/540
CH063_11663	NS, cyto, no TMD	GLRG_01844	NS, cyto, noTMD	CE4	313/329
CH063_02556	NS, cysk, no TMD	GLRG_11102	NS, cysk	CE4	289/309
CH063_09392	NS, nucl, no TMD	NF		CE4	
NF		GLRG_11303	NS, cyto, noTMD	CE4	
Chitin binding protein without enzyme domain					
CH063_16082	SP, extra, no TMD	GLRG_06483	SP, extra, no TMD, no GPI	CBM18	154/185
CH063_09381	NS, nucl, no TMD	NF		CBM18-CBM18	
CH063_11773	SP, extra, no TMD, no GPI	NF		CBM18	
CH063_12050	NS, nucl, no TMD	NF		CBM18	
CH063_12226	NS, mito, TMD	NF		CBM18	
CH063_15028	NS, extra, no TMD	NF		CBM18	
NF		GLRG_10441	SP, extra, no TMD, no GPI	CBM18	
NF		GLRG_11588	SA, extra, TMD, no GPI	CBM18-CBM18	
Chitinase					
CH063_04609	SP, extra, no TMD, no GPI	GLRG_11925	SP, extra, no TMD, no GPI	GH18	379/914
CH063_13923	SP, extra, no TMD, no GPI	GLRG_03690	SP, extra, no TMD, no GPI	GH18-CBM18	271/368
CH063_14573	SP, extra, no TMD, no GPI	GLRG_05560	SA, extra, TMD	GH18	383/448
CH063_00825	SP, extra, no TMD, no GPI	GLRG_01574	NS, extra, no TMD	GH18	336/374
CH063_10854	SP, extra, no TMD, GPI ND	GLRG_02468	SP, extra, TMD, no GPI	GH18	306/350
CH063_13769	SP, extra, no TMD, no GPI	GLRG_01873	SP, extra, no TMD, no GPI	GH18	288/391
CH063_04207	U, extra, no TMD	GLRG_05986	SP, extra, no TMD, no GPI	GH18 (*Ch*); GH18-CBM18 (*Cg*)	100/126

CH063_06248	NS, cyto, no TMD	GLRG_07132	SP, extra, no TMD, no GPI	GH18	299/368
CH063_16029	NS, extra, no TMD	GLRG_05332	NS, extra, no TMD	GH18	257/307
CH063_09205	NS, mito, no TMD	GLRG_06660	SP, extra, no TMD, no GPI	GH18-CBM18	342/406
CH063_00260	NS, nucl, no TMD	GLRG_07794	NS, mito, no TMD	GH18-CBM18	445/1267
CH063_00260[c]	(NS, nucl, no TMD)	GLRG_05750	NS, cyto, no TMD	GH18-CBM18	161/534
CH063_13332	NS, mito, no TMD	GLRG_05960	NS, mito, no TMD, no GPI	GH18	321/342
CH063_15235	NS, mito, no TMD	GLRG_00384	NS, mito, no TMD	GH18	292/347
CH063_06857	NS, mito, no TMD	GLRG_11295	NS, mito, no TMD	GH18	301/338
CH063_02428	SP, extra, no TMD, no GPI	NF		GH18	
CH063_06344	SP, cyto, no TMD, no GPI	NF		GH18	
CH063_07095	NS, extr, no TMD	NF		GH18	
CH063_13275	NS, extra, no TMD	NF		GH18	
CH063_04496	NS, cyto, no TMD	NF		GH18	
CH063_11931	NS, mito, TMD	NF		GH18	
CH063_13902	NS, cyto, no TMD	NF		GH18	
CH063_06428	NS, cyto/nucl, no TMD	NF		GH18	
CH063_08635	NS, nucl, no TMD	NF		GH18	
NF		GLRG_06994	SP, extra, TMD, no GPI	GH18	
NF		GLRG_11080	SP, extra, no TMD, no GPI	GH18	
NF		GLRG_00924	NS, cyto, no TMD	GH18-CBM18	
NF		GLRG_09005	NS, cysk, no TMD	GH18	
NF		GLRG_02891	NS, mito, no TMD	GH18	
NF		GLRG_11796	NS, cyto, no TMD	GH18	

[a]Broad Institute identification number (http://www.broadinstitute.org/annotation/genome/colletotrichum_group/MultiHome.html). NF, not found.

[b]Predictions by SignalP: either SP, signal peptide, SA, signal anchor, or NS, non-secreted protein; the sequence categorized as U is missing the N terminus. Predictions by WoLF PSORT: extra, extracellular; plas, plasmamembrane; cyto, cytoplasm; cysk, cytoskeleton; mito, mitochondria; nucl, nucleus. Predictions by TMHMM for transmembrane helices: TMD, has >1 transmembrane helix. GPI lipid anchor prediction by big-PI Fungal Predictor was only used for putatively secreted proteins.

[c]CH063_00260 is an ortholog to two *C. graminicola* genes

Supplementary Table 15. Number of genes significantly regulated in pair-wise comparisons between samples (p<0.05; log2 fold-change>2).

a. *C. higginsianum*

Ratio	VA vs.	PA vs.	BP vs.	NP vs.
VA		1,515	2,251	2,690
PA	773		711	1,380
BP	1,425	377		904
NP	2,961	2,702	2,003	

VA: *in vitro* appressoria, PA: *in planta* appressoria, BP: biotrophic phase, NP: necrotrophic phase

b. *C. graminicola*

Ratio	PA vs.	BP vs.	NP vs.
PA		379	964
BP	440		615
NP	779	255	

PA: *in planta* appressoria, BP: biotrophic phase, NP: necrotrophic phase

Nature Genetics: doi:10.1038/ng.2372

Supplementary Table 16. 100 most highly induced *C. higginsianum* genes (p<0.05) in (**a**) *in planta* appressoria (PA) vs. *in vitro* appressoria (VA), (**b**) biotrophic phase (BP) vs. *in planta* appressoria (PA) and (**c**) necrotrophic phase (NP) vs. biotrophic phase (BP).

a. 100 most highly induced *C. higginsianum* genes *in planta* appressoria (PA) vs. *in vitro* appressoria (VA)

