

# Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi

Pamela Gan<sup>1</sup>, Kyoko Ikeda<sup>2</sup>, Hiroki Irieda<sup>2</sup>, Mari Narusaka<sup>3</sup>, Richard J. O'Connell<sup>4</sup>, Yoshihiro Narusaka<sup>3</sup>, Yoshitaka Takano<sup>2</sup>, Yasuyuki Kubo<sup>5</sup> and Ken Shirasu<sup>1</sup>

<sup>1</sup>Plant Science Center, RIKEN, Yokohama, Japan; <sup>2</sup>Graduate School of Agriculture, Kyoto University, Kyoto, Japan; <sup>3</sup>Research Institute for Biological Sciences, Okayama Prefectural Technology Center for Agriculture, Forestry and Fisheries, Okayama, Japan; <sup>4</sup>Max Planck Institute for Plant Breeding Research, Cologne, Germany; <sup>5</sup>Graduate School of Life and Environmental Sciences, Kyoto Prefectural University, Kyoto, Japan

Author for correspondence:

Ken Shirasu

Tel: +81 45 503 9574

Email: ken.shirasu@psc.riken.jp

Received: 28 August 2012

Accepted: 5 November 2012

*New Phytologist* (2013) **197**: 1236–1249

doi: 10.1111/nph.12085

**Key words:** cell wall-degrading enzymes, *Colletotrichum*, effectors, genome, hemibiotrophy, plant pathogenomics, protease, secondary metabolite.

## Summary

- Hemibiotrophic fungal plant pathogens represent a group of agronomically significant disease-causing agents that grow first on living tissue and then cause host death in later, necrotrophic growth. Among these, *Colletotrichum* spp. are devastating pathogens of many crops. Identifying expanded classes of genes in the genomes of phytopathogenic *Colletotrichum*, especially those associated with specific stages of hemibiotrophy, can provide insights on how these pathogens infect a large number of hosts.
- The genomes of *Colletotrichum orbiculare*, which infects cucurbits and *Nicotiana benthamiana*, and *C. gloeosporioides*, which infects a wide range of crops, were sequenced and analyzed, focusing on features with potential roles in pathogenicity. Regulation of *C. orbiculare* gene expression was investigated during infection of *N. benthamiana* using a custom microarray.
- Genes expanded in both genomes compared to other fungi included sequences encoding small, secreted proteins (SSPs), secondary metabolite synthesis genes, proteases and carbohydrate-degrading enzymes. Many SSP and secondary metabolite synthesis genes were upregulated during initial stages of host colonization, whereas the necrotrophic stage of growth is characterized by upregulation of sequences encoding degradative enzymes.
- Hemibiotrophy in *C. orbiculare* is characterized by distinct stage-specific gene expression profiles of expanded classes of potential pathogenicity genes.

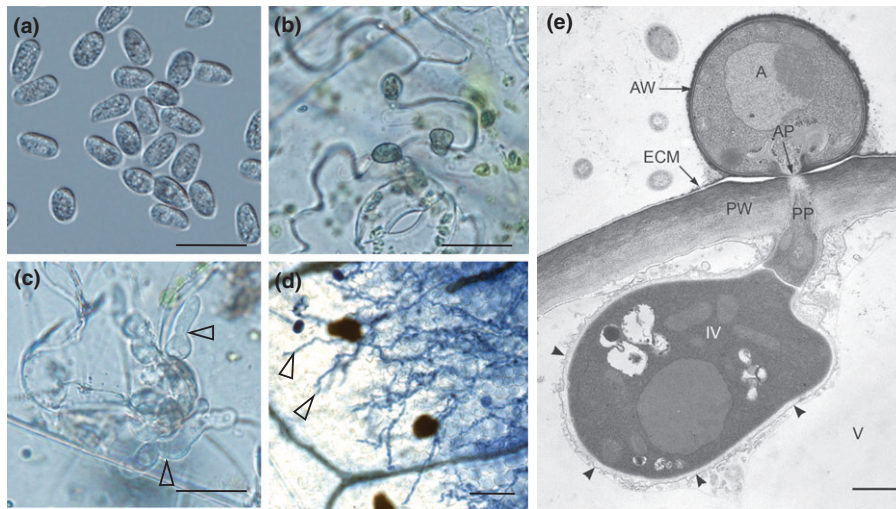
## Introduction

Members of the ascomycete genus *Colletotrichum* represent a group of plant pathogenic fungi that can infect a wide range of plant species including many commercially important crops. One example is *Colletotrichum orbiculare*, which is known as the causal agent of anthracnose on cucurbits including cucumbers, melons and watermelons (Hyde *et al.*, 2009). In addition, *Colletotrichum gloeosporioides* is a notable pathogen associated with > 470 different host species and is commercially significant as the major causal agent of post-harvest disease in fruits such as avocado, banana, mango, coffee and strawberries (Hyde *et al.*, 2009). For example, *C. gloeosporioides* contributes to 80–50% of plant deaths in commercially grown strawberry nurseries and 40–50% of yield losses in strawberry cultivation (Xie *et al.*, 2010).

Due to their economic importance, *Colletotrichum* spp. have become the subject of many studies on fungal pathogenicity (Perfect *et al.*, 1999; Münch *et al.*, 2008). Studies on *Colletotrichum* have revealed that many phytopathogenic *Colletotrichum* spp. adopt a hemibiotrophic lifestyle (Fig. 1). In

compatible interactions, germinating conidia form melanized appressoria that penetrate epidermal cells directly through the cuticle and cell wall. Following penetration, specialized infection vesicles form and begin the growth of biotrophic intracellular hyphae that develop inside living host cells surrounded by an intact host plasma membrane. The infection then enters a necrotrophic stage where the fungus forms morphologically distinct, secondary, necrotrophic hyphae and obtains nutrients from dead host cells (Perfect *et al.*, 1999; Shen *et al.*, 2001). However, the genus is highly diverse, with different subgroups within a single species complex possessing various host ranges and levels of virulence, from destructive pathogens to putative endophytes (Sharma *et al.*, 2011). Thus, study of this group of pathogens presents a unique opportunity to analyse features underlying a diverse range of plant–pathogen interactions.

In the case of other filamentous plant pathogenic hemibiotrophs, colonization and the initial biotrophic interaction with host cells is facilitated by pathogen-encoded small, secreted proteins termed effectors (Sharma *et al.*, 2011). Through such effectors, the pathogen is able to manipulate host metabolism and



**Fig. 1** Life-cycle stages of *Colletotrichum orbiculare*: *C. orbiculare* infection of *Nicotiana benthamiana*. Conidia of *C. orbiculare* (a) germinate and form melanized appressoria to penetrate the epidermal cells of plants (b). The fungus then proliferates as bulbous primary biotrophic hyphae (c) before later developing thin, filamentous secondary hyphae associated with the switch to necrotrophy (d). Images for (b) were taken at 1 d post-inoculation (dpi), for (c) were taken at 3 dpi and (d) at 7 dpi. (e) Biotrophic infection of cucumber epidermal cell by *C. orbiculare*. The appressorium (A) has a thick melanized cell wall (AW) surrounded by adhesive extracellular matrix (ECM). A penetration peg (PP) emerging from the appressorial pore (AP) has produced an infection vesicle (IV) inside the epidermal cell and is surrounded by an intact host plasma membrane (filled arrowheads). Unfilled arrowheads, hyphae. Bars: (a–c) 20  $\mu$ m; (d) 100  $\mu$ m; (e) 1  $\mu$ m. PW, plant cell wall; V, vacuole.

evade host immune responses that may be triggered by perception of conserved microbe-associated molecular patterns (MAMPS) such as chitin. As part of a ‘molecular-arms race’, plants have, in turn, evolved resistance proteins that can recognize specific effectors secreted by pathogens (Jones & Dangl, 2006; Dodds & Rathjen, 2010). Recognition of effectors by resistance proteins results in the death of infected cells thereby restricting the growth of the pathogen.

The tractability of different *Colletotrichum* species to *in vitro* culture and the development of techniques for gene disruption and complementation have facilitated their use as model pathogens (Perfect *et al.*, 1999). In particular, *Colletotrichum higginsianum*, which invades the model plant *Arabidopsis* (Narusaka *et al.*, 2004, 2009; O’Connell *et al.*, 2004), and *C. orbiculare*, which infects *Nicotiana benthamiana* and *Nicotiana tabacum* in addition to cucurbits (Shen *et al.*, 2001), have been well characterized. However, studies to identify effector proteins in commercially important phytopathogenic *Colletotrichum* have lagged behind experiments on pathogen biochemistry, including the analysis of metabolites produced by these fungi, and only a few effector proteins have been characterized so far (Kim *et al.*, 2000; Stephenson *et al.*, 2000; Kleemann *et al.*, 2012; Yoshino *et al.*, 2012). In addition, inventories of putative effectors have been recently predicted from the annotated genomes of *C. higginsianum* and *Colletotrichum graminicola*, which infects maize (O’Connell *et al.*, 2012).

This paper describes the sequencing and comparative annotation of the genomes of *C. orbiculare* 104-T, a strain used extensively as an experimental model for molecular analysis of fungal pathogenicity, and *C. gloeosporioides* strain Nara-gc5, which was isolated from strawberries (Okayama *et al.*, 2007). Transcriptomic analysis was also performed for *C. orbiculare* at different stages

of plant infection, revealing classes of genes and processes important for infection.

## Materials and Methods

### Genomic DNA extraction

Genomic DNA from *Colletotrichum orbiculare* (syn. *Colletotrichum lagenarium*) strain 104-T (MAFF 240422) and *C. gloeosporioides* Nara gc5 isolated from strawberry (Akai & Ishida, 1968; Okayama *et al.*, 2007) was obtained from cultures grown in liquid medium (Yoshimi *et al.*, 2004) using the QIAgen DNAeasy Plant Minikit (Qiagen).

### Sequencing and assembly

*Colletotrichum orbiculare* genomic DNA was sequenced to 55 $\times$  coverage using Illumina Genome Analyzer Ix (34 $\times$  coverage) and Roche 454 (22 $\times$  coverage) sequencers (Takara-Bio Inc., Otsu, Japan). Sequences obtained from 454-pyrosequencing were assembled using the GS De Novo Assembler (Newbler, Roche), followed by alignment of reads from Illumina using the BWA (Li & Durbin, 2009) and Bowtie (Langmead *et al.*, 2009) programs to correct homopolymer errors in the 454 data. Whole-genome shotgun (WGS) sequencing of *C. gloeosporioides* was performed on the Illumina HiSeq 2000 sequencer to 37 $\times$  coverage with 90 bp paired-end reads from a 500 bp insert library (BGI, Beijing, China). After removal of singletons, low quality and adaptor sequences, reads were assembled using SOAPdenovo (Li *et al.*, 2010). The completeness of each assembly was assessed using CEGMA v2.4 (Parra *et al.*, 2007). All sequences were submitted to NCBI under BioProject accession numbers PRJNA171217 and PRJNA171218.

