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.

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



Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/348

UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

EXPRESSION PROFILES AND THE IDENTIFICATION OF THE CLOCK-CONTROLLED GENES IN TWO TIME-OF-DAY-SPECIFIC cDNA LIBRARIES OF *NEUROSPORA CRASSA*

A Dissertation

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

By

Hua Zhu

Norman, Oklahoma

2001

UMI Number: 3014527



UMI Microform 3014527
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.

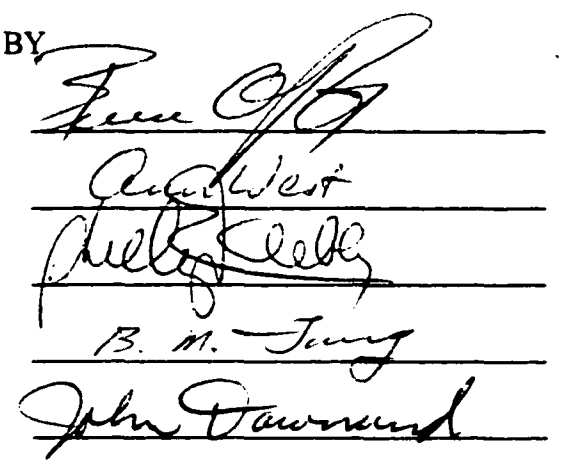
Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

EXPRESSION PROFILES AND THE IDENTIFICATION OF THE CLOCK-CONTROLLED GENES IN TWO TIME-OF-DAY-SPECIFIC cDNA LIBRARIES OF *NEUROSPORA CRASSA*

A Dissertation APPROVED FOR THE

DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

BY



© Copyright by Hua Zhu 2001
All Rights Reserved

Acknowledgements

I would like to express my sincere gratitude to my major professor, Dr. Bruce A. Roe, from my deep heart for everything he did for me. Without his guidance, knowledge, patience, and support, I could not have done anything. I am so grateful to him for his love and support of me and my family. It is my honor to study in his great and warm lab and that he brought me to my husband in front of God on my wedding.

Special thanks and appreciation also go to Dr. Scott Russell. He gave me the opportunity to study in his lab under his supervision for two years and helped me a lot in my life too. And I always remember and appreciated that he picked me up at the airport and arranged for an apartment for me when I arrived in the USA in 1994. Therefore, I did not experience the helplessness and feeling lost like most of the international students at the very beginning when they arrive in a new country. I also extend my appreciation to my committee members, Drs. Phil Klebba, Ann West, Bing M. Fung and John Downard for their help and encouragement during my study.

A big thanks goes to everyone in Dr. Roe's lab, both past and present, for all of their help, support, discussion and friendship. Thanks to: Feng Chen, Lingzhi Chu, Sandy Clifton, Judy Crabtree, Rose Morales-Diaz, Ann Do, Trang Do, Sara Downard, Mounir Elharam, Ying Fu, Jennifer Gray, Kaylynn Hale, Jennifer Hausner, Ping Hu, Xiaohong Hu, Honggui Jia, Xiuhong Jiang, Steve Kenton, Christopher Lau, Lisa Lane, Sharon Lewis, Phoebe Loh, Jianfeng Li, Shuling Li, Eda Malaj, Jami Milam, Ying Ni, Huaqin Pan, Sulan Qi, Linda Ray, Qun Ren, Steve Shaull, Lin Song, Runying Tian, Honggui Jia, Qiaoyang Wang, Yingping Wang, Zhili Wang, Hongmin Wu, Hui Wu, Doug White, Jim White, Dixie Wishnuck, Mary Williams, Heather Wright and her gel crew, Tammy Womack, Weihong Xu, Limei Yang, Ziyun Yao, YoungJu Yoon, Xiling Yuan, Min Zhan, Guozhong Zhang, Liping Zhou. A special thanks goes to Hongshing Lai, Doris Kupfer, Axin Hua, Shaoping Lin, and Fares Najar. Without their countless help, I could not finish this dissertation. Hongshing always encouraged me and was there for me whenever I need her help both mentally and computer technically. And many thanks and appreciation goes to Dr. Doris Kupfer for her discussions and correction of my dissertation.

I also thank Dr. Dunlap and his group for their help and support.

I could not forget to thank all friends in the Trinity Baptist Church in Norman, both American and Chinese friends. They helped me so much in my life and brought me to a new world and another big family.

Immense thanks and appreciation go to both my family members and my husband's family members. My mom, Ms. Yuanying Long, is always the person who raised my two babies when they were infants. She has been suffering so much to come to the USA to help me and take care of my child. Without her endless love and sacrifice, I could not imagine what I would be able to do. I could not ever return her and my father's greatest love to me in my life. I would like to express my great thanks and appreciation to my sister Xin Zhu, three brothers and sisters-in-law, my parents-in-law, and two sisters of my husband, Mingzhen Yao and Minghui Yao, as well. They always supported me, loved me and encouraged me in my study. Especially, they help me to raise my dear daughter Jessica Xiaoman Yao with their endless love.

Last but not least is my great thanks to my husband, Minghua Yao, and my dear daughters, Jessica Xiaoman Yao and Stephanie Xiaoru Yao, for their love, patience, and encouragement. I owe my dear daughter, Jessica Xiaoman Yao, for her sacrifice. Without their help and support, I could never be able to write this dissertation.

Everything for me in my life, if I have, is credit to my family members, my major advisor and my friends.

Table of contents

List of Tables X

List of Figures XII

Abstract XV

Chapter I: Introduction 1

1.1 *Neurospora crassa* and circadian rhythms 1
1.1.1 Introduction to *N. crassa* 1
1.1.2 Life cycle of *N. crassa* 2
1.1.3 *N. crassa* is an ideal model system for the study of circadian rhythms 5
1.1.4 Concept and characteristics of circadian rhythms 7
1.1.5 Studies of the clock-controlled genes in *N. crassa* 11

1.2 Expressed Sequenced Tags and cDNA library 14
1.2.1 DNA and RNA 14
1.2.2 Gene 14
1.2.3 Transcription and mRNA 16
1.2.4 cDNA and Expressed Sequenced Tags 18
1.2.4.1 The application of ESTs to the human genome project 22
1.2.4.2 The accuracy of EST sequences 24
1.2.4.3 Editing of EST data 25
1.2.4.4 EST analysis and the classification of biological functions 25
1.2.5 cDNA library construction 28
1.2.5.1 mRNA in eukaryotes 29
1.2.5.2 mRNA isolation 30
1.2.5.3 cDNA preparation 31
1.2.5.3.1 The synthesis of the first strand of cDNA 32
1.2.5.3.2 The synthesis of the second strand of cDNA 33
1.2.5.4 Vector for cDNA library construction 35
1.2.6 cDNA sequencing and the generating of EST sequences 37
1.2.6.1 Sanger enzyme method of DNA synthesis 39
1.2.6.2 Enzyme in DNA sequencing reaction 43
1.2.6.3 Labeling methods in DNA sequencing reactions 44
1.2.7 DNA sequence analysis 46

Chapter II: Materials and Methods 50

2.1 EST sequencing of two cDNA libraries from *N. crassa* 50
2.1.1 Strain and growth conditions 50
2.1.2 Construction of cDNA library 51
2.1.2.1 The synthesis of double stranded DNA and the λ primary cDNA library 51

2.1.2.2 Mass excision of the pBluescript phagemid from the Uni-Zap vector 54
2.1.2.3 Conversion of pBluescript phagemid DNA to plasmid DNA 55
2.1.3 Growth and amplification of cDNA clones 56
2.1.4 Alkaline lysis isolation of subclone DNA 56
2.1.5 cDNA sequencing 58
2.1.5.1 Enzyme in the EST sequencing 58
2.1.5.2 Primers in the EST sequencing 59
2.1.5.3 Thermocycler reaction and purification 59
2.1.6 Sample loading, electrophoresis, data collection and equipment 62
2.1.7 cDNA gap closure 63
2.1.7.1 Primer walking 64
2.1.7.2 PCR in the direct sequencing phase 65
2.1.8 The 96-well block retransformation 67
2.1.8.1 Preparation of glycerol stocks 67
2.1.8.2 96-well block transformation with $CaCl_2$ competent cells 68
2.1.8.2.1 Preparation of CaCl2 competent cells 68
2.1.8.2.2 The test of Calcium chloride competent cells 69
2.1.8.2.3 96-well block retransformation 71
2.1.9 Analysis of *N. crassa* EST sequences 72
2.1.9.1 Clip and clean of the raw EST sequences 73
2.1.9.2 Submission of ESTs to the dbEST of Genbank 77
2.1.9.3 The strategy for EST database analysis 83
2.1.9.4 Assembly of 3' ESTs 86
2.1.9.5 Assembly of both 3' ESTs and 5' ESTs 87
2.1.9.6 The BlastX and the annotation of the two assembled EST databases 88
2.1.9.7 The strategy of biological function assignment 93

2.2 Sequencing of two *N. crassa* cosmids nc14 and nc17 96
2.2.1 Overall sequencing strategy 97
2.2.1.1 Large scale DNA isolation 97
2.2.1.2 DNA physical shearing by nebulization 100
2.2.1.3 Random fragment end-repair and phosphorylation 101
2.2.1.4 Shotgun library construction 102
2.2.1.4.1 Ligation 103
2.2.1.4.2 Transformation 103
2.2.1.5 Semiautomated 96-well subcloned DNA isolation 104
2.2.1.6 Thermocycler sequencing reaction and data collection 105
2.2.1.7 Shotgun sequence edit and assembly 105
2.2.1.8 Gap closure of cosmid sequence 107
2.2.2 Sequencing analysis of the *N. crassa* cosmids 108

Chapter III: Results and Discussions 111

3.1 Two cDNA libraries and EST databases of *N. crassa* 111

3.1.1 EST database development 111
3.1.2 The summary of EST data of two *N. crassa* cDNA libraries 112
3.1.3 EST database analysis 114
3.1.3.1 Assembly of 3' ESTs 115
3.1.3.2 Assembly of both 3' ESTs and 5' ESTs 120
3.1.3.3 Estimation of gene numbers in two cDNA libraries 125
3.1.3.4 The redundancy of genes in two cDNA libraries 127
3.1.3.5 The ten most highly expressed genes in two cDNA libraries 133
3.1.4 Biological classification of expressed genes in two cDNA libraries of *N. crassa* 135
3.1.5 Comparison of two EST databases of *Neurospora crassa* 148
3.1.5.1 The combined EST database 150
3.1.5.2 Gene distribution in the combined EST database 151
3.1.5.3 Genes only expressed in the morning cDNA library 153
3.1.5.4 Genes only expressed in the evening cDNA library 157
3.1.5.5 Genes expressed in both of the cDNA libraries 161
3.1.6 *Neurospora* genes that have no sequence homolog 164

3.2 The clock-controlled genes in *Neurospora crassa* 166
3.2.1 The clock-controlled gene1 166
3.2.1.1 The *ccg-1* and catabolic repression in fungi 174
3.2.1.2 The relationship between glucose repression and conidiation in *N.crassa* 175
3.2.2 The clock-controlled gene 2 176
3.2.2.1 The *ccg-2* gene was isolated as *eas* and *bli-7* in the different screens 180
3.2.2.2 The regulation of *ccg-2* and the comparisons of the promoters of three *ccg* genes 184
3.2.3 The clock-controlled gene 7 186
3.2.4 The clock-controlled gene 4 198
3.2.4.1 Two putative proteins reported previously for *ccg-4* gene 199
3.2.4.2 Gap closure of the cDNA of the *ccg-4* gene 202
3.2.4.3 The identity of the *ccg-4* gene as *Neurospora* pheromone precursor homologue of *Sordaria macrospora* 205
3.2.4.4 The comparison and new annotation of the *ccg-4* gene and its cDNA sequences 210
3.2.4.5 The submission of the cDNA sequence of *Neurospora ppg-1* gene homologue 211
3.2.5 The clock-controlled gene 6 213
3.2.6 The clock-controlled gene 8 221
3.2.7 The clock-controlled gene 9 225
3.2.8 The identification of novel clock-controlled genes 238
3.2.9 Some common features of the *ccg* genes identified in *N. crassa* 240
3.2.9.1 The introns in the clock-controlled genes *N. crassa* 242
3.2.9.2 The promoters of several clock-controlled genes 243

3.2.9.3 The poly(A) signal of the *ccgs* in *N. crassa* 244

3.3 Analysis and annotation of the sequences of *Neurospora* cosmids 245
3.3.1 The annotation of the nc17 sequence with GENSCAN 247
3.3.1.1 *Neurospora Asm-1*, the first predicted gene 253
3.3.1.2 The second GENSCAN predicted gene 256

Chapter IV: Conclusion 265

Chapter V: Literature cited 279

Appendix 297
I. Keywords for *N. crassa* morning cDNA library category of cellular functions 297
II. Keywords for *N. crassa* evening cDNA library category of cellular functions 309
III. *N. crassa* morning cDNA library category of cellular functions 335
IV. *N. crassa* evening cDNA library category of cellular functions 356
V. Genes only expressed in NM cDNA library that contain 3' ESTs only 404
VI. Genes only expressed in NM cDNA library that contain 5' ESTs only 407
VII. Genes only expressed in NM cDNA library that contain both 3' and 5' ESTs only 410
VIII. Genes only expressed in NE cDNA library that contain 3' ESTs only 413
IX. Genes only expressed in NE cDNA library that contain 5' ESTs only 419
X. Genes only expressed in NE cDNA library that contain Both 3' and 5' ESTs only 426
XI. Genes expressed both in NM and NE cDNA libraries 435

List of Tables

Table 1.01 Summary of the clock-controlled genes of *Neurospora crassa* 13
Table 1.02 Comparative properties of eukaryotic RNA polymerase 30
Table 1.03 Blast programs and their descriptions 49
Table 2.01 Summary of *N. crassa* cDNA library 56
Table 2.02 Clip and clean scripts for EST processing 73
Table 2.03 The protocol of clean and clip 76
Table 2.04 Sequences used for screening mitochondrial and ribosomal sequences of ESTs 77
Table 2.05 The list of the EST numbers of different organisms in the Genbank 77
Table 2.06 The access numbers of *Neurospora* ESTs submitted to the dbEST of Genbank 81
Table 2.07 Three files resulting from the BlastX search 91
Table 2.08 DNA ligation protocol 103
Table 3.01 EST data summary of the two *N. crassa* cDNA libraries 112
Table 3.02 No pass ESTs in two *N. crassa* cDNA libraries 114
Table 3.03 3' EST assembly and the percent of new genes in *Neurospora* morning cDNA library 116
Table 3.04 3' EST assembly and the percent of new genes in *Neurospora* evening cDNA library 117
Table 3.05 Redundancy of the genes in the *Neurospora* morning cDNA library by cumulation of 3' ESTs 118
Table 3.06 Redundancy of the genes in the *Neurospora* evening cDNA library by cumulation of 3' ESTs 118
Table 3.07 The output files of the BlastX search of the GenBank nr protein database 121
Table 3.08 Estimation of the gene numbers in the assembled EST databases 125
Table 3.09 The complexity of *Neurospora* morning cDNA library 128
Table 3.10 The complexity of *Neurospora* evening cDNA library 129
Table 3.11 The contig compositions in two EST databases 130
Table 3.12 Gene expression levels in the morning cDNA library 132
Table 3.13 Gene expression levels in the evening cDNA library 132
Table 3.14 The ten most highly expressed genes in two cDNA libraries 134
Table 3.15 *N. crassa* morning cDNA library categories of cellular function 138
Table 3.16 *N. crassa* evening cDNA library categories of cellular function 140
Table 3.17 Genes in different biological pathway from two cDNA libraries of *Neurospora* 142
Table 3.18 Estimation of the genes in the combined EST database of *N. crassa* 150
Table 3.19 The distribution of ESTs from NM and NE libraries in the combined EST database 151
Table 3.20 Genes only expressed in one of the two different cDNA libraries 153
Table 3.21 The biological function classification of the genes that were only detected in the NM library 155
Table 3.22 Top ten highly expressed genes only detected in the NM cDNA library 157
Table 3.23 The biological function classification of the genes that were only detected in the NE library 158

Table 3.24 Top ten highly expressed genes only detected in the NE cDNA library 160
Table 3.25 Top ten highly expressed genes detected in both libraries 163
Table 3.26 Five potential novel genes expressed in both of cDNA libraries 166
Table 3.27 The contigs and the EST counts of *ccg-1* in combined EST database 167
Table 3.28 The contigs and the EST counts of *ccg-2* in combined EST database 177
Table 3.29 The summary of the promoters of four *ccg* genes 186
Table 3.30 The contigs and the EST counts of *ccg-7* in combined EST database 186
Table 3.31 The contigs and the EST counts of *ccg-4* in combined EST database 198
Table 3.32 BlastX of the contigs representing the *ccg-4* gene in the NE EST database 199
Table 3.33 The three new contigs representing the *ccg-4* gene 204
Table 3.34 The contigs and the EST counts of *ccg-6* in combined EST database 214
Table 3.35 The contigs and the EST counts of *ccg-8* in combined EST database 221
Table 3.36 The contigs and the EST counts of *ccg-9* in combined EST database 225
Table 3.37 The 26 cDNA clones used for RNA hybridization 239
Table 3.38 BlastX sequence homoiogs of the four new clock-controlled genes 240
Table 3.39 11 clock-controlled genes detected in *N. crassa* EST research 241
Table 3.40 Introns in the clock-controlled genes of *N. crassa* 243
Table 3.41 Summary of the promoters of the clock-controlled genes in *N. crassa* 243
Table 3.42 Ten predicted genes from nc17 by GENSCAN 247
Table 3.43 Ten predicted peptides from nc17 by GENSCAN 249
Table 3.44 The BlastP results of 10 predicted genes from nc17 251
Table 3.45 Two domains searched in peptide 6 using profileScan 260
Table 3.46 The PROSITE pattern searched in peptide 6 using ScanProsite 261
Table 3.47 Top five significant homologs of peptide#6 and the NE.Contig510 262
Table 4.01 The enzymes detected in three major bioenergetic pathways of *Neurospora crassa* two cDNA libraries 272
Table 4.02 The enzymes missed from the three major bioenergetic pathways of *Neurospora crassa* two cDNA libraries 272

List of Figures

Figure 1.01 Life cycle of *N. crassa* 3
Figure 1.02 SEM photos of *Neurospora* hyphae and spores 4
Figure 1.03 Race tube 7
Figure 1.04 Diagram showing the elements of oscillator and circadian rhythm in *N. crassa* 10
Figure 1.05 Putative mechanism of circadian rhythm in *N. crassa* 13
Figure 1.06 Central dogma of molecular biology 14
Figure 1.07 Typical structure of an eukaryotic gene and its expression 15
Figure 1.08 Typical structure of a mature mRNA 17
Figure 1.09 The GT-AG pair at the end of nuclear intron sequence 17
Figure 1.10 The synthesis of the first strand of cDNA 32
Figure 1.11 The synthesis of the second strand of cDNA 34
Figure 1.12 The lambda cDNA primary library 35
Figure 1.13 The strategy for sequencing ESTs from *N. crassa* 38
Figure 1.14 Assembly of both 3' ESTs and 5' ESTs into contig with Phrap and the stagger of 5' ESTs 40
Figure 1.15 Sanger enzyme method in DNA synthesis 42
Figure 2.01 Mycelial harvest protocol for clock-controlled genes 51
Figure 2.02 The synthesis of the double-stranded cDNA 52
Figure 2.03 Thermocycler reaction on the GeneAmp9600 60
Figure 2.04 Strategy for closing gap cDNA clones 64
Figure 2.05 Gap closure using primer walking 65
Figure 2.06 Flowchart of the $CaCl_2$ competent cell test 70
Figure 2.07 Three files attached to the ESTs in the EST submissions 80
Figure 2.08 Submission of ESTs of *N. crassa* to dbEST of GenBank 82
Figure 2.09 EST database analysis strategy 84
Figure 2.10 *N. crassa* biological function classification outline 94
Figure 2.11 Flowchart of the sequencing of *N. crassa* cosmids 98
Figure 3.01 3' ESTs assembly and the gene redundancy in two libraries 119
Figure 3.02 The assembly of 3' and 5' ESTs and the 5' stagger 124
Figure 3.03 The piechart view of the genes in different biological categories 144
Figure 3.04The distribution of genes in two cDNA libraries 152
Figure 3.05 New piechart of the genes in different biological categories after tBlastX search 165
Figure 3.06 Annotation of *ccg-1* gene 168
Figure 3.07 Sequence comparison between the contig1410 and the *ccg-1* mRNA 170
Figure 3.08 The translation products of five contigs representing the *ccg-1* gene 171
Figure 3.09 The alternative polyadenylation sites in five contigs of the *ccg-1* gene 173
Figure 3.10 Annotation of *ccg-2* gene 178
Figure 3.11 DNA sequence comparison between the contig1447 and the *ccg-2* mRNA 179
Figure 3.12 Protein sequence comparison between the contig1447 and the *ccg-2* mRNA 179
Figure 3.13 Protein sequence comparison between the *N. crassa* CCG2 and the *Aspergillus nidulans* RodA 181

Figure 3.14 The hydrophobicity plot of the *N. crassa* CCG2 protein 182
Figure 3.15 The hydrophobicity plot of the *Aspergillus nidulans* RodA 183
Figure 3.16 The sequence comparisons among seven conitgs of the *ccg-7* 189
Figure 3.17 The translated protein sequence comparisons among seven contigs of the *ccg-7* 190
Figure 3.18 Annotation of *ccg-7* gene 192
Figure 3.19 The DNA sequence comparison between the conti1448 and the *ccg-7* gene 195
Figure 3.20 Comparisons of the *ccg-7* translated product with other GAPDHs 196
Figure 3.21 The relationship tree of GAPDHs 197
Figure 3.22 Annotation and comparison of *ccg-4* gene 200
Figure 3.23 Two cDNAs of the ccg-4 gene represented by newccg4contig2 203
Figure 3.24 Positions of 3 ccg4contigs and the ORF for *Neurospora* pheromone homolog 204
Figure 3.25 Comparison between ccg4contig3 translated protein and *ccg-4* protein1, 2 and *Sordaria* pheromone precursor 206
Figure 3.26 Comparisons between ccg4contig3 translated product and *Sordaria* pheromone precursor 207
Figure 3.27 Hydrophobicity plot of *N. crassa* homolog of *Sordaria* pheromone precursor 209
Figure 3.28 Hydrophobicity plot of *Sordaria* pheromone precursor 210
Figure 3.29 The GenBank submission file of *Neurospora* cDNA for *ppg-1* homolog 213
Figure 3.30 Annotation of *ccg-6* gene 215
Figure 3.31 The DNA sequence comparison between the contig1435c and the *ccg-6* gene 217
Figure 3.32 The DNA sequence comparison between the contig1435c and the *ccg-6* mRNA 218
Figure 3.33 Comparison between contig1435c peptide and *ccg-6* protein 218
Figure 3.34 The trace files in the contig1435c 219
Figure 3.35 The translated peptide sequence comparison of 4 contigs represented the *ccg-6* gene 221
Figure 3.36 The different polyadenylation sites in the 4 contigs represented the *ccg-6* gene 221
Figure 3.37 Annotation of the *ccg-8* gene 223
Figure 3.38 The DNA sequence comparison among the *ccg-8* gene, the *ccg-8* mRNA and the contig1356 224
Figure 3.39 The protein sequence comparison of the *ccg-8* gene, the *ccg-8* mRNA and the contig1356 225
Figure 3.40 The BlastX results of two newccg9contigs 229
Figure 3.41 The sequence comparison among the *ccg-9* genomic DNA, the *ccg-9* mRNA and the newccg9contig2 233
Figure 3.42 Annotation of the *ccg-9* gene 234
Figure 3.43 The trace files in the newccg9contig2 represented the *ccg-9* gene 235
Figure 3.44 Comparison among the *ccg-9* gene protein, *ccg-9* mRNA protein, and *Neurospora* trehalose synthase 236
Figure 3.45 The trace files in the newccg9contig2 237

Figure 3.46 Comparison among the newccg9contig1, 2 and the *ccg-9* mRNA 238
Figure 3.47 Ten genes predicted by GENSCAN 249
Figure 3.48 powerBLAST and GENSCAN prediction of nc17 252
Figure 3.49 Annotation of *Neurospora Asm-1* gene 254
Figure 3.50 Sequence comparison among the homologs of *Neurospora* Asm-1 protein 256
Figure 3.51 Comparison between predicted peptide #6 and the translated product of NE.Contig510 258
Figure 3.52 Annotation of predicted gene #6 259
Figure 4.01 Gene expression profiles of the two cDNA libraries of *N. crassa* 267
Figure 4.02 Gene expression for the metabolism pathways of the two cDNA libraries of *Neurospora crassa* 268
Figure 4.03 Gene expression fir the bioenergetic pathways of the two cDNA libraries of *Neurospora crassa* 269
Figure 4.04 Diagram shows the different growth stage of two *N. crassa* strains at the time of harvest 274

Abstract

The ascomycete fungus *Neurospora crassa* is an ideal model system to study circadian rhythms, the widespread biological rhythms that are observed in a variety of organisms from *cyanobacteria* to human. Circadian rhythms are generated and controlled by an endogenous biological clock that has an innate, approximately daily periodicity. To study this phenomenon, time-of-day-specific cDNA libraries were constructed with the *N. crassa frq*$^{+}$ strain 30-7 (bd;A) and the long period mutant 695-425 (bd;*frq*^{7};A), resulting in a morning cDNA library (NM) equivalent to a clock time of one hour CT1 and an evening cDNA library (NE) equivalent to a clock time of 13 hours CT13, respectively. The sequences of both ends of the cDNAs from these two libraries represent their respective 3' ESTs and 5' ESTs.

A total of 10871 high quality ESTs obtained by end sequencing the morning library (NM) and 9148 high quality ESTs were generated from the evening cDNA library (NE) after removal of poor quality sequences, sequences with short inserts (< 100 bp), vector sequences, wrong end sequences (i.e. those with no poly T on 3' end), and contaminating *E. coli*, mitochondrial, and ribosomal RNA sequences. The 3' ESTs and 5' ESTs were assembled into an EST database using the Phred/phrap base calling/assembly programs at the high stringency parameters of a minimum-window size of 14 and a minimum score of 80. The resulting NM assembled-EST database contained 527 contigs and 78 singlets while the NE assembled-EST database contained 1126 contigs and 650 singlets. A BlastX search of both NM and NE databases revealed that a DnaJ-like protein homologue, a RNA helicase, a cyclophilin, a keratin-2-epidermis, a dehydrogenase and four additional genes with unknown function were highly expressed only in the morning library while a subunit of the mitochondrial F_1F_0-ATPase, two different ribosomal RNA proteins, calmodulin, the alpha chain of a mitochondrial precursor and five genes with unknown function were highly expressed only in the evening library. In addition, 281 genes in the assembled EST databases were expressed only in *N. crassa* cDNA morning library while 894 genes were expressed only in cDNA evening library. Another 281 genes were expressed in both cDNA libraries but at different levels.

In the group of genes that were detected from both cDNA libraries, the highest three of the top ten highly expressed genes were clock-controlled genes (*ccgs*). That these

ccgs are so highly expressed in *Neurospora* implies that circadian regulation of a gene expression may be more pervasive than previously anticipated. The identity of the previously unknown *ccg-4* gene as a *N. crassa* homolog of a *Sordaria macrospora* pheromone precursor may indicate that the sexual cycle and meiosis also may be clock-controlled.

Sequence comparison was performed between the *N. crassa* genomic DNA sequence and the cDNA sequences of the identified clock-controlled genes. The promoter sequences, protein-binding sites and putative control elements of these clock-controlled genes were also compared and conserved regions were identified.

During this work, several instances of alternative splicing and alternative polyadenylation sites were observed for the fungal cDNAs. For example, of the highly expressed *ccg-1* gene, 75 of 510 ESTs had alternatively spliced forms and five had different position of polyadenylation in the resulting mRNAs.

In addition, through these studies, 4 new clock-controlled genes which showed circadian expression were detected. Three of these, *ccg-13*, *ccg-14*, *ccg-15*, have no clear function while the fourth is a homologue of N,O-diacetylmuramidase (lysozyme). These studies provide for the first time a detailed understanding of the genes expressed during the circadian cycle and will allow further studies into their involvement with other clock-related genes, for example, the *frq* and the *wc1* and *wc2* genes and their products, to maintain and regulate circadian rhythms. All EST sequences and two assembled-EST databases are publicly available both on our web site at http://www.genome.ou.edu/fungus and in GenBank.

Chapter I

Introduction

1.1 *Neurospora crassa* and circadian rhythms

Neurospora crassa has been used as a model system for eukaryotic genetic studies for over 100 years (Rowland, 2000). In 1927, Shear and Dodge used *N. crassa* to study the segregation of fungal mating types (Shear and Dodge, 1927). Dodge earned his doctoral degree from Columbia University in 1912 and was an instructor there. At that time, he was a colleague of Dr. Thomas Hunt Morgan. When Morgan was working on his genetic studies with the fruit fly, Dodge was working on the genetic basis of fungal mating types using *N. crassa*. Later, Morgan's student, Dr. Carl C. Lindegren used the *Neurospora* that Dodge gave to Morgan to confirm the studies of Dodge and verify that *N. crassa* was a useful model organism for genetic studies (Rowland, 2000).

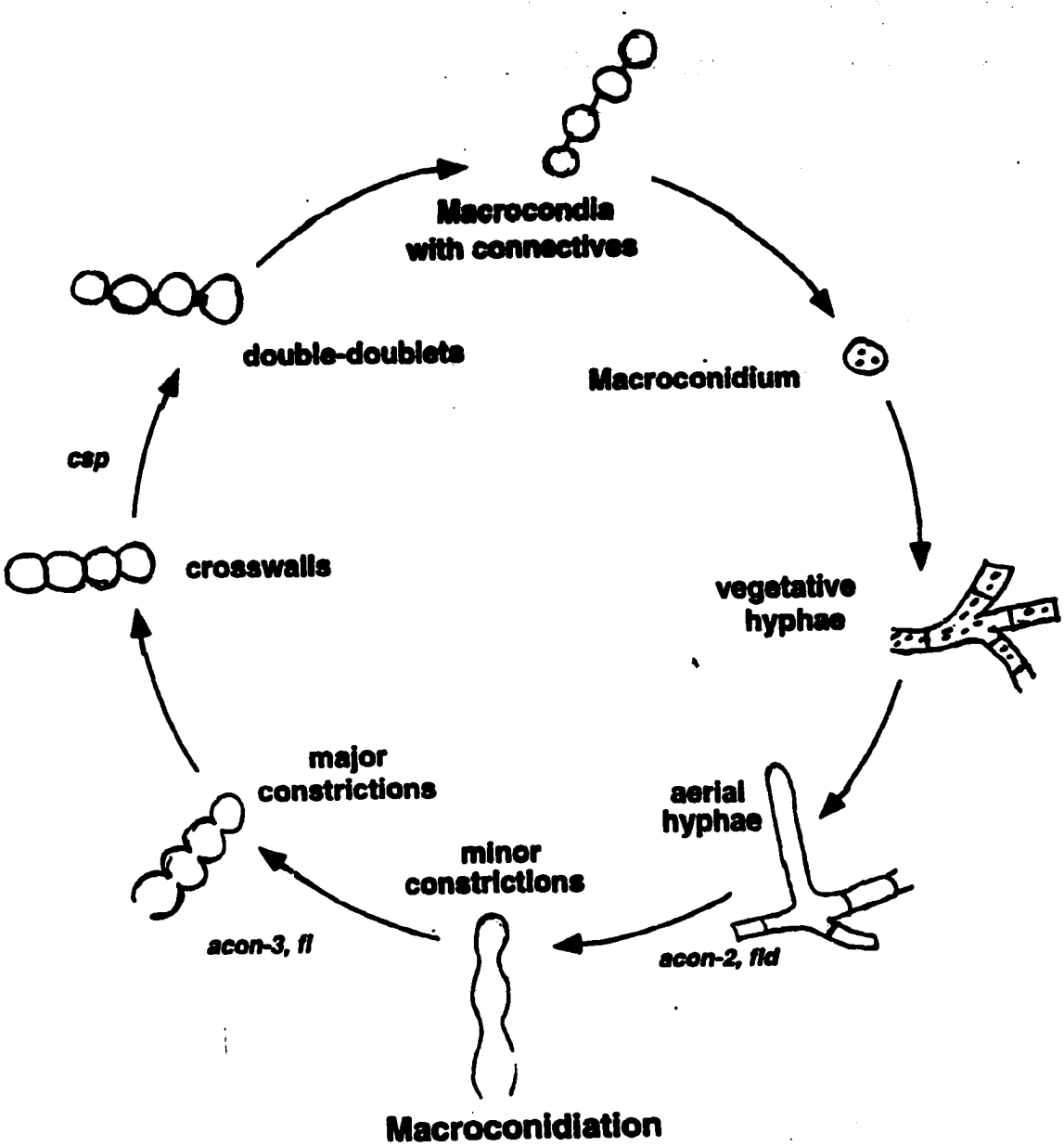
1.1.1 Introduction to *Neurospora crassa*

Neurospora crassa is a filamentous fungus. It belongs to the order Sordariales, subdivision ascomycotina (Ascomycetes) (Carlile and Watkinson, 1996). It is a type of multinucleate fungus. It also is a simple eukaryote. Generally speaking, filamentous fungi grow in the form of branched, thread-like cells called hyphae. During their growing period, the hyphae fuse and grow together as a vegetative mycelium. The typical haploid and multinucleate hyphae of filamentous fungi contain perforated crosswalls along their length that allow the cytoplasm, and even the nuclei, to travel from one compartment to the next. In nature there are many types of filamentous fungi (Turia and Blanchi,1972;

Lakin-Thomas et al., 1990), the majority of which have similar vegetative hyphae forms. However, because the sporulation mechanisms of different fungi are quite diverse and fascinating, these organisms are of interest to many mycologists, geneticists, and developmental biologists (Springer, 1993). *N. crassa* is an orange bread mold, which, more than 100 years after its identification, led Beadle and Tatum to perform their famous experiments on this fungus that led to their one-gene one-enzyme hypothesis, a milestone of modern biology (Beadle and Tatum, 1941).

1.1.2 Life cycle of *Neurospora crassa*

The ascomycete *Neurospora crassa* has three different sporulation pathways, which lead to the formation of macroconidia, microconidia, or ascospores (Springer, 1993). The macroconidia and microconidia result from asexual reproductive pathways but the ascospores result from the sexual pathway. Different environmental conditions lead to the different pathways. For example, the asexual sporulation process of macroconidiation is induced by desiccation or carbon deprivation (Mishra, 1991). Here, three cell structures, vegetative hyphae, aerial hyphae, and asexual spore conidia, are involved in the asexual reproduction of this ascomycete fungus (**Figure1.01**). During the macroconidia pathway, the aerial hyphae grow away from the substrate and produce conidiophores and aerial structures. An aerial structure consists of chains. These chains grow by budding and ultimately break apart into bright orange multinucleate macroconidia (**Figure1.02**).



Macroconidia (conidia) germinate and grow as vegetative hyphae. Macroconidia initiates by the formation of aerial hyphae. Eventually the aerial hyphae delineate conidiate by the formation of minor constriction, followed by the formation of major constrictions and crosswalls

Figure1.01 The macroconidia asexual reproduction cycle of *Neurospora crassa*

The microconidia are formed asexually within the vegetative hyphae and rupture through the hyphae cell wall when they mature. They are smaller, uninucleate conidia. These spores germinate at a low rate and are grayish brown in color. The sexual sporulation pathway of *Neurospora crassa* is induced from the vegetative hyphae under the condition of nitrogen limitation (Turian and Bianchi, 1972; Springer, 1993). Since the production of asexual conidia is under control of the circadian clock (Aroson, Johnson and Dunlap, 1994) and the macroconidia pathway is most obvious because of its orange colored conidia, the macroconidia pathway is used to study the circadian rhythms (Aroson, Johnson and Dunlap, 1994).

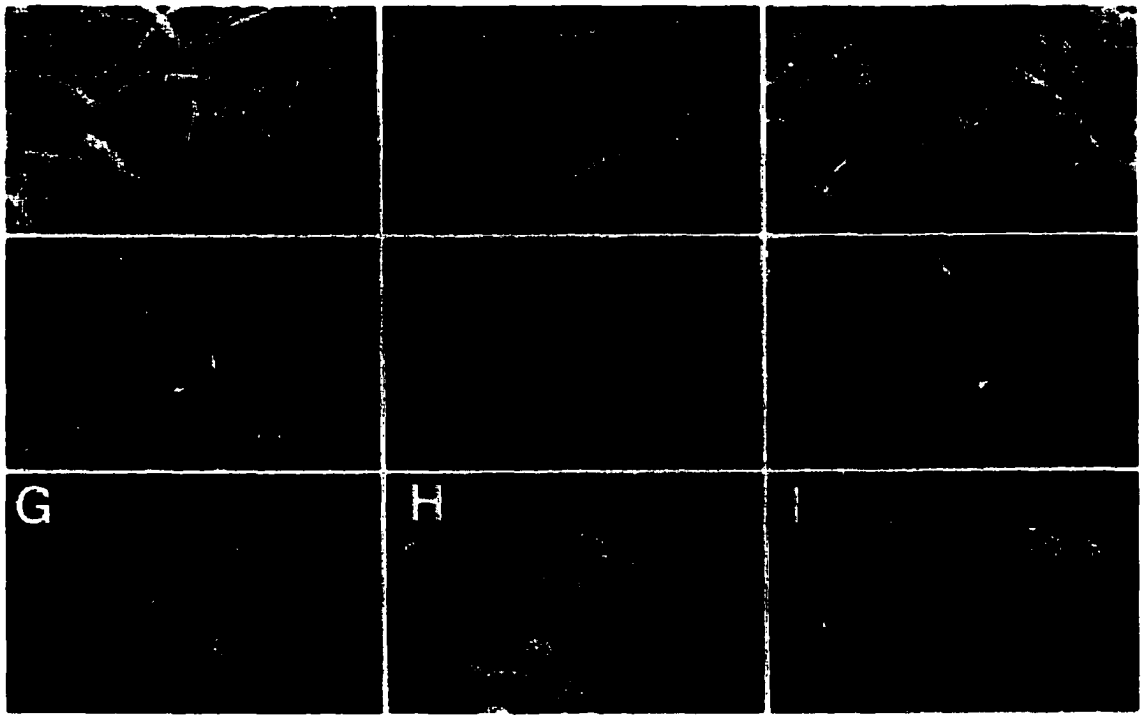


Figure 1.02 SEM photos of *Neurosproa crassa* hyphae and spores (Springer, 1993).

1.1.3 *Neurospora crassa* is an ideal model system for the study of circadian rhythms

Neurospora has been studied both genetically and biochemically. Its genome size is about 42.5 megabases (Radford and Parish, 1997; Dunlap, 1993; Dunlap, 1996), making it about 12 times that of *E. coli.* Ninety percent of *Neurospora*'s genomic sequence corresponds to non-repeated unique sequence while eight percent of the DNA contains ribosomal DNA repeats (Gurr et al., 1987; Selker, 1990). As is typical with eukaryotes, most of the genes have introns, but their size is relatively small, less than 500 nucleotides (nt), compared to other eukaryotic organisms but typical of fungal introns. *Neuropora crassa* has seven genetically identified chromosomes or linkage groups (Perkins, 1997). About 1000 map units are distributed over these seven cytological identified chromosomes and approximately 640 genetic loci identifying genes involved in the morphogenesis, development or metabolism of *Neurospora crassa* are mapped.

Neurospora has a very visible phenotype. The yellow-orange color of its macro conidias makes it very easy and convenient to observe the growth patterns of the yellow-orange colored conidia and the colorless vegetative mycelia (Lakin-Thomas et al., 1990). The switch of these two growth stages is under the control of *Neurospora*'s biological clock (Rosbash, 1995; Dunlap et al., 1996). *Neurospora* undergoes a daily cycle of differentiation. During this daily cycle, the clock controls the developmental switch of *Neurospora crassa.*

The circadian time, abbreviated CT, is used to measure the phase of a free-running rhythm. The unit of the circadian time is the circadian hour, equal to 1/24 the free-run period. The time from a light-dark cycle or from constant light to constant darkness is defined as CT12. In a free-running cycle, the period from CT0 to CT12 is the

subjective day. And the period from CT12 to CT24 is the subjective night (Bell-Pedersen et al., 1996b; Luo, Loros and Dunlap, 1998).

The *Neurospora* clock assays are typically carried out in a race tube (Dunlap, 1990; Bell-Pedersen, 1998; Bell-Pedersen et al., 1996a). A race tube is a hollow glass tube, which is about 10 cm long and 16 mm in diameter (**Figure1.03**). Both ends of this tube are bent upwards to keep the agar growth media in the tube. After inoculation and growth for a day in constant light, the position of the growth front of *Neurospora* is marked. Then, the culture is transferred to constant darkness. This transfer is called LD (light dark) transfer. It sets the clock running from CT12 and sets a developmental switch such that mycelia, as they are laid down, will not differentiate. Sometime later, at a time corresponding to late subjective night, the clock-controlled switch is thrown the other way so that the mycelia differentiate as they are laid down, and thus produce the aerial hyphae that eventually differentiate to produce asexual spores called conidia. It typically take one to several days for spores to appear (Loros et al., 1989; Dunlap, 1990).

Neurospora crassa is an ideal model for the molecular and genetic study of circadian rhythms because of its unique properties (Edmunds, 1988). During the vegetative growth on a surface (the late subjective night and early morning), the aerial hyphae emerge from the surface growth and give rise to asexual spores. At the other time of day, the aerial hyphae disappear and are no longer produced, and only the mycelia permanently remain as undifferential surface growth (Merrow and Dunlap, 1994). This cycle recurs every 21.6 hours in a wild-type strain at 25 ^{0}C. *Neurospora* is easy to cultivate on laboratory media using simple defined nutrients. The size of the haploid genome is 2-4$\times10^{7}$ bp, only one order of magnitude larger than that of bacteria (3.5$\times10^{6}$

bp), but its chromosome structure is like that of higher eukaryotes. It has seven chromosomes in the haploid form associated with histones. However, the *Neurospora* genome does not have the complex repeat sequences common in many higher eukaryotes (Gurr et al., 1987; Perkins, 1997). Taken together, the small genome size, obvious morphological changes, and low complexity of the genome make *Neurospora crassa* an ideal candidate to study the genetic and molecular components of the circadian clock (Bell-Pederson, Shinohara, Loros, Dunlap, 1996).

Figure1.03 The race tube used for measurement of the circadian clock exhibited in *Neurospora crassa*

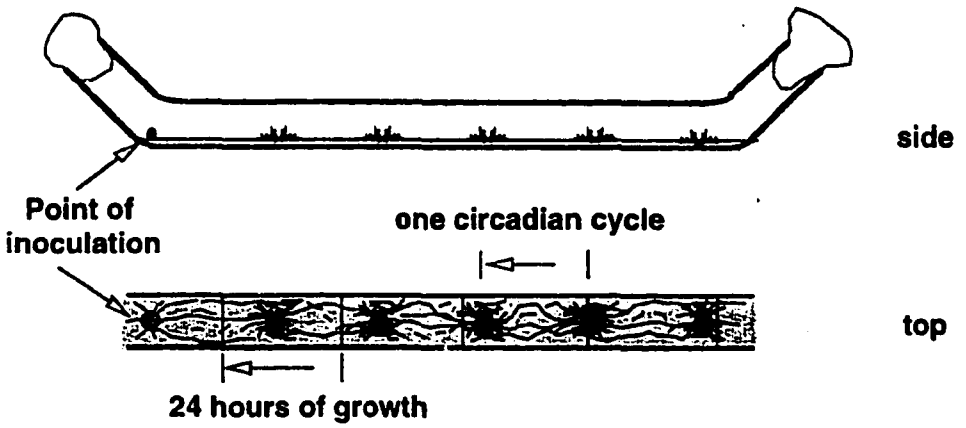


Figure1.03 The race tube used for measurement of the circadian clock exhibited in *Neurospora crassa*

1.1.4 Concept and characteristics of circadian rhythms

The term circadian comes from *circa diem*, meaning "about a day." The circadian rhythms are biological rhythms with innate periodicity of roughly one day. They are widespread throughout nature (Lakin-Thomas, Cote and Brody, 1990).

Most living organisms exhibit circadian rhythms, which allow them to adapt to an environment that varies with a periodicity of 24 hours. Such rhythms are ubiquitous and

govern so many key physiological functions that they have become synonymous with biological clocks (Gillette, 1997; Roenneberg and Merrow, 1999).

These rhythms are widespread through the plant and animal kingdoms, eucaryotic microorganisms, and the prokaryotes. Many biological phenomena such as sexual and asexual reproduction, bioluminescence, hormonal levels, and activity/rest cycles are under control of the circadian rhythms. The activities ranging from "gating" of cell division to complex behavioral changes within the metabolism of cells and organisms is also a manifestation of this rhythmicity (Roenneberg and Merrow, 1999).

Rhythms occur at all levels of biological organization, from unicellular to multicellular organisms, with periods ranging from fractions of a second to years. In humans, the cardiac and respiratory functions, sleep-wake cycles and nutrition cycles remind us every day of the fundamental physiological role of circadian rhythms in allowing us to cope with our periodically changing environment (McWatters, Dunlap and Millar, 1999). It is known that in humans circadian rhythms governed by the SCN (suprachiasmatic nucleus) affect a large number of physiological functions (Silver et al., 1999).

Many plant processes, including enzyme synthesis and activity, photosynthetic capacity, cell division, stomatal opening, flower opening, and odor production, are regulated in a circadian fashion. These rhythms are observed in species from unicellular to multi-cellular higher plants (Murtas and Millar, 2000).

The circadian clock regulates the process of photosynthetic carbon assimilation in plants (Milliar and Key, 1997; Sugano et al., 1998). Ribulose biphosphate carboxylase/oxygenase (RuBisCO) is a key element in the regulation of photosynthesis

because it catalyzes either the carboxylation or the oxygenation of ribulose bisphosphate, the initial steps in photosynthetic carbon assimilation, and photorespiration, respectively. In higher plants, rubisco is a heteromultimeric protein composed of eight large and eight small subunits. The *rbcL* gene, which encodes the large subunit, is part of the chloroplast genome, while *rbcS*, which encodes the small subunit, is nuclear-encoded. A number of studies have demonstrated that the circadian clock regulates the *rbcS* gene family (Giuliano et al., 1988).

Most organisms share the basic properties of all circadian rhythms (Dunlap, 1993). For example, circadian rhythms are an endogenous and self-sustaining process that is close to a 24-hour period. They have the property of temperature compensation, i.e. the period of the rhythms varies little with temperature and the pulse of light or temperature can change the phase of the rhythms.

In general, at least two elements are found in all circadian models: an input pathway and an output pathway. Through the input pathway, the environmental information is conveyed to the circadian pacemaker. Through the output pathway, the organisms or cells pass the timing information to drive cell activities in a rhythmic manner (Takahashi and Kornhayser, 1993). There most likely are hundreds of different output pathways controlled or regulated by biological clocks. Each organism has its own specific circadian clock system adapted to the specific needs of that organism (Dunlap, 1996). **Figure 1.04** shows the common element in the circadian oscillator of *Neurospora crassa*.

The mechanism of circadian rhythm is a biological sensory apparatus. Light or temperature (input) is a signal from outside (environment). This apparatus senses, accepts

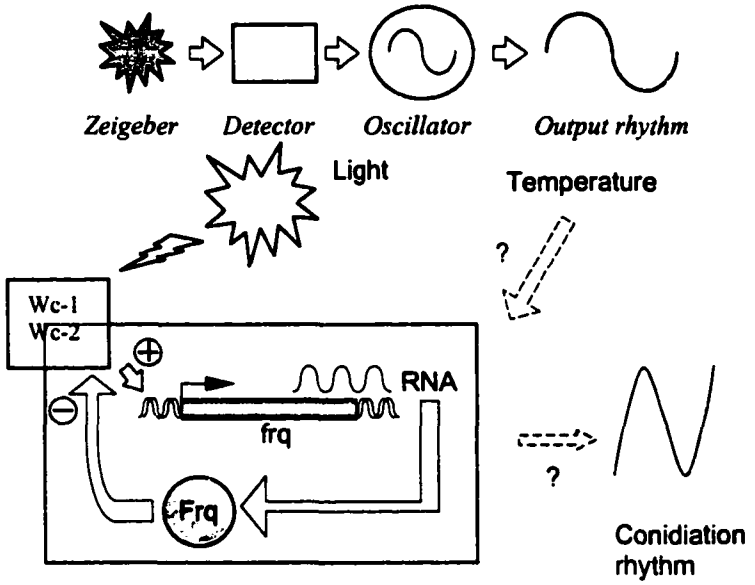


Figure1.04 Diagram showing the elements of oscillator and circadian rhythm in *Neurospora crassa* (adapted from Dunlap, 1999)

and transduces these signals into oscillators and then generates an output. An oscillator is a mechanism that generates the rhythmicity. In the cell, timing information is supplied to other processes through a circadian oscillator. Therefore, these processes are driven in a rhythmic way (Dunlap et al, 1996). Both a negative element and a positive element are required for oscillation. The product of a negative element feeds back to slow down the rate of the process itself while the execution of the feedback needs a delay. A positive element is necessary for a biological oscillator to activate the oscillator and keep it from winding down. The regulatory loop used by circadian oscillator is closed within cells. It does not require the interactions between cells. Inside that loop, the transcription of clock genes yield clock proteins (negative elements) that act in some way to block the action of positive elements. The role of positive elements is to activate the clock genes.

The circadian rhythms affect the physiological activities of organisms very pervasively. Therefore, the mechanisms about how the biological clock is generated and how the behavior of the organisms is controlled by the biological clock in response to the

changes in the environment have been a topic of study by biologists and geneticists for many years (Dunlap, 1993). Circadian rhythms remain fascinating because the fundamental mechanism of how circadian oscillations are generated, i.e. how the molecular time is generated and transduced into the cell remain unknown (Dunlap, 1999).

Clock genetics research has been very active in the past two decades. Genetic mutational analysis is commonly used to identify genes by comparing the phenotype of mutant strains to wild type strains. Comparison between mutations that have similar effects on the clock and a series of mutations having quite different phenotypes has uncovered some of the genes that are responsible for or related to the circadian clock (Dunlap and Feldman, 1988; Berlin and Yanofsky, 1985a, 1985b). Genetics analysis has been an effective approach to identify the clock-controlled genes and their products. After 20 years of study, the dominant question about the mechanism of circadian rhythms has switched from how does the clock work on the cellular level to how does organisms keep time on molecular level (Dunlap, 1999).

One way to address this issue is the characterization, identification, and analysis of genes and their protein products that are transcriptionally or transitionally regulated by the circadian clock is a very useful approach (Dunlap, 1998).

1.1.5 Studies of the clock-controlled genes in *Neurospora crassa*

The term "ccg" (clock-controlled gene) represents the genes whose transcription is regulated in a circadian way in *Neurospora* (Dunlap, 1999). Before the *ccgs* are discussed, it is really necessary to know what the clock gene and the clock-controlled gene (*ccg*) are and the difference between them.

A clock gene is one that encodes central components of an oscillatory loop within the circadian clock (Dunlap et al., 1995; Dunlap, 1999). The mutation of a clock gene will result in the loss of circadian rhythm. In *Neurospora*, *frq* is the clock gene.

From the reports of recent studies, the *frq* gene and its products are linked through the white-collar (WC) proteins, the products of two *white-collar* (*WC*) genes (Ballario and Macino, 1997; Crosthwaite, Dunlap, and Loros, 1997; McWatters, Dunlap and Millar, 1999; Dunlap, 1999). White-collar protein acts as transcriptional activators of *frq*. As we know, the FRQ proteins inhibit *frq* activation, which makes a negative feedback loop that drives a rhythm in *frq* RNA levels. The WC proteins mediate the rapid activation of *frq* transcription by light, which accounts for the entraining effects of light signals on the *Neurospora* circadian clock (Schwerdtfeger and Linden, 2000) (Figure1.05).

The clock-controlled genes are output regulatory targets of the oscillator and their transcription is rhythmically modulated. However, when *ccgs* are inactivated or mutated, the progress of the clock and the function of oscillator is not affected (Loros et al., 1989). This is the difference between the clock gene and the clock-controlled genes. The clock-controlled genes have been studied and identified in a wide range of organisms, that include the *Cyanobacteria* (Golden et al., 1997), plants (Carre & Key, 1995; Anderson & Key, 1997), *Drosophila* (Van Gelder & Krasnow, 1996) and vertebrates (Green & Besharse, 1996). To date, 11 *ccgs* have been identified (Loros et al., 1989; Bell-Pedersen et al., 1996c) and additional 4 new *ccgs* were identified in this present research (Zhu et al., 2001). Table1.01 is the summary of the 11 identified *ccgs* (Bell-Pedersen, 1998).

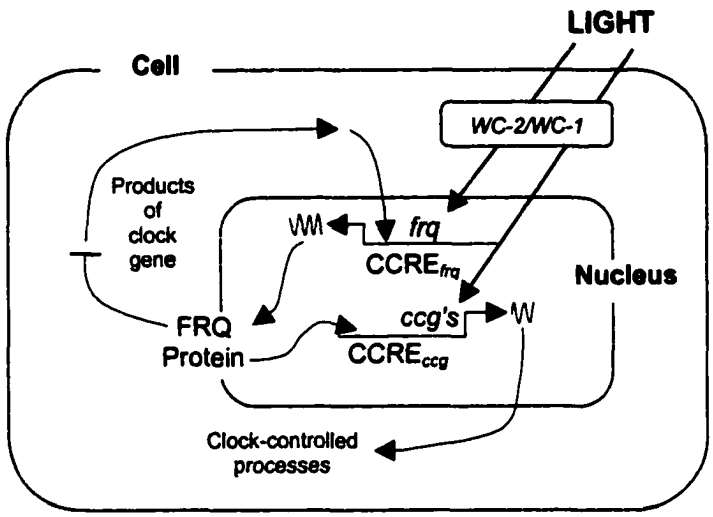


Figure1.05 Putative mechanism of circadian rhythm in *Neurospora crassa* (from Dunlap, cell 96:271-290,1998); CCRE: circadian clock responsive element.

Table1.01 Summary of the ccgs of *Neurospora crassa* (adapted from Bell-Pedersen1998)

Gene	Identity	Regulation		Reference
		Developmental	Light	
ccg-1	unknown	+	+	Loros et al., 1989
ccg-2(eas)	Hydrophobin	+	+	Loros et al., 1989; Bell-Pedersen et al., 1992
ccg-4	Pheromone precursor (new)	+	+	Bell-Pedersen et al., 1996
ccg-6	unknown	+	+	Bell-Pedersen et al., 1996
ccg-7	GAPDH	-	-	Bell-Pedersen et al.,1996; Shinohara et al.,1998
ccg-8	unknown	-	-	Bell-Pedersen et al., 1996
ccg-9	unknown	+	+	Bell-Pedersen et al., 1996
ccg-12(cmt)	CuMT	-	-	Bell-Pedersen et al.,1996;
al-3	GGPPS	+	+	Arpaia et al., 1995
con-6	unknown	+	+	Lauter & Yanofsky, 1993
con-10	unknown	+	+	Lauter & Yanofky, 1993

GAPDH, glyceraldehyde-3-phosphate dehydrogenase

CuMT, copper metallothionein

GGPPS, geranylgeranyl pyrophosphate synthase

1.2 Expressed sequenced tags (ESTs) and their significance in genome sciences

Expressed sequenced tags (ESTs) are the partial sequences generated from the ends of cDNAs. Since the 1990's, ESTs from many different organisms have provided extremely useful information relating gene expression and biological phenotype.

1.2.1 DNA and RNA

DNA, typically a double helix of deoxyribonucleic acid (Watson and Crick, 1953), is the genetic material of all known organisms and many viruses. RNA is a polynucleotide chain of ribonucleic acid. Its chemical formula is slightly different from that of DNA. DNA consists of Adenine (A), Guanine (G), Cytosine (C), Thymine (T) and the deoxyribose sugar. The nucleotide components in RNA are Adenine (A), Uracil (U), Guanine (G) and Cytosine (C) and the ribose sugar. Only some viruses use RNA as genetic materials. The central dogma relates the relationship between DNA, RNA, and protein (**Figure1.06**). It also states that information in nucleic acid (DNA and RNA) can be perpetuated or transferred but the transfer of information into protein is irreversible (Lewin, B., 2000).

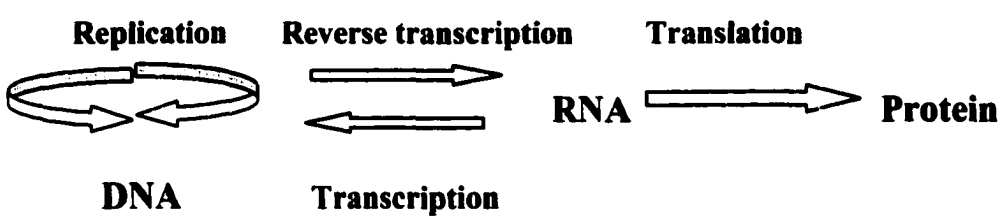


Figure 1.06 Central dogma of molecular biology

1.2.2 Gene

A gene is a region of DNA which is transcribed. There are three classes of genes in eukcaryotes (Singer and Berg, 1991). Genes that encode most ribosomal RNAs (5.8s, 18s, and 28s rRNA) are called class I genes. The enzyme for transcription of class one genes is RNA polymerase I. Genes that encode transfer RNA (tRNAs), 5s rRNA and other small cytoplasmic RNAs (scRNAs) are called class III genes. RNA polymerase III is responsible for the transcription of class III genes. Both of these classes of genes belong to structural genes. Another class of genes are class II genes or functional genes. These genes encode all the cytoplasmic messenger RNAs (mRNAs) and all but one of the U RNAs (U6) in snRNPs. RNA polymerase II performs the transcription of these genes. The structure of a typical eukaryotic gene is shown in **Figure 1.07**.

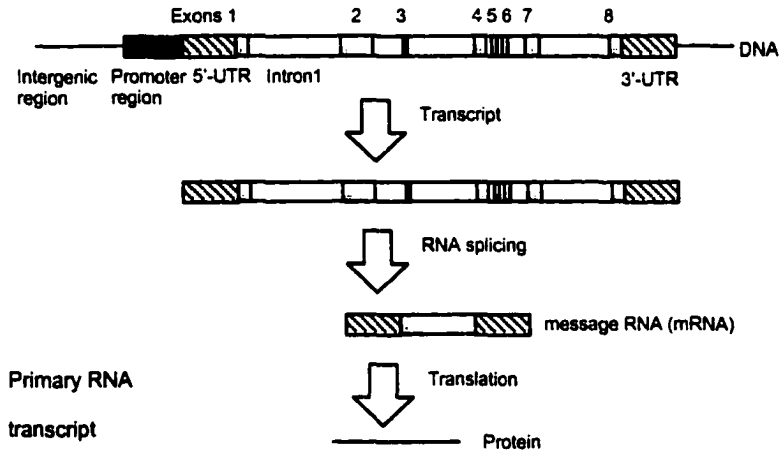


Figure 1.07 Typical structure of an eukaryotic gene and its expression

Generally, eukaryotic gene includes the following DNA segments: 1.) one or more transcription units which are contiguous sequences of DNA which are responsible for encoding the primary transcript. They include the coding sequences of both mRNA and protein product, the introns, the 5' UTRs, and the 3' UTRs; 2.) The promoter that is the minimal sequence needed to initiate correct transcription and the terminator that is required to terminate the transcription at a proper position; 3.) The regulator sequence

element such as enhancers and silencers. These factors regulate the rate of transcription initiation and influence transcription initiation from a distance.

1.2.3 Transcription and mRNA

Messenger RNA has a very important role in the generation of EST sequences, because cDNA is the complementary DNA of mRNA. The synthesis and structure of mRNA in prokaryotic and eukaryotic cells is very different (Stryers, 1995). In bacteria, transcription and translation are coupled and processed in the same single cellular compartment. Prokaryotic mRNA is relatively less stable than eukaryotic mRNA as it usually is translated into proteins while it is transcribed or with in a few minutes after RNA synthesis. In eukaryotic cells, the transcription and translation of mRNA are uncoupled. The synthesis and maturation of mRNA occur in the nucleus, but the translation takes place in the cytoplasm. mRNA is exported to the cytoplasm from the nucleus after it is mature. Eukaryotic mRNA is stable compared to prokaryotic mRNA. In some cases, it can be continually translated for up to several hours (Stryer, 1995).

All mRNAs consist of a coding region and noncoding region (**Figure1.08**). The coding region is a series of codons representing the amino acid sequence of the protein, start codon (AUG) and stop codons (UAA, UAG, UGA). The noncoding region includes the 5' untranslated region (leader sequence) and 3' untranslated region (tailer sequence) which do encode for protein. The primary mRNA transcripts, also termed precursor mRNA, pre-mRNA, hnRNA-heterogenous nuclear mRNA, need to go through a series of modifications to mature. These include splicing, addition of 5' CAPs and 3' polyadenylation.



Figure1.08 The typical structure of mRNA

The primary mRNA transcript contains both intron and exon sequences. Since the intron must be removed prior to producing a mature mRNA, a spliceosome catalyzes the splicing of mRNA precursors. The spliceosome is an assembly of small nuclear ribonucleoprotein (snRNPs) with mRNA precursors (Wolin and Walther, 1991).

The sequences at the ends of introns specify the splice sites in mRNA precursors. A common structural motif exists in the base sequences of thousands of intron-exon junctions of RNA transcripts in eukaryotes from yeast to mammals. This motif is the base sequences of an intron beginning with GU (GT) and ending with AG (**Figure1.09**).

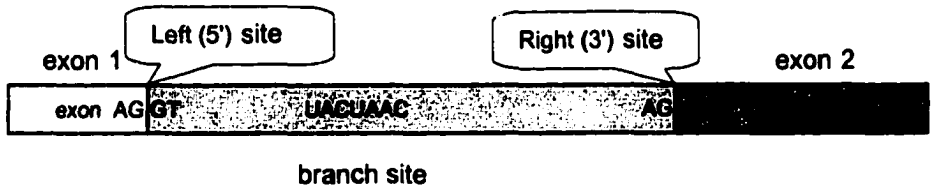


Figure1.09 The GT-AG pair at the ends of a nuclear intron sequence

The 5' splice site, 3' splice site and the branch site are necessary for correct splicing. Mutations in any of these three critical regions will lead to abnormal splicing (Sharp, 1985).