	Normalised read counts				Expression ratio (log2)		
Gene ID[a]	**VA**	**PA**	**BP**	**NP**	**PA/VA**	***P*-value**	**Description**
CH063_01084	1	125,050	114,092	13	17.2	1.5E-206	hypothetical protein (CSEP)
CH063_01083	1	29,872	20,205	1	15.8	8.7E-169	hypothetical protein
CH063_13925	0	15,891	15,547	2,005	15.3	1.9E-152	FAD binding domain-containing protein
CH063_14473	1	18,938	1,183	0	15.2	6.0E-157	hypothetical protein
CH063_13531	1	20,358	26,394	30	14.7	7.8E-159	hypothetical protein (CSEP)
CH063_04815	0	8,687	7,696	33	14.4	1.1E-136	DN3 putative cell death supressor
CH063_02506	0	5,683	720	5	13.8	1.0E-125	polyketide synthase
CH063_10755	1	8,241	6,385	1,125	13.1	2.4E-135	reducing polyketide synthase
CH063_00313	0	3,257	475	0	13.0	2.7E-111	ga4 desaturase family protein
CH063_03929	0	3,248	2,566	465	13.0	4.5E-111	polyketide synthase
CH063_09969	0	2,989	2,014	2	12.9	7.3E-109	secreted in xylem 6 effector protein
CH063_03928	3	19,290	23,308	2,622	12.8	1.1E-150	nonreducing polyketide synthase
CH063_08600	0	2,849	4,166	1	12.8	7.1E-108	short-chain dehydrogenase
CH063_09266	1	5,714	3,371	192	12.8	9.2E-126	FAD binding domain-containing protein
CH063_00051	0	2,580	2,936	2	12.7	2.8E-105	hypothetical protein
CH063_08811	0	2,656	3,847	1	12.7	4.5E-106	cytochrome p450 oxidoreductase
CH063_10362	1	3,545	6,749	1,449	12.7	1.9E-113	DJ-1/PfpI family protein
CH063_01082	1	3,265	2,425	1	12.6	2.9E-111	hypothetical protein
CH063_08042	0	2,391	1,903	907	12.6	3.2E-103	hypothetical protein (CSEP)
CH063_07865	1	3,107	8,331	4,953	12.5	4.4E-110	hypothetical protein
CH063_12562	3	14,831	11,368	1,479	12.5	4.0E-146	O-methyltransferase
CH063_04814	0	2,085	633	33	12.4	9.8E-100	cytosolic cu zn superoxide dismutase
CH063_08943	0	2,120	261	0	12.4	2.7E-100	ferric reductase transmembrane component 4
CH063_07049	1	5,315	8,019	6	12.3	2.9E-122	cytochrome p450 monooxygenase
CH063_08439	0	1,858	2,500	0	12.2	9.5E-97	hypothetical protein
CH063_01132	1	5,624	5,757	2,827	12.1	2.2E-125	hypothetical protein
CH063_00311	0	1,631	224	0	12.0	1.3E-93	short-chain dehydrogenase
CH063_00312	0	1,602	215	0	12.0	4.1E-93	hypothetical protein
CH063_00314	1	3,826	457	0	12.0	2.1E-115	polyamine transporter TPO1-like (MFS)
CH063_03737	0	1,663	1,079	1	12.0	7.3E-94	hypothetical protein
CH063_11682	0	1,591	4,419	7	12.0	8.8E-93	2og-fe oxygenase family oxidoreductase
CH063_00510	1	2,968	3,760	2	11.8	7.3E-109	toxin biosynthesis cytochrome p450
CH063_11846	4	15,964	19,415	3,111	11.8	6.3E-142	beta-carotene 15,15'-monooxygenase
CH063_12132	1	2,906	3,810	395	11.8	2.9E-108	polyketide synthase
CH063_04553	0	1,283	7,477	1,426	11.7	1.8E-87	alpha-ketoglutarate
CH063_07050	1	1,768	2,626	4	11.7	1.5E-95	polyketide synthase
CH063_11453	0	1,308	746	14	11.7	6.8E-88	mfs transporter
CH063_00511	1	3,673	6,169	4	11.6	2.2E-114	polyketide synthase
CH063_09159	0	1,196	2,623	3	11.6	1.4E-85	homocysteine synthase
CH063_07047	0	1,173	1,810	1	11.5	3.6E-85	cytochrome p450
CH063_01907	1	2,887	1,772	0	11.4	2.9E-106	hypothetical protein
CH063_03741	0	1,044	965	1	11.4	4.0E-82	hypothetical protein
CH063_07048	0	1,076	1,700	1	11.4	6.6E-83	citrinin biosynthesis oxidoreductase
CH063_11179	1	1,444	2,076	163	11.4	2.2E-90	zinc-binding dehydrogenase family
CH063_12567	0	1,087	1,635	1	11.4	3.5E-83	hypothetical protein
CH063_01906	1	1,357	804	0	11.3	1.2E-88	benzoate 4-monooxygenase cytochrome p450
CH063_10861	0	1,021	118	1	11.3	1.3E-81	polyketide synthase
CH063_01448	6	13,568	5,471	1	11.2	1.2E-137	tryptophan dimethylallyltransferase
CH063_12106	0	963	1,829	5	11.2	4.5E-80	cytochrome p450
CH063_12108	1	1,222	2,596	6	11.2	4.6E-86	o-acetylhomoserine o-acetylserine sulfhydrylase
CH063_13060	2	5,620	4,955	805	11.2	4.0E-122	reducing polyketide synthase
CH063_00336	0	842	642	0	11.1	1.9E-76	hypothetical protein
CH063_03740	0	865	716	1	11.1	3.0E-77	hypothetical protein

CH063_15744	0	865	715	1	11.1	3.0E-77	hypothetical protein
CH063_02970	0	764	270	10	10.9	2.2E-74	integral membrane protein
CH063_00534	0	718	459	26	10.8	8.6E-73	thioesterase domain containing protein
CH063_03738	1	2,472	1,701	1	10.8	2.5E-102	conserved hypothetical protein
CH063_00531	1	1,936	1,146	57	10.7	6.1E-98	polyketide synthase
CH063_08812	0	652	924	2	10.7	2.0E-70	toxin efflux pump (MFS)
CH063_12131	0	662	736	60	10.7	9.4E-71	non-ribosomal peptide synthetase
CH063_00535	1	1,698	994	31	10.6	1.5E-94	hypothetical protein
CH063_02504	0	563	84	0	10.5	7.2E-67	FAD binding domain-containing protein
CH063_12448	0	566	515	57	10.5	7.6E-67	non-ribosomal peptide synthetase
CH063_10640	0	544	520	9	10.4	6.7E-66	glutathionylspermidine synthase
CH063_03372	1	823	193	0	10.3	4.1E-76	mfs multidrug transporter
CH063_03641	1	985	231	0	10.3	1.6E-80	vacuolar 12-TM heavy metal transporter (ABC)
CH063_08043	0	473	399	75	10.2	1.8E-62	nad dependent epimerase dehydratase
CH063_09264	6	6,989	4,718	875	10.2	1.5E-119	extradiol ring-cleavage dioxygenase
CH063_00310	0	435	57	0	10.1	9.9E-61	ga4 desaturase family protein
CH063_00533	0	428	274	23	10.1	3.2E-60	hypothetical protein
CH063_05878	12	13,458	6,874	0	10.1	2.7E-126	cytochrome p450
CH063_08044	0	436	367	82	10.1	1.1E-60	multidrug transporter (MFS)
CH063_00532	0	417	324	20	10.0	1.1E-59	hypothetical protein
CH063_08464	0	398	703	1	10.0	4.4E-58	cytochrome p450
CH063_03373	0	388	99	0	9.9	8.4E-58	hypothetical protein
CH063_03735	1	633	487	1	9.9	1.4E-69	subtilisin-like protease (S08A)
CH063_13622	1	1,290	13,673	1,926	9.9	5.7E-86	hypothetical protein (CSEP)
CH063_00046	1	429	5,331	3	9.7	1.5E-59	hypothetical protein
CH063_11681	0	320	1,118	2	9.7	4.0E-53	hypothetical protein
CH063_03921	3	2,038	1,099	0	9.6	7.6E-96	nonribosomal peptide synthase
CH063_08639	0	302	145	2	9.6	7.5E-52	short-chain dehydrogenase reductase family
CH063_10944	1	409	4,379	344	9.6	1.9E-59	nudix domain-containing protein
CH063_11733	4	3,287	4,503	1	9.6	9.3E-104	DN3 putative cell death supressor
CH063_02502	0	292	28	0	9.5	3.2E-51	hypothetical protein
CH063_00050	0	263	846	0	9.4	1.4E-48	hypothetical protein
CH063_01081	2	1,633	813	1	9.4	2.8E-90	vacuolar 12-TM heavy metal transporter (ABC)
CH063_05927	0	276	286	36	9.4	7.6E-50	3-isopropylmalate dehydrogenase
CH063_06770	1	336	1,247	550	9.3	2.7E-54	drug resistance transporter (MFS)
CH063_02137	0	234	1,020	4	9.2	3.0E-46	putative transporter (MFS)
CH063_05929	1	388	329	40	9.2	9.0E-58	hypothetical protein
CH063_08303	0	228	2,293	274	9.2	2.7E-45	cytochrome p450 oxidoreductase
CH063_10865	1	315	38	0	9.2	5.1E-53	hypothetical protein
CH063_00048	0	214	1,013	1	9.1	6.6E-44	hypothetical protein
CH063_01047	0	224	693	3	9.1	5.4E-45	benzoate 4-monooxygenase cytochrome p450
CH063_02500	1	356	56	0	9.1	6.7E-56	multicopper oxidase
CH063_03286	0	216	2,156	193	9.1	8.7E-45	nudix domain-containing protein
CH063_07110	1	737	382	0	9.1	8.3E-72	alpha beta hydrolase
CH063_10662	10	5,617	2,805	86	9.1	1.9E-106	hypothetical protein
CH063_11725	0	222	240	27	9.1	1.2E-44	carboxypeptidase (S10)
CH063_15163	0	217	128	9	9.1	2.8E-44	hypothetical protein

