An ordered collection of expressed sequences from *Cryphonectria parasitica* and evidence of genomic microsynteny with *Neurospora crassa* and *Magnaporthe grisea*

Angus L. Dawe, Vanessa C. McMains, Maria Panglao, Shin Kasahara, Baoshan Chen and Donald L. Nuss

Center for Biosystems Research, University of Maryland Biotechnology Institute, 5115 Plant Sciences Building, College Park, MD 20742, USA

*Cryphonectria parasitica*, the causative agent of chestnut blight, has proven to be a tractable experimental system for studying fungal pathogenesis. Moreover, the development of infectious cDNA clones of *C. parasitica* hypoviruses, capable of attenuating fungal virulence, has provided the opportunity to examine molecular aspects of fungal plant pathogenesis in the context of biological control. In order to establish a genomic base for future studies of *C. parasitica*, the authors have analysed a collection of expressed sequences. A mixed cDNA library was prepared from RNA isolated from wild-type (virus-free) and hypovirus-infected *C. parasitica* strains. Plasmid DNA was recovered from individual transformants and sequenced from the 5’ end of the insert. Contig analysis of the collected sequences revealed that they represented approximately 2200 individual ORFs. An assessment of functional diversity present in this collection was achieved by using the BLAST software utilities and the NCBI protein database. Candidate genes were identified with significant potential relevance to *C. parasitica* growth, development, pathogenesis and vegetative incompatibility. Additional investigations of a 12-9 kbp genomic region revealed microsynteny between *C. parasitica* and both *Neurospora crassa* and *Magnaporthe grisea*, two closely related fungi. These data represent the largest collection of sequence information currently available for *C. parasitica* and are now forming the basis of further studies using microarray analyses to determine global changes in transcription that occur in response to hypovirus infection.

INTRODUCTION

Genomic studies have provided a wealth of information for numerous organisms, including many fungi (reviewed by Bennett & Arnold, 2001). The recent completion of the *Neurospora crassa* genome (Galagan et al., 2003) has provided the fungal research community with a vast resource of data. This success is being supplemented by efforts to make more widely available sequence data prepared in the private sector for the economically important rice pathogen *Magnaporthe grisea* (www-genome.wi.mit.edu/annotation/fungi/magnaporthe/) and the model fungus *Aspergillus nidulans* (www-genome.wi.mit.edu/annotation/fungi/aspergillus/index.html). As an alternative to sequencing a genome, analyses of expressed sequence tags (ESTs) have also provided a large quantity of gene identification information (reviewed by Skinner et al., 2001). Recent descriptions of EST sequencing projects from phytopathogenic fungi include *Phytophthora infestans* (Kamoun et al., 1999), *Mycosphaerella graminicola* (Keon et al., 2000), *Blumea graminis* (Thomas et al., 2001) and *Magnaporthe grisea* (Kim et al., 2001). This approach has the benefit of producing substantial quantities of data from organisms that may not be the focus of larger-scale genome sequencing ventures. The Internet has provided considerable utility as a universal method to disseminate such information, with many web sites such as COGEME (cogeme.ex.ac.uk: Soanes et al., 2002) that provide search-and-retrieve access to sequence information from one or several organisms.

---

**Correspondence**
Donald L. Nuss
nuss@umbi.umd.edu

**Present address:** Children’s National Medical Center, Center for Genetic Medicine, 111 Michigan Avenue NW, Washington DC 20010, USA.

**Present address:** Department of Molecular Biology, Keio University School of Medicine, 144-8 Ogura, Saiwai, Kawasaki, Kanagawa 212-0054, Japan.

**Present address:** Biotechnology Research Center, Guangxi University, 13 Xiling Road, Nanning, Guangxi 530005, PR China.

**Abbreviations:** EST, expressed sequence tag; GO, gene ontology.

The GenBank accession numbers for the sequences determined in this work are CB686454–CB690670.

Tables of clones classified according to molecular function and biological process are available as supplementary data with the online version of this paper at http://mic.sgmjournals.org.
The focus of our studies, the chestnut blight fungus *Cryphonectria parasitica*, was accidentally introduced into the United States in the early part of the twentieth century (Merkel, 1906) and has been responsible for the destruction of the native population of the American chestnut throughout almost the entire natural range of this hardwood species. *C. parasitica* is a Sordariomycete, a classification of ascomycetes that includes many phytopathogenic fungi, including the genera *Ophiostoma* (containing the species responsible for Dutch elm disease) and *Magnaporthe*. Further interest in *C. parasitica* stems from the observation that natural populations of the fungus can contain virulence-attenuating dsRNA elements, or hypoviruses (Anagnostakis & Day, 1979). The application of molecular biology techniques enabled the development of rigorously tested hypovirus and *C. parasitica* reverse genetics systems (reviewed by Dawe & Nuss, 2001). Subsequent studies led to the isolation and characterization of a number of genes important for virulence, including components of G-protein signalling pathways (Choi et al., 1995; Kasahara & Nuss, 1997; Kasahara et al., 2000; Parsley et al., 2003). Consistent with the prediction that hypoviruses alter host phenotype by modulating cellular signal transduction pathways, differential mRNA display (Chen et al., 1996) indicated that infection with hypovirus CHV1-EP713 results in a large change in the expression profile. While informative, differential display analyses do not readily reveal which host genes are differentially expressed. Cloning and investigation of the *C. parasitica* gene termed 13-J has allowed the construction of a promoter-based reporter system with which to further analyse viral determinants that influence cellular signal transduction pathways (Parsley et al., 2002). However, such studies are also limited due to the use of a single-gene readout to monitor pathway activity, which may not reflect larger, global, changes occurring through the whole transcriptome.

