Available online at www.sciencedirect.com
SCIENCE DIRECT®
Plant Science 167 (2004) 481–489
PLANT SCIENCE
www.elsevier.com/locate/plantsci

Comparative analysis of expressed sequence tags from *Malva pusilla*, *Sorghum bicolor*, and *Medicago truncatula* infected with *Colletotrichum* species

Paul H. Goodwin [a,*], Richard P. Oliver [b,c], Tom Hsiang [a]

[a] *Department of Environmental Biology, University of Guelph, Guelph, Ont., Canada N1G 2W1*
[b] *Australian Centre for Necrotrophic Fungal Pathogens, Division of Science and Engineering, Murdoch University, WA 6150, Australia*
[c] *Carlsberg Laboratory, Department of Physiology, Gl. Carlsbergvej 5-10, Valby, DK 2500, Denmark*

Received 4 February 2004; received in revised form 21 April 2004; accepted 21 April 2004

Available online 18 May 2004

Abstract

To assess relative gene expression, expressed sequence tag redundancy was compared between EST collections from susceptible *Malva pusilla* and *Medicago truncatula* inoculated with *Colletotrichum gloeosporioides* f. sp. *malvae* and *C. trifolii*, respectively, and resistant and susceptible *Sorghum bicolor* inoculated with *C. graminicola* (=*C. sublineolum*). EST redundancies from the fungal-inoculated *S. bicolor* and *M. truncatula* were also compared to healthy *S. bicolor* and *M. truncatula*. Several of the more redundant plant ESTs in the *C. gloeosporioides* f. sp. *malvae–M. pusilla* interaction represented genes encoding pathogenesis-related proteins, such as β-1,3-glucanase, osmotin and chitinase, but a number of other ESTs, such as those for cysteine proteinase, heat shock protein and glutathione *S*-transferase, were also relatively abundant. Differences in EST redundancy between different interactions included a greater abundance of heat shock protein ESTs in the susceptible *S. bicolor* interaction, and a greater abundance of cysteine proteinase ESTs in the resistant *S. bicolor* and susceptible *M. truncatula* interactions. Using EST redundancy to compare gene expression between different host plants interacting with *Colletotrichum* species provides a useful basis for selecting genes for further study in plant–*Colletotrichum interactions*.

Keywords: Expressed sequence tag; Hemibiotrophy; Resistance; Susceptibility

1. Introduction

Colletotrichum species are the causal agents of anthracnose, fruit rot and blight of a large number of crops [1,2]. Many cause disease through an intracellular hemibiotrophic infection strategy. The hyphae first develop biotrophically (i.e., feeding on living host cells) within the cell lumen without penetrating the cell membrane, and after several days, switch to growing necrotrophically (i.e., feeding on host cells previously killed by the pathogen) [3,4].

One example of an intracellular hemibiotrophic infection strategy is the infection of *Malva pusilla* (round-leaved mallow) by *C. gloeosporioides* f. sp. *malvae* [5]. After penetration, a large, rounded infection vesicle forms followed by the production of intracellular, large primary hyphae, which grow from cell to cell feeding biotrophically. At 72–96 h post-inoculation (HPI), *C. gloeosporioides* f. sp. *malvae* produces thin secondary hyphae from the large primary hyphae, which feed necrotrophically [5]. The infection of *Medicago sativa* and related hosts by *C. trifolii* shows the same pattern of infection with large primary and thin secondary hyphae, and *C. trifolii* is considered to be so similar to *C. gloeosporioides* f. sp. *malvae* that both these fungi have been suggested to be different forma specialis of the *C. orbiculare* aggregate species [4,6]. Infection of *Sorghum bicolor* by *C. sublineolum* also results in the growth of

Abbreviations: COG, cluster of orthologous groups; cDNA, complementary DNA; ESTs, expressed sequence tags; HPI, hours post-inoculation; GTP, guanosine triphosphate; NCBI, National Center for Biotechnology Information; mRNA, messenger RNA

* Corresponding author. Tel.: +1-519-824-4120x52754; fax: +1-519-837-0442.
E-mail addresses: pgoodwin@uoguelph.ca (P.H. Goodwin), roliver@central.murdoch.edu.au (R.P. Oliver), thsiang@uoguelph.ca (T. Hsiang).

0168-9452/$ – see front matter © 2004 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.plantsci.2004.04.014

intracellular hyphae spreading from host cell to cell with a distinct biotrophic phase that is similar to that of the *C. orbiculare* aggregate species, except that the difference between the infection vesicle and primary hyphae is less distinct with *C. sublineolum* [4,7].

One way of analyzing the types of genes being expressed and their level of expression is to conduct single-pass, partial sequencing of cDNA clones to generate expressed sequence tags (ESTs) [8]. While ESTs can be obtained from axenic cultures of fungi or from healthy plants, it is the genes expressed during a plant–pathogen interaction that are generally considered to be most relevant to understanding the molecular basis of a plant disease. These can be plant genes related to resistance and the response to infection, or pathogen genes related to virulence and the ability to grow and reproduce in the plant. Descriptions of several EST collections from fungal-infected plants have been published [9–15]. However, only 13 ESTs from a plant infected with a *Colletotrichum* species have been described thus far, and these ESTs from maize with anthracnose disease were obtained following sequential subtractive hybridization to select for up-regulated genes [16]. No characterizations of large-scale EST collections have yet been published for a plant disease caused by a *Colletotrichum* species nor have systematic cross comparisons ever been described between EST collections from different *Colletotrichum*-infected plants or any other fungal-infected plants.

An EST collection was developed from cloned cDNA of *M. pusilla* infected with *C. gloeosporioides* f. sp. *malvae* [17], and comparisons were made within and between this EST collection and three additional EST collections for two other plants species infected with *Colletotrichum* species. Two EST collections from *S. bicolor* inoculated with a sorghum isolate of *C. graminicola* (=*C. sublineolum*) and an EST collection from healthy *S. bicolor* plants were obtained from GenBank, and EST collections from *Medicago truncatula* (barrel medic) infected with *C. trifolii* and healthy *M. truncatula* were also obtained from GenBank.

