Identification of soluble secreted proteins from appressoria of *Colletotrichum higginsianum* by analysis of expressed sequence tags

Jochen Kleemann, Hiroyuki Takahara, Kurt Stüber and Richard O'Connell

Max-Planck-Institute for Plant Breeding Research, Department of Plant–Microbe Interactions, D-50829 Köln, Germany

Correspondence
Richard O'Connell
oconnel@mpiz-koeln.mpg.de

The hemibiotrophic ascomycete *Colletotrichum higginsianum* causes anthracnose disease on brassica crops and the model plant *Arabidopsis*. Melanized appressoria pierce the host cuticle and cell wall to form specialized biotrophic hyphae inside living epidermal cells. To identify proteins secreted by appressoria that may function as virulence effectors, a cDNA library was prepared from mature appressoria formed *in vitro*. Bidirectional sequencing of 980 clones generated 1442 high-quality expressed sequence tags (ESTs), comprising 518 unique sequences. BLASTX analysis showed that 353 (68 %) of these had significant similarity to entries in the NCBI non-redundant protein database, of which 49 were also homologous to experimentally verified fungal pathogenicity genes. ORFs were predicted *ab initio* from the unique sequences and screened for potential signal peptides using SignalP. Fifty-three unique sequences (10 %) were predicted to encode proteins entering the secretory pathway, of which 26 were likely to be soluble secreted proteins. For a selected subset of these, RT-PCR showed that seven genes that encode secreted proteins of unknown function, including two *Colletotrichum*-specific genes, are upregulated in appressoria and expressed early during plant infection, and therefore represent candidate effectors.

INTRODUCTION

The ascomycete genus *Colletotrichum* comprises one of the most economically damaging groups of plant-pathogenic fungi, causing anthracnose diseases and blights on an enormous range of dicot and monocot crop plants in temperate, tropical and subtropical regions (Bailey & Jeger, 1992). The crucifer anthracnose pathogen *Colletotrichum higginsianum* has a wide host range, attacking many cultivated forms of *Brassica* and *Raphanus* as well as wild crucifers, including the model plant *Arabidopsis thaliana* (Narusaka *et al.*, 2004; O'Connell *et al.*, 2004). *C. higginsianum* employs a two-stage, hemibiotrophic infection strategy: after melanized appressoria breach the host cuticle and cell wall, the fungus initially grows biotrophically inside living epidermal cells, before entering a destructive necrotrophic phase in which host tissues are killed and extensively macerated by cell-wall-degrading enzymes. During the biotrophic phase, *C. higginsianum* differentiates specialized intracellular primary hyphae

similar to haustoria, which expand and locally modify the host plasma membrane (Shimada *et al.*, 2006). However, in contrast to obligately biotrophic pathogens of *Arabidopsis*, *Colletotrichum* can be cultured axenically and stably transformed using a variety of methods; it is also a haploid organism with uninucleate conidia, which facilitates mutational analyses. This pathosystem therefore provides an attractive model for studying biotrophy, in which both partners in the interaction are genetically tractable.

Whole-genome sequencing projects have revealed that fungal and oomycete plant pathogens possess large repertoires of secreted proteins, which can play diverse roles in pathogenicity and interactions with host cells (Dean *et al.*, 2005; Kämper *et al.*, 2006; Tyler *et al.*, 2006). For example, during pre-penetration development, extra-cellular matrix proteins secreted by spores, germ-tubes and appressoria function in cell attachment, contact-sensing, and resistance to desiccation and toxic plant metabolites (Ahn *et al.*, 2004; Hughes *et al.*, 1999; Nicholson *et al.*, 1986). Secreted hydrolases likely contribute to penetration of the plant cuticle and cell wall, and to tissue maceration (Tyler *et al.*, 2006). Other secreted enzymes that modify pathogen cell walls, e.g. chitin deacetylase, and proteins that bind pathogen wall components, e.g. *Cladosporium fulvum* Avr4p, may be critical for evading host recognition

or lysis by host enzymes (van den Burg *et al.*, 2006; El Gueddari *et al.*, 2002). Secreted effector proteins manipulate host-cell structure and physiology to promote virulence, for example by allowing the pathogen to evade or disarm plant defence mechanisms and suppress programmed cell death. Some of these effector proteins exert their activity in the apoplast or plant–pathogen interface, where they may function in counter-defence, e.g. inhibiting plant proteases and lytic enzymes (Kamoun, 2006; Rep, 2005). A second class of secreted effectors enter into host cells, although the mechanism of translocation and their intracellular targets remain to be elucidated (Catanzariti *et al.*, 2006; Kemen *et al.*, 2005; Orbach *et al.*, 2000; Ridout *et al.*, 2006). Among anthracnose fungi, the only effector so far identified is *CgDN3* from *Colletotrichum gloeosporioides*, which encodes a small secreted protein (56 aa after signal peptide cleavage) that is essential for pathogenicity on *Stylosanthes* and which appears to suppress host-cell death during the initial biotrophic phase of infection (Stephenson *et al.*, 2000).

Several approaches have been used to characterize the secretome of fungal and oomycete plant pathogens. Classical proteomic analyses using 2D-gel electrophoresis coupled with MS can detect abundant polypeptides, e.g. in culture filtrates or fluids extracted from infected plants (Paper *et al.*, 2007; Houterman *et al.*, 2007), but protein identification from peptide mass fingerprints or sequence tags is challenging in organisms for which genome sequence information is unavailable. An immunochemical approach, based on the production of mAbs to fungal structures formed *in vitro* or *in planta*, identified some developmentally regulated secreted proteins in species of *Colletotrichum*, *Erysiphe* and *Phytophthora* (Green *et al.*, 1995). However, for the genome-scale analysis of pathogen secretomes, two approaches are currently available. The first exploits biocomputational algorithms to screen for canonical N-terminal signal peptides in proteins predicted from whole-genome sequences or ESTs, and has been used to identify secreted proteins from the oomycete *Phytophthora infestans* (Tian *et al.*, 2004; Torto *et al.*, 2003), the flax rust fungus *Melampsora lini* (Catanzariti *et al.*, 2006) and the rice blast fungus *Magnaporthe grisea* (Dean *et al.*, 2005). In contrast to this *in silico* approach, the 'signal sequence trap' is a yeast-based genetic screen for cDNAs that contain functional secretion signals, based on the complementation of a reporter gene lacking a signal peptide (Klein *et al.*, 1996). Signal sequence trapping has been applied successfully to a wide range of eukaryotes, including two plant pathogens, *Phytophthora sojae* (Lee *et al.*, 2006) and *Uromyces fabae* (Link & Voegele, 2008).

In addition to their critical role in host penetration, we speculate that *Colletotrichum* appressoria, and the penetration pegs that emerge from them, secrete soluble effector proteins that permit the fungus to overcome host defence responses and reprogram host cells for biotrophy. As a first step towards the discovery of secreted effectors in *C. higginsianum*, we have generated expressed sequence tags

(ESTs) from a stage-specific cDNA library prepared from mature appressoria grown *in vitro*. Analysis of these sequences with a range of biocomputational prediction tools resulted in the identification of 12 soluble secreted proteins with either homology to functionally uncharacterized fungal proteins or no homology to any known sequences. Genes encoding seven of these were expressed during plant infection and upregulated in appressoria formed *in vitro* and therefore represent candidate effectors for future functional analysis. In addition, our study is, to the best of our knowledge, the first EST-based analysis of the appressorial transcriptome for any *Colletotrichum* species, and has revealed a number of potential pathogenicity factors for future functional analysis.

METHODS

Fungal material. Conidia of *C. higginsianum* isolate IMI 349063A were harvested from 8- to 12-day-old cultures (O'Connell *et al.*, 2004) and washed twice by centrifugation (1000 g, 5 min) and resuspension in sterile distilled water. To produce a monolayer of appressoria, 45 ml of conidial suspension (2×10^6 spores ml^{-1}) was placed into a 15 cm diameter polystyrene Petri dish, and after allowing the spores to settle and attach to the polystyrene for 40 min, a disc of sterilized nylon mesh (50 μm pore-size) was applied to the liquid surface. The liquid was then decanted, leaving the nylon mesh on the base of the dish to provide a continuous thin film of water by capillary action. The dishes were incubated in a humid box at 25 °C for 18–22 h. All manipulations were performed aseptically. The developmental stage and average number of appressoria per dish were determined by differential-interference-contrast light microscopy. To assess the presence of penetration pores, appressoria were disrupted by scraping (see below), and cell wall fragments remaining attached to the polystyrene were air-dried, sputter-coated with platinum and examined at 10 kV using a Zeiss SUPRA 40VP field-emission scanning electron microscope.

Plant material. Plants of the *A. thaliana* glabrous mutant Col-gl1-1 were grown in a controlled environment (10 h light period, 180 μE m^{-2} s^{-1}, 22 °C, 65 % humidity). Fully expanded rosette leaves were excised from 5-week-old plants and the abaxial surface was inoculated with approximately 100 μl conidial suspension (5×10^6 conidia ml^{-1}). Inoculated leaves were incubated in humid boxes in the dark at 25 °C.

RNA isolation and cDNA library construction. Sterile distilled water (20 ml) was added to each dish to gently separate the nylon mesh from the cells without damaging them. After quickly decanting the water and removing residual liquid by vigorous shaking, 25 ml TRIzol reagent (Invitrogen) was immediately added. The appressoria were disrupted by scraping them off the surface into the TRIzol using a plastic cell scraper. A fresh aliquot of TRIzol reagent was used for every five dishes. Total RNA isolation was performed according to the manufacturer's instructions, including procedures to eliminate polysaccharide contamination. The total RNA from 60 dishes was pooled and purified with the RNeasy Plant Mini kit (Qiagen) before one round of enrichment of polyadenylated (polyA) RNA using the Oligotex mRNA Mini kit (Qiagen). A random-primed, bidirectional cDNA library was constructed from 1 μg polyA RNA using the SuperScript Choice System for cDNA Synthesis (Invitrogen). The resulting cDNA library was ligated into the pGEM-T Easy vector (Promega) using *Eco*RI restriction sites and transformed into *Escherichia coli* (ElectroMAX DH10B, Invitrogen). The cDNA inserts

were sequenced from the 5′ and 3′ termini using T7 and SP6 universal primers, respectively, on Applied Biosystems 3730XL sequencers (ADIS Laboratory, Max Planck Institute for Plant Breeding Research, Köln).

