UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

SEQUENCE AND ANALYSIS OF THE *PHYMATOTRICHOPSIS OMNIVORA* GENOME AND EXPRESSED SEQUENCE TAGS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

BY

SIMONE MACMIL
Norman, Oklahoma
2009

SEQUENCE AND ANALYSIS OF *PHYMATOTRICHOPSIS OMNIVORA* GENOME AND EXPRESSED SEQUENCE TAGS

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

By

_____
Dr. Bruce A. Roe, Chair

_____
Dr. Paul F. Cook

_____
Dr. Ann H. West

_____
Dr. Kathleen E. Duncan

_____
Dr. Charles V. Rice

Acknowledgements

I wish to express my sincere gratitude to my major advisor Dr. Bruce Roe for accepting me to his lab at the brink of his retirement, giving me the opportunity to work on this challenging yet rewarding project and for cheering me on through the many roadblocks and bumps that I encountered during the course of this study. His scientific accomplishments and discussions have always inspired me and have spurred me to reach higher in my goals and delve deeper in my understanding and it would be impossible to complete this project sans his patience, guidance and encouragement. I am extremely grateful to my advisory committee, Dr. Paul Cook, Dr. Kathleen Duncan, and Dr. Charles Rice for their advice, encouragement and participation in my research and dissertation. I would specially like to thank Dr. Ann West for chairing my committee during my candidacy exam and for her willingness to be present on my doctoral defense committee as well.

My special thanks also goes to Dr. Carolyn Young at the Noble Foundation for providing us with chromosomal bands, informative discussions and giving me the opportunity to learn at her lab. I am extremely grateful to Steve Kenton and Dr. Fares Najar for imparting to me some of their bioinformatic skills and for their assistance with the assembly and annotation. I would also like to extend my appreciation to all the members of Dr. Roe's lab; Graham Wiley, Doug White, Jim White, Yanbo Xing, Chunmei Qu, Ping Wang, Baifang Qin, Keqin Wang, Liping Zhou who worked together with me in processing pyrosequencing experiments on the 454 Genome Sequencer; Jiaxi Quan, Jing Yi, Ruihua Shi, Shaoping Lin, Fu

Ying, Axin Hua, Ziyun Yao, Limei Yang, Hongshing Lai, Dixie Wishnuck, Jennifer Lewis, Mounir Elharam and the many lab members who have now moved on for collectively making my experience at the lab memorable.

I wish to thank my brothers Llewellyn, Lyon, Warren, my friend Sini and my sister Sharon and her family for their love and support.

I dedicate this dissertation to my mum Patricia Rowe and dad Albert Macmil for having faith in my abilities and instilling in me a love for learning. Above all, I am eternally grateful to my heavenly Father, who daily loads me with His goodness, mercies and grace.

Table of Contents

List of Tables

List of Figures

Abstract

Phymatotrichopsis omnivora, a soilborne fungus confined to the southwestern regions of the United States, causes root rot in more than 2,000 species of dicotyledonous plants. To investigate the gene repertoire in this plant pathogenic fungus, a cDNA library containing expressed genes from the three distinct morphological stages in its life cycle and on exposure to three different nutrient conditions and a whole genome shotgun library was sequenced using massive parallel pyrosequencing and analyzed.

Unique expressed sequence tags (ESTs) obtained by sequencing the ends of the cDNA transcripts from each library were examined for homologs in GenBank, KEGG, KOG and COGEME using the Blast alignment program and categorized into groups based on biological function assignment. Normalization and comparison of the metabolic profile of the ESTs from each library revealed stage specific gene expression.

Analysis of the draft genomic sequence indicated that the ~74 Mbp assembly was approximately twice the size estimated by electrophoretic gel karyotyping from *P. omnivora* protoplasts supporting the hypothesis that this fungus is an obligate heterokaryon with several heterokaryotic nuclei. Approximately 22,000 genes were predicted with 9000 having homologs in GenBank and a further 12,000 sharing domains with proteins in the Pfam database. Annotation of the *P. omnivora* predicted proteins based on their biological function and metabolic pathways by comparison against the KOG, KEGG and COGEME databases revealed a comparable number of proteins involved in metabolism and cellular processes as

found in the well studied filamentous fungi *N. crassa* and *M. grisea*. Moreover, *P. omnivora* was found to encode slightly higher numbers of ABC-type transporters and calcium and heavy metal transporting P-type ATPases than *N. crassa* and *M. grisea.* These ATP-dependent proteins likely are involved in the survival of the pathogen in calcareous heavy metal containing soils and on exposure to plant toxins and fungicides. One such protein of particular interest is the homolog of Bcmfs1, a multidrug transporter involved in protection against natural toxins and fungicides. Since *P. omnivora* also is a filamentous fungus that lacks both functional conidia and an active sexual cycle, this is the first study to provide details of the life style and metabolic profile of fungi with a parasexual cycle.

Chapter1

Introduction

1.1 Fungi and their role in agriculture

Plant diseases have a significant impact on agriculture worldwide, affecting 25% of the yield in Western countries and almost 50% in developing countries (Agrios, 1997). Diseases caused by fungi in major crop plants (Strange and Scott, 2005) account for one third of all agricultural losses (Bowyer, 1999). The annual estimated damage caused by fungi in the United States exceeds $33 billion dollars resulting in more than $600 million dollars spent on fungicides (Madden and Wheelis, 2003). This high cost involved in treating infected soil, coupled with the need to avert chemical contamination of soil, makes it imperative to understand the biology of fungi as that information may provide the knowledge necessary to achieve targeted fungal control.

1.1.1 Introduction to Fungi

The Fungal Kingdom comprises a diverse group of eukaryotes consisting of more than 1.5 million members (Hawksworth, 1991) that have had a significant impact on humans since the dawn of civilization. Mushroom stones have been referenced in Greek literature far back as 1000-300 BC (Lowy, 1971), and the use of fungi in fermentation and brewing has been long practiced. Fungi that secrete enzymes such as cellulases, proteases, pectinases as well as secondary metabolites including for example antibiotics are of economic importance in the food and drug

industry. Most fungi are saprophytic, i.e. living on dead decaying organic matter, and play a vital role in the nutrient exchange cycle. Some fungi also share a symbiotic relationship with prokaryotes and eukaryotes i.e. plants and animals, while others can cause mycoses, plant diseases, and can also produce mycotoxins (Moss, 1987).

Fungi are heterotrophic organisms that occur in two morphological forms, either as free-living single cells, the most common example being the bakers and brewers yeast *Saccharomyces cerevisiae* or as multicellular filaments such as those observed in the rice blast fungus *Magnaporthe grisea* or the soil borne fungus *Phymatotrichopsis omnivora*, that causes cotton root rot and is the subject of this dissertation research.

Since fungi are well suited to grow in osmotrophic environments, they are the predominant biodegraders in all ecosystems (de Boer et al., 2005). A hybrid-type histidine kinase os-1/ Nik-1, that was shown to be involved in the osmosensitive signal transduction pathway in *Neurospora crassa,* also was found to be vital in adaptation to high osmolarity conditions (Alex et al., 1996). The *N. crassa* genome encodes twice as many genes as both *S. cerevisiae* and *S. pombe* with the gene complement displaying greater structural complexity, while the genomes of *N. crassa* and *M. grisea* genome possess 39 and 115 cytochrome P450 domain containing genes, a much greater amount than found in the other sequenced yeast genomes. Signaling pathways, including mitogen-activated protein kinases and cyclic AMP-dependant protein kinase, as well small GTPases of the Ras family and G-protein coupled receptors, also are predominantly present in several filamentous

fungi (Hynes, 2003). As shown in Table 1, several filamentous fungi that are found abundantly in the soil have been associated with plant diseases.

Table 1. Plant diseases caused by filamentous fungi.

Plant Disease	**Causative Fungal Agent**
Disease in rye and other grasses	*Claviceps purpurea*
Potato Blight	*Phytophthora infestans*
Rice Blast disease	*Magnaporthe grisea*
Chestnut Blight	*Cryptonectria parasitica*
Root rot in cotton and other dicots	*Phymatotrichopsis omnivora*
Dutch Elm disease	*Ophiostoma ulmi*

1.2 *Phymatotrichopsis omnivora* – the "cotton root rot fungus"

Phymatotrichopsis omnivora, a soil borne filamentous fungus, found predominantly in the southwestern United States and Mexico, is the causative agent of root rot by colonizing the tap root of over 2000 species of dicotyledonous plants (Damicone et al. 2003, Eaton and Rigler, 1946). The fungus, originally characterizd in cotton, either kills the plants before maturity, or arrests the growth of developing bolls in infected plants that survive until harvest. About 2% of the cotton yield in Texas each year is reportedly destroyed by *P. omnivora* (Watkins, 1981).

1.2.1 Life cycle of *Phymatotrichopsis omnivora*

P. omnivora exists in three morphological stages: vegetative, sclerotial and conidial. In the vegetative phase, it produces a network of root hair-like strands in soil until it comes in contact with descending roots. The strands encompass the root and grow toward the soil surface; and the fungus shows extensive cottony mycelial growth around the hypocotyl. The periderm then is destroyed and *P. omnivora* moves through the medullary rays until the vascular elements are occluded, blocking the flow of water and photosynthesis. The mycelial strands, as shown in Figure 1 where the life cycle of this fungus is depicted, are about 200 μm in diameter, composed of large central hypha entwined by many smaller hyphae, termed acicular hypha, that form distinctive cruciform branches emerging from the peripheral mycelium and are characteristic of this pathogen (Lyda, 1978).

In the vegetative stage, the fungus exists in the form of filamentous hyphae that consist of septate, 10 to 20 μm wide multinucleate cells that are perpendicular to each other, displaying characteristic cruciform branching (Hosford and Gries, 1966). It is in this stage that the fungus invades root of host plants causing plant death, following which the filamentous strands grow and enlarge followed with aggregation of the cells to form the sclerotial resting structures. Sclerotia have been recovered from as deep as 12 feet below the soil surface and have been tested to retain viability for at least five years (Streets and Bloss, 1973).

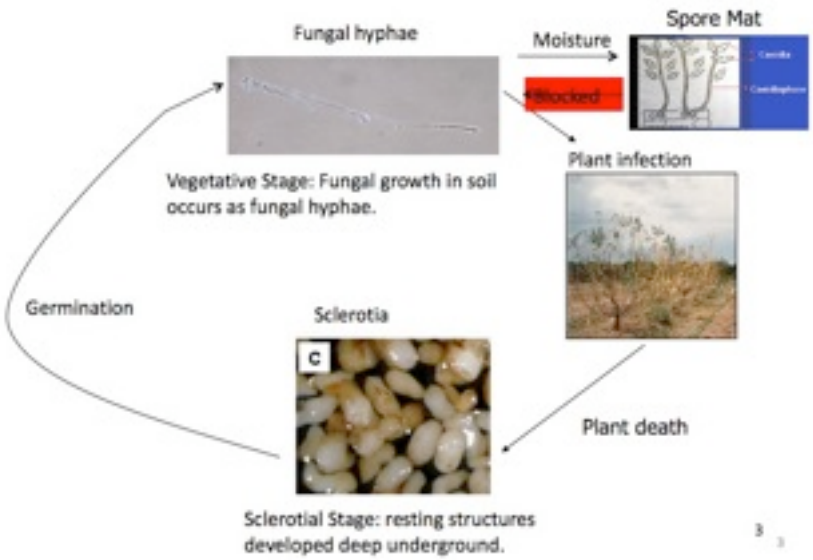

Figure 1. Life cycle of *P. ominvora* as it passes through the vegative, sclerotial and conidial stages.

Upon germination, the sclerotia act as primary inocula for vegetative growth and infection (Dunlap, 1941) and result in spore mats that form on the surface of mycelia infected soil following exposure to rain. These spore mats represent the conidial stage of this fungus and comprise spores formed at the tip of aerial spore bearing hyphae also known as conidiophore. *In vitro* germination of conidia has not been very successful, and although conidia subjected to drastic treatments such as sonication, resulted in only 60% germination (Kings et al., 1931), this is unlikely to occur in nature, hence the role of conidia in the fungal life cycle remains very unclear

Individual hyphal cells of *P. omnivora* have been reported to contain several nuclei, and this evidence coupled with the absence of a known sexual stage or functional asexual spore stage, has suggested the possibility of a parasexual cycle

within the fungus (Hosford and Gries, 1966). This ability of the fungus to maintain several slightly different copies of its genome may explain its survival and genetic flexibility as a pathogen.

1.2.2 Habitat of *Phymatotrichopsis omnivora*

P. omnivora affects plants growing in alkaline, calcareous soils under non-freezing conditions causing extensive crop damage resulting in large economic losses to the agricultural communities in Southern Oklahoma, Arizona, New Mexico, and thorough out most of Texas and northern regions of Mexico, as shown in Figure 2.

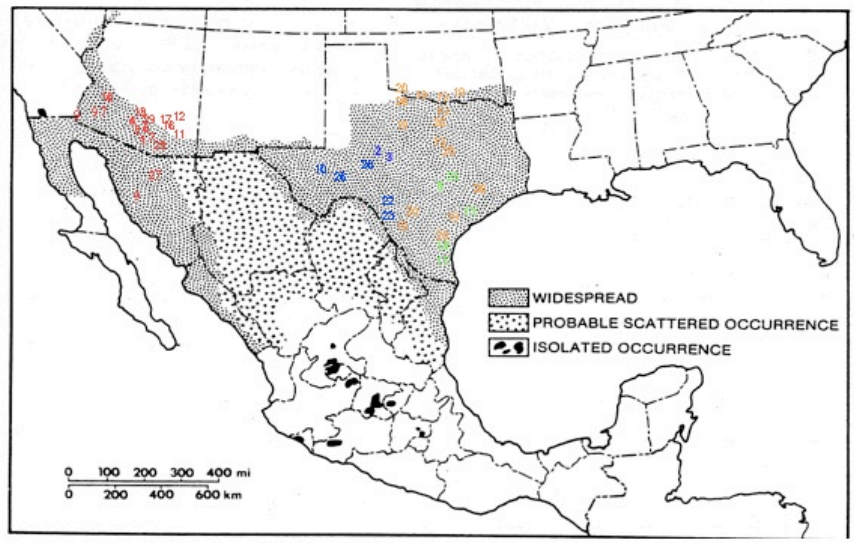

Figure 2. Reported distribution of *Phymatotrichopsis omnivora* in North America. Adapted from map by Streets and Bloss (1973).

The broad host range of this pathogen as well as its ability to thrive predominantly in the US and Mexico makes it also a potential agent for use in biological warfare it is listed as an A1 quarantine pest by the European and Mediterranean Plant Protection Organization (Damicone et al., 2003) and is resistant to several

fungicides.

Isolates obtained from roots of cotton, alfalfa, soybean and peach throughout the regions illustrated in Figure 2 have been classified in 29 haplotype groups based on the Maximum parsimony studies of their Internal Transcribed Spacer sequence (ITS). Of the 144 isolates studied, isolates from Texas were classified in 15 haplotype groups, moreover 13 haplotype groups were identified in Arizona, and 2 each from Mexico and Oklahoma (Marek et al., 2009). In an effort to understand the biology of this fungus in the wake of recent advances in fungal genomics and biology, I undertook whole genome shotgun and an expressed sequence tag (EST) sequencing project to obtain and characterize the resulting draft genomic sequence and expressed genes of *P. omnivora* OK-alf8.

1.3 Fungal Genomics

The availability of *S. cerevisiae* genome in 1996 (Goffeau et al. 1996) marked the advent of fungal genomics and laid the foundation for functional genomic studies in eukaryotes (Kastenmayer et al., 2006). A Fungal Genome Initiative was launched by the Broad Institute in Boston, MA to sequence genomes across the Fungal Kingdom to better understand eukaryotic biology, enhance comparative genomics studies, and aid in promoting genome-based evolutionary and pathogenicity studies (http://www.broad.mit.edu/). The Genomes OnLine Database ((http://www.genomesonline.org) reports that the sequene of 228 fungal genomes is in progress and that 24 have been sequenced and published to date, as shown in Table 2 below.

Table 2. List of selected completed and draft fungal genomes

Organism	Phylum	Size (Mb)	Number of chromosomes	Putative genes
Aspergillus fumigatus	Ascomycete	32	8	9,926
Ashbya gossypii	Ascomycete	9	7	4800
Aspergillus nidulans	Ascomycete	29	8	9,500
Aspergillus niger	Ascomycete	30	8	14,023
Aspergillus oryzae	Ascomycete	38	8	12,074
Candida albicans	Ascomycete	16	8	6500
Cryptococcus neoformans	Basidiomycete	21	14	6,967
Magnaporthe grisea	Ascomycete	40	7	11,074
Neurospora crassa	Ascomycete	40	7	10,082
Saccharomyces cerevisiae	Ascomycete	13	16	6500
Schizosaccharomyces pombe	Ascomycete	14	3	5000
Ustilago maydis	Basidiomycete	20	23	6,522

1.3.1 Fungal genomes used in this study

Saccharomyces cerevisiae has become one of the most important model organisms for studies of genetics and eukaryotic biology (Smutzer, 2001). As the

first eukaryote to have its genome completely sequenced (Goffeau et al. 1996), it also has become the model of choice for functional and comparative genomics. The Saccharomyces Genome Database (Hong et al., 2008) is regularly updated with information obtained from the functional studies from the 6,000 predicted open reading frames, protein-protein interaction studies providing a platform for comparative genomics.

Beadle and Tatum defined the role of genes in metabolism in the filamentous fungus *Neurospora crassa,* paving the way to the one-gene-one-enzyme hypothesis (Beadle and Tatum, 1941). Since substantial genetic, biochemical and molecular information exist from various isolates of the species, *Neurospora crassa* became the first model filamentous organism (Davis et al., 2002) and subsequently its genome was sequenced using a whole-genome shotgun strategy to obtain greater than 20-fold coverage. Although not completed to one gap-free sequence, the original assembly contained 38.6 Mb in 958 contigs encoding 10,082 genes including 424 tRNA genes (Borkovich et al., 2004) and the current assembly contains 39.23 Mb in 261 scaffolds on the Broad Institute web site at URL: http://www.broad.mit.edu/annotation/genome/neurospora/SingleGenomeIndex.html.

Magnaporthe grisea, the fungi that causes rice blast, resulting in significant economic problems worldwide, has been chosen as a model organism for studying fungal phytopathogenicity and host-parasite interactions. It is a haploid, filamentous fungus with a genome size of ~40 Mb contained in 7 chromosomes (Talbot et al. 1993; Orbach 1996) and is closely related to the non-pathogen *Neurospora crassa*, a leading model filamentous fungus for the study of eukaryotic genetics (Taylor et al.,

1993). A draft sequence of the *M. grisea* genome was completed by the Broad Institute, resulting in an assembled genome of 38.8 Mb in 2,273 contigs that encoded 11,074 proteins and 316 tRNAs (Dean et al., 2005). This assembly recently has been updated to 805 contigs in 30 scaffolds on the Broad Institute web site at URL: http://www.broad.mit.edu/annotation/genome/magnaporthe_grisea/AssemblyStats.html

Advances in high-throughput sequencing technology and WGS methods have aided the undertaking of fungal genomic sequencing on a large scale. However, the high numbers of repetitive sequence pose a major challenge to the genome assembly (Galagan et al., 2005) often making it extremely difficult to obtain a unique, contigous DNA sequence. These repetitive sequences also cause significant problems when attempting to assemble a diploid genome, such as that of *Candida albicans* (Jones et al., 2004; Braun et al., 2005). The variation in heterozygosity across chromosomal regions of diploid genomes results in an incorrect assembly of regions with low polymorphism causing highly polymorphic regions to remain separated, and has resulted in the developent of new assembly algorithms and an improved sequence assembly program (Vinson et al., 2005).

1.4 The discovery of nucleic acids and elucidation of the Central Dogma

The importance of DNA as the carrier of genetic information was established a century after its initial discovery. Frederich Miescher in 1869 isolated nuclei from pus cells and coined the high phosphorus organic content nuclein. Walther Fleming

in 1882, introduced the term chromatin suggesting its similarity to nuclein and Richard Altman in 1839 first isolated protein free nuclein, and was the first person to use the term nucleic acid (Dahm, 2005).

The validity of DNA as the genetic material was demonstrated in 1928 by Frederick Griffith by infecting mice with virulent type-R and avirulent type-S strains of *Streptococcus pneumoniae* (Griffith, 1928). Avery, MacLeod and McCarty in 1944 proved that DNA was the transferrable carrier of inheritable traits (Avery et al., 1944). That DNA and not protein was the genetic material was confirmed by Hershey and Chase in 1952 using radioactive $P^{32}$and S^{35} labeled media to grow and harvest bacteriphages (Hershey and Chase, 1952).

In 1860, Mendel studied the patterns of inheritance traits in pea lines and laid down the concept of hereditary units, Wilhelm Johannsen in 1909 termed these units "genes". At the turn of the last century, Morgan and his team studied the transfer of mutant traits in mutant fruit flies (*Drosophila Melanogaster)* leading to the identification of genes important in development and the idea of chromosomes. In 1913, Morgan and Sturtevant proposed the first genetic linkage map of *Drosophila melanogaster* displaying genes linearly arranged on chromosomes (Sturtevant, 1913). The next major breakthrough occurred when Beadle and Tatum proposed the one-gene-one-enzyme hypothesis based on their X-ray induced mutation studies in *Neurospora crassa* (Beadle and Tatum, 1941). The discovery that DNA was the hereditary material and its role in cell function paved the path to major milestones of the 20^{th} century. One such major landmark discovery of the structure (Figure 3) of DNA by Watson and Crick after their analysis of the X-ray

diffraction of salt crystallized DNA obtained by Rosalind Franklin and provided to them by Maurice Wilkins, led to the conclusion that the double-stranded DNA structure is composed of two antiparallel strands interwined in a double helix (Watson and Crick, 1953). These studies clearly demonstrated for the first time that each strand of double-stranded DNA is composed of nucleotide monomer units made up of phosphate groups and 2'-deoxyribose sugars joined by phosphodiester bonds and forms a backbone, where the four bases line the interior of the helix and the strands are base-paired with complementary bases from the opposite strand.

In 1970, the Central Dogma of Molecular Biology was proposed for understanding the transfer of genetic information (Crick, 1970). The dogma states that the general transfer of biological information occurs in three steps; from DNA to DNA (replication), DNA then is transcribed into mRNA (transcription) and proteins are synthesized from the information present in mRNA (translation). In special cases such as in retroviruses, genetic information in the form of RNA is reverse transcribed to DNA, RNA is copied to RNA (RNA replication in RNA viruses), and the translation of DNA directly to protein without an intermediate RNA has yet to be observed. More recently, the original Central Dogma has been modified to incorporate the evergrowing evidence of the importance of DNA base modifications, mRNA processing and protein post-translational modifications.

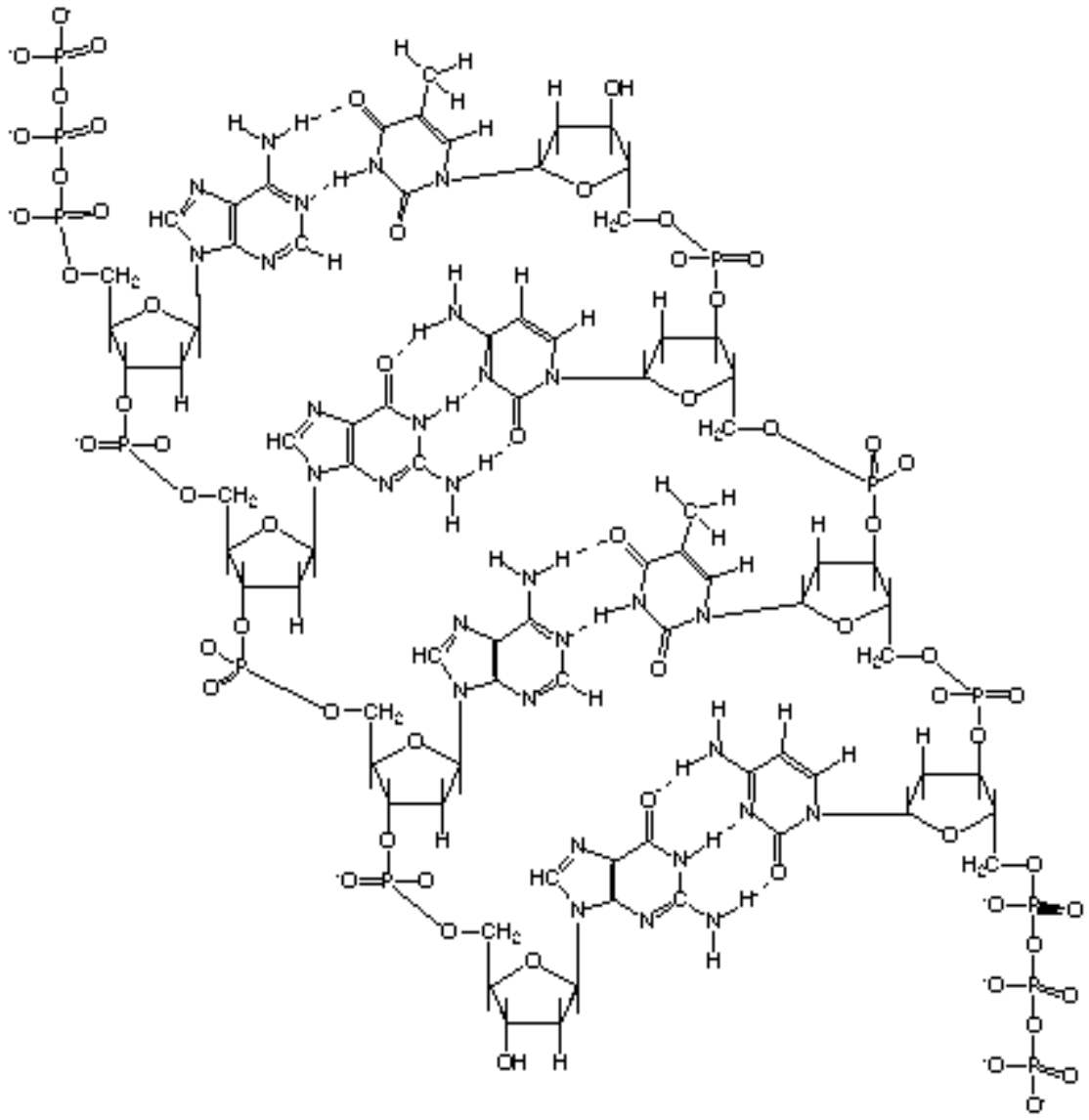

Figure 3: The structure of DNA (Watson and Crick, 1953)

1.5 Fungal Expressed Sequence Tags

An expressed sequence tag (EST) is a single pass DNA sequence obtained by sequencing one or both ends of a cDNA, e.g. a double-stranded copy of a mRNA, that represents a transcribed gene (Adams, et al., 1991). The information obtained from an EST sequence is important in mapping genes (Diener et al. 2004),

predicting novel genes (Zhu et al., 2001), predicting pathogenicity determinants (Soanes et al., 2006), improvement of functional gene assignments (Sims et al., 2004), identifying alternatively spliced transcripts (Ebbole et al., 2004) and determining patterns of genome evolution (Braun et al., 2000).

Since fungi are eukaryotes, their genes contain introns, that in contrast to other eukaryotes are short, and as is typical of eukaryotes, have both a 5'GU….. and a …..AG3' splice sites with a polypyrimidine tract in the region between the 5'end and the branch point. The primary transcript of fungi and other eukaryotic genes, a heteronuclear RNA, is post-transcriptionally modified by removal of the intron(s), polyadenylation, and the addition of a 5' cap sequence. A survey of large intron data sets among a diverse group of fungi indicated the potential of the fungal splicing mechanism to be slightly different from that of typical metazoans (Kupfer et al., 2004). Also, analysis of 15,435 unique ESTs obtained using conventional Sanger sequencing from the fungus *Aspergillus nidulans* during asexual development, revealed that only half the ESTs sharing homology with known genes in public databases, that the fungus accumulates stress response genes, and that a further 86% of fungal genes were associated with carbohydrate, amino acid, protein and peptide biosynthesis (Prade et al., 2000). In addition, EST sequences from several fungi also have been used to study fungal responses to host-related signals, one example of which is the demonstration that a set of functional group of genes more highly represented in phytopathogenic fungi than in non-pathogenic species, including gene products involved in lipid metabolism, ion-transporting ATPases, an alcohol dehydrogenase as well as genes of unknown function (Soanes and Talbot

2006).

Pyrosequencing using the 454 has been proved to be reliable in providing sequence depth and coverage, in a time, cost and labour effective manner when compared to classical Sanger dideoxynucleotide sequencing on capillary-based sequencers (Margulies et al., 2005; Wicker et al., 2006; Weber et al., 2007). Sequence reads obtained from a *Medicago truncatula* normalized cDNA library that compared the two approaches demonstrated that the 454 technology as effective as longer ESTs provided by conventional sequencing in mapping to a unique genome location, in the detecting new transcripts, and in improving gene prediction matrices (Cheung et al., 2006). In later studies, massively parallel pyrosequencing yielded an unbiased representation of nebulized 3'cDNA fragments from *Drosophila melanogaster* as 97% of the cDNA fragments were successfully mapped on to the *D.melanogaster* genome could be correlated with results from replicated microarray experiments (Torres et al., 2007). Also, similar 454-based pyrosequencing and the subsequent analysis of transcripts from *Arabidopsis thaliana* seedlings revealed deep coverage of the transcriptome with equal represenation of long, short and medium length ESTs significantly adding to the existing EST database and improving genome annotation (Weber et al., 2007).

In this present study, ESTs were sequenced from each stage in the *P. omnivora* life cycle in an attempt to understand the biochemical processes underlying phenotypic changes. ESTs also were sequenced from mRNA isolated from *P. omnivora* interacting with either host or non-host root exudates. This EST data then was used to determine the specific metabolic pathways and genes that are

expressed in the different stages of development and in response to different nutrient conditions.

1.6 DNA Sequencing Strategies

The quest to attain the $1000 human genome has revolutionized DNA sequencing and resulted in numerous sequencing strategies that include polony sequencing (Shendure et al., 2005), pyrosequencing using the luciferase-light detection system (Marguilles et al., 2005), sequence detection by incorporation of cleavable fluorescent nucleotide reversible terminators (Ju et al., 2006) and single-molecule sequencing (Harris et al., 2008, Korlach et al., 2008).

The polony sequencing method (Levene et al., 2003), that initially was used to detect SNPs in mRNA transcripts, has been further modified by Shendure and colleagues (Shendure et al., 2005) to accommodate bacterial genome resequencing. This method employs emulsion PCR based amplification of template DNA followed by immobilization of amplified and enriched beads on a 1.5 cm^2 slide layered with polyacrylamide gel. Sequencing is carried out by ligation of one of the uniquely fluorescently labeled degenerate nonamers to a hybridized anchor primer. Only the nonamer containing a base complementary to the query position is recognized and acted upon by the ligase. Subsequent epifluorescence imaging associates the fluorescence with the known sequence of the nonamer and identifies the base queried. Primer and nonamer complexes are stripped and new cycles are initiated.

In the sequencing strategy employing cleavable allyl carbamate linkers (Ju et al., 2006) with a unique fluorescent dye attached to the 5-position of dUTP and dCTP and 7-position of dGTP and dATP along with a cleavable *O*-allyl moiety to

cap the 3' hydroxyl group the template DNA is attached to a chip that contains a polyethylene glycol (PEG) linker. Following primer annealing, all four nucleotide analogues are added along with DNA polymerase in an extension reaction. Only one base is incorporated at each time, followed by fluorescence detection and base identity, the fluorophore and 3' cap is removed by a Pd- catalyzed deallylation reaction and subsequent washes. The cycle is reinitiated by addition of DNA polymerase and modified nucleotides.

The single-molecule sequencing approach was developed to avoid any errors that may be propagated in the amplified template DNA during PCR reactions. Here, the genomic DNA is fragmented, and a polyA tail is added to its 3'end, labeled and blocked. The template then is hybridized to a solid surface covalently bound with poly dT (50) oligonucleotide. The position of the template is detected by imaging of the labels, followed by addition of one of the fluorescently labeled nucleotides and DNA polymerase mixture. Following nucleotide incorporation, rinsing and imaging the fluorophore is chemically cleaved and the next nucleotide cycle is initiated (Harris *et al*., 2008). Similarly, although not yet commercialized, the use of optical nanostructures for detection of fluorescent base incorporation by immobilized DNA polymerase in high density arrays, are very promising as it has been reported to produce DNA sequences thousands of bases long (Korlach et al., 2008).

The pyrosequencing method described by Marguilles et al. (Marguilles et al., 2006), developed by 454 Life Sciences Inc., and now marketed by Roche Diagnostics, Inc., is based on measuring light intensity emitted by luciferase activity after incorporation of known nucleotide triphosphates using the high fidelity

Bacillus stearothermophilus DNA polymerase (Stenesh and Roe, 1972). In this method, randomly sheared genomic DNA is linked with adaptors and attached to beads, following which the sample is first amplified and enriched in a water-in–oil emulsion PCR reaction, followed by enrichment, primer annealing and deposition in picolitre sized wells of a fibre optic slide. Flow of each nucleotide across the slide results in incorporation of bases at the complementary position and release of pyrophosphate, the enzyme ATP sulfurylase then catalyzes the formation of ATP in the presence of APS (adenosine phosphosulfonate) and released pyrophosphate. The subsequent flow of substrate luciferin that then reacts with luciferase in the presence of ATP to produce oxyluciferin and light. The images from each flow are captured by a CCD camera and converted to flowgrams based on the light intensity produced from individual nucleotide addition. At the end of each cycle, the apyrase enzyme is flowed across the slide to remove excess ATP and APS and to avoid false base calls prior to initiating the next nucleotide flow cycle.

In contrast, during Sanger dideoxynucleotide sequencing, trace amounts of chain-terminating dideoxynucleotides are added to the 4 deoxynucleotide mixture to terminate DNA polymerase catalyzed DNA synthesis and obtain a DNA nested fragment set where each member differs the others by a single nucleotide and therefore allows base level resolution when resolved by electrophoresis (Sanger et al., 1977). The development of capillary array electrophoresis (Ueno and Yeung, 1994) and use of fluorescent dyes on primer or dideoxynucleotides, automated signal capture and improvement in equipment for data analysis (Smith et al., 1986, Hunkapiller et al., 1991) has made this method efficient in obtaining 500-1000 base

long reads and well suited to large-scale sequencing.

Since massively parallel pyrosequencing is an extremely useful and convenient method for determining large numbers of DNA sequences quite rapidly and inexpensively (Gharizadeh et al., 2006), we devised a strategy that combines 454 pyrosequencing and ABI 3730 Sanger sequencing to obtain the draft sequence of the *P. omnivora* genome. Also, during the course of this study we developed and implemented several improvements into the sequencing strategy that include using paired-end libraries to obtain paired-end sequence reads from the pyrosequencing method, manipulating pyrosequencing reagents to provide external reagent cooling coils, altering the machine run program script to generate longer read lengths, and optimizing the up-front chemistry for preparing samples prior to loading onto the pyrosequencer. These improvements have resulted in reducing the number of expensive ABI 3730 Sanger-based DNA sequence reads needed and thereby lowered the cost of obtaining the draft sequence the *P. omnivora* genome.

1.7 Bioinformatic Tools

1.7.1 BLAST

The Basic Local Alignment Search Tool (BLAST) suite of programs developed at the National Center for Biotechnology Information (NCBI) of the National Library of Medicine for finding ungapped, locally optimal sequences, is used to determine sequence similarity and to identify genes and genetic features by comparing a query sequence to a database of known DNA or protein sequences

(Altschul et al., 1990, 1997) such as GenBank, Kyoto Encyclopedia of Genes and Genomes (KEGG), Cluster of Orthologous Groups (COG) and Consortium for the Functional Genomics of Microbial Eukaryotes (COGEME). The BLAST family of programs is composed of a subset of five separate programs, blastn, blastx, blastp, tblastn, and tblastx, which are employed depending on the nature of the query sequence and a corresponding database. Blastn compares a nucleotide query sequence against a nucleotide database whereas tblastn compares a protein query against a nucleotide database dynamically translated in all six reading frames. The blastx program compares six-frame translations of a nucleotide query sequence against a protein sequence database whereas tblastx compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database and blastp compares an amino acid query sequence against a protein sequence database.

In blast searches, each High-scoring Segment Pair (HSP) consists of a segment from the query sequence and another segment from a database sequence of equal lengths, locally aligned with maximum alignment scores that are equal to or exceed a defined threshold. After the alignment of similar segments between the query and database sequences, the program evaluates the statistical significance of any matches that were found and reports the matches that comply with the user-defined threshold of significance. The parameter expect value or E value is defined as the maximum frequency at which a chance occurrence of an HSP (or set of HSPs) can be expected and can also be interpreted as the number of matches one expects to observe by chance when the query and database searches follow the random

sequence model (Karlin and Altschul, 1990). Although the default value for E typically is set to 10, it can be varied over a range of 0<E<=1000. In this present study, an E value of 0.00001 and 0.0000000001 was considered significant while searching for homology at the protein and nucleotide level respectively.

1.7.2 FgenesH

FgenesH, a based gene prediction program originally developed by V. Solovyev (Solovye and Salamov, 1997), and available from Softberry, Inc., (http://www.softberry.com), uses a hidden Markov model where, in contrast to the Markov model where the state is visible, the state is not visible is used to determine the known parameters influenced by one state or condition or another. Using this model, genes can be predicted based on information obtained from splice sites and information obtained from fully sequenced genes. The parameters can be altered to accommodate information obtained from genes sequenced from different organisms and typically the algorithm is built from information obtained from model organisms.

1.7.3 tRNA scan-SE

The tRNAscan-SE program detects 99-100% of transfer RNA genes in DNA sequences with high accuracy of only a single false detection per fifteen gigabases (Lowe and Eddy, 1997; http://lowelab.ucsc.edu/tRNAscan-SE/). It uses information obtained from the primary and secondary structures of functional RNAs, eukaryotic RNA polymerase III promoters and terminators to predict prokaryotic and eukaryotic tRNAs species including selenocysteine tRNA genes as well as pseudo

genes.

1.7.4 Databases

The Genbank database (Benson et al., 1996; http://www.ncbi.nlm.nih.gov) was established by the National Center for Biotechnology Information (NCBI) division of the National Library of Medicine (NLM) at the National Institute of Health (NIH) as a repository of all publicly available nucleotide and protein sequences along with their corresponding biological and bibliographic information. As of February 2009, 118 billion bases of Whole Genome Shotgun (WGS) sequence data has been recorded in addition to 101,467,270,308 bases for 101,815,678 sequences in the traditional Genbank divisions (ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt; Benson et al., 2009). This database contains information obtained directly from authors, daily exchange with international nucleotide sequence databases and data obtained by scanning research articles. DNA and protein sequence present in Genbank can be accessed using Entrez, which is NCBI's search and retrieval system.

The Consortium for the Functional Genomics of Microbial Eukaryotes or COGEME database currently contains 68,986 unique EST sequences obtained from eighteen species of plant pathogenic fungi, two species of phytopathogenic oomycetes and three species of saprophytic fungi (Soanes and Talbot, 2006; http://cogeme.ex.ac.uk/). Comparative analysis studies of unique EST sequences from 15 species of phytopathogenic fungi and three species of saprophytic fungi deposited in this database identified nineteen pathogen-specific genes (Soanes and Talbot, 2006; http://cogeme.ex.ac.uk/cgi-bin/path.pl) as well as the categorization of

fungal ESTs in fifteen functional classification groups that include disease and virulence genes, genes involved in metabolism, transcription, protein synthesis, cell division and growth are some of the other functional categories (Soanes et al., 2002).

The Kyoto Encyclopedia of Genes and Genomes or KEGG database is designed to facilitate the understanding of metabolic networks based on both *in silica* annotation of genes obtained from genomic sequences as well as information gleaned from functional genomic studies (Kanehisa et al., 2004; Kanehisa et al., 2006; http://www.genome.jp/kegg/). As of February 2009, the database contains 4,216,445 genes obtained from completed and draft genomes of 92 eukaryotes (12 filamentous fungi and ten unicellular fungal genomes), 846 prokaryotes and 67 plants and animal EST datasets. The most outstanding feature of this database is the generation of pathway maps organized in hierarchies, the first level of hierarchy comprises of five groups namely metabolism, genetic information processing, environmental information processing, cellular processes and human diseases. The KEGG Automatic Annotation Server (KAAS) generates KEGG pathway maps and KEGG orthology (KO) assignments by BLAST comparisons of predicted genes from completed and partial genomes against the manually curated KEGG Genes dataset (Moriya et al., 2007).

The Cluster of Orthologous Groups or COGs database (Tatusov et al., 2000; http://www.ncbi.nlm.nih.gov/COG) as of February 2009 contains proteins from 66 completed prokaryotic and eukaryotic genome sequences that have been classified into phylogenetic groups based on evolutionary descent. Currently, the database has

been modified to contain the eukaryotic orthologous groups (KOG) that is composed of clusters of predicted orthologs for seven completely sequenced and annotated eukaryotic genomes which includes the two yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. The present KOG contains 4,852 clusters of orthologs, which include 59,938 proteins (Tatusov et al., 2003). The proteins have been grouped to seventeen functional categories that fall under one of the five broad classes namely –information storage and processing, cellular processes and signaling, metabolism, poorly characterized and proteins with no functional similarities to KOG.

The Pfam database consists of protein domains and families sorted in multiple sequence alignments and by profile hidden Markov models (HMM) (Finn et al., 2008). As of March 2009, the most recent release contains 10,340 protein families based on metagenomic projects, the GenPept and UniProtKB sequence databases (http://pfam.sanger.ac.uk/). Based on the assumption that sequences with known protein structures have clear domain organization, non-redundant protein datasets deposited in the Protein Data Bank (PDB) are added to Pfam domains. In addition, Pfam classifies homologous proteins in clans based on related structure, function, significant matches of the sequences to HMMs from different families and protein-protein comparisons to allow better prediction of functions and structure for families of unknown functions.

Chapter2

Material and Methods

2.1 Construction of Shotgun Library for Sanger Sequencing and Paired-end Pyrosequencing

P. omnivora strain OKAlf 8 genomic DNA was extracted using the DNeasy Plant Maxi Prep kit (Qiagen) from hyphal strands of the fungus was provided by Dr. Stephen Marek at Oklahoma State University, Stillwater, OK. A whole genome shotgun strategy using both plasmid-based Sanger dideoxynucleotide dye terminator sequencing on the ABI 3730 and massively parallel pyrosequencing on the 454/Roche GS-20 was employed to obtain the draft sequence of the genome. In the Sanger dideoxynucleotide dye terminator method, a small-insert library was created and end-sequenced to obtain the genomic draft sequence while in the pyrosequencing method, a similar small insert library also was obtained by randomly shearing the genomic DNA to 2-4 kb sized fragments. However, after the DNA fragments were end-repaired and size selected, in the Sanger method the purified DNA then was cloned into pUC 18 vector and transformed into *E.coli* XL1-Blue MRF`. The sub-cloned DNA was isolated and purified followed by cycle sequencing using forward and reverse universal primers and fluorescently–labeled Taq ddNTP and resolving nested fragment set on ABI 3730 sequencers, converting to base calls with Phred and assembling with Phrap. In the pyrosequencing method, the sheared DNA was circularized followed by nebulization to create 300-500 bp fragments that are adaptor ligated and amplified in an emulsion PCR that covalently

attaches the DNA to small (~30 microns) plastic beads, with the subsequent sequencing of each base at a time by monitoring the release of pyrophosphate.

2.1.1 Fragmentation of Genomic DNA for both Sanger and Paired-end Pyrosequencing

Random 2-4 kb fragments of DNA suitable for cloning in pUC vectors or for use in paired-end pyrosequencing, as described below, can be obtained by several methods that include partial restriction enzyme digestion (Fitzgerald et al., 1992), sonication (Deininger, 1983), transposon insertion (Phadnis et al., 1989), nebulization (Bodenteich et al., 1994) and hydroshearing (Oefner et al., 1996). Since the Hydroshear method provided sharp distinct bands, it was used according to manufacturer's instructions, albeit with prior cooling at 4°C to generate a more random library (Roe 2004). Specifically, 100 µl of *P. omnivora* DNA (5-15 µg) was randomly sheared in a Gene Machine's Nebulizer followed by precipitation with 2.5 x volumes of ethanol-acetate (95% ethanol and 0.12 M sodium acetate) and a wash with 2.5 x volumes of 75% ethanol, vacuum dried and dissolved in 27 µl ddH_2O.

2.2 Sanger Sequencing

2.2.1 End Repair and Size Selection for Sanger sequencing

Ends with 5' and 3 ' overhangs and lacking phosphate groups at their termini often occur as a result of fragmentation and need to be made blunt ended prior to cloning. Hence, the sheared DNA was end repaired using Klenow DNA polymerase's 5-3' polymerization and T4 DNA polymerase's 3'5' exonuclease activity and the

resulting blunt ends were phosphorylated by T4 polynucleotide kinase by incubating 27 µl sheared DNA, 5 µl 10X kinase buffer (50 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 1 mM DTT, and 5 ug/ml BSA in double distilled water), 5 µl of 10 mM rATP, 7 µl of 0.25 mM dNTPs, 1 µl of 3 U/µl T4 DNA polymerase (New England Biolabs cat. # 302L), 2 µl of 5 U/µl Klenow DNA polymerase (New England Biolabs cat. # 210L) and T4 polynucleotide kinase (US Biochemicals cat # 70031) at 37°C for 30 minutes. The reaction mix then was loaded on to a 1% low-melt agarose gel alongside molecular size markers (Hind III-digested λ-DNA and HaeI-digested ϕX174-DNA, New England Biolabs, cat. # N3012L and N3026S). After electrophoresis at 120 mA for 1.5 hours, the 2-4 kb sized fragments were excised from the gel and frozen at -80°C. Following thawing at room temperature for 5 minutes and centrifugation in a table-top centrifuge at 13,000 rpm for 15 minutes, the supernatant containing DNA was transferred to a new tube and centrifugation was repeated with pooling of supernatants. The resulting DNA fragments were precipitated with 2.5 x volumes of ethanol-acetate (95% ethanol and 0.12 M sodium acetate) washed with 2.5 x volumes of 75% ethanol, vacuum dried, dissolved in 15 µl ddH_2O and stored frozed at -20°C.

2.2.2 Ligation of DNA fragements with pUC 18

The sheared and end-repaired fragments were ligated into Sma I cleaved calf intestine alkaline phosphatase-dephosphorylated pUC-18 vector (Pharmacia 27-4860-01) using T4 DNA ligase, an enzyme that catalyzes the formation of a phosphodiester bond between the 5' phosphate of DNA fragment and the 3' hydroxyl end of the vector (Pheiffer and Zimmerman, 1983). The ligation solution

contained 2 µl (~20 ng) of pUC-18 vector, 1 µl of 10x ligase buffer, 4 µl of sheared and repaired DNA, 2 µl ddH_2O and 1 µl of 400 U/µl T4 DNA ligase (New England Biolabs 202L).

2.2.3 Transformation

Recombinant pUC-18 insert carrying clones were transformed into electrocompetent *E.coli* XL1-Blue MRF' cells by electroporation. After mixing 2 µl of the ligation solution with 40 µl of electrocompetent cells in a cuvette on ice, the solution was placed in the electroporation chamber and an electric pulse of 2.5 kV was applied for 5 microseconds at 4°C. Immediately following electroporation 1 ml of cold YENB medium was added to the cuvette containing electroporated and transformed cells. The cell suspension then was transferred to a falcon tube (10 ml capacity) and allowed to grow at 37°C for 30 minutes with shaking at 250 rpm. Transformed cells were harvested by centrifugation at 2,000 rpm for 5 minutes, following which the cell pellet was resuspended in 30 µl of 25 mg/ml of 5-bromo-4-chloro-3-indoyl-D-galactoside (X-gal) and 30 µl of 25 mg/ml isopropyl-β-thiogalactopyranoside (IPTG). The cells were surface spread on ampicillin (100 µg/ml) containing LB agar plates and incubated at 37°C for 18 hours.

Both blue and white colonies were observed after incubation because during ligation, small oligonucleotides, with a length divisible by 3, allow translation of the *lac Z* gene region. Induction of the lac operon by IPTG then results in the transcription of the *lac Z* gene to produce functional β-galactosidase that in turn breaks X-gal down to form a blue coloured metabolite. Ligation of inserts whose length is not divisible by 3 into vector however results in inactivation of the N-

terminal fragment of β-galactosidase and hence recombinant vector carrying transformed cells lack the ability to cleave X-gal resulting in formation of white colonies. Insert containing white colonies were selected and inoculated into 384-well flat bottom microtiter plates containing ampicillin (100 µg/ml) supplemented TB media (8 µl of 10x TB salts made by adding 1.2 g potassium phosphate (monobasic) and 8.2 g potassium phosphate (dibasic) to 50 ml ddH_2O and 72 µl of TB media made by adding 6 g bacto-tryptone, 12 g bacto-yeast extract and 2 ml glycerol to 450 ml ddH_2O.) using Flexys colony picker. These plates were incubated in the HiGro incubator at 37°C with shaking at 350 rpm. Following 3.5 hours of shaking, an oxygen flow was initiated at 0.5 minutes intervals for 0.5 seconds. The cells were harvested by centrifugation at 3,000 rpm for 10 minutes following decantation of the supernatant and frozen overnight at -80°C.

2.2.4 Automatic Isolation of Subclone DNA

The sub-clone DNA was isolated using an automated single acetate cleared lysis method (Birnboim and Doly, 1979). This method entails cell lysis using SDS at a high pH to dissolve phospholipid and protein components of the *E. coli* cell membrane followed by treatment with RNase A and RNase T1 and potassium acetate (KOAc) to remove RNA and SDS/lipid/protein complexes. Treatment with potassium acetate also reduces the pH to neutral at which chromosomal DNA is trapped in the SDS/lipid/protein precipitate. Plasmid DNA in the solution is precipitated by isopropanol, washed with ethanol and disolved in sterile-distilled deionized water.

The automated procedure for isolation of subclone DNA involves the use of

the ZyMark robotic arm to transfer the cell pellets containing 384-well flat-bottom microtiter plates to the bed of the SciClone robot. Following which, cells in each well were suspended in 23µl of TE-RNase solution (50 mM Tris-HCl, pH 7.6, 0.5 M EDTA, 40 µg/ml RNase A (Sigma R-5500), and 0.04 U/µl RNase T1 (Sigma R-8251)). After 10 minutes of shaking at 1,000 rpm, 23 µl of lysis buffer (1% SDS and 0.2 M NaOH) was added to each well and the plates were subjected to shaking for another 10 minutes at 1000 rpm. Then 23 µl of 3 M KOAc (pH 4.5) was added and the plates followed by shaking for 10 minutes at 1,000 rpm and frozen at –80°C overnight. On the following day, plates were thawed and centrifuged at 3,000 rpm for 45 minutes in a Beckman C56R centrifuge. Using the Velocity 11 V-prep robot, 50µl of the resulting supernatant was transferred to a new 384-well plate, and DNA was precipitated by the addition of 50 µl of 100% isopropanol with mixing. After centrifugation at 3,000 rpm for 30 minutes in the Beckman C56R centrifuge, the obtained DNA pellet was washed with 50 µl of 70% ethanol. Following which, the DNA templates were dried in a vacuum dryer for 10 minutes and dissolved in 20 µl of molecular biology grade water. An aliquot then was evaluated by electrophoresis in a 1% agarose gel.