The goal of this work was to gain an overview of gene expression patterns in several *Colletotrichum*-infected plants by comparing ESTs from susceptible interactions of *M. pusilla* with *C. gloeosporioides* f. sp. *malvae* and *M. truncatula* with *C. trifolii*, as well as susceptible and resistant interactions of *S. bicolor* with *C. sublineolum*. ESTs from the interactions were first classified as to whether they were most likely of fungal or plant origin by making comparisons using Standalone TBLASTX with a database of the reported complete genomes of a plant and a fungus [18]. Homologues to the ESTs from each interaction were identified by comparison to the GenBank NR protein database using the BLASTX algorithm [19], and comparisons of EST frequency within and between the interactions were done using Standalone TBLASTX with databases of the ESTs from the four different plant–*Colletotrichum* interactions. Comparisons were also made with databases of ESTs from healthy *S. bicolor* and *M. truncatula* to determine if the redundancy of the ESTs from the *Colletotrichum*-inoculated tissues differed significantly from that of healthy tissues.

2. Materials and methods

2.1. Source of fungal and plant sequences

Comparisons were made of six EST libraries and were designated Cgm-*Malva*, Cs-2wk*Sorghum*, Cs-4wk*Sorghum*, Ct-*Medicago*, healthy *Medicago* and healthy *Sorghum*. The Cgm-*Malva* EST library was based on forward and reverse sequences from 840 clones of a cDNA library prepared from *M. pusilla* leaves inoculated with *C. gloeosporioides* f. sp. *malvae* ATCC20767 at 96 HPI [17]. Among 1610 sequences, those with 10% or more Ns or shorter than 100 bp were omitted, and the sequence from each pair of forward and reverse sequences that had a higher match in a BLAST search of the GenBank NR database was retained, resulting in 813 ESTs from the Cgm-*Malva* library. Contaminating vector sequences from the ESTs were trimmed. The 813 Cgm-*Malva* EST sequences described here have been deposited at GenBank (accession numbers CN445901–CN446713).

EST collections from other *Colletotrichum*-infected plants were directly downloaded from GenBank (Table 1). The 9533 ESTs from Cs-2wk*Sorghum* and 9660 ESTs from Cs-4wk*Sorghum* are described in GenBank as originating, respectively, from cDNA of 2- and 4-week-old *S. bicolor* cv. BTX 623 plants at 48 HPI with isolate FRM42I of *C.*

Table 1
Categorization of ESTs from infected plants as having a plant and fungal origin by Standalone TBLASTX (*e*-value ≤0.1) using a database composed of the genomes of *N. crassa* for *Colletotrichum* species and either *A. thaliana* for *M. pusilla* and *M. truncatula*, or *O. sativa* for *S. bicolor*

EST library (number of sequences)	Source	Standalone TBLASTX matches, number (%)		
		Plant	Fungal	Unknown
C. gloeosporioides f. sp. *malvae–M. pusilla* (813)	Current study	499 (73.7)	161 (19.8)	153 (18.8)
Colletotrichum sublineolum–S. bicolor, 2 weeks (9533)[a]	GenBank numbers BM322178–BM331713	8518 (89.4)	268 (2.8)	747 (7.8)
C. sublineolum–S. bicolor, 4 weeks (9660)	GenBank numbers BE363849–BM318787	8826 (91.4)	250 (2.6)	584 (6.0)
Colletotrichum trifolii–M. truncatula (6003)	GenBank numbers AW127313–BG604192	5543 (92.3)	85 (1.4)	375 (6.2)

[a] The interaction related to this EST library was described as a resistant reaction by the sequence generators, while the others have been described as susceptible.

graminicola (M.-M. Cordonnier-Pratt et al., Department of Plant Biology, University of Georgia, unpublished). This isolate originated from sorghum and therefore should be considered *C. sublineolum* [20]. Young 2-week-old sorghum seedlings were described as exhibiting a juvenile resistant reaction, which is an incompatible interaction, while 4-week or older plants were described as showing a susceptible reaction. In addition, an EST collection of 10043 sequences from healthy *S. bicolor* was obtained from GenBank (sequence numbers were not fully contiguous, but most were from AW282433 to AW287752) (M.-M. Cordonnier-Pratt et al., Department of Plant Biology, University of Georgia, unpublished). This cDNA was prepared from 10- to 14-day-old light-grown seedlings of *S. bicolor*.

The Ct-*Medicago* library was based on 6003 sequences selected from the interaction of *M. truncatula* genotype A17 with *C. trifolii* (M. Fedorova et al., Department of Agronomy and Plant Genetics, University of Minnesota, unpublished). These ESTs are described in GenBank as having been obtained from cDNA from cotyledons and primary leaves of *M. truncatula* harvested 5 and 8 days after inoculation with *C. trifolii*. In addition, an EST collection of 2143 sequences from healthy *M. truncatula* was also obtained from GenBank (sequence numbers were not fully contiguous, but most were from BQ151630 to BQ149890) from the same researchers who provided the *C. trifolii*-infected *Medicago* library. This cDNA was prepared from a mixture of cotyledons from 5-day-old seedlings and leaves from 2-week-old *M. truncatula* plants.

2.2. BLAST analyses of EST sequences

In addition to standard Internet access to BLAST at the National Center for Biotechnology Information (NCBI) of the National Library of Medicine, Washington, DC, USA, analyses were also done with Standalone BLAST, where both the BLAST programs (BLASTALL, Version 2.2.3) and sequence databases were downloaded and used locally. The Standalone BLAST program for Win32 systems was downloaded from NCBI (ftp://ftp.ncbi.nih.gov/blast/executables/blastz.exe). All BLAST analyses were run with default parameters, except for expectation values (*e*-values), which were specified for each analysis.

To distinguish fungal and plant sequences, the ESTs were queried with Standalone TBLASTX against a database composed of the reported genomes of *Arabidopsis thaliana* and *Neurospora crassa* for interactions involving dicotyledonous plants, and *Oryza sativa* and *N. crassa* for interactions involving monocotyledonous plants [18]. For each query sequence, the taxonomic origin was assigned based on the top match (*e*-value $\leq 10^{-1}$).