Sequence analysis and bioinformatics. After excluding low-quality DNA sequences with a phred score <20 (Ewing *et al.*, 1998), vector and adaptor sequences were trimmed using the Crossmatch program (http://www.phrap.org/) and the EST sequences were assembled into contigs using the CAP3 program (Huang & Madan, 1999). The resulting set of unique sequences was queried against the NCBI non-redundant protein database and the Pathogen–Host Interactions database (PHI-base; Baldwin *et al.*, 2006) using BLASTX, and against the Consortium for the Functional Genomics of Microbial Eukaryotes (COGEME) EST database (http://cogeme.ex.ac.uk/) using TBLASTX (Altschul *et al.*, 1990).

An in-house PERL script was developed to predict coding regions in all six reading frames *ab initio*: any sequence with a stop codon preceded by in-frame ATG codons was translated into protein sequence if it encoded at least 60 aa. When no stop codon was present, translation extended to the end of the unique sequence. All in-frame ATGs were considered as potential translation initiation sites to take account of possible upstream ATG codons. The amino acid translations were then scanned for potential signal peptides using both the hidden Markov model (HMM) and neural network (NN) modules of SignalP 3.0 with default settings (Emanuelsson *et al.*, 2007). As a low-stringency initial screen, protein sequences were considered to contain a signal peptide if the HMM prediction was 'signal peptide' and at least three out of five NN scores were positive. Proteins with predicted signal peptides were queried against the NCBI non-redundant database using BLASTP without the low-complexity filter. Amino acid translations with no significant BLASTP hits ($E > 1e-5$) were considered artefacts when the unique sequence from which they were derived had conflicting BLASTX matches ($E < 1e-5$) in a different reading frame. BESTORF (Softberry) was used to confirm the ORF of selected unique sequences. Proteins were considered to be soluble secreted proteins if TMHMM and Phobius did not identify transmembrane helices, and SignalP-NN and SignalP-HMM predictions were both positive (Emanuelsson *et al.*, 2007). Functional motifs were sought by querying the predicted protein sequences against the NCBI conserved domain database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and InterProScan (http://www.ebi.ac.uk/InterProScan/). Potential glycosyl-phosphatidy-linositol (GPI)-anchored proteins were identified using the Fungal Big-PI Predictor (Eisenhaber *et al.*, 2004).

Rapid amplification of cDNA 3′ ends (3′-RACE). A 1 μg sample of appressorial polyA RNA was used for reverse transcription with 200 U SuperScript reverse transcriptase (Invitrogen). The reaction mix contained 1× first strand buffer, 0.5 mM dNTP, 10 mM DTT, 0.2 μM OligodT-adaptor primer and 10 U RNase Inhibitor (Roche). Reverse transcription was for 5 min at 23 °C, 1 h at 42 °C and 10 min at 50 °C. After heat inactivation at 80 °C for 3 min, 1 μl of a 1 : 50 dilution was used in a first PCR using gene-specific RACE primer and adaptor primer (0.5 μM each) (Table 1). Thermal cycling included an initial cycle of 98 °C for 5 min, 61 °C for 2 min and 72 °C for 10 min, followed by 30 cycles of 94 °C for 20 s, 61 °C for 20 s and 72 °C for 3 min, and final extension at 72 °C for 10 min. After 1 : 20 dilution, 1 μl was used as template in a second nested PCR using the adaptor primer and an internal gene-specific primer (0.5 μM each). The second amplification was carried out as follows: initial denaturation at 98 °C for 5 min, followed by 30 cycles of 94 °C for 20 s, 61 °C for 20 s and 72 °C for 3 min, and final extension at 72 °C for 10 min.

Semiquantitative RT-PCR. First-strand cDNA was synthesized from 1 μg total RNA after DNase treatment as described for 3′-RACE.

cDNAs prepared from infected *A. thaliana* leaves at 24 h after inoculation, mock-inoculated leaves at the same time point, ungerminated spores, spores germinated for 5 h and for 18 h *in vitro* as above and mycelium were used as templates for RT-PCR. Mycelium grown on Mathur's agar medium was used for isolation of genomic DNA (O'Connell *et al.*, 2004). Ampliqon Taq DNA Polymerase was used in 20 μl reactions according to the manufacturer's instructions. *C. higginsianum* alpha tubulin was used as the control. The thermal cycling conditions were 95 °C for 5 min, and cycles of 94 °C for 30 s, 54.5–63 °C for 30s, 72 °C for 30s and final extension at 72 °C for 10 min. Primer sequences, annealing temperatures and numbers of PCR cycles are listed in Table 1.

RESULTS

Production of mature appressoria *in vitro* and RNA extraction

The timing and morphology of appressorial development by *C. higginsianum* on polystyrene were indistinguishable from those previously observed on *A. thaliana* leaf tissue (Narusaka *et al.*, 2004; O'Connell *et al.*, 2004). The generation of sufficient numbers of appressoria on polystyrene required the use of a nylon mesh to maintain a uniformly thin film of water over the highly hydrophobic plastic substratum. Under these conditions, fungal development was well synchronized, so that after 18 h approximately 97 % of conidia had germinated to form appressoria. These were considered fully mature because they had darkly melanized cell walls containing a basal penetration pore (Fig. 1a, b), the developmental stage that immediately precedes host penetration *in planta*. Inspection of mechanically disrupted appressoria at this time point by scanning electron microscopy (SEM) confirmed the presence of penetration pores, but the surface of the polystyrene within the pores was not indented, suggesting that fungal penetration pegs had not yet developed or were unable to penetrate this plastic substratum (Fig. 1c). Importantly for the construction of an appressorium-specific cDNA library, both germinated spores and germ tubes appeared to be empty of cytoplasm at 18 h (Fig. 1a), so the contribution of transcripts from earlier developmental stages to the total mRNA pool was likely to be small. The appressoria developed as a dense, uniform monolayer of cells, allowing approximately 8×10^7 mature appressoria to be harvested from each Petri dish.

cDNA library and analysis of the EST collection

Random primers were used for first-strand synthesis in order to increase the probability of obtaining complete 5′ ends, and thus detecting N-terminal signal peptides. The bidirectional cDNA library contained 3×10^4 independent recombinant clones with an average insert size of 1.2 kb. Sequencing the inserts from 980 randomly selected cDNA clones from both ends generated a total of 1442 high-quality ESTs which could be assembled into 276 contigs (including matching paired reads from one cDNA clone)

Table 1. Primers used in this study

Primer	5′→3′ sequence	Expected product size (bp)	Annealing temperature (°C)	Cycle number for RT-PCR*
Random primer	GGGGCTCGAGNNNNNN			
Oligo-dT adaptor	CGGCCGCGAATTCACTAGTG(T)$_{18}$			
3′-RACE adaptor	CGGCCGCGAATTCACTAGTG		61	
271 3′ RACE	CTACAACGACGGCGCCAATA		61	
	CCGTAGCACTCGCAACGTCT			
1-O01 3′ RACE	ATCTTACCCGCCATCGCTCT		61	
	TGTTCCTTTGATGCCATGTTTG			
A3 3′ RACE/RT-PCR	CCACCCGCAACAGCAATAAT	184	58	30
	CCTCCTGTTGCTGGATCTGG			
271 RT-PCR	ACTGTGGACGCGGGGTAAATGAG	285	58	30
	CCTTGCAGTTGGGGTAGTGGTTGTC			
16 RT-PCR	CGCGCTTCCGTCACTTCTGC	287	56	30
	GGTATCGCGGCCTTCATCTGGAC			
1-A05 RT-PCR	GCACCGTCGCTCCCATTGTCG	354	56	30
	TCGGGGTAGAAGAGGGAGCAGAAG			
1-O01 RT-PCR	CAAACACAATCGCCAAAAATGAAGTCC	323	58	30
	CTCGCGCCCTGCAACAATACCTG			
110 RT-PCR	GAGCACAGCCATCAAGTCTTCAGTTTTC	293	63	30
	GTCATCCTTCTTGTCACCCTTCTTATCACC			
75 RT-PCR	CGGCTTCGGAGGATTCGGATACG	431	60.5	30
	GGTGCGGGTGACGGTGACAGTG			
231 RT-PCR	ACGCGACCTCTCATTCCTT	231	58	30
	CAGTACAGGCATGGCAAGTG			
64 RT-PCR	AGCCGTTCCCCACCCAGCAGATAC	390	60.5	30
	TAACGCCGGGGATGAAGCCAGAG			
3-i20 RT-PCR	TCGAGGATCACCCTCACAAA	233	57	34
	CGGTCGCAGAATTCGACATA			
α-Tubulin RT-PCR	GCCCTATTCTCGCTCGTCTTCC	248	54.5	30
	GGGCTCCAAATCGCAGTAAATG			

*For fungal cell types differentiated *in vitro*.

and 242 singletons, giving a total of 518 unique sequences. The unique sequences had a mean length of 695 bp.