As part of our continuing efforts to understand the interactions of hypoviruses and their host, *C. parasitica*, and the factors that affect fungal virulence, we have undertaken a sequencing project to generate a collection of expressed sequences. At the time of writing, only about 40 *C. parasitica* genes were represented in the NCBI databases. In the absence of a large-scale sequencing effort, we describe below the identification of approximately 2200 new genes using a moderate-throughput approach. By comparison of the recovered sequences to the publicly available information from NCBI, we demonstrate that these ESTs relate to a wide variety of genes with known roles in other organisms and thus provide for preliminary functional assignment of 1255 genes. Further, we explore a small region of genomic microsynteny that appears conserved across three genera and may have functional relevance to G-protein signalling. We anticipate that this large increase in available sequence data for *C. parasitica* will prove of considerable utility to the phytopathogenic research community while also enabling future studies that will take advantage of new microarray technologies to examine large-scale transcriptional alterations under a variety of conditions.

**METHODS**

**Fungal strains and growth conditions.** *Cryphonectria parasitica* strain EP155 (ATCC 38755) and strain EP155 infected with the prototypic severe hypovirus CHV1-EP713 [designated EP713 (ATCC 52571)] were used as representatives of wild-type and virus-infected strains, respectively. Both strains were maintained on solid medium [3.9%, w/v, Difco potato dextrose agar (Becton Dickinson)]. Preparative cultures for RNA recovery were grown for 7 days at room temperature under ambient light with cellophane on the agar surface to permit mycelial harvest.

**RNA extraction and library construction.** Single-stranded RNA was prepared from solid-medium-grown EP155 and virus-infected cultures according to Chen et al. (1996). Poly(A)⁺ mRNA was isolated with oligo-d(T) cellulose (Gibco-BRL). Double-stranded cDNA was synthesized and SalI adapters added, then equal amounts of the cDNA from both EP155 and virus-infected EP155 were mixed prior to cloning using the SuperScript Lambda system (Gibco-BRL/Invitrogen) according to the manufacturer’s instructions. The in vitro packaging of the resulting phage constructs was accomplished with the Gigapack III procedure (Stratagene) and *Escherichia coli* strain Y1090 (ZL) for phage propagation and subsequent storage. Finally, the prepared phage were used to infect *E. coli* DH10B(Zip) (Gibco-BRL/Invitrogen), which permitted plasmid excision resulting in colonies containing cDNA constructs representing both mRNA populations. Individual colonies were transferred into 96-well microtitre plates for cataloguing and storage. Preliminary insert size verification was performed on 20 randomly selected colonies by miniprep and restriction digestion with NotI/SalI. For subsequent experiments not included in this paper, all the clones that provided good sequence data were PCR amplified using T7 and SP6 primers, permitting a larger sampling of insert size data. An additional 120 randomly selected clones were sized in this manner.

**cDNA sequence analyses.** Plasmids were prepared from cultures of stored colonies using the Qiaprep 8 system (Qiagen) with a vacuum manifold. Each preparation was checked for yield by electrophoresis, and then submitted to the DNA Sequencing Core Facility at the Center for Biosystems Research for analysis on ABI 3100 or 377 (Applied Biosystems) machines. Single sequence reads in one direction from the 5’ end of the cloned fragment were obtained using the M13 reverse primer. Results files were scanned using the SeqMan unit of DNAStar running on a Macintosh G4 computer for the presence of known vector sequence and quality of data so that erroneous results (such as those plasmids with no insert, or multiple templates per reaction) could be removed. Short sequences (less than 100 bp) were also eliminated at this time. Following cleanup, the sequences were assembled into contigs using SeqMan and minimum match percentage of 80% across at least 50 bp. The resulting contigs combined the similar sequences of duplicate clones or different clones of the same original cDNA product, thus providing an approximation of the number of individual sequences represented in the dataset. Analysis of the total sequence data was then accomplished in two stages, both performed on an IBM PC running the Red Hat Linux operating system and components of the BLAST program suite (Altschul et al., 1997). In the first stage, a library of DNA sequences was prepared from the known *C. parasitica* items in the NCBI database combined with unpublished sequences from our laboratory. These sequences were then compared, using the BLASTN algorithm, to all of the unknown individual EST sequences (not contigs) so that clones representing known *C. parasitica* genes could be identified. The second stage used
the EST sequences in a BLASTX analysis against the entire non-redundant protein database as available from NCBI. Tables that comprised lists of the clones and their matches, together with NCBI information and alignments for each hit, were generated from the BLAST output using BioPerl (http://www.bioperl.org) software tools. Gene ontology (GO) information was manually added using the Gene Ontology Consortium’s Amigo browser (http://www.godatabase.org/cgi-bin/go.cgi). The entire collection of redundant sequences has been submitted to the NCBI (accession numbers for dbEST: 17474474–17478690, corresponding to GenBank numbers CB686454–CB690670) and are also publicly searchable at the COGEME phytopathogen database (cogeme.ex.ac.uk).