The Cgm-*Malva* sequences were queried against the GenBank NR protein database using BLASTX to infer gene function. For each query sequence, the putative functional group was assigned based on the top match (*e*-value $\leq 10^{-1}$). These gene functions were then manually placed in the cluster of orthologous groups (COG) classification system [21].

To examine for clusters of homologous sequences within the Cgm-*Malva* ESTs, ESTs from Cgm-*Malva* were used to construct a BLAST database in a Standalone TBLASTX analysis and as query sequences compared against that database. These clusters were further combined by inferred gene functions, and those gene functions that had $\geq$0.37% redundancy (3 or more out of 813) were chosen for further study. Using Standalone TBLASTX analyses, these Cgm-*Malva* ESTs were then compared to databases composed of the other infected plant EST libraries to identify sequences in those EST libraries sharing homology with the Cgm-*Malva* ESTs that had $\geq$0.37% redundancy. To determine statistical significance, EST abundance was compared between Cgm-*Malva* and each of the other fungal-infected EST collections using Fisher's exact test as implemented in SAS Proc Freq (SAS Institute, Cary, NC) at $P = 0.05$.

The same procedure was employed to identify clusters of homologous sequences within and between the other *Colletotrichum*-infected plant libraries. Homology clusters in the other EST libraries, which also had a sequence abundance of $\geq$0.37%, were chosen for further study. A representative from each homology cluster was matched against the GenBank NR protein database using BLASTX, and inferred gene function was assigned based on best match (*e*-value $\leq 10^{-1}$). This representative sequence was also used as a query sequence against the respective healthy plant EST BLAST database created for Standalone TBLASTX analysis. In addition, the representative sequences were also used as query sequences to determine the abundance of homologues in other EST libraries by Standalone TBLASTX analysis. To determine statistical significance, EST abundance was compared between a fungal-infected EST collection and its respective healthy plant EST collection using Fisher's exact test as implemented in SAS Proc Freq (SAS Institute, Cary, NC) at $P = 0.05$.

2.3. Data extraction

Perl scripts were written to extract the following relevant data from BLAST output: query sequence name, query sequence length, top match, length of top match, *e*-value for top match, number of matches with *e*-value $\leq 10^{-1}$ or $\leq 10^{-5}$ and a list of the next five top matches. Separate scripts were written to parse output from Standalone BLAST or Internet BLAST, and they are available at http://www.uoguelph.ca/~thsiang/est/ or upon request to the last author. The Perl scripts generated a single line of output from each BLAST output file. For each data set, the single lines were combined into a single file, and the file imported into a spreadsheet program. The use of commas for delimiters as specified in the Perl scripts allowed for the spreadsheet file to be parsed within a spreadsheet program into columns of relevant data listed above.

3. Results

Separation of the ESTs from the fungal-infected plants into putative host and pathogen sequences showed that plant matches always predominated, ranging from 74 to 92% of the sequences (Table 1). There were more matches with fungal sequences in the Cgm-*Malva* interaction compared to the others, particularly the Ct-*Medicago* interaction. Although the Cs-2wk*Sorghum* ESTs were from an interaction reported to be resistant, the total number of ESTs with fungal matches was not significantly higher than that of the Cs-4wk*Sorghum* ESTs, which was a susceptible interaction.

Of the 813 sequences from the Cgm-*Malva* interaction, 79% could be manually placed into the COG functional gene classification system (Table 2), based on the function of the sequence with the best match in GenBank NR at *e*-value $\leq 10^{-1}$. However, when eukaryotic COG sequences (downloaded March 2003 from ftp.ncbi.nih.gov/pub/COG/KOG) were used as a matching database in Standalone BLASTX analysis, 75% of the Cgm-*Malva* sequences found a match at *e*-value $\leq 10^{-1}$, with 66% at *e*-value $\leq 10^{-5}$. By comparison, BLASTX analysis of these Cgm-*Malva* sequences against the GenBank NR database showed that 69% had matches at *e*-value $\leq 10^{-5}$, with 82% at *e*-value $\leq 10^{-1}$. Aside from the ESTs classified as having an unknown function, the largest group of ESTs was related to defense mechanisms (Table 2). These ESTs did not include any sequences that had been identified as having a fungal match, and the relatively high number of these ESTs indicated that there was a significant expression of putative host defense genes even in susceptible interactions. This COG class of defense mechanism genes included ESTs having GenBank NR matches with several of the well-known pathogenesis-related (PR) proteins, as well as less studied genes, such as harpin-induced protein, respiratory burst oxidase protein and pathogen-induced oxygenase. Among the ESTs related to cellular processes and signaling, the next largest group was related to signal transduction, which was mostly composed of various protein kinases. There was also a relatively large number of ESTs related to post-translational modification, protein turnover and chaperones, which were mostly different proteinase and chaperone/heat shock proteins. Among the ESTs classified in the metabolism group, many were associated with energy production and conversion and were primarily plant sequences related to photosynthesis. A number of ESTs were also related to different sugar, water and ion transporters, and these were in the COG class of secondary metabolites biosynthesis, transport and catabolism.

