INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI®

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

DEVELOPMENT, ANALYSIS AND USE OF AN EXPRESSED SEQUENCE TAG DATABASE FROM THE MULTICELLULAR ASCOMYCETE, *ASPERGILLUS NIDULANS*

A Dissertation

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

By

DORIS MARIE KUPFER

Norman, Oklahoma

1999

DEVELOPMENT, ANALYSIS AND USE OF AN EXPRESSED SEQUENCE TAG DATABASE FROM THE MULTICELLULAR ASCOMYCETE, *ASPERGILLUS NIDULANS*

A Dissertation APPROVED FOR THE

DEPARTMENT OF BOTANY AND MICROBIOLOGY

BY

Acknowledgements

I wish to thank my advisory committee, Drs. David McCarthy, Leonard Beevers, William Ortiz-Leduc, Ralph Tanner, and Bruce Roe for their continued support of my efforts to complete my dissertation work. A special thanks goes to Dr. McCarthy, my major professor, for his encouragement to continue my studies for a doctorate. The work for my dissertation was done entirely in the laboratory of Dr. Roe. I wish to express my gratitude to him for accepting me into his lab and providing me with the training and guidance which allowed me to complete my dissertation research.

All the members of Dr. Roe's lab at the Advanced Center for Genome Research, including the wonderful gel room crew, fellow graduate students and researchers deserve many thanks for their help. My thanks also to Fares Najar for his expertise and help. Hongshing Lai deserves special thanks for her patience with my very limited computer skills. I much appreciate her help and our many conversations. My thanks also to Jim White who was willing to make time to help with thorny data analysis problems. And to one of the lab members who has moved on, Sandra Clifton, goes my appreciation for her special friendship.

My husband, John, daughter, Sara and son, Scott, are the three I could not have done without. They deserve immense thanks for putting up with strange hours and poor cooking. Their love, care, support and John's gourmet skills made all the difference. I wish also to thank my brother, Paul, for always being interested and for being there when really needed. I also want to thank Marjorie and Cliff Downard for their love and support.

Finally, I want to recognize my mother who always knew I was the best at whatever I did and said so. I wish to dedicate this dissertation to her memory.

Eunice Bird Kupfer 1918-1999

Table of Contents

List of Tables

List of Figures

Abstract

A. nidulans is a well-studied model multicellular ascomycete closely related to both human and plant fungal pathogens. To investigate the expressed genes in this model fungus, a cDNA library containing clones from vegetatively and asexually expressed genes was sequenced from both the 3' and 5' ends. These cDNA end sequences called expressed sequence tags or ESTs reveal genes which were expressed at the time of mRNA harvest. The 3' ESTs were aligned using the multiple sequence assembly program Phrap to follow the level of sequence redundancy and to measure the percent of new genes detected during the sampling of the cDNA library. A Unigene database representing approximately 3200 genes was generated after assembly of the 3' and 5' ESTs using Phrap to cluster the EST sequences which were generated from the same gene. The Unigene sequences were examined for homologs in GenBank using the Blast alignment program and organized into groups based on a Biological Function Classification schema. Almost 56% of the Unigene database members had no homologs in GenBank and thus represented newly discovered genes. The Unigene database also was useful in annotating three *A. nidulans* chromosome VIII cosmids and a 60 Kbp region from chromosome IV containing the sterigmatocystin gene cluster. Through these studies, an average gene density of 1 gene/ 2.6 Kbp was obtained and a total number of 12,000 genes was predicted for the *A. nidulans* genome.

Since these ESTs are a rich resource of expressed gene sequences useful to the large community of fungal researchers, the resulting EST database was made publicly available by submission to GenBank and placement on the ACGT website. In addition, the cDNA clones were submitted to the Fungal Genetics Stock Center, University of Kansas Medical Center, Kansas City. These ESTs will provide a sequence and clone source for hybridization probes for *A. nidulans* expressed genes as well as for related fungi with industrial, agricultural and health application.

Chapter 1

Introduction

1.1 Overview

During the course of this study an *Aspergillus nidulans* expressed sequence tag database (EST) of 4.268 Mbp and a total of 12,485 EST sequences was created. Improved protocols for optimizing template production and data handling and analysis also were developed during this work. A Unigene database, which uses multiple sequence alignments to take advantage of clone redundancy, extended the quality and length of the single-pass EST sequences to produce a series of consensus sequences. This resulting Unigene database was useful for finding gene families, determining expression levels of detected genes, creating a biological function overview of the organism, and for annotating sequenced genomic DNA.

1.2 EST Background

Messenger RNA is the transcribed intermediate in the protein biosynthetic pathway of DNA to RNA to proteins. The discovery of reverse transcriptase and the ability to isolate mRNA has allowed the generation of a stable DNA copy of the less stable mRNA called a copy-DNA or cDNA. These cDNAs are a copy of processed expressed gene sequences since they lack introns. The mature mRNA from which they are produced provides the open reading frame for the translation product that includes both the 5' start and 3' stop codons. There is additional information such as the position of the ribosome binding site in the 5' untranslated region and the polyadenylation signal in the 3' untranslated region.

The value of cDNA sequences for annotation of genomic regions was realized early in the human genome project (Adams, 1991). The presence of variable-sized introns as well as repeat sequences has made annotation of human genomic sequences

difficult. To aid in annotating the human genome, a large number of tissue specific cDNA libraries was prepared by The Institute for Genome Research (Adams, 1991) and in conjunction with Merck and Company, at Washington University, St. Louis (Hillier, 1996). Since partial cDNA sequences could identify and tag coding regions of genomic sequence, high quality, full-length cDNA sequences were found to be unnecessary. This realization lead to an end-sequencing only approach and the generation of single-pass sequences, termed Expressed Sequence Tags (ESTs), from the ends of the cDNA templates (Adams, 1995). These sequences provide 1x or less coverage of a cDNA and are useful even though considered error prone (Marra, 1999).

Presented below is an overview of the large scale sequencing methods which have been used during this dissertation research to generate an EST database for the multicellular ascomycete fungus, *Aspergillus nidulans*. The computer-based methods developed to efficiently handle the large number of sequences generated will be outlined, as will the approaches taken for optimal use of the ESTs as a gene rich information source for annotation of genomic regions. These studies have provided a window into the classes of genes and the level of their expression in *A.nidulans*.

1.3 EST Database

ESTs represent sequences of expressed genes that are generated from both the 3' and 5' ends of randomly selected cDNAs (Adams, 1991). An interest in genome studies led to the realization that these single-pass sequences could be rapidly generated, and then used for gene discovery, constructing physical maps, studying expression patterns, and as an aid in sequence interpretation in large-scale genome projects (Hillier, 1996). There are growing numbers of human and other eucaryotic ESTs in a separate database at GenBank, dbEST (Benson, 1996).

Based on the intrinsic and extrinsic approaches developed by Borodovsky et al. (Borodovsky, 1994), which incorporate EST information into the analysis of completed

genomic sequences as a means of improving gene identification, potential exons can be predicted with great certainty using tools such as GeneMark (Borodovsky, 1994) since EST matches then can be used to verify the predictions. Thus, an EST database for a target organism is extremely useful as it complements the gene prediction algorithms based on ORFs with conserved characteristics such as codon usage and intron-exon borders. Those predicted exons which have an EST match are verified as such. Even though an EST is only a partial sequence of a cDNA, it is extremely useful for indicating the correct position of sequence landmarks such as transcription start, translation start and stop, intron-exon borders, and polyadenylation signals of the predicted genes. This information helps in predicting correct reading frames and eventually aids the construction of accurate physical maps of the genome. 3' ESTs have a uniform start site since they are anchored at the polyA tail by employing a polyT primer in first strand synthesis (Short, 1988 and see cDNA construction, below). The EST sequence then predictably will contain the 3' untranslated region (UTR). Even though this sequence may or may not be long enough to extend into the coding region, 3' ESTs represent gene specific sequences that have been used to provide genomic markers (Adams, 1991). The 5' ESTs are more variable in their start point because the length of first strand synthesis may not extend to the mRNA 5' end. Although the 5' EST does not necessarily contain the 5' UTR or the AUG start, it will contain coding sequence. The 5' ESTs therefore are considered gene family specific sequences rather than individual gene specific sequences (Hillier, 1996; Khan, 1992). The abundance of EST classes, which is proportional to their relative transcription levels can be determined from an EST database after assembly with multiple sequence alignment tools such as Phrap (Green, copyright 1994-1996). The EST database and the corresponding cDNA clones then serve as a source of expressed gene sequences for investigating specific genes. Therefore, the *A.nidulans EST* database created during the work described here will allow gene prediction, identification, annotation and suggest expression patterns as well as serve as a reservoir of vegetative

and asexually expressed gene sequences for immediate study.

1.4 cDNA Construction

In eucaryotic organisms the processes of transcription and translation are said to be uncoupled because the gene is transcribed in the nucleus and the mRNA passes through the nuclear pores into the cytoplasm before translation occurs. The coding regions in eucaryotic genes typically are not continuous but contain exons, regions of coding sequence interrupted by introns or noncoding sequences. The primary transcript, the heterogeneous nuclear RNA, found in the nucleus, contains both the intron and exon sequences (Abelson, 1979). Processing of the primary transcript to a messenger RNA requires the removal of the introns as well as the addition of a 5' cap (Furuichi, 1975) and a 3' poly A tail (Sachs, 1993) (Figure1).

Many methods of isolating RNA from sources as diverse as plant and mouse tissue have been developed but those including an isothiocyanate treatment step are commonly used (Chomczynski, 1987). Subsequent passage of the cellular RNA fraction over poly dT Sephadex enriches for the poly A containing mRNA (Chomczynski, 1987). The key discovery in cDNA synthesis was of an RNA-dependent DNA polymerase, reverse transcriptase (Baltimore, 1970) used for first-strand synthesis to produce a DNA copy of an mRNA (for example Houts, 1979) and RNase H (Krug, 1989), an RNA endonuclease associated with viral reverse transcriptase specific for DNA/RNA hybrids. The availability of these enzymes led to the development of protocols for synthesis of a double-stranded DNA copy (cDNA) and its subsequent insertion into a vector for maintenance and amplification. For example, the Lambda ZAP system was developed by Short et al. (Short, 1988), by adapting the methods of Gubler and Hoffman (Gubler, 1983), which uses poly dT primers containing a specific restriction enzyme sequence in first-strand synthesis to anchor the first strand synthesis start to the polyA tail. Second-

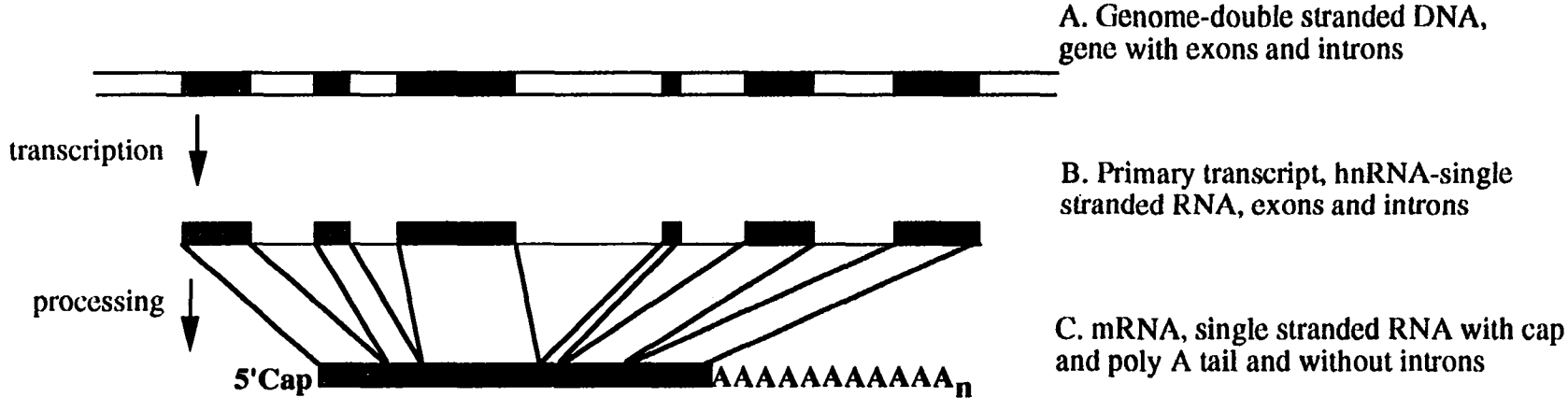


Figure 1. Steps required for generation of a processed mRNA ready for export to the cytoplasm and translation.

strand synthesis is performed by incubation with *E. coli* polymerase I in the presence of 5'-methyl dCTP. After the double-stranded cDNA is ligated to a second restriction enzyme cassette, it is digested with both enzymes. Since the methylated C residues prevent internal cleavage, only the sites flanking the cDNA are cleaved to create compatible ends on a double-stranded fragment which then can be directionally cloned into a suitable vector such as the *E. coli* bacteriophage lambda (Figure 2).

The lambda phage has been well characterized (Ptashne, 1992). A large region of this genome can be removed and replaced by ligation with foreign DNA inserts. The lambda Zap system, which is now a commercially available product , has been used for construction of a large number of cDNA libraries (Adams, 1995; Aramayo, 1990; Hillier, 1996; Nelson, 1997) and allows insertion of inserts up to 10 Kbp into the vector multicloning region (Short, 1988). The resulting molecule can be packaged in the presence of a "packaging mix" containing capsid proteins. This cell-free process prevents selection against any of the hybrid phage genomes. The lambda Zap vector also contains f1 and colEI origins of replication and the beta lactamase gene for ampicillin resistance which allows conversion of a hybrid phage genome to a plasmid, and the lacZ gene containing the multicoloning site allows selection for insert-containing plasmids.

Coinfection of the primary lambda library with an f1 helper phage, a filamentous single-stranded phage containing an amber mutation, allows excision of the cDNA from the lambda portion of the vector. A nonsupressing host then is infected with the f1 lysate which prevents replication of the f1 phage genome (Dotto, 1984). Since f1 replication is not supported and the single-stranded form of the cDNA carries not only the f1 origin but also the colE1 origin of replication, a double-stranded form of the cDNA contained in a plasmid vector containing an ampR gene and lacZ interrupted by the cDNA insert results (Short, 1988). This plasmid form of the cDNA easily is maintained in the laboratory. It can be stably grown and double-stranded sequencing template can be prepared from it.

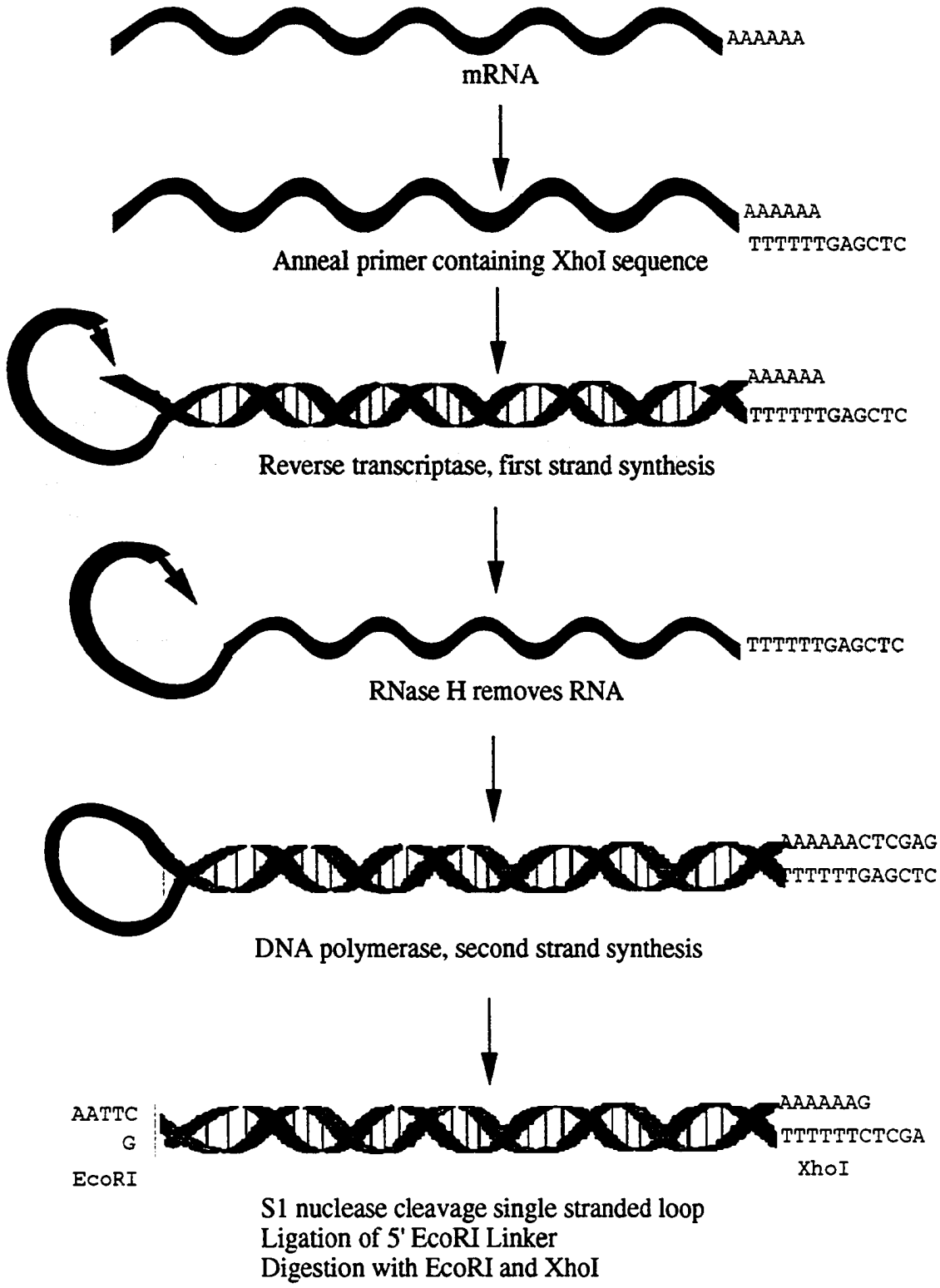


Figure 2. In vitro synthesis of double stranded cDNA from single stranded mRNA.

Figure 3 depicts a full length cloned cDNA with the position and direction of sequence for a pair of ESTs. Ideally, a full length cDNA will yield two ESTs where the5' EST covers the 5' UTR, translation start, gene coding region and the 3' EST starts in the poly A tail, includes the 3' UTR region, translation stop and coding region which may or may not overlap with the 5' EST. If the two ESTs do not overlap then custom synthetic primers can be synthesized and used for another round of DNA synthesis to obtain the full length cDNA sequence.

1.5 DNA Sequencing

Since the original Sanger dideoxynucleotide DNA sequencing method was reported (Sanger, 1977), modifications have been introduced to improve both its efficiency and ease of use, making it the primary sequencing technique employed today. Two recent key modifications will be mentioned here.

The Sanger method is an enzyme based synthesis method and the original enzyme used was the Klenow fragment of DNA polymerase I which contains both 5'-3' polymerizing activity and 3'-5' exonuclease activity but has the 5'-3' exonuclease domain removed (Klenow, 1971). The discovery of thermostable DNA polymerases, including Taq from *Thermus aquaticus* (Innis, 1988b), allowed the polymerization reactions to be incubated at high temperatures (72^0C for Taq) which inactivate Klenow enzyme but are required to reduce DNA secondary structure and GC compressions (Innis, 1988a). The original Taq polymerase has been modified to improve its efficiency by the removal of the 5' to 3' exonuclease domain and by a single amino acid change at position 667 from phenylalanine to tyrosine (Barnes, 1992; Tabor, 1995). This latter change reduced the discrimination against ddNTPs several thousand fold, a phenomenon first described by Tabor and Richardson (Tabor, 1995) who recognized that this discrimination could be altered by the presence of a tyrosine hydroxyl group at this position. The decreased discrimination against ddNTPs not only reduces the cost of DNA sequencing since much

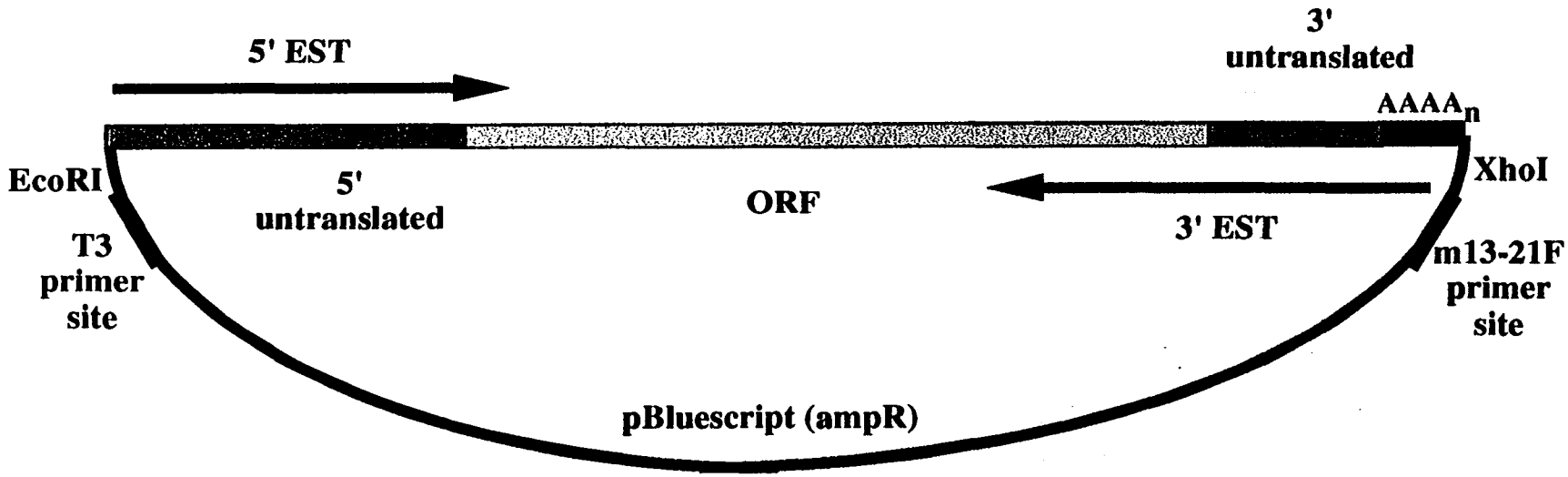


Figure 3. Position of an EST sequence pair on a cDNA clone. Blocks represent the cDNA insert. Red is the untranslated region, gray the coding region and green the polyA tail. The black line represents the cloning vector used in the *A. nidulans* library, pBluescript which contains the cloning sites, EcoRI and XhoI, and the sequencing primer sites (blue bars). The arrows represent the position and direction of the EST sequences.

less of the dideoxynucleotides need to be used in the cycle sequencing reactions but also improves the DNA sequence data quality and read length by producing more even signals and less background. Amplitaq-FS used in the work presented here is marketed by Applied Biosystems and contains both the Phe667Tyr mutation and an undisclosed N-terminal mutation to remove the 5' to 3' exonuclease activity (P. E. Applied Biosystems, 1998).

The introduction of fluorescent labeling and detection replaced the earlier radioisotope methods. Both fluorescent labeled primers and terminators now are available (Prober, 1987; Smith, 1986). For dye-labeled primers, the dyes are attached at the 5' most nucleotide base and for dye-terminator reactions, the fluorescent dyes are attached to the dideoxynucleotide base. Dye-labeled terminators have the advantage over labeled primers in that all four labeled dideoxynucleotides can be present in the same extension reaction, unlike labeled primer reactions where four separate reactions are required for each template. In addition, dye labeled dideoxynucleotides often eliminate any fold-back compressions because the large fluorescent group attached to the base of the ddNTP prevents base stacking on the 3' end of the newly synthesized chain. The background also is reduced because abortive stops which are not terminated with a labeled ddNTP are undetected (Lee, 1986). The fluorescent dyes typically used are fluorescein or rhodamine derivatives, each with a slightly different maximum emission wavelength. Applied Biosystems recently developed a dye set for terminator DNA sequencing which consists of energy-transfer dyes termed Big-Dye terminators, that use the 5-carboxy-d-rhodamine dyes as acceptor dyes coupled to 5- or 6-carboxy isomers of 4'-aminomethylfluorescein as the donor dye which results in a more even and intense signal (Rosenblum, 1997).

1.6 Instrumentation

Fluorescence-based DNA sequencers were first commercially developed by

Applied Biosystems (ABI 377 User manual, 1996) and all data reported in this dissertation was collected on the Applied Biosystems model ABI377. This instrument detects four different fluorescent dyes simultaneously and thus the four base-specific reactions can be pooled and electrophoresed in a single lane of the sequencing polyacrylamide gel. The ABI377 has a scanning argon laser and uses a CCD (charged coupled device) camera for detection. The argon laser excites the fluorophores and a series of lenses collect and focus the emitted light onto a spectrograph diffraction grating which separates the light based on wavelength into a predictably spaced pattern across the CCD camera. The collection software records the amplified emission signals and stores it in the Macintosh associated computer.

1.7 DNA sequence analysis

Several techniques were used to evaluate the sequence data in the EST and cosmid databases for quality and similarity with other sequences in the public databases.

Phred (Ewing, 1998; Ewing, 1998) and Phrap (Green, copyright 1994-1996) are companion programs developed in P. Green's group at the University of Washington. Phred examines the trace files obtained from the ABI377 by reading the raw DNA sequence trace data and then calls the bases. Phred also assigns sequence quality values to the bases by examining the four base traces in the region surrounding each point in the data set and predicting a series of evenly spaced locations. Phred then finds the center of the observed peaks in the trace files and the areas of these peaks relative to their neighbors. The observed peaks are compared to the predicted peaks to yield a calculated quality score (Green, copyright 1993-1996). Phrap, the assembly program, constructs contiguous sequences as a mosaic of the highest quality parts of reads using Phred quality values that are based on :

$q=-10 \log_{10} p$, where a quality value of 10, 20, 30 and 40 correspond to an error rate of 1/10, 1/100,1/1000, and 1/10,000 respectively. Phrap currently is the program of choice

for assembly of the human genome project sequences because it is the only public domain program which calculates individual quality scores.

Consed, a program for viewing and editing Phrap sequence assemblies, uses a color system to indicate the quality of the bases in the reads that are assembled as well as those of the resulting contiguous sequence (Gordon, 1998). Consed also contains an algorithm for determining quality of the contiguous sequence expressed in number of errors/10,000 bases. An error rate of less than 1 error /10,000 bases which generally corresponds to a coverage of each base once in both directions (B. A. Roe, personal communication), is considered acceptable for the human genome project and is the level to which the cosmids discussed in this work were finished.

A comparison of any DNA query sequence to the National Institutes of Health (NIH) database of all known nucleotide and protein sequences, GenBank (Benson, 1996), reveals if any previously sequenced nucleic acids have homology to the query sequence. GenBank is based at the National Center for Biotechnology Information (NCBI), a division of the National Institutes of Health (NIH) and is accessible through the web at URL ncbi.nlm.nih.gov. NCBI builds GenBank both from direct submissions of sequence data from authors and from scanning the journal literature. The GenBank database also includes the data deposited in the EMBL Data Library in the United Kingdom and the DNA Databank of Japan. Data is exchanged daily between the three international databases thus maintaining a comprehensive set of public sequence information. ESTs are the most rapidly-expanding source of new genes at GenBank and the dbEST division of GenBank has been established specifically for these sequences.

For sequence similarity searches, Blast (Basic Local Alignment Search Tool, developed at the NCBI) is a heuristic search used by the programs BlastP, which compares an amino acid query sequence against a protein sequence database; BlastN, which compares a nucleotide query sequence against a nucleotide sequence database; BlastX which compares the six-frame translation products of a nucleotide query sequence

against a protein sequence database; tBlastN which compares a protein query sequence against a nucleotide sequence database translated in all six reading frames; and tBlastX which compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database. These programs use the statistical methods of Karlin and Altschul (Altschul, 1990) to present the significant regions of similarity to a query sequence.

1.8 *Aspergillus nidulans* as a model organism.

The filamentous fungus, *Aspergillus nidulans*, is one of four historically important ascomycete models which include the yeasts, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* and a second filamentous species, *Neurospora crassa*. Seventy-five percent of the known fungi are filamentous. *A. nidulans* is unique among the four model organisms because of its extensive secondary metabolism. *A. nidulans* also is considered a model organism for investigation of molecular and genetic questions of closely related plant and human pathogens and industrially important members of the filamentous ascomycetes (Timberlake, 1990). Examples include *A. parasticus* and *A. flavus*, two of the producers of aflatoxin, a highly toxic DNA intercalating agent which contaminates harvested grain and nuts (Jimenez, 1991; Magnoli, 1998) and the industrially important *Penicillium chrysogenum*, the major producer of the antibiotics, penicillin (Josten, 1998) and griseofulvin (De Carli, 1998). Three Aspergillus species, *A. niger, A. flavus* and *A. fumigatus* , are common human opportunistic pathogens because they can infect at the site of a trauma such as a burn and can cause aspergillosis, a lung or sinus infection which can become systemic especially in immunocompromised individuals (Fenelong, 1999). *A. niger* and *A. oryzae* are two species used industrially to produce products such as ribonucleases, citric acid, alpha-amylase, invertase, pectinase, detergent enzymes, soy sauce and miso (a soy based food) (Carlile, 1994). Other filamentous ascomycetes such as *Magnaporte grisiae*, the causative agent in barley blight, are

responsible for significant crop losses worldwide (Bennett, 1985).

A. nidulans also is a eucaryotic model system for studying developmental regulation and cell differentiation (Morris, 1992). In the early 1950's, G. Pontecarvo selected *A. nidulans*, a nonpathogenic, easily grown member of the ascomycetes for intensive study, specifically to develop a well-characterized model system (Pontecorvo, 1953). His selection rationale was based on the criteria that *A. nidulans* has a well characterized sexual cycle, has a haploid nuclei for most of its life cycle, has unicellular ascospores, forms heterokaryon, grows on defined medium in liquid or on a solid surface, forms compact colonies on agar, has mature fruiting bodies which do not eject ascospores, and has easily micromanipulated asci from which a large number of ascospores can be obtained. Subsequent work has added two additional properties to this list, the ability of *A. nidulans* to grow as a synchronous culture for asexual development (Timberlake, 1980) and to be transformed after treatment with calcium chloride (Ballance, 1983) or by electroporation (Fromm, 1986)

Research findings have pointed to the success in using *A. nidulans* in a variety of studies. Dean and Timberlake (Dean, 1989) have demonstrated the plant pathogenic potential of *A. nidulans* by successfully establishing infections in cotyledons and fruit. This phytopathogenic character is reinforced by the findings that *A. nidulans* contains the inducible ability to catabolize quinate, a breakdown product of lignin (Hawkins, 1982) and to express two enzymes that are required in catabolism of pectin (Dean, 1989). The presence of the aflatoxin gene cluster lacking the last two enzymes in the pathway recently has been established (Adams, 1996). The entire *A. nidulans* penicillin biosynthetic cluster that originally was exploited in *Pennicillium* species has been isolated and characterized (MacCabe, 1990). *A. nidulans* now is considered the model organism for regulatory studies on both of these pathways. Many of the genes involved in *A. nidulans* mitosis also have been cloned and characterized. Several have homologs in other eucaryotes while others are unique to *A. nidulans* (Morris, 1992). *A. nidulans* also is a model system for

septation studies. Momany and Hamer recently found that as in *Drosophila*, the septin encoding gene is essential for *A. nidulans* growth. However, since neither *S. cerevisiae* nor *S. pombe* require septin (Brown, 1996) *A. nidulans* is a better model for studying this system.

Early sequencing results suggest that *A. nidulans* gene sequences will have application to other fungi and higher eucaryotes. For example, the results obtained from the analysis of *A. nidulans* cosmid W06E08 (Kupfer, 1997) showed only two ORFs with high homology to the completely sequenced *S.cerevisiae* ORFs, four with homology to other filamentous ascomycete genes, including the first transposase found in *Aspergillus* which is similar to a class found in *Drosophila*, two with homology to higher eucaryote genes (squid, and mouse). In addition, there are four predicted genes which have no significant similarity to any GenBank entries. An additional genome region of *A. nidulans* sequenced and analyzed as part of this dissertation work shows similar homology results as W06E08 and as well has homologous *A. nidulans* ESTs for ten of the 17 predicted genes. Thus, an EST sequencing project from an *A. nidulans* cDNA library most likely would have a broad application to other eucaryotes. In addition, there could be a number of new genes discovered during such an EST project that would complement the genome information from *S. cerevisiae* and provide a basis for further genomic structural and functional studies of this filamentous fungi.

Chapter II

Materials and Methods

Section 1. cDNA Library

The *Aspergillus nidulans* cDNA library that served as the source of template in the EST project was constructed by Rodolfo Aramayo, Texas A&M University. This library was constructed in 1990 using poly A RNA harvested from *A. nidulans*, strain FGSC A26 (veA1, bio), which had undergone development for 24 hours on a solid surface with an air interface and therefore contained cDNAs from both vegetative mycelial cells and cells involved in asexual reproduction. The cDNAs were cloned directionally such that the 5' end was adjacent to the Eco RI site and the 3' end the XhoI restriction site of the LambdaZap II vector arms (Stratagene Cloning Systems) (Aramayo, 1990).

The *A. nidulans* was grown in minimal medium (Kafer, 1977) plus biotin prepared as follows: 50 ml/l 20x salt solution (120 g/l $NaNO_3$, 10.4 g/l KCl, 10.4 g/l$MgSO_4.7H_2O$, 30.4 g/l KH_2PO_4), 1 ml/l trace elements (22 g/l $ZnSO_4.7H_2O$, 11 g/l H_3BO_3, 5 g/l $MnCl_2.4H_2O$, 5 g/l $FeSO_4.7H_2O$, 1.6 g/l $CoCl_2.5H_2O$, 1.6 g/l $CuSO_4.5H_2O$, 1.1 g/l $(NH_4)_6 Mo_7O_{24}.4H_2O$, 50 g/l Na_4 EDTA), 10 g/l D-glucose, 0.5 ml/l 10 mg/ml biotin in 95% ethanol. One liter of minimal medium was inoculated with $4x10^8$ conidia and shaken at 37^0C for 24 hours. Then 100 ml aliquots were harvested and transferred to Whatman #1 filter paper (9 cm diameter). The filters were laid down onto glass beads in a Petri dish containing 25 ml of minimal medium with biotin. Petri dishes were placed in a metal tray containing water to provide a moist atmosphere and incubated at 37^0C for 24 hours. The synchronously developing cells then were washed from the filter. The cells were lysed by grinding in liquid nitrogen. Total RNA was

isolated by the guanidine isothiocyanate method and the polyA RNA was purified twice by passage over oligo (dT) cellulose (Chomczynski, 1987). First and second strand synthesis, ligation of the cDNA into LambdaZap vectors and packaging was performed using the reagents and protocol provided in the ZAP-cDNA synthesis kit (Stratagene #200402, R. Aramayo, personal communication)

2.1 Characterization and Preparation of cDNAs

The cDNA library was converted from the LambdaZap form, provided by R. Aramayo and R. Prade, to a plasmid form in a mass excision by the following procedure:

1) f1 helper filamentous phage was added at a 1:1 phage:cell ratio and the lambda library at a 10:1 phage:cell ratio to strain XL1-Blue MRF' (an f1 filamentous phage host) using sufficient lambda library to equal ten times the primary library size of approximately $5x10^5$ plaques.

An overnight culture of XL1-Blue-MRF' cells grown in LB, 0.2% maltose and 10mM $MgSO_4$ was collected by centrifugation and resuspended at an A_{600} of 1.0 (approximately $8x10^8$ cells/ml). Cells, lambda library and helper phage were mixed and incubated in a 50 ml conical tube at 37^0C for 15 minutes. Twenty ml of LB broth was added and the cells were incubated for 3 hours at 37^0C with shaking. The tube was heated at 65^0C for 20 minutes to increase cell lysis.

2) The supernatant containing the released single-stranded filamentous phage containing the minus strand of the excised phagemid, pBlueScript SK- with the cDNA insert was centrifuged at 1000xg for 10 minutes to pellet cell debris and the supernatant was transferred to a fresh conical tube. The f1 supernatant could be stored for up to six months at 4^0C before any significant loss of titer was seen and an additional excision was necessary.

3) *E. coli* strain SOLR (not an f1 host) was infected with the filamentous phagemid in the absence of helper phage and plated on L broth solid medium containing 100 ug/ml ampicillin, 4 mg/ml Xgal, 5 mg/ml IPTG. To titer the excised phagemids, 5 ul

of the supernatant was mixed with 200 µl of a overnight culture of SOLR, collected and resuspended at one-half volume in 10 mM MgSO4, and incubated at 37^{0}C for 15 minutes. 100 µl of the mixture was plated onto a LB amp plate and incubated overnight at 37^{0}C. Generally, 2 µl to 5 ul of lysate per 200 µl of freshly prepared SOLR cells was used for single-colony isolation.

2.2 Growth of cDNA Clones and Preparation of Glycerol Stocks

The SOLR colonies containing cDNA phagemids were examined for insert based on the blue /white β-galactosidase selection where white indicates an insert in the multicloning site of the vector. The ratio of white to blue was noted (Table 1). White colonies were picked from the LB plates and inoculated into 1.5 ml Terrific Broth (12 g Bacto-tryptone, 24 g Bacto-yeast extract, 4 ml glycerol per 900 ml) with salts (2.31 g KH_2PO_4, 12.54 g K_2HPO_4 per 100 ml) containing 100 ug/ml ampicillin in sterile 96 well blocks. After incubation for 18 hours at 37^{0}C with aeration, the cells were placed on ice. Two microtiter plates per block were prepared as freezer stocks by adding 40 µl 50% sterile glycerol per well and 100 µl of cell culture transferred from each of the 96 wells of the blocks to the microtiter plate with a 12 channel pipettor for both transferring and mixing the cells with the glycerol. The microtiter plates were labeled and stored at -80^{0}C. One set of glycerol stocks was sent to the Fungal Genetics Stock Center, Kansas Medical School, Kansas City, KS for archiving and distribution. The remaining 1.3 ml of cell culture was collected via centrifugation at 3000 rpm for 5 minutes in a Beckman GS-6R centrifuge. The cell pellets were frozen at -20^{0}C.

Table 1. Summary of *A. nidulans* cDNA Library Characteristics

λ Library Titer	Phagemid Yield from F1 Supernatant	Percent Clones with Insert	Average Insert Size (36 Clones)
7x10^{9}pfu/ml	8x10^{5}cfu/ml	62%	1.0Kbp (0.5-3.0Kbp)

2.3 Semi-automated alkaline lysis isolation of small-insert DNA

The phagemids were isolated via a modified alkaline lysis protocol (Birnboim, 1979) using a Beckman Biomek 2000 automated workstation and a Hydra96 (Robbins Scientific) as described below.

The cell pellets were thawed for 15 minutes at room temperature and then 200 ul TE RNase A (50 mM Tris pH8, 10 mM EDTA, 100 μg/ml RNase A) was added to each well in the 96 well block by a Hydra96. The blocks were placed on a titer plate shaker for 10 minutes at setting 7.5 to resuspend the cell pellets.

A program originally written for the Biomek by Judy Crabtree (Crabtree, 1997) was modified to begin with the addition of 200 ul of alkaline lysis solution. The modified program "td sds-to-end" was selected and the blocks and reagents were placed on the worksurface as directed in the program layout. Following the addition of 200 μl of alkaline lysis solution (0.2 N sodium hydroxide and 1% dodecyl sulfate, SDS) and 100 ml of 3 M potassium acetate, pH 4.8, the blocks were removed from the Biomek and covered with an acetate sealer. Each block was vortexed for 60 seconds, a critical step for formation of a solid debris pellet and cleared lysate.

The blocks were placed on a titer plate shaker at a setting of 6 for 5', then stored at -20^0 C for 1 hour. This was followed by centrifugation at 3000 rpm for 45 minutes in the Beckman GS-6R centrifuge. The blocks were returned to the Biomek worksurface, and 400 μl of cleared lysate were transferred to four fresh blocks. One ml of 100% ethanol was added to each sample manually and the blocks were placed at -20^{0}C overnight. The precipitated DNA was collected by centrifugation for 30 minutes at 3000 rpm in the Beckman GS-6R. The pellets were washed twice with 70% ethanol and the blocks drained on paper towels. The Hydra was used to add 100 μl of sterile deionized water to the pellets which were stored overnight at 4^{0}C following 15 minutes of shaking on a titer plate shaker at a setting of 7.5. Two μl aliquots were examined on a 1% agarose gel using 0.2 μg pGEM as a vector standard. The presence of low molecular

weight RNA, insoluble material remaining in the wells and the relative concentration of the isolated cDNA template was examined. If insoluble material was present, the templates were treated by the addition of equal volumes of 7.5 M neutral potassium acetate. The mixture was vortexed for 30 sec. and placed at -20^{0}C for 30 min. The debris pellet was collected by centrifugation for 30 min. at 3000 rpm. The supernatant containing the cDNA was removed to a fresh block and two volumes of 100 % ethanol were added followed by incubation at -20^{0}C for 30 min. The DNA pellets were collected by centrifugation at 3000 rpm for 30 min and washed twice with 70 % ethanol. After drying in vacuo, the DNA was resuspended in 50 ul sterile deionized water and rechecked on an agarose gel before use in a sequencing reaction.

Thirty-six of the first block of clones isolated were examined for insert size by restriction analysis with XhoI and Eco RI. The approximate size of the insert was determined by comparison with the size markers 100 base-pair DNA ladder (Gibco BRL) and Hind III fragments of lambda DNA (Gibco BRL) (Table 1). When multiple bands appeared other than the vector, their sizes were added.

2.4 Optimum Sequencing Conditions

In trial sequencing reactions, a variety of parameters were examined to determine the optimum conditions for DNA sequencing. Both the quality of the sequence data and the number of bases obtained were examined because, for single-pass sequencing in the forward and reverse direction, the extent of data acquired should be maximized to improve the chance that the combined data for the forward and reverse reactions will yield a contiguous segment containing the entire coding region.

The thermocycling reaction employed Taq DNA polymerase catalyzed chain extension using fluorescent dye-terminators containing only one unlabeled primer. The polymerization product was a single-stranded DNA nested-fragment set ending in a dye-labeled dideoxynucleotide. Since dye terminators were used, all four bases can be

included in each reaction so that only a single sequencing reaction is required instead of four as is necessary in dye-labeled primers. In addition, false stops caused by termination when the enzyme prematurely dissociates from the template were not detected because of the lack of incorporated dye terminator and foldback compression was prevented when the extended DNA chain correctly terminated with the bulky fluorescent dideoxynucleotide.

In the trial reactions, two forms of Taq DNA polymerase were examined. The first, Amplitaq FS (Applied Biosystems) contains two modifications: (a) an N-terminal deletion which eliminates the 5'-3' exonucleolytic activity and results in greater processivity of the enzyme, and (b), a residue change from phenylalanine to tyrosine in the ribose selectivity site which increases the affinity for dideoxynucleotides (Tabor, 1995). The thermocycling reactions using Amplitaq FS were performed in the presence of a heat-stable pyrophosphatase which hydrolyzes pyrophosphate maintaining favorable conditions for DNA polymerization. The second, KlentaqTR enzyme, has both modifications as described for Amplitaq FS, however the N-terminal deletion is not identical (Barnes, 1995; Korolov, 1995). Reactions with this enzyme were performed in the absence of a pyrophosphatase.

Four sets of reactions were tested using the same 24 DNA samples; each enzyme in the presence and absence of 5% dimethlysulfoxide (DMSO) a denaturing agent which can reduce secondary structure in the DNA. Primer,0.65 μM and 200-500ng DNA template were used in all cases. The primers used were the M13 -21 universal forward (5'-TGT AAA ACG ACG GCC AGT3') and SK- (5'-CGC TCT AGA ACT AGT GGA TC-3'). The universal forward primer sequence is located in the pBluescript vector 48 bp upstream of the 3' end of the cDNA insertion in the XhoI site and the SK- primer site is located 12 bp upstream of the 5' end of the cDNA insertion in the Eco RI site. Amplitaq FS was a commercially available mixture (PerkinElmer) containing: Amplitaq DNA polymerase FS, Tris-HCl pH 9.2, dCTP, dATP, dTTP, dITP, ddATP-dye terminator

(R6G), ddGTP-dye terminator (R110), ddTTP-dye terminator (6-TAMRA), ddCTP-dye terminator (6-ROX) all rhodamine fluorescent dyes, magnesium chloride, thermal-stable pyrophosphatase. The concentrations for each reaction component are not given by the manufacturer (PerkinElmer, 1995). All Amplitaq reactions were in a final volume of 5 ul. The Klentaq TR reaction mix contained:50.0 mM Tris-HCl pH9.2, 16.0 mM ammonium sulfate, 3.5 mM $MgCl_2$, 1.0 mM $MnSO_4$, 150 uM dATP, dTTP, dCTP, 450 uM dITP, 1/2000 dilution of Perkin Elmer #401384 "dyedeoxy terminators", 1/200 dilution of KlenTaq TR as supplied (Wayne M. Barnes). Twenty µl reactions were used. FS-containing samples were thermocycled for 25 cycles of 96^0C for 10 sec, 50^0C for 5 sec and 60^0C for 4 minutes while the Klentaq-containing samples were incubated for 25 cycles of 97^0C for 50 sec and 65^0C for 4 minutes.

Unincorporated dye-labeled dideoxynucleotides were removed by chromatography through Sephadex G-50 into a 96 well microtiter plate The filtered samples were dried in a vacuum oven at room temperature before addition of 1 µl of a 5:1 formamide:0.1% blue dextran, 1 mM EDTA loading buffer. Slab gel electrophoresis was on 48 cm, low fluorescent glass plates separated by 0.2 mm spacers and containing 5.3% Long-Ranger acrylamide gel mix (FMC) with 8 M urea. Electrophoresis was performed at 2.3 kV, 52^0C for 10 hr in 1x TBE buffer (10.8 g Tris base, 5.5 g boric acid, 8.4 g EDTA in 1 L) on an Applied Biosystems ABI377 automated sequencer. The results summarized below show the number of high quality bases read. A high quality base is defined as the sequence length up to the position where two bases in five were called as N using the ABI basecalling software.

TABLE 2-Preliminary Sequencing Results-24 Samples

	Taq FS		KlenTaq TR	
	+DMSO	/-DMSO	+DMSO	/-DMSO
Average # of bases	715	650	586	540

The optimum runs were achieved with the Taq FS chemistry. Since 5% DMSO

has positive effects on sequences which exhibit secondary structure and does not appear to have any negative effect on Taq FS reactions, 5% DMSO was used routinely for sequencing the cDNAs.

Two reverse primers were examined, SK- , 12 bases upstream of the Eco RI site and T3 (5'-CGA AAT TAA CCC TCA CTA AAG3') which is 64 bp upstream. Both were used successfully, however the position of the SK- primer was too close to the insert to enable the program used later in analysis to routinely find the position of the Eco RI site which flagged the beginning of the insert sequence. This was because the signal generated for the first 20-25 bases often was too high to be accurately basecalled. Therefore, the T3 primer was used routinely for the 5' EST sequences.

Roughly half of the ESTs were sequenced with the rhodamine fluorescent dyes described above. Following the introduction by Perkin-Elmer Biosystems of the energy transfer dyes, termed BigDye energy transfer-based terminators (Rosenblum, 1997), the EST sequencing reactions were thermocycled using a 1:16 dilution of the Perkin-Elmer recommended reaction mix of BigDyes containing Amplitaq FS enzyme. The energy transfer dyes are constructed by linking one of two donor dyes, the 5- or 6-carboxy isomer of 4' aminomethylfluorescein (5 Fam or 6 Fam) with a 5-carboxy-dichloro-rhodamine dye. The ddATP contains the donor dye 5-FAM linked to the ddATP acceptor dye dR110-2. 6-FAM is linked to ddCTP acceptor dye dROX-2, ddGTP acceptor dye dR110-2 or ddTTP acceptor dye dTAMRA-2. A typical reaction contained 2-3 µl template DNA, 1 µl 25% DMSO, 1 µl 13 uM universal forward or 26 µM reverse primer, 2 µl diluted terminator:enzyme mix (1 µl terminator:enzyme mix :3 µl 5x buffer: 50 mM Tris, pH9.0, 10 mM $MgCl_2$). The average number of bases, 556 (average of 24 samples) cannot be compared directly to that seen with the dRhodamine dyes since the run time was changed from 10 hours to 6 and the average number of bases obtained with a six hour run was reduced to 550-600.

2.5 Cycle Sequencing Conditions

The dilution of the BigDye components in the thermocycling reactions, was extremely cost effective but required an additional number of cycles to achieve the same intensity of signal seen with the undiluted D-rhodamine dye mix. Therefore, 60 cycles instead of the ABI recommended 30 cycles were used with the same temperature and time as used previously for TaqFS: 95^0C for 10 seconds, 50^0C for 5 seconds, 60^0C for 4 minutes.

All thermocycling reactions were performed in 384 well plates (Cycleplate 384, Robbins Scientific) on the Perkin-Elmer 9600 or 9700 thermocycler. Since the 384 well plates have 96 wells each divided into quadrants, samples from two blocks of 96 cDNA templates could be thermocyled with both the forward and reverse universal sequencing primers simultaneously. The reactions were prepared by a semi-automated procedure using the Hydra96.

The pBluescript-based DNA sequencing template required for the reaction was transferred from the first 96 well DNA template block and dispensed to two of the four quadrants of the 384 plate using the Hydra 96. This process was repeated to transfer a second block of templates into the last two quadrants of the 384 plate. A sufficient volume of the reaction mix for two forward reactions was aliquoted to 96 wells of a 96-well thermocycle plate (Robbins Scientific). This was transferred into two quadrants of the 384 plate. The reverse reaction mix then was dispensed into the remaining two quadrants. Thus, a single 384 well plate contained both the forward and reverse sequencing reactions for each sample in two 96 well sets of subclones or cDNA.

2.6 Unincorporated Dye Terminator Removal.

Following thermocycling, unincorporated terminators and buffer salts were removed from the reaction by gel filtration through Sephadex G-50 mini-columns prepared in 96-well filter plates (Millipore, MASVN6550). To prepare the mini-filtration

columns, 200 µl dry Sephadex G-50 was added to each well. Then 400 µl of distilled-deionized water (dH_2O) was dispensed into each well using the Hydra96. The plates were allowed to hydrate at 4^0C for two hours to overnight. Each 96-well plate was centrifuged for 2.5 minutes at 1500 rpm. An additional 100 µl dH_2O then was added followed by centrifugation for 2.5 minutes at 1500 rpm. Ten microliters of dH_20 was added to each sample in the 384 well cycle plate once cycling was completed to dilute the reaction. The 384 well plate was centrifuged for 2 minutes at 1500 rpm to collect the liquid to the bottom of the well. The entire volume for each reaction was transferred to the top of one of four 96-well Sephadex G-50 column filter plates using the Hydra96. A 96 well microtiter plate was secured under each filter plate and the filter plate and attached microtiter plate were centrifuged for 3 minutes at 1500 rpm to collect the purified reaction products in the microtiter plate. The reactions then were dried under vacuum without heat and stored at -20^0C until loading dye was added prior to loading onto the sequencing gels.

2.7 Automated Data Collection

Electrophoresis was performed as described above for six hours on an ABI 377 automated sequencer. Data was automatically collected and analyzed by the Data Collection and Data Analysis software on a Macintosh computer associated with the ABI377 sequencing equipment. Analyzed data then was transferred to the networked Unix-based Sun workstations for assembly and analysis.

Section 2. *Aspergillus nidulans* Cosmids

2.8 Background-Method of Construction of Cosmids

Pulsed-field gel electrophoresis (PFGE) technology has allowed the separation of linear double-stranded molecules up to 10 megabase pairs (Mbp) (Vollrath, 1987). A version of PFGE, contour-clamped homogeneous field electrophoresis (CHEF) was used

successfully by Brody and Carbon (Brody, 1989) to separate the eight chromosomes of *A. nidulans*. The CHEF method employs a hexagonal array of 24 electrodes which produce a homogeneous electric field at an angle of 120^0 (Vollrath, 1987). The 2.9 Mbp chromosome IV was well resolved from the other chromosomes with this procedure.

Using the CHEF method developed by Brody and Carbon, Prade, et al. (Prade, 1997) isolated chromosomal material for electrophoresis from *A. nidulans* mycelial cells which were treated to form protoplasts. The protoplasts were mixed with an equal volume of 1.4% agarose in a plug mold that then was mixed with an SDS lysis solution, incubated and washed with EDTA. Plugs were inserted into a 0.8% agarose gel and electrophoresed with three pulse intervals, each with increasing duration of the forward pulse, for a period of 156 hours (Brody, 1989). Using this approach, Prade, et al. (Prade, 1997) isolated chromosomal DNA for creating a tiled set of cosmids for each *A. nidulans* chromosome, sub-cloning into the cosmid vector, pWE15, which encodes both kanomycin and ampicillin resistance and contains the colE1 origin of replication.

The three cosmids sequenced in this study were from the chromosome IV tiled set.

Cosmid W06E08 was obtained from Rolf Prade, Oklahoma State University, as dried DNA. The pellets were resuspended in 10:0.1 TE and transformed into $CaCl_2$ competent ED8767 cells. Cosmids W02H02 and W30B01 were received from Nancy Keller, Texas A&M University, as *E. coli* transformed cells.

2.9 Large scale DNA Template preparation

During the course of this research, two large scale isolation methods were used. Cosmid W06E08 was isolated using the diatomaceous earth-based method (Carter, 1993). This method which is based on the high salt binding and low salt release of DNA is less costly and hazardous than isolation using equilibrium centrifugation in cesium chloride-ethidium bromide gradients (Tanaka, 1977). However, the diatomaceous earth based method had two disadvantages. First, the yield of DNA often was low (Chen,

1997) and second, the method was not amenable to automation. Therefore a modification of the alkaline lysis procedure (Birnboim, 1979), termed a double acetate procedure, also was employed to successfully isolate cosmid DNA since this protocol gave more consistent results than the diatomaceous earth procedure described above and could be successfully automated. Chen and Kupfer (unpublished results) subsequently developed a protocol using the Beckman Biomek 2000 to automate the large scale cosmid, BAC and PAC DNA isolation. Cosmids W02H02 and W30B01 were isolated using this semi-automated double acetate method.

2.10 Large-insert DNA Isolation Using Diatomaceous Earth

A smear of colonies of cosmid W06E08 were picked and used to inoculate a 12x75 mm Falcon tube containing 3 ml LB medium (10g Bacto-Tryptone, 5 g Bacto-yeast extract and 10 g NaCl in 1 liter deionized water, autoclaved) with 100 ug/ml ampicillin. After incubation at 37^{0}C for 8 hours with shaking at 250 rpm, the culture was transferred to a 250 ml flask containing 50 ml of the same medium and incubated for an additional 8 hours. All 53 ml of the culture was transferred into a 2 liter flask containing 1 liter of the same medium and incubated for an additional 8 hours.

The cells were harvested by centrifugation at 6000 rpm for 15 minutes in 500 ml bottles in the Sorvall RC5-B using the GS-3 rotor. Cell pellets were frozen at -70^{0}C until processed. The cells were thawed and resuspended completely in 35 ml of GET (50 mM glucose, 25 mM Tris-HCl, pH 8.0 and 10 mM EDTA). To the resuspended cells was added 70 mg of lysozyme for a final concentration of 2 mg/ml. After gentle mixing to dissolve the lysozyme, the solution was incubated at room temperature for ten minutes.

Cells were lysed by addition of 70 ml of alkaline lysis solution (0.2 N sodium hydroxide, NaOH, and 1% sodium doedecyl sulfate, SDS) followed by gently swirling until the solution was homogenous. After the mixture was incubated for five minutes on an ice-water bath, 70 ml of 3 M sodium acetate, pH 4.8 was added with gentle inversion

of the bottle several times. The bottle then was placed on an ice-water bath for 60 minutes. The lysate was cleared from precipitated SDS, proteins, membranes and chromosomal DNA by passing through a double layer of cheesecloth into a sterile 250 ml centrifuge bottle, followed by centrifugation at 10,000 rpm for 30 minutes at 4^0C in the Sorvall RC5-B centrifuge using the GSA rotor. The supernatant was transferred into a 250 ml sterile centrifuge bottle and 20 mg/ml DNase-free RNase A (20mg/ml RNase A in 1 mM, pH 4.5, boiled for ten minutes and cooled slowly to room temperature) was added to a final concentration of 100 ug/ml. Following incubation at 37^0C for 30 minutes, the supernatant was divided equally between two 250 centrifuge bottles, an equal volume of isopropanol was added to each and mixed by inversion. The solution was held at room temperature for five minutes. The DNA pellet was collected by centrifugation for 30 minutes at 9,000 rpm for 30 minutes. The supernatant was decanted and the pellets drained. Each DNA pellet was dissolved in 10 ml of 10:1 TE (10mM Tris-HCl pH 8.0 and 1 mM EDTA) and transferred to a 50 ml Corning glass centrifuge tube. To each was added 20 ml of defined diatomaceous earth-guanidine-HCl (defined diatomaceous earth suspended as 100 mg/ml in 6 M guanidine-HCl, 50 mM Tris-HCl, 20 mM EDTA, final pH 6.4). This mixture was inverted several times during a five minute incubation at room temperature followed by centrifugation in the Beckman GS-6R tabletop centrifuge at 3000 rpm for ten minutes. The supernatant was decanted and 40 ml of wash buffer (equal volume of 10:1 TE and ethanol) was added to each tube to resuspend the diatomaceous earth-DNA complex. After an additional centrifugation and decanting of the supernatant, the diatomaceous earth-DNA complex was suspended in 40 ml of acetone and centrifuged for 10 minutes. The supernatant was decanted and the pellets were dried in a vacuum oven.

The DNA was eluted from the dry diatomaceous earth-DNA after resuspension in 20 ml of 10:1 TE buffer and incubation at 65^0C for ten minutes. The diatomaceous earth was collected by centrifugation for 15 minutes at 3000 rpm in the GS-6R. The

supernatant containing the DNA was decanted and the elution step was repeated. The two supernatants were pooled. The DNA was precipitated by adding of two volumes of 95% ethanol containing 0.12 M sodium acetate (NaOAc) and incubating in an ice-water bath for 45 minutes. The DNA pellet was collected by centrifugation for 15 minutes at 3000 rpm. The final pellet was washed with one volume of 70% ethanol and dried. The dried pellet was dissolved in 2 ml of dH_2O and the DNA concentration was determined by measuring the absorbance at 260 nm.

2.11 A Protocol for the Semi-automated Double Acetate Isolation of Large-insert Template DNA

The alternative protocol, a semiautomated version of the alkaline lysis protocol was used to isolate cosmid DNA from W02H02 and W30B01. This protocol, described below, includes an acidic sodium acetate precipitation and a second neutral potassium acetate precipitation. Both appear to be necessary and sufficient for a DNA preparation which can be used to generate shotgun subclones and which, with an additional phenol/chloroform treatment, can be used as a template for direct DNA sequencing.

For each cosmid, a smear of colonies was used to inoculate 3 ml of LB medium containing the 100 μg/ml ampicillin in a 12 x75 mm Falcon tube. The culture was incubated at 37°C for 8-10 hr with 250 rpm shaking. The entire culture was transferred to a 250 ml flask containing 50 ml of the same medium and incubated for 8-10 hr under the identical conditions. All 53 ml of the culture was added to a 2 liter flask containing 1 liter LB medium and antibiotic and incubated for additional 8-10 hr.

The cells were harvested by centrifugation at 7000 rpm for 15 min. in a 500 ml bottle in the RC5-B using the GS3 rotor. After the cell pellets were frozen and stored at -70°C, they were thawed on ice and resuspended in 80 ml of 50:10 TE (50 mM Tris-HCl, pH 8.0, and 10 mM of EDTA, pH 8.0) by gently pipetting up and down. The resuspended cells were dispensed in 200 μl aliquots into four 96-deep well blocks with

repeat pipet.

A series of programs baciso1-baciso5 were written to perform portions of the isolation protocol using the Beckman Biomek 2000. The Biomek was turned on and calibrated. The program baciso1 was opened, the MP200 pipet head, one block of 250 ul tips, and four blocks of resuspended cells were placed on the platform of the Biomek as shown on the program layout. baciso1 adds 200 ul of alkaline lysis solution and 200 ul 3M NaOAc pH4.8 to each well. These steps take 45 min. Reagents for this step are 100 ml of alkaline lysis solution (0.2 N NaOH, 1% SDS) and 100 ml of 3 M NaOAc, pH 4.8 which were added to the two designated reservoirs. The four blocks were removed following the addition of the lysis solution and placed on a titer plate shaker for 5 min. (setting of 5) to mix the lysis solution and to shear the genomic DNA to improve the efficiency of clearing the lysate. The blocks then were returned to the Biomek and the program continued by adding 200 µl sodium acetate to each well. The blocks were removed from the Biomek platform and again placed on a titer plate shaker for 5 min. (setting of 5) and frozen at -20°C for 40 min. Precipitated SDS, cell membrane and chromosomal DNA were collected by centrifuging the blocks in the Beckman GS-6R centrifuge at 3000 rpm for 30 min.

The program baciso2 was selected, and the blocks were returned to the Biomek platform along with four new blocks as shown on the program layout on the computer screen. The Biomek transferred 400 µl of cleared lysate from each well to an empty well in a new block over a 30 minute period. If floculent material (typically inadequately sheared genomic DNA and protein) inadvertently was transferred to the fresh blocks, the transfer blocks were placed on titer plate shaker for 5 min. and returned to -20^{0}C for 20 min. The 30 min. centrifugation was repeated, the blocks were placed on the Biomek and the baciso2 transfer to fresh blocks was repeated. The blocks containing the cleared lysate were removed from the Biomek and 250 µl of 100% isopropanol was added to each well manually using a repeat pipet. The blocks were covered with acrylic sealers and inverted

three times to mix. The sealers were removed immediately and the blocks incubated at room temperature for 5 min. To collect the DNA pellets, the blocks were centrifuged in a Beckman GS-6R centrifuge at 3000 rpm for 30 min. The isopropanol was decanted and the blocks drained by inversion on paper towels for 10 min.

Program baciso3 resuspended the DNA pellets and added 7.5 M potassium acetate (KOAc) for a second acetate precipitation. The blocks were returned to the Biomek platform as shown on the program layout. Fifty ml of 10:1 TE-RNaseA (10 mM Tris-HCl, pH 7.6, 1 mM of EDTA, pH 8.0, 100 ug/ml RNaseA) and 25 ml 7.5 M KOAc were prepared and placed in the designated reservoirs. The Biomek added 100 µl 10:1 TE to each well followed by a gentle pipetting up and down twice and then paused. This step took 15 min. The blocks were incubated at 37^{0}C for 10 min. allowing better resuspension of the DNA pellets and RNaseA activity. Following incubation, the blocks were placed on the titer plate shaker at speed 5 for 2 min. to make sure the pellets were completely dissolved. The blocks were returned to the Biomek and baciso3 was continued. The Biomek added 50 µl of 7.5 M KOAc to each well over a 10 min. period. Then the solution contained in the four blocks was pooled into one block by baciso4, a 15 min. step. Following the pooling there was only one block which contains DNA-acetate solution. The block was placed at -20°C for 45 min. The potassium acetate-cellular debris pellet was collected by centrifugation in the Beckman GS-6R centrifuge at 3000 rpm for 20 min. The block was returned to the Biomek and baciso5 was selected to transfer 550 µl of DNA-containing supernatant from each well of the old block to a fresh block concluding the Biomek portion of the protocol.

The block then was removed from the Biomek and 1.25 ml of cold 100% ethanol was added to each well of the block using the repeat pipet. The block was covered with an acetate plate sealer and inverted three times to mix. The sealer was removed immediately and the block was incubated in an ice-water bath for 45 minutes. The DNA pellet was collected by centrifugation at 3000 rpm for 30 minutes in the Beckman GS-6R

centrifuge. Each well was washed with 1 ml 70% ethanol and dried briefly in the vacuum oven. To each well was added 40 ul of dH_2O to dissolve the DNA. The block was placed on a titer plate shaker for 10-15 min. and then placed over night at 4^0C to allow the pellets to completely resuspend. The block was briefly centrifuged at 1000 rpm to collect the solution to the bottom of each well. The DNA solution was pooled into a 12 x75 mm Falcon tube.

For final purification, the DNA was brought to 50 mM Tris-HCl pH 8.0 with the addition of 1M Tris-HCl pH 8.0. The solution was divided into 0.5 ml aliquots in 1.5 ml Eppendorf tubes and 500 ul equilibrated phenol (in 50mM Tris-HCl pH 8.0) was added to each tube and mixed gently by inversion. The phases were separated by centrifugation for at 12,000 rpm for 5 min. at room temperature in a microcentrifuge. The upper aqueous phase was remove to a fresh 1.5 ml Eppendorf tube and 500 µl 1:1 equilibrated phenol:chloroform was added to each tube followed by gentle inversion and centrifugation at 12,000 rpm for 5 min. at room temperature. Again, the upper aqueous phase was removed to a fresh 1.5 ml Eppendorf tube and 500 µl isopropanol was added to precipitate the DNA. The tubes were inverted gently several times and incubated for 5 min. at room temperature. This was followed by centrifugation for 5 min. at 12,000 rpm at room temperature. The DNA pellets were washed twice with 500 µl 70% ethanol, dried in a vacuum chamber for 5-10 min. and resuspended in 400 µl 10:0.1 TE. Any DNA remaining on the sides of the tube was dissolved by pipetting buffer over the sides then the tubes were incubated at 37^0C for 15 min. and stored over night at 4^0C to ensure that all the large insert DNA was dissolved completely .

The DNA concentration was measured by determining the absorption at 260 nm. An uncut aliquot and an Eco RI restriction digest of 10-15 ul of the DNA preparation were examined on a 1% agarose gel for chromosomal contamination, presence of RNA and restriction digestion fragment pattern. If this DNA contained little visually detected genomic DNA, then it was suitable for nebulization and as a template for direct

sequencing. Typical yields per liter of original cell growth for the cosmids was roughly 2.5 mg for the cosmids.

2.12 Cosmid Sequencing Strategy

A blending of sequencing strategies was developed over the period of this dissertation research and was optimized for the three cosmids sequenced in this study. The random shotgun sequencing method was used for the initial phases of sequencing and the directed sequencing approach was used for closure of cosmids W02H02 and W30B01. The directed approach uses specific subclones or the cosmid as template for thermocycling in conjunction with region- specific DNA primers.

2.13 Random Shotgun DNA Subclone Library Construction

The cosmid DNA was randomly sheared by nebulization (Roe, 1995). The size range of the sheared fragments was controlled by the percent of glycerol in which the cosmid DNA was suspended, the temperature, and the pressure of nitrogen gas passed through the DNA solution. The resulting fragments that have "ragged" ends with either a 3' or a 5' overhanging end were made blunt-ended by treatment with the Klenow fragment of DNA polymerase I and T4 DNA polymerase. The fragments were further treated with kinase to make them suitable for blunt end ligation with Sma I cut pUC18 treated with calf alkaline phosphatase. After end-repair and phosphorylation, the fragments were size-selected by electrophoresis on a low melt agarose gel using appropriate size markers. The DNA was recovered from the agarose by melting, treatment with phenol, and then treatment with phenol/chloroform. The fragments were used in a ligation reaction with Sma I cut, dephosphorylated pUC18 and electrotransformed into competent *E. coli.*

2.14 DNA Shearing by Nebulization

The nebulizer (IPI Medical Products, Inc., Chicago, IL) is a commercial product originally designed for use in creating fine droplets from liquid medications or nutrients allowing them to be aspirated by a patient.

A solution of large insert cloned DNA was placed in the bottom of the nebulizer bowl. Nitrogen gas was introduced from the top. The pressurized nitrogen gas was channeled into a small chamber that was in turn connected to a tube which extends through the sample to the bottom of the bowl. The flowing gas creates a vacuum, drawing the sample up the tube and out into the upper chamber where its collides with a protruding plastic surface breaking into fine droplets. The size of the droplets was inversely proportional to the flow rate of the nitrogen and the fragment size is inversely proportional to the viscosity and the temperature. Therefore, glycerol was added to the DNA sample to increase viscosity as well as to prevent sample freezing since the temperature of the nebulizer was reduced (Bodenteich, 1993).

Specifically, nebulization was performed on 50-100 µg of cosmid DNA in a final sample volume of 2 ml with 25% glycerol 1x TM buffer (50 mM Tris-HCl pH8.0, 15 mM $MgCl_2$) in a -20^0 C dry ice-ethanol bath. After 2.5 min. at 8 psi nitrogen, the sample was removed from the chamber, divided into four aliquots, ethanol precipitated and subsequently resuspended in 27 µl 1x TM buffer.

2.15 End-fill and Phosphorylation

The resuspended DNA fragments were treated with T4 polynucleotide kinase and Klenow DNA polymerase and T4 DNA polymerase by addition of 5 µl 10x kinase buffer, 5 ul 10 mM rATP, 0.25 mM dNTPs, 1 µl 3 U/ul T4 polynucleotide kinase, 2 ul 5 U/ul Kelnow DNA polymerase, 1 µl 5 U/ul T4 DNA polymerase and incubation at 37^0C for 30 minutes. The reaction was halted by heating at 70^0C for 10 minutes and addition of 10 µl agarose gel loading dye (0.02% bromphenol blue, 5 mM EDTA pH 8.0, 50%

glycerol).

2.16 DNA Fragment Size Selection

The end-repaired DNA fragments were size selected by separation on a 1% low-melt agarose gel. Electrophoresis was performed at 100 mA for 90 minutes with Hind III cut lambda DNA (Gibco BRL) and 100 basepair ladder (Gibco BRL) as size markers. Samples were removed by slicing agarose gel pieces from the 1-2 Kbp and 2-4 Kbp ranges. The gel pieces were heated to melt the agarose, equal volumes of TE equilibrated phenol was added and the sample tubes were vortexed. The extraction was repeated with 1:1 phenol:chloroform. The aqueous phase was ethanol precipitated and the DNA pellet was washed twice with 70% ethanol. After drying in vacuo, resulting pellets were resuspended in 11 µl dH_2O.

2.17 Ligation of Fragments and Transformation of Plasmids

Examination of the relative concentration of the recovered sample was performed by electrophoresis of 1 µl of the DNA fragment with 0.1 and 0.2 µg pGEM as concentration standards and Hind III digested lambda and 100 basepair ladder as size standards. The remaining 10 µl of fragment DNA was diluted by two-fold and a series of ligation reactions were set up using 2 µl, 1 ul and 0.5 ul as the DNA fragment template volume with 2 ul Sma I cut CIP dephosphorylated pUC18 (10 ng/µl), 1 ul 10x ligation buffer, 1 µl T4 DNA ligase (400 U/µl) in a 10 ul reaction volume. Incubation at 4^0C for 16 hours was followed by transformation into *E. coli* strain XL1-Blue MRF' by electrotranformation. This procedure allowed the uptake of DNA by the passage of an electrical current at 2.5 KV through a chamber containing cells and DNA suspended in 50% glycerol. The procedure was performed at 4^0C. Two ul of ligation mix was added to 40 ul of electrocompetent cells and incubated on ice for one minute. The cells and DNA were transferred to a cuvette and placed in the pulser chamber (Bio-Rad).

Immediately following the five microsecond pulse the cells were diluted with 1 ml cold YENB medium (7.5 gm yeast extract, 15 gm Bactotryptone in 1 liter) and transferred to a small Falcon tube. Recovery was for 60 minutes at 37^0 C with shaking. Following recovery the cells were collected by centrifugation for 3 minutes at 3000 rpm. The medium was poured off and 200 ul fresh YENB was added along with 30 ul 20 mg/ml 5-bromo-4 chloro-3 indolyl β-D-glactosidase in dimethylformamide (XGal) and 30 ul 25 mg/ml isopropyl-thiogalactoside (IPTG). One hundred µl of the cell mixture was plated on a LB plate containing 100 µg/ml ampicillin (LB amp100). White colonies were picked to 1.5 ml LB amp100 medium in 96 deep well blocks following 18 hours incubation at 37^0C. The blocks were incubated for 18 hours at 37^0C with shaking at 300 rpm for maximum aeration. The cell pellets were collected by centrifugation at 3000 rpm for 5 minutes and stored at -20^0C.

2.18 Isolation of Subclone Template DNA

The same protocol and method was used to isolate cosmid subclone templates in a 96-well format as was used for isolation of cDNA template DNA.

2.19 Directed Phase of Cosmid Sequencing-Closure

The criteria for finishing was modified after W06E08 was sequenced and prior to sequencing W020H02 and W30B01 due to the availability of a program, primOU, modified in our laboratory by Steven Kenton, ACGT, and originally obtained from Southwest Medical Center, Dallas as the primo program. In addition, the acquisition of a MerMade (Southwest Medical Center, Avantech, Dallas) oligonucleotide synthesizer enabled the inexpensive synthesis of large numbers of single-stranded oligomers which could be used as primers in thermocycling reactions.

For W06E08 shotgun sequence data was collected from 980 reactions. This was sufficient to allow the data to be assembled into a database using Phred/Phrap (Ewing,

1998; Green, copyright 1994-1996) which required no custom synthetic primer-directed sequencing for finishing.

For cosmids W020H02 and W30B01, 768 and 576 thermocycle sequencing reactions were completed, respectively. Fifty-two 20-mer primers were designed by primOU for W02H02 and 61 for W30B01 with a melting temperature between 59-61^{0}C, and a distance from the ends of the contiguous sequences of not less than 90 bases and not more than 1000 bases for closure.

Section 3. Computer Methods for Data Analysis

2.20 EST Sequence Analysis

The collection of EST sequences presented a data type which differed from genomic sequences in several ways. Although each sequence represented an expressed gene, in most cases there were two sequences from each of the expressed genes only some of which overlapped. As additional cDNAs were sequenced, the number of contiguous sequences increased rather than decreased in contrast to a genomic sequencing project. To accommodate this difference, each EST was handled as a separate file which allowed each to be submitted to the GenBank, analyzed for homologs in the GenBank databases, and made publicly available on the ACGT website (http://www.genome.ou.edu). However, the EST sequences also were assembled to allow ESTs from the same gene to align and to generate a consensus sequence in the Unigene database. The strategy for analyzing the EST database is shown in Figure 4. Three directions were taken to yield: 1) high quality ESTs for submission to GenBank 2) a method for determining the endpoint of library sampling and 3) an assembled Unigene database reduced in sequence redundancy.

For each EST, an experiment file (Staden) was created containing the trace file (the raw sequence data) and library specific information. Automatic processing of each EST sequence was done via a series of scripts obtained from LaDeana Hillier,

Table 3. Clip and Clean EST processing scripts.

1. embellish_template	extract information from the template name, get library name from experiment file
2. reformat-scf.uwphred	reformat the trace file
3. the-big-one	call bases with ABI and phred and determines which sequences have overall poor quality (N ratio of 1:5). Makes the quality start and stop estimates based on trace quality, cut at first base <15
4. getscf_field2expfile	add the information from the trace to the experiment file
5. embellish_template_2	take the dye terminator information and the library name to extract information about the vector, adapter sequence and primer position.
6. clip-seq-vec	use vep-vector end point (Staden, 1992) to find the sequencing vector and mark those sequences which are completely sequencing vector
7. clip_left_seq_vec	repeat attempt to find the left cutoff point, using adapter sequence information and distance from primer. Tag if the poly T is not found on 3' end.
8. clip-seq-wep-left	cut the vector sequence off the left end
9. clip-seq-wep-right	cut the vector on the right end if detected, this indicates short enough insert to read through in single pass
10. check-wrong-adapter	is the wrong adapter sequence present, example if 5' adapter sequence seen and name indicates a 3' EST. A tag fails the EST.
11. blastn_vec_check	check for the vector again, trim sequence if necessary BlastN: S=133 S2=133 M=5 N=-11 W=8
12. extend_seq	can sequence be extended past the conservative initial quality estimate to the second tier of reasonable quality, are there high quality bases to the right of the first base with quality 15.
13. check_processor	checks for sequence length <100 bases, fail the short sequences.
14. screen.p	BlastN against the *E. coli* genome database, the *A. nidulans* mitochondrial sequence, *A. nidulans* ribosomal sequences. Contaminants are failed. S=170, S2=150, M=5, N=-11

15. expESTBlastx	BlastX against non-redundant protein database Matrix=blosum62, filter=seq
16. reversed	checks for reversed clones
17. check_processor_2	check if traces judged by the-big-one "overall poor quality" are worth keeping by checking similarity information with other ESTs. Use Blast information to extend the estimate of good quality sequence.
18. flip_qz_qr	bookkeeping to ensure the QR(quality right cut) always contains the right most cutoff point of the sequence, that if there is a QZ(quality extend) it is the hiqual_stop.
19.exp2dbest	create a dbEST submission file and place in directory for GenBank submission and placement on website

Table 4. Sequences used for screening mitochondrial and ribosomal sequences.

	Accession number	definition
Ribosomal sequences:	u77377	18S
	u40122	26S, partial
	u29859	28S, regions D1, D2
	u93686	5.8S
	x03564	5S, type 1
	x03567	5S, type 2
Mitochondrial sequences:		
	x00790	cytochrome oxidase subunit 1
	x01507	ATPase subunit 6, tRNA Arg, Asn
	x06960	cytochrome oxidase subunit 3, tRNA tyr
	x06961	L-rRNA, ATPase subunit 6, cytochrome oxidase subunit 3, 7 ORFs, tRNA pro, thr
	x07795	tRNA met , his
	x15441	NADH dehydrogenase subunit 3, cytochrome oxidase subunit 2
	x15442	ORF A3

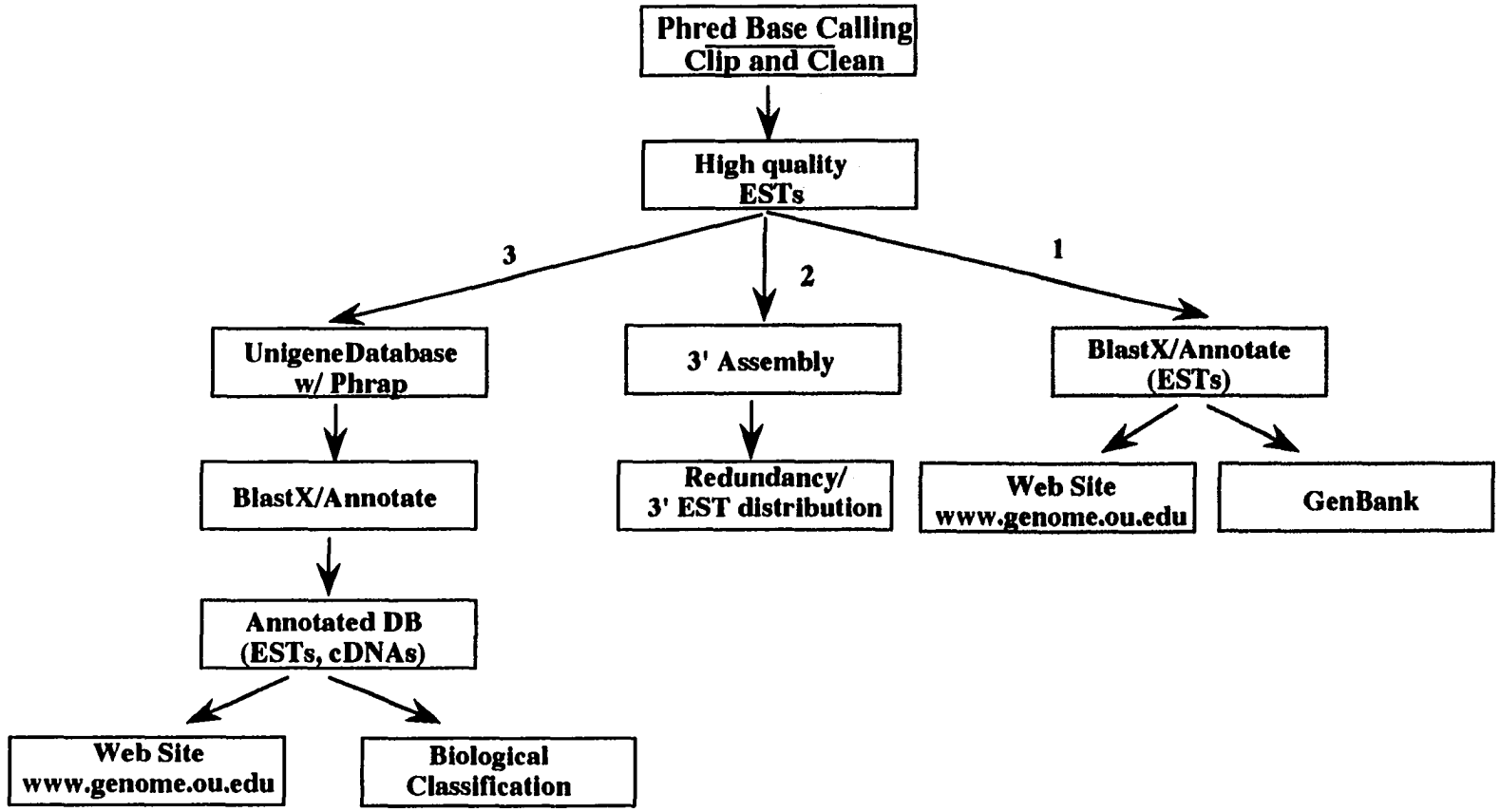


Figure 4. ESTdatabase analysis strategy. The arrows indicated the three directions for analyzing the *A. nidulans* ESTs. 1. The process of placing the high quality ESTs on the ACGT website and submitting them to Genbank. 2. 3' EST only assemblies for determination of library redundancy as sampling progressed. 3. The creation and annotation of the Unigene database. See text for detailed explanation.

Washington University, St. Louis and modified by Honshing Lai, ACGT. These scripts were linked or piped to allow passage of the EST from one script to the next. The scripts were piped in the order shown in Table 3, which also gives a brief description of the function of each script. "The-big-one" is a key script which recalls the bases using the editing program Phred (Ewing, 1998) and applies a quality value to each base called from 1-100. This quality evaluation was used in several scripts for an overall quality evaluation and endpoint determination. Table 4 lists the mitochondrial and ribosomal sequences obtained from GenBank used to screen against contaminants. The *E. coli* genome sequence was obtained from F. Blattner, University of Wisconsin prior to its publication. Figure 5 shows in a graphical format the order in which the ESTs sequences contained in the experiment files were examined by the piped scripts in the processing protocol to yield the high quality ESTs. In addition, the steps in the processing where a sequence could be removed from further consideration and moved automatically to a no-pass directory are noted.

Those ESTs which fail to pass the scripts were removed to specific no-pass directories that contained vector only, low quality, short, wrongend, ribosomal, mitochondrial, and *E. coli* sequences. Those which passed were termed the high quality ESTs.

The programs used in the Clip and Clean processing were the sequence editing program Phred (Ewing, 1998) which ranks each called base on a scale of 1-100 with a larger value indicating a higher quality. The Phred algorithm is based on peak height/width ratio and relative spacing as described earlier. BlastX (Altschul, 1997), which compares the six-frame conceptual translation products of a nucleotide query sequence against a protein sequence database and BlastN (Altschul, 1990), which compares a nucleotide query sequence against a nucleotide sequence database. The latter two programs use the Basic Local Alignment Search Tool algorithm. The parameters used with BlastN were:

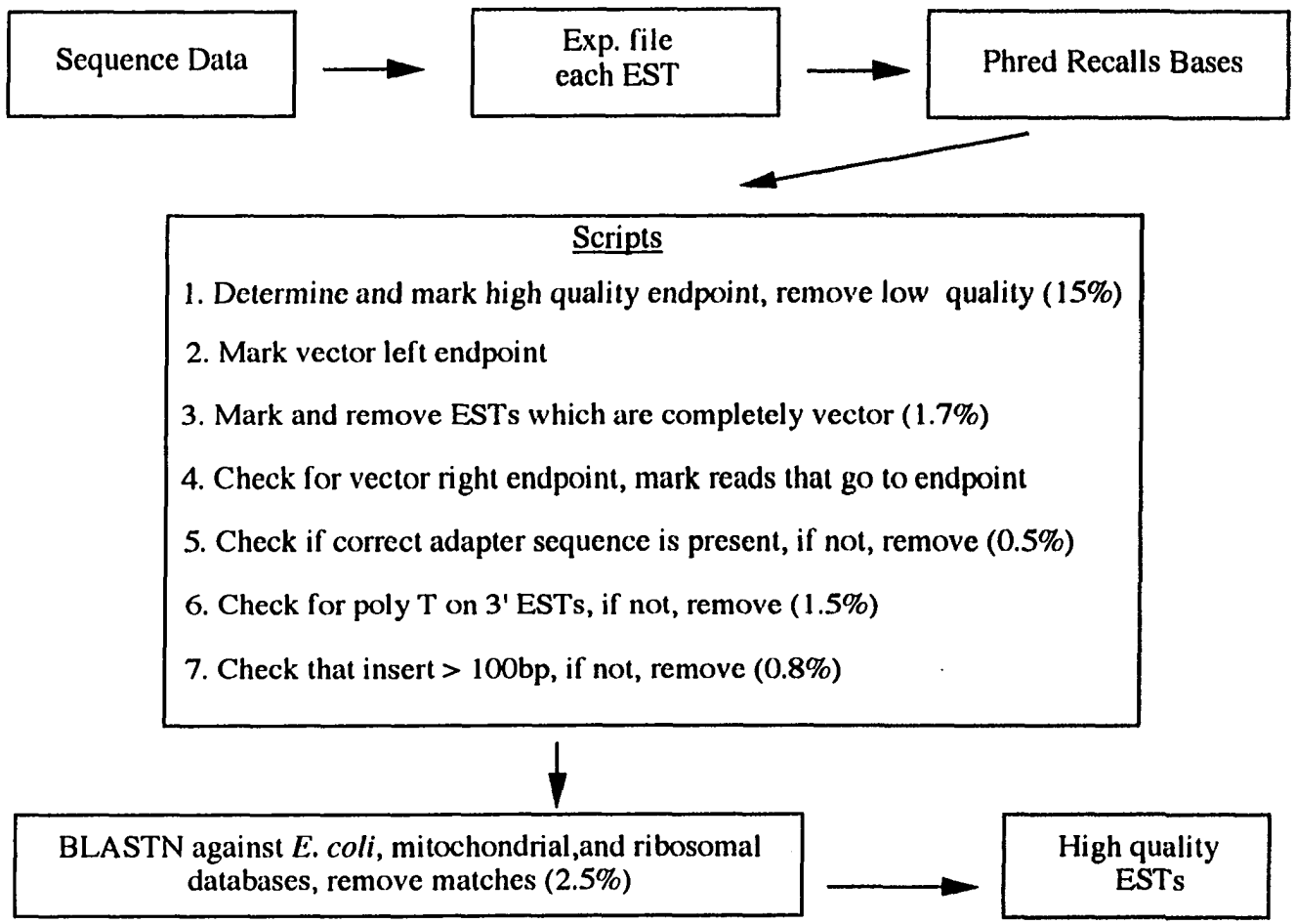


Figure 5. The "Clip and Clean" EST processing protocol. The boxes delineate steps in the processing of the EST sequence data to identify those sequences which pass the established criteria for high quality ESTs, see text.

S-limits the reported database sequence to those with at least one High Scoring Sequence Pairs (HSP) of the S setting,

S2-limits the reporting of individual HSPs to those with the S2 setting

M-score for a match

N-penalty for a mismatch

and the BlastX parameters were:

Matrix-amino acid blocks substitution matrix BLOSUM62 (Henikoff, 1991), the Blast default matrix, based on observed substitutions between segments that were less than 62% identical.

Seg filter, masked the low complexity sequences in the query sequence, default filter for BlastX

S-the cutoff for HSP score

Those sequences which were placed into the wrongend and low quality categories were examined manually. Those in the wrong end directory are checked for: 1). incorrect naming, a set of reactions identified by the wrong primer name. 2). A very short or absent poly A sequence tract with the expected 3' end primer and cloning sequence 3). a 5' reaction with the correct primer or cloning sequence not identified by the scripts. In all three instances the sequences showing these characteristics were removed from the wrong end directory and placed in the high quality database and in the first instance the EST name was also corrected. The ESTs in the low quality directory were examined to determine if reactions from a particular 96-well block of cDNA templates were of low quality. If so, the templates were reprecipitated with 7 M neutral potassium acetate as described in section 2.3 and the sequencing reaction repeated.

The high quality ESTs were submitted to the GenBank EST database (dbEST) by batch submission. Concurrently, a BlastX homology search was performed on each EST sequence. The complete report was saved and the top five homology hits with an

HSP of > 99 and a p value of $<e^{-4}$ were reported in the EST file that was placed on the ACGT website.

2.21 3' Assemblies of the EST Database

The level of redundancy in the cDNA library sampling was determined by cumulative 3' EST assemblies using the assembly program, Phrap (Green, copyright 1994-1996). The parameters of minmatch score=14, minscore=80 were empirically determined to be optimal for the assembly of the EST sequences, which were all greater than 100 bases in length, into relatively short consensus sequences (as compared to genomic) that required high stringency match. The percent redundancy was determined first by obtaining the number of individual gene alignments produced by each assembly and then by calculating the percent which represented new sequences. The inverse of this percentage gave the percent of redundant sequences seen. A best fit curve was generated from plotting the number of 3' ESTs vs the percent redundant sequences using CA-Cricket Graph III (Computer Associates Inc.). A report showing these results is presented in the Results and Discussion section.

2.22 3' and 5' Assemblies of the EST Database

Both the 3' and 5' EST sequences were assembled with the same minmatch and minscore as described above. This yielded a Unigene database assembly that was saved in two directories. Those ESTs which did not assemble with any others were considered singlets and stored in a singlets_dir; those that assembled with other EST sequences were considered members of a cluster, ESTs representing the expressed gene. The consensus sequences derived from the clustered ESTs were given "contig" identification names and were stored in a contigs_dir. The singlet and consensus cluster sequences (the contigs) were submitted for a BlastX search using the default Blosum 62 matrix, seg filter, High Scoring Sequence Pair (HSP) >99 and p value of $<e^{-4}$. Following the

BlastX search each of the two directories contained three files for every Unigene member. One file contained the sequence of the EST or consensus sequence in fasta format, a second file contained the complete BlastX output, and the third file contained a table with the BlastX header lines only.

2.23 Biological Function Outline

The analysis of the *A. nidulans* Unigene database entailed the organizing of the singlet and cluster members into a functionally related schema based on that developed by Monica Riley for *E. coli* (Riley, 1997)and improved by Selkov, et al.(Selkov Jr., 1998). This schema resulted in an *Aspergillus nidulans* Categories of Cellular Function as outlined in Appendix II. A BlastX search was performed on all members of the Unigene database, and the results were determined to be significant if they had a HSP >99 and ρ value<e-4 . The process of developing the Biological Function Classification for the Unigene database is shown graphically in Figure 6. Initially, a preliminary keyword list was developed, termed Keyword list 1, using a variety of sources that included the major metabolic pathway enzymes from biochemical texts (Bohinski, 1987; Stryer, 1995), pathway enzymes from the Metabolic Pathways Database (Selkov Jr., 1998) as well as reports from GenBank (Benson, 1996) Medline and Entrez (Schuler, 1996). The list was used with a series of scripts designed by James White, ACGT, to automatically scan the Unigene files containing the BlastX output descriptor lines for matching keywords to create an output file termed Keyword.hits that was organized into contig numerical order displaying the match. Following manual editing to examine for the multiple presence of a contig in the output file when a BlastX output which matched more than one keyword, the keyword list was corrected to yield Keyword list 2 and three output files were created: Keyword.hits 2, a list of contigs and BlastX descriptor lines which had matching keywords; Keyword.nohit, a list of those contigs which were not selected by a keyword, and Keyword.better, an exception file containing contigs with keyword matches but with

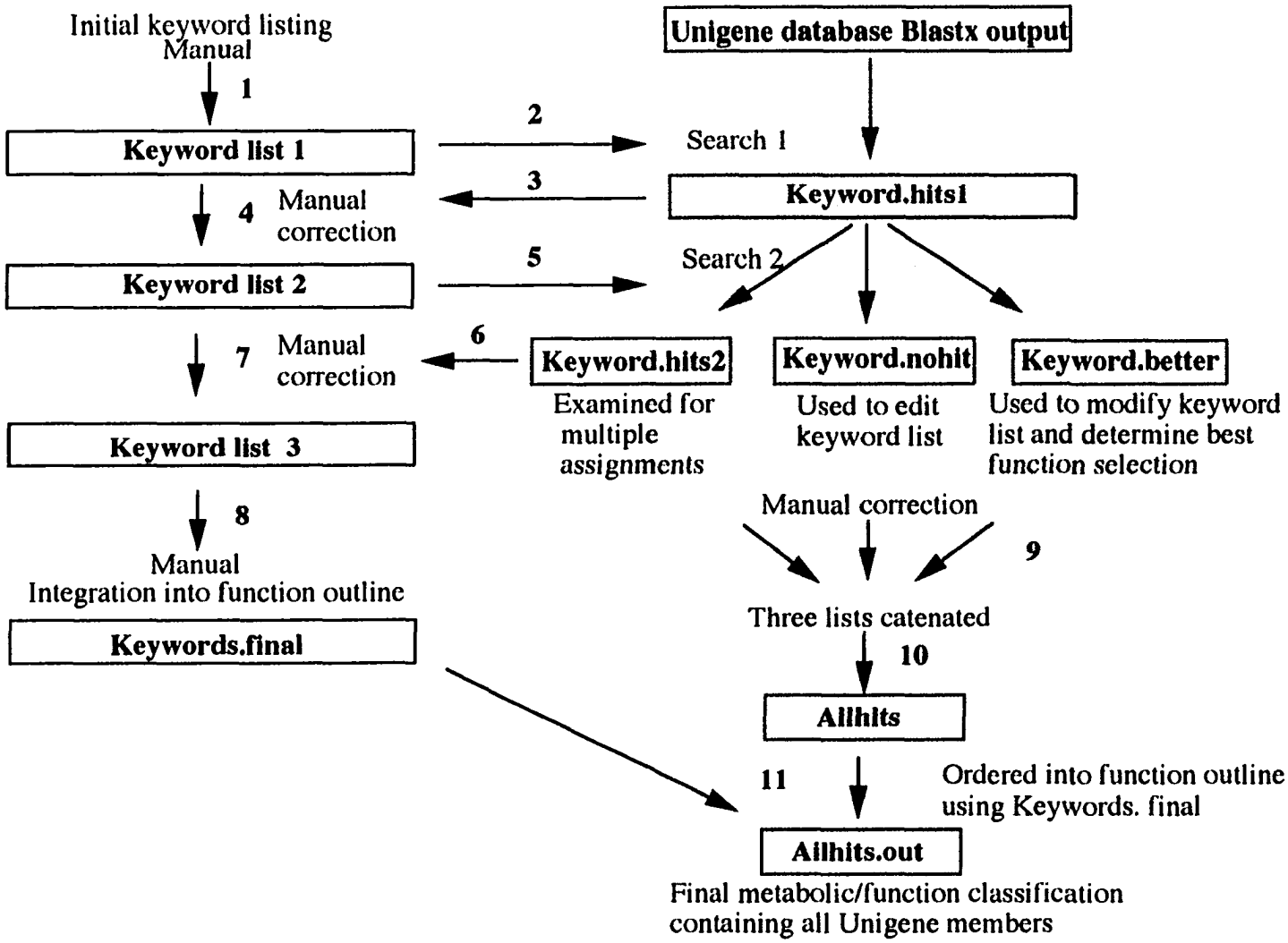


Figure 6. Steps used in creating the *Aspergillus nidulans* Biological Function Outline. Manual steps are noted. Other steps were done automatically with Perl scripts. In bold are the files created in the steps.

significant hits with better HSP and ρ-values for homologies not included in the keyword list. The Keyword.nohit and Keyword.better were examined manually for either addition or modification of the keyword list, yielding Keyword list 3. The Keyword list 3 was manually integrated into the Categories of Cellular Functions outline creating Keywords.final (Appendix I) containing 1375 keywords, 1041 primary keywords, 478 of which contain comments and 334 alternate keywords which allowed combining BlastX headers containing different spellings or terms for the same gene, protein or enzyme. The three files, Keyword.hits2, Keyword.nohit and Keyword.better were manually edited and then catenated into one file, Allhits. This file was automatically ordered into Keywords.final to yield Allhits.out which contained the Biological Function Classification including both the Contigs and singlet members of the Unigene database. The output from this is reported in Results and Discussion and listed in Appendix II.