## Repeat finding

*De novo* repeats were identified using RepeatScout (Price *et al.*, 2005) using default criteria. Repeat elements were then filtered to exclude elements that were >50% low-complexity and that occurred <10 times in the respective genome. Consensus sequences containing putative repeat elements were classified using TEclass (Abrusán *et al.*, 2009). Sequences were divided into blocks according to GC content using the program IsoFinder (Oliver *et al.*, 2004). For calculation of RIP indices, dinucleotide frequencies were determined using the RIPCAL program (Hane & Oliver, 2008).

## Gene prediction

Genes were predicted using a combination of Augustus (Stanke *et al.*, 2008), GeneMark-ES (Ter-Hovhannisyán *et al.*, 2008) and Conrad (DeCaprio *et al.*, 2007). GeneMark-ES v2 was trained on the unmasked sequence of *C. orbiculare*, whereas predictions were made using Augustus trained on *Magnaporthe oryzae*. Gene model predictions were compiled using EvidenceModeler (EVM) including evidence from sequenced transcripts from *in vitro*-grown *C. orbiculare* assembled into 96 131 contigs using the Program to Assemble Spliced Alignments feature (Haas *et al.*, 2008). This process was reiterated using Augustus and Conrad trained on the predicted transcripts and on *Neurospora crassa*. For *C. gloeosporioides*, gene predictions were made using GeneMark-ES v2 trained on its own genome sequence and Augustus trained on *N. crassa*, *M. oryzae*, *C. orbiculare* and Conrad trained on *C. orbiculare*. Gene models were compiled using EVM. Nucleotide and amino acid sequences were manually inspected using CLC Genomics Workbench (CLC bio, Tokyo, Japan). Multiple sequence alignments between the *C. orbiculare* and *C. gloeosporioides* genomes were performed and visualized using Geneious (Biomatters Ltd, Auckland, New Zealand).

## Subcellular localization predictions

The predicted subcellular localizations were determined using SignalP (Nielsen *et al.*, 1997) to identify N-terminal signal peptides and then TMHMM (Krogh *et al.*, 2001) and Fungal BIG-PI (Eisenhaber *et al.*, 2004) to exclude sequences with transmembrane domains and GPI-anchors. Sequences were then submitted to TargetP (Emanuelsson *et al.*, 2000) analysis to identify predicted sequences for extracellular localization.

## Functional annotation of predicted genes

Proteins were annotated using BLAST2go on the NCBI nonredundant protein database (Conesa *et al.*, 2005) and using the PFAM server (<http://pfam.sanger.ac.uk/>). Predicted proteins were classified as proteases by querying the MEROPS database (Rawlings *et al.*, 2012) using a BLASTp cut-off *E*-value of 1E-10. Sequences with homology for complete protease domains but with mutated active sites were further excluded as proteases. In addition, transporters were annotated by a BLAST search of the transporter classification database (Saier *et al.*, 2009) using a BLASTp *E*-value cut-off of 1E-5. Potential secondary metabolite

clusters were identified using SMURF (Khaldi *et al.*, 2010). Carbohydrate active enzymes were classified using the dbCAN HMMer-based classification system (Yin *et al.*, 2012) applying an *E*-value cut-off of 10E-5. For comparisons, similar analyses were performed in parallel for other sequenced fungi using sequences of proteins from *N. crassa* OR74A (version 10) *M. oryzae* 70-15 (version 6), *Fusarium oxysporum* f. sp. *lycopersici* 4287 (version 2) and *Fusarium graminearum* ph-1 (version 3) downloaded from the Broad Institute of Harvard and MIT. Sequences were clustered with MCL using BioLayout Express 3D (Theocharidis *et al.*, 2009) using BLASTp with 10E-5 as a cutoff. Syntenic regions were detected using the DAGchainer program (Haas *et al.*, 2004) with settings to search for regions of four or more collinear genes.

## Prediction of noncoding RNAs

Transfer RNAs were predicted using tRNAScan-SE (Lowe & Eddy, 1997) and ribosomal RNAs were predicted using RNAmmer (Lagesen *et al.*, 2007). Other noncoding RNAs were predicted using Erpin v5.4 (Gautheret & Lambert, 2001) that predicts ncRNAs based on RNA primary and secondary structure signatures from the RFAM 9.1 database.

## Microarray analysis of gene expression

A custom 8 × 60 k microarray with 60-mer probes against exons from *C. orbiculare* predicted genes was designed using the eArray custom microarray design tool (<http://earray.chem.agilent.com/earray/>) using the best probe methodology and manufactured by Agilent Technologies. Probes were designed against all predicted *C. orbiculare* exons. In addition, EST sequences of *N. benthamiana* were used as queries in the design step to reduce the likelihood of designing probes that would cross-hybridize with host sequences. Two to three CDS-specific probes were included for each predicted gene. For 15 probes, five replicates were included as internal controls. Further, 19 493 probes against ESTs from *N. benthamiana* were also included. A total of 11 152 fungal transcripts were indicated to be 'detected' in at least one stage. RNA was extracted from ungerminated conidia, infected epidermal peels after 1 d post-inoculation (dpi) and 3 dpi, from infected leaves at 7 dpi and from vegetative mycelia grown *in vitro* on complete media. RNA from uninfected leaves and epidermal peels were extracted as negative controls. Total RNA from three biological replicates was extracted using the Plant RNeasy Mini kit with DNaseI treatment (Qiagen) followed by linear amplification and labelling with Cy3 using the Low Input Quick Amp kit (Agilent Technologies, USA). Arrays were scanned using an Agilent Technologies DNA Microarray Scanner with Scan Control software (Agilent) and data were analysed using Feature Extraction Software (Agilent). Signal intensities from probes against fungal sequences were normalized to the 75th percentile and baseline transformed to the median of all samples using the GeneSpringGX12 software suite (Agilent). Quantitative PCR was performed on selected genes to validate microarray results using the Thunderbird SYBR qPCR mix (ToYoBo). Reactions were run on an Mx3000P QPCR system and analyzed with MxPro QPCR

software (Stratagene). The GEO accession number for the microarray data is GSE39714.

## Results and Discussion

### Genome sequencing and assembly

The *Colletotrichum orbiculare* genome was sequenced using both the Illumina Genome Analyzer IIx and Roche 454 sequencers and assembled into 525 scaffolds (N50: 428.89 kb) with an estimated total size of 88.3 Mb (Table 1), which is larger than the other sequenced *Colletotrichum* genomes (57.4 and 53.4 Mb in *C. graminicola* and *C. higginsianum*, respectively; O'Connell *et al.*, 2012). The *C. gloeosporioides* genome was sequenced on the Illumina Hiseq 2000 sequencer and assembled into 1241 scaffolds (N50: 112.81 kb) with an estimated total size of 55.6 Mb. Homologs to a set of conserved eukaryotic genes (Parra *et al.*, 2007) were identified to provide an estimate of assembly completeness. According to this analysis, the assemblies for *C. orbiculare* and *C. gloeosporioides* cover 97.98% and 96.37% of the total gene space, respectively.

### Extensive expansion of AT-rich sequences in *C. orbiculare*

Approximately 8.3% and 0.75% of the assembled *C. orbiculare* and *C. gloeosporioides* sequences consist of repeat elements with the majority of elements being the relics of transposable elements rather than encoding active protein-coding sequences. Strikingly, in addition to the annotated interspersed repeat elements, the *C. orbiculare* genome is enriched with low complexity sequences with higher concentrations of AT nucleotides. By dividing the genome into blocks based on GC content, 5514 blocks with < 50% GC content (AT blocks) were identified in *C. orbiculare* with an average size of 7.8 kb per block (Table 2). In total, these AT blocks make up 43.4 Mb (49.2%) of the genome, with an average GC% of 19.25% compared to 55.12% in the rest of the genome. In comparison, only 27.4% of the *C. gloeosporioides* genome assembly consists of AT blocks.

**Table 1** General features of the *Colletotrichum orbiculare* and *C. gloeosporioides* genome assemblies

	<i>C. orbiculare</i>	<i>C. gloeosporioides</i>
Assembly size (Mb)	88.3	55.6
Number of scaffolds	525	1241
Median length (N50)	428.89 kb	112.81 kb
Contigs	10545	5335
GC%	37.52%	53.58%
Number of predicted genes	13 479	15 469
GC% of genes	57.09%	55.99%
tRNAs	459	384
rRNAs	58	58
Repeat elements	8.30%	0.75%
Coverage of genome coding region (complete/partial)*	97.98%/98.8%	96.37%/98.8%

\*Genome coding region coverage was estimated by CEGs (Core Eukaryotic Genes) using the CEGMA pipeline (Parra *et al.*, 2007).

**Table 2** Properties of AT and GC blocks in *Colletotrichum orbiculare* and *C. gloeosporioides*

	<i>C. orbiculare</i>		<i>C. gloeosporioides</i>	
	AT block	GC block	AT block	GC block
Average GC (%)	19.25	55.12	48.19	57.07
Median GC (%)	18.28	54.72	48.99	55.74
Total size (bp)	43 416 476	42 166 855	2 014 730	53 388 137
Average size (bp)	7874	48 635	19 188	16 277
Median size (bp)	5052	27 339	12 898	8508
Number of genes	540	12 939	602	14 786
Number of SSP encoding genes	31	669	46	729

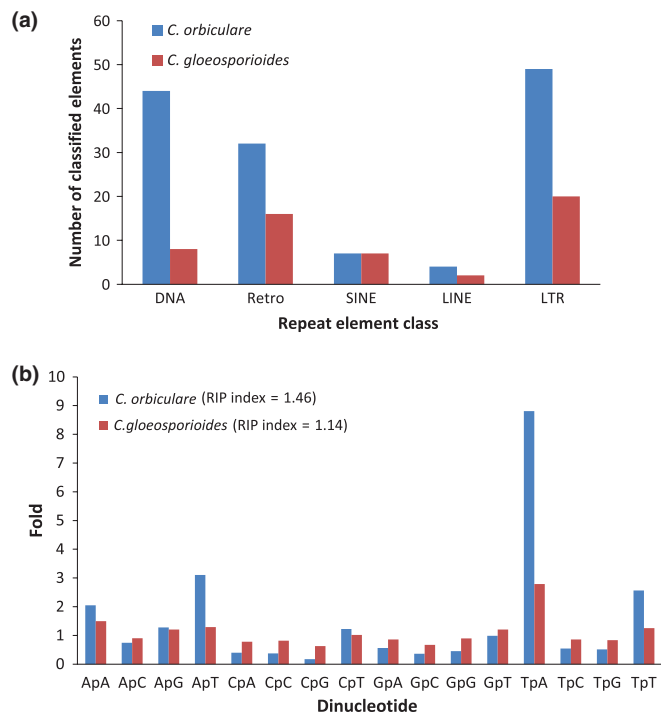
SSP, small, secreted proteins with < 300 amino acids predicted length.

