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.

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.

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

UMI®

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/304

UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

DEVELOPMENT AND ANALYSIS OF AN EXPRESSED SEQUENCE TAG DATABASE FROM *FUSARIUM SPOROTRICHIOIDES*

A Dissertation
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy

By
QUN REN
Norman, Oklahoma
2001

UMI Number: 3005138

UMI®

UMI Microform 3005138

Copyright 2001 by Bell & Howell Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

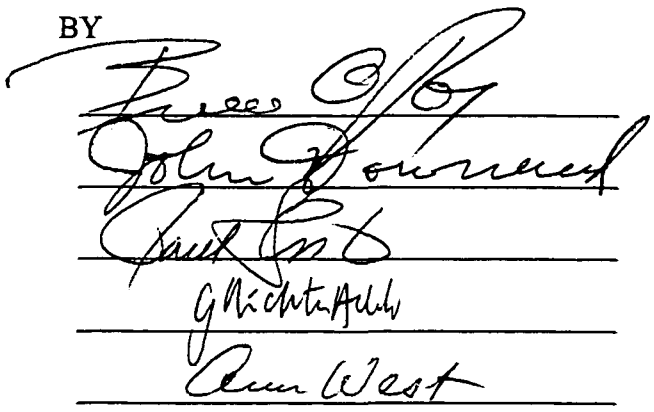
Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

© Copyright by Qun Ren 2001
All rights reserved.

DEVELOPMENT AND ANALYSIS OF AN EXPRESSED SEQUENCE TAG DATABASE FROM *FUSARIUM SPOROTRICHIOIDES*

A Dissertation APPROVED FOR THE
DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

BY



Acknowledgements

I am very grateful to my committee members, Dr. Bruce A. Roe, Dr. Paul F. Cook, Dr. Ann H. West, Dr. George B. Richter-Addo and Dr. John S. Downard for their kind help during my study here. Dr. Cook was particularly helpful throughout my program in providing encouragement and insights. I also wish to thank Dr. Phil E. Klebba for his help when I prepared for my general exam.

My special thanks go to Dr. Roe, my major professor, who has always provided constant encouragement, excellent advice, and generous financial support for my study here over the past five years. I also appreciate his patience and challenge to me.

I wish to thank everyone in Dr. Roe's lab for giving me advice, support and help. Thanks especially to Mueed Ahmad, Dennis Burian, Linda Cantu, Feng Chen, Lingzhi Chu, Sandra Clifton, Stephane Deschamps, Sara Downard, An Do, Trang Do, Angela Dorman, Mounir Elharam, Fang Fang, Ying Fu, Kaylynn Hale, Jennifer Gray, Li Hang, Jennifer Hausner, Ping Hu, Xiaohong Hu, Axin Hua, Honggui Jia, Emily Huang, Steve Kenton, Doris Kupfer, Hongshing Lai, Sean Meadows, Lisa Lane, Victoria Lao, Christopher Lau, Sharon Lewis, Shaoping Lin, Phobe Loh, Eda Malaj, Rose Morales-Diaz, Wendy Martin, Jami Milam, Fares Najar, Thuan Nguyen, Shelly Oommen, Huaqin Pan, Andy Peterson, Yudong Qian, Sulan Qi, Murli Rao, Lin Song, Jing Tian, Runying Tian, Yonathan Tilahun, Qiaoyan Wang, Ying-Ping Wang, Doug White, Rusty Wayt, Zhili Wang, Jim White, Mary Catherine Williams, Dixie Wishnuck, Tammy Womack, Heather Wright, Hui Wu, Limei Yang, Ziyun Yao, Xuling Yuan, Min Zhan, Guozhong Zhang and Hua Zhu.

I greatly appreciate my parents, Qiyu Ren and Yiying Zhou, my brother Min Ren, and my sister-in-laws Zhaomei Zhang and Lanying Han. Their love and help enabled my husband and I to come to the United States of America to study.

Finally, I dedicate this dissertation to my husband, Zhaojie Zhang and my daughter, Renfei Zhang, who gave me constant support, encouragement and love.

Table of Contents

List of Figures x
List of Tables xii
List of Abbreviations xiv
Abstract xviii
Chapter I. Introduction 1
1.1 DNA,RNA and Genes 1
1.1.1 DNA 1
1.1.2 RNA 4
1.1.3 Genes 5
1.1.3.1 Definition of gene 5
1.1.3.2 Classification of genes 5
1.1.3.3 Structures of genes 6
1.1.3.4 Pseudogenes 7
1.2 mRNA 8
1.2.1 Splicing 8
1.2.2 5' cap 8
1.2.3 3' poly(A) tail 10
1.2.4 Mature mRNA 10
1.3 cDNA library 11
1.3.1 Advantages of constructing an cDNA libray 11
1.3.2 General procedures for construction of cDNA library 12
1.3.3 The λ ZAP system as the vector for cDNA library 15
1.4.EST and UniGene databases 16
1.4.1 Beginning of large scale EST sequencing 17
1.4.2 Recent advances in large scale EST sequencing 17
1.4.3 3'EST and 5'EST 19
1.4.3.1 5'EST is more gene family specific 19
1.4.3.2 3' EST is more gene specific 22
1.4.4 UniGene database 24
1.4.5 Applications of EST and UniGene database 25
1.5 DNA sequencing 27
1.5.1 Basic methods for DNA sequencing 27
1.5.2 Improvements on Sanger method 28
1.5.3 DNA sequencing instruments 31
1.5.4 Shortgun strategy for large scale DNA sequencing 32
1.6 Sequence analysis software and data submission web site 33
1.6.1 Phred and Phrap 33
1.6.2 Cross_match, Dotter, Bestfit, Gap and Sim4 34
1.6.3 Consed 34
1.6.4 BLAST 35
1.6.5 Powerblast 36
1.6.6 Genscan and Xgrail, NNPP, Promoter 2.0 and Repeatmasker 36
1.6.7 GenBank, Entrez, Sequin and dbEST 37
1.7 *Fusarium sporotrichioides* cDNA library 39

1.8 Inborn gap in the published megabase sequence of the Ig light chain genes..... 44
1.9 William Syndrome is associated with deletion of chromosome band 7q11.23... 46

Chapter II. Materials and Methods (Part 1)-Sequence and analysis of ESTs... 48
2.1 Construction of cDNA library..... 48
2.1.1 Source of cDNA library..... 48
2.1.2 cDNA synthesis..... 48
2.1.2.1 Synthesis of the first strand cDNA..... 48
2.1.2.2 Synthesis of the second strand cDNA..... 49
2.1.3 Insert the cDNA into the vector and amplify the plasmid..... 51
2.1.4 Amplification of the cDNA clones..... 52
2.2 Growth of cDNA clones and isolation of templates..... 53
2.2.1 Growth of cDNA clones..... 53
2.2.2 Isolation of cDNA templates..... 53
2.3 Sequencing of ESTs..... 55
2.3.1 Set up reactions..... 55
2.3.2 Incubate reactions..... 56
2.3.3 Purify reactions..... 56
2.3.4 Collect data..... 56
2.4 Computer programs used in data analysis..... 57
2.4.1 Phred, Phrap and Cross_match..... 57
2.4.1.1 Use of Phred to generate .fasta file..... 58
2.4.1.2 Use of cross_match to screen out vector sequence and generate fasta.screen file..... 60
2.4.1.3 Algorithm used by Phred to calculate quality values..... 61
2.4.1.4 Use of Phred to generate .screen.qual file and .phd files..... 63
2.4.1.5 Use of Phrap to assemble the sequences..... 65
2.4.2 Consed..... 67
2.4.2.1 Display error rate of the contiguous sequence..... 68
2.4.2.2 Viewing quality of each read and consensus sequence..... 68
2.4.2.3 Viewing assembly information..... 70
2.4.2.4 Navigating to the weak region..... 70
2.4.2.5 Viewing and comparing traces..... 71
2.4.2.6 Aligning and comparing two region..... 71
2.4.2.7 Picking primers..... 71
2.4.2.8 Search for homologous sequence regions..... 72
2.4.3 BLAST..... 72
2.4.3.1 The parameter settings..... 73
2.4.3.2 Query sequence format..... 75
2.4.3.3 Interpreting sequence identifiers..... 75
2.5 EST data analysis..... 76
2.5.1 Clip and Clean..... 76
2.5.2 Cumulative 3' assembly of the EST database..... 80
2.5.3 3' and 5' assembly of the EST database..... 80
2.5.4 Biological function assignments..... 81
2.5.4.1 BlastX search output and primary keyword list..... 82

2.5.4.2 Edit the keyword list 83
2.5.4.3 Print the final form 84
2.5.5 tBlastX against dbEST 84
2.6 Full length cDNA sequence and submission 89
2.6.1 Picking primers for gap closures for full-length cDNA sequences 89
2.6.2 Analysis and submission of cDNA sequences 90

Chapter III. Material and Methods (Part 2)-Sequence and analysis of human genome projects 91
3.1 Overview of the Sequencing Strategy 91
3.2 Large scale DNA isolation 92
3.2.1 Culture the cells 92
3.2.2 Isolation target DNA 93
3.2.3 Purification of DNA 94
3.2.4 Degradation of RNA 94
3.2.5 Phenol extraction 94
3.3 Shotgun library construction 95
3.3.1 Physical shearing 95
3.3.2 End repair and size fractionation 96
3.3.3 Ligation 96
3.3.4 Transformation 97
3.3.5 Growth of the shotgun library 98
3.4 Isolation and sequencing the sub-clone library 99
3.5 Sequence assembly, data analysis and submission 100
3.6 Gap closure, proofreading, finishing and submission 101
3.6.1 Re-sequence 102
3.6.1.1 Re-sequence the clone 102
3.6.1.2 Re-sequence the contig ends 102
3.6.2 Primer walking 103
3.6.3 PCR (polymerase chain reaction) 105
3.6.4 MultiPlex PCR-based Method 107
3.6.5 Proofreading, finishing and submission 108

Chapter IV. EST data analysis and Discussion 109
4.1 Estimating EST quality 109
4.2 Estimating library redundancy 111
4.3 Genes represented in EST database 113
4.3.1 Number of genes identified in the database 113
4.3.2 Gene expression level 117
4.3.3 Top 12 highly expressed genes 118
4.3.3.1 Top 12 highly expressed genes in *F. sporotrichioides* 118
4.3.3.2 Comparison of the highly expressed genes from different cDNA libraries 124
4.3.4 Gene expression levels for each trichothecene biosynthetic gene 126
4.4 Submission of ESTs to dbEST 130
4.5 Biological function assignments 131

4.5.1 The last version of keyword list and the form of biological function assignments 131
4.5.2 Outline of biological function assignments 134
4.5.3 Comparison of outline of biological function assignments from four EST databases 135
4.6 tBlastX against dbEST 142
4.6.1 tBlastX against dbEST for members with no BlastX hit against nr 142
4.6.2 tBlastX against dbEST for members with BlastX hit against nr 143
4.7 Submission of full length cDNA sequences to GenBank 144

Chapter V. Analysis of contigs of human genome sequences 152
5.1 Analysis of 322f3 152
5.1.1 9 kb deletion in AC009286 152
5.1.2 Repeat elements in AC009286 155
5.1.3 Gene prediction in AC009286 156
5.1.4 Significance of discovering the phenotypically silent inborn gaps. 158
5.2 Analysis of 239c10 159
5.2.1 Human neutrophil cytosol factor 1 (NCF1) gene 159
5.2.2 Human *hPMS* gene 166
5.2.2.1 Human *hPMS* gene family 166
5.2.3.2 239c10 has homologues to three members in HMS gene family 167
5.2.3 Human Bruton's tyrosine kinase-associated protein-135 gene (BAP-135) 174
5.2.4 Human prohibitin pseudogene 179
5.2.5 Repetitive elements in 239c10 180

Chapter VI Conclusion 185
Literature cited 193
Appendix I Keyword list of *Fusarium sporotrichioides* categories of biological functions 212
Appendix II *Fusarium sporotrichioides* EST biological function assignments 247

List of Figures

Fig. 1.1	Flow of genetic information from DNA to RNA to protein	2
Fig. 1.2	DNA structure showing the antiparallel nature of DNA and base pairings between Adenine (A) and Thymine (T) and between Guanine (G) and Cytosine (C)	3
Fig. 1.3	The structure of the 5' cap of eukaryotic mRNAs	9
Fig. 1.4a	Construction of a cDNA clone	13
Fig. 1.4b	Construction of a cDNA clone	14
Fig. 1.5	3' EST and 5' EST	18
Fig. 1.6	A set of cDNA inserts from the same gene with different insert lengths and cDNA consensus construction by 3' EST and 5' EST assemblies	21
Fig. 1.7	Improved Sanger dideoxy method for DNA sequencing	29
Fig. 1.8	Spores of *Fusarium sporotrichioides*	41
Fig. 1.9	Examples of 4 trichothecenes sharing the same trycyclic nucleus	42
Fig. 1.10	Trychothecene biosynthetic pathway in *Fusarium* species	42
Fig. 1.11	World distribution of the deletion phenotype as detected by the R271 pair of primers	45
Fig. 2.1	Synthesis of a cDNA with different cohesive ends	50
Fig. 2.2	View quality of each read and consensus sequence with consed	69
Fig. 2.3	Steps in biological function assignments	82
Fig. 4.1	Cumulative 3' EST assembly for *F. sporotrichioides* library	115
Fig. 4.2	Biological function classification	140
Fig. 4.3	Comparison of biological function classification results	141
Fig. 4.4	tBlastX homologues against both *A. nidulans* and *N.crassa* EST databases	151
Fig. 5.1	Diagrammatic sketch of chromosome 22 showing from left to right the region 22q11.2, the Igλ V and IgλJ&C gene regions, the region containing clone D86999 and the newly sequenced clone AC009286	153
Fig. 5.2	Dotter plotting comparison between BAC 322f3 (AC009286) and the human chromosome 22 reference sequence (NT_001454)	154
Fig. 5.3	Gene prediction, repeat elements and GC level in AC009286	157
Fig. 5.4	Gene prediction and annotation of AC004166	160
Fig. 5.5	Sim4 alignment of the human NCF1 mRNA and genomic sequence AC004166	163
Fig. 5.6	Comparison of four hPMS mRNA sequences using Pileup from GCG	170
Fig. 5.7	GAP align of BAP-135 and SPIN protein sequences	177
Fig. 5.8	Sim4 alignment of human prohibitin mRNA and AC004166	181
Fig. 5.9	Gene prediction and repeat elements and GC level in AC004166	184

List of Tables

Table 1.1	Part of the EST data released on 04-13-2001 on NCBI's dbEST web site	20
Table 1.2	Primers used in detecting the presence or absence of the amplification products for all deletions encountered at the 22q11.2	46
Table 2.1	Output files in edit_dir for *F. sporotrichioides* EST project after using Phred, Phrap and cross_match	58
Table 2.2	A file in FS.fasta for a 5' EST in fasta format	59
Table 2.3	A file in FS.fasta.screen for a 5' EST after screen out the vector sequence	61
Table 2.4	One of the phd.1 file in phd_dir	64
Table 2.5	A file in FS.fasta.screen.qual for quality	65
Table 2.6	The syntax of sequence identifiers used by the NCBI BLAST	75
Table 2.7	Clip and Clean EST processing scripts (Adapted from Kupfer, 1999)	78
Table 2.8	Mitochondrial and ribosomal sequences used in Clip and Clean processes	79
Table 2.9.	A portion of the BlastX results for Contig 1002 showing that many BlastX hits listed in score order	85
Table 2.10	Result of Contig1002 in file "keywordhits" showing that one contig has several keyword hits	86
Table 2.11	A portion of file "nonkeywordhits"	87
Table 2.12	A portion of the file "nonkeywordhits" after adding the new key words in the angle bracket	88
Table 4.1.	*F. sporochioides* cDNA library and EST database quality summary	111
Table 4.2.	Comparing qualities for EST databases from different laboratories.	111
Table 4.3	The determination of library redundancy by cumulative 3' EST assemblies	114
Table 4.4	3' EST and 5' EST Phrap assembly and BlastX results summary	116
Table 4.5	Cluster sizes and the frequency of clusters in each size after 3' and 5' EST assembly	119
Table 4.6	Summary of gene expression level in *F.sporotrichioides* cDNA library	119
Table 4.7	Comparison of gene expression levels between *F.sporotrichioides* and *A. nidulans* cDNA libraries	119
Table 4.8	Top ten highly expressed gene members in *F.sporotrichioides* cDNA library	120
Table 4.9	Heat shock protein in *F. sporotrichioides* EST database	126
Table 4.10	Gene expression level for trichocheecene biosynthesis genes	128
Table 4.11	An EST homology file	132
Table 4.12	Three files required for batch submission of EST to dbEST	133

Table 4.13 Examples of report from dbEST 133
Table 4.14 Distribution of sequence lengths in three projects 134
Table 4.15 Outline of the categories of biological functions and number of database members falling in each category 135
Table 4.16 Percentage of database members falling in each of the biological function categories and percentage of those without known homology 139
Table 4.17 Comparison of biological function classification results from 4 EST databases at ACGT 142
Table 4.18 tBlastX result showing the ESTs which had homologs in *A. nidulus* database 144
Table 4.19 tBlastX result showing the ESTs which had homologs in *N. crassa* database 144
Table 4.20 tBlastX result showing the ESTs which had homologs in both *A. nidulus* and *N. crassa* databases 145
Table 4.21 Database members which have both BlastX hits against non-redundant protein (nr) database and tBlastX hits against *A. nidulans* EST database(Total: 341) 145
Table 4.22 Database members which have both BlastX hits against non-redundant protein (nr) database and tBlastX hits against *N. crassa* EST database (Total: 202) 147
Table 4.23 Database members which have BlastX hits against non-redundant protein (nr) database and tBlastX hits against both *A. nudulans* and *N. crassa* EST databases (Total: 201) 148
Table 4.24 Summary of tBlastX results 149
Table 4.25 The percentage of members falling in each of the biological function categories for the group of ESTs which had BlastX homologs against nr and tBlastX homologs against both *A. nidulus and N. crassa* databases 150
Table 4.26 The report containing "GenBank_Accn" for submitted cDNA 150

Table 5.1 Repetitive elements in AC009286 (BAC 322f3) 155
Table 5.2 The ranges and sizes of exons and introns of the NCF1 gene 162
Table 5.3 Amino acid sequence of NCF1 (protein_id=AAA57209.1) 165
Table 5.4 Members in human *PMS* gene family 167
Table 5.5 The ranges and sizes of exons and introns of the hPMS4 gene 169
Table 5.6 The ranges and sizes of exons and introns of the hPMS2L13 gene 169
Table 5.7 The ranges and sizes of exons and introns of the hPMS7 gene 167
Table 5.8 The ranges and sizes of exons and introns of the BAP-135 gene 176
Table 5.9 The ranges and sizes of exons and introns of the SPIN gene 176
Table 5.10 The ranges and sizes of exons and introns of prohibitin gene 180
Table 5.11 Repetitive elements in AC004166 (BAC 239c10) 183

List of Abbreviations

2xTY	-medium containing tryptone, yeast extract and salt
Aspergillus nidulans	*- A. nidulans*
ABI	-Applied Biosystems Incorporated
ACGT	-Advanced Center for Genome Technology, University of Oklahoma, Norman, OK
ALU	-a human DNA repeat element with homology to 7SL RNA
BAC	-bacterial artificial chromosome
BLAST	-basic local alignment search tool
bp	-base pair
CCD	-charge couple device
DNA	-deoxyribonucleotide acid
cDNA	-complementary deoxyribonucleic acid
Consed	-consensus editor. Graphical user interface for use with databases generated by the Phred/Phrap assembly program
contig	-a continuous sequence fragment assembled from the end-sequence reactions of several subclone
CpG islands	-a region with the higher than statistically expected occurrence of the CG dinucleotide
dbEST	-a Genbank database of ESTs
ddNTP	-dideoxynucleotide
ddATP	-2', 3'-deoxy-adenosine triphosphate
dGTP	-2'-deoxy-guanosine triphosphate

dITP -2'-deoxy-inosine triphosphate

DMSO -dimethyl sulfoxide

DNA -deoxyribonucleic acid

DNase -deoxyribonuclease

dNTP -deoxynucleotide

Escherichia coli - *E. coli*

EDTA -disodium ethylenediamine tetraacetate

EMBL -European Molecular Biology Laboratory

EST -expressed sequence tag

Fusarium sporotrichioides - *F. sporotrichioides*

FISH -fluorescence *in situ* hybridization

GCG -a software package from the Genetic Computer Group in Madison, WI

GenScan -an exon and gene prediction program

GRAIL -the Gene Recognition and Analysis Internet Link program for gene prediction

GTE -buffer containing glucose, Tric-HCl and EDTA

HIV -human immunodeficiency virus

HMG-CoA synthase -hydroxymethylglutaryl-CoA synthase

HSP -High-score Segment Pair

IPTG -isopropyl-β-D-thiogalactopyranoside

kb -kilobase

lacZ -gene coding for β-galactosidase

LB -medium containing tryptone, yeast extract and salt

LINE(L1) -a long interspersed repeat element. A human DNA repeat element

MER -a medium element of repeat. A human DNA repeat element

MIR -mammalian-wide interspersed repeats

MLT -a human repeat element

mRNA -messenger ribonucleotide acid

NCBI -National Center for Biotechnology Information

NCF1 -neutrophil cytosolic factor 1

NLM -National Library of Medicine

NNPP -Neural Net Promoter Prediction program

Northern blot -a probing of cellular RNA

Neurospora crassa - *N. crassa*

ORF -open reading frame

PCR -polymerase chain reaction

PEG -poly(ethylene glycol)

Phred/Phrap -Phil's read editor and assembly program for DNA sequences

PUC -the plasmid vector used to produce the shotgun sequence library

RACE -rapid amplification of cDNA ends

RAD -representational differential analysis

RepeatMasker -a computer program that finds interspersed repeat elements in genomic DNA sequence

RNA -ribonucleic acid

RNase -ribonuclease

rRNA -ribosomal ribonucleic acid

SDS	-sodium dodecyl sulfate
snRNA	-small nuclear RNA
STS	-sequence tagged site
SwissProt	-a protein database collection
Taq	-*Thermus aquaticus*
TB	-terrific broth
TBE	-buffer containing Tris-HCl, boric acid and EDTA
TE	-buffer containing Tris-HCl and EDTA
TED	-trace editor An editor for DNA sequencing trace data
TM	-melting temperature or buffer containing Tris-HCl and magnesium
Tri gene	-the trichothecene biosynthetic gene
tRNA	-transfer ribonucleic acid
UTR	-untranslated region
UV	-ultraviolet
WMS	-William Syndrome
X-gal	-5-bromo-4-chloro-3-indolyl-b-D-galactoside
XGAP	-X windows based gene recognition and analysis internet link used for sequence feature prediction
XGrail	-X windows based gene recognition and analysis internet link. Used for sequence feature prediction
YENB	-medium containing yeast extract and nutrient broth

ABSTRACT

A *Fusarium sporotrichioides Tri10* over-expressed cDNA library was sequenced and its EST database was constructed in this dissertation research. The EST database was made publicly available by submission to GenBank dbEST and publication on the ACGT web site. This database serves as a foundation for annotating and understanding gene expression in *F. sporotrichioides* and related fungi.

During the *F. sporotrichioides* EST project, 7495 high quality ESTs were obtained. The high quality ESTs were assembled using Phrap and then analyzed by a BlastX homology search against the GenBank non-redundant protein database. In total, 2181 singlets and 1057 contigs were obtained in the assembled database and 2139 genes were represented in this database. Several computer programs have been used to enable a semi-automated process of biological function assignments for each *F. sporotrichioides* EST database member that has a significant BlastX homologue. 50% of the ESTs that had significant homologues in GenBank non-redundant protein database were placed into the seven Riley categories. 50% of the ESTs had no significant homologues in GenBank non-redundant protein database. They may represent undiscovered genes.

To date, twelve genes in the trichothecene biosynthesis pathway have been identified in *F. sporotrichioides*. Eleven of the twelve genes products were found in the *F. sporotrioides* EST database. In total, 541 ESTs, 7.22% of the total 7495 ESTs in the database, represented genes involved in trichothecene biosynthesis pathway. Two of the twelve genes are genes newly defined during this EST projects by our collaborators and three other genes that have the subtle *Tri* patterns are being studying further.

The comparison of four other fungal cDNA libraries sequenced in our lab, *F. sporotrichioides* library, two *Neurospora crassa* libraries and one *Aspergillus nidulans* library, shows that the percentage of biological function divisions in each library is remarkably similar in spite of the diverse nature of the libraries. Using the group of ESTs that have no significant homologues to GenBank non-redundant protein database from *F. sporotrichioides* database to perform tBlastX against dbEST, it was found that there are 27 singlets/clusters that have homologs with both *A. nidulus* and *N. crassa* ESTs. These 27 new genes that present in all four libraries are valuable candidate unknown genes for further studies.

In this dissertation research, seven BACs, PAC and cosmids from the Human Genome Project also were sequenced to contiguity with an individual base error rate of less than 1 error every 10,000 bases. Among them, two BAC sequences were thoroughly analyzed. Sequence of BAC 322f3 revealed the sequence of the 9098 bp absent in the Ig λ region of the reference sequence for human chromosome 22. The discovery of this phenotypically silent inborn gap was a trigger to make a plea to search for deletion polymorphism through genome scans in population. BAC 239c10 covering the William Syndrome region encoded three genes, human neutrophil cytosol factor 1 (NCF1) gene, human hPMS gene and human Bruton's tyrosine kinase-associated protein-135 (BAP-135) gene. NCF1 gene and hMSP gene have been mapped to 7q11.23 before, but BAP-135 has not been mapped to 7q11.23 in the previous studies. A pseudogene of the human prohibitin gene, which is related to breast cancer and has been mapped to 17q21, also was found in BAC 239c10 sequence.

Chapter I

Introduction

1.1 DNA, RNA and Genes

1.1.1 DNA

According to the Central Dogma of Molecular Biology (Crick, 1958), the genetic information flows from DNA to RNA to protein, except in retroviruses where it flows from RNA to DNA. DNA is the template for replication of itself and transcription of RNAs, and mRNA serves as the template for translation of the protein (Fig. 1.1).

DNA, and in some instances, RNA, is the hereditary material. DNA is a polymer of deoxyribonucleotide units (Fig. 1.2). Each nucleotide is composed of a base, a sugar, and a phosphate group. The sugar in a deoxyribonucleotide is deoxyribose. Four types of bases are present in DNA. Two of them are purines: adenine (A), guanine (G), while the other two are pyrimidines: thymine (T) and cytocine (C). N-9 of a purine or N-1 of a pyrimidine is attached to C-1 of deoxyribose to form a deoxyribonucleoside. A deoxyribonucleotide is a phosphate ester of a deoxyribonucleoside. The site of esterification is the oxygen of the hydroxyl group attached to C-5 of the deoxyribose. The backbone of DNA consists of deoxyribose residues linked by phosphate residues. The 3'-hydroxyl of the sugar residue of one deoxyribonucleotide is joined to the 5'-hydroxyl of the adjacent sugar by a phosphodiester bridge. The 5' end of DNA either has a free hydroxyl or a phosphate group on the 5' carbon of the sugar. The 3' end of DNA generally has a free hydroxyl group on the 3' carbon of the sugar.

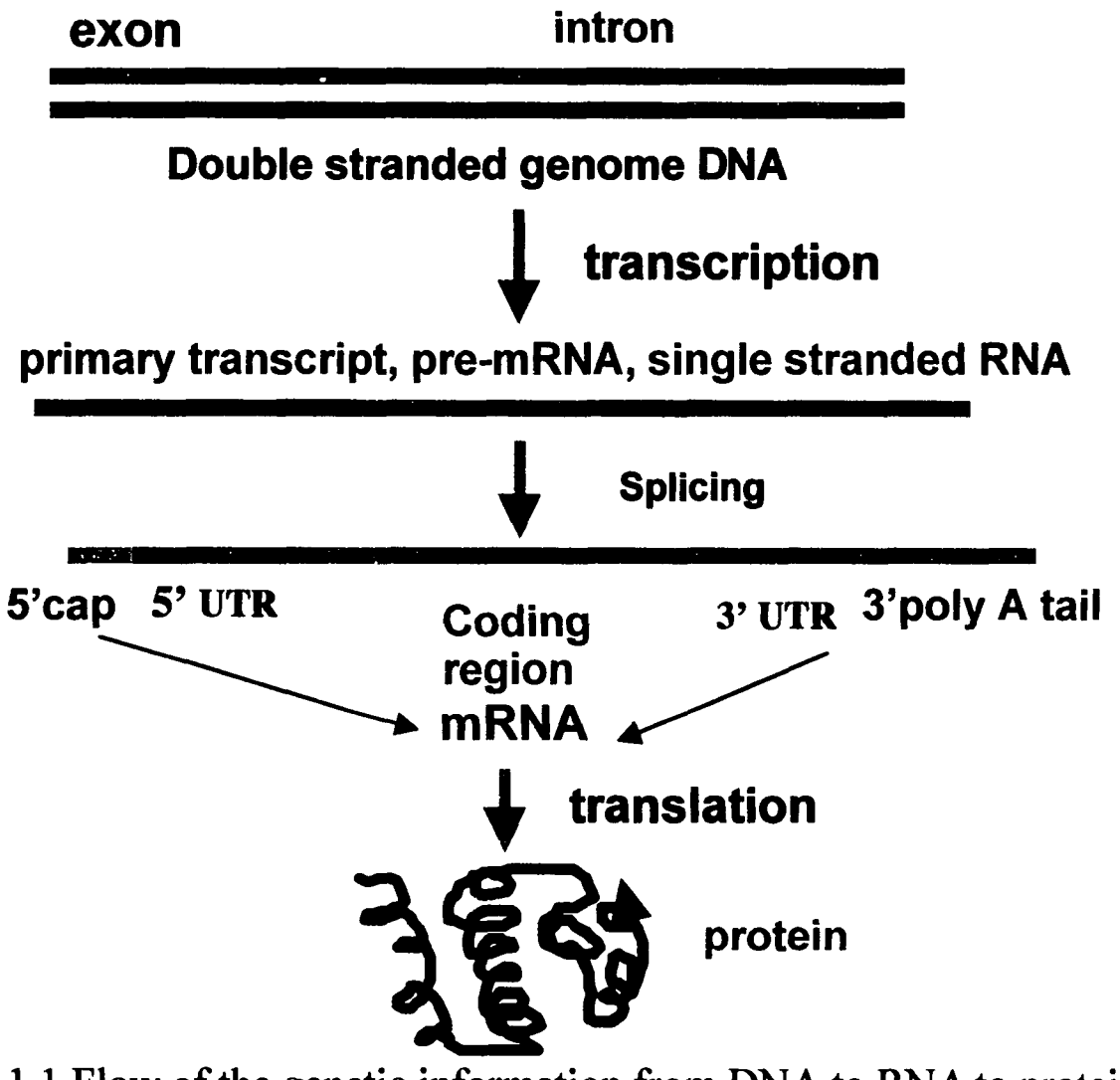


Fig. 1.1 Flow of the genetic information from DNA to RNA to protein

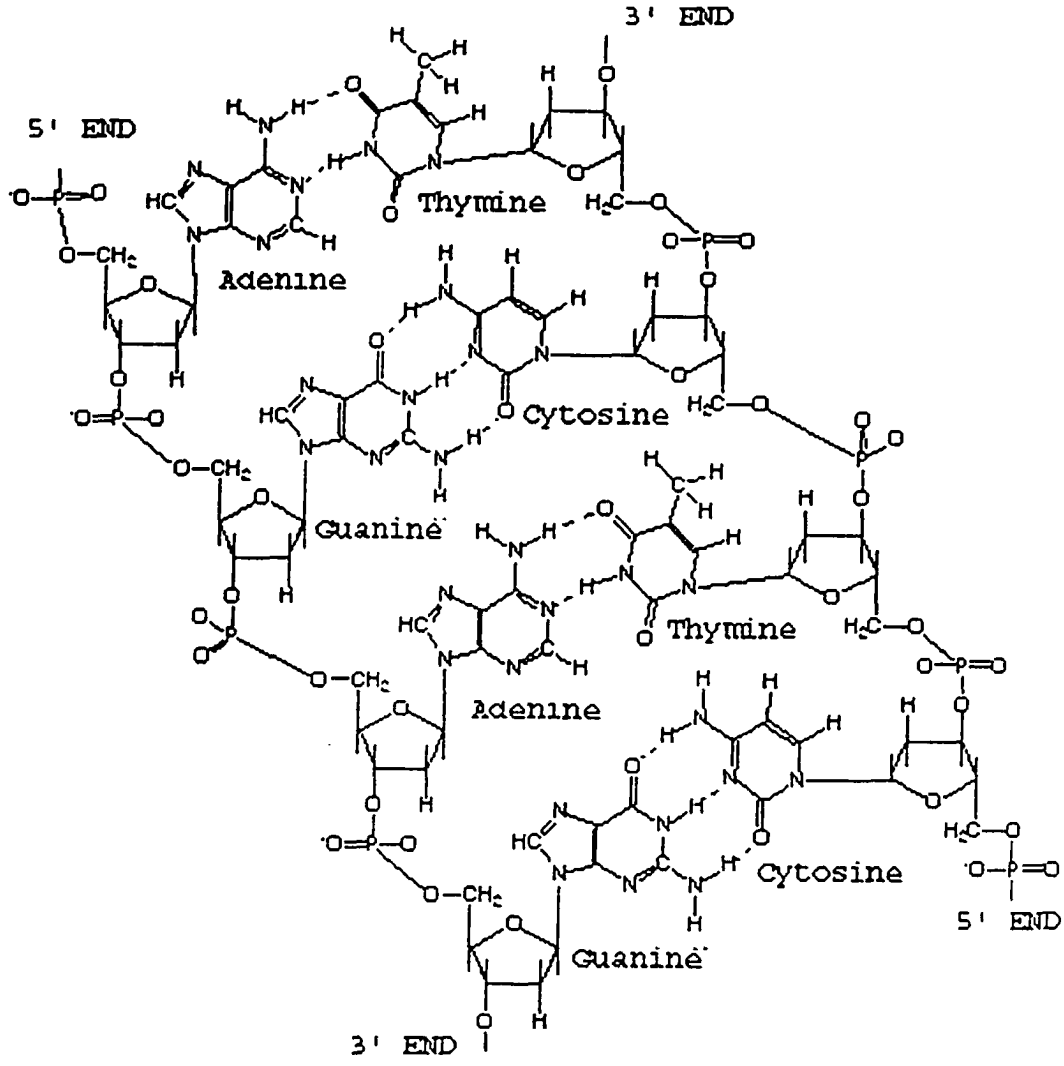


Fig. 1.2 DNA structure showing antiparallel nature of DNA and base pairing between Adenine (A) and Thymine (T) and between Guanine (G) and Cytosine (C)

One of the most significant discoveries in this past century was the determination of the 3-D structure of B-form DNA by James Watson and Francis Crick in 1953 (Watson and Crick, 1953). B-form DNA has two antiparallel helical polynucleotide chains coiled around a common axis. The two chains are held together by hydrogen bonds between pairs of bases. Adenine is always paired with thymine while guanine is always paired with cytosine (Fig. 1.2). The precise sequence of bases carries the genetic information. Under physiologic conditions, most of the DNA is in the Watson-Crick B form, but there are other helix forms such as A-DNA (Wang et al., 1982; Dickerson, 1983; Conner et al., 1984) and Z-DNA (Poll and Jovin, 1972; Wang et al., 1979; Rich et al., 1984).

The replication of DNA molecules is semiconservative.

1.1.2 RNA

RNA is a polymer of ribonucleotide units. The sugar unit in RNA is ribose. The four major bases in RNA are adenine (A), uracil (U), guanine (G) and cytosine (C). RNA molecules usually are single-stranded, except in some hairpin loop regions or in those virus genomes that have double stranded RNA. When RNA is double-stranded, adenine pairs with uracil and guanine pairs with cytosine. RNAs with several different functions have been observed. Among them, messenger RNAs (mRNAs) are the information-carrying intermediates in protein synthesis while other RNAs such as transfer RNAs (tRNA), ribosomal RNAs (rRNAs), small nuclear RNA (snRNA) are part of the protein-synthesizing process. All cellular RNAs are synthesized by RNA polymerases directed by the information given in DNA templates. The sequence in the DNA templates is fully

conserved in RNA synthesis, but many RNA molecules are modified after transcription. The detailed features and functions of mRNA are discussed in 1.2.

1.1.3 Genes

1.1.3.1 Definition of genes

The concept of a gene began with Gregor Mendel in 1865 when he summarized his experiments on pea genetics. In his two laws, a gene was recognized as a "particulate factor" passing unchanged from parent to progeny. However, the word "gene" had not been coined until twentieth century when it was introduced by W. Johannsen in 1909 (Johannsan, 1909). He defined a gene as a hypothetical unit of information governing the inheritance of individual characteristics in an organism.

With the rapid advances in biochemistry and molecular biology, it has been established that DNA is the major carrier of genetic information (genes) in most organisms. The genetic code is the relation between the sequence of bases in DNA or mRNA and the sequence of amino acids in the protein. Groups of three bases (called codons) code amino acids.

1.1.3.2 Classification of genes

In procaryotes, a single RNA polymerase is responsible for transcription. But in eucaryotes, different types of RNA are transcribed by different types of RNA polymerase. Eukaryotic genes are classified into three classes according to the differences in gene product and the enzyme required for gene transcription. Class I genes encode the 5,8S, 18S and 28S rRNAs; Class II genes encode all mRNAs and a variety of small nuclear RNAs (snRNAs); Class III genes encode tRNAs, 5S rRNAs and certain small

cytoplasmic RNAs (scRNAs). Gene classes I, II, and III are transcribed by RNA polymerase I, II, and III respectively.

1.1.3.3 Structure of genes

Prokaryotic and eukaryotic genes share the same basic overall design, as for example, they both use a triplet genetic code which is nearly universal. But the differences in molecular detail are still substantial. One difference is that most prokaryotic genes are polycistronic (one transcription unit contains coding sequences for more than one type of protein or RNA), while eukaryotic genes are monocistronic (one transcription unit contains only a single coding sequence for a protein or stable RNA). Another difference, which is one of the greatest differences between procaryotic and eucaryotic structure genes, is that most eucaryotic genes are discontinuous, while the procaryotic genes are continuous. In eucaryotic genomes, the discontinuous genomic regions that are represented in the mature RNA products are called exons while regions missing from the mature RNA products are called introns. In 1977, investigators in several laboratories discovered that exons are interspersed with introns in most eucaryotic genes (Berget et al., 1977; Brack et al., 1977; Breathnach et al., 1977; Chow et al., 1977). The number of exons per gene varies from one to more than 70 (Matsuo, 1995). The length of exons varies between several base pairs to several thousand base pairs, averaging about 130 bp in human exons (Zhang, 1998). The lengths of introns are short in yeast and fungi, typically less than 50 bases, while in humans the introns frequently are several thousand base pairs.

There is no universal definition for a eukaryotic gene with which everyone is satisfied. However, it generally is agreed that the following DNA segments should be included in a gene:

1) The sequences that will appear in hnRNA (see 1.2). Specifically, introns and exons, 5' untranslated regions (5' UTRs) and 3' untranslated regions (3' UTRs).

2) The promoter (the sequences needed to initiate correct transcription) and terminator (the sequence needed to terminate transcription).

3) The sequence elements that regulate the rate of transcription initiation such as enhancers (the sequence that increases the utilization of promoters).

Special short sequence patterns may appear in these segments. For example, in a promoter there may be TATA box, CCAAT box, GC box etc., while in a terminator a repeat sequence of AATAAA is found. The core sequence found in an enhancer is a sequence of nucleotides (GTGGAAAG or GTGGTTTG). Introns always seemed to begin with GT and end with AG. Beside the GT-AG rule, longer consensus sequences can be written for both the 5' and 3' splicing junctions e.g.: 5'(---exon---CAG)(GTAAGT ---intron---TAG)(G ---exon---)3'. The consensus sequences of special short sequence patterns have been summarized and used in the programs to predict genes based on DNA sequences (Xu et al., 1994; Burge and Karlin, 1997; Hua, 1999).

1.1.3.4 Pseudogene

Pseudogenes are regions of DNA that are structurally similar to specific functional genes, but are not able to yield functional gene products. Pseudogenes either contain mutations or premature stop codons in the coding or regulatory regions,

especially in transcribed pseudogene, or contain only part of the coding or regulatory regions, or lack functioning promoters, as is the case for non-transcribed pseudogenes.

1.2 mRNA(messenger RNA)

1.2.1 Splicing

In prokaryotes, transcription and translation are coupled and the primary transcript is translated directly on the ribosome. However, in eucaryotic organisms, transcription and translation are uncoupled. Class II genes in double stranded genomic DNA are transcribed into single stranded primary transcripts called pre-mRNA (precursor mRNA) or hnRNA (heterogeneous nuclear RNA) in nucleus (Figure 1.1). The hnRNA contains both intron and exon sequences. Introns will be removed from hnRNA in order to form mature mRNA by a procedure called splicing (Padgett et al., 1986; Maniatis and Reed, 1987). The spliceosome is the macromolecular complex which is responsible for hnRNA splicing (Wise, 1993). Since 1977, several small nuclear RNAs (snRNAs) and many proteins have been identified as essential components of the spliceosome (Wise, 1993).

1.2.2 5' cap

The 5' end of mRNA is modified by addition of a guanine nucleotide (Furuichi et al., 1975; Learned et al., 1985; Breathnach et al., 1981; Nevins, 1983; McKnight et al., 1982). However, the linkage is not a conventional 3'-5' phosphodiester bond. Instead, GTP reacts with triphosphate at the 5'end of a mRNA chain in a 5'-5' condensation. The structure 3'-G-5'ppp5'-N-3' has a 5'-5' triphosphate linkage and is known as a 5' cap

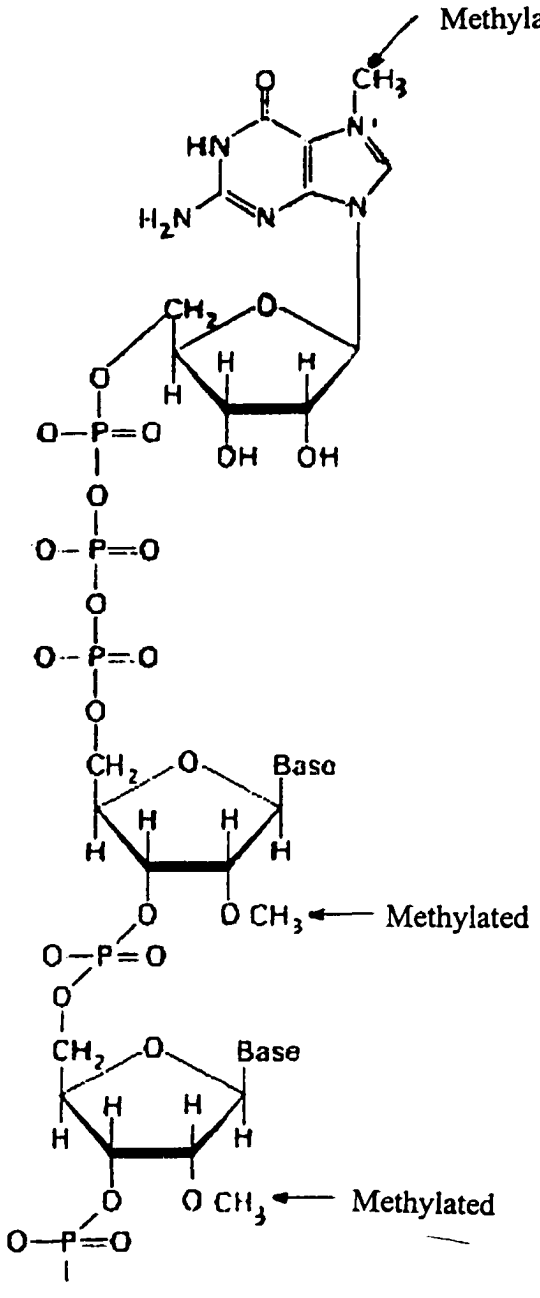


Fig. 1.3 The structure of the 5' cap in the eukaryotic mRNA

(Fig. 1.3). In many instances, a methyl group then is added to the 7-N of guanine residue and also to the 2'-OH groups of the first and sometimes also the second adjacent nucleotides. At this point, the 5' end of mRNA is blocked from 5' exonuclease degradation. It is generally accepted that the cap structure, together with cap binding protein, helps ribosomes attach to mRNA chain in 5' UTR to start translation, while it also contributes to the stability of mRNA by protecting its 5' end from phosphatases and nucleases.

1.2.3 3' poly(A) tail

The 3' end of mature mRNA is generated in the nucleus through several steps (Sachs and Wahle, 1993). At the 3' end, the pre-mRNA has hundreds of nucleotides extending beyond the mature 3' end. This region will be cleaved endonucleolytically and then a poly(A) tail will be extended by esterfication of several hundred of ATPs to the 3'-OH group of the 3' terminal nucleotide. Two RNA sequences will influence the cleavage reaction. One is highly conserved sequence AAUAAA 10-30 nucleotides upstream of the cleavage site. Another is poorly defined GU- or U-rich sequence 10-30 nucleotides downstream of the cleavage site. It is now generally accepted that poly(A) tail has two functions. First, it allows the mRNA to be transported from the nucleus and second, it protects mRNA from degradation in cells. It also has been reported that translation is most efficient with, and sometimes dependent on, the presence of a poly(A) tail (Jackson and Standart, 1990).

1.2.4 Mature mRNA

The mRNAs in eukaryotes are monocistronic. The sequence in mature mRNA can be divided into two types, translated region and untranslated region (UTR) (Fig. 1.1).

Translated regions (coding region) consist of a series of codons representing the amino acid sequence of the protein. They start with AUG, which represents not only the codon for the amino acid methionine, but also, the start site of translation of the protein. Coding regions end with a stop codon UAA, or UAG, or UGA, which represents the signal to stop translation. UAA, UAG and UGA are the only three codons that do not specify any amino acids. Untranslated non-coding regions are present at both ends of coding region. The untranslated region preceding the start codon is called the 5' UTR while the additional sequence following the termination signal is the 3' UTR.

The mRNA transported out of the nucleus is capped as it exits the nuclear pores and then serves as the templates for the translation of proteins in cytosol. Therefore mRNA molecules represent genes without introns and most of them have cap on their 5' end and a long (about 250 bases) polyadenylate (poly(A)) tail at their 3' end.

1.3. cDNA (complementary DNA) library

1.3.1 The cDNA libray

Since mRNA molecules are extremely labile and difficult to amplify, the information encoded by a mRNA molecule often is converted into a stable DNA duplex (complementary DNA, cDNA) and this cDNA duplex then is inserted into a self-replicating vector. A cell containing cDNA is called a cDNA clone if it contains at least one vector that contains duplex DNA sequence representing an mRNA. A collection of such cDNA clones is called a cDNA library. A cDNA library represents almost all of the information encoded in mRNA molecules, that is, a cDNA library represents the group of genes that were expressed in the cells from which the mRNA was isolated. Therefore, a

cDNA library is a snapshot for the expression pattern for the cells at that specific time point and it can provide a way to examine and interpret characteristics of transcription in a specific organism or in a specific developmental stage of an organism.

1.3.2 General procedures for construction of cDNA library

A cDNA library is constructed from the mRNA isolated from cells of interest. Many methods have been developed to isolate RNA (Chomczynski et al., 1987, Liu and Raghothama, 1996) and after the total RNA is isolated, it is passed through oligo(dT)-coated magnetic beads (Hornes and Korsnes, 1990) or oligo(dT) coated latex particles (Matsubara and Okubo, 1993) to select mRNA which contain poly(A) tails. The poly(A) tails of the mRNA will bind the poly(dT) tails in the column and thus separate the mRNA molecules from other RNA molecules. These purifed mRNAs then are used to construct a cDNA library.

The key enzyme in the construction of cDNA clone is the reverse transcriptase (RNA-directed DNA polymerase) (Temin and Mizutani, 1970; Baltimore,1970). One enzyme in this enzyme group was first found in virions of *Rous sarcoma* virus in 1970 (Temin and Mizutani, 1970). The *Rous sarcoma* viral genome was one of the first noted exceptions to Central Dogma of Molecular Biology since it has an RNA genome which is reverse transcribed to DNA prior to transcription and translation.

If a DNA primer that is base-paired to the RNA molecule and that contains a free 3'-OH group is provided, reverse transcriptase can use any RNA molecule as a template to synthesize a complementary DNA strand, yielding an RNA-DNA hybrid (Fig. 1.4(a)). Therefore, this enzyme is used to synthesize the first strand of cDNA from mRNA by

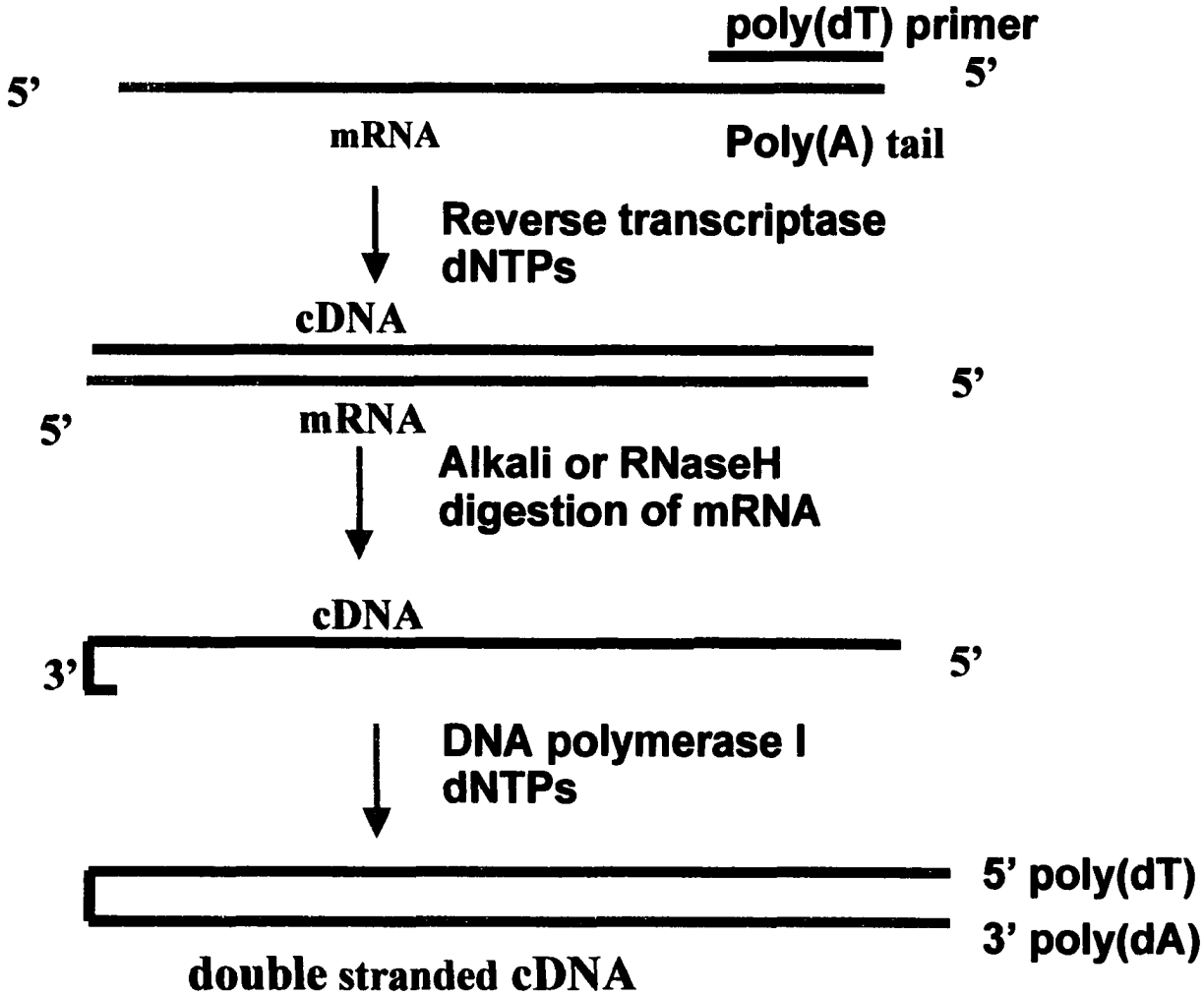


Fig. 1.4(a) Construction of a cDNA clone

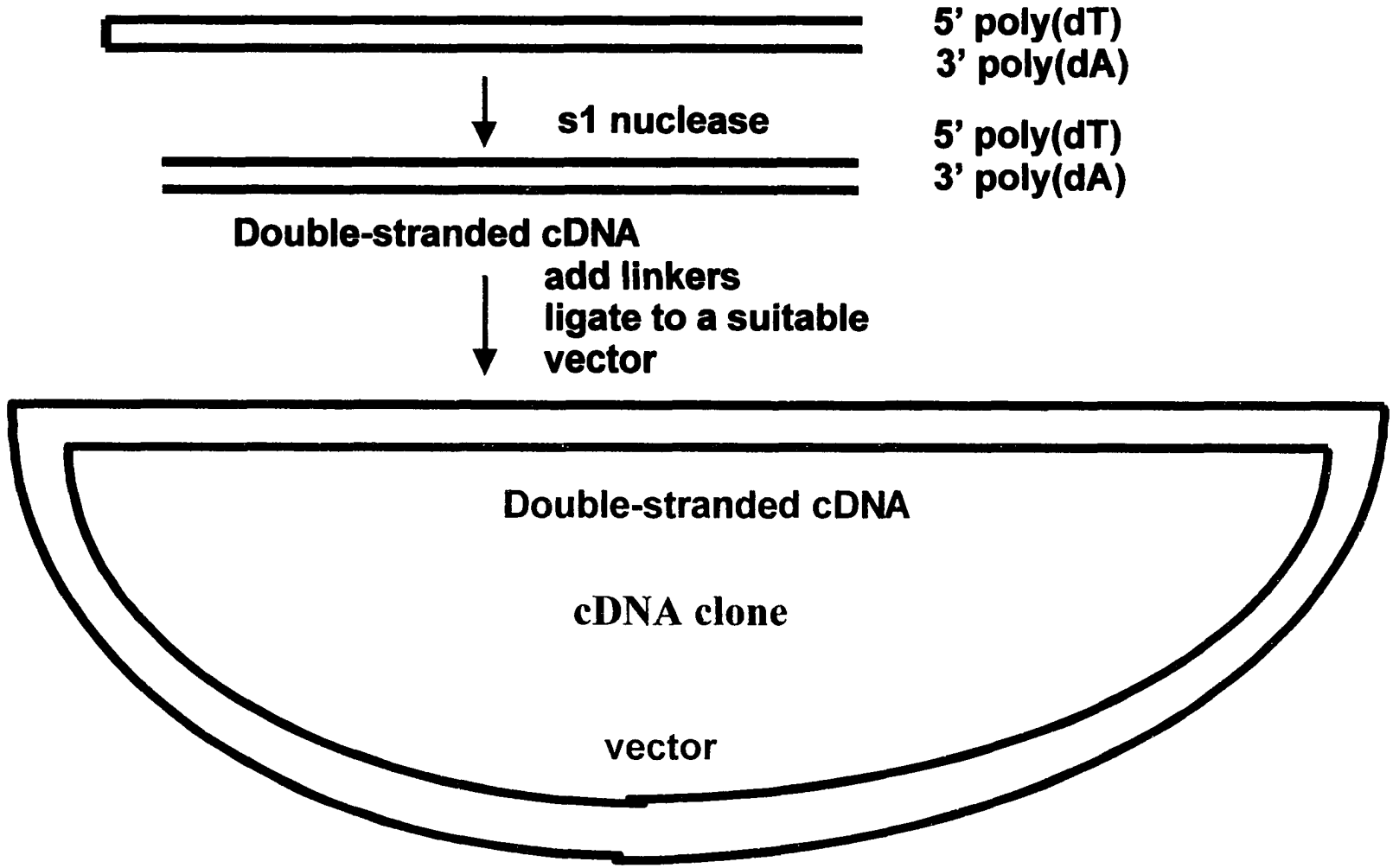


Fig 1.4 (b) Construction of a cDNA clone

providing an oligo(dT) primer that is paired with the poly(A) tail at the 3' end of the mRNA molecule.

After the mRNA-cDNA hybrid is formed, alkali or Ribonuclease H (RNase H) is used to degrade the mRNA strand. RNaseH is an RNA endonuclease (Krug, 1989) which will specifically digest RNA strand in DNA/RNA hybrid molecules. Then, the first DNA strand acts as a template for the synthesis of its complementary DNA strand by DNA directed DNA polymerase (e.g. DNA polymerase I). The 3' end of the first strand cDNA will form a hairpin loop which will be used as the primer for the synthesis of the second strand of cDNA. Thus, a double-stranded cDNA copy for that mRNA molecule is produced.

S1 nuclease (Rushizky et al., 1975) is a single-stranded specific endonuclease which can be used to cleave the single-stranded loop at one end of the double-standed cDNA (Fig. 1.4(b)). The cleavages of phosphodiester bonds by this enzyme produce 5'-monophosphates and 3'-hydroxy ends. Then, the blunt-ended double-stranded cDNA can be inserted into a vector either with or without added linkers for maintenance and amplification.

1.3.3 The λ ZAP system as the vector for cDNA library

Lambda (λ) phage is a widely used cloning vector. Lambda phage contains one double-stranded DNA molecule encapsulated in an icosahedral head, with a tubular tail projecting from the head (Hendrix et al., 1983). Lambda has two life styles: lytic pathway and lysogenic pathway. In lytic pathway, lambda virus reproduces itself very rapidly and destroys its host and releases its progeny virus particles by causing lysis. In lysogenic

pathway, the phage DNA is inserted into the host cell genome and replicates as part of the bacterial chromosome for many generations inactively.

The efficiency of lambda phage infection of *E. coli* exceeds that of plasmid transformation 100-fold (Short et al., 1988). Many different mutant lambda phages have been constructed. Among which is the lambda ZAP system (Short, 1988) commercially available from Stratagen that is used extensively to construct cDNA libraries (Hillier et al., 1996; Adams et al., 1991, 1995; Khan et al., 1992). This system has several advantages (Short, 1988, Statagene). First, it allows efficient *in vivo* excision and recircularization of pBluscript phagemid from lambda vector containing the cloned insert. This allows the insert DNA to be recovered and manipulated in a plasmid system. The partial *LacZ* gene in the phagemid provides α-complementation for the blue-white color selection of recombinant phagemid. The vector encoded ampicillin resistance gene enables antibiotic selection of the host cells that contain the phagemid vectors. Therefore, this system has combined the advantages of high efficiency of lambda library construction with the convenience of the plasmid system. Second, the multiple cloning site contains twenty-one restriction sites, thereby allowing many choices for cloning site. Third, the lambda vector allows inserts as large as 10 kb to be inserted. Fourth, it can be used efficiently as an expression vector.

A helper phage is used for the efficient excision of the pBluescript phagemid from the lamda ZAP vector. The helper phage contains an amber mutation in its replicon region that prevents it from replication and eliminates co-infection.

1.4. EST and UniGene databases

When a primer, whose sequence is complimentary to universal site in the vector is used to sequence a randomly chosen cDNA clone, single pass end sequences will be produced. These single pass, partial sequences of random cDNA clones from one or both ends are called Expressed Sequence Tags (ESTs) (Adams, 1991) (Fig. 1.5). Since the λ-ZAP vector contains T3 and M13 universal specific sequences, these sites were used for unique primer binding for the synthesis of ESTs in this present project.

1.4.1 Beginning of large scale EST sequencing

Early in 1983, it was found that single-pass partially sequencing of random cDNA clones was a relatively inexpensive and rapid means to access expressed genes (Putney et al., 1983; Milner and Sutcliffe, 1983; Costanzo et al., 1983). However, a large-scale EST database was not produced until recently in Human Genome Project. In 3 billion base pairs of human genome, only about 3% actually encodes proteins. It earlier was estimated human genome contains about 50,000 to 100,000 genes (Antequera and Bird, 1994; Nowak, 1994). Based on the human genome working draft reported recently (Lander et al., 2001; Venter et al., 2001), there mostly likely are only 30,000 to 40,000 human genes with significant alterative splicing. Because only 2000 human genes had been identified in 1990, it was suggested at that time that large-scale cDNA sequences be generated and added as a component of the Human Genome Project (Brenner, 1990), and subsequently the first EST project was begun (Adams, 1991).

1.4.2 Recent advances in large scale EST sequencing

It has been now widely realized that high accuracy, full length cDNA sequences are not necessary for identification of genes. For ESTs to be of use, they need only be sufficiently long (longer than 100 bp) to specify a unique sequence in the genome and be

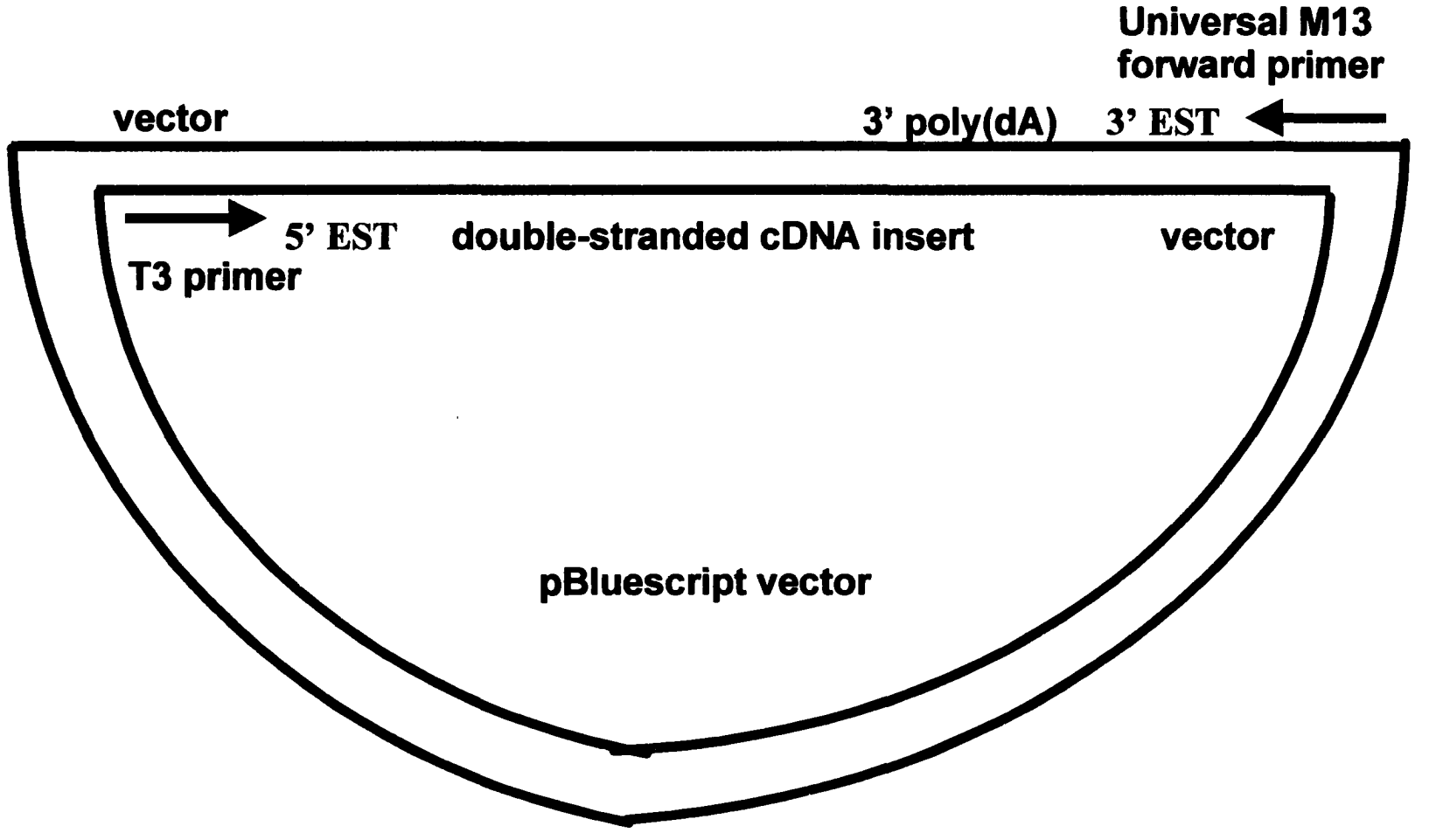


Fig. 1.5 3' EST and 5' EST

accurate enough to allow recognition of similarity by computer programs (Adams et al., 1991; Hillier et al., 1996). In recent years, many papers have been published reporting the results from EST projects on humans (Khan et al., 1999), mouse (Bailey, Jr., et al., 1998, Marra et al., 1999), rice (Yamamoto et al., 1997), *Arabidopsis thaliana* (Delseny et al., 1997), soybean and wheat (Rafalski, 1998), and a number of parasitic organisms including *Plasmodium falciparum*, *Leishmania major* and *Trypanosoma brucei* (Lawson, 1999).

According to NCBI's web site (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html) report, the number of public entries for EST in dbEST (see 1.6.7) is 7,780,212. Table 1.1 shows a portion of this table as released on the web on 04-13-2001.

1.4.3 3' EST and 5' EST

The 3' ESTs begins with a poly(dT) sequence since it corresponds to the sequence of the mRNA contiguous with the poly(A) tail. Therefore, a 3' EST has the complimentary sequence with the 3' end of its corresponding mRNA. The 5' EST has the same sequence with the 5' end of its corresponding mRNA (Fig. 1.5). In the *Fusarium sporotrichioides* project, 3' ESTs (labeled as .f1 sequence) were produced by universal M-13 forward primer and 5' ESTs (labeled as .r1 sequence) were produced by the universal T3 primer.

1.4.3.1 5' EST is more gene family specific

5' EST may not contain a start codon or 5' UTR because reverse transcriptase may dissociate from mRNA before it reaches the very end of 5' end during the

Table 1.1 Part of the EST data released on 04-13-2001 on NCBI's dbEST web site
Number of public entries: 7,780,212

Organisms	Number of ESTs*
Homo sapiens (human)	3,413,006
Mus musculus + domesticus (mouse)	1,959,672
Rattus sp. (rat) 273,740 Bos taurus (cattle)	162,037
Glycine max (soybean)	161,707
Medicago truncatula (barrel medic)	117,498
Drosophila melanogaster (fruit fly)	116,471
Arabidopsis thaliana (thale cress)	113,000
Caenorhabditis elegans (nematode)	109,215
Lycopersicon esculentum (tomato)	107,226
Zea mays (maize)	89,125
Xenopus laevis (African clawed frog)	88,954
Danio rerio (zebrafish)	85,586
Oryza sativa (rice)	72,744
Hordeum vulgare (barley)	68,480
Chlamydomonas reinhardtii	64,973
Sorghum bicolor (sorghum)	64,573
Sus scrofa (pig)	61,017
Triticum aestivum (wheat)	58,141
Solanum tuberosum (potato)	38,074
Pinus taeda (loblolly pine)	34,977
Neurospora crassa	28,089
Lotus japonicus	27,078
Brugia malayi (parasitic nematode)	22,441
Gossypium arboreum	20,978
Sorghum propinquum	20,798
Dictyostelium discoideum	19,183
Gallus gallus (chicken)	18,735
Bombyx mori (domestic silkworm)	14,849
Onchocerca volvulus	14,608
Schistosoma mansoni (blood fluke)	14,039
Mesembryanthemum crystallinum (common ice plant)	14,033
Emericella nidulans	12,993
Oryzias latipes (Japanese medaka)	12,733
Toxoplasma gondii	12,177
Strongyloides stercoralis	10,979
Ciona intestinalis	10,347
Porphyra yezoensis	10,184
Trypanosoma cruzi	10,133
Physcomitrella patens	9,850
Eimeria tenella	9,637
Gossypium hirsutum (upland cotton)	9,438
Lycopersicon pennellii	8,346
Secale cereale	8,123
Schizosaccharomyces pombe (fission yeast)	8,118
Meloidogyne incognita (southern root-knot nematode)	6,626
.....	

*Multiple copies of highly expressed genes contains

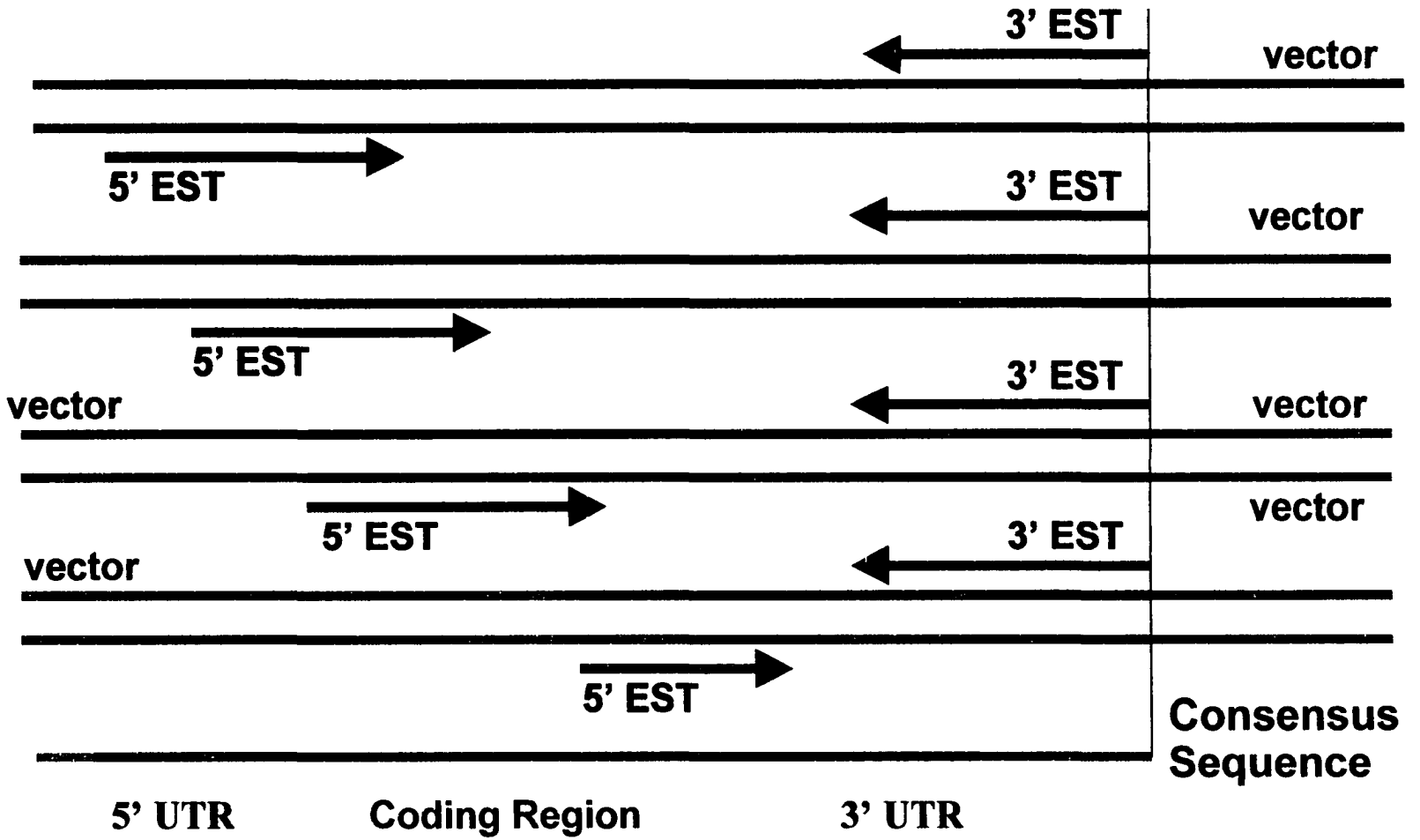


Fig. 1.6 A set of cDNA inserts from the same gene with different insert lengths and cDNA consensus construction by 3' EST and 5' EST assemblies

transcription of the first strand cDNA. It means that, although some cDNA clones may represent the same gene, the length of their cDNA insertions may be different. These cDNAs thus have the same 3' start point but different 5' stop point (Fig. 1.6). Their 3' EST sequence overlap with each other since they all begin at the 3' poly A region but the 5' ESTs might not overlap because of different positions of reverse transcriptase termination.

However, in most situations, 5' ESTs contain protein-coding regions. Genes that belong to the same family may share the same conserved functional motif and thus may have the homologous 5' EST sequences. Therefore, 5' ESTs are gene family specific. A 5' EST could be used as a query sequence to compare with a "model" sequence to decide which gene family the 5' EST belongs to. The "model" sequence is constructed from a sequence alignment of multiple members of that specific protein family (Hillier et al., 1996).

1.4.3.2 3' EST is more gene specific

3' ESTs that represent the same gene have the same starting point and share the overlapped sequences (Fig. 1.6). The sequencing primer for 3' EST (f1 primer in this project) binds to the vector, sequencing through the poly(dA) region of cDNA, then sequencing into 3' untranslated regions (3'UTR). 3' EST may or may not contain the protein-coding region depending on the length of the 3' UTR. There are several advantages for sequencing the 3' UTR (Adams et al., 1991; Wilcox et al., 1991; Khan et al., 1992; Berry et al., 1995; Kupfer, 1999) even though it may not contain a coding region:

1) During the construction of cDNA library, poly(dT) primers of the first cDNA strand are anchored on the corresponding positions of the poly(A) tails, of the mRNAs. Therefore, 3' ESTs starting from the regions upstream of poly(A) tails contain polyadenylation signals and 3' UTR of mRNA. Same gene should have overlapping 3' EST sequences except for the differing lengths of poly(A) sequence, although they might be derived from different copies of mRNA or different copies of cDNA. Thus, different gene would have different 3' EST sequences. 3' EST with homologous 3'end sequences can be aligned into clusters after assembly with sequence alignment tools such as Phrap (Fig. 1.6). Each cluster will represent a set of clones transcribed and reverse transcribed from the same gene. The number of ESTs in each cluster is an estimate of the abundance of a gene's transcript present in the cDNA library. Therefore the abundance of 3' ESTs in each EST cluster for a specific gene is proportional to this gene's relative transcription level. This property is used to study the gene expression patterns for an organism of interest or for its specific development stage.

2) 3' ESTs always contain 3' UTR sequence but may not include any coding region sequence. The sequences for 3' UTR are not as well conserved as coding sequences. Therefore, 3' EST are more gene-specific than 5' EST. They are more useful for discrimination among individual genes that belong to the same gene family and that may share the same conserved functional motif in coding region.

3) Almost all genome regions corresponding to 3'UTRs are intronless. Consequently, the 3' EST sequences usually are identical to the corresponding genomic sequences. In most cases, 200 to 500 bp of PCR sequence define an STS (Sequence-tagged-site) that is unique in the human genome (Olson, 1989). STSs are becoming

standard markers for physical mapping. ESTs can serve the same purpose as STSs because 3' EST size is the same with PCR product size and corresponding genomic sequence size. Therefore, a 3' EST database is useful for gene-based map development (Berry et al., 1995) and for building UniGene databases (Schuler et al., 1996).

1.4.4 UniGene database

The UniGene database collection began with the need to develop a transcript or expression map for the human genome, as it will be an invaluable annotation aid before completely sequencing the 3 billion nucleotides of human genome (Schuler et al., 1996). Since GenBank is a "historical archive", a collection of public data with a large degree of internal redundancy, any gene may be represented in the database by different types of sequences, e.g. EST sequences, mRNA sequences and genomic sequences. One gene also may have been sequenced and submitted to GenBank multiple times by different labs. The advance of EST technology, on one hand, has largely increased the number of genes in GenBank and has made it feasible to develop a transcription map (Schuler et al., 1996). On the other hand, it also has increased the redundancy and overlap of genes in GenBank (Table 1.1) and made it difficult to identify unique markers for mapping. Therefore, an automatic system to generate a non-redundant set of gene sequences was needed.

The NCBI (http://www.ncbi.nlm.nih.gov/UniGene/), defines a UniGene database as "an experimental system for automatically partitioning GenBank sequences into a non-redundant set of gene-oriented clusters." The main procedure for UniGene collection for ESTs (http://www.ncbi.nlm.nih.gov/UniGene/build.html) is to find the 3' UTRs that share significant sequence similarity and assign them to the same cluster. This set of clusters then is submitted to the UniGene database and compared with the already

existing set there. The new EST submissions that do not match any sequences in the UniGene database are considered to be new genes. The matched clusters are merged to their matched sequences. In this way, all representations of a single gene are collected in a single cluster. As one could imagines, some clusters contain only 1 EST while some may contain hundreds or thousands of ESTs.

At present, only EST sequences from human, rat, mouse and zebrafish have been processed in GenBank's UniGene database as it is presented on NCBI's web site. The procedures for automated sequence clustering are still under development and not yet in common use.

1.4.5 Applications of EST and UniGene databases

ESTs and UniGene clusters have become an increasingly important resource for many investigators in different fields especially for researchers for genomic sequencing projects and biomedical projects.

Although an EST is only a partial sequence of a cDNA, its information can be incorporated with the information from completed genomic sequences to indicate the correct positions of sequence landmarks such as transcription start sites, translation start and stop sites, intron-exon borders, alternative splice sites, overlapping transcription units and other distinction regions. ESTs also are extremely useful for predicting or verifying predicted open reading frames in the genome and in helping construct accurate physical and genetic maps of the genome. Existing EST and UniGene databases have already been applied to the genome sequence annotations (Bailey et al., 1998; Jiang et al., 1998) and used for human gene-based map development (Berry et al., 1995; Schuler et al., 1996).

EST sequences also have been used to predict new proteins and identify new genes. The discoveries of mammalian secretary proteins with possible pharmacological potential are but just examples of several new proteins discovered by an EST project (Ladunga, 2000). Another example is that the large-scale determination of sporozoite ESTs have increased the number of *Cryptosporidium parvum* genes identified (Abrahamsen, 1999). The application of EST sequencing also has contributed to the discovery of novel genes, including the discovery of new genes associated with osteoblast differentiation (Carulli, 1998), DiGeorge syndrome (Sutherland et al., 1998), and more than 10 new human chemokine genes (Wells et al., 1997).

Human EST sequences have been used to measure expression patterns of thousands of genes in the human brain (Colantuoni et al., 2000), and now are being used to build microarrays which are enabling the determining the genome-wide expression profile of thousands of genes in one experiment in cancer and other biomedical related research (Khan et al., 1999; Schena, 1998). Large-scale generation of ESTs also has been employed to establish gene expression profiles and to identify potentially significant apoptotic regulators in the cardiovascular system (Rezvani et al., 2000). A gene expression profile also has been reported for various stage of soybean seed development using EST data derived from different developmental stages (Rafalski, 1998).

By comparing the sequences of redundant ESTs, gene-associated single nucleotide polymorphisms (SNPs) can be found (Gu et al., 1998).

EST sequencing also has contributed to the revolution of pathology (Going et al., 1999). EST-based technologies are being used to understand disease processes and to find better disease treatments (Fannon, 1996; Zweiger et al., 1997). It has been estimated

that in the next few years, ESTs representing all sequences expressed in humans would be determined and their genomic position would be defined (STSs) (Pratt et al., 1999). These and related studies will result in a rapid growth of the functional genomic database, EST, SAGE and microarray techniques, while accelerating the rate of finding new high-value peptides or proteins including antibodies, vaccines, enzymes and therapeutic peptides (Ryu et al., 2000; Carulli et al., 1998).

1.5 DNA sequencing

1.5.1 Basic methods for DNA sequencing

Two basic methods of DNA sequencing were developed in 1977. One depends on specific chemical degradation (Maxam and Gilbert, 1977) and the other depends on enzymatic synthesis (Sanger et al., 1977).

In the Maxam-Gilbert method, polynucleotide kinase is used to add ^{32}P at the 5'-hydroxyl terminus. Then, labeled DNA is preferentially broken at one of the four nucleotides using specific chemical cleavage reagents. The reactions corresponding to different chemical cleavage are separately electrophoresed. By carefully choosing conditions, an average of one cleavage is made per DNA chain. In the reaction mixture for a given base, each broken chain yields a fragment extending from the ^{32}P label to one of the nucleotide positions of that base. By loading the A, C, G, and T specific cleavage reactions in different lanes on the same polyacrylamide gel, fragments in each mixture can be resolved by electrophoresis and the template DNA sequence can be deduced by comparing the four corresponding lanes in an autoradiogram.

In the early Sanger dideoxy method, DNA polymerase I and a short oligonucleotide primer were used to copy a complementary DNA template. The substrates were four deoxyribonucleoside triphosphates (dATP, dTTP, dCTP and dGTP) where the dATP was labeled with ^{32}P at the alpha phosphorus, and one of the four 2'3'-dideoxy analogs (ddATP, ddTTP, ddCTP and ddGTP). Because ddNTP lacks the 3'-hydroxyl terminus needed to form the next phosphodiester bond, once it is incorporated at the 3' end, the chain could not be extended. In reaction mixtures containing a given ddNTP, the chain-terminated fragments of various length are produced, with each fragment extending from the primer to one of the positions represented by the base in ddNTP. The four reactions, each contains one type of ddNTP, are incubated separately. By loading aliquot of each of the four reaction products into different lanes on the same polyacrylamide gel, the fragments can be separated by electrophoresis and the template DNA sequence can be deduced by comparing the four corresponding lanes in an autoradiogram. This method has been widely used because of its easy of use and clear distinction between synthesized DNA fragments ending in one of the four DNA bases.

Several improvements to this method have been made since 1977. The modified Sanger method discussed below was the method of choice in this dissertation.

1.5.2 Improvements on Sanger method

The first improvement was to replace the radiolabeled primer or α ^{32}P-dATP with either fluorescent labeled primers (Smith, 1986) or terminators (Prober, 1987). Since chemically differently fluorescent tags can be attached to primers or terminators (Fig. 1.7) for each of the four different chain-termination reactions, the four reaction mixtures can be combined and electrophoresed in one lane on the gel. Using dye-terminators has

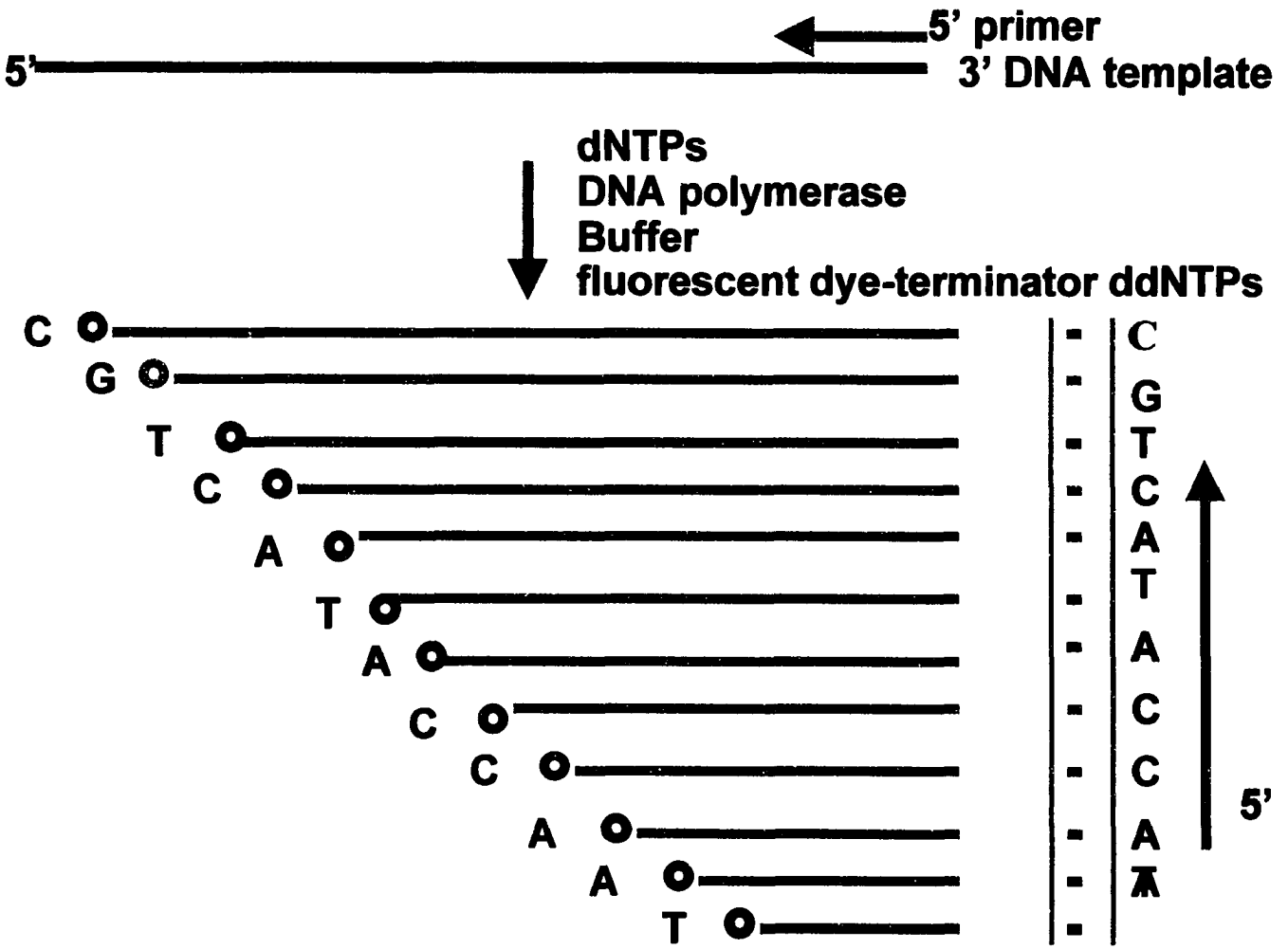


Fig. 1.7 Improved Sanger dideoxy method for DNA sequencing

an advantage over using dye-primers. First, if fluorescent tags are on dideoxy-terminators, the four base-specific reactions can be incubated in a single tube. The sequence is determined by measuring the fluorescence directly from the gel indicating the sequence of the bases copied from the DNA template. This improvement on the Sanger method not only eliminated the use of radioactive reagents, but also made it possible for the sequencing to be automated. Second, using dye-terminators reduces both the time and money required for making the fluorescent labeled custom primers needed for primer walking during gap closure.

The second important improvement to Sanger method was the development of different enzymes used in DNA sequencing. The enzyme originally used in Sanger method was Klenow fragment of *E.coli* DNA polymerase I which has its 5'-3' exonuclease activity domain removed but retains 5'-3' polymerizing activity and 3'-5' exonuclease activity (Klenow, 1971). T7 DNA polymerase was developed to replace the Klenow fragment (Tabor and Richardson, 1989) because it had a higher polymerization efficiency and a low dissociation constant for the fluorescent dye-labeled dideoxynucleotides. More recently, Taq (*Thermus aquaticus*) DNA polymerase (Innis et al., 1988) and several mutant forms of Taq polymerase have been widely used in DNA sequencing. Since Taq DNA polymerase has a high optimum reaction temperature (72°C), this enzyme is an ideal enzyme for cycle sequencing (Rosenthal et al., 1993). The wild type Taq polymerase had much higher dissociation constants for ddNTPs than dNTPs and therefore, a mutant Taq F667Y was produced (Tabor and Richardson, 1995), with the phenylalanine residue in the wild type replaced by a tyrosine residue in Taq F667Y at position 667, with a higher efficiency for incorporating ddNTPs than wild-type.

Subsequently the enzymes KlenTaq-TR (Barnes et al., 1995) and AmpliTaq-FS (Applied Biosystem) were developed from Taq F667Y. During this dissertation research, AmpliTaq-FS was the exclusive enzyme of choice as it is available as a pre-mix with the appropriate dye-labeled terminator and dNTPs.

Replacement of linear constant temperature sequencing with cycle sequencing improved results obtained by the Sanger method. Cycle sequencing (Craxton et al., 1993) was based on a strategy similar to PCR except that it uses only one primer. In cycle sequencing, the sequencing reactions are incubated for several cycles consisting of three different incubation temperatures, one for denaturation of double-stranded DNA, one for primer annealing and one for chain elongation. Repeated use of template DNA primer, and thermostable DNA polymerase allows cycle sequencing to yield more reaction product than constant temperature sequencing does.

1.5.3 DNA sequencing instruments

Although fluorescent labeled DNA sequencers were first described in work from Lee Hood's laboratory (Smith et al., 1986), it was not until 1989 that they were commercialized by Applied Biosystems Incorporated (ABI), and thus became widely used. The initial sequencers now have been updated first from ABI 370A to ABI 373 (Perkin Elmer, 1994) to ABI 377 (Perkin Elmer, 1996) and currently to ABI 3700 (Perkin Elmer, 2000). Most data collected in this dissertation was from ABI 377, except the very early data which was from ABI 373 and very late data obtained from ABI 3700.

ABI 377 has many improvements as compared with ABI 373. ABI 377 has a charge coupled device (CCD) camera to detect light emitted from laser-excited fluorescent dyes, which is a more highly sensitive detection system as compared with the

photomultiplier tube used in ABI 373. The gels used in ABI 377 are longer and thinner (48 cm long and 0.2 mm thick) as compared to those in ABI 373 (36 cm long and 0.4 mm thick). The number of the built-in heat-transfer plate increases from one (in the back of a mounted gel) to two (on both sides of the mounted gel). ABI 377 gains two folds of increase in the speed of the detector's left to right movement and the intensity of focus of lens in the detector. Those changes in ABI 377 result in an increase of the well-to-read length per lane from approximately 400 bases to about 550 bases. Throughput also was increased when number of lanes per gel increases from 36 to 48 and to 64.

ABI 3700 is an automated capillary gel electrophoresis system. It is much more efficient as compared with ABI 377. On the ABI 3700, 96 samples are electrophoresis simultaneously and 200 bases per hour per lane are detected with at least 98% accuracy. ABI 3700 can produce data in excess of 500 bases per lane in 2 hours for 96 samples while ABI 377 requires 8 hours to obtain a similar read length from 64 samples. In the ABI 3700, all four sequencing reactions are electrophoresised in a single capillary rather than in a single lane on a slab gel, therefore eliminating the time for manually tracking which is required by the other multisample ABI models. Finally, only a fraction of the sample is loaded, which makes it possible to reload the reaction should a machine failure occur.

1.5.4 Shotgun strategy for large scale DNA sequencing

Random shotgun sequencing strategy is widely accepted in sequencing DNA clones whose insert sizes ranging from 5 to 300 kb. A DNA clone can be randomly broken into short fragments of 2-4 kb via sonication (Deininger, 1983; Bankier, 1987), nebulization (Bodenteich, 1994), partial restriction enzyme digestion (Anderson et al.,

1981;) or transposon insertion (Phadnis et al., 1989). Since the sizes of fragments produced via nebulization are larger and more uniform than those produced via sonication (Bodenteich, 1994), and the nebulization method is easier to manipulate than enzyme digestion or transposon insertion method, nebulization is the chosen method in this dissertation.

After end repair, fragments are subcloned into plasmid sequencing vectors. The subclone is sequenced from both ends of insertion with oligonucleotide primers corresponding to vector sequences flanked on both ends of the insertion. The end sequence data are collected automatically from DNA sequencers and assembled via computer programs into a large sequence that represents the original insert sequence. Although at least three-fold sequence redundancy is required for each base to assure a 99.99% accurate final sequence, higher redundancies often are required to ensure the high overall accuracy of the entire final sequence. The gaps which cannot be covered by the end sequences from subclones can be closed by primer walking via custom synthesized primer, PCR or other methods which will be discussed in Chapter 2.

1.6 Sequence analysis software and web site

1.6.1 Phred and Phrap

Phred (Ewing et al., 1998, Ewing and Green, 1998) and Phrap (Green, copyright 1994-1996) are software developed in Phil Green's group at the University of Washington. Phred examines the trace files obtained from sequencers (e.g. ABI377, ABI3700), calls the bases and assigns quality values to the bases. Phrap is a program for assembling the sequence data. Using Phred quality values, Phrap assembles sequences by

combining the identical high quality regions to generate contiguous sequence segments or contigs. Phrap refers to the Phred quality value and uses its own algorithm to assign a quality value to each base in the consensus sequence. Their algorithms, function as well as input/output file formats will be discussed in chapter 2.

1.6.2. Cross_match, Dotter, Bestfit, Gap and Sim4

Cross_match is a sequence alignment tool in the PhredPhrap program package that is used to compare and align two DNA sequences. Dotter (Sonnhammer and Durbin,1996) is a graphical dot-matrix plots on X-windows which can compare DNA or protein sequences. Bestfit and Gap are programs in GCG (Genetics Computer Group) computer package that also can be used to align two sets of DNA or protein sequences. Sim4 is a public domain computer program developed by Florea et al. (1998) which can align a cDNA or mRNA sequence with a genomic DNA sequence containing that gene when there are introns in genomic sequence and sequencing errors in either or both sequences. Sim4 program was obtained by anonymous ftp from globin.cse.psu.edu or over the World Wide Web from http://globin.cse.psu.edu/

1.6.3 Consed

Consed is software for viewing and editing Phrap assemblies (Gordon et al., 1998). Consed requires at least three types of input files. First, the chromatogram files (one for each sequenced DNA) created by the sequencer which contain the fluorescence trace profiles; Second, the phd files (one for each sequenced DNA) created by Phred which contain the base calls, quality values, and trace peak positions for the phred called bases, and tags attached to each individual sequence. Third, a .ace file created by Phrap

which contains the assembly information that includes the sequence and quality values of the contigs, tags attached to the contigs, sequence alignment information.

Consed has many functions, including displaying the error rate of the contiguous sequence, viewing quality of each read and consensus sequence, viewing assembly information of each read and consensus sequence, viewing and comparing traces, aligning and comparing two regions, picking primers and looking for homologous sequence regions. A further description of how consed was used to analyze data and to aid in finishing a sequenced region will be described in chapter 2.

1.6.4 BLAST

The BLAST (Basic Local Alignment Search Tool) is a search algorithm for finding ungapped, locally optimal sequence. It is used for sequence similarity searching and identifying genes and genetic features. It was developed at the National Center for Biotechnology Information (NCBI) of the National Library of Medicine (Altschul et al., 1990, 1997). BLAST can execute a sequence searches for an individual EST sequence against the entire DNA database in GenBank (see 1.6.7) in about 15 seconds on a Sun Ultra5 computer with a 360 MHC CPU.

The BLAST family of programs actually is five separate programs for rapid sequence database search, BlastN, BlastX, BlastP, TblastN, TblastX which use many of the statistical methods of Karlin and Altschul (1990, 1993). BLASTN compares a nucleotide query sequence against a nucleotide sequence database; TBLASTN compares a protein query sequence against a nucleotide sequence database dynamically translated in all six reading frames; BLASTX compares the six-frame translations of a nucleotide query sequence against a protein sequence database. TBLASTX compares the six-frame

translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database. BLASTP compares an amino acid query sequence against a protein sequence database. The parameter setting, input/output file format will be discussed in the chapter 2.

1.6.5 Powerblast

Powerblast (Zhang and Madden, 1997) was developed to meet the increasing demand for a more efficient annotation tool resulting from recent advances in large-scale genomic sequencing. It performs various types of query "masking" for repetitive elements and low complexity regions to reduce or eliminate misleading results. The long genomic sequence is split into short pieces so that the length of query sequences will not exceed the practical memory limitations imposed by the BLAST server. After performing BLAST search for shorter pieces, Powerblast postprocesses the BLAST search results by merging them to produce a final result for the whole single large sequence.

1.6.6 Genscan and Xgrail, NNPP, Promoter 2.0 and Repeatmasker

Xgrail (X-window based Gene Recognition and Analysis Internet Link) (Xu et al., 1994) is one of the most popular programs for computational exon prediction for genomic sequence as it provide an excellent visualization of the gene feature it predicts. Genscan (Burge and Karlin, 1997) is another gene prediction program and may have advantages over other computational gene prediction programs (Claverie, 1997). The output from both Xgrail and Genscan can be printed to a text file with the corresponding predicted peptide sequences, but only Xgrail creats a graphical output to display the location and DNA strand of each predicted feature.

A eukaryotic promoter is a region in genomic DNA where RNA polymerase binds and initiates transcription. RNA polymerases and transcription factors are binding on multiple functional sites in DNA, such as transcription start site, TATA-box, CAAT-box and GC-box. Many programs have been written to predict the promoter sequences, however, their success in making correct prediction is less than 50% (Fickett and Hatzigeorgiou, 1997). Two of the programs, Neural Network Promoter Prediction (NNPP) and Promoter 2.0, were used for promoter prediction during this dissertation research. According to the on-line documentation, NNPP (Waibel et al., 1989; Reese and Eeckman, 1995) was able to recognize 50% of the given promoters in a test with a set containing 42 known human gene promoters and 84 random DNA sequences. More information about NNPP is available on: http://www.fruitfly.org/seq_tools/nnppAbst.html. According to the on-line document for Promoter 2.0, "Promoter 2.0 is a program for predicting transcription start sites of vertebrate *Pol*II promoters in DNA sequences. It has been developed as an evolution of simulated transcription factor that interact with sequence in promoter region." The tested performance was 42% (Fickett and Hatzigeorgiou, 1997).

Repeatmasker, a public domain program written by A.F.A. Smit & P. Green (http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker), was used to search for repetitive elements in the sequences.

1.6.7 GenBank, Entrez, Sequin and dbEST

GenBank (Benson et al., 1996; http://www.ncbi.nim.nih.gov) is database established by the National Library of Medicine (NLM) at the National Institution of Health (NIH) as a repository for all publicly available nucleotide and protein sequences

and their corresponding biological and bibliographic information. The National Center for Biotechnology Information (NCBI) was created to be a division of NLM at the NIH in 1988 and GenBank has been based at NCBI since 1992. GenBank obtains data from three different sources: 1) Data submitted directly by authors; 2) Data exchanged with other international nucleotide sequence databases, such as European Molecular Biology Laboratory (EMBL) and the DNA Database of Japan (DDBJ); 3) Data obtained by scanning research journals.

Each GenBank entry includes the following information: 1) GenBank accession number, which is assigned to the gene when the author submitted it; 2) Concise description of the sequence; 3) Scientific name and taxonomy of the source organism; 4) Bibliographic references. 5) A table of features that identifies coding regions and other sites of biological significance such as transcription units, mutations, repeats and protein translation for coding region. 6) Comment with a GenBank identifier; 7) DNA sequence.

Entrez is NCBI's search and retrieval system that provides users with integrated access to DNA and protein sequence data, mapping, taxonomy, and structural data. Entrez also provides graphical views of sequences and chromosome maps. The references are linked to PubMed, a web search interface that provides access to the 9 million journal citations in MEDLINE and also contains links to full-text articles at participating publishers' web sites. 3-D structures of macromolecules also are available through Molecular Modeling Database (MMDB) which is based on the data in the Brookhaven Protein Data Bank (PDB).

Sequin is a stand-alone software tool developed by NCBI for submitting and updating entries to GenBank, EMBL, or DDBJ sequence databases. In this disseratation

research, cDNA sequences and genomic sequence are submitted to GenBank by Sequin, which is organized into a series of forms to enter the authors, the organisms, the sequences of nucleotides and amino acids, the gene and protein names, the annotation and the submission date. Sequin normally read sequence file in fasta format (section 2.4.3.2). More information is available at URL:

http://www.ncbi.nlm.nih.gov/Genbank/index.html#ref3

EST data cannot be submitted through the regular GenBank submission procedure (via BankIt or Sequin), except when a sequence is later characterized and annotated with biological features such as a coding region, 5'UTR, or 3'UTR. Thus, EST data is submitted through the database of Expressed Sequence Tags (dbEST) system. dbEST (Boguski et al., 1993) is a division of GenBank that contains EST sequence data and its annotation from a number of organisms. The latest information about dbEST is available at URL:

http://www.ncbi.nlm.nih.gov/dbEST/index.html.

Because EST projects generally result in a large numbers of sequences with a great deal of redundancy in the citation, submitter and library information, a special streamlined submission process and data format was designed by NCBI to improve the efficiency of submission and is available at URLs:

http://www.ncbi.nlm.nih.gov/dbEST/how_to_submit.html

and

http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html

1.7. *Fusarium sporotrichioides* cDNA library

Fusarium sporotrichioides is a filamentous fungus (Fig. 1.8) that produces several trichothecenes. Trichothecenes are compounds (Fig. 1.9) which are toxic to almost all animals including humans (Desjardins et al., 1993). Trichothecene toxic symptoms vary widely including vomiting, weight loss, hemorrhage, aleukia, immunosuppression, skin inflammation, feed refusal in cattle, pigs and poultry, and death in humans and other animals (Ueno 1980; Fort et al., 1993; Desjardins et al., 1993). *F. sporotrichioides* also is a maize, barley, rye, wheat and rice pathogen (Desjardins et al., 1993; Alexander et al. 1999). Contamination of grains with trichothecenes produced by *F. sporotrichioides* substantially reduces crop value. Contamination of trichothecenes in human foods and animal feeds is a continuing worldwide problem. Therefore, understanding of the molecular biology of trichothecene production is essential for the development of practical control strategy for trichochecene toxin production.

Almost all of the isolated trichothecene biosynthetic genes in *F. sporotrichioides* are located in a coordinately regulatory gene cluster within a 27-kb region (Desjardins et al., 1993; Hohn et al., 1995; Keller and Hohn, 1997; Alexander et al., 1999). Before this EST project, the only trichothecene biosynthetic gene that is not closely linked to other pathway genes is a 3-*O*-acetyltransferase encoded by *Tri101* that not only plays a acetyltransferase function but also a self-protection against trichothecenes (McCormick et al., 1999; Hohn et al., 1998; Kimura et al., 1998a, 1998b). For many pathway genes in the cluster, the gene products and functions have been determined and are listed as follows:

1) Trichodiene oxygenase (cytochrome P450 58) encoded by *Tri4* that function in oxygenation of trichodiene to yield a product of unknown structure (Hohn et al., 1995a);

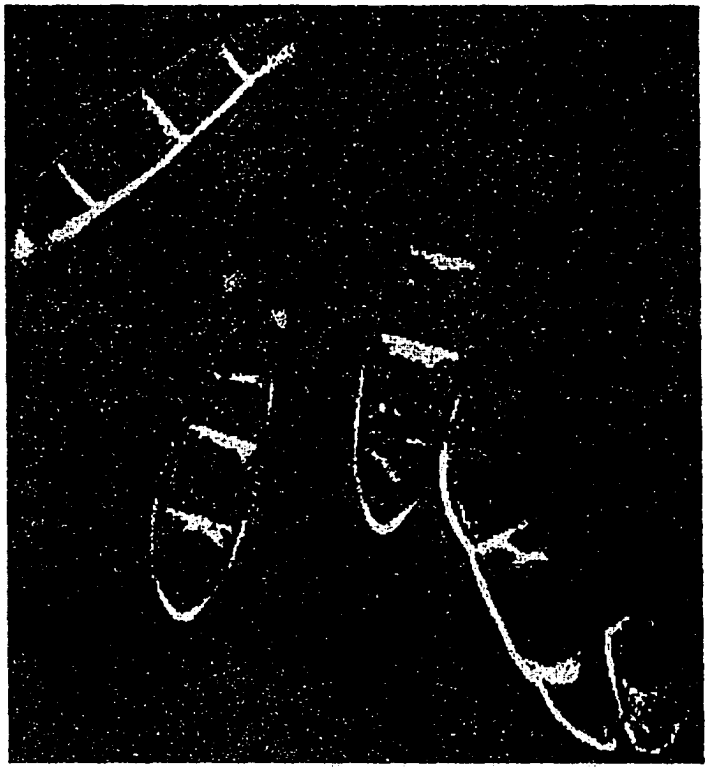


Fig. 1.8 Spores of *Fusarium sporotrichioides*

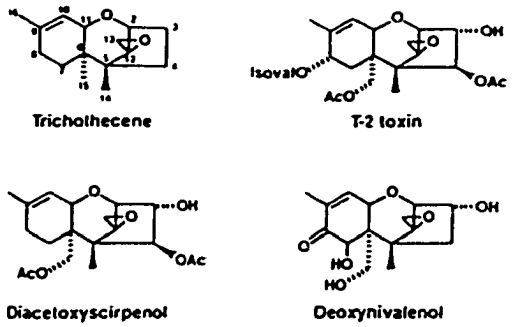


Fig. 1.9 Examples of four trichothecenes sharing the same tricyclic nucleus (Adapted from Desjardins, 1993)

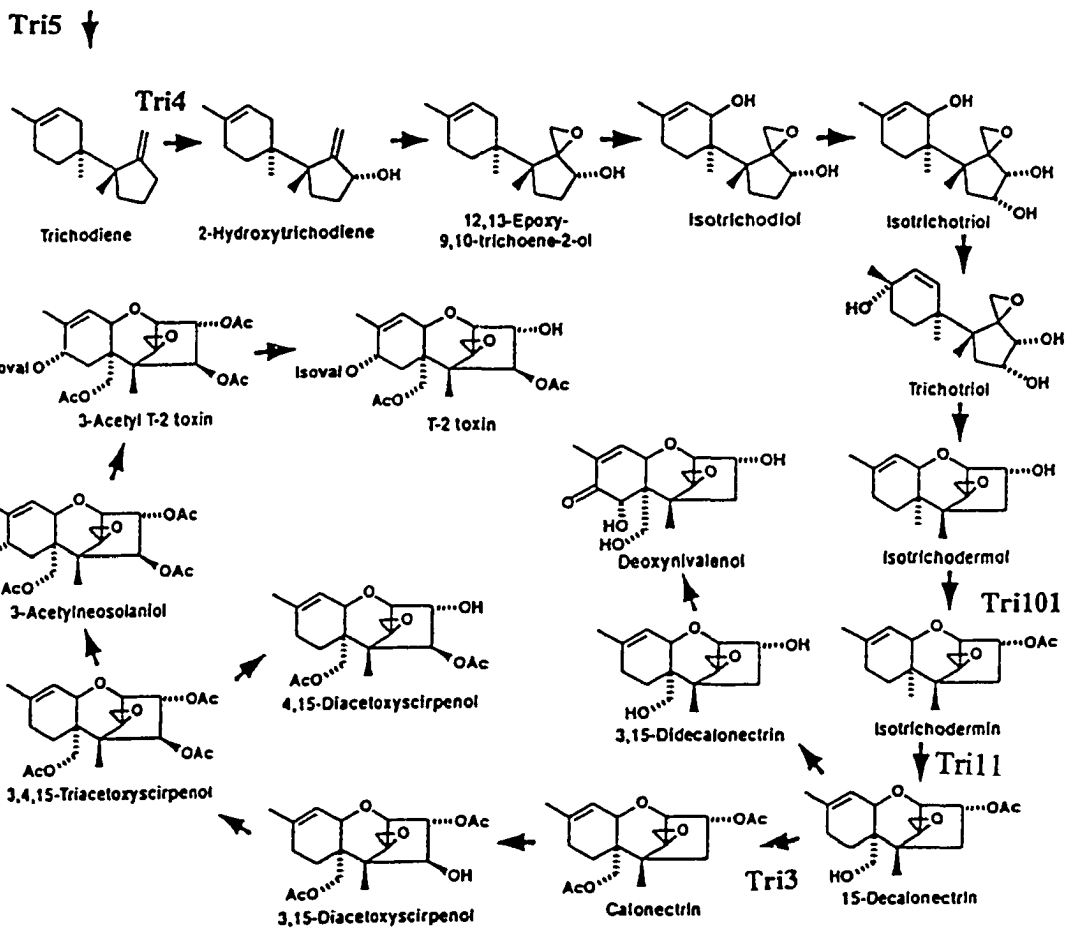


Fig. 1.10 Trichothecene biosynthetic pathway in *Fusarium* species (Adapted from Desjardins, 1993)

2) Trichodiene synthase encoded by *Tri5* that catalyzes the cyclization of farnesyldiphosphate (Hohn and Beremand, 1989);

3) Three acetyltransferases encoded by Tri3, Tri7 and Tri8 respectively, that are involved in the acetylation of the trichothecene hydroxyl groups (McCormick et al., 1996);

4) A transcription factor encoded by *Tri6* that is required for pathway gene expression (Proctor et al., 1995);

5) Isotrichodermin C-15 hydroxylase (cytochrome P450 65A1) encoded by *Tri11* that function in addition of a hydroxyl group at C-15 (Hohn et al., unpublished);

6) An efflux pump encoded by *Tri12* that is a transport protein (Alexander et al., 1999).

7) A regulatory gene, *Tri10*, within the cluster and a dramatic increase or decrease in toxin production resulted by mutations in *Tri10* gene (Peplow et al., 1997; Gurifulina et al., 1998; Tag et al., 1998).

Several new Tri family genes were identified by our collaborators using the result of this research, which will be discussed later (section 4.3).

The *F. sporotrichioides* cDNA library studied in this dissertation was constructed from a 23-hour culture containing an over producing mutation in the *Tri10* gene. The cDNA library has about 2-fold enrichment for the genes under the control of the regulatory *Tri10* gene (Beremand et al., unpublished). Sequencing ESTs from this cDNA library and constructing an EST database will help to identify other genes that also are controlled by the *Tri10* gene product. This data then will be used to construct a DNA microarray to study how *Tri10* controlled genes are regulated under different

environmental and genetic conditions. In addition, the new genes discovered during the EST project in this dissertation will provide additional information for further genomic and functional studies of this filamentous fungus.

1.8 Inborn gap in the published megabase sequence of the Ig λ light chain gene region of human chromosome 22

Now that the complete sequence of the first human chromosome, human chromosome 22, is available (Dunham et al., 1999), it is possible to ask global population genetics questions and determine the exact genomic structural changes that cause observed, altered phenotypes. By representational differential analysis (RAD), it was found that a deletion (R271) of about 200 nucleotides was present in cells of a kidney tumor (Lisitsyn et al., 1993; 1995). This deletion has been mapped to human chromosome 22q11.2 within the cluster of immunoglobulin λ light chain genes and pseudogenes via fluorescent in situ hybridization studies with molecular probes known to include the undeleted region (Esposito et al.). R271 deletion is one example of a common deletion polymorphism that occurs in various worldwide populations (Fig. 1.11). Since the human chromosome 22 reference sequence (Dunham et al., 1999) and the underlying immunoglobulin λ light chain regions (Kawasaki et al., 1997) are deleted in this region, these sequences must be from one or more individuals with a deletion-bearing chromosome. To search for the clones containing the deleted region, a BAC-contig panel, covering the entire long arm of chromosome 22 (Kim et al., 1996), was screened by PCR with a set of primer pairs (R271, AVA2, H61 shown in Table 1.2) (Siniscalco et al, 2000). Fortunately, these probes produced a large R271 amplification product from BAC

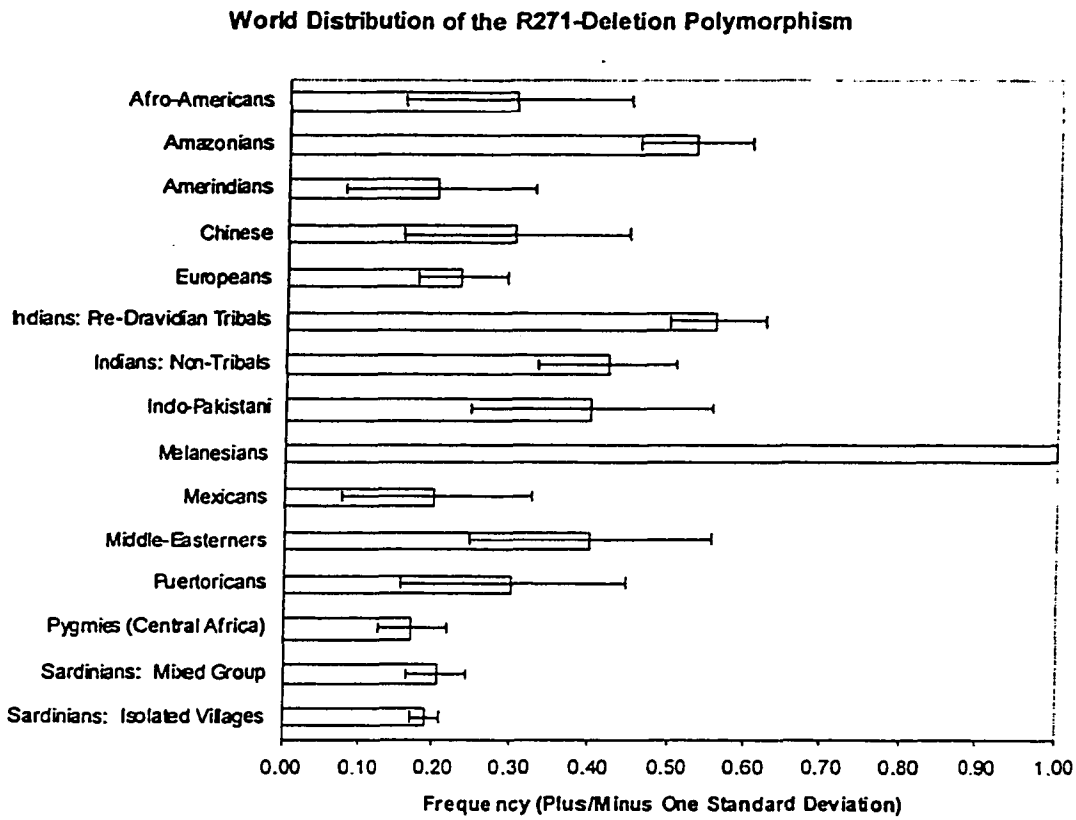


Fig. 1.11 World distribution of the deletion phenotype as detected by the R271 pair of primers. (Adapted from Siniscalco et al., 2000)

322F3 from the CalTech library. To investigate the sequence of this usually deleted region, BAC 322F3 subsequently was sequenced, analyzed and deposited into GenBank (AC009286). A manuscript describing this work recently was published (Siniscalco et al., 2000), and the genomic features present in BAC 322F3 will be described in chapter 5.

Table 1.2 Primers used in detecting the presence or absence of the amplification products for all deletions encountered at the 22q11.2:

		5'	3'
R271	Forward primer:	CTCAGCTAAGAATCCTCAGAGGATTG	
	Reverse primer:	GCCATCTTCCATTTTGGTATCAGTGC	
Ava2	Forward primer:	TGTGCCCGGGAAAGAGTTAG	
	Reverse primer:	TGGCCACCTCCTCTTTATTTC	
H61	Forward primer:	AGACAGAACCTGATACCGACCACA	
	Reverse primer:	AACACAAAGTCACTCCATTGCTG	

1.9 William Syndrome is associated with deletion of chromosome band 7q11.23.

William Syndrome (WMS), a genetic disorder that occurs in about one in twenty to fifty thousand births, is characterized by a set of physical features including mild mental retardation, congenital supravalvular aortic stenosis, a hoarse voice, transient neonatal hypercalcemia, and a set of facial features including broad forehead, medial eyebrow flare, flattened bridge of the nose and prominent lips with open mouth (Morris et al., 1988, 1993).

WMS has been found to be associated with a deletion of chromosome band 7q11.23. The smallest WMS deletion is ~2kb and includes both the elastin (Robinson et al., 1996; Gilbert-Dussardler et al., 1995; Nickerson et al., 1995; Ewart et al., 1993) and the LIM-kinase 1 gene (Frangiskakis et al., 1996). The elastin gene has been implicated in congenital heart disease (Ewart et al., 1994; Curran et al., 1993) while the LIM-kinase 1 contributes in part to the spatial deficit (Frangiskakis et al., 1996). Other genes mapped

to the 7q11.23 region and included in the larger region deletion in some WMS patients are: *RPC2* (Peoples et al., 1996), *WSCR1-5* (Osborne et al., 1996), *FZD9* (Wang et al., 1997), *STX1A* (Osborne et al., 1997), *GTF2I* (Perez-Jurado et al., 1998), *WS-betaTRP, WS-bHLH* and *BCL7B* (Meng et al., 1998a), *WSTF* (Lu et al., 1998), *FKBP6* and *CYLN2* (Hoogenraad et al., 1998), *CPE-R* and *RVP1* (Paperna et al., 1998). The gene for a neutrophil cytosolic factor 1 (*NCF1*) also has been mapped in this region but is outside of the larger WMS deletion region (Francke et al., 1990). Although WMS is caused by the direct and downstream effects of genes located within the commonly deleted regions, specific genes responsible for many of the major features of WMS remain unknown. As part of the effort to identify additional genes and other genomic features in the larger deleted region associated WMS and to develop a phenotypic cognitive map, the sequence of BAC 239c10 covering the WMS region in human chromosome 7, was completed and is included as part of this dissertation research.

Chapter II

Materials and Methods (Part 1)

Sequence and analysis of ESTs

2.1 Construction of cDNA library

2.1.1 Source of cDNA library

The *Fusarium sporotrichioides* cDNA library was constructed by A.W. Peplow, A.G.Tag, and M.N. Beremand in the Department of Plant Pathology & Microbiology, Texas A&M University in 1999. This cDNA library served as the source of cDNA templates in the EST project described in this dissertation.

The cDNA library was constructed from a 23 hour culture of an over producing mutant (^*Tri10*). It had about 2-fold enrichment for the genes under the control of the regulatory gene *Tri10* (Beremand, personal communication).

To provide the background information for this project, the construction of the *Fusarium sporotrichioides* cDNA library using the ZAP-cDNA Synthesis Kit (Catalog #200400) from Stratagene (11011 North Torrey Pines Road, La Jolla, CA 92037. techservices@stratagene.com) is briefly summarized in the following sections.

2.1.2 cDNA synthesis

2.1.2.1 Synthesis of the first strand cDNA

The mRNAs were islated by our collaborators at Texas A&M University and sent to Stratagene to construct the cDNA library. The first strand cDNA synthesis was accomplished by Stratagene by incubating the four deoxy nucleotides, the buffer, a poly (dT) primer and reverse transcriptase (RT) in the presence of an aliquot of the isolated *F.*

sporotrichioides mRNA (Fig.2.1). The nucleotide mixture for the first strand cDNA contained dATP, dGTP, dTTP and the analog 5-methyl dCTP.

The 5-methyl dCTP replaces dCTP to yield a first strand containing a methyl group on each cytosine base, which protects this strand from the digestion by the restriction enzymes used in the subsequent steps.

The primer used was a linker-primer which was a 50-base oligonucleotide containing a "GAGA" sequence, the *Xho* I restriction enzyme recognition site and an 18-base poly(dT). The sequence was as following:

5' GAGAGAGAGAGAGAGAGAGAACTAGT**CTCGAG**$(T)_{18}$ 3'

Xho I

The "GAGA" sequence was designed to protect the *Xho* I site between "GAGA" sequence and poly (dT) sequence. The *Xho* I site allows the final double stranded cDNA produced after the second strand synthesis and *Xho*I and *Eco*RI digest to be inserted into the Uni-ZAP vector in a designed orientation with respect to the *lacZ* promotor. The poly(dT) sequences allowed the binding of the primers with the 3' poly(A) sequences of the mRNA molecules. The RT used was MMLV-RT (Moloney murine leukemia virus reverse transcriptase).

2.1.2.2 Synthesis of the second strand cDNA

The second strand cDNA synthesis began with the addition of RNase H which digests the original mRNA strand in the RNA-cDNA hydbrid helix to produce a set of RNA fragments which served as the primers for synthesis of the second stand cDNA by DNA polymerase I. In contrast to the first strand synthesis, the nucleotide mixture for second strand DNA contained dCTP instead of the 5-methyl dCTP to ensure that the

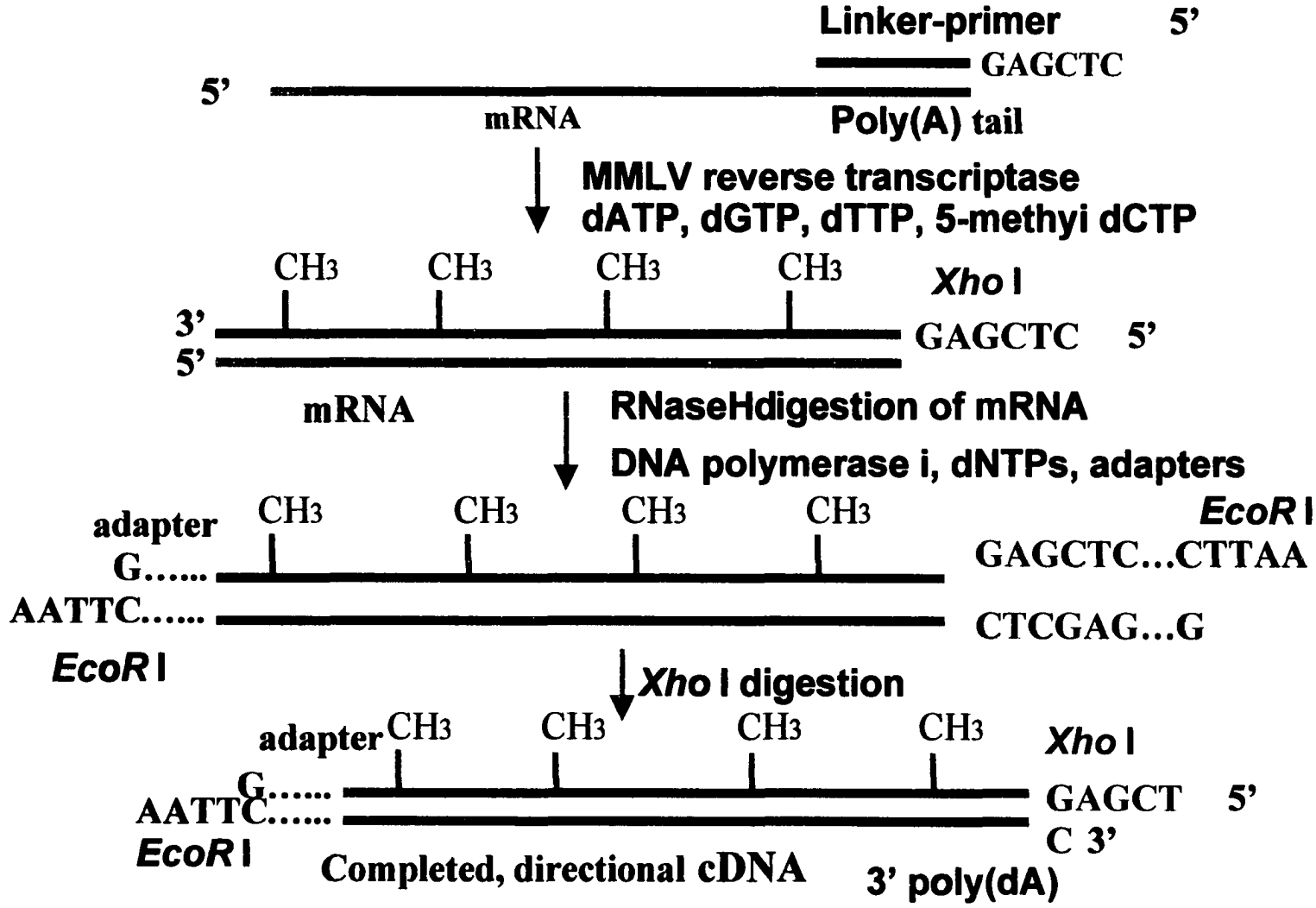


Fig 2.1 Synthesis of a cDNA with different cohesive ends

resulting restriction site in the linker-primer region would be accessible to restriction enzyme digestion. After the single strand overhangs on the termini of the double-stranded cDNA was made blunt ended by incubation with *Pfu* DNA polymerase, *Eco*R I adapters were ligated to the blunt ends. The sequence of the *Eco*R I adapter was:

5' **AATTC**GGCACGAG 3'

3' **G**CCGTGCTC 5'

*Eco*R I

The double-stranded cDNAs were digested with *Xho* I and *Eco*R I to release the *Eco*R I adapter and "GAGA" residues at the 3' end of the cDNA molecules so the resulting cDNA has different cohesive termini on each end. The 5' end has an *EcoR* I cohesive termini and the 3' end has the *Xho* I cohesive termini (Fig. 2.1), which orients the cDNA when they are inserted into the vectors.

2.1.3 Insert the cDNA into the vector and amplify the plasmid

The vector used in this project was Uni-ZAP XR system, which has the advantages of the high efficiency of a lambda library construction and the convenience of a plasmid system. Prior to cDNA insertion, the lambda phage must be cleaved into three segments by treatment with *EcoR* I and *Xho* I. After cleavage, the middle section was removed and the cDNA was ligated between two remaining arms to replace the middle portion of the phage DNA. This linear molecular then was packaged in vitro into phage particles to give the primary library.

The primary library was mixed with ExAssist helper phage, which is a mutated f1 phage containing an amber mutation. The mix was used to infect XL1-Blue-MRF` cell, which was an ExAssist phage host because it was an suppressing *E. coli* strain. The f1

polymerase produced by ExAssist phage replicated part of lambda molecule including pBlueScript SK- and the insert cDNA, and the resulting molecule was packaged into an ExAssist helper phage particle. The resulting particle had an ExAssist helper phage at outer part and a pBlueScript with cDNA in inside part. ExAssist helper phage also produced gene II product for circularization of pBlueScript. Therefore, ExAssist helper phage helped the efficient excision of the pBluescript phagemid from the lambda phage.

Then the lysate, the mix of ExAssist helper phage and pBlueScript (with the cDNA insert) packaged in ExAssist helper phage, were isolated and used to infect SOLR strain, which was not an ExAssist helper phage host. ExAssist helper phage could not replicate in SOLR cells because ExAssist helper phage had an amber mutation but SOLR cells were nonsuppressing *E.coli* strain. But pBlueScript could replicate because it had ColE1 origin, which was the plamid origin of replication used in the absence of helper phage. The pBluescript phagemid is derived from pUC19 and is 2958 bp without inserts. It has *lacZ* gene, which provided the α-complementation for blue-white color selection of recombinant phagemid. It has an Ampicillin-resistance gene for antibiotic selection of the phagemid vector.

2.1.4 Amplification of the cDNA clones

Cells which contained the pBlueScript with the cDNA inserts were selected and cultured in 384 well plates in Freezer Medium, which was prepared in a ratio of 9 volumes of LB and 1 volume of 10X FM and Ampicillin(100 μg/ml). LB medium contains 10g/l NaCl, 10g/l tryptone, yeast extract 5 g/l, and is autoclaved 20 minutes at 121°C. 10X FM medium contains 62.7 g/l K_2HPO_4, 17.96 g/l KH_2PO_4, 5 g/l sodium citrate, 0.98g/l $MgSO_4 \cdot 7H_2O$, 8.98 g/l $(NH_4)_2SO_4$ and 440 ml/l glycerol and is sterilized

by filtration through 0.2μm filter. Cells were frozen in -70°C after growth at 37 °C for 20 hours and then were sent to our laboratory.

2.2 Growth of cDNA clones and isolation of templates

2.2.1 Growth of cDNA clones

About 2 μl of each cDNA clone culture was transferred from a 384 well plate to a 96 deep well microtiter plate containing in each well 1.5 ml TB media and Ampicillin(100 μg/ml). TB media contains 2.31g/l KH_2PO_4, 12.54 g/l K_2HPO_4, 12 g/l Bacto-tryptone, 24 g/l yeast extract, 4 ml glycerol. After incubating by shaking for 18-20 hours at the 350 rpm at 37°C, the cells were harvested by centrifugation at 1800 rpm for 7 minutes. The supernatant was decanted and pellets were drained. The blocks containing the cell pellets were stored at -20°C.

2.2.2 Isolation of cDNA templates

The cDNA templates were isolated by single or double acetate cleared lysis procedures using the Biomek 2000 (Beckman) and automated Hydra 96 (Robbins). The following steps are identical to both the single and double acetate protocols.

The pellets were thawed and resuspended in 200 μl TE-RNaseA using the Biomek 2000 which was programmed to mix pellets and solution by pipetting the cell solution up and down 20 times. TE-RNaseA contains 10 mM Tris-HCl, pH 7.6, 1 mM EDTA, 42 μg/μl RNase A. Then 200 μl of lysis solution which contains 0.2N NaOH and 1 % SDS was added to the resuspended cells using the programmed Hydra. After incubated at 37°C in the shaker at 250 rpm for 10-15 minutes, 200 μl 3M NaOAc or KOAc, pH 4.5 was added to each well again using the Hydra. Each block was sealed,

vortexed vigorously and shaken at 37°C at 350 rpm for 20 minutes. The blocks were stored at -70°C overnight. The blocks were removed from the freezer and thawed before the lysate was cleared by centrifugation at 4250 rmp for 45 minutes in the JOUAN KR422 Centrifuge. Then the Hydra was used to remove the upper 400 µl of the cleared lysate supernant to another block. Manually, 1 ml 100% ethanol was added to each well using a repeater (Eppendorf). The blocks were incubated in an ice/water bath for 15-30 minutes and centrifuged at 3000 rpm for 30 minutes at 4 °C. The supernatant was decanted and the DNA pellet was washed with 500 µl 70% ethanol and centrifuged again for 10 to 15 minutes at 3000 rpm. The pellets were dried after decanting the supernatant and dissolved in 100 µl of doubly distilled water if the single acetate protocol was followed. If double acetate protocol was chosen, the pellet was not dissolved in ddH_2O, but instead, was dissolved in 200 µl of 10:1 TE pH 8.0 and 100 µl of 7.5M KOAc. After freezing the block at -20 °C or -70 °C overnight, it was removed from the freezer, thawed and centrifuged for 45 minutes at 4250 rpm. The top 200 µl of supernatant was transferred to another block and the DNA was precipitated with 500 µl of 100% ethanol at room temperature and centrifuged at 3000 rpm for 10 minutes at 4 °C. After decanting the ethanol, the pellet was washed with 500 µl of 70% ethanol two times and centrifuged at 3000 rpm for 20 minutes at 4 °C. The pellets were dried and dissolved in 100 µl of doubly distilled water.

The purity of DNA isolated using the double acetate procedures was much higher than that isolated using the single acetate procedures. Therefore, template prepared by double acetate procedure was suitable for sequencing reactions that were to be loaded on 3700 capillary DNA sequencers.

2.3 Sequencing of ESTs

2.3.1 Set up reactions

Following ethanol precipitation, drying, and re-suspending the templates in 100 ul of doubly distilled water, the concentration of cDNAs were estimated by agarose gel electrophoresis to decide the amount of template used in the following thermocycle reactions.

Using a Robbins Hydra robot, 2-4 ul (200-400 ng) of sequencing template was robotically transferred into a 384 well Robbins microtube plate (TM384ET). The hydra which pipets 96 samples at a time also is equipped with a rotating stage to accommodate a 384 well plate. Subsequently, 4 μl sequencing reaction mix was robotically pipetted by the Robbins Hydra 96 or added in by repeat pipetter and the plate was incubated in a thermocycler.

4 μl sequencing reaction mix contains:

1) 1 μl (30 pmoles) universal forward primer solution or t3 primer solution,

2) 2 μl 1:3 diluted ABI BigDye stock (dilute 2 μl of ABI stock with 4 μl of 5x reaction buffer). 5x reaction buffer contains 400mM Tris-HCl and 10mM $MgCl_2$, pH 9.0,

3) 1 μl 25% DMSO,

in a final reaction volume of 6-8 μl.

ABI PRISM BigDye stock contains: BigDye terminators, dNTP, AmpliTaq DNA polymerase FS, MgCl2 and reaction buffer. The concentrations of each chemicals are not made available to the users by the manufacture.

BigDye terminators are single energy transfer molecules. An energy transfer linker couples the donor fluorescein and acceptor dRhodamine dyes for every efficient energy transfer in a single dye molecule, which makes BigDye terminator a very bright fluorecent terminator. Additional information is available at URL:

http://www.appliedbiosystems.com/MolecularBiology/about/dna/dna_seq/bdterm.html

2.3.2 Incubate reactions

The reactions were incubated in a Perkin Elmer GeneAmp PCR system 9600 thermal cycler with the FS thermo-cycling conditions for 60 or 99 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes.

2.3.3 Purify reactions

Once the cycling reactions were completed, the unincorporated dye terminators were removed by chromatography through Sephadex G-50 spin columns made in a 96-well micro-titer filter plate (Millipore, MASVN6550). The G-50 columns were hydrated at least 3 hours before use by adding 250 ml ddH_2O to 20 g Sephadex G-50 powder. The purified reactions were collected by centrifugation at 1450 rpm, dried in a vaccum oven at room temperature and then stored at -20°C. All the required steps from preparing the columns to transferring the reaction mix from 384 well plates to 96 well columns were done with the Robbins Hydra 96. More recently, this spin column purification was replaced by a room temperature precipitation with 95% ethanol/acetate followed by rinsing with 70% ethanol.

2.3.4 Collect data

To prepare samples for the ABI 377 DNA sequencer, each reaction was disolved in 1 μl loading buffer and heated at 95°C for 3 minutes. The loading buffer contains 5mM EDTA and 10 mg/ml blue dextran in deionized formamide. The samples were loaded onto a polyacrylamide gel (5-5.5%). After 6-8 hours of electrophoresis and data collection on the ABI 377 sequencer, the resulting data was re-tracked manually, analyzed with the ABI supplied software. The analyzed data were transferred via Fetch to networked Sun workstations for sequence assembly and analysis.