2.2.5 Reaction and Clean Up

The DNA templates were sequenced using the cycle sequencing method (Mardis and Roe, 1989; Chissoe, et al., 1991) in which the sequencing reaction is incubated for several cycles, where each cycle consisted of denaturation of double-stranded DNA (95°C), primer annealing (50°C) and elongation (60°C).

The sequencing reaction mix was prepared by combining 150-200 ng of sub-

clone DNA, 2µl of 6.5 µM universal forward or universal reverse primer and 2 µl of the 20× diluted ET reaction kit containing Ampli*Taq* FS, dATP, dCTP, dTTP (100 µM each), dITP (500 µM), ddATP, ddCTP, ddTTP, and ddGTP (~0.11 µM each) and then incubated for 60 cycles of 95°C for 30 seconds, 50°C for 20 seconds, and 60°C for 4 minutes in an Perkin-Elmer thermocycler. Once the cycling reaction was complete, the unincorporated terminators were removed from the sequencing reactions by ethanol-acetate (95% ethanol and 0.12 M sodium acetate) precipitation, followed by a 70% ethanol rinse. The plates then were dried for 10 minutes at room temperature and stored at –20°C until ready for loading onto the sequencer.

2.2.6 Sequencing on the ABI 3730

Sequencing reaction products were dissolved in 20 µl of 0.1 mM EDTA, and loaded on the ABI 3730 DNA sequencer. After 2.5 hours of electrophoresis at 6.5 kV, DNA sequencing data was collected automatically and analyzed using the ABI base caller on the attached computer. The resulting trace files then were transferred to a Unix-based SUN work station for further base calling by Phred and assembled by Phrap that could be viewed by Consed (Gordon *et al*., 1998).

2.3 Pyrosequencing on the 454

The massively parallel pyrosequencer from 454/Roche Life Sciences has the ability to generate millions of bases per hour in a single instrument run (Wicker et al., 2006). The system captures the real-time release of pyrophosphate produced by DNA chain elongation catalyzed by DNA polymerase (Ronaghi et al., 1996). Sulfurylase enzyme present in the sequencing reaction utilizes the released

pyrophosphate to convert APS to ATP which in turn is converted to light by the activity of luciferase on it's substrate luciferin.

The advantage of sequencing genomes using the 454 pyrosequencing technology alleviates the need for cloning or colony picking as DNA sequences from some genomes are known to be detrimental to insert carrying bacterial colonies, and therefore are excluded by traditional Sanger sequencing methodology.

The original 454 sequencing protocol involves unidirectional sequencing from libraries generated by random fragmentation and hence was used primarily for construction of EST libraries. However, since generation of mate pairs is known to alleviate the perplexities involved with assembly of complex genomes and aids in ordering and orienting contigs, the initial shotgun protocol (Marguiles et al., 2005 and Ng et al., 2006) has been modified in our lab to yield a hybrid shotgun and paired-end library (Wiley et al., 2007) and was used to sequence the *P. omnivora* genome. In this process, the DNA first is sheared to 2-4 kb fragments, circularized by reconstitution of a common central adaptor followed by fragmentation to 500 bp pieces that then are processed using a shotgun library protocol modified from the manufacturers recommendation as described below and shown in Figure 4.

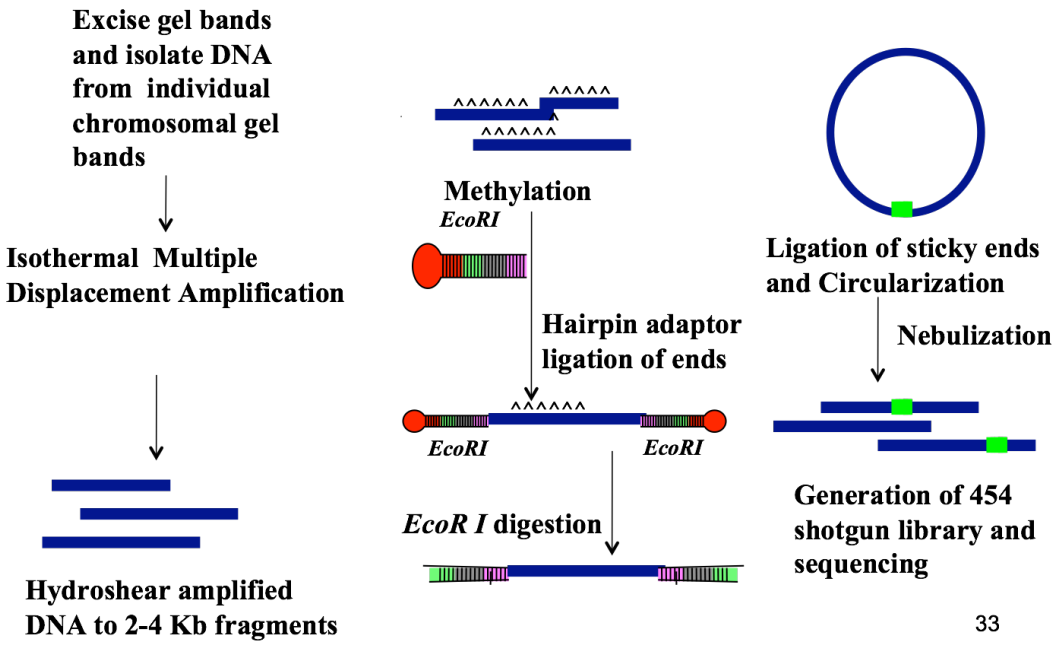

Figure 4. Schematic for generating a mixed paired-end library of gel amplified chromosomes

2.3.1 DNA fragmentation, SPRI bead purification and size selection

As described above with the Sanger dideoxynucleotide sequencing method, the genomic DNA first was fragmented to 2-4 kb pieces in the hydroshear, but instead of agarose gel electrophoresis to purify these fragments, the smaller fragments were removed by treating with Ampure SPRI magnetic beads (Agencourt, Inc. # 000132) that bind double-stranded DNA depending on the ionic strength (Hawkins, et al., 1994). At a SPRI-to-sample ratio (v/v) of 0.7, DNA fragments greater than 300 bp were preferentially bound resulting in elimination of shorter fragments. Subsequent washing with 75% ethanol removed unbound small fragments of DNA, and the DNA was eluted from beads in a molecular biology grade water or low salt containing buffer. Here, the volume of hydrosheared DNA

retrieved was measured and 100 µl of Ampure SPRI beads were added to each 200 µl of the DNA solution. Following mixing and incubating at 25°C for 5 minutes, the mixture was attracted to the wall of a Magnetic Particle collector (MPC). After discarding the supernatant, the beads were washed twice with 75% ethanol, and vacuum dried for 10 minutes. Dried SPRI beads then were resuspended in 50 µl of 10 mM Tris-HCl, pH 8.0 and once again pelleted against the wall of the MPC. The DNA containing supernatant was transferred to a new microfuge tube.

2.3.2 Methylation of the *Eco*RI site, shearing and blunt-ending

*Eco*RI methylase was used to prevent cleavage of *Eco*RI restriction sites in the sample DNA fragments prior to generating *Eco*RI cohesive ends following adaptor ligation. Here, a reaction mixture containing 30 µl molecular biology grade water, 10 µl EcoRI methylase buffer (10 mM EDTA, 50 mM sodium chloride, 50 mM Tris-HCl, pH 8.0), 5 µl SAM (S-adenosyl methionine, 3.2 mM), 5 µl of EcoRI methylase and 50 µl sheared and purified DNA was incubated at 37°C for 30 minutes. 100 µl of the resulting methylated, hydrosheared DNA was purified by treating with 70 µl of Ampure SPRI beads for 5 minutes, followed by washes with 75% ethanol and elution of DNA in 26 µl of 10 mM Tris, pH 8.0 as described earlier.

Since as described above, mechanical shearing of DNA results in both 3’ and 5’ overhanging ends of double-stranded DNA, the resulting DNA fragments were made blunt ended by T4 DNA polymerase and phosphorylated by T4 polynucleotide kinase prior to blunt ended ligation with adaptors. In this and subsequent steps, all reagents were obtained as components in a Roche Library Kit. In this first step, 26

µl of sheared methylated DNA was incubated with 5 µl 10x polynucleotide kinase (PNK) buffer, 5 µl bovine serum albumin (BSA,) 1mg/ml, 5 µl 10 mM ATP (10 mM), 2 µl dNTPs (10 mM each), 5 µl T4 DNA polymerase (3,000 U/ml), 2 µl T4 polynucleotide kinase (30,000 U/ml) for 15 minutes at 12°C followed by incubation at 25°C for 15 minutes. The DNA then was purified by treatment with 35 µl of Ampure SPRI beads and eluted in 15 µl of 10 mM Tris-HCl, pH 8.0 as described above.

2.3.3 Hairpin adaptor ligation, exonuclease and *EcoR1* digestion

Ligation of hairpin adaptors containing *EcoRI* sites to the ends of the sample DNA protects them from exonuclease digestion used to remove unligated fragments as well as facilitates the generation of sticky ends for DNA circularization using reagents supplied in the Roche Library Kit. To accomplish this, 50 µl 2x rapid ligase buffer (132 mM Tris-HCl, pH 7.6, 20 mM $MgCl_2$, 2 mM ATP, 2 mM DTT, 15% PEG 6000), 30 µl hairpin adaptor, 15 µl sheared, methylated polished DNA and 5 µl T4 DNA ligase were combined and incubated at 25°C for 15 minutes. After adding 100 µl of the ligated product with 2 µl of λ exonuclease, 2 µl of T7 exonuclease and 4 µl of exonuclease I, the mixture was incubated at 37°C for 30 minutes, and then the hairpin adaptor ligated methylated DNA product was purified by treatment with 76 µl of Ampure SPRI beads and elution in 50 µl of 10 mM Tris, pH 8.0 as described above.

Since digestion with *EcoRI* removes the hairpin structures at the ends of the fragments leaving cohesive ends, a reaction mixture containing 30 µl ddH_2O, 10 µl 10 x SuRE/ Cut buffer H (Tris-HCl, $MgCl_2$, NaCl, pH 7.5) and 50 µl of hairpin

adaptor ligated product (all components of the Roche Library Kit), were incubated with 10 µl EcoRI (20,000 U/ml) at 37°C for 16 hours. The EcoRI digested products then were purified by treatment with 70 µl of Ampure SPRI beads and eluted in 50 µl of 10 mM Tris, pH 8.0 as described above.

2.3.4 DNA circularization, DNA Nebulization and Concentration

Intramolecular ligation of the cohesive *EcoRI* ends generated by restriction enzyme digestion of the hairpin adaptors results in circularization of the fragments at the reconstituted *EcoRI* site (44 bp total) flanked by sample DNA on either ends and the exonuclease digestion removes non circularized DNA. DNA circularization was performed by ligating 50 µl (100 ng) of *EcoRI* digested product to itself in the presence of 123 µl molecular biology grade water, 20 µl 10x NE buffer 4 (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM dithiothreitol, pH 7.9 at 25°C), 2 µl ATP (100 mM) by the activity of 5 µl of Rapid ligase (all components in the Roche Library Kit) and incubating at 25°C for 60 minutes. This was followed by exonuclease digestion with 2 µl of λ exonuclease, 2 µl of T7 exonuclease and 4 µl of exonuclease I (10 U/µl) at 37°C for 30 minutes. The ligated circularized DNA then was purified by treatment with 146 µl of Ampure SPRI beads and eluted in 50 µl of 10 mM Tris, pH 8.0 as described earlier.

Hydrodynamic shearing was used to shear the circularized DNA obtained for paired end sequence as well as in earlier studies to shear the DNA used to create non-paired end shotgun libraries. In the case of paired-end sequencing, the circularized DNA molecules were sheared by nebulization such that the resulting fragments contained either, none or both ends of the reconstituted *EcoRI* adaptor.

For both paired-end shotgun and shotgun only DNA pyrosequencing the DNA was suspended to a final volume of 100 µl in TE buffer (10 mM Tris, 1mM EDTA) mixed with 500 µl of ice cold nebulization buffer (53% glycerol, 37 mM Tris-HCl, 5.5 mM EDTA, pH 7.5) and fragmented using a nebulizer (Alliance Medical, Russelville, MO) under 45 psi of nitrogen for 2.5 minutes at -20^{0}C to produce DNA fragments averaging around 700 bases. This nebulized DNA then was concentrated using Qiagen's MinElute PCR purification kit (Qiagen, Valencia, CA) according to the manufacturer's instructions and the purified DNA was eluted with 25 µl of Qiagen's EB Buffer (10 mM Tris, pH 8.0). After determining the concentration of the nebulized, purified DNA on the Caliper AMS 90 using the SE30 DNA Labchip, the DNA was end-repaired, adaptors were ligated and repaired by the fill-in reaction described above in section 2.3.4.

2.3.5 Adaptor Ligation and Quantitation

Ligation of the 44mer 454 adaptors A and B or the 20mer Multiple Identifier (MID) tags (possessing PCR primers, sequencing primer and tag) at either end of the nebulized, polished and purified fragments was carried out to enable amplification of the nebulized library. This reaction was performed by mixing 15 µl of DNA, 20 µl of 2X ligase buffer (132 mM Tris-HCl pH 7.6, 20 mM $MgCl_2$, 2 mM ATP, 2 mM DTT, 15% PEG 6000), 1 µl of an equimolar mixture of adaptors A and B (200 pmol/µl each) or 5 µl of each MID tag and 4 µl of DNA ligase at 25^{0}C for 15 minutes. All reagents were supplied as components in the Roche MID Kit) Adaptor ligated DNA fragments then were recovered by treatment with 28 µl ampure SPRI beads, eluted with 25 µl of of 10 mM Tris, pH 8.0 and end repaired as described

above.

The amount of purified product then was determined on a Calliper AMS 90 using the SE30 DNA Labchip. The average number of molecules/µl of the library was calculated using the formula:

$$\text{Molecules/µl} = \frac{(\text{Sample conc.ng/µl}) \times (6.022 \times 10^{23})}{(656.6 \times 10^{9}) \times (\text{avg. fragment length})}$$

The library then was diluted to $2x\ 10^{8}$ molecules/µl in TE and stored at -20^{0}C.
The single stranded DNA isolation and enrichment of fragments containing A and B adaptor at either end mentioned in the 454 Library preparation protocol recommended by Roche was omitted because during the subsequent emulsion PCR step, double-stranded DNA was melted to single strands during the initial hot start step. Strands carrying the same adaptors at either ends (AA or BB) loop around the annealed A or B primer, resulting in amplification failure.

2.4 Emulsion PCR

The adaptor carrying genomic DNA and EST library fragments required emulsion PCR amplification to yield detectable numbers of copies by pyrosequencing on the GS-FLX sequencer. This was done by incubating a single molecule of DNA carrying bead with high fidelity *Taq* polymerase, dNTPs and amplification primers in the presence of 454 B adaptor complement bound silicon beads in micelles of water in oil emulsions followed by PCR amplification.

First, appropriate amounts of library DNA ($2x10^{5}$ molecules/µl calculated as described above) were mixed with capture beads (600,000 per emulsion) in the ratio that results in the majority of beads containing only a single molecule of. A “live

amplification mix" was prepared by mixing 180 µl of "mock amplification mix" provided with the Roche emPCR Kit along with 10 µl Mg_2SO_4 (2.5 mM), 2 µl amplification primers (0.625 µM forward and 0.039 µM reverse), 6 µl HiFi *Taq* polymerase (0.15 U/µ), 0.3 µl thermostable pyrophosphatase (0.003 U/µl) per emulsion. In the following emulsification step, 500 µl of emulsion oil was added to 240 µl of mock amplification mix and mixed by shaking at a speed of 25 rounds per second for 5 minutes in the Tissue Lyser rack. DNA library carrying beads were mixed with 160 µl of live amplification mix and added to the mock amplification-emulsion oil mix with further shaking at 15 rounds per second for 5 minutes. This emulsion then was carefully dispensed (100 µl/ well) in to a 96 well plate with subsequent amplification in a thermocycler (95^0C hot start initiation for 4 minutes, 40 amplification cycles of alternating 94^0C for 30 seconds, 68^0C for 90 seconds, 13 hybridization extension cycles of alternating 94^0C for 30 seconds, 58^0C for 6 minutes, hold at 4^0C).

2.4.1 Bead recovery, enrichment and sequencing primer annealing

The amplified DNA beads containing micelles were disrupted, recovered with subsequent annealing to sequencing primers prior to sequencing by mixing 100 µl of the emulsion with 200 µl of isopropanol in a 50 ml Corning tube and followed by centrifugation at 3200 rpm for 4 minutes in the Beckman GS6R centrifuge thrice to obtain bead pellets. Following which the bead pellets were washed twice with bead wash buffer and once with 1X Roche "Enhancing Fluid" (2M NaCl, 10 mM Tris-HCl, 1 mM EDTA, pH 7.5) and resupended in 1 ml of the same. Enrichment of DNA carrying beads was performed by adding 100 µl of streptavidin coated

magnetic enrichment beads to the dissolved DNA bead pellet and mixing on the LabQuake tube roller for 5 minutes at room temperature. Amplified DNA carrying beads possess biotin-attached primer at their ends that bind to streptavidin coat on the magnetic beads and are retained against the wall of the magnetic particle collector (MPC).

Following two washes with "Enhancing Fluid", the beads were dissolved in 700 µl of Roche "Melt Solution" (0.125N NaOH, 0.2 M NaCl). The resulting single stranded DNA released into the supernatant was collected by centrifugation, rinsed twice with Roche "Annealing Buffer" (20 mM Tris-HCl pH 7.6, 5 mM magnesium acetate). Annealing of sequencing primer was performed by adding 3 µl of 100 µM sequencing primer and 15 µl of annealing buffer to the pellet of enriched DNA beads and running it on the thermocycler at 65^0C for five minutes, followed by a 0.1^0C /sec temperature drop to 50^0C, 1 minute hold at 50^0C, 0.1^0C per second temperature drop to 40^0C, 1 minute hold at 40^0C with a subsequent 0.1^0C per second temperature drop to 15^0C. Sequencing primer annealed DNA beads were washed twice with annealing buffer and resuspended in 200 µl of the same. The amount of DNA beads was determined using a Beckman Coulter Counter Multisizer 3 and then the DNA beads were stored at 4^0C.

2.4.2 Loading the PicoTiterPlate for Sequencing

Sequencing on the 1.8 million wells containing Roche PicoTiterPlate (PTP) where each well had the dimensions 44 µm x 50 µm x 55 µm, was carried out by sequentially depositing DNA containing beads (30 µm), packing beads and enzyme beads. Here, the PTP first was soaked in Roche "Assay Buffer" containing 25 mM

Tricine, 5 mM magnesium acetate, 8.5 units/L apyrase, 1 mM dithiothreitol, 0.1% BSA, pH 7.8) for 10 minutes at room temperature followed by centrifugation at 2800 rpm for 10 minutes at 25^0C followed by removal of the supernatant prior to bead deposition. The DNA containing beads were mixed with Roche control beads and both the sample DNA bead and the Control bead mix were incubated with 7000 units of Bst DNA polymerase and cofactor in buffer containing 25 mM Tricine, 5 mM magnesium acetate, 8.5 units/L apyrase, 1 mM dithiothreitol, 0.4 mg/ml polyvinyl pyrrolidone, 0.01% Tween 20, 0.1% BSA, 175 μg of *E.coli* single-stranded binding protein, pH 7.8 and incubated on a lab rotator (LabQuake, Thermolyne) for 30 minutes, prior to deposition on the PTP. After allowing deposition of DNA containing beads to settle by standing for 10 minutes, the supernatant was mixed with appropriate amounts of packing beads and centrifuged at 2800 rpm for 10 minutes at 25^0C in a Beckman centrifuge. After withdrawing the supernatant the enzyme bead suspension was deposited onto the PTP by centrifugation at 2800 rpm for 10 minutes at 25^0C. Upon completion of the pre-run rinse, the GS-20 or GS-FLX the reagent cassette kit was inserted into the 454/Roche sequencer followed by additions of apyrase (8.5 units/L) and α-thio dATP (50 μM).

2.4.3 Sequencing and signal processing on the 454 Genome Sequencer.

The signal intensity and position of each nucleotide incorporation of the DNA adhering bead was captured by the CCD camera juxtaposed to the PTP and was processed by the computer software on the GS-20 or GS-FLX. The output was

generated in a standard flowgram format by normalizing the signal of each well by detecting the difference in signal intensity of each base call of the Control bead sequence with that of its known output and assigning Phred-like quality values to each base call.

2.5 Assembling the massively parallel pyrosequencing data for cDNA and genomic DNA

The flowgrams containing the outputs of light intensity measurement for single nucleotide or homopolymer stretches were processed by the 454/Roche GS Run Processer and outputs the sequence reads as individual Standard Flogram Format (SFF) files that then were assembled using the 454/Roche Newbler program by first creating seeds of overlapping flowgram reads and aligning them to generate consensus contigs with final base-calling and quality scores for each base.

In the case of 454 sequenced cDNAs, to obtain expressed sequence tags or ESTs, the reads obtained from each library were assembled using Newbler version 2.0 under default factory settings after trimming for Clonetech SMART CDS 3 ‘ and 5’ PCR primers that were used in each library construction. The ESTs also were screened for the Clonetech SMART CDS 3’ and 5’ reverse complement sequences prior to assembly. Repeats and singletons obtained from each library were aligned to Newbler generated contigs using crossmatch (Green, 1994) at the threshold of 50 nucleotides with a score value of 100.

The *P. omnivora* genome, was assembled was using Newbler version 2.0.0 under default factory settings. Clone-end reads obtained from Sanger sequencing along with the GS 20 generated shotgun and GS FLX mixed paired-end reads were

screened for *E.coli* genomic sequences and assembled by Newbler 2.0.0 by using the option for large genome assemblies.

2.6 Isolation of Individual *P. omnivora* Chromosomes

To aid in the genomic sequence assembly process, individual chromosomes were isolated from protoplast plugs of *P. omnivora* strain OKAlf 8 on CHEF (Contour Clamped Homogenous Electrophoretic Field) gels using chromosomal grade agarose with electrophoresis for 10 days in the cold room. The resulting chromosomal bands were excised from the gel and provided to us by Dr. Carolyn Young at the Noble Foundation. To extract the individual chromosomes from the CHEF gels, the DNA bearing gel pieces were frozen at -20^{0}C for at least an hour and then melted at 65^{0}C. Agarose and DNA binding proteins were precipitated by adding 500 µl TE saturated phenol to equal volumes of melted DNA containing agarose solutions after mixing by vortexing and refreezing at -20^{0}C. The aqueous DNA containing layer was separated from the organic phase by centrifugation for 5 minutes at 25^{0}C at 12,000 rpm in a table-top centrifuge. This aqueous layer was washed twice with equal volumes of water-saturated ether in order to get rid of phenol droplets. The DNA then was subjected to precipitation with 2.5 volumes of 95% ethanol acetate, followed by a wash with 70% ethanol and drying.

The resulting purified, concentrated individual chromosome DNA then was dissolved in 10 µl of 10:0.1TE and subjected to multiple displacement amplification reaction using the REPLI-g Mini kit from Qiagen as per manufacturer's instructions. Here, 2.5 µl of the purified chromosomal DNA was mixed with equal volumes of denaturation buffer for 3 minutes at 25^{0}C (RT) for 3 minutes followed by mixing

with 5 µl of neutralization buffer. A master mix containing 10 µl nuclease-free water, 29 µl reaction buffer (containing dNTPs and exonuclease-resistant primers) and 1 µl of Qiagen's processive, high fidelity, proprietary DNA polymerase capable of replicating up to 100 kb DNA fragments without dissociation was added to the treated chromosomal DNA and isothermal amplification was carried out at 30^0C. The resulting individual chromosomes were verified by electrophoresing on a 1% agarose gel and then subjected to the mixed shotgun paired-end library making protocol illustrated in Figure 4 and described above.

2.7 Analysis of cDNAs

2.7.1 Biological function assignments

To assign biological function to ESTs from *P. omnivora* cDNA libraries, BLAST (Karlin and Altshul, 1990) homology searches were performed for each singleton, repeat and contig using blastx against Genbank non-redundant protein database, KEGG (Kanehisa et al., 2006) and KOG (Tatusov et al., 2003) databases and tblastx against the COGEME (Soans and Talbot, 2006) database. Homology was determined to be significant if the expect value (E value) was less than $1x\ 10^{-4}$. In many cases, singletons, repeats and contigs had homologies to more than one database entries that were equal to and higher than this threshold. Perl scripts, written by Fares Najar included summarize_blast_results that was used to summarize blast results obtained from searches against Genbank database, and an extract_kegg_kog and extract_cogeme that were used to extract and summarize homology results from searches against the KEGG, COG and COGEME databases

based on functional classifications to generate detailed metabolic reconstruction as listed in Appendix Tables 1-6.

2.7.2 Determining EST abundance in each library

Transcript redundancy in each library was determined using an in-house script written by Fares Najar to reveal the overall expression profile of cDNAs from each library. First, the singletons and repeats that aligned to contigs using crossmatch were appended to their respective contigs using the append_redundancytoblast perl script written by Fares Najar. Then, the expression level of each gene was determined using the EST_FZN_expression script also written by Fares Najar that sums the frequency of reads in individual contigs with queries bearing the same KEGG, KOG, COGEME and Genbank accession numbers. The overall expression profile was generated by summarizing the frequency of transcripts involved in each metabolic process based on KEGG annotation and processes involved in information storage and processing and cellular processes from KOG and COGEME annotation. To compare the metabolic profiles of the ESTs derived from each cDNA library the overall expression profile of ESTs from each cDNA library involved in each metabolic process was obtained as described above and a global normalization was carried out by dividing the total number of reads involved in each process by the total number of reads obtained from sequencing of that library on the GS-20 or GS-FLX and then multiplying it 100,000-fold.

2.8 Analysis of the *P. omnivora* genome

2.8.1 Gene prediction and annotation

The first step in annotating the genomic sequences was to perfom a blastn reciprocal homology search of the draft *P. omnivora* genomic sequence using the EST sequences at a cut off expect value (E value) of 10^{-10}. In addition, the *P. omnivora* genomic draft sequence contigs were assigned to chromosomes based on top matches from blastn homology searches with sequence contigs obtained from individual chromosomes at a threshold expect value (E value) of $1x\ 10^{-10}$. Once ordered, the contigs were catenated and genes were predicted using FgenesH (Softberry, Inc., Mount Kisco, New York) that had been trained on several fungi. Predicted proteins were analyzed further by blastp homology searches against Genbank, KEGG and KOG databases and the homology was determined to be significant if the expect value (E value) was less than $1x\ 10^{-4}$. *P. omnivora* predicted proteins with no significant homology to known sequences then were also searched against the Pfam database (Finn et al., 2008) and a *P. omnivora* metabolic map was drawn using the KEGG Automated Annotation Server (http://www.genome.jp/kegg/kaas/). *P. omnivora* tRNAs were predicted by tRNA Scan-SE (Lowe and Eddy, 1997) under default settings for determining eukaryotic tRNA sequences.

2.8.2 Comparison of *P. omnivora* predicted proteins with that of other fungi

S. cerevisiae, M. grisea, N. crassa genomic and protein sequences were downloaded from the Saccharomyces Genome Database available at URL: http://downloads.yeastgenome.org/sequence/, the *Magnaporthe grisea* database at URL: http://www.broad.mit.edu/cgi-bin/annotation/magnaporthe/ and the *Neurospora crassa* database at URL: http://www.broad.mit.edu/annotation/genome/neurospora, respectively.

Predicted proteins from *S. cerevisiae, M. grisea, N. crassa* and *P. omnivora* were searched by tblastn homology against the COGEME database at a threshold expect value (E value) of $1x\ 10^{-5}$. The results were extracted and summarized using the extract_cogeme perl script based on COGEME classifications and listed in an excel spread.

Chapter3

Results and Discussion

3.1 Sequencing and analyzing the ESTs from three different life stages of *P. omnivora* and after exposure to different environmental conditions.

The total RNA isolated from each of the three major *P. omnivora* life stages, and from *P. omnivora* exposed to three different environmental conditions were reverse transcribed from their polyA regions in the presence of deoxynucleotide triphosphates by RNA dependent DNA polymerase and made double-stranded by DNA dependent DNA polymerase to produce six cDNA libraries that were sequenced using massively parallel pyrosequencing on the 454 GS 20 and FLX. As a result, 304,200 EST sequences were obtained, approximately 80% of which had a blastn homology with the draft *P. omnivora* genomic sequence with a predicted genome size of 35-40 Mb as calculated below from Figure 16 A.

3.1.1.1 Assembly of ESTs

As described above, following a series of normalization, correction, and quality filtering algorithms, base calls were generated by the 454/Roche GS Run Processor that produces individual reads with associated quality scores in Standard Flowgram Format (SFF) files. After the Newbler assembler identified pair-wise overlaps with 90% identity over an aligned region of at least 40 bases, the resulting contigs and singletons, (reads that did not overlap to form contigs) and reads with identical signals that the assembler deems to be from repeat regions, were quantitated as shown in Table 3.

Table 3. Assembly statistics of *P. omnivora* cDNA libraries by Newbler

	C/N starved	Non-host root exudate	Host root exudate	Sclerotia	Spore mat	Vegetative
Total #reads	10720	41743	42177	60493	61702	84708
# Contigs	438	1022	1079	1748	2232	707
# Repeats	1073	10644	8601	30627	39768	60855
# Singletons	3569	5888	7467	8659	4179	8419

Transcript abundance was determined using the blastx homology results for contigs, singletons and repeats for each of the six cDNA libraries is shown below in Figure 5.

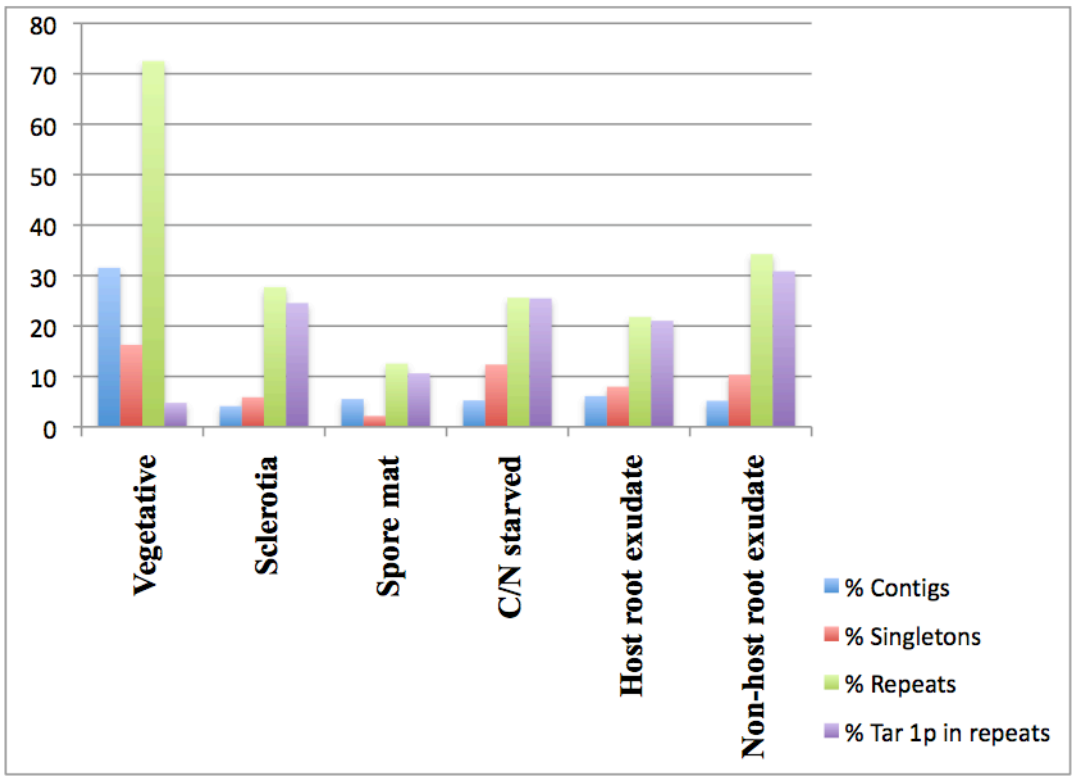

Figure 5. Percent (Y-axis) of contigs, singletons, repeats with blastx homology in GenBank.

As can be seen in Figure 5, 5 to 31% of the contigs, 2 to 16% of the singletons and 12% to 72% of repeats had homologs in GenBank. The spore mat ESTs had the least similarity to GenBank sequences, and since the spore mats of *P. omnivora* differ from those of other fungi in that it fails to germinate, the functions of most of the spore mat ESTs is unknown.

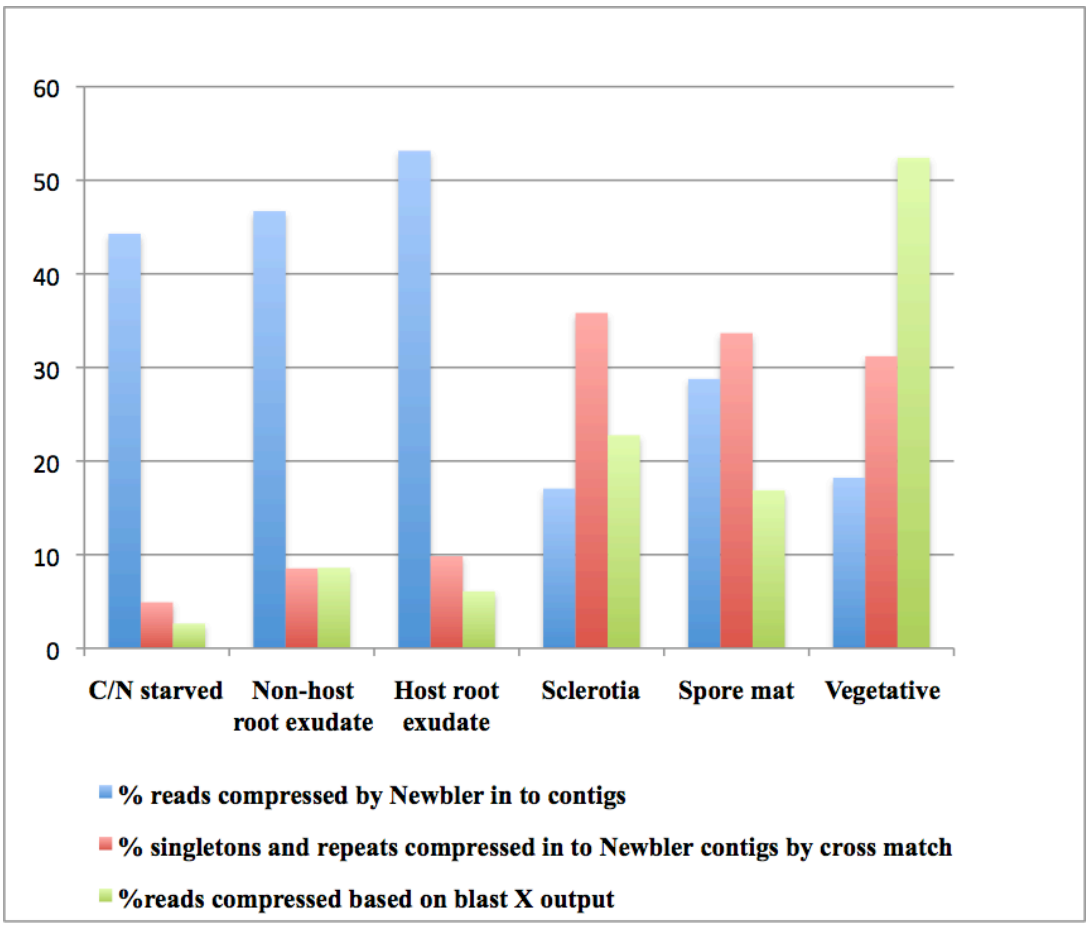

Figure 6. Compression of 454 generated reads from each EST library based on Newbler, Crossmatch and blastx homology. (Y-axis represents percentage as indicated in the footnote to this figure)

Since the 454 sequences were obtained after shearing the cDNA transcripts, those ESTs that overlapped each other with more than 90% identity were assembled by Newbler into contigs. For those that failed to assemble, we used cross match to identify sequence reads that extended the transcripts as well as a perl script to identify reads that belonged to the same transcript based on their blast homology to similar output accession numbers. This allowed additional joining of reads into a final set longer "compressed" ESTs. A comparison of this final set of "compressed" reads from each library is illustrated in Figure 6, where it can be seen that the

percentage of reads that assembled into contigs by Newbler were higher in EST libraries obtained from carbon/nitrogen starved mycelia, from the mycelial interaction with host and from the non-host root exudates. However a greater percentage of ESTs obtained from the vegetative library, sclerotia and spore mats was compressed based on identical accession numbers from blastx homology results. This indicates that the cDNAs from these libraries were fragmented during the library making process resulting in sequencing of different regions of the transcript belonging to the same cDNA. As with other fungal EST studies, see for example Akao et al. 2007 where they observed over 50% of the 21,446 *Aspergillus oryzae* ESTs had no similarity in GenBank, we also observed that a majority of the (~75%) of the 304,208 ESTs also lacked homology to sequences in the GenBank non-redundant public database.

3.1.1.2 The possible role of Tar 1p ESTs in *P. omnivora* cDNA libraries

Another observation that was unusual is that a transcript homologous to the *S. cerevisiae* mitochondrial protein Tar 1p was so abundant in all the libraries, that it was classified as a repeat. The percent of repeats homologous to Tar 1p is indicated in Figure 5. Tar 1p is encoded in the antisense strand of the nuclear 25S rDNA in the rDNA repeat region of Chromosome XII in *S. cerevisiae* and suppresses impaired growth caused by the R129D point mutation in the yeast mitochondrial RNA polymerase gene (*RPO41*) at 30°C and 37°C (Coelho et al., 2002). The *rpo41-R129D* mutation also interferes with mitochondrial translation in *S. cerevisiae* resulting in increased reactive oxygen species, reductions in respiration, slow

growth rate in glycerol and a severely decreased lifespan (Bonawitz et al., 2007). This study also indicated that Tar 1p expression is maintained at very low steady-state levels but increases sharply after post-diauxic shift from anaerobic metabolism (glucose repression) to aerobic respiration under the influence of the non-fermentable carbon source glycerol but is down regulated in response to mitochondrial dysfunction in order to protect from damage caused by the propagation of deleterious reactive oxygen species. Hence, optimal levels of Tar 1p are essential to promoting respiration and in mitigating the damage caused by reactive oxygen species in mitochondria. In addition, yeast two- hybrid assays have revealed that Tar 1p interacts with Coq5p methyl transferase enzyme involved in the synthesis of the antioxidant ubiquinone (Coenzyme Q), and further emphasizes the role of Tar 1p in enhancing the ability of cells to counteract reactive oxygen species (Bonawitz et al., 2008). Bonawitz and co-workers also speculated that the transcription of *TAR1* is antagonistic to the transcription of rRNA transcription in the opposite direction and is reflected by the up-regulation of Tar 1p in glycerol medium where rDNA transcription and biogenesis is down-regulated as a result of slower growth rate implying that the energy intensive process of ribosomal biogenesis is regulated by the site of energy production in the mitochondria (Bonawitz et al., 2008).

3.1.2 EST Sequencing of mycelia grown on M1078 medium

P. omnivora exists as mycelia in the vegetative phase where the fungus moves through the soil and causes infections in encountered plant roots. To understand the cellular and metabolic processes that are active during this phase, 84,708 EST

sequences were obtained from mycelia grown on M1078 media, that upon further assembly, "compression" and analysis, were determined to represent only 857 unique known genes with GenBank homologs.

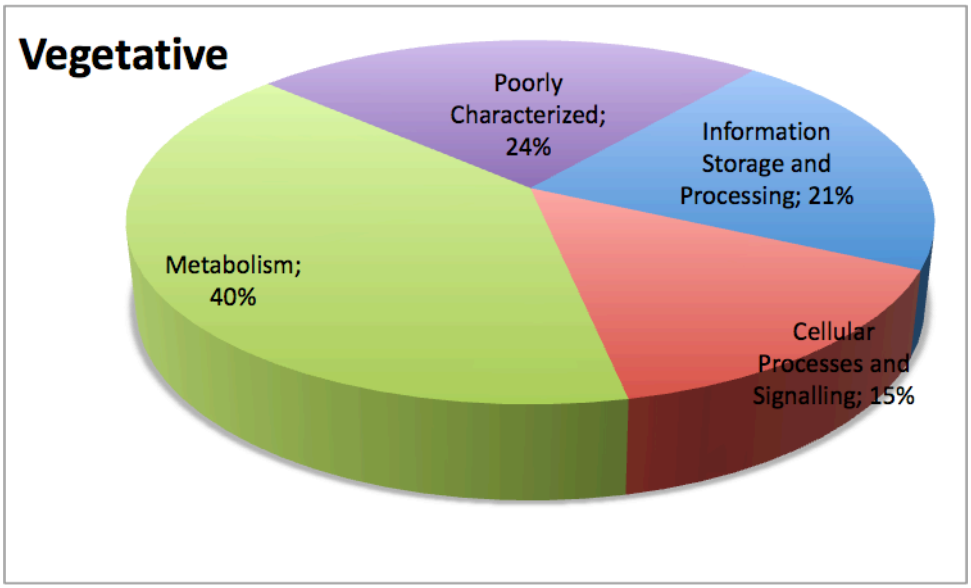

Figure 7. Distribution of genes involved in cellular and metabolic processes of ESTs obtained from mycelia grown on M1078 medium.

In an additional analysis of these 857 mycelia ESTs with blastx homology to genes of known functions, we compared these ESTs to the Kyoto Encyclopedia of Genes and Genomes (KEGG), the Consortium for the functional Genomics of Microbial Eukaryotes (COGEME) phytopathogen database, and the Cluster of Eukaryotic Orthologous Genes (KOG). The results shown in Figure 7 and detailed in Appendix Table 1, reveal that 20% of these ESTs represent genes that are involved in information storage and processing, 40% in metabolism, 21% in cellular processes and signaling and a further 24% are poorly characterized. In addition, genes involved in various pathways associated with carbohydrate, energy, nucleotide, amino acid, glycan and co-factors metabolism also were well

represented in this EST library. Of the genes involved in metabolism, majority of the expressed genes were assigned to carbohydrate and energy metabolism pathways, indicating that the fungus utilizes carbon and expends energy as it propagates via this phase. Also, genes involved in translation, transcription, splicing, recombination and repair, intracellular trafficking, post translational modification and transport facilitation were expressed at similar levels, implying that the propagating mycelia requires the activity of typical housekeeping genes. Since, several genes involved in signal transduction also were observed in this EST library, it is very likely that mycelia have an active cellular response to environmental factors.

Analysis of most highly expressed EST contigs, i.e. transcripts, given above in Figure 7 and shown below in Table 4 reveals EST contigs for four likely fungal RNA dependent RNA polymerases of mitovirus origin, one viral RNA dependent RNA polymerase, Tar 1p, NADH dehydrogenase subunit1, ribosome-associated protein RAP1-like protein, I-PcI endonuclease, and an uncharacterized protein antisense to the ribosomal RNA transcript protein 3. Nine of the top 20 expressed EST contigs have homology to unknown hypothetical proteins in GenBank, indicating that these proteins are expressed proteins with as yet unknown functions. Based on the most highly expressed ESTs, it is evident that high level of mitovirus RNA dependant RNA polymerase activity and mitochondrial respiration and biogenesis occurs in the vegetative mycelia.

Table 4. Description of top 20 most highly expressed transcripts in vegetative mycelia with homology in GenBank

Contig Name	Description	Organism	Score bit	E-value	% Identity
contig00704 length=102 numreads=4794	hypothetical protein NFIA_061320	*Neosartorya fischeri NRRL 181*	71.2	2.00E-11	96
contig00583 length=127 numreads=3338	hypothetical protein CHLREDRAFT_155068	*Chlamydomonas reinhardtii*	67.4	3.00E-10	96
contig00021 length=129 numreads=2242	ref|NP_690845.1| Tar 1p	*Saccharomyces cerevisiae*	62	1.00E-08	82%
contig00610 length=276 numreads=2241	hypothetical protein ACLA_028940	*Aspergillus clavatus NRRL 1*	122	1.00E-26	82%
contig00566 length=161 numreads=1377	hypothetical protein NEMVEDRAFT_v1g225775	*Nematostella vectensis*	65.5	5.00E-15	87%
contig00701 length=295 numreads=553	hypothetical protein CHGG_11103	*Chaetomium globosum CBS 148.51*	48.5	4.00E-11	65%
contig00476 length=3248 numreads=513	RNA-dependent RNA polymerase	*Ophiostoma mitovirus 1b*	233	7.00E-59	35%
contig00511 length=2512 numreads=439	RNA-dependent RNA polymerase	*Ophiostoma mitovirus 3b*	199	7.00E-49	52%
contig00621 length=393 numreads=380	unnamed protein product	*Kluyveromyces lactis*	107	2.00E-22	61%
contig00700 length=429 numreads=358	RNA-dependent RNA polymerase	*Botrytis cinerea debilitation-related virus*	79.7	1.00E-21	49%
contig00706 length=900 numreads=290	RdRp-like protein	*Sclerotinia homoeocarpa mitovirus*	68.2	8.00E-10	50%
contig00656 length=2514 numreads=155	dsRNA viral RNA-dependent RNA polymerase	*Thanatephorus cucumeris*	62.4	2.00E-07	33%
contig00696 length=363 numreads=81	ART3_YEAST Uncharacterized protein ART3 (Antisense to ribosomal RNA transcript protein 3)gb|AAL79278.1| unknown	*Saccharomyces cerevisiae*	91.3	2.00E-17	91%
contig00577 length=102 numreads=47	hypothetical protein	*Vitis vinifera*	53.5	5.00E-06	80%
contig00113	hypothetical protein	*Bacteroides*	57.8	3.00E-07	96%

length=240 numreads=45	BACCAP_02874	*capillosus ATCC29799*			
contig00055 length=868 numreads=27	hypothetical protein	*Yarrowia lipolytica*	106	2.00E-21	56%
contig00509 length=733 numreads=29	NU1M_NEUCR NADH-ubiquinone oxidoreductase chain 1 (NADH dehydrogenase subunit1)	*Neurospora crassa*	72.8	2.00E-11	82%
contig00163 length=130 numreads=25	ribosome-associated protein RAP1-like protein	*Epichloe festucae*	85.5	1.00E-15	93%
contig00628 length=390 numreads=26	I-PcI endonuclease	*Podospora curvicolla*	84.3	5.00E-18	51%

3.1.3 EST Sequencing of Sclerotia

Sclerotia are aggregated mycelial structures that constitute the resting stage of the fungus and enables it to survive deep below the soil surface for extended lengths of time. A cDNA library was constructed using RNA obtained from sclerotia grown in sterile soil and harvested after 14, 28, 42 and 56 days. Sequencing the cDNA library on the GS 20 resulted in 60,493 reads, that on further assembly, "compression" and analysis represented 1,013 unique genes with homology to known genes present in the GenBank non-redundant database.

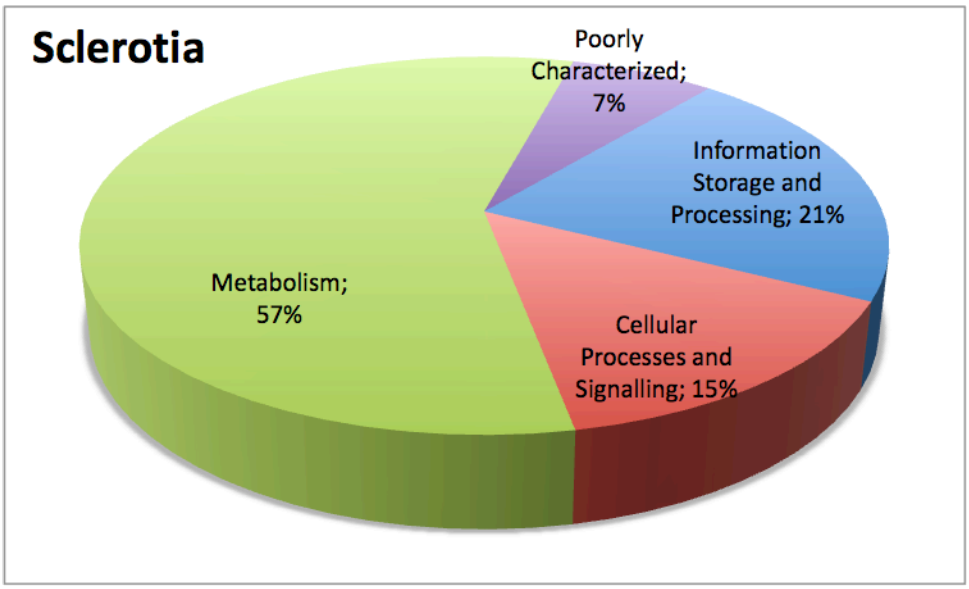

Figure 8. Distribution of ESTs involved in cellular and metabolic processes of ESTs obtained from sclerotia.

An analysis of the results of a blastx comparison of these known function genes against the KEGG, COGEME and the KOG databases is given in Appendix Table 2 and in Figure 8 revealed that 21% of these ESTs were involved in information storage and processing, 57% in metabolism, 15% in cellular processes and signaling and 7% were poorly characterized. Of those involved in metabolism, majority of the ESTs were involved in metabolism of carbohydrates, amino acids, vitamin co factors and lipids while 32% and 10% were involved in glycogen and lipid metabolism respectively. This observation is consistent with the findings of Ergle and Blank (Ergle and Blank, 1947) that by dry weight, sclerotia were composed of 37% of glycogen and 5% lipids. Moreover, several ESTs involved in metabolic pathways for glycolysis and pentose phosphate sythesis, as well as starch, sucrose, pyruvate, propanoate, and butanoate metabolism were observed. Since sclerotia are buried deep below the soil surface, the discovery of ESTs of formate

dehydrogenase that is involved in methane and dicarboxylate metabolism, likely indicates that those compounds could be possible energy sources under anaerobic conditions.