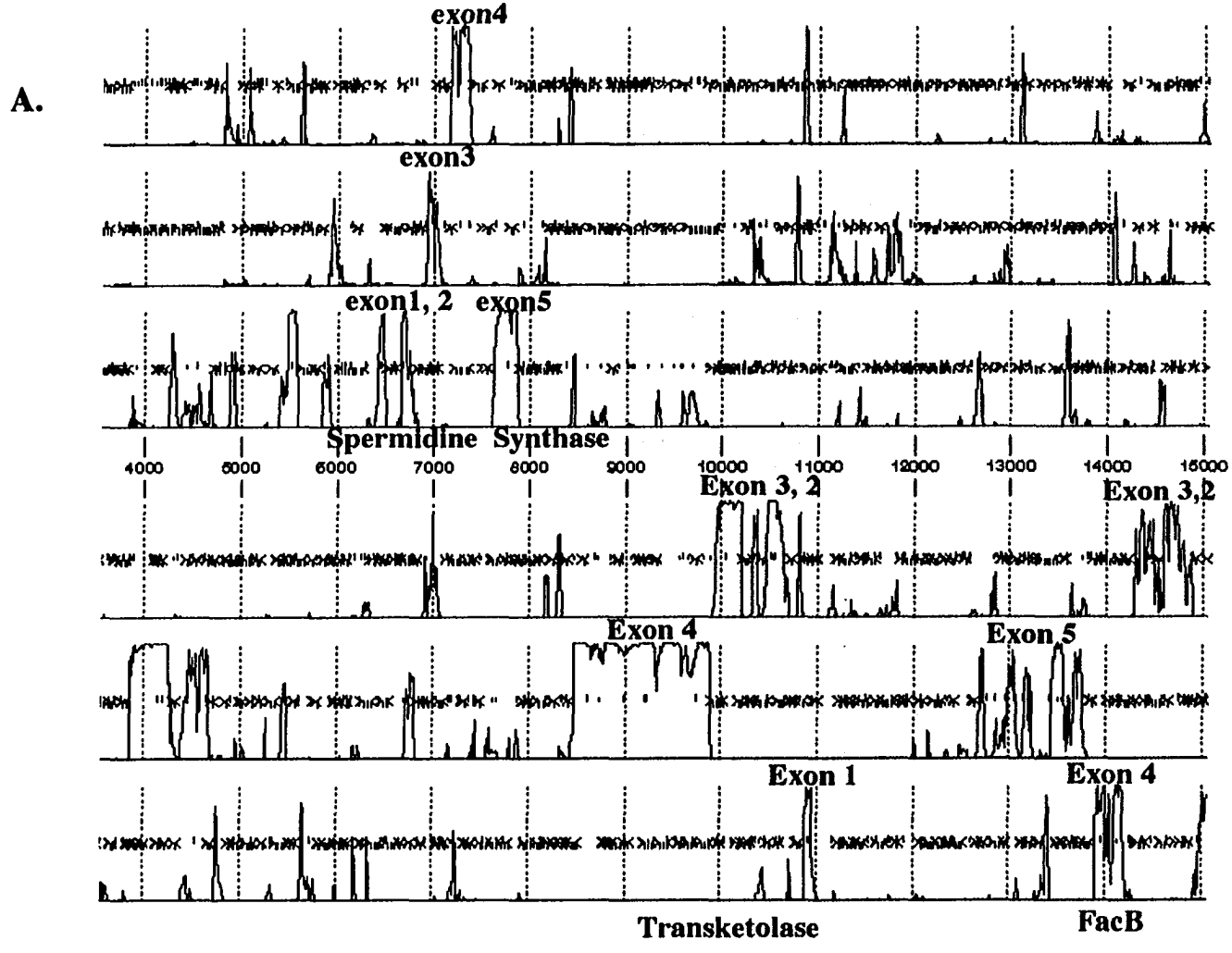
Assignment of function was done using the homolog with the best HSP and ρ values, i.e. the best hit, as the source of gene assignment. The assignments are based solely on sequence homology as no biological testing was performed and therefore are "predicted" until biologically confirmed. In some cases, the best hit was a hypothetical protein or ORF while other hits in the blast report were to genes with known functions. Each instance of this was examined in the Keyword.better file. If a less significant homolog showed sequence similarity over the same region as the best hit and had a HSP value of >99 and p value < 10^{-5} then it was used to assign a function to the Unigene member. A notation (*) was placed next to the contig name to designate when this occurred (Appendix II).

2.24 Annotation of *Aspergillus nidulans* Chromsome IV Cosmids

The sequence of cosmids W30B01 and W02H02 was completed as part of this dissertation research. Because the cosmids had a large overlap the data from both projects was combined to form a contiguous sequence for this genomic region for further

analysis (Results and Discussion).

Gene detection typically is more accurate if both intrinsic and extrinsic approaches are employed. The GeneMark program allowed an intrinsic approach to gene prediction which included evaluating particular properties of the DNA sequence without explicit referral to other sequences (Borodovsky, 1994). This approach required a training set based on *A. nidulans* non-redundant protein database entries to produce a matrix, E.nidulans_3.mat, that was kindly provided by Dr. M. Borodovsky. GeneMark produced a set of predicted exons that then were submitted for a Powerblast (Zhang, 1977) search using both BlastN and BlastX algorithms of the GenBank databases. The output from both the GenMark analysis and the Powerblast search were visualized by Diana, a program developed at the Sanger Centre, Hinxton, UK and modified by James White, ACGT. This program allows examination of a DNA sequence when provided with files containing intrinsic data, such as that provided by GeneMark and extrinsic data, such as that obtained from BlastX results. Figure 7, part A shows the six frame output from GeneMark for a region containing the exons for three predicted genes, spermidine synthase, transketolase and acetate regulatory factor (FacB). The peaks indicate regions of coding potential as defined by GeneMark, the larger ticks indicate potential methionine ATG translation start codons while the smaller ticks indicate the alternative GTG codon which can serve as an alternative start codon. The lower ticks indicate one of the three stop codons TAA, TAG, TGA, and < or > mark potential intron/exon borders. Figure 7, part B shows the graphic output from Diana after integration of both the BlastX information and the GeneMark exon data for the same region as an example of the data used for annotation of the cosmid sequences reported in Results and Discussion.


A.
exon4
exon3
exon1, 2
exon5
Spermidine Synthase
4000
5000
6000
7000
8000
9000
10000
11000
12000
13000
14000
15000
Exon 3, 2
Exon 3,2
Exon 4
Exon 5
Exon 1
Exon 4
Transketolase
FacB

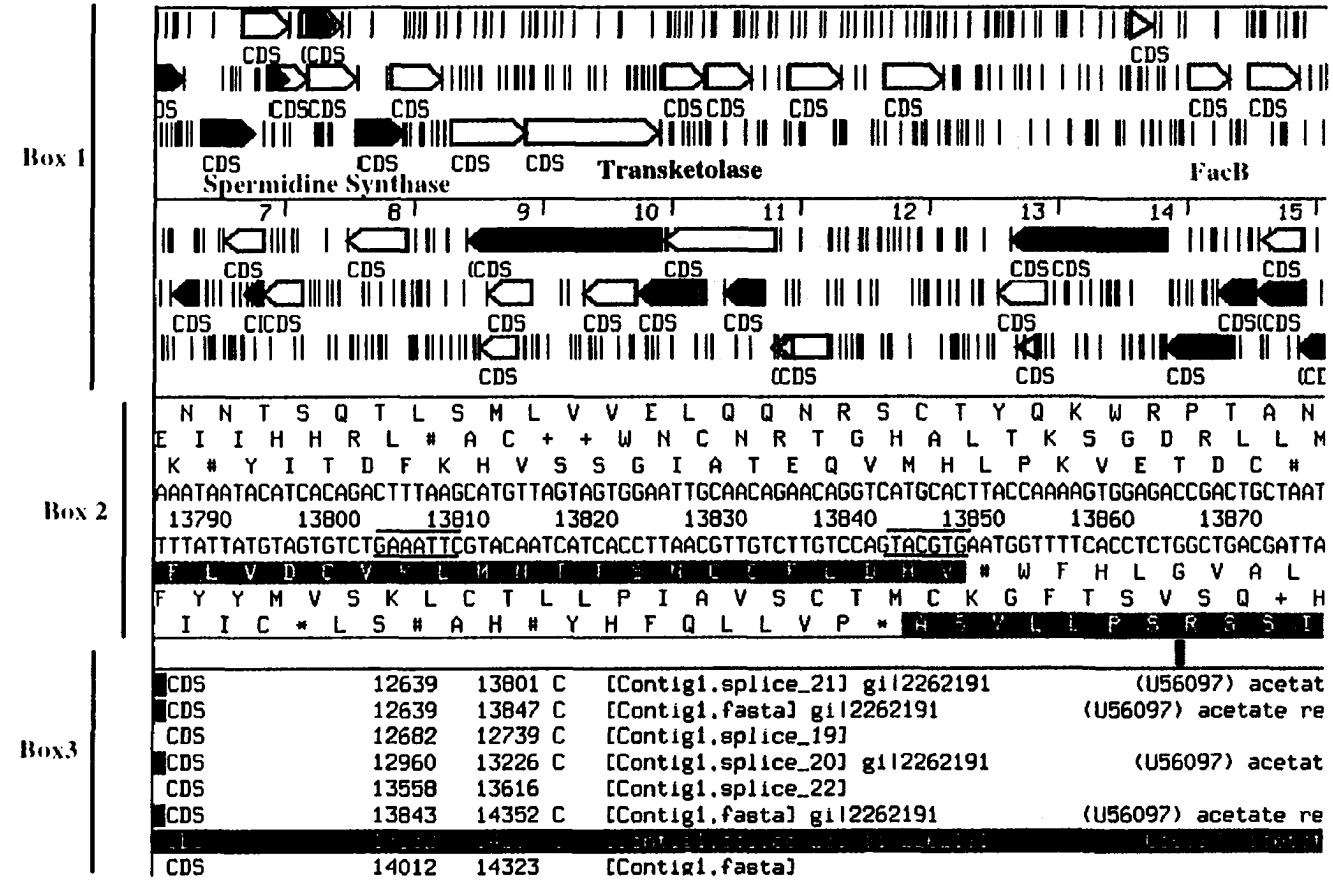


Figure 7. A comparison of the output from Genemark, in six frames showing regions covering three genes. The exons are labeled. An output for the same region in Diana showing the exons (Box1), the nucleotide sequence and the six frame translation (Box2) and the incorporation of information from Blastx (Box3) of a region of cosmid W30B01. Peaks in GeneMark represent ORFs with coding potential determined by the *A. nidulans* specific matrix used. For Diana the blue boxes represent exons called by GeneMark, red are the ORFs called by Diana. In Box2 the highlighted protein sequences represent the ends of FacB exon 4 and 5, the underlined sequences are intron borders.

Chapter III

Results and Discussion

3.1 Introduction

The goals of the EST project were to obtain 3' and 5' end sequences of a majority of the expressed genes represented in the *Aspergillus nidulans* vegetative and asexual cDNA library, to determine their relative expression level and their probable identity, to establish a metabolic overview based on the probable identity of the expressed genes based on the EST sequences, and to demonstrate the usefulness of the EST sequences in annotation of genomic sequences.

3.2 EST database overview

The *A. nidulans* EST database was created by obtaining the sequence of both the 3' and 5' ends of isolated cDNA clones. Each EST trace file was assessed for quality by a series of automated, sequential scripts. A file, containing the sequence and all additional information pertaining to that EST was created for each sequence. After the sequences were examined for homology to the non-redundant protein database at the NCBI, this information was added to the EST file. The separate files then were combined into a single directory on the Unix-based Sun computer to form the EST database. The database was made publicly available by placing it on the Advanced Center for Genome Research (ACGT) website (http://www.genome.ou.edu) and depositing it into the GenBank EST database (dbEST).

A double-stranded cDNA does not necessarily represent a complete copy of the mRNA from which it was synthesized. Variable 5' ends can result if the first strand synthesis reaction does not result in a copy of the entire mRNA. In contrast, the 3' end which contains the 3' untranslated region and 3' end of the coding region, is anchored in a fixed position at the poly A tail since a poly dT primer was used to begin the first strand synthesis. Thus, because the 3' end sequences, which represent a fixed region of the

cDNA and include the 3' untranslated region are gene specific sequences, they were used to determine the complexity of the cDNA library and also the sequence redundancy at various points in sampling the library. The homologous 3' ESTs were aligned into groups where each group represents a cDNA family, i. e. a set of clones transcribed from the same gene. The number of ESTs in each group was an estimate of the abundance of a gene's transcript present in the library. In addition, the probability of detecting a new gene was estimated by periodically examining the redundancy in the 3' assemblies. A plotting of this probability helped determine when the productive endpoint of library sampling was reached.

Because the 5' ESTs were not anchored by the poly A tail as were the 3' ESTs and therefore were from variable positions of the original mRNA, they either could represent the 5' untranslated region and/or the coding region depending on the efficiency of the first strand synthesis and the size of the cDNA insert. The sequence of an entire cDNA could be obtained if the assembly of 5' and 3' ESTs resulted in a consensus cDNA sequence from an abundant cDNA and the 5' ends of the aligned EST sequences were sufficiently staggered. Since one goal of this study was to determine the abundance and distribution of EST species, library subtraction to eliminate redundant copies was not attempted. Thus, these results yield an estimate of the relative level of transcription of the corresponding genes at the time of mRNA harvest. In addition, the ESTs and the resulting consensus sequences provided a means for gene identification which facilitated annotation of genomic sequences because the ESTs often reveal the correct translation start site and intron-exon borders. This was demonstrated by the annotation of three cosmids from *A. nidulans* Chromosome VIII sequenced as part of this dissertation research and a gene cluster region of Chromsome VI.

Several conclusions could be drawn from analysis of EST databases. First, the abundance of different EST species reflected the level of representation of that cDNA in the library. Second, alignment of ESTs into contiguous sequences was extremely useful

in predicting genes in genomic sequences. Third, determining the sequence homologs in the public databases not only classified the cDNA represented by the EST but allowed the organization of the data into a biological function-based scheme such as that designed by Monica Riley for *E. coli* (Riley, 1997). This approach has been used as an organizational scheme for the genes identified during the genomic sequencing of *Haemophilus influenzae* (Fleischmann, 1995), *Methanococcus jannaschii* (Bult, 1996), and *Mycoplasma genitalium* (Bult, 1996; Fraser, 1995) and for the *Saccharomyces cerevisiae* genome (http//:wit.mcs.anl.gov/WIT2). Classification of the expressed cDNA therefore presents a snapshot of the genes expressed at the time of harvest and serves as an index of the *A. nidulans* expressed genes.

3.3 EST Database Quality Summary.

Table 5a and b presents the EST sequencing results. Table 5a lists the number of 3' and 5' reactions which passed the Clip and Clean processing scripts called the high quality ESTs, as well as the no-pass sequences, those which did not (see Clip and Clean in Materials and Methods). Table 5b breaks down the no-pass ESTs into the specific categories in which each did not pass the processing criteria. The success rate for the 3' sequences was lower than that for the 5' sequences since the presence of the Poly A tract on the 3' end of a cDNA resulted in a number of lower quality sequence reads. This is because enzyme slippage occurs in the poly A homopolymer regions and the resulting shift of one or more bases in the sequence beyond the homopolymer region caused the trace to appear to have multiple overlapping sequences (PerkinElmer, 1995). The difference in success rate between 5' and 3' sequences also has been noted previously by Hillier et al.(Hillier, 1996) where their average success rate was 76% for 5' and 63.5 % for 3' cDNA sequences from 20 human tissue libraries. As can be seen in Table 5a, the present study had an efficiency of 89% and 78% for 5' and 3' end sequences, respectively. The resulting *A. nidulans* EST database represents 8645 cDNA clones.

Both the 3' and the 5' ESTs are present in the high quality database for 3847 cDNAs, only 3' ESTs are present for 1987 cDNAs and only the 5' ESTs are present for 2811 cDNAs. Sequences falling into the overall low quality no-pass category were examined manually and the 5' sequencing reactions were repeated (Materials and Methods). The continued failure of a subset of these reactions may be due to GC or AT rich sequence regions through which the 1:16 BigDye ampliTaq FS reaction chemistry used for this work was unable to sequence.

Table 5. *Aspergillus nidulans* 24 hour vegetative/asexual cDNA library data summary. Final release 12-15-98

A.	Total Number of ESTs		12,482
	Number of megabases sequenced		4.268
		5'	3'
	High quality	6658	5824
	No pass	792	1611
	Total	7450	7435
	Success	89%	78%
B.	Clip/Clean no pass:		
	Completely vector	69	109
	Too short (<100b)	92	313
	Wrong end	33	240
	Overall low qual.	450	880
	E. coli	141	68
	Mitochondrial	5	-
	Ribosomal	2	1

3.4 Submission of EST Data

The Clip and Clean processing created two files, a homology file which was placed on the ACGT publicly accessible web site and a GenBank entry file, which lacked the homology information. Figure 8 shows an example of the output created for each EST sequence. The homology information includes the ten highest homology sequence

```
TYPE: EST
STATUS: New
CONT_NAME: Bruce Roe, University of Oklahoma, broe@ou.edu
CITATION:
LIBRARY: Aspergillus nidulans 24hr mixed vegetative and
          developmental cDNA
EST#:   c9a04a1.r1
CLONE:  Aspergillus nidulans c9a04a1
SOURCE:  Rodolfo Aramayo, Texas A&M University, raramayo@bio.tamu.edu
P_END: 5'
SEQ_PRIMER: T3
HIQUAL_START: 1
HIQUAL_STOP: 225
DNA_TYPE: cDNA
PUBLIC: unpublished
COMMENT: This clone is available from the Fungal Genetics Stock Center;
         contact the curator, Dr. Kevin McCluskey
         (fgsc@kuhub.cc.ukans.edu), for further information.
HOMOLOGY:
gnl|d1007414 (D32070) heat-shock protein 30 (HSP30..+1 262  5.5e-22   1
gi|168820    (M55672) heat shock protein 30 [Neuro..+1 212  2.1e-16   1
SEQUENCE:
ATTAGACGCTTCAAAACAAAGCATCCTCGCAATCCACAAGGCATCAATTTCCTCGACATT
CTCATAAACAAACCCACGTCACAGGTACAATGGCTTTCTTCCCCCGCTACTGCTCAGGCG
ACTTCGCCCCTTTGTTTCAGCTCCTCGACGACTACGATATGCACCAGGCCACCCGCCGAC
CAAACAAGAAGGTCACCAACGTGAGAACATTTGTTCCTAAATTTGACGTCTACG
||
```

Figure 8. An EST homology file. This format without the homology information is used to prepared a GenBank entry file for submission to the Genbank dbEST.

matches found by a BlastX search of the NCBI nonredundant protein database. To date five batch submissions have been made to the NCBI EST database (dbEST) where they have been assigned accession numbers (Table 6). The April 1999 release from the NCBI dbEST listed the *A. nidulans* ESTs as the ninth largest collection of organism-specific ESTs.

Table 6. EST Accession Numbers and Deposit Dates into the GenBank dbEST

Accession Numbers	Submission Date
AA783056-AA788459 AA788523-AA788570	February 1998
AA965289-AA966707 AA966906-AA966919	May 1998
AI007487-AI007508	June 1998
AI209337-AI214027	October 1998
AI327546-AI328148	December 1998

3.5 Assessing Library Redundancy.

Since the *A. nidulans* library was oligo (dT) primed from the mRNA template's poly (A) tails, the 3' EST sequence provides the 3' end of a gene. Assembly of the 3' ESTs into homologous clusters revealed the complexity of the library as well as the number of new genes detected while library sampling was in progress. These 3' assemblies were based on the presence of the 3' untranslated region which was assumed to be gene specific. Those sequences which aligned during a Phrap assembly were considered to be transcripts from the same gene (Figure 9) and therefore grouped into a cDNA cluster.

The end sequencing of the cDNA library is an example of sampling with replacement. This differs from the sequence data obtained for a cloned genomic region where each position of the sequence is represented in the randomly generated subclones and the distribution of sequences can be mathematically expressed by a Poisson

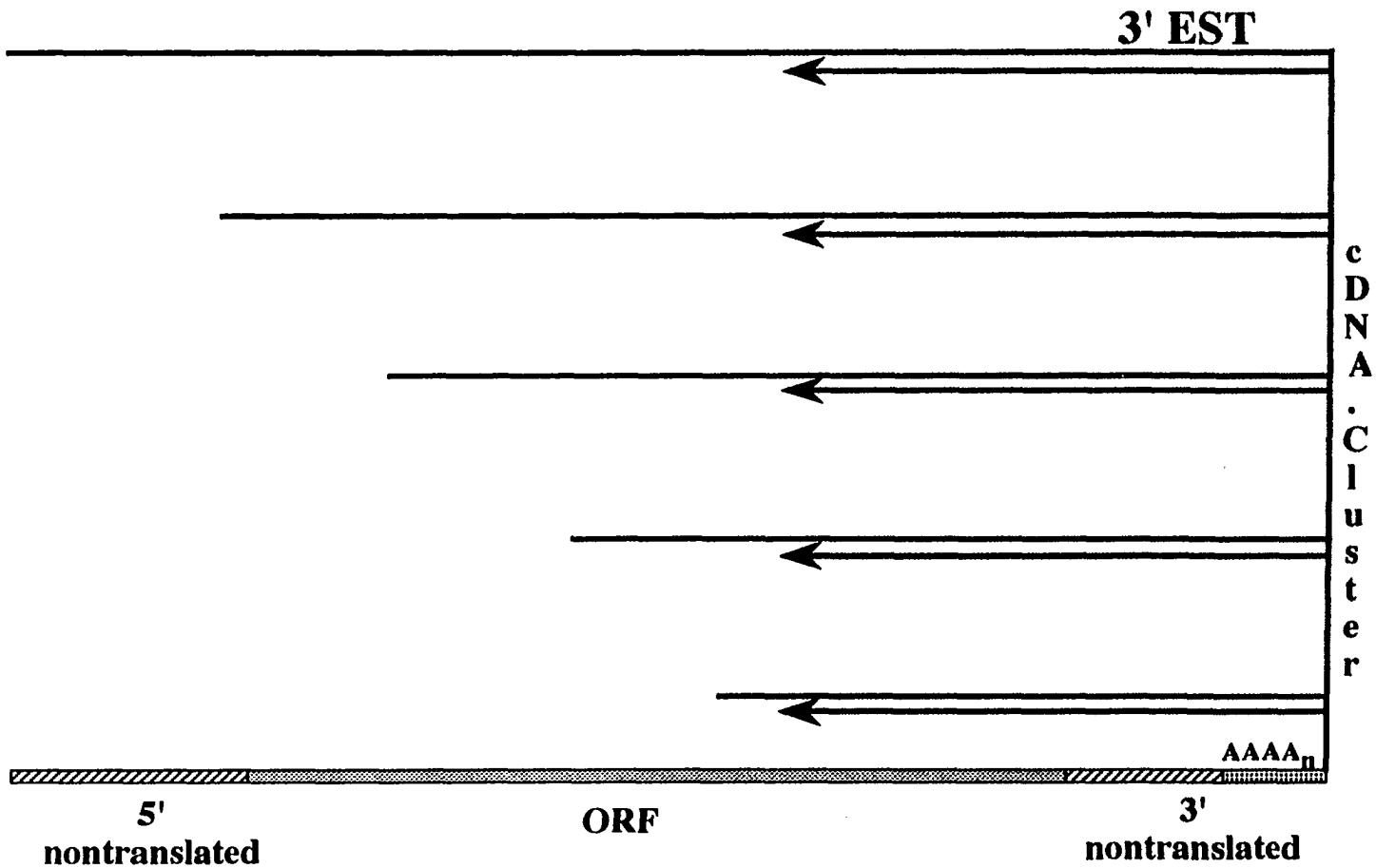


Figure 9. 3' EST Assembly of ESTs into a cDNA cluster. The lines represent the relative length of a cDNA to the entire length of the complete mRNA (Bar). The arrows represent the 3' ESTs sequenced from the cDNA templates.

distribution as described by Lander and Waterman (Fleischmann, 1995; Lander, 1988). Here, the calculations yield the theoretical number of sequences required for a certain fold coverage of a target sequence and the probability of completely sequencing the target region with that fold coverage. However, a cDNA library represents only transcripts and hence only the genes. Populations will be differently represented depending on the number and distribution of the mRNAs at the time of harvest, which in turn is dependent on the growth conditions. Finally, the probability of selecting and end sequencing representative cDNAs must be considered.

Table 7 lists the number of singlets, i.e. those ESTs which were not grouped with others and therefore represent nonredundant sequences, and clusters, i.e. those ESTs which were aligned or "clustered" by Phrap and represent transcripts from the same gene. Cumulative assemblies were performed at intervals once 200 new 3' EST sequences were obtained. The number of genes represented at each assembly point is given in Table 7 where the percent redundant sequence is simply the inverse of the percent of new genes found. A plot of the total number of 3' ESTs against the percent of redundant sequences is shown in Figure 10. As was expected, the percent of resampled sequences initially was low since most of the EST sequences represented newly sampled genes. As the sample size increased, the number of redundant sequences also increased and many of the new EST sequences could be clustered. This suggested that as expected, a population of redundant sequences was present in the library. Extrapolation of the data showed that a plateau occurred when an approximately 70-75% redundancy level was reached. The experimental evidence was consistent with the theoretical calculation since with 5824 3' EST sequences the redundancy approaches 70%. Since new genes were still being detected, but at a fairly low level, at least several thousand additional ESTs would have to be sampled to exhaust the library and find the rare cDNAs (Figure 10). It was not the purpose of these present studies to discover all the expressed genes in *A. nidulans*, but rather to determine the identity of the genes expressed in the

Table 7. Determination of the percent new genes by cumulative 3' EST assemblies of the *Aspergillus nidulans* database.

$$\% \text{ New Genes} = \frac{(\text{New S} + \text{C}) - (\text{Old S} + \text{C}) \times 100}{\text{New Reads}} = \frac{(300) - (165) \times 100}{200} = 67.5$$

Total Reads	Singlets (S)	Clusters (C)	S + C =# genes	% New Genes	% Redundant
200	144	21	165	82.5	17.5
400	249	51	300	67.5	32.5
600	354	72	426	63.0	37
800	443	95	538	56.0	44
1000	526	120	646	54.0	46
1200	586	168	754	54.0	46
1400	630	214	844	45.0	55
1600	696	243	939	47.5	52.5
1800	755	272	1027	45.5	54.5
2000	799	311	1110	41.5	58.5
2200	857	336	1193	41.5	58.5
2400	920	362	1282	44.5	55.5
2600	971	398	1369	43.5	56.5
2800	1005	437	1442	36.5	63.5
3000	1053	474	1527	42.5	57.5
3200	1106	503	1609	41.0	59
3400	1131	552	1683	37.0	63
3600	1106	610	1716	16.5	69
3800	1161	630	1791	37.5	62.5
4000	1205	661	1866	37.5	62.5
4200	1229	695	1924	29.0	71
4400	1272	725	1997	36.5	63.5
4600	1316	747	2063	33.0	67
4800	1325	788	2113	25.0	75
5000	1376	815	2191	39.0	61
5200	1403	839	2242	25.5	74.5
5400	1429	867	2296	30.5	69.5
5600	1469	888	2357	27.0	73
5800	1507	911	2418	30.5	69.5

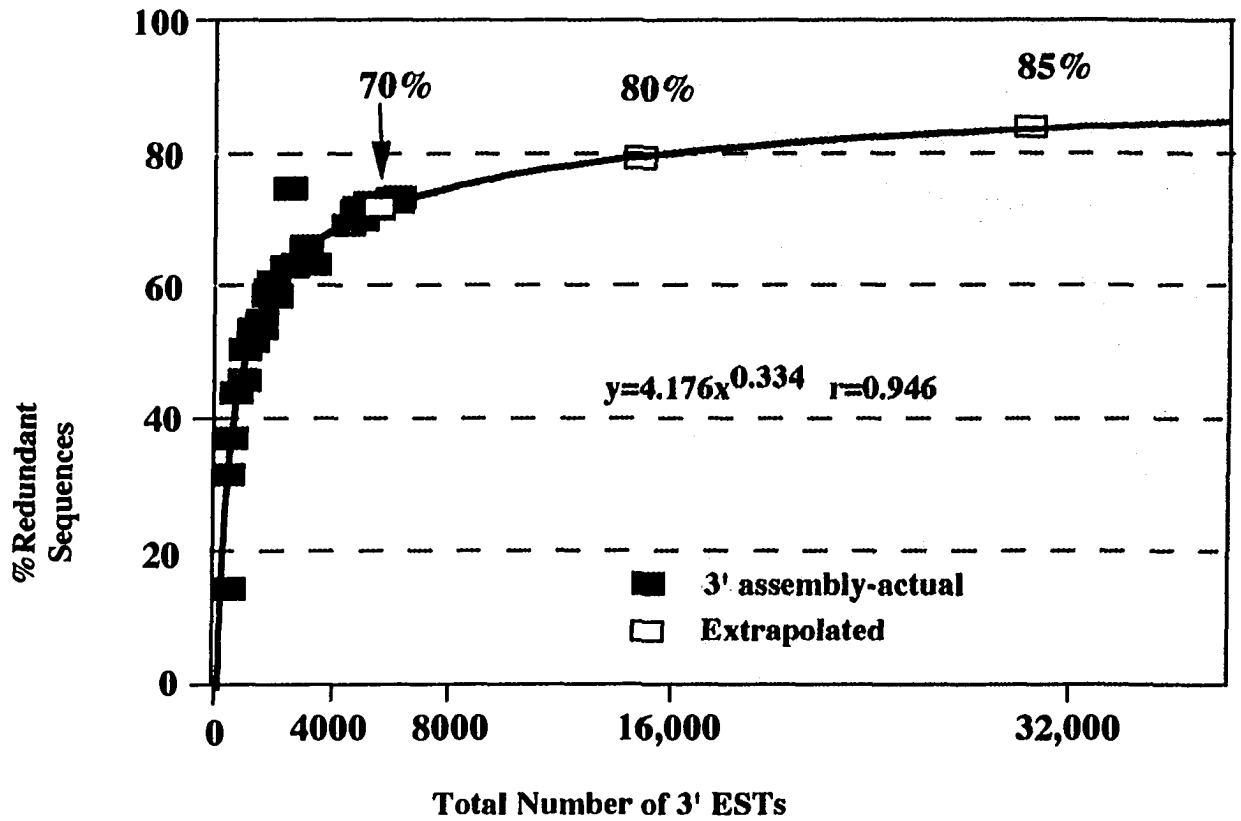


Figure 10. The percent of redundant sequences determined by cummulative 3' EST assemblies. The gray boxes represent the percent redundant at each assembly as determined in Table 8. the white boxes are extrapolated points determined by the formula given which was generated from the actual data points.

vegetative/asexual cDNA library. This library remains a good source of new gene information but the new information only can be obtained by acquiring a high number of redundant sequences or by switching to more efficient subtracted libraries. In addition, this redundancy would be expensive, and yield few new genes if sampling extended beyond the 70% redundancy level. As a comparison, in the libraries examined at Washington University Genome Sequencing Center, sampling also was discontinued when the level of redundancy reached 70% (Hillier, 1996).

Although beyond the scope of this present research, new, previously unidentified genes in the library could be detected by subtractive hybridization of the highly redundant members of the cDNA library (Adams, 1995) or by library normalization to reduce the frequency of the highly expressed populations (Bonaldo, 1996). Both of these techniques would allow a deeper examination of the library but would not yield the relative expression information which was a goal of the present research.

3.6 The Unigene Database.

Figure 11 shows the same set of theoretical cDNAs that represent the cDNA cluster shown in Figure 9 but with both the 3' and 5' ESTs aligned. The 5' ends terminate at variable positions depending on the success of the first strand synthesis so that staggered 5' ends often are present when the homologous EST sequences are assembled with Phrap. Such a multiple sequence alignment assembly yields a consensus sequence that is more informative than the individual EST sequences in two ways. First, the redundancy often yields a consensus sequence with multiple coverage that because of multiple sequence reads results in a more accurate sequence, and second, the consensus sequence could be extended in the 5' direction to yield a larger contiguous sequence than that obtained with the individual staggered 5' end sequences.

In this present study, the database assembled as described above and in Materials and Methods contained 1866 clusters with from two to 363 ESTs and 2429 single ESTs

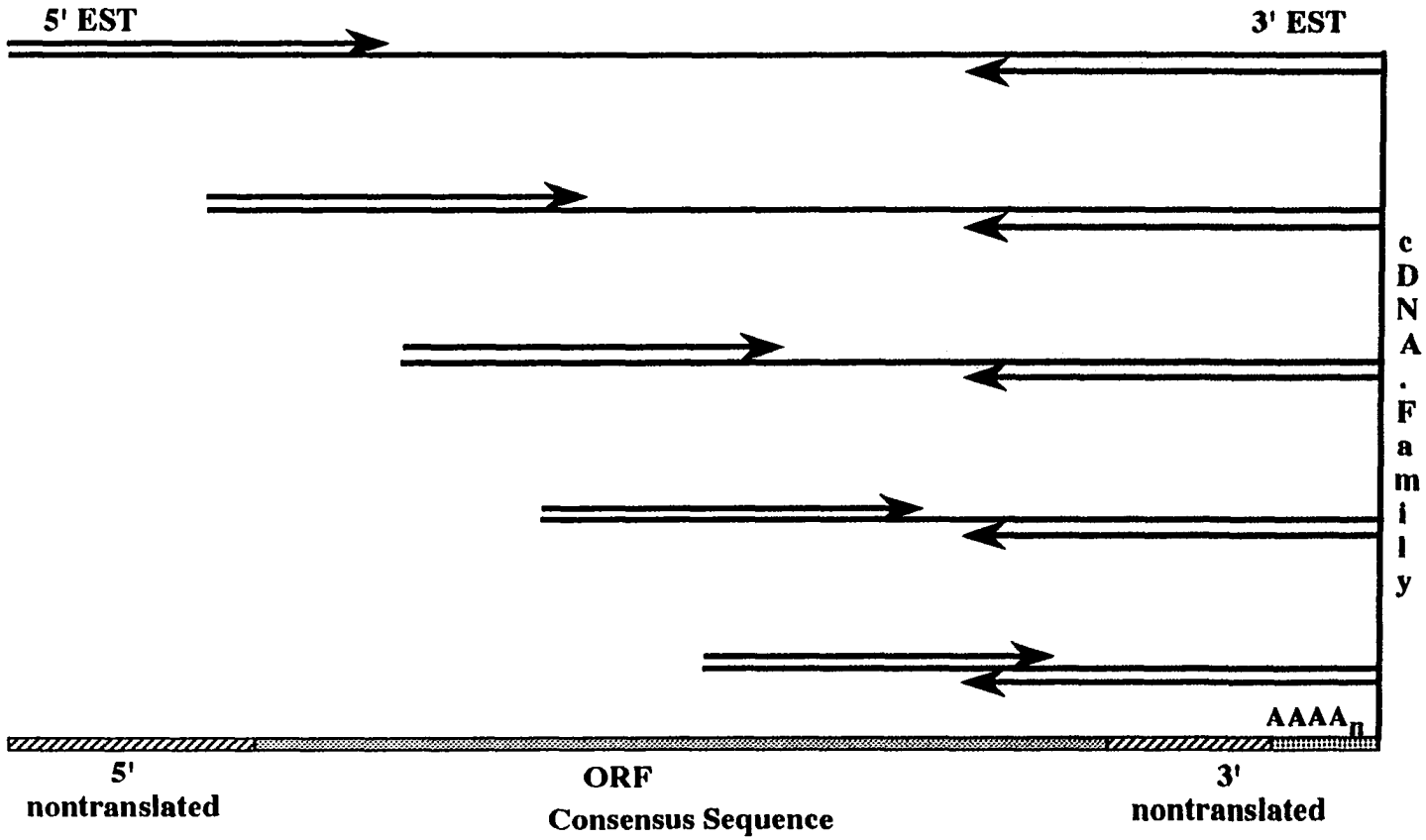


Figure11. cDNA consensus construction by assembly into a unigene database containing both 3' and 5' ESTs. This figure is related to figure 9. The arrows represent the pairs of EST sequences generated for each cDNA. The lines represent the 3' anchored cDNAs and their relative coverage of the complete transcript. The bar represents the consensus sequence resulting from assembly of the ESTs.

resulting in a total of 4295 unique members (Table 8). This Phrap assembled data results in a so called Unigene database, a term borrowed from the grouped human ESTs found at the NCBI (Boguski, 1995). The final *A. nidulans* Unigene database had 1219 clusters (65%) that contain both 5' and 3' ESTs and therefore a consensus sequence containing at least a portion of the 5' end of a cDNA sequence, although it should be noted that this is not necessarily the complete 5' end of the gene. There also were 647 consensus sequences which were not assembled from both 5' and 3' sequences, 427 (23%) consensus sequences assembled from only 3' reads and 220 (12%) consensus sequences assembled from only 5' reads. Note that the higher number of 3' only EST sequences was expected since they have a common end, contiguous with the poly A tail, while the 5' ends were staggered and not necessarily overlapping.

Previous studies have determined common characteristics of mRNA and EST databases (Bishop, 1974 ; Soares, 1994). Reassociation kinetics analysis studies of mammalian somatic cell mRNA indicated that the population was distributed into three frequency classes, a very abundant class that consists of 10-15 mRNA species which represent 10-20% of the total mRNAs, an intermediate abundant class with 1000-2000 mRNAs that represent approximately 50% of the total mRNAs and a low abundance class containing the remaining 30-40% of the mRNAs (Bishop, 1974). The percent redundancy in an EST database therefore is expected to become greater than 60% once most of the mRNAs from the first two abundance classes are identified (Soares, 1994).

The *A. nidulans* Unigene database also has the three abundance classes suggested by Bishop (Bishop, 1974) (Table 9). The very abundant class was a group of 15 EST clusters, each of which consisted of greater than 0.5% of the ESTs in the Unigene database . The abundant class was made up of clusters of greater than two ESTs and represented genes sampled more than once. The contiguous sequences containing one (singlets) or two ESTs (representing a clone pair), made up the low abundance class and represented genes which were sampled only once.

Table 8. Assessing library complexity using assembly of both 3' and 5' EST sequences with Phrap version 98. The three frequency classes total 4295 unique Unigene members. Cluster size gives the number of ESTs belonging to the cluster. Horizontal lines delineate the three frequency classes.

Cluster Size	Frequency of Cluster		Cluster Size	Frequency of Cluster	
1	2429	Low	33	3	Abund-
2	855		34	2	ant
3	316	Abund-	36	1	
4	176	ant	37	1	
5	110		39	2	
6	55		40	2	
7	47		41	1	
8	28		43	1	
9	46		45	2	
10	21		47	1	
11	20		51	1	
12	14		52	4	
13	13		53	2	
14	4		54	1	
15	9		55	1	
16	12		57	1	
17	3		63	1	High
18	4		64	1	
19	5		65	2	
20	4		71	1	
21	7		77	1	
22	2		80	1	
23	4		81	1	
24	2		83	1	
25	1		86	1	
27	2		88	1	
28	3		112	1	
29	2		124	1	
30	1		137	1	
31	1		145	1	
			363	1	

Table 9. Gene expression levels by EST abundance in the Unigene database for 12,490 entries.

Total Reads	Class		# Clusters/Singlets
1704 (13.6%)	Very abundant	63-363 ESTs > 0.5%	15
7602 (60.9%)	Abundant	3-62 ESTs 0.02-0.5%	1038
3184 (25.5%)	Low Abundance	1-2 ESTs < 0.02%	755 +2429 Singlets

The 15 members of the very abundant class represented almost 14% of the ESTs and were noticed early in the library sampling. The low abundance class represented just over 25% of the ESTs suggesting that additional groups of genes represented in the library at a much lower level were sampled but at a level lower than the 30-40% suggested by Bishop (Bishop, 1974). However, the 70% redundancy found by 3' EST assembly for the last sampling point (figure 10) suggested that close to 100% of the cDNAs from the first two abundance classes have been found (Soares, 1994).

3.6 Determining the Number of Genes Identified in the Unigene Database.

To obtain an estimate of the number of genes represented in the EST database, both the 5' and 3' EST sequences were aligned using Phrap (Green, copyright 1994-1996). As described above, the resulting Unigene database contained two types of data. The first were the sequences that occurred only once in the Unigene database and are termed the singlets. The second was the contigs which are the consensus sequences derived from the clusters, the aligned EST sequences which have homology as determined by the Phrap assembly criteria (Material and Methods section 2.20) . In addition, 367 of the 2429 EST sequences in the singlet category had their clone pair sequence aligned in a consensus sequence. There also were 718 clone pair members in

the singlets group whose sequences did not overlap and so could not align with each other to allow formation of a consensus but represent sequences from the 5' and 3' end of the same cDNA clone. Both of these categories were taken into account in determining that 1703 genes were represented in the singlets data category as shown in Table 10. An examination of the ESTs contributing to the consensus sequences showed that 10,061 ESTs were assembled into 1866 clusters. There were 742 clusters which shared a clone pair EST with other clusters. These were subtracted from the total to yield 1495 aligned EST consensus sequences representing 1495 unique genes. By totaling the 1703 genes represented in the singlets and the 1495 genes represented by the aligned ESTs it was concluded that 3198 genes were represented in the Unigene database. Since the estimate for the total number of genes present in the *A. nidulans* genome ranges from 8,000 to 12,000 (Kupfer, 1997; Timberlake, 1978) and see 3.30 Determining Gene Density), these 3198 genes in the Unigene database represent roughly one third of the *A. nidulans* predicted genes. Therefore, approximately one-third of the *A. nidulans* genes are expressed under the specific growth conditions which yielded the cells harvested for the cDNA library, a number consistent with the typical mammalian cDNA library which, even after subtraction and/or normalization yields only slightly greater than 30% of the postulated number of genes (Marra, 1999).

3.7 Identity of the Unigene Database Members, a Biological Classification for Cataloging the Expressed Genes.

An important goal of this study was to identify the genes represented in this vegetative/asexual *A. nidulans* cDNA library by comparison to known genes in the public databases. The predicted identity of the sequenced cDNAs and thus the expressed genes allows grouping by cellular roles and molecular families to facilitate data analysis. Several organizational schemes have been developed for this process (Riley, 1997;

Table 10. An estimate of the gene number in the Unigene database. containing 12,490 ESTs.

Total ESTs in Singlets-		2429
Singlets with pair member in Clusters-		-367
		2062
3'and 5' EST pairs in Singlets-	718/2=	-359 pairs
Number of genes represented in Singlets-		1703
10,061 aligned ESTS-		1866 Clusters
Number of Clusters sharing pair member	742/2=	-371
Number of genes represented in Clusters-		1495
		1703
		1495
Total genes represented by Unigene database		3198

White, 1997; Selkov, 1997). The Riley scheme was the earliest and one of the most comprehensive. It was designed for *E. coli* and provided for an exact cataloging of all proteins, and was based on both sequencing data as well as the extensive biological information available for this well-studied bacterium. More recently, genomic data from several additional organisms has been published and incorporated into versions of this classification scheme modified to reflect organism-specific information (Bult, 1996; Fleischmann, 1995; Fraser, 1995). Similarly, the genomic data for *Saccharomyces cerevisiae* and more than 30 bacterial and archeal species have been grouped into a comparable schema developed by Overbeek, et al. (WIT2 version 2.3, http://www.cme.msu.edu/WIT2/;Overbeek, 1997) called WIT2. In the study presented here, the seven major classes from the organizational scheme of Riley were used and a recent WIT2 analysis has been performed by Overbeek's group (Ross Overbeek, personal communication). When useful, each class was expanded to provide categories for all the Unigene members (Table 11). The bacterial genomic classification scheme of Riley lacks categories for asexual development, and a large number of metabolic functions that reflect the chemoheterotrophic and adsorptive nature of fungi. Although no scheme can be complete and different interpretations, such as placement of some functions in more than one category can be argued, this present fungal specific scheme given below with Table 11 and detailed in Appendices I and II followed established models and allowed presentation of the assembled *A. nidulans* EST data in an organized fashion that easily was accessed, modified, and analyzed.

Each cluster in the *A. nidulans* Unigene database was examined for homologs in the GenBank non-redundant protein database using the same parameters as for the individual ESTs (Materials and Methods). The singlets were not reexamined since the results of a BlastX homology search was completed earlier on the individual ESTs. Appendix II contains the complete cataloging of all the database members, both clusters (consensus) and the singlets. As can be seen, each of the 1788 Unigene

Table 11. *Aspergillus nidulans* categories of cellular functions.

I. Bioenergetics and Metabolism
  A. Metabolism of Carbohydrates
    1. macromolecules
    2. energy reserve biosynthesis
    3. sugars
    4. Calvin cycle
  B. Metabolism of Amino Acids
    1. biosynthesis
    2. degradation
  C. Metabolism of Nucleotides and Nucleic Acids, Purines, Pyrimidines
    1. purine metabolism
    2. pyrimidine metabolism
    3. nucleotide metabolism
  D. Biosynthesis of Lipids, Fatty Acids, Sterols
  E. Aromatic Compound Metabolism
  F. Sulfur Metabolism
  G. Phosphate Metabolism
  H. Nitrogen Metabolism
  I. Metabolism of Cofactors
  J. Energy
    1. glycolysis
    2. gluconeogenesis
    3. pentose-phosphate pathway
    4. tricarboxylic acid pathway
    5. fermentation
    6. glyoxylate cycle
    7. beta-oxidation of fatty acids
    8. metabolism of energy reserves (glycogen, starch, trehalose)
    9. respiration
II. Cell Growth, Cell Division
  A. Cell Walls
  B. Biomembranes
  C. Cytoskeleton, Organelle Biogenesis
  D. Cell Cycle Control
  E. Mitosis/Cytokinesis
  F. Meiosis
III. DNA Synthesis
  A. DNA Replication
  B. DNA Modification and repair
  C. DNA Packaging
IV. Gene Expression
  A. Transcription
    1. RNA polymerase
    2. regulation, includes asexual development regulatory pathway
    3. processing
    4. tRNA synthesis
    5. degradation
  B. Protein Biosynthesis
    1. Initiation, elongation, termination factors
    2. ribosomal proteins
    3. post-translational modifications

4. folding and targeting
5.Turnover

V. Cell Processes
A. Cell rescue, cell defense, osmotic adaptation, starvation response, development (asexual, sexual), includes antibiotics, toxins
B. Cell signaling, Signal transduction
1. kinases and second messengers
2. G proteins
3. cAMP
C. Transmembrane Transport
D. Classes of Enzymes, general cellular role
1. oxidoreductases
2. transferases
3. hydrolases
4. lyases
5. isomerases
6. ligases
7. synthetases
E. Non-enzymatic classes
1. zinc finger motif
2. leucine zipper motif
3. other regulatory proteins

VI. Unclassified (significant homolog but function uncertain in *Aspergillus nidulans*)
VII. Unidentified (significant match with ORFs)
VIII. No significant homolog

sequences with a significant homolog in the non-redundant protein databases was assigned a potential function based on BlastX scores (see Materials and Methods) and placed into the categories of cellular functions (Table 12). Those Unigene sequences with no significant homolog were not listed individually because of their number but are enumerated in section VIII. Because the listing is presented in an outline form, keywords, which represent the words used to search the BlastX descriptors in each report are shown in brackets (< >) under each section to distinguish them from headings. They are the words used to search the descriptors in each BlastX report (Material and Methods). The clusters were identified by a contig number followed by a contig identifier, for example Contig750_q0b02a1.f1, while the singlets were listed by their EST identifier, for example c4d02a1.r1.

Figure 12 displays a graphical representation of the percentage of Unigene members in each of the seven Riley categories plus those with no known homology (i.e. the no match category). Only 31% of the Unigene members could be given a functional assignment based on sequence homology. An additional 4% were unclassified, i.e. those with a homolog having a defined function in another organism but an unclear role in *A. nidulans*; while 9% had homology with ORFs or predicted proteins, which in the case of strong matches with Unigene members lent validation to the predicted genes , but gave no indication of the function in either organism. The remaining 56% of the Unigene sequences had no significant homology to any gene or protein in the GenBank databases. These results were in good agreement with recent findings from *Neurospora crassa*, a related multicellular ascomycete where Nelson et al. (Nelson, 1997) reported that 57% of the ESTs from three tissue specific libraries had no significant homologs in the databases using a significance cutoff of $p<10^{-4}$, a somewhat less stringent criteria than that used for *A. nidulans* (HSP>99 and $p<10^{-4}$). Therefore, the present data suggests that a significant number of the sequences in the *A. nidulans* Unigene database represent newly discovered expressed genes.

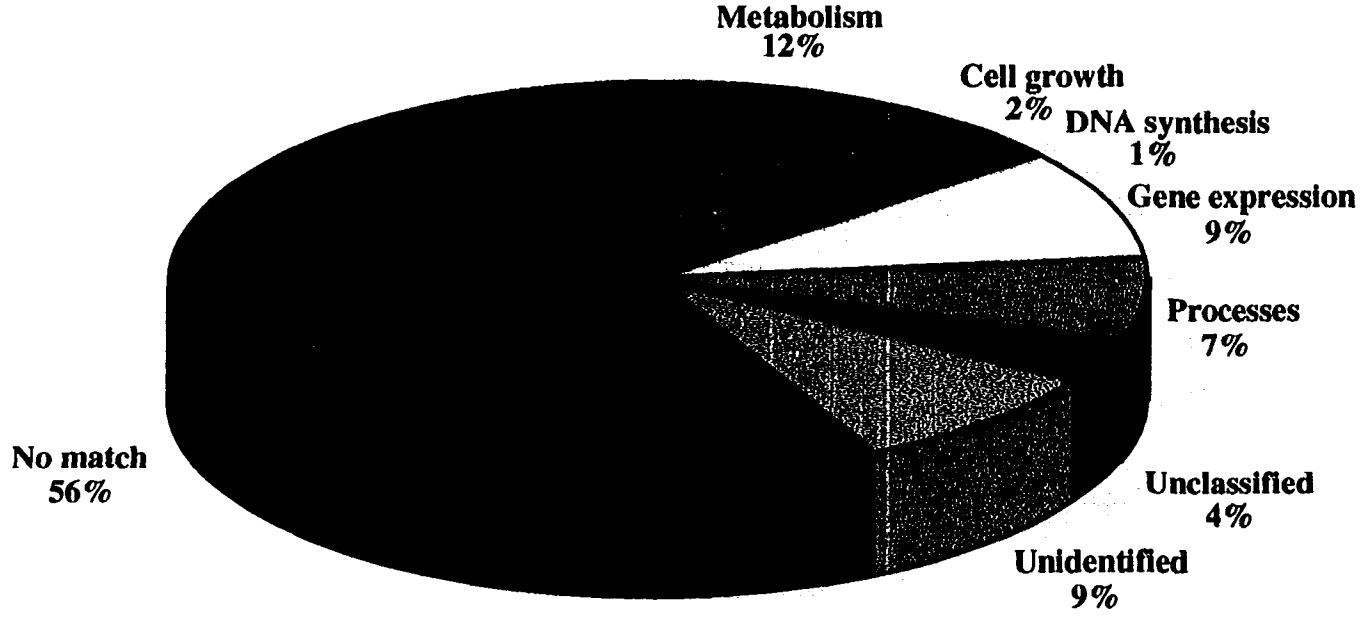


Figure 12. The Unigene database biological function classification by percent of members falling into each of eight catagories

3.8 Unigene Representation in the *A. nidulans* Biological Classification Scheme.

In addition to the rapid detection of expressed genes, EST sequencing allows for determining the relative levels of gene expression by their representation in the EST library. The alignments which generated the *A. nidulans* EST Unigene database also revealed the number of homologous EST sequences in each cluster as well as the relative position of each EST in the consensus sequence as shown below in Figure 13.

```
Sequence Contig923
Assembled_from z4b03a1.f1.comp -557 382
Assembled_from z4b03a1.r1 -73 516
Assembled_from a5f07a1.f1 .comp -46 1057

Sequence Contig1255
Assembled_from d3c05a1.r1.comp -430 770
Assembled_from d3c05a1.f1 -81 464
Assembled_from k0f07a1.f1 -68 524
Assembled_from k0f07a1.r1.comp 200 642

Sequence Contig1453
Assembled_from v4g03a1.f1 -117 723
Assembled_from v3g03a1.f1 -75 567
Assembled_from b0h05a1.f1 -65 466
Assembled_from u4b10a1.r1.comp 376 1063
Assembled_from b0h05a1.r1.comp 563 1006
```

Figure 13. An excerpt from the relationships file of the Unigene database showing the identity of the EST members in selected clusters and their relative position in the alignment used to generate a consensus sequence. Comp indicates that the complimentary sequence was required for alignment.

Thus, the percentage of genes assigned to each of the seven categories could be compared to the percentage of transcripts as represented by the EST clone pair sequences. In general, as shown in Figure 14 there was a direct correlation between the percent of the total genes and their relative expression levels. The notable exception was with the highly expressed genes in the "Gene Expression" category. Here, the percent of assigned genes (21.7%) was lower than the percent of total transcripts (32.6%). Table 12 indicates the number and percent of genes as well as the gene expression levels in the subcategories. A large number of genes, 286, were assigned to the protein biosynthesis category where the

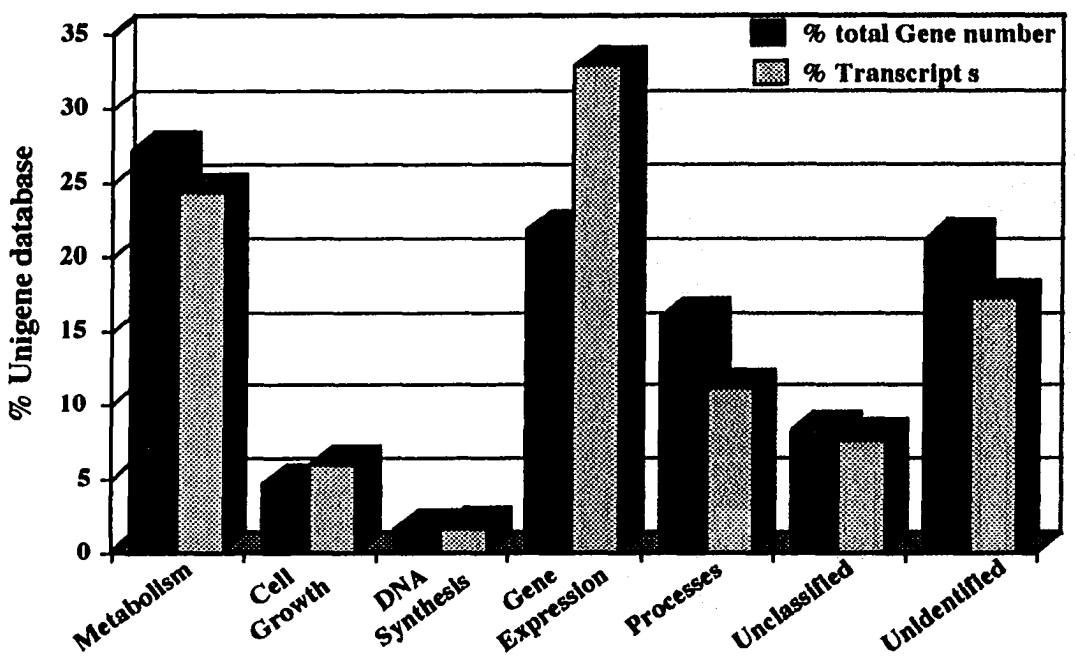


Figure 14. Gene vs transcript representation in the seven function catagories for 3198 genes and 8612 cDNAs.

Table 12. Comparison of gene number and gene expression level for the categories of cellular functions for 1775 genes and 4604 cDNA clones . The percent values in the subcategories are the percent of genes or cDNA clones in seven categories.

											Total	%Assigned
I. Bioenergetics and Metabolism												
	Carbo-hydrates	Amino Acids	Purines, Pyrim-idines	Lipids, Fatty Acids, Sterols	Aromatic Com-pounds	Sulfur	PO_4-	Nitrogen	Cofactors	Energy		
Genes	79	47	22	43	6	2	1	7	27	247	481	27.1
	4.4%	2.6%	1.2%	2.4%	0.3%	0.1%	.05%	0.4%	1.5%	14.0%		
cDNAs	232	100	45	84	12	2	2	33	64	551	1125	24.4
	5.7%	2.5%	1.1%	2.0%	0.3%	0.05%	.05%	0.8%	1.6%	13.6%		
II. Cell Growth, Cell Division												
	Cell Walls	Bio-membranes	Organelle, Cytoskeleton	Cell Cycle	Mitosis, Cytokinesis	Meiosis						
Genes	12	7	30	16	14	2					81	4.6
	0.68%	0.4%	1.7%	0.9%	0.8%	0.1%						
cDNAs	14	14	64	23	20	2					137	3.0
	0.3%	0.3%	1.6%	0.6%	0.5%	0.05%						
III. DNA Synthesis												
	Replication	Modification Repair	Packaging									
Genes	9	8	9								26	1.5
	0.5%	0.45%	0.5%									
cDNAs	15	8	50								73	1.7
	0.4%	0.2%	1.2%									
IV. Gene Expression												
	Transcription	Protein Biosynthesis										
Genes	100	286									386	21.7
	5.6%	16.1%										
cDNAs	193	1308									1501	32.6
	4.2%	28.4%										

V. Processes	Adaptation	Signalling	Transport	Enzymes	Non-Enzyme		
Genes	73	76	96	25	11	281	15.8
	6.2%	6.5%	8.1%	2.1%	0.9%		
cDNAs	142	130	168	47	16	503	11.0
	3.5%	3.2%	4.1%	1.1%	0.4%		
VI. Unclassified							
Genes						145	8.2
cDNAs						347	7.5
VII. Unidentified							
Genes						375	21.1
cDNAs						788	17.1

Total in no assignment category:
genes: 2034
cDNA:4008

level of measurable gene expression is high at 15.2% of the total assigned clones. Ribosomal proteins and chaperones are two classes of highly expressed genes in the protein biosynthesis category that contributed to this high level, representing 4.9% and 6.1%, respectively, of the total cDNA clone population. The other highly represented class was bioenergetics and metabolism, where over one-fourth of the genes were assigned and nearly one quarter of the transcripts were represented by cDNA clones.

Comparison of the *A. nidulans* Unigene database with the 1879 EST sequences reported for *Neurospora crassa* (Nelson, 1997) showed a slightly different gene expression pattern. Three hundred seventy-three clones from the three *N. crassa* libraries examined were placed into the seven Riley categories with the result that 45% of their cDNAs were metabolism related and 32.6% were involved in protein synthesis. It is interesting to note that no chaperone proteins were discovered in the *N. crassa* libraries and that the protein synthesis category consisted primarily of ribosomal proteins. Nelson et al. (Nelson, 1997) suggested that because only a small number of sequences were collected, only the very highly expressed genes have been observed. However, it remains a notable difference that no chaperone class of proteins was detected since this class represented a significant fraction of both the very abundant and abundant groups in the *A. nidulans* library.

3.10 An examination of the highly expressed Unigene members.

Table 13 lists the significant BlastX homologs for each of the 15 clusters making up the very abundantly expressed gene category. Nine consensus sequences have a significant homolog but only seven have predicted functions. Three of the clusters are members of the HSP30 family, a class of small chaperone proteins of the *S. cerevisiae* HSP20 family (Kusakabe, 1994), one cluster represents ubiquitin, a protein involved in protein turnover (Ozkaynak, 1984), one is a homolog of the spore-wall fungal hydrophobin, dewA, produced during asexual development (Stringer, 1995), one cluster

Table 13. The homologs of and EST representation in the very abundantly expressed class of the Unigene database.

Contig Number	Total ESTs in Contig	% of ESTs	Homologs with BLASTX HSP>99
1866	363	2.9	Heat shock protein 30, *Aspergillus nidulans*
1865	145	1.2	Ubiquitin, *S. cerevisiae, Candida albicans*
1864	137	1.1	Heat shock protein 30, *A. nidulans*
1863	124	1.08	Metalloproteinase, *A. fumigatus*
1862	136	1.0	Heat shock protein 30, *A. nidulans*
1861	112	0.9	Spore-wall fungal hydrophobin, *A. nidulans*
1860	88	0.7	ORF W02A2.g *C. elegans*
1859	86	0.69	Chitinase, *A. nidulans*
1858	83	0.66	Glucose repressible gene, *Neurospora crassa*
1857	81	0.65	no significant homologs
1856	80	0.64	no significant homologs
1855	77	0.62	no significant homologs
1854	65	0.52	no significant homologs
1853	64	0.51	no significant homologs
1852	63	0.5	no significant homologs

was chitinase, an intracellular enzyme which may be involved in cell wall remodeling, growth and development (Takaya, 1998), and one was a metalloproteinase, which is an extracellular peptidase involved in protein degradation (Ramesh, 1995). Two of the clusters had homology to genes which have no known function; one to a *C. elegans* ORF and the other to a glucose repressible gene of *N. crassa*. The remaining six very abundantly expressed clusters had no significant matches in the GenBank databases. Each cluster represents expressed sequences from a single *A. nidulans* gene. Therefore, roughly 40% of the very abundantly expressed genes found in the *A. nidulans* library have no known homologs. It is noteworthy that 39 of the 100 most highly expressed genes represented in the Unigene database have no significant matches in the GenBank databases, suggesting that a large number of new genes are represented in the *A. nidulans* Unigene database. An expanded discussion of the members of the Unigene very abundant class is given below.

3.11 Heat Shock Protein 30 Representation in the Unigene Database.

The Heat shock protein (HSP) 30 representation in the *A. nidulans* cDNA library was striking, as it included three of the 15 very abundant class and three additional HSP30 homologs from the abundantly expressed class. Although all six had their highest alignment score with accession number D32070, the only *A. nidulans* HSP30 in GenBank, they represented six different HSP30 genes as discussed below and contributed significantly to the protein expression category. A total of 684 HSP30 sequences were present which represented 5.5% of the total EST population from only six distinct EST clusters (Table 14). The relationship between these six *A. nidulans* families and to other HSP30s also was examined. Heat shock proteins in general serve as molecular chaperones that are reversibly associated with nascent proteins to assist their transport in the cell, their translocation across membranes and their proper folding. They also are produced in response to a variety of cell stresses and growth conditions such as

Table 14. HSP30 representation in the EST Database.

Family	EST#	Percent in Database
Contig 1866	363	2.9%
Contig 1864	137	1.1%
Contig 1862	112	0.9%
Contig 1820	28	0.22%
Contig 1812	23	0.18%
Contig 1800	21	0.17%
Total ESTs	684	5.5%

elevated temperature, glucose starvation, and chemical treatments (Gething, 1992; Lindquist, 1988). All HSPs have well conserved primary sequences and have been classified into either low molecular weight, HSP30, or the high molecular weight, HSP60, 70 and 90, families (Kusakabe, 1994). HSP30 occurs abundantly in plants while HSP70 predominates in animals (Lindquist, 1988). Kusakabe et al. (Kusakabe, 1994) isolated and sequenced a cDNA (Accession number D32070) from a heat stressed *A. nidulans* culture. Northern hybridization experiments revealed that there was a high level of HSP30 mRNA in *A. nidulans* grown at normal culturing temperature (30^0C). Kusakabe et al. (Kusakabe, 1994) also found that the corresponding gene did not have heat shock response elements upstream of the coding region, suggesting constitutive expression.

To further investigate these HSP30 ESTs, each of the six HSP30 contiguous sequences in the Unigene database was examined for an open reading frame and translated. An amino acid alignment was done using the GCG Bestfit program (Genetics Computer Group, Ver. 5) for each against D32070. The percent identity for each alignment are shown in Table 15 along with the HSP and ρ values from the BlastX search.

Table 15. Unigene HSP30 members comparison with GenBank entry D32070.

Unigene identifier	HSP	p value	% identity to D32070
Contig1866	816	1.2 e-80	87.6
Contig1864	933	4.8 e-93	100
Contig1862	291	2.9 e-43	60
Contig1820	832	2.4 e-82	90
Contig1812	612	3.7 e-59	92
Contig1800	855	8.9 e-85	91.6

Clearly, Contig1864 represents the same gene as accession number D32070 characterized by Kusakabe since it displayed 100% identity. A multiple alignment, performed to examine the six for regions of sequence identity using the GCG program Pileup (Genetics Computer Group, 1996), is shown in Figure 15. When analyzed by a Blocks search

(Henikoff, 1991), each HSP30 matched block BL010031B that contains an FPK motif and domain 473, a Prosite (Bairoch, 1995) pattern for the HSP30 family which contains a consensus sequence also found in alpha-crystallin (Plesofsky-Vig, 1995). Both characteristics were marked in Figure 15 and showed that six families of closely related HSP30 proteins were present in *A. nidulans*. The closest homolog to the *A. nidulans* HSP30 family in the databases was *Neurospora crassa* accession number P19752 which had a 42% sequence identity to Contig 1864 and a 43.4% sequence identity to Contig1862, the most distantly related contig in the *A. nidulans* Unigene family. The *N. crassa* homolog was found in association with the mitochondria during periods of glucose limitation and elevated temperature (40^0C) (Plesofsky-Vig, 1995). It is tempting to speculate that since the *A. nidulans* culture from which the library was grown on minimal media with 1% glucose at high temperature (37^0C) (R. Aramayo, personal communication) rather than 30^0C (Kusakabe, 1994), that under respiratory limitations at the elevated temperature populations of HSP30 were induced. In contrast, the *N. crassa* cDNA libraries studied by Nelson et al. (Nelson, 1997) were grown at 25^0C and no HSP30s were represented in the ESTs which they generated. Thus, in addition to the constitutive population seen by Kusakabe, additional HSP30s possibly were induced at the elevated temperature. It would be interesting to examine directly the HSP30 population in an developing culture grown at 25^0C or 30^0C as well as a vegetative only culture grown on rich media with 2% glucose to investigate the conditions which modulate HSP30 induction.

The relationship of the *A. nidulans* HSP30 Unigene members to HSP30 genes in other organisms was examined using the clusters of orthologous groups (COGS) provided by NCBI. Clusters of orthologous groups describe divergent evolutionary relationships between proteins with a common ancestor. The clonal theory of evolution is based on the analyses of rRNA sequences and holds that all genes are passed directly to each succeeding generation (Lake, 1999). Variations in these genes result in altered Fig

```
          1                                                          50
D32070    MSLFRTIPTP GDFAPLFRLL DDYDNHRSAR G...H.ASVQ SFAPRFDVRE
  1864    MSLFRTIPTP GDFAPLFRLL DDYDNHRSAR G...H.ASVQ SFAPRFDVRE
  1820    MSLFRTIPTP GEFAPLFRLL DDYDVHRSTR G...Q.TVVQ SFAPRFDVRE
  1866    MSLFRTIPTP GEFAPLFRLL DDYDVHRSTR G...Q.TVVQ SFAPRFDVRE
  1800    MSLFRTIPTP GDFAPLFRLL DDYDNHRSAR G...H.ASVQ SFAPRFDVRE
  1812    MSLFTTTPSV SSFAPLFNLL DDYDNHLASR NWGHH.TSVR SFSPRFDVRE
  1862    M~AFFPRYCS GDFAPLFQLL DDYDMHQATR RPNKKVTNVR TFVPKFDVYE

                              FPK motif
          51                  F  P     K                        100
D32070    SNEAYHLDGE LPGIPQSNID IEFTDPQTLV IKGRSEREYH SSSDDNKNDQ
  1864    SNEAYHLDGE LPGIPQSNID IEFTDPQTLV IKGRSEREYH SSSDDNKNDQ
  1820    SNEAYHLDGE LPGIPQSNIE IEFTDPQTLV IKGRSEREYH .SNDENKAEQ
  1866    SNEAYHLDGE LPGIPQSNIE IEFTDPQTLV IKGRSEREYH .SNDENKAEQ
  1800    SNEAYHLDGE LPGIPQSNID IEFTDPQTLV IKGRSEREYH .SNDENKAEQ
  1812    TSDTYHLDGE VPGVAQKDID IEFTDPQTLV IKGRVERQYH SGNTDDTGKQ
  1862    QGDRYYLDGE LPGVSQSNIE IEFTDPQTLV IKGHSKRNYH HKSEPDTDDK

          101                                                        150
D32070    ADTE..... ......NQAR GESSEVAKTG .........E KQVSTKKAAN
  1864    ADTE..... ......NQAR GESSEVAKTG .........E KQVSTKKAAN
  1820    AETE..... ......KPVQ GESSEVAKTG .........E KQISTKKAAN
  1866    AETE..... ......KPVQ GESSEVAKTG .........E KQISTKKAAN
  1800    AETE..... ......KPVQ GESSEVAKTG .........E KQISTKKAAN
  1812    RQVE..... ......DENE SSSNEVAKTS .........E KQMTKSASSE
  1862    SETSSVKSLQ PTVEDWDEME DATPAVEQTP SLGPKEKAVE KNSSTRSQEP

          151                    Domain hsp20                        200
D32070    KSRYWVSERS VGEFQRTFTF PTRVNQDDVK ASLKDGILSL VVPKAVPPTA
  1864    KSRYWVSERS VGEFQRTFTF PTRVNQDDVK ASLKDGILSL VVPKAVPPTA
  1820    KPRYWVSERS VGEFQRTFTF PTRVNQDDVK ASLKDGILSL VVPKAVPPTA
  1866    KPRYWVSERS VGEFQRTFTF PTRVNQDDVK ASLKDGILSV IVPKAVAPSA
  1800    KPRYWVSERS VGEFQRTFTF PTRVNQDDVK ASLKDGILSV IVPKAVAPSA
  1812    KPRYWVSERS VGEFQRTFSF PSRVDQDRVR ASLRDGILSV VVPKEAPPNA
  1862    AYKFWASERL VGEFSRTFAF PTRVDQDAVR ASLNNGILSV VLPKEPAPQL
Consensus KWHRMERS SGKFMRRFRL PENVKVDEIK ASMENGVLTV TVPK
          201
D32070    KKITIQ
  1864    KKITIQ
  1820    KKITIQ
  1866    KKITIQ
  1800    KKITIQ
  1812    KKITIQ
  1862    KKVRVE
```

Figure 15. The alignment of six HSP30 families found in the Unigene database aligned with the single representatiave from Genbank, AC# D32070. In red are the positions conserved in all *A. nidulans* members. The Prosite HSP30 consensus shown has the identity positions highlighted in blue.

phenotypes and the appearance of new species. The evidence from completed genome sequences has argued against the clonal theory as being the only method of gene acquisition. The examination of functional groups of proteins from eucaryotes, procaryotes and archaea suggest that both eucaryotic and prokaryotic genomes may be chimeras, obtaining transcription and translation related genes from an archaea while eucaryotes obtained the housekeeping genes from a procaryote (Rivera, 1998). Tatusov et al.(Tatusov, 1997) describe a comparison of proteins encoded in eight complete genomes from five phylogenetic lineages and delineated 824 clusters of orthologous groups (COGs) based on consistent patterns of protein sequences. These 824 COGs have between three and over 100 ortholog members, where orthologs are genes in different species that evolved from a common ancestral gene and result from a horizontal transfer (Koonin, 1997). Paralogs are genes that have arisen within a genome by duplication (Henikoff, 1997). Orthologs and hence the COG members are assumed to have retained the same function while paralogs have evolved new functions which may or may not be related to the original.

Koonin's group have determined a series of COGs and made them publicly available on the National Center for Biotechnology Information (NCBI) web site (http://www.ncbi.nlm.nih.gov/COG/cognitor.html). Any sequence can be compared to the COG database to determine its similarities to previously determined orthologous groups. Comparing individual HSP30 members of the *A. nidulans* Unigene database with the COG database revealed that the HSP30 Unigene members had significant homology to COG0071 a molecular chaperone COG containing four members, one each from *Methanococcus jannacheii*, an Archaea, *Senechocystis* a cyanobacteria, and two paralogs from *Saccharomyces cerevisiae*, the Ascomycete yeast. All four COG members had the conserved FPK motif within their consensus regions. The GCG multiple alignment program Pileup (Genetics Computer Group, 1996) was used to create a relational dendrogram for the *A. nidulans* HSP30 family, the *N. crassa* HSP30 and the

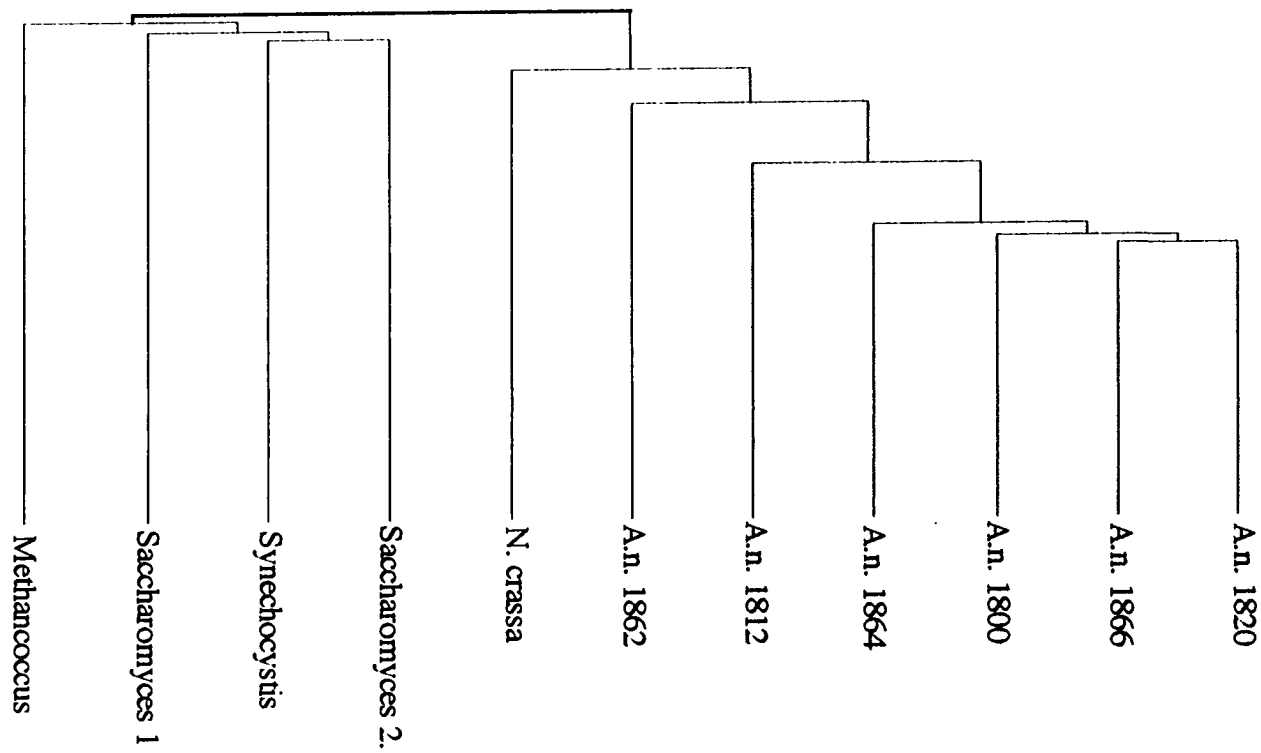


Figure 16. A dendrogram showing the sequence based relationship of the filamentous fungal HSP30 members to the members of the NCBI chaperone COG. A.n. is *Aspergillus nidulans*.

COG members. The results are shown in Figure 16. The filamentous ascomycete members, *N. crassa* and *A. nidulans*, formed a well-defined group and appeared to be comparably related to all four of the NCBI COG members. Since *S. cerevisiae* is not only a fungus but also an ascomycete, it is surprising that there is no difference in the filamentous ascomycete relatedness to the two *S. cerevisiae* and either the bacterial or archael representatives as might be expected. As seen by the distances from branchpoints in the dendrogram (Figure 16), the *Saccharomyces* HSP members of the COG were as closely related to the other COG members as they were to the filamentous ascomycete HSP30s. This observation suggested that the HSP30 progenitor had entered all three evolutionary groups by horizontal gene transfer. However, since there is a distant relationship between the HSPs of the multicellular ascomycete and the two *Saccharomyces* HSPs, this transfer most likely occurred after the evolutionary divergence of single and multicellular ascomycetes that occurred approximately 2.5 million years ago (Berbee, 1992).

3.12 Ubiquitin

Another member of the highly expressed clusters, Contig1865, had equivalent and high BlastX homology to three GenBank members, *Candida albicans* (Z54197), *Saccharomyces cerevisiae* (D29456), and *Nicotiana tabacum* (AJ223328) all three had an HSP of 1457 and $p=1.4e^{-148}$ and are ubiquitin homologs. In eucaryotes, ubiquitin is a highly conserved 76 residue protein involved in targeting intracellular proteins for degradation by conjugation of ubiquitin to the proteolytic substrates in an ATP-dependent reaction (Ozkaynak, 1984). The *A. nidulans* homolog had four perfect 76 amino acid repeats arranged in the head-to-tail fashion also seen in the yeasts and tobacco (Ozkaynak, 1984). This is an unusual structure for a precursor protein. Generally, spacer sequences are found between each copy which are lost during processing. There is a single amino acid difference, a serine for a threonine, in three of the repeats in the other homologs. In

the fourth repeat this residue is a threonine in all four organisms leading to 301/304 identical amino acids in this *A. nidulans* polyubiquitin compared to the fungal and tobacco homologs. It was not surprising to find that ubiquitin was expressed at high levels in the *A. nidulans* library since it is a key component in protein turnover. It is interesting to speculate that since the culture from which the library was derived was grown on minimal medium for 24 hours that the high level of ubiquitin was necessary to prepare the cells for this growth stage or that it could be part of a mechanism for providing carbon and nitrogen for a nutrient limited culture.

3.13 Metalloproteinase

Contig1863 was homologous to an *A. fumigatus* mep20 metalloproteinase (accession number JC4379). This metalloproteinase is a thermostable, zinc binding, extracellular protease which cleaves basic proteins such as histones and protamine. Comparison of the mep20 amino acid sequence with the Unigene homolog protein revealed a 57% sequence identity with the *A. fumigatus* sequence and conservation of the zinc binding domain AQDQATTTL**HEFTH**APGUY including the active site shown in bold type. The presence of this domain suggests that the *A. nidulans* Unigene member is a homolog of the *A. fumigatis* mep20 gene. The biological role of mep20 (Ramesh, 1995) and another homolog in *A. oryzae* (Tatsumi, 1991) is not clear but Tatsumi assumes that mep20 "serves to utilize substrates in the environment" since *A. oryzae* secretes a large amount, 300mg/liter, under normal culture conditions. Its high level of representation in the EST database from a culture entering carbon and nitrogen starvation is consistent with action as a "scavenger" during a time of increased cell death and lysis degrading external carbon and nitrogen sources for their absorption into the starving cell.

3.14 Dew A, a fungal hydrophobin

Contig1861, another very abundantly expressed cluster member has 100% amino

acid and DNA homology with the *A. nidulans dew*A protein, accession number P52750. The dewA protein is a spore wall specific hydrophobic protein which appears on the surface of the conidia and is one of several fungal hydrophobins needed for cell-surface hydrophobicity (Stringer, 1995). Since the *A. nidulans* culture was harvested during conidiation, it is not surprising that this gene was highly expressed in the cDNA library.

3.15 Chitinase

Contig1860 matched with 100% identity over 398 of 416 amino acids of the *A. nidulans* chitinase, GenBank accession number D87063. This nonsecreted chitinase which depolymerizes fungal wall chitin efficiently resulting in high molecular weight polysaccharides (Reyes, 1989) is thought to be expressed during morphogenesis (asexual development) (Takaya, 1998). This high degree of homology suggests that Contig1860 represents a chitinase similar to D87063 and is highly expressed because of the asexual development the *A. nidulans* culture was undergoing at the time of harvest. However, as can be seen in Figure 17A, the alignment of the first 42 residues of Contig1860 with D87063 only is possible if there is an 18 residue gap in Contig1860. The remainder of the alignment was 100% identical. The alignment of the cDNA version of the chitinase with the genomic conceptual translation suggests that there may be alternately spliced versions of this chitinase gene or that an intron, in the same reading frame, was included in the genomic translation and the cDNA reveals the true translation product. The second possibility is strongly suggested by examination of the nucleotide sequence shown in Figure 17B from accession number D87063 for the region covering the 18 residue gap. Both the 5' and 3' ends show the conserved intron border sequences, shown in bold (Mount, 1982).

3.16 The glucose-repressible gene

Another very abundantly expressed Unigene member, Contig 1859 which has

A. dbj|BAA35140| (D87063) chitinase [Emericella nidulans]
Length = 416

Score = 831 bits (2124), Expect = 0.0
Identities = 398/398 (100%), Positives = 398/398 (100%), Gaps = 18/398 (4%)

```
Contig1860   MSGYKTVGYFVNW----------------AIYGRNYNPQDLPAEKLTHILYAFANVRP 42
             MSGYKTVGYFVNW                AIYGRNYNPQDLPAEKLTHILYAFANVRP
D87063       MSGYKTVGYFVNWVRTSCLLPIYISFTNDRQAIYGRNYNPQDLPAEKLTHILYAFANVRP 60
```

```
          V  R  T  S  C  L  L  P  I  Y  I  S  F  T  N  D  R  Q
B.  5'  gtacgt acttcctgtctcctaccaatatatatatctttcactaacgat cgacag  3'
```

Figure 17. Comparison of Contig1860 with *A. nidulans* chitinase, accession number D87063. A. partial amino acid comparison of the translated Unigene chitinase homolog and the GenBank *A. nidulans* chitinase, chiB. The dashes represent the gap in the sequence alignment in Contig1860 required to align to D87063. B. The genomic region corresponding to amino acids 14-31 of accession number D87063. The translation is shown above the nucleotide sequence. Letters in bold are the conserved 5' and 3' intron border sequences.

59% amino acid identity and 73% conservation with grg-1, glucose repressible gene of *N. crassa* (Figure 18). A search of the Blocks database gave no matches in the database so it is not known which regions are important for function. This glucose repressible gene codes for a protein of unknown function but the level of grg-1 mRNA increases over 50 fold during the first hour of glucose deprivation (McNally, 1988). Since in *N. crassa*, a number of activities are expressed when glucose is limiting including invertase, amylases and high-affinity glucose transport systems (McNally, 1988) the overexpression of a grg-1 homolog is not surprising. If contig1859 is indeed a homolog of *N. crassa* grg-1 its presence in the very abundant expression group would be consistent with the decreased levels of available glucose 24 hours after plating on the minimal medium.

```
sp|P22151|GRG1_NEUCR GLUCOSE-REPRESSIBLE GENE PROTEIN >gi|3014|emb|CAA32907|
(X14801)
          grg1 [Neurospora crassa]
          Length = 71aa

 Score = 85.8 bits (209), Expect = 1e-16
 Identities = 41/69 (59%), Positives = 51/69 (73%)
 Frame = -3

1859: 344 METVKNAVNYVSESVQGAGATASKETNKNVAKDSDASLTSRATAAKDAVVDKKDEKSHDA 165
          M+T+KNA NYV + VQGA ATASKE NK+VAKDS+  + +R  AA DA+ DK  E  HDA
grg-1:1   MDTLKNAANYVGDKVQGATATASKEANKDVAKDSNQGVGTRLNAAGDAISDKVSENKHDA 60

`1859: 164 KADVHKEAA 138
           KA+ HK+ A
grg-1: 61  KAEAHKQGA 69
```

Figure 18. Alignment of the *A. nidulans* grg-1 homolog (1859) with the *N. crassa* grg-1 protein. (+) means a conservative change in sequence. Alignment shows high level of identity especially at 5' end.

3.17 C. elegans ORF W02A2.g

Contig1858 represents a gene which has homology with 13 GenBank entries. All of the entries are ORFs either from *C. elegans*, *S. cerevisiae* or *E. coli* with significance scores ranging from $1.3e^{-13}$ to $1.4\ e^{-08}$. However, the actual homology was in all cases

only over a small 5' region of 54 amino acids from the 198 amino acid translation product from Contig1858. The best alignment was with the *C. elegans* ORF W02a2.g shown in figure 19 A. A search of the Blocks database showed a match with block BL01309, that represented a domain from the uncharacterized protein family UPF0057 (figure 19 B). Although the function of the gene product represented by Contig1858 in the Unigene database is not clear, it does represent a new *A. nidulans* gene belonging to a protein family which has not been previously reported in a multicellular fungus.

A.

```
      >gnl|PID|e1188370 (Z82286) W02A2.g [Caenorhabditis elegans]
          Length = 594

 Minus Strand HSPs:

 Score = 193 (67.9 bits), Expect = 1.3e-13, P = 1.3e-13
 Identities = 35/54 (64%), Positives = 45/54 (83%)

C 1858:  1 MPFTASDICKLIFAFILPPLGVFLERGCGADFLINICLTILGWIPGIIHAIYII 54
           MP T +DI K I A +LPP+GV++E+GCGAD +INI  LTILG+IPG+IHA +II
W02A2.g: 1 MPITCTDIPKFICALLLPPIGVWMEKGCGADLVINIVLTILGFIPGVIHACFII 54
```

B.

```
BL01309  INILLTILGYIPGIIHACYVI
         ||| ||||| |||||||  | |
C 1858   INICLTILGWIPGIIHAIYII
```

Figure 19. The characterization of Unigene member Contig 1858. A. The small region of homology between Contig1858 and the *C. elegans* ORF W02A2.g is shown. No other region in the *A. nidulans* 198 residue translation product aligned. The highlighted region is the conserved uncharacterized domain found in all other homologs. B. Shows the alignment of the domain found in Contig1858 with the Block BL01309 representing an uncharacterized protein family (Henikoff, 1991).

3.18 Additional examination of "no match" Unigene members.

It is clear that Unigene members with no significant homologs in the GenBank databases represent new genes and that it would be useful to obtain some indication of their cellular function. Therefore, each of the six Unigene members in the highly abundant class which showed no matches with a BlastX score of 100 or greater and a ρ value $<e^{-4}$ (Table 13) were examined in greater detail. To examine further if there were

any significant homologs in the databases, each was translated and the translation product was used for a BlastP search without the low complexity default masking function seg. For Contig1857, the BlastP showed a weak similarity to mucin and a blocks database search (Henikoff, 1991) showed a partial match with a subset of the selectin superfamily motifs that represented a transmembrane domain termed sushi:

```
mouse-FSFSSQCT
      |   ||||
1857 -FCYRSQCT
```

Thus, there is some evidence that Contig1857 may be a membrane protein with a transmembrane domain.

The BlastX report for Contig1852 showed homology to amino acids 144-298 of a thyroid hormone responsive gene and the Downs syndrome region 1 alternatively spliced exon 1 with a score of 50 and e value $< 5e^{-5}$. A blocks search indicated a match with two regions of block BL00354 from the high mobility group 1 (HMG) DNA binding domain, a cell cycle dependent human embryogenesis transcription factor:

```
human-KRGRGRPRKQ
       | || || | The RGR motif
 1852-RRRRGQPRQQ

human-AEEYGNTSSDSSDED
       || |      | |   The ESE motif
1852 -EEESGGGGGKGSDEQ
```

A ProDom search (Sonnhammer, 1994) revealed a different domain in Contig1852 that had homology to domain 82878 of the *Drosophila* DNA binding protein k10. This domain contains the Wilms tumor motif that has been associated with developmental transcriptional regulation and binds to the DNA sequence CGC CCC CGC. In addition, since the domain 82878 overlaps block BL00354, Contig1852 may encode a gene product which may be developmentally regulated, have DNA binding capability, and thus may itself regulate transcription.

The remaining four Unigene members had no significant regions of homology to any known proteins, domains or COGs as revealed by BlastP, domain searches, or

search with the NCBI COGs database and thus their function remains unknown. Finally, all six Unigene member where examined with the Prosite program (Bairoch, 1995) from which Blocks is derived and the Stanford University emotifs Identity program (Nevill-Manning) but this gave no additional information for any of the six Unigene members examined.

3.19 Summary

Over 8500 clones were isolated and almost 13,000 high quality ESTs were generated by sequencing both ends of the cloned inserts during the course of this dissertation research. A series of automated scripts was developed in collaboration with Hongshing Lai, ACGT informatics group, to trim each EST to high quality endpoints and to remove any contaminating sequences. A GenBank submission file was created for each high quality EST that was submitted to the GenBank dbEST database. In addition, each high quality EST was examined by a BlastX search against the nonredundant protein databases and the sequence as well as the homology information was made publicly available on the ACGT website to facilitate their further study.

The 3' ESTs were assembled into a Phred/Phrap database that allowed construction of a graph which displayed the percent redundant sequences, was a measure of library sampling and a means to determine a sampling endpoint.

A Phred/Phrap assembly of both the 5' and 3' ESTs create the *A. nidulans* Unigene database from which the approximate number of genes, 3198, represented in the library could be determined. A further analysis of this database revealed three expression levels, approximately 14% of the Unigene members belonged to the abundantly expressed group, 61% belonged to a moderately expressed group and 25% belonged a rarely expressed group. When the Unigene database members with known functions were placed into biological function groups, 44% of the Unigene members could be assigned functional groups such as bioenergetics, gene expression or cellular processes. The majority of the Unigene members, approximately 56%, had no homologs and

represented new genes.

An examination of the highly expressed genes revealed that six of the 15 very abundantly expressed Unigene members had no homolog in the GenBank databases, consistent with the observation that a significant number of new genes were represented in the Unigene database. Of the members with significant homologs, many were members of the HSP30 gene family. A study of these HSP30 related ESTs showed that the assembly had aligned them into six families, each representing a gene, five which had not been previously reported. A comparison of these six HSP30 sequences to the NCBI COG database revealed their homology to the chaperone COG but surprisingly they were not closely related to the *S.cerevisiae* orthologs. Six of the other very abundantly expressed genes had homologs which, in general, were related to a cellular response to growth under conditions of carbon and nitrogen limitation, leading to asexual development. Finally, an attempt to determine any potential function for the six remaining very abundantly expressed Unigene members via a conserved domain search had mixed results since two of the six showed either conserved motifs or domains that suggested a possible sub-cellular location and/or function while four did not.

3.20 *Aspergillus nidulans* Cosmid Sequencing.

Since one purpose of the above EST studies was to aid in genomic sequence annotations, the next section will describe the sequencing and annotation results for three *A. nidulans* chromosome VIII cosmids. Each completed cosmid initially was examined by a BlastN search of the GenBank nonredundant database which included the *A. nidulans* ESTs sequenced as part of this dissertation research. The results demonstrate the application of EST sequences to annotation of genomic DNA sequences.

3.21 Cosmid W06E08 Analysis

Sequencing of the Chromsome VIII cosmid was undertaken as part of a pilot

project to determine if the minimum tiled library created by Prade et al. (Prade, 1997) was suitable for the proposed genomic sequencing of *A.nidulans* . In addition, the sequence was used to estimate the approximate gene density and total gene number for *A. nidulans*. This particular cosmid was selected since it had been shown by hybridization to contain the development specific transcription factor, brlA (R. Prade, personal communication) whose sequence was known (Prade, 1993). The cosmid sequence determined by fluorescent sequencing methods could be compared with the region previously sequenced by radiolabelled methods to check the accuracy of both sequences. The environment of the brlA gene also was of interest to compare the sequence of the region with the existing genetic map (Clutterbuck, 1997) as well as for comparison with the sequences in the EST Unigene database.

The sequence of the 38,807 basepair cosmid, W06E08, was completed (Table 16) and, after annotation using the gene-finding program, GeneMark (Borodovsky, 1994) coordinated with the results of a BlastX (Altschul, 1990) search, was submitted to GenBank (accession number AC000133). A summary of this work was published (Kupfer, 1997). However, when the EST database became available, the cosmid sequence was examined for homology to these ESTs. Subsequently two clones, c5h06 and o4h01, were found to align with the brlA gene region. Included here is a summary taken from the published work as well as an additional study done on the two brlA cDNA clones.

A GeneMark examination of the W06E08 sequence suggested thirteen ORFs. Eight of the ORFs had homologs in the GenBank non-redundant database. The five others had coding potential as revealed by examination using the *A. nidulans*-specific GeneMark matrix E.nidulans_mat3 (figure 20). Thus, the gene density for this cosmid was calculated to be approximately 1/ 2.9 Kbp.

Table 16. DNA sequencing summary for *A. nidulans* cosmid W06E08

A. nidulans strain:	FGSC4	Cosmid Vector:	pWE15
Chromsome source:	VIII	Subcloning Vector:	pUC18

Total thermocycling reactions:	980
Forward:	490
Reverse :	490
Gel readings in database:	851
Gel readings in contiguous sequence:	842
Forward:	475
Reverse:	367
Unpassed gel readings:	
(pUC vector or partial pUC)	51
E. coli	9
Insert size for cosmid W06E08:	38878 bp

3.22 ORFs and Exon Identification

The following is an expanded discussion of the homologies found to each of the predicted open reading frames of W06E08 also presented in Figure 20 and Table 17.

The *opa* ORF showed homology to a number of proteins with CAX repeats that encode glutamine residues. Because it was possible that this was a spurious match due to the presence of repeated DNA the glutamine residues in the translated sequence were masked and the sequence again was analyzed by a Blast search. However, a similar set of BLAST matches was found that included the mouse *mopa* gene as the best match homolog. Gerber et al.(Gerber, 1994) found that over 80% of the genes encoding a (CAX)n repeat were transcription factors and that the glutamine repeats reside within the trans-activating domain. This (CAX)n motif, also called an opa repeat, was first identified in homeotic genes of Drosophila (Schneuwly, 1986) and subsequently have been found in developmentally regulated proteins such as notch, engrailed and bithorax (Wharton, 1985). The *A. nidulans* homolog contains one 16 residue glutamine repeat as well as a 19 glutamine interrupted repeat and a total of 137 glutamine residues in the 483 amino acid polypeptide. This glutamine signature suggests that the *opa* ORF most likely encodes a polypeptide involved in transcription activation, and is related to genes primarily known in metazoans.

The pot ORF was flanked by 47 bp perfect inverted repeats and showed homology to three fungal transposases of the mariner class first identified in *Drosophila* (Farman, 1966). This is the first time the presence of a transposon-like element has been detected in *A. nidulans* and a recent finding suggests that additional copies are present in the genome (Li, M. G., Nicosia, D., Scazzocchio, C., personal communication) (see also discussion of cosmid sequence W02H02-W30B01 below).

The brlA ORF corresponds to bristle, a transcription factor that is necessary and sufficient for asexual development in *A. nidulans* (Adams, 1988; Prade, 1993) that had been previously reported (Prade, 1993). The gene region contained two overlapping Fig

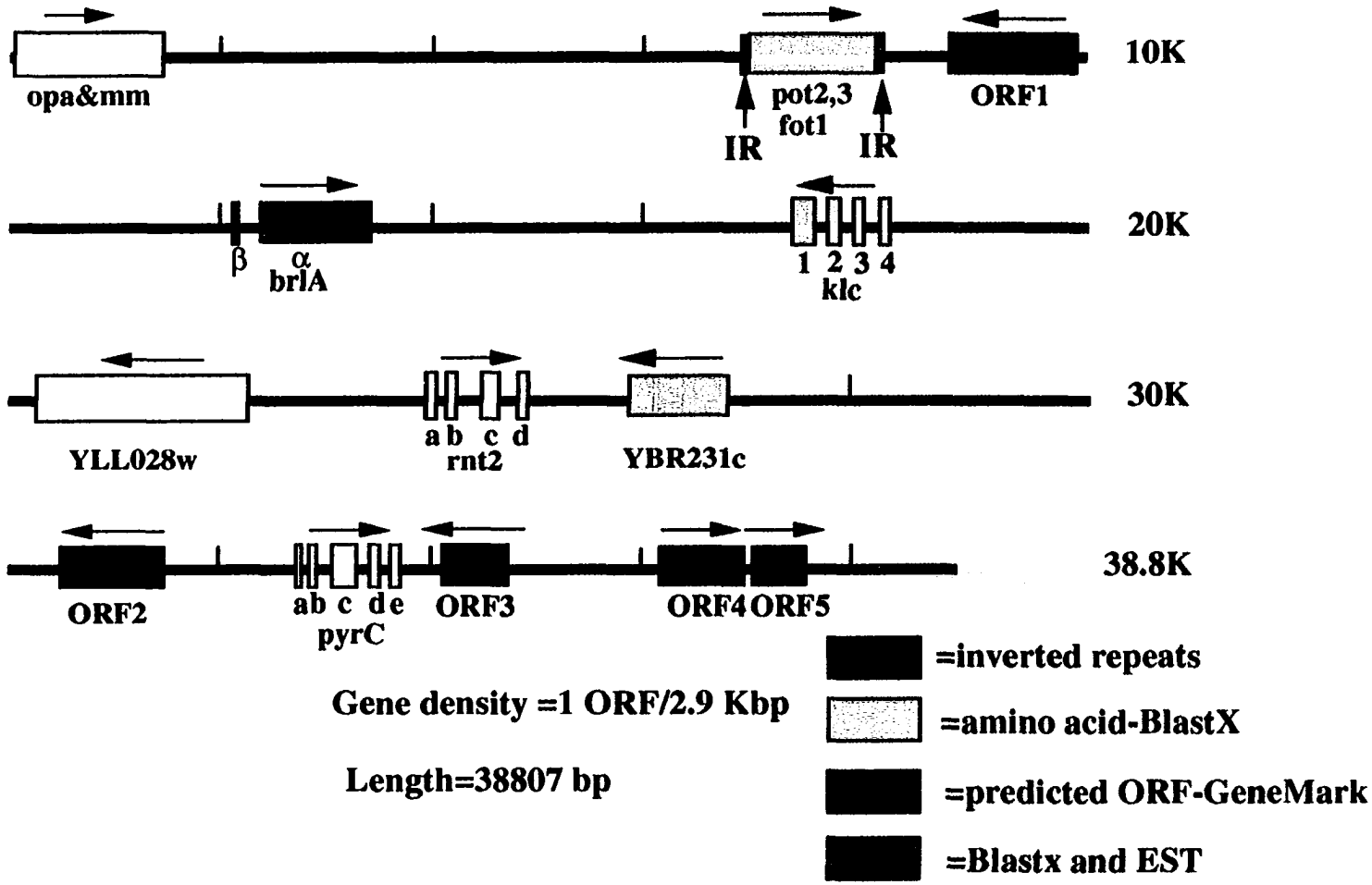


Figure 20. The ORF map of *Aspergillus nidulans* cosmid W06E08. Bars represent open reading frames or exons. Arrows indicate direction of transcription.

Table 18. Annotation of *A. nidulans* cosmid W06E08.

Significant Sequence Homology and Accession Number	% Similarity	Translation End Points	Size (a. a.)	Comments
mopa, differentially expressed Murine mRNA (M16362) mastermind-*Drosophila viridis* (M92914)	51 47	2>1453	483	diff. expressed in fetal and adult tissue possible transcription factor.
pot3, transposase, *M. griseq* (U60989) *pot2*, putative transposase, *M. gridea* (Z33638) *fot1*, transposon, *Fusarium oxysporum* (X64799)	62.7 49.3 45	6669>8335	555	*A.nidulans* inverted repeats (6575-6622 and 8397-8443)
brlA, regulator of conidiation, *A. nidulans* (L25858), cDNAs:c5h06, o4h01	100	*a-11176>11302 b-12315>13615	432	Transcription factor.
klc, kinesin light chain, *Loligo paelii* (squid) (L24440)	49.2 44.3 43.5 45.7	*1-17640<17950 2-17296<17702 3-16946<17308 4-16712<16926	103 136 124 71	4 ORFs with homology to *KLC*, coiled-coil and imperfect repeat region. Potential pseudogenes.
YLL028w, *S. cerevisiae* (Z73133) cyhr, *Candida maltosa* (cyhr_canma)	64.7 55.8	20577<22107	510	ORF of unknown function Cycloheximide resistant protein
rnt2, ribonuclease T2, *A. oryzae* (P10281)	73.4	a-24269>24520 b-24571>24780 c-24815>25135 d-25203>25255	277	fungal ribonuclease of T2/RH/M family, conserved cys residues.
YBR231c, *S. cerevisiae* (S46107)	51.1	25850<27008	386	ORF of unknown function.