[a]Broad Institute identification number (http://www.broadinstitute.org/annotation/genome/colletotrichum_group/MultiHome.html).

b. 100 most highly induced *C. higginsianum* genes in biotrophic phase (BP) vs. *in planta* appressoria (PA)

	Normalised read counts				Expression ratio (log2)		
Gene ID[a]	**VA**	**PA**	**BP**	**NP**	**BP/PA**	***P*-value**	**Description**
CH063_06094	1	3	696	716	7.6	2.2E-56	alcohol dehydrogenase
CH063_09569	0	1	221	7,749	7.5	7.6E-37	hypothetical protein
CH063_00788	1	1	217	5,049	7.2	1.9E-36	conserved hypothetical protein
CH063_05963	0	1	127	417	6.4	1.0E-26	hypothetical protein
CH063_04499	0	1	100	1,076	6.1	5.7E-23	hypothetical protein (CSEP)
CH063_07921	1	1	65	239	5.8	5.8E-18	radh flavin-dependent halogenase
CH063_07765	1	1	65	2,205	5.8	7.8E-18	cell wall glycosyl hydrolase (GH105)
CH063_08831	1	1	64	577	5.7	1.9E-17	napsin a aspartic peptidase (A01A)
CH063_05787	0	2	104	2,016	5.7	1.2E-23	hypothetical protein

CH063_00987	5	2	103	4,598	5.7	2.7E-22	cell wall glycosyl hydrolase (GH105)
CH063_11299	1	4	162	606	5.5	2.3E-27	gmc oxidoreductase
CH063_00789	0	7	317	7,509	5.4	9.3E-36	conserved hypothetical protein
CH063_04375	1	1	51	134	5.4	1.1E-14	zinc-binding dehydrogenase family
CH063_05657	0	25	1,048	145	5.4	2.8E-49	hypothetical protein
CH063_01323	0	2	80	2,129	5.3	1.9E-19	alkaline serine protease (S08A)
CH063_09735	113	12	454	7,108	5.3	1.4E-39	hypothetical protein
CH063_11272	1	2	91	271	5.3	1.7E-20	lovastatin nonaketide synthase
CH063_09960	0	1	44	223	5.2	5.1E-13	hypothetical protein
CH063_14294	5	7	231	1,367	5.1	5.5E-31	hypothetical protein (CSEP)
CH063_09104	0	22	764	6,911	5.1	1.1E-43	FMN-dependent dehydrogenase
CH063_10447	0	1	40	2	5.1	4.6E-12	hypothetical protein (CSEP)
CH063_12770	0	1	46	51	5.0	3.6E-13	polyketide synthase
CH063_03242	0	2	64	382	5.0	1.8E-15	subtilisin-like protease (S08A)
CH063_11281	0	5	144	948	5.0	8.9E-24	periplasmic beta-glucosidase beta-xylosidase (GH3)
CH063_14434	4	3	88	6	5.0	4.8E-18	hypothetical protein
CH063_07093	0	3	107	938	4.9	2.4E-21	hypothetical protein
CH063_15802	0	4	123	19	4.9	6.8E-22	cytochrome p450
CH063_12564	1	6	194	724	4.9	2.2E-28	amine flavin-containing superfamily
CH063_03128	0	5	138	152	4.9	4.3E-23	hypothetical protein (CSEP)
CH063_07918	1	14	381	2,613	4.8	4.4E-34	serin endopeptidase (S08A)
CH063_03704	0	2	54	91	4.7	2.7E-14	cytochrome p450 monooxygenase
CH063_15781	0	3	68	295	4.7	3.2E-16	hypothetical protein
CH063_14852	1	2	41	577	4.7	1.6E-11	hypothetical protein
CH063_09959	0	1	32	115	4.7	1.0E-09	hypothetical protein
CH063_05684	0	3	82	342	4.7	4.0E-18	polyketide synthase
CH063_00026	0	1	38	2,185	4.7	9.7E-12	quinate transporter (MFS)
CH063_03405	1	1	30	17	4.7	1.0E-09	hypothetical protein
CH063_12502	2	4	103	133	4.7	4.2E-19	pectate lyase (PL3)
CH063_13286	1	1	36	25	4.6	9.3E-11	homeobox domain-containing protein
CH063_11613	1	2	38	109	4.6	1.4E-10	fad-dependent monooxygenase
CH063_15832	1	2	56	2	4.6	6.5E-14	sugar isomerase
CH063_01629	1	8	193	4	4.6	1.1E-25	hypothetical protein
CH063_13603	0	2	36	255	4.5	4.1E-10	hypothetical protein
CH063_08273	0	3	59	2	4.5	1.7E-13	hypothetical protein
CH063_08573	4	5	103	144	4.5	2.4E-18	1-aminocyclopropane-1-carboxylate deaminase
CH063_13146	2	101	2,238	30,539	4.5	6.1E-44	carbonic anhydrase
CH063_08733	1	3	69	2,404	4.5	1.8E-15	cell wall glycosyl hydrolase (GH105)
CH063_04316	1	20	452	5	4.5	6.2E-34	hypothetical protein
CH063_13750	0	6	142	13	4.5	3.9E-22	glycoside hydrolase family (GH16)
CH063_13023	2	47	1,037	1,412	4.5	5.7E-40	domain-containing protein (CBM50-CBM50)
CH063_11355	0	1	32	107	4.4	4.4E-10	polyketide synthase
CH063_09152	1	3	68	362	4.4	8.1E-15	bifunctional catalase-peroxidase cat2
CH063_06129	19	16	344	1,323	4.4	8.0E-31	pectate lyase (PL9)
CH063_11943	0	7	146	279	4.4	1.2E-22	pyoverdine dityrosine biosynthesis
CH063_05961	0	1	26	126	4.4	9.1E-08	hypothetical protein
CH063_02659	9	15	322	27	4.4	1.8E-30	glucan -beta-glucosidase
CH063_05163	0	3	55	1,345	4.4	5.7E-13	endo-beta-xylanase (GH43)
CH063_04001	1	5	113	4,007	4.4	7.8E-20	hypothetical protein
CH063_01004	0	6	125	110	4.4	3.3E-20	hypothetical protein (CSEP)
CH063_14495	0	5	107	31	4.4	3.1E-18	hypothetical protein
CH063_04757	0	34	689	388	4.3	7.3E-36	hypothetical protein (CSEP)
CH063_04107	1	11	228	409	4.3	3.9E-26	hypothetical protein (CSEP)
CH063_07516	1	1	24	712	4.3	1.4E-07	endonuclease exonuclease phosphatase
CH063_14240	1	2	35	894	4.3	5.6E-09	exopolygalacturonase (GH28)
CH063_03469	1	3	50	34	4.3	1.1E-12	trichothecene C-15 hydroxylase
CH063_02153	21	362	7,102	886	4.3	3.3E-44	hypothetical protein (CSEP)
CH063_13004	0	3	68	498	4.3	7.2E-14	xylan-beta-xylosidase (GH3)
CH063_14677	0	1	23	479	4.3	2.1E-07	aminopeptidase y (M28A)
CH063_04611	6	4	68	1	4.3	7.2E-14	alkaline serine protease (S08A)
CH063_12273	0	1	28	127	4.2	1.1E-08	hypothetical protein
CH063_13466	1	7	133	203	4.2	1.5E-20	geranylgeranyl pyrophosphate synthase