Analysis of the 20 most highly expressed sclerotia EST contigs, i.e. transcripts, as shown in Table 5, revealed Tar 1p, senescence associated protein, and an uncharacterized protein antisense to the ribosomal RNA transcript protein 3. Twelve of the top 20 contigs representing the highest number of transcripts were homologous to hypothetical proteins in GenBank. Based on the most highly expressed ESTs it is evident that high levels of mitochondrial biogenesis, DNA condensation aided by histone H4, protein synthesis as well as senescence occurs within the sclerotia.

Table 5. Description of top 20 most highly expressed transcripts in sclerotia with homology in GenBank

Contig Name	Description	Organism	Score bit	E-value	% Identity
contig01755 length=422 numreads=153	ref|XP_001267666.1| hypothetical protein NFIA_061330	*Neosartorya fischeri NRRL 181*	128	8.00E-29	82%
contig01756 length=209 numReads=49	ref|XP_001214110.1| hypothetical protein ATEG_04932	*Aspergillus terreus NIH2624*	65	7.00E-10	66%
contig01855 length=396 numreads=40	ref|XP_001244658.1| predicted protein	*Coccidioides immitis RS*	56	4.00E-07	38%
contig01805 length=175 numReads=34	ref|XP_767406.1| hypothetical protein GLP_748_1200_211	*Giardialamblia ATCC 50803*	59	6.00E-08	69%
contig01744 length=196 numreads=19	sp|Q8TGM5|ART3_YEAST Uncharacterized protein ART3 (Antisense to ribosomal RNA transcript protein 3)	*Saccharomyces cerevisiae*	89	6.00E-17	93%
contig01701 length=382	ref|XP_001248375.1| hypothetical protein	*Coccidioides immitis RS*	64	1.00E-09	80%

numReads=11	CIMG_02146				
contig00163 length=232 numReads=7	sp|P00411|COX2_NEUCR Cytochrome c oxidase subunit 2 (Cytochrome c oxidase polypeptide II)	*Neurospora crassa*	65	9.00E-10	72%
contig01854 length=220 numReads=6	ref|XP_001239814.1| hypothetical protein CIMG_09435	*Coccidioides immitis RS*	57	3.00E-07	83%
contig00123 length=253 numReads=6	sp|P52809|RL44_PICJA 60S ribosomal protein L44 (60S ribosomalprotein L41)	*Neurospora crassa*	111	1.00E-23	94%
contig01840 length=237 numReads=6	ref|XP_385667.1| H4_NEUCR Histone H4	*Gibberella zeae PH-1*	89	5.00E-17	100%
contig00983 length=220 numReads=6	ref|XP_960578.1| 60S RIBOSOMAL PROTEIN L5 (CPR4)	*Neurospora crassa OR74A*	87	2.00E-16	97%
contig00404 length=234 numreads=1289	ref|NP_690845.1| Tar 1p	*Saccharomyces cerevisiae*	52	4.00E-09	78%
contig01201 length=237 numReads=4	gb|EAT83844.1| hypothetical protein SNOG_08676	*Phaeosphaeria nodorum SN15*	55	1.00E-06	78%
contig00959 length=224 numReads=5	ref|XP_359942.1| hypothetical protein MG11013.4	*Magnaporthe grisea70-15*	52	1.00E-05	82%
contig01452 length=136 numreads=69	ref|XP_001267676.1| hypothetical protein NFIA_043490	*Neosartorya fischeri NRRL 181*	55	7.00E-07	96%
contig01668 length=208 numReads=6	ref|XP_453836.1| unnamed protein product	*Kluyveromyces lactis*	56	5.00E-07	80%
contig01725 length=203 numreads=30	ref|XP_729762.1| senescence-associated protein	*Plasmodium yoelii yoelii str. 17XNL*	60	4.00E-08	81%
contig01782 length=243 numreads=10	ref|XP_001241782.1| conserved hypothetical protein	*Coccidioides immitis RS*	119	4.00E-26	94%
contig01856 length=222 numReads=7	ref|XP_453836.1| unnamed protein product	*Kluyveromyces lactis*	85	1.00E-15	90%
contig01857 length=370 numReads=9	ref|XP_001269594.1| hypothetical protein ACLA_028940	*Aspergillus clavatus NRRL 1*	111	9.00E-24	78%

3.1.4 EST Sequencing of Conidial Spore mat

Upon exposure to moisture, *P. omnivora* forms spore mats that are 2 to 16 inches in diameter, white to tan in color, and are contained in large branched aerial fungal strands called conidia that seem to be a “dead end” (see Figure 2 in the Introduction) as attempts to germinate conidia have not been successful. To investigate the metabolic and cellular processes active in this stage, a cDNA library constructed from spore mats isolated from the soil surface was sequenced and resulted in 61,702 sequence reads, that upon further assembly, “compression”, and analysis revealed 366 unique expressed genes with homologs in the GenBank non-redundant database. The results of a blastx analysis of these known function genes against the KEGG, COGEME, and KOGs databases, given in Appendix Table 3 and shown below in Figure 9 indicates that 56% of the ESTs represent expressed genes have significant homology to genes that are involved in information storage and processing, 26% in metabolism of which 30% and 20% are carbohydrate and energy metabolism genes, 12% in cellular processes and signaling and 6% are poorly characterized.

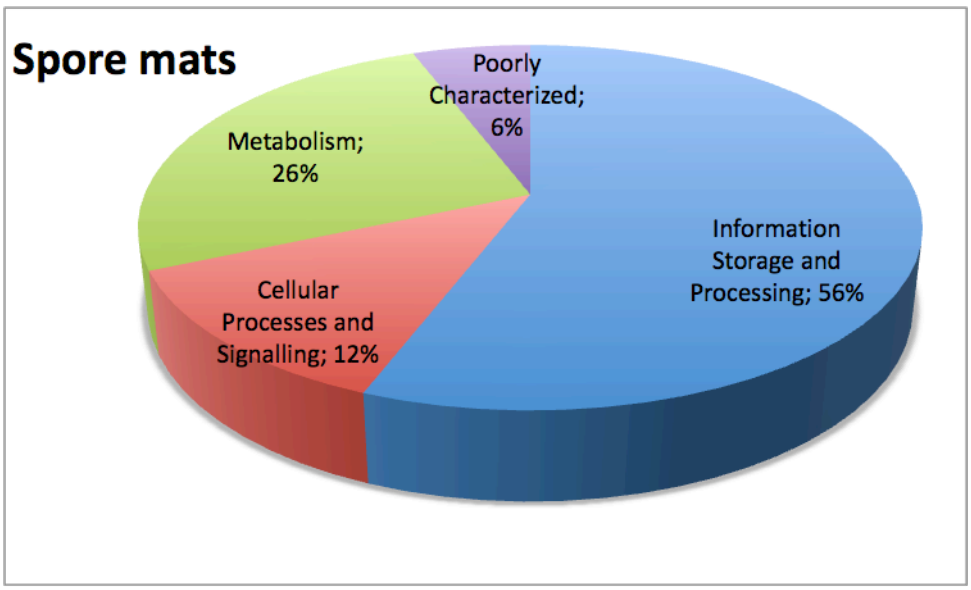

Figure 9. Distribution of genes involved in cellular and metabolic processes in ESTs obtained from spore mats.

Of the ESTs involved in information storage and processing, 20% were observed to be involved in protein translation while 74% of the ESTs were categorized in chromatin structure and dynamics with a majority of the transcripts encoding for histone H4. Of the ESTs involved in cellular processes and signaling, 50% and 12% belonged to genes responsible for posttranslational modification and signal transduction respectively.

A similar EST analysis from activated spores of the arbuscular mycorrhizal fungus *Gigaspora rosea* indicates a high level of gene expression also for proteins mainly involved in translation and protein processing, replication, cell cycle and signal transduction as well as a metallothionein-encoding gene involved in metal binding (Stommel et al., 2001). In addition, expression of a gene homologous to *E. nidulans* methyltransferase that negatively regulates sexual development also was present and unique to this EST library, an interesting observation since *P. omnivora* is not known to pass through a sexual phase. High expression levels of the conidial

hydrophobin found in cell walls of fungal conidia and histone H4 involved in chromosome structure and dynamics are uniquely characteristic to this life stage. The expression of genes involved in carbohydrate, nucleotide, lipid, amino acid and energy metabolism indicates that spores are not dormant and utilize energy gained from metabolism in order to complete the sporulation process.

Analysis of most highly expressed EST contigs, i.e. transcripts, as shown in Table 6, reveals expression of a conidial hydrophobin, histones H4 and H3, senescence associated protein, ubiquitin-ribosome fusion protein, Tar 1 and transcript antisense to the ribosomal RNA transcript, although twelve of the top 20 contigs are homologous to hypothetical and unnamed proteins in GenBank. Based on the most highly expressed ESTs it can be said that spore mats are mainly involved in conidial protein production, DNA condensation by histones H3 and H4, senescence and mitochondrial biogenesis.

Table 6. Description of top 20 most highly expressed transcripts in spore mats with homology in GenBank.

Contig Name	Description	Organism	Score bit	E-value	% Identity
contig01253 length=489 numreads=206	sp|P23750|H41_EMENI Histone H4.1Histone H4H4.1	*Emerciella nidulans*	149	9.00E-34	100
contig01991 length=375 numReads=153	sp|Q8TGM5|ART3_YEAST Uncharacterized protein ART3(Antisense to ribosomal RNA transcript protein 3)unknown	*Saccharomyces cerevisiae*	90.9	2.00E-17	88%
contig02179 length=440 numreads=121	ref|XP_001259730.1| conidial hydrophobin Hyp1/RodA	*Neosartorya fischeri NRRL 181*	52.4	9.00E-06	38%

contig02147 length=355 numReads=20	ref|XP_505708.1 | hypothetical protein	*Yarrowia lipolytica*	72.8	7.00E-12	60%
contig02030 length=153 numReads=18	ref|XP_729762.1 | senescence-associated protein	*Plasmodium yoelii yoelii str. 17XNL*	53.1	5.00E-06	77%
contig02105 length=176 numReads=17	ref|NP_001078516.1| histone H3.2	*Arabidopsis thaliana*	76.3	6.00E-13	92%
contig02132 length=353 numreads=17	ref|XP_715467.1 | ubiquitin-ribosomal protein fusionS27a	*Candida albicans SC5314*	119	1.00E-30	92%
contig02052 length=337 numReads=15	dbj|BAE57827.1| unnamed protein product	*Aspergillus oryzae*	119	8.00E-26	76%
contig02074 length=545 numReads=14	ref|XP_369461.1 | hypothetical protein MGG_06003	*Magnaporthe grisea 70-15*	48.5	4.00E-12	67%
contig02200 length=284 numReads=14	ref|XP_453852.1 | unnamed protein product	*Kluyveromyces lactis*	65.1	1.00E-09	93%
contig01593 length=193 numReads=11	gb|EAT85053.1| hypothetical protein SNOG_07587	*Phaeosphaeria nodorum SN15*	65.5	1.00E-09	88%
contig02031 length=382 numReads=9	ref|XP_001248375.1| hypothetical protein CIMG_02146	*Coccidioides immitis RS*	59.7	3.00E-08	82%
contig02127 length=182 numReads=9	gb|AAX30301.1| unknown	*Schistosoma japonicum*	74.7	2.00E-12	87%
contig01536 length=236 numreads=2437	ref|NP_690845.1 Tar 1p	*Saccharomyces cerevisiae*	65.5	1.00E-09	78%
contig02038 length=547 numReads=240	ref|XP_453843.1 | unnamed protein product	*Kluyveromycesl actis*	96.3	8.00E-19	70%
contig01464 length=339 numreads=301	ref|XP_001624691.1| predicted protein	*Nematostellave ctensis*	79	9.00E-14	73%
contig02146 length=159 numReads=26	gb|AAX30301.1| unknown	*Schistosoma japonicum*	68.6	1.00E-10	91`%
contig02157 length=497 numreads=78	ref|XP_956002.1 | HISTONE H4	*Neurospora crassa OR74A*	92.4	4.00E-36	100%

contig02194 length=260 numreads=15	ref|XP_001269594.1| hypothetical protein ACLA_028940	*Aspergillus clavatus NRRL 1*	112	6.00E-24	82%
contig02195 length=273 numReads=60	ref|XP_001618200.1| hypothetical proteinNEMVEDRAFT_v1g155353	*Nematostella vectensis*	77.4	3.00E-13	63%
contig02218 length=144 numReads=13	ref|XP_001624579.1| predicted protein	*Nematostellavectensis*	66.6	5.00E-10	91%
contig02050 length=215 numreads=10	ref|XP_724982.1| hypothetical protein PY04653	*Plasmodium yoelii yoelii str. 17XNL*	66.6	5.00E-10	81%

3.1.5 EST Sequencing of Carbon and Nitrogen Starved mycelia

To better understand the cellular and metabolic processes that are active in response to starvation, cDNAs were obtained from mycelia grown on M1078 media deprived of either carbon or nitrogen and then pooled. Sequencing of the cDNA library on the GS 20 resulted in 10,720 reads, which on further assembly, "compression" and analysis were found to represent 429 unique known genes homologous to genes present in the GenBank non-redundant database.

A blastx analysis of known function genes against KEGG, KOG and COGEME databases depicted in Figure 10 and listed in Appendix Table 4 revealed that 28% of the observed ESTs had homology to genes that were involved in information storage and processing, 60% in metabolism, 9% in cellular processes and signaling and 3% were poorly characterized. Of the ESTs representing genes involved in metabolism, a majority of them were involved in carbohydrate, nucleotide and amino acid metabolism indicating that although the fungus is metabolically active it responds to carbon or nitrogen deprivation by increasing the

expression of selected metabolic pathways. The lack of ESTs representing genes involved in secondary metabolism and expression of the C2H2 type Zn finger protein, and relative transcript abundance of histone H4 implies that the fungus regulates the expression of genes at the transcriptional level.

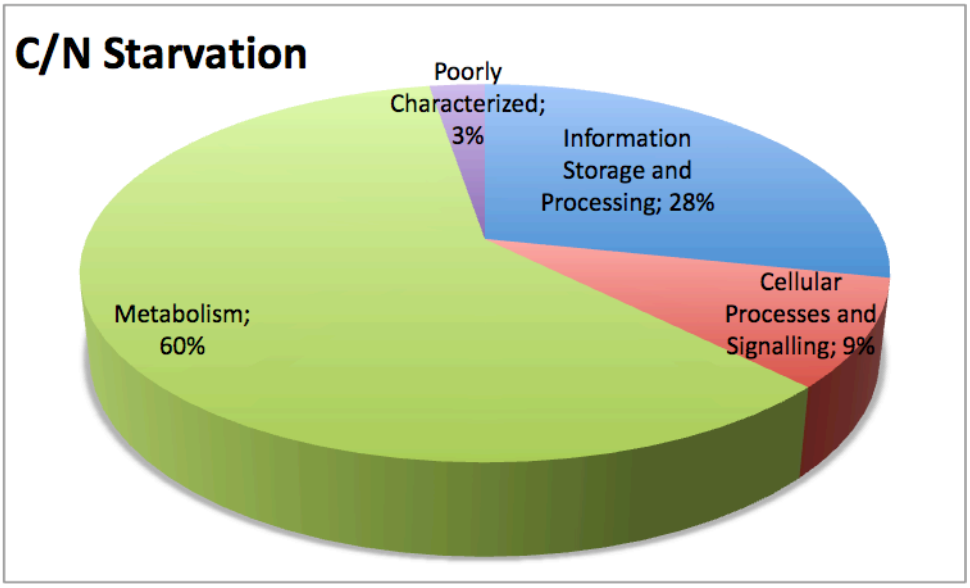

Figure 10. Distribution of genes involved in cellular and metabolic processes in ESTs obtained from Nitrogen and Carbon starved mycelia

Analysis of most highly expressed EST contigs, i.e. transcripts, as listed in Table 7, revealed expressions of ESTs representing enolase, 40S ribosomal protein, cytochrome oxidase subunit II, NADH dehydrogenase, a fungal RNA dependant RNA polymerase, cell cycle check point homologue (*CHK1*), and histone H4 genes. Similar to our findings in the other EST libraries, thirteen of the top 20 expressed EST contigs were found to be similar to hypothetical and unnamed proteins in GenBank. The high abundance of ESTs representing the yeast checkpoint homologue CHK1 indicates tight cell cycle regulation.

Table 7. Description of top 20 most highly expressed transcripts in carbon/nitrogen starved mycelia with homology in GenBank

Contig Name	Description	Organism	Score bit	E-value	% Identity
contig00011 length=173 numreads=86	ref|XP_001241787.1| conserved hypothetical protein	*Coccidioides immitis RS*	61	2.00E-08	93%
contig00262 length=464 numreads=78	emb|CAJ83813.1| CHK1 checkpoint homolog (*S. pombe*)	*Xenopus tropicalis*	64	2.00E-09	85%
contig00034 length=216 numreads=12	ref|YP_667832.1| cytochrome oxidase subunit II	*Verticillium dahliae*	58	1.00E-07	62%
contig00008 length=281 numreads=11	ref|XP_001267666.1| hypothetical protein NFIA_061330	*Neosartorya fischeri NRRL 181*	101	1.00E-20	79%
contig00250 length=373 numreads=8	ref|XP_001269594.1| hypothetical protein ACLA_028940	*Aspergillus clavatus NRRL 1*	155	5.00E-37	84%
contig00039 length=213 numreads=7	gb|EAT78193.1| hypothetical protein SNOG_14322	*Phaeosphaeria nodorum SN15*	76	5.00E-13	55%
contig00329 length=352 numreads=7	ref|XP_453836.1| unnamed protein product	*Kluyveromyces lactis*	101	9.00E-21	88%
contig00051 length=166 numreads=6	dbj|BAE56329.1| unnamed protein product	*Aspergillus oryzae*	63	5.00E-09	87%
contig00056 length=133 numreads=5	gb|EAT90946.1| hypothetical protein SNOG_01297	*Phaeosphaeria nodorum SN15*	70	4.00E-11	86%
contig00124 length=215 numreads=4	ref|NP_775398.1| NADH dehydrogenase subunit 3	*Lecanicillium muscarium*	60	4.00E-08	57%
contig00014 length=159 numreads=3	ref|XP_755719.1| histone H4	*Aspergillus fumigatus Af293*	54	3.00E-06	96%
contig00079 length=227 numreads=3	dbj|BAE57827.1| unnamed protein product	*Aspergillus oryzae*	82	5.00E-15	94%
contig00088 length=149 numreads=3	gb|EAT90242.1| predicted protein	*Phaeosphaeria nodorum SN15*	62	8.00E-09	77%
contig00246 length=155 numreads=3	ref|XP_382717.1| RS14_NEUCR 40S ribosomal protein S14 (CRP2)	*Gibberella zeae PH-1*	56	6.00E-07	100%
contig00464 length=172 numreads=3	emb|CAJ32468.1| RNA-dependent RNA polymerase	*Ophiostoma mitovirus3b*	70	4.00E-11	60%
contig00071 length=100	ref|XP_001239814.1| hypothetical protein	*Coccidioides immitis RS*	56	5.00E-07	81%

numreads=2	CIMG_09435				
contig00161 length=172 numreads=2	sp|P42040|ENO_CLAHE Enolase (2-phosphoglycerate dehydratase)(2-phospho-D-glycerate hydro-lyase) (Allergen Cla h 6) (Cla h VI)	*Aspergillus fumigatus*	93	3.00E-18	97%
contig00337 length=214 numreads=2	ref|XP_363655.1| hypothetical protein MG01581.4	*Magnaporthe grisea70-15*	82	9.00E-15	97%
contig00351 length=180 numreads=2	ref|XP_661620.1| hypothetical protein AN4016.2	*Aspergillus nidulans FGSC A4*	67	2.00E-10	66%
contig00066 length=110 numreads=1	ref|XP_001216358.1| conserved hypothetical protein	*Aspergillus terreus NIH2624*	54	2.00E-06	75%

3.1.6 EST Sequencing of mycelia exposed to host root exudates

To study the interaction of *P. omnivora* with host root proteins, roots were obtained from *M. truncatula* and *M. sativa*, both host plants for *P. omnivora*, were macerated and exudates were incubated with fungal mycelia.

When a cDNA library was constructed and sequenced from total RNA of mycelia exposed to host root exudates, 42,177 sequence reads were obtained, which upon further assembly, “compression” and analysis, were found to represent 627 unique genes with GenBank homologs.

A blastx analysis of known function genes against the KEGG, COGEME and KOGs databases listed in Appendix Table 5 and in Figure 11, revealed that 40% of the ESTs represented were involved in information storage and processing, 44% in metabolism, 13% in cellular processes and signaling and a further 3% were poorly characterized. Observing genes involved involved in carbohydrate, nucleotide, amino acid, vitamins, co-factors and energy metabolism indicates that the fungus avails itself of these nutrients from the host root exudate. Expression of transcripts

involved in cellular detoxification via catalase and perriredoxin and expression of pectin degradation, NADPH oxidase, structural protein homologue as well as G protein coupled signaling transcripts are indicative of the cellular response to host proteins present in the root exudate and also contribute to the pathogenic traits of *P. omnivora*. Remarkably, in response to *P. omnivora* infection, *M. truncatula* roots were found to upregulate class I and class IV chitinase as well as genes involved in reactive oxygen species generation and phytohormone signaling (Uppalapati et al., 2009).

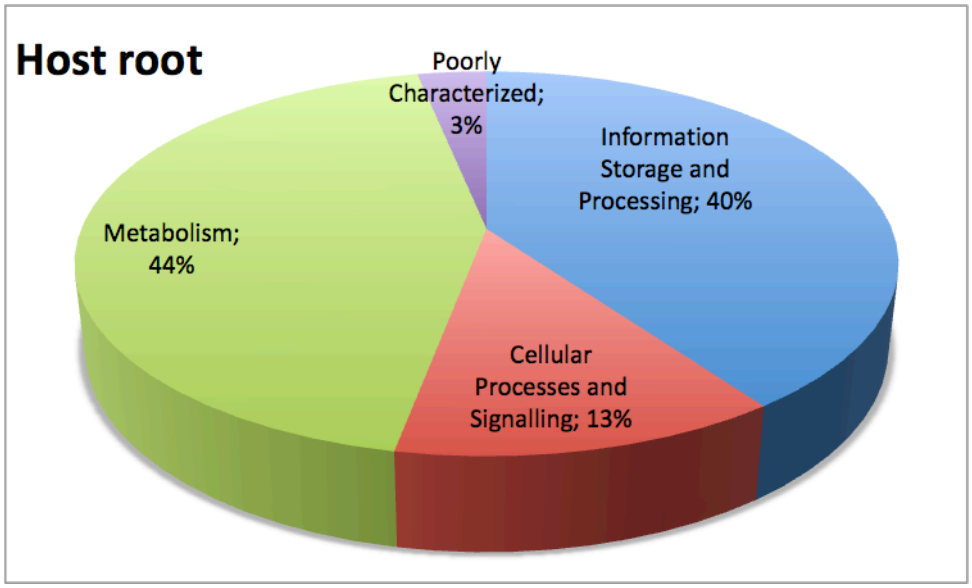

Figure 11. Distribution of genes involved in cellular and metabolic processes in ESTs obtained from mycelia exposed to host root exudate.

Analysis of most highly expressed ESTs, i.e. transcripts, as shown in Table 8, reveals the expression of 60S ribosomal proteins, cytosolic ribosomal protein rps29 , NADH dehydrogenase subunit 3, cytochrome oxidase I intronic ORF 5, cellular checkpoint protein CHK1, ARP2/3 complex 20 kDa subunit, Cytochrome c

oxidase polypeptide II, antimicrobial resistance protein, Tar 1p, an uncharacterized protein antisense to the ribosomal RNA transcript protein 3 and eleven ESTs that are homologous to GenBank hypothetical and unnamed proteins.

Table 8. Description of top 20 most highly expressed transcripts in mycelia exposed to host root exudate with homology in GenBank

Contig Name	Description	Organism	Score bit	E-value	% Identity
contig00994 length=190 numreads=191	sp|Q8TGM5|ART3_YEAST Uncharacterized protein ART3 (Antisense toribosomal RNA transcript protein 3)	*Saccharomyces cerevisiae*	91	1.00E-17	91%
contig00820 length=241 numreads=140	emb|CAJ83813.1| CHK1 checkpoint homolog (S. pombe)	*Xenopus tropicalis*	52	7.00E-06	92%
contig00032 length=674 numreads=62	ref|XP_001005117.1| PREDICTED: hypothetical protein	*Mus musculus*	164	4.00E-39	61%
contig00481 length=536 numreads=56	ref|XP_453836.1| unnamed protein product	*Kluyveromyces lactis*	92	1.00E-17	87%
contig00821 length=557 numreads=54	ref|XP_001217456.1| conserved hypothetical protein	*Aspergillus terreus NIH2624*	93	7.00E-18	54%
contig01155 length=242 numreads=51	gb|EAT89272.1| hypothetical protein SNOG_04067	*Phaeosphaeria nodorum SN15*	121	1.00E-26	80%
contig00989 length=351 numreads=23	ref|XP_001241782.1| conserved hypothetical protein	*Coccidioides immitis RS*	118	1.00E-25	91%
contig00259 length=551 numreads=22	sp|P15956|NU3M_EMENI NADH-ubiquinone oxidoreductase chain 3 (NADHdehydrogenase subunit 3)	*Emerciella nidulans*	81	2.00E-14	50%
contig00982 length=193 numreads=20	ref|XP_453849.1| unnamed protein product	*Kluyveromyces lactis*	88	1.00E-16	95%
contig00843 length=406 numreads=12	ref|XP_381438.1| hypothetical protein FG01262.1	*Gibberella zeaePH-1*	82	7.00E-15	73%
contig00177 length=258 numreads=11	ref|XP_001267837.1| conserved hypothetical protein	*Aspergillus clavatus NRRL 1*	67	3.00E-10	60%
contig00523 length=180 numreads=8	ref|XP_001237702.1| ENSANGP00000030087	*Anopheles gambiae str.PEST*	59	5.00E-08	45%
contig00134	ref|NP_013263.1|	*Saccharomyces*	57	2.00E-07	61%

length=231 numreads=6	Putative protein of unknown function;overexpression confers resistance to the antimicrobial peptide MiAMP1	*cerevisiae*			
contig00147 length=203 numreads=6	ref|XP_747840.1| ARP2/3 complex 20 kDa subunit (p20-ARC), putative	*Aspergillus fumigatus Af293*	65	9.00E-10	85%
contig00345 length=209 numreads=6	ref|XP_751150.1| cytosolic ribosomal protein (Rps29a), putative	*Aspergillus fumigatus Af293*	60	2.00E-08	77%
contig00007 length=187 numreads=5	ref|XP_364314.1| hypothetical protein MG09159.4	*Magnaporthe grisea70-15*	58	1.00E-07	72%
contig01046 length=157 numreads=5	ref|NP_690845.1| Tar 1p	*Saccharomyces cerevisiae*	59	5.00E-08	70%
contig00041 length=419 numreads=27	prf||1703265F cytochrome oxidase I intronic ORF 5	*Podospora anserina*	55	9.00E-07	62%
contig00086 length=275 numreads=14	sp|P00411|COX2_NEUCR Cytochrome c oxidase subunit 2 (Cytochrome c oxidase polypeptide II)	*N. crassa*	91	2.00E-17	71%
contig00502 length=445 numreads=23	ref|XP_001241783.1| hypothetical protein CIMG_05679	*Coccidioidesimmitis RS*	58	8.00E-08	75%

3.1.7 EST Sequencing of mycelia exposed to non-host root exudates

Since *P. omnivora* does not infect monocots, root exudates of sorghum were exposed to the fungal mycelia and a cDNA library was constructed and sequenced to study the nature of the interaction between the fungus and a non-host plant. A total of 41,743 sequence reads were obtained, and when they were assembled, "compressed" and analyzed, 613 unique genes with GenBank homologs were present.

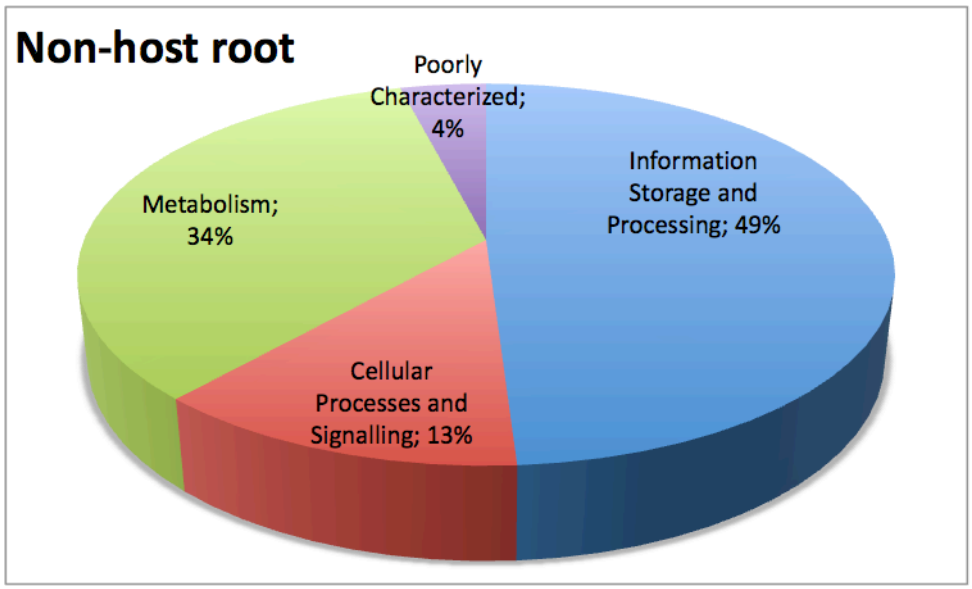

Figure 12. Distribution of genes involved in cellular and metabolic processes in ESTs obtained from mycelia exposed to non-host root exudate.

The blastx analysis of these known function genes against the KEGG, KOG and COGEME databases is presented in Figure 12 and detailed in Appendix Table 6 reveals that 49% of the ESTs represented genes involved in information storage and processing, 34% in metabolism, 13% in cellular processes and signaling and 4% were poorly characterized. Of the detected metabolic related genes, a majority of them were involved in carbohydrate metabolism. ESTs representing genes belonging to amino acid, nucleotides, glycans, co-factors, lipids and energy metabolic pathways also were observed, indicating that the fungus utilizes nutrients available from the root exudates of non-host plants. The observed ESTs of stress responsive proteins and rad16 nucleotide excision repair along with 1,4-beta-D-glucan cellobiohydralase cell wall degrading protein implies that the organism may be combating toxic products present in the host root while availing itself of nutrients.

Analysis of most highly expressed EST contigs, i.e. transcripts, as shown in Table 9, reveals expression of a fungal RNA dependant RNA polymerase, NADH dehydrogenase subunit 1, I-SceI DNA endonuclease, endonuclease encoded by the mitochondrial group I intron of the 21S_rRNA gene, senescence-associated protein, alpha-1,4-glucan lyase, NADP-dependent mannitol dehydrogenase, cytochrome oxidase I intronic ORF 5, antisense to ribosomal RNA transcript protein 3, cytochrome c oxidase subunit 2 and a maturase, while nine of the top 20 contigs were homologous to hypothetical and unknown function proteins present in GenBank.

Table 9. Description of top 20 most highly expressed transcripts in mycelia exposed to non-host root exudate with homology in GenBank

Contig Name	Description	Organism	Score bit	E-value	% Identity
contig01050 length=1063 numreads=182	ref|XP_001269594.1| hypothetical protein ACLA_028940	*Aspergillus clavatus NRRL 1*	179	4.00E-43	79%
contig00992 length=468 numreads=137	gb|AAX30301.1| unknown	*Schistosoma japonicum*	107	1.00E-22	90%
contig00044 length=689 numreads=90	ref|XP_453836.1| unnamed protein product	*Kluyveromyces lactis*	150	8.00E-35	77%
contig00030 length=554 numreads=31	ref|XP_001265421.1| conserved hypothetical protein	*Neosartorya fischeri NRRL 181*	79	1.00E-13	48%
contig00068 length=561 numreads=29	sp|P15956|NU3M_EMENI NADH-ubiquinone oxidoreductase chain 3 (NADHdehydrogenase subunit 3)	*Emerciella nidulans*	70	4.00E-11	42%
contig00624 length=515 numreads=24	ref|NP_009324.1| I-SceI DNA endonuclease, encoded by themitochondrial group I intron of the 21S_rRNA gene; mediates gene conversion that propagates the intron into intron-less copies of the21S_rRNA gene	*Saccharomyces cerevisiae*	76	8.00E-13	59%

contig00050 length=376 numreads=19	sp|P00411|COX2_NEUCR Cytochrome c oxidase subunit 2 (Cytochrome coxidase polypeptide II)	*Neurospora crassa*	94	1.00E-18	80%
contig00505 length=206 numreads=12	ref|XP_364314.1| hypothetical protein MG09159.4	*Magnaporthe grisea70-15*	74	2.00E-12	73%
contig00989 length=190 numreads=10	gb|AAB84211.1| putative maturase	*Cryphonectria parasitica*	63	3.00E-09	60%
contig00455 length=223 numreads=8	ref|XP_001229052.1| predicted protein	*Chaetomium globosum CBS148.51*	75	1.00E-12	84%
contig00642 length=217 numreads=7	ref|XP_001210155.1| conserved hypothetical protein	*Aspergillus terreus NIH2624*	61	2.00E-08	60%
contig00004 length=271 numreads=65	ref|XP_729762.1| senescence-associated protein	Plasmodium yoeliiyoelii str. 17XNL	60	3.00E-08	81%
contig00006 length=525 numreads=52	emb|CAB52260.1| alpha-1,4-glucan lyase	*Morchella costata*	89	9.00E-17	58%
contig00162 length=190 numreads=11	gb|ABB55877.1| NADP-dependent mannitol dehydrogenase	*Tuberborchii*	63	3.00E-09	87%
contig00447 length=147 numreads=1771	ref|NP_690845.1| Tar 1p	*Saccharomyces cerevisiae*	67	2.00E-10	79%
contig00612 length=115 numreads=10	ref|ZP_01142625.1| hypothetical protein GuraDRAFT_1187	*Geobacteruraniumreducens Rf4*	57	2.00E-07	67%
contig00820 length=730 numreads=45	prf||1703265F cytochrome oxidase I intronic ORF 5	*Podospora anserina*	87	7.00E-16	37%
contig00921 length=307 numreads=67	sp|Q8TGM5|ART3_YEAST Uncharacterized protein ART3 (Antisense to ribosomal RNA transcript protein 3)	*Saccharomyces cerevisiae*	57	2.00E-07	96%
contig00948 length=167 numreads=7	sp|P00411|COX2_NEUCR Cytochrome c oxidase subunit 2 (Cytochrome coxidase polypeptide II)	*N. crassa*	110	3.00E-23	92%
contig00986 length=168 numreads=99	ref|XP_001267666.1| hypothetical protein NFIA_061330	*Neosartoryafischeri NRRL 181*	91	1.00E-17	83%

3.2 Comparative analysis of P. omnivora EST libraries

3.2.1 Comparison of metabolic profile of ESTs derived from different life stages and environments

Digital expression profiling of ESTs is based on the assumption that the *in vivo* transcript copies of a given gene are directly proportional to *in vitro* synthesized cDNAs that are randomly sequenced from a non-normalized library. Hence the total number of ESTs in a given population reflects an estimate of the expression levels (Kozian and Kirschbaum, 1999). As described above, transcripts from six different libraries were normalized, sequenced and analyzed. The comparative metabolic profile of the three distinct *P. omnivora* morphological stages-vegetative mycelia, sclerotia and spore mat is shown in Figure 13 reveals that expression of genes involved in carbohydrate metabolism is quite high in all three morphological stages of *P. omnivora* development.

The relative number of transcripts involved in carbohydrate, energy, nucleotides, amino acids, glycans, co-factors and vitamin metabolism were the highest in vegetative mycelia, implying that the fungus avails itself of nutrients and expends energy, to grow and produce the newly branching mycelia while actively propagating. During the sclerotial phase, the fungus actively converts host derived nutrients for storage in the form of carbohydrates and lipids, as the relative number of observed ESTs involved in lipid metabolism was higher in sclerotia than in any other stage and most likely aids in maintaining cell viability during adverse conditions.

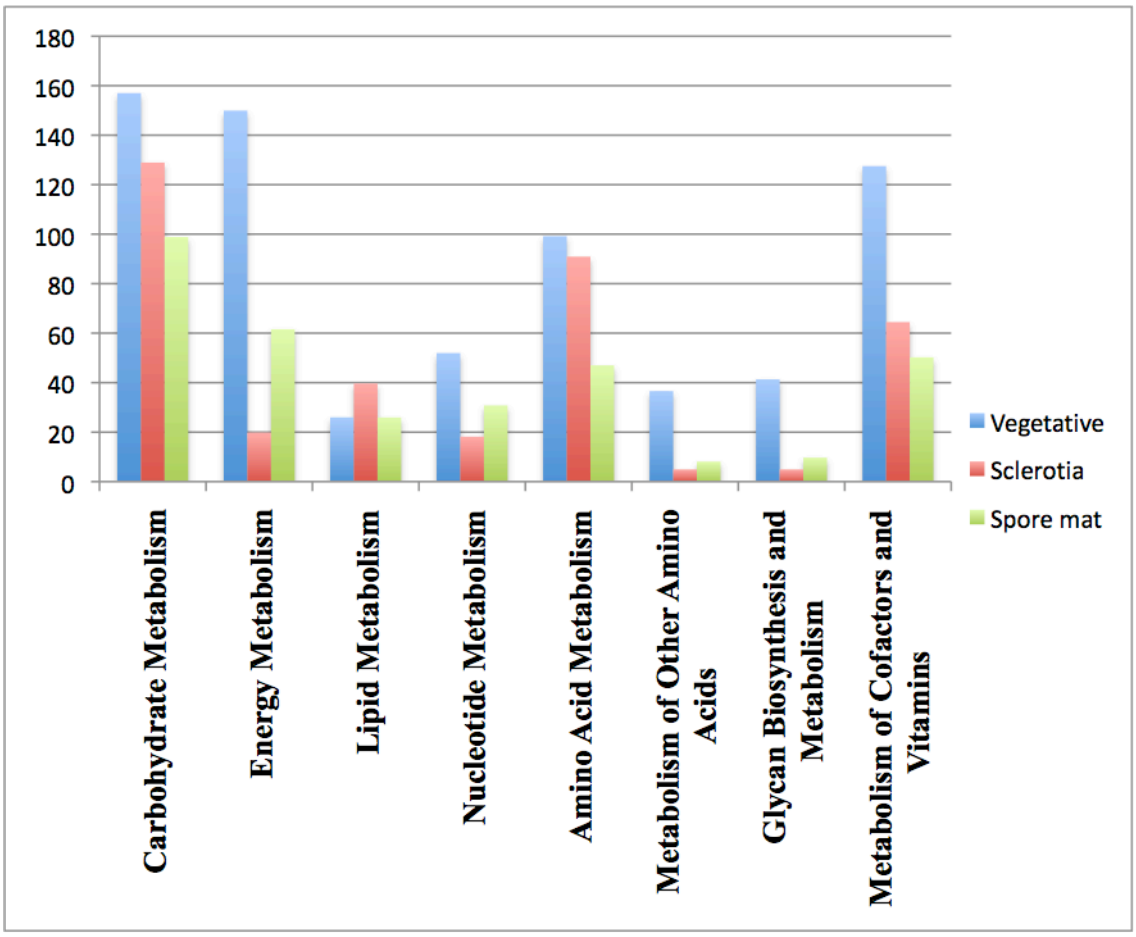

Figure 13. Comparison of metabolic profile from vegetative mycelia, sclerotia and spore mat of *P. omnivora*

In the spore mats, the mycelia branch upward to produce aerial conidia bearing spores with utilization of energy for sporulation. However, since during the sclerotial phase the fungus is resting and storing nutrients, it likely expends the least amount of energy. This is consistent with the observation that ESTs involved in energy metabolism were relatively less abundant in spore mats and least observed in sclerotia. ESTs involved in glycan biosynthesis also were least observed in sclerotia and spore mats, indicating that the fungus is not actively producing newly branching mycelia during these stage of the life cycle. A similar comparative analysis of cDNA

from the entomopathogenic fungus *Beauvaria bassiana* also revealed stage specific gene expression in the aerial conidia, *in vitro* blastospores and submerged conidia (Cho et al., 2006.

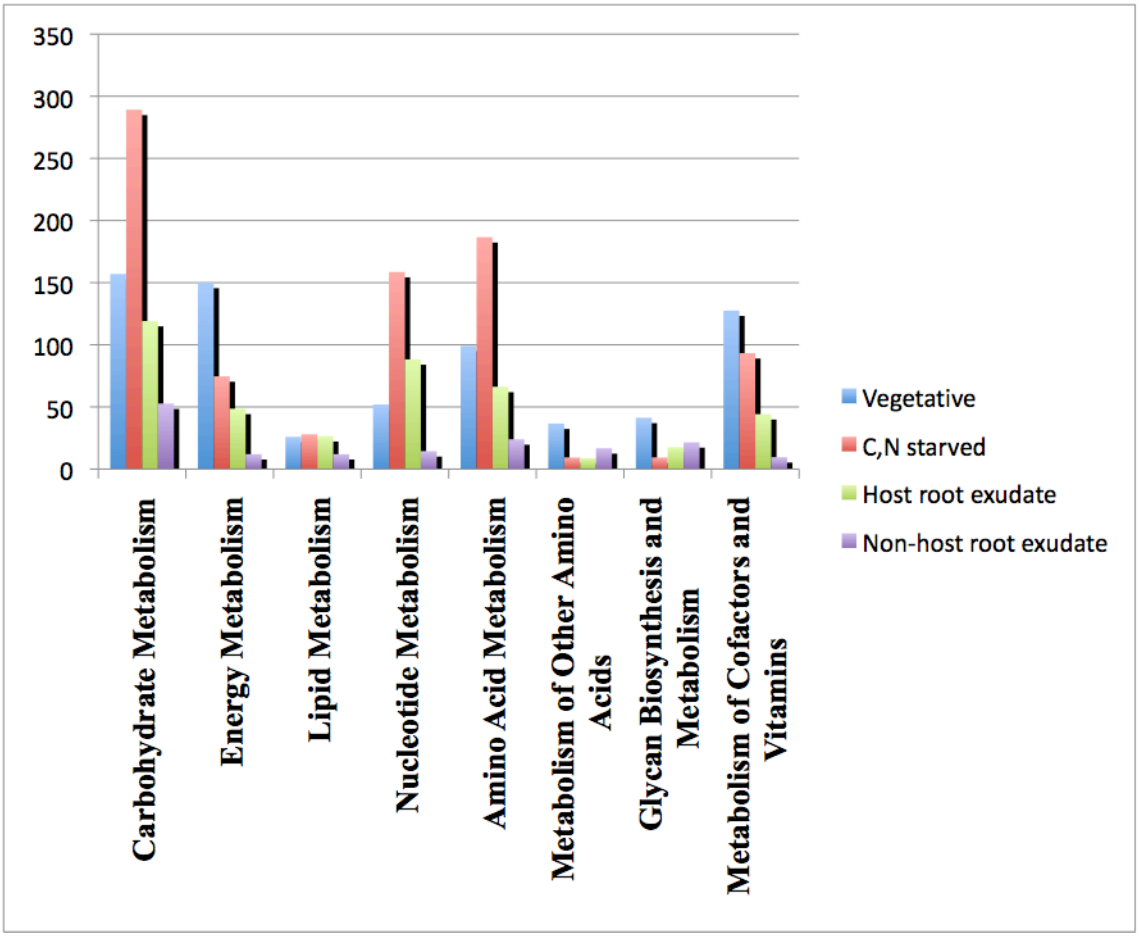

Figure 14. Comparison of metabolic profile of mycelia in the vegetative stage and in response to carbon/nitrogen starvation, host and non-host root exudate.

The relative number of EST's derived from mycelia exposed to different environmental conditions were compared, and is illustrated in Figure 14. Comparison of the metabolic profile of mycelia exposed to different nutrient sources with that of mycelial response to media deficient in carbon or nitrogen, indicates

that the fungus produces relatively higher number of transcripts involved in carbohydrate, amino acid and nucleotide metabolism in response to starvation. This may be a result of the fungus utilizing endogenous carbon and nitrogen reserves to maintain viability during adverse conditions. It is however not clear if the fungus responds to carbon starvation by actively metabolizing amino acids and nucleotides since the EST's derived from mycelia exposed to carbon and nitrogen starvation were obtained separately and then pooled. This was done because very low amount of cDNA was acquired from mycelia exposed to each of the two starvation condition and may be reflective of the fungus shutting down transcription of most genes. Comparison of mycelia exposed to host root exudate with that of non-host root exudate reveals abundant ESTs representing genes involved in all of the metabolic processes depicted in Figure 14 and is reflective of the fungus's adaptation to utilize host root derived nutrients. Since mycelia grown on M1078 medium expressed relatively higher number of metabolic transcripts than those grown on host and non-host root exudates, since the nutrients availability in the M1078 medium likely is much higher than available in the host or non-host root exudates. Interestingly, *P. omnivora* ESTs representing formate dehydrogenase were observed only in the sclerotial stage while the aldehyde dehydrogenase ESTs involved in detoxification were observed in response to both host and non-host root exudate and all three morphologically distinct stages.

In EST studies in the fungus *Metarhizium anisopliae var acridium*, a fungal species known to infect the locust *Schistocerca gregaria* but not the beetle *Leptinotarsa decimlineata*. Expression patterns in response to host and non-host

extracts also showed upregulated levels of genes involved in metabolism, utilization of cuticle, cell survival, detoxification, and signal transduction. Moreover, genes expressed in response to host extract showed a higher number of genes involved in cell division and accumulation of cell mass, whereas genes involved in detoxification and redox processes were more abundant in response to non-host extract, signal transduction genes involved in plant pathogenicity however were only found upregulated in response to host extract (Wang and Leger, 2005).

From our analysis, it is evident that in response to host root exudates, *P. omnivora* expresses ten percent more genes involved in metabolism as compared to non-host. Mycelia exposed to host and non-host root exudate also were expressed at a comparable level for genes involved in replication, chromatin dynamics, transcription, translation, protein turnover, intracellular trafficking and cytoskeleton, indicating a likely ability of *P. omnivora* to adapt to its non-host. However genes involved in RNA processing and modification and cell cycle control were greater in response to non-host, indicating regulation of transcription and the cell cycle. Signal transduction often has been implicated with fungal pathogenesis (Lev et al., 1999; Catlett et al., 2003), and although sensory serine/threonine kinases were in both, the MEKK related kinases as well as a polyketide synthase gene were only were expressed in response to host root exudate. ESTs for the pathogenicity related Snod Protein1 and a structural toxin homologue also were expressed only in response to host root, further establishing host-specific virulence of *P. omnivora*.

3.2.2 Does oxidative stress trigger metamorphosis in *P. omnivora*?

It has been proposed that microbial cell differentiation is initiated to counteract adverse environmental conditions (Moore, 1998). Under oxidative stress, the cell is unable to neutralize all of its free radicals causing it to lower its intracellular oxygen concentrations by reducing or limiting its entry into the cell (Aguirre et al., 2005). This also includes isolation from external water and other sources of soluble oxygen, and causes the cell to rely on intracellular sources of reducing power. When the ability of the cell to reduce molecular oxygen is exhausted (inavailability of electron donor molecules), it adapts to a physiologically stable state that is least permeable to molecular oxygen. Inability of the cell to adapt to production of reactive oxygen species by increasing its capacity to reduce oxygen or prevent its entry leads to cell death, whereas restoration of its capacity to counteract reactive oxygen species results in reversal of its primary state (Georgiou et al., 2006).

Since sclerotia are composed of highly aggregated mycelia that can bury itself deep below the soil surface (Streets and Bloss, 1973) away from molecular oxygen and conidiation of mycelia to form sporemats is triggered by mycelia exposed to moisture/humid conditions (Dunlap, 1941), it was of interest to investigate the relative abundance of Tar 1p ESTs in each library. Tar1p though well known to suppress the mitochondrial RNA polymerase R129D mutation in *S. cerevisiae* is maintained at low steady state levels in the cell to prevent the propagation of reactive oxygen species caused by mitochondrial dysfunction (Bonawitz et al., 2008). As illustrated in Figure 15 below, the % ESTs with Tar 1p

homologs were highest in sclerotia (12%) followed by sporemats (7%) and least in vegetative mycelia (3%), implying that it is undergoing the least oxidative stress. This analysis therefore supports the observation that fungal metamorphosis is triggered by oxidative stress (Georgiou et al., 2006). It is interesting to note that Tar 1p ESTs comprise as high as 80 % and 69% of sclerotial and sporemat ESTs that have GenBank homologs but only 6% of the ESTs of vegetative mycelial that share GenBank homologs indicating that many of the genes involved in fungal metamorphosis remain to be studied.

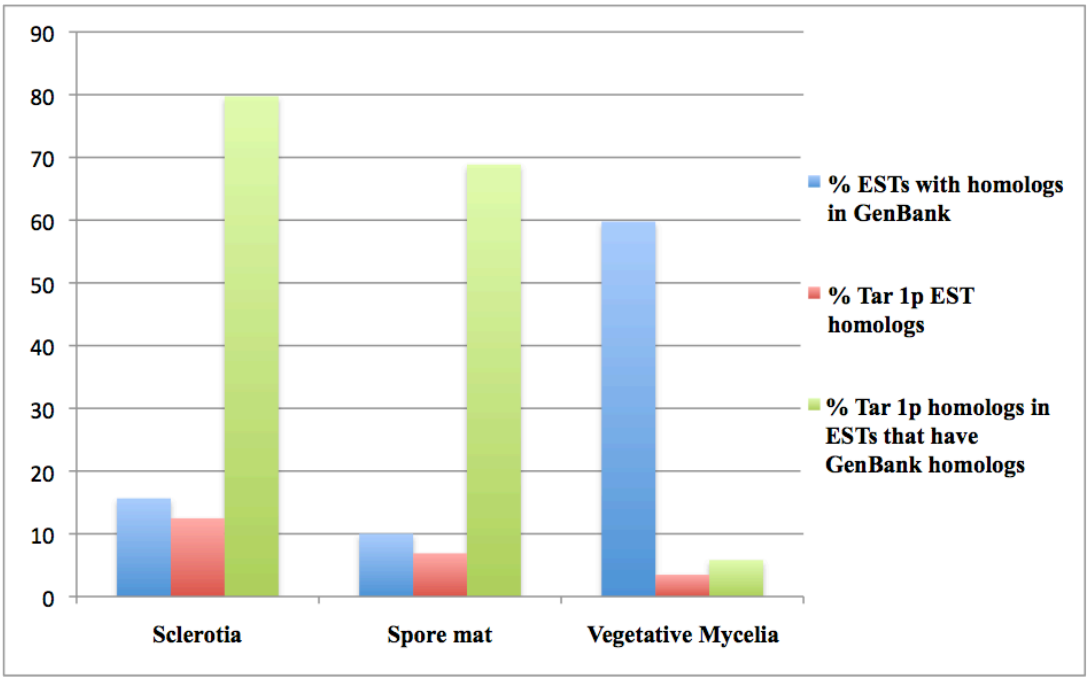

Figure 15. Distribution of Tar1 p ESTs in *P. omnivora* cDNA libraries obtained from sclerotia, sporemat and vegetative mycelia.

