Mol. Cells *29*, 131-144, February 28, 2010
DOI/10.1007/s10059-010-0018-y

Molecules and Cells

Comparative Analysis of Expressed Sequence Tags from the White-Rot Fungi (*Phanerochaete chrysosporium*)

Dae-Won Kim[1,4], Aeri Kim[1,2,4], Ryong Nam Kim[1], Seong-Hyeuk Nam[1,2], Aram Kang[1,2], Wan-Tae Chung[3], Sang-Haeng Choi[1], and Hong-Seog Park[1,2,*]

Comprehensive analysis of the transcriptome of the *P. chrysosporium* is a useful approach to improve our understanding of its special and unique enzyme system and fungal evolution in molecular and industrial aspects. In order to unveil the functional diversity of this white-rot fungus in gene level and the expression patterns of its genes, in this study we carried out sequencing and annotation of 4,917 *P. chrysosporium* expressed sequence tags (ESTs). Through our bioinformatic ESTs analysis, we elucidated that 1,751 genes were derived from the present dataset of 4,917 ESTs, based on clustering and comparative genomic analyses of the ESTs. Of the 1,751 unique ESTs, 1,006 (57.5%) had homologues and orthologues in similarity searches. Our *P. chrysosporium* ESTs showed many genes for encoding 23 secreted proteins, many proteins for the degradation of cellulose and hemicelluloses, and heat shock proteins for stress resistance, which explain the reason why *P. chrysosporium* is very important and unique white-rot fungus in dealing with contaminated resources and in degrading lignin and in applying this organism to several industrial aspects. In addition, comparative analysis has shed the fresh light on the mystery about how its unique enzyme system and stress resistance have been evolved differently from its closest relatives.

INTORDUCTION

Biomass is materials derived from not only living organisms, such as plants and animals, but also garden waste, manure and crop residues. It is a renewable energy source that influences the carbon cycling. Thus, biomass energy and fuel are becoming more and more indispensable materials for a sustainable future. Among the materials, woody biomass of numerous types of plants, including sorghum, corn, hemp, switchgrass, poplar and willow, with appropriate development of harvest and storage strategies are standing out conspicuously for large-scale production of bioenergy.

Generally, woody lignocellulosic biomass is complex mixture consisting of three polymers in close association: hemicellulose, cellulose and lignin. These polymer substances constitute the cell walls of wood and form strength and stiffness of the stem (Chabannes et al., 2001). Unlike the most abundant natural biopolymers hemicellulose and cellulose, lignin is highly resistant towards biological and chemical degradation, and provides mechanical resistance to wood. Because lignin is an integral part of the secondary cell walls of plants and give strength to wood, which consist principally of syringyl (S), guaiacyl (G) with various ether bonds and condensed linkages (Boerjan et al., 2003; Chabannes et al., 2001). That is, insoluble and amorphous aromatic material lacks stereoregularity and is not susceptible to hydrolytic attack in lignin (Boerjan Ralph et al., 2003).

To develop a practicable biological delignification process, lignin-degrading microorganisms were coordinated (Kumar et al., 2008). Basidiomycetes, called the "Higher Fungi" within the Kingdom Fungi, are used as the most conspicuous wood rotters due to their ability to modify or degrade lignin. Wood-rotting basidiomycetes are generally classified as white-rot, brown-rot, soft rot and staining fungi based mainly on different types of wood decaying (Martinez et al., 2005). Among them, a white-rot fungus (*P. chrysosporium*) was the most intensively studied until now. Because it secretes lignin and manganese peroxidases (LiP and MnP) which are extracellular enzymes in charge of an initial attack on lignin for the biodegradation of lignin (Kirk and Farrell, 1987; Gold and Alic, 1993; Gold et al., 2000; Martinez et al., 2005). In addition, *P. chrysosporium* has been widely known as model organism to study biochemistry, physiology and diverse toxic environmental pollutants for new industrial application (Decelle et al., 2004; Duranova et al., 2009; Wesenberg et al., 2003). For this reason, the *P. chrysosporium* genome (approximately 30 million base pairs) has been recently sequenced and assembled for studying efficient degradation of all wood components and treatment of a variety of chemical pollutants (Martinez et al., 2004). Moreover, current

[1]Genome Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 305-806, Korea, [2]Department of Functional Genomics, University of Science and Technology, Daejeon 305-333, Korea, [3]Technology Services Division, National Institute of Animal Science, Rural Development Administration, Suwon 441-706, Korea, [4]These authors contributed equally to this work.
*Correspondence: hspark@kribb.re.kr

Received July 30, 2009; revised October 22, 2009; accepted October 26, 2009; published online January 8, 2010

Keywords: comparative genomics, enzyme, expressed sequence tags, secretome, white-rot fungi

technological advances in genomics provided exciting opportunities for exploring basically genetic, biological and biochemical aspects of the *P. chrysosporium*. For example, long serial analysis of gene expression (LongSAGE) of the model fungus, *P. chrysosporium* has revealed some significant up or down regulation of genes during monitoring gene expression (Minami et al., 2007). The most recent paper described a novel method to identify extracellular enzyme produced during growth on woody biomass using 2D gel electrophoresis and liquid chromatography (LC)/MS/MS (Matsuzaki et al., 2008; Sato et al., 2007).

However, available information about transcripts that are expressed in *P. chrysosporium* is still limited. Thus, to reveal the genetic repertoire of *P. chrysosporium*, we sequenced and analyzed 4,992 cDNA at first step toward clarifying the role of individual gene. In the present study, we provide a first insight into the transcriptome of the *P. chrysosporium* EST sequencing and apply a newly established bioinformatics pipeline for the clustering and comparative analyses of data set against a wide range of organisms. The representative EST from this dataset has been functionally annotated in protein level to assign its putative function. Additionally, we constructed gene ontology, pathway and secretome using this dataset.

MATERIALS AND METHODS

P. chrysosporium cultivation

The white-rotting fungus, *P. chrysosporium* strain 20741 (ATCC 24752), was obtained from Korean Forest Research Institute. Fungal cultures were grown either in liquid or on solid substrates, as previously described (Sato et al., 2007). Solid substrate (wood) cultures were either grown in polypropylene bags in which the water content was 50% or as 'submerged' cultures. The submerged cultures included largely liquid medium, containing a 1% carbon source of glucose, cellulose or wood. The liquid stationary culture was grown at 37°C, and was flushed with water-saturated O_2 on every 3 days. When grown on solid substrate medium, *P. chrysosporium* inoculum was prepared by growth in 250 ml Erlenmeyer flasks containing 20 g millet, 10 g wheat bran and 30 ml water at 30°C for 7 days. These cultures were used to inoculate polypropylene growth bags containing 850 g red oak sawdust, 100 g millet, 50 g wheat bran and 1 L distilled water.

RNA isolation and cDNA library Construction

P. chrysosporium mycelia were submerged in liquid nitrogen in a pre-chilled grinding jars and grinding ball on a bed of dry ice. The *P. chrysosporium* put in pre-chilled grinding jars used pulverized Mixer Mill MM301 (Retsch GmbH, Germany). *P. chrysosporium* transferred 15-ml polypropylene tube filled with liquid nitrogen and stored at -80°C. Total mRNA extracted from fragmented frozen tissue using the TRI reagent (MRCgene, USA). Total RNA was purified from total RNA (100 ug) using the absolutely mRNA Purification Kit (Stratagene, USA) according to manufacturer's instruction.

To construct cDNA library, a directional λ ZAP cDNA synthesis/Gigapack III gold cloning kit (Stratagene, USA) was used. Reverse transcriptase of mRNA for production of first stand cDNA synthesis was primed from the poly-A tail using an oligo-dT linker-primer containing an *Xho*I cloning site. Following second strand synthesis, *Eco*RI linkers were ligated to the 5′-termini. The *Xho*I digestion releases the *Eco*RI adapter and residual linker–primer from the 3′ end of the cDNA. These two fragments are separated on a drip column containing Sepharose® CL-2B gel filtration medium. The size fractionated cDNA (above 500 bp) is then precipitated and ligated to the ZAP Express vector (pBK-CMV). The primary library produced by in vitro packaging of the ligation product with a ZAP Express cDNA Gigapack III Gold cloning Kit (Stratagene, USA).

Plasmid isolation and cDNA sequencing

The cDNAs were plated to LB-kanamycin plate (Rectangle, 23.5 cm × 23.5 cm) with X-gal/IPTG for blue/white selection. White colonies were randomly manually picked and inoculated into 15 384-well plate (Corning, USA) containing 40 μl TB/kanamycin, then incubated for 16 h at 37°C with fixation culture. Sequences of cDNAs inserts were determined from their 5′ end of clones using a BigDye Terminator Sequencing Kit ver 3.1 (Applied Biosystems, USA) and a 3730XL DNA analyser (Applied Biosystems).