To prepare samples for the ABI 3700, samples were disolved in 20 μl ddH_2O. After two hours of electrophoresis and data collection on ABI 3700, the resulting sequences were transferred from ABI 3700 to the Sun workstations for further analysis.

2.4 Computer programs used in data analysis

Phred, Phrap, Cross_match, consed and BLAST are computer programs that are frequently used during data analysis of the EST project and human BAC, PAC and cosmid squencing projects described in this dissertation. They are discussed in detail here and were briefly introduced earlier in section 1.6.

2.4.1 Phred , Phrap and Cross_match

Phred (Ewing et al., 1998, Ewing and Green, 1998), Phrap and cross_match (Green, copyright 1994-1996) are software developed in Phil Green's group at the University of Washington. Phred examines the trace files obtained from DNA sequencers (e.g. ABI377, ABI3700), calls bases and assigns quality values to bases. Phrap is a program for assembling sequence data. Using Phred quality values, Phrap assembles sequences by aligning the homologous high quality regions and then constructing

contiguous, consensus sequences as a mosaic of the sequence data. Phrap refers to the Phred quality values and also uses its own algorithm to assign a quality value to each base in the consensus sequence. Cross_match is a general-purpose utility for comparing any two sets of DNA sequences. A script, phredPhrap (David Gordon, 1995), automatically processes data through Phred and Phrap after it is transferred from DNA sequencers. The results of the assembly are stored in files in the projects' edit_dir. Table 2.1 is an example of the output files produced by phredPhrap after processing *F. sporotrichioides* ESTs.

Table 2.1 Output files in edit_dir for *F. sporotrichioides* EST project after using Phred, Phrap and cross_match

FS.fasta
FS.fasta.screen
FS.screen.out
FS.fasta.screen.qual
FS.fasta.screen.singlets
FS.fasta.screen.contigs
FS.fasta.screen.contigs.qual
FS.fasta.screen.ace
FS.fasta.log
FS.fasta.screen.log
FS.phrap.out
FS.tar

2.4.1.1 Use of Phred to call bases and generate fasta file

Although DNA sequencers (e.g. 377 and 3700) have their own base calling software, Phred generally gives more accurate base calls than the ABI software does (Ewing et al., 1998; Ewing and Green, 1998) and is the default standard in almost all genome centers.

Three directories, chromat_dir, edit_dir and phd_dir, are created before executing Phred and Phrap. The chromat files (trace files) representing the raw sequence data are transferred from the sequencers and stored in chromat_dir. Phred automatically detects the sequencer chromat file format and uncompresses the compressed data files.

Phred uses simple Fourier methods in its algorithms in determine a sequence of base-calls from the processed trace files. The basic procedures used by Phred include:

First, predicting the idealized peak locations (representing bases) based on the fact that fragments are locally relatively evenly spaced in most of regions of the gel.

Second, identifying the observed peaks in the trace file.

Third, matching the observed peaks with the predicted peaks.

Fourth, rechecking the unmatched peaks. If any unmatched peak meet required conditions (Ewing et al., 1998), an additional predicted peak is created and the observed peak is assigned to it.

After calling the bases, Phred writes the sequences to files in fasta format and stores the results in fasta file (e.g. FS.fasta in *F. sporotrichioides* project) in edit_dir (Table 2.1). Table 2.2 is an example of a subfile in fasta format for one EST in .fasta file. Each sequence has a single header line with leading character ">" followed by the read name and descriptive information about the read. The sequence with an arbitrary length follows the header starting in a separate line.

Table 2.2 A subfile in FS.fasta for a 5' EST in fasta format

>a1a08fs.r1 CHROMAT_FILE: a1a08fs.r1 PHD_FILE: a1a08fs.r1.phd.1 TIME: Thu Feb 25 08:33:30 1999
gaagttggacccgttcatgaaacaaaagctggagctccaccgcggtggcg
gccgctctagaactagtggatcccccgggctgcaggccttattctgcctc

tattggtattatttccttccgaaatatccttcacactggttgtcaagtac
aaggtttgttgagcatgaagatctcaacctctgttgtcaccggggccttc
gccttttcgctatagcacctgcatcagccatcggtagactcgtcgattc
ctatcccaaatacattcgcgaggccgaaccaacatgtataactcctggta
tcattcccaacatcaccggatctcagatccgtaacgttggggttgtcctt
ttccaggcattcgatatgatggacgtctttggccctctggatcctctcca
gctcatctcgctaggcgtccagaagctgaaccttcacctcatagccgaaa
cactcgatccagtcaccacagcacctgtggctatgaacaagttcgactca
ag

2.4.1.2 Use of cross_match to screen out vector sequence and generate .fasta.screen file

Any contaminating vector sequences are screened out using the cross_match feature embedded in Phred. A fasta format file called "vector.fasta" containing all the vector sequences is compared with each individual sequence by Cross_match using the command:

$ cross_match FS.fasta vector.fasta -minmatch 12 -minscore 20 -screen > screen.out

The parameter minmatch defines the minimum length of a matching word at the beginning of the comparison. The default value for minmatch is 14. If minmatch is 0, a full comparison is done. The parameter minscore defines minimum comparison score. The default value is 30. It was earlier determined by others in our laboratory that these more stringent parameter settings are more efficient in removing vector sequences than the default settings.

The screened sequence files are generated and stored in FS.fasta.screen, which contains a "vector masked" version of the original sequence. In this version, any region that matches any part of a vector sequence is replaced by corresponding numbers of X. Table 2.3 is an example of a subfile in FS.fasta.screen file. The results of cross_match are recorded in FS.screen.out.

Table 2.3 A subfile in FS.fasta.screen for a 5' EST after screening out the vector sequence

>a1a08fs.r1 CHROMAT_FILE: a1a08fs.r1 PHD_FILE: a1a08fs.r1.phd.1 TIME: Thu Feb 25 08:33:30 1999
GAAGTTGGACCCGTTCATGA*XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCCTTATTCTGCCTC
TATTGGTATTATTTCCTTCCGAAATATCCTTCACACTGGTTGTCAAGTAC
AAGGTTTGTTGAGCATGAAGATCTCAACCTCTGTTGTCACCGGGGCCTTC
GCCTTTTTCGCTATAGCACCTGCATCAGCCATCGGTAGACTCGTCGATTC
CTATCCCAAATACATTCGCGAGGCCGAACCAACATGTATAACTCCTGGTA
TCATTCCCAACATCACCGGATCTCAGATCCGTAACGTTGGGGTTGTCCTT
TTCCAGGCATTCGATATGATGGACGTCTTTGGCCCTCTGGATCCTCTCCA
GCTCATCTCGCTAGGCGTCCAGAAGCTGAACCTTCACCTCATAGCCGAAA
CACTCGATCCAGTCACCACAGCACCTGTGGCTATGAACAAGTTCGACTCA
AG

*These 20 bases were vector sequences that had not been screened out, because the quality values were too low for the sequences to be recognized as the vector sequences.

2.4.1.3 Algorithm used by Phred to calculate quality values

The method used by Phred to determine the base calling error probability is heuristic as it does not require the knowledge of the true sequence. Instead it is based on the trace characteristics, where Phred computes an error probability p for each base. If one base is misinterpreted, such as where a background peak is misinterpreted as a sample peak, then the distance between this misinterpreted peak and the following sample peak changes. This spacing change is an indication of an error. Generally speaking, misinterpretation of peaks in a region of the trace would cause presentations of indications of errors in the vicinity of the erroneous peak, although not at the peak itself (Ewing and Green, 1998). Four trace parameters are employed to discriminate incorrect from correct base-calls. These parameters are peak spacing, uncalled/called ratio in a

window of seven peaks, uncalled/called ratio in a window of three peaks, and peak resolution. In each case, the smaller the parameters, the smaller the p values.

Peak spacing is defined as the ratio of the largest peak-to-peak spacing to the smallest peak-to-peak spacing in a window of seven peaks centered on the current one. The smaller the ratio, the more evenly the peaks spaced, and the lower the p values.

The uncalled/called ratio is defined as the ratio of the amplitude of the largest uncalled peak to the smallest called peak, in a window of seven peaks centered on the current one. If the called base is an N, value of 100.0 is assigned. If there is no uncalled peaks, a value of 0 is assigned. This ratio is calculated again using a window of three peaks.

Peak resolution is defined as the inverse of the number of bases between the current base and the nearest unresolved base. The minimum possible value should be half the number of bases in the trace times -1 and the maximum should be 0.

Error probability p is converted to a quality value q using the transformation:

$q = -10 \times \log_{10}(p)$

That is:

$p = 10^{-(q/10)}$

Quality values q should be an integer between 0 and 99. A quality value of 10, 20, 30 corresponds to an error probability of 1/10, 1/100, 1/1000 respectively. The quality value of 99 is reserved for base calls that have been edited and the user has visually inspected and verified it as "highly accurate" during editing (high quality edit). A quality value of 98 is given to bases that have been manually edited. However, if it was not

possible to determine a base accurately, this quality value is set to 0 in Phrap as it is a low quality edit.

2.4.1.4 Use of Phred to generate .screen.qual file and .phd files

The quality value of each base is written into a .phd file and stored in the phd_dir. Table 2.4 gives an example for quality output for one EST stored in one .phd file. After the header information, there are many lines, each line has one letter and two numbers. The first letter is the base-call, the first number is the q value and the second number is the location of the peak on the trace file.

The quality output also is stored in FS.fasta.qual. The quality values are written into FS.fasta.screen.qual by moving or copying FS.fasta.qual to FS.fasta.screen.qual using the command

$ mv FS.fasta.qual FS.fasta.screen.qual

or

$ cp FS.fasta.qual FS.fasta.screen.qual

Table 2.5 gives an example for quality output for one 5' EST stored in FS.fasta.screen.qual. The format is similar to that of the corresponding fasta file. Each sequence has a header line identical (except for the "description" field) to that in the fasta file. The quality value for each base is followed the header starting in a separate line. The order of reads in the .fasta file and the order of reads in corresponding the .fasta.qual file must be identical.

Base "N" (a base which cannot be assigned a name confidently) or "X" (a vector sequence) are automatically assigned quality 0 in the .ace file (section 2.4.1.5). The

FS.fasta.qual files are appropriate quality files for input to Phrap. Both files,

FS.fasta.screen and FS.fasta.screen.qual, are used for all subsequent analysis.

Table 2.4 One of the phd.1 files in phd_dir

BEGIN_SEQUENCE a1a08fs.r1
BEGIN_COMMENT
CHROMAT_FILE: a1a08fs.r1
ABI_THUMBPRINT: 00000000000000000000000000000000
PHRED_VERSION: 0.961028
CALL_METHOD: phred
QUALITY_LEVELS: 99
TIME: Thu Feb 25 08:33:30 1999
TRACE_ARRAY_MIN_INDEX: 0
TRACE_ARRAY_MAX_INDEX: 6448
END_COMMENT

BEGIN_DNA
g 6 7
a 8 21
a 8 32
g 6 43
t 9 62
t 8 65
g 6 75
g 8 88
a 6 102
c 6 103
c 8 117
c 8 135
g 7 143
t 7 160
t 6 168
c 4 179
a 4 201
t 4 203
g 6 220
a 10 229
a 7 242
.....
g 32 6341
a 24 6354
c 32 6366

t 26 6382
c 25 6397
a 25 6409
a 11 6425
g 9 6438
END_DNA
END_SEQUENCE

Table 2.5 A subfile in FS.fasta.screen.qual for quality

>a1a08fs.r1 PHD_FILE: a1a08fs.r1.phd.1
6 8 8 6 9 8 6 8 6 6 8 8 7 7 6 4 4 4 6 10 7 4 10 10
15 14 8 8 14 18 18 20 17 17 17 20 23 24 24 24 24 34
37 49 49 49 45 45 45 40 40 37 37 37 37 45 45 40 40
40 40 40 40 40 40 40 42 45 42 42 42 42 42 45 45 49
49 49 49 49 49 49 49 49 45 49 49 45 45 45 45 45 45
45 45 45 45 45 45 40 40 40 40 40 40 49 49 45 45 45
45 45 45 49 49 49 49 49 49 49 49 49 49 49 45 45 45
45 49 45 45 49 49 45 45 45 45 45 45 45 45 45 45 45
45 45 45 45 45 45 49 51 51 51 51 37 40 40 40 40 42
51 51 51 51 51 51 51 51 41 41 45 45 45 45 40 42 42
40 38 38 39 39 39 39 39 43 43 40 42 40 49 49 49 51
51 51 51 39 40 40 42 40 42 51 51 49 40 40 40 40 40
40 45 41 41 45 45 45 45 45 45 41 41 45 51 51 51 51
51 51 51 45 45 39 39 39 40 45 45 39 39 39 40 40 42
39 40 45 45 51 51 51 51 51 51 51 51 45 45 45 45 45
45 45 39 33 33 33 33 33 39 39 39 39 39 39 39 39 39
39 39 45 45 45 51 51 51 51 45 45 45 51 51 45 51 51
51 51 51 51 40 40 42 42 42 42 45 45 43 43 43 45 45
40 40 40 40 39 39 45 45 41 51 51 51 51 45 39 39 39
39 39 39 45 45 45 45 51 51 51 51 39 39 39 39 39 41
51 43 45 41 41 41 37 37 51 45 45 38 37 37 37 37 37
41 51 51 51 51 41 37 37 37 37 37 37 51 45 41 37 37
37 37 40 45 45 45 45 45 45 45 45 45 45 45 45 45 45
45 45 45 45 45 45 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 32 37 34 34 37 40 37 34 34 34 32 33
34 37 37 38 38 37 37 37 38 37 37 34 34 34 34 34 37
37 37 37 36 34 34 37 38 38 34 34 34 34 37 37 37 37
37 37 37 37 37 37 37 37 37 34 34 34 34 34 34 34 34
34 33 33 33 33 29 33 32 33 32 32 24 32 26 25 25 11
9

2.4.1.5 Use of Phrap to assemble the sequences

The input files for Phrap are of two types, sequence file (e.g. FS.fasta.screen) and quality file (e.g. FS.fasta.screen.qual).

Phrap uses the SWAT (Smith and Waterman, 1981) algorithm to assemble sequences. Using Phred quality values, Phrap assembles sequences by combining the homologous high quality parts together and then constructing contiguous sequences as a mosaic to generate consensus sequences called contigs.

Phrap uses the Phred quality information to generate its own quality measures on the basis of read-read confirmation information. If a base call at a given position is confirmed by the base call from an opposite-strand or one from a different-chemistry, that position is given a quality which is the sum of the two input base call qualities. If more than one opposite-strand base call confirms a given one, only the highest single quality value is used for the calculation of the sum. The quality value assigned to a contig position is adjusted upward to take the highest quality value of any base call at that position and is adjusted downward to take into account any discrepancies with base calls.

The output files generated from a Phrap assembly are:

(1) a singlets file (e.g. FS.fasta.screen.singlets)

This fasta file contains the sequences of the singlets with no match to any other sequences.

(2) a contigs file (e.g. FS.fasta.screen.contigs)

This fasta file contains the consensus sequences of the contigs.

(3) a contigs.qual file (e.g. FS.fasta.screen.contigs.qual)

The corresponding quality file generating by the method described above.

(4) an ace file (e.g. FS.fasta.screen.ace)

This file is one of the input files for consed, software for viewing and editing assemblies, contains the assembly information including sequences and quality values of the contigs, tags attached to the contig sequences, and read alignment information. Tags are annotations relevant to the assembly.

(5) a log file (e.g. FS.fasta.log and FS.fasta.screen.log)

This is a summary of how the assembly or screen proceeded.

The files produced by Phred, Phrap and cross_match during data analysis are summarized in Table 2.1.

2.4.2 Consed

Consed is software for viewing and editing assemblies (Gordon et al., 1998). It is especially helpful in the finishing stage of human sequencing projects as well as in closing gaps for ESTs to obtain full-length cDNAs. Consed requires at least three types of input files. First, the chromatogram files located in the chromat_dir which contain the fluorescence trace profiles (one for each read) created by the DNA sequencer. Second, the phd files in the phd_dir (one for each read) created by Phred which contain the base calls, quality values, and trace peak positions for the read bases, and tags attached to the read. Third, the ace file in edit_dir created by Phrap which contains the assembly information including sequences and quality values of the contigs, tags attached to the contigs, and read alignment information.

Upon Consed the startup, after the user has selected the appropriate ace file from the available ace files, Consed will read the ace file and the associated phd files, and a list of all contigs in one window and a list of individual sequences used to construct the

contigs in another window. When a contig or a sequence name is double clicked, Consed will display the selected contig or the contig containing the sequence in the aligned reads window (Fig. 2.2) with the underlying individual aligned sequence. Several features of Consed are described in the following section.

2.4.2.1. Display error rate of the contiguous sequence

Phred uses trace parameters to calculate the error probabilities associated with each base and phrap uses these error probabilities together with the read alignments to attach an error probability to each base of the concensus sequence. All sequencing projects should have a predefined accuracy target for the finishing sequence. For human genome projects, the acceptable error rate is less than 1 error per 10,000 bases (Dunham et al., 1999). This rate corresponds to a quality of 50 according to $q = -10 \log_{10}(p)$ where p is 10^{-5}. Consed will calculate the expected number of errors in the entire consensus sequence by generating a sum of the per-base error probabilities from all bases. Then, Consed displays the over all error probability in a window in the unit of error per 10,000 bases, which is very convenient to the user for checking closure stage for shotgun sequencing projects (Fig. 2.2).

2.4.2.2. Viewing quality of each read and consensus sequence

There are several choices of color mode for the aligned read window. If the "color means quality and tags" mode is chosen, different shades of background colors around the bases will be displayed (Fig. 2.2). They have the following meanings:

consed.colorMeansQuality0_4: grey39 (very dark)

consed.colorMeansQuality5_9: grey47

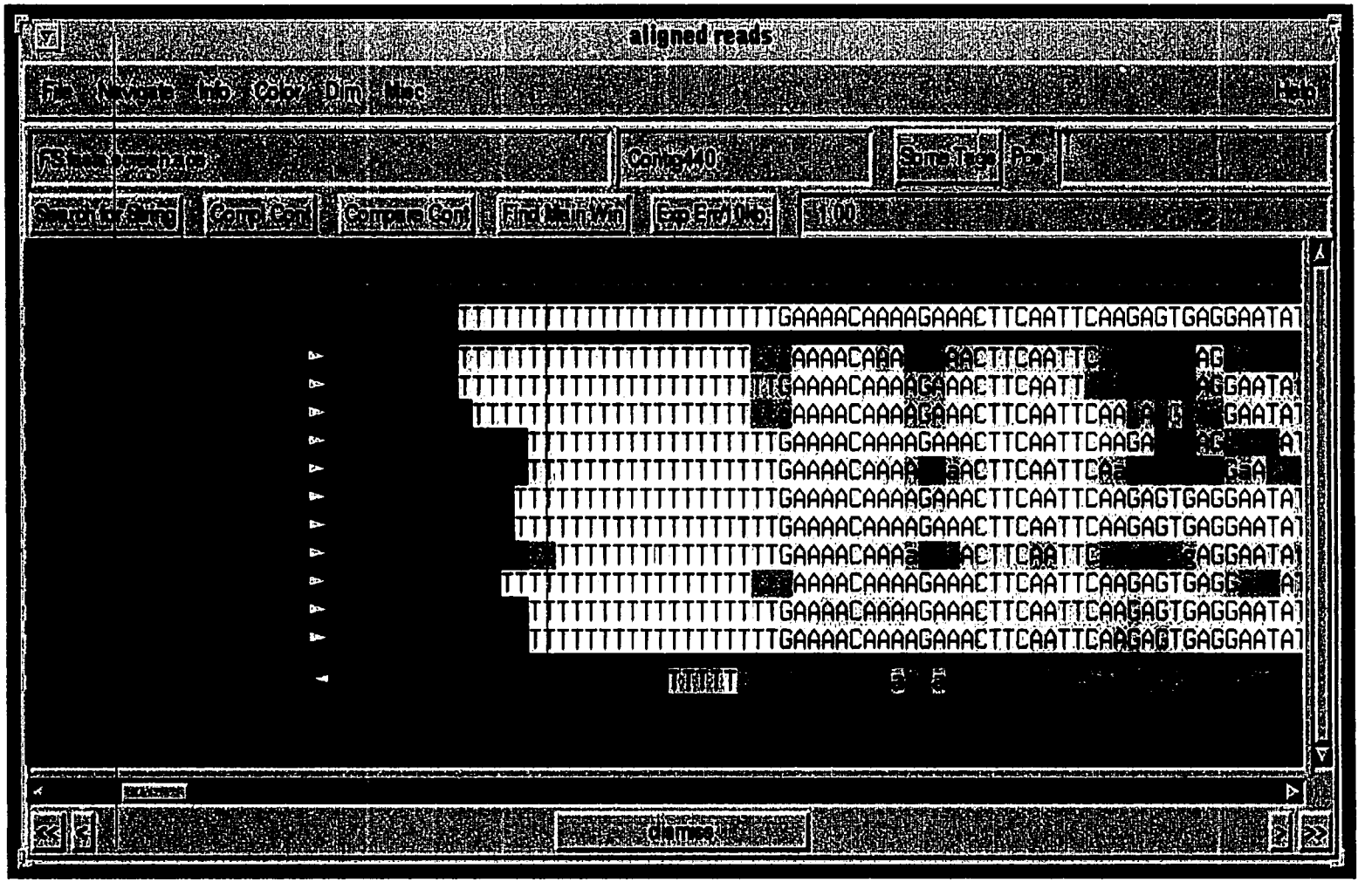


Fig. 2.2 View quality of each read and consensus sequence with consed

consed.colorMeansQuality10_14: grey54

consed.colorMeansQuality15_19: grey60

consed.colorMeansQuality20_24: grey66

consed.colorMeansQuality25_29: grey77

consed.colorMeansQuality30_34: grey86

consed.colorMeansQuality35_39: grey91 (close to white)

consed.colorMeansQuality40_97: white

consed.colorMeansQuality98: grey39 (low quality edit)

consed.colorMeansQuality99: white (high quality edit)

A quality value of 10, 20, 30 corresponds to an error probability of 1/10, 1/100, 1/1000 respectively. Therefore, the whiter the background around the base, the higher its quality value is.

2.4.2.3 Viewing assembly information

If "color means match" is selected, the bases or background have different colors. The colors have the following meaning:

Blue: agree with consensus

Orange: disagree with consensus

Yellow: read used to form the consensus

Grey: trimmed off

2.4.2.4 Navigating to the weak region

By using the navigation bar, the user scans the whole contig to look for the regions requiring inspection because of potential misassemblies, low quality data or other

problems. Consed also provides a navigation window listing the features including low-quality regions requiring additional data, high-quality regions disagreed with consensus, misassembly errors and bases that have been edited. Each item listed in the navigation window is visited by doubly clicking the item.

2.4.2.5 Viewing and comparing traces

Clicking on a base with the middle mouse button brings up the trace of the read in a trace window. If a second trace window for a different base is brought up, it will align with the first one. The traces of the different reads will scroll together. A vertical line in each trace indicates the position of the peak that corresponds to the cursor location in the aligned trace windows. Comparison of the traces at the same positions but in different reads helps make a decision about the correctness of the contig sequence at that position.

2.4.2.6 Aligning and comparing two region

By selecting two regions in the same or different contigs, traces for the reads in those regions then can be displayed to investigate possible joins which were not made by the assembly program Phrap.

2.4.2.7 Picking primers

Primers can be selected automatically using Consed for closing gaps for human genome projects as well as for cDNA sequences. The main criteria for picking primers are as follows:

1) Primer melting temperature should be in an acceptable range. Although the default is 50°C-55°C, it is acceptable to change to 50°C-65°C when needed.

2) Every primer base should have a corresponding consensus quality value at or above the threshold (default 30).

3) The primer should have low propensity to anneal with other locations on the sequencing template, or the clone vector (for genome sequencing, also subclone vector), or itself. A primer is rejected if some site other than the correct one has a score exceeding the threshold (default 17) or if it anneals to itself with a score exceeding the threshold (default 3). The scores are calculated as follows: +2 is awarded for each A/T pair, +3 for a G/C pair and –6 for a mismatch.

4) The desired region for picking primers can be specified (default 450 base).

In the EST project, some contigs contain only one or two sequences. The various primer picking programs, consed and PrimOU, cannot pick desired primers very accurately because there are too few bases in any appropriate site have quality values greater than 30. In this situation, either more sequences are needed to raise the consensus quality, or a special procedure for picking EST primers is used (see section 2.6).

2.4.2.8 Search for homologous sequence regions

Clicking the "Search for String" button on the main window will open a window in which a query sequence can be typed. Alternatively, the query region can be highlighted and after clicking the middle mouse button, the query sequence is written in the window for searching. If the same or similar sequences are found, the region is reported in another popup window. This feature can be used to investigate possible joins, which were not made by the assembly program.

2.4.3 BLAST

The BLAST family of programs actually is five separate programs for rapid sequence database searching, and include BlastN, BlastX, BlastP, TblastN, TblastX. These programs use the statistical methods described by Karlin and Altschul (1990,

1993). BlastN compares a nucleotide query sequence against a nucleotide sequence database. TblastN compares a protein query sequence against a nucleotide sequence database dynamically translated in all six reading frames and BlastX compares the six-frame translations of a nucleotide query sequence against a protein sequence database. TblastX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database and BlastP compares an amino acid query sequence against a protein sequence database. BlastX was used in the *F. sporotrichioides* EST project to find homology between the ESTs and any homologous sequences in the nonredundant protein database in GenBank. BlastN was used to screen out vector sequences, *E. coli* genome DNA, mitochondrial DNA and ribosomal gene sequences in the Clip and Clean procedure discussed in section 2.5.1.

2.4.3.1 The parameter settings

In Blast searches, each High-scoring Segment Pair (**HSP**) consists of a segment from the query sequence and another segment from a database sequence of equal lengths, their alignment is locally maximal, and their alignment score meets or exceeds a threshold or cutoff score.

Blast examines sequences for similarity in several steps. It first looks for similar segments (HSPs) between the query sequence and a database sequence. It then evaluates the statistical significance of any matches that were found and reports those matches that satisfy a user-selectable threshold of significance.

The parameter **E** is interpreted as the upper bound of the expected frequency of the chance occurrence of an HSP (or set of HSPs) when searching the entire database. E also may be thought of as the number of matches one expects to observe by chance

during the database search when other parameters are sensitive enough and the query and database sequences follow the random sequence model of Karlin and Astschul (1990). Any database sequence whose match satisfies E will be reported in the program output. The default value for E is 10, and E may be varied over a range of 0< E <= 1000. In the *F. sporotrichioides* project, the E value was set at 0.15 when performed BlastX to reduce the number of observed homologies, thereby increasing the stringency required for a significant match. However, only the homologies that had E values less than 0.0001 was consider significant. Similarly, the E value also was set at 0.15 for PowerBlast for genome shotgun projects.

The parameter **S** represents the score at which a single HSP would by itself satisfy the significance threshold E. The higher the S value, the lower the probability of chance occur, and less error possibility. The value of E is used to calculate S.

The parameter **E2** is the expected number of HSPs that will be found when comparing two sequences that each has the same length, either 300 amino acids or 1000 nucleotides. The default value for E2 is about 0.15. The cutoff score, which defines HSPs, is parameterized as **S2** and the default value for S2 is calculated from E2.

The parameter **W** is defined as the starting word length. The task of finding HSPs begins with identifying short words of length W in the query sequence that either match or satisfy some threshold score when aligned with a word of the same length in a database sequence. These word hits served as seeds for finding longer HSPs in the neighborhood. The default value for W is 3 amino acids for BlastP, BlastX, TblastN and TblastX and 11 nucleotides for BlastN.

The parameter **M** is the score for match and parameter **N** is the penalty for mismatch.

Blosum62 (Henikoff, 1992) is the default amino acid blocks substitution matrix for BlastX which is based on observed substitutions between segments that are less than 62% identical. Seg (Wootton and Federhen, 1993) is the default filter for BlastX that masked the low complexity sequences in the query sequence.

2.4.3.2 Query sequence format

The query sequence were in fasta format. For example:

```
>fsEST
aaaaccgetttaaaccctgtgtatcttcccaaagggttggccccccccaaa
accctgtgtatcttccatcttcccaaagggttggccccccaaaccgettta
ggtaatacgtttt
```

The first line of a fasta format file is a descriptor line which begins with a greater than sign followed by a short description or sequence name. The following lines contain the sequence in 5' to 3' orientation and may be either lower or upper case.

2.4.3.3 Interpreting Sequence Identifiers

The syntax of sequence identifiers used by BLAST depends on the source database where the sequence was obtained (Altschul et al., 1990). Table 2.6 summarized several databases and their identifier syntax.

Table 2.6 The syntax of sequence identifiers used by the NCBI BLAST

Database Name	Identifier Syntax
GenInfo Integrated Database	gi | number
GenBank	gb | accession | locus
"GenPept" derivative of GenBank	gp | accession | locus_cds#
EMBL Data Library	emb | accession | name
DDBJ, DNA Database of Japan	dbj | accession | name

NBRF PIR	pir | accession | entry
SWISS-PROT	sp | accession | name
Brookhaven Protein Data Bank	pdb | name | chain
Kabat's Sequences of Immuno..	qnl | kabat | name
TFD	gnl | tfd | name
Eukaryotic Promotor Database	gnl | epd | name

The following is an example of the BlastX (Matrix is blosum62, filter is Seq) result:

Contig1056 2015 1.6e-207 67 1188 gb|AAD13657.1| (M27246) trichodiene synthase [Fusarium sporotrichioides]

In this example, "Contig1056" is the ID of the contig used to do the BlastX homology search. "2015" is S score (section 2.4.3.1) and 1.6e-207 is the E value (section 2.4.3.1). High S score or low E value indicates less probability of error. "67 1188" is the sequence homology endpoints. The homologous sequence identifier is "gb|AAD13657.1| (M27246)", where the "gb" indicates that the identifier refers to a GenBank sequence, "AAD13657.1" is its GenBank Accession, and "M27246" is the GenBank Locus. At the end, "trichodiene synthase" is the description of homology and "[Fusarium sporotrichioides]" is the organism.

2.5 EST data analysis

2.5.1 Clip and Clean

After the gel image analysis and DNA sequence extraction, an experiment file was created automatically for each EST (.exp file). The information included in the experiment file is the EST sequence, base quality values, sequencing primer type, and vector type. The Clip and Clean program removes the low quality sequences and non-EST related sequence. Clip and Clean actually is a series of linked scripts which

originally were obtained from LaDeana Hiller at Washington University at St. Louis and modified by Hongshing Lai for use at ACGT. Each script is linked with the next and all the scripts are piped in the order listed in Table 2.7 (Kupfer, 1999). The following functions are included in Clip and Clean scripts. (1) Assess sequence quality and remove low quality data. (2) Trim flanking vector sequences. (3) Remove short insert sequences. (4) Remove wrong end sequences. (5) Remove ribosome RNA sequences. (6) Remove mitochondrial sequences. (7) Remove *E.coli* genome sequences. The parameters used in the scripts were summarized as follows.

1) Assess sequence quality and remove the low quality sequences

After the Phred quality value of each base is calculated, sequences that had overall N ratio of 1:5 were removed from the data set. For sequences which had low quality regions between high quality sequences, the regions of high quality were marked and the regions of low quality, i.e. quality was lower than 15, also were recorded.

2) Remove the vector sequences

Vector sequences which were contiguous with the EST sequences were located using the programs Vep-vector End Point (VEP) (Dear and Staden, 1991) and BLASTN (S = 133, S2 = 133, M = 5, N = -11, W = 8). Information about adapter sequences and the poly(dT) sequence also were used to decide the vector-EST junction. The entire sequence was removed from the database if the poly(dT) was not found on a 3' EST or the poly(dT) was found on a 5'EST, or an adapter sequence was seen on a 3' EST, or the whole sequence was vector sequence. The vector sequence was removed from both ends of the EST. If the sequence length was less than 100 bps after vector removal, these short insert sequences also were removed from the database.

Table 2.7. Clip and Clean EST processing scripts (Adapted from Kupfer, 1999)

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). Make 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 (Dear and Staden, 1991) to find the sequence 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(dT) is not found on 3' end and remove the read from the database
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	if the wrong adapter sequence present, e.g. if a 5' adapter sequence was seen in 3' EST, this 3' EST was labeled failed and remove from the database
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	Check the sequence quality again. If there are high quality sequence to the right of the first base with quality 15, add them to the high quality sequence.
13. check_processor	checks for sequence length < 100 bases, the short sequences are marked and removed from the database
14.screen.p	BLASTN against the *E.coli* genome database (S = 170, S2 = 150, M = 5, N = -11) and both mitochondrial and ribosomal sequences (S = 133, S2 = 133, M = 5, N = -11, W = 8). These non EST related sequences are removed from the database
15. expESTBlastx	BLASTX against non-redundant protein database. Matrix is blosum62, filter is seq filter.
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. exp2best	create a dbEST submission file and place in directory for GenBank submission and placement on website.

Table 2.8 Mitochondiral and ribosomal sequences used in Clip and Clean processes

gi|4160479|gb|AF111055.1|AF111055[4160479]
Fusarium sporotrichioides strain BBA69073 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence

gi|4160477|gb|AF111053.1|AF111053[4160477]
Fusarium sporotrichioides strain RUS446 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence

gi|2677811|gb|U85524.1|FSU85524[2677811]
Fusarium sporotrichioides 28S ribosomal RNA gene, partial sequence

gi|2674289|gb|U85541.1|FSU85541[2674289]
Fusarium sporotrichioides internal transcribed spacer ITS1, 5.8S ribosomal RNA gene, and internal transcribed spacer ITS2, complete sequence

gi|2957211|gb|AF006348.1|AF006348[2957211]
Fusarium sporotrichioides NRRL 25479 internal transcribed spacer 1, 5.8S ribosomal RNA gene; and internal transcribed spacer 2, complete sequence

gi|2957191|gb|AF006328.1|AF006328[2957191]
Fusarium sporotrichioides 28S ribosomal RNA gene, partial sequence

gi|2674306|gb|U85558.1|FSU85558[2674306]
Fusarium sporotrichioides mitochondrial small subunit ribosomal RNA, mitochondrial gene, partial sequence

gi|2104831|emb|X80806.1|FS28SRRA[2104831]
Fusarium sporotrichioides 28S rRNA gene

gi|1054925|gb|U38553.1|FSU38553[1054925]
Fusarium sporotrichioides 5.8S ribosomal RNA gene, complete sequence and internal transcribed spacers 1 and 2

gi|174601|gb|M85202.1|FSORRNAA[174601]
Fusarium sporotrichioides large subunit ribosomal RNA, partial

3) Remove the contaminating sequences

ESTs similar to ribosomal sequences or mitochondria sequences were identified using BLASTN (S = 133, S2 = 133, M = 5, N = -11, W = 8). The sequences used for screening mitochondrial sequences and ribosomal sequences are list in Table 2.8. ESTs similar to *E.coli* genome sequence were identified using BLASTN (S = 170, S2 = 150, M = 5, N = -11). Any identified ribosomal, mitochondrial or bacterial host genomic contaminating sequences then were removed from the database.

2.5.2 Cumulative 3' assembly of the EST database

During the construction of the cDNA library, poly(dT) primers of the first cDNA strand are anchored on the corresponding positions of the poly(A) tails, of the mRNAs. Therefore, 3' ESTs starting from the regions upstream of poly(A) tails contain polyadenylation signals and 3' UTR of mRNA. One gene may have several identical and overlapping 3' EST with differing lengths of poly(A) sequences, as they might be derived from different copies of mRNA or different copies of cDNA. Since different genes would have different 3' EST sequences, 3' ESTs with homologous 3'end sequences can be aligned into clusters after assembly with the sequence alignment tools such as Phrap (Fig. 1.6). Each cluster then would represent a set of clones transcribed and reverses transcribed from the same gene. The number of ESTs in each cluster represents an estimate of the abundance of a gene's transcript present in the cDNA library. Therefore, the abundance of 3' ESTs in each EST cluster for a specific gene is proportional to this gene's relative transcription level and reveals the gene expression levels for an organism of interest or for its specific development stage.

2.5.3 3' and 5' assembly of the EST database

After removing the failed sequences to a separate directory, the high quality sequences were assembled by Phrap. The parameters used in the Phrap assembly were minmatch 14 and minscore 80. ESTs that had homologous sequence will align into contigs. These contigs also can be called clusters, although the former emphasizes the consensus sequence derived from the aligned ESTs and the latter emphasizes the individual ESTs that are aligned. The stringency used in this project was high enough to allow only those 3' or 5' ESTs that were from the same gene to assemble into a contig. Only occasion 5' EST with same functional motif from the different genes were miss-aligned. In many cases, 3'ESTs and 5'ESTs were assembled into the same contig (Fig. 1.6), even though they did not have the 5' UTR sequence. However, once a Phrap contig was formed, it was given a unique "contigID" and was stored in a "contig_dir" (directory for contigs). Those ESTs that would not assemble with any other ESTs were called singlets and were stored in a "singlet_dir" (directory for singlets). All singlets and contigs were submitted for a BlastX search against GenBank non-redundant protein database. The results of BlastX for each singlet or contig were stored in the blastx_dir. Each singlet or contig had three files in the blastx_dir, one for its sequence (e.g. Contig365), one for complete BlastX output (e.g. Contig365.table) and one for BlastX header lines (e.g. Contig365.table_trim). This data was used for the analysis of biological function assignments for each EST in the library in the next step.

2.5.4 Biological function assignments

If possible, each singlet/contig was assigned a biological function at the end of the data analysis stage. Several procedures were performed in order to fulfill this task (Fig. 2.3).

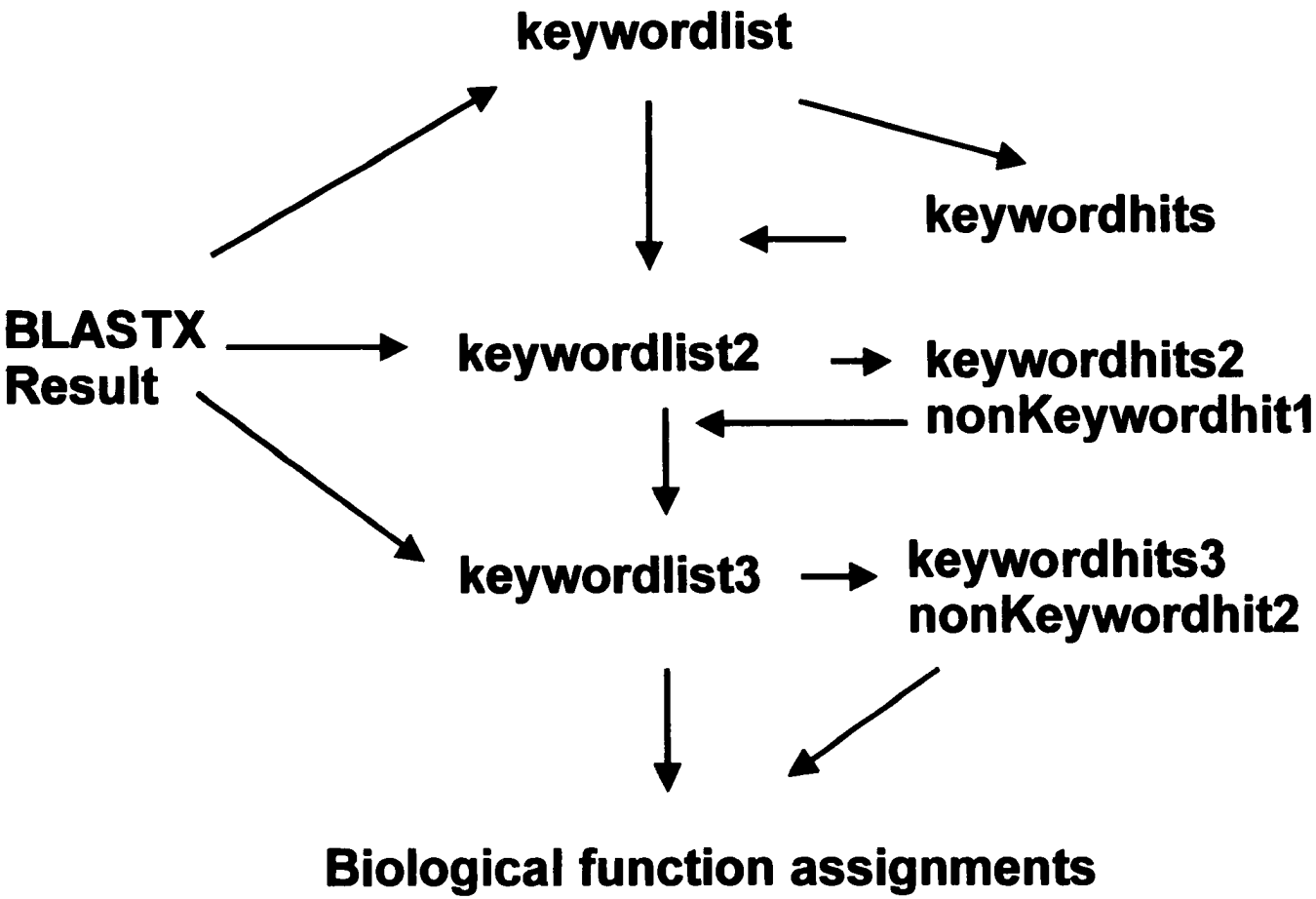


Fig. 2.3 Steps in biological function assigments

2.5.4.1 BlastX search output and primary keyword list

To assign biological functions to the ESTs from the *F. sporotrichioides* cDNA library, a BlastX search was performed for each singlet and contig in the database. The homology was determined to be significant if its S score (section 2.4.3.1) was higher than 99 or has an E value (section 2.4.3.1) was less than 1×10^{-4}. In many cases, a singlet or contig had Blast homologies to more than one database entries that were on or higher than this significance threshhold. The multiple Blast results were listed in score order with the highest homology list first (Table 2.9).

A preliminary keyword list was developed based on a list created by Doris Kupfer for *Aspergillus nidulans* (Kupfer, 1999), which was derived from the gene functions for *E. coli* (Reily, 1997; Selkov, 1998). The preliminary keyword list later was revised several times, and the most recent version is presented in Appendix I.

2.5.4.2 Edit the keyword list

A series of programs had been written by James White in ACGT at OU to analyze the BlastX homology data in the assembled EST database. The first of them, blast_best_keyword, was executed to read the preliminary keyword list and the BlastX results to search for matches of the protein or enzyme names in the BlastX results and those in the keyword list. If a match was found, the keyword was enclosed in angle brackets < > before that BlastX result. This program selected and kept in the final output only the top hits and matches. The output file was named "keywordhits" which listed the top blast hits, sorted by both singlet/contig and BlastX score order (Table 2.10). If a protein or enzyme hit carried a higher score in the BlastX results but had no match in the keyword list, then a new keyword would be added manually to the keyword list as well as

in the brackets before "keywordhits" line. Those addition steps were repeated until no new keywords were found. Many resources were referred to when a protein or enzyme was assigned to a pathway or a classification including biochemistry texts, Metabolic Pathway Database (MPD) (Selkov Jr. 1998), GenBank (Benson, 1996) and KEGG's web site (http://star.scl.kyoto-u.ac.jp/kegg/kegg2.html).

If no new keyword was found, a program blast_best_nonkeyword was executed to look for singlets and contigs whose BlastX results had no matches in the keyword list. The output file was named "nonkeywordhits" (Table 2.11). Then new keywords were added to the keyword list as well as in the brackets before the BlastX hit as described before (Table 2.12).

After editing the keyword list, blast_best_keyword and blast_best_nonkeyword programs again were re-run. The above procedures were repeated until there were no additional new keywords to be added to the keyword list.

2.5.4.3 Print the final form

The database result with the highest homology was selected manually and duplicates were removed for each singlet and contig in keywordhits file and nonkeywordhits file. Both files were combined to give the resulting homology with comments in keyword order using a third program, blast_print_keywords. Each keyword was followed by a list of singlets and contigs that showed homology to this database entry in their BlastX results (Appendix II).

2.5.5 tBlastX against dbEST

Over 50% of the assembled members in *F. sporotrichioides* had no significant homologues in the GenBank non-redundant protein database (nr) and may represent

Table 2.9 A portion of the BlastX results for Contig1002 showing that many BlastX hits listed in score order

```
  917  3.5e-91   Contig1002    120  1034 sp|P34054|INA1_TRIHA AMINO-ACID PERMEASE INDA1 >pir||S33212 INDA1
protein -fungus (Trichoderma harzianum) >emb|
  847  9.3e-84   Contig1002    177  1034 gb|AAB61277.1|       (AF001032) amino acid permease [Neurospora
crassa]
  606  3.2e-58   Contig1002    102  1052 pir||T39122          amino-acid permease - fission yeast
(Schizosaccharomyces pombe)(fragment) >emb|CAB60021.1|
  534  1.4e-50   Contig1002    240  1055 gi|6322892           ref|NP_012965.1|GAP1| general amino acid
permease; Gap1p>sp|P19145|GAP1_YEAST GENERAL AMIN
  527  7.5e-50   Contig1002    240  1055 emb|CAA36858.1|      (X52633) GAP1 protein (AA 1-601)
[Saccharomyces cerevisiae]
  517  8.6e-49   Contig1002    231   959 pir||T39829          amino-acid permease - fission yeast
(Schizosaccharomyces pombe)>emb|CAA19126.1| (AL023594)
  501  4.3e-47   Contig1002    156  1025 sp|P40901|ISP5_SCHPO SEXUAL DIFFERENTIATION PROCESS PUTATIVE
AMINO-ACIDPERMEASE ISP5 >pir||S45492 isp5 protein
  498  8.9e-47   Contig1002    198  1025 emb|CAB63545.1|      (AL133521) sexual differentiation process
putative amino-acidpermease isp5 [Schizosaccharo
  490  6.3e-46   Contig1002    180  1055 gi|6321629           ref|NP_011707.1|HIP1| histidine permease;
Hip1p>sp|P06775|HIP1_YEAST HISTIDINE PERMEASE >p
  489  8.0e-46   Contig1002    192  1049 gi|6324553           ref|NP_014622.1|TAT2| Tryptophan permease,
high affinity; Tat2p>sp|P38967|TAT2_YEAST TRYPT
  483  3.5e-45   Contig1002    192  1049 dbj|BAA03811.1|      (D16304) LTG3 protein [Saccharomyces
cerevisiae]>prf||2105281A LTG3 gene [Saccharomyces ce
  483  3.5e-45   Contig1002    180  1055 prf||1112180A        permease,His [Saccharomyces cerevisiae]
  474  3.1e-44   Contig1002    204   959 emb|CAB86887.1|      (AL163525) amino acid permease
[Schizosaccharomyces pombe]
  474  3.1e-44   Contig1002    204   959 dbj|BAA13506.1|      (D87954) amino acid permease
[Schizosaccharomyces pombe]
  465  2.8e-43   Contig1002    261  1043 gi|6322967           ref|NP_013039.1|MMP1| High affinity S-
methylmethionine permease;Mmp1p >pir||S50959 probabl
  465  2.8e-43   Contig1002    252  1043 gi|6324981           ref|NP_015049.1|SAM3| High affinity S-
adenosylMethionine Permease;Sam3p >pir||S65307 proba
  442  1.4e-40   Contig1002     96  1055 gi|6320717           ref|NP_010796.1|GNP1| high-affinity glutamine
permease; Gnp1p>sp|P48813|GNP1_YEAST HIGH-AF
  441  1.9e-40   Contig1002     96  1055 gb|AAB48002.1|       (U21643) high-affinity glutamine permease
[Saccharomycescerevisiae]
  423  7.9e-39   Contig1002    240  1028 sp|P25737|LYSP_ECOLI LYSINE-SPECIFIC PERMEASE >gb|AAA60532.1|
(U00007)lysine-specific permease [Escherichia col
  423  7.9e-39   Contig1002    240  1028 pir||C64984          lysine-specific permease - Escherichia coli
>gb|AAA17053.1|(M89774) lysine specific permea
```

Table 2.10 Result of Contig1002 in file "keywordhits" showing that one contig has several keyword hits

<AMINO-ACID PERMEASE>
917 3.5e-91 Contig1002 120 1034 sp|P34054|INA1_TRIHA AMINO-ACID PERMEASE INDA1 >pir||S33212 INDA1 protein -fungus (Trichoderma harzianum) >emb|CAA80308.1| (Z22594) INDA1[Trichoderma harzianum]

<AMINO-ACID PERMEASE>
847 9.3e-84 Contig1002 177 1034 gb|AAB61277.1| (AF001032) amino acid permease [Neurospora crassa]

<transport protein>
534 1.4e-50 Contig1002 240 1055 gi|6322892 ref|NP_012965.1|GAP1| general amino acid permease; Gap1p>sp|P19145|GAP1_YEAST GENERAL AMINO-ACID PERMEASE GAP1 >pir||S38111amino acid transport protein GAP1 - yeast (Saccharomycescerevisiae) >emb|CAA82113.1| (Z28264) ORF YKR039w [Saccharomycescerevisiae]

<unknown function>
534 1.4e-50 Contig1002 240 1055 gi|6322892 ref|NP_012965.1|GAP1| general amino acid permease; Gap1p>sp|P19145|GAP1_YEAST GENERAL AMINO-ACID PERMEASE GAP1 >pir||S38111amino acid transport protein GAP1 - yeast (Saccharomycescerevisiae) >emb|CAA82113.1| (Z28264) ORF YKR039w [Saccharomycescerevisiae]

<tryptophan permease>
489 8e-46 Contig1002 192 1049 gi|6324553 ref|NP_014622.1|TAT2| Tryptophan permease, high affinity; Tat2p>sp|P38967|TAT2_YEAST TRYPTOPHAN PERMEASE (TRYPTOPHAN AMINO ACIDTRANSPORTER) >pir||S46273 tryptophan transport protein - yeast(Saccharomyces cerevisiae) >gb|AAA60324.1| (L33461) tryptophanpermease [Saccharomyces cerevisiae] >emb|CAA55777.1| (X79150)tryptophan amino acid permease [Saccharomyces cerevisiae]>emb|CAA99020.1|

Table 2.11 A portion of the file "nonkeywordhits"

```
<>
367  8.2e-32 b4a06fs.f1                7 387   emb|CAA06786.1|    (AJ005963) 100 kDa protein [Ajellomyces
capsulatus]

<>
125  2.2e-05 b4a06fs.f1    88 243   sp|P25823|TUD_DROME MATERNAL TUDOR PROTEIN >pir||A41519 posterior-
groupprotein tudor - fruit fly (Drosophila melanogaster)>emb|CAA44286.1| (X62420) tudor protein [Drosophila
melanogaster]

<>
125  2.2e-05 b4a06fs.f1    88 243   gb|AAF46693.1|     (AE003453) tud gene product [Drosophila
melanogaster]

<>
105  0.00072 b4a06fs.f1    40 417   gb|AAD37448.1|AF1538 (AF153880) A-kinase-anchor-protein 84 [Fugu
rubripes]
```

Table 2.12 A portion of the file "nonkeywordhits" after adding the new key words in the angle bracket

```
<unknown>
367  8.2e-32 b4a06fs.f1                  7 387    emb|CAA06786.1|     (AJ005963) 100 kDa protein [Ajellomyces
capsulatus]

<TUD_DROME MATERNAL TUDOR PROTEIN>
125  2.2e-05 b4a06fs.f1     88 243    sp|P25823|TUD_DROME MATERNAL TUDOR PROTEIN >pir||A41519 posterior-
groupprotein tudor - fruit fly (Drosophila melanogaster)>emb|CAA44286.1| (X62420) tudor protein [Drosophila
melanogaster]

<tud gene product>
125  2.2e-05 b4a06fs.f1     88 243    gb|AAF46693.1|      (AE003453) tud gene product [Drosophila
melanogaster]

<A-kinase-anchor-protein 84>
105  0.00072 b4a06fs.f1     40 417    gb|AAD37448.1|AF1538 (AF153880) A-kinase-anchor-protein 84 [Fugu
rubripes]
```

undiscovered genes. Since these genes or their homologues also may exist in other species, ESTs representing them may present in dbEST. Therefore, singlets and contigs that had no significant BlastX homology to the nr database were further processed by a tBlastX search against dbEST to search for homologues. The *F. sporotrichioides* ESTs with tBlastX homologues with ESTs from *Aspergillus nidulans* (Kupfer, 1999) and *Neurospora crassa* (Zhu, unpublished) but not with non-fungal ESTs, most likely represent fungi specific genes.

The group of *F. sporotrichioides* ESTs that had significant homologs in GenBank nr similarly were processed through the tBlastX search against dbEST. Homologues among the four EST database, *F. sporotrichioides* database, two *N. crassa* databases and *A. nidulus* database, were subtracted from the tBlastX results. Although *F. sporotrichioides, A. nidulus* and *N. crassa*, represent three different families, Hypocreaceae, Trichocomaceae and Sordariaceae respectively, they belong to the same phylum, Ascomycota. The comparison of the similarities and diversities of the known genes represented in all three fungal species may help us to further define those genes which result in the unique fungal biological function similarities as well as the differences among these three families.

2.6 Full length cDNA sequence and submission

When a full-length cDNA sequence was required to be obtained for further studies, the primer walking method (Section 3.6) was used to obtain this sequence by extending the EST end sequences from the original cDNA clone.

2.6.1 Picking primers for gap closures to get a full-length cDNA sequence

Consed was used for picking the primers needed for this primer walking closure strategy (section 2.4.2.7). Although ace file is required as the input file for Consed (section 2.4.2), since many EST sequences were not assembled into the ace file because they did not overlap with others to form a contig, an alternative approach was chosen. Here, since Consed 7.0 provides an alternative procedure to produce the ace file for individual singlet or non-assembled EST, it can be used to view those sequences and pick primers for closing gaps between non-contig bound EST sequences. This procedure is outline here.

1. Make the edit_dir, phd_dir and chromat_dir;
2. Put the chromatogram into chromat_dir.
3. Go to edit_dir, run phredPhred to generate the phd file which goes into phd_dir;
4. Go to edit_dir and run phd2Ace.perl (name of the phd file)
5. Start consed and input that ace file produced by the above step.

Most primers could be picked out by using the above method. However, if the base quality values were low than 30 for the bases located at the desired primer binding site, additional sequences were obtained and added to the database to raise the consensus quality and subsequently raise the accuracy of the primer sequence for more efficiently primer binding.

2.6.2 Analysis and submission of cDNA sequences

Each final complete, unambiguous cDNA sequence was analyzed by performing BlastX against GenBank nr and other methods described in section 1.6. and submitted to GenBank by Sequin (section 1.6.7).

Chapter III

Material and Methods (Part 2)

Sequence and analysis of human genome projects

3.1 Overview of the Sequencing Strategy

The sequencing approach used in human genome projects in this dissertation was a shotgun-based sequencing strategy (section 1.5.4). It began with the large scale isolation of cosmid, PAC or BAC DNA, followed by extremely random nebulizer-based physical shearing. The resulting fragments were size-selected with low-melting agarose gel electrophoresis and 1.5-4 kb fragments were end-repaired, inserted into pUC18 vectors, and transformed into *E. coli* XL1-Blue cells. After the sub-cloned DNA molecule was isolated, cycle sequencing was performed with forward and reverse universal primers and fluorescent-labeled Taq ddNTP terminators. End sequences of the shotgun DNA sub-clones were collected on either an ABI377 or ABI 3700 sequencer and assembled with the Phred/Phrap programs. After all gaps were closed and the error rate was less than 1 error per 10,000 bases, the final sequence then was analyzed by Xgrail and GenScan to predict exons and other genetic features and by Powerblast and BlastX to search for regions with homology in the GenBank database. Finally, the resulting sequence and regions of structural and biological interest were annotated and the data was submitted to GenBank.

Protocols used in this dissertation research were from the "Protocols for Recombinant DNA Isolation, Cloning, and Sequencing" (Roe, 1997). They are also

available from the Advanced Center for Genome Technology web-site at http://www.genome.ou.edu.

3.2 Large scale DNA isolation

Large scale DNA isolation is aimed at obtaining sufficient quantities (micrograms) of purified BAC, PAC, or Cosmid DNA. The method developed and used for the large-scale DNA isolation in our laboratory is a modified alkaline lysis method (Birnboim and Doly, 1979; Roe, 1995). We have used two methods for large-scale isolation. They differ slightly in the steps included in the purification. One protocol uses a diatomaceous earth mix to purify the DNA. The detailed protocols for this are available on our Web Site (http://www.genome.ou.edu/proto.html). The other is a double acetate protocol, which was developed at the Washington University Genome Sequencing Center in St. Louis, MO. This protocol also is available on our web site at URL:

http://www.genome.ou.edu/DblAcetateProcV3.html

The second method is the method presently used in our laboratory. It also was the method used in this dissertation research.

3.2.1 Culture the cells

The first step was to pick a colony or a smear of *E. coli* colonies with a sterile toothpick. Then after placing the toothpick into 3 ml of LB medium (section 2.1.5) supplemented with 100 μg/ml ampicillin in 12x75 mm Falcon tube, the inoculated culture was incubated for 8-10 hours at 37 °C with shaking at 250 rpm. The 3 ml culture was transferred to 50 ml of the same medium in a 250 ml Erlenmeyer flask and incubated a further 8-10 hours under the same conditions. The 50 ml culture then was divided and

transferred into two flasks, each containing 1 liter of LB medium supplemented with ampicillin. After further incubation for another 8-10 hours under the same conditions, cells were harvested by centrifugation at 7000 rpm for 20 minutes in 500 ml centrifuge bottles in a RC5-B centrifuge in GS3 rotor and the cell pellets were frozen and stored at –70°C. Generally, cells in one-liter medium were collected by centrifugation in two 500 ml bottles.

3.2.2 Isolation of target DNA

The next step was to isolate target DNA from the host cells. The frozen cells were thawed and resuspended in 35 ml of GET solution (50 mM glucose, 25 mM Tris-HCl, pH 8.0, 10 mM EDTA) in each bottle. To avoid shearing the *E. coli* genomic DNA, this step should be done very gently. Once the pellets were completely resuspended, 70 mg lysozyme was added to a final concentration of 2 mg/ml and the cells were incubated at room temperature for 10 minutes. Then 70 ml of alkaline lysis buffer (200 mM NaOH, 1% sodium dodecylsulfate (SDS)) was added. The solution was gently mixed and incubated for 5 minutes in an ice-water bath. Then 52.5 ml of 3M NaOAc or KOAc was added, gently mixed and incubated in an ice-water bath for 30 minutes. This step precipitated the cellular proteins, membranes and genomic DNA. The lysate was cleared by filtration through a double-layer of cheesecloth into clean 500 ml centrifuge bottles. The filtrate was centrifuged at 10,000 rpm for 30 minutes at 4 °C using the RC5-B centrifuge to clear any small particulates that may not have been cleared by the cheesecloth. This step was repeated until all insoluble visible were removed.

The supernatant was transferred into a clean 500 ml bottle. An equal volume of isopropanol was added, mixed by inverting the bottle, and the solution was incubated at

room temperature for 5 minutes followed by centrifugation at 9000 rpm for 30 minutes in the RC5-B centrifuge. The supernatant was decanted and the DNA pellet drained.

3.2.3 Purification of DNA

The DNA pellet was dissolved in 18 ml of 10:1 TE (10 mM Tris-HCl, pH 7.6, 1 mM of EDTA, pH 8.0) and divided into two 50 ml Corning centrifuge tubes. 4.5 ml of 7.5 M KOAc was added into each tube. After mixing, the tubes were kept at –70°C for 30 minutes. After the solution was thawed, it was centrifuged in Beckman GS-6R centrifuge at 2000 rpm for 10 minutes. The supernatant of each tube was transferred into a clean 50 ml Corning centrifuge tube and 30 ml of 100% cold ethanol was added into each tube. The solution was mixed by inverting and the tubes were incubated in an ice-water bath for 15 minutes. The solution was centrifuged at 3000 rpm for 25 minutes in Beckman GS-6R centrifuge. The 30 ml of 70% ethanol was added to wash the pellets and the pellets were dried overnight in vaccum oven.

3.2.4 Degradation of RNA

The dried pellet in each tube was dissolved in 1 ml of ddH_2O, transferred to a microcentrifuge tube and DNase-free RNase A was added to a final concentration of 100 μg/ml. The tube was incubated in a 37°C water bath for 1 hour and then, phenol extracted to remove the RNase A and any remaining protein material from the DNA sample.

3.2.5 Phenol extraction

The DNA solution in each microcentrifuge tube was divided into two tubes and an equal volume of TE saturated phenol (prepared by adding equal volume of 10 mM Tris-HCl, pH 7.5-8.0, 1 mM Na_2EDTA to phenol, allowed phase to separate) was added to the tube. The mixture was vigorously vortexed for half a minute, and then centrifuged

for 5 minutes to separate the phases. The upper aquous layer was carefully removed to a new tube. This step should be done very carefully to avoid proteins at the interface of the aqueous and phenol. Then an equal volume of 1:1 TE-saturated phenol:chloroform was added to the tube, and the tube was vortexed, centrifuged for 5 minutes and the upper aquous layer was removed to a clean tube as above. Then an equal volume of chloroform was added, vortex briefly, and centrifuged for 3 minutes at room temperature. Then 2.5 volumes of ethanol-acetate (95% ethanol and 0.12M sodium acetate) were added to the tube, the tube was incubated in an ice-water bath for 10-15 minutes and centrifuged for 15 minutes at 4°C at 12,000 rpm. After the supernatant was decanted, the tube was drained on a paper towel. 80% ethanol was added to the tube, the tube was incubated at room temperature for 5 minutes before centrifugation again for 5 minutes. The tube with the purified DNA pellet was dried in a Savant Speed-Vac for 5-10 minutes and dissolved in ddH_2O. The DNA isolated from 1-liter medium was dissolved in 2 ml of ddH_2O and the typical yield was 500 μg/liter of culture.

3.3 Shotgun library construction

3.3.1 Physical shearing

The nebulizer-based method for physically shearing DNA results in a highly random population of sheared fragments where the sizes of the resulting fragments are inversely proportion to the gas pressure used. The solution for nebulization contained approximately 50 μg of DNA, 500 μl of sterile 100% glycerol, 200 μl of 10X TM buffer (500 mM Tris-HCl, pH 8.0, 150 mM $MgCl_2$). Sterile ddH_2O was added to a final volume of 2 ml. The solution was mixed and transferred to the cup of a nebulizer. The nebulizer

was connected to a high-pressure nitrogen gas source and the lower part of the nebulizer was incubated in a – 20 °C isopropanol dry ice bath. The DNA was nebulized for 2.5 minutes with nitrogen pressure at between 5-10 psi, depending on the desired size range of DNA fragments. After nebulization, the nebulizer was briefly centrifuged at 1500 rpm to collect the DNA sample to the bottom of the nebulizer. The DNA sample then was precipitated with ethanol-acetate, washed using 80% ethanol, dried in a vacuum and dissolved in a total of 27 μl of 1X TM buffer (50 mM Tris-HCl, pH 8.0, 15 mM $MgCl_2$).

3.3.2 End repair and size fractionation

End repair was performed to convert the single-stranded ends generated by nebulization to double-stranded blunt ends. The end repair solution contained 27 μl of the nebulized DNA, 5 μl of 10x polynucleotide kinase buffer, 5 μl of 10 mM rATP, 7 μl of 0.25 mM dNTPs, 1 μl of T4 polynucleotide kinase (30U/ul), 2 μl of Klenow DNA polymerase (5U/μl) and 3 μl T4 DNA polymerase (3U/μl). The mixture was incubated for 30 minutes at 37 °C and loaded into a 0.7% low melting point agarose gel. *Hind* III-digested λ DNA and *Hae* III digested ϕX174 DNA were used as size markers. Electrophoresis was performed at 120 mA for one hour. Fragments with the size between 1.5-4 kb were excised from the gel into a microcentrifuge tube and heated in a 70 °C water bath for about 10 minutes to melt the gel. Then, phenol extraction was performed as described in section 3.2.5. After the DNA pellet was dried, it was dissolved in 20 μl of 10:0.1 TE buffer (10 mM Tris-HCl, pH 7.6, 0.1 mM EDTA) to give a final concentration of 500-1000 ng/μl.

3.3.3 Ligation

The ligation reaction is used to join the sheared fragments with the vectors to obtain the sub-clones. The ligation reaction solution contained 2 μl (containing 20 ng) of *Sma*I cut pUC18, 1 μl of 10X ligation buffer and 1 μl of T4 DNA ligase. A range of volumes of nebulized DNA (such as 0.5 μl, 1 μl, 2 μl corresponding to 0.25, 0.5, 1 μg of DNA) were added to a set of ligation reaction to test for the optimal concentrations. Finally, ddH_2O was added to the final volume of 10 μl. The ligation solution was incubated at room temperature for at least 3 hours or stored in 4 °C cold room overnight before transformation. The optimized nebulized DNA concentration was used for additional ligation reactions after the initial one. Several parallel ligations also were performed as both positive and negative controls. An *Alu*I-digested cosmid DNA or a known blunt-ended insert replaced the nebulized DNA in the ligation solution to serve as a positive control to test the quality of the end repair reactions. A parallel reaction in the absence of insert DNA also was used as a negative control to test the rate of vector self ligation due to inefficiently phosphatased vector.

3.3.4 Transformation

The recombinant plasmids were transformed into the *E. coli* cells by electroporation.

The *E. coli* cells used for transformation were specifically prepared to increase their ability to take up foreign DNA. To prepare the electro-competent cells, an aliquot of the glycerol stock of XL1 Blue MRF` cells were spread on a tetracycline (20μg/ml) LB plate (10 g/L Bacto-Tryptone, 5 g/L Bacto-yeast extract, 10 g/L NaCl, 15 g/L Bacto-agar). The plate was incubated at 37 °C overnight. Colonies from the plate were picked with a sterile toothpick to start 3 ml YENB-tetracycline broth (7.5 g/l yeast extract and 15

g/l Bacto-tryptone with 20 μl/ml tetracycline). After growing at 37 °C overnight shaking at 250 rpm, the 3 ml culture was incubated into 1 liter of YENB medium at the same condition until the OD_{600} for the culture was 0.5. The culture was distributed in four sterile 500 ml centrifuged bottles and centrifuged for 10 minutes at 5000 rpm at 4 °C in the RC5-B in GS3 rotor. The cell pellets were resuspended into 100 ml of cold sterile water and centrifuged as above to remove the medium. This step was repeated once with sterile water and once with 10% cold sterile glycerol solution. Then, the cells were resuspended in 2 ml of 10% cold sterile glycerol solution and aliquoted into Eppendorf tubes with 40 μl in each tube. These Eppendorf tubes were placed on dry ice until the cell medium was frozen. The competent cells were stored in a –70 °C freezer immediately.

Transformation was performed on an *E. coli* Pulser (Bio-Rad) set for a 2500 kV pulse. The electro-competent cells were thawed on ice and 2 μl of ligation solution was added to 40 ul of competent cells in the cold room. The cells were mixed, transferred to an electroporation chamber and a 2500 KV pulse was applied for 8 second. 1 ml of YENB medium was immediately added to the chamber. After mixing, the cells were transferred to a small Falcon tube and were incubated at 37 °C with shaking at 250 rpm for 30 minutes. The cells were collected by centrifugation at 2000 for 5 minutes. The media was decanted, 25 μl isopropyl-thiogalactoside (IPTG, 25 mg/ml) and 25 μl 5-bromo-4-chloro-3-indolyl g-D-galactosidase (X-gal, 25 mg/ml in N,N-dimethylformamide) were added. These tubes then were briefly vortexed to resuspend the cells and the mixture was poured onto the surface of an LB plate supplemented with 100 μg/ml ampicillin. These inoculated plates were incubated for 16-20 hours at 37 °C.

3.3.5 Growth of the shotgun library

1.5 ml of TB media (section 2.2.1) supplemented with 100 μg/ml ampicillin was aliquoted to each well of a 96-well microtiter deep well block (Beckman). The white colonies that appeared on the LB plates were picked with sterile toothpicks and placed in each well. After 10 minutes, the toothpicks were removed from the blocks and the blocks were incubated for 18-20 hours at 37 °C shaking at 350 rpm. The blocks were centrifuged at 2500 rpm for 7 minutes to harvest the cells. Blocks with cell pellets were stored at –20 °C freezer. They were used as a shotgun sub-clone library for one BAC, or PAC or cosmid.

3.4 Isolation and sequencing the sub-clone library

The shotgun sub-clones were semi-automatedly isolated by single acetate cleared lysis procedures using the Biomek 2000 (Beckman) and automated Hydra 96 (Robbins). The detailed procedures were the same as described in section "2.2.2 Isolation of cDNA templates".

The detailed procedures for setting up cycle sequencing reaction using AmpliTaq-FS and BigDye-labeled terminators were the same as described in section "2.3 Sequencing of ESTs". The primers used in the human genome projects were M13 universal forward primer (5' TGTAAAACGACGGCCAGT 3') and M13/pUC reverse primer (5' CAGGAAACAGCTATGACC 3'). At times, it was useful to use the KlenTaq-TR DNA polymerase instead of the AmpliTaq-FS, together with either of two other types of dye-labeled terminators, the Rhodamine-labelled or the dRhodamine-labeled terminators. When KlenTaq-TR enzyme mix was used in cycle sequencing, 1 μl of

enzyme mix was added to one reaction. The other reaction components and procedures were the same as described in section 2.3.

Data collection procedures were the same as described in section "2.3.4 Collect data".

3.5 Sequence assembly, data analysis and submission

After electrophoresis and data collection, the resulting data was transferred via Fetch from the Macintach data collection computer to a Sun Sparc workstation. For each project, once the first 96 sequences were obtained, the data was analyzed via Powerblast versus the *E. coli* database to check the percentage of bacterial host contamination. If the contamination was less than 5 percent, then additional data was collected. However, if the bacterial host contamination was greater than 5%, the BAC clone was re-grow, and a new shotgun library was created.

The computer script phredPhrap was used to remove vector sequences and assemble the short end sequences into longer contigs. The algorithms, functions, input/output file formats and examples about the computer programs for assembly were described in section "2.4.1 Phred, Phrap and Cross_match". The phredPhrap assembly resulted in producing contigs containing overlapping end sequences of the sub-clones. The resulting contigs greater than 2kb were submitted to GenBank as Level 1 data within 24 hours of being generated.

When the average sequence redundancy was 3 or 4, the initial closure and proof reading began by employing semi-automated scripts written by Steve Kenton at ACGT to aid in picking primers for primer walking on subclones, or directly walk on cosmids,

BACs or PAC. Additional gap closure, proofreading, finishing and polishing methods will be described in section 3.6. The computer program Consed also was useful for viewing the phredPhrap assembly and aiding in picking primers and finishing a sequence. Section "2.4.2 Consed" contains a discussion how Consed was used to analyze data. When the contigs were ordered, the project was submitted to GenBank and upgraded to Level 2.

Then, final closure and proof reading was performed. Although at least three-fold sequence redundancy is required for each base to assure a 99.99% accurate final sequence, higher redundancies often are necessary to ensure the high overall accuracy of the entire final sequence. If a sequence redundancy follows the "Rule of 3", typically the error rate will be less than 1 error per 10,000 bases. Once a project was aligned into one contiguous sequence, this final sequence then was analyzed via programs such as XGrail, GenScan (section 1.6.6), Powerblast (section 1.6.5), and BlastX (section 1.6.4 and 2.4.3) to search regions of structural and biological interest. Finally, the resulting sequence and its annotation was submitted to GenBank as Level 3 data.

3.6 Gap closure, proofreading, finishing and submission

The gap closure step is one of the most challenging and time consuming steps in a large-scale sequencing project. During this dissertation research, beside the project BAC 322f3, I also closed the gaps and reduced the error rates for several other projects that had previously been worked on by others. These included cosmids 92f5, c48 and cosB1, PAC p856 and BAC 239c10 whose shotgun phase was completed by Dennis Burian, but not ligned to a consed derived error rate of less than 1/10,000. When I took over these

projects, c48 was in 11 contigs, p856 was in 9 contigs, cosb1 was 12 contigs. 92f5 was in one piece, and the error rate was higher than 1 error per 10,000 bp. 239c10 had 252 contigs. BAC mn1_90kb, a project that was began by Judy Crabtree and whose shotgun phase completed by Sharon Lewis, had 10 contigs.

3.6.1 Re-sequence

3.6.1.1 Re-sequence the clone

For these older projects, before using the strategies described below, additional shotgun data was added as a direct and efficient procedure for closing some gaps, for reducing error rates and to obtain additional shotgun clones for primer-walking-based closure. Most of the old data gathered before 1998 was collected from 36 cm long gels using KlenTaq TR DNA polymerase from short insert subclones (1-3 kb). Therefore, longer readings were obtained during this new shotgun phase to close some of the gaps and efficiently reduced the error rates. Here the entire cosmid, BAC or PAC clones were nebulized using a lower pressures (3-5 psi) to generate larger fragments (3-5 kb), subcloning these fragments and sequencing the subclones in the presence of 5%-10% DMSO (Winship, 1989) on a 48 cm gel with AmpliTaq FS DNA polymerase and ABI BigDye-terminators. Furthermore, as stated above, these subclones with a larger insert served as efficient templates for primer walking in the further closure steps.

3.6.1.2 Re-sequence the contig ends

Using both forward and reverse universal primers, many gaps were closed by re-sequencing the sub-clones at the ends of the contigs with different chemicals, longer gels, or different sequencers.

LICOR dye-terminator sequencing reactions also were used to re-sequence the contig ends to generate longer readings. Because all four types of LICOR dye-terminators contain the same dye label, each of the four reactions were incubated and loaded separately. LICOR achieves longer read lengths by avoiding the 'spectral deconvolution' required when all four reactions are combined. The protocol for this method is available at URL:

http://www.genome.ou.edu/LICORTerminatorSeqProtocol.html

The newly available dGTP Big Dye kits were useful to sequence through GC rich region. These kits are different from the standard kits because they have dGTP instead of dITP in their reaction pre-mix. Generally, dITP is used to limit G:C compressions. However, DNA polymerase incorporates dITP much more slowly than it incorporates dGTP. The combination of a hairpin and the dITP will slow down the polymerase and cause it to disassociate from the template. Therefore, dGTP kits were designed to overcome this type of early termination in sequencing. More information is available at URL:

http://www.genome.wustl.edu/gsc/TechD/question1.htm

For the regions that contained polyA or polyT trac, dRhodamine terminators were used in the reaction premix to efficiently extend through the early stop sequences that occurred when the BigDye mixes were used.

3.6.2 Primer walking

To elongate an end sequence for a sub-clone, a primer that was about 100 upstream or downstream of sequence end was picked. This custom synthesized primer then was used to "Primer walk" on the sub-clone. If shotgun sub-clones were available to

cover given gap regions, custom synthesized primers were used with the template sub-clones, which were picked using either Consed or the program "exgap" written by A. Hua in ACGT.

In many cases, shotgun sub-clones were not available to cover the given gap regions. For the smaller insert clones, for example, cosmids c48 and cosb1 which were less than 100 kb, their large scale isolated DNA were used directly as a template for primer walking with custom synthesized primers. To ensure the success for direct primer walking using cosmid DNA as the sequencing template, the "double acetate" protocol was required for large scale isolation of genomic DNA. Generally speaking, twice the amount of enzyme was used for the primer walking reaction premix. For clones that had inserts larger than 100 kb, such as BACs, where direct primer walking was not practical, PCR reactions were used as described in section 3.6.3 and 3.6.4 to generate a template for primer walking.

If the routine primer walking with AmpliTaq FS DNA polymerase and ABI BigDye-terminator was not successful, different chemicals described in 3.6.1 such as 5-10% DMSO, dGTP kits, dRhodamine terminators, also were used to improve primer walking processing by replacing the forward and reverse universal primers with custom synthesized primers.

Consed (section 2.4.2) was a very useful tool to view a phredPhrap assembly and to pick most of the primers for primer walking and PCR in this dissertation research. PrimOU, a program that was modified from the Whitehead Institute Primer and the South West Medical Center Primo by Steve Kenton, also was used as a new primer-picking tool.

In the early years of this dissertation research, primers used for primer walking and PCR were synthesized with the Beckman 1000M eight-column oligo synthesizer using phosphoramidite chemistry. The oligonucleotide was cleaved from the support beads and deprotected by gas-phase AMA, eluted with ddH_2O and measured A_{260} before using.

With the introduction of the MerMade, an oligonucleotide synthesis instrument that was developed at South West Medical Center in Dallas, oligonucleotides were synthesized faster and at lower cast. Primer walking with these inexpensive, custom synthetic primers allowed a more direct and efficient closure and proof reading in this dissertation research.

3.6.3 PCR (polymerase chain reaction)

PCR was performed using the entire cosmid, BAC or PAC as the template to amplify the given gap regions where no subclones were available and where sequencing directly off the large insert cosmid, BAC or PAC did not yield suitable results. Primers for PCR reactions were designed to bind at a location approximately 100 bp upstream or down stream of the gaps. A 50 µl PCR solution typically contains 10 ng of template DNA, 50 pmol of each primer, 10 nmol of each deoxynucleotide, 5 µl of 10x PCR buffer (containing 500 mM KCl, 100 mM Tris-HCl, 15 mM $MgCl_2$, pH 8.5) and 2.5 units of AmpliTaq DNA polymerase. The solution was incubated at 95 °C for 5 minutes before thermal cycling began to completely denature the template DNA. Then the reaction was incubated for 25 cycles of 95 °C for 1 minute, 55 °C for 1 minute and 72 °C for 2 minutes. The reaction was held at 4 °C for further processing. 10 µl of PCR product was loaded on a 0.7% agarose gel to estimate the product concentration and size. Then, 5U of

exonuclease I and 1U of shrimp alkaline phosphatase was added to 10 μl PCR product to degrade and dephosphylate unreacted primers and dNTPs (http://genome.wustl.edu/gsc/Protocols/EXOSAP.shtml). The clean up reaction was programmed to incubate in the thermocycler at 37 °C for 30 minutes, 80 °C for 15 minutes and then held at 4 °C. Low melting temperature agarose gel electrophoresis and phenol extraction (section 3.2.5) was an alternative method for purifying PCR products that was used in MultiPlex PCR described in section 3.6.4. After either clean up procedure, the PCR product was used for primer walking using PCR primers and other custom synthesized primers. Here, the sequence reaction mix contained 2-3 μl of PCR product, 1-2 μl of primer used in the PCR reaction, 1 μl 25% DMSO and 1 μl of BigDye mix. Thermocycling conditions were the same as described in section 2.3.2 for BigDye thermocycling reactions.

The PCR product generated by AmpliTaq was generally 1-4 kb long while the GeneAmp XL PCR kit (Perkin-Elmer) was used to generate a PCR product that was longer than 5 kb.

For templates that had a GC-rich simple repeat region, a reaction mixture without KCl but containing 5% DMSO was added to the reaction buffer to successfully obtain an amplified product.

In some instances, either difficult regions or when long range PCR was used, the PCR products were sheared into small (~500 bp) fragements using the nebulizer and sub-cloned into the pUC 18 vector for sequencing. End sequences of these "shatter" library clones were assembled separately to obtain a concensus sequence and the concensus

sequence subsequently was copied into the shotgun library database to cover the difficult region.

3.6.4 MultiPlex PCR-based Method

If the contigs could not be ordered, MultiPlex PCR, developed by Fares Najar in our lab, was performed to produce multiple PCR products within one tube. If fewer than 10 primers were used in multiplex PCR, 1 μl of each Mermade-synthesized primer (~20 pmole/ul) were pooled and used directly for multiplex PCR. If however, more than 10 primers were used, 1 μl of each primer was pooled into a microcentrifuge tube and dried in the vacuum drier. The dried primers then were dissolved in 10 μl sterile distilled water and used for multiplex PCR. The solution for a 50 μl MultiPlex PCR reaction contained 10 μl of 5X PCR buffer (83 mM $(NH_4)_2SO_4$, 335 mM Tris-HCl (pH 9.0), 33.5 mM $MgCl_2$, 50 mM b-mercaptoethanol, and 850 ug/ml bovine serum albumin), 10 μl of primers (see above) (~20 pmole each), ~250 ng template DNA, 12 μl of dNTP mix (25 mM each), 2 units of Taq polymerase XL (Perkin-Elmer), 2.5 μl 25% DMSO. This mixture was incubated in a thermocycler for 6 minutes at 94 °C and then for 30-40 cycles of 94 °C for 30 seconds, 55 °C for 30 seconds and 65 °C for 4 minutes followed by holding at 4 °C until further processing. Then, one of two alternative methods were used to purify the Multiplex PCR products. One method was to add 10 units of shrimp alkaline phosphatase and 100 units of exonuclease I to the product and incubate at the same conditions described in section 3.6.3. The other method was to purify the PCR products by low melting temperature agarose gel electrophoresis described in 3.2.5. The subsequent cycle sequence reaction mix contained 2 μl of PCR product, 2 μl of primer used in the PCR reaction, 1 μl 25% DMSO and 1 μl of BigDye mix. Typically, 60 cycles

in the thermocycler under the thermocycle conditions as described in section 2.3.2 for BigDye thermocycling reaction.

3.6.5 Proofreading, finishing and submission

As mentioned above, Consed (section 2.4.2) was a very useful tool to view and proofread the final sequences. After all the gaps were closed, Consed was used to scan for the low-quality regions requiring additional data, high-quality regions that disagreed with the consensus and assembly errors. Once a weak region was found, it was treated as a "gap" and the methods described in the above sections, re-sequencing, primer walking or PCR, were used to add more accurate data to the weak region to raise its quality value. In many cases, gap closure, proofreading and polishing were processed at the same time.

Finished sequences were analyzed by performing Powerblast against GenBank nr and other methods (section 1.6.) and submitted to GenBank by Sequin (section 1.6.7).

Chapter IV

EST data analysis and discussion

During the *Fusarium sporotrichioides* EST project, 7495 high quality ESTs were generated and released on our ftp site at URL:

ftp://ftp.genome.ou.edu/pub/fs

and also on the oracle database available on the ACGT at URL:

http://www.genome.ou.edu/cgi-bin/db/queryFusa.pl

BlastX homolog search of the GenBank non redundant protein database indicated that 49.38% of the ESTs had GenBank homologs and represented at least 2139 genes, while 50.62% did not have GenBank homologs and thus represented newly discovered genes.

4.1 Estimating EST quality

F. sporotrichioides EST database quality is summarized in Table 4.1.

After the gel image analysis and DNA sequence extraction, the low quality sequences and the contaminated sequences were removed from the database via the Clip and Clean scripts (section 2.5.1), which assessed the sequence quality and remove low quality data, trimmed the flanking vector sequences, removed the short insert sequences, wrong end sequences, ribosomal sequences, mitochondrial sequences and *E.coli* genome sequences. After a Phred quality value was determined for each base, the sequences that had an overall N ratio equal or larger than 1:5 also were removed. For the remaining sequences, the high quality start and stop positions were marked and low quality border

marked at the first base whose quality was lower than 15. If the high quality region was less than 100 bps, the whole sequence also was removed.

In this *F. sporotrichioides* EST project, 10,256 sequences were generated. Among them, 595 overall low quality sequences were removed. The primary reasons responsible for failure were long GC repeat sequence regions or multiple clones in one well. 728 3'ESTs were removed because of the stuttering in sequences due to the long poly (dT) repeat sequence at the beginning of the EST. Since DNA polymerase can slip out of register in long poly (dA) regions of the cDNA template during polymerization, and yield multiple lengths of the resulting poly (dT) sequences in a single reaction resulting the overlapping trace sequences in one lane on the sequencing gel (PerkinElmer, 1995). This was the main reason for difference in success rates between 5' ESTs (82.55%) and 3' ESTs (63.50%). This difference also has been noted in the human EST projects (Hillier et al., 1996) where rate for 5' EST was 76% and that for 3' EST was 63.5%.

The cDNA library also contained about 4.97% complete vector, 4.74% short inserts (< 100 bp), 2.14% wrong end inserts, 0.18% mitochondria, 0.54% ribosomal and 1.39% *E.coli* sequences. A total of 7495 high quality sequences (73.08%) passed Clip and Clean process, were placed on ACGT website, and used for further analysis.

The quality summaries for several EST libraries generated by different laboratories were compared and listed in Table 4.2. Since the Alu family, a set of approximately 300 bp dispersed, related sequences, occurs only in the human genome, in human EST projects, ESTs that have alu sequences as their inserts are removed from the high quality EST database. However, since alu sequences were not observed in the fungal EST projects, there was no need to remove them.

Table 4.1. *F. sporochioides* cDNA library and EST database quality summary

	3'(.f1)	5' (.r1)	Total #	Percentage
High quality	3238	4257	7495	73.08%
Low quality	301	294	595	5.8%
Long poly(dT)	728	--	728	7.10%
E. coli sequences	33	110	143	1.39%
Mitochondrial RNA	6	18	24	0.18%
Ribosomal RNA	17	38	55	0.54%
Complete vector	256	254	510	4.97%
Short insert (<100)	375	111	486	4.74%
Wrong end	141	79	220	2.14%

Total number of sequences: 10, 256

Table 4.2. Comparing qualities for EST databases from different laboratories

	Fusarium	***Aspergillus****	**A***	**B***	**C***
	10,256	14,885	77,922	3,321	25,461
High quality	73.08%	83.85%	83.33%	53.76%	88%
Low quality	5.8%	8.93%	6.94%	22.9%	3.02%
Long poly(dT)	7.10%	--	--	--	--
E. coli sequences	1.39%	1.40%	--	--	--
Mitochondril sequ.	0.18%	0.03%	0.04%	0.24%	0.01%
Ribosomal sequ.	0.54%	0.02%	0.01%	0.01%	0.01%
Complete vector	4.97%	1.20%	1.68%	2.29%	2.16%
Short insert (<100)	4.74%	2.72%	2.4%	20.8%	3.82%
Wrong end	2.14%	1.83%	--	--	--
Alu	--	--	5.6%	0.40%	3.0%

*, *Aspergillus nidulus* (Kupfer, 1999).
A*, Washington University-Merck EST project (Adams et al., 1995).
B*, Human liver cell EST (Okubo, 1992, 1994).
C*, Human cDNA libraries from neuromuscular tissue (Auffray et al., 1995).

4.2 Estimating library redundancy

Once every 100 new 3' EST sequences were obtained, cumulative 3' assemblies using Phrap were performed to determine the number of new genes represented. (section 2.5.2). Two methods were used to determine the redundancy of the library.

In the first method (Kupfer, 1999), the percent new genes was determined by the following formula:

% New genes = ((#New C/S – #Old C/S)/#New reads)x100

"#New C/S" was the number of singlets and contigs in the entire database summed up at the end of the present interval. "#Old C/S" was the number of the singlets and contigs in the entire database summed up at the end of the last interval. Therefore, "#New C/S – #Old C/S" was the number of the newly sequenced singlets and contigs in the present interval. "#New reads" was the number of the new sequences (not singlets or contigs) included in newly sequenced singlets and contigs. In this project, "#New reads" was 100 in each interval.

The percent redundancy was determined by:

% Redundancy = 1 - % New genes

Table 4.3 shows the cumulative percent new genes determined in batches of 100 in 3' EST assemblies.

The second method to calculate redundancy was a computer method developed by James White at ACGT. Rao's polynomial growth curve model (Rao, 1959) was implemented using computer statistical analysis system (SAS) to analyze repeated measured data (Qu and Palta, 1996), profile data and data for growth curve (Allen, 1983; Schneiderman et al., 1985, 1991), and to predict growth (Schneiderman et al. 1993). To assess the redundancies for EST projects, Jim White wrote a computer program using the similar SAS curve fit method which generates constants G and r from the 3' EST assembled data and uses these constants in the following formula to calculate redundancy:

$$\text{Redundancy} = 1 - r/(s/G + r)^2$$

In the above formula, "G" is the expected number of genes in the library, "s" is the number of 3' EST sequences and "r" is a redundancy factor. This program was used to predict the redundancy in this project as well as other EST projects in ACGT (Kupfer et al., 2000; Hua, 2001).

When the number of 3' ESTs was 3243, r was 1.3474 and G was 4176.7, the corresponding redundancies were calculated and are displayed in table 4.3. Fig. 4.1 show the plot of the number of EST reads against the percent of redundancy in the *F. sporotrichioides* database. The *F. sporotrichioides* EST project had a redundancy of 70% when 3243 3' ESTs were obtained. A redundancy of 70% was suggested as the point to discontinued for human EST projects (Hillier, 1996). Although new genes still could be detected after this point, it was decided to end the data collection because the small number of new ESTs that would result would be relatively expensive while generating a large number of redundant sequences.

The advantage of the second, newer method, developed by Jim White, is that a smoother curve results because the statistical analysis system used filters out short-term instantaneous changes and thus it more accurately captures the overall increases in redundancy.

4.3 Genes represented in EST database

4.3.1 Number of genes identified in the database

After performing 3' EST and 5' EST Phrap assembly and a BlastX homolog search against the non-redundant protein database (section 2.5.3), the number of genes represented in the EST database was calculated and summarized in Table 4.4. Totally,

Table 4.3 The determination of the library redundancy by cumulative 3' EST assemblies

Total Reads	CLUSTERS (C)	SINGLETS (s)	NewS/C-OldS/C	(NewS/C-OldS/C)/100	Redundancy*	
FS_100	7	79	88	0.88	12	28.35
FS_200	24	140	164	0.76	24	30.79
FS_300	44	194	238	0.74	26	35.31
FS_400	60	242	302	0.64	36	37.40
FS_500	77	289	366	0.64	36	39.40
FS_600	92	328	420	0.54	46	41.29
FS_700	103	371	474	0.54	46	43.11
FS_800	121	406	527	0.53	47	44.84
FS_900	143	428	571	0.44	56	46.49
FS_1000	160	465	625	0.54	46	48.07
FS_1100	170	506	676	0.51	49	49.58
FS_1200	189	539	728	0.52	48	51.02
FS_1300	206	568	774	0.46	54	52.41
FS_1400	216	603	819	0.45	55	53.73
FS_1500	229	639	868	0.49	51	55.00
FS_1600	241	676	917	0.49	51	56.22
FS_1700	260	710	970	0.53	47	57.40
FS_1800	268	738	1006	0.36	54	58.52
FS_2000	300	792	1057	0.51	49	59.60
FS_2100	314	823	1092	0.35	65	60.64
FS_2200	330	849	1137	0.45	55	61.64
FS_2300	357	845	1179	0.42	58	62.6
FS_2400	368	880	1202	0.23	77	63.53
FS_2500	393	886	1248	0.46	54	64.42
FS_2600	410	893	1279	0.31	69	65.28
FS_2700	428	918	1346	0.24	76	66.11
FS_2800	446	933	1379	0.43	57	66.91
FS_2900	456	962	1418	0.33	67	67.68
FS_3000	473	985	1458	0.40	60	68.42
FS_3100	493	1001	1494	0.36	64	69.14
FS_3243	518	1030	1548	0.37**	63	70.13

*The first value was the result from the first method and the second value was the result from the second method

**Calculated by (#New C/S - #Old C/S)/143 because the last interval was 143 instead of 100

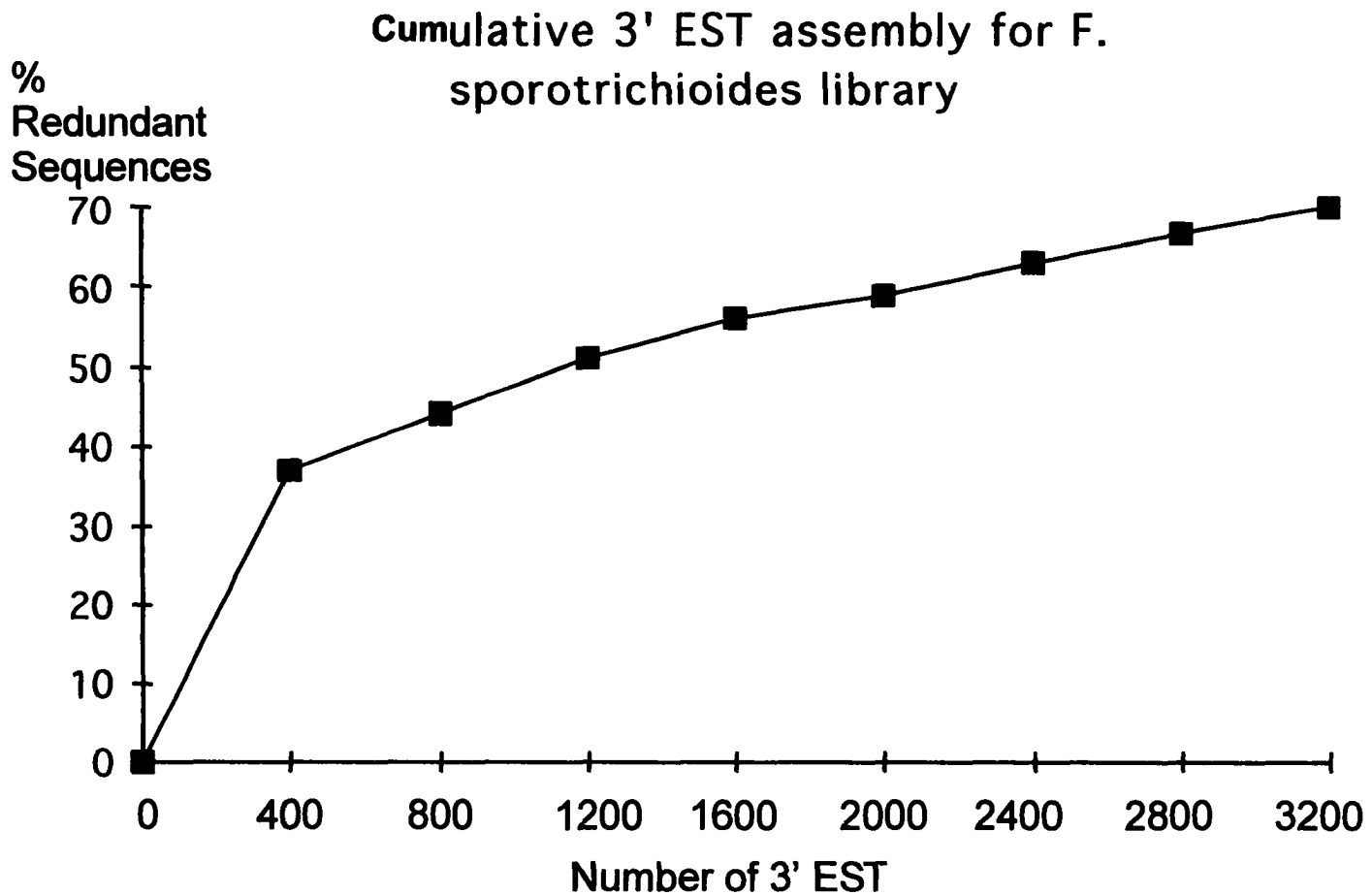


Fig. 4.1 Cumulative 3' EST assembly for *F. sporotrichioides* library

2181 singlets and 1057 contigs were obtained in the assembled database and 966 of the 2181 singlets and 633 of the 1057 contigs had significant BlastX homologs (Table 4.4).

To obtain the correct estimate of the expressed gene number, the singlets or contigs that had pair members in other singlets or contigs were subtracted from the total, because those paired singlets or contigs represented the same mRNAs and hence the same gene. Four situations were considered, singlets with pair member in singlets, clusters with pair member in clusters, clusters with pair member in singlets, singlets with pair member in clusters. If any of the above were observed, those singlets or clusters were only counted once instead of twice. Totally, 1099 duplications were found (Table 4.4). Substracting 1099 from the total singlet and contig number of 3238 yielded the conclusion that 2139 genes were represented in this database (Table 4.4).

During construction of the NIH's human unigene database, two clusters were merged only when "at least two 5' ends to a single cluster which contains at least two 3' ends from the same clones are found" (http://www.ncbi.nlm.nih.gov/UniGene/build.html). In *F. sporotrichioides* EST library, only 168 contigs fell into this category. Taking this into consideration, the total number of espressed genes in the database would be 3238 – 168 = 3052.

Table 4.4 3' EST and 5' EST Phrap assembly and BlastX results summary

	BlastX hit	**No hit**	**Total**
Singlets	966	1215	2181
Clusters	633	424	1057
Total	1599	1639	3238
Singlets with pair member in singlets		491	
Clusters with pair member in clusters		312	

Clusters with pair member in singlets 216
Singlets with pair member in clusters 80
Total 1099
Total genes represented by database: 3238 – 1019 = 2139

Clusters with two pair members in another clusters 186
Total genes represented by the database: 3238 – 186 = 3052

4.3.2 Gene expression level

3' and 5' assembly of the EST database for high quality sequences were performed by Phrap as described in section 2.5.3. Those ESTs that represented the same gene were assembled together to form a cluster. Table 4.5 shows the cluster sizes and the frequency of the clusters in each size.

Generally speaking, if one EST is not homologous with another EST, they will not be assembled into a cluster and therefore they will remain as singlets. However, when Phrap recognizes one singlet is somewhat homologous with another singlet or cluster, but the homologous score is less than border line score to assemble them into one cluster, Phrap places the sequence in a "contig" containing only one sequence. Therefore, in Table 4.5, although the largest cluster contains 164 ESTs, the smallest clusters contain only one EST.

Previous studies (Bishop et al., 1974; Soares et al., 1994) suggested that the population distribution of clones in cDNA library be divided into three frequency classes. Clusters in the low abundant class had one or two ESTs. Clusters in intermediate abundant class had more than two ESTs and clusters in the high abundance class had 10-15 highly expressed genes.

Following the classification method of Soares, the *F. sporotrichioides* ESTs population also fell into three classes (Table 4.6). The high abundance class consisted of

12 contigs and represented 11.57% of the ESTs in the database. The intermediate abundance class consisted of 496 contigs and represented 45.08% of ESTs in the database. The low abundance class consisted of 2182 singlets and 549 contigs and represented about 43.35% of ESTs in the database.

The population distribution of *F.sporotrichioides* ESTs was very similar to that described by Bishop and Soares. They characterized that the very abundant class consisted of 10-15 clusters that represented 10-20% of the total mRNA, the abundant class represented approximately 50% of the total mRNA and a low abundant class represented 30-40% of the mRNA (Bishop et al., 1974; Soares et al., 1994).

Table 4.7 compared EST distributions of the three abundance classes between *F.sporotrichioides* EST database and *A. nidulans* EST database (Kupfer, 1999). In *A. nidulans,* the percentage of ESTs in the intermediate abundant class (60.9%) was higher than that in *F.sporotrichioides* (45.08%) and the percentage of ESTs in the low abundant class was lower (25.5%) than that in *F.sporotrichioides* (43.35%). This result is in agreement with the conclusion that *A. nidulans* cDNA library was sequenced more deeply than the *F.sporotrichioides* cDNA library.

4.3.3 Top 12 highly expressed genes

4.3.3.1 Top 12 highly expressed genes in *F. sporotrichioides*

A BlastX homology search was performed with each member of the assembled EST database as described in section 2.5.3. The BlastX results for the top twelve most highly expressed gene members were summarized in Table 4.8. Five of the twelve highly expressed genes were genes that were involved in the biosynthesis of trichothecenes.

Table 4.5 Cluster sizes and the frequency of clusters in each size after 3' and 5' EST assembly

Cluster Size	Frequency of Cluster	Cluster Size	Frequency of Cluster
1	30*	25	1
2	519	27	1
3	164	28	3
4	102	29	2
5	57	31	1
6	41	33	2
7	27	34	1
8	17	36	1
9	11	40	2
10	4	41	1
11	8	42	1
12	7	44	2
13	6	45	1
14	7	46	1
15	6	47	1
16	5	49	1
17	1	50	1
18	1	54	2
19	4	65	1
20	2	71	1
21	3	76	1
23	2	90	1
24	2	101	1
		164	1

*There were 2128 singlets in the database besides those 30 "contigs" with single read

Table 4.6 Summary of gene expression level in *F.sporotrichioides* cDNA libray

Cluster size	classification	Total ESTs	#singlets/clusters
1-2 ESTs	Low abundant	3249 (43.35%)	549+2181=2730
3-45 ESTs	Abundant	3379 (45.08%)	496
46-164 ESTs	High abundant	867 (11.57%)	12

Table 4.7 Comparison of gene expression levels between *F.sporotrichioides* and *A. nidulans* cDNA libraries

Classification	Cluster size	*F.sporotrichioides*	Cluster size	*A. nidulans*
Low abundant	1-2 ESTs	3249 (43.35%)	1-2 ESTs	3184 (25.5%)
Abundant	3-45 ESTs	3379 (45.08%)	3-62 ESTs	7602 (60.9%)
High abundant	46-164 ESTs	867 (11.57%)	63-363	1704 (13.6%)

Table 4.8 Top ten highly expressed gene members in *F. sporotrichioides* cDNA libray

Contig ID	# of ESTs	Score	E value	homology start/end base, GenBank entries Names of genes
1057	164	2592	1.2e-268	41-1579, sp|Q12612|TRI4_FUSSP Trichodiene oxygenase, *F.sprorotrichioides*
1056	101	2015	1.6e-207	67-1188, sp|P13513|TRI5_FUSSP Trichodiene synthase, *F.sprorotrichioides*
1055	90	2220	2.8e-229	82-1410, sp|P34825|EF1A_TRIRE Elongation factor, *Trichoderma reesei*
1054	76	829	7.9e-82	213-1259, pir||B31776 LAC12 3' region *Kluyveromycesmarxianus var. lactis* (yeast)
1053	71	2259	2.3e-233	120-1496, gb|AAD19745.1| Trichothecene 3-O-acetyltransferase, *F.sprorotrichioides*
1052	65	846	1.3e-83	128-1597, sp|P52718|PEPF_ASPNG Serine-type carboxypeptidase, *Aspergillus niger*
1051	54	2900	2.4e-301	208-2169, emb|CAA12224.1| ATP citrate lyase, *Sordaria macrospora*
1050	54	1473	3.8e-150	153-1517, sp|P54874|HMCS_SCHPO Hydroxymethylglutaryl-CoA synthase, *Schizosaccharomyces pombe* (yeast)
1049	50	667	1.1e-64	1188-1748, gb|AAB61278.1| V-ATPase, *Neurospora crassa*
1048	49	563	1.2e-53	248-853, gb|AAF34754.1|AF2218 Acid protease, *Sclerotinia sclerotiorum*
1047	47	763	0.0	102-1223, gb|AF326571.1| *Tri14* *F.sprorotrichioides*
1046	46	2120	1.2e-128	50-1390, gb|AAD13655.1|U22463 T-2 toxin biosynthesis protein (*Tri8*), *F.sprorotrichioides*

One of the twelve genes are new genes defined during this EST projects by our collaborators.

1) **Contig1057** consists of 164 ESTs that are homologous to trichodiene oxygenase. Trichodiene oxygenase (cytochrome P450 58) is encoded by *Tri4*. It functions in the oxygenation of trichodiene to yield a product of unknown structure in trichothecene biosynthesis pathway (Hohn et al., 1995a). Contig1057 is 1763 bp long and it has 1538 bp (from base 41 to 1579) overlapping the coding region (1560 bp) of trichodiene oxygenase from *F.sprorotrichioides* (Q12612). According to the protein database record (AAB72032), the protein for Q12612 is 520 amino acid long and it belongs to the cytochrome P450 family.

2) **Contig1056** consists of 101 ESTs that were homologous to trichodiene synthase. Trichodiene synthase is encoded by *Tri5* and it catalyzes the cyclization of farnesyldiphosphate in trichothecene biosynthesis pathway (Hohn and Beremand, 1989). Contig1056 is 1339 bp long and it has 1121 bp (from base 67 to 1188) overlapping the coding region (1122 bp) of trichodiene synthase from *F.sprorotrichioides* of GenBank entry sp|P13513|TRI5_FUSSP. According to the protein database record, trichodiene synthase from *F.sprorotrichioides* is 374 amino acid long.

3) **Contig1055** consists of 90 ESTs that are homologous to translation elongation factor eEF-1 alpha chain. In eukaryotes, the factor eEF-1 is responsible for bringing aminoacyl-tRNA to the ribosome in the translation process (Nakari et al., 1993). Contig1055 is 1695 bp long and it has1328 bp (from base 82 to 1410) overlapping the coding region (1380 bp) of elongation factor from *Trichoderma reesei* (GenBank entry

sp|P34825|EF1A_TRIRE). According to the protein database record, this elongation factor is 460 amino acid long.

4) **Contig1054** consists of 76 ESTs that are homologous to the LAC12 gene of *Kluyveromyces lactis* coding for an inducible lactose permease. Lactose permease is a transporter that is driven by the proton gradient across the cell membrane to cotransport H^+ and lactose (Chang and Dickson, 1988). Contig1054 is 1428 bp long, and it has 1046 bp (from base 213 to 1259) overlapping the coding region (1266 bp) of LAC12 gene of *Kluyveromyces lactis* (GenBank entry pir||B31776). According to the protein database record, this lactose permease from *Kluyveromyces lactis* is 422 amino acid long.

5) **Contig1053** consists of 71 ESTs that are homologous to trichothecene 3-*O*-acetyltransferase that plays a role in self-protection against trichothecenes (McCormick et al., 1999; Hohn et al., 1998; Kimura et al., 1998a, 1998b) and is encoded by *Tri101*. Contig1053 is 1571 bp long and it has 1376 bp (from base 120 to 1496) overlapping the coding region (1377 bp) of trichothecene 3-O-acetyltransferase from *F.sprorotrichioides* of GenBank entry gb|AAD19745.1|. According to the GenBank record, this trichothecene 3-O-acetyltransferase from *F.sprorotrichioides* is 459 amino acid long.

6) **Contig1052** consists of 65 ESTs that are homologous to Serine-type carboxypeptidase, an enzyme catalyzing the hydrolysis of peptide bonds (Mehta, 1996). Contig1052 is 1812 bp long and has 1470 bp (from base 128 to 1597) overlapping the coding region (1593 bp) of Serine-type carboxypeptidase from *Aspergillus niger* of GenBank entry sp|P52718|PEPF_ASPNG. According to the GenBank record, this Serine-type carboxypeptidase from *Aspergillus niger* is 531 amino acid long.

7) **Contig1051** consists of 54 ESTs that are homologous to the enzyme ATP citrate lyase. ATP citrate lyase catalyzes the breakdown of citrate to form cytosolic acetyl-CoA, which is required by processes of fatty acid biosynthesis and cholesterol biosynthesis in the cytosol (Nowrousian et al., 1999). Contig1051 is 2390 bp long. It has 1961 bp (from base 208 to 2169) overlapped the coding region, which is 2022 bp long in total, with ATP citrate lyase from *Sordaria macrospora* of GenBank entry emb|CAA12224.1|. According to the GenBank record, this ATP citrate lyase is 674 amino acid long.

8) **Contig1050** consists of 54 ESTs and it is homologous to hydroxymethylglutaryl-CoA synthase (HMG-CoA synthase). HMG-CoA synthase is an enzyme that catalyze the condensation of the acetoacetyl-CoA with an acetyl-CoA to form β-hydroxy-β-methylglutaryl-CoA (HMG-CoA) (EC 4.1.3.5) in ketogenesis (Katayama et al., 1995). Contig1050 is 1927 bp long and it has 1364 bp (from base 153 to 1517) overlapping the coding region, which is 1341 bp in total, of hydroxymethylglutaryl-CoA synthase from *Schizosaccharomyces pombe* (fission yeast) of GenBank entry sp|P54874|HMCS_SCHPO. According to the GenBank record, this Hydroxymethylglutaryl-CoA synthase from *Schizosaccharomyces pombe* is 447 amino acid long.

9) **Contig1049** consists of 50 ESTs that are homologous to *Neurospora crassa* putative 20kDa subunit of the V-ATPase gene. V-ATPase, which is located in plant vacuolar membranes, drives a Ca^{2+} pump and has a H^{+}/Ca^{2+} antiporter that maintains cytosolic Ca^{2+} levels in vacuole, the major site of intracellular Ca^{2+} storage (Foster and Kane, 2000). Contig1049 is 2083 bp long and it has 560 bp (from base 1188 to 1748)

overlapping the coding region, which is 528 bp long in total, of V-ATPase from *Neurospora crassa*, GenBank entry gb|AAB61278.1|. According to the GenBank record, this V-ATPase is 176 amino acid long.

10) **Contig1048** consists of 49 ESTs and it is homologous to the acid protease that catalyzes the hydrolysis of peptide bonds in acidic environment. Contig1048 is 1023 bp long and it has 605 bp (from base 248 to 853) overlapping the coding region of the acid protease, which is 756 bp long in total, from *Sclerotinia sclerotiorum*, GenBank entry gb|AAF34754.1|AF2218. According to the GenBank record, this acid protease from *Sclerotinia sclerotiorum* is 252 amino acid long.

11) **Contig1047** consists of 47 ESTs and it is homologous to *Tri14*, a gene recently submitted to GenBank by our collaborator. *Tri14* is a gene that located in trichothecene pathway gene cluster and under the positive control of *Tri10*. The function of *Tri14* is not yet known. Contig 1047 is 1428 bp long and it has 1121 bp (from base 102 to 1223) overlapping the coding region of the *Tri14* from *F. sporotrichioides*, GenBank entry gb|AAG46054.1|AF326571.1.

12) **Contig1046** consists of 46 ESTs and it is homologous to an acetyltransferases encoded by *Tri8*, that is involved in the acetylation of the trichothecene hydroxyl group (McCormick et al., 1996). Contig 1046 is 1485 bp long and it has 1340 bp (from base 50 to 1390) overlapping with the *Tri8* from *F. sporotrichioides*, GenBank entry gb|AAD136555.1|U22463.

4.3.3.2 Comparison of the highly expressed genes from different cDNA libraries

Among the top twelve highly expressed genes in *F. sporotrichioides* EST database, five of them were the trichothecene biosynthesis pathway genes. The total number of ESTs for those five highly expressed genes, trichodiene oxygenase (164 ESTs), trichodiene synthase (101 ESTs) and trichothecene 3-O-acetyltransferase (71 ESTs), T-2 toxin biosynthesis protein (46 ESTs) and Tri14 (47 ESTs), is 429, which represented 5.72% of the total EST population in *F. sporotrichioides* database. Overview the whole *F. sporotrichioides* database, 7.62% of the total 7495 ESTs in the database represented the genes involved in trichothecene biosynthesis pathway. The expression levels of the trichothecene biosynthesis pathway genes will be further discussed in section 4.3.4. It was anticipated, and subsequently confirmed, that the trichothecene biosynthesis pathway genes would be highly expressed because the library was constructed from an *F. sporotrichioides* over producing mutant of the *Tri10* gene, which regulates trichothecene pathway gene expression.

The top twelve highly expressed genes in *F. sporotrichioides* (Table 4.8) are different from the top ten in *A. nidulans* (Kupfer, 1999). In *A. nidulans,* three of the ten highly expressed genes were Heat Shock Protein 30 (HSP30), which represented 5.5% of the total EST population in the *A. nidulans* EST database. In the *F. sporotrichioides* database, no HSP30 homologue was found. However, as listed in Table 4.9, 58 ESTs of heat shock protein were present in the *F. sporotrichioides* EST database and these only represented 0.78% of the total EST population in *F. sporotrichioides* database.

Heat Shock Proteins are synthesized by prokaryotic and eukaryotic cells as molecular chaperones or in response to stresses in the environment such as an increase in temperature. It was suggested that because the cells were grown at high temperature

(37°C) instead of normal temperature (20°C, Kusakabe, 1994) that HSP30 was highly expressed in *A. nidulans* cDNA library (Kupfer, 1999).

In *N. crassa*, the top ten highly expressed genes were either unknown genes or clock control genes (Zhu, unpublished). Approximately 9% of the population in the two *N. crassa* databases representing clock control genes (Zhu, unpublished).

Table 4.9 Heat shock protein in *F. sporotrichioides* EST database

Protein name	**singlets/cluseters**
Heat shock protein 88	Contig428
Heat shock protein 60	n2f06fs.f1, n2f06fs.r1
Heat shock protein 70	Contig731, Contig7, m1h04fs.r1, j2e03fs.r1
Activator of Hsp70 and Hsp90 chaperones	Contig449, r4b05fs.f1
Heat shock protein DDR48	Contig1045
Total: 58	

4.3.4 Gene expression levels for trichothecene biosynthetic genes

The total number of naturally occurring trichothecenes known today exceeds 60 (Besjardins, 1993). All trichothecenes share a tricyclic trichothecene nucleus (Fig. 1.9). The trichothecene biosynthetic pathway in *Fusarium* species is shown in Fig. 1.10 (Desjardins et al., 1993).

To date, fourteen genes in the trichothecene biosynthesis pathway have been identified in *F. sporotrichioides*. Thirteen of the fourteen genes products were found in the *F. sporotrioides* EST database (Table 4.10). Totally, 571 ESTs, 7.62% of the total 7495 ESTs in the database, represented genes involved in trichothecene biosynthesis pathway. This total is similar to the sum of ESTs involved in metabolism of carbohydrates, amino acids, nucleotides, nucleic acids, fatty acid and sterols (474 ESTs).

Interestingly, the *Tri6* gene product was not found in the *F. sporotrichioides* database. *Tri6* encodes a transcription activator factor that is required for trichothecene

biosynthesis pathway gene expression (Proctor et al., 1995). Since Matsumoto et al. (1999) reported that "*Tri6* appeared to be expressed for only a limited period prior to the toxin production.", it is likely that this gene product is either at low level or absent in the cDNA library studied.

The expression level for each trichothecene biosynthetic gene is summarized in Table 4.10.

1) *Tri3* encodes an acetyltransferase that converts 15-decalonectrin to calonectrin (McCormick et al., 1996). Disruption of the *Tri3* gene results in the accumulation of 15-decalonectrin instead of T-2 toxin. In this study, 11 ESTs were homologous to *Tri3* product.

2) 181 ESTs represented trichodiene oxygenase (cytochrome P450 58), a protein encoded by the *Tri4* gene. This protein functions in oxygenation of trichodiene to yield a product of unknown structure (Hohn et al., 1995a);

3) 110 ESTs represented trichodiene synthase that catalyzes the cyclization of farnesyldiphosphate to trichodiene (Hohn and Beremand, 1989) and trichodiene synthase is encoded by the *Tri5* gene;

4) 11 ESTs represented the *Tri7* gene product and 50 ESTs represented the *Tri8* gene product, both of which encode two acetyltransferases. Acetyltransferases are involved in the acetylation of the trichothecene hydroxyl groups (McCormick et al., 1996);

5) 5 ESTs represented the *Tri9* gene product, of which the function is unknown.

6) 36 ESTs represented the isotrichodermin C-15 hydroxylase (cytochrome P450 65A1) which is encoded by *Tri11*. This protein catalyzes the addition of a hydroxyl group

Table 4.10 Gene expression level for trichocheecene biosynthesis genes

ContigID	EST#	Score	E value	GenBank entries
15-decalonectrin 15-O-acetyltransferase (*Tri3*) (Total ESTs: 11)				
Contig797	4	1413	1.1e-143	gb|AAD13653.1|
Contig916	7	522	2.7e-49	gb|AAD13653.1|
Trichodiene oxygenase (Cytochrome p450 58, *Tri4*) (Total ESTs: 181)				
Contig1057	164	2592	1.2e-268	sp|Q12612|TRI4_FUSSP
Contig199	2	730	2.3e-71	sp|Q12612|TRI4_FUSSP
Contig988	13	500	5.6e-47	sp|Q12612|TRI4_FUSSP
Contig230	2	403	9.8e-37	sp|Q12612|TRI4_FUSSP
Trichodiene synthase (*Tri5*) (Total ESTs: 110)				
Contig1056	101	2015	1.6e-207	sp|P13513|TRI5_FUSSP
Contig899	6	584	7.2e-56	sp|P13513|TRI5_FUSSP
Contig12	1	473	4.1e-44	sp|P13513|TRI5_FUSSP
Contig14	1	282	7.4e-24	sp|P13513|TRI5_FUSSP
Contig27	1	266	3.7e-22	sp|P13513|TRI5_FUSSP
T-2 toxin biosynthesis protein (*Tri7*) (Total ESTs: 11)				
Contig947	8	1105	4.5e-111	gb|AAD13654.1|
Contig582	3	375	4.1e-42	gb|AAD13654.1|
T-2 toxin biosynthesis protein (*Tri8*) (Total ESTs: 50)				
Contig1046	46	2120	1.2e-218	gb|AAD13655.1|
Contig793	4	724	1.1e-70	gb|AAD13655.1|
Tri9 (Total ESTs: 5)				
Contig830	5	660	1.0e-51	gi|13621063|gb|AF359360.1|
Tri10 (Total ESTs: 2)				
j3h01fs.r1				gi|13621063|gb|AF359360.1|
d3g05fs.f1				gi|13621063|gb|AF359360.1|
Trichothecene 3-O-acetyltransferase (*Tri101*) (Total ESTs: 71)				
Contig1053	71	2259	2.3e-233	gb|AAD19745.1|
Isotrichodermin C-15 hydroxylase (Cytochrome p450 65A1, *Tri11*) (Total ESTs: 36)				
Contig928	7	1699	5.1e-174	sp|O13317|TR11_FUSSP
Contig844	5	408	3.4e-37	sp|O13317|TR11_FUSSP
Contig1026	24	271	2e-20	sp|O13317|TR11_FUSSP
Trichothecene efflux pump (*Tri12*) (Total ESTs: 12)				
Contig917	7	1104	5.8e-111	gb|AAD12756.1|
Contig673	3	1101	1.2e-110	gb|AAD12756.1|
Contig145	2	758	2.6e-74	gb|AAD12756.1|
Tri13 (Total ESTs: 25)				
Contig995	14	2300	1.2e-185	gi|13194731|gb|AF330109.1|
Contig976	11	2130	1.9e-171	gi|13194731|gb|AF330109.1|
Tri14 (Total ESTs: 47)				
Contig1047	47	763	0.0	gb|AF326571.1|
Tri16 (Total ESTs: 10)				
Contig936	10	603	e-171	gb|AF327521.1|
Total ESTs: 571				

at C-15 of isotrichoderm (Hohn et al., unpublished);

7) 12 ESTs represented an efflux pump, which is encoded by *Tri12* and is a transport protein (Alexander et al., 1999).

8) 2 ESTs represented the regulatory gene, *Tri10* (Andrew Tag, personnel communication). Mutations in *Tri10* gene will cause a dramatic increase or decrease in toxin production (Peplow et al., 1997; Gurifulina et al., 1998; Tag et al., 1998). The cDNA library studied in this dissertation contained a *Tri10* gene that was mutated such that the cDNA library had about a 2-fold enrichment in genes under the control of the *Tri10* gene product (Beremand et al., unpublished).

9) 71 ESTs represented the trichothecene 3-*O*-acetyltransferase which is encoded by *Tri101*. This protein converts isotrichodermol to isotrichodermin and is required for the biosynthesis of T-2 toxin (McCormick et al., 1999). Disruption of this gene leads to a slightly reduced growth on trichothecene-containing media, which suggests this gene may play a self-protection role against trichothecenes (Alexander, 1999). This is consistent with the observation that the homologous *Tri101* in F. graminearum plays a role in self-protection against trichothecenes (Hohn et al., 1998; Kimura et al., 1998a, 1998b). *Tri101* is the only known trichothecene biosynthetic gene that is not closely linked to other pathway genes in the gene cluster.

10) 25 ESTs represented *Tri13*, a gene recently submitted to GenBank by our collaborator. *Tri13*, a gene that locates in trichothecene pathway gene cluster, is a putative cytochrome p450 monooxygenase under the positive control of *Tri10* (Peplow et al., unpublished).

11) 47 ESTs represented *Tri14*, a gene recently submitted to GenBank by our collaborator. *Tri14*, a gene that locates in trichothecene pathway gene cluster, is under the positive control of *Tri10* (Peplow et al., unpublished), although its function presently is not known. This contig was sequenced into a cDNA sequence and submitted to GenBank (section 4.7).

12) 10 ESTs represented *Tri16*, a gene recently submitted to GenBank by our collaborator. *Tri16* encodes a putative C2H2 zinc finger protein that is under the positive control of *Tri10* (Peplow et al., unpublished).

The expression levels for trichothecine pathway genes varied over a wide range, as from 2 to 164 ESTs were observed for members of this pathway. This suggests that the regulatory mechanism for expression of the pathway genes is complex as besides *Tri10* and *Tri6*, other additional regulatory factors also may be involved.

4.4 Submission of ESTs to dbEST

EST data typically is submitted to the GenBank database of Expressed Sequence Tags (dbEST) system. Because EST projects generally contain large numbers of sequences with a great deal of redundancy in the citation, submitter and library information. Therefore, a special streamlined submission process and data format was designed by NCBI to improve the efficiency of submission. The relevant information is on the web site:

http://www.ncbi.nlm.nih.gov/dbEST/how_to_submit.html

After performing Clip and Clean and BlastX against GenBank as described in sections 2.5.1 and 2.5.3, a homology file was produced for each EST (Table 4.11). The

sequence, BlastX result, and library information were included in this file. In addition, the homology file for each EST was placed on ACGT's ftp site.

ESTs were submitted to dbEST through batch submission. For batch submission, each single EST was required to have a file of its own that was the same as the homology file described above (Table 4.11), except only the best homolog was kept in the "homology" portion. At the same time, all the ESTs from the same library shared three files that provided the library information, publication information and contact information. Examples of these three files for *F. sporotrichioides* are given in Table 4.12.

F. sporotrichioides ESTs were submitted to dbEST via email. The first step was going to the website: http://www.ncbi.nlm.nih.gov/dbEST/how_to_submit.html Then select batch-sub@ncbi.nlm.nih.gov. Then the files described above were attached to the email. The reply email address (if it was different from the sending address) also was provided in the email.

After GenBank received and accepted the submission requirement, a report was sent back from NCBI which contained the "DbEST_Id", "User_Id" and "GenBank_Accn" for each submitted EST (Table 4.13).

The distribution of sequence lengthes of submitted ESTs are summarized in Table 4.14.

4.5 Biological function assignments

At the end of the data analysis stage, each singlet or contig was assigned a biological function based on the results of the BlastX homology search. This procedure was performed as described in section 2.5.4.

Table 4.11 An EST homology file

TYPE: EST
STATUS: New
CONT_NAME: Bruce A. Roe, University of Oklahoma, broe@ou.edu
CITATION:
LIBRARY: Fusarium sporotrichioides Tri 10 overexpressed cDNA library
EST#: o1b05fs.r1
CLONE: o1b05fs
SOURCE: M.N. Beremand. Department of Plant Pathology, Texas A&M University, College Station, TX (e-mail address mnb3107@acs.tamu.edu)
SOURCE_DNA:
SOURCE_INHOST:
OTHER_EST:
INSERT:
ERROR:
P_END: 5'
SEQ_PRIMER: T3
HIQUAL_START: 1
HIQUAL_STOP: 428
DNA_TYPE: cDNA
PUBLIC:
COMMENT:
Contact Dr. Marian Beremand regarding clone availability
HOMOLGY:
8.1e-73 sp|P13513|TRI5_F TRICHODIENE SYNTHASE (SESQUITERPENE CYCLASE) (TS ....
1.7e-70 sp|P27679|TRI5_G TRICHODIENE SYNTHASE (SESQUITERPENE CYCLASE) (TS ....
2.8e-70 sp|Q00909|TRI5_G TRICHODIENE SYNTHASE (SESQUITERPENE CYCLASE) (TS ....
1.2e-69 sp|Q00835|TRI5_F TRICHODIENE SYNTHASE (SESQUITERPENE CYCLASE) (TS ....
1.0e-58 sp|O59947|TRI5_S TRICHODIENE SYNTHASE (SESQUITERPENE CYCLASE) (TS ....
3.9e-57 sp|O13489|TRI5_M TRICHODIENE SYNTHASE (SESQUITERPENE CYCLASE) (TS ....

SEQUENCE:
TTGTTACCAATACACCTTGGCCAATATGGAGAACTTTCCCACCGAATACTTCCTTAACAC
TACCGTGCGCCTTCTCGAGTATATACGATACCGAGATAGCAACTACACTCGAGAAGAGCG
TATCGAGAATTTGCACTATGCTTATAACAAGGCTGCTCATCACTTTGCTCAGCCACGCCA
ACAGCAGCTACTCAAGGTAGACCCCAAGCGACTACAGGCTTCCCTCCAAACAATCGTTGG
CATGGTCGTGTACAGTTGGGCAAAGGTCTCTAAAGAATGCATGGCCGATCTATCTATACA
TTATACCTACACTCTCGTTTTGGACGATAGCAAGGATGACCCGTACCCAACCATGGTAAA
CTATTTCGACGATCTTCAAGCTGGACGCGAACAAGCCCATCCGTGGTGGGCACTCGTCAA
CGAGCACTTTCCCAATG

|Table 4.12 Three files required for batch submission of EST to dbEST

File 1: Library information

TYPE:	Lib
NAME:	Fusarium sporotrichioides Tri 10 overexpressed cDNA library
ORGANISM:	Fusarium sporotrichioides
STRAIN:	^Tri 10
TISSUE:	
VECTOR:	pBlueScript SK-
V_TYPE:	phagemid
RE_1:	5' end of cDNA cloned into EcoRI site of pBluescript
RE_2:	3' end of cDNA cloned into XhoI site of pBluescript

File 2: Publication information

TYPE:	Pub
MEDUID:	
TITLE:	A Fusarium sporotrichioides EST Database
AUTHORS:	Qun Ren, *Andrew Peplow, *Andrew Tag, Andy Peterson, Hongshing Lai, Doris Kupfer, *Marian Beremand, Bruce Roe. University of Oklahoma, *Texas A&M University
JOURNAL	Unpublished
VOLUME:	
PAGES:	
YEAR:	
STATUS:	Unpublished

File 3: Contaction information

TYPE:	Cont
NAME:	Bruce A. Roe, University of Oklahoma, broe@ou.edu
FAX:	405 325 7762
TEL:	405 325 4912
EMAIL:	broe@ou.edu
LAB:	Department of Chemistry and Biochemistry
INST:	Advanced Center for Genome Technology, University of Oklahoma
ADDR:	620 Parrington Oval, Norman, OK 73019

Table 4.13 Examples of report from dbEST

DbEST_Id	User_Id	GenBank_Accn
5814087	b4b04fs.f1	BE605248
5814088	b4b04fs.r1	BE605249
5814089	e4g01fs.f1	BE605250
5814090	e4g01fs.r1	BE605251
5814091	j2a05fs.f1	BE605252
5814092	j2a05fs.r1	BE605253
5814093	m4b07fs.r1	BE605254
5814094	m4b07fs.f1	BE605255

Table 4.14 Distribution of sequence lengths in three projects

Project name	EST	322f3	239c10
Total	7495	3689	9250
50+	0	90	570
100+	6	93	789
150+	33	127	929
200+	88	171	945
250+	231	170	941
300+	558	304	842
350+	1011	454	924
400+	1397	812	1062
450+	1584	1150	1125
500+	1659	304	738
550+	816	13	330
600+	112	0	51
650+	0	1	4
Average	430	421	353

4.5.1 The last version of keyword list and the form of biological function assignments

The primary keyword list was developed based on a list created by Doris Kupfer for the *Aspergillus nidulans* EST database (Kupfer, 1999). After several revisions (section 2.5.4.2), a final version was obtained and it is presented in Appendix I. Because one key word may have one or several variations, the variations are listed after the parent keyword behind an "&" symbol. All entries matching the variations are placed under the parent keyword. The headings are in bold to distinguish them from keywords.

The best BlastX hit for each singlet or contig was selected manually. After combining the BlastX results with the last version of the keyword list (section 2.5.4.2), the final biological function assignments were generated, This listing is presented in Appendix II. Each of the 1599 singlets or clusters that had significant BlastX hits in the non-redundant protein database were assigned a potential function and placed in the appropriate category. Key words are shown in brackets "< >" and a list of singlets/contigs

that had this keyword in their highest scored BlastX results is provided under each keyword. Those ESTs that had no significant homolog were not listed individually, but included in a "no significant homolog" category in Table 4.15.

4.5.2 Outline of biological function assignments

An outline of the categories of biological functions and the number of siglets/contigs falling in each category is presented in Table 4.15 and summarized in Table 4.16. Table 4.16 also shows the percentage of database members falling in each of the biological function categories and the percentage of those with no database homology. A summary representation of this data is given in Fig. 4.2.

4.5.3 Comparison of outline of biological function assignments from four EST databases

The semi-automated process of biological function assignments also was used in analysis of three other EST databases in ACGT (Kupfer, 1999; Zhu, 2001). Table 4.17 shows the comparison of biological function classification results from the 4 libraries analyzed here. Fig. 4.3 is a graphical summary of the data presented in Table 4.17.

The comparison of the four libraries showed that the percentage of genes in each biological division is remarkably similar in spite of the diverse sources of the libraries. The percentages for no match in GenBank ranged from 50% to 58%. The percentages for metabolism pathway ranged from 11% to 15%. The percentages for genetic information processing ranged from 9% to 15%, while those for cell growth were either 2% or 4% and the percentages for other cellular processes ranged from 6% to 10%.

