

Expressed sequence tags-based identification of genes in the biocontrol agent *Chaetomium cupreum*

HaiYan Zhang · Qian Yang

Received: 3 August 2006 / Revised: 24 September 2006 / Accepted: 26 September 2006 / Published online: 13 January 2007
© Springer-Verlag 2007

Abstract *Chaetomium cupreum* has a potential as biocontrol agent against a range of plant pathogens on the basis of production of antifungal metabolites, mycoparasitism, competition for space and nutrients, or various combinations of these. To explore genes expressed in *C. cupreum*, a cDNA library was constructed from mycelium and 3,066 expressed sequence tags (ESTs) were generated. Clusters analysis enabled the identification of 1,471 unigenes with 392 contigs and 1,079 singleton sequences. Putative functions were assigned to 874 unigenes that exhibited strong similarity to genes/ESTs in public databases putatively containing genes involved in cellular component, molecular function, and biological process. Other 597 ESTs representing novel genes showed no significant similarity to public database resource of NCBI. A proportion of genes was identified related to degradation of pathogen cell wall, antifungal metabolite production, as was estimated in the biocontrol fungus. The paper described is a first step towards the knowledge of the *C. cupreum* genome. The results present the useful application of EST analysis on *C. cupreum* and provide a preliminary indication of gene expression putatively involved in biocontrol.

Keywords *Chaetomium cupreum* · Biocontrol · Expressed sequence tags

Introduction

Currently, much attention is given worldwide to the biological and integrated control means of plant fungal

diseases, and there is thus need for the biocontrol agents (Bourbos and Skoudridakis 1994). *Chaetomium cupreum* has considerable potential for the biocontrol of plant pathogens (Manandhar et al. 1998; Soyong 2003). *Chaetomium* spp. belongs to Ascomycetes, *Chaetomiaceae*, *Chaetomium*, normally exists in soil or organic compost (Seonju and Richard 1999). Some species of *Chaetomium* are very good cellulase and laccase producers and therefore they are important for the biotechnological industry (Ankudimova et al. 1999; Mimura et al. 1999; Chefetz et al. 1998). Agricultural significance of *Chaetomium* spp. lies in that some strains can suppress plant pathogen infection (Hubbard et al. 1982; Aggarwall et al. 2004; Pietro et al. 1992). *C. cupreum* also can antagonizes a wide set of plant pathogens like *Pyricularia oryzae*, *Rhizoctonia solani*, and *Curvularia lunata* (Soyong 2003; Yang 2003; Liao 2002). Biocontrol mechanisms of *Chaetomium* spp. generally are colonizing the plant root or seeds to avoid parasitism of deleterious rhizosphere microorganisms, secreting a variety of degrading enzymes (Inglis and Kawchuk 2002) and antibiotics (Kobayashi et al. 1996; Jiao et al. 2004; Kanokmedhakul et al. 2006), stimulating plant growth, and inducing resistance of plant host (Soyong 2003). Knowledge about the integrated biocontrol mechanism is critical for increasing biocontrol ability of *C. cupreum* and also as a resource for further research on biocontrol fungi privacy. To obtain a more comprehensive understanding of biocontrol mechanism of *C. cupreum*, it is important to identify genes expressed with biocontrol function. ESTs analysis has been shown to allow for rapid, large-scale, functional elucidation of genes (Adams et al. 1991). Nowadays, the application of ESTs project on fungus biology focuses on identifying genes expressed in plant pathogens, for example, identifying genes expressed related to mycelial development and pathogenicity of *Gibberella zeae* (Frances et al. 2003), during the course of *Sclerotinia*

H. Zhang · Q. Yang (✉)
Department of Life Science and Engineering,
Harbin Institute of Technology,
Harbin 150001, People's Republic of China
e-mail: yangqian84@hotmail.com

Sclerotium infecting host (Rugang et al. 2004), involved in the pathogenesis of *Phytophthora parasitica* (Franck et al. 2005) (Fig. 1). EST studies in biocontrol agents has been carried out on *Trichoderma harzianum*; of the 1,740 unigene sequences derived from a *T. harzianum* mycelial cDNA library, 55 genes were identified that may be involved in the process of biocontrol (Liu and Yang 2005). The ESTs approach was also successfully used on *T. harzianum* strain CECT 2413. Three thousand four hundred seventy-eight unique genes from eight different cDNA libraries were constructed. The study offered a base for expression profiling, enabling the identification of genes involved in specific physiological processes containing biocontrol process (Vizcaino et al. 2006). In this paper, the ESTs expressed in mycelium of *C. cupreum* were reported. This is the first report which focuses on research of *C. cupreum* gene expression, which will make a guidance towards the knowledge of the *C. cupreum* genome and provide potentially useful information for elucidating the molecular mechanisms underlying biocontrol competence as a biocontrol agent against plant pathogens.

