Lifestyle transitions in plant pathogenic Colletotrichum fungi deciphered by genome and transcriptome analyses

Richard J O'Connell, Michael R Thon, Stéphane Hacquard, Stefan G Amyotte, Jochem Kleeman, Maria F Torres, Ulrike Damm, Ester A Buiate, Lynn Epstein, Noam Alkan, et al.

nature genetics

Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses

Richard J O'Connell[1,31], Michael R Thon[2,31], Stéphane Hacquard[1], Stefan G Amyotte[3], Jochen Kleemann[1], Maria F Torres[3], Ulrike Damm[4], Ester A Buiate[3], Lynn Epstein[5], Noam Alkan[6], Janine Altmüller[7], Lucia Alvarado-Balderrama[8], Christopher A Bauser[9], Christian Becker[7], Bruce W Birren[8], Zehua Chen[8], Jaeyoung Choi[10], Jo Anne Crouch[11], Jonathan P Duvick[12,30], Mark A Farman[3], Pamela Gan[13], David Heiman[8], Bernard Henrissat[14], Richard J Howard[12], Mehdi Kabbage[15], Christian Koch[16], Barbara Kracher[1], Yasuyuki Kubo[17], Audrey D Law[3], Marc-Henri Lebrun[18], Yong-Hwan Lee[10], Itay Miyara[6], Neil Moore[19], Ulla Neumann[20], Karl Nordström[21], Daniel G Panaccione[22], Ralph Panstruga[1,23], Michael Place[24], Robert H Proctor[25], Dov Prusky[6], Gabriel Rech[2], Richard Reinhardt[26], Jeffrey A Rollins[27], Steve Rounsley[8], Christopher L Schardl[3], David C Schwartz[24], Narmada Shenoy[8], Ken Shirasu[13], Usha R Sikhakolli[28], Kurt Stüber[26], Serenella A Sukno[2], James A Sweigard[12], Yoshitaka Takano[29], Hiroyuki Takahara[1,30], Frances Trail[28], H Charlotte van der Does[1,30], Lars M Voll[16], Isa Will[1], Sarah Young[8], Qiandong Zeng[8], Jingze Zhang[8], Shiguo Zhou[24], Martin B Dickman[15], Paul Schulze-Lefert[1], Emiel Ver Loren van Themaat[1], Li-Jun Ma[8,30] & Lisa J Vaillancourt[3]

Colletotrichum **species are fungal pathogens that devastate crop plants worldwide. Host infection involves the differentiation of specialized cell types that are associated with penetration, growth inside living host cells (biotrophy) and tissue destruction (necrotrophy). We report here genome and transcriptome analyses of *Colletotrichum higginsianum* infecting *Arabidopsis thaliana* and *Colletotrichum graminicola* infecting maize. Comparative genomics showed that both fungi have large sets of pathogenicity-related genes, but families of genes encoding secreted effectors, pectin-degrading enzymes, secondary metabolism enzymes, transporters and peptidases are expanded in *C. higginsianum*. Genome-wide expression profiling revealed that these genes are transcribed in successive waves that are linked to pathogenic transitions: effectors and secondary metabolism enzymes are induced before penetration and during biotrophy, whereas most hydrolases and transporters are upregulated later, at the switch to necrotrophy. Our findings show that preinvasion perception of plant-derived signals substantially reprograms fungal gene expression and indicate previously unknown functions for particular fungal cell types.**

The genus *Colletotrichum* (Sordariomycetes, Ascomycota; **Fig. 1a**) comprises ~600 species[1] attacking over 3,200 species of monocot and dicot plants (ARS Fungal Databases, see URLs). These pathogens use a multistage hemibiotrophic infection strategy[2]: dome-shaped appressoria first puncture host surfaces using a combination of mechanical force and enzymatic degradation, bulbous biotrophic hyphae enveloped by an intact host plasma membrane then develop inside living epidermal cells, and finally, the fungus switches to necrotrophy and differentiates thin, fast-growing hyphae that kill and destroy host tissues (**Fig. 1b**,**c**). We sequenced two *Colletotrichum* species with different host specificities and infection strategies: *C. higginsianum* attacks several members of *Brassicaceae*, including *Arabidopsis*, and has emerged as a tractable model for studying fungal pathogenicity and plant immune responses[3–5]. Biotrophy in this fungus is confined to the first invaded host cell and is followed by a complete switch to necrotrophy[5] (**Supplementary Fig. 1**). In contrast, *C. graminicola* primarily infects maize (*Zea mays*), causing annual losses of approximately 1 billion dollars in the United States alone[6]. In this species, biotrophy extends into many host cells and persists at the advancing colony margin while the center of the colony becomes necrotrophic[7] (**Supplementary Fig. 2**).

Optical mapping showed that the genomes of the two species are similar in size and structure. *C. graminicola* has a 57.4-Mb genome that is distributed among 13 chromosomes, including three minichromosomes less than 1 Mb in size, whereas *C. higginsianum* has a 53.4-Mb genome comprising 12 chromosomes, including two minichromosomes (**Supplementary Table 1**). We sequenced the *C. graminicola* genome using Sanger and 454 platforms, which provided a high-quality reference assembly of 50.9 Mb. We sequenced the *C. higginsianum* genome using 454 and Illumina platforms, yielding an assembly of 49.3 Mb (**Supplementary Table 2**). Repetitive DNA comprises 12.2% of the *C. graminicola* genome assembly and 1.2% of the *C. higginsianum*

A full list of affiliations appears at the end of the paper.

Received 19 December 2011; accepted 5 July 2012; published online 12 August 2012; doi:10.1038/ng.2372

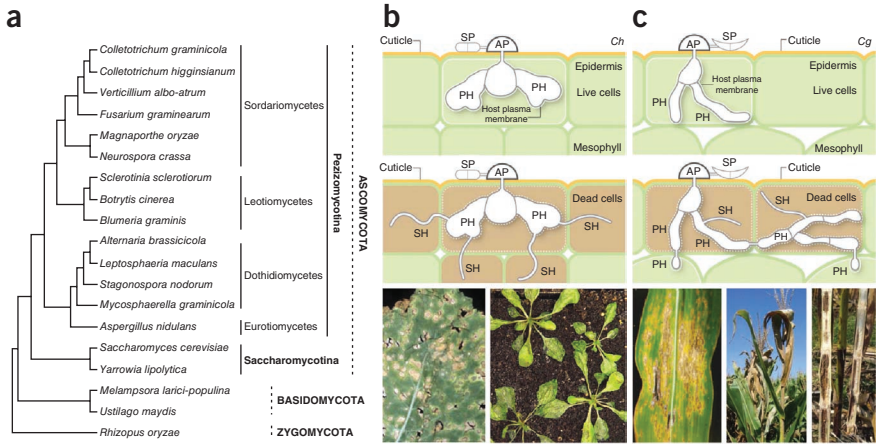

Figure 1 Phylogeny and infection of the two *Colletotrichum* species analyzed in this study. (**a**) Cladogram showing the phylogenetic relationship of *Colletotrichum* to other sequenced fungi, including 13 species used for comparative analyses (see **Fig. 3**). The unscaled tree was constructed using CVTree[34] with *Rhizopus oryzae* as the outgroup. (**b**) Infection process of *C. higginsianum* (*Ch*) and leaf anthracnose symptoms on *Brassica* and *Arabidopsis*. The *Brassica* image is reproduced with permission of University of Georgia Plant Pathology Archive (bugwood.org/). (**c**) Infection process of *C. graminicola* (*Cg*), and leaf-blight, top die-back and stalk-rot symptoms on maize. SP, spore; AP, appressorium; PH, biotrophic primary hyphae; SH, necrotrophic secondary hyphae.