Genomic sequence analyses. An EcoRI–EcoRI fragment approximately 15 kb in length isolated during characterization of the C. parasitica G-protein β-subunit gene (cpgb-1; Kasahara & Nuss, 1997) was subcloned into the pBluescript II SK+ (Stratagene). Following restriction mapping, this insert was divided into smaller sections and subcloned into pBluescript II SK+ or pUC18 to facilitate sequencing, with specific primers being designed to read sequence where required. Sequencing was performed at the CBR core facility and at Keio University, Tokyo, Japan. Assembly of the sequence information with reads in both directions was accomplished using the SeqMan utility of DNASTar. The single contig obtained was determined to accurately represent this region of the C. parasitica genome by comparing the location of the restriction sites predicted with those determined experimentally above. All genomic information for Neurospora crassa and Magnaporthe grisea was obtained from the publicly available databases at the Whitehead Institute/MIT Center for Genome Research (www-genome.wi.mit.edu). Data for Saccharomyces cerevisiae and Schizosaccharomyces pombe were obtained from publicly available information at the Saccharomyces genome resource (genome-www.stanford.edu/Saccharomyces/) and the Sanger Institute (http://www.sanger.ac.uk/Projects/S_pombe/), respectively.

RESULTS AND DISCUSSION

Sequence analyses

Given the lack of sequence data available for C. parasitica, we embarked upon the creation of a library to permit larger-scale transcriptional profiling in future studies. A library designed to reflect the genes expressed in the presence or absence of hypovirus was prepared by mixing cDNA derived from two sources of mRNA: wild-type C. parasitica strain EP155 and EP155 infected with the prototypic virulence-attenuating hypovirus, CHV1-EP713 (strain EP713). This approach was taken to ensure representation within the library of genes whose expression was greatly reduced in either circumstance. Repeated rounds of infection with the phage library and colony picking were performed to generate between 10 and 20 new microtitre plates on each occasion until 60 plates were completed, representing 5760 cDNA clones. Thus catalogued by plate number and coordinate, plasmid minipreps were performed on each, except for plates 38–44, which generated consistently poor yields and were discarded. Six hundred and thirty-four additional colony losses were accounted for by poor growth, insufficient plasmid yield or handling error. By analysing PCR products or restriction-digested plasmid preparations from a random selection of 140 constructs, 95.7% of the inserts were found to be larger than 500 bp, with the vast majority being in the 500–2000 bp range (data not shown). All sequence data were generated by single-pass reads of the insert from the 5’ end. In one case (53F04) a chimeric insert was found to contain two unrelated cDNA sequences (cryparin and a sequence highly similar to kinesin-related protein KIF1C from N. crassa), but no other cloning anomalies were found. The mean read length after vector removal and quality analysis was 608 bp, often including poly(A) regions at the 3’ end. Sequence files that were not included in the following analyses because they were determined to be of insufficient quality, or did not contain an insert of at least 100 bp, totalled 231. Following vector and quality trimming, a total of 2703 587 bp of DNA sequence were included, representing data derived from 4216 clones. Assembly of this sequence information into contigs permitted an estimation of the number of individual sequences included. Totalling the number of contigs present under the particular match parameters as described (see Methods) indicated that 2200 individual gene transcripts were represented but, since the number of contigs obtained in this manner can fluctuate according to the exact matching parameters applied and there is the possibility that two clones may represent different portions of the same coding sequence without overlapping, this number can only be regarded as an approximate value. More than 1600 of these were present only once (unisequences), while the multiply-represented clones followed the distribution shown in

Fig. 1. (a) Distribution of the number of clones per contig following assembly. Omitted for clarity are the contigs corresponding to the two highly abundant sequences (see text) that contain more than 270 clones each. (b) Distribution of the individual hit E values returned from the BLASTX analysis.
Fig. 1(a), with only 28 contigs containing more than 10 clones each. Two of these contigs represented highly abundant mRNA species described below.

Galan et al. (2003) reported that the genome size of \textit{N. crassa} was approximately 40 Mb, similar to the \textasciitilde 45 Mb predicted for \textit{C. parasitica} (Dr Bradley Hillman, Rutgers University, personal communication). Given the close phylogenetic relationship of these two fungi, it is reasonable to postulate that the number of genes in \textit{C. parasitica} is comparable to the 10 082 currently predicted for \textit{N. crassa}. Therefore, in terms of the genome, the collection described here probably represents between 20 and 25 \% of the total gene coding potential of \textit{C. parasitica}.

**Identification of known \textit{C. parasitica} sequences**

A BLASTN analysis was performed using a specific collection of DNA sequences consisting of the publicly available \textit{C. parasitica} information, as well as unpublished data from the Nuss laboratory, in comparison to our entire EST sequence data. In this manner we determined that 598 clones represented 13 known \textit{C. parasitica} genes as detailed in Table 1, 12 from NCBI submitted data and one from unpublished information. Two cDNA sequences were present in far greater abundance than any others and represented almost 6 \% of the total sequences each. These corresponded to ORF B of hypovirus CHV1-EP713 and cryparin, a secreted hydrophobin that has previously been shown to be almost 6 \% of the total sequences each. These corresponded to ORF B of hypovirus CHV1-EP713 and cryparin, a secreted hydrophobin that has previously been shown to be highly expressed (Zhang et al., 1994). Of the 40 sequences from NCBI that were included in this analysis, the fact that 12 were contained within our collection correlates with the projection of approximately 25 \% coverage above.