CH063_15598	1	1	28	807	4.2	1.1E-08	exopolygalacturonase (GH28)
CH063_11640	0	6	109	185	4.2	2.6E-19	hypothetical protein
CH063_10492	4	16	298	10	4.2	4.5E-28	hypothetical protein
CH063_10353	1	33	596	40	4.2	1.7E-33	hypothetical protein (CSEP)
CH063_15468	1	10	187	223	4.2	4.4E-23	hypothetical protein
CH063_01499	1	54	956	137	4.2	2.0E-36	hypothetical protein (CSEP)
CH063_15180	1	6	107	580	4.2	8.2E-18	beta-glucosidase 2 (GH3)
CH063_02158	2	36	641	580	4.1	7.4E-34	hypothetical protein
CH063_07488	0	4	65	107	4.1	4.2E-14	hypothetical protein (CSEP)
CH063_10341	0	2	31	233	4.1	1.8E-08	hypothetical protein
CH063_02054	0	178	3,128	1,240	4.1	1.8E-40	hypothetical protein (CSEP)
CH063_11996	1	25	437	24	4.1	8.4E-31	monooxygenase
CH063_04538	6	70	1,212	646	4.1	7.7E-37	hypothetical protein
CH063_14119	2	3	60	28	4.1	2.6E-12	putative polyamin transporter (MFS)
CH063_05510	1	3	54	186	4.1	1.1E-11	pectate lyase (PL1)
CH063_01399	2	198	3,412	2,922	4.1	1.8E-40	hypothetical protein (CSEP)
CH063_10359	1	8	129	11	4.1	1.9E-19	glycoside hydrolase family 16 (GH16)
CH063_06485	1	15	264	10,416	4.1	1.7E-26	quinate transporter (MFS)
CH063_03960	12	4	63	14	4.1	4.8E-13	hypothetical protein (CSEP)
CH063_02114	0	22	376	1,066	4.1	3.3E-29	hypothetical protein
CH063_00219	0	9	147	192	4.1	1.6E-21	hypothetical protein
CH063_04772	0	2	40	198	4.1	2.3E-10	polyketide synthase
CH063_10748	1	13	222	2	4.1	9.6E-25	hypothetical protein
CH063_02843	1	8	142	2,361	4.1	6.2E-21	hypothetical protein
CH063_01169	2	25	423	281	4.1	3.9E-30	hypothetical protein
CH063_14178	0	14	235	209	4.1	5.4E-25	alpha-ketoglutarate dependent xanthine dioxygenase
CH063_04647	2	37	610	26	4.0	1.3E-32	necrosis inducing protein
CH063_13692	1	1	20	13	4.0	1.6E-07	hypothetical protein
CH063_15857	3	10	157	107	4.0	3.3E-21	hypothetical protein

[a]Broad Institute identification number (http://www.broadinstitute.org/annotation/genome/colletotrichum_group/MultiHome.html).

c. 100 most highly induced *C. higginsianum* genes in necrotrophic phase (NP) vs. biotrophic phase (BP)

	Normalised read counts				Expression ratio (log2)		
Gene ID[a]	**VA**	**PA**	**BP**	**NP**	**NP/BP**	***P*-value**	**Description**
CH063_04608	0	1	1	131	7.2	1.3E-36	LysM domain-containing protein
CH063_03114	0	2	1	156	7.1	1.2E-38	choline dehydrogenase
CH063_10976	0	1	4	506	6.9	2.5E-58	pectate lyase (PL3)
CH063_00965	5	26	63	5,180	6.4	8.2E-78	Na+/Pi cotransporter (PiT)
CH063_07719	0	1	1	90	6.2	1.4E-28	beta-galactosidase (GH2)
CH063_08917	0	1	1	84	5.9	1.8E-26	LysM domain-containing protein (CBM50-CBM50)
CH063_01117	1	3	1	70	5.9	1.4E-24	sugar hexose transporter (MFS)
CH063_00988	1	1	3	176	5.9	5.2E-36	hypothetical protein
CH063_05318	0	1	11	649	5.9	1.7E-55	periplasmic beta-glucosidase beta-xylosidase (GH3)
CH063_15707	0	1	4	223	5.9	2.0E-40	lipolytic enzyme (CBM1)
CH063_03488	0	1	1	51	5.9	9.9E-22	pectate lyase (PL3)
CH063_00026	0	1	38	2,185	5.8	7.2E-66	quinate transporter (MFS)
CH063_06211	0	1	1	67	5.8	8.3E-24	pectate lyase (PL1)
CH063_02546	7	3	16	876	5.8	2.0E-57	pectate lyase (PL1)
CH063_16097	0	1	1	46	5.7	9.0E-21	hypothetical protein
CH063_15004	1	1	4	218	5.7	1.3E-39	high affinity nicotinic acid transporter
CH063_05959	0	1	3	169	5.7	1.7E-34	alpha amylase (GH13)
CH063_14530	0	1	1	74	5.6	1.4E-23	hypothetical protein (CSEP)
CH063_00255	0	2	8	376	5.6	2.2E-45	integral membrane protein
CH063_00987	5	2	103	4,598	5.5	3.7E-64	cell wall glycosyl hydrolase (GH105)
CH063_03563	7	25	76	3,350	5.5	4.6E-63	general amino acid permease (APC)
CH063_04205	1	1	4	186	5.5	5.6E-35	methyltransferase domain-containing protein
CH063_12616	0	1	1	39	5.5	6.7E-18	hypothetical protein
CH063_12650	8	3	28	1,226	5.5	5.6E-57	hypothetical protein
CH063_10011	1	1	7	302	5.4	3.1E-41	serine endopeptidase (S8A)
CH063_15801	1	1	4	160	5.3	2.6E-32	LysM domain-containing protein
CH063_06485	1	15	264	1,0416	5.3	3.6E-63	sugar or quinate transporter (MFS)