Materials and methods

Fungal strain

C. cupreum isolate was kindly provided by King Mongkut's Institute of Technology Ladkrabang of Thailand.



Fig. 1 The confrontation of *C. cupreum* with *S. sclerotiorum*

Construction of cDNA library and DNA sequencing

C. cupreum was grown on PD (potato dextrose) medium for 60 h at 27 °C, 140 rpm, then the mycelium was harvested and immediately frozen at –80 °C. Approximately 0.5 g sample were ground to a fine powder in liquid nitrogen and total RNA was extracted with TriZol Reagent (Life Technologies). The mRNA was isolated by means of Oligotex mRNA kits (Qiagen) according to the manufacturer's instructions. The cDNA double strands fragments with size of above 700 bp were ligated with the vector of pBluescriptII and cloned into the *Escherichia coli* DH10B. An aliquot at different dilution of the library was incubated on Luria broth medium at 37 °C in the presence of X-gal and IPTG for overnight. Plasmid DNA templates for the ESTs sequencing were random selected from the clones on agar plate and inoculated into 96-well plate with 5 ml culture of Luria broth containing ampicillin for overnight. Plasmid DNA was isolated by an alkaline lysis procedure and templates for DNA sequencing were prepared by specific polymerase chain reaction (PCR) amplification. Fragments of cDNA was sequenced by means of the T3 primers unidirectional with MegaBASE1000 (Amersham Pharmacia Biotech) DNA sequencer.

Data processing and analysis

Raw sequences files were imported a Unix server and changed into FASTA format by the Phred program (Ewing et al. 1998), the Crossmatch program was used to trim vector sequences, sequences containing more than 5% ambiguous base and shorter than 100 bp. Sequences were further assembled by the Phrap program (<http://www.phrap.org/>) and accuracy of contigs was confirmed with Consed program (Gordon et al. 1998). All unique sequences were compared against the *C. globosum* ESTs, *T. harzianum* ESTs deposited in our laboratory that were derived from the cDNA library of mycelium which was grown on PDA medium and other fungi ESTs sequences that were publicly available in the GenBank database using BlastN (E -value $<10^{-5}$). The functional annotation of ESTs was done to identify putative genes similar in the nonredundant (nr) protein sequence database of GenBank using BlastX algorithm with default parameters and significant similarity was declared when the ESTs with E -value lower than 10^{-10} . Known unigene matches were classified into functional groups according to Gene mapping provided by the GO website with the E -value lower than 10^{-5} (GO, <http://www.geneontology.org/>) (Ashburner et al. 2000). The unigenes were selected for KEGG pathway analysis when the length was more than 30 amino acids, the protein similarity was more than 30%, E -value was lower than 10^{-5} compared to the protein sequence database of KEGG, and then assorted to

different metabolism pathways defined by KEGG according to its corresponding enzyme number (Kyoto Encyclopedia of Genes and Genomes, <http://www.genome.ad.jp/kegg/>) (Kanehisa and Goto 2000). All the sequenced genes have been deposited in the GenBank database with following accession nos.: DV544375–DV547440.

Results

Characterization of the cDNA library

The cDNA library of *C. cupreum* was generated successfully with the following characteristics: the titer of the unamplified cDNA library was 0.93×10^6 pfu/ml. Blue/white plaques selection indicated the recombinant rate of the library was 94.2%. PCR amplification of the random selected recombinant plaques was used to assess the quality of library. The PCR products showed that all the plaques had inserts and the insert size range from 500 to 3,000 bp, most of them was above 1.2 kb. Partial DNA sequences analysis was performed on random selected cDNA clones to assess the quality of cDNA library.

ESTs analysis

Clones of cDNA with insert size more than 700 bp were selected for sequencing from the 5' end of the cDNA fragments. Three thousand sixty-six high quality sequences with a minimum of 100 bp were achieved from 3,647 cDNA clones for further analysis after removing sequences representing ribosomal, mitochondrial, and vector sequences. Minimum, average, and maximum lengths of high quality sequences were 102, 501, and 773 bp, respectively. Using Phrap and Consed program, 3,066 ESTs were assembled into 1,471 unigenes with 392 contigs and 1,079 singletons. BlastX analysis of the high quality sequences revealed that 874 (59.4%) ESTs could be assigned a putative identity based on strong sequence similarity to proteins in the GenBank. Because nonnormalized primary cDNA libraries were used, the number of cDNA clones derived from mRNA could illuminate the expression level of the gene in the mycelial growth stage. The contigs containing ten or more ESTs were listed in Table 1. The most prevalent contig consist of 109 ESTs and had significant similarity to glyceraldehyde-3-phosphate dehydrogenase. The second most abundant contig (63 ESTs) was homologous to coproporphyrinogen oxidase. The third most abundant contig showed no similarity with any known genes in the database representing predicted