To assign putative functions to the *C. higginsianum* unique sequences, they were queried against the NCBI non-redundant protein database using BLASTX. Out of the 518 unique sequences, 353 (68 %) showed significant (E <1e-5) similarity to known sequences. For 87 % of these, the best BLASTX matches were to fungal proteins annotated as 'hypothetical', 'predicted' or 'putative', i.e. proteins of unknown function that were predicted from automated whole-genome sequencing and annotation projects. To identify *C. higginsianum*-specific ESTs, the set of unique sequences without significant BLASTX matches was queried with TBLASTX against over 60 000 ESTs derived from 15 species of phytopathogenic fungi and oomycetes and three saprophytic fungi in the COGEME database. This revealed that 138 unique sequences (27 %) had no matches at the stringent cut-off of E <1e-4 and are therefore likely to be *C. higginsianum* orphan sequences. Three of these (contigs 271, 84 and 26) were also moderately to highly represented in the cDNA library (see below). This proportion of

orphan sequences is similar to that found in analyses of *Botrytis* (26 %) and *Mycosphaerella* (33 %) ESTs (Keon *et al.*, 2005; Viaud *et al.*, 2005).

To identify putative pathogenicity factors, the set of unique sequences was queried against the PHI-base database of experimentally verified pathogenicity, virulence and effector genes from fungal and oomycete pathogens of plants, animals and fungi (Baldwin *et al.*, 2006). Forty-nine (9 %) *C. higginsianum* unique sequences showed significant (E <1e-5) similarity to PHI-base entries implicated in diverse biological processes (Table 2). As expected, many of these homologues are involved in appressorium formation and function, including transcription factors, signalling pathway components, enzymes required for melanin biosynthesis and genes of uncharacterized function (*GAS1, GAS2* and *ORP1* from *Magnaporthe grisea* and *CAP20* from *C. gloeosporioides*).

To estimate the relative abundance of cDNAs in the library, we calculated the number of ESTs per contig. Most contigs (83 %) contained only two ESTs, usually the overlapping forward and reverse sequences from a single cDNA. Given
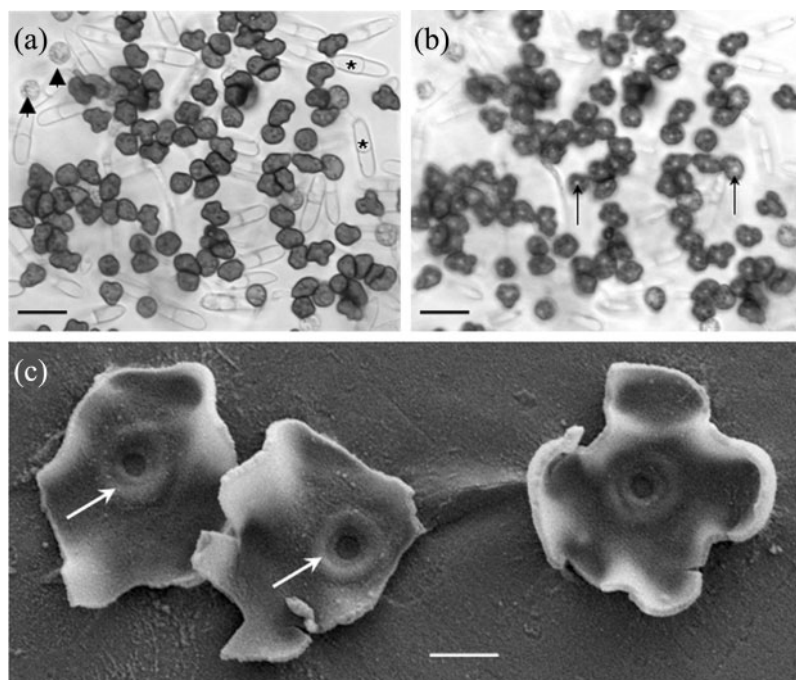
Fig. 1. Mature appressoria of *C. higginsianum* formed *in vitro*. (a, b) Light micrographs showing the monolayer of darkly melanized appressoria formed on a polystyrene Petri dish after incubation for 18 h. Bars, 10 μm. (a) Focal plane near the middle of the appressoria. Germinated conidia appear empty and are clearly septate (asterisks). Immature (non-melanized) appressoria are very infrequent (arrowheads). (b) Focal plane at the base of the appressoria. Note the basal penetration pores (arrows). (c) Scanning electron micrograph showing fragments of mature *C. higginsianum* appressoria remaining attached to the surface of a polystyrene Petri dish after mechanical disruption by scraping. The upper, domed part of the cell has been removed, revealing the presence of a penetration pore in the basal cell wall attached to the substratum. The pore is surrounded by an annular wall thickening (arrows). Within the pore, the surface of the polystyrene has not been indented by the fungal penetration peg. Bar, 2 μm.

the small sample size, we considered any sequence that was represented by three or more independent clones to be moderately to highly expressed (Table 3). The two largest contigs were also the most highly represented cDNAs (31 % of the EST collection) and encode rRNAs. This level of rRNA contamination is to be expected because only one round of polyA RNA enrichment was performed prior to library construction by random priming. The most abundant cDNA with an ORF (contig 259) showed high homology to a putative calpain protease of the rice blast fungus *M. grisea* (hypothetical protein XP_362911), which was the only homologue of fungal origin found in the NCBI non-redundant protein database. This contig was represented by 52 ESTs originating from 31 cDNA clones (3.2 %), suggesting that the gene is highly expressed in mature appressoria. Three of the most abundant cDNAs were *C. higginsianum* orphan sequences (see above). Contigs 278 and 96 had top BLASTX matches to housekeeping proteins, namely translation elongation factor 1α and histone H4, respectively. Contigs 246, 225, 283 and 198 all gave top BLASTX matches to fungal hypothetical proteins, but functionally annotated homologues of these encode glutathione *S*-transferase, *S*-formylglutathione hydrolase, mitochondrial 2-oxoglutarate/malate translocator and glucose-repressible gene (GRG1), respectively.

Mining the EST collection for putative soluble secreted proteins

After predicting ORFs *ab initio* from six-frame translations of each unique sequence, we obtained more than 2200 potential ORFs, most of which must be artefactual. However, after analysis of these with SignalP, only 167 ORFs were predicted to contain a signal peptide. These were then cross-referenced with BLASTX analyses of the respective unique sequences to eliminate artefactual amino acid translations: to merit further analysis, candidate translations had to be either orphan sequences (see above) or have consistent BLASTX matches, i.e. a significant ($E < 1e-5$) alignment to a homologous sequence using the same reading frame. A total of 72 potential ORFs with predicted signal peptides met these criteria. Ten of these ORFs were derived from unique sequences without any significant database match. To assign directionality to these unique sequences and to obtain independent support for the predicted ORFs, we used 3′-RACE to amplify further sequence. We were able to validate three out of the 10 no-homology ORFs (from 1-O01-T7, A3-T7 and contig 271) because their 3′-RACE products overlapped significantly with their respective unique sequences and the identified 3′ polyA tail correlated with the predicted ORF (data not shown). Moreover, these three ORFs could be independently identified using the BESTORF prediction program. In the case of contig 271 and 1-O01-T7, the 3′-RACE products did not provide additional sequence information apart from the polyA stretch, indicating that these unique sequences are C-terminally complete. However, the 3′-RACE product from A3-T7 revealed an additional 520 bp of sequence, and the deduced protein sequence was in-frame with the initially identified signal peptide. The extended sequence now displayed homology to a hypothetical protein from *M. grisea* (XP_364918, $E = 2e-17$), which also has a predicted signal peptide.

Table 2. *C. higginsianum* unique sequences homologous or similar to verified pathogenicity and virulence genes