EST cleaning and clustering

The ESTs were initially analysed and annotated using PESTAS, an automated EST analysis platform (unpublished, http://pestas.kribb.re.kr). In our study, the analysis pipeline is consisting of three steps (Fig. 1). In STEP I, *P. chrysosporium* EST trace data were base-called using the program Phred from trace chromatogram data using Phred score of 13 (Ewing and Green, 1998; Ewing et al., 1998). The sequences were then processed with the Cross_Match (http://www.phrap.org), RepeatMasker (http://www.repeatmasker.org/) and SeqClean (http://compbio.dfci.harvard.edu/tgi/software/) to filter out the sequences comprising vectors, *E-coli*, repetitive elements and mitochondrial DNA. Trimmed sequences over the 100 bp in length were clustered and assembled into putative unique EST objects (uniESTs) by TGICL (Pertea et al., 2003) and CAP3 (Huang and Madan, 1999), employing default options.

Functional categorization of ESTs

In STEP II, to assign a putative function to *P. chrysosporium*, we took into account BLASTX bets-hit descriptions and subsequent alignments with an e-value below 1E-10, at least 30% identity along 30 amino acid exactly matches, resulting from sequence comparison against non-redundant protein database (NR) at NCBI. And then, since a large portion of these ESTs has not yet been annotated, we further characterized domain/family in uniESTs using InterPro database version 17 (HMMPfam, HMMSmart, HMMTigr, HMMPanther and SuperFamily) (flagged as true by InterProScan with E-value < 1e-2) (Hunter et al., 2009). We also classified our uniESTs associated with Gene Ontology (GO) terms at the Protein-level annotation using BLAST2GO (cut-off value ≤ 1e-10) (Conesa et al., 2005). In order to predict pathway in the *P. chrysosporium*, we mined the enzyme information from the annotated results of UniProtKB database (2008) that contained an EC number in description with matching E-value <= 1e-10. To infer the existing functional knowledge via homology, we use eggNOG database (http://eggnog.embl.de/) (Jensen et al., 2008) that contains more fine-purified groups for selected subsets of organisms. To analysis differential gene expression, we implemented the web version of IDEG6 software (Romualdi et al., 2003), utilizing the Audic-Claverie approach with a significance threshold of 0.01 (http://telethon.bio.unipd.it/bioinfo/IDEG6). The final 4,917 set of quality ESTs reported in this paper have been deposited in DDBJ databases under accession numbers FS233675-FS238591.

Comparison of *P. chrysosporium* ESTs to *N. crassa*, *S. cerevisiae* and *S. pombe*

Protein database, *N. crassa* (28,090 ESTs), *S. cerevisiae*

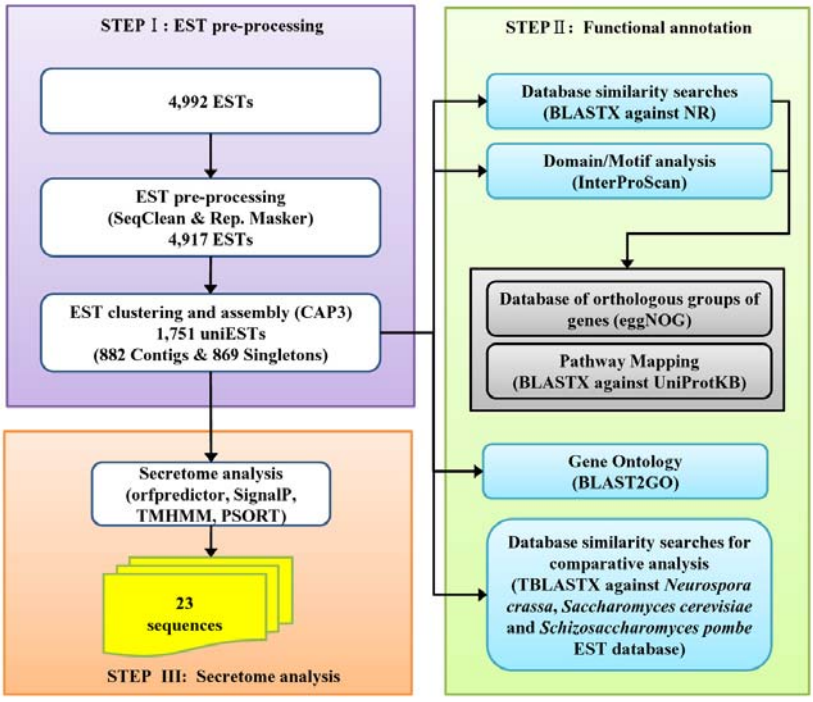

Fig. 1. Overall strategy of Bioinformatics analysis for *P. chrysosporium* ESTs

(34,915 ESTs) and *S. pombe* (8,123 ESTs) from GenBank (20 November 2008), for comparative transcriptome were generated in-house for homologue search. A local search against the new fungal database was made using TBLASTX algorithm. A cut off score 50 was set as stringency threshold.

Secretome analysis

From the ORFs inferred from uniESTs, secreted proteins were predicted using a combination of four programs [ORFpredictor (Min et al., 2005), SingalP (Bendtsen et al., 2004), TMHMM (Krogh et al., 2001) and PSORT II (Nakai and Horton, 1999)], to minimize the number of false positive predictions. Firstly, we identified protein-coding regions ORFs in uniESTs with exactly starting initiation codon encoding the amino acid methionine (Met), employing ORFpredictor. Secondly, SingalP 3.0 was used to predict the presence of secretory signal peptides and signal anchor for each predicted uniEST proteins, using both the neural network and Hidden Markov model. In order to exclude the erroneous prediction of putative transmemberane (TM) sequences as signal sequence, TMHMM, a membrane topology prediction program, was the applied. We further validated the list of secreted proteins, using extracellular localization using PROST II.

RESULTS AND DISSCUSION

Overview of *P. chrysosporium* EST analysis

Of 4,992 clones sequenced, a total of 4,917 high-quality ESTs were obtained with achieving a 98.5% of sequencing success rate (STEP I, Fig. 1). These pre-processed ESTs ranged from 105 to 846 bp, with a mean of 695 bp and a standard deviation (S.D.) was 101 bp. After clustering, the mean length of the contigs increased to 711 bp. The cluster analysis of the 4,917 ESTs from *P. chrysosporium* yielded 1,751 unique ESTs (uniESTs) (882 contigs and 869 singleton sequences), 1,305 of which were aligned against JGI predicted genes (Table 1).

Table 1. *P. chrysosporium* transcriptome features

	Numbers
Total sequence reads	4,992
Total analyed reads (average size)	4,917 (695 bp)
Total number of assembled sequences (average size)	1,751 (711 bp)
Contigs	882
Singlets	869
JGI predicted genes hit[a]	1,305
Total annotated genes	1,006
BLASTX	928
InterProScan	78

[a], JGI predicted genes include 10,048 putative *P. chrysosporium* genes.

To assign putative functions to uniESTs, we combined each uniEST's closest homolog search using BLAST algorithm against non-redundant (NR) protein database at NCBI, InterProScan database (HMMPfam, HMMSmart, HMMTigr, HMMPanther and SuperFamily) version 17 and Hidden Markov Model (HMM) domain/families search that are informative for a given putative function (STEP II, Fig. 1) (Hunter et al., 2009). Firstly, each uniEST was given the functional annotation from the best match identified after BLASTX search against NR protein database at NCBI. Non-assigned uniESTs after BLASTX were annotated using the results of InterProScan (see Supplementary Table 1). Of the 1,751 uniESTs, 57.5% (1,006) have significant sequence similarity to NR protein database and/or conserved protein domain at InterPro scanning. As shown at Supplementary Table 1, since a number of clusters had significant alignments to predicted hypothetical proteins according to the result of BLASTX at NR protein database, we added extra annotation description from BLAST2GO (Conesa et al., 2005), UniProtKB

(2008) and eggNOG (Jensen et al., 2008) database to take into account various function definitions with which each uniESTs actually was annotated (STEP II, Fig. 1). The remaining 745 uniESTs after primary annotation, accounting for 42.5% of total uniESTs, were classified as unknown ones. Among our data, unknown genes have taken considerable portion of uniESTs. However, this result is not surprising, because there are a limited number of annotated Basidiomycota nucleotide and protein sequences currently deposited in the public databases.

Highly represented P. chrysosporium ESTs

EST copy number can be used to estimate gene expression diversities in an organism, although there is an artificial collection of cloned cDNA fragments that might limit the estimation of over-expression of certain transcripts. From a virtual expression analysis of the *P. chrysosporium*, 882 clustered uniESTs from 4,048 high-quality ESTs were analysed. In our study, 47 most abundant transcripts in the *P. chrysosporium* EST collection, having eleven or more EST copies in each cluster, were identified based on their identity putatively assigned by BLASTX analysis of the assembled contigs (Table 2). If there are a number of uniESTs corresponding to the same accession number, we chose one uniEST with the highest cluster number of representative cluster ID. As a result, the majority of annotated genes with integrated molecular functions were divided into two class such as house-keeping genes (chaperone protein, ribosomal protein, transcription/elongation factor), and enzymes (phosphatase, glycoside hydrolase, aryl-alcohol dehydrogenase, oxidoreductase and ubiquitin-conjugating enzyme). We also found 13 unknown uniESTs which have no any sequence similarity in current databases based on our cut off value (see "Materials and Methods"). Majority of these genes have been identified in the *coprinopsis cinerea okayama* 7#130 fungus (an inky cap mushroom), requiring detailed molecular characterisation to understand their functions in the *P. chrysosporium*.