**Assignment of putative function by BLASTX**

Putative identification of all of the unknown clones was achieved using the BLASTX software tool after removal of the known \textit{C. parasitica} sequences. The remaining 3618 individual EST sequences were used for this analysis instead of contigs to avoid any problems created by artifacts from the contig assembly, although this approach lengthened processing time considerably. Once completed, the data were sorted using the expect \((E)\) value as a measure of significance of the hit, with \(E > 1 \times 10^{-2}\) classed as insignificant. This yielded 3009 clones with hits and 609 with no significant hit. This list was then trimmed manually to remove multiple representatives of the same hit. Thus, 1734 individual hits were recovered distributed across a range of \(E\) values as described in Fig. 1(b). To further sort the data and provide some relevant functional information, this list was truncated to include only those 1255 examples whose hits returned an \(E\) value of \(1 \times 10^{-10}\) or less. Ontology information was then added manually to create two tables of potential groupings relating to both molecular function and biological process according to the definitions outlined by the Gene Ontology Consortium (www.geneontology.org). Both entire lists are available as electronic supplementary data with the online version of this article (mic.sgmjournals.org). It should be noted, however, that this provides only an indication of functional activity or relevance to a particular process, but does not reflect fully the non-redundant and dynamic nature of the GO database terms. These GO terms were developed such that one protein may well fit into several categories (Ashburner et al., 2000), thus necessitating a subjective assessment of the most appropriate single group for assignment in our studies. To examine the groups of sequences isolated, the complete dataset has been used to generate pie charts that illustrate the distribution of the hits obtained across the various molecular function and biological process (Fig. 2) categories. In both cases, a significant number (almost 30 \%) of the hits related strongly to known genes of unknown function. In the case of molecular function, enzymes constituted a further 40-9 \%, of which the hydrolases (27-6 \%) and oxidoreductases (28-2 \%) were most prevalent (Fig. 2a). Of the structural proteins identified, the majority (almost 80 \%) were ribosomal subunits, with cytoskeleton and cell wall structural components.

**Table 1.** Known \textit{C. parasitica} sequences identified in the CEST collection by BLASTN

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Description</th>
<th>No. of sequences</th>
<th>NCBI gi no.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF B</td>
<td>Hypovirus CHV1-EP713</td>
<td>277</td>
<td>9626820</td>
<td>Shapira et al. (1991)</td>
</tr>
<tr>
<td>crp-1</td>
<td>Hydrophobin</td>
<td>270</td>
<td>1706154</td>
<td>Zhang et al. (1994)</td>
</tr>
<tr>
<td>cmcC</td>
<td>Acid proteinase</td>
<td>17</td>
<td>2133287</td>
<td>Jara et al. (1996)</td>
</tr>
<tr>
<td>cpa-1</td>
<td>Endothiapepsin</td>
<td>9</td>
<td>416748</td>
<td>Razanamparany et al. (1992)</td>
</tr>
<tr>
<td>gpd-1</td>
<td>Glyceroldehyde-3-phosphate dehydrogenase</td>
<td>8</td>
<td>120690</td>
<td>Choi &amp; Nuss (1990)</td>
</tr>
<tr>
<td>cpc-1</td>
<td>Cross-pathway control protein</td>
<td>6</td>
<td>4033378</td>
<td>Wang et al. (1998)</td>
</tr>
<tr>
<td>7-8</td>
<td>Uncharacterized</td>
<td>2</td>
<td>Not submitted</td>
<td>Nuss lab. (unpublished)</td>
</tr>
<tr>
<td>cpg-2</td>
<td>Gx subunit</td>
<td>2</td>
<td>563253</td>
<td>Choi et al. (1995)</td>
</tr>
<tr>
<td>cpg-1</td>
<td>MAP kinase kinase</td>
<td>2</td>
<td>4106374</td>
<td>Unpublished</td>
</tr>
<tr>
<td>eapB</td>
<td>Acid proteinase</td>
<td>2</td>
<td>2133286</td>
<td>Jara et al. (1996)</td>
</tr>
<tr>
<td>cpg-1</td>
<td>Gx subunit</td>
<td>1</td>
<td>563251</td>
<td>Choi et al. (1995)</td>
</tr>
<tr>
<td>lac-1</td>
<td>Laccase</td>
<td>1</td>
<td>250714</td>
<td>Choi et al. (1992)</td>
</tr>
<tr>
<td>Small subunit</td>
<td>rRNA gene</td>
<td>1</td>
<td>13540017</td>
<td>Zhang &amp; Blackwell (2001)</td>
</tr>
</tbody>
</table>
making up a further 19%. In terms of the biological processes with which these proteins are associated, more than two-thirds fell into the category of cell growth and maintenance, with far fewer grouped in cell communication (3.1%) or development (0.8%; Fig. 2b). Three-quarters of those sequences classified in the cell growth and maintenance cluster were involved in metabolic processes. From the combined observations in Fig. 2, we conclude that the C. parasitica sequences identified are broadly representative of the entire population of expressed genes in this organism.

**Identification of genes relevant to growth, development and pathogenicity of C. parasitica**

Specific details of the following clones that have returned hits of potential interest are noted in Table 2. This table does not represent an exhaustive list of proteins from any functional category; for that information the reader is referred to the electronic supplementary data (mic.sgmjournals.org).