Unique sequence ID	Function/gene name	Organism	E value
Contig 35	Polyketide synthase *PKS1*	*Colletotrichum lagenarium*	1e-144
Contig 132	Phosphoenolpyruvate carboxykinase *PCK1*	*Cryptococcus neoformans*	1e-137
Contig 140	Copper-transporting ATPase *CLAP1*	*Colletotrichum lindemuthianum*	1e-132
Contig 37	Polyketide synthase *PKS1*	*Colletotrichum lagenarium*	1e-126
1-M17-T7	Polyketide synthase *PKS1*	*Colletotrichum lagenarium*	1e-116
2-B11-SP6	Ste12-like transcription factor *CST1*	*Colletotrichum lagenarium*	1e-112
Contig 127	Topoisomerase *TOP1*	*Candida albicans*	1e-104
Contig 72	Class I chitin synthase *WdCHS2*	*Wangiella dermatitidis*	1e-104
Contig 247	PAK family kinase *CHM1*	*Magnaporthe grisea*	1e-103
3-K21-SP6*	Methionine synthase *MSY1*	*Fusarium graminearum*	4e-84
Contig 100	G protein α subunit *MAGC*	*Magnaporthe grisea*	2e-81
Contig 220	Virulence factor *GAS1*	*Magnaporthe grisea*	1e-78
2-K02-T7	Methylcitrate synthase *MCSA*	*Aspergillus fumigatus*	2e-77
2-H18-SP6	Superoxide dismutase *SOD1*	*Botrytis cinerea*	9e-72
2-H08-T7	CREB-like transcription factor *CPTF1*	*Claviceps purpurea*	3e-67
2-C10-T7	Virulence factor *GAS1*	*Magnaporthe grisea*	1e-65
2-O08-T7	Transcription factor *ZIF1*	*Fusarium graminearum*	3e-62
Contig 128	Cyclophilin A *BCP1*	*Botrytis cinerea*	4e-62
Contig 80	Virulence factor *GAS2*	*Magnaporthe grisea*	2e-48
1-D04-SP6	Laccase *BcLCC2*	*Botrytis cinerea*	2e-41
Contig 134	Cell surface hydrophobicity protein *CSH1*	*Candida albicans*	8e-39
Contig 31	Virulence factor *CAP20*	*Colletotrichum gloeosporioides*	2e-34
Contig 15	GTP-binding protein *RAS2*	*Ustilago maydis*	4e-34
Contig 233	Laccase *BcLCC2*	*Botrytis cinerea*	6e-30
2-B02-SP6	Kinesin *KIN2*	*Ustilago maydis*	6e-28
3-C09-T7	ATPase *ssaN*	*Salmonella enterica*	4e-27
1-P06-SP6	Major facilitator superfamily transporter *BcMFS1*	*Botrytis cinerea*	3e-24
2-E21-SP6	Cell surface hydrophobicity protein *CSH1*	*Candida albicans*	9e-23
3-K21-T7*	Methionine synthase *MSY1*	*Fusarium graminearum*	1e-22
Contig 221	G protein β subunit *CGB1*	*Cochliobolus heterostrophus*	2e-20
1-A08-T7	pH response regulator *RIM8* (*PRR1*, *palF*)	*Candida albicans*	3e-20
Contig 190	Secreted aspartyl protease *SAP3*	*Candida albicans*	9e-12
1-L01-T7	Calcineurin binding protein *CBP1*	*Cryptococcus neoformans*	1e-13
1-L21-SP6	MAPKKK *STE11a*	*Cryptococcus neoformans*	1e-13
3-G12-SP6	Transcription factor *NRG1*	*Cryptococcus neoformans*	9e-13
2-A10-SP6	pH response regulator *RIM8* (*PRR1*, *palF*)	*Candida albicans*	1e-12
1-B24-T7	Fructose transporter *FRT1*	*Botrytis cinerea*	2e-12
Contig 168	P-type ATPase *PMR1*	*Candida albicans*	7e-12
Contig 113	Secreted aspartyl protease *SAP3*	*Candida albicans*	9e-12
Contig 70	Peroxin *PEX6*	*Magnaporthe grisea*	2e-10
Contig 68	Plasma membrane protein *PTH11*	*Magnaporthe grisea*	6e-10
Contig 34	Virulence factor *ORP1*	*Magnaporthe grisea*	9e-10
Contig 193	Cytochrome P450 monooxygenase *BcBOT1*	*Botrytis cinerea*	3e-09
1-I09-T7	Methionine synthase *MSY1*	*Fusarium graminearum*	2e-08
Contig 171	Oxidoreductase *THR1*	*Colletotrichum lagenarium*	2e-06
Contig 241	PAK family kinase *CLA4*	*Ustilago maydis*	5e-06
Contig 199	*SCH9* protein kinase homologue	*Cryptococcus neoformans*	8e-06
3-M06-T7	Pectin methyl esterase *BcPME1*	*Botrytis cinerea*	1e-05
1-L06-SP6	Lysine/glutamate-rich plasma membrane protein *KER1*	*Candida albicans*	1e-05

*Two unique sequences from non-matching paired reads of one cDNA clone.

Fig. 2 summarizes the biocomputational strategy used to identify and differentiate soluble secreted proteins from integral membrane proteins, based on recently published guidelines (Emanuelsson *et al.*, 2007). Thus, ORFs containing predicted signal peptides were checked for the presence of downstream transmembrane domains (type I membrane

Table 3. Redundancy of the cDNA library; contigs represented by three (0.3 %) or more independent cDNA clones

Contig ID	EST count*	Best BLAST match	Organism	Accession number	E value
152	412	rRNA	*Hypocrea jecorina*	AF510497	0.0
203	87	rRNA	*Colletotrichum dematium*	AJ301954	0.0
259	52	Hypothetical protein	*Magnaporthe grisea*	XP_362911.1	0.0
271	11	No match			
84	10	No match			
198	8	Hypothetical protein	*Aspergillus terreus*	XP_001216808	2e-18
278	7	Translation elongation factor 1α	*Cordyceps bassiana*	AAU95496.1	0.0
246	6	Hypothetical protein	*Aspergillus terreus*	XP_001213924	3e-36
225	6	Hypothetical protein	*Magnaporthe grisea*	XP_359460	1e-117
26	6	No match			
220	5	CAS1	*Glomerella cingulata*	ABK41436.1	1e-100
96	5	Histone H4	*Magnaporthe grisea*	XP_368084	2e-38
283	4	Hypothetical protein	*Chaetomium globosum*	XP_001226880.1	1e-78

*Including forward and reverse sequences, giving paired reads.

proteins) and inverted signal anchors (type III membrane proteins). We also used BLASTX alignments and ORF prediction with the BESTORF program to investigate whether the first methionine within the amino acid translation represented the true N-terminal methionine. A striking finding was that in 15 cases where the ORF used for signal peptide prediction appeared to be N-terminally incomplete, the first methionine was invariably located within a predicted transmembrane domain, which had been spuriously recognized as a signal peptide by SignalP (Table 4). One further false positive, showing homology to a cytosolic or mitochondrial fumarate reductase of *Saccharomyces cerevisiae* (NP_010867), probably arose from the inability of SignalP to differentiate signal peptides from mitochondrial targeting peptides in some cases (Torto *et al.*, 2003).

Fig. 2. Strategy used to identify *C. higginsianum* soluble secreted proteins from ESTs.

Based on these analyses, a total of 28 unique sequences were predicted to encode integral membrane proteins, while 26 were predicted to encode soluble secreted proteins. Significant BLASTP matches could be used to assign probable subcellular localizations to 11 of the soluble secreted proteins (Table 4). Two of these have homologues that are secreted to the lumena of endomembrane compartments, namely a vacuolar carboxypeptidase from *Candida albicans* and a putative α-1,2-mannosidase from *Neosartorya fischeri*, involved in proteolysis and post-translational modifications, respectively. A further nine secreted proteins showed homology to proteins that are most likely extracellular in location. These included a putative GPI-anchored cell wall mannoprotein and a protein containing a fasciclin domain. Hydrolytic enzymes included aspartyl protease, β-1,3-glucanase, acid phosphatase and pectin methyl esterase. Oxidoreductases were represented by two different laccases and a manganese/lignin peroxidase. Based on the available ESTs, three *C. higginsianum* homologues of the *M. grisea* virulence factors GAS1 and GAS2 were also identified as soluble secreted proteins, although in *M. grisea* GFP fusion proteins have been found to be localized in the cytoplasm (Xue *et al.*, 2002).

It was impossible to extrapolate a subcellular localization for two soluble secreted proteins (contigs 271 and 1-O01-T7) which appear to be *C. higginsianum*-specific, and 10 others that resemble fungal hypothetical proteins predicted from automated whole-genome sequencing and annotation projects. After signal peptide cleavage, the two novel proteins encoded by contigs 271 and 1-O01-T7 were predicted to form mature proteins of 185 and 44 amino acids containing 12 and six cysteines, respectively. These 12 putative secreted proteins did not contain any known functional domains apart from the protein encoded by contig 194. This displayed an RPS-BLAST match below the default cut-off of E=1e-2, and was predicted to contain a
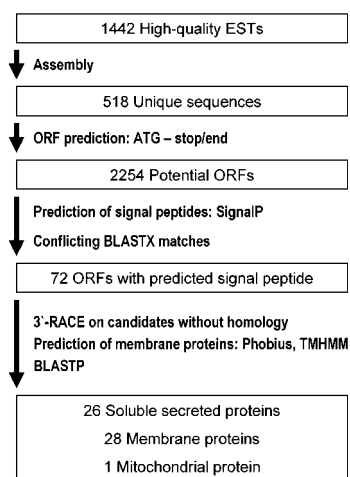
Table 4. *C. higginsianum* unique sequences providing protein translations with predicted signal peptides