As can be deduced from Table 2, *P. chrysosporium* ESTs having a largest cluster number of 8 clusters containing 290 ESTs (5.9% of total EST sequences) are heat shock protein 30 (HSP30) genes, which help a living organisms in maintaining homeostasis of life (Santoro, 2000) and in keeping a constant thermotolerance (Plesofsky-Vig and Brambl, 1995). Except HSP30, we identified three more uniESTs, which belong to molecular chaperones family (EPM005LAAA11C000322, EPM 005LAAA11C000358 and EPM005LAAA11C000356). The heat shock proteins (HSPs) are a subset within a larger group of genes encoding for stress response proteins as molecular chaperones and show dramatically increased expression when physiological conditions are substantially stimulated by an extremely higher temperature than optimum value for living (Vorob'eva, 2004). The results obtained in this study show that many highly expressed uniESTs of the *P. chrysosporium* respond to the thermal stress with the induction of HSPs, including small heat shock proteins (HSP20, HSP30 and so on). These proteins can be associated with the defence mechanisms of thermal stress, being part of the physiological adaptation and alteration of this organism in the environment with extreme temperatures' variations and other stress conditions. Thus, these results explain the reason why *P. chrysosporium* shows the greater capacity of adaptation to stress conditions, especially temperature changes in the environment.

In addition, we found a lot of antioxidative proteins such as Mn-superoxide dismutase, thioredoxin reductase, thioredoxin peroxidase, glutathione S-transferase and catalase, even though low expression. It suggested that these uniESTs seem to be up-regulated as an antioxidative response against the severe environmental change such as oxidative stress caused by hazardous chemicals (Iimura and Tatsumi, 2002; Lee et al., 1998).

Abundance of conserved domains in P. chrysosporium ESTs

We analysed the most abundant conserved domain from the *P. chrysosporium* using Pfam, and found that 34.7 (%) of the 1,751 uniESTs encoded proteins similar to members of 380 Pfam protein families (E-value < 1e-2). Pfam domain families were ranked according to the number of the *P. chrysosporium* uniESTs containing every Pfam domain, and the top 10 were represented in Table 3. Among these Pfam families, "WD40 repeat" domain, known to regulate a complex cell differentiation process (Poggeler and Kuck, 2004), was ranked in the top, and followed by "RNA recognition motif, RNP-1", "Serine/threonine protein kinase-related" and "Ubiquitin". Interestingly, we also found that several kinds of heat shock proteins are encoded by uniESTs (Tables 2 and 3). In addition, we had compared previously reported Pfam annotation results of *P. chrysosporium* genome with our Pfam domain analysis results of *P. chrysosporium* uniESTs (Martinez et al., 2004). To identify the correlation between genome data and transcriptome data, we used IDEG6 online software (http://telethon.bio.unipd.it/bioinfo/IDEG6_form/index. html). This comparison revealed that nine genes were differently classified among the domain family categories, such as "WD40 repeat", "RNA recognition motif, RNP-1", "Ubiquitin", "Ankyrin", "Heat shock protein Hsp20", "Heat shock protein 70", "RNA polymerase II, heptapeptide repeat, eukaryotic", "Heat shock factor (HSF)-type, DNA-binding", "FAD dependent oxidoreductase" and "UTP-glucose-1-phosphate uridylyltransferase". As a result, "WD40 repeat", "RNA recognition motif, RNP-1" and "Serine/ threonine protein kinase-related" were highly ranked between genome and transcriptome annotation data simultaneously, but "Ubiquitin", "Ankyrin" and "Heat shock protein" domain families are different in their rankings between them.

Degradation of cellulose and hemicellulose and lignin

P. chrysosporium is known to express genes encoding a wide spectrum of enzymes for lignin degradation (Minami et al., 2007). The genome harbours many putative carbohydrate-active enzymes including 166 glycoside hydrolases, 57 glycosyltransferases and 14 carbohydrate esterases, comprising 69 families within its genome (Martinez et al., 2004). Our annotation results also revealed 26 carbohydrate- and lignin-active enzymes such as esterase, exo-1,3-β-glucanase, β-1,6-glucan synthetase and α-1,2-mannosyltransferase (Table 4). Especially, we found that 17 glycoside hydrolase family (GH) members were expressed, and they are divided into ten families (GH5, GH10, GH13, GH16, GH18, GH37, GH61, GH63, GH71, GH72), most of which have been implicated in the degradation of hemicelluloses or pectin. These results show that *P. chryso-sporium* genome contains the genetic information encoding a large group of glycoside hydrolases. Furthermore, some previous studies reported the expressions of genes encoding various enzymes including exoglucanase and xylanase in protein level (Abbas et al., 2005; Wymelenberg et al., 2005). Taken together, these results suggest that the transcriptome of *P. chrysosporium* contains many genes encoding proteins that might play key roles in cellular process for degradation of the major polymers of wood.

In addition, we have identified that our EST data included several cytochrome P450 family members, which belong to the up-regulated genes possibly related to lignin degradation (Supplementary Table 1). Very interestingly, most of the above-mentioned genes, which have been known to be involved in the degradation of cellulose, hemicelluloses and lignin, have been

Table 2. The most abundant transcripts in *P. chrysosporium*

Number	Representative cluster ID	# of ESTs[a]	# of cluster	Database	Accession ID[b]	E-value[b]	Description from NR or InterProScan best hit[b]	Description from eggNOG, BLAST2GO and UniProtKB
1	EPM005LAAA11C000005	290 (5.9%)	8	NR	BAA76589.1	8.70E-94	Heat shock protein 30 [Trametes versicolor]	Heat shock protein 30 - Trametes versicolor (White-rot fungus) (Coriolus versicolor) (UniProtKB)
2	EPM005LAAA11C000311	84 (1.71%)	3	NR	EAU84739.1	1.50E-37	Predicted protein [Coprinopsis cinerea okayama7#130]	Predicted protein - Laccaria bicolor (strain S238N-H82) (Bicoloured deceiver) (Laccaria laccata var. bicolor) (UniProtKB)
3	EPM005LAAA11C000312	74 (1.51%)	1	NR	XP_001223637.1	1.30E-48	Hypothetical protein CHGG_04423 [Chaetomium globosum CBS 148.51]	Predicted CDS Pa_6_8780 - Podospora anserine (UniProtKB)
4	EPM005LAAA11C000313	55 (1.12%)	2	NR	XP_384017.1	3.20E-23	Hypothetical protein FG03841.1 [Gibberella zeae PH-1]	Glycine-rich rna-binding (BLAST2GO)
5	EPM005LAAA11C000314	46 (0.94%)	1	Unknown	-	-	-	-
6	EPM005LAAA11C000315	33 (0.68%)	1	NR	XP_710281.1	4.40E-34	Hypothetical protein CaO19.6835 [Candida albicans SC5314]	Predicted protein (BLAST2GO)
7	EPM005LAAA11C000345	28 (0.57%)	2	INTERPRO	SSF52821	0.0018	Rhodanese/Cell cycle control phosphatase	-
8	EPM005LAAA11C000317	28 (0.57%)	1	Unknown	-	-	-	-
9	EPM005LAAA11C000322	27 (0.55%)	1	NR	EAU83184.1	2.80E-28	Hypothetical protein CC1G_07866 [Coprinopsis cinerea okayama7#130]	DnaJ-class molecular chaperone with C-terminal Zn finger domain (eggNOG)
10	EPM005LAAA11C000323	27 (0.55%)	2	NR	XP_572310.1	2.70E-46	Hypothetical protein [Cryptococcus neoformans var. neoformans JEC21]	Glycoside hydrolase family 16 protein (BLAST2GO)
11	EPM005LAAA11C000327	25 (0.51%)	2	NR	EAU83952.1	4.20E-32	Predicted protein [Coprinopsis cinerea okayama7#130]	c6 transcription factor (BLAST2GO)
12	EPM005LAAA11C000331	25 (0.51%)	2	Unknown	-	-	-	-
13	EPM005LAAA11C000332	24 (0.49%)	3	NR	EAU84230.1	6.70E-86	Hypothetical protein CC1G_08160 [Coprinopsis cinerea okayama7#130]	Protein involved in mRNA catabolism, deadenylation-dependent decay (eggNOG)
14	EPM005LAAA11C000358	21 (0.43%)	6	NR	EAU90954.1	1.40E-43	Hypothetical protein CC1G_02341 [Coprinopsis cinerea okayama7#130]	Molecular chaperone (small heat shock protein) (eggNOG)
15	EPM005LAAA11C000336	21 (0.43%)	5	NR	XP_001269594.1	8.90E-43	Hypothetical protein ACLA_028940 [Aspergillus clavatus NRRL 1]	Putative uncharacterized protein - Aspergillus clavatus (UniProtKB)
16	EPM005LAAA11C000330	21 (0.43%)	1	Unknown	-	-		-