One possible mechanism that could have given rise to the non-homogeneous AT-content of the *C. orbiculare* genome is repeat-induced point mutations (RIP). RIP is a mechanism of defence against transposable elements in ascomycete fungi, in which C and G nucleotides are preferentially mutated to T and A nucleotides in duplicated sequences (Galagan & Selker, 2004). RIP has been hypothesized to generate a similarly compartmentalized genome in the phytopathogenic fungus *Leptosphaeria maculans* (Rouxel *et al.*, 2011). The components required for RIP are conserved in both *Colletotrichum* genomes (Supporting Information Table S1), and RIP indices (TpA/ApT dinucleotide ratios) of the consensus sequences of identified repeats in both genomes indicate RIP activity in both species (Fig. 2). It is also possible that an unknown mechanism is responsible for the expansion of AT-rich sequences.

### Gene content

A total of 13 479 and 15 469 genes are predicted in the *C. orbiculare* and *C. gloeosporioides* genomes, respectively (Table 1), similar to the number of genes predicted in the *C. graminicola* and *C. higginsianum* genomes (12 006 and 16 172, respectively; O'Connell *et al.*, 2012). Thus, despite the larger genome size of *C. orbiculare* relative to the other three genomes, the coding capacity of the genome is not significantly increased. In *C. orbiculare*, genes are predominantly located in GC blocks (306.9 genes per Mb), with AT blocks being largely gene-sparse regions (12.44 genes per Mb).

*Colletotrichum gloeosporioides* possesses dispensable, potentially pathogenicity-associated chromosomes that can be horizontally transferred between compatible strains (Masel, 1996). A partial sequence of one of these chromosomes from another isolate of *C. gloeosporioides* had previously been determined (NCBI Accession AF448489.1) Six scaffolds from the *C. gloeosporioides* genome sequenced in our study show similarity to this sequence. Notably, the gene densities of these scaffolds are below the genome average. Also, only two of these scaffolds have GC contents of below 50% unlike in *Mycosphaerella graminicola* where the GC content of small, dispensable, pathogenicity-related chromosomes are lower than the rest of the genome (Goodwin *et al.*, 2011). The majority of genes identified on these scaffolds encode proteins with no



**Fig. 2** Analysis of repeats in *Colletotrichum orbiculare* and *C. gloeosporioides*. (a) The *C. orbiculare* genome was shown to be enriched with LTR and DNA transposon-like elements compared to the *C. gloeosporioides* genome. (b) Fold differences of the dinucleotide frequencies of *Colletotrichum* repeat elements relative to the control. Nonduplicated genes show that both genomes have an enrichment of TpA and a depletion of CpA dinucleotides, which is characteristic of RIP.

homology to known sequences in the NCBI nonredundant database (Table S2). No *C. orbiculare* scaffolds have significant homology to these sequences, indicating that the region is not conserved and is likely to be species-specific.

The two genomes share many homologous genes, with 8183 (60.7%; *C. orbiculare*) and 8200 (53.0%; *C. gloeosporioides*) genes, respectively, grouping into 8166 orthologous groups as defined by an adaptive reciprocal best blast hit strategy (Lechner *et al.*, 2011). Of these, a total of 6769 (50.2%; *C. orbiculare*) and 6793 (43.9%; *C. gloeosporioides*) genes group together with 6764 orthologous groups from *C. higginsianum* and *C. graminicola*. Furthermore, 6187 (40.0%) *C. gloeosporioides* genes are organized into syntenic blocks that are collinear with four or more *C. orbiculare* genes. Among the *C. orbiculare* genes located in syntenic blocks, the majority are associated with GO terms for primary metabolism and macromolecule biosynthesis, where the conservation of gene order can be important for the regulation of gene co-expression (Fig. S1). A total of 1022 of these collinear genes encode hypothetical proteins identified in other fungal species; however, 59 encode genes with no known homolog in the NCBI nonredundant database.

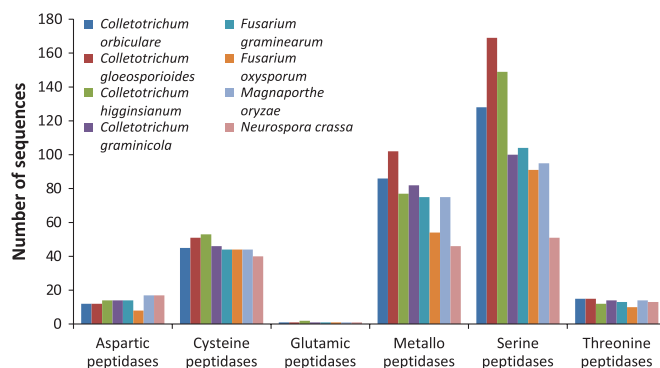
### Expansion of proteases

A total of 287 and 350 protease-encoding sequences are present in *C. orbiculare* and *C. gloeosporioides* respectively. These numbers

of proteases are higher compared to other fungi including other hemibiotrophic plant pathogens such as *F. graminearum* and *M. oryzae* (Fig. 3, Table S3), and is largely due to expansions in the serine protease and metalloprotease families. In *C. orbiculare* and *C. gloeosporioides*, 32% and 36.9% protease sequences were predicted to be secreted. The majority of secreted proteases in both genomes are subtilisins (MEROPS family S8A), a family of serine proteases with optimum activity at alkaline pH. This finding is intriguing given that during *C. gloeosporioides* infections, alkalinisation of host tissue has been associated with virulence (Prusky *et al.*, 2001). Furthermore, subtilisins may be important in penetration and colonization because they could potentially degrade plant cell wall glycoproteins or pathogenesis-related proteins (Olivieri *et al.*, 2002). Subtilisins were similarly found to be expanded in *C. higginsianum* (O'Connell *et al.*, 2012). Interestingly, clustering analysis revealed that some encoded subtilisins have greater homology to plant rather than fungal subtilisins, suggesting the possible acquisition of these genes by horizontal gene transfer from a host (Fig. S2). Furthermore, metalloproteases, especially from the M43 and M20 families, are highly expanded in both *Colletotrichum* species. M43 metalloproteases are also expanded in the genome of the endophyte *Piriformospora indica*, where members are upregulated during colonization of dead roots suggesting a role for these proteases in degrading plant materials (Zuccaro *et al.*, 2011).

### Expansion of carbohydrate active enzymes and LysM-domain containing proteins

The plant cell wall consists mainly of three types of polysaccharides: cellulose; hemicelluloses, consisting of xylans, xyloglucans and mannans; and pectins, consisting of rhamnogalacturonans, homogalacturanans and xylogalacturanans (Caffall & Mohnen, 2009). Carbohydrate active enzymes (CAzymes) that are able to break down these polysaccharides are thus important in establishing infection and also in accessing nutrients during necrotrophic and saprophytic growth. Consistent with this, *C. orbiculare* and *C. gloeosporioides* encode a large arsenal of plant cell



**Fig. 3** Proteases in *Colletotrichum orbiculare* and *C. gloeosporioides*. Number of predicted proteases encoded by the *C. orbiculare* and *C. gloeosporioides* genomes in comparison to other fungi. The *Colletotrichum* genomes have high numbers of putative metalloproteases and serine proteases.

wall-degrading enzymes (327 and 364, respectively; Table S4). Both genomes contained more cellulose-degrading enzymes compared to other ascomycete fungi, similar to *C. higginsianum* and *C. graminicola* (Table 3). In addition, the three dicot-infecting species, *C. orbiculare*, *C. gloeosporioides* and *C. higginsianum*, possess an expanded number of pectin-degrading enzymes, consistent with an adaptation to dicot plant cell walls which contain a greater proportion of pectins compared to monocots (Vogel, 2008). In addition, *C. gloeosporioides* showed an expansion in the number of genes encoding enzymes involved in the digestion of hemicellulose and/or pectin.

Another notable expansion was found in proteins containing the CBM50 chitin-binding module that corresponds to the lysin motif (LysM) domain (Table S4). *C. orbiculare* and *C. gloeosporioides* encode a total of 12 and 20 proteins containing this module, respectively, of which 10 and 15, respectively, are predicted to be secreted. A similar expansion of CBM50 domain proteins was also noted in *C. higginsianum* and *C. graminicola* (O'Connell *et al.*, 2012). This expansion seems specific to LysM-domain containing proteins because the numbers of proteins containing the chitin-binding module CBM18 is similar to those found in *Fusarium* spp (Table S4). Interestingly, whereas the CBM18 proteins are associated with chitin-degrading enzymatic domains, the majority of the LysM-domain containing proteins (11 in *C. orbiculare* and 18 of the *C. gloeosporioides* genes) are not, indicating that these proteins could be involved in protection against chitin-degradation by plant enzymes or may function to sequester chitin fragments released during infection to evade perception by host chitin receptors (de Jonge & Thomma, 2009; de Jonge *et al.*, 2010).