During transcription, mRNA precursors add a 5' CAP structure to protect from 5' exonuclease, phosphatases, and nucleases (Gallie, 1991). The 5' CAP structure is an unusual 5'-5' triphosphate linkage between the α phosphorus atom of GTP and the diphosphate on the 5' end of the nascent RNA chain, which results from a hydrolysis

reaction, in which a phosphate was hydrolyzed from the 5' triphosphate end of newly transcribed RNA chain.

A polyadenylation (poly(A)) tail is added to most mRNAs after the mRNA precursors were cleaved by an endonuclease. An internal AAUAAA sequence serves as the cleavage site. Roughly 250 A (adenine) residues (donated by ATP) are added to the 3' end of mRNA by Poly (A) polymerase (PAP). The poly (A) tail is required for the export of mRNAs from the nucleus to the cytoplasm and it affects the efficiency of protein synthesis. mRNA with a poly (A) tail is more effective as a translation template and the tail protects mRNA from digestion by nucleases (Jackson and Standart, 1990; Sachs and Wahle, 1993).

1.2.4 cDNA and expressed sequence tags

From the central dogma, we already know that selected portions of the DNA nucleotide sequence are copied into a corresponding RNA nucleotide sequence that eventually is processed to a form which either encodes a protein (mRNA), forms a transfer RNA (tRNA) or ribosomal RNA (rRNA) molecule. A cDNA library is a collection of cDNA clones each derived from the mRNA or present when the cDNA (complementary DNA) copy of the mRNA is synthesized by an RNA-dependent DNA polymerase (reverse transcriptase). Therefore, a cDNA clone only contains that region of the genome that has been processed into mRNA. Because the cells of different tissues produce distinct sets of mRNA molecules, cDNA libraries obtained from different types of cells used to prepare the library will be different. In contrast, a genomic DNA library is an entire collection of recombinant plasmids, which contains the genomic DNA insert

that was prepared by enzymatic digestion or physical shearing. Because the genomic DNA is cut into many small fragments, only some pieces will contain gene(s). Genomic clones represent all of the DNA sequences in an organism. Therefore, except in the case of immune cells, all of the DNA sequences in an eukaryotic organism will be the same regardless of the cell type used to prepare them.

A DNA sequence provides the primary identification of gene structure. Since one mRNA species represents one gene, the distribution of mRNA species in a cell provides key information for the analysis of the physiological activity of that cell. Automated sequencing technology with randomly selected cDNA clones from libraries that represent mRNA distributions makes it possible to construct a very detailed picture of the transcriptional activity of a cell or tissue. It includes not only the identification but also the abundance level of transcribed genes, and ultimately the degree of overlap in gene expression among various tissues (Boguski and Tolstoshev, 1994). In practice, partial cDNA sequences or sequencing the expressed sequence tags (ESTs) is one useful way to obtain gene-specific sequence data for a large number of independent cDNA clones (Boguski and Schuler, 1995; Bargmen, 1992).

First, what is an EST? An EST is an *Expressed Sequence Tag* that is obtained by a single-pass sequence generated from either end of a randomly selected cDNA clone. A cDNA is the complementary DNA of an mRNA synthesized by reverse transcriptase. The EST strategy was developed to allow rapid identification of expressed genes by sequence analysis and generating unique sequence tags for each clone in a cDNA library.

Large scale EST sequencing was first championed by Craig Venter of the Institute for Genomic Research Center in Gaithersburg, Maryland, in 1991(Boguski and Schuler,

1995). Since then, the amount of EST data has increased dramatically in the dbEST section of Genbank with ESTs being submitted at the rate of 1,500 sequences per day (Boguski, 1995). In 1995, more than 150,000 sequences were submitted to GenBank. These sequences accounted for more than half the sequences recorded in Genbank that year. These data were utilized by researchers so heavily that the EST division of GenBank had to process more than 150,000 queries consisting of homology searches, e-mail retrievals, anonymous ftp file transfers and accesses via the World Wide Web everyday (Boguski, 1995). In 2000, the dbEST of GenBank released 6,299,038 DNA sequences from 301 different organisms. The top four organisms were human (50%), mice (36%), cattle (3%), soybean (2%).

In an attempt to assess the usefulness of ESTs for gene discovery, Boguski and Tolstoshev (1994) cloned 32 human disease genes by either the positional cloning or positional candidate methods. The positional candidate approach is a method that was developed to identify new genes (Collins, 1992; Ballabio, 1993). Positional cloning is a method that locates the responsible gene solely on the basis of map position. It is opposite to the method of functional cloning, which identifies the gene causing a human disease based on the fundamental information of the basic biochemical defect, without reference to chromosomal map position (Collins, 1995). They performed sequence homology searching against dbEST. The results showed 38% of these human genes had exact and often multiple matches in dbEST, and 47% of them had homologies in other organisms. Only five of the 32 human disease genes had no convincing matches with ESTs. A conclusion was made that 85% of these 32 human disease genes have homologous partial cDNA sequences in the public domain (Boguski and Tolstoshev, 1994).

With the rapid expansion of EST data, the strategies used by molecular geneticists for identifying and cloning novel genes was changed. It became apparent to most biologists that even though this type of data is incomplete and inaccurate, it is rapidly and inexpensively obtained and it is quite useful (Sikela and Auffray, 1993; Bargmann, 1992). For example, ESTs can be used for discovering new members of gene families involved in human disease (Brody, L. C. *et al.*, 1995), for the identification of exons in vast expanses of genomic DNA (Boguski and Schuler, 1995). ESTs produce very important gene markers that can be used in genome physical maps (Link and Olson,1991), and in producing cDNA microarray (Duggan et al., 1999; Khan et al., 1999; Gerhold, Rushmore and Caskey, 1999).

Expressed sequence tags (ESTs) represent the largest amount of information possible per raw base sequenced. Compared to highly redundant contigs of overlapping sequences, single-pass sequencing is less accurate. But it is accurate enough for very sensitive similarity searches (Banfi, Guffani and Borsani, 1998). Without redundancy, a large number of independent clones can be analyzed, rather than analyzing tens to hundreds of overlapping clones to determine the sequence of a single gene. Although, an initially reported (Martin-Gallardo *et al.*, 1992), over 3000 sequencing reactions are needed to sequence a 106,000 base genomic region, which on average would contain 5 genes. Over 2000 different genes would be identified if the same number of sequencing reactions were applied to independent cDNA clones.

1.2.4.1 The application of ESTs in the human genome project

The purpose of the human genome project is to obtain and analyze the genetic blueprint of our genome and that of selected model organisms as well as to elucidate the mechanism of control of gene expression (Adams et al., 1995). Although most sequencing efforts currently are directed toward elucidating the physical structure of the genome, functional analyses of a genome are needed if gene expression and its control is to be understood. The goal of gene expression analyses is not only to uncover which genes are expressed, but also to determine to what extent they are expressed in any given cell at a given time. When these data are obtained from different cells and tissues at various stages of differentiation or under different physiological conditions, a "body map of expressed human genes" can be constructed (Boguski, 1995). Collecting cDNA sequences and obtaining the statistics of their frequencies in a particular cell type is the first step toward this goal. cDNA sequences also will provide markers for mapping genes along with chromosomes and reference sequences for identifying genes along the DNA. These ESTs will complement the physical analysis of the genome by providing direct data of biologically and maybe even industrially interesting gene products (McCombie et al., 1992; Waterson et al., 1992).

ESTs have several applications in human genome project. First, the ESTs can be used as a gene marker in the physical mapping of human genome. For example, when the master set of the cDNAs expressed in human brain was generated, when the precise chromosomal location of each cDNA and the complete nucleotide sequence of the coding region of each expressed human brain gene is known, a tremendous resource to human

biology and neuroscience will be available (Wilcox et al., 1991; Khan et al., 1992). Second, the single pass sequence of the 3' untranslated region (3' UTR) of human brain cDNAs was used to identify specific genes and to develop gene-based sequence-tagged-sites (STSs) which, in turn, was used to assign cDNAs to specific human chromosomes (Matoba, 1994; Berry et al., 1995). Third, the 3' UTR sequence can be used to design PCR primers in the application of a general PCR-based gene mapping strategy (Wilcox et al., 1991). Primers designed according to 3' UTR sequences have several important advantages. First, the 3' UTR almost always lacks introns. Therefore, the cDNA sequence is identical to the corresponding genomic sequence. When 3' UTR primers are used to amplify a single product from genomic DNA, this product will be the same size as that predicted from the cDNA. Each primer set then defines an sequence-tagged-site (STS). In addition, when compared to STSs obtained from regions of the genome for which no function is known, STSs derived from 3'UTR regions would have the advantage of being known to be part of expressed genes, facilitating the development of STSs. Second, sequences derived from the 3' UTR are unique for each gene. So it represents a common reference point for gene-to-gene comparison as compared to the 5' sequence which is not so gene-specific as the 3'UTR sequence. It allows discrimination among genes that are within the same gene family or that share conserved functional motifs. Human cDNAs can provide an abundant source of gene-associated markers for genetic mapping and demonstrate how such polymorphic cDNAs can be converted to highly informative genetic markers. By combining knowledge of the map position of a disease locus with the assessment of cloned candidate gene sequences, either cDNAs or

ESTs that map to the same chromosome region, disease genes can be efficiently identified (Sikela, and Auffray, 1993; Boguski and Tolstoshev, 1994).

1.2.4.2 The accuracy of EST sequences

The accuracy of EST sequences is important in an EST project. Because the rapidly accumulated EST sequence data are obtained using only single sequencing runs on both ends of many different clones, if the redundancy of one gene is low, it means that ambiguous base calls will have no chance to be confirmed by a second reaction or by sequencing the other strand (Adams et al., 1993a). This is why EST data is not as accurate as the genomic DNA sequences. The strategy used to estimate the accuracy of the sequences is different. It depends on the different usage of the sequence data. If the sequence data is used for predicting translation products, sequencing errors should be minimized by multiple sequencing efforts. On the other hand, if the sequence data is used only for identifying mRNA species, handling a large number of samples is more important than the accuracy of the sequence data (Adams et al., 1991, 1992)

Several factors affect the accuracy of sequence data in different stages in practice. During the stage of generating and collecting EST data, the factors that affect the accuracy of EST data include the quality of the template DNA, the fidelity of the sequencing reaction, and the sequence-dependent characteristics such as poly (A) stretches and G/C compressions. During the stage of editing of EST data, consistent data editing procedures are preferred in attempt to get high accuracy value EST data. This will ensure that sequence ends are trimmed at comparable breakpoints between reliable and unreliable base calls.

1.2.4.3 Editing of EST data

EST editing is a process in which all the ESTs generated through large scale EST sequencing program go through a series of examines to select high quality ESTs before the analysis of EST data. This is necessary to acquire highly accurate EST data for customers. During this editing procedure of new ESTs, the polylinker sequence, sequence of low quality, vector sequence, ribosomal and mitochondrial sequences are removed. Then, the high quality EST data can be used to search against the public nucleotide or peptide databases (with six-frame translations) using the BLAST (Altschul, *et al.*, 1990) network server at the National Center for Biotechnology Information (NCBI), National Library of Medicine.

1.2.4.4 EST analysis and the classification of biological functions

The 3' region of mRNA is a unique nucleotide sequence that often is used as the signature to identify a gene. The 3'-end ESTs of a cDNA clone contains the 3' untranslated region and 3'-end of the coding region and is anchored in a fixed position during the first strand synthesis of cDNA. The 3' end of a mature eukaryotic mRNA contains a poly A tail. The poly (T) primer used in the synthesis of the first strand of double strand cDNA results in a fixed start position of 3' end EST sequence. All 3' ESTs from the same or similar mRNAs start at same position and they are characterized by their poly (T) sequence. This unique property of 3' end ESTs makes the 3' untranslated region of a cDNA clone gene specific, a feature that can be used as the signature to identify a gene. Gene expression profiles in cells can be obtained by studying the

frequency of the appearance of 3' ESTs in the cDNA library (Matoba et al., 1994). On the other hand, 3' EST (upstream of the poly (A) tail) provides a more constant reference point for clone-to-clone comparisons. Because of the paucity of introns in the 3'untranslated region, 3' ESTs were more useful for sequence tag site (STS) generation and polymerase chain reaction (PCR)-based mapping. Sequencing of 3' directed libraries is thought to provide more reliable estimates of the relative frequencies of the mRNA species present. Therefore, the assembly of 3' end ESTs can be used to accurately calculate the frequency of the clone in the library (Matoba et al., 1994) and to deduce the abundance of the mRNA in the library. Furthermore, comparison of EST frequencies between tissues permits a detailed examination of tissue specificity and the number of genes that are in fact ubiquitously expressed in these tissues (Okubo et al., 1991, 1995). In contrast, the 5' end EST is more likely to provide protein coding information and thus provide a hint as to functional class a cDNA may belong to.

The diversity and activity of expressed genes in a higher eukaryote, can be estimated by collecting ESTs from as many different cells and tissues as possible to form a possible complete collection of expressed genes (Okubo, and Matsubara, 1993). Similarly a knowledge of cell- and time-dependent control of gene expression also can be gained by comparing data from the same cell placed under different physiological conditions, or from two similar cells at different stages of differentiation (Okubo, et al., 1992). Generally speaking, random sequencing of about 1000 clones is sufficient to characterize the expression profiles of genes in a given cell type (Okubo et al., 1991, 1995). A significant fraction of middle and highly abundant cell-specific genes can be identified in this kind of sample of cDNA clones.

From the EST sequencing, relative levels of gene expression can be determined for a cell-line or tissue and compared to similar information from other tissues. However, to analyze gene expression on a large-scale, a system that focuses on describing the physiological activity or biochemical pathways in a tissue or cell is very helpful. Such a system was established by grouping genes by biochemical processes and their biological functions in cell (Gribskov, 1999) and several such categorization schemes are available currently. A system developed by Riley was used in the analysis of bacterial genome function (Riley and Labedan, 1997). Riley's system attempts to classify the biological functions of genes by grouping genes into energy metabolism, macromolecular synthesis, genetic information processes and several other categories of regulatory processes such as cell growth, cell division and cell processes. Since the annotations of EST sequences matched in GenBank or other databases do not give a sufficient explanation to classify the function of the encoded protein adequately, several other informational resources are used to help the classification of the functions of the encoded proteins of these ESTs or genes. Since the NCBI Entrez program which links the GenBank sequences and references, it is frequently used to obtain abstracts from published papers. The neighboring scheme of the Entrez program also allows related articles, books, or sequences to be retrieved. Additional information for adequate functional classification usually is obtained from the full text papers and text books (Adams et al., 1995; Riley and Labedan, 1997; Kupfer, 1999).

1.2.5 cDNA library construction

The selection and construction of the cDNA library are the key aspects of an EST project. In early studies, a cDNA library usually was constructed only to find one or a few particularly interesting gene(s). Since the idea of ESTs was introduced in 1991, cDNA libraries have been used to find a large number of the genes expressed in a particular cell or tissue (Adams, et al., 1991, 1992; Khan et al., 1992; McCombie et al., 1992; Okubo et al., 1992). In most cases, the cDNA library can be constructed from any organism, tissue, cell type, or developmental stage. Most of the time, a library consists of oligo-dT-primed and directional cloned cDNAs. This type of library permits 3' anchored sequencing. It means each cDNA can be sequenced from the extreme 3' end, and then permits accurate analysis of the mRNA distribution. The 5' end sequences of cDNA clones of a directional library are much more likely to reveal the coding sequence in the cDNA clone and therefore are especially valuable when studying fungal and higher vertebrate gene expression.

A directionally cloned cDNA library usually is primed with oligo-dT coupled to an adapter/linker (Stratagene). This also is the location of the 5'end and 3' ends of the cDNA clone to be known in a directionally cloned library. In practice, if the length of the poly (A) tail in a cDNA clone is less than 50 bp, it will be easier to sequence the 3' end. A poly (A) tail that is too long will stall the 3' end sequencing reaction and result in a failed reaction. Finally, all clones should in fact have poly (A) tails; if cDNA synthesis was not initiated at the oligo-dT portion, the library will not be anchored at the 3'end even though the library will still be directional.

Several different strategies can be used for constructing cDNA libraries and which strategy is used can greatly influence the resulting information (Moreno-Palanques and Fuldner, 1993). First, selecting a specific cell or organ to prepare cDNA library so that the constituents of the cDNA library can reflect the physiology of this cell or organ. Alternatively, the cDNA library can be prepared from several sub-libraries from different tissues to maximize the number of gene transcripts. Second, the library can be designed either to faithfully represent the abundance of gene transcripts in the original mRNA population or to represent a collection of non-overlapping transcripts. The latter is often called a single book library or normalized cDNA library. It can be used for surveying as many gene transcripts as possible (Patanjali, Parimoo, and Weissman, 1991).

Single-pass automated sequencing of randomly chosen clones from cDNA libraries has proven to be an efficient strategy for identifying unknown genes (Adams *et al.*, 1991, 1992). However, the quality of data generated in an EST study is greatly dependent upon the quality of the cDNA library. The ideal cDNA library should have the following features (Moreno-Palanques and Fuldner, 1993). First, it is representative. A cDNA library should contain all sequences present in the initial poly (A) population in the same relative frequencies. Second, it is unidirectional. The orientation of each cDNA should be known. Third, it contains a high proportion of long or full length cDNA inserts. Fourth, it is clean. No genomic DNA, mitochondrial or ribosomal RNA inserts contamination exist in the cDNA library. Finally, the shorter the poly (A) tail, the better the quality of a cDNA library.

1.2.5.1 mRNA in eucaryotes

Eukaryotic cells have several different RNA polymerases, each with a specialized function (Table 1.02).

Table 1.02 Comparative properties of eukaryotic RNA polymerase

Polymerase	Location	RNA Synthesized
I	Nucleus	pre-rRNA (except 5S)
II	Nucleus	pre-mRNA, small nuclear RNAs
III	Nucleus	pre-tRNA, 5S rRNA, other small RNAs

RNA polymerase II transcribes mRNA in the nucleus. mRNA produced in the nucleus in eukaryotic cells must be exported into the cytosol for translation. The coding regions or exons in eukaryotic genes are discontinuous as they are interrupted by introns. The initial product of transcription (pre-mRNA) includes all of the introns and substantial flanking regions. Before translation can occur in the cytoplasm, the pre-mRNA has to undergo mRNA processing, which involves 5'-capping, splicing and addition of a poly (A) tail. Only when the capping, poly (A) tailing and splicing are complete, can the newly processed mRNA be exported from the nucleus into the cytosol through the nuclear pores and then it can begin the translation process.

1.2.5.2 mRNA preparation

Guanidinium salts are used to isolate total RNA from unfractionated cells and tissues (Chirgwin et al., 1979; McDonald et al., 1987) since they rapidly inhibit endogenous ribonucleases (McDonald et al., 1987). RNA also can be isolated from subcellular fractions such as nuclei (Nevins, 1987), cytoplasm (Berger, 1987b), membrane bound or free polysomes (Mechler, 1987; Berger, 1987b). After the total RNA

is isolated, it is necessary to select poly $(A)^+$ RNA. Oligo (dT) cellulose matrix (Invitrogen) or oligo(dT)-coated magnetic beads (Jakobsen et al., 1990) can be used to isolated poly$(A)^+$ mRNA directly from tissues or cells with a commercially available kit with standard protocols. This method is based on pairing between the poly $(A)^+$ residues at the 3'-end of mRNAs and the oligo d(T) residues coupled to the solid support. mRNA with a poly (A) tail is bound and non-poly $(A)^+$ species are washed through. The bound mRNA is subsequently eluted and collected in a low salt buffer. Since this method does not need centrifugation or filtration steps, it is rapid and the risk of physical or enzymatic degradation of mRNA also is reduced.

The quality of all the mRNAs used to construct a cDNA library need to be checked. Electrophoresis fractionation of mRNA in a denaturing gel or northern blotting and hybridization with a probe of known size common to most cells, such as actin or glyceraldehyde 3-phosphate dehydrogenase, are commonly used to achieve this goal. A distinct band on the gel indicates that the mRNA, which is homologous to the probe without a large amount of degraded product, is intact (Berger, 1987a).

1.2.5.3 cDNA preparation

The first stage of the construction of a cDNA library is to convert polyadenylated mRNA to clonable ds cDNA. Oligo (dT) primers can bind to the 3' poly (A) tail and thereby potentially generate a full-length DNA copy (Krug & Berger, 1987).

The first enzyme used in cDNA construction is an RNA-dependent DNA Polymerase or reverse transcriptase. First discovered by Temin and Baltimore in 1970 in the virions of an RNA tumor virus (Temin and Baltimore, 1972), reverse transcriptase

can synthesize a DNA strand complementary to an RNA (cDNA) when a free 3'- OH group or a primer is provided. Two different reverse transcriptases, avian reverse transcriptase and murine reverse transcriptase that are isolated from avian retrovirus and murine leukemia virus, respectively, are commercially available (Sambrook et al., 1989). Both of these enzymes have several enzymatic activities including RNA-dependent synthesis of DNA, DNA-dependent synthesis of DNA, and the capacity to degrade RNA in DNA: RNA hybrids.

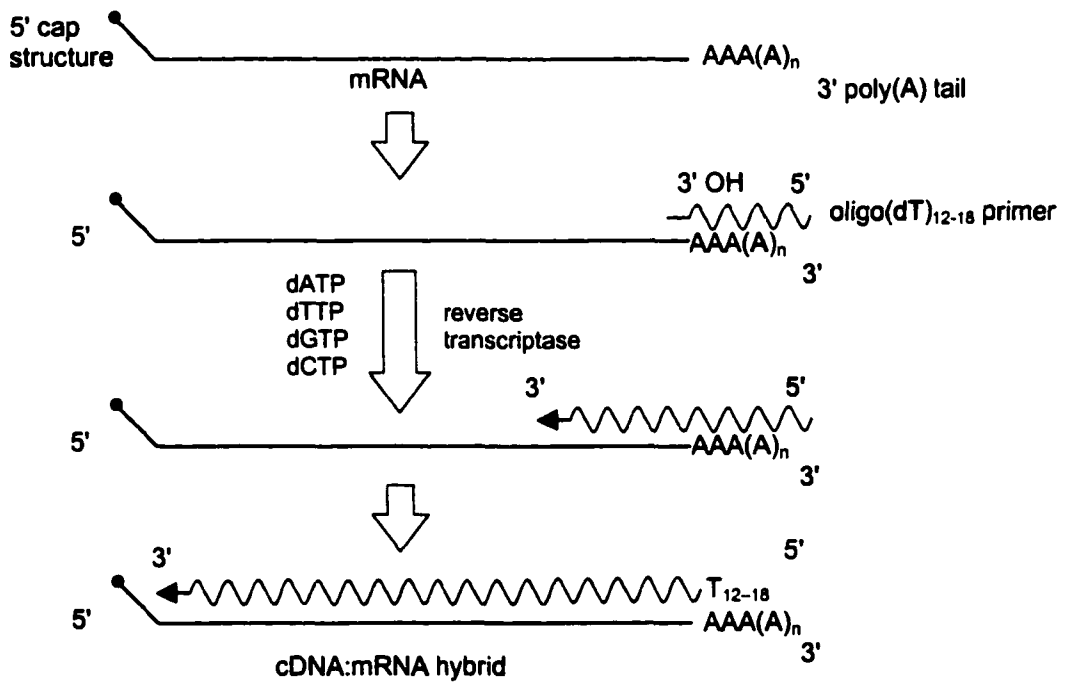


Figure1.10 Synthesis of the first strand of cDNA using an oligo(dT) primer and reverse transcriptase(Sambrook et al., 1989).

1.2.5.3.1 The synthesis of the first strand of cDNA

In first strand synthesis, dCTP is replaced by 5'-methyl dCTP to protect the DNA from restriction enzymes used later. In some protocols, strong denaturants or methyl mercury hydroxide is used to treat the mRNA solution to reduce the extensive secondary

structure. **Figure 1.10** shows the synthesis of the first strand of cDNA. mRNA is the template. A 12-18-base oligo (dT) with a 3' free hydroxyl terminus (OH) is a primer and it binds to the 3' poly(A) region of the mRNA template. When the mixture of dNTPs is added, the reverse transcriptase starts to synthesize the first-strand cDNA.

1.2.5.3.2 The synthesis of the second strand of cDNA and cloning

The second strand is synthesized by standard methods (**Figure1.11**). RNase H that nicks the RNA of the RNA-cDNA hybrid is needed for the efficient synthesis of the cDNA second strand. The transient hairpin loop at the 3'-end of the newly synthesized first strand can also be used as primer to generate the second strand cDNA after the RNA fragments are removed with alkali. After the mixture of dNTPs is added, the Klenow fragment of *E. coli* DNA polymerase I catalyzes the synthesis of the second strand of the cDNA. Then, the hairpin on the double-stranded cDNA is removed using nuclease S1. After the dscDNA is synthesized, the construct is blunt ended, size selected. The synthesized cDNA then usually are size-fractionated by Sephacryl columns, which removes inserts with less than 500 bp. One advantage of this method is that the cDNA is never exposed to a restriction enzyme. The size-selected cDNA is ligated to lambda vector arms and purified. After the lambda phage with cDNA inserts are packaged, the lambda cDNA primary library is established (**Figure 1.12**). Cloning can be carried out as needed, followed by a conversion to phagemid form by *in vivo* excision without prior amplification of the library. The resulting plasmid library is used to transform competent cells.

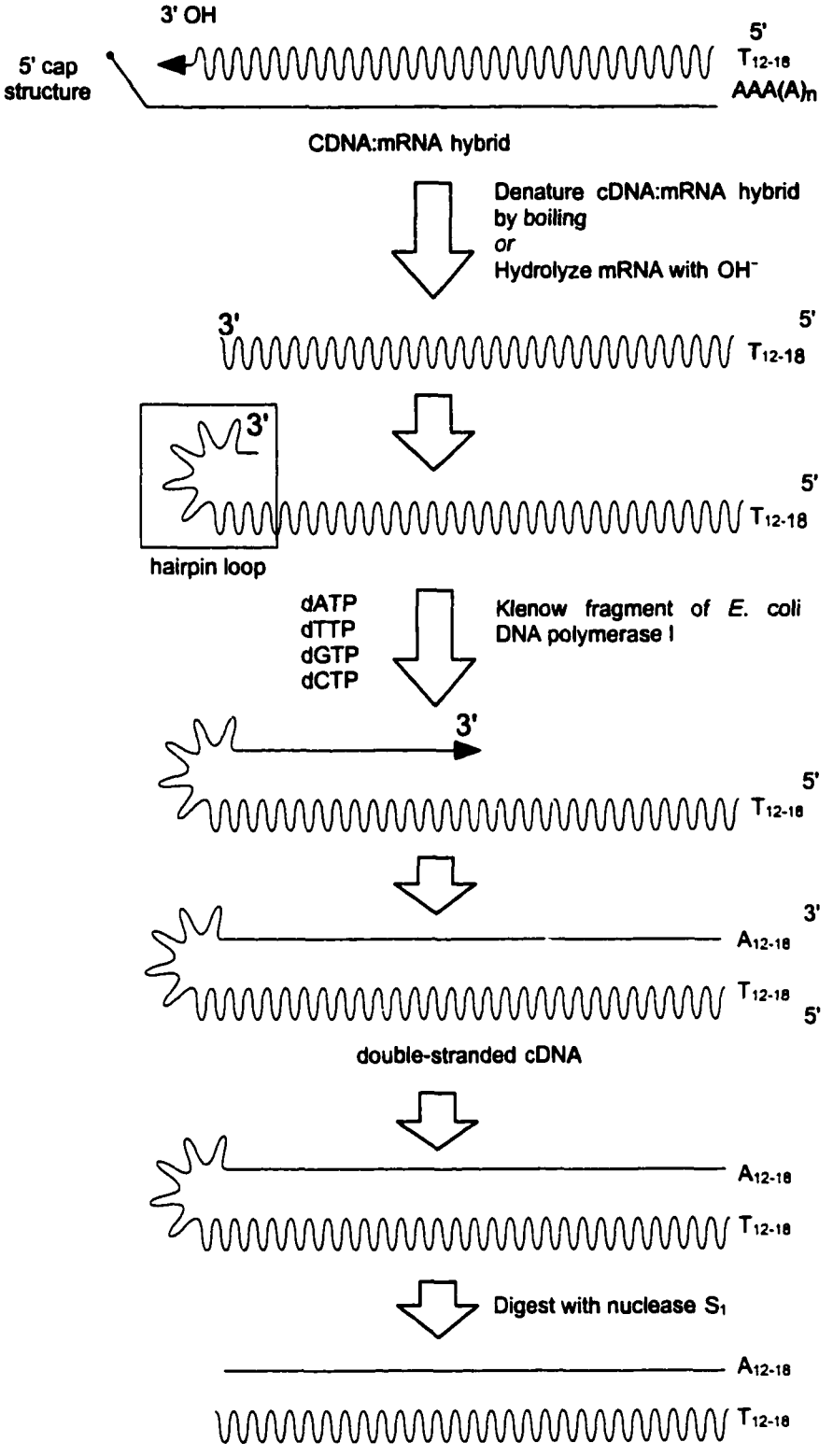


Figure1.11 The synthesis of double strand cDNA with hair-pin primer (Sambrook et al.,1989)

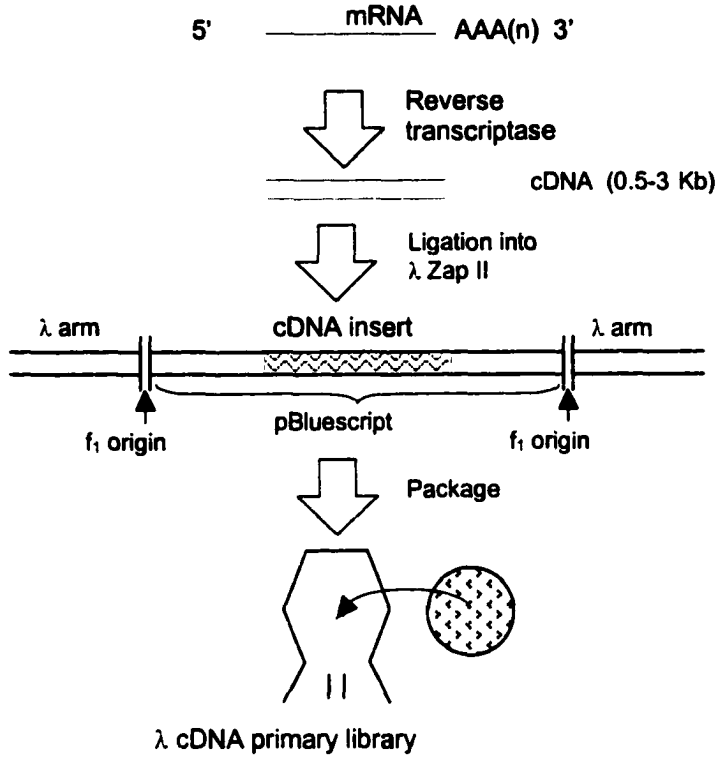


Figure 1.12 Lambda cDNA primary library

1.2.5.4 Vector for cDNA library construction

A vector is a DNA that contains the ability to be replicated, into which a foreign DNA fragment can be inserted. A vector can replicate autonomously inside the host cell which allows the insert fragment to be propagated (Pan, 1996).

Choosing the right vector system is very important for cloning efficiency. Multiple types of vectors may be used to construct a cDNA library. However, most of the large-scale cDNA sequencing studies are based on λ and phagemid libraries (Hillier et al., 1996; Sambrook, Fritsch, and Maniatis, 1989). The lambda phage has been widely used as a vector to clone both cDNAs and genomic DNA. The lambda phage genome is a double-stranded DNA molecule encapsulated in an icosahedral head (Singer and Berg, 1991). A tubular tail projects from its head. Lambda phage has alternative life styles: the

lytic pathway or the lysogenic pathway. In the lytic pathway, it replicates itself quickly inside the host cell and destroys the host cell. The progeny virus particles are released. The host cell is lysed. In the lysogenic pathway, λ DNA is integrated into the host-cell genome and is reproduced as a part of the host-cell DNA. Then the lambda phage keeps inactive for many generations until changes in the environment trigger the excision of this dormant viral DNA to begin a lytic pathway via transfection (Stryer, 1995). High efficiency in transfection and the convenience of constructing cDNA library outside of cell (*in vitro*) are the most important characters that attract people to use it. Modified λ phage can enter bacteria much more efficiently (30-100 fold) than plasmids (Short et al., 1988; Welsh et al., 1990). The high transfection efficiency greatly facilitates the construction of cDNA libraries that contain more than $1x10^6$ independent clones. The probability of isolating rare clones also is improved by this increased efficiency. Therefore, a lot of mutant λ phages have been constructed and designed as cloning vectors. Bacteriophage λ also has a disadvantage as a cloning vector. Since it is quite large, restriction mapping and the sequence analysis of genes cloned in lambda is complicated. Therefore, DNA inserts are usually subcloned into plasmid vectors after initially the cDNA library is created in a λ phage vector. By excising the cDNAs into plasmids they also usually are more stable and easier manipulate than they would be if maintained in λ phage.

Lambda ZAP is a bacteriophage λ expression vector that is often used as a cDNA cloning vector. The λ ZAP system is designed to have both the characters of bacteriophage and plasmid (it is called phagemid) and can rescue DNA from lambda to plasmid cloning vector *in vivo* rapidly and efficiently. It is designed to combine both the

advantages of high efficiency of λ library construction and the high convenience of the plasmid system (Stratagene, 1996).

λ Zap II vector also allows the identification of nonrecombinant phages. The *LacZ* gene in the phagemid works as a blue-white color selection system to perform this task. The ampicillin resistance gene on its arm enables antibiotic selection of the host cell that contains the phagemid vector. It has a large capacity and can accommodate DNA inserts from 0-10 kb in length. The phagemid form generated from excision has bacteriophage f1 origin of replication, allowing rescue of single strand DNA. Lambda Zap also can be used in various subtraction procedures. In the EST studies, clones obtained from λ cDNA libraries may either be subcloned into a plasmid vector prior to sequencing or sequenced directly using a polymerase chain reaction (PCR) to generate an adequate amount of insert DNA (Tracy & Mulcahy, 1991).

1.2.6 cDNA sequencing and the generating of EST sequences

The strategy used to sequence cDNA and generate ESTs is to run a single sequence reaction on both ends of cDNA clones (**Figure1.13**). An insert of a full length double strand cDNA is cloned into pBluescript vectors between the X*ba*I and E*co*RI restriction sites. Sequencing both ends of a cDNA will generate a pair of ESTs. The partial sequence from the 3' end of a cDNA clone is called the 3' EST and the sequence from the 5' end of cDNA clone is called the 5' EST. The 3' EST always starts at the poly (A) tail. It includes the 3' untranslated region (3' UTR), translation stop codon, and may or may not contain coding region. If the cDNA clone is a full length copy of information of its mRNA, then, its 5' EST should cover the 5' untranslated region (5' UTR),

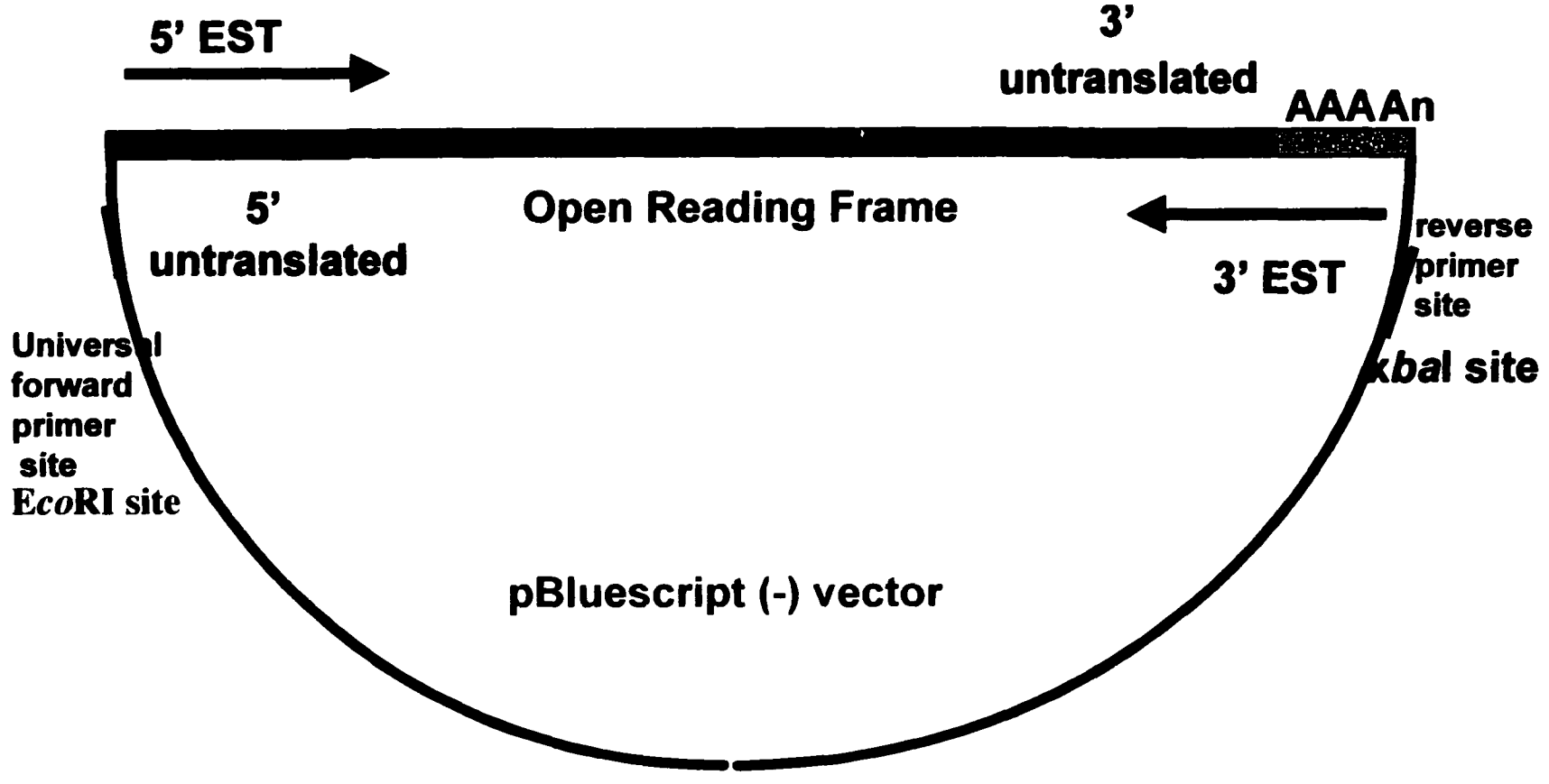


Figure1.13 The strategy for generating ESTs from two cDNA libraries of *Neurospora crassa*

translation start codon (AUG), and protein coding region. The sequence components of 3'ESTs and 5' ESTs can be simplified as follows:

1.) If the EST is not long enough, it may only have the UTR region. In this case, there is no homologue except self in the nr database of GenBank.

2.) If the mRNA is not very large, its cDNA sequence is not very long. Its two EST pairs (3'EST and 5' EST) overlap each other. The whole sequence of this cDNA is available for further studies.

3.) If the mRNA is a large one, its cDNA sequence is long. The 3' EST does not overlap with its pair 5' EST. At this case, custom synthetic primers can be synthesized and used for primer walking or another round of DNA synthesis to obtain the full length cDNA sequence.

4.) If the mRNA is a large one, but the synthesis of the first strand cDNA is stopped before it copied the full length of the mRNA for some reason, then if several cDNAs of different length are copied from same gene and sequenced (if it is redundant gene), their 5' ESTs will start at different positions on cDNA. After assembly, a 5' EST stagger will appear in the assembled EST database (**Figure 1.14**).

1.2.6.1 Sanger enzyme method of DNA synthesis

The details of methods and procedures of generating EST sequences will be discussed in chapter two. Here only a brief introduction of the background in cDNA sequencing is given.

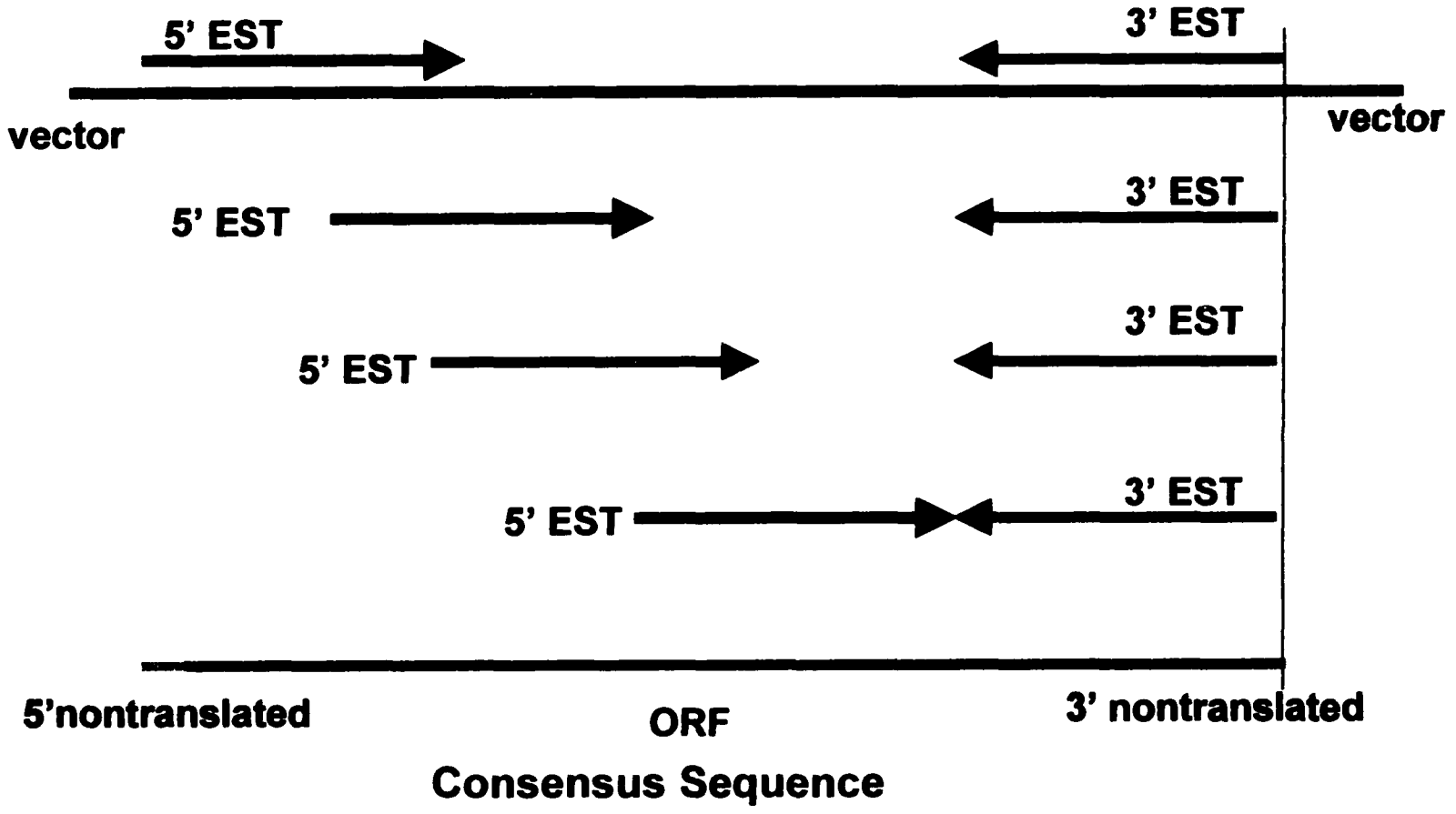


Figure 1.14 Assembly of both 3' EST and 5' EST into Contig with Phrap and the stagger of 5'ESTs

The principle used in generating ESTs is based on DNA synthesis. DNA synthesis is a process of elongation of a newly formed strand of DNA. During this process, four deoxynucleoside triphosphates (dNTPs) are incorporated into the 3' hydroxyl group (OH) of a short DNA primer by DNA polymerase. Generally, there are two methods of DNA synthesis: the chemical method (Maxam and Gilbert, 1977) and Sanger's enzyme method (Sanger et al., 1977). The Sanger dideoxynucleotide DNA sequencing method was used in the generating of ESTs. It is used in almost all automated DNA sequencing currently.

The Sanger method requires a single strand of DNA template as well as a short polydeoxynucleotide primer complementary to a small portion of the single strand template. Four sets of separate DNA synthesis reactions take place in one tube. During the reaction, DNA polymerase synthesizes labeled single-stranded fragments of varying lengths. Each fragment is complementary to the template strand and extends from the primer to an occurrence of the base that is complementary to one of four dideoxynucleotide residues. Four dideoxynucleoside triphosphates (ddATPs, ddGTPs, ddCTPs, and ddTTPs) are included in the reaction. The incorporation of a dideoxynucleotide residue to the reaction chain causes the reaction to stop (**Figure1.15**). Because the dideoxynucleotide residue lacks a 3'-hydroxyl group, it prohibits further extension of the chain. A 3' hydroxyl group is necessary for the phosphodiester linkage formation that joins adjacent nucleotides. These fragments are separated by gel electrophoresis according to the relative sizes and the identity of the final base of each fragment. The base sequence of the template is inferred by the order of these fragments on the gel.

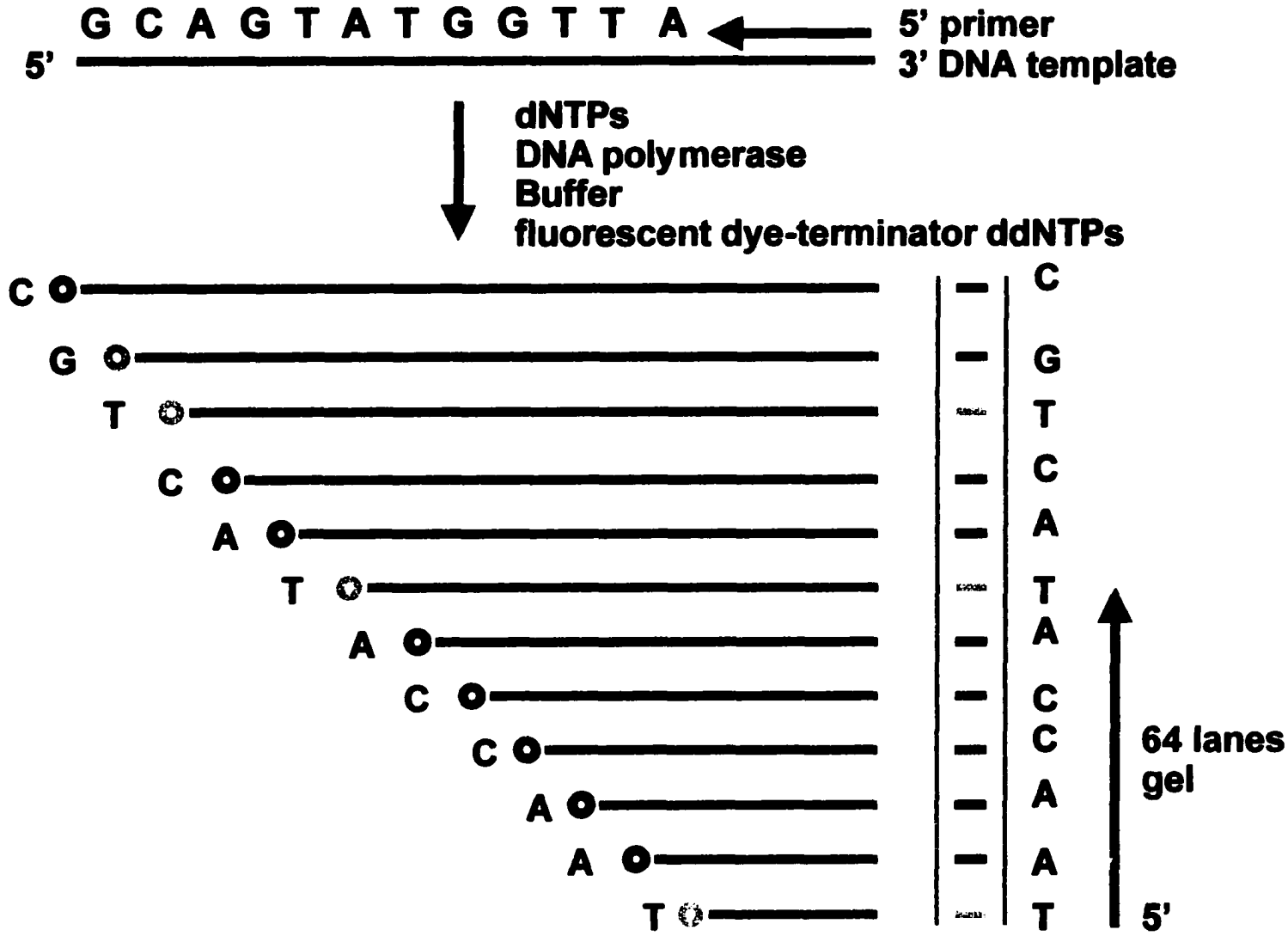


Figure1.15 Sanger enzyme method for DNA synthesis in DNA sequencing reactions

1.2.6.2 Enzymes used in DNA sequencing

The enzyme initially used in DNA synthesis was the Klenow fragment of *E. coli* DNA polymerase I. *Escherichia coli* DNA polymerase I consists of three domains: one has the polymerization and the 3'-5' exonuclease activity and a second one has the 5'-3' exonuclease activity. The Klenow fragment is the C-terminal of DNA polymerase I obtained by treatment of the wild type enzyme with subtilisin and has both the DNA polymerization and the 3'-5' proofreading exonuclease activities but lacks the N-terminal region which contains the 5'-3' exonuclease (Klenow, Overgaard-Hansen, Patkar, 1971).

DNA polymerases from bacteriophage T7 (Sequenase) (Tabor and Richardson, 1987), *Thermus aquaticus* (Taq) (Innis et al., 1988), and *Bacillus stearothermophilus* (Bst) (Stenesh and Roe, 1972) also have been used as sequencing enzymes. Since enzymes isolated from different organisms have different properties, these enzymes are different in their exonuclease activity, thermostability, and their relative ability to distinguish between dNTPs and ddNTPs (Tabor and Richardson, 1987; Innis et al., 1988; Stenesh, 1972). Among them, Taq DNA polymerase has high thermostability (optimal polymerization temperature 72° C) and a high reaction temperature has the potential to reduce DNA secondary structure and GC compression.

Using modern biotechnology and molecular genetics techniques, several groups have developed a novel set of enzymes to meet the needs of the DNA sequencing community and to reduce the cost and improve the efficiency of polymerization. Thermo Sequenase (Amersham) was developed by Tabor and Richardson. It has a Phe/Tyr 667 mutation and a short deletion in the N-terminal region. The replacement of the Phe with Tyr at the position 667 of Taq DNA polymerase decreased discrimination against ddNTP

by 250- to 800-fold (Tabor and Richardson, 1995) thereby allowing lower amounts of ddNTP in sequencing reactions and reducing the sequencing dramatically. At the same time, the quality and read length of sequence data were improved due to the more even signal and less background (Barnes, 1992; Tabor and Richardson, 1989). "Taquenase" or "Klen Taquenase TR" developed by Wayne Barnes is a double mutant Taq DNA polymerase. It is different from wild type Taq DNA polymerase (AmpliTaq) in that it not only has the Phe/Tyr 667 mutation but also has two N-terminal deletions to remove the 5'-3' exonuclease activity. AmpliTaq FS enzyme is the enzyme currently used in our lab. It was developed by Applied Biosystem Inc (ABI) to have the Phe/Tyr 667 mutation and an unknown mutation in the N-terminal region to remove the 5'-3' exonuclease activity (P.E. Applied Biosystems, 1998).

1.2.6.3 Labeling methods in DNA sequencing reaction

In automated sequencing (Smith *et al.*, 1986), two labeling approaches are used: the fluorescent dyes are attached to the primer (dye-primer chemistry) or to the dideoxy chain-terminating nucleotide (dye-terminator chemistry) (Prober *et al.*, 1987). The method of using fluorescent labeled terminators (ddNTP) has been used for sequencing reactions since 1987 (Prober *et al.*, 1987) and has essentially replaced the fluorescent 5'-end labeled primer method (Smith *et al.*, 1986).

Each of the four reactions (for A, C, G, and T) uses a different dye. For both fluorescent-labeled primers and terminators, the fluorescent dyes typically used are fluorescein or rhodamine derivatives. The maximum emission wavelengths of fluorescein and rhodamine derivatives are slightly different. In dye-terminator reactions, four

rhodamine derived dyes are R110, R6G, 6-TAMRA and 6-ROX. Each dye is excited by laser light at a different wavelength and thus all of four wavelengths of the four different bases can be detected and distinguished in a single gel lane. The BigDye terminators are energy transfer compounds as they have two fluorescent dyes attached to each dideoxynucleotide terminator. These energy transfer dyes were introduced commercially by ABI (ABI BigDye terminator protocol, 2000) and have a 5-carboxy-d-rhodamine dye used as the acceptor for all 4 bases, and 5- or 6-carboxy isomers of 4'-aminomethyl fluorescein used as the donor dyes. Since both dyes are attached to the dideoxy terminator, the BigDye terminators can produce a more even and intense signal (Rosenblum, 1997) and reduce noise of background.

For dye-terminator chemistry, all four reactions are carried out in a single tube. 48 or 64 lanes of data can be collected on the same slab gel. A laser which excites the fluorescent dye attached to the fragment is located at the bottom of the gel. When a fragment passes in front of laser, the laser scans back and forth across the read region of the gel. The fluorescence of the dyes is excited by the laser and a PMT collects the emission intensities at 4 different wavelengths. During the electrophoresis, laser and detector scan the bottom of the gel continuously to build a gel image. Each lane of the gel image consists of a ladder-like pattern of bands of four different colors. Each band represents a particular length of fragment. One base large than the preceding band.

Fluorescent labeled terminators were used for sequencing in this dissertation's research. The initial first half of the EST data in this research was collected using rhodamine dye and the later half of the EST data and all sequence data for the *Neurospora crassa* cosmids were collected using Bigdye terminators in the sequencing

reactions. That were resolved and detected mainly on ABI 377 slab gel sequencer but very recently on the 96 lane ABI3700 capillary sequencer.

Converting the gel image to an inferred base sequence (or read) for each template is performed by computer analysis. There are four distinct steps in computer analysis: lane tracking, lane profiling, trace processing and base calling (ABI377 user's manual, 1999). Lane tracking is to identify gel lane boundaries. In the step of lane profiling, each of the four signals is summed across the lane width to create a profile, or trace. it consists of a set of four arrays that indicate signal intensities at several thousand uniformly spaced time points during the gel run. Trace processing uses signal-processing methods to smooth the signal estimates, reduce noise, and correct for dye effects on fragment mobility and for long-range electrophoretic trends. Base calling programs translate the processed trace into a sequence of bases. There are four kinds of curves of different colors in chromatograms. Each curve represents the signal of one of the four bases. Evenly spaced and non-overlap peaks represent ideal trace quality.

1.2.7 DNA sequence analysis

DNA sequence analysis is performed with a variety of computer programs. In this present research, Phred and Phrap were used for editing and assembly of DNA or EST sequences (Green, 1994-1999, Phredphrap documentation). The BLAST programs were used for database searches and comparisons. Crossmatch and Consed, programs included in the Phred/Phrap package, were used for sequence comparing and sequence editing/viewing respectively.

Phred (Ewing and Green, 1998; Ewing et al., 1998) is a base-calling program that uses a four-phase procedure to examine the trace files obtained from the ABI. It reads the raw DNA sequence trace data first and then determines a sequence of base-calls. The first phase of Phred is to determine the idealized peak locations or predict peaks. The second phase is to identify observed peaks in the trace. The third phase is to match the observed peak with the predicted peak location, omit some peaks and split others. The final phase is to check the uncalled (or unmatched) but observed peaks to see if there is any peak that appears to represent a base but could not be assigned to a predicted peak in the third phase. If any of this kind of peak is found, it will be inserted into the read sequence at this stage. A sequence quality value also is assigned to each base after Phred examines the four base traces in the region surrounding each point in the data set. Finally, a calculated quality score is yielded after the observed peaks are compared to the predicted peaks. Phred produces an appropriate quality input file for phrap. It gives a more accurate base call overall for a longer distance, compared to ABI software. An error probability (p) in the base call at each position can be computed with Phred where p is converted to a quality value q using the relationship $q = -10 \text{Log}_{10} (p)$. When a quality value of a base is 10, 20, 30, and 40, it corresponds to an error rate of 1/10, 1/100, 1/1000, and 1/10000 respectively.

Phrap (Green, 1994-1999, Phrap documentation) is an assembly program developed by Phil Green at the University of Washington. The task of Phrap is to construct a contiguous sequence as a mosaic of the highest quality parts of reads using Phred quality values.

Cross_match is a general purpose utility in the Phredphrap program package that is designed for comparing any two sets of DNA sequences no matter if the query DNA sequences is long or short. The basis of cross_match is an efficient implementation of the Smith_Waterman algorithm (SWAT). Compared to BLASTN, cross_match is slower but more sensitive.

Consed is a program to view and edit sequences developed by P. Green's group (Gordon, Abajian and Green, 1998). It is designed to use in conjunction with the base-calling program Phred and the assembler Phrap. Consed needs three types of data input files: chromatogram files, phd files, and an ace files. Each read needs one chromatogram file, which contains the fluorescent trace profiles. Phred creates phd files which contain the base calls, quality values, and trace peak positions for the read bases, and any tags attached to the read. An ace file is created initially by Phrap, which contains assembly information including the contig sequences and quality values, tags attached to the contig sequences, and read alignment information. Consed's most distinctive feature is the use of the error probabilities as a primary tool for guiding the entire finishing process.

BLAST (Basic Local Alignment Search Tool) is a heuristic search algorithm. The BLAST programs are widely used tools for searching protein and DNA database for sequence similarities (Zhang and Madden, 1997). They allow comparison of protein or DNA queries with protein or DNA databases in any combination. The BLAST algorithm was written to meet the need of balancing speed and increased sensitivity for distant sequence relationships. Opposite to multiple sequence alignment programs, which rely on global alignments, BLAST focuses on detecting the similarity among sequences in the regions of local alignment (Altschul et al., 1990). BLAST is not only a simple tool to

view sequences aligned with each other and calculate percent homology, but also a program that can locate regions of sequence similarity which suggest the structure and function.

Table 1.03 shows the several Blast methods that are available for sequence similarity searches. They are BlastP, BlastN, BlastX, tBlastN, and tBlastX. How to use BLAST will also be briefly introduced in chapter two.

Table1.03 Blast programs and their descriptions

Programs	Description
BlastP	using an amino acid as query to screen against a protein sequence database
BlastN	using a nucleotide sequence query to search against a nucleotide sequence database
BlastX	using the six-frame translation products of a nucleotide query sequence to search against protein sequence database
tBlastN	using a protein sequence as query to screen against a nucleotide sequence database translated in all six reading frame
tBlastX	comparing the six-frame translations of a nucleotide query sequence with the six-frame translation of a nucleotide sequence database

Most DNA sequence comparison is an examination between the query sequence and the public sequence databases. Public databases of both nucleotide sequences and protein sequences are available for users all over the world. However, the databases of GenBank of NCBI (http://www.ncbi.nlm.nih.gov), the EMBL Data Library of United Kingdom (http://www.embl.org) and the DNA Databank of Japan (http://ddbj.nig.ac.jp) are most commonly used sources for sequence comparing.

Chapter II

Materials and Methods

2.1 Sequencing of the ESTs of two cDNA libraries from *Neurospora crassa*

Two cDNA libraries were constructed from the mycelium of two *Neurospora crassa* strains. Wild type frq^+ strain with a 21.6 hour period was used for construction of the *Neurospora crassa* evening cDNA library (NE) while long period strain frq^7 with 29 hour period was used for the construction of the *Neurospora crassa* morning cDNA library (NM). At the time of harvesting cells for isolating RNAs, the circadian time (CT) for frq^+ strain was CT13 (it equals the subjective evening) while for frq^7 mutation strain it was CT1 (it equals the subjective morning). Single sequencing reactions were run for both ends of each cDNA insert to generate the EST pairs, the 3' EST and the 5' EST.

2.1.1 Strains and growth conditions

Two *Neurospora crassa* strains were used in this study: the frq strain 30-7 (bd; A), and the long period mutant 695-425 (bd; frq^7; A). The band (bd) mutation enhances the clock cycle itself (Sargent, Briggs & Woodward, 1966).

In the preparation of the tissues for constructing the cDNA libraries for this research, conidia were inoculated into a high concentration glucose-arginine liquid medium, with the concentration of approximately 6×10^7 conidia per milliliter. The conidia germinated and grew in constant light (LL) at 22°C until a uniform mycelial mat was formed. Before conidiation started, disks were cut from a mycelial mat by a cork

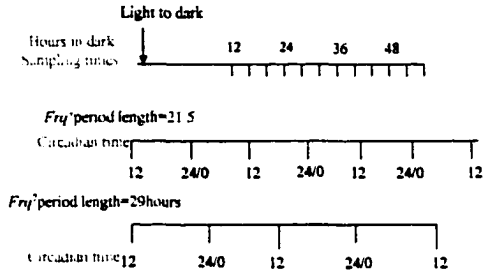


Figure2.01 Mycelial harvest protocol for clock-controlled genes (Sceience243:385-389).

borer, and then transferred to a low-concentration glucose-arginine shaking culture in constant darkness (DD).

Poly $(A)^+$ RNA used for cDNA synthesis was isolated from *frq*$^+$ and *frq7* cultures which grew in DD for 43 hours, representing approximately CT 13 (dusk) in *frq*$^+$ and approximately CT1 (dawn) in *frq*^{7} (Figure2.01). The method of isolating poly $(A)^+$ RNA (2μg) was described earlier (Chirgwin, Przybyla, MacDonald & Rutter, 1979).

2.1.2 Construction of cDNA library

The construction of the cDNA library was done using the lambda Zap II vector (Stratagene) after the two strands of cDNA were synthesized (Bell-Pedersen et al., 1996a). The insert was cloned between *Xba*I and *Eco*RI restriction sites on the lambda vector arm.

2.1.2.1 The synthesis of double stranded DNA and construction of λ primary cDNA library

Dr. Debrorah Bell-Perdersen of Dr. Jay C. Dunlap's lab, Dartmouth Medical School, constructed the *Neurospora crassa* evening and morning cDNA libraries (Bell-Pedersen et al., 1996c).