Table 4.15 Outline of the categories of biological functions and number of database members falling in each category

PART I. Metabolic Pathways

I. Metabolism of Carbohydrates(for glucose see energy) [Total 53*]
1. Chitin metabolism (6*)
2. Cellulose metabolism (8)
3. Cutin metabolism (4)
4. Polysaccharide metabolism (3)
5. Galactose metabolism (10)
6. Mannitol and mannose metabolism (14)
7. Quinate metabolism (4)
8. Sorbitol metabolism (2)
9. others (2)

II. Metabolism of Amino acids and Related Molecules [Total 64]
1. arginine metabolism (9)
2. asparagine metabolism (3)
3. aspartic acid metabolism (5)
4. cysteine metabolism and biosynthesis (4)
5. glutamine metabolism and biosynthesis (4)
6. glycine metabolism (6)
7. histidine metabolism (6)
8. isoleucine metabolism (2)
9. methionine metabolism (2)
10. tryptophan metabolism and synthesis (3)
11. aromatic amino acid metabolism (1)
12. glutamate metabolism (8)
13. phenylalanine metabolism (3)
14. lysine metabolism (4)
15. tyrosine metabolism (1)
16. alanine metabolism (1)
17. others (3)

III. Metabolism of Nucleotides and Nucleic Acids, Purines, Pyrimidines [Total 18]
1. Nucleotide metabolism (5)
2. Purine metabolism (9)
3. Pyrimidine metabolism (3)
4. Salvage of the bases (1)

IV. Metabolism of Lipids, Fatty Acids, Sterols-See also fatty acid degradation [Total 53]
1. Fatty acid biosynthesis (12)
2. sterols (23)
3. lipids (18)

V. Sulfur, Phosphate and Nitrogen Metabolism [Total 23]
1. Sulfur Metabolism (7)
2. Nitrogen Metabolism (see also amino acid metabolism) (16)
-urea related

VI. Metabolism of Cofactors, prosthetic groups [Total 33]
1. nicotinamide coenzymes (5)

2. biocytin (biotin) (2)
3. thiamine (3)
4. coenzyme A (3)
5. flavins (1)
6. heme (14)
7. molybdopterin (1)
8. PMP (3)
9. others (1)

VII. Energy [Total 233]

VII.1. Carbohydrate as energy source
1. Glycolysis (27)
2. Gluconeogenesis (2)
3. Pentose-phosphate pathway (12)
4. Pyruvate dehydrogenase-three kinds of enzymes (7)
5. Tricarboxylic acid (TCA) cycle (21)
6. related reactions (3)
7. Fermentation, alcoholic (10)
8. Fermentation, other (1)
9. Monocarbon metabolism (5)
10. Metabolism of energy reserves (glycogen, starch, trehalose) (12)

VII.2. fatty acid as energy source
1. lipase-triacylglycerols to glycerol+FA (2)
2. beta-oxydation of fatty acids (6)
3. Ketone body metabolism (2)

VII.3. Metabolism of other energy sources

VII.4. Electron transport
1. Complex I-NADH-ubiquinone (23)
2. Complex II-Succinate-ubiquinone (6)
3. Complex III-Ubiquinone to cytochrome C (3)
4. Other electron transport pathways (29)
5. ATP synthase and ADP, AMP (25)
6. Alternative respiratory path (2)
7. Reducing carriers (6)

Part II. Regulatory Pathways

I. Genetic information Processing [Total 289]

I.1. DNA replication
1. DNA synthesis (9)
2. DNA packaging (16)

I.2. Transcription
1. RNA Polymerase (4)
2. Regulation (27)
3. Processing (14)
4. tRNA synthetase and ligase (19)
5. Degradation (6)

6. RNA binding (5)
I.3. Translation
1. initiation (19)
2. elongation (17)
3. termination (0)
4. Ribosomal proteins (33)
5. Post-translational modifications (14)
6. Folding and Targeting (36)
7. Turnover-protein degradation-including vacuolar (68)
8. protein binding (2)

II. Cell Growth, Cell Division, Mating and Morphogenesis [Total 127]

II.1. Cell walls, Biomembranes and Cytoskeleton
1. Cell walls (17)
2. Biomembranes (29)
3. Cytoskeleton (55)
4. organelle (5)

II.2. cell cycle control (9)
II.3. Mitosis/cytokinesis
1. mitosis (11)
2. cytokinesis (3)

II.4. Meiosis (2)
II.5. Cell death (1)

III. Processes

III.1. Cell rescue, defense, osmotic adaptation, starvation response, development [Total 126]

1. development (24)
2. defense and secondary metabolites (57)
3. detoxification (4)
4. salt tolerance (3)
5. starvation response (2)
6. DNA repair (11)
7. allergen and immune system proteins (9)
8. tumor protein and tumor suppressor (8)
9. multidrug resistance (7)
10. cell reaction to environment (1)

III.2. Cell signalling, signal transduction and secondary messenger [Total 60]

1. phosphatases (14)
2. Kinases (25)
3. cAMP-secondary messenger (3)
4. G protein (13)
5. Inositol triphosphate-secondary messenger (2)
6. other (3)

III.3. Transmembrane transport [Total 146]

1. secretion (3)
2. exoenzymes (3)
3. membrane transport (6)

4. mitochondrial transport (12)
5. transporting for sugar, cation, anion, protein, fatty acid etc.
5.1. sugar transport (8)
5.2. cation transport (46)
5.3. Anion transport (18)
5.4. Protein, amino acid transport (21)
5.5. fatty acid trasport (2)
5.6. ABC transporter family (11)
5.7. other (16)

Part III Unclassified

I. Classes of Enzymes [Total 57]
1. Oxidoreductases (39)
2. Transferases (7)
3. Hydrolases (8)
4. Lyases (2)
5. Isomerases (0)
6. Ligases (0)

II. Non-enzymatic classes (not in defined pathways) [Total 27]

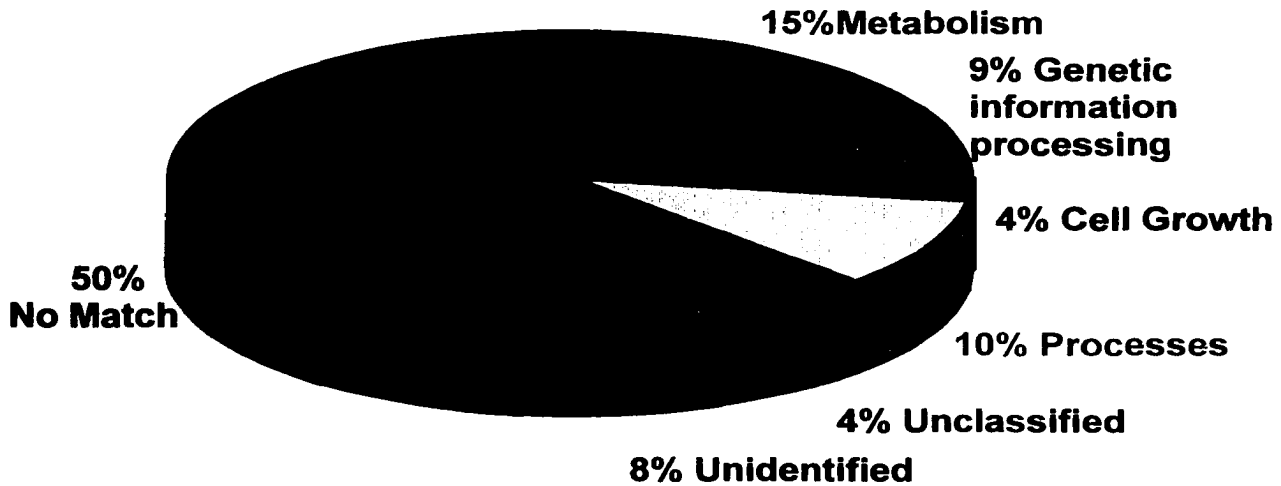
Part IV. Unidentified (includes significant match with ORFs) [Total 256]

PartV. No significant homolog
-Contigs[Total 424]
-Singlets[Total 1215]

* number of database members falling in each category

Table 4.16 Percentage of database members falling in each of the biological function categories and percentage of those without known homology

Categories	#ESTs	Percentage
Metabolic pathway	477	14.73%
Cell growth	127	3.92%
Genetic information processing	289	8.93%
Processes	332	10.25%
Unclassified	118	3.64%
Unidentified	256	7.91 %
Unknown	1639	50.61%
Total	3238	100%



Biological Function Classification for the *F. sporotrichioides* EST database

Fig. 4.2 Biological function classification

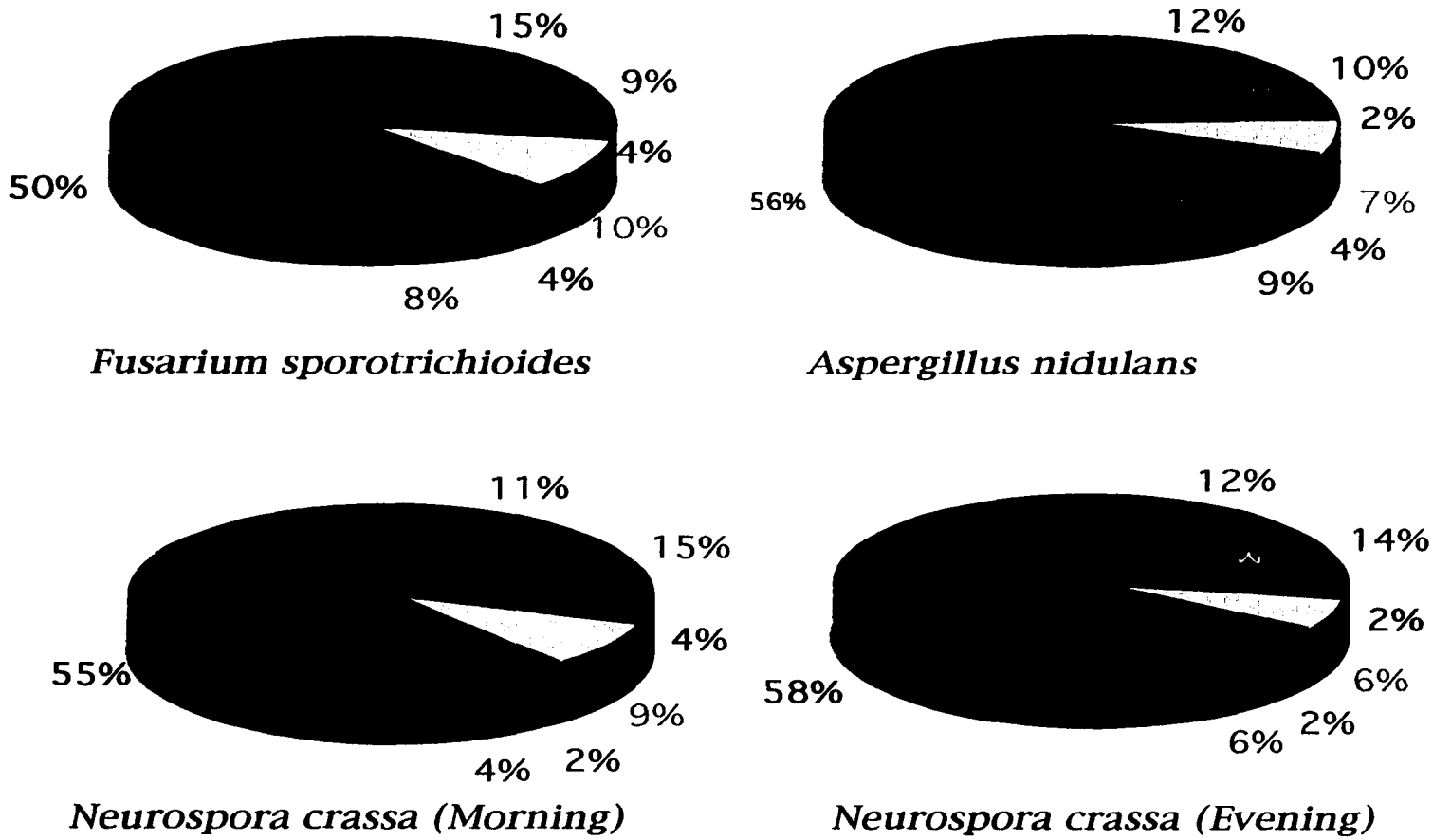


Fig. 4.3 Comparison of biological function classification results

Table 4.17 Comparison of biological function classification results from 4 EST databases at ACGT

Categories	*F. sporotrichioides*	*A. nidulans*	*N. crassa*(m)	*N. crassa*(e)
Metabolic pathway	15%	12%	11%	12%
Cell growth	4%	2%	4%	2%
Genetic information	9%	10%	15%	14%
Processes	10%	7%	9%	6%
Unclassified	4%	4%	2%	2%
Unidentified	8%	9%	4%	6%
Unknown	50%	56%	55%	58%

4.6 tBlastX against dbEST

The tBlastX program compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database and is a useful method for comparing one set of ESTs with another.

4.6.1 tBlastX against dbEST for members with no BlastX homologs against nr

Since 50% of the assembled members in *F.sporotrichioides* database had no significant homologues in the GenBank non-redundant protein (nr) database, these 1639 singlets/clusters likely represent previously un-discovered genes. When this group of *F. sporotrichioides* ESTs was compared against dbEST with tBlastX using the same stringency used before with BlastX (section 2.5.4.1), we observed that 323 database members had homologues in dbEST. Among them, 80 had their homologues in the *A. nidulans* database (Table 4.18 and Table 4.20) and 91 in the *N. crassa* databases (Table 4.19 and Table 4.20). Furthermore, 27 had homologues found in both the *A. nidulans* and *N. crassa* databases (Table 4.20). These 27 genes present in all four libraries most likely represent filamentous fungal specific genes (Table 4.24).

Two on-line computer programs, Pfam (http://pfam.wustl.edu/) and PROSITE (http://www.expasy.ch/prosite/), were used to analyze these 27 EST sequences and search

the functional protein domain matches, but the results were inadequate to be transformed into unambiguous understandable information. One of the limitations and fallibility that occurred in human genome annotations (Lander et al., 2001; Venter et al., 2001; Shoemaker et al., 2001) also occurred here, that is, computational methods alone are not efficient enough to reveal the precise structure and functions of the genes. Other experimental approaches -- full-length cDNA sequencing, gene lock out, mutagenesis, microarray technologies, have to be employed to explore the function of those genes.

4.6.2 tBlastX against dbEST for members with BlastX homologs against nr

When the *F. sporotrichioides* ESTs that have significant BlastX GenBank nr homologues (1599 singlets/clusters) were compared against dbEST using tBlastX, 550 singlets/clusters (34.40%) in the *F. sporotrichioides* database had observed homologues in the *A. nidulans* database (Table 4.21 and Table 4.23) and 411 singlets/clusters (25.70%) had homologues in the *N. crassa* databases (Table 4.22 and Table 4.23). Additionally, 201 singlets/clusters (12.57%) from this group also had homologues in both the *A. nidulans* and *N. crassa* databases (Table 4.23 and Appendix IV) and may represent fungal specific genes. Table 4.25 shows the percentage of members of this group that could be assigned into biological function categories. A graphical representation of this data is given in Fig. 4.4.

Although *F. sporotrichioides, A. nidulans* and *N. crassa* belong to the same phylum, Ascomycota, they represent three different families, Hypocreaceae, Trichocomaceae and Sordariaceae, respectively. Comparison of the expressed genes from these three species has helped us further understand the biological function similarities and differences among these Ascomycetes.

4.7 Submission of full-length cDNA sequences to GenBank

When a full-length cDNA sequence is useful for further study, it can be obtained by primer walking (Section 3.6) to extend the end sequences of ESTs from the same cDNA clone. Here, primers were picked with a special method as described in section 2.6.1. Once a full-length cDNA was obtained, it was compared against GenBank nr using the Blast programs described in section 1.6. and submitted to GenBank after annotation in Sequin (section 1.6.7).

Table 4.26 is the report sent back from NCBI containing the "GenBank_Accn" for each submitted full-length cDNA.

Table 4.18 tBlastX result showing the ESTs which had homologs in *A. nidulus* database

a1b08fs.fl	a1g02fs.rl	a2d05fs.rl	b1a12fs.rl
b1h05fs.rl	b3e06fs.fl	b4f01fs.rl	c1a01fs.fl
c2c12fs.fl	c2e01fs.fl	Contig10	Contig173
Contig194	Contig213	Contig359	Contig47
Contig473	Contig477	Contig498	Contig558
Contig586	Contig59	Contig599	Contig613
Contig628	Contig667	Contig732	Contig809
Contig821	Contig837	Contig885	Contig909
e4f11fs.rl	h4a04fs.rl	h4c01fs.rl	h4f04fs.rl
i1g09fs.rl	i2e08fs.fl	i3c04fs.rl	j1a04fs.fl
j1a05fs.fl	j3d02fs.fl	k2a03fs.rl	k4f07fs.rl
k4f11fs.fl	m1a12fs.rl	n1e10fs.rl	n2b06fs.fl
o2d04fs.fl	p3g03fs.rl	r4d11fs.fl	s2f10fs.rl
t4b08fs.fl			

Table 4.19 tBlastX result showing the ESTs which had homologs in *N. crassa* database

a2d05fs.rl	a2g04fs.rl	a2h07fs.rl	a4e06fs.rl
c2c12fs.fl	c3c10fs.rl	c4e11fs.rl	Contig121
Contig130	Contig227	Contig26	Contig272
Contig280	Contig290	Contig309	Contig314
Contig338	Contig339	Contig436	Contig453
Contig5	Contig508	Contig525	Contig53
Contig724	Contig740	Contig745	Contig758

Contig819	Contig822	Contig91	Contig912
Contig932	Contig962	Contig992	d4b08fs.r1
d4c11fs.r1	d4h12fs.r1	e2b06fs.f1	e3g05fs.f1
e4f01fs.r1	g1c10fs.f1	g3f08fs.r1	g4a01fs.r1
g4h12fs.r1	h4c12fs.f1	j2g06fs.r1	j3b03fs.f1
j4e11fs.f1	k2b06fs.f1	k2c11fs.f1	k3g04fs.f1
l2h12fs.r1	m2d03fs.r1	m3e05fs.f1	m3e05fs.r1
m4f03fs.r1	o1e01fs.r1	o2d09fs.f1	o3g06fs.f1
p4d05fs.f1	q2b12fs.r1	r4d02fs.f1	s3e07fs.r1

Table 4.20 tBlastX result showing the ESTs which had homologs in both *A. nidulus* and *N. crassa* databases

a1c01fs.r1	b3a05fs.r1	Contig1023	Contig278
Contig306	Contig335	Contig442	Contig443
Contig479	Contig564	Contig599	Contig613
Contig757	Contig905	e1h04fs.r1	h1c06fs.f1
h4h01fs.r1	i3f10fs.r1	j3e10fs.f1	j4c06fs.r1
m1b04fs.r1	m1f09fs.r1	o2d09fs.r1	o2e08fs.r1
o3g06fs.r1	r4c09fs.f1	t4a09fs.f1	

Table 4.21 Database members which have both BlastX hits against non-redundant protein (nr) database and tBlastX hits against *A. nidulans* EST database(Total: 341)

Contig100	Contig1002	Contig1004	Contig1011
Contig102	Contig1020	Contig1024	Contig1031
Contig1038	Contig1042	Contig1044	Contig1048
Contig1057	Contig108	Contig11	Contig122
Contig128	Contig15	Contig150	Contig154
Contig16	Contig178	Contig18	Contig199
Contig2	Contig202	Contig219	Contig231
Contig235	Contig242	Contig249	Contig253
Contig260	Contig263	Contig268	Contig279
Contig29	Contig294	Contig308	Contig313
Contig315	Contig326	Contig343	Contig344
Contig347	Contig351	Contig370	Contig374
Contig376	Contig380	Contig394	Contig398
Contig399	Contig410	Contig411	Contig42
Contig428	Contig435	Contig449	Contig452
Contig456	Contig46	Contig482	Contig488
Contig489	Contig491	Contig492	Contig496
Contig509	Contig51	Contig512	Contig514
Contig515	Contig529	Contig531	Contig532
Contig539	Contig554	Contig568	Contig573
Contig586	Contig59	Contig590	Contig592
Contig600	Contig602	Contig603	Contig608
Contig61	Contig616	Contig619	Contig627

Contig628	Contig635	Contig65	Contig662
Contig696	Contig700	Contig709	Contig715
Contig730	Contig74	Contig749	Contig752
Contig755	Contig756	Contig761	Contig771
Contig772	Contig777	Contig792	Contig817
Contig818	Contig82	Contig825	Contig83
Contig831	Contig837	Contig84	Contig850
Contig851	Contig858	Contig862	Contig87
Contig874	Contig880	Contig889	Contig9
Contig902	Contig908	Contig918	Contig928
Contig931	Contig941	Contig944	Contig951
Contig952	Contig955	Contig964	Contig971
Contig979	Contig99	Contig991	Contig993
a1e10fs.r1	a1f12fs.r1	a1g01fs.r1	a1h04fs.r1
a1h08fs.r1	a2b03fs.r1	a2e05fs.f1	a2f03fs.r1
a3e12fs.r1	a4a06fs.r1	a4c06fs.r1	a4g02fs.f1
a4g02fs.r1	b1e03fs.r1	b1g12fs.r1	b2e06fs.r1
b2g09fs.r1	b3b03fs.r1	b3f12fs.f1	b4a09fs.r1
b4e04fs.r1	b4g09fs.r1	b8h04fs.r1	c1f08fs.f1
c2a03fs.f1	c3b06fs.f1	c3h10fs.f1	c4c02fs.f1
c4c02fs.r1	c4f02fs.r1	c4h04fs.f1	c4h04fs.r1
d1a06fs.r1	d1e12fs.r1	d2c05fs.f1	d2c05fs.r1
d2h03fs.r1	d2h10fs.r1	d2h11fs.r1	d3a03fs.f1
d3a10fs.r1	d3b10fs.r1	d3c12fs.r1	d3e03fs.f1
d3h06fs.r1	d4a05fs.r1	d4b04fs.r1	d4c07fs.r1
d4d10fs.f1	d4g11fs.r1	d4h10fs.f1	e1g04fs.f1
e1h12fs.f1	e2a08fs.f1	e2f11fs.r1	e3f07fs.r1
e4b10fs.r1	e4g06fs.f1	e4g06fs.r1	f1c07fs.r1
f3a06fs.r1	f3b02fs.f1	f3f04fs.r1	f3f09fs.f1
g1b12fs.f1	g1c07fs.r1	g2a03fs.r1	g2c09fs.r1
g2h08fs.r1	g3b10fs.f1	g3c01fs.f1	g3d11fs.f1
g3e05fs.r1	g3g03fs.f1	g3g03fs.r1	g3g09fs.r1
g3g10fs.f1	g3h07fs.f1	g3h07fs.r1	g4b12fs.r1
g4c09fs.r1	g4h08fs.f1	g4h08fs.r1	h1d03fs.f1
h3b10fs.f1	h3f10fs.r1	h3g04fs.r1	h4b08fs.r1
i1b09fs.r1	i1c10fs.r1	i1g04fs.r1	i2c10fs.f1
i2e07fs.r1	i3g11fs.r1	i3h11fs.r1	i4g02fs.r1
i4h07fs.r1	j1a01fs.r1	j1d08fs.r1	j2c03fs.r1
j2f04fs.f1	j3d02fs.r1	j3h07fs.r1	j3h08fs.r1
k1a06fs.r1	k1b02fs.r1	k1e07fs.r1	k2a03fs.r1
k2c05fs.r1	k2h05fs.r1	k4b07fs.r1	k4f11fs.r1
k4g05fs.r1	k4h02fs.f1	l1b03fs.r1	l1h10fs.r1
l2d11fs.f1	l2h04fs.f1	l3a03fs.r1	l3h11fs.f1
l3h11fs.r1	l4e05fs.r1	m1a06fs.f1	m1f04fs.f1
m2d05fs.r1	m2d11fs.f1	m2e12fs.r1	m2h11fs.f1
m3a05fs.r1	m3a09fs.r1	m4a09fs.f1	m4g09fs.f1

n1e06fs.r1	n1f02fs.f1	n1h05fs.r1	n2b06fs.r1
n2d10fs.f1	n2d10fs.r1	n2e06fs.f1	n2e06fs.r1
n2f05fs.r1	n2h10fs.r1	n3a01fs.f1	n4c08fs.r1
n4e07fs.r1	n4h10fs.r1	o1a07fs.f1	o1a07fs.r1
o1b11fs.r1	o2a01fs.r1	o2d01fs.f1	o2e01fs.f1
o2e01fs.r1	o2g04fs.f1	o3a07fs.f1	o3e09fs.f1
o3g03fs.f1	o4b09fs.r1	o4e01fs.r1	o4e10fs.f1
o4f05fs.r1	o4g11fs.f1	p1h07fs.f1	p3g07fs.f1
p4a06fs.f1	q2c07fs.r1	q2e08fs.r1	q2f04fs.f1
q2f11fs.f1	q3a04fs.f1	r3h07fs.r1	r3h10fs.r1
r4d11fs.f1	r4e10fs.f1	r4f12fs.f1	r4f12fs.r1
r4g07fs.f1	s1b07fs.r1	s1c08fs.r1	s1e03fs.r1
s1f06fs.f1	s1f08fs.f1	s1g08fs.f1	s1g08fs.r1
s1g09fs.f1	s2c04fs.r1	s2c06fs.r1	s3b01fs.r1
s3f06fs.r1	s4a03fs.f1	t2b01fs.f1	t2c04fs.f1
t2e10fs.f1	t2h03fs.f1	t2h03fs.r1	t4b08fs.f1
t4b10fs.f1			

Table 4.22 Database members which have both BlastX hits against non-redundant protein (nr) database and tBlastX hits against *N. crassa* EST database (Total: 202)

Contig1003	Contig1006	Contig1007	Contig1012
Contig1014	Contig1018	Contig1019	Contig1034
Contig1036	Contig1040	Contig1050	Contig110
Contig13	Contig136	Contig142	Contig155
Contig182	Contig190	Contig197	Contig203
Contig206	Contig224	Contig232	Contig240
Contig254	Contig258	Contig261	Contig277
Contig289	Contig293	Contig310	Contig319
Contig322	Contig328	Contig349	Contig36
Contig364	Contig375	Contig391	Contig395
Contig415	Contig418	Contig437	Contig438
Contig469	Contig504	Contig518	Contig522
Contig540	Contig542	Contig555	Contig563
Contig565	Contig567	Contig577	Contig629
Contig632	Contig663	Contig680	Contig694
Contig706	Contig708	Contig711	Contig722
Contig731	Contig740	Contig743	Contig770
Contig774	Contig785	Contig800	Contig807
Contig808	Contig827	Contig861	Contig863
Contig866	Contig867	Contig869	Contig876
Contig891	Contig896	Contig898	Contig90
Contig901	Contig903	Contig912	Contig919
Contig921	Contig933	Contig954	Contig958
Contig960	Contig969	Contig97	Contig974
Contig978	Contig980	Contig981	Contig982

Contig998	a1g04fs.r1	a3b09fs.f1	b1a10fs.f1
b1a10fs.r1	b1d05fs.r1	b2d01fs.r1	b3g05fs.r1
b4a06fs.f1	b4d01fs.r1	b4d08fs.f1	b4f06fs.f1
c3d09fs.f1	c3d09fs.r1	c3h08fs.f1	c4d08fs.r1
c4f09fs.f1	c4f09fs.r1	d1c06fs.r1	d1c08fs.r1
d1g06fs.f1	d2h05fs.r1	d3c01fs.r1	e1e09fs.r1
e1g04fs.r1	e2e06fs.r1	e2e10fs.f1	e3b09fs.f1
e3b09fs.r1	e3d05fs.f1	e3g05fs.r1	e4d03fs.r1
e4d07fs.r1	e4e07fs.r1	e4g01fs.f1	f1d06fs.r1
f1h08fs.r1	f3b01fs.f1	f3c07fs.r1	g1e01fs.r1
g2e06fs.r1	g2f12fs.r1	g3a03fs.r1	g3a09fs.r1
g3b02fs.r1	g3f07fs.r1	g4h10fs.r1	h3g06fs.r1
h4h05fs.f1	i1a03fs.r1	i1b05fs.r1	i2d01fs.f1
i2g06fs.r1	i3e12fs.r1	i3g09fs.r1	i4f05fs.r1
i4f06fs.r1	j2h03fs.r1	j3e10fs.r1	k1a05fs.r1
k2c10fs.f1	k2c11fs.f1	k3c06fs.r1	k3c11fs.r1
k4b02fs.r1	k4g11fs.r1	l1a11fs.r1	l1b10fs.f1
l1h11fs.f1	l2a10fs.r1	l3a02fs.f1	l3h04fs.r1
l4c07fs.r1	l4d04fs.r1	m1a10fs.r1	m2f08fs.r1
m4d02fs.f1	m4d03fs.r1	n3d02fs.f1	n3e12fs.r1
o2f03fs.r1	o3d10fs.f1	o4b01fs.r1	o4f04fs.r1
p1b07fs.r1	p1c08fs.f1	p1h04fs.f1	p3f03fs.f1
p4b01fs.f1	p4c04fs.f1	q4d04fs.r1	q4e04fs.r1
r3a11fs.f1	r3g06fs.f1	r4b05fs.f1	r4d09fs.f1
s1c02fs.r1	s2a07fs.r1	t2d01fs.r1	t2f03fs.r1
t4f06fs.f1	t4h02fs.f1		

Table 4.23 Database members which have BlastX hits against non-redundant protein (nr) database and tBlastX hits against both *A. nudulans* and *N. crassa* EST databases (Total: 201)

Contig1001	Contig1008	Contig1009	Contig1010
Contig1013	Contig1022	Contig1027	Contig1029
Contig1035	Contig1037	Contig1049	Contig1051
Contig1052	Contig112	Contig114	Contig131
Contig141	Contig168	Contig183	Contig200
Contig216	Contig222	Contig241	Contig250
Contig259	Contig278	Contig291	Contig3
Contig304	Contig306	Contig334	Contig340
Contig352	Contig353	Contig356	Contig358
Contig371	Contig385	Contig386	Contig388
Contig390	Contig393	Contig403	Contig41
Contig414	Contig45	Contig459	Contig530
Contig533	Contig560	Contig569	Contig571
Contig574	Contig576	Contig589	Contig591
Contig594	Contig610	Contig611	Contig612

Contig615	Contig617	Contig631	Contig637
Contig648	Contig661	Contig669	Contig672
Contig693	Contig705	Contig712	Contig723
Contig727	Contig728	Contig742	Contig748
Contig75	Contig750	Contig753	Contig768
Contig769	Contig776	Contig778	Contig786
Contig787	Contig8	Contig815	Contig820
Contig835	Contig844	Contig846	Contig847
Contig848	Contig865	Contig870	Contig879
Contig88	Contig881	Contig886	Contig887
Contig893	Contig895	Contig904	Contig913
Contig920	Contig923	Contig925	Contig926
Contig937	Contig938	Contig942	Contig943
Contig953	Contig957	Contig959	Contig963
Contig968	Contig972	Contig973	Contig975
Contig977	Contig983	Contig987	Contig990
Contig994	Contig996	Contig999	a1d02fs.f1
a1d02fs.r1	a3h03fs.r1	b1d10fs.f1	b1e03fs.f1
b2e04fs.f1	b2e04fs.r1	b2g04fs.r1	b2h06fs.f1
b3e06fs.r1	b4d01fs.f1	c1g11fs.f1	c3b10fs.r1
d1d10fs.r1	d2h03fs.f1	d3b12fs.r1	e1d05fs.f1
e1d05fs.r1	e2e06fs.f1	e3h05fs.f1	e4a04fs.f1
e4a04fs.r1	e4c02fs.r1	e4g01fs.r1	f3c07fs.f1
f3f09fs.r1	g2e10fs.r1	g3a06fs.r1	g3c04fs.f1
g4a02fs.f1	h1h01fs.f1	i3d01fs.r1	j1e03fs.f1
j1g12fs.f1	j1h10fs.f1	j2c03fs.f1	j2e03fs.r1
j2g02fs.f1	j3a08fs.r1	j4f07fs.r1	j4h02fs.r1
k3f06fs.f1	k3h10fs.r1	k4d02fs.f1	k4d02fs.r1
k4h03fs.r1	l1h11fs.r1	l3c02fs.f1	l3c02fs.r1
l3d11fs.r1	m1f04fs.r1	m1f09fs.r1	m3a08fs.r1
m3a09fs.f1	m3c05fs.r1	m3e09fs.f1	n1f05fs.f1
n1f12fs.r1	n3e02fs.f1	o1d09fs.r1	o1h09fs.f1
o2b03fs.f1	o2c12fs.r1	o2e06fs.f1	o2e06fs.r1
o4c04fs.f1	p3f09fs.r1	p4c11fs.f1	q4f04fs.r1
r3b02fs.r1	r3e06fs.f1	r3h05fs.r1	r4c09fs.f1
s1e05fs.r1	s1e12fs.f1	s1f09fs.f1	s1g06fs.r1
s1h03fs.f1	s3d02fs.r1	s3h06fs.r1	t2e11fs.f1
t4h09fs.f1			

Table 4.24 Summary of tBlastX results

	BlastX hit	**BlastX No hit**
Total in *F.sporotrichioides*	1599	1639
TBlastX hit against *A. nidulus*	550 (34.40%)	80 (4.88%)
TBlastX hit against *N. crassa*	411 (25.70%)	91 (5.55%)

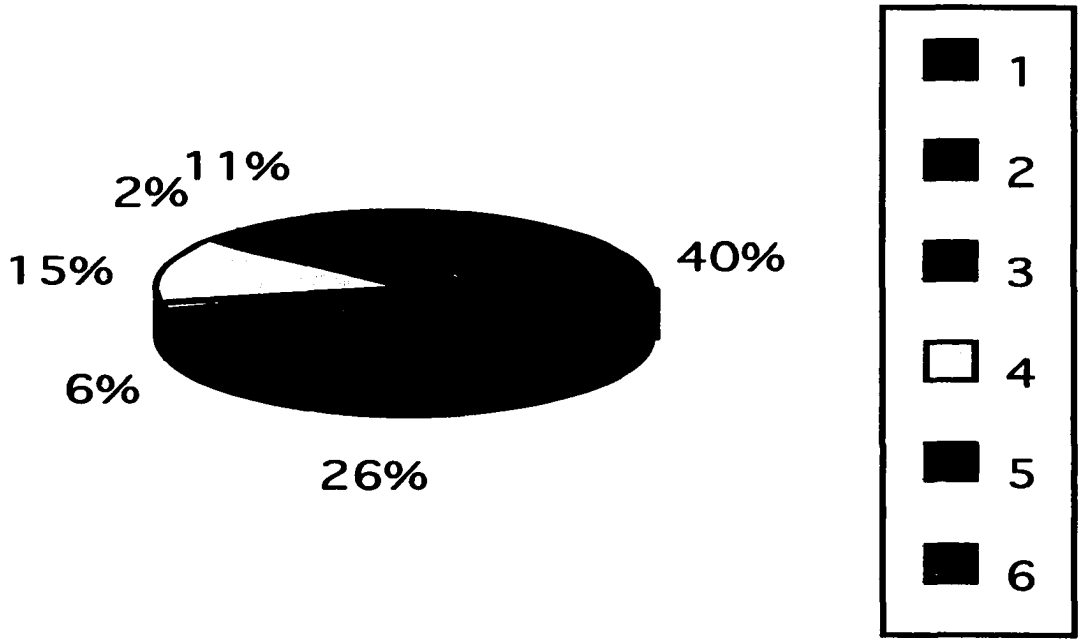
TBlastX hit against both	201 (12.57%)	27 (1.65%)

Table 4.25 The percentage of members falling in each of the biological function categories for the group of ESTs which had BlastX homologs against nr and tBlastX homologs against both *A. nidulus and N. crassa* databases

Metobolic Pathway	38.6%
Genetic information process	26.2%
Cell growth	5.94%
Process	15.3%
Unclassified	2.4%
Unidentified	11.4%

Table 4.26 The report containing "GenBank_Accn" for submitted cDNA

GTP-binding protein cDNA	AY032742
cytochrome p450 cDNA	AY032743
Tri8cDNA	AY032744
Tri5cDNA	AY032745
Tri14cDNA	AY032746
Tri101cDNA	AY032747



1. Metabolic pathway
2. Genetic information processing
3. Cell growth
4. Processing
5. Unclassified
6. Unidentified

Fig. 4.4 tBiastX homologues against both *A. nidulans* and *N.crassa* EST databases

Chapter V

Analysis of contigs of human genome sequences

5.1 Analysis of 322f3

A deletion (R271) of about 200 nucleotides originally was detected in a kidney tumor by representational differential analysis (RAD) and the deletion has been mapped to 22q11.2 by florescent *in situ* hybridization studies (section 1.8). Since a search for the complete sequence of the undeleted region in the human chromosome 22 reference sequence did not yield positive results (Marcello et al., 2000), BAC 322f3 subsequently was sequenced completely and deposited into GenBank (AC009286). A BlastX search of the Genbank database of this 121,871 bp sequence revealed the presence of several putative genes based on cDNA and EST similarities and the results from computational gene prediction programs.

5.1.1 an additional 9 kb region contained in BAC 322f3

Figure 5.1 is a diagrammatic sketch of chromosome 22 showing from left to right the region 22q11.2, the Igλ V and IgλJ&C gene regions, the region containing clone D86999 sequenced by Kawarated et al., (1997) and the newly sequenced clone BAC 322f3 (AC009286), with numbers indicating the nucleotide position. The regions of complete homology between the two clones are indicated in black and highlighted by dotted lines. The 9 kb sequence between 24803 and 33901 in clone AC009286 is absent in clone D86999. Clone AC009286 and the corresponding GenBank human chromosome 22 concensus contig NT_001454 shows a partial homology (84%) between the two bracketed fragments located on either side of the deletion. The nucleotide sequence

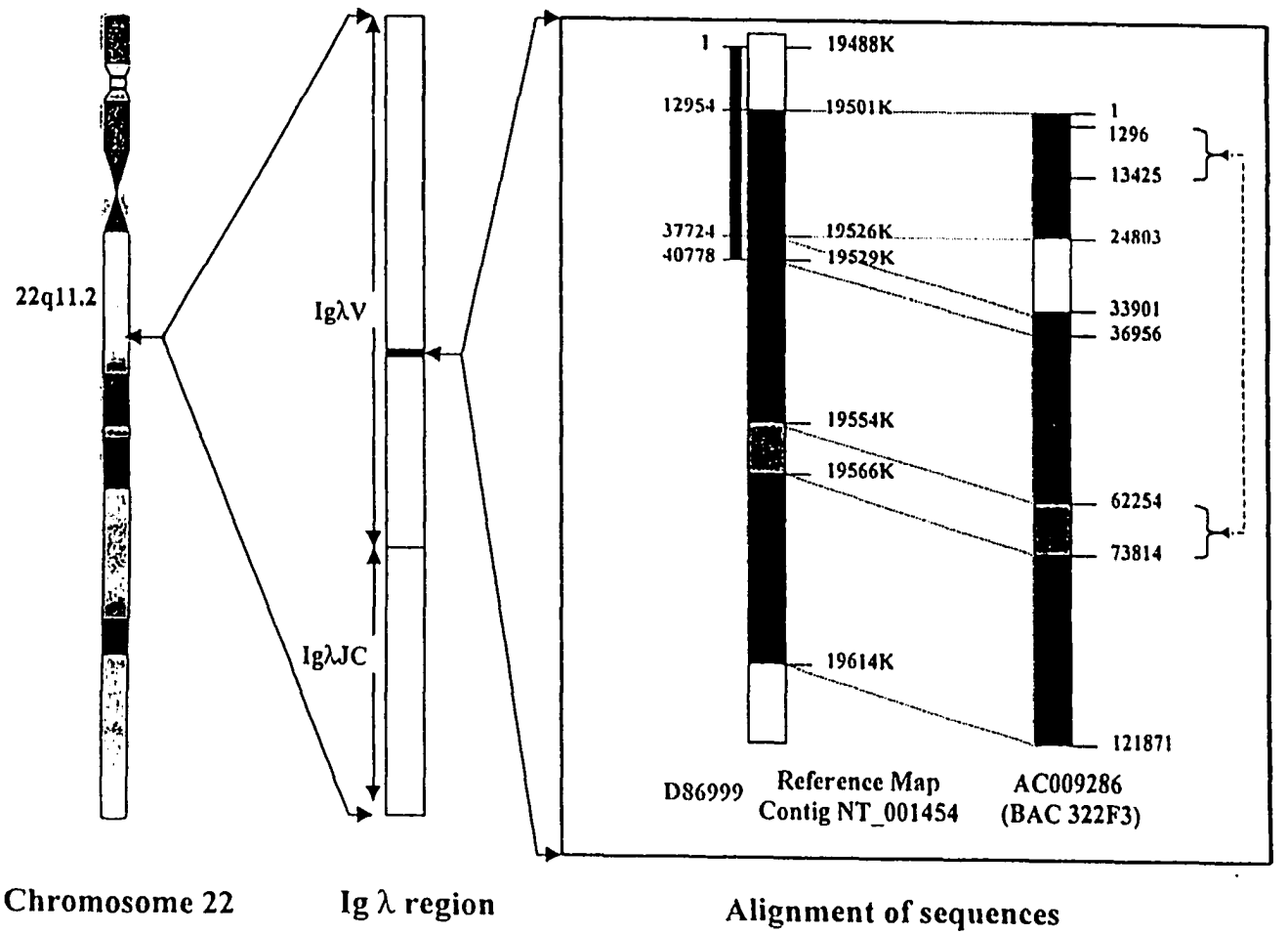


Fig. 5.1 Diagrammatic sketch of chromosome 22 showing from left to right the region 22q11.2, the Igλ V and IgλJ&C gene regions, the region containing clone D86999 and the newly sequenced clone AC009286.

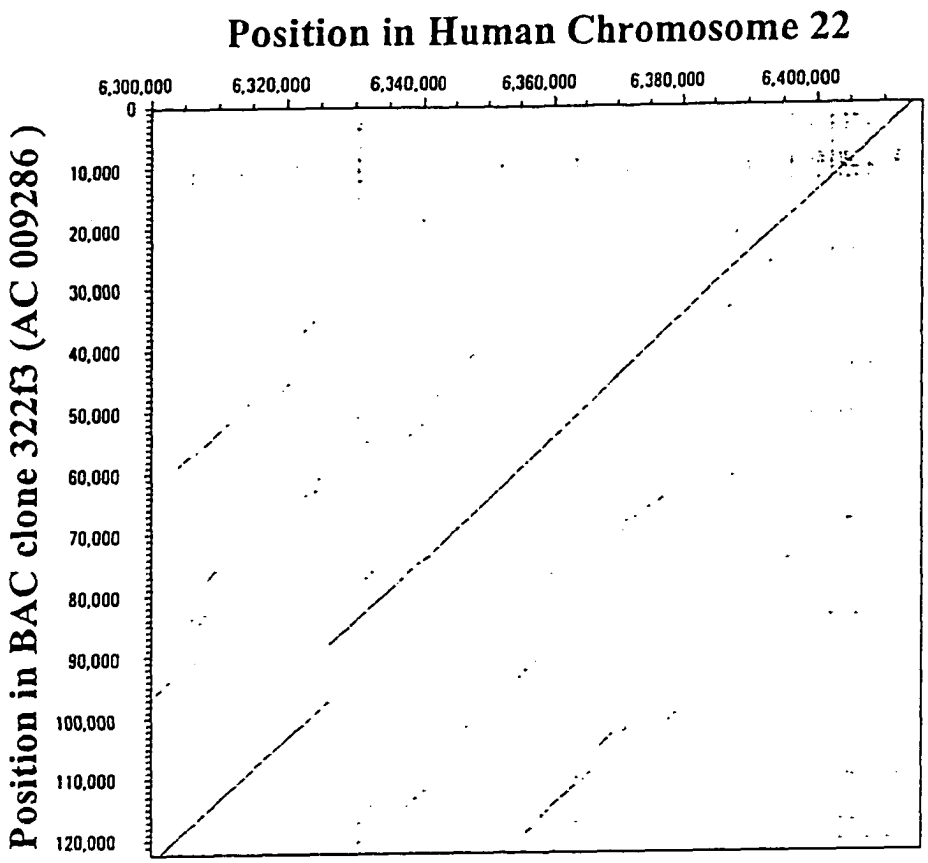


Fig. 5.2 Dotter plotting comparison between BAC 322f3 (AC009286) and the human chromosome 22 reference sequence (NT_0011454)

numbering for BAC 322f3 is derived from the sequence reported in GenBank for accession number AC009286 on July 26, 2000.

Figure 5.2 is a Dotter (section 1.6.2) comparison between BAC 322f3 (AC009286) and the human chromosome 22 reference sequence (NT_001454, Dunham et al., 1999). The deleted region is observed in the interrupted diagonal which otherwise indicates almost complete homology between the two sequences. Each dot in figure 5.2 represents an identity of 14 nucleotides in both sequences over a contiguous in a window of 20 nucleotides.

5.1.2 Repeat elements in AC009286

A RepeatMasker (section 1.6.6) analysis of the repetitive elements in AC009286 (BAC 322f3) revealed 46.86% of this 121,871 bp region contained known repeat elements and has an overall G+C content of 44.14% (Table 5.1). Figure 5.3 shows the positions of the most prevalent repeat elements in this sequence as plotted by the Geneliner program (Hua, 1999). Alu repeat elements, L1 repeat elements and LTR repeat elements are the main repeat elements present in the amount of 9.33%, 15.54% and 18.36% each respectively.

Table 5.1 Repetitive elements in AC009286 (BAC 322f3)*

Total length: 121871 bp
GC level: 44.14 %
Repeat bases masked: 57110 bp (46.86 %)

repeat element types	number of elements*	length occupied	percentage of sequence
SINEs:	45	11582 bp	9.50 %
ALUs	43	11369 bp	9.33 %
MIRs	2	213 bp	0.17 %
LINEs:	27	20324 bp	16.68 %

LINE1	21	18935 bp	15.54 %
LINE2	6	1389 bp	1.14 %
LTR elements:	17	22375 bp	18.36 %
MaLRs	3	1381 bp	1.13 %
ERVL	3	6003 bp	4.93 %
ERV_classI	8	12941 bp	10.62 %
ERV_classII	3	2050 bp	1.68 %
DNA elements:	6	675 bp	0.55 %
MER1_type	3	254 bp	0.21 %
MER2_type	1	243 bp	0.20 %
Satellites:	1	244 bp	0.20 %
Simple repeats:	9	305 bp	0.25 %
Low complexity:	8	311 bp	0.26 %
Unclassified:	1	1304 bp	1.07 %
Total interspersed repeats:		56260 bp	46.16 %

*RepeatMasker version 08/14/2000, default mode
run with cross_match version 0.990319
RepBase version 06/31/2000

5.1.3 Gene prediction in AC009286

Genscan predicts 3 genes in this segment of human chromosome 7 (Fig. 5.3). A Powerblast search against GenBank nr and dbEST revealed several cDNA and EST homologues (Fig. 5.3), but the similarity percentages are low. Upon Sim4 (section 1.6.2) alignment, a 95%-100% identity was observed to three putative mRNAs predicted by GenomeScan, representing human pre-B lymphocyte gene 1 mRNA (gi|10879160|XM_000806, gi|10879152|XM_000802) in the region 45466-71752, human immunoglobulin lambda joining 3 mRNA (gi|10879150| XM_000801) at position 39908-51013, and a human immunoglobulin kappa chain V region S211 mRNA (gi|10879156| XM_000804) at position 2565-14969. No other genes are encoded in 322f3.

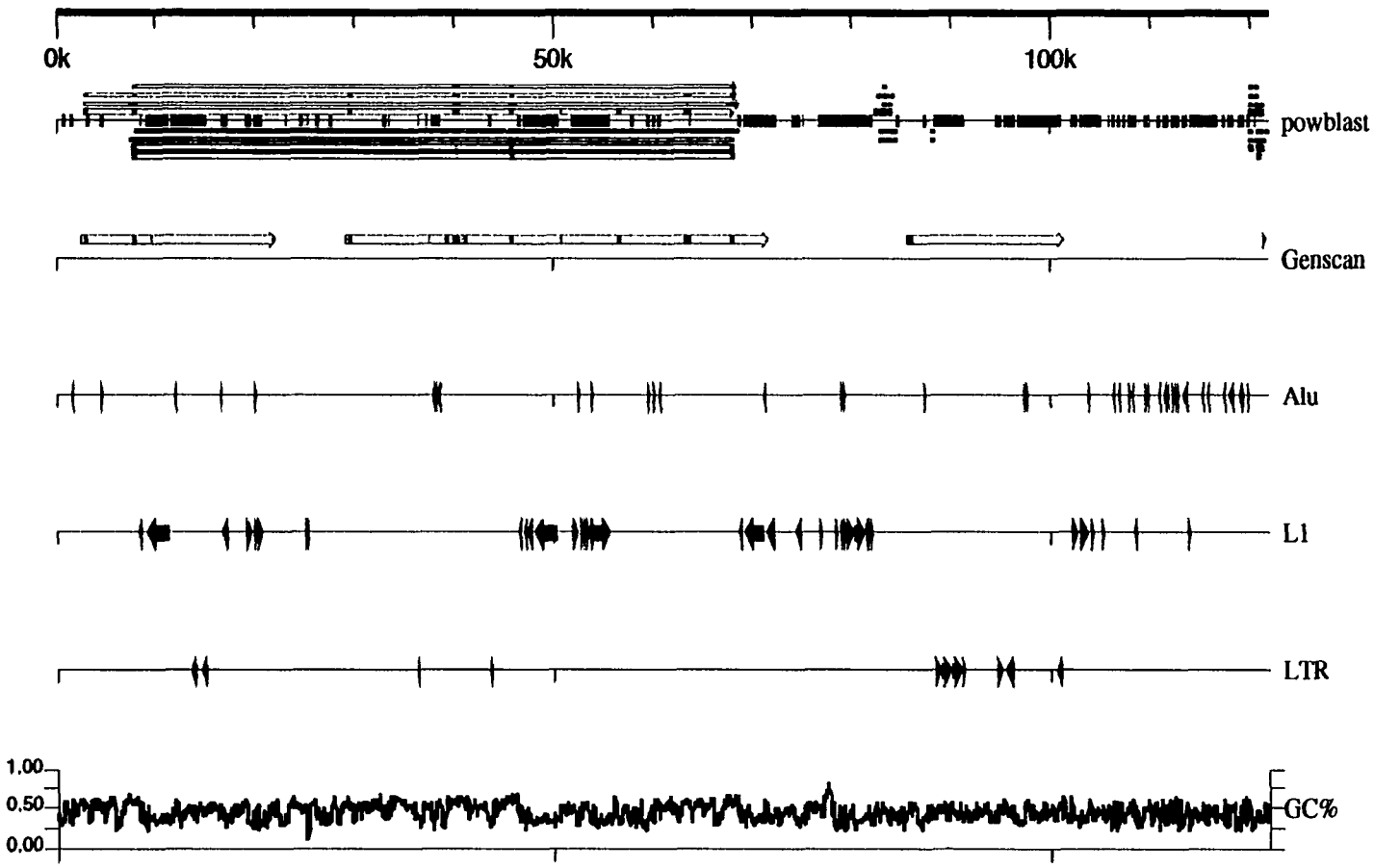


Fig. 5.3 Gene prediction, repeat elements and GC level in AC009286

5.1.4 Significance of discovering the phenotypically silent inborn gaps

Sequence of BAC 322f3 revealed the sequence of the 9098 bp absent in the Ig λ region of the reference sequence for human chromosome 22. Although this region has varying deletions (Robledo et al.), the most prevalent deletions can be quantitated by the R271 (Table 1.2) PCR probes. As seen in Fig. 1.11, pre_Dravidian Indians, Amazonians and Melanesians almost totally lack this 9kb region and therefore has shorter intron region between two putative genes, human immunglobin kappa chain V region S211 gene and human immunoglobulin lambda joining 3 genes. Two segments within this contig have 84% homology and are located a few thousand kb away from the ends of the 9 kb deletion (Fig. 5.2), one of them contains the human immunoglobulin lambda joining 3 pseudogene. Since the immunoglobulin gene rich region is known to contain a high number of duplicated genes and pseudogenes (Kawasaki et al., 1997), it may be that this R271 related deletion resulted from inaccurate cross over during replication (Siniscalco et al., 2000).

The observation that the deletions occurred more frequently among three tribal populations living in forest habitats (Fig. 1.11), points to the possibility that individuals who are simple or compound heterozygotes for these deletions may have a higher biological fitness in stressful ecological environments (Siniscalco et al., 2000). Since survival in a forest habitat is presumably a function of a well working immunological system, the overlapping of the described deletion polymorphism with the Ig λ light chain genes may be more than just a coincidence.

Numerous instances of deletions have been observed in human chromosome 22 that are associated with an abnormal phenotype, such as DiGeorge syndrome (Chen,

1997; Zhang, 1997; Gong et al., 1997; Kimber et al., 1999), Cat Eye Syndrome (Hough et al., 1995; Crabtree, 1997; Johnson et al., 1999) and Meningioma (Pan, 1996; Ruttledge et al., 1994; Metzger et al., 1999) as well as the recent report of the absence of a functional gene (*CYP2D6*5*), gene which related to various cancers and Parkinson's diseased (Idle et al., 2000). Evan so, the deletion polymorphism described here is the first example of a very common set of DNA deletion of relatively large size found among normal adults and seemingly associated with higher biological fitness.

5.2 Analysis of 239c10

The 275,197 bp BAC 239c10 that contains a genomic DNA insert from the WMS region in human chromosome 7 encodes at least three genes, Human neutrophil cytosol factor 1 (NCF1) gene, human hPMS gene, human Bruton's tyrosine kinase-associated protein-135 (BAP-135) gene and a human prohibitin pseudogene. Each of these genes and their related gene products will be discussed in detail below. In addition, several other cDNA and EST homologues have been found with homology in this region indicating the possible presence of additional genes or pseudo genes (Fig. 5.4).

5.2.1 Human neutrophil cytosol factor 1 (NCF1) gene

The human neutrophil cytosol factor 1 (NCF1, NCF-47k) is required for activation of nicotinamide adenine dinucleotide phosphate (NADPH) oxidase (Volpp et al., 1989). NADPH oxidase is required for generate superoxide anion ($O_2^{\cdot}$) that is converted to hydrogen peroxide and other microbicidal oxygene products in an activated phagocytic cells. NCF1 is absent in most patients with autosomal recessive chronic

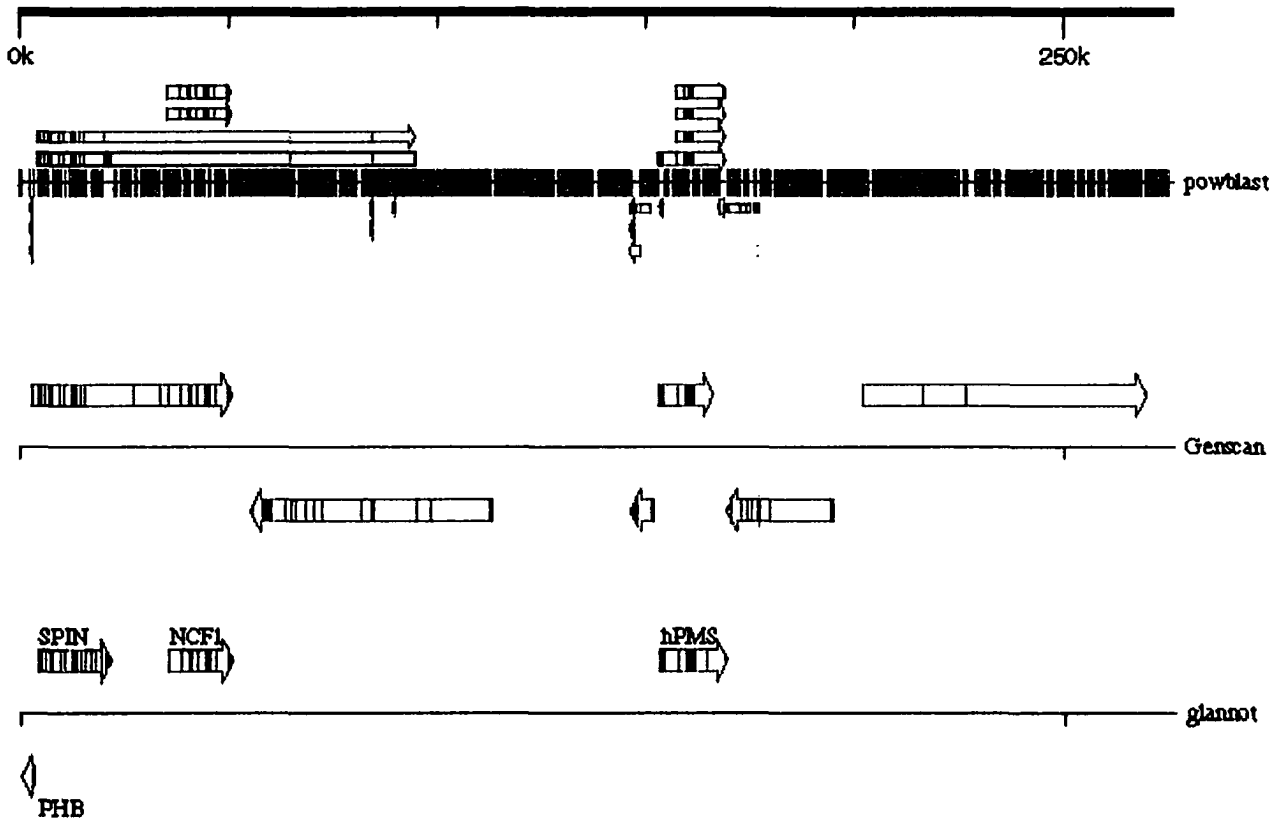


Fig. 5.4 Gene prediction and annotation of AC004166

granulomatous disease (AR-CGD) (Nunoi et al., 1988) and a recombinant NCF1 is available which restores NADPH oxidase in chronic granulomatous disease (Lomax et al., 1989).

The NCF1 gene was cloned and sequenced from the expression libraries by two different laboratories in the same year (Lomax et al., 1989; Volpp et al., 1989; gi|189107|HUMNCF1). The mRNA is 1339 bp long and has a functional expression of the 47-kilodalton protein which has 390 amino acids. In 1990, Rodaway et al. sequenced this gene again and submitted it to GenBank with the name "Human 47-kd autosomal chronic granulomatous diseases protein mRNA" in another entry (Rodaway et al., 1990; gi|189050|HUMANDPHO). A Bestfit alignment of HUMNCF1 and HUMANDPHO reveals that both genes have exact the same nucleotide sequences except that HUMANDPHO has 10 more bases at 5' end of untranslation region (5'UTR). The only discrepancy among the well-aligned bases is that an A on position 399 of HUMNCF1 is substituted for a G in HUMANDPHO. This discrepancy likely represents polymorphism (Volpp et al., 1989).

A Sim4 (Florea et al., 1998) alignment of the NCF1 mRNA (gi|189107|) and genomic sequence AC004166 shows that there are 11 exons spanning a region of about 15 kb(Fig. 5.5, Table 5.2). The ranges of exons and the exon and intron sizes of the NCF1 gene are listed in Table 5.2. The sizes of introns range from 101 bp to 3194 bp and the sizes of exons range from 56 bp to 276 bp. The average similarity between NCF1 gene mRNA and AC004166 is more than 99%.

The alignment of the NCF1 cDNA with the genomic sequence reveals seven discrepancies, at positions 40859 and 45013 where a G in genomic sequence is

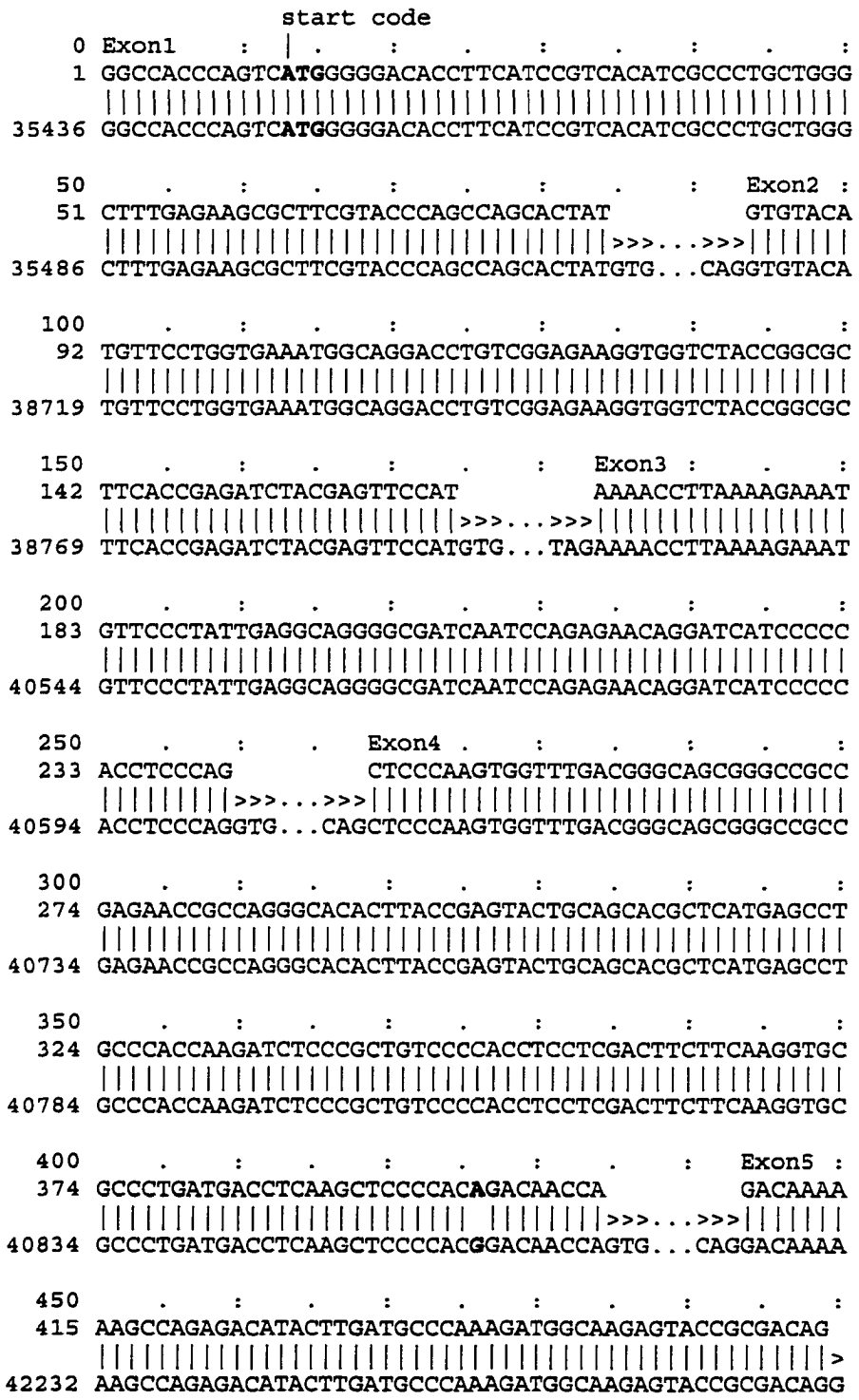
substituted for an A in the mRNA sequence, at positions 44425, 44487 and 49475 where an A is substituted for a G in the mRNA sequence and at positions 49451 and 50032 where a C is substituted for a T in the mRNA sequence. These mismatches do not change the coded amino acids for NCF1 except at position 44425 where an Asn is substituted for an Asp. All the seven discrepancies were noted by Volpp et al. (Volpp et al., 1989) and were classified as polymorphisms. Lomax et al. also noticed some of these discrepancies as well (Lomax et al., 1989).

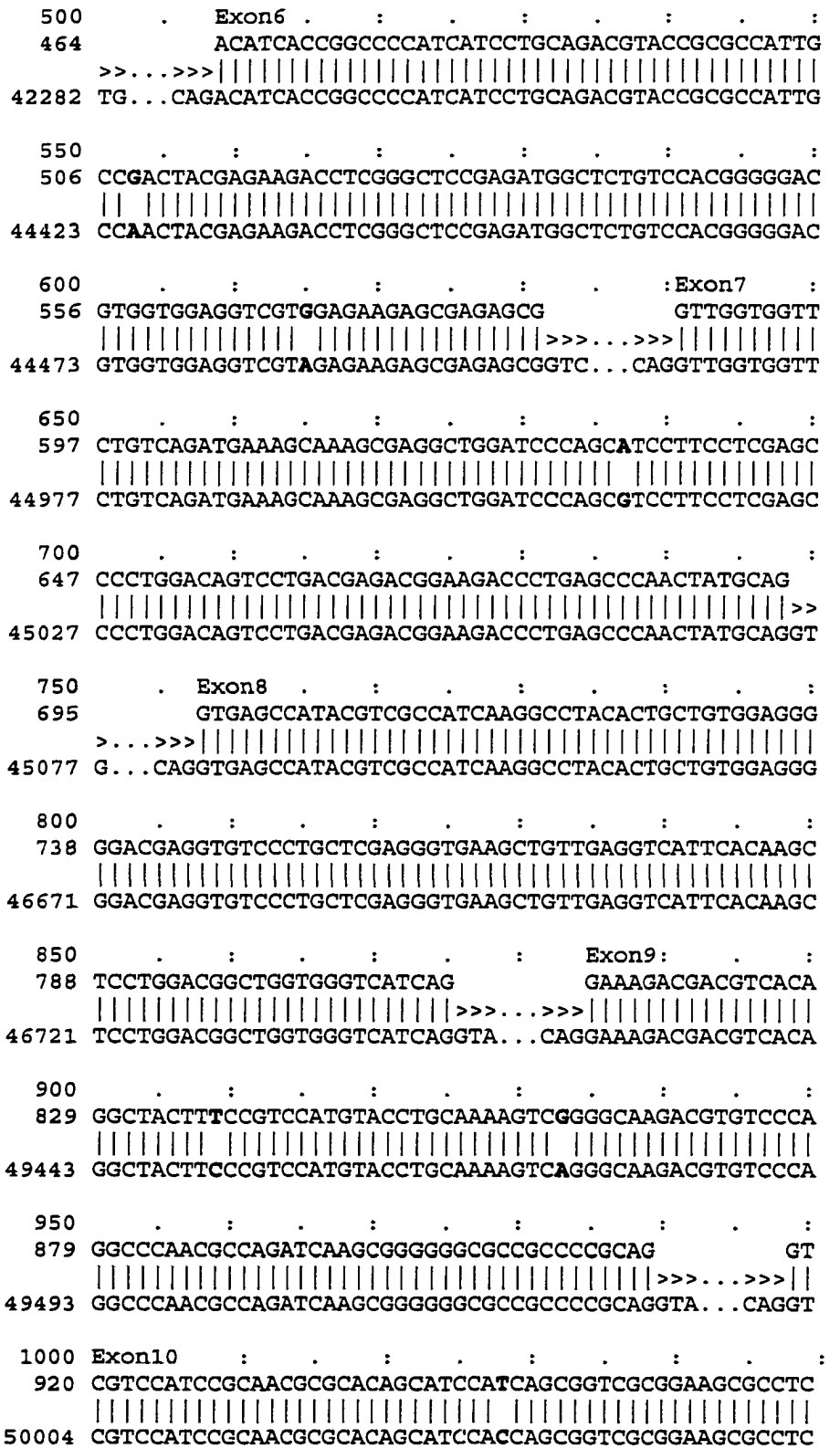
All of the introns of the NCF1 gene begin at GT and end at (T/C)AG. Translation begins at position 13 and ends at 1185 in the mRNA, which corresponds to positions 35448 and 60603 in the sequenced genomic DNA of BAC 239C10. The amino acid sequence of the translated NCF-47k protein is showed at Table 5.3.

Table 5.2 The ranges and sizes of exons and introns of NCF1 gene

No.	cDNA range	Exon range (239c10)	Exon size(bp)	Intron size(bp)	Similarity
1	1-84	35436-35519	84	3194	100%
2	85-165	38712-38792	81	1736	100%
3	166-241	40527-40602	76	101	100%
4	242-407	40702-40867	166	1359	99%
5	408-463	42225-42280	56	2102	100%
6	464-586	44381-44503	123	465	98%
7	587-694	44967-45074	108	1555	99%
8	695-812	46628-46745	118	2683	100%
9	813-917	49427-49531	105	472	98%
10	918-1063	50002-50147	146	336	99%
11	1064-1339	50482-50757	276		100%

seq1 = NCF1.fa (HUMNCF1A), 1339 bp,
seq2 = 239c10.fa (AC004166), 275197 bp




500 . Exon6 . : . : . : . :
464 ACATCACCGGCCCCATCATCCTGCAGACGTACCGCGCCATTG
>>...>>>
42282 TG...CAGACATCACCGGCCCCATCATCCTGCAGACGTACCGCGCCATTG
550 . : . : . : . : . :
506 CCGACTACGAGAAGACCTCGGGCTCCGAGATGGCTCTGTCCACGGGGGAC
44423 CCAACTACGAGAAGACCTCGGGCTCCGAGATGGCTCTGTCCACGGGGGAC
600 . : . : . : . :Exon7 :
556 GTGGTGGAGGTCGTGGAGAAGAGCGAGAGCG GTTGGTGGTT
>>>...>>>
44473 GTGGTGGAGGTCGTAGAGAAGAGCGAGAGCGGTC...CAGGTTGGTGGTT
650 . : . : . : . : . :
597 CTGTCAGATGAAAGCAAAGCGAGGCTGGATCCCAGCATCCTTCCTCGAGC
44977 CTGTCAGATGAAAGCAAAGCGAGGCTGGATCCCAGCGTCCTTCCTCGAGC
700 . : . : . : . : . :
647 CCCTGGACAGTCCTGACGAGACGGAAGACCCTGAGCCCAACTATGCAG
>>
45027 CCCTGGACAGTCCTGACGAGACGGAAGACCCTGAGCCCAACTATGCAGGT
750 . Exon8 . : . : . : . :
695 GTGAGCCATACGTCGCCATCAAGGCCTACACTGCTGTGGAGGG
>...>>>
45077 G...CAGGTGAGCCATACGTCGCCATCAAGGCCTACACTGCTGTGGAGGG
800 . : . : . : . : . :
738 GGACGAGGTGTCCCTGCTCGAGGGTGAAGCTGTTGAGGTCATTCACAAGC
46671 GGACGAGGTGTCCCTGCTCGAGGGTGAAGCTGTTGAGGTCATTCACAAGC
850 . : . : . : Exon9: . :
788 TCCTGGACGGCTGGTGGGTCATCAG GAAAGACGACGTCACA
>>>...>>>
46721 TCCTGGACGGCTGGTGGGTCATCAGGTA...CAGGAAAGACGACGTCACA
900 . : . : . : . : . :
829 GGCTACTTTCCGTCCATGTACCTGCAAAAGTCGGGGGCAAGACGTGTCCCA
49443 GGCTACTTCCCGTCCATGTACCTGCAAAAGTCAGGGCAAGACGTGTCCCA
950 . : . : . : . : . :
879 GGCCCAACGCCAGATCAAGCGGGGGGCGCCGCCCCGCAG GT
>>>...>>>
49493 GGCCCAACGCCAGATCAAGCGGGGGGCGCCGCCCCGCAGGTA...CAGGT
1000 Exon10 : . : . : . : . :
920 CGTCCATCCGCAACGCGCACAGCATCCATCAGCGGTCGCGGAAGCGCCTC
50004 CGTCCATCCGCAACGCGCACAGCATCCACCAGCGGTCGCGGAAGCGCCTC

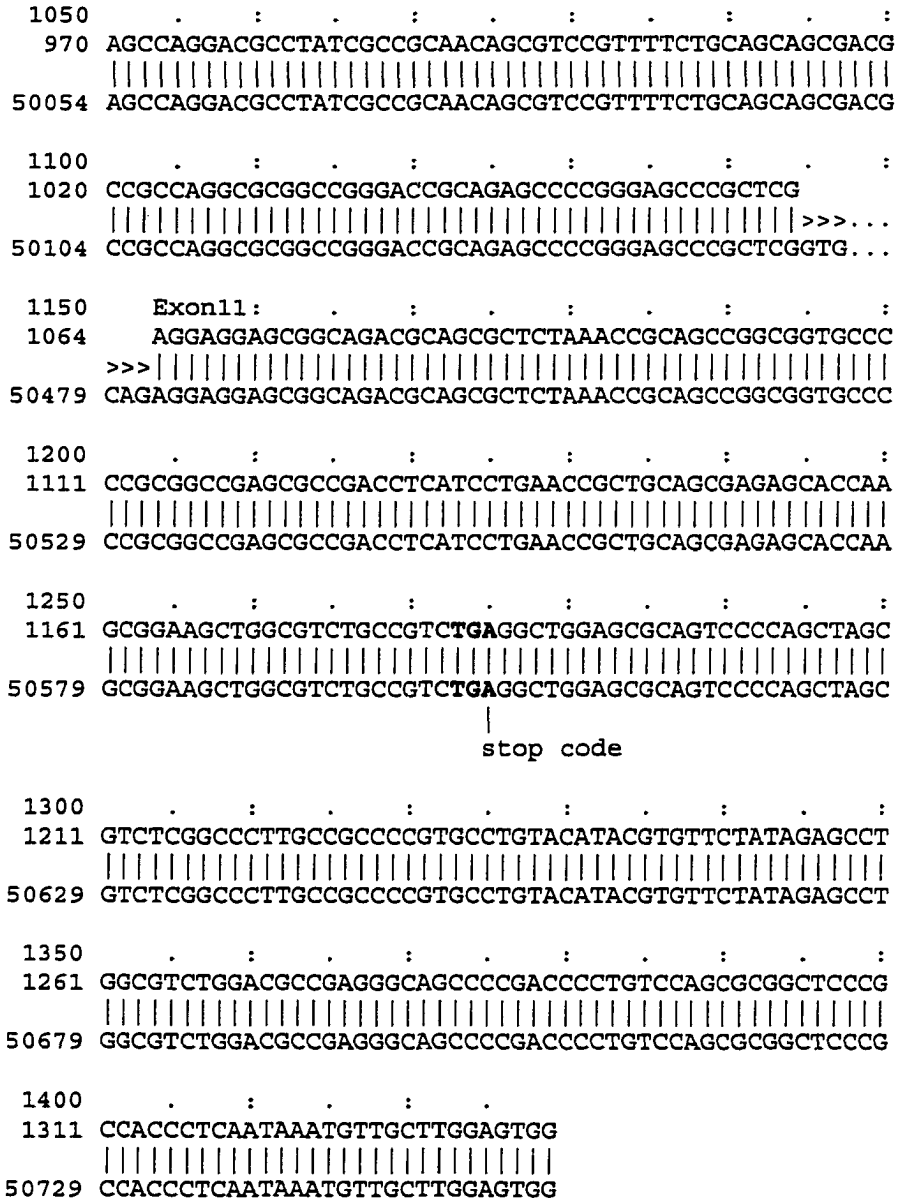


Fig. 5.5 Sim4 alignment of human NCF1 cDNA and genomic sequence AC004166 (Start and stop codons and discrepancies are highlighted in bold)

Table 5.3 Amino acid sequence of NCF1 (protein_id=AAA57209.1)

MGDTFIRHIALLGFEKRFVPSQHYVYMFLVKWQDLSEKVVYRRFTEIYEFHKTLKEMF
PIEAGAINPENRIIPHLPAPKWFDGQRAAENRQGTLTEYCSTLMSLPTKISRCPHLLD
FFKVRPDDLKLPTDNQTKKPETYLMPKDGKSTATDITGPIILQTYRAIADYEKTSGSE
MALSTGDVVEVVEKSESGWWFCQMKAKRGWIPASFLEPLDSPDETEDPEPNYAGEPYV
AIKAYTAVEGDEVSLLEGEAVEVIHKLLDGWWVIRKDDVTGYFPSMYLQKSGQDVSQA
QRQIKRGAPPRRSSIRNAHSIHQRSRKRLSQDAYRRNSVRFLQQRRRQARPGPQSPGS
PLEEERQTQRSKPQPAVPPRPSADLILNRCSESTKRKLASAV

Regulatory element: The promoter prediction program Neural Network Promoter Prediction (NNPP) (section 1.6.6) predicts a promoter region about 774 bp upstream the first exon at contig positions 34612-34662. Another program Promoter 2.0 (section 1.6.6) predicts a promoter 700 upstream the first exon. The first exon resides in a CpG island (35343-35520) as predicted by Xgrail. However, no TATA box was found within 1 kb upstream of the first exon. The 3' exon of NCF1 gene encodes a 154 bp 3' UTR that is GC-rich (67% GC bases). A putative polyadenylation signal (AATAAA) is found 21 bp upstream the end of last 3' exon, at positions 50737-50742. A second polyadenylation signal (GT rich segment) is found 71 bp downstream of the last 3' exon beginning at position 50828.

5.2.2 Human *hPMS* gene

5.2.2.1 Human *hPMS* gene family

In yeast, it was found that disruptions on any one of the three genes (*MSH2*, *MLH1* and *PMS1*) involved in DNA mismatch repair would lead to high degree of mutation rate (Strand et al., 1989). It was shown that defects in genes involved in the DNA mismatch repair system increased the rate for mutation in oncogenes and tumor suppressor genes, which might trigger a normal cell to transform into a cancer cell (Cleaver, 1994). Human homologues of yeast *MSH2* and *MLH1*, *hMSH2* and *hMLH1*, were cloned and mutations in those two genes were found in patients in some hereditary nonpolyposis colorectal cancer (HNPCC) families (Fishel et al., 1993; Leach et al., 1993; Bronner et al., 1994; Papadopoulos et al., 1994). As the candidate genes responsible for a part of HNPCC families, human homologues of yeast *PMS1* also were cloned and located on chromosomal band 7q11.23 and 7q22 by fluorescent *in situ* hybridization (FISH)

(Horii et al., 1994, Papadopoulos et al., 1994). Up to now, at least 11 genes of human *PMS* gene family (Table 5.4) were identified.

Table 5.4 Members in human *PMS* gene family

Human homolog of yeast mutL (*hPMS1*) gene, complete cds
gi|535512|gb|U13695.1|HSU13695[535512]
Homo sapiens hPMS1 gene, promoter region and exon 1
gi|2970045|dbj|AB006462.1|AB006462[2970045]
Human homolog of yeast mutL (*hPMS2*) gene, complete cds
gi|535514|gb|U13696.1|HSU13696[535514]
Human DNA mismatch repair gene homologue (*hPMS2*) mRNA, complete cds
gi|557469|gb|U14658.1|HSU14658[557469]
Homo sapiens *PMS2L13* mRNA, partial cds
gi|4239949|dbj|AB017004.1|AB017004[4239949]
Homo sapiens *hPMS3* mRNA, partial cds
gi|600590|dbj|D38435.1|D38435[600590]
Homo sapiens *hPMS4* mRNA, partial cds
gi|600591|dbj|D38436.1|HUMMRB[600591]
Human PMS4 mRNA, partial cds (C-terminal region)
gi|559337|dbj|D38502.1|HUMPMS1E[559337]
Homo sapiens *hPMS5* mRNA, partial cds
gi|600592|dbj|D38437.1|HUMMRC[600592]
Homo sapiens *hPMS6* mRNA, partial cds
gi|600593|dbj|D38438.1|HUMMRD[600593]
Homo sapiens *hPMS7* mRNA, partial cds
gi|600594|dbj|D38439.1|HUMMRE[600594]
Homo sapiens *hPMS8* mRNA, partial cds
gi|600595|dbj|D38440.1|HUMMRF[600595]

5.2.3.2 239c10 has homologues to three members in HMS gene family

239c10 are homologous to four GenBank entries in HMS gene family, hPMS4 (gi|600591 and gi|559337), hPMS2L13 (gi|4239949) and hPMS7 (gi|600594) (Fig. 5.3). The GCG Pileup program was used to compare the four mRNA sequences from GenBank (Fig. 5.6). All of the four mRNA sequences only contain the region that codes for the C-terminal protein of the protein. All three, PMS4, hPMS4 and PMS2L13, have a

stop codon at the same position (1170) while fourth one has a stop codon at position 1075 (Fig. 5.6). All four mRNA contain at least 1000 bp of identical sequence.

The human hPMS4 gene has two entries in the GenBank, gi|600591|HUMMRB and gi|559337|HUMPMS1E. These two entries were submitted by the same group of authors and the mRNA sequences of the two entries are exactly the same (Fig 5.5). Therefore, they are the same gene. The mRNA of hPMS4 is 1151 bp long and 760 bp at its 5' end codes for a putative translation product of 252 amino acids which represents the C-terminal region of the protein. The remaining 391 bp at 3' end is 3' untranslate region (3'UTR). A Sim4 alignment of the hPMS4 mRNA and the genomic sequence AC004166 shows that there are 7 exons spanning a region of about 15 kb (Table 5.5).

A Sim4 alignment of the hPMS2L13 mRNA and the genomic sequence AC004166 (Table 5.6) shows that hPMS4 gene and hPMS2L13 gene share five exons and six introns that have exactly the same range on the genomic DNA (Table 5.5 and Table 5.6). Gene hPMS2L13 has one more 5' exon than hPMS4.

A Sim4 alignment of the hPMS7 mRNA and the genomic sequence AC004166 (Table 5.7) shows that hPMS4 gene and hPMS7 gene share three exons and three introns that have exactly the same range on the genomic DNA (Table 5.6 and Table 5.7). The hPMS7 gene has one fewer exon than hPMS4 at 5' end, one fewer exon than both hPMS4 and hPMS2L13 in the middle of coding region, and one more exon than both hPMS4 and hPMS2L13 representing the 3'end untranslated region (3' UTR).

Regulatory element: The promoter prediction program NNPP (section 1.6.6) predicts three promoter regions at 765, 876 and 1234 bp upstream the first exon at contig positions 153055, 152944 and 152586 respectively. The Promoter 2.0 program predicts

Table 5.5 The ranges and sizes of exons and introns of the hPMS4 gene

No.	mRNA range	Exon range	Exon (bp)	Intron (bp)	similarity
1	(sequence is unknown in mRNA)				
2	1-60	153821-**153880***	60	**3269**	95%
3	61-200	**157148-157287**	**140**	**1892**	99%
4	201-287	**159178-159264**	**87**	**178**	99%
5	288-390	**159441-159543**	**103**	**1141**	98%
6	391-574	**160683-160866**	**184**	**3054**	99%
7	575-663	**163919-164007**	**89**	**4592**	100%
8	664-1151	**168598**-169083	488		100%

*Bold numbers indicates numbers that are the same in Table 5.5 and Table 5.6

Table 5.6 The ranges and sizes of exons and introns of the hPMS2L13 gene

No.	mRNA range	Exon range	Exon (bp)	Intron (bp)	similarity
1	(sequence is unknown in mRNA)				
2	1-106	153337-153442	106	75	97%
3	107-470	153517-**153880***	364	**3269**	96%
4	471-610	**157148-157287**	**140**	**1892**	96%
5	611-697	**159178-159264**	**87**	**178**	99%
6	698-800	**159441-159543**	**103**	**1141**	98%
7	801-984	**160683-160866**	**184**	**3054**	96%
8	985-1073	**163919-164007**	**89**	**4592**	96%
9	1074-1244	**168598**-168768	170		99%

*Bold numbers indicates numbers that are the same in Table 5.5 and Table 5.6

Table 5.7 The ranges and sizes of exons and introns of the hPMS7 gene

No.	mRNA range	Exon range	Exon (bp)	Intron (bp)	similarity
1	(sequence is unknown in mRNA)				
2	1-131	157157-**157287***	131	**1892**	100%
3	132-218	**159178-159264**	**87**	1419	100%
4	219-397	**160683-160866**	179	**3054**	97%
5	398-485	**163919-164007**	**89**	**4592**	98%
6	486-568	**168598**-168680	83	87	100%
7	569-1078	168707-169216	510		99%

*Bold numbers indicates numbers that are the same in Table 5.5 and Table 5.7

```
Pms = gi|559337|HUMPMS1E, Human PMS4 mRNA
Hpm4 = gi|600591|HUMMRB, Homo sapiens hPMS4 mRNA
Pms2 = gi|4239949|AB017004, Homo sapiens PMS2L13 mRNA
Hpms = gi|600594|HUMMRE, Homo sapiens hPMS7 mRNA

      1                                                            50
 Pms  ---------- ---------- ---------- ---------- ----------
Hpm4  ---------- ---------- ---------- ---------- ----------
Pms2  GCCGCTCCTG CCGTGCATGT TGGGGAGCCA GTACATGCAG GTGGGCTCCA
Hpms  ---------- ---------- ---------- ---------- ----------

      51                                                           100
 Pms  ---------- ---------- ---------- ---------- ----------
Hpm4  ---------- ---------- ---------- ---------- ----------
Pms2  CACGGAGAGG GGCGCCGACC CCGTGATAGG GCTTTACCTG GTACATCGGG
Hpms  ---------- ---------- ---------- ---------- ----------

      101                                                          150
 Pms  ---------- ---------- ---------- ---------- ----------
Hpm4  ---------- ---------- ---------- ---------- ----------
Pms2  GTGGCGCGTG CCAGACACCA ACGGTCGGAA ACCGCCAGAC ACCAACGCTC
Hpms  ---------- ---------- ---------- ---------- ----------

      151                                                          200
 Pms  ---------- ---------- ---------- ---------- ----------
Hpm4  ---------- ---------- ---------- ---------- ----------
Pms2  GGAATCCACG CCAGGCCACG ACGGAGGGCG ACTACCTCCC TTCTGACCCT
Hpms  ---------- ---------- ---------- ---------- ----------

      201                                                          250
 Pms  ---------- ---------- ---------- ---------- ----------
Hpm4  ---------- ---------- ---------- ---------- ----------
Pms2  GCTGCTGGCG TTCGGAAAAA ACGCAGTCCG GTGTGCTCTG ATTGGTCCAG
Hpms  ---------- ---------- ---------- ---------- ----------

      251                                                          300
 Pms  ---------- ---------- ---------- ---------- ----------
Hpm4  ---------- ---------- ---------- ---------- ----------
Pms2  GCTCTTTGAC GTCACGGACT CGACCTTTGA CAGAGCCACT AGGCGAAAAG
Hpms  ---------- ---------- ---------- ---------- ----------

      301                                                          350
 Pms  ---------- ---------- ---------- ---------- ----------
Hpm4  ---------- ---------- ---------- ---------- ----------
Pms2  GAGAGACGGG AAGTATTTTT TCCGCCCCGC CCGGAAAGGG TGGAGCACAA
Hpms  ---------- ---------- ---------- ---------- ----------

      351                                                          400
 Pms  ---------- ---------- ---------- ---------- ----------
Hpm4  ---------- ---------- ---------- ---------- ----------
Pms2  CGTCGAAAGC AGCCGTTGGG AGCCCAGGAG GCGGGGCGCC TGTGGGAGCC
Hpms  ---------- ---------- ---------- ---------- ----------
```

```
      401                                                        450
 Pms  ---------- CTTTTCCAGT TCCCGAGGCG GATCCGGTGT TGCATCCTTG
Hpm4  ---------- CTTTTCCAGT TCCCGAGGCG GATCCGGTGT TGCATCCTTG
Pms2  GTGGAGGGAA CTTTCCCAGT CCCCGAGGCG GATCCGGTGT TGCATCCTTG
Hpms  ---------- ---------- ---------- ---------- ----------

      451                                                        500
 Pms  GAGCGAGCTG AGAGCTGGAG TACAGAACCT GCTAAGGCCA TCAAACCTAT
Hpm4  GAGCGAGCTG AGAGCTGGAG TACAGAACCT GCTAAGGCCA TCAAACCTAT
Pms2  GAGCGAGCTG AGAGCTCGAG TACAGAACCT GCTAAGGCCA TCAAACCTAT
Hpms  ---------- ---------- ----------T GCTAAGGCCA TCAAACCTAT

      501                                                        550
 Pms  TGATCGGAAG TCAGTCCATC AGATTTGCTC TGGGCCGGTA GTACTGAGTC
Hpm4  TGATCGGAAG TCAGTCCATC AGATTTGCTC TGGGCCGGTA GTACTGAGTC
Pms2  TGATCGGAAG TCAGTCCATC AGATTTGCTC TGGGCCGGTG GTACCGAGTC
Hpms  TGATCGGAAG TCAGTCCATC AGATTTGCTC TGGGCCGGTA GTACTGAGTC

      551                                                        600
 Pms  TAAGCACTGC GGTGAAGAAG ATTGTAGAAA ACAGTCTGGA TGCTGGTGCC
Hpm4  TAAGCACTGC GGTGAAGAAG ATTGTAGAAA ACAGTCTGGA TGCTGGTGCC
Pms2  TAAGCACTGC GGTGAAGGAG TTAGTAGAAA ACAGTCTGGA TGCTGGTGCC
Hpms  TAAGCACTGC GGTGAAGAAG ATGGTAGAAA ACAGTCTGGA TGCTGGTGCC

      601                                                        650
 Pms  ACTAATATTG ATCTAAAGCT TAAGGACTAT GGAGTGGATC TCATTGAAGT
Hpm4  ACTAATATTG ATCTAAAGCT TAAGGACTAT GGAGTGGATC TCATTGAAGT
Pms2  ACTAATATTG ATCTAAAGCT TAAGGACTAT GGAGTGGATC TCATTGAAGT
Hpms  ACTAATATTG ATCTAAAGCT TAAGGACTAT GGAATGGATC TCATTGAAGT

      651                                                        700
 Pms  TTCAGGCAAT GGATGTGGGG TAGAAGAAGA AAACTTCGAA GGCTTAACTC
Hpm4  TTCAGGCAAT GGATGTGGGG TAGAAGAAGA AAACTTCGAA GGCTTAACTC
Pms2  TTCAGGCAAT GGATGTGGGG TAGAAGAAGA AAACTTCGAA GGCTTAACTC
Hpms  TTCAGGCAAT GGATGTGGGG TAGAAGAAGA AAACTTCGAA GGCTTAA...

      701                                                        750
 Pms  TGAAACATCA CACATCTAAG ATTCAAGAGT TTGCCGACCT ACCTCAGGTT
Hpm4  TGAAACATCA CACATCTAAG ATTCAAGAGT TTGCCGACCT ACCTCAGGTT
Pms2  TGAAACATCA CACATCTAAG ATTCAAGAGT TTGCCGACCT ACCTCAGGTT
Hpms  .......... .......... .......... .......... ..........

      751                                                        800
 Pms  GAAACTTTTG GCTTTCGGGG GGAAGCTCTG AGCTCACTTT GTGCACTGAG
Hpm4  GAAACTTTTG GCTTTCGGGG GGAAGCTCTG AGCTCACTTT GTGCACTGAG
Pms2  GAAACTTTTG GCTTTCGGGG GGAAGCTCTG AGCTCACTTT GTGCACTGAG
Hpms  .......... .......... .......... .......... ..........

      801                                                        850
 Pms  TGATGTCACC ATTTCTACCT GCCACGTATC GGCGAAGGTT GGGACTCGAC
Hpm4  TGATGTCACC ATTTCTACCT GCCACGTATC GGCGAAGGTT GGGACTCGAC
Pms2  TGATGTCACC ATTTCTACCT GCCATGTATC GGCGAAGGTT GGGACTCGAC
Hpms  TGATGTCACC ATTTCTACCT GCCACGTATC GGCGAAGGTT GGGACTCGAC
```

```
      851                                                        900
 Pms  TGGTGTTTGA TCACGATGGG AAAATCATCC AGAAAACCCC CTACCCCCAC
Hpm4  TGGTGTTTGA TCACGATGGG AAAATCATCC AGAAAACCCC CTACCCCCAC
Pms2  TGGTGTTTGA TCACTATGGG AAAATCATCC AGAAAACCCC CTACCCCCAC
Hpms  TGGTGTTTGA TCACGATGGG AAAATCATCC AGAAAACCCC CTACCCCCAC

      901                                                        950
 Pms  CCCAGAGGGA CCACAGTCAG CGTGAAGCAG TTATTTTCTA CGCTACCTGT
Hpm4  CCCAGAGGGA CCACAGTCAG CGTGAAGCAG TTATTTTCTA CGCTACCTGT
Pms2  CCCAGAGGGA TGACAGTCAG TGTGAAGCAG TTATTTTCTA CGCTACCTGT
Hpms  CCCAGA.GGA CCACAGTCAG CGTGAAGCAG TTA.TTTCTA CGTACTGTGC

      951                                                       1000
 Pms  GCACCATAAA GAATTTCAAA GGAATATTAA GAAGAAACGT GCCTGCTTCC
Hpm4  GCACCATAAA GAATTTCAAA GGAATATTAA GAAGAAACGT GCCTGCTTCC
Pms2  GCACCATAAA GAATTTCAAA GGAATATTAA GAAGAAACGT GCCTGCTTCC
Hpms  G...CATAAG GAATTTCAAA GGAATATTAA GAAGAAACGT GCCTGCTTCC

      1001                                                      1050
 Pms  CCTTCGCCTT CTGCCGTGAT TGTCAGTTCC TTGAGGGCTC CCCAGCCATG
Hpm4  CCTTCGCCTT CTGCCGTGAT TGTCAGTTCC TTGAGGGCTC CCCAGCCATG
Pms2  CCTTCGCCTT CTGCCGTGAT TGTCAGTTTC CTGAGGCCTC CCCAGCCATG
Hpms  CCTTCGCCTT CTGCCGT.AT TGTCAGTTTC TTGAGGGCTC CCCAGCCATG

      1051                                                      1100
 Pms  CTTCCTGTAC AGCCTGCAAA ACTGACTCCT AGAAGTACCC CACCCCACCC
Hpm4  CTTCCTGTAC AGCCTGCAAA ACTGACTCCT AGAAGTACCC CACCCCACCC
Pms2  CTTCCTGTAC AGCCTGCAGA ACTGACTCCT AGAAGTACCC CACCCCACCC
Hpms  CTTCCTGTAC AGCCTGCAAA ACTGACTCCT AGAAGTACCC CACCCCACCC

      1101                                                      1150
 Pms  CTGCTCCTTG GAGGACAACG TGATCACTGT ATTCAGCTCT GTCAAGAATG
Hpm4  CTGCTCCTTG GAGGACAACG TGATCACTGT ATTCAGCTCT GTCAAGAATG
Pms2  CTGCTCCTTG GAGGACAACG TGATCACTGT ATTCAGCTCT GTCAAGAATG
Hpms  CTGCTCCTTG GAGGACAACG TGATCACTGT ATTCAGCTCT GTCAAGAATG

      1151       stop code|                                   1200
 Pms  GTCCAGGTTC TTCTAGATGA TCTGCACAAA TGGTTCCTCT CCTCCTTCCT
Hpm4  GTCCAGGTTC TTCTAGATGA TCTGCACAAA TGGTTCCTCT CCTCCTTCCT
Pms2  GTCCAGGTTC TTCTAGATGA TCTGCACAAA TGGCTCCTCT CCTCCTTCCT
Hpms  GTCCAGGT.. .......... .......... ....TCCTCT CCTCCTTCCT

      1201                                                      1250
 Pms  GATGTCTGCC ATTAGCATTG GAATAAAGTT CCTGCTGAAA ATCCACATCT
Hpm4  GATGTCTGCC ATTAGCATTG GAATAAAGTT CCTGCTGAAA ATCCACATCT
Pms2  GATGTCTGCC ATTAGCACTG GAATAAAGTT CCTGCTGAAA ATCC------
Hpms  GATGTCTGCC ATTAGCATTG GAATAAAGTT CCTGCTGAAA ATCCACATCT

      1251                                                      1300
 Pms  CCCCTGGGTC CGGTGTTCTG GAAGTGAGAG AGACAATGTC ACACTTCAAG
Hpm4  CCCCTGGGTC CGGTGTTCTG GAAGTGAGAG AGACAATGTC ACACTTCAAG
Pms2  ---------- ---------- ---------- ---------- ----------
Hpms  CCCCTGGGTC CGGTG.TCTG GAAGTGAGAG AGACAATGTC ACACTTCAAG
```

```
      1301                                                       1350
 Pms  GAGGCAGCTC TCTAGACAGG AAGGTTATTC ACGTCCCATG TCAAGTCTAG
Hpm4  GAGGCAGCTC TCTAGACAGG AAGGTTATTC ACGTCCCATG TCAAGTCTAG
Pms2  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
Hpms  GAGGCAGCTC TCTAGACAGG AAGGTTATTC ACGTCCCATG TCAAGTCTAG

      1351                                                       1400
 Pms  CTAGAGTTCA GAGCAATTGA GAAGTGCAAT TTTATCTCCT GCCTTTCATT
Hpm4  CTAGAGTTCA GAGCAATTGA GAAGTGCAAT TTTATCTCCT GCCTTTCATT
Pms2  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
Hpms  CTAGAGTTCA GAGCAATTGA GAAGTGCAAT TTTATCTCCT GCCTTTCATT

      1401                                                       1450
 Pms  CTATACCCTG CTTCTGAACC ATCGTGTTCA ACTGTGAAAC TCACACTTTG
Hpm4  CTATACCCTG CTTCTGAACC ATCGTGTTCA ACTGTGAAAC TCACACTTTG
Pms2  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
Hpms  CTATACCCTG CTTCTGAACC ATCGTGTTCA ACTGTGAAAC TCACACTTTG

      1451                                                       1500
 Pms  GTGACCCTGA CTCCAAAACT TAATACACCC AAGGTCAGCC CCAGTGATCT
Hpm4  GTGACCCTGA CTCCAAAACT TAATACACCC AAGGTCAGCC CCAGTGATCT
Pms2  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
Hpms  GTGACCCTGA CTCCAAAACT TAATACACCC AAGGTCAGCC CCAGTGATCT

      1501                                                       1550
 Pms  GCTTCATAGC AAGGACTTTG GGTGGGTCTT CCCAGGGAGT AGGGCACCCT
Hpm4  GCTTCATAGC AAGGACTTTG GGTGGGTCTT CCCAGGGAGT AGGGCACCCT
Pms2  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
Hpms  GCTTCATAGC AAGGACTTTG GGTGGGTCTT CCCAGGGAGT AGGGCACCCT

      1551                                                       1600
 Pms  TCAGAGAATT G~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
Hpm4  TCAGAGAATT G~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
Pms2  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
Hpms  CAGAGAATGT GGCTTTGGAC TTCATCACAG CTGGGGCCTT TTGTGTCACT

      1601                                                       1650
 Pms  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
Hpm4  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
Pms2  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
Hpms  TCAGATCTAA ACTTGTAACC GTGCTAGATC TGTTTCTAAC GTGACAACAT

      1651                                                 1693
 Pms  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
Hpm4  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
Pms2  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~
Hpms  CACGAACCAC GAGTCCAGAA GCCTAATCCA TAATCCTCCC CCA
```

Fig. 5.6 Comparison of four hPMS mRNA sequences using Pileup from GCG

two promoter regions, one at position 152121 and another at 151321. It was reported that the hPMS1 gene lacks TATA-boxes within 1.4 kb of the 5' region (Yanagisawa, 1998). In the present research, however, the region approximately one kb upstream of the 5' exon is an AT rich region (90% AT from 152707 to 152843) and contains six TATAA sequences. A CpG island is located in the region upstream of the first exon, confirming the observation of Yanagisawa, 1998. Finally, a putative polyadenylation signal (AATAAA) is found in each of above hPMS genes upstream of 3' end of last exon at positions 168746-168751 (Fig. 5.6).

5.2.3 Human Bruton's tyrosine kinase-associated protein-135 gene (BAP-135)

When B cells are stimulated by antigens, Src-related tyrosine kinase is immediately activated, followed by the activation of Bruton's tyrosine kinase (Btk) and the tyrosine kinase Syk (Saouaf, 1994). Btk is essential for B cell activation and mutations on Btk are associated with agammaglobulinemia in human (Tsukada et al., 1993; Vetrie et al., 1993) and impaired B cell proliferation in mice (Thomas et al., 1993; Rawlings et al., 1993). It was found that a protein, Bruton's tyrosine kinase-associated protein-135 (BAP-135), has been associated *in vivo* with Btk and mutations that impair Btk activation also abolish Btk-dependent phosphorylation of BAP-135 (Yang and Desiderio, 1997). BAP-135 may be the downstream target of Btk in a signaling pathway originating at the engagement of B cell antigen receptor (Yang and Desiderio, 1997).

The BAP-135 cDNA (gi|1870687|HSU77948; Yang and Desiderio, 1997) is 3328 bp long and codes for a 958 amino acid protein. A Sim4 alignment of BAP-135 cDNA and the genomic DNA shows that the genomic BAC 239c10 encodes only the 16 C-

terminal exons in the 3' half of the gene (Table 5.8), which span a region of about 16.5 kb at the beginning of contig 239c10 (Table 5.8).

Since the BAP-135 enzyme is a multifunctional protein, the sequence of its cDNA also was submitted to GenBank and named as SRF-Phox1 interacting protein (SPIN protein) by another group (gi|2440077|HSY14946; Grueneberg et al., 1997). It is likely that this SPIN mRNA which is 2874 bp long and has 99.9% identity to BAP-135 mRNA (position 7-2880), represents the same BAP-135 protein coding region. When aligned, the sequence of BAC clone 239c10 differed from the BAP-135 in only three positions (Table 5.8) and there were no discrepancies when compared to the sequence of SPIN cDNA (Table 5.9). Fig. 5.7 shows the GAP alignment of the BAP-135 and SPIN proteins, indicating that they are 99.6% identitical with discrepancies only at positions 174, 178 and 593, 919. Also, the BAP-135 cDNA sequence is larger, beginning in the 5' UTR and ending in the 3' UTR, while the SPIN cDNA sequence starts at start codon and ends at stop codon.

While the BAP-135 protein associates with the Btk protein-tyrosine kinase (Yang and Desiderio, 1977), it also has been reported that SPIN is one of the proteins involved in the formation of an active promoter complex near the c-fos gene (Grueneberg et al., 1997). Here, a serum response element (SRE), an element within the enhancer of the proto-oncogene c-fos, is the target for activation of c-fos transcription by multiple signal transduction pathways (Treisman, 1985, 1986; Rivera et al., 1990). Also, serum response factor (SRF), a protein that is critical in the activation of mammalian genes and the Phox1 protein originally identified in the yeast, was found to interact with SRF and thereby

Table 5.8 The ranges and sizes of exons and introns of the BAP-135 gene

No.	Range of cDNA	Range of Genomic DNA	Exon size Size (bp)	Intron Size (bp)	Similarity
1	1558-1633*	4842-4917	76	1240	100%
2	1634-1817	6156-6339	184	943	99%
3	1818-1876	7281-7339	59	397	100%
4	1877-1951	7735-7809	75	1604	100%
5	1952-2053	9412-9513	102	910	100%
6	2054-2119	10422-10487	66	179	100%
7	2120-2303	10665-10848	184	1906	100%
8	2304-2359	12753-12808	56	675	100%
9	2360-2440	13482-13562	81	925	100%
10	2441-2524	14486-14569	84	670	100%
11	2525-2708	15238-15421	184	1481	100%
12	2709-2737	16901-16929	29	1284	100%
13	2738-2779	18212-18253	42	1100	98%
14	2780-2821	19352-19393	42	778	100%
15	2822-2897	20170-20245	76	694	100%
16	2898-3320	20938-21360	423		100%

*The N-terminal exons are not located in this contig.

Table 5.9 The ranges and sizes of exons and introns of the SPIN gene

No.	Range of cDNA	Range of Genomic DNA	Exon size Size (bp)	Intron Size (bp)	Similarity
1	1552-1627*	4842-4917	76	1240	100%
2	1628-1811	6156-6339	184	943	100%
3	1812-1870	7281-7339	59	397	100%
4	1871-1945	7735-7809	75	1604	100%
5	1946-2047	9412-9513	102	910	100%
6	2048-2113	10422-10487	66	179	100%
7	2114-2297	10665-10848	184	1906	100%
8	2298-2353	12753-12808	56	675	100%
9	2354-2434	13482-13562	81	925	100%
10	2435-2518	14486-14569	84	670	100%
11	2519-2702	15238-15421	184	1481	100%
12	2703-2731	16901-16929	29	1284	100%
13	2732-2773	18212-18253	42	1100	100%
14	2774-2815	19352-19393	42	778	100%
15	2816-2874	20170-20228	58	694	100%

*The N-terminal exons are not located in this contig

```
                .         .         .         .         .
bap-135  1 MAQVAMSTLPVEDEESSESRMVVTFLMSALESMCKELAKSKAEVACIAVY 50
           |||||||||||||||||||||||||||||||||||||||||||||||||||
SPIN     1 MAQVAMSTLPVEDEESSESRMVVTFLMSALESMCKELAKSKAEVACIAVY 50
                .         .         .         .         .
        51 ETDVFVVGTERGRAFVNTRKDFQKDFVKYCVEEEEKAAEMHKMKSTTQAN 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
        51 ETDVFVVGTERGRAFVNTRKDFQKDFVKYCVEEEEKAAEMHKMKSTTQAN 100
                .         .         .         .         .
       101 RMSVDAVEIETLRKTVEDYFCFCYGKALGKSTVVPVPYEKMLRDQSAVVV 150
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       101 RMSVDAVEIETLRKTVEDYFCFCYGKALGKSTVVPVPYEKMLRDQSAVVV 150
                .         .         .         .         .
       151 QGLPEGVAFKHPENYDLATLKWIGENKGGISFIIKRPFLEPKKHVGGRVM 200
           |||||||||||||||||||||| ||| ||||||||||||||||||||||||
       151 QGLPEGVAFKHPENYDLATLKWILENKAGISFIIKRPFLEPKKHVGGRVM 200
                .         .         .         .         .
       201 VTDADRSILSPGGSCGPIKVKTEPTEDSGISLEMAAVTVKEESEDPDYYQ 250
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       201 VTDADRSILSPGGSCGPIKVKTEPTEDSGISLEMAAVTVKEESEDPDYYQ 250
                .         .         .         .         .
       251 YNIQGSHHSSEGNEGTEMEVPAEDDDYSPPSKRPKANELPQPPVPEPANA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       251 YNIQGSHHSSEGNEGTEMEVPAEDDDYSPPSKRPKANELPQPPVPEPANA 300
                .         .         .         .         .
       301 GKRKVREFNFEKWNARITDLRKQVEELFERKYAQAIKAKGPVTIPYPLFQ 350
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       301 GKRKVREFNFEKWNARITDLRKQVEELFERKYAQAIKAKGPVTIPYPLFQ 350
                .         .         .         .         .
       351 SHVEDLYVEGLPEGIPFRRPSTYGIPRLERILLAKERIRFVIKKHELLNS 400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       351 SHVEDLYVEGLPEGIPFRRPSTYGIPRLERILLAKERIRFVIKKHELLNS 400
                .         .         .         .         .
       401 TREDLQLDKPASGVKEEWYARITKLRKMVDQLFCKKFAEALGSTEAKAVP 450
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       401 TREDLQLDKPASGVKEEWYARITKLRKMVDQLFCKKFAEALGSTEAKAVP 450
                .         .         .         .         .
       451 YQKFEAHPNDLYVEGLPENIPFRSPSWYGIPRLEKIIQVGNRIKFVIKRP 500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       451 YQKFEAHPNDLYVEGLPENIPFRSPSWYGIPRLEKIIQVGNRIKFVIKRP 500
                .         .         .         .         .
       501 ELLTHSTTEVTQPRTNTPVKEDWNVRITKLRKQVEEIFNLKFAQALGLTE 550
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       501 ELLTHSTTEVTQPRTNTPVKEDWNVRITKLRKQVEEIFNLKFAQALGLTE 550
                .         .         .         .         .
       551 AVKVPYPVFESNPEFLYVEGLPEGIPFRSPTWFGIPRLERIVHGSNKIKF 600
           |||||||||||||||||||||||||||||||||||||||||| |||||||
       551 AVKVPYPVFESNPEFLYVEGLPEGIPFRSPTWFGIPRLERIVRGSNKIKF 600
                .         .         .         .         .
       601 VVKKPELVISYLPPGMASKINTKALQSPKRPRSPGSNSKVPEIEVTVEGP 650
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       601 VVKKPELVISYLPPGMASKINTKALQSPKRPRSPGSNSKVPEIEVTVEGP 650
                .         .         .         .         .
       651 NNNNPQTSAVRTPTQTNGSNVPFKPRGREFSFEAWNAKITDLKQKVENLF 700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       651 NNNNPQTSAVRTPTQTNGSNVPFKPRGREFSFEAWNAKITDLKQKVENLF 700
```

```
                .         .         .         .         .
701 NEKCGEALGLKQAVKVPFALFESFPEDFYVEGLPEGVPFRRPSTFGIPRL 750
    |||||||||||||||||||||||||||||||||||||||||||||||||||
701 NEKCGEALGLKQAVKVPFALFESFPEDFYVEGLPEGVPFRRPSTFGIPRL 750
                .         .         .         .         .
751 EKILRNKAKIKFIIKKPEMFETAIKESTSSKSPPRKINSSPNVNTTASGV 800
    |||||||||||||||||||||||||||||||||||||||||||||||||||
751 EKILRNKAKIKFIIKKPEMFETAIKESTSSKSPPRKINSSPNVNTTASGV 800
                .         .         .         .         .
801 EDLNIIQVTIPDDDNERLSKVEKARQLREQVNDLFSRKFGEAIGMGFPVK 850
    |||||||||||||||||||||||||||||||||||||||||||||||||||
801 EDLNIIQVTIPDDDNERLSKVEKARQLREQVNDLFSRKFGEAIGMGFPVK 850
                .         .         .         .         .
851 VPYRKITINPGCVVVDGMPPGVSFKAPSYLEISSMRRILDSAEFIKFTVI 900
    |||||||||||||||||||||||||||||||||||||||||||||||||||
851 VPYRKITINPGCVVVDGMPPGVSFKAPSYLEISSMRRILDSAEFIKFTVI 900
                .         .         .         .         .
901 RPFPGLVINNQLVDQSESKGPVIQESAEPSQLEVPATEEIKETDGSSQIK 950
    |||||||||||||||||.|||||||||||||||||||||||||||||||||
901 RPFPGLVINNQLVDQSESEGPVIQESAEPSQLEVPATEEIKETDGSSQIK 950

951 QEPDPTW 957
    |||||||
951 QEPDPTW 957
```

Fig. 5.7 GAP align of BAP-135 and SPIN protein sequences

enhance the rate at which SRF could bind to SRE (Grueneberg et al., 1992, 1995). The SRF-Phox1 interacting protein (SPIN) also reportedly interacts with SRF and Phox1 *in vitro* and *in vivo* (Grueneberg et al., 1997). Furthermore, SPIN was observed to bind to multiple sequences in the c-fos promoter, interacted cooperatively with both SRF and Phox1 to promote the formation of stable higher-order complexes of SRF and Phox1, and promoted the transcription of a reporter gene driven by the c-fos SRE (Grueneberg et al., 1997).

Regulatory element: The promoter regions for the BAP-135 and SPIN genes are not located in BAC clone 239c10 as it lacks the 5' exons and upstream region of the BAP-135 gene. Therefore, only regulatory elements in the 3' most exon of BAP-135 (or SPIN) gene, which encodes a 449 bp AT-rich (63% AT bases) region were studied. Here, a putative polyadenylation signal (AATAAA) is observed 57 bp upstream 3' end of last exon at positions 20171-20177. A second polyadenylation signal (GT rich segment) also was observed 29 bp downstream of the 3' end of last exon beginning at position 20257.

5.2.4 Human prohibitin pseudo gene

Breast cancer is a common cancer in women and its highest risk factor is the family history (Anderson, 1972), which may be correlated with inherited susceptibility. Previous studies identified 17q21 as one of the commonly deleted region in breast tumors (Sato et al., 1990, 1991). Here, a rat prohibitin gene which has the ability to negatively regulate cell proliferation (Nuell et al., 1991), was cloned (Nuell et al., 1991) and subsequently used to isolate a fragment of human genomic DNA. This fragment was mapped by *in situ* hybridization to human chromosome 17 region 17q12-17q21 (White et al., 1991). Then, the human prohibitin gene, reported to be one of the genes responsible

for hereditary breast cancer (Sato et al., 1992), was cloned and sequenced (gi|246482|S85655).

The human prohibitin cDNA is 1043 bp long, having a coding region between positions 51-869, which results in a translated protein that is 272 amino acids long (Sato et al., 1992). Sim4 was used to align human prohibitin gene and BAC 239c10 (Fig. 5.8). Since the similarity only is 84% at positions 423-451 and 79% at positions 2600-3519 (Table 5.10), this suggests that this region of human chromosome 7 may contain a prohibitin pseudogene.

Table 5.10 The ranges and sizes of exons and introns of prohibitin gene

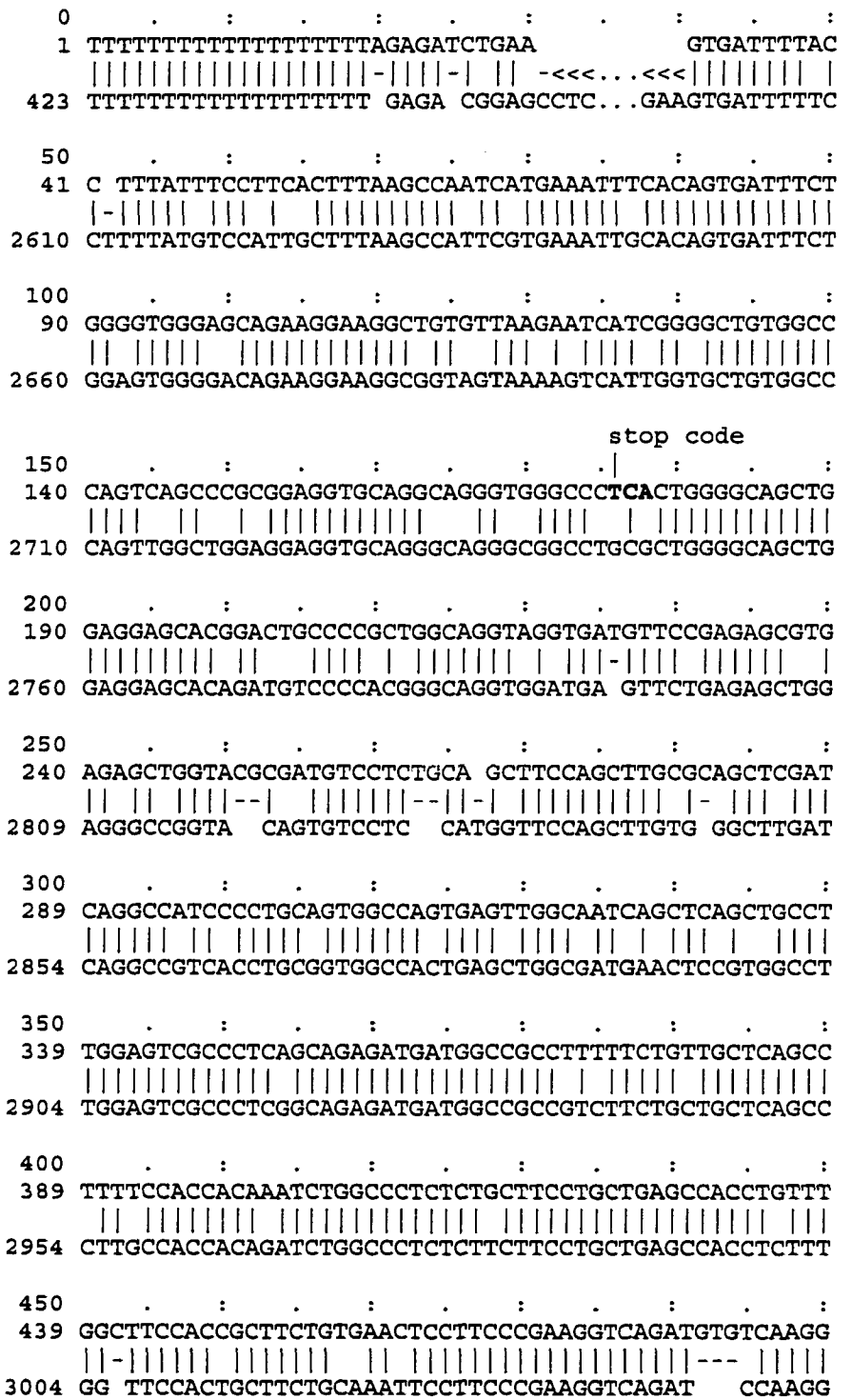
No.	Range of mRNA	Range of Genomic DNA	Exon size Size (bp)	Intron Size (bp)	Similarity
1	1-30	423-451	30	2149	84%
2	31-967	2600-3519	>919		79%

(The N-terminal exons are not located in this contig)

Regulatory element: The promoter region for the prohibitin gene could not be defined because the 5' end of mRNA was not contained within this BAC clone.

5.2.5 Repetitive elements in 239c10

A RepeatMasker (section 1.6.6) analysis revealed that BAC 239c10 contained a total of 61.12 % repeat elements (Table 5.11; Fig. 5.9). Alu repeats were the most prevalent and they represented 46.50% of the total sequence. Since, on average, the percentage of Alu sequence in the human genome is ~ 30%, this region contains a higher than usual concentration of Alu elements and correspondingly lower percentages of simple repeats and low complexity repeats.


stop code

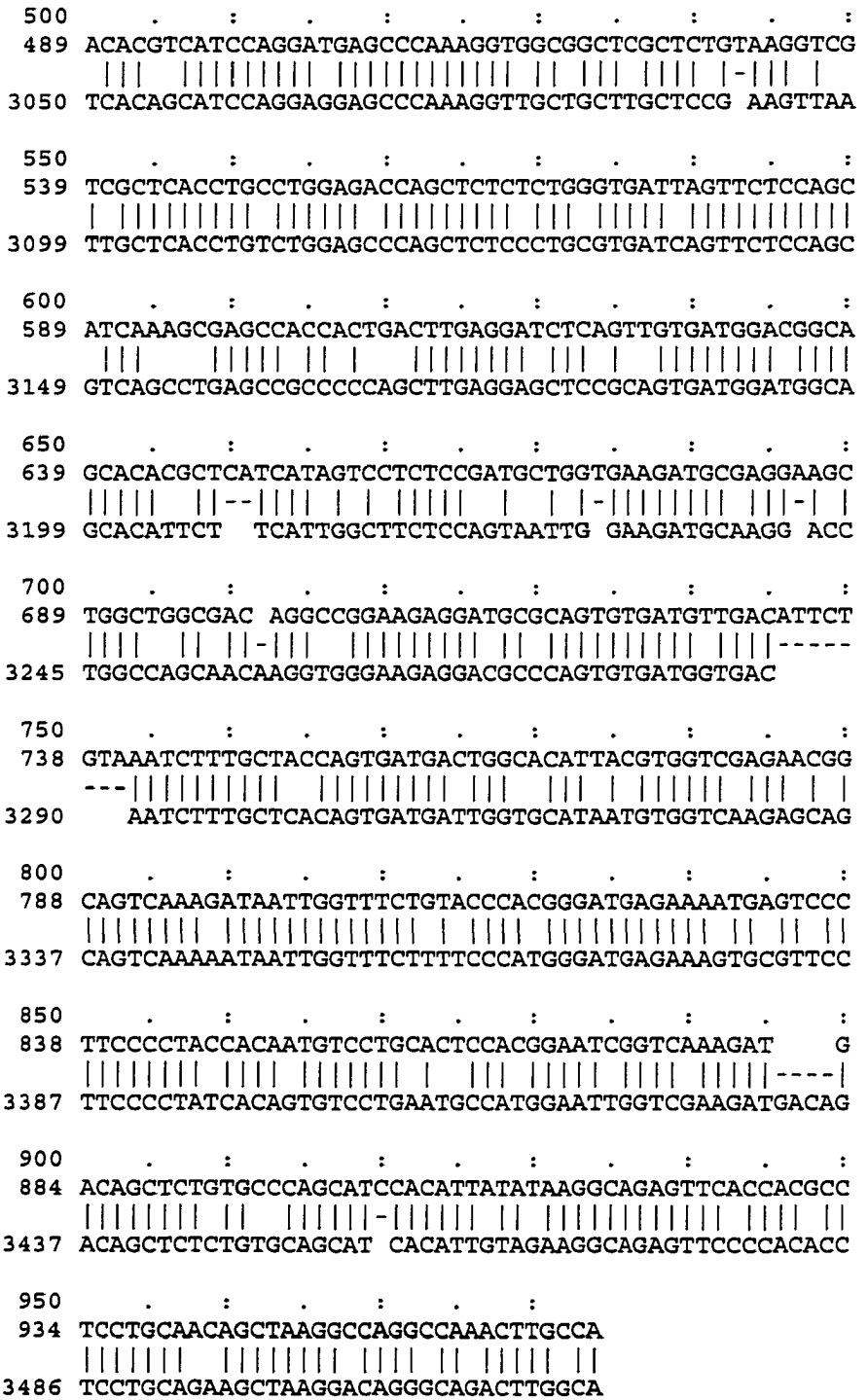


Fig. 5.8 Sim4 alignment of human prohibitin mRNA and AC004166

Table 5.11 Repeatitive elements in AC004166 (BAC 239c10)*

GC level: 48.47 %
bases masked: 168190 bp (61.12 %)

	number of elements*	length occupied	percentage of sequence
SINEs:	505	132647 bp	48.20 %
ALUs	467	127960 bp	46.50 %
MIRs	38	4687 bp	1.70 %
LINEs:	56	17938 bp	6.52 %
LINE1	37	14760 bp	5.36 %
LINE2	17	2710 bp	0.98 %
L3/CR1	2	468 bp	0.17 %
LTR elements:	25	6903 bp	2.51 %
MaLRs	15	4692 bp	1.70 %
ERVL	7	1377 bp	0.50 %
ERV_classI	3	834 bp	0.30 %
DNA elements:	24	6649 bp	2.42 %
MER1_type	14	4137 bp	1.50 %
MER2_type	4	1648 bp	0.60 %
Total interspersed repeats:		164137 bp	59.64 %
Small RNA:	1	96 bp	0.03 %
Simple repeats:	46	2306 bp	0.84 %
Low complexity:	39	1669 bp	0.61 %

*RepeatMasker version 08/14/2000 , default mode
run with cross_match version 0.990319
RepBase version 06/31/2000

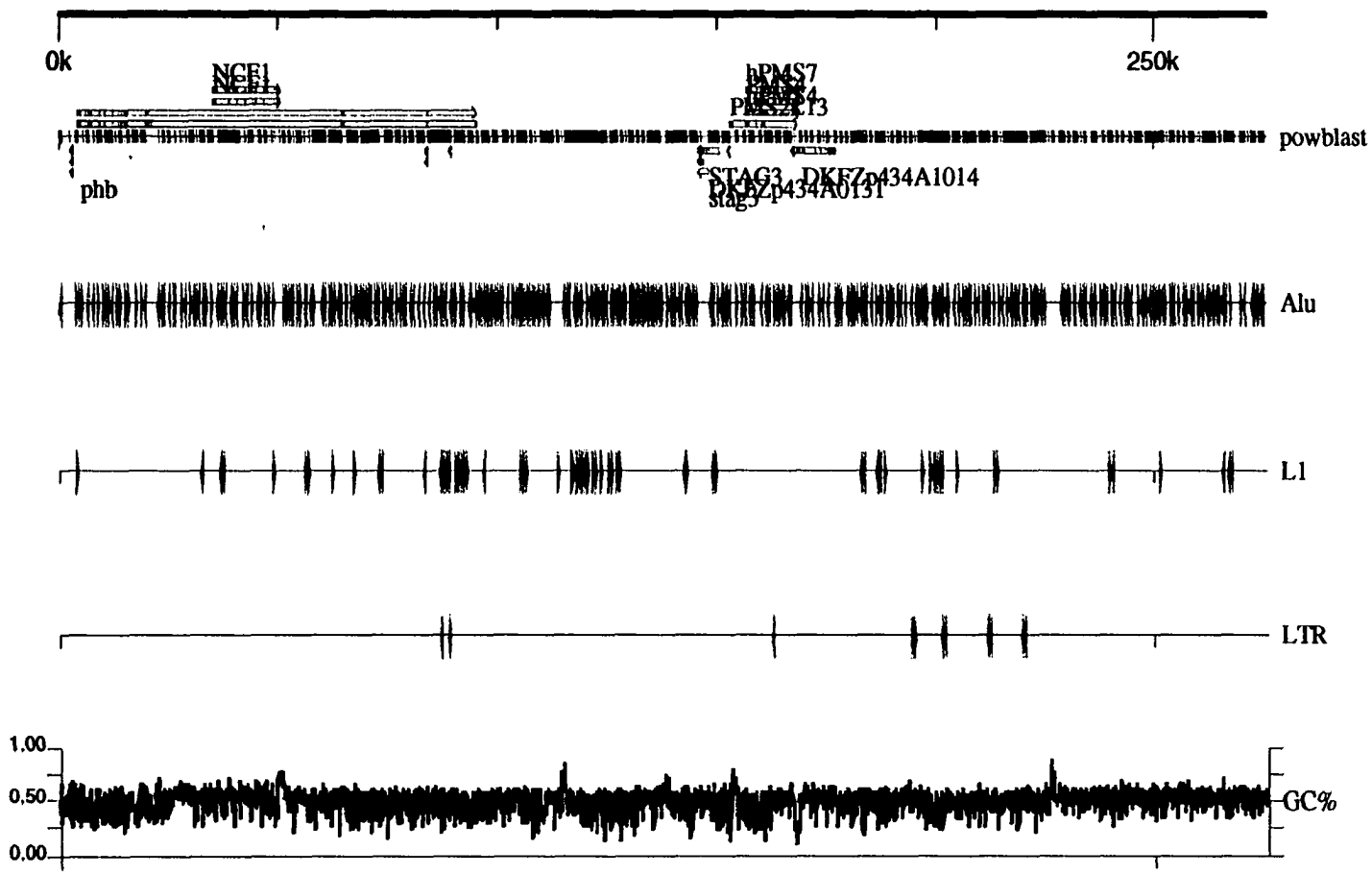


Fig. 5.9 Gene prediction, repeat elements and GC level in AC004166

Chapter VI

Conclusion

Although it was reported over eighteen years ago that single-pass partially sequencing of random cDNA clones was a relative inexpensive and rapid means to determine when and where specific genes were expressed (Putney et al., 1983, Milner and Sutcliffe, 1983; Costanzo et al., 1983), it was not until ten years later that the first large scale EST database was produced (Adams, 1991). Since then, ESTs have been extremely valuable for both gene discovery and for measure gene expression levels in a wide variety of species (Adams et al., 1995; Hillier et al., 1996; Khan et al., 1999; Marra et al., 1999; Bailey, Jr., et al., 1998; Yamamoto et al., 1997, Delseny et al., 1997). Now, as of 04-13-01, over 7.78 million of ESTs were deposited into dbEST on the NCBI's web site.

Fusarium sporotrichioides, a filamentous fungus which is pathogenic to maize, barley, rye, wheat and rice (Desjardins et al., 1993; Alexander et al. 1999), produces several trichothecenes, compounds that are toxic to humans and almost all animals tested. The vast majority of the *F. sporotrichioides* trichothecene biosynthetic genes are located in a coordinately regulatory gene cluster within a 27-kb region (Desjardins et al., 1993; Hohn et al., 1995; Keller and Hohn, 1997; Alexander et al., 1999). Expression of genes in this region is controlled by *Tri10*, a regulatory gene which is located within the cluster and which dramatically increases or decreases toxin production when it is mutated (Peplow et al., 1997; Gurifulina et al., 1998; Tag et al., 1998). In this dissertation research, an EST database was produced from a *F. sporotrichioides Tri10* overexpressed cDNA library constructed by Dr. Marian Beremand of Texas A&M. Analysis of this data

reveals the gene expressions patterns as represented by the mRNA levels in the library, as well as identifying additional genes that also are controlled by the *Tri10* gene product. This EST data presently are being used by Dr. Beremand to construct a DNA microarray that will be used to study *F. sporotrichioides* gene expression under different environmental and genetic conditions.

During the *F. sporotrichioides* EST project, 10,256 raw sequences were produced. After removing the low quality sequences, non-EST sequences, short insert and wrong end sequences, 7,495 high quality ESTs were obtained.

During the data collecting stage, once every 100 new 3' EST sequences were obtained, cumulative 3' assemblies using Phrap were performed to determine the number of new genes represented and the gene redundancy of the library. This analysis revealed that the *F. sporotrichioides* EST project had a redundancy of 70% once 3243 3' ESTs were obtained.

The high quality ESTs were assembled into 3' EST and 5' EST Phrap database and then analyzed by a BlastX homology search against the non-redundant (nr) protein database. Totally, 2,181 singlets and 1,057 contigs were assembled to form the database. To obtain an accurate estimate of the expressed gene number, the singlets and contigs that had pair members in the other singlets or contigs were subtracted from the total, because those paired singlets or contigs represented the same mRNAs and hence the same gene. 2139 genes were represented in this database after the 1,099 duplications were subtracted from 3,238 total singlets and contigs. Interestingly, only approximately 50% of these genes had significant BlastX homologues in the GenBank nr protein database.

Because the level of expression can vary widely for different genes, it often is useful to analyze the relative levels of gene expression in EST databases in classes. Here, the *F. sporotrichioides* EST population fell into three classes. 12 contigs were in the high abundance class which represented 11.57% of the ESTs in the database, with 46-164 ESTs in each contig. The intermediate abundance class consisted of 496 contigs, with 3-45 ESTs in each contig, and represented 45.08% of ESTs in the database. The low abundance class consisted of 2182 singlets and 549 contigs and represented about 43.35% of ESTs in the database. This population distribution was very similar to that described by Bishop and Soares (Bishop et al., 1974; Soares et al., 1994) for their EST studies.

Among the top twelve highly expressed genes in *F. sporotrichioides* EST database, five of them were the trichothecene biosynthesis pathway genes. To date, fourteen genes in the trichothecene biosynthesis pathway have been identified in *F. sporotrichioides* (section 4.3.4). Thirteen of the fourteen genes products were present in the *F. sporotrioides* EST database. Totally, 571 ESTs or 7.62% of the total 7495 ESTs in the database, represented genes involved in trichothecene biosynthesis pathway. Only the *Tri6* gene product was not found in the *F. sporotrichioides* database. It is likely that this gene product is either at low level or absent in the cDNA library studies since "*Tri6* appeared to be expressed for only a limited period prior to the toxin production." (Matsumoto et al., 1999). Three of the fourteen genes are new genes that were discovered through this EST project collaborately with our collaborators.

Several computer programs have been implemented to enable a semi-automated process of biological function assignments for *F. sporotrichioides* ESTs that have

significant BlastX homology. At the end of the data analysis stage, each singlet or contig was assigned a biological function based on the results of the BlastX homology search. The biological function schema developed by Monica Reily was used to generate a primary keyword list. Then, after several reiterations, a final keyword list was obtained which included approximately 750 key words. The 7495 *F. sporotrichioides* ESTs were placed into the seven Riley categories yielding 15% metabolism related ESTs, 9% genetic information processing related, 4% cell growth related, 10% other processes related, 4% unclassified, 8% unidentified and 50% previously unknown as they had no significant homologs in the GenBank non-redundant protein database.

Basing these assignments on the schema developed by Monica Reily allowed an easy comparison of the overall gene expression patterns among the different EST databases. The comparison of four EST databases determined in our laboratory, *F. sporotrichioides*, two *Neurospora crassa* libraries and one *Aspergillus nidulans*, shows that the percentage of biological function divisions in each library is remarkably similar in spite of their diverse source. However, the most highly expressed genes in each library were quite different, with 5.5% of the total EST population in *A. nidulus* representing heat shock protein 30, 7.62% of the total EST population in *F. sporotrichioides* database representing trichothecene biosynthesis pathway genes and approximately 9% of the total EST population in *N. crassa* database representing clock control genes.

The singlets and clusters that have no significant homologs by BlastX against the GenBank non-redundant protein database clearly represent newly discovered genes. Using this group of ESTs from *F. sporotrichioides* database to perform a tBlastX search against dbEST revealed 323 singlets and clusters which had homologs in dbEST. Among

them, 80 singlets and clusters in the *F. sporotrichioides* EST database had homologs in the *A. nidulus* EST database and 91 singlets and clusters had homologs in the *N. crassa* EST database. Furthermore, there were 27 singlets and clusters that had homologs in both *A. nidulus* and *N. crassa* EST databases. These 27 new genes that were present in all of the four fungal libraries in ACGT are valuable candidates for further studies.

As state in the first sentence of a Nature paper, "The human genome holds an extraordinary trove of information about human development, physiology, medicine and evolution" (Lander et al., 2001). The Human Genome Project was launched in 1990 and since then, 20 groups of scientists from the United States, the United Kindom, Japan, France, Germany and China have been involved in this object. Up to now, a draft sequence that covers about 94% of the human genome has been produced (Lander et al., 2001).

In this dissertation research, seven BACs, PAC and cosmids from Human Genome Project were sequenced into single piece and error rate decreased to less than 1 error every 10,000 bases. Among them, two BAC sequences were thoroughly analyzed, one is BAC 322f3 from human chromosome 22 and another is BAC 239c10 from human chromosome 7.

The sequence of BAC 322f3 revealed the presence of a 9098 bp region which was absent in the Ig λ region of the reference sequence for human chromosome 22. A paper in Nature (Roach et al., 1999) emphasized the urgency of filling the innumerable sequence gaps of the human genome that may be resulted from the shortcomings of the sequencing strategies used. The likely common occurrence of inborn deletions, such as the one

illustrated with this deleted region, may make it difficult to construct a truly representative human genomic map. The detection of these types of deletion and of their normal variants is practically impossible without a specially designed strategy. Fortunately, the molecular tools now are available for these population studies. These tools include the application of DNA microarray systems and of microchip technologies combined with multiple PCR analysis that when used for the systematic screening for deletion polymorphism, may reveal new chromosomal sites which are prone to genetic recombination events.

The sequence of BAC239c10 confirms the presence of several genes that have been mapped to chromosome band 7q11.23, the band related to William Syndrome (section 1.9). BAC 239c10 encodes three genes, the human neutrophil cytosol factor 1 (NCF1) gene, the human hPMS gene and the human Bruton's tyrosine kinase-associated protein-135 gene (BAP-135). NCF1, which is required for activation of nicotinamide adenine dinucleotide phosphate (NADPH) oxidase (Volpp et al., 1989), has been mapped in 7q11.23 and is not deleted in WMS (Francke et al., 1990). The human hPMS gene family which is related to hereditary nonpolyposis colorectal cancer (HNPCC) (Fishel et al., 1993; Leach et al., 1993; Bronner et al., 1994; Papadopoulos et al., 1994), has been mapped to chromosomal bands 7q11.23 and 7q22 by fluorescent in situ hybridization (FISH) (Horii et al., 1994, Papadopoulos et al., 1994). The human Bruton's tyrosine kinase associate protein 135, a multifunctional protein termed BAP-135, may be the downstream target of Btk in a signaling pathway originating at the engagement of B cell antigen receptor (Yang and Desiderio, 1997). Another gene, termed SPIN, which has 99.9% identity to BAP-135, contains the same protein-coding region as BAP-135. SPIN

is one of the coordinators for the formation of an active promoter complex at the c-fos gene (Grueneberg et al., 1997). Either BAP-135 or SPIN has not been mapped to 7q11.23 in the previous studies. Finally, a prohibitin peudogene was found in BAC239c10. This pseudogene has 80% identity to the human prohibitin gene that maps to 17q21 (Sato et al., 1990, 1991) and is in one of the commonly deleted regions in breast tumors.

From the rediscovery of the Mendel's laws in the first month of the 20th century to the publications of the human genome sequence draft in the second month of the 21th century (Lander et al., 2001; Venter et al., 2001), many exciting scientific advances have been made. The most significant progresses have occurred in the last two decades because of the development of three complementary sequencing approaches, whole genome sequencing, full-length cDNA sequencing and EST sequencing. Using these approaches has dramatically increased the deciphering of amount of genetic information coding in a large numbers of organisms. The whole genome sequence provides a broad landscape of the heredity basis while the full-length cDNA and EST data changes this silent picture into a talking book. That is, genes contained in the entire genome sequence that are predicted and identified by bioinformatics and comparative genomics approaches, now seem to come to life when their full-length cDNA and ESTs are described. These expression-based studies confirm and refine the gene structures, reveal which gene is active and how active it is at a particular development stage or a particular growth condition. However, because cDNA and EST studies give only snapshots of life processes under a certain condition, this snapshot is converted into a movie, when many

conditions are compared, as for example, using microarray based proteomics technologies (O'Donovan et al., 2001; Pradet-Balade et al., 2001).

As mention above when discussed BAC322f3, the truly representative human genomic map will be difficult to construct without in depth studies that include, for example, population genomics. Similarly, although several genes were defined in 7q11.23 after sequencing of BAC239c10, the biological mechanisms and pathways that lead from genes to WMS symptoms still will be difficult to define unless they are combined with proteomic analysis of clinical samples. Finally, the 27 EST database members that are present in all three filamentous fungus, but not in *Saccharomyces cerevisiae* or any other organisms studies to date, may represent the key genes that are responsible for maintaining the very unique properties of filamentous fungi. Therefore, nature is an amazing masterpiece. The more we learn, the more there is to explore, and the more is work there that remains to be done to decipher the codes written within it.