Table 2
Number of ESTs from *M. pusilla* infected by *C. gloeosporioides* f. sp. *malvae* manually placed into the putative protein classes of the cluster of orthologous groups (COG) classification system based on the gene function of the top match in GenBank number at *e*-value $\leq 10^{-1}$

COG classification[a]	Number
Information storage and processing	
[J] Translation, ribosomal structure and biogenesis	38
[A, K, L] RNA processing and modification, transcript, replication, recombination and repair	36
[B] Chromatin structure and dynamics	0
Cellular processes and signaling	
[D] Cell cycle control, cell division, chromosome partitioning	0
[Y] Nuclear structure	0
[V] Defense mechanisms	70
[T] Signal transduction mechanisms	49
[M] Cell wall/membrane/envelope biogenesis	9
[N] Cell motility	0
[Z] Cytoskeleton	14
[W] Extracellular structures	0
[U] Intracellular trafficking, secretion, and vesicular transport	12
[O] Post-translational modification, protein turnover, chaperones	23
Metabolism	
[C] Energy production and conversion	58
[G] Carbohydrate transport and metabolism	44
[E] Amino acid transport and metabolism	34
[F] Nucleotide transport and metabolism	0
[H] Coenzyme transport and metabolism	0
[I] Lipid transport and metabolism	22
[P] Inorganic ion transport and metabolism	39
[Q] Secondary metabolites biosynthesis, transport and catabolism	38
Poorly characterized	
[R] General function prediction only	2
[S] Function unknown	157
Total	645

[a] See [21] or http://www.ncbi.nlm.nih.gov/COG concerning COG classification.

ESTs with the same inferred function were divided into those with relatively higher and lower levels of expression using the number of matching sequences within an EST library as a measure of expression. Inferred functions with ≥0.37% matches of the ESTs from the Cgm-*Malva* interaction were selected as having relatively higher levels of expression, and these ESTs comprised approximately 30% of all the Cgm-*Malva* ESTs. Based on GenBank NR homology and Table 1 results, there were 39 different plant and 8 different fungal inferred gene functions in the Cgm-*Malva* interaction with relatively higher gene expression levels (Table 3). For these ESTs, the diversity of the sequences in each inferred gene function was estimated by the number of homology clusters, which were based on the number of ESTs with a high degree of similarity in each inferred gene function (Table 3). A comparison of the Cgm-*Malva* ESTs was made with the ESTs of Cs-2wk*Sorghum*, Cs-4wk*Sorghum* and Ct-*Medicago* to determine the relative abundance of gene expression between the different plant–Coll*etotrichum* interactions (Table 3). The ESTs in the *Colletotrichum–Sorghum* and *Medicago* interactions were also compared to the ESTs from healthy *Sorghum* and *Medicago*, respectively. Some of the highly expressed defense mechanism genes in the susceptible Cgm-*Malva* interaction were for the PR proteins, chitinase, osmotin and β-1,3-glucanase, and among these PR proteins, the highest number of ESTs was for β-1,3-glucanase. This was also true for the susceptible Cs-4wk*Sorghum* interaction, although there were more β-1,3-glucanase ESTs in the resistant Cs-2wk*Sorghum* interaction. In both susceptible and resistant *Sorghum*, there were significantly more β-1,3-glucanase ESTs than in healthy *Sorghum*. Another common group of ESTs in the Cgm-*Malva* interaction were heat shock proteins, which all matched chaperones of the hsp70 family. Heat shock protein ESTs were also common in the susceptible Cs-4wk*Sorghum* interaction but not in the susceptible Ct-*Medicago* interaction. Cysteine proteinase and glutathione *S*-transferase ESTs were abundant in both the susceptible Cgm-*Malva* and Ct-*Medicago* interactions. However, cysteine proteinase ESTs were also abundant in the resistant Cs-2wk*Sorghum* interaction, whereas glutathione *S*-transferase ESTs were relatively rare in both the Cs-2wk*Sorghum* and Cs-4wk*Sorghum* interactions.

Since the EST database of the Cgm-*Malva* interaction was smaller than those of the other EST collections examined in Table 3, some of the more highly expressed genes in a *Colletotrichum*–plant interaction may not have been detected. Therefore, for each EST database of the other three *Colletotrichum*–plant interactions, a search was made for inferred gene functions having the same level (≥0.37%) of redundancy as in the Cgm-*Malva* interaction. As a result, a total of 26,009 ESTs from *Colletotrichum*-inoculated plants were examined for EST abundance (Tables 3 and 4). All the ESTs in Table 4 most closely matched plant rather than fungal sequences. Many of these ESTs were significantly less redundant in the Cgm-*Malva* compared to the other interactions. Some inferred functions appeared to be relatively host–pathogen specific among the ESTs, such as an auxin-binding protein and germin-like protein genes, which were absent in all the *Colletotrichum* interactions except the Ct-*Medicago* interaction, where they were also significantly more abundant compared to healthy *Medicago*. ESTs for peroxidase and calmodulin were more common in the susceptible Cs-4wk*Sorghum* and susceptible Ct-*Medicago* interactions than in the resistant Cs-2wk*Sorghum* interaction. Some ESTs matched unknown plant genes that had been identified based on uncharacterized plant EST or genomic sequences. These were EST AU32777 and OSJNBa0032E21.09, which were only common in the susceptible Cs-4wk*Sorghum* interaction, and PnFL-2, which was only common in the susceptible Ct-*Medicago* interaction.

4. Discussion

EST libraries are a useful means to discover new genes and provide information about gene expression levels because they represent a sample of the genes being expressed at a particular time [22]. However, to effectively use the large inventories of gene sequences in EST collections, the data sets must be organized and categorized. The approach taken in this work was to determine the abundance of homologous ESTs with Standalone BLAST analyses using several EST collections as both the query sequences and the matching databases. Although homologous gene function is not necessarily implied by sequence similarity, these types of analyses can give insights into the types and extent of similar-sequence gene expression in different fungal–plant interactions. In this case, several interactions between plants and *Colletotrichum* species that have an intracellular hemibiotrophic infection pattern were selected.