3.3 The *P. omnivora* genome and assembly

The *Phymatotrichopsis omnivora* genome has been reported to be an obligate heterokaryon belonging to the class Pezizomycetes (Hosford and Gries, 1966; Marek et al., 2009). Failure to obtain a *P. omnivora* hyphal tip culture and lack of germination of conidia both of which contain between at least 2-3 nuclei in their cells, and the observation of anastomosing hyphae, abundance of diploid, aneuploid and polyploid nuclei, strain differences and attenuation indicates that a parasexual cycle coupled with the presence of several heterokaryotic nuclei within individual hyphae may be responsible for the sustainance and genetic flexibility of the pathogen in the absence of a sexual stage or functional asexual conidial spores (Hosford and Gries, 1966). The ploidy level and number of chromosomes in fungi are known to range between 1X-50X, and 2-20 respectively, and the average fungal genome size ranges between 10-60 Mb (Gregory et al., 2007), although there are reports of fungi with genomes as large as 795 Mb (Hijri and Sanders, 2005). Though little is known about the genome size of *P. omnivora* as it has not been studied earlier, in this present study, a total of 9,141,261 sequence reads representing ~1,125 x 10^6 total bases of the *P. omnivora* genome were obtained and assembled to yield a genome size of ~74 Mb, as shown in Table 10.

Table 10. *P. omnivora* whole genome shotgun sequencing and assembly statistics

	Assembly with Newbler
Cumulative reads	9,141,261 reads totaling ~1,125 Mbp
# Assembled	4,732,716
# Singletons	1,217,668
# Repeats	2,439,326
# Large Contigs	48,259
# Bases in large contigs	43,954,605
Size of largest contig	31,764
Total # contigs	168,644
Total #bases in contigs	73,674,409

3.3.1 Sequencing individual *P. omnivora* chromosomes

In an effort to estimate more accurately the genome size of *P. omnivora*, our collaborator, Dr. Carolyn Young at the Noble Foundation isolated protoplasts and resolved the seven individual chromosomes on CHEF gels by Pulse Field Gel Electrophoresis as shown in Figure 16. Bands representing each of the putative chromosomes were excised from the gel and supplied to us. Subsequently, six of the seven individual chromosomes were purified from the gel, amplified and used to construct a 454 paired-end library that when sequenced resulted in 4,388 genomic contigs that could be assigned to individual chromosomes.

As shown in Figure 16 A, B although the diffuse chromosome banding pattern is characteristic of heterokaryotic nuclei as opposed to sharp distinct bands from the yeasts *S.cerevisiae*, *S.pombe* and the filamentous fungus *Neotyphodium* hybrid isolates and thus the individual bands may include aneuploid chromosomes or isochromosomes as has been reported in other fungi, such as azole drug resistant strains of *C.albicans* (Selmecki *et al.,* 2006). Summing the estimated Mb size for each of the seven *P. omnivora* chromosomes resolved on this pulse field gel gives a genome size that was estimated to be at least 35-40 Mbp. In addition, based on the EST study shown above in section 3.1, where 304,200 EST sequences were obtained with approximately 80% having a blastn homology to the genomic sequence assembly shown in Table 10, it is clear that the overwhelming majority of this genomic sequence is represented in this most recent assembly.

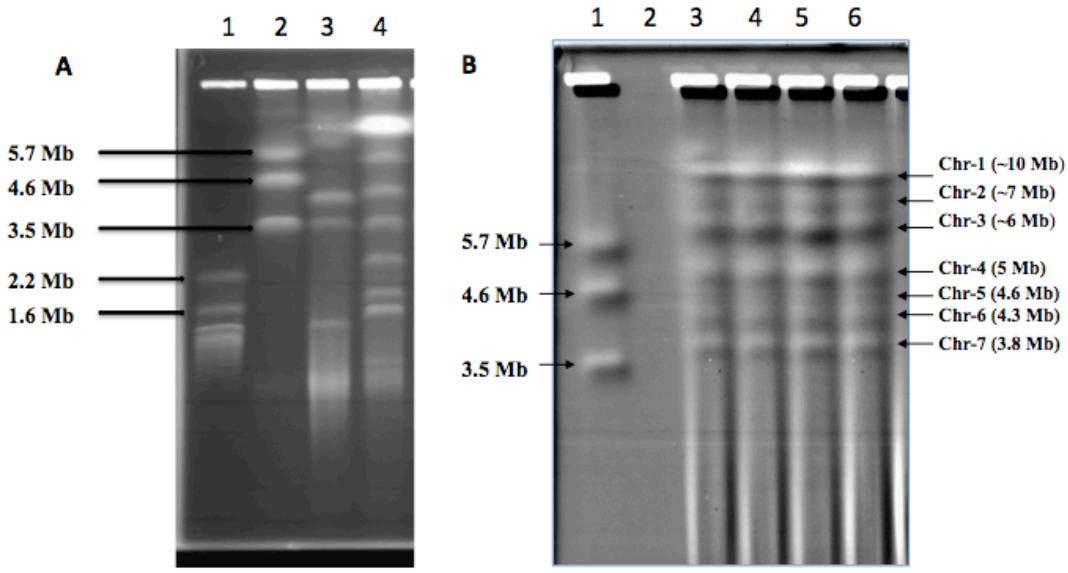

Figure 16. Pulse Field Gel Electrophoresis gel, using a contour-clamped homogeneous electric field (CHEF) of fungal protoplasts (A) Lanes 1- *S. cerevisiae*, 2- *S. pombe*, 3-*Neotyphodium* hybrid 1001, 4-*Neotyphodium* hybrid1002, (B) Lanes 1- *S. pombe,* Lane 2- blank, Lane 3-6 *P. omnivora* OK alf-8 to determine the size of its seven chromosomes

To obtain sequence data from each of the individual chromosomes, six of the seven diffuse bands were isolated from the gel, amplified and sequenced on the Roche/GS FLX. Although much less than one-fold coverage of each chromosome was obtained, it was possible to assemble the data using Newbler and the assembly statistics for each chromosome as well as the number of genomic contigs and bases that were assigned to each chromosome are shown in Table 11. Clearly additional data is needed if one were to sequence each chromosome, but at this stage, those experiments must await the longer reads on the GS-FLX using the recently released Titanium chemistry (Roche Diagnostics, personal communication).

Table 11. Preliminary sequencing results of *P. omnivora* amplified chromosomes

Chromosome	2	3	4	5	6	7
Cumulative reads	93,722 ~20Mb	66,486 ~13 Mb	59,760 ~11.3Mb	177,144 ~32.5Mb	51,323 ~9.3 Mb	18,714 ~3.7Mb
# Assembled	21,569	10,220	16,787	37,180	9,140	4,457
# Singletons	4,035	3,915	3,788	11,058	3,285	5,389
# Repeats	61,475	48,613	33,959	116,359	34,329	1,424
# Large Contigs	45	82	29	126	19	151
# Bases in large contigs	30,888	98,043	20,271	109,513	17,334	116,992
Size of largest contig	2,125	11,077	2,556	7,413	2,555	2,686
Total # contigs	1,087	1,218	1,042	2,534	936	1,852
Total #bases in contigs	66,434	343,852	242,325	637,219	213,166	472,213

As mentioned above, we estimated that *P. omnivora* has an approximate genome size of at least 35-40 Mbp. However, this approach of determining total genome size by adding the sizes of chromosomes resolved on a CHEF gel is fairly accurate only if during vegetative growth, fungal chromosomes and chromosomal segments are not randomly lost resulting in differences in estimated genome sizes (Beadle et al., 2003) and if there is no co-migration of different individual chromosomes that are of approximately the same size.

3.4 Predicted protein profile of *P. omnivora*

Most gene prediction is based on homology mapping to experimentally verified genes in model organisms. With the availability of systematic gene deletion studies in *S.cerevisiae*, a definitive map of essential genes has been developed and characteristic sequence features associated with essential genes has been adapted to gene prediction programs to accurately identify essential genes in the

Saccharomyces species (Seringhaus *et al.* 2006). This is a first step in the accurate heterogenomic gene prediction of essential genes among the unicellular yeasts, and is yet to be extrapolated to filamentous fungi.

Using our combined genomic and chromosomal sequence data, the *P. omnivora* genes were predicted using FgenesH and the gene prediction matrix of the filamentous fungi *Aspergillus nidulans*. Approximately 22,000 genes were predicted, 8,974 of which shared homology with proteins in the GenBank non-redundant database and a further 12,857 showed homology to proteins in the pfam database, albeit at very low stringency. The distribution of predicted proteins in functional categories is illustrated in Figure 17. About 20% of the proteins that showed homology to proteins in the pfam database were viral elements and a further 20% were classified as proteins of unknown function that were expressed in other organisms. Also, a total of 173 aminoacyl-tRNAs were predicted in the genome using tRNA ScanSE, in addition to 113 pseudo-tRNAs.

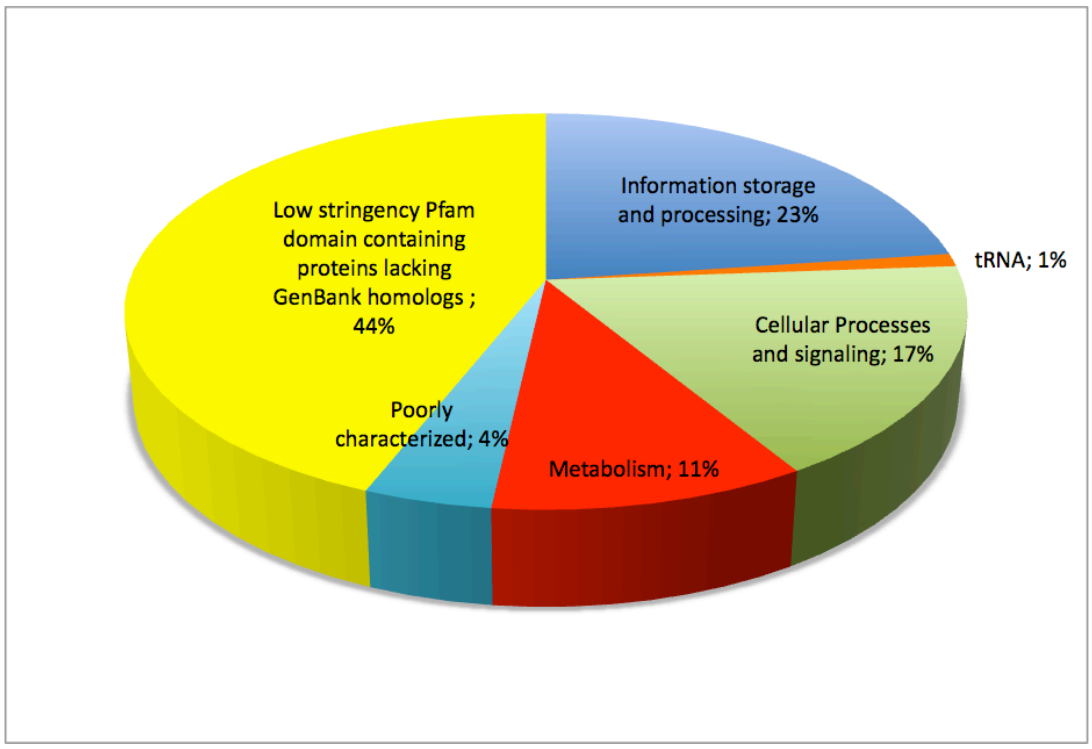

Figure 17. Distribution of *P. omnivora* predicted genes in functional categories based on tRNA ScanSE, KOG and KEGG annotation.

3.4.1 Analysis of *P. omnivora* predicted proteins

3.4.1.1 Comparison of *P. omnivora* predicted proteins with that of other fungi

The *P. omnivora* predicted genes, along with *M. grisea*, *S. cerevisiae* and *N. crassa* predicted proteins, were searched separately against the COGEME database to determine and compare the functional proteins predicted in these four fungal species. These results, shown in Figure 18, indicate that *P. omnivora* contains a comparable number of proteins involved in metabolism and cellular processes as the other filamentous fungi, *M. grisea* and *N. crassa,* but a slightly higher number than

the unicellular yeast *S. cerevisiae*.

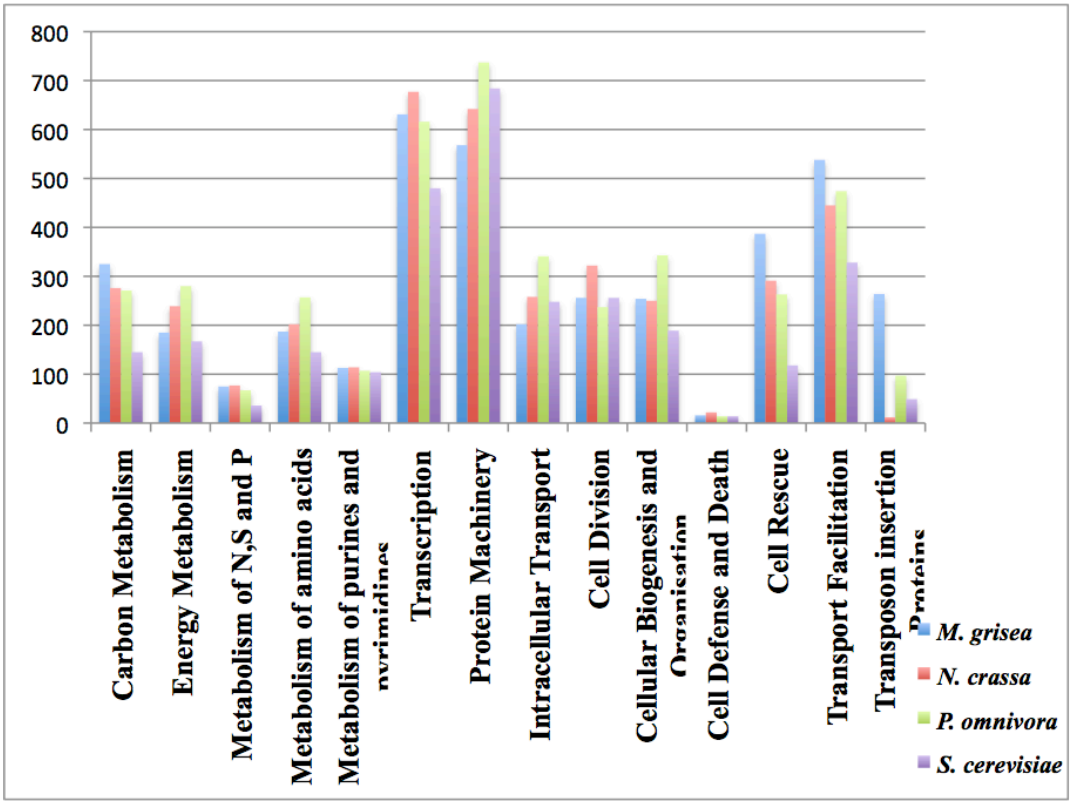

Figure 18. Relative numbers of *P. omnivora, M. grisea, N. crassa* and *S. cerevisiae* predicted proteins involved in various cellular processes based on homology with the COGEME database.

Predicted proteins involved in cell rescue by facilitating DNA repair, ageing, polysaccharide degradation and detoxification were more prevalent in *M. grisea* as compared to *P. omnivora* and included higher numbers of cytochrome P450s, superoxide dismutases, mono and dioxygenases as well as other cell rescue proteins. A significantly higher number of transposon insertion sequence proteins also were observed in *M. grisea,* the well-studied rice blast pathogen. The number of *P. omnivora* predicted and annotated proteins involved in energy metabolism, amino acid metabolism, protein synthesis, intracellular transport, and cellular biogenesis

and organization were higher than observed in *M. grisea, N. crassa* and *S. cerevisiae* indicating a plausible role of proteins involved in those processes in the survival and ability of the fungus to cause infection.

3.4.1.2 Metabolism in *P. omnivora*

Based on the annotated predicted proteins, a metabolic profile map of the *P. omnivora* genome was obtained using the biochemical pathways function on the Kegg Annotation Server. The pathways involved in carbohydrate metabolism, e.g. glycolysis and the citric acid cycle, the pentose phosphate pathway as well as those pathways for galactose, fructose and mannose metabolism, fatty acid metabolism and steroid biosynthesis, and the nucleotide and amino acid biosynthesis pathways in *P. omnivora* were analysed and are discussed below.

P. omnivora proteins predicted from genomic contigs and involved in various enzymatic reactions are represented in the figures below by their Enzyme Commission numbers (EC) in green shaded boxes, enzymes found in singletons are enclosed in green boxes, and enzymes that were not found in the assembly are encircled in red.

3.4.1.2.1 Glycolysis, the pentose phosphate pathway and gluconeogenesis.

Glycolysis is the series of enzymatic reactions that converts hexoses such as glucose and fructose to acetyl-CoA or pyruvate prior to oxidation via the citrate cycle or fermentation to ethanol or lactate. Among the three glycolysis pathways described (Griffen, 1994), only the Emben-Meyerhof-Parnas (EMP) and Hexose

Monophosphate (HM) pathways (also known as the oxidative arm of the pentose phosphate pathway) are prevalent in most fungi while the Entner-Doudoroff (ED) pathway has been observed in only a few fungal species (Griffen, 1994). The sorbitol bypass pathway proposed to be involved in the conversion of sorbitol to triose phosphate in animal cells has been extended to fungi as well (Jeffrey and Jornwall, 1988) and is represented in Figure 19. The enzymes involved in this conversion are aldo reductase, sorbitol dehydrogenase and hexokinase.

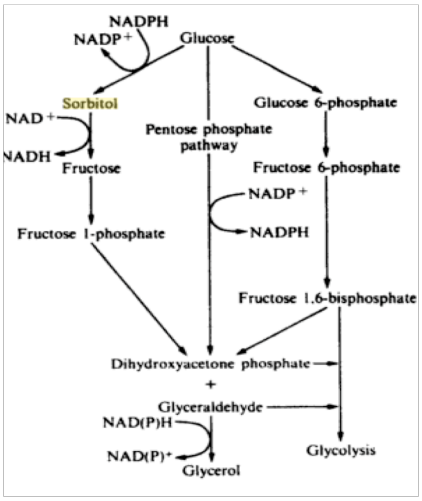

Figure 19. The sorbitol (glycitol) bypass and the conversion of glucose to dihydroxyacetone phosphate as it may apply to fungal metabolism (adapted from Jennings, 1984)

The EM pathway utilizes NAD as the electron acceptor to produce NADH, whereas HM glycolysis uses NADP as the electron acceptor to produce NADPH. ATP and pyruvate is produced as a result of the EM pathway, and the pyruvate formed is further converted to acetyl-CoA that is used in the citric acid cycle and fatty acid biosynthesis. The latter half of the HM pathway, known as the non-oxidative arm of the pentose phosphate pathway, is involved in the inter-conversion

of sugar-phosphate intermediates provides precursors for the synthesis of nucleotides, aromatic amino acids and sugar alcohols (Griffen, 1994).

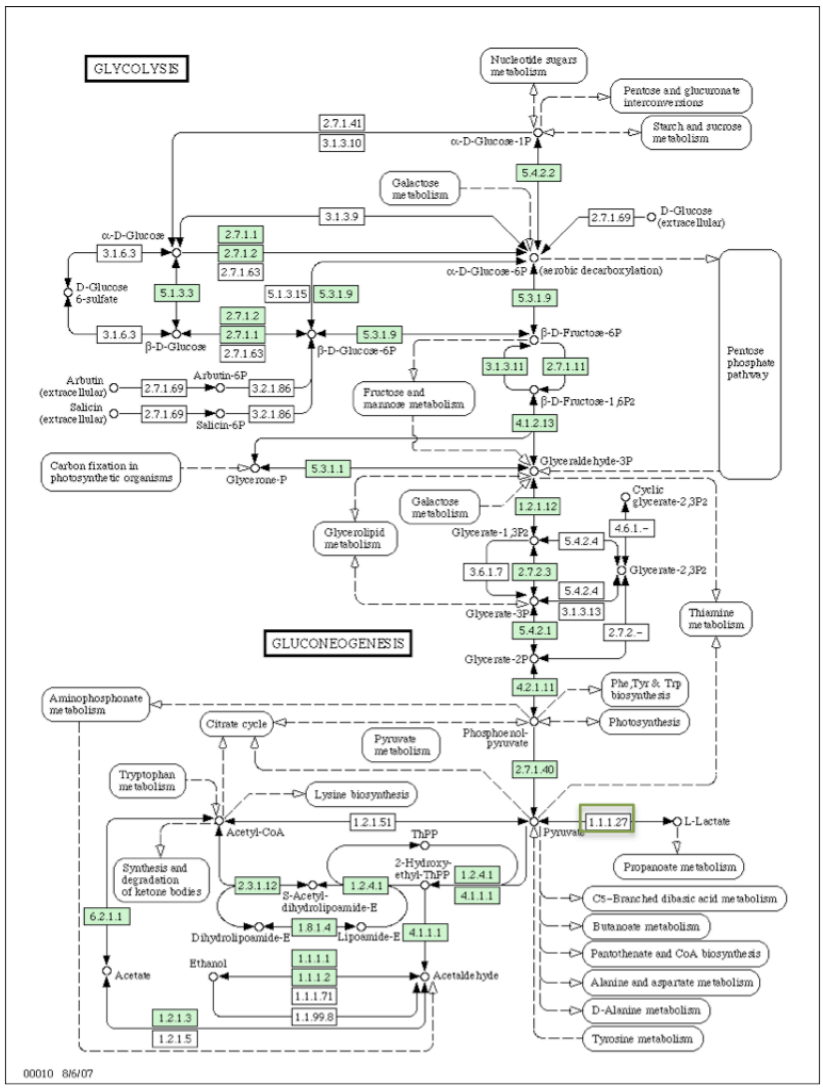

Figure 20. Predicted glycolytic and gluconeogenesis pathway in *P. omnivora*

Gluconeogenesis permits the utilization of non-carbohydrate energy sources, this occurs by the reversible activity of the enzymes of EM glycolysis with the exceptions of pyruvate kinase and 6-phosphofructokinase, which is overcome by the activity of phosphoenolpyruvate kinase involved in the conversion of oxaloacetate to phosphoenolpyruvate with the consumption of ATP and bisphosphatase

(hydrolyzes fructose bisphosphate to form fructose-6-phosphate).

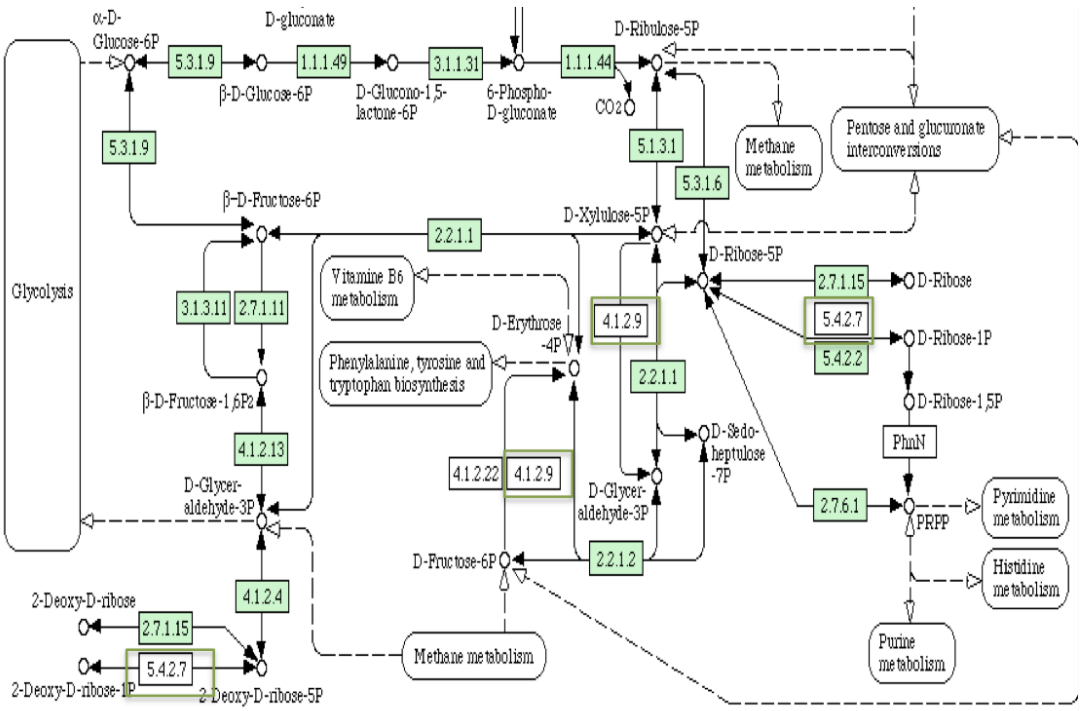

Figure 21. Predicted pentose phosphate pathway in *P. omnivora*

As observed in Figures 20 and 21 all of the enzymes involved in glycolysis via the EMP pathway, gluconeogenesis as well as the pentose phosphate pathway has been observed encoded in the *P. omnivora* genome along with aldo reductase (EC: 1.1.1.21) and sorbitol dehydrogenase (EC: 1.1.1.14) enzymes involved in the glycitol bypass. Glucose, fructose and mannose are phosphorylated by hexokinases (EC: 2.7.1.1), hexokinases as well as glucokinase (Glk1p; EC: 2.7.1.2) were found encoded in the *P. omnivora* genome. Unlike in the fungal genomes of *N. crassa*, *S. cerevisiae* and *S. pombe* enzymes involved in the ED pathway (Borkovich et al., 2004) were not found in our analysis implying that *P. omnivora* does not utilize the ED pathway for glycolysis.

Pyruvate produced as a result of glycolysis is oxidized to CO_2via the TCA cycle or converted to ethanol under anaerobic conditions by the enzyme alcohol dehydrogenases (EC: 1.1.1.1) encoded in the *P. omnivora* genome.

3.4.1.2.2 Tricarboxylic Acid (TCA) cycle and the glyoxylate shunt

The TCA cycle produces reducing equivalents (NADH and $FADH_2$) for the electron transport chain and provides anabolic precursors to different amino acid synthetic pathways. All of the enzymes involved in the TCA cycle were observed in the *P. omnivora* annotated proteins and is presented in Figure 22.

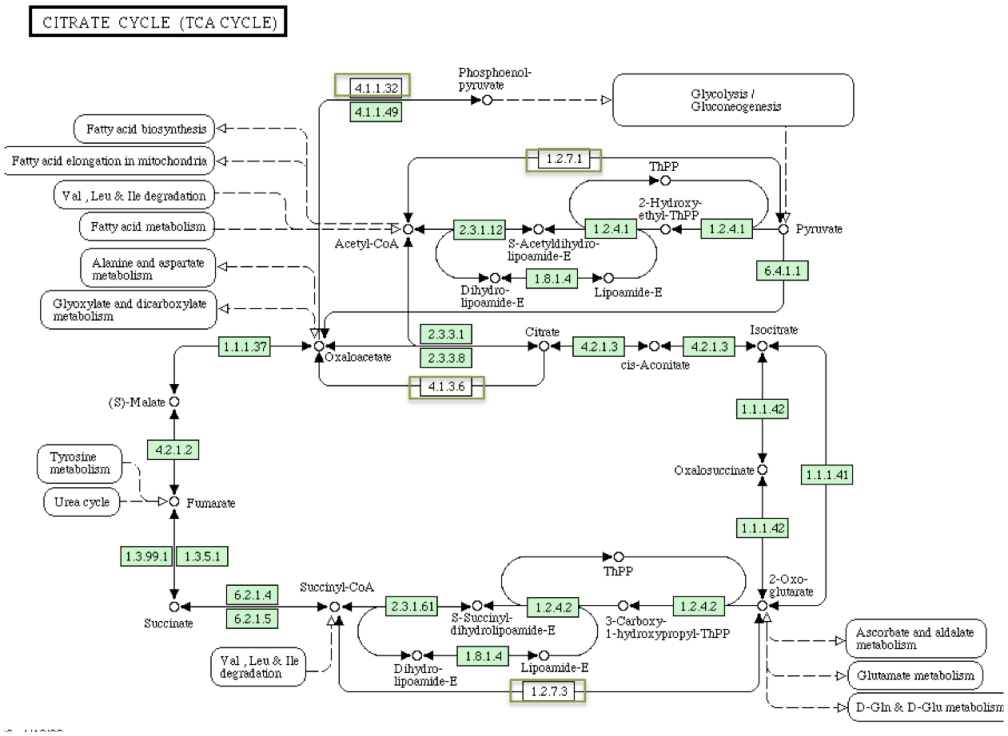

Figure 22. Predicted tricarboxylic acid cycle in *P. omnivora*

The enzymes involved in the gloxylate shunt (EC: 4.1.3.1) isocitrate lyase and (EC: 2.3.3.9) malate synthase have also been detected in the *P. omnivora* genome. The glyoxylate pathway is associated with fungal pathogenesis as it

enables growth on acetate or fatty acids as the sole carbon source and *M. grisea* mutants lacking isocitrate lyase lacked the ability to cause infection (Soloman et al., 2004).

3.4.1.2.3 Fructose, mannose and Galactose metabolism

The enzymes involved in fructose and mannose metabolism were encoded in the *P. omnivora* genome and include (EC:1.1.1.17) mannitol-1-phosphate 5-dehydrogenase involved in the inter-conversion of mannitol-1-phosphate to β-D fructose-6 phosphate, which is then converted by (EC:5.3.1.8) mannose-6-phosphate isomerase to mannose-6-phosphate. The enzyme (EC: 5.4.2.8) phosphomannomutase then converts mannose-6-phosphate to mannose-1-phosphate which is later converted to GDP-D-mannose by the action of (EC: 2.7.7.13) mannose-1-phosphate guanylyltransferase for the synthesis of N-glycans.

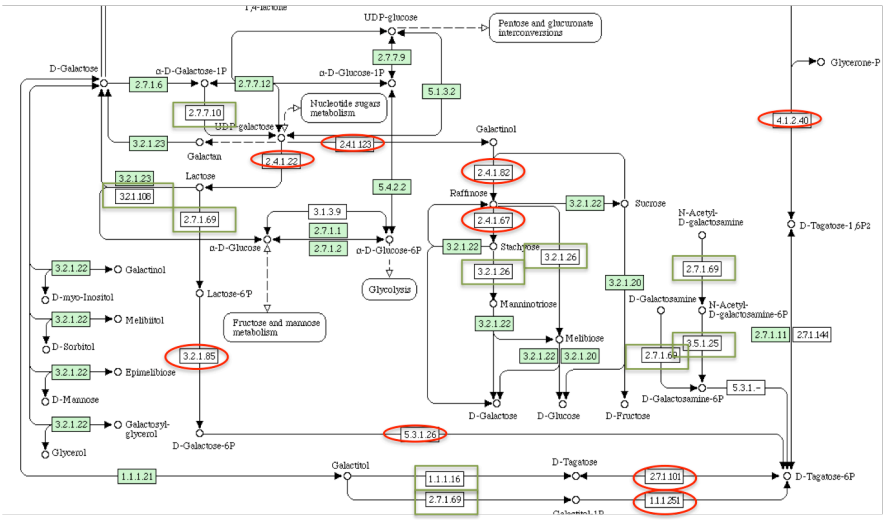

Figure 23. Predicted galactose metabolism in *P. omnivora*

The enzymes found in *P. omnivora* dedicated to galactose metabolism are detailed in Figure 23. UDP-galactose conversion to lactose and galacitol most likely

does not occur as *P. omnivora* lacks the genes for these enzymes (EC: 2.4.1.22 and EC: 2.4.123) involved in its conversion. It also lacks the enzymes for the synthesis of raffinose (EC: 2.4.1.82), stachyose (EC: 2.4.1.67), tagatose-6-phosphate conversions (EC: 2.7.1.101; EC: 1.1.1.251; EC: 3.2.1.85; EC: 5.3.1.26) and glycerone-phosphate synthesis (EC: 4.1.2.40).

3.4.1.2.4 Oxidative phosphorylation

The transfer of electrons from NADH to molecular oxygen occurs via four electron-transferring oligomers located in the inner membrane of the mitochondria. As a result of the transfer of electrons, protons are pumped across the membrane, and generates an electrochemical gradient for the synthesis of ATP by ATPase (Hatefi, 1985). All of the enzymes involved in oxidative phosphorylation via complexes I, II, III, IV and V were observed encoded in the *P. omnivora* genome as illustrated in Figure 24.

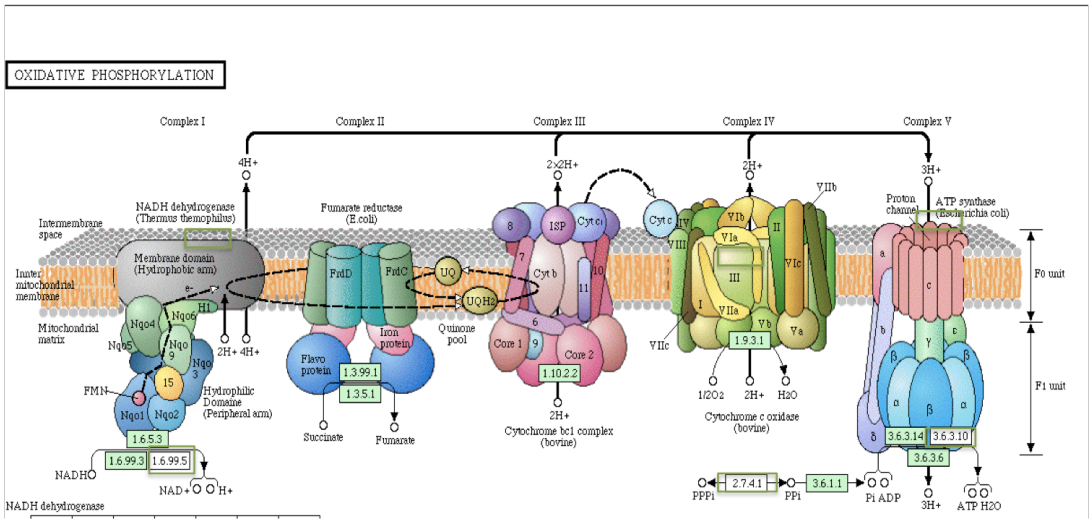

Figure 24. Predicted oxidative phosphorylation pathway in *P. omnivora*

In addition, the presence of the gene for formate dehydrogenase suggests that formate oxidation to carbon dioxide may occur under anaerobic conditions.

Lactate, obtained from anaerobic cleavage of pyruvate, also can donate electrons through lactate dehydrogenase to quinone. NADH generated from different metabolic pathways donates its electrons to NADH dehydrogenase which reduces quinone that passes the electron to succinate:ubiquinone oxidoreductase that, in turn passes the electron to the terminal electron acceptor.

3.4.1.2.5 Lipid metabolism

P. omnivora encodes the multifunctional enzyme complex FAS I (Fatty Acid Synthase), containing several globular domains that are involved in the enzymatic reactions for fatty acid synthesis. Malonyl-CoA (ACP) formation first is catalyzed by acetyl-CoA carboxylase, and acetyl-CoA and malonyl-CoA then are converted to fatty acid (such as palmitic acid (C_{16})), by a series of sequential reactions. The enzymatic steps of FAS involve decarboxylative condensation, reduction, dehydration and another reduction and result in a saturated acyl moiety, with two additional methylene groups at the end of the cycle.

The yeast FAS I complex contains two non-identical subunits (α and β) that form complexes ($\alpha_6\beta_6$) (Singh et al. 1985). ACP is associated with the α– subunit that additionally sustains β-ketoacyl synthase and β-ketoacyl reductase activities. The β-subunit is required for acetyl and malonyl transacylase, palmitoyl transacylase, dehydratase and enoyl reductase activities. Both NADPH and FMN act as cofactors for the activity of yeast enoyl-ACP reductase (Singh et al. 1985). Since all the enzymes needed for fatty acid synthesis and degradation through the β-oxidation pathway via acyl co-A dehydrogenase are encoded in the *P. omnivora* genome, under limiting oxaloacetate conditions, as shown in Figure 25, *P. omnivora*

seems to have the ability to synthesize and degrade ketone bodies.

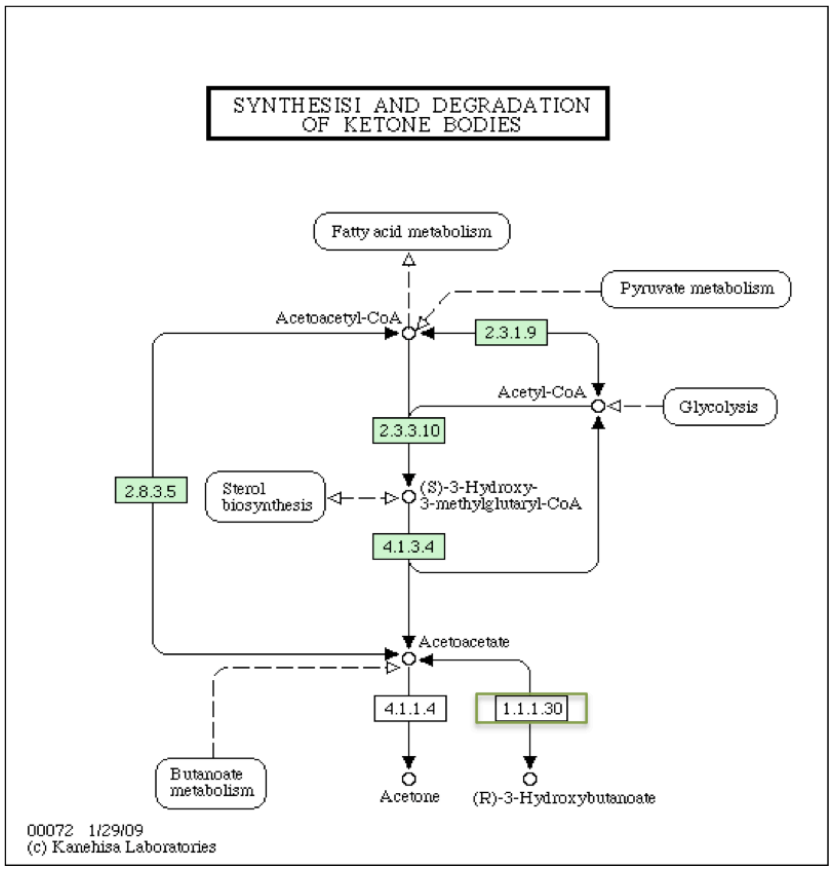

Figure 25. Predicted ketone synthesis and degradation pathway in *P. omnivora.*

Also, as shown in Figure 26, *P. omnivora* encodes all of the enzymes of the mevalonate pathway involved in ergosterol and cholesterol biosynthesis. In addition, the fungus can convert glycolate to glyoxylate (EC: 1.1.1.26), and cytochrome P450 (EC:1.14.14.1) and alkane monooxygenase (EC: 1.14.15.3) involved in the formation of dicarboxylic acids and all but two enzymes involved in propanoate metabolism (EC:1.1.1.59; EC1.2.1.18) are encoded in the genome. Since dicarboxylic acids are used as raw materials for the manufacturing of perfumes, polymers and adhesives, a closer look at the genes involved in dicarboxylic acid metabolism in *P. omnivora* may provide alternative sources for its synthesis.

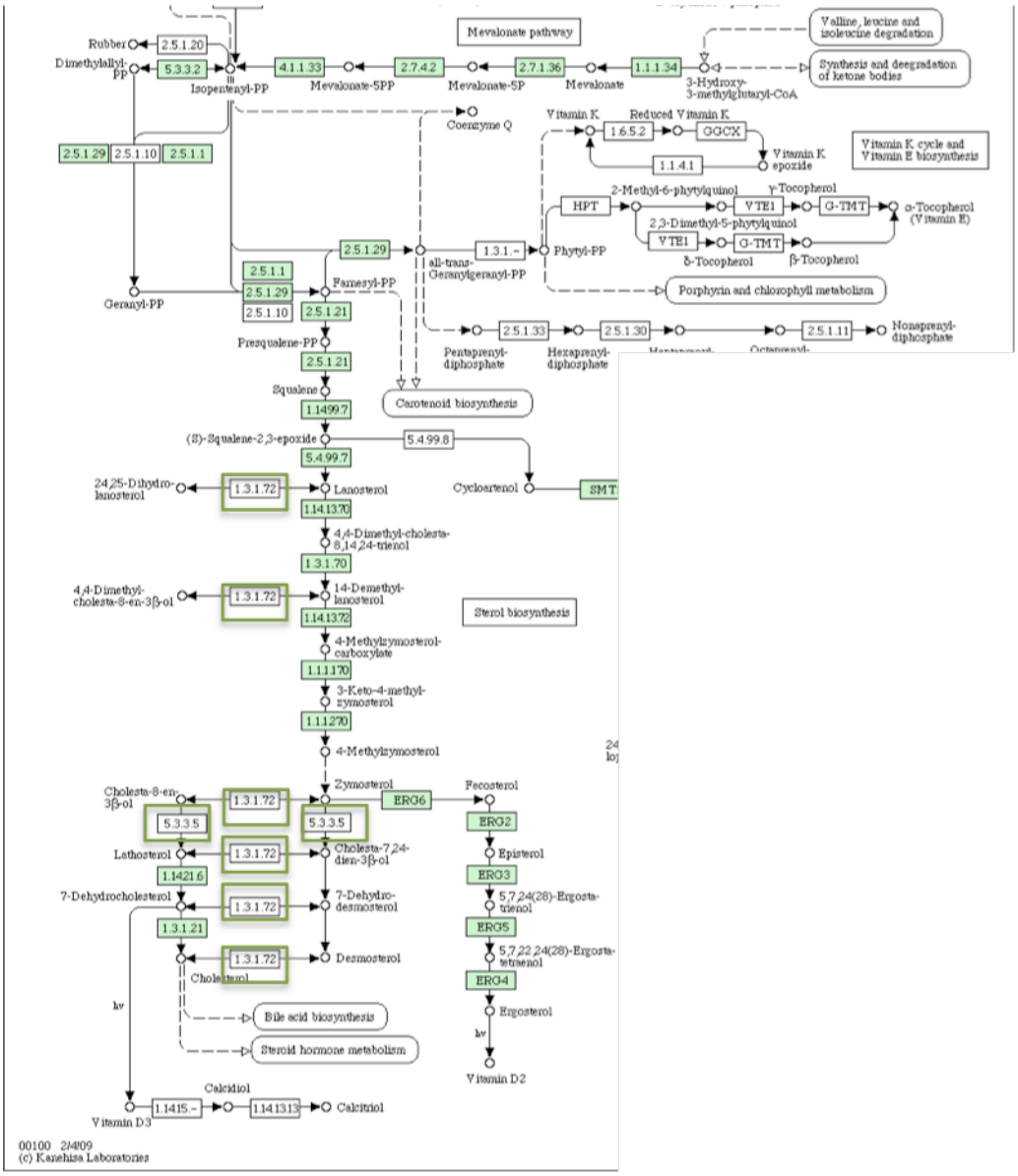

Figure 26. Predicted sterol biosynthesis pathway in *P. omnivora*

3.4.1.2.6 Amino acid biosynthesis

Amino acids contribute to protein synthesis as well as contribute to the nitrogenous bases of nucleic acids, coenzyme electron carriers. Following deamination they act as intermediates of the citric acid cycle and are oxidized to carbon dioxide and water or act as precursors in the synthesis of other biomolecules. Based on our analysis, *P.*

omnivora has the potential to synthesize all twenty amino acids.

In *P. omnivora,* alanine is most likey synthesized from pyruvate and aspartate from oxaloacetate by the transamination activity of alanine transaminase (EC: 6.1.1.7) and aspartate aminotransferase (EC: 2.6.1.1) respectively. Asparagine is synthesized from L-aspartate by asparagine synthase (EC:6.3.5.4) glutamate from α-ketoglutarate by glutamate dehydrogenase (EC:1.4.2.1), and glutamine is obtained from glutamate by the transamination activity of glutamine synthetase (EC:2.7.7.42).

The biosynthesis of tyrosine, tryptophan and phenylalanine occur by the condensation of phosphoenolpyruvate and D-erythrose-4-phosphate leading to the synthesis of chorismate, which acts as the precursor of prephanate for the synthesis of tyrosine and phenylalanine and anthranillate for the synthesis of tryptophan. All of the enzymes required for the biosynthesis of tyrosine, tryptophan and phenylalanine are encoded in the *P. omnivora* genome as shown in Figure 27.

With the exception of (EC: 3.5.4.19) phosphoribosyl-AMP cyclohydrolase, the genes for all the enzymes necessary for histidine biosynthesis were observed in the *P. omnivora* genome as is illustrated in Figure 28.

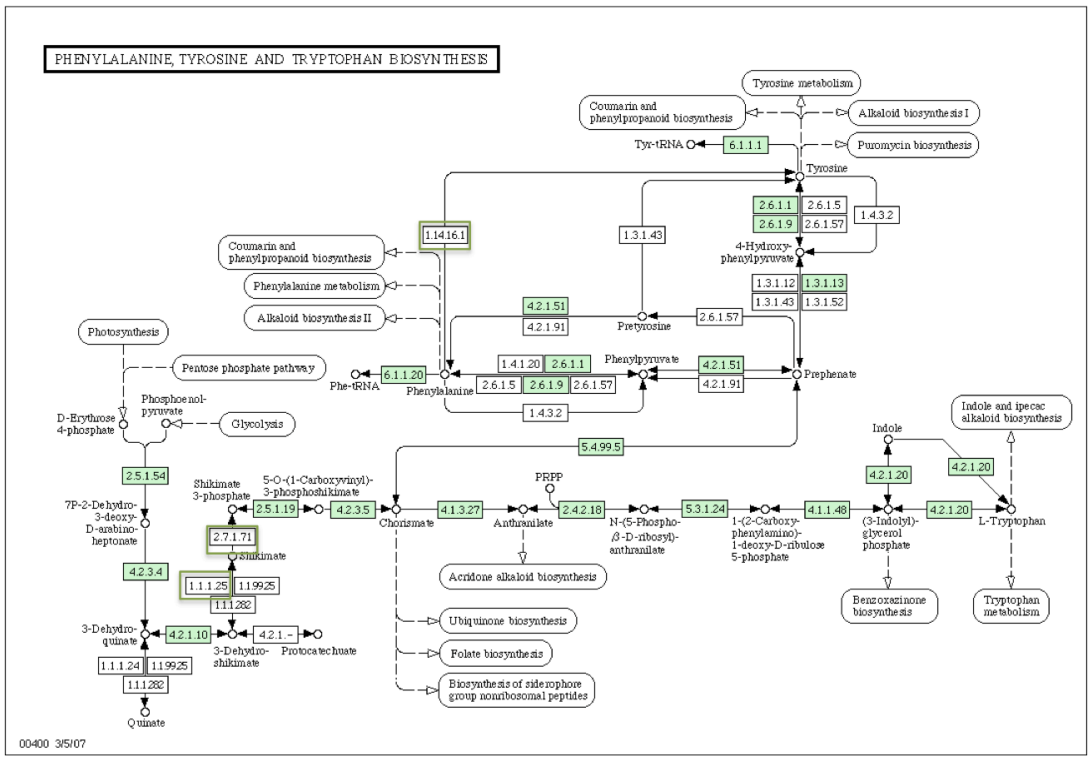

Figure 27. Predicted phenylalanine, tyrosine and tryptophan biosynthesis pathway in *P. omnivora*

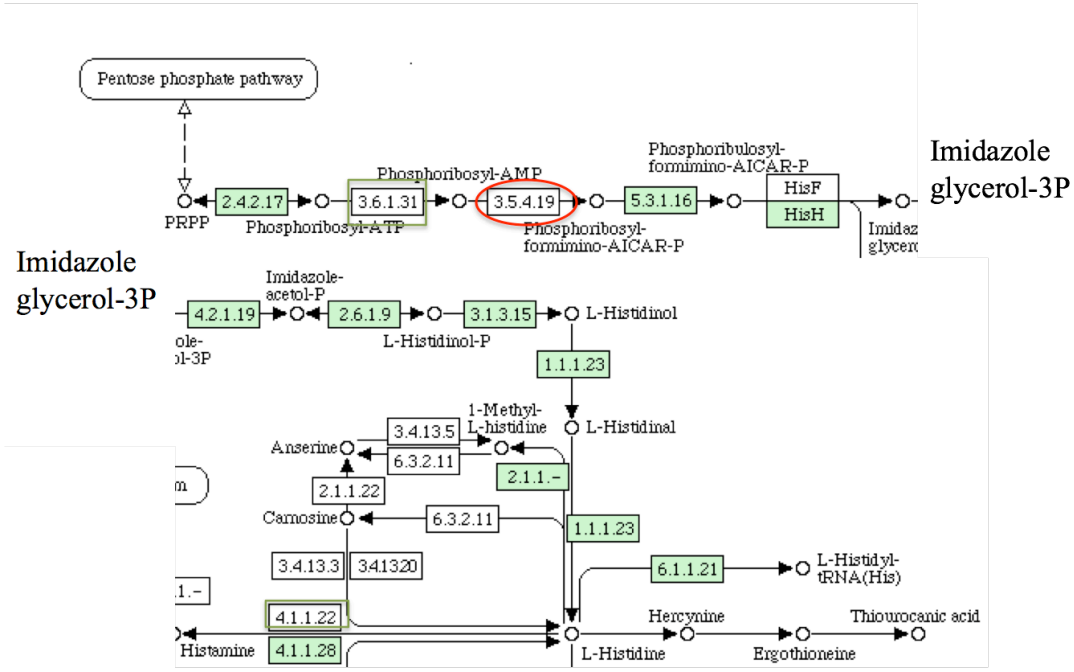

Figure 28. Predicted histidine biosynthesis pathway in *P. omnivora*

The genes for all of the enzymes involved in synthesis of branched chain amino acids: valine, leucine and isoleucine also were present in the *P. omnivora* genome as illustrated in Figure 29.