### Greater capacity for secondary metabolite production

Secondary metabolites produced by *Colletotrichum* are known to contribute to pathogenicity. For example, treatment with the

**Table 3** Comparisons of the numbers of plant cell wall-degrading enzymes in *Colletotrichum orbiculare* and *C. gloeosporioides* and other Ascomycete fungi

Fungal species	Cell wall component			
	Pectin	Hemicellulose/ pectin	Hemicellulose	Cellulose
<i>Colletotrichum orbiculare</i>	103	74	33	46
<i>Colletotrichum gloeosporioides</i>	117	90	35	44
<i>Colletotrichum higginsianum</i>	79	59	29	38
<i>Colletotrichum graminicola</i>	46	58	32	42
<i>Fusarium oxysporum</i>	79	81	23	27
<i>Fusarium graminearum</i>	49	60	20	20
<i>Magnaporthe oryzae</i>	21	63	30	35
<i>Neurospora crassa</i>	15	33	13	25

*Colletotrichum* toxin colletotrichin is sufficient to cause anthracnose disease-like symptoms including plasma membrane damage (Duke *et al.*, 1992) and secondary metabolite profiling has been used to distinguish between pathogenic and nonpathogenic isolates of *C. gloeosporioides* (Abang *et al.*, 2009). *C. orbiculare* possesses a number of secondary metabolite backbone-forming proteins, including 24 polyketide synthases (PKS), 1 PKS-like, 11 nonribosomal peptide synthases (NRPS), 9 NRPS-like, 3 PKS-NRPS hybrid backbone synthases and 11 demethylallyl tryptophan synthases (DMAT) (Table 4). *Colletotrichum gloeosporioides* appears to have an even greater capacity for secondary metabolite production with 34 PKS, 10 PKS-like, 14 NRPS, 10 NRPS-like, 6 PKS-NRPS hybrids and 8 DMAT. As reported for *C. higginsianum* and *C. graminicola* (O'Connell *et al.*, 2012), the total number of these key enzymes is much higher than those identified from other fungal genomes such as *M. oryzae* (32) and *F. graminearum* (37) indicating that the *Colletotrichum* genus as a whole has evolved a greater capacity for secondary metabolite production. In addition, both genomes encode a large number of cytochrome P450 proteins, comprising a total of 1.06% and 0.74% of all proteins in *C. gloeosporioides* and *C. orbiculare*, respectively. In other fungi, cytochrome P450 enzymes contribute to both primary and secondary metabolism and are involved in the production of mycotoxins and detoxification of host metabolites (Črešnar & Petrič, 2011).

As in other fungi, many of the genes encoding secondary metabolism key enzymes are organized into clusters together with

**Table 4** Number of genes predicted to be involved in the synthesis of secondary metabolite backbones identified by the secondary metabolite unique regions finder (SMURF)

Organism	PKS	PKS-like	NRPS	NRPS-like	Hybrid	DMAT	Total
<i>Aspergillus nidulans</i>	24	4	11	12	1	6	58
<i>Stagnospora nodorum</i>	12	7	11	5	0	2	37
<i>Sclerotinia sclerotiorum</i>	16	2	5	5	0	1	29
<i>Magnaporthe oryzae</i>	12	3	5	6	3	3	32
<i>Neurospora crassa</i>	6	2	3	3	0	1	15
<i>Fusarium graminearum</i>	14	1	10	11	1	0	37
<i>Ustilago maydis</i>	3	2	3	6	0	2	16
<i>Colletotrichum graminicola</i>	39*	–	7*	–	0	7*	53
<i>Colletotrichum higginsianum</i>	58*	–	12*	–	4	10*	84
<i>Colletotrichum gloeosporioides</i>	34	4	14	10	6	8	76
<i>Colletotrichum orbiculare</i>	24	1	11	9	3	6	54

NRPS, nonribosomal peptide synthase; PKS, polyketide synthase; DMAT, demethylallyl tryptophan synthase.

\*Numbers from manual annotations of the *C. higginsianum* and *C. graminicola* genomes (O'Connell *et al.*, 2012).

other genes that may be involved in secondary metabolism such as transporters, oxidoreductases and cytochrome P450s. A total of 38 such clusters were predicted in *C. orbiculare* and 53 in *C. gloeosporioides*. Of the 38 *C. orbiculare* clusters, 25 are in regions syntenic with *C. gloeosporioides* secondary metabolite gene clusters, as assessed by the conservation of genes (Table S20–23). This analysis showed that *C. orbiculare* has more secondary metabolism gene clusters in common with *C. gloeosporioides* than with *C. higginsianum* or *C. graminicola* (O'Connell *et al.*, 2012). In total, only seven metabolism gene clusters were found to be common between all four species, including *C. higginsianum* cluster 10, which is one of the most highly upregulated gene clusters in *C. higginsianum* during biotrophy (O'Connell *et al.*, 2012).

### Potential host manipulation by phytohormone production

The biosynthesis of the auxin indole acetic acid (IAA) and its intermediates has been experimentally demonstrated in *C. gloeosporioides* and related *Colletotrichum* species to occur mainly via the indole-3-acetamide (IAM) pathway (Robinson *et al.*, 1998; Chung *et al.*, 2003). Homologs of the *iaaM* and *iaaH* genes required for the synthesis of auxin via IAM in *Fusarium proliferatum* ET1 have also recently been identified in *C. gloeosporioides* and *C. graminicola* (Tsavkelova *et al.*, 2012). Interestingly, comparison of the genomic regions containing these genes in the two species revealed a lack of conserved flanking genes. Furthermore, the *iaaM* and *iaaH* genes (here identified as Cggc5\_4414 and Cggc5\_4413) are absent from *C. orbiculare* despite the conservation of flanking genes in the *C. gloeosporioides* genome (Fig. S3). This may reflect their distinct host ranges or independent acquisition of these genes in different *Colletotrichum* lineages. Nevertheless, both genomes contain two genes encoding proteins with homology to members of the auxin efflux carrier superfamily (Cggc5\_1694, Cggc5\_13472, Cob\_5997 and Cob\_6211). Thus, IAA may be synthesized via another intermediate in *C. orbiculare*.

Recently, a secreted chorismate mutase was demonstrated to contribute to *Ustilago maydis* virulence by reducing the available chorismate, a substrate for salicylic acid (SA) biosynthesis in host cells, which affected SA defense in plants (Djamei *et al.*, 2011). Although no secreted chorismate mutases were found in the two *Colletotrichum* genomes, a secreted isochorismatase is present in *C. orbiculare* and two such sequences are predicted in *C. gloeosporioides*. Like chorismate mutase, isochorismatases also metabolize chorismate and, although present in all filamentous ascomycetes, are predicted to be secreted only in phytopathogens (Soanes *et al.*, 2008).

### Comparative secretomes reveal candidate effectors

The secretome of a pathogen includes proteins that are deployed to the host–pathogen interface during invasion of the host and are thus of particular interest for identifying potential virulence factors. A total of 1557 (11.6%) and 2042 (13.2%) proteins in *C. orbiculare* and *C. gloeosporioides*, respectively, are predicted to

be secreted proteins. An analysis of significantly over-represented GO terms reveals that the majority of annotated secreted proteins have enzymatic activity (Fig. S4), especially in carbohydrate and protein degradation. Consistent with this finding, PFAM domain analysis reveals that out of the 335 and 496 PFAM domains identified in the predicted secreted proteins, only 129 and 190 are not associated with enzymatic functions in *C. orbiculare* and *C. gloeosporioides*, respectively (Tables S5–S7).

The *Colletotrichum* secretomes contain homologs of known effectors from other phytopathogens, indicating the use of conserved infection strategies. These proteins include homologs of the necrosis inducing protein (NPP1) from *Phytophthora* spp., the Biotrophy-Associated Secreted (BAS) protein 2 from *M. oryzae* (Mosquera *et al.*, 2009) and the *F. oxysporum* Secreted In Xylem (SIX) 5 protein (Lievens *et al.* 2009). Interestingly, the host range determining *Fusarium* spp. pea pathogenicity protein PEP1 is also present in both species (Coleman *et al.*, 2011). Further, the *C. orbiculare* genome also encodes small proteins similar to SIX1 and SIX6 proteins of *F. oxysporum*. A homolog of the Ave1 avirulence effector in the pathogens *Verticillium dahliae*, *F. oxysporum* and *C. higginsianum* that was hypothesized to have been horizontally transferred from plants (De Jonge *et al.*, 2012) was also detected only in *C. orbiculare* at the end of a region with collinearity to *C. gloeosporioides* (Fig. S5).

Respectively, *Colletotrichum orbiculare* and *C. gloeosporioides* possess 372 and 355 cysteine-rich secreted proteins (cysteine content of 3% or more). This is interesting given that the effectors of many other fungal phytopathogens are cysteine rich and this feature may be important for maintaining protein structure given the relatively small size of these proteins (Stergiopoulos & de Wit, 2009). In comparison, the median cysteine content of all predicted proteins in the *C. orbiculare* and *C. gloeosporioides* genomes are 1.1% and 1.2%, respectively. Notably, the majority of cysteine-rich secreted sequences (318 and 288) are SSPs (< 300 amino acids in length). This is consistent with findings that a large proportion of the candidate-secreted effectors identified in *C. graminicola* and *C. higginsianum* are cysteine rich (O'Connell *et al.*, 2012).

A total of 700 and 755 SSPs are predicted in the *C. orbiculare* and *C. gloeosporioides* genomes respectively. Although some of these sequences are associated with host interaction as predicted secreted proteases and polysaccharide-degrading enzymes, the majority (505 in *C. orbiculare* and 492 in *C. gloeosporioides*) are of unknown function. Of these sequences, 276 (54.6%) in *C. orbiculare* and 243 (49.4%) in *C. gloeosporioides* encoded cysteine-rich proteins. Amongst all the SSPs, only 142 *C. orbiculare* SSPs and 134 *C. gloeosporioides* SSPs could be clustered into tribes containing three or more proteins, representing potential gene families. (Table S8). In both genomes, only four of these genes cluster exclusively with other sequences within the same genome according to TribeMCL analysis. This finding contrasts with the expansion of lineage-specific effectors in the genomes of rust fungi (Duplessis *et al.*, 2011). Possibly this result reflects the diversification of some rust effectors to adapt to specific hosts. Comparison to the other two sequenced *Colletotrichum* species revealed that 195 (27.9%) *C. orbiculare* SSPs and 208 (26.8%)

*C. gloeosporioides* SSPs were conserved in all four species (Table 5). Interestingly, only 279 *C. orbiculare* SSPs possessed homologs in *C. higginsianum*. This number was similar to the number of SSPs with homologs in *C. graminicola* despite the fact that *C. graminicola* has a different host range, infecting only monocot plants. By contrast, 410 SSPs in *C. orbiculare* were found to have homologs in *C. gloeosporioides*. This difference may reflect a closer phylogenetic relationship between *C. orbiculare* and *C. gloeosporioides* relative to the other two species, which themselves are more closely related, belonging to the graminicola and destructivum sister clades (Cannon *et al.*, 2012).

### *Colletotrichum orbiculare* transcriptome analysis reveals a significant hemibiotrophic stage shift

A custom microarray was designed to test for the expression of the annotated *C. orbiculare* genes *in planta* at different stages of infection on *N. benthamiana* leaves. The sampled stages (Fig. 1a–d) were as follows: ungerminated conidia, appressorial penetration/early infection (1 dpi), established biotrophic growth (3 dpi), late necrotrophy (7 dpi), as well as *in vitro*-grown hyphae. A total of 9651 genes were differentially regulated (fold change  $\geq 2$ ,  $P \leq 0.05$ ) in conidia, 1, 3 and 7 dpi compared to *in vitro*-grown hyphae.