Table 18R.continued

pyrC, dihydroorotase, *Ustilago maydis* (X63181) 68.2	a-32883>33167 b-32994>33087 c-33178>33633 d-33685>34163	450	third step in pyrimidine biosynthesis.
ORF1-No significant homology	9163<9826	220	
ORF2-No significant homology	36479>37069	197	
ORF3-No significant homology	37090>37737	216	
ORF4-No significant homology	34206>34878	223	
ORF5-No significant homology	31062<32134	356	

*-letters indicate exons and numbers indicate pseudogenes.

transcription units, α and β, one intron, and one ORF of unknown function)mORF) (Prade, 1993). The two transcription units result in two forms of the protein, the α form, which predominates during vegetative growth and the β form, which is the predominant species during asexual development. The β form differs from the α form by the addition of 13 amino acids at the 5' end.

A BlastX examination of this region with the EST Unigene database showed two EST clones, c5h06 and o4h01 which had homology to the brlA gene. Each clone therefore was resequenced from both the 3' and 5' ends, and the sequences were assembled using Phrap to improve the accuracy of the cDNA sequences. Although neither clone contained the full length cDNA and thus the amino acid alignment was incomplete at the 5' end, the sequences that were present matched the genomic sequence with 100% identity. Because of the incomplete 5' end, it could not be determined whether the α or the β form was represented by these EST clones.

The klc region of this cosmid was a complex site consisting of four ORFs each with homology to a coiled-coiled motif and four 42-residue imperfect tandem repeats which are present in all kinesin light chain proteins (Wedaman, 1993). Kinesin light chains associate with heavy chains and, when coupled with ATP hydrolysis, move vesicles and organelles unidirectionally along microtubules. Since two of the klc ORFs overlap in different reading frames (ORFs 3 and 4 in Table 18) and a third region (ORF2) contains a stop codon, these three ORFs most likely are pseudogenes. ORF1 remains a potential gene coding region for kinesin light chain however, no ESTs had homology with this region to support this interpretation.

The drt region had its highest BlastX homology with a *S. cerevisiae* ORF, YLL028w, that encoded a product of unknown function. However, significant HSP scores also were found for a number of integral membrane drug resistance translocase genes, including *Candida maltosa* shown in Table 18. The highest alignment scores were with transmembrane proteins, where eleven of the twelve found in the *C. maltosa*

cycloheximide resistant protein (Sasnauskas, 1992) are present in the drt ORF. The hydrophobic residues characteristic of a transmembrane region are conserved in the drt ORF, suggesting that it most likely is an integral membrane protein belonging to the major facilitator drug resistance translocase family.

The rnt2 region encoded a polypeptide with significant amino acid sequence similarity and conservation of cysteine residues involved in disulfide bonds characteristic of the *A. oryzae* ribonuclease rnt. This ribonuclease belongs to the fungal T2/H/M RNase family which preferentially cleaves 3' to A residues (Kawata, 1988).

The next ORF had homology to YBR231c, an *S. cerevisiae* ORF with unknown function. Since no other homolog was observed in the GenBank, no potential function could be assigned to the protein encoded by this ORF.

The adjacent region had homology to the *Usilago maydis* pyrC gene which produces dihydroorotase, the enzyme catalyzing the third step in pyrimidine biosynthesis, i.e. the closure of the pyrimidine ring in a dehydration reaction. The location of this gene downstream of the brlA gene corrected the relative positions of these two genes on the *A. nidulans* genetic map (Clutterbuck, 1997).

Finally, none of the five remaining ORFs that could encode proteins longer than 190 amino acid residues, had any significant homology to any entries in the GenBank databases.

3.23 Cosmids W02H02 and W30B01 sequence analysis

Cosmid W02H02 and W30B01 are overlapping cosmids from Chromosome VIII of *A. nidulans* that were of interest because hybridization studies revealed that they most likely contained genes involved in spermidine biosynthesis (N. Keller, personal communication). These two cosmids were from the minimum tiled cosmid library created by Prade at al. (Prade, 1997) and mapped distal to brlA on the same arm of Chromosome VIII.

The polyamine, spermidine, is involved in an early step of the sterigmatocystin biosynthetic pathway. This pathway has been partially characterized and the region of Chromosome IV containing this gene cluster has been sequenced (Brown, 1996 and section 3.26). The related species *A. parasiticus* and *A. flavus* are responsible for enormous grain crop losses each year because they produce the related toxin, aflatoxin (Lee, 1992) in infected harvested grain. The region of *A. nidulans* Chromosome VIII containing the spermidine synthase gene, which catalyzes the fifth and last step in spermidine production from arginine and methionine was of interest for three reasons. First, several aflatoxin genes from both *A. parasiticus* and *A. flavus* have been studied and found to occur in a gene cluster similar to the *A. nidulans* sterigmatocystin cluster. It therefore was of interest to learn the precise position and sequence of any additional genes involved in producing the toxin. The second was to obtain the sequence of the encoded toxin genes for future mutational studies. The third was to determine the gene environment of the *A. nidulans* spermidine synthetase gene as a prelude to sequencing the similar region of *A. parasiticus* which recently began in the Roe laboratory.

These overlapping cosmids span 44,998 basepairs and subsequent analysis of the results of a BlastX search of the non-redundant GenBank databases and the Unigene database revealed the structural features discussed below.

Table 18 describes the number of reactions needed to complete the sequence of each of these cosmids to an error rate of fewer than one base in 10,000 as determined by Phrap version 98.

A slightly different approach was used to sequence these cosmids than that used for W06E08. The recent availability of inexpensive internal primers and programs for optimum primer selection (Materials and Methods) allowed a shift from shotgun (random) sequencing to a directed phase earlier in a sequencing project and more rapid completion of the project using fewer sequencing reactions. Table 18 lists those reactions containing the universal forward and reverse primers used in the shotgun phase of sequencing and

Table 18. DNA sequencing summary for *A. nidulans* cosmids W30B01 and W02H02.

A. nidulans strain: FGSC4 Cosmid Vector: pWE15
Chromsome source: VIII Subcloning Vector: pUC18

	W02H02	W30B01
Total thermocycling reactions:	830	641
Forward:	384	288
Reverse :	384	288
Internal Primer:	52	65
Gel readings in database:	786	554
Gel readings in contiguous sequence:	605	470
Forward:	245	189
Reverse:	240	220
Internal primer:	52	61
Unpassed gel readings (pUC vector or partial pUC):	155	60
E. coli:	26	9
low quality:	0	15
Final quality- in errors/10,000 bases	0.66	0.21
Insert size for cosmid	40554 bp	37312 bp
Accession number	AC005299	AC004395

the number of primers used in directed sequencing. Comparing the total number of sequencing reactions needed to finish the cosmids shows that 22 reactions/Kbp were required to complete W06E08 as compared to 19 reactions/Kbp for W02H02 and 15 reactions/Kbp for W30B01. A directed approach was used after 4x96 random samples were cycle sequenced for W02H02 and 3x96 samples for W30B01. These results indicate that adding a directed phase to the sequencing approach earlier resulted in fewer reactions being needed to obtain the finished sequence. Since there was a higher percent of *E. coli* and vector contamination in the W02H02 subclones than in W30B01, the data suggested that a shift to directed cycling after the same number of random samples as performed for W30B01 would result in fewer reactions to reach a similar error rate for W02H02 and the original database would have fewer contaminating sequences.

The cosmid sequences were aligned and extensive overlap found. Since 7685 base pairs are unique to W02H02 and 4443 base pairs are unique to W30B01, there is an overlap of 32,870 base-pairs and a combined length of 44998 base-pairs of Chromosome VIII. There was a single difference between the two cosmids in the overlap region where the G residue inserted in clone W02H02 position 21459, was missing from W30B01 position 13774. An example of a subclone sequence from each cosmid is shown in the Consed (Gordon, 1998) view of each region in Figure 21 where the arrow indicates the G residue. Each position was sequenced from three separate subclones verifying that both sequences were correct and that these are valid sequence differences between the two cosmid clones.

3.24 ORFs and Exon Identification

Table 19 and Figure 22 give a summary of the annotated features in the 45 Kbp region defined by the overlapping cosmids W02H02 and W30B01. Both BlastX and GeneMark (Borodovsky, 1994) were used in a combined extrinsic and intrinsic examination to determine these genomic features (Material and Methods). The features

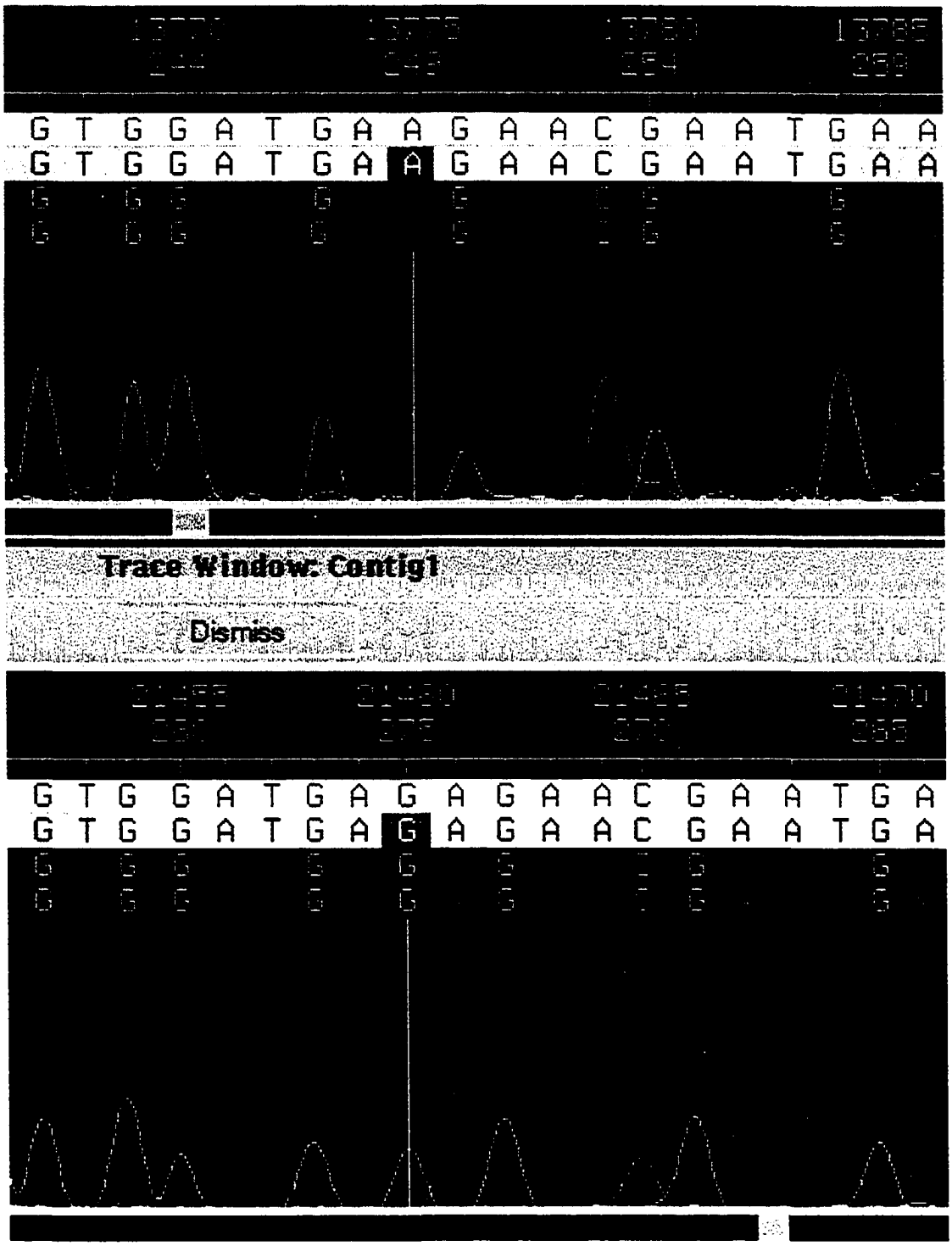


Figure 21. Comparison of two subclone sequences a2e12h2.f1 and a2a06b1.f1 from overlapping cosmids W30B01 (top panel) and W02H02 (bottom panel) showing an equivalent region in each. Cosmid W30B01 shows a deletion of a G residue between residues 13773 and 13774 (highlighted in red) which is present at postion 26459 (highlighted in red) in Cosmid W02H02.

Table 19. Annotation of the 45 Kbp region of ***Aspergillus nidulans*** chromosome VIII defined by overlapping cosmids pW02H02 and pW30B01. Probability is given for each exon when multiples are present. EST homology is for each cDNA aligning with the given cosmid region.

Gene Homolog/Accession #	Probability Scores HSP p value	Translation endpoints	Size in amino acids	cDNA	Unigene sequence endpoints	EST % Homology
1.Transposase, tc1-like z29098 *D. hydei*,w02h02 only	68 4e-10 68 4e-10	2987>3901 3945>4366	446	e0b11a1	11631, 12103	365/376 97%
2.Glucoamylase precursor p08640 *S. cerevisiae*, w02h02 only	144 2e-6	5895>8549	885	-		351/351 100%
3.yLR063w s61636, *S. cerevisiae*	182 2e-19	10716>11069	118	j4c04a1	10734, 11103	351/351 100%
4.Prohibitin-like p40961, *S. cerevisiae*	591 5e-95 390 3e-36	12074<12397 11498<12013	280	y3b11a1	12397, 11749	338/338 100%
5.CytoC oxidoreductase subunit VII u20790, *N. crassa*	70 1e-10	13833>13905	24	-		-
6.Spermidine synthase q09741, *S. pombe*	ND 176 9e-16 112 4e-6 264 7e-23 302 6e-27	14120>14150 14310>14490 14576>14724 14832>15051 15326>15604	320	h4a05a1	14331, 15957	303/306 98%
7.Transketolase q12630, *Kluyveromyces lactis*	1553 1e-159 1553 1e-159 1553 1e-159	18549<18602 18143<18285 18004<18088 15415<17934	934	y4a09a1 m0d06a1	18608, 15414	381/381 100% 381/381 100%

8.Acetate reg. DNA binding protein (FacB) u56097, *A. nidulans*	731 2e-72 413 1e-74 606 5e-59 606 5e-59 1986 3e-205	22655<23138 22282<22595 21943<22227 21537<21894 20324<21486	868	z4e02a1 20841, 20323	471/473 99%
9.spac3c7.01c z99568, *S. pombe*	none 123 4e-7 118 1e-6	26871<27161 26306<26776 25618<26229	458	-	
10.Thioredoxin-2 GTP binding protein u40843, p42942 *S. cerevisiae*	none 394 1e-36	29968<30060 29414<29632	40	-	
11.Expressed Gene 1	none none	30952>31065 31116>31409	136	d1d11a1 30882, 31485 i8e05a1 z1e12a1 m7b11a1 s3d05a1 z1a05a1 i3f04a1 c7g05a1	393/393 100% 361/362 99% 271/273 99% 524/525 99% 517/517 100% 468/469 99% 448/448 100% 369/369 100%
12.Expressed Gene 2	none	32285>33019	245	w7d08a1 32210, 32872 e7e03a1	182/184 98% 493/493 100%
13.High mobility group-like protein 2 p32495, *S. cerevisiae*	none 250 e-21	34052<34333 33532<33888	213	c5f04a1 34178, 33268	374/375 99%
14.Spindle assembly check-point MAD1 p40957, *S. cerevisiae*	281 5e-36	35602>37383	594	c5g07a1 35406, 37317	260/267 97%

15.yCR030 al021730, *S. cerevisiae*	164 2e-9	37755<38882	376	-
16.ORF1	none	39612<40355	248	-
17.ORF2 (w30b01 only)	none	42378<44555	726	-

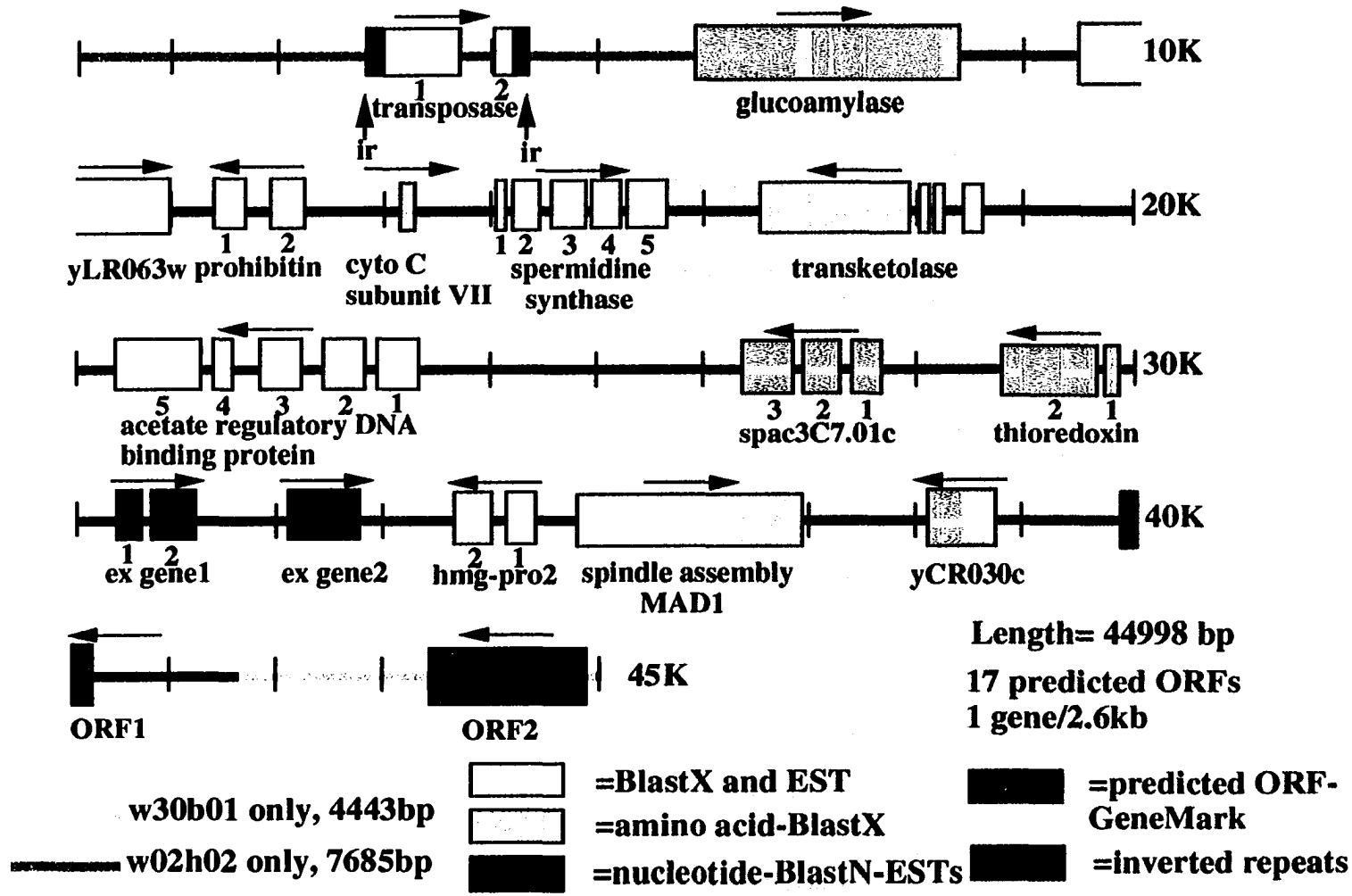


Figure22. ORF Map of *Aspergillus nidulans* Cosmids w30b01-w02h02. Bars indicated the cosmid sequence. Boxes indicate coding regions of genes and exons. Arrows indicate direction of transcription. Key defines the colors used for boxes and bars.

listed in Table 19 represent the BlastX homologies with the highest HSP and best probability scores. The importance of the EST database for genomic annotation was confirmed by a BlastN search showing a number of homologs from the *A. nidulans* EST database to the 45 Kbp W30B1-W02H02 region.

These matches are listed in Table 20 along with the percent of the individual EST homology. The corresponding Unigene member also was identified and, in all cases, additional sequencing of at least one of the clones was completed when the 3' and 5' sequences did not overlap each other in the Unigene database. Figure 22 displays the annotated region covered by the two overlapping cosmids with the predicted ORFs and their proposed gene assignment. The overlapping cosmid annotation is discussed below.

ORF1 had significant homology to a transposable element from *Drosophila hydei* which is a member of the Tc1-like transposons first characterized in *Caenorhabditis elegans* (Franz, 1992; Rosenzweig, 1983). The *D. hydei* transposable element, *Minos*, has perfect 255 nucleotide inverted repeats and two non-overlapping open reading frames which code for a 346 amino acid protein (Franz, 1992). Analysis of ORF1 revealed that perfect 244 nucleotide inverted repeats were present at the ends of two adjacent open reading frames with an ATG at the 5' end of the first, and the expected splice sites at the proposed intron/exon borders (Table 21). These two exons coded for a protein with 446 amino acids. The GCG Pileup alignment (Figure 23) of *minos* and the proposed transposon showed a 37% identity between the two amino acid sequences. There was conservation of uncharged residues, primarily in the 5' end of the *A. nidulans* protein which may be important for correct protein folding. This is the first report of a transposase of this class in *A. nidulans*.

A comparison of the tc-1 homolog nucleotide sequence with the Unigene database revealed a cDNA with high homology to the second exon. The 3' and 5' EST sequences of clone e0b11a1 overlapped to yield a single sequence whose GCG Bestfit alignment with the transposon is shown in figure 24. Mismatches between single pass EST

```
Percent Similarity: 36.986   Percent Identity: 30.822

                    | = IDENTITY
                    : = Similarity

          1                                                  50
   Minos  MSQYSMQKNF RLLQISRSLA TMVRGKPISK EIRVLIRDYF KSGKTLTEIS
                                 | ||      |:||:       |    :||
Transpos  ---------- ---------- -MPRGGFHPV ELRVQVLTLS AIGFSTEKIS

          51                                                 100
   Minos  KQLNLPKSSV HGVIQIFKKN G......... ..NIENNIAN RGRTSAITPR
          | |||   .|   :       |                      ||   ||
Transpos  KSLNLSPRTV QSIVKKGRDR GYRPEVSLRV QLEFVEDRKR SGRPVEITEA

          101                                                150
   Minos  DKRQLAKIVK ADRRQSLRNL ASKWSQQLAK LSSESGRDKL KSIGYGFYKA
               .  |  |||    |    |      |  :|  |    |  | |    |
Transpos  TQNTVITSVT ADR..AGREK LSEILAYEAG ISHSSVLCIL HSHGFVIAKP

          151                                                200
   Minos  KEKPLLTLRQ KKKRLQWARE RMSWTQRQWD TIIFSDEAKF DVSVGDTRKR
            || ||      :|         ||   |   :|| ||     .       |
Transpos  SWKPGLTEAA CLRRLEFCLA HQHWTLEDWK RVIFTDETGV ILGHRRGAIR

          201                                                250
   Minos  VIRKRSETYH KDCLKRTTKF PASTMVWGCM SAKGLGKLHF IEGTVNAEKY
          | |   :.   : | :|  |      |||||  |    | ||   :    |
Transpos  VWRTVKDSHT RNCVRRRWKA CSDFMVWGCF SYNKKGPLHI YKPETAAMRK

          251                                                300
   Minos  INILQDSLLP SIPKLLDCGE FTFQQDGASS HTAKRTKNWL QYNQMEVLDW
             :            |  |            |             |  |
Transpos  QADIEIEAMN RELEPLCREE WELATGLSRV HLRPNRGRVP KWNWNEKNGK

          301                                                350
   Minos  PSNSPDLSPI ENIWWLMKNQ LRNEPQRNIS DLKIKLQEMW DSISQEHCKN
             :                  :       :   | |          |
Transpos  LIRKGKGGID WWRYQTVCSL ISIILYYRLK PLLIPFAKEC .MIERPNTIV

          351                                                400
   Minos  LLSSMPKRVK CVMQAKGDVT QF--------- ---------- ----------
          |  | |
Transpos  LEDSAPAHCH RIQQHVYKAE DVQKILDWPG NSPDLNAIEP CWAWMKKRTT

          401                                                450
   Minos  ---------- ---------- ---------- ---------- ----------
Transpos  SRGAPRDKKT GEAEWRQAWA DLPQETIQHW IERLICHIQI VIELEGGNEY

          451              469
   Minos  ---------- ----------
Transpos  KEGREDRDTR SWAGRRIKG
```

Figure 23. The Pileup alignment of the *A. nidulans* Tc-1 element and the *Drosophila Minos* transposon. Highlighted bases are conserved uncharged residues.

```
           Length:    140
             Gaps:      0
Percent Similarity: 92.143   Percent Identity: 90.714

                  | = IDENTITY
                  : =   2
                  . =   1

Transposon.P x E0b11fr.P

                    .          .          .          .          .
    306 LKPLLIPFAKECMIERPNTIVLEDSAPAHCHRIQQHVYKAEDVQKILDWP 355
        |||||| ||||||||| |||:|||||||||||.||||:|||||.|||||||
      2 LKPLLILFAKECMIERLNTIILEDSAPAHCHQIQQHIYKAEDMQKILDWP 51
                .          .          .          .          .
    356 GNSPDLNAIEPCWAWMKKRTTSRGAPRDKKTGEAEWRQAWADLPQETIQH 405
        || ||||||.|||||||| | || ||||||||| | ||||||||||||||
     52 GNLPDLNAIKPCWAWMKKHTISRSAPRDKKTGEVECRQAWADLPQETIQH 101
                .          .          .          .
    406 WIERLICHIQIVIELEGGNEYKEGREDRDTRSWAGRRIKG 445
        |||||||||||||||||||||||||||||||||||||||||
    102 WIERLICHIQIVIELEGGNEYKEGREDRDTRSWAGRRIKG 141
```

Figure 24. The Bestfit sequence alignment of the *Aspergillus nidulans* tc-1-like transposon and cDNA clone e0b11a1 translation product showing strong homology with the second exon of the genomic transposon with the overlapping EST sequences of e0b11a1.

sequences and a genomic sequence are not unexpected. As can be seen in Figure 24, the two sequences have significant homology, this genomic region most likely encodes an expressed tc-1-like transposase.

This result is of great interest since there has been no evidence given in the literature for an active transposon in *A. nidulans*. An alignment with the fot1 class transposon discovered in cosmid W06E08 showed no sequence homology to this tc-1 class *A. nidulans* transposon confirming that these are representatives of two distinct classes of transposons, and that the tc-1 homolog appears to be actively transcribed. Interestingly, although Fot1 homologs has been found in related ascomycetes (Nyyssonen, 1996), homologs for the second, tc-1, have been detected only in higher eucaryotes (Merriman, 1995)

ORF2 had a significant homology with a glucoamylase of *S. cerevisiae*. This is a secreted enzyme which attacks both α-1,4 and α-1,6 linkages of starch to yield glucose and is used commercially to produce high glucose syrup (Carlile, 1994). The alignment of the *A. nidulans* and *S. cerevisiae* glucoamylase proteins shows a conservation of the serine/threonine residues in the region from 210-572 of the yeast glucoamylase and an overall sequence similarity of 31% (Figure 25). Thus, this open reading frame encodes a possible glucoamylase that had no corresponding ESTs in the *A. nidulans* Unigene database. This was not unexpected since, at the time of mRNA harvest for the library, glucose was the sole carbon source and glucoamylase is generally an inducible enzyme.

ORF3 showed homology to yLR063w, a *S. cerevisiae* open reading frame with unknown function. ORF3 also had significant homology to the EST clone j4c04a1. Eighteen bases at the 5' end of the EST clone are missing but the otherwise perfect nucleotide match indicates no introns in ORF3 nucleotide sequence and verifies this region as containing an actively transcribed gene.

ORF4 has significant homology to an*S. cerevisiae* prohibitin, a protein with anti-proliferative activity (Berger, 1998) that is important in determining the length of the cell

```
Identities = 72/372 (19%), Positives = 121/372 (32%), Frame = +3

cosmid: 5754  GGEVTLVWRSSDVRSLHQLKTTSESSTSTSSTSVSSTTTSSTSQT*RMASISLPLQPQTA 5933
              GG  +    SS   S     +TS  SST+TSSTS  SSTTTSSTS++   +S + P  P T
S. cere: 206  GGTKSSTTTSSTSESSTTTSSTSESSTTTSSTSESSTTTSSTSESSTSSSTTAPATPTTT 265

cosmid: 5934  -------PSPPHYFDSADPAPSGHRRNLTMSTPLPNPPFVFPARDHDEPDKQDLDTTTNGR 6095
                    P+PP          P+    + T  T                  K     T   +
S. cere: 266  SCTKEKPTPPTTTSCTKEKPTPPHHDTTPCTK----------------KKTTTSKTCTKK 309

cosmid: 6096  PPLPAFSFNPGSVGSNQAPAPAPSNSRM-SGHRRQYSEFVGGDQLIIPGNTAAGQTSDET 6272
                 P  + +  +  S+ AP P PS+S   S       S          +P  +++   S
S. cere: 310  TTTPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTESSSA 369

cosmid: 6273  PMASSTTVSSSVFRWSSTTXXSSQDLDTISSVDLTAISNALDLKPYVESAPCTSADMTR 6452
              P+ SSTT SSS    SSTT  SS    T SS   ++ S          SAP TS+
S. cere: 370  PVTSSTTESSSAPVTSSTTESSSAPVPTPSSSTTESSSAVTSSTTESSSAPVTSSTTES 429

cosmid: 6453  ERRP-----SLKPSQLLPHSATVLSRPTPPASPQLNLNEASPSSQIPKNERLENPRLCTPT 6620
                 P    + E S     S+T  S   P  +P  +  E+S +          +  + TP+
S. cere: 430  SSAPVTSSTTESSSAPVTSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVPTPS 489

cosmid: 6621  SFLAPESQNATRSSVLARKSDTAITSPKAESSAASQQPRPRPRTADA-SLMLELSGSTMT 6797
              S    ES +A  +S     S   + +P + ++ +S  P P P ++   S    ++ ST
S. cere: 490  SSTT-ESSSAPVTSSTTESSSAPVPTPSSSTTESSSAPAPTPSSSTTESSSAPVTSSTTE 548

cosmid: 6798  DNSSPTKRPNSAAGHSRSHKSMSS 6869
               +S+P   P+S+   S S    SS
S. cere: 549  SSSAPVPTPSSSTTESSSTPVTSS 572
```

Figure 25. Comparison of the *A. nidulans* glucoamylase homolog with *S. cerevisiae* AC# P08640 showing the conservation of threonine and serine residues (in bold) in a thr/ser signature region.

cycle (Coates, 1997). In yeast, the accumulation of the prohibitin protein has been associated with a shift to senescence, and in humans, prohibitin has been characterized as an antitumor factor (Coates, 1997). The *A. nidulans* protein encoded by ORF4 has 70% identity with the yeast prohibitin protein, (GenBank Accession number p40961). EST clone y3b11a1 included the entire mRNA sequence. Aligning the consensus Unigene sequence with the genomic sequence defined an intron, and matched with 100% accuracy the first exon and 98% the second. A 280 amino acid protein which was similar in size to the 284 amino acid yeast protein was predicted.

The region from 13834-13905 defined a small ORF which GenMark indicated as having coding potential. There was a single match in the GenBank databases which had low homology to the *N. crassa* subunit VII of cytochrome C oxidoreductase (Lobo-Hajdu, 1996). There were no Unigene sequence homologies to verify that this region was expressed, so no additional conclusions could be drawn.

Since both cosmids previously had hybridized to a spermidine synthase probe (Nancy Keller, personal communication) it was expected that the region coding for this gene would be detected. Therefore, it was not surprising when ORF6 had homology with an *Schizosaccharomyces pombe* spermidine synthase in GenBank. The EST clone, h4a05a1, also aligned with this region and the sequence revealed four of the five predicted introns. However, this EST clone did not include the extreme 5' end where the first small exon is located. GeneMark did not predict this exon and it only was determined based on the alignment of the translated products of the *S. cerevisiae* and *S. pombe* spermidine synthases. Figure 26 shows that this 10 amino acid-encoding exon contains the methionine start codon and had a sequence that was conserved between both the spermidine synthase of *S. cerevisiae* (accession number Q09741) and that of *S.pombe* (accession number Q12074).

ORF7 had a very significant level of BlastX homology to the 5' region of the transketolase of the single-celled ascomycete, *Kluyveromyces*. This enzyme catalyzes

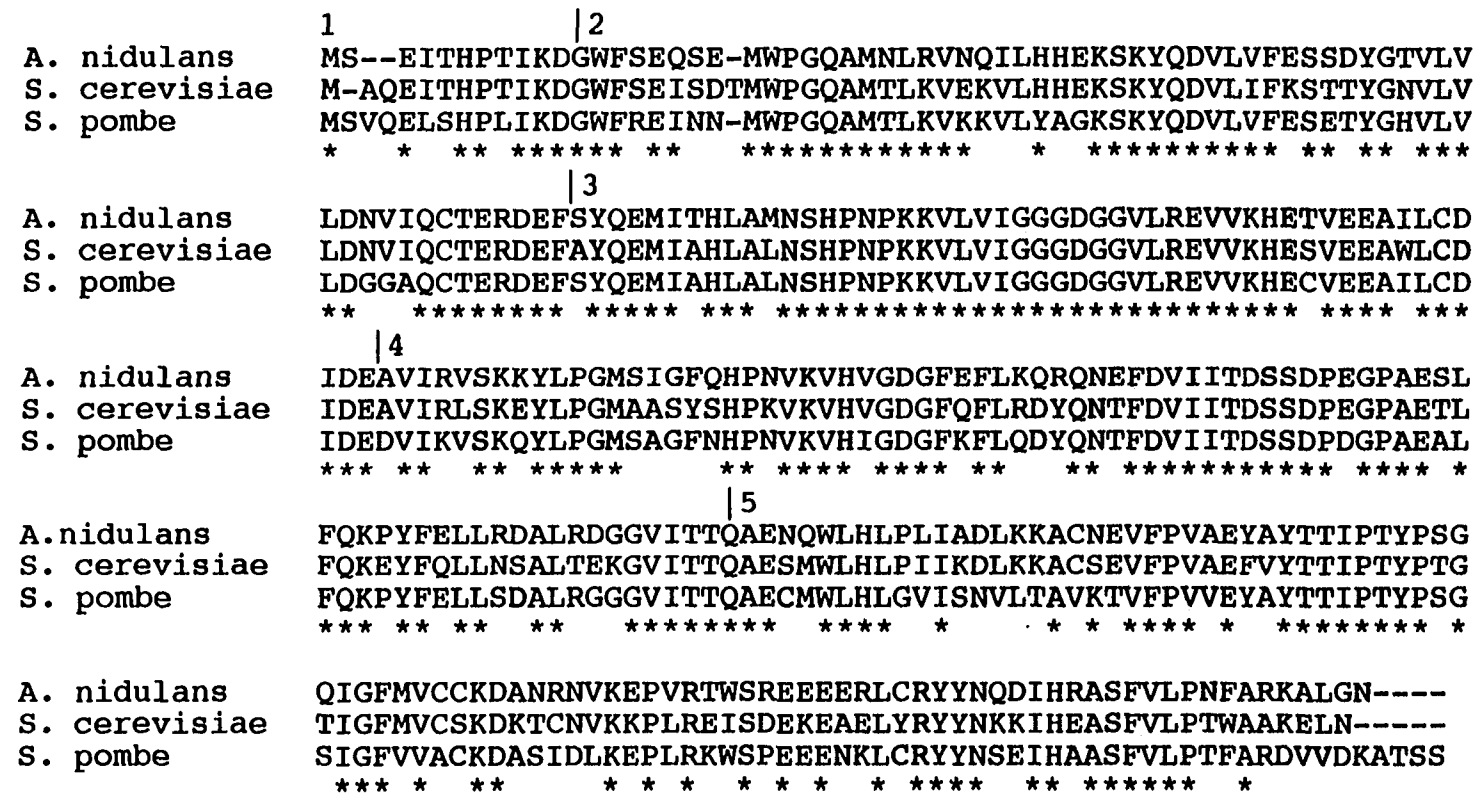


Figure 26. GCG Pileup alignment of the spermidine synthase protein products from three ascomycetes showing 193/240 or 80% conserved amino acids. The numbers and bars (|) indicate the five exons in *A. nidulans*. Exon 1 was determined by the alignment of the three orthologs of spermidine synthase. (*) indicate conserved amino acids.

the step of the pentose phosphate athway that links this pathway to glycolysis and provides four carbon sugars. Two cDNA clones, y4a09a1 and m0d06a1, aligned with this cosmid region. The h4a05a1 ESTs also aligned with the ESTs from these two clones in the Unigene database. However, it was apparent that this was not an alignment of clones from the same gene but rather an alignment of clones from overlapping convergent transcription units because the 3' sequences overlapped but in the incorrect orientation and the 5' EST sequences did not align (Figure 27). Further examination revealed that there was a 23 basepair overlap between the two regions in the 3' untranslated regions of both genes (Table 19).

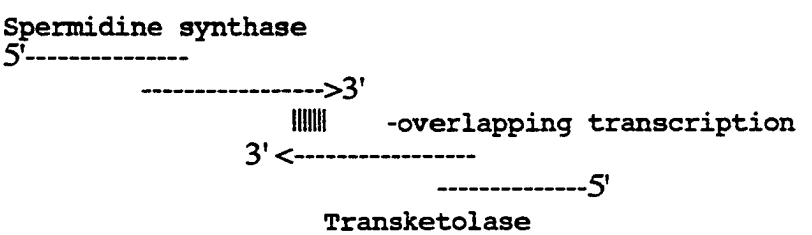


Figure 27. Alignment of the ESTs from clones y4a09a1 and m0d06a1 for transketolase and h4a05a1 from spermidine synthesis showing overlapping convergent transcription units.

Alignment of the Unigene consensus sequence containing the y4a09a1 and m0d06a1 ESTs with the *A. nidulans* cosmid sequences revealed three introns, three small exons, and one large exon (Table 19). The sequence alignment with the *Kluyveromyces* transketolase had 61% identity (Figure 28) but only in the region of the large 3' exon from cosmid nucleotide position 17,500 to 16,121. This region when analyzed with Blocks (Henikoff, 1991), revealed homology to three of six transketolase specific domains (block BL00801d-f). Since the remaining three domains were contained in the 5' end of the *Kluyveromyces* transketolase and the 5' region of the *A. nidulans* gene showed no homology to any GenBank entries and has no Blocks matches, it appeared

```
 sp|Q12630|TKT1_KLULA TRANSKETOLASE (TK) >gi|1488336 (U65983)
transketolase [Kluyveromyces lactis] Length = 679

 Score =  570 bits (1453), Expect = e-162
 Identities = 281/460 (61%), Positives = 353/460 (76%), Gaps = 2/460
 Frame = -1
 Query=2h2-30b1 subject=transketolase

cosmid: 17500 GDNDLEGIEAAIKEAQAVKDKPSMIRLTTTIGFGSKLQGTGGVHGNPLKADDIEGVKKRF 17321
              G++DL+ I  A+++A+ + D+P++I+LTTTIGFGS   G+  VHG PLKADD++ +K +F
Kluyv:  220   GNDDLDAISKALEQAK-LSDRPTLIKLTTTIGFGSLNAGSHSVHGAPLKADDVKQLKVKF 278

cosmid: 161320 GFDPAQSFVVPQQVYDLYHKHA-EEGAAQEQEWNQLLQKYAGEYPNEHADLTRRLSGKLP 17144
                GF+P +SFVVPQ+VYDLY+K   E G    ++W+ LL  Y G++P   A++ RRL+G+ P
Kluyv:  279    GFNPEESFVVPQEVYDLYNKSTIEPGIEANKQWDALLDAYVGQFPELGAEVKRRLAGEFP 338

cosmid: 161143 EGWEKSLPVYKPSDPAIASRKLSEAVLEKIHSVIPELLSGSADLTGSNNTRWKNAVDFQP 16964
                EGWE  LP Y P D A+ASRKLSE VL+ +   +PELL GSADLT SN TR K AVDFQP
Kluyv:  339    EGWESKLPTYTPEDSAVASRKLSEIVLDNVFDTLPELLGGSADLTPSNLTRSKGAVDFQP 398

cosmid: 16963  PEYGIGEWSGRYLRYGVREHAMAAIMNGLAAYGTVI-PAAGTFLNFVSYAAGAVRLSALS 16787
               P  G+G++SGRY+RYGVREH M AIMNG++A+G    P  GTFLNFVSYA+GAVRLSALS
Kluyv:  399    PITGLGDYSGRYIRYGVREHGMGAIMNGISAFGANYRPYGGTFLNFVSYASGAVRLSALS 458

cosmid: 16786  RVRAIHVATHDSIGLGEDGPTHQPIETLAHFRALPNCMVWRPADGNETSAAYYSALTAKH 16607
                  I VATHDSIGLGEDGPTHQPIETLAHFRA+PN  VWRPADGNE +AAY  ALT KH
Kluyv:  459    GHPVIWVATHDSIGLGEDGPTHQPIETLAHFRAIPNLQVWRPADGNEVTAAYKVALTNKH 518

cosmid: 16606  TPSILALTRQNLPQLENSSIEAALKGAYPVFEAADAKVTIISTGSEVSICIDAAKYLQEK 16427
               TP+I+AL+RQNLPQL+ SS+E A+KG Y + +     + I+STGSEV I ++AAK L EK
Kluyv:  519    TPAIIALSRQNLPQLQGSSVEKAVKGGYILQDVDQPDLAIVSTGSEVGIAVEAAKVLAEK 578

cosmid: 16426  HGVVARVVSIPCFEVFDAQDKEYRLKVLPDGIPILSVEVCSTMGWERYSHEQFGLNRFGA 16247
               + + AR+VS+P F  F  Q KEY+L V PDG+PILSVEV +T GW +Y+H+ FGL+RFGA
Kluyv:  579    N-IKARIVSLPDFHSFGQQSKEYQLSVFPDGVPILSVEVLATSGWSKYAHQSFGLDRFGA 637

cosmid: 16246  SGPYKEVYAKFEFTPEGISKRALATIDFYKGVPVRSPINRAF 16121
                SG    VY KFEFTP+GI+ RA  T++FYKG  V SP+N AF
Kluyv:  638    SGKGPAVYEKFEFTPQGIATRAEKTVEFYKGKQVISPLNTAF 679
```

Figure 28. BlastX alignment of W30B01-W02H02 ORF7 with *Kluyveromyces* transketolase showing alignment only of the 3' end of the proteins beginning at amino acid 220 of this yeast homolog.

likely that the *A. nidulans* ORF7 did not code for a homolog of this transketolase but rather coded for a gene with a related function or another transketolse with a unique 5' domain.

The region designated ORF8 had significant homology to the *A. nidulans* acetate regulatory DNA binding protein (facB). FacB is a transcriptional activator involved in acetamide and acetate utilization (Todd, 1997). The nucleotide sequences of the cosmid ORF8 differed only slightly from the GenBank entry, showing two base substitutions, one deletion and three insertions, all in intronic regions. This may be due to strain differences as the *A. nidulans* strain which was the source of the U56097 sequence was not reported. Interestingly, this gene included the single base difference between cosmid W02H02 and W30B01 that was discussed above and which occurs in the 3' most exon of facB. The W02H02 cosmid sequence matches the GenBank entry at position 21,460. The lack of the G in cosmid W30B01 results in a frameshift within the last exon that affects a coding region for facB and suggests that the difference between the two cosmid sequences may have occurred during the growth and maintenance of the cosmids under conditions where there was no selective pressure to maintain this gene. Even so, this difference is curious, especially since both cosmids were constructed from the same *A. nidulans* strain and by the same researcher (Prade, 1997).

A single facB cDNA clone, z4e02a1, was present in the Unigene database which covered only the 3' most exon but did not include the deletion site. It was surprising to find a clone for facB in the cDNA library since the culture producing the library was grown on glucose as the sole carbon source. However, it is known that fungi often ferment sugars to ethanol, glycerol, acetate, butanediol and additional products (Carlile, 1994). Since the Unigene database contains a butanediol dehydrogenase, Contig 1693, involved in butanediol production from glucose, it may be that *A. nidulans* fermented at least part of the available glucose and thus this gene is expressed because its product is required for utilization of acetate.

ORF 9 had homology to a hypothetical *S. pombe* protein, spac3c7.01c. There appeared to be three exons in this *A. nidulans* gene that would encode for a protein with 458 amino acids. The two 3' most exons had homology with the spac3c7.01c ORF. Since no ESTs aligned with this ORF, additional studies would be needed to verify if this hypothetical gene is expressed.

ORF10 encodes a region with homology to *S. cerevisiae* thioredoxin-2, a reducing power carrier, which is involved in control of the cell cycle S phase (Muller, 1991). No ESTs were observed for this region but the alignment with the *S. cerevisiae* thioredoxin-2 homolog predicts two exons that would encode a 104 amino acid protein, which is the same size as that predicted for the *S. cerevisiae* gene.

ORF 11 was defined by eight cDNA clones and was matched with 100% identity by their Unigene consensus sequence. Since the Unigene consensus defined two exons and an apparent 5' end, a protein of 136 amino acids was predicted (Figure 29). In addition, GeneMark defined both exons with "good" coding potential. Although there are no homologs to this gene in the GenBank databases, analysis of the eight homologous cDNAs revealed that the entire coding region and the intron were defined by the Unigene consensus sequence aligning with this region that included an intron with consensus borders (Table 21). Even though this gene was a member of the abundantly expressed class of genes in the *A. nidulans* Unigene database there were no known homologs in the public databases and thus this represented a new gene encoding a 136 amino acid polypeptide with unknown function.

ORF 12 is similar to ORF11 in that GeneMark indicated a region with coding potential that aligned with two cDNA clones. Again, there were no homologs found by a BlastX search. The ORF12 region had a single open reading frame but only the 5' end was present in the Unigene consensus sequence. Thus, this represents a new gene that encodes a protein with 245 predicted amino acids defined by the alignment of the cDNA clones.

```
Query file(s):  Contig2h2-30b1.fasta (ORF11)
Subject file(s): Unigene Contig19.fasta

minmatch: 14, minscore: 30

Exon 1
 CONTIG30B1   23197 GTACCAACAAACCATTAACCTAACACCACAACTCTATCTCTGCATTCAAC 23246
 Contig19         1 GTACCAACAAACCATTAACCTAACACCACAACTCTATCTCTGCATTCAAC 50
                                5'-> M  S  D  S  T  F  H  T  T  I
 CONTIG30B1   23247 GAAACATATCTCTATCACAATGTCTGACAGCACCTTCCACACCACCATTC 23296
 Contig19        51 GAAACATATCTCTATCACAATGTCTGACAGCACCTTCCACACCACCATTC 100
                     Q  D  I  R  K  P  E  S  H  A  S  H  A  A  K  G  N
 CONTIG30B1   23297 AGGACATTCGCAAGCCAGAGTCTCACGCTTCCCATGCTGCTAAGGGCAAC 23346
 Contig19       101 AGGACATTCGCAAGCCAGAGTCTCACGCTTCCCATGCTGCTAAGGGCAAC 150
                      T  P  K  D  S  N  V  S  A  M  K
 CONTIG30B1   23347 ACTCCTAAGGATTCTAATGTCTCCGCAATGAAG 23379
 Contig19       151 ACTCCTAAGGATTCTAATGTCTCCGCAATGAAG 183


Exon 2
                        S  I  I  D  Q  N  T  D  K  Q  A  D  I  E  K  T
 CONTIG30B1   23430   TCCATTATCGACCAGAACACAGACAAGCAAGCCGACATCGAAAAGACC 23477
 Contig19       184   TCCATTATCGACCAGAACACAGACAAGCAAGCCGACATCGAAAAGACC 231
                      K  A  N  L  P  L  P  D  Q  P  P  V  A  S  D  W  N
 CONTIG30B1   23478 AAGGCCAACCTGCCATTACCAGACCAGCCCCCTGTCGCTAGTGACTGGAA 23527
 Contig19       232 AAGGCCAACCTGCCATTACCAGACCAGCCCCCTGTCGCTAGTGACTGGAA 281
                        S  A  D  Q  R  A  V  N  V  G  S  G  G  I  E  G
 CONTIG30B1   23528 TTCCGCCGATCAGCGAGCTGTCAATGTTGGTTCCGGCGGTATCGAAGGAC 23577
 Contig19       282 TTCCGCCGATCAGCGAGCTGTCAATGTTGGTTCCGGCGGTATCGAAGGAC 331
                       P  I  S  G  E  N  N  S  A  L  R  G  P  A  T  A  S
 CONTIG30B1   23578 CCATCTCAGGCGAGAACAACTCTGCTCTCCGAGGTCCAGCCACAGCCTCA 23627
 Contig19       332 CCATCTCAGGCGAGAACAACTCTGCTCTCCGAGGTCCAGCCACAGCCTCA 381
                       S  S  A  R  E  V  G  E  E  T  H  R  N  T  Q  P  T
 CONTIG30B1   23628 AGCAGTGCTCGCGAGGTCGGAGAGGAGACGCACAGGAACACACAGCCAAC 23677
 Contig19       382 AGCAGTGCTCGCGAGGTCGGAGAGGAGACGCACAGGAACACACAGCCAAC 431
                          S  N  V  G  R  G  D  L  P  A  D  A  Q  A  R ->3'
 CONTIG30B1   23678 TAGCAATGTTGGTCGGGGAGACCTCCCTGCCGATGCTCAGGCTCGGTAAC 23727
 Contig19       432 TAGCAATGTTGGTCGGGGAGACCTCCCTGCCGATGCTCAGGCTCGGTAAC 481

 CONTIG30B1   23728 CCATGATTCTCATTGTTTGCAGCATAGCGATGTGATACGAACAAAACGAA 23777
 Contig19       482 CCATGATTCTCATTGTTTGCAGCATAGCGATGTGATACGAACAAAACGAA 531

 CONTIG30B1   23778 GACATAATGATGATTTCTCCATG 23800
 Contig19       532 GACATAATGATGATTTCTCCATG 554
```

Figure 29. Crossmatch comparison of the Unigene sequence Contig19 and a region of composite sequence of cosmids W30B01 and W02H02 showing 100% identity. This defines a new gene containing one exon with no homologs in the GenBank databases. The proposed tranlation start and stop are highlighted.

ORF13 had homology to a cDNA clone, c5f04a1, with a significant BlastX similarity to high mobility group-like nuclear protein2 (Kolodrubet, 1991). Although the function of this protein is unknown, it is essential in *S. cerevisiae* and had 62% identity with the *A. nidulans* ORF13. Comparison of the cDNA and genomic sequences revealed two exons that would encode a 213 amino acid protein.

ORF 14 is a homolog of the *S. cerevisiae* spindle assembly checkpoint gene MAD1 (Hardwick, 1995). The ESTs from clone c5g07a1 aligned with the large 3' exon but not with the smaller 5' exon. The BlastX alignment of ORF 14 with the *S. cerevisiae* MAD1 revealed that the *A. nidulans* homolog did not have the asparagine rich domain found in the smaller exon of the yeast homolog. In addition, the *A. nidulans* 5' exon defined by GeneMark does not align with the yeast homolog. The *A. nidulans* exons encoded a 594 amino acid protein, smaller in size than the 749 amino acid yeast homolog. This result was similar to that observed with the transketolase gene homolog reported above since the 5' end of the gene either coded for a domain unique to the *A. nidulans* homolog or defined a gene with a different function which shared a domain with MAD1.

The last two ORFs did not have any significant GenBank homologs as revealed by a BlastX search, had no homologs in the Unigene database, and had coding potenital suggested only by GeneMark. Thus it was unclear whether these ORFs actually encode their predicted proteins.

3.25 Cosmid W02H02-W30B01 Conclusion.

Seventeen ORFs were assigned to this 45 Kbp region of *A. nidulans* chromosome VIII by a combination of GeneMark, BlastX homology to the GenBank databases and by homology to the *A. nidulans* EST Unigene database. GeneMark ORF analysis was for all six frames, using a coding potential matrix specifically designed for *A. nidulans* based on codon usage. A BlastX search of the nonredundant protein database with the entire 45 Kbp region was examined and a BlastN search of the *A. nidulans* entries in the dbEST

also was performed. The results from GeneMark and BlastX gave support for the presence of a gene but neither was sufficient to allow more than a predicted, putative gene assignment to a region. However, the alignment of one or more ESTs could confirm that a region contained a transcribed gene. Based on these assumptions, the 45 Kbp W02H02-W30B01 region had 10 of the 17 genes predicted by GeneMark and/or BlastX verified as expressed genes by homology with one or more ESTs.

The difference between the number of ESTs which align with this region and to the SW06E08 cosmid region where only one in 13 predicted genes was aligned with an EST, is striking and it presumably was fortuitous that the W02H02-W30B01 region contained genes which were transcriptionally active at the time of library mRNA harvest.

3.26 The Sterigmtocystin Gene Cluster

Accession number U34740 represented a 60 Kbp region of Chromosome IV containing the 25 gene cluster for sterigmatocystin biosynthesis, a secondary metabolite produced by a complex pathway which includes a polyketide synthase (Figure 30). This fungal toxin is of extreme interest to agricultural scientists because of the enormous grain crop loss after infection by either *A. paraciticus* and *A. flavus* (Jimenez, 1991; Magnoli, 1998).

Polyketides, a structurally diverse class of secondary metabolites produced by fungi, bacteria and plants, are synthesized by polyketide synthase through a complex series of condensation reactions of small carboxylic acids (Katz, 1993). Sterigmatocystin and aflatoxin are related polyketide toxins produced by members of the *Aspergillus* family. These two toxins are among the most highly toxic and carcinogenic naturally produced products. *A. nidulans* does not produce aflatoxin because it lacks the last two enzymes in the pathway but does produce sterigmtocystin from which aflatoxin is derived. As Figure 30 shows, understanding of the pathway is incomplete and the

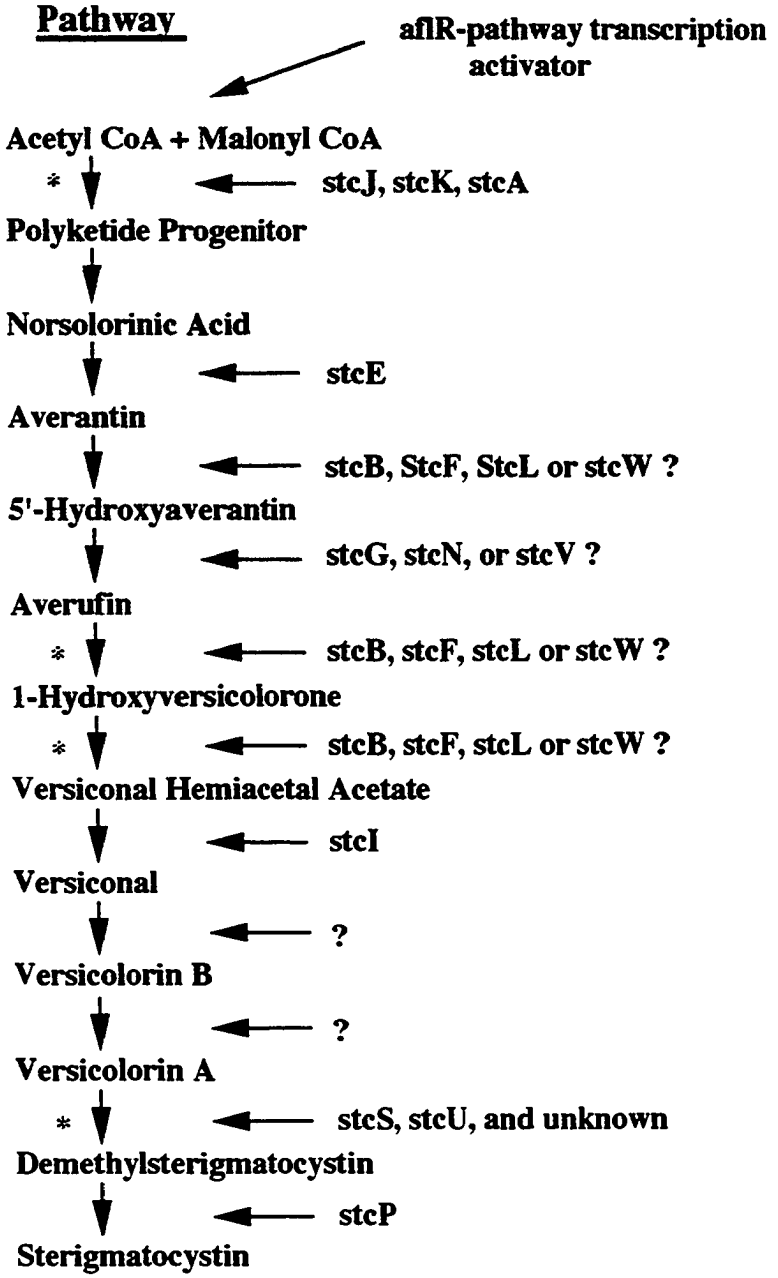


Figure 30. The proposed sterigmatocystin biosynthetic pathway. The cluster genes whose product is potentially responsible for particular steps are shown. * indicates a step for which no biochemical evidence has been found. ? indicates that the gene product involved in the step is not known or it is not clear which of several with similar activities are involved in a pathway step (adapted from Brown et al., 1996).

subject of current study (Chan, 1999; Fernandes, 1998; Guzman-de-Pena, 1998). The sterigmatocystin pathway of *A. nidulans* thus can serve as a model system to study regulation of aflatoxin production in other members of the *Aspergillus* family, including *A. parasiticus*, *A. flavus*, and *A. fumigatus* which do not have well studied genetic systems (Adams, 1998; Guzman-de-Pena, 1998) but have been shown to contain at least part of the aflatoxin gene cluster (Keller, 1997). Infection of crops such as corn by these organisms leads to large crop losses annually. For example, in Texas during the 1990 and 1995 growing seasons it was estimated that 50% and 30% respectively of the $400 million crop was lost due to unacceptable levels of aflatoxin contamination (Breaux, 1995). In addition, in low income nations, including regions of China and sub-Saharan Africa, consumption of infected grain leads to high incidents of liver cancer marked by a specific G-T transversion in the p53 tumor-suppressor gene (Bressac, 1991; Hsu, 1991). Finally, both *A. fumigatus* and *A. flavus* increasingly were reported as the causative agents of systemic fungal infections of immunocompromised patients (Iwen, 1997; Nenoff, 1996) where the possibility of low level exposure to aflatoxin increases the already serious nature of the disease.

In fungi, pathway genes for secondary metabolites are often clustered within a gene dense region as occurred for the genes belonging to the sterigmatocystin pathway. Recently, the region of *A. nidulans* chromosome IV containing the sterigmatocystin gene cluster was sequenced and revealed many of the members of the sterigmatocystin/aflatoxin pathway (Brown, 1996). Since this pathway was of considerable interest to health and agricultural researchers, the EST database was examined to determine if the cDNA library contained members derived from this cluster. Although, the RNA for this cDNA library was harvested at 24 hours into development, a time earlier than st/afl RNA had been detected by hybridization (Brown, 1996), EST sequences for members of this cluster were present in the database at relatively high levels. This allowed the comparison of those homologous ESTs from the Unigene

database with the *A. nidulans* sterigmatocystin gene cluster genomic sequence and resulted in several annotation corrections and the identification of an new cluster gene.

The genes in this cluster are named stcA through stcX in physical order arbitrarily from left to right for this region of the genome, except for the sixth gene in the cluster, aflR, the pathway transcriptional regulator (Brown, 1996). A number of the sterigmatocystin genes were represented by multiple cDNA clones in the EST database. Although three of the proposed ORFs in the cluster, stcD, stcO, and stcX, lacked EST homologs the remaining 21 proposed genes in the cluster were represented by between one and 22 cDNA clones. Interestingly, ORFs stcH, stcM, and stcR were assigned but the function or role of their gene products in the sterigmatocystin pathway were unknown and it was unclear if these genes were expressed. The presence of ESTs for these ORFs supported the idea that they were functional genes. The identification of homologous cDNAs indicated that stcH, M, Q and R were genes located in the cluster region with no identified function in the sterigmatocystin biosynthetic pathway but which were expressed at the same time as many of the cluster genes. This observation suggested that they indeed may be members of the sterig/afla pathway as originally proposed (Brown, 1996). A detailed discussion of the alignments of the ESTs and EST consensus sequences with homology to the sterigmatocystin region GenBank accession number U34740 (Brown, 1996) is included below where all references to genomic sequence were to the region described in this GenBank entry. Table 20 at the end of section 3.26 gives a summary of the finding of the Unigene database comparison to the stc cluster sequence.

stcA, the first gene in the cluster, encodes the polyketide synthase responsible for assembling the sterigmatocystin carbon skeleton (Katz, 1993). The consensus sequence obtained by the alignment of the three matching clones covered the 3' end of the gene. Although only two introns had been predicted, the EST consensus sequence reveals a third with conserved intron-exon borders, lengthening the predicted protein by 16 amino acids at the C-terminal.

stcB is a p450 monooxygenase homolog, which along with stcF, stcL, and stcW did not have an assigned role in the cluster. Two cDNAs with homology to stcB confirmed the presence of one of three predicted introns. A small region 14 Kbp to the right of stcB, between stcI and stcK at position 24,274-24368, also aligned with 93% identity to the stcB cDNAs and the stcB ORF. This indicated that at least a portion of the stcB gene might have been duplicated during the evolution of this gene cluster.

stcC, a predicted peroxidase, had homology to one cDNA. When aligned with the genomic sequence revealed a single large ORF as reported earlier (Brown, 1996).

stcE, a putative ketoreductase had homology to two EST clone pairs in the Unigene database. The Unigene consensus differed slightly from the genomic sequence by the insertion of an A residue at 14,899, a T residue at 14,189 and the deletion of a G residue at 14,893. These changes result in an additional intron and a product that differed by 5 amino acids and added 3 additional amino acids to the orginally predicted protein (figure 31). The final predicted protein would have 263 amino acids.

Additional investigation of the EST sequences which aligned with the stcE region showed an unexpected result. p0b05a1.r1, a 3' EST, could not be assembled with the three other ESTs aligning in this region even though it aligned with 99% identity to the stcE genomic region. When p0b05a1.r1 was compared with the EST consensus, the alignment revealed that p0b05a1.r1 contained the first intron and probably represented an incompletely spliced mRNA.

Another interesting observation was that the ESTs from cDNA, r1a06a1, aligned with the stcE Unigene consensus sequence but in the opposite orientation and was found instead to represent the 3' end of the adjacent aflR gene. The position of r1a06a1 showed that for these convergently transcribed genes, the 3' exon of aflR overlapped the 3'exon of stcE for a 36 basepair region (Figure 32). Overlapping 3' ends of convergently transcribed genes were also seen in the spermidine synthase and transketolase regions of W02H02-W30B01.

```
     Percent Similarity: 96.923   Percent Identity: 96.923
                    | = IDENTITY
                    : =   2
                    . =   1

stcE.EST x stcE.cosmid

                .         .         .         .         .
EST     6 MPSAAVSVPEVPSSDRKTVYLVTGASRGLGRGLVQAFLLRPNSIVIAGLR 55
          |||||||||||||||||||||||    .|   ||||||||||||||||||
Cosmid  1 MPSAAVSVPEVPSSDRKTVYLVNRCQQG...GLVQAFLLRPNSIVIAGLR 47
                .         .         .         .         .
     56 NRTSQAGALDALPRGENSSLIAVQLDSGSKSDPADAVSILQRDYGITHLD 105
        |||||||||||||||||||||||||||||||||||||||||||||||||||
     48 NRTSQAGALDALPRGENSSLIAVQLDSGSKSDPADAVSILQRDYGITHLD 97
                .         .         .         .         .
    106 VVIANAAIAANYGPASTMPLEYLETHMQINAYAALLLFQATRVLLQAAKS 155
        |||||||||||||||||||||||||||||||||||||||||||||||||||
     98 VVIANAAIAANYGPASTMPLEYLETHMQINAYAALLLFQATRVLLQAAKS 147
                .         .         .         .         .
    156 PQFICVGAPISTITEMESCARAPLTNYALSKLAACYLVRKIHFENKWLVA 205
        |||||||||||||||||||||||.|||||||||||||||||||||||||||
    148 PQFICVGAPISTITEMESCARAPVTNYALSKLAACYLVRKIHFENKWLVA 197
                .         .         .         .         .
    206 YIVDPGHIQSDMGAQAARLFGRKEAPTTIEESVAGICARMTEADKNTTSG 255
        ||||||||||||||..|||||||||||||||||||||||||||||||||||
    198 YIVDPGHIQSDMGARSARLFGRKEAPTTIEESVAGICARMTEADKNTTSG 247
                .
    256 RFILFSDGSDVPW 268
        |||||||||||||
    248 RFILFSDGSDVPW 260
```

Figure 31. A Bestfit alignment of the translation product of the stcE Unigene consensus (stcE.EST) with the U34740 stcE translation product (stcE.cosmid) showing the protein alignments.

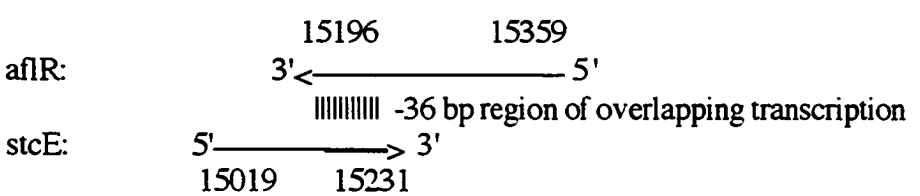


Figure 32. Region of overlapping convergent transcription between the aflR and stcE genes of the sterigmatocystin gene cluster. Overlap was defined by the homologous ESTs for these two regions, see text. Vertical lines indicate the overlap region. Numbers indicate exon endpoint positions in U34740.

Four cDNAs matched stcF, a predicted p450 monooxygenase, and confirmed the presence of its single intron but revealed its correct position at nucleotides 19946-19865, rather than 19996-19915 as reported (Brown, 1996). The entire coding region of stcF also was represented by the Unigene consensus sequence.

The EST analysis of stcG, a predicted dehydrogenase, revealed the position of its introns which were not determined by the earlier work and allowed prediction of a complete protein sequence. This 305 amino acid protein shared significant identity with the adhA gene from the *A. paraciticus* aflatoxin gene cluster that appeared to be regulated by nitrate and nitrite assimilation (Ehrlich, 1999). The Unigene consensus sequence differed from the cosmid sequence as two A residues occurred in the cosmid sequence at positions 21326-27 which introduced a stop codon, while the EST consensus had only one A residue which allowed the prediction of a larger open reading frame and the 305 amino acid protein with homology to the *A. paraciticus* homolog.

The ORF for stcH was not presented in Brown et al. (Brown, 1996). They originally mapped a 600 bp transcript to this region but could find no definitive ORF or any predicted protein with a homolog in GenBank (Brown, personal communication). The cDNA clone, h4b11a1, which mapped to stcH, was 480 bp in length and appeared to contain the complete coding region of stcH from U34740 position 22750-22372. The Unigene EST sequence differed from the genomic sequences by G deletions at position 22746 and at position 22632. Based on the EST sequence information, two introns with conserved intron-exon borders and an 83 amino acid protein which has no homolog in the databases was predicted.

ESTs from seven cDNA clones mapped to the stcI region and revealed a new intron and a new C-terminal exon which was not predicted previously. The translation of stcI thus was extended by the addition of eleven 5' amino acids from this additional 5' exon. The Unigene consensus allowed determination of the correct protein sequence as

well as indicating that this was an abundant transcript in the library.

ESTs from two cDNAs (Table 20) aligned with stcJ, a predicted alpha subunit of fatty acid synthase. The 3' third of the gene had 100% homology with the consensus sequence.

The g9b02a1 EST sequences had an exact match with the genomic stcK sequence which coded for a fatty acid synthase beta subunit. Although, the complete 5' end was not contained in the cDNA clone, the presence of two introns is confirmed.

The p450 monooxygenase coded for by stcL is represented by eight homologous cDNA clones which placed its representation in the abundant class of transcripts in the Unigene database. Their consensus sequence confirmed the one predicted intron.

stcM also was represented in the abundantly expressed class of the Unigene database by 15 clones whose ESTs showed homology to stcM. The consensus provided the entire coding region for this gene. This was another region that was not discussed in Brown et al. since they had no knowledge of the direction of transcription and the predicted ORF had no GenBank homolog. The presence of homologous ESTs predicted a gene with one intron that would encode a 148 amino acid protein that was transcribed in the same direction as is stcL. The Unigene consensus also differed from the published sequence in two positions and an intron was predicted in the 5' UTR.

stcN also was very highly represented in the EST database with 22 cDNA clones. The consensus sequence predicted two introns and had a 97 bp 3' overlap with the 3' end of the stcO gene. Interestingly, region 895-2024 of the stcN EST consensus sequence showed an exact match to a region between genes stcO and stcP, positions 41601-42612 of the genome cluster region, an apparently duplicated region within the sterigmatocystin cluster.

One cDNA showed homology to stcO. As mentioned above, this gene shared a 97 bp 3' overlap with stcN as predicted by Brown et al. and the stcN EST consensus. The 3' end of the stcO gene also showed an extensive perfect match with the same region

as stcN. No gene had been characterized in this duplication region that began at position 41601, one base from the 5' end of the stcO gene and extended to position 43449. The 5' end of the stcO EST did not align to the genomic stcO sequence but instead aligned, with one G-C transversion, to the second region. This suggested that the stcO region was not the gene defined by the matching cDNA but that the duplicated region containing an alternative ORF was more likely the transcribed region. This new region, stcO*, had a three prime end sequence that was identical to the stcO gene but had a unique 5' end (Figure 33). Neither StcO or stcO* had any homologs in the GenBank databases.

The consensus sequence for stcP revealed an intron in the 5' UTR. U34740 position 44901-44937 which defines exon 1, showed a significant sequence difference from the Unigene consensus sequence of 14 insertions not present in the ESTs (Figure 34). This region was immediately upstream of and included the predicted AUG start site. Therefore, the correct start codon occurred 36 bases downstream at position 44872 which shortened the predicted protein by 12 amino acids from that described by Brown, et al.

There is only one Unigene cDNA clone with homology to stcQ. The EST sequences agreed well with the cosmid sequence and revealed no introns. The direction of transcription is the same for both stcP and stcQ. 161 bases of the 5' end of stcP was overlapped by the 3' end of stcQ from positions 45071-44911. Since the overlap for both genes was in their respective 5' and 3' untranslated regions, the coding sequence would be unaffected.

StcR was represented by nine cDNA clones but had no sequence similarity to any GenBank entries. The Unigene consensus sequence covered the coding region and confirms the two predicted introns, one of which was in the 5' UTR region.

StcS was present as a large, 1560 bp, ORF and was represented by two cDNA clones with homology to the stcS p450 monooxygenase. This was one of the few genes in the sterigmatocystin gene cluster without an intron.

D. Brown (personal communication) predicted a single intron containing

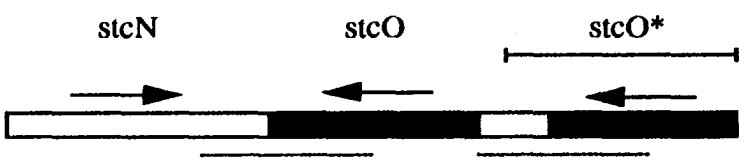


Figure 33. Sterigmatocystin gene cluster duplication region. Yellow indicates the sequences either coding for the stcN product or part of the the incomple duplication. Green represents the 97 bp overlap between the stcN and stcO genes, also found in the duplicated region. stcO* is shown as a composite region. Red is the region of gene stcO* which is not homologous to stcO and contains a region homologous to clone z7a04a1 not present in stcO. Red lines indicate the region of dupliction, black line the stcO* ORF, arrows indicate the direction of transcription.

```
C U34740    45071 GTTTCATTTTTGCCCGTGGACTTACCGGTTTCTTGGTGAATATAAGTTGC 45022
  stcP        41 GTTTCATTTTTGCCCGTGGACTTACCGGTTTCTTGGTGAATATAAGTTGC 90

C U34740    45021 AGTCGAGAAAGGAGAGGGACAAGCAAGCTTAGGCAGACTCCATATTTCAC 44972
  stcP        91 AGTCGAGAAAGGAGAGGGACAAGCAAGCTTAGGCAGACTCCATATTTCAC 140

C U34740    44971 CGTTCTCCGTACTTATTACTTCCCATTTTCGGAGGTGAAACTATTTACTC 44922
                                                  v--- v i  --  vi
  stcP       141 CGTTCTCCGTACTTATTACTTCCCATTTTCGGAGC---ATCCAT--ACAT 185

C U34740    44921 TTGTCTGATCTATCTATGCATTTGGTTGTTGACTGCACGTTTATCAATAA 4487
                    i  - --  iv---  ---
  stcP        186 TTATC-G--CTGG---TG---TTGGTTGTTGACTGCACGTTTATCAATAA 226
                   ***
C U34740    44871 TATGGACGCCATCTTCAAGCAAATCAAAGATGAGTACGCCCGTGCCGACG 4482
  stcP        227 TATGGACGCCATCTTCAAGCAAATCAAAGATGAGTACGCCCGTGCCGACG 276

C U34740    44821 AGCATGGCAAGCGAGAGATTCAAGGCTATATCCGCGAGTTGCAGGTTGGC 4477
  stcP        277 AGCATGGCAAGCGAAAGATTCAAGGCTATATCCGCGAGTTGCAGGTTGGC 326

C U34740    44771 TTCTATTCGGATTGGGATGTGGTGATGCGGTTGAGCAGTGGT 44730
  stcP        327 TTCTATTCGGATTGGGATGTGGTGATGCGGTTGAGCAGTGGT 368

Transitions / transversions = 1.25 (5 / 4)
Gap_init rate = 0.02 (6 / 328), avg. gap size = 2.33 (14 / 6)
```

Figure 34. Exon 1 of the stcP gene sequence from U34740 aligned with the consensus Unigene stcP sequence showing the sequence differences (underlined). The U34740 predicted AUG start site is highlighted and shown matching the corresponding EST consensus region. The new predicted start site based on the EST consensus sequence is also highlighted and marked with a (*).

conserved intron/exon borders for stcT. This was confirmed by the matching EST sequences from the single cDNA with homology to stcT but the positions of the ORF changed from the previously reported 48798-48863 to 50348-50401.

Five cDNA clones had homology to stcU. The consensus sequence extended to the 5' end of the gene and the two predicted introns were verified.

The Unigene consensus sequence generated from the three stcV homolog cDNA clones spanned the entire gene and confirmed that stcV had two predicted introns.

Brown et al. (Brown, 1996) stated that the stcW gene had homology to an FAD monooxygenase. However, no annotation of this region was given in the U34740 accession since the position of the introns was not clear (D. Brown, personal communication). Eight cDNA clones matched the stcW gene region and the Unigene consensus sequences revealed the positions of four introns (Table 21) and a final 424 amino acid protein.

3.27 Sterigmatocystin gene cluster relative expression level

It was possible to compare relative expression levels by examining the number of ESTs which were homologs to the predicted genes in the sterigmatocystin gene cluster. The various levels of individual gene expression ranged from no ESTs, to one and up to 22 ESTs (Figure 35). Although the direction of transcription for the genes in this cluster was determined, see Figure 35, it does not appear to correlate with the observed expression levels or their assigned role in the proposed biosynthetic pathway (Figure 30). There were four genes whose expression was exceptionally high, stcL, stcM, stcN and stcR. Since three of the genes were adjacent to each other and it has been suggested that selected enzymes in the sterigmatocystin biosynthetic pathway may have multiple functions or may be involved in another pathway (D. Brown, unpublished results), these four genes, two of which have no known function, would be candidates for investigation as a microcluster within the sterigmatocystin biosynthetic cluster.

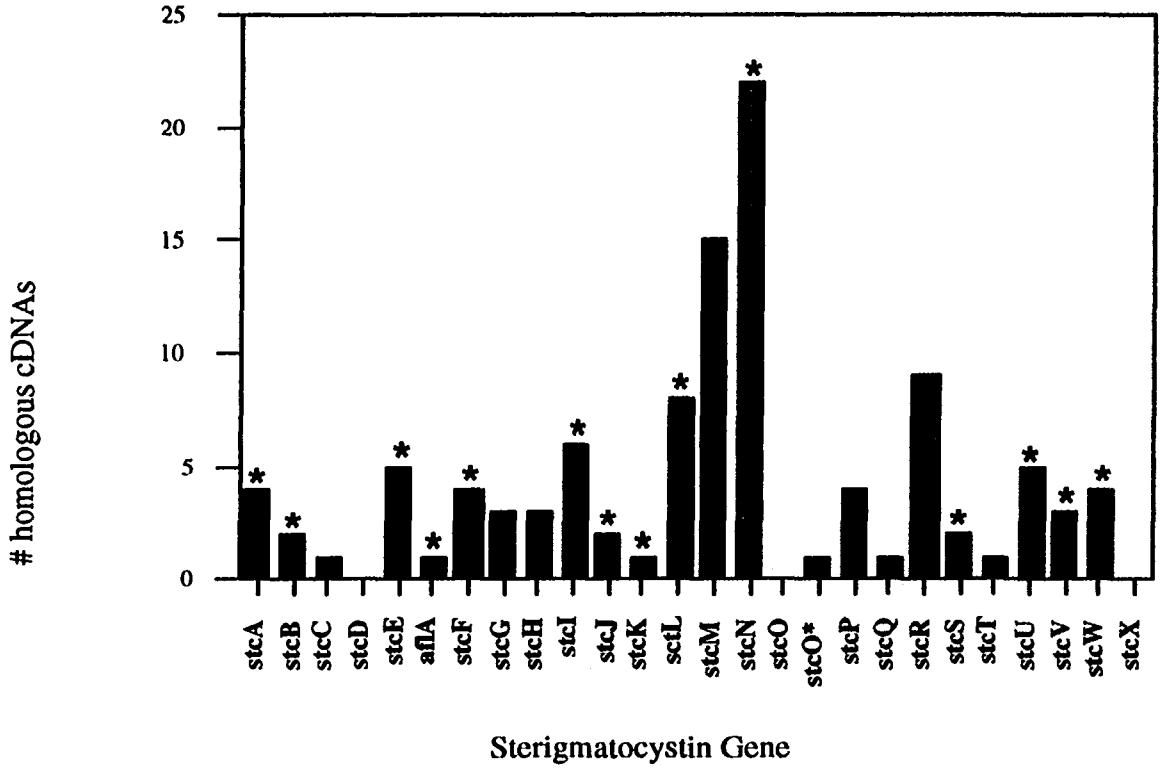


Figure 35. Histogram showing cDNA representation for each member of the sterigmatocystin gene cluster. Red bars represent genes transcribed to an arbitrary right, blue are genes are transcribed to the left. Genes are listed in order as they appear in the cluster. Stars indicate those genes assigned specifc function in the sterigmatocystin biosynthetic pathway.

3.28 Sterigmatocystin summary.

Previously it had been thought that expression of the sterigmatocystin pathway did not begin until after 24 hours of development when the sterigmatocystin toxin could be detected (Brown, 1996). The presence of expressed pathway members therefore was surprising since this library was constructed from RNA harvested at 24 hours into asexual development. A total of 107 cDNA clones or 1.2% of the total genes in the EST database represented sterigmatocystin pathway members.

Table 20 gives a summary of the cDNA clones which had homology to the sterigmatocystin gene cluster. The genes are listed in the order in which they are present in the cluster. A brief summary of the result of the Unigene comparison to the specific genes in the cluster described in detail in section 3.26 is also given in Table 20 indicating whether a change to the cluster sequence was suggested by the comparison. The sterigmatocystin EST sequences were confirmed by resequencing the cDNA templates. In ten cases, sequences of the ESTs differed from the previously reported sequence (accession number U34740). The corresponding intron-exon borders sites were determined which allowed predicting correct amino acid sequences. This resulted in demonstrating overlapping transcription units and gene duplications. In one instance, the stcO* gene was found in a duplicated region that was shown to encode a transcribed gene which had not been predicted previously. Since the previously predicted stcO gene did not have any matching ESTs but the stcO* gene did, it remains to be determined whether only one or both of the regions contain actively transcribed genes.

3.29 A Consensus Intron-Exon Splice Site for *Aspergillus nidulans*.

No consensus intron-exon splice site sequence has been published for *A. nidulans*. However, now that a large number of ESTs have been obtained from this present study along with the genomic sequences of cosmids W02H02, W30B01, W06E08, and the sterigmatocystin biosynthetic gene cluster, it was possible to generate

Table 20. cDNAs with homology to *A. nidulans* sterigmatocystin biosynthetic gene cluster, accession number U34740. The gene is listed with its proposed function assignment. The homologous cDNA for each are listed. The corrections or confirmations of the reported annotation are given.

Pathway gene, proposed function	Homologous cDNAs in Unigene consensus	Alignment results
StcA, polyketide synthase	c8f01, g5g07, g9h10, c0d01	*third exon found, corrected protein product
StcB, p450 monoxygenase	d2f11, y8c09	one intron confirmed, duplication shown
stcC, peroxidase	h4eo6	Confirmed reported sequence
stcD, ORF	-	not detected
StcE, ketoreductase	r4c09, p0b05, z5d08, r3b09, b0g10	*corrected sequencing error, additional intron found, corrected protein product
aflA, transcription factor	r1a06	convergent transcription with stcE shown to have 44 bp overlap of 3' ends
StcF, p450 monooxygenase	c9f01, h0f02, n5a05, a1f01	*corrected reported intron position
StcG, ORF	o6d04, o3d02, c7h01	*sequencing error corrected
StcH, ORF	i3d06, h4b11, y6h07	*sequencing error corrected, predicted two intron, 83 amino acid protein
StcI, lipase/esterase	g1d11, d4h03, r4g11, z6a11, e4a01, i7g11	*additional intron found, addition of 11 amino acids at 5' end

StcJ, fatty acid synthase-alpha subunit	n2e04, n2e10	Confirmed reported sequence
StcK, fatty acid synthase-beta subunit	g9b02	Confirmed reported sequence
StcL, p450 monooxigenase	w6h04, u4e02, m8f09, g2d05, h8g08, z5d12, f1b07, h0a05	Confirmed reported sequence
StcM, ORF	a1f04, z2f12, e6h09, s8h07, k8a06, u4e12, m9c04, p0f09, f1f08, h0b10, c4h10, y1e12, i0h09, y9e07, n3g07	*Sequencing error corrected, and 5' UTR intron shown.
StcN, GMC Oxydoreductase	u4g01, o0d06, h4g08, g9b11, x5h05, m6d05, y4b10, c9g02, c0g01, k5d04, m7f06, z3f08, h4a10, i8c11, v3b12, r8c03, g0f08, b0a06, n3c04, r8c03, e9c04, o9g06	Duplication found between stcO and stcP, includes 97 bp 3' overlap with 3' end of stcO
StcO, ORF	-	Not detected
StcO*, ORF	z7a04	*New gene identified, in partially duplicated region
StcP, O- methyl transferase	y4f12, c5e01, g6g10, c0c04	*Possible sequence error, correction shortens protein by 12 amino acids
StcQ, ORF	z4c07	3' end of stcQ overlaps 5' end of stcP
StcR, ORF	g4e02, r4g01, v1g04, x7g01, g2c03, j9b06, y6g09, h4f02, f1b10	Confirmed reported sequence

StcS, p450 monooxygenase	y1h01, x7g01	Confirmed reported sequence
StcT, Elongation factor 1 gamma	h4g09	*Corrected reported intron position
StcU, ketoreductase (verA)	d1e06, z4b06, f2c06, j7f11, w8e11	Confirmed reported sequence
StcV, dehydrogenase	o4e08, p0d10, b0f03	Confirmed reported sequence
StcW, FAD-containing monooxygenase	w6h01, g5a07, m8f12, u4b09	*4 introns found and 424 amino acid protein predicted.
StcX, ORF	-	not detected

* Unigene homolgs suggested a correction or addition to the sterigmatocystin cluster sequence U34740

the consensus splice sites for *A. nidulans* as presented in Table 22. The consensus derived from the examination of 49 introns was very similar to the eucaryotic consensus determined earlier by Mount (Mount, 1982) with the exception of position eight where a T residue was more prevalent in *A. nidulans* but no preference was noted in the Mount consensus and at position nine, where either a T residue or an A occurs in *A. nidulans* differs from the eucaryotic consensus.

3.30 Gene Density and Total Gene Number.

The prediction of gene density by combined EST and genomic sequencing allows a more accurate estimation of the number of genes contained in an organism's genome (Kupfer, 1997). The *A. nidulans* genome size and gene number have been independently estimated (Kupfer, 1997; Timberlake, 1978; Timberlake, 1980). Timberlake hybridized excess mRNA from varying growth conditions to single copy DNA to estimate that 13% of the genome was transcribed suggesting that *A. nidulans* contained about 6000 transcribed genes. More recently, Kupfer, et al. estimated the number of genes in the genome by applying the gene density determined for cosmid W06E08 (1 ORF/2.9 Kbp) to the 31 Mbp *A. nidulans* genome for an estimate of 10,000 genes. The sequence of the two overlapping cosmids W30b01 and W02H02 from a 45Kbp genomic region and the availability of the reannotated 60 Kbp genomic region of the sterigmatocystin cluster gave two additional genomic regions to further refine the estimate of average gene density and total gene number. The gene density for W30B01-W02H02 with its 17 predicted ORFs, was 1 gene/2.6 Kbp and the sterigmatocystin cluster region with 26 ORFs, including stcO*, was 1 gene/2.3 Kbp. An average of all three genomic regions gave a total of 56 genes in 144 Kbp or 1 gene/2.57 Kbp, suggesting that there are approximately 12,000 genes in the 31 Mbp genome. If this is a reasonable estimate, then the 3198 genes which were detected by this EST sequencing project represent approximately one third of the gene complement of *A. nidulans*.

Table 21. *A. nidulans* intron-exon splice sites for three chromosome VIII cosmids and the sterigmatocystin gene cluster of chromosome IV showing a consensus derived from 26 genes and 49 introns.

Gene Homolog	Left Position	Left Exon	Left Splice Junction	Intron Size(bp)	Right Splice Junction	Right Exon	Right Position
Cosmid W06e08							
brlA	11303	β	CCCTCA**GTCAGT**	**1011**	**TCGAAC**ACATGT	α	12314
ribonuclease T2	24521	1	CCTCTG**GTGCGC**	**50**	**GGCCAG**ATAACT	2	24570
	24779	2	AGCACG**GTACTT**	**34**	**TAACAG**GAACTT	3	24814
	25136	3	AGAAGC**GTTAGT**	**67**	**CTCTAG**TGACCC	4	25202
dihydroorotase	33166	1	.TCCACG**GTGCGT**	**172**	**AAATAG**TCCACC	2	32993
	33088	2	GTTATG**GTAAGG**	**90**	**TGATAG**CCAAAT	3	33177
	33634	3	GGAACA**GTAAAT**	**52**	**TTGTAG**TTCCG	4	33682
Cosmids W02H02 and W30B01							
prohibitin	12073	1	TATCAG**GTTCGT**	**60**	**TTTTAG**TCCTAC	2	12014
spermidine synthase	14151	1	TCAAGG**GTATGC**	**159**	**ATCCAG**ATGGCT	2	14309
	14491	2	GTTCTC**GTATGC**	**86**	**TTCCAG**CTACCA	3	14575
	14725	3	GATGAG**GTAACA**	**147**	**TTGCAG**GCCGTC	4	14831
	15052	4	ATGGAG**GTGTCA**	**274**	**TTACAG**CCGAAA	5	15325
Transketolase	18548	1	CTTGCG**GTATGC**	**262**	**TTCCAG**GTTGAT	2	18286
	18142	2	CCTCTC**GTATGA**	**54**	**CTATAG**CAACGG	3	18089
	18003	3	TTCCGT**GTAAGC**	**69**	**TTATAG**CAACTC	4	17935
Acetate reg. DNA binding	22654	1	TTGCTG**GTATGG**	**482**	**GTCAAG**CTGTCT	2	22596
	21942	2	GCTATT**GTAAGT**	**312**	**CTGTAG**TCGAGA	3	22228

protein (facB)	21536	3	CGGTTC**GTATCC..283..TCTTAG**TCTTTG	4	21895
	20323	4	CTTTTG**GTAAGT...50..TTAAAG**TCTGTG	5	21487
Spac3c7.01c	26870	1	GGGAAG**GTATGG...95..ATCCAG**ATCTCG	2	26777
	26305	2	GAGGAA**GTTACT...76..GGTCAG**GATAAA	3	26230
Exp. gene 1	31066	1	ATGAAG**GTAAGC...50..GTGCAG**TCCATT	2	31115
HMG-like protein 2	34051	1	AGAAGG**GTAAGC..128..CTACAG**CTGCGG	2	33889
Sterigmatocystin Gene Cluster					
stcA	1907	1	CTGATG**GTATGT...62..TCACAG**ACGAAG	2	1846
stcB	10187	1	AGACAC**GTAAGT...72..GGCTAG**GTTCGA	2	10258
StcF	19946	1	GGAGGG**GTAAGT...83..TTACAG**ACGACC	2	19865
StcG	22062	1	TACATG**GTACAT...69..CTACAG**GACGTG	2	22130
StcH	22730	1	ACGTCG**GTATGC...53..CGACAG**CCCTAT	2	22678
	22457	2	TGACCA**GTAAGT...58..CGACAG**AGTTCA	3	22400
StcI	22975	1	CAGCAG**GTACAG...48..TACCAG**TTCATA	2	23022
StcK	35585	1	GATCAG**GTATCT...50..TCGTAG**ACTTCA	2	35634
	35751	2	GGCCAT**GTAGTA...67..GAACAG**GTCCGC	3	35817
StcL	36770	1	GCTGAG**GTGTGT...66..CTCTAG**ACCCGC	2	36705
StcM	38588	1	TGTCCG**GTAATG...46..CACCAG**GTGCAA	2	38543
StcN	39281	1	CAGCCA**GTATGC...47..TTTCAG**GGGCTG	2	39327
	39411	2	TCATCG**GTATCT...44..AGACAG**GGGCTC	3	39454

StcP	44729	1	AGTGGT**GTATGT...47..GTGTAG**CCCTTG	2	44683
	44555	2	TGCCTT**GTAAGT...51..CAACAG**GGCGCA	3	44505
	44380	3	ACAGCT**GTATGT...50..TCTCAG**ATTCGA	4	44331
stcR	46167	1	TCCATA**GTAAGT...45..GTACAG**ATCACA	2	46211
	46517	2	ACTCTG**GTAGGT...46..AGGCAG**GTTTTC	3	46562
stcT	50348	1	TCTATG**GTACCA...54..GTATAG**TAACAT	2	50401
stcU	51601	1	CCAGAT**GTATGC...51..CAGCAG**GAATTC	2	51551
	51226	2	GATGAG**GTACGT...46..GGACAG**TGTGCC	3	51181
STCV	52738	1	CCGAGA**GTATGT...48..TACCAG**CCTATC	2	
STCW	53875	1	GCACTT**GTTTGT...55..TCGAAG**TACCCC	2	53929
	54179	2	CCAAAG**GTGCGC...46..CTGCAG**TGGCCA	3	54224
	55174	3	GCAGCT**GTACGT...95..CACCAG**GGTACA	4	55268
	55495	4	AGTTTG**GTAAGT...46..CAATAG**GTTCAA	5	55540

Summary for 26 genes and 49 introns:
Average intron size-114 bp, range 45-1011 bp

Eucaryotic consensus:				C			C		A	T			
	AG-	G	T	A	A	G	T...T	N	C	C	A	G-	N
[Mount, 1982 #140]					T			G		G			
					C			C		T			
A. nidulans	NG-	G	T	A	A	G	T...T	T	N	C	A	G-	N
Consensus motif:					T			G					

percent bases at	1	2	3	4	5	6	7	8	9	10	11	12
position for 49 introns:	q-	G	T	A	C\A\T	G	T....C/T/G	T	N	T\C	A	G-
	56	100	100	82	95.5	75	58	95.5	53	-	64	100	100

Chapter IV

Summary and Conclusion

During this work the sequence of over 12,000 ESTs were generated from an *A. nidulans* cDNA library that contained mRNAs from both vegetative and asexually expressed genes. To process the data in a rapid and orderly fashion, an automated screening method for handling a large number of sequences was developed. The sequences of each EST were released on the ACGT website and deposited in the GenBank dbEST.

The individual ESTs were aligned into a Unigene database with the Phred/Phrap sequence assembly programs (Green, copyright 1993-1996; Green, copyright 1994-1996) at various stages of the data collection. This allowed following the sampling progress of the cDNA library by estimating the efficiency of the sampling rate for unique 3' sequences and determining the endpoint of sampling by measuring how frequently new sequences were seen as the project progressed. In addition, the Unigene database aligned the respective 3' and 5' EST sequences into gene families yielding a consensus sequence with a greater accuracy than was obtained for each EST individually because of multiple overlapping EST sequences.

The Unigene consensus sequences also were examined for homologs via BlastX against the GenBank nonredundant protein database. These studies helped organize the data based on biological function. The resulting metabolic outline revealed a glimpse into the biochemistry of *A. nidulans*. These results were placed on the ACGT website where they are available for keyword searches and a summary of this data is presented in the appendices. Developing a Unigene database and the Biological Function outline allowed both estimating the numbers of genes expressed in the *A. nidulans* cell during vegetative and asexual development, as well as their role in the cell and revealed several examples of paralogous genes.

The heat shock protein 30 (HSP30) represented such a paralogous gene group in *A. nidulans* that contained at least six genes. Comparative analysis of the *A. nidulans* HSP30 paralogs revealed that the *N. crassa* and *A. nidulans* HSP30s are as closely related to *S. cerevisiae* as they are to bacteria or archae HSP30s.

Examination of the biological function outline shown in Appendix II also reveals that a number of other gene families including chitinases, serine/threonine kinases, and phosphatases as well as other classes of HSPs are present in this Unigene database. Thus, as with the HSP30 genes, new predictions based on sequencing, which otherwise were unavailable, now can be tested with more traditional genetic and molecular methods.

Examination of the very abundantly expressed Unigene members with no GenBank homologs suggested that additional studies will be required to identify conserved domains or motifs which may suggest the function or cellular location of EST sequences with no homologs in GenBank.

An important application of ESTs also has been demonstrated for genomic sequence annotation. The three cosmids which were sequenced and annotated supplied 83 Kbp of Chromosome VIII data for comparison to the EST database. These studies elucidated the structure of new genomic regions such as those for a transposon belonging to a class not found previously in fungus as well as the consensus sequence for fungal intron/exon borders and sequences of several new genes. The Chromosome IV region encoding the sterigmatocystin pathway had a large number of homologous ESTs. This EST information allowed reexamination of the sterigmatocystin cluster transcription units and revealed several new exons, a new gene and corrected previous sequencing errors in this region. The presence of the sterigmatocystin ESTs also suggested that the time of appearance of the toxin, sterigmatocystin, occurs more than 12 hours after the mRNAs in the pathway are synthesized (Brown, 1996; N. Keller, personal communication).

Intron/exon borders identified in the cosmids and sterigmatocystin cluster region were confirmed by the EST sequences and a more representative set of 3' and 5' splice

site consensus sequences was developed that was determined to be quite similar to those seen in mammalian systems. Finally, the present EST and genomic sequencing study provided the data needed to give a more accurate estimate (~12,000) of the number of genes in the *A. nidulans* genome.

The EST database produced during this dissertation research and the genomic regions sequenced provided a valuable tool for determining both gene content of this eucaryotic organism and for its genomic annotation. The study presented here will serve as a foundation for dissecting the molecular biology of *A. nidulans*. With the growing number of sequences available in the public databases, fruitful homology comparisons will continue to be made. As the fungal community looks toward the future when the entire sequence of *A. nidulans* and other fungal genomes will be completed, these *A. nidulans* ESTs will be an extremely useful resource for annotating future genomic sequences and beginning to understand gene expression in this and related fungi.

Chapter V

Literature Cited

(1996). Applied Biosystems Incorporated Users Manual ABI 377

(1998). AmpliTaq DNA polymerase protocol for use in double-stranded cycle sequencing. In Protocol: PE Applied Biosystems, pp. 45.

(1998). Amplitaq-FS DNA polymerase protocol for double-stranded cycle sequencing: Applied Biosystems Inc.

Abelson, J. (1979). RNA processing and the intervening sequences problem. Annual Review of Biochemistry *48*, 1035-1069.

Adams, M. D., Kelley, J.M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno R. F., Kerlavage, A. R., McCombie, W. R., Venter, J. C. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. Science *252*, 1651-1656.

Adams, M. D., 84 Authors (1995). Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature *377-suppliment*, 3-147.

Adams, T., Keller, T. (1996). Why *Aspergillus nidulans*? from: An initiative to sequence the genome of the filamentous fungus *Aspergillus niduans*, 1-5.

Adams, T. H., Boylan, M. T., Timberlake, W. E. (1988). brlA is necessary and sufficient to direct conidiophore development in *Aspergillus nidulans*. Cell *54*, 353-362.

Adams, T. H., Yu, J. H. (1998). Coordinate control of secondary metabolite production and asexual sporulation in *Aspergullus nidulans*. Current Opinions in Mcrobiology *1*, 674-677.

Altschul, S. F., Gish, JW., Miller, W., Myers, E. W., Lipman, D. J. (1990). Basic local alignment search tool. Journal of Molecular Biology *215*, 403-410.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research *25*, 3389-3402.

Aramayo, R. T., W E. (1990). Sequence and molecular structure of the *Aspergillus nidulans* yA (laccase I) gene. Nucleic Acids Research *18*, 3415.

Bairoch, A., Bucher, P., Hofman, K. (1995). The prosite database-its status in 1995. Nucleic Acids Research *24*, 189-196.