assembly (**Supplementary Tables 2** and **3**). The repeats clustered in genomic regions with low GC content in *C. graminicola* (**Supplementary Fig. 3**), similar to the AT-rich isochores found in *Leptosphaeria maculans*[8]. Including unassembled genomic regions (mostly repeats, such as ribosomal DNA, telomeres, centromeres and transposons), repetitive DNA was estimated to total 22.3% of the *C. graminicola* genome and 9.1% of the *C. higginsianum* genome. The two *Colletotrichum* species diverged relatively recently (~47 million years ago), after the separation of monocots and dicots 140–150 million years ago[9] (**Supplementary Fig. 4**). Although *C. graminicola* and *C. higginsianum* belong to sister clades within the genus (**Supplementary Fig. 5**), only 35% of the two genomes are syntenic (**Supplementary Table 4**), which is less than the synteny between *Botrytis cinerea* and *Sclerotinia sclerotiorum*[10]. Nevertheless, an analysis of synteny between the two *Colletotrichum* genomes identified homologous chromosomes and revealed that major intrachromosomal rearrangements have occurred in one or both species (**Fig. 2a** and **Supplementary Table 5**). The minichromosomes do not contain homologous sequences (**Fig. 2b** and **Supplementary Table 4**), suggesting that they are lineage-specific innovations, and in *C. graminicola*, the minichromosomes are enriched with repetitive DNA (averaging 23%) compared to the core genome (averaging 5.5%).

We predicted the existence of 12,006 protein-coding genes in *C. graminicola* compared to 16,172 in *C. higginsianum* (**Supplementary Table 2**). Having been compiled from short-read data only, the *C. higginsianum* assembly is more fragmented than that of *C. graminicola*, resulting in some genes (5.2%) being split into two or more gene models, whereas others (4%) are truncated versions of the complete gene (**Supplementary Note**). After correcting for this fragmentation, the estimated gene content of *C. higginsianum* (15,331) is still markedly larger than that of *C. graminicola*. The two species share 9,795 orthologous genes. Using Markov clustering (MCL)[11] to analyze the proteomes, we found that 10,077 *C. higginsianum* genes belong to multicopy gene clusters, compared to 5,342 genes in *C. graminicola*, suggesting that the greater gene content of *C. higginsianum* results partly from gene duplication (**Supplementary Table 6**). The MCL analysis also revealed that gene clusters encoding serine proteases, methyl transferases, polyketide synthases, cytochrome P450 enzymes and small-molecule efflux pumps are expanded in *C. higginsianum* compared to *C. graminicola* (**Supplementary Fig. 6**), which we verified by manual inspection (**Supplementary Tables 7** and **8** and **Supplementary Fig. 7**). Clusters that are expanded in *C. graminicola* relative to *C. higginsianum* include a family of genes encoding atypical cellulases (glycoside hydrolase GH61, described below) and another encoding secreted histidine acid phosphatases, which probably mobilize phytic acid, the main form of stored phosphorus in plants[12].

C. higginsianum and *C. graminicola* are particularly well equipped with genes encoding carbohydrate-active enzymes (CAZymes)[13] that potentially degrade the plant cell wall[14] (**Fig. 3a** and **Supplementary Table 9**) and modify the fungal cell wall (**Supplementary Tables 10** and **11**). Both species encode more CAZymes than 13 other fungal genomes we examined. These expanded CAZyme arsenals are more similar to those of other hemibiotrophic and necrotrophic pathogens than to the highly reduced set found in biotrophs such as *Melampsora* and *Blumeria* (**Fig. 3a**). The exceptionally large and diverse inventory of CAZymes encoded by both *Colletotrichum* genomes provides a rich source of enzymes for potential commercial exploitation[15]. *C. higginsianum* encodes over twice as many pectin-degrading enzymes as does *C. graminicola* (**Fig. 3b**), the majority (62%) of which are activated during necrotrophy (**Supplementary Fig. 8** and **Supplementary Note**). Conversely, although both species encode similar numbers of cellulases and hemicellulases, *C. graminicola* activates many more of these genes during necrotrophy (48%) than does *C. higginsianum* (26%), including 22 GH61 copper-dependent oxygenases, which act in concert with classical cellulases to enhance lignocellulose hydrolysis[16,17]. Thus, *C. graminicola* and *C. higginsianum* use very different strategies to deconstruct plant cell walls, reflecting their host preferences: dicot cell walls are enriched with pectin (35% in dicots compared to 10% in maize), whereas the cell walls of grasses contain more hemicellulose (60% in grasses compared to 30% in dicots) and phenolics (up to 5%)[18].

Many phytopathogens secrete proteins known as effectors that facilitate infection by reprogramming host cells and modulating plant immunity[19]. By defining candidate secreted effectors (CSEPs) as predicted extracellular proteins without any homology to proteins outside the genus *Colletotrichum*, we found 177 CSEP-encoding genes in *C. graminicola*, 85 (48%) of which were species specific. In contrast, *C. higginsianum* encodes twice as many CSEPs (365), including more species-specific proteins (264, or 72%) (**Supplementary Fig. 9a**). The CSEPs are mostly small proteins (averaging 110 residues and 175 residues in *C. higginsianum* and *C. graminicola*, respectively) and are more cysteine rich than the total proteome (**Supplementary Fig. 9b**). CSEP-encoding genes are randomly distributed across the chromosomes of *C. graminicola*, with no evidence for clustering, enrichment on particular chromosomes or localization near transposable elements or telomeres, as has been reported for some other plant pathogens[8,20–22] (**Supplementary Note**). An MCL analysis revealed that relatively few *Colletotrichum* CSEPs (14% in both species) belong to small multigenic families with two to five members (**Supplementary Fig. 9c**). The larger, more diversified CSEP repertoire of *C. higginsianum* might be an adaptation to invade a broader range of host plants than *C. graminicola*, which is restricted to infection of *Zea* under field conditions[23].