Literature Cited

Abrahamsen, M.S. (1999) *Cryptosporidium parvum* gene discovery. *Adv. Exp. Med. Biol.* **473**, 241-247.

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., et al. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**(5013), 1651-1656.

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., et al. (1995) Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* **377**(6547 Suppl), 3-174.

Alexander, N.J., McCormick, S.P. and Hohn, T.M. (1999) Tri12, a trichothecene efflux pump from *Fusarium sporotrichioides*: gene isolation and expression in yeast. *Mol. Gen. Genet.* **261**, 977-984.

Allen, O.B. (1983) A guide to the analysis of growth curve data with special reference to SAS. *Comput. Biomed. Res.* **16**(2), 101-115.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.

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

Anderson, D.E. (1972) A genetic study of human breast cancer. *J. Natl. Cancer Inst.*, **48**, 1029-1034.

Anderson, S., Bamkier, A.T., Barrell, B.G., deBruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1981) Sequencing and organization of the human mitochondrial genome. *Nature*, **290**, 457-465.

Antequera, F. and Bird, A. (1994) Predicting the total number of human genes. *Nature Genet.* **8**, 114.

Auffray, C., Behar, G., Bois, F., Bouchier, C., Da Silva, C., Devignes, M.D., Duprat, S., Houlgatte, R., Jumeau, M.N., Lamy, B., et al. (1995) Molecular integration of the analysis of the human genome and its expression. *C. R. Acad. Sci.* III. **318**(2), 263-272.

Bailey, Jr., L.C., Searls, D.B. and Overton, G.C. (1998) Analysis of EST-driven gene annotation in human genomic sequence. *Genome Research* **8**, 362-376.

Baltimore, D. (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**(252), 1209-1211.

Bankier, A.T., Weston, K.M. and Barrell, B.G (1987) Random cloning and sequencing by the M13/dideoxynucleotide chain termination method. *Meth. Enzymeol.* **155**, 51-93.

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

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

Berget, S.M., Moor, C., Sharp, P.A. (1977) Spliced segments at the 5' terminus of adenovirus-2 late mRNA. *Proc. Natl. Acad. Sci. USA* **74**, 3171-3175.

Berry, R., Stevens, T.J., and other 8 authors (1995) Gene-based sequence-tagged-sites (STSs) as the basis for a human gene map. *Nature Genetics*, **10**, 415-423.

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

Brack, C., and Tonegawa, S. (1977) Variable and constant parts of the immunoglobulin light chain of a mouse myeloma cell are 1250 nontranslated base apart. *Proc. Natl. Acad. Sci. USA* **74**, 5652-5656.

Breathnach, R., Mandel, J.L. and Chambon, P. (1977) Ovalbumin gene is split in chicken DNA. *Nature* **170**, 314-319.

Bronner, C.E., Baker, S.M., Morrison, P.T., Warren, G., Smith, L.G., Lescoe, M.K., Kane, M., Earabino, C., Lipford, J., Lindblom, A., et al. (1994) Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* **368**(6468), 258-261.

Bodenteich, A., Chissoe, S., Wang, Y. –F., and Roe, B.A. (1994) Shotgun cloning as the strategy of choice to generate templates for high-throughput dideoxynucleotide sequencing. In Automated DNA sequencing and analysis. Adams, M.D., Fields, C., Venter, C., eds. Academic Press, New York, pp42-45.

Boguski, M.S., Lowe, T.M. and Tolstoshev, C.M. (1993) dbEST--database for "expressed sequence tags." *Nat. Genet.* **4**(4), 332-333.

Boguski, M.S. and Schuler G.D. (1995) ESTablishing a Human Transcript Map. *Nature Genetics* **10**, 369-371.

Burge, C. and Karlin, S. (1997) Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**(1), 78-94.

Carothers, A.M., Urlaub, G., Mucha, J., Grunberger, D. and Chasin, L.A. (1989) Point mutation analysis in a mammalian gene: rapid preparation of total RNA, PCR amplification of cDNA, and Taq sequencing by a novel method. *Biotechniques* **7**(5), 494-499.

Carulli, J.P., Artinger, M., Swain, P.M., Root, C.D., Chee, L., Tulig, C., Guerin, J., Osborne, M., Stein, G., Lian, J., Lomedico, P.T. (1998) High throughput analysis of differential gene expression. *J. Cell. Biochem. suppl.* **30-32**, 286-296.

Chang, Y.D. and Dickson, R.C. (1988) Primary structure of the lactose permease gene from the yeast *Kluyveromyces lactis*. Presence of an unusual transcript structure. *J. Biol. Chem.* **263**(32), 16696-16703.

Chen, F. (1997) Sequence and analysis of approximately 0.5 MB of the DiGeorge syndrome critical region in human chromosome 22q11 and a syntenic mouse BAC. Ph.D. Thesis, University of Oklahoma.

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

Chow, L.T., Gelinas, R.E., Broker, T.R. and Roberts, R.J. (1977) An amazing sequence arrangement at the 5' ends of adenovirus-2 messenger RNA. *Cell* **12**, 1-8.

Claverie, J.M. (1997) Computational methods for the identification of genes in vertebrate genomic sequences. *Hum. Mol. Genet.* **6**(10), 1735-1744.

Cleaver, J.E. (1994) It was a very good year for DNA repair. *Cell* **76**(1), 1-4.

Colantuoni, C., Purcell, A.E., Bouton, C.M., Pevsner, J. (2000) High throughput analysis of gene expression in the human brain. *J. Neurosci. Res.* **59**(1),1 -10.

Conner, B.N., Yoon, C., Dickerson, J.L. and Dickerson, R.E. (1984) Helix geometry and hydration in a A-DNA tetramer: CCGG. *J. Mol. Biol.* **174**, 663-695.

Costanzo, F., Castagnoli, L., Dente, L., Arcari, P., Smith, M., Costanzo, P., Raugel, G., Izzo, P., Pietronaolo, T.C., Bougueleret, L., Cimino, J., Salvatore, F. and Cortese, R. (1983) Cloning of several cDNA segments coding for human liver proteins. *EMBO J.* **2**, 57-61.

Crabtree J.S. (1997) Sequencing Analysis of Regions of Human Chromosomes 11 and 22: Meningioma (22), Cat Eye Syndrome (22) and Multiple Endocrine Neoplasia, Type 1 (11). *University of Oklahoma*, Norman, Oklahoma

Crick, F.H.C. (1958) On protein synthesis. *Symp. Soc. Exp. Biol.* **12**, 548-555.

Curran, M.E., Atkinson, D.L., Ewart, A.K., Morris, C.A., Leppert, M.F., Keating, M.T.. (1993) The elastin gene is disrupted by a translocation associated with supravalvular aortic stenosis. *Cell.* **73**(1), 159-68.

Dear, S. and Staden, R. (1991) A sequence assembly and editing program for efficient management of large projects. *Nucleic Acids Res.* **19**, 3907-3911.

Delseny, M., Cooke, R., Raynal, M., Grellet, F. (1997) The Arabidopsis thaliana cDNA sequencing projects. *FEBS Lett.* **405**(2), 129-132.

Desjardins, A.E., Hohn, T.M. and McCormick, S.P. (1993) Trichothecene Biosynthesis in *Fusarium* Species: Chemistry, Genetics, and Significance. *Microbiol. Rev.* **57**, 595-604.

Deininger, P.L., (1983) Random subcloning of sonicated DNA: application to shotgun sequence analysis. *Anal. Biochem.* **129**, 216-223.

Dickerson, R.E. (1983) Base sequence and helix structure variation in B and A DNA. *J. Mol. Biol.*, **166**, 419-441.

Dunham, I., Shimizu, N., Roe, B.A., Chissoe, S., et al. (1999) The DNA sequence of human chromosome 22. *Nature* **402**, 489-495.

Esposito, D., Muresu, R., Volpi, E., Rocchi, M. and Baldini, A. (unpublished observation)

Ewart, A.K., Morris, C.A., Atkinson, D., Jin, W., Sternes, K., Spallone, P., Stock, A.D., Leppert, M. and Keating, M.T. (1993) Hemizygosity at the elastin locus in a developmental disorder, Williams syndrome. *Nat. Genet.* **5**(1), 11-16.

Ewart, A.K., Jin, W., Atkinson, D., Morris, C.A. and Keating, M.T. (1994) Supravalvular aortic stenosis associated with a deletion disrupting the elastin gene. *J. Clin. Invest.* **93**(3), 1071-1077.

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

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

Fannon, M.R. (1996) Gene expression in normal and disease states—identification of therapeutic targets. *Trends Biotechnol.* **14**(8), 294-298.

Fickett, J.W. and Hatzigeorgiou, A.Z. (1997) Eukaryotic promoter recognition. *Genome Res.* **7**(9), 861-878.

Fitzgerald, M.C., Skowron, P., van Etten, J.L., Smith, L.M. and Mead, D.A. (1992) Rapid shotgun cloning utilizing the two base recognition endonuclease CviJI. *Nucl. Acids Res.* **20**, 3753-3762.

Fishel, R., Lescoe, M.K., Rao, M.R.S., Copeland, N.G., Jenkins, N.A., Garber, J., Kane, M., Kolodner, R. (1993) The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **75**(5), 1027-1038.

Florea, L., Hartzell, G., Zhang, Z., Rubin, G.M. and Miller, W. (1998) A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res.* **8**, 967-974.

Forster, C., Kane, P.M. (2000) Cytosolic Ca2+ homeostasis is a constitutive function of the V-ATPase in Saccharomyces cerevisiae. *J. Biol. Chem.* 275(49), 38245-38253.

Fort, D.M., Barnes, C.L. and Tempesta, M.C. (1993) Two new modified trichochecenes from *Fusarium sporotrichioides. J. Natural Products* **56**, 1890-1897.

Francke, U., Hsieh, C.L., Foellmer, B.E., Lomax, K.J., Malech, H.L. and Leto, T.L. (1990) Genes for two autosomal recessive forms of chronic granulomatous disease assigned to 1q25 (NCF2) and 7q11.23 (NCF1). *Am. J. Hum. Genet.* **47**(3), 483-492.

Frangiskakis, J.M., Ewart, A.K., Morris, C.A., Mervis, C.B., Bertrand, J., Robinson, B.F., Klein, B.P., Ensing, G.J., Everett, L.A., Green, E.D., Proschel, C., Gutowski, N.J., Noble, M., Atkinson, D.L., Odelberg, S.J. and Keating, M.T. (1996) LIM-kinase1 hemizygosity implicated in impaired visuospatial constructive cognition. *Cell* **86**(1), 59-69.

Garifullina, G.F., Tag, A.G. and Beremand, M.N. (1998) Regulation of Tri10 gene expression: a complex role of Tri10 in the regulation of trichothecene biosynthesis. Taxas Genetics Society, Austin, Taxas

Gibson, S. and Somerville, C. (1993) Isolating plant genes. *Trends Biotechnol.* **11**(7), 306-313.

Gilbert-Dussardler, B. et al. (1995) A novel microsatellite DNA marker at locus D791870 detects hemizygosity in 75% of patients with Williams syndrome. *American Journal of Human Genetics*, **56**, 542-546.

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

Going, J.J., Gusterson, B.A. (1999) Molecular pathology and future developments. *Eur. J. Cancer* **35**(14), 1895-1904.

Gong, W., Emanuel, B.S., Galili, N., Kim, D.H., Roe, B., Driscoll, D. and Budarf, M.L. (1997) Structural and mutational analysis of a conserved gene (DGSI) from the minimal DiGeorge syndrome critical region. *Human Molecular Genetics* **6**, 267-276.

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

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

Grueneberg, D.A., Natesan, S., Alexandre, C., Gilman, M.Z. (1992) Human and Drosophila homeodomain proteins that enhance the DNA-binding activity of serum response factor. *Science.* **257**(5073), 1089-1095.

Grueneberg, D.A., Simon, K.J., Brennan, K., Gilman, M. (1995) Sequence-specific targeting of nuclear signal transduction pathways by homeodomain proteins. *Mol Cell Biol.* **15**(6), 3318-3326.

Grueneberg, D.A., Henry, R.W., Brauer, A., Novina, C.D., Cheriyath, V., Roy, A.L., Gilman, M. (1997) A multifunctional DNA-binding protein that promotes the formation of serum response factor/homeodomain complexes: identity to TFII-I. *Genes. Dev.* **11**(19), 2482-2493.

Gu, Z., Hillier, L., Kwok, P.Y. (1998) Single nucleotide polymorphism hunting in cyberspace. *Hum. Mutat.* **12**(4), 221-225.

Hendrix, R.W., Roberts, J.W., Stahl, F.W. and Weisberg, R.A. (1983) *Lambda II.* Cold Spring Harbor Laboratory

Henikoff, S., and Henikoff, J.G. (1992) Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* **89**, 10915-10919.

Hohn, T.M. and Beremand, P.D. (1989) Isolation and nucleotide sequence of a sesquiterpene cyclase gene from the trichothecene-producing fungus *Fusarium sporotrichioides. Gene* **79**, 131-138.

Hohn, T.M., Desjardins, A.E. and McCormick, S.P. (1995a) The *Tri4* gene of *Fusarium sporotrichioides* encodes a cytochrome P450 monooxygenase involved in trichothecene biosynthesis. *Mol. Gen. Genet.* **248**, 95-102.

Hohn, T.M., McCormick, S.P., Alexander, N.J., Desjardins, A.E. and. Proctor R.H. (1998) Function and biosynthesis of chichothecenes produced by *Fusarium* species. In: Kohmoto K. Yoder OC (eds) Molecular genetics of host-specific toxins in plant disease. Kluwer, Dordrecht, pp17-24.

Hoogenraad, C.C., Eussen, B.H., Langeveld, A., van Haperen, R., Winterberg, S., Wouters, C.H., Grosveld, F., De Zeeuw, C.I. and Galjart, N. (1998) The murine CYLN2 gene: genomic organization, chromosome localization, and comparison the human gene that is located within the 7q11.23 Williams syndrome critical region. *Genomics.* **53**(3), 348-358.

Horii, A., Han, H.J., Sasaki, S., Shimada, M. and Nakamura, Y. (1994) Cloning, characterization and chromosomal assignment of the human genes homologous to yeast PMS1, a member of mismatch repair genes. *Biochem. Biophys. Res. Commun.* **204**(3), 1257-1264.

Hornes, E. and Korsnes, L. (1990) Magnetic DNA hybridization properties of oligonucleotide probes attached to superparamagnetic beads and their use in the isolation of poly(A) mRNA from eukaryotic cells. *Genet. Anal. Tech. Appl.* 7(6), 145-150.

Hough, C.A., White, B.N., Holden, J.J. (1995) Absence of lambda immunoglobulin sequences on the supernumerary chromosome of the "cat eye" syndrome. *Am. J. Med. Genet.* **58**(3), 277-281.

Idle, J., Corchero, J., Gonzalez, F.J. (2000) Medical implication of HGP's sequence of chromosome 22. *Lancet*, **355**, 319.

Hua, A. (1999) Comparative sequencing of human and mouse syntenic regions from DGCR and MND2. Ph.D. Thesis, University of Okalahoma, Norman, Oklahoma

Jackson, R. and Standart, N. (1990) Do the poly(A) tail and 3' untranslated region control mRNA translation? *Cell* **62**(1), 15-24.

Jiang, J. and Jacob, H.J. (1998) EbEST: an automated tool using expressed sequence tags to delineate gene structure. *Genome Res.* **8**, 268-275.

Johnson, A., Minoshima, S., Asakawa, S., Shimizu, N., Shizuya, H., Roe, B.A. and McDermid, H.E. (1999) A 1.5 Mb contig within the cat eye syndrome critical region at human chromosome 22q11.2. *Genomics* **57**, 306-309.

Johannsen, W. (1909) Elemente der Exakten Erblichkeitslehre. Fischer. Jena

Katayama,S., Adachi,N., Takao,K., Nakagawa,T., Matsuda,H. and Kawamukai,M. (1995) Molecular cloning and sequencing of the hcs gene, which encodes 3-

hydroxy-3-methylglutaryl coenzyme A synthase of Schizosaccharomyces pombe. *Yeast* **11** (15), 1533-1537.

Kawasaki, K. et al. (1997) One-Megabase sequence analysis of the human immunoglobulin λ gene locus. *Genome Res.* **7**, 250-261.

Keller, N.P., Hohn, T.M. (1996) Metabolic pathway gene clusters in filamentous fungi. *Fungal Genetics and Biology* **21**, 17-29.

Khan, J., Saal. L.H., Bittner, M.L., Chen, Y., Trent, J.M. and Meltzer, P.S. (1999) Expression profiling in cancer using cDNA microarrays. *Electrophoresis* **20**(2), 223-229.

Kim, U. et al. (1997) A bacterial artificial chromosome-based framework contig map of human chromosome 22q. *Proc. Natl. Acad. Sci. USA* **93**, 6297-6301.

Kimber, W.L., Hsieh, P., Hirotsune, S., Yuva-Paylor, L., Sutherland, H.F., Chen, A., Ruiz-Lozano, P., Hoogstraten-Miller, S.L., Chien, K.R., Paylor, R., Scambler, P.J., Wynshqo-Boris, A., (1999) Deletion of 150 kb in the minimal DiGeorge/velocardiofacial syndrome critical region in mouse. *Hum. Mol. Genet.* **8**(12), 2229-2237.

Kimura, M., Kaneko, I., Komiyama, M., Takatsuki, A., Koshino, H., Yoneyama, K. and Yamaguchi, I. (1998a) Trichothecene 3-O-acetyltransferase protects both the producing organism and transformed yeast from related mycotoxins. Cloning and characterization of *Tri101*. *J. Biol. Chem.* **273**, 1654-1661.

Kimura, M., Matsumoto, G., Shingu, Y., Yoneyama, K. and Yamaguchi, I., (1998b) The mystery of the trichothecene 3-O-acetyltransferase gene. Analysis of the region around *Tri101* and characterization of its homologue from *Fusarium sporotrichioides*. *FEBS Lett.* **435** (2-3), 163-168.

Klenow, H., Overgaard-Hansen, K. and 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 accompanying cleavage. *Eur. J. Biochem*, **22**, 371-381.

Korenberg, J.R., Chen, X., Hirota, H., Lai, Z., Bellugi, U., Burian, D., Roe, B. and R. Matsuoka (2000) Genome structure and cognitive map of Williams Syndrome. *J. Cognitive Neuroscience*, **12**(3), 1-19.

Krug, M. S. and Berger, S. L. (1989) Ribonuclease H activities associated with viral reverse transcriptases are endonucleases. *Proc. Natl. Acad. Sci. USA*, **92**, 3539-3543.

Kufper, D.M.(1999) Development, analysis and use of EST database for the multicellular ascomycete, *Aspegillus nidulans*. University of Okalahoma, Norman, Oklahoma

Kupfer D, Ren Q, Lai H, White J and BA Roe (2000) Analysis of fungal EST databases. *Advances in Genome Biology and Technology I.* Marco Island, Florida, Poster 30.

Kusakabe, T., Koga, K., and 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.

Ladunga, I. (2000) Large-scale predictions of secretory proteins from mammalian genomic and EST sequences. *Curr Opin Biotechnol.* **11**(1), 13-8.

Lawson, D. (1999) Data mining parasite genomes: haystack searching with a computer. *Parasitology,* **118** Suppl:S15-18.

Lander, E.S., Rogers, J., Waterson, R. Hawkins, H., Gibbs, T., R., Sakaki, Y., weissenbach, J., Smith, D. R., Rosenthal, A., Yang, H., Hood, L., Davis, R. W. Myers, R. M., Olson, M. V., Shimizu, N., Evans, G. A., Roe, B. A., Ramser, J., McCombie, W. R., Blocker, H., Agarwala, R., Collins, F. and other members of the International Human Genome Sequencing Consortium (2001) Initial Sequencing and Analysis of the Human Genome. *Nature* **409**, 860-921.

Leach, F.S., Nicolaides, N.C., Papadopoulos, N., Liu, N.B., Jen, J., Parsons, R., Peltomäki, P., Sistonen, P., Aaltonen, L.A. Nyström-Lahti, M. et al. (1993) Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75**(6), 1215-1225.

Lisitsyn, N., Lisitsyn, N. and Wigler, M. (1993) Cloning the differences between tow complex genome. *Science* **259**, 946-951.

Lisitsyn, N.A., Lisitsina, N.M., Dalbagni, G., Barker, P., Sanchez, C.A., Gnarra, J., Linehan, W.M., Reid, B.J. and Wigler MH. (1995) Comparative genomic analysis of tumors: detection of DNA losses and amplification. *Proc. Natl. Acad. Sci. USA.* **92**(1), 151-155.

Liu, C., Raghothama, K.G. (1996) Practical method for cloning cDNAs generated in an mRNA differential display. *Biotechniques* **20**(4), 576-580.

Lomax, K.J., Leto, T.L., Nunoi, H, Gallin, J.I. and Malech, H.L. (1989) Recombinant 47-kilodalton cytosol factor restores NADPH oxidase in chronic granulomatous disease. *Science* **245**(4916), 409-412.

Lu, X., Meng, X., Morris, C.A. and Keating, M.T. (1998) A novel human gene, WSTF, is deleted in Williams syndrome. *Genomics.* **54**(2), 241-249.

Maniatis, T. and Reed, R. (1987) The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. *Nature* **325**, 673-678.

Marra, M., Hillier, L., Kucaba, T., Allen, M. and other 38 authers (1999) An encyclopedia of mouse genes. *Nature Geneics* **21**, 191-194.

Matsuo, M (1995) Duchenne muscular dystrophy. *Southeast Asian J. Trop. Med. Public Health.* **26** Suppl 1, 166-171.

Matsubara, K. and Okubo, K. (1993) cDNA analyses in the human genome project. *Gene* **135**(1-2), 265-274.

Matsumoto, G, Wuchiyama, J., Shingu, Y., Kimura, M, Yoneyama, K. and Yamaguchi, I. (1999) The trichothecene biosynthesis regulatory gene from the type B producer *Fusarium* strains: sequence of *Tri6* and its expression in *Escherichia coli. Biosci. iotechnol. Biochem.* **63**(11), 2001-2004.

Maxam, A.M. and Gilbert, W. (1977) A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA*, **74**(2), 560-564.

McCormick, S.P., Hohn, T.M. and Desjardins, A.E. (1996) Isolation and characterization of *Tri3*, a gene encoding 15-O-acetyltransferase from *Fusarium sporotrichioides. Appl. Enviro. Microbiol.* **62**, 353-359.

McCormick, S.P., Alexander, N.J., Trapp, S.E. and Hohn, T.M. (1999) Disruption of Tri101, the gene encoding trichothecene 3-O-acetyltransferase, from *Fusarium sporotrichioides. Appl. Enviro. Microbiol.* **65**(12), 5252-5256.

Mehta, R.A., Warmbardt, R.D., Mattoo, A.K. (1996)Tomato (Lycopersicon esculentum cv. pik-red) leaf carboxypeptidase: identification, N-terminal sequence, stress-regulation, and specific localization in the paraveinal mesophyll vacuoles. *Plant Cell Physiol.* **37**(6), 806-815.

Meng, X., Lu, X., Li, Z., Green, E.D., Massa, H., Trask, B.J., Morris, C.A. and Keating, M.T. (1998a) Complete physical map of the common deletion region in Williams syndrome and identification and characterization of three novel genes. *Hum. Genet.* **103**(5), 590-599.

Meng, X., Lu, X., Morris, C.A. and Keating, M.T. (1998b) A novel human gene FKBP6 is deleted in Williams syndrome. *Genomics* **52**(2), 130-137.

Metzger, A.K., Minn, Y.A., Lieberman, D., Bollen, A.W., Wilson, C.B., Feuerstein, B.G. (1999) CNS Resident Award. Malignant meningiomas frequently lose multiple chromosomal regions in addition to 22q: putative meningioma progression loci identified by comparative genomic hybridization. *Clin. Neurosurg.* **45**, 6-7.

Milner, R.J. and Sutcliffe, J.G. (1983) Gene expression in rat brain. *Nucleic Acids Res.* **11**, 5497-5520.

Morris, C.A., Leonard, C.O. and Dilates, C. (1988) Natural history of Williams syndrome: Physical characteristics. *Journal of Pediatrics*, **113**, 318-325.

Morris, C.A., Loker, J., Ensing, G. and Stock, A.D. (1993) Supravalvular aortic stenosis cosegregates with familial 6; 7 translocation which disrupts the elastin gene. *Am. J. Med. Genet.* **46**(6), 737-744.

Nakari,T., Alatalo,E. and Penttila,M.E. (1993) Isolation of Trichoderma reesei genes highly expressed on glucose-containing media: characterization of the tef1 gene encoding translation elongation factor 1 alpha *Gene* **136** (1-2), 313-318.

Nickerson, E., Greenberg, F., Keating, M.T., McCaskill, C., Shaffer, L.G. (1995) Deletions of the elastin gene at 7q11.23 occur in approximately 90% of patients with Williams syndrome. *Am. J. Hum. Genet.* **56**(5), 1156-1161.

Nowak, R. (1994) Mining treasures from 'junk DNA'. *Science* **263**, 608-610.

Nowrousian, M., Masloff, S., Poggeler, S. and Kuck, U. (1999) Cell differentiation during sexual development of the fungus Sordaria macrospora requires ATP citrate lyase activity *Mol. Cell. Biol.* **19** (1), 450-460 (1999).

Nuell, M.J., Stewart, D.A., Walker, L., Friedman, V., Wood, C.M., Owens, G.A., Smith, J.R., Schneider, E.L., Dell'Orco, R., Lumpkin, C.K., Danner, D.B. and McClung, J.K. (1991) Prohibitin, an evolutionarily conserved intracellular protein that blocks DNA synthesis in normal fibroblasts and HeLa cells. *Mol. Cell Biolo.* **11**, 1372-1381.

Nunoi, H., Rotrosen, D., Gallin, J.I. and Malech, H.L. (1988) Two forms of autosomal chronic granulomatous disease lack distinct neutrophil cytosol factors. *Science* **242**(4883), 1298-1301.

O'Donovan, C., Apweiler, R. and Bairoch, A. (2001) The human proteomics initiative (HPI). *Trends Biotechnol.* **19**(5), 178-181.

Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y., Matsubara, K., (1992) Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nat. Genet.* **2**(3), 173-179.

Okubo, K., Yoshii, J., Yokouchi, H., Kameyama, M. and Matsubara, K. (1994)An expression profile of active genes in human colonic mucosa. *DNA Res.* **1**(1), 37-45.

Olson, M., Hood, L., Cantor, C. and Botstein, D. (1989) A common language for Physical mapping of the human genome. *Science* **245**, 1434-1435.

Osborne, L.R., Martindale, D., Scherer, S.W., Shi, X.M., Huizenga, J., Heng, H.H.Q., Costa, T., Pober, B., Lew, L., Brinkman, J., Rommens, J., Koop, B. and Tsui, L.C. (1996) Identification of genes from a 500-kb region at 7q11.23 that is commonly deleted in Williams syndrome patients. *Genomics* **36**(2), 328-336.

Osborne, L.R., Soder, S., Shi, X.M., Pober, B., Costa, T., Scherer, S.W. and Tsui, L.C. (1997) Hemizygous deletion of the syntaxin 1A gene in individuals with Williams syndrome. *Am. J. Hum. Genet.* **61**(2), 449-452.

Papadopoulos, N., Nicolaides, N.C., Wei, Y.-F., Ruben, S.M., Carter, K.C., Rosen, C.A., Haseltine, W.A., Fleischmann, R.C., Fraser, C.M., Adams, M.D. et al., (1994) Mutation of a mutL homolog in hereditary colon cancer. *Science* **263**(5153), 1625-1629.

Paperna, T., Peoples, R., Wang, Y.K., Kaplan, P. and Francke, U. (1998) Genes for the CPE receptor (CPETR1) and the human homolog of RVP1 (CPETR2) are localized within the Williams-Beuren syndrome deletion. *Genomics* **54**(3), 453-459.

Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P.A. (1986) Splicing of messenger RNA precursors. *Ann. Rev. Biochem.* **55**, 1119-1150.

Pan, H. (1996) Half Million Bases of DNA Sequence in Two Meningioma Deletion Regions in Chromosome 22. *University of Oklahoma*, Norman, Oklahoma

Peoples, R., Perez-Jurado, L., Wang, Y.K., Kaplan, P. and Francke, U. (1996) The gene for replication factor C subunit 2 (RFC2) is within the 7q11.23 Williams syndrome deletion. *American Journal of Human Genetics*. **58**, 1370-1373.

Peplow, A.W., Tag, A.G., Garifullina, G.F. and Beremand, M.N. (1997) The trichothecene Tri10 gene: isolation of cDNA and effect of gene disruption on Tri gene expression and toxin production. Taxas A&M, Houston, TX, 1997.

Perez-Jurado, L.A., Wang, Y.K., Peoples, R., Coloma, A., Cruces, J. and Francke, U. (1998) A duplicated gene in the breakpoint regions of the 7q11.23 Williams-Beuren syndrome deletion encodes the initiator binding protein TFII-I and BAP-135, a phosphorylation target of BTK. *Hum. Mol. Genet.* **7**(3), 325-734.

Phadnis, S.H., Huang, H.V. and Berg, D.E. (1989) Tu5supF, A 264 base pair transposon derived from Tn5 for insertional mutagenesis and sequencing DNAs cloned in Phage λ. *Pro. Natl. Acad. Sci. USA* **86**, 5908-5912.

Pratt, R.E. and Dzau, V.J. (1999) Genomics and hypertension: concepts, potentials and opportunities. *Hypertension* **33**(1 Pt 2), 238-247.

Pradet-Balade, B., Boulme, F., Beug, H., Mullner, E.W., Garcia-Sanz, J.A. (2001) Translation control bridging the gap between genomics and proteomics? *Trends Biochem. Sci.* **26**(4), 225-229.

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

Proctor, R.H., Hohn, T.M., McCormick, S.P. and Desjardins, A.E. (1995) *Tri6* encodes an unusal zinc finger protein involved in the regulation of trichothecene biosynthesis in *Fusarium sporotrichioides. Appl. Environ. Microbiol.* **61**, 1923-1930.

Poll, F.M. and Jovin, T.M. (1972) Salt-induced co-operative conformational change of a synthetic DNA: Equilibrium and kinetic studies with poly(dG-dC). *J. Mol. Biol.* **67**, 375-396.

Putney, S.D., Herlihy, W.C. and Schimmel, P. (1983) A new troponin T and cDNA clones for 13 different muscle proteins, found by shotgun sequencing. *Nature* **302**, 718-721.

Qu, R.P. and Palta, M. (1996) Using projection for testing goodness-of-fit in regression models for repeated measures. *Biometrics*, **52**(4), 1259-1267.

Rafalski, J.A., Hanafey, M., Miao, G.H., Ching, A., Lee, J.M., Dolan, M. and Tingey, S. (1998) New experimental and computational approaches to the analysis of gene expression. *Acta. Biochim. Pol.* **45**(4), 929-934.

Rao, C.R. (1959) Some problems involving linear hypotheses in multivariate analysis. *Biometrika*, 46, 49-58.

Rawlings, D.J., Saffran, D.C., Tsukada, S., Largaespada, D.A., Grimaldi, J.C., Cohen, L., Mohr, R.N., Bazan, J.F., Howard, M., Copeland, N.G., et al., Mutation of unique region of Bruton's tyrosine kinase in immunodeficient XID mice. *Science* **261**(5119), 358-361.

Reese, M.G. and Eeckman, F.H. (1995) Novel Neural Network Algorithms for Improved Eukaryotic Promoter Site Recognition. Accepted talk for The seventh international Genome sequencing and analysis conference, Hyatt Regency, Hilton Head Island, South Carolina September 16-20, 1995

Rezvani, M., Barrans, J.D., Dai, K.S. and Liew, C.C. (2000) Apoptosis-related genes expressed in cardiovascular development and diseases: an EST approach. *Cardiovasc. Res.* **45**(3), 621-629.

Rich, A., Nordheim, A. and Wang, A.H.-J. (1984) The chemistry and biology of left-handed Z-DNA. *Ann. Rev. Biochem.* **53**, 791-846.

Riley, M. and 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, V.M., Sheng, M. and Greenberg, M.F. (1990) The inner core of the serum response element mediates both the rapid induction and subsequent repression of c-fos transcription following serum stimulation. *Genes. Dev.* **4**(2), 255-268.

Roach., J. C., Siegal, A.F., van den Engh, G., Trask, B., Hood, L., (1999) Gaps in the human genome project. *Nature* **401,** 843-845.

Robinson, W.P., Waslynka, J., Bernasconi, F., Wang, M., Clark, S., Kotzot, D. and Schinzel, A. (1996) Delineation of 7q11.2 deletions associated with Williams-Beuren syndrome and mapping of a repetitive sequence to within and to either side of the common deletion. *Genomics* **34**(1), 17-23.

Robledo, R., Orru, S., Lucito, R., Grimaldi, M.C., Giuditta, R., Carcassi, C., Bernini, L., Beck, J.C., Rinaldi, A., Contu, L., Kidd, J., Kidd, K., Lisitsyn, N., Wigler, M., Siniscalco, M. (unpublished observation)

Rodaway, A.R., Teahan,C.G., Casimir,C.M., Segal,A.W. and Bentley,D.L. (1990) Characterization of the 47-kilodalton autosomal chronic granulomatous disease protein: tissue-specific expression and transcriptional control by retinoic acid. *Mol. Cell. Biol.* **10** (10), 5388-5396.

Roe, B.A. (1997) Protocols for recombinant DNA isolation, cloning, and sequencing. The University of Oklahoma, Norman

Rosenthal, A., Coutelle, O. and Craxton, M. (1993) Large-scale production of DNA sequencing templates by microtitre format PCR. *Nucleic Acids Res.* **21**(1), 173-174.

Rushizky, G.W., Shaternikov, V.A., Mozejko, J.H. and Sober, H.A. (1975) S1 nuclease hydrolysis of single-stranded nucleic acids with partial double-stranded configuration. *Biochemistry*, **14**(19), 4221-4226.

Ruttledge, M.H., Xie, Y.G., Han, F.Y., Peyrard, M., Collins, V.P., Nordenskjold, M., Dumanski, J.P. (1994) Deletions on chromosome 22 in sporadic meningioma. *Genes Chromosomes Cancer.* **10**(2), 122-130.

Ryu, D.D. and Nam, D.H. (2000) Recent progress in biomolecular engineering. *Biotechnol. Prog.* **16**(1), 2-16.

Sachs, A. and Wahle, E. (1993) Poly(A) tail metabolism and function in eucaryotes. *J. Biol. Chem.*, **268**(31), 22955-22958.

Sanger, F., Nicklen, S. and Coulson, A.R. (1997) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, **74**(12), 5463-5467.

Sato, T., Tanigami, A., Yamakawa, K., Akiyama, F., Kasumi, F., Sakamoto, G., and Nakamura, Y. (1990) Allelotype of breast cancer: Cumulative allele losses promote tumor progression in primary breast cancer. *Cancer Res.* **50**, 7184-7189.

Sato, T., Akiyama, F., Sakamoto, G., Kasumi, F. and Nakamura, Y. (1991) Accumulation of genetic alterations and progression of primary breast cancer. *Cancer Res.* **51**, 5794-5799.

Sato, T., Saito, H., Swensen, J., Olifant, A., Wood, C., Danner, D., Sakamoto, T., Takita, K., Kasumi, F., Miki, Y., et al. (1992) The human prohibitin gene located on chromosome 17d21 is mutated in sporadic breast cancer. *Cancer Res.* **52**(6), 1643-1646.

Saouaf, S.J., Mahajan, S, Rowley, R.B., Kut, S.A., Fargnoli, J., Burkhardt, A.L., Tsukada, S., Witte, O.N., Bolen, J.B. (1994) Temporal differences in the activation of three classes of non-transmembrane protein tyrosine kinases following B-cell antigen receptor surface engagement. *Proc. Natl. Acad. Sci. USA.* **91**(20), 9524-9528.

Schena, M., Heller, R.A., Theriault, T.P., Konrad, K., Lachenmeier, E. and Davis, R.W. (1998) Microarrays: biotechnology's discovery plateform for functional genomics. *Trends Biotechnol.* **16**(7), 301-306.

Schneiderman, E.D. and Kowalski, C.J. (1985) Implementation of Rao's one-sample polynomial growth curve model using SAS. *Am. J. Phys. Anthropol.* **67**(4), 323-333.

Schneiderman, E.D. , Willis, S.M., Ten Have, T.R. and Kowalski, C.J. (1991) Rao's polynomial growth model for unequal-time intervals: a menu-driven GAUSS program. *Int. J. Biomed. Comput.* **29**(3-4), 235-244.

Schneiderman, E.D. , Willis, S.M., Kowalski, C.J. and Ten Have, T.R. (1993) Longer-Term Growth prediction using GAUSS. *Comput. Biol. Med.* **23**, 149-154.

Schuler, G.D. et al. (1996) A gene map of the human genome. *Science* **274**, 540-546.

Selkov, E. Jr, Grechkin, Y., Mikhailova, N. and Selkov, E., (1998) MPW: the Metabolic Pathways Database. *Nucleic Acids Res.* **26**(1), 43-45.

Shoemaker, D.D., Schadt, E.E., Armour, C.D., He, Y.D., Garrett-Engele, P., McDonagh, P.D., Loerch, P.M., Leonardson, A., Lum, P.Y., Cavet, G. et al. (2001) Experimental annotation of the human genome using microarray technology. *Nature* **409**,922-927.

Short, J.M., Fernandez, J.M., Sorge, J.A. and Huse, W.D. (1988) λ Zap: a bacteriophage λ expression vector with *in vivo* excision properties. *Nucleic Acids Res.* **16**, 7583-7600.

Siniscalco, M., Robledo, R., Orru, S., Contu, L., Yadav, P., Ren, Q., Lai, H. and Roe, B.A. (2000) A phenotypically silent inborn gap of 9kb is detected in the reference megabase sequence of the Ig λ light chain genes: a plea to search for deletion polymorphism through genome scans in populations. *Trends in Genetics.* **16**, 435-467.

Smith L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B. and Hood, L.E. (1986) Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674-679.

Smith, T.F. and Waterman, M.S. (1981) Identification of common molecular subsequences. *J. Mol. Biol.* **147**(1), 195-197.

Smith, D.P., Johnstone, E.M., Little, S.P. and Hsiung, H.M. (1990) Direct DNA sequencing of cDNA inserts from plaques using the linear polymerase chain reaction. *Biotechniques*, **9**(1), 48-52.

Soares, M.B., Bonaldo, M.F., Jelenc, P., Su, L., Lawton, L. and Efstratiadis, A. (1994) Construction and characterization of a normalized cDNA library. *Proc. Nat. Acad. Sci., USA.* **91**, 9228-9232.

Sonnhammer, E.L. and Durbin, R. (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene*, **167**(1-2), GC1-10.

Strand, M., Prolla, T.A., Liskay, R.M., and Petes, T.D. (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**(6443), 274-276.

Sutherland, H.F., Kim, U.J. and Scambler, P.J. (1998) Cloning and comparative mapping of the DiGeorge syndrome critical region in the mouse. *Genomics* **52**(1), 37-43.

Tabor, S. and Richardson, C.C. (1989) Selective inactivation of the exonuclease activity of bacteriophage T7 DNA polymerase by in vitro mutagenesis. *J. Biol. Chem.* **264**(11), 6447-6458.

Tag, A.G. and Beremand, M.N. (1998) Effect of constitutive expression of the trichothecene Tri10 gene on T-2 Toxin production. Taxas Genetics Society, Austin, TX

Temin, H.M. and Mizutani, S. (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**(252), 1211-1213.

Thomas, J.D., Sideras, P., Smith, C.I., Vorechovsky, I., Chapman, V., Paul, W.E. (1993) Colocalization of X-linked agammaglobulinemia and X-linked immunodeficiency genes. *Science* **261**(5119), 355-358.

Treisman R (1985) Transient accumulation of c-fos RNA following serum stimulation requires a conserved 5' element and c-fos 3' sequences. *Cell* **42**(3), 889-902.

Treisman R (1986) Identification of a protein-binding site that mediates transcriptional response of the c-fos gene to serum factors. *Cell* **46**(4), 567-574.

Tsukada, S., Saffran, D.C., Rawlings, D.J., Parolini, O., Allen, R.C., Klisak, I., Sparkes, R.S., Kubagawa, H., Mohandas, T., Ouan, S., et al (1993) Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia. *Cell* **72**(2), 279-290.

Ueno, Y. (1980) Trichothecene mycotoxins: mycology, chemistry, and toxicology. *Adv. Nutr. Sci.* **3**, 301-351.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. et al. (2001) The sequence of the human genome. *Science* **291**,1304-1351.

Vetrie, D., Vorechovsky, I., Sideras, P., Holland, J., Davies, A., Flinter, F., Hammarstrom, L., Kinnon, C., Levinsky, R., Bobrow, M., et al. (1993) The gene involved in X-linked agammaglobulinaemia is a member of the src family of protein-tyrosine kinases. *Nature* **361**(6409), 226-233.

Volpp, B.D., Nauseef, W.M., Donelson, J.E., Moser, D.R. and Clark, R.A. (1989) Cloning of the cDNA and functional expression of the 47-kilodalton cytosolic component of human neutrophil respiratory burst oxidase. *Proc. Nat. Acad. Sci., USA.* **86**(23), 7195-7199. Correction. *Proc. Nat. Acad. Sci. USA* **86**, 9563.

Waibel, A. H., Hanazawa, T., Hinton, G. E., Shikano, K., Lang, K. J. (1989) Phoneme Recognition Using Time-Delay Neural Networks. *IEEE Transactions on Acoustic, Speech, and Signal Processing.* 37(3), 328-339.

Wang, A.H.-J., Fujii, S., van Boom, J.H. and Rich, A. (1982) Molecular structure of the octamer d(G-G-C-C-G-G-C-C-): modified A-DNA. *Proc. Nat. Acad. Sci. USA* **79**, 3968-3972.

Wang, A.H.-J., Quigley, G.J., Kolpak, F.J., Crowford, J.L., van Boom, J.H., van der Marel, G. and Rich, A. (1979) Molecular structure of a left-handed DNA fragment at atomic resolution. *Nature* **282**, 680-686.

Wang, Y.K., Samos, C.H., Peoples, R., Perez-Jurado, L.A., Nusse, R. and Francke, U. (1997) A novel human homologue of the Drosophila frizzled wnt receptor gene binds wingless protein and is in the Williams syndrome deletion at 7q11.23. *Hum. Mol. Genet.* **6**(3), 465-472.

Watson, J.D. and Crick, F.H.C. (1953) Molecular structure of nucleic acid. A structure of the deoxyribose nucleic acid. *Nature* **171**, 737-738.

Wells, T.N. and Peitsch, M.C. (1997) The chemokine information source: identification and characterization of novel chemokines using the WorldWideWeb and expressed sequence tag databases. *J. Leukoc. Biol.* **61**(5), 545-550.

White, J.J., Ledbetter, D.H., Eddy, R.L., Shows, T.B., Stewart, D.A., Nuell, M.J., Friedman, V., Wood, C.M., Owens, G.A., McClung, J.K., Danner, D.B., and Morton, C.C. (1991) Assignment of the human prohibitin gene (PHB) to chromosome 17 and identification of a DNA polymorphism. *Genomics*, **11**, 228-230.

Wigler, M. and Lisitsyn, N: (unpublished observation)

Wise, J.A. (1993) Guide to the heart of the splicesome. *Science* **262**, 1978-1979.

Wootton, J.C., and Federhen, S., (1993) Statistics of local complexity in amino acid sequences and sequence databases. *Comput. Chem.* **17**, 191-201.

Xu, Y., Mural, R.J., Shah, M. and Uberbacher, E.C. (1994) Recognizing exons in genomic sequence using GRAIL II. *Genetic Engineering* **16**, 241-253.

Yamamoto, K. and Sasaki, T. (1997) Large scale EST sequence in rice. *Plant Mol. Biol.* **35**(1-2), 135-144.

Yanagisawa, Y., Ito, E., Iwahashi, Y., Akiyama, Y., Yuasa, Y. and Maruyama, K. (1998) Isolation and characterization of the 5' region of the human mismatch repair gene hPMS1. *Biochem. Biophys. Re.s Commun.* **243**(3), 738-743.

Yang, W. and Desiderio, S. (1997) BAP-135, a target for Bruton's tyrosine kinase in response to B cell receptor engagement. *Proc. Natl. Acad. Sci. USA* **94**(2), 604-609.

Zhang, G. (1997) Sequence and Analysis of the Human Chromosome 22 Band Q11 DiGeorge Syndrome Region University of Oklahoma, Norman, Oklahoma

Zhang, J. and Madden, T.L. (1997) PowerBLAST: a new network BLAST application for interactive or automated sequence analysis and annotation. *Genome Research*, **7**, 649-656.

Zweiger, G. and Scott, R.W. (1997) From expressed sequence tags to 'epigenomics': an understanding of disease processes. *Curr. Opin. Biotechnol.* **8**(6), 684-687.

Appendix I. Keyword list of *Fusarium sporotrichioides* categories of biological functions. Keywords are in standard type and headings are in bold. A keyword after the & indicates that it is a variation of the keyword in the line above. One keyword may have one to several variations. All entries matching the variations for one keyword will be placed under this higher keyword in appendix II.

PART I. Metabolic Pathways

I. Metabolism of Carbohydrates(for glucose see energy)

1. Chitin metabolism (6)
chitinase
&EC 3.2.1.14
chitinase precursor
chitin synthase

2. Cellulose metabolism (8)
beta glucosidase-breakdown of cellulose
&beta glucosidase
&beta-glucosidase
&glucoside glucohydrolase
glucanase
&GNS1 PROTEIN
&beta glucanase
endo-beta-1,6-glucanase
cellulase
&ENDOGLUCANASE TYPE K PRECURSOR
&ENDO-1,4-BETA-GLUCANASE
bcsABII-B-cellulose synthase genes
&bcsABII-B

3. Cutin metabolism (4)
cutinase G-box binding protein

4. Polysaccharide metabolism (3)
UDP-glucose dehydrogenase
&UDP-glucose 6-dehydrogenase
utp--glucose-1-phosphate uridylyltransferase

5. Galactose metabolism (10)
galactose oxidase
GALACTOSE OXIDASE PRECURSOR (GAO)
GALACTOSIDE O-ACETYLTRANSFERASE
&THIOGALACTOSIDEACETYLTRANSFERASE
alpha-1,4 polygalactosaminidase
GALACTOSYLTRANSFERASE

beta-galactosidase
UDP glucose NAD dependant epimerase/dehydratase

6. Mannitol and mannose metabolism (14)
mannose-1-phosphate guanylyltransferase
&mannose-1-phosphate gaunyl transferase
dolichyl phosphate-D-mannose:proteinO-D-mannosyltransferase
&DOLICHYL-PHOSPHATE-MANNOSE:PROTEIN MANNOSYLTRANSFERASE
&DOLICHYL-PHOSPHATE-MANNOSE--PROTEIN MANNOSYLTRANSFERASE
&(EC 2.4.1.109)
mannose-6-phosphate isomerase
&MANNOSE-6-PHOSPHATE ISOMERASE
mannosyl-oligosaccharide1,2-alpha-mannosidase
Alg2-encoding a mannosyltransferas
&Alg2
endoplasmic reticulum alpha-mannosidase

7. Quinate metabolism (4)
quinate
&quinate
pentafunctional arom polypeptide

8. Sorbitol metabolism (2)
SORBITOL UTILIZATION PROTEIN

9. others (2)
aldose reductase
L-sorbose dehydrogenase
D-xylose reductase
squalene synthase
B2-aldehyde-forming enzyme

II. Metabolism of Amino acids and Related Molecules

1. arginine metabolism (9)
arginine N-methyltransferase

1.1 arginine anabolism-glutamine, CO2 to arginine
ARGININOSUCCINATE SYNTHASE
&argininosuccinate synthetase
ARGININOSUCCINATE LYASE

1.2 arginine catabolism-arginine to proline
ARGINASE-also see urea cycle
&ARGINASE
ARG-6 PROTEIN PRECURSOR

2. asparagine metabolism (3)
asparagine synthetase
L-asparaginase II
L-ASPARAGINASE

3. aspartic acid metabolism (5)
ASPARTATE AMINOTRANSFERASE

4. cysteine metabolism and biosynthesis (4)
S-adenosyl-L-homocysteine hydrolase
CYSTATHIONINE GAMMA-SYNTHASE
CYSTEINE SYNTHASE
&O-ACETYLSERINE SULFHYDRYLASE
&O-ACETYLSERINE (THIOL)-LYASE
&CSASE

5. glutamine metabolism and biosynthesis (4)
glutamine synthase
glutamine amidotransferase
glutamyl-trna synthetase

6. glycine metabolism (6)
THREONINE ALDOLASE
THREONINE SYNTHASE
GLYCINE DEHYDROGENASE

7. histidine metabolism (6)
HISTIDINOL-PHOSPHATASE
HISTIDINOL-PHOSPHATE AMINOTRANSFERASE
imidazole glycerol phosphate dehydratase-biosynthesis of histidine

8. isoleucine metabolism (5)
2,3-DIHYDROXYACID HYDROLYASE-4th step in iso & val biosyn
&DIHYDROXY-ACID DEHYDRATASE
acetolactate synthase-biosynthesis of isoleucine, leucine and valine
&ACETOLACTATE SYNTHASE
&acetolactate synthase
ISOPROPYLMALATE DEHYDRATASE-biosynthesis of isoleucine, leucine and valine
&ISOPROPYLMALATE ISOMERASE

9. methionine metabolism (4)
methionine synthase-last step in met biosynthesis
&methionine synthase
S-adenosylmethionine synthetase-(EC 2.5.1.6)

10. tryptophan metabolism and synthesis (3)
anthranilate phosphoribosyltransferase-2nd step in tryp biosyn
&anthranilate phosphoribosyltransferase
&phosphoribosylanthranilate transferase
anthranilate synthase Component I-synthesis of tryptophan
&anthranilate synthase Component I

11. aromatic amino acid metabolism (1)
aromatic-L-amino-acid decarboxylase

12. glutamate metabolism (8)
glutamate synthase(NADH)precursor
glutamate synthase
glutamic acid decarboxylase
acetylglutamate kinase

13. phenylalanine metabolism (3)
phosphoethanolamine cytidylyltransferase
&ethanolamine-phosphate cytidylyltransferase
&(EC 2.7.7.14)
HOMOGENTISATE 1,2-DIOXYGENASE

14. lysine (4)
alpha-aminoadipate reductase
homocitrate synthase-for biosynthesis of lysine

15. tyrosine metabolism (1)
TYROSINASE
&MONOPHENOL MONOOXYGENASE

16. alanine metabolism (1)
alanine racemase

17. others (3)
orfT-transaminase type I
&orfT
aminobutyrate aminotransferase
& 4-AMINOBUTYRATE AMINOTRANSFERASE
&GAMMA-AMINO-N-BUTYRATETRANSAMINASE
&GABA TRANSAMINASE
&GABA AMINOTRANSFERASE

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

1. Nucleotide metabolism (5)
NUCLEOSIDE DIPHOSPHATE KINASE

ribose-phosphate pyrophosphokinase-purine,pyrimidine biosyn, also his and tryptophan biosyn
ribonucleotide reductase

2. Purine metabolism (9)

2.1. inosine mono phosphate de novo biosynthesis
amidophosphoribosyltransferase
&AMIDOPHOSPHORIBOSYLTRANSFERASE
&GLUTAMINE PHOSPHORIBOSYLPYROPHOSPHATE AMIDOTRANSFERASE
&ATASE
&phosphoribosylpyrophosphate amidotransferase
inosine 5'-monophosphate dehydrogenase
5-aminoimidazole-4-carboxamideribonucleotide (AICAR) transformylase-biosynthesis of inosine monophosphate

2.2. other purine metabolic enzymes
INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE
aminoimidazole ribonucleotide carboxylase-purine biosynthesis enzyme
&phosphoribosylaminoimidazole carboxylase
&aminoimidazole ribonucleotide carboxylase
&EC 4.1.1.21
&&PHOSPHORIBOSYLAMINOIMIDAZOLE CARBOXYLASE

3. Pyrimidine metabolism (3)
orotate reductase
&(EC 1.3.1.14)
orotidine-5'-phosphate decarboxylase
&EC 4.1.1.23

4. Salvage(reuse) of the bases (1)
ATP PHOSPHORIBOSYLTRANSFERASE-synthesizes nucleoside 5'-phosphates directly from free base
&ATP PHOSPHORIBOSYLTRANSFERASE

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

1. Fatty acid biosynthesis (13)
acetyl-CoA carboxylase
&(EC 6.4.1.2)
fatty acid synthase
&fatty acid synthase, beta subunit
stearoyl-CoA desaturase-adds double bonds to fatty acyl coA
&stearoyl-CoA desaturase
fatty acid (lipid) desaturase
beta-ketoacyl reductase

2. sterols (23)
C-3 sterol dehydrogenase
sterol C-14 reductase
&C-14 sterol reductase
sterol reductase
delta-12 desaturase
delta7-sterol-C5-desaturase

2.1. General
sterol c-methyltransferase
&STEROL C-METHYLTRANSFERASE
HYDROXY-3-METHYLGLUTARYL-COENZYME A REDUCTASE-also mevalonate biosyn to isoprenoids
&HMG-CoA-reductase
&HYDROXY-3-METHYLGLUTARYL-COENZYME A REDUCTASE

2.2. Farnesol biosynthesis
GERANYLGERANYL PYROPHOSPHATE SYNTHETASE
&dimethylallyltransferase
hydroxysteroid dehydrogenase
FARNESYL PYROPHOSPHATE SYNTHETASE
&FPP SYNTHETASE
&FPS
&FARNESYL DIPHOSPHATE SYNTHETASE

2.3. cholesterol and steroids
C-4 METHYL STEROL OXIDASE-cholesterol biosynthesis
&C-4 METHYL STEROL OXIDASE
isopentenyl-diphosphate Delta-isomerase
&ISOPENTENYL-DIPHOSPHATE DELTA-ISOMERASE-ISOPRENE-CHOLESTEROL BIOSYNTHESIS
&ISOPENTENYL-DIPHOSPHATE DELTA-ISOMERASE
&(EC 5.3.3.2)
ergosterol synthesis
ergosterol biosynthesis
phosphomevalonate kinase
MEVALONATE KINASE
&mevalonate kinase
diphosphomevalonate decarboxylase
&DIPHOSPHOMEVALONATE DECARBOXYLASE
&MEVALONATEPYROPHOSPHATE DECARBOXYLASE

3. lipids (18)

3.1. phospholipid metabolism
phospholipid methyltransferase

phosphoinositide-specific phospholipase
lysophospholipase
phosphatidyl synthase
phosphatidylserine decarboxylase
&PHOSPHATIDYLSERINE DECARBOXYLASE proenzyme1 precursor
myo-inositol phosphate synthase-biosynthesis of inositol containing phospholipids
&myo-inositol-1-phosphate synthase
&myo-inositol phosphate synthase
&myo-inositol 1-phosphate synthase
&myo-inositol-3-phosphate synthase
inositol 1-phosphate synthase-first enzyme in inositol biosynthetic pathway
&inositol 1-phosphate synthase
1-phosphatidylinositol phosphodiesterase
&(EC 3.1.4.10)
CDP-diacylglycerol synthase
annexin V-binding protein-calcium-dependent phospholipid binding proteins
&annexin V-binding protein

3.2. sphingolipids
serine palmitoyltransferase
sphingosine phosphate lyase
&dihydrosphingosine phosphate lyase

3.3. lipopolysaccharide biosyn-biomembrane precursors
UDP-glucose:sterol glucosyltransferase

V. Sulfur, Phosphate and Nitrogen Metabolism

1. Sulfur Metabolism (7)
sulphur metabolite repression
&sconCp
sulfate adenylyltransferase-leads to biosynthesis of cys&met
&ATP-SULFURYLASE
&SULFATE ADENYLATETRANSFERASE
&SAT
sulfite reductase
&Sulfite Oxidase

2. Nitrogen Metabolism (see also amino acid metabolism) (16)
nitrite reductase
spermidine synthase
4-nitrophenylphosphatase
&NIPSNAP
MeaB protein-affecting nitrogen metabolite repression
DRAP deaminase
2-methyl-3-hydroxypyridine-5-carboxylic acidoxygenase

-urea related
urease
urea amidolyase
URICASE
UREA ACTIVE TRANSPORTER
ureidoglycolate hydrolase
&UREIDOGLYCOLATE HYDROLASE

VI. Metabolism of Cofactors, prosthetic groups

1.nicotinamide coenzymes (5)
NICOTINATE-NUCLEOTIDE PYROPHOSPHORYLASE-DE NOVO BIOSYNTHESIS OF NAD AND NADP
kynureninase-biosyn of NAD cofactors
&kynureninase
&alpha-aminoadipate aminotransferase
nicotinate phosphoribosyltransferase
&nicotinate phosphorybosyltranferase
sir2-related protein type 7-metabolizes NAD
&sir2-related protein type 7

2.biocytin (biotin) (2)
BIOTIN SYNTHASE
&BIOTIN SYNTHETASE

3.thiamine (3)
thiamine-4
THIAMINE BIOSYNTHETIC BIFUNCTIONAL ENZYME

4.coenzyme A (3)
acetyl-coenzyme A synthetase
&ACETYL-COENZYME A SYNTHETASE
&ACETATE-COA LIGASE
&ACYL-ACTIVATING ENZYME

5.flavins (1)
GTP cyclohydrolase

6. heme (14)
heme protein precursor
siroheme synthase
-iron uptake
ferric reductase
&FERRIC REDUCTASE TRANSMEMBRANE COMPONENT 2
&FERRIC REDUCTASE TRANSMEMBRANE COMPONENT 1 PRECURSOR

delta-aminolevulinic acid dehydratase-heme biosynthesis
&DELTA-AMINOLEVULINIC ACID DEHYDRATASE
&PORPHOBILINOGENSYNTHASE
&ALADH
flavohemoglobin
Iron-sulfur cluster nifU-like protein
AMINOLEVULINIC ACID DEHYDRATASE-heme biosynthesis
&AMINOLEVULINIC ACID DEHYDRATASE
&PORPHOBILINOGEN
&porphobilinogen synthase
&5-AMINOLEVULINIC ACID SYNTHASE
siderophore biosynthesis protein pbsC

7. Molybdopterin (1)
MOLYBDOPTERIN BIOSYNTHESIS CNX1 PROTEIN
&MOLYBDENUMCOFACTOR BIOSYNTHESIS ENZYME CNX1

8. PMP (3)
PYRIDOXAMINE 5'-PHOSPHATE OXIDASE
&PNP/PMP OXIDASE

9. others (1)
cofactor required for Sp1 transcriptionalactivation subunit 2

VII. Energy

VII.1. Carbohydrate as energy source

1. Glycolysis (27)
hexokinase
GLUCOKINASE
&glucose kinase
6-PHOSPHOFRUCTOKINASE
&PHOSPHOFRUCTOKINASE
&PHOSPHOHEXOKINASE
glucose-6-phosphate isomerase
aldolase
&fructose-bisphosphate aldolase
triose-phosphate isomerase
glyceraldehyde-3-phosphate dehydrogenase
&GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE 2
&glyceraldehyde-3-phosphate dehydrogenase
phosphoglycerate kinase
pyruvate kinase
2,3-bisphosphoglycerate-independentphosphoglycerate mutase
enolase

&2-PHOSPHO-D-GLYCERATE HYDRO-LYASE
&2-PHOSPHOGLYCERATE DEHYDRATASE

2. Gluconeogenesis (2)
pyruvate carboxylase
&pyruvate carboxylase
fructose-1,6-bisphosphatase
&FRUCTOSE-1,6-BISPHOSPHATASE
&FRUCTOSE-1,6-BISPHOSPHATASE
&FRUCTOSE-1,6-BISPHOSPHATAS
&fructose-bisphosphatase

3. Pentose-phosphate pathway (12)
glucose-6-phosphate dehydrogenase
&GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE
&G6PD
6-phosphogluconolactonase
&gluconeolactonase
&6-PHOSPHOGLUCONOLACTONASE
phosphogluconate dehydrogenase
&6-PHOSPHOGLUCONATE DEHYDROGENASE
ribulose-phosphate 3-epimerase
&ribulose-5-phosphate 3-epimerase
transketolase
transaldolase

4. Pyruvate dehydrogenase-three kinds of enzymes (7)
pyruvate dehydrogenase
&pyruvatedehydrogenase complex
&PYRUVATEDEHYDROGENASE COMPLEX
dihydrolipoamide acetyltransferase
dihydrolipoamide dehydrogenase

5. Tricarboxylic acid (TCA) cycle (21)
citrate synthase
aconitase
isocitrate dehydrogenase
alpha-ketoglutarate dehydrogenase
succinyl-coa synthetase
&succinyl-CoA ligase
FUMARATE HYDRATASE
malate dehydrogenase

6. related reactions (3)
ATP citrate lyase
#&CITRATE (PRO-S-)-LYASE

7. Fermentation, alcoholic (10)
pyruvate decarboxylase
alcohol dehydrogenase
toluenesulfonate zinc-independent alcohol dehydrogenase

8.Fermentation, other (1)
LACTATE DEHYDROGENASE-pyruvate to lactate
&LACTATE DEHYDROGENASE
butanediol dehydrogenase

9. Monocarbon metabolism (5)
formate dehydrogenase

C1 Metabolism
methylenetetrahydrofolate reductase
&METHYLENETETRAHYDROFOLATE REDUCTASE
&FADH2
c-1-tetrahydrofolate synthase

10. Metabolism of energy reserves (glycogen, starch, trehalose) (12)

10.1. Glycogen degradation
glycogen phosphorylase
phosphoglucomutase-glycogen deg, glu1PO4 to glu6PO4
&PHOSPHOGLUCOMUTASE
&GLUCOSE PHOSPHOMUTASE
GLYCOGEN DEBRANCHING ENZYME
&4-alpha-glucanotransferase

10.2. Starch degradation
mannosyl-oligosaccharide glucosidase
&MANNOSYL-OLIGOSACCHARIDE GLUCOSIDASE
&PROCESSINGA-GLUCOSIDASE I
nucleoside diphosphate-sugar hydrolase

10.3. Trehalose degradation
trehalose-6-phosphate phosphatase

VII.2. fatty acid as energy source

1. lipase-triacylglycerols to glycerol+FA (2)
lipase

2. beta-oxydation of fatty acids (6)
long-chain-fatty-acid-CoA ligase

&long-chain-fatty-acid--coa ligase
&long-chain fatty acid--CoA ligase andsynthetase 4
carnitine/acyl carnitine carrier
L-carnitine dehydratase (caiB-2)
3-ketoacyl-CoA thiolase
&3-keto-acyl-CoA thiolase
&3-KETOACYL-COA THIOLASE

3. Ketone body metabolism (2)
SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASE-acetoacetate to acetoacyl-coA
&SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASE
HYDROXYMETHYLGLUTARYL-COA SYNTHASE-ketogenesis
&HMG-COA SYNTHASE
&3-HYDROXY-3-METHYLGLUTARYL COENZYME A SYNTHASE

VII.3. Metabolism of other energy sources (29)
glutathione-dependent formaldehyde dehydrognease
&GLUTATHIONE-DEPENDENT FORMALDEHYDE DEHYDROGENASE
&FDH
&FALDH
&EC 1.1.1.1
aldehyde reductase
&aldehydereductase
glycerol-3-phosphate dehydrogenase
acetamidase-allows acetamide and formamide as sole C or N source
&ACETAMIDASE
&amidase
ALDEHYDE DEHYDROGENASE-broad substrate specificity
&ALDEHYDE DEHYDROGENASE
aldehyde-dehydrogenase-like protein
KETOL-ACID REDUCTOISOMERASE PRECURSOR
&alpha-keto-beta-hydroxylacylreductoisomerase
&ALPHA-KETO-BETA-HYDROXYLACIL REDUCTOISOMERASE
Acetyl-CoA-Acetyltransferase
&acetyl-coa acetyltransferase
OMEGA-6 FATTY ACID DESATURASE
&ENDOPLASMIC RETICULUMISOZYME
succinate-semialdehyde dehydrogenase

VII.4. Electron transport

1. Complex I-NADH-ubiquinone (23)
NADH:Ubiquinone Oxidoreductase
&NADH dehydrogenase
&NADH-dehydrogenase

&ubiquinone
&NADH-UBIQUINONE OXIDOREDUCTASE(EC 1.6.5.3)
&NADH-UBIQUINONE OXIDOREDUCTASE B22 SUBUNIT
&COMPLEXI-B22
&CI-B22
&NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 2
&NADH-UBIQUINONE OXIDOREDUCTASE 30.4 KD SUBUNIT PRECURSOR
&(EC 1.6.5.3)
electron transfer protein

2. Complex II-Succinate-ubiquinone (6)
succinate dehydrogenase

3. Complex III-Ubiquinone to cytochrome C (3)
cytochrome b5
cytochrome c oxidase

4. Other electron transport pathways (29)
NADH oxidase
&NADH OXIDASE
GLUTATHIONE REDUCTASE
NADH-cytochrome b5 reductase
electron transfer flavoprotein alpha-subunit precursor
&ELECTRON TRANSFER FLAVOPROTEIN ALPHA-SUBUNITPRECURSOR
ubiquinone biosynthesis
ubiquinone/menaquinone biosynthesis methyltransferase
Thioredoxin Reductase
NADPH-ferrihemoprotein reductase
&(EC 1.6.2.4)
Nfrl-neurula-specific ferredoxin reductase-like protein
&Nfrl
ubiquinol-cytochrome c reductase complex
&ubiquinol-cytochrome c reductase
&ubiquinol-cytochrome c reductase iron-sulfur subunit
&UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX CORE PROTEIN
cytochrome oxidase
flavoprotein
NADH-dependent flavin oxidoreductase

5. ATP synthase and ADP, AMP (25)
ATP SYNTHASE
&ATP SYNTHASE ALPHA CHAIN
&ATP SYNTHASE PROTEIN 9, MITOCHONDRIAL PRECURSOR (LIPID-BINDING PROTEIN)
&ATP synthase subunit 5
adenylate kinase-formation of ADP from AMP

&adenylate kinase 2
&adenylatekinase 2
&ATP-AMP TRANSPHOSPHORYLASE
&ADENYLATE KINASE CYTOSOLIC
&(EC 2.7.4.3)

6. Alternative respiratory path (2)
ALTERNATIVE OXIDASE PRECURSOR
mitochondrial respiratory function protein

7.Reducing carriers (6)
7.1. glutaredoxin
glutaredoxin
7.2. gluathione
GLUTATHIONE PEROXIDASE
7.3. thioredoxin
thioredoxin
thioredoxin peroxidase PMP20

Part II. Regulatory Pathways

I. Genetic information Processing

I.1. DNA replication
1. DNA synthesis (9)
DNA replication initiation protein
&DNA replication initiationprotein
origin recognition complex subunit 1
&ORIGIN RECOGNITION COMPLEX SUBUNIT 1
Single-stranded DNA-binding protein
&SSB
&SINGLE-STRANDEDDNA-BINDING PROTEIN
DNA helicase
topoisomerase ii associated protein
SNA41-it is involved in DNA replication
MCM initiator complex protein-replication licensing factor
&MCM initiator complex protein

2. DNA packaging (16)
2.1. Histone
#Histones, class H1 (or I, or f1)
#Histones, class H2a (or IIb1, or f2a2)
#Histones, class H2b (or IIb2, or f2b)
#Histones, class H4 (or IV, or f2a1)
histone deacetylase
2.2. nonhistone chromosomal protein

nonhistone chromosomal protein NHP6B
chromosome condensation regulator protein
nucleosome assembly protein
&nucleosomeassembly protein
&nucleosome assemblyprotein
High mobility group-like nuclear protein
&HIGH MOBILITY GROUP-LIKE NUCLEAR PROTEIN
structure recognition/chromatin-associated HMG protein

I.2. Transcription
1. RNA Polymerase (4)
#RNA POLYMERASE I, rRNA
#RNA POLYMERASE II, mRNA
#RNA POLYMERASE III, tRNA

2. Regulation (27)
araC-family transcription regulator
transcription factor
transcription regulator
Cap1-transcription factor
&Cap1
fork head-related transcription factor
transcription factor HAP3
transcription initiation factor TFIIF
TRANSCRIPTION INITIATION FACTOR TFIID
HAC1 protein-unfolded protein response pathway, transcrip activation,also see leucine zipper
&HAC1
transcription activator
TRANSCRIPTIONAL REPRESSOR
TIP120-stimulates basal transcription
&TIP120
rho gdp dissociation inhibitor
RNA helicase
YptA-secretory gene product for transcriptional regulation of the secretory pathway
&YptA

3. Processing (14)
3.1. Spliceosome
splicing factor
small nuclear ribonucleoprotein
&snRNP protein
U5 snRNP-specific 200kd protein
RNA12 PROTEIN
RNA splicing protein
U3 smallnucleolar ribonucleoprotein protein IMP4
&U3 SMALLNUCLEOLAR RIBONUCLEOPROTEIN PROTEIN IMP4

3.2. polyA addition
poly(A) polymerase
&POLY(A) POLYMERASE
cleavage and polyadenylation specificity factor

3.3. other
FIBRILLARIN
MINOR CAPSID PROTEIN C
tRNA nucleotidyltransferase

4. tRNA synthetase and ligase (19)
tRNA synthetase
&TRNA SYNTHETASE
prolyl-trna synthetase
trytophanyl-tRNA synthetase
Phenylalanyl-tRNA synthetase
lysyl-tRNA synthetase
leucyl-tRNA synthetase
Glycyl-tRNA synthase
phenylalanine-tRNA ligase
&(EC 6.1.1.20)
tryptophan tRNA ligase
histidine-tRNA ligase precursor
aminoacyl-tRNAsynthetase
THREONYL-TRNA SYNTHETASE

5. Degradation (6)
ribonuclease
&ribonuclease F1
&RIBONUCLEASE TRV
nuclear RNase P and RNase MRP
NUCLEASE S1

6. RNA binding (5)
RNA binding protein
&rna binding protein
&RNA-binding protein

I.3. Translation

1. initiation (19)
RNA recognition motifs
EUKARYOTIC TRANSLATION INITIATION
EUKARYOTIC TRANSLATION INITIATION FACTOR
INITIATION FACTOR

EUKARYOTIC TRANSLATION INITIATION FACTOR 6
&EIF-6
EUKARYOTIC INITIATION FACTOR 4A
&EIF-4A
translation initiation factor 4e
translation release factor subunit 1
INITIATION FACTOR 5A
&EIF-5A
TRANSLATION INITIATION FACTOR EIF3
nuclear cap binding protein
Tif2p-translation initiation factor
&Tif2p

2. elongation (17)
ELONGATION FACTOR
&EF-1-ALPHA
&ELONGATION FACTOR 1-ALPHA
&ELONGATION FACTOR 1-ALPHA 1
&EF-TU
ELONGATION FACTOR 1-GAMMA
&EF-1-GAMMA
elongation factor 2-yeast
&Elongation factor 2
translation elongation factor eEF-3
GTPase activating protein
GTP-binding nuclear protein SPI1
&GTP-BINDING NUCLEAR PROTEIN SPI1
guanine nucleotide releasing factor 1-induces the GTP-binding protein EF-Tu to exchange its bound GDP for GTP
&guanine nucleotide releasing factor 1

3. termination (0)
PEPTIDE CHAIN RELEASE FACTOR

4. Ribosomal proteins (33)
ribosomal protein
40S ribosomal protein
60s ribosomal protein l2
60S ribosomal protein L24
60S ribosomal protein
60S ribosomal protein L3
60S RIBOSOMAL PROTEIN L27A
Ribosomal protein S4B
Ribosomal protein L23A
Ribosomal protein S18A
mitochondrial ribosomal protein

&Mitochondrial ribosomal protein MRPL10
&MITOCHONDRIAL RIBOSOMAL PROTEIN
&mitochondrial S4 ribosomal protein

5. Post-translational modifications (14)
5.1. methylation
SERINE HYDROXYMETHYLTRANSFERASE

5.2. myristoylization
N-myristoyl transferase

5.3. other
peptidylprolyl isomerase
cyclophilin
acid protease
protein disulphide isomerase
protein disulphide isomerase precursor
NosA-nostopeptolide biosynthetic gene cluster of Nostoc sp. peptide synthetase
&NosA
mitochondrial processing peptidase
& MITOCHONDRIAL PROCESSING PEPTIDASE
ModA-responsible for Trimming Asparagine-linked Oligosaccharides on Glycoproteins
&ModA
rehydrin-like protein

6. Folding and Targeting (36)
6.1. folding
CALNEXIN HOMOLOG-folding of glycoproteins
&CALNEXIN HOMOLOG
PEPTIDYL-PROLYL CIS-TRANS ISOMERASE-catalyzes folding
FK506-BINDING PROTEIN-protein folding inhibitor
&FK506-BINDING PROTEIN
6.2. chaperones
chaperone
&chaperonin
&ER chaperone
prefoldin-chaperone which delivers unfolded proteins to another chaperonin
&prefoldin
T-COMPLEX PROTEIN-chaperone of actin, tubulin
&T-COMPLEX PROTEIN
heat_shock_protein
heat shock protein Hsp88
heat shock protein 60
heat shock protein 70
DnaJ protein-heat shock protein
&DnaJ

activator of Hsp70 and Hsp90 chaperones
TPR domain-containing protein-elements in the assembly of the Hsp70-Hsp90 multichaperone machine.
&TPR domain-containing protein
heat-shock protein
heat shock protein DDR48
6.3. protein sorting and targeting
vacuolar protein sorting
vacuolar sorting protein
CARBOXYPEPTIDASE Y-sorting of vacuolar protein
&CARBOXYPEPTIDASE Y
ADP-ribosylation factor GTPase-activatingprotein
ADP-RIBOSYLATION FACTOR
COATOMER ZETA SUBUNIT-trafficing to golgi, nonclathrin vesicles
&COATOMER ZETA SUBUNIT
coatomer complex COPI delta-COP subunit
BET3-targeting and fusion of ER to Golgi transport vesicles
&BET3

7. Turnover-protein degradation-including vacuolar (68)
PROTEASE REGULATORY SUBUNIT
proteinase
proteasome
regulatory particle of the proteasome
&proteasome regulatory subunit

ubiquitin-eukaryotic protein involved in protein degradation
ubiquitin fusion degradation protein
N-END-RECOGNIZING PROTEIN
ubiquitin-activating enzyme e1
ubiquitin carboxyl-terminal hydrolase
Ubiquitin-specifc processing protease
ubiquitin ligase
ubiquitin conjugating enzyme
ubiquitin fusion protein
polyubiquitin
pepsinogen
PALB-cysteine protease
&PALB
aspartic protease
&aspartyl protease
&ASPARTIC PROTEINASE
&ASPARTYL PROTEINASE
&aspartic proteinase precursor
aminopeptidase

iminopeptidase
MepB-encodes an 82 kDa intracellular metalloproteinase structurally related to mammalian thimet oligopeptidases
&MepB
methionine metallopeptidase
proteinase T precursor
ca dependent protease
Subtilisin-like protease PR1H
L-KYNURENINE HYDROLASE
microsomal dipeptidase precursor
aspartylproteinase
carboxypeptidase s precursor
carboxypeptidase D
&(EC3.4.16.6)
serine-type carboxypeptidase
&EC3.4.16.1
3-hydroxybutyryl-CoA dehydrogenase-degradation of the branched-chain amino acids
&3-hydroxybutyryl-CoA dehydrogenase
&3-HYDROXYBUTYRYL-COA DEHYDROGENASE
3-hydroxyacyl-CoA dehydrogenase
PEPTIDASE
dipeptidase

8. protein binding (19)
ACYL-COA-BINDING PROTEIN HOMOLOG
oxysterol-binding protein
amiloride-binding protein
methyl-CpG binding protein MBD4
saccharide-binding protein
a-agglutinin attachment subunit precursor
lectin precursor
DNA-binding proteins
CURVED DNA-BINDING PROTEIN
TGACG-motif binding protein
DNA-binding protein HEXBP-specific DNA-binding protein
Zinc finger motif-DNA binding
zinc-finger transcription factor
zinc finger protein
&Zn-finger protein
&zinc-finger protein
&C2H2-type zinc finger protein
&Zinc finger
zinc finger suppressor

II. Cell Growth, Cell Division, Mating and Morphogenesis

II.1. Cell walls, Biomembranes and Cytoskeleton

1. Cell walls (17)

cell wall protein
cell wall biogenesis protein
QI74 protein-cell wall protein of Trichoderma harzianum
&QI74
Psu1 protein-cell wall synthesis protein psu1
&Psu1 protein
GEL1 protein-Biosynthesis of the Fungal Cell Wall
&GEL1 protein
septin
&G-septin gamma
adhesion regulating molecule ARM-1
mycelial surface antigen CSA1 precursor-cell wall protein
&mycelial surface antigen CSA1 precursor
Glutamine--fructose-6-phosphate amidotransferase
LysB-murein hydrolase
&LysB

2. Biomembranes (34)

membrane protein
&membraneprotein
&MEMBRANECOMPONENT
integral membrane protein
transmembrane protein
vacuolar membrane protein
&vacuolar membraneprotein
peripheral vacuolar membraneprotein
annexin XIV-calcium-dependent phospholipid binding proteins
&annexin XIV
opsin-1-a seven transmembrane helix retinal-binding protein
&opsin-1
peroxisomal membrane protein
peroxisomal membrane protein per10
&PEROXISOMAL MEMBRANE PROTEIN PER10

3. Cytoskeleton (55)

peroxisomal hydratase-dehydrognease-epimerase
Golgi complex-associated protein
Proteasome subunit
endoplasmic reticulum associated protei
&HDEL receptor
KINESIN
KINESIN-LIKE PROTEIN KLPA
tubulin

dynactin
&DYNACTIN-microtubule-binding

3.1. actin-see also mitosis
actin
gamma-actin
PROFILIN-assembly of actin monomers
&PROFILIN
ARP2/3 COMPLEX-actin polymerization
fimbrin-actin bundling
&fimbrin
COFILIN-actin binding
&cofilin
ACTIN-LIKE PROTEIN
exocyst complex component sec3

3.2. myosin
myosin I myoA-the heavy chain of muscle protein
&myosin I myoA
myosin II heavy chain
&MYOSIN II HEAVY CHAIN
tropomyosin I

3.3. choline
choline dehydrogenase
choline oxidase
cholinephosphate cytidylyltransferase
PHOSPHATIDYLINOSITOL/PHOSPHATIDYL-CHOLINE TRANSFER PROTEIN
phosphatidylcholine-sterol acetyltransferase

3.4. other
Glycosyltransferase-glycopeptidolipid biosyn
&Glycosyltransferase

4. organelle (5)
mitochondrial carrier protein
mitochondrial protein
mitochondrial morphologyprotein MMM1

II.2. cell cycle control (9)

cell division control protein
cell division control protein cdc15
cell division cycle CDC50
cullin1-neg regulator of cell cycle
clock-controlled gene-6 protein

B-type cyclin
cell cycle regulator p21 protein

II.3. Mitosis/cytokinesis

1. Mitosis (11)
CHROMOSOME SEGREGATION PROTEIN-with microtubule, migration of chromosomes
&CHROMOSOME SEGREGATION PROTEIN
dynamin-molecular motor, associated with mocrotuble, endocytosis
&dynamin
dynein-molecular motor
&dynein
&DYNEIN HEAVY CHAIN
spindle assembly checkpoint protein SLDA
SSD1 PROTEIN
syntaxin-essential for membrane fusion events critical for cell division
&syntaxin

2. cytokinesis (3)
F-ACTIN CAPPING PROTEIN ALPHA-2 SUBUNIT
F-ACTIN CAPPING PROTEIN BETA SUBUNIT ISOFORMS 1 AND 2

II.4. Meiosis (2)
pelota protein-a protein required for meiotic cell division
&pelota protein/yeast dom34 homolog
&pelota

II.5. Cell death (1)
CELLULAR APOTOSIS SUSCEPTIBILITY PROTEIN

III. Processes

III.1. Cell rescue, defense, osmotic adaptation, starvation response, development (includes antibiotics, toxins)

1. development (24)
growth regulation protein
growth regulation protein
&growth regulation
&involved in growth regulation
-sexual
mating and morphogenesis protein
&SCD1 PROTEIN
tol protein-a mediator of mating-type-associated vegetative incompatibility in fungus
&tol protein
SEXUAL DIFFERENTIATION PROCESS PROTEIN-expressed during sexual diff in S. pombe

pigment production
GREEN PIGMENT SYNTHASE
coproporphyrinogen III oxidase precursor
porphyrinogen oxidase
&diphenol oxidase
brown 2-A developmentally regulated gene cluster involved in conidial pigment biosynthesis
&brown 2
PvcA-Pseudomonas aeruginosa pyoverdine chromophore biosynthesis gene cluster
&PvcA
osmotic growth protein
osmotic growth protein
imbibition protein
organism development
epd1 protein precursor-essential for pseudohyphaldevelopment 1
CAP20-One of the genes expressed uniquely in C. gloeosporioides during appressorium formation
&CAP20
ascospore maturation 1 protein
genitaliumglycerol-3-phospate dehydrogenase
BIGH3-transforming growth factor-beta induced gene
&BIGH3
Bdf1p-required for sporulation
&Bdf1p
TRANSFORMING GROWTH FACTOR BETA 2 PRECURSOR(TGF-BETA 2)
involved in pseudohyphal growth

2.defense and secondary metabolites (57)
2.1. trichothecine biosynthesis pathway in F. sporotrichioides
15-decalonectrin 15-O-acetyltransferase
&TRI3
TRICHODIENE OXYGENASE
&CYTOCHROME P450 58
&EC 1.14.-.-
&TRI4
trichodiene synthase
&TRICHODIENE SYNTHASE
&SESQUITERPENE CYCLASE
&TRI5
Tri6
T-2 toxin biosynthesis protein
&TRI7
&TRI8
trichothecene 3-O-acetyltransferase
&Tri101
isotrichodermin C-15 hydroxylase
&Tri11 product
&ISOTRICHODERMIN C-15 HYDROXYLASE

&CYTOCHROME P450 65A1
trichothecene efflux pump
&Tri12 product
Tri13
Tri14
Tri16

2.2. other secondary metabolites

cytochrome P450 sterol 14-demethylase
CYTOCHROME P450 3A28
aflatoxin biosynthesis protein
sterigmatocystin 7-o-methyltransferase precursor
sterigmatocystin biosynthesis protein
sterigmatocystin biosynthesis p450 monoosygenase stcb
daunorubicin C-13 ketoreductase
gibberellin 7-oxidase
&EC 1.14.11.-
gibberellin 20-oxidase
&EC 1.14.11.-
gibberellin biosynthesis-related
thiazole biosynthetic enzyme procursor
cephalosporin C acetylhydrolase
leukotriene A-4 hydrolase
NosD-nostopeptolide biosynthetic gene cluste
&NosD
1-hydroxy-2-naphthoate dioxygenase-A phenanthrene degradative gene cluster
&1-hydroxy-2-naphthoate dioxygenase
deacetylcephalosporin C acetyltransferase
PHYTOENE SYNTHASE
phytoene dehydrogenase
6,7-dimethyl-8-ribityllumazine synthase
PhzG-pyocyanine biosynthesis operon
&PhzG
snodprot1-a protein produced during infection of wheat leaves by Phaeosphaeria nodorum
&snodprot1
versicolorin B synthase
HYGROMYCIN-B KINASE
&HYGROMYCIN B PHOSPHOTRANSFERASE

2.3. defense

VEGETABLE INCOMPATIBILITY PROTEIN HET-E-1-vegetative incompatibility,defense
pisatin demethylase-inactivates plant compound pisatin
&pisatin demethylase
D-AMINO ACID OXIDASE-oxidation of cephalosporin C
&D-AMINO ACID OXIDASE
oligomycin sensitivity conferring protein
aureobasidin-resistance protein

Syringomycin response protein 2
66 KD STRESS PRITEIN

3. detoxification (4)
catalase
manganese superoxide dismutase precursor
epoxide hydrolase
spindle poison sensitivity protein

4.salt tolerance (3)
heavy metal tolerance protein precursor
halotolerance protein
metal resistance protein

5. starvation response (2)
RVS167 protein-reduces viability upon starvation
&RVS167 protein

6.DNA repair (11)
UV-endonuclease-UV damage DNA repair
&UV-endonuclease
CDC20 protein-is required for recovery from DNA damage
&CDC20 protein
extracellular putative DNase-induces disease resistance responses in peas
&extracellular putative DNase
Hmp1-mismatch base pair and cruciform DNA recognition protein
&Hmp1
exonuclease
DNA REPAIR PROTEIN
&UV EXCISION REPAIR PROTEIN
uv excision repair protein rad23
dna repair helicase

7. allergen and immune system proteins (9)
7.1. allergen
ALLERGEN ALT A 7
rAsp f 7
rAsp f 9-Recombinant Aspergillus fumigatus allergens
&rAsp f 9
7.2. immune system proteins
immunoglobulin heavy chain binding protein homolog
NAALADase II protein-a marker of prostatic carcinomas
&NAALADase II protein

8. tumor protein and tumor suppressor (8)
8.1. tumor protein

cortactin-oncogene EMS1 product
&cortactin
tumor protein homolog
&TUMOR PROTEIN HOMOLOG
&RAB
25 kda ras-related protein-proto-oncogene product
&25 kda ras-related protein
hepatopoietin

8.2. tumor suppressor

tumor suppressor protein
&tumorsuppressor
saframycin Mx1 synthetase A-saframycin Mx1 is a DNA-binding antibiotic and antitumour agent
&saframycin Mx1 synthetase A
gene N33 protein-N33, candidate tumor suppressor gene
&39 kDa encoded by N33

9. multidrug resistance (7)

multiple drug resistance protein MDR
multidrug resistance related protein 2
TETRACYCLINE RESISTANCE PROTEIN
&tetracycline efflux protein
mfs-multidrug-resistance transporter
facilitator family multi-drug resistance protein
protein involved in drug resistance
&drug resistance

10. cell reaction to environment (1)

wc-1-the central regulator of blue light responses
&wc-1

III.2. Cell signalling, signal transduction and secondary messenger

1. PHOSPHATASES (14)

PROTEIN PHOSPHATASE
serine/threonine phosphatase
&serine/threonine protein phosphatase PPT1
&serine/threonine protein phosphatase
&Ser/Thr protein phosphatase
&serine/threonine protein phosphatase type 2A regulatory chain A
&SERINE/THREONINE PROTEIN PHOSPHATASE 2B CATALYTIC SUBUNIT
protein phosphotase 2a 65kd regulatory
dual-specificity map kinase phosphatase
&orf 5' to PPH3

2. Kinases (25)

protein kinase

protein kinase C
MAP kinase
MAP KINASE HOG1
CAMP-DEPENDENT PROTEIN KINASE
serine/threonine-protein kinase
&SERINE/THREONINE-PROTEIN KINASE, casein kinase acts on acidic proteins
&serine/threonine protein kinase
&SERINE/THREONINE PROTEIN KINASE
&casein kinase
calmodulin-dependent protein kinase
&CALCIUM/CALMODULIN-DEPENDENT PROTEIN KINASE
&CALMODULIN-DEPENDENT PROTEIN KINASE
&calmodulin
PduW-propionate kinase
&PduW
GM3 synthase
protein tyrosine kinase 9 related protein
Guanylate Kinase
muscle-specific serine kinase 1
casein kinase-1

3. cAMP-secondary messenger (3)
14-3-3-interacts with CAP (adenylyl cyclase-associated protein)
&14-3-3

4. G protein (13)
GTP-binding protein
GUANINE NUCLEOTIDE-BINDING PROTEIN
GTPASE-ACTIVATING PROTEIN-neg regulator of Ras1, play antagonistic role with rho-gdp dissociation inhibitor
&GTPASE-ACTIVATING PROTEIN
Ras protein

5. Inositol triphosphate-secondary messenger (2)
inositol regulator
Inositol polyphosphate phosphatase

6. other (3)
palA product-pH signal transduction pathway gene product
&palA product
&palA
carbon catabolite repression regulator
carbon catabolite repressor protein(CCR4)

III.3. Transmembrane transport

1. secretion (3)
ALPHA-ADAPTIN
large secreted protein
secreted glucosidase

2. exoenzymes (3)
alanyl dipeptidyl peptidase
GDP dissociation inhibitor

3. membrane transport (6)
nuclear transport factor 2
pre-tRNA nuclear export receptor
membrane transport protein
&membrane transporter
rab6-play an essential role in targeting and fusion of transport vesicles with their appropriate acceptor membranes
&rab6

4. mitochondrial transport (12)
ADP,ATP CARRIER PROTEIN
OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN
MITOCHONDRIAL PRECURSOR PROTEINS IMPORT RECEPTOR
&72 KDMITOCHONDRIAL OUTER MEMBRANE PROTEIN
&MITOCHONDRIAL IMPORTRECEPTOR FOR THE ADP/ATP CARRIER
&TRANSLOCASE OF OUTER MEMBRANETOM70
MITOCHONDRIAL NUCLEASE
mitochondrial import receptor-a mitochondrial phosphate transport protein
&MITOCHONDRIAL IMPORT RECEPTOR
mitochondrial intermediate peptidase percursor
Mitochondrial matrix protein involved inprotein import

5. transporting for sugar, cation, anion, protein, fatty acid etc.
5.1. sugar transport (8)
sugar transporter
&sugar transport
hexose transporter
&hexose transport
GLUCOSE TRANSPORTER
UDP-galactose transporter related protein 1
maltose permease
Golgi GDP-mannose transporter

5.2. cation transport (46)
5.2.1. ATPase family-transmemberane protein that transport cations
P Type Copper ATPase
ATPase proteolipid

member of the AAA ATPase family of proteins
membrane-spanning ATPase
DNA-dependent ATPases
calcium-transporting ATPase
&CALCIUM-TRANSPORTING ATPASE
&Ca2+-transporting ATPase
&calcium P-type ATPase
&Ca++ ATPase
transitional endoplasmic reticulum ATPase
&TRANSITIONAL ENDOPLASMIC RETICULUM ATPASE
sodium/potassium-transporting ATPase
&SODIUM/POTASSIUM-TRANSPORTING ATPASE
&NA+/K+ ATPASE
PLASMA MEMBRANE ATPASE (PROTON PUMP)
&(EC 3.6.1.35)
&H+-transportingATPase
&H+-transporting ATPase
&plasma membrane H(+)ATPase
V-ATPase
cation-transporting P-ATPase
&P-ATPase
vacuolar ATPase subunit H

5.2.2. facilitator protein
major facilitator superfamily protein
major facilitator protein
Mfs1.1-member of the major facilitator superfamily
&Mfs1.1
5.2.3. others
potassium channel protein
&K+ channel protein
&potassium channel subunit
&potassium channel beta subunit
&potassium channel beta 2
&potassium channel
Arabidopsis ATHKCP-potassium channel protein
&Arabidopsis ATHKCP
K+/H+-antiporter
H/K ATPase
copper transporter protein
&coppertransporter protein
&COPPER TRANSPORTER PROTEIN
&COPPER TRANSPORT
&copper transporter
calcium binding protein
&CALCIUM-BINDING PROTEIN

&calcium binding 140k protein
calcium/proton exchanger
&H+/Ca2+ exchanger
purine-cytosine permease
ammonia transport protein
&ammonium transporter MEPa
&AMMONIUM TRANSPORTER MEP1
&AMMONIUM TRANSPORTER MEP2
&AMMONIUM TRANSPORTER MEP3
Low-affinity zinc transport protein

5.3. Anion transport (18)
anion transporter
phosphate permease
&phosphate transport
&phosphate transporter
allantoate permease
&allantoate transport
&ALLANTOATE PERMEASE
quinate permease
&QUINATE PERMEASE
malate permease
pantothenate permease
&panF-1
acetate permease
tricarboxylate transport protein
mitochondrial dicarboxylate transportprotein
SutA-sulfate transporter
&SutA

5.4. Protein, amino acid transport (21)
PROTEIN TRANSPORT PROTEIN
&EXPORT PROTEIN NCE2
peptide transporter
&PEPTIDE TRANSPORTER
&PEPTIDE PERMEASE
peptide transporter MTD1
AMINO-ACID PERMEASE
&amino acid permease
ARGININE PERMEASE
importin beta-2 subunit
importin beta subunit
&IMPORTIN BETA-2 SUBUNIT-transportin
N AMINO ACID TRANSPORT SYSTEM PROTEIN
&METHYLTRYPTOPHANRESISTANCE PROTEIN
GABA permease-gamma-amino-n-butyrate permease

&GABA permease
gamma-glutamyltranspeptidase precursor
Opt1p-oligopeptide transport gene
&Opt1p
methionine permease
soluble NSF attachment protein-a soluble transport factor
&soluble NSF attachment protein
SLS1 PROTEIN PRECURSOR-an endoplasmic reticulum component involved in protein translocation process
&SLS1 PROTEIN PRECURSOR
synaptobrevin-protein trafficing
&synaptobrevin

5.5. fatty acid trasport (2)
fatty acid transporter protein
&fatty acid transport pro
phosphatidylglycerol/phosphatidylinositol transfer protein
&phosphatidylglycerol/phosphatidylinositol transferprotein

5.6. ABC transporter family (11)
ABC transporter protein-Osmoregulated and lipid transport
&ABC transporter protein
&ABC1 transporter
&ABC-type ATPase
&ATP-DEPENDENT TRANSPORTER
&ATP binding cassette (ABC) transporter
PDR-like ABC transporter

5.7. other (16)
transport protein
&transporter protein
&transporter
&translocation protein
metabolite transport protein
OLIGOMYCINRESISTANCE ATP-DEPENDENT PERMEASE
nitrogen permease

Part III Unclassified protein

I. Classes of Enzymes (from M. Reily and KEGG)

1. Oxidoreductases (39)
reductase
dimethylaniline monooxygnease
&DIMETHYLANILINE MONOOXYGENASE
OXIDOREDUCTASE

MONOOXYGENASE
GMC oxidoreductase
dihydroxy-acid dehydratase precursor
PHENOL 2-MONOOXYGENASE
&PHENOL HYDROXYLASE
amine oxidase-OXIDATIVE DEAMINATION of AMINES
fructosyl amino acid oxidase
fructosyl amine:oxygen oxidoreductase
cytochrome P450 monooxygenase
NADH-dependent butanol dehydrogenase
ascorbate oxidase
L-gulonolactone oxidase
desaturase
2-oxoglutarate dehydrogenase e1 component
short chain dehydrogenase

2. Transferases (7)
methyl chloride transferase
NrgA-N-acetyltransferase homolog
&NrgA
o-methyltransferase
methyltransferase

3.Hydrolases (8)
alkaline phosphatase
secreted acid phosphatase 2
acid phosphatase
dihydroxyacid dehydratase
esterase D
esterase

4. Lyases
LACTOYLGLUTATHIONE LYASE
&METHYLGLYOXALASE
&ALDOKETOMUTASE
&GLYOXALASE I

5. Isomerases

6. Ligases

II. Non-enzymatic classes (not in defined pathways)
glycoprotein 900
proline-rich protein
trp-asp repeat protein
selenoprotein P precursor

&SELENOPROTEIN P PRECURSOR
F-box protein Fbl7
GNS1/SUR4 family protein
WD repeat protein
C2-domain synaptic vesicle protein
nmt1 protein-the expression of this gene has been shown to be completely inhibited by thiamine
&nmt1 protein
regulatory protein
putative gamma-adaptin
ring-box protein 1
nebulette
SWH1 protein-from yeast encodes a candidate nuclear factor containing
ankyrin repeats and showing homology to mammalian oxysterol-binding protein
&SWH1
GlxB6-Escherichia coli glyoxylate induced protein
&GlxB6
CROSS-PATHWAY CONTROL PROTEIN 1

Part IV. Unidentified (includes significant match with ORFs) (256)

unknown function
&unknown
&unclear function
&hypothetical protein
&hypothetical
&HYPOTHETICAL
&prediction
&unnamed protein product
&ORF
&Unknown protein
HET-C protein
acr-2 protein
MYO2
SCD6 protein
CG6198 gene product
CG7459 gene product
CG13902 gene product
CG9953 gene product
CG10627 gene product
CG14542 gene product
CG8665 gene product
CG9318 gene product
CG5340 gene product
CG10882 gene product
CG7759 gene product
Spx gene product
Slh gene product

CG8031 gene product
CG3172 gene product
PDI related protein A
BG:DS02740.2 gene product
CG7828 gene product
CG8430 gene product
pac2 protein
CG8430 gene product
beta'Cop gene product
QM protein
CG12110 gene product
CG15015 gene product
CG10306 gene product
CG7218 gene product
&HOST SPECIFICITY PROTEIN J
Xenopus 14s cohesin smc1 subunit homolog
KIAA0829 protein
KIAA0585 protein
KIAA0713 protein
KIAA0416
CG10843 gene product
CG17661 gene product
CG10843 gene product
ORF YOL057w
ORF YGR159c
ORF YGR054w
01232
HSPC263
similar to S. cerevisiae PKR1
INTEGUMENTARY MUCIN A.1 PRECURSOR

protein disulphide isomerase

Part V. No significant homolog

#NONE
-Contigs 424
-Singlets 1215

Appendix II

Fusarium sporotrichioides ESTs biological function assingments

The following is an example of the BlastX result.

Contig1056 2015 1.6e-207 67 1188 gb|AAD13657.1| (M27246) trichodiene synthase [Fusarium sporotrichioides]

In this example, "Contig1056" is the ID of the contig used to do the BlastX homology search. "2015" is S score (section 2.4.3.1) and 1.6e-207 is the E value (section 2.4.3.1). High S score or low E value indicates less probability of error. "67 1188" is the sequence homology endpoints. The homologous sequence identifier is "gb|AAD13657.1| (M27246)", where the "gb" indicates that the identifier refers to a GenBank sequence, "AAD13657.1" is its GenBank Accession, and "M27246" is the GenBank Locus. At the end, "trichodiene synthase" is the description of homology and "[Fusarium sporotrichioides]" is the organism.

Please refer to section 2.4.3.3 for "Interpreting Sequence Identifiers".

PART I. Metabolic Pathways

I. Metabolism of Carbohydrates(for glucose see energy)

1. Chitin metabolism (6)

<chitinase>

j3h07fs.r1	223	3.4e-16	187	549	dbj|BAA88380.1| (AB024740) chitinase [Pyrococcus kodakaraensis]
b3b03fs.r1	165	9e-11	117	422	gb|AAF02299.1|AF0983 (AF098302) chitinase [Brassica juncea]
g3h07fs.r1	133	4.6e-08	341	439	gb|AAB48566.1| (U60806) complement-fixation chitinase [Coccidioides immitis]

<chitinase precursor>

j3h07fs.f1	143	5.5e-08	143	409	sp|P20533|CHI1_BACCI CHITINASE A1 PRECURSOR >pir||A38368 chitinase (EC3.2.1.14) precursor - Bacillus circulans >gb|AAA81528.1| (M57601)chitinase A1 [Bacillus circulans]
g3h07fs.f1	344	1.9e-30	155	514	pir||JC4565 chitinase (EC 3.2.1.14) 1 precursor - Coccidioides immitis

<chitin synthase>

g1c05fs.r1	562	1.5e-52	11	466	sp|O13395|CHS6_USTMA CHITIN SYNTHASE 6 (CHITIN-UDP ACETYL-GLUCOSAMINYLTRANSFERASE 6) (CLASS-V CHITIN SYNTHASE 6) >pir||T42022 probablechitin synthase (EC 2.4.1.16), class IV - smut fungus (Ustilagomaydis) >gb|AAB84285.1| (AF030554) class V chitin synthase[Ustilago maydis]

2. Cellulose metabolism (8)

<beta glucosidase-breakdown of cellulose>

c4f02fs.r1	638	1.3e-61	18	455	dbj|BAA74958.1| (AB003109) beta-glucosidase [Humicola grisea var. thermoidea]
r3h10fs.r1	278	1.9e-23	174	605	dbj|BAA19145.1| (AB000539) glucan 1,3-beta-glucosidase precursor[Schizosaccharomyces pombe]

<glucanase>

Contig756	570	2.2e-54	243	1049	pir||T39920 probable glucanase precursor - fission yeast (Schizosaccharomycespombe) >emb|CAB57923.1| (AL121794) putative glucanase precursor[Schizosaccharomyces pombe]
Contig897	256	7.9e-40	487	702	sp|P45699|GUNK_FUSOX PUTATIVE ENDOGLUCANASE TYPE K PRECURSOR(ENDO-1,4-BETA-GLUCANASE) (CELLULASE) >gb|AAA65589.1| (L29381)K-family cellulase homologue [Fusarium oxysporum]
i2e04fs.f1	301	7.1e-26	180	446	gb|AAB51451.1| (U14948) endo-beta-1,4-glucanase [*Macrophomina phaseolina*]

<endo-beta-1,6-glucanase>

k4d12fs.r1	530	3.5e-50	108	530	pir||S55325 endo-beta-1,6-glucanase - fungus (Trichoderma harzianum)>emb|CAA55789.1| (X79197) glucan endo-1,6-beta-glucosidase[Trichoderma harzianum]
<cellulase>					
Contig967	830	5.6e-82	110	589	sp|P45699|GUNK_FUSOX PUTATIVE ENDOGLUCANASE TYPE K PRECURSOR(ENDO-1,4-BETA-GLUCANASE) (CELLULASE) >gb|AAA65589.1| (L29381)K-family cellulase homologue [Fusarium oxysporum]
<bcsABII-B-cellulose synthase genes>					
c4h03fs.r1	187	3e-12	19	426	dbj|BAA77600.1| (AB015804) bcsABII-B [Acetobacter xylinus]
3. Cutin metabolism (4)					
<cutinase G-box binding protein>					
k3h03fs.r1	649	9.1e-63	51	527	gb|AAB04132.1| (U61841) cutinase G-box binding protein [Haematonectriahaematococca]
d1a06fs.r1	494	2.4e-46	3	317	gb|AAB04132.1| (U61841) cutinase G-box binding protein [Haematonectriahaematococca]
g1a09fs.r1	460	8.9e-43	10	465	gb|AAB04132.1| (U61841) cutinase G-box binding protein [Haematonectriahaematococca]
d1a06fs.r1	494	2.4e-46	3	317	gb|AAB04132.1| (U61841) cutinase G-box binding protein [Haematonectriahaematococca]
4. Polysaccharide metabolism (3)					
<UDP-glucose dehydrogenase>					
Contig262	186	6.1e-13	119	307	pir||A54926 UDPglucose 6-dehydrogenase (EC 1.1.1.22) - bovine >gb|AAB32227.1|UDP-glucose dehydrogenase, UDPGDH=52 kda subunit {EC1.1.1.22} [cattle, liver, Peptide, 468 aa]
a2e07fs.r1	100	0.00043	332	451	gb|AAF26173.1|AC0082 (AC008261) putative UDP-glucose 6-dehydrogenase[Arabidopsis thaliana]
<utp--glucose-1-phosphate uridylyltransferase>					
Contig554	668	7.5e-65	201	779	pir||T40935 probable utp--glucose-1-phosphate uridylyltransferase fissionyeast (Schizosaccharomyces pombe) >emb|CAA22857.1| (AL035259)probable utp--glucose-1-phosphate uridylyltransferase[Schizosaccharomyces pombe]
5. Galactose metabolism (10)					
<galactose oxidase>					
j1a08fs.r1	499	6.2e-47	10	426	pdb|1GOF| Galactose Oxidase (E.C.1.1.3.9) (Ph 4.5) >pdb|1GOG|GalactoseOxidase (E.C.1.1.3.9) (Ph 7.0) >pdb|1GOH| Galactose Oxidase(E.C.1.1.3.9) (Apo Form)
<GALACTOSE OXIDASE PRECURSOR (GAO)>					

Contig970	1306	2.1e-132	220	1365	sp|Q01745|GAOA_DACDE GALACTOSE OXIDASE PRECURSOR (GAO)>gb|AAA16228.1|(M86819) galactose oxidase [Hypomyces rosellus]
Contig688	352	1.9e-30	92	517	pir||A38084 galactose oxidase (EC 1.1.3.9) precursor - fungus (Cladobotryumdendroides)
<GALACTOSIDE O-ACETYLTRANSFERASE>					
s3b12fs.r1	161	4.7e-11	236	562	sp|P07464|THGA_ECOLI GALACTOSIDE O-ACETYLTRANSFERASE (THIOGALACTOSIDEACETYLTRANSFERASE) >pir||XXECTG galactoside O-acetyltransferase (EC2.3.1.18) - Escherichia coli >gb|AAA24055.1| (J01636)thiogalactoside acetyltransferase (ttg start codon) [Escherichiacoli] >emb|CAA36162.1| (X51872) thiogalactoside transacetylase[Escherichia coli] >gb|AAC73445.1| (AE000141)thiogalactosideacetyltransferase [Escherichia coli]
<alpha-1,4 polygalactosaminidase>					
Contig893	498	8.7e-47	201	872	dbj|BAA03574.1| (D14846) endo alpha-1,4 polygalactosaminidase precusor[Pseudomonas sp.]
i4c12fs.r1	227	4.5e-18	91	465	dbj|BAA03574.1| (D14846) endo alpha-1,4 polygalactosaminidase precusor[Pseudomonas sp.]
<GALACTOSYLTRANSFERASE>					
a1f04fs.r1	242	7.5e-33	213	431	gi|6320451 ref|NP_010531.1|MNN10| galactosyltransferase; Mnn10p>sp|P50108|MN10_YEAST GALACTOSYLTRANSFERASE MNN10 (BUD EMERGENCEDELAY PROTEIN 1) >pir||S54541 BED1 protein - yeast (Saccharomycescerevisiae)>emb|CAA89731.1| (Z49701) unknown [Saccharomycescerevisiae] >gb|AAB48372.1| (L42540) Mnn10p [Saccharomycescerevisiae]
<beta-galactosidase>					
o3d03fs.r1	315	4.4e-26	66	518	sp|P00723|BGAL_KLULA BETA-GALACTOSIDASE (LACTASE)>pir||JC1266beta-galactosidase (EC 3.2.1.23) - yeast (Kluyveromyces marxianusvar. lactis) >gb|AAA35265.1| (M84410) beta-D-galactosidase [Kluyveromyces lactis]
<UDP glucose NAD dependant epimerase/dehydratase>					
s1c02fs.r1	385	8.3e-35	92	538	pir||T40321 UDP glucose NAD dependant epimerase/dehydratase - fission yeast(Schizosaccharomyces pombe) >emb|CAB44766.1| (AL078627) UDP glucoseNAD dependant epimerase/dehydratase [Schizosaccharomyces pombe]
s1c02fs.f1	237	3.7e-19	250	489	pir||T40321 UDP glucose NAD dependant epimerase/dehydratase - fission yeast(Schizosaccharomyces pombe) >emb|CAB44766.1| (AL078627) UDP glucoseNAD dependant epimerase/dehydratase [Schizosaccharomyces pombe]

6. Mannitol and mannose metabolism (14)

<mannose-1-phosphate guanylyltransferase>

n4e12fs.r1	673	2.2e-65	22	462	gb|AAC39498.1| (U89991) mannose-1-phosphate guanylyltransferase [Hypocreajecorina]
o2c12fs.r1	429	1.5e-39	117	497	pir||T39403 probable mannose-1-phosphate gaunyl transferase - fission yeast(Schizosaccharomyces pombe) >emb|CAA18655.1| (AL022600) putativemannose-1-phosphate guanyltransferase. [Schizosaccharomyces pombe]

<dolichyl phosphate-D-mannose:proteinO-D-mannosyltransferase>

g2f12fs.r1	478	3.6e-44	16	432	gi|6322603 ref|NP_012677.1|PMT4| dolichylphosphate-D-mannose:proteinO-D-mannosyltransferase; Pmt4p>sp|P46971|PMT4_YEASTDOLICHYL-PHOSPHATE-MANNOSE--PROTEIN MANNOSYLTRANSFERASE4>pir||S60415 dolichyl-phosphate-mannose--proteinmannosyltransferase (EC2.4.1.109) PMT4 - yeast (Saccharomycescerevisiae) >emb|CAA58729.1| (X83798)PMT4 [Saccharomycescerevisiae] >emb|CAA89676.1| (Z49643) ORF YJR143c
d1c06fs.r1	386	7.5e-34	35	484	sp|O74189|PMT1_CANAL DOLICHYL-PHOSPHATE-MANNOSE—PROTEIN MANNOSYLTRANSFERASE 1>gb|AAC31119.1| (AF000232) protein mannosyltransferase 1 [Candidaalbicans]
b2b09fs.r1	313	3.3e-26	23	397	gi|6322603 ref|NP_012677.1|PMT4| dolichylphosphate-D-mannose:proteinO-D-mannosyltransferase; Pmt4p>sp|P46971|PMT4_YEASTDOLICHYL-PHOSPHATE-MANNOSE--PROTEIN MANNOSYLTRANSFERASE 4>pir||S60415 dolichyl-phosphate-mannose--proteinmannosyltransferase (EC2.4.1.109) PMT4 - yeast (Saccharomycescerevisiae) >emb|CAA58729.1| (X83798)PMT4 [Saccharomycescerevisiae] >emb|CAA89676.1| (Z49643) ORF YJR143c
i4d08fs.r1	290	1.1e-23	29	445	gi|6322603 ref|NP_012677.1|PMT4| dolichyl phosphate-D-mannose:proteinO-D-mannosyltransferase; Pmt4p>sp|P46971|PMT4_YEASTDOLICHYL-PHOSPHATE-MANNOSE--PROTEIN MANNOSYLTRANSFERASE4>pir||S60415 dolichyl-phosphate-mannose--proteinmannosyltransferase (EC 2.4.1.109) PMT4 - yeast (Saccharomycescerevisiae) >emb|CAA58729.1| (X83798)PMT4 [Saccharomycescerevisiae] >emb|CAA89676.1| (Z49643) ORF YJR143c
o4c04fs.f1	168	1.4e-10	307	534	gi|6319296 ref|NP_009379.1|PMT2| dolichyl phosphate-D-mannose:proteinO-D-mannosyltransferase; Pmt2p>sp|P31382|PMT2_YEASTDOLICHYL-PHOSPHATE-MANNOSE--PROTEIN MANNOSYLTRANSFERASE2>pir||S36711 dolichyl-phosphate-mannose--proteinmannosyltransferase (EC 2.4.1.109) PMT2 - yeast

					(Saccharomycescerevisiae) >gb|AAC04934.1| (L05146)Pmt2p: proteinO-D-mannosyltransferase [Saccharomyces cerevisiae]
<mannose-6-phosphate isomerase>					
p1h06fs.f1	375	9.5e-34	179	526	sp|P29951|MANA_EMENI MANNOSE-6-PHOSPHATE ISOMERASE (PHOSPHOMANNOSE ISOMERASE)(PMI) (PHOSPHOHEXOMUTASE) >pir||A56239 mannose-6-phosphateisomerase (EC 5.3.1.8) - Emericella nidulans >gb|AAA33319.1|(M85239) phosphomannose isomerase [Emericella nidulans]
<mannosyl-oligosaccharide1,2-alpha-mannosidase>					
l3h11fs.f1	249	1.6e-19	225	524	gb|AAF16414.1|AF1265 (AF126550) mannosyl-oligosaccharide1,2-alpha-mannosidase [Glycine max]
g1c07fs.r1	226	4.7e-17	39	458	gb|AAF16414.1|AF1265 (AF126550) mannosyl-oligosaccharide1,2-alpha-mannosidase [Glycine max]
e1h12fs.f1	134	4.2e-07	238	450	gb|AAF16414.1|AF1265 (AF126550) mannosyl-oligosaccharide1,2-alpha-mannosidase [Glycine max]
<Alg2-encoding a mannosyltransferas>					
g3g03fs.r1	336	1.2e-29	84	428	dbj|BAA34296.1| (AB015054) Alg2 [Rhizomucor pusillus]>dbj|BAA34297.1|(AB015055) Alg2 [Rhizomucor pusillus]
<endoplasmic reticulum alpha-mannosidase>					
l3h11fs.r1	237	4.6e-18	124	420	gi|6005808 ref|NP_009161.1|| endoplasmic reticulum alpha-mannosidase 1>gb|AAD45504.1|AF145732_1 (AF145732) endoplasmic reticulumalpha-mannosidase I [Homo sapiens]
l3h11fs.f1	225	8.4e-17	225	524	gi|6005808 ref|NP_009161.1|| endoplasmic reticulum alpha-mannosidase 1>gb|AAD45504.1|AF145732_1 (AF145732) endoplasmic reticulumalpha-mannosidase I [Homo sapiens]
7. Quinate metabolism (4)					
<quinate repressor>					
p3g01fs.r1	245	1.1e-18	10	513	sp|P11637|QA1S_NEUCR QUINATE REPRESSOR >pir||S04255 regulatory protein qa-1S -Neurospora crassa >emb|CAA32753.1| (X14603) repressor [Neurosporacrassa]
<pentafunctional arom polypeptide>					
Contig42	889	1.2e-93	6	887	sp|P07547|ARO1_EMENI PENTAFUNCTIONAL AROM POLYPEPTIDE [INCLUDES:3-DEHYDROQUINATE SYNTHASE ; 3-DEHYDROQUINATE DEHYDRATASE(3-DEHYDROQUINASE); SHIKIMATE 5-DEHYDROGENASE ; SHIKIMATE KINASE;EPSP SYNTHASE] >emb|CAA28836.1| (X05204) arom polypeptide[Emericella nidulans]

m4c09fs.r1	598	5.4e-56	36	494	emb|CAB75770.1| (AL157734) pentafunctional arom polypeptide [includes:3-dehydroquinate synthase(ec 4.6.1.3); 3-dehydroquinate dehydratase[Schizosaccharomyces pombe]
Contig35	185	5.8e-12	208	516	gi|6320332 ref|NP_010412.1|ARO1| pentafunctional arom polypeptide (contains:3-dehydroquinate synthase, 3-dehydroquinate dehydratase(3-dehydroquinase), shikimate 5-dehydrogenase, shikimate kinase,and epsp synthase); Aro1p >sp|P08566|ARO1_YEAST PENTAFUNCTIONALAROM POLYPEPTIDE [INCLUDES: 3-DEHYDROQUINATE SYNTHASE ;3-DEHYDROQUINATE DEHYDRATASE (3-DEHYDROQUINASE); SHIKIMATE5-DEHYDROGENASE ; SHIKIMATE
8. Sorbitol metabolism (2)					
<SORBITOL UTILIZATION PROTEIN>					
Contig983	653	2.8e-63	5	844	sp|P87219|SOU1_CANAL SORBITOL UTILIZATION PROTEIN SOU1 >gb|AAC24463.1|(AF002134) Sou1p [Candida albicans]
Contig615	400	2.2e-36	118	489	sp|P87219|SOU1_CANAL SORBITOL UTILIZATION PROTEIN SOU1>gb|AAC24463.1|(AF002134) Sou1p [Candida albicans]
9. others (6)					
<aldose reductase>					
Contig380	368	5.3e-33	20	514	pir||T37996 probable aldose reductase - fission yeast (Schizosaccharomycespombe) >emb|CAB10120.1| (Z97209) putative aldose reductase[Schizosaccharomyces pombe]
Contig381	175	2.1e-12	184	507	pir||T37996 probable aldose reductase - fission yeast (Schizosaccharomycespombe) >emb|CAB10120.1| (Z97209) putative aldosereductase[Schizosaccharomyces pombe]
<L-sorbose dehydrogenase>					
s3g05fs.r1	165	1.7e-10	116	400	dbj|BAA13145.1| (D86622) L-sorbose dehydrogenase, FAD dependent [Gluconobacteroxydans]
<D-xylose reductase>					
h4b07fs.r1	358	5.2e-32	93	389	gb|AAF61912.1|AF2196 (AF219625) D-xylose reductase [Aspergillus niger]
<squalene synthase>					
i3e11fs.r1	178	4.7e-12	185	463	gb|AAD22408.1|AF0924 (AF092497) squalene synthase [Yarrowia lipolytica]
<B2-aldehyde-forming enzyme>					
Contig1016	104	0.019	641	880	gb|AAB62250.1| (AF005405) B2-aldehyde-forming enzyme [Schizophyllum commune]

II. Metabolism of Amino acids and Related Molecules

1. arginine metabolism (9)

<arginine N-methyltransferase>					
Contig1011	102	0.11	169	414	emb|CAB63498.1| (AL133498) probable arginine N-methyltransferase[Schizosaccharomyces pombe]
1.1 arginine anabolism-glutamine, CO2 to arginine					
<ARGININOSUCCINATE SYNTHASE>					
Contig286	534	1.2e-50	50	517	gi|6324514 ref|NP_014583.1|ARG1| arginosuccinate synthetase; Arg1p>sp|P22768|ASSY_YEAST ARGININOSUCCINATE SYNTHASE(CITRULLINE-ASPARTATE LIGASE) >pir||AJBYRS argininosuccinatesynthase (EC 6.3.4.5) - yeast (Saccharomyces cerevisiae)>emb|CAA62528.1| (X91067) argininosuccinate synthase [Saccharomycescerevisiae] >emb|CAA99067.1| (Z74800) ORF YOL058w [Saccharomycescerevisiae]
Contig285	307	1.4e-26	141	518	gb|AAA34437.1| (M35237) argininosuccinate synthetase (ARG1; E.C.6.8.4.5)[Saccharomyces cerevisiae]
<ARGININOSUCCINATE LYASE>					
Contig472	523	2.2e-49	275	814	sp|P50514|ARLZ_SCHPO PROBABLE ARGININOSUCCINATE LYASE (ARGINOSUCCINASE) (ASAL)>pir||T39462 probable argininosuccinate lyase (EC 4.3.2.1) -fission yeast (Schizosaccharomyces pombe) >gb|AAA58961.1| (U13259)L-argininosuccinate lyase [Schizosaccharomyces pombe]>emb|CAB51335.1| (AL096874) probable argininosuccinate lyase (EC4.3.2.1) [Schizosaccharomyces pombe]
Contig891	448	1.9e-41	169	654	sp|P50514|ARLZ_SCHPO PROBABLE ARGININOSUCCINATE LYASE (ARGINOSUCCINASE) (ASAL)>pir||T39462 probable argininosuccinate lyase (EC 4.3.2.1) -fission yeast (Schizosaccharomyces pombe) >gb|AAA58961.1| (U13259)L-argininosuccinate lyase [Schizosaccharomyces pombe]>emb|CAB51335.1| (AL096874) probable argininosuccinate lyase (EC4.3.2.1) [Schizosaccharomyces pombe]
1.2 arginine catabolism-arginine to proline					
<ARGINASE-also see urea cycle>					
j2h06fs.r1	395	6.7e-36	112	468	sp|P33280|ARGI_NEUCR ARGINASE >gb|AAC82503.1| (L20687) arginase 36 kDa isoform[Neurospora crassa]
j2h06fs.r1	395	6.7e-36	112	468	gb|AAC82504.1| (L20687) arginase 41kDa isoform [Neurospora crassa]
r3h07fs.r1	339	6.7e-30	215	478	pir||T39168 arginase family protein - fission yeast (Schizosaccharomycespombe)
<ARG-6 PROTEIN PRECURSOR>					
o4g11fs.f1	407	4e-36	229	519	sp|P54898|AR56_NEUCR ARG-6 PROTEIN PRECURSOR [CONTAINS:N-ACETYL-GAMMA-GLUTAMYL-PHOSPHATE REDUCTASE (N-ACETYL-GLUTAMATESEMIALDEHYDE DEHYDROGENASE) (NAGSA DEHYDROGENASE); ACETYLGLUTAMATEKINASE (NAG KINASE) (AGK) (N-ACETYL-L-GLUTAMATE5-PHOSPHOTRANSFERASE)]

					>pir||A53429 acetylglutamate kinase (EC2.7.2.8) / N-acetyl-gamma-glutamyl-phosphate reductase (EC1.2.1.38) precursor, mitochondrial - Neurospora
2. asparagine metabolism (3)					
<asparagine synthetase>					
d2f09fs.r1	537	6.2e-51	82	501	pir||T39308 asparagine synthetase - fission yeast (Schizosaccharomyces pombe)>emb|CAA17925.1| (AL022117) asparagine synthetase[Schizosaccharomyces pombe]
<L-asparaginase II>					
b1f10fs.r1	336	1.3e-29	15	428	gb|AAF00929.1|AF1814 (AF181498) L-asparaginase II [Rhizobium etli]
<L-ASPARAGINASE>					
c4f04fs.r1	424	6.5e-39	31	465	sp|O34482|ASG2_BACSU PROBABLE L-ASPARAGINASE (L-ASPARAGINE AMIDOHYDROLASE)>pir||F69754 asparaginase homolog yccC - Bacillus subtilis>dbj|BAA22230.1| (AB000617) YccC [Bacillus subtilis]>emb|CAB12063.1| (Z99105) similar to asparaginase [Bacillussubtilis]
3. aspartic acid metabolism (5)					
<ASPARTATE AMINOTRANSFERASE>					
Contig825	671	4e-65	266	907	sp|P00503|AATC_PIG ASPARTATE AMINOTRANSFERASE, CYTOPLASMIC (TRANSAMINASE A)(GLUTAMATE OXALOACETATE TRANSAMINASE-1) >pir||XNPGDC aspartatetransaminase (EC 2.6.1.1), cytosolic - pig >gb|AAA53531.1| (M24088)cytosolic aspartate aminotransferase [Sus scrofa]
Contig646	397	4.8e-36	70	489	gb|AAB19394.1| aspartate aminotransferase [Saccharomyces cerevisiae, PeptidePartial, 414 aa]
Contig492	381	1.8e-34	104	445	pir||T15494 hypothetical protein C14F11.1 - Caenorhabditis elegans>gb|AAA80361.1| (U39645) similar to aspartate aminotransferase[Caenorhabditis elegans]
Contig266	226	2.2e-17	347	547	sp|P12344|AATM_BOVIN ASPARTATE AMINOTRANSFERASE, MITOCHONDRIAL PRECURSOR(TRANSAMINASE A) (GLUTAMATE OXALOACETATE TRANSAMINASE-2)>pir||S35960 aspartate transaminase (EC 2.6.1.1) precursor,mitochondrial - bovine >emb|CAA80960.1| (Z25466) aspartateaminotransferase [Bos taurus]
n2g11fs.r1	170	2.5e-11	24	389	sp|O67781|AAT_AQUAE ASPARTATE AMINOTRANSFERASE (TRANSAMINASE A) (ASPAT)>pir||A70469 aspartate aminotransferase - Aquifex aeolicus>gb|AAC07746.1| (AE000766) aspartate aminotransferase [Aquifexaeolicus]
4. cysteine metabolism and biosynthesis (4)					
<S-adenosyl-L-homocysteine hydrolase>					

Contig1043	1795	3e-184	28	1374	gi|6320882 ref|NP_010961.1|SAH1| putative S-adenosyl-L-homocysteine hydrolase;Sah1p >sp|P39954|SAHH_YEAST DENOSYLHOMOCYSTEINASE(S-ADENOSYL-L-HOMOCYSTEINE HYDROLASE) (ADOHCYASE)>pir||S50546adenosylhomocysteinase (EC 3.3.1.1) - yeast (Saccharomycescerevisiae) >gb|AAB64578.1| (U18796) Sam1p:Adenosylhomocysteinase[Saccharomyces cerevisiae]
<CYSTATHIONINE GAMMA-SYNTHASE>					
b4b12fs.r1	553	1.4e-52	31	450	sp|P38675|MET7_NEUCR CYSTATHIONINE GAMMA-SYNTHASE (O-SUCCINYLHOMOSERINE(THIOL)-LYASE) >pir||JQ1524 O-succinylhomoserine (thiol)-lyase (EC4.2.99.9) met-7 chain - Neurospora crassa
Contig476	484	2.9e-45	225	629	sp|P38675|MET7_NEUCR CYSTATHIONINE GAMMA-SYNTHASE (O-SUCCINYLHOMOSERINE(THIOL)-LYASE) >pir||JQ1524 O-succinylhomoserine (thiol)-lyase (EC4.2.99.9) met-7 chain - Neurospora crassa
<CYSTEINE SYNTHASE>					
c3a07fs.f1	125	1.9e-06	34	126	sp|P50867|CYSK_EMENI CYSTEINE SYNTHASE (O-ACETYLSERINE SULFHYDRYLASE)(O-ACETYLSERINE (THIOL)-LYASE) (CSASE) >gb|AAC06128.1| (U19395)cysteine synthase [Emericella nidulans]
5. glutamine metabolism and biosynthesis (4)					
<glutamine synthase>					
Contig815	969	1.1e-96	170	736	gb|AAD52617.1|AF1754 (AF175498) glutamine synthase [Haematonectriahaematococca]
r3b02fs.r1	607	2.5e-58	49	465	gb|AAD52617.1|AF1754 (AF175498) glutamine synthase [Haematonectriahaematococca]
<glutamine amidotransferase>					
n3a03fs.f1	97	0.13	152	493	emb|CAA07652.1| (AJ007747) putative glutamine amidotransferase [Bordetellabronchiseptica]
<glutamyl-trna synthetase>					
t1h11fs.r1	556	6.7e-53	17	544	pir||T37830 glutamyl-trna synthetase - fission yeast (Schizosaccharomycespombe) >emb|CAB11515.1| (Z98849) glutamyl-trnasynthetase[Schizosaccharomyces pombe]
6. glycine metabolism (6)					
<THREONINE ALDOLASE>					
o1b12fs.r1	405	6.8e-37	92	511	emb|CAA06545.1| (AJ005442) threonine aldolase [Eremothecium gossypii]
Contig330	316	1.8e-27	63	494	pir||T38302 probable threonine aldolase - fission yeast (Schizosaccharomycespombe) >emb|CAB16235.1| (Z99163) putative threonine aldolase[Schizosaccharomyces pombe]

Contig239	204	3.8e-15	71	520	pir||T38302 probable threonine aldolase - fission yeast (Schizosaccharomycespombe) >emb|CAB16235.1| (Z99163) putative threonine aldolase[Schizosaccharomyces pombe]
<THREONINE SYNTHASE>					
o1b03fs.r1	517	8.1e-49	41	514	sp|Q42598|THRC_SCHPO THREONINE SYNTHASE >pir||S49036 threonine synthase (EC4.2.99.2) - Arabidopsis thaliana >pir||T39213 threonine synthase -fission yeast (Schizosaccharomyces pombe) >emb|CAA86405.1| (Z46263)threonine synthase [Schizosaccharomyces pombe] >emb|CAB16415.1|(Z99262) threonine synthase [Schizosaccharomyces pombe]
o1b03fs.f1	265	2e-21	215	556	sp|Q42598|THRC_SCHPO THREONINE SYNTHASE >pir||S49036 threonine synthase (EC4.2.99.2) - Arabidopsis thaliana >pir||T39213 threonine synthase -fission yeast (Schizosaccharomyces pombe) >emb|CAA86405.1| (Z46263)threonine synthase [Schizosaccharomyces pombe] >emb|CAB16415.1|(Z99262) threonine synthase [Schizosaccharomyces pombe]
<GLYCINE DEHYDROGENASE>					
Contig154	276	5.9e-22	315	530	sp|Q09785|GCSP_SCHPO PUTATIVE GLYCINE DEHYDROGENASE [DECARBOXYLATING],MITOCHONDRIAL PRECURSOR (GLYCINE DECARBOXYLASE) (GLYCINE CLEAVAGESYSTEM P-PROTEIN) >pir||S62435 hypothetical protein SPAC13G6.06c-fission yeast (Schizosaccharomyces pombe) >pir||T37641 probableglycine dehydrogenase (decarboxylating) - fission yeast(Schizosaccharomyces pombe) >emb|CAA91099.1| (Z54308) putativeglycine dehydrogenase
7. histidine metabolism (6)					
<HISTIDINOL-PHOSPHATASE>					
Contig228	263	7.5e-22	6	452	sp|O14059|HIS9_SCHPO PROBABLE HISTIDINOL-PHOSPHATASE >pir||T41045histidinol-phosphatase - fission yeast (Schizosaccharomyces pombe)>emb|CAA20439.1| (AL031324) histidinol-phosphatase[Schizosaccharomyces pombe]
n2d03fs.r1	223	1.2e-17	201	497	sp|O14059|HIS9_SCHPO PROBABLE HISTIDINOL-PHOSPHATASE >pir||T41045histidinol-phosphatase - fission yeast (Schizosaccharomyces pombe)>emb|CAA20439.1| (AL031324) histidinol-phosphatase[Schizosaccharomyces pombe]
<HISTIDINOL-PHOSPHATE AMINOTRANSFERASE>					
Contig413	429	1.8e-39	27	497	sp|P56099|HIS8_CANMA HISTIDINOL-PHOSPHATE AMINOTRANSFERASE (IMIDAZOLEACETOL-PHOSPHATE TRANSAMINASE) >pir||A48329 histidinol-

					phosphatetransaminase (EC 2.6.1.9) - yeast (Candida maltosa)>emb|CAA35189.1| (X17310) histidinol-phosphate aminotransferase[Candida maltosa]
p1h01fs.f1	263	7.5e-22	144	527	sp|P56099|HIS8_CANMA HISTIDINOL-PHOSPHATE AMINOTRANSFERASE (IMIDAZOLEACETOL-PHOSPHATE TRANSAMINASE) >pir||A48329 histidinol-phosphatetransaminase (EC 2.6.1.9) - yeast (Candida maltosa)>emb|CAA35189.1| (X17310) histidinol-phosphate aminotransferase[Candida maltosa]
<imidazole glycerol phosphate dehydratase-biosynthesis of histidine>					
k4h02fs.r1	573	9.5e-55	77	508	sp|P34041|HIS7_TRIHA IMIDAZOLEGLYCEROL-PHOSPHATE DEHYDRATASE (IGPD)>pir||S26196 imidazoleglycerol-phosphate dehydratase (EC 4.2.1.19)-fungus (Trichoderma harzianum) >emb|CAA77617.1| (Z11528)imidazoleglycerolphosphate [Trichoderma harzianum]
k4h02fs.f1	195	1.2e-14	229	660	sp|O42621|HIS7_MAGGR IMIDAZOLEGLYCEROL-PHOSPHATE DEHYDRATASE (IGPD)>gb|AAB88888.1| (AF027980) imidazole glycerol phosphate dehydratase[Magnaporthe grisea]
8. isoleucine metabolism (5)					
<2,3-DIHYDROXYACID HYDROLYASE-4th step in iso & val biosyn>					
r4e10fs.f1	404	8.6e-37	162	521	gi|6322476 ref|NP_012550.1|ILV3| dihydroxyacid dehydratase;Ilv3p>sp|P39522|ILV3_YEAST DIHYDROXY-ACID DEHYDRATASE,MITOCHONDRIALPRECURSOR (DAD) (2,3-DIHYDROXY ACID HYDROLYASE)>pir||S55205dihydroxy-acid dehydratase (EC 4.2.1.9) - yeast (Saccharomycescerevisiae) >emb|CAA60939.1| (X87611) dihydroxyacid dehydratase[Saccharomyces cerevisiae] >emb|CAA89540.1| (Z49516) ORF YJR016c[Saccharomyces cerevisiae]
<acetolactate synthase-biosynthesis of isoleucine, leucine and valine>					
m4d02fs.f1	320	6.4e-28	129	446	gi|6319837 ref|NP_009918.1|ILV6| Small regulatory subunit of Acetolactatesynthase; Ilv6p >sp|P25605|ILV6_YEAST ACETOLACTATE SYNTHASE SMALLSUBUNIT PRECURSOR (AHAS) (ACETOHYDROXY-ACID SYNTHASE SMALL SUBUNIT)(ALS) >emb|CAA42350.1| (X59720) YCL009c, len:309 [Saccharomycescerevisiae]
<ISOPROPYLMALATE DEHYDRATASE>					
s1b10fs.r1	492	1.2e-45	121	528	sp|Q00464|LEU2_CANMA 3-ISOPROPYLMALATE DEHYDRATASE (ISOPROPYLMALATE ISOMERASE)(ALPHA-IPM ISOMERASE) (IPMI) >gb|AAB03335.1| (U60167) alphaisopropylmalate isomerase [Candida maltosa]
e1a08fs.r1	468	6e-43	145	489	sp|P49601|LEU2_USTMA 3-ISOPROPYLMALATE DEHYDRATASE (ISOPROPYLMALATE ISOMERASE)(ALPHA-IPM ISOMERASE) (IPMI) >gb|AAA34226.1| (L20832) LEU1[Ustilago maydis]

s1b10fs.f1 224 1.4e-16 2 187 sp|P55251|LEU2_RHIPU 3-ISOPROPYLMALATE DEHYDRATASE (ISOPROPYLMALATE ISOMERASE)(ALPHA-IPM ISOMERASE) (IPMI) >dbj|BAA11052.1| (D67033) LeuA[Rhizomucor pusillus]

9. methionine metabolism (4)