Table 1 Assembled contigs containing more than 10 ESTs

Contig no.	Redundancy	Putative function	Related taxon	E value
Contig428	109	Glyceraldehyde-3-phosphate dehydrogenase	<i>Chaetomium globosum</i>	5e-94
Contig63	72	Coproporphyrinogen oxidase	<i>Neurospora crassa</i>	5e-58
Contig30	56	Predicted protein	<i>Neurospora crassa</i>	3e-15
Contig433	54	C-4 sterol methyl oxidase	<i>Candida albicans SC5314</i>	1e-105
Contig313	36	Hypothetical protein	<i>Neurospora crassa</i>	4e-26
Contig2	27	Pyruvate decarboxylase	<i>Aspergillus nidulans</i>	3e-38
Contig39	23	Hypothetical protein	<i>Chaetomium globosum</i>	4e-84
Contig390	21	Xylulose-5-phosphate/fructose-6-phosphate phosphoketolase	<i>Methylococcus capsulatus</i>	5e-62
Contig392	19	EF1-alpha translation elongation factor	<i>Podospora curvicolle</i>	7e-99
Contig425	19	Stress-inducible protein sti35	<i>Gibberella zeae PH-1</i>	3e-84
Contig421	17	Actin	<i>Gaeumannomyces graminis</i>	1e-131
Contig420	17	Glutamine synthetase	<i>Tuber borchii</i>	5e-67
Contig422	16	Aspartic protease	<i>Neurospora crassa</i>	2e-51
Contig371	16	Ammonium transporter	<i>Tuber borchii</i>	3e-26
Contig419	16	ATP synthase protein 9	<i>Podospora anserina</i>	1e-23
Contig6	15	Alcohol dehydrogenase	<i>Neurospora crassa</i>	9e-60
Contig417	15	Hypothetical protein	<i>Magnaporthe grisea 70-15</i>	6e-69
Contig47	14	β -1,3-exoglucanase	<i>Neurospora crassa</i>	1e-38
Contig317	14	ATP citrate lyase	<i>Sordaria macrospora</i>	1e-119
Contig412	13	Histone H2B	<i>Rosellinia necatrix</i>	1e-48
Contig298	13	Pyruvate kinase	<i>Neurospora crassa</i>	5e-72
Contig410	11	Aspartate aminotransferase; complex with alpha-methyl	<i>Caenorhabditis elegans</i>	1e-62
Contig408	11	Heat shock protein 30	<i>Coriolus versicolor</i>	7e-34
Contig411	11	ADP-ATP translocase	<i>Neurospora crassa</i>	1e-144
Contig407	11	Cyclophilin	<i>Caenorhabditis elegans</i>	1e-48

protein. Glyceraldehyde-3-phosphate dehydrogenase, pyruvate kinase, pyruvate decarboxylase, and EF1- α translation elongation factor involved in protein translation, carbon metabolism, and energy production were also highly expressed which indicated that mycelium was in the process of metabolic fastigium. Other abundant ESTs encoded ATP synthase, alcohol dehydrogenase, ADP-ATP translocase implied energy metabolism was also active in the cell of *C. cupreum*. Of the unigenes, approximately 597 (40.6%) ESTs showed no or low similarity to protein sequences in the NCBI database according to BlastX algorithm. The isolation of expressed ESTs with no annotated function is also potentially an important find.

BlastN analysis revealed that 803 sequences had similarity to genes in the GenBank containing ESTs from other organisms. Comparative analysis of *C. cupreum* ESTs with fungi ESTs showed that 709 unigenes had significant similarity to fungal protein sequences. The comparison revealed that *C. cupreum* ESTs had 579 similar genes with *T. harzianum*. The abundant expressed genes in both *C. cupreum* and *T. harzianum* ESTs collections were cyclophilin (CB897191), short-chain dehydrogenase (CK906841), serine palmitoyl CoA transferase (CK907043), sterol 8-isomerase (CK909593), β -glucosidase (CK908283), and alcohol oxidase genes (CK908231). About 36 ESTs of *C. cupreum* had similarity with short-chain dehydrogenase of *T. harzianum*, 15 ESTs had similarity with sterol 8-isomerase, and 66 ESTs had similarity with hypothetical protein of *T. harzianum* (CK907495). The genes relative to biocontrol function expressed together in both *C. cupreum* and *T. harzianum* ESTs collections were class V chitinase, β -1,3-exoglucanase (CK909006), sterol 8-isomerase, and β -glucosidase gene. The comparison of *C. cupreum* ESTs to ESTs of *C. Globosum* using BlastN showed that 398 genes were homologous. The genes of cyclophilin (BP099197), related to nebula protein (BP113071), polyubiquitin, and three unnamed proteins (BP113269, BP113305, and BP113096) were highly abundant in both the EST collections.

Of the ESTs, 2,033 with similarity in BlastX analysis were organized into three parts according to GO classification involved in 60 different functions. More than half of ESTs (59.2%) showed similarity to genes involved in biological process and the most abundantly expressed ESTs associated with metabolism (22.7%); other highly expressed ESTs were associated with cellular physiological process (15.4%), regulation of physiological process (3.15%), response to stimulus (2.9%), and pigment (3.64%). The next most prevalent clones encoded proteins (29.33%) of *C. cupreum* were involved in molecular function including ligase activity (0.25%), transcription factor activity (0.4%), lyase activity (0.64%), and electron transporter activity (0.54%), especially clones with oxido-

reductase activity (11%) and nucleic acid binding functions (2.9%) were overexpressed. Importantly, six ESTs encoded immunomodulatory protein which related to toxin activity (0.3%). Two hundred and thirty-six ESTs (11.6%) were grouped into cellular component parts including respiratory chain complex I (0.1%), unlocalized protein complex (0.05%), intracellular organelle (3.7%), and membrane (0.1%) categories. The diversity of gene expression implied the complexity of metabolism in *C. cupreum* mycelium.