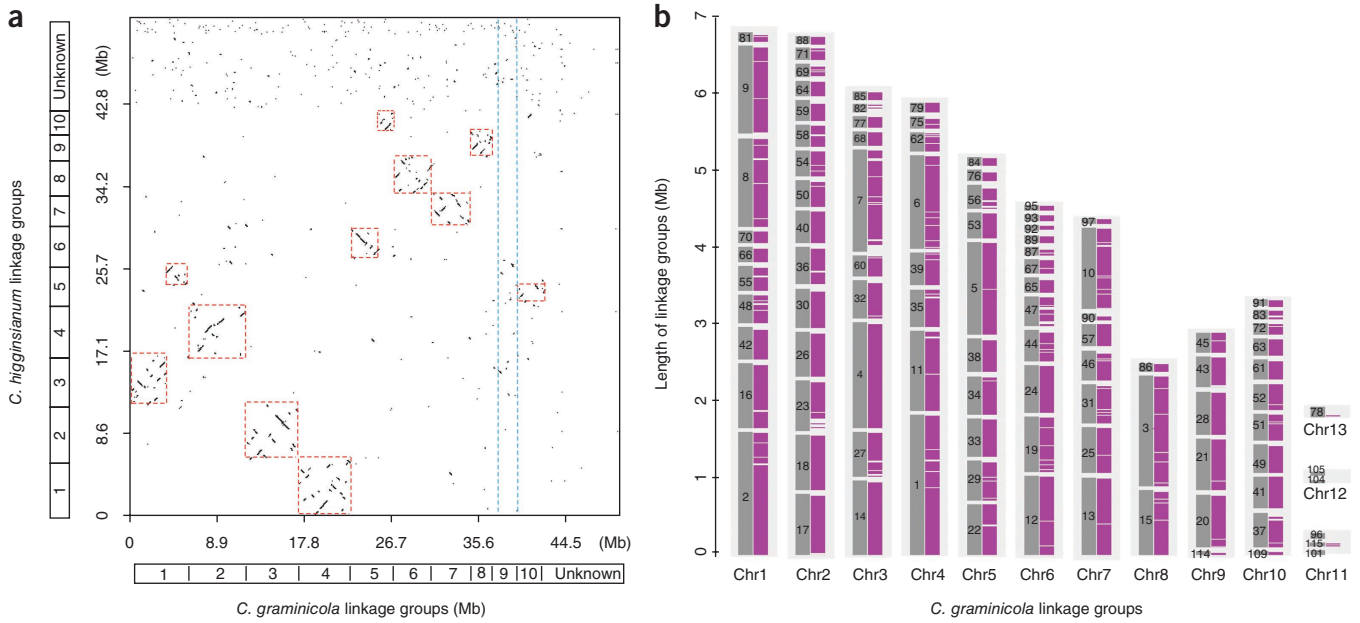

Figure 2 Conservation of synteny between the genomes of *C. graminicola* and *C. higginsianum*. (**a**) Dot plot showing the syntenic blocks between the 13 chromosomes (optical linkage groups) of *C. graminicola* (horizontal axis) and the 12 chromosomes of *C. higginsianum* (vertical axis). Homologies between chromosomes of each species are highlighted in red dashed boxes. Homologous sequences of *C. graminicola* chromosome 9, indicated between the blue dashed lines, are dispersed among many *C. higginsianum* chromosomes. (**b**) Global view of syntenic alignments between the genomes of *C. graminicola* and *C. higginsianum*. Linkage groups of *C. graminicola* are shown as the reference, with linkage group lengths defined by the *C. graminicola* optical map. For each chromosome, numbered genomic scaffolds (dark gray) positioned on the optical linkage groups are separated by scaffold breaks. The magenta blocks show syntenic mapping of the *C. higginsianum* sequences; notably, there is a near absence of homologous sequences among the minichromosomes.

Both *Colletotrichum* species encode markedly more secondary metabolism enzymes (103 in *C. higginsianum* and 74 in *C. graminicola*) than other sequenced fungi (2–58 in ascomycetes[24,25]) (**Supplementary Fig. 10a** and **Supplementary Table 8**). In fungi, secondary metabolism genes are typically located in clusters[26]; we found 42 of these clusters in *C. graminicola* and 39 in *C. higginsianum*, surpassing the numbers found in most other sequenced ascomycetes (**Supplementary Fig. 10b**). Only 11 secondary metabolism gene clusters are shared between the two *Colletotrichum* species (**Supplementary Table 12**), and only 6 of these clusters show limited synteny (**Fig. 4**). This cluster diversity seems to result from gene duplication or loss and chromosomal rearrangements and may be related to the association of secondary metabolism gene clusters (71% in *C. graminicola*) with repetitive DNA (**Supplementary Note**). Because each secondary metabolism gene cluster is probably involved in the biosynthesis of a specific metabolite[24], each *Colletotrichum* species can be expected to produce unusually large and divergent spectra of secondary metabolites, some of which may be previously unknown bioactive molecules.

To investigate how the fungal genetic program is deployed during host infection, we applied Illumina RNA sequencing to both pathosystems (**Supplementary Tables 13** and **14**). We collected samples from infected *Arabidopsis* or maize leaves at intervals corresponding to pre-penetration appressoria, the early biotrophic phase and the transition to necrotrophy and from *C. higginsianum* appressoria formed *in vitro* (**Fig. 5a**). Almost all the gene models were transcribed *in planta* (14,972 *C. higginsianum* genes, or 92%, and 10,812 *C. graminicola* genes, or 90%). However, this transcription was highly dynamic, particularly in *C. higginsianum*, where 7,162 genes (44%) were differentially regulated (log2 fold change >2, $P < 0.05$) between one or more of the infection stages (**Supplementary Tables 15a** and **16**). Fewer genes (2,619, or 22%) were differentially regulated in *C. graminicola*, which may reflect the contrasting biology of this species, where biotrophic and necrotrophic growth

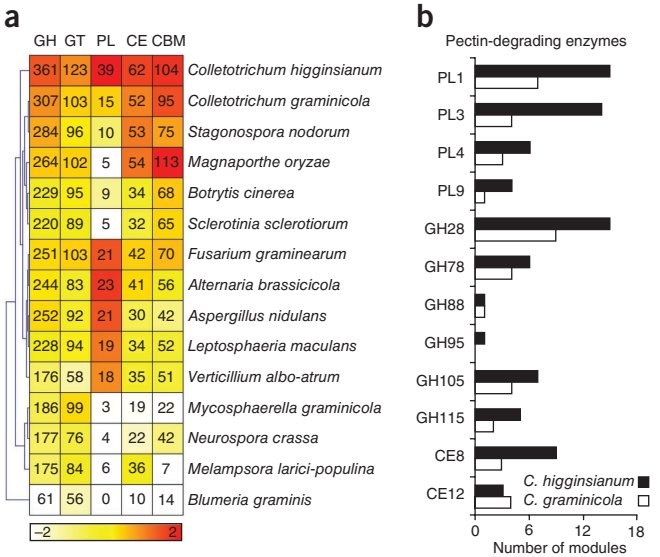

Figure 3 Comparison of fungal carbohydrate-active enzyme (CAZyme) repertoires. (**a**) Hierarchical clustering of CAZyme classes from *Colletotrichum* and 13 other fungal genomes. GH, glycoside hydrolase; GT, glycosyltransferase; PL, polysaccharide lyase; CE, carbohydrate esterase; CBM, carbohydrate-binding module. The numbers of enzyme modules in each genome are shown. Overrepresented (orange to red) and underrepresented modules (pale yellow to white) are depicted as fold changes relative to the class mean. (**b**) Comparison of the pectin-degrading enzyme repertoires of *C. higginsianum* and *C. graminicola* shown as the number of modules in each CAZyme family[11]. In total, *C. higginsianum* encodes 86 such modules, whereas *C. graminicola* encodes only 42.