<**methionine synthase**-last step in met biosynthesis>

Contig1038 2347 8.8e-243 8 1723 gb|AAF33834.1| (AF226997) methionine synthase [Cladosporium fulvum]

Contig951 533 2.3e-50 117 494 gb|AAF33834.1| (AF226997) methionine synthase [Cladosporium fulvum]

<**S-adenosylmethionine synthetase**-(EC 2.5.1.6)>

Contig1015 1431 1.2e-145 446 1477 sp|P48466|METK_NEUCR S-ADENOSYLMETHIONINE SYNTHETASE (METHIONINEADENOSYLTRANSFERASE) (ADOMET SYNTHETASE) >pir||S65800 methionineadenosyltransferase (EC 2.5.1.6) - Neurospora crassa>gb|AAC49260.1| (U21547) S-adenosylmethionine synthetase[Neurospora crassa] >prf||2210293A Met(S-adenosyl) synthetase[Neurospora crassa]

Contig1028 266 3.5e-22 650 826 sp|P48466|METK_NEUCR S-ADENOSYLMETHIONINE SYNTHETASE (METHIONINEADENOSYLTRANSFERASE) (ADOMET SYNTHETASE) >pir||S65800 methionineadenosyltransferase (EC 2.5.1.6) - Neurospora crassa>gb|AAC49260.1| (U21547) S-adenosylmethionine synthetase[Neurospora crassa] >prf||2210293A Met(S-adenosyl) synthetase[Neurospora crassa]

10. tryptophan metabolism and synthesis (3)

<**anthranilate phosphoribosyltransferase**-2nd step in tryp biosyn>

s1b01fs.r1 156 4.4e-09 122 433 pir||T01234 probable anthranilate phosphoribosyltransferase (EC 4.2.1.18)F6N23.8 - Arabidopsis thaliana >gb|AAC13630.1| (AF058919)F6N23.8gene product [Arabidopsis thaliana] >emb|CAB80879.1| (AL161472)putative phosphoribosylanthranilate transferase [Arabidopsisthaliana]

Contig349 145 1.1e-08 178 516 sp|O60122|TRPD_SCHPO ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE >pir||T39600phosphoribosylanthranilate transferase - fission yeast(Schizosaccharomyces pombe) >emb|CAA19028.1| (AL023554)phosphoribosylanthranilate transferase; tryptophan biosynthesispathway [Schizosaccharomyces pombe]

<**anthranilate synthase Component I**-synthesis of tryptophan>

Contig912 419 1.7e-93 348 773 gi|6320935 ref|NP_011014.1|TRP2| anthranilate synthase Component I; Trp2p>sp|P00899|TRPE_YEAST ANTHRANILATE SYNTHASE COMPONENT I>pir||NNBY1anthranilate synthase (EC 4.1.3.27) component I - yeast(Saccharomyces cerevisiae) >emb|CAA48402.1|(X68327)anthranilatesynthase (component 1)

					[Saccharomycescerevisiae]> gb|AAB64645.1|(U18839) Trp2p: anthranilate synthase component I [Saccharomycescerevisiae]
11. aromatic amino acid metabolism (1)					
<aromatic-L-amino-acid decarboxylase>					
a3c03fs.r1	431	1.2e-39	13	462	sp|P22781|DCD_CAVPO AROMATIC-L-AMINO-ACID DECARBOXYLASE (DOPADECARBOXYLASE)(DDC) >pir||DEGPA aromatic-L-amino-acid decarboxylase (EC4.1.1.28)- guinea pig >gb|AAA51530.1| (M58049) aromatic-L-amino aciddecarboxylase [Cavia porcellus]
12. glutamate metabolism (8)					
<glutamate synthase(NADH)precursor>					
Contig577	297	1.7e-25	184	492	dbj|BAA13827.1| (D89165) similar to Medicago sativa glutamate synthase(NADH)precursor, SWISS-PROT Accession Number Q03460 [Schizosaccharomycespombe]
o2a01fs.r1	380	1.4e-32	7	441	sp|Q03460|GLSN_MEDSA GLUTAMATE SYNTHASE [NADH] PRECURSOR (NADH-GOGAT)>pir||JQ1977 glutamate synthase (NADH) (EC 1.4.1.14) - alfalfa>gb|AAB46617.1| (L01660) NADH-glutamate synthase [Medicago sativa]
<glutamate synthase>					
c4c01fs.r1	568	3.4e-54	12	461	gb|AAD41651.1| (AF073360) glutamate synthase [Emericella nidulans]<NAD(+)-specific glutamate dehydrogenase>
o1c01fs.r1	502	3e-46	32	502	gb|AAB28355.1| (S66039) NAD(+)-specific glutamate dehydrogenase,NAD-GDH {EC1.4.1.2} [Neurospora crassa, Peptide, 1047 aa] >prf||1919235A Gludehydrogenase [Neurospora crassa]
<glutamic acid decarboxylase>					
Contig1031	1726	8.2e-187	238	1605	dbj|BAA88152.1| (AB025422) glutamic acid decarboxylase [Aspergillus oryzae]
Contig15	351	3.9e-31	125	448	dbj|BAA88152.1| (AB025422) glutamic acid decarboxylase [Aspergillus oryzae]
<acetylglutamate kinase>					
Contig446	421	1.2e-37	235	576	sp|P54898|AR56_NEUCR ARG-6 PROTEIN PRECURSOR [CONTAINS:N-ACETYL-GAMMA-GLUTAMYL-PHOSPHATE REDUCTASE (N-ACETYL-GLUTAMATESEMIALDEHYDE DEHYDROGENASE) (NAGSA DEHYDROGENASE); ACETYLGLUTAMATEKINASE (NAG KINASE) (AGK) (N-ACETYL-L-GLUTAMATE5-PHOSPHOTRANSFERASE)] >pir||A53429 acetylglutamate kinase (EC2.7.2.8) / N-acetyl-gamma-glutamyl-phosphate reductase (EC1.2.1.38) precursor, mitochondrial - Neurospora

o4b03fs.r1	396	5.6e-35	169	513	sp|P54898|AR56_NEUCR ARG-6 PROTEIN PRECURSOR [CONTAINS:N-ACETYL-GAMMA-GLUTAMYL-PHOSPHATE REDUCTASE (N-ACETYL-GLUTAMATESEMIALDEHYDE DEHYDROGENASE) (NAGSA DEHYDROGENASE);ACETYLGLUTAMATEKINASE (NAG KINASE) (AGK) (N-ACETYL-L-GLUTAMATE5-PHOSPHOTRANSFERASE)] >pir||A53429 acetylglutamate kinase (EC2.7.2.8) / N-acetyl-gamma-glutamyl-phosphate reductase (EC1.2.1.38) precursor, mitochondrial - Neurospora

13. phenylalanine metabolism (3)

<phosphoethanolamine cytidylyltransferase>

n1b09fs.r1	367	6.7e-33	152	481	pir||T37720 ethanolamine-phosphate cytidylyltransferase (EC2.7.7.14) -fission yeast (Schizosaccharomyces pombe)>emb|CAB52424.1|(AL109770) phosphoethanolamine cytidylyltransferase (EC2.7.7.14)[Schizosaccharomyces pombe]

<HOMOGENTISATE 1,2-DIOXYGENASE>

Contig39	542	2e-51	18	506	sp|Q00667|HGD_EMENI HOMOGENTISATE 1,2-DIOXYGENASE (HOMOGENTISICASE)(HOMOGENTISATE OXYGENASE) (HOMOGENTISIC ACID OXIDASE)>pir||A574353,4-dihydroxyphenylacetate 2,3-dioxygenase (EC 1.13.11.15)-Emericella nidulans >gb|AAC49071.1| (U30797) 2,5dihydroxyphenylacetate oxidase [Emericella nidulans]>emb|CAA05042.1| (AJ001836) homogentisate dioxygenase [Emericellanidulans]
l3b06fs.f1	476	1.9e-44	109	522	sp|Q00667|HGD_EMENI HOMOGENTISATE 1,2-DIOXYGENASE (HOMOGENTISICASE)(HOMOGENTISATE OXYGENASE) (HOMOGENTISIC ACID OXIDASE)>pir||A574353,4-dihydroxyphenylacetate 2,3-dioxygenase (EC 1.13.11.15)-Emericella nidulans >gb|AAC49071.1| (U30797) 2,5dihydroxyphenylacetate oxidase [Emericella nidulans]>emb|CAA05042.1| (AJ001836) homogentisate dioxygenase [Emericellanidulans]

14. lysine (4)

<alpha-aminoadipate reductase>

j3b04fs.r1	529	9.5e-49	18	545	emb|CAA74300.1| (Y13967) alpha-aminoadipate reductase large subunit[Penicillium chrysogenum]
j3b04fs.f1	216	2.2e-15	241	504	emb|CAA74300.1| (Y13967) alpha-aminoadipate reductase large subunit[Penicillium chrysogenum]

<homocitrate synthase-for biosynthesis of lysine>

Contig163	179	4e-12	233	409	sp|O94225|HOSM_PENCH HOMOCITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR>emb|CAA11503.1| (AJ223630) homocitrate synthase [Penicilliumchrysogenum]
Contig679	104	0.00044	433	525	gi|6320071 ref|NP_010151.1|LYS21| homocitrate synthase, highly homologous toYDL182W; Lys21p >sp|Q12122|HOSM_YEAST PROBABLE HOMOCITRATESYNTHASE, MITOCHONDRIAL PRECURSOR >pir||S67674 homocitrate synthasehomolog YDL131w - yeast (Saccharomyces cerevisiae) >emb|CAA65629.1|(X96876) putative ORF [Saccharomyces cerevisiae] >emb|CAA98700.1|(Z74179) ORF YDL131w [Saccharomyces cerevisiae]
15. tyrosine metabolism					
<TYROSINASE>					
l1h10fs.f1	147	1.6e-08	303	518	sp|Q92396|TYRO_PODAN TYROSINASE (MONOPHENOL MONOOXYGENASE)>gb|AAB07484.1|(U66807) tyrosinase [Podospora anserina] >gb|AAB07516.1| (U66808)tyrosinase [Podospora anserina]
16. alanine metabolism (1)					
<alanine racemase>					
o1b12fs.f1	230	5.5e-18	214	567	gb|AAD47837.1|AF1694 (AF169478) alanine racemase [Cochliobolus carbonum]
17. others (3)					
<orfT-transaminase type I>					
Contig454	177	1e-19	35	223	emb|CAA04685.1| (AJ001330) orfT [Lactobacillus sakei]
<aminobutyrate aminotransferase>					
Contig97	588	7.3e-65	234	725	sp|P14010|GATA_EMENI 4-AMINOBUTYRATE AMINOTRANSFERASE (GAMMA-AMINO-N-BUTYRATETRANSAMINASE) (GABA TRANSAMINASE) (GABA AMINOTRANSFERASE)>pir||JQ0197 4-aminobutyrate transaminase (EC2.6.1.19)-Emericella nidulans >emb|CAA33674.1| (X15647)gamma-amino-n-butyrate transaminase [Emericella nidulans]
b1d10fs.f1	322	7e-28	210	500	sp|P14010|GATA_EMENI 4-AMINOBUTYRATE AMINOTRANSFERASE (GAMMA-AMINO-N-BUTYRATETRANSAMINASE) (GABA TRANSAMINASE) (GABA AMINOTRANSFERASE)>pir||JQ0197 4-aminobutyrate transaminase (EC 2.6.1.19)-Emericella nidulans >emb|CAA33674.1| (X15647)gamma-amino-n-butyrate transaminase [Emericella nidulans]

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

1. Nucleotide metabolism (5)
<NUCLEOSIDE DIPHOSPHATE KINASE>

Contig904	627	1.5e-60	23	478	dbj|BAA83495.1| (D88148) nucleoside diphosphate kinase [Neurospora crassa]
<ribose-phosphate pyrophosphokinase-purine,pyrimidine biosyn, also his and tryptophan biosyn>					
t2h04fs.r1	501	4.1e-47	107	544	pir||T40366 probable ribose-phosphate pyrophosphokinase - fission yeast(Schizosaccharomyces pombe) >emb|CAB09126.1| (Z95620) putativeribose-phosphate pyrophosphokinase [Schizosaccharomyces pombe]
<ribonucleotide reductase>					
n1g05fs.r1	745	6.1e-73	7	486	gb|AAD49743.1| (AF171697) ribonucleotide reductase large subunit [Neurosporacrassa]
n1b05fs.r1	680	4.7e-66	11	484	gb|AAD49743.1| (AF171697) ribonucleotide reductase large subunit [Neurosporacrassa]
Contig276	121	1.6e-05	311	496	emb|CAB77640.1| (AJ390500) ribonucleotide reductase large subunit [Candidaalbicans]
2. Purine metabolism					
2.1. inosine mono phosphate de novo biosynthesis					
<amidophosphoribosyltransferase>					
k1b02fs.r1	441	1e-40	12	461	sp|P41390|PUR1_SCHPO AMIDOPHOSPHORIBOSYLTRANSFERASE (GLUTAMINEPHOSPHORIBOSYLPYROPHOSPHATE AMIDOTRANSFERASE) (ATASE)>pir||S43526amidophosphoribosyltransferase (EC 2.4.2.14) - fission yeast(Schizosaccharomyces pombe) >emb|CAA51034.1| (X72293) PRPPamidotransferase [Schizosaccharomyces pombe] >emb|CAB11280.1|(Z98602) amidophosphoribosyltransferase [Schizosaccharomyces pombe]
n4g10fs.r1	339	1e-29	130	489	gi|6323958 ref|NP_014029.1|ADE4| phosphoribosylpyrophosphate amidotransferase;Ade4p >sp|P04046|PUR1_YEAST AMIDOPHOSPHORIBOSYLTRANSFERASE(GLUTAMINE PHOSPHORIBOSYLPYROPHOSPHATE AMIDOTRANSFERASE) (ATASE)>pir||S53970 amidophosphoribosyltransferase (EC2.4.2.14) - yeast(Saccharomyces cerevisiae) >emb|CAA89133.1| (Z49212) Ade4p[Saccharomyces cerevisiae]
<inosine 5'-monophosphate dehydrogenase>					
e1d05fs.r1	511	3.7e-48	18	476	gb|AAF13230.1|AF1969 (AF196975) inosine 5'-monophosphatedehydrogenase[Pneumocystis carinii]
<5-aminoimidazole-4-carboxamideribonucleotide (AICAR) transformylase-biosynthesis of inosine monophosphate>					
Contig277	848	6.2e-84	4	696	gi|6323768 ref|NP_013839.1|ADE17|5-aminoimidazole-4-carboxamideribonucleotide (AICAR) transformylase/IMP cyclohydrolase; Ade17p>sp|P38009|PU92_YEAST BIFUNCTIONAL PURINE BIOSYNTHESIS

				PROTEINADE17 [INCLUDES:PHOSPHORIBOSYLAMINOIMIDAZOLECARBOXAMIDEFORMYLTRANSFERASE (AICAR TRANSFORMYLASE); IMP CYCLOHYDROLASE(INOSINICASE) (IMP SYNTHETASE) (ATIC)]>pir||S54489phosphoribosylaminoimidazolecarboxamide
a1c07fs.r1	454	3.8e-42	21 431	gi|6323768 ref|NP_013839.1|ADE17|5-aminoimidazole-4-carboxamideribonucleotide (AICAR) transformylase/IMP cyclohydrolase; Ade17p>sp|P38009|PU92_YEAST BIFUNCTIONAL PURINE BIOSYNTHESIS PROTEINADE17 [INCLUDES:PHOSPHORIBOSYLAMINOIMIDAZOLECARBOXAMIDEFORMYLTRANSFERASE (AICAR TRANSFORMYLASE); IMP CYCLOHYDROLASE(INOSINICASE) (IMP SYNTHETASE) (ATIC)] >pir||S54489phosphoribosylaminoimidazolecarboxamide
Contig393	291	4.3e-24	229 534	gi|6323768 ref|NP_013839.1|ADE17|5-aminoimidazole-4-carboxamideribonucleotide (AICAR) transformylase/IMP cyclohydrolase; Ade17p>sp|P38009|PU92_YEAST BIFUNCTIONAL PURINE BIOSYNTHESIS PROTEINADE17 [INCLUDES:PHOSPHORIBOSYLAMINOIMIDAZOLECARBOXAMIDEFORMYLTRANSFERASE (AICAR TRANSFORMYLASE); IMP CYCLOHYDROLASE(INOSINICASE) (IMP SYNTHETASE) (ATIC)]>pir||S54489phosphoribosylaminoimidazolecarboxamide

2.2. other purine metabolic enzymes

<INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE>

e1d05fs.f1	299	4.1e-25	216 491	gi|6323585 ref|NP_013656.1|YML056C| Yml056cp>sp|P50094|IMH3_YEAST PROBABLEINOSINE-5'-MONOPHOSPHATE DEHYDROGENASE (IMPDEHYDROGENASE) (IMPDH)(IMPD) >pir||S50890 IMP dehydrogenase (EC 1.1.1.205)YML056c - yeast (Saccharomyces cerevisiae) >emb|CAA86719.1| (Z46729)putativeinosine-5'-monophoshate dehydrogenase [Saccharomyces cerevisiae]

<aminoimidazole ribonucleotide carboxylase-purine biosynthesis enzyme>

e1e09fs.r1	285	1.6e-23	80 478	sp|Q01930|PUR6_PICME PHOSPHORIBOSYLAMINOIMIDAZOLE CARBOXYLASE (AIRCARBOXYLASE) (AIRC) >pir||S39112 phosphoribosylaminoimidazolecarboxylase (EC 4.1.1.21) - yeast (Pichia methanolica)>emb|CAA54041.1| (X76529)5-aminoimidazoleribonucleotide-carboxilase [Pichia methanolica]
e1e09fs.f1	272	7.1e-22	188 517	pir||T02535 phosphoribosylaminoimidazole carboxylase (EC 4.1.1.21)-Arabidopsis thaliana >gb|AAC23639.1|

					(AC004684)putativephosphoribosylaminoimidazole carboxylase [Arabidopsis thaliana]
3. Pyrimidine metabolism (3)					
<orotate reductase>					
Contig313	418	6e-54	187	621	pir||S72324 orotate reductase (NADH) (EC 1.3.1.14) - Emericella nidulans>gb|AAA86932.1| (U47318) dihydroorotate dehydrogenase [Emericellanidulans]
r3a11fs.f1	190	3e-13	92	259	pir||S72324 orotate reductase (NADH) (EC 1.3.1.14) - Emericella nidulans>gb|AAA86932.1| (U47318) dihydroorotate dehydrogenase [Emericellanidulans]
<orotidine-5'-phosphate decarboxylase>					
s2e10fs.r1	362	2.3e-32	341	613	pir||S14132 orotidine-5'-phosphate decarboxylase (EC 4.1.1.23) - fungus(Trichoderma reesei)
4. Salvage(reuse) of the bases (1)					
<ATP PHOSPHORIBOSYLTRANSFERASE-synthesizes nucleoside 5'-phosphates directly from free base>					
q3d05fs.f1	204	1.3e-15	303	503	sp|P40373|HIS1_SCHPO ATP PHOSPHORIBOSYLTRANSFERASE >pir||S55076 ATPphosphoribosyltransferase (EC 2.4.2.17) - fission yeast(Schizosaccharomyces pombe) >gb|AAA92790.1| (U07830) ATPphosphoribosyltransferase [Schizosaccharomyces pombe]>emb|CAA94634.1|(Z70691) atp phosphoribosyltransferase[Schizosaccharomyces pombe]
IV. Metabolism of Lipids, Fatty Acids, Sterols-See also fatty acid degradation					
1. Fatty acid biosynthesis (13)					
<acetyl-CoA carboxylase>					
q4d04fs.r1	612	3.1e-57	40	420	pir||T30568 acetyl-CoA carboxylase (EC 6.4.1.2) - Emericella nidulans>emb|CAA75926.1| (Y15996) acetyl-CoA carboxylase [Emericellanidulans]
f1d06fs.r1	582	5.1e-54	16	501	pir||T30568 acetyl-CoA carboxylase (EC 6.4.1.2) - Emericella nidulans>emb|CAA75926.1| (Y15996) acetyl-CoA carboxylase [Emericellanidulans]
Contig333	98	0.014	243	488	pir||T30568 acetyl-CoA carboxylase (EC 6.4.1.2) - Emericella nidulans>emb|CAA75926.1| (Y15996) acetyl-CoA carboxylase [Emericellanidulans]
<fatty acid synthase>					

dle12fs.r1	696	2.4e-66	25	495	gb|AAB41493.1| (U75347) fatty acid synthase, alpha subunit [Emericellanidulans]
m3d05fs.r1	671	1.1e-63	55	558	sp|P15368|FAS2_PENPA FATTY ACID SYNTHASE, SUBUNIT ALPHA [INCLUDES:EC1.1.1.100; EC 2.3.1.41] >pir||S01787 fatty-acid synthase (EC2.3.1.85) -Penicillium griseofulvum >gb|AAA33695.1| (M37461) FAS2protein [Penicillium griseofulvum]
k2c05fs.r1	615	9.8e-58	26	484	sp|P15368|FAS2_PENPA FATTY ACID SYNTHASE, SUBUNIT ALPHA [INCLUDES:EC1.1.1.100; EC 2.3.1.41] >pir||S01787 fatty-acid synthase (EC2.3.1.85) -Penicillium griseofulvum >gb|AAA33695.1| (M37461) FAS2protein [Penicillium griseofulvum]
Contig792	464	1.3e-41	242	691	sp|P15368|FAS2_PENPA FATTY ACID SYNTHASE, SUBUNIT ALPHA [INCLUDES:EC1.1.1.100; EC 2.3.1.41] >pir||S01787 fatty-acid synthase (EC2.3.1.85) - Penicillium griseofulvum >gb|AAA33695.1| (M37461) FAS2protein [Penicillium griseofulvum]
i1f01fs.r1	360	1.7e-30	12	431	gb|AAB41494.1| (U75347) fatty acid synthase, beta subunit [Emericellanidulans]
<stearoyl-CoA desaturase-adds double bonds to fatty acyl coA>					
Contig90	704	1.1e-68	3	470	pir||S52746 stearoyl-CoA desaturase (EC 1.14.99.5) - Ajellomyces capsulata(strain G217B) >emb|CAA59938.1| (X85962) delta-9 fatty aciddesaturase [Ajellomyces capsulatus]
Contig942	651	5.4e-63	172	726	pir||S52745 stearoyl-CoA desaturase (EC 1.14.99.5) - Ajellomyces capsulata(strain DOWNS) >emb|CAA59939.1| (X85963) delta-9 fatty aciddesaturase [Ajellomyces capsulatus]
Contig974	596	3.7e-57	445	876	pir||S52745 stearoyl-CoA desaturase (EC 1.14.99.5) - Ajellomyces capsulata(strain DOWNS) >emb|CAA59939.1| (X85963) delta-9 fatty aciddesaturase [Ajellomyces capsulatus]
<fatty acid (lipid) desaturase>					
c1f02fs.f1	120	4.7e-06	315	461	gi|4505193 ref|NP_003667.1|| membrane fatty acid (lipid) desaturase>gb|AAB62238.1| (AF002668) MLD [Homo sapiens]
<beta-ketoacyl reductase>					
Contig531	154	1.1e-09	1	234	gb|AAC62538.1| (AF052586) beta-ketoacyl reductase [Pseudomonas aeruginosa]
2. sterols (23)					
<C-3 sterol dehydrogenase>					
Contig488	158	3.5e-10	218	484	gi|6321437 ref|NP_011514.1|ERG26| C-3 sterol dehydrogenase; Erg26p>sp|P53199|YGA1_YEAST PUTATIVE 3BETA-HYDROXYSTEROIDDEHYDROGENASE/DELTA 5-->4-ISOMERASE (3BETA-HSD) [INCLUDES:3-BETA-HYDROXY-DELTA(5)-STEROID DEHYDROGENASE (3-BETA-

					HYDROXY-5-ENESTEROID DEHYDROGENASE) (PROGESTERONE REDUCTASE); STEROIDDELTA-ISOMERASE (DELTA-5-3-KETOSTEROI> >pir||S64003 hypotheticalprotein YGL001c - yeast (Saccharomyces
<sterol C-14 reductase>					
c1a11fs.f1	105	0.0015	388	465	gi|6324049 ref|NP_014119.1|ERG24| sterol C-14 reductase; Erg24p>sp|P32462|ER24_YEAST C-14 STEROL REDUCTASE (STEROL C14-REDUCTASE)>pir||S30769 probable C-14 sterol reductase (EC 1.1.-.-) - yeast(Saccharomyces cerevisiae) >gb|AAA18256.1| (M99419) C-14 sterolreductase [Saccharomyces cerevisiae] >emb|CAA96192.1| (Z71556) ORFYNL280c [Saccharomyces cerevisiae]
<sterol reductase>					
m1g01fs.r1	497	1.1e-46	31	447	sp|P36209|STS1_SCHPO C-24(28) STEROL REDUCTASE >pir||T38121 C-24(28) sterolreductase - fission yeast (Schizosaccharomyces pombe)>emb|CAB11256.1| (Z98600) c-24(28) sterol reductase (EC 1.-.-.-) [Schizosaccharomyces pombe]
<delta-12 desaturase>					
h3c02fs.r1	385	8.2e-35	112	528	gb|AAD55982.1| (AF161219) delta-12 desaturase [Mucor rouxii]
<delta7-sterol-C5-desaturase>					
d3d10fs.r1	220	2.2e-17	252	509	gb|AAF00544.1|AF1879 (AF187981) delta7-sterol-C5-desaturase [Homo sapiens]
2.1. General					
<sterol c-methyltransferase>					
Contig663	622	6.3e-60	146	769	sp|O74198|ERG6_CANAL DELTA(24)-STEROL C-METHYLTRANSFERASE >gb|AAC26626.1|(AF031941) sterol transmethylase [Candida albicans]
<HYDROXY-3-METHYLGLUTARYL-COENZYME A REDUCTASE-also mevalonate biosyn to isoprenoids>					
Contig817	1146	2.7e-128	9	707	sp|Q12577|HMDH_GIBFU 3-HYDROXY-3-METHYLGLUTARYL-COENZYME A REDUCTASE (HMG-COAREDUCTASE) >emb|CAA63970.1| (X94307) HMG-CoA-reductase [Gibberellafujikuroi]
Contig859	279	4e-30	363	530	sp|Q12577|HMDH_GIBFU 3-HYDROXY-3-METHYLGLUTARYL-COENZYME A REDUCTASE (HMG-COAREDUCTASE) >emb|CAA63970.1| (X94307) HMG-CoA-reductase [Gibberellafujikuroi]
2.2. Farnesol biosynthesis					
<GERANYLGERANYL PYROPHOSPHATE SYNTHETASE>					
k3h08fs.f1	163	1.2e-10	173	484	sp|P32434|CWG2_SCHPO TYPE I PROTEIN GERANYLGERANYLTRANSFERASE BETA SUBUNIT(TYPE I PROTEIN GERANYL-GERANYLTRANSFERASE BETA SUBUNIT)(GGTASE-I-BETA) (PGGT) >pir||S41686 geranylgeranyltransferase typeI (EC 2.5.1.-) beta chain - fission yeast (Schizosaccharomycespombe)>emb|CAA78143.1| (Z12155)

					dimethylallyltransferase[Schizosaccharomyces pombe] >emb|CAB86347.1| (AL163071) type iprotein geranylgeranyltransferase
<hydroxysteroid dehydrogenase>					
b1g12fs.r1	177	2.8e-12	87	431	pir||T40392 probable 3-beta-hydroxysteroid dehydrogenase / delta 5-4-isomerase- fission yeast (Schizosaccharomyces pombe)>emb|CAA17691.1|(AL022019) putative 3 beta-hydroxysteroid dehydrogenase/delta5-->4-isomerase(3beta-hsd) [Schizosaccharomyces pombe]
<FARNESYL PYROPHOSPHATE SYNTHETASE>					
Contig834	925	4.3e-92	61	621	sp|Q92235|FPPS_GIBFU FARNESYL PYROPHOSPHATE SYNTHETASE (FPP SYNTHETASE) (FPS)(FARNESYL DIPHOSPHATE SYNTHETASE) [INCLUDES:DIMETHYLALLYLTRANSFERASE ;GERANYLTRANSTRANSFERASE] >pir||S71435farnesyl-pyrophosphate synthetase -fungus (Gibberella fujikuroi)>emb|CAA65641.1| (X96940) Farnesylpyrophosphate Synthetase[Gibberella fujikuroi]
Contig836	766	3.1e-75	200	664	sp|Q92235|FPPS_GIBFU FARNESYL PYROPHOSPHATE SYNTHETASE (FPP SYNTHETASE) (FPS)(FARNESYL DIPHOSPHATE SYNTHETASE) [INCLUDES:DIMETHYLALLYLTRANSFERASE ; GERANYLTRANSTRANSFERASE] >pir||S71435farnesyl-pyrophosphate synthetase - fungus (Gibberella fujikuroi)>emb|CAA65641.1| (X96940) Farnesylpyrophosphate Synthetase[Gibberella fujikuroi]
2.3. cholesterol and steroids					
<C-4 METHYL STEROL OXIDASE-cholesterol biosynthesis>					
j4h02fs.r1	355	1.2e-31	197	511	pir||T38986 probable c-4 methyl sterol oxidase - fission yeast(Schizosaccharomyces pombe) >emb|CAB52730.1| (AL109832) putativec-4 methyl sterol oxidase [Schizosaccharomyces pombe]
h3b10fs.f1	174	4.1e-12	379	507	gi|6321497 ref|NP_011574.1|ERG25| C-4 sterol methyl oxidase; Erg25p>sp|P53045|ER25_YEAST C-4 METHYL STEROL OXIDASE >pir||S64354 ERG25protein - yeast (Saccharomyces cerevisiae) >gb|AAC49139.1| (U31885)C-4 sterol methyl oxidase [Saccharomyces cerevisiae]>emb|CAA97062.1| (Z72845) ORF YGR060w [Saccharomyces cerevisiae]
<isopentenyl-diphosphate Delta-isomerase>					
Contig991	757	3.3e-74	159	797	sp|Q10132|IDI1_SCHPO ISOPENTENYL-DIPHOSPHATE DELTA-ISOMERASE (IPP ISOMERASE)(ISOPENTENYL PYROPHOSPHATE ISOMERASE)>pir||A56442isopentenyl-diphosphate Delta-isomerase (EC 5.3.3.2) - fissionyeast (Schizosaccharomyces pombe) >pir||T37986isopentenyl-diphosphate delta-isomerase - fission

					yeast(Schizosaccharomyces pombe) >pir||T39272 isopentenyl-diphosphatedelta-isomerase - fission yeast (Schizosaccharomyces
Contig18	317	1.4e-27	272	526	sp|Q10132|IDI1_SCHPO ISOPENTENYL-DIPHOSPHATE DELTA-ISOMERASE (IPP ISOMERASE)(ISOPENTENYL PYROPHOSPHATE ISOMERASE)>pir||A56442isopentenyl-diphosphate Delta-isomerase (EC 5.3.3.2) - fissionyeast (Schizosaccharomyces pombe) >pir||T37986isopentenyl-diphosphate delta-isomerase - fission yeast(Schizosaccharomyces pombe) >pir||T39272 isopentenyl-diphosphatedelta-isomerase - fission yeast (Schizosaccharomyces
<ergosterol synthesis>					
d3d08fs.f1	119	4.5e-05	243	383	pir||T40135 probable involvement in ergosterol synthesis - fission yeast(Schizosaccharomyces pombe) >emb|CAB10154.2| (Z97211) probableinvolvement in ergosterol synthesis [Schizosaccharomyces pombe]
<ergosterol biosynthesis>					
l1h03fs.r1	320	1.6e-27	241	534	pir||T40584 probable involvement in ergosterol biosynthesis - fission yeast(Schizosaccharomyces pombe) >emb|CAA22812.1| (AL035216) possibleinvolvement in ergosterol biosynthesis [Schizosaccharomyces pombe]
<phosphomevalonate kinase>					
a1e02fs.r1	187	4.2e-13	103	408	gb|AAA34596.1| (M63648) phosphomevalonate kinase [Saccharomyces cerevisiae]
<MEVALONATE KINASE>					
Contig124	177	9.3e-13	338	577	dbj|BAA24409.1| (D78165) mevalonate kinase [Saccharomyces cerevisiae]
Contig906	175	8.8e-12	406	726	gi|6323864 ref|NP_013935.1|ERG12| mevalonate kinase; Erg12p>sp|P07277|KIME_YEAST MEVALONATE KINASE (MVK) >pir||BVBYR1mevalonate kinase (EC 2.7.1.36) - yeast (Saccharomyces cerevisiae)>emb|CAA39359.1| (X55875) mevalonate kinase [Saccharomycescerevisiae] >emb|CAA29487.1| (X06114) ORF1 (put. RAR1 protein) (AA1-443) [Saccharomyces cerevisiae]>emb|CAA89923.1| (Z49809) Rar1p[Saccharomyces cerevisiae]
<diphosphomevalonate decarboxylase>					
c4f09fs.r1	268	1.9e-22	107	445	pir||T38344 diphosphomevalonate decarboxylase - fission yeast(Schizosaccharomyces pombe) >emb|CAB11260.1| (Z98601)diphosphomevalonate decarboxylase [Schizosaccharomyces pombe]

c4f09fs.f1	151	2.9e-09	200	433	pir||T38344 diphosphomevalonate decarboxylase - fission yeast(Schizosaccharomyces pombe) >emb|CAB11260.1| (Z98601)diphosphomevalonate decarboxylase [Schizosaccharomyces pombe]

3. lipids

3.1. phospholipid metabolism (18)

<phospholipid methyltransferase>

b1c12fs.r1	226	5.7e-18	192	422	gb|AAB61409.1| (AF004112) phospholipid methyltransferase [Schizosaccharomycespombe]

<phosphoinositide-specific phospholipase>

f1e03fs.r1	513	1e-47	21	491	gb|AAC72385.1| (AF098645) phosphoinositide-specific phospholipase C[Pyricularia grisea]

<lysophospholipase>

Contig108	773	6.3e-76	7	645	gb|AAC03052.1| (AF045574) lysophospholipase [Neurospora crassa]
Contig908	615	3.8e-59	99	746	gb|AAC03053.1| (AF045575) lysophospholipase [Neurospora crassa]
Contig955	389	9.9e-35	212	568	gb|AAC03052.1| (AF045574) lysophospholipase [Neurospora crassa]

<phosphatidyl synthase>

Contig481	185	3.3e-12	537	809	pir||T38148 phosphatidyl synthase - fission yeast (Schizosaccharomyces pombe)>emb|CAB16578.1| (Z99295) CDP-alcohol phosphatidyltransferase[Schizosaccharomyces pombe]

<phosphatidylserine decarboxylase>

c3f02fs.f1	321	8.9e-27	19	474	pir||T38632 probable phosphatidylserine decarboxylase proenzyme - fissionyeast (Schizosaccharomyces pombe) >emb|CAB11699.1| (Z98979)phosphatidylserine decarboxylase proenzyme 2 precursor[Schizosaccharomyces pombe]
s1b01fs.f1	314	4.8e-26	3	428	pir||T38632 probable phosphatidylserine decarboxylase proenzyme - fissionyeast (Schizosaccharomyces pombe) >emb|CAB11699.1| (Z98979)phosphatidylserine decarboxylase proenzyme 2 precursor[Schizosaccharomyces pombe]
m4f07fs.f1	248	2.9e-20	100	492	pir||T39865 phosphatidylserine decarboxylase - fission yeast(Schizosaccharomyces pombe) >emb|CAA16834.1| (AL021746)phosphatidylserine decarboxylase proenzyme 1 precursor[Schizosaccharomyces pombe]

<myo-inositol phosphate synthase-biosynthesis of inositol containing phospholipids>

Contig953	326	2.9e-28	220	537	pir||T01647 myo-inositol-1-phosphate synthase (EC 5.5.1.4) - maize>gb|AAC15756.1| (AF056326) myo-inositol 1-phosphate synthase; INO1[Zea mays]

<inositol 1-phosphate synthase-first enzyme in inositol biosynthetic pathway>

Contig969	1169	7.1e-118	486	1631	gb|AAC33791.1| (AF078915) inositol 1-phosphate synthase [Pichia pastoris]
<1-phosphatidylinositol phosphodiesterase>					
Contig243	103	0.022	389	466	pir||S54403 1-phosphatidylinositol phosphodiesterase (EC 3.1.4.10) Listeriaivanovii >emb|CAA51230.1| (X72685) 1-phosphatidylinositolphosphodiesterase [Listeria ivanovii]
<CDP-diacylglycerol synthase>					
Contig427	423	6.9e-39	8	478	gi|6319503 ref|NP_009585.1|CDS1| CDP-diacylglycerol synthase, CTP-phosphatidicacid cytidylyltransferase, CDP-diglyceride synthetase; Cds1p>sp|P38221|CDS1_YEAST PHOSPHATIDATE CYTIDYLYLTRANSFERASE(CDP-DIGLYCERIDE SYNTHETASE) (CDP-DIGLYCERIDE PYROPHOSPHORYLASE)(CDP-DIACYLGLYCEROL SYNTHASE) (CDS) (CTP:PHOSPHATIDATECYTIDYLYLTRANSFERASE) (CDP-DAG SYNTHASE) >pir||S45885 probablemembrane protein YBR029c -
<annexin V-binding protein-calcium-dependent phospholipid binding proteins>					
b2a07fs.f1	456	2.5e-42	31	420	dbj|BAA20855.1| (D64062) annexin V-binding protein (ABP-10) [Rattusnorvegicus]
3.2. sphingolipids					
<serine palmitoyltransferase>					
d4d02fs.r1	406	5e-37	9	467	pir||T40091 probable serine palmitoyltransferase - fission yeast(Schizosaccharomyces pombe) >pir||T39753 serinepalmitoyltransferase subunit - fission yeast (Schizosaccharomycespombe) >emb|CAA18397.1| (AL022299) putative serinepalmitoyltransferase [Schizosaccharomyces pombe]>emb|CAA22662.1|(AL035077) serine palmitoyltransferase subunit [Schizosaccharomycespombe]
<sphingosine phosphate lyase>					
k3d11fs.r1	392	2.6e-35	24	458	gi|6320500 ref|NP_010580.1|DPL1| dihydrosphingosine phosphate lyase (alsoknown as sphingosine phosphate lyase); Dpl1p >pir||S70123 probablemembrane protein YDR294c - yeast (Saccharomyces cerevisiae)>gb|AAB64470.1| (U51031) Ydr294cp [Saccharomyces cerevisiae]
3.3. lipopolysaccharide biosyn-biomembrane precursors					
<UDP-glucose:sterol glucosyltransferase>					
n1e06fs.r1	372	6.7e-33	70	486	emb|CAB06081.1| (Z83832) UDP-glucose:sterol glucosyltransferase [Avena sativa]
b4g05fs.r1	151	1.5e-08	20	154	gb|AAD29570.1|AF0913 (AF091397) UDP-glucose:sterol glucosyltransferase[Pichia pastoris]

V. Sulfur, Phosphate and Nitrogen Metabolism

1. Sulfur Metabolism (7)

<sulphur metabolite repression>

o3g03fs.r1	503	2.5e-47	128	505	gb|AAB18274.2| (U75874) sconCp [Emericella nidulans]
o3g03fs.f1	152	4.2e-10	423	512	gb|AAB18274.2| (U75874) sconCp [Emericella nidulans]

<sulfate adenylyltransferase-leads to biosynthesis of cys&met>

i1b09fs.r1	644	2.8e-62	12	446	sp|Q12555|MET3_EMENI SULFATE ADENYLYLTRANSFERASE (SULFATE ADENYLATETRANSFERASE) (SAT) (ATP-SULFURYLASE) >pir||S55034 probable3'-phosphoadenosine-5'-phosphosulfate synthetase - Emericellanidulans >emb|CAA57891.1| (X82541) sulfate adenylyltransferase[Emericella nidulans]
i4f06fs.r1	385	1.2e-34	181	489	sp|Q12555|MET3_EMENI SULFATE ADENYLYLTRANSFERASE (SULFATE ADENYLATETRANSFERASE) (SAT) (ATP-SULFURYLASE) >pir||S55034 probable3'-phosphoadenosine-5'-phosphosulfate synthetase - Emericellanidulans >emb|CAA57891.1| (X82541) sulfate adenylyltransferase[Emericella nidulans]
Contig728	311	2.5e-26	305	541	sp|P56862|MET3_ASPTE SULFATE ADENYLYLTRANSFERASE (SULFATE ADENYLATETRANSFERASE) (SAT) (ATP-SULFURYLASE) >gb|AAF28890.1|AF123267_2(AF123267) sulfate adenylyltransferase [Aspergillus terreus]

<sulfite reductase>

t4f06fs.f1	286	7.6e-23	327	572	gi|6322597 ref|NP_012671.1|ECM17| Putative sulfite reductase; Ecm17p>sp|P47169|YJ9F_YEAST HYPOTHETICAL 161.2 KD PROTEIN IN NMD5-HOM6INTERGENIC REGION >pir||S57160 sulfite reductase homolog YJR137c-yeast (Saccharomyces cerevisiae) >emb|CAA89669.1| (Z49637) ORFYJR137c [Saccharomyces cerevisiae]
Contig152	133	3.9e-07	302	523	pdb|1SOX|A Chain A, Sulfite Oxidase From Chicken Liver >pdb|1SOX|B Chain B,Sulfite Oxidase From Chicken Liver

2. Nitrogen Metabolism (see also amino acid metabolism) (16)

<nitrite reductase>

e3b09fs.r1	498	1.4e-45	22	420	sp|P38681|NIR_NEUCR NITRITE REDUCTASE [NAD(P)H] >pir||A49848 nitrite reductase(NADH) (EC 1.6.6.4) - Neurospora crassa
e3b09fs.f1	297	4e-24	191	526	sp|P38681|NIR_NEUCR NITRITE REDUCTASE [NAD(P)H] >pir||A49848 nitrite reductase(NADH) (EC 1.6.6.4) - Neurospora crassa

<nitrogen regulation protein>

s3c09fs.r1	553	9.2e-52	11	568	sp|P78688|AREA_GIBFU NITROGEN REGULATION PROTEIN AREA >emb|CAA71897.1|(Y11006) AREA [Gibberella fujikuroi]
<spermidine synthase>					
Contig112	1040	2.6e-104	78	776	dbj|BAA81738.1| (AB001598) spermidine synthase [Neurospora crassa]
<4-nitrophenylphosphatase>					
Contig391	220	2.7e-17	266	502	emb|CAB56701.1| (AJ249796) NIPSNAP1 protein [Danio rerio]
<MeaB protein-affecting nitrogen metabolite repression>					
b1f03fs.r1	137	8.4e-08	244	432	emb|CAA66668.1| (X98065) MeaB protein [Emericella nidulans]
<DRAP deaminase>					
Contig189	283	3.6e-23	181	663	gi|6324506 ref|NP_014575.1|RIB2| DRAP deaminase; Rib2p >sp|Q12362|RIB2_YEASTDRAP DEAMINASE >pir||S50972 RIB2 protein - yeast (Saccharomycescerevisiae) >emb|CAA79742.1| (Z21618) DRAP deaminase [Saccharomycescerevisiae] >emb|CAA99076.1| (Z74808) ORF YOL066c [Saccharomycescerevisiae]
<2-methyl-3-hydroxypyridine-5-carboxylic acidoxygenase>					
a2g12fs.f1	96	0.025	113	370	gb|AAB60878.1| (AF001965) 2-methyl-3-hydroxypyridine-5-carboxylic acidoxygenase [Pseudomonas sp. MA-1]
-urea related					
<urease>					
j2b08fs.r1	485	1.2e-44	52	459	gb|AAC49868.1| (U81509) urease [Coccidioides immitis]
j2b08fs.f1	318	1.4e-26	161	484	gb|AAC49868.1| (U81509) urease [Coccidioides immitis]
<urea amidolyase>					
b1c09fs.r1	168	4.3e-10	22	435	prf||2009329A urea amidolyase [Pichia jadinii]
<URICASE>					
Contig469	441	9.8e-41	136	471	sp|Q00511|URIC_ASPFL URICASE (URATE OXIDASE) >pir||A38097 urate oxidase (EC1.7.3.3) - Aspergillus flavus >emb|CAA43895.1| (X61765) urateoxidase [Aspergillus flavus] >emb|CAA43896.1| (X61766) urateoxidase [Aspergillus flavus]
b4f06fs.f1	323	3.4e-28	165	416	sp|Q00511|URIC_ASPFL URICASE (URATE OXIDASE) >pir||A38097 urate oxidase (EC1.7.3.3) - Aspergillus flavus >emb|CAA43895.1| (X61765) urateoxidase [Aspergillus flavus] >emb|CAA43896.1| (X61766) urateoxidase [Aspergillus flavus]
<UREA ACTIVE TRANSPORTER>					
a2h11fs.r1	419	6e-38	22	459	pir||T39959 probable urea active transporter - fission yeast(Schizosaccharomyces pombe) >emb|CAA22629.1| (AL035065) putativeurea active transporter [Schizosaccharomyces pombe]

a2h11fs.f1 268 1.9e-21 110 406 pir||T39959 probable urea active transporter - fission yeast(Schizosaccharomyces pombe) >emb|CAA22629.1| (AL035065) putativeurea active transporter [Schizosaccharomyces pombe]

<ureidoglycolate hydrolase>

n3h12fs.r1 143 4e-09 236 475 pir||T37991 probable ureidoglycolate hydrolase - fission yeast(Schizosaccharomyces pombe) >emb|CAB10115.1| (Z97209) putativeureidoglycolate hydrolase [Schizosaccharomyces pombe]

VI. Metabolism of Cofactors, prosthetic groups

1.nicotinamide coenzymes (5)

<NICOTINATE-NUCLEOTIDE PYROPHOSPHORYLASE-DE NOVO BIOSYNTHESIS OF NAD AND NADP>

d4b04fs.r1 467 1.8e-43 9 521 sp|Q15274|NADC_HUMAN NICOTINATE-NUCLEOTIDE PYROPHOSPHORYLASE [CARBOXYLATING](QUINOLINATE PHOSPHORIBOSYLTRANSFERASE [DECARBOXYLATING])(QAPRTASE) >dbj|BAA11242.1| (D78177) quinolinate phosphoribosyltransferase [Homo sapiens]

<kynureninase-biosyn of NAD cofactors>

Contig263 178 5e-12 108 455 gi|6323261 ref|NP_013332.1|YLR231C| Ylr231cp >sp|Q05979|KYNU_YEAST PROBABLEKYNURENINASE (L-KYNURENINE HYDROLASE)>pir||S51453 probablemembrane protein YLR231c - yeast (Saccharomyces cerevisiae)>gb|AAB67417.1| (U19027) Weak similarity to kynureninase (rat, PIRaccession number PS0370) in small region of central portion ofprotein. [Saccharomyces cerevisiae]

<nicotinate phosphoribosyltransferase>

g4f07fs.r1 240 5.4e-19 17 295 emb|CAB62416.1| (AL133357) putative nicotinate phosphoribosyltransferase[Schizosaccharomyces pombe]

o2a12fs.r1 166 8.5e-11 307 504 emb|CAA85352.1| (Z36878) putative nicotinate phosphorybosyltranferase[Saccharomyces cerevisiae]

<sir2-related protein type 7-metabolizes NAD>

m3a05fs.r1 721 1.9e-70 50 547 gb|AAF43431.1|AF2333 (AF233395) sir2-related protein type 7 [Homo sapiens]

2.biocytin (biotin) (2)

<BIOTIN SYNTHASE>

Contig175 356 9.9e-32 104 418 gi|6321725 ref|NP_011802.1|BIO2| Biotin synthase; Bio2p>sp|P32451|BIOB_YEASTBIOTIN SYNTHASE (BIOTIN SYNTHETASE) >pir||S64621 biotin synthetase- yeast (Saccharomyces cerevisiae) >emb|CAA97318.1| (Z73071)ORFYGR286c [Saccharomyces cerevisiae]

Contig174	265	4.3e-22	230	619	gi|6321725 ref|NP_011802.1|BIO2| Biotin synthase; Bio2p >sp|P32451|BIOB_YEASTBIOTIN SYNTHASE (BIOTIN SYNTHETASE) >pir||S64621 biotin synthetase- yeast (Saccharomyces cerevisiae) >emb|CAA97318.1| (Z73071)ORFYGR286c [Saccharomyces cerevisiae]
3.thiamine (3)					
<thiamine-4>					
p1h07fs.r1	489	8.4e-46	4	507	dbj|BAA21049.1| (D45894) thiamine-4 [Neurospora crassa]
p1h07fs.f1	338	1e-29	218	517	dbj|BAA21049.1| (D45894) thiamine-4 [Neurospora crassa]
<THIAMINE BIOSYNTHETIC BIFUNCTIONAL ENZYME>					
Contig514	252	5.5e-20	147	659	sp|P40386|THI4_SCHPO PROBABLE THIAMINE BIOSYNTHETIC BIFUNCTIONAL ENZYME[INCLUDES: THIAMINE-PHOSPHATE PYROPHOSPHORYLASE (TMPPYROPHOSPHORYLASE) (TMP-PPASE); HYDROXYETHYLTHIAZOLE KINASE(4-METHYL-5-BETA-HYDROXYETHYLTHIAZOLE KINASE) (THZ KINASE) (THKINASE)]>pir||S44183 thiamin-phosphate pyrophosphorylase (EC2.5.1.3) /hydroxyethylthiazole kinase (EC 2.7.1.50) thi4 - fissionyeast (Schizosaccharomyces pombe)
4.coenzyme A (3)					
<acetyl-coenzyme A synthetase>					
o2e01fs.f1	726	6.3e-71	13	519	sp|P16928|ACSA_EMENI ACETYL-COENZYME A SYNTHETASE (ACETATE—COA LIGASE)(ACYL-ACTIVATING ENZYME) >pir||SYASAA acetate--CoA ligase (EC6.2.1.1)- Emericella nidulans >emb|CAA34858.1| (X16990)acetate--CoA ligase [Emericella nidulans]
n1h05fs.r1	545	8.4e-68	13	369	sp|P16928|ACSA_EMENI ACETYL-COENZYME A SYNTHETASE (ACETATE—COA LIGASE)(ACYL-ACTIVATING ENZYME) >pir||SYASAA acetate--CoA ligase (EC6.2.1.1)- Emericella nidulans >emb|CAA34858.1| (X16990)acetate--CoA ligase [Emericella nidulans]
o2e01fs.r1	440	1.1e-40	125	505	sp|P16929|ACSA_NEUCR ACETYL-COENZYME A SYNTHETASE (ACETATE—COA LIGASE)(ACYL-ACTIVATING ENZYME) >pir||SYNCAA acetate--CoA ligase (EC6.2.1.1) - Neurospora crassa >emb|CAA34857.1| (X16989) acetate--CoAligase [Neurospora crassa]
5.flavins (1)					
<GTP cyclohydrolase>					
o2f03fs.r1	706	8.4e-69	14	517	emb|CAB65619.1| (AL136078) putative GTP cyclohydrolase [Schizosaccharomycespombe]
7. heme (14)					
<heme protein precursor>					

Contig643	504	2e-47	75	539	sp|P07142|CY1_NEUCR CYTOCHROME C1, HEME PROTEIN PRECURSOR >pir||A27187ubiquinol--cytochrome-c reductase (EC 1.10.2.2) cytochrome c1precursor - Neurospora crassa >emb|CAA28860.1| (X05235) cytochromec1 precursor [Neurospora crassa]
Contig763	481	5.2e-45	193	531	sp|P07142|CY1_NEUCR CYTOCHROME C1, HEME PROTEIN PRECURSOR >pir||A27187ubiquinol--cytochrome-c reductase (EC 1.10.2.2) cytochrome c1precursor - Neurospora crassa >emb|CAA28860.1| (X05235) cytochromec1 precursor [Neurospora crassa]
<siroheme synthase>					
s2a11fs.f1	149	9.6e-09	251	454	gi|6322922 ref|NP_012995.1|MET1| siroheme synthase; Met1p>sp|P36150|SUMT_YEAST PROBABLE UROPORPHYRIN-III-METHYLTRANSFERASE(UROGEN III METHYLASE) (SUMT) (UROPORPHYRINOGEN III METHYLASE)(UROM) >pir||S38145 uroporphyrinogen methyltransferase homologYKR069w - yeast (Saccharomyces cerevisiae) >emb|CAA82148.1|(Z28294)ORF YKR069w [Saccharomyces cerevisiae]
-iron uptake					
<ferric reductase>					
s2h11fs.r1	172	5.3e-11	139	513	emb|CAB45649.1| (AJ387722) ferric reductase [Candida albicans]
b1h10fs.r1	147	2.4e-08	31	417	emb|CAB45649.1| (AJ387722) ferric reductase [Candida albicans]
b1g01fs.r1	110	0.00018	6	386	gi|6323243 ref|NP_013315.1|FRE1| Ferric (and cupric) reductase;Fre1p>sp|P32791|FRE1_YEAST FERRIC REDUCTASE TRANSMEMBRANE COMPONENT 1PRECURSOR >pir||S30075 ferric reductase (EC 1.6.99.-) FRE1 - yeast(Saccharomyces cerevisiae) >gb|AAA34608.1| (M86908) ferricreductase [Saccharomyces cerevisiae] >gb|AAB67424.1| (U14913)Fre1p: Ferric (and cupric) reductase [Saccharomyces cerevisiae]
<delta-aminolevulinic acid dehydratase-heme biosynthesis>					
e1f01fs.r1	161	1.7e-10	315	476	gi|6321398 ref|NP_011475.1|HEM2| delta-aminolevulinate dehydratase(porphobilinogen synthase); Hem2p >sp|P05373|HEM2_YEASTDELTA-AMINOLEVULINIC ACID DEHYDRATASE (PORPHOBILINOGEN SYNTHASE)(ALADH) >pir||S64042 porphobilinogen synthase (EC 4.2.1.24) - yeast(Saccharomyces cerevisiae) >emb|CAA96742.1| (Z72562) ORF YGL040c[Saccharomyces cerevisiae]
<flavohemoglobin>					
i3g09fs.r1	319	7.6e-28	18	461	dbj|BAA33011.1| (AB016807) flavohemoglobin [Fusarium oxysporum]<protoporphyrinogen oxidase>
<Iron-sulfur cluster nifU-like protein>					

j3a08fs.r1	417	3.1e-38	225	545	gi|6324800 ref|NP_014869.1|ISU2| Iron-sulfur cluster nifU-like protein; Isu2p>pir||S60953 nifU protein homolog YOR226c - yeast (Saccharomycescerevisiae) >emb|CAA63189.1| (X92441) YOR50-16 [Saccharomycescerevisiae] >emb|CAA99445.1| (Z75133) ORF YOR226c [Saccharomycescerevisiae]
<AMINOLEVULINIC ACID DEHYDRATASE-heme biosynthesis>					
m1g08fs.r1	612	8.1e-59	24	446	gb|AAD38391.1|AF1523 (AF152374) 5-aminolevulinic acid synthase[Aspergillus oryzae]
Contig636	431	1.1e-39	241	597	sp|O42768|HEM2_CANGA DELTA-AMINOLEVULINIC ACID DEHYDRATASE (PORPHOBILINOGENSYNTHASE) (ALADH) >gb|AAB94926.1| (AF038566)porphobilinogensynthase [Candida glabrata]
b2g11fs.r1	133	5.7e-07	219	374	sp|P38092|HEM1_EMENI 5-AMINOLEVULINIC ACID SYNTHASE, MITOCHONDRIAL PRECURSOR(DELTA-AMINOLEVULINATE SYNTHASE) (DELTA-ALA SYNTHETASE)>pir||S31846 5-aminolevulinate synthase (EC 2.3.1.37) precursor,mitochondrial - Emericella nidulans >emb|CAA45508.1| (X64170)5-aminolevulinic acid synthase [Emericella nidulans]
<siderophore biosynthesis protein pbsC>					
d4c05fs.r1	136	3.9e-07	59	409	pir||S45582 siderophore biosynthesis protein pbsC - Pseudomonas sp. (strainM114) >emb|CAA54778.1| (X77699) biosynthetic protein C [Pseudomonassp.]
8.Molybdopterin (1)					
<MOLYBDOPTERIN BIOSYNTHESIS CNX1 PROTEIN>					
k4f02fs.r1	151	7.4e-09	7	225	sp|Q39054|CNX1_ARATH MOLYBDOPTERIN BIOSYNTHESIS CNX1 PROTEIN (MOLYBDENUMCOFACTOR BIOSYNTHESIS ENZYME CNX1) >gb|AAA97413.1| (L47323)molybdenum cofactor biosynthesis enzyme [Arabidopsis thaliana]>emb|CAB38312.1| (AJ236870) molybdenum cofactor biosynthesis enzyme[Arabidopsis thaliana]
9. PMP (3)					
<PYRIDOXAMINE 5'-PHOSPHATE OXIDASE>					
e3a04fs.f1	345	1.4e-30	133	519	gi|6319509 ref|NP_009591.1|PDX3| pyridoxine (pyridoxiamine) phosphate oxidase;Pdx3p >sp|P38075|PDX3_YEAST PYRIDOXAMINE 5'-PHOSPHATE OXIDASE(PNP/PMP OXIDASE) >pir||S41301 pyridoxamine-phosphate oxidase (EC1.4.3.5) - yeast (Saccharomyces cerevisiae) >pdb|1CI0|A Chain A,Pnp Oxidase From Saccharomyces Cerevisiae >pdb|1CI0|B Chain B, PnpOxidase From Saccharomyces Cerevisiae >emb|CAA54295.1|

e3a04fs.r1	321	5.3e-28	146 508	emb|CAB60247.1| (AL132839) probable pyridoxamine 5'-phosphate oxidase[Schizosaccharomyces pombe]
o4e08fs.r1	250	1.8e-20	153 503	sp|P44909|PDXH_HAEIN PYRIDOXAMINE 5'-PHOSPHATE OXIDASE (PNP/PMP OXIDASE)>pir||H64098 probable pyridoxamine-phosphate oxidase (EC 1.4.3.5)HI0863 - Haemophilus influenzae (strain Rd KW20)>gb|AAC22522.1|(U32768) pyridoxamine phosphate oxidase (pdxH) [Haemophilusinfluenzae Rd]

##10. others (1)

<cofactor required for Sp1 transcriptionalactivation subunit 2>

o4d07fs.f1	130	8.5e-07	102 224	gi|6753526 ref|NP_036135.1|| cofactor required for Sp1 transcriptionalactivation subunit 2 (150 kDa) >dbj|BAA76611.1| (AB019029)similarto human EXLM1 gene [Mus musculus]

VII. Energy

VII.1. Carbohydrate as energy source

1. Glycolysis (27)

<hexokinase>

m4g07fs.r1	537	6.4e-51	68 481	emb|CAA08922.1| (AJ009973) hexokinase [Aspergillus niger]
Contig254	287	6.2e-24	261 518	emb|CAA08922.1| (AJ009973) hexokinase [Aspergillus niger]
b2g09fs.r1	92	0.005	154 303	gi|4557693 ref|NP_000212.1|| ketohexokinase, isoform a >sp|P50053|KHK_HUMANKETOHEXOKINASE (HEPATIC FRUCTOKINASE) >emb|CAA55347.1| (X78678)ketohexokinase [Homo sapiens]

<GLUCOKINASE>

f3d01fs.r1	538	5.5e-51	4 450	sp|Q92407|HXKG_ASPNG GLUCOKINASE (GLUCOSE KINASE) (GLK)>pir||S74210glucokinase (EC 2.7.1.2) - Aspergillus niger>emb|CAA67949.1|(X99626) glucokinase [Aspergillus niger]
g2e10fs.r1	298	3.8e-25	185 445	sp|Q92407|HXKG_ASPNG GLUCOKINASE (GLUCOSE KINASE) (GLK)>pir||S74210glucokinase (EC 2.7.1.2) - Aspergillus niger>emb|CAA67949.1|(X99626) glucokinase [Aspergillus niger]
i4c03fs.r1	199	4.7e-15	246 431	sp|Q10242|GNTK_SCHPO PROBABLE GLUCONOKINASE (GLUCONATE KINASE)>pir||T38871probable glucokinase - fission yeast (Schizosaccharomyces pombe)>emb|CAA93562.1| (Z69727) probable glucokinase [Schizosaccharomycespombe]
Contig322	128	1.4e-06	467 610	sp|Q92407|HXKG_ASPNG GLUCOKINASE (GLUCOSE KINASE) (GLK)>pir||S74210glucokinase (EC 2.7.1.2) - Aspergillus niger>emb|CAA67949.1|(X99626) glucokinase [Aspergillus niger]

<6-PHOSPHOFRUCTOKINASE>					
c4e09fs.r1	490	2.4e-45	12	428	sp|P78985|K6PF_ASPNG 6-PHOSPHOFRUCTOKINASE (PHOSPHOFRUCTOKINASE)(PHOSPHOHEXOKINASE) >emb|CAB01923.1| (Z79690)phosphofructokinase[Aspergillus niger]
s1d12fs.r1	402	1.6e-35	261	530	pir||T39624 6-phosphofructokinase beta subunit - fission yeast(Schizosaccharomyces pombe) >emb|CAA17900.1| (AL022104)6-phosphofructokinase beta subunit [Schizosaccharomyces pombe]
Contig435	300	1e-24	226	534	sp|P78985|K6PF_ASPNG 6-PHOSPHOFRUCTOKINASE (PHOSPHOFRUCTOKINASE)(PHOSPHOHEXOKINASE) >emb|CAB01923.1| (Z79690)phosphofructokinase[Aspergillus niger]
<glucose-6-phosphate isomerase>					
Contig853	755	4.4e-74	8	610	gi|6319673 ref|NP_009755.1|PGI1| Glucose-6-phosphate isomerase;Pgi1p>sp|P12709|G6PI_YEAST GLUCOSE-6-PHOSPHATE ISOMERASE (GPI)(PHOSPHOGLUCOSE ISOMERASE) (PGI) (PHOSPHOHEXOSE ISOMERASE) (PHI)>pir||NUBY glucose-6-phosphate isomerase (EC 5.3.1.9) - yeast(Saccharomyces cerevisiae) >emb|CAA32158.1| (X13977)phosphoglucoseisomerase (AA 1-554) [Saccharomyces cerevisiae]>gb|AAA34862.1|(M37267) phosphoglucose
Contig367	493	2.7e-46	19	498	pir||T39509 glucose-6-phosphate isomerase, cytosolic - fission yeast(Schizosaccharomyces pombe) >emb|CAA22338.1| (AL034433)glucose-6-phosphate isomerase, cytosolic [Schizosaccharomycespombe]
Contig877	363	3.3e-32	290	625	gi|6319673 ref|NP_009755.1|PGI1| Glucose-6-phosphate isomerase; Pgi1p>sp|P12709|G6PI_YEAST GLUCOSE-6-PHOSPHATE ISOMERASE (GPI)(PHOSPHOGLUCOSE ISOMERASE) (PGI) (PHOSPHOHEXOSE ISOMERASE) (PHI)>pir||NUBY glucose-6-phosphate isomerase (EC 5.3.1.9) - yeast(Saccharomyces cerevisiae) >emb|CAA32158.1| (X13977)phosphoglucoseisomerase (AA 1-554) [Saccharomyces cerevisiae]>gb|AAA34862.1|(M37267) phosphoglucose
<aldolase>					
b4e11fs.f1	149	1.7e-09	133	426	emb|CAB82026.1| (AL161755) putative aldolase [Streptomyces coelicolor A3(2)]
<triose-phosphate isomerase>					
Contig986	799	9e-79	226	957	sp|P04828|TPIS_EMENI TRIOSEPHOSPHATE ISOMERASE (TIM)>pir||ISASTNtriose-phosphate isomerase (EC 5.3.1.1) - Emericella nidulans>dbj|BAA00908.1| (D10019) triosephosphate isomerase [Emericellanidulans]
<glyceraldehyde-3-phosphate dehydrogenase>					

Contig1007	1151	4.4e-116	43	801	sp|P17730|G3P2_TRIKO GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE 2 (GAPDH2)>dbj|BAA03391.1| (D14518) glyceraldehydephosphate dehydrogenase [Trichoderma koningii]
Contig981	388	3.5e-35	272	523	sp|P35143|G3P_COLGL GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (GAPDH)>pir||JN0452 glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12)- anthracnose fungus (Colletotrichum gloeosporioides)>gb|AAA02485.1| (M88109)glyceraldehyde-3-phosphate dehydrogenase [Colletotrichum gloeosporioides]>gb|AAA02486.1| (M93427)glyceraldehyde 3-phosphate dehydrogenase [Colletotrichumgloeosporioides]
<phosphoglycerate kinase>					
Contig289	664	1.4e-64	2	460	sp|P24590|PGK_TRIVI PHOSPHOGLYCERATE KINASE >pir||S25381 phosphoglyceratekinase (EC 2.7.2.3) - fungus (Trichoderma viride)
Contig743	445	3.3e-41	181	516	sp|P24590|PGK_TRIVI PHOSPHOGLYCERATE KINASE >pir||S25381 phosphoglyceratekinase (EC 2.7.2.3) - fungus (Trichoderma viride)
<pyruvate kinase>					
Contig529	1269	1.6e-128	13	849	sp|P31865|KPYK_TRIRE PYRUVATE KINASE >pir||JN0780 pyruvate kinase (EC2.7.1.40) - fungus (Trichoderma reesei) >gb|AAA02922.1| (L07060)pyruvate kinase [Hypocrea jecorina]
Contig610	455	3e-42	194	520	sp|P31865|KPYK_TRIRE PYRUVATE KINASE >pir||JN0780 pyruvate kinase (EC2.7.1.40) - fungus (Trichoderma reesei) >gb|AAA02922.1| (L07060)pyruvate kinase [Hypocrea jecorina]
Contig748	319	2.4e-27	181	450	sp|P31865|KPYK_TRIRE PYRUVATE KINASE >pir||JN0780 pyruvate kinase (EC2.7.1.40) - fungus (Trichoderma reesei) >gb|AAA02922.1| (L07060)pyruvate kinase [Hypocrea jecorina]
<2,3-bisphosphoglycerate-independentphosphoglycerate mutase>					
Contig901	211	1.6e-15	295	516	gb|AAD26328.1|AF1200 (AF120091) 2,3-bisphosphoglycerate-independentphosphoglycerate mutase [Bacillus stearothermophilus]
<enolase>					
Contig1003	1869	4.3e-192	62	1375	sp|Q12560|ENO_ASPOR ENOLASE (2-PHOSPHOGLYCERATE DEHYDRATASE)(2-PHOSPHO-D-GLYCERATE HYDRO-LYASE)>pir||JC45426beta-hydroxyhyoscyamine epoxidase (EC 1.14.11.14) -Aspergillusoryzae >dbj|BAA09973.1| (D63941) enolase [Aspergillus oryzae]>dbj|BAA23760.1| (D64113) enolase [Aspergillus oryzae]>prf||2205241A enolase [Aspergillus oryzae]
Contig415	594	5.9e-57	28	459	sp|P42040|ENO_CLAHE ENOLASE (2-PHOSPHOGLYCERATE DEHYDRATASE)(2-PHOSPHO-D-GLYCERATE HYDRO-LYASE) (ALLERGEN CLA H 6) (CLA HVI)

Contig13	404	7.8e-37	204	479	sp|P42040|ENO_CLAHE ENOLASE (2-PHOSPHOGLYCERATE DEHYDRATASE) (2-PHOSPHO-D-GLYCERATE HYDRO-LYASE) (ALLERGEN CLA H 6) (CLA HVI)
Contig8	209	1.8e-15	200	352	sp|Q12560|ENO_ASPOR ENOLASE (2-PHOSPHOGLYCERATE DEHYDRATASE) (2-PHOSPHO-D-GLYCERATE HYDRO-LYASE) >pir||JC45426beta-hydroxyhyoscyamine epoxidase (EC 1.14.11.14) - Aspergillusoryzae >dbj|BAA09973.1| (D63941) enolase [Aspergillus oryzae]>dbj|BAA23760.1| (D64113) enolase [Aspergillus oryzae]>prf||2205241A enolase [Aspergillus oryzae]
2. Gluconeogenesis (2)					
<pyruvate carboxylase>					
Contig742	978	1.2e-97	3	737	gb|AAC69197.1| (AF097728) pyruvate carboxylase [Aspergillus terreus]
<fructose-1,6-bisphosphatase>					
n3a01fs.f1	411	1.4e-37	150	533	sp|Q05079|F16P_KLULA FRUCTOSE-1,6-BISPHOSPHATASE (D-FRUCTOSE-1,6-BISPHOSPHATE1-PHOSPHOHYDROLASE) (FBPASE) >pir||S29397 fructose-bisphosphatase(EC 3.1.3.11) - yeast (Kluyveromyces marxianus var. lactis)>emb|CAA49728.1| (X70181) fructose-bisphosphatase [Kluyveromyceslactis]
3. Pentose-phosphate pathway (12)					
<glucose-6-phosphate dehydrogenase>					
Contig279	958	1.6e-95	148	813	sp|P48826|G6PD_ASPNG GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (G6PD) >pir||S57485glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) - Aspergillusniger >emb|CAA61194.1| (X87942) glucose-6-phosphate 1-dehydrogenase[Aspergillus niger]
m3c05fs.r1	791	7.4e-78	49	534	emb|CAA54840.1| (X77829) glucose-6-phosphate 1-dehydrogenase [Aspergillusniger]
<6-phosphogluconolactonase>					
Contig61	410	1.7e-37	58	528	sp|O74455|6PGL_SCHPO PROBABLE 6-PHOSPHOGLUCONOLACTONASE (6PGL) >pir||T41100sol1 family protein - fission yeast (Schizosaccharomyces pombe)>emb|CAA20749.1| (AL031535) sol1 family protein[Schizosaccharomyces pombe]
Contig590	255	5.1e-21	457	717	sp|O74455|6PGL_SCHPO PROBABLE 6-PHOSPHOGLUCONOLACTONASE (6PGL) >pir||T41100sol1 family protein - fission yeast (Schizosaccharomyces pombe)>emb|CAA20749.1| (AL031535) sol1 family protein[Schizosaccharomyces pombe]
<phosphogluconate dehydrogenase>					

Contig998	1518	6.7e-155	78	1280	sp|O13287|6PGD_CANAL 6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING>dbj|BAA21690.1| (AB006102) 6-phosphogluconate dehydrogenase[Candida albicans]
Contig999	354	1.6e-31	252	530	gi|6321977 ref|NP_012053.1|GND1| 6-phosphogluconate dehydrogenase; probableGND gene; Gnd1p >sp|P38720|6PG1_YEAST 6-PHOSPHOGLUCONATEDEHYDROGENASE, DECARBOXYLATING 1 >pir||S46671 phosphogluconatedehydrogenase (decarboxylating) (EC 1.1.1.44) - yeast(Saccharomyces cerevisiae) >gb|AAB68452.1| (U00028) Yhr183wp[Saccharomyces cerevisiae] >emb|CAA86600.1| (Z46631)6-phosphogluconate dehydrogenase
<ribulose-phosphate 3-epimerase>					
c3b10fs.r1	317	1.4e-27	48	347	gi|6322341 ref|NP_012414.1|RPE1| D-ribulose-5-Phosphate 3-epimerase; Rpe1p>sp|P46969|RPE_YEAST RIBULOSE-PHOSPHATE 3-EPIMERASE(PENTOSE-5-PHOSPHATE 3-EPIMERASE) (PPE) (RPE) >pir||S51587 POS18protein - yeast (Saccharomyces cerevisiae) >pir||S72623ribulose-5-phosphate-epimerase (EC 5.1.-.-) - yeast (Saccharomycescerevisiae) >emb|CAA58554.1| (X83571)Ribulose-5-Phosphate-Epimerase [Saccharomyces
c3b10fs.f1	255	5.1e-21	223	453	gb|AAF01048.1|AF1893 (AF189365) D-ribulose-5-phosphate 3-epimerase [Oryzasativa]
<transketolase>					
Contig769	1129	1.1e-113	2	1006	sp|P34736|TKT_PICST TRANSKETOLASE (TK) >pir||S37439 transketolase (EC 2.2.1.1)- yeast (Pichia stipitis) >emb|CAA81260.1| (Z26486)transketolase[Pichia stipitis]
Contig631	734	8.5e-72	2	529	sp|Q12630|TKT1_KLULA TRANSKETOLASE (TK) >gb|AAB05935.1| (U65983) transketolase[Kluyveromyces lactis]
Contig968	344	1.6e-30	201	527	dbj|BAA13834.1| (D89172) similar to Saccharomyces serevisiae transketolase2(TK2), SWISS-PROT Accession Number P33315 [Schizosaccharomycespombe]
<transaldolase>					
Contig963	1065	6.6e-107	107	1066	pir||T40834 transaldolase - fission yeast (Schizosaccharomyces pombe)>emb|CAA18994.1| (AL023518) transaldolase [Schizosaccharomycespombe]
4. Pyruvate dehydrogenase-three kinds of enzymes (7)					
<pyruvate dehydrogenase>					
Contig865	412	1.1e-37	160	579	sp|Q10489|ODPA_SCHPO PYRUVATE DEHYDROGENASE E1 COMPONENT ALPHA SUBUNIT,MITOCHONDRIAL PRECURSOR (PDHE1-A) >pir||T38417

					pyruvatedehydrogenase complex alpha chain precursor, mitochondrial - fission yeast (Schizosaccharomyces pombe) >emb|CAA97360.1| (Z73100)pyruvate dehydrogenase e1 component alpha subunit, mitochondrialprecursor [Schizosaccharomyces pombe]
Contig589	306	1.9e-26	213	590	gi|6321026 ref|NP_011105.1|PDA1| alpha subunit of pyruvate dehydrogenase (E1alpha); Pda1p >pir||DEBYPA pyruvate dehydrogenase (lipoamide) (EC1.2.4.1) alpha chain precursor - yeast (Saccharomyces cerevisiae)>gb|AAB64705.1| (U18922) Pda1p: alpha subunit of pyruvatedehydrogenase [Saccharomyces cerevisiae]
Contig847	259	1.9e-21	428	700	gb|AAA34583.1| (M98476) pyruvate dehydrogenase E1-beta subunit [Saccharomycescerevisiae]
Contig722	252	9.3e-21	304	570	sp|Q09171|ODPB_SCHPO PYRUVATE DEHYDROGENASE E1 COMPONENT BETA SUBUNIT,MITOCHONDRIAL PRECURSOR (PDHE1-B) >pir||JC4080 pyruvatedehydrogenase (lipoamide) (EC 1.2.4.1) E1 beta chain - fissionyeast (Schizosaccharomyces pombe) >emb|CAA53303.1| (X75648)putative pyruvate dehydrogenase [Schizosaccharomyces pombe]>emb|CAB10808.1| (Z97992) pyruvate dehydrogenase e1 component betasubunit, mitochondrial precursor
Contig566	254	2.1e-20	338	619	sp|P20285|ODP2_NEUCR DIHYDROLIPOAMIDE ACETYLTRANSFERASE COMPONENT OF PYRUVATEDEHYDROGENASE COMPLEX, MITOCHONDRIAL PRECURSOR (E2) (PDC-E2) (MRP3)>pir||A30775 dihydrolipoamide acetyltransferase homolog - Neurospora crassa >gb|AAA60452.1| (J04432) ribosomal protein[Neurospora crassa]
<dihydrolipoamide acetyltransferase>					
Contig898	668	7.9e-65	142	960	sp|P20285|ODP2_NEUCR DIHYDROLIPOAMIDE ACETYLTRANSFERASE COMPONENT OF PYRUVATEDEHYDROGENASE COMPLEX, MITOCHONDRIAL PRECURSOR (E2) (PDC-E2) (MRP3)>pir||A30775 dihydrolipoamide acetyltransferase homolog - Neurospora crassa >gb|AAA60452.1| (J04432) ribosomal protein[Neurospora crassa]
<dihydrolipoamide dehydrogenase>					
Contig304	198	3.8e-14	408	557	gi|6321091 ref|NP_011169.1|LPD1| dihydrolipoamide dehydrogenase precursor(mature protein is the E3 component of alpha-ketoacid dehydrogenasecomplexes); Lpd1p >sp|P09624|DLDH_YEAST DIHYDROLIPOAMIDEDEHYDROGENASE, MITOCHONDRIAL PRECURSOR >pir||A30151dihydrolipoamide dehydrogenase (EC 1.8.1.4) precursor - yeast(Saccharomyces cerevisiae) >gb|AAA34565.1| (J03645)dihydrolipoamide dehydrogenase [Saccharomyces

5. Tricarboxylic acid (TCA) cycle (21)

<citrate synthase>					
Contig910	1278	2e-129	168	953	sp|P34085|CISY_NEUCR CITRATE SYNTHASE, MITOCHONDRIAL PRECURSOR >pir||S41563citrate (si)-synthase (EC 4.1.3.7), mitochondrial - Neurosporacrassa >gb|AAA16630.1| (M84187) mitochondrial citrate synthase[Neurospora crassa]
Contig818	445	3.3e-41	210	482	emb|CAB77625.1| (AJ243204) citrate synthase [Aspergillus niger]
<aconitase>					
Contig770	1337	9e-136	1	897	gb|AAC61778.1| (AF093142) aconitase [Aspergillus terreus]
Contig411	795	3e-78	6	581	gb|AAC61778.1| (AF093142) aconitase [Aspergillus terreus]
m1a10fs.r1	390	2e-34	192	458	gb|AAC61778.1| (AF093142) aconitase [Aspergillus terreus]
Contig937	344	1.6e-29	241	501	gb|AAC61778.1| (AF093142) aconitase [Aspergillus terreus]
<isocitrate dehydrogenase>					
Contig707	717	5e-70	83	616	sp|P79089|IDHP_ASPNG ISOCITRATE DEHYDROGENASE [NADP], MITOCHONDRIAL PRECURSOR(OXALOSUCCINATE DECARBOXYLASE) (IDH) (NADP+-SPECIFIC ICDH) (IDP)>dbj|BAA19073.1| (AB000261) NADP-dependent isocitrate dehydrogenaseprecursor [Aspergillus niger] >dbj|BAA19074.1| (AB000262)NADP-dependent isocitrate dehydrogenase precursor [Aspergillusniger]
a1g04fs.r1	655	1.9e-63	9	434	sp|P79089|IDHP_ASPNG ISOCITRATE DEHYDROGENASE [NADP], MITOCHONDRIAL PRECURSOR(OXALOSUCCINATE DECARBOXYLASE) (IDH) (NADP+-SPECIFIC ICDH) (IDP)>dbj|BAA19073.1| (AB000261) NADP-dependent isocitrate dehydrogenaseprecursor [Aspergillus niger] >dbj|BAA19074.1| (AB000262)NADP-dependent isocitrate dehydrogenase precursor [Aspergillusniger]
t2e11fs.f1	515	1.3e-48	125	535	gi|6322097 ref|NP_012172.1|LYS12| Homo-isocitrate dehydrogenase; ys12p>sp|P40495|YIJ4_YEAST HYPOTHETICAL 40.1 KD PROTEIN IN SGA1-KTR7INTERGENIC REGION >pir||S49786 3-isopropylmalate dehydrogenasehomolog YIL094c - yeast (Saccharomyces cerevisiae) >emb|CAA86700.1|(Z46728) YI9910.02c, orf, len: 371, CAI: 0.31, similar toisocitrate and isopropylmalate dehyrogenases [Saccharomycescerevisiae]
e2e06fs.r1	504	1.6e-47	142	483	gb|AAB63461.1| (AF009036) NAD(+)-isocitrate dehydrogenase subunit I[Ajellomyces capsulatus]
e2e06fs.f1	195	4.5e-14	351	509	gb|AAB63461.1| (AF009036) NAD(+)-isocitrate dehydrogenase subunit I[Ajellomyces capsulatus]
<alpha-ketoglutarate dehydrogenase>					

b1e07fs.r1	483	3.5e-44	35	445	gi|6322066 ref|NP_012141.1|KGD1| alpha-ketoglutarate dehydrogenase; Kgd1p>sp|P20967|ODO1_YEAST 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT,MITOCHONDRIAL PRECURSOR (ALPHA-KETOGLUTARATE DEHYDROGENASE)>pir||DEBY oxoglutarate dehydrogenase (lipoamide) (EC 1.2.4.2)precursor - yeast (Saccharomyces cerevisiae) >emb|CAA86867.1|(Z46833) 2-oxoglutarate dehydrogenase E1 component [Saccharomycescerevisiae]
Contig203	223	2.6e-16	306	539	gi|6322066 ref|NP_012141.1|KGD1| alpha-ketoglutarate dehydrogenase; Kgd1p>sp|P20967|ODO1_YEAST 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT,MITOCHONDRIAL PRECURSOR (ALPHA-KETOGLUTARATE DEHYDROGENASE)>pir||DEBY oxoglutarate dehydrogenase (lipoamide) (EC1.2.4.2)precursor - yeast (Saccharomyces cerevisiae)>emb|CAA86867.1|(Z46833) 2-oxoglutarate dehydrogenase E1 component [Saccharomycescerevisiae]
<succinyl-coa synthetase>					
d3b12fs.r1	381	2e-34	126	515	pir||T41038 atp-specific succinyl-coa synthetase beta subunit - fission yeast(Schizosaccharomyces pombe) >emb|CAA22492.1| (AL034491)atp-specific succinyl-coa synthetase beta subunit[Schizosaccharomyces pombe]
s3d02fs.r1	364	1.5e-32	121	621	gi|6324716 ref|NP_014785.1|LSC1| alpha subunit of succinyl-CoA ligase; Lsc1p>sp|P53598|SUCA_YEAST PROBABLE SUCCINYL-COA LIGASE [GDP-FORMING]ALPHA-CHAIN, MITOCHONDRIAL PRECURSOR (SUCCINYL-COA SYNTHETASE,ALPHA CHAIN) (SCS-ALPHA) >pir||S61696 succinate--CoA ligase (EC6.2.1.-) alpha chain - yeast (Saccharomyces cerevisiae)>emb|CAA64059.1| (X94335) YOR3352w [Saccharomyces cerevisiae]>emb|CAA99342.1|
<FUMARATE HYDRATASE>					
Contig74	637	1.4e-61	5	547	sp|P55250|FUMH_RHIOR FUMARATE HYDRATASE PRECURSOR (FUMARASE)>emb|CAA55314.1|(X78576) fumarase [Rhizopus oryzae]
Contig75	315	4.2e-27	305	589	pir||T00433 fumarate hydratase (EC 4.2.1.2) - Arabidopsis thaliana>gb|AAB71399.1| (AF020303) fumarase [Arabidopsis thaliana]>gb|AAC62859.1| (AC002535) putative fumarase [Arabidopsis thaliana]
<malate dehydrogenase>					
Contig669	609	1.3e-58	69	644	gi|6322765 ref|NP_012838.1|MDH1| mitochondrial malate dehydrogenase; Mdh1p>sp|P17505|MDHM_YEAST MALATE DEHYDROGENASE,MITOCHONDRIAL PRECURSOR>pir||DEBYMM malate dehydrogenase (EC 1.1.1.37)precursor,mitochondrial - yeast (Saccharomyces

					cerevisiae)>gb|AAA34759.1|(J02841) malate dehydrogenase [Saccharomyces cerevisiae]>emb|CAA81923.1| (Z28085) ORF YKL085w [Saccharomyces cerevisiae]
Contig812	457	1.8e-42	22	786	pir||D69655 malate dehydrogenase malS - Bacillus subtilis >gb|AAC00287.1|(AF008220) putative malolactic enzyme [Bacillus subtilis]>emb|CAB14966.1| (Z99119) malate dehydrogenase (decarboxylating) [Bacillus subtilis]
l1g12fs.r1	282	3.9e-23	129	521	gb|AAD40139.1|AF1494 (AF149413) similar to malate dehydrogenases; PfamPF00390, Score=1290.5. E=0, N=1 [Arabidopsis thaliana]
o1d09fs.r1	120	4.5e-06	8	124	gi|6322765 ref|NP_012838.1|MDH1| mitochondrial malate dehydrogenase; Mdh1p>sp|P17505|MDHM_YEAST MALATE DEHYDROGENASE, MITOCHONDRIAL PRECURSOR>pir||DEBYMM malate dehydrogenase (EC 1.1.1.37) precursor,mitochondrial - yeast (Saccharomyces cerevisiae)>gb|AAA34759.1|(J02841) malate dehydrogenase [Saccharomyces cerevisiae]>emb|CAA81923.1| (Z28085) ORF YKL085w [Saccharomyces cerevisiae]
6. related reactions (3)					
<ATP citrate lyase>					
Contig1051	2900	2.4e-301	208	2169	emb|CAA12224.1| (AJ224922) ATP citrate lyase [Sordaria macrospora]>emb|CAB76165.1| (AJ243817) ATP citrate lyase, subunit 1 [Sordariamacrospora]
Contig1008	1208	4.6e-122	253	1056	emb|CAB76164.1| (AJ243817) ATP citrate lyase, subunit 2 [Sordaria macrospora]
Contig1001	1025	1.2e-102	341	1036	emb|CAB76164.1| (AJ243817) ATP citrate lyase, subunit 2 [Sordaria macrospora]
7. Fermentation, alcoholic (10)					
<pyruvate decarboxylase>					
Contig1006	971	6e-97	263	1093	gb|AAD16178.1| (AF098293) pyruvate decarboxylase [Aspergillus oryzae]
m2e12fs.r1	453	5.3e-42	44	538	gb|AAD16178.1| (AF098293) pyruvate decarboxylase [Aspergillus oryzae]
Contig979	235	2.6e-18	238	537	pir||T38759 probable pyruvate decarboxylase (EC 4.1.1.1) - fission yeast(Schizosaccharomyces pombe) (fragment) >emb|CAA93158.1| (Z69086)putative pyruvate decarboxylase (EC 4.1.1.1) [Schizosaccharomycespombe]
<alcohol dehydrogenase>					
Contig975	554	1e-52	278	754	sp|P41747|ADH1_ASPFL ALCOHOL DEHYDROGENASE I >gb|AAA32684.1| (L27434) alcoholdehydrogenase [Aspergillus flavus]

Contig895	458	1.3e-42	115	534	sp|P41747|ADH1_ASPFL ALCOHOL DEHYDROGENASE I >gb|AAA32684.1| (L27434)alcoholdehydrogenase [Aspergillus flavus]
j3e10fs.r1	211	4.3e-16	160	570	pir||T39671 alcohol dehydrogenase - fission yeast (Schizosaccharomyces pombe)>emb|CAA21911.1| (AL033389) alcohol dehydrogenase[Schizosaccharomyces pombe]
b4a09fs.r1	173	2.5e-12	155	439	gb|AAB62297.1| (U24215) p-cumic alcohol dehydrogenase [Pseudomonas putida]
m1f09fs.r1	175	5.2e-12	86	445	pir||T39671 alcohol dehydrogenase - fission yeast (Schizosaccharomyces pombe)>emb|CAA21911.1| (AL033389) alcohol dehydrogenase[Schizosaccharomyces pombe]
Contig625	133	2.1e-07	221	475	pir||T39671 alcohol dehydrogenase - fission yeast (Schizosaccharomyces pombe)>emb|CAA21911.1| (AL033389) alcohol dehydrogenase[Schizosaccharomyces pombe]
<toluenesulfonate zinc-independent alcohol dehydrogenase>					
s3b01fs.r1	249	2.3e-20	24	470	gb|AAC44807.1| (U32622) toluenesulfonate zinc-independent alcoholdehydrogenase [Comamonas testosteroni]
8.Fermentation, other (1)					
<LACTATE DEHYDROGENASE-pyruvate to lactate>					
p4c04fs.f1	298	7.7e-25	226	525	sp|Q12627|DLD1_KLULA D-LACTATE DEHYDROGENASE [CYTOCHROME] PRECURSOR (D-LACTATEFERRICYTOCHROME C OXIDOREDUCTASE) (D-LCR) >pir||S51528 D-lactatedehydrogenase (cytochrome) (EC 1.1.2.4) - yeast (Kluyveromycesmarxianus var. lactis) >emb|CAA50635.1| (X71628) D-lactatedehydrogenase (cytochrome) [Kluyveromyces lactis]
9. Monocarbon metabolism (5)					
<formate dehydrogenase>					
t4b10fs.f1	644	2.8e-62	145	573	sp|Q07103|FDH_NEUCR FORMATE DEHYDROGENASE (NAD-DEPENDENT FORMATEDEHYDROGENASE) (FDH) >pir||A47117 formate dehydrogenase (EC1.2.1.2) -Neurospora crassa >gb|AAA99900.1| (L13964) formatedehydrogenase [Neurospora crassa]
t4b10fs.r1	141	3.5e-08	211	315	sp|Q07103|FDH_NEUCR FORMATE DEHYDROGENASE (NAD-DEPENDENT FORMATEDEHYDROGENASE) (FDH) >pir||A47117 formate dehydrogenase (EC1.2.1.2) - Neurospora crassa >gb|AAA99900.1| (L13964) formatedehydrogenase [Neurospora crassa]
C1 Metabolism					
<methylenetetrahydrofolate reductase>					
Contig156	370	1.1e-32	14	466	sp|Q10258|MTHR_SCHPO PROBABLE METHYLENETETRAHYDROFOLATE REDUCTASE 1>pir||T38920 methylenetetrahydrofolate reductase 2 - fission

					yeast(Schizosaccharomyces pombe) >emb|CAA93581.1| (Z69728)methylenetetrahydrofolate reductase 2 [Schizosaccharomyces pombe]
<c-1-tetrahydrofolate synthase>					
Contig681	628	3.4e-60	8	628	pir||T40723 c-1-tetrahydrofolate synthase - fission yeast (Schizosaccharomycespombe) >emb|CAA17019.1| (AL021816) c-1-tetrahydrofolate synthase[Schizosaccharomyces pombe] >emb|CAB46709.1| (AL096796)c-1-tetrahydrofolate synthase [Schizosaccharomyces pombe]
Contig537	321	8.8e-27	229	531	gi|6319558 ref|NP_009640.1|MIS1| mitochondrial C1-tetrahydroflate synthase;Mis1p >sp|P09440|C1TM_YEAST C-1-TETRAHYDROFOLATE SYNTHASE,MITOCHONDRIAL PRECURSOR (C1-THF SYNTHASE) [INCLUDES:METHYLENETETRAHYDROFOLATE DEHYDROGENASE ; METHENYLTETRAHYDROFOLATECYCLOHYDROLASE ; FORMYLTETRAHYDROFOLATE SYNTHETASE] >pir||A28174C1-tetrahydrofolate synthase precursor, mitochondrial -yeast(Saccharomyces cerevisiae)

10. Metabolism of energy reserves (glycogen, starch, trehalose) (12)

10.1. Glycogen degradation

<glycogen phosphorylase>					
Contig594	1038	5.5e-104	2	886	emb|CAA28273.1| (X04604) glycogen phosphorylase (AA 1-891) [Saccharomycescerevisiae] >prf||1212353A phosphorylase,glycogen [Saccharomycescerevisiae]
l3h04fs.r1	619	2.7e-59	11	538	gi|6325418 ref|NP_015486.1|GPH1| Glycogen phosphorylase; Gph1p >pir||S61144glycogen phosphorylase (EC 2.4.1.1) - yeast (Saccharomycescerevisiae) >gb|AAB68057.1| (U28371) Glycogen phosphorylase (SwissProt. accession number P06738) [Saccharomyces cerevisiae]
Contig293	285	5.5e-23	338	649	emb|CAA28273.1| (X04604) glycogen phosphorylase (AA 1-891) [Saccharomycescerevisiae] >prf||1212353A phosphorylase,glycogen [Saccharomycescerevisiae]
Contig848	280	1.8e-22	252	527	gi|6325418 ref|NP_015486.1|GPH1| Glycogen phosphorylase; Gph1p >pir||S61144glycogen phosphorylase (EC 2.4.1.1) - yeast (Saccharomycescerevisiae) >gb|AAB68057.1| (U28371) Glycogen phosphorylase (SwissProt. accession number P06738) [Saccharomyces cerevisiae]
<phosphoglucomutase-glycogen deg, glu1PO4 to glu6PO4>					

l2a10fs.r1	762	8.2e-75	3	530	gb|AAF36531.1|AF1352 (AF135264) phosphoglucomutase [Emericella nidulans]
m4c01fs.r1	525	1.2e-49	54	473	gb|AAF36531.1|AF1352 (AF135264) phosphoglucomutase [Emericella nidulans]
Contig45	426	3.3e-39	141	548	gb|AAF36531.1|AF1352 (AF135264) phosphoglucomutase [Emericella nidulans]
<GLYCOGEN DEBRANCHING ENZYME>					
o1b11fs.r1	571	3.8e-53	9	494	gi|6325442 ref|NP_015510.1|YPR184W| Ypr184wp >pir||S598414-alpha-glucanotransferase / amylo-1,6-glucosidase homolog YPR184w- yeast (Saccharomyces cerevisiae) >gb|AAB68117.1| (U25842) Highlysimilar to Glycogen debranching enzyme (4-alpha-glucanotransferase,Swiss Prot. accession number P35573) [Saccharomyces cerevisiae]
10.2. Starch degradation					
<mannosyl-oligosaccharide glucosidase>					
b4a07fs.f1	207	9.9e-15	205	429	sp|O14255|GCS1_SCHPO PROBABLE MANNOSYL-OLIGOSACCHARIDE GLUCOSIDASE (PROCESSINGA-GLUCOSIDASE I) >pir||T39059 probable mannosyl-oligosaccharideglucosidase (EC 3.2.1.106) - fission yeast (Schizosaccharomycespombe) >emb|CAB11295.1| (Z98603) mannosyl-oligosaccharideglucosidase [Schizosaccharomyces pombe]
<nucleoside diphosphate-sugar hydrolase>					
Contig504	356	1e-31	73	654	emb|CAB87374.1| (AL163702) nucleoside diphosphate-sugar hydrolase ofthe MutT(nudix) family [Schizosaccharomyces pombe]
10.3. Trehalose degradation					
<trehalose-6-phosphate phosphatase>					
f2c09fs.r1	333	4e-28	44	511	gi|6320279 ref|NP_010359.1|TPS2| Trehalose-6-phosphate phosphatase; Tps2p>sp|P31688|TPS2_YEAST TREHALOSE-PHOSPHATASE (TREHALOSE 6-PHOSPHATEPHOSPHATASE) (TPP) >pir||S48761 trehalose-phosphatase (EC3.1.3.12)- yeast (Saccharomyces cerevisiae) >emb|CAA86796.1| (Z46796)trehalose-phosphatase [Saccharomyces cerevisiae]>emb|CAA98893.1|(Z74370) ORF YDR074w [Saccharomyces cerevisiae]
b3c12fs.r1	216	1.3e-15	40	444	gi|6320279 ref|NP_010359.1|TPS2| Trehalose-6-phosphate phosphatase; Tps2p>sp|P31688|TPS2_YEAST TREHALOSE-PHOSPHATASE (TREHALOSE 6-PHOSPHATEPHOSPHATASE) (TPP) >pir||S48761 trehalose-phosphatase (EC 3.1.3.12)- yeast (Saccharomyces cerevisiae) >emb|CAA86796.1| (Z46796)trehalose-phosphatase [Saccharomyces

cerevisiae]>emb|CAA98893.1|(Z74370) ORF YDR074w [Saccharomyces cerevisiae]

VII.2. fatty acid as energy source

1. lipase-triacylglycerols to glycerol+FA (2)

<lipase>

Contig80	228	3.7e-18	181	573	gb|AAC38151.1| (AF034088) lipase [Pseudomonas sp. B11-1]
Contig79	165	4.7e-11	117	455	gb|AAC38151.1| (AF034088) lipase [Pseudomonas sp. B11-1]

2. beta-oxydation of fatty acids (6)

<long-chain-fatty-acid-CoA ligase>

e4a04fs.r1	465	4.1e-43	21	500	pir||T39766 probable long-chain-fatty-acid--coa ligase - fission yeast(Schizosaccharomyces pombe) >emb|CAA18399.1| (AL022304)putativelong-chain-fatty-acid--coa ligase [Schizosaccharomyces pombe]
e4a04fs.f1	111	0.0017	378	521	gi|6323903 ref|NP_013974.1|FAA4| long-chain fatty acid–CoA ligase andsynthetase 4; Faa4p >sp|P47912|LCF4_YEASTLONG-CHAIN-FATTY-ACID--COA LIGASE 4 (LONG-CHAIN ACYL-COA SYNTHETASE4) (FATTY ACID ACTIVATOR 4) >pir||S56060 long-chain-fatty-acid--CoAligase (EC 6.2.1.3) FAA4 - yeast (Saccharomyces cerevisiae)>emb|CAA88656.1| (Z48756) unknown [Saccharomyces cerevisiae]

<carnitine/acyl carnitine carrier>

i1h04fs.r1	302	5.2e-26	226	438	emb|CAB44434.1| (AJ011563) carnitine/acyl carnitine carrier [Emericellanidulans]

<L-carnitine dehydratase (caiB-2)>

c1c12fs.f1	185	3e-13	136	450	pir||F69373 L-carnitine dehydratase (caiB-2) homolog -Archaeoglobus fulgidus>gb|AAB90253.1| (AE001036) L-carnitine dehydratase (caiB-2) [Archaeoglobus fulgidus]

<3-ketoacyl-CoA thiolase>

b1e03fs.f1	395	7.8e-36	106	489	sp|Q05493|THIK_YARLI 3-KETOACYL-COA THIOLASE, PEROXISOMAL PRECURSOR(BETA-KETOTHIOLASE) (ACETYL-COA ACYLTRANSFERASE) (PEROXISOMAL3-OXOACYL-COA THIOLASE) >pir||S36838 acetyl-CoA C-acyltransferase(EC 2.3.1.16), peroxisomal - yeast (Yarrowia lipolytica)>emb|CAA49605.1| (X69988) acetyl-CoA acyltransferase [Yarrowialipolytica]
b1e03fs.r1	142	5.9e-17	161	316	sp|Q05493|THIK_YARLI 3-KETOACYL-COA THIOLASE, PEROXISOMAL PRECURSOR(BETA-KETOTHIOLASE) (ACETYL-COA ACYLTRANSFERASE)

					(PEROXISOMAL3-OXOACYL-COA THIOLASE) >pir||S36838 acetyl-CoA C-acyltransferase(EC 2.3.1.16), peroxisomal - yeast (Yarrowia lipolytica)>emb|CAA49605.1| (X69988) acetyl-CoA acyltransferase [Yarrowialipolytica]
3. Ketone body metabolism (2)					
<SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASE-acetoacetate to acetoacyl-coA>					
Contig89	247	1.8e-19	238	528	sp|Q09450|SCOT_CAEEL PROBABLE SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASEPRECURSOR (3-OXOACID COA-TRANSFERASE) >pir||T18942 hypotheticalprotein C05C10.3 - Caenorhabditis elegans >emb|CAA88202.1| (Z48178)similar to 3-oxoacid CoA-transferase; cDNA EST EMBL:Z14816 comesfrom this gene; cDNA EST EMBL:Z14946 comes from this gene; cDNA ESTEMBL:D69746 comes from this gene; cDNA EST yk219b6.3 comes fromthis gene;
<HYDROXYMETHYLGLUTARYL-COA SYNTHASE-ketogenesis>					
Contig1050	1473	3.8e-150	153	1517	sp|P54874|HMCS_SCHPO HYDROXYMETHYLGLUTARYL-COA SYNTHASE (HMG-COA SYNTHASE)(3-HYDROXY-3-METHYLGLUTARYL COENZYME A SYNTHASE)>pir||S61875hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - fission yeast(Schizosaccharomyces pombe) >gb|AAB17601.1| (U32187)3-hydroxy-3-methylglutaryl coenzyme A synthase [Schizosaccharomycespombe] >emb|CAB11060.1| (Z98530)hydroxymethylglutaryl-coa synthase(EC 4.1.3.5)
VII.3. Metabolism of other energy sources (29)					
<glutathione-dependent formaldehyde dehydrognease>					
a3h03fs.r1	577	3.7e-55	23	466	sp|Q06099|FADH_CANMA GLUTATHIONE-DEPENDENT FORMALDEHYDE DEHYDROGENASE (FDH)(FALDH) >pir||JN0447 alcohol dehydrogenase (EC 1.1.1.1) FDH1 -yeast (Candida maltosa) >gb|AAA34344.1| (M58332) encodingformaldehyde resistance [Candida maltosa]
<aldehyde reductase>					
q2e08fs.r1	414	6.8e-38	63	404	pir||S78113 aldehyde reductase (NADPH) (EC 1.1.1.-) - fungus (Sporidiobolussalmonicolor) >gb|AAB17362.1| (U26463) NADPH-dependent aldehydereductase [Sporidiobolus salmonicolor]
Contig752	263	6.9e-22	212	523	pir||S78113 aldehyde reductase (NADPH) (EC 1.1.1.-) - fungus (Sporidiobolussalmonicolor) >gb|AAB17362.1| (U26463) NADPH-dependent aldehydereductase [Sporidiobolus salmonicolor]
<glycerol-3-phosphate dehydrogenase>					

r4f10fs.f1	128	1.5e-06	203	523	pir||B75218 glycerol-3-phosphate dehydrogenase (glpa) PAB0183 - Pyrococcusabyssi (strain Orsay) >emb|CAB49193.1| (AJ248283)glycerol-3-phosphate dehydrogenase (glpA) [Pyrococcus abyssi]
<**acetamidase**-allows acetamide and formamide as sole C or N source>					
l1h11fs.r1	545	1e-51	41	511	sp|P08158|AMDS_EMENI ACETAMIDASE >pir||A26511 amdS protein - Emericellanidulans >gb|AAA33295.1| (M16371) acetamidase enzyme [Emericellanidulans]
l1h11fs.f1	420	1.7e-38	143	517	sp|P08158|AMDS_EMENI ACETAMIDASE >pir||A26511 amdS protein - Emericellanidulans >gb|AAA33295.1| (M16371) acetamidase enzyme [Emericellanidulans]
a4g06fs.r1	318	3.4e-27	3	470	pir||T41382 acetamidase - fission yeast (Schizosaccharomyces pombe)>emb|CAA19111.1| (AL023592) acetamidase [Schizosaccharomyces pombe]
a3e12fs.r1	197	6.1e-14	126	455	sp|P08158|AMDS_EMENI ACETAMIDASE >pir||A26511 amdS protein - Emericellanidulans >gb|AAA33295.1| (M16371) acetamidase enzyme [Emericellanidulans]
l1a12fs.f1	174	1.8e-11	145	519	pir||JS0633 amidase (EC 3.5.1.4) - Aspergillus oryzae>dbj|BAA01373.1|(D10492) acetamidase [Aspergillus oryzae]
a4g06fs.f1	169	6.5e-11	127	483	pir||T41382 acetamidase - fission yeast (Schizosaccharomyces pombe)>emb|CAA19111.1| (AL023592) acetamidase [Schizosaccharomyces pombe]
c1b01fs.f1	164	2.4e-10	202	465	pir||T39112 probable amidase - fission yeast (Schizosaccharomyces pombe)>emb|CAB60011.1| (AL132779) putative amidase [Schizosaccharomycespombe]
<**ALDEHYDE DEHYDROGENASE**-broad substrate specificity>					
Contig491	677	1e-65	26	760	sp|P40108|DHAL_CLAHE ALDEHYDE DEHYDROGENASE (ALDDH) (ALLERGEN CLA H 3) (CLA HIII) >pir||S43114 aldehyde dehydrogenase (NAD+) (EC 1.2.1.3) -fungus (Cladosporium herbarum) >emb|CAA55072.1| (X78228) aldehydedehydrogenase (NAD+) [Cladosporium herbarum]
Contig617	509	2.1e-63	271	663	sp|P41751|DHAL_ASPNG ALDEHYDE DEHYDROGENASE (ALDDH) >gb|AAA87596.1| (M32351)aldehyde dehydrogenase [Aspergillus niger]
Contig388	534	1.2e-50	111	518	sp|P40108|DHAL_CLAHE ALDEHYDE DEHYDROGENASE (ALDDH) (ALLERGEN CLA H 3) (CLAHIII) >pir||S43114 aldehyde dehydrogenase (NAD+) (EC 1.2.1.3) -fungus (Cladosporium herbarum) >emb|CAA55072.1| (X78228) aldehydedehydrogenase (NAD+) [Cladosporium herbarum]
Contig532	396	6e-36	8	667	gb|AAB62298.1| (U24215) p-cumic aldehyde dehydrogenase [Pseudomonas putida]

Contig532	396	6e-36	8	667	gb|AAB62298.1| (U24215) p-cumic aldehyde dehydrogenase [Pseudomonas putida]
p4a06fs.f1	373	1.6e-33	162	512	emb|CAB63554.1| (AL133522) probable aldehyde dehydrogenase[Schizosaccharomyces pombe]
Contig394	354	1.7e-31	109	648	gi|6323821 ref|NP_013892.1|ALD3| aldehyde dehydrogenase;Ald3p>sp|P54114|DHA4_YEAST ALDEHYDE DEHYDROGENASE [NAD(P)+] 1>pir||S54527 aldehyde dehydrogenase (NAD+) (EC 1.2.1.3) 3 - yeast(Saccharomyces cerevisiae) >emb|CAA89805.1| (Z49705) unknown[Saccharomyces cerevisiae]
p3f09fs.r1	320	1.3e-27	42	515	sp|P41751|DHAL_ASPNG ALDEHYDE DEHYDROGENASE (ALDDH) >gb|AAA87596.1| (M32351)aldehyde dehydrogenase [Aspergillus niger]
Contig835	302	1.1e-25	315	860	sp|P77674|YDCW_ECOLI PUTATIVE BETAINE ALDEHYDE DEHYDROGENASE (BADH)>pir||G64896 probable aldehyde dehydrogenase (EC 1.2.1.-) - Escherichia coli >dbj|BAA15073.1| (D90783) Betaine-aldehydedehydrogenase precursor (EC1.2.1.8) (BADH). [Escherichia coli]>dbj|BAA15079.1| (D90784)Betaine-aldehyde dehydrogenase precursor(EC 1.2.1.8) (BADH). [Escherichia coli] >gb|AAC74526.1| (AE000241)putative aldehyde dehydrogenase
Contig9	288	2.3e-24	207	590	pir||S45858 probable aldehyde dehydrogenase (EC 1.2.1.-) - yeast(Saccharomyces cerevisiae)
Contig2	245	1.8e-19	148	633	pir||S45858 probable aldehyde dehydrogenase (EC 1.2.1.-) - yeast(Saccharomyces cerevisiae)
<aldehyde-dehydrogenase-like protein>					
s2f01fs.r1	310	2.3e-26	95	466	emb|CAB76051.1| (AL157918) putative aldehyde-dehydrogenase-like protein[Schizosaccharomyces pombe]
<KETOL-ACID REDUCTOISOMERASE PRECURSOR>					
Contig1019	1777	2.4e-182	56	1270	sp|P38674|ILV5_NEUCR KETOL-ACID REDUCTOISOMERASE PRECURSOR (ACETOHYDROXY-ACIDREDUCTOISOMERASE) (ALPHA-KETO-BETA-HYDROXYLACIL REDUCTOISOMERASE)>pir||JC1428 ketol-acid reductoisomerase (EC 1.1.1.86) -Neurosporacrassa >gb|AAB00797.1| (M84189)alpha-keto-beta-hydroxylacylreductoisomerase [Neurospora crassa]
<Acetyl-CoA-Acetyltransferase>					
Contig1018	1163	2.8e-117	36	1202	pir||T39899 acetyl-coa acetyltransferase - fission yeast (Schizosaccharomycespombe) >emb|CAA22123.1| (AL033534) acetyl-coa acetyltransferase (EC2.3.1.9) [Schizosaccharomyces pombe]
j1h10fs.f1	599	1.8e-57	84	575	emb|CAB46281.1| (AJ243195) Acetyl-CoA-Acetyltransferase [Mycosphaerellagraminicola]
<OMEGA-6 FATTY ACID DESATURASE>					

Contig208	164	1.1e-10	302	523	sp|P48631|FD62_SOYBN OMEGA-6 FATTY ACID DESATURASE, ENDOPLASMIC RETICULUMISOZYME 2 >pir||T07688 omega-6 desaturase FAD2-2, microsomal-soybean >gb|AAB00860.1| (L43921) microsomal omega-6 desaturase[Glycine max]
<succinate-semialdehyde dehydrogenase>					
d2h03fs.f1	357	7.9e-32	118	447	emb|CAB65612.1| (AL136078) probable succinate-semialdehyde dehydrogenase[Schizosaccharomyces pombe]
d2h03fs.r1	206	4.9e-15	166	477	sp|P25526|GABD_ECOLI SUCCINATE-SEMIALDEHYDE DEHYDROGENASE [NADP+] (SSDH)>pir||F65045 succinate-semialdehyde dehydrogenase (NAD(P)+) (EC1.2.1.16) - Escherichia coli (strain K-12) >gb|AAC36831.1| (M88334)succinic semialdehyde dehydrogenase [Escherichia coli]>gb|AAC75708.1| (AE000351) succinate-semialdehyde dehydrogenase,NADP-dependent activity [Escherichia coli]
VII.4. Electron transport					
1. Complex I-NADH-ubiquinone (23)					
<NADH:Ubiquinone Oxidoreductase>					
o1g02fs.r1	898	4e-89	1	513	pir||S59926 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) 78K chain precursor -Neurospora crassa >gb|AAA98999.1| (L36813) NADH dehydrogenasesubunit [Neurospora crassa]
Contig360	789	1.2e-77	79	669	sp|P15578|NU2M_PODAN NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 2 >pir||S02154 NADHdehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - Podosporaanserina mitochondrion >emb|CAA32646.1| (X14485) ND2 (AA 1 - 556)[Podospora anserina]>emb|CAA38765.1| (X55026) NADH-ubiquinoneoxidoreductase subunit 2 [Podospora anserina]
Contig837	642	4.9e-62	152	649	pir||A36621 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) 22K chain precursor -Neurospora crassa
e2a09fs.f1	641	5.9e-62	150	545	sp|P23710|NUGM_NEUCR NADH-UBIQUINONE OXIDOREDUCTASE 30.4 KD SUBUNIT PRECURSOR(COMPLEX I-30KD) (CI-31KD) >pir||A35935 NADH dehydrogenase(ubiquinone) (EC 1.6.5.3) 31K chain precursor - Neurospora crassa
Contig875	584	7.3e-56	194	538	sp|P22142|NUCM_NEUCR NADH-UBIQUINONE OXIDOREDUCTASE 49 KD SUBUNIT PRECURSOR(COMPLEX I-49KD) (CI-49KD) >pir||S13801 NADH dehydrogenase(ubiquinone) (EC 1.6.5.3) 49K chain - Neurospora crassa>emb|CAA38368.1| (X54508) NADH dehydrogenase 49 kD subunit[Neurospora crassa]

q2g11fs.r1	567	4.3e-54	25 438	pir||S59926 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) 78K chain precursor -Neurospora crassa >gb|AAA98999.1| (L36813) NADH dehydrogenasesubunit [Neurospora crassa]
e2a09fs.r1	529	4.9e-50	26 487	sp|P23710|NUGM_NEUCR NADH-UBIQUINONE OXIDOREDUCTASE 30.4 KD SUBUNIT PRECURSOR(COMPLEX I-30KD) (CI-31KD) >pir||A35935 NADH dehydrogenase(ubiquinone) (EC 1.6.5.3) 31K chain precursor - Neurospora crassa
i1g03fs.r1	523	2e-49	36 416	sp|P24917|NUBM_NEUCR NADH-UBIQUINONE OXIDOREDUCTASE 51 KD SUBUNIT PRECURSOR(COMPLEX I-51KD) (CI-51KD) >pir||S17663 NADH dehydrogenase(ubiquinone) (EC 1.6.5.3) flavoprotein 1 precursor - Neurosporacrassa >emb|CAA39676.1| (X56227) 51 kD subunit of NADHdehydrogenase (ubiquinone) [Neurospora crassa]
r4c09fs.f1	504	2e-47	164 526	sp|Q02854|NUXM_NEUCR NADH-UBIQUINONE OXIDOREDUCTASE 21 KD SUBUNIT (COMPLEXI-21KD) (CI-21KD) >pir||S27171 NADH dehydrogenase (ubiquinone) (EC1.6.5.3) 20.9K chain - Neurospora crassa >emb|CAA43221.1| (X60829)NADH dehydrogenase, 21 kDa subunit [Neurospora crassa]
Contig628	494	5.1e-46	156 530	pir||S59926 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) 78K chain precursor -Neurospora crassa >gb|AAA98999.1| (L36813) NADH dehydrogenasesubunit [Neurospora crassa]
b1e08fs.r1	469	1e-43	15 437	sp|P25284|NUEM_NEUCR NADH-UBIQUINONE OXIDOREDUCTASE 40 KD SUBUNIT PRECURSOR(COMPLEX I-40KD) (CI-40KD) >pir||S13025 NADH dehydrogenase(ubiquinone) (EC 1.6.5.3) 40K chain - Neurospora crassa>emb|CAA39695.1| (X56238) 40 kD subunit of NADH dehydrogenase[Neurospora crassa]
Contig94	346	3.9e-41	3 290	sp|P15578|NU2M_PODAN NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 2 >pir||S02154 NADHdehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - Podosporaanserina mitochondrion >emb|CAA32646.1| (X14485) ND2 (AA 1 - 556)[Podospora anserina] >emb|CAA38765.1| (X55026) NADH-ubiquinoneoxidoreductase subunit 2 [Podospora anserina]
Contig892	402	1.3e-36	133 543	sp|P22142|NUCM_NEUCR NADH-UBIQUINONE OXIDOREDUCTASE 49 KD SUBUNIT PRECURSOR(COMPLEX I-49KD) (CI-49KD) >pir||S13801 NADH dehydrogenase(ubiquinone) (EC 1.6.5.3) 49K chain - Neurospora crassa>emb|CAA38368.1| (X54508) NADH dehydrogenase 49 kD subunit[Neurospora crassa]
Contig59	360	3.9e-32	268 825	sp|P25710|NUJM_NEUCR NADH-UBIQUINONE OXIDOREDUCTASE 21.3 KD SUBUNIT>pir||S14277 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3)

					21.3Kchain - Neurospora crassa >emb|CAA39949.1| (X56612) NADHdehydrogenase [Neurospora crassa]
Contig171	362	4.1e-32	277	870	pir||T11629 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) - fission yeast(Schizosaccharomyces pombe) >emb|CAB16382.1| (Z99260) putativenadh-dehydrogenase [Schizosaccharomyces pombe]
Contig740	347	9.1e-31	635	856	emb|CAB65525.1| (AJ250340) subunit NUKM of protein NADH:UbiquinoneOxidoreductase (Complex I) [Yarrowia lipolytica]
Contig642	308	1.3e-26	229	468	sp|Q12644|NUIM_NEUCR NADH-UBIQUINONE OXIDOREDUCTASE 23 KD SUBUNIT PRECURSOR(COMPLEX I-23KD) (CI-23KD) >emb|CAA64794.1| (X95547)ferredoxin-like iron-sulfur subunit of mitochondrial complex I[Neurospora crassa]
c3a11fs.f1	253	7.2e-21	295	456	sp|Q12644|NUIM_NEUCR NADH-UBIQUINONE OXIDOREDUCTASE 23 KD SUBUNIT PRECURSOR(COMPLEX I-23KD) (CI-23KD) >emb|CAA64794.1| (X95547)ferredoxin-like iron-sulfur subunit of mitochondrial complex I[Neurospora crassa]
t4b08fs.f1	231	1.5e-17	392	589	emb|CAA07265.1| (AJ006852) alternative NADH-dehydrogenase [Yarrowialipolytica]
o1a08fs.r1	176	1.2e-12	230	397	gb|AAB19471.1| 13 kDa-A polypeptide of iron-sulfur protein fraction of NADH:ubiquinone oxidoreductase [cattle, heart, PeptideMitochondrial Partial, 96 aa]
Contig306	149	8.7e-10	126	377	sp|Q02369|NI2M_BOVIN NADH-UBIQUINONE OXIDOREDUCTASE B22 SUBUNIT (COMPLEXI-B22) (CI-B22) >pir||S28256 NADH dehydrogenase (ubiquinone) (EC1.6.5.3) chain CI-B22 - bovine >emb|CAA46048.1| (X64836)NADH-ubiquinone oxidoreductase complex B22 subunit [Bos taurus]
Contig278	463	4.9e-43	97	489	sp|O47950|NUKM_NEUCR PROBABLE NADH-UBIQUINONE OXIDOREDUCTASE 19.3 KD SUBUNITPRECURSOR (COMPLEX I-19.3KD) (CI-19.3KD) >emb|CAA04802.1|(AJ001520) 19.3kD iron-sulfur subunit of mitochondrial complex I[Neurospora crassa]
<electron transfer protein>					
o1f03fs.f1	288	1.2e-23	242	562	sp|Q10361|YDBA_SCHPO HYPOTHETICAL 70.2 KD PROTEIN C22E12.10C IN CHROMOSOME I>pir||T38167 electron transfer protein - fission yeast(Schizosaccharomyces pombe) >emb|CAA93897.1| (Z70043) electrontransfer protein [Schizosaccharomyces pombe]
2. Complex II-Succinate-ubiquinone (6)					
<succinate dehydrogenase>					
Contig876	1295	3e-131	201	1139	emb|CAB61213.1| (AL132984) probable succinate dehydrogenase flavoproteinsubunit precursor(ec 1.3.5.1) [Schizosaccharomyces pombe]

Contig92	619	1.3e-59	226	696	sp|O42772|DHSB_MYCGR SUCCINATE DEHYDROGENASE [UBIQUINONE] IRON-SULFUR PROTEIN,MITOCHONDRIAL PRECURSOR (IP) >gb|AAB97419.1| (AF042062) succinatedehydrogenase iron-sulphur protein [Mycosphaerella graminicola]
g3a02fs.r1	598	2.3e-57	120	455	sp|P10444|DHSA_ECOLI SUCCINATE DEHYDROGENASE FLAVOPROTEIN SUBUNIT >pir||DEECSFsuccinate dehydrogenase (EC 1.3.99.1) flavoprotein - Escherichiacoli >dbj|BAA35390.1| (D90711) Succinate dehydrogenase (EC1.3.99.1) flavoprotein [Escherichia coli] >gb|AAC73817.1|(AE000175) succinate dehydrogenase, flavoprotein subunit[Escherichia coli]
Contig829	566	5.4e-54	305	829	emb|CAB61213.1| (AL132984) probable succinate dehydrogenase flavoproteinsubunit precursor(ec 1.3.5.1) [Schizosaccharomyces pombe]
k2c11fs.f1	291	4.3e-41	233	445	sp|O42772|DHSB_MYCGR SUCCINATE DEHYDROGENASE [UBIQUINONE] IRON-SULFUR PROTEIN,MITOCHONDRIAL PRECURSOR (IP) >gb|AAB97419.1| (AF042062) succinatedehydrogenase iron-sulphur protein [Mycosphaerella graminicola]
d1d10fs.r1	241	1.4e-19	186	491	emb|CAB66444.1| (AL136535) putative succinate dehydrogenase membrane anchorsubunit precursor [Schizosaccharomyces pombe]

3. Complex III-Ubiquinone to cytochrome C (3)

<cytochrome b5>

Contig964	309	8.8e-27	230	580	dbj|BAA82440.1| (AB022443) cytochrome b5 [Mortierella alpina] >dbj|BAA82441.1|(AB022444) cytochrome b5 [Mortierella alpina]

<cytochrome c oxidase>

Contig823	346	1e-30	258	668	gi|6321251 ref|NP_011328.1|COX4| subunit IV of cytochrome c oxidase; Cox4p>sp|P04037|COX4_YEAST CYTOCHROME C OXIDASE POLYPEPTIDE IV,MITOCHONDRIAL PRECURSOR >pir||OLBY4 cytochrome-c oxidase (EC1.9.3.1) chain IV precursor - yeast (Saccharomyces cerevisiae)>emb|CAA25665.1| (X01418) cytochrome c oxidase subunit IVprecursor [Saccharomyces cerevisiae]>emb|CAA62787.1| (X91489)cytochrome C oxidase chain IV
Contig807	340	4.6e-30	150	461	gi|6321842 ref|NP_011918.1|COX6| subunit VI of cytochrome c oxidase; Cox6p>sp|P00427|COX6_YEAST CYTOCHROME C OXIDASE POLYPEPTIDE VI,MITOCHONDRIAL PRECURSOR >pir||OTBY6 cytochrome-c oxidase (EC1.9.3.1) chain VI precursor - yeast (Saccharomyces cerevisiae)>gb|AAA66900.1| (M10138) cytochrome c oxidase subunit VI[Saccharomyces cerevisiae] >gb|AAB68899.1| (U00062) Cox6p:cytochrome c oxidase subunit VI

4. Other electron transport pathways (29)

<NADH oxidase>

m1h03fs.r1	382	1.8e-34	14	439	gb|AAF26274.1|AF1686 (AF168613) NADH oxidase [Aspergillus parasiticus]
m1h03fs.f1	144	2e-08	181	516	gb|AAF26274.1|AF1686 (AF168613) NADH oxidase [Aspergillus parasiticus]

<GLUTATHIONE REDUCTASE>

Contig755	1168	8.2e-118	80	1237	gi|6325166 ref|NP_015234.1|GLR1| Glutathione oxidoreductase; Glr1p>sp|P41921|GSHR_YEAST GLUTATHIONE REDUCTASE (GR) (GRASE)>pir||S61975 glutathione reductase (NADPH) (EC 1.6.4.2) - yeast(Saccharomyces cerevisiae)>gb|AAB68208.1| (U43281) Glr1p,Lpg17p[Saccharomyces cerevisiae]

<NADH-cytochrome b5 reductase>

Contig403	277	2.4e-23	265	525	dbj|BAA85586.1| (AB020034) NADH-cytochrome b5 reductase [Mortierella alpina]>dbj|BAA85587.1| (AB020035) NADH-cytochrome b5 reductase[Mortierella alpina]
d3c01fs.r1	188	7.9e-14	268	501	pir||T41677 probable nadh-cytochrome b5 reductase - fission yeast(Schizosaccharomyces pombe) >emb|CAA20696.1| (AL031530) putativenadh-cytochrome b5 reductase [Schizosaccharomyces pombe]

<electron transfer flavoprotein alpha-subunit precursor>

f3g04fs.f1	521	3.3e-49	95	514	sp|P78790|ETFA_SCHPO PROBABLE ELECTRON TRANSFER FLAVOPROTEIN ALPHA-SUBUNITPRECURSOR (ALPHA-ETF) >pir||T38439 electron transfer flavoproteinalpha-subunit precursor - fission yeast (Schizosaccharomyces pombe)>emb|CAA15825.1| (AL009227) electron transfer flavoproteinalpha-subunit precursor [Schizosaccharomyces pombe]
Contig316	316	1.8e-27	32	520	pir||T42416 probable electron transfer flavoprotein alpha chain precursor -fission yeast (Schizosaccharomyces pombe) (fragment)>dbj|BAA13801.1| (D89139) similar to Human electron transferflavoprotein alpha subunit precursor, SWISS-PROT Accession NumberP13804 [Schizosaccharomyces pombe]

<ubiquinone biosynthesis>

Contig73	447	2.3e-41	128	775	gi|6324699 ref|NP_014768.1|CAT5| may encode a protein involved in one or moremonoxygenase or hydroxylase steps of ubiquinone biosynthesis; Cat5p>sp|P41735|CAT5_YEAST CAT5 PROTEIN (UBIQUINONE BIOSYNTHESIS PROTEINCOQ7) >pir||S49912 CAT5 protein - yeast (Saccharomyces cerevisiae)>emb|CAA58105.1| (X82930) CAT5 [Saccharomyces cerevisiae]>emb|CAA62119.1| (X90518) putative [Saccharomyces

<NADPH DEHYDROGENASE>

<ubiquinone/menaquinone biosynthesis methyltransferase>

k4h03fs.r1	103	0.0035	171	347	pir||F75277 ubiquinone/menaquinone biosynthesis methyltransferase-Deinococcus radiodurans (strain R1) >gb|AAF11949.1|AE002071_3(AE002071)ubiquinone/menaquinone biosynthesis methyltransferase[Deinococcus radiodurans]

<Thioredoxin Reductase>

c2a04fs.f1	304	3.1e-26	203	460	sp|P51978|TRXB_NEUCR THIOREDOXIN REDUCTASE >dbj|BAA08090.1| (D45049)Thioredoxin Reductase (NADPH) [Neurospora crassa]

<NADPH-ferrihemoprotein reductase>

Contig684	882	1.9e-87	3	749	pir||S38427 NADPH--ferrihemoprotein reductase (EC 1.6.2.4) - Aspergillus niger>emb|CAA81550.1| (Z26938) NADPH cytochrome P450 oxidoreductase[Aspergillus niger] >prf||2119198A NADPH cytochrome P450 reductase[Aspergillus niger]
o4f11fs.r1	413	3.2e-37	9	497	pir||S38427 NADPH--ferrihemoprotein reductase (EC 1.6.2.4) - Aspergillus niger>emb|CAA81550.1| (Z26938) NADPH cytochrome P450 oxidoreductase[Aspergillus niger] >prf||2119198A NADPH cytochrome P450 reductase[Aspergillus niger]
Contig378	295	2.4e-24	196	453	pir||S38427 NADPH--ferrihemoprotein reductase (EC 1.6.2.4) - Aspergillus niger>emb|CAA81550.1| (Z26938) NADPH cytochrome P450 oxidoreductase[Aspergillus niger] >prf||2119198A NADPH cytochrome P450 reductase[Aspergillus niger]
Contig38	278	1.9e-22	295	525	pir||S38427 NADPH--ferrihemoprotein reductase (EC 1.6.2.4) Aspergillus niger>emb|CAA81550.1| (Z26938) NADPH cytochrome P450 oxidoreductase[Aspergillus niger] >prf||2119198A NADPH cytochrome P450 reductase[Aspergillus niger]
p1c08fs.f1	102	0.053	379	531	pir||T14081 NADPH--ferrihemoprotein reductase (EC 1.6.2.4) ATR2 - Arabidopsisthaliana >emb|CAB52465.1| (AL109796) NADPH-ferrihemoproteinreductase (ATR2) [Arabidopsis thaliana]>emb|CAB81014.1| (AL161576)NADPH-ferrihemoprotein reductase (ATR2) [Arabidopsis thaliana]

<Nfrl-neurula-specific ferredoxin reductase-like protein>

Contig774	265	3.1e-21	240	536	dbj|BAA22375.1| (D86491) Nfrl [Xenopus laevis]

<ubiquinol-cytochrome c reductase complex>

Contig36	678	7.4e-66	42	536	sp|P11913|MPPB_NEUCR MITOCHONDRIAL PROCESSING PEPTIDASE BETA SUBUNIT PRECURSOR(BETA-MPP) (UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX CORE PROTEINI)>pir||A29881 mitochondrial processing peptidase (EC 3.4.99.41)beta chain precursor - Neurospora crassa

				>gb|AAA33606.1| (M20928)processing enhancing protein precursor [Neurospora crassa]
o3b04fs.r1	471	6.1e-44	9 518	sp|P07056|UCRI_NEUCR UBIQUINOL-CYTOCHROME C REDUCTASE IRON-SULFUR SUBUNIT,MITOCHONDRIAL PRECURSOR (RIESKE IRON-SULFUR PROTEIN) (RISP)>pir||RDNCUF ubiquinol--cytochrome-c reductase (EC1.10.2.2)iron-sulfur protein - Neurospora crassa >emb|CAA26308.1| (X02472)cytochrome c reductase iron-sulfur subunit [Neurospora crassa]
Contig813	470	7.6e-44	112 498	sp|O60044|UCR2_NEUCR UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX CORE PROTEIN 2PRECURSOR >emb|CAA70067.1| (Y08841) core protein II [Neurosporacrassa]
Contig639	379	3.9e-34	324 527	sp|P07056|UCRI_NEUCR UBIQUINOL-CYTOCHROME C REDUCTASE IRON-SULFUR SUBUNIT,MITOCHONDRIAL PRECURSOR (RIESKE IRON-SULFUR PROTEIN) (RISP)>pir||RDNCUF ubiquinol--cytochrome-c reductase (EC 1.10.2.2)iron-sulfur protein - Neurospora crassa >emb|CAA26308.1| (X02472)cytochrome c reductase iron-sulfur subunit [Neurospora crassa]
Contig198	347	7.6e-31	10 465	sp|O60044|UCR2_NEUCR UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX CORE PROTEIN 2PRECURSOR >emb|CAA70067.1| (Y08841) core protein II [Neurosporacrassa]
Contig894	334	2e-29	149 523	sp|O60044|UCR2_NEUCR UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX CORE PROTEIN 2PRECURSOR >emb|CAA70067.1| (Y08841) core protein II [Neurosporacrassa]
c4d08fs.r1	113	5.5e-06	55 252	pir||T41058 ubiquinol-cytochrome c reductase complex 7.2 kd protein - fissionyeast (Schizosaccharomyces pombe) >emb|CAA20667.1| (AL031525)ubiquinol-cytochrome c reductase complex 7.2 kd protein[Schizosaccharomyces pombe]
<cytochrome oxidase>				
o1f04fs.r1	322	5.6e-28	278 523	gi|6320989 ref|NP_011068.1|COX15| cytochrome oxidase assembly factor; Cox15p>sp|P40086|COXW_YEAST CYTOCHROME C OXIDASE ASSEMBLY PROTEIN COX15>pir||S50644 cytochrome oxidase assembly factor COX15 - yeast(Saccharomyces cerevisiae) >gb|AAB64668.1| (U18917) Cox15p:Cytochrom oxidase assembly factor [Saccharomyces cerevisiae]>gb|AAA57471.1| (L38643)cytochrome oxidase assembly factor[Saccharomyces
<flavoprotein>				