Ballance, D. J., Buxton, F. P., Turner, G. (1983). Transformation of *Aspergillus nidulans* by the orotidine-5'-phosphate decarboxylase gene of *Neurospora crassa*. Biochemical and Biophysical Research Communication *112*, 284-289.

Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature *226*, 1209-1211.

Barnes, W. M. (1995). Thermostable DNA polymerase with enhanced thermostability and enhanced length and efficiency of primer extension. U. S. Patent No. 5,436,149.

Barnes, W. M. (1992). The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion. Gene *112*, 29-35.

Bennett, J. W. (1985). Molds, manufacturing and molecular genetics. In Molecular Genetics of Filamentous Fungi, W. Timberlake, ed.

Benson, D. A., Boguski, M. Lipman, D. J., Ostell, J. (1996). GenBank. Nucleic Acids Research *24*, 1-5.

Berbee, M. L., Taylor, J. W. (1992). Dating the evolutionary radiations of the true fungi. Canadian Journal of Botany *71*, 1114-1127.

Berger, K. H., Yaffe, M.P. (1998). Prohibitin family members interact genetically with mitochondrial inheritance components in *Saccharomoyces cerevisiae*. Molecular and Cellular Biology *18*, 4043-4052.

Birnboim, H. C., Doly, J. (1979). A rapid alkaline extraction procedure for screening recombianant plasmid DNA. Nucleic Acids Research 7, 1513-1523.

Birnboim, H. C., Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Research 7, 1513.

Bishop, J. O., Morton, J. G., Rosbash, M. Richardson, M. (1974). Three abundance classes in HeLa cell messenger RNA. Nature *250*, 199-204.

Bodenteich, A., Chissoe, S., Wang, Y. F. Roe, B. A. (1993). Shotgun cloning as the strategy of choice to generate tmplates for high-throughput dideoxynucleotide sequencing. In Automated DNA sequencing and analysis techniques, C. Venter, ed.: Academic Press, London).

Boguski, M. S., Schuler, G. D. (1995). ESTablishing a human transcript map. Nature Genetics *10*, 369-371.

Bohinski, R. C. (1987). Modern Concepts in Biochemistry, fifth Edition: Allyn and Bacon, Inc.).

Bonaldo, M. F., Lennon, G., Soares, M. B. (1996). Normalization and Subtraction: Two approaches to facilitate gene discovery. Genome Research *6*, 791-806.

Borodovsky, M., Rudd, K.E., Koonin, E.V. (1994). Intrinsic and extrinsic approaches for detecting genes in a bacterial genome. Nucleic Acids Research 22, 4756-4767.

Breaux, K. (1995). Hot weather causes aflatoxin in southeast Texas corn. In Knight-Ridder Financial News (College Station, pp. C-3.

Bressac, B., Kew M., Wands, J., Ozturk, M. (1991). Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. Nature *350*, 429-431.

Brody, H., Carbon, J. (1989). Electrophoretic karyotped of *Aspergillus nidulans*. Proceedings of the National Academy of Sciences *86*, 6260-6263.

Brown, D. W., Yu, J. H. Kelkar, H. S. Fernandes, M. Nesbitt, T.C. Keller, N. P. Adams, T. H. Leonard, T. J. (1996). Twenty-five coregulated transcripts define a sterigmatocystin gene cluster in *Aspergillus nidulans*. Proceeding of the National Academy of Science, USA *93*, 1418-1422.

Bult, C. J., 339 other authors (1996). Complete genome sequence of the Methanogenic archaeon, *Methanococcus jannaschii*. Science *273*, 1058-1073.

Carlile, M. J., Watkinson, S. C. (1994). The Fungi, 3rd Edition: Academic Press, Harcourt Brace & Company).

Carter, J. M., Milton, I. D. (1993). An inexpensive and simple method for DNA purifications on silica particles. Nucleic Acids Research *21*, 1044-1046.

Chang, P. K., Yu, J., Bhatnagar, D., Cleveland, T. E. (1999). The carboxy-terminal portion of the aflatoxin pathway regulatory protein AFLR of *Aspergillus parasiticus* activates GAL::lacZ gene expression in *Saccharomyces cerevisiae*. Applied Environmental Microbiology *65*, 2508-2512.

Chen, F. (1997). Sequence and analysis of approximately 0.5 mega-basepair of the DiGeorge syndrome critical region on human chromosome 22 band Q11 and a syntenic mouse BAC. In Department of Chemistry and Biochemistry (Norman: University of Oklahoma), pp. 248.

Chomczynski, P., Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium. Analytical Biochemistry *162*, 156-159.

Clutterbuck, A. J. (1997). The validity of the *Aspergillus nidulans* linkage map. Fungal Genetics and Biology. *21*, 267-277.

Coates, P. J., Jamieson, D.J., Smart, K., Prescott, A. R., Hall, P. A. (1997). The prohibitin family of mitochondrial proteins regulate replicative lifespan. Current Biology *7*, 607-610.

Crabtree, J. S. (1997). Sequencing regions of human chromosome 11 and 22: a meningioma deletion region (22), the cat eye syndrome region (22) and the multiple

endocrine neoplasm region (11). In Department of Chemistry and Biochemistry (Norman: University of Oklahoma).

De Carli, L., Larizza, L. (1998). Griseofulvin. Mutation Research *195*, 91-126.

Dean, R. A., Timberlake, W. E. (1989). Production of cell wall-degrading enzymes by *Aspergillus nidulans*: A model system for fungal pathogenesis of plants. The Plant Cell *1*, 265-273.

Dotto, G. P., Horiuchi, K., Zinder, N. D. (1984). The functional origin of bacteriophage f1 DNA replication. Its signals and domains. Journal of Molecular Biology *172*, 507-521.

Ehrlich, K. C., Montalbano, B. G., Cary, J. W. (1999). Binding of the C6-zinc cluster protein, aflR, to the promoters of aflatoxin pathway biosynthesis genes in *Aspergillus paraciticus*. Gene *230*, 249-257.

Ewing, B., Hillier, L., Wendl, M. C. Green, P. (1998). Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Research *8*, 175-185.

Ewing, B. G., P. (1998). Base-calling of automated sequencer traces using Phred. II. Error probabilities. Genome Research *8*, 186-194.

Farman, M. L., Taura, S., Leong, S. A. (1966). The *Magnaporthe grisea* DNA fingerprinting probe MGR 586 contains the 3' end of an inverted repeat transposon. Molecular and General Genetics *251*, 675-681.

Fenelong, L. E., Hamilton, A. J., Figueroa, J. I., Bartholomew, M. A., Allen M. H., McCarthy, P., Hay, P. J. (1999). Production of specific monoclonal antibodies to Aspergillus species and their use in immunohistochemical identification of aspergillosis. Journal of Clinical Microbiology *37*, 1221-1223.

Fernandes, M., Keller, N. P., Adams, T. H. (1998). Sequence-specific binding by Aspergillus nidulans AflR, a C6 zinc cluster protein regulating mycotoxin biosynthesis. Molecular Microbiology *28*, 1355-1365.

Fleischmann, R. D., 39 other authors (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science *269*, 496-512.

Franz, G., Savakis, C. (1992). *Minos*, a new transposable element from *Drosophila hydei*, is a member of the Tc1-like family of transposons. Nucleic Acids Research *19*, 6646.

Fraser, C., M., 28 other authors (1995). The minimal gene complement of *Mycoplasma genitalium*. Sscience *270*, 397-403.

Fromm, M. E., Taylor, L. P., Walbot, V. (1986). Stable transformation of maize after gene transfer by electroporation. Nature *319*, 791-793.

Furuichi, Y., Morgan, M., Muthukrishman, S., Shatkin, A. J. (1975). Reovirus messenger RNA contains a methylated blocked 5'-termini structure: m7G(5')ppp(5')-GmpCp. Proceedings of the National Academy of Sciences *72*, 362-356.

Gerber, H.-P., Seipel, K., Georgiev, O., Hofferer, M., Hug, M., Rusconi, S., Schaffner, W. (1994). Transcriptional activation modulated by homopolymeric glutamine and proline stretches. Science *263*, 808-811.

Gething, M. J., Sambrook, J. (1992). Protein folding in the cell. Nature *355*, 33-45.

Gordon, D., Abajian, C., Green, P. (1998). Consed: A graphical tool for sequence finishing. Genome Research *8*, 195-202.

Green, P. (copyright 1994-1996). PHRAP documentation.

Green, P., Ewing, B. (copyright 1993-1996). PHRED documentation.

Genetics Computer Group, Inc. (GCG), Wisconsin Package Version 9.0, Copyright 1982-1996 Madison, Wisc.

Gubler, U., Hoffman, B. J. (1983). A simple and very efficient method for generating cDNA libraries. Gene *25*, 263-269.

Guzman-de-Pena, D., Aguirre, J., Ruiz-Herrera, J. (1998). Correlation between the regulation of sterigmatocystin biosynthesis and asexual and sexual sporulation in *Emericella nidulans*. Antonie Van Leeuwenhoek *73*, 199-205.

Hardwick, K. G., Murray, A. W. (1995). Mad1p, a phosphoprotein component of the spindle assembly checkpoint in budding yeast. Journal of Cell Biology *131*, 709-720.

Hawkins, A. R., Giles, N. H., Kinghorn, J. R. (1982). Genetical and Biochemical Aspects of Quinate Breakdown in the Filamentous Fungus *Aspergillus nidulans*. Biochemical Genetics *20*, 271-286.

Henikoff, S., Henikoff, J. G. (1991). The blocks database and its application. Nucleics Acids Research *19*, 6565.

Henikoff, S., Greene, E. A., Pietrokovski, S., Bork, P., Attwood, T. K., Hood, L. (1997). Gene Families: The taxonomy of protein paralogs and chimeras. Science *278*, 609-614.

Hillier, L., thirty other authors (1996). Generation and analysis of 280,000 human expressed sequence tags. Genome Research *6*, 807-828.

Houts, G. E., Miyagi, M., Ellis, C., Beard, D., Beard, J. W. (1979). Reverse transcriptase from avian myeloblastosis virus. Journal of Virology *29*, 517-522.

Hsu, I. C., Metcalf, R. A., Sun T., Welsh, J. A., Wang, N. J., Harris, C. C. (1991). Mutational hostspot in the p53 gene in human hepatocellular carcinomas. Nature *350*, 427-428.

Innis, M. A., Myambo, K. B., Gelfand, D. H., Drow, M. A. D. (1988). DNA seqquencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. Proceedings of the National Academy of Sciences *85*, 9436-9440.

Iwen, P. C., Rupp, M. E., Hinrichs, S. H. (1997). Invasive mold sinusitis: 17 cases in immunocompromised patients and review of the literature. Clinical and Infective Diseases

24, 1178-1184.

Jimenez, M., Mateo, R., Querol, A., Huerta, T., Hernandez, E. (1991). Mycotoxins and mycotoxigenic moulds in nuts and sunflower seeds for human consumption. Mycopathologia *115*, 121-127.

Josten, P., Paul, G. C., Nienow, A. W ., Thomas, C. R. (1998). Dependence of penicillium chrysogenum growth, morphology, vacuolation, and productivity in fed-batch fermentations on impeller type and agitation intensity. Biotechnology and Bioengineering *59*, 762-775.

Kafer, E. (1977). Meiotic and mitotic recombination in Aspergillus and its chromosomal aberrations. Advanced Genetics *19*, 33-131.

Katz, L., Donadio, S. (1993). Polyketide synthesis: prospects for hybrid antibiotics. In Annual Review of Microbiology *47*, 875-912.

Kawata, Y., Sakiyama, F., Tamaoki, H. (1988). Amino-acid sequence of ribonuclease T2 from *Aspergillus oryzae*. European Journal of Biochemistry *176*, 683-697.

Keller, N. P., Hohn, T. M. (1997). Metabolic pathway gene clusters in filamentous fungi 21, 17-29.

Khan, A. S., Wilcox, A. S., Polymeropoulos, M. H., Hopkins, J. A., Stevens, T. J., Robinson, M., Orpana, A.K., Sikela, J. M. (1992). Single pass sequencing and physical and genetic mapping of human brain cDNAs. Nature Genetics 2, 180-185.

Klenow, H., Overgaard-Hansen, K., Patkar, S. A., (1971). Proteolytic cleavage of native DNA polymerase into two different catalytic fragments. Influence of assay conditions on the change of exonuclease activity and polymerase activity accompapnying cleavage. European Journal of Biochemistry 22, 371-381.

Kolodrubetz, D., Burgum, A. (1991). Sequence and genetic analysis of NHP2: a moderately abundant high mobility group-like nuclear protein with an essential function in *Saccharomyces cerevisiae*. Yeast 7, 79-90.

Koonin, E. V., Mushegian, A. R. Galperin, M. Y. , Walker, D. R. (1997). Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. Molecular Microbiology *25*, 619-637.

Korolov, S., Nayal, M., Barnes, W. M., DiCera, E., Wuksman, G. (1995). Crystal structure of the large fragment of *Thermus aquaticus* DNA polymerase I at 2.5A resolution: structural basis for thermostabililty. Proceeding of the National Academy of Sciences *92*, 9264-9268.

Krug, M. S., Berger, S. L. (1989). Ribonuclease H activities associated with viral reverse transcriptases are endonucleases. Proceedings of the National Academy of Sciences *86*, 3539-3543.

Kupfer, D. M., Reece, C. A., Clifton, S. W., Roe, B. A., and Prade, R. A. (1997). Multicellular Ascomoycetous Fungal Genomes Contain More Than 8000 Genes. Fungal Genetics and Biology *21*, 364-372.

Kusakabe, T., Koga, K., Sugimoto, Y. (1994). Isolation and characterization of cDNA and genomic promoter region for a heat shock protein 30 from *Aspergillus nidulans*. Biochimica et Biophysica Acta *1219*, 555-558.

Lake, J. A., Jain, R., Rivera, M. C. (1999). Mix and Match in the tree of life. Science *283*, 2027-2028.

Lander, E. S., Waterman, M. S. (1988). Genomic mapping by fingerprinting random clones:a mathematical analysis. Genomics 2, 231-239.

Lee, L. G., Connell, C. R., Woo, S. L., Cheng, R. D., McArdle, B. F., Fuller, C. W., Halloran, N. D., Wilson, R. K. (1986). Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-=7-deazaguanosine triphosphate in place of dGTP. Nucleic Acids Research *14*, 1319-1324.

Lee, L. S., Bayman, P., Bennett, J. W. (1992). Mycotoxins. In Biotechnology of Filamentous Fungi: Technology and Products, D. B. Finkelstein, Ball, C., ed. (Boston: Butterworth-Heinemann), pp. 463-503.

Lindquist, S. C., E. A. (1988). The heat-shock proteins. Annual Review of Genetics *22*, 631-677.

Lobo-Hajdu, G., Braun, H. P., Romp, N., Grivell, L. A., Berden, J. A., Schnitz, U. K. (1996). Subunit VII of ubiquinol:cytoochrome-c oxidoreductase from *Neurospora crassa* is functional in yeast and has an N-terminal extension that is not essential for mitochondrial targeting. Journal of Biochemistry *320*.

MacCabe, A. P., Riach, M. B. R., Unles, S. E., and Kinghorn, J. R. (1990). The *Aspergillus nidulans* npeA locus consists of three contiguous genes required for penicillin biosynthesis. EMBO Journal *9*, 279-287.

Magnoli, C., Dalcero, A. M., Chiacchiera, S. M., Miazzo, R., Saenz, M. A. (1998). Enumeration and identification of Aspergillus group and Penicillium species in poultry feeds from Argentina. Mycopathologia *142*, 27-32.

Marra, M., and 41 additional authors (1999). An encyclopedia of mouse genes. Nature Genetics *21*, 191-194.

McNally, M. T., Free, S. J. (1988). Isolation and characterization of a Neurospora glucose-repressible gene. Current Genetics *14*, 545-551.

Merriman, P. J., Grimes, C. D., Ambroziak, J., Hackett, D. A., Skinner, P., Simmons, M. J. (1995). S elements: a family of Tc1-like transposons in the genome of *Drosophila melanogaster*. Genetics *141*, 1425-1438.

Morris, N. R., Pallone-Enos, A. (1992). Mitotic gold in a mold: Aspergillus genetics and the biology of mitosis. Trends in Genetics *8*, 32-37.

Mount, S. M. (1982). A catalogue of splice junction sequences. Nucleic Acids Research *10*, 459-472.

Muller, E. G. (1991). Thioredoxin deficiency in yeast prolongs S phase and shortens the G1 interval of the cell cycle. Journal of Biological Chemistry *266*, 9194-9202.

Nelson, M. A., 26 additional authors (1997). Expressed sequences from conidial, mycelial and sexual stages of *Neurospora crassa*. Fungal Genetics and Biology *21*, 348-363.

Nenoff, P., Horn, L. C., Schwenke, H., Mierzwa, M., Rieske, K., Houstein, U. F. (1996). Invasive mold infections in the university clinics of Leipzig in the period from 1992-1994. Mycoses *39 suppliment*, 107-112.

Nevill-Manning, C., Wu, T., Brutlag, D. Emotif (Department of Biochemistry, Stanford Univsity, Palo Alto, CA.

Nyyssonen, E., Amutan, M., Enfield, L., Stubbs, J., Dunn-Coleman, N. S. (1996). The transposable element Tan1 of Aspergillus niger var. awamori, a new member of the Fot1 family. Molecular and General Genetics *27*, 50-56.

Overbeek, R., Larsen, N., Smith, W., Maltsev, N., Selkov, E. (1997). Representation of Function: The next step. Gene *191*, 1-9.

Ozkaynak, E., Finley, D., Varshavsky, A. (1984). The yeast ubiquitin gene: head-to-tail repeats encoding a polyubiquitin precursor protein. Nature *312*, 663-666.

PerkinElmer. (1995). ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit, pp. 1-20.

Plesofsky-Vig, N., Brambl, R. (1995). Disruption of the gene for hsp30, an alpha-crystallin-related heat shock protin of *Neurospora crassa*, causes defects in thermotolerance. Proceedings of the National Academy of Sciences *92*, 5032-5036.

Pontecorvo, G. (1953). The Genetics of Aspergillus nidulans. In Advances in Genetics, M. Demerec, ed. (New York: Academic Press Inc.), pp. 142-235.

Prade, R. A., Timberlake, W. E. (1993). The *Aspergillus nidulans brlA* regulatory locus consists of overlapping transcription units that are individually required for conidiophore development. The European Molecular Biology Organization Journal, EMBO *12*, 2439-2447.

Prade, R. A., Griffith, J., Kochut, K., Arnold, J., Timberlake, W. E. (1997). In vitro reconstruction of the *Aspergillus* (=Emericella) *nidulans* genome. Proceedings of the National Academy of Sciences, USA *94*, 14564-14569.

Prober, J. M., Trainor, G. L., Dam. R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., Cocuzza, A. J., Jensen, M. A., Baumeister, K. (1987). A system for rapid DNa seqeuncing with fluorescent chain-terminating dideoxynucleotides. Science *238*, 336-341.

Ptashne, M. (1992). A Genetic Switch: Phage Lambda and Higher Organisms, 2nd Edition (Cambridge, MA: Cell Press & Blackwell Scientific Publications).

Ramesh, M. V., Sirakova, T. D., Kolattukudy, P. E. (1995). Cloning and characterization of the cDNAs and genes (mep20) encoding homologous metalloproteinases from *Aspergillus flavus* and *A. fumigatus*. Gene *165*, 121-125.

Reyes, F., Calatayud, J., Martinez, M. J. (1989). Endochitinase from *Aspergillus nidulans* implicated in the autolysis of its cell wall. FEMS Microbiological Letters *51*, 119-124.

Riley, M., Labedan, B. (1997). *Escherichia coli* gene products:physiological functions and common ancestries. In *Escherichia coli* and *Salmonella*, F. C. Neidhardt, ed.: ASM Press), pp. 2118-2202.

Rivera, M. C., Jain,R., Moore, J. E., Lake, J. A. (1998). Genomic evidence for two functionally distinct gene classes. Proceedings of the National Academy of Sciences *95*, 6239-6244.

Roe, B. A. (1995). Protocols for recombinant DNA isolation, cloning and sequencing. The University of Oklahoma.

Rosenblum, B. B., Lee, L. G., Spurgeon, S. L., Khan, S. H., Menchen, S. M., Heiner, C. R., Chen, S. M. (1997). New dye-ableled terminators for improved DNA sequencing patterns. Nucleic Acids Research *25*, 4500-4504.

Rosenblum, B. B., Lee, L. G., Spurgeon, S. L., Khan, S. H., Menchen, S. M.,

Heiner, C. R., Chen, S. M. (1997). New dye-labeled terminators for improved DNA sequencing patterns. Nucleic Acids Research *25*, 4500-4504.

Rosenzweig, B., Liao, L. W., Hirsh, D. (1983). Sequence of the *C. elegans* transposable element Tc1. Nucleic Acids Research *11*, 4201-4209.

Sachs, A., Wahle, E. (1993). Poly(A) tail metabolism and function in eucaryotes. The Journal of Biological Chemistry *268*, 22955-22958.

Sanger, F., Nicklen, S., Coulson, A. R. (1977). Dna sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Science *74*, 5463-5467.

Sasnauskas, K., Jomantiene, R., Lebediene, E., Lebedys, J., Januska, A., Janulaitis, A. (1992). Cloning and sequence analysis of a *Candida maltosa* gene which confers resistance to cycloheximide. Gene *116*, 105-108.

Schneuwly, S., Kuroiwa, A., Baumgartner, P., Gehring, W.J. (1986). Structural organisation and sequence of the homeotic gene Antennapedia of *Drosophila melanogaster*. EMBO Journal *5*, 733-739.

Schuler, G. D., Epstein, J. A., Ohkawa, H., Kans, J. A. (1996). Entrez: Molecular biology database and retrieval system. In Computer Methods for Macromolecular Sequence Analysis, R. F. Doolittle, ed. (San Diego: Acedemic Press), pp. 141-162.

Selkov, E., Maltsev, N., Olsen, G. J., Overbeek, R., Whitman, W. B. (1997). A reconstruction of the metabolism of *Methanococcus jannaschii* from sequence data. Gene *197*, 11-26.

Selkov Jr., E., Grechkin, Y., Mikhailova, N., Selkov, E. (1998). MPW: the Metabolic Pathways Database. Nucleic Acids Research *26*, 43-45.

Short, J. M., Fernandez, J. M., Sorge, J. A., Huse, W. D. (1988). Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties. Nucleic Acids Research *16*, 7583-7600.

Smith, L. M., Sander, J.Z., Kaiser, R. J., Huges, P., Dodd, C., Connell, C. R.,

Heiner, C., Kemt, S. B. H., Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. Nature *321*, 674-679.

Soares, M. B., Bonaldo, M. F., Jelenc, P., Su, L., Lawton, L., Efstratiadis, A. (1994). Construction and characterization of a normalized cDNA library. Proceeding of the National Academy of Sciences *91*, 9228-9232.

Sonnhammer, E. K., D. (1994). Homologous domains database of nonfragment protein sequences. Protein Science *3*, 482.

Staden, R., Dear, S. (1992). Indexing the sequence libraries: software providing a common indexing system for all the standard sequence libraries. DNA Sequencing 3, 99-105.

Stringer, M. A., Timberlake, W. E. (1995). dewA encodes a fungal hydrophobin component of the Aspergillus spore wall. Molecular Microbiology *16*, 33-44.

Stryer, L. (1995). Biochemistry, 4th Edition (New York: W. H. Freeman and Company).

Tabor, S., Richardson, C. C. (1995). A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. Proceedings of the National Academy of Sciences 92, 6339-6343.

Takaya, N., Yamazaki, D., Horiuchi, H., Ohta, A., Takagi, M. (1998). Cloning and characterization of a chitinase-encoding gene (chiA) from *Aspergillus nidulans*, disruption of which decreases germination frequency and hyphal growth. Bioscience and Biotechnology Biochemistry *62*, 60-65.

Tanaka, T., Kuroda, M., Sakaguchi, K. (1977). Isolation and characterization of four plasmids from *Bacillus subtilis*. Journal of Bacteriology *129*, 1487-1494.

Tatsumi, H. M., S., Tsuji, R. F., Ishida, Y., Murakami, K., Masaki, A., Kawabe, H., Arimura, H., Nakano, E., Motai, H. (1991). Cloning and expression in yeast of a cDNA clone encoding *Aspergillus oryzae* neutral protease II, a unique metalloprotease.

Molecular and General Genetics *228*, 97-103.

Tatusov, R. L., Koonin, E. V., Lipman, D. J. (1997). A genomic perspective on protein families. Science *278*, 631-637.

Timberlake, W. E. (1980). Developmental Gene Regulation in *Aspergillus nidulans*. Developmental Biology *78*, 497-510.

Timberlake, W. E. (1978). Low repetitive DNA content in *Aspergillus nidulans*. Science *202*, 773-775.

Timberlake, W. E. (1990). Molecular Genetics of Aspergillus Development. Annual Review of Genetics *24*, 5-36.

Todd, R. B., Murphy, R. L., Martin, H. M., Sharp, J. A., Davis, M. A., Katz, M. E., Hynes, M. J. (1997). The acetate regulatory gene facB of *Aspergillus nidulans* encodes a Zn(II)2Cys6 transcriptional activator. Molecular and General Genetics *20*, 495-504.

Vollrath, D., Davis, R W. (1987). Resolution of DNA molecules greater than 5 megabases by contour-clamped homogeneous electric fields. Nucleic Acids Research *15*, 7865-7875.

Wedaman, K. P., Knight, A.E., Kendrick-Jones, J., Scholey, J.M. (1993). Sequences of sea uchin kinesin light chain isoforms. Journal of Molecular Biology *231*, 155-158.

Wharton, K. A., Yedvobnic, B., Finnerty, V.G., Artavanis-Tsakonas, S. (1985). OPA: a novel family of transcribed repeats shared by the Notch locus and other developmentally regulated loci in D. melanogaster. Ce.. *40*, 55-62.

White, O., Kerlavage, A. (1997). EGAD, The Expressed Gene Anatomy Database (ftp://ftp.tigr.org/pub: The Institute for Genome Research).

Zhang, J., Madden, T. L. (1977). PowerBLAST: A new network BLAST application for interactive or automated sequence analysis and annotation. Genome Research *7*, 649-656.

Appendix I. *Aspergillus nidulans* categories of cellular functions with keywords. In bold are headings, in standard type are the keywords. & indicates a variation of the keyword in the line above, all entries matching the & keyword will be placed under the higher keyword. Some keywords are abreviated or capitalized as needed for use in searching the BlastX reports.

Categories of Cellular Functions

I. Bioenergetics and Metabolism

A. Metabolism of Carbohydrates(for glucose see energy)

1. Chitin metabolism
chitin
&chitinase

2. Cellulose degradation
beta glucosidase-breakdown of cellulose
&beta glucosidase
&beta-glucosidase
&glucoside glucohydrolase
glucanase
&GNS1 PROTEIN
&beta glucanase
cellulase
cellobiohydrolase

3. Pectin degradation
pectin
pectate
2-deoxy-D-gluconate 3-dehydrogenase

4. Cutin metabolism
cutin

5. Polysaccharide synthesis
UDP-glucose dehydrogenase

6. Energy reserve synthesis-see also energy reserve metabolism
GLYCOGEN (STARCH) SYNTHASE
&glycogen synthase
1,4-ALPHA-GLUCAN BRANCHING ENZYME
starch branching enzyme
starch synthase
trehalose synthase

7. Arabinose metabolism
arabin

8. Glucosamine
GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE
GLUCOSAMINE-6-PHOSPHATE ISOMERASE-GLUCOSAMINE UTILIZATION PATHWAY
&GLUCOSAMINE-6-PHOSPHATE ISOMERASE

&GLUCOSAMINE-6-PHOSPHATE DEAMINASE
beta glucosamine
&glycoamidase
GLUCOSAMINIDASE-degradation of glycans
&GLUCOSAMINIDASE
GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE

9. Aminosugar metabolism
PHOSPHOACETYLGLUCOSAMINE MUTASE

10. Sucrose metabolism
sucro
levanase-sucrose to glucose
&levanase

11. Galactose metabolism
galactose
GALACTOSE-1-PHOSPHATE URIDYLYLTRANSFERASE

12. Mannitol metabolism
manno
mannitol

13. Xylanose metabolism
xylosidase
xylanase
xylitol dehydrogenase

14. Quinate metabolism
quinate-utilization is in cluster
&quinate

15. Sorbitol metabolism
SORBITOL UTILIZATION PROTEIN
SORBITOL DEHYDROGENASE

16. Gluconate
Glucose Oxidase-first step of glucose + O2 to gluconic acid
&Glucose Oxidase

17. Pyranose metabolism
pyranose oxidase

18. Ribitol metabolism
ribitol kinase

19. Calvin cycle
RIBULOSE-PHOSPHATE 3-EPIMERASE-ribulose-5 PO4->xylulose-5 PO4
&RIBULOSE-PHOSPHATE 3-EPIMERASE

B. Metabolism of Amino acids and Related Molecules

1.Arginine metabolism

a. Arginine anabolism-glutamine, CO2 to arginine
ORNITHINE CARBAMOYLTRANSFERASE
ARGININOSUCCINATE SYNTHASE
ARGININOSUCCINATE LYASE
AGMATINASE

b. Arginine catabolism-arginine to proline
ARGINASE-also see urea cycle
&ARGINASE
ARG-6 PROTEIN
ACETYLORNITHINE AMINOTRANSFERASE
PYRROLINE-5-CARBOXYLATE REDUCTASE

2.Asparagine metabolism
ASPARAGINE SYNTHASE

3.Aspartic acid metabolism

aspartate anabolism-oxaloacetate, glutamate to aspartate
ASPARTATE AMINOTRANSFERASE
Aspartase

4.Cysteine metabolism
cysteine
O-ACETYLHOMOSERINESULFHYDRYLASE-also methionine biosyn
&O-ACETYLHOMOSERINESULFHYDRYLASE
&O-acetyl-L-homoserine sulfhydrylase
homocysteine synthase
CYSTATHIONINE GAMMA-LYASE

5.Glutamine metabolism
GLUTAMINE SYNTHETASE

6.Glycine metabolism

a. Serine to glycine
GLYCINE HYDROXYMETHYLTRANSFERASE

b. glycolate to glycine
GLYCERATE DEHYDROGENASE
SERINE--PYRUVATE AMINOTRANSFERASE

c. threonine to glycine
THREONINE ALDOLASE
PYRIDOXAMINE-PHOSPHATE AMINOTRANSFERASE

d. glycine catabolism

-glycine to serine
GLYCINE HYDROXYMETHYLTRANSFERASE

e. glycine decarboxylase complex-made up of P,T,L,H-removes amino group
GLYCINE CLEAVAGE SYSTEM H PROTEIN

&glycine cleavage system protein H
Glycine cleavage system T protein
&AMINOMETHYLTRANSFERASE

7.Histidine metabolism
HISTIDINE BIOSYNTHESIS

8.Isoleucine metabolism
2,3-DIHYDROXYACID HYDROLYASE-4th step in iso & val biosyn
&2,3-DIHYDROXYACID HYDROLYASE
&DIHYDROXY-ACIDDEHYDRATASE

-catabolism
propionyl-CoA carboxylase-also leucine and valine degradation
&propionyl-CoA carboxylase
methylcrotonyl-CoA carboxylase

9.Leucine metabolism
hydroxy-3-methylglutaryl-CoA lyase-FINAL STEP OF KETOGENESIS AND LEUCINE CATABOL
ISM
&hydroxy-3-methylglutaryl-CoA lyase
&HYDROXYMETHYLGLUTARYL-COALYASE

10.Lysine metabolism
HOMOCITRATEDEHYDRATASE
SACCHAROPINE DEHYDROGENASE

11.Methionine metabolism
methionine synthase
ORNITHINE AMINOTRANSFERASE
HOMOSERINE O-ACETYLTRANSFERASE
cystathionine beta-lyase-3rd step
&cystathionine beta-lyase
methionine synthase-last step in met biosynthesis
&methionine synthase
&5-METHYLTETRAHYDROPTEROYLTRIGLUTAMATE--HOMOCYSTEINEMETHYLTRANSFERASE

12.Phenylalanine metabolism
phenylalanine hydroxylase

13.Proline metabolism
proline dehydrogenase

14.Serine metabolism
PHOSPHOSERINE AMINOTRANSFERASE

15.Tryptophan metabolism
anthranilate phosphoribosyltransferase-2nd step in tryp biosyn
&anthranilate phosphoribosyltransferase
CATECHOL 1,2-DIOXYGENASE-tryp & lysine catabolism in KETOADIPATE PATHWAY
&CATECHOL 1,2-DIOXYGENASE

&hydroxyquinol-1, 2-dioxygenase

16.Tyrosine metabolism
prephenate dehydrogenase

17.Valine metabolism
hydroxyisobutyrate dehydrogenase
valine synthetase
METHYLMALONATE-SEMIALDEHYDEDEHYD

18.Aromatic amino acid metabolism
PHOSPHO-2-DEHYDRO-3-DEOXYHEPTONATEALDOLASE

19.Polyamine biosynthesis
polyamine
&spermidine

C. Metabolism of Nucleotides and Nucleic Acids, Purines, Pyrimidines

1. Nucleotide metabolism
NUCLEOSIDE DIPHOSPHATE KINASE
ribose-phosphate pyrophosphokinase-purine,pyrimidine biosyn, also his and trypto phan biosyn
&ribose-phosphate pyrophosphokinase

2. Purine metabolism

a. inosine mono phosphate de novo biosynthesis
amidophosphoribosyl transferase
phosphoribosylamine-glycine ligase
phosphoribosylglycinamide formyltransferase
PHOSPHORIBOSYLFORMYLGLYCINAMIDINE SYNTHASE
PHOSPHORIBOSYLFORMYLGLYCINAMIDINE CYCLO LIGASE
amidophosphoribosyltransferase
ADENYLOSUCCINATELYASE
ADENYLOSUCCINATE SYNTHETASE-first committed step to AMP biosyn
&ADENYLOSUCCINATE SYNTHETASE
&IMP--ASPARTATE LIGASE
PHOSPHORIBOSYLAMINOIMIDAZOLECARBOXAMIDE
FORMYLTRANSFERASE
&IMP cyclohydrolase

b. other purine metabolic enzymes
INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE
glutamine phosphoribosylpyro
xanthine dehydrogenase
amidophosphoribosyl transferase
phosphoribosoyl synthase
Purine Nucleoside Phosphorylase
AMP DEAMINASE
&MYOADENYLATEDEAMINASE
adenosine kinase-phosphorylates purine nucleoside
&adenosine kinase

3. Pyrimidine metabolism

a. de novo pyrimidine biosynthesis
PYRIMIDINE BIOSYNTHESIS ENZYME Thi5-thiamine biosyn
&PYRIMIDINE BIOSYNTHESIS ENZYME Thi5
carbamoyl-phosphate synthase
ASPARTATE CARBAMOYLTRANSFERASE
dihydroorotase
dihydroorotate
dihydroorotate dehydrogenase-4th step in pyr biosyn
&dihydroorotate dehydrogenase
orotate reductase
OROTATE PHOSPHORIBOSYLTRANSFERASE
orotidine

b. other pyrimidine metabolic enzymes
PHOSPHORIBOSYLPYROPHOSPHATE SYNTHETASE
DEOXYCYTIDYLATE DEAMINASE-degeradation to dUMP
&DEOXYCYTIDYLATE DEAMINASE
&DCMP DEAMINASE

D. Metabolism of Lipids, Fatty Acids, Sterols-See also fatty acid degradation

1. Fatty acid biosynthesis
a. ACETYL-COA CARBOXYLASE-yields malonylcoA,comitted step to FA biosyn.
&ACETYL-COA CARBOXYLASE

b. ACYL-CARRIER PROTEINS
ACYL-CARRIER-PROTEIN

c. FATTY ACID SYNTHASE
&FATTY ACID SYNTHASE
3-OXOACYL-[ACYL-CARRIER-PROTEIN] SYNTHASE
[ACYL-CARRIER PROTEIN] S-MALONYLTRANSFERASE
CROTONOYL-[ACYL-CARRIER PROTEIN] HYDRATASE
ENOYL-[ACYL-CARRIER PROTEIN] REDUCTASE (NADH)
3-ketoacyl-acyl carrier protein reductase
3-HYDROXYPALMITOYL-[ACYL-CARRIER PROTEIN] DEHYDRATASE
3-HYDROXYDECANOYL-[ACYL-CARRIER PROTEIN] DEHYDRATASE

d. BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE-keto acids->short branch-chain fatty acids
&BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE

e. Other
stearoyl-CoA desaturase-adds double bonds to fatty acyl coA
&stearoyl-CoA desaturase
2. Sterols
a. sterol
sterol
steroid monooxygenase

LANOSTEROL SYNTHASE
&LANOSTEROL CYCLASE
glucuronidase
HYDROXY-3-METHYLGLUTARYL-COENZYME A REDUCTASE-also mevalonate biosyn->isoprenoids
&HMG-CoA-reductase
&HYDROXY-3-METHYLGLUTARYL-COENZYME A REDUCTASE
C-5 STEROL DESATURASE
sterol demethylase
&eburicol C14-alpha-demethylase

b. Farnesol biosynthesis
ISOPENTENYL-DIPHOSPHATE ISOMERASE
GERANYLGERANYL PYROPHOSPHATE SYNTHETASE
&dimethylallyltransferase
GERANYLGERANYL TRANSFERASE
hydroxysteroid dehydrogenase

c. cholesterol metabolism
C-4 METHYL STEROL OXIDASE-cholesterol biosynthesis
&C-4 METHYL STEROL OXIDASE
STEROL O-ACYLTRANSFERASE-esterification of cholesterol
&STEROL O-ACYLTRANSFERASE

3. Lipids

a.phospholipid biosynthesis
LYSOPHOSPHOLIPASE PRECURSOR
phospholipid biosynthesis
phosphatidyl synthase
PHOSPHATIDYLSERINE SYNTHASE
PHOSPHATIDYLSERINE DECARBOXYLASE
myo-inositol phosphate synthase-biosynthesis of inositol containing phospholipids
&myo-inositol phosphate synthase
&myo-inositol 1-phosphate synthase
&myo-inositol-3-phosphate synthase

b.SPHINGOLIPIDS
serine palmitoyltransferase

c.Lipopolysaccharide biosyn-biomembrane precursors
UDP-glucose:sterol glucosyltransferase
UDP-GLUCOSE PYROPHOSPHORYLASE
& UTP--GLUCOSE-1-PHOSPHATE URIDYLYLTRANSFERASE

E. Aromatic compound metabolism
4-coumarate--CoA ligase-thioester substrates for phenylpropanoid biosyn
&4-coumarate--CoA ligase
&coumarate:CoA ligase
&COUMARATE--COA LIGASE
chorismate
aminobutyrate aminotransferase
CARBOXYMUCONOLACTONE DECARBOXYLASE-aromatic hydrocarbon cat.

&CARBOXYMUCONOLACTONEDECARBOXYLASE

F. Sulfur Metabolism
SULFATE ADENYLYLTRANSFERASE
ADENYLYLSULFATE KINASE
sulphur metabolite repression-4 genes, met down, no S up
&sulphur metabolite repression
&sconC
-sulfate assimilation
sulfate adenylyltransferase-leads to biosynthesis of cys&met
&sulfate adenylyltransferase
&ATP-SULFURYLASE

G. Phosphate Metabolism
INORGANIC PYROPHOSPHATASE
&PYROPHOSPHATEPHOSPHO-HYDROLASE

H. Nitrogen Metabolism (see also amino acid metabolism)
nitrate reductase
nitrite reductase
NITROGEN METABOLIC REGULATION PROTEIN -NEGATIVE REGULATORY PROTEIN IN THE NITROG
EN CONTROL CIRCUIT
&NITROGEN METABOLIC REGULATION PROTEIN
&nitrogen metabolite repression regulator
cyanate lyase-cyanate, bicarbonate substrates
&cyanate lyase

-urea cycle
urea cycle
glutamate dehydrogenase
CARBAMOYL-PHOSPHATE SYNTHASE-also arginine and pyrimidine biosynthesis
&CARBAMOYL-PHOSPHATE SYNTHASE

I. Metabolism of Cofactors, prosthetic groups

1.Nicotinamide coenzymes
nicotinamide adenine dinucleotidephosphate
NICOTINATE-NUCLEOTIDE PYROPHOSPHORYLASE-DE NOVO BIOSYNTHESIS OF NAD AND NADP
&NICOTINATE-NUCLEOTIDE PYROPHOSPHORYLASE
kynureninase-biosyn of NAD cofactors
&kynureninase
&alpha-aminoadipate aminotransferase

2.Biocytin (biotin)
biotin carboxylase

3.Thiamine
thiamine
THIAMIN BIOSYNTHESIS
&THIAMIN-PHOSPHATE PYROPHOSPHORYLASE
&THIAMIN BIOSYNTHETIC BIFUNCTIONAL ENZYME

&NMT1 PROTEIN

4.Coenzyme A
acetyl-coenzyme A synthetase
acetyl coenzyme A acetyltransferase

5.fFavins
riboflavin synthase
GTP cyclohydrolase II-riboflavin biosyn
&GTP cyclohydrolase II

6.Folate-methyl donor
folate

7. Heme
heme
siroheme synthase
##-iron uptake
FERRIC REDUCTASE TRANSMEMBRANE COMPONENT 2

8.PANTOTHENATE
PANTOTHENATE SYNTHETASE
&PANTOATE--BETA-ALANINE LIGASE

9.Molybdopterin
molybdopterin biosynth
&molybdopterin converting factor

J. Energy
1. Glycolysis

a.hexokinase
&hexokinase

b.glucose-6-phosphate isomerase
&glucose-6-phosphate isomerase

c.fructose-6-phosphate2-kinase
&fructose-6-phosphate2-kinase
&PHOSPHOFRUCTOKINASE
&phosphofructo-2-kinase
&fructose-2,6-bisphosphate 2-phosphatase

d.fructose-bisphosphate aldolase-also gluconeogenesis, PP cycle, carbon fixation, fructose and mannose metab
&fructose-bisphosphate aldolase

e.triose-phosphate isomerase
&triose-phosphate isomerase

f.glyceraldehyde-3-phosphate dehydrogenase
&glyceraldehyde-3-phosphate dehydrogenase

g.phosphoglycerate kinase
&phosphoglycerate kinase

h.phosphoglycerate mutase
&phosphoglycerate mutase
&PHOSPHOGLYCERATEDEHYDRATASE

i.phosphopyruvate hydratase
&phosphopyruvate hydratase

j.pyruvate kinase
&pyruvate kinase

2. Gluconeogenesis
a.LACTATE DEHYDROGENASE
&LACTATEDEHYDROGENASE

b.pyruvate carboxylase
&pyruvate carboxylase

c.phosphoenolpyruvate carboxykinase
&phosphoenolpyruvate carboxykinase

d.FRUCTOSE-1,6-BISPHOSPHATASE
&FRUCTOSE-1,6-BISPHOSPHATASE
&FRUCTOSE-1,6-BISPHOSPHATAS
&fructose-bisphosphatase

3. Pentose-phosphate pathway
a.glucose-6-phosphate dehydrogenase
&glucose-6-phosphate dehydrogenase
&GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE

b. gluconeolactonase
&gluconeolactonase

c.phosphogluconate dehydrogenase
&phosphogluconate dehydrogenase

d.ribose 5-phosphate isomerase-nonoxidative PO4
&ribose 5-phosphate isomerase

e.ribulose-phosphate 3-epimerase
&ribulose-5-phosphate 3-epimerase

f.transketolase
&transketolase

g.transaldolase
&transaldolase

4. Pyruvate dehydrogenase
pyruvate dehydrogenase
pyruvate decarboxoylase

dihydrolipoyl transacetylase
&LIPOAMIDEACYLTRANSFERASE
dihydrolipoamide dehydrogenase
pyruvate dehydrogenase kinase-inhibits pyruvate dehyd by phos of E1 alpha subunit
&pyruvate dehydrogenase kinase

5. Tricarboxylic acid pathway
a.citrate synthase
&citrate synthase

b.aconitate hydratase
&aconitate hydratase

c.isocitrate dehydrogenase
&isocitrate dehydrogenase

d.alpha-ketoglutarate dehydrogenase
&alpha-ketoglutarate dehydrogenase

e.SUCCINYL-COA LIGASE
&SUCCINYL-COA LIGASE

f.succinate dehydrogenase
&succinate dehydrogenase

g.FUMARATEHYDRATASE
&FUMARATEHYDRATASE

h.malate dehydrogenase
&malate dehydrogenase

6. Related reactions
citrate lyase-citrate to oxaloacetate+acetylcoA
&citrate lyase
&citrate-lyase
&CITRATECLEAVAGEENZYME
&CITRATE(PRO-S-)-LYASE
carbonic anhydride
&carbonic

-related pathway
homoprotocatechuate degradation-degraded to krebs intermediates
&homoprotocatechuate degradation

7. Glyoxylate cycle
malate synthase
isocitrate lyase

8. Fermentation, alcoholic
a.pyruvate decarboxylase
&pyruvate decarboxylase

b.alcohol dehydrogenase
&alcohol dehydrogenase

9.Fermentation, other

LACTATE DEHYDROGENASE-pyruvate to lactate
&LACTATE DEHYDROGENASE

butanediol dehydrogenase

10. Monocarbon metabolism
formate dehydrogenase

-C1 Metabolism
ALCOHOL OXIDASE-first step-methanol utylization-> FORMALDEHYDE
&ALCOHOL OXIDASE

11. Metabolism of energy reserves (glycogen, starch, trehalose)

a. Glycogen degradation

glycogen phosphorylase
PHOSPHOGLUCOMUTASE-glycogen deg, glu1PO4 to glu6PO4
&PHOSPHOGLUCOMUTASE
&GLUCOSE PHOSPHOMUTASE

b. Starch degradation

alpha glucosidase
&alpha-glucosidase
&MALTASE
glucoamylase
&alpha-amylase

c. Trehalose degradation

trehalose-6-phosphate synthase
trehalase

12.Fatty acid degradation

a. lipase-triacylglycerols->glycerol+FA
&lipase

b. beta-oxydation of fatty acids
beta-oxidation

i. fatty acid activation-thiokinase
long-chain-fatty-acid-CoA ligase
medium-chain acyl-CoA ligase
LONG-CHAIN ACYL-COASYNTHETASE

ii.carnitine acetyl transferase
&carnitine acetyl transferase
carnitine racemase-d to l form
&carnitine racemase

iii.acyl-CoA dehydrogenase
&acyl-CoA dehydrogenase

iv.enoyl-CoA hydratase
&enoyl-CoA hydratase

v.3-hydroxyacyl-CoA dehydrogenase
&3-hydroxyacyl-CoA dehydrogenase

vi.HYDROXYBUTYRYL-COA DEHYDROGENASE
&HYDROXYBUTYRYL-COA DEHYDROGENASE

vii.3-ketoacyl-CoA thiolase
&3-ketoacyl-CoA thiolase
&3-keto-acyl-CoA thiolase

c. odd chain fatty acids
methylmalonyl carboxylase-also ile, thr, met, val degradation
&methylmalonyl carboxylase

d. Unsaturated fatty acid degradation
fatty acyl-CoA reductase

e. Branch chain fatty acid degradation
branched-chain enoyl CoA reductase
&branched-chain enoyl CoA reductase
&2-methyl branched-chain enoyl CoA reductase

f. Ketone body metabolism
3-OXOACID COA-TRANSFERASE
hydroxybutyrate dehydrogenase
ACETOACETYL-COA THIOLASE-acetyl-coA->acetoacyl-coA
&ACETOACETYL-COA THIOLASE
SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASE-acetoacetate->acetoacyl-coA
&SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASE

13. Metabolism of other energy sources

a. alcohol dehydrogenases
FORMALDEHYDE DEHYDROGENASE-long chain primary alcohol->aldehyde or ketone
&FORMALDEHYDE DEHYDROGENASE
aldehyde reductase
&aldehydereductase

b. GLYCEROL
glycerol
glycerol-3-phosphatase
glycerol-3-phosphate dehydrogenase

c. proprionate
PRPD PROTEIN

d. Other
diaminobutyrate decarboxylase-dia-butyrate to dia-propane
&diaminobutyrate decarboxylase
ACETAMIDASE-allows acetamide and formamide as sole C or N source
&ACETAMIDASE
&amidase
formamidase
acetate
ALDEHYDE DEHYDROGENASE-broad substrate specificity
&ALDEHYDEDEHYDROGENASE

14. Electron transport

a. Complex I-NADH-ubiquinone
NADH dehydrogenase
&ubiquinone

b. Complex II-Succinate-ubiquinone
succinate dehydrogenase

c. Complex III-Ubiquinone to cytochrome C
cytochrome b
cytochrome c
cytochrome oxidase
CBP4 PROTEIN-cytoC, ubiquinol assembly
&CBP4 PROTEIN

d. Other electron transport pathways
NADH OXIDASE
NADPH DEHYDROGENASE

f. Electron carriers
flavoprotein
FLAVOHEMOPROTEIN
QUINONE

g. Component enzymes and molecules
FLAVIN OXIDOREDUCTASE
&HYDROXYPHENYLPYRUVATE DIOXYGENASE
RESPIRATORY complex assembly
MITOCHONDRIAL FAD CARRIER PROTEIN

h. ATP synthase
ATP SYNTHASE
PLASMA MEMBRANE ATPASE (PROTON PUMP)
&H+-transportingATPase
&H+-transporting ATPase
&V-ATPase
&p-ATPase

i. Alternative respiratory path
ALTERNATIVE OXIDASE

15.Reducing carriers

a.glutaredoxin
&glutaredoxin

b.gluathione
gamma-glutamyl transpeptidase-synthesis and deg of glutathione
&gamma-glutamyl transpeptidase
glutathione S-transferase-reduces peroxides,reducing agent
&glutathione S-transferase

c.thioredoxin
&thioredoxin

II. Cell Growth, Cell Division
A. Cell walls
septin
RODLET PROTEIN-spore-wall fungal hydrophobin
&RODLET PROTEIN
SPORE-WALL FUNGAL HYDROPHOBIN-not rodlet
&SPORE-WALL FUNGAL HYDROPHOBIN
glucanosyltransferase-elongation of cell wall beta(1-3)glucan
&glucanosyltransferase
INTEGRIN ALPHA CHAIN-LIKE PROTEIN-cell adhesion
&INTEGRIN ALPHA CHAIN-LIKE PROTEIN

B. Biomembranes
phosphatidylethanolamine methyltransferase
ACETYLGLUCOSAMINE-6-PHOSPHATE DEACETYLASE
DIACETYLMURAMIDASE
Glycosyltransferase-glycopeptidolipid biosyn
&Glycosyltransferase

C. Cytoskeleton, organelle biogenesis
peroxisom
KINESIN
tubulin
ankyrin
VACUOLAR ASSEMBLY PROTEIN VPS39

1. Actin-see also mitosis
#actin
PROFILIN-assembly of actin monomers
&PROFILIN
ARP2/3 COMPLEX-actin polymerization
&ARP2/3 COMPLEX
&p21-Arc
fimbrin-actin bundling
&fimbrin
COFILIN-actin binding
&cofilin
ACTIN-BINDING PROTEIN

2.Choline

choline dehydrogenase
choline kinase
chitin synthase

3. Other
phosphatidylethanolamine methyltransferase
ACETYLGLUCOSAMINE-6-PHOSPHATE DEACETYLASE
DIACETYLMURAMIDASE
Glycosyltransferase-glycopeptidolipid biosyn
&Glycosyltransferase

D. cell cycle control
cell division protein
CELL DIVISION CONTROL PROTEIN
cell cycle protein
SCH9 protein-cell progress through G1
&SCH9 protein
G1/S-SPECIFIC CYCLIN-essential for movement from g1-S
&G1/S-SPECIFIC CYCLIN
cullin-neg regulator of cell cycle
&cullin
&cell cycle control
SENESCENCE
apoptosis
&DEFENDER AGAINST CELL DEATH
&DAD-1

E. Mitosis/cytokinesis

1.MITOSIS
MITOSIS
DNA DAMAGE CHECKPOINT PROTEIN-allows entry into Mitosis
&DNA DAMAGE CHECKPOINT PROTEIN
centromere
CHROMOSOME SEGREGATION PROTEIN-with microtubule, migration of chromosomes
&CHROMOSOME SEGREGATION PROTEIN
dynamin-molecular motor, associated with mocrotuble, endocytosis
&dynamin
dynein-molecular motor
&dynein
nuclear positioning
&apsB
DMR-N9 PROTEIN-regulation of mitosis
&DMR-N9 PROTEIN
CALTRACTIN-ASSOCIATED WITH THE POLES OF THE MITOTIC SPINDLES
&CALTRACTIN

2.Cytokinesis
cytokinesis
f-actin-contractile ring during cytokinesis
&f-actin
TROPOMYOSIN-component of contractile ring

&TROPOMYOSIN

F. Meiosis
Rad9-required for meiotic chromosome condensation and synapsis
&Rad9
condensin-chromosome condensation protein
&condensin

III. DNA synthesis

A. DNA replication
DNA POLYMERASE
replication factor
Single-stranded DNA-binding protein
&SSB
&SINGLE-STRANDEDDNA-BINDING PROTEIN
DnaJ protein

B. DNA modification and repair
DNA LYASE
endonuclease IV
DNA REPAIR PROTEIN
DNA methyltransferase
cytosine C5-DNA methyltransferase

C. DNA packaging
1.Histone
Histones, class H1 (or I, or f1)
Histones, class H2a (or IIb1, or f2a2)
Histones, class H2b (or IIb2, or f2b)
Histones, class H3 (or III, or f3)
Histones, class H4 (or IV, or f2a1)
histone

2. DNA-binding
DNA-binding protein amdA

IV. Gene Expression

A. Transcription
1. RNA Polymerase
RNA POLYMERASE I, rRNA
RNA POLYMERASE II, mRNA
RNA POLYMERASE III, tRNA
RNA POLYMERASE

2. Regulation
transcription factor
SUPRESSOR OF STEM-LOOP PROTEIN-transcription initiation, pol binding
&SUPRESSOR OF STEM-LOOP PROTEIN
&SSL1
TRANSCRIPTION INITIATION FACTOR TFIID
TRANSCRIPTION INITIATION PROTEIN SPT6
RNA helicase

HELICASE MOT1-essential
&HELICASE MOT1
regulatory protein creA-carbon catabolite repression
&creA
HAC1-unfolded protein response pathway, transcrip activation
&HAC1
TRANSCRIPTIONAL REPRESSOR
QutR protein-repressor protein in the quinic acid utilization pathway, cluster
&QutR protein
QUINATE REPRESSOR
PacC-factor regulates acid and base expressed genes
&PacC

-Asexual development-Central regulatory pathway
regulatory protein brlA-transcription factor
&regulatory protein brlA
REGULATORY PROTEIN WETA
STUA transcription factor
CELL PATTERN FORMATION-ASSOCIATED PROTEIN: SPATIAL LOCALIZATION OF ABAA AND BRLA
&CELL PATTERN FORMATION-ASSOCIATED PROTEIN

3. Processing
a. SPLICEOSOME
SPLICEOSOME ASSOCIATED PROTEIN
splicing factor
small nuclear ribonucleoprotein
U4/U6 SNRNA-ASSOCIATED SPLICING FACTOR PRP24
SPLICING FACTOR U2AF 65 KD SUBUNIT
MITOCHONDRIAL RNA SPLICING PROTEIN
&MRS2 protein
tRNA splicing endonuclease
HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN K

b. polyA addition
POLYA

c. 5' capping
MRNA CAPPING ENZYME

d. Other
nucleolin-rRNA processing
&nucleolin
3'-TERMINAL PHOSPHATE CYCLASE-RNA processing?
&3'-TERMINAL PHOSPHATE CYCLASE
FIBRILLARIN
MINOR CAPSID PROTEIN C
QUEUINE TRNA-RIBOSYLTRANSFERASE-tRNA, guanine modification
&QUEUINE TRNA-RIBOSYLTRANSFERASE
&Trna-Guanine Transglycosylase

4. tRNA synthetase
&trna synthetase
&tRNA ligase

5. Degradation
ribonuclease
NONSENSE-MEDIATED MRNA DECAY PROTEIN-DECAY OF MRNAS CONTAINING PREMATURE STOP CO
DONS
&NONSENSE-MEDIATED MRNA DECAY PROTEIN
&NMD3 protein

B.Protein Biosynthesis

1. Initiation
EUKARYOTIC TRANSLATION INITIATION
INITIATION FACTOR

2. Elongation
ELONGATION FACTOR, eucaryotic and archaeal
&ELONGATION FACTOR
EF-TU
elongation factor 2
TRANSLATION FACTOR

3. Termination
PEPTIDE CHAIN RELEASE FACTOR

4. Ribosomal proteins
a.40S ribosomal protein
&40S ribosomal protein

b.60S ribosomal protein
&60S ribosomal protein
ribosomal protein

5. Post-translational modifications

a.methylation
SERINE HYDROXYMETHYLTRANSFERASE
SERINE METHYLASE

b.glycosylation
glycosylation
GPI-ANCHOR TRANSMIDASE
MNN9 PROTEIN
GUANOSINE-DIPHOSPHATASE-glycosylation
&GUANOSINE-DIPHOSPHATASE
UDP-GLUCOSE:GLYCOPROTEIN GLUCOSYLTRANSFERASE

c.myristoylization
PEPTIDE N-MYRISTOYLTRANSFERASE
&GLYCYLPEPTIDE N-TETRADECANOYLTRANSFERAS

d.other
protein disulfide-isomerase
cyclophilin

6. Folding and Targeting

a. folding

CALNEXIN HOMOLOG-folding of glycoproteins
&CALNEXIN HOMOLOG

b.chaperones

chaperone
&chaperonin
prefoldin-chaperone which delivers unfolded proteins to another chaperonin
&prefoldin
heat shock protein Hsp88
heat-shock protein30
&30 KD HEAT SHOCK PROTEIN
&HSP30
heat shock protein 70
&DNAK Protein
&HEAT SHOCK 70
HEAT SHOCK PROTEIN HSP1
HEAT SHOCK PROTEIN 104
T-COMPLEX PROTEIN-chaperone of actin, tubulin
&T-COMPLEX PROTEIN
complex I intermediate associated protein-chaperone in assembly of NADH: Ubiquin one Oxidoreductase
&complex I intermediate associated protein

c.protein sorting

protein sorting
CARBOXYPEPTIDASE Y-sorting of vacuolar protein
&CARBOXYPEPTIDASE Y
MVP1 PROTEIN-vacuolar protein sorting
&MVP1 PROTEIN
vacuolar protein sorting homolog h-vps45
clathrin
RIBOSYLATION FACTOR
SIGNAL RECOGNITION PARTICLE
MITOCHONDRIAL IMPORT RECEPTOR
synaptobrevin-protein trafficing
&synaptobrevin
VESICLE TRANSPORT V-SNARE PROTEIN
COATOMER ALPHA SUBUNIT-trafficing to golgi, nonclathrin vesicles
&COATOMER ALPHA SUBUNIT
COATOMER BETA SUBUNIT-trafficing to golgi, nonclathrin vesicles
&COATOMER BETA SUBUNIT
&beta COP
&BETA-COAT PROTEIN
&beta prime coatomer protein
COATOMER ZETA SUBUNIT-trafficing to golgi, nonclathrin vesicles
&COATOMER ZETA SUBUNIT
NPL6 PROTEIN-nuclear protein localization
&NPL6 PROTEIN

7.Turnover-protein degradation-including vacuolar

protein-L-isoaspartate O-methyltransferase-esterification for degradation
&protein-L-isoaspartate O-methyltransferase
&PROTEIN-BETA-ASPARTATEMETHYLTRANSFERASE
PROTEASE REGULATORY SUBUNIT
proteasome
ubiquitin
UBIQUITIN-CONJUGATING ENZYME
&ubiquitin-protein ligase
pepsinogen
aspartic protease
&aspartyl protease
&ASPARTIC PROTEINASE
proline peptidase
aminopeptidase
iminopeptidase
serine protease
ASPARTATEPROTEASE
metallopeptidase
&metalloproteinase
NEUTRAL PROTEASE II
ca dependent protease
alkaline protease
&ALKALINEPROTEINASE
ACID PROTEASE A
&ASPERGILLOPEPSIN II
CAAX PRENYL PROTEASE-cleavage of alpha factor for activation
&CAAX PRENYL PROTEASE
insulinase-peptidase M16 family
&insulinase
&INSULIN-DEGRADING ENZYME
&ZINC-PROTEASE C2E11.12C
#Lon serine protease
&MITOCHONDRIAL ATP-DEPENDENT PROTEASE

V. Processes

A. Cell rescue, defense, osmotic adaptation, starvation response, development

(asexual, sexual)(includes antibiotics, toxins)see also B.cell signalling, signal transduction and C. transmembrane transport

1. Development
-asexual-conidiation
velvet A
CONIDIATION-SPECIFIC PROTEIN
SpoC1-C1C protein

-asexual-pigment production
GREEN PIGMENT SYNTHASE
PORPHYRIN
porphyrinogen oxidase
polyketide synthase
LACCASE

&LACCASE
&diphenol oxidase

2.Defense
clavulanate-inhibits beta-lactamases
&clavulanate
VEGETABLE INCOMPATIBILITY PROTEIN-vegetative incompatibility,intra-species
&VEGETABLE INCOMPATIBILITY PROTEIN
&VEGETATIBLE INCOMPATIBILITY PROTEIN HET-E-1
pisatin demethylase-inactivates plant compound pisatin
&pisatin demethylase
D-AMINO ACID OXIDASE-oxidation of cephalosporin C
&D-AMINO ACID OXIDASE

-Secondary metabolites
penicill
-sterigmatocystin biosynthesis
sterigmatocystin
&sterigmatocystin
&norsolorinic acid reductase
&(U34740)

3.Detoxification
oxygen resistance
SHO1 OSMOSENSOR
catalase
super oxide dismutase
&superoxide dismutase
peroxidase
epoxide hydrolase-+ water=glycol
&epoxide hydrolase

4.Salt tolerance
HALOTOLERANCE PROTEIN HAL2
&3'(2'),5-diphosphonucleoside 3'(2') phosphohydrolase

5.Starvation response
MAK16 PROTEIN-moves cytoplasmic proteins to vacuole-autophagocytosis
&MAK16 PROTEIN

B. Cell signalling, signal transduction

1. Kinases and second messengers

a.PHOSPHATASES
PROTEIN PHOSPHATASE
ca dependent protein phosphatase
PROTEIN-TYROSINE PHOSPHATASE
serine/threonine phosphatase
phosphatase regulator

b.Kinases
protein kinase
protein kinase C

cAMP-dependent protein kinase
MAP kinase
MAP KINASE KINASE 1
MAP KINASE HOG1
CALMODULIN-DEPENDENT PROTEIN KINASE
CAMP-DEPENDENT PROTEIN KINASE
SERINE/THREONINE-PROTEIN KINASE, casein kinase acts on acidic proteins
&SERINE/THREONINE-PROTEIN KINASE
&serine/threonine protein kinase
&casein kinase
serine/threonine-specific protein kinase KIN2
&PROTEIN KINASE KIN2
histidine kinase
&histidine kinase
nitrogen permease reactivator-on for nitrogenous transport systems/s-t kinase
&nitrogen permease reactivator

c.cAMP
adenyl cyclase
ADENYLYL CYCLASE-ASSOCIATED PROTEIN-CAP protein, binds cAMP to allow activation
&ADENYLYL CYCLASE-ASSOCIATED PROTEIN

2. G proteins
GTP-binding protein
GUANINE NUCLEOTIDE-BINDING PROTEIN
RHO1 GDP-GTP EXCHANGE PROTEIN:GUANINE-NUCLEOTIDE RELEASING FACTOR
&RHO1 GDP-GTP EXCHANGE PROTEIN
rho-gdp dissociation inhibitor-prevents cycling of GDP with GTP of rho protein family
&rho-gdp dissociation inhibitor
GTPASE-ACTIVATING PROTEIN-neg regulator of Ras1, play antagonistic role with rho
-gdp dissociation inhibitor
&GTPASE-ACTIVATING PROTEIN

C. Transmembrane transport

1.Secretion
secretion
&SEC10
KEXIN-proteinase secretion
&KEXIN

2.Exoenzymes
exoenzyme
dipeptidyl peptidase-exoenzyme
&dipeptidyl peptidase

3.Transport

a. sugar transport
sugar transport
GLUCOSE TRANSPORTER

GALACTOSE TRANSPORTER
inositol transport

b.multidrug resistance
multidrug resistance
multidrug transporter
oxytetracycline exporter
cycloheximide resistance protein

c.nuclear membrane
nuclear pore membrane protein
NUCLEAR TRANSPORT FACTOR 2

d. cation transport-ATPase, or major facilitator superfamily
cation transport
CATION-TRANSPORTING ATPASE
&cation transport-ATPase
CALCIUM-TRANSPORTING ATPASE
&Ca2+-transporting ATPase
sodium transport
sulfate transporter
cobalt transporter
COPPER TRANSPORT
zinc cadmium resistance
manganese resistance

e. Anion transport
arsenite translocating ATPase-anion transport,resistance to arsenite, antimonite, arsenate
&arsenite translocating ATPase
phosphate transporter
&phosphate transport
tartrate transport
CHOLINE TRANSPORT
allantoate transport
&ALLANTOATE PERMEASE

f. Protein and amino acid
PROTEIN TRANSPORT PROTEIN
PEPTIDE TRANSPORTER
&PEPTIDE PERMEASE
AMINO ACID TRANSPORTER
AMINO-ACID PERMEASE
ARGININE PERMEASE

g. mitochondrial transport
MITOCHONDRIAL PROTEIN IMPORT PROTEIN 2
MITOCHONDRIAL IMPORT RECEPTOR SUBUNIT TOM22-translocation of
cytosolic proteins
into mitochondria
&MITOCHONDRIAL IMPORT RECEPTOR SUBUNIT TOM22
mitochondrial transport protein amc-1
MITOCHONDRIAL 2-OXOGLUTARATE/MALATE CARRIER
&2-oxoglutarate/malate translocator
benzodiazepine receptor-TRANSPORT OF PORPHYRINS AND HEME, mitochondria

&benzodiazepine receptor
ADP,ATP CARRIER PROTEIN

h. ABC transporter family
ATP-DEPENDENT PERMEASE
ABC transporter
&ABC1 transporter
&ABC-type ATPase
&ATP-DEPENDENT TRANSPORTER

i. Other
transport protein
aquaporin

D. Classes of Enzymes-not in defined pathways (Reily,1997)

1. Oxidoreductases
OXIDOREDUCTASE
MONOOXYGENASE
cytochrome P450
&cytochrome p450
CR(VI) REDUCTASE-FLAVIN OXIDOREDUCTASE family
&CR(VI) REDUCTASE
AMINE OXIDASE-OXIDATIVE DEAMINATION of AMINES
&AMINE OXIDASE
&COPPER AMINE OXIDASE
SEXUAL DIFFERENTIATION PROCESS PROTEIN-expressed during sexual diff in S. pombe
&SEXUAL DIFFERENTIATION PROCESS PROTEIN
fructosyl amine:oxygen oxidoreductase
&fructosyl amino acid oxidase
chlorocatechol 1,2-dioxygenase-degradation
&chlorocatechol 1,2-dioxygenase
cytochrome-P450
&trichothecene biosynthesis
&trichodiene oxygenase 4
&isotrichodermin C-15 hydroxylase
HYDROXYACID DEHYDROGENASE

2. Transferases
ORNITHINE AMINOTRANSFERASE

3.Hydrolases
alkaline phosphatase

4. Lyases

5. Isomerases

6. Ligases

7. Synthetases
S-adenosylmethionine synthetase

E. Non-enzymatic classes (not in defined pathways)

1. Zinc finger motif-DNA binding
zinc finger protein
&zinc-finger protein
KIN17 protein-crossreacts w recA antibody
&KIN17 protein
&protein KIN17

2. Leucine zipper motif
leucine zipper

VI. Unclassified (significant homolog but function uncertain in *Aspergillus nidulans*)
uncertain function
&unclassified
bacteriorhodopsin
LEUKOTRIENE biosynthesis
&LEUKOTRIENE A-4 HYDROLASE
bleomycin hydrolase
&BLM HYDROLASE
transposase
&transposon
prohibitin
DIPHTHINE SYNTHASE
&morphine dehydrogenase
STRESS PROTEIN
PEPTIDASE
regulatory protein
short chain dehydrogenase

VII. Unidentified (includes significant match with ORFs)
unknown function
&unclear function
&hypothetical protein

VIII. No significant homolog
NONE
-Contigs:949
-Singlets:1419

Appendix II. *Aspergillus nidulans* Unigene database in the categories of cellular functions. Members in each catagory are listed in order of homology significance. Numbers in parentheses indicate the number of Unigene members belonging to a particular category, a gene represented by both 3' and 5' Unigene members is counted only once. Words in < > are the keywords used to identify functions from the Unigene database BlastX reports and can also be category headings. There are several catagories with no members, these were left to indicate that the category was considered in preparation of the outline. Contigs beginning with a * indicate that placement in the categories was made on the basis of a homology that was not the highest match, generally done when the best homolog was an ORF with no assigned function. The first number is the homology score followed by the probability of error. The next two numbers are the Unigene member sequence homolgy endpoints. The next column contains the code indicating into which database the orignal sequence deposit was made with accession numbers specifically for that database (ex: sp|P48825|BGL1_ASPAC, in the SwissProt database with ac# BGL1_ASPAC and GenBank ac# P48825). The final column contains the description of the homolog, organism if available, and may also list the GenBank accession number. The category, Classes of Enzymes are for those enzymes not in a defined pathway and are grouped in general functions suggested by Riley (Riley, 1997). Also refer to Figure 14 and Table 12.

I. Bioenergetics and Metabolism

A. Metabolism of Carbohydrates(for glucose see energy)

1.Chitin metabolism (6)

<chitin>

Contig1860_a5f03a1.f1	2005	1.2e-206	175	1329	gnl|PID|d1013927	(D87063) chitinase [Emericella nidulans]
Contig33_v7h11a1.r1	635	1.3e-89	184	612	gnl|PID|d1013927	(D87063) chitinase [Emericella nidulans]
Contig558_o5d02a1.r1	662	2.5e-64	471	836	gnl|PID|d1013927	(D87063) chitinase [Emericella nidulans]
Contig40_i8f07a1.r1	400	1.5e-36	300	530	gnl|PID|d1013927	(D87063) chitinase [Emericella nidulans]
Contig1503_c3g10a1.r1	221	6.8e-15	27	536	gnl|PID|e220269	(Z68924) Chitinase [Clostridium thermocellum]
Contig685_u4c02a1.f1	127	8.8e-07	161	232	gnl|PID|d1013927	(D87063) chitinase [Emericella nidulans]

2.Cellulose degradation (17)

<beta glucosidase-breakdown of cellulose>

m7d01a1.r1	977	1.1e-97	7	678	sp|P48825|BGL1_ASPAC	BETA-GLUCOSIDASE 1 PRECURSOR (GENTIOBIASE) (CELLOBIASE)(BETA-D-GLUCOSIDE GLUCOHYDROLASE) >
Contig1829_d5d08a1.f1	952	5e-95	5	1951	gi|3004863	(AF029354) exo-beta-1,3-glucanase [Ampelomyces quisqualis]
m7d01a1.f1	700	2.4e-68	105	659	sp|P48825|BGL1_ASPAC	BETA-GLUCOSIDASE 1 PRECURSOR (GENTIOBIASE) (CELLOBIASE)(BETA-D-GLUCOSIDE GLUCOHYDROLASE) >
Contig1715_e9e09a1.f1	651	3.8e-63	197	850	prf||1713235A	extracellular beta glucosidase [Trichoderma reesei]
*Contig1707_b0e09a1.f1	573	7.1e-55	375	1220	sp|P15703|BGL2_Y	GLUCAN 1,3-BETA-GLUCOSIDASE PRECURSOR(EXO-1,3-BETA-GLUCANASE) (GP29) >pir||A33499 beta-1,3-exoglucana
Contig7_e9e09a1.r1	274	1.9e-40	178	507	gi|493580	(U09580) beta-D-glucoside glucohydrolase [Trichoderma

reesei]

o5g11a1.r1	369	3.4e-33	8	436	gnl|PID|e218254 (X94986) beta glucosidase [Manihot esculenta]
t2h09a1.r1	317	5e-27	8	439	gi|534844 (U13672) beta-glucosidase [Candida wickerhamii] >prf||2107160Abeta-glucosidase [Candida wi
t2h09a1.f1	307	6e-26	177	455	gi|534844 (U13672) beta-glucosidase [Candida wickerhamii] >prf||2107160Abeta-glucosidase [Candida wi
Contig608_c3g03a1.f1	288	1.1e-24	127	612	sp|P15703|BGL2_Y GLUCAN 1,3-BETA-GLUCOSIDASE PRECURSOR(EXO-1,3-BETA-GLUCANASE) (GP29) >pir||A33499 beta-1,3-exoglucana
Contig1738_c9c10a1.f1	123	6e-06	402	545	sp|P49426|EXG1_C GLUCAN 1,3-BETA-GLUCOSIDASE PRECURSOR (EXO-BETA 1,3GLUCANASE) (1,3-BETA-D-GLUCANOHYDROLASE) >gi|10665

<glucanase>

Contig1263_m7b06a1.f1	316	1.2e-26	249	620	gi|3004863 (AF029354) exo-beta-1,3-glucanase [Ampelomyces quisqualis]
Contig974_r2f09a1.f1	277	1.6e-23	169	687	gi|2326188 (U81606) mixed-linked glucanase precursor [Cochliobolus carbonum]
Contig1017_c9g12a1.r1	255	3.6e-21	145	519	gi|2326188 (U81606) mixed-linked glucanase precursor [Cochliobolus carbonum]
Contig432_e0d11a1.r1	227	1.1e-17	310	606	gi|3152652 (AF064870) endo-1,3(4)-beta-glucanase [Phaffia rhodozyma]
Contig869_z2d08a1.f1	152	6.6e-08	579	863	sp|P25358|GNS1_Y GNS1 PROTEIN >pir||S12916 probable membrane proteinYCR034w - yeast (Saccharomyces cerevisiae) >gi|449

<cellobiohydrolase>

a1g02f2.f1	387	3.6e-35	250	693	gi|912494 (U25129) cellobiohydrolase [Cochliobolus carbonum]

3.Pectin degradation

<pectate>

4.Cutin metabolism (3)

<cutin>

Contig1651_j9c08a1.f1	633	3e-61	127	711	sp|P52956|CUTI_A CUTINASE PRECURSOR (L1) >gnl|PID|d1008007 (D38311)Cutinase [Aspergillus oryzae]
Contig291_g7h01a1.r1	402	9.1e-37	6	443	gi|1438949 (U61841) cutinase G-box binding protein [Fusarium solani f. sp.pisi]
h1c07a1.r1	268	2.3e-21	28	390	sp|P52959|CT1B_FUSSO CUTINASE TRANSCRIPTION FACTOR 1 BETA >gi|1262914 (U51672)cutinase transcription factor 1 [

5. Polysaccharide synthesis (1)

<UDP-glucose dehydrogenase>

*Contig602_c3h05a1.r1	188	2.2e-11	41	730	gi|3127129 (AF061017) UDP-glucose dehydrogenase [Mus musculus]

6.Energy reserve synthesis-see also energy reserve metabolism (10)

<GLYCOGEN (STARCH) SYNTHASE>

Contig1376_c6c08a1.r1	925	3.4e-92	23	871	sp|P23337|UGS1_Y GLYCOGEN (STARCH) SYNTHASE, ISOFORM 1

```
                                                    >pir||A38326UDPglucose--starch glucosyltransferase (EC 2.4.1.11
r5h03a1.r1                734  6.1e-72   168 794    sp|P49841|KG3B_HUMAN  GLYCOGEN SYNTHASE KINASE-3 BETA (GSK-3 BETA)
                                                    >pir||S53324protein kinase - human >gi|529237,serine/threonine protein
                                                      kinase
r5h03a1.f1                405  4.5e-37   381 776    sp|P49841|KG3B_HUMAN  GLYCOGEN SYNTHASE KINASE-3 BETA (GSK-3 BETA)
                                                    >pir||S53324protein kinase - human >gi|529237,serine/threonine protein
                                                      kinase
Contig1713_c6c08a1.f1      273  4.4e-22   250 552    sp|P27472|UGS2_Y      GLYCOGEN (STARCH) SYNTHASE, ISOFORM 2
                                                    >pir||S51396UDPglucose--starch glucosyltransferase (EC 2.4.1.11
w5c03a1.r1                213  1.2e-15   329 475    sp|P23337|UGS1_YEAST  GLYCOGEN (STARCH) SYNTHASE, ISOFORM 1
                                                    >pir||A38326UDPglucose--starch glucosyltransferase (

<1,4-ALPHA-GLUCAN BRANCHING ENZYME>
Contig1640_c5a02a1.f1     1633 3.3e-167   201 1571  sp|P32775|GLGB_Y      1,4-ALPHA-GLUCAN BRANCHING ENZYME (GLYCOGEN
                                                    BRANCHINGENZYME) >pir||S50448 1,4-alpha-glucan branching

<starch branching enzyme>
Contig1191_e9a06a1.r1      430  7.3e-39   119 640    gnl|PID|e1228556      (AJ000004) starch branching enzyme II, SBE-II
                                                      [Solanumtuberosum]

<trehalose synthase>
*Contig1396_l3f08a1.f1     831  3.1e-82   103 1026  gnl|PID|d1032303      (AB010104) trehalose synthase [Grifola
                                                    frondosa]>gnl|PID|d1032304 (AB010105) trehalose synthase [Grif
l3f08a1.r1                394  3.9e-35    22 594    gnl|PID|d1032303      (AB010104) trehalose synthase [Grifola
                                                    frondosa]>gnl|PID|d1032304 (AB010105) trehalose syn
Contig1266_m7e05a1.r1      158  7.9e-08   418 768    gnl|PID|d1032303      (AB010104) trehalose synthase [Grifola
                                                    frondosa]>gnl|PID|d1032304 (AB010105) trehalose synthase [Grif


7.Arabinose metabolism (5)
<arabin>
Contig1764_e9h03a1.r1      805  1.6e-79   366 1277  sp|P42256|ABNA_A      ARABINAN ENDO-1,5-ALPHA-L-ARABINOSIDASE A
                                                    PRECURSOR(ENDO-1,5-ALPHA-L-ARABINANASE A) (ABN A) >gi|44106
n8b03a1.f1                486  1.1e-45   144 521    gnl|PID|e257620       (Z78010) (1,4)-beta-D-arabinoxylan
                                                    arabinofuranohydrolase[Aspergillus tubingensis]
Contig1676_c5f01a1.f1      264  6.5e-22   634 990    sp|P50166|ARDH_C      D-ARABINITOL 2-DEHYDROGENASE (RIBULOSE FORMING)
                                                    (ARDH)>pir||JC4041 D-arabinitol dehydrogenase (EC 1.1
r7f11a1.f1                254  4.4e-21   225 749    sp|P43066|ARDH_CANAL  D-ARABINITOL 2-DEHYDROGENASE (RIBULOSE FORMING)
                                                    (ARDH)>gi|295568 (L16227) D-arabinitol deh
n8b03a1.r1                137  8.7e-06   413 568    gnl|PID|e257619       (Z78011) (1,4)-beta-D-arabinoxylan
                                                    arabinofuranohydrolase[Aspergillus niger]

8.Glucosamine (4)
<GLUCOSAMINE-6-PHOSPHATE ISOMERASE-GLUCOSAMINE UTILIZATION PATHWAY>
i7c04a1.r1                365  7.5e-33   100 474    sp|P09375|NAGB_ECOLI  GLUCOSAMINE-6-PHOSPHATE
                                                    ISOMERASE(GLUCOSAMINE-6-PHOSPHATE DEAMINASE) >pir||MUECNG probable

<beta glucosamine>
```

```
g0a07a1.f1                184  1.8e-11    29 244   gi|2731443           (U96923) [prot= cDNA of the glycoamidase gene
                                                    [Aspergillus niger]
<GLUCOSAMINIDASE-degradation of glycans>
Contig1463_n2f07a1.r1     716  4.4e-70    13 780   sp|P43077|HEX1_C     BETA-HEXOSAMINIDASE
                                                   PRECURSOR(N-ACETYL-BETA-GLUCOSAMINIDASE)
                                                   (BETA-GLCNACASE)(BETA-N-ACETYLHEXOSAMINI
<GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE>
Contig1420_c5a03a1.r1    1107 1.8e-111    25 1008  sp|P53704|GFA1_C     GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE
                                                   (ISOMERIZING) (HEXOSEPHOSPHATE AMINOTRANSFERASE)(D-

9.Aminosugar metabolism (1)
<PHOSPHOACETYLGLUCOSAMINE MUTASE>
o8f08a1.f1                353  1.6e-31   212 655   sp|Q09687|PCM1_SCHPO  PUTATIVE PHOSPHOACETYLGLUCOSAMINE
                                                   MUTASE(ACETYLGLUCOSAMINE PHOSPHOMUTASE)=S. pombe

10.Sucrose metabolism (1)
<sucro>
<levanase-sucrose to glucose>
Contig581_c4h06a1.r1      323  8.8e-58   619 1014  prf||1404371A        levanase [Bacillus subtilis]
Contig586_c4f06a1.f1      335  2.9e-29    87 431   gnl|PID|e1254710     (AJ000493) Sucrose:Sucrose
                                                   1-Fructosyltransferase[Aspergillus foetidus]

11.Galactose metabolism (2)
<galactose>
e0g08a1.f1                400  1.4e-36   109 483   sp|P08431|GAL7_YEAST  GALACTOSE-1-PHOSPHATE URIDYLYLTRANSFERASE
                                                   >pir||XNBYUGUDPglucose--hexose-1-phosphate uridy
<alpha-1,4 polygalactosaminidase>
Contig387_f2f12a1.f1      160    8e-11   139 462   gnl|PID|d1004085     (D14846) endo alpha-1,4 polygalactosaminidase
                                                   precusor[Pseudomonas sp.]

12.Mannitol metabolism (14)
<manno>
Contig1218_f2g08a1.r1    1143 2.6e-115    40 801   gi|2407176           (AF016850) alpha-mannosidase [Emericella nidulans]
Contig1260_a5h12a1.f1     596  2.5e-57    81 635   gnl|PID|d1009247     (D49827) alpha-mannosidase [Aspergillus phoenicis]
r2h02a1.r1                518  4.6e-49   186 677   gnl|PID|e1287777     (AL022600) putative mannose-1-phosphate gaunyl
                                                   transferase[Schizosaccharomyces pombe]
Contig453_d4h02a1.f1      445  2.6e-41    66 587   gnl|PID|e1287777     (AL022600) putative mannose-1-phosphate gaunyl
                                                   transferase[Schizosaccharomyces pombe]
Contig243_h4a11a1.r1      436  1.2e-39     9 464   sp|P31382|PMT2_Y     DOLICHYL-PHOSPHATE-MANNOSE--PROTEIN
                                                   MANNOSYLTRANSFERASE 2>pir||S36711 hypothetical protein YAL023 - y
Contig115_m5e09a1.r1      369  2.7e-33    46 783   gnl|PID|e1263907     (AL022103) mannose-6-phosphate isomerase
                                                   [Schizosaccharomycespombe]
Contig1306_c9e06a1.f1     336  5.8e-29    99 683   gi|2245570           (AF005035) alpha 1,2-mannosidase [Spodoptera
```

frugiperda]

e0f01a1.f1	289	3.3e-24	65	490	sp|P53966|KTR5_YEAST PROBABLE MANNOSYLTRANSFERASE KTR5 >pir||S62941 probablemembrane protein YNL029c - yeast (S
Contig1098_a1e06c9.r1	283	1.3e-23	218	862	sp|P31723|MA12_P MANNOSYL-OLIGOSACCHARIDE ALPHA-1,2-MANNOSIDASE PRECURSOR(MAN(9)-ALPHA-MANNOSIDASE) >pir||S58766 manno
Contig244_h4a11a1.f1	151	6.6e-20	165	380	sp|P31382|PMT2_Y DOLICHYL-PHOSPHATE-MANNOSE--PROTEIN MANNOSYLTRANSFERASE 2>pir||S36711 hypothetical protein YAL023 - y
Contig1340_m5e09a1.f1	146	9.8e-07	177	431	gnl|PID|e1263907 (AL022103) mannose-6-phosphate isomerase [Schizosaccharomycespombe]
<mannitol>					
Contig1790_f0h09a1.f1	642	3.3e-62	137	1237	sp|Q45421|MTLD_B MANNITOL-1-PHOSPHATE 5-DEHYDROGENASE >gi|1480431 (U18943)mannitol-1-phosphate dehydrogenase [Bacillus
*Contig1086_i3e06a1.r1	242	3.3e-19	414	884	gi|3128349 (AF010496) mannitol 2-dehydrogenase [Rhodobacter capsulatus]
Contig1601_a5b09a1.f1	161	1.5e-10	567	809	sp|Q02418|MTLD_S MANNITOL-1-PHOSPHATE 5-DEHYDROGENASE >pir||C44798mannitol-phosphate dehydrogenase MtlD - Streptococcu

13.Xylanose metabolism (5)

<xylosidase>					
l0h06a1.f1	497	7e-47	192	479	pir||JC5034 xylan endo-1,3-beta-xylosidase (EC 3.2.1.32) - Emericella nidulans>gi|1050888 (Z49894) xyl
<xylanase>					
i2e05a1.r1	850	2.7e-84	109	600	pir||S57397 xylanase - Emericella nidulans
Contig1167_h1e02a1.f1	206	2.1e-13	289	768	sp|P54865|XYND_C ENDO-1,4-BETA-XYLANASE D PRECURSOR (XYLANASE D) (XYLD)>pir||I40712 endo-1,4-beta-xylanase (EC 3.2.1.8
<xylitol dehydrogenase>					
Contig310_g6d05a1.f1	620	7.1e-60	84	770	gi|3264834 (AF072541) xylitol dehydrogenase; XDH [Galactocandidamastotermitis]
u4g02a1.r1	372	3e-49	37	369	gi|3264834 (AF072541) xylitol dehydrogenase; XDH [Galactocandidamastotermitis]

14.Quinate metabolism (2)

<quinate-utilization is in cluster>					
o4h02a1.r1	872	1.3e-86	22	522	pir||S11944 QUTG protein - Emericella nidulans
Contig394_f1g10a1.r1	235	1.8e-17	8	397	sp|P07547|ARO1_E PENTAFUNCTIONAL AROM POLYPEPTIDE (CONTAINS:3-DEHYDROQUINATE SYNTHASE , 3-DEHYDROQUINATE DEHYDRATASE(3

15.Sorbitol metabolism (4)

<SORBITOL UTILIZATION PROTEIN>					
h1e12a1.f1	492	2.7e-46	11	565	sp|P87218|SOU2_CANAL SORBITOL UTILIZATION PROTEIN SOU2 >gi|2183242 (AF002134)Sou2p [Candida albicans]
w7c03a1.r1	331	3e-29	36	509	sp|P87219|SOU1_CANAL SORBITOL UTILIZATION PROTEIN SOU1 >gi|2183243 (AF002134)Sou1p [Candida albicans]

a0e03a1.f1	275	2.7e-23	174	476	sp|P87218|SOU2_CANAL SORBITOL UTILIZATION PROTEIN SOU2 >gi|2183242 (AF002134)Sou2p [Candida albicans]
a0e03a1.r1	250	1.2e-20	35	463	sp|P87218|SOU2_CANAL SORBITOL UTILIZATION PROTEIN SOU2 >gi|2183242 (AF002134)Sou2p [Candida albicans]
<SORBITOL DEHYDROGENASE>					
i7c03a1.f1	185	7.4e-12	148	666	sp|P35497|DHSO_YEAST SORBITOL DEHYDROGENASE (L-IDITOL 2-DEHYDROGENASE)>pir||S55941 sorbitol dehydrogenase - yea

16.Gluconate (1)

<Glucose Oxidase-first step of glucose + O2 to gluconic acid>

f2f08a1.r1	401	1.5e-36	23	367	sp|P81156|GOX_PENAG GLUCOSE OXIDASE (GLUCOSE OXYHYDRASE) (GOD)(BETA-D-GLUCOSE:OXYGEN 1-OXIDO-REDUCTASE)=Penicillum

17.Pyranose metabolism (1)

<pyranose oxidase>

y6e01a1.r1	366	2.3e-32	88	630	gnl|PID|d1011780 (D73369) pyranose oxidase [Coriolus versicolor]

18. Ribitol metabolism (1)

<ribitol kinase>

Contig831_r2d11a1.f1	541	1.8e-51	240	1055	gi|2905643 (AF045244) ribitol kinase [Klebsiella pneumoniae]

19.Calvin cycle (1)

<RIBULOSE-PHOSPHATE 3-EPIMERASE-ribulose-5 PO4 to xylulose-5 PO4>

w8f11a1.r1	330	3.7e-29	198	620	sp|P46969|RPE_YEAST RIBULOSE-PHOSPHATE 3-EPIMERASE (PENTOSE-5-PHOSPHATE3-EPIMERASE) (PPE) (RPE) >pir||S51587 P

B. Metabolism of Amino acids and Related Molecules

1.alanine metabolism (0)

2.arginine metabolism(8)

a. arginine anabolism-glutamine, co2 to arginine

<ORNITHINE CARBAMOYLTRANSFERASE>

h1a01a1.f1	753	5.7e-74	106	558	sp|P11803|OTC_EMENI ORNITHINE CARBAMOYLTRANSFERASE PRECURSOR (OTCASE)(ORNITHINE TRANSCARBAMYLASE) >gi|168017 (
h1a01a1.r1	355	8.8e-32	37	297	pir||OWASN ornithine carbamoyltransferase (EC 2.1.3.3) precursor - Emericellanidulans
<AGMATINASE>					
m0g05a1.r1	271	7.2e-23	184	465	sp|Q10088|SPEB_SCHPO PUTATIVE AGMATINASE PRECURSOR (AGMATINE UREOHYDROLASE)(AUH) >gi|1107898 (Z68166) unknown [
m0g05a1.f1	225	1.4e-17	125	418	sp|Q10088|SPEB_SCHPO PUTATIVE AGMATINASE PRECURSOR (AGMATINE UREOHYDROLASE)(AUH) >gi|1107898 (Z68166) unknown [

b. arginine catabolism-arginine to proline
<ARGINASE-also see urea cycle>
Contig1156_x7b02a1.r1 174 6e-12 17 265 gnl|PID|e1250600 (AL021815) arginase family protein [Schizosaccharomycespombe]

<ARG-6 PROTEIN>
i2h03a1.r1 667 7.3e-65 34 597 sp|P54898|AR56_NEUCR ARG-6 PROTEIN PRECURSOR (CONTAINS:N-ACETYL-GAMMA-GLUTAMYL-PHOSPHATE REDUCTASE (N-ACETYL-GL
Contig1036_d1f07a1.f1 372 1.5e-32 223 486 sp|P54898|AR56_N ARG-6 PROTEIN PRECURSOR (CONTAINS:N-ACETYL-GAMMA-GLUTAMYL-PHOSPHATE REDUCTASE (N-ACETYL-GLUTAMATESEMI

<PYRROLINE-5-CARBOXYLATE REDUCTASE>
Contig905_y7b01a1.r1 268 1.4e-22 55 522 sp|P22008|PROC_P PYRROLINE-5-CARBOXYLATE REDUCTASE (P5CR) (P5C REDUCTASE)>pir||JQ0418 pyrroline-5-carboxylate reductas
y7b01a1.f1 227 3.3e-18 59 379 sp|P22008|PROC_PSEAE PYRROLINE-5-CARBOXYLATE REDUCTASE (P5CR) (P5C REDUCTASE)>pir||JQ0418 pyrroline-5-carboxyla

3.asparagine metabolism (0)

4.aspartic acid metabolism (3)

-aspartate anabolism-oxaloacetate, glutamate to aspartate
<ASPARTATE AMINOTRANSFERASE>
g6c09a1.r1 640 5.2e-62 152 769 gi|1049345 (U39645) similar to aspartate aminotransferase [Caenorhabditiselegans]
g5e09a1.r1 512 1.9e-48 4 561 gnl|PID|e1202255 (AL009197) hypothetical aspartate aminotransferase[Schizosaccharomyces pombe]
g5e09a1.f1 305 1.8e-26 309 599 gnl|PID|e1202255 (AL009197) hypothetical aspartate aminotransferase[Schizosaccharomyces pombe]

<Aspartase>

5.cysteine metabolism (2)
<O-ACETYLHOMOSERINESULFHYDRYLASE-also methionine biosyn>
Contig1712_g8g09a1.r1 2041 1.8e-210 27 1337 gi|2605905 (AF029318) O-acetyl-L-homoserine sulfhydrylase [Emericellanidulans]

<CYSTATHIONINE GAMMA-LYASE>
Contig1760_c6e02a1.f1 1228 2.8e-124 246 1391 sp|P31373|CYS3_Y CYSTATHIONINE GAMMA-LYASE (GAMMA-CYSTATHIONASE)>pir||S31228 cystathionine gamma-lyase (EC 4.4.1.1) -

6.glutamic acid metabolism (0)

7.glutamine metabolism (2)
<GLUTAMINE SYNTHETASE>

Contig1740_c6a12a1.r1 1383 1e-140 302 1336 sp|Q12613|GLNA_C GLUTAMINE SYNTHETASE (GLUTAMATE--AMMONIA LIGASE)>gi|1322275 (L78067) glutamine synthetase [Glomerella

Contig398_m7a03a1.r1 558 2.7e-53 133 588 sp|Q12613|GLNA_C GLUTAMINE SYNTHETASE (GLUTAMATE--AMMONIA LIGASE)>gi|1322275 (L78067) glutamine synthetase [Glomerella

8.glycine metabolism (5)

a. glycolate to glycine

<GLYCERATE DEHYDROGENASE>

Contig1128_m5d12a1.r1 376 5.3e-34 133 540 sp|P40054|SERX_Y PUTATIVE D-3-PHOSPHOGLYCERATE DEHYDROGENASE YER081W(PGDH) >pir||S50584 hypothetical protein YER081w -

d2e07a1.r1 125 2.3e-06 9 233 sp|O29445|SERA_ARCFU D-3-PHOSPHOGLYCERATE DEHYDROGENASE (PGDH) >gi|2649798(AE001048) phosphoglycerate dehydroge

b. glycine catabolism

-glycine decarboxylase complex-made up of P,T,L,H-removes amino group

<GLYCINE CLEAVAGE SYSTEM H PROTEIN>

b0e07a1.f1 282 4.2e-24 112 438 gnl|PID|e1184358 (Z99120) similar to glycine cleavage system protein H[Bacillus subtilis]

<Glycine cleavage system T protein>

f1d07a1.r1 234 1.4e-18 233 448 sp|P48015|GCST_YEAST AMINOMETHYLTRANSFERASE PRECURSOR (GLYCINE CLEAVAGE SYSTEMT PROTEIN) >pir||S54642 glycine c

f1d07a1.f1 184 3e-11 200 460 gnl|PID|e339946 (Z98979) aminomethyltransferase precursor [Schizosaccharomycespombe]

9.histidine metabolism (1)

<HISTIDINE BIOSYNTHESIS>

o5a04a1.r1 365 1.4e-32 73 525 sp|P33734|HIS5_YEAST HISTIDINE BIOSYNTHESIS BIFUNCTIONAL AMIDOTRANSFERASE /CYCLASE >pir||S46125 amidotransferas

10.isoleucine metabolism (3)

<2,3-DIHYDROXYACID HYDROLYASE-4th step in iso & val biosyn>

Contig760_z5f11a1.f1 552 1.2e-52 76 603 sp|P39522|ILV3_Y DIHYDROXY-ACID DEHYDRATASE PRECURSOR (DAD) (2,3-DIHYDROXYACID HYDROLYASE) >pir||S55205 dihydroxy-acid

-catabolism

<propionyl-CoA carboxylase-also leucine and valine degradation>

n1c10a1.r1 588 1.7e-56 25 552 gnl|PID|e290075 (Y07660) B subunit of propionyl-CoA carboxylase [Mycobacteriumtuberculosis]

<methylcrotonyl-CoA carboxylase>

*Contig665_a1h06c9.r1 273 4.4e-22 222 560 gi|533707 (U12536) 3-methylcrotonyl-CoA carboxylase precursor [Arabidopsisthaliana]

11.leucine metabolism (2)

<hydroxy-3-methylglutaryl-CoA lyase-FINAL STEP OF KETOGENESIS AND LEUCINE CATABOLISM>

w8e12a1.r1	203	1.8e-25	162	422	sp|P13703|HMGL_PSEMV HYDROXYMETHYLGLUTARYL-COA LYASE (HMG-COA LYASE) (HL)(3-HYDROXY-3-METHYLGLUTARATE-COA LYASE
w8e12a1.f1	275	2.7e-23	110	469	sp|P13703|HMGL_PSEMV HYDROXYMETHYLGLUTARYL-COA LYASE (HMG-COA LYASE) (HL)(3-HYDROXY-3-METHYLGLUTARATE-COA LYASE

12.Lysine metabolism (2)

<HOMOCITRATE DEHYDRATASE>

z8g05a1.f1	215	5.9e-17	175	516	sp|P40202|LYS7_YEAST HOMOCITRATE DEHYDRATASE >pir||S50245 LYS7 protein - yeast(Saccharomyces cerevisiae) >gi|59

<SACCHAROPINE DEHYDROGENASE>

w5f07a1.r1	306	1.4e-26	26	271	sp|P38997|LYS1_YARLI SACCHAROPINE DEHYDROGENASE (NAD+, L-LYSINE FORMING)(LYSINE--2-OXOGLUTARATE REDUCTASE) (SDH

13.methionine metabolism(5)

<HOMOSERINE O-ACETYLTRANSFERASE>

Contig185_i7e07a1.f1	362	1.6e-32	400	783	sp|P12917|MET2_A HOMOSERINE O-ACETYLTRANSFERASE (HOMOSERINEO-TRANS-ACETYLASE) >pir||XYIMHA homoserine O-acetyltransfer

<cystathionine beta-lyase-3rd step>

g5d05a1.r1	749	1.6e-73	166	702	gi|1399263 (U28383) cystathionine beta-lyase [Emericella nidulans]

<methionine synthase-last step in met biosynthesis>

g1a07a1.f1	680	2.9e-66	8	766	gnl|PID|d1014526 (D89167) similar to Saccharomyces cerevisiae 5-METHYLTETRAHYDROPTEROYLTRIGLUTAMATE--HOMOCYSTEINEMETHYLTRANSFERASE
g7c01a1.r1	487	2.8e-45	188	637	gi|609350 (U15099) methionine synthase [Saccharomyces cerevisiae]
Contig712_t2d10a1.f1	417	1.3e-37	358	753	gi|2738248 (U97200) cobalamin-independent methionine synthase [Arabidopsisthaliana]

14.phenylalanine metabolism (0)

15.proline metabolism (0)

16.serine metabolism (2)

<PHOSPHOSERINE AMINOTRANSFERASE>

Contig190_j9e07a1.f1	211	5.3e-16	137	442	sp|P33330|SERC_Y PHOSPHOSERINE AMINOTRANSFERASE >pir||S42680 phosphoserinetransaminase (EC 2.6.1.52) - yeast (Saccharo
Contig128_j9e07a1.r1	200	8.1e-15	21	308	sp|P33330|SERC_Y PHOSPHOSERINE AMINOTRANSFERASE >pir||S42680 phosphoserinetransaminase (EC 2.6.1.52) - yeast (Saccharo

17. threonine metabolism (0)

18.tryptophan metabolism (2)

<anthranilate phosphoribosyltransferase-2nd step in tryp biosyn>

f5b03a1.f1 133 5.3e-06 115 291 gnl|PID|e1292700 (AL023554) anthranilate phosphoribosyltransferase[Schizosaccharomyces pombe]

<CATECHOL 1,2-DIOXYGENASE-tryp & lysine catabolism in KETOADIPATE PATHWAY>

u4d08a1.f1 268 1.4e-22 316 612 gnl|PID|d1013794 (D86544) hydroxyquinol-1, 2-dioxygenase [Ralstonia pickettii]

19.tyrosine metabolism (2)

<prephenate dehydrogenase>

Contig920_c5c08a1.r1 525 7.5e-87 466 987 gnl|PID|e1295792 (AL023776) prephenate dehydrogenase [Schizosaccharomycespombe]

Contig1454_w8d03a1.f1 200 1.2e-14 310 642 gnl|PID|e1295792 (AL023776) prephenate dehydrogenase [Schizosaccharomycespombe]

20.valine metabolism (3)

<hydroxyisobutyrate dehydrogenase>

Contig1468_r5h09a1.f1 474 2.2e-44 356 1117 sp|P29266|D3HI_R 3-HYDROXYISOBUTYRATE DEHYDROGENASE PRECURSOR (HIBADH)>pir||A32867 3-hydroxyisobutyrate dehydrogenase

*Contig16_e4a03a1.r1 130 2.4e-05 88 717 sp|P28811|MMSB_P 3-HYDROXYISOBUTYRATE DEHYDROGENASE (HIBADH) >pir||C429023-hydroxyisobutyrate dehydrogenase (EC 1.1.1.