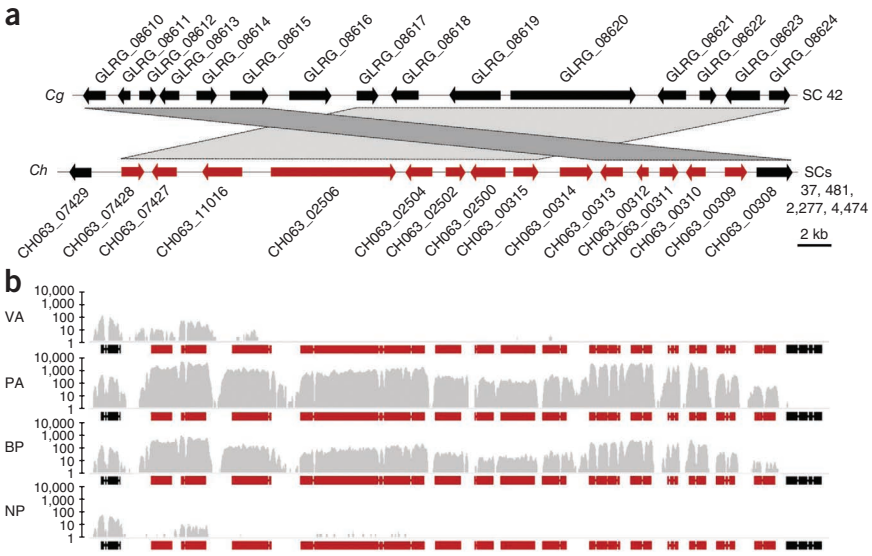

Figure 4 Structure and transcription of a secondary metabolism gene cluster. (**a**) Gene cluster 18 from *C. graminicola* (*Cg*) is orthologous to cluster 10 from *C. higginsianum* (*Ch*). The latter is split between four small scaffolds in the Broad Institute genome annotation (supercontigs (SCs) 37, 481, 2,277 and 4,474) and was reconstructed based on an improved genome assembly (**Supplementary Note**). Microsynteny is indicated by gray bars. The 14 genes highlighted in red in the *C. higginsianum* cluster are co-regulated. Functional annotation for the cluster genes is provided in **Supplementary Table 12**. (**b**) Visualization of RNA-Seq coverage across the *C. higginsianum* polyketide biosynthesis cluster. The gray curves indicate read coverage (log scale) for the four samples. Co-regulated gene models are highlighted in red. VA, *in vitro* appressoria; PA, *in planta* appressoria; BP, biotrophic phase; NP, necrotrophic phase.

occur simultaneously (**Supplementary Tables 15b** and **17**). The more clearly defined infection stages of *C. higginsianum* provided better temporal and spatial resolution of expression changes, and we therefore highlight our results for this species.

Five gene categories relevant to pathogenicity (encoding transcription factors, secondary metabolism enzymes, CSEPs, CAZymes and transporters) had markedly different expression patterns during infection (**Fig. 5b** and **Supplementary Fig. 11**). We distinguished three waves of gene activation corresponding to pathogenic transitions (**Fig. 5c**). Among the genes upregulated at the appressorial phase were those encoding CAZymes that are predicted to degrade cutin, cellulose, hemicellulose and pectin, which may contribute to initial host penetration, together with a larger set of enzymes that potentially remodel the fungal cell wall (**Fig. 5b** and **Supplementary Fig. 8a**). However, early during infection, the transcriptome of *C. higginsianum* was dominated by secondary metabolism genes, with 12 different secondary metabolism gene clusters being induced before penetration and during biotrophy (**Fig. 5b,c** and **Supplementary Table 13**). This indicates previously unsuspected roles for appressoria and biotrophic hyphae in synthesizing an array of small molecules for delivery to the first infected plant cells. Because these cells initially remain alive, such molecules are probably not toxins and may instead function in host manipulation, similar

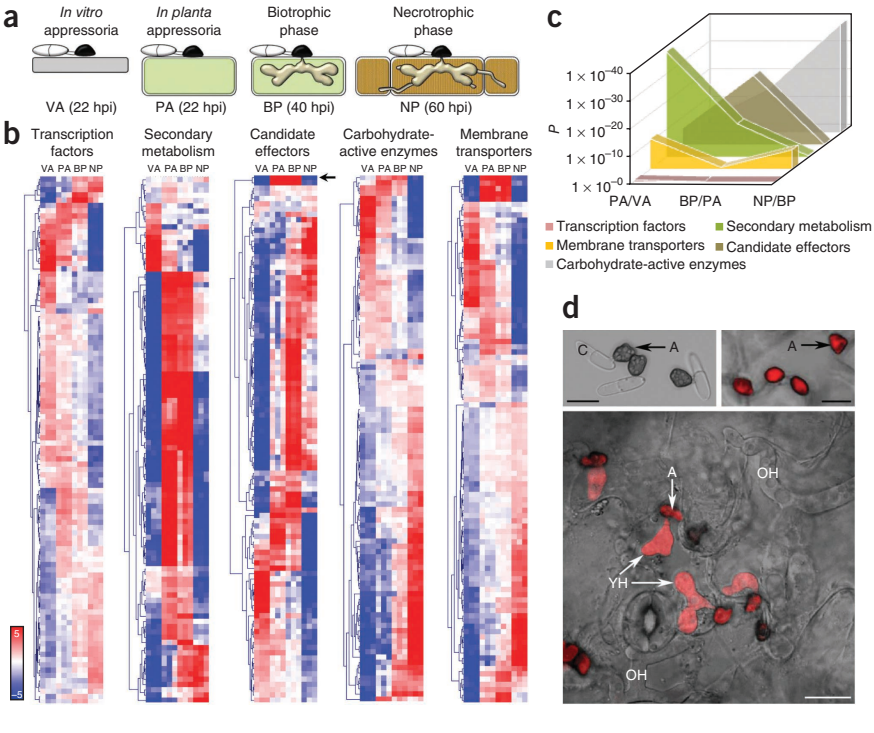

Figure 5 Expression profiling of pathogenicity-related genes in *C. higginsianum*. (**a**) Schematic representation of the four *C. higginsianum* developmental stages selected for RNA sequencing. Gray indicates polystyrene, green indicates living plant cell, and brown indicates dead plant cell. Hpi, hours post-inoculation. (**b**) Heatmaps of gene expression showing the 100 most highly expressed and significantly regulated genes (log2 fold change >2, $P < 0.05$) in five functional categories. Overrepresented (pale red to dark red) and underrepresented transcripts (pale blue to dark blue) are shown as log2 fold changes relative to the mean expression measured across all four stages. The arrow indicates the CSEP-encoding gene *ChEC6* (CH063_01084). (**c**) The statistical significance of gene induction (*y* axis) in five functional categories during fungal developmental transitions (*x* axis). The *P* values were calculated using a one-sided Fisher's exact test and represent the probability of observing the number of significantly induced genes for a specific category during a transition given the total number of significantly induced genes during that transition (log2 fold change >2, $P < 0.05$) and the total number of genes in the category. (**d**) Transcriptional regulation of the effector gene *ChEC6* by plant-derived signals. Confocal micrographs showing *C. higginsianum* expressing the *mCherry* reporter gene under the native *ChEC6* promoter (overlays of bright-field and fluorescence channels). Appressoria (A) formed on polystyrene are unlabeled (top left), whereas those on the leaf surface (top right) have fluorescent cytoplasm. After host penetration, labeling is visible in young biotrophic hyphae (YH) but not older biotrophic hyphae (OH) (bottom). Scale bars, 10 μm. C, conidium.

to protein effectors[27]. Remarkably, the *C. higginsianum* secondary metabolism gene cluster with the strongest activation at this stage was silent in *C. graminicola* at all the infection stages we examined (**Fig. 4** and **Supplementary Fig. 12**), suggesting that additional metabolite diversity is generated through transcriptional regulation.

Different sets of CSEP-encoding genes were expressed at each infection stage, but the majority of these genes were strongly induced during biotrophy (**Fig. 5b**,**c** and **Supplementary Table 16**). This suggests *Colletotrichum* requires a maximum capacity for host manipulation during intracellular colonization and that biotrophic hyphae provide a major interface for effector delivery to host cells. These specialized hyphae morphologically resemble the haustoria of obligate biotrophs, which function both as platforms for effector secretion and feeding structures for the uptake of sugars and amino acids[28,29]. However, we found no evidence for specific transcriptional reprogramming of nutrient transporters in *C. higginsianum* during biotrophy (**Fig. 5b**,**c** and **Supplementary Table 16**), suggesting that the biotrophic hyphae of this pathogen function primarily to deliver protein effectors and secondary metabolites to the plant cell.