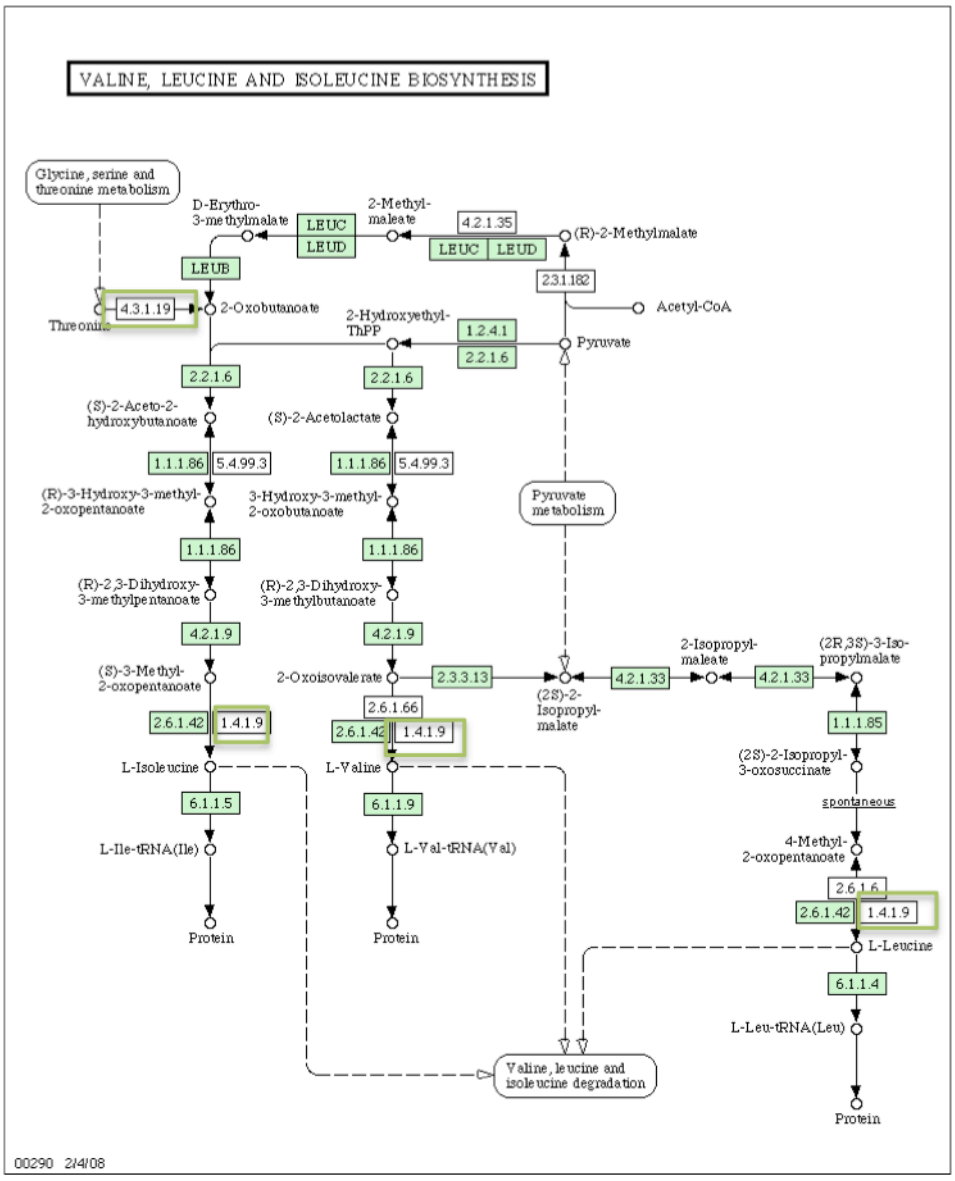

Figure 29. Predicted valine, leucine, isoleucine biosynthesis pathway in *P. omnivora*

Serine is synthesized from pyruvate and can be converted to cysteine and glycine, and threonine is synthesized from aspartate and can be converted to glycine. As illustrated in Figure 30, the genes encoding all of the enzymes necessary for serine, threonine and glycine biosynthesis were found in the *P. omnivora* genome.

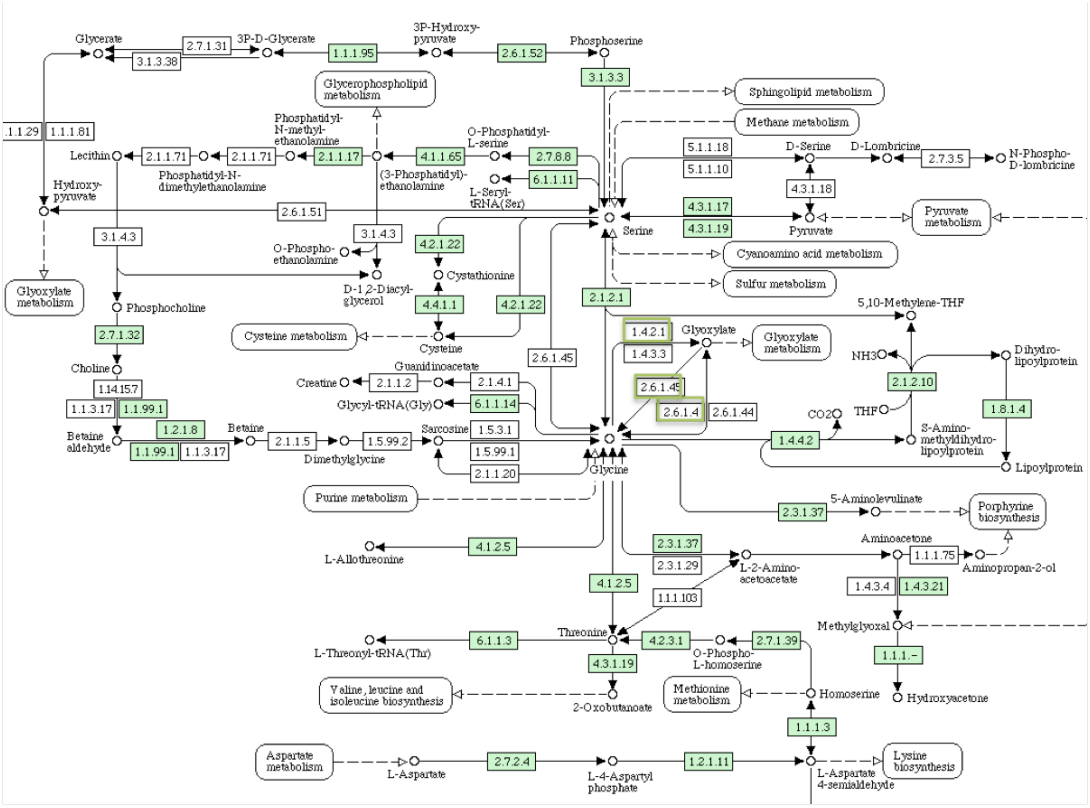

Figure 30. Predicted serine, threonine, glycine biosynthesis pathway in *P. omnivora*

Analysis of the sequence of the *P. omnivora* genome indicates that it contains the genes encoding proteins for synthesis of proline and arginine, and as shown in Figure 31, it also encodes all of the enzymes of the urea cycle: arginase, arginosuccinate lyase, arginosuccinate synthetase and ornithine carbamoyl transferase.

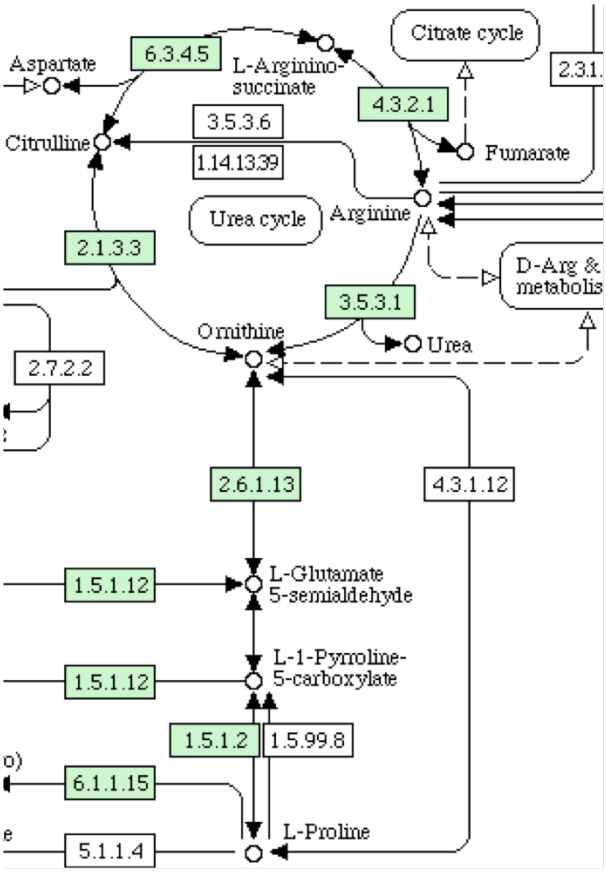

Figure 31. Predicted arginine and proline biosynthesis pathway in *P. omnivora.*

L-lysine is synthesized via the α- aminoadipate pathway from α-ketoglutarate in euglenoids and higher fungi and from aspartate via the diaminopimelate pathway in lower fungi, bacteria and plants (Zabriskie and Jackson, 2000). Interestingly, *P. omnivora* contains enzymes involved in lysine biosynthesis via both the alpha aminoadipate pathway and diaminopimelate pathway as shown in Figure 32.

P. omnivora encodes all of the enzymes involved in the alpha aminoadipate pathway, and all but one of the enzymes involved in the diaminopimelate pathway. The enzyme (EC:2.3.1.89; 2.3.1.117; 2.6.1.83) is involved in the conversion of the intermediate tetrahydrodipicolinate was not found in the present *P. omnivora*

assembly.

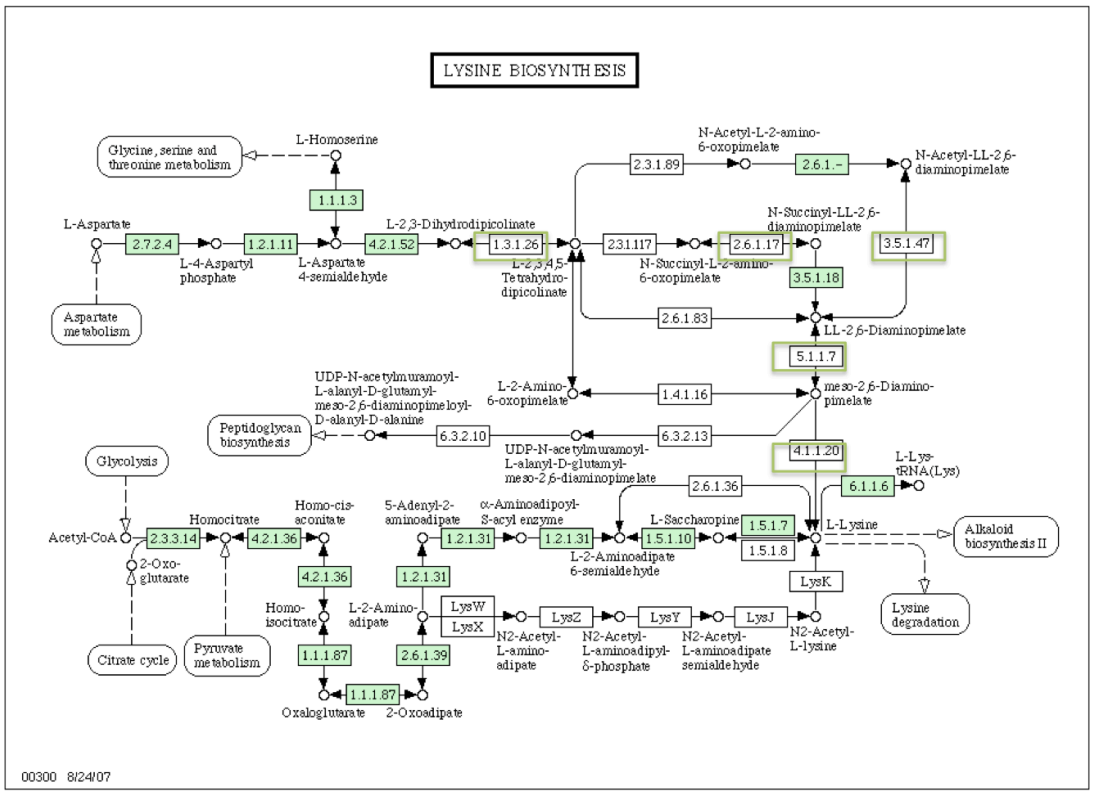

Figure 32. Predicted lysine biosynthesis pathway in *P. omnivora.*

Both cysteine and methionine are sulfur containing amino acids and for their synthesis, inorganic sulphate is reduced and incorporated into organic compounds by the assimilatory sulfate reduction pathways in most living organisms (Kopriva et al., 2002). Based on the annotated protein analysis, *P. omnivora* has the potential to synthesize cysteine from various sulfur sources. In this processes, arylsulfate is converted to sulfate by the action of arylsulfatase (EC: 3.1.6.1) which then is converted to APS (adenosine -5'-phosphosulfate) by (EC:2.7.7.4) ATP sulfurylase in the presence of ATP. PAPS (3'-phosphoadenosine-5'-phosphosulfate) formed by the phosphorylation of APS by adenylyl sulfate kinase (EC: 2.7.1.25) then is reduced to sulfite and further to sulfide by (EC: 1.8.4.8) 3'-phosphoadenosine-5'-

phosphosulfate reductase and (EC: 1.8.1.2) sulfite reductase respectively. Cysteic acid, taurine and alkylsulfonates are additional potential sources of sulfite in *P. omnivora* as they can be utilized by the encoded enzymes (EC: 4.1.1.15) cysteic acid decarboxylase, (EC: 1.14.11.17) taurine dioxygenase and (EC:1.14.14.5) alkanesulfonate monoxygenase respectively. Cysteine then is synthesized from sulfide and *O*-acetyl serine by cysteine synthase. Subsequently, cysteine and *O*-acetyl homoserine are transformed by cystathione γ-synthase (EC: 2.5.1.48) to form cystathionine, which is then acted upon by cystathione β-lyase (EC: 4.4.1.8) to generate homocysteine. 5-methyl tetrahydrofolate acts as the methyl donor for the synthesis of methionine from homocysteine by methionine synthase (EC: 2.1.1.13). Analysis of the *P. omnivora* genomic sequence reveals that it encodes genes for homologues of all the enzymes involved in the interconversion of methionine and cysteine.

3.4.1.2.7 Nucleotide biosynthesis

Analysis of the *P. omnivora* genomic sequence revealed the genes that encode all of the enzymes necessary for the *de novo* synthesis of purines and pyrimidines. *De novo* purine synthesis in *P. omnivora* begins with 5-phosphoribosyl-1-pyrophosphate (formed from the precursors ATP and ribose-5-phosphate formed in the pentose phosphate pathway) which is then converted to produce Adenine and Guanine as detailed in Figure 33. Adenine and Guanine are converted to deoxyadenine and deoxyguanine by the action of purine-nucleoside phosphorylase (EC: 2.4.2.1), and the deoxypurine monophosphates are then synthesized by AMP phosphohydrolase (EC:3.1.3.5).

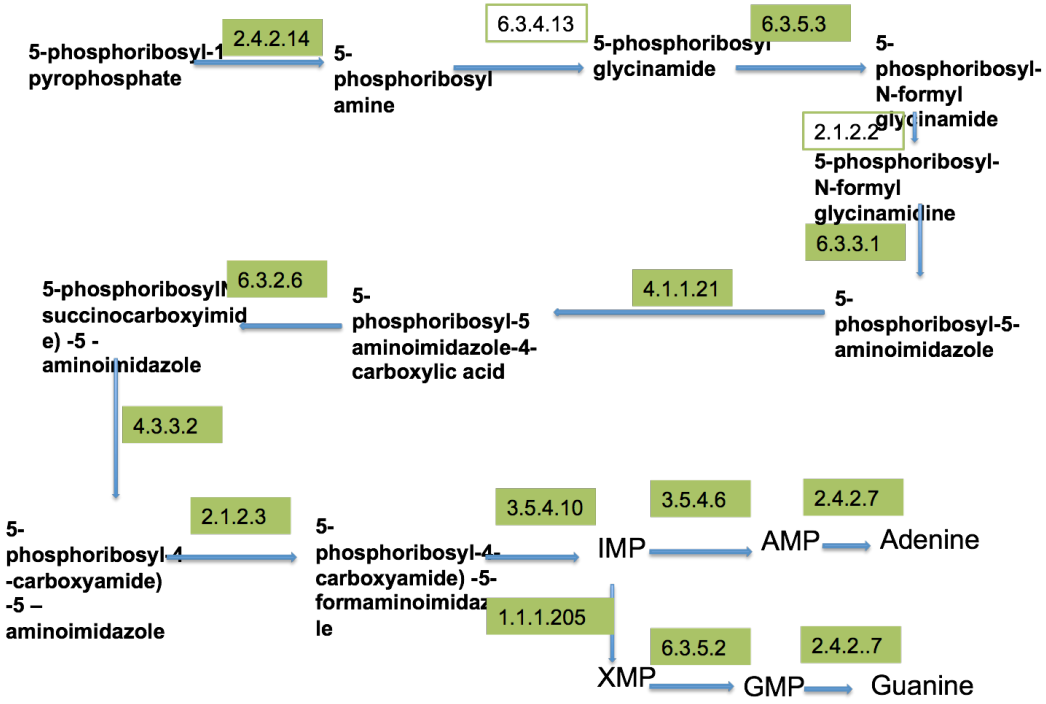

Figure 33. Predicted purine biosynthesis pathway in *P. omnivora*

The high energy required for synthesis of purines is overcome by the salvage pathway, here a free purine base that has been cleaved from a nucleotide can react with phosphoribosyl pyrophosphate to form the corresponding nucleotide. The genes for both enzymes-adenine phosphoribosyltransferase (EC: 2.4.2.7) and hypoxanthine-guanine phosphoribosyl transferase (EC:2.4.2.8) were identified in the *P. omnivora* genome indicating that it can utilize the salvage pathway for purine synthesis.

The *de novo* synthesis of pyrimidines begins with carbamoyl phosphate production and synthesis of the six membered pyrimidine ring that is attached to 5-phosphoribosyl-1-pyrophosphate. The genes encoding all of the enzymes necessary for pyrimidine biosynthesis were detected in the *P. omnivora* genome and is detailed

in Figure 34. Also, the genes encoding the enzymes of the pyrimidine salvage pathway that were identified in *P. omnivora* include (EC:2.7.1.48) uridine kinase and (EC: 2.7.1.21) thymidine kinase.

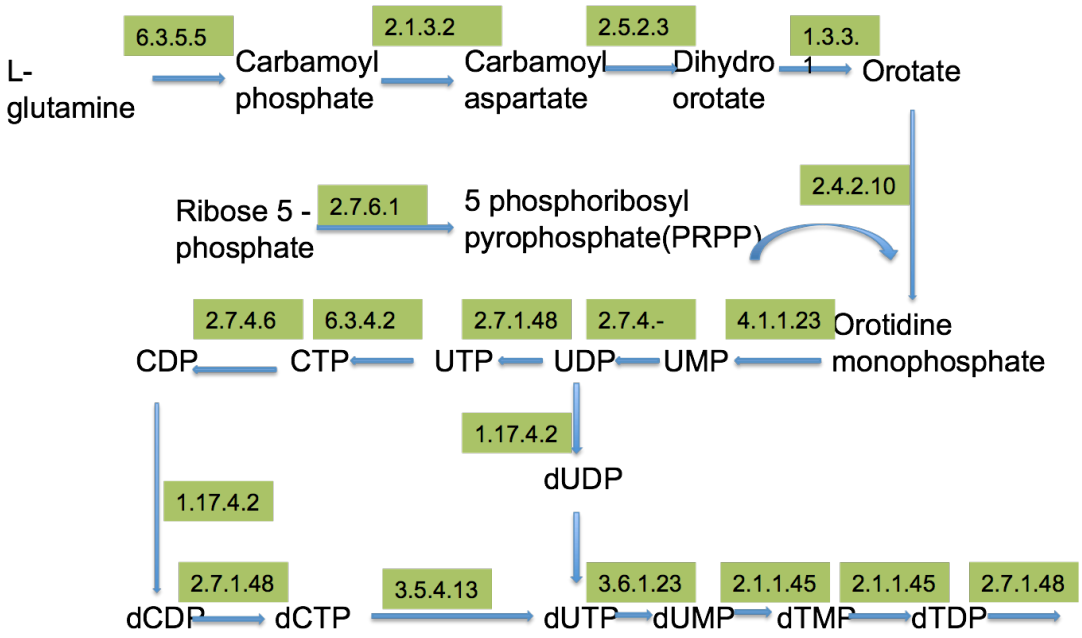

Figure 34. Predicted pyrimidine biosynthesis pathway in *P. omnivora*

3.4.1.2.8 Storage polysaccharides

The genes encoding the enzymes involved in the synthesis and breakdown of both glycogen and trehalose also were present in the *P. omnivora* genome, confirming that this fungi stores glycogen (Ergle and Blank, 1947). The glycogen synthesis pathway in *P. omnivora* thus involvers (EC:2.4.1.186) glycogenin glucosyltransferase and glycogen (starch) synthase (EC: 2.4.1.11), followed by the action of (EC: 2.4.1.18) 1,6-alpha-glucan branching enzyme. Trehalose synthesis

occurs by the action of trehalose-6-phosphate synthase which catalyzes the reaction between one molecule of glucose-6-phosphate and UDP-glucose, trehalose-6-phosphatase then cleaves the phosphate moiety to yield trehalose, which is acted upon by α, α-trehalase to yield two molecules of glucose (Jules et al., 2004).

P. omnivora genome also encodes all the enzymes necessary for the degradation of phenylacetonitrile and phenylacetamide to fumarate and the enzymes for the conversion of parathion to paraoxan and 4-nitrophenol to p-benzoquinone. It should be noted that phenylacetonitrile is considered toxic to humans and is used in the preparation of other chemicals (Kenneth Barbalace 2009).

3.4.1.3 Analysis of predicted transporters in the *P. omnivora* genome

P. omnivora transport facilitation proteins accounted for 12% of the proteins predicted from the genomic sequences with homology to GenBank genes with known functions. This agrees with the earlier report that filamentous fungi encode 25-30% more transporter systems than do the unicellular yeasts (Borkovich et al. 2004). By comparing predicted transporters in the unicellular *S. cerevisiae* with those in the filamentous fungi *P. omnivora, M. grisea* and *N. crassa*, we observed a greater number of major facilitator superfamily (MFS) and ATP binding cassette (ABC) family transporters in the filamentous fungi than in *S. cerevisiae* as shown below in Figure 35. We also observed that the rice blast pathogen *M. grisea* encodes 40%, 17% and 11% more predicted transporters than predicted in *S. cerevisae*, *N. crassa* and *P. omnivora*. In addition, *P. omnivora* encodes slightly higher numbers of ABC transporters and P-type ATPases than *M. grisea* but fewer numbers of MFS family proteins than found in both *M. grisea* and *N. crassa.*

Most of the *P. omnivora* predicted P-type ATPases identified were heavy metal and cation transporting ATPases involved in copper, arsenite and calcium transport and likely play a role in the maintaining viability of the fungus in calcareous heavy metal containing soils where it typically grows.

ABC transporters are well studied and are present from bacteria to humans and are involved in both influx and efflux mechanisms (Andrade et al., 2000). While bacteria use ABC transporters for both import and export, eukaryotes have been reported to use them mainly for export (Saurin et al., 1999) as most ABC transporters in plant pathogenic fungi are involved in secretion of toxins used in virulence, as well as for protection against phytoalexins and synthetic fungicides (De Waard, 1997). Recently the ABC type transporter Abc1 was shown to be required for host infection by *M. grisea* (Urban et al., 1999).

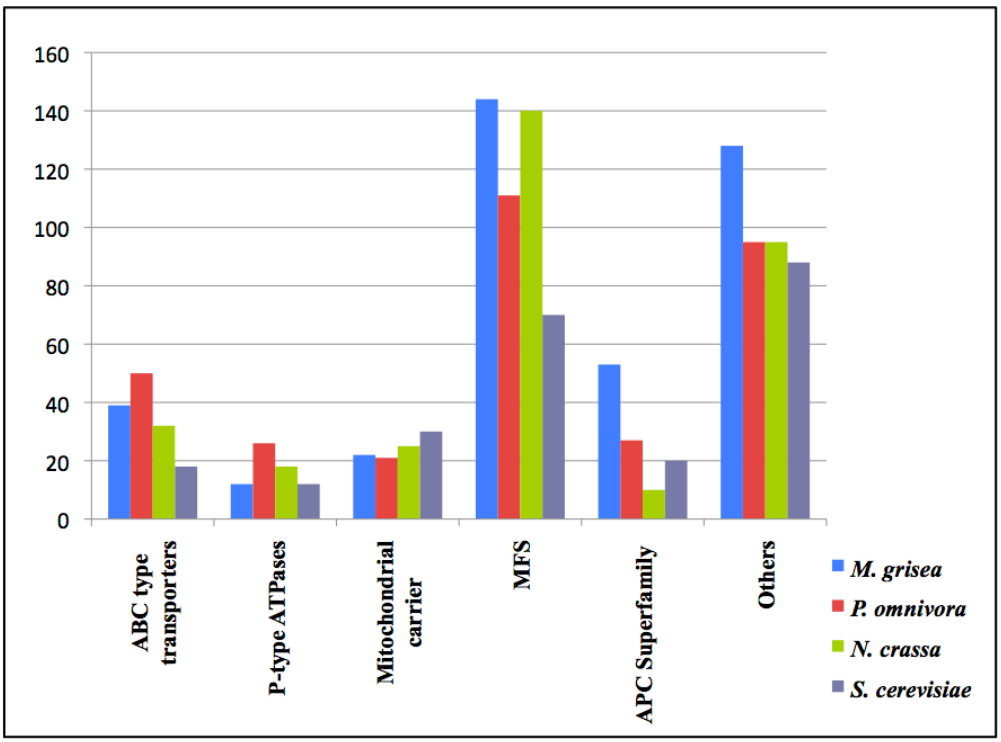

Figure 35. Relative numbers of *M. grisea, P. omnivora, N. crassa* and *S. cerevisiae* predicted transporters in various families.

The genes for twenty-eight predicted drug efflux proteins were identified in this study, among them a DHA14-like major facilitator superfamily –the Bcmfs1 multidrug transporter from *Bortrytis cinerea*, involved in protection against natural toxins and fungicides (Schoonbeck et al., 2001) also was found. Therefore, it is quite likely that the higher number of ABC transporters in *P. omnivora* contribute to the efflux of plant toxins and fungicides and aids in maintaining cell viability when these toxins are present.

3.4.1.4 Putative pathogenesis genes of *P. omnivora*

In the wake of functional genomics, several fungal genes have been studied

by tagged mutagenesis experiments for reduced or total loss of disease symptoms. Genes that when disrupted result in compromised disease symptoms and progression have been identified as pathogenic traits (Idnuram and Howlett, 2001). These include genes involved in the production of infection structures such as hydrophobin (mpg1) for appresorium in *M. grisea* (Talbot et al., 1983), melanin biosynthetic genes in *Cryptococcus neoformans* (Feng et al., 2001), cuticle and plant cell wall degrading enzymes such as cutinases and pectinases (Rogers et al., 2004; Enkerlii et al., 1999).

Two component signal transduction systems composed of sensor histidine kinases and response regulator domains often are involved in virulence in fungi and hence offer attractive targets for antifungal agents (Stock et al., 2000). In response to osmolarity, nutrients, oxygen levels, cellular redox status and light, the histidine kinase is autophosphorylated in an ATP-dependant manner on a conserved histidine residue and the phosphate then is transferred to the response regulator. Fungi posses a more complex version of the two-component phosphorelay pathway composed of hybrid sensors containing both histidine kinase and response regulator domains. The SLN1-YPD1-SSK1 phosphorelay system that regulates the high-osmolarity glycerol (HOG)1 response mitogen activated protein kinase (MAPK) cascade has been well-characterized in *S. cerevisiae* (Posas et al., 1996). Genome sequencing analyses revealed that *P. omnivora* has orthologous genes to all those of the high-osmolarity glycerol (HOG) response MAPK pathway of *S. cerevisiae.* Under low conditions of osmolarity, the Sln1p transmembrane histidine kinase autophosphorylates its histidine residue, this phosphate is transferred to its response regulator domain and

subsequently to the histidine residue of the histidine phosphotransfer domain protein Ypd1. The phosphate is finally transferred to the aspartate of the response regulator domain of Ssk1p, the phosphorylated Ssk1p is unable to activate the MAPKKKs Ssk2/22p, that activates MAPKK, Pbs2p which regulates the activity of the Hog1 MAPK (Hohmann, 2002; Posas et al., 1996).

Under increased osmolarity conditions, Sln1p is not phosphorylated and hence the phospho transfer to Ypd1 and Ssk1 does not occur, and Ssk1 in its unphosphorylated form activates the Hog1p, which modulates the transcription of glycerol synthesis in the cell.

Although, *P. omnivora* is not known to produce infection structures, the genes encoding the cuticle and cell wall degrading enzyme -cutinases, pectate lyases and endopolygalacturonases have been detected in the fungal genome. In addition pathogenicity genes involved in responding to the host environment such as pisatin demethylase (Wasmann and VanEtten, 1995), ABC transporters (Schoonbeek et al., 2001), neutral trehalase (Sweigard et al., 1988) alcohol oxidases (Sergers et al., 2001), also have been reported and our analysis of the *P. omnivora* genome reveals the genes for several pisatin demethylases, as well as ABC transporters, neutral trehalases and alcohol oxidases. Furthermore, several of the signal cascade components such as MAP kinases, cAMP dependant protein kinases, serine/threonine protein kinases, α and β subunits of the G proteins and other non classified pathogenicity genes such as the signal peptidase subunit (Thon et al., 2000) also are encoded in the *P. omnivora* genome.

P. omnivora proteins genes encoding proteins homologous to those involved

in disease and virulence present in the COGEME database included the superoxide generating NADPH oxidases, pathogenesis related Snod protein1, a structural toxin protein homologue, alcohol oxidase, a homologue of a *M. grisea* pathogenicity protein, and a CAP20-like protein involved in virulence in *Blumeria graminis* also were observed in the genome.

Implication of reactive oxygen species in *P. omnivora* morphogenesis indicates a pivotal role for antioxidant enzymes such as catalase and peroxidase in the survival of the fungus.

Functional genomic studies performing targeted gene disruptions of the above mentioned genes coupled with random gene disruption experiments is required to enforce the pathogenicity of these genes.

Chapter4

Conclusions

In this study, over 9,000,000 sequence reads and 300,000 ESTs were generated from the genomic DNA and six cDNA libraries of the fungal phytopathogen *P. omnivora*, respectively. Optimized automated next generation sequencing protocols, as well as an improved assembly and annotation scheme was developed to process the large number of genomic and individual EST reads from six different growth states and conditions. After predicting the encoded and expressed proteins, they were broadly classified into functional categories based on KEGG, KOG and COGEME annotation.

In the EST studies, stage specific gene expression was observed in the three distinct morphological stages-vegetative mycelia, sclerotia and spore mats of *P. omnivora* and on exposure of mycelia to different nutritional conditions. Majority of the annotated ESTs were involved in metabolism except for ESTs obtained from spore mats and mycelia exposed to non-host root exudates where majority of the ESTs are involved in information storage and processing. A comparative metabolic profile of the three morphological stages-vegetative mycelia, sclerotia and spore mat and on exposure of mycelia to different nutrient conditions indicated that the fungus uses carbohydrate metabolism in each stage of its life cycle. The highest proportion of ESTs involved in carbohydrate, energy, nucleotide, amino acids, glycan and cofactors observed in vegetative mycelia growing on M1078 medium was reflective of the actively propagating state of the fungus as it avails itself of nutrients available

in the medium. The sclerotial resting phase was defined by the highest proportion of ESTs involved in lipid metabolism and lowest in energy metabolism, nucleotide metabolism and glycan biosynthesis as compared to the other two life stages. The spore mats bearing conidia were metabolically active, although slightly less active than the other two stages in carbohydrate metabolism and displayed less than half the activity of vegetative mycelia in energy, amino acid and cofactor metabolism. A major proportion of the spore mats ESTs represent histones H3 and H4 involved in information storage and processing and is relevant to the condensation and packaging of DNA in the newly formed spores.

Mycelia deprived of either carbon or nitrogen in the medium respond by producing higher number of transcripts involved in carbohydrate, amino acid and nucleotide metabolism most likely by availing itself of endogenous carbon and nitrogen reserves. Analysis and comparison of ESTs obtained from mycelia exposed to host and non-host root exudate indicates that the pathogen is well adapted to utilize host root derived nutrients and has the potential to adapt to its non-host.

Interestingly, Tar 1p (Transcript antisense to ribosomal protein 1) and ART 3 (Antisense to ribosomal transcript 3) are both located on the antisense strand of nuclear encoded rDNA on chromosome XII of *S. cerevisiae* and high level expression of ESTs homologous to these two transcripts in the six *P. omnivora* cDNA library analyzed, along with subunits of the electron transport chain proteins indicates that the fungus is actively involved in mitochondrial respiration and the likely existence of rDNA transcription and mitochondrial function regulation.

Although ART 3 still remains uncharacterized, in wake of the recent functional studies conducted on Tar 1p in *S. cerevisiae* it can be concluded that high levels of the stringently controlled Tar 1p is indicative of active mitochondrial respiration and biogenesis and cellular oxidative stress when respiration is defective most likely in response to the changing cellular needs or energy demands under different types of growth conditions, such as during mitosis and or in aging.

Based on our analysis it also is evident that relatively higher proportions of Tar 1p expression occurs in the sclerotial and conidial stage as compared to the vegetative mycelia implying a likely role for reactive oxygen species in *P. omnivora* metamorphosis. This observation coupled with the report that artificial germination induction by sonication of conidia is only 60 % successful (Kings et al., 1931) further indicates that the fungus is compromised for energy metabolism in the presence of 2-3 nuclei present in conidia and that on exposure to moisture, metamorphosis by conidiation is the only alternative route to survival.

The draft sequence of the *P. omnivora* genome revealed a 74Mb assembled genome size of *P. omnivora*, approximately twice the size estimated by electrophoretic gel karyotyping of *P. omnivora* protoplasts. The assembled genome size and observation of diffuse banding patterns of putative *P. omnivora* chromosomes resolved by contour clamped homogenous electric field (CHEF) gel electrophoresis supports the hypothesis of Hosford and Gries which states that *P. omnivora* most likely exists as an obligate heterokaryon with the dependance on several heterokaryotic nuclei for maintenance of individual mycelial strand and sclerotium (Hosford and Gries, 1966).

A new method for the isolation, amplification and generation of 454 pyrosequencing based mixed paired-end library was developed during the course of this study and was tested to sequence the putative chromosomal bands of *P. omnivora* resolved by CHEF gel electrophoresis. This approach facilitated the ordering of genomic contigs simultaneously providing a platform for a future detailed study of each individual chromosome.

Approximately 9000 of the predicted 22,000 orfs encoded in the *P. omnivora* genome were homologous to proteins present in the GenBank non-redundant database at E-values less than 0.0001 and constitutes comparable number of *P. omnivora* proteins involved in metabolism and cellular processes as compared to the filamentous fungi *N. crassa* and *M.grisea*. The completeness of the metabolic pathways based on genomic and EST sequence data confirms that the vast majority, i.e. at least 80% and likely approaching over 90%, of the coding regions of the *P. omnivora*, were contained within the most recent draft of genomic sequence assembly.

High numbers of *P. omnivora* predicted copper, arsenite and calcium transporting P-type ATPases and ABC transporters were observed. These ATP-dependent proteins likely are involved in the survival of the pathogen in calcareous heavy metal containing soils and on exposure to plant toxins and fungicides. Predicted proteins homologous to a superoxide generating NADPH oxidase, a pathogenesis related Snod protein 1, a structural toxin protein homolog, an alcohol oxidase and a CAP20 like protein involved in disease and virulence, and the Bcmfs1 multidrug transporter involved in protection against natural toxins and fungicide

constitute potential *P. omnivora* pathogenicity genes were identified in this study that require further investigation including functional studies to define their role in the fungal life cycle and pathogenecity as well as the possibility of being a target for possible anti-fungal agents.

The high abundance of ESTs involved in mitochondrial respiration and relatively higher numbers of *P. omnivora* predicted proteins that are involved in energy metabolism, make it tempting to speculate that the levels of oxidative stress is linked with the burden of maintaining a heterokaryotic genome. It could be that the viability of the fungus in the presence of 2-3 nuclei as observed in hyphal tip cells and individual conidiospores is forfeited in the absence of adequate energy metabolism and its ability to thrive as a pathogen is endowed by its ability to combat oxidative stress in response to host induced reactive oxygen species. This area of research is yet to be explored while this proposition remains intriguing and remains aimed at answering fundamental question to the existence of multinucleate fungal cells.

The *P. omnivora* genome contains numerous endogenous fungal-specific viruses, many of which are yet to be characterized, as reflected by assignment of pfam motifs of viral proteins to 2500 of the predicted proteins that did not have GenBank homologs. This observation coupled with the high level expression of RNA dependant RNA polymerases of fungal mitoviruses in vegetative mycelial phase indicates a significant role for viruses in maintaining the flexibility of the *P. omnivora* genome and it's role in propagating mycelia.

Since approximately 75% of the ESTs analyzed in this study lacked

GenBank homologs, additional analysis is required to identify conserved domains or motifs needed to assign cellular functions. However, the assignment of pfam motifs at low stringency to 12,857 *P. omnivora* predicted proteins that lack GenBank homolog and examination of those ESTs that have high transcript abundance and are homologous to hypothetical proteins indicates that a vast majority of the *P. omnivora* encoded proteins remains to be explored by functional genomic studies.

The genomic sequence and the EST databases resulted from this dissertation research has revealed the gene content of this broad host range fungal phytopathogen. As this is the first study of a filamentous fungal genome that lacks both functional conidia and an active sexual cycle, it provides a peek into the life style and metabolic profile of fungi with a parasexual cycle.

References

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC.1991. Complementary DNA sequencing: expressed sequence tags and the human genome project. Science 252 (5013): 1651-1656

Agrios GN. 1997. Significance of plant diseases. In: Agrios GN editor. Plant Pathology 4th edition. San Diego: Academic Press. p. 24-34.

Aguirre J, Rios-Momberg M, Hewitt D, Hansberg W. 2005. Reactive oxygen species and development in microbial eukaryotes. Trends Microbiol. 13:111-118

Alex LA, Borkovich KA, Simon MI.1996. Hyphal development in *Neurospora crassa*: Involvement of a two-component histidine kinase. Proc. Natl. Acad. Sci. USA 93: 3416–3421.

Altschul S F, Gish W, Miller W, Myers E W, Lipman DJ.1990. Basic local alignment search tool. J. Mol. Biol. 215: 403-410

Altschul S F, Madden TL, Schaffer AA, Zhang J Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl. Acids Res. 25: 3389-3402

Andrade AC, Sorbo GD, Van Nistelrooy JGM, De Waard MA. 2000. The ABC transporter AtrB from *Aspergillus nidulans* mediates resistance to all major classes of fungicides and some natural toxic compounds. Microbiology 146: 1987-1997

Avery OT, MacLeod CM, McCarty M. 1944. Studies of the chemical nature of substance inducing transformation of pneumococcal types. Induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcal type

III. J. Exp. Med. 79: 137-158

Beadle GW, Tatum EL. Genetic Control of Biochemical Reactions in Neurospora. Proc. Natl. Acad. Sci. USA.1941 27(11) : 499-506

Beadle J, Wright M, McNeely L, Bennett JW. 2003. Electrophoretic karyotype analysis in fungi. Adv. Appl. Microbiol. 53: 243-270

Benson DA, Boguski M, Lipman DJ, Ostell J, 1996. GenBank. Nucleic Acids Res.24:1-5

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2009. GenBank. Nucleic Acids Res. 37: (Database Issue) D26-D31

Birnboim HC, Doly J.1979. A rapid alkaline extraction procedure for screeningrecombinant plasmid DNA. Nucl. Acid. Res. 7:1513-23

Bodenteich A, Chissoe S, Wang FY, Roe BA.1994. Shotgun cloning as the strategy of choice to generate templates for high throughput dideoxynucleotide sequencing. In Automated DNA sequencing and analysis techniques (ed. M.D. Adams, C. Fields and C. Venter), pp. 42-50. Academic Press, London, UK

Bonawitz ND, Chatenay-Lapointe M, Pan Y, Shadel GS. 2007. Reduced TOR signaling extends chronological lifespan via increased respiration and up-regulation of mitochondrial gene expression. Cell Metab. 5:265-277

Bonawitz ND, Chatenay-Lapointe M, Wearn CM, Shadel GS. 2008. Expression of the rDNA-encoded mitochondrial protein Tar 1p is stringently controlled and responds differentially to mitochondrial respiratory demand and dysfunction. Curr. Genet. 54(2): 83-94

Borkovich KA, Alex LA, Yarden O, Freitag M, Turner GE et al. 2004. Lessons from

the genome sequence of *Neurospora crassa*: tracing the path from genomic blueprint to multicellular organism. Microbiol Mol Biol Rev. 68:1-108

Bowyer P. 1999. Plant disease caused by fungi: Phytopathology. In: Oliver RP, Schweizer M. editors. Molecular Fungal Biology. Cambridge: Cambridge University Press. p. 294-321.

Braun BR, van Het Hoog M, d'Enfert C, Martchenko M, Dungan J, Kuo A, Inglis, DO, Uhl MA, Hogues H, Berriman M, et al. 2005. A human-curated annotation of the *Candida albicans* genome. PLoS Genet. 1: 36-57.

Braun EL, Halpern AL, Nelson MA, Natvig DO. 2000. Large-scale comparison of fungal sequence information: mechanisms of innovation in *Neurospora crassa* and gene loss in *Saccharomyces cerevisiae*. Genome Res.10: 416–430. doi: 10.1101/gr.10.4.416

Catlett NL, Yoder OC, Turgeon BG. 2003. Whole genome analysis of two-component signal transduction genes in fungal pathogens. Eukaryotic cell 2(6):1151-1161

Cheung F, Haas BJ, Goldberg SM, May GD, Xiao Y, Town CD. 2006. Sequencing *Medicago truncatula* expressed sequenced tags using 454 Life Sciences technology BMC Genomics 7: 272 .doi:10.1186/1471-2164-7-272

Chissoe SL, Wang YF, Clifton SW, Ma NT, Sun HJ, Lobsinger JS, Kenton SM, White JD, Roe BA.1991. Strategies for rapid and accurate DNA sequencing. Methods: a Companion to Meth. Enzymol. 3: 55-65

Cho E-M, Liu L, Farmerie W, Keyhani NO. 2006. EST analysis of cDNA libraries from the entomopathogenic fungus Beauveria (Cordyceps) bassiana. I. Evidence for

stage-specific gene expression in aerial conidia, in vitro blastospores and submerged conidia. Microbiology 152: 2843-2854

Coelho PS, Bryan AC, Kumar A, Shadel GS, Snyder M. 2002. A novel mitochondrial protein, Tar 1p, is encoded on the antisense strand of the nuclear 25S rDNA. Genes Dev. 16:2755-2760

Crick F.1970. Central Dogma of Molecular Biology. Nature 227:561-563

Dahm R. 2005. Friedrich Miescher and the discovery of DNA. Dev. Biol. 278:274-288

Damicone JP, Pratt PW, Conway KE . 2003. Phymatotrichum root rot. Oklahoma Cooperative Extension Service Fact Sheet. F-7621

Davis RH, Perkins DD. 2002. Neurospora: a model of model microbes. Nature Rev. Genet. 3: 7-13.

De Waard MA. 1997. Significance of ABC transporters in fungicide sensitivity and resistance. Pestic. Sci. 51:271-275

Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu JR, Pan H, et al. 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature. 434: 980-986. doi: 10.1038/nature03449.

Deininger PL. 1983. Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. Anal. Biochem. 129: 216–223

Diener SE, Dunn-Coleman N, Foreman P, Houfek TD, Teunissen PJM, van Solingen P, Dankmeyer L, Mitchell TK, Ward M, Dean RA. 2004. Characterization of the protein processing and secretion pathways in a comprehensive set of expressed sequence tags from *Trichoderma reesei*. FEMS Microbiol. Lett. 230 (2): 275-282

Dunlap AA. 1941. A convenient soil culture method for obtaining sclerotia of the cotton root rot fungus. Am J Bot. 28: 945-947.

Eaton FM, Rigler NE.1946. Influence of carbohydrate levels and rootsurface microfloras on phymatotrichum root rot in cotton and maize plants. J Agric Res 72: 137-161.

Ebbole DJ, Jin Y, Thon M, Pan H, Bhattarai E, Thomas T, Dean R. 2004. Gene discovery and gene expression in the rice blast fungus, *Magnaporthe grisea*: analysis of expressed sequence tags. Mol Plant Microbe Interact.17:1337-1347.

Enkerli J, Felix G, Boller T. 1999. The enzymatic activity of fungal xylanase is not necessary for its elicitor activity. Plant Physiol. 121:391-397

Ergle DM, Blank LM. 1947.A chemical study of the mycelium and sclerotia of *Phymatotrichum omnivorum*. Phytopathol 37:153-161

Feng B, Wang X, Hauser M, Kaufmann S, Jentsch S, Haase G, Becker JM, Szaniszla PJ. 2001. Molecular cloning and characterization of *WdPKS1*, a gene involved in dihydroxynaphthalene melanin biosynthesis and virulence in *Wangiella dermatitidis*. Infect. Immun 69: 1781-1794.

Finn RD, Tate J, Mistry J, Coggil PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A. 2008. The Pfam protein families database. Nucleic Acids Res. 36:D281–D288

Fitzgerald MC, Skowron P, Van Etten JL, Smith LM, Mead DA .1992. Rapid shotgun cloning utilizing the two base recognition endonuclease *CviJI*. Nucleic Acids Res. 20, 3753–62.

Galagan JE, Henn MR, Ma LJ, Cuoma CA, Birren B. 2005. Genomics of the fungal

kingdom: Insights into eukaryotic biology. Genome Res. 15:1620-1631

Georgiou CD, Patsoukis N, Papapostolou L, Zervoudakis G. 2006. Sclerotial metamorphosis in filamentous fungi is induced by oxidative stress. Integrative and Comparative Biology 46: 691-712

Gharizadeh B, Zheng B, Akhras M, Ghaderi M, Jejelowo O, Strander B, Nyren P, Wallin K L, Pourmand N. 2006. Sentinel-base DNA genotyping using multiple sequencing primers for high-risk human papilloma viruses; Mol. Cell Probes 20: 230-238

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG. 1996. Life with 6000 genes. Science. 274: 546, 563–7. doi: 10.1126/science.274.5287.546

Gordon D, Abajian C, Green P.1998. Basecalling of automated sequencer traces using phred.II. Error probabilities. Genome Res. 8:186-194

Green P. (Copyright 1994-1996) Phrap Documentation

Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K, Leitch IJ, Murray BG, Kapraun DF, Greilhuber J, Bennett MD. 2007. Eukaryotic genome size databases.

Griffen, DH. 1994. Primary metabolism. In: Griffen DH, editor. Fungal Physiology 2nd edition, New York: Wiley-Liss, Inc. p 215-245

Griffith F.1928. The significance of pneumococcal types. J. Hyg. 27: 113-159

Harris TD, Buzby PR, Babcock H, Beer P, Bowers J, Braslavsky I, Causey M, Colonnel J, Dimio J et al. 2008. Single-molecule DNA sequencing of a viral genome. Science 320: 106–109

Hatefi Y. 1985. The mitochondrial electron transport and oxidative phosphorylation system. Annu. Rev. Biochem. 54:1015-1069

Hawkins TL, O'Connor-Morin T, Roy A, Santillan C. 1994. DNA purification and isolation using a solid-phase. Nucleic Acids Res. 22(21): 4543-4544

Hawksworth DL. 1991. The fungal dimension of biodiversity: mag-nitude, significance, and conservation. Mycol Res. 6:641-55.

Hershey AD, Chase M.1952. Independent function of viral proteins and nucleic acid growth of bacteriophage. J. Gen. Physiol. 36:39-56

Hijri M, Sanders JR. 2005. Low gene copy number shows that arbuscular mycorrhizal fungi inherit genetically different nuclei Nature 433: 160-163

Hong EL, Balakrishnan R, Dong Q, Christie KR, Park J, Binkley G, Costanzo MC, et al. 2008. Gene Ontology annotations at SGD: new data sources and annotation methods. Nucleic Acids Res. 36: D577-81.

Hohmann S. 2002. Osmotic stress signaling and osmoadaptation in yeasts. Microbiol Mol Biol Rev 66:3 00-372

Hosford RM Jr, Gries GA. 1966. The nuclei and parasexuality in *Phymatotrichum omnivorum.* American Journal of Botany*,* 53: 570-79

Hunkapiller T, Kaiser RJ, Koop BF, Hood L. 1991.Large-scale and automated DNA sequence determination. Science. 254 (5028): 59-67

Hynes MJ. 2003. The *Neurospora crassa* genome opens up the world of filamentous fungi. Genome Biol. 4(6): 217.

Idnuram A, Howlett BJ. 2001. Pathogenicity genes of phytopathogenic fungi. Molecular Plant Pathol. 2: 241-255

Jeffrey J, Jornwall H. 1988. Sorbitol dehydrogenase. Recent Adv. Enzymol. 61:47-106

Jennings DH.1995. Carbon. In: Jennings DH, editor. The Physiology of fungal nutrition. Cambridge: Cambridge University Press p87-194

Jones T, Federspiel NA, Chibana H, Dungan J, Kalman S, et al. 2004. The diploid genome sequence of *Candida albicans*. Proc. Natl. Acad. Sci. USA 101: 7329-7334.

Ju J, Kim DH, Bi L, Meng Q, Bai X, Li Z, Li X, Marma MS, Shi S, Wu J, Edwards JR, Romu A, Turro NJ. 2006. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. Proc. Natl. Acad. Sci. USA 103 (52): 19635-19640

Jules M, Guillou V, Francois J, Parrou JL. 2004. Two distinct pathways for trehalose assimilation in the yeast *Saccharomyces cerevisiae.* Appl. Environ. Microbiol. 70:2771-2778

Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. 2006. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. 34: D354–D357

Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res. 32: D277–D280

Karlin S, Altschul S F.1990. Method for assessing the statistical significance of molecular sequence features by using general scoring schemes. Proc. Natl. Acad. Sci. USA 87: 2264-2268

Kastenmayer JP, Ni L, Chu A. et al. 2006. Functional genomics of genes with small open reading frames (sORFs) in S. cerevisiae. Genome Res.16: 365–373.

Kenneth Barbalace. Chemical Database - Phenylacetonitrile. EnvironmentalChemistry.com. 1995 - 2009. Accessed on-line: 4/7/2009 http://EnvironmentalChemistry.com/yogi/chemicals/cn/Phenylacetonitrile.html).

King CJ, Loomis HF, Hope C. 1931. Studies on sclerotia and mycelial strands of the cotton root-rot fungus. J Agric Res 42: 827-840.