The interaction between a pathogen and its host is dynamic, requiring the pathogen to continually adjust to changes in the microenvironment and to resistance factors produced by the host. Stress response proteins may fulfil roles to maintain homeostasis under these constantly variable conditions, thus contributing to virulence. This may be manifested in the adjustment of polyol accumulation to cope with hyperosmotic stress as is the case for Cladosporium fulvum (Clark et al., 2003) and M. grisea (Dixon et al., 1999) or production of superoxide dismutases by the human pathogen Candida albicans to protect against oxidative stress (Hwang et al., 2002). Stress-response proteins account for 1% of the hits classified as cell growth and maintenance (Fig. 2) and include a halotolerance protein from Sch. pombe involved in inositol metabolism, as well as superoxide dismutases from Claviceps purpurea, Glomerella cingulata, both plant pathogens, and A. nidulans. Furthermore, we noted the presence of genes corresponding to a general stress-response protein STI35 identified in several species of Fusarium (Choi et al., 1990), and a protein termed GRG-1 from Podospora anserina that is specifically enhanced under conditions of carbon starvation (Kimpel & Osiewacz, 1999). These sequences may represent genes involved in mechanisms of resistance and/or response to host factors that influence phytopathogenicity of C. parasitica.

Given the dynamic host–pathogen interface, the importance of pathways that transmit signals from the cell surface of the pathogen to the nucleus cannot be overstated. Our laboratory has previously identified a number of signal
Table 2. Details of the putative protein identifications based on BLASTX analysis described in the text

<table>
<thead>
<tr>
<th>CEST clone*</th>
<th>Name of match*</th>
<th>E value*</th>
<th>NCBI gi no.*</th>
<th>Organism*</th>
<th>Description/function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stress response</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>03D12</td>
<td>unnamed</td>
<td>$3 \times 10^{-40}$</td>
<td>19075730</td>
<td>Sch. pombe</td>
<td>Putative halotolerance protein, inositol metabolism</td>
<td>Wood et al. (2002)</td>
</tr>
<tr>
<td>14E02, 05D06, 27C05, 35B03, 54B01, 59C05, 11G08, 56F11, 27G10, 55C10, 13C07, 01B09</td>
<td>SOD1</td>
<td>$1 \times 10^{-74}$</td>
<td>15139865</td>
<td>C. purpurea and others</td>
<td>Superoxide dismutase</td>
<td>Unpublished</td>
</tr>
<tr>
<td>28A03</td>
<td>STI35</td>
<td>$1 \times 10^{-43}$</td>
<td>135112</td>
<td>F. oxysporum</td>
<td>Thiazole biosynthesis, stress-inducible protein</td>
<td>Choi et al. (1990)</td>
</tr>
<tr>
<td>03E10, 14F04, 16C07, 18E03, 27C11, 28C01, 46F11, 47E12</td>
<td>GRG-1</td>
<td>$6 \times 10^{-15}$</td>
<td>5051406</td>
<td>P. anserina</td>
<td>Induced upon carbon starvation</td>
<td>Kimpel &amp; Osiewacz (1999)</td>
</tr>
<tr>
<td>Cell signalling/growth regulation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11H07, 18H09, 26F07, 50E10</td>
<td>RHO1</td>
<td>$3 \times 10^{-84}$</td>
<td>15293426</td>
<td>B. graminis and others</td>
<td>Rho family small GTPase</td>
<td>Unpublished</td>
</tr>
<tr>
<td>51B05</td>
<td>Rho2p</td>
<td>$3 \times 10^{-34}$</td>
<td>19115481</td>
<td>Sch. pombe</td>
<td>Rho family small GTPase</td>
<td>Nakano &amp; Mabuchi (1995)</td>
</tr>
<tr>
<td>01D03</td>
<td>RH03</td>
<td>$2 \times 10^{-70}$</td>
<td>13432036</td>
<td>A. fumigatus</td>
<td>Rho family small GTPase</td>
<td>Beauvais et al. (2001)</td>
</tr>
<tr>
<td>32H06</td>
<td>Drab11</td>
<td>$4 \times 10^{-60}$</td>
<td>18376163</td>
<td>N. crassa</td>
<td>Probable GTP binding protein</td>
<td>Unpublished</td>
</tr>
<tr>
<td>06C10</td>
<td>RIP3</td>
<td>$3 \times 10^{-12}$</td>
<td>5453419</td>
<td>D. discoideum</td>
<td>Ras-interacting protein</td>
<td>Lee et al. (1999)</td>
</tr>
<tr>
<td>24A04</td>
<td>ALEA</td>
<td>$5 \times 10^{-32}$</td>
<td>18376134</td>
<td>N. crassa</td>
<td>Related to aimless RasGEF</td>
<td>Galagan et al. (2003)</td>
</tr>
<tr>
<td>36D07</td>
<td>Sbp1p</td>
<td>$1 \times 10^{-45}$</td>
<td>19111914</td>
<td>Sch. pombe</td>
<td>Affects mitosis-interphase transition</td>
<td>He et al. (1998)</td>
</tr>
<tr>
<td>Structural proteins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17E11, 60A05, 21C02, 32H02, 33D09, 58F08</td>
<td>TUB-1</td>
<td>$1 \times 10^{-104}$</td>
<td>12963224</td>
<td>C. lagenarium and others</td>
<td>x-Tubulin</td>
<td>Takano et al. (2001)</td>
</tr>
<tr>
<td>45E09, 46C02</td>
<td>TBG</td>
<td>$5 \times 10^{-41}$</td>
<td>1729857</td>
<td>N. crassa</td>
<td>γ-Tubulin</td>
<td>Heckmann et al. (1997)</td>
</tr>
<tr>
<td>26F08, 05B02, 57H08, 50H03, 60C01</td>
<td>ActA</td>
<td>$1 \times 10^{-121}$</td>
<td>14194449</td>
<td>A. chrysogenum</td>
<td>γ-Actin</td>
<td>Diez et al. (2000)</td>
</tr>
<tr>
<td>31B07, 51B07</td>
<td>DFG5</td>
<td>$8 \times 10^{-94}$</td>
<td>16944622</td>
<td>N. crassa</td>
<td>Cell wall biogenesis in S. cerevisiae</td>
<td>Galagan et al. (2003)</td>
</tr>
<tr>
<td>35A02</td>
<td>ASPB</td>
<td>$2 \times 10^{-96}$</td>
<td>1791305</td>
<td>A. nidulans</td>
<td>Septin B</td>
<td>Momany &amp; Hamer (1997)</td>
</tr>
<tr>
<td>04E11</td>
<td>PCD1</td>
<td>$1 \times 10^{-111}$</td>
<td>5725417</td>
<td>P. brassicae</td>
<td>Putative septin</td>
<td>Singh et al. (2000)</td>
</tr>
<tr>
<td>12C01</td>
<td>unnamed</td>
<td>$2 \times 10^{-11}$</td>
<td>19114968</td>
<td>Sch. pombe</td>
<td>Similar to Dam1p required for spindle maintenance in S. cerevisiae</td>
<td>Wood et al. (2002)</td>
</tr>
<tr>
<td>13B10</td>
<td>Cct7p</td>
<td>$2 \times 10^{-57}$</td>
<td>6322350</td>
<td>S. cerevisiae</td>
<td>Chaperonin component, affects cytoskeleton organization/biogenesis</td>
<td>Rasmussen (1995)</td>
</tr>
<tr>
<td>60G04</td>
<td>Sop2p</td>
<td>$1 \times 10^{-51}$</td>
<td>11359003</td>
<td>Sch. pombe</td>
<td>Arp2/3 complex subunit</td>
<td>Balasubramanian et al. (1996)</td>
</tr>
<tr>
<td>Light detection/response</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37B08</td>
<td>VVD</td>
<td>$1 \times 10^{-38}$</td>
<td>12831203</td>
<td>N. crassa</td>
<td>Clock-controlled repressor of light regulated processes</td>
<td>Heintzen et al. (2001)</td>
</tr>
<tr>
<td>47H02, 28C02</td>
<td>CCG6</td>
<td>$4 \times 10^{-23}$</td>
<td>8488971</td>
<td>N. crassa</td>
<td>Clock-controlled protein</td>
<td>Bell-Pedersen et al. (1996)</td>
</tr>
</tbody>
</table>
Table 2. cont.