For the susceptible Cgm-*Malva* interaction, mRNA was extracted from *C. gloeosporioides* f. sp. *malvae*-infected mallow leaves when the infection was switching from large primary biotrophic hyphae to thin necrotrophic hyphae [5,17]. Host necrosis is at the earliest stages of development at this time, and relatively few symptoms are visible. For sorghum, the EST collections were created from mRNA of resistant Cs-2wk*Sorghum* and susceptible Cs-4wk*Sorghum* interactions collected at 48 HPI, which corresponds to the timing of attempted penetration of resistant leaves and the period of biotrophic growth of susceptible leaves, respectively [7]. In susceptible sorghum tissue, intracellular infection vesicles were observed at 42 HPI, then intracellular primary hyphae spread to adjacent epidermal cells, and at 66 HPI, thin secondary hyphae developed from the primary hyphae [7]. However, in resistant sorghum, intracellular vesicles were restricted to the initially infected cell, which became filled with cytoplasmic inclusions by 42 HPI, and the cytoplasms of the host cell and fungal vesicle were disintegrating by 66 HPI [7]. For the ESTs from the Ct-*Medicago* interaction, mRNA was extracted from *M.*

Table 3
Percent abundance of sequences (≥0.37%) in the Cgm-*Malva* EST library separated by putative gene function based on highest GenBank match (*e*-value $\leq 10^{-5}$)

Putative gene function	COG[a]	Clusters[b]	Cgm-*Malva*	Cs-2wk*Sorghum*	Cs-4wk*Sorghum*	Ct-*Medicago*
Berberine bridge enzyme	C	2	0.37^{xyz}	0.00	0.00	0.07
Blue copper binding-like protein	C	1	0.37^{xy}	0.00	0.02	0.12
Chlorophyll A–B binding protein	C	5	$0.98^{(y)(z)}$	0.67	3.61	4.25
Ribulose bisphosphate carboxylase small chain	C	1	$1.23^{xy(z)}$	0.03	0.17	5.08
Glyceraldehyde 3-phosphate dehydrogenase	G	2	0.37	0.36	0.36	0.17
Fructose-bisphosphate aldolase	G	3	0.62^{x}	0.04	0.42^{d}	0.37
Acid phosphatase	G	1	0.37^{xyz}	0.01	0.00	0.00
Sugar transporter	G	5	0.62^{xz}	0.08^{d}	0.20^{d}	0.03
Fatty acid synthase[c]	I	1	0.37^{xyz}	0.00	0.00	0.00
Elongation factor 2	J	4	0.62^{yz}	0.28^{d}	0.13	0.03
60S Ribosomal protein	J	3	0.49^{xyz}	0.05	0.02	0.05
40S Ribosomal protein	J	4	0.49^{xyz}	0.05	0.05	0.00
60S Ribosomal protein[c]	J	7	0.98^{xyz}	0.01	0.04	0.18
40S Ribosomal protein[c]	J	6	0.74^{xyz}	0.10	0.06	0.00
DNA binding protein	K	6	0.86^{xy}	0.14	0.32	0.40
MYB-related protein	K	2	0.37^{xy}	0.01	0.06	0.22
Helicase	L	2	0.37^{xyz}	0.06	0.04	0.02
Extensin	M	8	0.98^{xyz}	0.00	0.00	0.00
Cellulose synthase	M	3	0.37^{xyz}	0.03	0.03	0.03
14-3-3-like protein[c]	O	1	0.37^{xy}	0.01	0.02	0.20
Cysteine proteinase	O	3	1.11^{xyz}	0.62^{d}	0.12	0.58^{d}
Ubiquitin[c]	O	3	0.49	0.45	0.67	0.60
Heat shock protein	O	5	1.23^{xz}	0.20	0.93^{d}	0.03
Ubiquitin	O	3	0.62	0.38	0.38	0.25
Glutathione *S*-transferase	O	4	0.74^{xy}	0.09	0.02	0.87^{d}
Catalase	P	2	0.37^{z}	0.23	0.27	0.02
ABC transporter	Q	3	0.37^{xyz}	0.03	0.05	0.02
Alcohol dehydrogenase	Q	1	0.37^{xy}	0.01	0.01	0.12
Cytochrome P450	Q	3	0.74^{xyz}	0.10	0.11	0.10
Zinc-finger protein	R	5	0.62^{xy}	0.04	0.03	0.25
Glycine-rich protein	S	2	0.37^{xyz}	0.02	0.00	0.02
Serine/threonine protein kinase	T	5	0.74^{xyz}	0.16	0.19^{d}	0.22
Serine/threonine protein phosphatase[c]	T	2	$0.37^{c,xyz}$	0.03	0.03	0.03
1-Aminocyclopropane-1-carboxylate oxidase	T	3	0.74^{xy}	0.00	0.03	0.65
Other kinase	T	9	1.35^{xyz}	0.03	0.01	0.08
Receptor protein kinase	T	4	0.49^{xz}	0.04	0.19^{d}	0.03^{d}
Auxin-responsive GH3-like protein	T	2	0.49^{xyz}	0.00	0.02	0.00
S-Adenosylmethionine decarboxylase	T	2	0.37^{z}	0.05	0.28^{d}	0
Calcium dependent protein kinase	T	3	0.62^{xyz}	0.19^{d}	0.13^{d}	0.08
GTP-binding protein	U	4	0.49^{xy}	0.06	0.08	0.37^{d}
GTPase activating protein[c]	U	1	0.37^{xyz}	0.00	0.02	0.00
GTP-binding protein	U	2	0.49^{xy}	0.06	0.14	0.35
β-1,3-Glucanase	V	3	1.48^{xz}	0.44^{d}	1.57^{d}	0.13
Caffeic acid *O*-methyltransferase	V	2	0.37^{xy}	0.00	0.00	0.08
Chitinase	V	4	0.62^{xz}	0.09	0.25	0.12
Non-race specific disease resistance protein	V	1	0.37^{xy}	0.00	0.00	0.00
Osmotin	V	2	0.37^{xz}	0.04	0.32^{d}	0.07

The Cs-2wk*Sorghum* interaction was described as resistant by the sequence generators, while the others have been described as susceptible. The COG classification, the number of homology clusters and the percent abundance of homologues in other EST libraries of interactions with a *Colletotrichum* species are also shown. Using Fisher's exact test, the percentage in the Cgm-*Malva* library was found to be significantly greater than that in Cs-2wk*Sorghum*, Cs-4wk*Sorghum*, or Ct-*Medicago* indicated by x, y, or z, respectively. "()" means significantly less. For each gene function, the relative abundance in the fungal-infected plant EST library was compared to the respective healthy plant EST library using Fisher's exact test ($P = 0.05$) and those followed by "d" are significantly higher or lower, respectively, compared to the healthy plant library.