Kopriva S, Suter M, von Ballmoos P, Hesse H, Krahenbuhl U, Rennenberg H, Brunold C. 2002. Interaction of sulfate assimilation with carbon and nitrogen metabolism in Lemma minor. Plant Physiol. 130: 1406-1413

Korlach J, Marks PJ, Cicero RL, Gray JJ, Murphy DL, Roitman DB, Pham TT, Otto GA, Foquet M, Turner SW. 2008. Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. Proc. Natl. Acad. Sci. USA 105: 1176-1181

Kozian DH, Kirschbaum BJ. 1999. Comparative gene-expression analysis. Trends Biotechnol. 17:73-78

Kupfer DM, Drabenstot SD, Buchanan KL, Lai H, Zhu H, Dyer DW, Roe BA, Murphy WJ. 2004. Introns and splicing elements of five diverse fungi. Eukaryotic Cell 3:1088–1100

Lev S, Sharon A, Hadar R, Ma H, Horwitz BA. 1999. A mitogen-activated protein kinase of the corn leaf pathogen Cochliobolus heterostrophus is involved in conidiation, appresorium formation, and pathogenicity: diverse roles for mitogen-activated protein kinase homologs in foliar pathogens. Proc. Natl. Acad. Sci. USA 96: 13542-13547

Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG, Webb WW. 2003.

Zero-mode waveguides for single-molecule analysis at high concentrations. Science 299: 682-686

Lowe, T.M., Eddy, S.R. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25: 955-964.

Lowy B. 1971. New records of mushroom stones from Guatemala. Mycologia 63: 983-999.

Lyda SD .1978. Ecology of *Phymatotrichum omnivorum*. Ann Rev Phytopathol. 16: 193-209.

Madden LV, Wheelis MA. 2003. The threat of plant pathogens as weapons against U.S. crops. Annual Review of Phytopathol. 41: 155-176.

Mardis ER, Roe BA. 1989. Automated methods for single-stranded DNA isolation and dideoxynucleotide DNA sequencing reactions on a robotic workstation. BioTechniques 7: 840-850

Marek SM, Hansen K, Romanish M, Thorn RG. 2009. Molecular systematics of the cotton root rot pathogen, *Phymatotrichopsis omnivora.* Persoonia 22: 63–74

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. 2005. Genome

sequencing in microfabricated high-density picolitre reactors. Nature. 437(7057): 376-80.

Moore D. 1998. Fungal morphogenesis. Cambridge, UK: Cambridge University Press.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M.2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35: W182–W185

Moss MO. 1987. Microbial food poisoning. In Essays in Agricultural and Food Microbiology Norris, J.R. Pettipher, G.L. pp. 369-399. Chichester : John Wiley and Sons.

Ng P, Tan JJ, Ooi HS, Lee YL, Chiu KP, Fullwood MJ, Srinivasan KG, Perbost C, Du L, Sung WK, Wei CL, Ruan Y .2006. Multiplex sequencing of paired-end ditags (MS-PET): a strategy for the ultra-high-throughput analysis of transcriptomes and genomes. Nucleic Acids Res. 34:e84

Nucleic Acids Res. 35: D332-D338

Oefner PJ, Hunicke-Smith SP, Chiang L, Dietrich F, Mulligan J, Davis RW. 1996. Efficient random subcloning of DNA sheared in a recirculating point-sink flow system. Nucleic Acids Res. 24:3879-3886

Orbach, M.J., F.G. Chumley, and B. Valent. 1996. Electrophoretic karyotypes of *Magnaporthe grisea* pathogens of diverse grasses. Molecular Plant-Microbe Interact 9: 261-271.

Phadnis SH, Huang HV, Berg DE. 1989. Tn5supF, a 264-base-pair transposon derived from Tn5 for insertion mutagenesis and sequencing DNAs cloned in phage

lambda. Proc. Natl. Acad. Sci. USA 86: 5908-5912

Pheiffer BH, Zimmerman SB. 1983. Polymer-Stimulated Ligation: Enhanced Blunt-End or Cohesive-End Ligation of DNA or Deoxyribooligonucleotides by T4-DNA Ligase in Polymer-Solutions. Nucleic Acids Res. 11:7853-7871

Posas F, Wurgler-Murphy SM, Maeda T, Witten EA, Thai TC, Saito H. 1996. Yeast HOG1 MAP kinase cascade is regulated by multistep phosphorelay mechanism in the SLN1-YPD1-SSK1 “two-component” osmosensor. Cell 86: 865-875

Prade R A, Ayoubi P, Krishnan S, Macwana S, Russel H. 2001. Accumulation of stress and cell wall degrading enzymes associated transcripts during asexual development in *Aspergillus nidulans*. Genetics 157:957-967

Roche Diagnostics, personal communication. Karow J. 2009. 454 targets longer reads, better *de novo* assembly; cites clinical apps of amplicon sequencing. In Sequence. Available from http://www.genomeweb.com/sequencing/454-develops-longer-reads-better-de-novo-assembly-cites-clinical-apps-amplicon-s

Roe B. 2004. Shotgun Library Construction for DNA Sequencing in Methods in Molecular Biology, vol 255: Bacterial Artificial Chromosomes, Volume 1: Library Construction, Physical Mapping, and Sequencing. S. Zhao and M. Stodolsky, eds., Humana Press Inc, Totowa, New Jersey pp. 171-187.

Rogers LM, Flaishman MA, Kolattukudy PE. 1994. Cutinase gene disruption in *Fusarium solani* f sp *pisi* decreases its virulence on pea. Plant cell. 6: 935-945

Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P.1996. Real-time DNA sequencing using detection of pyrophosphate release. Anal. Biochem. 242: 84-89

Sanger F, Nicklen S, Coulsen AR. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467

Saurin W, Hofnung M, Dassa E. 1999. Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. J. Mol. Evol. 48:22-41

Schoonbeek H, Del Sorbo G, De Waard MA. 2001. The ABC transporter BcatrB affects the sensitivity of Bortrytis cinerea to the phytoalexin reservatrol and the fungicide fenpiclonil. Mol Plant Microbe Interact 14:562-571

Segers G, Bradshaw N, Archer D, Blissett K, Oliver RP. 2001. Alcohol oxidase is a novel pathogenicity factor for *Cladosporium fulvum*, but aldehyde dehydrogenase is dispensible. Mol. Plant-Microbe Interact. 14: 367-377

Selmecki A, Forche A, Berman J. 2006. Aneuploidy and isochromosome formation in drug-resistant *Candida albicans*. Science 313: 367–370

Seringhaus M, Paccanaro A, Borneman A, Snyder M, Gerstein M. 2006. Predicting essential genes in fungal genomes. Genome Res. 16: 1126-1135

Shendure J, Porreca, GJ, Reppas, NB, Lin, X, McCutcheon, JP, Rosenbaum, AM, Wang, MD , Zhang, K, Mitra, RD, Church, GM. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309: 1728-1732

Sims AH, Gent ME, Robson GD, Dunn-Coleman NS, Oliver SG. 2004. Combining transcriptome data with genomic and cDNA sequence alignments to make confident functional assignments for *Aspergillus nidulans* genes. Mycol Res. 108: 853-857. doi: 10.1017/S095375620400067X

Singh S, Wakil JS. Stoops JK. 1985. Yeast fatty acid synthase: structure to function

relationship. Biochemistry 24: 6598-6602

Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE. 1986. Fluorescence detection in automated DNA sequence analysis. Nature 321: 674-679

Smutzer G . 2001.Yeast: an attractive, yet simple model. Scientist 15: 24-25

Soanes DM, Skinner W, Keon J, Hargreaves J, Talbot NJ. 2002. Genomics of phytopathogenic fungi and the development of bioinformatic resources. Molecular Plant-Microbe Interact 15: 421-427

Soanes DM, Talbot NJ. 2006. Comparative genomic analysis of phytopathogenic fungi using expressed sequence tag (EST) collections. Molecular Plant Pathol 7: 61-70

Soloman PS, Lee RC, Wilson TJG, Oliver RP. 2004. Pathogenicity of *Stagnospora nodorum* requires malate synthase. Mol. Microbiol. 53: 1065-1073

Solovyev V, Salamov A. 1997. The gene-ender computer tools for analysis of human and model organisms genome sequences. ISMB97, 294-302

Stenesh J, Roe BA. 1972. DNA Polymerase from Mesophilic and Thermophilic Bacteria I- Purification and Properties of the Enzyme. Biochem. Biophys. Acta. 272, 156-166

Stock AM, Robinson VL, Goudreau PN. 2000. Two-component signal transduction. Annu. Rev. Biochem.69: 183-215

Stommel M, Mann P, Franken P. 2001. Construction and analysis of an EST, library using RNA from dormant spores of the arbuscular mycorrhizal fungus *Gigaspora rosea*. Mycorrhiza 10: 281–785.

Strange RN, and Scott PR. 2005. Plant disease: A threat to global food security. Annual Review of Phytopathol 43:83-116.

Streets RB, Bloss HE. 1973. Phymatotrichum root rot. Am. Phytopathol. Soc. Monogr

Sturtevant AH. 1913. The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. J. Exp. Zool. 14:43-59

Sweigard JA, Carroll AM, Farrall L, Chumley F, Valent B. 1998. *Magnaporthe grisea* pathogenicity genes obtained through insertional mutagenesis. Mol. Plant-Microbe Interact. 11:404-412

Talbot NJ, Ebbole DJ, Hamer JE. 1993. Identification and characterization of MPG1, a gene involved in pathogenicity of the rice blast fungus *Magnaporthe grisea.*Plant Cell 5: 1575-1590.

Talbot, N.J., Y.P. Salch, M. Ma, and J.E. Hamer. 1993. Karyotypic variation with clonal lineages of the rice blast fungus, *Magnaporthe grisea*. Appl. Env. Microbiol. 59: 585-593

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, et al. 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4: 41.

Taylor, J.W., B.H. Bowman, M.L. Berbee, and T.J. White. 1993. Fungal Model Organisms: Phylogenetics of Saccharomyces, Aspergillus, and Neurospora. Syst. Biol. 42: 440-457.

Thon MR, Nuckles EM, Vaillancourt LJ. 2000. Restriction enzyme-mediated integration used to produce pathogenicity mutants of *Collectotrichum graminicola*. Mol. Plant-Microbe Interact. 13: 1356-1365

Torres TT, Metta M, Ottenwalder B, Sclottere C. 2008. Gene expression profiling by massively pyrosequencing Genome Res. 18:172-177

Ueno K, Yeung ES. 1994. Simultaneous Monitoring of DNA Fragments Separated by Capillary Electrophoresis in a Multiplexed Array of 100 Channels. Anal. Chem. 66: 1424-1431

Uppalapati SR, Marek SM, Lee HK, Nakashima J, Tang Y, Sledge MK, Dixon RA, Mysore KS. 2009. Global gene expression profiling during *Medicago truncatula – Phymatotrichopsis omnivora* interaction reveals a role for jasmonic acid, ethylene and the flavonoid pathway in disease development. Mol Plant Microbe Interact. 22 :7-17

Urban M, Bhargava T, Hamer JE. 1999. An ATP-driven efflux pump is a novel pathogenicity factor in rice blast disease. EMBO J 18:512-521

Vinson JP, Jaffe DB, O'Neill K, Karlsson EK, Stange-Thomann N, Anderson S, Mesirov JP, Satoh N, Satou Y, Nusbaum C, et al. (2005) Assembly of polymorphic genomes: Algorithms and application to *Ciona savignyi*. Genome Res. 15: 1127-1135

Wang C, St. Leger RJ. 2005. Developmental and transcriptional responses to host and non-host cuticles by the specific locust pathogen *Metarhizium anisopliae sf. acridum*. Eukaryotic Cell. 4: 937-947

Watkins GM.1981.Compendium of cotton diseases. APS Press, St Paul.

Watson JD, Crick FHC.1953. A structure for deoxyribose nucleic acid. Nature 171: 737-738

Weber APM, Weber KL, Carr K, Wilkerson C, Ohlrogge JB. 2007. Sampling the

Arabidopsis transcriptome with massively parallel pyrosequencing. Plant Physiol. 144: 32-42

Wicker T, Schlagenhauf E, Graner A, Close TJ, Keller B, Stein N. 2006. 454 sequencing put to the test using the complex genome of barley. BMC Genomics 7: 275-279

Wiley G, Macmil S, Qu C, Wang P, Xing Y, White D, White J, Roe BA. (In press) Methods for Generating Shotgun and Mixed Shotgun/Paired End Libraries for the 454 DNA Sequencer. In: Smith DR, editor. Current Protocols Human Genet. Winston-Salem, NC: John Wiley & Sons, Inc.

Zabriskie TM, Jackson MD.2000. Lysine biosynthesis and metabolism in fungi. Nat. Prod. Rep. 17:85-97.

Appendix Table 1. Putative identification and classification of EST's from the vegetative mycelia based on blast homology searches in KEGG, COG and COGEME databases.

Vegetative		
1. Carbohydrate Metabolism		
Glycolysis / Gluconeogenesis		
AFUA_1G06960	Pyruvate dehydrogenase E1 component alpha subunit, putative	7.00E-22
AFUA_6G07430	pyruvate kinase(EC:2.7.1.40)	2.00E-26
AN2875.2	similar to fructosebiphosphate aldolase	6.00E-26
AN3059.2	similar to phosphoglyceromutase	5.00E-26
AN5746.2	ENO_ASPOR Enolase (2-phosphoglyceratedehydratase)	5.00E-74
AN6037.2	G6PI_ASPORGlucose-6-phosphate isomerase (GPI) (Phosphoglucose	5.00E-36
AN8041.2	similar to glyceraldehyde 3-phosphatedehydrogenase	4.00E-31
AO090005001300	phosphoglycerate mutase	2.00E-22
AO090038000395	3-phosphoglycerate kinase	9.00E-32
UM03299.1	hypothetical protein triosephosphate isomerase (TIM) [EC:5.3.1.1]	4.00E-18
Citrate cycle (TCA cycle)		
AFUA_4G04520	succinyl-CoAsynthetase beta subunit (EC:6.2.1.4)	1.00E-27
AFUA_6G10660	ATP citrate lyase subunit(Acl), putative (EC:6.2.1.5)	9.00E-27
AN0896.2	hypothetical protein	2.00E-20
AO090009000285	citrate lyase beta subunit	2.00E-08
CAGL0E03850g	hypothetical protein succinate dehydrogenase (ubiquinone) iron-sulfur protein precursor [EC:1.3.5.1]	3.00E-27
DSY4204	hypothetical protein aconitate hydratase 1 [EC:4.2.1.3]	4.00E-31
Pentose phosphate pathway		
AN0688.2	similar to transketolase	4.00E-17
AN2875.2	similar to fructosebiphosphate aldolase	6.00E-26
AN6037.2	G6PI_ASPORGlucose-6-phosphate isomerase (GPI) (Phosphoglucose	5.00E-36
AN7588.2	hypothetical protein ribulose-	5.00E-27

	phosphate 3-epimerase [EC:5.1.3.1]	
MGG_02471	hypothetical protein transketolase [EC:2.2.1.1]	2.00E-23
Pentose and glucuronate interconversions		
AFUA_2G00760	pectate lyase A (EC:4.2.2.2)	5.00E-26
AN7588.2	hypothetical protein ribulose-phosphate 3-epimerase [EC:5.1.3.1]	5.00E-27
BSU35670	UTP-glucose-1-phosphateuridylyltransferase (EC:2.7.7.9)	2.00E-18
Fructose and mannose metabolism		
AN2875.2	similar to fructosebiphosphate aldolase	6.00E-26
CD2270	putative 1-phosphofructokinase(EC:2.7.1.56)	2.00E-18
SPAC1556.07	phosphomannomutase Pmm1 (EC:5.4.2.8)	2.00E-24
UM03299.1	hypothetical protein triosephosphate isomerase (TIM) [EC:5.3.1.1]	4.00E-18
Galactose metabolism		
4336515	hypothetical protein beta-fructofuranosidase [EC:3.2.1.26]	2.00E-06
AFUA_1G16250	alpha-glucosidase	3.00E-14
BSU35670	UTP-glucose-1-phosphateuridylyltransferase (EC:2.7.7.9)	2.00E-18
Starch and sucrose metabolism		
4336146	hypothetical protein beta-glucosidase [EC:3.2.1.21]	4.00E-21
4336515	hypothetical proteinbeta-fructofuranosidase [EC:3.2.1.26]	2.00E-06
AFUA_1G02410	AAA family ATPase Rvb2/Reptin (EC:3.6.1.-)	3.00E-08
AFUA_1G16250	alpha-glucosidase	3.00E-14
AFUA_5G03360	NADH pyrophosphatase(EC:3.6.1.-)	9.00E-06
AN6037.2	G6PI_ASPORGlucose-6-phosphate isomerase (GPI) (Phosphoglucose	5.00E-36
BSU35670	UTP-glucose-1-phosphateuridylyltransferase (EC:2.7.7.9)	2.00E-18
MGG_07289	glycogen synthase	5.00E-10
NT01CX_0970	glycogen phosphorylase (EC:2.4.1.1)	3.00E-14
SPAC1002.13c	beta-glucosidase Psu1(predicted) (EC:3.2.1.21)	4.00E-19
spr1709	sucrose phosphorylase (EC:2.4.1.5)	1.00E-12

Aminosugars metabolism		
AFUA_4G07850	endoglucanase (EC:3.2.1.-)	2.00E-06
MGG_09159	hypothetical protein chitin deacetylase [EC:3.5.1.41]	1.00E-11
Nucleotide sugars metabolism		
BSU35670	UTP-glucose-1-phosphateuridylyltransferase (EC:2.7.7.9)	2.00E-18
Pyruvate metabolism		
AFUA_1G06960	Pyruvate dehydrogenase E1 component alpha subunit, putative	7.00E-22
AFUA_5G12840	Hydroxyacyl glutathione hydrolase(EC:3.1.2.6)	7.00E-24
AFUA_6G07430	pyruvate kinase(EC:2.7.1.40)	2.00E-26
AN3901.2	hypothetical protein L-lactate dehydrogenase (cytochrome) [EC:1.1.2.3]	1.00E-25
H16_A1682	D-lactate dehydrogenase (EC:1.1.1.28)	5.00E-06
Glyoxylate and dicarboxylate metabolism		
DSY4204	hypothetical protein aconitate hydratase 1 [EC:4.2.1.3]	4.00E-31
Propanoate metabolism		
AFUA_4G04520	succinyl-CoAsynthetase beta subunit (EC:6.2.1.4)	1.00E-27
H16_A1909	2-methylcitrate dehydratase (EC:4.2.1.79)	6.00E-16
Butanoate metabolism		
AFUA_1G06960	Pyruvate dehydrogenase E1 component alpha subunit, putative	7.00E-22
AFUA_2G10650	enoyl-CoA hydratase (EC:4.2.1.-)	8.00E-18
AFUA_3G10660	hydroxymethylglutaryl-CoAsynthase Erg13 (EC:2.3.3.10)	3.00E-32
C5-Branched dibasic acid metabolism		
AFUA_4G04520	succinyl-CoAsynthetase beta subunit (EC:6.2.1.4)	1.00E-27
Inositol metabolism		
UM03299.1	hypothetical protein	4.00E-18
Inositol phosphate metabolism		
AO090701000359	myo-inositol-1-phosphate synthase	1.00E-22
1.2 Energy Metabolism		
Oxidative phosphorylation		
4339546	hypothetical protein	3.00E-39
AFUA_1G06610	NADH-quinone oxidoreductase, 23 kDasubunit (EC:1.6.5.3)	2.00E-27
AFUA_2G03010	cytochrome c subunit Vb	2.00E-25

AFUA_5G08560	vacuolar ATPase proteolipid subunit c	3.00E-17
AFUA_6G12790	NADH-ubiquinone oxidoreductase 39 kDasubunit (EC:1.6.5.3)	1.00E-19
AFUA_8G05440	mitochondrial ATPase subunit ATP4 (EC:3.6.3.14)	6.00E-10
AN0896.2	hypothetical protein	2.00E-20
AN6287.2	hypothetical protein	4.00E-09
AO090005000604	mitochondrial F1F0-ATP synthase,subunit b/ATP4	2.00E-26
AO090005000749	F0F1-type ATP synthase, gammasubunit	6.00E-25
AO090011000782	NADH: ubiquinone oxidoreductase,NDUFV2/24 kD subunit	7.00E-35
AO090120000313	vacuolar H+-ATPase V1 sector,subunit D	1.00E-20
AO090672000005	NADH-ubiquinone oxidoreductase, NUFS7/PSST/20 kDa subunit	2.00E-27
ArthMp006	NADH dehydrogenase subunit 5 (EC: 1.6.99.3)	4.00E-14
CAGL0E03850g	hypothetical protein	3.00E-27
Cthe_3020	NADH dehydrogenase subunit E (EC: 1.6.5.3)	1.00E-14
MGG_01111	cytochrome c oxidase polypeptide vib	7.00E-27
Q0070	Endonuclease I-SceIV, involved in intron mobility	2.00E-14
SPAC1782.07	ubiquinol-cytochrome-c reductase complex subunit 7 (EC:1.10.2.2)	9.00E-06
SPAC1B3.14	V-type ATPase subunit c (EC: 3.6.3.14)	6.00E-18
UsmafMp05	cytochrome c oxidase subunit 2	6.00E-16
UsmafMp10	NADH:ubiquinone oxidoreductase subunit 1	8.00E-08
UsmafMp12	apocytochrome b	9.00E-17
UsmafMp14	NADH:ubiquinone oxidoreductase subunit 5	1.00E-12
Methane metabolism		
AN5918.2	similar to AF316033_1 catalase C	6.00E-18
DEHA0F10593g	hypothetical protein	6.00E-21
SSON_4116	catalase	1.00E-36
Nitrogen metabolism		
AFUA_4G03950	cystathionine beta-lyase MetG (EC:4.4.1.8)	5.00E-11
AO090001000717	glutamate/leucine/phenylalanine /valinedehydrogenases	7.00E-11
MGG_14279	glutamine synthetase	3.00E-55

1.3 Lipid Metabolism		
Fatty acid metabolism		
AFUA_2G09910	fatty acid activator Faa4(EC:6.2.1.3)	5.00E-22
AN2762.2	hypothetical protein glutaryl-CoA dehydrogenase [EC:1.3.99.7]	9.00E-14
Synthesis and degradation of ketone bodies		
AFUA_3G10660	hydroxymethylglutaryl-CoA synthase Erg13 (EC:2.3.3.10)	3.00E-32
Biosynthesis of steroids		
AN5184.2	hypothetical protein carboxymethylene butenolidase [EC:3.1.1.45]	1.00E-18
Bile acid biosynthesis		
AFUA_2G10650	enoyl-CoA hydratase (EC:4.2.1.-)	8.00E-18
Glycerolipid metabolism		
AFUA_2G08380	diacylglycerol O-acyltransferase (DgaT) (EC:2.3.1.20)	9.00E-20
Glycerophospholipid metabolism		
AFUA_2G15970	Phosphatidylethanolamine methyltransferase	8.00E-15
AT3G15730	PLDALPHA1 phospholipase D	1.00E-34
Arachidonic acid metabolism		
AN5812.2	hypothetical protein [EC:3.3.2.6]	2.00E-22
CAGL0C01705g	hypothetical proteinK00432 glutathione peroxidase [EC:1.11.1.9]	5.00E-24
Biosynthesis of unsaturated fatty acids		
AN6731.2	similar to AF510861_1 stearic acid desaturase	1.00E-20
AO090005000456	fatty acid desaturase	3.00E-20
1.4 Nucleotide Metabolism		
AFUA_1G08840	guanylate kinase (EC:2.7.4.8)	5.00E-25
AFUA_6G07430	pyruvate kinase (EC:2.7.1.40)	2.00E-26
AFUA_6G08520	adenylate cyclase AcyA (EC:4.6.1.1)	5.00E-22
AFUA_7G02620	DNA-directed RNApolymerases N/8 kDa subunit superfamily	5.00E-10
AN5939.2	hypothetical protein 5'-nucleotidase [EC:3.1.3.5]	3.00E-08
Dvul_2495	adenine deaminase (EC:3.5.4.2)	7.00E-14
MGG_08622	nucleoside diphosphate kinase	6.00E-69
S0432	bifunctional UDP-sugarhydrolase/5'-nucleotidase periplasmic	9.00E-40
lwe1335	DNA polymerase III PolC (EC:2.7.7.7)	6.00E-20
Dpse_GA19108	GA19108 gene product from transcript GA19108-RA	2.00E-11

KLLA0E08030g	hypothetical protein cytidylate kinase [EC:2.7.4.14]	4.00E-16
MGG_01537	hypothetical protein aspartate carbamoyltransferase catalytic chain [EC:2.1.3.2]	2.00E-07
1.5 Amino Acid Metabolism		
Glutamate metabolism		
AGOS_AGR196W	glutathione reductase (NADPH) [EC:1.8.1.7]	2.00E-22
AN4159.2	similar to glutamine synthetase	7.00E-33
AO090001000717	glutamate/leucine/phenylalanine /valine dehydrogenases	7.00E-11
MGG_01537	hypothetical protein carbamoyl-phosphate synthase small chain [EC:6.3.5.5]	2.00E-07
MGG_14279	glutamine synthetase	3.00E-55
Alanine and aspartate metabolism		
AFUA_1G06960	pyruvatedehydrogenase E1 component alpha subunit, putative	7.00E-22
AGOS_AER230C	AER230Cp argininosuccinate synthase [EC:6.3.4.5]	9.00E-17
Cthe_3149	aminoacyl-histidine dipeptidase (EC:3.4.13.3)	1.00E-07
MGG_01537	hypothetical protein carbamoyl-phosphate synthase small chain [EC:6.3.5.5]	2.00E-07
Mmc1_1746	L-aspartate oxidase(EC:1.4.3.16)	2.00E-18
Glycine, serine and threonine metabolism		
AFUA_2G15970	Phosphatidyl ethanolamine methyl transferase	8.00E-15
AO090003000721	homoserine dehydrogenase	9.00E-22
CNG01110	copper amine oxidase	5.00E-22
MGG_06321	hypothetical protein glycyl-tRNA synthetase, class II [EC:6.1.1.14]	4.00E-16
MGG_11450	hypothetical protein homoserine dehydrogenase [EC:1.1.1.3]	4.00E-15
Methionine and Cysteine metabolism		
AFUA_4G03950	cystathionine beta-lyase MetG (EC:4.4.1.8)	5.00E-11
AN8277.2	CYSD_EMENI O-acetyl homoserine (Thiol)-Lyase	4.00E-31
BF3351	putative cysteine biosynthesisrelated protein	6.00E-34
Valine, leucine and isoleucine degradation		
AFUA_3G10660	hydroxymethylglutaryl-CoAsynthase Erg13 (EC:2.3.3.10)	3.00E-32
Valine, leucine and isoleucine biosynthesis		
AFUA_1G06960	Pyruvate dehydrogenase E1 component alpha subunit,	7.00E-22

	putative	
AO090005001122	3-isopropylmalate dehydrogenase	6.00E-26
CD2618	isoleucyl-tRNA synthetase (EC:6.1.1.5)	2.00E-22
Lysine biosynthesis		
AO090003000721	homoserine dehydrogenase	9.00E-22
Dred_3249	Orn/DAP/Arg decarboxylase 2	1.00E-28
MGG_08564	saccharopine dehydrogenase	1.00E-21
MGG_11450	hypothetical protein	4.00E-15
SPCG_2062	2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase,	2.00E-06
Lysine degradation		
AN2762.2	hypothetical protein glutaryl-CoA dehydrogenase [EC:1.3.99.7]	9.00E-14
MGG_08564	saccharopine dehydrogenase	1.00E-21
Arginine and proline metabolism		
AGOS_AER230C	AER230Cp argininosuccinate synthase [EC:6.3.4.5]	9.00E-17
TM1097	ornithine carbamoyltransferase	5.00E-09
Histidine metabolism		
BCE_0363	RNA methyltransferase	3.00E-08
CNG01110	copper amine oxidase	5.00E-22
Cthe_0724	histidinol phosphate phosphatase HisJfamily (EC:3.1.3.15)	2.00E-16
Cthe_3149	aminoacyl-histidine dipeptidase (EC:3.4.13.3)	1.00E-07
Dvul_2060	RNA methyltransferase, TrmA family	4.00E-13
Phenylalanine, tyrosine and tryptophan metabolism		
AFUA_2G04200	4-hydroxyphenylpyruvatedioxygenase (EC:1.13.11.27)	5.00E-19
AFUA_2G04220	homogentisate 1,2-dioxygenase(HmgA) (EC:1.13.11.5)	4.00E-25
AFUA_2G10650	enoyl-CoA hydratase (EC:4.2.1.-)	8.00E-18
AN1897.2	HGD_EMENI Homogentisate1,2-dioxygenase (Homogentisicase)	9.00E-22
BCE_0363	RNA methyltransferase	3.00E-08
CNG01110	copper amine oxidase	5.00E-22
Dvul_2060	RNA methyltransferase, TrmA family	4.00E-13
DEHA0F10593g	hypothetical protein peroxidase [EC:1.11.1.7]	6.00E-21
AN2762.2	hypothetical protein glutaryl-CoA dehydrogenase [EC:1.3.99.7]	9.00E-14
AN5918.2	similar to AF316033_1 catalase C	6.00E-18
MGG_05814	Hypothetical protein	2.00E-22
SSON_4116	catalase	1.00E-36

AO090120000438	chorismate mutase	1.00E-21
Urea cycle and metabolism of amino groups		
AGOS_AER230C	AER230Cp	9.00E-17
AO090701000729	putative glutamate/ornithineacetyltransferase	7.00E-17
CNG01110	copper amine oxidase	5.00E-22
Cthe_3149	aminoacyl-histidine dipeptidase (EC:3.4.13.3)	1.00E-07
MGG_04210	hypothetical protein glutamate N-acetyltransferase [EC:2.3.1.35]	2.00E-17
TM1097	ornithine carbamoyltransferase	5.00E-09
TTE2495	PLP-dependent aminotransferase	2.00E-16
1.6 Metabolism of Other Amino Acids		
beta-Alanine metabolism		
CNG01110	copper amine oxidase	5.00E-22
Cthe_3149	aminoacyl-histidine dipeptidase (EC:3.4.13.3)	1.00E-07
Aminophosphonate metabolism		
BCE_0363	RNAmethyltransferase	3.00E-08
Selenoamino acid metabolism		
AFUA_4G03950	cystathionine beta-lyase MetG (EC:4.4.1.8)	5.00E-11
BCE_0363	RNA methyltransferase	3.00E-08
Dvul_2060	RNA methyltransferase, TrmA family	4.00E-13
Cyanoamino acid metabolism		
4336146	hypothetical proteinbeta-glucosidase [EC:3.2.1.21]	4.00E-21
SPAC1002.13c	beta-glucosidase Psu1(predicted) (EC:3.2.1.21)	4.00E-19
D-Alanine metabolism		
Dred_1796	D-alanine--D-alanine ligase(EC:6.3.2.4)	5.00E-09
Glutathione metabolism		
AGOS_AGR196W	AGR196Wp	2.00E-22
AN4905.2	similar to AF425746_1 thetaclass glutathione S-transferase	6.00E-24
CAGL0C01705g	hypothetical protein glutathione peroxidase [EC:1.11.1.9]	5.00E-24
SSON_0935	aminopeptidase N	2.00E-18
1.7 Glycan Biosynthesis and Metabolism		
N-Glycan biosynthesis		
453895	asparagine-linked glycosylation 1homolog (S. cerevisiae,	4.00E-15
High-mannose type N-glycan biosynthesis		
AN7672.2	hypothetical protein mannan polymerase complexes MNN9 subunit [EC:2.4.1.-]	5.00E-16
AO090003001140	subunit of Golgimannosyltransferase	5.00E-18

	complex	
Glycan structures - biosynthesis 1		
453895	asparagine-linked glycosylation 1homolog (S. cerevisiae,	4.00E-15
AN7672.2	hypothetical protein	5.00E-16
AO090003001140	subunit of Golgimannosyltransferase complex	5.00E-18
1.9 Metabolism of Cofactors and Vitamins		
Riboflavin metabolism		
AO090003000004	6,7-dimethyl-8-ribityllumazinesynthase	1.00E-09
Vitamin B6 metabolism		
SPAC9E9.11	pyridoxal reductase (PMID 10438489) (EC:1.1.1.65)	7.00E-19
Nicotinate and nicotinamide metabolism		
AFUA_3G05730	nicotinate-nucleotidepyrophosphorylase (EC:2.4.2.19)	6.00E-23
AN5939.2	hypothetical protein	3.00E-08
Mmc1_1746	L-aspartate oxidase(EC:1.4.3.16)	2.00E-18
S0432	bifunctional UDP-sugarhydrolase/5'-nucleotidase periplasmic	9.00E-40
Pantothenate and CoA biosynthesis		
AO090003001332	halotolerance protein HAL3(contains flavoprotein domain)	3.00E-10
Folate biosynthesis		
AFUA_1G02410	AAA family ATPase Rvb2/Reptin (EC:3.6.1.-)	3.00E-08
AFUA_5G03360	NADH pyrophosphatase(EC:3.6.1.-)	9.00E-06
CPE1019	GTP cyclohydrolase I (EC:3.5.4.16)	2.00E-18
Ubiquinone biosynthesis		
Cthe_3020	ech hydrogenase subunit E(EC:1.6.5.3)	1.00E-14
UsmafMp10	NADH:ubiquinone oxidoreductasesubunit 1	8.00E-08
UsmafMp14	NADH:ubiquinone oxidoreductasesubunit 5	1.00E-12
Limonene and pinene degradation		
AFUA_2G10650	enoyl-CoA hydratase (EC:4.2.1.-)	8.00E-18
MGG_05814	hypotheticalprotein	2.00E-22
Phenylpropanoid biosynthesis		
4336146	hypothetical proteinbeta-glucosidase [EC:3.2.1.21]	4.00E-21
DEHA0F10593g	hypothetical protein peroxidase [EC:1.11.1.7]	6.00E-21
SPAC1002.13c	beta-glucosidase Psu1(predicted) (EC:3.2.1.21)	4.00E-19
Streptomycin biosynthesis		
AN7625.2	myo-inositol-1-phosphate synthase [EC:5.5.1.4]	6.00E-23

gamma-Hexachlorocyclohexane degradation		
AN5184.2	hypothetical protein delta24(24(1))-sterol reductase [EC:1.3.1.71]	1.00E-18
Styrene degradation		
AFUA_2G04220	homogentisate 1,2-dioxygenase(HmgA) (EC:1.13.11.5)	4.00E-25
AN1897.2	HGD_EMENI Homogentisate1,2-dioxygenase (Homogentisicase)	9.00E-22
1,4-Dichlorobenzene degradation		
AN5184.2	hypothetical protein delta24(24(1))-sterol reductase [EC:1.3.1.71]	1.00E-18
Benzoate degradation		
AFUA_2G10650	enoyl-CoA hydratase (EC:4.2.1.-)	8.00E-18
AN2762.2	hypothetical protein glutaryl-CoA dehydrogenase [EC:1.3.99.7]	9.00E-14
1- and 2-Methylnaphthalene degradation		
AFUA_2G10650	enoyl-CoA hydratase (EC:4.2.1.-)	8.00E-18
MGG_05814	hypotheticalprotein	2.00E-22
Metabolism of xenobiotics by cytochrome P450		
AFUA_1G12880	epoxide hydrolase (EC:3.3.2.9)	4.00E-12
AN4905.2	similar to AF425746_1 theta class glutathione S-transferase	6.00E-24
2. INFORMATION STORAGE AND PROCESSING		
Translation, ribosomal structure and biogenesis		
KOG0900	40S ribosomal protein S20	1.00E-27
KOG1560	Translation initiation factor 3, subunit h(eIF-3h)	3.00E-35
KOG0123	Polyadenylate-binding protein (RRMsuperfamily)	2.00E-09
KOG2187	tRNA uracil-5-methyltransferase and relatedtRNA-modifying enzymes	4.00E-07
KOG1697	Mitochondrial/chloroplast ribosomal proteinS9	5.00E-09
KOG0462	Elongation factor-type GTP-binding protein	8.00E-19
KOG0469	Elongation factor 2	2.00E-31
KOG1678	60s ribosomal protein L15	3.00E-12
KOG0434	Isoleucyl-tRNA synthetase	1.00E-11
KOG0901	60S ribosomal protein L14/L17/L23	3.00E-19
KOG0898	40S ribosomal protein S15	2.00E-24
KOG3677	RNA polymerase I-associated factor - PAF67	3.00E-11
KOG1670	Translation initiation factor 4F,cap-binding subunit (eIF-4E) and related cap-bindingproteins	2.00E-13
KOG2298	Glycyl-tRNA synthetase and related class IItRNA synthetase	1.00E-12
KOG3387	60S ribosomal protein 15.5kD/SNU13,NHP2/L7A	1.00E-12

	family (includes ribonuclease P subunit p38),involved in splicing	
KOG1646	40S ribosomal protein S6	3.00E-16
KOG1714	60s ribosomal protein L18	7.00E-12
KOG0878	60S ribosomal protein L32	2.00E-19
KOG3181	40S ribosomal protein S3	2.00E-34
KOG4163	Prolyl-tRNA synthetase	5.00E-29
KOG1770	Translation initiation factor 1(eIF-1/SUI1)	3.00E-14
KOG3301	Ribosomal protein S4	5.00E-28
KOG2523	Predicted RNA-binding protein with PUAdomain	3.00E-10
KOG1732	60S ribosomal protein L21	2.00E-34
KOG0052	Translation elongation factor EF-1 alpha/Tu	7.00E-36
KOG0434	Isoleucyl-tRNA synthetase	9.00E-14
KOG1570	60S ribosomal protein L10A	1.00E-11
KOG3185	Translation initiation factor 6 (eIF-6)	2.00E-11
KOG0900	40S ribosomal protein S20	3.00E-08
KOG0459	Polypeptide release factor 3	3.00E-19
KOG0877	40S ribosomal protein S2/30S ribosomal proteinS5	4.00E-09
KOG3421	60S ribosomal protein L14	2.00E-09
KOG3291	Ribosomal protein S7	5.00E-07
KOG0815	60S acidic ribosomal protein P0	2.00E-07
KOG3401	60S ribosomal protein L26	8.00E-27
KOG0893	60S ribosomal protein L31	2.00E-19
KOG3295	60S Ribosomal protein L13	3.00E-14
KOG1242	Protein containing adaptin N-terminal region	6.00E-26
KOG2334	tRNA-dihydrouridine synthase	2.00E-18
KOG0746	60S ribosomal protein L3 and related proteins	5.00E-28
KOG0460	Mitochondrial translation elongation factor Tu	5.00E-23
KOG0887	60S ribosomal protein L35A/L37	7.00E-12
KOG0402	60S ribosomal protein L37	2.00E-22
KOG0052	Translation elongation factor EF-1 alpha/Tu	3.00E-38
KOG0397	60S ribosomal protein L11	1.00E-20
KOG1749	40S ribosomal protein S23	1.00E-22
KOG0830S	ribosomal protein SA (P40)/Lamininreceptor 1	1.00E-14
KOG0407	40S ribosomal protein S14	8.00E-07
KOG0407	40S ribosomal protein S14	1.00E-15
KOG3271	Translation initiation factor 5A (eIF-5A)	2.00E-15
KOG2988	60S ribosomal protein L30	4.00E-40
KOG3412	60S ribosomal protein L28	2.00E-08

KOG1728	40S ribosomal protein S11	8.00E-51
KOG1646	40S ribosomal protein S6	3.00E-16
KOG0830	40S ribosomal protein SA (P40)/Lamininreceptor 1	1.00E-14
KOG3406	40S ribosomal protein S12	1.00E-48
KOG1767	40S ribosomal protein S25	2.00E-21
KOG3502	40S ribosomal protein S28	6.00E-15
KOG3504	60S ribosomal protein L29	2.00E-15
KOG1742	60s ribosomal protein L15/L27	1.00E-33
KOG3418	60S ribosomal protein L27	3.00E-45
KOG3204	60S ribosomal protein L13a	2.00E-28
KOG3486	40S ribosomal protein S21	3.00E-12
KOG0004	Ubiquitin/40S ribosomal protein S27a fusion	2.00E-30
KOG1754	40S ribosomal protein S15/S22	6.00E-29
KOG1678	60s ribosomal protein L15	8.00E-26
KOG1714	60s ribosomal protein L18	4.00E-18
KOG0469	Elongation factor 2	3.00E-29
KOG3344	40s ribosomal protein s10	3.00E-16
KOG1732	60S ribosomal protein L21	1.00E-06
KOG1570	60S ribosomal protein L10A	5.00E-26
RNA processing and modification		
KOG0123	Polyadenylate-binding protein (RRMsuperfamily)	2.00E-09
KOG1919	RNA pseudouridylate synthases	1.00E-09
KOG1644	U2-associated snRNP A' protein	3.00E-16
KOG3801	Uncharacterized conserved protein BCN92	2.00E-12
KOG0007	Splicing factor 3a, subunit 1	1.00E-15
KOG0345	ATP-dependent RNA helicase	1.00E-14
KOG3460	Small nuclear ribonucleoprotein (snRNP)LSM3	1.00E-12
KOG3448	Predicted snRNP core protein	5.00E-20
KOG4768	Mitochondrial mRNA maturase	7.00E-17
KOG0329	ATP-dependent RNA helicase	7.00E-11
TWOG0967	Mitochondrial mRNA maturase/Homingendonuclease	1.00E-15
TWOG0658	Mitochondrial mRNA maturase encoded bypartially processed COB mRNA	9.00E-06
TWOG0658	Mitochondrial mRNA maturase encoded bypartially processed COB mRNA	6.00E-08
KOG0131	Splicing factor 3b, subunit 4	2.00E-18
KOG1784	Small Nuclear ribonucleoprotein splicingfactor	1.00E-06
KOG3387	60S ribosomal protein 15.5kD/SNU13,NHP2/L7A family (includes ribonuclease P	1.00E-12

	subunit p38),involved in splicing	
KOG3172	Small nuclear ribonucleoprotein Sm D3	5.00E-21
KOG3482	Small nuclear ribonucleoprotein (snRNP) SMF	1.00E-24
TWOG0658	Mitochondrial mRNA maturase encoded bypartially processed COB mRNA	4.00E-06
KOG4768	Mitochondrial mRNA maturase	5.00E-22
TWOG0967	Mitochondrial mRNA maturase/Homingendonuclease	2.00E-19
Transcription		
KOG0483	Transcription factor HEX, contains HOX andHALZ domains	4.00E-11
KOG0323	TFIIF-interacting CTD phosphatases, includingNLI-interacting factor	9.00E-12
KOG2239	Transcription factor containing NAC and TS-Ndomains	7.00E-06
KOG0668	Casein kinase II, alpha subunit	1.00E-24
KOG3677	RNA polymerase I-associated factor - PAF67	3.00E-11
KOG4210	Nuclear localization sequence bindingprotein	7.00E-09
KOG3001	Dosage compensation regulatorycomplex/histone acetyltransferase complex, subunitMSL-3/MRG15/EAF3, and related CHROMO domain-containingproteins	1.00E-12
KOG1534	Putative transcription factor FET5	9.00E-14
KOG2240	RNA polymerase II general transcriptionfactor BTF3 and related proteins	3.00E-06
KOG1414	Transcriptional activator FOSB/c-Fos andrelated bZIP transcription factors	1.00E-09
KOG3497	DNA-directed RNA polymerase, subunit RPB10	3.00E-10
KOG2345	Serine/threonine protein kinase/TGF-betastimulated factor	7.00E-06
Replication, recombination and repair		
KOG1806	DEAD box containing helicases	3.00E-11
KOG3041	Nucleoside diphosphate-sugar hydrolase of theMutT (NUDIX) family	7.00E-09
KOG1433	DNA repair protein RAD51/RHP55	4.00E-10
KOG3752	Ribonuclease H	3.00E-06
KOG1275	PAB-dependent poly(A) ribonuclease, subunitPAN2	2.00E-26

KOG2249	3'-5' exonuclease	5.00E-18
KOG3176	Predicted alpha-helical protein, potentially involved in replication/repair	4.00E-10
TWOG0967	Mitochondrial mRNA maturase/Homingendonuclease	1.00E-15
TWOG0967	Mitochondrial mRNA maturase/Homingendonuclease	2.00E-19
KOG0219	Mismatch repair ATPase MSH2 (MutS family)	5.00E-13
Chromatin structure and dynamics		
KOG1279	Chromatin remodeling factor subunit andrelated transcription factors	2.00E-15
KOG3467	Histone H4	2.00E-37
KOG3001	Dosage compensation regulatorycomplex/histone acetyltransferase complex, subunitMSL-3/MRG15/EAF3, and related CHROMO domain-containingproteins	1.00E-12
KOG1745	Histones H3 and H4	8.00E-32
3. CELLULAR PROCESSES AND SIGNALING		
Cell cycle control, cell division, chromosome partitioning		
KOG1852	Cell cycle-associated protein	4.00E-12
KOG0668	Casein kinase II, alpha subunit	1.00E-24
KOG3484	Cyclin-dependent protein kinase CDC28, regulatory subunit CKS1, and related proteins	2.00E-28
KOG2655	Septin family protein (P-loop GTPase)	2.00E-12
Nuclear structure		
KOG2171	Karyopherin (importin) beta 3	1.00E-10
Signal transduction mechanisms		
KOG0500	Cyclic nucleotide-gated cation channel CNGA1-3and related proteins	6.00E-07
KOG1187	Serine/threonine protein kinase	5.00E-18
KOG1554	COP9 signalosome, subunit CSN5	1.00E-09
KOG0519	Sensory transduction histidine kinase	9.00E-06
KOG0703	Predicted GTPase-activating protein	2.00E-13
KOG0027	Calmodulin and related proteins (EF-Handsuperfamily)	2.00E-15
KOG4293	Predicted membrane protein, contains DoHand Cytochrome b-561/ferric reductase	1.00E-31

	transmembranedomains	
KOG0027	Calmodulin and related proteins (EF-Handsuperfamily)	8.00E-15
KOG1251	Serine racemase	2.00E-10
LSE0196	H S100 EF-hand calcium-binding protein	8.00E-10
KOG0027	Calmodulin and related proteins (EF-Handsuperfamily)	1.00E-31
KOG0668	Casein kinase II, alpha subunit	1.00E-24
KOG0279	G protein beta subunit-like protein	6.00E-26
KOG4561	Uncharacterized conserved protein, contains TBC domain	6.00E-13
KOG1354	Serine/threonine protein phosphatase 2A,regulatory subunit	4.00E-11
KOG3979	FGF receptor activating protein 1	3.00E-06
KOG3699	Cytoskeletal protein Adducin	4.00E-12
KOG0585	Ca2+/calmodulin-dependent protein kinasekinase beta and related serine/threonine protein kinases	5.00E-17
KOG0034	Ca2+/calmodulin-dependent protein phosphatase (calcineurin subunit B), EF-Hand superfamily protein	6.00E-42
KOG1435	Sterol reductase/lamin B receptor	1.00E-17
KOG3348	BolA (bacterial stress-inducedmorphogen)-related protein	7.00E-07
KOG1486	GTP-binding protein DRG2 (ODN superfamily)	6.00E-10
KOG1251	Serine racemase	4.00E-12
KOG0279	G protein beta subunit-like protein	2.00E-06
KOG2345	Serine/threonine protein kinase	7.00E-06
Cell wall/membrane/envelope biogenesis		
KOG3144	Ethanolamine-P-transferase GPI11/PIG-F,involved in glycosyl phosphatidyl inositol anchorbiosynthesis	7.00E-07
KOG4748	Subunit of Golgi mannosyl transferase complex	1.00E-14
Cytoskeleton		
KOG0836	F-actin capping protein, alpha subunit	1.00E-10
KOG0676	Actin and related proteins	4.00E-29
KOG3699	Cytoskeletal protein Adducin	4.00E-12
KOG0676	Actin and related proteins	1.00E-24
KOG2655	Septin family protein (P-loop GTPase)	2.00E-12

KOG1735	Actin depolymerizing factor	5.00E-25
KOG1755	Profilin	1.00E-20
Gz47835846	Arp2/3 complex subunit Arc16	1.00E-19
UmCon[0003]	cofilin, cortical cytoskeleton component, actin binding and severing protein	2.00E-25
mg[0026]	beta (1-3) glucanosyltransferase, glycophospholipid-anchored surface glycoprotein that regulates the crosslinking of beta-1,6-glucans in the cell wall	4.00E-15
Mag45392335	endochitinase	1.00E-07
W0AA026ZC01C1	involved in mitochondrial distribution and morphology	3.00E-17
PsCon[0140]	glycosyltransferase, involved in cellwall biogenesis	1.00E-09
BgCon[1705]	chitin deacetylase	2.00E-19
BfCon[0775]	cell wall biogenesis protein	3.00E-07
Gz22509250	Golgi reassembly stacking protein	2.00E-06
Cell death		
Gz31374047	apoptosis inducing factor (pyridine nucleotide-disulphide oxidoreductase)	2.00E-14
Mag30404439	DNA-binding apoptosis protein	8.00E-06
Cell rescue Polysacch degradation		
MagCon[3403]	beta-glucosidase	2.00E-17
GzCon[3452]	2-deoxy-D-gluconate3-dehydrogenase, pectin degradation	3.00E-21
mg[0971]	endoglucanase	9.00E-12
VD0104F03	pectate lyase	3.00E-25
GzCon[0313]	endo-1,4-beta-glucanase	4.00E-11
Cell rescue /Detoxification		
Mag3391888	catalase/peroxidase	2.00E-30
GzCon[0787]	cytochrome c peroxidase	4.00E-30
FsCon[1531]	glutathione-disulfide reductase	3.00E-24
mgb0524f	glutathione-peroxidase,involved in oxidative stress response	6.00E-31
mg[0705]	toluene sulfonate zinc-independent alcohol dehydrogenase	1.00E-07
SSPG680	glutathione S-transferase	2.00E-28
FsCon[1411]	lysyl oxidase	4.00E-23
BgCon[0791]	haloacid dehalogenase-like hydrolase	3.00E-32
Mag30403767	lactamase beta 2	4.00E-21
mg[1358]	thiol-specific antioxidantprotein	3.00E-28
mg[1108]	alpha-ketoglutarate-dependentsulphonate dioxygenase involved in sulphonate catabolism	4.00E-28

mg[0170]	Cu/Zn superoxide dismutase	2.00E-59
MagCon[3299a]	glutaredoxin	1.00E-18
mgc05e09f	pisatin demethylase(cytochrome P450)	5.00E-15
mgb0771f	peroxisomal membrane proteinPMP20, peroxiredoxin	9.00E-21
W0AA054ZF04C1	cytosolic Cu/Zn superoxidedismutase	1.00E-07
mg[0642]	Hydroxyacyl glutathionehydrolase / glyoxylase	2.00E-27
Intracellular trafficking, secretion, and vesicular transport		
KOG3316	Transport protein particle (TRAPP) complexsubunit	9.00E-18
KOG0934	Clathrin adaptor complex, small subunit	4.00E-24
KOG0094	GTPase Rab6/YPT6/Ryh1, small G proteinsuperfamily	1.00E-29
KOG3251	Golgi SNAP receptor complex member	6.00E-06
KOG2171	Karyopherin (importin) beta 3	3.00E-13
KOG2667	COPII vesicle protein	5.00E-09
KOG0070	GTP-binding ADP-ribosylation factor Arf1	1.00E-26
KOG4097	Succinate dehydrogenase membrane anchorsubunit and related proteins	3.00E-15
KOG0292	Vesicle coat complex COPI, alpha subunit	2.00E-16
KOG0072	GTP-binding ADP-ribosylation factor-likeprotein ARL1	2.00E-16
KOG2104	Nuclear transport factor 2	1.00E-15
KOG3065	SNAP-25 (synaptosome-associated protein) component of SNARE complex	8.00E-07
KOG2655	Septin family protein (P-loop GTPase)	2.00E-12
KOG1691	emp24/gp25L/p24 family of membrane trafficking proteins	8.00E-13
KOG0096	GTPase Ran/TC4/GSP1 (nuclear protein transportpathway), small G protein superfamily	2.00E-10
KOG0810	SNARE protein Syntaxin 1 and related proteins	8.00E-08
Posttranslational modification, protein turnover, chaperones		
KOG0549	FKBP-type peptidyl-prolyl cis-trans isomerase	4.00E-32
KOG0885	Peptidyl-prolyl cis-trans isomerase	5.00E-23
KOG4127	Renal dipeptidase	7.00E-11