CH063_13381	0	1	3	125	5.3	3.0E-30	pectate lyase (PL3)
CH063_07571	1	5	3	114	5.3	1.9E-29	histidine acid phosphatase
CH063_13093	2	3	10	393	5.3	1.1E-43	conserved hypothetical protein
CH063_14348	1	1	2	94	5.2	1.1E-24	integral membrane protein
CH063_00975	1	13	40	1,484	5.2	1.3E-55	putative D-galactonate transporter (MFS)
CH063_01116	0	1	10	359	5.2	2.4E-42	hypothetical protein
CH063_04323	3	11	11	413	5.2	4.1E-44	LysM domain-containing protein
CH063_11211	1	3	1	33	5.2	1.2E-15	hypothetical protein
CH063_14035	0	1	3	99	5.2	1.2E-25	hypothetical protein
CH063_07181	4	14	3	103	5.2	2.3E-27	Integral membrane peptide transporter (POT)
CH063_04001	1	5	113	4,007	5.1	5.8E-59	hypothetical protein
CH063_00570	0	2	18	646	5.1	4.7E-48	hypothetical protein (CSEP)
CH063_09569	0	1	221	7,749	5.1	4.3E-60	hypothetical protein
CH063_08733	1	3	69	2,404	5.1	7.9E-57	cell wall glycosyl hydrolase (GH105)
CH063_02950	2	1	1	30	5.1	5.2E-15	creatine transporter
CH063_08334	0	1	1	30	5.1	1.3E-14	hypothetical protein
CH063_05118	0	1	4	134	5.1	2.6E-29	carotenoid ester lipase precursor
CH063_01708	1	1	4	134	5.1	2.1E-29	glycosyl hydrolase (GH47)
CH063_11834	261	8	2	52	5.1	6.1E-18	3-phytase
CH063_01115	1	3	37	1,250	5.1	9.6E-53	hypothetical protein
CH063_07765	1	1	65	2,205	5.1	5.0E-56	cell wall glycosyl hydrolase (GH105)
CH063_07368	1	13	6	202	5.1	1.8E-34	pectate lyase (PL3)
CH063_08335	0	1	1	47	5.1	1.4E-17	hypothetical protein
CH063_09484	169	272	813	27,136	5.1	5.5E-60	P-type Na+-ATPase (P-ATPase)
CH063_02718	0	1	3	94	5.0	2.1E-25	glucose oxidase
CH063_04580	0	3	23	757	5.0	2.7E-48	hypothetical protein
CH063_03257	0	2	10	325	5.0	8.1E-40	metalloproteinase (M35)
CH063_13346	0	2	5	152	5.0	5.5E-31	putative transporter
CH063_08940	2	9	63	2,026	5.0	3.3E-54	hypothetical protein
CH063_12909	0	1	1	28	5.0	5.7E-14	hypothetical protein
CH063_13094	1	23	28	875	5.0	6.3E-49	putative D-galactonate transporter (MFS)
CH063_12401	6	2	3	100	5.0	1.7E-25	hypothetical protein (CSEP)
CH063_01837	1	12	19	575	4.9	8.4E-45	FAD binding domain-containing protein
CH063_07516	1	1	24	712	4.9	2.1E-46	endonuclease exonuclease phosphatase
CH063_06886	0	1	2	50	4.9	1.2E-18	subtilisin (S08A)
CH063_03707	1	1	2	46	4.9	1.9E-18	hypothetical protein
CH063_07889	1	1	1	41	4.9	5.9E-16	pectate lyase (PL3)
CH063_04418	4	2	1	34	4.9	7.5E-15	aldehyde dehydrogenase
CH063_06916	1	3	1	26	4.9	1.4E-13	serine endopeptidase (S8A)
CH063_15598	1	1	28	807	4.9	4.1E-47	exopolygalacturonase (GH28)
CH063_13607	0	1	3	82	4.8	8.5E-24	zinc carboxypeptidase (M14A)
CH063_13808	0	1	3	87	4.8	2.5E-23	pectate lyase (PL9)
CH063_09591	10	9	17	480	4.8	7.6E-42	nudix family protein
CH063_15582	0	1	5	150	4.8	2.9E-29	pectate lyase (PL3)
CH063_01626	14	25	38	1,045	4.8	4.0E-48	putative Hexose transporter (MFS)
CH063_00353	0	1	4	97	4.8	1.2E-24	pectate lyase (PL1)
CH063_08373	1	2	7	192	4.8	5.2E-32	gmc oxidoreductase
CH063_01323	0	2	80	2,129	4.7	8.2E-51	alkaline serine protease (S08A)
CH063_11130	0	1	1	31	4.7	3.7E-13	leucyl aminopeptidase (M28A)
CH063_13696	3	5	23	598	4.7	3.3E-43	glutamate carboxypeptidase (M28B)
CH063_14240	1	2	35	894	4.7	9.4E-46	exopolygalacturonase (GH28)
CH063_03871	1	3	24	595	4.7	1.7E-42	integral membrane protein
CH063_04129	2	1	10	247	4.6	3.1E-34	cellulose-binding family protein (CBM1)
CH063_15079	10	5	4	90	4.6	5.9E-23	cell wall glucanosyltransferase (GH16)
CH063_10254	0	1	1	21	4.6	1.1E-11	hypothetical protein (CSEP)
CH063_05163	0	3	55	1,345	4.6	3.0E-47	endo-beta-xylanase (GH43-GH43)
CH063_10719	2	2	5	123	4.6	4.6E-27	high-affinity nicotinic acid transporter (MFS)
CH063_14413	0	1	1	27	4.6	2.0E-12	hypothetical protein
CH063_00789	0	7	317	7,509	4.6	8.6E-51	conserved hypothetical protein
CH063_08376	0	1	7	160	4.6	1.4E-29	cellobiose dehydrogenase
CH063_05587	0	1	5	107	4.5	5.2E-25	chlorogenic acid esterase precursor
CH063_01213	154	18	18	426	4.5	9.9E-39	pectate lyase (PL1)