(continued)

Number	Representative cluster ID	# of ESTs[a]	# of cluster	Database	Accession ID[b]	E-value[b]	Description from NR or InterProScan best hit[b]	Description from eggNOG, BLAST2GO and UniProtKB
17	EPM005LAAA11C000325	18 (0.37%)	1	Unknown	-	-	-	-
18	EPM005LAAA11C000337	16 (0.33%)	1	NR	EAU83071.1	9.40E-72	Hypothetical protein CC1G_12379 [Coprinopsis cinerea okayama7#130]	U1 small nuclear ribonucleoprotein 70 kDa (eggNOG)
19	EPM005LAAA11C000347	16 (0.33%)	3	NR	EAU83368.1	2.40E-50	Predicted protein [Coprinopsis cinerea okayama7#130]	Predicted protein - Coprinopsis cinerea (strain Okayama-7 / 130 / FGSC 9003) (Inky cap fungus) (Hormographiella aspergillata) (UniProtKB)
20	EPM005LAAA11C000334	15 (0.31%)	2	NR	AAG53696.1	7.40E-18	Hypothetical protein [Schizophyllum commune]	Predicted protein - Laccaria bicolor (strain S238N-H82) (Bicoloured deceiver) (Laccaria laccata var. bicolor) (UniProtKB)
21	EPM005LAAA11C000342	15 (0.31%)	2	NR	XP_001263382.1	7.70E-34	D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase, putative [Neosartorya fischeri NRRL 181]	Phosphoketolase (eggNOG)
22	EPM005LAAA11C000338	15 (0.31%)	1	NR	XP_758812.1	2.20E-32	Hypothetical protein UM02665.1 [Ustilago maydis 521]	Protein translation factor sui1 (BLAST2GO)
23	EPM005LAAA11C000339	15 (0.31%)	1	Unknown	-	-	-	-
24	EPM005LAAA11C000340	15 (0.31%)	1	Unknown	-	-	-	-
25	EPM005LAAA11C000343	14 (0.29%)	2	NR	EAU91069.1	2.50E-15	Predicted protein [Coprinopsis cinerea okayama7#130]	Predicted protein - Laccaria bicolor (strain S238N-H82) (Bicoloured deceiver) (Laccaria laccata var. bicolor) (UniProtKB)
26	EPM005LAAA11C000341	14 (0.29%)	1	NR	EAU91394.1	3.80E-11	Predicted protein [Coprinopsis cinerea okayama7#130]	Predicted protein (BLAST2GO)
27	EPM005LAAA11C000368	14 (0.29%)	2	NR	Q01752	3.00E-94	Aryl-alcohol dehydrogenase [NADP+] (AAD)	Aryl-Alcohol dehydrogenase with similarity to *P. chrysosporium* aryl-alcohol dehydrogenase (eggNOG)
28	EPM005LAAA11C000502	14 (0.29%)	3	NR	XP_567862.1	1.50E-124	ATP-dependent protein binding protein [Cryptococcus neoformans var. neoformans JEC21]	Polyubiquitin - Cryptococcus neoformans var. neoformans B-3501A (UniProtKB)
29	EPM005LAAA11C000320	13 (0.27%)	1	NR	BAE93903.1	1.80E-33	Response regulator-like protein [Neurospora crassa]	Transcription factor (eggNOG)

(continued)

Number	Representative cluster ID	# of ESTs[a]	# of cluster	Database	Accession ID[b]	E-value[b]	Description from NR or InterProScan best hit[b]	Description from eggNOG, BLAST2GO and UniProtKB
30	EPM005LAAA11C000352	13 (0.27%)	2	NR	EAU85842.1	2.60E-13	Predicted protein [Coprinopsis cinerea okayama7#130]	Predicted protein (BLAST2GO)
31	EPM005LAAA11C000015	13 (0.27%)	3	NR	EAU91766.1	8.80E-36	Hypothetical protein CC1G_04534 [Coprinopsis cinerea okayama7#130]	Utp--Glucose-1-Phosphate uridylyltransferase (eggNOG)
32	EPM005LAAA11C000348	13 (0.27%)	1	NR	EAU91946.1	1.20E-23	Hypothetical protein CC1G_11132 [Coprinopsis cinerea okayama7#130]	Protein involved in amino acid salvage (eggNOG)
33	EPM005LAAA11C000350	13 (0.27%)	1	NR	O42820	7.50E-132	Elongation factor 1-alpha (EF-1-alpha)	Elongation factor 1 alpha (eggNOG)
34	EPM005LAAA11C000328	13 (0.27%)	1	Unknown	-	-	-	-
35	EPM005LAAA11C000349	13 (0.27%)	1	Unknown	-	-	-	-
36	EPM005LAAA11C000465	12 (0.25%)	2	NR	EAU89333.1	7.80E-14	Predicted protein [Coprinopsis cinerea okayama7#130]	Predicted protein - Coprinopsis cinerea (strain Okayama-7 / 130 / FGSC 9003) (Inky cap fungus) (Hormographiella aspergillata) (UniProtKB)
37	EPM005LAAA11C000351	12 (0.25%)	1	Unknown	-	-	-	-
38	EPM005LAAA11C000477	12 (0.25%)	2	INTERPRO	SSF103473	2.40E-04	MFS general substrate transporter	-
39	EPM005LAAA11C000354	12 (0.25%)	1	Unknown	-	-	-	-
40	EPM005LAAA11C000357	11 (0.23%)	1	NR	CAN70790.1	3.70E-13	Hypothetical protein [Vitis vinifera]	Gamma 1 (BLAST2GO)
41	EPM005LAAA11C000401	11 (0.23%)	2	NR	EAU82684.1	4.50E-85	Hypothetical protein CC1G_08841 [Coprinopsis cinerea okayama7#130]	Stomatin family protein(BLAST2GO)
42	EPM005LAAA11C000362	11 (0.23%)	1	NR	EAU83138.1	4.70E-12	Hypothetical protein CC1G_07820 [Coprinopsis cinerea okayama7#130]	NADPH:quinone reductase and related Zn-dependent oxidoreductases (eggNOG)
43	EPM005LAAA11C000448	11 (0.23%)	2	NR	EAU90003.1	1.30E-66	Hypothetical protein CC1G_05919 [Coprinopsis cinerea okayama7#130]	Ubiquitin-Conjugating enzyme E2 (eggNOG)
44	EPM005LAAA11C000359	11 (0.23%)	2	NR	EAU91229.1	1.10E-30	Predicted protein [Coprinopsis cinerea okayama7#130]	Protein involved in regulation of transcription (eggNOG)
45	EPM005LAAA11C000361	11 (0.23%)	1	Unknown	-	-	-	-
46	EPM005LAAA11C000308	11 (0.23%)	2	Unknown	-	-	-	-
47	EPM005LAAA11C000356	11 (0.23%)	1	INTERPRO	SSF49764	2.20E-12	HSP20-like chaperones	-

[a], Number of sequences in cluster

[b], Best protein match by BLASTX in the non-redundant protein database at NCBI or by InterProScan.

Table 3. The 10 most frequently occurring Pfam domains in the *P. chrysosporium* uniESTs

Protein family	Pfam accession number	*P. chrysosporium* uniESTs		*P. chrysosporium* genome[a]	
		Rank	No. of uniESTs	Rank	No. of genes
WD40 repeat‡	PF00400	1	30	2	116
RNA recognition motif, RNP-1‡	PF00076	2	21	9	59
Serine/threonine protein kinase-related	PF00069	3	17	1	120
Ubiquitin‡	PF00240	4	15	38	12
Ankyrin‡	PF00023	5	12	28	22
Mitochondrial substrate carrier	PF00153	6	9	19	34
Heat shock protein Hsp20‡	PF00011	6	9	44	6
Heat shock protein 70‡	PF00012	7	8	43	7
Ras	PF00071	7	8	23	28
Zinc finger, C2H2-type	PF00096	7	8	12	52
Cyclin-like F-box	PF00646	7	8	4	100
Peptidase A1	PF00026	8	7	11	53
Fungal transcriptional regulatory protein, N-terminal	PF00172	8	7	18	35
RNA polymerase II, heptapeptide repeat, eukaryotic‡	PF05001	8	7	49	1
Ubiquitin-conjugating enzyme, E2	PF00179	9	6	30	20
Aldo/keto reductase	PF00248	9	6	11	53
Heat shock factor (HSF)-type, DNA-binding‡	PF00447	9	6	46	4
Pyridoxal phosphate-dependent enzyme, beta subunit	PF00291	9	6	40	10
Zinc finger, RING-type	PF00097	9	6	24	26
Heat shock protein DnaJ, N-terminal	PF00226	10	5	28	22
Histone core	PF00125	10	5	34	16
FAD dependent oxidoreductase‡	PF01266	10	5	46	4
UTP--glucose-1-phosphate uridylyltransferase‡	PF01704	10	5	46	4
Calcium-binding EF-hand	PF00036	10	5	26	24

[a], These data was extracted on the basis of previous report by Diego Martinez (Martinez et al., 2004).
‡, indicate that 9 genes were significantly expressed based on IDEG6 online software analysis.