Unique sequence ID	Best BLASTP match with assigned function or top match			
	Annotation	Organism	Accession number	*E* value
Secreted proteins				
Intracellular*				
Contig 239	Carboxypeptidase Y precursor	*Candida albicans*	P30574	1e-98
2-n10-T7	Putative α-1,2-mannosidase	*Neosartorya fischeri*	XP_001264690	5e-31
Contig 220	Virulence factor GAS1	*Magnaporthe grisea*	AAL28112	1e-74
Contig 80	Virulence factor GAS2	*Magnaporthe grisea*	AAF74764	2e-38
2-c10-T7	Virulence factor GAS1	*Magnaporthe grisea*	AAL28112	9e-60
Extracellular*				
Contig 162	Fasciclin I family protein	*Aspergillus fumigatus*	CAF32145	6e-19
Contig 113	Aspartyl protease	*Sclerotinia sclerotiorum*	AAF76202	1e-60
2-k11-SP6/T7†	1,3-β-Glucanase Eng1p/Dse4p	*Saccharomyces cerevisiae*	NP_014465	1e-09
1-d04-SP6	Laccase	*Melanocarpus albomyces*	CAE00180	5e-55
Contig 233	Laccase Lcc5	*Fusarium oxysporum*	ABS19941	6e-103
Contig 275	Lignin peroxidase	*Ganoderma applanatum*	BAA88392	1e-06
Contig 137	Purple acid phosphatase PAP16	*Arabidopsis thaliana*	NP_187626	1e-22
2-k14-T7	Cell wall mannoprotein cwpA	*Aspergillus niger*	AAT09020	8e-07
3-m06-T7	Putative pectin methyl esterase	*Aspergillus fumigatus*	XP_754846	1e-20
Unknown localization				
Contig 194	Hypothetical protein	*Magnaporthe grisea*	XP_362227	2e-22
Contig 51	Hypothetical protein	*Sclerotinia sclerotiorum*	EDN98905	5e-13
Contig 64	Hypothetical protein	*Magnaporthe grisea*	XP_361472	3e-21
Contig 110	Hypothetical protein	*Ustilago maydis*	XP_756613	4e-25
Contig 16	Hypothetical protein	*Magnaporthe grisea*	XP_368999	1e-61
Contig 231	Hypothetical protein	*Magnaporthe grisea*	XP_363985	4e-32
Contig 75	Hypothetical protein	*Neurospora crassa*	XP_964889	3e-12
3-i20-SP6	Hypothetical protein	*Gibberella zeae*	XP_390345	3e-17
1-a05-T7/SP6†	Hypothetical protein	*Gibberella zeae*	XP_380470	2e-05
A3-T7	Hypothetical protein	*Magnaporthe grisea*	XP_364918	7e-17
Contig 271	No match			
1-O01-T7	No match			
Membrane proteins				
Type I membrane proteins: cleavable signal peptide and one or more downstream transmembrane domains				
Contig 280	Putative integral membrane protein Ptm1	*Aspergillus fumigatus*	XP_753413	4e-89
2-f17-T7	Putative endosomal cargo receptor Erp3	*Aspergillus fumigatus*	XP_748017	3e-62
1-j14-SP6	Cell wall integrity protein Wsc2	*Saccharomyces cerevisiae*	NP_014116	4e-11
3-k20-T7/SP6†	Calreticulin	*Penicillium chrysogenum*	AAX45072	1e-48
Contig 163	Hypothetical protein	*Chaetomium globosum*	EAQ92644	2e-30
Contig 68	PTH11-like hypothetical protein	*Magnaporthe grisea*	XP_365018	7e-14
Contig 17	Hypothetical protein	*Chaetomium globosum*	EAQ85891	3e-52
Type III membrane proteins with inverted signal anchors				
Contig 223	Hypothetical protein	*Magnaporthe grisea*	XP_364865	5e-31
Contig 25	Hypothetical protein	*Gibberella zeae*	XP_387661	2e-09
Contig 82	Hypothetical protein	*Gibberella zeae*	XP_386365	3e-42
Contig 87	Hypothetical protein	*Magnaporthe grisea*	XP_363445	1e-23
3-c03-T7	Hypothetical protein	*Magnaporthe grisea*	XP_362472	1e-22
1-b07-T7	Hypothetical protein	*Magnaporthe grisea*	XP_361092	2e-07
False positives				
SignalP inaccuracy				
Contig 112	Soluble fumarate reductase (mitochondrial)	*Saccharomyces cerevisiae*	NP_010867	4e-84
N-terminally incomplete ORFs with signal peptide-mimicking transmembrane domains‡				
Contig 166	Proteophosphoglycan Ppg1	*Leishmania major*	AAZ14282	9e-26
Contig 140	Copper transporter CLAP1	*Colletotrichum lindemuthianum*	AAN62846	6e-82
Contig 168	Plasma membrane H$^+$-ATPase	*Blumeria graminis*	AAK94188	6e-42
Contig 236	Vacuolar ATP synthase subunit	*Gibberella zeae*	XP_390178	2e-28
Contig 201	Integral membrane protein	*Aspergillus fumigatus*	XP_747263	8e-39

Table 4. cont.

Unique sequence ID	Best BLASTP match with assigned function or top match			
	Annotation	Organism	Accession number	E value
1-i04-SP6	UDP-Glc/Gal endoplasmic reticulum nuc-leotide sugar transporter	*Aspergillus fumigatus*	XP_750293	1e-29
1-b24-T7	Monosaccharide transporter	*Aspergillus niger*	AAL89826	3e-28
b2-SP6	Ceramide synthase membrane component Lag1	*Aspergillus fumigatus*	XP_750934	4e-10
1-a18-SP6	Integral membrane protein	*Aspergillus fumigatus*	XP_753437	8e-08
1-d19-T7	Prohibitin Phb2p (mitochondrial)	*Saccharomyces cerevisiae*	NP_011747	2e-30
Contig 149	Hypothetical protein	*Sclerotinia sclerotiorum*	EDO04501	4e-41
Contig 153	Hypothetical protein	*Chaetomium globosum*	EAQ88262	5e-15
Contig 66	Hypothetical protein	*Gibberella zeae*	XP_387166	3e-66
2-b07-T7	Hypothetical protein	*Neurospora crassa*	XP_959845	8e-05
3-m15-SP6	Hypothetical protein	*Magnaporthe grisea*	XP_368534	7e-17

*Subcellular localization inferred from BLAST matches.

†Non-matching paired reads derived from the same cDNA clone. The expect value shown belongs to the signal peptide-containing part of the protein.

‡SignalP analysis was performed on amino acid translations predicted *ab initio* by screening the unique sequences in all six reading frames for potential start codons. The BLASTX alignment and/or ORF prediction with BESTORF suggested that the actual ORF extends upstream of the amino acid translation predicted to contain a signal peptide. Analysis of the extended ORF with TMHMM showed the presence of an internal methionine associated with a signal peptide-mimicking transmembrane domain.

SET domain, frequently found in lysine methyl transferases. None of the 12 putative secreted proteins was predicted to contain C-terminal GPI-modification sites, which are involved in the covalent linkage of secreted proteins to the extracellular face of the plasma membrane or to β-1,6-glucans in the fungal cell wall (De Groot *et al.*, 2005). These 12 *C. higginsianum* secreted proteins of unknown function therefore appear to be truly soluble, extracellular proteins and we considered them to be candidate fungal effectors.

Expression analysis of candidate secreted effector proteins

Semiquantitative RT-PCR was used to verify the expression of candidate secreted effector proteins in different fungal cell types formed *in vitro*, namely undifferentiated saprophytic mycelium, ungerminated spores, germinated spores forming germ-tubes and immature appressoria (5 h), and germinated spores with mature appressoria (18 h). Expression was also analysed at an early stage of plant infection (24 h after inoculation of *A. thaliana* leaves), when mature melanized appressoria were present on the leaf surface and approximately 10 % of appressoria had penetrated into host epidermal cells to form intracellular primary hyphae. The use of *C. higginsianum* genomic DNA as template yielded fragments that in some cases were larger than the cDNA amplicons, indicating that the RNA extracts were free from DNA contamination. Consistent and reproducible results were obtained for 11 out of the 12 candidate genes (Fig. 3). Among these, three different expression patterns could be distinguished. Two

genes (unique sequences 3-i20 and 1-a05) were constitutively expressed in saprophytic mycelia, and in both germinated and ungerminated spores. We consider that these are unlikely to function as biotrophy-related effectors. Five genes (unique sequences 64, 110, 51, A3 and 271) were induced in mature appressoria only. A further four genes (unique sequences 231, 75, 16 and 1-O01) were induced early in appressorium differentiation (germinated spores at 5 h), but were strongly upregulated in mature appressoria (germinated spores at 18 h). Although all the genes were expressed in appressoria formed *in vitro*, expression *in planta* was undetectable for three of the candidates (unique sequences 110, 64 and 3-i20). It is likely that their expression *in planta* was below the detection limit of semiquantitative RT-PCR. Overall, the expression of seven candidate genes, including two encoding *C. higginsianum*-specific proteins (unique sequences 271 and 1-O01), was upregulated in appressoria and detectable at early stages of host infection, consistent with a possible role in fungal pathogenicity or virulence.

DISCUSSION

The main aim of this study was to identify candidate secreted effector proteins from the appressoria of *C. higginsianum* that could be involved in establishing the initial biotrophic phase of infection on *Arabidopsis*. Our strategy was to construct a stage-specific cDNA library from mature appressoria formed *in vitro* and then use biocomputational prediction tools to scan the resulting ESTs for proteins containing signal peptides. In earlier
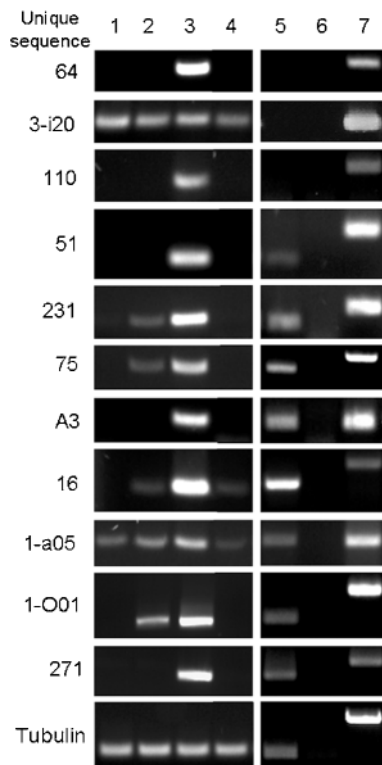
Fig. 3. Detection using semiquantitative RT-PCR of transcripts of *C. higginsianum* candidate secreted effectors during plant infection. Total RNA was extracted from fungal cell types differentiated *in vitro* (lanes 1–4), infected *A. thaliana* leaves at 24 h after inoculation (lane 5) and mock-inoculated leaves at the same time point (lane 6). *C. higginsianum* genomic DNA was included as a control template (lane 7). *C. higginsianum* α-tubulin was used as constitutively expressed control gene. Lanes: 1, ungerminated spores; 2, germinated spores (5 h) forming germ tubes with apical swellings (appressorium initials); 3, germinated spores (18 h) with mature melanised appressoria; 4, saprophytic mycelium.

studies, cDNA libraries derived from conidia undergoing appressorium differentiation have been prepared for *C. gloeosporioides*, *C. lagenarium* and *M. grisea*, and transcripts upregulated in appressoria were then enriched by differential screening or subtractive hybridization (Hwang & Kolattukudy, 1995; Inagaki *et al.*, 2000; Lu *et al.*, 2005). However, to the best of our knowledge, there has been no previous EST-based analysis of the appressorial transcriptome for any *Colletotrichum* species. Moreover, we believe this is the first attempt to systematically identify genes encoding secreted proteins in *Colletotrichum*.