Contig772	608	2.1e-58	128	916	pir||T38406 probable flavoprotein - fission yeast (Schizosaccharomyces pombe)(fragment) >emb|CAA97371.1| (Z73100) putative flavoprotein[Schizosaccharomyces pombe]
<NADH-dependent flavin oxidoreductase>					
Contig1042	899	3e-89	130	1332	pir||T39956 probable nadh-dependent flavin oxidoreductase - fission yeast(Schizosaccharomyces pombe) >emb|CAA22626.1| (AL035065) putativenadh-dependent flavin oxidoreductase [Schizosaccharomyces pombe]
Contig971	468	1.4e-43	218	829	pir||T39956 probable nadh-dependent flavin oxidoreductase - fission yeast(Schizosaccharomyces pombe) >emb|CAA22626.1| (AL035065) putativenadh-dependent flavin oxidoreductase [Schizosaccharomyces pombe]
d4h10fs.f1	191	9.4e-14	175	483	pir||H75303 probable NADH-dependent flavin oxidoreductase - Deinococcusradiodurans (strain R1) >gb|AAF11740.1|AE002052_3 (AE002052)NADH-dependent flavin oxidoreductase, putative [Deinococcusradiodurans]
Contig918	151	2.8e-09	282	518	pir||H75303 probable NADH-dependent flavin oxidoreductas Deinococcusradiodurans (strain R1) >gb|AAF11740.1|AE002052_3 (AE002052)NADH-dependent flavin oxidoreductase, putative [Deinococcusradiodurans]
7. ATP synthase and ADP, AMP (25)					
<ATP SYNTHASE>					
Contig1034	2262	9.9e-234	76	1614	sp|P23704|ATPB_NEUCR ATP SYNTHASE BETA CHAIN, MITOCHONDRIAL PRECURSOR>pir||JC1112 H+-transporting ATP synthase (EC 3.6.1.34) beta chain- Neurospora crassa >emb|CAA37756.1| (X53720) F(1)-ATPasebeta-subunit precursor (519 AA) [Neurospora crassa] >gb|AAA33562.1|(M84192) mitochondrial ATPase beta-subunit [Neurospora crassa]
Contig1014	1317	1.3e-133	95	1015	sp|P37211|ATPA_NEUCR ATP SYNTHASE ALPHA CHAIN, MITOCHONDRIAL PRECURSOR>pir||JC1111 H+-transporting ATP synthase (EC 3.6.1.34) alpha chain- Neurospora crassa >gb|AAA33560.1| (M84191) mitochondrial ATPasealpha-subunit [Neurospora crassa]
Contig858	763	6.3e-75	268	954	sp|Q01278|VATE_NEUCR VACUOLAR ATP SYNTHASE SUBUNIT E (V-ATPASE E SUBUNIT)(V-ATPASE 26 KD SUBUNIT) >gb|AAA87901.1| (U17641) vacuolar ATPase26 kDa subunit [Neurospora crassa]
e3c10fs.r1	686	1e-66	81	512	sp|P11593|VATB_NEUCR VACUOLAR ATP SYNTHASE SUBUNIT B (V-ATPASE 57 KD SUBUNIT)>pir||A30800 H+-transporting ATPase (EC 3.6.1.35) 57K

				chain,vacuolar - Neurospora crassa >gb|AAA33622.1| (J03956) vacuolarATPase vma-2 [Neurospora crassa]
Contig863	666	1.2e-64	100 525	sp|P53659|VATX_NEUCR VACUOLAR ATP SYNTHASE SUBUNIT AC39 (V-ATPASE AC39SUBUNIT) (V-ATPASE 41 KD SUBUNIT) >gb|AAB02771.1| (U36470) vacuolarATPase 41 kDa subunit [Neurospora crassa]
m2d02fs.r1	582	1.1e-55	156 497	sp|P21282|VATC_BOVIN VACUOLAR ATP SYNTHASE SUBUNIT C (V-ATPASE C SUBUNIT)>pir||A23671 H+-transporting ATPase (EC 3.6.1.35) chain C, vacuolar- bovine >gb|AAA30803.1| (J05681) H+ -ATPase C subunit [Bos taurus]
Contig190	403	7.3e-54	176 448	sp|P31413|VATL_NEUCR VACUOLAR ATP SYNTHASE 16 KD PROTEOLIPID SUBUNIT>pir||S43893 H+-transporting ATPase (EC 3.6.1.35)lipid-bindingprotein - Neurospora crassa >gb|AAA19974.1| (L07105)ATPaseproteolipid subunit [Neurospora crassa]
Contig575	488	1.1e-45	164 499	sp|P11592|VATA_NEUCR VACUOLAR ATP SYNTHASE CATALYTIC SUBUNIT A (V-ATPASE 67 KDSUBUNIT) >pir||PXNCV7 H+-transporting ATPase (EC 3.6.1.35),vacuolar, 67K chain - Neurospora crassa >gb|AAA33621.1| (J03955)vacuolar ATPase vma-1 [Neurospora crassa]
Contig833	483	3.7e-45	215 607	sp|P53659|VATX_NEUCR VACUOLAR ATP SYNTHASE SUBUNIT AC39 (V-ATPASE AC39SUBUNIT) (V-ATPASE 41 KD SUBUNIT) >gb|AAB02771.1| (U36470) vacuolarATPase 41 kDa subunit [Neurospora crassa]
Contig265	437	2.3e-40	204 497	sp|P11592|VATA_NEUCR VACUOLAR ATP SYNTHASE CATALYTIC SUBUNIT A (V-ATPASE 67 KDSUBUNIT) >pir||PXNCV7 H+-transporting ATPase (EC 3.6.1.35),vacuolar, 67K chain - Neurospora crassa >gb|AAA33621.1| (J03955)vacuolar ATPase vma-1 [Neurospora crassa]
Contig977	428	1.9e-39	260 550	sp|P37211|ATPA_NEUCR ATP SYNTHASE ALPHA CHAIN, MITOCHONDRIAL PRECURSOR>pir||JC1111 H+-transporting ATP synthase (EC 3.6.1.34) alpha chain- Neurospora crassa >gb|AAA33560.1| (M84191) mitochondrial ATPasealpha-subunit [Neurospora crassa]
Contig987	393	1.2e-35	93 530	sp|P00842|ATP9_NEUCR ATP SYNTHASE PROTEIN 9, MITOCHONDRIAL PRECURSOR(LIPID-BINDING PROTEIN) >pir||LWNCA H+-transporting ATP synthase(EC 3.6.1.34) lipid-binding protein precursor - Neurospora crassa
Contig677	371	2.2e-33	89 415	sp|P78713|VATG_NEUCR VACUOLAR ATP SYNTHASE SUBUNIT G (V-ATPASE 13 KD SUBUNIT)(VACUOLAR H(+)-ATPASE SUBUNIT G) >gb|AAB41886.1| (U84904) V-typeATPase subunit G [Neurospora crassa]
Contig569	370	3.1e-33	214 453	sp|P23704|ATPB_NEUCR ATP SYNTHASE BETA CHAIN, MITOCHONDRIAL PRECURSOR>pir||JC1112 H+-transporting ATP synthase (EC 3.6.1.34) beta chain-Neurospora crassa >emb|CAA37756.1| (X53720) F(1)-

				ATPasebeta-subunit precursor (519 AA) [Neurospora crassa] >gb|AAA33562.1|(M84192) mitochondrial ATPase beta-subunit [Neurospora crassa]
Contig141	285	2.7e-24	66 506	sp|O13349|ATPF_KLULA ATP SYNTHASE SUBUNIT 4, MITOCHONDRIAL PRECURSOR>gb|AAC64860.1| (AF019222) F1Fo-ATP synthase subunit 4[Kluyveromyces lactis]
Contig574	283	5.5e-24	158 433	sp|P49377|ATPG_KLULA ATP SYNTHASE GAMMA CHAIN, MITOCHONDRIAL PRECURSOR>pir||S56153 H+-transporting ATP synthase (EC 3.6.1.34) gamma chainprecursor - yeast (Kluyveromyces marxianus var. lactis)>emb|CAA57355.1| (X81711) gamma subunit of mitochondrial ATPsynthase [Kluyveromyces lactis]
Contig750	279	1.4e-23	200 550	sp|P49377|ATPG_KLULA ATP SYNTHASE GAMMA CHAIN, MITOCHONDRIAL PRECURSOR>pir||S56153 H+-transporting ATP synthase (EC 3.6.1.34) gamma chainprecursor - yeast (Kluyveromyces marxianus var. lactis)>emb|CAA57355.1| (X81711) gamma subunit of mitochondrial ATPsynthase [Kluyveromyces lactis]
Contig518	274	4.5e-23	336 578	pir||T38524 ATP synthase subunitH+-ATPase subunit, vacuolar - fission yeast(Schizosaccharomyces pombe) >emb|CAB16373.1| (Z99259) putativevacuolar ATP synthase subunit C [Schizosaccharomyces pombe]
Contig459	242	1.2e-19	329 478	sp|Q03672|ATP9_PODAN ATP SYNTHASE PROTEIN 9, MITOCHONDRIAL PRECURSOR(LIPID-BINDING PROTEIN) >pir||S17915 H+-transporting ATP synthase(EC 3.6.1.34) lipid-binding protein precursor - Podospora anserina>emb|CAA42471.1| (X59801) ATP synthase subunit 9 [Podosporaanserina]>gb|AAA05756.1| ATP synthase subunit 9 [Podosporaanserina, Peptide, 144 aa]
p2a11fs.f1	220	2.5e-17	310 522	sp|Q36918|ATP6_PENCH ATP SYNTHASE A CHAIN PRECURSOR (PROTEIN 6)>pir||S42271H+-transporting ATP synthase (EC 3.6.1.34) protein 6 - Penicilliumchrysogenum mitochondrion >emb|CAA80613.1| (Z23072) ATPase subunit6 [Penicillium chrysogenum] >gb|AAA58774.2| (L19866) ATPase subunit6 [Penicillium chrysogenum]
Contig612	185	1.4e-13	291 482	gi|6320504 ref|NP_010584.1|ATP5| ATP synthase subunit 5;oligomycinsensitivity-conferring protein; Atp5p >sp|P09457|ATPO_YEAST ATPSYNTHASE OLIGOMYCIN SENSITIVITY CONFERRAL PROTEIN PRECURSOR,MITOCHONDRIAL (OSCP) (ATP SYNTHASE CHAIN 5) >pir||PWBYDH+-transporting ATP synthase (EC 3.6.1.34) delta chain precursor -yeast (Saccharomyces cerevisiae)>emb|CAA30917.1| (X12356) ATPsynthase synthase 5 (AA 1-212)

<**adenylate kinase**-formation of ADP from AMP>

j1e02fs.r1	587	2.6e-56	6	458	gi|6320432 ref|NP_010512.1|ADK1| adenylate kinase; Adk1p >sp|P07170|KAD1_YEASTADENYLATE KINASE CYTOSOLIC (ATP-AMP TRANSPHOSPHORYLASE)>pir||KIBYAadenylate kinase (EC 2.7.4.3) - yeast (Saccharomyces cerevisiae)>emb|CAA29624.1| (X06304) adenylate kinase (AA1-222)[Saccharomyces cerevisiae] >gb|AAA66319.1| (M18455) adenylatekinase [Saccharomyces cerevisiae] >gb|AAC33143.1| (U13239)adenylate kinase
Contig127	227	4.9e-18	359	544	sp|P33075|KAD1_SCHPO ADENYLATE KINASE (ATP-AMP TRANSPHOSPHORYLASE)>pir||S31338 adenylate kinase (EC 2.7.4.3) 1 - fission yeast(Schizosaccharomyces pombe) >pir||A46718 adenylate kinase (EC2.7.4.3) 1 - fission yeast (Schizosaccharomyces pombe)>emb|CAA49826.1| (X70363) adenylate kinase [Schizosaccharomycespombe] >emb|CAA93553.1| (Z69727) adenylate kinase[Schizosaccharomyces pombe]
e4d03fs.r1	218	4.3e-17	297	515	gi|6321018 ref|NP_011097.1|ADK2| Adenylate kinase (mitochondrial GTP:AMPphosphotransferase); Adk2p >sp|P26364|KAD2_YEAST ADENYLATE KINASE 2(ATP-AMP TRANSPHOSPHORYLASE) >pir||S23568 adenylate kinase (EC2.7.4.3) ADK2 - yeast (Saccharomyces cerevisiae) >emb|CAA46254.1|(X65126) adenylate kinase [Saccharomyces cerevisiae]>gb|AAA34418.1| (M77757) adenylate kinase 2 [Saccharomycescerevisiae] >gb|AAB64697.1|
e3g05fs.r1	146	1.9e-09	350	505	gi|6321018 ref|NP_011097.1|ADK2| Adenylate kinase (mitochondrial GTP:AMPphosphotransferase); Adk2p >sp|P26364|KAD2_YEAST ADENYLATE KINASE 2(ATP-AMP TRANSPHOSPHORYLASE) >pir||S23568 adenylate kinase (EC2.7.4.3) ADK2 - yeast (Saccharomyces cerevisiae) >emb|CAA46254.1|(X65126) adenylate kinase [Saccharomyces cerevisiae]>gb|AAA34418.1| (M77757) adenylate kinase 2 [Saccharomycescerevisiae] >gb|AAB64697.1|
8. Alternative respiratory path (2)					
<ALTERNATIVE OXIDASE PRECURSOR>					
p4b01fs.f1	241	1.7e-19	305	523	sp|Q01355|AOX_NEUCR ALTERNATIVE OXIDASE PRECURSOR (ALTOX)>pir||S65752alternative oxidase precursor - Neurospora crassa>gb|AAC37481.1|(L46869) alternative oxidase [Neurospora crassa]
<mitochondrial respiratory function protein>					
g4b09fs.r1	203	5e-15	224	487	sp|Q10488|MRF1_SCHPO MITOCHONDRIAL RESPIRATORY FUNCTION PROTEIN HOMOLOG>pir||T38416 probable mitochondrial respiratory function protein-fission yeast (Schizosaccharomyces pombe) >emb|CAA97361.1|

					(Z73100)putative mitochondrial respiratory function protein[Schizosaccharomyces pombe]
9.Reducing carriers (6)					
9.1. glutaredoxin					
<glutaredoxin>					
j4f07fs.r1	246	4.6e-20	94	396	sp|P55143|GLRX_RICCO GLUTAREDOXIN >pir||S54825 glutaredoxin - castor bean
9.2. gluathione					
<GLUTATHIONE PEROXIDASE>					
14f04fs.r1	496	1.4e-46	68	532	gi|6319721 ref|NP_009803.1|GPX2| Probable glutathione peroxidase (EC1.11.1.9); Gpx2p >sp|P38143|GSHI_YEAST GLUTATHIONE PEROXIDASEHOMOLOG YBR244W >pir||S46121 probable glutathione peroxidase (EC1.11.1.9) - yeast (Saccharomyces cerevisiae) >emb|CAA85207.1|(Z36113) ORF YBR244w [Saccharomyces cerevisiae]
9.3. thioredoxin					
<thioredoxin>					
Contig168	271	1e-22	38	400	sp|P42115|THIO_NEUCR THIOREDOXIN >dbj|BAA08305.1| (D45892) thioredoxin[Neurospora crassa]
o1e07fs.f1	184	1.9e-13	220	534	pir||T40552 thioredoxin-like protein - fission yeast (Schizosaccharomycespombe) >emb|CAB54816.1| (AL1I0506) thioredoxin-like protein[Schizosaccharomyces pombe]
Contig334	101	0.00019	450	593	gb|AAB01771.1| (U42760) thioredoxin homolog [Naegleria fowleri]
<thioredoxin peroxidase PMP20>					
q4f04fs.r1	203	1.5e-15	78	416	gb|AAF04856.1|AF1979 (AF197952) thioredoxin peroxidase PMP20 [Homosapiens]
					Part II. Regulatory Pathways
I. Genetic information Processing					
I.1. DNA replication					
1. DNA synthesis (9)					
<DNA replication initiation protein>					
d3a11fs.f1	134	4.7e-07	278	430	gi|6323132 ref|NP_013204.1|CDC45| Chromosomal DNA replication initiationprotein; Cdc45p >sp|Q08032|CC45_YEAST CELL DIVISION CONTROL PROTEIN45 >pir||S64939 CDC45 protein - yeast (Saccharomyces cerevisiae)>emb|CAA97668.1| (Z73275) ORF YLR103c [Saccharomyces cerevisiae]>gb|AAC49620.1| (U65790) Cdc45p [Saccharomyces cerevisiae]>gb|AAB09053.1| (U56821) Cdc45p [Saccharomyces cerevisiae]

<origin recognition complex subunit 1>

i2a01fs.f1	259	2.6e-20	53	412	sp|O74270|ORC1_CANAL ORIGIN RECOGNITION COMPLEX SUBUNIT 1>emb|CAA76762.1|(Y17395) origin recognition complex 1 protein [Candida albicans]

<Single-stranded DNA-binding protein>

Contig694	146	1.8e-09	228	554	sp|P32445|RIM1_YEAST MITOCHONDRIAL SINGLE-STRANDED DNA-BINDING PROTEIN RIM1PRECURSOR >pir||S23548 RIM1 protein precursor - yeast(Saccharomyces cerevisiae) >gb|AAB22978.1| (S43128) single-strandedDNA binding protein, SSB [Saccharomyces cerevisiae=yeast, PeptideMitochondrial, 135 aa]

<DNA helicase>

d3a10fs.r1	485	2.1e-45	45	509	emb|CAA88537.1| (Z48618) DNA helicase type protein [Saccharomyces cerevisiae]

<topoisomerase ii associated protein>

o3b10fs.r1	200	4.7e-14	14	502	pir||T39841 topisomerase II associated protein pat1 homolog - fission yeast(Schizosaccharomyces pombe) >emb|CAA17064.1| (AL021839)topisomerase yeast pat1 homologue [Schizosaccharomyces pombe]

<SNA41-it is involved in DNA replication>

d3a11fs.r1	364	6.4e-32	29	487	dbj|BAA28947.1| (AB001739) SNA41 [Schizosaccharomyces pombe]<DEAD (aspartate-glutamate-alanine-aspartate) boxpolypeptide 3>
i3a01fs.r1	682	2.9e-66	100	480	gi|6753620 ref|NP_034158.1|| DEAD (aspartate-glutamate-alanine-aspartate) boxpolypeptide 3>sp|Q62167|DDX3_MOUSE DEAD BOX PROTEIN 3 (DEAD-BOXRNA HELICASE DEAD3) (MDEAD3) (EMBRYONIC RNA HELICASE) (D1PAS1RELATED SEQUENCE 2) >pir||I84741 RNA helicase - mouse>gb|AAA53630.1| (L25126) RNA helicase [Mus musculus]>emb|CAA86261.1| (Z38117) dead-box RNA helicase [Mus musculus]>prf||2115205A RNA helicase [Mus musculus]

<MCM initiator complex protein-replication licensing factor>

m2h05fs.f1	272	1.1e-21	187	489	gi|6323304 ref|NP_013376.1|CDC46| MCM initiator complex protein; Cdc46p>sp|P29496|MCM5_YEAST MINICHROMOSOME MAINTENANCE PROTEIN 5 (CELLDIVISION CONTROL PROTEIN 46) >pir||A39631 replication licensingfactor MCM5 - yeast (Saccharomyces cerevisiae) >gb|AAA18027.1|(U09242) Cdc46p [Saccharomyces cerevisiae] >gb|AAB67364.1| (U17245)Cdc46p [Saccharomyces cerevisiae]

2. DNA packaging (16)

2.1. Histone

<Histones, class H1 (or I, or f1)>					
r3h05fs.r1	242	1.2e-19	180	395	emb|CAB72936.1| (AJ011780) histone H1 [Emericella nidulans]
<Histones, class H2a (or IIb1, or f2a2)>					
Contig786	432	9.1e-40	238	552	emb|CAA75581.1| (Y15320) histone H2A [Aspergillus niger]
Contig224	409	2.2e-37	185	484	sp|P48003|H2AV_SCHPO HISTONE H2A VARIANT >pir||S52560 histone H2A variant Pht1- fission yeast (Schizosaccharomyces pombe) >gb|AAB32938.1|(S74633) histone H2A variant [Schizosaccharomyces pombe=fissionyeast, Peptide, 171 aa] >dbj|BAA21378.1| (AB004534) HISTONE H2AVARIANT [Schizosaccharomyces pombe]
<Histones, class H2b (or IIb2, or f2b)>					
Contig972	458	1.4e-42	180	479	sp|P23754|H2B_EMENI HISTONE H2B >pir||S11937 histone H2B - Emericella nidulans>emb|CAA39153.1| (X55547) H2B [Emericella nidulans] >prf||1707275Ahistone H2B [Emericella nidulans]
<Histones, class H4 (or IV, or f2a1)>					
Contig340	409	2.2e-37	200	445	sp|P23750|H41_EMENI HISTONE H4.1 >pir||S11939 histone H4.1 - Emericellanidulans >emb|CAA39155.1| (X55549) H4.1 [Emericella nidulans]>gb|AAA20820.1| (U12630) histone H4.1 [Emericella nidulans]>prf||1707275C histone H4.1 [Emericella nidulans]
Contig705	409	2.2e-37	142	387	sp|P23750|H41_EMENI HISTONE H4.1 >pir||S11939 histone H4.1 - Emericellanidulans >emb|CAA39155.1| (X55549) H4.1 [Emericella nidulans]>gb|AAA20820.1| (U12630) histone H4.1 [Emericella nidulans]>prf||1707275C histone H4.1 [Emericella nidulans]
<histone deacetylase>					
14g11fs.r1	374	1.3e-33	225	509	gi|6323999 ref|NP_014069.1|RPD3| histone deacetylase; Rpd3p>sp|P32561|RPD3_YEAST HISTONE DEACETYLASE RPD3 (TRANSCRIPTIONALREGULATORY PROTEIN RPD3) >pir||S22284 transcription regulator RPD3- yeast (Saccharomyces cerevisiae) >gb|AAB20328.1| (S66438) RPD3[Saccharomyces cerevisiae, Peptide, 433 aa] >emb|CAA58228.1|(X83226) global transcriptional regulator [Saccharomycescerevisiae] >emb|CAA96263.1|
2.2. nonhistone chromosomal protein					
<nonhistone chromosomal protein NHP6B>					
Contig870	290	9.2e-25	157	378	gi|6319565 ref|NP_009647.1|NHP6B| 11-kDa nonhistone chromosomal protein;Nhp6bp >pir||S78076 nonhistone chromosomal protein NHP6B - yeast(Saccharomyces cerevisiae) >emb|CAA85042.1| (Z35959) ORF YBR089c-a[Saccharomyces cerevisiae]

<chromosome condensation regulator protein>

j2f12fs.r1	260	6.1e-21	40	447	pir||T18221 chromosome condensation regulator protein - yeast (Candidaalbicans) >emb|CAA21948.1| (AL033396) regulator of chromosomecondensation [Candida albicans]

<nucleosome assembly protein>

Contig933	648	1e-62	2	604	pir||T41330 nucleosome assembly protein - fission yeast (Schizosaccharomycespombe) >emb|CAA18288.1| (AL022243) nucleosome assembly protein. [Schizosaccharomyces pombe]
Contig258	220	3.5e-17	141	464	dbj|BAA13932.1| (D89271) similar to Saccharomyces cerevisiae nucleosomeassembly protein, SWISS-PROT Accession Number p25293 [Schizosaccharomyces pombe]

<High mobility group-like nuclear protein>

a1h04fs.r1	253	6.3e-21	162	431	gi|6319993 ref|NP_010073.1|NHP2| HMG-like nuclear protein; Nhp2p>sp|P32495|NHP2_YEAST HIGH MOBILITY GROUP-LIKE NUCLEAR PROTEIN 2>pir||S67767 high mobility group-like protein NHP2 - yeast(Saccharomyces cerevisiae) >emb|CAA40885.1| (X57714) high mobilitygroup-like nuclear protein 2 [Saccharomyces cerevisiae]>emb|CAA67483.1| (X99000) high-mobility-group-like protein[Saccharomyces cerevisiae]

<structure recognition/chromatin-associated HMG protein>

Contig467	287	7.5e-24	85	438	pir||T40576 probable structure recognition/chromatin-associated HMG protein -fission yeast (Schizosaccharomyces pombe)>emb|CAA22834.1|(AL035226) similar to yeast POB3 protein that binds to DNApolymerase I; putative structure specific recognition protein[Schizosaccharomyces pombe]

I.2. Transcription

1. RNA Polymerase (4)

<RNA POLYMERASE I, rRNA>

l1h10fs.r1	144	9.3e-08	5	121	pir||T30515 DNA-directed RNA polymerase (EC 2.7.7.6) I 135K chain - Neurosporacrassa >dbj|BAA33445.1| (AB006052) RNA polymerase I second-largestsubunit [Neurospora crassa]

<RNA POLYMERASE II, mRNA>

Contig365	144	2.8e-09	186	401	gi|6321937 ref|NP_012013.1|RPC10| subunit of RNA polymerase II; Rpc10p>sp|P40422|RPCX_YEAST DNA-DIRECTED RNA POLYMERASES I, II, AND III7.7 KD POLYPEPTIDE (ABC10-ALPHA) >pir||S58932 DNA-directed RNApolymerase (EC 2.7.7.6) chain ABC10 alpha - yeast (Saccharomycescerevisiae) >gb|AAA64417.1| (U23378) RNA polymerase I,

					II and IIIsubunit ABC10 alpha [Saccharomyces cerevisiae] >gb|AAB68994.1|(U10397) Rpb12p:
<RNA POLYMERASE III, tRNA>					
p3g02fs.r1	347	8.6e-31	123	521	gi|6325445 ref|NP_015513.1|RPO26| subunit common to RNA polymerases I, II, andIII; Rpo26p >sp|P20435|RPB6_YEAST DNA-DIRECTED RNA POLYMERASES I,II, AND III 23 KD POLYPEPTIDE (ABC23) >pir||RNBYR6 DNA-directed RNApolymerase (EC 2.7.7.6) chain RPO26 - yeast (Saccharomycescerevisiae) >gb|AAA34989.1| (M33924) RNA polymerase II sixthsubunit (RPO26) [Saccharomyces cerevisiae] >emb|CAA37382.1|(X53288) RNA
d3c03fs.f1	162	3.8e-11	77	289	gi|6324215 ref|NP_014286.1|RPC19| subunit common to RNA polymerases I (A) andIII (C); Rpc19p >sp|P28000|RPC9_YEAST DNA-DIRECTED RNA POLYMERASESI AND III 16 KD POLYPEPTIDE (AC19) >pir||A39418 DNA-directed RNApolymerase (EC 2.7.7.6) I/III chain AC19 - yeast (Saccharomycescerevisiae) >gb|AAA34998.1| (M64991) AC19 RNA polymerase subunit[Saccharomyces cerevisiae] >emb|CAA93394.1| (Z69382) Subunit of
2. Regulation (27)					
<araC-family transcription regulator>					
Contig1021	109	0.018	672	941	pir||T36284 probable araC-family transcription regulator - Streptomycescoelicolor >emb|CAB42661.1| (AL049819) putative AraC-familytranscriptional regulator [Streptomyces coelicolor A3(2)]
<transcription factor>					
Contig653	957	2.1e-95	5	676	gb|AAC31206.1| (AF080600) homeodomain DNA-binding transcription factor[Emericella nidulans]
k2d01fs.r1	820	6.9e-81	27	482	pir||S35335 transcription factor MTF-1 - mouse
t4f10fs.f1	305	5.5e-25	63	566	sp|P28349|NIT4_NEUCR NITROGEN ASSIMILATION TRANSCRIPTION FACTOR NIT-4>pir||A41696 regulatory protein nit-4 - Neurospora crassa>gb|AAB21394.1| NIT4=protein product involved in nitrateassimilation [Neurospora intermedia, Peptide, 1090 aa]
Contig259	245	5.5e-20	92	448	pir||T40706 yeast mbf1 homolog, transcription factor - fission yeast(Schizosaccharomyces pombe) >emb|CAB36879.1| (AL035536) yeast mbf1homolog, transcription factor [Schizosaccharomyces pombe]
n4d03fs.r1	125	5.6e-05	177	485	gi|6325241 ref|NP_015309.1|SWI1| Zinc-finger transcription factor; Swi1p>sp|P09547|ADR6_YEAST TRANSCRIPTION REGULATORY PROTEIN ADR6(SWI/SNF COMPLEX COMPONENT ADR6) (REGULATORY PROTEIN SWI1)(REGULATORY PROTEIN GAM3) >pir||TNBYR6 transcription regulator SWI1- yeast (Saccharomyces cerevisiae) >emb|CAA31013.1| (X12493)

					put.transcription factor (AA 1 - 1314) [Saccharomyces cerevisiae]>gb|AAB68089.1|
q2h02fs.r1	104	0.0013	55	399	gi|6324483 ref|NP_014552.1|HAL9| putative transcription factor; contains azinc finger; Hal9p >pir||S57380 probable membrane protein YOL089c-yeast (Saccharomyces cerevisiae) >emb|CAA58190.1| (X83121) orf00938 [Saccharomyces cerevisiae] >emb|CAA99101.1| (Z74831) ORFYOL089c [Saccharomyces cerevisiae]
<transcription regulator>					
q2f04fs.f1	207	7.1e-15	337	486	pir||T38023 probable transcription regulator - fission yeast(Schizosaccharomyces pombe) >emb|CAB11234.1| (Z98598) putativetranscriptional regulator [Schizosaccharomyces pombe]
Contig364	124	0.00014	287	622	pir||T35271 probable transcription regulator - Streptomyces coelicolor>emb|CAB40696.1| (AL049587) putative transcriptional regulator[Streptomyces coelicolor A3(2)]
i3g11fs.r1	98	0.023	228	329	pir||T39677 transcription regulator, binuclear cluster zinc-finger protein -fission yeast (Schizosaccharomyces pombe)>emb|CAA21917.1|(AL033389) putative transcriptional regulator, zinc-finger,binuclear cluster domain containing. [Schizosaccharomyces pombe]
<Cap1-transcription factor>					
k2a02fs.r1	145	2.1e-08	37	354	gb|AAD00802.1| (U95611) Cap1 [Candida albicans] <fork head-related transcription factor>
i2h01fs.f1	94	0.0073	73	168	pir||T42234 fork head-related transcription factor homolog - Caenorhabditiselegans >gb|AAB84392.1| (AF020344) fork head-related transcriptionfactor DAF-16b [Caenorhabditis elegans]
<transcription factor HAP3>					
g4g04fs.r1	170	5.5e-12	248	448	pir||JC6080 transcription factor HAP3 - Emericella nidulans >gb|AAC49411.1|(U35341) HapC [Emericella nidulans]
<transcription initiation factor TFIIF>					
g2a03fs.r1	283	5.2e-24	43	381	gi|6325128 ref|NP_015196.1|ANC1| transcription initiation factor TFIIF smallsubunit; Anc1p >sp|P35189|T2FC_YEAST TRANSCRIPTION INITIATIONFACTOR TFIIF SMALL SUBUNIT (TRANSCRIPTION FACTOR G 30 KD SUBUNIT)(ANC1 PROTEIN) >pir||S38568 transcription initiation factor IIF 30Kchain - yeast (Saccharomyces cerevisiae) >emb|CAA81125.1| (Z26040)Anc1p [Saccharomyces cerevisiae] >gb|AAA61644.1|
<TRANSCRIPTION INITIATION FACTOR TFIID>					
g2h08fs.r1	108	6.9e-05	268	372	sp|Q12731|TF2D_EMENI TRANSCRIPTION INITIATION FACTOR TFIID (TATA-BOX FACTOR)(TATA SEQUENCE-BINDING PROTEIN) (TBP) >gb|AAB57874.1|

					(U28332)TATA-box binding protein [Emericella nidulans]>gb|AAB57876.1|(U28333) TATA-box binding protein [Emericella nidulans]
<HAC1 protein-unfolded protein response pathway, transcrip activation,also see leucine zipper>					
Contig920	98	0.061	680	838	gi|6321078 ref|NP_011156.1|HAC1| bZIP (basic-leucine zipper) protein; Hac1p>pir||S56223 HAC1 protein - yeast (Saccharomyces cerevisiae)
<transcription activator>					
n2d12fs.r1	115	5e-05	251	346	pir||T37604 probable transcription activator - fission yeast(Schizosaccharomyces pombe) >emb|CAB59617.1| (AL132667) putativetranscriptional activator [Schizosaccharomyces pombe]
<TRANSCRIPTIONAL REPRESSOR>					
p1d10fs.r1	585	5.6e-56	3	473	sp|P78706|RCO1_NEUCR TRANSCRIPTIONAL REPRESSOR RCO-1 >gb|AAB37245.1| (U57061)rco-1 gene product [Neurospora crassa]
Contig637	265	3.4e-21	424	597	sp|P78706|RCO1_NEUCR TRANSCRIPTIONAL REPRESSOR RCO-1 >gb|AAB37245.1| (U57061)rco-1 gene product [Neurospora crassa]
<TIP120-stimulates basal transcription>					
l2h07fs.r1	316	4e-26	3	536	dbj|BAA13432.1| (D87671) TIP120 [Rattus norvegicus]
<rho gdp dissociation inhibitor>					
k4d02fs.r1	224	9.6e-18	40	450	pir||T11657 rho GDP dissociation inhibitor. - fission yeast(Schizosaccharomyces pombe) >emb|CAB11090.1| (Z98533) rho gdpdissociation inhibitor. [Schizosaccharomyces pombe]
k4d02fs.f1	137	1.3e-07	300	425	pir||T11657 rho GDP dissociation inhibitor. - fission yeast(Schizosaccharomyces pombe) >emb|CAB11090.1| (Z98533) rho gdpdissociation inhibitor. [Schizosaccharomyces pombe]
<RNA helicase>					
g2b08fs.r1	546	7.5e-52	17	436	gi|6324778 ref|NP_014847.1|DED1| ATP-dependent RNA helicase of DEAD boxfamily; Ded1p >sp|P06634|DED1_YEAST PUTATIVE ATP-DEPENDENT RNAHELICASE DED1 >pir||S13653 ATP-dependent RNA helicase DED1 - yeast(Saccharomyces cerevisiae) >emb|CAA40546.1| (X57278) Ded1p (Spp81p)[Saccharomyces cerevisiae] >emb|CAA99419.1| (Z75110) ORF YOR204w[Saccharomyces cerevisiae]
Contig846	349	5.3e-31	196	450	gi|6320119 ref|NP_010199.1|SUB2| RNA helicase; Sub2p >sp|Q07478|HE47_YEASTPROBABLE ATP-DEPENDENT RNA HELICASE P47 HOMOLOG >pir||S67620hypothetical protein YDL084w - yeast (Saccharomyces cerevisiae)>emb|CAA98650.1| (Z74132) ORF YDL084w [Saccharomyces cerevisiae]

d1f06fs.r1	282	2.3e-23	25	504	sp|Q09747|YB66_SCHPO PUTATIVE ATP-DEPENDENT RNA HELICASE C12C2.06 >pir||T39375probable ATP-dependent RNA helicase - fission yeast(Schizosaccharomyces pombe) >emb|CAA90819.1| (Z54140) probableATP-dependent RNA helicase [Schizosaccharomyces pombe]
h1a02fs.f1	179	7.9e-12	156	515	pir||T39930 probable atp-dependent rna helicase - fission yeast(Schizosaccharomyces pombe) >emb|CAA18864.1| (AL023286) probableatp-dependent rna helicase [Schizosaccharomyces pombe]
<YptA-secretory gene product for transcriptional regulation of the secretory pathway>					
Contig856	906	5.3e-90	135	740	gb|AAF63333.1|AF2445 (AF244545) YptA [Aspergillus niger var. awamori]
3. Processing (14)					
3.1. Spliceosome					
<splicing factor>					
t4a01fs.f1	294	1.1e-38	326	553	pir||T41600 probable pre-mRNA splicing factor - fission yeast(Schizosaccharomyces pombe) >emb|CAA21234.1| (AL031825) rna bindingprotein-putative pre mrna splicing factor [Schizosaccharomycespombe]
<small nuclear ribonucleoprotein>					
Contig559	246	4.2e-20	85	402	pir||T10586 small nuclear ribonucleoprotein-associated protein homologF9F13.90 - Arabidopsis thaliana >emb|CAB45810.1| (AL080253)putative snRNP protein [Arabidopsis thaliana] >emb|CAB79044.1|(AL161553) putative snRNP protein [Arabidopsis thaliana]
<U5 snRNP-specific 200kd protein>					
o1b06fs.r1	548	2.1e-50	10	513	pir||T39188 probable U5 snRNP-specific 200kd protein - fission yeast(Schizosaccharomyces pombe) >emb|CAB57421.1| (AL121764) putative U5snRNP-specific 200kd protein [Schizosaccharomyces pombe]
o1b06fs.f1	299	5.8e-24	172	555	pir||T39188 probable U5 snRNP-specific 200kd protein - fission yeast(Schizosaccharomyces pombe) >emb|CAB57421.1| (AL121764) putative U5snRNP-specific 200kd protein [Schizosaccharomyces pombe]
<RNA12 PROTEIN>					
Contig371	216	1.3e-15	132	770	gi|6323960 ref|NP_014031.1|PRP12| Integral membrane mitochondrial protein;Prp12p >sp|P32843|RN12_YEAST RNA12 PROTEIN >pir||S20462 RNA12protein - yeast (Saccharomyces cerevisiae) >gb|AAB21991.1| (S92205)rna12+=pre-rRNA maturation [Saccharomyces cerevisiae,

					Peptide, 850aa] >emb|CAA89135.1| (Z49212) Rna12p [Saccharomyces cerevisiae]
<RNA splicing protein>					
f3c10fs.r1	232	1.5e-18	9	272	pir||T39149 probable RNA splicing proteinmitochondial carrier protein -fission yeast (Schizosaccharomyces pombe) >emb|CAB16300.1| (Z99168)putative RNA splicing protein [Schizosaccharomyces pombe]
rna12+=pre-rRNAmaturation					
<U3 smallnucleolar ribonucleoprotein protein IMP4>					
h4g08fs.r1	418	2.6e-38	24	395	gi|6324254 ref|NP_014324.1|IMP4| Imp4p >sp|P53941|IMP4_YEAST U3 SMALLNUCLEOLAR RIBONUCLEOPROTEIN PROTEIN IMP4 >pir||S53904 hypotheticalprotein YNL075w - yeast (Saccharomyces cerevisiae)>emb|CAA60184.1|(X86470) unknown [Saccharomyces cerevisiae]>emb|CAA95949.1|(Z71351) ORF YNL075w [Saccharomyces cerevisiae]
3.2. polyA addition					
<poly(A) polymerase>					
m4a12fs.r1	468	1.4e-43	34	495	gi|6322854 ref|NP_012927.1|PAP1| poly(A) polymerase; Pap1p>sp|P29468|PAP_YEAST POLY(A) POLYMERASE (PAP) (POLYNUCLEOTIDEADENYLYLTRANSFERASE) >pir||S19031 poly(A) polymerase - yeast(Saccharomyces cerevisiae) >emb|CAA46250.1| (X65124) polyApolymerase 1 [Saccharomyces cerevisiae] >emb|CAA42852.1| (X60307)poly(A)polymerase [Saccharomyces cerevisiae]
<cleavage and polyadenylation specificity factor>					
h1d08fs.f1	145	8.2e-08	191	466	sp|Q10569|CPSA_BOVIN CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR, 160 KDSUBUNIT (CPSF 160 KD SUBUNIT) >pir||S57335 cleavage andpolyadenylation specificity factor 160K chain - bovine>emb|CAA58152.1| (X83097) cleavage and polyadenylation specificityfactor, 160 kDa subunit [Bos taurus]
i4e01fs.r1	136	3.8e-07	113	448	pir||T39643 probable cleavage and polyadenylation specificity factor - fissionyeast (Schizosaccharomyces pombe) >emb|CAA21254.1| (AL031852)putative cleavage and polyadenylation specificity factor subunit.[Schizosaccharomyces pombe]
3.3. other					
<FIBRILLARIN>					
m2d11fs.f1	229	2.9e-18	289	483	sp|P35551|FBRL_SCHPO FIBRILLARIN >pir||S33690 fibrillarin - fission yeast(Schizosaccharomyces pombe) >emb|CAA49550.1| (X69930) fibrillarin[Schizosaccharomyces pombe] >emb|CAA21168.1| (AL031788)fibrillarin. [Schizosaccharomyces pombe]

<MINOR CAPSID PROTEIN C>

c1h10fs.f1	728	4.1e-71	10	462	sp|P03711|VCAC_LAMBD MINOR CAPSID PROTEIN C (GPC) [CONTAINS: CAPSID ASSEMBLYPROTEIN NU3] >pir||VHBPCL minor capsid protein precursor C - phagelambda >gb|AAA96537.1| (J02459) C (capsid component;439) [bacteriophage lambda]
a4a06fs.r1	633	4.6e-61	63	479	sp|P03711|VCAC_LAMBD MINOR CAPSID PROTEIN C (GPC) [CONTAINS: CAPSID ASSEMBLYPROTEIN NU3] >pir||VHBPCL minor capsid protein precursor C - phagelambda >gb|AAA96537.1| (J02459) C (capsid component;439) [bacteriophage lambda]

<tRNA nucleotidyltransferase>

g3h05fs.r1	222	1.2e-16	34	273	gi|6321016 ref|NP_011095.1|CCA1| tRNA nucleotidyltransferase (tRNACCA-pyrophosphorylase); Cca1p >sp|P21269|CCA1_YEAST TRNANUCLEOTIDYLTRANSFERASE PRECURSOR (TRNA ADENYLYLTRANSFERASE) (TRNACCA-PYROPHOSPHORYLASE) (CCA-ADDING ENZYME) >pir||S11180 tRNAadenylyltransferase (EC 2.7.7.25) - yeast (Saccharomycescerevisiae)>gb|AAA35160.1| (M59870) transfer RNAnucleotidyltransferase [Saccharomyces cerevisiae]

4. tRNA synthetase and ligase (19)

<tRNA synthetase>

Contig390	685	1.4e-66	15	761	gb|AAD21582.1| (AF113612) aspartyl tRNA synthetase [Drosophila melanogaster]
n4c08fs.r1	270	3e-22	36	494	sp|O14018|SYSC_SCHPO SERYL-TRNA SYNTHETASE, CYTOPLASMIC (SERINE--TRNA LIGASE)(SERRS) >pir||T38474 serine--tRNA ligase (EC 6.1.1.11), cytosolic-fission yeast (Schizosaccharomyces pombe) >emb|CAB10149.1| (Z97210)seryl-trna synthetase, cytoplasmic (EC 6.1.1.11) [Schizosaccharomyces pombe]
s2g11fs.f1	237	5.2e-18	254	613	gi|6324082 ref|NP_014152.1|YNL247W| Ynl247wp >sp|P53852|YNY7_YEAST PUTATIVECYSTEINYL-TRNA SYNTHETASE C29E6.06C (CYSTEINE--TRNA LIGASE) (CYSRS)>pir||S63220 probable membrane protein YNL247w - yeast(Saccharomyces cerevisiae) >emb|CAA65497.1| (X96722) ORF N0885[Saccharomyces cerevisiae] >emb|CAA96154.1| (Z71523) ORF YNL247w[Saccharomyces cerevisiae]
a3b09fs.f1	203	4.4e-14	229	477	sp|P10857|SYLC_NEUCR LEUCYL-TRNA SYNTHETASE, CYTOPLASMIC (LEUCINE--TRNALIGASE) (LEURS) >pir||SYNCLC leucine--tRNA ligase (EC 6.1.1.4),cytosolic - Neurospora crassa >gb|AAA33593.1| (M30473) leucyl-tRNAsynthetase [Neurospora crassa]
o2b03fs.f1	150	1.6e-08	290	502	gi|6324911 ref|NP_014980.1|ALA1| Cytoplasmic alanyl-tRNA synthetase gene;Ala1p >sp|P40825|SYAC_YEAST ALANYL-TRNA SYNTHETASE,

					CYTOPLASMIC(ALANINE--TRNA LIGASE) (ALARS) >pir||S62065 alanine–tRNA ligase(EC 6.1.1.7), cytosolic - yeast (Saccharomyces cerevisiae)>emb|CAA89980.1| (Z49821) ALA1 [Saccharomyces cerevisiae]>emb|CAA99658.1| (Z75243) ORF YOR335c [Saccharomyces cerevisiae]
<prolyl-trna synthetase>					
m1b05fs.f1	214	1.4e-15	269	520	pir||T39812 hypothetical protein SPBC19C7.06 - fission yeast(Schizosaccharomyces pombe) >emb|CAA19574.1| (AL023859) putativeprolyl-trna synthetase [Schizosaccharomyces pombe]
<trytophanyl-tRNA synthetase>					
b4e04fs.r1	156	6.8e-10	65	451	gb|AAC42246.1| (AC005395) putative trytophanyl-tRNA synthetase [Arabidopsisthaliana]
<Phenylalanyl-tRNA synthetase>					
Contig527	461	7.8e-43	189	503	gi|6321087 ref|NP_011165.1|FRS2| Phenylalanyl-tRNA synthetase, beta subunit,cytoplasmic; Frs2p >sp|P15625|SYFB_YEAST PHENYLALANYL-TRNASYNTHETASE BETA CHAIN CYTOPLASMIC (PHENYLALANINE–TRNA LIGASE BETACHAIN) >pir||YFBYAC phenylalanine--tRNA ligase (EC 6.1.1.20) betachain, cytosolic - yeast (Saccharomyces cerevisiae)>dbj|BAA09216.1| (D50617) cytoplasmic phenylalanyl-tRNA synthetasebeta chain
<lysyl-tRNA synthetase>					
p2c06fs.r1	459	1.1e-42	8	508	pir||T39726 probable lysyl-trna synthetase - fission yeast(Schizosaccharomyces pombe) >emb|CAB52801.1| (AL109846) putativelysyl-trna synthetase [Schizosaccharomyces pombe]
p2c06fs.f1	413	8.1e-38	225	524	gi|6320242 ref|NP_010322.1|KRS1| lysyl-tRNA synthetase; Krs1p>sp|P15180|SYKC_YEAST LYSYL-TRNA SYNTHETASE, CYTOPLASMIC(LYSINE–TRNA LIGASE) (LYSRS) >pir||SYBYKT lysine--tRNA ligase (EC6.1.1.6) - yeast (Saccharomyces cerevisiae) >gb|AAA66916.1|(J04186) lysyl-tRNA synthetase [Saccharomyces cerevisiae]>emb|CAA92376.1| (Z68196) Krs1p [Saccharomyces cerevisiae]>emb|CAA98863.1| (Z74333) ORF YDR037w
<leucyl-tRNA synthetase>					
d1e08fs.r1	613	4.4e-58	4	498	sp|P10857|SYLC_NEUCR LEUCYL-TRNA SYNTHETASE, CYTOPLASMIC (LEUCINE--TRNALIGASE) (LEURS) >pir||SYNCLC leucine--tRNA ligase (EC 6.1.1.4),cytosolic - Neurospora crassa >gb|AAA33593.1| (M30473)leucyl-tRNAsynthetase [Neurospora crassa]
<Glycyl-tRNA synthase>					

Contig65	522	2.6e-49	249	719	gi|6319597 ref|NP_009679.1|GRS1| Glycyl-tRNA synthase; Grs1p>sp|P38088|SYG_YEAST GLYCYL-TRNA SYNTHETASE (GLYCINE--TRNA LIGASE) (GLYRS) >pir||S48285 probable glycine--tRNA ligase (EC 6.1.1.14) GRS1 - yeast (Saccharomyces cerevisiae) >emb|CAA85078.1| (Z35990) ORF YBR121c [Saccharomyces cerevisiae] >emb|CAA55623.1| (X78993) probable transfer RNA-Gly synthetase [Saccharomyces cerevisiae]
<phenylalanine-tRNA ligase>					
Contig541	186	7.8e-13	109	483	pir||T11642 probable phenylalanine--tRNA ligase (EC 6.1.1.20) beta chain -fission yeast (Schizosaccharomyces pombe) >emb|CAA15915.1|(AL021046) phenylalanyl-trna synthetase beta chain cytoplasmic[Schizosaccharomyces pombe]
<tryptophan tRNA ligase>					
g1c08fs.f1	98	0.034	136	450	emb|CAA36356.1| (X52113) tryptophan tRNA ligase (AA 1-459) [Bos taurus]
<histidine-tRNA ligase precursor>					
d3c12fs.r1	437	2.4e-40	21	518	pir||T40151 histidine-tRNA ligase precursor, mitochondrial - fission yeast(Schizosaccharomyces pombe) >emb|CAA17892.1| (AL022103) histidyl-trna synthetase, mitochondrial precursor[Schizosaccharomyces pombe]
i4g02fs.r1	255	2.7e-20	282	455	pir||T40151 histidine-tRNA ligase precursor, mitochondrial - fission yeast(Schizosaccharomyces pombe) >emb|CAA17892.1| (AL022103) histidyl-trna synthetase, mitochondrial precursor[Schizosaccharomyces pombe]
Contig399	122	7.4e-06	305	481	pir||T40151 histidine-tRNA ligase precursor, mitochondrial - fission yeast(Schizosaccharomyces pombe) >emb|CAA17892.1| (AL022103) histidyl-trna synthetase, mitochondrial precursor[Schizosaccharomyces pombe]
<aminoacyl-tRNAsynthetase>					
m1b05fs.r1	168	9e-11	252	452	pir||T16915 hypothetical protein T20H4.3 - Caenorhabditis elegans>gb|AAA50660.1| (U00037) similar to multifunctional aminoacyl-tRNAsynthetase, especially to the prolyl-tRNA synthetase region[Caenorhabditis elegans]
<THREONYL-TRNA SYNTHETASE>					
n4f06fs.r1	618	1.7e-59	9	494	sp|P87144|SYTC_SCHPO THREONYL-TRNA SYNTHETASE, CYTOPLASMIC (THREONINE--TRNALIGASE) (THRRS) >pir||T39997 Ths1p - fission yeast(Schizosaccharomyces pombe) >emb|CAB08788.1| (Z95397) threonyl-trnasynthetase, cytoplasmic [Schizosaccharomyces pombe]

5. Degradation (6)

<ribonuclease>					
Contig941	921	1.3e-91	189	890	sp|P24657|RNTR_TRIVI RIBONUCLEASE TRV >pir||JX0197 ribonuclease TRV (EC3.1.27.-) - fungus (Trichoderma viride) >gb|AAB21597.1|ribonuclease Trv,RNase Trv [Trichoderma viride, Peptide, 234 aa]
Contig1000	468	1.3e-43	48	440	dbj|BAA31984.1| (AB006460) ribonuclease F1 [Gibberella fujikuroi]
<RNASE L inhibitor>					
r4d08fs.r1	630	1e-60	161	625	pir||T39452 probable RNASE L inhibitor - fission yeast (Schizosaccharomycespombe) >emb|CAA19324.1| (AL023780) putative RNASE L inhibitor[Schizosaccharomyces pombe]
<nuclear RNase P and RNase MRP>					
Contig48	131	1.4e-06	332	526	gi|6324108 ref|NP_014178.1|POP1| Component of nuclear RNase P and RNase MRP;Pop1p >sp|P41812|POP1_YEAST RNASES MRP/P 100.4 KD SUBUNIT (RNAPROCESSING PROTEIN POP1) >pir||A53901 ribonuclease P (EC 3.1.26.5)chain POP1 - yeast (Saccharomyces cerevisiae) >emb|CAA56589.1|(X80358) POP1 [Saccharomyces cerevisiae] >emb|CAA96124.1| (Z71497)ORF YNL221c [Saccharomyces cerevisiae]
<NUCLEASE S1 PRECURSOR>					
o3f06fs.f1	265	4.5e-22	171	530	dbj|BAA08310.1| (D45902) nuclease S1 precursor [Aspergillus oryzae]
<NUCLEASE S1>					
Contig137	262	9e-22	263	514	sp|P24021|NUS1_ASPOR NUCLEASE S1 (ENDONUCLEASE S1) (SINGLE-STRANDED-NUCLEATEENDONUCLEASE) (DEOXYRIBONUCLEASE S1)>gb|AAB20216.1| nuclease S1[Aspergillus oryzae, Peptide, 267 aa]

6. RNA binding (5)

<RNA binding protein>					
t2e12fs.f1	168	7.6e-20	455	532	sp|Q28009|FUS_BOVIN RNA-BINDING PROTEIN FUS/TLS (NUCLEAR ANTIGEN) (PIGPEN)>gb|AAC13543.1| (U26024) pigpen [Bos taurus]
g2e06fs.r1	208	4.3e-16	242	439	pir||T39586 rna binding protein - fission yeast (Schizosaccharomyces pombe)>emb|CAB16904.1| (Z99759) putative poly(a) binding protein[Schizosaccharomyces pombe]
Contig777	211	4.8e-14	118	390	gi|6324324 ref|NP_014394.1|HRB1| hypothetical RNA-binding protein; Hrb1p>sp|P38922|HRB1_YEAST HRB1 PROTEIN (TOM34 PROTEIN)>pir||S45459TOM34 protein - yeast (Saccharomyces cerevisiae)>gb|AAA64803.1|(U02536) Tom34p [Saccharomyces cerevisiae]>emb|CAA54378.1|(X77114) N2009 [Saccharomyces cerevisiae] >emb|CAA95863.1| (Z71280)ORF YNL004w [Saccharomyces cerevisiae]

d2e10fs.f1	190	1.1e-12	17	454	pir||T41389 rna binding protein - fission yeast (Schizosaccharomyces pombe)>emb|CAA19118.1| (AL023592) hypothetical KH domain RNA-bindingprotein [Schizosaccharomyces pombe]
o2h06fs.r1	174	1.5e-12	170	436	emb|CAA63557.1| (X92980) RNA-binding protein [Anabaena variabilis]

I.3. Translation

1. initiation (19)

<RNA recognition motifs>

l1h06fs.r1	180	2.6e-12	294	533	gi|6319835 ref|NP_009916.1|GBP2| Protein with RNA recognition motifs; Gbp2p>sp|P25555|GBP2_YEAST SINGLE-STRAND TELOMERIC DNA-BINDING PROTEINGBP2 (G-STRAND BINDING PROTEIN 2) (RAP1 LOCALIZATION FACTOR 6)>pir||S19338 hypothetical protein YCL011c - yeast (Saccharomycescerevisiae) >emb|CAA42348.1| (X59720) YCL011c,len:427[Saccharomyces cerevisiae]

<EUKARYOTIC TRANSLATION INITIATION>

m3b11fs.r1	544	1e-51	74	538	sp|O14164|IF38_SCHPO PROBABLE EUKARYOTIC TRANSLATION INITIATION FACTOR 3 93 KDSUBUNIT (EIF3 P93) >pir||T38786 translation intiation factor eif-3-fission yeast (Schizosaccharomyces pombe) (fragment)>emb|CAB11485.1| (Z98762) translation intiation factor eif-3[Schizosaccharomyces pombe]
a1e10fs.r1	412	5.4e-37	51	446	sp|Q10425|IF39_SCHPO PROBABLE EUKARYOTIC TRANSLATION INITIATION FACTOR 3 90 KDSUBUNIT (EIF3 P90) >pir||T38379 eukaryotic translation initiationfactor 3 beta subunit - fission yeast (Schizosaccharomyces pombe)>emb|CAA94637.1| (Z70691) eukaryotic translation initiation factor3 beta subunit [Schizosaccharomyces pombe]

<EUKARYOTIC TRANSLATION INITIATION FACTOR>

Contig515	469	1.1e-43	494	826	gi|4503499 ref|NP_001403.1|| eukaryotic translation initiation factor 1A>sp|P47813|IF1A_HUMAN EUKARYOTIC TRANSLATION INITIATION FACTOR 1A(EIF-1A) (EIF-4C) >gb|AAA19812.1| (L18960) protein synthesis factor[Homo sapiens]
s3a12fs.r1	247	3.5e-20	184	546	sp|P78795|IF34_SCHPO PROBABLE EUKARYOTIC TRANSLATION INITIATION FACTOR 3RNA-BINDING SUBUNIT (EIF-3 RNA-BINDING SUBUNIT) (EIF3 P33)(TRANSLATION INITIATION FACTOR EIF3, P33 SUBUNIT) >pir||T39767eukaryotic translation initiation factor 3 rna-binding subunit -fission yeast (Schizosaccharomyces pombe) >emb|CAA18400.1|(AL022304) eukaryotic translation initiation factor 3 rna-bindingsubunit [Schizosaccharomyces pombe]

p3a09fs.r1	256	4.9e-20	2	526	sp|O74760|IF3A_SCHPO PROBABLE EUKARYOTIC TRANSLATION INITIATION FACTOR 3 110KD SUBUNIT (EIF3 P110) (TRANSLATION INITIATION FACTOR EIF3, P110SUBUNIT)>pir||T39716 eukaryotic translation initiation factor 3subunit - fission yeast (Schizosaccharomyces pombe)>emb|CAA21076.1| (AL031739) probable eukaryotic translationinitiation factor 3 110 kd subunit(eif3 p110) [Schizosaccharomycespombe]
Contig591	219	3.9e-16	283	546	sp|Q10425|IF39_SCHPO PROBABLE EUKARYOTIC TRANSLATION INITIATION FACTOR 3 90 KDSUBUNIT (EIF3 P90) >pir||T38379 eukaryotic translation initiationfactor 3 beta subunit - fission yeast (Schizosaccharomyces pombe)>emb|CAA94637.1| (Z70691) eukaryotic translation initiation factor3 beta subunit [Schizosaccharomyces pombe]
<INITIATION FACTOR>					
k4g05fs.r1	545	7.6e-51	19	513	gi|6319282 ref|NP_009365.1|FUN12| 97 kDa protein; Fun12p >sp|P39730|IF2P_YEASTTRANSLATION INITIATION FACTOR IF-2 >pir||S70292 FUN12 protein -yeast (Saccharomyces cerevisiae) >gb|AAC04996.1| (U12980) Fun12p:97kDa protein, function unknown [Saccharomyces cerevisiae]
Contig654	342	3e-30	99	461	sp|P56329|IF2B_SCHPO PROBABLE EUKARYOTIC TRANSLATION INITIATION FACTOR 2 BETASUBUNIT (EIF-2-BETA) >pir||T39024 probable eukaryotic translationinitiation factor 2 beta subunit - fission yeast(Schizosaccharomyces pombe) (fragment) >emb|CAB11076.1| (Z98531)probable eukaryotic translation initiation factor 2 beta subunit[Schizosaccharomyces pombe]
n4c02fs.r1	125	1.7e-05	250	516	pir||T38663 probable translation initiation factor eif-2b beta subunit -fission yeast (Schizosaccharomyces pombe)>emb|CAB52277.1|(AL109739) putative translation initiation factor eif-2b betasubunit [Schizosaccharomyces pombe]
<EUKARYOTIC TRANSLATION INITIATION FACTOR 6>					
Contig701	640	8.2e-62	122	574	sp|O94476|IF6_SCHPO EUKARYOTIC TRANSLATION INITIATION FACTOR 6 (EIF-6)>pir||T41234 eukaryotic translation initiation factor 6 - fissionyeast (Schizosaccharomyces pombe) >emb|CAA22640.1| (AL035075)eukaryotic translation initiation factor 6 [Schizosaccharomycespombe]
<EUKARYOTIC INITIATION FACTOR 4A>					
Contig913	1387	5.3e-141	3	989	sp|P47943|IF4A_SCHPO EUKARYOTIC INITIATION FACTOR 4A (EIF-4A)>pir||S71745translation initiation factor eIF-4A - fission yeast(Schizosaccharomyces pombe) >emb|CAA56772.1| (X80796)

					translationinitiation factor eIF-4A [Schizosaccharomyces pombe]>gb|AAB61679.1| (L40627) cell cycle control protein eIF-4A[Schizosaccharomyces pombe] >emb|CAB60237.1| (AL132828) eukaryoticinitiation factor 4a [Schizosaccharomyces
f3c07fs.f1	272	8.3e-23	214	510	sp|P47943|IF4A_SCHPO EUKARYOTIC INITIATION FACTOR 4A (EIF-4A) >pir||S71745translation initiation factor eIF-4A - fission yeast(Schizosaccharomyces pombe) >emb|CAA56772.1| (X80796) translationinitiation factor eIF-4A [Schizosaccharomyces pombe]>gb|AAB61679.1| (L40627) cell cycle control protein eIF-4A[Schizosaccharomyces pombe] >emb|CAB60237.1| (AL132828) eukaryoticinitiation factor 4a [Schizosaccharomyces
<translation initiation factor 4e>					
i3e08fs.r1	225	7.5e-18	19	480	gb|AAC39871.1| (AF038957) translation initiation factor 4e [Homo sapiens]
<translation release factor subunit 1>					
k3h10fs.r1	529	4.5e-50	135	464	gb|AAC08410.1| (AF053983) translation release factor subunit 1 [Podosporaanserina]
<INITIATION FACTOR 5A>					
o4e01fs.r1	569	2.3e-54	39	485	sp|O94083|IF5A_CANAL INITIATION FACTOR 5A (EIF-5A) (EIF-4D)>gb|AAD10697.1|(U07366) eIF-5A [Candida albicans]
<TRANSLATION INITIATION FACTOR EIF3>					
Contig202	375	2.4e-43	68	349	sp|P79083|IF32_SCHPO EUKARYOTIC TRANSLATION INITIATION FACTOR 3 39 KD SUBUNIT(EIF3 P39) (TRANSLATION INITIATION FACTOR EIF3, P39 SUBUNIT)>pir||T38796 eukaryotic translation initiation factor 3 deltasubunit - fission yeast (Schizosaccharomyces pombe)>emb|CAA70722.1| (Y09529) SUM1 [Schizosaccharomyces pombe]>emb|CAB11277.1| (Z98602) eukaryotic translation initiation factor3 subunit 2 [Schizosaccharomyces pombe]
<nuclear cap binding protein>					
h3d06fs.r1	722	1.6e-70	43	510	sp|P52298|CB20_HUMAN 20 KD NUCLEAR CAP BINDING PROTEIN (NCBP 20 KD SUBUNIT)(CBP20) >pir||I37222 dimeric cap binding protein CBC - human>emb|CAA58962.1| (X84157) subunit of the dimeric cap bindingcomplex CBC [Homo sapiens] >prf||2118330A cap-binding protein[Homo sapiens]
<Tif2p-translation initiation factor>					
f3c07fs.r1	126	2.5e-07	382	498	gb|AAA91645.1| (U25436) Tif2p [Saccharomyces cerevisiae]
2. elongation (17)					
<ELONGATION FACTOR>					

Contig1055	2220	2.8e-229	82	1410	sp|P34825|EF1A_TRIRE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) >pir||S35772translation elongation factor eEF-1 alpha chain - fungus(Trichoderma reesei) >emb|CAA80554.1| (Z23012) translationelongation factor 1a [Hypocrea jecorina] >prf||2004295A elongationfactor 1alpha [Hypocrea jecorina]
Contig749	869	4.1e-86	330	956	sp|P25997|EF3_CANAL ELONGATION FACTOR 3 (EF-3) >emb|CAA77567.1| (Z11484)elongation factor 3 [Candida albicans]
d3a03fs.f1	414	9.5e-37	85	447	pir||T41396 probable translocation elongation factor-Tu fa mily - fissionyeast (Schizosaccharomyces pombe) >emb|CAA19260.1| (AL023704)elongation factor 2-like protein [Schizosaccharomyces pombe]
f3a06fs.r1	384	6.9e-34	19	501	gb|AAD13681.1| (AF035434) elongation factor 3 [Aspergillus fumigatus]
Contig661	374	1.1e-33	191	490	dbj|BAA11572.1| (D82574) elongation factor 1 beta [Schizosaccharomyces pombe]
Contig923	359	4.4e-32	301	525	sp|P34825|EF1A_TRIRE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA)>pir||S35772translation elongation factor eEF-1 alpha chain - fungus(Trichoderma reesei) >emb|CAA80554.1| (Z23012) translationelongation factor 1a [Hypocrea jecorina] >prf||2004295A elongationfactor 1alpha [Hypocrea jecorina]
Contig603	223	1.2e-17	200	466	sp|P34826|EF1B_RABIT ELONGATION FACTOR 1-BETA (EF-1-BETA)>pir||S37087translation elongation factor eEF-1 beta chain - rabbit>emb|CAA52741.1| (X74728) elongation factor 1 beta [Oryctolaguscuniculus]
s3g11fs.r1	148	5.1e-08	115	486	pir||T24472 hypothetical protein T04H1.2 - Caenorhabditis elegans>emb|CAB01578.1| (Z78200) predicted using Genefinder; Weaksimilarity to elongation factors; cDNA EST EMBL:D32732 comes fromthis gene; cDNA EST EMBL:D32745 comes from this gene; cDNA ESTEMBL:D33685 comes from this gene; cDNA EST EMBL:D34113 comes fromthis gene; cDNA>
<ELONGATION FACTOR 1-GAMMA>					
Contig563	580	1.6e-55	205	687	gi|6322769 ref|NP_012842.1|TEF4| Translation elongation factor EF-1gamma;Tef4p >sp|P36008|EF1H_YEAST ELONGATION FACTOR 1-GAMMA 2 (EF-1-GAMMA2) >pir||S37906 translation elongation factor eEF-1 gamma chainhomolog TEF4 - yeast (Saccharomyces cerevisiae) >emb|CAA81919.1|(Z28081) ORF YKL081w [Saccharomyces cerevisiae]
Contig997	390	2.6e-35	32	673	sp|P29694|EF1G_RABIT ELONGATION FACTOR 1-GAMMA (EF-1-GAMMA)>pir||S26649translation elongation factor eEF-1 gamma chain -

					rabbit>emb|CAA48242.1| (X68142) elongation factor 1 gamma [Oryctolaguscuniculus]
<elongation factor 2-yeast>					
Contig1012	1267	2.9e-128	286	1290	pir||T39256 elongation factor 2 - fission yeast (Schizosaccharomyces pombe)(fragment) >emb|CAB52147.1| (AL109734) elongation factor 2[Schizosaccharomyces pombe]
Contig960	981	4.7e-98	7	783	pir||T39256 elongation factor 2 - fission yeast (Schizosaccharomyces pombe)(fragment) >emb|CAB52147.1| (AL109734) elongation factor 2[Schizosaccharomyces pombe]
Contig136	304	9.4e-26	200	502	gb|AAB64821.1| (U28373) Etf1p: Elongation factor 2 (Swiss Prot. accessionnumber P32324). Note that the entire gene is not included in thiscosmid. [Saccharomyces cerevisiae]
<translation elongation factor eEF-3>					
f3a06fs.r1	254	1.1e-19	22	501	pir||S25363 translation elongation factor eEF-3 - yeast (Candida albicans)>emb|CAA78282.1| (Z12822) translation elongation factor 3 [Candidaalbicans]
<GTPase activating protein>					
l2a11fs.f1	409	2.6e-37	232	525	pir||T40517 GTPase activating protein - fission yeast (Schizosaccharomycespombe) >emb|CAA19167.1| (AL023634) TBC domainprotein.[Schizosaccharomyces pombe]
<GTP-binding nuclear protein SPI1>					
Contig437	212	1.9e-16	330	500	sp|P28748|SPI1_SCHPO GTP-BINDING NUCLEAR PROTEIN SPI1 >pir||A40039 gtp-bindingnuclear protein spi1 - fission yeast (Schizosaccharomyces pombe)>gb|AAB25844.1| GTPase=spi1 gene product [Schizosaccharomycespombe,Peptide, 216 aa] >emb|CAB38683.1| (AL035675) gtp-bindingnuclear protein spi1. [Schizosaccharomyces pombe]
<guanine nucleotide releasing factor 1-induces the GTP-binding protein EF-Tu to exchange its bound GDP for GTP>					
b8h04fs.r1	234	8.5e-19	85	492	gb|AAF08011.1| (AF094691) guanine nucleotide releasing factor 1 [Mus musculus]

3. termination
<PEPTIDE CHAIN RELEASE FACTOR>

4. Ribosomal proteins (33)
<ribosomal protein>

Contig828	737	3.9e-72	69 680	gi|6319399 ref|NP_009481.1|RPS8A| Ribosomal protein S8A (S14A) (rp19) (YS9);Rps8ap >gi|6320949 ref|NP_011028.1|RPS8B| Ribosomal protein S8B(S14B) (rp19) (YS9); Rps8bp >sp|P05754|RS8_YEAST 40S RIBOSOMALPROTEIN S8 (S14) (YS9) (RP19) >pir||S45591 ribosomal protein S8.e,cytosolic - yeast (Saccharomyces cerevisiae) >emb|CAA81525.1|(Z26879) ribosomal protein S8 [Saccharomyces cerevisiae]>emb|CAA84893.1|
Contig549	555	8.1e-53	253 633	gi|6321145 ref|NP_011223.1|RPL2A| Ribosomal protein L2A (L5A) (rp8) (YL6);Rpl2ap >gi|6322171 ref|NP_012246.1|RPL2B| Ribosomal protein L2B(L5B) (rp8) (YL6); Rpl2bp >sp|P05736|RL6_YEAST 60S RIBOSOMALPROTEIN L2 (YL6) (L5) (RP8) >pir||S50243 ribosomal protein L8.e,cytosolic - yeast (Saccharomyces cerevisiae) >gb|AAA92283.1|(U17359) ribosomal protein YL6 (L5) [Saccharomyces cerevisiae]>emb|CAA86974.1|
o4b09fs.r1	549	3.3e-52	2 439	pir||T40797 60s ribosomal protein l2 - fission yeast (Schizosaccharomycespombe) >emb|CAA21788.1| (AL032684) 60s ribosomal protein l2[Schizosaccharomyces pombe]
p1b07fs.r1	243	2.2e-42	35 214	gi|6321145 ref|NP_011223.1|RPL2A| Ribosomal protein L2A (L5A) (rp8) (YL6);Rpl2ap >gi|6322171 ref|NP_012246.1|RPL2B| Ribosomal protein L2B(L5B) (rp8) (YL6); Rpl2bp >sp|P05736|RL6_YEAST 60S RIBOSOMALPROTEIN L2 (YL6) (L5) (RP8) >pir||S50243 ribosomal protein L8.e,cytosolic - yeast (Saccharomyces cerevisiae) >gb|AAA92283.1|(U17359) ribosomal protein YL6 (L5) [Saccharomyces cerevisiae]>emb|CAA86974.1|
b1a10fs.r1	436	3.2e-40	127 435	sp|Q09781|RS3A_SCHPO 40S RIBOSOMAL PROTEIN S3AE-A (S1-A) >pir||S62431 40sribosomal protein s3ae (S1) - fission yeast (Schizosaccharomycespombe)>emb|CAA91095.1| (Z54308) 40s ribosomal protein s3ae (S1)[Schizosaccharomyces pombe]
t2e10fs.f1	406	4.7e-37	227 517	gb|AAD54383.1|AF1785 (AF178537) ribosomal protein L13a [Emericellanidulans]
e4f03fs.r1	181	3.4e-13	59 361	sp|Q09727|YA48_SCHPO HYPOTHETICAL 19.0 KD PROTEIN C31A2.08 IN CHROMOSOME I>pir||S58103 hypothetical protein SPAC31A2.08 - fission yeast(Schizosaccharomyces pombe) >pir||T38606 probable ribosomal proteinL23,mitochondrial - fission yeast (Schizosaccharomyces pombe)>emb|CAA90466.1| (Z50113) mitochondrial 60s ribosomal; 123 family[Schizosaccharomyces pombe]
p1a08fs.f1	152	4.4e-10	169 501	sp|P36256|RL1_STRGR 50S RIBOSOMAL PROTEIN L1 >pir||S32236 ribosomal protein L1- Streptomyces griseus >emb|CAA51298.1| (X72787) ribosomal

					proteinL1 [Streptomyces griseus] >dbj|BAA22446.1| (D87846) ribosomalprotein L1 [Streptomyces griseus]
<40S ribosomal protein>					
i4f05fs.r1	347	8.4e-31	77	274	sp|P21772|RS26_NEUCR 40S RIBOSOMAL PROTEIN S26E (CRP5) (13.6 KD RIBOSOMALPROTEIN) >pir||R4NC26 ribosomal protein S26.e - Neurospora crassa>emb|CAA39162.1| (X55637) ribosomal protein [Neurospora crassa]
b1a10fs.f1	335	1.6e-29	196	465	sp|P40910|RS3A_CANAL 40S RIBOSOMAL PROTEIN S3AE (S1) >pir||S49366 ribosomalprotein S0.e.B, cytosolic - yeast (Candida albicans)>emb|CAA57542.1| (X82017) ribosomal protein 10 [Candida albicans]
<60s ribosomal protein l2>					
f1h08fs.r1	189	1.7e-13	142	474	sp|P48535|RM02_KLULA 60S RIBOSOMAL PROTEIN L2, MITOCHONDRIAL PRECURSOR>gb|AAA79723.1| (U38369) 50S subunit ribosomal protein[Kluyveromyces lactis]
<60S ribosomal protein L24>					
s1g06fs.r1	425	4.5e-39	72	500	pir||T39071 60S ribosomal protein L24 - fission yeast (Schizosaccharomycespombe) >emb|CAB03611.1| (Z81317) 60S ribosomal protein L24[Schizosaccharomyces pombe]
<60S ribosomal protein>					
Contig46	835	1.7e-82	148	798	sp|O74836|R10B_SCHPO 60S RIBOSOMAL PROTEIN L1-B (L10A) >pir||T40848 60sribosomal protein l10a - fission yeast (Schizosaccharomyces pombe)>emb|CAA21088.1| (AL031740) 60s ribosomal protein l10a.[Schizosaccharomyces pombe]
Contig706	664	2.3e-64	38	547	sp|Q10157|RL11_SCHPO 60S RIBOSOMAL PROTEIN L11 >pir||T38395 60s ribosomalprotein L11 - fission yeast (Schizosaccharomyces pombe)>pir||T39733 60s ribosomal protein L11 - fission yeast(Schizosaccharomyces pombe)>emb|CAA93230.1| (Z69240) 60s ribosomalprotein L11 [Schizosaccharomyces pombe] >dbj|BAA31552.1| (AB016005)ribosomal protein L11 homolog [Schizosaccharomyces pombe]>emb|CAB52808.1| (AL109846) 60s ribosomal
c1g11fs.f1	319	8.1e-28	201	467	sp|O60143|RL7C_SCHPO 60S RIBOSOMAL PROTEIN L7-C >pir||T39776 60s ribosomalprotein l7-c - fission yeast (Schizosaccharomyces pombe)>emb|CAA18409.1| (AL022304) 60s ribosomal protein l7-c.[Schizosaccharomyces pombe]
Contig386	310	7.8e-27	421	672	sp|O44125|RL37_SCHMA 60S RIBOSOMAL PROTEIN L37 >gb|AAB88508.1| (AF035770)ribosomal protein L37 [Schistosoma mansoni]

j2e02fs.r1	209	4.1e-16	146 409	pir||T39349 probable 60s ribosomal protein - fission yeast(Schizosaccharomyces pombe) >emb|CAA22203.1| (AL034353) putativemitochondrial ribosomal protein [Schizosaccharomyces pombe]
j2e02fs.f1	135	2.4e-08	271 456	pir||T39349 probable 60s ribosomal protein - fission yeast(Schizosaccharomyces pombe) >emb|CAA22203.1| (AL034353) putativemitochondrial ribosomal protein [Schizosaccharomyces pombe]
<60S ribosomal protein L3>				
e2a08fs.r1	620	1e-59	58 486	gb|AAF15600.1|AF1984 (AF198447) 60S ribosomal protein L3 [Emericellanidulans]
e2a08fs.f1	394	8.6e-36	235 510	gb|AAF15600.1|AF1984 (AF198447) 60S ribosomal protein L3 [Emericellanidulans]
<60S RIBOSOMAL PROTEIN L27A>				
l4c07fs.r1	689	4e-67	15 461	sp|P78987|RL2A_ERYGR 60S RIBOSOMAL PROTEIN L27A (L29)>emb|CAA72204.1|(Y11394) 60S ribosomal protein L29 (L27A) [Blumeria graminis f. sp.hordei]
Contig182	496	1.4e-46	220 624	sp|O14388|R27A_SCHPO 60S RIBOSOMAL PROTEIN L27-A >pir||T40638 60s ribosomalprotein l27-a - fission yeast (Schizosaccharomyces pombe)>emb|CAB39364.1| (AL049474) 60s ribosomal protein l27-a.[Schizosaccharomyces pombe]
<Ribosomal protein S4B>				
j1g12fs.f1	475	2.5e-44	187 552	gi|6321997 ref|NP_012073.1|RPS4B| Ribosomal protein S4B (YS6) (rp5) (S7B);Rps4bp >gi|6322605 ref|NP_012679.1|RPS4A| Ribosomal protein S4A(YS6) (rp5) (S7A); Rps4ap >sp|P05753|RS4_YEAST 40S RIBOSOMALPROTEIN S4 (S7) (YS6) (RP5) >pir||S20054 ribosomal protein S4.e,cytosolic - yeast (Saccharomyces cerevisiae) >gb|AAA35011.1|(M64293) ribosomal protein S7 [Saccharomyces cerevisiae]>gb|AAA35012.1| (M64294)
<Ribosomal protein L23A>				
b4d01fs.r1	557	5e-53	60 440	gi|6319384 ref|NP_009466.1|RPL23A| Ribosomal protein L23A (L17aA) (YL32);Rpl23ap >gi|6320963 ref|NP_011042.1|RPL23B| Ribosomal protein L23B(L17aB) (YL32); Rpl23bp >sp|P04451|RL23_YEAST 60S RIBOSOMAL PROTEINL23 (L17) >pir||R5BY17 ribosomal protein L23.e, cytosolic - yeast(Saccharomyces cerevisiae) >emb|CAA25841.1| (X01694) ribosomalprotein L17 [Saccharomyces cerevisiae] >emb|CAA56018.1| (X79489)L23 B
b4d01fs.f1	135	2.8e-08	351 437	gi|6319384 ref|NP_009466.1|RPL23A| Ribosomal protein L23A (L17aA) (YL32);Rpl23ap >gi|6320963 ref|NP_011042.1|RPL23B| Ribosomal protein L23B(L17aB) (YL32); Rpl23bp >sp|P04451|RL23_YEAST 60S RIBOSOMAL

					PROTEINL23 (L17) >pir||R5BY17 ribosomal protein L23.e, cytosolic - yeast(Saccharomyces cerevisiae) >emb|CAA25841.1| (X01694) ribosomalprotein L17 [Saccharomyces cerevisiae] >emb|CAA56018.1| (X79489)L23 B
<Ribosomal protein S18A>					
Contig509	574	7.9e-55	339	749	gi|6320658 ref|NP_010738.1|RPS18A| Ribosomal protein S18A;Rps18ap >gi|6323615ref|NP_013686.1|RPS18B| Ribosomal protein S18B;Rps18bp>sp|P35271|RS18_YEAST 40S RIBOSOMAL PROTEIN S18 >pir||S50886ribosomal protein S18.e, cytosolic - yeast (Saccharomycescerevisiae) >emb|CAA86629.1| (Z46659) 40S ribosomal protein gene,len: 146, CAI: 0.74 [Saccharomyces cerevisiae] >gb|AAB64891.1|(U33007) Ydr450wp
<mitochondrial ribosomal protein>					
Contig253	388	4.1e-35	235	699	emb|CAB51776.1| (AJ243960) mitochondrial ribosomal protein L23 [Kluyveromyceslactis]
g3g10fs.f1	312	4.7e-27	111	509	gi|6319728 ref|NP_009810.1|MRPS5| Probable mitochondrial ribosomal protein S5;Mrps5p >sp|P33759|RT05_YEAST PROBABLE MITOCHONDRIAL 40S RIBOSOMALPROTEIN S5 >pir||S38374 probable ribosomal protein S5,mitochondrial - yeast (Saccharomyces cerevisiae) >gb|AAA65610.1|(L20296) homology with a procaryotic 30S ribosomal protein S5[Saccharomyces cerevisiae]>emb|CAA85214.1| (Z36120) ORF YBR251w[Saccharomyces
a4g01fs.f1	234	2.6e-18	40	318	sp|P23351|RMS5_NEUCR MITOCHONDRIAL RIBOSOMAL PROTEIN S5 >pir||A19079 23S rRNAintron protein - Neurospora crassa mitochondrion >gb|AAA31966.2|(K02658) ribosomal protein S5 (putative); putative [Neurosporacrassa]
m4a04fs.r1	213	1.6e-16	152	430	gi|6319759 ref|NP_009841.1|MRPL27| Mitochondrial ribosomal protein MRPL27(YmL27); Mrpl27p >sp|P36526|RM27_YEAST 60S RIBOSOMAL PROTEIN L27,MITOCHONDRIAL PRECURSOR (YML27) >pir||S27285 ribosomal proteinYmL27 precursor, mitochondrial - yeast (Saccharomyces cerevisiae)>emb|CAA53645.1| (X76053) YBR2019-ORF [Saccharomyces cerevisiae]>emb|CAA85246.1| (Z36151) ORF YBR282w [Saccharomyces
d3d07fs.f1	197	1.2e-14	138	428	gi|6324045 ref|NP_014115.1|MRPL10| Mitochondrial ribosomal protein MRPL10(YmL10); Mrpl10p >sp|P36520|RM10_YEAST 60S RIBOSOMAL PROTEIN L10,MITOCHONDRIAL PRECURSOR (YML10) >pir||S63258 ribosomal protein L15precursor, mitochondrial - yeast (Saccharomyces cerevisiae)>emb|CAA96198.1| (Z71560) ORF YNL284c [Saccharomyces cerevisiae]

Contig219	195	7.5e-14	202	456	gi|6324192 ref|NP_014262.1|NAM9| putative mitochondrial S4 ribosomal protein;Nam9p >sp|P27929|NAM9_YEAST NAM9 PROTEIN PRECURSOR >pir||S55146probable ribosomal protein S4 precursor, mitochondrial - yeast(Saccharomyces cerevisiae) >gb|AAA19439.1| (M60730) NAM9+ protein[Saccharomyces cerevisiae] >emb|CAA86888.1| (Z46843) mitochondrialribosomal protein (putative) [Saccharomyces cerevisiae]>emb|CAA96019.1|
m3f08fs.r1	179	4.2e-12	78	317	gi|6324192 ref|NP_014262.1|NAM9| putative mitochondrial S4 ribosomal protein;Nam9p >sp|P27929|NAM9_YEAST NAM9 PROTEIN PRECURSOR >pir||S55146probable ribosomal protein S4 precursor, mitochondrial - yeast(Saccharomyces cerevisiae) >gb|AAA19439.1| (M60730) NAM9+ protein[Saccharomyces cerevisiae] >emb|CAA86888.1| (Z46843) mitochondrialribosomal protein (putative) [Saccharomyces cerevisiae]>emb|CAA96019.1|

5. Post-translational modifications (14)

5.1. methylation

<SERINE HYDROXYMETHYLTRANSFERASE>

Contig110	684	1.5e-66	88	531	sp|P34898|GLYC_NEUCR SERINE HYDROXYMETHYLTRANSFERASE, CYTOSOLIC (SERINEMETHYLASE) (GLYCINE HYDROXYMETHYLTRANSFERASE) (SHMT)>pir||A42241glycine hydroxymethyltransferase (EC 2.1.2.1), cytosolic-Neurospora crassa >gb|AAA31967.2| (M81918) serinehydroxymethyltransferase [Neurospora crassa]

5.2. myristoylization

<N-myristoyl transferase>

Contig377	191	2.1e-13	279	512	dbj|BAA87865.1| (AB035414) N-myristoyl transferase [Aspergillus fumigatus]

5.3. other

<peptidylprolyl isomerase>

Contig982	862	2.3e-85	173	709	pir||JT0686 peptidylprolyl isomerase (EC 5.2.1.8) a, cytosolic - fungus(Fusarium sporotrichioides)

<cyclophilin>

a1d02fs.r1	386	5.4e-35	116	424	dbj|BAA34384.1| (AB019518) cyclophilin [Trichophyton mentagrophytes]
a1d02fs.f1	247	3.6e-20	308	511	gb|AAD17998.1| (AF107254) cyclophilin B; CYPB [Emericella nidulans]

<acid protease>

Contig1048 563 1.2e-53 248 853 gb|AAF34754.1|AF2218 (AF221843) acid protease [Sclerotinia sclerotiorum]

<protein disulphide isomerase precursor>

a3f04fs.r1 461 7.1e-43 120 473 sp|Q92249|ER38_NEUCR PUTATIVE DISULFIDE ISOMERASE ERP38 PRECURSOR>emb|CAA68847.1| (Y07562) ERp38 [Neurospora crassa]

d2h10fs.r1 394 9.6e-36 255 500 sp|P55059|PDI_HUMIN PROTEIN DISULFIDE ISOMERASE PRECURSOR (PDI) >pir||JC2291protein disulfide-isomerase (EC 5.3.4.1) precursor - Humicolainsolens >gb|AAC60578.1| (S74296) protein disulfide isomerase,PDI{EC 5.3.4.1} [Humicola insolens, KASI, Peptide, 505 aa]>prf||2018168A protein disulfide isomerase [Humicola insolens]

<protein disulphide isomerase>

s1h03fs.f1 373 1.4e-33 170 523 emb|CAA10978.1| (AJ222773) protein disulphide isomerase [Hypocrea jecorina]

<NosA-nostopeptolide biosynthetic gene cluster of Nostoc sp. peptide synthetase>

Contig33 248 3.7e-18 106 597 gb|AAF15891.2|AF2048 (AF204805) NosA [Nostoc sp. GSV224]

<mitochondrial processing peptidase>

Contig565 522 2.6e-49 164 535 sp|P11913|MPPB_NEUCR MITOCHONDRIAL PROCESSING PEPTIDASE BETA SUBUNIT PRECURSOR(BETA-MPP) (UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX CORE PROTEINI)>pir||A29881 mitochondrial processing peptidase (EC 3.4.99.41)beta chain precursor - Neurospora crassa >gb|AAA33606.1| (M20928)processing enhancing protein precursor [Neurospora crassa]

<ModA-responsible for Trimming Asparagine-linked Oligosaccharides on Glycoproteins>

k4g11fs.r1 417 3.3e-38 294 530 gb|AAF24513.1|AF2171 (AF217198) MODA [Emericella nidulans]>gb|AAF24514.1|AF217199_1 (AF217199) MODA [Emericella nidulans]

s1f09fs.f1 154 5.6e-09 256 513 gb|AAB18921.1| (U72236) ModA [Dictyostelium discoideum]

<rehydrin-like protein>

Contig232 223 1.1e-17 260 445 pir||T18224 rehydrin protein homolog - yeast (Candida albicans)>emb|CAA21951.1| (AL033396) rehydrin-like protein [Candidaalbicans]

6. Folding and Targeting (36)

6.1. folding

<CALNEXIN HOMOLOG-folding of glycoproteins>

Contig648 b 659 6.9e-64 19 840 sp|P36581|CALX_SCHPO CALNEXIN HOMOLOG PRECURSOR >pir||A56106 calnexin homologcnx1 - fission yeast (Schizosaccharomyces pombe) >pir||S56142calcium-binding protein precursor cnx1 - fission

					yeast(Schizosaccharomyces pombe) >gb|AAA79757.1| (M98799)calcium-binding protein [Schizosaccharomyces pombe] >gb|AAA68631.1|(U13389) Cnx1p [Schizosaccharomyces pombe] >emb|CAB16741.1|(Z99568) calnexin homolog precursor.
<PEPTIDYL-PROLYL CIS-TRANS ISOMERASE-catalyzes folding>					
Contig134	714	1.1e-69	54	869	sp|P26882|CYP4_BOVIN 40 KD PEPTIDYL-PROLYL CIS-TRANS ISOMERASE (PPIASE)(ROTAMASE) (CYCLOPHILIN-40) (CYP-40) (CYCLOPHILIN-RELATED PROTEIN)(ESTROGEN RECEPTOR BINDING CYCLOPHILIN) >pir||A46579 estrogenreceptor-binding cyclophilin - bovine >dbj|BAA03159.1| (D14074)cyclophilin [Bos taurus]
<FK506-BINDING PROTEIN-protein folding inhibitor>					
Contig902	341	3.8e-30	245	571	sp|O42993|FKBP_SCHPO FK506-BINDING PROTEIN (FKBP) (PEPTIDYL-PROLYL CIS-TRANSISOMERASE) (PPIASE) >pir||T40724 peptidyl-prolyl cis-transisomerase - fission yeast (Schizosaccharomyces pombe)>emb|CAA17020.1| (AL021816) peptidyl-prolyl cis-trans isomerase;fk506-binding protein [Schizosaccharomyces pombe] >emb|CAB46710.1|(AL096796) peptidyl-prolyl cis-trans isomerase; fk506-bindi ngprotein [Schizosaccharomyces pombe]

6.2. chaperones

<chaperone>

Contig131	907	3.5e-90	8	661	emb|CAA70090.1| (Y08867) putative ER chaperone [Aspergillus niger var.awamori] >emb|CAA70091.1| (Y08868) putative ER chaperone[Aspergillus niger]
g3a03fs.r1	440	1.3e-40	131	454	gi|6322524 ref|NP_012598.1|CCT5| subunit of chaperonin subunit epsilon; Cct5p>pir||S57083 t-complex-type molecular chaperone CCT5 - yeast(Saccharomyces cerevisiae) >emb|CAA89592.1| (Z49564) ORF YJR064w[Saccharomyces cerevisiae] >gb|AAB39290.1| (L47993) ORF YJR064w[Saccharomyces cerevisiae]
<prefoldin-chaperone which delivers unfolded proteins to another chaperonin>					
s2a07fs.r1	264	5.7e-22	1	426	pir||T39892 probable prefoldin subunit, molecular chaperone non-native actinbinding complex - fission yeast (Schizosaccharomyces pombe)>emb|CAA22116.1| (AL033534) putative prefoldin subunit; molecularchaperone non-native actin binding complex subunit[Schizosaccharomyces pombe]
<T-COMPLEX PROTEIN-chaperone of actin, tubulin>					
d4e04fs.r1	698	4.6e-68	11	517	pir||T37665 probable t-complex protein 1, epsilon subunit - fission yeast(Schizosaccharomyces pombe) >emb|CAB57321.1| (AL121732)

					probablet-complex protein 1, epsilon subunit [Schizosaccharomyces pombe]
k1e07fs.r1	551	2.2e-52	64	444	pir||T39383 t-complex protein 1 alpha chain homolog - fission yeast(Schizosaccharomyces pombe) >emb|CAA22677.1| (AL035085) t-complexprotein 1, alpha subunit [Schizosaccharomyces pombe]
Contig62	487	1.2e-45	23	526	sp|P47828|TCPQ_CANAL T-COMPLEX PROTEIN 1, THETA SUBUNIT (TCP-1-THETA)(CCT-THETA) >gb|AAC31764.1| (U37371) chaperonin [Candida albicans]
Contig67	226	4.4e-17	235	528	sp|P78921|TCPQ_SCHPO PROBABLE T-COMPLEX PROTEIN 1, THETA SUBUNIT (TCP-1-THETA)(CCT-THETA)
<prohibitin>					
Contig662	492	4.2e-46	151	528	gb|AAB82549.1| (AF022225) prohibitin [Pneumocystis carinii]
Contig660	346	1e-30	247	585	gb|AAB82549.1| (AF022225) prohibitin [Pneumocystis carinii]
c3d09fs.f1	256	3.9e-21	138	446	gi|6321670 ref|NP_011747.1|PHB2| mitochondrial protein, prohibitin homolog;homolog of mammalian BAP37 and S. cerevisiae Phb1p; Phb2p>sp|P50085|YG4W_YEAST HYPOTHETICAL 34.9 KD PROTEIN IN SMI1-PHO81INTERGENIC REGION >pir||S57696 prohibitin PHB2 - yeast(Saccharomyces cerevisiae) >emb|CAA61181.1| (X87941) ORF 315[Saccharomyces cerevisiae] >emb|CAA97259.1| (Z73016) ORF YGR231c[Saccharomyces cerevisiae]
heat_shock_protein					
<heat shock protein Hsp88>					
Contig428	422	3.6e-38	401	748	gb|AAC23862.1| (AF069523) heat shock protein Hsp88 [Neurospora crassa]
<heat shock protein 60>					
n2f06fs.f1	231	1.4e-17	347	508	gb|AAB46362.2| (L11390) heat shock protein 60 [Ajellomyces capsulatus]
<heat shock protein 70>					
Contig731	1376	7.3e-140	3	980	sp|Q01233|HS70_NEUCR HEAT SHOCK 70 KD PROTEIN (HSP70) >gb|AAA82183.1| (U10443)70 kDa heat shock protein [Neurospora crassa]
Contig7	677	7.8e-66	19	468	sp|Q01233|HS70_NEUCR HEAT SHOCK 70 KD PROTEIN (HSP70) >gb|AAA82183.1| (U10443)70 kDa heat shock protein [Neurospora crassa]
m1h04fs.r1	604	4.6e-58	43	456	emb|CAA67431.1| (X98931) heat shock protein 70 [Emericella nidulans]
j2e03fs.r1	554	8.9e-53	4	405	emb|CAA67431.1| (X98931) heat shock protein 70 [Emericella nidulans]
<DnaJ protein-heat shock protein>					

d3b02fs.r1	175	6.2e-12	228	431	gb|AAC95379.1| (AF106835) putative DnaJ [Methylovorus sp. SS1]
<activator of Hsp70 and Hsp90 chaperones>					
Contig449	539	4.1e-51	73	660	pir||T41531 activator of Hsp70 and Hsp90 chaperones - fission yeast(Schizosaccharomyces pombe) >emb|CAB39910.1| (AL049498) activatorof Hsp70 and Hsp90 chaperones [Schizosaccharomyces pombe]
r4b05fs.f1	269	1.1e-21	266	502	pir||T41531 activator of Hsp70 and Hsp90 chaperones - fission yeast(Schizosaccharomyces pombe) >emb|CAB39910.1| (AL049498) activatorof Hsp70 and Hsp90 chaperones [Schizosaccharomyces pombe]
<TPR domain-containing protein-elements in the assembly of the Hsp70-Hsp90 multichaperone machine >					
m2b10fs.r1	339	1.7e-28	12	521	pir||T41230 hypothetical TPR domain-containing protein - fission yeast(Schizosaccharomyces pombe) >emb|CAA22636.1| (AL035075)hypothetical TPR domain protein [Schizosaccharomyces pombe]
<heat-shock protein>					
n2f06fs.r1	714	9.2e-70	16	483	gb|AAD00521.1| (U81786) heat-shock protein [Coccidioides immitis]
<heat shock protein DDR48>					
Contig1045	257	6.3e-21	426	905	gi|6323826 ref|NP_013897.1|DDR48| flocculent specific protein;contains >35repeats of the amino acid sequence NNNDSYGS;Ddr48p>sp|P18899|DR48_YEAST DDR48 STRESS PROTEIN (DNA DAMAGE-RESPONSIVEPROTEIN 48) (DDRP 48) (YP 75) (FLOCCULENT SPECIFIC PROTEIN)>pir||HHBYD8 heat shock protein DDR48 - yeast (Saccharomycescerevisiae) >gb|AAB31954.1| (S73336) FSP=flocculent specificprotein [Saccharomyces
6.3. protein sorting and targeting					
<vacuolar protein sorting>					
Contig106	204	1.3e-15	590	844	gb|AAB61610.1| (AF004837) putative vacuolar protein sorting homolog[Aspergillus fumigatus]
<vacuolar sorting protein>					
i2b09fs.r1	148	1.2e-08	292	411	emb|CAB62829.1| (AL133441) probable vacuolar sorting protein; dynamin family[Schizosaccharomyces pombe] >emb|CAB62830.1| (AL133442) probablevacuolar sorting protein; dynamin family [Schizosaccharomycespombe]
<CARBOXYPEPTIDASE Y-sorting of vacuolar protein>					
Contig571	926	3.8e-92	100	855	sp|P30574|CBPY_CANAL CARBOXYPEPTIDASE Y PRECURSOR (CARBOXYPEPTIDASE YSCY)
Contig576	318	3.8e-27	2	610	sp|P30574|CBPY_CANAL CARBOXYPEPTIDASE Y PRECURSOR (CARBOXYPEPTIDASE YSCY)

Contig761	267	5.6e-21	153	473	pir||T37997 carboxypeptidase y - fission yeast (Schizosaccharomyces pombe)>emb|CAB10121.1| (Z97209) carboxypeptidase y [Schizosaccharomycespombe] >dbj|BAA25568.1| (D86560) carboxypeptidase Y[Schizosaccharomyces pombe]
<ADP-ribosylation factor GTPase-activatingprotein>					
Contig337	239	2.6e-19	341	739	gi|6319975 ref|NP_010055.1|GCS1| ADP-ribosylation factor GTPase-activatingprotein (ARF GAP); Gcs1p >sp|P35197|GCS1_YEAST ZINC FINGER PROTEINGCS1 >pir||S47006 zinc finger protein GCS1 - yeast (Saccharomycescerevisiae) >gb|AAA50389.1| (L24125) zinc finger protein[Saccharomyces cerevisiae] >emb|CAA98805.1| (Z74274) ORF YDL226c[Saccharomyces cerevisiae]
<ADP-RIBOSYLATION FACTOR>					
Contig934	863	1.9e-85	143	667	sp|P34727|ARF_AJECA ADP-RIBOSYLATION FACTOR >pir||D49993 ADP-ribosylationfactor - Ajellomyces capsulata >gb|AAA17548.1| (L25117)ADP-ribosylation factor [Ajellomyces capsulatus]
<COATOMER ZETA SUBUNIT-trafficing to golgi, nonclathrin vesicles>					
n3d02fs.f1	248	3e-20	237	563	sp|O74891|COPZ_SCHPO PROBABLE COATOMER ZETA SUBUNIT (ZETA-COAT PROTEIN)(ZETA-COP) >pir||T41417 coatomer zeta subunit - fission yeast(Schizosaccharomyces pombe) >emb|CAA21186.1| (AL031798) coatomerzeta subunit [Schizosaccharomyces pombe]
<coatomer complex COPI delta-COP subunit>					
Contig689	379	3.6e-34	8	898	gb|AAF14250.1| (AF110234) coatomer complex COPI delta-COP subunit [Drosophilamelanogaster]
Contig188	169	5.1e-11	212	592	gb|AAF14250.1| (AF110234) coatomer complex COPI delta-COP subunit [Drosophilamelanogaster]
<BET3-targeting and fusion of ER to Golgi transport vesicles>					
Contig209	395	7.4e-36	13	483	emb|CAB75995.1| (AL157874) similar to yeast BET3 involved in targeting andfusion of ER to Golgi transport vesicles [Schizosaccharomycespombe]

7. Turnover-protein degradation-including vacuolar (68)

<PROTEASE REGULATORY SUBUNIT>					
r4g07fs.f1	515	1.5e-48	135	527	sp|O14126|PRSA_SCHPO PROBABLE 26S PROTEASE REGULATORY SUBUNIT 6A >pir||T1163426S proteasome regulatory particle chain RPT5 - fission yeast(Schizosaccharomyces pombe) >emb|CAB16387.1| (Z99260) probable 26sprotease regulatory subunit 6a. [Schizosaccharomyces pombe]>dbj|BAA88693.1| (AB012136) regulatory subunit of 26S proteasome[Schizosaccharomyces pombe]

<proteinase>				
b3e06fs.r1	509	4.1e-47	19 459	pir||T39572 probable proteinase subunit - fission yeast (Schizosaccharomycespombe) >emb|CAB38512.1| (AL035637) putative chaperonin; heat shockprotein [Schizosaccharomyces pombe]
i4a09fs.r1	96	0.13	215 487	emb|CAB77335.1| (AL160331) putative proteinase [Streptomyces coelicolor A3(2)]
<proteasome>				
Contig672	795	3.1e-78	135 764	gi|6322459 ref|NP_012533.1|PRE3| Subunit of 20S proteasome;Pre3p>sp|P38624|PRCD_YEAST PROTEASOME COMPONENT PRE3 PRECURSOR(MACROPAIN SUBUNIT PRE3) (PROTEINASE YSCE SUBUNIT PRE3)(MULTICATALYTIC ENDOPEPTIDASE COMPLEX SUBUNIT PRE3) >pir||S61337multicatalytic endopeptidase complex (EC3.4.99.46) chain PRE3 -yeast (Saccharomyces cerevisiae) >emb|CAA89290.1| (Z49276) ORFYJL001w [Saccharomyces cerevisiae]
t4h02fs.f1	568	3.5e-54	125 574	gi|6320197 ref|NP_010277.1|RPT2| (putative) 26S protease subunit; Rpt2p>sp|P40327|PRS4_YEAST 26S PROTEASE REGULATORY SUBUNIT 4 HOMOLOG(TAT-BINDING HOMOLOG 5) >pir||S46613 26S proteasome regulatoryparticle chain RPT2 - yeast (Saccharomyces cerevisiae)>emb|CAA56957.1| (X81070) probable regulatory subunit of 26Sproteasome; homologue to S4 subunit of human 26S proteasome[Saccharomyces cerevisiae]
o2d01fs.f1	300	9.4e-26	155 487	gi|6320054 ref|NP_010134.1|RPN5| Subunit of the regulatory particle of theproteasome; Rpn5p >pir||S67695 26S proteasome regulatory particlechain RPN5 - yeast (Saccharomyces cerevisiae) >emb|CAA66344.1|(X97751) D1572 [Saccharomyces cerevisiae] >emb|CAA98721.1| (Z74195)ORF YDL147w [Saccharomyces cerevisiae]
Contig696	286	2.8e-24	179 472	gi|6319430 ref|NP_009512.1|PRE7| proteasome subunit; Pre7p>sp|P23724|PRC5_YEAST POTENTIAL PROTEASOME COMPONENT C5(MULTICATALYTIC ENDOPEPTIDASE COMPLEX SUBUNIT C5) >pir||S42436multicatalytic endopeptidase complex (EC 3.4.99.46) chain PRS3 -yeast (Saccharomyces cerevisiae)>gb|AAA68908.1| (M34777)proteasome subunit [Saccharomyces cerevisiae]>dbj|BAA00725.1|(D00845) proteasome subunit [Saccharomyces
b4d08fs.f1	135	1e-07	310 426	gi|6321116 ref|NP_011194.1|RPN11| Similar to S. pombe PAD1 gene product;Rpn11p >sp|P43588|MPR1_YEAST MPR1 PROTEIN >pir||S56259 26Sproteasome regulatory particle chain RPN11 - yeast (Saccharomycescerevisiae) >dbj|BAA09243.1| (D50617) YFR004W

					[Saccharomycescerevisiae] >emb|CAA56098.1| (X79561) mpr1 [Saccharomycescerevisiae]
<regulatory particle of the proteasome>					
s1e12fs.f1	440	1.3e-40	110	529	emb|CAB72236.1| (AL138854) 26S proteasome regulatory subunit[Schizosaccharomyces pombe]
Contig533	437	2.4e-40	113	751	emb|CAB72236.1| (AL138854) 26S proteasome regulatory subunit[Schizosaccharomyces pombe]
n4h10fs.r1	255	1.3e-20	183	494	gb|AAF27916.1|AF2201 (AF220199) 26S proteasome regulatory subunit 8 [Pinustaeda]
Contig119	249	3.8e-20	188	523	gi|6320635 ref|NP_010715.1|RPN9| Subunit of the regulatory particle of theproteasome; Rpn9p >pir||S69708 26S proteasome regulatory particlechain RPN9 - yeast (Saccharomyces cerevisiae)>gb|AAB64853.1|(U33007) Ydr427wp; CAI: 0.22 [Saccharomyces cerevisiae]
ubiquitin-eukaryotic protein involved in protein degradation					
<ubiquitin fusion degradation protein>					
Contig666	561	9.7e-53	10	981	gb|AAC80427.1| (AF059906) ubiquitin fusion degradation protein-2[Schizosaccharomyces pombe]
Contig507	301	7.2e-26	165	500	pir||T39543 ubiquitin fusion degradation protein - fission yeast(Schizosaccharomyces pombe) >emb|CAA06721.1| (AJ005824) Ufd1protein [Schizosaccharomyces pombe] >emb|CAA06722.1| (AJ005825)Ufd1 protein [Schizosaccharomyces pombe] >emb|CAB59876.1|(AL021748) ubiquitin fusion degradation protein[Schizosaccharomyces pombe]
<N-END-RECOGNIZING PROTEIN>					
h4b08fs.r1	164	1.3e-09	54	278	sp|O60152|UBR1_SCHPO PROBABLE N-END-RECOGNIZING PROTEIN (UBIQUITIN-PROTEINLIGASE E3 COMPONENT) (N- RECOGNIN) >pir||T39808 hypotheticalprotein SPBC19C7.02 - fission yeast (Schizosaccharomyces pombe)>emb|CAA19570.1| (AL023859) putative ubiquitin protein ligase[Schizosaccharomyces pombe]
<ubiquitin-activating enzyme e1>					
l3c02fs.r1	539	3.9e-51	8	517	pir||T39493 ptr3 or ubiquitin-activating enzyme e1 - fission yeast(Schizosaccharomyces pombe) (fragment) >emb|CAA22354.1| (AL034433)ptr3 or ubiquitin-activating enzyme e1 [Schizosaccharomyces pombe]
<ubiquitin carboxyl-terminal hydrolase>					

t2b01fs.r1	302	9.2e-26	27	362	sp|Q92353|UBPC_SCHPO PUTATIVE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE C6G9.08(UBIQUITIN THIOLESTERASE) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE)(DEUBIQUITINATING ENZYME) >pir||T39070 probable ubiquitincarboxyl-terminal hydrolase - fission yeast (Schizosaccharomycespombe) >emb|CAB03610.1| (Z81317) putative ubiquitincarboxyl-terminal hydrolase [Schizosaccharomyces pombe]
e1d04fs.r1	286	1e-23	22	471	pir||T40815 probable ubiquitin carboxyl-terminal hydrolase - fission yeast(Schizosaccharomyces pombe) >emb|CAA21806.1| (AL032684) possibleubiquitin carboxyl-terminal hydrolase [Schizosaccharomyces pombe]
t2b01fs.f1	245	5.9e-20	230	511	gb|AAC27499.1| (AF077976) putative ubiquitin carboxyl-terminal hydrolase[Schizophyllum commune]
Contig83	233	4.8e-18	177	482	sp|Q92353|UBPC_SCHPO PUTATIVE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE C6G9.08(UBIQUITIN THIOLESTERASE) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE)(DEUBIQUITINATING ENZYME) >pir||T39070 probable ubiquitincarboxyl-terminal hydrolase - fission yeast (Schizosaccharomycespombe) >emb|CAB03610.1| (Z81317) putative ubiquitincarboxyl-terminal hydrolase [Schizosaccharomyces pombe]
<Ubiquitin-specifc processing protease>					
d1f05fs.r1	145	1.5e-08	203	355	emb|CAB54286.1| (Z29095) similar to Ubiquitin-specifc processing protease;cDNA EST CEMSG64F comes from this gene; cDNA EST yk455f9.5 comesfrom this gene [Caenorhabditis elegans]
<ubiquitin ligase>					
c1f08fs.f1	145	2.2e-07	346	456	pir||T37964 probable ubiquitin ligase - fission yeast (Schizosaccharomycespombe) >emb|CAB16714.1| (Z99531) putative ubiquitin ligase[Schizosaccharomyces pombe]
<ubiquitin conjugating enzyme>					
o4a12fs.r1	634	3.5e-61	20	436	gi|4507795 ref|NP_003340.1|| ubiquitin-conjugating enzyme E2 variant 1>gb|AAB72016.1| (U39361) DNA-binding protein [Homo sapiens]
m3a08fs.r1	518	6.9e-49	220	531	emb|CAB75415.1| (AL139314) ubiquitin conjugating enzyme [Schizosaccharomycespombe]
s2f07fs.f1	407	4.2e-37	58	504	gi|4507793 ref|NP_003339.1|| ubiquitin-conjugating enzyme E2N (homologous toyeast UBC13) >sp|Q16781|UBCC_HUMAN UBIQUITIN-CONJUGATING ENZYMEE2-17 KD (UBIQUITIN-PROTEIN LIGASE) (UBIQUITIN CARRIER PROTEIN)(UBC13)>pir||JC4894 ubiquitin--protein ligase (EC 6.3.2.19) E2N -human >dbj|BAA11675.1| (D83004) ubiquitin-conjugating enzyme E2UbcH-ben [Homo sapiens]

i1a03fs.r1	386	6.5e-35	148	411	gb|AAD42941.1|AF0916 (AF091621) ubiquitin-conjugating enzyme E2 [Catharanthus roseus]
j4f03fs.r1	303	5.6e-25	10	495	gb|AAB49301.1| (U84404) E6-associated protein E6-AP/ubiquitin-protein ligase [Homo sapiens] >gb|AAB69154.1| (AF016708) E6-AP ubiquitin-proteinligase [Homo sapiens]
h1h01fs.f1	279	1.5e-23	350	511	gb|AAD00154.1| (U66493) ubiquitin conjugating enzyme [Metarhizium anisopliae]
<ubiquitin fusion protein>					
Contig560	659	7.5e-64	378	761	gb|AAC13689.1| (AF056623) ubiquitin fusion protein [Magnaporthe grisea]
<polyubiquitin>					
Contig1004	1145	2.2e-115	91	792	emb|CAA11267.1| (AJ223328) polyubiquitin [Nicotiana tabacum] >emb|CAA07773.1|(AJ007936) polyubiquitin [Gibberella pulicaris]
<pepsinogen>					
Contig943	959	1.2e-95	179	973	gb|AAA20876.1| (U03278) pepsinogen [Aspergillus niger]
<PALB-cysteine protease>					
i3e01fs.r1	152	8.5e-09	73	477	emb|CAA91013.1| (Z54244) PALB [Emericella nidulans]
<aspartic protease>					
Contig862	401	1.7e-36	37	480	gb|AAB57763.1| (U43775) secreted aspartic proteinase precursor [Glomerellacingulata]
Contig592	397	4.5e-36	116	556	gb|AAB57763.1| (U43775) secreted aspartic proteinase precursor [Glomerellacingulata]
Contig352	206	3.5e-15	158	523	gb|AAD33216.1|AF1153 (AF115320) secreted aspartic protease 2 [Candidatropicalis]
Contig843	158	3.4e-10	218	565	sp|P17576|CARP_POLTU POLYPOROPEPSIN (ASPARTIC PROTEINASE)>pir||PEIKLpolyporopepsin (EC 3.4.23.29) - Irpex lacteus>dbj|BAA00467.1|(D00589) aspartic proteinase precursor [Irpex lacteus]>prf||1512141A Asp protease [Irpex lacteus]
Contig676	137	7.4e-08	239	490	sp|P17576|CARP_POLTU POLYPOROPEPSIN (ASPARTIC PROTEINASE)>pir||PEIKLpolyporopepsin (EC 3.4.23.29) - Irpex lacteus>dbj|BAA00467.1|(D00589) aspartic proteinase precursor [Irpex lacteus]>prf||1512141A Asp protease [Irpex lacteus]
Contig353	107	0.0097	380	526	gi|6323149 ref|NP_013221.1|YPS1| Yps1p >sp|P32329|YAP3_YEAST ASPARTICPROTEINASE 3 PRECURSOR (YAPSIN 1) >pir||S64957 aspergillopepsin I(EC3.4.23.18) YAP3 precursor - yeast (Saccharomyces cerevisiae)>gb|AAB82367.1| (U53877) Yap3p: aspartic proteinase [Saccharomycescerevisiae]>emb|CAA61699.1| (X89514)

					Aspartyl protease[Saccharomyces cerevisiae]>emb|CAA97688.1| (Z73292) ORF YLR120c[Saccharomyces
<aminopeptidase>					
Contig235	262	4.7e-21	176	484	gi|6322746 ref|NP_012819.1|LAP4| vacuolar aminopeptidase ysc1; Lap4p>sp|P14904|AMPL_YEAST VACUOLAR AMINOPEPTIDASE I PRECURSOR(POLYPEPTIDASE) (LEUCINE AMINOPEPTIDASE IV) (LAPIV) (AMINOPEPTIDASEIII) (AMINOPEPTIDASE YSCI) >pir||A33879 aminopeptidase yscI (EC3.4.11.-) precursor, vacuolar - yeast (Saccharomyces cerevisiae)>gb|AAA34738.1| (M25548) aminopeptidase I [Saccharomycescerevisiae] >emb|CAA50454.1|
n3c06fs.f1	241	9.5e-19	278	541	pir||A42209 D-stereospecific aminopeptidase (EC 3.4.11.19) - Ochrobactrumanthropi >gb|AAA25519.1| (M84523) D-aminopeptidase [Ochrobactrumanthropi]
<iminopeptidase>					
Contig1017	392	1.6e-35	125	952	sp|P46542|PIP_LACDL PROLINE IMINOPEPTIDASE (PROLYL AMINOPEPTIDASE)>pir||A59087 prolyl aminopeptidase (EC 3.4.11.5) - Lactobacillusdelbrueckii (strain DSM7290) >emb|CAA81556.1| (Z26948) prolineiminopeptidase [Lactobacillus delbrueckii]
<MepB-encodes an 82 kDa intracellular metalloproteinase structurally related to mammalian thimet oligopeptidases>					
Contig144	323	2.6e-27	127	513	gb|AAB66656.1| (U85769) MepB [Aspergillus fumigatus]
Contig143	225	9.6e-17	282	542	gb|AAB66656.1| (U85769) MepB [Aspergillus fumigatus]
<methionine metallopeptidase>					
Contig700	512	3.1e-48	281	880	pir||T39431 probable methionine metallopeptidase - fission yeast(Schizosaccharomyces pombe) >emb|CAA18421.1| (AL022305) putativemethionine metallopeptidase [Schizosaccharomyces pombe]
i2c10fs.f1	375	1e-33	172	462	pir||T39431 probable methionine metallopeptidase - fission yeast(Schizosaccharomyces pombe) >emb|CAA18421.1| (AL022305) putativemethionine metallopeptidase [Schizosaccharomyces pombe]
s1g09fs.f1	194	7.1e-14	235	495	pir||T39431 probable methionine metallopeptidase - fission yeast(Schizosaccharomyces pombe) >emb|CAA18421.1| (AL022305) putativemethionine metallopeptidase [Schizosaccharomyces pombe]
<proteinase T precursor>					
j2g02fs.f1	386	5.7e-35	125	481	sp|P20015|PRTT_TRIAL PROTEINASE T PRECURSOR >pir||JQ0380 proteinase T (EC3.4.21.-) - imperfect fungus (Tritirachium album (fragment)>gb|AAA34204.1| (M54900) proteinase [Tritirachium album]>gb|AAA34205.1| (M54901) proteinase [Tritirachium album]
<Ca dependent protease>					

m2d05fs.r1	130	1.8e-06	330	488	prf||1613155A Ca dependent Cys protease p94 [Rattus norvegicus]
<Subtilisin-like protease PR1H>					
Contig881	1386	6.6e-141	99	1223	emb|CAB63907.1| (AJ251921) Subtilisin-like protease PR1H [Metarhiziumanisopliae var. anisopliae]
Contig958	329	1.5e-28	236	583	emb|CAB63913.1| (AJ251965) subtilisin-like protease PR1H [Metarhiziumanisopliae var. anisopliae]
<L-KYNURENINE HYDROLASE>					
a2e05fs.f1	172	2.2e-11	141	524	gi|4504937 ref|NP_003928.1|| kynureninase >sp|Q16719|KYNU_HUMAN KYNURENINASE(L-KYNURENINE HYDROLASE) >pir||G02652 kynureninase (EC 3.7.1.3)-human >gb|AAC50650.1| (U57721) L-kynurenine hydrolase [Homosapiens]
<microsomal dipeptidase precursor>					
Contig598	379	4e-34	52	699	sp|P22412|MDP1_PIG MICROSOMAL DIPEPTIDASE PRECURSOR (MDP *(DEHYDROPEPTIDASE-I)(RENAL DIPEPTIDASE)* (RDP) >pir||JS0759 membrane dipeptidase (EC3.4.13.19) precursor - pig >emb|CAA37762.1| (X53730) dipeptidase[Sus scrofa] >dbj|BAA02433.1| (D13142) dipeptidase precursor [Susscrofa]
<aspartylproteinase>					
Contig623	172	2.6e-11	13	396	gi|6322230 ref|NP_012305.1|YPS6| Yps6p >sp|P40583|YIV9_YEAST PUTATIVE ASPARTYLPROTEINASE YIR039C PRECURSOR >pir||S50344 aspergillopepsin homologYIR039c - yeast (Saccharomyces cerevisiae)
<carboxypeptidase S1>					
Contig850	420	1e-69	105	575	sp|P34946|CPS1_PENJA CARBOXYPEPTIDASE S1 >pir||S38953 carboxypeptidase D (EC3.4.16.6) - Penicillium janthinellum >gb|AAB28596.1|carboxypeptidase S1,CPD-S1 [Penicillium janthinellum, Peptide, 423aa] >prf||1923269A carboxypeptidase S1 [Penicillium janthinellum]
Contig944	480	7.5e-45	204	542	sp|P34946|CPS1_PENJA CARBOXYPEPTIDASE S1 >pir||S38953 carboxypeptidase D (EC3.4.16.6) - Penicillium janthinellum >gb|AAB28596.1|carboxypeptidase S1,CPD-S1 [Penicillium janthinellum, Peptide, 423aa] >prf||1923269A carboxypeptidase S1 [Penicillium janthinellum]
<carboxypeptidase s precursor>					
g3g05fs.r1	217	5.2e-16	117	446	pir||T38349 carboxypeptidase s precursor - fission yeast (Schizosaccharomycespombe) >emb|CAB11265.1| (Z98601) carboxypeptidase s precursor[Schizosaccharomyces pombe]
<serine-type carboxypeptidase>					

Contig1052	846	1.3e-83	128	1597	sp|P52718|PEPF_ASPNG SERINE-TYPE CARBOXYPEPTIDASE F PRECURSOR (PROTEINASE F)(CPD-II) >emb|CAA56075.1| (X79541) serine carboxypeptidase[Aspergillus niger]
Contig978	565	7.4e-54	301	1221	pir||S78072 serine-type carboxypeptidase (EC 3.4.16.-) I - Aspergillus niger
g4c09fs.r1	400	2.2e-36	90	494	pir||S55328 serine-type carboxypeptidase (EC 3.4.16.1) precursor - Aspergillusphoenicis
Contig573	295	8.7e-25	192	482	gi|6319615 ref|NP_009697.1|YBR139W| Probable serine-type carboxypeptidase (EC3.4.16.1); Ybr139wp >sp|P38109|YBY9_YEAST PUTATIVE SERINECARBOXYPEPTIDASE IN ESR1-IRA1 INTERGENIC REGION >pir||S46008probable carboxypeptidase C (EC 3.4.16.5) YBR139w - yeast(Saccharomyces cerevisiae)>emb|CAA53497.1| (X75891) YBR1015[Saccharomyces cerevisiae] >emb|CAA85097.1| (Z36008) ORF YBR139w[Saccharomyces cerevisiae]
Contig989	265	1.5e-21	182	544	pir||S78072 serine-type carboxypeptidase (EC 3.4.16.-) I - Aspergillus niger
Contig950	192	1.9e-13	331	735	gi|6319615 ref|NP_009697.1|YBR139W| Probable serine-type carboxypeptidase (EC3.4.16.1); Ybr139wp >sp|P38109|YBY9_YEAST PUTATIVE SERINECARBOXYPEPTIDASE IN ESR1-IRA1 INTERGENIC REGION >pir||S46008probable carboxypeptidase C (EC 3.4.16.5) YBR139w - yeast(Saccharomyces cerevisiae)>emb|CAA53497.1| (X75891) YBR1015[Saccharomyces cerevisiae] >emb|CAA85097.1| (Z36008) ORF YBR139w[Saccharomyces cerevisiae]
<3-hydroxybutyryl-CoA dehydrogenase-degradation of the branched-chain amino acids>					
f3f09fs.f1	337	1.1e-29	214	543	sp|P45856|MMGB_BACSU PROBABLE 3-HYDROXYBUTYRYL-COA DEHYDROGENASE(BETA-HYDROXYBUTYRYL-COA DEHYDROGENASE) (BHBD)>gb|AAB09614.1|(U29084) similar to Clostridium acetobutylicum NAD-dependentbeta-hydroxybutyrl-CoA dehydrogenase, PIR Accession Number A43723[Bacillus subtilis]
<3-hydroxyacyl-CoA dehydrogenase>					
f3f09fs.r1	211	3.1e-15	136	510	pir||C75389 enoyl-CoA hydratase/3,2-trans-enoyl-CoAisomerase/3-hydroxyacyl-CoA dehydrogenase - Deinococcus radiodurans(strain R1) >gb|AAF11052.1|AE001993_2 (AE001993) enoyl-CoAhydratase/3,2-trans-enoyl-CoA isomerase/3-hydroxyacyl-CoAdehydrogenase [Deinococcus radiodurans]
j4c03fs.r1	166	1.7e-10	123	503	pir||E69285 3-hydroxyacyl-CoA dehydrogenase (hbd-2) homolog - Archaeoglobusfulgidus >gb|AAB90948.1| (AE001085) 3-hydroxyacyl-CoA dehydrogenase(hbd-2) [Archaeoglobus fulgidus]

<PEPTIDASE>					
Contig576	318	3.8e-27	2	610	sp|P30574|CBPY_CANAL CARBOXYPEPTIDASE Y PRECURSOR (CARBOXYPEPTIDASE YSCY)
Contig602	189	4.1e-13	264	758	gi|6321118 ref|NP_011196.1|YFR006W| Yfr006wp >sp|P43590|YFH6_YEASTHYPOTHETICAL 61.8 KD PEPTIDASE IN MPR1-GCN20 INTERGENIC REGION>pir||S56261 probable membrane protein YFR006w - yeast(Saccharomyces cerevisiae) >dbj|BAA09245.1| (D50617) YFR006W[Saccharomyces cerevisiae]
<dipeptidase>					
i3h10fs.r1	229	8.3e-18	153	470	pir||T41665 probable dipeptidase - fission yeast (Schizosaccharomyces pombe)>emb|CAA19072.1| (AL023590) putative dipeptidase[Schizosaccharomyces pombe]
8. protein binding (19)					
<ACYL-COA-BINDING PROTEIN HOMOLOG>					
Contig556	178	6.7e-13	93	368	sp|P31824|ACBP_MANSE ACYL-COA-BINDING PROTEIN HOMOLOG (ACBP) (DIAZEPAM BINDINGINHIBITOR HOMOLOG) (DBI) >gb|AAA29309.1| (L11449) diazepam bindinginhibitor-like peptide [Manduca sexta] >gb|AAB28237.2| (S65642)diazepam binding inhibitor; DBI [Manduca sexta]
<oxysterol-binding protein>					
i3h11fs.r1	571	1.7e-54	95	460	emb|CAA73224.1| (Y12693) oxysterol-binding protein [Neurospora crassa]
<amiloride-binding protein>					
Contig225	146	2.9e-08	26	304	gb|AAA58358.1| (M55602) amiloride-binding protein [Homo sapiens]
<methyl-CpG binding protein MBD4>					
a2g06fs.r1	154	3.1e-09	133	429	gi|4505121 ref|NP_003916.1|| methyl-CpG binding domain protein 4>gb|AAC68879.1| (AF072250) methyl-CpG binding protein MBD4 [Homosapiens]>gb|AAD22195.1|AF114784_1 (AF114784) methyl-CpG bindingendonuclease [Homo sapiens] >gb|AAD50374.1| (AF120999) methyl-CpGbinding protein 4 [Homo sapiens]
saccharide-binding protein					
<a-agglutinin attachment subunit precursor>					
Contig362	103	0.0011	38	361	gi|6324372 ref|NP_014442.1|AGA1| anchorage subunit of a-agglutinin; Aga1p>sp|P32323|AGA1_YEAST A-AGGLUTININ ATTACHMENT SUBUNIT PRECURSOR>pir||A41258 a-agglutinin core protein AGA1 - yeast (Saccharomycescerevisiae) >gb|AAA34382.1| (M60590) a-agglutinin core

					subunit[Saccharomyces cerevisiae] >emb|CAA96325.1| (Z71659) ORF YNR044w[Saccharomyces cerevisiae]
<lectin precursor>					
n2d02fs.r1	330	5.4e-29	14	505	pir||T40912 probable lectin precursor - fission yeast (Schizosaccharomycespombe) >emb|CAA22477.1| (AL034490) putative lectin precursor;possible vesicular protein [Schizosaccharomyces pombe]
DNA-binding proteins					
<CURVED DNA-BINDING PROTEIN>					
i1b05fs.r1	291	7.1e-25	100	435	sp|Q09184|CDB4_SCHPO CURVED DNA-BINDING PROTEIN (42 KD PROTEIN)>pir||S46583442K curved dna-binding protein - fission yeast(Schizosaccharomyces pombe) >dbj|BAA03607.1| (D14907) 42K-protein[Schizosaccharomyces pombe] >emb|CAB11663.1| (Z98977) curveddna-binding protein [Schizosaccharomyces pombe]
<TGACG-motif binding protein>					
Contig751	108	0.0087	9	383	pir||T08591 TGACG-motif binding protein STF1 - soybean >gb|AAC05017.1|(L28003) TGACG-motif binding factor [Glycine max]
<DNA-binding protein HEXBP-specific DNA-binding protein>					
Contig395	344	1.6e-30	111	677	sp|Q04832|HEXP_LEIMA DNA-BINDING PROTEIN HEXBP (HEXAMER-BINDING PROTEIN)>pir||A47156 hexamer-binding protein HEXBP - Leishmania major>gb|AAA29245.1| (M94390) HEXBP DNA binding protein [Leishmaniamajor]
Zinc finger motif-DNA binding					
<zinc-finger transcription factor>					
p3b09fs.r1	207	1.1e-14	5	517	pir||T39608 inc finger transcription factor - fission yeast(Schizosaccharomyces pombe) >emb|CAA19035.1| (AL023554) fungalZn(2)-Cys(6) binuclear cluster zinc finger transcription factor[Schizosaccharomyces pombe]
<zinc finger protein>					
m1e04fs.r1	285	1.5e-23	106	441	gi|6319716 ref|NP_009798.1|YBR239C| Probable Zn-finger protein; Ybr239cp>sp|P38140|YB89_YEAST PUTATIVE 60.3 KD TRANSCRIPTIONAL REGULATORYPROTEIN IN PRP5-THI2 INTERGENIC REGION >pir||S46116 probableregulatory protein YBR239c - yeast (Saccharomyces cerevisiae)>emb|CAA85202.1| (Z36108) ORF YBR239c [Saccharomyces cerevisiae]
Contig936	193	7.4e-14	178	573	gb|AAD18121.1| (AC006403) putative C2H2-type zinc finger protein [Arabidopsisthaliana] >gb|AAD25324.1|AF095588_1 (AF095588) C2H2 zinc fingerprotein FZF [Arabidopsis thaliana]

d1a06fs.r1	179	7.3e-13	3	119	gb|AAD22484.1|AF1244 (AF124404) C2H2-type zinc finger protein Mhy1p[Yarrowia lipolytica]
d1a06fs.r1	179	7.3e-13	3	119	gb|AAD22484.1|AF1244 (AF124404) C2H2-type zinc finger protein Mhy1p[Yarrowia lipolytica]
d1a06fs.r1	179	7.3e-13	3	119	gb|AAD22484.1|AF1244 (AF124404) C2H2-type zinc finger protein Mhy1p[Yarrowia lipolytica]
m4g09fs.f1	136	2.2e-08	267	413	gb|AAF35997.1| (AC024200) contains similarity to several zinc finger proteinsbut not to the zinc finger domains [Caenorhabditis elegans]
g4b12fs.r1	135	2.2e-07	37	138	pir||T41718 hypothetical fungal Zn(2)-Cys(6) zinc-finger protein - fissionyeast (Schizosaccharomyces pombe) >emb|CAB57441.1| (AL121770)hypothetical fungal Zn(2)-Cys(6) zinc-finger protein[Schizosaccharomyces pombe]
r3h02fs.r1	92	0.66	92	169	pir||T39478 zinc-finger protein - fission yeast (Schizosaccharomyces pombe)>emb|CAA20477.1| (AL031349) hypothetical zinc-finger protein[Schizosaccharomyces pombe]
<zinc finger suppressor>					
o4e06fs.r1	748	3.1e-73	12	428	gb|AAD05020.1| (U56732) KRAB/zinc finger suppressor protein 1 [Rattusnorvegicus]

II. Cell Growth, Cell Division, Mating and Morphogenesis

II.1. Cell walls, Biomembranes and Cytoskeleton

1. Cell walls (17)

<cell wall protein>					
n1f05fs.f1	222	4e-17	179	511	gi|6323964 ref|NP_014035.1|SCW10| soluble cell wall protein; Scw10p>sp|Q04951|YM8Z_YEAST HYPOTHETICAL 40.5 KD PROTEIN IN UBP15-GAS1INTERGENIC REGION PRECURSOR >pir||S53975 probable membrane proteinYMR305c - yeast (Saccharomyces cerevisiae) >emb|CAA89138.1|(Z49212) unknown [Saccharomyces cerevisiae]
<cell wall biogenesis protein>					
Contig1040	324	3.6e-28	248	1276	pir||T38309 probable cell wall biogenesis protein - fission yeast(Schizosaccharomyces pombe) (fragment) >emb|CAB11673.1| (Z98977)putative cell wall biogenesis protein [Schizosaccharomyces pombe]
<QI74 protein-cell wall protein of Trichoderma harzianum>					
Contig1039	173	3.4e-09	751	1014	emb|CAA64974.1| (X95671) QI74 protein [Trichoderma harzianum]
<Psu1 protein-cell wall synthesis protein psu1>					

Contig919	402	1.3e-36	606	1106	pir||JC7092 Psu1 protein - fission yeast (Schizosaccharomyces pombe)>dbj|BAA83907.1| (AB009980) PSU1 [Schizosaccharomyces pombe]>emb|CAB65613.1| (AL136078) cell wall synthesis protein psu1[Schizosaccharomyces pombe]
<GEL1 protein-Biosynthesis of the Fungal Cell Wall>					
Contig973	611	9.7e-59	544	1161	gb|AAC35942.1| (AF072700) GEL1 protein [Aspergillus fumigatus]
Contig896	195	6.2e-14	339	533	gb|AAC35942.1| (AF072700) GEL1 protein [Aspergillus fumigatus]
<septin>					
Contig1027	630	8.9e-61	10	531	emb|CAB61437.1| (AJ251017) putative septin [Mucor circinelloides]
d1b07fs.r1	471	5.6e-44	217	498	gb|AAD21037.1| (AF111181) G-septin gamma [Rattus norvegicus]
g3a06fs.r1	354	1.3e-31	176	454	emb|CAB61437.1| (AJ251017) putative septin [Mucor circinelloides]
b1a07fs.r1	138	7.8e-08	39	140	gb|AAB41233.1| (U83489) septin B [Emericella nidulans]
<adhesion regulating molecule ARM-1>					
Contig345	91	0.2	281	427	gb|AAF33401.1|AF2259 (AF225959) adhesion regulating molecule ARM-1 [Musmusculus]
<mycelial surface antigen CSA1 precursor-cell wall protein>					
d2h05fs.r1	149	2.1e-08	116	457	pir||T17415 mycelial surface antigen CSA1 precursor - yeast (Candida albicans)>gb|AAC29486.1| (AF080221) mycelial surface antigen precursor[Candida albicans]
<Glutamine--fructose-6-phosphate amidotransferase>					
s3f06fs.r1	433	5.9e-50	14	358	sp|P53704|GFA1_CANAL GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE[ISOMERIZING] (HEXOSEPHOSPHATE AMINOTRANSFERASE)(D-FRUCTOSE-6-PHOSPHATE AMIDOTRANSFERASE) (GFAT)>pir||JC6012glutamine--fructose-6-phosphate transaminase (isomerizing) (EC2.6.1.16) - yeast (Candida albicans) >emb|CAA64380.1| (X94753)glucosamine--fructose-6-phosphate aminotransferase (isomerizing)[Candida albicans]
Contig795	447	5.3e-41	247	642	sp|Q09740|GFA1_SCHPO PUTATIVE GLUCOSAMINE--FRUCTOSE-6-PHOSPHATEAMINOTRANSFERASE [ISOMERIZING] (HEXOSEPHOSPHATE AMINOTRANSFERASE)(D-FRUCTOSE-6-PHOSPHATE AMIDOTRANSFERASE) (GFAT) >pir||T11674glutamine--fructose-6-phosphate transaminase (isomerizing) (EC2.6.1.16) - fission yeast (Schizosaccharomyces pombe)>pir||T39370glucosamine--fructose-6-phosphate aminotransferase - fission yeast(Schizosaccharomyces pombe)
n1a08fs.r1	412	3.8e-37	23	484	sp|Q09740|GFA1_SCHPO PUTATIVE GLUCOSAMINE--FRUCTOSE-6-PHOSPHATEAMINOTRANSFERASE [ISOMERIZING] (HEXOSEPHOSPHATE AMINOTRANSFERASE)(D-FRUCTOSE-6-PHOSPHATE AMIDOTRANSFERASE) (GFAT) >pir||T11674glutamine--fructose-6-phosphate transaminase (isomerizing) (EC2.6.1.16) - fission yeast (Schizosaccharomyces

					pombe)>pir||T39370glucosamine--fructose-6-phosphate aminotransferase - fission yeast(Schizosaccharomyces pombe)
Contig326	161	7.1e-10	350	469	sp|P53704|GFA1_CANAL GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE[ISOMERIZING] (HEXOSEPHOSPHATE AMINOTRANSFERASE)(D-FRUCTOSE-6-PHOSPHATE AMIDOTRANSFERASE) (GFAT) >pir||JC6012glutamine--fructose-6-phosphate transaminase (isomerizing) (EC2.6.1.16) - yeast (Candida albicans) >emb|CAA64380.1| (X94753)glucosamine--fructose-6-phosphate aminotransferase (isomerizing)[Candida albicans]
<LysB-murein hydrolase>					
p2g11fs.f1	125	2.3e-06	263	412	gb|AAA20878.1| (U04309) LysB [Bacteriophage phi-LC3]
2. Biomembranes (34)					
<membrane protein>					
o1g08fs.r1	352	7.7e-31	4	513	gi|6324450 ref|NP_014519.1|SMF1| has been localized to both the plasmamembrane and the mitochondrial membrane; Smf1p>sp|P38925|SMF1_YEAST TRANSPORTER PROTEIN SMF1/ESP1 >pir||S58647vacuolar transport protein ESP1 - yeast (Saccharomyces cerevisiae)>gb|AAB48984.1| (U15929) plasma membrane manganese transporter[Saccharomyces cerevisiae] >emb|CAA99141.1| (Z74864) ORF YOL122c[Saccharomyces cerevisiae]
g2e09fs.r1	311	9.4e-26	88	426	gi|6319862 ref|NP_009943.1|YCR017C| Ycr017cp >sp|P25618|YCQ7_YEASTHYPOTHETICAL 107.9 KD PROTEIN IN POL4-SRD1 INTERGENIC REGION>pir||S19427 probable membrane protein YCR017c - yeast(Saccharomyces cerevisiae) >emb|CAA42308.1| (X59720) YCR017c,len:953 [Saccharomyces cerevisiae]
m4a09fs.r1	286	4e-23	146	502	gi|6324008 ref|NP_014078.1|YNL321W| Ynl321wp >sp|P42839|YN61_YEASTHYPOTHETICAL 102.5 KD PROTEIN IN KRE1-HXT14 INTERGENIC REGION>pir||S51293 probable membrane protein YNL321w - yeast(Saccharomyces cerevisiae) >emb|CAA86376.1| (Z46259) NO339[Saccharomyces cerevisiae]>emb|CAA96252.1| (Z71597) ORF YNL321w[Saccharomyces cerevisiae]
r4f12fs.f1	271	4.5e-22	137	529	emb|CAA21926.1| (AL033391) hypothetical membrane protein [Candida albicans]
o1g08fs.f1	272	4.9e-22	151	558	gi|6324450 ref|NP_014519.1|SMF1| has been localized to both the plasmamembrane and the mitochondrial membrane; Smf1p>sp|P38925|SMF1_YEAST TRANSPORTER PROTEIN SMF1/ESP1 >pir||S58647vacuolar transport protein ESP1 - yeast (Saccharomyces

				cerevisiae)>gb|AAB48984.1| (U15929) plasma membrane manganese transporter[Saccharomyces cerevisiae] >emb|CAA99141.1| (Z74864) ORF YOL122c[Saccharomyces cerevisiae]
s1g08fs.r1	264	1.8e-21	89 655	gi|6320310 ref|NP_010390.1|YDR105C| Ydr105cp >pir||S51256 probable membraneprotein YDR105c - yeast (Saccharomyces cerevisiae)>emb|CAA87681.1|(Z47746) unknown [Saccharomyces cerevisiae]>emb|CAA88659.1|(Z48758) unknown [Saccharomyces cerevisiae]
f3e05fs.f1	227	4.8e-18	311 511	gi|6321475 ref|NP_011552.1|ORM1| Orm1p >sp|P53224|ORM1_YEAST ORM1 PROTEIN>pir||S64329 probable membrane protein YGR038w - yeast(Saccharomyces cerevisiae) >emb|CAA97026.1| (Z72823) ORF YGR038w[Saccharomyces cerevisiae]
m1a06fs.f1	235	8.7e-18	271 507	gi|6323872 ref|NP_013943.1|SKY1| Serine Protein Kinase; Sky1p>sp|Q03656|KM65_YEAST PROBABLE SERINE/THREONINE-PROTEIN KINASEYMR216C >pir||S55098 probable membrane protein YMR216c - yeast(Saccharomyces cerevisiae) >emb|CAA89931.1| (Z49809) unknown[Saccharomyces cerevisiae]
m4a09fs.f1	236	9.6e-18	243 509	gi|6324008 ref|NP_014078.1|YNL321W| Ynl321wp>sp|P42839|YN61_YEASTHYPOTHETICAL 102.5 KD PROTEIN IN KRE1-HXT14 INTERGENIC REGION>pir||S51293 probable membrane protein YNL321w - yeast(Saccharomyces cerevisiae) >emb|CAA86376.1| (Z46259) NO339[Saccharomyces cerevisiae]>emb|CAA96252.1| (Z71597) ORF YNL321w[Saccharomyces cerevisiae]
Contig255	203	2.5e-14	2 454	gi|6319265 ref|NP_009348.1|YAL053W| Yal053wp >sp|P39719|YAF3_YEASTHYPOTHETICAL 87.5 KD PROTEIN IN ACS1-GCV3 INTERGENIC REGION>pir||S51968 probable membrane protein YAL053w - yeast(Saccharomyces cerevisiae) >gb|AAC04980.1| (U12980) Yal053wp[Saccharomyces cerevisiae]
Contig86	196	7.2e-14	253 516	gi|6324882 ref|NP_014951.1|YOR306C| Yor306cp >pir||S67210 probable membraneprotein YOR306c - yeast (Saccharomyces cerevisiae)>emb|CAA99626.1|(Z75214) ORF YOR306c [Saccharomyces cerevisiae]
o3d10fs.f1	188	1.2e-12	218 508	gi|6325452 ref|NP_015520.1|YPR194C| Ypr194cp >pir||S58824 probable membraneprotein YPR194c - yeast (Saccharomyces cerevisiae)>gb|AAB64623.1|(U25841) Similar to S. cerevisiae hypothetical protein HRD799 (PIRaccession number S45161) and S. pombe hypothetical protein (PIRaccession number S43741) [Saccharomyces cerevisiae]

r4f12fs.r1	167	9.4e-11	35 376	emb|CAA21926.1| (AL033391) hypothetical membrane protein [Candida albicans]
b3e12fs.r1	164	3.2e-10	124 417	gi|6320427 ref|NP_010507.1|YDR221W| Ydr221wp >pir||S59428 probable membraneprotein YDR221w - yeast (Saccharomyces cerevisiae)>emb|CAA88501.1|(Z48612) unknown [Saccharomyces cerevisiae]
o2f01fs.r1	164	3.6e-10	3 509	gi|6319265 ref|NP_009348.1|YAL053W| Yal053wp >sp|P39719|YAF3_YEASTHYPOTHETICAL 87.5 KD PROTEIN IN ACS1-GCV3 INTERGENIC REGION>pir||S51968 probable membrane protein YAL053w - yeast(Saccharomyces cerevisiae) >gb|AAC04980.1| (U12980) Yal053wp[Saccharomyces cerevisiae]
Contig242	159	7.6e-10	258 527	gi|6321118 ref|NP_011196.1|YFR006W| Yfr006wp >sp|P43590|YFH6_YEASTHYPOTHETICAL 61.8 KD PEPTIDASE IN MPR1-GCN20 INTERGENIC REGION>pir||S56261 probable membrane protein YFR006w - yeast(Saccharomyces cerevisiae) >dbj|BAA09245.1| (D50617) YFR006W[Saccharomyces cerevisiae]
Contig644	134	1e-07	687 878	pir||S61100 probable membrane protein YNL044w - yeast (Saccharomycescerevisiae) >emb|CAA64238.1| (X94547) N2650 [Saccharomycescerevisiae]
Contig356	108	2e-05	176 301	gi|6323191 ref|NP_013263.1|YLR162W| Ylr162wp >pir||S68478 probable membraneprotein YLR162w - yeast (Saccharomyces cerevisiae)>gb|AAB67486.1|(U51921) L9632.2 gene product [Saccharomyces cerevisiae]
e2f11fs.r1	111	0.002	198 362	gi|6679020 ref|NP_032702.1|| next to the Brca1>sp|P97432|M172_MOUSE MEMBRANECOMPONENT, CHROMOSOME 17, SURFACE MARKER 2 (NEXT TO BRCA1 GENE 1PROTEIN) >gb|AAC53025.1| (U73039) Nbr1 [Mus musculus]
g1a02fs.r1	104	0.0063	365 469	gi|6320419 ref|NP_010499.1|YDR213W| Ydr213wp >pir||S61580 probable membraneprotein YDR213w - yeast (Saccharomyces cerevisiae)>emb|CAA92364.1|(Z68195) unknown [Saccharomyces cerevisiae]>emb|CAA92356.1|(Z68194) unknown [Saccharomyces cerevisiae]
<integral membrane protein>				
Contig195	555	8.1e-53	37 696	pir||T40742 hypothetical integral membrane protein - fission yeast(Schizosaccharomyces pombe) >emb|CAA21902.1| (AL033388)hypothetical integral membrane protein [Schizosaccharomyces pombe]
Contig327	165	2e-10	198 473	pir||T40742 hypothetical integral membrane protein - fission yeast(Schizosaccharomyces pombe) >emb|CAA21902.1|