1). The BRL cDNA synthesis system was used to synthesize the first- and second-strand cDNA. An oligo (dT) linker-primer that contains an X*ba*I site was used in the synthesis of the first-strand cDNA. The poly(A) mRNA was the template. The poly (dT) region bound to the 3' poly (A) region of the mRNA and guaranteed the direction of the synthesized cDNA (**Figure 2.02**). Poly $(A)^+$ RNA (2μg) was annealed to 0.6 μg of oligo

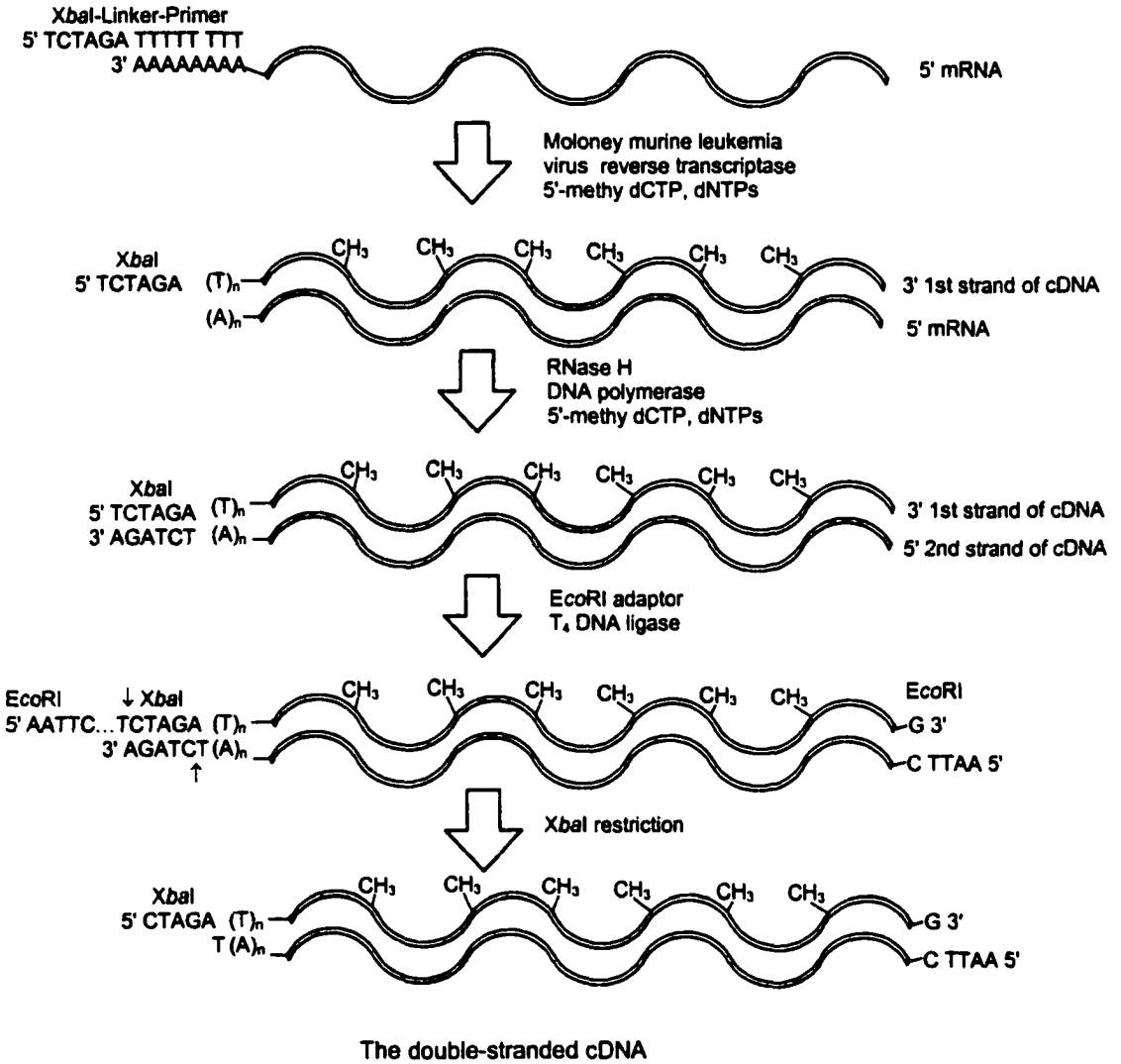


Figure2.02 The synthesis of the double-stranded cDNA

(dT)-X*ba*I primer-adapter (Promega) by heating the reaction to 70°C for 3 minutes, then cooling down on ice immediately. 5'-methyl dCTP was used to protect the

cDNA from digestion from the restriction endonuclease, *Xba*I and *Eco*RI, used later as it is an equally acceptable substrate as dCTP for Moloney murine leukemia virus reverse transcriptase (MMLV-RT).

2). RNase H was used to nick the RNA bound to the first-strand cDNA to produce a fragment that served as primer for DNA polymerase I to start the synthesis of the second-strand of cDNA.

3.) The blunt ends of double strand cDNA (ca. 250ng) was ligated to 10 pM phosphorylated *Eco*RI adapter (Promega). After phenol extraction, the product was then digested with 10 units of *Xba*I. The sequence of the *Eco*RI adapters is as follow:

5' AATTCGGCACGAG3'

3' GCCGTGCTC5'

4.) Digestion with *Xba*I released the *Eco*RI adapter and residual linker-primer from the 3' end of the cDNA. The 5'-methy CTP on the first strand of cDNA allows that only the unmethylated site within the linker-primer is cleaved during the digestion by *Xba*I.

5.) Sephacryl S-400 spin column (Promega) was used to purify cDNA and to select cDNA molecules between 0.5 kb and 5 kb in size.

6.) The resulting cDNA then was inserted into the Uni-ZAP XR vector in a sense orientation (*Xba*I – *Eco*RI) with respect to the lac *Z* promoter by ligation was done overnight at 5° C at a 1:1 molar ratio of cDNA to λ Zap II arms.

The construct needs to be packaged with packaging mixture containing phage head protein and packing protein. The packaging was done *in vitro* with Promega's package extracts and plated on *Escherichia coli* strain XL1-Blue. The percent of non-

recombinant background plaques was determined by plating an aliquot of the library on LB plates containing 10 mM IPTG (isopropyl-β-D-galactopyranoside) and 5 mg/ml of X-Gal (5-bromo-4-chloro-indolyl--β-D-galactopyranoside). The ratio of blue/white plaques were counted. A recombinant plaque is white while a non-recombinant plaque is blue. A cDNA library at that stage is called a cDNA lambda primary library.

The *E. coli* host strain XL1-Blue MRF' was supplied with the ZAP-cDNA synthesis kit. The amplified library can grow very efficiently on XL1-Blue MRF' and produces blue plaques with non-recombinants and white plaques with recombinants. So XL1-Blue MRF' is suitable for being used as both the determination of the ratio of non-recombinants to recombinants with the blue-white color selection system in the primary library again and for screening after the library has been amplified in XL1-Blue MRF' cells.

2.1.2.2 Mass excision of the pBluescript phagemid from the Uni-Zap XR vector

For the reasons discussed in the introduction and in order to manipulate the DNA conveniently, the insert cDNA and pBluescript SK(+/-) vector system must be excised from the lambda ZAP vector. Any cloned insert contained within the lambda vector can be excised *in vivo* and recircularized to form a phagemid that contains this cloned insert in a simple and efficient way. This step was done by Dr. Bell-Pedersen using M13 helper phage 408 as described in the Lambda ZapII Cloning Kit (Stratagene) (Bell-Pedersen et al., 1996a). Single-stranded circularized phagemid DNA resulted from this excision step. These phagemids with cDNA inserts were then packaged as single-stranded filamentous phage particles.

2.1.2.3 Conversion of pBluescript phagemid DNA to plasmid DNA

Since plasmid DNA is more stable and easier to manipulate in the lab than phagemid, the pBluescript phagemid DNA was converted to plasmid DNA by infecting *E. coli* Su^- (nonsuppressing) SOLR strain with phagemid DNA without the help of f1/M13 helper phage. Since the helper phage has an amber mutation and lacks the ampicillin-resistance gene, it can not grow and replicate its genome in SOLR cells. Since they do not support the growth of lambda phage, lambda phage contamination was avoided. My research started at this stage. All the work before this stage was done by Dr.Dunlap's lab at Dartmouth Medical School.

Preparation of SOLR cells was done as follows. 20 ml of autoclaved LB medium was put in a 12×75 Facon tube. Kanamycin was added to a final concentration of 20μg/ml. The cap on the top of the tube was pierced to allow aeration. After one clone was inoculated into media, the tube was incubated in a 37 °C shaker for 20 hours at 250 rpm. Cells were collected by centrifugation 10 minutes at 2500 rpm on the Beckman GPR. Then the cell pellets were resuspended with 8 ml sterile 10mM Mg_2SO_4. Cells were stored in a 4 °C cold room for later use.

Conversion of pBluescript phagemid DNA to plasmid DNA was performed by incubating 250 μl of the SOLR cell solution containing 1μl of f1 lysate from the cDNA library at a suitable concentration in a 37 °C water bath for 20 minutes. To each tube 62.5 μl of 25 mg/ml IPTG and 62.5 μl of 20 mg/ml X-Gal was added. Each of 100 μl of the cell mix was plated on LB plate containing 100 μg/ml ampicillin. Plates were incubated at 37° C for 16-18 hours. The blue/white β-galactosidase selection system was used to

check the cDNA plasmid contained in the SOLR colonies. The ratio of white to blue colonies of the two cDNA libraries was examined (Table2.01)

Table 2.01 Summary of *Neurospora crassa* cDNA library

cDNA library	% clones with insert (white colony)
Neurospora morning cDNA library	24
Neurospora evening cDNA library	63.2

2.1.3 Growth and amplification of cDNA clones

To grow and amplify the cDNA clones, white colonies picked from the LB ampicillin plates were inoculated into 1.5 ml of Terrific Broth media (TB) with 100 μg/ml ampicillin. One liter TB contains 900 ml terrific broth (12 g Bacto-tryptone, 24 g Bacto-yeast extract, 4 ml glycerol per 900 ml) and 100 ml of TB salts (2.31 g KH_2PO_4, 12.54 K_2HPO_4 per 100ml). 10 ml ampicillin (10 mg/ml) is added into above one liter TB solution to reach a final concentration of 100 μg/ml. 1.5 ml TB was aliquoted into each well of a sterile 96 well block. The cells were incubated for 18 to 22 hours at 37°C in a shaker at 350 rpm. The cells were collected by centrifugation at 2000 rpm for 10 minutes in a Beckman GS-6R centrifuge. The cell pellets were stored at –20°C.

2.1.4 Alkaline lysis isolation of plasmid DNA

The plasmid DNA templates were isolated using a Beckman Biomek 2000 automated workstation. The details of the procedure for isolation of the double stranded DNA templates are described as follows.

The cell pellets were thawed completely at room temperature for at least 30 minutes. Then, the program "td TE-RNase A -to-End" on the Biomek2000 was selected

to isolate plasmid DNA template. First, 200 µl of TE_RNAseA (50mM Tris, pH8.0, 10mM EDTA, 100 µg/ml RNase A) was added to each well of 96 well-block to resuspend cell pellets (20 times). Second, 200 µl alkaline lysis solution (0.2 N sodium hydroxide or 1% sodium dodecyl sulfate, SDS) was added to lyse the cells. After mixing 10 times, 200 µl of sodium acetate (3M, pH 4.8) solution was added to each well of the blocks to precipitate the lysate. Each block was removed from the Biomek. The blocks were shaken in a 350 rpm shaker for 10-30 minutes to make sure that the solutions were mixed very well. This step is important to help to form a solid debris pellet and cleared lysate. Then, the blocks were stored at -20°C over night. The next day, the lysate was thawed at room temperature and the blocks were centrifuged at 3000 rpm for 45 minutes in the Beckman GS-6R centrifuge to get the cleared lysate. 400 µl of the top portion of the cleared lysate was transferred to four fresh blocks. 1 ml of 100% ethanol was added to each well by the Biomek 2000 or manually. Then, after sealing the block tightly with foiled-paper, the block was inverted 2-3 times to mix ethanol and sample. After that, the block was stored in -20°C overnight. The DNA was precipitated by centrifugation for 30 minutes at 3000 rpm in the Beckman GS-6R. The pellets were washed with 75% ethanol by centrifugating for 15 minutes at 3000 rpm in Beckman GS-6R. The block was drained on paper towel and dried in a vacuum for more than 30 minutes, 100 µl of sterile deionized water was added by the Hydra 96 to resuspend the DNA pellets. Then, a 2µl aliquot from each well was examined by electrophoresis on a 0.8% agarose gel before it was used for sequencing. The size of the cDNA insert could be checked at this step. In this research, the average size of cDNA inserts was estimated from the sequencing results. The consensus sequences of 30 contigs from each cDNA library were used to

count the average size of the cDNA inserts. All these contigs cover the complete sequences of the cDNA inserts. The average size of the inserts from the morning cDNA library was 978 bp and the average size of the inserts from the evening cDNA library was 1032 bp after examine 30 cDNA clones from each cDNA library.

2.1.5 cDNA sequencing

Before the large scale cDNA sequencing, a variety of parameters were examined to determine the optimum conditions for EST sequencing. A single-pass sequencing reaction in both the forward and the reverse direction was performed to gain an EST pair from each cDNA clone.

2.1.5.1 Enzyme used in the EST sequencing

The enzyme used in this dissertation research was AmpliTaq FS (Applied Biosystem Inc.). A small part of the EST data was collected using Rhodamine dye fluorescent terminator in the sequencing reaction but most EST data was collected using BigDye terminator in the sequencing reaction.

The following components were included in Perkin-Elmer Bigdye energy transfer-based terminator: AmpliTaq DNA polymerase FS with thermally stable pyrophosphatase, Tris-HCl (pH9.0), dNTPs (dITP, dUTP, dATP, dCTP), ddNTPs terminators, magnesium chloride, thermal-stable pyrophosphatase (Rosenblum, 1997). A 1/12 dilution (1x terminator/enzyme mix: 2x of 5x reaction buffer 50mM Tris-HCl, pH 9.2, 10mM $MgCl_2$) of the Perkin-Elmer reaction mix of BigDyes containing AmpliTaq FS enzyme was used in EST sequencing reactions. Each reaction volume contains 2-3μl

of template DNA, 1 μl of 25% DMSO (the final concentration of DMSO is 5%-10%), 1μl of 13 μM universal forward or reverse primer, 2μl of 1/12 diluted terminator/enzyme mix. The reaction was cycled for 60 to 99 cycles of a three-temperature profile programmed on the thermocycler GeneAmp9600/2400. The three temperature profiles are 50°C for 5 seconds, 96°C for 10 seconds and 60°C for 4 minutes (Roe et al., 1997).

2.1.5.2 Primers used in the EST sequencing

Several primers were tested and compared before the large-scale EST sequencing. T3, T7, KS, Universal reverse and M13-20 Universal forward primers were used for sequencing reaction and compared. Finally, Universal reverse primer was chosen for reverse reactions. The reverse reaction was sequencing the 3' ends of cDNA clones, which yielded .r1 ESTs that have poly (T) tail in their beginning sequence. The M13-20 universal forward primer was used for the forward sequencing reaction. It sequenced the 5'ends of cDNA clones and yielded .f1 ESTs. The sequence of the 18 mer universal reverse primer is 5' GGAAACAGCTATGCCATG 3'. The sequence of M13-20 universal forward primer is 5' GTAAAACGACGGCCAGT 3'.

DMSO (dimethly sulfoxide) is a denaturing agent that can reduce the secondary structure in the DNA. 5-10% DMSO was used in the EST sequencing reactions routinely since at that concentration it has only positive effects on sequences of secondary structure and does not have a negative effect on Taq FS enzyme.

2.1.5.3 Thermocycler reaction and the purification of reaction products

Both 384-well plates and 96-well plates (Cycleplate, 384ET or 96ET, Robbin Scientific) were used for sequencing reactions. Reaction samples were prepared by a series of semi-automated procedures programmed on the Hydra 96 (Robbin Scientific). After the reaction plates were ready, they were put on the Perkin-Elmer9600 or 9700 thermocycler. Different reaction conditions for both BigDye and dRhodamine were programmed on the thermocycler. The program of 25 cycles or 45 cycles was selected for the dRhodamine dye-terminator reactions while 60 cycles or 99 cycles was used for Bigdye dye-terminator reactions in this research. Three temperature files were linked for each cycle on Perkin-Elmer GeneAmp PCR system9600 or 9700 thermocycler. Typically, they were at 96°C for 10 seconds (denaturation), 50°C for 5 seconds (primer annealing) and 60°C for 4 minutes (chain elongation)(**Figure2.03**).

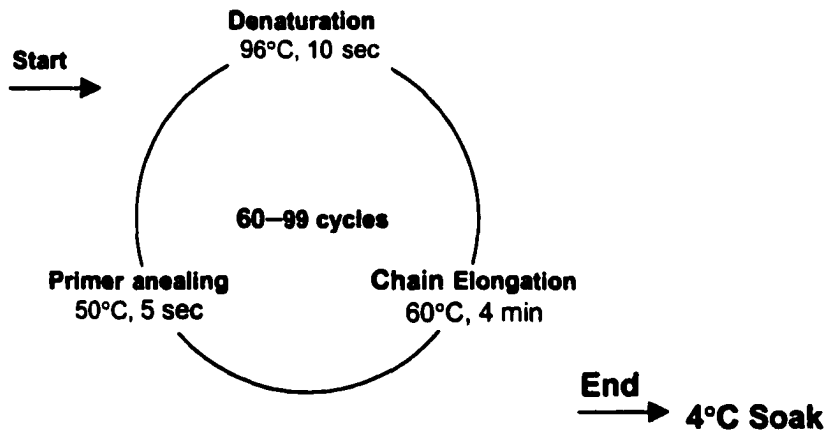


Figure2.03 Thermocycle reaction on the GeneAmp9600

The thermocycling amplification reaction products were treated by either gel filtration through Sephadex G-50 mini-columns or by ethanol precipitation to remove unincorporated terminators and buffer salts. The mini-filtration columns were prepared at least 3 hours in advance using a semi-automated procedure programmed on the Hydra96. 200 μl of dry Sephadex G-50 was added manually to each well of a 96-well filter plate

(Unifilter 350, Polyfiltronics) with the help of a multiscreen column loader (Multiscreen Filtration System, Millipore, CAT# MACL09645). Next, 290 µl of dd water was dispensed into each well by a Hydra96. The plates were stored in the cold room for more than 3 hours or over night. Before the filtration plates were used to filter DNA products, each 96-well filter plate was centrifuged for 2.5 minutes at 1500 rpm. Once the cycling of thermocycling reactions was completed, the cycle plate was centrifuged for 2 minutes at 1500 rpm to collect the liquid to the bottom of well. Then, ten microliters of dd H_2O was added to each sample of the 384 /96 cycle plates to dilute the reaction. Next, all the liquid from each reaction well was transferred to the top of a 96-well Sephadex G-50 column filter plates by Hydra 96. A correct labeled and clean microtiter plate was attached to the bottom of each filter plate to collect purified reaction products. The cleaned reaction products were collected by centrifugation for 3 minutes at 1500 rpm on Beckman GRP. The samples were dried under vacuum. Then, they were ready for loading and were stored at –20 °C before being loaded onto an ABI Sequencer.

Clean-up with 95% ethanol precipitation does not require filter plates and G-50 columns. Before transferring the reactions to a 96-well microtiter plate, 10 µl of ddH_2O was added to each well using Hydra96. 60 µl of 95% ethanol with 0.12 M NaOAc was added to the reaction sample of each well with the semi-automated program on the Hydra96. After incubation at room temperature for 30 minutes, the microtiter plate was centrifuged for 30 minutes at 3000 rpm on Beckman. Then, 100 µl of 75% ethanol was added to wash sample followed by centrifugation for 15 minutes at 3000 rpm on the Beckman. The cleaned reaction samples were dried in vacuum and stored at –20°C before loading on ABI377 sequencer.

2.1.6 Sample loading, electrophoresis, automated data collection and equipments

Before the samples were loaded onto an ABI377 Sequencer, 1 μl loading buffer (10 mg/ml blue dextran and 5mM EDTA in deionized formamide) was added into each well to resuspend the sample. Samples were heated at 95°C for 2 minutes and loaded on a long-ranger polyacrylamide gel (5-5.5%). Vertical electrophoresis was performed for 7 or 10 hours on an ABI 377 sequencer. The data were then automatically collected and preliminarily analyzed by ABI software on a Macintosh computer associated with the ABI 377 sequencer. Then, the raw sequence trace files were transferred to a SUN workstation for assembly and further analysis.

The ABI 377 sequencer is a microprocessor-electrophoresis and fluorescence detection system (ABI 377 users manual, 1999). It has 4 major kinds of components: loading system, separation system, detecting system and control systems. When DNA fragments labeled with fluorescent dye pass through the read region of a gel, the argon ion laser excites the fluorophores. The gel is held at the focal point of the laser beam by a positioning pin in the electrophoresis chamber. The emitted light is collected by a series of lenses and focused into a spectrograph. The spectrograph separates lights (based on wavelength) across a charge coupled device (CCD) camera. The data collection software will collect the light intensities from the appropriate areas of the CCD camera. This software has stored information about the dyes that are being used and their emission light wavelength. Then, the analysis software will process, analyze and translate the information into base sequences. The CCD camera in the Model 377 leads to a higher detection sensitivity compared to the old Model ABI373. Moreover, the ABI377 has several other improvements. First, one more heat-transfer plate was added to the front of

the gel plate in Model377 instead of only one heat-transfer plate in the back of a mounted gel plate in the Model 373. This brings more efficient temperature control and makes it possible for the electrophoresis to be performed at higher voltage. In addition, more numbers of lanes per gel (from 48 to 64, and 96) result from a 2-fold increased movement speed of the detector and a 2-fold more focused lens on the detector.

2.1.7 cDNA gap closure

For some cDNA clones of special interest, their full sequences were obtained by a primer walking approach. Generally, the average size of a cDNA is 500-2000 bp. The possibility of a gap larger than 2 kb is very rare. The strategy used to close a cDNA gap is described as follow (**Figure2.04**):

1). If the cDNA clone that needs gap closure had both 3' and 5' EST clusters, the primers flanking the gap were determined using primOU and a primer-picking program attached with consed and custom synthesized in our lab. Normal primer walking was performed to close gap and the cDNA clone was used as template. If a GC-rich region was hard to walk through, PCR was performed and the PCR product was sequenced.

2). If the cDNA clone had only one EST, or only had one pair of 3' EST plus 5' EST, the sequence of this cDNA clone was enhanced at first through resequencing it at both forward and reverse directions. This was achieved by repeating the normal fluorescent-labeled terminator cycling sequencing reaction using universal primers. After ESTs from the both end of this cDNA clone were obtained, the sequence obtained in this way was crossmatched with its original sequence to check if the resequenced clone was the correct

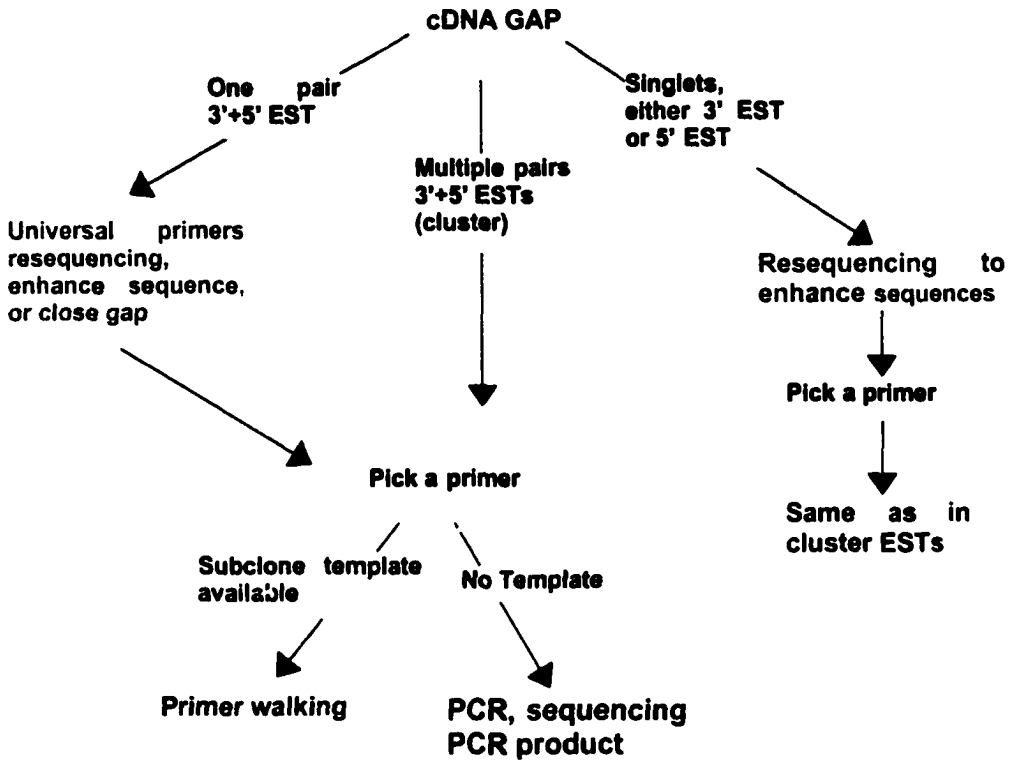


Figure 2.04 Strategy for closing gap in cDNA clones

sequence. Sometime, a tracking problem occurred causing the ESTs to be incorrectly named so that a pair of ESTs did not come from the same clone. If the clone resequenced was the correct sequence, the sequences obtained were assembled to yield the contigs of this cDNA clone. Then, if the gap is not closed, new primers were picked and additional sequencing was done. Usually one or two additional rounds of primer walking were sufficient to complete the sequence of an entire cDNA clone.

2.1.7.1 Primer walking

Primer walking is a strategy that is commonly used in the sequencing closure phase. In this method, the sequencing reaction is performed directly on the template using a pair of primers flanking the gap region (**Figure 2.05.**)

The reaction mixture contained 1-2 µl DNA template, 1µl of 26µm primer (forward or reverse direction), 1 µl BigDye-terminator sequencing reaction mix (or dRhodamine-terminator sequencing reaction mix, or dGTP mix), 1µl DMSO (final to 5%

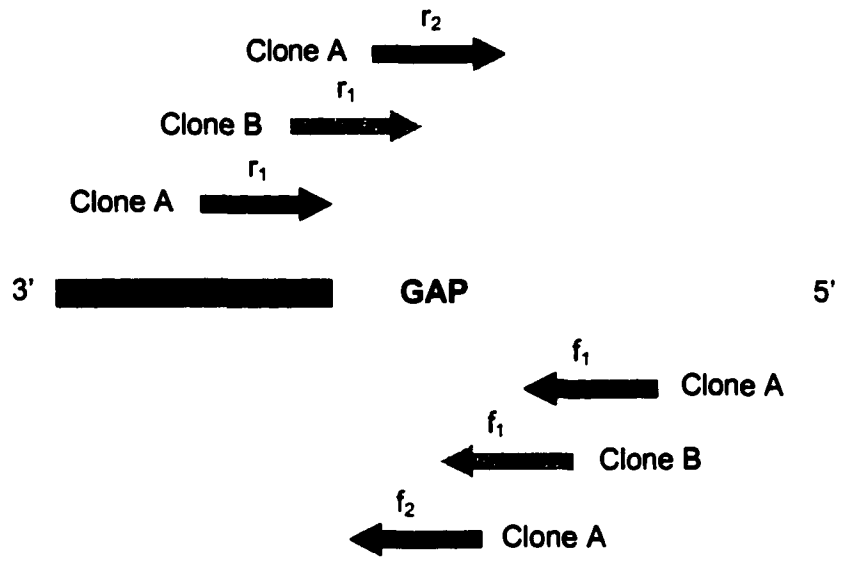


Figure 2.05 Gap closure using primer walking

5% in reaction volumn). The reaction was cycled for 60 to 99 cycles at the same cycle condition as used in normal BigDye- terminator sequencing reactions (section 2.1.5.1, chapter two). After cycling, 10 µl ddH_2O was added. Excess terminators were removed either by passage through a G-50 column or by ethanol precipitation (with 95% ethanol containing 0.12% NaoAC followed by washing the collected pellet with 70% ethanol, drying in vacuum and then dissolving in dd H_2O) prior to loading on the ABI 377. As mentioned above, one or more rounds of primer walking usually were required to close the gap, depending on the size of gap and the length of sequence extension of each walk.

2.1.7.2 PCR in direct sequencing phase

The Polymerase Chain Reaction (PCR) (Mullis and Faloona, 1987) was performed in either the Perkin-Elmer Cetus (PE) DNA thermal Cycler or the Perkin-Elmer Cetus Cycler 9600 to amplify the DNA fragments. PCR was performed by mixing 10-20 ng target DNA, 2.5 μl of each primer, 1 μl of AmpliTaq DNA polymerase (5U), 10 μl 2 mM deoxynucleotide (20 nmole), 10 μl 10X PCR buffer (500 mM KCl, 100mM Tris-HCl, pH8.5, 15 mM $MgCl_2$), and sterile water. The reaction volume was 50 μl. The thermocycling conditions were as follows: 95°C for 1 minute for strand denaturation, 55°C for 1 minute for primer annealing, and 72°C for 2 minutes for primer extension. Reaction was cycled for 25 cycles. The cycling profile was linked to a 4°C soak file. After cycling, 4-10 μl of the reaction was loaded on a 1.2% agarose gel to estimate the size and the amount of PCR product. The PCR product was purified either by low-melting agarose gel electrophoresis followed by phenol-chloroform-ether extraction or by enzyme treatment to digest the extra primers using 1 μl of Shrimp alkaline phosphatase (SAP) (1 unit/μl USB # E779Y) and 1 μl of Exonuclease-I (10 units/μl USB#700073Z). The cleanup reaction mix was incubated at 37°C for 30 minutes and then heated to 80°C for 20 minutes to inactivate the enzymes. An ethanol precipitation or a phenol:chloroform extraction, followed by ethanol precipitation to remove the enzymes (http://www.genome.ou.edu/protocol).

The oligonucleotide primers used both in primer walking and PCR were determined by the PrimOU program adapted from Primo from Southwest Medical Center, Dallas (Chen, Ph.D dissertation, 1997). The primer selected with PrimOU met several criteria. First, the primer was unique. It could only assemble to one desired site. Second, the G/C content of the primer was less than 50% and the melting temperature was between 50 °C and 55 °C. Third, it had no obvious reverted repeat motifs. Oligonucleotide

(primers) was synthesized on a Mermade (Avantech automation corp, New Braunfels, TX) using the oligonucleotide synthesize chemistry in our lab.

2.1.8 Preparation of glycerol stocks and 96-well block retransformation

In this research, three sets of glycerol stocks of each cDNA clone that was sequenced were prepared. One set of glycerol stocks was sent to Dr. Jay C. Dunlap's lab at the Dartmouth Medical School. Another set of glycerol stock was sent to the Fungal Genetics Stock Center, Kansas Medical School, Kansas City, for public requests. The third set was kept at the ACGT, the University of Oklahoma. Most of these stock plates were prepared by Jennifer Gray and other undergraduate students with the help of Dr. Kupfer. Two methods were used to prepare the glycerol stocks.

2.1.8.1 Preparation of glycerol stocks

In this method, the glycerol stock plates were prepared directly from the cDNA containing clones. Here each white colony that has a cDNA insert was picked from the LB amp plates and inoculated into one well of a 96 deep well Beckman block containing 1.5 ml Terrific Broth (TB) media. Cells were incubated in 37°C shaker for 20-23 hours at 350 rpm. Then, cells were incubated on ice and ready for transferring into microtiter plates.

For preparing microtiter plates, 40 µl 50% sterile glycerol (autoclaved) was aliquoted to each well of 96 well microtiter plates. 120 µl cell culture from each well of the 96 deep well TB block was transferred into the microtite plates with a 12-channel

pipette (Costar®). The cells were mixed 3 to 4 times with glycerol using a 12-channel pipette. The plates were labeled and stored at – 80 ° C immediately.

2.1.8.2 96-well block retransformation with calcium chloride competent cell

Retransformation was used when additional template DNA was needed for resequencing or to be sent to our collaborators as glycerol stocks. Since many templates needed to be retransformed, this process usually was done in a multi-sample format using the calcium chloride method (Cohen et al., 1972) to transform SOLR cells with previously isolated and sequenced cDNA clones, rather than the single-sample electroporation method (Sharma and Schimke, 1996) used in shotgun library construction.

2.1.8.2.1 Preparation of $CaCl_2$ competent cells

By treating *E. coli* cells with $CaCl_2$, the cells absorb water and swell. This causes the cell membrane to become porous and thereby allow passage of DNA molecules into the cell. A subsequent heat shock step causes the cells to shrink back to their normal size and trap the transformed DNA inside (Mandel and Higa,1970; Cohen, Chang and Hso, 1972). The protocol we used to prepare calcium chloride competent cell for large-scale retransformation is described as follow (Roe et al., 1997). A single colony from an XL1-Blue streak was inoculated in 3 ml LB in a 12x75 Falcon tube containing 30 µl Tet (Tetracycline) (10mg/ml) and incubated in a 37°C shaker for 8 hours at 250 rpm. 3 ml of above LB solution was transferred to a 50 ml LB containing 0.5 ml Tet (10 mg/ml) in a 250 ml flask. The flask was replaced in a 37°C shaker for 16 hours at 250 rpm. 12.5 ml

50% sterile glycerol was added and mixed very well. Then, 3.3 ml was aliquoted into small tube (12x75 Facon tube) and frozen at – 80 °C (overnight) for later use. After 3.3 ml above stock was thawed in 37°C water bath, the cells were poured into one liter prewarmed sterile 2XTY media (16 g Bacto-tryptone, 10 g Bacto-yeast extract, 5g $NaCl_2$, add water to one liter) immediately. The cells were incubated in a 37°C water bath for one hour. Special attention is needed when pouring 3.3 ml stock into one liter of 2XTY. The tops of both the 1-liter flask and small tube need to be dry and clean to avoid contamination. Incubated one-liter flask for 2.5-3 hours in 37 °C shaker at 250 rpm. Divided above one liter of 2XTY into two of 500 ml centrifuge bottles, and centrifuge 8 minutes at 3000 rpm on RC5-B using GS3 rotor (DuPont). After decanting the supernatant, 250 ml of cold 50 mM $CaCl_2$ was added into each bottle. The cell pellet was resuspended well very gently with repeat pipette up and down very slowly, and then incubated on ice for 20 minutes. Both bottles were centrifuged 8 minutes again at 3000 rpm on RC5-B using GS3 or GSA rotor. The supernatant was decanted and 50 ml of cold 50 mM $CaCl_2$ was added. The pellet was resuspended well very gently as in step 6 to yield the final competent cell suspension. Then, the competent cells were stored at 4°C where they were viable for at least one week.

2.1.8.2.2 The test of calcium chloride competent cell

All calcium chloride competent cells were tested for their viability, purity (if there is any contamination) and their capacity to take and maintain plasmid DNA (if they are competent) before they were used for transformation. Two positive controls and one negative control were designed in the test. The details of the protocol we used to test

competent cells are shown in the flow chart as follow (Figure2.06). The results were explained as follow:

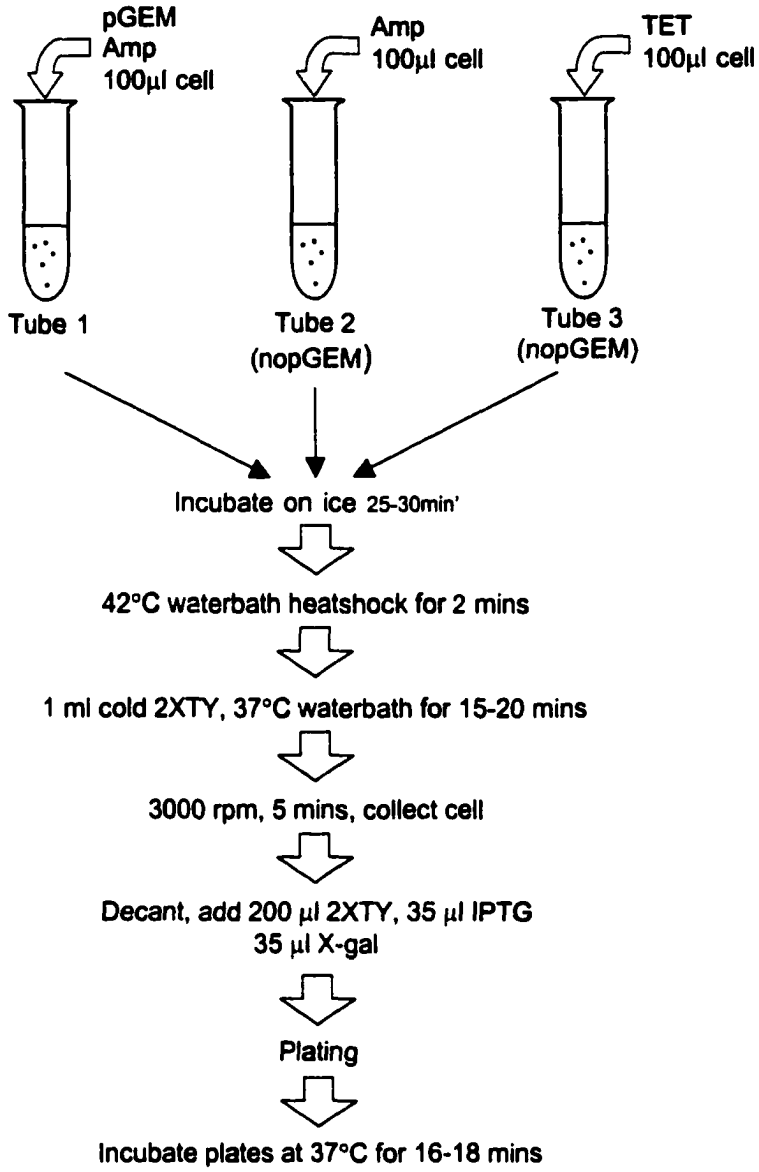


Figure2.06 The flowchart of the protocol of calcium competent cell test

1). In tube #1, standard pGEM DNA, ampicillin and the competent cells were added to test if the cells were competent as host cells for plasmid DNA uptake.

The higher the numbers of blue colonies that appeared on the plate, the better the competent cells were.

2). In tube #2, no DNA was added to test if the competent cells were contaminated. If cells appear on the plates, it means competent cells are contaminated and they should be discarded.

3). Tube #3 was designed to test the viability of the competent cells. Only the antibiotic Tet and competent cells were added into this tube. The result is positive if cells grow well on the LB Tet plate.

2.1.8.2.3 96-well block retransformation

The details of performing block transformation using $CaCl_2$ competent cells is described as follows:

1.) 2 µl of DNA was spotted on the wall 2/3 into each wall of a 96 well block by hand or a Hydra96;

2.) 100 µl of CaCl2 competent cells were added to the bottom of the block and the block tapped to bring the DNA down and mix very gently with the cells. No centrifugation or vortex was used at this step;

3.) The block was incubated on ice for 20 minutes;

4.) The block was put in a 42°C water bath for 2.5 minutes;

5.) 1 ml of cold 2XTY was added and the cells recovered in a 37°C water bath for 15-30 minutes;

6.) The block was centrifuged for 10 minutes at 1800 rpm on GPR Beckman;

7.) Supernatant was decanted and cells were resuspended in 1.5 ml of TB media containing 1x TB Salt and ampicillin (100μg/ml); 1x TB Salt was made from 900 ml TB media plus 100 ml 10x TB salt. 10x TB salt contains 2.31 g KH_2PO_4, 12.54 g K_2HPO_4 in 100 ml water.

8.) Cells were incubated for 24 hours in a 37 °C shaker at 350 rpm;

9.) 2 of 96-well microtiter plates with 40 μl of 50% glycerol in each well were prepared and 120 μl of cell solution from above incubated TB block were transferred into microtiter plates with 12-channel pipette. Cells were mixed well with glycerol by pipetting up and down 3-4 times;

10.) Plates were labeled clearly and frozen at – 80 °C immediately;

The rest of cells in 96-well deep block were collected by centrifuge for 10 minutes at 1800 rpm. The supernatant was decanted and the cell pellets were stored in a – 20 °C freezer.

Before step 10, all the clones that were not transformed successfully were recorded. After step 10, the cDNAs were isolated from above cells using normal cDNA isolation method and sequenced. These sequence were compared using crossmatch to the original sequence of the same cDNA clone to check if the correct cDNA clone was retransformed.

2.1.9 Analysis of *Neurospora crassa* EST sequences

The strategy for EST sequence analysis is based on the assumption that since ESTs represent partial sequences of the genes expressed in specific cells, their analysis will provide direct knowledge about the quantitative and qualitative difference of gene

expression and transcription in the different cells and tissues. To analyze the *Neurospora crassa* EST sequences, the clip and clean method was used to select high quality ESTs. Then, the PhredPhrap programs were used to call the bases and then to assemble the EST data. BLAST is used for the database similarity searches.

2.1.9.1 Clean and Clip of the raw EST sequences

The clean and clip processing of the raw EST sequence data (trace file) was processed automatically by a series of scripts that was modified by Hongshing Lai at ACGT center of OU from the scripts originally written by LaDeana Hillier, Washington University at St. Louis. These scripts initially were used by Dr. Kupfer to establish the EST database of *Aspergillus nidulans* (Kupfer, Ph.D dissertation, the University of Oklahoma, Norman, 1999). There are 19 scripts that perform the clip and clean task. Table 2.02 lists these scripts and their different functions.

Table 2.02 Clip and Clean EST processing scripts

embellish_template	extract information from the template name, get library name from experiment file
reformat-scf.uwphrep	reformat the trace files
the-big-one	call bases with ABI and phred and determines which sequences have overall poor quality (N ratio of 1:5). Makes the quality start and stop estimates based on trace quality, cut at first base < 15
getscf_fiels2expfile	add the information from the trace to the experiment file
embellish_template_2	take the dye terminator information and the library name to extract information about the vector, adapter sequence and primer position.
clip-seq-vec	use vep-vector end point(Staden, 1992) to

	find the sequencing vector and mark those sequences which are completely sequencing vector
clip_left_seq_vec	repeat attempt to find the left cutoff point, using adapter sequence information and distance from primer. Tag if the poly T is not found on 3' end.
clip-seq-wep-left	cut the vector sequence off the left end
clip-seq-wep-right	cut the vector on the right end if detected, this indicates short enough insert to read through in single pass
check-wrong-adapter	is the wrong adapter sequence present, example if 5' adapter sequence seen and name indicates a 3' EST. A tag fails the EST
blastn_vec_check	check for the vector again, trim sequence if necessary BlastN: S=133 S2=133 M=5 N=-11 W=8
extend_seq	can sequence be extended past the conservative initial quality estimate to the second tier of reasonable quality, are there high quality bases to the right of the first base with quality 15
check_processor	checks for sequence length < 100 bases, fail the short sequences.
screen.p	BlastN against the E.coli genome database, the *N. crassa* mitochondrial sequence, *N. crassa* ribosomal sequences. Contaminants are failed. S=170, S2=150, M=5, N=-11
expESTBlastx	BlastX against non-redundant protein database Matrix = blosum62, filter=seq
reversed	checks for reversed clones
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.
filp_qz_qr	bookkeeping to ensure the QR(quality right

	cut) always contains the right most cutoff point the sequence, that if there is a QZ (quality extend) it is the hiqual_stop.
exp2dbest	create a dbEST submission files and place in directory for GenBank submission and placement on webstie

All of these scripts are linked sequentially. A raw EST trace file is loaded as input into the first script, and the output of this first script will be loaded as the input of the second script automatically. The order of script processing is as listed in **Table 2.02**. The base calling and quality screen were performed Phred (Ewing and Green, 1998; Ewing et al., 1998). Here, the quality of each base was assigned a rank of 1-100, where the higher numerical number represented a higher confidence in the base called. **Table 2.03** is the clean and clip process protocol used in this research that was adapted from the protocol of Dr. Kupfer in her dissertation (Kupfer, 1999).

Any ESTs that failed to pass these scripts represented low quality ESTs and they were moved to a separate low-quality directory. These sequences included vector only ESTs, low quality ESTs (too many Ns in reads), short ESTs (<100 bp), wrong end ESTs, ribosomal sequence, mitochondrial sequence, and *E. coli* sequences. **Table 2.04** lists the sequences used to screen out the mitochondrial and ribosomal sequences of *Neurospora crassa*. All the ESTs that passed the scripts were designed high quality ESTs. These high quality ESTs then were used for further analysis via 3'end and 5'end assemblies to generate the assembled-EST database, which then was used in the BlastX homology search.

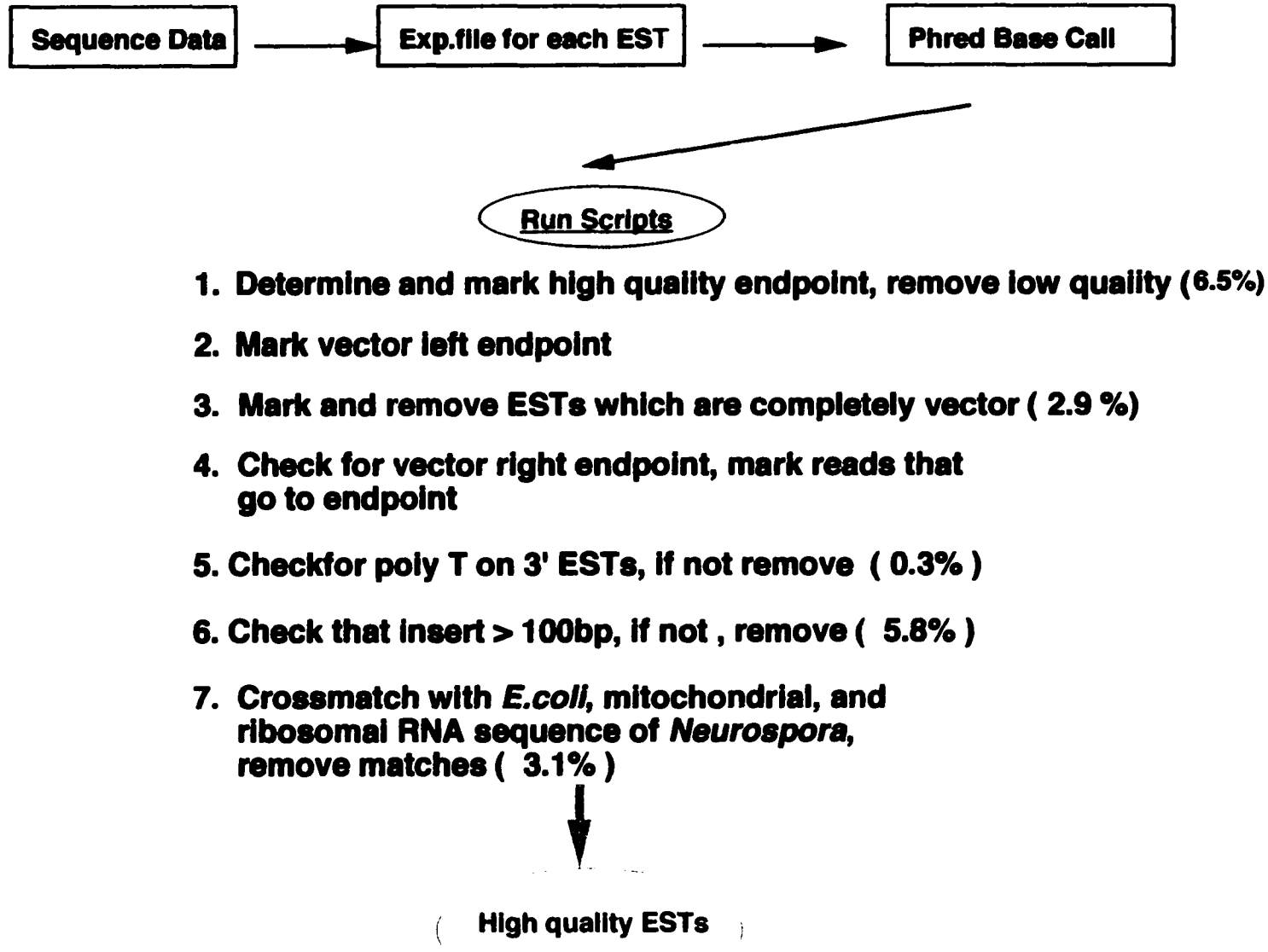


Table 2.03 The processing protocol for clip and clean EST, adapted from Dr. Kupfer.

Table 2.04 Sequences used for screening mitochondrial and ribosomal sequences

Ribosomal sequences:	Accession number	Definition
	x04971	18s rRNA
	u40124	26s rRNA
	x02447	5.8s rRNA
	M13906	5s 18s 26s rRNA
	x02705	5s rRNA
	k03162	5s rRNA
	M11398	5s rRNA
Mitochondrial sequences:		
	x00790	cytochrome oxidase subunit 1, ORF,ATPase subunit6
	x01507	ATPase subunit6,
	x06960	cytochrome oxidase subunit3
	x04161	ATPase, subunit6, 2ORFs, tRNA-Arg, tRNA-Asp
	x07795	cytochrome oxidase subunit 1, ORF
	x15441	NADH dehydrogenase subunit 4, cytochrome oxidase subunit 1
	x15442	NADH dehydrogenase subunit 3 cytochrome oxidase subunit 2

2.1.9.2 Submission of ESTs to the dbEST of Genbank

The EST database is a rapidly growing division of GenBank. As of Jan. 26, 2001, 7,249,092 EST entries had been deposited in dbEST of GenBank. *Neurospora crassa* ESTs represented 25,400 entries, which ranks 23rd among 301 organisms, and is about 0.4 % of the total EST entries in Genbank. Our laboratory has submitted over 20,000 of the ESTs of *Neurospora crassa* to dbEST (**Table 2.05**).

Table 2.05 The list of the EST numbers of several organisms in the GenBank.

dbEST release 012601

Summary by Organism - January 26, 2001

Number of public entries: 7,249,092

Homo sapiens (human)	3,129,328
Mus musculus + domesticus (mouse)	1,919,955
Rattus sp. (rat)	263,234
Bos taurus (cattle)	158,593
Glycine max (soybean)	144,217
Drosophila melanogaster (fruit fly)	116,471
Arabidopsis thaliana (thale cress)	112,500
Caenorhabditis elegans (nematode)	109,215
Medicago truncatula (barrel medic)	101,752
Lycopersicon esculentum (tomato)	94,544
Danio rerio (zebrafish)	79,237
Zea mays (maize)	76,069
Oryza sativa (rice)	69,693
Hordeum vulgare (barley)	68,665
Chlamydomonas reinhardtii	64,973
Xenopus laevis (African clawed frog)	62,601
Sus scrofa (pig)	57,060
Triticum aestivum (wheat)	54,701
Sorghum bicolor (sorghum)	50,825
Pinus taeda (loblolly pine)	31,828
Lotus japonicus	27,078
Gossypium arboreum	26,630
Neurospora crassa	**25,407**
Brugia malayi (parasitic nematode)	22,392
Solanum tuberosum (potato)	20,679
Dictyostelium discoideum	19,183
Bombyx mori (domestic silkworm)	14,849
Onchocerca volvulus	14,347
Schistosoma mansoni (blood fluke)	14,039
Gallus gallus (chicken)	13,512
Emericella nidulans	12,993
Mesembryanthemum crystallinum (common ice plant)	12,879
Sorghum propinquum	12,634
Toxoplasma gondii	12,177
Ciona intestinalis	10,347
Porphyra yezoensis	10,184
Trypanosoma cruzi	10,133
Gossypium hirsutum (upland cotton)	9,438
Schizosaccharomyces pombe (fission yeast)	8,118
Secale cereale	7,886

All ESTs from the two cDNA libraries of *Neurospora crassa* were submitted in two batches either on December, 1998 or on April, 2000. All the high quality ESTs that passed clip and clean processing were submitted to dbEST of GenBank. Each batch of EST files was attached to 3 files according to the NCBI instructions for the submission of ESTs: the pub file (publication), the lib file (library), and the con file (contact). The information about the cDNA library is addressed in the lib file (library file) while the information including EST sequence, cDNA clone, sequencing reaction, sequencing primer and the source that was used for preparing cDNA library is specified in EST file. **Figure 2.07** shows the three files attached to each EST submission. A batch of ESTs with three files attached to each EST file was compressed in a .tar.z file. This .tar.z file was sent to the dbEST of Genbank with the address at: batch-sub@ncbi.nlm.nih.gov through email on Internet.

Contact File:

```
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
```

Library file:

```
TYPE:       Lib
NAME:       Neurospora crassa evening cDNA library
ORGANISM:   Neurospora crassa
STRAIN:           Strain 30-7 (bd; A)
TISSUE:           tissue harvested following 22hr growth in dark
VECTOR:           pBlueScript SK-
V_TYPE:           phagemid
RE_1:       XbaI
```

RE_2: EcoRI
DESCR: See: Bell-Pedersen,D., et al. PNAS 93:13096,1996.
5' end of cDNA cloned into XbaI site of pBluescript
3' end of cDNA cloned into EcoRI site of pBluescript

TYPE: Lib
NAME: *Neurospora crassa* morning cDNA library
ORGANISM: *Neurospora crassa*
STRAIN: Strain bd, frq7 A
TISSUE: tissue harvested following 22hr growth in dark
VECTOR: pBlueScript SK-
V_TYPE: phagemid
RE_1: XbaI
RE_2: EcoRI
DESCR: See: Bell-Pedersen, D.; et al. PNAS 93:13096,1996.
5' end of cDNA cloned into XbaI site of pBluescript
3' end of cDNA cloned into EcoRI site of pBluescript

Publication file:
TYPE: Pub
MEDUID:
TITLE: Two *Neurospora crassa* EST Databases
AUTHORS: Zhu, H.; Lai, H.; Kupfer, D.; Bell-Pedersen, D.; Loros, J; Dunlap, J.C.; Roe, B.A.
VOLUME:
PAGES:
YEAR: 1998
STATUS: 1

Figure 2.07 Three files attached to the EST file in the submission of ESTs of *Neurospora crassa*.

Each EST that was submitted to Genbank was assigned a separate accession number and an EST ID number in the dbEST of Genbank. The accession numbers assigned to the ESTs of two cDNA libraries from *Neurospora crassa* that were submitted to the dbEST of GenBank by our laboratory are listed in Table 2.06.

Table 2.06 Accession numbers of *Neurospora crassa* ESTs submitted to the GenBank dbEST

cDNA Library	Date of Submission	Accession Number of ESTs	ESTs# of submitted
Neurospora crassa cDNA Morning Library (NM)	12/18/1998	AI318679-AI322045	3349
	04/19/2000	AW714914-AW717690	2777
		AW717691-AW719192	1502
		AW721859-AW725138	3279
Neurospora crassa cDNA Evening Library (NE)	12/21/1998	AI328149-AI330327	2179
	04/19/2000	AW708067-AW710278	2212
		AW710279-AW712534	2256
		AW712535-AW714913	2379

The following is an example of the format of the EST submission file (Figure2.08).

AI328667 a5g08ne.r1 *Neurospora crassa* evening cDNA library *Neurospora crassa* cDNA clone a5g08ne 3', mRNA sequence

IDENTIFIERS

dbEST Id: 2112409

EST name: a5g08ne.r1

GenBank Acc: AI328667

GenBank gi: 4065226

CLONE INFO

Clone Id: a5g08ne (3')

Source: J. C. Dunlap, Dept. of Biochemistry, Dartmouth Medical School, Hanover. (Jay.C.Dunlap@Dartmouth.EDU)

Other ESTs on clone:a5g08ne.f1

DNA type: cDNA

PRIMERS

Sequencing: Universal Reverse Primer

PolyA Tail: Unknown

SEQUENCE

AAGGAAAGTATTCGAGGCTGCCTTCGATCGGCAAGCAGCCTCATCCTTGAATTCAGTGTC
CGGGTGAGCGGGAAAGGTTCAGAACTAGCTAAGTACAGGCATGCTCCAATACCCCCCTTC
CCATCCCACTTGAAGAATAATTGATAAAAACAGGATTGCACAAGTATTGACCTATCCAAT
CCCTTTGTGTTGAGTGTGAACCCAAAAAGCTTCTGTGGTGTATGTAGATGTGTTGTACTT
TTTTTTTTCCAGAAAGAAGTGAAGTGTGTAAAGAGGGAGCAAAAGTGAT

Quality:	High quality sequence stops at base: 182
Entry Created:	Dec 28 1998
Last Updated:	Dec 28 1998
COMMENTS	
	We anticipate the future release of the cDNA clones to the Fungal Genetics Stock Center
LIBRARY	
Lib Name:	*Neurospora crassa* evening cDNA library
Organism:	*Neurospora crassa*
Strain:	Strain 30-7 (bd; A)
Tissue type:	tissue harvested following 22hr growth in dark
Vector:	pBlueScript SK-
R. Site 1:	XbaI
R. Site 2:	EcoRI
Description:	See: Bell-Perdersen,D., et al. PNAS 93:13096,1996. 5' end of cDNA cloned into XbaI site of pBluescript; 3' end of cDNA cloned into EcoRI site of pBluescript
SUBMITTER	
Name:	Bruce A. Roe, University of Oklahoma, broe@ou.edu
Lab:	Department of Chemistry and Biochemistry
Institution:	Advanced Center for Genome Technology, University of Oklahoma
Address:	620 Parrington Oval, Norman, OK 73019, USA
Tel:	405 325 4912
Fax:	405 325 7762
E-mail:	broe@ou.edu
CITATIONS	
Title:	Two *Neurospora crassa* EST Databases
Authors:	Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C., Roe,B.A.
Year:	1998
Status:	Unpublished

Figure 2.08 Submission of ESTs of *Neurospora crassa* to dbEST of Genbank

2.1.9.3 The strategy for EST database analysis

A three part strategy was used to analyze the two EST databases of *Neurospora crassa* (**Figure 2.09**). Each part had a different analysis goal. The goal of the first part was to release all of the high quality EST data and databases into GenBank and on our lab website at URL: http://www.genome.ou.edu/fungal.html. After a BlastX search against the non-redundant GenBank protein database using the sequence of the high quality ESTs as queries, these results also were posted on our website where a search of the representative ESTs can be performed using either a keyword or EST clone identifier such as a1b01nm.r1 as the query. The second part of the EST analysis was the assembly of 3' ESTs. By plotting the cumulative 3' ESTs against the redundancy of genes represented in the assembled EST database, the curve of the assembly of 3' end ESTs gives the potential to detect new genes in future sequencing. This graph can be used to determine when it is not productive to collect additional EST sequence data. Since the purpose was to find the genes that were expressed differentially in two time-specific growing stages of *Neurospora crassa*, the gene expressions in these two stages need to be analyzed both quantitatively and qualitatively. Therefore, it requires as many genes as possible to be detected even though they have a low abundance in the library. Thus, the data collection and the EST assembly in this research continued until the redundancy was higher than 90% and the libraries essentially were exhausted of new genes. This is different from two other EST research projects have been performed in our laboratory (Kupfer, 1999; Ren, 2001). In these projects, the ESTs were determined for the two fungi *Aspergillus nidulans and Fusarium sporotrichioides*, to identify the genes expressed in their respective vegetative cDNA libraries. Discovering all genes in these fungi was not

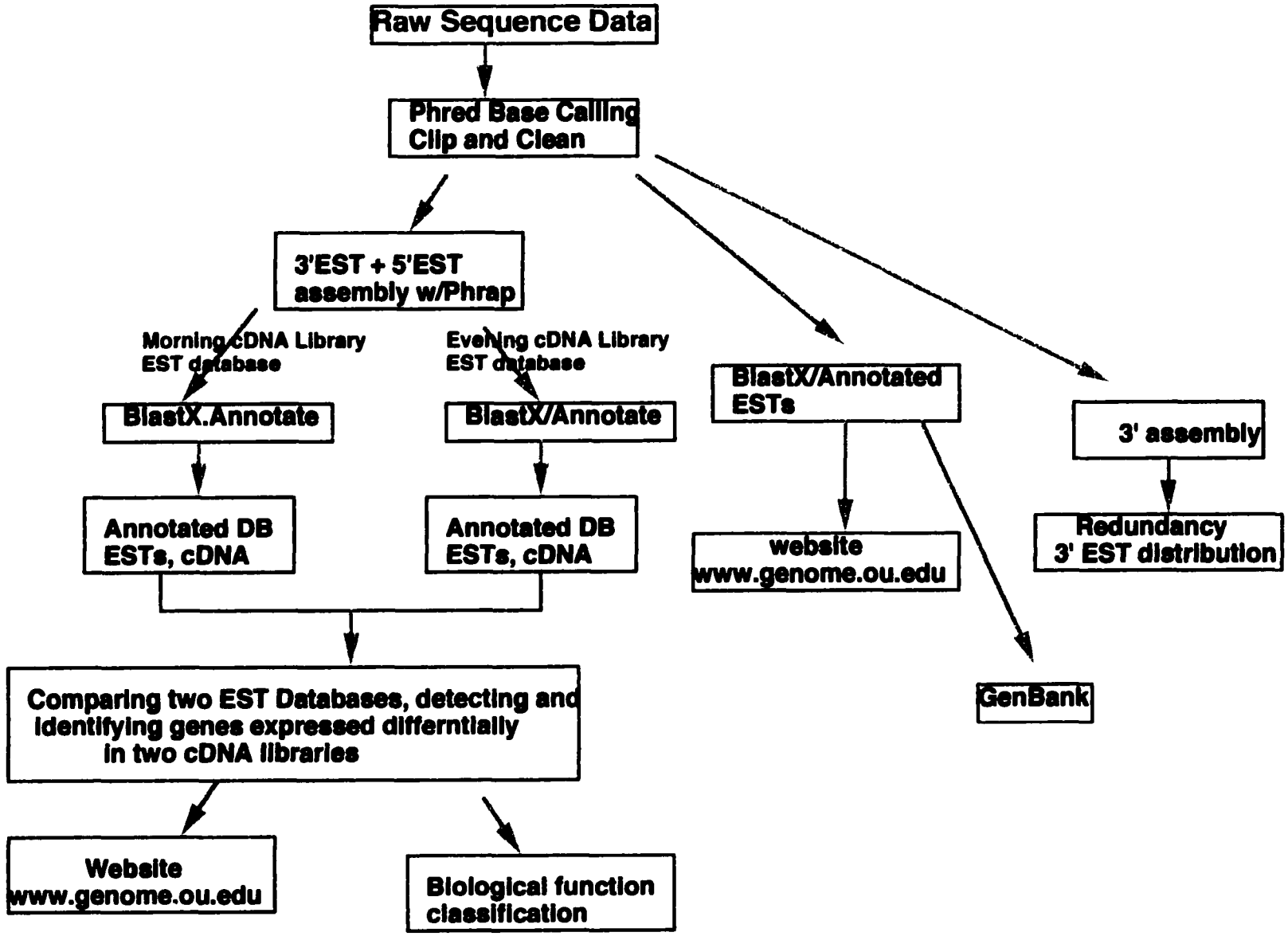


Figure 2.09 EST database analysis strategy

the goal of these two EST projects. Therefore, the assembly of ESTs stopped when redundancy was up to 70% (Kupfer, 1999; Ren, unpublished dissertation data; Hillier, 1996). After the redundancy level is beyond 70%, the yield of new genes is very low and not very efficient in gathering new data by further sampling. Therefore, as noted before, different strategies can be used for the different study aims (Lee et al., 1995).