Transcripts encoding a vast array of lytic enzymes are induced at the transition to necrotrophy, when the pathogen uses dead and dying host cells as a nutrient source to support rapid colonization and sporulation (**Fig. 5b**,**c** and **Supplementary Fig. 8**). These enzymes include 44 putative secreted proteases and 146 CAZymes that potentially cleave all major polysaccharides in the host wall (**Supplementary Figs. 8** and **11**). Concomitantly, numerous genes encoding plasma membrane transporters that may be required for assimilating the products of this degradative activity, for example, oligopeptides, amino acids and sugars, are also induced (**Fig. 5b**,**c**). In fungi, genes encoding secreted proteases, CAZymes and permeases are often subject to pH regulation[30]. Consistent with this, we found evidence that necrotrophy in *C. higginsianum* is associated with local alkalinization of *Arabidopsis* tissue, probably resulting from fungal ammonia secretion[31], but tissue alkalinization was less pronounced in maize colonized by *C. graminicola* at this stage (**Supplementary Fig. 13**).

Notably, although appressoria *in vitro* are morphologically indistinguishable from those *in planta*, their transcriptomes are substantially different, with 1,515 genes significantly induced by host contact (**Supplementary Table 15**). One of these, the CSEP-encoding gene called *ChEC6* (CH063_01084)[32], is the most highly and significantly induced of all *C. higginsianum* genes (>50,000-fold) compared to appressoria *in vitro*. To experimentally verify this expression pattern at the cellular level, we generated transgenic *C. higginsianum* strains expressing a reporter gene under the control of the *ChEC6* promoter (**Fig. 5d**). Using this method, we confirmed that the transcription of *ChEC6* was plant specific, starting in the appressorium before penetration and continuing in young biotrophic hyphae, but it was switched off before the hyphae were fully expanded, indicating that its expression is transient and tightly regulated. The large-scale reprogramming of appressorial gene expression *in planta* shows that these specialized cells are highly responsive to host-derived cues that are perceived before penetration. Long regarded as organs of attachment and penetration[33], our findings assign a previously unsuspected sensory function to fungal appressoria, enabling the pathogen to prepare for the subsequent invasion of living host cells.

Major hemibiotrophic plant pathogens such as *Colletotrichum* and the rice blast fungus *Magnaporthe oryzae* undergo major transformations in cell morphology and infection mode when switching from growth on the plant surface to intracellular biotrophy and from biotrophy to necrotrophy. Genome sequencing combined with high-throughput transcriptome sequencing revealed the transcriptional dynamics underlying these transitions and led us to redefine the functions of appressoria and intracellular hyphae. Despite their similar morphologies, a genomic comparison of *C. higginsianum* and *C. graminicola* uncovered major differences in their gene content. We propose that the diversification of functions required for host interaction, notably, the secretion of small-molecule and protein effectors and the degradation of plant polymers, allows *C. higginsianum* to colonize a wider range of plant species. In contrast, *C. graminicola*, a pathogen that is adapted to a narrow range of hosts, has maintained a more targeted arsenal of virulence factors.

URLs. Broad Institute Colletotrichum Genome Database, http://www.broadinstitute.org/annotation/genome/colletotrichum_group; Max Planck Institute for Plant Breeding Research Fungal Genomes Database, http://www.mpipz.mpg.de/14157/fungal_genomes; Fungal Transcription Factor Database, http://ftfd.snu.ac.kr/; Fungal Cytochrome P450 Database, http://p450.riceblast.snu.ac.kr/species.php; Transporter Classification Database, http://www.tcdb.org/; *Saccharomyces* Genome Database, http://www.yeastgenome.org/; NCBI Conserved Domains Database, http://www.ncbi.nlm.nih.gov/cdd; SAMtools, http://samtools.sourceforge.net/; Broad Institute Integrative Genomics Viewer (IGV) browser, http://www.broadinstitute.org/igv/; Geneious v5.5., http://www.geneious.com/; RepeatMasker Open-3.0, http://www.repeatmasker.org/; InterProScan, http://www.ebi.ac.uk/Tools/pfa/iprscan/; ARS Fungal Databases, http://nt.ars-grin.gov/fungaldatabases/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. The *C. graminicola* and *C. higginsianum* genome assemblies have been deposited in NCBI's Whole-Genome Shotgun Project with accession numbers ACOD01000000000 and CACQ02000000000, respectively. The RNA-Seq data for *C. graminicola* and *C. higginsianum* have been deposited in the NCBI Gene Expression Omnibus under GEO Series accession numbers GSE34632 and GSE33683, respectively.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

This manuscript is dedicated to the memory of Robert Hanau. This work was primarily supported by US Department of Agriculture (USDA)–Cooperative State Research, Education and Extension Service (CSREES) grant 2007-35600-17829 (L.J.V., M.B.D., S.R. and 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.), the 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. Shirasu). This is manuscript number 12-12-011, published with the approval of the Director of the University of Kentucky Agricultural Experiment Station. See the **Supplementary Note** for a full list of acknowledgments.

AUTHOR CONTRIBUTIONS

L.J.V., R.J.O., M.R.T., L.-J.M., P.S.-L. and M.B.D. initiated the project. R.J.O. and L.J.V. coordinated the project. M.R.T., E.V.L.v.T. and L.-J.M. coordinated bioinformatics. L.-J.M., B.W.B., C.A.B., U.D., Y.T., R.R., E.V.L.v.T., K. Stüber, H.T., J.K., L.J.V., A.D.L., J.P.D., J.A.S., C.L.S. and S.R. performed DNA and expressed sequence tag sequencing. D.H., S.Y., L.-J.M., C.A.B., E.V.L.v.T., K.N. and K. Stüber performed sequence assembly. S.Z., D.C.S., J.Z., D.H., L.-J.M., J.A.R., M.P. and S.G.A. performed optical map and synteny analyses. Q.Z., Z.C., L.A.-B., N.S., L.-J.M., E.V.L.v.T., S.G.A., N.M., M.R.T. and B.K. performed gene predictions and managed the data. E.A.B., M.R.T., S.A.S., L.E., M.F.T., R.J.O., L.J.V., K. Stüber, E.V.L.v.T., J.K. and H.C.v.d.D. manually curated gene models. U.D., M.R.T., Y.-H.L. and G.R. performed phylogeny and evolution analyses. M.R.T. and G.R. performed comparative genomics and MCL analyses. S.G.A., M.L.F., M.R.T. and E.V.L.v.T. analyzed repeats. S.H., E.V.L.v.T., J.K., H.C.v.d.D., B.K., J.A., C.B. and R.J.O. performed and analyzed *C. higginsianum* RNA-Seq. E.A.B., M.F.T., S.G.A., L.J.V. and M.K. performed and analysed

C. graminicola RNA-Seq. L.M.V. and C.K. annotated transporters. L.J.V., M.F.T., M.-H.L., Y.K., D.G.P., R.H.P., C.L.S., S.H., R.J.O., M.R.T. and S.G.A. annotated secondary metabolism genes. B.H., L.E., H.T. and M.-H.L. annotated CAZymes and chitin-binding proteins. M.R.T., S.A.S. and S.H. annotated secretomes and effectors. P.G. and K. Shirasu annotated phytohormone genes. M.R.T. and R.J.O. annotated proteases. Y.-H.L. and J.C. annotated transcription factors and P450 proteins. J.A.C., F.T. and U.R.S. annotated mating-type genes. J.K. and I.W. performed reporter assays. D.P., N.A. and I.M. performed pH and ammonia assays. S.H., U.N., E.V.L.v.T., R.J.H., J.K., L.J.V., S.G.A., U.D., M.R.T., L.-J.M., E.A.B. and R.J.O. prepared tables, figures and text. R.J.O. wrote and edited the paper with input from L.J.V., L.-J.M., M.R.T., P.S.-L., R.P., S.H., J.K., M.-H.L., L.E., E.V.L.v.T., E.A.B. and S.G.A.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/ng.2372.