<table>
<thead>
<tr>
<th>CEST clone*</th>
<th>Name of match*</th>
<th>E value*</th>
<th>NCBI gi no.*</th>
<th>Organism*</th>
<th>Description/function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>19F01</td>
<td>NOP-1</td>
<td>$1 \times 10^{-35}$</td>
<td>5565892</td>
<td><em>N. crassa</em></td>
<td>Opsin</td>
<td>Bieszke et al. (1999)</td>
</tr>
<tr>
<td>12A07, 55F11, 12E11</td>
<td>HET-C</td>
<td>$7 \times 10^{-47}$</td>
<td>2133323</td>
<td><em>P. anserina</em></td>
<td>Glycolipid transfer</td>
<td>Mattjus et al. (2003)</td>
</tr>
<tr>
<td>19B12</td>
<td>HET-D</td>
<td>$2 \times 10^{-44}$</td>
<td>17225210</td>
<td><em>P. anserina</em></td>
<td>β-Transducin-like WD-40 protein</td>
<td>Espagne et al. (2002)</td>
</tr>
<tr>
<td>13B11, 12D10</td>
<td>MOD-A</td>
<td>$5 \times 10^{-16}$</td>
<td>7452504</td>
<td><em>P. anserina</em></td>
<td>Proline-rich SH-3 domain protein</td>
<td>Barreau et al. (1998)</td>
</tr>
<tr>
<td>60H02, 06E05, 25F07</td>
<td>MOD-E</td>
<td>$1 \times 10^{-103}$</td>
<td>6016265</td>
<td><em>P. anserina</em></td>
<td>HSP90 homologue</td>
<td>Loubradou et al. (1997)</td>
</tr>
<tr>
<td>21B02, 18G09, 12G12</td>
<td>HET-6</td>
<td>$3 \times 10^{-15}$</td>
<td>6578941</td>
<td><em>N. crassa</em></td>
<td>TOL/HET-6/HET-E domain, linked to other incompatibility genes</td>
<td>Smith et al. (2000)</td>
</tr>
</tbody>
</table>