Table 4. Putative carbohydrate- and lignin-active enzymes identified in *P. chrysosporium*

Enzyme	CAZy family	# of contigs	Contig ID
Esterase	CE1	1	EPM005LAAA11C000046
Carboxylesterase	CE1	1	EPM005LAAA11C000569
N-acetylglucosamine-6-phosphate deacetylase	CE9	1	EPM005LAAA11S000465
Exo-1,3-β-glucanase	GH5	2	EPM005LAAA11C000017, EPM005LAAA11S000585
Endo-1,4-β-xylanase	GH10	1	EPM005LAAA11C000837
α-amylase	GH13	1	EPM005LAAA11C000499
α-glucosidase	GH13	1	EPM005LAAA11C000427
β-1,6-glucan synthetase	GH16	2	EPM005LAAA11C000114, EPM005LAAA11S000101
Endo-1,3(4)-β-glucanase	GH16	2	EPM005LAAA11C000323
Chitinase	GH18	1	EPM005LAAA11C000066
Trehalase	GH37	1	EPM005LAAA11C000100
Endo-1,4-β-glucanase	GH61	1	EPM005LAAA11C000562
Endoglucanase	GH61	1	EPM005LAAA11C000879
Glucosidase	GH63	1	EPM005LAAA11S000301
Endo-(1,3)-α-glucanase	GH71	1	EPM005LAAA11S003097
1,3-β-glucanosyltransferase	GH72	2	EPM005LAAA11C000397, EPM005LAAA11S002847
Trehalose-phosphatase	GT20	1	EPM005LAAA11S004620
Trehalose synthase	GT20	1	EPM005LAAA11S001118
Trehalose 6-phosphateglycosyltransferase	GT20	1	EPM005LAAA11S004620
α-1,2-mannosyltransferase	GT4	1	EPM005LAAA11C000385
Glycogenin	GT8	1	EPM005LAAA11C000612
UDP-N-acetylglucosamine pyrophosphorylase	GT41	1	EPM005LAAA11C000746

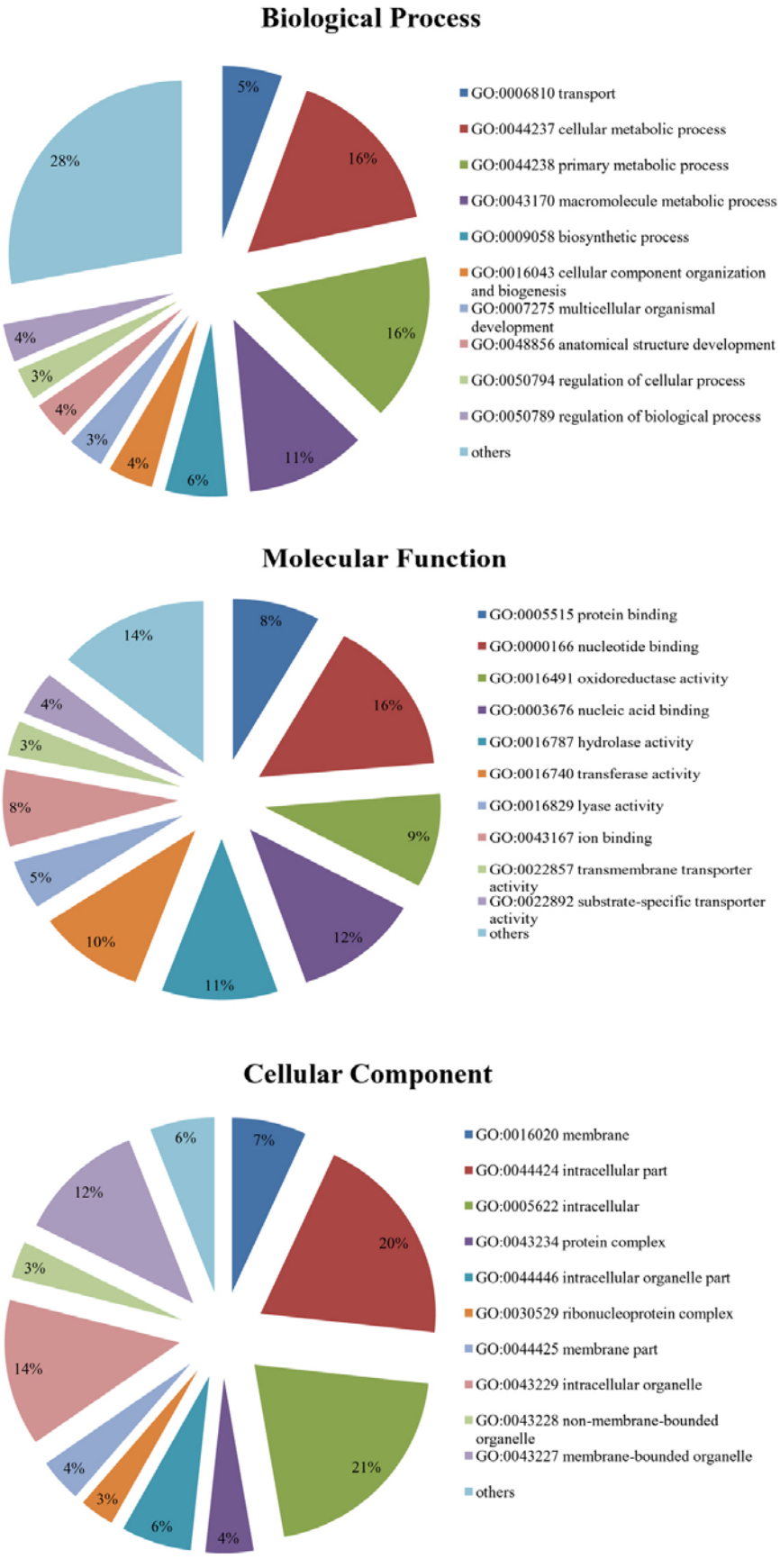

Fig. 2. Gene Ontology mappings for *P. chrysosporium* uniESTs by using BLAST2GO. The genes were functionally categorized according to the Gene Ontology Consortium and level three of the assignment results was charted here. 41% (718 of a total 1,751) uniESTs from *P. chrysosporium* were classified by GO. Note that since a gene product could be assigned to more than one GO term, the percentages in each main category could be added up over 100%.

Table 5. The biological process and molecular function GO terms with the highest 15 scores

	Level	GO ID	GO terms	Representation	% Representation of total	Score‡
Biological process	4	GO:0043581	Mycelium development	37	10.82%	37
	7	GO:0044408	Growth or development of symbiont on or near host surface	24	7.02%	24
	3	GO:0006950	Response to stress	34	9.94%	20
	3	GO:0006810	Transport	70	20.47%	18
	4	GO:0015031	Protein transport	22	6.43%	16
	7	GO:0006355	Regulation of transcription, DNA-dependent	20	5.85%	15
	5	GO:0006412	Translation	27	7.89%	13
	4	GO:0006118	Electron transport	13	3.80%	13
	7	GO:0006468	Protein amino acid phosphorylation	12	3.51%	12
	6	GO:0007264	Small GTPase mediated signal transduction	11	3.22%	9
	4	GO:0006811	Ion transport	19	5.56%	8
	7	GO:0008380	RNA splicing	12	3.51%	8
	7	GO:0006397	mRNA processing	12	3.51%	8
	3	GO:0044237	Cellular metabolic process	202	59.06%	7
	5	GO:0006350	Transcription	33	9.65%	7
Molecular function	7	GO:0005524	ATP binding	49	12.50%	49
	3	GO:0005515	Protein binding	46	11.73%	27
	3	GO:0000166	Nucleotide binding	89	22.70%	24
	4	GO:0003677	DNA binding	29	7.40%	20
	3	GO:0016491	Oxidoreductase activity	54	13.78%	19
	3	GO:0003676	Nucleic acid binding	67	17.09%	19
	3	GO:0016787	Hydrolase activity	61	15.56%	17
	7	GO:0005525	GTP binding	14	3.57%	14
	4	GO:0046872	Metal ion binding	42	10.71%	13
	4	GO:0003723	RNA binding	16	4.08%	13
	3	GO:0016740	Transferase activity	57	14.54%	12
	7	GO:0004674	Protein serine/threonine kinase activity	12	3.06%	10
	3	GO:0003735	Structural constituent of ribosome	10	2.55%	10
	3	GO:0003700	Transcription factor activity	8	2.04%	8
	6	GO:0005216	Ion channel activity	8	2.04%	7

Note that individual GO categories can have multiples mappings. The representation means that the number of uniESTs can be mapped to GO term. The representation percentage is based on the total number of GO mappings in each of the three major ontologies (biological process: 342, molecular function: 392).