The third part of the EST analysis was to create and annotate an assembled EST database for the two *Neurospora* cDNA libraries. These results were posted on the ACGT website and subsequently used for further biological studies (**Figure 2.09**).

The assembled-EST database was created by automatically partitioning EST sequence into non-redundant sets of gene-oriented clusters. Each assembled-EST database member contained a consensus sequence that represented a unique gene, as well as related information such as the conditions under which the gene has been expressed. Clustering is the process of finding subsets of EST sequences that belong to the same genes within each *Neurospora crassa* cDNA library. Two sequences are considered linked as a cluster or contig if their similarity exceeds a threshold. The stringency used for creating an assembled-EST database with Phrap in this research is Minimum window_14, Miniscore_80.

A contig or cluster is a set of gel readings that are related to one another by overlap of their sequences. However, the concept of cluster and contig is not same. A cluster consists of a group of sequences that are overlap or related each other. The consensus is the sequence derived from the overlapping of all sequences of a cluster. All clustered gel readings belong to one and only one contig even though a contig might contain only one gel reading. The gel readings in a cluster can be summed to produce a

continuous consensus sequence, the contig. At any stage of a sequencing project, the data will be comprised of a number of contigs. In a genomic DNA sequencing project, when a project is complete, those should be one contig and its consensus will be the finished sequence. In contrast, in an EST sequence project, the number of contigs will increase with the more ESTs being sequenced.

The assembled-EST databases of the *Neurospora crassa* EST project were established after Phrap assembled all high quality EST data. Both 3' and 5' EST sequences were assembled with phredphrap98. Phrap ("phragment assembly program or phil's revised assembly program") is the major program used to assemble shotgun DNA sequence data, in the Human Genome Sequencing project and in the biotech industry. Phrap allows the use of the entire read (or more precisely, as much as is accurate enough to align against other reads) in the assembly. In the Human Genome Project and other large genomic DNA sequencing projects, the contig sequence is constructed as a mosaic of the highest quality parts of reads. However, there is no correspondent standard for the assembly of ESTs yet, and Phrap was not written specifically for the assembly of ESTs and needed modification for this purpose.

2.1.9.4 Assembly of 3' ESTs

Since one of goals of this EST project was to compare the difference in the gene expression profile in two *Neurospora* cDNA libraries, it was critical to obtain EST data for as many genes as possible to ensure that a representation of the cDNA library was obtained. Therefore, it needed to be determined how many ESTs would represent the library complexity correctly.

By assembling the 3' end ESTs of one cDNA library, the redundancy and complexity of different genes in this library could be determined. The 3' ESTs generated from cDNA clones of identical genes can be grouped together as a cluster. Those that have no significant matches to other ESTs within the library were considered singlets. The higher proportion of singlets a library exhibits, the higher complexity this library was considered to have (Hillier et al., 1996).

The percent of new genes can be viewed by determining the redundancy of the cDNA clones. The redundancy of a cDNA library can be determined by the output of 3' assembly of ESTs. Redundancy is the copies of one specific gene in one cell or tissue of one organism. Redundancy of gene expressed in some cell or tissue can be deduced from the 3' ESTs of the specific cDNA library of this cell or tissue. The redundancy of assembled ESTs was determined using formula (2) that was developed by Jim White in our informatics group.

$$\%\text{Redundancy} = \{1 - \frac{r}{(S/G + r)^2}\} \times 100 \qquad (2)$$

S : number of 3' EST sequences

G: the expected number of genes in the library

r: a redundancy factor, r>1. The larger the r is, the large the number of EST samples is required to reach a given sampling % of the library.

2.1.9.5 **Assembly of both 3' ESTs and 5' ESTs**

Even though 3' ESTs can be used as the signature to identify the genes, the assembly of only 3' ESTs has a disadvantage. 3' EST has limited coding information. It can not be used to reliably predict the function of novel sequences.

To acquire as much coding information and extend the function of novel sequences, both 3' and 5' EST sequences were assembled with the same high stringency parameters (minimum window size 14, minimum score 80). Two directories were produced following assembly of 3' and 5' ESTs: contigs_dir and singlets_dir, where the ESTs whose sequences did not overlap with the sequence of any other ESTs were stored in a singlets_dir and those ESTs whose sequences overlap with were assembled to generate a contig and stored in contigs_dir. The overall sequence of a contig is called a consensus sequence. The database that consists of both singlets and contigs represents the assembled-EST database. A partial annotation of the assembled EST database was performed by doing a BlastX analysis against the non-redundant protein database of GenBank using the singlet sequences and consensus sequences of the contigs as the queries. Only the BlastX search results of HSP>99 or p-value < e^{-4} were reported.

2.1.9.6 The BlastX search and the annotation of the two assembled EST databases

An important goal of any sequencing project is to find out as much biological function information as possible from these nucleotide sequences, such as the identity of genes and/or the information related to genes. A sequence similarity search is often used in the above attempts. BLAST (Basic Local Alignment Search Tool) is a power tool that is frequently used in all kinds of sequence similarity searches. Unlike a global alignment, in which a global similarity score between the entire lengths of the sequences being

compared is calculated, the local alignment method focuses on shorter regions of local similarity. It is a better and faster method compared to the global alignment method. It is also more sensitive for highly diverged sequences (Bioinformatics: a Biologist's Guide to Biocomputing and the Internet, unpublished, S.M., Brown, 2000). There are five BLAST programs and the BlastX is used to compare a nucleotide query sequence that is translated into all 6 reading frames against a protein sequence database. The purpose of a BlastX search therefore is to find potential translation products of an unknown nucleotide sequence based on the sequence similarity. The matches that BlastX finds from the searched databases represent the potential products of an EST sequence.

Several factors were considered when deciding to use BlastX as the sequence similarity tool in this EST research project. First, since all EST sequences in this research are partial sequences of cDNA clones, each EST represents only a partial region of an expressed gene. This information is different from the entire genomic sequence of a gene, as it does not contain information such as codon region, noncodon region and regulation regions. It should be noted that although the genomic DNA sequence of *Neurospora crassa* was not yet available when this work was done, there is a project at the Whitehead Institute to sequence this genome and partial genomic sequence data only recently has become available. For the sequence of protein, the same amino acid in different organisms may have different codon usage. Therefore, comparing proteins instead of the nucleotide sequence (DNA) is a more effective way of sequence similarity search. Second, the sequence of DNA is composed of only 4 bases (A, C, G, T) while the sequence of a protein could be composed of 20 different amino acids. Therefore, protein-protein comparison has a greater ability to find all related homologies than does

comparing to DNA-DNA. The Blast search using either a 6 frame translated DNA sequence as a query against a protein databank or a protein query against an all frame translated DNA databank is much more sensitive than BlastN search (DNA to DNA). In addition, the protein coding regions are usually conserved over longer evolutionary periods than the noncoding parts of the genome. The homologue based function prediction methods are more effective for protein sequences than for DNA sequences.

After the nr protein database (nr = all nonredundant GenBank CDS translations + PDB + Swiss-PROT) was downloaded to the Sun workstation, BlastX was automatically performed by running a script that was written by Hongshing Lai in our informatics group, the University of Oklahoma. Through running this script, the sequence of each assembled EST database member (the sequence of singlets or the consensus sequence of a contig) was used as query to search against the nr protein database one by one. The sequence of the query is in FASTA format. FASTA format is a method that stores DNA and protein sequences in a compact and simple way. It is a text file that can be read by virtually all molecular biology programs. In FASTA format text file, a sequence begins with a single-line header description. A ">" symbol is followed by this single-line header description. The sequence data follows this description line. After running BlastX, all the members of assembled database have three files, fasta format sequence file, .x file, and .table file (**Table 2.07**). This three file format was created by Hongshing Lai in our informatics group. The detail of these three files will be discussed in Chapter three (section 3.1.3.2). The output of a BlastX search includes program introduction, histogram of expectations, one-line summaries, and alignments. High-scoring Segment Pairs (HSP) is the fundamental unit of Blast algorithm output (Blast help manual on WWW BLAST

interface). An HSP consists of a pair of sequence fragments that are arbitrary but equal length: one from a database sequence and one is the query sequence. A local alignment of these two sequences is maximum in score. Their alignment score meets or exceeds a threshold or cutoff score. The modified Smith-Waterman or Sellers algorithms (Blast help manual on WWW BLAST interface) will identify these segment pairs, and the EXPECT or E-Value is the statistical significant threshold for reporting matches against database sequence.

Table 2.07 Three files resulting from the BlastX search of the GenBank nr protein database

NE.Contig1 file:

```
>NE.Contig1
TGAACTGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
TTTTTTTTTTGGGCAGAAAGCCGGGTCAATTCCTAGCTAGAACGGACCAG
TGTCCATGCCTATGTAATAGGGGTATCGGTTCTGGGGTAATGCTAAAGCT
GCCCTTCGAAATTCCAAGCCAAGTATGTCATAAAGAAGCCCTTGGCGAAC
TCCTTAGCCGGCCATTATCATCAAGTCATGACTGATAAAAAGGAAATGGA
AAACGACTGTTGGTAAAGTCGCTTAATGGGTAGCACCCTGCTTGGGGGCC
TCGGCCTTAACGTCGTGCTTGTTCTCGGAGACCTTGTCGGAGATGGCATC
ACCGGCAGCGTTGAAACAAGTGCCGACGCCCTGGTTGGAGTCCTTGGCAA
CGTCCTTGTTGGCCTCCTTGGAAGCAGTGGCGGTAGCGCCCTGGACCTTG
TCACCGACGTAATTGGCAGCGTTCTTG
```

NE.Contig1.x file:

```
Query=  Contig1
        (467 letters)
Translating both strands of query sequence in all 6 reading frames
Database:  /est/EST/nr
           356,412 sequences; 108,900,803 total letters.
Searching....10....20....30....40....50....60....70....80....90....100% done

                                                      Reading  High  Probability
Sequences producing High-scoring Segment Pairs:         Frame Score  P(N)       N

sp|P22151|GRG1_NEUCR GLUCOSE-REPRESSIBLE GENE PROTEIN ... -2   313  2.4e-27    1
>sp|P22151|GRG1_NEUCR GLUCOSE-REPRESSIBLE GENE PROTEIN >gi|3014 (X14801) grg1
            gene product [Neurospora crassa]
            Length = 71
  Minus Strand HSPs:
 Score = 313 (110.2 bits), Expect = 2.4e-27, P = 2.4e-27
 Identities = 63/67 (94%), Positives = 63/67 (94%), Frame = -2
```

```
Query:   466 KNAANYVGDKVQGATATASKEANKDVAKDSNQGVGTCFNAAGDAISDKVSENKHDVKAEA 287
             KNAANYVGDKVQGATATASKEANKDVAKDSNQGVGT  NAAGDAISDKVSENKHD KAEA
Sbjct:     5 KNAANYVGDKVQGATATASKEANKDVAKDSNQGVGTRLNAAGDAISDKVSENKHDAKAEA 64
Query:   286 PKQGATH 266
              KQGATH
Sbjct:    65 HKQGATH 71
```

NE.Contig1.table file:

```
 313   2.4e-27    Contig1        266      466 sp|P22151|GRG1_NEUCR GLUCOSE-REPRESSIBLE GENE
PROTEIN >gi|3014 (X14801) grg1 gene product [Neurospora crassa]
```

The default E-value is 10. It means that 10 matches are expected merely by chance, according to the stochastic model of Karlin and Altschul (Altschul,1990).

P-value is the probability of finding at least one HSP with score >=S.

$P = 1\text{-}e^{-E}$ (P-value associated with the score S)

If three HSPs with score >=S are expected, then the probability of finding at least one is 0.95. The lower E-values result in a higher stringency and lead to fewer chances that the matches are reported. The cutoff score also is a threshold for reporting high-scoring segment pairs. The default cutoff score is calculated from the EXPECT value. The higher CUTOFF value, the more stringent it is. Fewer chance matches will be reported. The S score of the HSP is a scale of which a single HSP satisfy the significance EXPECTATION E by itself. The higher the S value it is, the lower the probability of chance matches.

The description of the sequence and the summaries of blast research results are provided in one line which is useful information in the biological function assignment (Audic and Claverie, 1997; Pandey and Lewitter, 1999; Schmitt et al., 1999; Okubo et al., 1995; Lee et al., 1995). The first column is the identifier for the database sequences. The following is the list of several commonly used identifiers for the database sequences.

gi GenInfo Integrated database

gb	GenBank at NCBI
gp	GenPept of GenBank
emb	EMBL
dbj	DDBJ
sp	Swiss-PROT
pdb	Brookhaven Protein DataBank

The "High Score" column represents the score (S score) of the highest-scoring HSP that was found with each database sequence (subject sequence). "P (N)" column contains the lowest P-value ascribed to any set of HSPs for each database sequence. N is the number of HSPs in the set that was ascribed to the lowest P-value. P-values are a function of N as used in Karlin-Altschul "sum" statistics. The probability ranges 0-1.

2.1.9.7 The strategy of biological function assignment

The first step in the process of the analysis of the *Neurospora crassa* assembled EST databases was to organize the singlet and cluster members of the two assembled-EST databases by their biological function in the metabolism of the fungal organism. This was done based on the Riley schema. Prior to this step, all assembled EST database members of the two *Neurospora crassa* cDNA libraries were examined by a BlastX search. If the HSP > 99 or the p value < e-4, this query sequence was thought to have a significant homologue in Genbank. **Figure 2.10** shows the process of developing the biological function classification for the assembled EST database used in this research.

Initially, a preliminary keyword list called keyword list #I was obtained from the final keyword list used in *Aspergillus nidilans*. The primary keyword list used in the

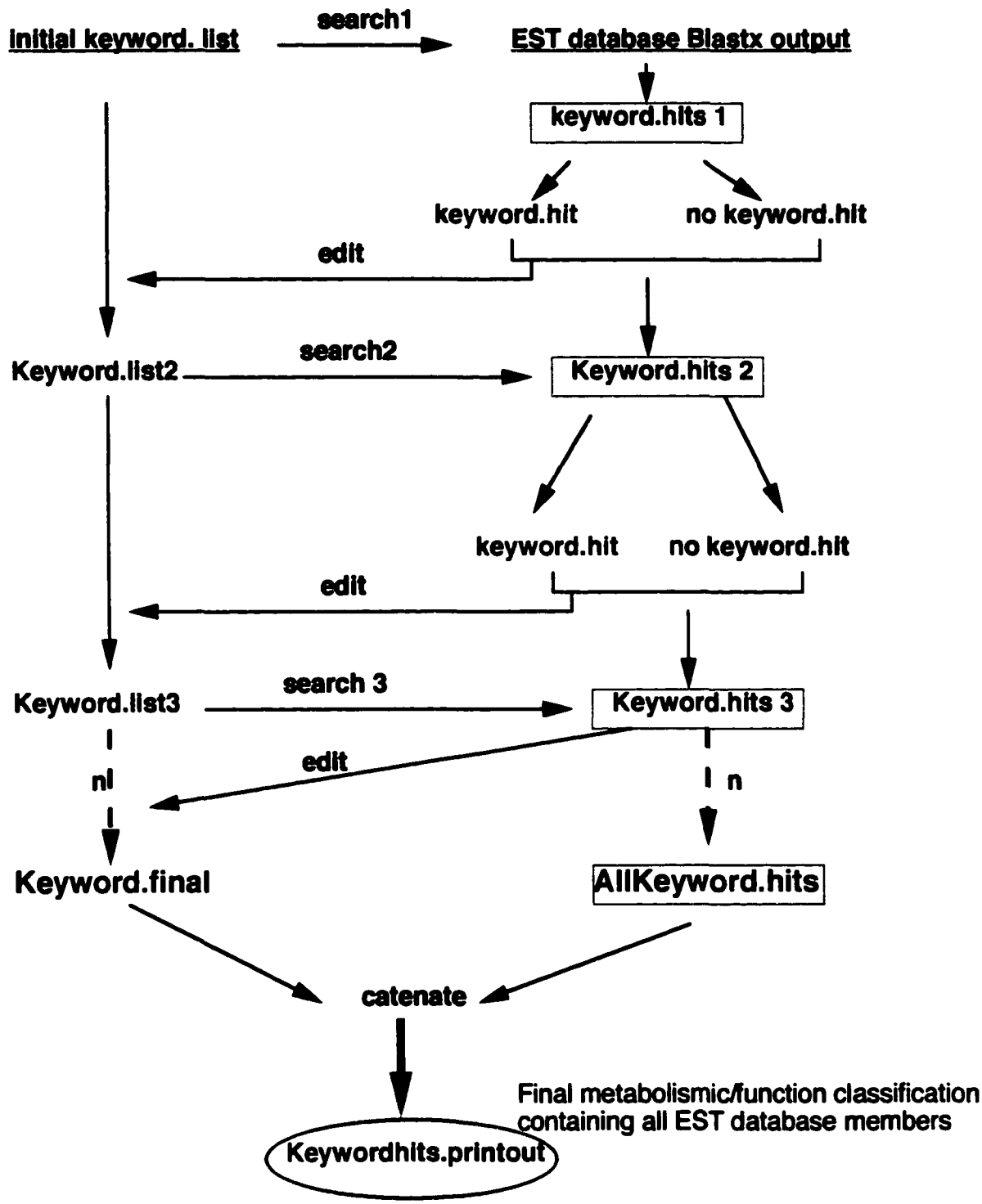


Figure2.10 *Neurospora crassa* biological function classification outline. Scripts and searches were done by Jim White in ACGT, the University of Oklahoma. n indicates the repeated cycles of keyword edit.

analysis of *Aspergillus nidulans* was developed by using a variety of sources such as the major metabolic pathway and enzymes from biochemical textbooks (Stryer, 1995), pathway enzymes from the metabolic pathways database (Selkov. Jr., 1998) and the pathway enzymes of KEGG (Kyoto Encyclopedia of Genes and Genomes) (http://www.genome.ad.jp/kegg). A series of scripts designed by Jim White in our informatics group, was performed to automatically screen the EST BlastX .table output files containing the BlastX output descriptors for the 5 top hits by using the above preliminary keyword list as query. If the BlastX output descriptor in these files matched the keyword in keyword lists, a new output file named keyword.hits that organized the contigs and singlets in numerical order was created to display the matches. For some contigs and singlets, there is no keyword in the new output file. That was indicated by a blank brace "<>". Therefore, manual editing was involved to examine the Blast descriptor lines for a new keyword. This was selected and put in "<>" as keyword for this contig. This new keyword was inserted into the right position in the keyword list according to its biological function. Due to the possibility of multiple functions of one enzyme or protein in the cell activities and physiological activities of organisms, one keyword might be assigned into different biological function categories during the manual editing process of keyword list. When the descriptor in the BlastX output file of a contig or singlet matches more than one keyword, only the top one which has the highest HSP score and the smallest p value and was the most specific in function description was kept in the keyword.hits file. All the contigs and singlets in the keyword.hits file which had significant homologs were examined. All the "<>" were filled with a selected best word as their keywords. Then, the selected keywords were put into a specific position in

the keyword list after carefully searching for the biological functions of the homologues of these contigs or singlets. The keyword list was corrected to yield the keyword list#2. The new keyword list#2 was used as query to screen again against the EST database BlastX output descriptor lines by using James White's scripts to search the matching keywords and produced the new output file named keyword.hits#2. After two cycles of this process, another output file named nonkeyword.hits was resulted as well. Keyword.hits is a list of contigs and singlets that have matching keywords (best words) in their BlastX descriptor lines. Nonkeyword.hits is a list of those contigs and singlets that have no matching keywords. The manual examination and editing were repeated for several cycles (n in the figure 2.10 is the number of cycles repeated) until no further "<>" appeared in the new output file of both keyword.hits and nonkeyword.hits. The keywords.final file was produced when the last keyword list was integrated into the categories of cellular functions outline by manual editing. The last keyword.hits and nonkeyword.hits were catenated into an All.hits file. Finally, the All.hits file was merged with the keywords.final to create the morning.printout and evening.printout separately, which contained the biological function classification including both the contig and singlet members of the assembled EST databases of the two *Neurospora crassa* cDNA libraries. The files of keywords.final for the morning and evening cDNA libraries are listed in Appendix I and II. The resulting output files, the morning.printout (Appendix III) and evening.printout (Appendix IV) can be accessed on our website at http://www.genome.ou.edu/oracle-test/.

2.2 Sequencing of two *Neurospora crassa DNA cosmids nc14, nc17*

The initial part of sequencing of these two cosmids was done by Jennifer Gray. Since the phrap coverage was only 3.0, these sequences were completed as part of this dissertation research. To complete these sequences, new subclone libraries were established and sequenced.

2.2.1 Overall sequencing strategy

The random shotgun sequencing approach was used in the sequencing of *Neurospora crassa* cosmids. **Figure 2.11** shows a summary of the processes of sequencing used in this research. The isolated cosmid DNA was nebulized, size-selected, subcloned, amplified, isolated and sequenced from the both ends of the DNA inserts. The details of each step are explained in the following sections. All of protocols used in this research were posted on the website at http://www.genome.ou.edu/protocol.

2.2.1.1 Large scale DNA isolation

Cosmid DNAs were isolated using a cleared lysate method followed by double acetate precipitation (Chen, 1997; Roe, 1997). A smear of colonies of the cosmid were picked and transferred into a 12×75 mm Falcon tube containing 3 ml of LB medium (10 g Bacto-Tryptone, 5g Bacto-yeast extract, and 10 g NaCl in 1L H_2O, autoclaved) with the appropriate antibiotic. After incubating at 37°C for 8-10 hours with 250 rpm shaking, the culture was transferred to a 250 ml flask containing 50 ml of the same medium and incubated for 8-10 hours under the same conditions. Then, the 53 ml of the cell culture was transferred into a 2 liter flask containing 1 liter of the same medium and incubated for an additional 8-10 hours. After harvesting the cells by centrifugation at 5000 rpm for

15 minutes in a 500 ml bottle in a Sorvall RC5-B using the GS3 rotor, cell pellets were frozen and stored at –70°C.

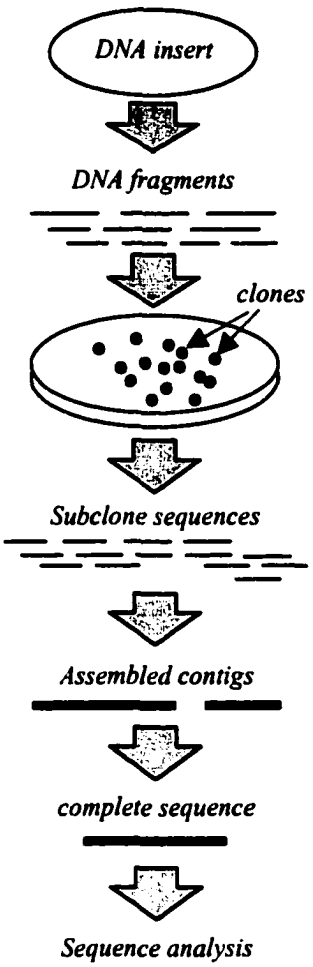


Figure 2.11 Flowchart of the sequencing of *Neurospora crassa* cosmids

Prior to use, the cells from 500 ml growth were thawed and resuspended in 20 ml of 10mM of EDTA, pH 8.0 by gently pipetting up and down with a 10ml pipette without vortex. Cells were resuspended completely. After mixing gently, the solution was incubated at room temperature for 5 minutes.

To resuspend cells, 40 ml of alkaline lysis solution (0.2 N NaOH and 1% SDS) was added. Gently swirling cells until the solution was homogenous, it was incubated for

5 minutes at room temperature. Immediately, 30 ml of cold 2 M KOAc (made by mixing 50ml of 7.5M KOAc with 23 ml of HOAc and 127 ml of ddH2O, stored at 4°C) was added and mixed very gently by swirling the bottle several times. Then the bottle was placed in an ice-water bath for at least 5 minutes.

The above solution was filtered through cheese cloth. Then, the lysate was cleared from precipitated SDS, protein, membranes, and chromosomal DNA by centrifuging at 9000 rpm for 30 minutes at 4°C in the RC5-B using the GSA rotor. An additional centrifugation was performed to ensure that all insoluble were removed. The supernatants of two tubes (from one liter original cell grow) were transferred into one 500 ml bottle and then an equal volume of isopropanol was added and mixed by swirling. After centrifugation at 5000 rpm for 20 minutes in RCC5-B using GS3 rotor, the supernatant was decanted and the pellet drained.

The DNA pellet was quickly and gently dissolved in 9 ml of 10:1 TE (10 mM Tris-HCl, pH7.6, 1 mM of EDTA, pH8.0) 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 stored at –70°C for 30 minutes or over night. After the solution was thawed and centrifuged in Beckman GS-6R centrifuge at 2000 rpm for 10 minutes, the supernatant of each tube was transferred into 50 ml Corning centrifuge tube. Then, DNase-free RNase A (100 μg/ml) +T1 (40 units/100 ml) was added, followed by incubation in 37°C water bath for 1 hour. Then, 30 ml of 100 % cold ethanol was added into each tube. After mixing by inverting, the tubes were incubated in an ice-water bath for 15 minutes to overnight. The DNA pellet was formed by spinning at 3000 rpm for 25 minutes in Beckman GS-6R centrifuge. The pellet in each tube was washed with 30 ml of 70% ethanol and dried in

vacuum oven. The dried pellet usually was dissolved in 1ml of ddH_2O (generally, the DNA pellet from one liter original growth was dissolved in 2 ml ddH_2O). Next, the size of the cosmid DNA was estimated by agarose gel electrophoresis using HindIII digested λ DNA and HaeIII digested ϕX174 DNA as size marker.

2.2.1.2 DNA physical shearing by nebulization

The large cosmid DNA molecules were sheared physically into fragments of 1.5-4Kb using the nebulization method. 2 ml of sample containing 500μl large DNA insert, 500 μl glycerol and 100 μl ddH_2O were put into the bottom of a nebulizer (IPI medical products, INC, Chicago, IL). After mixing the glycerol with the DNA sample and water very gently and completely, the nebulization was performed at the condition of –5 °C, 6-8 psi for 2.5 minutes. Pressurized nitrogen gas was introduced from the top of the nebulizer. The sample solution was pushed by the flowing gas and collided with a protruding plastic surface and then broken into fine droplets. The size of the DNA fragments was inversely proportional to the viscosity and the temperature. The purpose of glycerol was to increase viscosity of sample solution as well as to prevent DNA's freezing when the temperature of the nebulizer was below 0°C (Bodenteich, 1993). Sodium chloride was used to reduce the temperature of ice water to –12°C. The nebulized sample was then collected to the bottom of the nebulizer unit by centrifugation at 1500 rpm in a tabletop centrifuge. The DNA sample was divided into 4 Eppendorf tubes with 450 μl in each tube. 1.0 ml of 95% ethanol (0.12 M NaOAc, pH 4.5) was added to precipitate DNAs. The tube was incubated on ice for 15-30 minutes, then, centrifuged for

30 minutes at 12,000 rpm. DNA sample was washed using 75% ethanol, then centrifuged again for 10-15 minutes at 12,000 rpm. The pellet was dried in a vacuum for 4 hours.

2.2.1.3 Random fragment end-repair and phosphorylation

The DNA fragments generated by nebulization have single-stranded ends. These ends were repaired and phosphorylated before they were ligated into pUC vectors.

DNAs in 4 above tubes were resuspended in	27 µl	ddH2O.
10X Kinase buffer	5 µl	
10 mM rATP	5 µl	
.25 mM dNTPs	7 µl	
T4 polynucleotide kinase	1 µl	(3 units/µl)
Klenow DNA polymerase	2 µl	(5 units/µl)
	47 µl	

Totally, 47 µl solution was in each tube. Each tube was incubated in a 37°C water bath for 30 minutes to ensure the phosphorylation of the 5' ends of DNA fragments.

The reaction mixture from the above tube was loaded into one well of a 1% low melting temperature agarose gel (300 ml of 1X TAE, 3g of low melting temperature agarose, 70 µl of ethidium bromide (EtBr)) with 10 µl of agarose loading dye (0.02% bromphenol blue, 5 mM EDTA, pH8.0, 50% glycerol). HidIII digested λ DNA and HaeIII digested ϕX174 DNA mixture was loaded in parallel with the sample lane as size markers. A distance between the nebulized DNA sample and size-marker DNA was needed to avoid contamination. Size fractionation was achieved by electrophoresis on a low melting temperature agarose gel for 1.5 hours at 120-150 mA. DNA fragments with the size in the range of 1-2 Kb and 2-4 Kb were quickly cut from the gel and put into clean 1.5 ml microcentrifuge tubes under a long wave length UV light in a dark room.

The gel pieces were stored at –70°C for at least 15 minutes or over night and then melted in the 70°C H_2O bath for 10 minutes. The melted agarose solution with the DNA fragment was divided into Eppendorf tubes equally with about 450μl in each tube. Then, equal volumes of TE-saturated phenol were added to each tube and the tube was vortexed at least 5 minutes in room temperature. Then, the tube was centrifuged at 12,000 rpm for 15 minutes at room temperature to separate the aqueous and phenol organic phases. The upper aqueous phase with the DNA sample was transferred into a clean tube. The phenol extraction was repeated at least twice. Then, the final aqueous solution was extracted at least twice by adding equal volumes of ether. The ether was saturated with an equal volume of ddH_2O. At this step, the DNA sample was at the bottom of tube. Finally, the purified sample was dried in a vacuum overnight to allow the ether to evaporate. 450 μl of dd H_2O was added to the Eppendorf tubes to resuspend DNAs. Then, 2–2.5X volume of 95 % ethanol/0.12M NaOAc was added to precipitate DNA. After centrifuging at 4°C for 30 minutes at 12,000 rpm, the DNA pellet was washed with 75% ethanol in room temperature for 5-10 minutes. Then, the clean DNA pellet was collected using centrifuge again at 4°C for 5-10 minutes at 12,000 rpm. After drying, the DNA sample pellet was resuspended in 20 μl –30 μl of ddH_2O. At this stage, the DNA sample was ready for ligation.

2.2.1.4 Shotgun library construction: ligation and transformation

The DNA insert was ligated to its vector (pUC). The recombinant DNA molecule was cloned into *E. coli* host cells for amplification through transformation.

2.2.1.4.1 Ligation

The ligation was performed according to the protocol as follows (table2.08).

Table 2.08 DNA Ligation Protocol (unit:μl)

DNA template	1	1(1/2)	1(1/4)	0	1pGEM
Vector pUC18	2	2	2	2	2
10Xligation buffer	1	1	1	1	1
T4 DNA ligase	1	1	1	1	1
Sterile H_2O	5	5	5	5	5
Total	10	10	10	10	10

The reaction mixture containing 100-1000 ng of DNA fragment, 20 ng of SmaI-linearized and calf intestinal alkaline phosphatase, dephosphorylated pUC18 vector, 1μl of 10X ligation buffer and 1μl of T4 DNA ligase, with sterile ddH_2O added to a total volume of 10 μl, was incubated at 4°C overnight. Different DNA concentration were used in ligation to obtain an optimal result. pGEM was used as positive control while a negative control ligation was also designed in the absence of insert DNA fragment to test the possibility of vector self ligation.

2.2.1.4.2 Transformation

The recombinant plasmids were transformed into the *E. coli* XL1 Blue strain to obtain colonies from the shotgun library for sequencing template isolation. The details for competent cell preparation for electroporation (Sharma and Schimke, 1996) were described before (Roe et al., 1997; Pan, 1996). The competent cells used in this study were prepared by Mounir Elharam and Rose Morales-Diaz and stored at –80°C..

Protocol for <u>electrotransformation:</u>

1. 1 to 2.5µl of DNA was added to one 40 µl aliquot of competent cells. The tube was mixed well and incubated on ice for approximately 1 minute.
2. The mixture was transferred to a cold 0.2 cm electroporation cuvette (Bio-Rad, Hercules, CA, USA) and electroporated in a Bio-Rad Gene Pulser at 2.5 V.
3. The cuvette was removed from the chamber and 1 ml of YENB medium was immediately added to the cuvette to resuspend the cells. Then, the cell suspension was transferred to a small Falcon tube (12×75mm).
4. The cell was incubated in 37°C shaker for 30 minutes at 250rpm and harvested by centrifuging for 10 minutes at 2500 rpm.
5. 30 µl IPTG (25mg.ml) and 30 µl X-gal (20mg/ml) were added to the tube to resuspend the pellet. Then, cells were plated on LB plates with 100µg/ml ampicillin.

The ampicillin resistance gene present on the pUC vectors allows transformed cells to grow in the presence of ampicillin while nontransformed cell can not grow in the media containing ampicillin. The white colonies containing recombinant plasmids were selected and grown for further use.

2.2.1.5 Semi-automated 96-well subcloned DNA isolation

The strategy used to isolate the subcloned DNA was similar to that used for isolation of plasmid DNA templates in the EST project that was described in Section 2.1.4. All subclone DNA template isolations were performed on the Biomek2000 with the program "td TE_RNaseA-to-END".

2.2.1.6 Thermocycler sequencing reaction and data collection

The thermocycler sequencing reaction and data collection used for sequencing *Neurospora crassa* cosmid subclones also were the same for the two cDNA libraries of *Neurospora crassa* and were described in section 2.1.5.3 and section 2.1.6.

2.2.1.7 Shotgun sequence edit and assembly

The random shotgun sequences generated from the subcloned DNA averaged about 500 bp in length but because their positions in their original large insert DNAs are uncertain. The Phrap computer program was used to assembly these shotgun sequences into larger continuous sequences and build a consensus sequence. To estimate the progress of sequencing, the information for the average coverage of each base by counting gel reading and the number of sequence gaps remaining resulted from an assembly was used to determine whether or not more shotgun sequencing is needed and when the gap closure stage should begin. Gap closure for this project was started when the Phrap coverage was 5-6 fold. In this part of the dissertation research, the programs Phred and Phrap developed by Phil Green at University of Washington were used as assembly tools (Green, 1998; Green and Ewing, 1998). Both Phred and Phrap are in a program package that also includes Cross-match program and Consed (Consensus sequence editing program) (Gordon, Abajian and Green, 1998).

Phred is the editing part of the phred/phrap package that contains the base calling program. Phred also calculates the ratio of signal to noise for each base, and the ratio of the peak to peak spacing to produce a numerical quality value for each base. This quality score is used in the input file for Phrap, the assembly program. Therefore, the Phrap also

contains quality screen for each base, compared to other assembly tools such as XGAP (X-window genome assembly program)(Gleeson and Staden, 1991), which do not contain base quality value.

Phrap, the assembly program, therefore uses the quality values generated from Phred, to combine the overlapping high quality sequence reads. The mosaic that is constructed by these contiguous sequences is called a Contig.

Cross-match also is a program that was developed by P. Green which is designed for homolog comparison. It can be used to compare any two sets of DNA sequences and do alignment while performing comparison. All vector sequences contained in a vector file in fasta format, mitochondria sequence, ribosomal RNA sequence and *E. coli* sequences were screened out using cross-match.

Consed is a program that can view the output from phred/phrap and edit the result of the assembly. To view and edit the assembly, at least three types of input files are required by consed: the chromatogram file, the phd file, and the .ace file. The chromatogram file is generated by the sequencer such as ABI 377, and contains the fluorescence trace profiles. The phd file is generated by Phred, and contains the quality value, trace peak positions for the read bases, any tags attached to each read and the phred base calls. Each sequence read has its own separate chromatogram file and phd file. The .ace file is the third Phrap output file. All the assembly information such as sequences and quality values of contigs, and read information, is contained in this .ace file. After reading the .ace files and the phd files, consed (Gorden et al., 1998) will open a window to display the list of contigs and reads. The quality of read and consensus sequence is displayed to users by the color of each base in window. For example, an

orange colored base disagrees with consensus and a white colored base represents a high quality match. The whiter the background on the window, indicates a higher quality alignment.

2.2.1.8 Gap closure of cosmid sequencing

Since the sequencing data collected in the shotgun stage were generated from the randomly cloned fragments, some regions of the original DNA insert were not covered in sequenced subclones. In addition, GC-rich or polynucleotide repeat regions often are difficult to sequence because of their possible secondary structure in sequence. These factors are the reasons why the sequence gaps exist.

Primer walking and PCR-based methods (individual PCR and multiplex PCR (MPCR)) are the basic strategies used to close gaps. The closure and finishing process begins by using consed to view the database and examine the gel reading flanking a gap to see whether the gap resulted from GC-rich region or polynucleotide repeat region that was difficult for the DNA polymerase to read through. In addition, the program exgap written by Dr. Axin Hua in Dr.Roe's lab is used to find any available subclones covering the gap that can be used as templates for primer walking. If there was no subclone covering the gap, polymerase chain reaction will be performed to amplify the region using custom synthesized primers picked from the both ends of the gap using an automatic program joined with consed and the cosmid as template. By primer walking and PCR, the gaps in these two cosmids were reduced to one. The last gap is a GC-rich region.

PCR reaction mixture contains:

Cosmid DNA template 1 μl (1-20 ng)

Primers	1 µl of each(13µM or 26 µM)
AmpliTaq polymerase	1 µl (5U)
10X PCR buffer	10 µl (500 mM KCl; 100 mM Tris-HCL, pH7.6; 10mM MgCl)
2 mM dNTP	10 µl
DMSO (final 5%)	3 µl
H_2O	24 µl
	50µl

When needed, additional sequencing under different reaction conditions were performed to close difficult regions. These conditions included, increasing the reaction temperature, denaturation of secondary structure formed in the GC-rich or repeat region in the template using 5%-10% DMSO (Winship, 1989), formamide or glycerol may allow DNA polymerase to read through a difficult region. Also, a new reaction premix in which dGTP replaced dITP in the regular reaction pre-mix was introduced during this research period and was useful to sequence through G-rich regions.

2.2.2 Sequence analysis of the *Neurospora crassa* cosmid projects

The programs used in this part are different from that in the analysis of ESTs project. The sequences from cosmid projects are genomic DNA sequences. Like all other eukaryote genomes, the genome of fungi also has repeat sequences and noncoding sequences which usually are not present in ESTs, since they have been removed by splicing. In an EST project, the main purpose of sequence analysis is to characterize the gene products identified, to study gene expression and relate the distribution of expressed genes to biological function. In the analysis of genomic DNA sequences, generally speaking, after the final high quality sequences are available, these sequences are

analyzed for the presence of genomic feature, i.e. to predict genes, predict protein motifs or function domains, to find any cis-acting regulatory elements and to locate any trans-acting factor binding sites in the DNA sequence. Genome sequencing a tool to study the biological functions implied by these sequences. Thus it is a key to open the door of the kingdom of the knowledge about the structure and functions of a gene and its protein product in the cell.

Several different computer programs have been developed for EST and genomic analysis. In this research, emphasis was placed on the identification of genes expressed in two different cDNA libraries of *Neurospora crassa* and comparison of the genes differentially expressed in these two libraries. The following computer programs were used in the analysis of the EST and cosmid sequence data: BLAST (Altschul et al., 1990), powerblast (Zhang and Madden, 1997), profileScan (Gribskov et al., 1987; Bucher et al., 1996), NNPP (Neural Network Promoter Prediction) (Reese and Eeckman, 1995), TSSW (Solovyev and Salamov, 1997) and Promoter2.0, GENSCAN (Burge and Karlin, 1997), GenLiner (Hua, 1999), hydrophobicity plot (Kyte and Doolittle, 1982), BLOCK (Henikoff and Henikoff, 1994) and a series of GCG programs including the BESTFIT, COMPARE, TRANSLATE, and the DOTPLOT program (GCG Wisconsin Package version 8.0, Genetics Computer Group, Madison, Wisconsin) (Sonnhammer and Durbin, 1995).

1.) BLAST and powerblast are programs used for database similarity searches. They help reveal matches to known genes from other species. If a homology exists, the genomic structure of those genes will be deduced by sequence alignment of genomic sequences with sequence of genes and or cDNA in the databases in GenBank.

2.) The BESTFIT program was used for DNA or amino acid sequence alignment. TRANSLATE was used to translate DNA sequence to amino acid sequence. The COMPARE and DOTPLOT were used to graphically display the homology comparison for two sequences.

3.) profileScan was used in exploring potential features and/or functions of known or predicted genes such as protein motif.

4.) NNPP, TSSW (Solovyev and Salamov, 1997) and Promoter2.0 were used for promoter predicton from the upstream sequence of genes. NNPP predicts promoters using a neural network. TSSW and other promoter-prediction program predict promoters based on the sequence features of TATA, CAAT and other transcription factor binding sites. The results of these programs are quite different when predicting promoter from same DNA sequence.

5.) The BLOCK database was used to search possible domain or motif from protein.

6.) Hydrophobicity plots were used to predict the possible membrane-spanning segments from protein sequence.

7.) GENSCAN was used to predict genes or exons. GenLiner was mainly used together with GENSCAN to view and edit the output of GENSCAN, and to view predicted genes and exons.

8.) Sequin was used to edit and submit cDNA sequences and genomic DNA sequences to GenBank.

Chapter III

Results and Discussions

3.1 Two cDNA libraries and EST databases of *Neurospora crassa*

In an attempt to study the biological clock in *N.crassa*, the difference in gene expression at two different strains were analyzed by sequencing ESTs from the two representative EST databases were established and analyzed. The wild type strain *frq*+ was used for the construction of the evening cDNA library an the long period strain frq^7 was used for the construction of the morning cDNA library. At continuous darkness for 43 hours, the wild type *frq*+ represents clock time 13 (CT13) while the frq^7 strain is equivalent to a clock time of 1 (CT1).

3.1.1 EST database development

The first step in establishing an EST database is to generate EST sequences. As described in the Materials and Methods section in the second chapter of this thesis, these ESTs were generated by sequencing both the 3'-end and 5'-ends of one cDNA clone. The sequence from the 3'-end and 5'-end of one cDNA clone are called 3' EST and 5' EST, respectively. The quality of each EST was checked and all low quality ESTs were screened out before further analysis. The sequences that passed the quality check were used to search for homology in the non-redundant protein database at the NCBI. A separate EST database was established for both the *Neurospora crassa* evening cDNA library and the morning cDNA library.

3.1.2 The summary of EST data of two *Neurospora crassa* cDNA libraries

A summary of EST sequences resulting from two cDNA libraries is shown in Table 3.01. All ESTs referred to from this point on in this dissertation are high quality ESTs except where indicated. 10871 high quality ESTs were generated from the *Neurospora* morning cDNA library by end sequencing 7187 cDNA clones, while 9148 high quality ESTs were generated from the *Neurospora* evening cDNA library from 5922 cDNA clones. The percent of high quality ESTs in the *Neurospora* morning and evening

Table 3.01 EST data summary of the two *N.crassa* cDNA libraries

	Morning library		**Evening library**	
Total number of ESTs	10871		9148	
5′ EST only	2110		1502	
3′ EST only	1388		1194	
5′ + 3′ ESTs	3689		3226	
#clones sequenced	7187		5922	
	5′	**3′**	**5′**	**3′**
High quality	5799	5077	4728	4420
No pass	711	538	1462	1069
Total	6505	5615	6190	5489
Success	89%	90%	76%	81%
Clip/Clean no pass				
Completely vector	194	61	164	169
Too Short (<100bp)	205	175	368	432
Poor quality	197	278	492	346
Wrong end	18	3	43	6
E.coli	22	2	43	12
Mitochondrial	0	0	3	0
Ribosomal	75	19	349	104

cDNA libraries were ~90% and ~78.5%, respectively. In the morning cDNA library, 2110 clones had only 5' ESTs while 1388 clones had just 3' ESTs and 3689 clones had both 5' ESTs and 3' ESTs. In the evening cDNA library, 1502 cDNA clones generated only successful 5' ESTs, 1194 clones had only 3' ESTs and 3226 clones had both the 3' ESTs and 5' ESTs. Totally, 5077 of 3'ESTs and 5799 of 5' ESTs were produced from the morning cDNA library while 4420 3' ESTs and 4728 5' ESTs were generated from the *Neurospora* evening cDNA library. The number of 5' ESTs is slightly higher than that of 3' ESTs in both cDNA libraries. This phenomenon, which also has been observed before (Kupfer, 1999; Hillier, 1996), is due in part to the lower successful sequencing ratio of 3' ESTs because the poly A region (20-100 residue A) at the 3' end of a cDNA clone is difficult to sequence through.

From **table 3.01**, it can be seen that several hundred to more than one thousand ESTs did not pass the clip and clean processing scripts in the two cDNA libraries. These include the ESTs that are categorized as completely vector, as too short a sequence read length (<100 bps), poor quality sequences, *E. coli* host genomic sequences, mitochondrial sequences and ribosomal RNA sequences. From the categories of no pass ESTs in **table 3.02**, it is obvious that there are many more ESTs sequences where are vector and ribosomal RNA sequences in evening library (333 vectors, 453 ribosomal RNA sequences) than in morning cDNA library (255 vectors and 94 ribosomal RNA sequence). This result may indicate that the cDNA morning library is of higher quality, i.e. more representative of cDNAs, than that of the evening cDNA library. This agrees with the data presented above that showed that the morning library had a higher % of

cDNA insert (68%) than did the evening cDNA library (24%), and the morning library has lower % of vector that the evening library does.

Table 3.02: No pass ESTs in two *Neurospora crassa* cDNA libraries

category	EST# (NM)	% (NM)	EST# (NE)	% (NE)
completely vector	255	2.35	333	3.64
Too short (<100bp)	380	3.50	800	8.75
Poor quality	475	4.37	838	9.16
Wrong end	21	0.19	49	0.54
E. coli	24	0.22	55	0.60
Mitochondrial	0	0	3	0.03
Ribosomal	94	0.86	453	4.95
Total	1249	11.49	2531	27.67

3.1.3 EST database analysis

The establishment of the EST database and the strategy of EST database analysis are the most fundamental and important parts of this research. How the EST database is produced will affect the quality or reliability of its EST data. The strategy of analysis of the EST database will determine or restrict the quantity and quality of the information extracted from these EST data. Therefore, a model the model of the EST database and EST database analysis used in this research was adapted from that of the Genome Research Center of Washington University at St. Louis after modifications by Dr. Kupfer and James White in our laboratory at the University of Oklahoma. The main goal of this research was to study the gene expression profiles in two time-of-day-specific cDNA libraries from *Neurospora crassa*. This was accomplished by employing 3' alone, 3' and 5' combined EST assemblies. Finally, the differences of the expression of the clock-

controlled genes in these two cDNA libraries were compared.

3.1.3.1 Assembly of 3' ESTs

Table 3.03 and **3.04** list the number of singlets and clusters after every 200 3' ESTs sequences were collected and assembled from the two *Neurospora crassa* cDNA libraries. Since singlets are ESTs that their sequences do not overlap with that of other ESTs, as they could not be grouped with others during the alignment of the EST into one cluster or contig. Therefore, the number of singlets plus the number of contig pairs represents the numbers of genes expressed in this cDNA library.

Table 3.05 and 3.06 are the deduced redundancy of the genes sampled in two cDNA libraries of *Neurospora crassa* by cumulation of 3' ESTs. The last value of the G and r in the table 3.05 and 3.06 were used in the calculation of the redundancy. G indicates the expected number of genes in the library. r is a redundancy factor. As more the EST samples are assembled, the values of these two parameters become more representative. Therefore, only the final G and r values were used in the calculation.

Figure 3.01 shows the plots of the total number of 3' ESTs against the percent of redundant sequences in two Neurospora cDNA libraries. The tendency of redundancy is inversed to that of the possibility of finding new genes in the future assembly of this library. From the **Figure 3.01**, it can be seen that at the beginning of sampling, because most of the EST sequences were newly detected they represented new genes, the rate of the increase of the redundancy is very fast. Therefore, the shape of curve is very steep. This implies that the large redundancy is gained very quickly with smaller size of

Table 3.03 3' EST assembly and the percent of new genes in *Neurospora* cDNA morning library

Total Reads	Singlets (S)	Contigs (C)	S + C	% New Genes
200	61	47	108	54.0
400	69	100	169	30.5
600	70	126	196	13.5
800	86	150	236	20.0
1000	97	175	272	18.5
1200	84	201	285	6.5
1400	91	216	307	11.0
1600	93	231	324	8.5
1800	90	244	334	5.0
2000	85	255	340	3.0
2200	85	265	350	5.0
2400	82	273	355	2.5
2600	82	276	358	1.5
2800	84	282	366	4.0
3000	84	288	372	3.0
3200	79	292	371	0.5
3400	78	299	377	3.0
3600	76	307	383	3.0
3800	71	313	384	0.5
4000	66	318	384	0
4200	61	327	388	2.0
4400	61	329	390	1.0
4600	60	333	393	1.5
4800	58	335	393	0
5000	59	341	400	3.5

$$\% \text{ New} = \frac{\{(\text{New } (S+C) - \text{old } (S+C)\} X\ 100}{\#\text{ of New Reads}}$$

Table 3.04 3' EST assembly and the percent of new genes in *Neurospora* cDNA evening library

Total 3' ESTs	Singlets (S)	Contigs (C)	S + C	% New Genes
200	123	26	149	74.5
400	212	58	270	60.5
600	264	92	356	43.0
800	289	138	427	35.5
1000	330	176	506	39.5
1200	346	213	559	26.5
1400	404	249	653	47.0
1600	444	291	735	41.0
1800	481	331	812	38.5
2000	500	373	873	30.5
2200	506	415	921	24.0
2400	497	452	949	14.0
2600	488	483	971	11.0
2800	484	508	992	10.5
3000	476	531	1007	7.5
3200	465	558	1023	8.0
3400	461	578	1039	8.0
3600	447	610	1057	9.0
3800	449	626	1075	9.0
4000	441	649	1090	7.5
4200	441	670	1111	10.5
4400	437	688	1125	7.0

$$\% \text{ New Gene} = \frac{\{\text{New } (S + C)\text{-old } (S+C)\} \times 100}{\#\text{ of New Reads}}$$

Table 3.05 Redundancy of the genes sampled in the *Neurospora* cDNA morning library by cumulation of 3' ESTs

Total 3' ESTs	G	r	Redundancy (%)
500	388	1.3368369	77.8894
1000	437	1.5067519	89.0258
1500	458	1.5670153	93.4611
2000	471	1.6073594	95.6645
2500	469	1.5977894	96.9163
3000	463	1.5752405	97.6590
3500	460	1.5561679	98.2120
4000	457	1.5423101	98.5728
4500	456	1.5321786	98.8344
5000	**454**	**1.5244313**	99.0302

Table 3.06 Redundancy of the genes sampled in the *Neurospora* cDNA evening library by cumulation of 3' ESTs

Total 3' ESTs	G	r	Redundancy (%)
500	1437	1.203082	46.7689
1000	1226	1.184551	62.9600
1500	1903	1.357425	72.7480
2000	2263	1.437325	79.1128
2500	2237	1.430456	83.4827
3000	1942	1.337851	86.6121
3500	1809	1.280970	88.9296
4000	1704	1.225988	90.6935
4500	**1671**	**1.205911**	92.0672

$$\text{Redundancy} = \{ 1 - \frac{r}{(S/G + r)^2} \} \times 100$$

S: number of 3' EST sequences
G: the expected number of genes in the library
r: a redundancy factor, r>1. Larger r indicates an increased number of samples needed to reach a given sampling % of the library
Note: final value of G and r in the table (bold) are used in the redundancy calculation.

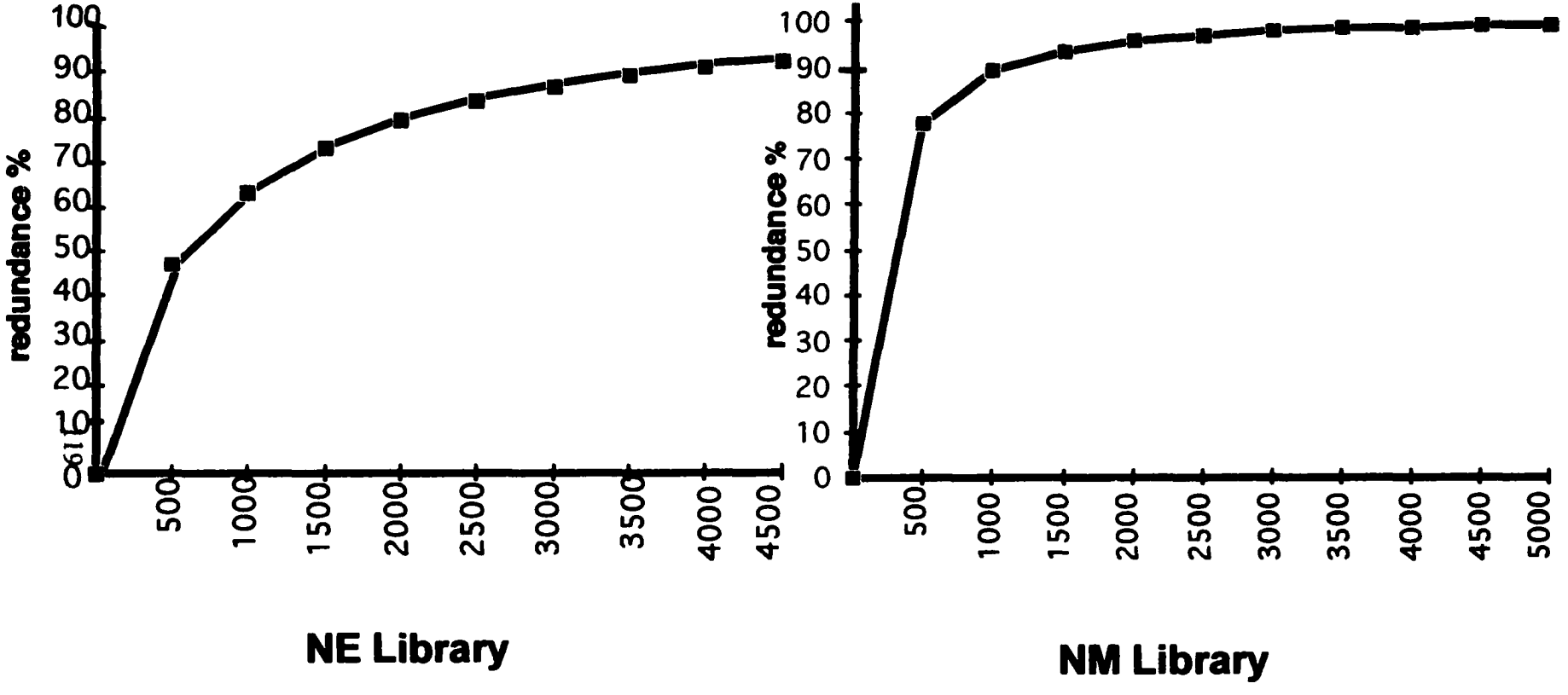


Figure 3.01 3' ESTs Assembly with Phrap and the Redundancy of the Libraries

resampled ESTs. When the sample size increase beyond 1000, the rate of increase in the number of redundant sequences is much slower. When the sampling size reached 4500 ESTs, the redundancies of both libraries were greater than 90%. However, the shape of curve of these plots is quite different for the two cDNA libraries. In the *Neurospora* cDNA morning library, when the sample size is 500, the redundancy is about 80% and after the number of 3' ESTs assembled exceeded 1000, the redundancy had almost achieved its maximum. This means that the library was exhausted of new genes because when 5000 3' ESTs were assembled, the redundancy was 99.2 %. Therefore, the variety of gene species in this library was very limited and several mRNA species had a very high copy number in this library . In the *Neurospora* cDNA evening library, when the sample size reached 2000, the redundancy level was 70%. After 4500 3' ESTs were assembled, the redundancy was beyond 92%. At this point, a plateau still did not occur as it did in the cDNA evening library. This indicates that the variety of genes species in the morning cDNA library was much larger than it was in the evening cDNA library.

One purpose of this research was to identify the genes expressed under control of *Neurospora* clock both quantitatively and qualitatively, therefore to discover new clock-controlled genes (*ccgs*). Thus, an attempt was made to assemble as many samples as possible and thereby obtain rarely expressed genes without using normalized and subtracted libraries (Bonaldo et al., 1996; Adams et al, 1995).

3.1.3.2 Assembly of both 3' ESTs and 5' ESTs

Through the assembly of both 3' ESTs and 5' ESTs, an assembled-EST database

was established for both cDNA libraries. The assembled-EST database consists of contigs and singlets. Using scripts developed by Hongshing Lai, a BlastX search was performed against the non-redundant protein database of GenBank in batch mode.

After the BlastX search, each singlet and contig resulted in three files. For example, with NE.Contig1, the first contig in the evening library, the 3 files in the Neurospora cDNA evening assembled EST database were NE.Contig1, NE.Contig1.x, NE.Contig1.table. The first file, NE.Contig1 contained the consensus sequence of this contig in fasta format. The second file, NE.Contig1.x, contained the complete BlastX output of NE.Contig1, the length of the consensus sequence, all the homolog identity information found in GenBank including its HSP score and their p-values as well as the information of the assignment of amino acid sequence between the subject and the query. The third file, NE.Contig1.table, contained a table with the BlastX header lines only (Table 3.07).

Table 3.07 The output files of the BlastX search of the nr protein database of Genbank

NE.Contig1 file:

```
>NE.Contig1

TGAACTGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
TTTTTTTTTTGGGCAGAAAGCCGGGTCAATTCCTAGCTAGAACGGACCAG
TGTCCATGCCTATGTAATAGGGGTATCGGTTCTGGGGTAATGCTAAAGCT
GCCCTTCGAAATTCCAAGCCAAGTATGTCATAAAGAAGCCCTTGGCGAAC
TCCTTAGCCGGCCATTATCATCAAGTCATGACTGATAAAAAGGAAATGGA
AAACGACTGTTGGTAAAGTCGCTTAATGGGTAGCACCCTGCTTGGGGGCC
TCGGCCTTAACGTCGTGCTTGTTCTCGGAGACCTTGTCGGAGATGGCATC
ACCGGCAGCGTTGAAACAAGTGCCGACGCCCTGGTTGGAGTCCTTGGCAA
CGTCCTTGTTGGCCTCCTTGGAAGCAGTGGCGGTAGCGCCCTGGACCTTG
TCACCGACGTAATTGGCAGCGTTCTTG
```

NE.Contig1.x file:

```
Query= NE.Contig1 (467 letters)
```

```
  Translating both strands of query sequence in all 6 reading frames
>sp|P22151|GRG1_NEUCR GLUCOSE-REPRESSIBLE GENE PROTEIN >gi|3014 (X14801) grg1
            gene product [Neurospora crassa]
            Length = 71
 Minus Strand HSPs:
 Score = 313 (110.2 bits), Expect = 2.4e-27, P = 2.4e-27
 Identities = 63/67 (94%), Positives = 63/67 (94%), Frame = -2
Query:   466 KNAANYVGDKVQGATATASKEANKDVAKDSNQGVGTCFNAAGDAISDKVSENKHDVKAEA 287
             KNAANYVGDKVQGATATASKEANKDVAKDSNQGVGT  NAAGDAISDKVSENKHD KAEA
Sbjct:     5 KNAANYVGDKVQGATATASKEANKDVAKDSNQGVGTRLNAAGDAISDKVSENKHDAKAEA 64
Query:   286 PKQGATH 266
              KQGATH
Sbjct:    65 HKQGATH 71
```

NE.Contig1.table file:

```
 313  2.4e-27   Contig1       266      466 sp|P22151|GRG1_NEUCR GLUCOSE-
REPRESSIBLE  GENE  PROTEIN  >gi|3014  (X14801)  grg1  gene  product
[Neurospora crassa]
```

The symbol "X" in the NE.Contig1 file indicates the vector sequences that were masked by script. The four numbers in the NE.Contig1.table file are a summary of the information in the NE. Contig1.x file and list the following four scores from the homology results. The score=313 represents the HSP (High-scoring Segment Pair) score. The higher the HSP score, the better the local alignment between this pair of sequences is. The expect=2.4^{e-27} indicates the E-Value. The smaller the E-value, the better the match of the query with the searched database is. The numbers 266 and 466 indicate that the bases between 266 bp and 466 bp of the query sequence align with the subject sequence from the searched databases.

In the assembly of 3'ESTs plus 5' ESTs, several small groups of 5' ESTs

belonging to the same gene terminated at different positions and with different sequence length were observed because of 5' end stagger of a gene transcript (**Figure 3.02**). This is a result of different stop positions which occurred during the synthesis of the first strand cDNA. In some instance, the cDNAs from one gene may have a complete 5' UTR (untranslated region) and a completed coding region (CDS) while in other instance the same cDNAs may have no 5' UTR region or only have a part of the 5' UTR region sequence of this gene. This occurs when a 3' EST of one cDNA clone has no homology, but the 5' EST of this same cDNA clone has significant homology to some genes of other organisms and this usually is observed when the sequence of the 3' EST has only 3' UTR region of this cDNA clone, but the 5' EST sequence contain a part of this gene's coding. However, since all the synthesis of cDNA for these two libraries always started at the 3' end of mRNA with polyA tail, the cDNAs have unique, transcript-specific information at their 3' end. If the full length of mRNA was sequenced, the start codon for its open reading frame (ORF) can be found.

There are several advantages to doing multiple sequence alignment assembly with phrap. First, the sequence of an individual EST can be confirmed by alignment of other multiple sequences. The consensus is a more accurate and high quality sequence than a single EST sequence. Second, it yields a longer contiguous sequence than that obtained from the individual 5' end staggers. The second advantage can be used to obtain information about the differentially genes expressed in the two cDNA libraries (Schmitt et al., 1999).