Metabolism pathways analysis

ESTs that showed similarities to known proteins was classified into different metabolisms pathways (Table 2). The most represented pathway was glycolysis showing the important roles in cell metabolism. The second most represented category was porphyrin and chlorophyll metabolism; obviously, fungi cannot produce chlorophyll, but they have heme biosynthetic pathway. The main gene, found in the *C. cupreum* cDNA library relative to heme biosynthesis, was coproporphyrinogen oxidase, which was represented by 72 ESTs. It catalyzes the conversion of coproporphyrinogen-III to protoporphyrinogen-IX and is an essential enzyme in the heme biosynthetic pathway. Forty-seven ESTs (18%) belonged to the citrate cycle category with role in the energy metabolism. Thirty-four ESTs were associated with electron transport and oxidative phosphorylation which also played a key in cell energy metabolism. A lot of genes found in the cDNA library were connected with protein biosynthesis and degradation pathways including glutamate, methionine, cysteine, arginine and praline, histidine, tyrosine, and tryptophan metabolism; phenylalanine, valine, leucine, and isoleucine degradation; valine, leucine and isoleucine biosynthesis, tyrosine, and tryptophan biosynthesis. The proteolytic system found in these categories may play an important role in *C. cupreum* biocontrol ability by proteases taking part in the host cell wall breakdown of pathogens. Some pathways related to saccharide metabolism were also obtained, for example, galactose, fructose, and mannose, starch and sucrose, nucleotide sugars, aminosugars metabolism, glycoprotein, and peptideprotein biosynthesis. We can then proceed to analyze the relationship between the assignment of a metabolic pathway for a protein and its structure, considered in terms of both its conformational fold and its evolutionary superfamily.

Candidate genes associated with biocontrol function

The genes identified among mycelial growth reflected the biocontrol functions of *C. cupreum*. The expressed genes encoding biocontrol-associated proteins were related to cell-wall degradation, antifungal metabolite production,

Table 2 KEGG biochemical pathway mappings of *C. cupreum*

Map	Pathways	ESTs	Percentage	Map	Pathways	ESTs	Percentage
00010	Glycolysis	188	17.52	00480	Glutathione metabolism	8	0.75
00020	Citrate cycle	47	4.38	00500	Starch and sucrose metabolism	16	1.49
00030	Pentose phosphate cycle	3	0.28	00510	Glycoprotein biosynthesis	1	0.09
00031	Inositol metabolism	3	0.28	00520	Nucleotide sugars metabolism	7	0.65
00040	Pentose and glucuronate interconversions	6	0.56	00530	Aminosugars metabolism	7	0.65
00051	Fructose and mannose metabolism	4	0.37	00550	Peptide protein biosynthesis	17	1.58
00052	Galactose metabolism	7	0.65	00561	Glycerolipid metabolism	5	0.47
00053	Ascorbate and aldarate metabolism	5	0.47	00562	Inositol phosphate metabolism	4	0.37
00061	Fatty acid biosynthesis(path 1)	3	0.28	00600	Sphingo glycolipid metabolism	4	0.37
00071	Fatty acid metabolism	24	2.24	00620	Pyruvate metabolism	35	3.26
00100	Sterol, vitamin K, vitamin E, carotenoids biosynthesis	1	0.09	00621	Biphenyl degradation	1	0.09
00120	Bile acid biosynthesis	19	1.77	00622	Toluene degradation	1	0.09
00130	Ubiquinone biosynthesis	14	1.30	00624	Xylene degradation	1	0.09
00150	Androgen and estrogen metabolism	4	0.37	00626	Nitrobenzene degradation	1	0.09
00190	Electron transport and oxidative phosphorylation	34	3.17	00628	Fluorene degradation	1	0.09
00220	Urea cycle and metabolism of amino groups	13	1.21	00630	Glyoxylate and dicarboxylate metabolism	9	0.84
00230	Purine metabolism	41	3.82	00631	1,2-Dichloroethane degradation	4	0.37
00240	Pyrimidine metabolism	6	0.56	00640	Propanoate metabolism	9	0.84
00251	Glutamate metabolism	6	0.56	00650	Butanoate metabolism	13	1.21
00252	Alanine and aspartate metabolism	14	1.30	00670	One carbon pool by folate	6	0.56
00260	Glycine, serine and threonine metabolism	2	0.19	00680	Methane metabolism	35	3.26
00271	Methionine metabolism	12	1.11	00720	Reductive carboxylate cycle in photosynthetic bacteria	6	0.56
00272	Cysteine metabolism	6	0.56	00740	Riboflavin metabolism	5	0.47
00280	Valine, leucine and isoleucine degradation	6	0.56	00760	Nicotinate and nicotinamide metabolism	4	0.37
00290	Valine, leucine and isoleucine biosynthesis	12	1.11	00770	Pantothenate and CoA biosynthesis	9	0.84
00300	Lysine biosynthesis	4	0.37	00790	Folate biosynthesis	4	0.37
00310	Lysine degradation	6	0.56	00860	Porphyry and chlorophyll metabolism	180	16.78
00330	Arginine and proline metabolism	15	1.40	00910	Nitrogen metabolism: reduction and fixation	22	2.05
00340	Histidine metabolism	4	0.37	00920	Sulfur metabolism: reduction and fixation	2	0.19
00350	Tyrosine metabolism	25	2.25	00930	Caprolactam degradation	3	0.28
00351	1,1,1-trichloro-2,2-bis-(4-chlorophenyl)ethane degradation	1	0.09	00940	Flavonoid, stilbene and lignin biosynthesis	3	0.28
00360	Phenylalanine metabolism	9	0.84	00950	Alkaloid biosynthesis I	7	0.65
00361	Gamma-hexachlorocyclohexane degradation	5	0.47	00970	Aminoacyl-tRNA biosynthesis	11	1.03
00380	Tryptophan metabolism	22	2.05	03020	RNA polymerase	3	0.28
00400	Phenylalanine, tyrosine and Tryptophan biosynthesis	13	1.21	03030	DNA polymerase	4	0.37
00410	B-alanine metabolism	5	0.47	03130	Cytochrome C oxidase	1	0.09
00440	Aminophosphonate metabolism	4	0.37	03140	Cytochrome C reductase	10	0.90
00450	Selenoamino acid metabolism	14	1.30	03150	Succinate dehydrogenase	3	0.28
00460	Cyanoamino acid metabolism	1	0.09	00472	d-arginine and ornithine metabolism	2	0.19