<METHYLMALONATE-SEMIALDEHYDE DEHYD>

*Contig100_k9d06a1.f1 249 4.1e-20 224 469 sp|Q02252|MMSA_H METHYLMALONATE-SEMIALDEHYDE DEHYDROGENASE (ACYLATING)(MMSDH) >gi|188696 (M93405) methylmalonate semia

21.aromatic amino acid metabolism (2)

<PHOSPHO-2-DEHYDRO-3-DEOXYHEPTONATE ALDOLASE>

w5e04a1.r1 618 1.1e-59 8 481 sp|P79023|AROG_CANAL PHOSPHO-2-DEHYDRO-3-DEOXYHEPTONATE ALDOLASE,TYROSINE-INHIBITED (PHOSPHO-2-KETO-3-DEOXYHEPT

f0b04a1.r1 435 2.9e-40 155 592 sp|P34725|AROF_CANAL PHOSPHO-2-DEHYDRO-3-DEOXYHEPTONATE ALDOLASE,PHENYLALANINE-INHIBITED (PHOSPHO-2-KETO-3-DEOX

22.polyamine biosynthesis(3)

<polyamine>

Contig1761_c1h02a1.f1 574 4.9e-55 275 1102 gnl|PID|e1286476 (AJ002204) polyamine oxidase [Zea mays]

h4a05a1.r1 458 1e-42 13 450 sp|Q12074|SPEE_YEAST SPERMIDINE SYNTHASE (PUTRESCINE AMINOPROPYLTRANSFERASE)(SPDSY) >pir||S54090 SPE3 protein -

h1h03a1.r1 125 0.00047 135 401 gnl|PID|e1286476 (AJ002204) polyamine oxidase [Zea mays]

C. Metabolism of Nucleotides and Nucleic Acids, Purines, Pyrimidines

1. Nucleotide metabolism (5)

<NUCLEOSIDE DIPHOSPHATE KINASE>

Contig1402_m3d02a1.r1 557 2.8e-53 122 565 sp|P19804|NDKB_R NUCLEOSIDE DIPHOSPHATE KINASE B (NDK B) (NDP KINASE

B)(P18) >pir||A38369 nucleoside-diphosphate kinas

<ribose-phosphate pyrophosphokinase-purine,pyrimidine biosyn, also his and tryptophan biosyn>

t2f05a1.f1	469	7.4e-44	83	478	sp|P32895|KPR1_YEAST RIBOSE-PHOSPHATE PYROPHOSPHOKINASE 1 (PHOSPHORIBOSYLPYROPHOSPHATE SYNTHETASE 1) >pir||S305
z8e10a1.f1	419	1.5e-38	149	532	pir||S60393 ribose-phosphate pyrophosphokinase (EC 2.7.6.1) PRS1 - yeast(Candida albicans)
f0e10a1.r1	401	1.2e-36	13	618	pir||S61716 ribose-phosphate pyrophosphokinase PRPS3 homolog YOL061w - yeast(Saccharomyces cerevisiae)
t2f05a1.r1	154	1.1e-09	206	433	sp|P32895|KPR1_YEAST RIBOSE-PHOSPHATE PYROPHOSPHOKINASE 1 (PHOSPHORIBOSYLPYROPHOSPHATE SYNTHETASE 1) >pir||S305

2. Purine metabolism (12)

a. inosine mono phosphate de novo biosynthesis

<amidophosphoribosyl transferase-1st step>

Contig382_f5e07a1.r1	378	3.1e-34	25	495	sp|Q12698|PUR1_S AMIDOPHOSPHORIBOSYLTRANSFERASE (GLUTAMINEPHOSPHORIBOSYLPYROPHOSPHATE AMIDOTRANSFERASE) (ATASE) >gi|98
r5c10a1.r1	318	2.1e-27	448	795	sp|P41390|PUR1_SCHPO AMIDOPHOSPHORIBOSYLTRANSFERASE (GLUTAMINEPHOSPHORIBOSYLPYROPHOSPHATE AMIDOTRANSFERASE) (AT

<PHOSPHORIBOSYLAMINOIMIDAZOLECARBOXAMIDE FORMYLTRANSFERASE-9th step>

n1b11a1.r1	583	5.8e-56	49	450	sp|P38009|PU92_YEAST PHOSPHORIBOSYLAMINOIMIDAZOLECARBOXAMIDE FORMYLTRANSFERASE2 (AICAR TRANSFORMYLASE) / IMP CY
Contig900_t2g09a1.f1	138	1.1e-07	139	276	sp|P38009|PU92_Y PHOSPHORIBOSYLAMINOIMIDAZOLECARBOXAMIDE FORMYLTRANSFERASE2 (AICAR TRANSFORMYLASE) / IMP CYCLOHYDROLAS

b. other purine metabolic enzymes

<INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE>

Contig1754_c8d12a1.f1	1706	5.5e-175	244	1845	sp|P50094|IMH3_Y PROBABLE INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (IMPDEHYDROGENASE) (IMPDH) (IMPD) >pir||S50890 hypoth

<xanthine dehydrogenase>

m6e07a1.r1	1208	3.5e-122	74	832	sp|Q12553|XDH_EMENI XANTHINE DEHYDROGENASE (PURINE HYDROXYLASE I) >pir||A55875xanthine dehydrogenase (EC 1.1.1
m6e07a1.f1	915	1.4e-90	52	621	sp|Q12553|XDH_EMENI XANTHINE DEHYDROGENASE (PURINE HYDROXYLASE I) >pir||A55875xanthine dehydrogenase (EC 1.1.1

<ADENYLOSUCCINATE SYNTHETASE-first committed step to AMP biosyn>

s9f08a1.r1	448	1e-41	37	516	sp|P80210|PURA_YEAST ADENYLOSUCCINATE SYNTHETASE (IMP--ASPARTATE LIGASE)>pir||S48515 adenylosuccinate synthase

<Purine Nucleoside Phosphorylase>

Contig1116_d5b06a1.f1	220	8.7e-36	553	840	sp|Q05788|PNPH_Y PROBABLE PURINE NUCLEOSIDE PHOSPHORYLASE (INOSINEPHOSPHORYLASE) (PNP) >pir||S48560 hypothetical prote

<AMP DEAMINASE>

Contig992_c3h03a1.f1	1551	1.6e-158	61	1197	sp|P15274|AMDM_Y AMP DEAMINASE (MYOADENYLATE DEAMINASE) >pir||S49744 AMPdeaminase (EC 3.5.4.6) - yeast (Saccharomyces

<adenosine kinase-phosphorylates purine nucleoside>
*Contig838_y6e09a1.r1 405 4e-37 46 609 gnl|PID|e1198603 (Y15430) adenosine kinase [Physcomitrella patens]
y6e09a1.f1 145 7.1e-07 385 618 sp|P47143|ADK_YEAST PUTATIVE ADENOSINE KINASE >pir||S57126 ribokinase homologYJR105w - yeast (Saccharomyces ce

3. Pyrimidine metabolism (4)

a. de novo pyrimidine biosynthesis
<dihydroorotate dehydrogenase-4th step in pyr biosyn>
w6f03a1.r1 682 1.8e-66 81 551 gi|1181887 (U47318) dihydroorotate dehydrogenase [Emericella nidulans]
Contig138_j9d03a1.f1 430 9.5e-40 103 348 gi|1181887 (U47318) dihydroorotate dehydrogenase [Emericella nidulans]
w6f03a1.f1 232 5.8e-18 338 577 gi|1181887 (U47318) dihydroorotate dehydrogenase [Emericella nidulans]

b. other pyrimidine metabolic enzymes
<DEOXYCYTIDYLATE DEAMINASE-degeradation to dUMP>
Contig885_r7c07a1.f1 556 3.8e-67 91 549 gnl|PID|e1263904 (AL022103) deoxycytidylate deaminase [Schizosaccharomycespombe]

D. Metabolism of Lipids, Fatty Acids, Sterols-See also fatty acid degradation

1. Fatty acid biosynthesis (7)
<a. ACETYL-COA CARBOXYLASE-yields malonylcoA,comitted step to FA biosyn.>

<b. ACYL-CARRIER PROTEINS>
Contig449_d5b09a1.f1 285 2.2e-24 158 865 gnl|PID|e1185182 (Z99112) 3-ketoacyl-acyl carrier protein reductase [Bacillussubtilis]
Contig140_j7h12a1.f1 271 6.9e-23 92 433 gi|2827320 (AF042860) 3-oxoacyl-[acyl-carrier-protein]-reductase [Neurosporacrassa]

<c. FATTY ACID SYNTHASE>
Contig950_r5g09a1.f1 785 1.3e-169 777 1226 gi|1805261 (U75347) fatty acid synthase, alpha subunit [Emericella nidulans]
r7b10a1.r1 1012 3e-100 54 656 gi|1805261 (U75347) fatty acid synthase, alpha subunit [Emericella nidulans]

<d. BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE-keto acids to short branch-chain fatty acids>
Contig1741_f5h02a1.f1 824 1.6e-81 106 954 sp|P50136|ODBA_M 2-OXOISOVALERATE DEHYDROGENASE ALPHA SUBUNIT PRECURSOR(BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE C
Contig1560_i2a01a1.r1 343 1.6e-30 175 612 sp|P50136|ODBA_M 2-OXOISOVALERATE DEHYDROGENASE ALPHA SUBUNIT PRECURSOR(BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE C

e. other

<stearoyl-CoA desaturase-adds double bonds to fatty acyl coA>

j5f02a1.r1	485	1.4e-45	321	755	pir||S52746 stearoyl-CoA desaturase (EC 1.14.99.5) - Ajellomyces capsulata>gi|757862 (X85962) delta-9

2. sterols (19)

a. sterol metabolism

<sterol>

*Contig1334_g3e09a1.r1	182	2e-10	28	693	gnl|PID|e314043 (Y12693) oxysterol-binding protein [Neurospora crassa]
*Contig1364_c1g09a1.f1	133	5.4e-05	425	820	sp|Q02318|CP27_H STEROL 26-HYDROXYLASE MITOCHONDRIAL PRECURSOR (VITAMIND(3) 25-HYDROXYLASE)(5-BETA-CHOLESTANE-3-ALPHA,

<steroid monooxygenase>

d5g12a1.r1	246	1.9e-19	262	717	gnl|PID|d1025370 (AB010439) steroid monooxygenase [Rhodococcus rhodochrous]

<LANOSTEROL SYNTHASE>

Contig252_h1b11a1.f1	639	7.2e-62	129	749	sp|P38604|ERG7_Y LANOSTEROL SYNTHASE (OXIDOSQUALENE--LANOSTEROL CYCLASE)(2,3-EPOXYSQUALENE--LANOSTEROL CYCLASE) (OSC)

<glucuronidase>

t2f06a1.r1	175	1.1e-11	52	420	pir||S43555 beta-glucuronidase - Escherichia coli >gi|475169 (Z32701)beta-glucuronidase [expression ve
Contig1101_a1b06c9.r1	126	1.6e-05	72	392	gi|529326 (U12638) beta-glucuronidase [Cloning vector pdeltagusBin19]>gi|529330 (U12639) beta-glucuronidase [Cl

<HYDROXY-3-METHYLGLUTARYL-COENZYME A REDUCTASE-also mevalonate biosyn to isoprenoids>

q0d11a1.r1	423	4.8e-39	27	365	gnl|PID|e233478 (X94308) HMG-CoA-reductase [Sphaceloma manihoticola]

<C-5 STEROL DESATURASE>

q0g11a1.f1	261	9e-42	290	523	sp|P32353|ERG3_YEAST C-5 STEROL DESATURASE >pir||JQ1146 C-5 sterol desaturase(EC 1.-.-.-) - yeast (Saccharomyce
q0g11a1.r1	173	1.6e-12	213	488	gnl|PID|d1019713 (D85181) fungal sterol-C5-desaturase homolog [Homo sapiens]

<sterol demethylase>

w9e04a1.r1	617	1.6e-59	34	636	gi|2406574 (U72657) eburicol C14-alpha-demethylase [Uncinula necator]>gi|2406576 (U72658) eburicol 14
w9e04a1.f1	275	1.2e-22	244	582	gi|2406574 (U72657) eburicol C14-alpha-demethylase [Uncinula necator]>gi|2406576 (U72658) eburicol 14

b.Farnesol biosynthesis

<GERANYLGERANYL PYROPHOSPHATE SYNTHETASE>

Contig1202_g8c10a1.r1	885	6.3e-88	61	831	sp|P24322|GGPP_N GERANYLGERANYL PYROPHOSPHATE SYNTHETASE (GGPP SYNTHETASE)(DIMETHYLALLYLTRANSFERASE / GERANYLTRANSTRAN
Contig727_w6a02a1.f1	217	1.4e-16	342	536	sp|P24322|GGPP_N GERANYLGERANYL PYROPHOSPHATE SYNTHETASE (GGPP SYNTHETASE)(DIMETHYLALLYLTRANSFERASE / GERANYLTRANSTRAN

<GERANYLGERANYL TRANSFERASE>

d5a07a1.f1	306	1.4e-26	340 594	sp|P20133|BET2_YEAST TYPE II PROTEINS GERANYLGERANYLTRANSFERASE BETA SUBUNIT(TYPE II PROTEIN GERANYL-GERANYLTRA
d5a07a1.r1	289	8.9e-25	231 641	sp|P53611|PGTB_HUMAN GERANYLGERANYL TRANSFERASE TYPE II BETA SUBUNIT (RABGERANYLGERANYLTRANSFERASE BETA SUBUNIT
<hydroxysteroid dehydrogenase>				
m6f07a1.r1	305	1.6e-26	320 679	sp|P51652|PE2R_RAT 20-ALPHA-HYDROXYSTEROID DEHYDROGENASE (20-ALPHA-HSD) (HSD1)>gnl|PID|d1003827 (D14424) 20-a
g4f11a1.f1	186	1.2e-11	159 470	gnl|PID|e1254577 (AL022019) putative 3-beta-hydroxysteroid dehydrogenase[Schizosaccharomyces pombe]

c.cholesterol metabolism

<C-4 METHYL STEROL OXIDASE-cholesterol biosynthesis>				
Contig1481_a0g03a1.f1	559	2.1e-53	260 649	gi|2970627 (AF051914) C-4 methyl sterol oxidase [Candida albicans]
i0b10a1.r1	427	2.1e-39	15 470	sp|P53045|ER25_YEAST C-4 METHYL STEROL OXIDASE >pir||S64354 ERG25 protein -yeast (Saccharomyces cerevisiae) >gi
<STEROL O-ACYLTRANSFERASE-esterification of cholesterol>				
o0d02a1.r1	151	4.3e-09	16 180	sp|P53629|ARE2_YEAST STEROL O-ACYLTRANSFERASE 2 (STEROL-ESTER SYNTHASE 2)>pir||S63350 probable membrane protein
<ISOPENTENYL-DIPHOSPHATE DELTA-ISOMERASE-ISOPRENE and CHOLESTEROL BIOSYNTHESIS>				
Contig1411_c7a08a1.f1	727	3.1e-71	133 783	sp|Q10132|IPPI_S ISOPENTENYL-DIPHOSPHATE DELTA-ISOMERASE (IPP ISOMERASE)>pir||A56442 isopentenyl-diphosphate Delta-iso

3. lipids (25)

a. phospholipid metabolism

<LYSOPHOSPHOLIPASE PRECURSOR>				
o8d03a1.r1	636	1.5e-61	29 583	sp|P39457|PLB1_PENNO LYSOPHOSPHOLIPASE PRECURSOR (PHOSPHOLIPASE B)>pir||S29318 lysophospholipase (EC 3.1.1.5) -
Contig635_o8d03a1.f1	633	3e-61	155 706	sp|P39457|PLB1_P LYSOPHOSPHOLIPASE PRECURSOR (PHOSPHOLIPASE B)>pir||S29318 lysophospholipase (EC 3.1.1.5) - Penicilliu
g5d06a1.r1	611	6.6e-59	34 624	sp|P39457|PLB1_PENNO LYSOPHOSPHOLIPASE PRECURSOR (PHOSPHOLIPASE B)>pir||S29318 lysophospholipase (EC 3.1.1.5) -
<CDP-DIGLYCERIDE PYROPHOSPHORYLASE-PHOSPHOLIPID BIOSYNTHESIS>				
Contig129_j9d09a1.f1	387	3.4e-35	399 797	sp|O04940|CDS1_S PHOSPHATIDATE CYTIDYLYLTRANSFERASE (CDP-DIGLYCERIDESYNTHETASE) (CDP-DIGLYCERIDE PYROPHOSPHORYLASE) (C
*Contig1059_d0f03a1.f1	222	1.5e-14	639 872	gnl|PID|d1033240 (AB010810) phospholipase D [Candida albicans]
<phosphatidyl synthase>				
Contig1664_m3d12a1.r1	388	3.5e-35	8 691	gnl|PID|e349677 (Z99295) phosphatidyl synthase [Schizosaccharomyces pombe]
d4b08a1.f1	153	1.8e-07	207 566	gnl|PID|e349677 (Z99295) phosphatidyl synthase [Schizosaccharomyces pombe]
<PHOSPHATIDYLSERINE SYNTHASE>				

```
Contig1674_a0a05a1.f1     402  4.7e-57   498 950   sp|P08456|PSS_YE      CDP-DIACYLGLYCEROL--SERINE
                                                    O-PHOSPHATIDYLTRANSFERASE(PHOSPHATIDYLSERINE SYNTHASE) >pir||S00080CDPdiac
k0a08a1.f1                438  1.4e-40    17 550   sp|P79001|PEL1_SACPS  PUTATIVE
                                                    CDP-DIACYLGLYCEROL--SERINEO-PHOSPHATIDYLTRANSFERASE (PHOSPHATIDYLSERINE
                                                      SYNTHASE)=Saccaromyces pateurianus
Contig1324_d4b03a1.r1     412  7.5e-38   164 616   sp|P08456|PSS_YE      CDP-DIACYLGLYCEROL--SERINE
                                                    O-PHOSPHATIDYLTRANSFERASE(PHOSPHATIDYLSERINE SYNTHASE) >pir||S00080CDPdiac
k0a08a1.r1                200  1.6e-14    35 376   sp|P79001|PEL1_SACPS  PUTATIVE
                                                    CDP-DIACYLGLYCEROL--SERINEO-PHOSPHATIDYLTRANSFERASE (PHOSPHATIDYLSERINE
                                                      SYNTHASE)
<PHOSPHATIDYLSERINE DECARBOXYLASE>
m0g03a1.f1                181  3.6e-13   223 483   gi|3329153            (AE001340) Phosphatidylserine Decarboxylase [Chlamydia
                                                      trachomatis]
<ETHANOLAMINE KINASE-PHOSPHATIDYLETHANOLAMINE SYNTHESIS>
Contig1288_j4e01a1.f1     282  1.6e-23   160 756   pir||A54980           easily shocked protein - fruit fly (Drosophila
                                                    melanogaster)>gi|532128 (L35604) ethanolamine kinase [
<myo-inositol phosphate synthase-biosynthesis of inositol containing phospholipids>
g5a11a1.f1                534  9.3e-51   102 749   sp|P42803|INO1_SPIPO  MYO-INOSITOL-1-PHOSPHATE SYNTHASE (IPS)
                                                    >pir||S60302D-myo-inositol-3-phosphate synthase =Spirodela
                                                      polyrrhiza,duckweed
g5a11a1.r1                469  7.4e-44   123 638   gi|973313             (U30250) myo-inositol 1-phosphate synthase isozyme-2
                                                      [Arabidopsisthaliana]

b.  SPHINGOLIPIDS
<serine palmitoyltransferase>
g4a07a1.r1                230  8.9e-18   245 589   gnl|PID|e1285366      (AL022299) putative serine
                                                    palmitoyltransferase[Schizosaccharomyces pombe]

c.  lipopolysaccharide biosyn-biomembrane  precursors
<UDP-glucose:sterol glucosyltransferase>
v7h12a1.r1                340  1.4e-29     5 439   gnl|PID|e1169031      (Z83832) UDP-glucose:sterol glucosyltransferase
                                                      [Avenasativa]
<UDP-GLUCOSE PYROPHOSPHORYLASE>
*Contig1575_c5f09a1.f1    719  2.1e-70    64 681   sp|P32861|UDPG_Y      PROBABLE UTP--GLUCOSE-1-PHOSPHATE
                                                    URIDYLYLTRANSFERASE(UDP-GLUCOSE PYROPHOSPHORYLASE) (UDPGP) >pir||S3
x1e10a1.r1                407  2.6e-37   195 710   sp|P32861|UDPG_YEAST  PROBABLE UTP--GLUCOSE-1-PHOSPHATE
                                                    URIDYLYLTRANSFERASE(UDP-GLUCOSE PYROPHOSPHORYLASE) (UDPG

E.  Aromatic  compound  metabolism  (6)
<4-coumarate--CoA ligase-thioester substrates for phenylpropanoid biosyn>
f5b09a1.r1                311  4.2e-27    40 585   sp|P31687|4CL2_SOYBN  4-COUMARATE--COA LIGASE 2 (4CL) (CLONE
                                                    4CL16)>pir||PQ0772 4-coumarate--CoA ligase soybean pathogen resistance
                                                      response(EC 6.2.1687
```

m0f05a1.r1 261 5e-21 26 469 gi|2911799 (AF008184) 4-coumarate:CoA ligase 1 [Populus balsamifera subsp.trichocarpa X Populus delto-a fatty acid coA?

<chorismate>
Contig968_m5b08a1.f1 444 2.8e-41 10 600 sp|P32178|CHMU_Y CHORISMATE MUTASE (CM) >pir||A45921 chorismate mutase (EC5.4.99.5) - yeast (Saccharomyces cerevisiae)
n3f08a1.r1 166 6.2e-10 109 312 gi|2983461 (AE000715) chorismate mutase/prephenate dehydratase [Aquifexaeolicus]

<aminobutyrate aminotransferase>
Contig1708_c7h01a1.f1 875 6.9e-87 163 783 sp|P14010|GATA_E 4-AMINOBUTYRATE AMINOTRANSFERASE (GAMMA-AMINO-N-BUTYRATETRANSAMINASE) (GABA TRANSAMINASE) (GABA AMINO

<CARBOXYMUCONOLACTONE DECARBOXYLASE-aromatic hydrocarbon cat.>
m7e09a1.f1 207 2.1e-15 229 477 gnl|PID|e1313496 (AL031155) 3-oxoadipate enol-lactonehydrolase/4-carboxymuconolactone decarboxylase [Strept

F. Sulfur Metabolism (3)
<ADENYLYLSULFATE KINASE>
<sulphur metabolite repression-4 genes, methionine-down, no S-up>
Contig776_z1c07a1.f1 260 9.6e-22 223 372 gi|1658298 (U75874) sconCp [Emericella nidulans]

-sulfate assimilation
<sulfate adenylyltransferase-leads to biosynthesis of cys&met>
g4d04a1.r1 811 4.1e-80 23 538 pir||S55034 sulfate adenylyltransferase (EC 2.7.7.4) - Emericella nidulans>gi|572513 (X82541) sulfate
g4d04a1.f1 471 4.4e-44 150 428 pir||S55034 sulfate adenylyltransferase (EC 2.7.7.4) - Emericella nidulans>gi|572513 (X82541) sulfate

G. Phosphate Metabolism (1)
<INORGANIC PYROPHOSPHATASE>
Contig903_x1d03a1.r1 618 1e-59 15 482 sp|O13505|IPYR_P INORGANIC PYROPHOSPHATASE (PYROPHOSPHATEPHOSPHO-HYDROLASE) (PPASE) >gnl|PID|e1180018 (AJ001000) inorg

H. Nitrogen Metabolism (see also amino acid metabolism) (6)
<nitrite reductase>
x1d07a1.r1 285 4.9e-23 8 166 sp|P22944|NIR_EMENI NITRITE REDUCTASE (NAD(P)H) >pir||JH0181 nitrite reductase(NADH) (EC 1.6.6.4), long form =E. nidulans
<NITROGEN METABOLIC REGULATION PROTEIN -NEGATIVE REGULATORY PROTEIN IN THE NITROGEN CONTROL CIRCUIT>
h1a08a1.r1 622 4.5e-60 30 518 gi|3015626 (AF041976) nitrogen metabolite repression regulator NmrA[Emericella nidulans

<cyanate lyase-cyanate, bicarbonate substrates>
m3f04a1.r1 111 3.4e-05 35 160 gi|2055402 (U90436) cyanate lyase; cyanase [synthetic construct]

-urea cycle
<glutamate dehydrogenase>

```
Contig1839_c1c09a1.f1    3021 2.8e-314    12 2426  bbs|138429          (S66039) NAD(+)-specific glutamate dehydrogenase,
                                                    NAD-GDH {EC1.4.1.2} [Neurospora crassa, Peptide, 10
Contig17_x1a06a1.r1       461  4.1e-74   244 588   bbs|138429          (S66039) NAD(+)-specific glutamate dehydrogenase,
                                                    NAD-GDH {EC1.4.1.2} [Neurospora crassa, Peptide, 10
<CARBAMOYL-PHOSPHATE SYNTHASE-also arginine and pyrimidine biosynthesis>
d3g03a1.r1                752  7.5e-74    32 604   sp|P87183|CARA_TRIVE  CARBAMOYL-PHOSPHATE SYNTHASE, ARGININE-SPECIFIC,
                                                    SMALLCHAIN PRECURSOR (ARGININE-SPECIFIC C
d3g03a1.f1                288    2e-24   182 496   sp|P22572|CARA_NEUCR  CARBAMOYL-PHOSPHATE SYNTHASE, ARGININE-SPECIFIC,
                                                    SMALLCHAIN PRECURSOR (ARGININE-SPECIFIC C


I. Metabolism of Cofactors, prosthetic groups

1.nicotinamide coenzymes (5)
<NICOTINATE-NUCLEOTIDE PYROPHOSPHORYLASE-DE NOVO BIOSYNTHESIS OF NAD AND NADP>
y8c08a1.f1                298  3.9e-36    87 389   sp|P43619|NADC_YEAST  PUTATIVE NICOTINATE-NUCLEOTIDE
                                                    PYROPHOSPHORYLASE(CARBOXYLATING) (QUINOLINATE PHOSPHORIBOSY
y8c08a1.r1                182  2.6e-13   316 534   sp|P43619|NADC_YEAST  PUTATIVE NICOTINATE-NUCLEOTIDE
                                                    PYROPHOSPHORYLASE(CARBOXYLATING) (QUINOLINATE PHOSPHORIBOSY
<kynureninase-biosyn of NAD cofactors>
Contig1362_c5h01a1.f1     525  8.4e-50   194 883   pir||S59898           kynureninase (EC 3.7.1.3) - rat >bbs|171864
                                                    kynureninase,L-kynurenine hydrolase {EC 3.7.1.3} [rats, 1
Contig539_c7e07a1.f1      516  7.5e-49    71 1006  gi|1050752            (Z50144) kynurenine/alpha-aminoadipate
                                                    aminotransferase [Rattusnorvegicus]
c5h01a1.r1                384  7.5e-35   138 767   pir||S59898           kynureninase (EC 3.7.1.3) - rat >bbs|171864
                                                    kynureninase,L-kynurenine hydrolase {EC 3.7.1.

2.biocytin (biotin) (0)

3.thiamine (4)
<thiamine>
Contig146_m6c09a1.f1      410  1.3e-37   321 728   gnl|PID|d1019649      (D45894) thiamine-4 [Neurospora crassa]
<THIAMIN BIOSYNTHESIS>
Contig1797_d4c08a1.f1    1666 9.5e-171   145 1167  sp|P42882|NMT1_A      NMT1 PROTEIN HOMOLOG >pir||S53697 nmt1 protein
                                                    -Aspergillus parasiticus >gi|557050 (U15196) the expre
Contig318_g5h01a1.r1      231  7.3e-18    81 515   sp|P40386|THI4_S      PROBABLE THIAMIN BIOSYNTHETIC BIFUNCTIONAL
                                                    ENZYME(CONTAINS: THIAMIN-PHOSPHATE PYROPHOSPHORYLASE (TMPP
Contig319_g5h01a1.f1      164  1.4e-10   125 463   sp|P40386|THI4_S      PROBABLE THIAMIN BIOSYNTHETIC BIFUNCTIONAL
                                                    ENZYME(CONTAINS: THIAMIN-PHOSPHATE PYROPHOSPHORYLASE (TMPP

4.coenzyme A (2)
<acetyl-coenzyme A synthetase>
Contig1515_f2c11a1.f1    1316 1.3e-133   442 1209  sp|P16928|ACSA_E      ACETYL-COENZYME A SYNTHETASE (ACETATE--COA
```

```
                                                                    LIGASE)(ACYL-ACTIVATING ENZYME) >pir||SYASAA acetate--CoA
y4b06a1.r1                469  2.4e-69     8 292   sp|P16928|ACSA_EMENI  ACETYL-COENZYME A SYNTHETASE (ACETATE--COA
                                                    LIGASE)(ACYL-ACTIVATING ENZYME) >pir||SYASAA ac

<acetyl coenzyme A acetyltransferase>

5.flavins  (2)
<riboflavin synthase>
Contig102_k8h04a1.f1      269  1.1e-22   151 390   sp|P50861|RIB4_Y      6,7-DIMETHYL-8-RIBITYLLUMAZINE SYNTHASE (DMRL
                                                   SYNTHASE)(LUMAZINE SYNTHASE) (RIBOFLAVIN SYNTHASE BETA

<GTP cyclohydrolase II-riboflavin biosyn>
g6e06a1.r1                508  5.1e-48   137 736   gnl|PID|e1291629      (AL023287) GTP cyclohydrolase II [Schizosaccharomyces
                                                    pombe]

6.folate-methyl donor  (5)
<folate>
Contig1825_z4c10a1.f1     306  1.4e-26   283 717   gi|2565196            (AF000381) non-functional folate binding protein [Homo
                                                    sapiens]
Contig1487_e9g09a1.f1     264  3.9e-22   254 508   gi|2565196            (AF000381) non-functional folate binding protein [Homo
                                                    sapiens]
Contig1786_a0h08a1.r1     235  4.7e-19   286 507   gi|2565196            (AF000381) non-functional folate binding protein [Homo
                                                    sapiens]
Contig148_j4g02a1.f1      234  9.9e-18   281 628   sp|P28037|FTDH_R      10-FORMYLTETRAHYDROFOLATE DEHYDROGENASE (FBP-CI)
                                                   >gi|908915(M59861) 10-formyltetrahydrofolate dehydro
Contig1124_i0c01a1.f1     182  1.3e-12   276 431   gi|2565196            (AF000381) non-functional folate binding protein [Homo
                                                    sapiens]

7. heme  (3)
<heme>
Contig669_a1c01c9.r1      270  9.4e-23   263 517   gnl|PID|e1284430      (AL022245) ferrochelatase [Schizosaccharomyces
                                                   pombe]=(PROTOHEME FERRO-LYASE) (HEMESYNTHETASE)

<siroheme synthase>
c5c01a1.r1                263  4.9e-22    85 600   gi|2983676            (AE000730) siroheme synthase [Aquifex aeolicus]
-iron uptake
<FERRIC REDUCTASE TRANSMEMBRANE COMPONENT 2>
o5f06a1.r1                168  8.3e-11    17 505   sp|P36033|FRE2_YEAST  FERRIC REDUCTASE TRANSMEMBRANE COMPONENT 2
                                                   PRECURSOR>pir||S38063 ferric reductase FRE2 pre

8.PANTOTHENATE  (1)
<PANTOTHENATE SYNTHETASE>
g2d12a1.f1                279  9.8e-24   122 715   sp|P40459|PANC_YEAST  PUTATIVE PANTOATE--BETA-ALANINE LIGASE
                                                   (PANTOTHENATESYNTHETASE) (PANTOATE ACTIVATING ENZYM
g2d12a1.r1                198  3.8e-15   232 420   sp|P56061|PANC_HELPY  PANTOATE--BETA-ALANINE LIGASE (PANTOTHENATE
                                                   SYNTHETASE)(PANTOATE ACTIVATING ENZYME) >gi|23
```

9.Molybdopterin (3)
<molybdopterin biosynth>

c9e12a1.rl	217	2.1e-16	208	564	gnl|PID|e349592 (Z99258) molybdopterin biosynthesis [Schizosaccharomycespombe]
Contig567_c5e06a1.fl	166	9.1e-12	131	433	gi|2984359 (AE000776) molybdopterin converting factor subunit 2 [Aquifexaeolicus]
Contig265_m8c12a1.fl	138	3e-06	410	745	sp|P12281|MOEA_E MOLYBDOPTERIN BIOSYNTHESIS MOEA PROTEIN >pir||A32352molybdopterin-converting factor chlE - Escherichi

J. Energy
1. Glycolysis (16)

<a.hexokinase>

n0c07a1.rl	139	1.5e-07	220	429	sp|P27926|HXK3_RAT HEXOKINASE TYPE III (HK III) >pir||S13913 hexokinase (EC2.7.1.1) III - rat >gi|1658068 (U7

<b.glucose-6-phosphate isomerase>

<c.fructose-6-phosphate2-kinase>

Contig1540_c3b10a1.rl	607	6.8e-103	513	974	sp|P32604|F26_YE FRUCTOSE-2,6-BISPHOSPHATASE >pir||S56938fructose-2,6-bisphosphate 2-phosphatase (EC 3.1.3.46) - yeast
Contig1737_m8e11a1.fl	679	4.1e-66	388	1233	sp|P40433|6P21_Y 6-PHOSPHOFRUCTO-2-KINASE 1 (PHOSPHOFRUCTOKINASE 2 I)(6PF-2-K 1) >pir||S48465 6-phosphofructo-2-kinase
Contig1005_c1e06a1.fl	620	7.5e-60	2	742	sp|P32604|F26_YE FRUCTOSE-2,6-BISPHOSPHATASE >pir||S56938fructose-2,6-bisphosphate 2-phosphatase (EC 3.1.3.46) - yeast
g0e07a1.fl	222	1.8e-16	219	449	gi|172136 (M80801) 6-phosphofructo-2-kinase [Saccharomyces cerevisiae]

<d.fructose-bisphosphate aldolase-also gluconeogenesis, PP cycle, carbon fixation, fructose and mannose metab>

Contig455_d4g02a1.rl	600	8.3e-58	113	682	sp|P14540|ALF_YE FRUCTOSE-BISPHOSPHATE ALDOLASE >pir||ADBY2fructose-bisphosphate aldolase (EC 4.1.2.13) II - yeast(Sac
c5b07a1.rl	251	7.8e-21	112	381	sp|P36580|ALF_SCHPO FRUCTOSE-BISPHOSPHATE ALDOLASE >gnl|PID|d1004756 (D17415)fructose 1,6-bisphosphate aldolas

<e.triose-phosphate isomerase>

m7h07a1.fl	468	8.4e-44	217	492	sp|P04828|TPIS_EMENI TRIOSEPHOSPHATE ISOMERASE (TIM) >pir||ISASTNtriose-phosphate isomerase (EC 5.3.1.1) - Emer

<f.glyceraldehyde-3-phosphate dehydrogenase>

Contig1588_c5c07a1.rl	987	9e-99	44	622	pir||DEASG3 glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12)

-Emericella nidulans >gi|168049 (M19694) glyce

Contig1637_a5b02a1.f1	569	1.8e-54	212	739	pir||DEASG3 glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12)

-Emericella nidulans >gi|168049 (M19694) glyce

Contig570_c5d10a1.f1	380	1.9e-34	118	537	gi|551682 (L07497) glyceraldehyde-3-phosphate dehydrogenase [Anabaenavariabilis]

<g.phosphoglycerate kinase>

<h.phosphoglycerate mutase>

z3b09a1.r1	626	1.6e-60	11	754	gi|2773203 (AF039713) Similar to phosphoglycerate mutase; coded for by C.elegans cDNA yk357d11.5; cod
Contig1282_h1h12a1.f1	341	2.5e-30	222	443	sp|Q12560|ENO_AS ENOLASE (2-PHOSPHOGLYCERATE DEHYDRATASE)(2-PHOSPHO-D-GLYCERATE HYDRO-LYASE) >pir||JC45426beta-hydroxy

<i.phosphopyruvate hydratase>

<j.pyruvate kinase>

Contig176_i8e01a1.r1	850	3.1e-84	344	934	sp|P22360|KPYK_E PYRUVATE KINASE (PK) >pir||S27364 pyruvate kinase (EC2.7.1.40) - Emericella nidulans >gi|168074 (M369
o5g10a1.r1	821	3.7e-81	11	499	sp|P22360|KPYK_EMENI PYRUVATE KINASE (PK) >pir||S27364 pyruvate kinase (EC2.7.1.40) - Emericella nidulans >gi|1
Contig972_i8e01a1.f1	321	9.1e-28	379	555	sp|P22360|KPYK_E PYRUVATE KINASE (PK) >pir||S27364 pyruvate kinase (EC2.7.1.40) - Emericella nidulans >gi|168074 (M369

2. Gluconeogenesis (6)

<a.LACTATE DEHYDROGENASE>

<b.pyruvate carboxylase>

h1c05a1.r1	433	8.4e-39	24	350	sp|P32327|PYC2_YEAST PYRUVATE CARBOXYLASE 2 (PYRUVIC CARBOXYLASE 2) (PCB 2)>pir||S46094 pyruvate carboxylase (E
Contig1375_d1d10a1.f1	407	5.1e-36	326	712	gnl|PID|d1011901 (D78170) pyruvate carboxylase [Schizosaccharomyces pombe]

<c.phosphoenolpyruvate carboxykinase>

Contig1602_d2g02a1.r1	779	2.9e-101	164	808	gi|2738107 (U88575) phosphoenolpyruvate carboxykinase [Kluyveromyces lactis]
Contig1620_d2g02a1.f1	659	7.9e-89	484	957	gi|2738107 (U88575) phosphoenolpyruvate carboxykinase [Kluyveromyces lactis]

<d.FRUCTOSE-1,6-BISPHOSPHATASE>

Contig292_g7g01a1.r1	544	7.2e-52	253	855	sp|P09201|F16P_Y FRUCTOSE-1,6-BISPHOSPHATASE (D-FRUCTOSE-1,6-BISPHOSPHATE1-PHOSPHOHYDROLASE) (FBPASE) >pir||PABY fruct
x8e05a1.f1	387	3.6e-35	171	554	sp|P09201|F16P_YEAST FRUCTOSE-1,6-BISPHOSPHATASE

(D-FRUCTOSE-1,6-BISPHOSPHATE1-PHOSPHOHYDROLASE) (FBPASE) >pir|

3. Pentose-phosphate pathway (9)
<a.glucose-6-phosphate dehydrogenase>

n3f11a1.r1	1132	4.1e-114	31	681	sp|P41764|G6PD_EMENI GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (G6PD)>gnl|PID|e99568 (X77830) glucose-6-phosphate 1-d
Contig619_c2a06a1.f1	737	2.7e-72	227	685	sp|P41764|G6PD_E GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (G6PD)>gnl|PID|e99568 (X77830) glucose-6-phosphate 1-dehydrogenas

<b. gluconeolactonase>

<c.phosphogluconate dehydrogenase>

m2g12a1.f1	503	1.7e-47	190	597	sp|P53319|6PG2_YEAST 6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING 2>pir||S64588 phosphogluconate dehydroge
n8f05a1.f1	213	5.8e-16	120	518	gi|3322609 (AE001213) phosphogluconate dehydrogenase (gnd) [Treponemapallidum]

<d.ribose 5-phosphate isomerase-nonoxidative PO4>

Contig1751_d4f10a1.f1	276	1.8e-23	121	570	gi|2983605 (AE000725) ribose 5-phosphate isomerase B [Aquifex aeolicus]

<e.ribulose-phosphate 3-epimerase>

<f.transketolase>

y4a09a1.r1	549	2.4e-52	19	468	sp|Q12630|TKT1_KLULA TRANSKETOLASE (TK) >gi|1488336 (U65983) transketolase[Kluyveromyces lactis]
m0d06a1.r1	482	2.6e-45	11	448	sp|Q12630|TKT1_KLULA TRANSKETOLASE (TK) >gi|1488336 (U65983) transketolase[Kluyveromyces lactis]
Contig958_h4a05a1.f1	297	9.5e-23	998	1267	sp|Q12630|TKT1_K TRANSKETOLASE (TK) >gi|1488336 (U65983) transketolase[Kluyveromyces lactis]

<g.transaldolase>

Contig1698_d4f04a1.f1	1097	1.6e-110	142	1107	gnl|PID|e1292580 (AL023518) Tal1p transaldolase [Schizosaccharomyces pombe]

4. Pyruvate dehydrogenase (5)
<pyruvate dehydrogenase>

o0a06a1.r1	609	1e-58	17	466	sp|P16387|ODPA_YEAST PYRUVATE DEHYDROGENASE E1 COMPONENT, ALPHA SUBUNITPRECURSOR (PDHE1-A) >gi|298059 (X71664)
y3g01a1.f1	489	5e-46	177	680	gi|2623175 (AF030425) pyruvate dehydrogenase E1 component alpha subunit[Pichia stipitis]

<pyruvate decarboxoylase>
<dihydrolipoyl transacetylase>

Contig1423_f5f03a1.f1 549 2.4e-52 77 811 sp|P53395|ODB2_M LIPOAMIDE ACYLTRANSFERASE COMPONENT (E2) PRECURSOR OFBRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE COM

<dihydrolipoamide dehydrogenase>
Contig1537_h8f04a1.r1 1470 5.6e-150 9 1370 gi|1911177 (L40360) dihydrolipoamide dehydrogenase [Schizosaccharomyces pombe]

<pyruvate dehydrogenase kinase-inhibits pyruvate dehyd by phos of E1 alpha subunit>
v1e08a1.r1 169 1.1e-09 47 598 gnl|PID|e1285215 (AJ001418) pyruvate dehydrogenase kinase-like protein [Musmusculus]

5. Tricarboxylic acid pathway (18)

<a.citrate synthase>
d3a06a1.r1 952 4.5e-95 43 636 sp|O00098|CISY_EMENI CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR >gi|2138332(U89675) citrate synthase [Emericella
Contig1245_d5a04a1.r1 607 1.7e-58 84 821 sp|P79024|CISY_C CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR>gnl|PID|d1020163 (AB001565) citrate synthase [Candida tropi
d3a06a1.f1 428 1.6e-39 196 483 sp|O00098|CISY_EMENI CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR >gi|2138332(U89675) citrate synthase [Emericella
Contig1092_d5a04a1.f1 358 3.7e-32 133 459 sp|P43635|CISX_Y CITRATE SYNTHASE 3 >pir||S52814 citrate (si)-synthase (EC4.1.3.7) - yeast (Saccharomyces cerevisiae)

<b.aconitate hydratase>
Contig1726_c1g04a1.f1 1464 2.5e-149 45 1178 sp|P19414|ACON_Y ACONITATE HYDRATASE, MITOCHONDRIAL PRECURSOR (CITRATEHYDRO-LYASE) (ACONITASE) >pir||S50387 aconitate

<c.isocitrate dehydrogenase>
Contig119_k0c02a1.f1 528 3.4e-50 183 515 sp|P79089|IDHP_A ISOCITRATE DEHYDROGENASE (NADP), MITOCHONDRIAL PRECURSOR(OXALOSUCCINATE DECARBOXYLASE) (IDH) (NADP+-S
k0c02a1.r1 469 6.5e-44 4 342 sp|P79089|IDHP_ASPNG ISOCITRATE DEHYDROGENASE (NADP), MITOCHONDRIAL PRECURSOR(OXALOSUCCINATE DECARBOXYLASE) (ID
Contig1171_e7a09a1.f1 437 1.7e-40 125 736 gi|1182011 (X87172) NAD+-isocitrate dehydrogenase, alpha subunit [Macacafascicularis]
Contig543_c7b05a1.f1 367 4.6e-33 234 464 gi|2266941 (AF009036) NAD(+)-isocitrate dehydrogenase subunit I [Ajellomycescapsulatus]
n0h09a1.r1 327 8.1e-29 141 452 gi|2266941 (AF009036) NAD(+)-isocitrate dehydrogenase subunit I [Ajellomycescapsulatus]

<d.alpha-ketoglutarate dehydrogenase>

<e.SUCCINYL-COA LIGASE>
c3f09a1.r1 771 7e-76 18 755 sp|P53587|SUCB_NEOFR SUCCINYL-COA LIGASE (GDP-FORMING), BETA-CHAIN PRECURSOR(SUCCINYL-COA SYNTHETASE, BETA CHAI
i7e06a1.r1 452 4.5e-42 17 325 sp|O13750|SUCA_SCHPO PROBABLE SUCCINYL-COA LIGASE (GDP-FORMING), ALPHA-CHAINPRECURSOR (SUCCINYL-COA SYNTHETASE,

l3e03a1.r1	334	1.5e-29	125	433	sp|O13750|SUCA_SCHPO PROBABLE SUCCINYL-COA LIGASE (GDP-FORMING), ALPHA-CHAINPRECURSOR (SUCCINYL-COA SYNTHETASE,
<f.succinate dehydrogenase>					
<g.FUMARATE HYDRATASE>					
Contig1327_f1h10a1.f1	575	1.4e-98	536	982	sp|P08417|FUMH_Y FUMARATE HYDRATASE, MITOCHONDRIAL PRECURSOR (FUMARASE)>pir||UFBYM fumarate hydratase (EC 4.2.1.2) pre
<h.malate dehydrogenase>					
Contig1670_e9b05a1.f1	870	2.2e-86	78	1046	sp|O02640|MDHM_C PROBABLE MALATE DEHYDROGENASE, MITOCHONDRIAL PRECURSOR>gi|2076896 (AF002197) Similar to malate dehydr
h8e02a1.r1	629	7.7e-61	31	492	bbs|179680 (S83228) beta-isopropylmalate dehydrogenase [Aspergillus niger,strain A733, Peptide, 363 a
m8h04a1.r1	518	4e-49	127	612	sp|P17505|MDHM_YEAST MALATE DEHYDROGENASE, MITOCHONDRIAL PRECURSOR>pir||DEBYMM malate dehydrogenase (EC 1.1.1.3
h8e02a1.f1	213	4.2e-28	139	306	bbs|179680 (S83228) beta-isopropylmalate dehydrogenase [Aspergillus niger,strain A733, Peptide, 363 a
6. related reactions (6)					
<citrate lyase-citrate to oxaloacetate+acetylcoA>					
Contig1615_d5a08a1.r1	957	1.3e-95	242	1066	gnl|PID|e349683 (Z99295) citrate lyase [Schizosaccharomyces pombe]
k9f05a1.f1	562	9.8e-53	10	582	sp|P16638|ACLY_RAT ATP-CITRATE (PRO-S-)-LYASE (CITRATE CLEAVAGE ENZYME)>pir||A35007 ATP citrate (pro-S)-lyase
x7b07a1.r1	424	7.3e-49	32	472	gnl|PID|e349683 (Z99295) citrate lyase [Schizosaccharomyces pombe]
j0h12a1.f1	353	1.3e-31	213	458	gnl|PID|d1014552 (D89194) similar to Rat ATP citrate-lyase, SWISS-PROTAccession Number P16638 [Schizosaccha
<carbonic anhydride>					
Contig1792_c3a07a1.r1	542	1.3e-51	165	788	gnl|PID|d1013665 (D86050) carbonic anhydrase [Porphyridium purpureum]
n0d03a1.f1	220	1.3e-16	108	452	gnl|PID|d1013666 (D86051) carbonic anhydrase [Porphyridium purpureum]
7. glyoxylate cycle (5)					
<malate synthase>					
o6e06a1.r1	931	6.9e-93	1	534	sp|P28344|MASY_EMENI MALATE SYNTHASE, GLYOXYSOMAL >pir||S17773 malate synthase(EC 4.1.3.2) - Emericella nidulan
o6h07a1.r1	718	2.9e-70	82	552	sp|P28344|MASY_EMENI MALATE SYNTHASE, GLYOXYSOMAL >pir||S17773 malate synthase(EC 4.1.3.2) - Emericella nidulan
Contig290_g8a02a1.f1	487	8.2e-46	190	480	sp|P28344|MASY_E MALATE SYNTHASE, GLYOXYSOMAL >pir||S17773 malate synthase(EC 4.1.3.2) - Emericella nidulans >gi|2702
<isocitrate lyase>					
Contig1259_h0h01a1.r1	1396	3.8e-142	6	866	pir||S26857 isocitrate lyase (EC 4.1.3.1) - Emericella nidulans
Contig1173_h0h03a1.f1	839	3.8e-83	109	585	sp|P28298|ACEA_E ISOCITRATE LYASE (ISOCITRASE) (ISOCITRATASE)

(ICL)>gi|2317 (X62696) isocitrate lyase [Emericella nidu

8. Fermentation, alcoholic (15)
<a.pyruvate decarboxylase>

Contig1526_d2a11a1.f1	731	4.2e-113	497	940	sp|P87208|DCPY_E PYRUVATE DECARBOXYLASE >gi|2160688 (U73194) pyruvatedecarboxylase [Emericella nidulans]

<b.alcohol dehydrogenase>

Contig1832_a1a03c9.r1	1694	1e-173	60	1109	sp|P08843|ADH1_E ALCOHOL DEHYDROGENASE I >pir||A29054 alcoholdehydrogenase (EC 1.1.1.1) - Emericella nidulans >gi|1680
Contig848_z3d06a1.r1	783	4.2e-101	17	538	sp|P07754|ADH3_E ALCOHOL DEHYDROGENASE III >pir||A24648 alcoholdehydrogenase (EC 1.1.1.1) III - Emericella nidulans >g
Contig354_g3d03a1.f1	605	2.7e-58	73	810	gi|1790870 (U32622) toluenesulfonate zinc-independent alcohol dehydrogenase[Comamonas testosteroni]
Contig1275_g2b06a1.f1	540	2.2e-51	144	875	gi|1790870 (U32622) toluenesulfonate zinc-independent alcohol dehydrogenase[Comamonas testosteroni]
z5a05a1.f1	492	2.6e-46	133	444	sp|P07754|ADH3_EMENI ALCOHOL DEHYDROGENASE III >pir||A24648 alcoholdehydrogenase (EC 1.1.1.1) III - Emericella
j5d06a1.r1	423	5.2e-39	48	833	sp|P00332|ADH_SCHPO ALCOHOL DEHYDROGENASE >pir||DEZPA alcohol dehydrogenase(EC 1.1.1.1) - fission yeast (Schiz
Contig1703_c4c04a1.r1	361	1.8e-32	50	997	gnl|PID|e209885 (X92868) NADP-dependent alcohol dehydrogenase [Bacillussubtilis] >gnl|PID|e1183931 (Z99117) NADP-depe
Contig186_i7c03a1.r1	337	6.5e-30	28	627	sp|P39714|YAG0_Y HYPOTHETICAL ZINC-TYPE ALCOHOL DEHYDROGENASE-LIKE PROTEININ GDH3-CNE1 INTERGENIC REGION >pir||S51962
*Contig411_e9g02a1.r1	239	1.6e-19	28	507	gi|1934622 (U93875) alcohol dehydrogenase [Bacillus subtilis]
c4b11a1.r1	230	3.3e-18	187	648	sp|P39714|YAG0_YEAST HYPOTHETICAL ZINC-TYPE ALCOHOL DEHYDROGENASE-LIKE PROTEININ GDH3-CNE1 INTERGENIC REGION >p
Contig1508_d4d04a1.r1	170	1.5e-09	94	675	sp|P42328|ADH3_B ALCOHOL DEHYDROGENASE (ADH-HT) >pir||S45605 alcoholdehydrogenase (EC 1.1.1.1) - Bacillus stearothermo
Contig406_f0c08a1.f1	135	2.6e-06	384	665	gi|1118145 (U41749) strong similarity to the insect-type alcoholdehydrogenase/ribitol dehydrogenase family [Caen
w9c01a1.r1	122	6.3e-05	348	566	gnl|PID|e1183501 (Z99113) similar to alcohol dehydrogenase [Bacillus subtilis]>gnl|PID|e1185316 (Z99114) si
j7g12a1.f1	123	0.00013	127	417	sp|P42327|ADH2_BACST ALCOHOL DEHYDROGENASE (ADH) >pir||S47643 alcoholdehydrogenase (EC 1.1.1.1) - Bacillus stea

9.Fermentation, other (2)

<LACTATE DEHYDROGENASE-pyruvate to lactate>

r4g05a1.r1	386	4.4e-35	113	655	sp|P52643|LDHD_ECOLI D-LACTATE DEHYDROGENASE (D-LDH) >gi|1049265 (U36928)D-lactate dehydrogenase [Escherichia c

<butanediol dehydrogenase>

Contig1693_a5d12a1.f1 438 1.3e-40 145 945 gnl|PID|d1013772 (D86412) meso-2,3-butanediol dehydrogenase (D-acetoinforming) [Klebsiella pneumoniae]

10. Monocarbon metabolism (2)

<formate dehydrogenase>
Contig1117_d5c06a1.f1 917 2.2e-91 143 748 sp|Q07103|FDH_NE FORMATE DEHYDROGENASE (NAD-DEPENDENT FORMATEDEHYDROGENASE) (FDH) >pir||A47117 formate dehydrogenase (

C1 Metabolism
<ALCOHOL OXIDASE-first step-methanol utilization to FORMALDEHYDE>
x8a04a1.r1 174 1.6e-11 1 528 sp|P04841|ALOX_PICAN ALCOHOL OXIDASE (AOX) (METHANOL OXIDASE) (MOX)>pir||OXHQAP alcohol oxidase (EC 1.1.3.13) -

11. Metabolism of energy reserves (glycogen, starch, trehalose) (19)

a. Glycogen degradation

<glycogen phosphorylase>
Contig122_j9h10a1.r1 519 2.1e-48 37 594 pir||S61144 glycogen phosphorylase (EC 2.4.1.1) - yeast (Saccharomycescerevisiae) >gi|849168 (U28371) Glycogen ph
Contig1329_c4e04a1.r1 509 2.7e-47 28 504 pir||S61144 glycogen phosphorylase (EC 2.4.1.1) - yeast (Saccharomycescerevisiae) >gi|849168 (U28371) Glycogen ph
<PHOSPHOGLUCOMUTASE-glycogen deg, glu1PO4 to glu6PO4>
Contig519_c8e10a1.f1 431 7e-40 84 722 sp|P37012|PGM2_Y PHOSPHOGLUCOMUTASE 2 (GLUCOSE PHOSPHOMUTASE 2) (PGM 2)>pir||S41200 phosphoglucomutase (EC 5.4.2.2) PG
<GLYCOGEN DEBRANCHING ENZYME>
*Contig1382_c4a08a1.f1 867 5.6e-85 148 1257 sp|P35574|GDE_RA GLYCOGEN DEBRANCHING ENZYME (4-ALPHA-GLUCANOTRANSFERASE(OLIGO-1,4-1,4-GLUCANTRANSFERASE) / AMYLO-1,6-

b. Starch degradation

<alpha glucosidase>
l3h03a1.r1 621 2.3e-59 30 539 sp|Q12558|AGLU_ASPOR ALPHA-GLUCOSIDASE PRECURSOR (MALTASE) (AGL) >pir||JC4217alpha-glucosidase (EC 3.2.1.20) -
r5e04a1.r1 528 2.3e-49 5 652 sp|O04893|AGLU_SPIOL ALPHA-GLUCOSIDASE PRECURSOR (MALTASE) >gnl|PID|d1020713(D86624) alpha-glucosidase precours, glucoamylase
o4g07a1.r1 278 6.6e-23 41 439 pir||S44188 alpha-glucosidase (EC 3.2.1.20) - Staphylococcus xylosus>gi|474177 (X78853) alpha-D-1,4-gl
l0f02a1.f1 167 1.6e-10 85 480 sp|P38138|YB79_YEAST PUTATIVE FAMILY 31 GLUCOSIDASE IN PCS60-ABD1 INTERGENICREGION >pir||S46105 glucan 1,4-alpha=glucosidase

<glucoamylase>
Contig1841_a1e02c9.r1 1195 5.2e-201 736 1914 dbj||AB008370_1 (AB008370) acid-stable alpha-amylase [Aspergillus kawachii]

y4b12a1.r1	750	1.3e-73	8 649	sp|P36914|AMYG_ASPOR	GLUCOAMYLASE PRECURSOR (GLUCAN 1,4-ALPHA-GLUCOSIDASE)(1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE) >
Contig1310_h4b03a1.f1	544	8.2e-52	146 979	pir||JN0588	alpha-amylase (EC 3.2.1.1) precursor - Aspergillus oryzae
y4b12a1.f1	523	1.4e-49	101 661	sp|P36914|AMYG_ASPOR	GLUCOAMYLASE PRECURSOR (GLUCAN 1,4-ALPHA-GLUCOSIDASE)(1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE) >
h4b03a1.r1	331	4.4e-29	83 448	sp|Q02905|AMYA_ASPAW	ALPHA-AMYLASE A PRECURSOR (1,4-ALPHA-D-GLUCANGLUCANOHYDROLASE A) >pir||A48305 alpha-amylas
Contig213_i0g04a1.f1	268	2.4e-21	316 834	sp|P22861|AMYG_D	GLUCOAMYLASE 1 PRECURSOR (GLUCAN 1,4-ALPHA-GLUCOSIDASE)(1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE) >pir||JN01
z8c10a1.f1	154	6.8e-08	183 527	gi|3420947	(AF082188) glucoamylase [Candida albicans]

c. Trehalose degradation

<trehalose-6-phosphate synthase>

Contig1709_c5f08a1.f1	1999	7.2e-260	485 1696	gi|3170246	(AF043230) trehalose-6-phosphate synthase subunit 1 [Emericellanidulans]
Contig1789_c7d11a1.f1	761	8.5e-75	704 1780	gnl|PID|e339278	(Z98850) hypothetical alpha-trehalose-phosphate synthase[Schizosaccharomyces pombe] >gnl|PID|e1314275

<trehalase>

Contig1448_f5f04a1.f1	998	6.3e-100	291 845	gi|2827392	(AF043229) neutral trehalase [Emericella nidulans]
r3b05a1.r1	298	1.4e-33	392 580	gi|2827392	(AF043229) neutral trehalase [Emericella nidulans]

12.fatty acid degradation (35)

a. lipase-triacylglycerols to glycerol+FA

r5h04a1.r1	203	2.4e-25	68 304	gnl|PID|d1033240	(AB010810) phospholipase D [Candida albicans]
c7c07a1.f1	224	1.2e-16	322 768	gi|2773042	(AF038440) phospholipase D2 [Homo sapiens]
*Contig1148_e9g05a1.r1	160	1.1e-10	348 608	gi|2853612	(AF034088) lipase [Pseudomonas sp. B11-1]

b. beta-oxydation of fatty acids

<beta-oxidation>

Contig1498_c8d05a1.f1	755	3.1e-74	128 841	pir||S54786	multifunctional beta-oxidation protein - Neurospora crassa>gi|510867 (X80052) multifunctional beta-ox
r5d03a1.r1	396	4.3e-35	23 406	pir||S54786	multifunctional beta-oxidation protein - Neurospora crassa>gi|510867 (X80052) multifunctio

i. fatty acid activation-thiokinase

<long-chain-fatty-acid-CoA ligase>

o9c08a1.r1	443	4.2e-41	12 527	gnl|PID|e316918	(Z95556) fadD35 [Mycobacterium tuberculosis]

<medium-chain acyl-CoA ligase>

*Contig351_g3f05a1.f1	216	3.6e-16	105 416	gi|2650449	(AE001093) medium-chain acyl-CoA ligase (alkK-1) [Archaeoglobusfulgidus]

*Contig215_m7d10a1.f1	129	8.8e-07	3	143	gi|2648519 (AE000963) medium-chain acyl-CoA ligase (alkK-5) [Archaeoglobusfulgidus]
<LONG-CHAIN ACYL-COASYNTHETASE>					
Contig1122_m1b06a1.r1	590	1.1e-56	13	741	gnl|PID|e1285372 (AL022304) fatty acid coa ligase [Schizosaccharomyces pombe]
Contig1087_o6c03a1.r1	518	4.7e-49	12	1079	sp|P41215|LCFA_H LONG-CHAIN-FATTY-ACID--COA LIGASE 1 (LONG-CHAIN ACYL-COASYNTHETASE 1) (LACS 1) (PALMITOYL-COA LIGASE)
g4g11a1.r1	496	9.7e-47	38	517	sp|P30624|LCF1_YEAST LONG-CHAIN-FATTY-ACID--COA LIGASE 1 (LONG-CHAIN ACYL-COASYNTHETASE 1) (FATTY ACID ACTIVATO
Contig1303_n2b07a1.r1	340	2.9e-29	203	805	sp|P39518|LCF2_Y LONG-CHAIN-FATTY-ACID--COA LIGASE 2 (LONG-CHAIN ACYL-COASYNTHETASE 2) (FATTY ACID ACTIVATOR 2) >pir||
Contig1701_c6g08a1.f1	179	5.1e-12	505	711	sp|P39002|LCF3_Y LONG-CHAIN-FATTY-ACID--COA LIGASE 3 (LONG-CHAIN ACYL-COASYNTHETASE 3) (FATTY ACID ACTIVATOR 3) >pir||
<fatty acid coa ligase>					
Contig1120_m1c06a1.r1	353	1.4e-31	353	937	gi|2648302 (AE000952) 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase (hpcE-2)[Archaeoglobus fulgidus]
<ii.carnitine acetyl transferase>					
Contig1254_m6h06a1.r1	1647	1.1e-168	103	1185	gi|2511761 (AF023156) carnitine acetyl transferase FacC [Emericella nidulans]
Contig1138_m8c01a1.f1	427	1.9e-39	332	820	gi|2688966 (AF027979) carnitine acetyl transferase [Magnaporthe grisea]
m8c01a1.r1	268	1e-21	417	671	gi|2688966 (AF027979) carnitine acetyl transferase [Magnaporthe grisea]
<carnitine racemase-d to l form>					
h0f12a1.f1	194	9.6e-15	224	649	gnl|PID|e326890 (Z97336) carnitine racemase homolog [Arabidopsis thaliana]
<iii.acyl-CoA dehydrogenase>					
m8b09a1.r1	394	6.3e-36	168	719	gi|2736364 (AF039038) Similar to acyl-coA dehydrogenase; coded for by C.elegans cDNA yk335a7.3; coded
r2g12a1.f1	292	4e-25	192	617	gnl|PID|e306530 (Z92770) fadE2 [Mycobacterium tuberculosis]
Contig976_m8b09a1.f1	292	4.2e-25	314	724	gi|2736364 (AF039038) Similar to acyl-coA dehydrogenase; coded for by C.elegans cDNA yk335a7.3; coded for by C.
<iv.enoyl-CoA hydratase>					
Contig1024_n0g10a1.r1	485	1.5e-45	76	783	gi|755067 (L39265) enoyl CoA hydratase [Rhizobium meliloti]
<v.3-hydroxyacyl-CoA dehydrogenase>					
Contig747_q0a04a1.r1	205	8.4e-15	115	498	gnl|PID|e1294492 (AL023702) fatty acid oxidation complex alpha-subunit[Streptomyces coelicolor]
<vi.HYDROXYBUTYRYL-COA DEHYDROGENASE>					

```
Contig70_m0d04a1.f1      405  4.4e-37    91 471   sp|P45856|MMGB_B      PROBABLE 3-HYDROXYBUTYRYL-COA
                                                  DEHYDROGENASE(BETA-HYDROXYBUTYRYL-COA DEHYDROGENASE) (BHBD) >gi|881605

<vii.3-ketoacyl-CoA  thiolase>
c6b01a1.r1               517  5.8e-49    86 655   sp|Q05493|THIK_YARLI  3-KETOACYL-COA THIOLASE PEROXISOMAL
                                                  PRECURSOR(BETA-KETOTHIOLASE) (ACETYL-COA ACYLTRANSFERA
r2e10a1.f1               461  5.2e-43   128 610    pir||JT0551           acetyl-CoA C-acyltransferase (EC 2.3.1.16),
                                                  peroxisomal - rat>gi|205049 (M32801) 3-ketoacyl-coA thiolase
c6b01a1.f1               394    6e-36   130 483   sp|P07871|THI1_RAT    3-KETOACYL-COA THIOLASE PEROXISOMAL B
                                                  PRECURSOR(BETA-KETOTHIOLASE B) (ACETYL-COA ACYLTRANS

c. odd chain fatty acids
<methylmalonyl carboxylase-also ile, thr, met, val degradation>
c9h06a1.f1               153  5.3e-10   151 351   gnl|PID|d1031330      (AP000005) 149aa long hypothetical
                                                  methylmalonyl-CoAdecarboxylase gamma chain [Pyrococcus

d. Unsaturated fatty acid degradation
<fatty acyl-CoA reductase>
c6g04a1.r1               184  5.4e-12   124 438   gi|1684886            (U77680) fatty acyl-CoA reductase [Acinetobacter
                                                   calcoaceticus]

e. Branch chain fatty acid degradation
<branched-chain enoyl CoA reductase>
Contig556_c6b03a1.f1     169  8.9e-10   145 642   gi|2407655            (AF019136) 2-methyl branched-chain enoyl CoA reductase
                                                  isoform I[Ascaris suum]

f. Ketone body metabolism
<3-OXOACID COA-TRANSFERASE>
Contig219_i0c10a1.r1     716  4.6e-70     5 748   sp|P55809|SCOT_H      SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASE
                                                  PRECURSOR(SUCCINYL COA:3-OXOACID COA-TRANSFERASE) (OXC
Contig750_q0b02a1.f1     218  1.9e-16   180 401   sp|Q09450|SCOT_C      PROBABLE SUCCINYL-COA:3-KETOACID-COENZYME A
                                                  TRANSFERASEPRECURSOR (3-OXOACID COA-TRANSFERASE) >gi|6656
<hydroxybutyrate dehydrogenase>
<ACETOACETYL-COA THIOLASE-acetyl-coA to acetoacyl-coA>
w8a06a1.r1               441  6.2e-41    21 503   sp|P17764|THIL_RAT    ACETYL-COA ACETYLTRANSFERASE PRECURSOR,
                                                  MITOCHONDRIAL(ACETOACETYL-COA THIOLASE) >pir||XXRT
w8a06a1.f1               310  4.9e-27    97 411   sp|P17764|THIL_RAT    ACETYL-COA ACETYLTRANSFERASE PRECURSOR,
                                                  MITOCHONDRIAL(ACETOACETYL-COA THIOLASE) >pir||XXRT
x3h11a1.r1               260  2.2e-20   198 569   sp|P17764|THIL_RAT    ACETYL-COA ACETYLTRANSFERASE PRECURSOR,
                                                  MITOCHONDRIAL(ACETOACETYL-COA THIOLASE) >pir||XXRT
<SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASE-acetoacetate to acetoacyl-coA>

13. Metabolism of other energy sources (27)
```

a.alcohol dehydrogenases

<FORMALDEHYDE DEHYDROGENASE-long chain primary alcohol to aldehyde or ketone>

Contig1273_h4c10a1.f1	661	3.2e-64	7	471	sp|Q06099|FADH_C GLUTATHIONE-DEPENDENT FORMALDEHYDE DEHYDROGENASE (FDH)(FALDH) >pir||JN0447 FDH1 protein - yeast (Cand
j9h02a1.r1	539	2.5e-51	137	484	sp|Q06099|FADH_CANMA GLUTATHIONE-DEPENDENT FORMALDEHYDE DEHYDROGENASE (FDH)(FALDH) >pir||JN0447 FDH1 protein -
Contig1690_e9a05a1.f1	484	1.9e-45	178	1086	sp|P47734|FADH_M GLUTATHIONE-DEPENDENT FORMALDEHYDE DEHYDROGENASE (FDH)(FALDH) >gi|496118 (L33464) alcohol dehydrogena
Contig1186_e9a05a1.r1	198	1.7e-14	267	497	sp|P47734|FADH_M GLUTATHIONE-DEPENDENT FORMALDEHYDE DEHYDROGENASE (FDH)(FALDH) >gi|496118 (L33464) alcohol dehydrogena

<aldehyde reductase>

Contig879_r4e05a1.f1	507	6.9e-48	138	902	sp|P47137|YJ66_Y PROBABLE OXIDOREDUCTASE YJR096W >pir||S57117 aldehydereductase homolog YJR096w - yeast (Saccharomyces
*Contig770_s9g09a1.f1	471	4.4e-44	118	831	sp|P23901|ALDR_H ALDOSE REDUCTASE (AR) (ALDEHYDE REDUCTASE) >pir||S15024aldose reductase-related protein - barley >gi|
e4b11a1.r1	341	2.3e-30	25	564	gi|1142698 (U26463) NADPH-dependent aldehyde reductase [Sporidiobolussalmonicolor]
Contig599_c4a06a1.r1	161	1.3e-20	195	422	sp|P47137|YJ66_Y PROBABLE OXIDOREDUCTASE YJR096W >pir||S57117 aldehydereductase homolog YJR096w - yeast (Saccharomyces
Contig1561_h1h04a1.f1	175	2.5e-12	295	615	gi|1142698 (U26463) NADPH-dependent aldehyde reductase [Sporidiobolussalmonicolor]

b.GLYCEROL

<glycerol>

*Contig48_w8h05a1.f1	119	9.3e-06	201	410	gnl|PID|d1029223 (AB004569) glycerol kinase [Thermus aquaticus flavus]

<glycerol-3-phosphatase>

Contig1355_a1b02c9.r1	294	2.4e-25	185	610	sp|P41277|GPP1_Y (DL)-GLYCEROL-3-PHOSPHATASE 1 >gnl|PID|d1009695 (D50471)unknown [Saccharomyces cerevisiae]
Contig1_e0h11a1.r1	208	3.4e-16	120	578	sp|P41277|GPP1_Y (DL)-GLYCEROL-3-PHOSPHATASE 1 >gnl|PID|d1009695 (D50471)unknown [Saccharomyces cerevisiae]

<glycerol-3-phosphate dehydrogenase>

Contig1076_r5a09a1.r1	384	6e-45	288	806	sp|P41911|GPD2_Y GLYCEROL-3-PHOSPHATE DEHYDROGENASE (NAD+) 2 >pir||S61719glycerol-3-phosphate dehydrogenase (NAD+) (EC
v3f01a1.r1	145	8.8e-09	9	116	sp|P21696|GPDA_SCHPO GLYCEROL-3-PHOSPHATE DEHYDROGENASE (NAD+), CYTOPLASMIC(GPD-C) (GPDH-C) >gi|4952 (X56162) g

c.proprionate

<PRPD PROTEIN>

i0h10a1.f1	309	8.8e-27	108	506	sp|P77243|PRPD_ECOLI PRPD PROTEIN >gi|1657530 (U73857) similar to yqiP of B.subtilis [Escherichia coli] >gi|178

d. other

<diaminobutyrate decarboxylase-dia-butyrate to dia-propane>

Contig933_o6g06a1.f1	135	1.5e-05	602	835	gi|1573971 (U32776) L-2,4-diaminobutyrate decarboxylase [Haemophilusinfluenzae Rd]

<ACETAMIDASE-allows acetamide and formamide as sole C or N source>

Contig1638_c1h08a1.f1	170	3.2e-09	15	443	pir||JS0633 amidase (EC 3.5.1.4) - Aspergillus oryzae >gnl|PID|d1001845(D10492) acetamidase [Aspergillus oryzae]
p0e07a1.r1	146	1.3e-08	171	449	sp|P08158|AMDS_EMENI ACETAMIDASE >pir||A26511 amdS protein - Emericellanidulans >gi|168015 (M16371) acetamidase

<formamidase>

Contig978_c8h05a1.f1	606	9.6e-104	430	1011	gnl|PID|e256826 (X99632) formamidase [Methylophilus methylotrophus]

<acetate>

z4e02a1.r1	909	1.7e-90	13	528	gi|2262191 (U56097) acetate regulatory DNA binding protein FacB [Emericellanidulans]
w9h04a1.r1	875	7e-87	9	509	gi|1130507 (L41670) fumarylacetoacetate hydrolase [Emericella nidulans]
w9h04a1.f1	631	5.2e-61	65	457	gi|1130507 (L41670) fumarylacetoacetate hydrolase [Emericella nidulans]
l5g04a1.f1	355	3.5e-43	209	439	gnl|PID|e1249838 (AJ001836) maleylacetoacetate isomerase [Emericella nidulans]>gnl|PID|e1249842 (AJ001837)
n0c08a1.r1	301	4.1e-26	24	455	gnl|PID|e1285330 (AJ004870) acetate kinase [Thermoanaerobacteriumthermosaccharolyticum]

<ALDEHYDE DEHYDROGENASE-broad substrate specificity>

Contig1813_a1h01f2.f1	2420	1.1e-250	152	1639	sp|P08157|DHAL_E ALDEHYDE DEHYDROGENASE (ALDDH) >pir||A29055 aldehydedehydrogenase (NAD+) (EC 1.2.1.3) - Emericella ni
Contig147_j4g02a1.r1	663	2e-64	41	859	sp|P38067|YBL6_Y HYPOTHETICAL ALDEHYDE-DEHYDROGENASE LIKE PROTEIN INCOQ1-HHF1 INTERGENIC REGION >pir||S45858 probable
Contig521_c8e01a1.r1	569	1.7e-54	176	1141	sp|P33008|DHAL_P PROBABLE ALDEHYDE DEHYDROGENASE >pir||S27652 aldehydedehydrogenase - Pseudomonas sp >gi|151586 (M9144
c4h09a1.f1	423	5.5e-39	46	540	sp|P54885|PROA_YEAST GAMMA-GLUTAMYL PHOSPHATE REDUCTASE (GPR)(GLUTAMATE-5-SEMIALDEHYDE DEHYDROGENASE)(GLUTAMYL-
c4h09a1.r1	349	3.9e-31	136	642	sp|P54885|PROA_YEAST GAMMA-GLUTAMYL PHOSPHATE REDUCTASE (GPR)(GLUTAMATE-5-SEMIALDEHYDE DEHYDROGENASE)(GLUTAMYL-
n1c01a1.r1	276	1.8e-23	91	426	gi|1399099 (U44901) aspartate semialdehyde dehydrogenase [Ustilago maydis]
z1g07a1.r1	150	4e-09	311	490	sp|P25526|GABD_ECOLI SUCCINATE-SEMIALDEHYDE DEHYDROGENASE (NADP+) (SSDH)>gi|147901 (M88334) succinic semialdehy

14. Electron transport (78)

a.Complex I-NADH-ubiquinone

<NADH dehydrogenase>

f0b07a1.r1	619	9.1e-60	49	615	gnl|PID|e1227831 (AJ001520) 19.3kD iron-sulfur subunit of mitochondrialcomplex I [Neurospora crassa]

g8f09a1.r1	529	3.2e-50	65 601	sp|Q02854|NUXM_NEUCR NADH-UBIQUINONE OXIDOREDUCTASE 21 KD SUBUNIT (COMPLEXI-21KD) (CI-21KD) >pir||*S27171 NADH* d
o8h07a1.f1	488	6.6e-46	203 673	sp|P21976|NUPM_NEUCR NADH-UBIQUINONE OXIDOREDUCTASE 20.8 KD SUBUNIT >gi|168781(M55323) complex I 22-kDa polypep
j4e05a1.f1	489	1.3e-45	212 652	sp|P24918|NUAM_NEUCR NADH-UBIQUINONE OXIDOREDUCTASE 78 KD SUBUNIT PRECURSOR(COMPLEX I-78KD) (CI-78KD) >pir||S17=NADH dehydrogenase
o9g03a1.f1	483	2.5e-45	86 568	sp|Q08822|ETFD_YEAST PROBABLE ELECTRON TRANSFER FLAVOPROTEIN-UBIQUINONEOXIDOREDUCTASE PRECURSOR (ETF-QO) (ETF-U
m7e03a1.r1	360	4.6e-32	14 466	gnl|PID|e349376 (Z99260) ubiqinone reductase [Schizosaccharomyces pombe]
l5b07a1.f1	354	1.1e-31	362 598	sp|P22142|NUCM_NEUCR NADH-UBIQUINONE OXIDOREDUCTASE 49 KD SUBUNIT PRECURSOR(COMPLEX I-49KD) (CI-49KD) >pir||S13
Contig506_c9f11a1.f1	343	1.6e-30	93 503	sp|P25710|NUJM_N NADH-UBIQUINONE OXIDOREDUCTASE 21.3 KD SUBUNIT>pir||S14277 NADH dehydrogenase (ubiquinone) (EC 1.6.5.
j4e05a1.r1	341	2.3e-29	9 341	pir||S59926 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) 78K chain precursor -Neurospora crassa >gi|55
z7e09a1.r1	294	2.5e-25	69 404	sp|P24919|NUFM_NEUCR NADH-UBIQUINONE OXIDOREDUCTASE 29.9 KD SUBUNIT PRECURSOR(COMPLEX I-29.9KD) (CI-29.9KD) >pi
Contig1077_o6a09a1.r1	264	3.8e-22	117 347	pir||S47150 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) 14K chain -Neurospora crassa >gi|*499315 (Z18945) NADH*:ub
Contig317_g5h08a1.f1	264	1.4e-21	203 607	sp|P53318|COQ6_Y UBIQUINONE BIOSYNTHESIS MONOOXGENASE COQ6 >pir||S64587hypothetical protein YGR255c - yeast (Saccharom
w8g05a1.r1	239	1.5e-19	219 419	sp|P24919|NUFM_NEUCR NADH-UBIQUINONE OXIDOREDUCTASE 29.9 KD SUBUNIT PRECURSOR(COMPLEX I-29.9KD) (CI-29.9KD) >pi
Contig1249_m6a09a1.r1	238	1.6e-18	48 278	gnl|PID|e349376 (Z99260) ubiqinone reductase [Schizosaccharomyces pombe]
g5h08a1.r1	200	7.4e-13	30 518	sp|P53318|COQ6_YEAST UBIQUINONE BIOSYNTHESIS MONOOXGENASE COQ6 >pir||S64587hypothetical protein YGR255c - yeast
Contig813_w8f02a1.f1	153	2.3e-10	145 378	sp|Q02369|NI2M_B NADH-UBIQUINONE OXIDOREDUCTASE B22 SUBUNIT (COMPLEXI-B22) (CI-B22) >pir||S28256 NADH dehydrogenase (u

b.Complex II-Succinate-ubiquinone
<succinate dehydrogenase>

c.Complex III-Ubiquinone to cytochrome C
<cytochrome b>

Contig715_p0e03a1.r1	755	3.1e-74	10 489	sp|P00161|CYB_EM CYTOCHROME B >pir||CBASN ubiquinol--cytochrome-c *reductase(EC 1.10.2.2) cytochrome b - Emericella nid*
p0e03a1.f1	727	3e-71	24 464	sp|P00161|CYB_EMENI CYTOCHROME B >pir||CBASN ubiquinol--cytochrome-c reductase(EC 1.10.2.2) cytochrome b - Eme
h0a08a1.f1	244	4.9e-20	434 649	sp|P40312|CYB5_YEAST CYTOCHROME B5 >pir||S63052 cytochrome b5 - yeast(Saccharomyces cerevisiae) >gnl|PID|e22183
j4c03a1.f1	234	5.7e-19	283 531	sp|P38626|NC5R_YEAST PUTATIVE NADH-CYTOCHROME B5 REDUCTASE (P35)

					>pir||S49935cytochrome-b5 reductase (EC 1.6.2.
Contig1096_m5d06a1.f1	186	6.9e-14	171	428	prf||751845A cytochrome b2 1-103 [Saccharomycetales]
Contig20_m2d10a1.f1	183	1.4e-13	177	407	prf||751845A cytochrome b2 1-103 [Saccharomycetales]
x3a05a1.f1	163	4.6e-11	186	314	sp|P36060|MCR1_YEAST NADH-CYTOCHROME B5 REDUCTASE PRECURSOR (P34 / P32)>pir||S37800 cytochrome-b5 reductase (EC
g9e09a1.f1	175	7.6e-11	316	606	sp|P36060|MCR1_YEAST NADH-CYTOCHROME B5 REDUCTASE PRECURSOR (P34 / P32)>pir||S37800 cytochrome-b5 reductase (EC
z8c03a1.r1	156	9.4e-08	190	477	sp|P00175|CYB2_YEAST CYTOCHROME B2 PRECURSOR (L-LACTATE DEHYDROGENASE(CYTOCHROME)) (L-LACTATE FERRICYTOCHROME C
Contig254_h1a12a1.f1	107	1.7e-05	1	87	gi|2062405 (U79011) cytochrome b5 [Borago officinalis]
<cytochrome c>					
Contig531_o4f04a1.r1	602	4.5e-58	71	409	sp|P38091|CYC_EM CYTOCHROME C >gi|1899007 (M83141) cytochrome c [Emericellanidulans]
Contig723_u4g07a1.r1	451	5.4e-42	13	444	sp|P32891|DLD1_Y D-LACTATE DEHYDROGENASE (CYTOCHROME) PRECURSOR (D-LACTATEFERRICYTOCHROME C OXIDOREDUCTASE) (D-LCR) >p
Contig477_d3b01a1.f1	317	8.9e-28	397	711	sp|P04037|COX4_Y CYTOCHROME C OXIDASE POLYPEPTIDE IV PRECURSOR >pir||OLBY4cytochrome-c oxidase (EC 1.9.3.1) chain IV p
Contig732_x7b08a1.f1	288	9.9e-25	222	419	sp|Q01519|COXG_Y CYTOCHROME C OXIDASE POLYPEPTIDE VIB (AED) >pir||S31256cytochrome-c oxidase (EC 1.9.3.1) chain VIb -
Contig175_m6h01a1.f1	213	8.9e-17	259	588	sp|P32799|COXE_Y CYTOCHROME C OXIDASE POLYPEPTIDE VIA PRECURSOR>pir||A48520 cytochrome-c oxidase (EC 1.9.3.1) chain VI
Contig640_c0e03a1.f1	212	1.2e-16	224	541	pdb|2CYP| Cytochrome c Peroxidase (E.C.1.11.1.5) (Ferrocytochrome c (Colon)H2O2 Reductase)
Contig963_k5a06a1.f1	176	7.5e-13	377	556	gnl|PID|e1256442 (AL021767) possible ubiquinol-cytochrome c reductasecomponent [Schizosaccharomyces pombe]
o5b10a1.r1	175	1.1e-12	159	308	sp|Q12287|COX5_YEAST CYTOCHROME C OXIDASE ASSEMBLY PROTEIN COX17 >pir||S62056COX17 protein - yeast (Saccharomyc
<cytochrome oxidase>					
<CBP4 PROTEIN-cytoC, ubiquinol assembly>					
Contig1400_w9h07a1.f1	121	5.2e-06	248	562	sp|P37267|CBP4_Y CBP4 PROTEIN PRECURSOR >pir||S64488 regulatory proteinCBP4 precursor - yeast (Saccharomyces cerevisia
d.Other electron transport pathways					
<NADH OXIDASE>					
*Contig1365_o9c07a1.f1	182	2.4e-12	208	660	sp|P32382|NADO_T NADH OXIDASE >pir||S25102 NADH oxidase -Thermoanaerobacter brockii >gi|48123 (X67220) NADH oxidase[Th
<NADPH DEHYDROGENASE>					
x1h09a1.r1	385	5.5e-35	136	729	gi|2232254 (AF005237) old-yellow-enzyme homolog [Catharanthus roseus]
e.Electron carriers					
<flavoprotein>					
Contig1240_m2d10a1.r1	751	8.3e-74	28	891	gnl|PID|e349663 (Z99292) flavoprotein [Schizosaccharomyces pombe]

Contig64_m0e09a1.r1	402	8.4e-37	1 606	sp|Q10499|YDGE_S PUTATIVE FLAVOPROTEIN C26F1.14C
Contig50_h8g02a1.r1	325	2.7e-28	104 508	gnl|PID|e349663 (Z99292) flavoprotein [Schizosaccharomyces pombe]
Contig1250_g7a10a1.r1	172	1.9e-12	64 252	sp|P53575|ETFB_B ELECTRON TRANSFER FLAVOPROTEIN BETA-SUBUNIT (BETA-ETF)(ELECTRON TRANSFER FLAVOPROTEIN SMALL SUBUNIT)
<FLAVOHEMOPROTEIN>				
i8f11a1.r1	455	2e-42	85 567	sp|P39662|HMPA_ALCEU FLAVOHEMOPROTEIN (HAEMOGLOBIN-LIKE PROTEIN)(FLAVOHEMOGLOBIN) >pir||A53396 flavohemoprotein
w9e03a1.r1	206	1.8e-15	46 327	sp|P24232|HMPA_ECOLI FLAVOHEMOPROTEIN (HAEMOGLOBIN-LIKE PROTEIN)(FLAVOHEMOGLOBIN) (DIHYDROPTERIDINE REDUCTASE (
<QUINONE>				
w8a03a1.r1	361	2e-32	4 531	sp|P38230|QOR_YEAST PROBABLE QUINONE OXIDOREDUCTASE (NADPH:QUINONE REDUCTASE)>pir||S45904 quinone oxidoreducta

f.Component enzymes and molecules

<FLAVIN OXIDOREDUCTASE>				
Contig1495_d3e02a1.f1	835	1.1e-82	53 598	sp|Q00415|HPPD_C 4-HYDROXYPHENYLPYRUVATE DIOXYGENASE (4HPPD) (HPD) (T-CELLREACTIVE PROTEIN) >gi|601846 (L38493) T-cell
Contig1360_c5g11a1.r1	622	4.3e-60	58 912	sp|Q00415|HPPD_C 4-HYDROXYPHENYLPYRUVATE DIOXYGENASE (4HPPD) (HPD) (T-CELLREACTIVE PROTEIN) >gi|601846 (L38493) T-cell
Contig11_y8g03a1.r1	263	4.8e-22	176 490	sp|P54550|YQJM_B PROBABLE NADH-DEPENDENT FLAVIN OXIDOREDUCTASE YQJM>gnl|PID|d1013299 (D84432) YqjM [Bacillus subtilis]
Contig191_m8f03a1.r1	139	3.1e-08	312 476	sp|P54550|YQJM_B PROBABLE NADH-DEPENDENT FLAVIN OXIDOREDUCTASE YQJM>gnl|PID|d1013299 (D84432) YqjM [Bacillus subtilis]
<RESPIRATORY complex assembly>				
i2e08a1.r1	783	4e-77	10 606	sp|P40341|RCA1_YEAST MITOCHONDRIAL RESPIRATORY CHAIN COMPLEXES ASSEMBLYPROTEIN RCA1 (TAT-BINDING HOMOLOG 12) >p
<MITOCHONDRIAL FAD CARRIER PROTEIN>				
s9f04a1.f1	247	2.4e-20	208 567	gnl|PID|d1022297 (AB004539) MITOCHONDRIAL FAD CARRIER PROTEINFAD CARRIER PROTEIN FLX1[Schizosaccharomyces pombe] >gnl|PID|e125

g.ATP synthase

<ATP SYNTHASE>				
y8h10a1.r1	867	4.1e-86	16 540	sp|P23704|ATPB_NEUCR ATP SYNTHASE BETA CHAIN, MITOCHONDRIAL PRECURSOR>pir||JC1112 H+-transporting ATP synthase
k5d06a1.r1	743	6.2e-73	7 504	sp|P22550|VATB_CANTR VACUOLAR ATP SYNTHASE SUBUNIT B (V-ATPASE 57 KD SUBUNIT)>pir||S13080 H+-transporting ATPas
Contig1172_h1f03a1.r1	719	2.4e-70	13 570	sp|P24487|ATPA_S ATP SYNTHASE ALPHA CHAIN, MITOCHONDRIAL PRECURSOR>pir||A39036 H+-transporting ATP synthase (EC 3.6.1.
Contig256_h0h10a1.f1	501	3.3e-62	249 701	sp|P37211|ATPA_N ATP SYNTHASE ALPHA CHAIN, MITOCHONDRIAL PRECURSOR>pir||JC1111 H+-transporting ATP synthase (EC 3.6.1.
Contig763_s8f09a1.f1	613	3.9e-59	321 725	sp|P23704|ATPB_N ATP SYNTHASE BETA CHAIN, MITOCHONDRIAL PRECURSOR>pir||JC1112 H+-transporting ATP synthase (EC 3.6.1.3
Contig1377_e9e12a1.r1	564	5e-54	55 726	sp|O13349|ATPF_K ATP SYNTHASE SUBUNIT 4, MITOCHONDRIAL

					PRECURSOR>gi|2425071 (AF019222) F1Fo-ATP synthase subunit 4 (Kl
j0b09a1.r1	543	1.1e-51	12	425	sp|P11592|VATA_NEUCR VACUOLAR ATP SYNTHASE CATALYTIC SUBUNIT A (V-ATPASE 67 KDSUBUNIT) >pir||PXNCV7 H+-transpor
Contig1510_h1f11a1.f1	505	1e-47	88	516	sp|P16000|ATPL_E ATP SYNTHASE PROTEIN 9, MITOCHONDRIAL PRECURSOR(LIPID-BINDING PROTEIN) >gi|168019 (M30144) mitochondr
Contig1289_i8d01a1.f1	490	3.4e-46	9	515	sp|O13350|ATP7_K ATP SYNTHASE D CHAIN, MITOCHONDRIAL >gi|2425073(AF019223) F1Fo-ATP synthase subunit 7 (Kluyveromyces
w4g11a1.r1	475	1.5e-44	117	488	sp|P56525|ATPD_NEUCR ATP SYNTHASE DELTA CHAIN, MITOCHONDRIAL PRECURSOR
g8b10a1.r1	455	2.2e-42	107	514	sp|Q01278|VATE_NEUCR VACUOLAR ATP SYNTHASE SUBUNIT E (V-ATPASE E SUBUNIT)(V-ATPASE 26 KD SUBUNIT) >gi|600167 (U
j0b09a1.f1	410	1.9e-37	138	461	sp|P11592|VATA_NEUCR VACUOLAR ATP SYNTHASE CATALYTIC SUBUNIT A (V-ATPASE 67 KDSUBUNIT) >pir||PXNCV7 H+-transpor
o5f02a1.r1	383	9.2e-35	176	517	sp|P39111|VATF_YEAST VACUOLAR ATP SYNTHASE 14 KD SUBUNIT (V-ATPASE F SUBUNIT)>pir||A55118 H+-transporting ATPas
Contig647_b0h11a1.f1	296	1.5e-25	166	498	sp|P38077|ATPG_Y ATP SYNTHASE GAMMA CHAIN, MITOCHONDRIAL PRECURSOR>pir||S55891 H+-transporting ATP synthase (EC 3.6.1.
b0h11a1.r1	268	1.3e-22	13	384	sp|P49377|ATPG_KLULA ATP SYNTHASE GAMMA CHAIN, MITOCHONDRIAL PRECURSOR>pir||S56153 H+-transporting ATP synthase
Contig433_n8c05a1.f1	211	8.7e-16	178	600	sp|P41807|VM13_Y VACUOLAR ATP SYNTHASE 54 KD SUBUNIT (V-ATPASE 54 KDSUBUNIT) >pir||A47429 H+-transporting ATPase (EC 3
m8f04a1.r1	135	1.2e-07	17	145	sp|P09457|ATPO_YEAST ATP SYNTHASE OLIGOMYCIN SENSITIVITY CONFERRAL PROTEINPRECURSOR, MITOCHONDRIAL (OSCP) (ATP
Contig1686_a5b12a1.f1	113	7.3e-06	142	291	pir||S53404 ATP synthase regulatory factor homolog YLR327c - yeast(Saccharomyces cerevisiae) >gi|662125 (U20618)
<PLASMA MEMBRANE ATPASE (PROTON PUMP)>					
Contig1349_c5b10a1.f1	1142	2.9e-115	27	737	gi|3366659 (AF036763) P-ATPase [Emericella nidulans]
m7e01a1.r1	755	3.6e-74	30	590	gi|3366659 (AF036763) P-ATPase [Emericella nidulans]=(PROTON PUMP)
Contig1421_c8b08a1.f1	542	1.2e-51	312	815	gi|2197050 (AF001033) putative 20kDa subunit of the V-ATPase [Neurosporacrassa]
q0g10a1.f1	546	1.7e-51	4	504	gi|1237128 (U36396) vacuolar ATPase 98 kDa subunit [Neurospora crassa]-H+-transporting
n3f03a1.r1	459	6.4e-42	30	401	sp|P49380|PMA1_KLULA PLASMA MEMBRANE ATPASE (PROTON PUMP) >gi|598435 (L37875)proton-ATPase [Kluyveromyces lacti
Contig757_v1e09a1.r1	448	1.3e-41	7	480	sp|P31413|VATL_N VACUOLAR ATP SYNTHASE 16 KD PROTEOLIPID SUBUNIT>pir||S43893 H+-transporting ATPase (EC 3.6.1.35) lipi
n3f03a1.f1	438	1.4e-39	204	773	sp|Q07421|PMA1_AJECA PLASMA MEMBRANE ATPASE (PROTON PUMP) >gi|409249 (L07305)ATPase [Ajellomyces capsulatus] >p
Contig962_j7d08a1.f1	392	1.4e-34	24	470	gi|3366659 (AF036763) P-ATPase [Emericella nidulans]
d2c08a1.f1	324	1.7e-28	352	759	gi|1208770 (U48365) V-type ATPase 16 kDa proteolipid subunit [Pleurochrysiscarterae] >gi|2149129 (U81
q0g10a1.r1	329	5.7e-28	10	495	gi|1237128 (U36396) vacuolar ATPase 98 kDa subunit [Neurospora crassa]

h.Alternative respiratory path
<ALTERNATIVE OXIDASE>
Contig340_g4b11a1.f1 778 1.4e-76 1 735 gnl|PID|d1032995 (AB016540) alternative oxidase [Aspergillus niger]

15.Reducing carriers (8)

<a.glutaredoxin>
Contig1357_a1b04c9.r1 276 2e-23 130 426 gi|3249567 (AF047694) glutaredoxin [Vernicia fordii]
*Contig1556_c9f10a1.r1 239 1.6e-19 137 448 gi|3335374 (AC003028) glutaredoxin-like protein [Arabidopsis thaliana]

<b.gluathione>
<gamma-glutamyl transpeptidase-synthesis and deg of glutathione>
r4h11a1.r1 471 4.3e-44 46 822 sp|P19440|GGT1_HUMAN GAMMA-GLUTAMYLTRANSPEPTIDASE 1 PRECURSOR(GAMMA-GLUTAMYLTRANSFERASE 1) >pir||EKHUEXgamma-gl
<glutathione S-transferase-reduces peroxides,reducing agent>
Contig1732_c8d03a1.f1 133 5.7e-08 256 654 gi|2583081 (AF026977) microsomal glutathione S-transferase 3 [Homo sapiens]
*Contig1246_v3f09a1.r1 136 1.6e-06 208 426 sp|P04907|GTH3_M GLUTATHIONE S-TRANSFERASE III (GST-III) (CLASS PHI)>pir||XUZM32 glutathione transferase (EC 2.5.1.18)
<GLUTATHIONE PEROXIDASE>
Contig1403_c7g02a1.f1 518 4.7e-49 231 719 sp|P38143|GSHI_Y GLUTATHIONE PEROXIDASE HOMOLOG YBR244W >pir||S46121probable glutathione peroxidase (EC 1.11.1.9) - ye