KOG3144	Ethanolamine-P-transferase GPI11/PIG-F,involved in glycosylphosphatidylinositol anchorbiosynthesis	7.00E-07
KOG1554	COP9 signalosome, subunit CSN5	1.00E-09
KOG1051	Chaperone HSP104 and related ATP-dependentClp proteases	3.00E-16
KOG1812	Predicted E3 ubiquitin ligase	8.00E-08
KOG0174	20S proteasome, regulatory subunit beta typePSMB6/PSMB9/PRE3	7.00E-35
KOG1752	Glutaredoxin and related proteins	4.00E-06
KOG0841	Multifunctional chaperone (14-3-3 family)	6.00E-06
KOG0880	Peptidyl-prolyl cis-trans isomerase	3.00E-12
KOG3946	Glutaminyl cyclase	8.00E-08
KOG1760	Molecular chaperone Prefoldin, subunit 4	3.00E-09
KOG1812	Predicted E3 ubiquitin ligase	2.00E-12
KOG2012	Ubiquitin activating enzyme UBA1	2.00E-23
KOG1769	Ubiquitin-like proteins	3.00E-06
KOG2941	Beta-1,4-mannosyltransferase	2.00E-16
KOG0546	HSP90 co-chaperone CPR7/Cyclophilin	7.00E-19
KOG0173	20S proteasome, regulatory subunit betatype PSMB7/PSMB10/PUP1	1.00E-10
KOG1153	Subtilisin-related protease/Vacuolarprotease B	1.00E-09
KOG0183	20S proteasome, regulatory subunit alphatype PSMA7/PRE6	6.00E-27
KOG0417	Ubiquitin-protein ligase	2.00E-37
KOG3457	Sec61 protein translocation complex, betasubunit	4.00E-14
KOG0184	20S proteasome, regulatory subunit alphatype PSMA3/PRE10	9.00E-17
KOG0541	Alkyl hydroperoxide reductase/peroxiredoxin	1.00E-15
KOG0896	Ubiquitin-conjugating enzyme E2	2.00E-18
KOG2754	Oligosaccharyltransferase, beta subunit	7.00E-07
KOG1641	Mitochondrial chaperonin	2.00E-10
KOG0854	Alkyl hydroperoxide reductase, thiol specificantioxidant and related enzymes	1.00E-14
KOG0359	Chaperonin complex component, TCP-1 zetasubunit (CCT6)	3.00E-08
KOG0880	Peptidyl-prolyl cis-trans isomerase	2.00E-20
KOG2195	Transferrin receptor and related	2.00E-17

	proteinscontaining the protease-associated (PA) domain	
KOG1769	Ubiquitin-like proteins	4.00E-15
KOG1153	Subtilisin-related protease/Vacuolar protease B	4.00E-09
KOG0177	20S proteasome, regulatory subunit beta typePSMB2/PRE1	3.00E-06
KOG1555	26S proteasome regulatory complex, subunitRPN11	2.00E-08
KOG0182	20S proteasome, regulatory subunit alpha typePSMA6/SCL1	1.00E-17
KOG1651	Glutathione peroxidase	1.00E-22
KOG0729	26S proteasome regulatory complex, ATPase RPT1	2.00E-18
KOG0001	Ubiquitin and ubiquitin-like proteins	1.00E-45
KOG0907	Thioredoxin	2.00E-21
KOG0865	Cyclophilin type peptidyl-prolyl cis-transisomerase	1.00E-44
KOG0181	20S proteasome, regulatory subunit alpha typePSMA2/PRE8	1.00E-22
KOG1047	Bifunctional leukotriene A4hydrolase/aminopeptidase LTA4H	9.00E-22
KOG3355	Mitochondrial sulfhydryl oxidase involved inthe biogenesis of cytosolic Fe/S proteins	1.00E-13
Transport facilitation		
Mag30417961	sugar ABC transporter	7.00E-08
BfCon[1206]	putrescine transport protein	2.00E-12
MagCon[1406]	copper transporter	8.00E-16
Gz22504204	4-nitrophenylphosphatase domainand non-neuronal SNAP25-like protein 1 (NIPSNAP1)	1.00E-27
Ct21907372	hexose transporter	1.00E-36
mg[0852]	calcium-related spray protein	3.00E-26
SSPG1149F	vacuolar ATP synthase (vacuolarproton pump) 16 kDa proteolipid subunit	1.00E-25
Um37410097	multidrug resistant protein	4.00E-08
BfCon[1033]	MRP-like ABC transporter	8.00E-32
BfCon[1708]	stomatin	2.00E-33
BgCon[2034]	transporter	3.00E-08
Mag30415231	MFS transporter	2.00E-07
CpCEST-19-F-02	tricarboxylate transportprotein	3.00E-08
GzCon[2954]	killer toxin resistance	2.00E-09
MagCon[0480]	mitochondrial phosphate carrier	1.00E-26
W0AA050ZE11C1	vacuolar proton pump D subunit	2.00E-27
Gz15771683	Na-Ca exchanger	3.00E-16
GzCon[0657]	MFS transporter	7.00E-07

CpCon[0735]	inorganic phosphate transporter	3.00E-11
GzCon[2761]	phosphate permease	6.00E-33
BfCon[1054]	vacuolar ATP synthase subunit G	4.00E-08
MagCon[0264a]	acyl-coenzyme A bindingprotein	3.00E-06
VD0201B05	vacuolar ATPase V0 domain subunit	2.00E-20
Um34332829	V-type ATPase, ATP synthase jchain	2.00E-15
GzCon[7196]	plasma membrane zinc ion transporter	4.00E-18
mga0204f	carboxylic acid transporter protein(pyruvate and lactate/H[+] symporter)	2.00E-14
MagCon[7556a]	mitochondrial phosphate transporter	7.00E-30
Um34330445	outer mitochondrial membraneprotein porin, voltage-dependent anion-selective channel	6.00E-28
SSPG256F	vacuolar-ATPase	3.00E-14
UmCon[0408]	Heterokaryon incompatibility, het-c	2.00E-23
Transposon insertion sequence proteins		
Gz22509405	intron derivedmaturase	2.00E-07
Bg27453265	pol polyprotein	2.00E-06
Gz22509405	intron derivedmaturase	2.00E-06
CpCon[0027]	ORF B (Cryphonectria hypovirus 1)	1.00E-07
BfCon[1613]	intron derivedmaturase	2.00E-11

Appendix Table 2. Putative identification and classification of EST's from sclerotia based on blast homology searches in KEGG, COG and COGEME databases.

Sclerotia		
1.1 Carbohydrate Metabolism		
Glycolysis / Gluconeogenesis		
BgCon[1917]	phosphoglycerate kinase	5.00E-10
GzCon[1766]	6-phosphofructokinase alphasubunit	3.00E-15
17544880	probable transmembrane aldehydedehydrogenase oxidoreductase protein	8.00E-06
71001262	glucose-6-phosphate isomerase[EC5319]	3.00E-09
27383009	alcoholdehydrogenase [EC1111 111284]	4.00E-06
19112945	hypothetical protein[EC5422]	2.00E-07
54023581	putative pyruvatedehydrogenase E1 component [EC1241]	3.00E-06
67538912	acetyl-coenzyme A synthetase[EC6211]	4.00E-16
126133587	phosphoglucomutase[EC5422]	2.00E-07
Citrate cycle (TCA cycle)		
114328732	2-oxoglutarate dehydrogenase E1component [EC1242]	6.00E-06
	aconitase	1.00E-06
Pentose phosphate pathway		
71001262	glucose-6-phosphate isomerase[EC5319]	3.00E-09
GzCon[0258]	L-xylulose reductase	9.00E-09
Um37414661	sorbitol-utilisation protein	3.00E-11
mg[0122]	transaldolase, component ofnon-oxidative part of pentose-phosphate pathway	9.00E-08
SSPG592	transketolase	1.00E-09
SSPG741	6-phosphogluconate dehydrogenase	2.00E-12
56421765	ribokinase [EC27115]	8.00E-06
19112945	hypothetical protein [EC5422]	2.00E-07
110639997	ribose 5-phosphate isomerase B [EC5316]	8.00E-06
126133587	Phosphoglucomutase [EC5422]	2.00E-07
Fructose and mannose metabolism		
108760539	glycosyl transferase group 2 familyprotein [EC241-]	3.00E-06
83770940	fructose-6-phosphate 2-kinasefructose-26-biphosphatase	7.00E-12
Galactose metabolism		
113478008	alpha-glucosidase [EC32120]	4.00E-08
19112945	hypothetical protein[EC5422]	2.00E-07
126133587	phosphoglucomutase[EC5422]	2.00E-07

Ascorbate and aldarate metabolism		
17544880	probable transmembrane aldehyde dehydrogenase oxidoreductase protein	8.00E-06
Starch and sucrose metabolism		
83771290	glycogen synthase [EC24111]	1.00E-05
71001262	glucose-6-phosphate isomerase[EC5319]	3.00E-09
107025858	Alpha alpha-trehalase [EC32128]	1.00E-05
113478008	alpha-glucosidase [EC32120]	4.00E-08
47212375	hypothetical protein	3.00E-06
146322636	glycogen phosphorylase GlpVGph1 putative[EC2411]	8.00E-09
19112945	hypothetical protein[EC5422]	2.00E-07
83776215	14-alpha-glucan branching enzymestarchbranching enzyme II	1.00E-06
126133587	phosphoglucomutase[EC5422]	2.00E-07
67526543	hypothetical protein [EC24134]	1.00E-18
Aminosugars metabolism		
70991353	glucosamine-fructose-6-phosphateaminotransferase [EC26116]	5.00E-07
113868431	soluble lytic murein transglycosylase orrelatedregulatory protein	1.00E-05
83767858	NADH-cytochrome b-5 reductase [EC1622]	8.00E-08
Nucleotide sugars metabolism		
120402727	dTDP-4-dehydrorhamnose reductase[EC111133]	2.00E-06
15836863	dTDP-4-keto-L-rhamnose reductase[EC111133]	3.00E-06
86359939	dTDP-glucose 46-dehydratase protein[EC42146]	6.00E-06
Pyruvate metabolism		
17544880	probable transmembrane aldehydedehydrogenase oxidoreductase protein	8.00E-06
67537984	hypothetical protein[EC1241]	5.00E-12
54023581	putative pyruvatedehydrogenase E1 component [EC1241]	3.00E-06
67538912	acetyl-coenzyme A synthetase[EC6211]	4.00E-16
56478499	phenylglyoxylateacceptoroxidoreductase [EC1271]	6.00E-06
Glyoxylate and dicarboxylate metabolism		
	conserved hypothetical protein	1.00E-06
SSPG547	glutathione-dependent formaldehydedehydrogenase	6.00E-28
cal:CaO19.8252	Formate Dehydrogenase NAD-Dependant	5.00E-08
Propanoate metabolism		
17544880	probable transmembrane aldehydedehydrogenase oxidoreductase protein	8.00E-06
	enoyl-CoA hydrataseisomerase family	
118380817	enoyl-CoA hydrataseisomerase family protein [EC42117]	1.00E-05

67538912	acetyl-coenzyme A synthetase[EC6211]	4.00E-16
56478499	phenylglyoxylateacceptoroxidoreductase [EC1271]	6.00E-06
Butanoate metabolism		
17544880	probable transmembrane aldehydedehydrogenase oxidoreductase protein	8.00E-06
118380817	enoyl-CoA hydrataseisomerase family protein [EC42117]	1.00E-05
67537984	hypothetical protein[EC1241]	5.00E-12
54023581	putative pyruvatedehydrogenase E1 component [EC1241]	3.00E-06
56478499	phenylglyoxylateacceptoroxidoreductase [EC1271]	6.00E-06
70997956	short chain dehydrogenasereductase familyoxidoreductase putative	4.00E-08
1.2 Energy Metabolism		
Oxidative phosphorylation		
		8.00E-06
19112733	hypothetical protein [EC11022]	9.00E-10
75858994	ATP synthase subunit 8 [EC36314]	5.00E-08
P04037	Saccharomyces cerevisiaeYGL187c COX4	9.00E-08
Methane metabolism		
27383009	alcoholdehydrogenase [EC1111 111284]	2.00E-06
cal:CaO19.8252	Formate Dehydrogenase NAD-Dependant	5.00E-08
1.3 Lipid Metabolism		
Fatty acid metabolism		
67538756	hypothetical protein [EC111100]	1.00E-05
	hypothetical protein	9.00E-08
KOG0301	Phospholipase A2-activating protein (containsWD40 repeats)	2.00E-09
17544880	probable transmembrane aldehydedehydrogenase oxidoreductase protein	8.00E-06
27383009	Alcohol dehydrogenase [EC1111 111284]	4.00E-06
118380817	enoyl-CoA hydrataseisomerase family protein [EC42117]	1.00E-05
94967641	acyl-CoA dehydrogenase-like[EC13997]	8.00E-14
Synthesis and degradation of ketone bodies		
70997956	short chain dehydrogenasereductase familyoxidoreductase putative	4.00E-08
Biosynthesis of steroids		
118362896	myosin [EC5335]	4.00E-06
19115654	Sterol reductase/lamin B receptor	2.00E-06
119943112	7-dehydrocholesterol reductase [EC13121]	6.00E-06
73983551	similar to 7-dehydrocholesterol reductase 7-DHCreductase Sterol	3.00E-06
Glycerolipid metabolism		

17544880	probable transmembrane aldehydedehydrogenase oxidoreductase protein	8.00E-06
27383009	Alcohol dehydrogenase [EC1111 111284]	4.00E-06
146329820	glycosyl transferase family protein [EC241-]	8.00E-06
Glycerophospholipid metabolism		
15605561	CDP-diacylglycerol--serineO-phosphatidyltransferase [EC2788]	4.00E-06
Linoleic acid metabolism		
115454785	lipoxygenase [EC:1.13.11.12]	6.00E-06
1.4 Nucleotide Metabolism		
78778386	DNA polymerase III beta subunit[EC2777]	2.00E-06
86608913	polyribonucleotide nucleotidy ltransferase [EC2778]	6.00E-06
17544914	ribonucleoside-diphosphate reductase alpha chain[EC11741]	8.00E-06
145589407	tRNA pseudouridine synthase B [EC42170]	8.00E-06
15643514	uridine kinase [EC27148]	4.00E-06
PsCon[10781]	metal-dependent Rnase	2.00E-12
SSPG21	nucleoside-diphosphate kinase	8.00E-19
71891910	carbamoyl-phosphate synthase large chain[EC6355]	3.00E-06
1.5 Amino Acid Metabolism		
Glutamate metabolism		
91782644	putative succinyldiaminopimelate transaminase [EC2611]	4.00E-06
70991353	glucosamine-fructose-6-phosphateaminotransferase [EC26116]	5.00E-07
77359884	glutamate--cysteine ligase, gamma-glutamylcysteine synthetase	8.00E-06
27378288	Probable argininelysineornithine decarboxylases[EC41117	1.00E-05
71891910	carbamoyl-phosphate synthase large chain[EC6355]	3.00E-06
Alanine and aspartate metabolism		
mg[1412]	alanine glyoxylate aminotransferase(serine pyruvate aminotransferase)	8.00E-07
91782644	putative succinyldiaminopimelate transaminase [EC2611]	4.00E-06
67537984	hypothetical protein[EC1241]	5.00E-12
54023581	putative pyruvate dehydrogenase E1 component [EC1241]	3.00E-06
Glycine, serine and threonine metabolism		
15605561	CDP-diacylglycerol--serineO-phosphatidyltransferase [EC2788]	4.00E-06
72391536	threonine synthase putative [EC4231]	8.00E-06
Methionine metabolism		
70994984	methionyl-tRNA synthetase [EC61110]	1.00E-28

Cysteine metabolism		
91782644	putative succinyldiaminopimelate transaminase [EC2611]	4.00E-06
Valine, leucine and isoleucine degradation		
17544880	probable transmembrane aldehydedehydrogenase oxidoreductase protein	8.00E-06
118380817	enoyl-CoA hydrataseisomerase family protein [EC42117]	1.00E-05
Valine, leucine and isoleucine biosynthesis		
33601432	putative dihydroxy-acid dehydratase [EC4219]	1.00E-05
67537984	hypothetical protein[EC1241]	5.00E-12
54023581	putative pyruvatedehydrogenase E1 component [EC1241]	3.00E-06
Lysine Metabolism		
CtCon[0055]	dihydrodipicolinate synthase,lysine synthesis	6.00E-07
	dihydrodipicolinate reductase [EC13126]	6.00E-06
114328732	2-oxoglutarate dehydrogenase E1component [EC1242]	6.00E-06
17544880	probable transmembrane aldehydedehydrogenase oxidoreductase protein	8.00E-06
118380817	enoyl-CoA hydrataseisomerase family protein [EC42117]	1.00E-05
94967641	acyl-CoA dehydrogenase-like[EC13997]	8.00E-14
Arginine and proline metabolism		
91782644	putative succinyldiaminopimelate transaminase [EC2611]	4.00E-06
Histidine metabolism		
VD0211H07	imidazoleglycerol-phosphatedehydratase, histidine biosynthesis	5.00E-07
17544880	probable transmembrane aldehyde dehydrogenase oxidoreductase protein	8.00E-06
Tyrosine metabolism		
SSPG365	mandelate racemase, aromatic amino acid catabolism	2.00E-13
91782644	putative succinyldiaminopimelate transaminase [EC2611]	4.00E-06
27383009	Alcohol dehydrogenase [EC1111 111284]	4.00E-06
83767644	phosphoketolase [EC412-]	2.00E-06
Phenylalanine metabolism		
91782644	putative succinyl diaminopimelate transaminase [EC2611]	4.00E-06
75911029	amino acid adenylation [EC51111]	8.00E-06
Tryptophan metabolism		
114328732	2-oxoglutarate dehydrogenase E1component [EC1242]	6.00E-06
	probable transmembrane aldehyde	
17544880	probable transmembrane aldehyde dehydrogenase oxidoreductase protein	8.00E-06

118380817	enoyl-CoA hydratase isomerase family protein [EC42117]	1.00E-05
94967641	acyl-CoA dehydrogenase-like[EC13997]	8.00E-14
Phenylalanine, tyrosine and tryptophan biosynthesis		
91782644	putative succinyldiaminopimelate transaminase [EC2611]	4.00E-06
108562557	tyrosine-regulated 3-deoxy-D-arabino-heptulosonate7-phosphate	4.00E-06
15618944	dehyroquinate synthase [EC4234]	3.00E-06
1.6 Metabolism of Other Amino Acids		
beta-Alanine metabolism		
17544880	probable transmembrane aldehydedehydrogenase oxidoreductase protein	8.00E-06
118380817	enoyl-CoA hydrataseisomerase family protein [EC42117]	1.00E-05
Selenoamino acid metabolism		
70994984	methionyl-tRNA synthetase [EC61110]	1.00E-28
1.7 Glycan Biosynthesis and Metabolism		
Glycan structures - biosynthesis 2		
74318716	probable 3-deoxy-D-manno-octulosonic-acidtransferase transmembrane	9.00E-07
108760539	glycosyl transferase group 2 familyprotein [EC241-]	3.00E-06
1.9 Metabolism of Cofactors and Vitamins		
Thiamine metabolism		
AFU3248	hypothetical protein	8.00E-06
Vitamin B6 metabolism		
72391536	threonine synthase putative [EC4231]	8.00E-06
Pantothenate and CoA biosynthesis		
33601432	putative dihydroxy-acid dehydratase [EC4219]	1.00E-05
Folate biosynthesis		
27366277	branched-chain amino acid aminotransferase [EC41338]	6.00E-06
119906475	similar to human gamma-glutamyl hydrolase [EC34199]	6.00E-06
Limonene and pinene degradation		
17544880	probable transmembrane aldehydedehydrogenase oxidoreductase protein	8.00E-06
118380817	enoyl-CoA hydrataseisomerase family protein [EC42117]	1.00E-05
Alkaloid biosynthesis I		
91782644	putative succinyldiaminopimelate transaminase [EC2611]	4.00E-06
1,2-Dichloroethane degradation		
GzCon[1324]	2-nitropropane dioxygenase (nitroalkane oxidase)	8.00E-15
	probable transmembrane	
17544880	probable transmembrane aldehydedehydrogenase oxidoreductase protein	8.00E-06

Metabolism of xenobiotics by cytochrome P450		
UmCon[1240]	benzoate 4-monooxygenasecytochrome P450	1.00E-08
27383009	alcoholdehydrogenase [EC1111 111284]	4.00E-06
2. INFORMATION STORAGE AND PROCESSING		
Translation, ribosomal structure and biogenesis		
KOG3421	60S ribosomal protein L14	3.00E-08
KOG0901	60S ribosomal protein L14/L17/L23	4.00E-07
KOG3475	60S ribosomal protein L37	5.00E-16
KOG1732	60S ribosomal protein L21	2.00E-06
KOG1750	Mitochondrial/chloroplast ribosomal proteinS12	2.00E-13
KOG3291	Ribosomal protein S7	6.00E-06
KOG1779	40s ribosomal protein S27	7.00E-07
KOG0002	60s ribosomal protein L39	6.00E-06
KOG0688	Peptide chain release factor 1 (eRF1)	3.00E-12
KOG2988	60S ribosomal protein L30	4.00E-09
KOG3434	60S ribosomal protein L22	9.00E-12
KOG0004	Ubiquitin/40S ribosomal protein S27a fusion	9.00E-25
KOG1247	Methionyl-tRNA synthetase	6.00E-27
KOG1628	40S ribosomal protein S3A	2.00E-07
KOG2145	Cytoplasmic tryptophanyl-tRNA synthetase	5.00E-07
KOG0469	Elongation factor 2	4.00E-06
KOG0746	60S ribosomal protein L3 and relatedproteins	9.00E-15
KOG1742	60s ribosomal protein L15/L27	6.00E-12
KOG0402	60S ribosomal protein L37	1.00E-05
KOG0875	60S ribosomal protein L5	2.00E-15
Um34331998	translation initiation factor 3 (eIF3)	1.00E-05
mg[0975]	eukaryotic translation initiation factor1A	6.00E-12
BfCon[1283]	translation elongation factor eEF-2	2.00E-11
Mag30414692	eukaryotic translation initiationfactor eIF4G	5.00E-08
MagCon[10330a]	eukaryotic translation initiationfactor 3 subunit 6 interacting protein	3.00E-08
VD0107F06	translation elongation factor 2 (EF-2)	2.00E-11
MagCon[4208]	eukaryotic translation initiationfactor 4A (eIF-4A)	2.00E-17
Mag23356336	eukaryotic peptide chain releasefactor subunit 1	3.00E-14
KOG0327	Translation initiation factor 4F, helicasesubunit (eIF-4A) and related helicases	3.00E-12
VD0211H09	chaperonin, T-complex protein	7.00E-11
Gz22503933	protein disulphide isomerase	2.00E-16
	subunit beta of the	
BgCon[0990]	subunit beta of the cytosolicchaperonin Cct ring complex	5.00E-16

GzCon[2746]	NifU-like iron-sulphur clusterassembly protein, iron homeostasis	1.00E-19
VD0106H02	heat shock protein 80	2.00E-07
CtCon[0214]	heat shock protein 78,chaperonin	1.00E-07
MagCon[10827a]	heat shock protein 70	2.00E-07
Mag30417250	peroxisomal targeting signal receptor (Peroxin-5) (PTS1 receptor)	9.00E-12
RNA processing and modification		
TWOG0967	Mitochondrial mRNA maturase/Homingendonuclease	1.00E-10
KOG0331	ATP-dependent RNA helicase	2.00E-08
VD0209A07	transcription elongation complex subunit,global regulator of transcription	6.00E-08
Ps38115386	transcription factor	7.00E-12
GzCon[6207]	regulatory protein	1.00E-08
BfCon[1177]	pre-mRNA splicing factor	4.00E-09
GzCon[7413]	RNA polymerase II transcription factor	9.00E-23
MagCon[11562a]	DNA helicase, chromatin modeling	3.00E-09
MagCon[8173a]	QDE2 homologue, RNA interference(RNAi), mechanism through which double-stranded RNAssilence cognate genes by mRNA	5.00E-10
VD0105B12	11-kDa nonhistone chromosomal protein,involved in transcriptional activation of a number ofgenes	4.00E-09
KOG2441	mRNA splicing factor/probable chromatinbinding snw family nuclear protein	3.00E-07
TWOG0967	Mitochondrial mRNA maturase/Homingendonuclease	5.00E-08
KOG1098	Putative SAM-dependent rRNA methyltransferaseSPB1	9.00E-06
KOG0119	Splicing factor 1/branch point bindingprotein (RRM superfamily)	7.00E-06
TWOG0658	Mitochondrial mRNA maturase encoded bypartially processed COB mRNA	7.00E-13
KOG1781	Small Nuclear ribonucleoprotein splicing factor	1.00E-07
Transcription		
KOG0216	RNA polymerase I, second largest subunit	8.00E-13
KOG0015	Regulator of arginine metabolism and relatedMADS box-containing transcription factors	5.00E-09
KOG0668	Casein kinase II, alpha subunit	3.00E-06
KOG1952	Transcription factor NF-X1, contains NFX-typeZn2+-binding and R3H domains	2.00E-10
Replication, recombination and repair		

TWOG0967	Mitochondrial mRNA maturase/Homingendonuclease	1.00E-10
KOG0388	SNF2 family DNA-dependent ATPase	3.00E-06
Chromatin structure and dynamics		
KOG2441	mRNA splicing factor/probable chromatinbinding snw family nuclear protein	3.00E-07
KOG3467	Histone H4	7.00E-17
3. CELLULAR PROCESSES AND SIGNALING		
Cell cycle control, cell division, chromosome partitioning		
KOG0668	Casein kinase II, alpha subunit	3.00E-06
KOG2166	Cullins	1.00E-12
KOG0590	Checkpoint kinase and related serine/threonineprotein kinases	1.00E-07
KOG0987	DNA helicase PIF1/RRM3	6.00E-06
Signal transduction mechanisms		
KOG0034	Ca2+/calmodulin-dependent proteinphosphatase (calcineurin subunit B), EF-Hand superfamilyprotein	9.00E-12
KOG1435	Sterol reductase/lamin B receptor	1.00E-07
KOG0668	Casein kinase II, alpha subunit	3.00E-06
KOG2675	Adenylate cyclase-associated protein(CAP/Srv2p)	5.00E-06
KOG0660	Mitogen-activated protein kinase	3.00E-10
KOG3348	BolA (bacterial stress-inducedmorphogen)-related protein	2.00E-06
KOG1354	Serine/threonine protein phosphatase 2A,regulatory subunit	1.00E-23
KOG4076	Regulator of ATP-sensitive K+ channelsAlpha-endosulfine/ARPP-19 and related cAMP-regulatedphosphoproteins	3.00E-11
KOG0598	Ribosomal protein S6 kinase and relatedproteins	8.00E-10
KOG3417	Ras1 guanine nucleotide exchange factor	9.00E-07
KOG2265	Nuclear distribution protein NUDC	6.00E-08
Cell wall/membrane/envelope biogenesis		
KOG1268	Glucosamine 6-phosphate synthetases, containamidotransferase and phosphosugar isomerase domains	7.00E-06
KOG3396	Glucosamine-phosphate N-acetyltransferase	1.00E-06
KOG0916	1,3-beta-glucan synthase/callose synthasecatalytic subunit	8.00E-16
Cytoskeleton		
PsCon[0019]	actin	1.00E-07
KOG2675	Adenylate cyclase-associated protein (CAP/Srv2p)	5.00E-06
KOG1755	Profilin	2.00E-13
KOG0303	Actin-binding protein Coronin, containsWD40 repeats	6.00E-06
KOG1735	Actin depolymerizing factor	6.00E-16
MagCon[8326a]	peroxisome assembly protein,peroxin-1	3.00E-08
mg[0214]	chitin synthase	8.00E-07
mgc19f06f	coronin, actin-binding protein	6.00E-10
MagCon[9590a]	required for F-actin regulation	8.00E-19
Lm13259633	profilin (actin-binding protein)	5.00E-09
	cofilin, cortical cytoskeletoncomponent, actin	

KOG0929	Guanine nucleotide exchange factor	3.00E-11
KOG0070	GTP-binding ADP-ribosylation factor Arf1	4.00E-13
KOG4076	Regulator of ATP-sensitive K+ channelsAlpha-endosulfine/ARPP-19 and related cAMP-regulatedphosphoproteins	3.00E-11
KOG1107	Membrane coat complex Retromer, subunit VPS35	7.00E-12
Posttranslational modification, protein turnover, chaperones		
KOG1864	Ubiquitin-specific protease	2.00E-10
KOG3359	Dolichyl-phosphate-mannose:protein O-mannosyltransferase	3.00E-09
KOG0190	Protein disulfide isomerase (prolyl4-hydroxylase beta subunit)	1.00E-11
KOG0362	Chaperonin complex component, TCP-1 thetasubunit (CCT8)	1.00E-05
KOG0544	FKBP-type peptidyl-prolyl cis-trans isomerase	5.00E-06
KOG0177	20S proteasome, regulatory subunit beta typePSMB2/PRE1	1.00E-06
KOG0363	Chaperonin complex component, TCP-1 betasubunit (CCT2)	6.00E-11
KOG2292	Oligosaccharyltransferase, STT3 subunit	5.00E-06
KOG2062	26S proteasome regulatory complex, subunitRPN2/PSMD1	4.00E-09
KOG0841	Multifunctional chaperone (14-3-3 family)	6.00E-15
KOG0687	26S proteasome regulatory complex, subunitRPN7/PSMD6	3.00E-06
KOG0001	Ubiquitin and ubiquitin-like proteins	2.00E-10
KOG1439	RAB proteins geranylgeranyltransferasecomponent A (RAB escort protein)	3.00E-07
KOG0182	20S proteasome, regulatory subunit alpha typePSMA6/SCL1	4.00E-15
KOG1051	Chaperone HSP104 and related ATP-dependentClp proteases	2.00E-06
KOG0001	Ubiquitin and ubiquitin-like proteins	2.00E-07
KOG1769	Ubiquitin-like proteins	6.00E-08
SSPG264F	alpha-1,2-galactosyltransferase	1.00E-06
Um37401875	Oligosaccharyl transferase	8.00E-10
Ps38064043	p70 ribosomal protein S6 kinase	2.00E-06
mgc04g06f	dolichyl phosphate-D-mannose:proteinO-D-mannosyltransferase	4.00E-21
BfCon[1949]	ubiquitin-like protein, homologue of humanSUMO-1 protein	3.00E-13
Transport facilitation		

Transposon insertion sequence proteins		
Gz22509405	intron derivedmaturase	1.00E-06
mg[0007]	Retrotransposon polyprotein	6.00E-07
Gz22509405	intron derivedmaturase	3.00E-11

Appendix Table 3.Putative identification and classification of EST's from spore mats based on blast homology searches in KEGG, COG and COGEME databases.

Spores		
1.1 Carbohydrate Metabolism		
Glycolysis / Gluconeogenesis		
Tbd_0162	pyruvate kinase [EC27140]	1.00E-05
mga0634f	glyceraldehyde-3-phosphatedehydrogenase	3.00E-07
VD0202C03	Phosphoenolpyruvate carboxykinase, rate limiting gluconeogenic enzyme	2.00E-12
mg[0595]	phosphoglycerate kinase	2.00E-10
MagCon[0423]	hexokinase	4.00E-06
AN57462	enolase [EC42111]	5.00E-11
Pnap_0655	glucose-6-phosphate isomerase [EC5319]	8.00E-06
Citrate cycle (TCA cycle)		
AFUA_6G07720	Phosphoenol pyruvate carboxy kinase, AcuF [EC41149]	7.00E-10
Rru_A1213	2-oxoglutarate dehydrogenase E1component [EC1242]	8.00E-06
p2A13	2-ketoglutarate NADP oxidoreductase alpha subunit [EC1273]	3.00E-06
NCU030312	succinate dehydrogenase membrane anchor subunit	1.00E-09
Pentose phosphate pathway		
Mmar10_2728	gluconolactonase [EC31117]	6.00E-06
Pnap_0655	glucose-6-phosphate isomerase[EC5319]	8.00E-06
Fructose and mannose metabolism		
DNO_0097	glycosyl transferase family protein [EC241-]	1.00E-06
RHA1_ro02811	fructokinase [EC2714]	3.00E-06
DNO_0097	glycosyl transferase family protein [EC241-]	6.00E-06
AFUA_3G09190	aldehyde reductase I ARI putative [EC111-]	1.00E-13
AFUA_3G03940	23-diketo-5-methylthio-1-phosphopentane phosphatase putative	3.00E-09
lin2813	similar to sorbitol dehydrogenase [EC11114]	6.00E-06
AO090003000189	GDP-mannose pyrophosphorylase [EC27713]	5.00E-13
CPS_2648	carbohydrate kinase PfkB family [EC2714]	3.00E-06

MagCon[0916a]	D-arabinitol dehydrogenase	4.00E-06
Um37414661	sorbitol-utilisation protein	5.00E-06
Galactose metabolism		
AFUA_3G09190	aldehyde reductase I ARI putative [EC111-]	1.00E-13
Ascorbate and aldarate metabolism		
Mmar10_2728	gluconolactonase [EC31117]	6.00E-06
Starch and sucrose metabolism		
RHA1_ro02811	fructokinase [EC2714]	3.00E-06
AT1G27680	APL2 large subunit of AGP 2 [EC27727]	2.00E-06
Acel_1821	glucose-1-phosphate adenylyl transferase [EC27727]	6.00E-06
PSEEN2046	14-alpha-glucan branching enzyme [EC24118]	6.00E-10
Pnap_0655	glucose-6-phosphate isomerase[EC5319]	8.00E-06
CPS_2648	carbohydrate kinase PfkB family [EC2714]	3.00E-06
Aminosugars metabolism		
Dmel_CG18140	Chitinase 3 [EC32114]	6.00E-06
AFUA_2G13430	chitin synthase putative [EC24116]	5.00E-07
AN04322	hypothetical protein [EC1622]	3.00E-06
AFUA_3G03940	23-diketo-5-methylthio-1-phosphopentane phosphatase putative	3.00E-09
Nucleotide sugars metabolism		
Mvan_1727	dTDP-4-dehydrorhamnose reductase[EC111133]	4.00E-07
Pyruvate metabolism		
Tbd_0162	pyruvate kinase [EC27140]	1.00E-05
BT_3693	acetate kinase [EC2721]	3.00E-06
Adeh_1984	hydroxyacylglutathione hydrolase [EC3126]	8.00E-06
AFUA_6G07720	Phosphoenol pyruvate carboxykinase, AcuF [EC41149]	7.00E-10
Bcep18194_B2932	2-isopropylmalate synthase [EC23313]	6.00E-06
Glyoxylate and dicarboxylate metabolism		
RSP_1821	molybdopterin-containing oxidoreductase probable formate	8.00E-06
Propanoate metabolism		
BT_3693	acetate kinase [EC2721]	3.00E-06
BBta_7124	Putative long-chain-fatty-acid--CoA ligase long-chain acyl-CoA	3.00E-06
150989	Hypothetical protein	1.00E-05
Butanoate metabolism		
AFUA_5G00960	feruloyl esterase putative [EC311-]	4.00E-06
150989	Hypothetical protein	1.00E-05
117147	acyl-CoA synthetase medium-chain family	6.00E-06

	member 1[EC6212]	
Noc_2520	acetolactate synthase large subunitbiosynthetic type [EC2216]	1.00E-05
AFUA_3G09190	aldehyde reductase I ARI putative [EC111-]	1.00E-13
Bxe_A3826	glutamate decarboxylase[EC41115]	9.00E-07
Inositol phosphate metabolism		
444238	MGC80809 protein [EC27168]	4.00E-06
Acid345_2392	inositol-3-phosphate synthase [EC5514]	8.00E-06
AFUA_3G10530	protein serine threonine kinase Ran1 putative [EC271-]	1.00E-07
1.2 Energy Metabolism		
Oxidative phosphorylation		
AO090010000482	F0F1-type ATP synthase beta subunit [EC36314]	1.00E-15
BBta_4549	NADH-quinone oxidoreductase chain M [EC169951653]	4.00E-06
RoseRS_2230	proton-translocating NADH-quinone oxidoreductasechain N	8.00E-06
NCU023732		8.00E-06
YBL099W	Alpha subunit of the F1 sector of mitochondrial F1F0ATP synthase	7.00E-10
NCU030312		1.00E-09
Sde_2107	cytidylate kinase [EC1351]	2.00E-14
Cag_0643	proton-translocating NADH-quinone oxidoreductase chain M	1.00E-05
Methane metabolism		
RPE_0253	Catalase peroxidase HPI [EC11116]	1.00E-05
RSP_1821	molybdopterin-containing oxidoreductase probable formate	8.00E-06
Nitrogen metabolism		
YALI0F21406g	O94255 *Schizosaccharomyces pombe* Carbonic anhydrase [EC:4.2.1.1]	2.00E-12
1.3 Lipid Metabolism		
Fatty acid metabolism		
AGOS_AFL138W	fatty-acyl-CoA synthase, subunit alpha [EC:2.3.1.86]	7.00E-11
HCH_05141	short-chain alcohol dehydrogenase-like protein [EC111100]	1.00E-05
BfCon[0166]	acyl-CoA synthetase (long-chain fattyacid CoA ligase)	1.00E-20
PsCon[0448]	propionyl-CoA carboxylase alpha	3.00E-11
MagCon[4018]	Peroxisomal hydratase-dehydrogenase-epimerase (HDE) (multifunctional beta-oxidation protein)	5.00E-10
Bpro_4161	AMP-dependent synthetase and ligase [EC6213]	1.00E-08
Biosynthesis of steroids		
Dgeo_0180	4-diphosphocytidyl-2C-methyl-D-erythritol kinase[EC271148]	3.00E-06

YALI0D19206g	P36209 *Schizosaccharomyces pombe* delta 24 (24(1))-sterol reductase [EC:1.3.1.71]	2.00E-16
Glycerolipid metabolism		
511277370	phosphatidic acid phosphatase [EC3134]	6.00E-06
DNO_0097	glycosyl transferase family protein [EC241-]	1.00E-06
Glycerophospholipid metabolism		
511277370	phosphatidic acid phosphatase [EC3134]	6.00E-06
YALI0D03480g	Q872A4 *Neurospora crassa* phosphatidyl serine decarboxylase [EC: 4.1.1.65]	1.00E-07
Linoleic acid metabolism		
AFUA_3G09190	aldehyde reductase I ARI putative [EC111-]	1.00E-13
1.4 Nucleotide Metabolism		
Purine and Pyrimidine metabolism		
Tbd_0162	pyruvate kinase [EC27140]	1.00E-05
MK1649	glutamine phosphoribosylpyrophosphate amidotransferase [EC24214]	8.00E-06
STH1512	DNA polymerase III alpha subunit[EC2777]	1.00E-05
AN68152	hypothetical protein [EC3543]	3.00E-06
Kwal_23216	hypothetical protein	4.00E-07
SPAC14403	hypothetical protein [EC6344]	1.00E-23
BB4400	Diadenosine tetraphosphatase [EC36141]	8.00E-06
NCU078532	uricase (urate oxidase)	1.00E-12
NCU033502	xanthine dehydrogenase	2.00E-14
mg[1246]	nucleoside diphosphate kinase	4.00E-13
PsCon[10781]	metal-dependent Rnase	4.00E-14
HQ2647A	uridine phosphorylase [EC2423]	1.00E-05
TM0751	uridine kinase [EC27148]	1.00E-06
1.5 Amino Acid Metabolism		
Glutamate metabolism		
MK1649	glutamine phosphoribosyl pyrophosphate amidotransferase [EC24214]	8.00E-06
TM1040_0865	aminotransferase class I and II [EC2611]	3.00E-06
150989	K08231 MFS transporter, MCP family, solute carrier family 16	1.00E-05
Bxe_A3826	glutamate decarboxylase [EC41115]	9.00E-07
Alanine and aspartate metabolism		
TM1040_0865	aminotransferase class I and II [EC2611]	3.00E-06
MG_036	aspartyl-tRNA synthetase [EC61112]	6.00E-06
150989	K08231 MFS transporter, MCP family, solute carrier family 16	1.00E-05
SPAC14403	hypothetical protein [EC6344]	1.00E-23
Bxe_A3826	glutamate decarboxylase [EC41115]	9.00E-07

Glycine, serine and threonine metabolism		
YALI0F13453g	P17423 *Saccharomyces cerevisiae* phosphatidylserine decarboxylase [EC:4.1.1.65]	2.00E-15
IL1708	L-serine deaminase [EC43117]	8.00E-06
YALI0D03480g	Q872A4 Neurospora crassaCAD70830 CAD70830	1.00E-07
AFUA_3G09190	aldehyde reductase I ARI putative [EC111-]	1.00E-13
Pcal_1612	serine--glyoxylate transaminase [EC26145]	3.00E-06
SACOL1362	homoserine dehydrogenase [EC1113]	6.00E-06
Cysteine metabolism		
TM1040_0865	aminotransferase class I and II [EC2611]	3.00E-06
IL1708	L-serine deaminase [EC43117]	8.00E-06
Valine, leucine and isoleucine biosynthesis		
15795	Hypothetical protein	1.00E-05
Noc_2520	acetolactate synthase large subunit biosynthetic type [EC2216]	1.00E-05
Bcep18194_B2932	2-isopropylmalate synthase [EC23313]	6.00E-06
Lysine biosynthesis		
CtCon[0055]	dihydrodipicolinate synthase,lysine synthesis	2.00E-07
SACOL1362	homoserine dehydrogenase [EC1113]	6.00E-06
Lysine degradation		
AFUA_3G09190	aldehyde reductase I ARI putative [EC111-]	1.00E-13
Rru_A1213	2-oxoglutarate dehydrogenase E1component [EC1242]	8.00E-06
Arginine and proline metabolism		
TM1040_0865	aminotransferase class I and II [EC2611]	3.00E-06
Tyrosine metabolism		
SSPG168F	2-hydroxyhepta-2,4-diene-1,7-dioateisomerase	3.00E-07
TM1040_0865	aminotransferase class I and II [EC2611]	3.00E-06
Phenylalanine metabolism		
TM1040_0865	aminotransferase class I and II [EC2611]	3.00E-06
Tryptophan metabolism		
RPE_0253	Catalase peroxidase HPI [EC11116]	1.00E-05
Rru_A1213	2-oxoglutarate dehydrogenase E1component [EC1242]	8.00E-06
Phenylalanine, tyrosine and tryptophan biosynthesis		
TM1040_0865	aminotransferase class I and II [EC2611]	3.00E-06
1.6 Metabolism of Other Amino Acids		
beta-Alanine metabolism		
575768	similar to MGC81821 protein [EC1312]	8.00E-06
150989	Hypothetical protein	1.00E-05

Bxe_A3826	glutamate decarboxylase[EC41115]	9.00E-07
Taurine and hypotaurine metabolism		
BT_3693	acetate kinase [EC2721]	3.00E-06
Bxe_A3826	glutamate decarboxylase[EC41115]	9.00E-07
Glycan structures - biosynthesis 2		
AFUA_6G05260	transferase Gpi7 putative [EC27--]	4.00E-16
AN64962	hypothetical protein [EC27--]	1.00E-07
1.9 Metabolism of Cofactors and Vitamins		
Thiamine metabolism		
AFUA_3G03940	23-diketo-5-methylthio-1-phosphopentane phosphatase putative	3.00E-09
Riboflavin metabolism		
UmCon[0365]	B2-aldehyde-forming enzyme, riboflavin 5'-hydroxymethyl oxidation	2.00E-07
Mag30412750	3,4-dihydroxy-2-butanone4-phosphate synthase, riboflavin biosynthesis	1.00E-06
MagCon[2773]	GTP cyclohydrolase; riboflavin biosynthesis	1.00E-12
AFUA_3G03940	23-diketo-5-methylthio-1-phosphopentane phosphatase putative	3.00E-09
Nicotinate and nicotinamide metabolism		
AFUA_3G03940	23-diketo-5-methylthio-1-phosphopentane phosphatase putative	3.00E-09
AFUA_3G10530	protein serine/ threonine kinaseRan1 putative [EC271-]	1.00E-07
Pantothenate and CoA biosynthesis		
575768	similar to MGC81821 protein [EC1312]	8.00E-06
Noc_2520	acetolactate synthase large subunit biosynthetic type [EC2216]	1.00E-05
Sden_0156	pantothenate kinase [EC27133]	6.00E-06
Limonene and pinene degradation		
BBta_7124	putativelong-chain-fatty-acid--CoA ligase long-chain acyl-CoA	3.00E-06
Alkaloid biosynthesis II		
AFUA_5G00960	feruloyl esterase putative [EC311-]	4.00E-06
BBta_7124	Putative long-chain-fatty-acid--CoA ligase long-chain acyl-CoA	3.00E-06
Streptomycin biosynthesis		
Mvan_1727	dTDP-4-dehydrorhamnose reductase[EC111133]	4.00E-07
DET0979	myo-inositol-1-phosphate synthase family protein [EC5514]	6.00E-06
Acid345_2392	inositol-3-phosphate synthase [EC5514]	8.00E-06
Novobiocin biosynthesis		

TM1040_0865	aminotransferase class I and II [EC2611]	3.00E-06
Caprolactam degradation		
Mmar10_2728	gluconolactonase [EC31117]	6.00E-06
BBta_7124	Putative long-chain-fatty-acid--CoA ligase long-chain acyl-CoA	3.00E-06
2. INFORMATION STORAGE AND PROCESSING		
Translation, ribosomal structure and biogenesis		
KOG3504	60S ribosomal protein L29	1.00E-09
KOG3506	40S ribosomal protein S29	4.00E-08
KOG0004	Ubiquitin/40S ribosomal protein S27a fusion	6.00E-21
KOG1779	40s ribosomal protein S27	4.00E-23
KOG3411	40S ribosomal protein S19	4.00E-08
KOG1762	60s acidic ribosomal protein P1	2.00E-06
KOG0875	60S ribosomal protein L5	4.00E-10
KOG1732	60S ribosomal protein L21	1.00E-09
KOG0901	60S ribosomal protein L14/L17/L23	2.00E-09
KOG2298	Glycyl-tRNA synthetase and related class IItRNA synthetase	2.00E-15
KOG3406	40S ribosomal protein S12	4.00E-09
KOG0002	60s ribosomal protein L39	3.00E-13
KOG0378	40S ribosomal protein S4	1.00E-19
KOG0052	Translation elongation factor EF-1 alpha/Tu	4.00E-14
KOG0887	60S ribosomal protein L35A/L37	2.00E-10
KOG3486	40S ribosomal protein S21	7.00E-12
KOG0469	Elongation factor 2	3.00E-10
KOG0746	60S ribosomal protein L3 and relatedproteins	2.00E-07
KOG2988	60S ribosomal protein L30	7.00E-12
KOG3464	60S ribosomal protein L44	4.00E-06
KOG3283	40S ribosomal protein S8	3.00E-09
KOG3464	60S ribosomal protein L44	3.00E-15
KOG1694	60s ribosomal protein L6	2.00E-07
KOG0898	40S ribosomal protein S15	6.00E-11
KOG3486	40S ribosomal protein S21	3.00E-11
KOG0009	Ubiquitin-like/40S ribosomal S30 proteinfusion	8.00E-13
KOG1765	Regulator of ribosome synthesis	8.00E-08
KOG3464	60S ribosomal protein L44	2.00E-26
KOG3280	Mitochondrial/chloroplast ribosomal proteinL17	8.00E-08
KOG0878	60S ribosomal protein L32	1.00E-08
KOG3181	40S ribosomal protein S3	4.00E-14
KOG0900	40S ribosomal protein S20	7.00E-09
KOG3475	60S ribosomal protein L37	4.00E-14
mga0457f	translation elongation factor 3 (EF-3)	1.00E-12
Ct21906489	translation elongation factor 2 (EF-2)	1.00E-20

Mag23356336	eukaryotic peptide chain releasefactor subunit 1	9.00E-14
VD0109E12	eukaryotic translation initiation factor1A	3.00E-09
RNA processing and modification		
KOG1780	Small Nuclear ribonucleoprotein G	1.00E-11
KOG1774	Small nuclear ribonucleoprotein E	3.00E-06
MagCon[2278a]	mRNA splicing factor	8.00E-06
MagCon[10359a]	11-kDa nonhistone chromosomalprotein, involved in transcriptional activation of anumber of genes	3.00E-15
Ps38115386	transcription factor	1.00E-17
PsCon[6247]	chromatin remodeling factor	4.00E-07
Gz48690244	splicing factor 3B subunit 1	1.00E-21
CpCon[0312]	single stranded DNA binding protein	4.00E-09
MagCon[6948]	11-kDa nonhistone chromosomal protein,involved in transcriptional activation of a number ofgenes	1.00E-10
KOG3448	Predicted snRNP core protein	1.00E-06
KOG0343	RNA Helicase	1.00E-07
KOG0951	RNA helicase BRR2, DEAD-box superfamily	1.00E-07
KOG0331	ATP-dependent RNA helicase	7.00E-22
KOG3167	Box H/ACA snoRNP component, involved inribosomal RNA pseudouridinylation	1.00E-06
KOG1780	Small Nuclear ribonucleoprotein G	8.00E-11
TWOG0967	Mitochondrial mRNA maturase/Homingendonuclease	1.00E-09
KOG3459	Small nuclear ribonucleoprotein (snRNP) Smcore protein	6.00E-10
Gz47835873	rRNA processing protein	3.00E-06
SSPG371	HMG-like nuclear protein, rRNA modificationand processing	1.00E-07
KOG3218	RNA polymerase, 25-kDa subunit (common topolymerases I, II and III)	2.00E-07
KOG3490	Transcription elongation factor SPT4	2.00E-08
KOG3473	RNA polymerase II transcription elongationfactor Elongin/SIII, subunit elongin C	1.00E-06
Trnscrption developmental		
Gz15771564	homologue of UMTA (E. nidulans),methyltransferase, negatively regulates sexualdevelopment	2.00E-09
Replication, recombination and repair		
TWOG0967	Mitochondrial mRNA maturase/Homingendonuclease	1.00E-09
KOG0958	DNA damage-responsive repressor GIS1/RPH1,jumonji superfamily	2.00E-07
KOG1942	DNA helicase, TBP-interacting protein	3.00E-20
Chromatin structure and dynamics		
KOG3467	Histone H4	6.00E-35
KOG1745	Histones H3 and H4	2.00E-14