proteolytic function, production of enhancing plant resistance substance, and self-metabolism resistance (Table 3). The mycoparasitic process is presumed to involve various fungal cell wall degrading enzymes, including chitinase, glucanase, and glucosidase. Indeed, cDNAs encoding cell wall hydrolases were found in the library of *C. cupreum*. Of the 142 ESTs associated with biocontrol, 35 (24.65%) ESTs encoding cell wall degrading enzymes accounted for the highest percentage including β -1, 3-exoglucanase (DV546424, DV544842, DV544713, DV546012, DV544470, DV545432, DV544477, DV547403, DV544454, DV544484, DV545129, DV546364, DV545072, DV544562, DV545952, DV544979, DV546676, DV546546, DV546460, DV545542, DV544737, DV544416), endoglucanase IV (DV545380), β -glucosidase 5 (DV545481, DV545793, DV545126, DV546743, DV546476, DV548542), β -glucosidase (DV546469), β -glucosidase 6 (DV547247), and chitinase (DV546055, DV544732, DV544989, DV544376). Of them, ESTs similar to β -1,3-exoglucanase gene was found 22 times, ESTs similar to β -glucosidase was observed eight times and chitinase was observed four times. Chitinase is considered as potential biocontrol agents against fungi and plays a key role in antagonism of some plant pathogens. β -1,3-glucanases (exo or endo) and β -glucosidase is a major wall-lytic enzyme responsible for the dissolution of the β -1,3-glucan in host cell wall during mycoparasitism (Figs. 2 and 3). Proteases are also regarded as antifungal proteins secreted as part of the defense responses by the biocontrol

fungus. The subtilisin-like serine protease (DV546459, DV546294, DV544423, DV546484) genes and aspartic proteinase genes (DV545776, DV547231, DV546784, DV546176, DV546884, DV544891, DV547261, DV547321, DV545627, DV544971, DV546366, DV546126, DV544706, DV545010, DV544844, DV545924, DV546967, DV545636, DV546174, DV546528, DV547357) also were identified in the cDNA library of *C. cupreum*. In our studies, 25 ESTs were identified related to proteolytic degradation including four ESTs from subtilisin-like serine protease and 21 ESTs from aspartic proteinase which accounted for about 19.72% in total biocontrol ESTs. It is also known that the ability of *Chaetomium* spp. to produce antifungal metabolites such as chetomin, rotiorinols A–C, multidrug resistance protein, (DV545494, DV544640, DV544681, DV546719, DV545349, DV547367), polyketide synthase (DV547431, DV546945), isopenicillin N synthase, and related dioxygenases genes (DV546194, DV546282, DV544663) were obtained in the cDNA library of *C. cupreum*. Terpene compounds are involved in the biocontrol process due to their antifungal properties. C-4 sterol methyl oxidase (DV544986, DV544993, DV545404, etc.), C-8 sterol isomerase (DV544900, DV547178), and sterol C-22 desaturase (DV546589) are three enzymes in the biosynthesis of one kinds of triterpene derivatives—ergosterol. The xylanase genes of *C. cupreum* were obtained which related to induction of plant resistance. Fungus itself has resistance ability to outside stress. DHA14-like major

Table 3 The ESTs associated with biocontrol function of *C. cupreum*

EST cluster	Redundancy	Putative function	Related taxon	E value
Cell wall degradation	35			
Contig360	22	β -1, 3-exoglucanase	<i>Neurospora crassa</i>	3e-44
ma1222	1	Endoglucanase IV	<i>Hypocrea jecorina</i>	3e-52
Contig143	6	β -glucosidase 5	<i>Coccidioides posadasii</i>	3e-14
ma2809	1	β -glucosidase	<i>Aspergillus niger</i>	3e-39
ma5255	1	β -glucosidase 6	<i>Coccidioides immitis</i>	2e-28
Contig320	4	Chitinase	<i>Aspergillus nidulans</i>	1e-83
Proteolytic degradation	25			
Contig207	4	Subtilisin-like serine protease	<i>Podospira anserina</i>	1e-111
Contig422	21	Microbial aspartic proteinases	<i>Cryphonectria parasitica</i>	9e-50
Antifungal metabolites	68			
Contig56	6	Mfs-family multidrug resistance protein	<i>Aspergillus fumigatus</i>	8e-67
Contig65	2	Polyketide synthase	<i>Cochliobolus heterostrophus</i>	1e-11
Contig251	3	Isopenicillin N synthase and related dioxygenases	<i>Pseudomonas fluorescens</i>	4e-40
Contig85	2	C-8 sterol isomerase erg-1	<i>Neurospora crassa</i>	5e-67
ma2962	1	Sterol C-22 desaturase	<i>Symbiotaphrina buchneri</i>	8e-22
Contig433	54	C-4 sterol methyl oxidase	<i>Candida albicans SC5314</i>	1e-105
Enhancing plant resistance	3			
Contig252	3	G/11 family xylanase	<i>Scytalidium thermophilum</i>	9e-91
Self-metabolism resistance	11			
Contig66	2	DHA14-like major facilitator	<i>Botryotinia fuckeliana</i>	3e-43
Contig393	9	β -1, 3-glucan binding protein	<i>Neurospora crassa</i>	1e-11



Fig. 2 The mycoparasitism of *C. cupreum* on *P. capsici*

facilitator (DV544766, DV546717) can provide the tolerance of fungus to toxic compounds and two ESTs of *C. cupreum* encoding DHA14-like major facilitator genes were found. Nine ESTs encoding β -1,3-glucan binding protein (DV544687, DV545882, DV546061, DV546443, DV546757, DV546969, DV547432, DV548098, and DV548106) are involved in the recognition of invading microorganisms that acts to protect itself.