CH063_00788	1	1	217	5,049	4.5	6.3E-50	conserved hypothetical protein
CH063_02625	0	3	10	242	4.5	1.6E-33	beta-galactosidase (GH35)
CH063_04581	6	2	25	577	4.5	3.9E-41	hypothetical protein (CSEP)
CH063_14192	6	5	1	26	4.5	6.8E-12	fad-dependent oxygenase
CH063_11903	1	2	9	199	4.5	7.8E-31	hypothetical protein
CH063_08354	1	15	27	618	4.5	4.6E-41	mfs transporter
CH063_02790	1	3	12	260	4.5	7.0E-34	endoribonuclease
CH063_09636	0	1	1	33	4.5	5.4E-13	hypothetical protein
CH063_14109	1	1	8	180	4.5	6.3E-30	pectinesterase (CE8-PL1)
CH063_15656	1	4	3	62	4.5	2.7E-19	trehalose transporter
CH063_06953	0	1	3	66	4.5	4.1E-19	pectate lyase (PL3)

[a]Broad Institute identification number (http://www.broadinstitute.org/annotation/genome/colletotrichum_group/MultiHome.html).

Supplementary Table 17. 100 most highly induced *C. graminicola* genes (p<0.05) in (**a**) biotrophic phase (BP) vs. *in planta* appressoria (PA) and (**b**) necrotrophic phase (NP) vs. biotrophic phase (BP).

a. 100 most highly induced *C. graminicola* genes in biotrophic phase (BP) vs. *in planta* appressoria (PA)

	Normalised read counts			Expression ratio (log2)		
Gene ID[a]	PA	BP	NP	BP/PA	P-value	Description
GLRG_05319	0	332	698	7.9	2.7E-20	subtilase
GLRG_06511	0	337	657	7.9	2.8E-20	metalloprotease
GLRG_09294	2	349	229	6.9	2.2E-19	hypothetical protein
GLRG_07653	2	283	1,384	6.9	2.2E-18	hypothetical protein
GLRG_06285	0	162	651	6.8	3.5E-15	peptidase family M28
GLRG_10439	17	1,716	3,651	6.8	1.4E-27	fungal cellulose binding domain-containing protein
GLRG_07845	1	155	370	6.4	2.2E-14	tat pathway signal sequence
GLRG_04091	1	155	560	6.4	2.8E-14	peptidase family M28
GLRG_02172	0	121	125	6.4	1.1E-12	hypothetical protein
GLRG_03688	10	818	1,824	6.3	5.3E-23	hypothetical protein
GLRG_00048	2	294	805	6.3	1.1E-17	methyltransferase domain-containing protein
GLRG_01850	1	139	344	6.2	2.0E-13	hypothetical protein
GLRG_11551	0	107	788	6.2	1.3E-11	fungal cellulose binding domain-containing protein
GLRG_01903	1	183	456	6.0	1.4E-14	hypothetical protein
GLRG_07698	3	285	708	6.0	2.7E-17	peptidase family M28
GLRG_06249	2	190	274	5.9	2.2E-14	aldo/keto reductase
GLRG_01430	1	113	423	5.9	1.5E-11	methyltransferase domain-containing protein
GLRG_06286	13	573	1,547	5.8	2.8E-19	metalloprotease
GLRG_10440	3	134	212	5.8	2.1E-12	hypothetical protein
GLRG_06543	10	475	1,401	5.8	5.7E-19	fungalysin metallopeptidase
GLRG_10700	11	506	1,553	5.8	1.8E-18	glycosyl hydrolase family 61
GLRG_10056	0	77	130	5.8	8.7E-10	glycosyl Hydrolase Family 88
GLRG_01014	0	72	608	5.7	1.4E-09	high-affinity methionine permease
GLRG_01785	1	99	737	5.7	7.0E-11	hypothetical protein
GLRG_02084	7	289	240	5.7	1.1E-16	FMN-dependent dehydrogenase
GLRG_00979	1	97	293	5.7	8.8E-11	hypothetical protein
GLRG_00795	0	70	238	5.7	3.5E-09	glycosyl hydrolase family 3
GLRG_04432	0	77	415	5.7	3.5E-09	integral membrane protein
GLRG_11725	1	96	132	5.6	1.9E-10	pectate lyase
GLRG_02090	10	556	4,606	5.6	2.2E-19	hypothetical protein
GLRG_11593	0	65	350	5.6	6.6E-09	subtilase
GLRG_01005	0	70	554	5.6	4.8E-09	major facilitator superfamily transporter
GLRG_09426	7	287	282	5.5	8.2E-16	hypothetical protein
GLRG_01192	1	93	100	5.5	6.6E-10	hypersensitive response-inducing protein
GLRG_11729	2	107	419	5.5	2.7E-11	glycosyl hydrolase family 61
GLRG_05154	3	147	115	5.5	1.5E-12	ricin B lectin
GLRG_09293	49	1,634	1,165	5.4	7.0E-21	GMC oxidoreductase
GLRG_00455	7	264	108	5.4	1.6E-15	polyprenyl synthetase
GLRG_07173	8	292	716	5.3	1.0E-15	peptidase family M28
GLRG_10930	14	453	1,301	5.3	4.5E-17	subtilase
GLRG_00978	0	57	205	5.3	1.4E-07	integral membrane protein
GLRG_08992	5	167	691	5.3	1.7E-12	OPT oligopeptide transporter
GLRG_09837	3	99	551	5.1	2.7E-10	hypothetical protein
GLRG_09951	7	242	896	5.1	1.4E-14	major facilitator superfamily transporter
GLRG_06496	2	81	112	5.1	1.8E-09	phytase
GLRG_06274	3	118	606	5.1	8.8E-11	glycosyl hydrolase family 61
GLRG_05268	56	1,478	2,643	5.1	2.2E-19	methyltransferase domain-containing protein
GLRG_01028	0	48	457	5.1	6.3E-07	hypothetical protein
GLRG_10480	4	103	71	5.0	3.1E-10	deuterolysin metalloprotease
GLRG_09299	22	636	1091	5.0	2.7E-17	archaeal flagellin
GLRG_01075	3	106	197	5.0	5.1E-10	5'-nucleotidase domain-containing protein
GLRG_00232	0	42	204	4.9	2.7E-06	amino acid adenylation domain-containing protein
GLRG_03066	0	42	197	4.9	3.3E-06	hypothetical protein
GLRG_11302	2	76	165	4.9	1.2E-08	hypothetical protein