In the *Neurospora* morning cDNA library, 527 contigs and 78 singlets were

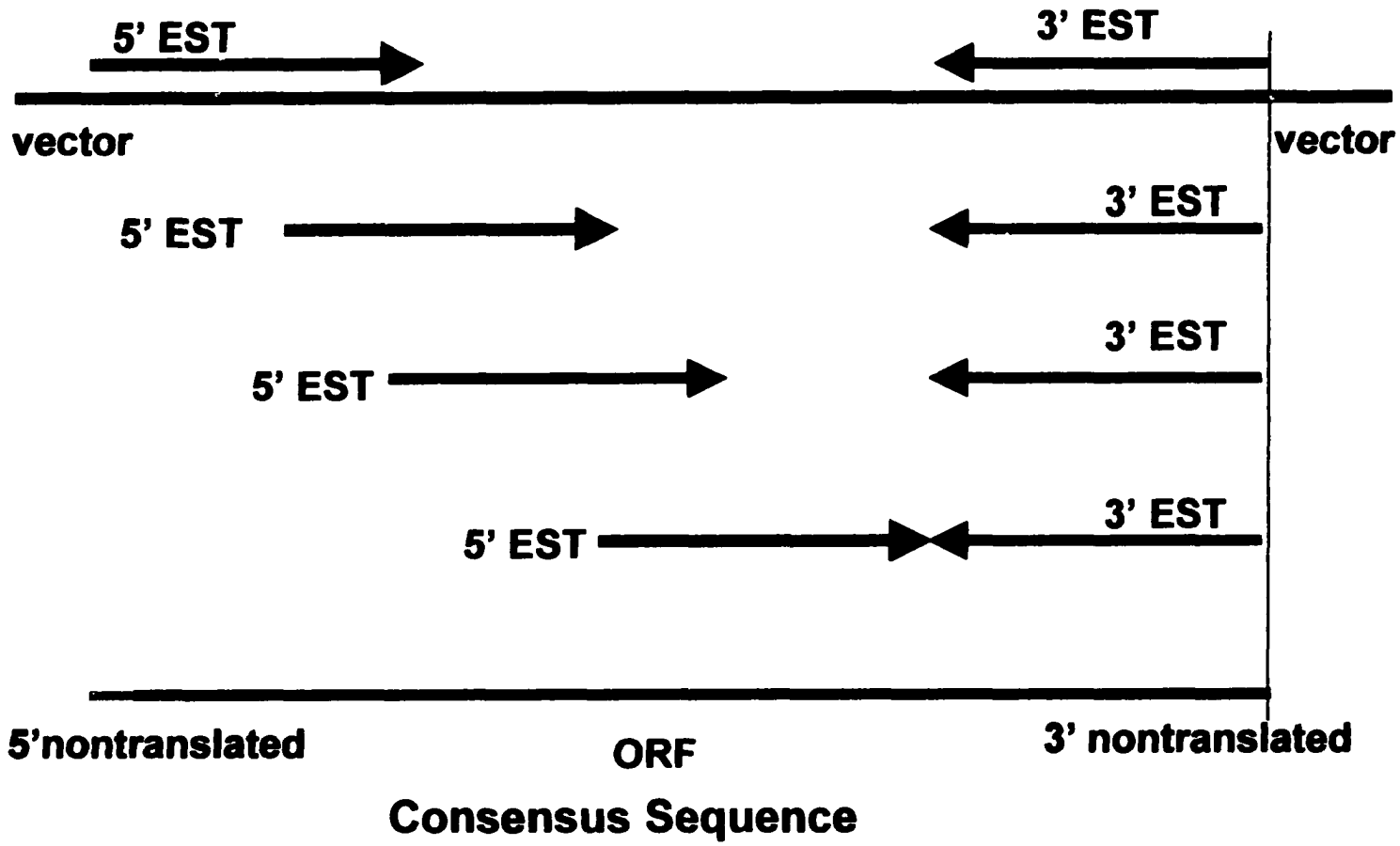


Figure 3.02 Assembly of both 3' EST and 5' EST into Contig with Phrap and the stagger of 5'ESTs

generated while 1126 contigs and 650 singlets were generated in the assembly of ESTs from the *Neurospora* evening cDNA library.

3.1.3.3 Estimation of gene numbers in two cDNA libraries

Since each cDNA represents a gene product, the distribution of cDNA species in a cDNA library reflects the number of copies genes expressed in the library. After assembling both 3' ESTs and 5' ESTs with Phrap (Phil Green, copyright 1994-1999), the number of genes represented in the EST database can be estimated by the numbers of cluster members and singlet members in the assembled-EST database.

Table 3.08 Estimation of the gene number in the assembled-EST databases of *Neurospora* NM and NE cDNA libraries

	NM	**NE**
Total ESTs in Singlets	78	650
3' EST Singlets with pair member in clusters	-10	- 89
5' EST Singlets with pair member in clusters	-16	- 67
3' and 5' EST pairs in Singlets	- 8/2	- 111/2
Number of genes represented in Singlets	**48**	**439**
Total number of ESTs in clusters	**10587** aligned ESTs	**8491** aligned ESTs
Number of clusters	527	1126
Number of clusters sharing pair number	-198	- 263
Number of genes represented in clusters	**329**	**87**
Total genes represented by assembled-EST database	**377**	**1302**

After assembly of both 3' and 5' ESTs into aligned contigs, the complexity of component of cDNA library can be reduced because the EST homologues from the same

gene are aligned together. However, because of the possibility of a large gene transcript or a multi gene family, the 3' and 5' ESTs of the same cDNA clone may not overlap each other, even though these two ESTs represent the same gene transcript. In this case, the two ESTs/contigs are matched clone pairs, and therefore, the numbers of genes represented by EST/contigs with shared clone pairs in the library were subtracted. **Table 3.08** shows the result of the estimation of gene numbers in the assembled EST databases of two *Neurospora* cDNA libraries. In the NE library, 650 singlets represent 439 genes and 1127 contigs represent 863 genes after subtraction of the numbers of the singlets and contigs that have pair sharing with others. Totally, 1302 genes were detected from the NE library after normalization. In the same way, 78 singlets represent 48 genes and 527 contigs represent 329 genes in the NM library. Totally, 377 genes were detected from the NM library. On the other hand, 8491 ESTs were aligned into 1126 clusters in the NE assembled EST database while 10587 ESTs were aligned into 527 clusters in the NM assembled EST database. Taking into account the two cDNA libraries together, 1679 total genes were observed. Although the estimation of total genes in the *Neurospora crass* genome ranges from 8000 to 10000 (Dunlap, 1996; Kupfer et al., 1997), the expressed gene number in vegetative tissue in *Neurospora cassias* estimated at about 2000 (Bell-Pedersen, 1996c). Since the sources of these two cDNA libraries are vegetative mycelium, approximately, 4.7% (377/8000) of the genes were observed in the cells when they were harvested for the morning cDNA library while 16.2% (1302/8000) genes were found in the cells at the time of harvest for the evening library. The possible reason why these two numbers are so different is that the sources used for the

construction of the cDNA libraries were different. The strain used for the construction of the morning cDNA library contains a *frq*^{7} long period mutation while the strain used for the construction of the evening cDNA library has a wild type *frq*+. Although these two strains should have no difference in the expression of genes (Bell-Pedersen et al., 1996c), the presence of the mutation might be one factor causing the difference in gene expressions seen in these two cDNA libraries.

3.1.3.4 The redundancy of genes in the two cDNA libraries

Besides revealing more information of the coding region of genes in the library, the assembly of both 3' and 5' ESTs has another advantage. It can lead to studies of the population of the cDNA library and the representation of the mRNA population. Generally speaking, the number of clones represented in one contig corresponds to the redundancy of this gene. The distribution of contigs with different clone number representation can reveal the mRNA population in the cell that was used to construct the cDNA library. **Table 3.09** and **table 3.10** show the information of the distribution of contigs with different cluster size and the frequency of different size contigs in these two cDNA libraries after 3' and 5' ESTs were assembled with Phrap 98. It can reflect the complexity of the components of a cDNA library. Complexity means the varieties of the numbers and species of mRNAs in a cDNA library. The cluster size is determined by the number of ESTs aligned to this cluster. The frequency of a cluster is the number of contigs that have similar cluster size. It indicates the number of genes that have the same ratio of mRNA copies in a cell.

Table 3.09 Assessing library complexity of Neurospora NM library using assembly of both 3' and 5' EST sequences with Phrap version 98.

Cluster Size	Frequency of Cluster	Cluster Size	Frequency of Cluster
1	81	41	1
		43	1
2	66	44	1
3	41	45	1
4	42	46	2
5	33	47	3
6	24	48	3
7	26	49	2
8	21	50	1
9	18	51	3
10	16	52	2
11	15	53	3
12	19		
13	14	57	1
14	5	58	1
15	11	60	1
16	15	64	2
17	4	65	2
18	7	68	2
19	7	77	1
20	7	78	1
21	9	83	1
22	8	86	1
23	8	91	1
24	3	92	1
25	5	99	1
26	8	100	1
27	3	106	1
		112	1
28	3	113	1
29	3	117	1
30	5		
31	1	130	1
32	5	137	1
33	3	140	1
34	1	142	1
35	5	145	1
36	1	152	1
37	5	171	1
38	2	235	1
39	2	258	1
40	1	411	1
		623	1

Table 3.10 Assessing library complexity of *Neurospora crassa* NE library using assembly of both 3' and 5' EST sequences with Phrap version 98.

Cluster Size	Frequency of Cluster	Cluster Size	Frequency of Cluster
1	650	29	3
2	335	30	2
3	169	31	1
4	116	32	4
5	89	33	1
6	71	34	2
7	48	35	2
8	35	36	2
9	31	38	2
10	35	39	1
11	17		
12	16	42	2
13	17	44	1
14	13	45	1
15	11	47	1
16	10	48	2
17	7	50	1
		52	1
18	11	67	1
19	7	69	1
20	5		
21	5	83	1
22	10	147	1
23	5	148	1
24	3	154	1
25	1	155	1
26	6		
27	5	215	1
28	5		

In the morning library, 601 unique clusters were produced from the assembly. Among them, 243 clusters contain both 5' and 3' ESTs, where 138 clusters were assembled from only 3' ESTs and 146 clusters were assembled from only 5' ESTs. In the evening cDNA library, 1760 unique members were produced from the assembly of both 5' and 3' ESTs. Among them, 553 contigs contained both 3' and 5' ESTs. 285 contigs contained only 3' ESTs while 288 contigs contained only 5' ESTs. **Table 3.11** shows the distribution of the contigs with different EST components in the morning library and evening library of *Neurospora crassa*.

Table 3.11 The contig composition in two EST databases

NE Library	Contig#	EST counts	NM Library	Contig#	EST counts
3' EST only	285	1282	3' EST only	138	1250
5'EST only	288	1307	5'EST only	146	1466
3'+5' EST	553	5902	3'+5' EST	243	7871
Total	1126	8491	Total	527	10587

Table 3.09 and 3.10 show that, several contigs occur in large clusters. In the assembled EST database of morning library, the number of ESTs in 16 contigs is larger than 100 and these 16 contigs contain 3092 ESTs in total. In the assembled EST database of the evening library, each of the largest 16 contigs have more than 40 ESTs and together they contain 1456 ESTs. These represent 29.6% and 16.9% of the total mRNA of these two cDNA libraries, respectively. These are the very abundant mRNA class according to the classification of the mRNA abundance proposed by Bishop (Bishop et al., 1974; Hastie and Bishop, 1976) who suggested that the mRNAs in each cell were

organized into three abundance classes rather than as a continuum with respect to concentration. In Bishop's studies, a kinetic analysis of hybridization of mRNA to a highly labeled cDNA copy was used to estimate of the complexity of an mRNA population. When the amount of mRNA was in vast excess, the hybridization reaction follows first-order kinetics. In the reaction, the percentage of cDNA made double-stranded is a function of the logarithm of Rot (RNA concentration in moles per liter x time in second). In an ideal condition, the $Rot_{1/2}$ (the Rot value sufficient to hybridize 50% of the cDNA) is proportional to the complexity of the mRNA and inversely proportional to the rate constant of the reaction. In an attempt to obtain the complexity of an unknown mRNA population, the hybridization of the unknown mRNA to a cDNA copy was compared with the hybridization of a standard mRNA of known complexity to a cDNA copy. According to the reassociation kinetics analysis in the studies of Hela cell and mouse tissues (Hastie and Bishop,1976; Lewin, 2000), the very abundant class consists of 15 mRNA species. Each mRNA has 12,000 copies per cell. These few mRNA species represent 10-20% of the total mRNA. Abundance is the average number of molecules of each mRNA per cell. The low abundant class consists more than 10,000 mRNA species. Each species has 1-15 copies per cell. This class of mRNA represents 40-50% of total mRNAs. In contrast, the intermediate abundant class contains 500-800 mRNA species, each has about 300 copies per cell, and it represents 40-45% of total mRNA.

The two *Neurospora* assembled EST databases also have three abundance classes **(Table3.12 and 3.13)**. The very abundant class consists of 15-16 contigs. It represents

~17% mRNA in the evening library and ~30% mRNA in morning cDNA library. In the morning library, each mRNA of this class has 100-600 copies per cell. Each contig represents >0.9% ESTs in database. While in the evening library, each of these 16 contigs contains 45-215 molecular copies per cell, which represents >0.5% ESTs in assembled EST database. The lower abundance class in these two cDNA libraries was made up of singlets and clusters of only 2 ESTs. It means these genes were only sampled one to two times in the assembled database. In the morning library, the low abundance class represents 12.3 % ESTs of this assembled EST database. About 30% genes are very abundant and more than 60% genes are intermediate abundant in the cDNA morning library. In the evening cDNA library, 16.4% of genes are very abundant and 14.5% of

Table 3.12 Gene expression levels by EST abundance in the EST database of the morning cDNA library

Total Reads	Class		# Contigs/ Singlets
3141 (29.6 %)	Very abundant	100-623 ESTs > 0.9%	16 Contigs
6219 (58.1%)	Medium abundance	10-99 ESTs 0.09%-0.9%	240 Contigs
1308 (12.3%)	Low abundance	1-9 ESTs < 0.1 %	81singlets 271 Contigs

Table 3.13 Gene expression levels by ESTs abundance in the EST database of the evening cDNA library

Total Reads	Class		# Contigs/ Singlets
1503 (16.4%)	Very abundant	45-215 ESTs > 0.5 %	14 Contigs
6315 (69.1%)	Medium abundance	3-44 ESTs 0.03% - 0.5 %	771 Contigs

1323 (14.5%)	Low abundance	1-2 ESTs < 0.03 %	650 singlets 335 contigs

genes are low abundant. ~70% of genes from this cDNA library are of medium abundance. Therefore, the variety of expressed genes in the morning cDNA library is very limited compared to that in the evening cDNA library.

3.1.3.5 The ten most highly expressed genes in two cDNA libraries of *Neurospora crassa*

The BlastX search against the nr protein database reveals that 286 contigs and 45 singlets in *Neurospora* cDNA morning library have no significant homolog in the nr protein database of GenBank while 594 contigs and 428 singlets in the NE assembled database have no homolog. This indicates that 54% of genes in NM library and 58% of genes in NE library are potential novel genes. Most of them have high redundancy in the *Neurospora* libraries.

The remaining 241 contigs and 34 singlets from the NM assembled EST database and 532 contigs and 222 singlets from the NE assembled EST database have significant matches (HSP score>100 or $E<10^{-4}$) with the genes of other organisms in the searched database. **Table 3.14** shows the top 10 highly expressed genes in these two cDNA libraries. From the table3.14, the first three highly expressed genes in both cDNA libraries were found to be clock-controlled genes *ccg-7*, *ccg-2*, and *ccg-1*, each with a different expression ratio. The detail of the similarity and difference between these two cDNA libraries will be discussed in the later sections of this dissertation.

#ctg/#ESTs	NE library	#ctg/#ESTs	NM library
1110/215	Hydrophobin precusor, *ccg-2*	523/593	glyceraldehyde 3-phosphate dehydrogenase, *ccg-7*
1109b/155	glyceraldehyde 3-phosphate dehydrogenase, *ccg-7*	522/465	hydrophobin precusor, *ccg-2*
1108/155	Glucose-repressible gene (*grg1*) product, *ccg-1*	521/346	glucose-repressible gene product, *ccg-1* (*grg1*)
1106/148	No homology	520/296	no homology
1107a/128	No homology	519/275	no homology
1104/83	*ccg-6* gene product	518/221	no homology
1102/69	phosphoglycerate kinase	517/202	N,O-Diacetymuramidase, lysozyme (new CCG gene)
1101/67	No homology	516/175	no homology
1105b/52	No homology	515/160	YBS8-yeast hypothetical 48.3kD protein
1100/50	Alcohol dehydrogenase I	514/160	IgE-binding protein

Table 3.14 The ten most highly expressed genes in two cDNA libraries of *Neurospora crassa*

3.1.4 Biological classification of expressed gene in the two cDNA libraries of *Neurospora crassa*

The final purpose of any sequencing project is to predict the identity of genes from nucleic acid sequences. In large scale genomic DNA sequencing or partial cDNA sequencing research, many genes will be predicted from the nucleic acid sequence in each organism. To identify and fully characterize the complete network of interactions among genes and their products in an organism is the ultimate challenge of a molecular biologist. However, our ability to extract biologically important information about gene interactions from genome sequences is still quite limited. Classifying the expressed genes according to their biological function, cellular roles and molecular families will significantly facilitate the data analysis and the mining of the biological function information from the DNA sequences (Jiang and Jacob, 1998).

One of most important advantage of sequencing the Expressed Sequenced Tags is that it can find new genes very quickly and inexpensively compared to genomic DNA sequencing. The mRNA population used for construction of the cDNA library can be very cell or tissue specific. Therefore, it is very useful to study the possible biological functions of genes expressed in this cell or tissue, and the possible metabolism they are involved in or regulate. Further biological function study is another important goal of an EST project. Whenever a specific gene product of interest is found, this cDNA clone can be fully sequenced. By fully determining this cDNA sequence, the structure and functions of this gene product in cell's physiological and biochemical activities may be implied.
The Riley scheme was used in the biological function classification in this study (Riley

and Labedan, 1997). The Riley scheme was designed originally for the *E. coli* genomic classification, a well-studied bacteria, and improved with KEGG (Kyoto Encyclopedia of Genes and Genomes, at http://www.genome.ad.jp/kegg) to provide a complete and exact categorization of all proteins known in *E. coli*. However, one disadvantage of this scheme is that it lacks categories for asexual development and some metabolic functions that are specifically involved in the development of fungi (Kupfer, 1999). Therefore, several additions to this category system were made to meet the requirement of the organism-specific information found in the *N.crassa* databases. The functions of a cell are quite complex that there are many ways to organize them. Since it is difficult to create an exact category classification because the boundaries between functional categories are artificial due to the complexly intertwined cellular functions and processes, no category classification system is complete and perfect, and sometime, one gene product can belong to more than one category. For example, the enzyme acetolactate synthase represented by cDNA clone c6e12ne.f1 from the evening library is involved in the metabolisms of isoleucine and/or leucine. The keyword <Acetolactate synthase> can be put either in the category of isoleucine metabolism or the leucine metabolism. In this dissertation, one cDNA only was assigned one biological function. Since the redundancy of each expressed gene in the two cDNA libraries was counted, duplicate assignments would result in false information about the expression levels. Nevertheless, a classification scheme does allow an assignment of a major cellular role to each subsequent gene and gene product. Thus, it allows a view of the activities of groups of genes and gene products of the whole organism.

The 7 Riley categories of cellular functions of the *Neurospora crass* genes identified in this research are listed in Table 3.15 and 3.16. Based on *Aspergillus nidulans* categories of cellular functions (Kupfer, 1999), several new subcategories were added and other subcategories were deleted when no genes were assigned into these function categories.

The classifications of genes identified in this research, ordered according to their cellular roles and gene families, are presented in Table3.17, using combined Riley and the KEGG (Kanehisa, 1997) category system. The detail results are accessible on our website *http://www.genome.ou.edu/oracle-test*. Each number in the left column represents the number of contigs plus singlets that represent the genes whose function in cell are assigned under this category. In this table, some genes may be counted twice if the 3' and 5' contigs or ESTs of these genes were not merged.

Figure 3.03 is the presentation of this result in a pie chart. From this figure, it can be seen that the overall levels of biological function roles of the assembled EST database members from two libraries are very similar. Of note is the observation that only 46% of assembled-EST database members in the NM library and 42% of assembled-EST database members in the NE library have a significant biological functional assignment. 2 % of members in NE library and 3 % members in NM library are unclassified, e.g., each of these members has a sequence homolog with a protein of another organism and the homologs have a defined biological function, but the query member's biological role in *Neurospora crassa* is not clear. Another 6 % from the NE library and 4 % from the NM library are categorized as unidentified. This indicates that those query members have

Table 3.15 Neurospora crassa morning cDNA library categories of cellular functions

I. Metabolism and Energy

Part one: Metabolism

- A. Metabolism of carbohydrates
- B. Metabolism of amino acids
- C. Metabolism of nucleic acid, nucleotides, purines, pyrimidines
- D. Cofactors, prosthetic groups

Part two: Energy

- A. Carbohydrate as energy source
  1. Glycolysis
  2. Gluconeogenesis
  3. Pentose-phosphate pathway
  4. Pyruvate metabolism
  5. Tricarboxylic acid (TCA) cycle
  6. Fermentation, alcoholic
  7. Energy reserves metabolism
  8. Related reaction
- B. Other energy resource
- C. Electron transport
- D. Reducing carriers

II. Gene expression and genetic information process

- A. DNA synthesis
- B. Gene expression
  1. transcription
  2. protein biosynthesis
     2.1 translation
     a. initiation
     b. elongation
     c. ribosomal protein
     2.2 post-translational modification
     2.3 folding and targeting
     a. folding
     b. chaperones
     c. protein sorting and targeting
     2.4. turnover-protein degradation

III. Cell growth ,cell process, and cell division

- A. Cell growth, cell division and cytoskeleton
  1. Cell walls
  2. Biomembranes

3. Cytoskeleton, organelle
4. Cell cycle control
5. Other

B. Cell process

1. Cell rescue, defense, osmotic adaptation, starvation response, and development
    1.1 development
    1.2. detoxification
    1.3. oxidative stress
    1.4. stress-induced protein
    1.5. clock-controlled genes

2. Cell signaling, signal transduction

3. Transmembrane transport
    3.1 transport
    a. sugar transport
    b. cation transport-ATPase or major facilitator super family
    c. anion transport
    d. protein, amino acid transport

IV. Unclassified, unidentified, no significant homolog

A. Class of enzyme (Riley and Kegg)
B. Non-enzymatic class
C. Unclassified
D. Unidentified
E. No significant homolog
Contigs
Singlets

Table 3.16 *Neurospora crassa* evening cDNA library categories of cellular functions

I. Metabolism and energy

Part one: Metabolism

- A. Metabolism of carbohydrates
- B. Metabolism of amino acids
- C. Metabolism of nucleic acid, nucleotides, purines, pyrimidines
- D. Metabolism of lipids, fatty acids, sterols
- E. Aromatic compound metabolism
- F. Sulfur metabolism
- G. Phosphate metabolism
- H. Nitrogen metabolism
- I. Cofactors, prosthetic groups

Part two: Energy

- A. Carbohydrate as energy source
  1. Glycolysis
  2. Gluconeogenesis
  3. Pentose-phosphate pathway
  4. Pyruvate metabolism
  5. Tricarboxylic acid (TCA) cycle
  6. Related reaction
  7. Glyoxylate cycle
  8. Fermentation, alcoholic
  9. Energy reserves metabolism
- B. Fatty acid resource
- C. Other energy resource
- D. Electron transport
- E. Reducing carriers

II. Gene expression and genetic information process

- A. DNA synthesis
- B. Gene expression
  1. transcription
  2. protein biosynthesis

  2.1. translation
    - a. initiation
    - b. elongation
    - c. termination
    - d. ribosomal protein

  2.2. post-translational modification

  2.3. folding and targeting
    - a. folding

b. chaperones
c. protein sorting and targeting

2.4. turnover-protein degradation

III. Cell growth , cell process, cell division

A. Cell growth, cell division and cytoskeleton
1. cell walls
2. Biomembranes
3. Cytoskeleton, organelle
4. Cell cycle control
5. Mitosis/cytokinesis
6. Other

B. Cell process
1. Cell rescue, defense, osmotic adaptation, starvation response, and development
1.1. development
1.2. defense
1.3. detoxification
1.4. desiccation tolerance
1.5. oxidative stress
1.6. clock-controlled genes
1.7. tumor protein and tumor suppressor
1.8. multidrug resistance
1.9. other
2. Cell signaling, signal transduction
3. Transmembrane transport
3.1. secretion
3.2. transport
a. sugar transport
b. cation transport-ATPase or major facilitator super family
c. anion transport
d. protein, amino acid transport
e. mitochondrial transport
f. other

IV. Unclassified, unidentified, no significant homolog
A. Class of enzyme (Riley and Kegg)
B. Non-enzymatic class
C. Unclassified
D. Unidentified
E. No significant homolog
Contigs
Singlets

Table 3.17 Genes in the different biological pathways from two cDNA libraries of *Neurospora crassa*

	NM	NE
I: Metabolism and Energy	**73**	**216**
Part one: Metabolism	**25**	**93**
A. Metabolism of Carbohydrates	5	7
B. Metabolism of amino acids	11	35
C. Metabolism of nucleic acid, nucleotides, purines, pyrimidines	5	5
D. Metabolism of lipids, fatty acids, sterols	0	19
E. Aromatic compound metabolism	0	2
F. Sulfur metabolism	0	2
G. Phosphate metabolism	0	2
H. Nitrogen metabolism	0	5
1. Cofactors, prosthetic groups	4	16
Part II: Energy	**48**	**123**
A: Carbohydrate as energy source	**37**	**55**
1. Glycolysis	17	13
2. Gluconeogenesis	1	5
3. Pentose-phosphate	0	4
4. Pyruvate metabolism	2	3
5. Tricarboxylic acid (TCA) cycle	5	11
6. Related reaction	1	1
7. Glyoxylate cycle	0	5
8. Fermentation , alcoholic	4	8
9. Energy reserves metabolism	7	5
B. Fatty acid as resource	**0**	**3**
C. Other energy resource	**1**	**13**
D. Electron transport	**8**	**51**
E. Reducing carriers	**2**	**1**
II: Gene expression and genetic information process	**91**	**248**
A. DNA synthesis	**7**	**16**
B. Gene expression	**84**	**232**
1. transcription	14	34
2. Protein biosynthesis	70	198
2.1. translation	38	120
a. initiation	2	6
b. elongation	3	9
c. termination	0	1
d. ribosomal protein	32	102
2.2 post-translational modification	2	10
2.3. folding and targeting	19	41
a. folding	4	10

b. chaperones	12	14
c. protein sorting and targeting	3	17
2.4. Turnover-protein degradation	11	27
III: Cell growth, cell division and cell process	**77**	**153**
A. Cell Growth, Cell Division and Cytoskeleton	**28**	**45**
1. cell walls	10	13
2. Biomembranes	7	8
3. Cytoskeleton, organelle	7	15
4. Cell cycle control	2	4
5. Mitosis/ cytokinesis	1	3
6. Other	1	2
B. Cell process	**49**	**108**
1. Cell rescue, defense, osmotic adaptation, starvation response, development	**28**	**49**
1.1. development	6	10
1.2. defense	0	5
1.3. detoxification	6	8
1.4. desiccations tolerance	0	1
1.5. oxidative stress	3	4
1.6. stress-inducible protein	3	0
1.7. clock-controlled genes	10	13
1.8. other	0	3
1.9. tumor protein and tumor suppressor	0	2
1.10. multidrug resistance	0	3
2. Cell signaling, signal transduction	**9**	**22**
3. Transmembrane transport	**12**	**37**
3.1. secretion	0	2
3.2 transport	12	35
a. sugar transport	6	6
b. nuclear membrane transport	0	10
c. cation transport-Atpase or major facilitator superfamily	1	7
d. anion transport	2	2
e. protein, amino acid transport	3	10
f. mitochondrial transport	0	8
IV : Unclassified, unidentified, no significant homolog	**365**	**1161**
A. Class of enzyme (Riley and Kegg)	4	3
B. Non-enzymatic class		
Leucine zipper motif	1	0
Zinc finger motif	0	1
C. Unclassified	5	30
D. Unidentified	25	105
E. No significant homolog	330	1022
Contigs	285	594
Singlets	45	428

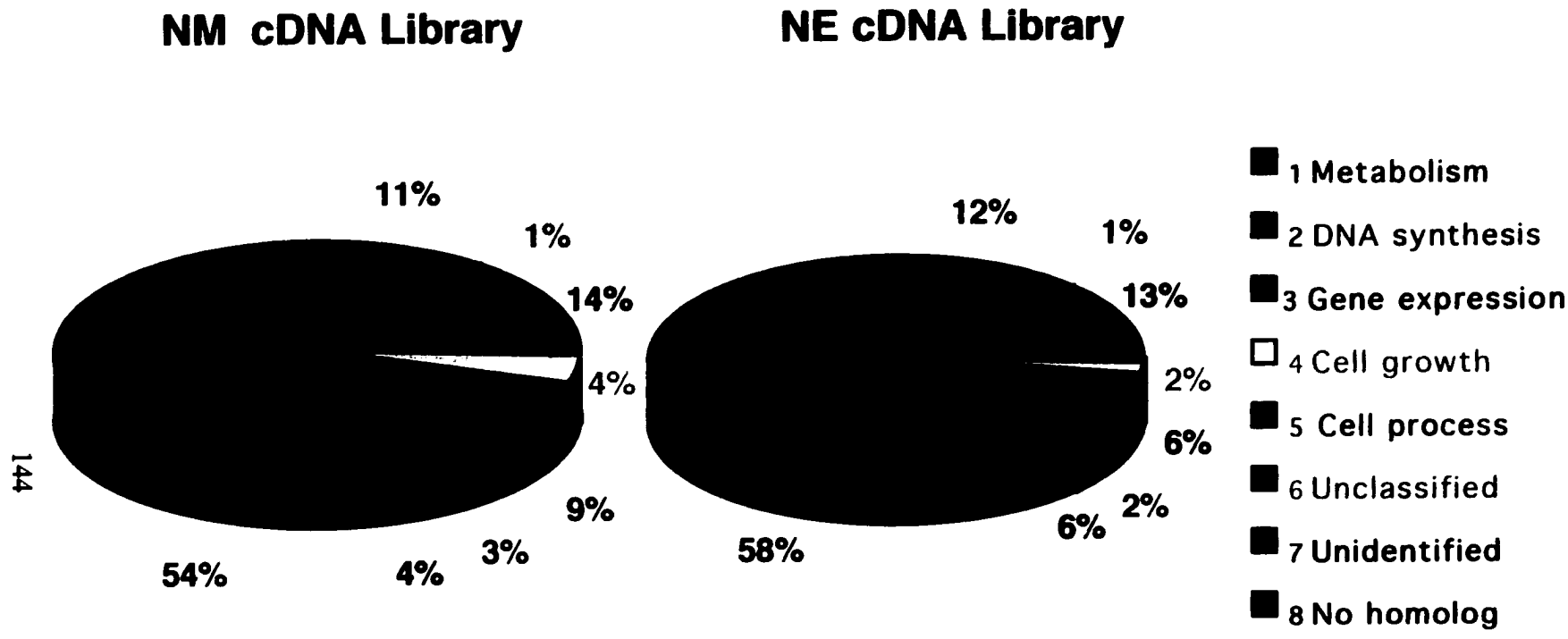


Figure 3.03 Genes in different categories of celluiar functions from two cDNA iibraries of *Neurospora crassa*

homologs or strong matches with ORFs or predicted genes/proteins from other organisms, but no clear biological functions were assigned to them. More than 50% of the members in both libraries have no significant homologs to any gene or protein in the Genbank database. They potentially represent novel genes and this observation will be discussed in detail below.

The percentage of novel genes found in different EST projects vary with the species, the tissues studied and the nature of library used (Retief, et al., 1999). In human tissue adult prostate, the percentage of novel genes is 6% (Nelson, et al., 1998) while it is 51.7% in fetal heart (Hwang et al., 1995). Plants show higher percentages of novel genes in the libraries examined with 75% in rice (Yamamota and Sasaki, 1997) and 91% in rape (Park et al., 1993).

However, the percentages of novel genes found in the EST studies of several different fungi are very similar and are in good agreement with each other. In the EST studies from three tissue-specific libraries of *Neurospora crassa*, the percentage of novel gene was 57% (Nelson et al., 1997). In the EST studies of *Aspergillus nidulans*, the percentage of novel genes was 56% (Kupfer, 1999), and in the EST studies of *Fusarium sporotrichioides,* the percentage of novel genes was 54% (Ren, unpublished data). The database search stringency criteria used in *A. nidulans*, *F. sporotrichioides* and this research are the same (HSP>99 and $P<10^{-4}$). From the data discussed previously and here, it is clear that the percentages of new genes observed varies with the species or tissues studied.

Since the early 1990's, a high number of ESTs has been generated and the

number of genes in database has grown exponentially. Although the BLAST programs (Altschul et al., 1990) are sensitive, there still are a high percentage of ESTs which have no match in the database. The major reason for this is likely that these ESTs may represent fungal specific genes that are not present in the public databases because there are a limited member of fungal species whose ESTs and/or genome have been studied. Another possible reason for such a large percentage of ESTs with no known homologies is that there is a high frequency of non-full-length cDNAs in the studied cDNA libraries and therefore, portions of a gene in question are not present in Genbank. The other possible reason for the existence of the large amount of non- homologous EST data is that these ESTs represent new, previously unknown genes, gene families or uncharacterized branches of known families. Alternative splicing in different species or tissues of same genes and the inaccurate sequences also may be possible factors that result in the lack of homolog of genes in question in Genbank databases (Cirera and Winter and Fredholm, 2000). Finally, since ESTs are generated by only single-pass sequencing of one strand of the cDNA, they are prone to errors in their sequences.

Although the percent of identified genes assigned into the 8 major categories for two cDNA libraries in *Neurospora crass* are very similar, the absolute numbers of genes belonging to these categories are quite different. For example, the percentages of genes located in DNA synthesis are 1% in both cDNA libraries in *Neurospora*. However, only 7 genes were detected in this category in NM library while 16 genes were detected in NE library. Another 84 genes were involved in gene expression in NM library while 228 genes are in this category in the NE library. Ribosomal protein genes are highly

represented in the category of protein biosynthesis in both libraries. However, only 32 ribosomal proteins (5.3%) were detected in NM library, and 99 genes (5.6%) encoding ribosomal proteins were detected in NE library. Genes assigned into the category of bioenergetics and metabolism is another class that is highly expressed in both libraries. Here, 66 genes were detected in NM library and 212 genes were detected in NE libraries, representing 10.1% and 12.03% of the total identified genes in these two libraries, respectively.

Comparison of the *A. nidulans* and *N.crassa* EST data reveals difference in two classes of genes, the genes encoding chaperons and the genes under control of the biological clock. Compared to the assembled EST database of *A. nidulans*, the genes coding for chaperones are not as highly expressed in *N.crassa*, as the *A. nidulans*, chaperones represent 6.1% of its total cDNA clone population and the *N.crassa* chaperone genes represent 1.9% and 0.8% of the total cDNA population in the NM and NE cDNA libraries, respectively. In contrast, 15 and 12 the known clock-controlled genes (*ccgs*) were detected in the NM and NE libraries of *Neurospora,* respectively. Interestingly, in the NM library, the clock-controlled genes represent more than 9.9 % of total cDNA clone population. However, the *ccgs* were not significantly over expressed but two *ccgs* (the conidiation specific gene *con-10* and the *ccg-1*gene) (Ebbole, 1998) were observed in the moderately expressed EST in this *A. nidulans* library. The details about the clock-controlled genes expressed in two cDNA libraries of *Neurospora crass* will be discussed in detail later.

3.1.5 Comparisons of two EST databases from *Neurospora crassa*

For comparison between sequences, either EST or genomic is one of the most informative activities in computational molecular biology. As long as a pair of sequences has been determined to be homolgous, a precise map of the conserved feature can be created by a sequence comparison. Functional information, such as the location of the secondary structural elements, domains, active sites, and regulatory regions from a well studied molecule can be compared to a new sequence to map these features.

Comparisons of genes from different genomes reveal that the conserved genes during evolution typically are the housekeeping and regulatory genes (Eizinger et al., 1999). Conservation of these genes may reflect transfer of genetic information between species needed for the survival of the species during evolution. There are two types of homologous gene families: paralogs and orthologs. Paralogs are genes that descend from the same ancestral gene but have new functions by duplication and divergence during the course of evolution that occur in same species. In contrast, the homolog genes from different species are called orthologs. They also have evolved from a common ancestral gene, but orthologs normally have the same or similar function in different species. The comparison between orthologous genes from different organism can reveal conserved functions since these genes have a conserved sequence in both species.

EST database comparisons also can provide a rapid and efficient way to characterize gene function and tissue physiology. One of main purposes of this present research was to identify the genes that are expressed under control of the *Neurospora* clock and the genes that might interact with clock gene by comparing the expression

profiles of two cDNA libraries. Therefore, the comparisons of gene expression between these two cDNA libraries can greatly aid these studies.

In recent years there has been an exponential growth in genomic DNA sequencing. This large amount of data challenges both computational and experimental biologists to create new methods to decipher the secrets of the encoded proteins. In addition, a variety of large-scale methods to assay both gene expression and protein-protein interaction have made new functional-related data available (Ohlrogge and Benning, 2000).

Thus, massive amounts of detailed information now is becoming available describing the molecular basis of metabolism, regulation, intracellular and intercellular interactions, and this information us driving the rapid growth of the electronic data resources. Genbank (Bensen et al., 1998), EMBL (Stoesser et al., 1998), SWISS-PROT (Bairoch & Apweilr, 1997), PIR (George et al., 1996), and PDB (Abola et al., 1987) are the examples of the first-generation of molecular biological data resource. The European Bioinformatics Institute (EBI, *http://www.ebi.ac.uk/*), the National Center for Biotechnology Information (NCBI, *http://ww.ncbi.nlm.nih.gov/*) and the DNA Data Bank of Japan (DDBJ, *http://www.ddbj.nig.ac.jp/*) collaborate to produce the nucleic acid databases EMBL, Genbank and DDBJ. These first-generation databases are broad but shallow since they contain sequence information, but provide only a limited amount of additional information or references to related metabolic information. In contrast, KEGG (Kanehisa, 1997) is a second generation database, which contains additional detailed annotation added to the information present in the usual first generation databases.

3.1.5.1 The combined EST database

Because the two cDNA libraries of *Neurospora crass* used in this research are from the same organism but at different growth stages, most of the observed gene expression should be same although the genes under control of the *Neurospora* clock may be expressed differently. By combining all of the ESTs from both cDNA libraries the accuracy of similar sequence can be improved, and the length of identical cDNAs can be extended, resulting in a more complete cDNA sequence. Therefore, before doing individual comparisons, all of the ESTs from both NM and NE EST databases were merged. These ESTs were assembled with Phrap version 98 with the same stringency criteria that were used in the previous assemblies of ESTs in this research. After assembly, the combined EST database contained 1456 contigs and 664. The all-frame translated sequences of each contig then were used for sequence similarity searches against the nr protein database of Genbank using the Blast tools.

After correction and normalization, in total 1431 genes were present in the combined EST database (Table 3.18) and several ESTs aligned into clusters after these ESTs were merged. In the combined database, all 19093 ESTs were aligned into clusters. It is larger than the sum of ESTs that were aligned in the NM and NE. 10587 of ESTs and 8491 of ESTs

Table 3.18 Estimation of the gene number in the combined EST database of *Neurospora crassa*

	Combined EST database (NMNE)
Total ESTs in Singlets	664
3' EST Singlets with pair member in clusters	-73
5' EST Singlets with pair member in clusters	-92

3' and 5' EST pairs in Singlets	-108/2
Number of genes represented in Singlets	445
Total Number of ESTs In Clusters	19093
Number of Clusters	1456
Number of clusters sharing pair number	470
Number of genes represented in Cluster	986
Total genes represented by combined EST database	**1431**

were aligned into clusters in NM and NE assembled databases, respectively. This increase is due to the alignment of singlets from each library matching to form a contig.

3.1.5.2 Gene distribution in the combined EST database

In order to study the distribution of genes in these two cDNA libraries, a script was written by Hongshing Lai in our informatics group to sort the clusters according the EST components from the NM or NE cDNA library. The results of this sorting is summarized in the Table 3.19 and visualized in the Figure 3.04.

Table 3.19 The distribution of ESTs from NM and NE libraries in the combined EST database.

NM%	0	10	20	30	40	50	60	70	80	90	99	100
#Contigs	894	25	25	30	24	18	24	28	38	49	29	281

From this table it can be seen that there were 894 contigs in the combined database which were only from NE data and 281 contigs solely from NM data. Therefore, 894 contigs represented 894 genes that only are expressed in the NE cDNA library while there are 281 only expressed in the NM cDNA library. The data in between these two

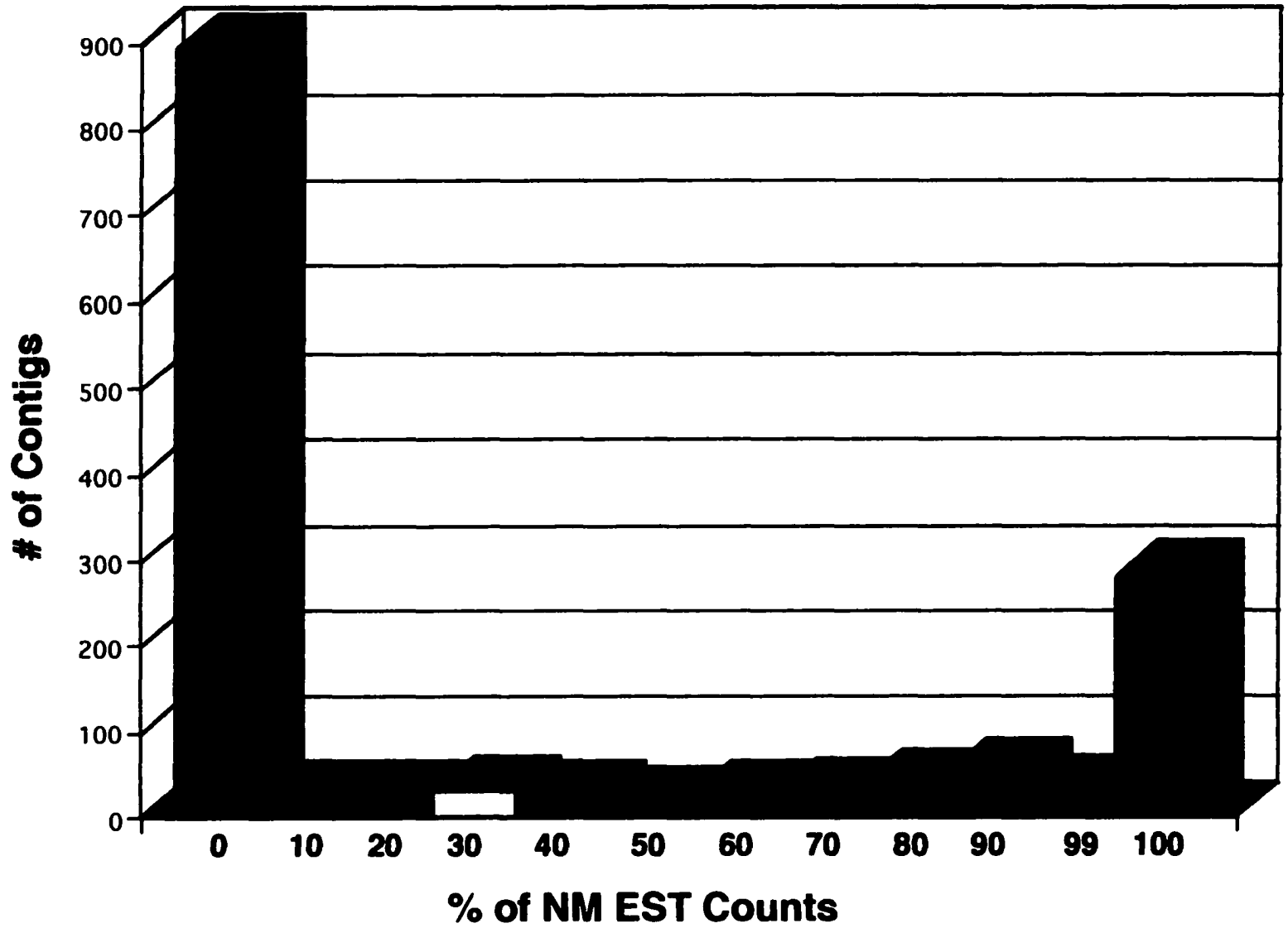


Figure 3.04 The distribution of genes in two cDNA libraries of *Neurospora crassa*

extremes consists of the ESTs from both the NE library and the NM library. The percent indicates the ratio of the ESTs component from the NM library in this group of contigs. For example, in the second column, the NM EST percent is 10 and the number of contigs is 25. This means that in each contig of this group, the ESTs from the NM library is 10% and the rest of 90% ESTs are from the NE library. Totally, 25 different genes represented by this group of contigs have the same expression ratio in the combined database. All the members of groups of contigs between 10% NM to 99% NM represent the genes that were expressed both in the morning library and in the evening library. In total, there were 290 genes expressed in both cDNA libraries of *Neurospora crassa* but with different expression ratios. Table 3.20 shows the numbers of contigs that represent genes solely expressed in the two different cDNA libraries. A detailed analysis of these three groups of genes (genes only expressed in NM, NE and genes expressed in both of NM and NE libraries) is discussed in the following sections.

Table 3.20 The numbers of genes expressed only in one of the two different libraries

Genes only expressed in the NE library			Genes only expressed in the NM library		
	Ctg numbers	EST Counts		Ctg numbers	EST Counts
3' only	253	1058	3' only	92	753
5' only	243	996	5' only	97	820
3'+5'	398	2781	3'+5'	92	1620
total	894	4835	total	281	2943

3.1.5.3 Genes only expressed in the morning cDNA library

Because one of the purposes of this research was to detect and compare the

difference of gene expression and transcription level in two cDNA libraries, the genes that only were expressed in each library were determined. Here, by combining the ESTs from the two cDNA libraries, assembling by phrap with same stringency criteria, examining and locating the compositions of each assembled EST cluster, provides one way to detect the genes that only were expressed in NM or NE library. From table 3.19, it is clear that 281 genes only were expressed in NM library while 894 genes only were detected in the NE library.

Among the 281 genes expressed only in the NM library, 156 of them had significant homologues and a clear biological function assignment. In contrast, 125 had no sequence similarity to any gene product in the nr protein database of Genbank and therefore may be potential new, previously unexpected novel genes.

The genes represented by the 92 of 281 contigs that have only 3' ESTs are listed in Appendix V. These 92 contigs contain 753 ESTs, which were from the reverse sequencing reactions of 753 cDNA clones and represent 3.94% of total ESTs (19093) of the combined EST database. 55 of these 92 contigs have no sequence homolog in the GenBank nr protein database while 41 of the 92 contigs had significant homologies in the GenBank nr protein database.

The genes represented by the 97 of 281 contigs which have only 5' ESTs are listed in Appendix VI. These 97 contigs contain 820 of 5' ESTs from 820 cDNA clones. They represent 4.29% of the total ESTs of the combined EST database. 50 of these 97 contigs have no sequence homolog in the searched database while 47 of them had significant hits in the GenBank nr protein database.

Another 92 contigs consisted of both 3' ESTs and 5' ESTs. The genes represented by this group of contigs are listed in Appendix VII. These 92 contigs contain 1620 ESTs, represent 8.48% of the total ESTs of the combined EST database, and 27 of these 92 contigs had no homolog in GenBank nr protein database, while 65 of them had significant homologies in GenBank.

Totally, there were 2943 ESTs in these 281 genes that only were expressed in the *Neurospora* cDNA morning library. They represent 16.71% of the total ESTs of this combined-EST database.

The biological functions were assigned to these 281 contigs as described in the previous section of this dissertation. The result of this classification of genes represented by these contigs is listed in table3.21 as follow.

Table 3.21 The biological function classification of the genes that were only detected in the NM library

Pathway	# of Contig	%NM	%NMNE
Metabolism	12	2.3	0.8
Bioenergetics	18	3.4	1.2
DNA synthesis	0	0	0
Transcription	9	1.7	0.6
Protein synthesis	19	3.6	1.3
Cell growth, cell division	10	1.9	0.7
Cell process	19	3.6	1.3
Unclassified	17	3.2	1.2
Unidentified	15	2.8	1.0
No homology	163	30.9	11.1

%NM was based on the total NM contig numbers 527; %NMNE was based on the total NMNE contig number 1456.

The largest segment of this group of genes only were expressed in the NM library.

Because they have no homolog in the GenBank nr protein database, they represent potentially novel genes. Both transcription and cell process are in the second largest group of genes. Under the category of protein synthesis, 7 of 19 contigs represented different heat shock proteins or chaperones.

The clock-controlled gene 8 observed in contig1356 only was detected in the NM library. The gene encoding arylalkylamine N-acetyltransferase (AA-NAT enzyme) only was detected in the NM library. This enzyme is responsible for the synthesis of melatonin, a hormone that secreted primarily from the pineal gland and plays an important role in the synchronizing biological rhythms (Pevet, 2000; Vanecek and Watanabe, 1999). The circadian rhythm in SCN (suprachiasmatic nucleus) of mammalian or human is related to the change of the level of melatonin (Hunt, Al-Ghoul, Mu and Dubocovich, 2001). It was present in an EST singlet h6a11nm.f1. Another three clock-controlled genes, the *ccg-1* present in contig48, the *ccg-7* present in contig53 and the *ccg-9* present in contig739 and contig1332 were included in this group of genes. However, these three *ccgs* also were detected in the NE library but did not aligned with the ESTs from these three genes in the NE library. This observation will be discussed in the next section.

Table 3.22 shows the top 10 highly expressed genes that only were expressed in the NM library. A DNA-like protein homologue, 2 RNA helicases, a keraton 2 epidermis, a cyclophilin, a dehydrogenase and four additional genes with unknown function were ten genes that have highest frequency in this group of genes that only were expressed in the NM library.

Table 3.22 Top 10 highly expressed genes only detected in the NM library

# contig	Gene	#ESTs	%NMNE
1434	Psiprotein/dnaJ-like protein homolog	134	0.70
1402	similar to dehydrogenase	60	0.31
1392	cyclophilin	53	0.27
1390	keratin 2 epidermis	53	0.27
1388	no homolog	52	0.27
1386	no homolog	51	0.27
1383	no homolog	49	0.26
1380	no homolog	48	0.25
1373	putative RNA helicase	44	0.23
1370	putative pre-mRNA splicing factor	41	0.21

- The number of total ESTs of NMNE is 19093

3.1.5.4 Genes only expressed in the evening cDNA library

894 of contigs consist of ESTs that were generated from the cDNA clones only from the NE cDNA library. They represent the genes that only were expressed in the NE cDNA library.

From table 3.20 in section 3.1.5.2, among 894 contigs, 253 contigs only had 3' ESTs that were generated from the reverse sequencing reactions from cDNA clones of the NE library. 144 of them had no sequence homolog in the GenBank nr protein database while 109 of them had significant homologies. 1058 of ESTs were aligned into these 253 clusters, which represents 5.54 % of total ESTs of the combined database and 8.07% of the total cDNA clone population (13019) of these two cDNA libraries. The genes represented by these contigs are listed in Appendix VIII.

243 of 894 contigs contain only 5' ESTs that were generated from the forward sequencing reaction of cDNA clones. 89 of them have no sequence homolog while 154 of them have significant hits in the GenBank nr protein database. 996 of these ESTs were aligned into these 243 contigs and represent 5.22 % of total ESTs of the combined EST database and 7.65% of cDNA clone population in the combined NMNE library. The genes assigned to these contigs are listed in Appendix IX.

The remaining 398 of 894 contigs consist of both 3' ESTs and 5' ESTs that were generated from both forward and reverse sequencing reaction of cDNA clones. 195 of them had no sequence homolog in Genbank while 203 of them had significant homolog in the GenBank nr protein database. 2781 of ESTs were aligned into these 398 contigs, which represent 14.57% of total ESTs of the combined EST database. The genes represented by these 398 contigs are listed in Appendix X.

Totally, 4835 of ESTs were assembled into genes that only were expressed in the NE cDNA library and they represent 25.33% of the total ESTs of the combined NMNE EST database.

The biological functions were assigned to these 894 contigs as described in the previous section of this dissertation. The results of this classification of genes represented by these contigs is as follows (table.3.23). The percent of these contigs in the NE is based on the 1126 contigs in the NE library. The %NMNE based on 1456 total contigs in the combined database.

Table 3.23 The biological function classification of the genes that were only detected in the NE library

Pathway	# of Contig	%NE	%NMNE

Metabolism	57	5.1	3.9
Bioenergetics	81	7.2	5.5
DNA synthesis	9	0.9	0.7
Transcription	14	1.3	0.9
Protein synthesis	105	9.3	7.2
Cell growth, cell division	26	2.3	1.8
Cell process	49	4.4	3.3
Unclassified	77	6.8	5.3
Unidentified	46	4.1	3.1
No homology	430	38.2	29.4

%NE was based on the total NE contig numbers 1126; %NMNE was based on the total NMNE contig number 1456

As observed with the NM library, the largest group of genes had no homolog in the GenBank nr protein database, and represent 29.4% of the total contigs of the NMNE. The second largest observed gene group was in the category of protein synthesis, and 32 of these 105 contigs represented the different ribosomal proteins. 9 contigs of this group in the NE library represented genes required for DNA synthesis. However, in correspond part of the table 3.21, the number of the contigs represented genes required for DNA synthesis in the NM library was zero.

The clock-controlled gene 4, represented by contigs 605, 1002, and 1262, was only detected in the NE library. Several other important genes related to the circadian rhythm in *Neurospora crass* were also only were detected in the NE library. They are the clock-controlled gene 6 (*ccg-6*), represented by contigs 1122 and 1165; the conidiation specific gene 6 (*con-6*) and gene 10 (*con-10*), represented by contigs 1098 and 1116, respectively; the *bli-3* gene represented by contig724 and the gene encoding spermidine

synthtase represented by contig300. Interestingly, spermidine was reported to determine the sensitivity of the conidiation rhythm to the calmodulin antagonist, chlorpromazine (Suzuki, Katagiri, and Nakashima, 1996; Katagiri, Onai and Nakashima, 1998). In addition, the genes encoding the enzymes involved in the biosynthesis of sterigmatocystin, the sterigmatocystin 7-O-methyltransferase (ST-OMTase) (P55790) (Liu, Bhatnagar and Chu, 1999) and the sterigmatocystin biosynthesis protein (Q00717), only were detected in the NE library as two EST singlets, b8e06ne.r1 and the g5h05ne.f1, respectively (Appendix II and III). The sterigmatocystin is the precursor of a mycotoxin, aflatoxin, that has been identified in several species of a pathogenic *Aspergillus* fungi (Klich, Mullaney, Daly and Cary, 2000; Begum and Samajpati, 2000). The gene encoding the RCO-1 regulator for the transcription of *grg1* gene also was detected only in the NE library and was represented by an EST singlet b5c06ne.r1.

Table 3.24 shows the top 10 highly expressed genes that only were detected in the NE library and includes a subunit E of mitochondrial F_1F_0-ATPase, two different ribosomal RNA proteins, calmodulin, an alpha chain of mitochondrial precursor and 5 genes with unknown function.

Table 3.24 Top ten highly expressed genes only detected in the NE library

Contig#	Gene	#ESTs	%NMNE
1374	subunit E of mitochondrial F1F0-ATPase; Tim11p	45	0.24
1367	alpha chain, mitochondrial precursor	40	0.21
1353	no homolog	36	0.19
1342	no homolog	34	0.18
1339	no homolog	32	0.17
1338	ribosomal protein S25B(S31B)(rp45); Rps25bp	32	0.17

1337	no homolog	32	0.17
1319	no homolog	29	0.15
1317	calmodulin	29	0.15
1316	ribosomal protein CRP7	28	0.15

3.1.5.5. Genes expressed in both of the cDNA libraries

The 290 contigs that consist of the ESTs from both NM and NE libraries represent genes that were detected in both of cDNA libraries. Totally, these contigs contained 11012 ESTs from both the NM and the NE cDNA libraries. Therefore, 290 similar genes were detected in both of cDNA libraries. The ratio of the EST totals from the NM and the NE libraries in one contig indicates the expression ratio of a gene in two different libraries. Among these 290 genes listed in Appendix XI, 161 of them have at least one significant sequence homologue in the GenBank nr protein database. While 129 genes in this group have no significant similarity to any sequences in the GenBank nr protein database, and might be potential new genes specifically expressed in *Neurospora*.

From the above table, it can be seen that the 50 highly expressed genes are represented by the largest 50 contigs of the combined EST database, with the exception of the contig 1434. Contig 1434 represents the highest expressed gene that only was detected in the NM library. Since the largest 49 contigs in combined EST database represented the top 49 highly expressed genes in this group of genes, the very abundant mRNA species in combined EST database were those that were expressed at both growing stages.

Table 3.25 shows the top 10 highly expressed genes in the group of genes that were expressed in both of cDNA libraries in *Neurospora crassa*. From this table, it can be seen that four members of these 10 genes are clock-controlled genes. They are present in contig1442, contig1446, contig1447, and contig1448, that together contain 1890 ESTs aligned into these 4 clock-controlled genes, representing 9.9% of the total ESTs in the combined EST database. The redundancy of several clock-controlled genes in these two time-of-day-specific cDNA libraries is high. This is the major unique feature of these two cDNA libraries as they were constructed to study the clock-controlled genes. Similarly the EST databases from two other fungal species studied in our laboratory (Kupfer, 1999; Run., et al, 2001), also had phenotypic gene expression. The heat-shock genes were predominant in the stress-induced cDNA library of *Aspergillus* studied by D. Kupfer, the trichloride genes were the most highly expressed genes in the trichloride-enriched cDNA library of *Fusarium sporotrichioides* studied by Q. Ren.

In addition, 5 of the top 10 highly expressed genes in this group have no homologue in the GenBank nr protein database, are present in more than 165 ESTs. It is also a very high frequency, may represent new genes only expressed in *Neurospora*. Interestingly, these genes were expressed at both growth stages of these two *Neurospora* strains with a high frequency. Since had no sequence similarity to any other known sequence in the GenBank nr protein database, further studies are required to uncover their identity and function.

Table 3.25 Top 10 genes expressed in both of cDNA libraries of *Neurospora crassa*

#Contig	Gene	P Value	# ESTs	Ratio of NM/NE	NM%
1448	Glyceraldehyde 3-phosphate dehydrogenase, ccg-7	8.9e-178	669	537/132	4.94
1447	Hydrophobin precursor(rodlet protein) blue light induced proteinn 7, ccg-2	6.7e-42	664	457/207	4.20
1446	Giucose-repressible gene protein *ccg-1*	4.2e-32	371	253/118	2.33
1444	No homolog		252	234/18	2.15
1443	No homolog		194	169/25	1.55
1442	N,O-Diacetylmuramidase (Lysozyme CH), new ccg gene protein	2.1e-79	186	152/34	1.40
1441	No homolog		175	27/148	0.25
1440	IgE-binding protein	4.0e-17	174	140/34	1.29
1439	No homolog		168	147/21	1.35
1438	No homolog		165	143/22	1.32

3.1.6 *Neurospora* genes that have no sequence homologue in the nr protein database of Genbank

As described in the previous section, 54% of genes in the NM library and 58% of genes in the NE library had no homology in the GenBank nr protein database and likely represent *Neurospora crassa* specific genes that are present 285 contigs and 45 in the NM library and 594 of contigs and 428 of singlets in the NE library. The genes represented by these contigs and singlets from both cDNA libraries of *Neurospora* are listed in Appendix XII and XIII.

A tBlastX search against the dbEST database of GenBank, which did not contain the *Neurospora crassa* ESTs determined during this present research, was performed using above contigs and singlets that had no homolog in the GenBank nr protein database. tBlastx is a more sensitive sequence similarity search program because it uses the six-frame translations of a nucleotide query sequence to compare with the six-frame translation of a nucleotide sequence database even though the stringency parameters used were the same stringency as used in BlastX. The tBlastx search results with an HSP>99 or an E-value of $<e^{-4}$ revealed that among the above 285 contigs and 45 singlets in the NM EST database, 202 contigs and 78 singlets had no homolog in the dbEST database. Therefore only 8% of contigs and singlets that had no homolog in the GenBank nr protein database had homolog in the dbEST database. Among the 1022 contigs and singlets of the NE EST database that had no homolog in the nr protein database of GenBank, 482 contigs and 373 singlets had no homolog in the dbEST database (**Figure3.05**).