1. Crous, P.W., Gams, W., Stalpers, J.A., Robert, V. & Stegehuis, G. MycoBank: an online initiative to launch mycology into the 21st century. *Stud. Mycol.* **50**, 19–22 (2004).
2. 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).
3. 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).
4. Narusaka, M. *et al. RRS1* and *RPS4* provide a dual resistance-gene system against fungal and bacterial pathogens. *Plant J.* **60**, 218–226 (2009).
5. 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).
6. 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).
7. 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).
8. Rouxel, T. *et al.* Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by repeat-induced point mutations. *Nat. Commun.* **2**, 202 (2011).
9. Chaw, S.M., Chang, C.C., Chen, H.L. & Li, W.H. Dating the monocot-dicot divergence and the origin of core Eudicots using whole chloroplast genomes. *J. Mol. Evol.* **58**, 424–441 (2004).
10. Amselem, J. *et al.* Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genet.* **7**, e1002230 (2011).
11. 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).
12. Oh, B.C., Choi, W.C., Park, S., Kim, Y.O. & Oh, T.K. Biochemical properties and substrate specificities of alkaline and histidine acid phytases. *Appl. Microbiol. Biotechnol.* **63**, 362–372 (2004).
13. Cantarel, B.L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37**, D233–D238 (2009).
14. van den Brink, J. & de Vries, R.P. Fungal enzyme sets for plant polysaccharide degradation. *Appl. Microbiol. Biotechnol.* **91**, 1477–1492 (2011).
15. King, B.C. *et al.* Arsenal of plant cell wall degrading enzymes reflects host preference among plant pathogenic fungi. *Biotechnol. Biofuels* **4**, 4 (2011).
16. Quinlan, R.J. *et al.* Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc. Natl. Acad. Sci. USA* **108**, 15079–15084 (2011).
17. 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).
18. Vogel, J. Unique aspects of the grass cell wall. *Curr. Opin. Plant Biol.* **11**, 301–307 (2008).
19. Stergiopoulos, I. & de Wit, P.J.G.M. Fungal effector proteins. *Annu. Rev. Phytopathol.* **47**, 233–263 (2009).
20. Ma, L.-J. *et al.* Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**, 367–373 (2010).
21. Farman, M.L. Telomeres in the rice blast fungus *Magnaporthe oryzae*: the world of the end as we know it. *FEMS Microbiol. Lett.* **273**, 125–132 (2007).
22. Kämper, J. *et al.* Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* **444**, 97–101 (2006).
23. Crouch, J.A. & Beirn, L.A. Anthracnose of cereals and grasses. *Fungal Divers.* **39**, 19–44 (2009).
24. 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).
25. 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).
26. Khaldi, N. *et al.* SMURF: genomic mapping of fungal secondary metabolite clusters. *Fungal Genet. Biol.* **47**, 736–741 (2010).
27. Collemare, J. & Lebrun, M.-H. Fungal secondary metabolites: ancient toxins and novel effectors in plant-microbe interactions. in *Effectors in Plant-Microbe Interactions* (eds. Martin, F. & Kamoun, S.) 379–402 (Wiley-Blackwell, Oxford, 2011).
28. Voegele, R.T. & Mendgen, K.W. Nutrient uptake in rust fungi: how sweet is parasitic life? *Euphytica* **179**, 41–55 (2011).
29. Catanzariti, A.-M., Dodds, P.N., Lawrence, G.J., Ayliffe, M.A. & Ellis, J.G. Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *Plant Cell* **18**, 243–256 (2006).
30. Peñalva, M.A. & Arst, H.N. Regulation of gene expression by ambient pH in filamentous fungi and yeasts. *Microbiol. Mol. Biol. Rev.* **66**, 426–446 (2002).
31. 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).
32. Kleemann, J. *et al.* Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen. *Colletotrichum higginsianum*. *PLoS Pathog.* **8**, e1002643 (2012).
33. Frank, B. Ueber einige neue und weniger bekannte Pflanzenkrankheiten. *Ber. Deutsch. Bot. Gesell.* **1**, 29–34 (1883).
34. Wang, H., Xu, Z., Gao, L. & Hao, B. A fungal phylogeny based on 82 complete genomes using the composition vector method. *BMC Evol. Biol.* **9**, 195 (2009).

[1]Department of Plant Microbe Interactions, Max Planck Institute for Plant Breeding Research, Cologne, Germany. [2]Centro Hispano-Luso de Investigaciones Agrarias, Departamento de Microbiología y Genética, Universidad de Salamanca, Villamayor, Spain. [3]Department of Plant Pathology, University of Kentucky, Lexington, Kentucky, USA. [4]Centraalbureau voor Schimmelcultures, Koninklijke Nederlandse Akademie van Wetenschappen, Fungal Biodiversity Centre, Utrecht, The Netherlands. [5]Department of Plant Pathology, University of California, Davis, California, USA. [6]Department of Postharvest Science of Fresh Produce, Agricultural Research Organization, The Volcani Center, Bet Dagan, Israel. [7]Cologne Center for Genomics, University of Cologne, Cologne, Germany. [8]Broad Institute, Cambridge, Massachusetts, USA. [9]GATC Biotech AG, Konstanz, Germany. [10]Department of Agricultural Biotechnology, Center for Fungal Genetic Resources, Seoul National University, Seoul, Korea. [11]Systematic Mycology and Microbiology Laboratory, US Department of Agriculture, Agricultural Research Service, Beltsville, Maryland, USA. [12]Pioneer Hi-Bred International, DuPont Agricultural Biotechnology, Wilmington, Delaware, USA. [13]Plant Immunity Research Group, RIKEN Plant Science Center, Yokohama, Japan. [14]Laboratoire Architecture et Fonction des Macromolécules Biologiques, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 7257, Université Aix-Marseille, Marseille, France. [15]Department of Plant Pathology and Microbiology, Institute for Plant Genomics and Biotechnology, Borlaug Genomics and Bioinformatics Center, Texas A&M University, College Station, Texas, USA. [16]Department of Biology, Division of Biochemistry, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany. [17]Laboratory of Plant Pathology, Graduate School of Life and Environmental Sciences, Kyoto Prefectural University, Kyoto, Japan. [18]Institut National de la Recherche Agronomique, Biologie et Gestion des Risques en Agriculture—Champignons Pathogènes des Plantes, Thiverval-Grignon, France. [19]Department of Computer Science, University of Kentucky, Lexington, Kentucky, USA. [20]Central Microscopy, Max Planck Institute for Plant Breeding Research, Cologne, Germany. [21]Department of Plant Developmental Biology, Max Planck Institute for Plant Breeding Research, Cologne, Germany. [22]Division of Plant and Soil Sciences, West Virginia University, Morgantown, West Virginia, USA. [23]Institute for Biology I, Unit of Plant Molecular Cell Biology, Rheinisch-Westfälische Technische Hochschule Aachen University, Aachen, Germany. [24]Laboratory for Molecular and Computational Genomics, University of Wisconsin-Madison, Madison, Wisconsin, USA. [25]US Department of Agriculture, Agriculture Research Service, National Center for Agricultural Utilization Research, Peoria, Illinois, USA. [26]Max Planck Genome Centre Cologne, Cologne, Germany. [27]Department of Plant Pathology, University of Florida, Gainesville, Florida, USA. [28]Department of Plant Biology, Michigan State University, East Lansing, Michigan, USA. [29]Laboratory of Plant Pathology, Graduate School of Agriculture, Kyoto University, Kyoto, Japan. [30]Present addresses: Department of Genetics, Development and Cell Biology, Iowa State University, Ames, Iowa, USA (J.P.D.), Department of Bioproduction Science, Faculty of Bioresources and Environmental Sciences, Ishikawa Prefectural University, Ishikawa, Japan (H.T.), Department of Plant Pathology, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands (H.C.v.d.D.) and The College of Natural Sciences, University of Massachusetts Amherst, Amherst, Massachusetts, USA (L.-J.M.). [31]These authors contributed equally to this work. Correspondence should be addressed to R.J.O. (oconnel@mpipz.mpg.de) or L.J.V. (vaillan@uky.edu).