GLRG_06219	7	186	884	4.9	1.6E-12	zinc carboxypeptidase
GLRG_04806	0	42	276	4.9	4.1E-06	major facilitator superfamily transporter
GLRG_00454	16	418	152	4.9	1.1E-15	cytochrome P450
GLRG_06648	7	179	415	4.8	3.5E-12	peptidase family M28
GLRG_06247	2	96	146	4.8	6.4E-09	phosphate transporter
GLRG_05336	54	1,276	2,315	4.8	1.1E-17	hypothetical protein
GLRG_07971	0	38	40	4.8	7.8E-06	cytochrome P450
GLRG_00260	6	161	318	4.7	8.1E-12	fungal cellulose binding domain-containing protein
GLRG_01902	20	461	855	4.7	3.5E-15	hypothetical protein
GLRG_09541	5	183	40	4.7	1.4E-12	major facilitator superfamily transporter
GLRG_04762	27	757	1,208	4.7	3.7E-17	TfdA family Taurine catabolism dioxygenase TauD
GLRG_11600	1	44	62	4.7	2.8E-06	necrosis inducing protein
GLRG_10716	2	68	709	4.7	3.1E-08	potassium/sodium efflux P-type ATPase
GLRG_07794	1	51	135	4.7	2.8E-06	glycosyl hydrolase family 18
GLRG_11454	5	94	303	4.6	5.7E-09	sulfate permease
GLRG_07978	3	57	147	4.6	8.1E-07	C2 domain-containing protein
GLRG_07167	57	1,364	3,282	4.6	2.7E-17	TfdA family Taurine catabolism dioxygenase TauD
GLRG_05283	9	231	4,390	4.5	6.1E-13	hypothetical protein
GLRG_01889	2	54	255	4.5	6.1E-07	glycosyl hydrolase family 61
GLRG_09078	24	577	760	4.5	1.3E-15	high-affinity methionine permease
GLRG_01071	0	29	48	4.5	1.0E-04	endonuclease/Exonuclease/phosphatase
GLRG_07539	7	149	288	4.4	7.0E-11	arabinosidase
GLRG_00289	0	30	162	4.4	1.3E-04	parallel beta-helix repeat protein
GLRG_08241	12	253	700	4.4	7.5E-13	OPT oligopeptide transporter
GLRG_09274	2	53	103	4.4	2.8E-06	hypothetical protein
GLRG_01804	318	5,553	3,844	4.3	2.2E-16	hypothetical protein
GLRG_08924	1	57	235	4.3	1.6E-06	hypothetical protein
GLRG_11022	1	33	112	4.3	7.7E-05	hypothetical protein
GLRG_09371	1	36	140	4.3	2.4E-05	mechanosensitive ion channel
GLRG_10949	19	327	521	4.2	7.8E-13	short chain dehydrogenase
GLRG_09468	6	87	286	4.2	4.7E-08	glycosyl hydrolase family 76
GLRG_10476	6	80	193	4.2	1.8E-07	endonuclease/Exonuclease/phosphatase
GLRG_11299	2	46	51	4.2	7.6E-06	major facilitator superfamily transporter
GLRG_09952	19	399	1,254	4.2	1.2E-13	2OG-Fe(II) oxygenase superfamily protein
GLRG_07276	4	76	102	4.2	1.8E-07	hypothetical protein
GLRG_03485	1	33	83	4.2	4.8E-05	hypothetical protein
GLRG_09805	3	55	705	4.2	1.7E-06	amino acid permease
GLRG_03486	29	406	828	4.2	1.3E-12	hypothetical protein
GLRG_09704	0	24	105	4.1	1.0E-03	CVNH domain-containing protein
GLRG_10901	0	23	41	4.1	1.0E-03	aldo/keto reductase
GLRG_00994	167	2,727	1,881	4.1	4.8E-15	hypothetical protein
GLRG_10900	1	35	57	4.1	6.1E-05	dienelactone hydrolase
GLRG_10818	15	225	813	4.1	4.1E-11	hypothetical protein
GLRG_11874	0	26	364	4.1	1.3E-03	hypothetical protein
GLRG_01017	3	49	13	4.0	5.5E-06	glycosyl hydrolase family 71
GLRG_05524	4	105	1,296	4.0	9.2E-09	glycosyl hydrolase family 11

[a]Broad Institute identification number (http://www.broadinstitute.org/annotation/genome/colletotrichum_group/MultiHome.html).

b. 100 most highly induced *C. graminicola* genes in necrotrophic phase (NP) vs. biotrophic phase (BP)

	Normalised read counts			Expression ratio (log2)		
Gene ID[a]	**PA**	**BP**	**NP**	**NP/BP**	**P-value**	**Description**
GLRG_09602	15	0	628	9.4	2.37E-29	hypothetical protein
GLRG_09715	1	0	272	8.2	3.83E-22	beta-ketoacyl synthase domain-containing protein
GLRG_11446	3	0	339	8.1	1.35E-23	hypothetical protein
GLRG_08211	4	0	250	8.1	2.45E-21	GDSL-like Lipase/Acylhydrolase
GLRG_09215	1	1	325	8.0	7.29E-23	glycosyl hydrolase family 61
GLRG_09714	1	0	247	7.7	2.54E-21	alcohol dehydrogenase GroES-like
GLRG_09165	1	2	692	7.6	3.31E-28	glycosyl hydrolase family 61
GLRG_04764	3	0	156	7.4	1.01E-17	glycosyl hydrolase family 62