Among the top 10 highly expressed genes that were expressed both in the NM cDNA library and in the NE cDNA libraries (Table 3.25), 5 had no sequence homolog in

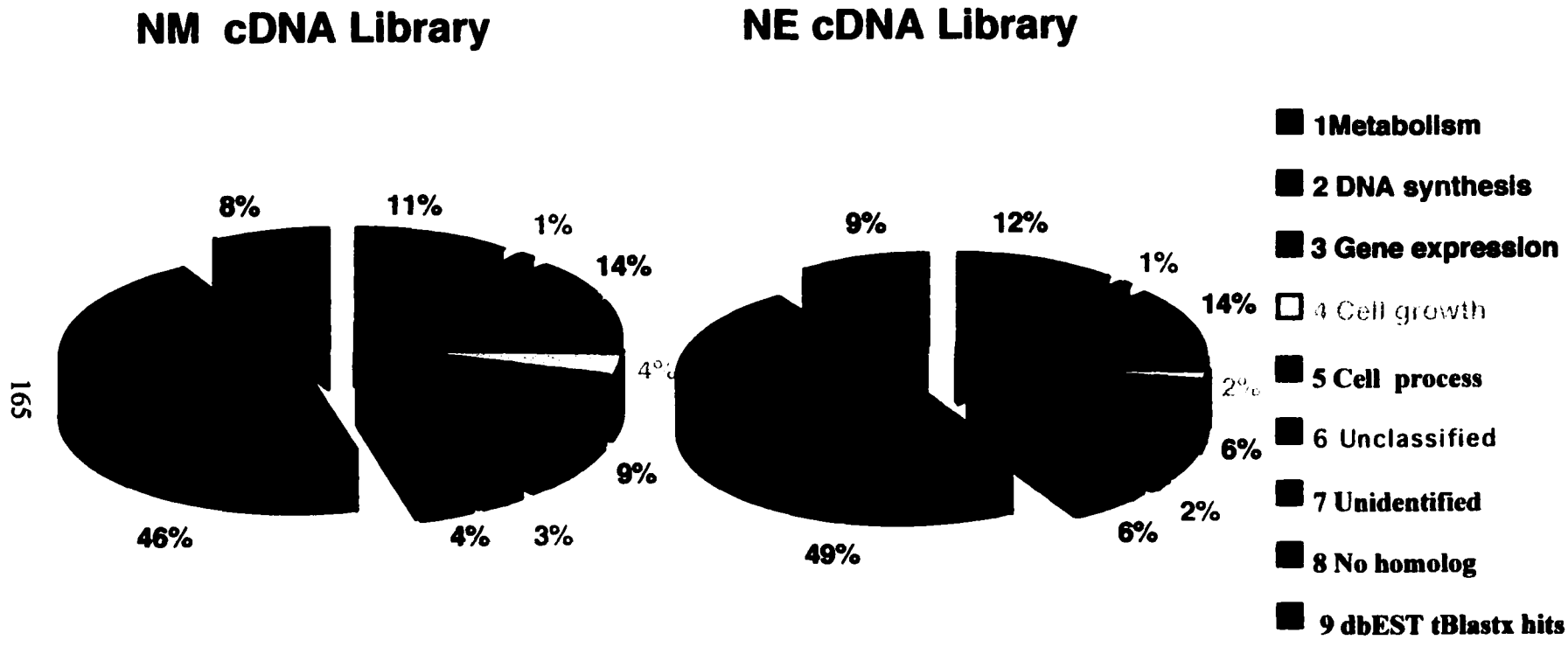


Figure 3.05 Genes in different categories of cellular functions from two cDNA libraries of *Neurospora crassa*

the GenBank nr protein database (Table 3.26). Of those with no homolog, four of the five had much higher expression ratios in the NM cDNA library than in the NE cDNA library. It is quite possible that these highly expressed cDNAs all could be potential novel clock-controlled genes in *Neurospora*, because the clock could be the factor that induces the transcription of these novel genes.

Table 3.26 Five potential novel genes expressed in both of cDNA libraries

Contig#	EST total	NM/NE
1444	252	234/22=10.8
1443	194	169/30=5.6
1441	175	27/178=0.15
1439	168	147/25=5.86
1438	165	143/26=5.42

3.2 The clock-controlled genes in *Neurospora crassa*

Studying cock-controlled genes was an important goal of this research. Even though the background about these clock-controlled genes was addressed in the previous part of this writing, more detail about the clock-controlled genes identified and the related studies of these *ccgs* will be addressed here, with emphasis on the three most redundant *ccgs* in these two cDNA libraries, *ccg-1*, *ccg-2*, *ccg-7*.

3.2.1 The clock-controlled gene 1

ccg-1 gene was one of the most highly expressed clock-controlled genes observed in the two cDNA libraries studied although it had higher expression in the morning cDNA library. The contigs and their original EST counts representing this gene in the

combined EST database were listed in table 3.27.

ccg-1, a morning specific clock controlled gene, was one of the initially identified clock-controlled genes in *Neurospora crassa* (Loros et al., 1989). The *ccg-1* gene's message RNA is 1888 nucleotides length. The molecular mass of the protein encoded by its mRNA is 7200 Da (Wang et al., 1994) and it contains 71 amino acid residues. The *ccg-1* was identified independently by McNally and Free (McNally and Free, 1988) as a glucose-repressible gene (*grg-1*). The *N.crassa ccg-1* gene is an excellent example of a glucose repressible gene as its mRNA level increases over 500-fold within an hour after

Table 3.27 Individual EST totals and contigs for *ccg-1* in combined EST database

NMNE Contig#	EST Totals		alignments		% homolog
	NM	NE	DNA	Amino acid	
Contig1446	254	117	452-240	1-71	100
Contig1410	57	18	535-413	1-41	100
Contig1399	39	19	422-255	1-71	100
Contig634	0	4	477-265	1-71	98
Contig48	2	0	457-245	1-71	91
Total	351	159	NM/NE=351/159*1.2=1.8		

*The total number of ESTs in the NM is 10871. The total EST number in the NE is 9148. Therefore, all the total EST for each *ccg* genes in the NE library were corrected by multiply 1.2. 10871/9148=1.2

glucose starvation and modulates between 5 and 10 fold during the circadian cycle under conditions of glucose deprivation (McNally and Free, 1988). The genomic DNA sequence of this gene (L14464) and the alignment to cDNAs sequences is shown in **Figure 3.06.**

There were a total 510 ESTs for *ccg-1* that were assembled into five different contigs, contig48, contig634, contig1339, contig1410, and contig1446. The alignments

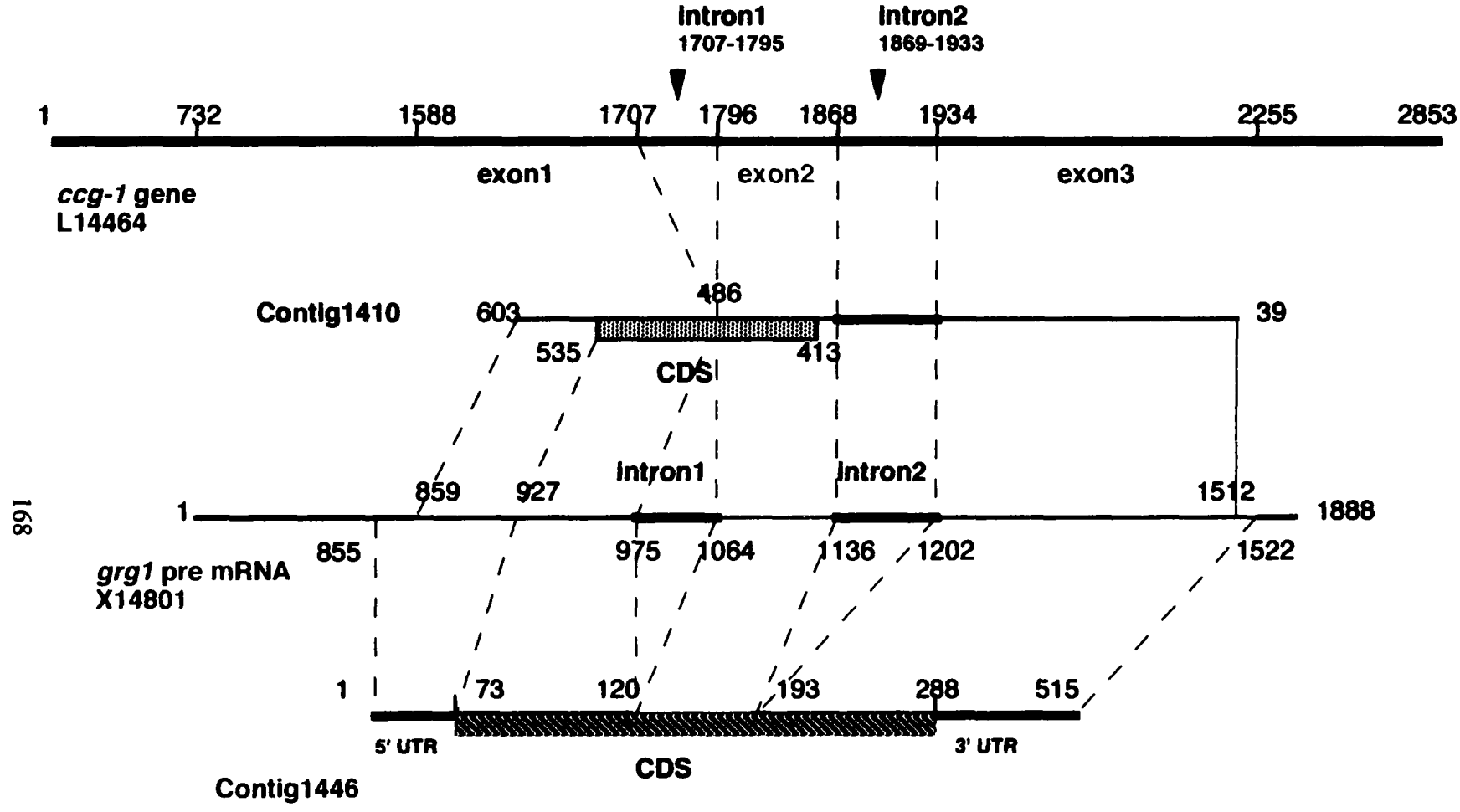


Figure3.06 The annotation of *ccg-1* gene, *ccg-1* mRNA and cDNAs represented by Contig1446, Contig1410. Two introns of the *ccg-1* gene were edited from the Contig1446. Intron2 was not edited from the Contig1410. Dashed line points the positions of alignments.

between these contigs and the *ccg-1* protein and the ratios of homology between them are listed in Table 3.27. Except for contig1410, all the four remaining contigs at amino acid residues 1 through 71 with the *ccg-1* gene protein (X14801). Analysis of this alignment indicates that the mRNA represented by the contig1410 and contig1446 was alternatively spliced. Two introns, located between 975bp and 1064 bp and between 1136 bp and 1202 bp in the sequence of *ccg-1* pre mRNA (X14801), were lacking in contig1446, while only one intron, located between 975bp and 1064 bp on the sequence of the *ccg-1* pre mRNA sequence (X14801), was not present in the sequence of the contig1410 (**Figure3.06 and Figure3.07**). The bases between 409bp to 350 bp on the sequence of contig1410 specify the second intron that was lacking in the contig1446. These bases are 3' $_{350}$CTGAGAGGAAGATGAATGTTAGTGTCTGATGATGGCTGTGAGAGTGGAAG ATGATGGAGCTTAC$_{412}$ 5'. The branch sites on this intron are "CA (=GT)...TC(=AG)", the consensus sequence found in most eukaryote intron splice sites. The three bases ATT (=TAA on its another strand) on the 5' branch site is a stop codon and the translation of mRNA represented by contig1410 therefore would end at this position, explaining why the translated product of contig1410 had only 41 amino acid residues. All five contigs represent gene products from the ccg-1 gene. Except for contig1410, their translation products are exactly same (Figure3.08) and matched with the *ccg-1* gene protein. However, the cDNAs represented by these five contigs have alternative polyadenylation sites (**Figure3.09**). Although all the coding regions of these cDNAs are identical, they had different poly (A) addition sites, and most likely the reason why Phrap could not assembly them together even though they represent the same gene

```
C gi|3013|emb|X14   1512 TTAATTGGCAGAGAGCCGGGTCAATTCCTAGCTAGAACGGACCAGTGTCC 1463

  NMNE.fasta.scre     39 TTAATTGGCAGAGAGCCGGGTCAATTCCTAGCTAGAACGGACCAGTGTCC 88

C gi|3013|emb|X14   1462 ATGCCTATGTAATAGGGGTATCGGTTCTGGTGTAATGCTAAAGCTGCCCT 1413

  NMNE.fasta.scre     89 ATGCCTATGTAATAGGGGTATCGGTTCTGGTGTAATGCTAAAGCTGCCCT 138

C gi|3013|emb|X14   1412 TCGAAATTCCAAGCCAAGTATGTCATAGAGAAGCCCTTGGCGAACTCCTT 1363

  NMNE.fasta.scre    139 TCGAAATTCCAAGCCAAGTATGTCATAGAGAAGCCCTTGGCGAACTCCTT 188

C gi|3013|emb|X14   1362 AGCCGGCCATTATCATCAAGTCATGACTGATAAAAAGGAAATGGAAAACG 1313

  NMNE.fasta.scre    189 AGCCGGCCATTATCATCAAGTCATGACTGATAAAAAGGAAATGGAAAACG 238

C gi|3013|emb|X14   1312 ACTGTTGGTAAAGTCGCTTAGTGGGTAGCACCCTGCTTGTGGGCCTCGGC 1263

  NMNE.fasta.scre    239 ACTGTTGGTAAAGTCGCTTAGTGGGTAGCACCCTGCTTGTGGGCCTCGGC 288

C gi|3013|emb|X14   1262 CTTAGCGTCGTGCTTGTTCTCGGAGACCTTGTCGGAGATGGCATCACCGG 1213

  NMNE.fasta.scre    289 CTTAGCGTCGTGCTTGTTCTCGGAGACCTTGTCGGAGATGGCATCACCGG 338

C gi|3013|emb|X14   1212 CAGCGTTGAGACTGAGAGGAAGATGAATGTTAGTGTCTGATGATGGCTGT 1163

  NMNE.fasta.scre    339 CAGCGTTGAGACTGAGAGGAAGATGAATGTTAGTGTCTGATGATGGCTGT 388

C gi|3013|emb|X14   1162 GAGAGTGGAAAAGTGATGGAGCTTACCGAGTGCCGACGCCCTGGTTGGAG 1113
                              i -
  NMNE.fasta.scre    389 GAGAGTGGAAGA-TGATGGAGCTTACCGAGTGCCGACGCCCTGGTTGGAG 437

C gi|3013|emb|X14   1112 TCCTTGGCAACGTCCTTGTTGGCCTCCTTGGAAGCAGTGGCGGTAGCGCC 1063

  NMNE.fasta.scre    438 TCCTTGGCAACGTCCTTGTTGGCCTCCTTGGAAGCAGTGGCGGTAGCGCC 487

gi|3013|emb|X14801.1|NCGRG1G       859    976 (912)   C NMNE.fasta.screen.Contig1410    (0)
603   486

C gi|3013|emb|X14    976 CCCTGGACCTTGTCACCGACGTAGTTGGCAGCGTTCTTGAGGGTATCCAT 927

  NMNE.fasta.scre    486 CCCTGGACCTTGTCACCGACGTAGTTGGCAGCGTTCTTGAGGGTATCCAT 535

C gi|3013|emb|X14    926 TTTGGTTGATGTGAGGGGTTGTGAAAGTGGTTTGAGTGTTGTGTTGATTT 877

  NMNE.fasta.scre    536 TTTGGTTGATGTGAGGGGTTGTGAAAGTGGTTTGAGTGTTGTGTTGATTT 585

C gi|3013|emb|X14    876 GAAGTAGTGAAGATGTGA 859

  NMNE.fasta.scre    586 GAAGTAGTGAAGATGTGA 603
```

Figure3.07 The cross_match between the contig1410 and the *ccg-1* premRNA sequence to show the intron between 975bp and 1064bp on the *ccg-1* premRNA sequence. The up sequence is from the *ccg-1* premRNA (X14801). The bottom sequence is from the contig1410.

product. Several *ccg* genes identified in this research also have alternative polyadenylation sites. Alternative polyadenylation is common in the human 3' mRNA

processing (Pauws et al., 2001) as it was reported that 29% of the human mRNAs have more than one polyadenylation site. The reason why alternative polyadenylation happens is not known but several groups have proposed explanation. One reason proposed was that the enzyme of polyadenylation uses different pre-RNA cleavage sites in the different development or tissue states (Moreira et al., 1998) via an "alternative cleavage site selection" machinery (Pauws et al., 2001). Another proposal is that the polyadenylation mechanism just randomly uses different polyadenylation hexanucleotide signals (Edwards-Gilbert et al., 1997). In either case, just as has been observed in higher eukaryotes, *N.crassa* mRNAs do contain instance of both alternative splicing and the alternative polyadenylation.

```
            1                                                   50

Nmnectg1399  MDTLKNAANY VGDKVQGATA TASKEANKDV AKDSNQGVGT RLNAAGDAIS
Nmnectg1446  MDTLKNAANY VGDKVQGATA TASKEANKDV AKDSNQGVGT RLNAAGDAIS
Nmnectg1410  MDTLKNAANY VGDKVQGATA TASKEANKDV AKDSNQGVGT R---------
 Nmnectg634  MDTLKNAANY VGDKVQGATA FASKEANKDV AKDSNQGVGT RLNAAGDAIS
  Nmnectg48  MDTLKNGANY VGDKVQGGTA TGSKEANKDV AKDSNQGVGT CFNAAGNAIS

            51                  71
Nmnectg1399  DKVSENKHDA KAEAHKQGAT H
Nmnectg1446  DKVSENKHDA KAEAHKQGAT H
Nmnectg1410  ---------- ---------- -
 Nmnectg634  DKVSENKHDA KAEAHKQGAT H
  Nmnectg48  DKVSENKHDA KAEAHKQGAT H
```

Figure 3.08 The translation products of the five contigs represented the *ccg-1* gene

```
              1                                                   50
  Nmnectg1446  ---------- ---------- ---TTTTT TTTTTGTGAA GGTATTAATT
Nmnectg1399un  ---------TTTTTTTTTT TTTTTTTTT CAGATGTGAA GGTATTAATT
   Nmnectg634  TTTTTTTTTTTTTTTTTTT AGGAAAAAAACAGATGTGAA GGTATTAATT
    Nmnectg48  ---------- -------   TTTTTTTTTTTTTTTTTTTT TTTTTTTTT
  Nmnectg1410  -----TTTT TTTTTTTTT TTTTTTTTTTTTTTTTTTTT TTTTTTAATT

             51                                                  100
  Nmnectg1446  GGCAGAGAGC CGGGTCAATT CCTAGCTAGA ACGGACCAGT GTCCATGCCT
Nmnectg1399un  GGCAGAGAGC CGGGTCAATT CCTAGCTAGA ACGGACCAGT GTCCATGCCT
   Nmnectg634  GGCAGAGAGC CGGGTCAATT CCTAGCTAGA ACGGACCAGT GTCCATGCCT
    Nmnectg48  GGCAAAAAGC CGGGTCAATT CCTAGCTAAA ACGGACCAGT GTCCATGCCT
  Nmnectg1410  GGCAGAGAGC CGGGTCAATT CCTAGCTAGA ACGGACCAGT GTCCATGCCT
```

```
                101                                                   150
  Nmnectg1446   ATGTAATAGG GGTATCGGTT CTGGTGTAAT GCTAAAGCTG CCCTTCGAAA
Nmnectg1399un   ATGTAATAGG GGTATCGGTT CTGGTGTAAT GCTAAAGCTG CCCTTCGAAA
   Nmnectg634   ATGTAATAGG GGTATCGGTT CTGGTGTAAT GCTAAAGCTG CCCTTCGAAA
    Nmnectg48   ATGTAATAGG GGTATCGGTT CTGGTGTAAT GCTAAAGCTG CCCTTCGAAA
  Nmnectg1410   ATGTAATAGG GGTATCGGTT CTGGTGTAAT GCTAAAGCTG CCCTTCGAAA

                151                                                   200
  Nmnectg1446   TTCCAAGCCA AGTATGTCAT AGAGAAGCCC TTGGCGAACT CCTTAGCCGG
Nmnectg1399un   TTCCAAGCCA AGTATGTCAT AGAGAAGCCC TTGGCGAACT CCTTAGCCGG
   Nmnectg634   TTCCAAGCCA AGTATGTCAT AGAGAAGCCC TTGGCGAACT CCTTAGCCGG
    Nmnectg48   TTCCAAGCCA AGTTTGTCAT AAAAAACCCC TTGGCGAACT CCTTAGCCGG
  Nmnectg1410   TTCCAAGCCA AGTATGTCAT AGAGAAGCCC TTGGCGAACT CCTTAGCCGG

                201                                                   250
  Nmnectg1446   CCATTATCAT CAAGTCATGA CTGATAAAAA GGAAATGGAA AACGACTGTT
Nmnectg1399un   CCATTATCAT CAAGTCATGA CTGATAAAAA GGAAATGGAA AACGACTGTT
   Nmnectg634   CCATTATCAT CAAGTCATGA CTGATAAAAA GGAAATGGAA AACGACTGTT
    Nmnectg48   CCATTATCAT CAAGTCATGA CTGATAAAAA GGAAATGGAA AACGACTGTT
  Nmnectg1410   CCATTATCAT CAAGTCATGA CTGATAAAAA GGAAATGGAA AACGACTGTT

                251                                                   300
  Nmnectg1446   GGTAAAGTCG CTTAGTGGGT AGCACCCTGC TTGTGGGCCT CGGCCTTAGC
Nmnectg1399un   GGTAAAGTCG CTTAGTGGGT AGCACCCTGC TTGTGGGCCT CGGCCTTAGC
   Nmnectg634   GGTAAAGTCG CTTAGTGGGT AGCACCCTGC TTGTGGGCCT CGGCCTTAGC
    Nmnectg48   GGTAAAGTCC CTTAGTGGGT AGCACCCTGC TTGTGGGCCT CGGCCTTAGC
  Nmnectg1410   GGTAAAGTCG CTTAGTGGGT AGCACCCTGC TTGTGGGCCT CGGCCTTAGC

                301                                                   350
  Nmnectg1446   GTCGTGCTTG TTCTCGGAGA CCTTGTCGGA GATGGCATCA CCGGCAGCGT
Nmnectg1399un   GTCGTGCTTG TTCTCGGAGA CCTTGTCGGA GATGGCATCA CCGGCAGCGT
   Nmnectg634   GTCGTGCTTG TTCTCGGAGA CCTTGTCGGA GATGGCATCA CCGGCAGCGT
    Nmnectg48   GTCGTGCTTG TTTTCGGAAA CCTTGTCGGA AATGGCATTA CCGGCAGCGT
  Nmnectg1410   GTCGTGCTTG TTCTCGGAGA CCTTGTCGGA GATGGCATCA CCGGCAGCGT

                351                                                   400
  Nmnectg1446   TGAGA..... .......... .......... .......... ..........
Nmnectg1399un   TGAGA..... .......... .......... .......... ..........
   Nmnectg634   TGAGA..... .......... .......... .......... ..........
    Nmnectg48   TGAAA..... .......... .......... .......... ..........
  Nmnectg1410   TGAGACTGAG AGGAAGATGA ATGTTAGTGT CTGATGATGG CTGTGAGAGT

                401                                                   450
  Nmnectg1446   .......... .........C GAGTGCCGAC GCCCTGGTTG GAGTCCTTGG
Nmnectg1399un   .......... .........C GAGTGCCGAC GCCCTGGTTG GAGTCCTTGG
   Nmnectg634   .......... .........C GAGTGCCGAC GCCCTGGTTG GAGTCCTTGG
    Nmnectg48   .......... .........C AAGTGCCGAC CCCCTGGTTG GAGTCCTTGG
  Nmnectg1410   GGAAGATGAT GGAGCTTACC GAGTGCCGAC GCCCTGGTTG GAGTCCTTGG

                451                                                   500
  Nmnectg1446   CAACGTCCTT GTTGGCCTCC TTGGAAGCAG TGGCGGTAGC GCCCTGGACC
Nmnectg1399un   CAACGTCCTT GTTGGCCTCC TTGGAAGCAG TGGCGGTAGC GCCCTGGACC
   Nmnectg634   CAACGTCCTT GTTGGCCTCC TTGGAAGCAA AGGCGGTAGC GCCCTGGACC
    Nmnectg48   CAACGTCCTT GTTGGCCTCC TTGGAACCAG TGGCGGTACC CCCCTGAACC
  Nmnectg1410   CAACGTCCTT GTTGGCCTCC TTGGAAGCAG TGGCGGTAGC GCCCTGGACC

                501                                                   550
  Nmnectg1446   TTGTCACCGA CGTAGTTGGC AGCGTTCTTG AGGGTATCCA TTTTGGTTGA
Nmnectg1399un   TTGTCACCGA CGTAGTTGGC AGCGTTCTTG AGGGTATCCA TTTTGGTTGA
   Nmnectg634   TTGTCACCGA CGTAGTTGGC AGCGTTCTTG AGGGTATCCA TTTTGGTTGA
    Nmnectg48   TTGTCACCGA CGTAATTGGC ACCGTTCTTG AGGGTATCCA TTTTGGTTGA
```

```
  Nmnectg1410  TTGTCACCGA CGTAGTTGGC AGCGTTCTTG AGGGTATCCA TTTTGGTTGA

               551                                                600
  Nmnectg1446  TGTGAGGGGT TGTGAAAGTG GTTTGAGTGT TGTGTTGATT TGAAGTAGTG
Nmnectg1399un  TGTGAGGGGT TGTGAAAGTG GTTTGAGTGT TGTGTTGATT TGAAGTAGTG
   Nmnectg634  TGTGAGGGGT TGTTGAAATC GGTTGAATGG TTTGGTGATT AAAACGGGGA
    Nmnectg48  TTTTAGGGGT TGTAAAAGTG GTTTGAATGT TTGTATTATT TAAATTTTTT
  Nmnectg1410  TGTGAGGGGT TGTGAAAGTG GTTTGAGTGT TGTGTTGATT TGAAGTAGTG

               601
  Nmnectg1446  AAGATGTGAT
Nmnectg1399un  AAGATGTGA-
   Nmnectg634  AAAATGGGAA
    Nmnectg48  TAATATACTA
  Nmnectg1410  AAGATGTGA-
```

Figure 3.09 The alternative polyadenylation sites in the five contigs represented the *ccg-1* gene. The translation start site (pink color font) and stop sites (blue color font) are bolded. The intron is underlined and the different polyadenylation sites are also marked by red "T"s.

The observed *ccg-1* EST totals from the NE library were 159 while those from the NM library were 349. The ratio of *ccg-1* between the NM library and the NE library is 1.8 after correction. Totally, the morning cDNA library had 10871 ESTs while the NE cDNA library had 9148 ESTs, a ratio of ~1.2. Therefore, the EST totals from the NE cDNA library could be normalized by multiplying by 1.2. This result further confirmed that *ccg-1* is a morning-specific clock-controlled gene, when coupled with the observation that RNA Northern blot analysis reveals a peak for *ccg-1* mRNA transcription in strain *frq*^{7} at DD43 (switch to constant dark 43 hrs), while the peak for *ccg-1* mRNA is 180° out of phase for *frq*+ strain (Bell-Pedersen et al., 1996c).

ccg-1 has been independently isolated using two different screening approaches, either a screen for circadian clock regulated genes (Loros et al., 1989) or a screen for glucose repressible genes (*grg-1*) (McNally and Free, 1988). Even though the level of the *ccg-1* gene expression is under tight negative control by glucose, it also is induced by

light and this induction is independent of the direct effects of light on the clock. In a *Neurospora* mutant strain lacking a functional clock (a null mutation at the *frq* locus), the expression of *ccg-1* fluctuates in concert with changes in developmental potential. But this fluctuation is not constitutive (Aronson et al., 1994), but requires the products of the *wc* genes.

3.2.1.1 The *ccg-1* and carbon catabolitic repression in fungi

ccg-1 is a glucose-repressible gene in *N.crassa* (McNelly and Free, 1988). Glucose repression affects a large number of genes that are required for the metabolism of alternative carbon sources. For example, enzymes required for metabolism of the sugars sucrose, maltose, and galactose, and non-fermentable carbon source such as glycerol, ethanol, and acetate, as well as gluconeogenic and respiratory enzymes are under control of glucose. Glucose repression in the cells is a good example of the coordinate control of a large number of genes by an environmental signal, the external glucose concentration. When cells grown on glucose media, many enzymes that are needed to utilize alternate carbon sources have lower levels. When the glucose in medium is exhausted, the syntheses of these enzymes is induced, by a phenomenon called glucose repression or carbon catabolic repression. The genes that are repressed by the appearance of glucose and induced by the exhaust of glucose are called glucose-repressible genes. In many cases, glucose repression acts at the level of transcription. This glucose repression mechanism also seems to be conserved in both insects and yeast (Ebbole, 1998).

ccg-1 gene in *N.crassa* is repressed by the addition of glucose to the medium.

Mutations in the *ccg-1* gene can effect both the blue-light and circadian response of *Neurospora crass* too. Two cis-acting sites required for this repression have been identified in the upstream untranscribed region of the *ccg-1* gene (Wang and Free, 1994).

3.2.1.2 The relationship between glucose repression and conidiation in *N.crassa*

As described above, since the *rco-1* gene (b5c06ne.r1), two conidiation-specific genes, *con-6* (contig1098) and *con-10* (contig1116), only were detected in the NE cDNA library, the *rco-1* gene product may play a role in inducing expression of the *ccg-1* gene in the *Neurospora crassa* cDNA libraries. It is known that there is a strong correlation between carbon catabolite repression and development of macroconidia exists in *N.crassa* (Ebbole, 1998). Also, desiccation and the deprivation of carbon and nitrogen are other factors that activate or stimulate conidiation (Springer, 1993). Under conditions of drying or nutrient starvation, aerial hyphae grow upward from the substrate that leads to the formation of macroconidia in mycelia (Springer and Yanofsky, 1989) and *rco-1* gene product links carbon catabolite repression and conidiation. Therefore, *rco-1* (regulators of conidiation) is a regulatory gene that mediates the repression of conidiation gene expression under stress (Lee and Ebbole, 1998). For example, for *con-10* and *ccg-1* of *Neurospora*, *rco-1* releases the repression and allows them a tissue-specific response to environmental stress and permits the induction of these conidiation-specific genes (Yamashiro et al., 1996). The *rco-1* gene product also is responsible for regulating genes that are essential for developmental pathways. Therefore, *con-10*, which is expressed in all three sporulation pathways (Corrochano et al., 1995), and *ccg-1* gene, are the targets

that are regulated by *rco-1* gene (Yamashiro, et al., 1996; Madi et al., 1997).

Acetyl-CoA hydrolase of *Saccharomyces cerevisiae,* is an enzyme that catalyzes the hydrolysis of acetyl-CoA encoded by a glucose repressible gene (Lee, F-J.S. et al., 1990). A series of acetyltransferases including yeast N-acetyltransferase, rat brain pyruvate carboxylase, choline acetyltransferase, and [acyl-carrier-protein] acetyltransferase are inhibited by acetyl-CoA hydrolase. It has been reported that arylalkylamine N-acetyltransferase in the rat pineal gland plays a key role in generation of the circadian rhythmicity of melatonin synthesis (Pevet, 2000). The level of melatonin also plays a key role in regulating gene expression and transcription by the clock in mammalian suprachiasmatic nucleus (SCN). Thus, acetyl-CoA hydrolase may be involved in the overall regulation of circadian rhythms in mammalian (Lee, F-J. S. et al., 1990).

The relationship between the level of glucose and the regulation of circadian rhythm is not know, however, if it is assumed that the glucose-repressible gene (including *grg-1* in *N.crassa*), acetyl-CoA hydrolase, and N-acetyltransferase may interact in some way to affect the concentration of melatonin, this may cause the oscillation in mammals and regulate their circadian clock.

3.2.2 The clock-controlled gene-2

The *ccg-2* gene is one of the highly expressed clock-controlled genes in both cDNA libraries studied. Like the *ccg-1* gene, it also is highly transcribed in the morning cDNA library confirming earlier reports that it is a morning specific gene (Bell-Pedersen,

et al., 1996c). The annotation of *ccg-2* and Contig1447 is showed in **Figure 3.10.** The expression level of this gene in these two cDNA libraries is presented in Table 3.28.

Table 3.28 The contig and the EST totals for the *ccg-2* in combined EST database

Contig#	EST totals		alignments		% homolog
	NM	NE	DNA	Amino acid	
NMNE Contig1447	457	207	104-427	1-108	79
Total	457	207	NM/NE=457/207*1.2=1.8		

From table 3.28, it is apparent that after normalization 457/207*1.2, the ratio of the gene expression of *ccg-2* in these two libraries is ~1.8. The sequence comparison between the contig1447 and the *ccg-2* mRNA (X67339) shows that both DNA sequences and the translated protein sequences are perfectly matched (Figure1.11 and 3.12) with a similarity of 100%.

```
percent Similarity: 100.000   Percent Identity: 100.000
NMNEctg1447.Gcg x Ccg2mRNA.Gcg February 14, 2001
        .         .         .         .         .
  1 CAACATCTTCACTTCACAACATCTTCTCAACCTTCCAACTCACCTTCCAA 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 12 CAACATCTTCACTTCACAACATCTTCTCAACCTTCCAACTCACCTTCCAA 61

 51 ACCACCTTCAAAACCAACTCCCAGCTTCTTTCAGCAAACCCCCAACCGCC 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 62 ACCACCTTCAAAACCAACTCCCAGCTTCTTTCAGCAAACCCCCAACCGCC 111
        .         .         .         .         .
101 AAAATGCAGTTCACCAGCGTCTTCACCATCCTCGCCATTGCCATGACCGC 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
112 AAAATGCAGTTCACCAGCGTCTTCACCATCCTCGCCATTGCCATGACCGC 161
        .         .         .         .         .
151 CGCTGCGGCCCCGGCTGAGGTTGTTCCCCGCGCCACCACCATCGGCCCCA 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
162 CGCTGCGGCCCCGGCTGAGGTTGTTCCCCGCGCCACCACCATCGGCCCCA 211
        .         .         .         .         .
201 ACACCTGCTCCATCGACGACTACAAGCCTTACTGCTGCCAGTCTATGTCC 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
212 ACACCTGCTCCATCGACGACTACAAGCCTTACTGCTGCCAGTCTATGTCC 261
        .         .         .         .         .
251 GGCCCCGCCGGCTCCCCTGGTCTCCTCAACCTCATCCCCGTCGACCTCAG 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
262 GGCCCCGCCGGCTCCCCTGGTCTCCTCAACCTCATCCCCGTCGACCTCAG 311
```

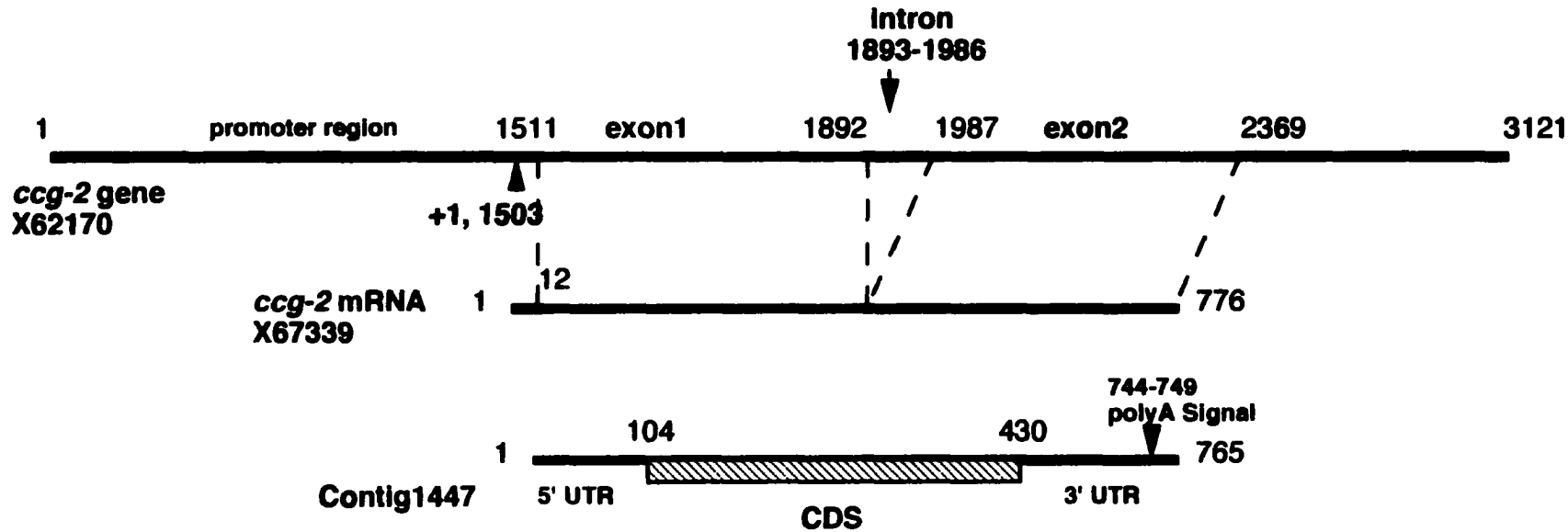


Figure 3.10 The annotation of *ccg-2* gene, *ccg-2* mRNA, and the cDNA represented by contig1447

```
301 CGCCTCGCTCGGCTGCGTTGTCGGTGTCATCGGCTCCCAATGTGGTGCCA 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
312 CGCCTCGCTCGGCTGCGTTGTCGGTGTCATCGGCTCCCAATGTGGTGCCA 361
           .         .         .         .         .
351 GCGTCAAGTGCTGCAAGGACGATGTTACCAACACCGGCAACTCCTTCCTC 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||
362 GCGTCAAGTGCTGCAAGGACGATGTTACCAACACCGGCAACTCCTTCCTC 411
           .         .         .         .         .
401 ATCATCAACGCTGCCAACTGCGTTGCCTAAGTGTTTACGCGGCAACAGCG 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||
412 ATCATCAACGCTGCCAACTGCGTTGCCTAAGTGTTTACGCGGCAACAGCG 461
           .         .         .         .         .
451 CAAAGTCTAGGCAATGCCTTGTTCTCAACGCTGCTGCCAGTCCAGCACCC 500
    ||||||||||||||||||||||||||||||||||||||||||||||||||
462 CAAAGTCTAGGCAATGCCTTGTTCTCAACGCTGCTGCCAGTCCAGCACCC 511
           .         .         .         .         .
501 CCCTTCTGCAGCAAGGAGCCCCCTTCTGCTGGACTGGCAGCACAACGAGC 550
    ||||||||||||||||||||||||||||||||||||||||||||||||||
512 CCCTTCTGCAGCAAGGAGCCCCCTTCTGCTGGACTGGCAGCACAACGAGC 561
           .         .         .         .         .
551 TGCTACTACAACACAAGCATCATGCCTGGACGCAACAGAAGCCGATAATC 600
    ||||||||||||||||||||||||||||||||||||||||||||||||||
562 TGCTACTACAACACAAGCATCATGCCTGGACGCAACAGAAGCCGATAATC 611
           .         .         .         .         .
601 TTGGGGTTTGGTTTTGGGGGATGAAGGTGATGAGTTGATGGATTGGATCG 650
    ||||||||||||||||||||||||||||||||||||||||||||||||||
612 TTGGGGTTTGGTTTTGGGGGATGAAGGTGATGAGTTGATGGATTGGATCG 661
           .         .         .         .         .
651 ATATCTTACAATGCGTGTCTCTTCCTGTTAAGATCTGCTTTACTATTTTC 700
    ||||||||||||||||||||||||||||||||||||||||||||||||||
662 ATATCTTACAATGCGTGTCTCTTCCTGTTAAGATCTGCTTTACTATTTTC 711
           .         .         .         .         .
701 CTATTTTCTTTTACACATAGCTATGTATCACTAAGGCCTGGTGATTAATA 750
    ||||||||||||||||||||||||||||||||||||||||||||||||||
712 CTATTTTCTTTTACACATAGCTATGTATCACTAAGGCCTGGTGATTAATA 761
           .
751 CACTCTCTTAACCCT 765
    |||||||||||||||
762 CACTCTCTTAACCCT 776
```

Figure 3.11 The DNA sequence comparison between the NMNEcontig1447 and the *ccg-2* mRNA(X67339)

```
NMNEctg1447 1 MQFTSVFTILAIAMTAAAAPAEVVPRATTIGPNTCSIDDYKPYCCQSMSG 50
              ||||||||||||||||||||||||||||||||||||||||||||||||||
ccg2 mRNA     MQFTSVFTILAIAMTAAAAPAEVVPRATTIGPNTCSIDDYKPYCCQSMSG 50

           51 PAGSPGLLNLIPVDLSASLGCVVGVIGSQCGASVKCCKDDVTNTGNSFLI 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||
           51 PAGSPGLLNLIPVDLSASLGCVVGVIGSQCGASVKCCKDDVTNTGNSFLI 100

          101 INAANCVA 108
              ||||||||
          101 INAANCVA 108
```

Figure 3.12 The translated protein sequence comparison between the NMNEcontig1447 and the *ccg-2* gene mRNA (X67339).

3.2.2.1 The *ccg-2* gene was also isolated as *eas* and *bli-7* genes in the different screens

Light and the circadian clock regulate the transcription level of the *ccg-2* gene independently (Arpaia et al., 1993; 1995). Both *ccg-2* and *ccg-1* are active in the subjective morning. This is the time of day when *Neurospora* undergoes asexual differentiation that results in the production of conidia.

The eas (*ccg-2*) locus was originally identified through a mutation in a morphologically abnormal strain with easily-wettable conidiospores (Beever and Dempsey, 1978; Lauter et al., 1992). In separated screens, the *eas* gene was independently isolated on the basis of rhythmic abundance of transcript as *ccg-2* and as a blue-light-induced gene *bli-7* (*ccg-2=eas=bli-7*). This *ccg-2* gene encodes a fungal hydrophobin that is the major constitute of the hydrophobic, rodlet coating on the surface of maturing asexual conidiospores (Bell-Pedersen et al., 1992; Templeton et al., 1995). The mature message encoded by *ccg-2* is 776 nucleotides length, which specifics an ORF for a polypeptide of 108 amino acids. In *ccg-2* gene product, about 40-45% residues are hydrophobic. The *ccg-2* gene is transcriptionally activated by the circadian clock in a time-of-day-specific manner. Analysis of the sequence of *ccg-2* results reveals that the polypeptide sequence of *ccg-2* is significantly similar to a class of hydrophobic, low molecular weight, cysteine-rich proteins, such as a small hydrophobic protein, RodA (Figure3.13) of *Aspergillus nidulans* (M61113) (Stringer et al., 1991). RodA, a hydrophobin, is encoded by the *rodA* gene and is developmentally regulated. Hydrophobin is an integral component of fungal asexual spores, which is important for spore dispersal (Kershaw and Talbot, 1998). In *CCG-2* and RodA, the mostly conserved

residues are hydrophobic. Figure 3.14 and 3.15 shows the hydrophobicity plots of *CCG-2* and RodA obtained using the Kyte-Doolittle algorithm with a window size of 7. In this plot, points below the x-axis indicate hydrophilic domains and hydrophobic region occur above the x-axis. Both *CCG-2* and RodA contain a strong hydrophobic amino terminus (characteristic of a signal sequence) and a central hydrophobic domain (Stringer et al., 1991). Both polypeptides have 8 conserved cysteine residues which may be involved in the formation of the disulfide bonds necessary for cross-linking the individual components. The structure of *CCG-2* and RodA are very similar, suggesting that they have similar functions in both *N.crassa* and *A. nidulans*. Inactivation of the *rodA* gene or *ccg-2* gene give a mutant phenotype in which the cells of conidiophores and mature conidia lack the rodlet layer, making them easy to wet than in wild type and resulting in inefficient spore dispersal in nature.

```
CCG2     1 MQFTSV.FTILAIAMTAAA.APA...E.....V.......V..PRATTIG 31
RodA     1 MKF.SIAAAVVAFAASVAALPPAHDSQFAGNGVGNKGNSNVKFP....V. 44

        32 P.N.T...CSID...DYKPY.CC.Q.SMSGPA.....G..SPGLLN.LIP 62
        45 PENVTVKQAS.DKCGDQAQLSCCNKATYAGDTTTVDEGLLS.GALSGLIG 92

        63 VDLS.A.SLG....C..V.VGV.IG.....SQ.CGASVKCCKDDVTNT.G 95
        93 AG.SGAEGLGLFDQCSKLDVAVLIGIQDLVNQKCKQNIACCQNSPSSADG 141

        96 NSFLI.INAANCVA..... 108
       142 N..LIGVGLP.CVALGSIL 157
```

Figure3.13 The protein sequence comparison between the *Neurospora crassa* CCG2 protein and the *Aspergillus nidulans* RodA (M61113) using GCG program GAP. The conserved 8 cysteines are bolded and underlined.

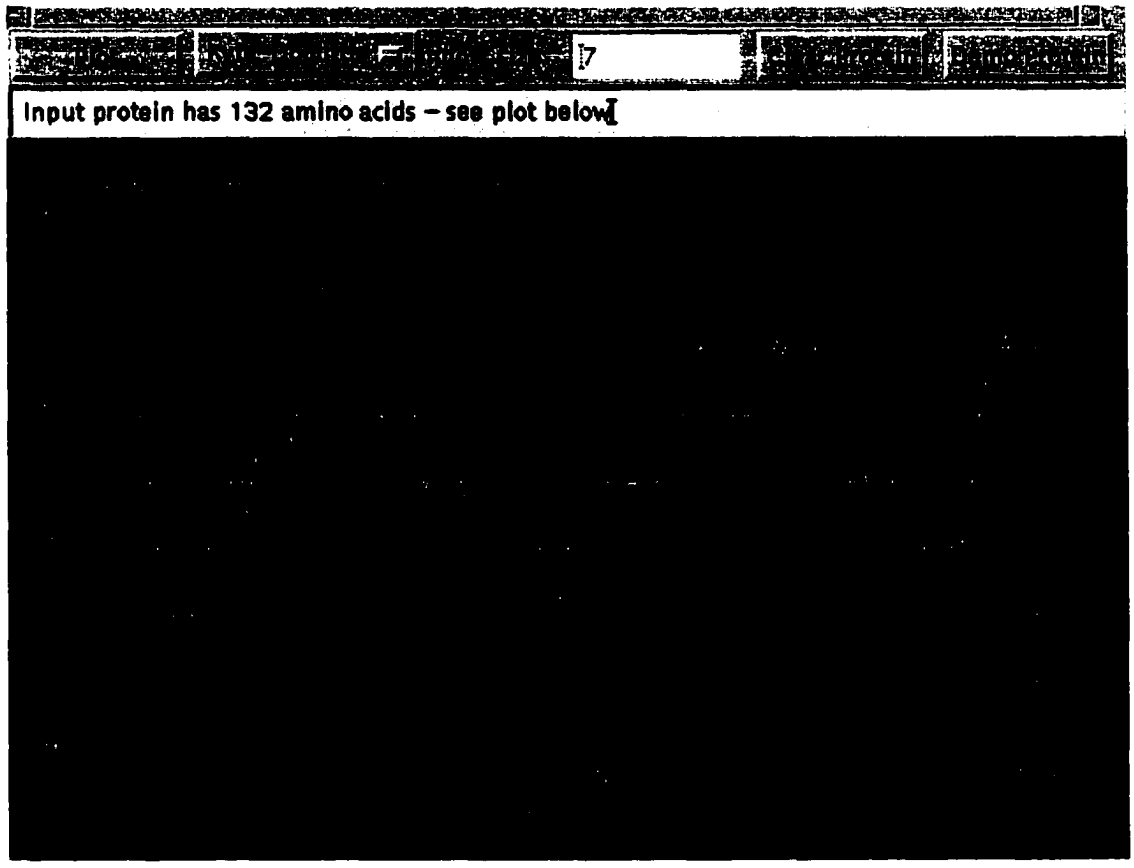


Figure 3.14 The hydrophobicity plot of the *Neurospora crassa* CCG2 protein (108 amino acids) using Kyte-Doolittle method. The residues above the X-axis are hydrophobic.

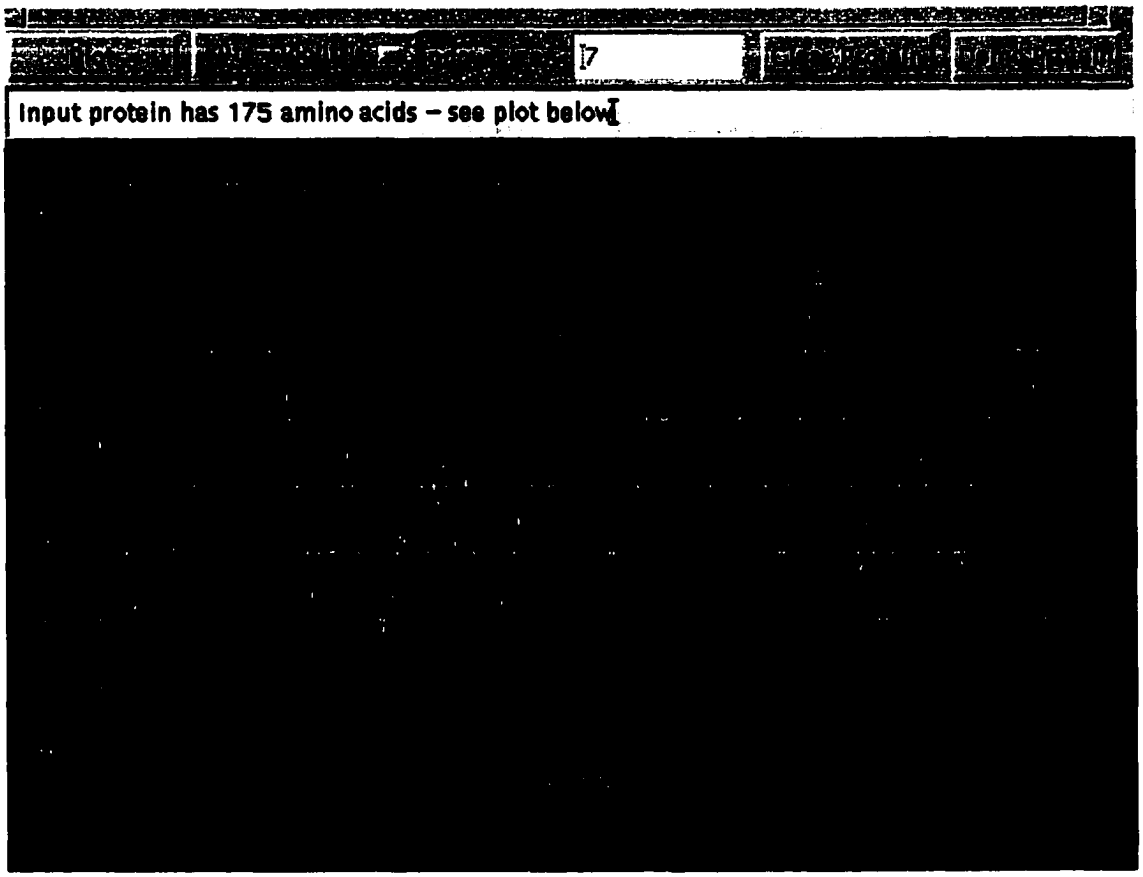


Figure 3.15 The hydrophobicity plot of the *Aspergillus nidulans* RodA protein (M61113, 157 amino acids) using Kyte-Doolittle method. The hydrophobic residues are above the X-axis.

3.2.2.2 The regulation of *ccg-2* gene and the comparisons of the promoters of three *ccg* genes

The expressions of three clock-controlled genes that were detected in this research, *ccg-1, con-10, ccg-2* are regulated by both light and development and their expression undergoes a circadian oscillation in constant darkness (Loros et al., 1989) which also is regulated by changes in the ambient light intensity (Arpaia et al., 1993).

The sequences of the promoters of the *ccg-2* and *ccg-1* genes have 73.7% similarity. TATA box is a feature of inducible genes, which is absent in housekeeping genes (Dynan, 1989), and it is present in the promoter region of all three genes. The promoter region of the *ccg-2* gene also contains two repeat units, CAACATATTC and ACCTTCCAA. Although the function of these repeats is not clear, one of them is highly conserved between *ccg-1* and *ccg-2*, this region may be involved in regulating the expression of these genes.

ccg-2 is another gene that not only is transcriptionally regulated by the circadian clock but also is positively regulated by light. Therefore, knowing the basis of this complex regulation will be very important and useful to understand the interaction between the endogenous biological clock and outside environmental stimulation factors such as light, temperature and nutrition in regulating transcription and expressions of these genes.

Two light elements, a positive light element and a light response element, and the Activating Clock Element (ACE) are present in the promoter of *ccg-2* gene. An 8-nucleotide region GTTGGGAT located inside the ACE box of *ccg-2* promoter, occurs in the promoter region of all three clock-controlled genes. The function of this 8-nucleotide

box is not clear although ACE is a unique sequence element that has been shown by EMSA (electrophoresis mobility shift assay) to bind a nuclear factor from nuclear protein extracts of *N.crassa* (Bell-Pedersen et al., 1996a).

In addition to fungi, the dual clock and light regulation also has been observed in the CAB2 gene in plants (Anderson and Kay, 1995) and in the N-acetyltransferase and retinal transducin genes in human (Brann and Cohen, 1987; Zate, 1993). The CAB2 gene, a light-harvesting chlorophyll a/b-binding protein of chloroplasts, is a well-understood example of a light and clock regulated gene (Anderson et al., 1994). CUF-1 (*CAB* upstream factor 1) and CGF-1 (*CAB* GATA factor 1) are two factors isolated from tobacco nuclear proteins that are involved in the regulation of CAB2 gene (Anderson and Kay, 1995). The mutational studies reveal that the CUF-1 only functions as a positive factor to increase *CAB2* expression levels while CGF-1 plays a role in response to phytochrome activation and contributes to the light-induced high-amplitude circadian oscillation in CAB2 expression.

A GATA repeated motif that is well conserved in the CAB promoters of several species (Gidoni, D. et al., 1989) is the binding site for CGF-1. This GATA motif is located in a 78 bp region of the Cab2 promoter that is required for both light and clock regulation of *Arabidopsis* CAB2 gene (Anderson et al., 1994). The CGF-1 binding site and this 78 bp region are proposed the possible genomic targets for the circadian clock and/or light signal transduction pathways because the GATA motif is conserved in the light-responsive promoters for the chloroplast PSBD-PSBC operon from five cereals and three dicot species (Christopher et al., 1992), and in the light-induced chloroplast PETE

promoter (Haley and Bogorad, 1990). A possible GATA motif also was identified in the upstream sequence of the *ccg-2, con-10 and CAB2* genes. However, no GATA motif sequence was detected in the promoter of *ccg-1*. In *Neurospora crassa*, there is mounting evidence that indicates that the FRQ protein interacts with GATA motif (Dunlap, 1999; Talora et al., 1999; Denault et al., 2001). The features of the promoters of these ccg genes are summarized in the table 3.29.

Table3.29 The summary of the features of the promoters of four *ccg* genes

ccg genes	GATA motif	ACE	CGF-1	GTTGGGAT	CAACATATTC	positive light element	light response element
ccg-1				+	+		
ccg-2	+	+		+	+	+	+
con-10	+			+			
CAB2	+		+				

3.2.3 The clock-controlled gene 7

The clock-controlled gene *7* was the most highly expressed clock-controlled gene in these two cDNA libraries. Totally, there were 870 *ccg-7* EST in the combined EST database (Table 3.30). 692 of these ESTs were from the NM cDNA library while 178 were from the NE cDNA library to give a ratio of the redundancy of this gene in these two cDNA libraries of 3.2. As with the two other highly expressed *ccgs* in *Neurospora*, *ccg-1* and *ccg-2* gene, *ccg-7* also is considered a morning specific gene. Since it has occur more often in the NM cDNA library.