ONLINE METHODS

Sequencing and assembly. *C. graminicola* strain M1.001 (M2) was collected in Missouri from infected maize (Fungal Genetics Stock Center culture 10212). *C. higginsianum* strain IMI349063 was isolated from *Brassica campestris* in Trinidad and Tobago (CABI culture collection, Wallingford, UK). The genome assemblies of *C. graminicola* were generated at the Broad Institute by combining data from Sanger and 454 pyrosequencing using a Newbler hybrid approach. Paired-end reads from 468,734 plasmids and 67,151 fosmids improved the continuity of the assembly (**Supplementary Table 2**). In the assembled genome, more than 98.5% of the sequence bases had quality scores >40. The *C. higginsianum* genome assembly was generated by GATC Biotech AG (Konstanz, Germany) by combining 454 GS-FLX shotgun reads and Illumina GAII mate-pair reads. Additionally, 864 fosmids were end-sequenced with Sanger technology. After removing dinucleotide repeats, the 454 reads and the fosmid end sequences were coassembled using the SeqMan NGen assembler (DNAStar Inc., USA). Contigs were then sorted into scaffolds using the paired-end information derived from an Illumina 3-kb–insert mate-pair library (2 × 36 bp reads). Scaffolds were manually edited to correct falsely joined contigs and falsely arranged scaffolds. To correct homopolymer sequencing errors in the 454 data, the 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**).

Gene annotation. A total of 28,424 expressed sequence tags (ESTs) from two *C. graminicola* complementary DNA libraries and 828,592 ESTs from six *C. higginsianum* libraries were used to provide a training set for the gene-calling pipeline and for validating the gene models (**Supplementary Note**). Protein-coding genes were annotated in *C. graminicola* using multiple lines of evidence from BLAST, PFAM searches and EST alignments, as described previously[20]. Gene structures were predicted using the Broad Institute automated gene-calling pipeline[35] based on a combination of gene models predicted by the programs FGENESH (Softberry Inc., USA), GENEID[36], GeneMark[37], SNAP[38] and Augustus[39] together with EST-based and manually curated gene models. GENEID, FGENESH, SNAP and Augustus were trained using a set of high-confidence EST-based gene models generated by clustering Blat-aligned species-specific ESTs. By combining BLAST, EST and *ab initio* predictions, annotators manually built additional gene models that were otherwise missed by the automated annotation. *C. graminicola* was predicted to have 12,006 gene models, 39% of which were verified by the alignment of 13,600 Sanger EST reads. The *C. higginsianum* gene set was created similarly and was filtered using TBLASTN alignments from 10,661 of the *C. graminicola* gene models ($<1 \times 10^{-10}$). Another 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. *C. higginsianum* was predicted to have 16,172 protein-coding genes, 89% of which were validated by the alignment of 135,923 ESTs from 454 sequencing.

Optical mapping. *C. graminicola* and *C. higginsianum* protoplasts[40] were lysed and prepared for optical mapping[41] using *Mlu*I (with an average fragment size for both genomes of 9.2 kb). Raw datasets comprising single DNA molecule maps (Rmaps; 300× coverage per genome) were assembled into genome-wide contig maps spanning each chromosome using divide-and-conquer[41] and iterative assembly strategies[42]. PROmer from the MUMmer package[43] was used to conduct pairwise comparisons between the *C. graminicola* and *C. higginsianum* genomes (**Fig. 2a** and **Supplementary Table 4**). The synteny map (**Fig. 2b**) was generated using the Argo browser[44].

Transposable element analysis. Repetitive DNA elements were identified by performing a self BLASTN of each genome and processing the output with a custom Perl script (available on request), which identified multicopy sequences and organized them into nonredundant families. Consensus sequences of these families were then used to generate a custom library for RepeatMasker (see URLs) to scan both genome assemblies. The distributions of the genes, the transposable elements and the GC content were examined within a 100-kb window, sliding 10 kb across each chromosome.

Orthology and multigene families. To identify differences in gene family size between *C. graminicola* and *C. higginsianum*, we clustered their proteomes using the Markov clustering program MCL[11]. An all-versus-all BLASTP search was performed using default parameters, followed by clustering with MCL using an inflation value of 2.0. We also included the proteomes of 13 additional fungal species (**Fig. 3**). Sequences were aligned using MAFFT[45], and phylogenetic trees were constructed using the neighbor-joining method, followed by a bootstrap test with 100 replications. Sequence editing and alignment and phylogenetic analyses were performed using Geneious Pro (version 5.5; see URLs).

Annotation of specific gene categories. Secretomes of both *Colletotrichum* species were predicted using WoLF-PSORT[46]. CSEPs were defined as extracellular proteins with no significant BLAST homology (expect value $<1 \times 10^{-3}$) to sequences in the UniProt database (SwissProt and TrEMBL components). Homologs of proteins from outside the genus *Colletotrichum* were excluded. Genes encoding putative carbohydrate-active enzymes were identified using the CAZy annotation pipeline[13]. To identify secreted peptidase genes, sequences of predicted extracellular proteins were subjected to a MEROPS Batch BLAST analysis[47]. Membrane transporters were identified from BLAST searches against the Transporter Collection Database (see URLs) and *Saccharomyces* Genome Database (see URLs). Secondary metabolism genes were initially identified using MCL and gene family searches using the Broad Institute *Colletotrichum* database, BLAST searches against GenBank and InterproScan analysis (see URLs). The Secondary Metabolite Unknown Region Finder (SMURF)[26] was used to predict secondary metabolism gene clusters. SMURF was applied to the Velvet assembly of *C. higginsianum* (**Supplementary Note**). Candidate genes identified using automated searches were inspected manually, including protein sequence alignments to known enzymes and searches against the NCBI Conserved Domain Database (see URLs). Further details of the secondary metabolism gene annotation are presented in the **Supplementary Note**. The Fungal Cytochrome P450 Database (see URLs) and Fungal Transcription Factor Database (see URLs) were used to annotate cytochrome P450 enzymes and transcription factors, respectively.