Fig. 3 The mycoparasitism of *C. cupreum* on *R. solani*

Discussion

EST analysis has been proven to be an efficient approach to identify genes expressed under a wide variety of conditions and systems about fungi. In vascular wilt pathogen *Verticillium dahlia* (Neumann and Dobinson 2003), more than 2,000 ESTs were generated from two cDNA libraries. In *Mycosphaerella graminicola*, 704 unigenes were identified from budding conidia cDNA library (Keon et al. 2000). A lot of emphases were placed on the identification of pathogenetic genes, however, work on the biocontrol fungus were poorly studied. Knowledge about the molecular aspects of biocontrol fungus to plant pathogens is critical for increasing resistance of biocontrol fungus and plant through genetic engineering. In this study, a cDNA library was constructed from *C. cupreum* mycelium; 3,066 ESTs assembled and 1,471 unigenes were generated. BlastX analysis showed that 59.4% (874) unigenes exhibited strong similarity to genes/ESTs in public databases. As expected, about half of genes showed high similarity to the special genes of fungi that reflected the fungal feature of *C. cupreum*. It included class V chitin synthase (DV546782) and glucan synthase (DV547356). The cell wall of fungus was mainly composed of glucan and chitin and the acquisition of genes related to fungus cell wall biosynthesis could offer clues for understanding mechanism of fungus cell wall biosynthesis and help to develop new methods for plant fungi disease control. Contig building reduced the total number of nucleotides used for further analysis and increased the length and quality of the sequence from 423 to 1,732 nucleotides. Eleven ESTs matched with heat shock protein 30 which is associated with cell resistance. Contigs had significant similarity to fungal genes in the GenBank and they encode proteins with interesting functions containing C-4 sterol methyl oxidase (contig433, 54 copies), aspartic protease (contig422, 16 copies), and β -1,3-exoglucanase (contig47, 14 copies), which may play role in the antagonisms of biocontrol fungus.

Biocontrol is a complex process including the release of hydrolase, production of antifungal substance, and further penetration into the host mycelium. The strong biodegradation and substrate colonization performances of *C. cupreum* are the result of metabolic versatility and high secretory potential that leads to the production of diversified sets of hydrolytic enzymes. Similarly, the direct attack of *C. cupreum* to phytopathogen is based on the secretion of complex cocktails of enzymes involved in the degradation and further penetration of the fungal host cell wall. The results showed that the EST approach is a mean of discovering potentially genes that may be involved in the biocontrol mechanism. Most studies have proven that chitinase (Bruce et al. 1995; Rachel and Ilan 1998) and

β -1,3-exoglucanase (Chiu and Tzean 1995) are classical enzymes connected with fungus biocontrol. ESTs obtained from cDNA library of *C. cupreum* had high similarity to Endoglucanase IV of *Hypocrea jecorina* (3e-52). Chitinase can decompose pathogen cell wall which catalyzes the hydrolysis of chitin, an unbranched polymer of β -1,4-N-acetylglucosamine. Especially, it can destroy neonatal chitin on mycelia tip to inhibit pathogens growth, thereby generating oligosaccharides containing GlcNAC that may act as elicitors for inducing general antifungal response of biocontrol fungus (Susanne et al. 1999). Moreover, chitinase also plays an important physiological and ecological role in ecosystems as recyclers of chitin by generating carbon and nitrogen sources. Twenty-two ESTs is close to β -1,3-exoglucanase from *Neurospora crassa* (3e-44) which can break down β -1,3-glucan of pathogen cell wall, and experiments showed that it could inhibit mycelial growth of fungus obviously. More importantly, the oligosaccharide that was released from pathogen cell wall also can induce resistance of host plant to pathogens. Owing to there is β -1,3-glucan and lipoid layer outer chitin at fungus mycelium tip, the cooperation of chitinase and β -1,3-glucanase has stronger inhibitory effects than any of them. β -glycosidase is also strongly antifungal against a wide range of plant pathogens by decompose cell wall of fungus and it is synergistic with other cell wall degradation enzymes (Giuliano et al. 2001). Proteases can take part in the host cell wall breakdown course or act as proteolytic inactivators of pathogen enzymes (Kapteyn et al. 1996; Elad and Kapat 1999). Protease has biocontrol function including aspartic proteinase, subtilisin (Maria et al. 2004; Yang et al. 2005), and trypsin-like proteases. Elicitation of plant defense response by a 18-kDa protein from *T. virens* with similarity to serine proteases has been recently described (Hanson and Howell 2004). Aspartic proteinase related to mycoparasitic and plant root colonization activities also expressed in *T. asperellum* T-203 (Ada et al. 2004). The Aspartic proteinase of *T. harzinum* induced by fungal cell wall suggested that it might participate in early stages of the mycoparasitic process (Suarez et al. 2005). The acquisition of aspartic protease gene from *C. cupreum* provided base for further study of its function during biocontrol. Polyketide synthase is involved in biosynthesis of diverse carbon skeletons from simple activated carboxylic acid units. The products of the complex pathways possess a wide range of pharmaceutical properties, including antibiotic, antifungal, and immunosuppressive activities. For example, it plays a key role in cercosporin biosynthesis by *Cercospora nicotianae* and the antibiotic erythromycin biosynthesis by *Saccharopolyspora erythraea*. Three ESTs were obtained for encoding isopenicillin N synthase and related dioxygenases which were the key enzymes of penicillin and cephalosporins biosynthesis. However, whether *C. cupreum* can produce