In *L. maculans*, effector genes are associated with AT blocks, and many SSPs within these blocks are upregulated *in planta*. Given the similar compartmentalization of the *C. orbiculare* genome and the greater representation of novel SSP genes in AT-rich regions, we analyzed the expression of genes according to their localization in AT or GC blocks. This revealed that a greater proportion of genes located in AT-rich regions were not upregulated *in planta* compared to *in vitro* (Table 6). This was also observed for the genes encoding SSPs within these regions, suggesting the suppression of genes within AT-rich compartments.

GO terms that are over-represented among the sequences upregulated *in planta* compared to *in vitro*, are associated with carbohydrate binding and degradation, interactions with the host, protein degradation and transmembrane transport (Tables

**Table 5** Features of small, secreted proteins (SSPs), < 300 amino acids

	<i>C. orbiculare</i>	<i>C. gloeosporioides</i>
Secreted proteins	1557	2042
Small, secreted proteins	700	775
With homologs in <i>C. orbiculare</i>	–	419
With homologs in <i>C. gloeosporioides</i>	410	–
With homologs in <i>C. higginsianum</i>	279	277
With homologs in <i>C. graminicola</i>	270	290
With homologs in <i>C. higginsianum</i> , <i>C. gloeosporioides</i> and <i>C. orbiculare</i>	239	251
With homologs in <i>C. orbiculare</i> , <i>C. gloeosporioides</i> , <i>C. higginsianum</i> and <i>C. graminicola</i> .	195	208

BLAST was performed on SSPs from the genomes of *Colletotrichum orbiculare*, *C. gloeosporioides*, *C. higginsianum* and *C. graminicola* (Cut-off = 10E-5).

**Table 6** Percentages of genes in AT blocks and GC blocks upregulated > 10-fold in conidia and at different time points during infection of *Colletotrichum orbiculare* of *Nicotiana benthamiana* compared to *in vitro*

	All genes		SSPs	
	AT block	GC block	AT block	GC block
Upregulated > 10 times at 1 dpi	3.54%	9.91%	7.41%	21.82%
Upregulated > 10 times at 3 dpi	2.59%	5.47%	7.41%	19.73%
Upregulated > 10 times at 7 dpi	3.07%	9.24%	11.11%	14.04%
Upregulated > 10 times in conidia	3.54%	7.02%	3.70%	7.97%

dpi, days post-inoculation; SSPs, small, secreted proteins.

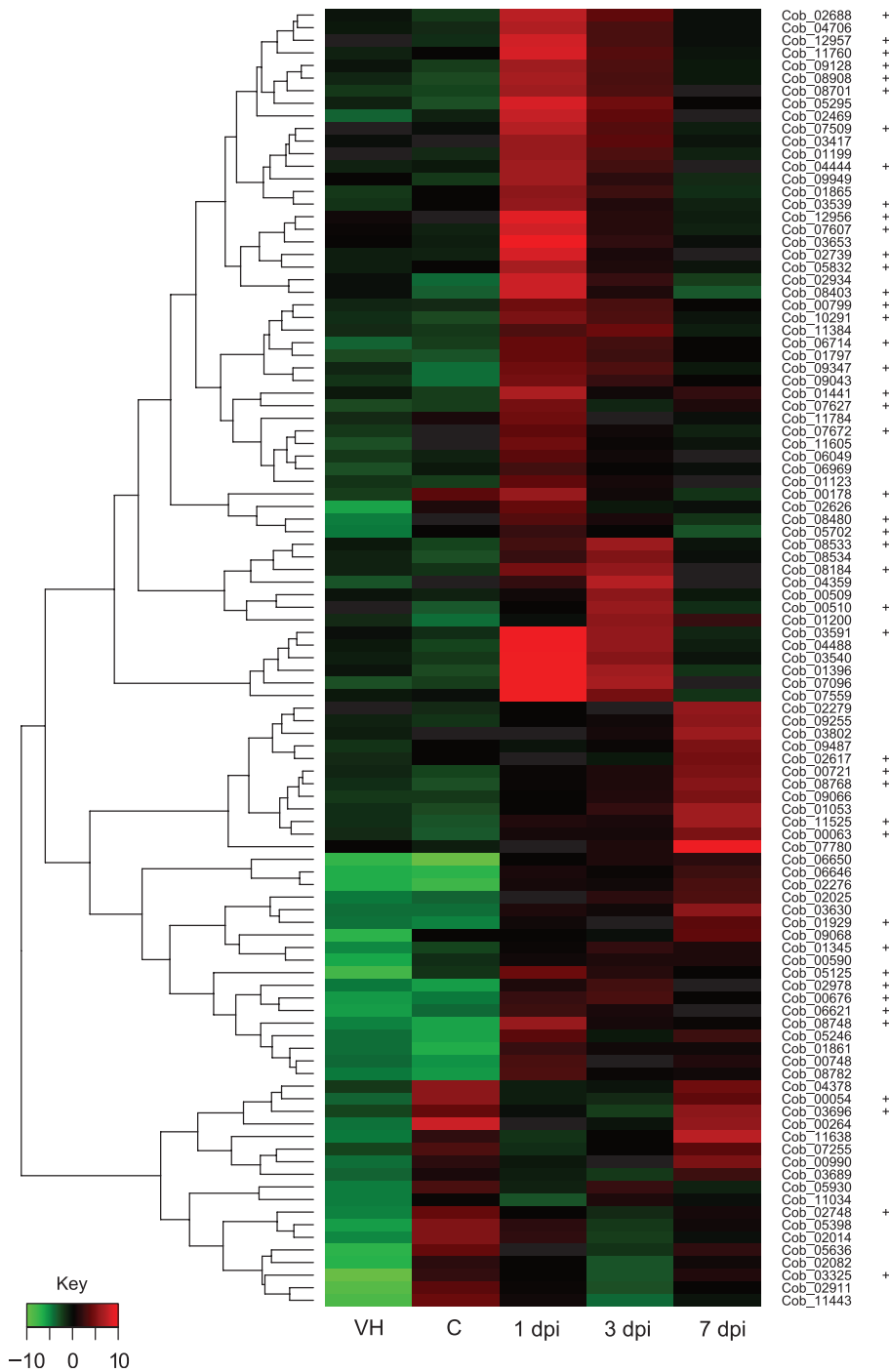
S9–12). A total of 405 SSPs were upregulated > 2-fold *in planta* compared to *in vitro* ( $P$ -value  $\leq 0.05$ ). Of these, 291 were not associated with any PFAM domain. Further, 165 genes have no known homolog in the NCBI database, indicating that these genes may represent truly novel effectors. The expression patterns of nine selected SSPs were assessed by QPCR (Fig. S6) confirming their individual expression profiles *in planta*.

Analysis of the transcriptome revealed that the initial colonization stage (1 dpi) is characterized by the upregulation of gene models encoding SSPs (Fig. 4). A total of 28 of the top 100 genes expressed are SSP-encoding genes. In comparison, only 11, 7 and 12 of the top 100 genes expressed in late necrotrophy, conidia and *in vitro* hyphae, respectively, encode SSPs (Tables S13–S17). Among 69 SSP sequences that were highly upregulated (> 50-fold) during early infection (1 dpi), only 25 were also induced to similar levels at later stages of infection (3 and 7 dpi), indicating that the majority of highly expressed SSP genes are stage-specific. By contrast, relatively few SSP-encoding genes were specifically induced at 3 and 7 dpi. SSPs that had peak expression during early infection (1 dpi) included sequences with homology to the effectors of other phytopathogens such as *Fusarium* SIX proteins, the *Magnaporthe* BAS2 protein, as well as two NPP1 domain-containing proteins (Table S18). However, as in *C. higginsianum* (Kleemann *et al.*, 2012), these two NPP1 homologs do not encode proteins with well-conserved necrosis-inducing consensus motifs.

Among the 10 genes encoding secreted chitin-binding proteins, four had higher expression *in planta* at the early invasion stage (1 dpi), when they may interfere with the perception of chitin degradation products by host chitin receptors. In addition, three out of four chitin deacetylases (Cob\_748, Cob\_1749 and Cob\_5573) were also upregulated from this early stage of infection onwards. These proteins may potentially function in modifying chitin in fungal cell walls into chitosan to evade host defense responses, as hypothesized for *C. graminicola* (El Gueddari *et al.*, 2002).

Examination of sequences with homology to MFS transporters revealed that 49 genes (55.7%) were upregulated > 2-fold during penetration/early infection (Table S19). Among these upregulated genes were seven sequences related to hexose transport and six sequences with homology to transporters of other monosaccharides. Other transporters upregulated *in planta* included proteins related to quinate transport. Four quinate transporters were



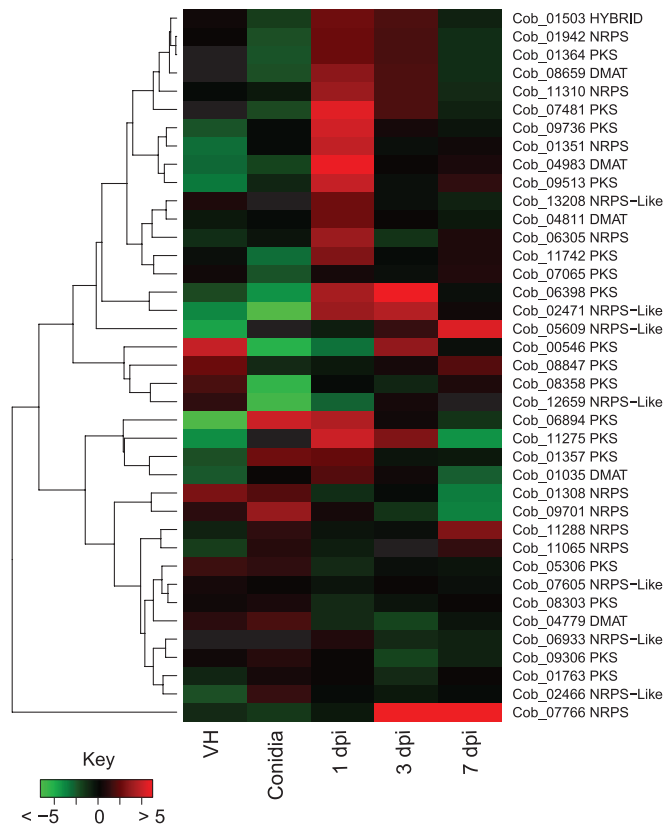


**Fig. 4** Expression of small, secreted proteins (SSPs) *in planta*. Heatmap of SSPs with at least a 50-fold upregulation *in planta* compared to *in vitro*-grown hyphae (VH). This analysis revealed the stage-specific expression of SSPs during infection, particularly during early host invasion (1 d post-inoculation (dpi)). The levels of expression represent normalized probe intensity values ( $\log_2$ ). Genes with similar patterns of expression are clustered. Cysteine-rich sequences are denoted with '+' after the gene model name. C, conidia

upregulated > two-fold during initial invasion and biotrophic hyphal growth, but during necrotrophy only one such transporter was upregulated > two-fold. Quinate is a polyol present in decaying plant material that can be used as a carbon source by fungi such as *N. crassa* (Giles *et al.*, 1985). A recent study suggests that in the early stages of *M. oryzae* infection, host quinate concentrations are manipulated by the pathogen at the expense of the host shikimate pathway, which synthesizes defense-related phenylpropanoids (Parker *et al.*, 2009; Soanes *et al.*, 2012). The expression pattern of genes in *C. orbiculare* encoding quinate

transporters is consistent with the importance of quinate utilization during early infection, although genes in the quinate utilization cluster of *C. orbiculare* had peak expression during late necrotrophy. This finding indicated that quinate may also be important as a carbon source on dying host tissue.