[a] COG classification was described in Table 2.

[b] Number of homology clusters within each putative gene function in the Cgm-*Malva* library (*e*-value $\leq 10^{-5}$).

[c] Based on Standalone TBLASTX analysis described in Table 1, representative sequences matching these functions have their highest match with a fungal sequence, while the others have their highest match with a plant sequence.

Table 4
Percent abundance of sequences (≥0.37%) in at least one of the EST libraries of an interaction with a *Colletotrichum* species, but not abundant in the Cgm-*Malva* EST library

Putative gene function[a]	COG[b]	Cgm-*Malva*	Cs-2wk *Sorghum*	Cs-4wk *Sorghum*	Ct-*Medicago*
Carbonate dehydratase	C	0.00$^{(z)}$	0.01^{d}	0.07^{d}	0.55^{d}
Oxygen-evolving enhancer protein 1	C	0.12	0.01	0.13	0.47^{d}
Photosystem I antenna protein	C	0.25$^{(yz)}$	0.22^{d}	1.82	2.92^{d}
Photosystem I subunit XI	C	0.25^{x}	0.02^{d}	0.26^{c}	0.52^{d}
Photosystem II 22 kDa protein	C	0.00$^{(z)}$	0.00	0.06	0.53^{c}
Phytochrome A	C	0.00$^{(xy)}$	0.58^{c}	0.49^{c}	0.00
Plastidic aldolase	C	0.25^{x}	0.01^{d}	0.18	0.37^{d}
Aquaporin	G	0.12	0.02^{d}	0.49	0.70^{d}
Pyruvate orthophosphate dikinase	G	0.00^{x}	0.55^{c}	0.34^{c}	0.03
Ethylene responsive element binding factor 5	K	0.12	0.05	0.08	0.48^{c}
Ubiquitin conjugating enzyme	O	0.12	0.18	0.46^{c}	0.45^{c}
Peroxisomal ascorbate peroxidase	Q	0.00	0.07	0.04	0.42^{d}
Thiazole biosynthetic enzyme	Q	0.12	0.00	0.00	0.42^{c}
Auxin-binding protein ABP19a	R	0.00$^{(z)}$	0.00	0.00	0.50^{c}
Germin-like protein	R	0.00$^{(z)}$	0.00	0.00	0.50^{c}
Glycine-rich RNA-binding protein	R	0.12$^{(y)}$	0.18	0.73^{c}	0.50^{c}
Peroxidase 1B	R	0.00$^{(z)}$	0.05	0.34^{c}	0.58^{c}
Regulator of gene silencing	R	0.00	0.14	0.18	0.43^{c}
EST AU032777	S	0.00	0.10	0.42^{c}	0.00
OSJNBa0032E21.09	S	0.00	0.10	0.42^{c}	0.00
PnFL-2	S	0.00	0.02	0.00	0.40^{c}
Calmodulin	T	0.00	0.17	0.26^{c}	0.42^{c}
Ethylene-forming-enzyme-like dioxygenase	T	0.00	0.02	0.04	0.43^{c}

The Cs-2wk*Sorghum* interaction was described as resistant by the sequence generators, while the others have been described as susceptible. Sequences were separated by putative gene function based on top GenBank match (*e*-value $\leq 10^{-5}$) and COG classification. Using Fisher's exact test, the percentage in the Cgm-*Malva* library was found to be significantly more than that in Cs-2wk*Sorghum*, Cs-4wk*Sorghum*, or Ct-*Medicago* indicated by x, y, or z, respectively. "()" means significantly less. For each gene function, the relative abundance in the fungal-infected plant EST library was compared to the respective healthy plant EST library using Fisher's exact test ($P = 0.05$), and those followed by "c" or "d" are significantly higher or lower, respectively, compared to the healthy plant library.

[a] Putative gene function was derived from BLASTX analysis using the GenBank NR database of the sequence which showed the greatest number of homologues within each library. All sequences matched plant sequences in Standalone TBLASTX analysis described in Table 1.

[b] COG classification was described in Table 2.

truncatula 5–8 days after inoculation, which would have been during necrotrophic growth and thus after necrosis had appeared.

Although photosynthesis and chloroplast-related ESTs were relatively abundant in all the EST collections, differences in the number of some of these ESTs, such as those for ribulose bisphosphate carboxylase, fructose-bisphosphate aldolase, carbonate dehydratase, chlorophyll A–B binding protein, two different photosystem I proteins and a photosystem II protein, were observed between the interactions. In the resistant Cs-2wk*Sorghum* interaction, many of those ESTs were at levels several times less than those found in the susceptible Cs-4wk*Sorghum* interaction, which were generally similar to those found in healthy *Sorghum*. This difference between the resistant and susceptible *Sorghum* interactions was not related to the 2-week age difference in the plants as the healthy *Sorghum* EST collection also included ESTs from 2-week-old plants. The general reduction in photosynthesis and chloroplast-related transcripts in the resistant *Sorghum* interaction may be related to plant cell disintegration that occurred in the resistance response to *C. sublineolum*. In contrast, the mRNA in the susceptible *Sorghum* interaction was sampled during the biotrophic phase of infection prior to plant cell death, and the similarity in the number of photosynthesis and chloroplast-related ESTs between the susceptible and healthy *Sorghum* indicated that transcription of at least some photosynthesis-related genes was not affected prior to necrotrophy. However, there was a reduction in several photosynthesis and chloroplast-related ESTs in the susceptible Ct-*Medicago* interaction compared to the healthy *Medicago*, which may also be related to the development of host necrosis, such as in the resistant *Sorghum* interaction, since this mRNA was sampled during the necrotrophic phase of infection.

Excluding ESTs related to photosynthesis and other aspects of central plant metabolism, some of the most abundant ESTs from the Cgm-*Malva* interaction were related to biotic stress, such as β-1,3-glucanase and osmotin, and these were also highly represented in the susceptible Cs-4wk*Sorghum* interaction, but not in healthy *Sorghum*. PR proteins, such as β-1,3-glucanase, chitinase and osmotin, are well known to increase in many plants following infection by a variety of pathogens, and many have been shown to have antimicrobial activity [23].