<c.thioredoxin>
Contig1775_d1a07a1.r1 560 1.6e-53 269 595 sp|P29429|THIO_E THIOREDOXIN >pir||S27053 thioredoxin - Emericellanidulans >bbs|120057 thioredoxin [Aspergillus nidula
j9b04a1.r1 515 8.8e-49 50 484 sp|P18408|MT16_YEAST PHOSPHOADENOSINE PHOSPHOSULFATE REDUCTASE (PAPSREDUCTASE, THIOREDOXIN DEPENDENT) (PADOPS R
j9b04a1.f1 255 3.5e-21 350 583 sp|P18408|MT16_YEAST PHOSPHOADENOSINE PHOSPHOSULFATE REDUCTASE (PAPSREDUCTASE, THIOREDOXIN DEPENDENT) (PADOPS R

II. Cell Growth, Cell Division

A. Cell walls (12)
<septin>
Contig1351_c5d08a1.r1 1174 1.4e-118 226 1284 sp|P48009|SPN4_S SEPTIN HOMOLOG SPN4 >gi|987283 (U29890) septin homolog[Schizosaccharomyces pombe] >gnl|PID|e1168676 (
Contig447_d5d02a1.r1 1173 1.8e-118 15 773 gi|1791305 (U83489) septin B [Emericella nidulans]
d5d02a1.f1 352 1.7e-31 247 480 gi|1791305 (U83489) septin B [Emericella nidulans]
<RODLET PROTEIN-spore-wall fungal hydrophobin>
Contig1836_a1b02f2.f1 545 5.8e-52 90 560 sp|P28346|RODL_E RODLET PROTEIN PRECURSOR >pir||A40323 Rodletless

					protein- Emericella nidulans >gi|168086 (M61113) rod
Contig1791_c6e09a1.f1	355	8.5e-32	975	1325	sp|P28346|RODL_E RODLET PROTEIN PRECURSOR >pir||A40323 Rodletless
					protein- Emericella nidulans >gi|168086 (M61113) rod
Contig971_k0g04a1.r1	165	1.2e-11	6	101	sp|P28346|RODL_E RODLET PROTEIN PRECURSOR >pir||A40323 Rodletless
					protein- Emericella nidulans >gi|168086 (M61113) rod
Contig1330_g6b08a1.f1	162	2.5e-11	213	308	sp|P28346|RODL_E RODLET PROTEIN PRECURSOR >pir||A40323 Rodletless
					protein- Emericella nidulans >gi|168086 (M61113) rod
<SPORE-WALL FUNGAL HYDROPHOBIN-not rodlet>					
Contig1806_b0d04a1.r1	381	1.3e-34	75	479	sp|P52750|DEWA_E SPORE-WALL FUNGAL HYDROPHOBIN DEWA PRECURSOR
					>pir||S67924spore-wall fungal hydrophobin DewA - Emerice
Contig1711_b0c12a1.f1	381	1.4e-34	175	579	sp|P52750|DEWA_E SPORE-WALL FUNGAL HYDROPHOBIN DEWA PRECURSOR
					>pir||S67924spore-wall fungal hydrophobin DewA - Emerice
Contig1863_a1a04f2.f1	381	1.4e-34	186	590	sp|P52750|DEWA_E SPORE-WALL FUNGAL HYDROPHOBIN DEWA PRECURSOR
					>pir||S67924spore-wall fungal hydrophobin DewA - Emerice
<INTEGRIN ALPHA CHAIN-LIKE PROTEIN-cell adhesion>					
g4g04a1.r1	257	8.6e-20	96	560	sp|P53705|INT1_CANAL INTEGRIN ALPHA CHAIN-LIKE PROTEIN
					(ALPHA-INT1)>gi|1144531 (U35070) integrin-like protein a
g7a09a1.r1	159	9.9e-08	235	447	sp|P53705|INT1_CANAL INTEGRIN ALPHA CHAIN-LIKE PROTEIN
					(ALPHA-INT1)>gi|1144531 (U35070) integrin-like protein a=C. albicans
B. Biomembranes (see also D,3.c biomembrane precursors),Cytoskeleton, organelle biogenesis (40)					
<peroxisom>					
Contig1486_r2g02a1.f1	315	1.4e-27	130	588	gnl|PID|e1169881 (AJ002536) PMP20=PEROXISOMAL MEMBRANE PROTEIN
					[Schizosaccharomyces pombe]
w7h12a1.r1	184	7.8e-25	357	521	sp|P78723|PEXE_PICAN PEROXISOMAL MEMBRANE PROTEIN PER10
					(PEROXIN-14)>gi|1777749 (U46195) Per10p [Pichia angusta
Contig1043_d2h01a1.r1	251	9.8e-21	248	541	gnl|PID|e1132728 (Z99167) putative peroxisomal organisation and
					biogenesisprotein [Schizosaccharomyces pombe]
Contig866_y6h08a1.f1	176	4.4e-12	3	203	sp|P21245|P47A_C PEROXISOMAL MEMBRANE PROTEIN PMP47A >pir||A23667
					47Kperoxisomal membrane protein - yeast (Candida boi
Contig1058_o0h05a1.f1	142	3.2e-09	292	435	gi|2769700 (U58050) peroxisomal-like protein [Aspergillus
					fumigatus]
Contig194_i3c05a1.r1	157	3.3e-09	37	564	sp|Q12462|PEXB_Y PEROXISOMAL MEMBRANE PROTEIN PMP27
					(PEROXIN-11)>pir||A56509 peroxisomal membrane protein PMP27 - yeas
Contig1023_d1a12a1.f1	137	4.7e-06	12	218	gi|2992543 (AF021797) peroxisomal receptor for PTS2-containing
					proteins Pex7p[Pichia pastoris]
<KINESIN>					
x1d12a1.r1	562	1.3e-52	19	348	sp|P17120|BIMC_EMENI KINESIN-LIKE PROTEIN BIMC >pir||A34795
					kinesin-relatedprotein bimC - Emericella nidulans >
Contig755_q0b12a1.f1	309	6.9e-26	353	538	sp|P28739|KLPA_E KINESIN-LIKE PROTEIN KLPA >pir||A44337
					kinesin-relatedprotein KLPA - Emericella nidulans >gi|2704 (X6
<tubulin>					
r3f05a1.f1	484	1.9e-45	3	290	sp|P18695|TBG_EMENI TUBULIN GAMMA CHAIN >pir||S03916 tubulin gamma chain

					-Emericella nidulans >gi|2363 (X15479
c6h05a1.r1	280	1.5e-23	23	184	sp|P24633|TBA1_EMENI TUBULIN ALPHA-1 CHAIN >pir||S13336 tubulin alpha-1 chain- Emericella nidulans
m5g06a1.r1	196	5e-12	128	319	sp|P18695|TBG_EMENI TUBULIN GAMMA CHAIN >pir||S03916 tubulin gamma chain -Emericella nidulans >gi|2363 (X15479
<ankyrin>					
c8f04a1.f1	274	3.5e-23	151	711	gi|2447128 (U42580) contains 10 ankyrin-like repeats; similar to humanankyrin, corresponds to Swiss-P
c3c08a1.r1	261	1.2e-20	18	617	gi|1841966 (U65916) ankyrin [Rattus norvegicus]
<VACUOLAR ASSEMBLY PROTEIN VPS39>					
m6b12a1.r1	146	1.8e-06	260	637	sp|Q07468|VP39_YEAST VACUOLAR ASSEMBLY PROTEIN VPS39 (VACUOLAR MORPHOGENESISPROTEIN VAM6) >pir||S67613 probable
<cytoskeleton assembly control protein>					
Contig1107_c6c12a1.f1	259	1.8e-28	216	563	pir||S63211 cytoskeleton assembly control protein SLA2 - yeast (Saccharomycescerevisiae) >gnl|PID|e239710 (Z71519=TRANSMEMBRANE PROTEIN MOP2
1.actin-see also mitosis					
<actin>					
m8d09a1.r1	916	3.1e-91	29	664	sp|P78712|ARP3_NEUCR ACTIN-LIKE PROTEIN 3 >gi|1718497 (U79737) actin-relatedprotein 3 [Neurospora crassa]
z5e10a1.f1	438	7.1e-40	117	659	sp|P53946|ARP5_YEAST ACTIN-LIKE PROTEIN ARP5 >pir||S58718 probable nuclearprotein YNL059c - yeast (Saccharomyce
Contig1684_a1c05f2.f1	386	4.6e-35	567	788	sp|P20359|ACTG_E ACTIN, GAMMA >pir||JT0385 actin gamma - Emericellanidulans >gi|168005 (M22869) gamma-actin [Emericell
Contig42_z5a09a1.f1	121	0.00015	214	366	sp|P53459|ACT6_D ACTIN 6 >gi|1098579 (U27837) actin [Diphyllobothriumdendriticum]
<PROFILIN-assembly of actin monomers>					
*Contig1634_c8a05a1.f1	243	5.5e-28	1287	1559	pdb|1HLU|A Chain A, Structure Of Bovine Beta-Actin-Profilin Complex With ActinBound Atp Phosphates Solvent Acces
Contig1788_g5b09a1.f1	283	3.6e-24	181	546	sp|P39825|PROF_S PROFILIN >pir||A53952 profilin - fission yeast(Schizosaccharomyces pombe)
<ARP2/3 COMPLEX-actin polymerization>					
Contig1692_c4c11a1.f1	813	2.6e-80	174	1052	sp|O14241|AR34_S PROBABLE ARP2/3 COMPLEX 34 KD SUBUNIT (P34-ARC)>gnl|PID|e339970 (Z98981) probable Arp2-3 complex subu
m6f08a1.r1	617	1.5e-59	60	563	sp|O15509|AR20_HUMAN ARP2/3 COMPLEX 20 KD SUBUNIT (P20-ARC) >gi|2282040(AF006087) p20-Arc [Homo sapiens] >gi|24
Contig495_o1b04a1.f1	499	5e-47	25	582	sp|P78774|AR41_S PROBABLE ARP2/3 COMPLEX 41 KD SUBUNIT (P41-ARC)
c0h11a1.r1	464	2.3e-43	98	484	sp|P32381|ARP2_YEAST ACTIN-LIKE PROTEIN ARP2 >pir||S20225 actin-like proteinACT2 - yeast (Saccharomyces cerevis
Contig726_w6b05a1.f1	181	9.5e-24	116	361	sp|O15145|AR21_H ARP2/3 COMPLEX 21 KD SUBUNIT (P21-ARC) >gi|2282038(AF006086) p21-Arc [Homo sapiens]
<fimbrin-actin bundling>					

m6e11a1.r1	507	6.8e-48	202 840	sp|P32599|FIMB_YEAST FIMBRIN (ABP67) >pir||S29320 fimbrin - yeast(Saccharomyces cerevisiae) >gi|4420 (X63867) f
Contig229_h8h08a1.f1	278	1e-22	378 614	sp|P32599|FIMB_Y FIMBRIN (ABP67) >pir||S29320 fimbrin - yeast(Saccharomyces cerevisiae) >gi|4420 (X63867) fimbrin [Sac
<COFILIN-actin binding>				
Contig998_r7c08a1.f1	248	1.8e-20	358 798	sp|Q03048|COFI_Y COFILIN >pir||A44397 cofilin - yeast (Saccharomycescerevisiae) >pdb|1CFY|A Saccharomyces cerevisiae >
<ACTIN-BINDING PROTEIN>				
Contig248_h1d05a1.f1	270	8.9e-23	179 625	sp|P25229|CAPA_X ACTIN-BINDING PROTEIN CHAIN A (ABP-A) >pir||S07105actin-binding protein chain a, nuclear - African cl
2.choline				
<choline dehydrogenase>				
f2b11a1.r1	269	5.8e-22	80 571	gi|1657509 (U73857) choline dehydrogenase [Escherichia coli]
<chitin synthase>				
Contig542_c7b10a1.f1	780	1.7e-89	242 691	pir||JC2315 chitin synthase (EC 2.4.1.16) chsB - Emericella nidulans>gnl|PID|d1005340 (D21269) chitin synthase [E
f5c06a1.f1	246	5.5e-19	268 528	pir||JC2408 chitin synthase (EC 2.4.1.16), class I - Emericella nidulans>prf||2102237A chitin synthase
3. other				
<phosphatidylethanolamine methyltransferase>				
Contig1042_o4a10a1.f1	184	6.4e-11	296 721	gi|2209107 (AF004113) phosphatidylethanolamine methyltransferase[Schizosaccharomyces pombe]
<ACETYLGLUCOSAMINE-6-PHOSPHATE DEACETYLASE>				
Contig1645_g2a12a1.f1	351	2.3e-31	249 902	sp|P34480|NAGA_C PUTATIVE N-ACETYLGLUCOSAMINE-6-PHOSPHATE DEACETYLASE(GLCNAC 6-P DEACETYLASE) >pir||S31124 hypothetica
Contig1139_g3h11a1.r1	348	4.8e-31	442 1053	sp|P34480|NAGA_C PUTATIVE N-ACETYLGLUCOSAMINE-6-PHOSPHATE DEACETYLASE(GLCNAC 6-P DEACETYLASE) >pir||S31124 hypothetica
<DIACETYLMURAMIDASE>				
Contig1142_g2c04a1.f1	614	3.2e-59	83 679	sp|P00721|LYCH_C N,O-DIACETYLMURAMIDASE (LYSOZYME CH) >pir||MUKAD lysozyme(EC 3.2.1.17) - fungus (Chalara sp.)
Contig520_o4b02a1.r1	380	4.1e-52	90 407	sp|P00721|LYCH_C N,O-DIACETYLMURAMIDASE (LYSOZYME CH) >pir||MUKAD lysozyme(EC 3.2.1.17) - fungus (Chalara sp.)
<Glycosyltransferase-glycopeptidolipid biosyn>				
n8d01a1.r1	377	4.2e-34	48 488	gi|871531 (X87947) glycosyltransferase [Saccharomyces cerevisiae]
n8d01a1.f1	249	6.6e-20	165 512	gnl|PID|d1022233 (AB004534) Glycosyltransferase [Schizosaccharomyces pombe]
C. cell cycle control (17)				
<cell division protein>				
c6f09a1.r1	423	3.4e-58	17 454	gnl|PID|d1022237 (AB004534) cdc2 kinase homologue [Schizosaccharomyces

pombe]

Contig930_c5g03a1.f1	242	8.4e-20	656	1009	sp|P51953|CDK7_C CELL DIVISION PROTEIN KINASE 7 (40 KD PROTEIN KINASE)(P40 MO15) (CDC2/CDK2,4-ACTIVATING KINASE) >gnl|
<CELL DIVISION CONTROL PROTEIN>					
Contig583_c4g06a1.r1	815	1.6e-80	276	1301	sp|P53699|CC4_CA CELL DIVISION CONTROL PROTEIN 4 >gnl|PID|e234056 (X96763)CDC4 gene product [Candida albicans]
p0g11a1.r1	709	2.6e-69	7	465	sp|P25694|CC48_YEAST CELL DIVISION CONTROL PROTEIN 48 >pir||S67669 celldivision control protein CDC48 - yeast (
o8a05a1.r1	646	1.3e-62	40	606	sp|P25694|CC48_YEAST CELL DIVISION CONTROL PROTEIN 48 >pir||S67669 celldivision control protein CDC48 - yeast (
j7e08a1.r1	426	6e-38	40	543	sp|P43069|CC25_CANAL CELL DIVISION CONTROL PROTEIN 25
Contig1269_u4c02a1.r1	247	6.7e-20	142	507	gnl|PID|e334053 (Z98533) hypothetical cell division control protein[Schizosaccharomyces pombe]
<cell cycle protein>					
m6h05a1.r1	480	5.1e-45	345	695	gnl|PID|d1033372 (AB000281) krev-1 [Neurospora crassa]
c4f02a1.r1	449	9.3e-42	43	735	sp|P50582|HSK1_SCHPO CELL CYCLE PROTEIN KINASE HSK1 >pir||S56143 proteinkinase hsk1 (EC 2.7.1.-) - fission yeas
d4e12a1.f1	346	7.7e-30	47	517	gnl|PID|e1291642 (AL023288) AAA family ATPase protein [Schizosaccharomycespombe]
Contig593_c4d04a1.f1	292	1.3e-24	166	534	sp|P50582|HSK1_S CELL CYCLE PROTEIN KINASE HSK1 >pir||S56143 proteinkinase hsk1 (EC 2.7.1.-) - fission yeast (Schizosa
*Contig997_j7h03a1.f1	126	9.5e-07	316	555	gi|2697005 (U59435) cell cycle protein p38-2G4 homolog [Homo sapiens]
<SCH9 protein-cell progress through G1>					
Contig554_c6c06a1.f1	212	1.2e-34	295	564	gi|4426 (X12560) SCH9 protein (AA 1-824) [Saccharomyces cerevisiae]
<G1/S-SPECIFIC CYCLIN-essential for movement from g1-S>					
r5d06a1.r1	152	2.9e-08	423	734	sp|P24867|CG16_YEAST G1/S-SPECIFIC CYCLIN PCL1 (CYCLIN HCS26) >pir||A40027cyclin G1 homolog HCS26 - yeast (Sacc
<cullin-neg regulator of cell cycle>					
f0f02a1.r1	433	2.3e-39	11	592	sp|Q13616|CUL1_HUMAN CULLIN HOMOLOG 1 G1 to G0 phase(CUL-1) >gi|1381142 (U58087) Hs-CUL-1[Homo sapiens]
<SENESCENCE>					
*Contig1193_x3c04a1.f1	286	1.9e-24	414	950	sp|Q64374|SM30_M SENESCENCE MARKER PROTEIN-30 (SMP-30) (REGUCALCIN) (RC)>gi|1144000 (U28937) senescence marker protein
<apoptosis>					
Contig890_v7d01a1.f1	200	2.1e-15	382	606	sp|P46966|DAD1_H DEFENDER AGAINST CELL DEATH 1 (DAD-1) >pir||A54437apoptotic cell death regulator DAD1 - human >gnl|PI

E. Mitosis/cytokinesis (15)

1.*MITOSIS*

<MITOSIS>

c4f05a1.r1 981 4.1e-98 19 567 sp|P30303|MPIP_EMENI M-PHASE INDUCER PHOSPHATASE >pir||S24395protein-tyrosine-phosphatase (EC 3.1.3.48) nimT -

<DNA DAMAGE CHECKPOINT PROTEIN-allows entry into Mitosis>

*Contig1730_c1a04a1.f1 659 5.6e-85 115 612 sp|P42656|RA24_S DNA DAMAGE CHECKPOINT PROTEIN RAD24 >gnl|PID|e1251047(AL021817) dna damage checkpoint protein [Schizo

<centromere>

Contig1160_w5c12a1.f1 938 1.4e-93 333 980 sp|O43100|CBF5_E CENTROMERE/MICROTUBULE BINDING PROTEIN CBF5(CENTROMERE-BINDING FACTOR 5) (NUCLEOLAR PROTEIN CBF5) >gi

<CHROMOSOME SEGREGATION PROTEIN-with microtubule, migration of chromosomes>

Contig627_c1b12a1.r1 672 1.7e-64 132 704 sp|P38989|SMC2_Y CHROMOSOME SEGREGATION PROTEIN SMC2 (DA-BOX PROTEIN SMC2)>pir||A56157 chromosome segregation protein

o6g05a1.r1 299 1.2e-24 12 536 sp|P33307|CSE1_YEAST CHROMOSOME SEGREGATION PROTEIN CSE1 >pir||A48083chromosome segregation protein CSE1 - yeas

o6g05a1.f1 197 1e-13 68 679 sp|P33307|CSE1_YEAST CHROMOSOME SEGREGATION PROTEIN CSE1 >pir||A48083chromosome segregation protein CSE1 - yeas

<dynamin-molecular motor, associated with mocrotuble, endocytosis>

Contig862_r7d07a1.f1 920 1.1e-91 13 708 sp|P21576|VPS1_Y VACUOLAR SORTING PROTEIN 1 >pir||S25820 dynamin-relatedprotein VPS1 - yeast (Saccharomyces cerevisiae

g4c03a1.r1 401 7.2e-36 92 412 sp|P54861|DNM1_YEAST DYNAMIN-RELATED PROTEIN DNM1 >pir||S64742 dynamin-relatedprotein DNM1 - yeast (Saccharomyc

Contig339_g4c03a1.f1 309 6.7e-26 172 498 sp|Q09748|YB68_S DYNAMIN-LIKE PROTEIN C12C2.08 >gi|984696 (Z54140)dynamin-related protein [Schizosaccharomyces pombe]

<dynein-molecular motor>

Contig171_m6g04a1.f1 123 7.8e-07 310 405 gnl|PID|e326960 (Z97340) strong similarity to dynein light chain [Arabidopsisthaliana]

<nuclear positioning>

e7d01a1.r1 927 1.9e-92 24 620 gnl|PID|e1286087 (AJ003163) apsB [Emericella nidulans]

<DMR-N9 PROTEIN-regulation of mitosis>

k5h10a1.r1 294 1.2e-24 9 497 sp|Q09019|DMR9_HUMAN DMR-N9 PROTEIN (PROTEIN 59) >pir||A49364 59 protein,brain - human (fragment) >gi|306712 (L

<CALTRACTIN-ASSOCIATED WITH THE POLES OF THE MITOTIC SPINDLES>

s8b04a1.f1 172 1.9e-12 163 582 sp|P53441|CATR_NAEGR CALTRACTIN (CENTRIN) >gi|972963 (U21725) centrin[Naegleria gruberi]

2.cytokinesis

<cytokinesis>

k9g02a1.f1 264 4e-22 362 619 gnl|PID|e347848 (AJ001587) sid3 [Schizosaccharomyces pombe]-septum deposition

<f-actin-contractile ring during cytokinesis>

Contig989_j9f08a1.f1 152 5.7e-09 145 420 sp|P20111|AACS_C ALPHA-ACTININ, SKELETAL MUSCLE ISOFORM (F-ACTIN CROSSLINKING PROTEIN) >pir||S02032 alpha-actinin 2, s

<TROPOMYOSIN-component of contractile ring>

o4g05a1.r1 241 8.9e-20 128 508 sp|Q02088|TPM_SCHPO TROPOMYOSIN >pir||S27127 tropomysin -

yeast(Schizosaccharomyces pombe) >gi|173517 (L04126)

F. Meiosis (2)

<Rad9-required for meiotic chromosome condensation and synapsis>

w7a08a1.f1	172	9.6e-09	9	455	gi|1353390	(U34998) Rad9 [Coprinus cinereus]

<condensin-chromosome condensation protein>

j0a03a1.r1	154	2.5e-09	124	330	gi|3298547	(AC004681) putative condensin protein [Arabidopsis thaliana

III. DNA synthesis

A. DNA replication (8)

<DNA POLYMERASE>

Contig467_o0b09a1.f1	539	2.6e-50	3	410	sp|P46588|DPOD_C	DNA POLYMERASE DELTA LARGE CHAIN (DNA POLYMERASE III)>gi|951001 (X88804) DNA-directed DNA polymerase

<replication factor>

Contig793_v3e05a1.f1	836	9.6e-83	128	868	sp|P35250|AC14_H	ACTIVATOR 1 40 KD SUBUNIT (REPLICATION FACTOR C 40 KDSUBUNIT) (A1 40 KD SUBUNIT) (RF-C 40 KD SUBUNIT)=human
Contig443_d5e10a1.f1	694	1.1e-67	194	1006	sp|Q92372|RFA1_S	REPLICATION FACTOR-A PROTEIN 1 (SINGLE-STRANDEDDNA-BINDING PROTEIN P68 SUBUNIT) >gi|1502413 (U59385)s

<Single-stranded DNA-binding protein>

<DnaJ protein>

Contig1596_i2e07a1.r1	485	6.2e-67	808	1233	sp|P25491|MAS5_Y	MITOCHONDRIAL PROTEIN IMPORT PROTEIN MAS5 (PROTEIN YDJ1)>pir||S26703 dnaJ protein homolog YDJ1 - yeas
m5e08a1.r1	322	2.8e-28	145	807	sp|P25491|MAS5_YEAST	MITOCHONDRIAL PROTEIN IMPORT PROTEIN MAS5 (PROTEIN YDJ1)>pir||S26703 dnaJ protein homolog
Contig1123_x8c07a1.f1	292	4.2e-25	260	745	sp|Q09912|PSI_SC	PSI PROTEIN >pir||S55900 DNAJ-like protein homolog -fission yeast (Schizosaccharomyces pombe) >gi|953
n8g02a1.r1	138	2.1e-08	206	394	gi|1127833	(U40992) heat shock protein hsp40 homolog [Homo sapiens]>gi|1518918 (U41290) DNAJ homolog
h4a08a1.f1	128	0.00016	268	483	gi|1707079	(U80451) contains strong similarity to a DNAJ-like domain(PS:PS00636) [Caenorhabditis eleg

B. DNA modification and repair (7)

<DNA LYASE>

x1g07a1.r1	362	1.5e-32	55	588	sp|P22936|APN1_YEAST	DNA-(APURINIC OR APYRIMIDINIC SITE) LYASE (APENDONUCLEASE) (APURINIC-APYRIMIDINIC ENDONUCL

<endonuclease IV>

<DNA REPAIR PROTEIN>

w6c06a1.r1	737	2.9e-72	8	544	sp|P41410|RA54_SCHPO	DNA REPAIR PROTEIN RHP54 >gnl|PID|e325327 (Z97208) Rad54p[Schizosaccharomyces pombe]
d4c06a1.r1	359	1e-30	102	539	sp|Q03468|ERC6_HUMAN	EXCISION REPAIR PROTEIN ERCC-6 (COCKAYNE SYNDROME PROTEINCSB) >pir||A44224 DNA repair heli

w8d07a1.r1	234	1.4e-17	22 480	sp|P53692|RA18_SCHPO DNA REPAIR PROTEIN RAD18 >gi|1150622 (X80929) rad18 geneproduct [Schizosaccharomyces pombe
Contig654_b0g01a1.f1	229	5.8e-17	242 469	sp|P12753|RA50_Y DNA REPAIR PROTEIN RAD50 (153 KD PROTEIN) >pir||BWBYDLRAD50 protein - yeast (Saccharomyces cerevisiae
*Contig756_v1g09a1.r1	185	1.3e-10	22 342	gnl|PID|e349697 (Z99292) dna repair protein [Schizosaccharomyces pombe]
j7f09a1.r1	151	1.9e-06	89 460	sp|P12753|RA50_YEAST DNA REPAIR PROTEIN RAD50 (153 KD PROTEIN) >pir||BWBYDLRAD50 protein - yeast (Saccharomyces

C. DNA packaging (7)

1.Histone

<Histones, class H1 (or I, or f1)>

Contig1551_g2h03a1.f1	180	2.4e-13	86 298	sp|P53551|H1_YEA HISTONE H1 >pir||S69056 histone H1 - yeast (Saccharomycescerevisiae) >gi|1244786 (U43703) Lpi17p [Sac

<Histones, class H2a (or IIb1, or f2a2)>

Contig1795_k8h08a1.f1	500	3.7e-47	216 611	sp|P08844|H2A_EM HISTONE H2A >pir||A27332 histone H2A - Emericella nidulans>gi|168053 (M18258) histone H2A [Emericella
*Contig1045_n8g06a1.r1	175	8.6e-10	169 708	gi|3395780 (AF058445) histone macroH2A1.1 [Gallus gallus]

<Histones, class H2b (or IIb2, or f2b)>

Contig1765_g4e06a1.f1	443	3.7e-41	121 540	sp|P23754|H2B_EM HISTONE H2B >pir||S11937 histone H2B - Emericella nidulans>gi|296335 (X55547) H2B [Emericella nidulan

<Histones, class H3 (or III, or f3)>

Contig1767_d1h05a1.f1	669	4.3e-65	96 503	sp|P23753|H3_EME HISTONE H3 >pir||S11938 histone H3 - Emericella nidulans>gi|296337 (X55548) H3 [Emericella nidulans]

<Histones, class H4 (or IV, or f2a1)>

Contig1427_i8e09a1.r1	409	1.6e-37	141 386	sp|P23750|H41_EM HISTONE H4.1 >pir||S11939 histone H4.1 - Emericellanidulans >gi|296341 (X55549) H4.1 [Emericella nidu
Contig1739_j7e02a1.f1	406	3e-37	237 482	sp|P23751|H42_EM HISTONE H4.2 >pir||S11940 histone H4.2 -

2. DNA-binding (3)

<DNA-binding protein amdA>

m3b09a1.r1	374	9.6e-33	53 592	pir||S61908 DNA-binding protein amdA - Emericella nidulans >gi|454103 (L28810)regulatory protein [Emer

<NONHISTONE>

Contig1353_g2f05a1.f1	280	7e-24	56 316	sp|P11633|NHPB_Y NONHISTONE CHROMOSOMAL PROTEIN 6B >pir||B35072 nonhistonechromosomal protein NHP6B - yeast (Saccharom

<SINGLE-STRAND TELOMERIC DNA-BINDING PROTEIN>

Contig332_g4e10a1.f1	166	1.8e-09	510 776	sp|P25555|GBP2_Y SINGLE-STRAND TELOMERIC DNA-BINDING PROTEIN GBP2(G-STRAND BINDING PROTEIN 2) (RAP1 LOCALIZATION FACTO

IV. Gene Expression
A. Transcription
1. RNA Polymerase (9)

<RNA POLYMERASE I, rRNA>				
c7b07a1.f1	352	3.8e-30	389 703	sp|P22138|RPA2_YEAST DNA-DIRECTED RNA POLYMERASE I 135 KD POLYPEPTIDE (A135)(RNA POLYMERASE I SUBUNIT 2) >pir||
Contig80_13b11a1.r1	397	1e-34	24 551	sp|P15398|RPA1_S DNA-DIRECTED RNA POLYMERASE I 190 KD POLYPEPTIDE>pir||JS0080 DNA-directed RNA polymerase (EC 2.7.7.6)
13b11a1.f1	225	2.1e-16	248 481	sp|P15398|RPA1_SCHPO DNA-DIRECTED RNA POLYMERASE I 190 KD POLYPEPTIDE>pir||JS0080 DNA-directed RNA polymerase (
Contig479_d2h10a1.f1	228	2.6e-18	136 486	sp|P32529|RPA9_Y DNA-DIRECTED RNA POLYMERASE I 13.7 KD POLYPEPTIDE (A12.2)>pir||A48107 DNA-directed RNA polymerase (EC
<RNA POLYMERASE II, mRNA>				
a0b11a1.f1	476	2.2e-43	129 485	sp|P08518|RPB2_YEAST DNA-DIRECTED RNA POLYMERASE II 140 KD POLYPEPTIDE (B150)(RNA POLYMERASE II SUBUNIT 2) >pir
<RNA POLYMERASE III, tRNA>				
Contig535_c8b02a1.f1	342	2.2e-30	269 805	sp|P32910|RPC6_Y DNA-DIRECTED RNA POLYMERASE III 36 KD POLYPEPTIDE (C34)>pir||A45107 DNA-directed RNA polymerase (EC 2
<RNA Polymerase>				
Contig1398_j9h11a1.f1	220	1.8e-17	117 548	gi|3372230 (AF017074) RNA polymerase I, II and III 16.5 kDa subunit[Arabidopsis thaliana]

2. Regulation (48)

<transcription factor>				
x1g11a1.r1	1084	4.9e-109	9 635	gi|3411264 (AF080600) homeodomain DNA-binding transcription factor [Emericellanidulans]
s3f07a1.r1	912	8.4e-91	28 681	sp|P52958|CT1A_FUSSO CUTINASE TRANSCRIPTION FACTOR 1 ALPHA >gi|1262912(U51671) cutinase transcription factor 1
x3c12a1.r1	372	1.4e-33	68 415	gnl|PID|e1311364 (Z99168) putative heat shock transcription factor[Schizosaccharomyces pombe]
*Contig380_f5h12a1.r1	307	4.7e-26	63 506	gi|1100209 (L49345) transcription factor ZFM1 [Homo sapiens]
*Contig286_g8f03a1.f1	310	6.8e-26	237 857	gnl|PID|e280495 (Z46606) helicase-like transcription factor [Homo sapiens]
g6f09a1.r1	289	4.5e-24	60 296	sp|P52890|ATF1_SCHPO TRANSCRIPTION FACTOR ATF1 (TRANSCRIPTION FACTOR MTS1)(PROTEIN SSS1) >pir||S66147 transcrip
*Contig1222_g6h05a1.r1	219	5.9e-14	55 537	sp|P48361|AS10_Y ASK10 PROTEIN >pir||S64402 probable transcription

					factorASK10 - yeast (Saccharomyces cerevisiae) >gnl
Contig927_c4c12a1.f1	182	1.8e-13	311	541	sp|P40096|NCB1_Y CLASS 2 TRANSCRIPTION REPRESSOR >pir||S50662 hypotheticalprotein YER159c - yeast (Saccharomyces cerev
Contig1252_g4c01a1.r1	150	1e-06	11	292	sp|P28349|NIT4_N NITROGEN ASSIMILATION TRANSCRIPTION FACTOR NIT-4>pir||A41696 regulatory protein nit-4 - Neurospora cr
*Contig132_m5h10a1.r1	139	2.4e-05	190	669	pir||S43749 transcription factor nft1 - fission yeast (Schizosaccharomycespombe)
o9c12a1.r1	123	4.6e-05	376	561	pir||S61704 probable transcription factor YPL230w - yeast (Saccharomycescerevisiae) >gi|1181258 (X9456
n8h07a1.r1	132	0.0002	24	500	sp|P28349|NIT4_NEUCR NITROGEN ASSIMILATION TRANSCRIPTION FACTOR NIT-4>pir||A41696 regulatory protein nit-4 - Ne
Contig1461_o6c05a1.f1	130	0.00021	548	1069	sp|P28348|NIRA_E NITROGEN ASSIMILATION TRANSCRIPTION FACTOR NIRA>pir||A41697 nitrate assimilation regulatory protein n
Contig546_c6g06a1.f1	126	0.00024	571	720	sp|P56095|AP1_KL AP-1-LIKE TRANSCRIPTION FACTOR >gi|2245654 (AF006499)transcription factor KlYAP1 [Kluyveromyces lacti
<SUPRESSOR OF STEM-LOOP PROTEIN-transcription initiation, pol binding>					
i0d11a1.r1	230	3.6e-18	261	473	gnl|PID|d1032706 (AB016221) SSL1 [Schizosaccharomyces pombe]=SUPRESSOR OF STEM-LOOP PROTEIN, translation initiation and UV resistance
<TRANSCRIPTION INITIATION FACTOR TFIID>					
Contig1018_j7a02a1.f1	632	3.8e-61	522	899	sp|Q12731|TF2D_E TRANSCRIPTION INITIATION FACTOR TFIID (TATA-BOX FACTOR)(TATA SEQUENCE-BINDING PROTEIN) (TBP) >gi|8878
x1d01a1.r1	520	8.2e-49	7	636	sp|P38129|T2D4_YEAST TRANSCRIPTION INITIATION FACTOR TFIID 90 KD SUBUNIT(TAFII-90) >pir||S34023 TATA box-bindin
Contig1605_e9a11a1.r1	490	4.3e-46	441	863	sp|Q12731|TF2D_E TRANSCRIPTION INITIATION FACTOR TFIID (TATA-BOX FACTOR)(TATA SEQUENCE-BINDING PROTEIN) (TBP) >gi|8878
Contig1682_g9c11a1.f1	409	1.7e-37	512	751	sp|Q12731|TF2D_E TRANSCRIPTION INITIATION FACTOR TFIID (TATA-BOX FACTOR)(TATA SEQUENCE-BINDING PROTEIN) (TBP) >gi|8878
f0d08a1.r1	259	9.7e-33	96	539	sp|P53040|T2D5_YEAST TRANSCRIPTION INITIATION FACTOR TFIID 60 KD SUBUNIT(TAFII-60) >pir||S64120 TATA box-bindin
Contig1037_c1e03a1.f1	217	2.4e-24	498	725	sp|P38129|T2D4_Y TRANSCRIPTION INITIATION FACTOR TFIID 90 KD SUBUNIT(TAFII-90) >pir||S34023 TATA box-binding protein-a
i8b08a1.r1	144	1.7e-06	185	484	sp|P49846|T2D4_DROME TRANSCRIPTION INITIATION FACTOR TFIID 85 KD SUBUNIT (P85)(TAFII-80) >pir||A54593 TFIID sub
<TRANSCRIPTION INITIATION PROTEIN SPT6>					
n8a10a1.r1	352	5.4e-30	81	650	sp|P23615|SPT6_YEAST TRANSCRIPTION INITIATION PROTEIN SPT6 >pir||A36468 SPT6protein - yeast (Saccharomyces cere
<HAP3 protein-transcription factor>					
Contig1371_e9d05a1.r1	459	5.5e-60	18	296	pir||JC6080 HAP3 protein - Emericella nidulans >gi|1017716 (U35341) HapC[Emericella nidulans]
<RNA helicase>					
g6h04a1.r1	821	3.6e-81	55	825	gi|2408027 (Z99162) atp-dependent rna helicase [Schizosaccharomyces pombe]
c7d09a1.r1	640	5.6e-62	24	716	sp|Q09747|YB66_SCHPO PUTATIVE ATP-DEPENDENT RNA HELICASE C12C2.06

```
                                                                >gi|984214(Z54140) probable ATP-dependent RNA
w5b11a1.r1                542  4.7e-51    14 412   sp|Q09719|YA47_SCHPO  PUTATIVE ATP-DEPENDENT RNA HELICASE-translation
                                                   initiation C31A2.07C>pir||S59645 hypothetical protein SPAC31A2.07
i8b11a1.r1                518  4.7e-49    12 392   sp|P24782|DBP2_SCHPO  P68-LIKE PROTEIN >pir||S14048 RNA helicase dbp2 -
                                                   fissionyeast (Schizosaccharomyces pombe)
j9a01a1.r1                527    1e-48    51 485   gi|172764             (M83755) STH1 protein, [Saccharomyces cerevisiae],snf2
                                                    RNA helicase homolog?
Contig1348_e7a04a1.r1     489  5.5e-46   159 845   sp|P20447|DBP3_Y      PROBABLE ATP-DEPENDENT RNA HELICASE DBP3 (HELICASE
                                                   CA3)>pir||S30805 probable RNA helicase CA3 - yeast
n2a07a1.r1                473  9.3e-44   213 668   sp|Q09903|YAJ3_SCHPO  PUTATIVE ATP-DEPENDENT RNA HELICASE
                                                   C30D11.03>pir||S62561 hypothetical protein SPAC30D11.0
g6h04a1.f1                420  1.1e-38   193 531   sp|Q07478|HE47_YEAST  PROBABLE ATP-DEPENDENT RNA HELICASE P47
                                                   HOMOLOG>pir||S67620 hypothetical protein YDL084w -
Contig1557_g3e04a1.f1     275  3.7e-35   405 716   sp|P20447|DBP3_Y      PROBABLE ATP-DEPENDENT RNA HELICASE DBP3 (HELICASE
                                                   CA3)>pir||S30805 probable RNA helicase CA3 - yeast
Contig1337_w4b05a1.f1     181  2.5e-10   577 825   gnl|PID|e248472       (X98130) RNA helicase [Arabidopsis thaliana]
                                                   >gnl|PID|e245943(X97970) RNA helicase [Arabidopsis thali
l5f03a1.f1                162  2.7e-09   389 601   sp|Q09719|YA47_SCHPO  PUTATIVE ATP-DEPENDENT RNA HELICASE
                                                   C31A2.07C>pir||S59645 hypothetical protein SPAC31A2.07
<HELICASE MOT1-essential>
g3h10a1.r1                709  6.7e-68    15 749   sp|P32333|MOT1_YEAST  PROBABLE HELICASE MOT1 >pir||S22775 MOT1 protein -
                                                   yeast(Saccharomyces cerevisiae) >gi|171
y8h08a1.f1                207  2.8e-13   214 546   sp|P32333|MOT1_YEAST  PROBABLE HELICASE MOT1 >pir||S22775 MOT1 protein -
                                                   yeast(Saccharomyces cerevisiae) >gi|171
<regulatory protein creA-carbon catabolite repression>
Contig1541_g5c01a1.f1     825  1.4e-81    10 588   sp|Q01981|CREA_E      DNA-BINDING PROTEIN CREA >gi|168035 (L03563)
                                                    CREA[Emericella nidulans]
<HAC1-unfolded protein response pathway, transcrip activation>
Contig1782_c7a10a1.f1     130  1.2e-05   435 527   pir||S56223           HAC1 protein - yeast (Saccharomyces cerevisiae)
<TRANSCRIPTIONAL REPRESSOR>
k5g05a1.f1                451  5.9e-42    10 384   sp|P78706|RCO1_NEUCR  TRANSCRIPTIONAL REPRESSOR RCO-1 >gi|1698504
                                                   (U57061)rco-1 gene product [Neurospora crassa]
Contig1520_g6a01a1.f1     258  9.8e-21   646 930   sp|O14063|IMA1_S      IMPORTIN ALPHA SUBUNIT (KARYOPHERIN ALPHA
                                                   SUBUNIT)(SERINE-RICH RNA POLYMERASE I SUPPRESSOR PROTEIN) >
Contig673_a0c03a1.r1      959  8.5e-96   131 877   sp|O14063|IMA1_S      IMPORTIN ALPHA SUBUNIT (KARYOPHERIN ALPHA
                                                   SUBUNIT)(SERINE-RICH RNA POLYMERASE I SUPPRESSOR PROTEIN) >

<QUINATE REPRESSOR>-repressor protein in the quinic acid utilization pathway, cluster>

d4a10a1.f1                410  1.5e-36   104 658   pir||JH0262           QutR protein QUINATE REPRESSOR - Emericella nidulans
                                                   >gi|168084 (M77664) repressorprotein [Emericella nidula
<PacC-factor regulates acid and base expressed genes>
Contig573_c5d05a1.f1      138  1.3e-07   196 276   pir||S54308           DNA binding protein - Emericella nidulans >gi|695416
```

						(Z47081) DNAbinding protein [Emericella nidulans=pacC=transcription factor mediating pH regulation
<RNA-BINDING POST-TRANSCRIPTIONAL REGULATOR>						
Contig82_l3b05a1.r1	451	5.6e-42	136	654	sp|O13759|CSX1_S	RNA-BINDING POST-TRANSCRIPTIONAL REGULATOR CSX1>gnl|PID|e1198263 (Z99292) rna binding post-transcript
<HapE-transcription regulation, penicillin and acetamidase biosynthesis>						
Contig1480_h4c03a1.f1	636	3e-111	149	556	gi|2098795	(U96847) HapE [Emericella nidulans]
-Asexual development-Central regulatory pathway						
<regulatory protein brlA-transcription factor>						
o4h01a1.r1	913	6.2e-91	11	511	sp|P10069|BRLA_EMENI	REGULATORY PROTEIN BRLA (BRISTLE A PROTEIN) >pir||A28913regulatory protein brlA - Emericel
c5h06a1.r1	897	3.2e-89	131	778	sp|P10069|BRLA_EMENI	REGULATORY PROTEIN BRLA (BRISTLE A PROTEIN) >pir||A28913regulatory protein brlA - Emericel
<REGULATORY PROTEIN WETA>						
<STUA transcription factor>						
<CELL PATTERN FORMATION-ASSOCIATED PROTEIN: SPATIAL LOCALIZATION OF ABAA AND BRLA>						
Contig1369_d5d07a1.f1	543	1e-51	542	874	sp|P36011|STUA_E	CELL PATTERN FORMATION-ASSOCIATED PROTEIN >pir||A44068cell pattern formation-associated protein - Eme
p0c08a1.r1	133	3.4e-07	412	486	sp|P36011|STUA_EMENI	CELL PATTERN FORMATION-ASSOCIATED PROTEIN >pir||A44068cell pattern formation-associated pr-E. nidulans
3. Processing (19)						
a.SPLICEOSOME						
<SPLICEOSOME ASSOCIATED PROTEIN>						
l0f01a1.f1	224	6.2e-18	5	202	gnl|PID|e340019	(Z98979) putative splicosome associated protein[Schizosaccharomyces pombe]
<splicing factor>						
Contig231_h8h02a1.r1	592	6.2e-57	50	832	gi|2911284	(U97681) putative splicing factor [Schizosaccharomyces pombe]
Contig1332_a1d02c9.r1	243	5.4e-19	375	797	gnl|PID|e325342	(Z97209) splicing factor [Schizosaccharomyces pombe]
n3e12a1.r1	226	7.3e-18	396	683	gi|2749972	(AF012278) putative pre-mRNA splicing factor [Schizosaccharomycespombe]
<small nuclear ribonucleoprotein>						
Contig71_m0c10a1.r1	344	1.3e-30	19	441	gnl|PID|e339912	(Z98974) putative small nuclear ribonucleoprotein[Schizosaccharomyces pombe]
Contig1261_s9f03a1.f1	310	5.2e-27	365	607	gnl|PID|e349593	(Z99259) small nuclear ribonucleoprotein [Schizosaccharomycespombe]
Contig328_m7c12a1.r1	264	3.9e-22	166	396	gnl|PID|e1292641	(AL023534) small nuclear ribonucleoprotein F[Schizosaccharomyces pombe]
Contig907_r5d10a1.f1	234	5.6e-19	252	455	gnl|PID|e1294550	(AL023706) small nuclear ribonucleoprotein[Schizosaccharomyces pombe]

<U4/U6 SNRNA-ASSOCIATED SPLICING FACTOR PRP24>

m6f12a1.r1	259	3.7e-21	65	637	sp|P49960|PR24_YEAST U4/U6 SNRNA-ASSOCIATED SPLICING FACTOR PRP24 (U4/U6SNRPPROTEIN) >pir||S54480 U6 snRNP prot

<SPLICING FACTOR U2AF 65 KD SUBUNIT>

y6c08a1.r1	316	1.7e-27	9	530	sp|P26369|U2AF_MOUSE SPLICING FACTOR U2AF 65 KD SUBUNIT (U2 AUXILIARY FACTOR65 KD SUBUNIT) (U2 SNRNP AUXILIARY
y6c08a1.f1	259	4.9e-21	180	467	sp|P26369|U2AF_MOUSE SPLICING FACTOR U2AF 65 KD SUBUNIT (U2 AUXILIARY FACTOR65

<MITOCHONDRIAL RNA SPLICING PROTEIN>

i7h06a1.r1	224	2.8e-17	18	314	sp|Q01926|MRS2_YEAST MITOCHONDRIAL RNA SPLICING PROTEIN MRS2 PRECURSOR>pir||S62064 MRS2 protein - yeast (S. cerevisiae

<tRNA splicing endonuclease>

n5h01a1.r1	241	8.7e-29	7	351	gnl|PID|e1298615 (AL023859) putative tRNA splicing endonuclease gamma subunit[Schizosaccharomyces pombe]

<HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN K>

Contig1199_e9b12a1.f1	143	2.1e-08	266	424	sp|Q07244|ROK_HU HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN K (HNRNP K)(DC-STRETCH BINDING PROTEIN) (CSBP) (TRANSFORMATIO

<hnRNP-E1 protein>

Contig116_k0g05a1.r1	268	1.5e-22	92	559	pir||S42472 hnRNP-E1 protein - human >pir||S65678 PCBP-1 protein - human>gi|460771 (X78137) hnRNP-E1 [Homo sapien

<NAM8 PROTEIN-MITOCHONDRIAL SPLICING>

Contig1512_i2h05a1.f1	254	2.6e-18	1210	1413	gi|4026 (X64763) NAM8 gene product [Saccharomyces cerevisiae] >prf||1814447BNAM8 gene [Saccharomyces cerevisi

b.polyA addition

<POLYA>

l3h08a1.r1	260	3.4e-20	29	607	sp|Q10569|CPSA_BOVIN CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR, 160 KDSUBUNIT (CPSF 160 KD SUBUNIT) >pir|
n1e12a1.r1	208	2.2e-15	194	430	gi|632500 (U17394) polyadenylation factor 64 kDa subunit [Xenopus laevis]
l3h08a1.f1	165	4.3e-10	168	431	sp|Q10570|CPSA_HUMAN CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR, 160 KDSUBUNIT (CPSF 160 KD SUBUNIT) >gi|1

c.5' capping

<MRNA CAPPING ENZYME>

c8e06a1.r1	377	3.8e-34	59	604	sp|P40997|MCE1_SCHPO MRNA CAPPING ENZYME (MRNA GUANYLYLTRANSFERASE) (GTP--RNAGUANYLYLTRANSFERASE) >gi|562123 (U

d.other

<nucleolin-rRNA processing>

Contig1368_e9a08a1.f1	357	5.3e-32	341	781	bbs|110564 NSR1=nucleolin homolog [Saccharomyces cerevisiae, Peptide, 249 aa]

<3'-TERMINAL PHOSPHATE CYCLASE-RNA processing?>

h1d06a1.r1	264	3.9e-22	61 480	sp|Q08096|RTC1_YEAST RNA 3'-TERMINAL PHOSPHATE CYCLASE (RNA-3'-PHOSPHATECYCLASE) (RNA CYCLASE) >pir||S66692 hyp
<FIBRILLARIN>				
Contig350_g3f08a1.f1	380	1.6e-34	184 483	sp|P15646|FBRL_Y FIBRILLARIN (NUCLEOLAR PROTEIN 1) >pir||S25421 nucleolarprotein NOP1 - yeast (Saccharomyces cerevisia
<MINOR CAPSID PROTEIN C>				
j4f08a1.r1	755	3.5e-74	27 548	sp|P03711|VCAC_LAMBD MINOR CAPSID PROTEIN C (GPC) (CONTAINS: CAPSID ASSEMBLYPROTEIN NU3) >pir||VHBPCL minor cap
<QUEUINE TRNA-RIBOSYLTRANSFERASE-tRNA, guanine modification>				
Contig336_g4d01a1.r1	473	2.8e-44	10 609	sp|P15178|SYD_RA ASPARTYL-TRNA SYNTHETASE (ASPARTATE--TRNA LIGASE) (ASPRS)>pir||SYRTDT aspartate--tRNA ligase (EC 6.1.
Contig63_m0f06a1.r1	418	1.9e-38	210 737	gnl|PID|e1250585 (AL021813) phenylalanyl-trna synthetase alpha chain[Schizosaccharomyces pombe]
Contig181_m7a04a1.f1	357	3.5e-31	73 387	sp|P46655|SYEC_Y GLUTAMYL-TRNA SYNTHETASE, CYTOPLASMIC (GLUTAMATE--TRNALIGASE) (GLURS) (P85) >pir||S53934 probable glu
Contig534_c8b04a1.f1	344	1.3e-30	342 776	sp|P07284|SYSC_Y SERYL-TRNA SYNTHETASE, CYTOPLASMIC (SERINE--TRNA LIGASE)(SERRS) >pir||YSBYC serine--tRNA ligase (EC 6
Contig839_v3f08a1.f1	337	1.7e-29	152 601	gnl|PID|e1263903 (AL022103) histidyl-trna synthetase [Schizosaccharomycespombe]
Contig572_o5f10a1.r1	317	1.4e-26	214 726	sp|P40825|SYAC_Y ALANYL-TRNA SYNTHETASE, CYTOPLASMIC (ALANINE--TRNALIGASE) (ALARS) >pir||S62065 alanine--tRNA ligase (
r8a04a1.f1	160	1.8e-10	9 170	sp|Q23623|TGT_CAEEL PUTATIVE QUEUINE TRNA-RIBOSYLTRANSFERASE (TRNA-GUANINETRANSGLYCOSYLASE) (GUANINE INSERTION
4. tRNA synthetase (12)				
g8c04a1.r1	735	4.9e-72	25 558	sp|P40825|SYAC_YEAST ALANYL-TRNA SYNTHETASE, CYTOPLASMIC (ALANINE--TRNALIGASE) (ALARS) >pir||S62065 alanine--tR
Contig1136_u4f08a1.f1	675	4.7e-65	102 1055	sp|O13651|SYIC_S ISOLEUCYL-TRNA SYNTHETASE, CYTOPLASMIC (ISOLEUCINE--TRNALIGASE) (ILERS) >gnl|PID|d1022285 (AB004538)
y8e05a1.r1	648	7.4e-63	12 545	gnl|PID|e1240164 (Y12589) phenylalanyl-tRNA synthetase [Candida albicans]
v3f08a1.r1	526	7e-50	42 545	gnl|PID|e1263903 (AL022103) histidyl-trna synthetase [Schizosaccharomycespombe]
c8b04a1.r1	511	2.5e-48	76 651	sp|O14018|SYSC_SCHPO SERYL-TRNA SYNTHETASE, CYTOPLASMIC (SERINE--TRNA LIGASE)(SERRS) >gnl|PID|e351309 (Z97210)
m1h01a1.r1	496	9.7e-47	106 651	sp|P38088|SYG_YEAST GLYCYL-TRNA SYNTHETASE (GLYCINE--TRNA LIGASE) (GLYRS)>pir||S48285 probable glycine--tRNA l
n0h04a1.f1	449	9.6e-42	75 437	sp|P04802|SYDC_YEAST ASPARTYL-TRNA SYNTHETASE, CYTOPLASMIC (ASPARTATE--TRNALIGASE) (ASPRS) >pir||SYBYDC asparta
r1a02a1.r1	402	1.5e-36	207 641	sp|Q05506|SYRC_YEAST PROBABLE ARGINYL-TRNA SYNTHETASE, CYTOPLASMIC(ARGININE--TRNA LIGASE) (ARGRS) >pir||S70106
i0b08a1.r1	398	2.5e-36	12 443	sp|P04803|SYWM_YEAST TRYPTOPHANYL-TRNA SYNTHETASE, MITOCHONDRIAL(TRYPTOPHAN--TRNA LIGASE) (TRPRS) >pir||YWBYM t

r1a02a1.f1	244	4.4e-19	151	366	sp|Q05506|SYRC_YEAST PROBABLE ARGINYL-TRNA SYNTHETASE, CYTOPLASMIC(ARGININE--TRNA LIGASE) (ARGRS) >pir||S70106
m7a04a1.r1	223	1e-16	85	624	gnl|PID|e339154 (Z98849) putative glutamyl-trna synthetas [Schizosaccharomycespombe]
Contig1126_u4c01a1.r1	201	3.5e-15	169	531	sp|P43835|SYW_HA TRYPTOPHANYL-TRNA SYNTHETASE (TRYPTOPHAN--TRNA LIGASE)(TRPRS) >pir||C64083 tryptophan--tRNA ligase (E

5. RNA Degradation (3)

<ribonuclease>

n0b06a1.r1	401	1.1e-36	35	433	sp|P24657|RNTR_TRIVI RIBONUCLEASE TRV >pir||JX0197 ribonuclease T2 (EC3.1.27.1) - fungus (Trichoderma viride) >
o4d08a1.r1	314	2.9e-26	14	463	gnl|PID|e349364 (Z99259) ribonuclease II RNB family protein[Schizosaccharomyces pombe]

<NONSENSE-MEDIATED MRNA DECAY PROTEIN-DECAY OF MRNAS CONTAINING PREMATURE STOP CODONS>

w4a10a1.r1	471	4.6e-44	61	528	pir||S48909 NMD3 protein - yeast (Saccharomyces cerevisiae) >gi|458900(U00027) Nmd3p: Putative Upf1p =NONSENSE-MEDIATED MRNA DECAY PROTEIN 3

B.Protein Biosynthesis

1. initiation (16)

<EUKARYOTIC TRANSLATION INITIATION>

Contig1022_n1e01a1.r1	851	2.3e-84	45	1043	sp|Q10425|IF3X_S PROBABLE EUKARYOTIC TRANSLATION INITIATION FACTOR 3 BETASUBUNIT (EIF-3 BETA) >gnl|PID|e1168609 (Z7069
Contig1049_m3c05a1.r1	522	1.7e-49	154	573	sp|P56286|IF2A_S EUKARYOTIC TRANSLATION INITIATION FACTOR 2 ALPHA SUBUNIT(EIF-2-ALPHA) >gnl|PID|e1216795 (AL021046) tr
Contig1590_a1e03f2.f1	500	3.7e-47	104	499	sp|P47813|IF1A_H EUKARYOTIC TRANSLATION INITIATION FACTOR 1A (EIF-1A)(EIF-4C) >gi|306725 (L18960) protein synthesis fa
u4a05a1.f1	410	1.3e-37	282	593	sp|P40217|IF34_YEAST EUKARYOTIC TRANSLATION INITIATION FACTOR 3 DELTA SUBUNIT(EIF-3 DELTA) (EIF3 P39) (TRANSLAT
Contig1493_d4b03a1.f1	323	1.7e-27	487	846	sp|Q10425|IF3X_S PROBABLE EUKARYOTIC TRANSLATION INITIATION FACTOR 3 BETASUBUNIT (EIF-3 BETA) >gnl|PID|e1168609 (Z7069
z6f08a1.r1	240	1.3e-19	207	521	sp|P78954|IF4E_SCHPO EUKARYOTIC TRANSLATION INITIATION FACTOR 4E (EIF-4E)(EIF4E) (MRNA CAP-BINDING PROTEIN) (EI

<INITIATION FACTOR>

Contig299_g7c05a1.r1	654	1.7e-63	4	495	sp|P47943|IF4A_S EUKARYOTIC INITIATION FACTOR 4A (EIF-4A) >gnl|PID|e114182(X80796) translation initiation factor eIF-4
Contig1750_h4c11a1.f1	594	3.9e-57	266	742	sp|P23301|IF52_Y INITIATION FACTOR 5A-2 (EIF-5A) (EIF-4D) (HYPUSINECONTAINING PROTEIN HP2) >pir||FIBYA1 translation in
r3d06a1.r1	530	2.5e-50	29	853	gnl|PID|e1291633 (AL023287) hypothetical translation initiation factor[Schizosaccharomyces pombe]
Contig1153_h4c11a1.r1	367	4.6e-33	320	550	sp|P23301|IF52_Y INITIATION FACTOR 5A-2 (EIF-5A) (EIF-4D) (HYPUSINECONTAINING PROTEIN HP2) >pir||FIBYA1 translation in
j5c01a1.f1	313	2.5e-27	56	508	sp|P32502|E2BB_YEAST TRANSLATION INITIATION FACTOR EIF-2B DELTA

					SUBUNIT(EIF-2B GDP-GTP EXCHANGE FACTOR) (GUANIN
d5c04a1.f1	247	2.4e-20	254	565	sp|P32774|TOA2_YEAST TRANSCRIPTION INITIATION FACTOR IIA SMALL CHAIN (TFIIA13.5 KD SUBUNIT) >pir||A41810 transc
p0g09a1.r1	200	7.4e-14	72	470	sp|Q10475|YDF3_SCHPO PROBABLE EUKARYOTIC INITIATION FACTOR C17C9.03>gnl|PID|e241757 (Z73099) probable initiatio
j9a04a1.r1	161	2.3e-10	16	339	sp|P56288|E2BG_SCHPO PROBABLE TRANSLATION INITIATION FACTOR EIF-2B GAMMASUBUNIT (EIF-2B GDP-GTP EXCHANGE FACTOR
*Contig614_c2e07a1.f1	120	1.5e-05	396	590	gnl|PID|e1291633 (AL023287) hypothetical translation initiation factor[Schizosaccharomyces pombe]
l3f02a1.r1	128	7.1e-05	172	420	sp|P32502|E2BB_YEAST TRANSLATION INITIATION FACTOR EIF-2B DELTA SUBUNIT(EIF-2B GDP-GTP EXCHANGE FACTOR) (GUANIN

2. elongation (16)

<ELONGATION FACTOR, eucaryotic and archaeal>

Contig1720_c8h08a1.f1	1458	1.8e-211	94	1032	sp|Q01765|EF1A_P ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) >gnl|PID|e229326(X96614) EF1-alpha translation elongation fact
Contig1745_g7f04a1.r1	825	1.2e-81	146	790	sp|P53978|EF3B_Y ELONGATION FACTOR 3B (EF-3B) >pir||S62926 probablemembrane protein YNL014w - yeast (Saccharomyces cer
o9f08a1.r1	695	3.9e-67	26	580	gnl|PID|e212187 (X92971) translocation elongation factor [Saccharomycescerevisiae]
Contig938_e7d04a1.r1	678	1.9e-65	112	1035	sp|P53978|EF3B_Y ELONGATION FACTOR 3B (EF-3B) >pir||S62926 probablemembrane protein YNL014w - yeast (Saccharomyces cer
o9f08a1.f1	524	7.6e-49	78	557	gnl|PID|e1294522 (AL023704) translocation elongation factor[Schizosaccharomyces pombe]
z5g01a1.f1	504	6.3e-47	289	708	sp|O14460|EF2_SCHPO ELONGATION FACTOR 2 (EF-2) >gnl|PID|d1024469 (D83975)elongation factor 2 [Schizosaccharomy
Contig1388_c3h01a1.f1	483	2.1e-45	167	817	gnl|PID|d1032532 (AB016046) elongation factor 1 beta [Schizosaccharomycespombe]
Contig847_w9b11a1.r1	434	5e-39	8	565	pir||S25363 translation elongation factor eEF-3 - yeast (Candida albicans)>gi|2521 (Z12822) translation elongatio
Contig1190_r7h02a1.r1	418	1.8e-38	269	997	sp|P49373|TFS2_S TRANSCRIPTION ELONGATION FACTOR S-II (TFIIS) >pir||S63845transcription elongation factor TFIIS - fiss
o6a10a1.r1	345	8.6e-30	8	262	sp|P25039|EFG1_YEAST ELONGATION FACTOR G, MITOCHONDRIAL 1 PRECURSOR (MEF-G-1)>pir||S61642 translation elongation
w5g09a1.f1	325	1.1e-27	327	602	gi|2832315 (AF045014) translation release factor eRF3 [Podospora anserina]
*Contig1412_f1d12a1.f1	251	1.9e-19	436	636	gnl|PID|e351430 (Z99568) elongation factor [Schizosaccharomyces pombe]
Contig4_v7g03a1.r1	184	1.2e-12	239	679	sp|P49373|TFS2_S TRANSCRIPTION ELONGATION FACTOR S-II (TFIIS) >pir||S63845transcription elongation factor TFIIS - fiss
j7e11a1.r1	169	6e-09	157	408	sp|P25997|EF3_CANAL ELONGATION FACTOR 3 (EF-3) >gi|2498 (Z11484) elongationfactor 3 [Candida albicans]
<elongation factor 2>					
n3d10a1.r1	591	4.4e-56	21	686	gnl|PID|d1005229 (D21163) similar to human elongation factor 2 mRNA

(HSEF2).[Homo sapiens]

<TRANSLATION FACTOR>

g5f06a1.f1	232	9.9e-19	343	567	sp|P79060|SUI1_SCHPO PROTEIN TRANSLATION FACTOR SUI1

3. termination (1)

<PEPTIDE CHAIN RELEASE FACTOR>

Contig810_r8e11a1.r1	1398	2.6e-142	27	965	gi|2996008 (AF053983) translation release factor subunit 1 [Podosporaanserina]

4. Ribosomal proteins (88)

<a.40S ribosomal protein> (35)

Contig1727_c8e12a1.f1	806	1.3e-79	53	568	sp|P52810|RS9_PO 40S RIBOSOMAL PROTEIN S9 (S7) >gnl|PID|e242707 (X96613)cytoplasmic ribosomal protein S7 [Podospora an
Contig1574_c9a10a1.f1	788	1e-77	349	1023	sp|P05752|RS6_SC 40S RIBOSOMAL PROTEIN S6 >pir||R3ZP6E ribosomal proteinS6.e, cytosolic - fission yeast (Schizosacchar
Contig1529_a1c06c9.r1	787	1.3e-77	68	694	sp|P26783|RS5_YE 40S RIBOSOMAL PROTEIN S5 (RP14) (YS8) >pir||S55720ribosomal protein S5.e - yeast (Saccharomyces cerev
Contig59_m1g12a1.r1	770	8.5e-76	29	634	sp|P40910|RS3A_C 40S RIBOSOMAL PROTEIN RP10 >pir||S49366 ribosomal proteinS0.e.B, cytosolic - yeast (Candida albicans)
Contig1219_d1c01a1.f1	742	7.5e-73	125	721	sp|P05754|RS8_YE 40S RIBOSOMAL PROTEIN S8 (S14) (YS9) (RP19) >pir||S45591ribosomal protein S8.e, cytosolic - yeast (Sa
Contig1770_d1h11a1.r1	717	3.4e-70	368	787	sp|P27073|RS19_E 40S RIBOSOMAL PROTEIN S19 (S16) >pir||JQ1349 ribosomalprotein S19.e, cytosolic - Emericella nidulans
Contig1573_d1c08a1.f1	656	9.9e-64	283	762	sp|P26781|RS41_Y 40S RIBOSOMAL PROTEIN RP41 (YS12) (S18A / S18B)>pir||S41784 ribosomal protein S11.e, cytosolic - yeas
Contig1297_g7h08a1.f1	653	2.1e-63	299	757	sp|P34737|RS15_P 40S RIBOSOMAL PROTEIN S15 (S12) >pir||A53793 ribosomalprotein S12, cytosolic - Podospora anserina >gi
Contig485_d1h08a1.r1	627	1.1e-60	223	840	sp|P25443|RS4_YE 40S RIBOSOMAL PROTEIN S4 (OMNIPOTENT SUPRESSOR PROTEINSUP44) (RP12) (S2E) >pir||R3BYS2 ribosomal prot
Contig1106_a1c02c9.r1	610	7.9e-59	315	737	sp|P34737|RS15_P 40S RIBOSOMAL PROTEIN S15 (S12) >pir||A53793 ribosomalprotein S12, cytosolic - Podospora anserina >gi
Contig1417_r2f10a1.f1	600	7.4e-58	161	571	sp|P27770|RS17_N 40S RIBOSOMAL PROTEIN S17 (CRP3) >pir||S34441 ribosomalprotein L17.e - Neurospora crassa >gi|168796 (
Contig1491_d5g06a1.f1	596	2.3e-57	58	486	sp|P40213|R61A_Y 40S RIBOSOMAL PROTEIN RS16 HOMOLOG (RP61R HOMOLOG)>pir||S67619 ribosomal protein S16.e.B - yeast (Sac
Contig1399_d2h09a1.f1	579	1.4e-55	212	580	gi|3114615 (AF052483) 40S ribosomal protein S12 [Erysiphe graminis f. sp.hordei]
Contig1639_h1e05a1.r1	547	3.7e-52	228	626	sp|P35271|RS18_Y 40S RIBOSOMAL PROTEIN S18E >pir||S50886 ribosomal proteinS18.e, cytosolic - yeast (Saccharomyces cere
Contig1524_u4e07a1.r1	508	4.8e-48	53	466	sp|P19115|RS14_N 40S RIBOSOMAL PROTEIN S14 (CRP2) >gi|2995 (X53734)ribosomal protein crp-2 [Neurospora crassa]
*Contig1089_f0f06a1.r1	448	1.1e-41	3	458	sp|Q10101|RS7_SC PROBABLE 40S RIBOSOMAL PROTEIN S7
Contig1779_c6f04a1.f1	395	4.8e-36	94	399	sp|P26782|RS24_Y 40S RIBOSOMAL PROTEIN S24E (RP50) >pir||S48410

				ribosomalprotein S24.e - yeast (Saccharomyces cerevisi
Contig1593_g6c01a1.f1	392	9.6e-36	47 331	sp|Q08745|YO93_Y PUTATIVE 40S RIBOSOMAL PROTEIN IN SNF2-CPA1 INTERGENICREGION >pir||S67197 ribosomal protein S10.e.A -
Contig993_m8f08a1.r1	384	6.7e-35	106 351	gnl|PID|e1313483 (AL031154) 40s ribosomal protein s27 type[Schizosaccharomyces pombe]
Contig1570_c9b09a1.f1	371	1.5e-33	134 457	sp|P23403|RS20_X 40S RIBOSOMAL PROTEIN S20 (S22) >pir||A37974 ribosomalprotein S20 - African clawed frog >gi|214758 (M
Contig311_n0e01a1.f1	365	7.2e-33	164 436	sp|P05753|RS4E_Y 40S RIBOSOMAL PROTEIN S4 (S7) (YS6) (RP5) >pir||S20054ribosomal protein S4.e, cytosolic - yeast (Sacc
Contig1571_d5d04a1.f1	238	2e-19	54 233	sp|Q12087|RS30_Y 40S RIBOSOMAL PROTEIN S30 >pir||S67074 ribosomal proteinS30.e, cytosolic - yeast (Saccharomyces cerev
Contig980_n0f03a1.f1	226	3.9e-18	252 413	sp|Q10421|RS28_S PROBABLE 40S RIBOSOMAL PROTEIN S28 (S33)>gnl|PID|e1168607 (Z70691) ribosomal protein S28[Schizosaccha
Contig873_r4c07a1.f1	210	1.8e-16	192 368	sp|P53733|YN8L_Y PUTATIVE 40S RIBOSOMAL PROTEIN YNR037C >pir||S63368probable ribosomal protein S19, mitochondrial - ye
Contig1691_c3g10a1.f1	200	6.6e-15	76 291	sp|P79009|RS25_S 40S RIBOSOMAL PROTEIN S25 (S31) >gnl|PID|d1019840(AB000398) ribosomal protein S31 [Schizosaccharomyce
y6c06a1.r1	747	2.2e-73	88 615	sp|Q01291|RSP4_NEUCR 40S RIBOSOMAL PROTEIN SA HOMOLOG (RIBOSOME-ASSOCIATEDPROTEIN 1) >gi|1039443 (U36470) putat
w8d11a1.r1	621	5.6e-60	61 492	gnl|PID|e1293406 (AL023594) 40s ribosomal protein s4 type [Schizosaccharomycespombe]
a1d06f2.f1	501	2.8e-47	163 549	sp|P04648|RS22_YEAST 40S RIBOSOMAL PROTEIN S22 (YS24) (YP58) >pir||R4BY24ribosomal protein S15a.e.c10 - yeast (
w4e01a1.r1	433	4.6e-40	58 528	sp|P32899|YHU8_YEAST PUTATIVE 40S RIBOSOMAL PROTEIN YHR148W >pir||S33911hypothetical protein YHR148w - yeast (S
Contig182_i7h11a1.f1	245	3.8e-20	141 494	gnl|PID|e1287762 (AL022598) 40s mitochondrial ribosomal protein[Schizosaccharomyces pombe]
y6c06a1.f1	218	2.8e-17	159 512	sp|Q01291|RSP4_NEUCR 40S RIBOSOMAL PROTEIN SA HOMOLOG (RIBOSOME-ASSOCIATEDPROTEIN 1) >gi|1039443 (U36470) putat
o6g03a1.f1	191	1.4e-13	186 386	sp|P27929|NAM9_YEAST NAM9 PROTEIN PRECURSOR >pir||S55146 ribosomal proteinS4.e precursor - yeast (Saccharomyces
o5e07a1.r1	186	2.9e-13	69 290	sp|P32902|RT04_YEAST MITOCHONDRIAL 40S RIBOSOMAL PROTEIN MRP4 >pir||A42115ribosomal protein S2, mitochondrial -
w4e03a1.f1	184	2e-11	164 352	sp|Q10234|RT05_SCHPO PROBABLE MITOCHONDRIAL 40S RIBOSOMAL PROTEIN S5>gnl|PID|e223738 (Z69727) probable ribosoma
i3f02a1.r1	126	8.6e-07	7 93	sp|P02382|RMS5_EMENI MITOCHONDRIAL RIBOSOMAL PROTEIN S5 >pir||QXASRI 21S rRNAintron protein - Emericella nidulans
<b.60S ribosomal protein>	**(50)**			
Contig1635_a1e02f2.f1	959	8e-96	235 843	sp|O13418|RL15_A 60S RIBOSOMAL PROTEIN L15 >gnl|PID|e1181730 (Y15321)putative ribosomal protein L15 [Aspergillus niger
Contig1080_i7d05a1.r1	800	5e-79	45 860	sp|P35679|RL2_SC 60S RIBOSOMAL PROTEIN L2 >gi|312177 (X73146) ribosomalprotein L2 [Schizosaccharomyces pombe]

Contig1656_c1c10a1.f1	790	5.9e-78	62 712	gnl|PID|e330379 (Z97992) 60s ribosomal protein L10 [Schizosaccharomyces pombe]
Contig1757_d5g11a1.f1	747	2.5e-73	111 710	sp|P26784|R13A_Y 60S RIBOSOMAL PROTEIN L13A (RP22) >pir||S48401 ribosomalprotein L16.e.A, cytosolic - yeast (Saccharom
Contig1652_o6h02a1.f1	722	9.4e-71	44 793	gnl|PID|e1285382 (AL022304) 60s ribosomal protein [Schizosaccharomyces pombe]
Contig1436_r6h09a1.r1	628	8.9e-61	249 767	sp|P47913|RL1X_Y 60S RIBOSOMAL PROTEIN L18A >pir||S59848 ribosomal proteinL18a.e.c15 - yeast (Saccharomyces cerevisiae
Contig1192_g4c02a1.f1	625	2.2e-60	179 868	sp|P31334|RM09_Y MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L9 PRECURSOR (YML9)>pir||R5BYL3 ribosomal protein L3 precursor, m
Contig1536_c8f11a1.f1	619	8.6e-60	503 994	sp|P35979|RL12_M 60S RIBOSOMAL PROTEIN L12 >pir||JN0778 ribosomal proteinL12 - mouse >gi|398048 (L04280) ribosomal pro
Contig1395_a1b03c9.r1	598	1.3e-57	202 699	sp|P26321|RL1_YE 60S RIBOSOMAL PROTEIN L1 (L5) (YL3) (RIBOSOMAL 5 SRNA-BINDING PROTEIN) >gi|173232 (M65056) 5S ribosom
Contig1091_b0f01a1.f1	561	1.2e-53	204 608	sp|P04451|RL1A_Y 60S RIBOSOMAL PROTEIN L17 >pir||R5BY17 ribosomal proteinL23.e, cytosolic - yeast (Saccharomyces cerev
Contig955_b0a07a1.r1	406	6.9e-52	328 645	sp|O13418|RL15_A 60S RIBOSOMAL PROTEIN L15 >gnl|PID|e1181730 (Y15321)putative ribosomal protein L15 [Aspergillus niger
Contig1603_c4f01a1.f1	533	1.1e-50	31 504	sp|Q02753|R21A_Y 60S RIBOSOMAL PROTEIN L21E A >pir||S28921 ribosomalprotein L21.e.A, cytosolic - yeast (Saccharomyces
Contig1759_j7c10a1.f1	524	9.3e-50	48 365	sp|P52809|RL44_P 60S RIBOSOMAL PROTEIN L44 (L41) >gnl|PID|d1011717(D67040) ribosomal protein L41 [Candida utilis]
Contig1608_c8d07a1.f1	513	1.3e-48	362 925	sp|P05735|RL19_Y 60S RIBOSOMAL PROTEIN L19 (L23) (YL14) (RP33) (RP15L)>pir||S44597 ribosomal protein L19.e, cytosolic
Contig1344_c1c05a1.f1	508	4.8e-48	198 644	sp|P17076|RL4A_Y 60S RIBOSOMAL PROTEIN L7A-2 (L4-2) (YL5) (RP6)>pir||R5BY7A ribosomal protein L7a.e.A - yeast (Sacchar
Contig575_c5b05a1.r1	441	5.8e-41	62 430	gnl|PID|e349694 (Z99296) 60s ribosomal protein L32 [Schizosaccharomyces pombe]
Contig816_y6a08a1.f1	422	5.6e-39	90 422	sp|P40525|YIF2_Y PROBABLE 60S RIBOSOMAL PROTEIN YIL052C >pir||S48427ribosomal protein L34.e.B, cytosolic - yeast (Sacc
Contig1499_n8g05a1.r1	420	1.1e-38	103 411	sp|P41056|R372_Y 60S RIBOSOMAL PROTEIN L37B (YL37) (RP47) >pir||S44069ribosomal protein L35a.e.c15 - yeast (Saccharomy
Contig1449_e4g11a1.r1	387	3e-35	63 395	sp|P04649|RL34_Y 60S RIBOSOMAL PROTEIN L34 (YL28) >pir||R5BY1E ribosomalprotein L31.e.A, cytosolic - yeast (Saccharomy
Contig1705_h1g02a1.f1	387	3.4e-35	129 473	gi|292435 (L07287) ribosomal protein L26 [Homo sapiens]
Contig1803_c8a04a1.f1	387	3.4e-35	71 376	sp|P52808|RL30_S 60S RIBOSOMAL PROTEIN L30 (L32) >gi|1621046 (U52080)ribosomal protein Rpl32p [Schizosaccharomyces pom
Contig1696_l3e05a1.f1	377	4e-34	50 301	gi|2665824 (AF035770) ribosomal protein L37 [Schistosoma mansoni]
Contig682_y8h12a1.f1	376	4.4e-34	80 352	sp|P49631|R37A_Y PROBABLE 60S RIBOSOMAL PROTEIN L37A >pir||S54068ribosomal protein L37a.e - yeast (Saccharomyces cerev
Contig1513_d4c07a1.f1	373	9.8e-34	540 923	sp|Q12690|R13D_Y PROBABLE 60S RIBOSOMAL PROTEIN L13E B >pir||S67618ribosomal protein L13.e.A, cytosolic - yeast (Sacch
Contig1125_r5f10a1.f1	368	3.2e-33	334 609	sp|P38665|RL24_K 60S RIBOSOMAL PROTEIN L24 (L30) >gi|173317

					(L05777)ribosomal protein L30 [Kluyveromyces lactis]
Contig1019_c1h09a1.f1	353	1.2e-31	150	410	sp|P06380|RL16_Y 60S RIBOSOMAL PROTEIN L16 (YL16) (39A) (RP39)>pir||S59767 ribosomal protein L11.e.A, cytosolic - yeas
Contig38_z3e04a1.r1	350	2.6e-31	180	512	sp|Q02753|R21A_Y 60S RIBOSOMAL PROTEIN L21E A >pir||S28921 ribosomalprotein L21.e.A, cytosolic - yeast (Saccharomyces
Contig1168_u4f09a1.r1	343	1.5e-30	157	477	sp|Q02753|R21A_Y 60S RIBOSOMAL PROTEIN L21E A >pir||S28921 ribosomalprotein L21.e.A, cytosolic - yeast (Saccharomyces
Contig1799_d3h04a1.f1	295	1.9e-25	62	451	sp|O14069|YEA4_S PROBABLE 60S RIBOSOMAL PROTEIN C2E11.04 >gnl|PID|e339274(Z98850) ribosomal protein [Schizosaccharomyc
Contig267_h0a09a1.f1	269	1.1e-22	90	470	sp|P35996|RM38_Y MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L38 (YML38)>pir||S38000 ribosomal protein L14, mitochondrial - ye
Contig653_b0g02a1.f1	266	2.2e-22	261	470	sp|P08978|RL2A_N 60S RIBOSOMAL PROTEIN L27A (L29) (CRP1) >pir||R6NC7Aribosomal protein L27a.e - Neurospora crassa >gi|
Contig754_t2e02a1.f1	248	1.6e-20	122	412	sp|P05745|RL39_Y 60S RIBOSOMAL PROTEIN YL39 >pir||S50922 ribosomal proteinL36.e.A, cytosolic - yeast (Saccharomyces ce
Contig645_c0a03a1.r1	234	5.1e-19	287	478	sp|P24049|RL17_R 60S RIBOSOMAL PROTEIN L17 (L23) (AMINO ACIDSTARVATION-INDUCED PROTEIN) (ASI) >pir||R5RT17 ribosomal p
Contig1622_u4f10a1.r1	193	1.2e-14	72	398	gi|3098460 (AF040713) 60S ribosomal protein P2 [Cryptochiton stelleri]
Contig1441_e9c05a1.f1	169	4.2e-12	8	181	sp|P05740|RL7A_Y 60S RIBOSOMAL PROTEIN YL17-A >pir||S38012 ribosomalprotein L17.e.A, cytosolic - yeast (Saccharomyces
Contig1284_j4h04a1.r1	142	6.3e-09	344	493	sp|P36525|RM24_Y MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L24 PRECURSOR (YML24)>pir||S50921 ribosomal protein YmL24, mitoch
Contig1552_c5c05a1.f1	140	2.6e-08	33	188	sp|P05739|R16B_Y 60S RIBOSOMAL PROTEIN YL16B >pir||S55970 ribosomalprotein L6.e.B, cytosolic - yeast (Saccharomyces ce
Contig551_c6d06a1.r1	153	4.5e-08	361	573	sp|O14337|RM07_S PROBABLE MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L7 PRECURSOR>gnl|PID|e325415 (Z97211) probable mitochond
*Contig1149_w8a04a1.r1	120	0.00016	324	488	sp|P36517|RM04_Y MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L4 PRECURSOR (YML4)>pir||S59407 ribosomal protein YmL4 precursor,
o4g08a1.r1	443	3.8e-41	135	500	sp|P51997|RL2B_PUCGR 60S RIBOSOMAL PROTEIN L23A (L25) >gi|1707876 (U44800)ribosomal protein L23a [Puccinia gram
v1a02a1.r1	433	4.7e-40	84	524	sp|P53875|YNS5_YEAST PUTATIVE 60S MITOCHONDRIAL RIBOSOMAL PROTEIN YNL185C>pir||S63140 probable ribosomal protei
x1d02a1.r1	364	8.3e-33	6	296	sp|P27659|RL3_MOUSE 60S RIBOSOMAL PROTEIN L3 (J1 PROTEIN) >gi|52741 (Y00225)J1 protein [Mus musculus] >prf||16
c8b05a1.r1	351	2.2e-31	104	454	gnl|PID|e1285350 (AL022299)60S ribosomal protein [Schizosaccharomyces pombe]>gnl|PID|e1291883 (AJ001133) ribos
e4b10a1.r1	315	1.4e-27	30	335	sp|Q09668|RL22_SCHPO 60S RIBOSOMAL PROTEIN L22 >gnl|PID|e334258 (Z98595) 60sribosomal protein l22 [Schizosaccha
Contig1679_d5d09a1.f1	174	1.2e-12	375	632	sp|P50344|RLA1_C 60S ACIDIC RIBOSOMAL PROTEIN P1 (ALLERGEN CLA H 12) (CLAH XII) >gi|1143425 (X85180) ribosomal protein
z3g02a1.f1	151	3.6e-10	201	437	sp|P22354|RM20_YEAST MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L20 PRECURSOR (YML20)>pir||S38163 ribosomal protein Ym

z5g03a1.f1 148 6.8e-10 102 290 sp|P41805|RL10_YEAST 60S RIBOSOMAL PROTEIN L10 (L9) (UBIQUINOL-CYTOCHROME CREDUCTASE COMPLEX SUBUNIT VI REQUIRI

c9e07a1.f1 162 3e-09 108 395 sp|P05317|RLA0_YEAST 60S ACIDIC RIBOSOMAL PROTEIN P0 (L10E) >gi|4371 (X06959)ribosomal protein A0 (AA 1-312) [S

Contig1435_j4b03a1.f1 145 1.7e-09 310 423 sp|P73300|RL36_S 50S RIBOSOMAL PROTEIN L36 >gnl|PID|d1018061 (D90905) 50Sribosomal protein L36 [Synechocystis sp.]

n8f06a1.f1 142 3.1e-09 355 447 sp|P05747|RL43_YEAST 60S RIBOSOMAL PROTEIN YL43 >pir||S71066 ribosomal proteinL29.e, cytosolic - yeast (Sacchar

<ribosomal protein>

c9e07a1.r1 459 8.4e-43 79 588 pir||R5BY0E acidic ribosomal protein P0.e, cytosolic - yeast (Saccharomycescerevisiae) >gi|171806 (M37

o6f09a1.f1 446 1.5e-41 112 489 gi|3003044 (AF054907) putative 5S rRNA binding ribosomal protein [Neurosporacrassa]

Contig1007_f0f06a1.f1 330 3.8e-29 257 523 gi|2737908 (U73847) ribosomal protein [Neurospora crassa]

5. Post-translational modifications (14)

a.methylation

<SERINE HYDROXYMETHYLTRANSFERASE>

g4f08a1.r1 900 8.2e-99 36 692 sp|O13426|GLYC_CANAL SERINE HYDROXYMETHYLTRANSFERASE, CYTOSOLIC (SERINEMETHYLASE) (GLYCINE HYDROXYMETHYLTRANSFE

Contig1363_f1e12a1.f1 663 3.1e-94 561 1067 sp|O13426|GLYC_C SERINE HYDROXYMETHYLTRANSFERASE, CYTOSOLIC (SERINEMETHYLASE) (GLYCINE HYDROXYMETHYLTRANSFERASE) (SHMT

g4f08a1.f1 144 1.7e-06 389 592 sp|P34898|GLYC_NEUCR SERINE HYDROXYMETHYLTRANSFERASE, CYTOSOLIC (SERINEMETHYLASE) (GLYCINE HYDROXYMETHYLTRANSFE=N. crassa

b.glycosylation

<glycosylation>

c7g06a1.f1 492 2.6e-46 195 668 sp|P32621|GDA1_YEAST GUANOSINE-DIPHOSPHATASE (GDPASE) >pir||A40732guanosine-diphosphatase (EC 3.6.1.42) - yeast

<GPI-ANCHOR TRANSMIDASE>

Contig26_d1f08a1.r1 216 1.7e-16 166 321 sp|P49018|GPI8_Y GPI-ANCHOR TRANSMIDASE >pir||S59796 probable membraneprotein YDR331w - yeast (Saccharomyces cerevisia

<MNN9 PROTEIN>

j4f07a1.f1 464 2.4e-43 182 595 sp|P53697|MNN9_CANAL MNN9 PROTEIN >gi|1488302 (U63642) Mnn9p [Candidaalbicans]

<UDP-GLUCOSE:GLYCOPROTEIN GLUCOSYLTRANSFERASE>

w5a11a1.r1 353 3.6e-30 192 467 gi|860712 (U28735) coded for by C. elegans cDNA cm06e4; coded for by C.elegans cDNA CEESP39F; coded

c.myristoylization

<PEPTIDE N-MYRISTOYLTRANSFERASE>

n0g11a1.r1 495 1.2e-46 8 451 sp|P34763|NMT_AJECA GLYCYLPEPTIDE N-TETRADECANOYLTRANSFERASE (PEPTIDEN-MYRISTOYLTRANSFERASE) (MYRISTOYL-COA:PR Ajellomyces capsulatus

d.other

<protein disulfide-isomerase>

Contig1743_e4a07a1.r1 1365 8.1e-139 260 1276 sp|Q12730|PDI_AS PROTEIN DISULFIDE ISOMERASE PRECURSOR (PDI) >pir||S57942protein disulfide-isomerase (EC 5.3.4.1) - As

Contig89_m5b09a1.r1 320 9.9e-28 172 453 sp|Q00248|PDI_AS PROTEIN DISULFIDE ISOMERASE PRECURSOR (PDI)>gnl|PID|d1013598 (D85900) protein disulfide isomerase [As

<cyclophilin>

Contig1618_g6g11a1.f1 636 1.1e-61 90 563 sp|P18253|CYPH_S PEPTIDYL-PROLYL CIS-TRANS ISOMERASE (PPIASE) (ROTAMASE)(CYCLOPHILIN) (CYCLOSPORIN A-BINDING PROTEIN)

Contig1641_e9e07a1.f1 416 2.9e-38 355 744 gi|3288923 (AF071225) cyclophilin B [Rattus norvegicus]

k0f12a1.f1 379 2.6e-34 177 488 sp|P38911|FKB3_YEAST FK506-BINDING NUCLEAR PROTEIN (PEPTIDYL-PROLYL CIS-TRANSISOMERASE) (PPIASE) (PROLINE ROTAMase)(cyclophilin)

Contig111_k5c02a1.f1 251 8.3e-21 325 990 pir||S62327 cyclophilin-like protein wis2 - fission yeast (Schizosaccharomycespombe) >gnl|PID|e205292 (X91981) wi

6. Folding and Targeting (68)

a. folding

<CALNEXIN HOMOLOG-folding of glycoproteins>

o4g04a1.r1 510 3e-48 18 515 sp|Q39817|CALX_SOYBN CALNEXIN HOMOLOG PRECURSOR >gi|669003 (U20502) calnexin[Glycine max]

Contig835_r2e08a1.r1 430 9.5e-40 213 734 sp|P36581|CALX_S CALNEXIN HOMOLOG PRECURSOR >pir||A56106 calnexin homologcnx1 - fission yeast (Schizosaccharomyces pom

<PEPTIDYL-PROLYL CIS-TRANS ISOMERASE (catalyzes folding)>

<FK506-BINDING PROTEIN-protein folding inhibitor>

Contig1483_c4f02a1.f1 394 6.3e-36 172 492 sp|P48375|FKB2_D 12 KD FK506-BINDING PROTEIN (FKBP) (PEPTIDYL-PROLYLCIS-TRANS ISOMERASE) (PPIASE) (MACROLIDE BINDING P

b.chaperones

<chaperone>

Contig547_c6f12a1.r1 922 6.8e-92 7 741 gnl|PID|e290095 (Y08867) putative ER chaperone [Aspergillus awamorii]>gnl|PID|e290123 (Y08868) putative ER chaperone

o0h11a1.r1 693 1.3e-67 10 441 gnl|PID|e290095 (Y08867) putative ER chaperone [Aspergillus awamorii]>gnl|PID|e290123 (Y08868) putative ER

Contig548_c6f12a1.f1 408 6e-48 148 417 gnl|PID|e290095 (Y08867) putative ER chaperone [Aspergillus awamorii]>gnl|PID|e290123 (Y08868) putative ER chaperone

Contig991_c4a06a1.f1 370 2.1e-33 103 321 gi|2367623 (AF016187) chaperone/heat shock protein [Emericella nidulans]

Contig312_g6c05a1.f1 336 8.9e-30 204 404 gi|2367623 (AF016187) chaperone/heat shock protein [Emericella nidulans]

<prefoldin-chaperone which delivers unfolded proteins to another chaperonin>

Contig522_c8d04a1.f1	219	2.1e-17	156	467	gnl|PID|e1297429	(Y17393) prefoldin subunit 2 [Mus musculus]

<heat shock protein Hsp88>

Contig603_c3h04a1.r1	895	5.2e-89	255	1115	gi|3242972	(AF069523) heat shock protein Hsp88 [Neurospora crassa]
Contig540_c7e04a1.f1	142	4.8e-08	408	524	gi|3242972	(AF069523) heat shock protein Hsp88 [Neurospora crassa]

<heat-shock protein30>

Contig1840_c1c01a1.f1	933	4.7e-93	84	626	sp|P40920|HS30_E	30 KD HEAT SHOCK PROTEIN >pir||S50131 heat-shock protein30 - Emericella nidulans >gnl|PID|d1007414 (D
Contig1864_a0e10a1.f1	933	4.8e-93	323	865	sp|P40920|HS30_E	30 KD HEAT SHOCK PROTEIN >pir||S50131 heat-shock protein30 - Emericella nidulans >gnl|PID|d1007414 (D
Contig1800_c6d03a1.f1	855	8.9e-85	451	990	sp|P40920|HS30_E	30 KD HEAT SHOCK PROTEIN >pir||S50131 heat-shock protein30 - Emericella nidulans >gnl|PID|d1007414 (D
Contig1820_a5c12a1.f1	832	2.4e-82	101	640	sp|P40920|HS30_E	30 KD HEAT SHOCK PROTEIN >pir||S50131 heat-shock protein30 - Emericella nidulans >gnl|PID|d1007414 (D
Contig1866_c1c02a1.f1	816	1.2e-80	342	881	sp|P40920|HS30_E	30 KD HEAT SHOCK PROTEIN >pir||S50131 heat-shock protein30 - Emericella nidulans >gnl|PID|d1007414 (D
Contig1772_a0a02a1.r1	798	9.6e-79	88	615	sp|P40920|HS30_E	30 KD HEAT SHOCK PROTEIN >pir||S50131 heat-shock protein30 - Emericella nidulans >gnl|PID|d1007414 (D
Contig1612_g7f07a1.r1	733	7.5e-72	95	583	sp|P40920|HS30_E	30 KD HEAT SHOCK PROTEIN >pir||S50131 heat-shock protein30 - Emericella nidulans >gnl|PID|d1007414 (D
Contig1812_g4b07a1.f1	612	4.8e-59	249	800	sp|P40920|HS30_E	30 KD HEAT SHOCK PROTEIN >pir||S50131 heat-shock protein30 - Emericella nidulans >gnl|PID|d1007414 (D
Contig1862_c1b11a1.f1	342	2.1e-30	261	758	sp|P19752|HS30_N	30 KD HEAT SHOCK PROTEIN >pir||A38360 heat shock protein30 - Neurospora crassa >gi|168820 (M55672) he
s8b01a1.f1	143	1.1e-22	119	280	sp|P40920|HS30_EMENI	30 KD HEAT SHOCK PROTEIN >pir||S50131 heat-shock protein30 - Emericella nidulans >gnl|PID|
Contig1576_f5d04a1.f1	189	3.4e-14	311	520	sp|P40920|HS30_E	30 KD HEAT SHOCK PROTEIN >pir||S50131 heat-shock protein30 - Emericella nidulans >gnl|PID|d1007414 (D
Contig917_c3g11a1.f1	181	2.4e-13	312	521	sp|P40920|HS30_E	30 KD HEAT SHOCK PROTEIN >pir||S50131 heat-shock protein30 - Emericella nidulans >gnl|PID|d1007414 (D

<heat shock protein 70>

Contig1610_a1d03f2.f1	1682	1.9e-172	19	1257	gi|1498496	(U64207) heat shock protein 70 [Penicillium citrinum]
Contig1144_x7a06a1.f1	882	1.2e-87	176	712	gnl|PID|e267541	(X98931) heat shock protein 70 [Emericella nidulans]
d2d10a1.f1	634	7.6e-83	155	703	gnl|PID|e267541	(X98931) heat shock protein 70 [Emericella nidulans]
r5g03a1.r1	667	6.4e-65	84	671	sp|P22774|HS7M_SCHPO	MITOCHONDRIAL HEAT SHOCK 70 KD PROTEIN PRECURSOR>pir||S18670 heat shock protein SSP1 precu
*Contig1366_c3d01a1.r1	356	3.9e-31	229	513	sp|P22774|HS7M_S	MITOCHONDRIAL HEAT SHOCK 70 KD PROTEIN PRECURSOR>pir||S18670 heat shock protein SSP1 precursor - fiss
Contig1567_c3d01a1.f1	127	9.5e-05	449	664	sp|P22774|HS7M_S	MITOCHONDRIAL HEAT SHOCK 70 KD PROTEIN PRECURSOR>pir||S18670 heat shock protein SSP1 precursor - fiss