‡, Score was calculated by BLAST2GO according to number of different sequences annotated at a child GO term and distance to node of the child GO term (Refer to BLAST2GO).

up-regulated in response to their substrates in the media (Vanden Wymelenberg et al., 2009).

Functional categorization of *P. chrysosporium* uniESTs

To determine whether there are distinctly functional groups among *P. chrysosporium* uniESTs, we made putative functional assignments to these uniESTs using bioinformatics information-processing pipeline which has been designed by ourselves (Fig. 1). In our pipeline, gene ontology (GO) assignment for each uniEST is based upon the most significant match results obtained from BLASTX search against NCBI non-redundant (NR) database using BLAST2GO (Conesa et al., 2005), pathway mapping using KEGG (Kyoto Encyclopedia of Genes and Genomes) (Ogata et al., 1999), analyses of the *P. chrysosporium* secretome using Orfpredictor (Min et al., 2005), SignalP (Bendtsen et al., 2004), TMHMM (Krogh et al., 2001) and PSORTII (Nakai and Horton, 1999). Results of these analyses are described in the following three sections.

Gene ontology

The most widely used method to predict gene families and functions among EST sequences is gene ontology (GO). GO provides a dynamic controlled vocabulary and hierarchy that consist of three major ontologies (biological processes, molecular functions and cellular components). To functionally categorize 1,751 uniESTs using BLAST homology searches, we used BLAST2GO program (Conesa et al., 2005). During functional categorization of these uniESTs, 718 (41%) of 1,751 uniESTs were assigned to three major ontologies, biological processes (n = 342), molecular functions (n = 392) and cellular component (n = 241) (Fig. 2).

Table 5 shows descriptions to functional categories, which have been classified during the assignments of GO two major ontologies (biological processes and molecular functions) to our *P. chrysosporium* uniESTs. In particular, uniESTs, which had been classified into GO categories of mycelium development (GO:0043581), Growth or development of symbiosis on or near host surface (GO:0044408) and Response to stress (GO:0006950) in biological process and ATP binding (GO:0005524), Protein binding (GO:0005515) and Nucleotide binding (GO:0000166) in molecular function were significantly sufficient among our *P. chrysosporium* uniESTs. Interestingly, these results were very consistent with the above-mentioned annotation data in Tables 2 and 3. A complete listing of GO mappings assigned for uniESTs is provided in Supplementary Table 2. Taken together, these results demonstrate that the high expression patterns of genes encoding heat shock proteins in mycelium development and Response to stress (Table 5) were

Table 6. The top 20 most abundant EC numbers in *P. chrysosporium* using UniProtKB

Number	EC	Domain	ESTs sequence count	Unique enzymes
1	EC 1.14.11.-	With 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors	27	1
2	EC 6.3.2.-	Acid--D-amino-acid ligases (peptide synthases)	17	4
3	EC 1.-.-.-	Oxidoreductases	16	7
4	EC 1.13.11.-	With incorporation of two atoms of oxygen	16	1
5	EC 6.3.2.19	Ubiquitin--protein ligase	15	4
6	EC 4.2.1.11	Phosphopyruvate hydratase	12	2
7	EC 2.7.1.-	Phosphotransferases with an alcohol group as acceptor	11	8
8	EC 3.1.1.3	Triacylglycerol lipase	11	2
9	EC 3.6.1.-	In phosphorous-containing anhydrides	10	10
10	EC 1.14.13.1	Salicylate 1-monooxygenase	10	2

due to the adaptation for surviving dramatically changing and fluctuating environmental conditions.

Pathway analysis using KEGG assignments

In order to figure out the phenotype of any living system, it is essential to investigate not only molecular functions of its genes, but also the specific metabolic pathway diversity of the organism of interest, ideally in comparison with other organisms. In our study, biochemical pathways were predicted by assigning Enzyme commission (EC) numbers from the results of UniProtKB database with an E-value cut-off 1e-10 to pathways. A total of 186 (10.62%) uniESTs were given assignments to unique 120 EC numbers. The top 10 (highly represented) EC numbers are shown in Table 6. Among the 186 identified proteins and the sequences mapped to KEGG pathways, Oxidoreductases (EC 1.14.11.-) involved in catalyzing the hydrolysis of the glycosidic linkage in order to generate two smaller sugars were also highly expressed (Table 6). In addition, unique enzymes, phosphorus-containing anhydrides (EC 3.6.1.-) and Phosphotransferases with an alcohol group as an acceptor (EC 2.7.1.-) domain were very abundant.

Especially, we have carefully considered the pathways catalysed by the 34 most highly represented carbohydrate metabolism-associated enzymes in our *P. chrysosporium* EST database using KEGG assignments (Supplementary Table 3). During the analysis, in our EST data we have identified the existence of the pathways mediated by plant cell wall-degrading enzymes, cellulase and chitinase, whose activities are very characteristic for the *P. chrysosporium*. In addition, many other pathways, including glycolysis/gluconeogenesis, galactose metabolism, starch and sucrose metabolism, amino sugar and nucleotide sugar metabolism, TCA cycle, pentose and glucuronate interconversions, fructose and mannose metabolism and pyruvate metabolism, all of which are closely associated with the carbohydrate metabolism, have been predicted during the analysis using KEGG assignments. These results strongly suggest that the energy-producing and carbohydrate metabolisms, including catalytic reactions for degradation of cellulose, starch, glycogen and chitin are being highly activated in the *P. chrysosporium*.

Secretome analysis

Extracellular enzyme system utilized by the fungus, especially, the *P. chrysosporium*, is generally believed to degrade all major components of plant cell walls (1990). To profile ligninolytic enzymes and to identify novel extracellular enzymes, we carried out signalling peptide predictions of more than 20 amino acids beginning from first methionine start codon within the N-terminal region of ORFs, which have more than one transmembrane domains. In the present data set (1,751 uniESTs), we identified 23 putatively secreted proteins, which are corresponding to putative carbohydrate- and lignin-active enzymes, such as "1,3-β-Glucanosyltransferase", "α-Amylase" "aspartic peptidase a1" and "Chitinase", recognized in the *P. chrysosporium* genome (Table 7) (Martinez et al., 2004). Of them, four unknown uniESTs were among the secretome subset. These unknown uniESTs may encode extracellular enzymes, which appeared to be unique to the *P. chrysosporium*.

Comparative analyses with EST data of *Neurospora crassa*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*

Comprehensive comparisons with expressed sequence tags (EST) data from other fungus *Neurospora crassa* and yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* may well yield important clues about their origins and evolutionary diversity compared with other organisms including the gain, loss and development of orthologous genes in different organisms. *P. chrysosporium* is a member of filamentous fungus among basidimocycetes derived from the dikarya clade, which includes the two phyla Ascomycota and Basidiomycota. In order to survey evolutionarily conserved genes from *P. chrysosporium*, we queried uniESTs against three EST databases entries of a filamentous fungi (*N. crassa*; 28,090 ESTs), two yeasts (*S. cerevisiae*; 34,915 ESTs and *S. pombe*; 8,123 ESTs), because these organisms represent the best characterized fungus and yeasts in many respects, particularly in terms of their genome sequences, biology, biochemistry and physiology, as well as representatives of meiotic spores called ascospores, which are divided into three monophyletic subphyla: Pezizomycotina (*N. crassa*), Saccharomycotina (*S. cerevisiae*) and Taphrinomycotina (*S. pombe*) in the fungi phylogenetic tree (James et al., 2006). Therefore, they provide a wide opportunity for studying the evolutionary divergence of individual genes and gene families.