GLRG_08974	0	5	822	7.3	1.62E-27	glycosyl hydrolase family 7
GLRG_09681	1	2	235	7.3	1.05E-19	glycosyl hydrolase family 61
GLRG_01818	22	14	1,411	7.2	1.26E-27	linoleate diol synthase
GLRG_09214	1	0	116	7.0	3.27E-15	glycosyl hydrolase family 61
GLRG_03683	3	1	223	7.0	1.13E-19	hypothetical protein
GLRG_08534	3	0	137	6.8	3.39E-16	cellulase
GLRG_10074	7	1	293	6.8	2.53E-21	fungal hydrophobin
GLRG_09712	2	3	627	6.7	8.74E-26	alpha/beta hydrolase
GLRG_11951	4	2	216	6.7	1.30E-18	glycosyl hydrolase family 12
GLRG_11420	0	0	118	6.6	6.41E-15	hypothetical protein
GLRG_09732	0	0	79	6.4	3.88E-12	glycosyl hydrolase family 61
GLRG_00903	16	6	524	6.3	1.19E-22	hypothetical protein
GLRG_08518	0	1	95	6.3	8.59E-13	glycosyl hydrolase family 6
GLRG_01857	1	1	226	6.2	4.90E-19	glycosyl hydrolase family 11
GLRG_11469	3	0	85	6.1	1.10E-12	glycosyl hydrolase family 61
GLRG_05713	0	0	82	6.1	2.09E-12	RTA1 like protein
GLRG_02944	4	2	183	6.1	6.66E-17	glycosyl hydrolase family 45
GLRG_04553	1	0	78	6.0	5.17E-12	cytochrome P450
GLRG_00951	23	10	722	6.0	3.46E-23	glycosyl hydrolase family 6
GLRG_03376	4	1	186	6.0	2.56E-17	multicopper oxidase
GLRG_09713	1	0	54	5.9	8.96E-10	HypE protein
GLRG_01819	15	11	698	5.9	1.82E-22	hypothetical protein
GLRG_04126	5	4	289	5.8	4.16E-19	conidiation protein 6
GLRG_00775	2	0	51	5.8	1.56E-09	PHB depolymerase family esterase
GLRG_07406	1	2	213	5.8	7.54E-18	fungal cellulose binding domain-containing protein
GLRG_09977	0	0	49	5.8	3.25E-09	integral membrane protein
GLRG_11170	1	1	113	5.8	8.37E-14	hypothetical protein
GLRG_11038	0	3	190	5.7	2.67E-16	glycosyl hydrolase family 7
GLRG_08966	6	6	384	5.6	7.48E-20	glycosyl hydrolase family 10
GLRG_00823	1	0	43	5.6	2.76E-08	glycosyl hydrolase family 61
GLRG_10298	7	1	108	5.5	1.83E-13	hypothetical protein
GLRG_02945	0	1	133	5.5	1.97E-14	hypothetical protein
GLRG_07430	11	0	40	5.5	2.65E-08	cytochrome P450
GLRG_11201	7	13	686	5.5	5.29E-21	glycosyl hydrolase family 7
GLRG_10282	1	5	374	5.4	1.08E-19	AMP-binding enzyme
GLRG_05332	7	0	45	5.2	1.26E-08	glycosyl hydrolase family 18
GLRG_09709	11	8	307	5.1	5.73E-17	hypothetical protein
GLRG_04694	1	0	30	5.1	1.38E-06	hypothetical protein
GLRG_00811	1	0	30	5.1	1.82E-06	hypothetical protein
GLRG_09760	0	0	29	5.0	1.82E-06	hypothetical protein
GLRG_04805	2	0	39	5.0	9.70E-08	hypothetical protein
GLRG_10499	0	0	29	5.0	2.39E-06	cytochrome P450
GLRG_05915	8	4	169	5.0	1.97E-14	3-dehydroquinate synthase
GLRG_03065	0	0	28	5.0	2.76E-06	DJ-1/PfpI family protein
GLRG_11240	10	43	1,452	5.0	1.08E-19	glycosyltransferase family 2
GLRG_09749	15	2	143	4.9	7.35E-14	HK97 family phage prohead protease
GLRG_01933	0	0	26	4.9	9.04E-06	hypothetical protein
GLRG_05509	3	1	60	4.8	1.85E-09	epoxide hydrolase
GLRG_06525	0	3	86	4.8	8.47E-11	hypothetical protein
GLRG_10477	0	0	25	4.8	7.54E-06	cytochrome P450
GLRG_06646	0	1	66	4.8	8.11E-10	glycosyl hydrolase family 61
GLRG_01035	3	0	24	4.8	1.04E-05	ATPase
GLRG_04158	18	45	1,338	4.7	1.97E-18	multicopper oxidase
GLRG_09339	0	2	41	4.7	2.80E-07	multicopper oxidase
GLRG_11156	4	2	71	4.7	8.96E-10	hypothetical protein
GLRG_01366	0	0	23	4.7	1.22E-05	hypothetical protein
GLRG_09112	3	0	23	4.7	1.45E-05	ATPase
GLRG_10288	0	0	31	4.7	2.25E-06	major facilitator superfamily transporter
GLRG_05664	12	5	118	4.7	2.39E-11	alcohol dehydrogenase GroES-like
GLRG_00800	6	2	111	4.7	3.64E-12	glycosyl hydrolase family 28
GLRG_07123	16	6	158	4.6	1.64E-12	high-affinity methionine permease
GLRG_05995	0	0	29	4.6	2.08E-06	hypothetical protein

GLRG_06813	1	1	72	4.6	3.62E-10	glycosyl hydrolase family 35
GLRG_11440	26	156	3,796	4.6	5.78E-18	hypothetical protein
GLRG_09070	0	0	21	4.6	4.60E-05	glycosyl hydrolase family 61
GLRG_11486	1	0	28	4.6	3.68E-06	stress-induced bacterial acidophilic repeat domain-containing protein
GLRG_11837	0	0	21	4.6	5.49E-05	RadH flavin-dependent halogenase
GLRG_00813	1	3	104	4.6	3.94E-11	peroxidase
GLRG_00801	1	4	153	4.6	2.00E-13	sodium/hydrogen exchanger family protein
GLRG_10266	4	0	21	4.5	3.90E-05	hypothetical protein
GLRG_00140	13	6	144	4.5	8.17E-12	hypothetical protein
GLRG_10919	4	11	289	4.5	2.21E-14	hypothetical protein
GLRG_04691	3	0	27	4.5	4.92E-06	hypothetical protein
GLRG_09601	2	1	40	4.5	1.49E-07	Sodium:neurotransmitter symporter family protein
GLRG_10073	322	27	786	4.5	4.86E-17	WSC domain-containing protein
GLRG_10894	5	0	33	4.5	9.01E-07	SPFH domain/Band 7 family protein
GLRG_10836	12	0	19	4.5	1.1E-04	sulfate permease
GLRG_05284	181	1704	40,712	4.4	1.23E-17	hypothetical protein
GLRG_07531	7	11	230	4.4	8.56E-13	integral membrane protein
GLRG_11839	0	0	19	4.4	1.6E-04	cytochrome P450
GLRG_02125	2	1	38	4.4	3.75E-07	ergot alkaloid biosynthesis protein
GLRG_00481	8	1	44	4.4	8.23E-08	ATPase
GLRG_02416	1	3	108	4.4	2.07E-11	hypothetical protein
GLRG_04695	1	1	37	4.4	8.77E-07	hypothetical protein
GLRG_04282	1	0	24	4.4	1.97E-05	hypothetical protein
GLRG_01912	37	4	57	4.3	1.57E-07	hypothetical protein
GLRG_06868	0	0	18	4.3	2.8E-04	GMC oxidoreductase
GLRG_10067	0	0	30	4.3	3.37E-06	POT family protein
GLRG_09761	0	0	18	4.3	2.8E-04	hypothetical protein
GLRG_07529	42	8	178	4.3	3.45E-12	short-chain dehydrogenase/reductase
GLRG_11052	5	1	35	4.3	1.14E-06	O-acetylhomoserine aminocarboxypropyltransferase/cysteine synthase
GLRG_02116	0	16	373	4.3	2.21E-14	zinc carboxypeptidase

[a]Broad Institute identification number (http://www.broadinstitute.org/annotation/genome/colletotrichum_group/MultiHome.html).

Supplementary Table 18. Primer pairs used for amplification and construction of pChEC6::mCherry fluorescent reporter.

Primer pairs		Sequence (5'→3')	Remarks
Primer pair 1	Forward	CACCGGATCCTGTGGGTGAGATTAAGGATAGG TTTGAT	BamHI site underlined
	Reverse	TCCTCGCCCTTGCTCACCATGTTTTCTGTAAGG GGT	hybrid primer, N-terminus of mCherry underlined
Primer pair 2	Forward	ATGGTGAGCAAGGGCGAGGAG	
	Reverse	TTACTTGTACAGCTCGTCCATGCCG	
Primer pair 3	Forward	CGGCATGGACGAGCTGTACAAGTAATTTAATAG CTCCATGTCAACAAG	hybrid primer, C-terminus of mCherry underlined
	Reverse	TCTTGAATTCCGTCTAGAAAGAAGGATTACCTC	EcoRI site underlined