Mining ESTs for extracellular secreted proteins

Most available software tools for the prediction of secreted proteins from EST sequences scan amino acid translations derived from predicted coding regions for the presence of a canonical N-terminal signal peptide (Klee & Ellis, 2005). The success of this approach depends upon the accurate identification of translation initiation sites, including the N-terminal methionine. However, many ORF-prediction algorithms either do not identify translation initiation sites with high accuracy or were not trained on fungal sequences (Nadershahi *et al.*, 2004). Other prediction tools that rely upon BLASTX-based similarity searches to identify protein coding regions, e.g. TargetIdentifier (Min *et al.*, 2005), are unlikely to uncover novel genes. We therefore designed in-house software to identify putative ORFs *ab initio* (see Methods). In contrast to the PEXfinder algorithm for mining secreted proteins from EST sequences (Torto *et al.*, 2003), we also took into account downstream ATG codons to allow for the possible presence of ATG codons within 5′ untranslated regions, which may be the case in 30 % of fungal transcripts (Mignone *et al.*, 2002).

The majority of false positives in our analysis arose from incorrect ORF prediction in unique sequences encoding membrane proteins, whereby internal transmembrane domains can be identified as signal peptides by SignalP. The non-cytoplasmic ends of transmembrane helices often resemble signal peptide cleavage sites, and SignalP was not trained to distinguish these (Emanuelsson *et al.*, 2007). The signal peptide-mimicking transmembrane domains of truncated proteins have generated false positives in computational EST analyses (Klee *et al.*, 2004) and in functional screens for secreted proteins based on the yeast signal sequence trap (Klein *et al.*, 1996). On the other hand, some false negatives may have arisen in our analysis due to the fragmentary nature of EST data, whereby a proportion of the unique sequences may be N-terminally incomplete. Although we used random primers to prime cDNA synthesis to increase the probability of cloning the complete 5′ end of transcripts, we found several examples of N-terminally incomplete unique sequences that evaded signal peptide prediction, whereas their BLASTX matches were predicted to be secreted. These included a putative glycosyl hydrolase (contig 24), serine protease (1-B12-T7), carboxypeptidase (3-i23-T7) and glutaminyl cyclase (2-K06-SP6). Finally, we cannot exclude the possibility that some fungal effector proteins lack prototypical signal peptides and would therefore evade capture by our computational approach. Although most effectors reported from fungal and oomycete plant pathogens contain N-terminal signal peptides for secretion via the endoplasmic reticulum, AVR_{a10} and AVR_{k1} from the powdery mildew fungus *Blumeria graminis* do not, and are presumably secreted through non-classical pathways (Ridout *et al.*, 2006).

Discovery of candidate secreted effectors

In spite of the above-mentioned limitations to our *in silico* approach and the relatively small number of EST sequences analysed, we were able to identify 53 *C. higginsianum* proteins that are likely to be co-translationally imported

into the ER, of which 26 were predicted to encode soluble secreted proteins without membrane-spanning domains. We assume that fungal effectors are most likely to be soluble, extracellular secreted proteins that do not become cross-linked into the fungal cell wall, and which show no homology to proteins of known function. We identified 12 secreted proteins that met these criteria, including 10 proteins homologous to functionally uncharacterized proteins from other fungi and two novel proteins without homology to any known proteins, which may be *Colletotrichum*-specific. The two novel secreted proteins encoded by contigs 271 and 1-O01-T7 were both predicted to be small polypeptides with an even number of cysteine residues, which are typical features of a subset of fungal and oomycete effectors, especially those acting in the plant apoplast (Kamoun, 2006; Rep, 2005). Moreover, although 3′-RACE suggested that the C termini of these proteins were complete, no recognizable structural motifs or protein domains were found that could provide clues to their function, which is also a common feature of fungal effectors (Rep, 2005). Such structural hallmarks were also absent from most of the secreted hypothetical proteins, but it remains to be determined whether all of these sequences are C-terminally complete. Similarly, we cannot exclude the possibility that some of these proteins contain transmembrane domains.

EST analyses of the post-penetration infection stages of rust fungi revealed a number of genes that are exclusively expressed during interaction with the host (Hahn & Mendgen, 1997), some of which may encode secreted effector proteins (Catanzariti *et al.*, 2006). Clearly, such plant-induced genes would not be captured by our analysis of *in vitro*-grown appressoria. However, we identified seven genes encoding secreted proteins that were not only specifically induced or upregulated in mature appressoria but also expressed during the early stages of plant infection. These genes may therefore encode developmentally regulated effectors, with their expression coupled to the appressorium morphogenesis programme.

Soluble secreted proteins with annotated functions

Seven of the soluble secreted proteins identified in this study showed homology to known extracellular enzymes, some of which could contribute to pathogenicity. For example, contig 113 was homologous to aspartyl proteases from *Glomerella cingulata* (anamorph=*C. gloeosporioides*) and other fungal plant and animal pathogens (Hube *et al.*, 1997; Plummer *et al.*, 2004). Secreted proteases could degrade plant cell wall proteins during penetration or interfere with host-defence-related proteins (ten Have *et al.*, 2004), although in *G. cingulata* an aspartyl protease expressed in appressoria was dispensable for pathogenicity (Plummer *et al.*, 2004). We also identified a putative pectin methyl esterase that could assist host penetration. Demethylation of highly esterified cell wall pectins may

be a prerequisite for their degradation by fungal endo-polygalacturonases (Lionetti *et al.*, 2007), and a pectin methyl esterase was required for the full virulence of *Botrytis cinerea* on *Arabidopsis* (Valette-Collet *et al.*, 2003). Furthermore, our analysis suggested that *C. higginsianum* secretes a putative lignin peroxidase. Such enzymes utilize H_2O_2 for the oxidative breakdown of lignin, melanin and other aromatic polymers (Martinez *et al.*, 2005; Woo *et al.*, 2004). We also found a putative β-1,3-glucanase with homology to *Saccharomyces cerevisiae* Eng1p/Dse4p, which functions in cell separation and localizes asymmetrically to the daughter cell side of the septum (Baladron *et al.*, 2002). In *C. higginsianum* appressoria, such an enzyme could function to remodel the fungal cell wall during morphogenesis.

Two of the predicted soluble secreted proteins may become incorporated into the fungal cell wall or extracellular matrix. For example, we found a homologue of the *Aspergillus niger* mannoprotein CwpAp, a GPI-anchored structural protein that is covalently cross-linked to cell wall glucans (Damveld *et al.*, 2005). We also identified a secreted protein containing a fasciclin domain that was homologous to fasciclin-like proteins of many other fungi. In bacteria, algae, lichens, animals and higher plants, such proteins are known to promote cell-to-cell adhesion through fasciclin domain-mediated homophilic interactions, but their function in fungi has yet to be determined (Miyazaki *et al.*, 2007). Conceivably, the *C. higginsianum* fasciclin domain protein could be involved in assembly of the fungal cell wall or extracellular matrix, or in attachment to host surfaces.

Putative virulence and pathogenicity genes

An in-depth functional annotation of all the unique sequences was beyond the scope of the present study, which aimed principally to identify appressorial cDNAs encoding secreted proteins. Nonetheless, our EST analysis uncovered a remarkable number of sequences homologous to genes with experimentally verified or potential roles in fungal pathogenicity and virulence, suggesting that the appressorial transcriptome is enriched for pathogenicity-related sequences.

Genes encoding components of cell signalling pathways were particularly well represented in the cDNA library. Thus, we discovered homologues for six components of mitogen-activated protein kinase (MAPK) signalling cascades, namely G protein α- and β-subunits, Ras2 GTP-binding protein, Ste20 p21-activated kinase (PAK) family kinase, Ste11 MAPKKK, and the Ste12 transcription factor. Some of these proteins play pivotal roles in appressorium formation and invasive growth in *Magnaporthe* and *Colletotrichum* species (Caracuel-Rios & Talbot, 2007). The cDNA library also contained a homologue of the *S. cerevisiae* plasma membrane sensor Wsc2, which operates upstream of protein kinase C and the MAPK pathway, controlling cell wall integrity (Lommel *et al.*, 2004). Two components of the calcium signalling pathway were

identified, namely homologues of Ca^{2+}/calmodulin-dependent protein kinase (contig 89) and phosphoinositide-specific phospholipase C (contig 1-B17-T7/SP6). Studies on several *Colletotrichum* species indicate that Ca^{2+}/calmodulin signalling is essential for germination, appressorium formation and perception of hard-surface contact by conidia (Ahn *et al.*, 2003). In addition, we found a homologue of cyclophilin A, a regulator of calcineurin signalling that is required for penetration peg formation and turgor generation in *M. grisea* appressoria (Viaud *et al.*, 2002). Contig 68 showed similarity to the G protein-coupled receptor *PTH11*, an integral plasma membrane protein implicated in surface sensing that acts upstream of the cAMP signalling pathway and is essential for appressorium formation and virulence in *M. grisea* (DeZwaan *et al.*, 1999).

Although appressorial melanization appeared complete when mRNA was harvested, we found homologues of several genes involved in melanin biosynthesis that may contribute to pathogenicity in *C. higginsianum*. Polyketide synthase, 1,3,8-trihydroxynaphthalene reductase and the *PEX6* peroxin have all been shown to be essential for the melanization and penetration ability of *C. lagenarium* appressoria (Kimura *et al.*, 2001; Perpetua *et al.*, 1996; Takano *et al.*, 1995). The immediate precursor of melanin, 1,8-dihydroxynaphthalene (DHN), is believed to be oxidatively polymerized in the appressorial cell wall through the activity of laccases (Langfelder *et al.*, 2003). However, no DHN oxidase has yet been characterized from either *Colletotrichum* or *Magnaporthe*, and the *C. lagenarium* *LAC1* laccase is dispensable for melanin biosynthesis, presumably due to redundancy between laccase isoenzymes (Tsuji *et al.*, 2001). In the present study, we identified two putative secreted laccases that differed in amino acid sequence (data not shown), suggesting that *C. higginsianum* deploys at least two laccase isoenzymes. We also found a homologue of the copper-transporting ATPase *CLAP1*, which is required for appressorial melanization in *C. lindemuthianum* and is thought to deliver copper to secreted cuproenzymes such as laccases (Parisot *et al.*, 2002).