<HEAT SHOCK PROTEIN HSP1>

Contig113_k5b07a1.f1	275	1.3e-49	205	465	sp|P40292|HS82_A HEAT SHOCK PROTEIN HSP1 (65 KD IGE-BINDING PROTEIN)>gi|1930153 (U92465) heat shock protein [Aspergill
k5b07a1.r1	371	1.6e-33	1	306	sp|P40292|HS82_ASPFU HEAT SHOCK PROTEIN HSP1 (65 KD IGE-BINDING PROTEIN)>gi|1930153 (U92465) heat shock protein
Contig677_a0d04a1.f1	299	6.4e-26	22	195	sp|P40292|HS82_A HEAT SHOCK PROTEIN HSP1 (65 KD IGE-BINDING PROTEIN)>gi|1930153 (U92465) heat shock protein [Aspergill
<HEAT SHOCK PROTEIN 104>					
Contig1831_c3b01a1.r1	1354	1.2e-137	3	1691	sp|P31539|H104_Y HEAT SHOCK PROTEIN 104 >pir||S61476 heat shock protein104 - yeast (Saccharomyces cerevisiae) >gi|5578
*Contig981_e9e06a1.r1	386	5.2e-34	240	668	sp|P31539|H104_Y HEAT SHOCK PROTEIN 104 >pir||S61476 heat shock protein104 - yeast (Saccharomyces cerevisiae) >gi|5578
Contig1489_d4a01a1.r1	231	3.6e-16	74	631	sp|P31539|H104_Y HEAT SHOCK PROTEIN 104 >pir||S61476 heat shock protein104 - yeast (Saccharomyces cerevisiae) >gi|5578
<T-COMPLEX PROTEIN-chaperone of actin, tubulin>					
m8h03a1.r1	750	1.2e-73	63	647	sp|P39078|TCPD_YEAST T-COMPLEX PROTEIN 1, DELTA SUBUNIT (TCP-1-DELTA)(CCT-DELTA) >pir||S67690 chaperonin ANC2,
c9e10a1.r1	622	4.5e-60	24	482	sp|P87153|TCPH_SCHPO PROBABLE T-COMPLEX PROTEIN 1, ETA SUBUNIT (TCP-1-ETA)(CCT-ETA) >gnl|PID|e315886 (Z95397) C
k5e01a1.r1	584	4.8e-56	3	521	sp|P12612|TCPA_YEAST T-COMPLEX PROTEIN 1, ALPHA SUBUNIT (TCP-1-ALPHA)(CCT-ALPHA) >pir||A39793 TCP1 protein - ye
Contig83_l5b12a1.f1	470	5.5e-44	204	719	sp|P87153|TCPH_S PROBABLE T-COMPLEX PROTEIN 1, ETA SUBUNIT (TCP-1-ETA)(CCT-ETA) >gnl|PID|e315886 (Z95397) Cct7p [Schiz
k5e01a1.f1	230	6.1e-32	395	607	sp|P28769|TCPA_ARATH T-COMPLEX PROTEIN 1, ALPHA SUBUNIT (TCP-1-ALPHA)(CCT-ALPHA) >pir||JN0448 t-complex polypep, chaperonin for acitn, tubulin
x9d05a1.r1	348	1.3e-30	52	348	sp|P87153|TCPH_SCHPO PROBABLE T-COMPLEX PROTEIN 1, ETA SUBUNIT (TCP-1-ETA)(CCT-ETA) >gnl|PID|e315886 (Z95397) C
Contig697_p0a05a1.f1	132	4.2e-07	150	236	gnl|PID|e1314064 (AL031174) t-complex protein 1 gamma subunit homolog[Schizosaccharomyces pombe]
<complex I intermediate associated protein-chaperone in assembly of NADH: Ubiquinone Oxidoreductase>					
w4h08a1.r1	162	1.4e-23	88	354	gnl|PID|e1198586 (AJ001726) complex I intermediate associated protein CIA35[Neurospora crassa]
w4h08a1.f1	217	3.5e-17	167	430	gnl|PID|e1198586 (AJ001726) complex I intermediate associated protein CIA35[Neurospora crassa]
<DNAJ-LIKE PROTEIN>					
Contig844_x8c07a1.r1	130	2.4e-07	7	141	sp|P50027|DNJH_S DNAJ-LIKE PROTEIN SLR0093 >gnl|PID|d1011217 (D64004) DnaJ[Synechocystis sp.]
c.protein sorting					
<protein sorting>					
d3c09a1.r1	344	1e-28	94	873	sp|Q07878|VP13_YEAST VACUOLAR PROTEIN SORTING-ASSOCIATED PROTEIN VPS13>pir||S64791 probable membrane protein YL
<CARBOXYPEPTIDASE Y-sorting of vacuolar protein>					

```
Contig1585_c3d11a1.r1    1431 8.1e-146     9 1142  sp|P30574|CBPY_C     CARBOXYPEPTIDASE Y PRECURSOR (CARBOXYPEPTIDASE YSCY)
<MVP1 PROTEIN-vacuolar protein sorting>
Contig275_g9f06a1.r1      393  7.9e-36    29 700   sp|P40959|MVP1_Y     MVP1 PROTEIN >pir||S53033 MVP1 protein -
                                                   yeast(Saccharomyces cerevisiae) >gi|728651 (Z48613) unknown[
Contig1203_g9f06a1.f1     225  3.3e-17   241 633   sp|P40959|MVP1_Y     MVP1 PROTEIN >pir||S53033 MVP1 protein -
                                                   yeast(Saccharomyces cerevisiae) >gi|728651 (Z48613) unknown[
<vacuolar protein sorting homolog h-vps45>
f5f09a1.r1                463  3.2e-43    31 699   gi|1477466           (U35246) vacuolar protein sorting homolog h-vps45
                                                    [Homo sapiens]
<clathrin>
g3d10a1.r1                650    1e-61    28 768   sp|P22137|CLH_YEAST   CLATHRIN HEAVY CHAIN >pir||A36349 clathrin heavy chain
                                                    1 -yeast (Saccharomyces cerevisiae)
Contig1038_o5c12a1.r1     219  2.1e-17   478 774   gnl|PID|d1026032      (AB011822) clathrin light chain [Schizosaccharomyces
                                                    pombe]
Contig504_c9g08a1.f1      210  1.1e-15   140 277   sp|Q00776|AP54_Y      CLATHRIN COAT ASSEMBLY PROTEIN AP54 (CLATHRIN
                                                   COATASSOCIATED PROTEIN AP54) (GOLGI ADAPTOR AP-1 54 KD
Contig826_w4e09a1.f1      147  3.9e-08   134 343   sp|Q10161|CLH_SC      PROBABLE CLATHRIN HEAVY CHAIN >gnl|PID|e220677
                                                   (Z69240)clathrin heavy chain [Schizosaccharomyces pomb
<RIBOSYLATION FACTOR>
Contig1661_k5d07a1.r1     904  5.7e-90   101 643   sp|P34727|ARF_AJ      ADP-RIBOSYLATION FACTOR >pir||D49993
                                                   ADP-ribosylationfactor - Ajellomyces capsulata >gi|407693 (L2511
<SIGNAL RECOGNITION PARTICLE>
Contig1155_x7d07a1.f1     538  3.2e-51   428 778   sp|Q00179|SR54_A      SIGNAL RECOGNITION PARTICLE 54 KD PROTEIN
                                                   HOMOLOG>pir||JC4572 signal recognition particle 54K protein
r1f06a1.f1                130  1.1e-07   228 596   sp|P36057|SRPB_YEAST  PUTATIVE SIGNAL RECOGNITION PARTICLE RECEPTOR
                                                   BETASUBUNIT (SR-BETA) >pir||S37984 probable
<MITOCHONDRIAL IMPORT RECEPTOR>
Contig294_n0a07a1.f1      114  2.9e-06   501 659   gnl|PID|d1022291      (AB004538) MITOCHONDRIAL IMPORT RECEPTOR SUBUNIT
                                                    TOM40[Schizosaccharomyces pombe]
<synaptobrevin-protein trafficing>
Contig1563_e4a05a1.f1     375  6.5e-34   434 709   gi|2769755           (AF010288) synaptobrevin [Aspergillus parasiticus]
*Contig965_m3h03a1.r1     195  2.5e-14   186 653   sp|Q16943|VP33_A      VESICLE-ASSOCIATED MEMBRANE PROTEIN/SYNAPTOBREVIN
                                                   BINDINGPROTEIN (VAP-33) >pir||A57245 VAMP-binding p
Contig830_z4b01a1.f1      156    1e-10   356 550   sp|P47192|SYBR_A      SYNAPTOBREVIN-RELATED PROTEIN >gi|600710
                                                   (M90418)formerly called HAT24; synaptobrevin-related protein
<VESICLE TRANSPORT V-SNARE PROTEIN>
*Contig987_c3f10a1.r1     319  5.6e-28   262 891   sp|P78768|VTI1_S      VESICLE TRANSPORT V-SNARE PROTEIN VTI1
                                                   HOMOLOG>gnl|PID|d1014475 (D89116) similar to Saccharomyces cer
<COATOMER ALPHA SUBUNIT-trafficing to golgi, nonclathrin vesicles>
Contig577_c5b03a1.f1     1366 6.5e-139     3 815   gi|3170523           (AF053883) coatomer alpha subunit [Emericella
                                                    nidulans]
<COATOMER BETA SUBUNIT-trafficing to golgi, nonclathrin vesicles>
r2a12a1.f1                566    2e-53   278 805   sp|P23514|COPB_RAT    COATOMER BETA SUBUNIT (BETA-COAT PROTEIN)
```

					(BETA-COP)>pir||S13520 beta-COP protein - rat >g
g1d10a1.r1	398	2.9e-35	29	679	sp|P41810|COPB_YEAST COATOMER BETA SUBUNIT (BETA-COAT PROTEIN) (BETA-COP)>pir||S54534 coatomer complex beta cha
i7g05a1.r1	302	3.9e-26	1	468	gi|2809537 (AF043120) beta prime coatomer protein [Mus musculus]
<COATOMER ZETA SUBUNIT-trafficing to golgi, nonclathrin vesicles>					
k9f10a1.f1	137	4.1e-08	293	535	sp|P53600|COPZ_YEAST COATOMER ZETA SUBUNIT (ZETA-COAT PROTEIN) (ZETA-COP)>pir||S52521 hypothetical protein YPL0
<COATOMER GAMMA SUBUNIT>					
Contig178_i8d09a1.r1	171	5.2e-11	9	371	sp|P87140|COPG_S PROBABLE COATOMER GAMMA SUBUNIT (GAMMA-COAT PROTEIN)(GAMMA-COP) >gnl|PID|e316116 (Z95396) unknown [Sc
<NPL6 PROTEIN-nuclear protein localization>					
Contig174_i8g02a1.r1	166	2.3e-09	282	518	sp|P32832|NPL6_Y NPL6 PROTEIN >pir||S30792 NPL6 protein - yeast(Saccharomyces cerevisiae) >gi|172050 (M98434) nuclear

7.Turnover-protein degradation-including vacuolar (91)

<protein-L-isoaspartate O-methyltransferase-esterification for degradation>					
Contig1569_g3g03a1.f1	490	4.4e-46	761	1474	sp|Q27869|PIMT_D PROTEIN-L-ISOASPARTATE(D-ASPARTATE) O-METHYLTRANSFERASE(PROTEIN-BETA-ASPARTATE METHYLTRANSFERASE) (PI
<PROTEASE REGULATORY SUBUNIT>					
Contig1452_m2d09a1.f1	1039	2.8e-104	42	767	sp|P33299|PRS7_Y 26S PROTEASE REGULATORY SUBUNIT 7 HOMOLOG (CIM5 PROTEIN)(TAT-BINDING HOMOLOG 3) >pir||S34354 tat-bind
Contig636_c0h10a1.f1	960	6.9e-96	108	845	sp|P41836|PRS8_S 26S PROTEASE REGULATORY SUBUNIT 8 HOMOLOG (LET1 PROTEIN)>pir||S45176 transcription factor SUG1 homolo
Contig656_o8g12a1.r1	851	2.4e-84	7	822	sp|P33297|PRSA_Y 26S PROTEASE REGULATORY SUBUNIT S6A (TAT-BINDING PROTEINHOMOLOG 1) (TBP-1) >pir||S46605 YTA1 protein
Contig1319_g6f01a1.r1	810	5.2e-80	102	653	sp|P78578|PRS6_A 26S PROTEASE REGULATORY SUBUNIT 6B HOMOLOG >gi|1777414(U15601) 26S proteasome subunit [Aspergillus ni
o8g12a1.f1	588	1.7e-56	129	614	sp|P33297|PRSA_YEAST 26S PROTEASE REGULATORY SUBUNIT S6A (TAT-BINDING PROTEINHOMOLOG 1) (TBP-1) >pir||S46605 YT
Contig634_c0h10a1.r1	581	9.6e-56	126	821	sp|P47210|PRS8_H 26S PROTEASE REGULATORY SUBUNIT 8 (PROTEASOME SUBUNITP45) (THYROID HORMONE RECEPTOR INTERACTING PROTE
Contig717_u4g08a1.r1	338	5.6e-30	151	477	sp|P33299|PRS7_Y 26S PROTEASE REGULATORY SUBUNIT 7 HOMOLOG (CIM5 PROTEIN)(TAT-BINDING HOMOLOG 3) >pir||S34354 tat-bind
<proteasome>					
Contig1183_g6f01a1.f1	930	1.1e-92	203	826	sp|P78578|PRS6_A 26S PROTEASE REGULATORY SUBUNIT 6B HOMOLOG >gi|1777414(U15601) 26S proteasome subunit [Aspergillus ni
Contig808_z4f05a1.f1	780	7.6e-77	106	735	sp|Q09682|PRC9_S PUTATIVE PROTEASOME COMPONENT C9/Y13 (MACROPAIN SUBUNIT)(MULTICATALYTIC ENDOPEPTIDASE COMPLEX SUBUNIT
Contig1547_m8e08a1.f1	696	6.2e-68	139	867	sp|P25786|PRC2_H PROTEASOME COMPONENT C2 (MACROPAIN SUBUNIT C2)(PROTEASOME NU CHAIN) (MULTICATALYTIC ENDOPEPTIDASE COM
Contig990_m8e02a1.f1	691	2.1e-67	27	683	gnl|PID|e1263959 (AL022117) putative proteosome subunit [Schizosaccharomycespombe]
Contig1453_u4b10a1.r1	681	2.5e-66	85	786	sp|P23724|PRC5_Y POTENTIAL PROTEASOME COMPONENT C5

(MULTICATALYTICENDOPEPTIDASE COMPLEX SUBUNIT C5) >pir||S42436 multi

Contig721_t2e09a1.f1	560	7.4e-53	126	677	gnl|PID|d1011888 (D78151) human 26S proteasome subunit p97 [Homo sapiens]
Contig1033_d1f02a1.f1	544	8.2e-52	106	639	gi|476044 (X78991) proteasomal subunit Pre3 [Saccharomyces cerevisiae]>gi|854572 (X87611) proteasome component
n0f06a1.r1	438	1.3e-40	119	457	sp|P21242|PRC8_YEAST PROTEASOME COMPONENT C1 (MACROPAIN SUBUNIT C1)(PROTEINASE YSCE SUBUNIT 1) (MULTICATALYTIC
15h10a1.f1	325	1.3e-28	90	518	gnl|PID|d1020529 (AB003102) 26S proteasome subunit p44.5 [Homo sapiens]
z2a09a1.f1	280	1.8e-23	154	504	gnl|PID|d1020530 (AB003103) 26S proteasome subunit p55 [Homo sapiens]
k5a04a1.r1	262	1.2e-21	156	614	gi|3450889 (AF083890) 20S proteosome subunit 9 [Arabidopsis thaliana]
n0f06a1.f1	216	3.7e-17	205	429	sp|P21242|PRC8_YEAST PROTEASOME COMPONENT C1 (MACROPAIN SUBUNIT C1)(PROTEINASE YSCE SUBUNIT 1) (MULTICATALYTIC

<CLPB PROTEIN>

Contig1439_n3d06a1.f1	281	8.9e-23	261	623	sp|P53532|CLPB_C CLPB PROTEIN >gi|1163118 (U43536) heat-inducibleexpression; two ATP-binding domains; ClpB homolog, si

<ubiquitin>

Contig1865_a1g03f2.f1	1457	1.4e-148	142	1023	pir||D29456 ubiquitin precursor UBI4 - yeast (Saccharomyces cerevisiae)>gi|4734 (X05731) ubiquitin (AA 1-381) [Sa
Contig1048_d3c02a1.r1	1091	9.2e-110	104	1201	sp|P22515|UBA1_Y UBIQUITIN-ACTIVATING ENZYME E1 1 >pir||S38048ubiquitin--protein ligase (EC 6.3.2.19) - yeast (Sacchar
Contig1479_j7d10a1.f1	1084	2.6e-107	271	1098	gnl|PID|e351301 (Z99531) ubiquitin system protein [Schizosaccharomyces pombe]
Contig1475_a0f02a1.r1	958	1e-95	198	782	pir||D29456 ubiquitin precursor UBI4 - yeast (Saccharomyces cerevisiae)>gi|4734 (X05731) ubiquitin (AA 1-381) [Sa
o8f12a1.r1	896	4.3e-89	3	554	gi|2262193 (U62795) ubiquitin ligase Publ [Schizosaccharomyces pombe]>gi|2408007 (Z99161) ubiquitin ligase
Contig1650_d3c02a1.f1	856	7e-85	130	1023	gi|4715 (X55386) ubiquitin-activating enzyme [Saccharomyces cerevisiae]
m8d11a1.r1	788	1.1e-77	42	827	sp|Q09879|UBPB_SCHPO PUTATIVE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE C8A4.01C(UBIQUITIN THIOLESTERASE) (UBIQUITI
Contig1093_k0e12a1.f1	469	1.5e-74	293	577	gnl|PID|e1251311 (AJ223328) polyubiquitin [Nicotiana tabacum]>gnl|PID|e1311722 (AJ007936) polyubiquitin [Gibberella pu
Contig1509_n8a02a1.r1	622	3.9e-60	22	390	pir||A29456 ubiquitin / ribosomal protein CEP52 - yeast (Saccharomycescerevisiae) >gi|4728 (X05728) ubiquitin [Sa
Contig621_o8a06a1.r1	609	9.5e-59	181	561	gi|166336 (M88684) polyubiquitin [Aglaothamnion neglectum] >prf||1908440Apoly-ubiquitin [Aglaothamnion neglectu
Contig271_g9h01a1.f1	521	2.2e-49	348	803	gi|3309661 (AF075599) ubiquitin conjugating enzyme 12 [Homo sapiens]
Contig53_m3f08a1.r1	437	1.8e-40	24	356	gi|1143188 (U32627) ubiquitin precursor [Candida albicans]
Contig376_g1d09a1.r1	436	2.7e-40	29	403	gi|2408071 (Z99166) ubiquitin fusion degradation protein [Schizosaccharomycespombe]
Contig24_a0f02a1.f1	388	2.6e-35	242	484	gi|571519 (U16852) polyubiquitin [Gracilaria verrucosa]

				>prf||2109223Apoly-ubiquitin [Gracilaria verrucosa]
c4a04a1.f1	355	8.6e-32	139 558	sp|P14682|UBC3_YEAST UBIQUITIN-CONJUGATING ENZYME E2-34 KD (UBIQUITIN-PROTEINLIGASE) (UBIQUITIN CARRIER PROTEIN
Contig389_f2b09a1.r1	349	3.5e-30	323 754	sp|P39940|RSP5_Y UBIQUITIN--PROTEIN LIGASE RSP5 >pir||S43217 hypotheticalprotein YER125w - yeast (Saccharomyces cerevi
g3h09a1.r1	330	3.8e-29	80 505	sp|Q92353|UBPC_SCHPO PUTATIVE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE C6G9.08(UBIQUITIN THIOLESTERASE) (UBIQUITIN
Contig1105_c6c07a1.f1	324	1.7e-28	3 299	gi|3265058 (AF060232) monoubiquitin/carboxy extension protein fusion[Botryotinia fuckeliana]
Contig98_k9g12a1.f1	320	4.4e-28	335 562	gnl|PID|e354806 (AJ003818) ubiquitin-like protein [Schizosaccharomyces pombe]
r6c08a1.r1	283	9.4e-23	143 541	sp|P39538|UBPC_YEAST UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 12 (UBIQUITINTHIOLESTERASE 12) (UBIQUITIN-SPECIFIC P
Contig1277_j4a05a1.f1	257	2.1e-21	173 424	sp|Q12306|SMT3_Y UBIQUITIN-LIKE PROTEIN SMT3 >pir||S63999 SMT3 protein -yeast (Saccharomyces cerevisiae) >gi|881372 (U
Contig1352_g2f11a1.f1	262	7.8e-21	452 706	gi|2262193 (U62795) ubiquitin ligase Pubi [Schizosaccharomyces pombe]>gi|2408007 (Z99161) ubiquitin ligase [Schi
d3d05a1.f1	247	3.7e-19	93 359	gi|1843535 (U82122) E6-AP ubiquitin-protein ligase [Mus musculus]
Contig1818_d2g09a1.f1	226	3.8e-18	231 377	pir||S62680 ubiquitin-extension protein - Emericella nidulans
z7e03a1.f1	226	3.8e-18	185 604	sp|P15374|UBL3_HUMAN UBIQUITIN CARBOXYL-TERMINAL HYDROLASE ISOZYME L3 (UCH-L3)(UBIQUITIN THIOLESTERASE L3) >pir
v3g05a1.f1	229	8.2e-18	212 556	sp|Q09738|UBPA_SCHPO PUTATIVE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE C13A11.04C(UBIQUITIN THIOLESTERASE) (UBIQUI
e9g11a1.f1	230	2.1e-17	135 515	gnl|PID|e1251102 (AL021838) ubiquitin carboxyl-terminal hydrolase[Schizosaccharomyces pombe]
Contig436_e0c04a1.f1	220	4e-17	210 500	sp|P48510|DSK2_Y UBIQUITIN-LIKE PROTEIN DSK2 >gi|786151 (L40587)ubiquitin-like protein [Saccharomyces cerevisiae]
o0f09a1.r1	213	8.8e-15	31 429	gnl|PID|e351301 (Z99531) ubiquitin system protein [Schizosaccharomyces pombe]
e0e09a1.r1	182	1.2e-12	318 608	sp|Q92353|UBPC_SCHPO PUTATIVE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE C6G9.08(UBIQUITIN THIOLESTERASE) (UBIQUITIN
Contig527_c8b10a1.f1	174	1.8e-09	330 785	gnl|PID|e1298610 (AL023859) putative ubiquitin protein ligase[Schizosaccharomyces pombe]
*Contig1346_a1b01c9.r1	142	7e-09	189 515	gi|2668744 (AF034946) ubiquitin conjugating enzyme [Zea mays]
Contig18_q0b07a1.f1	114	2.9e-06	427 510	pir||S43786 ubiquitin--protein ligase (EC 6.3.2.19) - Arabidopsis thaliana
<UBIQUITIN-CONJUGATING ENZYME>				
Contig1649_m8h12a1.f1	632	3.8e-61	394 942	sp|P16577|UBC4_W UBIQUITIN-CONJUGATING ENZYME E2-23 KD (UBIQUITIN-PROTEINLIGASE) (UBIQUITIN CARRIER PROTEIN) >pir||A34
Contig939_c7g10a1.f1	318	1.5e-55	450 635	gi|3323498 (AF030296) ubiquitin conjugating enzyme UBC1 [Glomerella cingulata]
d3d05a1.r1	293	4.3e-24	129 602	gi|1872514 (U84404) E6-associated protein E6-AP/ubiquitin-protein ligase [Homosapiens] >gi|2361031 (A

Contig1112_c6d08a1.f1	206	5.5e-16	501 677	sp|O00102|UBC7_S	UBIQUITIN-CONJUGATING ENZYME E2-18 KD (UBIQUITIN-PROTEINLIGASE) (UBIQUITIN CARRIER PROTEIN) >gnl|PID|
*Contig797_y6b01a1.f1	195	8e-15	95 217	sp|P52493|UBC2_N	UBIQUITIN-CONJUGATING ENZYME E2-17 KD (UBIQUITIN-PROTEINLIGASE 2) (UBIQUITIN CARRIER PROTEIN) >gnl|PI
<pepsinogen>					
Contig200_i2g09a1.r1	918	1.9e-91	72 653	gi|530795	(U03278) pepsinogen [Aspergillus niger]
m3c08a1.r1	288	1.1e-24	277 567	gi|530795	(U03278) pepsinogen [Aspergillus niger]
e4h01a1.f1	247	4.7e-20	16 159	gi|530795	(U03278) pepsinogen [Aspergillus niger]
<aspartic protease>					
m5f01a1.f1	181	4.6e-11	205 642	sp|P32329|YAP3_YEAST	ASPARTIC PROTEINASE 3 PRECURSOR (YAPSIN 1) >pir||S64957aspergillopepsin I (EC 3.4.23.18) Y
Contig117_m5f01a1.r1	119	0.0001	299 547	sp|Q12303|YL12_Y	PUTATIVE ASPARTYL PROTEASE YLR121C PRECURSOR >pir||S64958probable membrane protein YLR121c - yeast (S
<proline peptidase>					
<aminopeptidase>					
*Contig1816_a1h01c9.r1	991	3.7e-99	668 1759	sp|P38174|AMP2_Y	METHIONINE AMINOPEPTIDASE 2 (METAP 2) (PEPTIDASE M 2)>gi|1045302 (U17437) methionine aminopeptidase 2
Contig1347_g3b06a1.f1	915	4.1e-91	172 1263	sp|P37302|APE3_Y	AMINOPEPTIDASE Y PRECURSOR >pir||A54134 aminopeptidase Y(EC 3.4.11.-) - yeast (Saccharomyces cerevisi
Contig1504_o3d01a1.f1	624	2.6e-60	237 1124	gi|3366	(Y07522) aminopeptidase ysc1 (AA 1-514) [Saccharomyces cerevisiae]
Contig946_r7a12a1.r1	586	5.9e-56	30 1268	gi|3368	(X63998) aminopeptidase yscII [Saccharomyces cerevisiae]
p0f10a1.f1	428	1.6e-39	65 490	gnl|PID|e339951	(Z98980) aminopeptidase i precursor [Schizosaccharomycespombe]
Contig501_c9h01a1.f1	399	1.9e-36	4 504	sp|P38174|AMP2_Y	METHIONINE AMINOPEPTIDASE 2 (METAP 2) (PEPTIDASE M 2)>gi|1045302 (U17437) methionine aminopeptidase 2
c9a01a1.r1	319	6.2e-27	58 531	sp|P18962|DAP2_YEAST	DIPEPTIDYL AMINOPEPTIDASE B (DPAP B) (YSCV) >pir||S46780hypothetical protein YHR028c - yea
Contig210_m7d03a1.f1	286	1.9e-24	170 418	pir||S45411	methionyl aminopeptidase (EC 3.4.11.18) MAP2 - yeast(Saccharomyces cerevisiae) >gi|496684 (X79489) D-
w6h10a1.f1	294	3.3e-24	9 470	sp|P18962|DAP2_YEAST	DIPEPTIDYL AMINOPEPTIDASE B (DPAP B) (YSCV) >pir||S46780hypothetical protein YHR028c - yea
j7a03a1.r1	226	1.2e-15	118 528	gi|2773225	(AF039716) Similar to aminopeptidase; coded for by C. elegans cDNAyk91g4.3; coded for by C
j7a03a1.f1	148	1.8e-07	98 358	gnl|PID|e1169567	(X95762) Aminopeptidase P-like [Homo sapiens]
<iminopeptidase>					
g7b07a1.r1	378	3.3e-34	70 597	sp|P46547|PIP_AERSO	PROLINE IMINOPEPTIDASE (PROLYL AMINOPEPTIDASE)>pir||JC4184 prolyl aminopeptidase (EC 3.4.1
g7b07a1.f1	255	8.2e-21	163 552	sp|P46547|PIP_AERSO	PROLINE IMINOPEPTIDASE (PROLYL AMINOPEPTIDASE)>pir||JC4184 prolyl aminopeptidase (EC 3.4.1
<serine protease>					
Contig322_g5d04a1.r1	742	8.4e-73	89 760	gnl|PID|e319086	(Y13338) intracellular vacuolar serine proteinase

[Aspergillusfumigatus]

n0c11a1.r1	602	5e-58	4	462	gnl|PID|e319086 (Y13338) intracellular vacuolar serine proteinase [Aspergillusfumigatus]
<ASPARTATE PROTEASE>					
o8c12a1.r1	391	1.2e-35	104	550	gi|1469396 (U43775) secreted aspartic proteinase precursor [Glomerellacingulata]
<metallopeptidase>					
Contig1854_a5c06a1.f1	625	5.2e-78	271	957	sp|P46073|ME24_A 24 KD METALLOPROTEINASE PRECURSOR (DEUTEROLYSIN)>pir||JC4378 metalloproteinase (EC 3.4.-.-) 23K - Asp
Contig1687_j5b07a1.f1	608	1.4e-58	257	1042	gnl|PID|e1293248 (AL023590) putative metallopeptidase [Schizosaccharomycespombe]
Contig1855_a1e05c9.r1	524	9.9e-50	99	659	pir||JC4379 metalloproteinase (EC 3.4.-.-) 23K - Aspergillus fumigatus>gi|780794 (U24146) MEP20 [Aspergillus fumi
Contig235_h4h04a1.f1	462	3.8e-43	277	615	sp|P46073|ME24_A 24 KD METALLOPROTEINASE PRECURSOR (DEUTEROLYSIN)>pir||JC4378 metalloproteinase (EC 3.4.-.-) 23K - Asp
<ca dependent protease>					
l3d08a1.r1	132	1.4e-05	235	378	prf||1613155A Ca dependent Cys protease p94 [Rattus norvegicus]
<alkaline protease>					
Contig1828_a5e07a1.f1	1909	1.7e-196	205	1413	gi|470731 (L31778) alkaline protease [Aspergillus nidulans]
<ACID PROTEASE A>					
x1b07a1.r1	347	5.4e-31	7	444	sp|P24665|PRTA_ASPNG ASPERGILLOPEPSIN II PRECURSOR (ACID PROTEASE A) (PROCTASEA) >pir||A41025 aspergillopepsin
<CAAX PRENYL PROTEASE-cleavage of alpha factor for activatio>					
Contig421_e9b02a1.r1	487	8.5e-46	8	544	sp|P47154|ST24_Y CAAX PRENYL PROTEASE 1 (PRENYL PROTEIN-SPECIFICENDOPROTEASE 1) (PPSEP 1) (A-FACTOR CONVERTING ENZYME)
n3b02a1.f1	167	5.3e-09	221	427	sp|P47154|ST24_YEAST CAAX PRENYL PROTEASE 1 (PRENYL PROTEIN-SPECIFICENDOPROTEASE 1) (PPSEP 1) (A-FACTOR CONVERT
<insulinase-peptidase M16 family>					
j0g08a1.r1	374	1.1e-32	73	501	sp|O14077|YEAC_SCHPO PUTATIVE ZINC-PROTEASE C2E11.12C >gnl|PID|e339165(Z98850) hypothetical protease=insulinase family
<Lon serine protease>					
y7c06a1.f1	140	1e-05	373	555	sp|P36775|LONM_YEAST MITOCHONDRIAL ATP-DEPENDENT PROTEASE PRECURSOR>pir||S43938 proteinase PIM1 (EC 3.4.21.-)=LON gene serine protease

V. Processes
A. Cell rescue, defense, osmotic adaptation, starvation response, development (asexual, sexual) includes antibiotics, toxins.
See also B.cell signalling, signal transduction and C. transmembrane transport

1. development (8)
a. conidiation-asexual
<velvet A>

Contig1621_u4c01a1.f1 970 6.1e-97 256 1089 gi|3329358 (U95045) velvet A [Emericella nidulans]
<CONIDIATION-SPECIFIC PROTEIN 6>
r3c09a1.r1 164 1.5e-11 88 324 sp|P34762|CON6_NEUCR CONIDIATION-SPECIFIC PROTEIN 6 >gi|415714 (L26036)conidiation protein [Neurospora crassa]

b. pigment production
<GREEN PIGMENT SYNTHASE>
g9a11a1.r1 668 1.9e-63 11 478 sp|Q03149|WA_EMENI CONIDIAL GREEN PIGMENT SYNTHASE >pir||S28353 probablepolyketide synthase - Emericella nidu

<PORPHYRIN>
<porphyrinogen oxidase>
Contig458_d4d03a1.r1 522 1.8e-49 116 874 sp|P35055|HEM6_S COPROPORPHYRINOGEN III OXIDASE PRECURSOR(COPROPORPHYRINOGENASE) (COPROGEN OXIDASE) >pir||S39905coprop
Contig857_r3e10a1.f1 314 1.8e-27 110 484 gi|171316 (J03873) coproporphyrinogen oxidase (EC 1.3.3.3) [Saccharomycescerevisiae]

<polyketide synthase>
Contig1214_g9a11a1.f1 664 9.1e-109 483 950 gi|3136092 (AF025541) polyketide synthase [Aspergillus fumigatus]
Contig295_g7e09a1.f1 279 5.5e-22 1 528 pir||S60224 melanin biosynthetic polyketide synthase PKS1 - Colletotrichumlagenarium >gnl|PID|d1019697 (D83643) p

<LACCASE>
k5d05a1.f1 255 5.2e-20 191 502 sp|P17489|LAC1_EMENI LACCASE PRECURSOR (BENZENEDIOL:OXYGEN OXIDOREDUCTASE)(URISHIOL OXIDASE) (LACCASE I) >gi|24

2.defense (41)

a.Secondary metabolites
-Penicullin production
<penicill>
Contig416_n5f11a1.r1 1022 1.8e-102 87 1016 sp|P21133|AAAA_E ACYL-COENZYME A:6-AMINOPENICILLANIC-ACID-ACYLTRANSFERASEPRECURSOR (ISOPENICILLIN N ACYLTRANSFERASE) >
r3c06a1.f1 405 4.5e-37 283 513 sp|P21133|AAAA_EMENI ACYL-COENZYME A:6-AMINOPENICILLANIC-ACID-ACYLTRANSFERASEPRECURSOR (ISOPENICILLIN N ACYLTRA

-sterigmatocystin biosynthesis
<sterigmatocystin>
Contig1677_g2d05a1.f1 1543 1.1e-157 72 983 sp|Q00707|STCL_E PROBABLE STERIGMATOCYSTIN BIOSYNTHESIS P450 MONOOXYGENASESTCL >gi|1235628 (U34740) putative p450 mono
Contig1811_a1g02c9.r1 1444 3.2e-147 5 1033 gi|1293655 (U51327) versicolorin B synthase [Aspergillus parasiticus]>gi|1293657 (U51328) versicolorin B synthas
Contig1291_c8f01a1.f1 1328 7.6e-134 260 1003 sp|Q12397|STCA_E PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS POLYKETIDESYNTHASE (PKS) >gi|972728 (L39121) polyketide syntha
Contig529_o4e08a1.r1 1310 5.5e-133 49 849 sp|Q00727|STCV_E PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS DEHYDROGENASE STCV>gi|1235634 (U34740) putative dehydrogenase
Contig1627_a1c06f2.f1 1231 1.2e-124 61 852 sp|Q00791|STCU_E VERSICOLORIN REDUCTASE >gi|1235633 (U34740)

					putativeketoreductase [Emericella nidulans]
Contig1600_m8f12a1.f1	1125	2.3e-113	533	1156	sp|Q00730|STCW_E PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS MONOOXYGENASE STCW>gi|1235635 (U34740) putative FAD-containing
Contig1527_r3b09a1.r1	1107	1.8e-111	268	963	sp|Q00674|STCE_E PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS KETOREDUCTASE STCE>gi|1235622 (U34740) putative ketoreductase
Contig1079_a1f01c9.r1	845	4.4e-110	66	584	sp|Q12609|STCF_E PROBABLE STERIGMATOCYSTIN BIOSYNTHESIS P450 MONOOXYGENASESTCF >gi|1235624 (U34740) putative p450 mono
c8f01a1.r1	1098	3.4e-109	11	658	sp|Q12397|STCA_EMENI PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS POLYKETIDESYNTHASE (PKS) >gi|972728 (L39121) polyke
Contig1442_e9c04a1.r1	1066	3.9e-107	23	847	gi|1293655 (U51327) versicolorin B synthase [Aspergillus parasiticus]>gi|1293657 (U51328) versicolorin B synthas
n2e10a1.r1	1059	7.8e-106	28	669	sp|Q00681|STCJ_EMENI PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS FATTY ACIDSYNTHASE ALPHA SUBUNIT >gi|1235626 (U3474
Contig1158_g5a07a1.r1	1033	1.1e-103	119	715	sp|Q00730|STCW_E PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS MONOOXYGENASE STCW>gi|1235635 (U34740) putative FAD-containing
Contig736_x7g01a1.f1	993	2.2e-99	45	782	sp|Q00714|STCS_E PROBABLE STERIGMATOCYSTIN BIOSYNTHESIS P450 MONOOXYGENASESTCS >gi|1235631 (U34740) putative p450 mono
Contig21_r4c09a1.r1	973	2.8e-97	85	792	sp|Q00674|STCE_E PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS KETOREDUCTASE STCE>gi|1235622 (U34740) putative ketoreductase
g9b02a1.f1	939	2.2e-92	120	713	sp|Q00706|STCK_EMENI PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS FATTY ACIDSYNTHASE BETA SUBUNIT >gi|1235627 (U34740
Contig1009_c9f01a1.f1	744	1.8e-85	150	752	sp|Q12609|STCF_E PROBABLE STERIGMATOCYSTIN BIOSYNTHESIS P450 MONOOXYGENASESTCF >gi|1235624 (U34740) putative p450 mono
Contig1406_d2f11a1.r1	496	2.6e-84	5	283	sp|Q12608|STCB_E PROBABLE STERIGMATOCYSTIN BIOSYNTHESIS P450 MONOOXYGENASESTCB >gi|1235620 (U34740) putative p450 mono
g9h10a1.f1	844	3.9e-82	2	541	sp|Q12397|STCA_EMENI PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS POLYKETIDESYNTHASE (PKS) >gi|972728 (L39121) polyke
Contig452_d4h03a1.r1	818	7.3e-81	106	639	sp|Q00675|STCI_E PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS LIPASE/ESTERASESTCI >gi|1235625 (U34740) putative lipase/ester
z4c07a1.r1	800	6.1e-79	29	514	sp|Q00713|STCQ_EMENI PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS PROTEIN STCQ>gi|1235630 (U34740) similar to A. para
g9b02a1.r1	799	2.1e-77	13	468	sp|Q00706|STCK_EMENI PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS FATTY ACIDSYNTHASE BETA SUBUNIT >gi|1235627 (U34740
c0d01a1.r1	788	4.1e-76	12	476	sp|Q12397|STCA_EMENI PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS POLYKETIDESYNTHASE (PKS) >gi|972728 (L39121) polyke
u4e02a1.r1	547	7.4e-71	47	340	sp|Q00707|STCL_EMENI PROBABLE STERIGMATOCYSTIN BIOSYNTHESIS P450 MONOOXYGENASESTCL >gi|1235628 (U34740) putativ
Contig1592_e4e01a1.f1	701	1.9e-68	217	600	sp|Q00675|STCI_E PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS LIPASE/ESTERASESTCI >gi|1235625 (U34740) putative lipase/ester
z7a04a1.r1	630	6.4e-61	19	453	sp|Q00710|STCO_EMENI PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS PROTEIN STCO>gi|1235629 (U34740) similar to A. para
Contig657_b0f03a1.f1	610	7.9e-59	133	513	sp|Q00727|STCV_E PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS DEHYDROGENASE STCV>gi|1235634 (U34740) putative dehydrogenase

x7g01a1.r1	562	1e-53	45 572	sp|Q00714|STCS_EMENI PROBABLE STERIGMATOCYSTIN BIOSYNTHESIS P450 MONOOXYGENASESTCS >gi|1235631 (U34740) putativ
Contig961_m3e08a1.r1	532	1.5e-50	62 595	sp|Q00727|STCV_E PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS DEHYDROGENASE STCV>gi|1235634 (U34740) putative dehydrogenase
Contig1619_c5e01a1.f1	529	3.2e-50	547 906	sp|Q00713|STCQ_E PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS PROTEIN STCQ>gi|1235630 (U34740) similar to A. parasiticus put
h4g09a1.f1	397	3e-36	206 463	sp|Q00717|STCT_EMENI PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS PROTEIN STCT>gi|1235632 (U34740) putative translati
c0f02a1.f1	391	1.2e-35	171 704	gi|1200177 (U24698) norsolorinic acid reductase [Aspergillus parasiticus]
Contig779_z5d08a1.r1	302	3.5e-26	3 188	sp|Q00674|STCE_E PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS KETOREDUCTASE STCE>gi|1235622 (U34740) putative ketoreductase
h8e06a1.f1	226	4.3e-18	169 456	sp|Q00668|STCC_EMENI PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS PEROXIDASE STCC>gi|1235621 (U34740) putative peroxi
d5d05a1.r1	174	4.1e-10	46 468	sp|Q12608|STCB_EMENI PROBABLE STERIGMATOCYSTIN BIOSYNTHESIS P450 MONOOXYGENASESTCB >gi|1235620 (U34740) putativ

b.other

<clavulanate-inhibits beta-lactamases>

<VEGETABLE INCOMPATIBILITY PROTEIN-vegetative incompatibility,defense>

Contig820_z4c11a1.f1	134	8.6e-07	6 251	sp|Q00808|HET1_P VEGETATIBLE INCOMPATIBILITY PROTEIN HET-E-1 >gi|607003(L28125) beta transducin-like protein [Podospor
x5f10a1.r1	146	6.5e-06	81 326	sp|Q00808|HET1_PODAN VEGETATIBLE INCOMPATIBILITY PROTEIN HET-E-1 >gi|607003(L28125) beta transducin-like protei

<pisatin demethylase-inactivates plant compound pisatin>

x8d06a1.f1	263	2.3e-21	168 575	pir||S45583 pisatin demethylase (a cytochrome p450)- fungus (Nectria haematococca) >gi|487426(L20976)-defense,pea phytoalexin detoxifying
r8g06a1.f1	169	3.5e-11	15 458	sp|P38364|PIDE_NECHA PISATIN DEMETHYLASE >pir||S34286 pisatin demethylase -fungus (Nectria haematococca) >gi|31

<D-AMINO ACID OXIDASE-oxidation of cephalosporin C>

c1d08a1.f1	151	1.4e-07	297 470	sp|Q99042|OXDA_TRIVR D-AMINO ACID OXIDASE (DAMOX) (DAO) (DAAO)>gnl|PID|e187982 (Z50019) D-amino acid oxidase [T

3.detoxification (9)

<oxygen resistance>

Contig103_k5g12a1.f1	372	1.3e-33	164 499	gi|2979688 (AF035619) singlet oxygen resistance protein [Cercosporanicotianae]

<SHO1 OSMOSENSOR>

Contig843_z3e05a1.f1	133	1.5e-07	377 475	sp|P40073|SS81_Y SSU81 PROTEIN (SHO1 OSMOSENSOR) >pir||S50621 SSU81protein - yeast (Saccharomyces cerevisiae) >gi|6033

<catalase>

Contig404_f0d12a1.r1	1242	8.3e-126	8 703	sp|P55305|CATA_E CATALASE A (SPORE-SPECIFIC CATALASE) >pir||S68115catalase (EC 1.11.1.6) - Emericella nidulans >gi|109

Contig1345_e4g06a1.f1	650	8.7e-125	470 916	sp|P55305|CATA_E CATALASE A (SPORE-SPECIFIC CATALASE) >pir||S68115catalase (EC 1.11.1.6) - Emericella nidulans >gi|109
c1g07a1.r2	705	3.4e-108	66 482	sp|P55305|CATA_EMENI CATALASE A (SPORE-SPECIFIC CATALASE) >pir||S68115catalase (EC 1.11.1.6) - Emericella nidul
f0f12a1.r1	994	1.7e-99	10 612	sp|P78619|CATB_EMENI CATALASE B >gi|1737449 (U80672) catalase [Emericellanidulans]
Contig1099_o8b12a1.f1	735	4.1e-72	78 533	sp|P78619|CATB_E CATALASE B >gi|1737449 (U80672) catalase [Emericellanidulans]
<super oxide dismutase>				
Contig1507_g4h09a1.f1	609	8.9e-59	100 696	sp|P00447|SODM_Y SUPEROXIDE DISMUTASE PRECURSOR (MN) >pir||DSBYNsuperoxide dismutase (EC 1.15.1.1) (Mn) precursor - ye
<epoxide hydrolase-+ water=glycol>				
g3d12a1.r1	274	8.1e-23	74 580	gi|1465805 (U64852) coded for by C. elegans cDNA cm17d4; Similar to epoxidehydrolase. [Caenorhabditis

4.salt tolerance (1)

<HALOTOLERANCE PROTEIN HAL2>				
g9e05a1.f1	295	1.9e-25	127 531	gi|1109672 (U33283) 3'(2'),5-diphosphonucleoside 3'(2') phosphohydrolase[Oryza sativa] >prf||2204308A=homolog of HALOTOLERANCE PROTEIN HAL2 of S. cerevisiae, homology is same
g9e05a1.r1	179	1.3e-12	221 448	sp|P32179|HAL2_YEAST HALOTOLERANCE PROTEIN HAL2 >pir||S35318 HAL2 protein -yeast (Saccharomyces cerevisiae) >gi

5.starvation response (1)

<MAK16 PROTEIN-moves cytoplasmic proteins to vacuole-autophagocytosis>				
c3f08a1.r1	249	5.9e-30	19 264	sp|P40344|AUT1_YEAST AUTOPHAGOCYTOSIS PROTEIN AUT1 >pir||S45130 hypotheticalprotein YNR007c - yeast (Saccharomy

B. Cell signalling, signal transduction

1. Kinases phosphatases and second messengers

a.PHOSPHATASES (15)

<PROTEIN PHOSPHATASE>				
m5b10a1.r1	860	2.7e-85	37 648	gi|2429085 (U59418) protein phosphatase 2A B'alpha3 regulatory subunit [Musmusculus]
o5a03a1.r1	672	1.1e-77	25 447	gi|2290382 (U89985) serine/threonine protein phosphatase PPT1 [Neurosporacrassa]
i2c04a1.r1	460	6.2e-43	27 497	sp|Q09172|P2C2_SCHPO PROTEIN PHOSPHATASE 2C HOMOLOG 2 (PP2C-2) >pir||S54297protein phosphatase 2C - fission yea
*Contig287_g8d11a1.r1	394	1.3e-34	16 612	gi|2459997 (AF012898) protein phosphatase Ssd1 homolog [Candida albicans]
z6e07a1.r1	249	3.5e-20	87 380	pir||S54298 protein phosphatase 2C - fission yeast

(Schizosaccharomyces pombe)>gi|609658 (L34882) prot

Contig1195_v3g09a1.f1	237	2.3e-18	8	217	gnl|PID|d1010595 (D63916) protein phosphotase 2A 65kD regulatory sububit (Asubunit) [Schizosaccharomyces pombe]
c3g08a1.r1	156	1.3e-15	319	543	sp|P40371|P2C1_SCHPO PROTEIN PHOSPHATASE 2C HOMOLOG 1 (PP2C-1) >pir||A56058phosphoprotein phosphatase (EC 3.1.3
Contig941_o3b04a1.f1	173	2e-10	324	791	sp|P40152|YNV7_Y HYPOTHETICAL 37.2 KD PROTEIN IN ALG9-RAP1 INTERGENICREGION >pir||S50713 phosphoprotein phosphatase ho
Contig1224_c4a09a1.r1	143	1.6e-07	666	869	gnl|PID|d1016373 (D90827) Serine/Threonine protein phosphatase (EC 3.1.3.16).[Escherichia coli] >gi|1788143 (AE000278)
<ca dependent protein phosphatase>					
o1b08a1.r1	577	2.4e-55	199	597	gi|2645886 (AF034089) calcineurin subunit B [Neurospora crassa]
<PROTEIN-TYROSINE PHOSPHATASE>					
Contig1586_o1f04a1.f1	305	5.8e-26	403	1011	sp|P34137|PTP1_D PROTEIN-TYROSINE PHOSPHATASE 1(PROTEIN-TYROSINE-PHOSPHATE PHOSPHOHYDROLASE 1) >gi|167862 (L07125)prot
<serine/threonine phosphatase>					
Contig1494_m7c01a1.f1	718	2.9e-70	723	1238	sp|P78968|PPZ_SC SERINE/THREONINE PROTEIN PHOSPHATASE PP-Z >gi|1763281(U73689) PPZ protein phosphatase [Schizosaccharo
r2h10a1.f1	622	4.5e-60	160	633	sp|P36614|PPE1_SCHPO SERINE/THREONINE PROTEIN PHOSPHATASE PPE1 (PHOSPHATASEESP1) >pir||A47727 cell shape contro
<PRL1/PRL2-LIKE PROTEIN>					
Contig403_f0e04a1.r1	969	7.7e-97	24	728	sp|O13615|PRL1_S PRL1/PRL2-LIKE PROTEIN >gnl|PID|d1022249 (AB004535) PRL1[Schizosaccharomyces pombe]
Contig260_i8d10a1.r1	413	6.1e-38	184	528	sp|O13615|PRL1_S PRL1/PRL2-LIKE PROTEIN >gnl|PID|d1022249 (AB004535) PRL1[Schizosaccharomyces pombe]
b.Kinases (35)					
<protein kinase>					
Contig1226_g7h05a1.r1	668	6.1e-65	13	711	sp|P32490|MKK1_Y PROTEIN KINASE MKK1/SSP32 >pir||A48069 protein kinaseMKK1 (EC 2.7.1.-) - yeast (Saccharomyces cerevis
Contig701_w6h11a1.f1	278	6.1e-23	416	739	gnl|PID|e1293569 (AL023634) protein kinase dsk1 [Schizosaccharomyces pombe]
Contig1159_g3d04a1.r1	271	5.5e-20	54	635	gi|532798 (U13398) protein kinase [Saccharomyces cerevisiae]
<protein kinase C>					
Contig281_g9b09a1.r1	981	3.8e-98	9	623	sp|Q00078|KPC1_A PROTEIN KINASE C-LIKE >pir||S61917 protein kinase Chomolog PKCA - Aspergillus niger >gi|507900 (U1054
Contig1630_c4c06a1.f1	702	1.4e-68	81	920	sp|Q10138|YAS2_S HYPOTHETICAL 51.5 KD PROTEIN C3H8.02 IN CHROMOSOME I>gi|1177660 (Z69086) unknown [Schizosaccharomyces
Contig1176_e7a01a1.r1	510	4.1e-47	341	658	sp|Q00078|KPC1_A PROTEIN KINASE C-LIKE >pir||S61917 protein kinase Chomolog PKCA - Aspergillus niger >gi|507900 (U1054
<MAP kinase>					
k8b03a1.f1	152	4.6e-10	55	162	gi|496307 (L26523) stem loop mutation suppressor [Saccharomyces cerevisiae]-MAP kinase regulator
<MAP kinase HOG1>					

Contig1648_g4a02a1.f1	821	3.7e-81	420	1109	sp|P32485|HOG1_Y MITOGEN-ACTIVATED PROTEIN KINASE HOG1 (MAP KINASE HOG1)(OSMOSENSING PROTEIN HOG1) >pir||S64950 protei
<CALMODULIN-DEPENDENT PROTEIN KINASE>					
i3h06a1.r1	815	1.5e-80	23	493	sp|Q00771|KCC1_EMENI CALCIUM/CALMODULIN-DEPENDENT PROTEIN KINASE (CMPK)>pir||JN0323 Ca2+/calmodulin-dependent p
i7a07a1.r1	332	2.2e-29	158	367	sp|Q00771|KCC1_EMENI CALCIUM/CALMODULIN-DEPENDENT PROTEIN KINASE (CMPK)>pir||JN0323 Ca2+/calmodulin-dependent p
i3h06a1.f1	188	3.4e-12	342	497	sp|Q00771|KCC1_EMENI CALCIUM/CALMODULIN-DEPENDENT PROTEIN KINASE (CMPK)>pir||JN0323 Ca2+/calmodulin-dependent p
Contig29_d1f08a1.f1	145	9.6e-09	274	414	gi|2654181 (AF034963) calmodulin-dependent protein kinase; CgCMK [Glomerellacingulata]
<CAMP-DEPENDENT PROTEIN KINASE>					
n8d03a1.r1	700	2.5e-68	32	502	gi|516040 (U12335) cAMP-dependent protein kinase catalytic subunit[Magnaporthe grisea]
Contig1318_i0a06a1.f1	683	1.5e-66	445	849	gi|3170248 (AF043231) cAMP-dependent protein kinase regulatory subunit[Emericella nidulans]
f2h07a1.r1	383	8.6e-34	15	605	sp|P11792|SCH9_YEAST CAMP-DEPENDENT PROTEIN KINASE SCH9 >pir||S48986 probableprotein kinase SCH9 (EC 2.7.1.-) -
<SERINE/THREONINE-PROTEIN KINASE, casein kinase acts on acidic proteins>					
Contig1695_d2d12a1.f1	1127	1.4e-113	727	1581	sp|P25333|KCR8_Y PROBABLE SERINE/THREONINE-PROTEIN KINASE YCR8W>pir||OKBY8W probable protein kinase YCR008w (EC 2.7.1.
Contig1717_c5c12a1.r1	757	2.2e-74	40	987	sp|P38623|RCK2_Y SERINE/THREONINE-PROTEIN KINASE RCK2 (CAM KINASE-LIKEPROTEIN KINASE CLK1) >gi|733594 (U23464) CaM kin
Contig1440_r2c01a1.f1	488	5.7e-74	712	1227	sp|Q01389|BCK1_Y SERINE/THREONINE PROTEIN KINASE BCK1/SLK1/SSP31>pir||S20117 protein kinase BCK1 (EC 2.7.1.-) - yeast(
k5d01a1.r1	602	5.9e-58	16	510	gi|2911462 (AF046923) serine/threonine protein kinase [Colletotrichumtrifolii]
Contig1673_g3g11a1.r1	554	7.1e-53	290	1252	sp|O14019|KDPG_S PROBABLE SERINE/THREONINE-PROTEIN KINASE C29A4.16>gnl|PID|e325359 (Z97210) protein kinase [Schizosacc
Contig881_r4f03a1.f1	502	2.4e-47	28	600	sp|Q07538|PRP4_S SERINE/THREONINE-PROTEIN KINASE PRP4 >gi|1857026 (L10739)serine/threonine kinase [Schizosaccharomyces
m6f09a1.r1	454	4.1e-41	42	413	sp|Q08217|KOE5_YEAST PROBABLE SERINE/THREONINE-PROTEIN KINASE YOL045W>pir||S66730 hypothetical protein YOL045w
i8h08a1.r1	403	1.2e-36	149	490	sp|P33886|WIS1_SCHPO SERINE/THREONINE PROTEIN KINASE WIS1 >pir||S18648 protein kinase wis1 (EC2.7.1.-) - fission yeast (Schizosa
v1h01a1.r1	313	4.9e-26	207	587	sp|P32361|IRE1_YEAST SERINE/THREONINE-PROTEIN KINASE IRE1 PRECURSOR>pir||A47541 protein kinase IRE1-required for iniositol uptake (EC 2.7.1.-
u4h01a1.f1	250	2.7e-19	108	506	sp|P32361|IRE1_YEAST SERINE/THREONINE-PROTEIN KINASE IRE1 PRECURSOR>pir||A47541 protein kinase IRE1 (EC 2.7.1.-
n1a03a1.r1	110	1.7e-05	102	269	gnl|PID|d1033019 (AB014506) non receptor serine/threonine kinase [Dugesiajaponica]

<serine/threonine-specific protein kinase KIN2>
j5c10a1.r1 151 1.2e-17 10 168 sp|P13186|KIN2_YEAST PROTEIN KINASE KIN2 >gi|171789 (M69018) protein kinase 2[Saccharomyces cerevisiae]=serine/threonine-specific protein kinase KIN2,EC 2.7.1

<casein kinase-ser/thr protein kinase, acts on acidic proteins>
m7a09a1.f1 233 7.4e-19 272 499 sp|P40232|KC2B_SCHPO CASEIN KINASE II BETA CHAIN (CK II) >gi|452290 (X74274)casein kinase II beta subunit [Schi
z7e04a1.f1 137 2.9e-07 496 603 sp|Q08466|KC22_ARATH CASEIN KINASE II, ALPHA CHAIN 2 (CK II) >pir||S31099casein kinase II (EC 2.7.1.-) alpha-ty

<histidine kinase>
Contig528_c8b06a1.r1 397 1.9e-35 66 539 gi|1679757 (U77605) two-component histidine kinase CHK-1 [Glomerellacingulata] >gi|1679760 (U77606) two-componen
c5e08a1.r1 241 1.1e-30 374 709 sp|P39928|SLN1_YEAST OSOMOLARITY TWO-COMPONENT SYSTEM PROTEIN SLN1>pir||S48387 SLN1 protein - yeast (Saccharomy
j7e12a1.r1 275 7.3e-22 2 241 gi|1262210 (U50264) Nik-1 histidine kinase [Neurospora crassa]
g2e07a1.r1 220 3.5e-16 104 508 gnl|PID|d1018731 (D90910) sensory transduction histidine kinase [Synechocystissp.]
Contig1187_g2a10a1.f1 146 1.2e-06 245 601 gi|2313493 (AE000555) histidine kinase (cheA) [Helicobacter pylori]
g2c06a1.f1 149 1.6e-06 209 523 gi|3243089 (U69886) histidine kinase [Candida albicans]

<nitrogen permease reactivator-on for nitrogenous transport systems/s-t kinase>
Contig555_o5c08a1.r1 336 9.1e-29 4 420 sp|P22211|NPR1_Y NITROGEN PERMEASE REACTIVATOR PROTEIN >pir||S63138probable protein kinase NPR1 (EC 2.7.1.-) - yeast (

c.cAMP (1)
<adenyl cyclase>
<ADENYLYL CYCLASE-ASSOCIATED PROTEIN-CAP protein, binds cAMP to allow activation>
j9f01a1.f1 258 8.1e-21 179 706 sp|P17555|CAP_YEAST ADENYLYL CYCLASE-ASSOCIATED PROTEIN (CAP) >pir||A34896adenylate cyclase-associated protein
j9f01a1.r1 218 2e-16 125 481 sp|P36621|CAP_SCHPO ADENYLYL CYCLASE-ASSOCIATED PROTEIN (CAP) >pir||A60047adenylyl cyclase-associated protein

2. G protein (24)
<GTP-binding protein>
Contig73_m3f12a1.r1 725 6.5e-80 93 602 sp|P52886|SARA_A GTP-BINDING PROTEIN SARA >gi|1061034 (Z67742) sarA[Aspergillus niger]
n1h10a1.r1 767 1.8e-75 20 604 sp|P53742|YN8U_YEAST HYPOTHETICAL GTP-BINDING PROTEIN IN POP2-HOL1 INTERGENICREGION >pir||S63384 hypothetical p
Contig833_w4d03a1.r1 764 3.5e-75 7 489 sp|P32835|GSP1_Y GTP-BINDING NUCLEAR PROTEIN GSP1/CNR1 >pir||S35504GTP-binding protein GSP1 - yeast (Saccharomyces cer
Contig1665_l3g11a1.f1 736 3.8e-72 201 800 sp|P36586|YPT5_S YPT1-RELATED PROTEIN 5 >pir||S34729 GTP-binding

```
                                                                    proteinypt5 - fission yeast (Schizosaccharomyces pomb
o5f05a1.r1                  699  3.1e-68     7 489   sp|P22129|R11B_DISOM  RAS-RELATED PROTEIN RAB-11B (ORA3)
                                                    >pir||C38625GTP-binding protein ora3 - electric ray
*Contig1458_g6b11a1.f1      546  5.1e-52   611 1159  sp|P01122|RHO_AP      RAS-LIKE GTP-BINDING PROTEIN RHO >pir||TVGAAC
                                                    transformingprotein rho - California sea hare >gi|15580
f0h02a1.r1                  473  2.7e-44   180 611    pir||I38176          ragA - human >pir||I84474 RagA (ras-related,
                                                    alternatively splicedGTPase A) - rat >gi|1063
h1a06a1.f1                  433  4.9e-40     9 359   gnl|PID|e353254      (Z99753) rho protein GTP-binding [Schizosaccharomyces
                                                     pombe]
*Contig383_f5a04a1.f1       245  3.3e-39   502 843    gi|173183           (M33315) GTP-binding protein (VPS1) [Saccharomyces
                                                     cerevisiae]
Contig1227_g7a07a1.f1       398  2.4e-36   295 576   sp|P36017|YP51_Y     GTP-BINDING PROTEIN YPT51/VPS21 >pir||S43399
                                                    GTP-bindingprotein VPS21 - yeast (Saccharomyces cerevisi
c3f01a1.r1                  355  8.4e-32    19 342   sp|P32235|GTP1_SCHPO  GTP-BINDING PROTEIN 1 >pir||JT0741 GTP-binding protein
                                                     1- fission yeast (Schizosaccharomyc
y6a01a1.r1                  170    2e-25   512 649    pir||S71233          GTP-binding protein 2 - Arabidopsis thaliana
                                                    >pir||S71585 Rab2homolog GTP-binding protein
j9c11a1.f1                  244  2.8e-19   226 615   sp|P40010|YEJ6_YEAST  HYPOTHETICAL GTP-BINDING PROTEIN IN PMI40-PAC2
                                                    INTERGENICREGION >pir||S50464 hypothetical
*Contig199_i3a04a1.f1       198  1.3e-14   288 593    gnl|PID|e1294581     (AJ006412) putative GTP-binding protein [Homo sapiens]
y6a01a1.f1                  189  3.2e-14   128 442    pir||E42148          GTP-binding protein rab14 - rat
m8d10a1.r1                  148    2e-08   417 785    pir||A53778          GTP-binding protein A-ras - Emericella nidulans
                                                    >gi|458024(U03023) Ras-like protein [Asper
<GUANINE NUCLEOTIDE-BINDING PROTEIN>
Contig163_j0e04a1.r1        253  5.5e-21   126 287   sp|Q01369|GBLP_N     GUANINE NUCLEOTIDE-BINDING PROTEIN BETA
                                                    SUBUNIT-LIKEPROTEIN (CROSS-PATHWAY CONTROL WD-REPEAT PROTEIN
<RHO1 GDP-GTP EXCHANGE PROTEIN:GUANINE-NUCLEOTIDE RELEASING FACTOR>
g4e12a1.r1                  278  3.7e-22    32 511   sp|P51862|ROM2_YEAST  RHO1 GDP-GTP EXCHANGE PROTEIN 2 >pir||S51389 ROM2
                                                    protein- yeast (Saccharomyces cerevisiae
Contig1206_g4e12a1.f1       207  1.6e-12   624 836   sp|P51862|ROM2_Y     RHO1 GDP-GTP EXCHANGE PROTEIN 2 >pir||S51389 ROM2
                                                    protein- yeast (Saccharomyces cerevisiae) >gi|60940
<rho-gdp dissociation inhibitor-prevents cycling of GDP with GTP of rho protein family>
b0c03a1.r1                  256  2.4e-21    58 378   gnl|PID|e334111      (Z98533) rho-gdp dissociation inhibitor
                                                     [Schizosaccharomycespombe]
b0c03a1.f1                  239  1.7e-19   224 478   gnl|PID|e334111      (Z98533) rho-gdp dissociation inhibitor regulates rho
                                                    protein needed for actin reorganization[Schizosaccharomycespombe]
<RHO2 PROTEIN>
Contig1386_e9f01a1.r1       723  8.5e-71   405 971   sp|Q10133|RHO2_S     RHO2 PROTEIN >pir||JC4045 Rho2 protein - fission
                                                    yeast(Schizosaccharomyces pombe) >gnl|PID|d1007956 (
<GTPASE-ACTIVATING PROTEIN-neg regulator of Ras1, play antagonistic role with rho-gdp dissociation inhibitor>
o4e01a1.r1                  379  1.9e-33    22 480   sp|P33277|GAP1_SCHPO  GTPASE-ACTIVATING PROTEIN >pir||A40258
                                                    RASGTPase-activating protein sar1 - fission yeast (
<RAS-2 PROTEIN-GTP-BINDING>
```

Contig1415_o9d09a1.r1	829	5.1e-82	129 662	sp|Q09914|RHO1_S RHO1 PROTEIN >pir||JC4044 Rho1 protein - fission yeast(Schizosaccharomyces pombe) >pir||S62576 hypoth
Contig1216_g8h09a1.r1	407	2.7e-37	8 571	sp|Q01387|RAS2_N RAS-2 PROTEIN >gnl|PID|d1004223 (D16137) NC-ras-2 protein[Neurospora crassa]

C. Transmembrane transport

1.secretion (6)

<secretion>

c7e05a1.r1	237	4e-18	32 595	gnl|PID|e1316739 (AL031324) subunit of the final step of the secretory pathway[Schizosaccharomyces pombe]
c7e05a1.f1	200	7.4e-13	89 616	gnl|PID|e1316739 (AL031324) subunit of the final step of the secretory pathway[Schizosaccharomyces pombe]
d2f12a1.r1	136	1.3e-05	486 659	sp|Q06245|SC10_YEAST EXOCYST COMPLEX COMPONENT SEC10 >pir||S68482 probablemembrane protein YLR166c - yeast (Sac

<KEXIN-proteinase secretion>

m7g09a1.r1	341	8.2e-39	99 578	sp|O13359|KEX2_CANAL KEXIN PRECURSOR (KEX2 PROTEASE) >gi|2511732 (AF022372)proteinase secretion [Candida albicans]

<VESICULAR-FUSION PROTEIN SEC17>

Contig1446_g7e01a1.r1	349	1.4e-45	82 681	gi|1711132 (U79186) sec17-like protein [Coprinus cinereus]

<SECRETORY COMPONENT PROTEIN SHR3>

Contig165_j0c07a1.f1	171	2.5e-12	349 675	sp|Q02774|SHR3_Y SECRETORY COMPONENT PROTEIN SHR3 >gi|172572 (L01264)secretory component [Saccharomyces cerevisiae]

2.exoenzymes (5)

<exoenzyme>

e0c10a1.r1	720	1.5e-70	26 610	gi|2340046 (L48074) secreted dipeptidyl peptidase [Aspergillus fumigatus]

<dipeptidyl peptidase-exoenzyme>

i8f01a1.r1	691	2.1e-67	20 595	gi|2340046 (L48074) secreted dipeptidyl peptidase [Aspergillus fumigatus]
Contig679_u4e06a1.r1	480	1.6e-44	9 350	gi|2351700 (U87950) dipeptidyl-peptidase IV [Aspergillus fumigatus]
*Contig729_w7f07a1.f1	383	9e-34	24 584	gnl|PID|d1025528 (D89340) dipeptidyl peptidase [Rattus norvegicus]
Contig1312_c9a01a1.f1	121	1e-05	134 298	gnl|PID|e1254390 (AJ002369) prolyl dipeptidyl peptidase [Aspergillus oryzae]

3.transport (85)

a. sugar transport

<sugar transport>

o0f08a1.r1	435	2.7e-40	21 416	gi|2306977 (AF010145) hexose transporter [Aspergillus parasiticus]

s3g03a1.f1	360	5.1e-32	55 561	sp|P39932|STL1_YEAST SUGAR TRANSPORTER STL1 >pir||S69591 sugar transportprotein STP1 - yeast (Saccharomyces cer
*Contig1230_k0h11a1.f1	329	1.4e-28	15 1067	gi|409547 (L07492) sugar transport protein [Saccharomyces cerevisiae]
d5a05a1.r1	310	1.8e-26	68 646	gi|409547 (L07492) sugar transport protein [Saccharomyces cerevisiae]
c6g07a1.r1	289	3.3e-24	11 640	gi|409547 (L07492) sugar transport protein [Saccharomyces cerevisiae]
o0f08a1.f1	263	2.3e-21	184 447	gi|2306977 (AF010145) hexose transporter [Aspergillus parasiticus]
<GLUCOSE TRANSPORTER>				
*Contig1735_d5f08a1.r1	719	2.4e-70	353 1351	sp|Q12300|RGT2_Y HIGH-AFFINITY GLUCOSE TRANSPORTER RGT2 >pir||S67684probable membrane protein YDL138w - yeast (Sacchar
m7a06a1.f1	197	4e-14	206 490	sp|P49374|HGT1_KLULA HIGH-AFFINITY GLUCOSE TRANSPORTER >gi|726336 (U22525)high affinity glucose transporter [Kl
Contig57_o0b10a1.r1	137	1.2e-07	253 414	sp|P32465|HXT1_Y LOW-AFFINITY GLUCOSE TRANSPORTER HXT1 >pir||S38798 hexosetransport protein HXT1 - yeast (Saccharomyce
Contig924_o0b10a1.f1	121	9e-06	232 408	sp|Q12300|RGT2_Y HIGH-AFFINITY GLUCOSE TRANSPORTER RGT2 >pir||S67684probable membrane protein YDL138w - yeast (Sacchar
<GALACTOSE TRANSPORTER>				
r7d01a1.r1	245	1.5e-19	25 717	dbj||AB007638_9 (AB007638) metabolite transport protein [Bacillus subtilis]>gnl|PID|e1182602 (Z99107) simi
<inositol transport>				
Contig1074_f0a04a1.r1	329	7e-29	9 926	gnl|PID|e353127 (Z99708) myo-inositol transport protein homolog [Arabidopsisthaliana]
o5h09a1.r1	208	8.6e-16	27 506	gnl|PID|d1014610 (D89252) similar to Saccharomyces cerevisiae myo-inositoltransporter 1
b. multidrug resistance				
<multidrug resistance>				
r1g01a1.f1	763	4.8e-74	3 581	gi|2673951 (U62933) multidrug resistance protein 1 [Aspergillus fumigatus]>gi|2673953 (U62934) multid
j9b01a1.r1	603	9.5e-57	57 479	gnl|PID|d1032208 (AB010442) PMR1 [Penicillium digitatum]
h8h04a1.r1	599	2.6e-56	39 467	gnl|PID|d1032208 (AB010442) PMR1 MULTIDRUG RESISTANCE[Penicillium digitatum]
r2c11a1.f1	389	4.7e-51	49 381	gi|2673955 (U62935) multidrug resistance protein 2 [Aspergillus fumigatus]>gi|2673957 (U62936) multid
q0g01a1.f1	539	6e-50	2 454	gnl|PID|d1032208 (AB010442) PMR1 [Penicillium digitatum]
o8f05a1.f1	371	4.7e-32	157 666	gi|2668553 (U62929) multidrug resistance protein 1 [Filobasidiella neoformans]>gi|2668555 (U62930) mu
o6b10a1.r1	232	6.1e-18	37 528	gnl|PID|e1313752 (Y14703) multidrug resistance protein [Candida albicans]
<multidrug transporter>				

h0d01a1.f1	605	5.5e-57	44 673	gnl|PID|e219956	(Z68904) ATP-binding cassette multidrug transporter[Emericella nidulans]
y4h01a1.f1	229	1.8e-15	332 541	gnl|PID|e219956	(Z68904) ATP-binding cassette multidrug transporter[Emericella nidulans]
<oxytetracycline exporter>					
l0e10a1.r1	122	4.7e-06	309 476	gi|3108177	(AF061335) oxytetracycline exporter [Streptomyces rimosus]
<cycloheximide resistance protein>					
*Contig1115_r2h04a1.f1	174	1.3e-11	62 571	pir||JC1173	cycloheximide resistance protein - yeast (Candida maltosa)
a1f03f2.f1	161	5.2e-09	33 773	pir||JC1173	cycloheximide resistance protein - yeast (Candida maltosa)
<HOL1 PROTEIN-DRUG RESISTANCE TRANSLOCASE FAMILY, MAJOR FACILITATOR FAMILY>					
Contig1295_m8d08a1.f1	251	6.8e-20	166 648	gi|825501	(L42348) HOL1 [Saccharomyces cerevisiae]

c. nuclear membrane transport

<nuclear pore membrane protein>					
Contig1041_d2b04a1.r1	415	8.7e-37	33 884	sp|P39685|P152_Y	NUCLEAR ENVELOPE PORE MEMBRANE PROTEIN POM152 (P150)>pir||A53824 nuclear pore membrane protein POM152
<NUCLEAR TRANSPORT FACTOR 2>					
Contig472_o0c11a1.f1	394	6e-36	292 660	sp|P33331|NTF2_Y	NUCLEAR TRANSPORT FACTOR 2 (NTF-2) (NUCLEAR TRANSPORTFACTOR P10) >pir||S50467 hypothetical protein YE

d.cation transport-ATPase, or major facilitator superfamily

<cation transport>					
z5c08a1.r1	539	7e-50	12 587	sp|P32660|ATC5_YEAST	PROBABLE CALCIUM-TRANSPORTING ATPASE 5 >pir||S50669hypothetical protein YER166w - yeast (S
e6h10a1.r1	324	3.3e-27	11 472	sp|P54678|ATC1_DICDI	CATION-TRANSPORTING ATPASE PAT1 >pir||S57726 PAT1 protein- slime mold (Dictyostelium disco
d3a10a1.r1	175	2.5e-09	27 278	sp|P40527|ATC7_YEAST	PROBABLE CALCIUM-TRANSPORTING ATPASE 7 >pir||S48431probable membrane protein YIL048w - yea
<CATION-TRANSPORTING ATPASE>					
k0c11a1.r1	440	1.3e-39	37 417	sp|P54678|ATC1_DICDI	CATION-TRANSPORTING ATPASE PAT1 >pir||S57726 PAT1 protein- slime mold (Dictyostelium disco
<CALCIUM-TRANSPORTING ATPASE>					
Contig642_c0a04a1.f1	238	5.4e-18	95 349	sp|P39986|ATC6_Y	PROBABLE CALCIUM-TRANSPORTING ATPASE 6 >pir||S50428hypothetical protein YEL031w - yeast (Saccharomyce
m8b04a1.f1	103	4.1e-05	7 195	pir||A35731	Ca2+-transporting ATPase (EC 3.6.1.38), cardiac and slow skeletalmuscle - rabbit (fragment
<sodium transport>					
Contig1736_i8e08a1.f1	1073	7e-108	44 952	gi|1438947	(U61840) sodium transport ATPase FST [Fusarium solani f. sp. pisi]
<sulfate transporter>					

h1c01a1.f1	300	6.1e-25	15 590	sp|P53394|SULX_YEAST PUTATIVE SULFATE TRANSPORTER YPR003C >pir||S52816probable membrane protein YPR003c - yeast
<cobalt transporter>				
Contig1391_n0d02a1.f1	247	7.8e-20	208 720	sp|P32798|COT1_Y COBALT UPTAKE PROTEIN COT1 >pir||S58327 COT1 protein -yeast (Saccharomyces cerevisiae) >gi|940847 (X9
<COPPER TRANSPORT>				
Contig430_e0f09a1.r1	195	7.6e-15	144 341	sp|Q06686|CTR3_Y COPPER TRANSPORT PROTEIN CTR3 (COPPER TRANSPORTER 3)>pir||S59377 probable membrane protein YLR411w -
<zinc cadmium resistance>				
Contig221_i0b06a1.r1	460	6.4e-43	86 589	sp|P20107|ZRC1_Y ZINC/CADMIUM RESISTANCE PROTEIN >pir||S56057 heavy metalion resistance protein ZRC1 - yeast (Saccharo
o5d12a1.r1	232	2e-18	69 419	gnl|PID|e334142 (Z98559) probable zinc cadmium resistance protein[Schizosaccharomyces pombe]
<manganese resistance>				
Contig1525_g1g10a1.r1	370	2.3e-33	9 359	gnl|PID|e1169877 (AJ001272) manganese resistance 1 protein [Saccharomycescerevisiae]
m6d08a1.f1	287	2.4e-23	193 585	sp|P35724|MNR2_YEAST MANGANESE RESISTANCE PROTEIN >pir||S37886 hypotheticalprotein YKL064w - yeast (Saccharomyc
Contig1478_a0c02a1.f1	178	2.4e-12	315 482	gnl|PID|e1169877 (AJ001272) manganese resistance 1 protein [Saccharomycescerevisiae]

e.Anion transport

<arsenite translocating ATPase-anion transport,resistance to arsenite, antimonite, arsenate>				
c0h09a1.f1	441	6.7e-41	103 594	gi|2905657 (AF047469) arsenite translocating ATPase [Homo sapiens]
c0h09a1.r1	133	1.3e-07	325 486	gi|1616741 (U60276) hASNA-I [Homo sapiens]
<phosphate transporter>				
*Contig1457_j9a02a1.f1	571	1.1e-54	124 693	gnl|PID|d1032546 (AB016066) mitochondrial phosphate transporter [Arabidopsisthaliana]
m2g02a1.f1	545	6.6e-52	61 636	gnl|PID|d1032546 (AB016066) mitochondrial phosphate transporter [Arabidopsisthaliana]
Contig1511_d1f10a1.f1	507	6.5e-48	278 880	pir||S60949 phosphate transport protein, mitochondrial - yeast (Saccharomycescerevisiae) >gi|1050774 (X92441) YOR
Contig91_l3g04a1.r1	319	5.7e-28	300 596	gnl|PID|d1032543 (AB016063) mitochondrial phosphate transporter [Glycine max]
<tartrate transport>				
*Contig258_m8b10a1.f1	316	2.3e-27	349 1344	gi|805291 (U25634) putative tartrate transporter; inducible by tartrate;Method: conceptual translation supplied
m1c01a1.r1	262	1.4e-21	143 667	gi|805291 (U25634) putative tartrate transporter; inducible by tartrate;Method: conceptual translati
<CHOLINE TRANSPORT>				
r4a12a1.r1	313	1e-26	76 798	sp|P19807|HNM1_YEAST CHOLINE TRANSPORT PROTEIN >pir||S11175 choline transportprotein - yeast (Saccharomyces cer

r4a12a1.f1	125	0.00019	425 763	sp|P19807|HNM1_YEAST CHOLINE TRANSPORT PROTEIN >pir||S11175 choline transportprotein - yeast (Saccharomyces cer
<allantoate transport>				
Contig880_r5f08a1.f1	355	1.6e-31	135 929	sp|P15365|DAL5_Y ALLANTOATE PERMEASE >pir||A28671 allantoate transportprotein - yeast (Saccharomyces cerevisiae) >gi|2
g7c12a1.r1	325	4.1e-28	18 767	sp|P15365|DAL5_YEAST ALLANTOATE PERMEASE >pir||A28671 allantoate transportprotein - yeast (Saccharomyces cerevi
d4e02a1.r1	180	1.3e-10	13 639	sp|P15365|DAL5_YEAST ALLANTOATE PERMEASE >pir||A28671 allantoate transportprotein - yeast (Saccharomyces cerevi
f.Protein, amino acid transport				
<PROTEIN TRANSPORT PROTEIN>				
z4b10a1.r1	685	9.1e-67	11 649	sp|P15303|SC23_YEAST PROTEIN TRANSPORT PROTEIN SEC23 >pir||BVBY23 proteintransport protein SEC23 - yeast (Sacch
a0d10a1.f1	402	9.1e-37	133 489	sp|Q04491|SC13_YEAST PROTEIN TRANSPORT PROTEIN SEC13 >pir||A45442 SEC13protein - yeast (Saccharomyces cerevisia
Contig1341_e4b02a1.f1	127	0.0002	289 735	sp|P40357|SEC9_Y PROTEIN TRANSPORT PROTEIN SEC9 >pir||A55100 SEC9 protein- yeast (Saccharomyces cerevisiae) >gi|508620
<EXPORT PROTEIN NCE2-protein transport, w/o signal sequence>				
Contig1262_c2c07a1.f1	217	3.5e-17	36 473	sp|Q12207|NCE2_Y NON-CLASSICAL EXPORT PROTEIN NCE2 >pir||S69036 NCE2protein - yeast (Saccharomyces cerevisiae) >gi|106
<PEPTIDE TRANSPORTER>				
Contig973_m6b08a1.f1	337	3.2e-29	130 654	sp|P46031|PT2A_A PEPTIDE TRANSPORTER PTR2-A >gi|575427 (U01171) similar toS. cerevisiae PTR2 gene, GenBank Accession N
m0h08a1.r1	200	2.5e-14	14 466	sp|P46030|PTR2_CANAL PEPTIDE TRANSPORTER PTR2 >gi|806693 (U09781) peptidetransporter [Candida albicans]
<AMINO ACID TRANSPORTER>				
*Contig1165_e0f01a1.r1	307	3.2e-26	18 587	sp|P51906|EAT3_M EXCITATORY AMINO ACID TRANSPORTER 3 (SODIUM-DEPENDENTGLUTAMATE/ASPARTATE TRANSPORTER 3) (EXCITATORY A
<AMINO-ACID PERMEASE>				
15b08a1.f1	276	9.8e-23	130 420	sp|Q09887|YAHB_SCHPO HYPOTHETICAL AMINO-ACID PERMEASE C8A4.11 >pir||S62527hypothetical protein SPAC8A4.11 - fis
w7b01a1.f1	185	8.8e-13	198 497	sp|P38971|ALP1_YEAST BASIC AMINO-ACID PERMEASE >pir||S60912 probable transportprotein ALP1 - yeast (Saccharomyc
p0h04a1.r1	154	2.1e-09	132 461	sp|P36029|YKR4_YEAST HYPOTHETICAL AMINO-ACID PERMEASE IN STE3-GIN10 INTERGENICREGION >pir||S38004 probable tran
<ARGININE PERMEASE>				
w7b01a1.r1	438	1.4e-40	6 530	sp|P04817|CAN1_YEAST ARGININE PERMEASE >pir||QRBYPR arginine transport protein- yeast (Saccharomyces cerevisiae
<importin beta-4 subunit, nuclear protein import>				
Contig1370_c4a02a1.f1	432	7.2e-39	47 760	gnl|PID|e1295815 (AL023780) putative importin beta-4 subunit[Schizosaccharomyces pombe]

g.mitochondrial transport

<MITOCHONDRIAL PROTEIN IMPORT PROTEIN 2>

Contig1606_e9d06a1.f1	588	1.8e-56	252	698	sp|P39515|IM17_Y	MITOCHONDRIAL IMPORT INNER MEMBRANE TRANSLOCASE SUBUNITTIM17 (MITOCHONDRIAL PROTEIN IMPORT PROTEIN 2)

<MITOCHONDRIAL IMPORT RECEPTOR SUBUNIT TOM22-translocation of cytosolic proteins into mitochondria>

f2f07a1.r1	185	1e-13	200	535	sp|Q07335|OM22_NEUCR	MITOCHONDRIAL IMPORT RECEPTOR SUBUNIT TOM22(MITOCHONDRIAL 22 KD OUTER MEMBRANE PROTEIN) =N. crassa

<mitochondrial transport protein amc-1>

Contig1668_c4b09a1.f1	418	1.9e-38	171	521	gi|1621438	(U71603) mitochondrial transport protein amc-1 [Emericellanidulans]
g5b11a1.r1	272	5.3e-23	363	542	gi|1621438	(U71603) mitochondrial transport protein amc-1 [Emericellanidulans]

<MITOCHONDRIAL 2-OXOGLUTARATE/MALATE CARRIER>

r5e09a1.r1	353	1.5e-31	86	703	pir||S65040	2-oxoglutarate/malate translocator (clone OMT134), mitochondrialmembrane - proso -millet

<benzodiazepine receptor-TRANSPORT OF PORPHYRINS AND HEME, mitochondria>

v3c06a1.r1	173	1.7e-12	63	482	pir||A53405	peripheral-type benzodiazepine receptor 1 isoquinoline-bindingprotein - mouse =MITOCHONDRIAL

<ADP,ATP CARRIER PROTEIN>

Contig1208_s9c05a1.f1	*781*	*5e-77*	*72*	*626*	*sp|P02723|ADT_NE*	ADP,ATP CARRIER PROTEIN (ADP/ATP TRANSLOCASE) (ADENININENUCLEOTIDE TRANSLOCATOR) (ANT) >pir||XWNC ADP,A
Contig1293_i8a04a1.r1	580	1.2e-55	258	650	sp|Q09188|ADT_SC	ADP,ATP CARRIER PROTEIN (ADP/ATP TRANSLOCASE) (ADENININENUCLEOTIDE TRANSLOCATOR) (ANT) >gnl|PID|e186767

h.ABC transporter family

<ABC transporter>

Contig1663_d5h04a1.f1	1385	6.3e-141	311	2029	sp|P36619|PMD1_S	LEPTOMYCIN B RESISTANCE PROTEIN PMD1 >pir||S20548leptomycin B resistance protein - fission yeast(Schi
Contig1432_c3a08a1.r1	765	5.4e-74	2	883	gi|2625138	(AF032443) ABC1 transporter; ABC-type ATPase [Magnaporthe grisea]
Contig1558_c9a02a1.f1	709	2.6e-69	363	893	sp|P40024|YEM6_Y	PROBABLE ATP-DEPENDENT TRANSPORTER YER036C >pir||S50539hypothetical protein YER036c - yeast (Saccharo
c9a02a1.r1	705	6.8e-69	13	540	sp|P40024|YEM6_YEAST	PROBABLE ATP-DEPENDENT TRANSPORTER YER036C >pir||S50539hypothetical protein YER036c - yeas
*Contig1477_s3f01a1.f1	527	1.4e-48	244	792	gi|2625138	(AF032443) ABC1 transporter; ABC-type ATPase [Magnaporthe grisea]
f0d04a1.r1	500	1.1e-45	10	600	gi|2625138	(AF032443) ABC1 transporter; ABC-type ATPase [Magnaporthe grisea]
Contig1000_c9e04a1.f1	372	3.4e-32	6	440	gnl|PID|e1316128	(AL031307) leptomycin B resistance protein, ABC transporter[Schizosaccharomyces pombe]
Contig428_e4a06a1.r1	309	6.4e-26	57	707	gnl|PID|e1285355	(AL022299) ABC transporter [Schizosaccharomyces pombe]
*Contig1647_c3a02a1.f1	239	4.8e-16	900	1337	gi|2622773	(AE000923) ABC transporter [Methanobacterium thermoautotrophicum]

d3g08a1.f1	169	1.1e-10	155 412	sp|P25371|ADP1_YEAST PROBABLE ATP-DEPENDENT PERMEASE PRECURSOR >pir||S19421ATP-dependent permease ADP1 precurso
<ATP-DEPENDENT PERMEASE>				
Contig464_d3g08a1.r1	412	1.1e-36	235 729	sp|P25371|ADP1_Y PROBABLE ATP-DEPENDENT PERMEASE PRECURSOR >pir||S19421ATP-dependent permease ADP1 precursor - yeast (
*Contig1339_g3c12a1.r1	255	1.1e-19	4 258	sp|P51533|PDRA_Y ATP-DEPENDENT PERMEASE PDR10 >pir||S55517 probabletransport protein PDR10 - yeast (Saccharomyces cere

i.other

<transport protein>				
Contig1361_c5g12a1.f1	506	8.4e-48	360 908	gi|409547 (L07492) sugar transport protein [Saccharomyces cerevisiae]
<aquaporin>				
c5f02a1.r1	284	2.8e-24	172 744	sp|P53386|AQPL_YEAST AQUAPORIN-LIKE PROTEIN YPR192W >pir||S58822 probablemembrane protein YPR192w - yeast (Sacc

D. Classes of Enzymes-No pathway specified

1. Oxidoreductases (24)

<OXIDOREDUCTASE>				
Contig884_y8f01a1.f1	682	2e-66	101 964	gi|2407193 (AF017151) oxidoreductase [Aspergillus parasiticus]
Contig1472_e0e10a1.r1	668	5.5e-65	48 917	sp|O07575|YHDF_B HYPOTHETICAL OXIDOREDUCTASE IN CITA-SSPB INTERGENICREGION >gnl|PID|e1191878 (Y14082) hypothetical pro
Contig1579_d5d03a1.f1	387	3.2e-35	82 492	sp|P11943|ACPM_N ACYL CARRIER PROTEIN, MITOCHONDRIAL PRECURSOR (ACP)(NADH-UBIQUINONE OXIDOREDUCTASE 9.6 KD SUBUNIT) >p
Contig1776_a1g04f2.f1	376	5.1e-34	37 660	sp|Q09851|YAEB_S HYPOTHETICAL OXIDOREDUCTASE C23D3.11 IN CHROMOSOME I>pir||S62502 hypothetical protein SPAC23D3.11 - f
Contig465_d3f01a1.f1	335	1.1e-29	176 703	sp|Q39172|P1_ARA PROBABLE NADP-DEPENDENT OXIDOREDUCTASE P1 >pir||S57611zeta-crystallin homolog - Arabidopsis thaliana
Contig1418_c8g11a1.f1	307	1.1e-26	584 1117	sp|P76113|YNCB_E PUTATIVE NADP-DEPENDENT OXIDOREDUCTASE IN TEHB-RHSEINTERGENIC REGION >gnl|PID|d1015800 (D90784) Possi
d3f01a1.r1	297	1.2e-25	194 652	dbj||D86417_36 (D86417) YfmJ [Bacillus subtilis] >gnl|PID|e1182735 (Z99108)similar to quinone oxidoreduct=PROBABLE NADP-DEPENDENT OXIDOREDUCTASE
*Contig1393_k9c01a1.f1	247	3.7e-20	365 913	sp|P25145|YINL_L HYPOTHETICAL OXIDOREDUCTASE IN INLA 5'REGION (ORFA)>gi|149673 (M67471) ORFA [Listeria monocytogenes]
Contig1296_g5h07a1.f1	203	1.6e-15	266 760	gnl|PID|e1287847 (AL022603) putative NADPH quinone oxidoreductase [Arabidopsisthaliana]
Contig1046_c0e04a1.f1	163	2.1e-11	120 569	sp|P42317|YXJF_B HYPOTHETICAL OXIDOREDUCTASE IN PEPT-KATB INTERGENICREGION >gnl|PID|d1012374 (D83026) homologous to ma
Contig1119_d5c07a1.r1	156	2.9e-10	124 627	gi|3293547 (AF072709) putative oxidoreductase [Streptomyces lividans]

<MONOOXYGENASE>

g8g03a1.r1	271	5.9e-22	30	650	sp|P55487|Y4ID_RHISN PROBABLE MONOOXYGENASE Y4ID >gi|2182441 (AE000078) Y4iD[Rhizobium sp. NGR234]
p0f05a1.r1	181	1.8e-12	179	430	sp|P17549|CP53_ASPNG BENZOATE 4-MONOOXYGENASE (BENZOATE-PARA-HYDROXYLASE)>pir||S12015 benzoate 4-monooxygenase

<cytochrome P450>

i8e02a1.r1	436	2.3e-40	9	482	gnl|PID|e339903 (Z98974) putative cytochrome p450 [Schizosaccharomyces pombe]
f5b05a1.f1	238	1.3e-18	211	729	pir||S57337 trichodiene oxygenase 4 - fungus (Fusarium sporotrichioides)>gi|837032 (U22462)= tri4 gene is p450 monooxygenase of trichothecene biosyn
g6d12a1.f1	160	3.5e-08	186	476	sp|P48416|CP10_LYMST CYTOCHROME P450 X >pir||JX0225 cytochrome P450 CYP10 -great pond snail >bbs|115322 (S46130

<CR(VI) REDUCTASE-FLAVIN OXIDOREDUCTASE family>

i2h01a1.f1	187	8e-14	21	452	sp|P96977|CHRR_PSESP CR(VI) REDUCTASE >gnl|PID|d1012488 (D83142) Cr(VI)reductase [Pseudomonas sp.]