Using their EST databases as databases for BLAST search against our *P. chrysosporium* ESTs makes it possible to do a relatively comprehensive transcriptome analysis. Figure 3 summarizes the distribution of *P. chrysosporium* ESTs by TBLASTX results (cut-off score of 50). During this comparison, sequence similarity searches of the 1,751 uniESTs resulted in 400 (22.84%) common homologues to these three organisms, 363 (20.73%) homologues to *N. crassa*, 310 (17.7%) homologues to *S. cerevisiae*, 158 (9.02%) homologues to *S. pombe*,

Table 7. Putative secretory proteins predicted by the ORFpredictor, SignalP, TMHMM and PSORT II programs

Cluster ID	# of ESTs	Accession ID	E-value	Gene description (NR top hit)	Description from eggNOG, BLAST2GO and UniProtKB
EPM005LAAA11C000406	8	EAU91186.1	3.50E-12	Hypothetical protein CC1G_06821 [Coprinopsis cinerea okayama7#130]	Lysophospholipase plb1 (BLAST2GO)
EPM005LAAA11C000397	8	XP_568714.1	1.10E-66	1,3-beta-glucanosyltransferase [Cryptococcus neoformans var. neoformans JEC21]	1,3-Beta-Glucanosyltransferase (eggNOG)
EPM005LAAA11C000480	6	EAU86123.1	8.30E-62	Hypothetical protein CC1G_07202 [Coprinopsis cinerea okayama7#130]	pa domain protein (BLAST2GO)
EPM005LAAA11C000472	6	AAT11911.1	1.50E-35	Immunomodulatory protein [Antrodia camphorata]	Putative uncharacterized protein – Botryotinia fuckeliana (strain B05.10) (Noble rot fungus) (Botrytis cinerea) (UniProtKB)
EPM005LAAA11C000499	5	XP_001271889.1	2.80E-73	Alpha-amylase, putative [Aspergillus clavatus NRRL 1]	Alpha-Amylase (eggNOG)
EPM005LAAA11C000562	4	EAU87051.1	2.00E-54	Hypothetical protein CC1G_12388 [Coprinopsis cinerea okayama7#130]	Glycoside hydrolase family 61 protein (BLAST2GO)
EPM005LAAA11C000370	4	EAU85842.1	6.50E-15	Predicted protein [Coprinopsis cinerea okayama7#130]	Predicted protein (BLAST2GO)
EPM005LAAA11C000039	4	EAU91335.1	5.60E-70	Hypothetical protein CC1G_07370 [Coprinopsis cinerea okayama7#130]	Palmitoyl-Protein thioesterase (eggNOG)
EPM005LAAA11C000009	4	XP_774481.1	2.70E-11	Hypothetical protein CNBG1270 [Cryptococcus neoformans var. neoformans B-3501A]	Putative uncharacterized protein - Cryptococcus neoformans (Filobasidiella neoformans) (UniProtKB)
EPM005LAAA11C000699	3	ZP_01464234.1	2.40E-30	Protein TOS1 [Stigmatella aurantiaca DW4/3-1]	Protein tos1 (BLAST2GO)
EPM005LAAA11C000663	3	EAU86015.1	1.20E-16	Predicted protein [Coprinopsis cinerea okayama7#130]	Aspartic peptidase a1 (BLAST2GO)
EPM005LAAA11C000066	3	EAU84319.1	1.20E-26	Hypothetical protein CC1G_01315 [Coprinopsis cinerea okayama7#130]	Chitinase (eggNOG)
EPM005LAAA11C000838	2	No hit	-	-	-
EPM005LAAA11C000815	2	EAU85245.1	3.50E-57	Predicted protein [Coprinopsis cinerea okayama7#130]	Predicted protein (BLAST2GO)
EPM005LAAA11C000210	2	XP_568191.1	8.30E-29	Hypothetical protein [Cryptococcus neoformans var. neoformans JEC21]	Predicted unsaturated glucuronyl hydrolase involved in regulation of bacterial surface properties, and related proteins (eggNOG)
EPM005LAAA11C000209	2	XP_567939.1	2.10E-40	Chaperone regulator [Cryptococcus neoformans var. neoformans JEC21]	Chaperone regulator (BLAST2GO)
EPM005LAAA11S004880	1	XP_391381.1	2.90E-26	Hypothetical protein FG11205.1 [Gibberella zeae PH-1]	Epl1 protein - Trichoderma atroviride (Hypocrea atroviridis) (UniProtKB)
EPM005LAAA11S003406	1	No hit	-	-	-
EPM005LAAA11S003184	1	XP_566828.1	6.50E-39	Hypothetical protein [Cryptococcus neoformans var. neoformans JEC21]	wsc domain protein (BLAST2GO)
EPM005LAAA11S002155	1	No hit	-	-	-
EPM005LAAA11S001892	1	XP_001266622.1	7.80E-14	Conserved hypothetical protein [Neosartorya fischeri NRRL 181]	Similarity to hypothetical protein dag11 - Agaricus bisporus precursor - Aspergillus niger (UniProtKB)
EPM005LAAA11S000108	1	No hit	-	-	-
EPM005LAAA11S000065	1	XP_569376.1	4.50E-12	Hypothetical protein [Cryptococcus neoformans var. neoformans JEC21]	Expansin family protein (BLAST2GO)

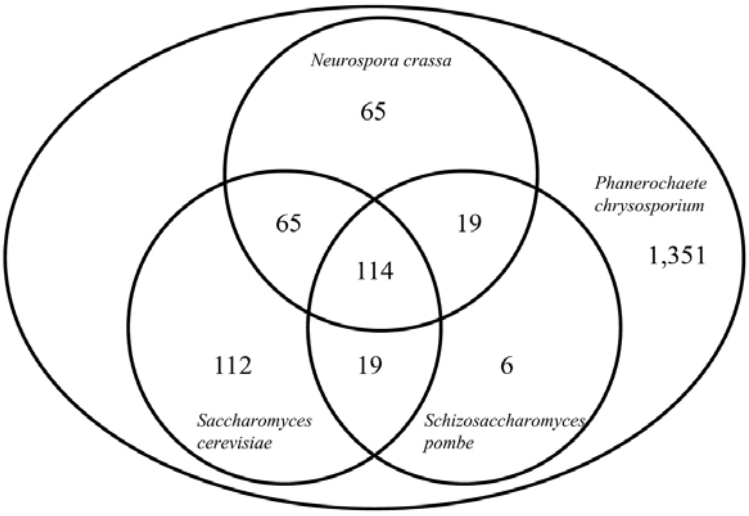

Fig. 3. A Venn diagram showing the distribution of *P. chrysosporium* uniEST TBLASTX matches by databases. The 1,751 translated *P. chrysosporium* non-redundant ESTs were used as queries in homology searches against *N. crassa*, *S. cerevisiae* and *S. pombe* EST databases, respectively. The three circles and the intersected areas among them contain the numbers of *P. chrysosporium* uniESTs that share TBLASTX similarity with *N. crassa*, *S. cerevisiae* and *S. pombe* ESTs.

and 1,351 (77.16%) had no significant similarity to EST sequences of these organisms. In our study, the Venn diagram (Fig. 3) shows that, based on the current database contents, the 114 uniESTs sequences from *P. chrysosporium* are equally conserved to these three organisms. Especially, these uniESTs evolutionarily well conserved to other three organisms revealed well-characterized housekeeping genes associated with various biological processes, including ribosomal protein, heat shock protein, ubiquitin and chromatin-related protein (see Supplementary Table 4).

Moreover, we found that 65, 112 and 6 uniESTs had similarity only with *N. crassa*, *S. cerevisiae* and *S. pombe*, respectively. These uniESTs may represent the conserved group during evolution after the divergence of Dikarya, or those that were lost from the genome contexts (see Supplementary Table 4). Among these, it is possible to identify candidates that might have contributed to the industrial and medical successes, including those involved in functions of enzymes such as 2-nitropropane dioxygenase (EPM005LAAA11C000012) (Ha et al., 2006), Alpha-1,2-Mannosyltransferase (EPM005LAAA11C000385) (Hausler et al., 1992), Glycerol-3-Phosphate dehydrogenase (EPM005LAAA11C000103) (DosSantos et al., 2003) and Homoserine O-acetyltransferase (EPM005LAAA11S004051) (Han et al., 2004). However, the fact that 1,351 uniESTs have no significant similarities to these three EST databases under our TBLASTX criteria (cut-off score 50), appeared to be due to relatively small data contents. *P. chrysosporium* is one of the most well studied white-rot fungi and its genome sequence already had been released as the first complete genome of a member of the Basidiomycota phylum. So far, the other fungal transcriptomes had been published from the Ascomycota phylum, such as *N. crassa*, *S. cerevisiae* and *S. pombe*. Taken together, this study offers deeper insights of gene evolution into the Basidomycota, which diverged from the Ascomycota 550 million years ago.

CONCLUSION

We sequenced 4,992 *P. chrysosporium* ESTs, putatively representing 1,751 uniESTs. Our data showed that 57.5% of the *P. chrysosporium* ESTs had homologs and orthologs in the well-studied model species and 47 genes, including genes encoding the most highly expressed heat shock proteins, were highly expressed. In addition, we profiled 26 putative carbohydrate- and lignin-active enzymes and 23 secreted proteins. By comparing *P. chrysosporium* ESTs with publicly available ESTs from the *N. crassa*, *S. cerevisiae* and *S. pombe*, we identified a set of highly conserved 114 genes. Of these, the most part was housekeeping genes for encoding heat shock protein, ribosomal proteins and transcription factors. In addition, we found lineage-specific ESTs overlapped with uniESTs from *P. chrysosporium*. These genes appeared to be good candidates for playing an important role in the evolution of the exceptional diversity after deriving from the dikarya clade. Taken together, our *P. chrysosporium* EST database provides novel molecular insight into the unique abilities and characteristics of this organism in degrading lignin and contaminated resources, and in surviving severe environmental changes and fluctuations, and will serve as a valuable resource for researchers to study evolutionary fungal genomics, particularly in the comparison with *P. chrysosporium* genome sequence and the genomes of its closest relatives.

Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).

ACKNOWLEDGMENTS

This study was supported by grant M10752000001-07N5200-00110 from the Ministry of Education, Science and Technology and grant KGM1230812 from the Korea Research Institute of Bioscience and Biotechnology.

REFERENCES

Abbas, A., Koc, H., Liu, F., and Tien, M. (2005). Fungal degradation of wood: initial proteomic analysis of extracellular proteins of Phanerochaete chrysosporium grown on oak substrate. Curr. Genet. *47*, 49-56.

Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. *340*, 783-795.

Boerjan, W., Ralph, J., and Baucher, M. (2003) Lignin biosynthesis. Annu. Rev. Plant Biol. *54*, 519-546.

Chabannes, M., Ruel, K., Yoshinaga, A., Chabbert, B., Jauneau, A., Joseleau, J.P., and Boudet, A.M. (2001). In situ analysis of lignins in transgenic tobacco reveals a differential impact of individual transformations on the spatial patterns of lignin deposition at the cellular and subcellular levels. Plant J. *28*, 271-282.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation,

visualization and analysis in functional genomics research. Bioinformatics *21*, 3674-3676.
Decelle, B., Tsang, A., and Storms, R.K. (2004). Cloning, functional expression and characterization of three Phanerochaete chrysosporium endo-1,4-beta-xylanases. Curr. Genet. *46*, 166-175.
DosSantos, R.A., Alfadda, A., Eto, K., Kadowaki, T., and Silva, J.E. (2003). Evidence for a compensated thermogenic defect in transgenic mice lacking the mitochondrial glycerol-3-phosphate dehydrogenase gene. Endocrinology *144*, 5469-5479.
Duranova, M., Spanikova, S., Wosten, H.A., Biely, P., and de Vries, R.P. (2009). Two glucuronoyl esterases of Phanerochaete chrysosporium. Arch. Microbiol. *191*, 133-140.
Eriksson, K.-E.L., Blanchette, R.A., and Ander, P. (1990). Microbial and Enzymatic Degradation of Wood and Wood Components (Berlin: Springer Verlag).
Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. *8*, 186-194.
Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. *8*, 175-185.
Gold, M.H., and Alic, M. (1993). Molecular biology of the lignin-degrading basidiomycete Phanerochaete chrysosporium. Microbiol. Rev. *57*, 605-622.
Gold, M.H., Youngs, H.L., and Gelpke, M.D. (2000). Manganese peroxidase. Met. Ions Biol. Syst. *37*, 559-586.
Ha, J.Y., Min, J.Y., Lee, S.K., Kim, H.S., Kim do, J., Kim, K.H., Lee, H.H., Kim, H.K., Yoon, H.J., and Suh, S.W. (2006). Crystal structure of 2-nitropropane dioxygenase complexed with FMN and substrate. Identification of the catalytic base. J. Biol. Chem. *281*, 18660-18667.
Han, Y.K., Lee, T., Han, K.H., Yun, S.H., and Lee, Y.W. (2004). Functional analysis of the homoserine O-acetyltransferase gene and its identification as a selectable marker in *Gibberella zeae*. Curr. Genet. *46*, 205-212.
Hausler, A., Ballou, L., Ballou, C.E., and Robbins, P.W. (1992). Yeast glycoprotein biosynthesis: MNT1 encodes an alpha-1,2-mannosyltransferase involved in O-glycosylation. Proc. Natl. Acad. Sci. USA *89*, 6846-6850.
Huang, X., and Madan, A. (1999). CAP3: A DNA sequence assembly program. Genome Res. *9*, 868-877.
Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., et al. (2009). InterPro: the integrative protein signature database. Nucleic Acids Res. D211-5.
Iimura, Y., and Tatsumi, K. (2002). Structure of genes for Hsp30 from the white-rot fungus Coriolus versicolor and the increase of their expression by heat shock and exposure to a hazardous chemical. Biosci. Biotechnol. Biochem. *66*, 1567-1570.
James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., et al. (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature *443*, 818-822.
Jensen, L.J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., and Bork, P. (2008). eggNOG: automated construction and annotation of orthologous groups of genes. Nucleic Acids Res. *36*, D250-254.
Kirk, T.K., and Farrell, R.L. (1987). Enzymatic "combustion": the microbial degradation of lignin. Annu. Rev. Microbiol. *41*, 465-505.
Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. *305*, 567-580.
Kumar, R., Singh, S., and Singh, O.V. (2008). Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. J. Ind. Microbiol. Biotechnol. *35*, 377-391.
Lee, Y.J., Galoforo, S.S., Berns, C.M., Chen, J.C., Davis, B.H., Sim, J.E., Corry, P.M., and Spitz, D.R. (1998). Glucose deprivation-induced cytotoxicity and alterations in mitogen-activated protein kinase activation are mediated by oxidative stress in multidrug-resistant human breast carcinoma cells. J. Biol. Chem. *273*, 5294-5299.
Martinez, D., Larrondo, L.F., Putnam, N., Gelpke, M.D., Huang, K., Chapman, J., Helfenbein, K.G., Ramaiya, P., Detter, J.C., Larimer, F., et al. (2004). Genome sequence of the lignocellulose degrading fungus Phanerochaete chrysosporium strain RP78. Nat. Biotechnol. *22*, 695-700.
Martinez, A.T., Speranza, M., Ruiz-Duenas, F.J., Ferreira, P., Camarero, S., Guillen, F., Martinez, M.J., Gutierrez, A., and del Rio, J.C. (2005). Biodegradation of lignocellulosics: microbial, chemical, and enzymatic aspects of the fungal attack of lignin. Int. Microbiol. *8*, 195-204.
Matsuzaki, F., Shimizu, M., and Wariishi, H. (2008). Proteomic and metabolomic analyses of the white-rot fungus Phanerochaete chrysosporium exposed to exogenous benzoic acid. J. Proteome Res. *7*, 2342-2350.
Min, X.J., Butler, G., Storms, R., and Tsang, A. (2005). OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res. *33*, W677-680.
Minami, M., Kureha, O., Mori, M., Kamitsuji, H., Suzuki, K., and Irie, T. (2007). Long serial analysis of gene expression for transcriptome profiling during the initiation of ligninolytic enzymes production in Phanerochaete chrysosporium. Appl. Microbiol. Biotechnol. *75*, 609-618.
Nakai, K., and Horton, P. (1999). PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem. Sci. *24*, 34-36.
Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. *27*, 29-34.
Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., et al. (2003). TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics *19*, 651-652.
Plesofsky-Vig, N., and Brambl, R. (1995). Disruption of the gene for hsp30, an alpha-crystallin-related heat shock protein of Neurospora crassa, causes defects in thermotolerance. Proc. Natl. Acad. Sci. USA *92*, 5032-5036.
Poggeler, S., and Kuck, U. (2004). A WD40 repeat protein regulates fungal cell differentiation and can be replaced functionally by the mammalian homologue striatin. Eukaryot. Cell *3*, 232-240.
Romualdi, C., Bortoluzzi, S., D'Alessi, F., and Danieli, G.A. (2003). IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments. Physiol. Genomics *12*, 159-162.
Santoro, M.G. (2000). Heat shock factors and the control of the stress response. Biochem. Pharmacol. *59*, 55-63.
Sato, S., Liu, F., Koc, H., and Tien, M. (2007). Expression analysis of extracellular proteins from *Phanerochaete chrysosporium* grown on different liquid and solid substrates. Microbiology *153*, 3023-3033.
The UniProt Consortium (2008). The universal protein resource (UniProt). Nucleic Acids Res. *36*, D190-195.
Vanden Wymelenberg, A., Gaskell, J., Mozuch, M., Kersten, P., Sabat, G., Martinez, D., and Cullen, D. (2009). Transcriptome and secretome analyses of Phanerochaete chrysosporium reveal complex patterns of gene expression. Appl. Environ. Microbiol. *75*, 4058-4068.
Vorob'eva, L.I. (2004). Stressors, stress reactions, and survival of bacteria (a review). Prikl. Biokhim. Mikrobiol. *40*, 261-269.
Wesenberg, D., Kyriakides, I., and Agathos, S.N. (2003). White-rot fungi and their enzymes for the treatment of industrial dye effluents. Biotechnol. Adv. *22*, 161-187.
Wymelenberg, A.V., Sabat, G., Martinez, D., Rajangam, A.S., Teeri, T.T., Gaskell, J., Kersten, P.J., and Cullen, D. (2005). The Phanerochaete chrysosporium secretome: database predictions and initial mass spectrometry peptide identifications in cellulose-grown medium. J. Biotechnol. *118*, 17-34.
Youn, H.S., Saitoh, S.I., Miyake, K., and Hwang, D.H. (2006). Inhibition of homodimerization of Toll-like receptor 4 by curcumin. Biochem. Pharmacol. *72*, 62-69.