Other abundant ESTs in the Cgm-*Malva* interaction may also be related to biotic stress, including glutathione

S-transferase and 1-aminocyclopropane-1-carboxylate oxidase, which were also abundant in the Ct-*Medicago* interaction. Glutathione *S*-transferases facilitate detoxification of molecules by conjugation with glutathione or detoxification of specific active oxygen species [24]. An examination of the expression patterns of four glutathione *S*-transferases genes of *M. pusilla* obtained using the Cgm-*Malva* EST collection described in this paper showed that three of these genes were significantly induced following inoculation by *C. gloeosporioides* f. sp. *malvae*, which would be consistent with these ESTs being relatively abundant in ESTs from *Colletotrichum*–plant interactions [25]. In plants, 1-aminocyclopropane-1-carboxylate oxidase converts 1-aminocyclopropane-1-carboxylate to ethylene, and increased ethylene production and expression of a 1-aminocyclopropane-1-carboxylate oxidase gene has been observed in tobacco leaves infected with *C. destructivum* [26].

Cysteine proteinase ESTs were abundant in the Cgm-*Malva* interaction as well as in the susceptible Ct-*Medicago* and resistant Cs-2wk*Sorghum* interactions compared to the susceptible Cs-4wk*Sorghum* interaction, healthy *Medicago* and healthy *Sorghum*. This greater abundance may be related to plant cell death, regardless of whether it was due to resistance in the Cs-2wk*Sorghum* interaction or susceptibility in the Ct-*Medicago* interaction. Cysteine proteinases have been implicated in programmed cell death in both resistant and susceptible plant–pathogen interactions [13,27].

In contrast, the abundance of ESTs for heat shock protein appeared to be inversely related to plant cell death. The highest redundancy of heat shock protein ESTs was found in the susceptible Cs-4wk*Sorghum* interaction compared to healthy *Sorghum*, resistant Cs-2wk*Sorghum* and susceptible Ct-*Medicago* interactions. The higher level of heat shock protein transcripts, therefore, may be related to stresses created during the biotrophic stage of infection in the Cs-4wk*Sorghum* interaction. Heat shock protein ESTs were also abundant in the Cgm-*Malva* interaction where the mRNA was extracted at the transition between biotrophy and necrotrophy. Heat shock proteins in the hsp70 and hsp90 families have been found to be related to host resistance and are believed to be involved in signal transduction for plant defense responses [28].

Although most of the ESTs examined in this work had counterparts in the other *Colletotrichum*–plant interactions, there were also many that had very few or no matches between the EST collections. For example, ESTs for GTP-binding protein, auxin-binding protein 19a and a germin-like protein occurred with relatively high redundancy in the Ct-*Medicago* interaction but were either not observed or observed much less frequently in the other EST collections examined in this study. GTP-binding proteins are involved in cytosolic signal transduction proteins and are required for fungal infections [29], auxin-binding proteins could be associated with auxin imbalances produced during some fungal infections [30], and germin-like proteins are associated with the extracellular matrix and have a variety of functions in plant development and defense [31]. Although such differences may reflect unique aspects of the response of *M. truncatula* to infection, the lack of matching sequences in other EST collections could also be due to a variety of other factors, including a relatively high degree of sequence divergence for certain genes, the expression of an unusual variant of a gene or poor quality or incomplete sequence data.

This work has shown that there are many similarities and differences in EST redundancies between various interactions of plants with intracellular hemibiotrophic *Colletotrichum* species. Using Standalone BLAST analyses to compare within and between EST collections gives a broad overview of the types of genes being expressed during plant infections. Despite the limitations of *in silico* analysis, the putative identification of ESTs as having a fungal or plant origin, assignment of ESTs to COG protein classes and comparison of homologous EST abundance can provide an important starting point for selecting genes related to infection for further study. An example of such a study was the examination of the expression of several glutathione *S*-transferases genes of *M. pusilla*, which were selected because of their relative abundance in the Cgm-*Malva* EST collection, and were shown to have different patterns of induction following infection by *C. gloeosporioides* f. sp. *malvae* [25]. The comparative EST analysis in this study has identified a number of genes, such as those for auxin-binding protein, cysteine proteinase, germin-like protein, heat shock protein, osmotin and others, that appear to be promising candidates for future analysis of diseases caused by intracellular hemibiotrophic *Colletotrichum* species.

Acknowledgements

Funding for this study was provided by the Natural Science and Engineering Research Council of Canada. We wish to thank Yi-Jun Grace Chen for her technical assistance, and Drs. Marie-Michele Cordonnier-Pratt and Deborah Samac for their acknowledgment of the use of their EST libraries. We are grateful to Ms. Bibi S. Andersen and Dr. Søren W. Rasmussen of the Carlsberg Laboratory for assistance in generating and analyzing DNA sequence data.

References

[1] J.A. Bailey, M.J. Jeger, *Colletotrichum*: Biology, Pathology and Control. CAB International, Wallingford, UK, 1992.
[2] D. Prusky, S. Freeman, M.B. Dickman, *Colletotrichum*: host specificity, pathology and host–pathogen interaction, APS Press, St. Paul, USA, 2000.
[3] J.A. Bailey, R.J. O'Connell, R.J. Pring, C. Nash, Infection strategies of *Colletotrichum* species, in: J.A. Bailey, M.J. Jeger (Eds.), *Colletotrichum*: Biology, Pathology and Control, CAB International, Wallingford, UK, 1992, pp. 88−120.