Table 3.30 The contigs and the EST totals of *ccg-7* gene in combined EST database

NMNE contig#	EST totals	Alignments	% homolog

	NM	NE	DNA	amino acid	
Contig1448	537	132	216-1229	1-338	99
Contig1425	84	18	480-217	251-338	100
Contig1394	31	23	591-199	208-338	99
Contig1149	13	1	366-190	280-338	100
Contig975	6	2	556-176	212-338	93
Contig273	0	2	375-109	250-338	95
Contig53	2	0	568-335	212-289	83
			348-187	285-338	96
Total	692	178	NM/NE=692/178*1.2=692/214=3.2		

In the combined EST database, there are seven different contigs representing the *ccg-7* gene. They all could be aligned to the *ccg-7* gene sequence (U65745). Both the DNA sequences and the translated protein sequences of these seven contigs were compared each using the GCG multiple sequence compare program Pileup. These results are shown in the Figure 3.16 and 3.17. Except for the poly (T) region, the coding regions are exactly same for all six contigs, and their translated protein sequences also are identical. However, contig53 had an observed frame shift probably because this sequence was of low quality. Even so, the 5' sequence of contig53 had 96% homolog to the *ccg-7*

```
                  1                                                  50
Nmneccg7ctg1448   -TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTCTGGA
 Nmneccg7ctg273   ---------- ---------- ---------- ---------- ----------
 Nmneccg7ctg975   ---------- ---------- ---------- ----------TTTTTTTTT
  Nmneccg7ctg53   ---------- ---------- ---------- --TTTTTTT TTTTTTTGGA
Nmneccg7ctg1394   ---------- ---------TT TTTTTTTTT TTTTTAGGAA CAGAACTGGA
Nmneccg7ctg1425   TTTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTTT TTTTTTTGGA
Nmneccg7ctg1149   ---------- ---------- -------TTT TTTTTTTTT TTTTTTTGGA

                  51                                                 100
Nmneccg7ctg1448   AGTGAACTTG GTATCTATTT GAGTTCGGAT ATTATTTTAC TTTAAGCAGG
 Nmneccg7ctg273   ---------- ---------- ---------- ---------- ----------
 Nmneccg7ctg975   TTTTTTTTT TTTTCTATTT GAGTTCGGAT ATTATTTTAC TTTAAGCAGG
  Nmneccg7ctg53   AGTGAACTTG GTATCTATTG GAGTTCGGAT ATTATTTTAC TTTAAGCAGG
Nmneccg7ctg1394   AGTGAACTTG GTATCTATTT GAGTTCGGAT ATTATTTTAC TTTAAGCAGG
Nmneccg7ctg1425   AGTGAACTTG GTATCTATTT GAGTTCGGAT ATTATTTTAC TTTAAGCAGG
Nmneccg7ctg1149   AGTGAACTTG GTATCTATTT GAGTTCGGAT ATTATTTTAC TTTAAGCAGG

                  101                                                150
```

```
Nmneccg7ctg1448  ACTGCCCGAG ATCCAACATG CTCGAGTTCA ATCCCACGGA ATCATTGCCC
 Nmneccg7ctg273  --------TT TTTTTTTTTT TTTTTTTTTT TTCCCACGGA ATCATTGCCC
 Nmneccg7ctg975  ACTGCCCGAG ATCCAACATG CTCGAGTTCA ATCCCACGGA ATCATTGCCC
  Nmneccg7ctg53  ACTGCCCGAG ATCCAACATG CTCGAGTTCA ATCCCACGGA ATCATTGCCC
Nmneccg7ctg1394  ACTGCCCGAG ATCCAACATG CTCGAGTTCA ATCCCACGGA ATCATTGCCC
Nmneccg7ctg1425  ACTGCCCGAG ATCCAACATG CTCGAGTTCA ATCCCACGGA ATCATTGCCC
Nmneccg7ctg1149  ACTGCCCGAG ATCCAACATG CTCGAGTTCA ATCCCACGGA ATCATTGCCC

                 151                                             200
Nmneccg7ctg1448  TTTTTTTTTC TTCATATCTA GGAAAAGACC ATTACTTCAT AACCGTGCGG
 Nmneccg7ctg273  TTTTTTTTTC TTCATATCTA GGAAAAGACC ATTACTTCAT AACCGTGCGG
 Nmneccg7ctg975  TTTTTTTTTC TTCATATCTA GGAAAAGACC ATTACTTCAT AACCGTGCGG
  Nmneccg7ctg53  TTTTTTTTTC TTCATATCTA GGAAAAGACC ATTACTTCAT AACCGTGCGG
Nmneccg7ctg1394  TTTTTTTTTC TTCATATCTA GGAAAAGACC ATTACTTCAT AACCGTGCGG
Nmneccg7ctg1425  TTTTTTTTTC TTCATATCTA GGAAAAGACC ATTACTTCAT AACCGTGCGG
Nmneccg7ctg1149  TTTTTTTTTC TTCATATCTA GGAAAAGACC ATTACTTCAT AACCGTGCGG

                 201                                             250
Nmneccg7ctg1448  GTACGCAACC GATTTAAGCC TTCTTGGCAT CGACCTTGGA GATGTAGGAG
 Nmneccg7ctg273  GTACGCAACC GATTTAAGCC TTCTTGGCAT CGACCTTGGA GATGTAGGAA
 Nmneccg7ctg975  GTACGCAACC GATTTAAGCC TTCTTGGCAT CGACCTTGGA GATGTAGGAG
  Nmneccg7ctg53  GTACGCAACC GATTTAAGCC TTCTTGGCAT CGACCTTGGA AATGTAGGAA
Nmneccg7ctg1394  GTACGCAACC GATTTAAGCC TTCTTGGCAT CGACCTTGGA GATGTAGGAG
Nmneccg7ctg1425  GTACGCAACC GATTTAAGCC TTCTTGGCAT CGACCTTGGA GATGTAGGAG
Nmneccg7ctg1149  GTACGCAACC GATTTAAGCC TTCTTGGCAT CGACCTTGGA GATGTAGGAG

                 251                                             300
Nmneccg7ctg1448  ATGAGGTCGA GGACACGGCG AGAGTAGCCC CACTCGTTGT CGTACCAGGA
 Nmneccg7ctg273  ATGAGGTCGA GGACACGGCG AGAGTAGCCC CACTCGTTGT CGTACCAGGA
 Nmneccg7ctg975  ATGAGGTCGA GGACACGGCG AGAGTAGCCC CACTCGTTGT CGTACCAGGA
  Nmneccg7ctg53  ATGAGGTCGA GGACACGGCA AGAGTACCCC CACTCGTTGT CGTACCAGGA
Nmneccg7ctg1394  ATGAGGTCGA GGACACGGCG AGAGTAGCCC CACTCGTTGT AGTACCAGGA
Nmneccg7ctg1425  ATGAGGTCGA GGACACGGCG AGAGTAGCCC CACTCGTTGT CGTACCAGGA
Nmneccg7ctg1149  ATGAGGTCGA GGACACGGCG AGAGTAGCCC CACTCGTTGT CGTACCAGGA

                 301                                             350
Nmneccg7ctg1448  GACAAGCTTG ACGAAGTTCT TGTTGAGGGA GATACCAGCC TTGGCATCGA
 Nmneccg7ctg273  GACAAGCTTG ACGAAGTTCT TGTTGAGGGA GATACCAGCC TTGGCATCCA
 Nmneccg7ctg975  GACAAGCTTG ACGAAGTTCT TGTTGAGGGA GATACCAGCC TTGGCATCGA
  Nmneccg7ctg53  GACAAGCTTG ACAAAGTTCT TGTTGAGGGA GATACCAGCC TTGGCATCGA
Nmneccg7ctg1394  GACAAGCTTG ACGAAGTTCT TGTTGAGGGA GATACCAGCC TTGGCATCGA
Nmneccg7ctg1425  GACAAGCTTG ACGAAGTTCT TGTTGAGGGA GATACCAGCC TTGGCATCGA
Nmneccg7ctg1149  GACAAGCTTG ACGAAGTTCT TGTTGAGGGA GATACCAGCC TTGGCATCGA

                 351                                             400
Nmneccg7ctg1448  AGATGGAGGA GGCGGGGTTG CCGTTC.ATG TCGGAAGAGA CAACCTCATC
 Nmneccg7ctg273  AAATGGAGGA GGCGGGGTTG CCGTTC.ATG TCGGAAGAAA CAACCTCATC
 Nmneccg7ctg975  AGATGGAGGA GGCGGGGTTG CCGTTC.ATG TCGGAAGAGA CAACCTCATC
  Nmneccg7ctg53  AAATGGAGGA GGCGGGGTTG CCGTTCAATG TCGGAAGAGA CAACCTTATC
Nmneccg7ctg1394  AGATGGAGGA GGCGGGGTTG CCGTTC.ATG TCGGAAGAGA CAACCTCATC
Nmneccg7ctg1425  AGATGGAGGA GGCGGGGTTG CCGTTC.ATG TCGGAAGAGA CAACCTCATC
Nmneccg7ctg1149  AGATGGAGGA GGCGGGGTTG CCGTTC.ATG TCGGAAGAGA CCACATAACG

                 401                                             450
Nmneccg7ctg1448  CTCGGTGTAG GCAAGGATGC CAGCGAGGGG ACCCTCAGAG GCCTTCTTGA
 Nmneccg7ctg273  CTCGGTGTAG GCAAGGATGC CACCCAGGGG ACCCTCAAAG GCCTTCTTGA
 Nmneccg7ctg975  CTCGGTGTAG GCAAGGATGC CAGCGAGGGG ACCCTCAGAG GCCTTCTTGA
  Nmneccg7ctg53  CTCGGTGTAG GCAAGGATGC CAGCAAGGGG ACCCTAAAAG GCCTTCTTGA
Nmneccg7ctg1394  CTCGGTGTAG GCAAGGATGC CAGCGAGGGG ACCCTCAGAG GCCTTCTTGA
Nmneccg7ctg1425  CTCGGTGTAG GCAAGGATGC CAGCGAGGGG ACCCTCAGAG GCCTTCTTGA
Nmneccg7ctg1149  ATTTCACACG TCTCGTTCAA CTACCAGACA AACATATAAT CGCAACATCA
```

```
                451                                                 500
Nmneccg7ctg1448 TGACCTCCTT GATCTCATCG TAGGTAGCAC CCTTCTCGA. .TGCGGGCAG
 Nmneccg7ctg273 TGACCTCCTT GATCTCATCG GAGGTAGCAC CCTTCA---- ----------
 Nmneccg7ctg975 TGACCTCCTT GATCTCATCG TAGGTAGCAC CCTTCTCGA. .TGCGGGCAG
  Nmneccg7ctg53 TGACCTCCTT GATCTCATCG TAGGTAGCAC CCTTTTTAA. .TGCGGGCAT
Nmneccg7ctg1394 TGACCTCCTT GATCTCATCG TAGGTAGCAC CCTTCTCGA. .TGCGGGCAG
Nmneccg7ctg1425 TGACCTCCTT GATCTCATCG TAGGTAGCAC CCGTTCCCG. .ACAACGGAA
Nmneccg7ctg1149 ATATATAAAG CCGAAAACAA CAGTCAGCCA GGCTACTCAA CACAAAACAA

                501                                                 550
Nmneccg7ctg1448 TAAGATCGAC AACGGAGACG TTGGCGGTGG GGACACGCAT GGCCATACCA
 Nmneccg7ctg273 ---------- ---------- ---------- ---------- ----------
 Nmneccg7ctg975 TAAGATCGAC AACGGAGACG TTGGCGGTGG GGACACCCAT GGCCATTCCA
  Nmneccg7ctg53 TAAAATCGAC AACGGAAACT TTGGCGGTGG GGACACGCAT GGCCATACCA
Nmneccg7ctg1394 TAAGATCGAC AACGGAGACG TTGGCGGTGG GGACACGCAT GGCCATACCA
Nmneccg7ctg1425 ATTATGAGCA TGACTATGAG GGAGCAGACG ATATG----- ----------
Nmneccg7ctg1149 AACAACACAC GAACAACCCA AACAAACAGC AAAAATGCAG TTCAAGACCC

                551                                                 600
Nmneccg7ctg1448 GTGAGCTTGC CGTTGAGGTC GGGGATGACC TTGCCGACGG CCTTGGCGGC
 Nmneccg7ctg273 ---------- ---------- ---------- ---------- ----------
 Nmneccg7ctg975 ATGAGCTTGC CCTTGAAGTC GGGGATGAAC TTTGCGACAG GCTT.GCGGG
  Nmneccg7ctg53 ATGAACTTGC CGTTAAGGTC GGGGATAACC TTGCCCACGG CCTT GGCGG
Nmneccg7ctg1394 GTGAGCTTGC CGTTGAGGTC GGGGATGACC TTGCCGACGG CCTTGGCGGC
Nmneccg7ctg1425 ---------- ---------- ---------- ---------- ----------
Nmneccg7ctg1149 TCATCGTCTC CACCCTCGCC GGCCTTGCCG GTGGCCAAAA CACCGTCCCC

                601
Nmneccg7ctg1448 ACCAGTGCTG CTGGGAATGA
 Nmneccg7ctg273 ---------- ----------
 Nmneccg7ctg975 ACCAAGGCTG CTGGGAATTG
  Nmneccg7ctg53 ACCATTGCTT CTGGGATTGA
Nmneccg7ctg1394 ACCAGTGCTG CTGGTGGTGC
Nmneccg7ctg1425 ---------- ----------
Nmneccg7ctg1149 TCTATCTACC GG--------
```

Figure 3.16 The DNA sequence comparisons among the 7 contigs of the *ccg-7*. The alternative polyadenylation sites were showed and the stop codon was marked with bold font ATT(TAA).

```
            201                                                 250
Nmneccg7ctg1448 AAQNIIPSST GAAKAVGKVI PDLNGKLTGM AMRVPTANVS VVDLTARIEK
Nmneccg7ctg1149 ---------- ---------- ---------- ---------- ----------
Nmneccg7ctg1425 ---------- ---------- ---------- ---------- ----------
Nmneccg7ctg1394 -------SST GAAKAVGKVI PDLNGKLTGM AMRVPTANVS VVDLTARIEK
 Nmneccg7ctg975 ---------- -SRKPVAKFI PDFKGKLIGM AMGVPTANVS VVDLTARIEK
 Nmneccg7ctg273 ---------- ---------- ---------- ---------- ---------K
  Nmneccg7ctg53 ---------- -SAKAVGKVI PDLNGKFIGM AMRVPTAKVS VVDFNARIKK

                251                                                 300
Nmneccg7ctg1448 GATYDEIKEV IKKASEGPLA GILAYTEDEV VSSDMNGNPA SSIFDAKAGI
Nmneccg7ctg1149 ---------- ---------- ---------V VSSDMNGNPA SSIFDAKAGI
Nmneccg7ctg1425 GATYDEIKEV IKKASEGPLA GILAYTEDEV VSSDMNGNPA SSIFDAKAGI
Nmneccg7ctg1394 GATYDEIKEV IKKASEGPLA GILAYTEDEV VSSDMNGNPA SSIFDAKAGI
 Nmneccg7ctg975 GATYDEIKEV IKKASEGPLA GILAYTEDEV VSSDMNGNPA SSIFDAKAGI
 Nmneccg7ctg273 GATSDEIKEV IKKAFEGPLG GILAYTEDEV VSSDMNGNPA SSILDAKAGI
  Nmneccg7ctg53 GATYDEIKEV IKKAF*GPLA GILAYTEDKV VSSDIERQP~ ----------
```

```
                301                                         338
Nmneccg7ctg1448  SLNKNFVKLV SWYDNEWGYS RRVLDLISYI SKVDAKKA
Nmneccg7ctg1149  SLNKNFVKLV SWYDNEWGYS RRVLDLISYI SKVDAKKA
Nmneccg7ctg1425  SLNKNFVKLV SWYDNEWGYS RRVLDLISYI SKVDAKKA
Nmneccg7ctg1394  SLNKNFVKLV SWYYNEWGYS RRVLDLISYI SKVDAKKA
 Nmneccg7ctg975  SLNKNFVKLV SWYDNEWGYS RRVLDLISYI SKVDAKKA
 Nmneccg7ctg273  SLNKNFVKLV SWYDNEWGYS RRVLDLISYI SKVDAKKA
  Nmneccg7ctg53  ---------- ---------- ---------- --------
```

Figure 3.17 The translated protein sequence comparisons among the 7 contigs of the *Neurospora crassa ccg-7* gene. The first part sequence of the contig1448 was omitted since the translated peptides of the other six contigs are not completed and only match with it's rare part sequence. The back part amino acid sequence of the contig53 was not showed in this figure. It matches with the *ccg-7* protein from the 285 aa to 338 aa of 96% similarity.

gene. Except for the contig1448, the other six contigs were not full length cDNA sequences as they lacked 5' end sequences of their mRNA. However, it is obvious that these seven contigs represent same gene, the *Neurospora crassa ccg-7* gene. Contig1425, contig1149 and contig53 have a same polyadenylation site but with different length of the poly (A) tail. Contig1448, contig1394, contig975 and contig273 had different polyadenylation start sites during their RNA processing. Therefore, taking into account the alternative polyadenylation, the *ccg-7* gene had five different gene products in the two *Neurospora crassa* cDNA libraries.

The *ccg-7* encodes *N.crassa* GAPDH (glyceraldehyde-3-phosphate dehydrogenase), the first energy-harvesting enzyme in the glycolytic pathway. Almost all previously isolated *ccg* gene transcripts of *N.crassa* peak in late night to early morning, a time corresponding to the time of day at which conidiation takes place. Therefore, all these *ccgs* were assumed to be directly associated with this asexual developmental pathway. This is different from the expression of "housekeeping genes" that are expected

to be expressed in a relatively time-invariant manner.

The transcript encoded by *ccg-7* is 1226 nucleotides long and contains a short 5'-untranslated region (43 nucleotide) followed by a single ORF of 338 codons. Its primary transcript has 2 introns (**Figure3.18**). The intron1 is 567 nucleotide long and the intron2 is 76 nucleotide long. This was confirmed by the sequence comparison between the *ccg-7* gene (U67457) and the contig1448 (Figure3.19). The cross_match result showed that except for the two introns on the sequence of the *ccg-7* gene, the sequence of the contig1448 matched perfectly with the region from 249 bp to 2100 bp of the *ccg-7* gene sequence (U67457). Among GAPDH genes of several filamentous ascomycetes, the presence and position of the shorter intron within their ORF is conserved (Jungehulsing et al., 1994). However, the longer intron located within the 5' UTR region was not observed in GAPDH genes of other fungal species as shown in **Figure 3.20,** a comparison of *ccg-7* translated protein with GAPDHs from other 7 species. These seven species are *Saccharomyces cerevisae* Tdh2p (X60157, 63% identity), *Arabidopsis thaliana* cytosolic GapC (M64116, 72% identity), *Mus musculus* (M32599, 71% identity), Human (P04406, 70% identity), *Cryphonectria parasitical* (X53996, 86% identity), *Aspergillus nidulans* (M19694, 79% identity) and *E. coli* (X02662, 67% identity). *Cryphonectria* GAPDH has the closest identity to the GAPDH of *Neurospora crassa* **(Figure 3.21).** A GAPDH motif is universally conserved (ASCTTNCL) at the position 155-162 and contains the catalytic site cys-157. From the sequence comparison and alignment of several GAPDHs, the catalytic residues His-184, Thr-187, Lys-191, NAD^+-binding residue Asp-37 and Phe-106 are found to be conserved in *ccg-7*

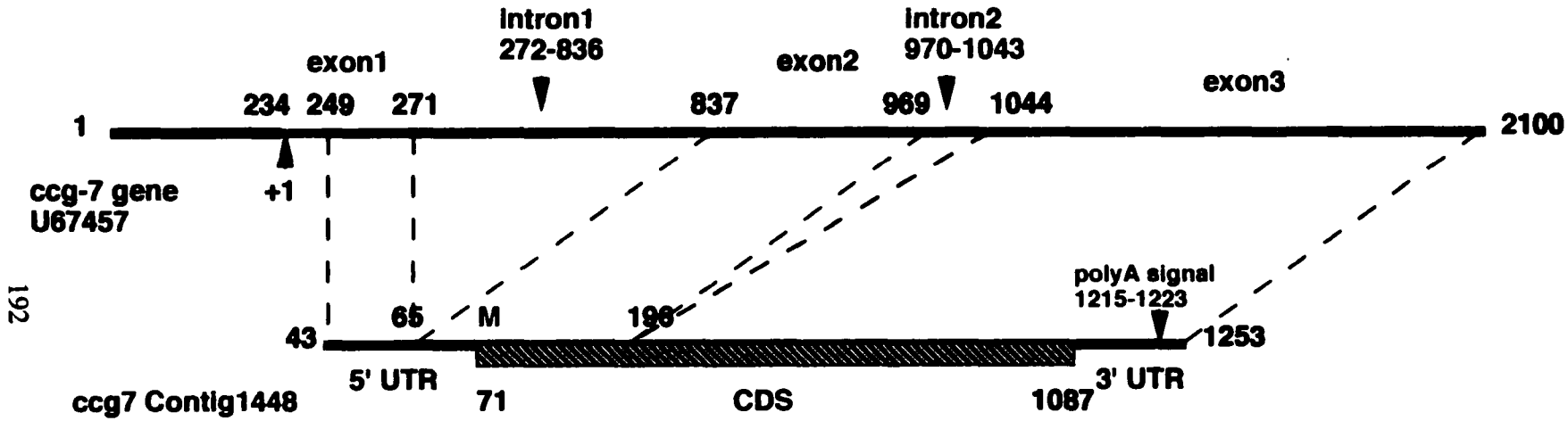


Figure 3.18 The annotation of *ccg-7* gene and the *ccg-7* cDNA represented by Contig1448

(Templeton et al., 1992). The conservation between these seven GAPDHs is high, including the important residues of GAPDH. Since GAPDH is a housekeeping gene, it plays an important role in glycolysis and a number of non-glycolytic pathways. These non-glycolytic activities include DNA repair (Meyer-Sieger et al, 1991), DNA, mRNA, and protein binding (Nagy and Rigby, 1995), protein kinase activity (Kawamoto and Caswell, 1986) and the interaction between microtubules and cell membranes (Allen et al 1987). Changes in GAPDH activity would affect many cellular activities (Shinohara et al., 1998).

Unlike other *ccg* genes in N.crassa, nitrogen and carbon starvations have no effect on the expression of *ccg-7* (Shinohara et al., 1998). Therefore, *ccg-7* is different from

```
cross_match -alignments ccg7Ctg1448.fa  ccg7.fa
Contig1448                              1    65 (1234)
gi|1532188|gb|U67457.1|NCU67457       207   271 (1829)

  Contig1448            1 CTTCATCATCATCCTCGCGATACCTTCATCTTTCCAACCAAAACCCTTCT 50

  gi|1532188|gb|U     207 CTTCATCATCATCCTCGCGATACCTTCATCTTTCCAACCAAAACCCTTCT 256

  Contig1448           51 TCCAAACCACATCAG 65                Intron 1

  gi|1532188|gb|U     257 TCCAAACCACATCAG 271

Contig1448                              64   197 (1102)
gi|1532188|gb|U67457.1|NCU67457       837   970 (1130)

  Contig1448           64 AGTCAACATGGTCGTCAAGGTCGGCATCAACGGTTTCGGCCGTATCGGTC 113

  gi|1532188|gb|U     837 AGTCAACATGGTCGTCAAGGTCGGCATCAACGGTTTCGGCCGTATCGGTC 886

  Contig1448          114 GCATTGTCTTCCGCAATGCCATTGAGCACGATGACATCCACATCGTCGCT 163

  gi|1532188|gb|U     887 GCATTGTCTTCCGCAATGCCATTGAGCACGATGACATCCACATCGTCGCT 936

  Contig1448          164 GTCAACGACCCCTTCATTGAGCCCAAGTACGCTG 197   Intron 2

  gi|1532188|gb|U     937 GTCAACGACCCCTTCATTGAGCCCAAGTACGCTG 970

Contig1448                             197  1253 (46)
gi|1532188|gb|U67457.1|NCU67457       1044  2100 (0)

  Contig1448          197 GCTTACATGCTCCGCTACGACACCACCCACGGCAACTTCAAGGGCACCAT 246

  gi|1532188|gb|U    1044 GCTTACATGCTCCGCTACGACACCACCCACGGCAACTTCAAGGGCACCAT 1093

  Contig1448          247 CGAGGTTGACGGTGCTGACCTCGTCGTCAACGGCAAGAAGGTCAAGTTCT 296

  gi|1532188|gb|U    1094 CGAGGTTGACGGTGCTGACCTCGTCGTCAACGGCAAGAAGGTCAAGTTCT 1143
```

```
Contig1448          297 ACACTGAGCGCGACCCCGCTGCCATCCCCTGGTCCGAGACCGGTGCCGAC 346
gi|1532188|gb|U    1144 ACACTGAGCGCGACCCCGCTGCCATCCCCTGGTCCGAGACCGGTGCCGAC 1193
Contig1448          347 TACATTGTCGAGTCCACTGGTGTCTTCACCACCACCGAGAAGGCCTCCGC 396
gi|1532188|gb|U    1194 TACATTGTCGAGTCCACTGGTGTCTTCACCACCACCGAGAAGGCCTCCGC 1243
Contig1448          397 CCACTTGAAGGGTGGTGCCAAGAAGGTCATCATCTCTGCCCCCTCTGCTG 446
gi|1532188|gb|U    1244 CCACTTGAAGGGTGGTGCCAAGAAGGTCATCATCTCTGCCCCCTCTGCTG 1293
Contig1448          447 ATGCCCCCATGTACGTTATGGGTGTCAACAACGAGACCTACGATGGCTCC 496
gi|1532188|gb|U    1294 ATGCCCCCATGTACGTTATGGGTGTCAACAACGAGACCTACGATGGCTCC 1343
Contig1448          497 GCCGACGTCATCTCCAACGCCTCTTGCACCACCAACTGCTTGGCTCCCCT 546
gi|1532188|gb|U    1344 GCCGACGTCATCTCCAACGCCTCTTGCACCACCAACTGCTTGGCTCCCCT 1393
Contig1448          547 CGCCAAGGTCATCCACGACAACTTCACCATCGTCGAGGGTCTCATGACCA 596
gi|1532188|gb|U    1394 CGCCAAGGTCATCCACGACAACTTCACCATCGTCGAGGGTCTCATGACCA 1443
Contig1448          597 CCGTCCACTCCTACACCGCCACCCAGAAGACCGTCGATGGTCCTTCCGCC 646
gi|1532188|gb|U    1444 CCGTCCACTCCTACACCGCCACCCAGAAGACCGTCGATGGTCCTTCCGCC 1493
Contig1448          647 AAGGACTGGCGCGGTGGCCGCACTGCTGCTCAGAACATCATTCCCAGCAG 696
gi|1532188|gb|U    1494 AAGGACTGGCGCGGTGGCCGCACTGCTGCTCAGAACATCATTCCCAGCAG 1543
Contig1448          697 CACTGGTGCCGCCAAGGCCGTCGGCAAGGTCATCCCCGACCTCAACGGCA 746
                              v
gi|1532188|gb|U    1544 CACTGGTTCCGCCAAGGCCGTCGGCAAGGTCATCCCCGACCTCAACGGCA 1593
Contig1448          747 AGCTCACTGGTATGGCCATGCGTGTCCCCACCGCCAACGTCTCCGTTGTC 796
gi|1532188|gb|U    1594 AGCTCACTGGTATGGCCATGCGTGTCCCCACCGCCAACGTCTCCGTTGTC 1643
Contig1448          797 GATCTTACTGCCCGCATCGAGAAGGGTGCTACCTACGATGAGATCAAGGA 846
gi|1532188|gb|U    1644 GATCTTACTGCCCGCATCGAGAAGGGTGCTACCTACGATGAGATCAAGGA 1693
Contig1448          847 GGTCATCAAGAAGGCCTCTGAGGGTCCCCTCGCTGGCATCCTTGCCTACA 896
gi|1532188|gb|U    1694 GGTCATCAAGAAGGCCTCTGAGGGTCCCCTCGCTGGCATCCTTGCCTACA 1743
Contig1448          897 CCGAGGATGAGGTTGTCTCTTCCGACATGAACGGCAACCCCGCCTCCTCC 946
gi|1532188|gb|U    1744 CCGAGGATGAGGTTGTCTCTTCCGACATGAACGGCAACCCCGCCTCCTCC 1793
Contig1448          947 ATCTTCGATGCCAAGGCTGGTATCTCCCTCAACAAGAACTTCGTCAAGCT 996
gi|1532188|gb|U    1794 ATCTTCGATGCCAAGGCTGGTATCTCCCTCAACAAGAACTTCGTCAAGCT 1843
Contig1448          997 TGTCTCCTGGTACGACAACGAGTGGGGCTACTCTCGCCGTGTCCTCGACC 1046
gi|1532188|gb|U    1844 TGTCTCCTGGTACGACAACGAGTGGGGCTACTCTCGCCGTGTCCTCGACC 1893
Contig1448         1047 TCATCTCCTACATCTCCAAGGTCGATGCCAAGAAGGCTTAAATCGGTTGC 1096
gi|1532188|gb|U    1894 TCATCTCCTACATCTCCAAGGTCGATGCCAAGAAGGCTTAAATCGGTTGC 1943
Contig1448         1097 GTACCCGCACGGTTATGAAGTAATGGTCTTTTCCTAGATATGAAGAAAAA 1146
gi|1532188|gb|U    1944 GTACCCGCACGGTTATGAAGTAATGGTCTTTTCCTAGATATGAAGAAAAA 1993
```

```
Contig1448        1147 AAAAGGGCAATGATTCCGTGGGATTGAACTCGAGCATGTTGGATCTCGGG 1196

gi|1532188|gb|U   1994 AAAAGGGCAATGATTCCGTGGGATTGAACTCGAGCATGTTGGATCTCGGG 2043

Contig1448        1197 CAGTCCTGCTTAAAGTAAAATAATATCCGAACTCAAATAGATACCAAGTT 1246

gi|1532188|gb|U   2044 CAGTCCTGCTTAAAGTAAAATAATATCCGAACTCAAATAGATACCAAGTT 2093

Contig1448        1247 CACTTCC 1253

gi|1532188|gb|U   2094 CACTTCC 2100
```

Figure 3.19 The sequence comparison between the contig1448 and the *ccg-7* gene genomic DNA. Two introns were marked. The first intron was not found in other fungi.

```
                    1                                                  50
       Scgdphaa  ----MVRVAI NGFGRIGRLV MRIALQRKNV EVVALNDPFI SNDYSAYMFK
      Athgapdaa  MADKKIRIGI NGFGRIGRLV ARVVLQRDDV ELVAVNDPFI TTEYMTYMFK
       Musgapdh  ----MVKVGV NGFGRIGRLV TRAAICSGKV EIVAINDPFI DLNYMVYMFQ
       Humgapdh  --MGKVKVGV NGFGRIGRLV TRAAFNSGKV DIVAINDPFI DLNYMVYMFQ
Nmneccg7ctg1448  ---MVVKVGI NGFGRIGRIV FRNAIEHDDI HIVAVNDPFI EPKYAAYMLR
       Crygapdh  ---MVVKVGI NGFGRIGRIV FRNAHEHSDV EIVAVNDPFI EPHYAAYMLK
       Aspgapdh  ---MAPKVGI NGFGRIGRIV FRNAIEAGTV DVVAVNDPFI ETHYAAYMLK
      Ecogapdaa  ---MTIKVGI NGFGRIGRIV FRAAQKRSDI EIVAIND.LL DADYMAYMLK

                   51                                                 100
       Scgdphaa  YDSTHGRYA. GEVSHDDKHI IVDGHK.IAT FQERDPANLP WASLNIDIAI
      Athgapdaa  YDSVHGQWKH NELKIKDEKT LLFGEKPVTV FGIRNPEDIP WAEAGADYVV
       Musgapdh  YDSTHGKF.N GTVKAENGKL VING.KPITI FQERDPTNIK WGEAGAEYVV
       Humgapdh  YDSTHGKF.H GTVKAENGKL VING.NPITI FQERDPSKIK WGDAGAEYVV
Nmneccg7ctg1448  YDTTHGNF.K GTIEVDGADL VVNG.KKVKF YTERDPAAIP WSETGADYIV
       Crygapdh  YDSQHGNF.K GDVTVEGSDL VVGG.KKVRF YTERDPAAIP WSETGADYIV
       Aspgapdh  YDSQHGQF.K GTIETYDEGL IVNG.KKIRF HTERDPANIP WGQDGAEYIV
      Ecogapdaa  YDSTHGRF.D GTVEVKDGHL IVNG.KKIRV TAERDPANLK WDEVGVDVVA

                  101                                                 150
       Scgdphaa  DSTGVFKELD TAQKHIDAGA KKVVITAPSS T.APMFVMGV NEEKYTSDLK
      Athgapdaa  ESTGVFTDKD KAAAHLKGGA KKVVISEPSK D.APMFVVGV NEHEYKSDLD
       Musgapdh  ESTGVFTTME KAGAHLKGGA KRVIISAPSA D.APMFVMGV NHEKYDNSLK
       Humgapdh  ESTGVFTTME KAGAHLQGGA KRVIISAPSA D.APMFVMGV NHEKYDNSLK
Nmneccg7ctg1448  ESTGVFTTTE KASAHLKGGA KKVIISAPSA D.APMYVMGV NNETYDGSAD
       Crygapdh  ESTGVFTTTE KAKAHLKGGA KKVIISAPSA D.APMYVMGV NEKTYDGSGM
       Aspgapdh  ESTGVFTTQE KASAHLKGGA KKVVISAPSA D.APMFVMGV NNETYKKDIQ
      Ecogapdaa  EATGLFLTDE TARKHITAGA KKVVMTGPSK DNTPMFVKGA NFDKYAGQ.D

                  151                                                 200
       Scgdphaa  IVSNASCTTN CLAPLAKVIN DAFGIEEGLM TTVHSMTATQ KTVDGPSHKD
      Athgapdaa  IVSNASCTTN CLAPLAKVIN DRFGIVEGLM TTVHSITATQ KTVDGPSMKD
       Musgapdh  IVSNASCTTN CLAPLAKVIH DNFGIVEGLM TTVHAITATQ KTVDGPSGKL
       Humgapdh  IISNASCTTN CLAPLAKVIH DNFGIVEGLM TTVHAITATQ KTVDGPSGKL
Nmneccg7ctg1448  VISNASCTTN CLAPLAKVIH DNFTIVEGLM TTVHSYTATQ KTVDGPSAKD
       Crygapdh  VISNASCTTN CLAPLAKVIN DEFKIIEGLM TTVHSYTATQ KTVDGPSAKD
       Aspgapdh  VLSNASCTTN CLAPLAKVIN DNFGIIEGLM TTVHSYTATQ KVVDGPSAKD
      Ecogapdaa  IVSNASCTTN CLAPLAKVIN DNFGIIEGLM TTVHATTATQ KTVDGPSHKD

                  201                                                 250
       Scgdphaa  WRGGRTASGN IIPSSTGAAK AVGKVLPELQ GKLTGMAFRV PTVDVSVVDL
      Athgapdaa  WRGGRAASFN IIPSSTGAAK AVGKVLPALN GKLTGMSFRV PTVDVSVVDL
       Musgapdh  WRDGRGAAQN IIPASTGAAK AVGKVIPELN GKLTGMAFRV PTPNVSVVDL
```

```
       Humgapdh  WRDGRGALQN IIPASTGAAK AVGKVIPELN GKLTGMAFRV PTANVSVVDL
Nmneccg7ctg1448  WRGGRTAAQN IIPSSTGAAK AVGKVIPDLN GKLTGMAMRV PTANVSVVDL
       Crygapdh  WRGGRTAAQN IIPSSTGAAK AVGKVIPELN GKLTGMSMRV PTSNVSVVDL
       Aspgapdh  WRGGRTAATN IIPSSTGAAK AVGKVIPSLN GKLTGMAMRV PTSNVSVVDL
      Ecogapdaa  WRGGRGASQN IIPSSTGAAK AVGKVLPELN GKLTGMAFRV PTPNVSVVDL

                 251                                                 300
       Scgdphaa  TVKLNKETTY DEIKKVVKAA AEGKLKGVLG YTEDAVVSSD FLGDSNSSIF
      Athgapdaa  TVRLEKAATY EEIKKAIKEE SEGKLKGILG YTEDDVVSTD FVGDNRSSIF
       Musgapdh  TCRLEKPAKY DDIKKVVKQA SEGPLKGILG YTEDQVVSCD FNSNSHSSTF
       Humgapdh  TCRLEKPAKY DDIKKVVKQA SEGPLKGILG YTEHQVVSSD FNSDTHSSTF
Nmneccg7ctg1448  TARIEKGATY DEIKEVIKKA SEGPLAGILA YTEDEVVSSD MNGNPASSIF
       Crygapdh  TVRIEKGATY EQIKTAVKKA ADGPLKGVLA YTEDDVVSTD MNGNPNSSIF
       Aspgapdh  TVRTEKAVTY DQIKDAVKKA SENELKGILG YTEDDIVSTD LNGDTRSSIF
      Ecogapdaa  TVRLEKAATY EQIKAAVKAA AEGEMKGVLG YTEDDVVSTD FNGEVCTSVF

                 301                                                 344
       Scgdphaa  DAAAGIQLSP KFVKLVSWYD NEYGYSTRVV DLVEHVAKA- ----
      Athgapdaa  DAKAGIALSD KFVKLVSWYD NEWGYSSRVV DLIVHMSKA- ----
       Musgapdh  DAGAGIALND NFVKLISWYD NEYGYSNRVV DLMAYMASKE ----
       Humgapdh  DAGAGIALND HFVKLISWYD NEFGYSNRVV DLMAHMASKE ----
Nmneccg7ctg1448  DAKAGISLNK NFVKLVSWYD NEWGYSRRVL DLISYISKVD AKKA
       Crygapdh  DAKAGISLND HFVKLVSWYD NEWGYSRRVL DLISHVAKVD GNA-
       Aspgapdh  DAKAGIALNS NFIKLVSWYD NEWGYSRRVV DLITYISKVD AQ--
      Ecogapdaa  DAKAGIALND NFVKLVSWYD NETGYSNKVL DLIAHISK-- ----
```

Figure 3.20 The multiple sequence comparison of *Neurospora* GAPDH with another 8 GAPDHs from different organisms. The GAPDH-specific catalytic motif is boxed.

other *ccg* genes that are involved in organism-specific behaviors. These *ccg* genes are, for example, the *CAB2* gene in *Arabidopsis* involved in photosynthesis, the *con-10* gene of *Neurospora crass* involved in the development, and the *ccg-4* gene or pheromone gene involved in reproduction. According to the literature, *ccg-7* mRNA appears to be modulated solely by the biological clock (Shinohara et al., 1998). However, in the promoter region of *ccg-7* gene, there are 4 putative GATA boxes. The GATA box is conserved in several plant *CAB2* genes and also found in several *Neurospora ccg* genes (section3.2.2.2). In *Arabidopsis CAB2 gene*, the GATA box is inside the CGF-1 binding site, which maybe the regulation site for the circadian clock and/or light signal transduction pathway (Anderson and key, 1995). This may imply that the *ccg-7* may be

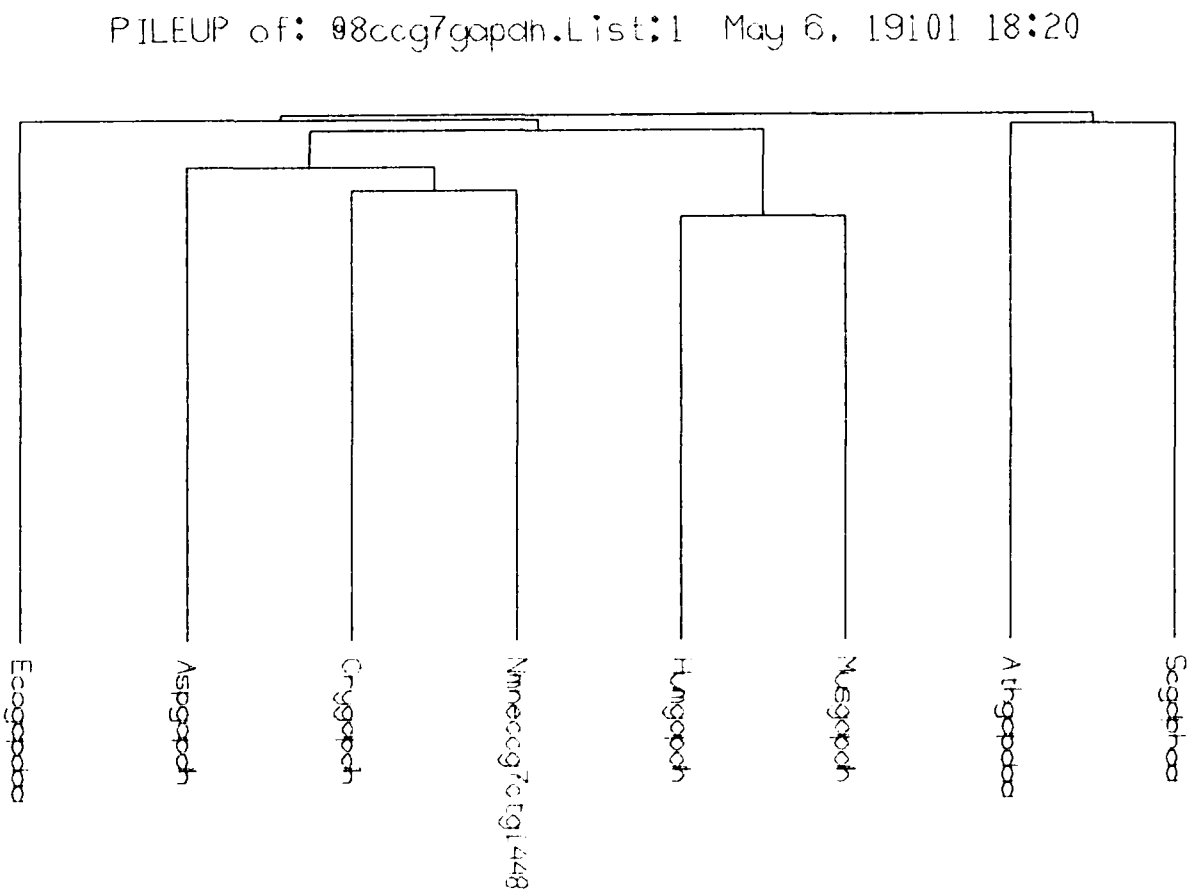


Figure 3.21 The comparisons of *N. crassa* GDPAH with other 7 GDPAHS from different organisms. *N. crassa* GDPAH was represented by NMNEctg1448.

regulated by light too.

Since GAPDH is a key enzyme involved in the glycolytic pathway, the regulation of GAPDH by the circadian clock will influence the organism's fundamental metabolic activities. Thus, both clock regulation of transcription and expression and expression of *ccg-7* in *N.crassa* support the idea that the circadian clock has much more common influences in the cell at the cellular level than previously expected.

In summary, the conceptual amino acid sequence of *ccg-7* is highly similar to GAPDH peptides from diverse organisms. Compared to other *ccg* genes, only *ccg-7* has the potential to be ubiquitous. Since all GAPDH sequences are highly conserved, all modern GAPDHs likely arose from a common ancestral enzyme (Martinez et al., 1989).

3.2.4 The clock-controlled gene 4

The clock-controlled gene 4 is one of the two special *ccg* genes that have unusual representation in the two cDNA libraries of *Neurospora crassa.* Generally, most *ccgs* have higher redundancy in the morning library than in the evening cDNA library. But the *ccg-4* is an exception.

From **table 3.31**, the EST totals of the *ccg-4* in the morning library is zero while in the evening library there are 42 *ccg-4* ESTs. This is different from all other *Neurospora crassa ccgs*. The ratio of the EST total of the other identified *ccgs* in the NM and NE libraries are greater than 1.0.

Table 3.31 The contigs and the EST totals of *ccg-4* gene in combined EST database

NE ctg#	EST totals	NM ctg#	EST totals	NMNE ctg#	EST totals

Contig512 Contig872 Contig1042	4 9 22			Contig605 Contig1002 Contig1262	4 9 22
Total	35 35x1.2=42		0		35 NM/NE=0

Although the function of the *ccg-4* gene product is not known in *Neurospora crassa*, it was found from the screen against the time-of-day-specific *Neurospora* cDNA library used to detect the other 4 *ccgs (ccg-7, ccg-8, ccg-9, ccg-12)* (Bell-Pedersen et al., 1996c). The genomic DNA sequence and the cDNA sequence of *ccg-4* gene are compared in **Figure 3.22** and the crossmatch results show that the cDNA sequence matches the genomic DNA sequence of the *ccg-4* gene 100%.

3.2.4.1 Two putative proteins were reported previously for *ccg-4* gene

Two putative polypeptides were reported for *ccg-4*. Contig 872 and contig512 match with polypeptide 1 from amino acid 22 to amino acid 84 with 98% and 100% identities. Contig1042 matches with polypeptide 2 from amino acid 1 to amino acid 87 with 98% identity (Table 3.32).

Table 3.32 BlastX of the ctgs representing *ccg-4* gene in the NE EST database

Ctg#	EST#	cDNA clones	Ctg length	Alignment region	frame	Blastx hits
512	4	2	463	50-238	-1	Gi|1184782(U46085) *ccg-4* putative protein 1, 105 aa
872	9	9	1002	270-482	2	Gi|1184782 (U46085) *ccg-4* putative protein 1, 105 aa
1042	22	21	1334	731-991	2	Gi|1184781 (U46085) *ccg-4* protein2, 87 aa Gi|1184782(U46085), *ccg-4* protein1, 105aa Gi2290408(U92042), *Cryphonectria parasitical* mating type 1 pheromone precursor

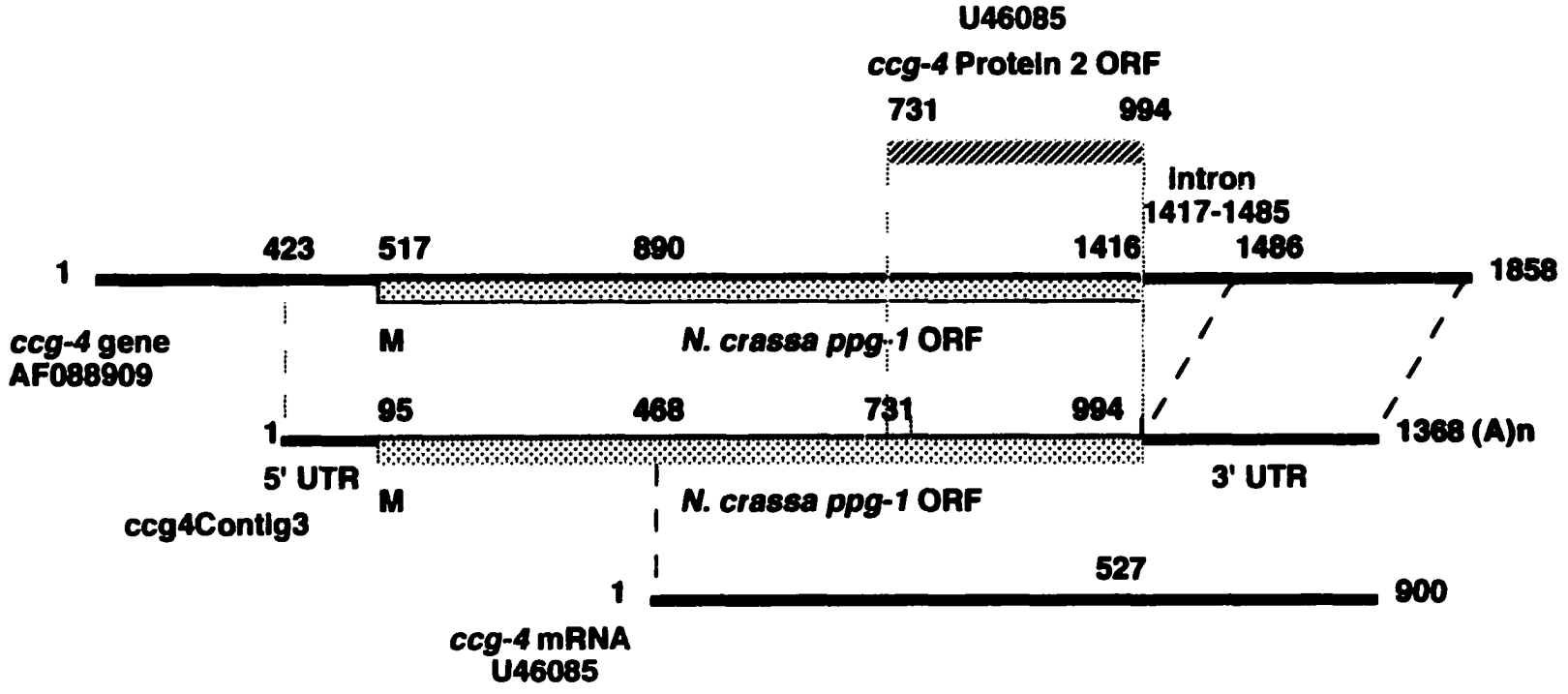


Figure 3.22 The annotation of *ccg-4* gene, *ccg-4* mRNA and the ccg4contig3

The BlastX result revealed that contig1042 has 3 homologous sequences in Genbank: *ccg-4* protein2, *ccg-4* protein 1 and *Cryphonectria parasitical* mating type 1 pheromone. The alignments between NE.Contig1042 and the three Genbank homologues are listed below with contig1042 being the query sequence.

```
Query=  Contig1042
        (1334 letters)

>gi|1184781 (U46085) ccg-4 putative polypeptide 2;  Method: conceptual
            translation supplied by author. [Neurospora crassa]

            Length = 87
Query: 731  MTAIQSAEAESALLLRDTTFSPVDRVGKRDPQVCNMRLHPKKVCWKRDASPEAACNAPDG 910
            MTAIQSAEAESALLLRDTTFSPVDRVGKRDPQVCNMRLHPKKVCWKRDASPEAACNAPDG
Sbjct:  1    MTAIQSAETESALLLRDTTFSPVDRVGKRDPQVCNMRLHPKKVCWKRDASPEAACNAPDG 60

Query:   911 SCTKATRDLHAMYNVARAILTAHSDEN 991
             SCTKATRDLHAMYNVARAILTAHSDEN
Sbjct:    61 SCTKATRDLHAMYNVARAILTAHSDEN 87

>gi|1184782 (U46085) ccg-4 putative polypeptide 1;  Method: conceptual
translation supplied by author. [Neurospora crassa] >gi|3746901 (AF088909) clock-
controlled gene-4 protein [Neurospora crassa] Length = 105

Query: 930 VTCTPCTTWLVPSSLLTPMRTRCFPEIFFPRRGYDKANQKGAEKRLGYITFAPSLHTSLL 1109
           VTCTPCTTWLVPSSLLTPMRTRCFPEIFFPRRGYDKANQKGAEKRLGYITFAPSLHTSLL
Sbjct: 20  VTCTPCTTWLVPSSLLTPMRTRCFPEIFFPRRGYDKANQKGAEKRLGYITFAPSLHTSLL 79

Query:  1110 SGKKK 1124
             SGKKK
Sbjct:    80 SGKKK 84

Query:  1027 GTTKQTKRGQRKGLDISXXXXXXXXXXXXEKKKYNTFTYTTEAFWVHTQHKGIG1188
             G  K  ++G  K L                KKK NT TYTTEAFWVHTQHKGIG
Sbjct:    52 GYDKANQKGAEKRLGYITFAPSLHTSLLSGKKKCNTSTYTTEAFWVHTQHKGIG 105

>gi|2290408 (U92042) mating type 1 pheromone precursor [Cryphonectria
            parasitica]
            Length = 530

Query:  158 PNAEAEAQWCRIHGQSCWKVKRVADAFANAIQGMGGLPP-RD-ESGHQ---PAQVAKRQV 322
            P A+ EA WC  HG+ CWKVKR   AFANAI+G  G+P  R  E  +     A  AKR V
Sbjct:  141 PEADPEA-WCLFHGEGCWKVKRAVYAFANAIRGAAGIPESRSAEISNMRGGAAYNAKRAV 199

Query: 323 DELAGIIALTQEDVNAYYDSLSLQEKFAPSTXXXXXXXXXXXXXXXXXXQWCRIHGQSCW 502
            ++A ++A    +   +   L + E F P                      S
Sbjct: 200 QDIATMMAGRTSEPPEFLKQLFILEHFGPDANITAFGPPPTDADAAPIST----DDDSTT 255

Query: 503 KKREAEAQ-WCRIHGQSCWKRD-----------ALPEAEPQ---WCRIHGQSCWKKRDXX 637
            KR+A+A+ WC  HG+ CWKR            A  +A P    +C   G S
Sbjct: 256 TKRDADAEPWCLFHGEGCWKRSEEAGSSAATSLATRDASPAAAAFCPFEGSSTC-----Y 310

Query: 638 XXXXXXXXXNPQWCRIHGQSCWKAKRAAEAVMTAIQSAEAESALLLRDTTFSPVDRVGKR 817
                  + + C   G++C+ A+ AAEA++T I S             ++P  R
Sbjct: 311 ASKRDFAAADKRACNQPGEACYVARCAAEAIVTEIAS------------WAPAKR--SA 355
```

The *ccg-4* protein2 (1-87 amino acid residues) is 100% identical to that of the

translated product of NE.Contig1042 between 731 bp and 991 bp. However, a frame shift exists in the reported translation of the *ccg-4* protein 1 gene as 65 amino acid residues of the *ccg-4* protein 1 were translated from positions 930 to 1124 of contig1042 in frame +3. The identity of these 65 amino acid residues is 100%. In contrast, the amino acid residues from 52 to 105 of *ccg-4* protein1 were translated from position 1027 to 1188 of the same contig1042 in frame +1 and their identity was only about 50% (27/54). In addition, the DNA sequences for *ccg-4* protein 1 and protein 2 partially overlaps indicating that the *ccg-4* cDNA was fully sequenced.

3.2.4.2 Gap closure of the cDNA of the *ccg-4* gene

From the Consed window (**Figure 3.23**), it can be seen that NE.contig512 is composed of 2 pairs of ESTs from 2 different cDNA clones. Both ends of this contig begin at the cloning sites. Therefore, this cDNA is a shorter form of *ccg-4* gene transcript as it lacks a complete 5' end. Since NE.contig872 was produced from the 3' end of the ESTs from 9 cDNAs, 6 of which had clone pairs in NE.contig1042. Clearly there was a gap between the NE.contig872 and the NE.contig1042 that was spanned by the six cDNA clones a5g08ne, a7b07ne, a8e05ne, b7c02ne, b8g04ne, and c6h05ne, which subsequently were directly sequenced by primer walking. The primers were selected from the 5' direction of NE.Contig1042 and the 5' direction of the NE.Contig872 with the primer-picking program associated with Consed. The sequences of these primers were as follows.

*Ccg*4ctg1042F(419-438): GAGGAGAAGAAGA*CCG*AGAA

*Ccg*4ctg872F(388-407): TGAAGGTG GTG TCACGGAGA

After the six clones were used as templates for primer walking, the primer walking sequences were copied into the database, and all the ESTs from the three contigs plus the primer walking reads from the *ccg-4* gene also were copied into a separate directory. After running phredPhrap98, three new contigs formed in this *ccg4* directory (**Table 3.33**). Four ESTs from two cDNA clones of the old NE.Contig512 assembled into *ccg*4contig2, while the nine ESTs of the old NE.Contig872 were assembled into a

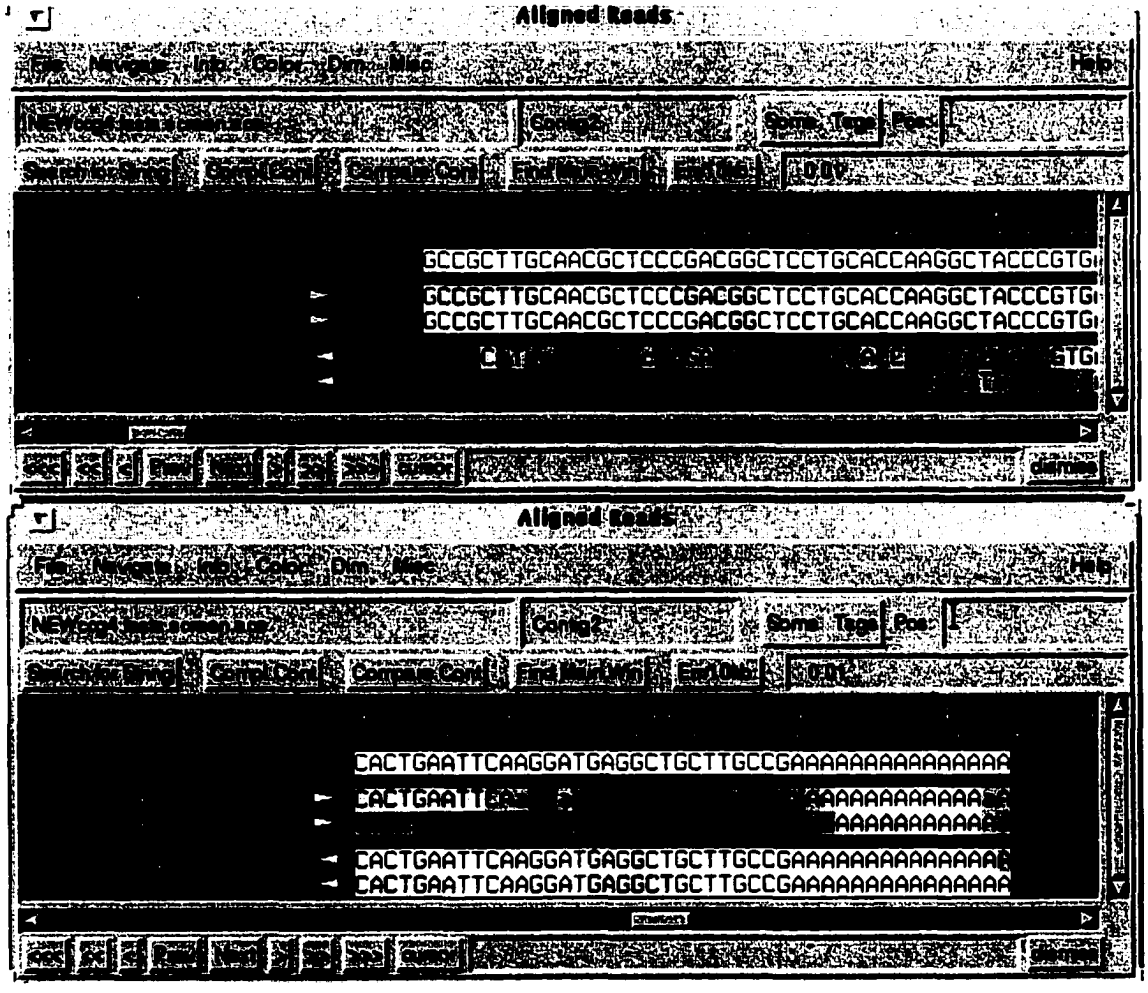


Figure3.23 Two cDNAs of the *ccg-4* gene represented by newccg4contig2

separate contig. However, three of the 3' ESTs of the original NE.Contig1042 were not assembled into ccg4contig3, but rather made their own contig, ccg4contig1.

Table 3.33 Three new *ccg4*contigs representing the *ccg-4* gene

NE.Ctg#	EST#	*ccg4*Ctg#	EST#
512	4	*ccg4*contig2	4 (exactly same as NE.contig512)
872	9	*ccg4*contig3	9+19 (9 from ctg872, 19 from NE.ctg1042)+5 pwk product.
1042	22	*ccg4*contig1	3 (from NE.ctg1042)

Sequence comparison of these three ccg4contigs was performed (Figure 3.24). The result revealed that except for the region of the poly (A) tail, the sequences of these three ccg4contigs were 100% identical. Ccg4contig1 aligned with ccg4contig3 from 864 bp to 1334 bp, with the sequences from 1298 bp to 1334 bp of this contig containing the polyA tail. Ccg4contig2 aligned with ccg4contig3 from 887 bp to 1350 bp, with region from 1335 bp to 1350 bp containing the polyA tail of ccg4contig2 and the region from 1369 to 1393 of ccg4contig3 contained its polyA tail. Therefore, the phredPhrap assembly produced 3 alternatively 3' polyadenylated mRNAs for the *ccg-4* gene. This is the first observation that *N.crassa* produces mRNAs with alternatively 3' polyadenylation positions.

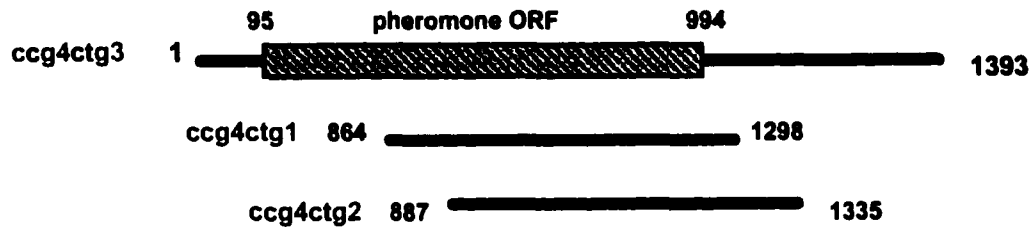


Figure 3.24 The positions of three new ccg4contigs and the ORF for *Neurospora* pheromone

3.2.4.3 The identity of the *ccg-4* gene is a *Neurospora* pheromone precursor homologue of the *ppg-1* gene from *Sordaria macrospora*

The BlastX search revealed that the sequence of ccg4contig3 at position 95-994 was highly homologous to that of the *ppg-1* gene of *Sordaria macrospora*. In an attempt to determine the relationship between the *ccg-4* protein1, *ccg-4* protein 2, *Sordaria* pheromone precursor 1 protein and the translated protein product of ccg4contig3, they were aligned at the protein sequence level using the GCG Pileup program. *ccg-4* protein 2 is a part of the ccg4contig3 translated product and *ccg-4* protein1 has no relationship to ccg4contig3 translated product since it is translated in a different frame. In contrast, the Sordaria pheromone precursor is highly homologous to the translated product of ccg4contig3. In addition, when the translated peptide sequences of ccg4contig3 was used as a query in a BlastP search against the nr protein database of Genbank, it also had 3 significant GenBank homologies, to the following; *Sordaria macrospora* pheromone precursor (CAB96172), the *ccg-4* protein 2(T47209), and the *ccg-4* protein1 (T47210).

As shown in Figure3.25, both the *ccg-4* protein 1 and the *ccg-4* protein 2 can be translated from the ccg4contig3. However, it appears that *ccg-4* protein 2 is a partial product of the ccg4contig3 translation product, and the *ccg-4* protein1 was incorrectly

```
            1                                                    50
Ccg4protin2 ---------- ----------
    Ctg3cut ---------- ---------- ---------- ---------- ----------
   Phrompre ---------- ---------- ---------- ---------- ----------
Ccg4protin1 MPPLRPLATL PTAPAPRLPV TCTPCTTWLV PSSLLTPMRT RCFPEIFFPR

            51                                                   100
Ccg4protin2 ---------- ---------- ---------- ---------- ----------
    Ctg3cut ---------- ---------- ---------- ---------- ----------
   Phrompre ---------- ---------- ---------- ---------- ----------
Ccg4protin1 RGYDKANQKG AEKRLGYITF APSLHTSLLS GKKKCNTSTY TTEAFWVHTQ

            101                                                  150
Ccg4protin2 ---------- ---------- ---------- ---------- ----------
    Ctg3cut ---MKFTLPL VIFAAVASAT PVAQPNAEAE AQWCRIHGQS CWKVKRVADA
   Phrompre ---MKFTLPL VIFAAVASAT PVAQPIAEAE AQWCRIHGQS CWKVKRVAEA
Ccg4protin1 HKGIG----- ---------- ---------- ---------- ----------
```

```
            151                                                   200
Ccg4protin2 ---------- ---------- ---------- ---------- ----------
    Ctg3cut FANAIQGMGG LPPRDESGHQ PAQVAKRQVD ELAGIIALTQ EDVNAYYDSL
   Phrompre FSTAIQGMGG LPTSDESGHL PAQVAKRQVD ELAGIIALTQ EDVNAYYDSL
Ccg4protin1 ---------- ---------- ---------- ---------- ----------

            201                                                   250
Ccg4protin2 ---------- ---------- ---------- ---------- ----------
    Ctg3cut SLQEKFAPST EEEKKTEKVA KREAEAEAQW CRIHGQSCWK KREAEAQWCR
   Phrompre NLQDKFAPST EEEKKDEKVA KRDAEAEAQW CRIHGQSCWK K.........
Ccg4protin1 ---------- ---------- ---------- ---------- ----------

            251                                                   300
Ccg4protin2 ---------- ---------- ---------- ---------- ----------
    Ctg3cut IHGQSCWKRD ALPEAEPQWC RIHGQSCWKK RDAAPEAAPE AEANPQWCRI
   Phrompre .......... AKREAEAQWC RIHGQSCWKK RDAAPEAAP. .EANPQWCRI
Ccg4protin1 ---------- ---------- ---------- ---------- ----------

            301                                                   350
Ccg4protin2 ---------- -----MTAIQ SAETESALLL RDTTFSPVDR VGKRDPQVCN
    Ctg3cut HGQSCWKAKR AAEAVMTAIQ SAEAESALLL RDTTFSPVDR VGKRDPQVCN
   Phrompre HGQSCWKAKR AAEAVMTAIQ SAEAESALLL RDTTFSPVDR VGKREADPQW
Ccg4protin1 ---------- ---------- ---------- ---------- ----------

            351                                                   400
Ccg4protin2 MRLHPKKVCW KRDASPEAAC NAPDGSCTKA TRDLHAMYNV ARAILTAHSD
    Ctg3cut MRLHPKKVCW KRDASPEAAC NAPDGSCTKA TRDLHAMYNV ARAILTAHSD
   Phrompre CRIHGQS.CW KRYASPEAAC NAPDGSCTKA TRDLHAMYNV ARAILTAHSD
Ccg4protin1 ---------- ---------- ---------- ---------- ----------

            401
Ccg4protin2 EN
    Ctg3cut EN
   Phrompre EN
Ccg4protin1 --
```

Figure3.25 The comparison of *ccg4ctg3* translation product with *ccg4* protein1, *ccg4* protein2, and *Sordaria macrospora* pheromone precursor.

translated from ccg4contig3 in another frame by Bell-Pedersen, et al. A personal communication with Dr. Bell-Pedersen who submitted these two putative *ccg-4* proteins, concluded with the belief that their translation of the *ccg-4* protein 1 is a mistake and the correct translation is given by the *ccg-4* protein 2. This was confirmed by the observation that the *Neurospora crassa* homolog of the *Sordaria* pheromone precursor (*ppg-1* gene product) is correctly translated in the EST of the *ccg-4* gene represented by ccg4contig3.

The Bestfit result of the compare between the *Neurospora crass* pheromone homolog and the *Sordaria* pheromone precursor 1 protein revealed that these two peptide sequences are 90.6% identical and 92.4% similar (**Figure 3.26**). On the DNA sequence

level, their sequences have 87.1% similarity and 87.1% identity.

```
Sordaria Pher  1 MKFTLPLVIFAAVASATPVAQPIAEAEAQWCRIHGQSCWKVKRVAEAFST 50
                 |||||||||||||||||||||| ||||||||||||||||||||||:||.
ccg4ctg3pep   32 MKFTLPLVIFAAVASATPVAQPNAEAEAQWCRIHGQSCWKVKRVADAFAN 81
                   .         .         .         .         .
               51 AIQGMGGLPTSDESGHLPAQVAKRQVDELAGIIALTQEDVNAYYDSLNLQ 100
                  |||||||||  ||||| |||||||||||||||||||||||||||||.||
               82 AIQGMGGLPPRDESGHQPAQVAKRQVDELAGIIALTQEDVNAYYDSLSLQ 131
                  .         .         .         .         .
              101 DKFAPSTEEEKKDEKVAKRDAEAEAQWCRIHGQSCWKK..........  138
                  :|||||||||||| ||||||:||||||||||||||||
              132 EKFAPSTEEEKKTEKVAKREAEAEAQWCRIHGQSCWKKREAEAQWCRIHG 181
                  .         .         .         .         .
              139 .......AKREAEAQWCRIHGQSCWKKRDAAPEAAP..EANPQWCRIHGQ 179
                        |  ||| ||||||||||||||||||||  ||||||||||||
              182 QSCWKRDALPEAEPQWCRIHGQSCWKKRDAAPEAAPEAEANPQWCRIHGQ 231
                  .         .         .         .         .
              180 SCWKAKRAAEAVMTAIQSAEAESALLLRDTTFSPVDRVGKREADPQWCRI 229
                  ||||||||||||||||||||||||||||||||||||||||:      |:
              232 SCWKAKRAAEAVMTAIQSAEAESALLLRDTTFSPVDRVGKRDPQVCNMRL 281
                  .         .         .         .
              230 HGQS.CWKRYASPEAACNAPDGSCTKATRDLHAMYNVARAILTAHSDEN 277
                  | .  |||| ||||||||||||||||||||||||||||||||||||||
              282 HPKKVCWKRDASPEAACNAPDGSCTKATRDLHAMYNVARAILTAHSDEN 330
```

Figure 3.26 The comparison of *Neurospora* pheromone precursor homolog with *Sordoria* pheromone precursor 1 protein. The 16-residue repeats and 12-residue repeats are marked with color and bold font.

From the Bestfit comparison, a 16 amino acid residue region EAEAQWCRIHGQSCWK was found to be repeated three times in *Neurospora* and 2 times in the *Sordaria* pheromone precursor. A 12-residue unit QWCRIHGQSCWK is repeated twice in *Neurospora* and three times in the *Sordaria* pheromone precursor. These regions were marked in Figure3.26.

Very interestingly, the region EAQWCRIHGQSCWKKREAEAQWCRI is a possible domain of an uncharacterized protein family UPF034 (IPB001269E) determined by a search of the BLOCK database (Henikoff and Henikoff, 1994). The Block E-value is 1.4.

NIR3_RHILP|P41504: G I **Q** L V T **I H G** R T R M Q F Y **E** G R A D **W** G A **I**

*ccg*4ctg3: 124 E A**QW** C r **I H G** q s C **W** k k r **E** A E A q **W** c **R I** 148

Another possible domain found in the ccg4contig3 protein product after searching BLOCK database are a pheromone/general odorant binding protein signature, Block ID:PR00484E, Block E-value is 1.4.