Interestingly, the early invasion stage (1 dpi; Fig. 1b) is also characterized by the upregulation of many secondary metabolite backbone synthesis genes (Fig. 5). Although expression of secondary metabolite synthesis genes is commonly associated with necrotrophic fungi (Pinedo *et al.*, 2008; Amselem *et al.*, 2011),



**Fig. 5** Expression of secondary metabolite backbone synthesis genes *in planta*. Heatmap of secondary metabolite backbone synthesis genes showing differential expression *in planta* and in conidia compared to *in vitro*-grown hyphae (VH). This analysis revealed the increased expression of secondary metabolite backbone synthesis genes during early invasion. The levels of expression represent normalized probe intensity values (log<sub>2</sub>). Genes with similar patterns of expression are clustered. dpi, days post-inoculation.

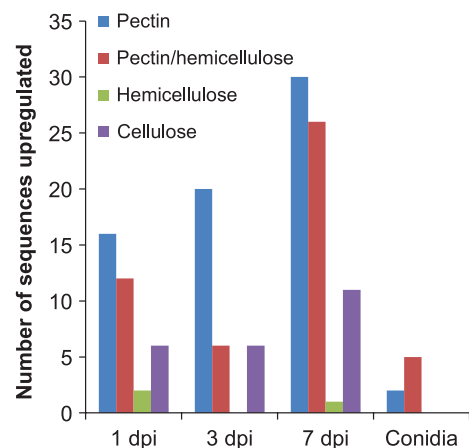
among the 54 *C. orbiculare* secondary metabolite backbone synthesis genes, 22.2% were upregulated >10-fold at this stage compared to *in vitro* hyphae, four of which did not have detectable expression *in vitro*. By contrast, only 9.3% secondary metabolism genes were upregulated to the same extent during the necrotrophic stage. This result indicates that the products of these enzymes are not necessarily toxins and that the fungus may be manipulating its host via secondary metabolites during biotrophic growth. Secondary metabolite backbone synthesis genes were also among the top 100 most highly expressed genes during early penetration and biotrophic growth in *C. higginsianum* (O'Connell *et al.*, 2012), suggesting that this strategy may be common in other *Colletotrichum* species. Alternatively, these secondary metabolites may play important roles within the fungus to aid infection. One gene that falls into this category is the *PKS1* gene (Cob\_9513), required for melanin biosynthesis, which was upregulated >90 times at this stage. This result is consistent with the importance of melanin in appressorial penetration of host epidermal cells.

The 3 dpi time point corresponds to a period of intracellular biotrophic hyphal growth (Fig. 1c). Transcriptionally, this stage is similar to the early invasion stage although in general genes

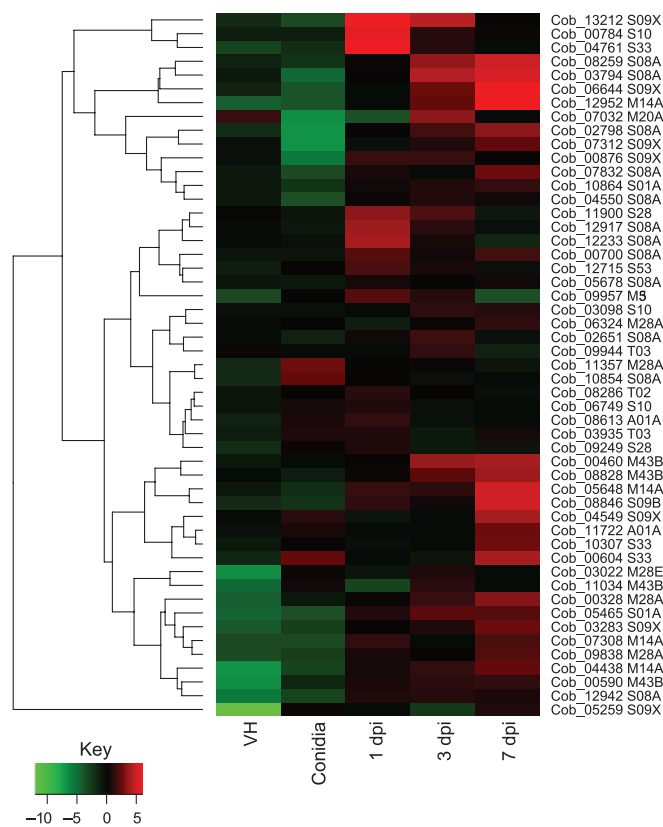
were not as highly upregulated relative to *in vitro*-grown hyphae. As in the early invasion stage, a large number of SSPs were highly expressed. In particular, 32 of the top 100 expressed genes were SSPs (Table S14). Furthermore, as in the earlier stage of infection (1 dpi), many secondary metabolite backbone synthesis genes were also upregulated at 3 dpi. Interestingly, these included a gene with homology to an NRPS, one of the most highly expressed genes during biotrophic hyphal growth. This gene is in a predicted secondary metabolite cluster and another member of the cluster, encoding an MFS drug efflux transporter, was also upregulated at this phase suggesting that the product of the enzyme may be targeted to the host. This cluster was not identified in the other three *Colletotrichum* species, indicating that the product could be lineage-specific to *C. orbiculare*.

The necrotrophic phase of growth at 7 dpi (Fig. 1d) was notable for the expression of degradative enzymes such as carbohydrate-degrading enzymes and proteases. Specifically, in contrast to the early invasion and biotrophic phases of growth, where only 8% of genes encoding plant cell wall-degrading enzymes were upregulated 10-fold or more compared to *in vitro* hyphae, the proportion of such enzymes upregulated at 7 dpi was 15%. In particular, this increase can be attributed to the upregulation of pectin-degrading enzyme sequences, consistent with the abundance of pectins in dicot cell walls (Fig. 6).

Analysis of the expression profiles of genes encoding secreted proteases in *C. orbiculare* revealed that although some proteases were expressed from early infection at 1 dpi, many sequences had peak expression only at the necrotrophic stage (Fig. 7). Furthermore, the most upregulated protease-encoding sequences *in planta* belong to the metallo- and serine-protease families. Three out of five zinc carboxypeptidases from the M14 family were upregulated >10-fold *in planta* from 1 dpi, with expression peaking at the late necrotrophic stage. In addition, four out of five M43 metalloproteases were upregulated >10-fold during infection, with expression also peaking during late necrotrophy,



**Fig. 6** Plant cell wall-degrading genes upregulated *in planta* compared to *in vitro*-grown hyphae. Number of predicted plant cell wall-degrading enzyme genes upregulated 10-fold or more compared to *in vitro*-grown hyphae at different time points during infection of *Nicotiana benthamiana* and in conidia, classified according to carbohydrate substrate. dpi, days post-inoculation.



**Fig. 7** Expression of proteases *in planta*. Heatmap of proteases showing at least two-fold upregulation *in planta* compared to *in vitro*-grown hyphae (VH). The expression of proteases during infection shows that the majority of upregulated protease-encoding genes have peak expression during late necrotrophy. The levels of expression represent normalized probe intensity values (log<sub>2</sub>). Genes with similar patterns of expression are clustered. The putative class of each protease is indicated after the gene ID. dpi, days post-inoculation.

consistent with roles in the breakdown of plant material and the drastic change from biotrophic to necrotrophic growth. However, this expression pattern does not apply to all protease-encoding genes. For example, among the 12 subtilisin genes that were upregulated two-fold or more *in planta*, four showed highest expression during early infection, three showed peak expression during biotrophic growth, whereas three others were preferentially expressed during necrotrophy. Such contrasting expression profiles suggest that particular subtilisins may have specific roles at different points in infection.

In *C. gloeosporioides* and *Colletotrichum coccodes* infections, accumulation and secretion of ammonia are important for virulence through the alkalization of host tissue (Prusky *et al.*, 2001; Alkan *et al.*, 2008; Miyara *et al.*, 2010). Analysis of predicted ammonia transporter genes provides evidence for the upregulation of all four genes *in planta* relative to *in vitro*-grown hyphae (Table S19); however, the highest expression levels of all four genes were at the conidial stage. Appressorium formation by germinating *C. gloeosporioides* conidia is stimulated by exogenous ammonium (Miyara *et al.*, 2010) and the high expression of ammonium transporters in *C. orbiculare* conidia may reflect the

importance of the perception and accumulation of external nitrogen sources before conidial germination.

### Concluding remarks

Recent studies have indicated that genome expansion via the action of repeat elements in filamentous plant pathogens can drive the replication and diversification of pathogen-encoded virulence factors. This can result in irregular organization of the genome, as well as the enrichment of virulence factors within regions or compartments of the genome (Raffaele & Kamoun, 2012). One of the remarkable findings of our study is that within the same fungal genus, the genome organizations of two *Colletotrichum* pathogens are quite different. The *C. orbiculare* genome is significantly expanded relative to that of *C. gloeosporioides* and is organized into blocks of sequences with strikingly distinct AT and GC contents, with gene-sparse AT blocks potentially arising as the result of transposable element inactivation via RIP. In *L. maculans*, RIP-affected regions are sources of novel effectors where the mutation of duplicated sequences drives effector diversification (Rouxel *et al.*, 2011). Although the *C. orbiculare* genes located within AT blocks had higher sequence variation compared to the rest of the genome, as indicated by the lack of known homologs, very few of these genes are highly upregulated *in planta*, suggesting that these genomic regions are not enriched with effector genes. However, we cannot exclude the possibility that these genes are expressed in hosts other than *N. benthamiana*.