					(AL033388)hypothetical integral membrane protein [Schizosaccharomyces pombe]
n1c08fs.r1	152	5.1e-09	14	343	gb|AAD30438.1|AF1196 (AF119672) integral membrane protein [Magnaporthegrisea]
Contig234	146	2.4e-08	54	461	gb|AAD30438.1|AF1196 (AF119672) integral membrane protein [Magnaporthegrisea]
n3e12fs.r1	99	0.037	301	513	gi|6323589 ref|NP_013660.1|SUR7| putative integral membrane protein; Sur7p>sp|P54003|SUR7_YEAST SUR7 PROTEIN >pir||S49809 probable membraneprotein YML052w - yeast (Saccharomyces cerevisiae)>emb|CAA86724.1|(Z46729) unknown [Saccharomyces cerevisiae]
<transmembrane protein>					
a1c11fs.r1	570	4.6e-53	7	441	emb|CAB65007.1| (Y17316) transmembrane protein [Erysiphe pisi]
s1c08fs.r1	411	4.2e-36	74	640	emb|CAB65007.1| (Y17316) transmembrane protein [Erysiphe pisi]
d3b08fs.r1	131	2.6e-06	309	488	emb|CAB65007.1| (Y17316) transmembrane protein [Erysiphe pisi]
<vacuolar membrane protein>					
Contig261	190	3.4e-12	147	482	gi|6322988 ref|NP_013060.1|VPS13| component of peripheral vacuolar membraneprotein complex; Vps13p >sp|Q07878|VP13_YEAST VACUOLAR PROTEINSORTING-ASSOCIATED PROTEIN VPS13 >pir||S64791 VPS13 protein - yeast(Saccharomyces cerevisiae) >emb|CAA97491.1| (Z73145) ORF YLL040c[Saccharomyces cerevisiae] >gb|AAC08284.1| (AF001317) Soi1p[Saccharomyces cerevisiae]
<peripheral vacuolar membraneprotein>					
a2e11fs.r1	268	1.8e-20	17	445	gi|6322988 ref|NP_013060.1|VPS13| component of peripheral vacuolar membraneprotein complex; Vps13p >sp|Q07878|VP13_YEAST VACUOLAR PROTEINSORTING-ASSOCIATED PROTEIN VPS13 >pir||S64791 VPS13 protein - yeast(Saccharomyces cerevisiae) >emb|CAA97491.1| (Z73145) ORF YLL040c[Saccharomyces cerevisiae] >gb|AAC08284.1| (AF001317) Soi1p[Saccharomyces cerevisiae]
<annexin XIV-calcium-dependent phospholipid binding proteins>					
Contig501	359	5.1e-32	231	533	gb|AAC09237.1| (AF036871) annexin XIV [Neurospora crassa]
d4e10fs.f1	212	1e-15	127	483	gb|AAC09237.1| (AF036871) annexin XIV [Neurospora crassa]
<opsin-1-a seven transmembrane helix retinal-binding protein>					
b2e04fs.r1	460	7.3e-43	39	413	gb|AAD45253.1| (AF135863) opsin-1 [Neurospora crassa]
b2e04fs.f1	151	1.5e-09	331	501	gb|AAD45253.1| (AF135863) opsin-1 [Neurospora crassa]
<peroxisomal membrane protein>					
d2g08fs.r1	104	0.00079	315	461	sp|P21245|P47A_CANBO PEROXISOMAL MEMBRANE PROTEIN PMP47A >pir||A23667 47Kperoxisomal membrane protein - yeast (Candida

					boidinii)>gb|AAA63791.1| (J05672) peroxisomal membrane protein [Candidaboidinii]
<peroxisomal membrane protein per10>					
b4g09fs.r1	111	5.2e-05	142	276	sp|P78723|PEXE_PICAN PEROXISOMAL MEMBRANE PROTEIN PER10 (PEROXIN-14)>gb|AAB40596.1| (U46195) Per10p [Pichia angusta]
3. Cytoskeleton, organelle (55)					
<peroxisomal hydratase-dehydrognease-epimerase>					
g3b10fs.r1	477	1e-43	117	482	sp|Q01373|FOX2_NEUCR PEROXISOMAL HYDRATASE-DEHYDROGENASE-EPIMERASE (HDE)(MULTIFUNCTIONAL BETA-OXIDATION PROTEIN) (MFP) [INCLUDES:2-ENOYL-COA HYDRATASE ; D-3-HYDROXYACYL COA DEHYDROGENASE]>pir||S54786 multifunctional beta-oxidation protein - Neurosporacrassa >emb|CAA56355.1| (X80052) multifunctional beta-oxidationprotein [Neurospora crassa]
g3b10fs.f1	380	3.1e-33	204	506	sp|Q01373|FOX2_NEUCR PEROXISOMAL HYDRATASE-DEHYDROGENASE-EPIMERASE (HDE)(MULTIFUNCTIONAL BETA-OXIDATION PROTEIN) (MFP) [INCLUDES:2-ENOYL-COA HYDRATASE ; D-3-HYDROXYACYL COA DEHYDROGENASE]>pir||S54786 multifunctional beta-oxidation protein - Neurosporacrassa >emb|CAA56355.1| (X80052) multifunctional beta-oxidationprotein [Neurospora crassa]
<Golgi complex-associated protein>					
d1b11fs.r1	634	2.5e-59	23	502	pir||JC5837 364K Golgi complex-associated protein - rat >dbj|BAA05026.1|(D25543) rat GCP360 [Rattus rattus]
<Proteasome subunit>					
Contig323	661	4.5e-64	381	875	gi|6325360 ref|NP_015428.1|PRE2| proteasome subunit; Pre2p>sp|P30656|PRCE_YEAST PROTEASOME COMPONENT PRE2 PRECURSOR(MACROPAIN SUBUNIT PRE2) (PROTEINASE YSCE SUBUNIT PRE2)(MULTICATALYTIC ENDOPEPTIDASE COMPLEX SUBUNIT PRE2) >pir||A45411multicatalytic endopeptidase complex (EC3.4.99.46) chain PRE2precursor - yeast (Saccharomyces cerevisiae)>emb|CAA48628.1|(X68662) proteasome Pre2 subunit [Saccharomyces
Contig759	585	5.4e-56	111	689	gi|6320849 ref|NP_010928.1|PRE1| 22.6 kDa proteasome subunit;Pre1p>sp|P22141|PRCG_YEAST PROTEASOME COMPONENT C11 (MACROPAIN SUBUNITC11) (PROTEINASE YSCE SUBUNIT 11) (MULTICATALYTIC ENDOPEPTIDASECOMPLEX SUBUNIT C11) >pir||S50470 multicatalytic endopeptidasecomplex (EC 3.4.99.46) chain PRE1 - yeast (Saccharomycescerevisiae) >pdb|1RYP|K Chain K, Crystal Structure Of The 20sProteasome From Yeast At 2.4

o2e06fs.r1	545	9.5e-52	61	513	pir||T39054 probable proteasome component - fission yeast (Schizosaccharomycespombe) >emb|CAB11290.1| (Z98603) proteasome subunit C2[Schizosaccharomyces pombe]
d1d04fs.r1	404	8.1e-37	144	494	gi|6321427 ref|NP_011504.1|SCL1| Proteasome subunit YC7alpha/Y8 (protease yscEsubunit 7); Scl1p >sp|P21243|PRCI_YEAST PROTEASOME COMPONENTC7-ALPHA (MACROPAIN SUBUNIT C7-ALPHA) (PROTEINASE YSCE SUBUNIT 7)(MULTICATALYTIC ENDOPEPTIDASE COMPLEX C7) (COMPONENT Y8) (SCL1SUPPRESSOR PROTEIN) >pir||SNBYS1 multicatalytic endopeptidasecomplex (EC 3.4.99.46) chain YC7-alpha - yeast (Saccharomycescerevisiae)
p1h04fs.r1	399	2.9e-36	82	507	sp|P24495|PRC3_XENLA PROTEASOME COMPONENT C3 (MACROPAIN SUBUNIT C3)(MULTICATALYTIC ENDOPEPTIDASE COMPLEX SUBUNIT C3) >pir||JH0421proteasome chain XC3 - African clawed frog >gb|AAB19485.1| (S51111)proteasome subunit XC3 [Xenopus laevis, Peptide, 234 aa]
o2e06fs.f1	299	1.1e-25	180	527	pir||T39054 probable proteasome component - fission yeast (Schizosaccharomycespombe) >emb|CAB11290.1| (Z98603) proteasome subunit C2[Schizosaccharomyces pombe]
p1h04fs.f1	217	5.2e-17	260	541	gi|4506181 ref|NP_002778.1|| proteasome (prosome, macropain) subunit, alphatype, 2 >sp|P25787|PRC3_HUMAN PROTEASOME COMPONENT C3 (MACROPAINSUBUNIT C3) (MULTICATALYTIC ENDOPEPTIDASE COMPLEX SUBUNIT C3)>pir||SNHUC3 multicatalytic endopeptidase complex (EC 3.4.99.46)chain C3 - human >dbj|BAA00657.1| (D00760) proteasome subunit C3[Homo sapiens]
Contig250	207	5.8e-16	211	471	gi|6319430 ref|NP_009512.1|PRE7| proteasome subunit; Pre7p>sp|P23724|PRC5_YEAST POTENTIAL PROTEASOME COMPONENT C5(MULTICATALYTIC ENDOPEPTIDASE COMPLEX SUBUNIT C5) >pir||S42436multicatalytic endopeptidase complex (EC 3.4.99.46) chain PRS3 -yeast (Saccharomyces cerevisiae)>gb|AAA68908.1| (M34777)proteasome subunit [Saccharomyces cerevisiae]>dbj|BAA00725.1|(D00845) proteasome subunit [Saccharomyces
m1e01fs.r1	176	4.7e-12	209	451	emb|CAB63792.1| (AL135751) putative 26s proteasome subunit[Schizosaccharomyces pombe]
<endoplasmic reticulum associated protein>					
t2f03fs.r1	185	1.3e-13	228	458	gb|AAA68907.1|(M34777) endoplasmic reticulum lumen protein retaining receptor[Saccharomyces cerevisiae]>emb|CAA84860.1| (Z35801) ORF YBL040c[Saccharomyces

<KINESIN>					
p4h04fs.f1	170	3.9e-11	125	532	gb|AAB87735.1| (U78597) kinesin light chain [Plectonema boryanum]
<KINESIN-LIKE PROTEIN KLPA>					
j1e03fs.f1	285	4.3e-23	287	481	sp|P28739|KLPA_EMENI KINESIN-LIKE PROTEIN KLPA >pir||A44337 kinesin-relatedprotein KLPA - Emericella nidulans >emb|CAA45887.1| (X64603)KLPA[Emericella nidulans]
<tubulin>					
Contig649	1192	2.2e-120	164	892	sp|Q92335|TBA_SORMA TUBULIN ALPHA CHAIN >emb|CAA94304.1| (Z70290)alpha-tubulin [Sordaria macrospora]
Contig580	1175	1.5e-118	92	808	sp|P33127|TBB_ACRCO TUBULIN BETA CHAIN >pir||S29625 tubulin beta chain-fungus (Acremonium coenophialum) >emb|CAA40178.1| (X56847)beta-tubulin [Neotyphodium coenophialum]
Contig826	540	2.9e-51	348	743	sp|P24634|TBA2_EMENI TUBULIN ALPHA-2 CHAIN >pir||S13337 tubulin alpha-2 chain- Emericella nidulans
f1a03fs.r1	536	8.3e-51	137	499	sp|P38668|TBA1_NEUCR TUBULIN ALPHA-A CHAIN >emb|CAA55940.1| (X79403)alpha-tubulin A [Neurospora crassa]
Contig457	493	3.1e-46	258	536	gb|AAB18275.1| (U27303) beta-tubulin [Gibberella fujikuroi]
Contig29	347	8.1e-31	298	537	sp|P24634|TBA2_EMENI TUBULIN ALPHA-2 CHAIN >pir||S13337 tubulin alpha-2 chain- Emericella nidulans
Contig22	298	1.9e-25	289	480	sp|P10875|TBB_CANAL TUBULIN BETA CHAIN >pir||UBCKBA tubulin beta chain - yeast(Candida albicans) >gb|AAA34375.1| (M19398) beta-tubulin [Candidaalbicans]
<dynactin>					
Contig57	1067	4.7e-107	3	860	sp|Q01397|DYNA_NEUCR DYNACTIN, 150 KD ISOFORM (150 KD DYNEIN-ASSOCIATEDPOLYPEPTIDE) (DP-150) (DAP-150) (P150-GLUED) >pir||T18364 ro-3protein - Neurospora crassa >gb|AAA80458.1| (L48661) productp150Glued [Neurospora crassa]
3.1. actin-see also mitosis					
<actin>					
Contig292	777	2.3e-76	63	503	gb|AAD41038.1|AF1125 (AF112537) actin [Colletotrichum gloeosporioides f.sp.malvae]
Contig294	576	5.1e-55	77	502	pir||T38191 actin-like protein - fission yeast (Schizosaccharomyces pombe)>emb|CAB52711.1| (AL109831) actin-like protein [Schizosaccharomycespombe]
Contig295	474	3e-44	127	474	gi|6320175 ref|NP_010255.1|ARP2| actin-related protein; Arp2p>sp|P32381|ARP2_YEAST ACTIN-LIKE PROTEIN ARP2 >pir||S20225actin-like protein ACT2 - yeast (Saccharomyces

					cerevisiae) (strainX2180)>emb|CAA43718.1| (X61502) actin-like protein [Saccharomycescerevisiae] >emb|CAA96460.1| (Z71781) actin-like protein ACT2[Saccharomyces cerevisiae] >emb|CAA98588.1| (Z74077) ORF YDL029w[Saccharomyces cerevisiae]
Contig886	373	1.6e-33	285	503	dbj|BAA74960.1| (AB003111) actin [Humicola grisea var. thermoidea]
<gamma-actin>					
Contig1025	1946	3.2e-200	98	1222	gb|AAF00008.1|AF0569 (AF056976) gamma-actin [Acremonium chrysogenum]
<PROFILIN-assembly of actin monomers>					
Contig216	299	1e-25	261	653	sp|P39825|PROF_SCHPO PROFILIN >pir||A53952 profilin - fission yeast(Schizosaccharomyces pombe) >emb|CAB38578.1| (Z98762) profilin.[Schizosaccharomyces pombe]
<ARP2/3 COMPLEX-actin polymerization>					
Contig925	922	1e-91	64	1005	sp|O14241|AR34_SCHPO PROBABLE ARP2/3 COMPLEX 34 KD SUBUNIT (P34-ARC)>pir||T39044 probable Arp2-3 complex subunit - fission yeast(Schizosaccharomyces pombe) >emb|CAB11733.1| (Z98981) probableArp2-3 complex subunit [Schizosaccharomyces pombe]
j4a05fs.r1	505	1.6e-47	9	488	sp|P78774|AR41_SCHPO PROBABLE ARP2/3 COMPLEX 41 KD SUBUNIT (P41-ARC)>emb|CAA70202.1| (Y08998) putative Arp2/3 complex 41kD subunit[Schizosaccharomyces pombe]
<fimbrin-actin bundling>					
g4h10fs.r1	805	2.7e-79	20	490	emb|CAA10667.1| (AJ132432) fimbrin [Gibberella pulicaris]
m1f04fs.f1	606	3.4e-58	153	521	emb|CAA10667.1| (AJ132432) fimbrin [Gibberella pulicaris]
m1f04fs.r1	227	4.4e-17	49	438	gi|6320334 ref|NP_010414.1|SAC6| fibrim homolog (actin-filament bundlingprotein); Sac6p >sp|P32599|FIMB_YEAST FIMBRIN (ABP67) >pir||S29320fimbrin - yeast (Saccharomyces cerevisiae)>emb|CAA45346.1|(X63867) fimbrin [Saccharomyces cerevisiae]>emb|CAA88210.1|(Z48179) Sac6p [Saccharomyces cerevisiae] >prf||1802390A fimbrin[Saccharomyces cerevisiae]
<COFILIN-actin binding>					
Contig827	337	1e-29	179	625	sp|P78929|COFI_SCHPO COFILIN >pir||T38120 cofilin - fission yeast(Schizosaccharomyces pombe) >dbj|BAA14039.1| (D89939)actindepolymerazing factor [Schizosaccharomyces pombe]>emb|CAB11258.1|(Z98600) cofilin [Schizosaccharomyces pombe]
<ACTIN-LIKE PROTEIN>					
e4a10fs.r1	411	1.6e-37	283	531	sp|P38673|ACTZ_NEUCR ACTIN-LIKE PROTEIN (CENTRACTIN)>pir||A54802actin-related protein ro-4 - Neurospora crassa >gb|AAA64907.1|(L31505) centractin [Neurospora crassa]

e4a10fs.f1	262	9.8e-22	373	513	sp|P38673|ACTZ_NEUCR ACTIN-LIKE PROTEIN (CENTRACTIN)>pir||A54802actin-related protein ro-4 - Neurospora crassa>gb|AAA64907.1|(L31505) centractin [Neurospora crassa]
<exocyst complex component sec3>					
j1a04fs.r1	102	0.0089	30	305	gi|6320845 ref|NP_010924.1|SEC3| SEC3 encodes the 144 kD and 91 kD componentsof the Exocyst complex; the 91 kD component is a C-terminalproteolytic breakdown product of full length Sec3p; Sec3p>sp|P33332|SEC3_YEAST EXOCYST COMPLEX COMPONENT SEC3 (PSL1 PROTEIN)>pir||S41794 SEC3 protein - yeast (Saccharomyces cerevisiae)>gb|AAB49380.1| (L22204) Psl1p [Saccharomyces cerevisiae]>gb|AAB64541.1| (U18778)
3.2. myosin					
<myosin I myoA-the heavy chain of muscle protein>					
r4f11fs.r1	686	1.1e-65	1	594	pir||A56511 myosin I myoA - Emericella nidulans >gb|AAA67877.1| (U12427)myosin I heavy chain [Emericella nidulans]
<myosin II heavy chain>					
g4h07fs.r1	124	2.1e-05	23	361	sp|P08799|MYS2_DICDI MYOSIN II HEAVY CHAIN, NON MUSCLE >pir||A26655 myosinheavy chain - slime mold (Dictyostelium discoideum)>gb|AAA33227.1|(M14628) myosin heavy chain [Dictyostelium discoideum]
<tropomyosin I>					
Contig539	216	6.5e-17	179	565	gi|6324250 ref|NP_014320.1|TPM1| tropomyosin I; Tpm1p >sp|P17536|TPM1_YEASTTROPOMYOSIN 1 >pir||A32183 tropomyosin TPM1 - yeast (Saccharomycescerevisiae) >gb|AAA35174.1| (M25501) tropomyosin (TPM1)[Saccharomyces cerevisiae] >emb|CAA60179.1| (X86470) tropomyosin[Saccharomyces cerevisiae] >emb|CAA95953.1| (Z71355) ORF YNL079c[Saccharomyces cerevisiae]
3.3. choline					
<choline dehydrogenase>					
e4g06fs.f1	187	7.1e-13	169	510	sp|P54223|BETA_RHIME CHOLINE DEHYDROGENASE (CHD)
e4g06fs.r1	156	1.6e-09	208	522	gb|AAD23901.1|AF0094 (AF009415) choline dehydrogenase [Staphylococcusxylosus]
<choline oxidase>					
Contig55	278	9.2e-23	104	505	pir||S52489 choline oxidase (EC 1.1.3.17) - Arthrobacter globiformis>pir||S62689 choline oxidase (EC 1.1.3.17) - Arthrobacterglobiformis >emb|CAA59321.1| (X84895) choline oxidase [Arthrobacterglobiformis]

Contig84	270	6.8e-22	209	529	pir||S52489 choline oxidase (EC 1.1.3.17) - Arthrobacter globiformis>pir||S62689 choline oxidase (EC 1.1.3.17) - Arthrobacterglobiformis >emb|CAA59321.1| (X84895) choline oxidase [Arthrobacterglobiformis]
<cholinephosphate cytidylyltransferase>					
j3f01fs.r1	425	4.4e-39	14	487	pir||T41163 cholinephosphate cytidylyltransferase - fission yeast(Schizosaccharomyces pombe)
<PHOSPHATIDYLINOSITOL/PHOSPHATIDYL-CHOLINE TRANSFER PROTEIN>					
Contig800	265	4.2e-22	210	563	sp|Q10137|SC14_SCHPO PUTATIVE SEC14 CYTOSOLIC FACTOR(PHOSPHATIDYLINOSITOL/PHOSPHATIDYL-CHOLINE TRANSFER PROTEIN) (PI/PCTP)>pir||T38768 probable sec14 cytosolic factor - fission yeast(Schizosaccharomyces pombe) >emb|CAA93167.1| (Z69086) putativesec14 cytosolic factor [Schizosaccharomyces pombe]
Contig685	229	2.9e-18	539	715	sp|Q10137|SC14_SCHPO PUTATIVE SEC14 CYTOSOLIC FACTOR(PHOSPHATIDYLINOSITOL/PHOSPHATIDYL-CHOLINE TRANSFER PROTEIN) (PI/PCTP)>pir||T38768 probable sec14 cytosolic factor - fission yeast(Schizosaccharomyces pombe) >emb|CAA93167.1| (Z69086) putativesec14 cytosolic factor [Schizosaccharomyces pombe]
g2d10fs.r1	117	8.2e-06	99	428	sp|Q10137|SC14_SCHPO PUTATIVE SEC14 CYTOSOLIC FACTOR(PHOSPHATIDYLINOSITOL/PHOSPHATIDYL-CHOLINE TRANSFER PROTEIN) (PI/PCTP)>pir||T38768 probable sec14 cytosolic factor - fission yeast(Schizosaccharomyces pombe) >emb|CAA93167.1| (Z69086) putativesec14 cytosolic factor [Schizosaccharomyces pombe]
<phosphatidylcholine-sterol acetyltransferase>					
b3h11fs.r1	319	5.1e-27	15	395	pir||T40685 phosphatidylcholine-sterol acetyltransferase homolog - fissionyeast (Schizosaccharomyces pombe) >emb|CAA22887.1| (AL035263) weaksimilarity to chick phosphatidylcholine-ste rol acetyltransferase[Schizosaccharomyces pombe]
Contig480	234	7.7e-18	267	518	pir||T40685 phosphatidylcholine-sterol acetyltransferase homolog - fissionyeast (Schizosaccharomyces pombe) >emb|CAA22887.1| (AL035263) weaksimilarity to chick phosphatidylcholine-ste rol acetyltransferase[Schizosaccharomyces pombe]

3.4. other

<Glycosyltransferase-glycopeptidolipid biosyn>					
g3g03fs.f1	259	9.2e-21	112	519	dbj|BAA21387.2| (AB004534) glycosyltransferase [Schizosaccharomyces pombe]

i1c10fs.f1	239	7.2e-19	72	437	pir||F75587 probable glycosyltransferase - Deinococcus radiodurans (strain R1)>gb|AAF12451.1|AE001863_76 (AE001863) glycosyltransferase,putative[Deinococcus radiodurans]
4. organelle (5)					
<mitochondrial carrier protein>					
s1f08fs.f1	237	4.2e-19	221	514	sp|P32331|YMC1_YEAST MITOCHONDRIAL CARRIER PROTEIN YMC1 PRECURSOR>emb|CAA47602.1| (X67122) mitochondrial carrier protein[Saccharomyces cerevisiae]
p3c04fs.r1	132	2.2e-07	8	148	gb|AAC82534.1| (AC003083) mitochondrial carrier protein-like; similar toQ09461 (PID:g2497990) [Homo sapiens]
<mitochondrial protein>					
Contig414	728	3.9e-71	8	808	emb|CAA56954.1| (X81067) probable mitochondrial protein; nearly identical toYME1 [Saccharomyces cerevisiae]
l1g02fs.r1	234	1.3e-17	8	466	emb|CAA56955.1| (X81068) probable mitochondrial protein [Saccharomycescerevisiae]
<mitochondrial morphologyprotein MMM1>					
p1b06fs.r1	328	9.9e-29	160	462	gb|AAF43713.1|AF2384 (AF238480) maintaining mitochondrial morphologyproteinMMM1 [Neurospora crassa] >gb|AAF43714.1|AF239620_1(AF239620) maintaining mitochondrial morphology protein MMM1[Neurospora crassa]
II.2. cell cycle control (9)					
<cell division control protein>					
m2h05fs.r1	375	4.4e-33	34	495	pir||T38702 hypothetical protein SPAC3F10.01 - fission yeast(Schizosaccharomyces pombe) (fragment) >emb|CAA93299.1| (Z69369)cell division control protein nda4 [Schizosaccharomyces pombe]
j3c03fs.r1	305	4.7e-25	62	538	pir||T40813 probable cell division control protein 68/transcription activatorhomolog - fission yeast (Schizosaccharomyces pombe)>emb|CAA21804.1| (AL032684) putative yeast cell division controlprotein 68 homolog (CDC68), putative transcriptional activator[Schizosaccharomyces pombe]
<cell division control protein cdc15>					
k4c10fs.f1	176	2.2e-11	314	520	pir||A57087 cell division control protein cdc15 - fission yeast(Schizosaccharomyces pombe)
<cell division cycle CDC50>					

p1a06fs.r1	221	6.1e-17	10	351	pir||T39676 probable yeast cell division cycle CDC50 homolog - fission yeast(Schizosaccharomyces pombe) >emb|CAA21916.1| (AL033389) similar to yeast cdc50 [Schizosaccharomyces pombe]
<cullin-neg regulator of cell cycle>					
r3f09fs.r1	467	7.6e-43	266	613	gb|AAD34471.1|AF1364 (AF136441) cullin 1 [Mus musculus]
<clock-controlled gene-6 protein>					
Contig980	172	1.4e-11	124	489	gb|AAC64287.1| (AF088908) clock-controlled gene-6 protein [Neurospora crassa]
Contig1029	147	2.9e-09	186	356	gb|AAC64287.1| (AF088908) clock-controlled gene-6 protein [Neurospora crassa]
<B-type cyclin>					
a4b02fs.r1	300	2.6e-25	70	456	gb|AAC79857.1| (AF094507) B-type cyclin [Candida albicans]
<cell cycle regulator p21 protein>					
b3g05fs.r1	199	4.2e-15	119	460	pir||T39220 cell cycle regulator p21 protein, Wos2p - fission yeast(Schizosaccharomyces pombe) >gb|AAA64891.1| (L41166) p21 protein[Schizosaccharomyces pombe] >emb|CAB16411.1| (Z99262) cell cycleregulator p21 protein, Wos2p [Schizosaccharomyces pombe]

II.3. Mitosis/cytokinesis

1. MITOSIS

<CHROMOSOME SEGREGATION PROTEIN-with microtubule, migration of chromosomes>					
Contig150	385	8.5e-35	135	509	sp|Q10113|MAL3_SCHPO MAL3 PROTEIN >pir||T37928 probable chromosome segregationprotein - fission yeast (Schizosaccharomyces pombe)>emb|CAA92392.1| (Z68198) putative chromosome segregation protein[Schizosaccharomyces pombe] >emb|CAA70707.1| (Y09518) MAL3 protein[Schizosaccharomyces pombe]
<dynamin-molecular motor, associated with mocrotuble, endocytosis>					
b4g06fs.r1	282	8.5e-23	238	447	sp|Q09748|YB68_SCHPO DYNAMIN-LIKE PROTEIN C12C2.08 >pir||T39373dynamin-related protein - fission yeast (Schizosaccharomyces pombe)>emb|CAA90821.1| (Z54140) dynamin-related protein[Schizosaccharomyces pombe]
b2f10fs.f1	112	0.00015	299	415	emb|CAB75996.1| (AL157874) dynamin-related protein [Schizosaccharomyces pombe]
<dynein-molecular motor>					
b4h06fs.r1	651	5.7e-61	7	459	sp|P78716|DYHC_FUSSO DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC) >gb|AAC33176.1|(U84215) cytoplasmic dynein heavy chain [Haematonectriahaematococca]

e2b07fs.r1	486	3.4e-44	20	478	sp|Q01397|DYNA_NEUCR DYNACTIN, 150 KD ISOFORM (150 KD DYNEIN-ASSOCIATEDPOLYPEPTIDE) (DP-150) (DAP-150) (P150-GLUED) >pir||T18364 ro-3protein - Neurospora crassa >gb|AAA80458.1| (L48661) productp150Glued [Neurospora crassa]
p3a02fs.r1	399	2.6e-36	208	474	gb|AAD00525.1| (U81827) 8 kDa cytoplasmic dynein light chain [Emericellanidulans]
e2b07fs.f1	200	1e-13	205	492	sp|Q01397|DYNA_NEUCR DYNACTIN, 150 KD ISOFORM (150 KD DYNEIN-ASSOCIATEDPOLYPEPTIDE) (DP-150) (DAP-150) (P150-GLUED) >pir||T18364 ro-3protein - Neurospora crassa >gb|AAA80458.1| (L48661) productp150Glued [Neurospora crassa]
b4h06fs.f1	197	8.8e-13	255	419	sp|P78716|DYHC_FUSSO DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC)>gb|AAC33176.1|(U84215) cytoplasmic dynein heavy chain [Haematonectriahaematococca]
<spindle assembly checkpoint protein SLDA>					
Contig68	562	2e-52	63	533	pir||T09224 spindle assembly checkpoint protein SLDA - Emericella nidulans>gb|AAC39457.1| (AF032987) spindle assembly checkpoint protein SLDA[Emericella nidulans]
<SSD1 PROTEIN>					
o4d08fs.f1	133	1.4e-06	250	525	gi|6320499 ref|NP_010579.1|SSD1| Ssd1p >sp|P24276|SSD1_YEAST SSD1 PROTEIN(SRK1 PROTEIN) >pir||A39578 SSD1 protein - yeast (Saccharomycescerevisiae) >gb|AAA35047.1| (M60318) SSD1 protein [Saccharomycescerevisiae] >gb|AAA35089.1| (M63004) SRK1 [Saccharomycescerevisiae] >gb|AAB64469.1| (U51031) Ssd1p [Saccharomycescerevisiae]
<syntaxin-essential for membrane fusion events critical for cell division>					
Contig785	154	2.5e-08	641	844	pir||T41624 probable syntaxin - fission yeast (Schizosaccharomyces pombe)>emb|CAB58411.1| (AL122011) putative syntaxin [Schizosaccharomycespombe]

2. cytokinesis (3)

<F-ACTIN CAPPING PROTEIN ALPHA-2 SUBUNIT>					
Contig851	482	4.6e-45	27	827	sp|P28497|CAZ2_CHICK F-ACTIN CAPPING PROTEIN ALPHA-2 SUBUNIT (CAPZ 36/32)(BETA-ACTININ SUBUNIT I) >pir||S36093 actin-capping protein alpha-2chain - chicken >gb|AAA48656.1| (M80589) capping protein alpha 2isoform [Gallus gallus]
<F-ACTIN CAPPING PROTEIN BETA SUBUNIT ISOFORMS 1 AND 2>					
j2g10fs.f1	379	3.8e-34	121	537	sp|P14315|CAPB_CHICK F-ACTIN CAPPING PROTEIN BETA SUBUNIT ISOFORMS 1 AND 2(CAPZ 36/32) (CAPZ B1 AND B2) (BETA-ACTININ SUBUNIT

					II)>pir||A34335 Z line actin-capping protein beta chain, form 1 - chicken >gb|AAA49144.1| (J04959) actin-capping protein Z betasubunit [Gallus gallus]
b4h11fs.r1	325	1.7e-28	85	456	sp|P14315|CAPB_CHICK F-ACTIN CAPPING PROTEIN BETA SUBUNIT ISOFORMS 1 AND 2(CAPZ 36/32) (CAPZ B1 AND B2) (BETA-ACTININ SUBUNIT II)>pir||A34335 Z line actin-capping protein beta chain, form 1 - chicken >gb|AAA49144.1| (J04959) actin-capping protein Z betasubunit [Gallus gallus]

II.4. Meiosis (2)

<pelota protein-a protein required for meiotic cell division>

o4e09fs.r1	332	3.8e-29	3	506	sp|P48612|PELO_DROME PELOTA PROTEIN >gb|AAC46879.1| (U27197) pelota[Drosophila melanogaster]
o4e09fs.f1	219	9e-17	161	526	pir||T41199 dom34 protein homolog - fission yeast (Schizosaccharomyces pombe)>emb|CAB52153.1| (AL109736) pelota protein/yeast dom34 homolog;probable involvement in meiotic and mitotic divisions[Schizosaccharomyces pombe]

II.5. Cell death (1)

<CELLULAR APOTOSIS SUSCEPTIBILITY PROTEIN>

g1c12fs.r1	287	3.7e-24	30	464	dbj|BAA21462.1| (AB004539) CELLULAR APOTOSIS SUSCEPTIBILITY PROTEIN[Schizosaccharomyces pombe]

III. Processes

III.1. Cell rescue, defense, osmotic adaptation, starvation response, development ## (includes antibiotics, toxins)

1. development (24)

growth regulation protein

<growth regulation protein>

a2b06fs.r1	111	0.00029	92	343	gi|6324617 ref|NP_014686.1|WHI2| involved in growth regulation; Whi2p>sp|P12611|WHI2_YEAST GROWTH REGULATION PROTEIN >pir||COBYW2 WHI2protein - yeast (Saccharomyces cerevisiae) >emb|CAA99234.1|(Z74951) ORF YOR043w [Saccharomyces cerevisiae]

-sexual

<mating and morphogenesis protein>

m1c06fs.r1	155	3.9e-09	17	214	sp|P40995|SCD1_SCHPO SCD1 PROTEIN >pir||T37789 mating and morphogenesisprotein Scd1p - fission yeast (Schizosaccharomyces pombe)>emb|CAB11037.1| (Z98529) mating and morphogenesis protein Scd1p.[Schizosaccharomyces pombe] >gb|AAA50556.2| (U12538) Scd1 protein[Schizosaccharomyces pombe]
<tol protein-a mediator of mating-type-associated vegetative incompatibility in fungus>					
g1b06fs.f1	195	2.7e-13	15	482	pir||T17430 tol protein - Neurospora crassa >gb|AAC64945.1| (AF085183) TOL[Neurospora crassa]
i3a12fs.r1	155	5.2e-09	50	442	pir||T17430 tol protein - Neurospora crassa >gb|AAC64945.1| (AF085183) TOL[Neurospora crassa]
<SEXUAL DIFFERENTIATION PROCESS PROTEIN-expressed during sexual diff in S. pombe>					
i4e07fs.r1	410	1.2e-36	31	477	sp|P40900|ISP4_SCHPO SEXUAL DIFFERENTIATION PROCESS PROTEIN ISP4 >pir||S45495isp4 protein - fission yeast (Schizosaccharomyces pombe)>dbj|BAA03147.1| (D14061) ORF [Schizosaccharomyces pombe]
pigment production					
<GREEN PIGMENT SYNTHASE>					
a3c02fs.r1	303	2e-24	77	475	sp|Q03149|WA_EMENI CONIDIAL GREEN PIGMENT SYNTHASE >pir||S28353 probablepolyketide synthase - Emericella nidulans >emb|CAA46695.1| (X65866)putative polyketide or fatty acid synthase [Emericella nidulans]>prf||1905375A wA gene [Emericella nidulans]
<coproporphyrinogen III oxidase precursor>					
Contig183	407	4.2e-37	220	621	sp|P36551|HEM6_HUMAN COPROPORPHYRINOGEN III OXIDASE PRECURSOR(COPROPORPHYRINOGENASE) (COPROGEN OXIDASE) (COX) >pir||I37259coproporphyrinogen oxidase (EC 1.3.3.3) precursor - human(fragment) >emb|CAA82250.1| (Z28409) coproporphyrinogen oxidase[Homo sapiens]
b2g04fs.r1	162	1.7e-10	236	418	gb|AAD28474.1|AF1336 (AF133671) coproporphyrinogen III oxidase precursor[Chlamydomonas reinhardtii] >gb|AAD28475.1|AF133672_1 (AF133672)coproporphyrinogen III oxidase precursor [Chlamydomonasreinhardtii]
<brown 2-A developmentally regulated gene cluster involved in conidial pigment biosynthesis>					
g3c01fs.r1	336	5.4e-29	9	455	gb|AAF03349.1|AF1048 (AF104823) brown 2 [Aspergillus fumigatus]
g3c01fs.f1	245	4.5e-19	304	534	gb|AAF03349.1|AF1048 (AF104823) brown 2 [Aspergillus fumigatus]
<PvcA-Pseudomonas aeruginosa pyoverdine chromophore biosynthesis gene cluster>					
b4f04fs.r1	203	2.6e-15	149	445	gb|AAC21671.1| (AF002222) PvcA [Pseudomonas aeruginosa]
b3f10fs.r1	111	5.3e-05	205	432	gb|AAC21671.1| (AF002222) PvcA [Pseudomonas aeruginosa]

<osmotic growth protein>

Contig128	252	5.1e-20	178	408	gi|6322511 ref|NP_012585.1|OSM1| osmotic growth protein; Osm1p>sp|P21375|OSM1_YEAST OSMOTIC GROWTH PROTEIN 1 >pir||S46591fumarate reductase (EC 1.3.99.1) homolog OSM1 precursor - yeast(Saccharomyces cerevisiae) >gb|AAA62859.1| (L26347) orf gtB501[Saccharomyces cerevisiae]>emb|CAA89579.1| (Z49551) ORF YJR051w[Saccharomyces cerevisiae]>gb|AAA88754.1| (L36344) ORF; putative[Saccharomyces cerevisiae]
Contig616	137	1.5e-07	91	477	gi|6322511 ref|NP_012585.1|OSM1| osmotic growth protein; Osm1p>sp|P21375|OSM1_YEAST OSMOTIC GROWTH PROTEIN 1 >pir||S46591fumarate reductase (EC 1.3.99.1) homolog OSM1 precursor - yeast(Saccharomyces cerevisiae) >gb|AAA62859.1| (L26347) orf gtB501[Saccharomyces cerevisiae] >emb|CAA89579.1| (Z49551) ORF YJR051w[Saccharomyces cerevisiae]>gb|AAA88754.1| (L36344) ORF; putative[Saccharomyces cerevisiae]

<imbibition protein>

g1h07fs.r1	260	2e-20	22	462	emb|CAB66109.1| (AL133248) imbibition protein homolog [Arabidopsis thaliana]

organism development

<epd1 protein precursor-essential for pseudohyphaldevelopment 1>

Contig1037	1209	3.2e-122	274	1635	sp|P56092|EPD1_CANMA EPD1 PROTEIN PRECURSOR (ESSENTIAL FOR PSEUDOHYPHALDEVELOPMENT 1) >dbj|BAA21103.1| (AB005130) EPD1 [Candida maltosa]

<**CAP20**-One of the genes expressed uniquely in C. gloeosporioides during appressorium formation>

Contig370	378	4.4e-34	187	579	gb|AAA77678.1| (U18061) CAP20 [Colletotrichum gloeosporioides]
Contig811	106	0.0047	718	852	gb|AAA77678.1| (U18061) CAP20 [Colletotrichum gloeosporioides]

<ascospore maturation 1 protein>

g1e01fs.r1	490	6.5e-46	79	465	gb|AAB06995.1| (U51117) ascospore maturation 1 protein [Neurospora crassa]

<genitaliumglycerol-3-phospate dehydrogenase>

r4f10fs.r1	173	1.6e-11	179	403	gb|AAF60576.1| (AC006768) contains similarity to Mycoplasma genitaliumglycerol-3-phospate dehydrogenase (SW:P47285) [Caenorhabditiselegans]

<BIGH3-transforming growth factor-beta induced gene>

Contig218	104	0.0001	287	490	gb|AAC24944.1| (AC005219) BIGH3 [Homo sapiens]

<**Bdf1p**-required for sporulation>

h4g12fs.r1	178	9.8e-12	4	276	gb|AAA89115.1| (U18116) Bdf1p [Saccharomyces cerevisiae]

<TRANSFORMING GROWTH FACTOR BETA 2 PRECURSOR(TGF-BETA 2)>

s3b10fs.r1	586	5.6e-67	183	608	gi|6678317 ref|NP_033393.1|| transforming growth factor, beta 2>sp|P27090|TGF2_MOUSE TRANSFORMING GROWTH FACTOR BETA 2 PRECURSOR(TGF-BETA 2) >pir||WFMSB2 transforming growth factor beta-2precursor - mouse >emb|CAA40672.1| (X57413) transforming growthfactor-beta2 precurser [Mus musculus]

<involved in pseudohyphal growth>

Contig291	183	1.4e-12	227	562	gi|6323894 ref|NP_013965.1|DFG5| involved in pseudohyphal growth; Dfg5p>sp|Q05031|YM77_YEAST HYPOTHETICAL 50.5 KD PROTEIN IN RNA1-RNT1INTERGENIC REGION >pir||S57605 probable membrane protein YMR238w-yeast (Saccharomyces cerevisiae) >emb|CAA90209.1| (Z49939) unknown[Saccharomyces cerevisiae]

2.defense and secondary metabolites (63)

2.1. trichothecine biosynthesis pathway in F. sporotrichioides

<15-decalonectrin 15-O-acetyltransferase> (Tri3)

Contig797	1413	1.1e-143	115	924	gb|AAD13653.1| (U22463) 15-decalonectrin 15-O-acetyltransferase [Fusarium sporotrichioides]
Contig916	522	2.7e-49	219	551	gb|AAD13653.1| (U22463) 15-decalonectrin 15-O-acetyltransferase [Fusarium sporotrichioides]

<TRICHODIENE OXYGENASE> (Tri4)

Contig1057	2592	1.2e-268	41	1579	sp|Q12612|TRI4_FUSSP TRICHODIENE OXYGENASE (CYTOCHROME P450 58) >pir||S57337trichodiene oxygenase (EC 1.14.-.-) cytochrome P450 CYP58 - fungus(Fusarium sporotrichioides) >gb|AAB72032.1| (U22462)trichodieneoxygenase [Fusarium sporotrichioides]
Contig199	730	2.3e-71	8	430	sp|Q12612|TRI4_FUSSP TRICHODIENE OXYGENASE (CYTOCHROME P450 58) >pir||S57337trichodiene oxygenase (EC 1.14.-.-) cytochrome P450 CYP58 - fungus(Fusarium sporotrichioides) >gb|AAB72032.1| (U22462) trichodieneoxygenase [Fusarium sporotrichioides]
Contig988	500	5.6e-47	209	550	sp|Q12612|TRI4_FUSSP TRICHODIENE OXYGENASE (CYTOCHROME P450 58) >pir||S57337trichodiene oxygenase (EC 1.14.-.-) cytochrome P450 CYP58 - fungus(Fusarium sporotrichioides) >gb|AAB72032.1| (U22462) trichodieneoxygenase [Fusarium sporotrichioides]
Contig230	403	9.8e-37	218	463	sp|Q12612|TRI4_FUSSP TRICHODIENE OXYGENASE (CYTOCHROME P450 58) >pir||S57337trichodiene oxygenase (EC 1.14.-.-) cytochrome P450

					CYP58 - fungus(Fusarium sporotrichioides) >gb|AAB72032.1| (U22462) trichodieneoxygenase [Fusarium sporotrichioides]
<trichodiene synthase> (Tri5)					
Contig1056			2015	1.6e-207	67 1188 sp|P13513|TRI5_FUSSP TRICHODIENE SYNTHASE (SESQUITERPENE CYCLASE) (TS)>pir||SYFUTP trichodiene synthase (EC 4.1.99.6) - fungus (Fusarium sporotrichioides) >gb|AAD13657.1| (M27246) trichodiene synthase[Fusarium sporotrichioides]
Contig899	584	7.2e-56	260	613	sp|P13513|TRI5_FUSSP TRICHODIENE SYNTHASE (SESQUITERPENE CYCLASE) (TS)>pir||SYFUTP trichodiene synthase (EC 4.1.99.6) - fungus (Fusariumsporotrichioides) >gb|AAD13657.1| (M27246) trichodiene synthase[Fusarium sporotrichioides]
Contig12	473	4.1e-44	241	510	sp|P13513|TRI5_FUSSP TRICHODIENE SYNTHASE (SESQUITERPENE CYCLASE) (TS)>pir||SYFUTP trichodiene synthase (EC 4.1.99.6) - fungus (Fusariumsporotrichioides) >gb|AAD13657.1| (M27246) trichodiene synthase[Fusarium sporotrichioides]
Contig14	282	7.4e-24	273	437	sp|P13513|TRI5_FUSSP TRICHODIENE SYNTHASE (SESQUITERPENE CYCLASE) (TS)>pir||SYFUTP trichodiene synthase (EC 4.1.99.6) - fungus (Fusariumsporotrichioides) >gb|AAD13657.1| (M27246) trichodiene synthase[Fusarium sporotrichioides]
Contig27	266	3.7e-22	258	419	sp|P13513|TRI5_FUSSP TRICHODIENE SYNTHASE (SESQUITERPENE CYCLASE) (TS)>pir||SYFUTP trichodiene synthase (EC 4.1.99.6) - fungus (Fusariumsporotrichioides) >gb|AAD13657.1| (M27246) trichodiene synthase[Fusarium sporotrichioides]
<Tri6>					
<T-2 toxin biosynthesis protein> (Tri7 and Tri8)					
Contig1046	2120	1.2e-218	50	1390	gb|AAD13655.1| (U22463) T-2 toxin biosynthesis protein; TRI8 [Fusarium sporotrichioides]
Contig947	1105	4.5e-111	58	696	gb|AAD13654.1 (U22463) T-2 toxin biosynthesis protein; TRI7 [Fusarium sporotrichioides]
Contig793	724	1.1e-70	93	500	gb|AAD13655.1| (U22463) T-2 toxin biosynthesis protein; TRI8 [Fusarium sporotrichioides]
Contig582	375	4.1e-42	294	503	gb|AAD13654.1| (U22463) T-2 toxin biosynthesis protein; TRI7 [Fusarium sporotrichioides]
<trichothecene 3-O-acetyltransferase> (Tri101)					

Contig1053	2259	2.3e-233	120	1496	gb|AAD19745.1| (AF127176) trichothecene 3-O-acetyltransferase [Fusarium sporotrichioides]
<isotrichodermin C-15 hydroxylase> (Tri11)					
Contig928	1699	5.1e-174	56	1042	sp|O13317|TR11_FUSSP ISOTRICHODERMIN C-15 HYDROXYLASE (CYTOCHROME P450 65A1)>gb|AAD12755.1| (AF011355) isotrichodermin C-15 hydroxylase[Fusarium sporotrichioides]
Contig844	408	3.4e-37	283	510	sp|O13317|TR11_FUSSP ISOTRICHODERMIN C-15 HYDROXYLASE (CYTOCHROME P450 65A1)>gb|AAD12755.1| (AF011355) isotrichodermin C-15 hydroxylase[Fusarium sporotrichioides]
Contig1026	271	2e-20	140	1246	sp|O13317|TR11_FUSSP ISOTRICHODERMIN C-15 HYDROXYLASE (CYTOCHROME P450 65A1)>gb|AAD12755.1| (AF011355) isotrichodermin C-15 hydroxylase[Fusarium sporotrichioides]
<trichothecene efflux pump> (Tri12)					
Contig917	1104	5.8e-111	88	750	gb|AAD12756.1| (AF011355) trichothecene efflux pump [Fusarium sporotrichioides]
Contig673	1101	1.2e-110	8	643	gb|AAD12756.1| (AF011355) trichothecene efflux pump [Fusarium sporotrichioides]
Contig145	758	2.6e-74	3	482	gb|AAD12756.1| (AF011355) trichothecene efflux pump [Fusarium sporotrichioides]
<Tri10>					
j3h01fs.r1					
d3g05fs.f1					
<Tri13>					
Contig976-995					gi|13194731|AF330109.1
<Tri14>					
Contig1047	763	0.0	102-1223		gb|AF326571.1|
<Tri16>					
Contig936-697	603	e-171	85-1043		gb|AF327521.1|
2.2. other secondary metabolites					
<cytochrome P450 sterol 14-demethylase>					
e4g01fs.r1	593	7.4e-57	8	502	gb|AAD55135.1|AF0420 (AF042067) eburicol 14-alpha demethylase; cytochromeP450 sterol 14-alpha demethylase; Erg11 [Uncinula necator]
e4g01fs.f1	289	5.4e-24	266	520	gb|AAC97606.1| (AF052515) eburicol 14alpha demethylase; CYP51; cytochrome P450sterol 14-demethylase [Blumeria graminis f. sp. hordei]
<CYTOCHROME P450 3A28>					

Contig976	160	1.5e-08	435	818	sp|P79102|CP3S_BOVIN CYTOCHROME P450 3A28 (CYPIIIA28)>emb|CAA71266.1|(Y10214) cytochrome P450 [Bos taurus]
<aflatoxin biosynthesis protein>					
Contig686	234	3.2e-18	257	721	gb|AAC62645.1| (AF077975) aflatoxin biosynthesis protein; AflJ [Aspergillus flavus]
<sterigmatocystin 7-o-methyltransferase precursor>					
j1h12fs.f1	153	2.1e-09	252	560	sp|Q12120|OMTA_ASPPA STERIGMATOCYSTIN 7-O-METHYLTRANSFERASE PRECURSOR>gb|AAA32699.1| (L25835) O-methyltransferase [Aspergillus flavus]>gb|AAA99509.1| (L25834) O-methyltransferase [Aspergillusparasiticus] >gb|AAA32697.1| (L22091) O-methyltransferase[Aspergillus parasiticus]
<sterigmatocystin biosynthesis protein>					
Contig542	410	1.5e-37	34	597	sp|Q00717|STCT_EMENI PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS PROTEIN STCT>gb|AAC49204.1| (U34740) putative translation elongation factor 1gamma [Emericella nidulans]
<sterigmatocystin biosynthesis p450 monoosygenase stcb>					
d3e03fs.r1	312	4.8e-27	122	487	sp|Q12608|STCB_EMENI PROBABLE STERIGMATOCYSTIN BIOSYNTHESIS P450 MONOOXYGENASESTCB (CYTOCHROME P450 62) >gb|AAC49192.1| (U34740) putative p450monooxygenase [Emericella nidulans]
d3e03fs.f1	296	2.3e-25	73	447	sp|Q12608|STCB_EMENI PROBABLE STERIGMATOCYSTIN BIOSYNTHESIS P450 MONOOXYGENASESTCB (CYTOCHROME P450 62) >gb|AAC49192.1| (U34740) putative p450monooxygenase [Emericella nidulans]
<daunorubicin C-13 ketoreductase>					
Contig384	160	3.1e-10	108	485	pir||C75365 daunorubicin C-13 ketoreductase - Deinococcus radiodurans (strainR1) >gb|AAF11255.1|AE002011_6 (AE002011) daunorubicin C-13ketoreductase [Deinococcus radiodurans]
<gibberellin 7-oxidase>					
b4b04fs.f1	180	9.7e-13	129	431	pir||T09683 gibberellin 7-oxidase (EC 1.14.11.-) - winter squash>gb|AAB64346.1| (U61386) gibberellin 7-oxidase [Cucurbita maxima]
<gibberellin 20-oxidase>					
f3b02fs.r1	123	3.3e-06	140	385	pir||T10222 gibberellin 20-oxidase (EC 1.14.11.-) - Arabidopsis thaliana>emb|CAB45519.1| (AL079350) gibberellin 20-oxidase-Arabidopsisthaliana >emb|CAB81353.1| (AL161563) gibberellin20-oxidase-Arabidopsis thaliana
<gibberellin biosynthesis-related>					
e1b01fs.r1	265	4.8e-22	151	498	gb|AAB06951.1| (U61530) gibberellin biosynthesis-related [Gibberellafujikuroi]

<thiazole biosynthetic enzyme precursor>

k3c06fs.r1	861	3e-85	9	527	sp|P23618|THI4_FUSOX THIAZOLE BIOSYNTHETIC ENZYME PRECURSOR (STRESS-INDUCIBLEPROTEIN STI35) >pir||B37767 stress-inducible protein sti35 - fungus(Fusarium oxysporum) >gb|AAA33341.1| (M33643) STI35 protein[Fusarium oxysporum] >dbj|BAA85305.1| (AB033416) stress-responsivegene product [Fusarium oxysporum]
Contig418	723	1.2e-70	326	787	sp|P23618|THI4_FUSOX THIAZOLE BIOSYNTHETIC ENZYME PRECURSOR (STRESS-INDUCIBLEPROTEIN STI35) >pir||B37767 stress-inducible protein sti35 - fungus(Fusarium oxysporum) >gb|AAA33341.1| (M33643) STI35 protein[Fusarium oxysporum] >dbj|BAA85305.1| (AB033416) stress-responsivegene product [Fusarium oxysporum]

<cephalosporin C acetylhydrolase>

Contig619	347	7.4e-31	181	501	emb|CAB87194.1| (AJ238108) cephalosporin C acetylhydrolase [Acremoniumchrysogenum]

<leukotriene A-4 hydrolase>

Contig489	244	6.5e-19	109	372	pir||T40936 probable leukotriene A-4 hydrolase - fission yeast(Schizosaccharomyces pombe) >emb|CAA22858.1| (AL035259) probableleukotriene a-4 hydrolase [Schizosaccharomyces pombe]

<NosD-nostopeptolide biosynthetic gene cluste>

gla08fs.r1	298	8.6e-24	12	458	gb|AAF17281.1| (AF204805) NosD [Nostoc sp. GSV224]

<1-hydroxy-2-naphthoate dioxygenase-A phenanthrene degradative gene cluster>

n3e02fs.f1	228	2.3e-17	252	578	dbj|BAA76331.1| (AB024945) 1-hydroxy-2-naphthoate dioxygenase [Alcaligenesfaecalis]

<deacetylcephalosporin C acetyltransferase>

q2c09fs.r1	265	4.8e-22	19	438	gb|AAD30471.1| (AF124929) putative deacetylcephalosporin C acetyltransferase[Streptomyces clavuligerus]
q2c09fs.f1	185	3.3e-13	136	480	gb|AAD30471.1| (AF124929) putative deacetylcephalosporin C acetyltransferase[Streptomyces clavuligerus]

<PHYTOENE SYNTHASE>

c2e03fs.f1	221	2e-16	153	458	gb|AAA19428.1| (L27652) phytoene synthase [Neurospora crassa]

<phytoene dehydrogenase>

r3b07fs.r1	533	1.9e-50	83	640	sp|P48537|CRTI_CERNC PHYTOENE DEHYDROGENASE (PHYTOENE DESATURASE)>gb|AAB86988.1| (U03903) phytoene dehydrogenase [Cercosporanicotianae]

<6,7-dimethyl-8-ribityllumazine synthase>

s1e05fs.r1	563	1.2e-53	88	594	gb|AAD55372.1|AF1484 (AF148449) 6,7-dimethyl-8-ribityllumazine synthase[Magnaporthe grisea]

<PhzG-pyocyanine biosynthesis operon>

o4e08fs.f1 125 4.1e-07 313 525 gb|AAC64493.2| (AF005404) PhzG [Pseudomonas aeruginosa]
<snodprot1-a protein produced during infection of wheat leaves by Phaeosphaeria nodorum>
Contig776 360 3.1e-32 235 606 sp|O74238|SNP1_PHANO PROTEIN SNODPROT1 PRECURSOR >gb|AAC26870.1| (AF074941)snodprot1 [Phaeosphaeria nodorum]

<versicolorin B synthase>
r4d11fs.f1 322 2.6e-27 135 518 gb|AAC49318.1| (U51327) versicolorin B synthase [Aspergillus parasiticus]>gb|AAC49319.1| (U51328) versicolorin B synthase [Aspergillusparasiticus] >gb|AAF26279.1|AF169016_2 (AF169016) versicolorin Bsynthase [Aspergillus parasiticus]

<HYGROMYCIN-B KINASE>
Contig710 375 9.6e-34 415 633 sp|P00557|KHYB_ECOLI HYGROMYCIN-B KINASE (HYGROMYCIN B PHOSPHOTRANSFERASE)(APH(7'')) >pir||WGECH hygromycin-B kinase (EC 2.7.1.119)-Escherichia coli plasmids >emb|CAA24743.1| (V01499) aph(4)[Escherichia coli] >gb|AAA92252.1| (K01193) hygromycin Bphosphotransferase [Plasmid pJR225] >emb|CAA61952.1| (X89856)hygromycin phosphotransferase [Vaccinia virus] >emb|CAA61953.1|(X89857) hygromycin phosphotransferase

2.3. defense
<VEGETABLE INCOMPATIBILITY PROTEIN HET-E-1-vegetative incompatibility,defense>
Contig506 343 1e-44 213 674 sp|Q00808|HET1_PODAN VEGETATIBLE INCOMPATIBILITY PROTEIN HET-E-1 >pir||T18521beta transducin-like protein - Podospora anserina >gb|AAA85775.1|(L28125) beta transducin-like protein [Podospora anserina]
i2c06fs.f1 184 5.9e-12 25 411 sp|Q00808|HET1_PODAN VEGETATIBLE INCOMPATIBILITY PROTEIN HET-E-1 >pir||T18521beta transducin-like protein - Podospora anserina >gb|AAA85775.1|(L28125) beta transducin-like protein [Podospora anserina]

<pisatin demethylase-inactivates plant compound pisatin>
a1g01fs.r1 143 3.8e-08 35 430 sp|P38364|PID6_FUSSO PISATIN DEMETHYLASE (CYTOCHROME P450 57A2) >pir||S34286pisatin demethylase PDA6-1 - fungus (Nectria haematococca)>emb|CAA51665.1| (X73145) pisatin demethylase [Haematonectriahaematococca]

<D-AMINO ACID OXIDASE-oxidation of cephalosporin C>
n2f04fs.f1 650 7e-63 120 509 sp|P24552|OXDA_FUSSO D-AMINO ACID OXIDASE (DAMOX) (DAO) (DAAO)>pir||JX0152D-amino-acid oxidase - fungus (Fusarium solani)>dbj|BAA00692.1|(D00809) D-amino acid oxidase [Haematonectria haematococca]

n2f04fs.r1	576	5.2e-55	96	515	sp|P24552|OXDA_FUSSO D-AMINO ACID OXIDASE (DAMOX) (DAO) (DAAO)>pir||JX0152D-amino-acid oxidase - fungus (Fusarium solani)>dbj|BAA00692.1|(D00809) D-amino acid oxidase [Haematonectria haematococca]
<oligomycin sensitivity conferring protein>					
Contig319	282	4.7e-24	131	487	gb|AAC64903.1| (AF008185) oligomycin sensitivity conferring protein[Kluyveromyces lactis]
<aureobasidin-resistance protein>					
Contig692	271	2.1e-22	381	827	gb|AAD22750.1|AF0766 (AF076692) aureobasidin-resistance protein; Aur1homolog; Aur1 [Aspergillus fumigatus]
o4a03fs.f1	175	8.7e-12	252	527	gb|AAD22750.1|AF0766 (AF076692) aureobasidin-resistance protein; Aur1homolog; Aur1 [Aspergillus fumigatus]
<Syringomycin response protein 2>					
r3e09fs.r1	102	0.028	211	552	gi|6320503 ref|NP_010583.1|SUR2| Syringomycin response protein 2; Sur2p>sp|P38992|SUR2_YEAST SUR2 PROTEIN (SYRINGOMYCIN RESPONSE PROTEIN2)>pir||S48533 SUR2 protein - yeast (Saccharomyces cerevisiae)>gb|AAA16608.1| (U07171) Sur2p [Saccharomyces cerevisiae]>gb|AAB64733.1| (U28374) Sur2p:syringomycin response protein 2[Saccharomyces cerevisiae] >gb|AAB41115.1| (U10427) Syr2p[Saccharomyces cerevisiae]
<66 KD STRESS PROTEIN>					
Contig268	240	1.7e-18	129	497	sp|P90587|WD66_PHYPO 66 KD STRESS PROTEIN (P66) >pir||JE0238 stress proteinp66 - slime mold (Physarum polycephalum) >gb|AAC26321.1| (U86011)66-kDa stress protein p66 [Physarum polycephalum]
a2d03fs.r1	226	5.4e-17	124	426	sp|P90587|WD66_PHYPO 66 KD STRESS PROTEIN (P66) >pir||JE0238 stress proteinp66 - slime mold (Physarum polycephalum) >gb|AAC26321.1| (U86011)66-kDa stress protein p66 [Physarum polycephalum]
3. detoxification (4)					
<catalase>					
n2f02fs.f1	429	8.4e-39	115	504	emb|CAA67268.1| (X98718) T-catalase [Mycobacterium smegmatis]
<manganese superoxide dismutase precursor>					
Contig1035	708	4.9e-69	144	824	gb|AAD28503.1|AF1188 (AF118809) manganese superoxide dismutase precursor[Neurospora crassa]
<epoxide hydrolase>					
Contig683	228	2.8e-17	145	450	pir||JC4711 epoxide hydrolase (EC 3.3.2.3) - human >emb|CAA65751.1| (X97024)epoxide hydrolase [Homo sapiens]
<spindle poison sensitivity protein>					

r4g06fs.r1 118 0.001 146 325 pir||T11624 spindle poison sensitivity protein - fission yeast(Schizosaccharomyces pombe) >emb|CAB16391.1| (Z99260) spindlepoison sensitivity related protein. [Schizosaccharomyces pombe]

4.salt tolerance (3)

<heavy metal tolerance protein precursor>

j2e05fs.r1 169 5.8e-11 2 472 pir||T41582 heavy metal tolerance protein precursor - fission yeast(Schizosaccharomyces pombe) (fragment) >emb|CAA20865.1| (AL031546)heavy metal tolerance protein precursor [Schizosaccharomyces pombe]

<halotolerance protein>

i4h07fs.r1 193 4.4e-14 161 496 pir||T41127 halotolerance protein - fission yeast (Schizosaccharomyces pombe)>emb|CAA22778.1| (AL035210) halotolerance protein homolog; putativeinositol metabolism [Schizosaccharomyces pombe]

<metal resistance protein>

l1b10fs.f1 371 8e-32 184 525 gi|6320339 ref|NP_010419.1|YCF1| metal resistance protein,similar tomultidrug resistance proteins and cystic fibrosis protein CFTR;Ycf1p >sp|P39109|YCFI_YEAST METAL RESISTANCE PROTEIN YCF1 (YEASTCADMIUM FACTOR 1) >pir||S51863 cadmium resistance protein YCF1 -yeast (Saccharomyces cerevisiae) >emb|CAA88217.1| (Z48179) unknown[Saccharomyces cerevisiae]

5. starvation response (2)

<RVS167 protein-reduces viability upon starvation>

g4h08fs.r1 449 1.5e-41 111 485 emb|CAB77648.1| (AJ390509) RVS167 protein [Candida albicans]

g4h08fs.f1 132 5e-07 404 490 pir||T11684 RVS167 protein homolog - fission yeast (Schizosaccharomyces pombe)>emb|CAA20768.1| (AL031536) hypothetical protein[Schizosaccharomyces pombe]

6.DNA repair (11)

<UV-endonuclease-UV damage DNA repair>

j2f08fs.r1 559 3.2e-53 24 470 dbj|BAA74539.1| (D11392) UV-endonuclease [Neurospora crassa]

j2f08fs.f1 270 1.1e-21 152 457 dbj|BAA74539.1| (D11392) UV-endonuclease [Neurospora crassa]

<CDC20 protein-is required for recovery from DNA damage>

Contig497 110 0.00033 441 527 pir||T41719 CDC20 protein homolog - fission yeast (Schizosaccharomyces pombe)>gb|AAC49621.1| (U77983) WD-domain protein [Schizosaccharomycespombe] >emb|CAB57442.1| (AL121770) wd-

					domain protein;CDC20/p55CDC/Fizzy homolog [Schizosaccharomyces pombe]
<extracellular putative DNase-induces disease resistance responses in peas>					
n4d11fs.r1	175	1.5e-12	284	499	gb|AAD53090.1| (AF175291) extracellular putative DNase [Fusarium solani]
<Hmp1-mismatch base pair and cruciform DNA recognition protein>					
t2d01fs.r1	126	2e-07	265	498	gb|AAA86754.1| (U39049) Hmp1 [Ustilago maydis]
<exonuclease>					
j3e05fs.f1	200	3.2e-14	14	322	pir||T39001 probable exonuclease - fission yeast (Schizosaccharomyces pombe)>emb|CAA22588.1| (AL034583) putative exonuclease[Schizosaccharomyces pombe]
j3e05fs.r1	120	8.5e-05	80	406	pir||T39001 probable exonuclease - fission yeast (Schizosaccharomyces pombe)>emb|CAA22588.1| (AL034583) putative exonuclease[Schizosaccharomyces pombe]
<DNA REPAIR PROTEIN>					
g1d11fs.r1	132	4.2e-07	87	332	sp|P26306|RAD9_SCHPO DNA REPAIR PROTEIN RAD9 >pir||S20986 rad9 protein -fission yeast (Schizosaccharomyces pombe) >pir||S16441 rad9 protein- fission yeast (Schizosaccharomyces pombe) >emb|CAA41189.1|(X58231) rad9 protein [Schizosaccharomyces pombe] >emb|CAA45919.1|(X64648) rad9 [Schizosaccharomyces pombe] >emb|CAA54491.1| (X77276)rad9 [Schizosaccharomyces pombe] >emb|CAB65808.1| (AL136235) DNArepair protein
<uv excision repair protein rad23>					
Contig407	205	1.1e-33	192	407	pir||T40115 uv excision repair protein rad23 homolog - fission yeast(Schizosaccharomyces pombe) >emb|CAA21170.1| (AL031788) nucleotideexcision repair protein yeast rad23/ human HHR23A homolog[Schizosaccharomyces pombe] >gb|AAD51975.1|AF174293_1 (AF174293)Rhp23 [Schizosaccharomyces pombe]
f3b06fs.f1	180	1.7e-12	300	503	pir||T40115 uv excision repair protein rad23 homolog - fission yeast(Schizosaccharomyces pombe) >emb|CAA21170.1| (AL031788) nucleotideexcision repair protein yeast rad23/ human HHR23A homolog[Schizosaccharomyces pombe] >gb|AAD51975.1|AF174293_1 (AF174293)Rhp23 [Schizosaccharomyces pombe]
<dna repair helicase>					
d3c04fs.r1	678	7.9e-66	8	505	pir||T37821 probable dna repair helicase - fission yeast (Schizosaccharomycespombe) >emb|CAB11506.1| (Z98849) putative dna repair helicase[Schizosaccharomyces pombe]

7. allergen and immune system proteins (9)

7.1. allergen

<ALLERGEN ALT A 7>

Contig869	685	1.2e-66	258	860	sp|P42058|ALA7_ALTAL MINOR ALLERGEN ALT A 7 (ALT A VII) >pir||S43111 minorallergen - Alternaria alternata >emb|CAA55069.1| (X78225) minorallergen [Alternaria alternata]

<rAsp f 7>

Contig41	147	1.4e-09	360	527	emb|CAA11255.1| (AJ223315) rAsp f 7 [Aspergillus fumigatus]

<rAsp f 9-Recombinant Aspergillus fumigatus allergens>

Contig1044	679	6.3e-66	259	984	emb|CAA11266.1| (AJ223327) rAsp f 9 [Aspergillus fumigatus]
e2b05fs.r1	417	3e-38	24	482	emb|CAA11266.1| (AJ223327) rAsp f 9 [Aspergillus fumigatus]
l2h04fs.r1	361	3e-32	102	539	emb|CAA11266.1| (AJ223327) rAsp f 9 [Aspergillus fumigatus]
l2h04fs.f1	344	1.9e-30	157	540	emb|CAA11266.1| (AJ223327) rAsp f 9 [Aspergillus fumigatus]
Contig1030	101	0.00045	504	590	emb|CAA11266.1| (AJ223327) rAsp f 9 [Aspergillus fumigatus]

7.2. immune system proteins (2)

<immunoglobulin heavy chain binding protein homolog>

Contig567	433	1.5e-39	150	539	sp|P78695|GR78_NEUCR 78 KD GLUCOSE-REGULATED PROTEIN HOMOLOG PRECURSOR (GRP78) (IMMUNOGLOBULIN HEAVY CHAIN BINDING PROTEIN HOMOLOG) (BIP)>emb|CAA70214.1| (Y09011) grp78 homologue [Neurospora crassa]

<NAALADase II protein-a marker of prostatic carcinomas>

g2c06fs.r1	213	2.1e-15	7	396	emb|CAB39967.1| (AJ012370) NAALADase II protein [Homo sapiens]

8. tumor protein and tumor suppressor

8.1. tumor protein (8)

<cortactin-oncogene EMS1 product>

d4d10fs.f1	193	1.6e-13	183	353	gi|4885205 ref|NP_005222.1|| cortactin; oncogene EMS1>sp|Q14247|SRC8_HUMANSRC SUBSTRATE CORTACTIN (AMPLAXIN) (EMS1 ONCOGENE)>pir||A48063mammary tumor/squamous cell carcinoma-associated protein EMS1-human >gb|AAA58455.1| (M98343) amplaxin [Homo sapiens]>gb|AAB26248.1| EMS1 gene product [human, Peptide, 550 aa]

<tumor protein homolog>

Contig155	524	1.4e-49	88	597	gi|6322794 ref|NP_012867.1|YKL056C| Ykl056cp>sp|P35691|TCTP_YEASTTRANSLATIONALLY CONTROLLED TUMOR PROTEIN HOMOLOG (TCTP)>pir||S37878 IgE-dependent histamine-releasing

factor homolog -yeast (Saccharomyces cerevisiae) >emb|CAA53416.1| (X75781) Humantumor protein homologue [Saccharomyces cerevisiae] >emb|CAA81893.1|(Z28056) ORF YKL056c [Saccharomyces cerevisiae] >prf||2206495L ORF[Saccharomyces

e1a10fs.r1 161 2.2e-09 120 473 pir||F71414 hypothetical protein - Arabidopsis thaliana >emb|CAB10288.1|(Z97337) hypothetical protein [Arabidopsis thaliana]>emb|CAB78551.1| (AL161540) hypothetical protein [Arabidopsisthaliana]

<25 kda ras-related protein-proto-oncogene product>

d2c01fs.f1 97 0.00028 400 456 gb|AAB20486.1| Ran=25 kda ras-related protein [human, Peptide Partial, 36 aa,segment 5 of 5]

<hepatopoietin>

h4e04fs.r1 159 8.1e-11 155 496 gb|AAD56538.1|AF1838 (AF183892) hepatopoietin [Homo sapiens]

8.2. tumor suppressor

<tumor suppressor protein>

j4b05fs.r1 241 3.5e-18 8 490 gi|6319362 ref|NP_009444.1|SRO77| yeast homolog of the Drosphila tumorsuppressor, lethal giant larvae; Sro77p >sp|P38163|SNI2_YEAST SNI2PROTEIN (SRO77 PROTEIN) >pir||S45389 probable membrane proteinYBL106c - yeast (Saccharomyces cerevisiae) >emb|CAA55989.1|(X79489) E-1010 protein [Saccharomyces cerevisiae] >emb|CAA84933.1|(Z35867) ORF YBL106c [Saccharomyces cerevisiae]

<saframycin Mx1 synthetase A-a DNA-binding antibiotic and antitumour agent>

f1c07fs.r1 128 1.2e-05 234 488 pir||T18552 saframycin Mx1 synthetase A - Myxococcus xanthus>gb|AAC44129.1|(U24657) saframycin Mx1 synthetase A [Myxococcus xanthus]

<gene N33 protein-N33, candidate tumor suppressor gene>

e4d07fs.r1 131 3.7e-07 271 519 sp|Q13454|N33_HUMAN N33 PROTEIN >pir||G02297 gene N33 protein - human>gb|AAB18374.1| (U42349) 39 kDa encoded by N33 [Homo sapiens]

9. multidrug resistance (7)

<multidrug resistance protein MDR>

c4h04fs.f1 378 1.2e-32 146 445 gb|AAD43626.1|AF0714 (AF071411) multidrug resistance protein MDR[Emericella nidulans] >gb|AAF29805.1|AF173826_1 (AF173826)ABC-transporter [Emericella nidulans]

<multidrug resistance related protein 2>

k2h08fs.f1 191 1.1e-12 119 394 gb|AAB07022.1| (U66261) multidrug resistance related protein 2 [Caenorhabditiselegans]

<TETRACYCLINE RESISTANCE PROTEIN>

Contig608	487	1.4e-45	656	1000	gb|AAC53625.1| (U35134) tetracycline resistance protein [Cloning vectorpBSL190] >gb|AAC53627.1| (U35135) tetracycline resistance protein[Cloning vector pBSL193]

<mfs-multidrug-resistance transporter>

Contig307	126	2.8e-06	39	170	emb|CAB69830.1| (AJ132188) mfs-multidrug-resistance transporter [Gibberellapulicaris]

<facilitator family multi-drug resistance protein>

Contig496	241	1.2e-18	259	516	pir||T39346 probable major facilitator family multi-drug resistance protein -fission yeast (Schizosaccharomyces pombe) >emb|CAA22200.1|(AL034353) putative MSF transporter [Schizosaccharomyces pombe]
Contig647	168	9.4e-11	259	480	pir||T39346 probable major facilitator family multi-drug resistance protein -fission yeast (Schizosaccharomyces pombe)>emb|CAA22200.1|(AL034353) putative MSF transporter [Schizosaccharomyces pombe]

<protein involved in drug resistance>

b2d08fs.r1	202	3.9e-14	48	356	gi|6324592 ref|NP_014661.1|ROD1| involved in drug resistance;Rod1p>sp|Q02805|ROD1_YEAST ROD1 PROTEIN >pir||S54624 ROD1 protein -yeast (Saccharomyces cerevisiae) >emb|CAA60767.1| (X87331) ORFOR26.08 [Saccharomyces cerevisiae] >emb|CAA99208.1| (Z74926) ORFYOR018w [Saccharomyces cerevisiae]

10. cell reaction to environment (1)

<wc-1-the central regulator of blue light responses>

i3e12fs.r1	386	1.3e-33	224	469	emb|CAA63964.2| (X94300) wc-1 [Neurospora crassa]

III.2. Cell signalling, signal transduction and secondary messenger

1. PHOSPHATASES (14)

<PROTEIN PHOSPHATASE>

d2h11fs.r1	315	2.2e-27	160	501	sp|Q09173|P2C3_SCHPO PROTEIN PHOSPHATASE 2C HOMOLOG 3 (PP2C-3)>pir||S62462hypothetical protein SPAC2G11.07c - fissionyeast(Schizosaccharomyces pombe) >pir||T38573 protein phosphatase 2chomolog 3 - fission yeast (Schizosaccharomyces pombe)>emb|CAA91172.1| (Z54354) protein phosphatase 2c homolog 3[Schizosaccharomyces pombe]

<serine/threonine phosphatase>

j1f04fs.r1	788	1.6e-77	23 460	sp|P48580|P2A1_NEUCR SERINE/THREONINE PROTEIN PHOSPHATASE PP2A CATALYTICSUBUNIT >pir||S60471 phosphoprotein phosphatase (EC 3.1.3.16) type2A catalytic chain - Neurospora crassa >emb|CAA58573.1| (X83593)phosphoprotein phosphatase [Neurospora crassa]
d2c06fs.r1	771	1.1e-75	15 479	gb|AAD09995.1| (AF071751) protein phosphatase-Z-like serine/threonine proteinphosphatase [Neurospora crassa] >gb|AAD09996.1| (AF071752) proteinphosphatase-Z-like serine/threonine protein phosphatase [Neurosporacrassa]
i1c10fs.r1	514	1.8e-48	21 377	sp|P49576|PPX1_PARTE SERINE/THREONINE PROTEIN PHOSPHATASE PP-X HOMOLOG>gb|AAA75081.1| (U31445) PPX homolog [Paramecium tetraurelia]
n1f02fs.r1	453	4.9e-42	8 490	dbj|BAA93675.1| (AB027711) Ser/Thr protein phosphatase 2A regulatory subunit A[Lentinula edodes]
q2d04fs.r1	357	7.8e-32	26 247	gb|AAC05275.1| (AF049853) serine/threonine protein phosphatase type 1[Neurospora crassa]
a3h11fs.r1	287	5.1e-24	238 459	gb|AAB65138.1| (U89985) serine/threonine protein phosphatase PPT1 [Neurosporacrassa]
Contig129	259	9.6e-21	293 451	sp|Q05681|P2B_NEUCR SERINE/THREONINE PROTEIN PHOSPHATASE 2B CATALYTIC SUBUNIT(CALMODULIN-DEPENDENT CALCINEURIN A SUBUNIT)>pir||A40942phosphoprotein phosphatase (EC 3.1.3.16) 3-alpha catalytic chain -Neurospora crassa >gb|AAA33565.1| (M73032) calmodulin-dependentprotein phosphatase [Neurospora crassa]
Contig572	170	2.5e-19	513 635	sp|Q05681|P2B_NEUCR SERINE/THREONINE PROTEIN PHOSPHATASE 2B CATALYTIC SUBUNIT(CALMODULIN-DEPENDENT CALCINEURIN A SUBUNIT)>pir||A40942phosphoprotein phosphatase (EC 3.1.3.16) 3-alpha catalytic chain-Neurospora crassa >gb|AAA33565.1| (M73032) calmodulin-dependentprotein phosphatase [Neurospora crassa]
a3h11fs.f1	228	2e-17	319 483	gb|AAB65138.1| (U89985) serine/threonine protein phosphatase PPT1 [Neurosporacrassa]
j1h11fs.r1	107	0.0003	316 426	pir||S71203 serine/threonine protein phosphatase type 2A regulatory chain A -Arabidopsis thaliana >gb|AAB60713.1| (U27299)serine/threonineprotein phosphatase type 2A regulatory subunit A [Arabidopsisthaliana]
<protein phosphotase 2a 65kd regulatory>				
n1f02fs.f1	256	2.9e-20	252 515	pir||T39246 protein phosphotase 2a 65kd regulatory sububit - fission yeast(Schizosaccharomyces pombe) >emb|CAB55176.1| (AL117210)proteinphosphotase 2a 65kd regulatory sububit [Schizosaccharomyces pombe]

<dual-specificity map kinase phosphatase>

b1d01fs.r1	125	9.4e-07	107	229	pir||T39517 dual-specificity map kinase phosphatase - fission yeast(Schizosaccharomyces pombe) >dbj|BAA22897.1| (D82022)dual-specificity MAP kinase phosphatase [Schizosaccharomyces pombe]>emb|CAA20049.1| (AL031154) dual-specificity map kinase phosphatase[Schizosaccharomyces pombe]
Contig675	96	0.0069	274	453	emb|CAA41661.1| (X58858) orf 5' to PPH3 [Saccharomyces cerevisiae]

2. Kinases (25)

<protein kinase>

n1a02fs.r1	276	8.8e-22	12	485	emb|CAA69030.1| (Y07750) protein kinase [Schizosaccharomyces pombe]
r3g06fs.f1	244	7e-20	329	517	gb|AAB72017.1| (U52963) mitogen-activated protein kinase [Fusarium solani]
i1d02fs.r1	176	1.3e-12	26	403	sp|P38937|CUT8_SCHPO CUT8 PROTEIN >pir||T11593 protein kinase (EC 2.7.1.37)-fission yeast (Schizosaccharomyces pombe) >dbj|BAA06550.1| (D31772)ORF [Schizosaccharomyces pombe] >emb|CAA97343.1| (Z73099) cut8protein [Schizosaccharomyces pombe]
i1b03fs.r1	155	8.1e-09	67	426	gb|AAA35214.1| (M59835) protein kinase [Saccharomyces cerevisiae]

<protein kinase C>

j2h03fs.r1	663	2.3e-63	8	409	sp|P87253|KPC1_NEUCR PROTEIN KINASE C-LIKE >emb|CAA72731.1| (Y12002)proteinkinase C homologue [Neurospora crassa]
b3e12fs.f1	145	2.2e-08	202	351	pir||T24944 hypothetical protein ZK1307.8 - Caenorhabditis elegans>emb|CAA87420.1| (Z47356) similar to protein kinase C substrate;cDNA EST EMBL:M75788 comes from this gene; cDNA EST EMBL:D71530comes from this gene; cDNA EST EMBL:C08471 comes from this gene;cDNA EST yk427a3.3 comes from this gene; cDNA EST yk427a3.5 comesfrom this> >emb|CAA87438.1| (Z47358)similar to protein kinase Csubstrate; cDNA EST

<MAP kinase>

s3h05fs.r1	455	3.3e-42	11	598	gb|AAC49521.2| (U70134) pathogenicity MAP kinase 1; Pmk1; MAP kinase homolog[Magnaporthe grisea]

<MAP KINASE HOG1>

p3g07fs.f1	150	3.6e-09	441	569	sp|Q92207|HOG1_CANAL MITOGEN-ACTIVATED PROTEIN KINASE HOG1 (MAP KINASE HOG1)>emb|CAA62214.1| (X90586) unnamed protein product [Candidaalbicans]

<serine/threonine-protein kinase>

f3f04fs.r1	798	1.6e-78	20	493	gb|AAB04130.1| (U61839) serine/threonine protein kinase FSK [Haematonectriahaematococca]
l3d11fs.r1	703	1.8e-68	5	466	gb|AAC04357.1| (AF046923) serine/threonine protein kinase [Colletotrichumtrifolii]
d3h06fs.r1	599	1.8e-57	16	447	sp|O14019|KDPG_SCHPO PROBABLE SERINE/THREONINE-PROTEIN KINASE C29A4.16>pir||T38473 probable serine/threonine-specific protein kinase (EC2.7.1.-) - fission yeast (Schizosaccharomyces pombe)>emb|CAB10142.1| (Z97210) probable serine/threonine-protein kinase(EC 2.7.1.-) [Schizosaccharomyces pombe]
e3h12fs.r1	580	2.2e-54	6	512	sp|Q92212|ST20_CANAL SERINE/THREONINE-PROTEIN KINASE STE20 HOMOLOG>pir||T18259 serine/threonine protein kinase homolog - yeast(Candida albicans) >gb|AAB38875.1| (U73457) Cst20p [Candidaalbicans]
g3a09fs.r1	484	2.6e-45	39	482	sp|O42626|NRC2_NEUCR SERINE/THREONINE-PROTEIN KINASE NRC-2 (NONREPRESSIBLECONIDIATION PROTEIN 2) >gb|AAC21677.1| (AF034260) protein kinaseNRC-2 [Neurospora crassa]
i2e07fs.r1	383	7.6e-34	9	416	sp|Q10156|KATB_SCHPO PROBABLE SERINE/THREONINE-PROTEIN KINASE C1D4.11C>pir||T38052 probable protein kinase - fission yeast(Schizosaccharomyces pombe) >emb|CAA93220.1| (Z69239) probableprotein kinase [Schizosaccharomyces pombe]
o4e03fs.r1	252	3.2e-20	165	515	gi|6319327 ref|NP_009410.1|KIN3| protein kinase; Kin3p>sp|P22209|KIN3_YEASTSERINE/THREONINE-PROTEIN KINASE KIN3 >pir||S23580 probable proteinkinase KIN3 (EC 2.7.1.-) - yeast (Saccharomyces cerevisiae)>gb|AAB22795.1| FUN52=protein kinase homolog [Saccharomycescerevisiae=yeast, Peptide, 435 aa] >emb|CAA43042.1| (X60549)non-essential protein kinase [Saccharomyces cerevisiae]>gb|AAC04964.1| (L22015)
t4f04fs.f1	237	8.2e-18	346	591	sp|O14427|CLA4_CANAL SERINE/THREONINE-PROTEIN KINASE CLA4>gb|AAB68613.1|(U87996) CLA4 protein kinase homolog [Candida albicans]
Contig287	139	1.3e-07	471	578	sp|O42626|NRC2_NEUCR SERINE/THREONINE-PROTEIN KINASE NRC-2 (NONREPRESSIBLECONIDIATION PROTEIN 2) >gb|AAC21677.1| (AF034260) protein kinaseNRC-2 [Neurospora crassa]
<calmodulin-dependent protein kinase>					
Contig753	1222	1.8e-123	205	1023	sp|O14408|KCC1_METAN CALCIUM/CALMODULIN-DEPENDENT PROTEIN KINASE>gb|AAB80685.1| (U28358) serine/threoninecalcium/calmodulin-dependent protein kinase [Metarhiziumanisopliae]

Contig866	757	2.9e-74	59	505	sp|Q02052|CALM_NEUCR CALMODULIN >pir||S58709 calmodulin - Neurospora crassa>emb|CAA50271.1| (X70923) calmodulin [Neurospora crassa]>gb|AAA33564.1| (L02964) calmodulin [Neurospora crassa]>gb|AAA51652.1| (U15993) calmodulin [Colletotrichum trifolii]>gb|AAC62516.1| (AF034964) calmodulin; CgCaM [Colletotrichumgloeosporioides]
m4a06fs.r1	236	1.6e-18	325	483	gb|AAC62515.1| (AF034963) calmodulin-dependent protein kinase; CgCMK[Colletotrichum gloeosporioides]
<PduW-propionate kinase>					
k2c10fs.f1	167	5.6e-11	152	523	gb|AAD39021.1| (AF026270) PduW [Salmonella enterica serovar Typhimurium]
<GM3 synthase>					
k2h01fs.r1	830	6.4e-82	18	488	dbj|BAA33491.1| (AB018048) GM3 synthase [Mus musculus]>dbj|BAA76467.1|(AB013302) alpha 2,3 sialyltransferase [Mus musculus]
<protein tyrosine kinase 9 related protein>					
i1b09fs.f1	205	2.1e-15	97	435	gi|6755224 ref|NP_036006.1|| protein tyrosine kinase 9 related protein>emb|CAB38083.1| (Y17808) A6 related protein [Mus musculus]
<Guanylate Kinase>					
s2g02fs.f1	286	2.6e-24	190	510	pdb|1GKY| Guanylate Kinase (E.C.2.7.4.8) Complex With GuanosineMonophosphate
<muscle-specific serine kinase 1>					
m1a06fs.r1	250	8.1e-20	217	444	gb|AAD00539.1| (U82808) muscle-specific serine kinase 1 [Homo sapiens]
<casein kinase-1>					
k1e05fs.r1	250	1.9e-20	292	447	sp|P40235|HHP1_SCHPO CASEIN KINASE I HOMOLOG HHP1 >pir||S46357 casein kinase-1homolog hhp1 - fission yeast (Schizosaccharomyces pombe)>pir||JC2547 casein kinase-1 homolog hhp1 - Yeast(Schizosaccharomyces pombe) >pir||T40400 casein kinase I homolog -fission yeast (Schizosaccharomyces pombe)>emb|CAA55473.1| (X78871)Hhp1 protein kinase [Schizosaccharomyces pombe]>gb|AAA21544.1|(U10863) casein kinase-1
3. cAMP-secondary messenger (3)					
<14-3-3-interacts with CAP (adenylyl cyclase-associated protein)>					
Contig139	627	1.7e-60	27	545	dbj|BAA89421.1| (AB029307) 14-3-3 [Lentinula edodes]>dbj|BAA89422.1|(AB029308) 14-3-3 [Lentinula edodes]
Contig712	518	6.2e-49	196	522	sp|Q99002|1433_TRIHA 14-3-3 PROTEIN HOMOLOG (TH1433) >gb|AAB17101.1| (U24158)14.3.3. protein [Trichoderma harzianum]

Contig711	226	5.4e-18	361	507	sp|Q99002|1433_TRIHA 14-3-3 PROTEIN HOMOLOG (TH1433) >gb|AAB17101.1| (U24158)14.3.3. protein [Trichoderma harzianum]
4. G protein (13)					
<GTP-binding protein>					
Contig747	965	2.7e-96	103	669	sp|P78976|SAR1_TRIRE GTP-BINDING PROTEIN SAR1 >emb|CAA69926.1| (Y08636)sar1[Hypocrea jecorina]
Contig921	811	5.4e-80	110	643	sp|Q09914|RHO1_SCHPO RHO1 PROTEIN >pir||JC4044 GTP-binding protein rho1-fission yeast (Schizosaccharomyces pombe) >pir||S62576 hypotheticalprotein SPAC1F7.04 - fission yeast (Schizosaccharomyces pombe)>dbj|BAA07377.1| (D38180) Rho1 [Schizosaccharomyces pombe]>emb|CAA91951.1| (Z67998) rho1 protein [Schizosaccharomyces pombe]
e4b10fs.r1	616	3e-59	62	523	gi|6324381 ref|NP_014451.1|YNR053C| Ynr053cp>sp|P53742|YN8U_YEASTHYPOTHETICAL GTP-BINDING PROTEIN IN POP2-HOL1 INTERGENIC REGION>pir||S63384 hypothetical protein YNR053c - yeast (Saccharomycescerevisiae) >emb|CAA96334.1| (Z71668) ORF YNR053c [Saccharomycescerevisiae]
d1c08fs.r1	570	2.1e-54	152	499	gi|6323324 ref|NP_013396.1|GSP1| GTP-binding protein; Gsp1p>sp|P32835|GSP1_YEAST GTP-BINDING NUCLEAR PROTEIN GSP1/CNR1>pir||S35504 GTP-binding protein GSP1 - yeast (Saccharomycescerevisiae) >gb|AAA34653.1| (L08690) GTP-binding protein[Saccharomyces cerevisiae] >emb|CAA50747.1| (X71945) CNR2[Saccharomyces cerevisiae] >gb|AAB67339.1| (U17243)GTP-bindingnuclear protein. Highly similar to GSP2_YEAST.
Contig421	547	5.2e-52	169	522	sp|P17609|YPT2_SCHPO YPT1-RELATED PROTEIN 2 >pir||S12790 GTP-binding proteinypt2 - fission yeast (Schizosaccharomyces pombe)>emb|CAA36707.1|(X52469) ypt2 gene product (AA 1-200) [Schizosaccharomyces pombe]>emb|CAA37045.1| (X52864) ypt-related protein (AA1-200)[Schizosaccharomyces pombe] >emb|CAB16405.1| (Z99262)ypt1-relatedprotein 2 [Schizosaccharomyces pombe]
l4h08fs.r1	417	8.9e-38	9	521	sp|O13869|YE14_SCHPO PUTATIVE GTP-BINDING PROTEIN C1B3.04C>pir||T38022probable GTP-binding protein - fission yeast (Schizosaccharomycespombe) >emb|CAB11233.1| (Z98598) putative gtp binding protein,gtpase; Elongation factor Tu family [Schizosaccharomyces pombe]
m2g02fs.r1	328	9.5e-29	137	496	gb|AAC33135.1| (AF035177) GTP-binding protein [Oncorhynchus tshawytsha]

a2b03fs.r1	234	8.5e-19	275	445	gi|6322073 ref|NP_012148.1|RHO3| ras homolog--GTP binding protein; Rho3p>sp|Q00245|RHO3_YEAST RHO3 PROTEIN >pir||S49891 GTP-binding proteinRHO3 - yeast (Saccharomyces cerevisiae)
Contig693	222	1.6e-17	337	492	sp|P36586|YPT5_SCHPO YPT1-RELATED PROTEIN 5 >pir||S34729 GTP-binding proteinypt5 - fission yeast (Schizosaccharomyces pombe)>emb|CAA80223.1|(Z22220) ypt5 protein [Schizosaccharomyces pombe]>emb|CAB11737.1|(Z98981) ras-like rab subfamily protein [Schizosaccharomyces pombe]
<GUANINE NUCLEOTIDE-BINDING PROTEIN>					
Contig781	854	1.7e-84	239	811	sp|O14435|GBB_CRYPA GUANINE NUCLEOTIDE-BINDING PROTEIN BETASUBUNIT>gb|AAC49838.1| (U95139) GTP binding protein beta subunit[Cryphonectria parasitica]
e3e11fs.r1	736	5.5e-72	71	508	sp|Q00580|GBA1_CRYPA GUANINE NUCLEOTIDE-BINDING PROTEIN ALPHA SUBUNIT>gb|AAA67706.1| (L32176) guanine nucleotide regulatory protein[Cryphonectria parasitica]
<GTPASE-ACTIVATING PROTEIN-neg regulator of Ras1, play antagonistic role with rho-gdp dissociation inhibitor>					
i3f01fs.r1	264	6.9e-21	12	473	sp|P33277|GAP1_SCHPO GTPASE-ACTIVATING PROTEIN >pir||A40258 RASGTPase-activating protein sar1 - fission yeast (Schizosaccharomycespombe)>pir||T40588 gtpase-activating protein - fission yeast(Schizosaccharomyces pombe) >dbj|BAA01251.1| (D10457)GTPase-activating protein [Schizosaccharomyces pombe]>gb|AAB19697.1| (S37449) sar1=RAS GTPase-activating protein[Schizosaccharomyces pombe, Peptide, 766 aa]
<Ras protein>					
g3f07fs.r1	300	8.2e-26	232	441	sp|P87018|RAS_BOTCI RAS-LIKE PROTEIN >gb|AAB51236.1| (U79558) Ras protein[Botryotinia fuckeliana]
5. Inositol triphosphate-secondary messenger (2)					
<inositol regulator>					
g2c09fs.r1	119	7.8e-06	302	433	pir||T39597 probable inositol regulator - fission yeast (Schizosaccharomycespombe) >emb|CAA19025.1| (AL023554) vesicle associated memebraneprotein; putative inositol regulator [Schizosaccharomyces pombe]
<Inositol polyphosphate phosphatase>					
i1e05fs.r1	353	4.8e-30	11	409	pir||T39233 probable Inositol polyphosphate phosphatase - fission yeast(Schizosaccharomyces pombe)

6. other (3)

<palA product-pH signal transduction pathway gene product>

h3d02fs.r1	496	5.7e-46	65	523	emb|CAB05920.1| (Z83333) palA [Emericella nidulans]

<carbon catabolite repression regulator>

d4a05fs.r1	663	3e-64	5	490	emb|CAA76330.1| (Y16626) carbon catabolite repression regulator [Gibberellafujikuroi]

<carbon catabolite repressor protein(CCR4)>

j4c07fs.f1	127	4.8e-07	319	465	gi|7019339 ref|NP_037486.1|| carbon catabolite repressor protein(CCR4)-associative factor 1 >gb|AAF01500.1|L46722_1 (L46722)BTG1binding factor 1 [Homo sapiens]

III.3. Transmembrane transport

1. secretion (3)

<ALPHA-ADAPTIN>

b3a04fs.r1	345	2.2e-29	6	467	gb|AAF51502.1| (AE003589) alpha-Adaptin gene product [Drosophila melanogaster]

<large secreted protein>

a1d06fs.f1	207	5.9e-15	93	509	pir||T36922 probable large secreted protein - Streptomyces coelicolor>emb|CAB46409.1| (AL096743) putative large secreted protein[Streptomyces coelicolor A3(2)]

<secreted glucosidase>

i2c03fs.r1	264	1.9e-21	89	457	pir||T35164 probable secreted glucosidase - Streptomyces coelicolor>emb|CAA19944.1| (AL031107) putative secreted glucosidase[Streptomyces coelicolor A3(2)]

2. exoenzymes (3)

<alanyl dipeptidyl peptidase>

k2h05fs.r1	157	2e-09	176	418	gb|AAD41777.1|AF1251 (AF125190) alanyl dipeptidyl peptidase [Aspergillusoryzae]

<GDP dissociation inhibitor>

g4a02fs.r1	419	2.1e-38	52	465	sp|Q10305|YD4C_SCHPO PUTATIVE SECRETORY PATHWAY GDP DISSOCIATION INHIBITOR>pir||T38215 probable secretory pathway GDP dissociation inhibitor-fission yeast (Schizosaccharomyces pombe) >emb|CAA93612.1|(Z69730) probable secretory pathway GDP dissociation inhibitor[Schizosaccharomyces pombe]
g4a02fs.f1	194	8.1e-14	279	473	gi|6320983 ref|NP_011062.1|GDI1| GDP dissociation inhibitor;Gdi1p>sp|P39958|GDI1_YEAST SECRETORY PATHWAY GDP DISSOCIATION INHIBITOR>pir||S44446 GDP dissociation inhibitor GDI1 -

					yeast (Saccharomycescerevisiae) >gb|AAB30540.1| (S69371) Gdi1p=GDP dissociationinhibitor [Saccharomyces cerevisiae, Peptide, 451 aa]>gb|AAC03234.1| (U18916) Gdi1p: secretory pathway GDP dissociationinhibitor [Saccharomyces
3. membrane transport (6)					
<nuclear transport factor 2>					
h3f10fs.r1	314	2.6e-27	131	502	sp|P87102|NTF2_NEUCR NUCLEAR TRANSPORT FACTOR 2 (NTF-2)>emb|CAA73689.1|(Y13237) putative nuclear transport factor 2 [Neurospora crassa]
<pre-tRNA nuclear export receptor>					
m1f03fs.r1	212	3.9e-15	14	415	pir||T40803 probable pre-tRNA nuclear export receptor - fission yeast(Schizosaccharomyces pombe) >emb|CAA21794.1| (AL032684)putativepre-tRNA nuclear export receptor [Schizosaccharomyces pombe]
<membrane transport protein>					
Contig926	397	4.7e-36	87	614	emb|CAB65616.1| (AL136078) probable membrane transporter [Schizosaccharomycespombe]
g3b02fs.r1	161	3.9e-10	31	441	emb|CAB65616.1| (AL136078) probable membrane transporter [Schizosaccharomycespombe]
s1b07fs.r1	126	0.0001	250	633	pir||T41604 probable membrane transport protein - fission yeast(Schizosaccharomyces pombe) >emb|CAA21238.1| (AL031825)putativemembrane transport protein [Schizosaccharomyces pombe]
<rab6-play an essential role in targeting and fusion of transport vesicles with their appropriate acceptor membranes>					
r4d09fs.f1	189	5e-14	292	513	dbj|BAA21707.1| (D84314) rab6 [Drosophila melanogaster]>gb|AAF53168.1|(AE003635) Rab6 gene product [Drosophila melanogaster]
4. mitochondrial transport (12)					
<ADP,ATP CARRIER PROTEIN>					
Contig1022	864	1.3e-85	136	789	sp|P02723|ADT_NEUCR ADP,ATP CARRIER PROTEIN (ADP/ATP TRANSLOCASE) (ADENINENUCLEOTIDE TRANSLOCATOR) (ANT) >pir||XWNC ADP,ATP carrier protein-Neurospora crassa >emb|CAA25104.1| (X00363) ADP/ATP carrier protein[Neurospora crassa]
Contig996	410	1.9e-37	309	575	sp|P02723|ADT_NEUCR ADP,ATP CARRIER PROTEIN (ADP/ATP TRANSLOCASE) (ADENINENUCLEOTIDE TRANSLOCATOR) (ANT) >pir||XWNC ADP,ATP carrier protein-Neurospora crassa >emb|CAA25104.1| (X00363) ADP/ATP carrier protein[Neurospora crassa]

<OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN>

Contig555	586	3.4e-56	60	479	sp|P07144|PORI_NEUCR OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN >pir||MMNCPporin - Neurospora crassa >emb|CAA29264.1| (X05824) major protein(AA 1-283) [Neurospora crassa]
Contig328	373	1.6e-33	220	468	sp|P07144|PORI_NEUCR OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN>pir||MMNCPporin - Neurospora crassa >emb|CAA29264.1| (X05824) major protein(AA 1-283) [Neurospora crassa]

<MITOCHONDRIAL PRECURSOR PROTEINS IMPORT RECEPTOR>

d4g12fs.r1	195	1.1e-13	98	472	sp|P23231|OM70_NEUCR MITOCHONDRIAL PRECURSOR PROTEINS IMPORT RECEPTOR (72 KDMITOCHONDRIAL OUTER MEMBRANE PROTEIN) (MITOCHONDRIAL IMPORTRECEPTOR FOR THE ADP/ATP CARRIER) (TRANSLOCASE OF OUTER MEMBRANETOM70) >pir||A36682 72K mitochondrial outer membrane protein -Neurospora crassa >emb|CAA37767.1| (X53735) mitochondrial outermembrane 72K protein [Neurospora crassa]>prf||1704253A ADP/ATPcarrier receptor

<MITOCHONDRIAL NUCLEASE>

b2h11fs.r1	346	1.1e-30	103	396	gi|6322253 ref|NP_012327.1|NUC1| mitochondrial nuclease; Nuc1p>sp|P08466|NUC1_YEAST MITOCHONDRIAL NUCLEASE >pir||NCBYN1 nucleaseNUC1 (EC 3.1.30.-) precursor, mitochondrial - yeast (Saccharomycescerevisiae)>emb|CAA29870.1| (X06670) nuclease [Saccharomycescerevisiae]>emb|CAA84003.1| (Z34098) ORF [Saccharomycescerevisiae] >emb|CAA54748.1| (X77688) mitochondrial nuclease[Saccharomyces cerevisiae]

<mitochondrial import receptor-a mitochondrial phosphate transport protein>

Contig1036	908	2.8e-90	141	1004	gi|6322537 ref|NP_012611.1|MIR1| Identified as mitochondrial import receptor(p32) and as PTP (PiC), a mitochondrial phosphate transportprotein.; Mir1p >sp|P23641|MPCP_YEAST MITOCHONDRIAL PHOSPHATECARRIER PROTEIN (PHOSPHATE TRANSPORT PROTEIN) (PTP) (MITOCHONDRIALIMPORT RECEPTOR) (P32) >pir||S12318 phosphate transport proteinMIR1, mitochondrial - yeast (Saccharomyces cerevisiae)>gb|AAA34782.1| (M54879)
Contig87	199	4.1e-15	210	479	sp|Q07335|OM22_NEUCR MITOCHONDRIAL IMPORT RECEPTOR SUBUNIT TOM22(MITOCHONDRIAL 22 KD OUTER MEMBRANE PROTEIN) (MOM22 PROTEIN)(TRANSLOCASE OF OUTER MEMBRANE 22 KD SUBUNIT)>pir||A40669mitochondrial receptor complex chain MOM22 - Neurospora crassa>emb|CAA50339.1| (X71021) mitochondrial 22 kDa outer membraneprotein [Neurospora crassa]

Contig11	162	1e-10	315	506	gi|6322537 ref|NP_012611.1|MIR1| Identified as mitochondrial import receptor(p32) and as PTP (PiC), a mitochondrial phosphate transportprotein.; Mir1p >sp|P23641|MPCP_YEAST MITOCHONDRIAL PHOSPHATECARRIER PROTEIN (PHOSPHATE TRANSPORT PROTEIN) (PTP) (MITOCHONDRIALIMPORT RECEPTOR) (P32) >pir||S12318 phosphate transport proteinMIR1, mitochondrial - yeast (Saccharomyces cerevisiae)>gb|AAA34782.1| (M54879)
<mitochondrial intermediate peptidase percursor>					
a2h09fs.r1	322	4.1e-27	20	451	sp|Q10415|PMIP_SCHPO PROBABLE MITOCHONDRIAL INTERMEDIATE PEPTIDASE PRECURSOR(MIP) >pir||T38081 probable mitochondrial intermediate peptidaseprecursor - fission yeast (Schizosaccharomyces pombe)>emb|CAA94628.1| (Z70690) putative mitochondrial intermediatepeptidase precursor [Schizosaccharomyces pombe]
m4g01fs.r1	124	7e-06	243	479	sp|P35999|PMIP_YEAST MITOCHONDRIAL INTERMEDIATE PEPTIDASE PRECURSO (MIP)>gb|AAA21278.1| (U10243) Mip1p [Saccharomyces cerevisiae]
<Mitochondrial matrix protein involved inprotein import>					
Contig51	697	6.2e-68	8	532	gi|6322505 ref|NP_012579.1|SSC1| Mitochondrial matrix protein involved inprotein import; subunit of SceI endonuclease; Ssc1p>sp|P12398|HS77_YEAST HEAT SHOCK PROTEIN SSC1, MITOCHONDRIALPRECURSOR (ENDONUCLEASE SCEI 75 KD SUBUNIT) >pir||HHBYS1 dnaK-typemolecular chaperone SSC1 precursor, mitochondrial - yeast(Saccharomyces cerevisiae)>gb|AAA63792.1| (M27229) heat shockprotein [Saccharomyces cerevisiae]
5. transporting for sugar, cation, anion, protein, fatty acid etc.					
5.1. sugar transport (8)					
<sugar transporter>					
t4h09fs.f1	251	7.7e-20	266	574	gb|AAD00266.1| (U77382) sugar transporter 1 [Pichia stipitis]
<hexose transporter>					
Contig820	727	4.9e-71	234	1070	gb|AAB65790.1| (AF010145) hexose transporter [Aspergillus parasiticus]
Contig206	256	2.3e-20	221	469	gb|AAB65790.1| (AF010145) hexose transporter [Aspergillus parasiticus]
<GLUCOSE TRANSPORTER>					
i1b04fs.r1	168	7.9e-11	101	424	gi|6321884 ref|NP_011960.1|HXT4| High-affinity glucose transporter; Hxt4p>pir||S46724 hexose transport protein HXT4 - yeast (Saccharomycescerevisiae) >gb|AAB68932.1| (U00060) Hxt4p: High-affinity glucosetransporter [Saccharomyces cerevisiae]
<UDP-galactose transporter related protein 1>					

c3f10fs.r1	624	4e-60	102	461	pir||JC5025 UDP-galactose transporter related protein 1 - mouse>dbj|BAA13526.1| (D87990) UGTrel1 [Mus musculus]
c3f10fs.f1	451	7.9e-42	150	452	pir||JC5025 UDP-galactose transporter related protein 1 - mouse>dbj|BAA13526.1| (D87990) UGTrel1 [Mus musculus]
<maltose permease>					
o4b01fs.r1	348	2.5e-30	6	509	emb|CAB46745.1| (AJ007636) maltose permease [Kluyveromyces lactis]
<Golgi GDP-mannose transporter>					
l1g10fs.r1	139	4.6e-08	374	517	gi|6321213 ref|NP_011290.1|GOG5| Golgi GDP-mannose transporter; Gog5p>sp|P40107|GOG5_YEAST VANADATE RESISTANCE PROTEIN GOG5/VRG4/VAN2>pir||S50238 vanadate resistance protein VAN2 - yeast(Saccharomyces cerevisiae) >gb|AAC37468.1| (L33915) vanadateresistant protein [Saccharomyces cerevisiae] >gb|AAA81537.1|(U15599) Van2p [Saccharomyces cerevisiae] >emb|CAA96941.1| (Z72747)ORF YGL225w [Saccharomyces cerevisiae]
5.2. cation transport (46)					
5.2.1. ATPase family-transmemberane protein that transport cations					
<P Type Copper ATPase>					
Contig207	118	3.7e-05	142	423	pir||T40072 P Type Copper ATPase - fission yeast (Schizosaccharomyces pombe)>emb|CAA18378.1| (AL022299) P Type Copper ATPase[Schizosaccharomyces pombe]
<ATPase proteolipid>					
Contig536	127	1.9e-07	436	522	emb|CAA24039.1| (V00667) ATPase proteolipid [Neurospora crassa]>prf||0808299AATPase subunit [Neurospora crassa]
<member of the AAA ATPase family of proteins>					
i2d01fs.r1	440	1.2e-39	17	406	gi|6320887 ref|NP_010966.1|SAP1| member of the AAA ATPase family of proteins;Sap1p >sp|P39955|SAP1_YEAST SAP1 PROTEIN >pir||S50550SIN1-associated protein SAP1 - yeast (Saccharomyces cerevisiae)>gb|AAB64582.1| (U18796) Yer047cp [Saccharomyces cerevisiae]
i2d01fs.f1	197	1.4e-13	241	438	gi|6320887 ref|NP_010966.1|SAP1| member of the AAA ATPase family of proteins;Sap1p >sp|P39955|SAP1_YEAST SAP1 PROTEIN>pir||S50550SIN1-associated protein SAP1 - yeast (Saccharomyces cerevisiae)>gb|AAB64582.1| (U18796) Yer047cp [Saccharomyces cerevisiae]
<membrane-spanning ATPase>					
j2d01fs.f1	225	1.4e-17	176	475	gi|6321465 ref|NP_011542.1|MSP1| 40 kDa putative membrane-spanning ATPase;Msp1p >sp|P28737|MSP1_YEAST MSP1 PROTEIN (TAT-BINDING HOMOLOG

					4)>pir||A49506 MSP1 protein - yeast (Saccharomyces cerevisiae)>emb|CAA48191.1| (X68055) MSP1 protein [Saccharomyces cerevisiae]>emb|CAA56956.1| (X81069) probable regulatory subunit of 26Sprotease [Saccharomyces cerevisiae] >emb|CAA97015.1| (Z72813) ORFYGR028w
<DNA-dependent ATPases>					
k1e08fs.r1	233	3.6e-17	12	398	gi|6321289 ref|NP_011365.1|INO80| similar to the Snf2p family of DNA-dependentATPases; Ino80p >sp|P53115|YGP0_YEAST HYPOTHETICAL 171.5 KDHELICASE IN NUT1-ARO2 INTERGENIC REGION >pir||S60416 DNA helicaseYGL150c - yeast (Saccharomyces cerevisiae) >emb|CAA96861.1|(Z72672) ORF YGL150c [Saccharomyces cerevisiae]
<calcium-transporting ATPase>					
Contig231	610	2e-57	2	514	emb|CAB65293.1| (AJ243515) putative calcium P-type ATPase [Neurospora crassa]
l1g01fs.f1	528	5.5e-49	144	518	emb|CAB65296.1| (AJ243518) putative calcium P-type ATPase [Neurospora crassa]
<transitional endoplasmic reticulum ATPase>					
Contig709	121	1.5e-05	295	495	sp|P54812|TER2_CAEEL TRANSITIONAL ENDOPLASMIC RETICULUM ATPASE HOMOLOG 2(P97/CDC48 HOMOLOG 2) >pir||T19879 hypothetical protein C41C4.8-Caenorhabditis elegans >emb|CAA88105.1| (Z48045) similar to P97protein;cDNA EST EMBL:M89239 comes from this gene; cDNA ESTEMBL:D28083 comes from this gene; cDNA EST EMBL:D28082 comes fromthis gene; cDNA EST EMBL:T00671 comes from this gene; cDNA ESTEMBL:D32722 comes from this
<sodium/potassium-transporting ATPase>					
t2a06fs.f1	167	2.6e-10	171	512	sp|P28774|ATNA_ARTSF SODIUM/POTASSIUM-TRANSPORTING ATPASE ALPHA CHAIN (SODIUMPUMP) (NA+/K+ ATPASE) >pir||JH0470 Na+/K+-exchanging ATPase (EC3.6.1.37) alpha chain (clone pArATNa136) - brine shrimp>emb|CAA39972.1| (X56650) alpha subunit of the Na/K ATPase [Artemiafranciscana]
<PLASMA MEMBRANE ATPASE (PROTON PUMP)>					
Contig1009	2105	4.3e-217	95	1648	sp|Q07421|PMA1_AJECA PLASMA MEMBRANE ATPASE (PROTON PUMP)>gb|AAB53772.1|(L07305) ATPase [Ajellomyces capsulatus] >prf||2004293A H ATPase[Ajellomyces capsulatus]
Contig385	732	1.4e-71	1	465	sp|P07038|PMA1_NEUCR PLASMA MEMBRANE ATPASE (PROTON PUMP)>pir||PXNCPH+-transporting ATPase (EC 3.6.1.35), plasma membrane - Neurosporacrassa >gb|AAA33561.1| (M14085) plasma membrane ATPase [Neurosporacrassa]

Contig1010	529	2.5e-49	195	518	sp|Q07421|PMA1_AJECA PLASMA MEMBRANE ATPASE (PROTON PUMP)>gb|AAB53772.1|(L07305) ATPase [Ajellomyces capsulatus] >prf||2004293A H ATPase[Ajellomyces capsulatus]
m2d02fs.f1	413	9.6e-38	206	475	gi|4502315 ref|NP_001686.1|| ATPase, H+ transporting, lysosomal (vacuolarproton pump) 42kD >sp|P21283|VATC_HUMAN VACUOLAR ATP SYNTHASESUBUNIT C (V-ATPASE C SUBUNIT) >pir||JN0907 H+-transporting ATPase(EC 3.6.1.35) chain C, vacuolar - human >emb|CAA48903.1| (X69151)vacuolar proton-ATPase [Homo sapiens]
Contig940	303	6.2e-25	205	519	sp|P07038|PMA1_NEUCR PLASMA MEMBRANE ATPASE (PROTON PUMP)>pir||PXNCPH+-transporting ATPase (EC 3.6.1.35), plasma membrane - Neurosporacrassa >gb|AAA33561.1| (M14085) plasma membrane ATPase [Neurosporacrassa]
<V-ATPase>					
Contig1049	667	1.1e-64	1188	1748	gb|AAB61278.1| (AF001033) putative 20kDa subunit of the V-ATPase [Neurosporacrassa]
Contig957	230	2.2e-18	338	520	gb|AAB61278.1| (AF001033) putative 20kDa subunit of the V-ATPase [Neurosporacrassa]
<cation-transporting P-ATPase>					
o1h09fs.f1	746	5e-73	5	574	gb|AAC27991.1| (AF036763) P-ATPase [Emericella nidulans]
o1h09fs.r1	426	4.4e-48	114	524	gb|AAC27991.1| (AF036763) P-ATPase [Emericella nidulans]
d3b10fs.r1	282	9.3e-23	168	506	pir||C69069 cation-transporting P-ATPase PacL - Methanobacteriumthermoautotrophicum (strain Delta H) >gb|AAB85991.1| (AE000912)cation-transporting P-ATPase PacL [Methanobacteriumthermoautotrophicum]
q3a04fs.f1	280	1.9e-22	164	508	gb|AAC27991.1| (AF036763) P-ATPase [Emericella nidulans]
<vacuolar ATPase subunit H>					
k3f06fs.f1	93	0.22	317	499	emb|CAB55499.1| (AJ249389) vacuolar ATPase subunit H [Manduca sexta]

5.2.2. facilitator protein

<major facilitator superfamily protein>					
b1f02fs.r1	273	9e-22	13	435	sp|Q09766|YA7D_SCHPO HYPOTHETICAL 98.4 KD PROTEIN C24H6.13 IN CHROMOSOME I>pir||S62415 major facilitator protein homolog - fission yeast(Schizosaccharomyces pombe) >emb|CAA90857.1| (Z54142) putativemajor facilitator superfamily protein [Schizosaccharomyces pombe]
b2b06fs.f1	102	0.0018	194	298	gi|6322399 ref|NP_012473.1|LAS21| putative membrane protein, a member of themajor facilitator super family; Las21p >sp|P40367|YJG2_YEASTHYPOTHETICAL 94.9 KD PROTEIN IN MRPL8-NUP82

					INTERGENIC REGION>pir||S50810 probable membrane protein YJL062w - yeast(Saccharomyces cerevisiae) >emb|CAA84061.1| (Z34288) HRC830[Saccharomyces cerevisiae] >emb|CAA89353.1| (Z49337) ORF YJL062w[Saccharomyces cerevisiae]
<major facilitator protein>					
Contig1020	708	5.2e-69	29	1726	pir||T40506 major facilitator protein homolog - fission yeast(Schizosaccharomyces pombe) >emb|CAA20729.1| (AL031534) MFS effluxtransporter of unknown specificity [Schizosaccharomyces pombe]
<Mfs1.1-member of the major facilitator superfamily>					
Contig197	318	3.5e-27	3	509	gb|AAF01426.1|AF1863 (AF186391) Mfs1.1 [Coprinus cinereus]
Contig482	243	5.8e-19	165	545	gb|AAF01426.1|AF1863 (AF186391) Mfs1.1 [Coprinus cinereus]
5.2.3. others					
<potassium channel protein>					
Contig632	545	9.7e-52	3	503	pir||T41659 probable potassium channel subunit - fission yeast(Schizosaccharomyces pombe) >emb|CAA19066.1| (AL023590) putativepotassium channel subunit [Schizosaccharomyces pombe]
Contig762	319	7.5e-28	151	513	pir||T41659 probable potassium channel subunit - fission yeast(Schizosaccharomyces pombe) >emb|CAA19066.1| (AL023590)putativepotassium channel subunit [Schizosaccharomyces pombe]
n2h02fs.r1	206	9.4e-15	207	506	gi|6322368 ref|NP_012442.1|TOK1| outward-rectifier potassium channel; Tok1p>sp|P40310|TOK1_YEAST OUTWARD-RECTIFIER POTASSIUM CHANNEL TOK1(TWO-DOMAIN OUTWARD RECTIFIER K+ CHANNEL YORK)>pir||S46585outward-rectifier potassium channel - yeast (Saccharomycescerevisiae) >emb|CAA54360.1| (X77087) potential membranouslocalization; J0911 ORF [Saccharomyces cerevisiae]>gb|AAC49070.1|(U28005) outward-rectifier
<Arabidopsis ATHKCP-potassium channel protein>					
Contig635	138	5e-08	187	477	gb|AAB80621.1| (AC002376) Match to Arabidopsis ATHKCP (gb|L40948).ESTsgb|ATTS0764, gb|R90646, gb|AA389809, gb|ATTS2615 come from thisgene. [Arabidopsis thaliana] >gb|AAC15999.1| (AF061570) potassiumchannel beta subunit homolog [Arabidopsis thaliana]
<K+/H+-antiporter>					
g3b03fs.r1	247	5.5e-19	141	446	emb|CAB76234.1| (AL157994) Probable K+/H+-antiporter [Schizosaccharomycespombe]
<H/K ATPase>					
d3b10fs.f1	140	2e-07	132	434	prf||2112199A H/K ATPase:SUBUNIT=alpha [Xenopus laevis]

<copper transporter protein>					
Contig990	305	2.6e-26	207	827	pir||T40958 high affinity copper transporter - fission yeast(Schizosaccharomyces pombe) >emb|CAB38165.1| (AL035592) highaffinity copper transporter [Schizosaccharomyces pombe]>emb|CAB52305.1| (AJ243833) high affinity copper transporter[Schizosaccharomyces pombe] >gb|AAD51064.1| (AF175405) Ctr4 protein[Schizosaccharomyces pombe]
p3f03fs.f1	126	1.2e-05	259	570	pir||T40958 high affinity copper transporter - fission yeast(Schizosaccharomyces pombe) >emb|CAB38165.1| (AL035592) highaffinity copper transporter [Schizosaccharomyces pombe]>emb|CAB52305.1| (AJ243833)high affinity copper transporter[Schizosaccharomyces pombe] >gb|AAD51064.1| (AF175405) Ctr4 protein[Schizosaccharomyces pombe]
Contig726	92	0.12	66	203	sp|Q39065|COPT_ARATH COPPER TRANSPORTER 1 >emb|CAA90018.1| (Z49859) coppertransporter protein [Arabidopsis thaliana]
<calcium binding protein>					
i3d02fs.r1	146	2.4e-08	180	473	pir||PC4014 calcium binding 140k protein - mouse (fragment)>gb|AAB35051.1|(S78797) CBP-140=calcium binding protein/heat shock protein homolog[mice, F9 embryonal carcinoma cells, Peptide Partial, 652 aa] [Mussp.]
<calcium/proton exchanger>					
h3g04fs.r1	449	1.4e-41	60	467	gb|AAC08353.1| (AF053229) calcium/proton exchanger [Neurospora crassa]
l2d11fs.f1	335	1.7e-29	218	529	gi|6320075 ref|NP_010155.1|VCX1| vacuolar H+/Ca2+ exchanger; Vcx1p>pir||S61933 Ca2+/H+-exchanging protein, vacuolar - yeast(Saccharomyces cerevisiae) >gb|AAB60313.1| (U36603) vacuolarH+/Ca2+ exchanger [Saccharomyces cerevisiae] >gb|AAC49550.1|(U18944) Hum1p [Saccharomyces cerevisiae] >emb|CAA98696.1| (Z74176)ORF YDL128w [Saccharomyces cerevisiae]
<purine-cytosine permease>					
Contig715	448	1.9e-41	77	982	gi|6320902 ref|NP_010981.1|FCY21| purine-cytosine permease; Fcy21p>sp|P40039|YEP0_YEAST HYPOTHETICAL 58.1 KD PROTEIN IN PET117-CEM1INTERGENIC REGION >emb|CAA66032.1| (X97346) FCYY [Saccharomycescerevisiae] >gb|AAB64596.1| (U18813) Fcy21p:Purine-cytosinepermease [Saccharomyces cerevisiae]
Contig471	134	3.6e-07	297	500	gi|6320897 ref|NP_010976.1|FCY2| purine-cytosine permease; Fcy2p>sp|P17064|FCY2_YEAST PURINE-CYTOSINE PERMEASE (PCP)(CYTOSINE/PURINE TRANSPORT PROTEIN) >pir||GRBYCP

					cytosine/purinetransport protein - yeast (Saccharomyces cerevisiae)>gb|AAB64592.1| (U18813) Fcy2p: purine-cytosine permease[Saccharomyces cerevisiae]
<ammonia transport protein>					
j3g06fs.r1	577	3.5e-55	26	544	gb|AAD40955.1|AF1595 (AF159568) ammonium transporter MEPa [Microbotryumviolaceum]
Contig438	237	1.9e-18	312	518	gi|6325396 ref|NP_015464.1|MEP3| NH4+ transporter, highly similar to Mep1p andMep2p; Mep3p >sp|P53390|MEP3_YEAST AMMONIUM TRANSPORTER MEP3>pir||S69027 ammonium transport protein MEP3 - yeast (Saccharomycescerevisiae) >gb|AAB68278.1| (U40829) Similar to B. subtilismembrane protein NrgA (Swiss Prot. accession number Q07429)[Saccharomyces cerevisiae]
j3g04fs.f1	163	2.4e-10	178	510	gi|6324187 ref|NP_014257.1|MEP2| Ammonia transport protein; Mep2p>sp|P41948|MEP2_YEAST AMMONIUM TRANSPORTER MEP2 >pir||S51089ammonium transport protein MEP2 - yeast (Saccharomyces cerevisiae)>emb|CAA58587.1| (X83608) ammonium transporter [Saccharomycescerevisiae] >emb|CAA86884.1| (Z46843) NH3 permease [Saccharomycescerevisiae] >emb|CAA96025.1| (Z71418)ORF YNL142w [Saccharomycescerevisiae]
<Low-affinity zinc transport protein>					
Contig746	345	1.5e-30	183	581	gi|6323159 ref|NP_013231.1|ZRT2| Low-affinity zinc transport protein; Zrt2p>pir||S59319 probable membrane protein YLR130c - yeast(Saccharomyces cerevisiae) >emb|CAA62642.1| (X91258) L3120[Saccharomyces cerevisiae] >gb|AAB82397.1| (U53881) Ylr130cp[Saccharomyces cerevisiae] >emb|CAA97701.1| (Z73302) ORF YLR130c[Saccharomyces cerevisiae]
a2b01fs.r1	172	1.6e-11	30	191	gi|6323159 ref|NP_013231.1|ZRT2| Low-affinity zinc transport protein; Zrt2p>pir||S59319 probable membrane protein YLR130c - yeast(Saccharomyces cerevisiae) >emb|CAA62642.1| (X91258)L3120[Saccharomyces cerevisiae] >gb|AAB82397.1| (U53881)Ylr130cp[Saccharomyces cerevisiae] >emb|CAA97701.1| (Z73302) ORF YLR130c[Saccharomyces cerevisiae]
5.3. Anion transport (18)					
<anion transporter>					
k2h08fs.r1	130	3.5e-06	94	300	dbj|BAA13016.1| (D86086) canalicular multispecific organic anion transporter[Rattus norvegicus]
<phosphate permease>					
Contig1032	2448	2.1e-253	59	1762	dbj|BAA33769.1| (AB011417) phosphate permease [Gibberella zeae]

Contig954	209	7.9e-14	581	886	emb|CAB68656.1| (AL137099) putative inorganic phosphate transporter[Schizosaccharomyces pombe]
Contig354	151	9.2e-10	30	473	gb|AAF40188.1| (AF229169) phosphate transporter [Oryza sativa]
l1a02fs.r1	118	2.1e-05	14	142	dbj|BAA33769.1| (AB011417) phosphate permease [Gibberella zeae]
<allantoate permease>					
Contig771	595	5.2e-57	13	915	pir||T41345 probable allantoate permease - fission yeast (Schizosaccharomycespombe) >emb|CAA22656.1| (AL035076) putative allantoate permease[Schizosaccharomyces pombe]
Contig109	185	1e-12	84	554	pir||T39680 probable allantoate permease - fission yeast (Schizosaccharomycespombe) >emb|CAA21920.1| (AL033389) putative allantoate permease[Schizosaccharomyces pombe]
Contig283	113	0.00011	221	517	sp|Q10097|YAOI_SCHPO PUTATIVE TRANSPORTER C11D3.18C >pir||T37527 probableallantoate transporter - fission yeast (Schizosaccharomyces pombe)>emb|CAA92319.1| (Z68166) putative allantoate transporter[Schizosaccharomyces pombe]
<quinate permease>					
b1c03fs.r1	146	1.8e-08	32	421	sp|P15325|QUTD_EMENI QUINATE PERMEASE (QUINATE TRANSPORTER)>pir||S08498quinate transport protein - Emericella nidulans>emb|CAA31879.1|(X13525) quinate permease [Emericella nidulans]
<malate permease>					
q2b07fs.r1	108	0.00017	85	324	pir||T41614 malate permease - fission yeast (Schizosaccharomyces pombe)>emb|CAA19133.1| (AL023595) malate permease [Schizosaccharomycespombe]
<pantothenate permease>					
e3c02fs.r1	141	5.3e-08	120	458	pir||E69383 pantothenate permease (panF-1) homolog - Archaeoglobus fulgidus>gb|AAB90171.1| (AE001029) pantothenate permease (panF-1)[Archaeoglobus fulgidus]
<acetate permease>					
t2c04fs.r1	599	1.4e-57	13	477	sp|P15937|ACU8_NEUCR ACETYL-COA HYDROLASE (ACETYL-COA DEACYLASE (ACETYL-COAACYLASE) (ACETATE UTILIZATION PROTEIN) >pir||A36316 acu-8 protein-Neurospora crassa >gb|AAA33554.1| (M31521) acetate permease (acu-8)[Neurospora crassa]
t2c04fs.f1	500	5.5e-47	194	526	sp|P15937|ACU8_NEUCR ACETYL-COA HYDROLASE (ACETYL-COA DEACYLASE) (ACETYL-COAACYLASE) (ACETATE UTILIZATION PROTEIN) >pir||A36316 acu-8 protein-Neurospora crassa >gb|AAA33554.1| (M31521) acetate permease (acu-8)[Neurospora crassa]
<tricarboxylate transport protein>					

g4f06fs.r1	425	5.1e-39	68	478	pir||T37992 probable tricarboxylate transport protein - fission yeast(Schizosaccharomyces pombe) >emb|CAB10116.1| (Z97209) putativetricarboxylate transport protein [Schizosaccharomyces pombe]
g4f06fs.f1	300	8.6e-26	205	483	pir||T37992 probable tricarboxylate transport protein - fission yeast(Schizosaccharomyces pombe) >emb|CAB10116.1| (Z97209)putativetricarboxylate transport protein [Schizosaccharomyces pombe]
<mitochondrial dicarboxylate transportprotein>					
Contig1	288	1.4e-24	100	468	gi|6323381 ref|NP_013452.1|DIC1| mitochondrial dicarboxylate transportprotein; Dic1p >pir||S51351 hypothetical protein YLR348c - yeast(Saccharomyces cerevisiae) >gb|AAB67266.1| (U19028)Ylr348cp[Saccharomyces cerevisiae] >gb|AAB71336.1| (U79459)dicarboxylatetransport protein [Saccharomyces cerevisiae]
Contig17	94	0.11	221	316	gi|6323381 ref|NP_013452.1|DIC1| mitochondrial dicarboxylate transportprotein; Dic1p >pir||S51351 hypothetical protein YLR348c - yeast(Saccharomyces cerevisiae) >gb|AAB67266.1| (U19028)Ylr348cp[Saccharomyces cerevisiae] >gb|AAB71336.1| (U79459)dicarboxylatetransport protein [Saccharomyces cerevisiae]
<SutA-sulfate transporter>					
k4a04fs.r1	486	4e-45	12	461	gb|AAF14540.1|AF1639 (AF163975) SutA [Penicillium chrysogenum]
5.4. Protein, amino acid transport (21)					
<PROTEIN TRANSPORT PROTEIN>					
r3c05fs.r1	606	3.6e-58	7	618	pir||T41295 protein transport protein sec23 homolog - fission yeast(Schizosaccharomyces pombe) >emb|CAA21224.1| (AL031824) proteintransport protein sec23 homolog [Schizosaccharomyces pombe]
b3a12fs.r1	163	2.6e-11	235	441	pir||T40142 probable protein transport protein - fission yeast(Schizosaccharomyces pombe) >emb|CAA17883.1| (AL022103) putativeprotein transport protein; yeast sec61 beta 2 subunit-like[Schizosaccharomyces pombe]
<peptide transporter>					
Contig1041	616	7.5e-115	1420	2379	gb|AAF26618.1|AF1250 (AF125094) peptide transporter MTD1 [Schizophyllumcommune]
c3h10fs.f1	247	2.8e-19	205	453	gi|6322946 ref|NP_013019.1|PTR2| Peptide transporter; Ptr2p>sp|P32901|PTR2_YEAST PEPTIDE TRANSPORTER PTR2 (PEPTIDE PERMEASEPTR2)>pir||S38171 peptide transport protein PTR2 - yeast(Saccharomyces cerevisiae) >emb|CAA51947.1| (X73541) ORF