KOG1756	Histone 2A	3.00E-09
KOG0933	Structural maintenance of chromosome protein 2(chromosome condensation complex Condensin, subunit E)	4.00E-07
2. CELLULAR PROCESSES AND SIGNALING		
Cell cycle control, cell division, chromosome partitioning		
KOG0933	Structural maintenance of chromosome protein 2(chromosome condensation complex Condensin, subunit E)	4.00E-07
KOG0018	Structural maintenance of chromosomeprotein 1 (sister chromatid cohesion complex Cohesin,subunit SMC1)	2.00E-08
Gz40384617	cell cycle inhibitor	3.00E-12
BfCon[1590]	DNA polymerase delta accessory protein (PCNA)	8.00E-10
KOG3694	Protein required for meiosis	7.00E-14
Signal transduction mechanisms		
KOG0078	GTP-binding protein SEC4, small G protein superfamily, and related Ras family GTP-binding proteins	3.00E-07
KOG0789	Protein tyrosine phosphatase	1.00E-06
KOG2125	Glycosyl phosphatidylinositol anchor synthesis protein	2.00E-07
KOG1435	Sterol reductase/lamin B receptor	1.00E-15
KOG4476	Gluconate transport-inducing protein	2.00E-08
KOG0748	Predicted membrane proteins, containhemolysin III domain	6.00E-08
KOG4369	RTK signaling protein MASK/UNC-44	1.00E-07
KOG0583	Serine/threonine protein kinase	2.00E-08
Cell wall/membrane/envelope biogenesis		
KOG1460	GDP-mannose pyrophosphorylase	2.00E-10
Cytoskeleton		
KOG1876	Actin-related protein Arp2/3 complex, subunitARPC4	5.00E-08
KOG1755	Profilin	7.00E-17
KOG1057	Arp2/3 complex-interacting protein VIP1/Asp1,involved in regulation of actin cytoskeleton	9.00E-12
KOG1735	Actin depolymerizing factor	2.00E-07
KOG0676	Actin and related proteins	4.00E-09
PsCon[10845]	loricrin-like protein, extracellular matrix	3.00E-06
KOG0239	Kinesin (KAR3 subfamily)	9.00E-11
KOG1654	Microtubule-associated anchor protein involved in autophagy and membrane trafficking	2.00E-08
Intracellular trafficking, secretion, and vesicular transport		
KOG0078	GTP-binding protein SEC4, small G protein superfamily, and related Ras family GTP-binding proteins	3.00E-07
KOG1691	emp24/gp25L/p24 family of membrane trafficking proteins	2.00E-09

KOG4580	Component of vacuolar transporter chaperone (Vtc) involved in vacuole fusion	7.00E-06
KOG3343	Vesicle coat complex COPI, zeta subunit	7.00E-06
KOG0861	SNARE protein YKT6, synaptobrevin/ VAMP superfamily	7.00E-13
KOG0985	Vesicle coat protein clathrin, heavy chain	3.00E-08
KOG0070	GTP-binding ADP-ribosylation factor Arf1	2.00E-25
KOG0096	GTPase Ran/TC4/GSP1 (nuclear proteintransport pathway), small G protein superfamily	7.00E-07
KOG4097	Succinate dehydrogenase membrane anchor subunit and related proteins	3.00E-07
KOG1985	Vesicle coat complex COPII, subunitSEC24/subunit SFB2	6.00E-06
W0AA035ZC05C1	GTPase-activator protein ofRab-like small GTPases	9.00E-11
MagCon[9838a]	exocyst complex component Sec15, protein involved in vesicle traffic between Golgi and plasma membrane	1.00E-13
VD0101A05	dynamin like protein, endocytosis	4.00E-06
CpCEST-23-E-10	vesicular integral-membrane protein	6.00E-08
Um37406272	member of the AAA family of ATPases, required for export of pre-ribosomal large subunits from the nucleus	3.00E-12
Mag30402679	actin-related protein(centractin)	6.00E-06
SSPG431	component of translocase of the outer mitochondrial membrane (TOM) complex	1.00E-08
Posttranslational modification, protein turnover, chaperones		
KOG0182	20S proteasome, regulatory subunit alpha typePSMA6/SCL1	5.00E-12
KOG0001	Ubiquitin and ubiquitin-like proteins	2.00E-29
KOG1641	Mitochondrial chaperonin	7.00E-06
KOG0549	FKBP-type peptidyl-prolyl cis-trans isomerase	2.00E-18
KOG1867	Ubiquitin-specific protease	2.00E-08
KOG4580	Component of vacuolar transporter chaperone (Vtc) involved in vacuole fusion	7.00E-06
KOG0001	Ubiquitin and ubiquitin-like proteins	2.00E-09
KOG0726	26S proteasome regulatory complex, ATPaseRPT2	7.00E-13
KOG2012	Ubiquitin activating enzyme UBA1	5.00E-08
KOG0009	Ubiquitin-like/40S ribosomal S30 proteinfusion	8.00E-13
KOG0100	Molecular chaperones GRP78/BiP/KAR2, HSP70superfamily	9.00E-09
KOG0730	AAA+-type ATPase	1.00E-22
KOG0714	Molecular chaperone (DnaJ superfamily)	3.00E-10
KOG1460	GDP-mannose pyrophosphorylase	2.00E-10
KOG3359	Dolichyl-phosphate-	6.00E-13

	mannose:protein O-mannosyltransferase	
KOG2581	26S proteasome regulatory complex, subunitRPN3/PSMD3	3.00E-11
KOG2100	Dipeptidyl aminopeptidase	1.00E-11
Bg27453551	regulator of ribosome synthesis	2.00E-12
mg[0368]	involved in cytochrome c oxidaseassembly	5.00E-12
MagCon[10827a]	heat shock protein 70	2.00E-07
VD0106H02	heat shock protein 80	3.00E-07
GzCon[5678]	cyclophilin-like peptidylprolyl cis-trans isomerase	2.00E-07
GzCon[1540]	ER protein involved inresponse to unfolded proteins	2.00E-12
GzCon[5678]	cyclophilin-like peptidylprolyl cis-trans isomerase	4.00E-06
CpCEST-53-C-03	vacuolar sortingprotein	2.00E-13
PsCon[0987]	involved in intramitochondrial sorting	2.00E-06
Transport facilitation		
Mag30403206	MFS transporter	1.00E-07
VD0200N16	amino acid permease	3.00E-12
Um37410201	sodium P-type ATPase	1.00E-06
GzCon[3208]	peroxisome membrane protein 47,carrier protein	7.00E-07
VD0105A01	plasma membrane H+ ATPase	1.00E-12
GzCon[2738]	acyl CoA binding protein	3.00E-22
GzCon[7100]	MFS-multidrug-resistancetransporter	1.00E-07
FsCon[0022]	amino acid permease	3.00E-09
CtCon[0287]	glucose-6-phosphate/phosphate-translocator	1.00E-06
Fs14664947	phosphate transport protein	2.00E-11
MagCon[4569]	ABC transporter	3.00E-26
W0AA005ZH12C1	ABC transporter	3.00E-06
BgCon[0931]	vacuolar ATPase V0 domainsubunit	1.00E-08
CpCEST-56-E-05	peroxisome membrane protein47, carrier protein	2.00E-15
GzCon[3445]	plasma membrane ATPase (proton pump)	1.00E-11
FsCon[0278]	plasma membrane H+ ATPase	7.00E-08
Transposon insertion sequence proteins		
Ps38056377	retroelement polpolyprotein	8.00E-16
Cell division mating determination		
UmCon[0408]	heterokaryonincompatibility, het-c	2.00E-16

Appendix Table 4. Putative identification and classification of EST's from the carbon, nitrogen starved mycelia based on blast homology searches in KEGG, COG and COGEME databases.

C/N Metabolism		
1.1 Carbohydrate Metabolism		
Glycolysis / Gluconeogenesis		
KOG2670	Enolase	2.00E-17
AO090003000725	fructose 16-bisphosphatealdolase [EC41213]	4.00E-09
Citrate cycle (TCA cycle)		
AO090005001404	NADP-dependent isocitratedehydrogenase [EC11142]	3.00E-09
Rru_A1213	2-oxoglutarate dehydrogenase E1component [EC1242]	2.00E-06
Pentose phosphate pathway		
SSPG967F	D-xylose reductase	2.00E-07
AO090003000725	fructose 16-bisphosphatealdolase [EC41213]	4.00E-09
CFF8240_1474	ribose-phosphate pyrophosphokinase [EC2761]	1.00E-05
BPSL1830	putative ribokinase [EC27115]	3.00E-06
Pentose and glucuronate interconversions		
Lxx03360	xylulose kinase [EC27117]	4.00E-06
Fructose and mannose metabolism		
AO090003000725	fructose 16-bisphosphatealdolase [EC41213]	4.00E-09
DNO_0097	glycosyl transferase family protein [EC241-]	2.00E-06
AN45912	hypothetical protein [EC5428]	4.00E-06
Galactose metabolism		
str1735	sucrose-6-phosphate hydrolase [EC32126]	3.00E-06
Tery_4624	alpha-glucosidase [EC32120]	1.00E-07
TTE0273	Galactose-1-phosphate uridylyltransferase[EC27710]	8.00E-06
Starch and sucrose metabolism		
Bcen_3502	Alpha alpha-trehalase [EC32128]	4.00E-06
str1735	sucrose-6-phosphate hydrolase [EC32126]	3.00E-06
10197		2.00E-06
KOG0470	1,4-alpha-glucan branching enzyme/starch branching enzyme II	7.00E-11
Tery_4624	alpha-glucosidase [EC32120]	1.00E-07

Aminosugars metabolism		
F59B23	F59B23 [EC35125]	8.00E-06
Nucleotide sugars metabolism		
XF0258	dTDP-4-keto-L-rhamnose reductase[EC111133]	8.00E-06
Pcryo_0624	dTDP-glucose 46-dehydratase [EC42146]	6.00E-06
TTE0273	Galactose-1-phosphate uridylyltransferase[EC27710]	8.00E-06
Pyruvate metabolism		
Gmet_0984	acetyl-CoA carboxylase biotincarboxylase [EC63414 6412]	4.00E-06
Propanoate metabolism		
Gmet_0984	acetyl-CoA carboxylase biotincarboxylase [EC63414 6412]	4.00E-06
Rfer_3612	methylmalonyl-CoA mutase [EC54992]	8.00E-06
1.2 Energy Metabolism		
Nitrogen metabolism		
ABO_2100	glutamate dehydrogenase fragment [EC1414]	1.00E-05
1.3 Lipid Metabolism		
Fatty acid metabolism		
Gmet_0984	acetyl-CoA carboxylase biotin carboxylase [EC63414 6412]	4.00E-06
Glycerolipid metabolism		
DNO_0097	glycosyl transferase family protein [EC241-]	2.00E-06
Glycerophospholipid metabolism		
GOX2215	glycerol-3-phosphate dehydrogenase [EC11995]	4.00E-06
1.4 Nucleotide Metabolism		
Francci3_0003	DNA polymerase III beta subunit [EC2777]	3.00E-06
Cag_0489	IMP dehydrogenase [EC111205]	4.00E-06
CFF8240_1474	ribose-phosphate pyrophosphokinase [EC2761]	1.00E-05
SSPG21	nucleoside-diphosphate kinase	
PsCon[10781]	metal-dependent Rnase	
KOG0888	Nucleoside diphosphate kinase	4.00E-11
AN05652	hypothetical protein [EC2132 3523 6355]	8.00E-06
1.5 Amino Acid Metabolism		
Glutamate metabolism		
ABO_2100	glutamate dehydrogenase fragment [EC1414]	1.00E-05
AN05652	hypothetical protein [EC2132 3523 6355]	8.00E-06
Alanine and aspartate metabolism		
AN05652	hypothetical protein [EC2132 3523 6355]	8.00E-06
Glycine, serine and threonine metabolism		

Kwal_11274		3.00E-06
AO090003000721	homoserine dehydrogenase [EC1113]	8.00E-06
CBG02570		3.00E-06
Methionine metabolism		
AFUA_1G13490	spermidine synthase [EC25116]	2.00E-13
Valine, leucine and isoleucine degradation		
tll1462	branched-chain amino acid aminotransferase [EC26142]	1.00E-05
Rfer_3612	methylmalonyl-CoA mutase [EC54992]	8.00E-06
Valine, leucine and isoleucine biosynthesis		
tll1462	branched-chain amino acid aminotransferase [EC26142]	1.00E-05
Mhun_2923	leucyl-tRNA synthetase [EC6114]	6.00E-06
PM1662	isoleucyl-tRNA synthetase [EC6115]	3.00E-06
Lysine biosynthesis		
Kwal_11274		3.00E-06
CtCon[0055]	dihydrodipicolinate synthase, lysine synthesis	
Kwal_11274		3.00E-06
AO090003000721	homoserine dehydrogenase [EC1113]	8.00E-06
Lysine degradation		
Rru_A1213	2-oxoglutarate dehydrogenase E1component [EC1242]	2.00E-06
Tryptophan metabolism		
Rru_A1213	2-oxoglutarate dehydrogenase E1component [EC1242]	2.00E-06
AFUA_2G15660	aldehyde dehydrogenase family protein putative [EC121-]	1.00E-18
Phenylalanine, tyrosine and tryptophan biosynthesis		
Bamb_5913	shikimate dehydrogenase [EC11125]	7.00E-07
1.7 Glycan Biosynthesis and Metabolism		
Glycan structures - biosynthesis 2		
YALI0A20922g	phosphatidylinositol glycan, class M [EC:2.4.1.-]	3.00E-06
1.8 Biosynthesis of Polyketides and Nonribosomal Peptides		
Streptomycin and Polyketide sugar unit biosynthesis		
XF0258	dTDP-4-keto-L-rhamnose reductase [EC111133]	8.00E-06
Pcryo_0624	dTDP-glucose 46-dehydratase [EC42146]	6.00E-06
1.9 Metabolism of Cofactors and Vitamins		
Pantothenate and CoA biosynthesis		
tll1462	branched-chain amino acid aminotransferase [EC26142]	1.00E-05
Biotin metabolism		

mll6270	dethiobiotin synthetase [EC6333]	3.00E-06
Nitrobenzene degradation		
AFUA_2G15660	aldehyde dehydrogenase family protein putative [EC121-]	1.00E-18
Metabolism of xenobiotics by cytochrome P450		
Reut_A1486	glutathione S-transferaseN-terminalglutathione S-transferase	9.00E-07
2. INFORMATION STORAGE AND PROCESSING		
Translation, ribosomal structure and biogenesis		
KOG0003	Ubiquitin/60s ribosomal protein L40 fusion	3.00E-11
KOG3291	Ribosomal protein S7	8.00E-08
KOG0407	40S ribosomal protein S14	3.00E-11
KOG3418	60S ribosomal protein L27	3.00E-08
KOG3255	60S ribosomal protein L9	4.00E-06
KOG1767	40S ribosomal protein S25	7.00E-06
KOG0004	Ubiquitin/40S ribosomal protein S27a fusion	2.00E-17
KOG0402	60S ribosomal protein L37	8.00E-07
MagCon[0380]	60S ribosomal protein L7	5.00E-14
UmCon[1715]	40S ribosomal protein S20	4.00E-12
PsCon[0062]	40S ribosomal protein S27A	1.00E-08
mg[1442]	40S ribosomal protein S14	2.00E-23
MagCon[0195]	60S ribosomal protein L23	5.00E-15
CpCEST-05-B-11	40S ribosomal protein S26	9.00E-19
UmCon[1296]	60S ribosomal protein L32	4.00E-07
UmCon[0855]	60S ribosomal protein L23	7.00E-16
BfCon[0339]	60S ribosomal protein L3	9.00E-08
Bg27453639	40S ribosomal protein S3	7.00E-31
CpCon[0075]	60S ribosomal protein L27	1.00E-08
GzCon[0097]	40S ribosomal protein S8	6.00E-19
SSPG78	40S ribosomal S12	1.00E-26
MagCon[0481]	polyubiquitin / ribosomal protein	3.00E-26
Um34332559	60S ribosomal protein L22	5.00E-12
SSPG347	glycyl tRNA synthetase	9.00E-06
Mag3391542	cytoplasmic alanyl-tRNAsynthetase	2.00E-06
mg[0073]	elongation factor 1-alpha	2.00E-12
UmCon[0131]	translation initiation factor 3 (eIF3)	9.00E-14
RNA processing and modification		
TWOG0658	Mitochondrial mRNA maturase encoded bypartially processed COB mRNA	8.00E-14
TWOG0967	Mitochondrial mRNA maturase/Homing endonuclease	2.00E-07
KOG4768	Mitochondrial mRNA maturase	8.00E-07
Transcription		
KOG3265	Histone chaperone involved in gene silencing	3.00E-19
KOG2462	C2H2-type Zn-finger protein	6.00E-08

MagCon[0847a]	splicing factor, RNA-binding protein	1.00E-19
CpCon[1190]	11-kDa nonhistone chromosomal protein,involved in transcriptional activation of a number ofgenes	4.00E-09
Gz22508234	ATP-dependent chromatin remodeling protein	7.00E-08
Chromatin structure and dynamics		
KOG3467	Histone H4	1.00E-07
KOG3265	Histone chaperone involved in gene silencing	3.00E-08
3. CELLULAR PROCESSES AND SIGNALING		
Signal transduction mechanisms		
KOG0279	G protein beta subunit-like protein	3.00E-10
Posttranslational modification, protein turnover, chaperones		
KOG0729	26S proteasome regulatory complex, ATPase RPT1	5.00E-12
KOG0101	Molecular chaperones HSP70/HSC70, HSP70superfamily	5.00E-12
KOG4295	Serine proteinase inhibitor (KU family)	2.00E-10
KOG0541	Alkyl hydroperoxide reductase/peroxiredoxin	1.00E-13
KOG0001	Ubiquitin and ubiquitin-like proteins	2.00E-20
Cell rescue/Detoxification		
BfCon[0734]	Hydroxyacylglutathione hydrolase	4.00E-10
mgb0771f	peroxisomal membrane proteinPMP20, peroxiredoxin	6.00E-19
BgCon[0797]	bleomycin hydrolase	2.00E-10
Cellular organisation		
PsCon[1116]	elicitin protein	5.00E-06
mg[0024]	alpha-tubulin	8.00E-09
GzCon[0013]	histone H3	1.00E-13
PsCon[0019]	actin	4.00E-07
MagCon[3262]	histone H4	9.00E-12
PsCon[10845]	loricrin-like protein,extracellular matrix	4.00E-06
Ionic homeostasis		
W0AA020ZD05C1	nickel-binding protein	1.00E-12
phosphate_metabolism		
UmCon[0408]	Heterokaryon incompatibility, het-c	2.00E-15
Transport facilitation		
GzCon[5039]	MFS-multidrug-resistance transporter	1.00E-06
BfCon[1976]	amino acid permease	4.00E-09
VD0105A01	plasma membrane H+ ATPase	5.00E-06
BfCon[1976]	amino acid permease	9.00E-14

Appendix Table 5.Putative identification and classification of EST's from the mycelia exposed to host root exudate based on blast homology searches in KEGG, COG and COGEME databases

Host		
1.1 Carbohydrate Metabolism		
Glycolysis / Gluconeogenesis		
Neut_0333	glyceraldehyde-3-phosphate dehydrogenase type I[EC12112]	3.00E-06
MagCon[0208]	phosphoglycerate kinase	3.00E-12
BgCon[0009]	enolase (2-phosphoglyceratedehydratase)	1.00E-21
AFUA_4G08600	aldehyde dehydrogenase putative[EC1213]	1.00E-14
Citrate cycle (TCA cycle)		
lp_0055	fumarate reductase flavoprotein subunit precursor [EC13991]	1.00E-05
NCU038572	NADP-dependent isocitrate dehydrogenase	5.00E-10
p2A13	2-ketoglutarate NADP oxidoreductase alpha subunit [EC1273]	4.00E-06
Pentose phosphate pathway		
mg[0122]	transaldolase, component of non-oxidative part of pentose-phosphate pathway	6.00E-17
SSPG967F	D-xylose reductase	3.00E-14
GzCon[7227]	D-arabinitol dehydrogenase	5.00E-08
Fructose and mannose metabolism		
APL_0652	phosphomannomutase [EC5428]	8.00E-06
Csal_0931	PfkB [EC2714]	8.00E-06
Galactose metabolism		
Tery_4624	alpha-glucosidase [EC32120]	2.00E-11
Ascorbate and aldarate metabolism		
AFUA_4G08600	aldehyde dehydrogenase putative [EC1213]	1.00E-14
Starch and sucrose metabolism		
Csal_0931	PfkB [EC2714]	8.00E-06
HQ2752A	Lhr-like helicase [EC361-]	3.00E-06
W08D27	W08D27 [EC361-]	8.00E-06
Tery_4624	alpha-glucosidase [EC32120]	2.00E-11
Aminosugars metabolism		
F59B23	glutaminase [EC35125]	2.00E-06

Nucleotide sugars metabolism		
Mvan_1727	dTDP-4-dehydrorhamnose reductase [EC111133]	2.00E-06
Pyruvate metabolism		
AFUA_4G08600	aldehyde dehydrogenase putative [EC1213]	1.00E-14
Propanoate metabolism		
AFUA_4G08600	aldehyde dehydrogenase putative [EC1213]	1.00E-14
Butanoate metabolism		
lp_0055	fumarate reductase flavoprotein subunit precursor [EC13991]	1.00E-05
AFUA_4G08600	aldehyde dehydrogenase putative[EC1213]	1.00E-14
Inositol phosphate metabolism		
PICST_82710	Inositol-145-triphosphate 5-phosphatase [EC31356]	1.00E-05
1.2 Energy Metabolism		
Oxidative phosphorylation		
Pro1067	NADPH-quinone oxidoreductase NdhD subunit [EC1653]	3.00E-06
CAGL0L06160g	P04037 *Saccharomyces cerevisiae* YGL187c COX4	8.00E-14
ST0676	hypothetical cytochrome c oxidase polypeptide I[EC1931]	3.00E-06
TK1614	NADH ubiquinone oxidoreductase subunit E [EC1653]	4.00E-06
lp_0055	fumarate reductase flavoprotein subunit precursor [EC13991]	1.00E-05
3283492	NADH dehydrogenase subunit 1 [EC1653]	1.00E-05
Methane metabolism		
AN59182	hypothetical protein [EC11116]	1.00E-11
1.3 Lipid Metabolism		
Fatty acid metabolism		
AFUA_4G08600	aldehyde dehydrogenase putative[EC1213]	1.00E-14
Glycerolipid metabolism		
SPO0104	glycerol kinase [EC27130]	8.00E-06
AFUA_4G08600	aldehyde dehydrogenase putative[EC1213]	1.00E-14
Glycerophospholipid metabolism		
PP_3664	CDP-diacylglycerol--serineO-phosphatidyl transferase [EC2788]	8.00E-06
Sphingolipid metabolism		
AO090003001164	glutamate decarboxylase sphingosine phosphate lyase [EC41227]	
1.4 Nucleotide Metabolism		
PMT9312_0001	DNA polymerase III beta subunit	2.00E-06

	[EC2777]	
SG0701	Phosphoribosyl amino imidazole carboxylase ATPase subunit	6.00E-06
Gz31373472	purine permease	6.00E-07
Gz31370931	Extracellular guanyl-specific ribonuclease	2.00E-12
MagCon[0864a]	nucleoside diphosphate kinase	2.00E-21
PsCon[10781]	metal-dependent Rnase	8.00E-06
AFUA_7G02620	DNA-directed RNA polymerases N8 kDa subunit superfamily	5.00E-08
RSP_3591	cytidylate kinase [EC27414]	4.00E-07
1.5 Amino Acid Metabolism		
Glycine, serine and threonine metabolism		
PP_3664	CDP-diacylglycerol--serineO-phosphatidyl transferase [EC2788]	8.00E-06
AAur_1482	glycyl-tRNA synthetase [EC61114]	2.00E-06
SPA2921	glycine dehydrogenase decarboxylating [EC1442]	1.00E-05
Methionine metabolism		
Sfum_2797	DNA-cytosine methyl transferase [EC21137]	6.00E-06
Valine, leucine and isoleucine degradation		
AN36392	hypothetical protein [EC231168]	2.00E-10
Pfl_0696	3-hydroxyisobutyrate dehydrogenase [EC11131]	4.00E-12
AFUA_4G08600	aldehyde dehydrogenase putative[EC1213]	1.00E-14
Valine, leucine and isoleucine biosynthesis		
PAB1782	leucyl-tRNA synthetase [EC6114]	1.00E-05
Lysine biosynthesis		
CtCon[0055]	dihydrodipicolinate synthase, lysine synthesis	1.00E-06
Lysine degradation		
NCU071172		7.00E-10
AFUA_4G08600	aldehyde dehydrogenase putative [EC1213]	1.00E-14
Histidine metabolism		
AT4G14910	IGPD Imidazoleglycerol-phosphate dehydratase [EC42119]	4.00E-06
AFUA_4G08600	aldehyde dehydrogenase putative [EC1213]	1.00E-14
Tryptophan metabolism		
Veis_1623	tryptophanyl-tRNA synthetase [EC6112]	3.00E-06
AN59182	hypothetical protein [EC11116]	1.00E-11
AFUA_4G08600	aldehyde dehydrogenase putative [EC1213]	1.00E-14
1.6 Metabolism of Other		

Amino Acids		
beta-Alanine metabolism		
AFUA_4G08600	aldehyde dehydrogenase putative [EC1213]	1.00E-14
D-Glutamine and D-glutamate metabolism		
BTH_I1116	UDP-N-acetylmuramoylalanine--D-glutamate ligase [EC6329]	1.00E-05
1.7 Glycan Biosynthesis and Metabolism		
Glycan structures - biosynthesis 2		
DDBDRAFT_0218103	hypothetical protein [EC27--]	6.00E-06
BT_2747	3-deoxy-D-manno-octulosonic-acid transferase [EC2---]	6.00E-06
1.8 Biosynthesis of Polyketides and Nonribosomal Peptides		
Polyketide sugar unit biosynthesis		
Mvan_1727	dTDP-4-dehydrorhamnose reductase [EC111133]	2.00E-06
1.9 Metabolism of Cofactors and Vitamins		
Biotin metabolism		
Hac_1730	biotin synthetase [EC2816]	6.00E-06
Folate biosynthesis		
HQ2752A	Lhr-like helicase [EC361-]	3.00E-06
W08D27	Hypothetical protein [EC361-]	8.00E-06
gamma-Hexachlorocyclohexane degradation		
NCU096002	Hypothetical protein	8.00E-06
Benzoate degradation via CoA ligation		
lp_0055	fumarate reductase flavoprotein subunit precursor [EC13991]	1.00E-05
Atrazine degradation		
Kwal_11951	Hypothetical protein	2.00E-06
2. INFORMATION STORAGE AND PROCESSING		
Translation, ribosomal structure and biogenesis		
YGL103w	60s ribosomal protein L15/L27	6.00E-18
YPL079w	60S ribosomal protein L21	2.00E-16
YPL013c	Mitochondrial/chloroplast ribosomal protein S16	1.00E-06
YJL136c	40S ribosomal protein S21	1.00E-13
YLR167w	Ubiquitin/40S ribosomal protein S27a fusion	1.00E-20
YPL143w	60S ribosomal protein L35A/L37	1.00E-12

YNL162w	60S ribosomal protein L44	5.00E-14
YER117w	60S ribosomal protein L14/L17/L23	1.00E-09
Hs14277700	40S ribosomal protein S12	3.00E-09
SPCC663.04	60s ribosomal protein L39	3.00E-10
Gz31372516	translation release factor erf3	5.00E-06
Mag23356336	eukaryotic peptide chain release factor subunit 1	1.00E-15
SPBC25H2.07	Translation initiation factor 1A (eIF-1A)	6.00E-11
SPCC285.15c	40S ribosomal protein S28	5.00E-15
YNL162w	60S ribosomal protein L44	3.00E-10
At5g59850	40S ribosomal protein S15/S22	1.00E-11
YHR010w	60S ribosomal protein L27	1.00E-12
SPBC16C6.11	60S ribosomal protein L32	1.00E-15
SPBC29A3.12	Ribosomal protein S4	6.00E-08
YOR293w	40s ribosomal protein s10	2.00E-07
SPCC1223.05c	60S ribosomal protein L37	5.00E-07
YDL061c	40S ribosomal protein S29	3.00E-07
7290855	40S ribosomal protein S14	9.00E-07
SPCC1183.08c	60S ribosomal protein L10A	3.00E-07
YFR032c-a	60S ribosomal protein L29	6.00E-10
YPR132w	40S ribosomal protein S23	4.00E-30
At3g62250	Ubiquitin/40S ribosomal protein S27a fusion	2.00E-27
Trnscrptn_rRNA		
W0AA058ZF07C1	rRNA methyltransferase	1.00E-06
UmCon[1919]	required for pre-rRNA pseudouridylationand processing	3.00E-18
Um37415431	involved in 35S primary transcriptprocessing	8.00E-06
RNA processing and modification		
Hs5729802	Component of the U4/U6.U5 snRNP/ mitosisprotein DIM1	4.00E-13
YOR046c	ATP-dependent RNA helicase	7.00E-06
SPBC20F10.09	U6 snRNA-associated Sm-like protein	3.00E-17
SPBC3E7.14	Small nuclear ribonucleoprotein (snRNP) SMF	6.00E-11
YMi017_2	Mitochondrial mRNA maturase/ Homing endonuclease	2.00E-06
YMi015_2	Mitochondrial mRNA maturase encoded by partially processed COB mRNA	2.00E-12
YHR072w-a	H/ACA snoRNP complex, subunit NOP10	1.00E-15
Transcription		
mgb0318f	small nuclear ribonucleoprotein polypeptide	4.00E-09
SSPG993	telomere-associated protein that binds single-stranded G-strand telomere sequence	5.00E-10
YDR045c	RNA polymerase III subunit C11	1.00E-13

YOR210w	DNA-directed RNA polymerase, subunit RPB10	7.00E-07
Replication, recombination and repair		
YDR013w	Predicted alpha-helical protein, potentially involved in replication/repair	4.00E-07
SPBC16D10.09	DNA polymerase delta processivity factor (proliferating cell nuclear antigen)	1.00E-08
YMi017_2	Mitochondrial mRNA maturase/ Homingendonuclease	2.00E-06
Chromatin structure and dynamics		
SPBPI060	Histones H3 and H4	2.00E-11
YNL030w	Histone H4	1.00E-07
2. CELLULAR PROCESSES AND SIGNALING		
Cell cycle control, cell division, chromosome partitioning		
Hs5729802	Component of the U4/U6.U5 snRNP/ mitosis protein DIM1	4.00E-13
Cell division		
CpCEST-24-E-02	subunit of the GINS complex required for chromosomal DNA replication	3.00E-16
BfCon[1590]	DNA polymerase delta accessory protein (PCNA)	5.00E-15
Gz40384617	cell cycle inhibitor	2.00E-08
Cell division/ mating sex specificity		
Um34330540	beta transducin-like vegetatible incompatibility protein	8.00E-08
Cell death		
MagCon[10719a]	programmed cell death protein	3.00E-31
Cell rescue/ Polysaccharide degradation		
GzCon[3452]	2-deoxy-D-gluconate3-dehydrogenase, pectin degradation	2.00E-09
Cell rescue/ detoxification		
SSPG103	catalase	4.00E-14
mgb0771f	peroxisomal membrane proteinPMP20, peroxiredoxin	4.00E-13
Cellular biogenesis		
W0AA017ZE03C1	mannosidase, glycosylphosphatidyl inositol (GPI)-anchored membrane protein required for cell wall biogenesis and filamentous growth	2.00E-08
Mag45419875	chitin deacetylase	1.00E-12
BfCon[1046]	actin related protein 2/3 complex, subunit 4	9.00E-22
Signal transduction mechanisms		

SPCC830.06	Ca2+/calmodulin-dependent protein phosphatase (calcineurin subunit B), EF-Hand superfamily protein	4.00E-16
7289349	Ca2+-binding protein, EF-Hand protein superfamily	1.00E-11
7290576	Ca2+/calmodulin-dependent protein phosphatase (calcineurin subunit B), EF-Hand superfamily protein	1.00E-06
Hs5174447	G protein beta subunit-like protein	1.00E-23
CE13902	Calmodulin and related proteins (EF-Handsuperfamily)	3.00E-25
Cytoskeleton		
Hs4501887	Actin and related proteins	6.00E-08
Hs5031595	Actin-related protein Arp2/3 complex, subunitARPC4	8.00E-09
Extracellular structures		
Hs4502955	Collagens (type IV and type XIII), and related proteins	9.00E-06
Intracellular trafficking, secretion, and vesicular transport		
CpCEST-36-C-02	protein transport protein sec23	6.00E-09
SPAC15E1.06	Membrane coat complex Retromer, subunitVPS29/PEP11	8.00E-08
YEL020w-a	Mitochondrial import inner membrane translocase, subunit TIM9	8.00E-06
SPBC4F6.18c	GTP-binding ADP-ribosylation factor Arf1	4.00E-22
W0AA005ZB04C1	kinesin light chain	4.00E-09
BfCon[0118]	dynein light chain	6.00E-09
SPBC4F6.18c	GTP-binding ADP-ribosylation factor Arf1	1.00E-15
7292782	Peptide exporter, ABC superfamily	1.00E-13
Posttranslational modification, protein turnover, chaperones		
SPBC28F2.03	Cyclophilin type peptidyl-prolyl cis-transisomerase	8.00E-16
SPAC1B3.03c	HSP90 co-chaperone CPR7/Cyclophilin	7.00E-22
SPAC3A11.10c	Renal dipeptidase	6.00E-11
YNL135c	FKBP-type peptidyl-prolyl cis-trans isomerase	5.00E-10
SPBC119.02	Ubiquitin-protein ligase	2.00E-20
7291296	FKBP-type peptidyl-prolyl cis-trans isomerase	7.00E-07
GzCon[4542]	vacuolar protein sorting29	4.00E-12
mgb0297f	protein-vacuolar targeting	3.00E-15
Um34331964	protein transport proteinsec61-gamma subunit	1.00E-12
Um34331109	tubulin binding protein, tubulin folding	1.00E-13
CE17506	Cyclophilin type peptidyl-prolyl cis-trans isomerase	2.00E-06

SPCC330.06c	Alkyl hydroperoxide reductase/ peroxiredoxin	2.00E-08
Disease virulence		
Gz22505271	structural toxin protein homologue	1.00E-10
MagCon[2881a]	NADPH oxidase	4.00E-11
Ionic homeostasis		
GzCon[5502]	L-ornithine N5-oxygenase, siderophore biosynthesis	2.00E-14
Nitrogen /sulphur metabolism		
Fs14664609	urease	2.00E-06
W0AA064ZC12C1	cyanate lyase	8.00E-15
Mag45375503	Ni-binding urease accessory protein G	5.00E-07
Transport facilitation		
BfCon[1629]	hexose transporter	1.00E-10
MagCon[1406]	copper transporter	6.00E-18
mga1511f	metal resistance protein, ABC transporter	6.00E-14
Um37412581	secretory pathway Ca2+-ATPase	2.00E-10
Ct21906599	ATP-binding cassette (ABC) transporter, multidrug resistance protein	2.00E-13
VD0210A09	vacuolar proton pump B subunit	3.00E-29
GzCon[0948]	neutral amino acid transporter	6.00E-12
SPCC757.07c	Catalase	1.00E-10
Transposon insertion sequence proteins		
Gz22509405	intron derived maturase	3.00E-07

Appendix Table 6. Putative identification and classification of EST's from the mycelia exposed to non-host root exudate based on blast homology searches in KEGG, COG and COGEME databases

Non host		
1.1 Carbohydrate Metabolism		
Glycolysis / Gluconeogenesis		
BfCon[0344]	enolase (2-phosphoglyceratedehydratase)	1.00E-29
MagCon[0623]	pyruvate kinase	4.00E-08
VD0202C03	Phosphoenolpyruvate carboxykinase, rate limiting gluconeogenic enzyme	1.00E-11
BfCon[0818]	phosphoglucomutase	4.00E-09
126273795	pyruvate decarboxylase [EC4111]	5.00E-08
Citrate cycle (TCA cycle)		
GzCon[2156]	malate dehydrogenase	3.00E-06
SSPG305	aldo/keto reductase	1.00E-12
114328732	2-oxoglutarate dehydrogenase E1component [EC1242]	6.00E-06
Pentose phosphate pathway		
SSPG592	transketolase	8.00E-09
Pentose and glucuronate interconversions		
MagCon[0916a]	D-arabinitol dehydrogenase	2.00E-08
mg[1220]	sorbitol utilisation protein	9.00E-11
Fructose and mannose metabolism		
108762403	glycosyl transferase group 2 family protein [EC241-]	2.00E-06
Galactose metabolism		
22299749	UDP-glucose 4-epimerase [EC5132]	1.00E-05
83772646	beta-galactosidase [EC32123]	5.00E-11
113478008	alpha-glucosidase [EC32120]	5.00E-11
Starch and sucrose metabolism		
67523717	glycogen branching enzyme	2.00E-11
15217670	APL2 large subunit of AGP 2 [EC27727]	8.00E-06
70990230	glycogen debranching enzyme Gdb1 putative [EC24125 32133]	7.00E-07
83764527	13-beta-glucan synthase/ callose synthase catalytic subunit	2.00E-10
113478008	alpha-glucosidase [EC32120]	5.00E-11
Aminosugars metabolism		
YLR307wCDA1	Q06702 *Saccharomyces cerevisiae* YLR307wCDA1	4.00E-06
Nucleotide sugars metabolism		
22299749	UDP-glucose 4-epimerase [EC5132]	1.00E-05

Pyruvate metabolism		
75909594	PEP-utilizing enzyme [EC2792]	1.00E-05
Butanoate metabolism		
70994460	succinate-semialdehyde dehydrogenase Uga2 putative [EC12124]	3.00E-07
73541826	alphabeta hydrolase fold poly-beta-hydroxy butyrate polymerase	4.00E-06
1.2 Energy Metabolism		
Oxidative phosphorylation		
67540170	hypothetical protein [EC1931]	4.00E-08
67902124	hypothetical protein [EC1653 16993]	6.00E-19
58618665	NADH dehydrogenase subunit 1 [EC1653]	2.00E-06
134098270	cytochrome b membrane protein [EC11022]	8.00E-06
1.3 Lipid Metabolism		
Fatty acid metabolism		
GzCon[6802]	phytanoyl-CoA dioxygenase, catalyses the first step of phytanic acid alpha-oxidation	1.00E-08
SSPG728	acyl-CoA dehydrogenase	1.00E-19
PsCon[1259]	fatty acid beta-oxidation-related protein	7.00E-06
	peroxisomal-coenzyme A synthase	1.00E-06
Synthesis and degradation of ketone bodies		
Biosynthesis of steroids		
77164555	Squalene phytoene synthase [EC25132]	1.00E-05
Glycerolipid metabolism		
108762403	glycosyl transferase group 2 family protein [EC241-]	2.00E-06
83772646	beta-galactosidase [EC32123]	5.00E-11
1.4 Nucleotide Metabolism		
PsCon[10781]	metal-dependent Rnase	7.00E-06
78046782	DNA polymerase III alpha chain protein [EC2777]	3.00E-06
59800692	putative thymidylate kinase [EC2749]	1.00E-05
86609722	dihydroorotate oxidase [EC1331]	4.00E-06
1.5 Amino Acid Metabolism		
Glutamate metabolism		
70994460	succinate-semialdehyde dehydrogenase Uga2 putative [EC12124]	3.00E-07
94987309	glutamate racemase [EC5113]	3.00E-06
Alanine and aspartate metabolism		
42524510	alanine racemase [EC5111]	8.00E-06
15673810	asparaginyl-tRNA synthetase [EC61122]	6.00E-06
Methionine metabolism		
76800795	O-acetylhomoserine aminocarboxypropyl transferase 2 methionine	9.00E-07
Cysteine		

metabolism		
76800795	O-acetylhomoserine aminocarboxypropyl transferase 2 methionine	9.00E-07
Lysine Metabolism		
CtCon[0055]	dihydrodipicolinate synthase, lysine synthesis	6.00E-08
114328732	2-oxoglutarate dehydrogenase E1component [EC1242]	6.00E-06
Arginine and proline metabolism		
123967984	delta 1-pyrroline-5-carboxylate reductase [EC1512]	6.00E-06
Tryptophan metabolism		
114328732	2-oxoglutarate dehydrogenase E1component [EC1242]	6.00E-06
Phenylalanine, tyrosine and tryptophan biosynthesis		
88602325	3-dehydroquinate synthase [EC4234]	6.00E-06
1.6 Metabolism of Other Amino Acids		
Taurine and hypotaurine metabolism		
74317428	gamma-glutamyl transpeptidase [EC2322]	6.00E-06
70992223	alpha-ketoglutarate-dependent taurine dioxygenase [EC1141117]	3.00E-10
D-Glutamine and D-glutamate metabolism		
94987309	glutamate racemase [EC5113]	3.00E-06
50955151	UDP-N-acetylmuramoyl alanine-D-glutamate ligase [EC6329]	1.00E-05
D-Alanine metabolism		
42524510	alanine racemase [EC5111]	8.00E-06
1.7 Glycan Biosynthesis and Metabolism		
N-Glycan biosynthesis		
22122355	beta galactoside alpha 26 sialyl transferase 1[EC24991]	6.00E-06
N-Glycan degradation		
83772646	beta-galactosidase [EC32123]	5.00E-11
83766236	alpha-mannosidase [EC32124]	7.00E-12
Glycan structures -		

biosynthesis 1		
22122355	beta galactoside alpha 26 sialyl transferase 1[EC24991]	6.00E-06
Glycan structures - biosynthesis 2		
108762403	glycosyl transferase group 2 family protein [EC241-]	2.00E-06
1.9 Metabolism of Cofactors and Vitamins		
Folate biosynthesis		
16801953	similar to para-aminobenzoate synthase component I [EC6358	8.00E-06
Porphyrin and chlorophyll metabolism		
54025707	putative protoporphyrinogen oxidase [EC1334]	9.00E-07
Terpenoid biosynthesis		
77164555	Squalene phytoene synthase [EC25132]	1.00E-05
gamma-Hexachlorocyclohexane degradation		
17560956	Hypothetical protein[EC31341]	2.00E-06
2. INFORMATION STORAGE AND PROCESSING		
Translation, ribosomal structure and biogenesis		
KOG3506	40S ribosomal protein S29	4.00E-11
KOG3504	60S ribosomal protein L29	1.00E-09
KOG0004	Ubiquitin/40S ribosomal protein S27a fusion	1.00E-17
KOG1570	60S ribosomal protein L10A	2.00E-06
KOG0378	40S ribosomal protein S4	7.00E-19
KOG3257	Mitochondrial/chloroplast ribosomal protein L11	6.00E-10
KOG3475	60S ribosomal protein L37	8.00E-21
KOG0688	Peptide chain release factor 1 (eRF1)	5.00E-18
KOG1628	40S ribosomal protein S3A	2.00E-09
KOG0901	60S ribosomal protein L14/L17/L23	8.00E-10
KOG0898	40S ribosomal protein S15	2.00E-18
KOG0052	Translation elongation factor EF-1 alpha/Tu	2.00E-08
MagCon[3339]	eukaryotic translation initiation factor 2 alpha subunit	4.00E-09
Um34332274	translation elongation factor EF-1beta	2.00E-15
Mag23356336	eukaryotic peptide chain release factor subunit 1	2.00E-15
KOG0893	60S ribosomal protein L31	2.00E-11
RNA		

processing and modification		
TWOG0967	Mitochondrial mRNA maturase/ Homing endonuclease	2.00E-23
KOG4768	Mitochondrial mRNA maturase	3.00E-06
TWOG0658	Mitochondrial mRNA maturase encoded by partially processed COB mRNA	2.00E-07
KOG3448	Predicted snRNP core protein	6.00E-14
KOG3482	Small nuclear ribonucleoprotein (snRNP) SMF	1.00E-11
KOG1781	Small Nuclear ribonucleoprotein splicing factor	1.00E-06
mgc18d12f	zinc finger protein	8.00E-16
KOG4768	Mitochondrial mRNA maturase	7.00E-07
KOG2151	Predicted transcriptional regulator	2.00E-06
KOG1668	Elongation factor 1 beta/delta chain	1.00E-08
KOG3473	RNA polymerase II transcription elongationfactor Elongin/SIII, subunit elongin C	1.00E-07
Replication, recombination and repair		
KOG0890	Protein kinase of the PI-3 kinase family involved in mitotic growth, DNA repair and meiotic recombination	6.00E-08
Chromatin structure and dynamics		
KOG3467	Histone H4	1.00E-09
3. CELLULAR PROCESSES AND SIGNALING		
Cell cycle control, cell division, chromosome partitioning		
KOG2151	Predicted transcriptional regulator	2.00E-06
KOG0890	Protein kinase of the PI-3 kinase family involved in mitotic growth, DNA repair and meiotic recombination	6.00E-08
Gz40384617	cell cycle inhibitor	5.00E-11
Mag45395388	Cwf15 / Cwc15 cell cycle control family protein	
Cell death		
Mag30404439	DNA-binding apoptosis protein	4.00E-07
Cell rescue/ Polysaccharide degradation		
SSPG615	1,4-beta-D-glucan cellobiohydrolase	1.00E-15
Cellular		

biogenesis		
Fs14666757	cofilin / tropomyosin-type actin binding protein	8.00E-12
UmCon[1917]	profilin 1B	1.00E-06
Fs14663871	required for F-actin regulation	9.00E-08
Mag45419875	chitin deacetylase	1.00E-17
Cell rescue/ DNA repair		
Bg13900918	rad16 nucleotide excision repair protein	2.00E-08
Cell rescue stress response		
MagCon[9987a]	stress-responsive protein	9.00E-08
Signal transduction mechanisms		
KOG0192	Tyrosine kinase specific for activated (GTP-bound) p21cdc42Hs	3.00E-07
KOG0027	Calmodulin and related proteins (EF-Handsuperfamily)	2.00E-20
KOG0890	Protein kinase of the PI-3 kinase family involved in mitotic growth, DNA repair and meiotic recombination	6.00E-08
KOG0027	Calmodulin and related proteins (EF-Handsuperfamily)	2.00E-08
Cytoskeleton		
PsCon[0019]	actin	9.00E-10
MagCon[8161a]	F-actin capping protein alpha-2 subunit	1.00E-05
PsCon[10845]	loricrin-like protein, extracellular matrix	5.00E-11
KOG1755	Profilin	4.00E-09
Intracellular trafficking, secretion, and vesicular transport		
W0AA070ZC03C1	syntaxin	1.00E-11
MagCon[9481a]	GTPase EF-Hand Protein of mitochondria, involved in vesicle-mediated transport	8.00E-11
GzCon[7347]	trafficking protein particle complex 3	1.00E-17
BfCon[0118]	dynein light chain	4.00E-07
BfCon[0384]	involved in membrane trafficking	2.00E-07
KOG4224	Armadillo repeat protein VAC8 required for vacuole fusion, inheritance and cytosol-to-vacuole protein targeting	4.00E-12
KOG0446	Vacuolar sorting protein VPS1, dynamin, and related proteins	8.00E-09
KOG4580	Component of vacuolar transporter chaperone (Vtc) involved in vacuole fusion	7.00E-14
KOG3330	Transport protein particle (TRAPP) complex subunit	7.00E-09
Posttranslational modification, protein		

turnover, chaperones		
KOG3478	Prefoldin subunit 6, KE2 family	7.00E-09
LSE0115	A Subtilisin-like proteinase	8.00E-10
KOG0182	20S proteasome, regulatory subunit alpha type PSMA6/SCL1	1.00E-15
KOG1661	Protein-L-isoaspartate (D-aspartate) O-methyltransferase	2.00E-08
KOG0181	20S proteasome, regulatory subunit alpha typePSMA2/PRE8	4.00E-06
KOG4146	Ubiquitin-like protein	1.00E-11
KOG0179	20S proteasome, regulatory subunit beta typePSMB1/PRE7	2.00E-10
KOG0357	Chaperonin complex component, TCP-1 epsilon subunit (CCT5)	9.00E-17
KOG2512	Beta-tubulin folding cofactor C	2.00E-06
KOG4580	Component of vacuolar transporter chaperone (Vtc) involved in vacuole fusion	7.00E-14
KOG0714	Molecular chaperone (DnaJ superfamily)	9.00E-07
KOG2195	Transferrin receptor and related proteins containing the protease-associated (PA) domain	7.00E-14
Protein modification		
SSPG214	multiple RNA binding domain, essential for synthesis of the small ribosomal subunit	7.00E-09
Protein modification sorting and translocation		
CpCEST-25-B-10	vacuolar protein required for vacuole fusion, involved in vacuolar inheritance and protein targeting from the cytoplasm to vacuole	2.00E-20
Prtn_mofifn		
mgc02e05f	ribophorin, part of oligosaccharyl transferase complex in ER	2.00E-09
MagCon[0500]	Mannosyl transferase complex subunit	2.00E-10
Disease virulence		
CpCon[0457]	pathogenesis related (SnodProt1)	4.00E-09
Ionic homeostasis		
W0AA020ZD05C1	nickel-binding protein	6.00E-08
Nitrogen/ sulphur metabolism		
BfCon[1566]	negative regulator of sulphur metabolism (sconCp homologue, *E. nidulans*)	4.00E-10
Transport facilitation		
MagCon[10315a]	transporter	2.00E-15
SSPG39F	hexose transporter	1.00E-06

mgb0154f	vacuolar-ATPase	1.00E-14
mg[1488]	ER / Golgi polyphosphoinositide phosphatase, required for transport of ATP into ER	1.00E-07
BfCon[1976]	amino acid permease	2.00E-07
MagCon[1406]	copper transporter	6.00E-16
Transposon insertion sequence proteins		
BfCon[1613]	intron derived maturase	2.00E-13
Gz22509405	intron derived maturase	1.00E-09