The genome-wide comparative analyses presented here revealed a number of genes that are potentially involved in pathogenesis. In particular, gene families encoding proteases and carbohydrate-degrading enzymes that are able to target a wide range of substrates, are highly expanded, possibly reflecting the broad host ranges of both species. The presence of large numbers of secondary metabolism genes, similar to the genomes of *C. higginsianum* and *C. graminicola* (O'Connell *et al.*, 2012), highlights the potential capacity of *Colletotrichum* species metabolites for host manipulation and/or detoxification of host antifungal compounds. In particular, the upregulation of *C. orbiculare* secondary metabolism-related genes during early infection of *N. benthamiana* coupled with the observation that many such genes are also upregulated in *C. higginsianum* during biotrophy (O'Connell *et al.*, 2012), suggests that small molecules are important in the establishment and/or maintenance of biotrophy. A number of SSP effector candidates are also present in the two genomes described in this study. Different expression profiles of secreted proteins during infection were observed, with more SSPs upregulated during the establishment of biotrophy followed by a shift to the upregulation of genes associated with degradative enzymes during necrotrophic growth. A similar shift in expression patterns was also noted in other *Colletotrichum* species, particularly in *C. higginsianum* (O'Connell *et al.*, 2012), underscoring the importance of these different groups of genes for the different stages of hemibiotrophic pathogen growth. Importantly, the list of SSPs that are highly upregulated *in planta* represents a base for future studies on effector biology of *C. orbiculare*. Finally, the genomic information presented in this study will be a

valuable resource for relating data from functional studies to pathogen genetics. This information is especially significant because *C. gloeosporioides* and *C. orbiculare* not only serve as model systems for plant–pathogen interactions, but also represent two economically significant pathogens where a deeper understanding of pathogen biology can directly inform strategies for targeted disease control.

## Acknowledgements

This work was supported by the Programme for Promotion of Basic and Applied Researches for Innovations in Bio-oriented Industry to K.S., Y.T., Y.N.; Grant-in-Aid for Scientific Research (KAKENHI) (24228008 to K.S., 21380031 to Y.K.) and by The Strategic Research Funds in Kyoto Prefectural University to Y.K. We thank Yoshihiko Hirayama for kindly providing *C. gloeosporioides* strain Nara gc5.

## References

- Abang MM, Abraham WR, Asiedu R, Hoffmann P, Wolf G, Winter S. 2009. Secondary metabolite profile and phytotoxic activity of genetically distinct forms of *Colletotrichum gloeosporioides* from yam (*Dioscorea* spp.). *Mycological Research* 113: 130–140.
- Abrusán G, Grundmann N, DeMester L, Makalowski W. 2009. TEclass – a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics* 25: 1329–1330.
- Akai S, Ishida N. 1968. An electron microscopic observation on the germination of conidia of *Colletotrichum lagenarium*. *Mycopathologia* 34: 337–345.
- Alkan N, Fluhr R, Sherman A, Prusky D. 2008. Role of ammonia secretion and pH modulation on pathogenicity of *Colletotrichum coccodes* on tomato fruit. *Molecular Plant-Microbe Interactions* 21: 1058–1066.
- Amselem J, Cuomo CA, van Kan JAL, Viaud M, Benito EP, Couloux A, Coutinho PM, de Vries RP, Dyer PS, Fillinger S *et al.* 2011. Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genetics* 7: e1002230.
- Caffall KH, Mohnen D. 2009. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydrate Research* 344: 1879–1900.
- Cannon PF, Damm U, Johnston PR, Weir BS. 2012. *Colletotrichum* – current status and future directions. *Studies in Mycology* 73: 181–213.
- Chung KR, Shilts T, Ertürk Ü, Timmer LW, Ueng PP. 2003. Indole derivatives produced by the fungus *Colletotrichum acutatum* causing lime anthracnose and postbloom fruit drop of citrus. *FEMS Microbiology Letters* 226: 23–30.
- Coleman JJ, Wasmann CC, Usami T, White GJ, Temporini ED, McCluskey K, VanEtten HD. 2011. Characterization of the gene encoding pisatin demethylase (FoPDA1) in *Fusarium oxysporum*. *Molecular Plant-Microbe Interactions* 24: 1482–1491.
- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676.
- Črešnar B, Petrič Š. 2011. Cytochrome P450 enzymes in the fungal kingdom. *Biochimica et Biophysica Acta (BBA)* 1814: 29–35.
- De Jonge R, Peter Van Esse H, Maruthachalam K, Bolton MD, Santhanam P, Saber MK, Zhang Z, Usami T, Lievens B, Subbarao KV *et al.* 2012. Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *Proceedings of the National Academy of Sciences, USA* 109: 5110–5115.
- DeCaprio D, Vinson JP, Pearson MD, Montgomery P, Doherty M, Galagan JE. 2007. Conrad: gene prediction using conditional random fields. *Genome Research* 17: 1389–1398.
- Djamei A, Schipper K, Rabe F, Ghosh A, Vincon V, Kahnt J, Osorio S, Tohge T, Fernie AR, Feussner I *et al.* 2011. Metabolic priming by a secreted fungal effector. *Nature* 478: 395–398.
- Dodds PN, Rathjen JP. 2010. Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature Reviews Genetics* 11: 539–548.
- Duke SO, Gohbara M, Paul RN, Duke MV. 1992. Colletotrichin causes rapid membrane damage to plant cells. *Journal of Phytopathology* 134: 289–305.
- Duplessis S, Cuomo CA, Lin Y-C, Aerts A, Tisserant E, Veneault-Fourrey C, Joly DL, Hacquard S, Amselem J, Cantarel BL *et al.* 2011. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences, USA* 108: 9166–9171.
- Eisenhaber B, Schneider G, Wildpaner M, Eisenhaber F. 2004. A sensitive predictor for potential GPI lipid modification sites in fungal protein sequences and its application to genome-wide studies for *Aspergillus nidulans*, *Candida albicans*, *Neurospora crassa*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *Journal of Molecular Biology* 337: 243–253.
- El Gueddari NE, Rauchhaus U, Moerschbacher BM, Deising HB. 2002. Developmentally regulated conversion of surface-exposed chitin to chitosan in cell walls of plant pathogenic fungi. *New Phytologist* 156: 103–112.
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* 300: 1005–1016.
- Galagan JE, Selker EU. 2004. RIP: the evolutionary cost of genome defense. *Trends in Genetics* 20: 417–423.
- Gautheret D, Lambert A. 2001. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *Journal of Molecular Biology* 313: 1003–1011.
- Giles NH, Case ME, Baum J, Geever R, Huiet L, Patel V, Tyler B. 1985. Gene organization and regulation in the qa (quinic acid) gene cluster of *Neurospora crassa*. *Microbiological Reviews* 49: 338–358.
- Goodwin SB, Ben M'Barek S, Dhillon B, Wittenberg AHJ, Crane CF, Hane JK, Foster AJ, Van der Lee TAJ, Grimwood J, Aerts A *et al.* 2011. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genetics* 7: e1002070.
- Haas BJ, Delcher AL, Wortman JR, Salzberg SL. 2004. DAGchainer: a tool for mining segmental genome duplications and synteny. *Bioinformatics* 20: 3643–3646.
- Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biology* 9: R7.
- Hane J, Oliver R. 2008. RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. *BMC Bioinformatics* 9: 478.
- Hyde K, Cai L, Cannon P, Crouch J, Crous P, Damm U, Goodwin P, Chen H, Johnston P, Jones E *et al.* 2009. *Colletotrichum* – names in current use. *Fungal Divers* 39: 147–182.
- Jones JDG, Dangl JL. 2006. The plant immune system. *Nature* 444: 323–329.
- de Jonge R, Thomma BPHJ. 2009. Fungal LysM effectors: extinguishers of host immunity? *Trends in Microbiology* 17: 151–157.
- de Jonge R, van Esse HP, Kombrink A, Shinya T, Desaki Y, Bours R, van der Krol S, Shibuya N, Joosten MHAJ, Thomma BPHJ. 2010. Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. *Science* 329: 953–955.
- Khalidi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, Fedorova ND. 2010. SMURF: genomic mapping of fungal secondary metabolite clusters. *Fungal Genetics and Biology* 47: 736–741.
- Kim Y-K, Liu Z-M, Li D, Kolattukudy PE. 2000. Two novel genes induced by hard-surface contact of *Colletotrichum gloeosporioides* conidia. *Journal of Bacteriology* 182: 4688–4695.
- Kleemann J, Rincon-Rivera LJ, Takahara H, Neumann U, van Themaat EVL, van der Does HC, Hacquard S, Stüber K, Will I, Schmalenbach W *et al.* 2012. Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. *PLoS Pathogens* 8: e1002643.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* 305: 567–580.

- Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. 2007. RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research* 35: 3100–3108.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10: R25.
- Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ. 2011. Proteinortho: detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics* 12: 124.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25: 1754–1760.
- Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K *et al.* 2010. *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Research* 20: 265–272.
- Lievens B, Houterman PM, Rep M. 2009. Effector gene screening allows unambiguous identification of *Fusarium oxysporum* f. sp. *lycopersici* races and discrimination from other formae speciales. *FEMS Microbiology Letters* 300: 201–215.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* 25: 955–964.
- Masel AM. 1996. Molecular evidence for chromosome transfer between biotypes of *Colletotrichum gloeosporioides*. *Molecular Plant-Microbe Interactions* 9: 339.
- Miyara I, Shafran H, Davidzon M, Sherman A, Prusky D. 2010. pH Regulation of ammonia secretion by *Colletotrichum gloeosporioides* and its effect on appressorium formation and pathogenicity. *Molecular Plant-Microbe Interactions* 23: 304–316.
- Mosquera G, Giraldo MC, Khang CH, Coughlan S, Valent B. 2009. Interaction transcriptome analysis identifies *Magnaporthe oryzae* BAS1-4 as biotrophy-associated secreted proteins in rice blast disease. *The Plant Cell Online* 21: 1273–1290.
- Münch S, Lingner U, Floss DS, Ludwig N, Sauer N, Deising HB. 2008. The hemibiotrophic lifestyle of *Colletotrichum* species. *Journal of Plant Physiology* 165: 41–51.
- Narusaka Y, Narusaka M, Park P, Kubo Y, Hirayama T, Seki M, Shiraishi T, Ishida J, Nakashima M, Enju A *et al.* 2004. RCH1, a locus in Arabidopsis that confers resistance to the hemibiotrophic fungal pathogen *Colletotrichum higginsianum*. *Molecular Plant-Microbe Interactions* 17: 749–762.
- Narusaka M, Shirasu K, Noutoshi Y, Kubo Y, Shiraishi T, Iwabuchi M, Narusaka Y. 2009. RRS1 and RPS4 provide a dual Resistance-gene system against fungal and bacterial pathogens. *Plant Journal* 60: 218–226.
- Nielsen H, Engelbrecht J, Brunak S, von Heijne G. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* 10: 1–6.
- O’Connell R, Herbert C, Sreenivasaprasad S, Khatib M, Esquerré-Tugayé M-T, Dumas B. 2004. A novel Arabidopsis–*Colletotrichum* pathosystem for the molecular dissection of plant-fungal interactions. *Molecular Plant-Microbe Interactions* 17: 272–282.
- O’Connell RJ, Thon MR, Hacquard S, Amyotte SG, Kleemann J, Torres MF, Damm U, Buiaite EA, Epstein L, Alkan N *et al.* 2012. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nature Genetics* 44: 1060–1065.
- Okayama K, Hirayama Y, Nishizaki M. 2007. Detection of benomyl-resistant *Glomerella cingulata* from latently infected strawberry [*Fragaria ananassa*] plants and infested nursery medium using a semiselective medium. *Japanese Journal of Phytopathology* 73: 155–161.
- Oliver JL, Carpena P, Hackenberg M, Bernaola-Galvan P. 2004. IsoFinder: computational prediction of isochores in genome sequences. *Nucleic Acids Research* 32: W287–W292.
- Olivieri F, Eugenia Zanetti M, Oliva CR, Covarrubias AA, Casalongué CA. 2002. Characterization of an extracellular serine protease of *Fusarium eumartii* and its action on pathogenesis related proteins. *European Journal of Plant Pathology* 108: 63–72.
- Parker D, Beckmann M, Zubair H, Enot DP, Caracuel-Rios Z, Overy DP, Snowdon S, Talbot NJ, Draper J. 2009. Metabolomic analysis reveals a common pattern of metabolic re-programming during invasion of three host plant species by *Magnaporthe grisea*. *The Plant Journal* 59: 723–737.
- Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23: 1061–1067.
- Perfect SE, Hughes HB, O’Connell RJ, Green JR. 1999. *Colletotrichum*: a model genus for studies on pathology and fungal–plant interactions. *Fungal Genetics and Biology* 27: 186–198.
- Pinedo C, Wang C-M, Pradier J-M, Dalmais B, Choquer M, Pêcheur PL, Morgant G, Collado IG, Cane DE, Viaud M. 2008. The sesquiterpene synthase from the Botrydial biosynthetic gene cluster of the phytopathogen *Botrytis cinerea*. *ACS Chemical Biology* 3: 791–801.
- Price AL, Jones NC, Pevzner PA. 2005. *De novo* identification of repeat families in large genomes. *Bioinformatics* 21: i351–i358.
- Prusky D, McEvoy JL, Leverenz B, Conway WS. 2001. Local modulation of host pH by *Colletotrichum* species as a mechanism to increase virulence. *Molecular Plant-Microbe Interactions* 14: 1105–1113.
- Raffaële S, Kamoun S. 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology* 10: 417–430.
- Rawlings ND, Barrett AJ, Bateman A. 2012. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research* 40: D343–D350.
- Robinson M, Riou J, Sharon A. 1998. Indole-3-acetic acid biosynthesis in *Colletotrichum gloeosporioides* f. sp. *aeschyromene*. *Applied and Environmental Microbiology* 64: 5030–5032.
- Rouxel T, Grandaubert J, Hane JK, Hoede C, van de Wouw AP, Couloux A, Dominguez V, Anthouard V, Bally P, Bourras S *et al.* 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nature Communication* 2: 202.
- Saier MH Jr, Yen MR, Noto K, Tamang DG, Elkan C. 2009. The Transporter Classification Database: recent advances. *Nucleic Acids Research* 37: D274–D278.
- Sharma KR, Bhagya N, Sheik S, Samhitha M. 2011. Isolation of endophytic *Colletotrichum gloeosporioides* Penz. from *Salacia chinensis* and its antifungal sensitivity. *Journal of Phytochemistry* 3: 20–22.
- Shen S, Goodwin P, Hsiang T. 2001. Infection of *Nicotiana* species by the anthracnose fungus, *Colletotrichum orbiculare*. *European Journal of Plant Pathology* 107: 767–773.
- Soanes DM, Alam I, Cornell M, Wong HM, Hedeler C, Paton NW, Rattray M, Hubbard SJ, Oliver SG, Talbot NJ. 2008. Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. *PLoS ONE* 3: e2300.
- Soanes DM, Chakrabarti A, Paszkiewicz KH, Dawe AL, Talbot NJ. 2012. Genome-wide transcriptional profiling of appressorium development by the rice blast fungus *Magnaporthe oryzae*. *PLoS Pathogens* 8: e1002514.
- Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntetically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* 24: 637.
- Stephenson SA, Hatfield J, Rusu AG, Maclean DJ, Manners JM. 2000. CgDN3: an essential pathogenicity gene of *Colletotrichum gloeosporioides* necessary to avert a hypersensitive-like response in the host *Stylosanthes guianensis*. *Molecular Plant-Microbe Interactions* 13: 929–941.
- Stergiopoulos I, de Wit PJGM. 2009. Fungal effector proteins. *Annual Review of Phytopathology* 47: 233–263.
- Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008. Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. *Genome Research* 18: 1979–1990.
- Theocharidis A, van Dongen S, Enright AJ, Freeman TC. 2009. Network visualization and analysis of gene expression data using BioLayout Express3D. *Nature Protocols* 4: 1535–1550.
- Tsavkelova E, Oeser B, Oren-Young L, Israeli M, Sasson Y, Tudzynski B, Sharon A. 2012. Identification and functional characterization of indole-3-acetamide-mediated IAA biosynthesis in plant-associated *Fusarium* species. *Fungal Genetics and Biology* 49: 48–57.
- Vogel J. 2008. Unique aspects of the grass cell wall. *Current Opinion in Plant Biology* 11: 301–307.
- Xie L, Zhang J, Wan Y, Hu D. 2010. Identification of *Colletotrichum* spp. isolated from strawberry in Zhejiang Province and Shanghai City, China. *Journal of Zhejiang University-Science B* 11: 61–70.

- Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* 40: W445–W451.
- Yoshimi A, Tsuda M, Tanaka C. 2004. Cloning and characterization of the histidine kinase gene *Dic1* from *Cochliobolus heterostrophus* that confers dicarboximide resistance and osmotic adaptation. *Molecular Genetics and Genomics* 271: 228–236.
- Yoshino K, Irieda H, Sugimoto F, Yoshioka H, Okuno T, Takano Y. 2012. Cell death of *Nicotiana benthamiana* is induced by secreted protein NIS1 of *Colletotrichum orbiculare* and is suppressed by a homologue of CgDN3. *Molecular Plant-Microbe Interactions* 25: 625–636.
- Zuccaro A, Lahrmann U, Güldener U, Langen G, Pfiffi S, Biedenkopf D, Wong P, Samans B, Grimm C, Basiewicz M *et al.* 2011. Endophytic life strategies decoded by genome and transcriptome analyses of the mutualistic root symbiont *Piriformospora indica*. *PLoS Pathogens* 7: e1002290.

## Supporting Information

Additional supporting information may be found in the online version of this article.

**Fig. S1** GO terms associated with collinear genes.

**Fig. S2** Relationship of secreted *Colletotrichum* subtilisins to plant subtilisins.

**Fig. S3** Genomic region with auxin synthesis genes.

**Fig. S4** Enriched GO terms in secretomes.

**Fig. S5** Genomic region with *Ave1*.

**Fig. S6** Quantitative PCR of selected small, secreted proteins.

**Table S1** Homologs of genes required for repeat-induced point mutation

**Table S2** Predicted genes on scaffolds with homology to the partial sequence of a *C. gloeosporioides* dispensible chromosome

**Table S3** Proteases encoded in the *C. orbiculare* and *C. gloeosporioides* genomes according to MEROPS family

**Table S4** CAZy family members associated with degradation of plant cell wall carbohydrates and chitin binding in *C. orbiculare* and *C. gloeosporioides*

**Table S5** Nonenzymatic PFAM domains present in *C. orbiculare* secreted proteins but not in nonsecreted proteins

**Table S6** Nonenzymatic PFAM domains present in *C. gloeosporioides* secreted proteins but not in nonsecreted proteins

**Table S7** Number of selected nonenzymatic PFAM domains associated with infection in other phytopathogens as well as the ratio of predicted secreted proteins predicted to be carrying such domains relative to the rest of the genome

**Table S8** List of *C. orbiculare* and *C. gloeosporioides* sequences in SSP tribes with three or more members

**Table S9** Enriched GO terms associated with sequences upregulated in conidia 10-fold or more compared to *in vitro*-grown hyphae

**Table S10** Enriched GO terms associated with sequences upregulated in 1 dpi 10-fold or more compared to *in vitro*-grown hyphae

**Table S11** Enriched GO terms associated with sequences upregulated in 3 dpi 10-fold or more compared to *in vitro*-grown hyphae

**Table S12** Enriched GO terms associated with sequences upregulated in 7 dpi 10-fold or more compared to *in vitro*-grown hyphae

**Table S13** Top 100 most highly expressed genes at 1 dpi

**Table S14** Top 100 most highly expressed genes at 3 dpi

**Table S15** Top 100 most highly expressed genes at 7 dpi

**Table S16** Top 100 most highly expressed genes in conidia

**Table S17** Most highly expressed genes in *in vitro*-grown hyphae

**Table S18** Expression of homologs of known effectors from other fungi with differential expression *in planta* compared to *in vitro*-grown hyphae

**Table S19** Expression values of major facilitator superfamily (MFS) transporters

**Table S20** Secondary metabolite clusters predicted in *C. orbiculare* by SMURF

**Table S21** Secondary metabolite clusters predicted in *C. gloeosporioides* by SMURF

**Table S22** Conservation of *C. orbiculare* secondary metabolite cluster genes in *C. gloeosporioides*, *C. higginsianum* and *C. graminicola*

**Table S23** Conservation of *C. gloeosporioides* secondary metabolite cluster genes in *C. orbiculare*, *C. higginsianum* and *C. graminicola*

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.