					YKR413[Saccharomyces cerevisiae]>emb|CAA82172.1| (Z28318) ORF YKR093w[Saccharomyces cerevisiae]
<peptide transporter MTD1>					
g4h02fs.r1	368	4.6e-32	10	462	gb|AAF26618.1|AF1250 (AF125094) peptide transporter MTD1 [Schizophyllumcommune]
g4h02fs.f1	220	3.3e-16	135	461	gb|AAF26618.1|AF1250 (AF125094) peptide transporter MTD1 [Schizophyllumcommune]
<AMINO-ACID PERMEASE>					
Contig1002	917	3.5e-91	120	1034	sp|P34054|INA1_TRIHA AMINO-ACID PERMEASE INDA1 >pir||S33212 INDA1 protein-fungus (Trichoderma harzianum) >emb|CAA80308.1| (Z22594) INDA1[Trichoderma harzianum]
Contig948	501	4.6e-47	82	534	gb|AAB61277.1| (AF001032) amino acid permease [Neurospora crassa]
<importin beta-2 subunit>					
e2d08fs.r1	367	7.7e-32	70	483	sp|O14089|IMB2_SCHPO PUTATIVE IMPORTIN BETA-2 SUBUNIT (KARYOPHERIN BETA-2SUBUNIT) (IMPORTIN 104) (TRANSPORTIN) (TRN) >pir||T38539 probableimportin beta-2 subunit (transportin) - fission yeast(Schizosaccharomyces pombe) >emb|CAB16272.1| (Z99165) putativeimportin beta-2 subunit (transportin) [Schizosaccharomyces pombe]
<importin beta subunit>					
m2g05fs.r1	106	0.0057	26	502	pir||T41171 importin beta subunit - fission yeast (Schizosaccharomyces pombe)>emb|CAA20126.1| (AL031179) importin beta subunit[Schizosaccharomyces pombe]
<N AMINO ACID TRANSPORT SYSTEM PROTEIN>					
n1d09fs.r1	450	1e-41	68	508	sp|P38680|MTR_NEUCR N AMINO ACID TRANSPORT SYSTEM PROTEIN (METHYLTRYPTOPHANRESISTANCE PROTEIN) >pir||S47892 neutral amino acid permease -Neurospora crassa >gb|AAA33600.1| (L34605) neutral amino acidpermease [Neurospora crassa]
Contig305	205	8.7e-16	101	529	pir||A54551 N amino acid transport system protein mtr - Neurospora crassa>gb|AAB21410.1| (S81767) N amino acid transport system=mtr[Neurospora crassa, Peptide, 261 aa]
<GABA permease-gamma-amino-n-butyrate permease>					
Contig344	287	7.9e-24	188	646	emb|CAB43936.1| (AJ131668) GABA permease [Emericella nidulans]
n2e06fs.r1	226	3.4e-17	7	486	emb|CAB43936.1| (AJ131668) GABA permease [Emericella nidulans]
<gamma-glutamyltranspeptidase precursor>					
m3e07fs.r1	301	4.7e-25	50	529	emb|CAB65810.1| (AL136235) putative gamma-glutamyltranspeptidase precursor[Schizosaccharomyces pombe]
<Optlp-oligopeptide transport gene>					

o3d10fs.r1 214 1.8e-15 167 502 gb|AAB69628.1| (U60973) Opt1p [Candida albicans]
<methionine permease>
n2b12fs.r1 129 1.4e-06 25 321 gi|6321492 ref|NP_011569.1|MUP1| high affinity methionine permease; Mup1p>sp|P50276|MUP1_YEAST HIGH AFFINITY METHIONINE PERMEASE>pir||S61943 methionine transport protein, high affinity - yeast(Saccharomyces cerevisiae) >gb|AAB63529.1| (U40316) high affinitymethionine permease [Saccharomyces cerevisiae]>emb|CAA97055.1|(Z72840) ORF YGR055w [Saccharomyces cerevisiae]

<soluble NSF attachment protein-a soluble transport factor>
m3a09fs.r1 206 8.1e-16 247 543 sp|P81126|SNAB_BOVIN BETA-SOLUBLE NSF ATTACHMENT PROTEIN (SNAP-BETA)>pir||S32368 beta-SNAP protein - bovine >gb|AAB25813.1| betasoluble NSF attachment protein, betaSNAP=N-ethyl-maleimide-sensitive fusion protein attachment protein[cattle, brain, Peptide, 298 aa]>prf||1910317B NSF attachmentprotein (SNAP):ISOTYPE=beta [Bos taurus]
m3a09fs.f1 128 5.2e-07 302 463 dbj|BAA19246.1| (AB001375) similar to soluble NSF attachment protein [Vitisvinifera]

<SLS1 PROTEIN PRECURSOR-an endoplasmic reticulum component involved in protein translocation process>
k4g08fs.r1 169 3.5e-11 74 505 sp|Q99158|SLS1_YARLI SLS1 PROTEIN PRECURSOR >pir||S58132 Sls1 proteinprecursor - yeast (Yarrowia lipolytica) >emb|CAA90516.1| (Z50154)Sls1 protein [Yarrowia lipolytica]

<synaptobrevin-protein trafficing>
Contig351 341 3.6e-30 184 456 sp|Q92356|SNC2_SCHPO PROBABLE SYNAPTOBREVIN HOMOLOG C6G9.11 >pir||T39073synaptobrevin homolog1 - fission yeast (Schizosaccharomyces pombe)>emb|CAB03613.1| (Z81317) synaptobrevin homolog1[Schizosaccharomyces pombe]

5.5. fatty acid trasport (2)
<fatty acid transporter protein>
b2g01fs.r1 166 1.7e-10 2 421 emb|CAA75802.1| (Y15839) fatty acid transporter protein [Cochliobolusheterostrophus]

<phosphatidylglycerol/phosphatidylinositol transfer protein>
Contig522 407 4.2e-37 173 589 gb|AAD16095.1| (AF089838) phosphatidylglycerol/phosphatidylinositol transferprotein [Aspergillus oryzae]

5.6. ABC transporter family (11)
<ABC transporter protein-Osmoregulated and lipid transport>

a2f03fs.r1	627	4.3e-59	20 433	pir||T30541 ABC1 transport protein - rice blast fungus>gb|AAB86640.1|(AF032443) ABC1 transporter; ABC-type ATPase [Magnaporthe grisea]
n2f05fs.r1	515	3.9e-47	14 484	pir||T30541 ABC1 transport protein - rice blast fungus >gb|AAB86640.1|(AF032443) ABC1 transporter; ABC-type ATPase [Magnaporthe grisea]
c4c02fs.r1	427	8.4e-38	10 459	sp|O74676|CDR4_CANAL ABC TRANSPORTER CDR4 >pir||T30550 ABC transport protein -yeast (Candida albicans) >gb|AAC72295.1| (AF044921) ABC transporter[Candida albicans]
Contig452	424	1.7e-37	73 495	pir||T30541 ABC1 transport protein - rice blast fungus>gb|AAB86640.1|(AF032443) ABC1 transporter; ABC-type ATPase [Magnaporthe grisea]
g1b12fs.r1	323	1.2e-26	61 471	gi|6322980 ref|NP_013052.1|YBT1| bile acid transporter of ABC family; Ybt1p>sp|P32386|YBT1_YEAST ATP-DEPENDENT BILE ACID PERMEASE >pir||S64800probable membrane protein YLL048c - yeast (Saccharomycescerevisiae) >emb|CAA97500.1| (Z73153) ORF YLL048c [Saccharomycescerevisiae]
k2h12fs.r1	283	1.8e-22	40 495	emb|CAB66463.1| (AL136538) putative ABC transporter [Schizosaccharomycespombe]
c4h04fs.r1	250	4.5e-19	13 282	emb|CAA08835.1| (AJ009799) ABC transporter protein [Gallus gallus]
Contig794	244	3.2e-18	141 374	pir||T30541 ABC1 transport protein - rice blast fungus>gb|AAB86640.1|(AF032443) ABC1 transporter; ABC-type ATPase [Magnaporthe grisea]
c4c02fs.f1	222	1.7e-17	225 455	emb|CAB76198.1| (AJ272521) putative PDR-like ABC transporter [Botryotiniafuckeliana]
e2d03fs.r1	183	4.4e-12	119 451	gb|AAC27077.1| (AF071203) ABC transporter MOAT-B isoform [Homo sapiens]
<PDR-like ABC transporter>				
Contig456	276	3e-23	225 473	emb|CAB76198.1| (AJ272521) putative PDR-like ABC transporter [Botryotiniafuckeliana]
5.7. other (16)				
<transport protein>				
Contig952	396	6e-36	241 924	sp|Q10097|YAOI_SCHPO PUTATIVE TRANSPORTER C11D3.18C >pir||T37527 probableallantoate transporter - fission yeast (Schizosaccharomyces pombe)>emb|CAA92319.1| (Z68166) putative allantoate transporter[Schizosaccharomyces pombe]

Contig530	331	4.2e-29	438 863	pir||T41634 probable transport protein - fission yeast (Schizosaccharomycespombe) >emb|CAB52881.1| (AL109850) putative transport protein[Schizosaccharomyces pombe]
o4f05fs.r1	288	5.8e-24	88 510	emb|CAB63540.1| (AL133521) putative transporter [Schizosaccharomyces pombe]
c3h10fs.r1	281	6.2e-23	77 436	gi|6322946 ref|NP_013019.1|PTR2| Peptide transporter;Ptr2p>sp|P32901|PTR2_YEAST PEPTIDE TRANSPORTER PTR2 (PEPTIDE PERMEASEPTR2)>pir||S38171 peptide transport protein PTR2 - yeast(Saccharomyces cerevisiae) >emb|CAA51947.1| (X73541) ORF YKR413[Saccharomyces cerevisiae]>emb|CAA82172.1| (Z28318) ORF YKR093w[Saccharomyces cerevisiae]
Contig524	196	9.5e-15	533 850	pir||T38535 probable translocation protein - fission yeast(Schizosaccharomyces pombe) >emb|CAB16260.1| (Z99165)putativetranslocation protein [Schizosaccharomyces pombe]
n4e06fs.r1	177	9.6e-12	193 495	pir||T38501 hypothetical protein SPAC29B12.14c - fission yeast(Schizosaccharomyces pombe) >emb|CAB16258.1| (Z99164) NCS1 uraciltransporter [Schizosaccharomyces pombe]
n2e06fs.f1	159	7.5e-10	98 508	gi|6321361 ref|NP_011438.1|HNM1| Transporter (permease) for choline andnitrogen mustard; share homology with UGA4;Hnm1p>sp|P19807|HNM1_YEAST CHOLINE TRANSPORT PROTEIN >pir||S11175choline transport protein - yeast (Saccharomyces cerevisiae)>gb|AAA34537.1| (J05603)choline transport protein [Saccharomycescerevisiae] >emb|CAA96782.1| (Z72599) ORF YGL077c [Saccharomycescerevisiae]
r3c05fs.f1	154	4.4e-09	180 368	pir||T40674 protein transport protein sec23 homolog - fission yeast(Schizosaccharomyces pombe) >emb|CAA22877.1| (AL035263)proteintransport protein sec23 homolog. [Schizosaccharomyces pombe]
o4e10fs.f1	143	2.6e-08	279 503	emb|CAB61275.1| (AL132991) putative transporter protein [Streptomycescoelicolor A3(2)]
e4e07fs.r1	113	7.7e-05	175 441	gi|6319871 ref|NP_009952.1|YCR023C| Membrane transporter;Ycr023cp>sp|P25351|YCR3_YEAST HYPOTHETICAL 69.2 KD PROTEIN IN HSP30-PMP1INTERGENIC REGION >pir||S19434 probable transport protein YCR023c-yeast (Saccharomyces cerevisiae) >emb|CAA42315.1| (X59720) YCR023c,len:611 [Saccharomyces cerevisiae]

<metabolite transport protein>

Contig831	201	2.2e-14	319	603	pir||T39345 probable metabolite transport protein - fission yeast(Schizosaccharomyces pombe) >emb|CAA22199.1| (AL034353)metabolitetransport protein [Schizosaccharomyces pombe]
Contig721	184	1.5e-12	250	513	pir||T39345 probable metabolite transport protein - fission yeast(Schizosaccharomyces pombe) >emb|CAA22199.1| (AL034353)metabolitetransport protein [Schizosaccharomyces pombe]
<OLIGOMYCINRESISTANCE ATP-DEPENDENT PERMEASE>					
Contig915	902	1.1e-88	9	1034	gi|6321720 ref|NP_011797.1|YOR1| Yor1p >sp|P53049|YOR1_YEAST OLIGOMYCINRESISTANCE ATP-DEPENDENT PERMEASE YOR1 >pir||S64616 YOR1 protein-yeast (Saccharomyces cerevisiae) >gb|AAB35750.1| oligomycinresistance 1 protein, yor1p=ATP-binding cassette transporter[Saccharomyces cerevisiae,Peptide, 1477 aa] >emb|CAA97312.1|(Z73066) ORF YGR281w [Saccharomyces cerevisiae]
g1b12fs.f1	206	2.8e-14	244	468	gi|6321720 ref|NP_011797.1|YOR1| Yor1p >sp|P53049|YOR1_YEAST OLIGOMYCINRESISTANCE ATP-DEPENDENT PERMEASE YOR1 >pir||S64616 YOR1 protein-yeast (Saccharomyces cerevisiae) >gb|AAB35750.1| oligomycinresistance 1 protein, yor1p=ATP-binding cassette transporter[Saccharomyces cerevisiae, Peptide, 1477 aa] >emb|CAA97312.1|(Z73066) ORF YGR281w [Saccharomyces cerevisiae]
<nitrogen permease>					
i1b02fs.f1	138	9.1e-08	127	534	pir||T38296 nitrogen permease regulator homolog - fission yeast(Schizosaccharomyces pombe) >emb|CAB16240.1| (Z99163) similar tonitrogen permease regulator. [Schizosaccharomyces pombe]
m1f03fs.f1	107	0.006	130	342	gb|AAF21172.1|AC0109 (AC010926) putative exportin, tRNA (nuclear exportreceptor for tRNAs) [Arabidopsis thaliana]

Part III Unclassified protein

I. Classes of Enzymes (from M. Reily and KEGG)

1. Oxidoreductases (39)

<reductase>					
n2d10fs.r1	669	6.4e-65	119	493	dbj|BAA33773.1| (AB014493) reductase [Gibberella zeae]
n2d10fs.f1	371	2.7e-33	303	542	dbj|BAA33773.1| (AB014493) reductase [Gibberella zeae]
<dimethylaniline monooxygnease>					
Contig244	233	7.4e-18	2	526	sp|P49109|FMO5_CAVPO DIMETHYLANILINE MONOOXYGENASE [N-OXIDE FORMING]5(HEPATIC FLAVIN-CONTAINING MONOOXYGENASE 5) (FMO 5)(DIMETHYLANILINE OXIDASE 5) >pir||S71617 dimethylanilinemonooxygenase (N-oxide-forming) (EC1.14.13.8) FMO5 - guinea pig>gb|AAA67848.1| (L37081) flavin containing monooxygenase 5 [Caviaporcellus]

<OXIDOREDUCTASE>				
Contig422	878	5.2e-87	7 789	pir||T37167 probable oxidoreductase - Streptomyces coelicolor >emb|CAB53292.1|(AL109972) putative oxidoreductase [Streptomyces coelicolor A3(2)]
Contig760	402	1.4e-36	177 590	emb|CAB61541.1| (AL133171) putative oxidoreductase [Streptomyces coelicolorA3(2)]
t4e11fs.f1	376	7.9e-34	182 577	pir||T37167 probable oxidoreductase - Streptomyces coelicolor>emb|CAB53292.1|(AL109972) putative oxidoreductase [Streptomyces coelicolor A3(2)]
Contig621	337	1.1e-29	78 482	pir||T39296 Oxidoreductase - fission yeast (Schizosaccharomyces pombe)>emb|CAA17915.1| (AL022105) Oxidoreductase [Schizosaccharomycespombe]
g3g09fs.r1	247	3.5e-20	38 406	emb|CAA22691.1| (AL035159) putative oxidoreductase [Mycobacterium leprae]
n4e07fs.r1	227	5.9e-18	140 490	gi|6319635 ref|NP_009717.1|YBR159W| Ybr159wp>sp|P38286|YB09_YEASTHYPOTHETICAL OXIDOREDUCTASE IN RPB5-CDC28 INTERGENIC REGION>pir||S46030 probable membrane protein YBR159w - yeast(Saccharomyces cerevisiae) >emb|CAA85118.1| (Z36028) ORF YBR159w[Saccharomyces cerevisiae]
Contig78	175	5.4e-12	95 433	sp|O32223|YVAA_BACSU HYPOTHETICAL OXIDOREDUCTASE IN FHUD-OPUBD INTERGENICREGION >pir||G70026 conserved hypothetical protein yvaA - Bacillussubtilis >emb|CAB15358.1| (Z99121) similar to hypothetical proteins[Bacillus subtilis]
b3d10fs.r1	170	2.5e-11	69 446	emb|CAB61541.1| (AL133171) putative oxidoreductase [Streptomyces coelicolorA3(2)]
<MONOOXYGENASE>				
a4c06fs.r1	307	3.7e-26	86 478	pir||C70655 probable monooxygenase - Mycobacterium tuberculosis (strain H37RV)>emb|CAB06212.1| (Z83864) hypothetical protein Rv3854c[Mycobacterium tuberculosis]
Contig804	198	4.4e-14	126 374	gi|4503755 ref|NP_002012.1|| flavin containing monooxygenase 1>sp|Q01740|FMO1_HUMAN DIMETHYLANILINE MONOOXYGENASE [N-OXIDEFORMING] 1 (FETAL HEPATIC FLAVIN-CONTAINING MONOOXYGENASE 1) (FMO1) (DIMETHYLANILINE OXIDASE 1) >pir||A40876 dimethylanilinemonooxygenase (N-oxide-forming) (EC1.14.13.8), hepatic 1 - human>gb|AAA52457.1| (M64082) flavin-containing monooxygenase [Homosapiens]
<GMC oxidoreductase>				

s2c04fs.r1	259	9.5e-21	81 413	pir||C75453 GMC oxidoreductase - Deinococcus radiodurans (strain R1)>gb|AAF10542.1|AE001949_1 (AE001949) GMC oxidoreductase[Deinococcus radiodurans]
<dihydroxy-acid dehydratase precursor>				
Contig874	862	2.4e-85	119 919	sp|Q10318|ILV3_SCHPO PUTATIVE DIHYDROXY-ACID DEHYDRATASE,MITOCHONDRIALPRECURSOR (DAD) (2,3-DIHYDROXY ACID HYDROLYASE)>pir||T37858probable dihydroxy-acid dehydratase precursor - fission yeast(Schizosaccharomyces pombe) >emb|CAA93689.1| (Z69795)putativedihydroxy-acid dehydratase precursor [Schizosaccharomyces pombe]
Contig267	114	0.0038	563 673	sp|Q10318|ILV3_SCHPO PUTATIVE DIHYDROXY-ACID DEHYDRATASE,MITOCHONDRIALPRECURSOR (DAD) (2,3-DIHYDROXY ACID HYDROLYASE)>pir||T37858probable dihydroxy-acid dehydratase precursor - fission yeast(Schizosaccharomyces pombe) >emb|CAA93689.1| (Z69795)putativedihydroxy-acid dehydratase precursor [Schizosaccharomyces pombe]
<PHENOL 2-MONOOXYGENASE>				
b3a06fs.r1	227	5.1e-17	5 343	sp|P15245|PH2M_TRICU PHENOL 2-MONOOXYGENASE (PHENOL HYDROXYLASE)>gb|AAA34202.1| (L04488) phenol hydroxylase [Trichosporon cutaneum]
Contig389	170	6.7e-11	109 462	sp|P15245|PH2M_TRICU PHENOL 2-MONOOXYGENASE (PHENOL HYDROXYLASE)>gb|AAA34202.1| (L04488) phenol hydroxylase [Trichosporon cutaneum]
<amine oxidase-OXIDATIVE DEAMINATION of AMINES>				
Contig82	468	5.2e-61	79 456	sp|Q12556|AMO1_ASPNG COPPER AMINE OXIDASE 1 >pir||S71320 amine oxidase(copper-containing) (EC 1.4.3.6) - Aspergillus niger>gb|AAB03385.1| (U31869) copper amine oxidase [Aspergillus niger]
Contig463	380	1.3e-33	11 448	sp|Q12556|AMO1_ASPNG COPPER AMINE OXIDASE 1 >pir||S71320 amine oxidase(copper-containing) (EC 1.4.3.6) - Aspergillus niger>gb|AAB03385.1| (U31869) copper amine oxidase [Aspergillus niger]
n3d08fs.f1	268	2e-21	196 528	sp|Q12556|AMO1_ASPNG COPPER AMINE OXIDASE 1 >pir||S71320 amine oxidase(copper-containing) (EC 1.4.3.6) - Aspergillus niger>gb|AAB03385.1| (U31869) copper amine oxidase [Aspergillus niger]
Contig275	260	1.9e-20	10 606	dbj|BAA88896.1| (AB019242) semicarbazide-sensitive amine oxidase [Bos taurus]

n3d12fs.r1	140	1.1e-07	87	533	sp|Q07121|AMO1_ARTS1 COPPER AMINE OXIDASE PRECURSOR (MAOXI)>gb|AAA22076.1|(L12983) amine oxidase [Arthrobacter sp.]
Contig729	106	0.029	341	607	gi|4502119 ref|NP_003725.1|| copper containing amine oxidase 3 precursor;vascular adhesion protein 1 >sp|Q16853|AOC3_HUMAN MEMBRANE COPPERAMINE OXIDASE (VASCULAR ADHESION PROTEIN-1) (VAP-1) (HPAO)>pir||JC5234 amine oxidase (copper-containing) (EC 1.4.3.6)precursor - human >gb|AAC50919.1| (U39447) copper monamine oxidase[Homo sapiens] >gb|AAC25170.1| (AF067406) vascular adhesionprotein-1; semicarbazide
<fructosyl amino acid oxidase>					
a4g02fs.f1	276	5.8e-23	109	489	emb|CAA70218.1| (Y09020) fructosyl amino acid oxidase [Aspergillus terreus]
e1g04fs.r1	161	2.7e-10	92	316	pir||T40295 fructosyl amine - fission yeast (Schizosaccharomyces pombe)>emb|CAA17815.1| (AL022071) putative fructosyl amino acid oxidase[Schizosaccharomyces pombe]
e1g04fs.f1	159	4.6e-10	316	495	pir||T40295 fructosyl amine - fission yeast (Schizosaccharomyces pombe)>emb|CAA17815.1| (AL022071) putative fructosyl amino acid oxidase[Schizosaccharomyces pombe]
<fructosyl amine:oxygen oxidoreductase>					
a4g02fs.r1	239	3.3e-32	174	470	gb|AAB88209.1| (AF035700) fructosyl amine:oxygen oxidoreductase [Aspergillusfumigatus]
<cytochrome P450 monooxygenase>					
Contig993	376	8e-34	320	799	emb|CAA75565.1| (Y15277) cytochrome P450 monooxygenase [Gibberella fujikuroi]
a1h08fs.r1	310	2.5e-26	23	424	emb|CAA75565.1| (Y15277) cytochrome P450 monooxygenase [Gibberella fujikuroi]
m2h11fs.f1	252	5.7e-20	157	498	emb|CAA75566.1| (Y15278) cytochrome P450 monooxygenase [Gibberella fujikuroi]
Contig956	242	7.8e-19	153	650	emb|CAA75565.1| (Y15277) cytochrome P450 monooxygenase [Gibberella fujikuroi]
<NADH-dependent butanol dehydrogenase>					
c4g09fs.r1	114	3.2e-05	145	423	gb|AAD19418.1| (AF102543) NADH-dependent butanol dehydrogenase [Zymomonasmobilis]
<ascorbate oxidase>					
i2f05fs.r1	343	6.6e-30	6	422	dbj|BAA24288.1| (AB010110) ascorbate oxidase [Acremonium sp.]
i2f05fs.f1	271	5.8e-22	212	466	dbj|BAA24288.1| (AB010110) ascorbate oxidase [Acremonium sp.]
<L-gulonolactone oxidase>					

e2c01fs.r1	240	7.2e-19	114	437	pir||OXRTGU L-gulonolactone oxidase (EC 1.1.3.8) - rat>dbj|BAA02232.1|(D12754) L-gulono-gamma-lactone oxidase [Rattus norvegicus]
<desaturase>					
Contig252 .	262	9.2e-22	144	398	gi|6323928 ref|NP_013999.1|SCS7| desaturase/hydroxylase enzyme;Scs7p>sp|Q03529|YM8I_YEAST HYPOTHETICAL 44.9 KD PROTEIN IN URA10-NRC1INTERGENIC REGION >pir||S54484 probable membrane protein YMR272c-yeast (Saccharomyces cerevisiae) >emb|CAA89255.1| (Z49260)unknown[Saccharomyces cerevisiae]
Contig315	175	6.8e-12	352	510	gb|AAC99343.1| (AF069752) C5,6 desaturase [Candida albicans]
<2-oxoglutarate dehydrogenase e1 component>					
o4f04fs.r1	603	3.3e-57	7	513	pir||T40412 2-oxoglutarate dehydrogenase e1 component - fission yeast(Schizosaccharomyces pombe) >emb|CAA20299.1| (AL031261)2-oxoglutarate dehydrogenase e1 component [Schizosaccharomycespombe]
<short chain dehydrogenase>					
k2a03fs.r1	185	3.8e-13	126	440	pir||T41570 hypothetical protein SPCC736.13 - fission yeast(Schizosaccharomyces pombe) >emb|CAA19277.1| (AL023705)hypothetical short chain dehydrogenase. [Schizosaccharomyces pombe]
2. Transferases (5)					
<methyl chloride transferase>					
n1b07fs.f1	140	9.5e-09	279	509	gb|AAC72357.1| (AF084829) methyl chloride transferase [Batis maritima]>gb|AAD26120.1|AF109128_1 (AF109128) methyl chloride transferase[Batis maritima]
<NrgA-N-acetyltransferase homolog>					
Contig408	101	0.016	147	317	gb|AAF03159.1|AF1907 (AF190732) NrgA [Bradyrhizobium japonicum]
<o-methyltransferase>					
Contig369	234	7.5e-19	84	485	pir||T39301 probable o-methyltransferase - fission yeast (Schizosaccharomycespombe) >emb|CAA17918.1| (AL022117) putative catecholo-methyltransferase [Schizosaccharomyces pombe]
e1h05fs.f1	136	5.1e-08	265	471	pir||T39301 probable o-methyltransferase - fission yeast (Schizosaccharomycespombe) >emb|CAA17918.1| (AL022117) putative catecholo-methyltransferase [Schizosaccharomyces pombe]
<methyltransferase>					
i2e09fs.r1	102	0.00034	304	420	emb|CAB76043.1| (AL157918) putative methyltransferase [Schizosaccharomycespombe]
3.Hydrolases (8)					
<alkaline phosphatase>					

t2e02fs.r1	690	3.9e-67	9	545	gb|AAA21821.1| (L27993) alkaline phosphatase [Neurospora crassa]
Contig122	529	4.4e-50	10	597	gi|6320689 ref|NP_010769.1|PHO8| repressible alkaline phosphatase; Pho8p>sp|P11491|PPB_YEAST REPRESSIBLE ALKALINE PHOSPHATASE PRECURSOR>pir||S69648 alkaline phosphatase (EC 3.1.3.1) - yeast(Saccharomyces cerevisiae) >gb|AAB64930.1| (U33050) Pho8p:repressible alkaline phosphatase; CAI: 0.16 [Saccharomycescerevisiae]
Contig849	242	8.4e-19	255	770	sp|P42251|PPBD_BACSU ALKALINE PHOSPHATASE D PRECURSOR (APASED) (RAN1) (BC6)>pir||D69676 phosphodiesterase/alkaline phosphatase phoD -Bacillussubtilis >gb|AAB47803.1| (U49060)phosphodiesterase/alkalinephosphatase D precursor [Bacillus subtilis]>emb|CAB12056.1|(Z99105) phosphodiesterase/alkaline phosphatase D [Bacillussubtilis]
Contig458	178	6.6e-12	286	459	pir||T39459 repressible alkaline phosphatase precursor - fission yeast(Schizosaccharomyces pombe) >emb|CAA19331.1| (AL023780)repressiblealkaline phosphatase precursor [Schizosaccharomyces pombe]
Contig735	108	0.0032	218	358	pir||T35097 probable secreted alkaline phosphatase - Streptomyces coelicolor>emb|CAB51460.1| (AL096884) putative secreted alkaline phosphatase[Streptomyces coelicolor A3(2)]
<secreted acid phosphatase 2>					
Contig841	178	3.4e-10	25	879	emb|CAA87091.1| (Z46970) secreted acid phosphatase 2 (SAP2) [Leishmaniamexicana]
<acid phosphatase>					
g2a07fs.r1	204	7.4e-15	106	456	pir||T40420 probable acid phosphatase - fission yeast (Schizosaccharomycespombe) >emb|CAB58405.1| (AL121863) putative acid phosphatase[Schizosaccharomyces pombe]
<dihydroxyacid dehydratase>					
Contig16	346	4e-30	117	431	gi|6322476 ref|NP_012550.1|ILV3| dihydroxyacid dehydratase;Ilv3p>sp|P39522|ILV3_YEAST DIHYDROXY-ACID DEHYDRATASE,MITOCHONDRIALPRECURSOR (DAD) (2,3-DIHYDROXY ACID HYDROLYASE)>pir||S55205dihydroxy-acid dehydratase (EC 4.2.1.9) - yeast (Saccharomycescerevisiae) >emb|CAA60939.1| (X87611) dihydroxyacid dehydratase[Saccharomyces cerevisiae] >emb|CAA89540.1| (Z49516) ORF YJR016c[Saccharomyces cerevisiae]
<esterase D>					
e2e10fs.f1	188	6.3e-14	306	548	sp|P10768|ESTD_HUMAN ESTERASE D >gb|AAC99788.1| (AF112219) esterase D [Homosapiens]

<esterase>

Contig864	275	4.1e-23	173	679	gb|AAC01724.1| (AF040570) esterase [Amycolatopsis mediterranei]
m3g06fs.r1	154	9.5e-10	280	492	pdb|1QLW|A Chain A, The Atomic Resolution Structure Of A Novel BacterialEsterase

4. Lyases

<LACTOYLGLUTATHIONE LYASE>

a1g08fs.f1	308	1.2e-26	151	519	sp|Q09751|LGUL_SCHPO PROBABLE LACTOYLGLUTATHIONE LYASE (METHYLGLYOXALASE) (ALDOKETOMUTASE) (GLYOXALASE I) (GLX I) (KETONE-ALDEHYDE MUTASE) (S-D-LACTOYLGLUTATHIONE METHYLGLYOXAL LYASE) >pir||T11675lactoylglutathione lyase (EC 4.4.1.5) - fission yeast (Schizosaccharomyces pombe) >pir||T39369 lactoylglutathione lyase-fission yeast (Schizosaccharomyces pombe) >emb|CAA90825.1| (Z54140) lactoylglutathione lyase
a1g08fs.r1	277	2.3e-23	124	438	gi|6323639 ref|NP_013710.1|GLO1| lactoylglutathione lyase (glyoxalase I);Glo1p >sp|P50107|LGUL_YEAST LACTOYLGLUTATHIONE LYASE (METHYLGLYOXALASE) (ALDOKETOMUTASE) (GLYOXALASE I) (GLX I) (KETONE-ALDEHYDE MUTASE) (S-D-LACTOYLGLUTATHIONE METHYLGLYOXALLYASE) >pir||S55115 GLO1 protein - yeast (Saccharomyces cerevisiae) >emb|CAA89948.1| (Z49810) unknown [Saccharomyces cerevisiae] >emb|CAA67622.1| (X99240)

5. Isomerases

6. Ligases

II. Non-enzymatic classes (not in defined pathways) (27)

<glycoprotein 900>

Contig77	100	0.28	70	717	pir||T31113 mucin-like glycoprotein 900 - Cryptosporidium parvum>gb|AAC98153.1| (AF068065) GP900; mucin-like glycoprotein [Cryptosporidium parvum]

<proline-rich protein>

Contig375	166	1.1e-11	303	563	emb|CAA07370.1| (AJ006984) proline-rich protein [Capsicum annuum]

<trp-asp repeat protein>

d1d03fs.r1	167	9.6e-11	26	496	pir||T38653 trp-asp repeat protein - fission yeast (Schizosaccharomyces pombe) >emb|CAB52267.1| (AL109739) trp-asp repeat protein [Schizosaccharomyces pombe]

<selenoprotein P precursor>

p3e06fs.r1	710	3.1e-69	65	523	pir||T10442 selenoprotein P precursor - mouse >emb|CAA68140.1| (X99807) Selenoprotein P [Mus musculus]

<F-box protein Fbl7>

n2f08fs.f1	95	0.13	287	499	dbj|BAA74863.1| (AB020647) KIAA0840 protein [Homo sapiens]>gb|AAF04514.1|AF174593_1 (AF174593) F-box protein Fbl7 [Homosapiens]
<GNS1/SUR4 family protein>					
b1d05fs.r1	158	3.4e-10	236	436	emb|CAB61470.1| (AL133157) GNS1/SUR4 family protein [Schizosaccharomycespombe]
<WD repeat protein>					
l3a03fs.r1	330	3.8e-28	8	511	emb|CAB76232.1| (AL157993) WD repeat protein [Schizosaccharomyces pombe]
<C2-domain synaptic vesicle protein>					
g4c03fs.f1	452	1.3e-40	8	478	pir||T41696 probable C2-domain synaptic vesicle protein - fission yeast(Schizosaccharomyces pombe) >emb|CAB58372.1| (AL121859)C2-domain,putative synaptic vesicle protein [Schizosaccharomyces pombe]
Contig249	100	0.004	169	327	pir||T41696 probable C2-domain synaptic vesicle protein - fission yeast(Schizosaccharomyces pombe) >emb|CAB58372.1| (AL121859)C2-domain,putative synaptic vesicle protein [Schizosaccharomyces pombe]
<nmt1 protein-the expression of this gene has been shown to be completely inhibited by thiamine>					
Contig3	481	5.8e-45	149	448	sp|P42882|NMT1_ASPPA NMT1 PROTEIN HOMOLOG >pir||S53697 nmt1 protein-Aspergillus parasiticus >gb|AAA70083.1| (U15196) the expression ofthis gene has been shown to be completely inhibited by thiamine aswas observed for the Schizosaccharomyces pombe nmt1, Swiss-ProtAccession Number P36597 [Aspergillus parasiticus] >prf||2109237Anmt1 gene [Aspergillus parasiticus]
Contig200	385	8.1e-35	1	285	sp|P42882|NMT1_ASPPA NMT1 PROTEIN HOMOLOG >pir||S53697 nmt1 protein-Aspergillus parasiticus >gb|AAA70083.1| (U15196) the expression ofthis gene has been shown to be completely inhibited by thiamine aswas observed for the Schizosaccharomyces pombe nmt1, Swiss-ProtAccession Number P36597 [Aspergillus parasiticus] >prf||2109237Anmt1 gene [Aspergillus parasiticus]
Contig727	272	8.2e-23	384	539	sp|P42882|NMT1_ASPPA NMT1 PROTEIN HOMOLOG >pir||S53697 nmt1 protein-Aspergillus parasiticus >gb|AAA70083.1| (U15196) the expression ofthis gene has been shown to be completely inhibited by thiamine aswas observed for the Schizosaccharomyces pombe nmt1, Swiss-ProtAccession Number P36597 [Aspergillus parasiticus] >prf||2109237Anmt1 gene [Aspergillus parasiticus]
<regulatory protein>					
p1g10fs.r1	248	2.7e-20	11	166	pir||S55723 pac2 protein - fission yeast (Schizosaccharomyces pombe)>pir||T38628 camp independent regulatory protein - fission

					yeast(Schizosaccharomyces pombe) >dbj|BAA07805.1| (D43748)Pac2p[Schizosaccharomyces pombe] >emb|CAB11695.1| (Z98979)camp-independent regulatory protein pac2. [Schizosaccharomycespombe]
k3e12fs.r1	241	2.4e-18	132	452	sp|P21228|ALCR_EMENI REGULATORY PROTEIN ALCR
j1d08fs.r1	113	0.00011	57	296	pir||T38690 probable regulatory protein - fission yeast (Schizosaccharomycespombe) >emb|CAB16735.1| (Z99568) putative regulatory protein; zincfinger [Schizosaccharomyces pombe]
<putative gamma-adaptin>					
b4b05fs.r1	240	3.4e-18	210	452	pir||T41685 probable gamma-adaptin - fission yeast (Schizosaccharomyces pombe)>emb|CAB54865.1| (AL117183) putative gamma-adaptin[Schizosaccharomyces pombe]
b4b05fs.f1	120	2.3e-05	225	425	pir||T41685 probable gamma-adaptin - fission yeast (Schizosaccharomyces pombe)>emb|CAB54865.1| (AL117183) putative gamma-adaptin[Schizosaccharomyces pombe]
<ring-box protein 1>					
Contig310	515	1.5e-48	190	480	gb|AAD29715.1|AF1405 (AF140598) ring-box protein 1 [Homo sapiens]>gb|AAD29716.1|AF140599_1 (AF140599) ring-box protein 1 [Musmusculus] >gb|AAD30146.1|AF142059_1 (AF142059) RING finger protein[Homo sapiens] >emb|CAB62925.1| (AL080242) bA554C12.1 (RBX1 or ROC1(ring-box or ring finger protein 1)) [Homo sapiens]
<nebulette>					
f1f11fs.f1	170	1.2e-10	109	312	gb|AAF24858.1|AF0473 (AF047368) nebulette [Homo sapiens]
f1c02fs.r1	158	2.5e-09	122	226	gb|AAF24858.1|AF0473 (AF047368) nebulette [Homo sapiens]
<SWH1 protein-from yeast encodes a candidate nuclear factor containing>					
<ankyrin repeats and showing homology to mammalian oxysterol-binding protein>					
d3d08fs.r1	326	3.9e-27	21	446	gi|6320185 ref|NP_010265.1|YDL019C| Ydl019cp >pir||S52500 SWH1 protein homologYDL019c - yeast (Saccharomyces cerevisiae)>emb|CAA88340.1|(Z48432) homolog of yeast SWH1 protein (X74552) [Saccharomycescerevisiae] >emb|CAA98578.1| (Z74067) ORF YDL019c [Saccharomycescerevisiae]
<GlxB6-Escherichia coli glyoxylate induced protein>					
f3g03fs.r1	131	9.8e-08	87	494	gb|AAB93856.1| (U89279) GlxB6 [Escherichia coli]
<CROSS-PATHWAY CONTROL PROTEIN 1>					
Contig959	255	5e-21	172	519	sp|P11115|CPC1_NEUCR CROSS-PATHWAY CONTROL PROTEIN 1 >gb|AAA33577.1| (J03262)cross-pathway control protein 1 [Neurospora crassa]

Contig938	182	1.6e-11	907	1149	pir||A30208 cross-pathway control protein 1 - Neurospora crassa

Part IV. Unidentified (includes significant match with ORFs) (256)

<unknown function>

Contig1024	1936	3.8e-199	61	2502	sp|Q10094|YAOF_SCHPO HYPOTHETICAL 143.7 KD PROTEIN C11D3.15 IN CHROMOSOME I>pir||T37524 probable oxoprolinase - fission yeast(Schizosaccharomyces pombe) >emb|CAA92316.1| (Z68166) putativeoxoprolinase [Schizosaccharomyces pombe]
Contig1013	1726	7e-177	9	1547	pir||S76896 hypothetical protein - Synechocystis sp. (strain PCC 6803)>dbj|BAA18808.1| (D90917) hypothetical protein [Synechocystis sp.]
Contig1054	829	7.9e-82	213	1259	pir||B31776 hypothetical protein (LAC12 3' region) - yeast (Kluyveromycesmarxianus var. lactis) >emb|CAA30054.1| (X06997) ORF II[Kluyveromyces lactis]
Contig887	766	3.2e-75	6	968	pir||T32749 hypothetical protein F57B10.3 - Caenorhabditis elegans>gb|AAB96720.1| (AF039713) Similar to phosphoglycerate mutase;coded for by C. elegans cDNA yk357d11.5; coded for by C. eleganscDNA yk387c10.5; coded for by C. elegans cDNA yk384f12.5; coded forby C. elegans cDNA cm10f9; coded for by C. elegans cDNA cm18g2;coded fo>
Contig222	741	1.6e-72	11	766	gi|6321159 ref|NP_011237.1|YFR044C| Yfr044cp >sp|P43616|YFL4_YEASTHYPOTHETICAL 52.9 KD PROTEIN IN SAP155-YMR31 INTERGENIC REGION>pir||S56299 hypothetical protein YFR044c - yeast (Saccharomycescerevisiae) >dbj|BAA09283.1| (D50617) YFR044C [Saccharomycescerevisiae] >dbj|BAA08010.1| (D44597) unknown [Saccharomycescerevisiae]
b2d01fs.r1	536	2.5e-50	22	423	pir||S76896 hypothetical protein - Synechocystis sp. (strain PCC 6803)>dbj|BAA18808.1| (D90917) hypothetical protein [Synechocystis sp.]
Contig374	525	1.1e-49	41	712	gi|6321079 ref|NP_011157.1|YFL030W| Yfl030wp>sp|P43567|YFD0_YEASTHYPOTHETICAL 41.9 KD PROTEIN IN HAC1-CAK1 INTERGENIC REGION>pir||S56224 hypothetical protein YFL030w - yeast (Saccharomycescerevisiae) >dbj|BAA09208.1| (D50617) YFL030W [Saccharomycescerevisiae]
13a02fs.r1	506	1.3e-47	9	512	sp|Q10211|YAY3_SCHPO HYPOTHETICAL 74.5 KD PROTEIN C4H3.03C IN CHROMOSOME I>pir||T38883 hypothetical protein SPAC4H3.03c - fission yeast(Schizosaccharomyces pombe) >emb|CAA93342.1| (Z69380) hypotheticalprotein [Schizosaccharomyces pombe]

Contig494	512	2.4e-47	17	499	gi|6319862 ref|NP_009943.1|YCR017C| Ycr017cp>sp|P25618|YCQ7_YEASTHYPOTHETICAL 107.9 KD PROTEIN IN POL4-SRD1 INTERGENIC REGION>pir||S19427 probable membrane protein YCR017c - yeast(Saccharomyces cerevisiae) >emb|CAA42308.1| (X59720) YCR017c,len:953 [Saccharomyces cerevisiae]
l3c02fs.f1	492	3.6e-46	77	520	sp|P52495|UBA1_CANAL UBIQUITIN-ACTIVATING ENZYME E1 1 >pir||T18215hypothetical protein - yeast (Candida albicans) (fragment)>gb|AAC49911.1| (U13193) similar to the 3' end of UBA1:Swiss-ProtAccession Number P22515 [Candida albicans]
Contig149	502	6.1e-46	23	433	gi|6322634 ref|NP_012707.1|YKL215C| Ykl215cp>sp|P28273|YKV5_YEASTHYPOTHETICAL 140.4 KD PROTEIN IN URA1-DOA1 INTERGENIC REGION>pir||S38058 hypothetical protein YKL215c - yeast (Saccharomycescerevisiae) >emb|CAA53558.1| (X75951) ORF4, F1286 [Saccharomycescerevisiae] >emb|CAA82060.1| (Z28215) ORF YKL215c [Saccharomycescerevisiae]
e4d09fs.r1	487	1.4e-45	20	523	pir||T39599 conserved hypothetical protein SPBC16G5.07c - fission yeast(Schizosaccharomyces pombe) >emb|CAA19027.1| (AL023554)conservedhypothetical protein. [Schizosaccharomyces pombe]
m4d03fs.r1	472	4.8e-44	56	484	emb|CAB06790.1| (Z86109) unknown [Saccharomyces pastorianus]
Contig880	469	9.3e-44	44	583	dbj|BAA13835.1| (D89173) similar to Saccharomyces cerevisiae hypothetical36.4KD protein in SOD1-CPA2 intergenic region, SWISS-PROT AccessionNumber P47143 [Schizosaccharomyces pombe]
Contig611	457	1.8e-42	178	615	pir||T39412 hypothetical protein SPBC13G1.11 - fission yeast(Schizosaccharomyces pombe) >emb|CAA18664.1| (AL022600)Synaptobrevin-like V snare protein [Schizosaccharomyces pombe]
Contig730	455	2.8e-42	111	602	emb|CAA60486.1| (X86790) N3810 [Saccharomyces cerevisiae]>emb|CAA96358.1|(Z71689) ORF YNR073c [Saccharomyces cerevisiae]
o1b08fs.r1	455	2.9e-42	19	375	dbj|BAA13821.1| (D89159) similar to Saccharomyces serevisiae hypothetical 75.5KD protein in CCT3-CCT8 intergenic region, SWISS-PROT AccessionNumber P47075 [Schizosaccharomyces pombe]
Contig768	440	1.4e-40	148	1236	pir||T39513 hypothetical protein SPBC1604.01 - fission yeast(Schizosaccharomyces pombe) (fragment) >emb|CAA22334.1| (AL034433)hypothetical protein [Schizosaccharomyces pombe]
Contig100	434	4.9e-40	123	485	pir||T40926 conserved hypothetical protein SPCC1281.07c - fission yeast(Schizosaccharomyces pombe) >emb|CAA22828.1| (AL035218)

					proteinwith Glutathione S transferase domain [Schizosaccharomyces pombe]
Contig347	431	1.1e-39	267	731	gi|6324771 ref|NP_014840.1|YOR197W| Yor197wp >pir||S67089 hypothetical proteinYOR197w - yeast (Saccharomyces cerevisiae)>emb|CAA99410.1|(Z75105) ORF YOR197w [Saccharomyces cerevisiae]
t2h03fs.r1	419	1.9e-38	73	552	pir||T40385 hypothetical protein SPBC3E7.11c - fission yeast(Schizosaccharomyces pombe) >emb|CAA19014.1| (AL023534) DNA Jdomain protein [Schizosaccharomyces pombe]
l2a11fs.r1	422	2.1e-38	18	521	gi|6324644 ref|NP_014713.1|GYP1| GTPase activating protein; Gyp1p>pir||S66953hypothetical protein YOR070c - yeast (Saccharomyces cerevisiae)>emb|CAA99263.1| (Z74978) ORF YOR070c [Saccharomyces cerevisiae]>emb|CAA94555.1| (Z70678) YOR29-21 [Saccharomyces cerevisiae]
Contig303	418	2.6e-38	117	557	sp|Q10243|YD1D_SCHPO HYPOTHETICAL 35.8 KD PROTEIN C4G9.13C IN CHROMOSOME I>pir||T38872 probable vacuolar sorting protein - fission yeast(Schizosaccharomyces pombe) >emb|CAA93563.1| (Z69727) putativevacuolar sorting protein [Schizosaccharomyces pombe]
b4d08fs.r1	415	5.9e-38	106	450	pir||T33344 hypothetical protein K07D4.3 - Caenorhabditis elegans>gb|AAC26287.1| (AF077534) similar to the proteasome regulatorysubunit [Caenorhabditis elegans]
Contig935	414	7.3e-38	31	822	pir||T30216 hypothetical protein Z - Streptomyces hygroscopicus>emb|CAA60450.1| (X86780) orfZ [Streptomyces hygroscopicus]
f2b11fs.r1	416	1.2e-37	32	640	pir||T32472 hypothetical protein F08F1.7 - Caenorhabditis elegans>gb|AAB71307.1| (AF026213) strong similarity to Saccharomycescerevisiae endosomal P24A protein (SP:P32802) [Caenorhabditiselegans]
j1a01fs.r1	406	5.3e-37	53	460	sp|O14179|YDS3_SCHPO HYPOTHETICAL 28.1 KD PROTEIN C4F8.03 IN CHROMOSOME I>pir||T38833 hypothetical UPF0023 family protein - fission yeast(Schizosaccharomyces pombe) >emb|CAB11050.1| (Z98530) hypotheticalUPF0023 family protein [Schizosaccharomyces pombe]
k3d03fs.r1	418	5.7e-37	31	474	gi|6322634 ref|NP_012707.1|YKL215C| Ykl215cp>sp|P28273|YKV5_YEASTHYPOTHETICAL 140.4 KD PROTEIN IN URA1-DOA1 INTERGENIC REGION>pir||S38058 hypothetical protein YKL215c - yeast (Saccharomycescerevisiae) >emb|CAA53558.1| (X75951) ORF4, F1286 [Saccharomycescerevisiae] >emb|CAA82060.1| (Z28215) ORF YKL215c [Saccharomycescerevisiae]

s1e03fs.r1	281	7.4e-37	11	310	sp|O14301|YE85_SCHPO HYPOTHETICAL 65.3 KD TRP-ASP REPEATS CONTAINING PROTEINC9G1.05 IN CHROMOSOME I >pir||T39228 beta-transducin - fissionyeast (Schizosaccharomyces pombe) >emb|CAB11489.1| (Z98763)putative WD repeat stress protein [Schizosaccharomyces pombe]
Contig861	409	2e-36	195	602	pir||S76896 hypothetical protein - Synechocystis sp. (strain PCC 6803)>dbj|BAA18808.1| (D90917) hypothetical protein [Synechocystis sp.]
g1g01fs.r1	397	3.8e-36	128	469	gb|AAC44550.1| (U34346) unknown [Paracoccus denitrificans]
l4e05fs.r1	390	2.4e-35	49	531	gi|6323882 ref|NP_013953.1|YMR226C| Ymr226cp>sp|Q05016|YM71_YEASTHYPOTHETICAL OXIDOREDUCTASE IN MRPL44-MTF1 INTERGENIC REGION>pir||S57593 hypothetical protein YMR226c - yeast (Saccharomycescerevisiae) >emb|CAA90197.1| (Z49939) unknown [Saccharomycescerevisiae]
k4f11fs.r1	389	3.1e-35	109	522	pir||T40726 hypothetical protein SPBC887.01 - fission yeast(Schizosaccharomyces pombe) >emb|CAA21886.1| (AL033388)hypothetical protein [Schizosaccharomyces pombe]
d2c05fs.r1	226	4.6e-35	18	215	pir||T41172 hypothetical protein SPCC1840.04 - fission yeast(Schizosaccharomyces pombe) >emb|CAA20127.1| (AL031179)hypothetical protein [Schizosaccharomyces pombe]
g4h08fs.r1	385	8.9e-35	111	488	pir||T11684 RVS167 protein homolog - fission yeast (Schizosaccharomyces pombe)>emb|CAA20768.1| (AL031536) hypothetical protein[Schizosaccharomyces pombe]
o2g01fs.r1	383	1.3e-34	10	462	pir||T39419 hypothetical coiled-coil protein - fission yeast(Schizosaccharomyces pombe) >emb|CAB46758.1| (AL096797)hypothetical coiled-coil protein [Schizosaccharomyces pombe]
Contig502	387	1.4e-34	92	475	pir||T40481 hypothetical protein SPBC4B4.09 - fission yeast(Schizosaccharomyces pombe) >emb|CAA19289.1| (AL023706)possiblepre-mRNA processing by similarity to yeast prp39[Schizosaccharomyces pombe]
i4e05fs.r1	319	1.7e-34	19	309	gi|6325116 ref|NP_015184.1|YPL141C| Ypl141cp >pir||S69044 hypothetical proteinYPL141c - yeast (Saccharomyces cerevisiae)>gb|AAB68219.1| (U43703)Lpi5p [Saccharomyces cerevisiae]
o1a07fs.f1	202	1.4e-33	316	456	sp|Q10138|YAS2_SCHPO HYPOTHETICAL 51.5 KD PROTEIN C3H8.02 IN CHROMOSOME I>pir||T38760 hypothetical protein SPAC3H8.02 - fission yeast(Schizosaccharomyces pombe) >emb|CAA93159.1| (Z69086)hypotheticalprotein [Schizosaccharomyces pombe]

l3a02fs.f1	377	2.5e-33	105 515	sp|Q10211|YAY3_SCHPO HYPOTHETICAL 74.5 KD PROTEIN C4H3.03C IN CHROMOSOME I>pir||T38883 hypothetical protein SPAC4H3.03c - fission yeast(Schizosaccharomyces pombe) >emb|CAA93342.1| (Z69380) hypotheticalprotein [Schizosaccharomyces pombe]
Contig765	366	9.3e-33	223 597	pir||G69896 conserved hypothetical protein yoaN - Bacillus subtilis>gb|AAB84450.1| (AF027868) YoaN [Bacillus subtilis]>emb|CAB13759.1| (Z99114) similar to hypothetical proteins[Bacillus subtilis]
Contig178	365	1.1e-32	44 742	pir||T16025 hypothetical protein F10D7.2 - Caenorhabditis elegans>gb|AAA81720.1| (U40945) coded for by C. elegans cDNA yk74b9.3;coded for by C. elegans cDNA yk74b9.5; similar to repeat ofcalcium channel alpha subunits; similar to tetracycline resistanceprotein; similar to hypothetical protein in HSP30-PMP1 region(SP:YCR3_YEAST>
Contig329	363	5.1e-32	2 286	pir||T16915 hypothetical protein T20H4.3 - Caenorhabditis elegans>gb|AAA50660.1| (U00037) similar to multifunctional aminoacyl-tRNAsynthetase, especially to the prolyl-tRNA synthetase region[Caenorhabditis elegans]
j3b10fs.f1	359	5.1e-32	289 549	sp|Q10243|YD1D_SCHPO HYPOTHETICAL 35.8 KD PROTEIN C4G9.13C IN CHROMOSOME I>pir||T38872 probable vacuolar sorting protein - fission yeast(Schizosaccharomyces pombe) >emb|CAA93563.1| (Z69727) putativevacuolar sorting protein [Schizosaccharomyces pombe]
b4a06fs.f1	367	8.2e-32	7 387	emb|CAA06786.1| (AJ005963) 100 kDa protein [Ajellomyces capsulatus]
j3d11fs.f1	355	1.2e-31	285 545	sp|Q09711|NCS1_SCHPO HYPOTHETICAL CALCIUM-BINDING PROTEIN C18B11.04 INCHROMOSOME I >pir||S58303 related to neuronal calcium sensor 1proteins - fission yeast (Schizosaccharomyces pombe)>emb|CAA90589.1| (Z50728) related to neuronal calcium sensor 1proteins [Schizosaccharomyces pombe]
g3c04fs.f1	352	2.8e-31	203 550	pir||T13417 hypothetical protein EG:171D11.1 - fruit fly (Drosophilamelanogaster) >emb|CAB41309.1| (AL009147) alternatively splicedform; /prediction=(method:""genefinder"", version:""084"");/prediction=(method:""genscan"",version:""1.0"");/match=(desc:""METHYLMALONATE-SEMIALDEHYDE DEHYDROGENASE PRECURSOR[ACYLATING] (EC 1.2.1.27) (MMSDH)"", spec>>gb|AAF45511.1|(AE003417) EG:171D11.1 gene product [alt
d2h06fs.f1	348	6.8e-31	79 447	emb|CAA89772.1| (Z49703) unknown [Saccharomyces cerevisiae]
l2f07fs.r1	347	8.3e-31	15 509	sp|P53990|Y174_HUMAN HYPOTHETICAL PROTEIN KIAA0174 >dbj|BAA11491.1| (D79996)KIAA0174 [Homo sapiens]

h3c07fs.r1	344	2.1e-30	62 511	sp|Q09923|YAKC_SCHPO HYPOTHETICAL 37.7 KD PROTEIN C1F7.12 IN CHROMOSOME I>pir||S62584 hypothetical protein SPAC1F7.12 - fission yeast(Schizosaccharomyces pombe) >pir||T38106 probable oxidoreductase-fission yeast (Schizosaccharomyces pombe) >emb|CAA91959.1| (Z67998)putative oxidoreductase [Schizosaccharomyces pombe]
i2g10fs.f1	340	5.3e-30	19 384	gi|6320168 ref|NP_010248.1|YDL036C| Ydl036cp >sp|Q12069|YD36_YEASTHYPOTHETICAL 53.4 KD PROTEIN IN PRP9-NAT1 INTERGENIC REGION>pir||S67569 hypothetical protein YDL036c - yeast (Saccharomycescerevisiae) >emb|CAA96453.1| (Z71781) unknown [Saccharomycescerevisiae] >emb|CAA98595.1| (Z74084) ORF YDL036c [Saccharomycescerevisiae]
g3f12fs.r1	328	9.2e-29	125 478	pir||T36421 hypothetical protein SCF34.22 - Streptomycescoelicolor>emb|CAB53333.1| (AL109974) hypothetical protein [Streptomycescoelicolor A3(2)]
g3c04fs.r1	329	1.5e-28	129 452	pir||T13417 hypothetical protein EG:171D11.1 - fruit fly (Drosophilamelanogaster) >emb|CAB41309.1| (AL009147) alternatively splicedform; /prediction=(method:""genefinder"", version:""084"");/prediction=(method:""genscan"",version:""1.0"");/match=(desc:""METHYLMALONATE-SEMIALDEHYDE DEHYDROGENASE PRECURSOR[ACYLATING] (EC 1.2.1.27) (MMSDH)"", spec>>gb|AAF45511.1|(AE003417) EG:171D11.1 gene product [alt
k4b07fs.r1	328	1.6e-28	16 381	gi|6321628 ref|NP_011705.1|CRH1| Cell wall protein;Crh1p>sp|P53301|YG46_YEAST HYPOTHETICAL 52.8 KD PROTEIN IN BUB1-HIP1INTERGENIC REGION >pir||S64507 probable membrane protein YGR189c-yeast (Saccharomyces cerevisiae) >emb|CAA97215.1| (Z72974) ORFYGR189c [Saccharomyces cerevisiae] >emb|CAA67525.1| (X99074) G7553[Saccharomyces cerevisiae]
g3d11fs.f1	328	1.7e-28	283 498	emb|CAA58825.1| (X84001) unnamed protein product [Emericella nidulans]
Contig512	324	2.5e-28	344 754	gi|6320246 ref|NP_010326.1|YDR041W| Ydr041wp >pir||S59279 hypothetical proteinYDR041w - yeast (Saccharomyces cerevisiae)>emb|CAA90780.1|(Z54075) unknown [Saccharomyces cerevisiae]
Contig368	324	2.6e-28	54 560	sp|Q22915|YC4P_CAEEL HYPOTHETICAL 34.1 KD PROTEIN C37C3.8 IN CHROMOSOME V>pir||T34398 hypothetical protein C37C3.8 - Caenorhabditis elegans>gb|AAC25863.1| (U64857) C37C3.8 gene product [Caenorhabditiselegans]

q2f11fs.f1	321	5e-28	101 493	sp|Q09923|YAKC_SCHPO HYPOTHETICAL 37.7 KD PROTEIN C1F7.12 IN CHROMOSOME I>pir||S62584 hypothetical protein SPAC1F7.12 - fission yeast(Schizosaccharomyces pombe) >pir||T38106 probable oxidoreductase-fission yeast (Schizosaccharomyces pombe) >emb|CAA91959.1| (Z67998)putative oxidoreductase [Schizosaccharomyces pombe]
Contig680	320	6.9e-28	81 389	pir||T40885 hypothetical protein SPCC1235.11 - fission yeast(Schizosaccharomyces pombe) >emb|CAA21115.2| (AL031764)conservedhypothetical protein [Schizosaccharomyces pombe]
Contig410	319	8.9e-28	228 521	gb|AAC15461.1| (AF060862) unknown [Homo sapiens]
Contig879	314	2.6e-27	227 514	sp|Q09896|YAI9_SCHPO HYPOTHETICAL 13.5 KD PROTEIN C24B11.09 IN CHROMOSOME I>pir||S62554 conserved hypothetical protein SPAC24B11.09 - fissionyeast (Schizosaccharomyces pombe) >emb|CAA91774.1| (Z67757)conserved hypothetical protein [Schizosaccharomyces pombe]
k1e01fs.r1	322	3.1e-27	52 447	gi|6320660 ref|NP_010740.1|YDR452W| Ydr452wp >pir||S69731 hypothetical proteinYDR452w - yeast (Saccharomyces cerevisiae)>gb|AAB64872.1| (U33007)Ydr452wp; CAI: 0.15 [Saccharomyces cerevisiae]
k3e01fs.r1	309	1e-26	165 515	pir||T41383 hypothetical protein SPCC550.08 - fission yeast(Schizosaccharomyces pombe) >emb|CAA19112.1| (AL023592)hypothetical protein [Schizosaccharomyces pombe]
a4e07fs.r1	317	1.2e-26	23 454	gi|6321705 ref|NP_011782.1|YGR266W| Ygr266wp>sp|P53326|YG5L_YEASTHYPOTHETICAL 81.2 KD PROTEIN IN MES1-FOL2 INTERGENIC REGION>pir||S64599 probable membrane protein YGR266w - yeast(Saccharomyces cerevisiae) >emb|CAA97296.1| (Z73051) ORF YGR266w[Saccharomyces cerevisiae]>emb|CAA69197.1| (Y07893) ORF YGR266w[Saccharomyces cerevisiae]
o2g04fs.f1	308	1.2e-26	185 529	gi|6320623 ref|NP_010703.1|YDR415C| Ydr415cp >pir||S69699 hypothetical proteinYDR415c - yeast (Saccharomyces cerevisiae)>gb|AAB64879.1| (U33007)Ydr415cp; CAI: 0.14 [Saccharomyces cerevisiae]
h1d03fs.f1	323	1.3e-26	166 525	gi|6323581 ref|NP_013652.1|YML059C| Yml059cp >sp|Q04958|YMF9_YEASTHYPOTHETICAL 187.1 KD PROTEIN IN OGG1-CNA2 INTERGENIC REGION>pir||S49802 probable membrane protein YML059c - yeast(Saccharomyces cerevisiae) >emb|CAA86716.1| (Z46729) unknown[Saccharomyces cerevisiae]
f1c05fs.r1	317	1.6e-26	39 476	gi|6321108 ref|NP_011186.1|YFL004W| Yfl004wp>sp|P43585|YFA4_YEASTHYPOTHETICAL 95.4 KD PROTEIN IN SEC4-

				MSH4 INTERGENIC REGION>pir||S56250 probable membrane protein YFL004w - yeast(Saccharomyces cerevisiae) >dbj|BAA09234.1| (D50617) YFL004W[Saccharomyces cerevisiae]>dbj|BAA08072.1| (D44604) unknown[Saccharomyces cerevisiae]
f3b11fs.r1	316	1.8e-26	19 387	gi|6322586 ref|NP_012660.1|YJR126C| Yjr126cp>sp|P47161|YJ96_YEASTHYPOTHETICAL 92.0 KD PROTEIN IN RPS5-ZMS1 INTERGENIC REGION>pir||S57149 probable membrane protein YJR126c - yeast(Saccharomyces cerevisiae) >emb|CAA89657.1| (Z49626) ORF YJR126c[Saccharomyces cerevisiae]
n3b09fs.r1	304	3.2e-26	43 546	gi|6322971 ref|NP_013043.1|YLL057C| Yll057cp>sp|Q12358|YL57_YEAST PUTATIVEDIOXYGENASE YLL057C >pir||S50963 hypothetical protein YLL057c -yeast (Saccharomyces cerevisiae) >emb|CAA88000.1| (Z47973)ORFL0572 [Saccharomyces cerevisiae] >emb|CAA97510.1| (Z73162) ORFYLL057c [Saccharomyces cerevisiae]
o1a07fs.r1	304	3.2e-26	35 508	sp|Q10138|YAS2_SCHPO HYPOTHETICAL 51.5 KD PROTEIN C3H8.02 IN CHROMOSOME I>pir||T38760 hypothetical protein SPAC3H8.02 - fission yeast(Schizosaccharomyces pombe) >emb|CAA93159.1| (Z69086)hypotheticalprotein [Schizosaccharomyces pombe]
o2b12fs.r1	302	5.6e-26	7 510	pir||T40424 hypothetical protein SPBC405.03c - fission yeast(Schizosaccharomyces pombe) >emb|CAB38602.1| (AL035655)hypothetical protein similar to yeast YML018C [Schizosaccharomycespombe]
Contig376	301	6.5e-26	54 500	gi|6321079 ref|NP_011157.1|YFL030W| Yfl030wp>sp|P43567|YFD0_YEASTHYPOTHETICAL 41.9 KD PROTEIN IN HAC1-CAK1 INTERGENIC REGION>pir||S56224 hypothetical protein YFL030w - yeast (Saccharomycescerevisiae) >dbj|BAA09208.1| (D50617) YFL030W [Saccharomycescerevisiae]
Contig181	300	8.6e-26	215 523	gi|6325087 ref|NP_015155.1|YPL170W| Ypl170wp>sp|Q12091|YP70_YEASTHYPOTHETICAL 16.8 KD PROTEIN IN MEX67-OYE3 INTERGENIC REGION>pir||S65181 hypothetical protein YPL170w - yeast (Saccharomycescerevisiae) >emb|CAA97876.1| (Z73526) ORF YPL170w [Saccharomycescerevisiae] >emb|CAA65551.1| (X96770) P2515 protein [Saccharomycescerevisiae]
j2c03fs.r1	309	1.2e-25	27 404	pir||T39973 hypothetical protein SPBC24C6.09c - fission yeast(Schizosaccharomyces pombe) >emb|CAA21153.1| (AL031786)hypothetical protein [Schizosaccharomyces pombe]

i1g12fs.r1	308	2.2e-25	7	453	pir||T04462 hypothetical protein F4D11.160 - Arabidopsis thaliana>emb|CAA18597.1| (AL022537) putative protein [Arabidopsis thaliana]
Contig903	299	5.7e-25	255	548	pir||T39513 hypothetical protein SPBC1604.01 - fission yeast(Schizosaccharomyces pombe) (fragment) >emb|CAA22334.1| (AL034433)hypothetical protein [Schizosaccharomyces pombe]
Contig596	294	5.7e-25	225	629	gi|6322584 ref|NP_012658.1|YJR124C| Yjr124cp >sp|P47159|YJ94_YEASTHYPOTHETICAL 49.7 KD PROTEIN IN RPS5-ZMS1 INTERGENIC REGION>pir||S57147 probable membrane protein YJR124c - yeast(Saccharomyces cerevisiae) >emb|CAA89655.1| (Z49624) ORF YJR124c[Saccharomyces cerevisiae]
Contig789	289	1.3e-24	153	572	pir||T40205 hypothetical protein SPBC31F10.02 - fission yeast(Schizosaccharomyces pombe) >emb|CAB10079.1| (Z97204)hypotheticalprotein [Schizosaccharomyces pombe]
Contig114	291	1.8e-24	190	513	sp|P77243|PRPD_ECOLI PRPD PROTEIN >pir||F64760 membrane protein prpD-Escherichia coli >gb|AAB18058.1| (U73857) similar to yqiP of B.subtilis [Escherichia coli] >gb|AAC73437.1| (AE000140) orf,hypothetical protein [Escherichia coli]
d3d03fs.r1	296	2.5e-24	57	488	gi|6319570 ref|NP_009652.1|YBR094W| Ybr094wp>sp|P38254|YBU4_YEASTHYPOTHETICAL 86.4 KD PROTEIN IN PHO5-VPS15 INTERGENIC REGION>pir||S48261 hypothetical protein YBR094w - yeast (Saccharomycescerevisiae) >emb|CAA55599.1| (X78993) hyp. protein [Saccharomycescerevisiae] >emb|CAA85047.1| (Z35963) ORF YBR094w [Saccharomycescerevisiae]
Contig568	282	6.6e-24	121	729	sp|O14359|YB4E_SCHPO HYPOTHETICAL 27.4 KD PROTEIN C30D10.14 IN CHROMOSOME II>pir||T40182 conserved hypothetical protein SPBC30D10.14 - fissionyeast (Schizosaccharomyces pombe) >emb|CAB10809.1| (Z97992)conserved hypothetical protein [Schizosaccharomyces pombe]
j2c01fs.r1	289	8.1e-24	138	416	emb|CAB63538.1| (AL133521) hypothetical protein [Schizosaccharomyces pombe]
Contig708	290	9e-24	192	539	gi|6324515 ref|NP_014584.1|YOL057W| Yol057wp >pir||S66749 hypothetical proteinYOL057w - yeast (Saccharomyces cerevisiae)>emb|CAA99066.1|(Z74799) ORF YOL057w [Saccharomyces cerevisiae]
e3h05fs.f1	282	1.5e-23	131	463	gi|6324771 ref|NP_014840.1|YOR197W| Yor197wp >pir||S67089 hypothetical proteinYOR197w - yeast (Saccharomyces

					cerevisiae)>emb|CAA99410.1|(Z75105) ORF YOR197w [Saccharomyces cerevisiae]
f3b01fs.f1	286	9.4e-23	304	558	pir||T38236 hypothetical protein SPAC23A1.17 - fission yeast(Schizosaccharomyces pombe) >emb|CAA16991.1| (AL021813)hypothetical protein; extensin-like; with SH3 Src homology domain[Schizosaccharomyces pombe]
g3e05fs.r1	271	1.1e-22	26	448	sp|O14351|YB45_SCHPO HYPOTHETICAL OXIDOREDUCTASE C30D10.05C IN CHROMOSOME II>pir||T40191 short chain dehydrogenase - fission yeast(Schizosaccharomyces pombe) >emb|CAB10800.1| (Z97992) short chaindehydrogenase [Schizosaccharomyces pombe]
Contig179	279	1.2e-22	7	426	gi|6322558 ref|NP_012632.1|YJR098C| Yjr098cp>sp|P47139|YJ68_YEASTHYPOTHETICAL 74.1 KD PROTEIN IN ACR1-YUH1 INTERGENIC REGION>pir||S57119 hypothetical protein YJR098c - yeast (Saccharomycescerevisiae) >emb|CAA89628.1| (Z49598) ORF YJR098c [Saccharomycescerevisiae]
s1g08fs.f1	272	2.5e-22	147	440	gi|6320310 ref|NP_010390.1|YDR105C| Ydr105cp >pir||S51256 probable membraneprotein YDR105c - yeast (Saccharomyces cerevisiae)>emb|CAA87681.1|(Z47746) unknown [Saccharomyces cerevisiae]>emb|CAA88659.1|(Z48758) unknown [Saccharomyces cerevisiae]
q2a04fs.r1	265	4.8e-22	1	156	sp|P03766|Y290_LAMBD HYPOTHETICAL NIN REGION PROTEIN ORF290>pir||QXBP5Lhypothetical nin region protein B-290 (nin region) - phage lambda>gb|AAA96588.1| (J02459) Nin 290 (pept unknown;290) [bacteriophagelambda]
Contig99	262	8.9e-22	128	490	pir||T40926 conserved hypothetical protein SPCC1281.07c - fission yeast(Schizosaccharomyces pombe) >emb|CAA22828.1| (AL035218) proteinwith Glutathione S transferase domain [Schizosaccharomyces pombe]
k3c11fs.f1	259	1.9e-21	385	615	gi|6322652 ref|NP_012725.1|YKT6| Ykt6p >sp|P36015|YKT6_YEAST HYPOTHETICAL 22.7KD PROTEIN IN PAS1-MST1 INTERGENIC REGION >pir||S38033 celldivision control protein SLY2 homolog YKL196c - yeast(Saccharomyces cerevisiae) >emb|CAA82040.1| (Z28196) ORF YKL196c[Saccharomyces cerevisiae]>gb|AAB32050.1| p26, Ykt6, YKL196cproduct=putative farnesylated VAMP homolog/v-SNARE component ofER-Golgi docking complex
Contig466	259	2.1e-21	73	456	sp|P45637|YPRA_CORGL HYPOTHETICAL 33.0 KD PROTEIN IN PROB-PROA INTERGENICREGION >gb|AAC44175.1| (U31230) unknown [Corynebacteriumglutamicum]

m3b12fs.f1	258	2.4e-21	40 465	pir||T19538 hypothetical protein K08H10.8 - Caenorhabditis elegans>emb|CAB02797.1| (Z81042) similar to Yeast hypothetical proteinYEY6 like; cDNA EST yk206h5.3 comes from this gene; cDNA ESTyk206h5.5 comes from this gene; cDNA EST yk303h1.3 comes from thisgene; cDNA EST yk303h1.5 comes from this gene; cDNA EST yk367e12.3comes from> >emb|CAB05549.1| (Z83113)similar to Yeast hypotheticalprotein YEY6
Contig629	258	6.6e-21	18 650	gi|6322068 ref|NP_012143.1|SIM1| (putative) invovled in control of DNAreplication; Sim1p >sp|P40472|SIM1_YEAST SIM1 PROTEIN>pir||S49886probable membrane protein YIL123w - yeast (Saccharomycescerevisiae) >emb|CAA86869.1| (Z46833) unknown [Saccharomycescerevisiae]
Contig736	252	1.1e-20	211 870	gb|AAD03446.1| (AF118223) T4B21.2 gene product [Arabidopsis thaliana]>emb|CAB80851.1| (AL161501) predicted protein of unknown function[Arabidopsis thaliana]
e4c12fs.r1	260	1.3e-20	78 530	pir||T18227 hypothetical protein - yeast (Candida albicans)>emb|CAA21985.1|(AL033501) CBS domain protein [Candida albicans]
Contig184	260	1.6e-20	142 399	sp|O13773|YE9C_SCHPO HYPOTHETICAL 79.5 KD PROTEIN C17A5.12 IN CHROMOSOME I>pir||T37827 hypothetical protein SPAC17A5.12 - fission yeast(Schizosaccharomyces pombe) >emb|CAB11512.1| (Z98849)hypotheticalprotein [Schizosaccharomyces pombe]
s4a10fs.f1	250	1.7e-20	263 580	pir||T39583 hypothetical protein SPBC16E9.09c - fission yeast(Schizosaccharomyces pombe) >emb|CAB16901.1| (Z99759)putativeinvolvement in vesicular protein trafficking via the ER[Schizosaccharomyces pombe]
t2f02fs.r1	255	2.5e-20	182 523	pir||T39542 hypothetical protein SPBC16A3.10 - fission yeast(Schizosaccharomyces pombe) >emb|CAA16861.1| (AL021748)hypothetical protein [Schizosaccharomyces pombe]
Contig717	248	2.9e-20	206 514	emb|CAB61587.1| (AL133210) hypothetical protein SCG11A.06c [Streptomycescoelicolor A3(2)]
n3f10fs.r1	244	1.4e-19	279 530	sp|Q09895|YAI8_SCHPO HYPOTHETICAL 43.7 KD PROTEIN C24B11.08C IN CHROMOSOME I>pir||S62553 hypothetical protein SPAC24B11.08c - fission yeast(Schizosaccharomyces pombe) >pir||T38335 conserved hypotheticalprotein SPAC24B11.08c - fission yeast (Schizosaccharomyces pombe)>emb|CAA91773.1| (Z67757) conserved hypothetical protein[Schizosaccharomyces pombe]
r3h08fs.r1	247	2.2e-19	232 447	pir||S52976 hypothetical protein 2 - Erwinia herbicola (fragment)

i4a04fs.r1	238	3.4e-19	104	442	gi|6321438 ref|NP_011515.1|YGR001C| Ygr001cp>sp|P53200|YG11_YEASTHYPOTHETICAL 22.2 KD PROTEIN IN ERP6-TFG2 INTERGENIC REGION>pir||S64290 hypothetical protein YGR001c - yeast (Saccharomycescerevisiae) >emb|CAA96984.1| (Z72786) ORF YGR001c [Saccharomycescerevisiae]
k3c11fs.r1	174	8.1e-19	25	210	pir||T40708 hypothetical protein SPBC839.01 - fission yeast(Schizosaccharomyces pombe) (fragment) >emb|CAA17004.1| (AL021816)hypothetical protein [Schizosaccharomyces pombe]>emb|CAB46694.1|(AL096796) hypothetical protein [Schizosaccharomyces pombe]
q2d10fs.r1	243	1e-18	10	438	gi|6320545 ref|NP_010625.1|YDR338C| Ydr338cp>sp|Q05497|YD38_YEASTHYPOTHETICAL 77.8 KD PROTEIN IN MRPS28-HXT7 INTERGENIC REGION>pir||S70103 probable membrane protein YDR338c - yeast(Saccharomyces cerevisiae) >gb|AAB64774.1| (U51032) Ydr338cp[Saccharomyces cerevisiae]
i2e09fs.f1	231	2.7e-18	257	451	gi|6324814 ref|NP_014883.1|YOR240W| Yor240wp >pir||S67133 probable membraneprotein YOR240w - yeast (Saccharomyces cerevisiae)>emb|CAA99461.1|(Z75147) ORF YOR240w [Saccharomyces cerevisiae]
d2h06fs.r1	227	4.5e-18	94	459	emb|CAA88787.1| (Z48952) unknown [Saccharomyces cerevisiae]
Contig627	228	1.3e-17	251	484	gi|6320119 ref|NP_010199.1|SUB2| RNA helicase; Sub2p>sp|Q07478|HE47_YEASTPROBABLE ATP-DEPENDENT RNA HELICASE P47 HOMOLOG>pir||S67620hypothetical protein YDL084w - yeast (Saccharomyces cerevisiae)>emb|CAA98650.1| (Z74132) ORF YDL084w [Saccharomyces cerevisiae]
k4f03fs.r1	150	1.4e-17	271	465	pir||T40219 hypothetical protein SPBC31F10.16 - fission yeast(Schizosaccharomyces pombe) >emb|CAB10092.1| (Z97204)hypotheticalprotein [Schizosaccharomyces pombe]
j3b02fs.r1	234	1.6e-17	17	520	sp|Q09778|YA8D_SCHPO HYPOTHETICAL 103.4 KD PROTEIN C22F3.13 IN CHROMOSOME I>pir||S62428 hypothetical protein SPAC22F3.13 - fission yeast(Schizosaccharomyces pombe) >pir||T38178 probable coiled coilprotein - fission yeast (Schizosaccharomyces pombe)>emb|CAA91078.1| (Z54285) putative coiled coil protein[Schizosaccharomyces pombe]
c4f04fs.f1	223	1.9e-17	150	455	gb|AAF13924.1| (AF192974) unknown [Schizosaccharomyces pombe]>emb|CAB69634.1|(AL137130) L-asparaginase precursor [Schizosaccharomyces pombe]

o3e09fs.r1	229	2.2e-17	15 341	pir||T40476 hypothetical protein SPBC4B4.04 - fission yeast(Schizosaccharomyces pombe) >emb|CAA19284.1| (AL023706)hypothetical protein. [Schizosaccharomyces pombe]
Contig102	221	2.2e-17	121 417	pir||G72387 conserved hypothetical protein - Thermotoga maritima (strain MSB8)>gb|AAD35454.1|AE001716_17 (AE001716) conserved hypotheticalprotein [Thermotoga maritima]
o2a08fs.r1	219	3.4e-17	87 497	gi|6320268 ref|NP_010348.1|YDR063W| Ydr063wp >pir||S54047 hypothetical proteinYDR063w - yeast (Saccharomyces cerevisiae)>emb|CAA58979.1|(X84162) unknown [Saccharomyces cerevisiae]>emb|CAA89092.1|(Z49209) unknown [Saccharomyces cerevisiae]>emb|CAA98881.1|(Z74359) ORF YDR063w [Saccharomyces cerevisiae]
p3g05fs.r1	222	4e-17	23 523	pir||H70548 hypothetical protein Rv0557 - Mycobacterium tuberculosis (strainH37RV) >emb|CAB08968.1| (Z95558) hypothetical protein Rv0557[Mycobacterium tuberculosis] >gb|AAC63250.1| (AF061562)alpha-D-mannose-alpha(1-6)phosphatidyl myo-inositol monomannosidetransferase [Mycobacterium tuberculosis]
j2c03fs.f1	228	5.5e-17	153 464	pir||T39973 hypothetical protein SPBC24C6.09c - fission yeast(Schizosaccharomyces pombe) >emb|CAA21153.1| (AL031786)hypothetical protein [Schizosaccharomyces pombe]
o1b08fs.f1	216	7.1e-17	261 557	gi|6320916 ref|NP_010995.1|NRF1| Homolog of S. pombe Nrf1 (97% identical inpredicted amino acid sequence), which was identified in a geneticscreen by its ability to reverse the Cdc42p suppression of acdc24-4ts mutant; Nrf1p >sp|P40046|YEQ2_YEAST HYPOTHETICAL 14.4 KDPROTEIN IN RNR1-ALD3 INTERGENIC REGION >pir||S50575 hypotheticalprotein YER072w - yeast (Saccharomyces cerevisiae) >gb|AAB64608.1|(U18813)
n3f06fs.f1	219	9e-17	191 469	sp|Q09895|YAI8_SCHPO HYPOTHETICAL 43.7 KD PROTEIN C24B11.08C IN CHROMOSOME I>pir||S62553 hypothetical protein SPAC24B11.08c - fission yeast(Schizosaccharomyces pombe) >pir||T38335 conserved hypotheticalprotein SPAC24B11.08c - fission yeast (Schizosaccharomyces pombe)>emb|CAA91773.1| (Z67757) conserved hypothetical protein[Schizosaccharomyces pombe]
Contig379	215	1.6e-16	293 646	pir||T05349 hypothetical protein F8B4.90 - Arabidopsis thaliana>emb|CAA22566.1| (AL034567) putative protein [Arabidopsis thaliana]>emb|CAB79956.1| (AL161581) putative protein [Arabidopsis thaliana]

s1c07fs.r1	211	2.3e-16	369 611	pir||T40210 hypothetical protein SPBC31F10.07 - fission yeast(Schizosaccharomyces pombe) >emb|CAB10084.1| (Z97204) hypotheticalprotein [Schizosaccharomyces pombe]
f3g03fs.f1	208	5.3e-16	299 520	pir||B75432 conserved hypothetical protein - Deinococcus radiodurans (strainR1) >gb|AAF10724.1|AE001964_8 (AE001964) conserved hypotheticalprotein [Deinococcus radiodurans]
a1b10fs.r1	207	6.4e-16	27 440	emb|CAB83178.1| (AL162692) hypothetical protein [Schizosaccharomyces pombe]
Contig325	206	8.2e-16	444 788	gb|AAD34553.1| (AF141925) unknown [Aspergillus terreus]
Contig88	204	1.3e-15	336 527	gi|6324626 ref|NP_014695.1|YOR052C| Yor052cp >pir||S66926 hypothetical proteinYOR052c - yeast (Saccharomyces cerevisiae)>emb|CAA99244.1|(Z74960) ORF YOR052c [Saccharomyces cerevisiae]>emb|CAA94537.1|(Z70678) YOR29-03 [Saccharomyces cerevisiae]
e2e10fs.r1	204	1.3e-15	233 478	gi|6322393 ref|NP_012467.1|YJL068C| Yjl068cp>sp|P40363|YJG8_YEASTHYPOTHETICAL 33.9 KD ESTERASE IN SMC3-MRPL8 INTERGENIC REGION>pir||S50803 hypothetical protein YJL068c - yeast (Saccharomycescerevisiae) >emb|CAA84054.1| (Z34288) 39.8% identity in 289 aaoverlap with human esterase D [Saccharomyces cerevisiae]>emb|CAA89359.1| (Z49343) ORF YJL068c [Saccharomyces cerevisiae]
t4f03fs.f1	204	1.4e-15	257 550	pir||T26112 hypothetical protein W02D9.2 - Caenorhabditis elegans>emb|CAB03470.1| (Z81137) Similarity to Yeast YIP1 protein(SW:P53039); cDNA EST EMBL:T01608 comes from this gene; cDNA ESTEMBL:C07393 comes from this gene; cDNA EST EMBL:C07550 comes fromthis gene; cDNA EST EMBL:C08746 comes from this gene; cDNA ESTEMBL:C08986 come>
g2c08fs.r1	213	1.5e-15	94 441	pir||T32708 hypothetical protein T22D1.4 - Caenorhabditis elegans>gb|AAB94272.1| (AF039052) similar to ribophorin I [Caenorhabditiselegans]
l4d04fs.r1	203	1.7e-15	87 503	gi|6321023 ref|NP_011102.1|YER175C| Yer175cp>sp|P32643|YE05_YEASTHYPOTHETICAL 34.8 KD PROTEIN IN RAD24-BMH1 INTERGENIC REGION>pir||S30861 hypothetical protein YER175c - yeast (Saccharomycescerevisiae) >gb|AAB64702.1| (U18922) Yer175cp [Saccharomycescerevisiae]
a1e11fs.r1	208	2.5e-15	19 417	pir||T33865 hypothetical protein H04M03.4 - Caenorhabditis elegans>gb|AAD12787.1| (AF125442) H04M03.4 gene product [Caenorhabditiselegans]

s1f06fs.f1	200	3.4e-15	234	518	sp|P77526|YFCG_ECOLI HYPOTHETICAL 24.5 KD PROTEIN IN PTA-FOLX INTERGENICREGION >pir||D65002 hypothetical protein b2302 - Escherichia coli(strain K-12) >gb|AAC75362.1| (AE000319) putative S-transferase[Escherichia coli] >dbj|BAA16139.1| (D90861) URE2 PROTEIN.[Escherichia coli] >dbj|BAA16148.1| (D90862) URE2 PROTEIN.[Escherichia coli]
Contig787	208	4e-15	460	633	pir||T26079 hypothetical protein W02A2.5 - Caenorhabditis elegans>emb|CAB05308.1| (Z82286) predicted using Genefinder[Caenorhabditis elegans]
g3e05fs.f1	199	4.5e-15	180	536	gi|6322227 ref|NP_012302.1|YIR036C| Yir036cp >sp|P40580|YIV6_YEASTHYPOTHETICAL OXIDOREDUCTASE IN LYS1-HYR1 INTERGENIC REGION>pir||S48498 oxidoreductase homolog YIR036c - yeast (Saccharomycescerevisiae) >emb|CAA86196.1| (Z38061) orf, len: 263, CAI:0.12,similar to orf (32877..33641) and BA71_EUBSP P07914 7-ALPHA-HYDROXYSTEROID DEHYDROGENASE [Saccharomyces cerevisiae]
n3b05fs.f1	204	5.3e-15	249	542	gi|6322971 ref|NP_013043.1|YLL057C| Yll057cp >sp|Q12358|YL57_YEAST PUTATIVEDIOXYGENASE YLL057C >pir||S50963 hypothetical protein YLL057c -yeast (Saccharomyces cerevisiae) >emb|CAA88000.1| (Z47973)ORFL0572 [Saccharomyces cerevisiae] >emb|CAA97510.1| (Z73162) ORFYLL057c [Saccharomyces cerevisiae]
n2b06fs.r1	208	6.1e-15	169	486	gi|6320812 ref|NP_010891.1|YEL023C| Yel023cp>sp|P39992|YEC3_YEASTHYPOTHETICAL 78.3 KD PROTEIN IN RIP1-GEA2 INTERGENIC REGION>pir||S50436 hypothetical protein YEL023c - yeast (Saccharomycescerevisiae) >gb|AAB64500.1| (U18530) Yel023cp [Saccharomycescerevisiae]
c3h11fs.f1	201	6.4e-15	309	449	gi|6322924 ref|NP_012997.1|YKR071C| Ykr071cp>sp|P36152|YK51_YEASTHYPOTHETICAL 38.5 KD PROTEIN IN MET1-SIS2 INTERGENIC REGION>pir||S38148 hypothetical protein YKR071c - yeast (Saccharomycescerevisiae) >emb|CAA82150.1| (Z28296) ORF YKR071c [Saccharomycescerevisiae]
d4g11fs.r1	204	8.2e-15	143	451	gi|6323877 ref|NP_013948.1|YMR221C| Ymr221cp>sp|Q04991|YM68_YEASTHYPOTHETICAL 56.2 KD PROTEIN IN ERG8-UBP8 INTERGENIC REGION>pir||S57589 probable membrane protein YMR221c - yeast(Saccharomyces cerevisiae) >emb|CAA90192.1| (Z49939) unknown[Saccharomyces cerevisiae]
l4a08fs.r1	211	9.4e-15	214	522	gi|6322598 ref|NP_012672.1|IML1| Iml1p >sp|P47170|YJ9G_YEAST HYPOTHETICAL182.0 KD PROTEIN IN NMD5-HOM6 INTERGENIC REGION >pir||S57161hypothetical protein YJR138w - yeast (Saccharomyces

				cerevisiae)>emb|CAA89670.1| (Z49638) ORF YJR138w [Saccharomyces cerevisiae]
p4d01fs.f1	196	1.1e-14	122 511	emb|CAB77115.1| (AX001421) unnamed protein product [unidentified]
g1g01fs.f1	195	1.1e-14	292 483	gb|AAC44550.1| (U34346) unknown [Paracoccus denitrificans]
Contig274	194	1.5e-14	236 652	pir||A70672 hypothetical protein Rv2972c - Mycobacterium tuberculosis (strainH37RV) >emb|CAB05439.1| (Z83018) hypothetical protein Rv2972c[Mycobacterium tuberculosis]
Contig778	193	1.7e-14	333 596	pir||T38915 hypothetical protein SPAC56F8.05c - fission yeast(Schizosaccharomyces pombe) >emb|CAA93576.2| (Z69728) hypotheticalprotein [Schizosaccharomyces pombe]
m2f08fs.r1	192	2.1e-14	192 473	pir||T11655 hypothetical protein SPAC6F12.04 - fission yeast(Schizosaccharomyces pombe) >emb|CAB11088.1| (Z98533) hypotheticalprotein [Schizosaccharomyces pombe]
p4c11fs.f1	189	4.5e-14	311 508	emb|CAA81838.1| (Z28005) ORF YKL006w [Saccharomyces cerevisiae]
j3d02fs.r1	198	5.3e-14	208 546	pir||T40996 conserved hypothetical protein SPCC1450.14c - fission yeast(Schizosaccharomyces pombe) >emb|CAB40181.1| (AL049559) conservedhypothetical protein [Schizosaccharomyces pombe]
Contig540	206	1.7e-13	167 880	gb|AAD29633.1|AF1168 (AF116827) unknown [Homo sapiens]
n4d08fs.f1	197	1.7e-13	225 653	pir||T32814 hypothetical protein H17B01.4 - Caenorhabditis elegans>gb|AAB94988.1| (AF040646) H17B01.4 gene product [Caenorhabditiselegans]
s3h06fs.r1	184	1.7e-13	123 275	pir||T14082 hypothetical protein F9N11.70 - Arabidopsis thaliana>emb|CAB52466.1| (AL109796) snRNP Sm protein F-like [Arabidopsisthaliana] >emb|CAB81015.1| (AL161576) snRNP Sm protein F-like[Arabidopsis thaliana]
Contig60	191	2.2e-13	116 469	gb|AAD34558.1| (AF141925) unknown [Aspergillus terreus]
Contig516	181	3.5e-13	190 483	gi|6323716 ref|NP_013787.1|YMR071C| Ymr071cp >sp|Q04767|YMW1_YEASTHYPOTHETICAL 18.7 KD PROTEIN IN HMS1-ABF2 INTERGENIC REGION>pir||S52831 probable membrane protein YMR071c - yeast(Saccharomyces cerevisiae) >emb|CAA88796.1| (Z48952) unknown[Saccharomyces cerevisiae]
s4a03fs.f1	186	3.7e-13	276 485	dbj|BAA91255.1| (AK000562) unnamed protein product [Homo sapiens]
a1g12fs.r1	181	3.7e-13	48 428	gb|AAF30968.1|AE0021 (AE002154) conserved hypothetical [Ureaplasmaurealyticum]
d4h06fs.r1	185	7.6e-13	46 489	gi|6324684 ref|NP_014753.1|TFC7| 55 kDa subunit of TFIIIC (tau55); Tfc7p>pir||S61668 hypothetical protein YOR110w - yeast (Saccharomycescerevisiae) >emb|CAA64030.1| (X94335) YOR3234w

				[Saccharomycescerevisiae] >emb|CAA99308.1| (Z75018) ORF YOR110w [Saccharomycescerevisiae]
t2h03fs.f1	179	1.9e-12	317 529	pir||T40385 hypothetical protein SPBC3E7.11c - fission yeast(Schizosaccharomyces pombe) >emb|CAA19014.1| (AL023534) DNA Jdomain protein [Schizosaccharomyces pombe]
l1g08fs.r1	180	2.5e-12	12 536	sp|Q10173|YAV4_SCHPO HYPOTHETICAL 51.9 KD PROTEIN C27F1.04C IN CHROMOSOME I>pir||T38462 nuf2-like coiled-coil protein - fission yeast(Schizosaccharomyces pombe) >emb|CAA93293.1| (Z69368) nuf2-likecoiled-coil protein [Schizosaccharomyces pombe]
Contig783	174	2.5e-12	250 669	gi|6321674 ref|NP_011751.1|YGR235C| Ygr235cp>sp|P50087|YG4Y_YEASTHYPOTHETICAL 26.9 KD PROTEIN IN YHB1-PFK1 INTERGENIC REGION>pir||S57700 probable membrane protein YGR235c - yeast(Saccharomyces cerevisiae) >emb|CAA61185.1| (X87941) ORF 233[Saccharomyces cerevisiae]>emb|CAA97263.1| (Z73020) ORF YGR235c[Saccharomyces cerevisiae]
Contig808	184	3e-12	306 506	pir||S76896 hypothetical protein - Synechocystis sp. (strain PCC 6803)>dbj|BAA18808.1| (D90917) hypothetical protein [Synechocystis sp.]
c3h08fs.f1	181	3.3e-12	242 460	emb|CAA57291.1| (X81635) unnamed protein product [Saccharomyces cerevisiae]
r4g12fs.f1	176	3.8e-12	110 394	pir||T35215 hypothetical protein SC5C7.08 SC5C7.08 - Streptomyces coelicolor>emb|CAA20620.1| (AL031515) hypothetical protein SC5C7.08[Streptomyces coelicolor A3(2)]
g2h01fs.f1	179	4.7e-12	234 425	pir||T40649 hypothetical protein SPBC6B1.08c - fission yeast(Schizosaccharomyces pombe) >dbj|BAA12034.1| (D83659) homology toSaccharomyces cerevisiae hypothetical protein YER049W[Schizosaccharomyces pombe] >emb|CAA17051.1| (AL021838)hypothetical protein [Schizosaccharomyces pombe]
s2g05fs.r1	177	5e-12	25 315	gi|6320901 ref|NP_010980.1|PCL6| PHO85 cyclin; Pcl6p>sp|P40038|YEO9_YEASTHYPOTHETICAL 47.0 KD PROTEIN IN PET117-CEM1 INTERGENIC REGION>pir||S50562 hypothetical protein YER059w - yeast (Saccharomycescerevisiae) >gb|AAB64595.1| (U18813) Yer059wp [Saccharomycescerevisiae]
Contig867	170	5.4e-12	369 629	pir||E72312 conserved hypothetical protein - Thermotoga maritima (strain MSB8)>gb|AAD36047.1|AE001759_11 (AE001759) conserved hypotheticalprotein [Thermotoga maritima]
q2e01fs.f1	181	7.8e-12	225 455	pir||T40715 hypothetical protein SPBC839.08c - fission yeast(Schizosaccharomyces pombe) >emb|CAA17011.1|

					(AL021816)hypothetical protein [Schizosaccharomyces pombe] >emb|CAB46701.1|(AL096796) hypothetical protein [Schizosaccharomyces pombe]
e3f02fs.f1	168	8.4e-12	179	514	dbj|BAA91215.1| (AK000508) unnamed protein product [Homo sapiens]
h3g06fs.r1	174	1e-11	29	307	gi|6320098 ref|NP_010178.1|QRI2| Qri2p >sp|P43124|QRI2_YEAST HYPOTHETICAL 46.1KD PROTEIN IN PHO2-POL3 INTERGENIC REGION >pir||S50739 QRI2 protein- yeast (Saccharomyces cerevisiae) >emb|CAA55925.1| (X79380)QRI2 [Saccharomyces cerevisiae] >emb|CAA64908.1| (X95644) ORF 2354 [Saccharomyces cerevisiae] >emb|CAA98672.1| (Z74153) ORF YDL105w [Saccharomyces cerevisiae]
o4b07fs.r1	167	1.1e-11	214	495	pir||T15586 hypothetical protein C24A3.4 - Caenorhabditis elegans>gb|AAA81457.1| (U40424) similar to Eubacterium sp bileacid-inducible operon protein F (SP:BAIF_EUBSP, P19413) and E. coliL-carnitine dehydratase (SP:CAIB_ECOLI, P31572) [Caenorhabditiselegans]
Contig260	166	1.3e-11	317	439	gb|AAB92222.1| (U94183) unknown [Colletotrichum gloeosporioides]
Contig931	167	3.7e-11	279	533	gi|6322565 ref|NP_012639.1|YJR105W| Yjr105wp >sp|P47143|ADK_YEAST PUTATIVEADENOSINE KINASE >pir||S57126 ribokinase homolog YJR105w - yeast(Saccharomyces cerevisiae) >emb|CAA89635.1| (Z49605) ORF YJR105w [Saccharomyces cerevisiae]
Contig889	165	4.5e-11	169	879	pir||H72277 hypothetical protein TM1254 - Thermotoga maritima (strain MSB8)>gb|AAD36329.1|AE001780_13 (AE001780) beta-phosphoglucomutase,putative [Thermotoga maritima]
s3g02fs.r1	161	4.8e-11	88	495	pir||T40346 hypothetical protein SPBC3B9.07c - fission yeast(Schizosaccharomyces pombe) >emb|CAA17787.1| (AL022070)dna-dependent rna polymerase polypeptide [Schizosaccharomycespombe]>gb|AAF40067.1|AF129106_1 (AF129106) RNA polymerase Isubunit Rpa43 [Schizosaccharomyces pombe]
j3a07fs.r1	169	1.2e-10	29	493	pir||T38528 conserved hypothetical protein SPAC2C4.17c - fission yeast(Schizosaccharomyces pombe) >emb|CAB16377.1| (Z99259)conservedhypothetical protein [Schizosaccharomyces pombe]
s2f03fs.r1	157	1.2e-10	185	445	pdb|1YAC|A Chain A, The 1.8 Angstrom Crystal Structure Of The Ycac GeneProduct From Escherichia Coli Reveals An Octameric Hydrolase OfUnknown Specificity >pdb|1YAC|B Chain B, The 1.8 Angstrom CrystalStructure Of The Ycac Gene Product From Escherichia Coli Reveals AnOctameric Hydrolase Of Unknown Specificity
Contig215	161	1.9e-10	292	486	gi|6324623 ref|NP_014692.1|YOR049C| Yor049cp >pir||S66923 hypothetical proteinYOR049c - yeast (Saccharomyces

					cerevisiae)>emb|CAA99241.1|(Z74957) ORF YOR049c [Saccharomyces cerevisiae]
Contig317	161	1.9e-10	197	553	pir||S76674 hypothetical protein - Synechocystis sp. (strain PCC 6803)>dbj|BAA10618.1| (D64004) hypothetical protein [Synechocystis sp.]
i4b06fs.r1	162	3.8e-10	68	484	pir||T04946 hypothetical protein F7J7.90 - Arabidopsis thaliana>emb|CAA17534.1| (AL021960) putative protein [Arabidopsis thaliana]>emb|CAB79115.1| (AL161554) putative protein [Arabidopsis thaliana]
g3g02fs.f1	151	5.3e-10	316	534	pir||T41359 hypothetical protein SPCC4G3.17 - fission yeast(Schizosaccharomyces pombe) >emb|CAB09764.1| (Z97052)hypotheticalprotein [Schizosaccharomyces pombe]
d4c07fs.r1	158	9.7e-10	15	356	pir||T40996 conserved hypothetical protein SPCC1450.14c - fission yeast(Schizosaccharomyces pombe) >emb|CAB40181.1| (AL049559)conservedhypothetical protein [Schizosaccharomyces pombe]
m2g08fs.r1	163	1.1e-09	1	375	gb|AAD27577.1|AF1141 (AF114171) hypothetical protein [Sorghum bicolor]
e1a10fs.f1	160	2.6e-09	126	518	pir||F71414 hypothetical protein - Arabidopsis thaliana>emb|CAB10288.1|(Z97337) hypothetical protein [Arabidopsis thaliana]>emb|CAB78551.1| (AL161540) hypothetical protein [Arabidopsisthaliana]
Contig343	153	2.9e-09	272	505	gi|6320765 ref|NP_010844.1|YEL070W| Yel070wp>gi|6324401ref|NP_014471.1|YNR073C| Ynr073cp>sp|P39941|YEI0_YEASTHYPOTHETICAL 56.5 KD PROTEIN IN HXT8 5'REGION AND IN PAU6 5'REGION>pir||S50519 hypothetical protein YEL070w - yeast (Saccharomycescerevisiae) >gb|AAB65017.1| (U18795) Yel070wp [Saccharomycescerevisiae] >emb|CAA96356.1| (Z71688) ORF YNR073c [Saccharomycescerevisiae]
Contig40	142	5e-09	299	538	emb|CAB63714.1| (AL133560) hypothetical protein [Homo sapiens]
n2h10fs.r1	143	6e-09	357	497	emb|CAB61455.1| (AL133156) hypothetical protein [Schizosaccharomyces pombe]
b3f12fs.f1	146	6.5e-09	259	441	pir||T37851 hypothetical protein SPAC17G6.19c - fission yeast(Schizosaccharomyces pombe) >emb|CAB16230.1| (Z99162)hypotheticalprotein [Schizosaccharomyces pombe] >emb|CAB77009.1| (AL159951)conserved hypothetical protein [Schizosaccharomyces pombe]
Contig70	156	6.8e-09	5	526	gi|6322421 ref|NP_012495.1|NUP192| Nup192p >sp|P47054|YJD9_YEAST HYPOTHETICAL191.5 KD PROTEIN IN NSP1-KAR2 INTERGENIC REGION

					>pir||S56811probable membrane protein YJL039c - yeast (Saccharomycescerevisiae) >emb|CAA89330.1| (Z49314) ORF YJL039c [Saccharomycescerevisiae]
Contig586	145	1.4e-08	253	501	gb|AAA86839.1| (U44916) ORF; unknown function [Trichomonas vaginalis]
o3e09fs.f1	145	1.4e-08	367	519	emb|CAA97056.1| (Z72840) ORF YGR054w [Saccharomyces cerevisiae]
c1d05fs.f1	149	4.5e-08	10	387	gb|AAF26098.1|AC0123 (AC012328) unknown protein [Arabidopsis thaliana]
i3d01fs.r1	133	4.6e-08	31	207	gi|6322504 ref|NP_012578.1|YJR044C| Yjr044cp>sp|P47111|YJ14_YEASTHYPOTHETICAL 15.7 KD PROTEIN IN NUP85-SSC1 INTERGENIC REGION>pir||S57063 probable membrane protein YJR044c - yeast(Saccharomyces cerevisiae) >emb|CAA89572.1| (Z49544) ORF YJR044c[Saccharomyces cerevisiae]>gb|AAA88746.1| (L36344) ORF; putative[Saccharomyces cerevisiae]
Contig158	137	8e-08	354	512	gi|6322923 ref|NP_012996.1|YKR070W| Ykr070wp>sp|P36151|YK50_YEASTHYPOTHETICAL 39.4 KD PROTEIN IN MET1-SIS2 INTERGENIC REGION>pir||S38147 hypothetical protein YKR070w - yeast (Saccharomycescerevisiae) >emb|CAA82149.1| (Z28295) ORF YKR070w [Saccharomycescerevisiae]
j1d05fs.r1	134	8.5e-08	49	459	dbj|BAA90956.1| (AK000120) unnamed protein product [Homo sapiens]
Contig725	137	9.5e-08	347	508	pir||E70047 conserved hypothetical protein yvrK - Bacillus subtilis>emb|CAB15314.1| (Z99120) similar to hypothetical proteins[Bacillus subtilis] >emb|CAB15329.1| (Z99121) similar tohypothetical proteins [Bacillus subtilis] >emb|CAA11727.1|(AJ223978) YvrK protein [Bacillus subtilis]
i2g06fs.r1	136	9.8e-08	212	430	pir||B70678 hypothetical protein Rv3553 - Mycobacterium tuberculosis (strainH37RV) >emb|CAB05054.1| (Z82098) hypothetical protein Rv3553[Mycobacterium tuberculosis]
l1d10fs.f1	138	1.3e-07	280	519	gi|6324882 ref|NP_014951.1|YOR306C| Yor306cp >pir||S67210 probable membraneprotein YOR306c - yeast (Saccharomyces cerevisiae)>emb|CAA99626.1|(Z75214) ORF YOR306c [Saccharomyces cerevisiae]
j3c07fs.f1	137	1.4e-07	300	449	gi|6322150 ref|NP_012225.1|YIL039W| Yil039wp>sp|P40533|YID9_YEASTHYPOTHETICAL 54.9 KD PROTEIN IN CBR5-NOT3 INTERGENIC REGION>pir||S49939 probable membrane protein YIL039w - yeast(Saccharomyces cerevisiae) >emb|CAA86912.1| (Z46861) unknown[Saccharomyces cerevisiae]
k2c04fs.r1	136	1.9e-07	392	496	emb|CAA96769.1| (Z72587) ORF YGL066w [Saccharomyces cerevisiae]

Contig240	126	2.3e-07	253	366	gi|6324172 ref|NP_014242.1|YNL157W| Ynl157wp>sp|P53897|YNP7_YEASTHYPOTHETICAL 18.1 KD PROTEIN IN YGP1-YCK2 INTERGENIC REGION>pir||S60970 hypothetical protein YNL157w - yeast (Saccharomycescerevisiae) >emb|CAA63282.1| (X92517) N1743 [Saccharomycescerevisiae] >emb|CAA96044.1| (Z71433) ORF YNL157w [Saccharomycescerevisiae]
p4b12fs.f1	133	2.6e-07	185	511	pir||T41456 probable phosphoslipase SPCC5E4.05c - fission yeast(Schizosaccharomyces pombe) >emb|CAA21960.1| (AL033406)hypothetical protein, possible phosphoslipase [Schizosaccharomycespombe]
Contig585	131	4e-07	351	506	gi|6322923 ref|NP_012996.1|YKR070W| Ykr070wp>sp|P36151|YK50_YEASTHYPOTHETICAL 39.4 KD PROTEIN IN MET1-SIS2 INTERGENIC REGION>pir||S38147 hypothetical protein YKR070w - yeast (Saccharomycescerevisiae) >emb|CAA82149.1| (Z28295) ORF YKR070w [Saccharomycescerevisiae]
Contig398	134	5e-07	351	452	gi|6324320 ref|NP_014390.1|YNL008C| Ynl008cp >sp|P53983|YNA8_YEASTHYPOTHETICAL 76.7 KD PROTEIN IN SPO1-SIS1 INTERGENIC REGION>pir||S62919 probable membrane protein YNL008c - yeast(Saccharomyces cerevisiae) >emb|CAA95868.1| (Z71284) ORF YNL008c[Saccharomyces cerevisiae]
b2h06fs.f1	125	7.8e-07	310	417	pir||D70635 hypothetical protein Rv1928c - Mycobacterium tuberculosis (strainH37RV) >emb|CAB06498.1| (Z84498) hypothetical protein Rv1928c[Mycobacterium tuberculosis]
l3b09fs.r1	127	9.7e-07	84	422	gi|6321889 ref|NP_011965.1|YHR097C| Yhr097cp>sp|P38809|YHP7_YEASTHYPOTHETICAL 40.7 KD PROTEIN IN HXT5-NRK1 INTERGENIC REGION>pir||S46727 hypothetical protein YHR097c - yeast (Saccharomycescerevisiae) >gb|AAB68935.1| (U00060) Yhr097cp [Saccharomycescerevisiae]
f3h07fs.f1	131	1.5e-06	263	478	pir||T40417 hypothetical protein SPBC4.03c - fission yeast(Schizosaccharomyces pombe) >emb|CAB58402.1| (AL121863)hypothetical protein [Schizosaccharomyces pombe]
Contig142	118	1.7e-06	245	541	gi|6325320 ref|NP_015388.1|YPR063C| Ypr063cp >pir||S54084 probable membraneprotein YPR063c - yeast (Saccharomyces cerevisiae)>emb|CAA89180.1|(Z49219) unknown [Saccharomyces cerevisiae]>emb|CAA95007.1|(Z71255) unknown [Saccharomyces cerevisiae]
h3h04fs.r1	125	2.6e-06	363	530	pir||T22632 hypothetical protein F54C9.6 - Caenorhabditis elegans>emb|CAA90252.1| (Z49967) ATP-binding protein

				(CDC48/PAS1/SEC18family) with strong similarity to the yeast BCS1 protein (SwissProt accession number P32839); cDNA EST EMBL:C07371 comes from thisgene; cDNA EST EMBL:C08716 comes from this gene [Caenorhabditiselegans]
j1a01fs.f1	120	2.9e-06	111 290	sp|Q23202|YOM4_CAEEL HYPOTHETICAL 29.1 KD PROTEIN W06E11.4 IN CHROMOSOME III>gb|AAA62303.1| (U20862) W06E11.4 gene product [Caenorhabditiselegans]
a2h03fs.r1	124	4.1e-06	220 441	gb|AAD34558.1| (AF141925) unknown [Aspergillus terreus]
b4e06fs.r1	118	4.1e-06	226 456	gb|AAB71479.1| (AC002294) Unknown protein [Arabidopsis thaliana]
g3e02fs.f1	120	6.7e-06	423 545	pir||T04917 hypothetical protein T10I14.190 - Arabidopsis thaliana>emb|CAA16786.1| (AL021712) putative protein [Arabidopsis thaliana]>emb|CAB79191.1| (AL161557) putative protein [Arabidopsis thaliana]
Contig193	132	8.6e-06	136 567	gi|6321548 ref|NP_011625.1|YGR110W| Ygr110wp>sp|P53264|YG2V_YEASTHYPOTHETICAL 52.0 KD PROTEIN IN CLB6-SHY1 INTERGENIC REGION>pir||S64418 hypothetical protein YGR110w - yeast (Saccharomycescerevisiae) >emb|CAA97118.1| (Z72895) ORF YGR110w [Saccharomycescerevisiae]
j2c01fs.f1	121	1.1e-05	140 439	emb|CAB63538.1| (AL133521) hypothetical protein [Schizosaccharomyces pombe]
l1b03fs.r1	125	1.4e-05	236 445	gb|AAB95634.1| (AC003982) unknown function; 60% similar to Z50177(PID:g927403) (PID:g927402) [Homo sapiens]
Contig241	113	2.2e-05	132 449	pir||T37851 hypothetical protein SPAC17G6.19c - fission yeast(Schizosaccharomyces pombe) >emb|CAB16230.1| (Z99162)hypotheticalprotein [Schizosaccharomyces pombe] >emb|CAB77009.1| (AL159951)conserved hypothetical protein [Schizosaccharomyces pombe]
j3b07fs.f1	114	4.8e-05	332 490	emb|CAB58086.1| (AL121806)/prediction=(method:""genefinder"",version:""084"", score:""54.56"");/prediction=(method:""genscan"",version:""1.0"", score:""105.80"");/match=(desc:""LP08751.5primeLP Drosophila melanogaster larval-early pupal pOT2 Drosophilamelanogas> >gb|AAF45595.1| (AE003420) CG11382 gene product[Drosophila melanogaster]
j1b04fs.f1	110	0.0001	289 477	pir||T16025 hypothetical protein F10D7.2 - Caenorhabditis elegans>gb|AAA81720.1| (U40945) coded for by C. elegans cDNA yk74b9.3;coded for by C. elegans cDNA yk74b9.5; similar to repeat ofcalcium channel alpha subunits; similar to tetracycline

					resistanceprotein; similar to hypothetical protein in HSP30-PMP1 region(SP:YCR3_YEAST>
g4g01fs.r1	100	0.00012	65	328	pir||T39576 hypothetical protein SPBC16E9.01c - fission yeast(Schizosaccharomyces pombe) (fragment) >emb|CAB16894.1| (Z99759)hypothetical protein [Schizosaccharomyces pombe]
Contig600	102	0.00014	371	463	gi|6321102 ref|NP_011180.1|YFL010C| Yfl010cp>sp|P43582|YFB0_YEASTHYPOTHETICAL 22.7 KD PROTEIN IN AUA1-CDC4 INTERGENIC REGION>pir||S48311 probable membrane protein YFL010c - yeast(Saccharomyces cerevisiae) >emb|CAA86342.1| (Z46255) orf, len: 211,CAI: 0.16 [Saccharomyces cerevisiae] >dbj|BAA09228.1| (D50617)YFL010C [Saccharomyces cerevisiae]>dbj|BAA06494.1| (D31600)unknown [Saccharomyces cerevisiae]
s3e12fs.r1	118	0.00021	19	222	gb|AAA18516.1| (L19300) ORF3 [Staphylococcus aureus]
d2c05fs.f1	107	0.00022	337	441	pir||T41172 hypothetical protein SPCC1840.04 - fission yeast(Schizosaccharomyces pombe) >emb|CAA20127.1| (AL031179)hypothetical protein [Schizosaccharomyces pombe]
Contig223	117	0.00034	44	535	pir||T19897 hypothetical protein C41G7.6 - Caenorhabditis elegans>emb|CAB02844.1| (Z81048) cDNA EST EMBL:C13282 comes from thisgene; cDNA EST yk232c5.3 comes from this gene; cDNA EST yk232c5.5comes from this gene; cDNA EST yk364d3.3 comes from this gene; cDNAEST yk386a7.3 comes from this gene; cDNA EST yk386a7.5 comes fromthis g>
t2g05fs.f1	109	0.00037	102	308	pir||T06297 hypothetical protein T9E8.140 - Arabidopsis thaliana>emb|CAB40775.1| (AL049608) putative protein [Arabidopsis thaliana]>emb|CAB78382.1| (AL161536) putative protein [Arabidopsis thaliana]
i3g12fs.r1	103	0.00043	85	168	pir||T16755 hypothetical protein R144.4 - Caenorhabditiselegans>gb|AAC46548.1| (U23515) weak similarity to adenylylcyclase-associated protein (CAP) and to P. chabaudi adami majormerozoite surfae antigen protein (PIR:A32555). Final exon overlapsgene predicted on other strand. [Caenorhabditis elegans]
s2e11fs.f1	120	0.00074	405	632	gi|6321906 ref|NP_011982.1|YHR114W| SH3 domain; Yhr114wp>sp|P38822|YHR4_YEASTHYPOTHETICAL 71.2 KD PROTEIN IN ERP5-ORC6 INTERGENIC REGION>pir||S48956 hypothetical protein YHR114w - yeast (Saccharomycescerevisiae) >gb|AAB68850.1| (U00059) Yhr114wp [Saccharomycescerevisiae]
Contig401	112	0.00082	350	508	gi|6322584 ref|NP_012658.1|YJR124C| Yjr124cp>sp|P47159|YJ94_YEASTHYPOTHETICAL 49.7 KD PROTEIN IN RPS5-

				ZMS1 INTERGENIC REGION>pir||S57147 probable membrane protein YJR124c - yeast(Saccharomyces cerevisiae) >emb|CAA89655.1| (Z49624) ORF YJR124c[Saccharomyces cerevisiae]
Contig698	113	0.0011	104 469	gi|6324623 ref|NP_014692.1|YOR049C| Yor049cp >pir||S66923 hypothetical proteinYOR049c - yeast (Saccharomyces cerevisiae)>emb|CAA99241.1|(Z74957) ORF YOR049c [Saccharomyces cerevisiae]
i4c07fs.r1	111	0.0011	284 457	dbj|BAA92008.1| (AK001969) unnamed protein product [Homo sapiens]>emb|CAB77023.1| (AJ275986) transcription factor [Homo sapiens]
k1a06fs.r1	99	0.0011	305 442	gi|6321523 ref|NP_011600.1|YGR086C| Ygr086cp >sp|P53252|YG2J_YEASTHYPOTHETICAL 38.3 KD PROTEIN IN RPL11B-PDC6 INTERGENIC REGION>pir||S64381 hypothetical protein YGR086c - yeast (Saccharomycescerevisiae) >emb|CAA97088.1| (Z72871) ORF YGR086c [Saccharomycescerevisiae]
t2h05fs.f1	111	0.002	246 530	pir||T41068 hypothetical protein SPCC1682.11c - fission yeast(Schizosaccharomyces pombe) >emb|CAA20677.1| (AL031525)hypothetical protein [Schizosaccharomyces pombe]
q2d11fs.r1	98	0.002	216 383	emb|CAB60686.1| (AL132870) hypothetical protein [Schizosaccharomyces pombe]
b2f01fs.r1	102	0.0021	259 411	sp|Q10328|YD73_SCHPO HYPOTHETICAL 104.0 KD PROTEIN C32A11.03C IN CHROMOSOME I>pir||T38649 hypothetical homeobox domain protein - fission yeast(Schizosaccharomyces pombe) >emb|CAA93700.1| (Z69796)hypotheticalhomeobox domain protein [Schizosaccharomyces pombe]
Contig994	114	0.0022	177 377	pir||T26267 hypothetical protein W07G1.3 - Caenorhabditis elegans>emb|CAB04935.1| (Z82076) predicted using Genefinder; cDNA ESTyk443d7.5 comes from this gene [Caenorhabditis elegans]
p3b02fs.r1	112	0.0032	11 289	gb|AAF24952.1|AC0123 (AC012375) T22C5.24 [Arabidopsis thaliana]
j3h08fs.r1	106	0.0039	242 541	pir||T41074 hypothetical protein SPCC16A11.01 - fission yeast(Schizosaccharomyces pombe) >pir||T41516 hypothetical proteinSPCC63.15 - fission yeast (Schizosaccharomyces pombe)>emb|CAB40019.1| (AL049522)hypothetical protein[Schizosaccharomyces pombe] >emb|CAB53073.1| (AL109957)hypothetical protein [Schizosaccharomyces pombe]
l4c03fs.r1	90	0.0044	426 515	emb|CAB51774.1| (AJ243960) hypothetical protein [Kluyveromyces lactis]

a1d05fs.f1	104	0.0049	18 275	gi|6324200 ref|NP_014270.1|YNL129W| Ynl129wp>sp|P53915|YNM9_YEASTHYPOTHETICAL 27.7 KD PROTEIN IN CPT1-SPC98 INTERGENIC REGION>pir||S55154 hypothetical protein YNL129w - yeast (Saccharomycescerevisiae) >emb|CAA86896.1| (Z46843) orf19 [Saccharomycescerevisiae] >emb|CAA96011.1| (Z71405) ORF YNL129w [Saccharomycescerevisiae]
f3b11fs.f1	109	0.0087	274 513	gi|6322586 ref|NP_012660.1|YJR126C| Yjr126cp>sp|P47161|YJ96_YEASTHYPOTHETICAL 92.0 KD PROTEIN IN RPS5-ZMS1 INTERGENIC REGION>pir||S57149 probable membrane protein YJR126c - yeast(Saccharomyces cerevisiae) >emb|CAA89657.1| (Z49626) ORF YJR126c[Saccharomyces cerevisiae]
r3e12fs.r1	94	0.0092	197 322	pir||T26607 hypothetical protein Y37A1B.14 - Caenorhabditis elegans>emb|CAA19485.1| (AL023835) predicted using Genefinder; similar toSrc homology domain 3 (2 domains) [Caenorhabditis elegans]
Contig579	107	0.021	224 355	gi|6323785 ref|NP_013856.1|GAT2| Gat2p >sp|P40209|YM19_YEAST HYPOTHETICAL 63.1KD PROTEIN IN REC114-PSO2 INTERGENIC REGION>pir||S50392hypothetical protein YMR136w - yeast (Saccharomyces cerevisiae)>emb|CAA87350.1| (Z47071) unknown [Saccharomyces cerevisiae]
Contig358	101	0.028	359 646	pir||T38995 hypothetical protein SPAC637.03 - fission yeast(Schizosaccharomyces pombe) >emb|CAA22582.1| (AL034583)hypothetical protein [Schizosaccharomyces pombe]
k4b02fs.r1	100	0.031	230 385	gb|AAF13090.1|AC0091 (AC009176) unknown protein [Arabidopsis thaliana]>gb|AAF21182.1|AC013483_6 (AC013483) unknown protein [Arabidopsisthaliana]
s1e01fs.r1	103	0.041	93 344	sp|Q09214|YP65_CAEEL HYPOTHETICAL 81.5 KD PROTEIN B0495.5 IN CHROMOSOME II>gb|AAA62531.1| (U21317) similar to SP:YR40_BACSU (P37512)hypothetical 78.8 kD protein in TETB-EXOA intergenic region[Caenorhabditis elegans]
o3a07fs.f1	95	0.046	425 529	gi|6324116 ref|NP_014186.1|YNL213C| Ynl213cp>sp|P40156|YNV3_YEASTHYPOTHETICAL 25.3 KD PROTEIN IN PEX17-MER1 INTERGENIC REGION>pir||S50718 hypothetical protein YNL213c - yeast (Saccharomycescerevisiae) >emb|CAA55495.1| (X78898) N1323 [Saccharomycescerevisiae] >emb|CAA96115.1| (Z71489) ORF YNL213c [Saccharomycescerevisiae]
n3g09fs.r1	95	0.056	378 536	pir||T41608 hypothetical protein SPCC790.03 - fission yeast(Schizosaccharomyces pombe) >emb|CAA21293.1| (AL031855)hypothetical protein [Schizosaccharomyces pombe]

m1b09fs.r1	97	0.077	203	439	gi|6320939 ref|NP_011018.1|YER093C| Yer093cp>sp|P40061|YES3_YEASTHYPOTHETICAL 164.4 KD PROTEIN IN MET6-PUP3 INTERGENIC REGION>pir||S50596 hypothetical protein YER093c - yeast (Saccharomycescerevisiae) >gb|AAB64648.1| (U18839) Yer093cp [Saccharomycescerevisiae]
c3b06fs.f1	99	0.078	347	472	sp|Q09706|YA2G_SCHPO HYPOTHETICAL 157.7 KD PROTEIN C2F7.16C IN CHROMOSOME I>pir||S58160 hypothetical protein SPAC2F7.16c - fission yeast(Schizosaccharomyces pombe) >pir||T38564 hypothetical proteinSPAC2F7.16c - fission yeast (Schizosaccharomyces pombe)>emb|CAA90503.1| (Z50142) putative phospholipase d1[Schizosaccharomyces pombe]
Contig461	98	0.19	610	798	pir||T05349 hypothetical protein F8B4.90 - Arabidopsis thaliana>emb|CAA22566.1| (AL034567) putative protein [Arabidopsis thaliana]>emb|CAB79956.1| (AL161581) putative protein [Arabidopsis thaliana]
<HET-C protein>					
j2f04fs.f1	645	2.3e-62	13	480	gb|AAD54275.1|AF1697 (AF169793) HET-C protein [Podospora anserina]
Contig166	356	7.6e-31	250	579	gb|AAD54275.1|AF1697 (AF169793) HET-C protein [Podospora anserina]
<acr-2 protein>					
d1g06fs.f1	120	1.2e-05	56	205	pir||S72537 acr-2 protein - Neurospora crassa
<MYO2>					
h4e10fs.f1	409	2.6e-36	99	602	emb|CAA89973.1| (Z49821) MYO2 [Saccharomyces cerevisiae]>emb|CAA99648.1|(Z75235) ORF YOR326w [Saccharomyces cerevisiae]
<SCD6 protein>					
l1a11fs.r1	313	3.5e-27	195	485	dbj|BAA13831.1| (D89169) similar to Saccharomyces cerevisiae SCD6 protein,SWISS-PROT Accession Number P45978 [Schizosaccharomyces pombe]
<CG6198 gene product>					
Contig308	319	8.5e-28	98	628	gb|AAF56252.1| (AE003746) CG6198 gene product [Drosophila melanogaster]
<CG7459 gene product>					
Contig521	149	8.3e-10	237	605	gb|AAF54223.1| (AE003678) CG7459 gene product [Drosophila melanogaster]
<CG13902 gene product>					
Contig76	172	6.5e-10	186	662	gb|AAF47420.1| (AE003469) CG13902 gene product [Drosophila melanogaster]
<CG9953 gene product>					

Contig845	212	1.3e-15	72	680	gb|AAF50628.1| (AE003560) CG9953 gene product [Drosophila melanogaster]
<CG10627 gene product>					
b2c06fs.r1	284	2.2e-23	97	384	gb|AAF49901.1| (AE003541) CG10627 gene product [Drosophila melanogaster]
<CG14542 gene product>					
e3f07fs.r1	259	2e-21	201	497	gb|AAF56559.1| (AE003754) CG14542 gene product [Drosophila melanogaster]
<CG8665 gene product>					
e4c02fs.r1	295	3.8e-24	88	480	gb|AAF53994.1| (AE003670) CG8665 gene product [Drosophila melanogaster]
<CG9318 gene product>					
f2b11fs.f1	197	8.9e-14	398	649	gb|AAF53917.1| (AE003667) CG9318 gene product [Drosophila melanogaster]
<CG5340 gene product>					
f3b02fs.f1	202	1.8e-14	120	548	gb|AAF56014.1| (AE003739) CG5340 gene product [Drosophila melanogaster]
<CG10882 gene product>					
f3h07fs.r1	202	6.3e-14	68	454	gb|AAF51283.1| (AE003583) CG10882 gene product [alt 2] [Drosophila melanogaster]
<CG7759 gene product>					
a2h01fs.r1	109	0.00016	33	356	gb|AAF58645.1| (AE003826) CG7759 gene product [Drosophila melanogaster]
<Spx gene product>					
g2e03fs.r1	314	3e-27	161	466	gb|AAF46136.1| (AE003437) Spx gene product [Drosophila melanogaster]
<Slh gene product>					
g3f06fs.f1	133	6.2e-07	87	290	gb|AAF51247.1| (AE003583) Slh gene product [Drosophila melanogaster]
<CG8031 gene product>					
h4h05fs.f1	158	2.4e-10	254	511	gb|AAF54889.1| (AE003698) CG8031 gene product [Drosophila melanogaster]
<CG3172 gene product>					
i1g04fs.r1	159	2.9e-10	85	429	gb|AAF55022.1| (AE003703) CG3172 gene product [Drosophila melanogaster]
<PDI related protein A>					
j4b04fs.r1	570	2e-54	1	504	gb|AAC77456.1| (AF095899) PDI related protein A [Aspergillus niger]
<BG:DS02740.2 gene product>					
k1a05fs.r1	136	6e-08	92	445	gb|AAF53509.1| (AE003650) BG:DS02740.2 gene product [Drosophila melanogaster]

<CG7828 gene product>					
m1g10fs.r1	300	3.2e-25	70	435	gb|AAF50102.1| (AE003546) CG7828 gene product [Drosophila melanogaster]
<CG8430 gene product>					
n1d10fs.r1	259	3.6e-21	75	497	gb|AAF58059.1| (AE003808) CG8430 gene product [Drosophila melanogaster]
q2c04fs.r1	186	5.4e-13	114	404	gb|AAF58059.1| (AE003808) CG8430 gene product [Drosophila melanogaster]
<beta'Cop gene product>					
q2c07fs.r1	423	6.9e-38	32	439	gb|AAF53294.1| (AE003639) beta'Cop gene product [Drosophila melanogaster]
<QM protein>					
q4e04fs.r1	571	1.7e-54	36	416	gb|AAC98301.1| (AF099012) QM protein [Bombyx mandarina]
<CG12110 gene product>					
s2c06fs.r1	176	3.5e-11	52	291	gb|AAF57264.1| (AE003784) CG12110 gene product [alt 1] [Drosophilamelanogaster] >gb|AAF57265.1| (AE003784) CG12110 gene product [alt3] [Drosophila melanogaster]
<CG15015 gene product>					
s2e11fs.r1	147	1.4e-08	56	604	gb|AAF47887.1| (AE003481) CG15015 gene product [Drosophila melanogaster]
<CG10306 gene product>					
s3h02fs.r1	163	2.7e-11	79	327	gb|AAF46753.1| (AE003454) CG10306 gene product [Drosophila melanogaster]
<CG7218 gene product>					
s2g12fs.r1	95	0.35	285	446	gb|AAF55495.1| (AE003721) CG7218 gene product [Drosophila melanogaster]
<Xenopus 14s cohesin smc1 subunit homolog>					
c2b03fs.f1	219	9.3e-16	218	463	pir||T40059 Xenopus 14s cohesin smc1 subunit homolog - fission yeast(Schizosaccharomyces pombe) >emb|CAA22432.1| (AL034463) Xenopus 14scohesin smc1 subunit homolog [Schizosaccharomyces pombe]
<KIAA0585 protein>					
j3a02fs.r1	312	4.9e-27	6	497	dbj|BAA25511.1| (AB011157) KIAA0585 protein [Homo sapiens]
<KIAA0713 protein>					
Contig133	111	0.00046	16	348	dbj|BAA34433.1| (AB018256) KIAA0713 protein [Homo sapiens]
<KIAA0416>					
j3e11fs.f1	99	0.058	375	539	dbj|BAA24846.1| (AB007876) KIAA0416 [Homo sapiens]
<CG17661 gene product>					

Contig34	121	1.3e-06	416	559	gb|AAF45393.1| (AE003000) CG17661 gene product [Drosophila melanogaster]
<CG10843 gene product>					
Contig1005	183	1.7e-12	184	516	gb|AAF58961.1| (AE003834) CG10843 gene product [Drosophila melanogaster]
Contig581	178	6.2e-12	286	612	gb|AAF58961.1| (AE003834) CG10843 gene product [Drosophila melanogaster]
<ORF YOL057w>					
Contig723	269	1.7e-22	130	525	emb|CAA99065.1| (Z74798) ORF YOL057w [Saccharomyces cerevisiae]
<ORF YGR159c>					
i1e04fs.r1	105	5.3e-05	21	398	emb|CAA97180.1| (Z72946) ORF YGR159c [Saccharomyces cerevisiae]
<01232>					
b2e06fs.r1	352	3.1e-31	29	388	emb|CAA62529.1| (X91067) 01232 [Saccharomyces cerevisiae]
<similar to S. cerevisiae PKR1>					
q4g04fs.r1	155	1.8e-10	36	227	emb|CAB66430.1| (AL136535) similar to S. cerevisiae PKR1 [Schizosaccharomycespombe]
<INTEGUMENTARY MUCIN A.1 PRECURSOR>					
g2a11fs.r1	174	7.3e-12	18	425	sp|P10667|MUA1_XENLA INTEGUMENTARY MUCIN A.1 PRECURSOR (FIM-A.1) (PREPROSPASMOLYSIN) >pir||A28172 spasmolysin precursor - Africanclawed frog >gb|AAA49960.1| (M19971) spasmolysin (put.);putative [Xenopus laevis]
g2f09fs.r1	152	2e-09	184	429	sp|P10667|MUA1_XENLA INTEGUMENTARY MUCIN A.1 PRECURSOR (FIM-A.1) (PREPROSPASMOLYSIN) >pir||A28172 spasmolysin precursor - Africanclawed frog >gb|AAA49960.1| (M19971) spasmolysin (put.);putative [Xenopus laevis]

<NONE>
-Contigs 424
-Singlets 1215

Part V. No significant homolog