cephalosporins remained to be determined. Some biocontrol fungus could improve plant host growth and induce plant resistance to a variety of plant pathogens. Xylanase could induce plant resistance by opening K^+ , H^+ , and Ca^{2+} ion channels, synthesizing PR protein and ethylene (Dean et al. 1989). DHA14-like major facilitator is a member of a novel major facilitator superfamily from *Botrytis cinerea* and provides tolerance to toxic compounds, such as camptothecin and cercoaporin, as well as certain fungicides (Fleissner et al. 2002); two ESTs of *C. cupreum* encoding DHA14-like major facilitator genes were found. β -1,3-glucan binding protein is involved in the recognition of invading microorganisms which binds specifically to β -1,3-glucan and lipoteichoic acid and causes aggregation of invading microorganism that acts to protect the cell of organism.

This is the first report of the use of high-throughput EST analysis to examine the genes expression in *C. cupreum* mycelium. The EST collection and its annotation provide a significant resource for fundamental and applied research of *C. cupreum*. Abundant expressed clones could be determined and their expression patterns and physiological significances are examined in more detail. Especially, with ESTs approach, 142 ESTs biocontrol-associated sequences were identified. The ESTs analysis exemplifies a fertile approach for the identification genes expressed during mycelium development in *C. cupreum* and investigation genes with biocontrol for further elucidation the integrated mechanism against fungal pathogens at the molecular level. The biocontrol genes are a valuable resource for further research. In the future, on the one hand, the mycelial cDNA library of *C. cupreum* will be constructed with plant fungus pathogens cell wall or chitin as inducement. On the other hand, the ESTs biocontrol-associated sequences will be cloned for the transform into the biocontrol fungi or plant for further applications on agriculture.

References

- Ada V, Michal H, Ilan C (2004) Isolation of two aspartyl proteases from *Trichoderma asperellum* expressed during colonization of cucumber roots. FEMS Microbiol Lett 238:151–158
- Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropolos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Mccombe WR, Venter JC (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252:1651–1656
- Aggarwall R, Tewari AK, Srivastava KD, Singh DV (2004) Role of antibiotics in the biological control of spot blotch (*Cochliobolus sativus*) of wheat by *Chaetomium globosum*. Mycopathologia 157:369–377
- Ankudimova NV, Baraznenok VA, Becker EG, Okunev ON (1999) Cellulase complex from *Chaetomium cellulolyticum*: isolation and properties of major components. Biochemistry 64:1068–1073

- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. *Nat Genet* 25:25–29
- Bourbos VA, Skoudridakis MT (1994) Integrated control of *Botrytis cinera* in non-heated greenhouse tomatoes. *Phytopathology* 9:327–328
- Bruce A, Srinivasan U, Staines HJ, Highley TL (1995) Chitinase and laminarinase production in liquid culture by *Trichoderma* spp. and their role in biocontrol of wood decay fungi. *Int Biodeterior Biodegrad* 47:337–352
- Chefetz B, Chen Y, Hadar Y (1998) Purification and characterization of laccase from *Chaetomium thermophilum* and its role in humification. *Appl Environ Microbiol* 64:3175–3179
- Chiu SC, Tzean SS (1995) Glucanolytic enzyme production by *Schizophyllum commune* Fr. during mycoparasitism. *Physiol Mol Plant Pathol* 46:83–94
- Dean JFD, Gamble HR, Anderson JD (1989) The ethylene biosynthesis-inducing xylanase: its induction in *Trichoderma viride* and certain plant pathogens. *Phytopathology* 79:1071–1078
- Elad Y, Kapat A (1999) The role of *Trichoderma harzianum* protease in the biocontrol of *Botrytis cinerea*. *Plant Pathol* 1105:177–189
- Ewing B, Hiller L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res* 8:175–185
- Fleissner A, Sopalla C, Weltring KM (2002) An ATP-binding cassette multidrug-resistance transporter is necessary for tolerance of *Gibberella pulicaris* to phytoalexins and virulence on potato tubers. *Mol Plant Microbe Interact* 15:102–108
- Frances T, Xu JR, Phillip SM, Robert GH, Kistler HC (2003) Analysis of expressed sequence tags from *Gibberella zeae* (anamorph *Fusarium graminearum*). *Fungal Genet Biol* 38:187–197
- Franck P, Joelle A, Eric G, Yanne J (2005) LB Gene identification in the oomycete pathogen *Phytophthora parasitica* during in vitro vegetative growth through expressed sequence tags. *Fungal Genet Biol* 42:611–623
- Giuliano BGD, Lorito M, Scala F, Harman GE (2001) Cloning, sequence and structure of a gene encoding an antifungal glucan 1,3- β -glucosidase from *Trichoderma atroviride* (*T. harzianum*). *Gene* 277:199–208
- Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. *Genome Res* 8:195–202
- Hanson LE, Howell CR (2004) Elicitor of plant defense responses from biocontrol strains of *Trichoderma virens*. *Phytopathology* 94:171–176
- Hubbard JP, Harman GE, Eckenrode CJ (1982) Interaction of a biological control agent *Chaetomium globosum* with seed coat Microflora. *Can J Microbiol* 28:431–437
- Inglis GD, Kawchuk LM (2002) Comparative degradation of oomycete, ascomycete, and basidiomycete cell walls by mycoparasitic and biocontrol fungi. *Can J Microbiol* 48:60–70
- Jiao W, Feng Y, Blunt JW, Cole AL, Munro MH (2004) Chaetoglobosins Q, R, and T, three further new metabolites from *Chaetomium globosum*. *Nat Prod* 67:1722–1725
- Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27–30
- Kanokmedhakul S, Kanokmedhakul K, Nasomjai P, Louangsysouphanh S, Soyong K, Isobe M, Kongsaree P, Prappai S, Suksamrarn A (2006) Antifungal azaphilones from the fungus *Chaetomium cupreum* CC3003. *J Nat Prod* 6:891–895
- Kapteyn JC, Montijn RC, Vink E (1996) Retention of *Saccharomyces cerevisiae* cell wall proteins through a phosphodiester-linked β -1,3/ β -1,6-glucan heteropolymer. *Glycobiology* 6:337–345
- Keon J, Bailly A, Hargreaves J (2000) A group of expressed cDNA sequences from the wheat fungal leaf blotch pathogen, *Mycosphaerella graminicola* (*Septoria tritici*). *Fungal Genet Biol* 29:118–133
- Kobayashi H, Namikoshi M, Yoshimoto T, Yokochi T (1996) A screening method for antimetabolic and antifungal substances using conidia of *Pyricularia oryzae*, modification and application to tropical marine fungi. *J Antibiot* 49:873–879
- Liao QS (2002) Study on antagonism to appressorium formation in *magnaporthe grisea*. Zhejiang University Press, Zhejiang University
- Liu PG, Yang Q (2005) Identification of genes with a biocontrol function in *Trichoderma harzianum* mycelium using the expressed sequence tag approach. *Res Microbiol* 156:416–423
- Maria JP, Baek JM, Juan M (2004) Functional analysis of tvsp1, a serine protease-encoding gene in the biocontrol agent *Trichoderma virens*. *Fungal Genet Biol* 41:336–348
- Manandhar JB, Thapliyal N, Cavanaugh KJ (1998) Interaction between pathogenic and saprobic fungi isolated from soybean roots and seeds. *Mycopathologia*:69–75
- Mimura S, Rao U, Yoshino S, Kato M, Tsukagoshi N (1999) Depression of the xylanase-encoding cgxA gene of *Chaetomium gracile* in *Aspergillus nidulans*. *Microbiol Res* 153:369–376
- Neumann MJ, Dobinson KF (2003) Sequence tag analysis of gene expression during pathogenic growth and microsclerotia development in the vascular wilt pathogen *Verticillium dahliae*. *Fungal Genet Biol* 38:54–62
- Pietro AD, Gut RM, Pachlatk JP, Schwinn, FJ (1992) Role of antibiotics produced by *Chaetomium globosum* in biocontrol of *Pythium ultimum*, a causal agent of damping off. *J Physiol Biochem* 82:131–135
- Cohen-Kupiec R, Chet I (1998) The molecular biology of chitin digestion. *Curr Opin Biotechnol* 9:270–277
- Rugang L, Roger R, Lone B, Andrew G, Sharpe Ginette SS, Cathy C, Dwayne DH (2004) Interaction of *Scletinia sclerotium* with a resistant *Brassica napus* cultivar: expressed sequence tag analysis identifies genes associated with fungal pathogenesis. *Fungal Genet Biol* 41:735–753
- Seonju L, Richard T (1999) Phylogenetic relationships of *Chaetomium* and similar genera based on ribosomal DNA sequences. *Mycologia* 91:434–442
- Soytong K (2003) Application of a new broad spectrum biological fungicide for environmental plant protection. In: Yang Q (ed) *Biological control and biotechnology*, 1st. Heilongjiang Science and Technology Press, Harbin, pp 70–85
- Suarez MB, Sanz L, Chamorro MI, Rey M, Gonzalez FJ, Llobell A, Monte E (2005) Proteomic analysis of secreted proteins from *Trichoderma harzianum* identification of a fungal cell wall-induced aspartic protease. *Fungal Genet Biol* 42:924–934
- Susanne Z, Christiane G, Kathrin P, Sheridan LW, Robert LM, Csaba F, Matteo L (1999) Chitinase gene expression during mycoparasitic interaction of *Trichoderma harzianum* with its host. *Fungal Genet Biol* 26:131–140
- Vizcaino JA, Gonzalez FJ, Suarez MB, Redondo J, Heinrich J, Delgado JJ, Hermosa R, Gutierrez S, Monte E, Llobell A, Rey M (2006) Generation, annotation and analysis of ESTs from *Trichoderma harzianum* CECT 2413. *BMC Genomics* 7–193
- Yang Q (2003) Evaluation of *Chaetomium* for biological control of *Fusarium* Wilt of tomato in P. R. China. In: Yang Q (ed) *Biological control and biotechnology*. Heilongjiang Science and Technology Press, Harbin, pp 25–36
- Yang JK, Huang XW, Tian BY, Wang M, Niu QH, Zhang KQ (2005) Isolation and characterization of a serine protease from the nematophagous fungus, *Lecanicillium psalliotae*, displaying nematocidal activity. *Biotechnol Lett* 27:1123–1128