Whole-genome transcriptome profiling. *Arabidopsis* leaves infected by *C. higginsianum* were obtained as described previously[48]. Sampling and RNA isolation of the pre-penetration stage (22 h hpi), the early biotrophic stage (40 hpi), the switch between biotrophy and necrotrophy (60 hpi) and *in vitro* appressoria (22 hpi) have been described previously[32,49]. Each experimental repetition of the *in planta* stages was based on RNA extracted from ~300 leaves. Maize leaf sheaths infected by *C. graminicola* were obtained as described previously[50]. Sheaths from the maize inbred line Mo940 at the V3 stage were cut into 5-cm–long segments and inoculated with two 10-µl drops of spore suspension (5×10^5 spores per ml). Sheaths containing mature pre-penetration appressoria (24 hpi), intracellular biotrophic hyphae (36 hpi) and necrotrophic hyphae with water-soaked lesions (60 hpi) were sampled. Each leaf sheath was trimmed to include only the inoculated area, and total RNA was extracted as described previously[51] (15 maize sheaths per experimental repetition). The RNA integrity of all samples was verified on an Agilent 2100 Bioanalyzer.

Twelve *C. higginsianum* libraries (four developmental stages and three biological replicates) and nine *C. graminicola* libraries (three developmental stages and three biological replicates) were prepared with the Illumina TruSeq RNA Sample Preparation Kit and sequenced using the Illumina Genome Analyzer IIx (single reads, 100 bp for *C. higginsianum* and 76 bp for *C. graminicola*). Further details are provided in the **Supplementary Note**. The RNA-Seq reads were mapped to the annotated genomes with TopHat ($a = 10$, $g = 5$)[52] and transformed into counts per annotated gene per sample with the 'coverageBed' function from the BEDtools suite[53] and custom R scripts. Differentially expressed genes between two developmental stages were detected using the 'exactTest' function from the R package EdgeR[54]. 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, and direct ratios (log2) were calculated between the different developmental stages. Transcripts with a significant P value (<0.05) and more than a twofold change (log2) in transcript level were considered to be differentially expressed. All P values were corrected for false discoveries resulting from multiple hypothesis testing using the Benjamini-Hochberg procedure. Heatmaps of gene expression profiles were generated with the Genesis expression analysis package[55]. All codes for the RNA-Seq

doi:10.1038/ng.2372

processing are available upon request. The *C. higginsianum* RNA-Seq data were also mapped onto the unannotated Velvet genome assembly (**Supplementary Note**) using bowtie[56] and visualized with SAMtools (see URLs) and the IGV browser (see URLs). RNA-Seq expression profiles were validated by quantitative RT-PCR (**Supplementary Note**).

Molecular phylogeny and evolutionary divergence date estimation. A whole-genome cladogram showing the phylogenetic relationships of *C. graminicola* and *C. higginsianum* to 17 other sequenced fungi was constructed with CVTree[34] (**Fig. 1**). A phylogeny was generated for the genus *Colletotrichum* based on sequencing five genes in 28 selected isolates (**Supplementary Fig. 5**), as described in the **Supplementary Note**. To estimate the evolutionary divergence date for *C. graminicola* and *C. higginsianum*, a phylogenetic analysis was performed using the 13 species shown in **Supplementary Figure 4**. The proteomes were clustered using MCL, and proteins in each cluster were aligned using MUSCLE. Sixty-four clusters containing only one protein from each species and having at least 80% average pairwise nucleotide identity were used for further analyses. Sequence alignments were concatenated, and a phylogenetic tree was constructed with MrBayes[57] using the WAG amino acid substitution model. Date estimates were computed using the program r8s[58] with the nonparametric rate smoothing (NPRS) method using date estimates by Lücking *et al.*[59].

Fluorescent reporter gene assay. The promoter of the CSEP-encoding gene *ChEC6* (CH063_01084) was fused to *mCherry*[60] and a transcriptional terminator by overlap fusion PCR[61] using the primer pairs shown in **Supplementary Table 18**. The genomic region between the *ChEC6* start codon and the stop codon of its upstream gene (1,198 bp) was amplified with primer pair 1. The *mCherry* gene was amplified with primer pair 2. The transcriptional terminator of *Aspergillus nidulans trpC* was amplified from the plasmid pBin-GFP-hph[5] with primer pair 3. After fusion, the insert was subcloned into the plasmid pENTR/D-TOPO (Invitrogen) and verified by sequencing. The insert was cut out with BamHI and EcoRI and ligated into the plasmid pBIGDR1, providing direct repeat recombination-mediated gene targeting[62]. A *ku70* mutant of *C. higginsianum* strain IMI349063 (ref. 62) was used for *Agrobacterium*-mediated transformation[3]. Confocal images of transformants were obtained using a Leica TCS SP2 confocal laser scanning microscope. Excitation for imaging mCherry fluorescence was at 563 nm, and emission was detected at 566–620 nm.

Host tissue alkalinization. The pH of the host cells during infection was measured using the cell-permeant pH-sensitive dye 2′,7′-bis(carboxyethyl)-5(6)-carboxyfluorescein (BCECF) for analysis by epifluorescence microscopy[31]. Fluorescence intensity values were correlated with direct pH determinations obtained with a piercing-tip pH electrode (Eutech, Singapore). Ammonia concentrations in infected maize and *Arabidopsis* leaf tissues were measured using a photometric ammonium assay kit (Merck, Germany).

35. Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32**, W309–W312 (2004).
36. Haas, B.J., Zeng, Q., Pearson, M.D., Cuomo, C.A. & Wortman, J.R. Approaches to fungal genome annotation. *Mycology* **2**, 118–141 (2011).
37. Parra, G., Blanco, E. & Guigó, R. GeneID in *Drosophila*. *Genome Res.* **10**, 511–515 (2000).
38. Borodovsky, M. & McIninch, J. GeneMark: parallel gene recognition for both DNA strands. *Comput. Chem.* **17**, 123–133 (1993).
39. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
40. 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).
41. Zhou, S. *et al.* Validation of rice genome sequence by optical mapping. *BMC Genomics* **8**, 278 (2007).
42. Zhou, S. *et al.* A single molecule scaffold for the maize genome. *PLoS Genet.* **5**, e1000711 (2009).
43. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
44. Engels, R. *et al.* Combo: a whole genome comparative browser. *Bioinformatics* **22**, 1782–1783 (2006).
45. 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).
46. Horton, P. *et al.* WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* **35**, W585–W587 (2007).
47. Rawlings, N.D., Barrett, A.J. & Bateman, A. MEROPS: the peptidase database. *Nucleic Acids Res.* **38**, D227–D233 (2010).
48. 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).
49. 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).
50. 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).
51. 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).
52. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
53. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
54. 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).
55. Sturn, A., Quackenbush, J. & Trajanoski, Z. Genesis: cluster analysis of microarray data. *Bioinformatics* **18**, 207–208 (2002).
56. 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).
57. Ronquist, F. & Huelsenbeck, J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
58. Sanderson, M.J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302 (2003).
59. Lücking, R., Huhndorf, S., Pfister, D.H., Plata, E.R. & Lumbsch, H.T. Fungi evolved right on track. *Mycologia* **101**, 810–822 (2009).
60. Shaner, N.C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.* **22**, 1567–1572 (2004).
61. Szewczyk, E. *et al.* Fusion PCR and gene targeting in *Aspergillus nidulans*. *Nat. Protoc.* **1**, 3111–3120 (2006).
62. 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).