<AMINE OXIDASE-OXIDATIVE DEAMINATION of AMINES>

k0g09a1.f1	516	6.9e-49	37	522	sp|Q12556|AMO1_ASPNG COPPER AMINE OXIDASE 1 >gi|1401157 (U31869) copper amineoxidase [Aspergillus niger]
k0g09a1.r1	*310*	*3.8e-26*	*136*	*396*	*sp|Q12556|AMO1_ASPNG* COPPER AMINE OXIDASE 1 >gi|1401157 (U31869) copper amineoxidase,HISTAMINE OXIDASE [Aspergillus niger]

<SEXUAL DIFFERENTIATION PROCESS PROTEIN-expressed during sexual diff in S. pombe>

m0e12a1.r1	113	3e-05	157	414	sp|P40902|ISP7_SCHPO SEXUAL DIFFERENTIATION PROCESS PROTEIN ISP7 >pir||S45496isp7 protein - fission yeast (Schi

<fructosyl amine:oxygen oxidoreductase>

n5a03a1.r1	561	1.3e-53	33	692	gi|2661130 (AF035700) fructosyl amine:oxygen oxidoreductase [Aspergillusfumigatus]

<chlorocatechol 1,2-dioxygenase-degradation>

k8a07a1.f1	211	1.6e-16	96	410	gnl|PID|d1013794 (D86544) hydroxyquinol-1, 2-dioxygenase [Ralstonia pickettii]
Contig489_o0g05a1.f1	159	4.5e-10	2	301	gi|2318013 (AF011355) isotrichodermin C-15 hydroxylase [Fusariumsporotrichioides]

<HYDROXYACID DEHYDROGENASE>

Contig1209_g7g12a1.f1	286	1.8e-24	275	604	sp|P30799|DDH_ZY 2-HYDROXYACID DEHYDROGENASE HOMOLOG >pir||D40649D-2-hydroxy-acid dehydrogenase (EC 1.1.99.6) - Zymomo

<SOL1 Protein>

Contig1090_o6d10a1.r1	502	2.3e-47	173	841	sp|P50278|SOL1_Y SOL1 PROTEIN >pir||S62015 SOL1 protein - yeast(Saccharomyces cerevisiae) >gi|1163192 (U43608) Sol1p[S

2. Transferases (1)

<ORNITHINE AMINOTRANSFERASE>

Contig1762_c1a08a1.f1	2172	2.4e-224	299	1660	sp|Q92413|OAT_EM ORNITHINE AMINOTRANSFERASE

(ORNITHINE--OXO-ACIDAMINOTRANSFERASE) >gi|1658173 (U74303) ornithine trans

3.Hydrolases (1)
<alkaline phosphatase>
j0g12a1.r1 369 5.3e-33 13 447 sp|P11491|PPB_YEAST REPRESSIBLE ALKALINE PHOSPHATASE PRECURSOR >pir||S69648alkaline phosphatase (EC 3.1.3.1) -

4. Lyases (0)

5. Isomerases (0)

6. Ligases (0)

7. Synthetases (2)
<S-adenosylmethionine synthetase>
z6d05a1.r1 401 1.1e-36 232 675 sp|P48466|METK_NEUCR S-ADENOSYLMETHIONINE SYNTHETASE (METHIONINEADENOSYLTRANSFERASE) (ADOMET SYNTHETASE) >pir||
Contig1039_s4b01a1.f1 264 4.2e-22 365 544 sp|P48466|METK_N S-ADENOSYLMETHIONINE SYNTHETASE (METHIONINEADENOSYLTRANSFERASE) (ADOMET SYNTHETASE) >pir||S65800 meth

E. Non-enymatic classes (not in defined pathways)

1. zinc finger motif-DNA binding (11)
<zinc finger protein>
n3g08a1.r1 570 1.4e-54 63 566 gnl|PID|e254304 (X99094) zinc finger protein [Ascobolus immersus]
z7f05a1.f1 331 3e-29 127 603 gi|1438877 (U41287) zinc finger protein [Mus musculus]
x9f12a1.r1 185 3.4e-13 215 436 gnl|PID|e223435 (X95455) RING zinc finger protein [Gallus gallus]>prf||2211437A RING finger protein [Gallu
Contig517_c8g02a1.f1 201 4.8e-13 281 997 gnl|PID|e1291640 (AL023288) Zinc finger protein [Schizosaccharomyces pombe]
v3f10a1.f1 174 1e-11 209 532 sp|P38682|GLO3_YEAST ZINC FINGER PROTEIN GLO3 >pir||S50625 GLO3 protein -yeast (Saccharomyces cerevisiae) >gi|6
z1h04a1.f1 157 8.3e-11 294 557 gnl|PID|d1021201 (D45213) zinc finger protein [Homo sapiens]
Contig1196_e9a09a1.f1 175 4.5e-10 318 626 sp|P34670|YO14_C HYPOTHETICAL ZINC FINGER PROTEIN ZK686.4 IN CHROMOSOMEIII >pir||S44909 ZK686.4 protein - Caenorhabdit
z7d04a1.f1 154 1.8e-09 199 348 gnl|PID|e1291655 (AL023290) zinc finger C3HC4 type [Schizosaccharomyces pombe]
*Contig749_y3d09a1.f1 143 2.1e-07 488 676 gi|3033395 (AC004238) putative zinc-finger protein [Arabidopsis thaliana]
o0f10a1.r1 120 7.6e-06 78 191 gnl|PID|e1316877 (AL031349) zinc-finger protein [Schizosaccharomyces pombe]
<KIN17 protein-crossreacts w recA antibody>
n8e10a1.r1 225 5.2e-18 25 426 prf||1713233A recA crossreacting protein KIN17 [Mus musculus]

2. Leucine zipper motif (1)

<leucine zipper>

Contig1642_c7b12a1.f1	140	6.3e-07	524 637	gnl|PID|e253252	(X99215) leucine zipper [Aspergillus niger]

VI. Unclassified (significant homolog but function uncertain in Aspergillus nidulans) (132)

<uncertain function>

g2c02a1.r1	908	2.3e-90	55 588	gi|3411262	(AF080599) Medusa [Emericella nidulans]
Contig1231_r5c11a1.f1	827	8.3e-82	10 921	gnl|PID|d1004374	(D16355) crm1-N1 protein [Schizosaccharomyces pombe]=CHROMOSOME REGION MAINTENANCE PROTEIN 1
o6g08a1.r1	813	2.3e-80	57 533	gi|168081	(M59935) unidentified gene; ORF [Emericella nidulans]
i3e01a1.r1	778	1.4e-76	14 487	sp|P03710|VMCB_LAMBD	PORTAL PROTEIN (GPB) (MINOR CAPSID PROTEIN B)>pir||VHBPBL minor capsid protein B - phage =possible contaminant from Lambda
n0f09a1.r1	777	1.7e-76	11 454	gi|1755051	(U56696) palFp [Emericella nidulans]=pH response
o4a08a1.r1	771	7.4e-76	6 443	sp|P03765|Y146_LAMBD	HYPOTHETICAL NIN REGION PROTEIN ORF-146 >pir||QXBP4Lhypothetical protein C-146 (nin region=lambda contaminant?
Contig1548_d3a07a1.f1	548	1e-75	698 1291	sp|P39743|R167_Y	REDUCED VIABILITY UPON STARVATION PROTEIN 167>pir||S40887 RVS167 protein - yeast (Saccharomyces cerev
Contig1773_d1d09a1.f1	742	8.5e-73	118 1080	gi|2352898	(AF012091) cystein rich protein [Metarhizium anisopliae]
Contig1718_c8g07a1.f1	735	3.1e-71	322 1182	pir||S62011	PHO85 protein - yeast (Saccharomyces cerevisiae) >gi|1163103(U43503) Lph16p [Saccharomyces cerevisiae
x7h06a1.r1	681	2.4e-66	5 607	gi|563253	(L32177) guanine nucleotide regulatory protein [Cryphonectriaparasitica]
Contig434_e0d10a1.f1	654	1.8e-63	226 639	sp|P19815|SC1D_E	CONIDIUM-SPECIFIC PROTEIN >pir||S12113 SpoC1-C1D protein- Emericella nidulans >gi|2419 (X54668) SpoC1
Contig466_d3e06a1.f1	645	1.6e-62	2 373	gi|215150	(J02459) O (DNA replication;299) [Bacteriophage lambda]
m0e10a1.r1	647	6.3e-62	22 459	pir||D45029	crm1+ protein - fission yeast (Schizosaccharomyces pombe)=CHROMOSOME REGION MAINTENANCE PROTEIN
Contig541_c7d04a1.r1	355	1.1e-61	31 498	pir||S51256	probable membrane protein YDR105c - yeast (Saccharomycescerevisiae) >gi|633641 (Z47746) unknown [Sacc
r1h07a1.r1	593	5.2e-57	164 667	sp|Q12499|NOP5_YEAST	NUCLEOLAR PROTEIN NOP5 >pir||S58322 hypothetical proteinYOR310c - yeast (Saccharomyces cer
c3d07a1.r1	561	1.3e-53	58 642	sp|P10962|MK16_YEAST	MAK16 PROTEIN >pir||BVBYK6 MAK16 protein - yeast(Saccharomyces cerevisiae) >gi|171880= may be concerned with nuclear events of cell cycle
z7g06a1.f1	534	9.1e-51	114 602	pir||S71153	het-c4 protein - Podospora anserina >gi|537939 (L36210) het-c[Podospora anserina]
Contig1455_c9f12a1.f1	523	1.2e-49	180 650	gnl|PID|e1285104	(AJ001732) rAsp f 4 [Aspergillus fumigatus]
u4a09a1.r1	475	1.7e-44	191 604	gi|1870215	(AC000133) ORF [Emericella nidulans]
w7c04a1.r1	465	1.8e-43	10 504	pir||JC4516	protein kinase (EC 2.7.1.37) - fission yeast

				(Schizosaccharomycespombe) >gnl|PID|d1008838
o5b02a1.r1	462	4e-43	13 459	sp|P40107|GOG5_YEAST VANADATE RESISTANCE PROTEIN GOG5/VRG4/VAN2 >pir||S50238vanadate resistance protein VAN2 -=REGULATION OF THE PHOSPHORYLATION OF A NUMBER OF PROTEINS
l3b09a1.r1	461	5.2e-43	142 594	sp|P33300|SUR1_YEAST SUR1 PROTEIN >pir||S41798 SUR1 protein - yeast(Saccharomyces cerevisiae) >gi|311163 (M9664
k5g12a1.r1	459	7.4e-43	121 525	sp|Q39963|ER1_HEVBR ETHYLENE-INDUCIBLE PROTEIN HEVER >pir||S60047ethylene-responsive protein 1 - Para rubber t
Contig537_c8a02a1.r1	431	7.2e-40	252 569	gnl|PID|e1250610 (AJ223315) rAsp f 7 [Aspergillus fumigatus
v8h01a1.r1	405	4.4e-37	8 829	sp|P38789|SSF1_YEAST SSF1 PROTEIN >pir||S46700 SSF1 protein - yeast(Sacoharomyces cerevisiae) >gi|487948 (U0006=possible role in mating
c4c08a1.r1	366	5.3e-33	169 603	gnl|PID|d1013767 (D86381) Ran/spi1 binding protein [Schizosaccharomyces pombe]=may regulate transport of protein across nuclear membrane as part of cell cycle
Contig911_y8b01a1.f1	365	7.2e-33	216 857	sp|Q13610|PWP1_H PERIODIC TRYPTOPHAN PROTEIN 1 HOMOLOG (KERATINOCYTEPROTEIN IEF SSP 9502) >gi|177765 (L07758) IEF SSP
Contig173_i8h09a1.f1	364	9.4e-33	205 474	gi|348156 (L16844) yps-3 [Histoplasma capsulatum]=yeast-phase-specific gene
Contig686_o9f10a1.f1	241	7.2e-32	454 642	pir||S49326 Nascent polypeptide associated complex alpha chain - human>gi|556642 (X80909) Nascent polypeptide ass
m7f02a1.f1	354	1.1e-31	29 622	sp|P25621|YCR8_YEAST PUTATIVE TRANSPORTER YCR28C >pir||S19439 probablemembrane protein YCR028c - yeast (Sacchar
m7f02a1.r1	348	5.9e-31	236 607	gi|2981103 (AF052688) putative transmembrane transporter Liz1p[Schizosaccharomyces pombe] >gnl|PID|e1
Contig1553_r1g04a1.f1	344	1.2e-30	324 836	gi|603050 (U18061) CAP20 [Glomerella cingulata]
Contig69_m3d06a1.r1	351	3.4e-30	76 576	gi|1261823 (L77234) glycine rich protein [Neurospora crassa]
y7h02a1.f1	342	8.6e-30	83 691	sp|P90587|WD66_PHYPO 66 KD STRESS PROTEIN (P66) >gi|1835727 (U86011) 66-kDastress protein p66 [Physarum polycep
b0e11a1.r1	335	1.1e-29	13 270	sp|P78712|ARP3_NEUCR ACTIN-LIKE PROTEIN 3 >gi|1718497 (U79737) actin-relatedprotein 3 [Neurospora crassa]
i7f03a1.f1	333	1.8e-29	151 627	sp|Q10494|YDG7_SCHPO PROBABLE OXIDOREDUCTASE C26F1.07 IN CHROMOSOME I>gnl|PID|e241770 (Z73100) unknown [Schizos
Contig435_e0c12a1.r1	315	1.1e-27	486 740	gnl|PID|e1285104 (AJ001732) rAsp f 4 [Aspergillus fumigatus]
Contig1793_c3c10a1.f1	310	4.8e-27	88 306	sp|P10713|CONX_N CONIDIATION-SPECIFIC PROTEIN 10 >pir||A31849conidiation-specific protein - Neurospora crassa >gi|1687
Contig420_n8a04a1.f1	305	4.2e-26	247 744	sp|Q10097|YAOI_S PUTATIVE TRANSPORTER C11D3.18C >gi|1107907 (Z68166)unknown [Schizosaccharomyces pombe]
c7e10a1.r1	292	4.2e-25	363 758	gi|1326076 (U28150) adrenoleukodystrophy related protein [Homo sapiens]
w9d08a1.r1	299	5e-25	84 440	gnl|PID|e1263921 (AL022104) kinase-binding protein 1 [Schizosaccharomycespombe]
y9h11a1.r1	290	7e-25	56 532	sp|P53859|YNX2_YEAST HYPOTHETICAL 31.6 KD PROTEIN IN SIN4-URE2 INTERGENICREGION >pir||S63198 hypothetical prote

i0c11a1.r1	288	1e-24	11 454	gi|1845332 (U60450) putative ATPase [Gigaspora rosea]
Contig951_c8b09a1.f1	176	5.3e-24	203 418	sp|Q02336|ADA2_Y POTENTIAL TRANSCRIPTIONAL ADAPTOR >pir||A43252 probabletranscriptional adaptor ADA2 - yeast (Saccharo
r7h01a1.r1	292	9.4e-24	412 726	gnl|PID|e325416 (Z97211) probable involvement in ergosterol synthesis[Schizosaccharomyces pombe]
v3g01a1.r1	270	8.6e-23	128 562	gnl|PID|d1031084 (AP000004) 387aa long hypothetical amidohydrolase [Pyrococcushorikoshii]
c0g09a1.f1	159	8.9e-23	369 575	sp|Q92377|MD12_SCHPO MITOCHONDRIAL INHERITANCE COMPONENT MDM12 >gi|1655884(U64674) required for mitochondrial inheritance in budding and fission yeast
w9a03a1.r1	279	1.6e-22	40 513	sp|P26674|STE6_SCHPO STE6 PROTEIN >pir||S28098 ste6 protein - fission yeast(Schizosaccharomyces pombe) >gi|5101
e7f02a1.r1	259	1.3e-21	202 489	gnl|PID|e1175783 (AJ002894) OsGRP2 [Oryza sativa]=RNA binding protein
g1g12a1.r1	264	7.3e-21	42 509	pir||S55945 STE23 protein - yeast (Saccharomyces cerevisiae) >gi|625109(U19729) Ste23p similar to insulinprotease
c5f04a1.r1	247	2.3e-20	67 423	sp|P32495|NHP2_YEAST HIGH MOBILITY GROUP-LIKE NUCLEAR PROTEIN 2 >pir||S67767high mobility group-like protein NH
i7f03a1.r1	247	2.4e-20	42 305	sp|Q12458|YPR1_YEAST PUTATIVE REDUCTASE 1 >pir||S61163 YPR1 protein - yeast(Saccharomyces cerevisiae) >gi|84918
c6c10a1.r1	251	6e-20	14 460	gnl|PID|e339034 (Z98762) hypothetical acetyl hydrolase [Schizosaccharomycespombe]=enzyme involved in antibiotic bialaphos biosyn
Contig913_c6h06a1.r1	244	8.7e-20	76 456	sp|P41890|SCN1_S SCN1 PROTEIN >pir||B55164 scn1 protein - fission yeast(Schizosaccharomyces pombe) >gnl|PID|d1007203 (
z6h10a1.r1	182	8.9e-20	186 542	sp|Q09698|YA27_SCHPO HYPOTHETICAL 68.8 KD PROTEIN C2F7.07C IN CHROMOSOME I>pir||S58151 hypothetical protein SPA
f5g01a1.f1	232	2.3e-18	154 525	sp|P53693|RDS1_SCHPO RDS1 PROTEIN >pir||S58477 rds1 protein - fission yeast(Schizosaccharomyces pombe) >gnl|PID
l0f04a1.f1	226	3.8e-18	252 443	gnl|PID|e1216790 (AL021046) probable involvement in transcription initiation[Schizosaccharomyces pombe]
t2c06a1.r1	230	8.9e-18	36 440	gi|3249066 (AC004473) Similar to S. cerevisiae SIK1P protein gb|984964. ESTsgb|F15433 and gb|AA395158
o8h04a1.f1	230	1.3e-17	269 706	gi|517205 (U09352) 67 kDa Myosin-crossreactive streptococcal antigen[Streptococcus pyogenes]
h1h05a1.r1	219	2.1e-17	74 547	gnl|PID|e1231246 (AJ001261) NIPSNAP2 protein [Mus musculus]
g0a04a1.f1	221	2.2e-17	98 529	gnl|PID|e1293292 (AL023592) rna binding protein [Schizosaccharomyces pombe]
c6d11a1.r1	221	6.2e-17	118 780	gnl|PID|e1314296 (AL031182) putative lipoprotein [Streptomyces coelicolor]
w8g04a1.r1	219	2.1e-16	19 576	gi|531469 (U12973) renal osmotic stress-induced Na-Cl organic solutecotransporter [Rattus norvegicus
Contig1276_j7b11a1.f1	215	2.6e-16	112 477	sp|P32783|ABD1_Y ABD1 PROTEIN >pir||S41782 ABD1 protein - yeast(Saccharomyces cerevisiae) >gi|170966 (L12000) ABD1 [Sa
g4d03a1.r1	205	2.1e-15	38 562	sp|Q15392|DIMH_HUMAN DIMINUTO-LIKE PROTEIN >gnl|PID|d1003311 (D13643)

				KIAA0018[Homo sapiens]
Contig118_k0e09a1.f1	211	4.3e-15	84 287	gi|2804455 (AF043699) similar to a human orf (GB:D13642) and human UV-damagedDNA binding factor (GB:U32986) in s
z5e10a1.r1	206	6.9e-15	23 616	sp|P53946|ARP5_YEAST ACTIN-LIKE PROTEIN ARP5 >pir||S58718 probable nuclearprotein YNL059c - yeast (Saccharomyce
m0f05a1.f1	120	7.5e-15	68 196	pir||S33788 luciferase - southern Russian firefly >bbs|133113 (S61961)luciferase [Luciola mingrelica=E
x9d07a1.r1	197	2.4e-14	13 486	sp|P32832|NPL6_YEAST NPL6 PROTEIN >pir||S30792 NPL6 protein - yeast(Saccharomyces cerevisiae) >gi|172050 (M9843=possible involvement in nuclear protein localisation
h1h05a1.f1	196	2.4e-14	352 801	gnl|PID|e1231231 (AJ001258) NIPSNAP1 protein [Homo sapiens]
Contig1121_u4c11a1.r1	208	6.1e-14	527 931	sp|P21560|CBP3_Y CBP3 PROTEIN PRECURSOR >pir||A34290 membrane protein CBP3- yeast (Saccharomyces cerevisiae) >gi|17117
Contig222_i0b03a1.f1	184	4.8e-13	273 587	sp|P41890|SCN1_S SCN1 PROTEIN >pir||B55164 scn1 protein - fission yeast(Schizosaccharomyces pombe) >gnl|PID|d1007203 (
i3g06a1.r1	191	7.1e-13	29 433	sp|P40340|TBP7_YEAST TAT-BINDING HOMOLOG 7 >pir||S64603 YTA7 protein - yeast(Saccharomyces cerevisiae) >gnl|PID
r7f01a1.r1	185	7.8e-13	389 766	sp|P40445|YIQ6_YEAST PUTATIVE TRANSPORTER YIL166C >pir||S50361 probablemembrane protein YIL166c - yeast (Saccha
Contig693_x8g03a1.r1	*184*	*1.1e-12*	*160 561*	*gnl|PID|d1012479 (D83125) secretory component [Sarcophaga peregrina]*
d0e03a1.f1	174	1.4e-12	193 435	sp|P36149|YK48_YEAST HYPOTHETICAL 22.1 KD PROTEIN IN CCP1-MET1 INTERGENICREGION >pir||S38144 =bet3 homolog-
g4c09a1.r1	186	2e-12	31 453	sp|P15904|PCR_AVESA PROTOCHLOROPHYLLIDE REDUCTASE (PCR)(NADPH-PROTOCHLOROPHYLLIDE OXIDOREDUCTASE) >pir||S08406electron transport in chlorophyll biosynthesis?
f0h04a1.r1	166	2.4e-11	219 554	gnl|PID|d1002364 (D11111) chloroplast 33 kd ribonucleoprotein (cp33)[Nicotiana sylvestris] >gi|311952 (X583
g3f09a1.r1	174	2.6e-11	30 422	sp|P40528|SYG1_YEAST SYG1 PROTEIN >pir||S49931 SYG1 protein - yeast(Saccharomyces cerevisiae) >gi|600001 (Z4686
c1f03a1.f1	169	6.6e-11	109 564	gi|2367392 (U82513) random slug cDNA25 protein [Dictyostelium discoideum]
q0b09a1.r1	166	7.3e-11	58 354	gnl|PID|e1316378 (AJ010169) Ariadne-2 protein [Drosophila melanogaster]
z6h07a1.r1	183	7.6e-11	50 538	gi|1098491 (U12141) Ynl2515p [Saccharomyces cerevisiae]
n3b09a1.r1	163	1.2e-10	365 589	sp|Q02336|ADA2_YEAST POTENTIAL TRANSCRIPTIONAL ADAPTOR >pir||A43252 probabletranscriptional adaptor ADA2 - yeas
g8b01a1.r1	184	1.4e-10	25 339	gi|2109297 (U97696) cyclooxygenase-2 [Oryctolagus cuniculus]required for prostiglandin synthesis
Contig1817_a1h05c9.r1	180	9.6e-10	752 1237	sp|Q03465|SON1_Y NUCLEAR PROTEIN SON1 (UB FUSION DEGRADATION PROTEIN 5)>pir||S41986 nuclear protein SON1 - yeast (Sacc
h1g12a1.r1	154	2.3e-09	109 276	gnl|PID|d1018112 (D90905) rehydrin [Synechocystis sp.]
Contig1217_n2g11a1.r1	141	3.9e-09	215 418	gi|3168840 (U88711) copper homeostasis factor [Arabidopsis thaliana]
r1e08a1.r1	142	7.2e-09	335 547	gnl|PID|e1316908 (AL031350) putative dehydrogenase [Streptomyces

				coelicolor]
i0a07a1.f1	142	8.2e-09	126 503	gnl|PID|d1016518 (D90835) H-NS-repressed protein, 30K [Escherichia coli]
c5d12a1.r1	144	3.6e-07	24 158	sp|P19541|YP33_YEAST PUTATIVE TRANSCRIPTIONAL REGULATORY PROTEIN IN MKK2-COX11INTERGENIC REGION >pir||S69051 hy
v8g01a1.r1	132	6.8e-07	7 150	sp|P52977|LON_CAUCR ATP-DEPENDENT PROTEASE LA >gi|1667399 (U56652) lonprotease [Caulobacter crescentus]-DNA methylation control
Contig1298_j4e02a1.f1	118	1e-06	152 307	pir||A54523 histidine-rich protein - Plasmodium lophurae (fragment) >gi|552196(M15317) histidine-rich protein [Pl
w7c04a1.f1	142	1.6e-06	430 570	pir||JC4516 protein kinase (EC 2.7.1.37) - fission yeast (Schizosaccharomycespombe) >gnl|PID|d1008838
z6h04a1.f1	138	2.2e-06	44 349	gnl|PID|e1252031 (AL021899) hypothetical protein Rv2052c [Mycobacteriumtuberculosis]
r2d07a1.r1	138	2.7e-06	427 681	sp|P54069|BE46_SCHPO BEM46 PROTEIN >gi|987287 (U29892) temperature sensitivesupressor of Saccharomyces cerevisi
Contig1532_d2g07a1.r1	135	3.9e-05	48 299	gi|172672 (M13629) sporulation protein [Saccharomyces cerevisiae]
Contig1109_u4a07a1.r1	126	4.1e-05	357 626	gnl|PID|e236571 (X96977) cell wall anchoring signal [Enterococcus faecalis]
v4g04a1.r1	127	8.9e-05	337 585	gi|1572821 (U70856) weak similarity to rat cytosolic acyl coenzyme A thioesterhydrolase (GB:U49694)C. elegans
Contig412_e9e05a1.r1	120	0.00024	309 452	sp|P53081|NIF3_Y NGG1-INTERACTING FACTOR 3 >pir||S64243 hypotheticalprotein YGL221c - yeast (Saccharomyces cerevisiae)
<bacteriorhodopsin>				
Contig1753_c8f10a1.f1	180	2.1e-11	367 1116	sp|P02945|BACR_H BACTERIORHODOPSIN PRECURSOR (BR) >pir||RAHSBbacteriorhodopsin precursor - Halobacterium halobium >pir
<LEUKOTRIENE biosynthesis>				
Contig246_h1g06a1.r1	723	9.1e-71	3 680	sp|Q10740|LKHA_Y PROBABLE LEUKOTRIENE A-4 HYDROLASE (LTA-4 HYDROLASE)(LEUKOTRIENE A(4) HYDROLASE) >pir||S61099 leukotr
Contig247_h1g06a1.f1	172	2.9e-11	193 342	sp|Q10740|LKHA_Y PROBABLE LEUKOTRIENE A-4 HYDROLASE (LTA-4 HYDROLASE)(LEUKOTRIENE A(4) HYDROLASE) >pir||S61099 leukotr
<bleomycin hydrolase>				
Contig476_d3b04a1.f1	499	4.8e-47	51 884	sp|Q13867|BLMH_H BLEOMYCIN HYDROLASE (BLM HYDROLASE) (BMH)>gnl|PID|e205512 (X92106) bleomycin hydrolase [Homo sapiens]
<transposase>				
*Contig1404_g5h09a1.f1	294	7e-24	503 898	pir||S60179 pol polyprotein homolog - fungus (Fusarium oxysporum)retrotransposon skippy >gi|510697 (L34658) pol p
w4c11a1.r1	201	2.9e-15	6 155	bbs|175409 (S80872) putative Tc1-mariner class transposase/IS630-Tc1 homolog[Aspergillus niger, chlor
g9e10a1.f1	201	4.6e-14	11 421	gnl|PID|e1273348 (AL022140) LTR retrotransposon like protein [Arabidopsisthaliana]
Contig1184_g6g01a1.r1	189	1.2e-12	271 741	pir||S32437 pol polyprotein - Volvox carteri f. nagariensis retrotransposonOsser >gi|288597 (X69552) gag,protease

<prohibitin>						
y3b11a1.r1	425	1e-68	39	386	gi|2582388	(AF022225) prohibitin [Pneumocystis carinii]
<DIPHTHINE SYNTHASE>						
Contig1021_s9d11a1.f1	531	7.3e-87	351	839	sp|P32469|DPH5_Y	DIPHTHINE SYNTHASE (DIPHTAMIDE BIOSYNTHESISMETHYLTRANSFERASE) >pir||S30890 methyltransferase DPH5 - y
<SpoC1-C1C protein>						
m5f12a1.r1	652	3.1e-63	61	510	pir||S27412	SpoC1-C1C protein - Emericella nidulans >gi|168091 (M83571)SpoC1-C1C [Emericella nidulans]
Contig1034_m5f12a1.f1	527	5.2e-50	127	444	pir||S27412	SpoC1-C1C protein - Emericella nidulans >gi|168091 (M83571)SpoC1-C1C [Emericella nidulans]
<PEPTIDASE>						
Contig1724_c1f11a1.f1	630	6.6e-61	72	857	sp|P43590|YFH6_Y	HYPOTHETICAL 61.8 KD PEPTIDASE IN MPR1-GCN20 INTERGENICREGION >pir||S56261 probable membrane protein
Contig1633_g0h01a1.f1	480	4.5e-45	93	707	gi|1763684	(U81483) pre-pro-penicillopepsin-JT2 [Penicillium janthinellum]=(PEPTIDASE A) >pdb|3APP| Acid Proteinase
Contig400_f0g06a1.f1	280	3.7e-23	463	978	sp|P43590|YFH6_Y	HYPOTHETICAL 61.8 KD PEPTIDASE IN MPR1-GCN20 INTERGENICREGION >pir||S56261 probable membrane protein
Contig1040_c6c03a1.f1	270	8.1e-23	7	312	gi|1763684	(U81483) pre-pro-penicillopepsin-JT2 [Penicillium janthinellum](PEPTIDASE A) >pdb|3APP| Acid Proteinase
Contig1328_d4b06a1.f1	270	8.2e-23	7	312	gi|1763684	(U81483) pre-pro-penicillopepsin-JT2 [Penicillium janthinellum]=(PEPTIDASE A) >pdb|3APP| Acid Proteinase
Contig1580_c1f11a1.r2	264	1.9e-21	36	647	sp|P43590|YFH6_Y	HYPOTHETICAL 61.8 KD PEPTIDASE IN MPR1-GCN20 INTERGENICREGION >pir||S56261 probable membrane protein
<regulatory protein>						
Contig395_f1d08a1.r1	235	4.6e-19	50	439	pir||A57145	regulatory protein prrC - Rhodobacter sphaeroides >gi|733128(U22347) membrane-anchored regulatory pro
Contig1020_c9h08a1.f1	165	2.6e-09	62	244	sp|P28348|NIRA_E	NITROGEN ASSIMILATION TRANSCRIPTION FACTOR NIRA>pir||A41697 nitrate assimilation regulatory protein n
Contig301_g7a11a1.f1	137	3.4e-06	390	593	pir||A61382	phosphorylation regulatory protein HP-10 - human
<MAL3 PROTEIN>						
Contig1424_c8b11a1.f1	372	3.3e-49	166	516	sp|Q10113|MAL3_S	MAL3 PROTEIN >gnl|PID|e213819 (Z68198) unknown[Schizosaccharomyces pombe] >gnl|PID|e282230 (Y09518) M
<GRR1 PROTEIN-GLUCOSE REPRESSION PATHWAY>						
Contig1516_c9h12a1.r1	734	2.9e-71	22	891	sp|P24814|GRR1_Y	GRR1 PROTEIN >pir||A41529 GRR1 protein - yeast(Saccharomyces cerevisiae) >gi|171617 (M59247) putative
<PIG8>						
Contig1742_g2b04a1.f1	326	9.5e-29	203	985	gi|1764133	(U81790) PIG8 [Uromyces fabae]
<PHO85 protein>						
Contig516_c8g07a1.r1	211	4.3e-15	37	402	pir||S62011	PHO85 protein - yeast (Saccharomyces cerevisiae) >gi|1163103(U43503) Lph16p [Saccharomyces cerevisiae
<Argonaute protein>						
Contig664_o8h09a1.f1	274	6.8e-22	215	715	gi|2149640	(U91995) Argonaute protein [Arabidopsis thaliana]

<BC-2 protein-putative breast adenocarcinoma marker>
Contig705_p0c12a1.f1 322 2.6e-28 325 729 gi|2828147 (AF042384) BC-2 protein [Homo sapiens]
<POS5 protein>
Contig787_s8a06a1.f1 321 3.5e-28 345 704 pir||S65200 POS5 protein - yeast (Saccharomyces cerevisiae) >gnl|PID|e247057(Z73544) ORF YPL188w [Saccharomyces c
<G10 PROTEIN>
Contig979_j9h07a1.f1 384 4.4e-48 122 454 sp|P12805|G10_XE G10 PROTEIN >pir||S05955 G10 protein - African clawed frog>gi|64704 (X15243) G10 protein (AA 1-144) [
<COLD SHOCK PROTEIN>
Contig1802_d1e12a1.f1 146 1e-09 147 338 sp|P95459|CSPA_P MAJOR COLD SHOCK PROTEIN CSPA >gi|1778825 (U82822) majorcold shock protein CspA [Pseudomonas aerugino
<GLUCOSE-REPRESSIBLE GENE PROTEIN>
Contig1809_c1a06a1.f1 199 2.4e-15 68 280 sp|P22151|GRG1_N GLUCOSE-REPRESSIBLE GENE PROTEIN >gi|3014 (X14801) grg1gene product [Neurospora crassa]
Contig1843_a5a07a1.f1 177 5.6e-13 87 293 sp|P22151|GRG1_N GLUCOSE-REPRESSIBLE GENE PROTEIN >gi|3014 (X14801) grg1gene product [Neurospora crassa]
Contig1859_a1a06f2.f1 177 5.6e-13 138 344 sp|P22151|GRG1_N GLUCOSE-REPRESSIBLE GENE PROTEIN >gi|3014 (X14801) grg1gene product [Neurospora crassa]
<EF hand protein>
Contig630_o8c05a1.f1 307 1.1e-26 210 599 gi|2459421 (AC002332) putative calcium-binding EF-hand protein [Arabidopsisthaliana]

VII. Unidentified (includes significant match with ORFs) (366)
<unknown function>
Contig557_c6b02a1.r1 1058 2.7e-106 1 1203 gnl|PID|d1022254 (AB004535) hypothetical protein YPR112c [Schizosaccharomycespombe]
Contig1626_c4c09a1.f1 945 2.5e-94 178 954 gnl|PID|d1014558 (D89200) similar to Saccharomyces cerevisiae reducedviability upon starvation protein 161, SWISS-PROT
Contig1719_f5d08a1.r1 937 1.9e-93 238 1266 sp|P53753|YN96_Y HYPOTHETICAL 121.1 KD PROTEIN IN BIO3-HXT17 INTERGENICREGION PRECURSOR >pir||S63399 probable membrane
Contig1798_c5e07a1.r1 927 2.1e-92 278 1207 pir||S67089 hypothetical protein YOR197w - yeast (Saccharomyces cerevisiae)>gnl|PID|e252390 (Z75105) ORF YOR197w
c6c01a1.r1 915 3.8e-91 19 837 sp|Q10178|YAV9_SCHPO HYPOTHETICAL 137.2 KD PROTEIN C27F1.09C IN CHROMOSOME I>gi|1182046 (Z69368) unknown [Schiz
n8b04a1.f1 893 8.9e-89 10 537 gi|1870209 (AC000133) ORF [Emericella nidulans]
Contig1587_i8b04a1.r1 838 5.1e-83 6 1124 sp|Q03655|YM64_Y HYPOTHETICAL 56.8 KD PROTEIN IN SCJ1-GUA1 INTERGENICREGION PRECURSOR >pir||S55097 probable membrane p
Contig1379_c6c09a1.f1 543 2e-78 440 904 sp|P49954|YL85_Y HYPOTHETICAL 32.5 KD PROTEIN YLR351C >pir||S51459hypothetical protein YLR351c - yeast (Saccharomyces
Contig1506_c7b08a1.f1 374 1.3e-73 695 1141 gnl|PID|e1287784 (AL022600) hypothetical protein [Schizosaccharomyces pombe]
h0h04a1.f1 748 4.6e-73 37 534 sp|Q04336|YM54_YEAST HYPOTHETICAL 126.6 KD PROTEIN IN RPL39-VTI1

					INTERGENICREGION >pir||S50925 hypothetical pro
Contig1545_h4b06a1.f1	737	3e-72	89	1252	gnl|PID|e1285394 (AL022305) hypothetical protein [Schizosaccharomyces pombe]
Contig1401_f0g09a1.r1	725	5.2e-71	30	887	pir||S59389 probable membrane protein YLR243w - yeast (Saccharomycescerevisiae) >gi|662338 (U20865) Ylr243wp [Sac
Contig1568_c5h09a1.f1	695	7.5e-68	188	910	sp|Q08193|YOD0_Y HYPOTHETICAL 51.9 KD PROTEIN IN MSE1-LAG2 INTERGENICREGION PRECURSOR >pir||S66713 hypothetical protei
Contig1729_a5h10a1.f1	681	2.4e-66	298	1077	sp|P53252|YG2J_Y HYPOTHETICAL 38.3 KD PROTEIN IN RPL16B-PDC6 INTERGENICREGION >pir||S64381 hypothetical protein YGR086
Contig530_o4f02a1.r1	681	2.4e-66	9	746	sp|P40055|YER2_Y HYPOTHETICAL 62.3 KD PROTEIN IN PTP3-ILV1 INTERGENICREGION >pir||S50585 hypothetical protein YER082c
Contig329_g4g01a1.r1	670	3.6e-65	33	917	sp|Q09885|YAH9_S HYPOTHETICAL 43.0 KD PROTEIN C8A4.09C IN CHROMOSOME I>pir||S62525 hypothetical protein SPAC8A4.09c -
Contig1384_c3c09a1.r1	649	6e-63	127	732	sp|O13917|YDWD_S HYPOTHETICAL 23.6 KD PROTEIN C23C11.13C IN CHROMOSOME I>gnl|PID|e334141 (Z98559) SPAC23C11.13c; len:2
c8a01a1.r1	622	4.5e-60	23	718	sp|Q92359|YDHE_SCHPO HYPOTHETICAL 73.3 KD PROTEIN C6G9.14 IN CHROMOSOME I>gnl|PID|e276617 (Z81317) serine rich
Contig883_r4f07a1.f1	400	8.5e-60	169	546	sp|Q92342|YDI4_S HYPOTHETICAL 50.4 KD PROTEIN C1F8.04C IN CHROMOSOME I>gnl|PID|e276494 (Z81312) unknown [Schizosacchar
Contig1565_r4b01a1.f1	614	3.1e-59	281	1081	gnl|PID|d1019605 (D90917) hypothetical protein [Synechocystis sp.]
x8f06a1.r1	606	2.2e-58	13	552	gnl|PID|e1294549 (AL023706) hypothetical protein [Schizosaccharomyces pombe]
m1d02a1.r1	603	4.6e-58	48	650	pir||S67033 hypothetical protein YOR145c - yeast (Saccharomyces cerevisiae)>gi|1293706 (U55020) O3513p
o8g08a1.r1	597	1.9e-57	16	552	gnl|PID|e334145 (Z98559) SPAC23C11.17; len:485aa, similar eg. to YPR125W,Q06493, chromosome xvi orf, (454a=S. pombe
y4g01a1.r1	595	3e-57	19	627	gnl|PID|e334163 (Y14554) pSI-7 protein [Cladosporium fulvum]
v7c12a1.r1	594	4.3e-57	13	621	gnl|PID|e330194 (Z93386) R11H6.1 [Caenorhabditis elegans]
n8d02a1.r1	584	4.6e-56	6	614	gnl|PID|e334344 (Z98598) hypothetical protein [Schizosaccharomyces pombe]
i7e12a1.f1	580	1.3e-55	120	683	pir||S61980 hypothetical protein YPL086c - yeast (Saccharomyces cerevisiae)>gi|1151240 (U43281) Lpg22p
w4d09a1.r1	582	6.1e-55	8	523	sp|Q10251|YD23_SCHPO HYPOTHETICAL 119.9 KD PROTEIN C56F8.03 IN CHROMOSOME I>gi|1204225 (Z69728) unknown [Schizo
Contig1844_c3c01a1.f1	569	4.3e-54	29	1360	sp|Q10327|YD72_S HYPOTHETICAL 97.1 KD PROTEIN C32A11.02C IN CHROMOSOME I>gi|1213266 (Z69796) unknown [Schizosaccharomy
f1g08a1.r1	565	5e-54	24	455	sp|P25586|YCF9_YEAST HYPOTHETICAL 37.2 KD PROTEIN IN CHA1-PRD1 INTERGENICREGION >pir||S19389 hypothetical protein
Contig1530_a5d09a1.f1	562	1e-53	412	915	sp|P53224|YG1Q_Y HYPOTHETICAL 25.2 KD PROTEIN IN ACB1-KSS1 INTERGENICREGION >pir||S64329 probable membrane protein YGR
Contig778_y4e04a1.r1	556	1e-52	118	690	pir||S74280 hypothetical protein YCL054w - yeast (Saccharomyces cerevisiae)>gnl|PID|e309034 (X59720) YCL054w, len
o1f05a1.f1	347	1.3e-52	15	293	sp|Q10325|YD6D_SCHPO HYPOTHETICAL 47.3 KD PROTEIN C17G8.13C IN CHROMOSOME

					I>gi|1213262 (Z69795) unknown [Schizo
d3b03a1.r1	530	2.4e-50	80	658	gnl|PID|e1249764 (AL021730) hypothetical protein [Schizosaccharomyces pombe]
y4h02a1.r1	522	1.8e-49	31	510	gnl|PID|e1256479 (AL022071) hypothetical protein [Schizosaccharomyces pombe]
c6h03a1.r1	515	8.8e-49	18	578	gnl|PID|e1287786 (AL022600) hypothetical protein [Schizosaccharomyces pombe]
o8g06a1.r1	511	2e-48	12	563	sp|P39941|YEI0_YEAST HYPOTHETICAL 56.5 KD PROTEIN IN HXT8 5'REGION AND IN PAU65'REGION >pir||S50519 hypothetica
t2c07a1.r1	509	4e-48	9	437	pir||S58091 probable membrane protein YDR091c - yeast (Saccharomyces cerevisiae) >gi|914875 (Z50111) un
m5a07a1.r1	409	5.7e-48	82	450	sp|P43547|YFF6_YEAST HYPOTHETICAL 23.9 KD PROTEIN IN THI5-AGP3 INTERGENICREGION >pir||S56199 hypothetical prote
i3a04a1.r1	516	8.3e-48	134	487	sp|Q10251|YD23_SCHPO HYPOTHETICAL 119.9 KD PROTEIN C56F8.03 IN CHROMOSOME I>gi|1204225 (Z69728) unknown [Schizo
j9b03a1.r1	519	8.4e-48	15	479	pir||S70099 hypothetical protein YDR334w - yeast (Saccharomyces cerevisiae)>gi|1230661 (U51032) Ydr334
Contig1381_c6d02a1.r1	500	3.4e-47	68	592	sp|P47096|YJZ5_Y HYPOTHETICAL 20.2 KD PROTEIN IN MER2-CPR7 INTERGENICREGION >pir||S57043 hypothetical protein YJR025c
m7e10a1.f1	498	5.9e-47	4	447	gnl|PID|e343814 (Z99126) hypothetical protein [Schizosaccharomyces pombe]
b0e08a1.f1	503	3.4e-46	11	487	gnl|PID|e291018 (Y08997) 146kDa nuclear protein [Xenopus laevis]
Contig1771_c9f07a1.f1	491	3.4e-46	116	1075	sp|O13716|YDZ9_S HYPOTHETICAL 44.5 KD PROTEIN C14C4.09 IN CHROMOSOME I>gnl|PID|e334267 (Z98596) hypothetical protein [
Contig536_c8a03a1.f1	491	3.4e-46	629	1030	pir||B26955 hypothetical protein - yeast (Yarrowia lipolytica) (fragment)
m2g01a1.f1	481	3.8e-45	197	715	pir||S67622 hypothetical protein YDL086w - yeast (Saccharomyces cerevisiae)>gnl|PID|e253022 (Z74134) O
m5a07a1.f1	474	2.1e-44	165	725	sp|P42884|YN71_YEAST HYPOTHETICAL 42.0 KD PROTEIN IN THI12-RPD3 INTERGENICREGION >pir||S51335 probable aryl-alc
Contig604_g2h07a1.f1	471	4.3e-44	108	842	sp|Q09686|YA14_S HYPOTHETICAL 28.0 KD PROTEIN C13C5.04 IN CHROMOSOME I>pir||S58096 hypothetical protein SPAC13C5.04 -
n8e12a1.r1	465	2e-43	130	699	pir||S67695 hypothetical protein YDL147w - yeast (Saccharomyces cerevisiae)>gnl|PID|e242702 (X97751) D
l0b09a1.r1	460	1.4e-42	122	511	pir||S61717 probable membrane protein YOL060c - yeast (Saccharomycescerevisiae) >gi|984180 (X91067) 01
Contig1428_c8c10a1.f1	451	5.3e-42	221	673	pir||S69049 hypothetical protein YPL135w - yeast (Saccharomyces cerevisiae)>gi|1244779 (U43703) Lpi10p [Saccharom
y4d02a1.r1	451	5.9e-42	37	504	pir||S69699 hypothetical protein YDR415c - yeast (Saccharomyces cerevisiae)>gi|927713 (U33007) Ydr415c
Contig851_z3b12a1.f1	433	4.7e-40	183	611	gi|2583079 (AF026816) putative oncogene protein [Homo sapiens]
Contig1405_c9b05a1.f1	432	6e-40	130	792	sp|P53290|YG3T_Y HYPOTHETICAL 38.6 KD PROTEIN IN TIF4631-KRE11 INTERGENICREGION >pir||S64474 probable membrane protein

Contig1838_c4g07a1.f1	428	1.3e-39	523 972	gnl|PID|e1283575 (AJ224865) IgE-binding protein [Aspergillus fumigatus]
z1e04a1.f1	428	2.9e-39	103 561	sp|Q09909|YAJ9_SCHPO HYPOTHETICAL 74.4 KD PROTEIN C30D11.09 IN CHROMOSOME I>pir||S62567 hypothetical protein SP
Contig207_i2a02a1.f1	318	3.6e-39	290 697	gi|2583216 (AF029913) unknown [Cochliobolus heterostrophus] >gi|2598190(AF027687) unknown [Cochliobolus heterost
Contig134_m6a03a1.f1	424	3.8e-39	219 938	gnl|PID|d1014592 (D89234) similar to Saccharomyces cerevisiae ORF YGR205W,EMBL Accession Number Z72990 [Schizosaccharo
o5f09a1.r1	421	8.7e-39	14 388	sp|Q03529|YM8I_YEAST HYPOTHETICAL 44.9 KD PROTEIN IN URA10-NRC1 INTERGENICREGION >pir||S54484 probable membrane
c6g01a1.r1	416	3.1e-38	335 832	sp|O14171|YE54_SCHPO HYPOTHETICAL 30.2 KD PROTEIN C4D7.04C IN CHROMOSOME I>gnl|PID|e334311 (Z98602) hypothetica
Contig390_n3d12a1.r1	415	3.9e-38	116 565	sp|Q10319|YD67_S HYPOTHETICAL 24.9 KD PROTEIN C17G8.07 IN CHROMOSOME I>gi|1213256 (Z69795) unknown [Schizosaccharomyce
r1a10a1.r1	415	3.9e-38	138 659	gnl|PID|e339276 (Z98850) hypothetical dehydrogenase [Schizosaccharomycespombe] >gnl|PID|e1314269 (AL031180
v7c12a1.f1	414	5e-38	136 639	sp|P43616|YFL4_YEAST HYPOTHETICAL 52.9 KD PROTEIN IN SAP155-YMR31 INTERGENICREGION >pir||S56299 hypothetical pr
Contig1485_c1d10a1.f1	412	7.2e-38	22 525	sp|Q03677|YMO9_Y HYPOTHETICAL 20.9 KD PROTEIN IN PLB1-HXT2 INTERGENICREGION >pir||S53039 hypothetical protein YMR009w
Contig1207_g7h02a1.r1	407	2.5e-37	200 769	sp|P38805|YHO8_Y HYPOTHETICAL 35.1 KD PROTEIN IN NAM8-GAR1 INTERGENICREGION >pir||S46718 hypothetical protein YHR088w
Contig1490_c3d05a1.r1	407	2.6e-37	283 864	sp|Q07821|YL27_Y HYPOTHETICAL 27.7 KD PROTEIN IN PRP19-HSP104 INTERGENICREGION >pir||S64778 hypothetical protein YLL02
e9b10a1.r1	407	2.8e-37	64 729	sp|O14057|YEB8_SCHPO HYPOTHETICAL 58.0 KD PROTEIN C2C6.08 IN CHROMOSOME I>gnl|PID|e339287 (Z98887) hypothetical
p0h07a1.r1	411	1.8e-36	36 506	pir||S52525 probable membrane protein YPL006w - yeast (Saccharomycescerevisiae) >gi|683784 (Z48483) un
Contig1659_c7f11a1.f1	399	1.8e-36	128 763	sp|O13716|YDZ9_S HYPOTHETICAL 44.5 KD PROTEIN C14C4.09 IN CHROMOSOME I>gnl|PID|e334267 (Z98596) hypothetical protein [
Contig856_x3c10a1.r1	398	2.3e-36	66 476	sp|O14155|YE72_S HYPOTHETICAL 15.9 KD PROTEIN C4A8.02C IN CHROMOSOME I>gnl|PID|e338958 (Z98762) hypothetical protein [
c7a12a1.r1	398	6.8e-36	37 666	sp|P40009|YEJ5_YEAST HYPOTHETICAL 71.9 KD PROTEIN IN PMI40-PAC2 INTERGENICREGION >pir||S50463 hypothetical prot
Contig777_y4d05a1.f1	401	1.3e-35	308 670	sp|P42839|YN61_Y HYPOTHETICAL 102.5 KD PROTEIN IN KRE1-HXT14 INTERGENICREGION >pir||S51293 probable membrane protein Y
z1c01a1.f1	392	1e-34	4 588	gnl|PID|e340027 (Z98980) hypothetical protein [Schizosaccharomyces pombe]
Contig594_c4d02a1.r1	391	1.3e-34	26 742	sp|P40164|YNU1_Y HYPOTHETICAL 98.1 KD PROTEIN IN SPX19-GCR2 INTERGENICREGION >pir||S50730 hypothetical protein YNL201c
o6b05a1.r1	380	2e-34	20 523	sp|O14249|YE63_SCHPO HYPOTHETICAL 48.7 KD PROTEIN C6G10.03C IN CHROMOSOME I>gnl|PID|e334323 (Z98603) hypothetic
e4e10a1.r1	387	8.2e-34	37 600	sp|Q10164|YAU9_SCHPO HYPOTHETICAL 143.6 KD PROTEIN C26A3.09C IN CHROMOSOME I>gnl|PID|e220681 (Z69240) hypotheti

Contig1629_a5b07a1.f1	365	7.2e-33	298 945	gi|2649154	(AE001006) membrane protein [Archaeoglobus fulgidus]
g0d02a1.f1	362	1.5e-32	215 697	gi|2983324	(AE000705) hypothetical protein [Aquifex aeolicus]
n8d02a1.f1	365	2.9e-32	261 518	gnl|PID|e334344	(Z98598) hypothetical protein [Schizosaccharomyces pombe]
m8a03a1.f1	358	3.8e-32	131 409	sp|P36156|YK56_YEAST	HYPOTHETICAL 43.3 KD PROTEIN IN SIS2-MTD1 INTERGENICREGION >pir||S38153 hypothetical prote
o1g04a1.f1	358	4.3e-32	130 624	sp|Q09923|YAKC_SCHPO	HYPOTHETICAL 37.7 KD PROTEIN C1F7.12 IN CHROMOSOME I>pir||S62584 hypothetical protein SPAC
Contig616_o6h08a1.r1	365	6e-32	128 532	sp|Q09782|YA93_S	HYPOTHETICAL 85.7 KD PROTEIN C13G6.03 IN CHROMOSOME I>pir||S62432 hypothetical protein SPAC13G6.3 - f
Contig984_c3e04a1.r1	356	7e-32	205 687	pir||S52527	hypothetical protein YPL004c - yeast (Saccharomyces cerevisiae)>gi|683786 (Z48483) unknown [Saccharom
v1c10a1.r1	363	8.2e-32	24 524	pir||S61717	probable membrane protein YOL060c - yeast (Saccharomycescerevisiae) >gi|984180 (X91067) 01
x8f04a1.f1	369	1e-31	270 593	sp|Q04958|YMF9_YEAST	HYPOTHETICAL 187.1 KD PROTEIN IN OGG1-CNA2 INTERGENICREGION >pir||S49802 probable membrane
j0b08a1.r1	362	1.1e-31	27 494	gnl|PID|e339160	(Z98850) hypothetical protein [Schizosaccharomyces pombe]>gnl|PID|e1314282 (AL031181) puta
Contig1095_o8e08a1.f1	356	1.1e-31	246 596	gnl|PID|e334108	(Z98532) hypothetical protein [Schizosaccharomyces pombe]
Contig1437_f5e03a1.f1	353	1.4e-31	26 826	gnl|PID|e1294540	(AL023705) hypothetical protein [Schizosaccharomyces pombe]
Contig1234_k8f08a1.f1	352	1.7e-31	306 1097	pir||S57377	probable membrane protein YOL092w - yeast (Saccharomycescerevisiae) >gi|600466 (X83121) orf 00929 gen
Contig598_c4a07a1.f1	363	2.3e-31	482 889	sp|P38144|YB95_Y	HYPOTHETICAL 131.1 KD HELICASE IN ALG7-ENP1 INTERGENICREGION >gnl|PID|e304681 (Z36114) ORF YBR245c [S
h1b12a1.r1	356	1.4e-30	41 385	pir||S65236	probable membrane protein YPL217c - yeast (Saccharomycescerevisiae) >gnl|PID|e246934 (Z735
w5a03a1.r1	343	1.6e-30	53 481	pir||JC4256	hypothetical 32.0k protein - Neurospora crassa >gi|773386 (L40806)open reading frame [Neur
m5e02a1.f1	342	1.9e-30	212 655	gnl|PID|d1025722	(AB010900) YNL123w homolog [Schizosaccharomyces pombe]
Contig441_n8e06a1.f1	225	2.2e-30	557 901	sp|Q10063|YAM8_S	HYPOTHETICAL 53.9 KD PROTEIN C1F5.08C IN CHROMOSOME I>gi|1103735 (Z68136) unknown [Schizosaccharomyce
n5d03a1.r1	338	5.4e-30	197 673	sp|P38286|YB09_YEAST	HYPOTHETICAL 38.7 KD PROTEIN IN RPB5-CDC28 INTERGENICREGION >pir||S46030 probable membrane
w9b04a1.r1	337	7.1e-30	74 628	sp|P47137|YJ66_YEAST	PROBABLE OXIDOREDUCTASE YJR096W >pir||S57117 aldehydereductase homolog YJR096w - yeast (Sa
Contig107_m5d10a1.f1	335	1.2e-29	188 856	sp|P36091|YKE6_Y	HYPOTHETICAL 49.6 KD PROTEIN IN ELM1-PRI2 INTERGENICREGION >pir||S37867 hypothetical protein YKL046c
i8e07a1.r1	335	1.2e-29	55 408	pir||S67038	hypothetical protein YOR150w - yeast (Saccharomyces cerevisiae)>gi|1293710 (U55020) O3530p
i0e05a1.r1	334	1.4e-29	11 397	sp|P32623|UTR2_YEAST	UTR2 PROTEIN (UNKNOWN TRANSCRIPT 2 PROTEIN)
Contig643_o8f01a1.f1	252	1.4e-29	588 863	pir||S51434	hypothetical protein YLR189c - yeast (Saccharomyces

				cerevisiae)>gi|577215 (U17246) Ylr189cp [Saccharo
w9c09a1.r1	333	1.8e-29	7 489	gnl|PID|e349610 (Z99262) hypothetical protein [Schizosaccharomyces pombe]
i3c03a1.r1	330	3.8e-29	16 483	sp|P43567|YFD0_YEAST HYPOTHETICAL 41.9 KD PROTEIN IN HAC1-CAK1 INTERGENICREGION >pir||S56224 hypothetical prote
g6g08a1.r1	329	5.1e-29	50 694	sp|Q10478|YDF6_SCHPO HYPOTHETICAL 51.8 KD PROTEIN C17C9.06 IN CHROMOSOME I>gnl|PID|e241976 (Z73099) hypothetica=S. pombe
f0g01a1.r1	337	5.4e-29	10 606	sp|Q09850|YAEA_SCHPO HYPOTHETICAL 77.9 KD PROTEIN C23D3.10C IN CHROMOSOME I>pir||S62501 hypothetical protein SP
i3h07a1.r1	336	6.9e-29	29 466	sp|O13965|YE45_SCHPO HYPOTHETICAL 79.3 KD PROTEIN C24C9.05C IN CHROMOSOME I>gnl|PID|e334374 (Z98601) hypothetic
f5c03a1.f1	325	1.2e-28	100 498	gnl|PID|d1018329 (D90907) hypothetical protein [Synechocystis sp.]
v3b07a1.r1	324	1.6e-28	149 514	sp|P53219|YG1L_YEAST HYPOTHETICAL 38.5 KD PROTEIN IN ERV1-GLS2 INTERGENICREGION >pir||S64322 probable membrane
Contig1338_g1d12a1.r1	330	3e-28	406 750	pir||S52675 probable membrane protein YDR109c - yeast (Saccharomycescerevisiae) >gi|747884 (Z48758) unknown [Sacc
Contig985_j9f11a1.f1	320	4.5e-28	397 774	gnl|PID|e1263895 (AL022103) hypothetical protein [Schizosaccharomyces pombe]
j7c01a1.r1	329	6.7e-28	38 289	gnl|PID|e1292820 (AJ005963) 100 kDa protein [Ajellomyces capsulatus]
r5d12a1.r1	316	1.1e-27	199 537	gnl|PID|e339959 (Z98980) transcription factor [Schizosaccharomyces pombe]
q0d09a1.r1	315	1.5e-27	47 379	sp|Q03691|YM56_YEAST HYPOTHETICAL 28.9 KD PROTEIN IN CLN1-RAD14 INTERGENICREGION
Contig611_c3a12a1.f1	324	1.6e-27	125 544	sp|P97739|ECE1_C ENDOTHELIN-CONVERTING ENZYME 1 (ECE-1) >bbs|178962(S82653) endothelin converting enzyme, ECE [guinea
c3e08a1.r1	313	2.6e-27	53 613	sp|P75791|YBIU_ECOLI HYPOTHETICAL 47.3 KD PROTEIN IN OMPX-MOEB INTERGENICREGION >gi|1787042 (AE000184) f421; Th
Contig605_c3g12a1.r1	312	2.9e-27	141 641	sp|P38260|YBV1_Y HYPOTHETICAL 32.6 KD PROTEIN IN VPS15-YMC2 INTERGENICREGION >pir||S48266 hypothetical protein YBR101c
Contig145_j5a05a1.f1	316	5.5e-27	220 684	gi|4088 (X69881) ORF2 [Saccharomyces cerevisiae]
c5h08a1.r1	317	7.1e-27	90 572	sp|P39992|YEC3_YEAST HYPOTHETICAL 78.3 KD PROTEIN IN RIP1-URA3 INTERGENICREGION >pir||S50436 hypothetical prote
Contig457_d4d11a1.f1	309	1.1e-26	50 604	sp|Q10478|YDF6_S HYPOTHETICAL 51.8 KD PROTEIN C17C9.06 IN CHROMOSOME I>gnl|PID|e241976 (Z73099) hypothetical protein [
z3a10a1.r1	201	1.1e-26	9 326	sp|P53326|YG5L_YEAST HYPOTHETICAL 81.2 KD PROTEIN IN MES1-FOL2 INTERGENICREGION >pir||S64599 probable membrane
Contig1778_a1g03c9.r1	312	1.2e-26	480 956	pir||S70114 probable membrane protein YDR284c - yeast (Saccharomycescerevisiae) >gi|1332640 (U51031) Ydr284cp [Sa
o6e11a1.r1	303	2.8e-26	70 531	gnl|PID|e1295796 (AL023776) hypothetical protein [Schizosaccharomyces pombe]
Contig114_l3a06a1.r1	302	3.1e-26	34 435	gnl|PID|e332207 (AJ000977) hypothetical protein [Rhodobacter sphaeroides]
Contig211_i0g12a1.r1	302	3.3e-26	18 455	gi|2266911 (AE001274) L4171.5 [Leishmania major]

Contig1309_j7h09a1.f1	301	4.5e-26	344 598	gi|2707187	(U94183) unknown [Glomerella cingulata]
Contig1200_g9g02a1.f1	310	6.5e-26	101 508	sp|Q04500|YMJ3_Y	HYPOTHETICAL 103.0 KD PROTEIN IN RAD10-PRS4 INTERGENICREGION >pir||S49634 hypothetical protein YML093
Contig503_c9g10a1.f1	299	7.2e-26	197 649	gnl|PID|e236467	(Z71178) B0024.12 [Caenorhabditis elegans]
Contig994_j9a11a1.f1	299	7.4e-26	498 794	gi|3094014	(AF060862) unknown [Homo sapiens]
p0f10a1.r1	302	8.1e-26	11 466	sp|P38821|YHR3_YEAST	HYPOTHETICAL 54.2 KD PROTEIN IN CDC12-ORC6 INTERGENICREGION >pir||S48955 hypothetical prot
Contig1598_m1f01a1.r1	297	1.1e-25	515 1165	gnl|PID|e321532	(Y13635) Vip1 protein [Schizosaccharomyces pombe]>gnl|PID|e1202248 (AL009197) hypothetical protein[Sc
Contig563_c5h03a1.f1	304	2.8e-25	438 656	gnl|PID|d1022254	(AB004535) hypothetical protein YPR112c [Schizosaccharomycespombe]
n3c09a1.r1	292	4.1e-25	101 325	gnl|PID|e275716	(Z81071) F28F8.3 [Caenorhabditis elegans]
v3a12a1.r1	291	5e-25	5 328	gnl|PID|e323034	(Z97052) hypothetical protein [Schizosaccharomyces pombe]
r7f02a1.r1	303	5.3e-25	232 744	gi|1197061	(L36344) ORF; putative [Saccharomyces cerevisiae]
Contig1407_c4h07a1.r1	296	5.4e-25	9 533	sp|P53189|YGC8_Y	HYPOTHETICAL 56.4 KD PROTEIN IN RPL32-CWH41 INTERGENICREGION PRECURSOR >pir||S64030 probable membrane
Contig1097_g9g10a1.r1	290	6.4e-25	116 808	gnl|PID|e1291650	(AL023290) hypothetical protein [Schizosaccharomyces pombe]
Contig629_c1b03a1.f1	300	6.8e-25	246 725	gnl|PID|e1313487	(AL031154) hypothetical protein [Schizosaccharomyces pombe]
z6a10a1.r1	290	1.5e-24	4 417	sp|P53285|YG3H_YEAST	HYPOTHETICAL 54.5 KD PROTEIN IN CBF2-SKN1 INTERGENICREGION >pir||S64450 probable membrane
10f05a1.f1	285	1.9e-24	103 474	pir||S61978	hypothetical protein YPL088w - yeast (Saccharomyces cerevisiae)>gi|1151238 (U43281) Lpg20p
Contig1236_g9d02a1.f1	285	2.3e-24	3 497	sp|P38278|YBZ1_Y	HYPOTHETICAL 38.5 KD PROTEIN IN IRA1-MAK5 INTERGENICREGION >pir||S46010 hypothetical protein YBR141c
n0f01a1.f1	285	2.3e-24	181 450	gnl|PID|e1285101	(AJ002026) rAsp f 13 [Aspergillus fumigatus]=human allergen
g7h07a1.r1	284	3.1e-24	222 686	sp|Q04179|YD40_YEAST	HYPOTHETICAL 42.3 KD PROTEIN IN YTA2-DIT1 INTERGENICREGION >pir||S69683 hypothetical prote
Contig613_c2f06a1.f1	191	3.5e-24	149 400	pir||S59765	hypothetical protein YPR100w - yeast (Saccharomyces cerevisiae)>gi|914971 (U32445) Note that there is
v3e02a1.r1	282	4.8e-24	303 566	gnl|PID|e331452	(Z98056) hyypothetical protein [Schizosaccharomyces pombe]
Contig1342_d4c01a1.f1	281	5.9e-24	354 863	pir||S67183	hypothetical protein YOR281c - yeast (Saccharomyces cerevisiae)>gnl|PID|e189406 (X89633) hypothetical
Contig262_a0e01a1.f1	281	6.1e-24	172 573	sp|Q40784|AAPC_P	POSSIBLE APOSPORY-ASSOCIATED PROTEIN C >gi|549984(U13148) possible apospory-associated protein [Penni
o8e11a1.f1	288	1.7e-23	10 312	sp|O14053|YEB4_SCHPO	HYPOTHETICAL 100.6 KD TRP-ASP REPEATS CONTAINING PROTEINC2C6.04C IN CHROMOSOME I >gnl|PID|
m5e02a1.r1	275	2.7e-23	27 482	gnl|PID|d1025722	(AB010900) YNL123w homolog [Schizosaccharomyces pombe]
Contig827_y4b09a1.f1	279	4.3e-23	209 631	gi|1067085	(Z47357) ZK1128.1 [Caenorhabditis elegans]

m7b03a1.r1	273	4.3e-23	38 364	pir||S61970 hypothetical protein YPL096w - yeast (Saccharomyces cerevisiae)>gi|1151230 (U43281) Lpg12p
k9d02a1.f1	272	5.4e-23	287 574	sp|P39519|YBDF_YEAST HYPOTHETICAL PROTEIN IN BDF1 5'REGION (ORF1) >gi|547573(Z18944) ORF1 [Saccharomyces cerevi
l3a06a1.f1	272	5.5e-23	249 521	sp|P36156|YK56_YEAST HYPOTHETICAL 43.3 KD PROTEIN IN SIS2-MTD1 INTERGENICREGION >pir||S38153 hypothetical prote
Contig525_c8c02a1.f1	280	6.4e-23	215 1033	gnl|PID|e1251101 (AL021838) hypothetical protein [Schizosaccharomyces pombe]
q0e02a1.r1	283	6.7e-23	27 521	gnl|PID|d1010417 (D63484) The KIAA0150 gene product is novel. [Homo sapiens]
w8g01a1.f1	278	9.1e-23	118 405	gi|736313 (Z48756) unknown [Saccharomyces cerevisiae]
o6h03a1.r1	268	1.4e-22	20 547	sp|P40402|YZEC_BACSU HYPOTHETICAL 41.8 KD PROTEIN (ORFM)
h1a09a1.f1	267	1.8e-22	198 626	sp|P47095|YJZ4_YEAST HYPOTHETICAL 27.4 KD PROTEIN IN MER2-CPR7 INTERGENICREGION >pir||S57042 hypothetical prote
r1d10a1.r1	266	2.2e-22	338 661	gi|2408032 (Z99162) hypothetical protein [Schizosaccharomyces pombe]
d5b06a1.r1	266	2.3e-22	39 338	gnl|PID|e1216801 (AL021046) SPAC3G9.15c; len:230aa; similarity: to YLR051C,Q12035, unclassified protein, (2
h1d10a1.f1	265	7.5e-22	1 546	sp|P35728|YKF9_YEAST HYPOTHETICAL 49.6 KD PROTEIN IN FBA1-TOA2 INTERGENICREGION >pir||S37881 hypothetical prote
m2f08a1.f1	272	8.9e-22	98 484	gi|1019710 (L47993) ORF YJR091c [Saccharomyces cerevisiae]
Contig1253_g7a02a1.r1	260	8.9e-22	70 303	pir||S72314 hypothetical protein YHR004c-a - yeast (Saccharomyces cerevisiae)>gnl|PID|e273884 (Z80875) Mrs11p [Sa
l3e10a1.r1	164	9.3e-22	34 309	pir||S61140 probable membrane protein YPR156c - yeast (Saccharomycescerevisiae) >gi|849164 (U28371) Si
z5a06a1.r1	272	9.9e-22	248 577	sp|Q09764|YA7B_SCHPO HYPOTHETICAL 107.1 KD PROTEIN C24H6.11C IN CHROMOSOME I>pir||S62413 hypothetical protein S
c0h04a1.r1	266	1e-21	195 476	sp|Q04991|YM68_YEAST HYPOTHETICAL 56.2 KD PROTEIN IN ERG8-UBP8 INTERGENICREGION >pir||S57589 probable membrane
p0e05a1.r1	278	1.2e-21	208 501	sp|Q10064|YAMB_SCHPO HYPOTHETICAL 420.8 KD PROTEIN C1F5.11C IN CHROMOSOME I>gi|1103738 (Z68136) unknown [Schizo
w4b10a1.r1	275	1.2e-21	11 520	sp|P38737|YHD0_YEAST HYPOTHETICAL 210.4 KD PROTEIN IN GUT1-RIM1 INTERGENICREGION >pir||S48938 hypothetical prot
y4c07a1.r1	258	1.7e-21	99 476	sp|Q09839|YADE_SCHPO HYPOTHETICAL 40.0 KD PROTEIN C4G8.14C IN CHROMOSOME I>pir||S62491 hypothetical protein SPA
Contig208_j7f01a1.f1	257	2.1e-21	136 465	sp|P87132|YDM1_S HYPOTHETICAL PROTEIN C57A7.01 IN CHROMOSOME I>gnl|PID|e316110 (Z95396) unknown [Schizosaccharomyces p
Contig809_r8d09a1.r1	270	3.6e-21	287 541	gnl|PID|e1314595 (AL031187) putative protein [Arabidopsis thaliana]
s9e01a1.r1	260	3.6e-21	118 597	sp|Q04371|YMR7_YEAST HYPOTHETICAL 54.1 KD PROTEIN IN PEX12-TAP42 INTERGENICREGION >pir||S54029 hypothetical pro
Contig1380_d5f07a1.r1	257	5e-21	271 837	sp|P40087|DDI1_Y DNA-DAMAGE INDUCIBLE PROTEIN DDI1 >pir||S50646hypothetical protein YER143w - yeast (Saccharomyces cer
x7g05a1.r1	249	1.5e-20	90 599	pir||S66832 hypothetical protein YOL135c - yeast (Saccharomyces

				cerevisiae)>gnl|PID|e252305 (Z74877) O
y6h11a1.r1	254	1.6e-20	181 558	sp|Q09844|YAE3_SCHPO HYPOTHETICAL 54.3 KD PROTEIN C23D3.03C IN CHROMOSOME I>pir||S62494 hypothetical protein SP
r7d03a1.r1	255	1.9e-20	56 679	gi|1707074 (U80450) M01E11.2 [Caenorhabditis elegans]
Contig201_i2g01a1.r1	258	3.4e-20	10 348	pir||S69079 hypothetical protein YPR097w - yeast (Saccharomyces cerevisiae)>gi|1230699 (U51033) P9513.1 gene prod
Contig699_t2g12a1.r1	247	4e-20	176 622	gnl|PID|e349695 (Z99296) hypothetical protein [Schizosaccharomyces pombe]
Contig41_c4d04a1.r1	154	7.9e-20	271 408	pir||S72314 hypothetical protein YHR004c-a - yeast (Saccharomyces cerevisiae)>gnl|PID|e273884 (Z80875) Mrs11p [Sa
r7g09a1.r1	240	1.3e-19	217 651	gi|927403 (Z50177) F46G10.3 [Caenorhabditis elegans]
Contig658_b0d01a1.f1	239	1.6e-19	132 395	gi|832882 (L42454) EF-hand protein [Schizosaccharomyces pombe]
Contig135_j9d07a1.f1	238	2.1e-19	289 552	sp|O13868|YE12_S HYPOTHETICAL 11.8 KD PROTEIN C1B3.02C IN CHROMOSOME I>gnl|PID|e334341 (Z98598) hypothetical protein [
c4h05a1.r1	249	2.4e-19	263 550	gnl|PID|e1292820 (AJ005963) 100 kDa protein [Ajellomyces capsulatus]
d3a01a1.r1	244	4.3e-19	82 642	sp|P53962|YND5_YEAST HYPOTHETICAL 43.8 KD PROTEIN IN NCE3-HHT2 INTERGENICREGION >pir||S62957 hypothetical prote
Contig623_c1f06a1.f1	239	7.5e-19	269 577	pir||S61029 hypothetical protein YPL235w - yeast (Saccharomyces cerevisiae)>gi|1061254 (Z67751) putative protein
Contig995_c9c03a1.f1	243	1.2e-18	69 536	pir||S55965 probable membrane protein YLR409c - yeast (Saccharomycescerevisiae) >gi|625119 (U19729) Ylr409cp [Sac
Contig306_g6f04a1.f1	230	1.5e-18	13 426	pir||S67607 probable membrane protein YDL072c - yeast (Saccharomycescerevisiae) >gnl|PID|e253013 (Z74120) ORF YDL
Contig1026_n3g10a1.r1	229	2.1e-18	101 316	gnl|PID|d1014541 (D89182) similar to Saccharomyces cerevisiae Lpg10p,SWISS-PROT Accession Number U43281 [Schizosacchar
z7g05a1.f1	228	2.3e-18	232 540	sp|P38068|YBM4_YEAST HYPOTHETICAL 22.6 KD PROTEIN IN IPP1-TTP1 INTERGENICREGION >pir||S45869 glutaredoxin homol
r1a10a1.f1	228	2.9e-18	53 424	sp|P53839|YN14_YEAST HYPOTHETICAL 38.8 KD PROTEIN IN MET2-SEC2 INTERGENICREGION >pir||S63248 hypothetical prote
c6f08a1.r1	227	3.3e-18	65 277	gnl|PID|e213997 (Z68218) K01H12.1 [Caenorhabditis elegans]
Contig609_c3f10a1.f1	235	6e-18	194 697	sp|Q09782|YA93_S HYPOTHETICAL 85.7 KD PROTEIN C13G6.03 IN CHROMOSOME I>pir||S62432 hypothetical protein SPAC13G6.3 - f
Contig1522_d3b06a1.f1	224	6.9e-18	563 949	sp|P40513|MA33_Y MITOCHONDRIAL ACIDIC PROTEIN MAM33 PRECURSOR >pir||S48409hypothetical protein YIL070c - yeast (Saccha
Contig79_l3c11a1.r1	231	9.3e-18	107 532	sp|Q09799|YAA5_S HYPOTHETICAL 69.5 KD PROTEIN C22G7.05 IN CHROMOSOME I>pir||S62449 hypothetical protein SPAC22G7.05 -
Contig1746_c7g11a1.f1	222	1e-17	431 619	pir||S66926 hypothetical protein YOR052c - yeast (Saccharomyces cerevisiae)>gnl|PID|e251973 (Z74960) ORF YOR052c
o6c10a1.r1	222	1e-17	10 408	sp|Q07651|YD22_YEAST HYPOTHETICAL 34.1 KD PROTEIN IN CDC13-GCS1 INTERGENICREGION >pir||S67785 probable membrane
j0h11a1.f1	222	1.1e-17	105 446	gnl|PID|d1019435 (D90916) hypothetical protein [Synechocystis sp.]
Contig891_w9e07a1.f1	219	2.2e-17	301 507	sp|O14056|YEB7_S HYPOTHETICAL 11.3 KD PROTEIN C2C6.07 IN CHROMOSOME I>gnl|PID|e339194 (Z98887) hypothetical protein [S

Contig1694_d1c06a1.f1	222	2.5e-17	10 489	gnl|PID|e1293607 (AL023635) hypothetical protein MLCB1243.36 [Mycobacteriumleprae]
g3e01a1.r1	226	6.3e-17	25 522	gnl|PID|d1019541 (D90917) hypothetical protein [Synechocystis sp.]
Contig491_d1f11a1.r1	210	1.9e-16	305 568	pir||S54502 probable membrane protein YPR028w - yeast (Saccharomycescerevisiae) >gi|809593 (Z49274) unknown [Sacc
g1d07a1.f1	212	2e-16	71 469	gnl|PID|e293680 (Z84498) hypothetical protein Rv1928c [Mycobacteriumtuberculosis]
Contig368_g2d09a1.f1	212	2.2e-16	364 822	gnl|PID|e1132792 (Z99163) hypothetical protein [Schizosaccharomyces pombe]
Contig845_r7g08a1.f1	208	3e-16	31 549	sp|P54168|YPGQ_B HYPOTHETICAL 23.1 KD PROTEIN IN BSAA-ILVD INTERGENICREGION >gi|1256633 (L77246) putative [Bacillus su
Contig454_n8h01a1.f1	208	3.4e-16	313 570	gi|949850 (Z50795) R166.3 [Caenorhabditis elegans]
m6d09a1.f1	221	3.6e-16	94 576	pir||S65157 hypothetical protein YPL146c - yeast (Saccharomyces cerevisiae)>gi|1244771 (U43703) Lpi2p
u4h11a1.r1	219	4.4e-16	67 507	sp|Q09731|YA4E_SCHPO HYPOTHETICAL 107.3 KD TRP-ASP REPEATS CONTAINING PROTEINC31A2.14 IN CHROMOSOME I >pir||S58
Contig1683_v7b02a1.f1	208	6.7e-16	586 1107	sp|P53721|YN89_Y HYPOTHETICAL 25.3 KD PROTEIN IN TIM23-ARE2 INTERGENICREGION >pir||S63349 probable membrane protein YN
r2g07a1.r1	220	6.8e-16	28 426	sp|P47045|YJF4_YEAST HYPOTHETICAL 54.2 KD PROTEIN IN BTN1-PEP8 INTERGENICREGION >pir||S56826 hypothetical prote
i0e06a1.f1	206	1.1e-15	255 515	gnl|PID|e1292638 (AL023534) hypothetical protein [Schizosaccharomyces pombe]
c5h03a1.r1	217	1.2e-15	28 576	gnl|PID|e1250351 (AL021768) putative protein [Arabidopsis thaliana]
m8h10a1.f1	202	1.2e-15	188 373	sp|O01578|YXX3_CAEEL HYPOTHETICAL 16.3 KD PROTEIN F53F10.3 IN CHROMOSOME I>gi|1943771 (U97191) similar to the r
Contig407_f0c01a1.r1	202	1.3e-15	2 181	sp|P34227|YBG4_Y HYPOTHETICAL 29.5 KD PROTEIN IN SEF1-KIP1 INTERGENICREGION >pir||S39825 hypothetical protein YBL064c
o0d01a1.r1	202	1.4e-15	241 450	gnl|PID|e349291 (Z99281) Y57G11C.13 [Caenorhabditis elegans]
k5b08a1.r1	211	1.8e-15	16 234	sp|P53968|YNC7_YEAST HYPOTHETICAL 76.3 KD ZINC FINGER PROTEIN IN KTR5-UME3INTERGENIC REGION >pir||S62939 hypoth
g7a11a1.r1	207	2.4e-15	149 496	pir||S66939 hypothetical protein YOR056c - yeast (Saccharomyces cerevisiae)>gnl|PID|e251975 (Z74964) O
o5a08a1.r1	210	2.6e-15	114 461	sp|P53326|YG5L_YEAST HYPOTHETICAL 81.2 KD PROTEIN IN MES1-FOL2 INTERGENICREGION >pir||S64599 probable membrane
j5g10a1.r1	199	3e-15	75 323	gnl|PID|d1018464 (D90908) hypothetical protein [Synechocystis sp.]
m8d11a1.f1	198	3.8e-15	239 718	sp|Q10433|YDD9_SCHPO HYPOTHETICAL 24.7 KD PROTEIN C1B9.09C IN CHROMOSOME I>gnl|PID|e235401 (Z70720) unknown [Sc
Contig386_f4b12a1.r1	219	3.9e-15	131 442	sp|P36114|YKZ8_Y HYPOTHETICAL 81.8 KD PROTEIN IN YPT52-DBP7 INTERGENICREGION >pir||S38087 hypothetical protein YKR018c
c1g08a1.r2	227	5.2e-15	36 353	gnl|PID|e315881 (Z95396) unknown [Schizosaccharomyces pombe]
Contig1554_c7c05a1.f1	204	5.7e-15	557 796	gnl|PID|e1251086 (AL021837) hypothetical protein [Schizosaccharomyces pombe]
c2f07a1.f1	195	7.6e-15	144 410	gi|809578 (Z49273) unknown [Saccharomyces cerevisiae]

f5c08a1.r1	205	1e-14	18 263	gnl|PID|e1250023	(AL021747) hypothetical protein [Schizosaccharomyces pombe]
Contig1108_g9e08a1.r1	194	1e-14	55 234	gnl|PID|e275630	(Z81038) C25A1.6 [Caenorhabditis elegans]
g4f12a1.r1	193	1.2e-14	311 586	gi|3378330	(AF079317) unknown [Sphingomonas aromaticivorans]
Contig408_e9h07a1.r1	200	1.4e-14	11 568	sp|Q10367|YDBH_S	HYPOTHETICAL 53.0 KD PROTEIN C22E12.17C IN CHROMOSOME I>gi|1220292 (Z70043) unknown [Schizosaccharomy
Contig1311_o4b12a1.f1	196	1.5e-14	198 596	pir||S61991	hypothetical protein YOR007c - yeast (Saccharomyces cerevisiae)>gi|1151003 (U43491) hypothetical prot
x1c08a1.r1	199	1.6e-14	220 630	gi|2622063	(AE000870) conserved protein [Methanobacterium thermoautotrophicum]
r1d02a1.r1	201	2.6e-14	30 260	gnl|PID|e339911	(Z98974) hypothetical protein [Schizosaccharomyces pombe]
l5a05a1.f1	200	2.6e-14	287 529	gnl|PID|e1251101	(AL021838) hypothetical protein [Schizosaccharomyces pombe]
Contig1474_h4a03a1.f1	199	2.6e-14	357 713	sp|P40055|YER2_Y	HYPOTHETICAL 62.3 KD PROTEIN IN PTP3-ILV1 INTERGENICREGION >pir||S50585 hypothetical protein YER082c
k0g06a1.r1	199	2.8e-14	5 385	gnl|PID|e1295804	(AL023777) hypothetical protein [Schizosaccharomyces pombe]
m0e09a1.f1	193	3.6e-14	183 425	gnl|PID|e241985	(Z73100) unknown [Schizosaccharomyces pombe]
Contig578_c5a08a1.f1	207	4e-14	864 1136	gi|1658377	(U69170) unknown [Pichia pastoris]
Contig1078_j4f01a1.f1	188	4e-14	624 800	gnl|PID|e1256502	(AL022072) hypothetical protein [Schizosaccharomyces pombe]
Contig1445_h8a07a1.f1	190	4.8e-14	465 767	pir||S66764	hypothetical protein YOL071w - yeast (Saccharomyces cerevisiae)>gnl|PID|e251875 (Z74813) ORF YOL071w
Contig670_a1b05c9.r1	191	5.3e-14	297 524	sp|P53290|YG3T_Y	HYPOTHETICAL 38.6 KD PROTEIN IN TIF4631-KRE11 INTERGENICREGION >pir||S64474 probable membrane protein
x3e04a1.f1	194	5.4e-14	266 580	gi|2689890	(AE000792) conserved hypothetical protein [Borrelia burgdorferi]
f1d11a1.r1	195	5.6e-14	18 410	gi|1173491	(U20390) ORF494 [Saccharomyces cerevisiae]
y6g07a1.r1	186	6.5e-14	13 450	sp|Q04272|YMW7_YEAST	HYPOTHETICAL 25.6 KD PROTEIN IN ABF2-CHL12 INTERGENICREGION >pir||S54452 hypothetical prot
r2c05a1.r1	209	9e-14	21 623	gnl|PID|e280810	(Z82015) yukK [Bacillus subtilis]-similar to Pristinamycin I synthase of Streptomyces
g7b05a1.r1	192	1.2e-13	149 772	pir||S53401	probable membrane protein YLR324w - yeast (Saccharomycescerevisiae) >gi|662138 (U20618) Yl
Contig1787_c0c10a1.r1	193	1.3e-13	98 259	gnl|PID|e1188370	(Z82286) W02A2.g [Caenorhabditis elegans]
Contig1861_c4g05a1.f1	193	1.3e-13	394 555	gnl|PID|e1188370	(Z82286) W02A2.g [Caenorhabditis elegans]
e7c01a1.r1	205	1.4e-13	131 610	gnl|PID|e334014	(Z98529) hypothetical protein [Schizosaccharomyces pombe]
g5b08a1.r1	203	1.4e-13	113 700	gnl|PID|e1315376	(AL031261) hypothetical protein [Schizosaccharomyces pombe]
m7c06a1.r1	185	1.5e-13	274 609	sp|P53337|YG5Y_YEAST	HYPOTHETICAL 35.0 KD PROTEIN IN BGL2-ZUO1 INTERGENICREGION >pir||S64619 probable membrane

m0h07a1.r1	189	1.9e-13	144 446	sp|O13909|YDW1_SCHPO HYPOTHETICAL 49.2 KD PROTEIN C23C11.01 IN CHROMOSOME I>gnl|PID|e334129 (Z98559) SPAC23C11.
i3e06a1.f1	188	2e-13	215 466	gnl|PID|d1032703 (AB016218) unknown: similar to human GA17 protein[Schizosaccharomyces pombe]
Contig1671_f5a07a1.r1	210	2.7e-13	726 1469	gnl|PID|e1292632 (AL023534) hypothetical protein [Schizosaccharomyces pombe]
g3b10a1.r1	136	2.9e-13	38 322	gi|2315350 (AF016439) No definition line found [Caenorhabditis elegans]
r5a04a1.r1	194	3.3e-13	96 407	sp|P40531|YIE1_YEAST 36.7 KD PROTEIN IN CBR5-NOT3 INTERGENIC REGION>pir||S49937 hypothetical protein YIL041w -
j0h09a1.r1	189	3.4e-13	20 328	sp|P25351|YCR3_YEAST HYPOTHETICAL 69.2 KD PROTEIN IN HSP30-PMP1 INTERGENICREGION >pir||S19434 probable transpor
Contig678_o9d08a1.r1	192	3.8e-13	302 610	sp|P47029|YJI4_Y HYPOTHETICAL 117.2 KD PROTEIN IN EXO70-ARP4 INTERGENICREGION >gi|895905 (X88851) hypothetical protein
m6e06a1.r1	212	4.5e-13	48 404	pir||S59841 probable membrane protein YPR184w - yeast (Saccharomycescerevisiae) >gi|786314 (U25842) Hi
z6g01a1.f1	178	4.8e-13	309 557	gnl|PID|e1198272 (Z99165) hypothetical protein [Schizosaccharomyces pombe]
y7g04a1.f1	202	5.7e-13	294 650	gnl|PID|e237905 (X97346) FCYX gene product [Saccharomyces cerevisiae]>gi|1381130 (U18813) Fcy22p: Puring-c
Contig1429_g6f08a1.f1	175	1e-12	184 423	sp|P47111|YJ14_Y HYPOTHETICAL 15.7 KD PROTEIN IN NUP85-SSC1 INTERGENICREGION >pir||S57063 probable membrane protein YJ
w4a10a1.f1	182	1.3e-12	109 507	sp|Q09817|YAC3_SCHPO HYPOTHETICAL 56.6 KD PROTEIN C16C9.03 IN CHROMOSOME I>pir||S62473 hypothetical protein SPA
Contig1643_o4g06a1.r1	173	1.6e-12	248 547	pir||S67201 hypothetical protein YOR297c - yeast (Saccharomyces cerevisiae)>gnl|PID|e252135 (Z75205) ORF YOR297c
Contig446_d5d06a1.f1	184	1.9e-12	273 494	gnl|PID|e351296 (Z99532) hypothetical protein [Schizosaccharomyces pombe]
Contig1225_g3e01a1.f1	196	2e-12	315 686	gnl|PID|d1019541 (D90917) hypothetical protein [Synechocystis sp.]
Contig1044_s9f01a1.f1	183	2.1e-12	483 662	pir||S37694 gene PC326 protein - mouse >gi|200241 (M95564) protein PC326 [Musmusculus]
j7c01a1.f1	183	2.8e-12	238 669	gnl|PID|e1292820 (AJ005963) 100 kDa protein [Ajellomyces capsulatus]
y6h06a1.r1	170	3.6e-12	85 276	gi|1350548 (L47609) heat shock-like protein [Picea glauca]
a0e01a1.r1	172	3.9e-12	158 406	sp|Q03161|YMY9_YEAST HYPOTHETICAL 34.0 KD PROTEIN IN CTF13-YPR2 INTERGENICREGION >pir||S55085 hypothetical prot
Contig724_u4g06a1.r1	169	4.4e-12	14 292	pir||S59397 probable membrane protein YLR251w - yeast (Saccharomycescerevisiae) >gi|662333 (U20865) Ylr251wp [Sac
Contig1081_g9g02a1.r1	181	4.6e-12	5 415	sp|Q04500|YMJ3_Y HYPOTHETICAL 103.0 KD PROTEIN IN RAD10-PRS4 INTERGENICREGION >pir||S49634 hypothetical protein YML093
g6a10a1.f1	183	5.6e-12	294 548	gnl|PID|e1292587 (AL023518) hypothetical protein [Schizosaccharomyces pombe]
f2b12a1.r1	193	5.8e-12	95 490	gnl|PID|e1294546 (AL023706) hypothetical protein [Schizosaccharomyces pombe]

i0h03a1.r1	174	8.1e-12	111 449	gnl|PID|e310346 (Z93386) R11H6.2 [Caenorhabditis elegans]
Contig876_w4a08a1.f1	166	9.5e-12	270 596	gnl|PID|e324205 (Z97185) hypothetical protein [Schizosaccharomyces pombe]
Contig1189_h0b12a1.f1	177	1.1e-11	51 557	sp|Q09686|YA14_S HYPOTHETICAL 28.0 KD PROTEIN C13C5.04 IN CHROMOSOME I>pir||S58096 hypothetical protein SPAC13C5.04 -
Contig663_a5e09a1.f1	171	1.2e-11	237 506	gnl|PID|e312773 (Z72840) ORF YGR054w [Saccharomyces cerevisiae]
s8a07a1.f1	183	1.4e-11	352 561	gnl|PID|e1285386 (AL022304) hypothetical protein [Schizosaccharomyces pombe]
k0f02a1.r1	163	3e-11	176 400	sp|P38838|YHT4_YEAST HYPOTHETICAL 30.6 KD PROTEIN IN ACT5-YCK1 INTERGENICREGION >pir||S48978 hypothetical prote
Contig344_g3h06a1.f1	161	3e-11	275 550	sp|P40156|YNV3_Y HYPOTHETICAL 25.3 KD PROTEIN IN PEX17-MER1 INTERGENICREGION >pir||S50718 hypothetical protein YNL213c
c6h09a1.r1	184	3.1e-11	98 436	gnl|PID|e339146 (Z98849) hypothetical protein [Schizosaccharomyces pombe]
h4c07a1.r1	171	3.1e-11	33 425	sp|P38731|YHE0_YEAST HYPOTHETICAL 70.9 KD PROTEIN IN CBP2 5'REGION>pir||S48928 hypothetical protein YHL040c - y
Contig641_c0c03a1.r1	172	5.1e-11	76 342	gi|2149640 (U91995) Argonaute protein [Arabidopsis thaliana]
Contig1473_x5f05a1.r1	186	5.8e-11	200 820	sp|P38724|YHE7_Y HYPOTHETICAL 71.6 KD PROTEIN IN CBP2 5'REGION>pir||S48921 hypothetical protein YHL047c - yeast (Sacch
Contig695_o9h08a1.f1	*158*	*6.1e-11*	*110 298*	*gi|1397277 (U61947) C06G3.11 gene product [Caenorhabditis elegans]*
h0g07a1.r1	160	6.3e-11	38 376	gi|2384956 (AF022985) No definition line found [Caenorhabditis elegans]
Contig388_f2d10a1.f1	164	7.7e-11	102 419	gi|666912 (M93129) [Mycobacterium tuberculosis DNA sequence, complete cds.],gene products [Mycobacterium tuberc
g7e08a1.r1	159	8.5e-11	355 711	gi|3128287 (AF010496) hypothetical protein [Rhodobacter capsulatus]
Contig1443_d4b07a1.f1	187	8.6e-11	442 1059	sp|Q10327|YD72_S HYPOTHETICAL 97.1 KD PROTEIN C32A11.02C IN CHROMOSOME I>gi|1213266 (Z69796) unknown [Schizosaccharomy
Contig775_x5a07a1.f1	160	9.6e-11	368 619	gnl|PID|e123998 (X82490) unnamed protein product [Fusarium oxysporum]
n8c01a1.r1	180	1e-10	21 326	sp|P48236|YG3L_YEAST HYPOTHETICAL 51.6 KD PROTEIN IN RPL30B-RSR1 INTERGENICREGION >pir||S60440 probable membran
u4d06a1.f1	156	1.1e-10	344 508	gi|3094014 (AF060862) unknown [Homo sapiens]
d3d03a1.f1	174	1.5e-10	55 489	gnl|PID|e1316738 (AL031324) conserved hypothetical protein[Schizosaccharomyces pombe]
i3f03a1.r1	160	2e-10	78 458	sp|Q10058|YAM3_SCHPO HYPOTHETICAL 41.5 KD PROTEIN C1F5.03C IN CHROMOSOME I>gi|1103730 (Z68136) unknown [Schizos
m5f02a1.r1	177	2.2e-10	70 531	pir||S67175 probable membrane protein YOR273c - yeast (Saccharomycescerevisiae) >gnl|PID|e189400 (X896
m7h11a1.f1	153	2.3e-10	273 461	sp|P87151|YB0A_SCHPO HYPOTHETICAL 20.9 KD PROTEIN C25H2.10C IN CHROMOSOME II>gnl|PID|e316124 (Z95397) unknownS. pombe
h4a03a1.r1	162	2.4e-10	142 423	sp|P40055|YER2_YEAST HYPOTHETICAL 62.3 KD PROTEIN IN PTP3-ILV1 INTERGENICREGION >pir||S50585 hypothetical prote

m3a11a1.r1	177	4.2e-10	124 285	gnl|PID|e1263973 (AL022117) hypothetical protein [Schizosaccharomyces pombe]
Contig766_v1c09a1.r1	162	5.9e-10	201 449	sp|P38163|YBK6_Y HYPOTHETICAL 111.7 KD PROTEIN IN PKC1 5'REGION>pir||S45389 probable membrane protein YBL106c - yeast(
Contig746_p0h09a1.f1	148	7e-10	233 376	gi|431953 (X76302) nucleic acid binding protein [Homo sapiens]
n2g05a1.r1	151	1.3e-09	79 318	sp|O14256|YE6A_SCHPO HYPOTHETICAL 22.4 KD PROTEIN C6G10.10C IN CHROMOSOME I>gnl|PID|e334330 (Z98603) hypothetic
e9e04a1.r1	162	1.4e-09	182 526	gi|3158469 (AF067216) No definition line found [Caenorhabditis elegans]
m7c06a1.f1	155	1.5e-09	250 492	sp|Q06567|YL53_YEAST HYPOTHETICAL 65.9 KD PROTEIN IN SSP120-HAP1 INTERGENICREGION >pir||S59398 probable membran
m0b08a1.r1	151	1.7e-09	12 161	sp|P38263|YBV5_YEAST HYPOTHETICAL 41.2 KD PROTEIN IN YMC2-CMD1 INTERGENICREGION >pir||S48270 hypothetical prote
Contig1356_d5f05a1.f1	146	1.9e-09	172 540	sp|O14068|YEA3_S HYPOTHETICAL 13.9 KD PROTEIN C2E11.03C IN CHROMOSOME I>gnl|PID|e339159 (Z98850) hypothetical protein
l3c11a1.f1	153	2.5e-09	231 440	sp|P42846|YN48_YEAST HYPOTHETICAL 68.7 KD PROTEIN IN STB1-MCK1 INTERGENICREGION >pir||S51303 hypothetical prote
Contig76_l3e08a1.f1	143	2.6e-09	243 467	sp|Q07549|YD23_Y HYPOTHETICAL 15.7 KD PROTEIN IN UBP1-HNT1 INTERGENICREGION >pir||S67666 probable membrane protein YDL
Contig1655_e9e11a1.f1	172	3e-09	698 1213	pir||S61185 hypothetical protein YDR299w - yeast (Saccharomyces cerevisiae)>gi|849214 (U28374) Ydr299wp [Saccharo
Contig308_g6e08a1.r1	168	3.7e-09	247 645	gi|2558956 (AF025475) Masc1 [Ascobolus immersus]
n5a02a1.r1	167	3.9e-09	66 509	sp|P38787|YHM3_YEAST HYPOTHETICAL 42.8 KD PROTEIN IN VMA22-RRP3 INTERGENICREGION >pir||S46711 hypothetical prot
Contig363_m2f12a1.f1	142	4.7e-09	327 590	sp|Q09730|YA4D_S HYPOTHETICAL 10.5 KD PROTEIN C31A2.13C IN CHROMOSOME I>pir||S59647 hypothetical protein SPAC31A2.13c
r4b10a1.f1	166	5.7e-09	327 656	gnl|PID|e223969 (Z69793) R03A10.3 [Caenorhabditis elegans]
j9f03a1.f1	159	7e-09	245 502	gnl|PID|d1022302 (AB004539) pi074 [Schizosaccharomyces pombe]>gnl|PID|e1250311 (AL021766) hypothetical prot
Contig696_p0a01a1.f1	139	7e-09	152 370	gi|2621836 (AE000853) conserved protein [Methanobacterium thermoautotrophicum]
Contig393_f1g11a1.f1	155	8.9e-09	442 681	gnl|PID|d1026035 (AB011825) CSH3 [Schizosaccharomyces pombe] >gnl|PID|e1263963(AL022117) hypothetical protein [Schizos
Contig277_g9e11a1.f1	153	1.4e-08	67 408	gnl|PID|e1292630 (AL023534) hypothetical protein [Schizosaccharomyces pombe]
Contig373_g2a02a1.r1	157	2e-08	72 473	gnl|PID|d1011269 (D64004) hypothetical protein [Synechocystis sp.]
j4h11a1.r1	135	2e-08	149 364	sp|Q10481|YDF9_SCHPO HYPOTHETICAL 10.7 KD PROTEIN C17C9.09C IN CHROMOSOME I>gnl|PID|e241761 (Z73099) hypothetic
x1g08a1.r1	141	2.2e-08	17 271	sp|P53938|YNI0_YEAST HYPOTHETICAL 41.7 KD PROTEIN IN PMS1-TPM1 INTERGENICREGION >pir||S53898 probable membrane
n3g12a1.r1	134	2.3e-08	122 319	pir||S66709 probable membrane protein YOL026c - yeast (Saccharomycescerevisiae) >gnl|PID|e252264 (Z747
g7e07a1.r1	162	2.4e-08	191 427	gi|939724 (U30858) putative sensor kinase; regulatory protein

				for production of antifungal antibiotic
e0a10a1.f1	142	3.5e-08	182 472	gnl|PID|e1295823 (AL023780) hypothetical protein [Schizosaccharomyces pombe]
x8f04a1.r1	146	5.1e-08	75 473	sp|Q04958|YMF9_YEAST HYPOTHETICAL 187.1 KD PROTEIN IN OGG1-CNA2 INTERGENICREGION >pir||S49802 probable membrane
Contig385_n3c12a1.r1	157	5.8e-08	559 813	sp|P53219|YG1L_Y HYPOTHETICAL 38.5 KD PROTEIN IN ERV1-GLS2 INTERGENICREGION >pir||S64322 probable membrane protein YGR
Contig72_13g10a1.f1	146	5.9e-08	134 463	gnl|PID|d1013422 (D85230) hypothetical protein [Plectonema boryanum]
t2f11a1.r1	136	6.6e-08	117 284	sp|P35735|YKF1_YEAST HYPOTHETICAL 40.5 KD PROTEIN IN NUP120-CSE4 INTERGENICREGION >pir||S37873 hypothetical pro
Contig1343_c1d09a1.f1	150	7e-08	157 813	gi|2190955 (AF002247) ORF4; Putative transmembrane protein [Rhodococcuserythropolis]
g8b06a1.f1	155	9.7e-08	318 581	pir||S66834 probable membrane protein YOL137w - yeast (Saccharomycescerevisiae) >gnl|PID|e252306 (Z748
z4a08a1.r1	137	1e-07	184 504	gnl|PID|d1017526 (D90900) hypothetical protein [Synechocystis sp.]
Contig935_13a07a1.f1	132	1.5e-07	288 461	sp|O13802|YE05_S HYPOTHETICAL 37.8 KD PROTEIN C17H9.05 IN CHROMOSOME I>gnl|PID|e334275 (Z98597) hypothetical protein [
Contig1444_c1f02a1.f1	124	2.7e-07	398 493	pir||S12206 hypothetical protein 2 (rRNA external transcribed spacer) - mouse
Contig497_d0c12a1.f1	152	3e-07	349 741	sp|Q09844|YAE3_S HYPOTHETICAL 54.3 KD PROTEIN C23D3.03C IN CHROMOSOME I>pir||S62494 hypothetical protein SPAC23D3.03c
Contig397_f0g10a1.f1	136	3.1e-07	201 329	gnl|PID|e340013 (Z98977) hypothetical protein [Schizosaccharomyces pombe]
y9d06a1.r1	128	3.3e-07	177 416	pir||S69568 hypothetical protein YDR511w - yeast (Saccharomyces cerevisiae)>gi|927780 (U33057) Ydr511w
g9b06a1.f1	157	3.5e-07	197 562	gnl|PID|e1314595 (AL031187) putative protein [Arabidopsis thaliana]
Contig897_v7g01a1.f1	154	3.8e-07	618 1025	sp|Q10218|YAYB_S HYPOTHETICAL 89.2 KD PROTEIN C4H3.11C IN CHROMOSOME I>gi|1184024 (Z69380) unknown [Schizosaccharomyce
m8a03a1.r1	145	4.2e-07	6 170	sp|P42620|YQJG_ECOLI HYPOTHETICAL 37.4 KD PROTEIN IN EXUR-TDCC INTERGENICREGION (O328) >gi|606043 (U18997) ORF_
n8f02a1.f1	130	5.7e-07	168 485	sp|O14011|YDP8_SCHPO HYPOTHETICAL 54.2 KD TRP-ASP REPEATS CONTAINING PROTEINC29A4.08C IN CHROMOSOME I >gnl|PID|
Contig1728_c7g04a1.f1	148	6e-07	283 705	gnl|PID|e339966 (Z98981) hypothetical protein [Schizosaccharomyces pombe]
i7e09a1.f1	143	9e-07	324 623	sp|Q09875|YAGC_SCHPO HYPOTHETICAL 35.8 KD PROTEIN C12G12.12 IN CHROMOSOME I>pir||S62543 hypothetical protein SP
Contig1783_c5d01a1.f1	152	1.2e-06	1149 1598	sp|P39992|YEC3_Y HYPOTHETICAL 78.3 KD PROTEIN IN RIP1-URA3 INTERGENICREGION >pir||S50436 hypothetical protein YEL023c
a1d05f2.f1	141	1.3e-06	89 319	gnl|PID|e1251050 (AL021817) hypothetical protein [Schizosaccharomyces pombe]
Contig1414_m5g11a1.f1	137	1.4e-06	195 500	gnl|PID|d1031256 (AP000005) 223aa long hypothetical protein [Pyrococcushorikoshii]
a0e05a1.r1	120	1.5e-06	136 456	sp|Q10430|YDD5_SCHPO HYPOTHETICAL 28.8 KD PROTEIN C1B9.05C IN CHROMOSOME

				I>gnl|PID|e235479 (Z70720) unknown [Sc
j0a03a1.f1	128	1.7e-06	247 420	sp|P38170|YBJ7_YEAST HYPOTHETICAL 83.0 KD PROTEIN IN ATP1-ROX3 INTERGENICREGION >pir||S45403 hypothetical prote
Contig1685_j9f07a1.f1	139	2.1e-06	538 816	pir||S61199 hypothetical protein YDR313c - yeast (Saccharomyces cerevisiae)>gi|849227 (U28374) YDR313C gene produ
c9d11a1.r1	151	2.2e-06	39 296	gnl|PID|d1010130 (D50929) The KIAA0139 gene product is related to mousecentrosomin B. [Homo sapiens] >gi|18
w5f12a1.r1	129	2.3e-06	3 236	sp|P48231|YNI7_YEAST HYPOTHETICAL 132.5 KD PROTEIN IN TOP2-MKT1 INTERGENICREGION >pir||S57535 probable membrane
c5g07a1.r1	142	2.6e-06	8 619	gnl|PID|e317345 (Z95620) unknown [Schizosaccharomyces pombe]
i0e06a1.r1	121	2.7e-06	274 447	gnl|PID|e1292638 (AL023534) hypothetical protein [Schizosaccharomyces pombe]
h1a09a1.r1	139	2.8e-06	222 386	sp|P47095|YJZ4_YEAST HYPOTHETICAL 27.4 KD PROTEIN IN MER2-CPR7 INTERGENICREGION >pir||S57042 hypothetical prote
Contig888_v3h04a1.f1	136	5e-06	419 505	gi|1870215 (AC000133) ORF [Emericella nidulans]
Contig1194_w6g09a1.f1	122	5.9e-06	481 642	gi|940146 (U30501) orf2; Method: conceptual translation supplied by author.[Thermotoga maritima]
n8d09a1.r1	136	6.9e-06	33 158	pir||S67247 hypothetical protein YOR338w - yeast (Saccharomyces cerevisiae)>gnl|PID|e223200 (X95720) O
o8h04a1.r1	127	1.5e-05	84 374	gi|310604 (L19300) ORF3 [Staphylococcus aureus]
a0a09a1.f1	120	2.2e-05	67 444	gnl|PID|e1295819 (AL023780) hypothetical protein [Schizosaccharomyces pombe]
r4h12a1.r1	128	3.9e-05	76 318	gnl|PID|e1250039 (AL021748) hypothetical protein [Schizosaccharomyces pombe]
e7d05a1.f1	132	4.7e-05	334 522	pir||S66834 probable membrane protein YOL137w - yeast (Saccharomycescerevisiae) >gnl|PID|e252306 (Z748
Contig1133_j7g02a1.f1	129	5.6e-05	296 556	gi|1109800 (U41528) C15C7.1 gene product [Caenorhabditis elegans]
Contig1706_c8b03a1.f1	124	5.8e-05	509 853	gnl|PID|e349611 (Z99262) hypothetical protein [Schizosaccharomyces pombe]
l3d09a1.r1	106	6.1e-05	127 279	gnl|PID|e324207 (Z97185) hypothetical protein [Schizosaccharomyces pombe]
Contig297_g7d08a1.r1	128	0.00011	621 815	sp|Q09875|YAGC_S HYPOTHETICAL 35.8 KD PROTEIN C12G12.12 IN CHROMOSOME I>pir||S62543 hypothetical protein SPAC12G12.12
w7f03a1.f1	128	0.00011	29 217	pir||S67290 probable membrane protein YOR378w - yeast (Saccharomycescerevisiae) >gnl|PID|e252199 (Z752
t2h07a1.f1	109	0.00011	186 452	sp|P53220|YG1M_YEAST HYPOTHETICAL 27.2 KD PROTEIN IN GLS2-RPL26B INTERGENICREGION >pir||S64324 probable membran
Contig698_r5a07a1.f1	126	0.00013	694 939	sp|P53267|YG2Y_Y HYPOTHETICAL 37.8 KD PROTEIN IN CLB6-SPT6 INTERGENICREGION >pir||S64421 hypothetical protein YGR113w
Contig1769_e9g10a1.r1	133	0.00015	395 580	gnl|PID|e1293563 (AL023634) hypothetical protein [Schizosaccharomyces pombe]
d1b10a1.r1	134	0.00016	43 477	gnl|PID|e349693 (Z99296) hypothetical protein [Schizosaccharomyces pombe]

Contig280_g9c05a1.f1	122	0.00026	312	500	gnl|PID|e349324 (Z99264) T05F1A.2 [Caenorhabditis elegans] >gnl|PID|e1247195(Z81586) T05F1.a [Caenorhabditis elegans]
Contig1164_g2a01a1.f1	118	0.00038	2	373	sp|P40526|YIE9_Y HYPOTHETICAL 30.3 KD PROTEIN IN GPP1-SYG1 INTERGENICREGION >pir||S48430 probable membrane protein YIL

VIII. No significant homolog

<NONE>

-949 Contigs

-1419 Singlets