*Where multiple clones matched the same entry repeatedly, or the same protein hit from different organisms, only the best match (lowest E value) is given with corresponding name, gi number and organism of origin. Subsequent CEST clones are listed in match order (increasing E value).
enhance the naturally occurring virulence attenuation characteristics of the virus family Hypoviridae as a method of biological control (reviewed by Dawe & Nuss, 2001). Effective dissemination of the hypovirus through a native population requires the passing of the dsRNA elements from infected to non-infected fungus by means of anastomosis (hyphal fusion; Anagnostakis & Day, 1979). This process is dependent upon the vegetative incompatibility (vic) loci of each strain, and virus transmission can be greatly inhibited by heteroallelism between pairs (Cortesi et al., 2001). At the time of writing, public information concerning the sequence of the vic genes was unavailable, but we have isolated five new C. parasitica sequences in this study that appear to be related to this important process. Genetic analysis of the fungus P. anserina has identified at least 10 loci (het-c, -d and -e; mod-a, -a and -e; idi-1, -2 and -3; pspA), as being involved in the vegetative incompatibility reaction (reviewed by Glass & Kaneko, 2003) and their characterization has revealed proteins probably involved in signal transduction and membrane structural modifications. The HET-C protein has recently been shown to increase glycosphingolipid transfer rates (Mattjus et al., 2003) and also is involved in growth and sporulation (Saupe et al., 1994). Our analysis yielded three EST clones corresponding to het-c. Complete sequencing of these inserts from both 5’ and 3’ ends has enabled us to construct a complete predicted amino acid sequence of the C. parasitica HET-C orthologue. When compared to the sequence for het-c of P. anserina and the predicted protein of N. crassa locus NCU07947.1, significant identity throughout the polypeptides is clear (Fig. 3). Each protein is 53–63% identical to the other two, suggesting considerable conservation of incompatibility components between these fungi. This is further supported by the presence of C. parasitica ESTs corresponding to the vegetative incompatibility proteins HET-D, a WD-40 domain protein of unknown function from P. anserina (Espagne et al., 2002) and HET-6 of N. crassa, whose function is also presently unspecified. The mod genes of P. anserina are believed to code for signalling components that mediate the incompatibility response that can lead to programmed cell death. MOD-A and MOD-E were represented in the C. parasitica EST library and a sixth C. parasitica gene, the Gx subunit CPG-2 identified by Choi et al. (1995), may also be relevant since it is closely related to the MOD-D protein (Loubradou et al., 1999). In contrast to the CPG-1 α subunit, which caused extensive phenotypic changes when absent, deletion of cpg-2 caused only minor changes from wild-type and therefore was not required for pigmentation, sporulation or virulence (Choi et al., 1995). Further analysis of the clones identified in this study may, in conjunction with characterized loci in C. parasitica and other organisms, provide additional insights into this phenomenon and its relevance to the engineering of hypoviruses for enhanced potential as biological control agents.

A representation of the genera of the organisms from which the best BLASTX results for the EST clones were derived is given in Fig. 4. To avoid skewing of the data due to the presence of large numbers of hits from highly analysed genomes, we have scored each genus only once. Thus, the pie charts depict a breakdown of the 132 individual genera represented in the hits analysed above. As expected, most of the highly homologous hits are obtained with sequences from ascomycetous fungi. Narrowing the list to follow the taxonomy of C. parasitica, it is evident that most of the hits are from phylogenetically related organisms. The final grouping represents the Sordariomycetes and contains many phytopathogenic fungi including the genera Claviceps, Glomerella, Gibberella, Nectria, Ophiostoma, Rosellinia and Magnaporthe. N. crassa, while not a pathogen, is also a member of this group. Of the 33 genera from the Pezizomycotina group of ascomycetes that yielded BLASTX hits of $E < 1 \times 10^{-10}$, 16 are plant pathogens and a further six are pathogens of other organisms. Continued study of the sequence data presented here may well illuminate specific pathways and genes that are common between pathogenic fungi and required for virulence.

**Genomic microsynteny among Sordariomycetes**

During investigations in our laboratory of signal transduction pathways that modulate virulence in C. parasitica, we have used the publicly available information to compare C. parasitica genes with those present in the N. crassa genome. Our previous studies had identified CPGB-1, a G-protein β-subunit (Kasahara & Nuss, 1997), and a novel phosducin-like protein called BDM-1 (Kasahara et al., 2000). While the exact roles of these two proteins have not been determined, it is clear from studies in C. parasitica that BDM-1 is essential for correct CPGB-1-mediated signalling (Kasahara et al., 2000), most likely through a direct interaction of the two proteins (A. L. Dawe & D. L. Nuss, unpublished observations), and that the pathway(s) affected subsequently modulate a wide variety of cellular processes including growth, pigmentation, sporulation and virulence.

---

**Fig. 3.** PILEUP comparison of HET-C protein sequences from P. anserina (P.a. HET-C; gi number 537934), N. crassa (NCU07947.1) and C. parasitica (C.p. HET-C; derived from three completely sequenced cDNA inserts). Only those amino acids conserved between all three proteins are highlighted, with dark blocks representing identical residues and grey blocks similar residues.
Similar developmental importance has been ascribed to the gnb-1 gene of N. crassa (Yang et al., 2002). Examination of the N. crassa genomic data revealed a gene, locus NCU00441.1, that appeared similar to bdm-1. Furthermore, this locus was situated in close proximity to gnb-1 (spanning a region of approximately 7 kb) and the ORFs were oriented in the same direction (Fig. 5a). Returning to the original λ phage clone from which the genomic sequence of cpgb-1 was obtained, we were able to establish that part of the ORF of bdm-1 was present at the opposite end to cpgb-1 (data not shown). Subsequent shotgun sequencing of this entire λ clone was then correlated with existing genomic data of the two genes and their immediate surroundings. A single contig of 12982 bp comprising 51 sequences (23 for the top strand, 28 for the bottom, resulting in 2.5-fold average coverage over the entire length) was found to contain the two genes oriented in the same manner as N. crassa (Fig. 5b).