Two of the most abundant cDNAs in the library (contigs 246 and 225) encoded a putative glutathione *S*-transferase and *S*-formylglutathione hydrolase, both of which are involved in the metabolism of glutathione, which functions in detoxification and the regulation of intracellular redox status (Belozerskaya & Gessler, 2007). Other transcripts with potential roles in metabolism of reactive oxygen species or redox homeostasis included a putative Cu/Zn superoxide dismutase, a catalase (1-L10-SP6/T7) and a homologue of the cAMP response element (CREB-like) transcription factor *CPTF1*, which regulates catalase activity and contributes to virulence in the biotrophic plant pathogen *Claviceps purpurea* (Nathues *et al.*, 2004).

The most abundant transcript in our cDNA library was a putative calpain protease, suggesting a high expression level in mature appressoria. The identified *C. higginsianum*

sequence had no fungal homologues in public protein databases except for a functionally uncharacterized hypothetical protein from *M. grisea*, with which it shared high sequence similarity. This protein may therefore play unique roles in these two appressorium-forming fungi and is an interesting target for functional analysis.

Conclusions

In this study, we have identified a set of seven soluble secreted proteins of unknown function from *C. higginsianum* appressoria, including two novel proteins, which are expressed at early stages of host infection. We consider these to be candidate effector proteins, and work is in progress to determine their localization *in planta* and evaluate their possible roles in overcoming host defence responses. For this, we are exploiting the genetic tractability of *Colletotrichum* to perform targeted gene knockouts, overexpression of candidate proteins and heterologous expression in non-adapted species. In addition, our analysis has provided the largest inventory of appressorium-expressed genes for any *Colletotrichum* species, and revealed many candidate pathogenicity genes for functional analysis by reverse genetics. Together with additional ESTs that we are generating from other developmental stages, the expressed genes identified in this study will be a valuable resource for future genome sequence annotation and will provide a set of characterized cDNAs for construction of a *C. higginsianum* microarray.

ACKNOWLEDGEMENTS

REFERENCES

Ahn, I. P., Uhm, K. H., Kim, S. & Lee, Y. H. (2003). Signaling pathways involved in preinfection development of *Colletotrichum gloeosporioides, C. coccodes,* and *C. dematium* pathogenic on red pepper. *Physiol Mol Plant Pathol* **63**, 281–289.

Ahn, N., Kim, S., Choi, W., Im, K. H. & Lee, Y. H. (2004). Extracellular matrix protein gene, *EMP1*, is required for appressorium formation and pathogenicity of the rice blast fungus, *Magnaporthe grisea*. *Mol Cells* **17**, 166–173.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Bailey, J. A. & Jeger, M. J. (1992). *Colletotrichum: Biology, Pathology and Control*. Wallingford, UK: CAB International

Baladron, V., Ufano, S., Duenas, E., Martin-Cuadrado, A. B., del Rey, F. & de Aldana, C. R. V. (2002). Eng1p, an endo-1,3-β-glucanase localized

at the daughter side of the septum, is involved in cell separation in *Saccharomyces cerevisiae. Eukaryot Cell* **1**, 774–786.

Baldwin, T. K., Winnenburg, R., Urban, M., Rawlings, C., Koehler, J. & Hammond-Kosack, K. E. (2006). The pathogen–host interactions database (PHI-base) provides insights into generic and novel themes of pathogenicity. *Mol Plant Microbe Interact* **19**, 1451–1462.

Belozerskaya, T. A. & Gessler, N. N. (2007). Reactive oxygen species and the strategy of antioxidant defense in fungi: a review. *Appl Biochem Microbiol* **43**, 506–515.

Caracuel-Rios, Z. & Talbot, N. J. (2007). Cellular differentiation and host invasion by the rice blast fungus *Magnaporthe grisea. Curr Opin Microbiol* **10**, 339–345.

Catanzariti, A. M., Dodds, P. N., Lawrence, G. J., Ayliffe, M. A. & Ellis, J. G. (2006). Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *Plant Cell* **18**, 243–256.

Damveld, R. A., Arentshorst, M., VanKuyk, P. A., Klis, F. M., van den Hondel, C. & Ram, A. F. J. (2005). Characterisation of CwpA, a putative glycosylphosphatidylinositol-anchored cell wall mannoprotein in the filamentous fungus *Aspergillus niger. Fungal Genet Biol* **42**, 873–885.

De Groot, P. W. J., Ram, A. F. & Klis, F. M. (2005). Features and functions of covalently linked proteins in fungal cell walls. *Fungal Genet Biol* **42**, 657–675.

Dean, R. A., Talbot, N. J., Ebbole, D. J., Farman, M. L., Mitchell, T. K., Orbach, M. J., Thon, M., Kulkarni, R., Xu, J. R. & other authors (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea. Nature* **434**, 980–986.

DeZwaan, T. M., Carroll, A. M., Valent, B. & Sweigard, J. A. (1999). *Magnaporthe grisea* Pth11p is a novel plasma membrane protein that mediates appressorium differentiation in response to inductive substrate cues. *Plant Cell* **11**, 2013–2030.

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. J Mol Biol* **337**, 243–253.

El Gueddari, N. E., Rauchhaus, U., Moerschbacher, B. M. & Deising, H. B. (2002). Developmentally regulated conversion of surface-exposed chitin to chitosan in cell walls of plant pathogenic fungi. *New Phytol* **156**, 103–112.

Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2**, 953–971.

Ewing, B., Hillier, L., Wendl, M. C. & Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**, 175–185.

Green, J. R., Pain, N. A., Cannell, M. E., Jones, G. L., Leckie, C. P., McCready, S., Mendgen, K., Mitchell, A. J., Callow, J. A. & O'Connell, R. J. (1995). Analysis of differentiation and development of the specialized infection structures formed by biotrophic fungal pathogens using monoclonal antibodies. *Can J Bot* **73**, S1, 408–417.

Hahn, M. & Mendgen, K. (1997). Characterization of *in planta* induced rust genes isolated from a haustorium-specific cDNA library. *Mol Plant Microbe Interact* **10**, 427–437.

Houterman, P. M., Speijer, D., Dekker, H. L., de Koster, C. G., Cornelissen, B. J. C. & Rep, M. (2007). The mixed xylem sap proteome of *Fusarium oxysporum*-infected tomato plants. *Mol Plant Pathol* **8**, 215–221.

Huang, X. Q. & Madan, A. (1999). CAP3: a DNA sequence assembly program. *Genome Res* **9**, 868–877.

Hube, B., Sanglard, D., Odds, F. C., Hess, D., Monod, M., Schafer, W., Brown, A. J. & Gow, N. A. (1997). Disruption of each of the secreted aspartyl proteinase genes SAP1, SAP2, and SAP3 of *Candida albicans* attenuates virulence. *Infect Immun* **65**, 3529–3538.

Hughes, H. B., Carzaniga, R., Rawlings, S. L., Green, J. R. & O'Connell, R. J. (1999). Spore surface glycoproteins of *Colletotrichum lindemuthianum* are recognized by a monoclonal antibody which inhibits adhesion to polystyrene. *Microbiology* **145**, 1927–1936.

Hwang, C. S. & Kolattukudy, P. E. (1995). Isolation and characterization of genes expressed uniquely during appressorium formation by *Colletotrichum gloeosporioides* conidia induced by the host surface wax. *Mol Gen Genet* **247**, 282–294.

Inagaki, A., Takano, Y., Kubo, Y., Mise, K. & Furusawa, I. (2000). Construction of an equalized cDNA library from *Colletotrichum lagenarium* and its application to the isolation of differentially expressed genes. *Can J Microbiol* **46**, 150–158.

Kamoun, S. (2006). A catalogue of the effector secretome of plant pathogenic oomycetes. *Annu Rev Phytopathol* **44**, 41–60.

Kämper, J., Kahmann, R., Bolker, M., Ma, L. J., Brefort, T., Saville, B. J., Banuett, F., Kronstad, J. W., Gold, S. E. & other authors (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis. Nature* **444**, 97–101.

Kemen, E., Kemen, A. C., Rafiqi, M., Hempel, U., Mendgen, K., Hahn, M. & Voegele, R. T. (2005). Identification of a protein from rust fungi transferred from haustoria into infected plant cells. *Mol Plant Microbe Interact* **18**, 1130–1139.

Keon, J., Antoniw, J., Rudd, J., Skinner, W., Hargreaves, J. & Hammond-Kosack, K. (2005). Analysis of expressed sequence tags from the wheat leaf blotch pathogen *Mycosphaerella graminicola* (anamorph *Septoria tritici*). *Fungal Genet Biol* **42**, 376–389.

Kimura, A., Takano, Y., Furusawa, I. & Okuno, T. (2001). Peroxisomal metabolic function is required for appressorium-mediated plant infection by *Colletotrichum lagenarium. Plant Cell* **13**, 1945–1957.

Klee, E. W. & Ellis, L. B. (2005). Evaluating eukaryotic secreted protein prediction. *BMC Bioinformatics* **6**, 256.

Klee, E. W., Carlson, D. F., Fahrenkrug, S. C., Ekker, S. C. & Ellis, L. B. M. (2004). Identifying secretomes in people, pufferfish and pigs. *Nucleic Acids Res* **32**, 1414–1421.

Klein, R. D., Gu, Q. M., Goddard, A. & Rosenthal, A. (1996). Selection for genes encoding secreted proteins and receptors. *Proc Natl Acad Sci U S A* **93**, 7108–7113.

Langfelder, K., Streibel, M., Jahn, B., Haase, G. & Brakhage, A. A. (2003). Biosynthesis of fungal melanins and their importance for human pathogenic fungi. *Fungal Genet Biol* **38**, 143–158.