[4] R. O'Connell, S. Perfect, B. Hughes, R. Carzaniga, J. Bailey, J. Green, Dissecting the cell biology of *Colletotrichum* infection processes, in: D. Prusky, S. Freeman, M.B. Dickman (Eds.), *Colletotrichum*: Host Specificity, Pathology and Host–Pathogen Interaction, APS Press, St. Paul, USA, 2000, pp. 57–77.
[5] Y.D. Wei, K.N. Byer, P.H. Goodwin, Hemibiotrophic infection of round-leaved mallow by *Colletotrichum gloeosporioides* f. sp. *malvae* in relation to leaf senescence and reducing agents, Mycol. Res. 101 (1997) 357–364.
[6] J. Bailey, C. Nash, L. Morgan, R. O'Connell, D. TeBeest, Molecular taxonomy of *Colletotrichum* species causing anthracnose on the Malvaceae, Phytopathology 86 (1996) 1076–1083.
[7] P.S. Wharton, A.M. Julian, A cytological study of compatible and incompatible interactions between *Sorghum bicolor* and *Colletotrichum sublineolum*, New Phytol. 134 (1996) 25–34.
[8] D.M. Soanes, W. Skinner, J. Keon, J. Hargreaves, N.J. Talbot, Genomics of phytopathogenic fungi and the development of bioinformatic resources, Mol. Plant–Microbe Interact. 15 (2002) 421–427.
[9] B. Fristensky, M. Balcerzak, D. He, P. Zhang, Expressed sequence tags from the defense response of *Brassica napus* to *Leptosphaeria maculans*, Mol. Plant Pathol. Online (1999), http://www.bspp.org.uk/mppol/1999/0301FRISTENSKY.
[10] S. Kim, I.P. Ahn, Y.H. Lee, Analysis of genes expressed during rice–*Magnaporthe grisea* interactions, Mol. Plant–Microbe Interact. 14 (2001) 1340–1346.
[11] P. Rauyaree, W. Choi, E. Fang, B. Blackmon, R.A. Dean, Genes expressed during early stages of rice infection with the rice blast fungus *Magnaporthe grisea*, Mol. Plant Pathol. 2 (2001) 347–354.
[12] C. Voiblet, S. Duplessis, N. Encelot, F. Martin, Identification of symbiosis-regulated genes in *Eucalyptus globulus–Pisolithus tinctorius* ectomycorrhiza by differential hybridization of arrayed cDNAs, Plant J. 25 (2001) 181–191.
[13] J. Kruger, C.M. Thomas, C. Goldstein, M.S. Dixon, M. Smoker, S. Tang, L. Mulder, J.D.G. Jones, A tomato cysteine protease required for *Cf*-2-dependent disease resistance and suppression of autonecrosis, Science 296 (2002) 744–747.
[14] M. Karlsson, A. Olson, J. Stenlid, Expressed sequences from the basidiomycetous tree pathogen *Heterobasidion annosum* during early infection of Scots pine, Fung. Genet. Biol. 39 (2003) 29–51.
[15] V.K. Thara, J.P. Fellers, J.M. Zhou, In planta induced genes of *Puccinia triticina*, Mol. Plant Pathol. 4 (2003) 51–56.
[16] J.A. Sugui, H.B. Deising, Isolation of infection-specific sequence tags expressed during early stages of maize anthracnose disease development, Mol. Plant Pathol. 3 (2002) 197–203.
[17] S. Jin, R. Xu, Y. Wei, P.H. Goodwin, Increased expression of a plant actin gene during a biotrophic interaction between round-leaved mallow, *Malva pusilla*, and *Colletotrichum gloeosporioides* f. sp. *malvae*, Planta 209 (1999) 487–494.
[18] T. Hsiang, P.H. Goodwin, Distinguishing plant and fungal sequences in ESTs from infected plant tissues, J. Microb. Meth. 54 (2003) 339–351.
[19] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[20] B.C. Sutton, The coelomycetes: fungi imperfect with pycnidia, acervuli and stromata, Commonwealth Mycological Institute, Kew, UK, 1980.
[21] H.H. Rashidi, L.K. Buehler, Bioinformatics Basics, CRC Press, Boca Raton, USA, 2000.
[22] J. Ohlrogge, C. Benning, Unraveling plant metabolism by EST analysis, Curr. Opin. Plant Biol. 3 (2000) 224–228.
[23] S.K. Datta, S. Muthukrishnan, Pathogenesis-related Proteins in Plants, CRC Press, Boca Raton, USA, 1999.
[24] D.P. Dixon, A. Lapthorn, R. Edwards, Plant glutathione transferases, Gen. Biol. 3 (2002) 3004.1–3004.10.
[25] J.D. Dean, P.H. Goodwin, T. Hsiang, *Colletotrichum gloeosporioides* infection induces differential expression of glutathione *S*-transferase genes in *Malva pusilla*, Funct. Plant Biol. 30 (2003) 821–828.
[26] N. Chen, P.H. Goodwin, T. Hsiang, The role of ethylene during the infection of *Nicotiana tabacum* by *Colletotrichum destructivum*, J. Exp. Bot. 54 (2003) 2449–2456.
[27] C. Richael, J.E. Lincoln, R.M. Bostock, D.G. Gilchrist, Caspase inhibitors reduce symptom development and limit bacterial proliferation in susceptible plant tissues, Physiol. Mol. Plant Pathol. 59 (2001) 213–221.
[28] H. Kanzaki, H. Saitoh, A. Ito, S. Fujisawa, S. Kamoun, S. Katou, H. Yoshioka, R. Terauchi, Cytosolic HSP90 and HSP70 are essential components of INF-1 mediated hypersensitive response and non-host resistance to *Pseudomonas cichorii* in *Nicotiana benthamiana*, Mol. Plant Pathol. 4 (2003) 383–392.
[29] H. Schultheiss, C. Dechert, K.H. Kogel, R.A. Huckelhoven, A small GTP-binding host protein is required for entry of powdery mildew fungus into epidermal cells of barley, Plant Physiol. 128 (2002) 1447–1454.
[30] B.A. Cohen, Z. Amsellem, R. Maor, A. Sharon, J. Gressel, Transgenically enhanced expression of indole-3-acetic acid confers hypervirulence to plant pathogens, Phytopathology 92 (2002) 590–596.
[31] F. Bernier, A. Berna, Germins and germin-like proteins: plant do-all proteins. But what do they do exactly? Plant Physiol. Biochem. 39 (2001) 545–554.