Regions of microsynteny have been previously noted between M. grisea and N. crassa, wherein orientations of ORFs and relative locations were maintained (Hamer et al., 2001). Therefore, further comparison was also made with available data for M. grisea, although the genomic information from release 2.1 did not represent a complete genome assembly (Whitehead Institute/MIT Center for Genome Research, January 2003). However, the orthologues of the Gβ subunit (locus MG05201.1; also known as Mgb1, NCBI gi number 21624377) and bdm-1 (locus MG05200.1) were found to be located on adjacent (linked) cosmids (Fig. 5c). The distance between the two genes cannot be accurately determined until the intervening sequence is released, but the relative orientation of the two reading frames appeared to be identical to that of C. parasitica and N. crassa. Using the COGEME phytopathogen EST database (cogeme.ex.ac.uk/) we have also noted the presence of EST sequences corresponding to orthologues of Gβ (clone Bc-con2700) and BDM-1 (clone Bc-con1300) from the plant pathogen Botryotinia fuckeliana. Presently, appropriate genome information is not available but it would be our prediction that the protein products of these genes will prove important to the correct function of the Gβ subunit given that this consistent arrangement of the two genes is maintained so rigorously.

Fig. 5. A region of microsynteny between three Sordariomycetes: conserved genomic arrangement of two genes that encode a G-protein β-subunit and a phosducin-like potential regulator of Gβ signalling in (a) N. crassa, (b) C. parasitica and (c) M. grisea.
The synteny appears absent in the genomes of *S. cerevisiae* and *Sch. pombe*, since the two genes of interest are at least 100 kb and 40 kb apart, respectively (data not shown). Furthermore, of the two phosducin-like proteins Plp1p (gi number 6320389; most similar to BDM-1) and Plp2p (gi number 6324856) in *S. cerevisiae* (Flanary et al., 2000), PLP2 is located closer to the β-subunit STE4. We have also noted an EST clone (26C10, tentatively termed *bdm-2*) from *C. parasitica* that appears to correspond to PLP2 (*E*= 9×10⁻³⁴), a gene that gave a lethal phenotype when knocked out in *S. cerevisiae* (Flanary et al., 2000). Examing the genome data available for *N. crassa* and *M. grisea* we have determined that sequences that resemble PLP2/bdm-2 are present in both of these fungi also (NCU00617.1 and MG02871.1, respectively) but are located on different linkage groups and supercontigs. Therefore, while both Plp1p and Plp2p were characterized as phosducin-like proteins from *S. cerevisiae*, it appears as if the functional relationship of the Plp2p family to Gβ subunits may be quite different from that which has been hypothesized for BDM-1 (Kasahara et al., 2000) and Plp1p (Flanary et al., 2000).

Of additional note, several *N. crassa* EST clones were found to map to a region approximately 700 bp downstream of the Gβ subunit, but comparison with the entire cDNA fragment of *gub-1* (Yang et al., 2002) indicated that they represented an untranslated region of this ORF (data not shown). There are no other ESTs or predicted ORFs that map to this region of the *N. crassa* genome. Similarly, we have not isolated any *C. parasitica* ESTs that correspond to the equivalent region in this organism. Therefore, it appears as if the two genes are maintained in this arrangement across genera without any intervening genes. Additional comparison of the intervening regions of *N. crassa* and *C. parasitica* failed to indicate any conserved regions of sequence identity or similarity (data not shown). This suggests that the physical proximity of the two genes is the most important feature, perhaps reflecting a requirement for co-segregation of these loci.

**Conclusions**

We have catalogued the largest collection of expressed genes thus far produced for *C. parasitica*, resulting in approximately 1.3 million bp of previously unknown sequence. We have demonstrated that the sequences represented a wide variety of molecular functions involved in many biological processes. The acquired data pointed to considerable similarity of expressed genes between related species and suggested that this may extend to conserved genomic arrangements. With the available database created in this study, used in conjunction with microarray technology, we now have the opportunity to explore on a large scale the transcriptional changes that occur during events such as hypovirus infection. This will undoubtedly lead to enhanced understanding of hypovirus-mediated modulation of host gene expression as well as provide opportunities for future studies of genes that are likely to be critical for fungal pathogenesis.

**ACKNOWLEDGEMENTS**

The authors would like to thank Steve Screen for assistance with setting up the BLAST utilities, Todd Allen for critical evaluation of the manuscript, and Ellen Rosenbloom and Sharmili Mathur for technical assistance. This work was supported in part by Public Health Service grant number GM55981 to D.L.N.

**REFERENCES**


Rasmussen, S. W. (1995). A 37.5 kb region of yeast chromosome X includes the SME1, MEP2, GSH1 and CSD3 genes, a TCP-1-related gene, an open reading frame similar to the DAL80 gene, and a tRNA(Arg). Yeast 11, 873–883.