Lee, S. J., Kelley, B. S., Damasceno, C. M. B., John, B. S., Kim, B. S., Kim, B. D. & Rose, J. K. C. (2006). A functional screen to characterize the secretomes of eukaryotic pathogens and their hosts in planta. *Mol Plant Microbe Interact* **19**, 1368–1377.

Link, T. & Voegele, R. T. (2008). Secreted proteins of *Uromyces fabae*: similarities and stage specificity. *Mol Plant Pathol* **9**, 59–66.

Lionetti, V., Raiola, A., Camardella, L., Giovane, A., Obel, N., Pauly, M., Favaron, F., Cervone, F. & Bellincampi, D. (2007). Overexpression of pectin methylesterase inhibitors in *Arabidopsis* restricts fungal infection by *Botrytis cinerea. Plant Physiol* **143**, 1871–1880.

Lommel, M., Bagnat, M. & Strahl, S. (2004). Aberrant processing of the WSC family and Mid2p cell surface sensors results in cell death of *Saccharomyces cerevisiae O*-mannosylation mutants. *Mol Cell Biol* **24**, 46–57.

Lu, J. P., Liu, T. B. & Lin, F. C. (2005). Identification of mature appressorium-enriched transcripts in *Magnaporthe grisea*, the rice

blast fungus, using suppression subtractive hybridization. *FEMS Microbiol Lett* **245**, 131–137.

Martinez, A. T., Speranza, M., Ruiz-Duenas, F. J., Ferreira, P., Camarero, S., Guillen, F., Martinez, M. J., Gutierrez, A. & del Rio, J. C. (2005). Biodegradation of lignocellulosics: microbial, chemical, and enzymatic aspects of the fungal attack of lignin. *Int Microbiol* **8**, 195–204.

Mignone, F., Gissi, C., Liuni, S. & Pesole, G. (2002). Untranslated regions of mRNAs. *Genome Biol* **3**, reviews0004.

Min, X. J., Butler, G., Storms, R. & Tsang, A. (2005). TargetIdentifier: a webserver for identifying full-length cDNAs from EST sequences. *Nucleic Acids Res* **33**, W669–W672.

Miyazaki, Y., Kaneko, S., Sunagawa, M., Shishido, K., Yamazaki, T., Nakamura, M. & Babasaki, K. (2007). The fruiting-specific *Le.flp1* gene, encoding a novel fungal fasciclin-like protein, of the basidiomycetous mushroom *Lentinula edodes*. *Curr Genet* **51**, 367–375.

Nadershahi, A., Fahrenkrug, S. C. & Ellis, L. B. M. (2004). Comparison of computational methods for identifying translation initiation sites in EST data. *BMC Bioinformatics* **5**, 14.

Narusaka, Y., Narusaka, M., Park, P., Kubo, Y., Hirayama, T., Seki, M., Shiraishi, T., Ishida, J., Nakashima, M. & other authors (2004). *RCH1*, a locus in *Arabidopsis* that confers resistance to the hemibiotrophic fungal pathogen *Colletotrichum higginsianum*. *Mol Plant Microbe Interact* **17**, 749–762.

Nathues, E., Joshi, S., Tenberge, K. B., von den Driesch, M., Oeser, B., Baumer, N., Mihlan, M. & Tudzynski, P. (2004). CPTF1, a CREB-like transcription factor, is involved in the oxidative stress response in the phytopathogen *Claviceps purpurea* and modulates ROS level in its host *Secale cereale*. *Mol Plant Microbe Interact* **17**, 383–393.

Nicholson, R. L., Butler, L. G. & Asquith, T. N. (1986). Glycoproteins from *Colletotrichum graminicola* that bind phenols: implications for survival and virulence of phytopathogenic fungi. *Phytopathology* **76**, 1315–1318.

O'Connell, R., Herbert, C., Sreenivasaprasad, S., Khatib, M., Esquerre-Tugaye, M. T. & Dumas, B. (2004). A novel *Arabidopsis–Colletotrichum* pathosystem for the molecular dissection of plant–fungal interactions. *Mol Plant Microbe Interact* **17**, 272–282.

Orbach, M. J., Farrall, L., Sweigard, J. A., Chumley, F. G. & Valent, B. (2000). A telomeric avirulence gene determines efficacy for the rice blast resistance gene *Pi-ta*. *Plant Cell* **12**, 2019–2032.

Paper, J. M., Scott-Craig, J. S., Adhikari, N. D., Cuom, C. A. & Walton, J. D. (2007). Comparative proteomics of extracellular proteins *in vitro* and *in planta* from the pathogenic fungus *Fusarium graminearum*. *Proteomics* **7**, 3171–3183.

Parisot, D., Dufresne, M., Veneault, C., Lauge, R. & Langin, T. (2002). *clap1*, a gene encoding a copper-transporting ATPase involved in the process of infection by the phytopathogenic fungus *Colletotrichum lindemuthianum*. *Mol Genet Genomics* **268**, 139–151.

Perpetua, N. S., Kubo, Y., Takano, Y. & Furusawa, I. (1996). Cloning and characterization of a melanin biosynthetic THR1 reductase gene essential for appressorial penetration of *Colletotrichum lagenarium*. *Mol Plant Microbe Interact* **9**, 323–329.

Plummer, K. M., Clark, S. J., Ellis, L. M., Loganathan, A., Al-Samarrai, T. H., Rikkerink, E. H. A., Sullivan, P. A., Templeton, M. D. & Farley, P. C. (2004). Analysis of a secreted aspartic peptidase disruption mutant of *Glomerella cingulata*. *Eur J Plant Pathol* **110**, 265–274.

Rep, M. (2005). Small proteins of plant-pathogenic fungi secreted during host colonization. *FEMS Microbiol Lett* **253**, 19–27.

Ridout, C. J., Skamnioti, P., Porritt, O., Sacristan, S., Jones, J. D. G. & Brown, J. K. M. (2006). Multiple avirulence paralogues in cereal powdery mildew fungi may contribute to parasite fitness and defeat of plant resistance. *Plant Cell* **18**, 2402–2414.

Shimada, C., Lipka, V., O'Connel, R., Okuno, T., Schulze-Lefert, P. & Takano, Y. (2006). Nonhost resistance in *Arabidopsis–Colletotrichum* interactions acts at the cell periphery and requires actin filament function. *Mol Plant Microbe Interact* **19**, 270–279.

Stephenson, S. A., Hatfield, J., Rusu, A., Maclean, D. J. & Manners, J. M. (2000). *CgDN3*: an essential pathogenicity gene of *Colletotrichum gloeosporioides* necessary to avert a hypersensitive-like response in the host *Stylosanthes guianensis*. *Mol Plant Microbe Interact* **13**, 929–941.

Takano, Y., Kubo, Y., Shimizu, K., Mise, K., Okuno, T. & Furusawa, I. (1995). Structural analysis of PKS1, a polyketide synthase gene involved in melanin biosynthesis in *Colletotrichum lagenarium*. *Mol Gen Genet* **249**, 162–167.

ten Have, A., Dekkers, E., Kay, J., Phylip, L. H. & van Kan, J. A. L. (2004). An aspartic proteinase gene family in the filamentous fungus *Botrytis cinerea* contains members with novel features. *Microbiology* **150**, 2475–2489.

Tian, M. Y., Huitema, E., Da Cunha, L., Torto-Alalibo, T. & Kamoun, S. (2004). A Kazal-like extracellular serine protease inhibitor from *Phytophthora infestans* targets the tomato pathogenesis-related protease P69B. *J Biol Chem* **279**, 26370–26377.

Torto, T. A., Li, S. A., Styer, A., Huitema, E., Testa, A., Gow, N. A. R., van West, P. & Kamoun, S. (2003). EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen *Phytophthora*. *Genome Res* **13**, 1675–1685.

Tsuji, G., Fujikawa, J., Ishida, H., Horino, O. & Kubo, Y. (2001). Laccase gene *LAC1* of *Colletotrichum lagenarium* is not essential for melanin biosynthesis and pathogenicity. *J Gen Plant Pathol* **67**, 182–190.

Tyler, B. M., Tripathy, S., Zhang, X. M., Dehal, P., Jiang, R. H. Y., Aerts, A., Arredondo, F. D., Baxter, L., Bensasson, D. & other authors (2006). *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* **313**, 1261–1266.

Valette-Collet, O., Cimerman, A., Reignault, P., Levis, C. & Boccara, M. (2003). Disruption of *Botrytis cinerea* pectin methylesterase gene *Bcpme1* reduces virulence on several host plants. *Mol Plant Microbe Interact* **16**, 360–367.

van den Burg, H. A., Harrison, S. J., Joosten, M., Vervoort, J. & de Wit, P. (2006). *Cladosporium fulvum* Avr4 protects fungal cell walls against hydrolysis by plant chitinases accumulating during infection. *Mol Plant Microbe Interact* **19**, 1420–1430.

Viaud, M. C., Balhadere, P. V. & Talbot, N. J. (2002). A *Magnaporthe grisea* cyclophilin acts as a virulence determinant during plant infection. *Plant Cell* **14**, 917–930.

Viaud, M., Legeai, F., Pradier, J. M., Brygoo, Y., Bitton, F., Weissenbach, J., Brunet-Simon, A., Duclert, A., Fillinger, S. & other authors (2005). Expressed sequence tags from the phytopathogenic fungus *Botrytis cinerea*. *Eur J Plant Pathol* **111**, 139–146.

Woo, S. H., Cho, J. S., Lee, B. S. & Kim, E. K. (2004). Decolorization of melanin by lignin peroxidase from *Phanerochaete chrysosporium*. *Biotechnol Bioproc Eng* **9**, 256–260.

Xue, C. Y., Park, G., Choi, W. B., Zheng, L., Dean, R. A. & Xu, J. R. (2002). Two novel fungal virulence genes specifically expressed in appressoria of the rice blast fungus. *Plant Cell* **14**, 2107–2119.

Edited by: H. A. B. Wösten