OBP1_ANTPE|P87508 128 DHCWR I LRVAECFKRSCQE

*ccg*4contig3 36 q sCWKVKRVADa Fa nA IQg 54

The possible hydrophobic and hydrophilic profiles in the ccg4contig3 protein product were analyzed further. Here a hydrophobicity plot displaying the distribution of polar and apolar residues in the peptide and to predicting the membrane-spanning segment or antigenic site from its sequences was drawn using the Kyte-Doolittle scale (http://arbl.cvmbs.colostate.edu/molkit)(Kyte and Doolittle, 1982). In this program, a hydropathic region achieves a positive value (symbol as CH3) and hydrophilic region has a negative value (symbol as H_2O). Window size of 5-7 was chosen because it has been shown by others to be a good value for predicting putative hydrophilic surface-exposed regions. Alternatively, a window size of 19-21 can be used to generate a plot in which a transmembrane domain can stands out sharply. **Figure 3.27 and 3.28** are the outputs of the hydrophobicity plots of *Neurospora ccg*4contig3 putative peptide and *Sordaria* pheromone precursor. The hydrophobic plot of the *Sordaria* pheromone precursor is highly similar to that of its *Neurospora* homolog.

Kyte-Doolittle Hydrophobicity Scales:

Ala (A)	1.8	Leu (L)	3.8
Arg (R)	4.5	Lys (K)	-3.9
Asn (N)	-3.5	Met (M)	1.9
Asp (D)	-3.5	Phe (F)	2.8

Cys (C)	2.5	Pro (P)	-1.6
Gln (Q)	-3.5	Ser (S)	-0.8
Glu (E)	-3.5	Thr (T)	-0.7
Gly (G)	-0.4	Trp (W)	-0.9

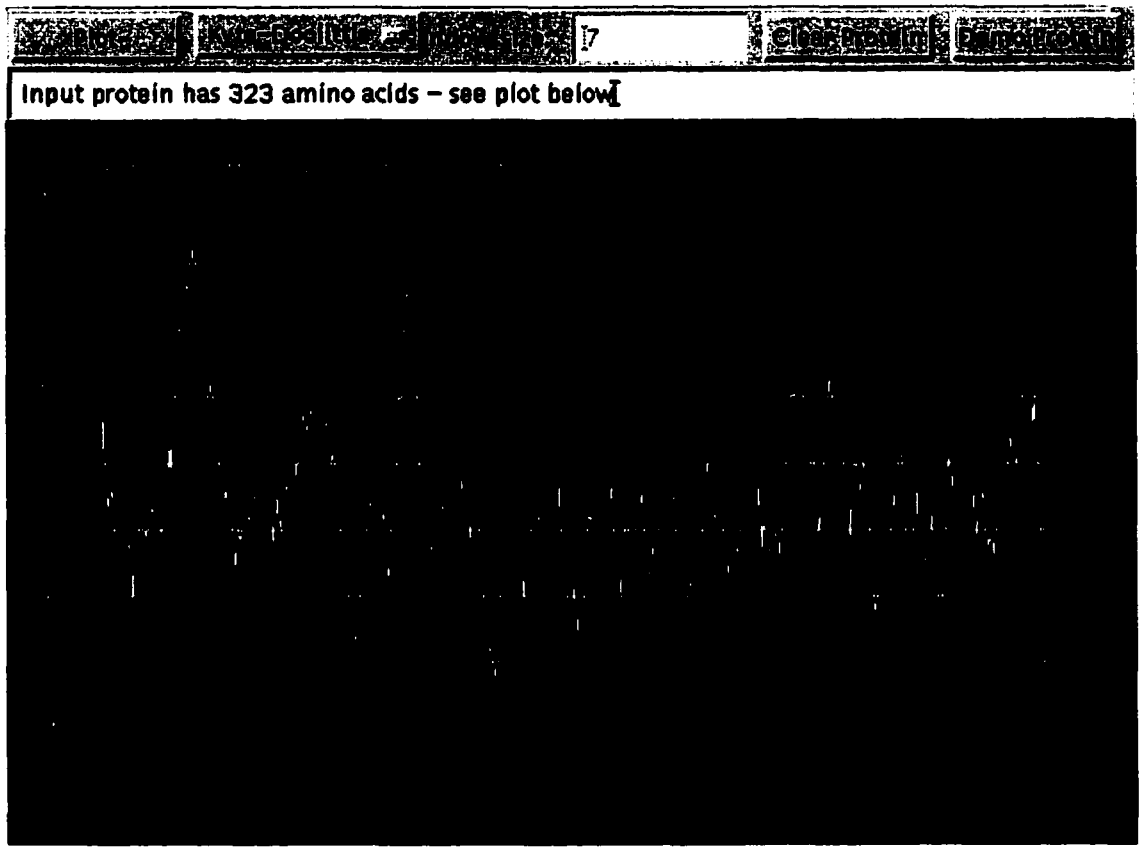


Figure 3.27 Hydrophobicity plot of the *Neurospora crassa* homolog of *Sordaria macrospora* pheromone precursor

His (H)	-3.2	Tyr (Y)	-1.3
Ile (I)	4.5	Val (V)	4.2

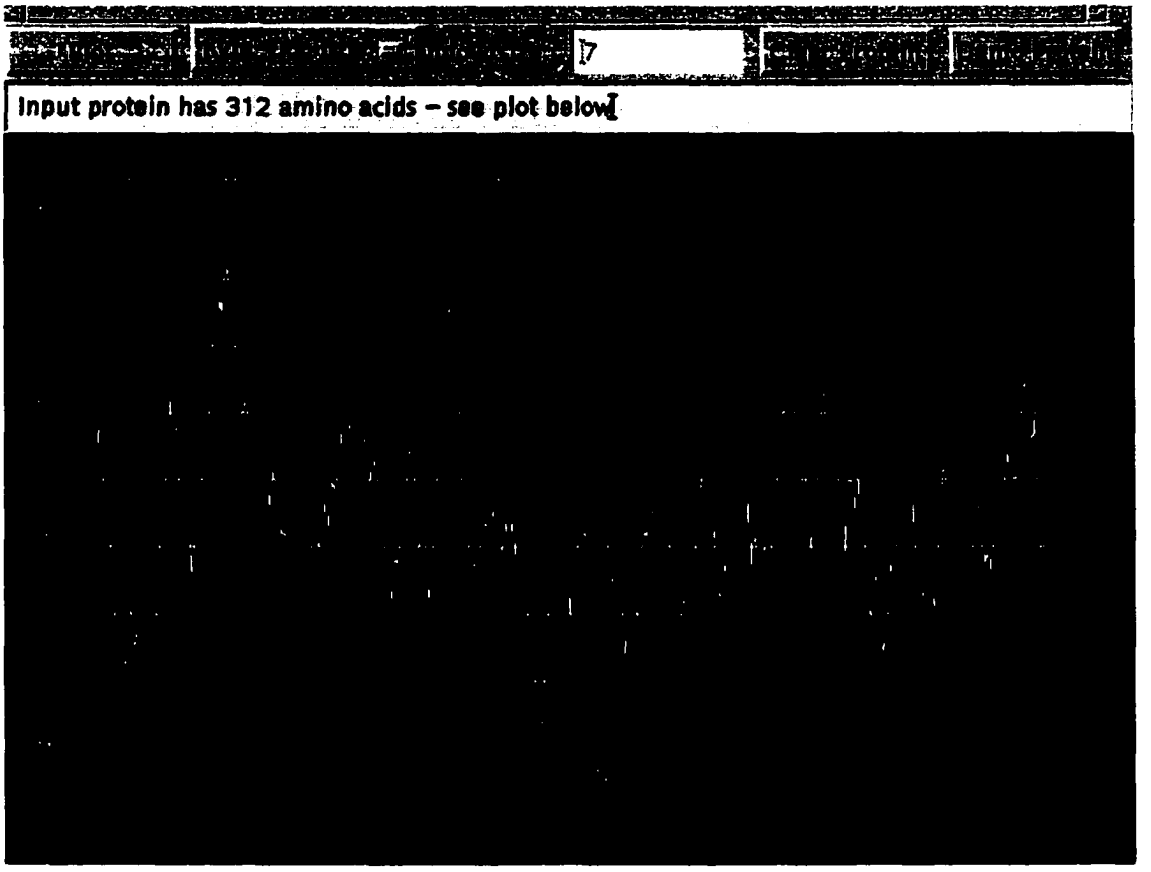


Figure 3.28 Hydrophobicity plot of Sordaria macrospora pheromone precursor

3.2.4.4 The comparison and the correct annotation of the *ccg-4* gene and its cDNA sequence

The genomic DNA sequence and the mRNA sequence for *ccg-4* gene have Genbank accession numbers AF088909 and U46085, respectively. When their sequences were compared to the sequence of *ccg*4ctg3, matched perfectly with the genomic DNA sequence of the *ccg-4* gene except the two GenBank entries lacked the first intron and 5' UTR for the *ccg-4* gene, as can be seen in the Figure 3.22. GenBank entries U46085 and AF088909 were corrected to give accurate version of the annotation of the *ccg-4* gene

genomic DNA sequence as follow:

5'UTR:	423..516
mRNA:	423..1858
gene:	423..1858 gene= "*Neurospora crassa* homolog of pheromone precursor *ppg1* gene of *Sordaria macrospora*"
CDS:	517..1416 gene= "*Neurospora crassa* homolog of pheromone precursor *ppg1* gene of *Sordaria macrospora*" Codon_start=1 Evidence=not_experimental Product=" *Neurospora crassa* homolog of pheromone precursor of *Sordaria macrospora*" Translation="MKFTLPLVIFAAVASATPVAQPNAEAEAQWCRIHGQSCWKVK RVADAFANAIQGMGGLPPRDESGHQPAQVAKRQVDELAGIIALTQEDVNAYYD SLSLQEKFAPSTEEEKKTEKVAKREAEAEAQWCRIHGQSCWKKREAEAQWCRI HGQSCWKRDALPEAEPQWCRIHGQSCWKKRDAAPEAAPEAEANPQWCRIHGQS CWKAKRAAEAVMTAIQSAEAESALLLRDTTFSPVDRVGKRDPQVCNMRLHPKK VCWKRDASPEAACNAPDGSCTKATRDLHAMYNVARAILTAHSDEN"
intron:	1417..1485
3'UTR:	1486..1858

The correctly translated gene product of the ccg-4 protein is the *Neurospora crassa* homolog of *Sordaria macrospora* pheromone precursor (CAB96172) as it perfectly matches with that of the translated protein product from the CDS of the *ccg-4* genomic DNA sequence.

3.2.4.5 The submission of the cDNA sequence of *Neurospora ppg-1* gene homologue

The annotation of *ccg*4ctg3 cDNA sequence and amino acid sequence were submitted to Genbank Jan. 29, 2001 using Sequin (Figure 3.29), the DNA sequence submission and update tool, developed by the NCBI for DNA sequence entries to the GenBank sequence database.

```
LOCUS       ccg4ppg1      1368 bp    mRNA                    29-JAN-2001
DEFINITION  Contig3.
```

ACCESSION *ccg*4ppg1
VERSION
KEYWORDS .
SOURCE *Neurospora crassa.*
ORGANISM *Neurospora crassa*
Unclassified.
REFERENCE 1 (bases 1 to 1368)
AUTHORS Zhu,H., Lai,H., Kupfer,D.M., Bell-Pedersen,D., Loros,J.J., Dunlap,J.C. and Roe,B.A.
TITLE *Neurospora crassa* cDNA Homolog of the Pheromone Precursor Gene,ppg-1, from *Sordaria macrospora*, complete cds
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1368)
AUTHORS Zhu,H., Lai,H., Kupfer,D.M., Bell-Pedersen,D., Loros,J.J., Dunlap,J.C. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-2001) Department of Biochemistry & Chemistry, The University of Oklahoma, 620 Parrington Oval, Norman, Oklahoma 73019, USA
FEATURES Location/Qualifiers
source 1..1368
/organism="*Neurospora crassa*"
5'UTR 1..94
/gene="*Neurospora crassa* ppg1 gene homolog"
gene 1..1368
/gene="*Neurospora crassa* ppg1 gene homolog"
CDS 95..994
/gene="*Neurospora crassa* ppg1 gene homolog"
/note="*Neurospora crass* pheromone precursor homolog of *Sordaria macrospora* ppg1(AC#:AJ249863). *Neurospora crassa ccg-4* protein 2 (AC#:T47209 or U46085) is a partial product of this gene. It starts at 731 bp and ends at 991 bp."
/codon_start=1
/evidence=not_experimental
/product="*Neurospora crassa* pheromone precursor homolog of *Sordaria macrospora* ppg-1"
/translation="MKFTLPLVIFAAVASATPVAQPNAEAEAQWCRIHGQSCWKVKRV
DAFANAIQGMGGLPPRDESGHQPAQVAKRQVDELAGIIALTQEDVNAYYDSLSLQEK
FAPSTEEEKKTEKVAKREAEAEAQWCRIHGQSCWKKREAEAQWCRIHGQSCWKRDALP
EAEPQWCRIHGQSCWKKRDAAPEAAPEAEANPQWCRIHGQSCWKAKRAAEAVMTAIQS
AEAESALLLRDTTFSPVDRVGKRDPQVCNMRLHPKKVCWKRDASPEAACNAPDGSCTK
ATRDLHAMYNVARAILTAHSDEN"
repeat_unit 167..214
/gene="*Neurospora crassa* ppg1 gene homolog"
/note="16 amino acid residues repeat 3 times in this protein:EAEAQWCRIHGQSCWK. A 12-residue region QWCRIHGQSCWK repeats twice. In the *Sordaria macrospora* ppg1, there are only two repeats of the above 16-residue unit. The 12-residue repeats three times. The positions of the 16-residue repeat unit are at 25-40aa, 122-137aa, 140-155aa. The 12-residue repeat units are at 165-176aa and

```
                     193-204aa."
                     /evidence=not_experimental
     repeat_unit     458..505
                     /gene="Neurospora crassa ppg1 gene homolog"
     repeat_unit     512..559
                     /gene="Neurospora crassa ppg1 gene homolog"
     repeat_unit     587..622
                     /gene="Neurospora crassa ppg1 gene homolog"
     repeat_unit     671..706
                     /gene="Neurospora crassa ppg1 gene homolog"
     3'UTR           995..1368
                     /gene="Neurospora crassa ppg1 gene homolog"
poly(A) Signal site   1351..1357
BASE COUNT      300 a    442 c    352 g    274 t
ORIGIN
    1 tgccaatcaa ccctcagagg tcttcattct ctcaatcaac agggtccttt cgttgacact
   61 ttttacattc ttcatccaag ccgttttgtt caagatgaag ttcactctcc ctcttgtcat
  121 cttcgccgcc gtggcctccg ccaccccggt cgcccagcca aatgccgagg ccgaagccca
  181 gtggtgccgg atccatggcc agtcctgctg gaaggtcaag cgtgttgccg atgccttcgc
  241 caacgccatc cagggcatgg gtggtctccc tccccgcgat gagtccggcc accagcccgc
  301 ccaggtcgcc aagcgccagg tcgacgagct tgccggcatc atcgccctca ctcaggagga
  361 cgttaacgcc tactacgact cgctcagcct ccaagagaag ttcgctccct ccaccgagga
  421 ggagaagaag accgagaagg tcgccaagcg tgaggccgag gccgaggcac aatggtgccg
  481 catccacggc caatcgtgct ggaagaagcg tgaggccgaa gcccagtggt gccggatcca
  541 cggccagtcg tgctggaagc gtgacgccct ccccgaggcc gagccccagt ggtgccgcat
  601 ccacggccag tcctgctgga agaagcgtga cgccgctccc gaggctgctc ccgaagctga
  661 ggccaacccg caatggtgcc gcatccacgg ccagtcctgc tggaaggcca agcgcgccgc
  721 cgaggccgtc atgaccgcca tccagtccgc cgaggccgag tccgcccttc ttctccgtga
  781 caccaccttc agccccgtcg accgcgtcgg caagcgtgat ccccaggtgt gcaacatgag
  841 gctccacccc aaaaaggttt gctggaagcg tgatgcctcc cctgaggccg cttgcaacgc
  901 tcccgacggc tcctgcacca aggctacccg tgacttgcac gccatgtaca acgtggctcg
  961 tgccatcctc actgctcact ccgatgagaa ctagatgttt ccccgaaatc ttcttcccta
 1021 ggagagggta cgacaaagca aaccaaaagg gggcagagaa aaggcttgga tatatcactt
 1081 ttgctccctc tttacacact tcacttcttt ctggaaaaaa aaaagtacaa cacatctaca
 1141 tacaccacag aagcttttg ggttcacact caacacaaag ggattggata ggtcaatact
 1201 tgtgcaatcc tgtttttatc aattattctt caagtgggat gggaaggggg gtattggagc
 1261 atgcctgtac ttagctagtt ttgaacctct cccgctcacc cggacactga attcaaggat
 1321 gaggctgctt gccgatcgaa ggcagcctcg aatactttcc ttgaaaaa
```

Figure 3.29 The GenBank submission file of the *Neurospora crassa* cDNA for the pheromone precursor homolog of *Sordaria macrospora*.

3.2.5 The clock-controlled gene 6

As mentioned previously, the *ccg-6* gene displayed an unexpected expression pattern in the *Neurospora crassa*. According to published data, it was supposed to have higher expression in the morning cDNA library (Bell-Pedersen et al., 1996c). However,

in this research, the expression of *ccg-6* in the NM cDNA library was much lower than it was in the NE cDNA library (Table 3.34).

Table 3.34 The contigs and the EST counts of *ccg-6* in combined EST database

NMNE ctg#	EST counts		Alignment		%Homology
	NM	NE	DNA	protein	
Contig1122	0	13	700-488	74-142	97
Contig1165	0	15	754-545	74-142	97
Contig1435c	24	81	798-1226	1-142	100
Contig1435d	1	10	609-256	26-142	98
Total	25	119	NM/NE=25/119*1.2=0.2		

There were 119 *ccg-6* ESTs in the NE library were 119 while there were 25 *ccg-6* ESTS observed in the NM library. The ratio of the redundancy of *ccg-6* specific ESTs in those two cDNA libraries is 0.2. Comparison of the genomic DNA sequence (AF088908) and the mRNA sequence (U46086) of the *ccg-6* (Figure3.30) with the sequence of contig1435c reveals that the contig is similar to the *ccg-6* genomic sequence except for two of the introns in the *ccg-6* genomic DNA sequence (Figure 3.31) and is 97% similar to the *ccg-6* mRNA sequences (Figure 3.32). The 5' UTR region between position 8 and 150 in the contig1435c matches the region between position 1 and 144 in the *ccg-6* mRNA sequence and the region from position 676 to 819 in the *ccg-6* genomic DNA sequence. The 3' UTR region between position 580 and 761 in the contig1435c matches the region between position 573 and 754 in the *ccg-6* mRNA sequence and the region from position 1567 to1748 in the *ccg-6* genomic DNA sequence. Two introns in the *ccg-6* gene genomic sequences were not present in the sequence both of the *ccg-6* mRNA

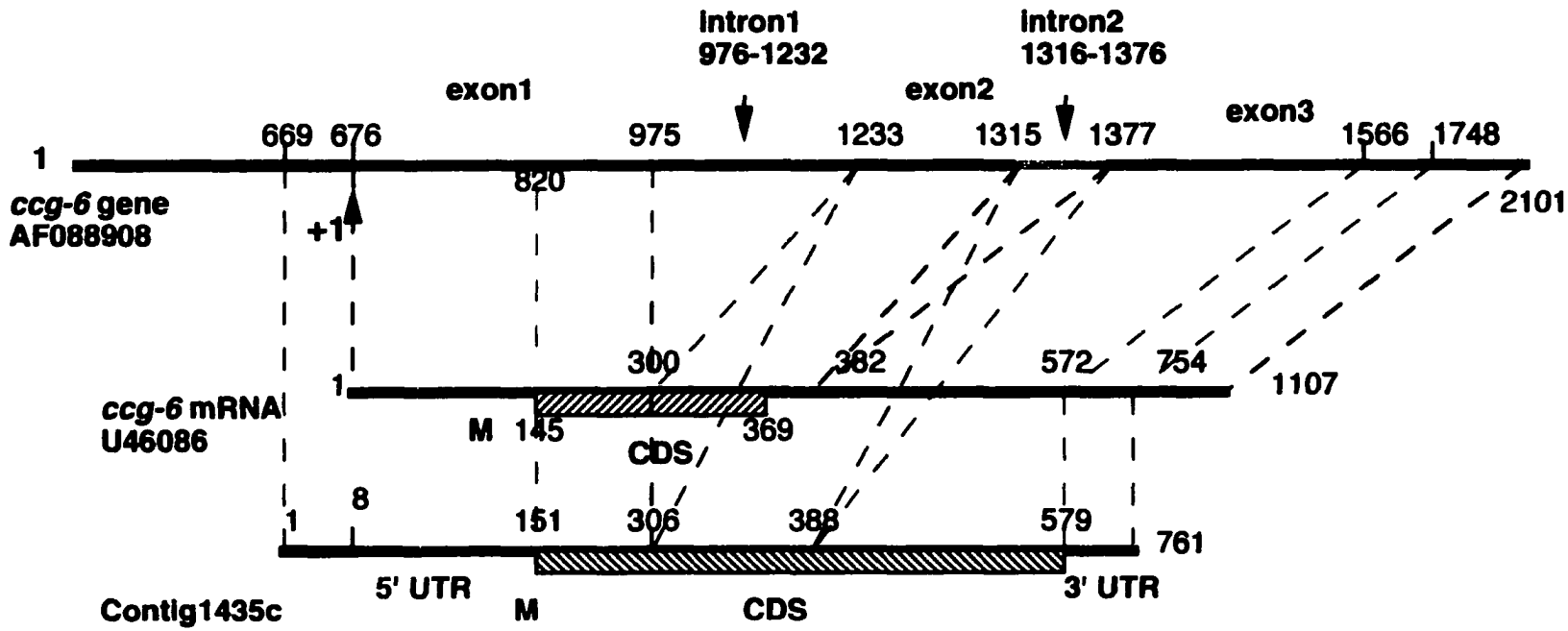


Figure 3.30 The annotation of the *ccg-6* gene, the *ccg-6* mRNA and the cDNA represented by Contig1435c . Same color indicates the similar DNA sequences

(U46086) and the contgi1435c. The peptide length of the protein encoded by the ORF of *ccg-6* (AF088908) is 142 amino acids. The translated peptide sequence of the ccg6contig1435c has 100% identity to the *ccg-6* gene protein sequence (AAC64287) (Figure3.33). However, the translated peptide of the ccg-6 mRNA has only 74 amino acid residues, and 12 of the 74 residues are different than those in the sequences of the contig1435c and the ccg-6 gene. These 12 residues are underlined in the Figure3.33. The DNA sequences of the contig1435c and the ccg-6 mRNA were carefully examined. Since

```
 273  2.61 0.33 0.00  Contig1435c        1   307 (1069)
gi|3746898|gb|AF088908.1|ccg-6 gene     669   976 (1125)

 Contig1435c          1 CTCACTCCACACATCTTGGCTTTTCAGTCAGTTCCATTCGTTTTAATATC 50
                                          vi                 iv
 gi|3746898|gb|A    669 CTCACTCCACACATCTTGGCTTGCCAGTCAGTTCCATTCGCGTTAATATC 718

 Contig1435c         51 TCAGCTGCCGACGT-TCGCCGTCCTCAACTTTTCTTACCCCCAACGACGA 99
                                     -                v?          ?
 gi|3746898|gb|A    719 TCAGCTGCCGACGTGTCGCCGTCCTCAACTTANCTTACCCCCANCGACGA 768

 Contig1435c        100 ATACACAACTCCTTCTCTGAGCACATAATCCTCACCAACAAATCAGCCAA 149

  gi|3746898|gb|A   769 ATACACAACTCCTTCTCTGAGCACATAATCCTCACCAACAATTCAGCCAA 818

 Contig1435c        150 AATGAAGTTCTCTGCTGCCGCCGTCCTTGCCGCTGCCGCTGGCGCCCACG 199

gi|3746898|gb|A    819 AATGAAGTTCTCTGCTGCCGCCGTCCTTGCCGCTGCCGCTGGCGCCCACG 868

 Contig1435c        200 CCTGGAGCAACGTGACCTACACCACCGAGATTGTCACTGCCGTCACCACC 249

 gi|3746898|gb|A    869 CCTGGAGCAACGTGACCTACACCACCGAGATTGTCACTGCCGTCACCACC 918

 Contig1435c        250 TACTGCCCTGGCCCCACTGAGATCACCCATGGCGGCAACACTTACACCGT 299

 gi|3746898|gb|A    919 TACTGCCCTGGCCCCACTGAGATCACCCATGGCGGCAACACTTACACCGT 968

 Contig1435c        300 CACCGAGG 307 intron1

 gi|3746898|gb|A    969 CACCGAGG 976

 Contig1435c                        305   389  (987)
 gi|3746898|gb|AF088908.1|ccg-6     1231  1315 (786) *

 Contig1435c        305 AGGCCACCACCTTGACCATCTCTGACTGCCCTTGCACTGTCACCAAGCCC 354

 gi|3746898|gb|A   1231 AGGCCACCACCTTGACCATCTCTGACTGCCCTTGCACTGTCACCAAGCCC 1280

 Contig1435c        355 ATCATCACCACCTCGTCAGTGATCTGCCACAGCTG 389 intron 2

 gi|3746898|gb|A   1281 ATCATCACCACCTCGTCAGTGATCTGCCACAGCTG 1315

  Contig1435c                     389   759  (617)
gi|3746898|gb|AF088908.1|ccg-6   1376  1746 (355)
```

```
Contig1435c        389 GCACCGGCTACGTCAACTCTACCATCCCTGCCCCCACCTCGGCCGGCTCC 438

gi|3746898|gb|A   1376 GCACCGGCTACGTCAACTCTACCATCCCTGCCCCCACCTCGGCCGGCTCC 1425

Contig1435c        439 GTTGGCACTGGCAGCGCCCCCGCTGTCGTCACTCCCACCGTCAGCCCCTC 488

gi|3746898|gb|A   1426 GTTGGCACTGGCAGCGCCCCCGCTGTCGTCACTCCCACCGTCAGCCCCTC 1475

Contig1435c        489 CGAGGTCCCCACCGCCGGTGCTGGCAAGGCTGCTGCCCTCTCTGGCGCCG 538

gi|3746898|gb|A   1476 CGAGGTCCCCACCGCCGGTGCTGGCAAGGCTGCTGCCCTCTCTGGCGCCG 1525

Contig1435c        539 GTCTCGTTGGTGTTCTCGGTCTCGCTGCCATCCTCCTCTAATTTCCTGTC 588

gi|3746898|gb|A   1526 GTCTCGTTGGTGTTCTCGGTCTCGCTGCCATCCTCCTCTAATTTCCTGTC 1575

Contig1435c        589 CATGCGTGACAGCCAAGCGTTTCGACTTCCATGTCTCATGATCCCTCGAT 638

gi|3746898|gb|A   1576 CATGCGTGACAGCCAAGCGTTTCGACTTCCATGTCTCATGATCCCTCGAT 1625

Contig1435c        639 TCTCTTGATCGGTAATAGTTAATACTGCCCCGTCAACTACCATGTTTCGG 688

gi|3746898|gb|A   1626 TCTCTTGATCGGTAATAGTTAATACTGCCCCGTCAACTACCATGTTTCGG 1675

Contig1435c        689 CATGACGTTGCGGCGCCAAGGGGGTTGCGCTTGGAGAACTCAAATGGCTG 738

gi|3746898|gb|A   1676 CATGACGTTGCGGCGCCAAGGGGGTTGCGCTTGGAGAACTCAAATGGCTG 1725

Contig1435c        739 TGGAGTTGGATGTATTATGGA 759

gi|3746898|gb|A   1726 TGGAGTTGGATGTATTATGGA 1746
```

Figure 3.31 The comparison of the DNA sequence between the contig1435c and the *ccg-6* gene (AF088908) using crossmatch. Two introns of the *ccg-6* gene are showed in this figure.

```
ccg-6 mRNA  1 CACACATCTTGGCTTGCCAGTCAGTTCCATTCGCGTTAATATCTCAGCTG 50
              |||||||||||||||  ||||||||||||||||  ||||||||||||||
contig1435c 8 CACACATCTTGGCTTTTCAGTCAGTTCCATTCGTTTTAATATCTCAGCTG 57

           51 CCGACGTGTCGCCGTCCTCAACTTAGCTTACCCCCATCGACGAATACACA 100
              |||||||  ||||||||||||||||  |||||||||  ||||||||||||
           58 CCGACGT.TCGCCGTCCTCAACTTTTCTTACCCCCAACGACGAATACACA 106

          101 ACTCCTTCTCTGAGCACATAATCCTCACCAACAATTCAGCCAAAATGAAG 150
              |||||||||||||||||||||||||||||||||  ||||||||||||||||
          107 ACTCCTTCTCTGAGCACATAATCCTCACCAACAAATCAGCCAAAATGAAG 156

          151 TTCTCTGCTGGCGACGTCCTTGCCGCTGCCGCTGGCGCCCACGCCTGGAG 200
              ||||||||||  || ||||||||||||||||||||||||||||||||||||
          157 TTCTCTGCTGCCGCCGTCCTTGCCGCTGCCGCTGGCGCCCACGCCTGGAG 206

          201 CAACGTGACCTACACCACCGAGATTGTCACTGTCGTCACCACCTACTGGC 250
              |||||||||||||||||||||||||||||||||  ||||||||||||||| |
          207 CAACGTGACCTACACCACCGAGATTGTCACTGCCGTCACCACCTACTGCC 256

          251 CTGGTCCCACTGAGATCACCCATGGCGGCAACACTTACACCGTCACCGAG 300
              ||||  ||||||||||||||||||||||||||||||||||||||||||||
          257 CTGGCCCCACTGAGATCACCCATGGCGGCAACACTTACACCGTCACCGAG 306

          301 GCCACCACCTTGACCATCTCTGACTGCCCTTGCACTGT.ACCAAGCCCAT 349
              |||||||||||||||||||||||||||||||||||||  ||||||||||
```

```
307 GCCACCACCTTGACCATCTCTGACTGCCCTTGCACTGTCACCAAGCCCAT 356
    .         .         .         .         .
350 CATCACCACCTCGTCAGTGATCTGCCACAGCTGCACCGGCTACGTCAACT 399
    ||||||||||||||||||||||||||||||||||||||||||||||||||
357 CATCACCACCTCGTCAGTGATCTGCCACAGCTGCACCGGCTACGTCAACT 406

400 CTACCATCCCTGCCCCCACCTCGGTCGGCTCCGTTGGCACTGGCAGTGCC 449
    |||||||||||||||||||||| |||||||||||||||||||| |||
407 CTACCATCCCTGCCCCCACCTCGGCCGGCTCCGTTGGCACTGGCAGCGCC 456
    .         .         .         .         .
450 CCCGCTGTCGTCACTCCCACCGTCAGCCCCTCCGAGGTCCCCACCGCCGG 499
    ||||||||||||||||||||||||||||||||||||||||||||||||||
457 CCCGCTGTCGTCACTCCCACCGTCAGCCCCTCCGAGGTCCCCACCGCCGG 506
    .         .         .         .         .
500 TGCTGGCAAGGCTGCTGCCCTCTCTGGCGCCGGTCTCGTTGGTGTTCTCG 549
    ||||||||||||||||||||||||||||||||||||||||||||||||||
507 TGCTGGCAAGGCTGCTGCCCTCTCTGGCGCCGGTCTCGTTGGTGTTCTCG 556
    .         .         .         .         .
550 GTCTCGCTGCCATCCTCCTCTAATTTCCTGTCCATGCGTGACAGCCAAGC 599
    ||||||||||||||||||||||||||||||||||||||||||||||||||
557 GTCTCGCTGCCATCCTCCTCTAATTTCCTGTCCATGCGTGACAGCCAAGC 606
    .         .         .         .         .
600 GTTTCGACTTCCATGTCTCATGATCCCTCGATTCTCTTGATCGGTAATAG 649
    ||||||||||||||||||||||||||||||||||||||||||||||||||
607 GTTTCGACTTCCATGTCTCATGATCCCTCGATTCTCTTGATCGGTAATAG 656
    .         .         .         .         .
650 TTAATACTGCCCCGTCAACTACCATGTTTCGGCATGACGTTGCGGCGCCA 699
    ||||||||||||||||||||||||||||||||||||||||||||||||||
657 TTAATACTGCCCCGTCAACTACCATGTTTCGGCATGACGTTGCGGCGCCA 706
    .         .         .         .         .
700 AGGGGGTTGCGCTTGGAGAACTCAAATGGCTGTGGAGTTGGATGTATTAT 749
    ||||||||||||||||||||||||||||||||||||||||||||||||||
707 AGGGGGTTGCGCTTGGAGAACTCAAATGGCTGTGGAGTTGGATGTATTAT 756

750 GGATG 754
    ||| |
757 GGACG 761
```

Figure 3.32 The sequence comparison between the contig1435c and the *ccg-6* mRNA (U46086) using GCG program Bestfit. The base C at the position 345bp of the contig1435c was missing in the sequence of the *ccg-6* mRNA. The translation start codon and stop codons are indicated in bold font.

```
                1                                                  50
Ccg6ctg1435c  MKFSAAAVLA AAAGAHAWSN VTYTTEIVTA VTTYCPGPTE ITHGGNTYTV
      Ccg6aa  MKFSAAAVLA AAAGAHAWSN VTYTTEIVTA VTTYCPGPTE ITHGGNTYTV
  Ccg6mrnaaa  MKFSAGDVLA AAAGAHAWSN VTYTTEIVTV VTTYWPGPTE ITHGGNTYTV

               51                                                 100
Ccg6ctg1435c  TEATTLTISD CPCTVTKPII TTSSVICHSC TGYVNSTIPA PTSAGSVGTG
      Ccg6aa  TEATTLTISD CPCTVTKPII TTSSVICHSC TGYVNSTIPA PTSAGSVGTG
  Ccg6mrnaaa  TEATTLTISD CPCTVPSPSS PPRQ------ ---------- ----------

              101                                        142
Ccg6ctg1435c  SAPAVVTPTV SPSEVPTAGA GKAAALSGAG LVGVLGLAAI LL
      Ccg6aa  SAPAVVTPTV SPSEVPTAGA GKAAALSGAG LVGVLGLAAI LL
  Ccg6mrnaaa  ---------- ---------- ---------- ---------- --
```

Figure3.33 The comparison of contig1435c peptide with *ccg-6* gene protein (AAC64287.1) and the *ccg-6* mRNA peptide (U46086).

their DNA sequences were only 97% identical, with the different bases marked in dark bolded fonts in the Figure3.32, the trace files of the contig1435c (Figure3.34) were checked. From examination of the trace files, it was shown that the sequence of the contig1435c is correct. Especially, a important base pair "C" at the position 345 of the contig1435c was missed from the earlier reported *ccg-6* mRNA. Therefore, the correct

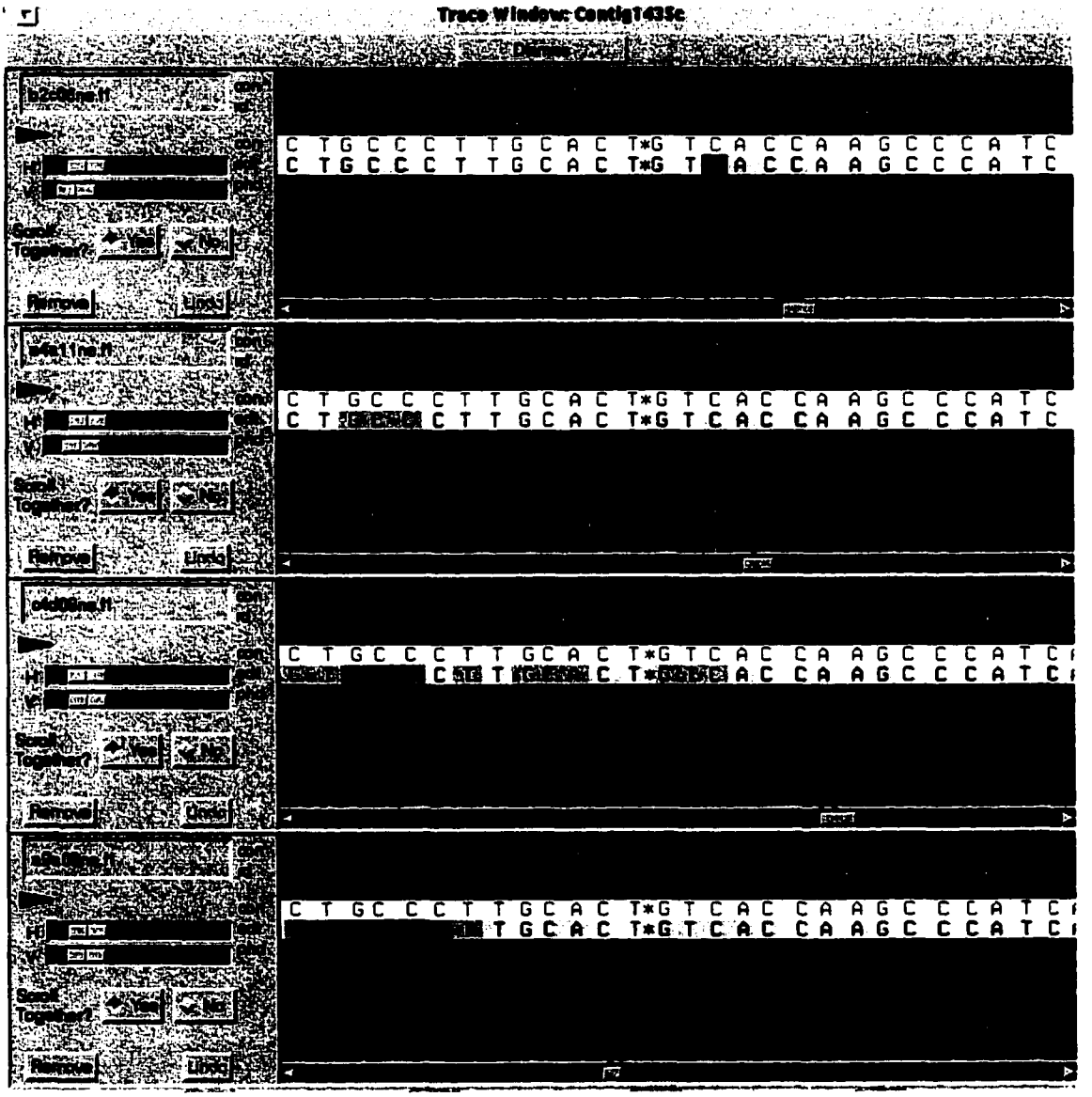


Figure 3.34 The trace files in the contig1435c show the base "C" at the position 345 bp. This base C was missing in the sequence of the *ccg-6* mRNA (u46086).

sequence in this region is $_{343}$GTC ACC AAG CCC ATC ATC ACC ACC TCG TCA GTG ATC TGC$_{381}$..... The correct translated peptide is VTKPIITTSSVIC....Since the "C" was missed from the sequence of the *ccg-6* mRNA, the sequence became GTA CCA AGC CCA TCA TCA CCA CCT CGT CAG TGA TCT GC.... Therefore, the translated peptide became VPSPSSPPRG. Then, the translation stopped at the stop codon TGA. This is why the translation product of the *ccg-6* mRNA (U46086) had only 74 residues. Actually, it is a incorrect translation product due to a sequencing error. Therefore, the accurate annotation for this mRNA sequence is corrected to:

3' UTR: 573...1106

5'UTR: 1...144

CDS: 145-572

The sequences of these 4 contigs representing ccg-6 were examined both on the level of nucleotide sequence level and on the protein sequence level. The translated protein sequence (Figure 3.35) reveals that both the contig1435c and the contig1122 represent a complete cDNA, but the 5' end of the cDNA represented by contig1122 is shorter than that of contig1435c. The contig1165 and contig1435d represent the incomplete cDNAs. They missed the 5' ends of their messages. However, all these 4 contigs contains the similar codons for the carboxyl end of the *ccg-6* gene protein (Figure3.36). However, the position of the poly (T) tails are different (Figure3.36). Therefore, alternative polyadenylation also exists in the RNA processing of the *ccg-6* gene in *Neurospora crassa*.

```
                1                                                  50

   Ccg6ctg1435c  MKFSAAAVLA AAAGAHAWSN VTYTTEIVTA VTTYCPGPTE ITHGGNTYTV
Nmneccg6ctg1435d ---------- ---------- -----EMVTA VTTYCPGPTE ITQGGNTYTV
```

```
Nmneccg6ctg1165  ---------- ---------- ---------- ---------- ----------
Nmneccg6ctg1122  ---------- ---------- ---------- ---------- ----------

                 51                                                  100
    Ccg6ctg1435c TEATTLTISD CPCTVTKPII TTSSVICHSC TGYVNSTIPA PTSAGSVGTG
Nmneccg6ctg1435d TEATTLTISD CPCTVTKPII TTSSVICHSC TG*VNSTIPA PTSAGSVGTG
 Nmneccg6ctg1165 ---------- ---------- ---SVICHSC TGYVNSTIPA PTSAGSVGTG
 Nmneccg6ctg1122 ---------- ---------- --MSVICHSC TGTSNSTIPA PTSAGSVGTG

                 101                                             142
    Ccg6ctg1435c SAPAVVTPTV SPSEVPTAGA GKAAALSGAG LVGVLGLAAI LL
Nmneccg6ctg1435d SAPAVVTPTV SPSEVPTAGA GKAAALSGAG LVGVLGLAAI LL
 Nmneccg6ctg1165 SAPAVVTPTV SPSEVPTAGA GKAAALSGAG LVGVLGLAAI LL
 Nmneccg6ctg1122 SAPAVVTPTV SPSEVPTAGA GKAAALSGAG LVGVLGLAAI LL
```

Figure 3.35 The translated peptide sequences comparison of 4 contigs represented the *ccg-6* gene in the combined EST database

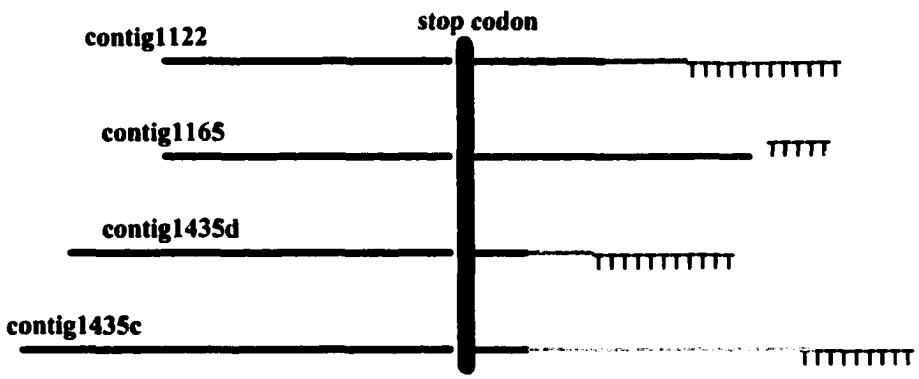


Figure 3.36 The different polyadenylation sites of the 4 contigs represented the *ccg-6* gene in the *Neurospora crassest* database. The sequences before the stop codon are same for 4 contigs. 4 different color before the poly(T) tail indicates the difference of the sequences for the 4 contigs. Same color means the DNA sequence is same.

3.2.6 The clock-controlled gene 8

No ESTs were observed for the *ccg-8* gene in the NE library although there were 37 ESTs from the *ccg-8* gene transcript in the morning library (Table3.35).

Table 3.35 The contigs and the EST counts of *ccg-8* in combined EST database

NMNE ctg#	ESTs count		Alignment		% homolog
	NM	NE	DNA	Protein	
Contig1356	37	0	3-497	110-274	98

Total	37	0	NM/NE=∝

The comparison of the *ccg-8* and contig1356 is presented in **Figure3.37**. The genomic DNA sequence of the *ccg-8* gene (AF088907) and another mRNA sequence (U46087) were downloaded from Genbank and used for comparison. The translated protein of the *ccg-8* gene is 274 amino acid in length (Q01306). However, the putative translated protein from the *ccg-8* mRNA (U46087) has only 142 amino acids.

From the studies of both the DNA sequences (Figure 3.38) and amino acid sequences (Figure3.39), it is obvious that the cDNA represented by contig1356 and the *ccg-8* mRNA (U46087) are the same product of the *ccg-8* gene as their DNA sequences and amino acid sequences matched exactly, and they have a same polyadenylation site. Even though the cDNA represented by contig1356 is several bases longer than that of the *ccg-8* mRNA, both will give the same protein.

```
          2521
   Ccg8dna2  GGCACCAACG AGCTGAGCTC AGTACTTGAT CAGTATGACA GGTCGACAAC 2570
   Ccg8mrna  ---ACCAACG AGCTGAGCTC AGTACTTGAT CAGTATGACA GGTCGACAAC 47
Ccg8ctg1356  ----CCAACG AGCTGAGCTC AGTACTTGAT CAGTATGACA GGTCGACAAC 46

   Ccg8dna2  CACGAGGGAT GGTGAAGACC ATGTCATGAC CAACGGGTCC GCGCCGGAAG 2620
   Ccg8mrna  CACGAGGGAT GGTGAAGACC ATGTCATGAC CAACGGGTCC GCGCCGGAAG 97
Ccg8ctg1356  CACGAGGGAT GGTGAAGACC ATGTCATGAC CAACGGGTCC GCGCCGGAAG 96

   Ccg8dna2  AAGATCGTAC GCGGCTGGTG GAGCGTATGA CTGAACTTCG GAAGAGCATT 2670
   Ccg8mrna  AAGATCGTAC GCGGCTGGTG GAGCGTATGA CTGAACTTCG GAAGAGCATT 147
Ccg8ctg1356  AAGATCGTAC GCGGCTGGTG GAGCGTATGA CTGAACTTCG GAAGAGCATT 146

   Ccg8dna2  TACTCCAACA TTAAGGACGT CTCTTCGGTT GTTGCCAATT ACACTGGCGC 2720
   Ccg8mrna  TACTCCAACA TTAAGGACGT CTCTTCGGTT GTTGCCAATT ACACTGGCGC 197
Ccg8ctg1356  TACTCCAACA TTAAGGACGT CTCTTCGGTT GTTGCCAATT ACACTGGCGC 196

   Ccg8dna2  TGCTTTGCCC GAAAATGCCG GCAACCTTGT TCGGCATCAT CTGCTCAGCC 2770
   Ccg8mrna  TGCTTTGCCC GAAAATGCCG GCAACCTTGT TCGGCATCAT CTGCTCAGCC 247
Ccg8ctg1356  TGCTTTGCCC GAAAATGCCG GCAACCTTGT TCGGCATCAT CTGCTCAGCC 246

   Ccg8dna2  TTCCCATGGT CTGGAGTCAG GCTAGCAAAA CCACTACCTC GAGCGAGCAA 2820
   Ccg8mrna  TTCCCATGGT CTGGAGTCAG GCTAGCAAAA CCACTACCTC GAGCGAGCAA 297
Ccg8ctg1356  TTCCCATGGT CTGGAGTCAG GCTAGCAAAA CCACTACCTC GAGCGAGCAA 296

   Ccg8dna2  CCAAACACGC AAGACCCGAA TGCACTTGTG CGCAAGGCTG CCAAGGTGGC 2870
   Ccg8mrna  CCAAACACGC AAGACCCGAA TGCACTTGTG CGCAAGGCTG CCAAGGTGGC 347
Ccg8ctg1356  CCAAACACGC AAGACCCGAA TGCACTTGTG CGCAAGGCTG CCAAGGTGGC 346
```

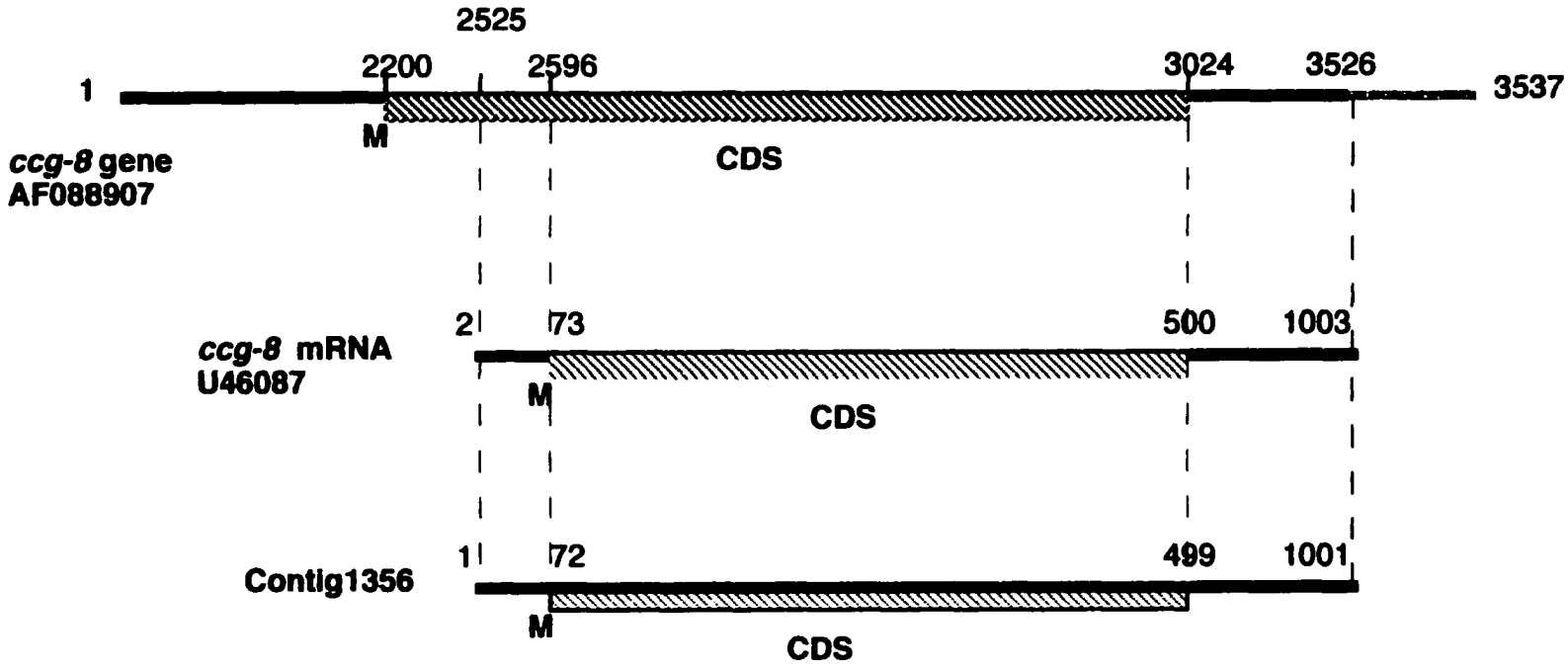


Figure 3.37 The annotation of *ccg-8* gene , *ccg-8* mRNA and Contig1356. The ORF sequences of the ccg-8 mRNA and the contig1356 are exactly same .

```
   Ccg8dna2  TCTCTCCTTT AGCAAAGAGG GATTGCACAT TTTCTCACAG ATCATGGAGA 2920
   Ccg8mrna  TCTCTCCTTT AGCAAAGAGG GATTGCACAT TTTCTCACAG ATCATGGAGA 397
Ccg8ctg1356  TCTCTCCTTT AGCAAAGAGG GATTGCACAT TTTCTCACAG ATCATGGAGA 396

 Ccg8dna2  TCATTCGTCC CGCGACCGAT CACGCGGAAG ACTGGAGGAA CAAGAAGATG 2970
   Ccg8mrna  TCATTCGTCC CGCGACCGAT CACGCGGAAG ACTGGAGGAA CAAGAAGATG 447
Ccg8ctg1356  TCATTCGTAC CGCGACCGAT CACGCGGAAG ACTGGAGGAA CAAGAAGATG 446

   Ccg8dna2  AACCAAATGA CCCCTGCGAA CGGCACAGAG CAGGAGATTC GACCTTTGAT 3020
   Ccg8mrna  AACCAAATGA CCCCTGCGAA CGGCACAGAG CAGGAGATTC GACCTTTGAT 497
Ccg8ctg1356  AACCAAATGA CCCCTGCGAA CGGCACAGAG CAGGAGATTC TACCTTTGAT 496

   Ccg8dna2  TTGACCCAAC CTCTTCCCCA GGTTGCTACC ATCAATGGTG ATATTCCCAT 3070
   Ccg8mrna  TTGACCCAAC CTCTTCCCCA GGTTGCTACC ATCAATGGTG ATATTCCCAT 547
Ccg8ctg1356  T..GACCAAC CTCTTCCCCA GGTTGCTACC ATCAATGGTG ATATTCCCAT 544

   Ccg8dna2  GAGGTAGAGA CTCTTGAAAT TTGTGAGGAA ACCGCAGCGA GCACCTATTG 3120
   Ccg8mrna  GAGGTAGAGA CTCTTGAAAT TTGTGAGGAA ACCGCAGCGA GCACCTATTG 596
Ccg8ctg1356  GAGGTAGAGA CTCTTGAAAT TTGTGAGGAA ACCGCAGCGA GCACCTATTG 594

   Ccg8dna2  TTCCAAAAAT TTGTTTGTTG TAAAAAAGAA GGATAGCTTC GGCGGTCAGG 3170
   Ccg8mrna  TTCCAAAAAT TTGTTTGTTG TAAAAAAGAA GGATAGCTTC GGCGGTCAGG 647
Ccg8ctg1356  TTTCAAAAAT TTGTTTGTTG TAAAAAAGAA GGATAGCTTC GGCGGTCAGG 644

   Ccg8dna2  GGTTTCGGCG ATTGAAAACT CATGTTATGA CTAGAGCTTT TTGGTTGGCG 3220
   Ccg8mrna  GGTTTCGGCG ATTGAAAACT CATGTTATGA CTAGAGCTTT TTGGTTGGCG 697
Ccg8ctg1356  GGTTTCGGCG ATTGAAAACT CATGTTATGA CTAGAGCTTT TTGGTTGGCG 694

   Ccg8dna2  TAAAGAGCTG GGCTTGGATA ATTGGTTGGG ATTTGAGAGA GACTTATGCC 3270
   Ccg8mrna  TAAAGAGCTG GGCTTGGATA ATTGGTTGGG ATTTGAGAGA GACTTATGCC 747
Ccg8ctg1356  TAAAGAGCTG GGCTTGGATA ATTGGTTGGG ATTTGAGAGA GACTTATGCC 744

   Ccg8dna2  AAGTCTCGGG ACATTAACAG TTATATACCA CTTTTCTTTT CGTTTGTTTT 3320
   Ccg8mrna  AAGTCTCGGG ACATTAACAG TTATATACCA CTTTTCTTTT CGTTTGTTTT 797
Ccg8ctg1356  AAGTCTCGGG ACATTAACAG TTATATACCA CTTTTCTTTT CGTTTGTTTT 794

   Ccg8dna2  ATTAGTTGAT CATATTTGGG GTGCCGACAA GTTCTGGCGT CTGGGAGTAC 3370
   Ccg8mrna  ATTAGTTGAT CATATTTGGG GTGCCGACAA GTTCTGGCGT CTGGGAGTAC 847
Ccg8ctg1356  ATTAGTTGAT CATATTTGGG GTGCCGACAA GTTCTGGCGT CTGGGAGTAC 844

   Ccg8dna2  TTTCCACACC ACATCACCCT TACTGGGACG GCGA.TTTGG TCTATGTACA 3419
   Ccg8mrna  TTTCCACACC ACATCACCCT TACTGGGACG GCGA.TTTGG TCTATGTACA 896
Ccg8ctg1356  TTTCCACACC ACATCACCCT TACTGGGACG GCGATTTTGG TCTATGTACA 894

   Ccg8dna2  TGGTTCATTT GTCCAAATTA CCACTTCTGC ATGTTTGTCG ACGTTTCCAT 3469
   Ccg8mrna  TGGTTCATTT GTCCAAATTA CCACTTCTGC ATGTTTGTCG ACGTTTCCAT 946
Ccg8ctg1356  TGGTTCATTT GTCCAAATTA CCACTTCTGC ATGTTTGTCG ACGTTTCCAT 944

   Ccg8dna2  TCTCGATCTT GGTCAATTCA AAAGCAGGCA GCTGAAATTG AAACTATCAC 3519
   Ccg8mrna  TCTCGATCTT GGTCAATTCA AAAGCAGGCA GCTGAAATTG AAACTATCAC 996
Ccg8ctg1356  TTTCGATCTT GGTCAATTCA AAAGCAGGCA GCTGAAATTG AAACTATCAC 994

   Ccg8dna2  GGATCCAAAA AAAAAACC-- ----------
   Ccg8mrna  GGATCCA--- ---------- ----------
Ccg8ctg1356  GGATCCAAAA AAAAAAAAAA AAAAAAAAA
```

Figure 3.38 The DNA sequence comparison among the *ccg-8* gene (AF088907), the *ccg-8* mRNA (U46087) and the contig1356. The putative translation start sites on the *ccg-8* mRNA and contig1356 and the stop codon were bolded and underlined. The numbers on the left side indicates the positions of bases in the sequences.

```
            1                                                  50
   Ccg8maa  ---------- ---------- ---------- ---------- ----------
```

```
        Ccg8aa  MTSRPVNGTG GEARFSEAPT ADASRRTSVS TIETLPAYDD QRSPAYSETV
Ccg8ctg1356pep  ---------- ---------- ---------- ---------- ----------

                51                                                  100
       Ccg8maa  ---------- ---------- ---------- ---------- ----------
        Ccg8aa  EQNGQAVDPK SALHAQLGNG RVQVTTSSLK ETMKEESLRS LKYVLETLRD
Ccg8ctg1356pep  ---------- ---------- ---------- ---------- ----------
                101                                                 150
       Ccg8maa  ---------- ---------- ---------- --MTNGSAPE EDRTRLVERM
        Ccg8aa  VTNTLQQGTN ELSSVLDQYD RSTTTRDGED HVMTNGSAPE EDRTRLVERM
Ccg8ctg1356pep  ----------N ELSSVLDQYD RSTTTRDGED HVMTNGSAPE EDRTRLVERM

                151                                                 200
       Ccg8maa  TELRKSIYSN IKDVSSVVAN YTGAALPENA GNLVRHHLLS LPMVWSQASK
        Ccg8aa  TELRKSIYSN IKDVSSVVAN YTGAALPENA GNLVRHHLLS LPMVWSQASK
Ccg8ctg1356pep  TELRKSIYSN IKDVSSVVAN YTGAALPENA GNLVRHHLLS LPMVWSQASK

                201                                                 250
       Ccg8maa  TTTSSEQPNT QDPNALVRKA AKVALSFSKE GLHIFSQIME IIRPATDHAE
        Ccg8aa  TTTSSEQPNT QDPNALVRKA AKVALSFSKE GLHIFSQIME IIRPATDHAE
Ccg8ctg1356pep  TTTSSEQPNT QDPNALVRKA AKVALSFSKE GLHIFSQIME IIRTATDHAE

                251                  274
       Ccg8maa  DWRNKKMNQM TPANGTEQEI RPLI
        Ccg8aa  DWRNKKMNQM TPANGTEQEI RPLI
Ccg8ctg1356pep  DWRNKKMNQM TPANGTEQEI LPLI
```

Figure3.39 The sequence comparison of the *ccg-8* gene protein (Q01306), the translated protein(U46087)of the *ccg-8* mRNA and the contig1356.

3.2.7 The clock-controlled gene 9

The *ccg-9* gene also is a *Neurospora crassa* morning specific clock-controlled gene. Although its identity and its function in cells are not known, based on homology to a GenBank entry, *ccg-9* may be a trehalose synthase. The ratio of the redundancy of these ESTs in the NM and NE cDNA libraries was ~2.5 (table 3.36). A gap exists in the cDNA sequence between contig1358 and contig1332 and the 5' ESTs and 3' ESTs of more than 10 cDNA clones occurred in both in contigs 1358 and 1332.

Table 3.36 The contigs and the EST totals of the *ccg-9* in combined EST database

NMNE contig#	EST counts		alignment		% homolog
	NM	NE	DNA	protein	

Contig1358 (5')	30	7	2-502,492-830,	3-169, 166-278	99%-100%
			887-1105	339-351	100%
Contig1332 (3')	31	0	606-94	561-731	99%
Contig1049 (3')	0	12	646-119	556-731	99%
Contig739 (3')	5	0	391-131	644-731	90%
Contig529 (5')	0	3	1-459	408-560	100%
	66	22	NM/NE=66/22*1.2=2.5		

The direct sequencing method using primer walking was performed to close the gap between the contig1358 and the contig1332. Primers were selected in the regions flanked the gap and prepared with the same method described in the previous part of this dissertation. The cDNA clones cover the gap were used as templates for the reactions.

Like the strategy used for the analysis of the *ccg-4* gene in this research, a separate directory called ccg9_dir was established to contain an edit_dir, chromat_dir and phd_dir. All the ESTs consisting of above 5 contigs were copied to the chromat_dir of the ccg9_dir. Since the gap region is large, primer walking was performed three times. Every time, the new primers flanked the gap region were prepared and used for the sequencing reactions. After all the final primer walking product sequences assembled into the database, the gap was covered when Phrap was run again. All the ESTs from above five contigs and the primer walking products went to the newccg9contig2 except three 3' ESTs went to a separate small cluster, the newccg9contig1. The consensus sequences of these two contigs were used to BlastX against the nr protein database of GenBank. The result was presented in Figure3.40. The newccg9contig1 hits the 3' of the ccg-9 gene protein and the protein of the ccg-9 mRNA. The newccg9contig2 hits the ccg-9 gene protein from amino acid from the residue 3 to 731 but still has a frame shift around the

base pair 492 in the sequence of the newccg9contig2. A stop codon at the 493 bp follows the residue methionine (M) at the 490 bp on the peptide of the translation with frame +2.

```
Newccg9ctg1:
>gi|3746895|gb|AAC64285.1| (AF088906) clock-controlled gene-9 protein [Neurospora crassa]
          Length = 731

 Score =  163 bits (412), Expect = 2e-39
 Identities = 77/78 (98%), Positives = 77/78 (98%)
 Frame = -1

Query: 364 LAWFYLAAKWTEVGVETSGKGGLKGNEQWVNDMARTEAGYLYTQEENRLPRHFTQRKPES 185
           LAWFYLAAKWTEVGVETSGKGGLKGNEQWVNDMARTEAGYLYTQEENRLPRHFTQRKPES
Sbjct: 654 LAWFYLAAKWTEVGVETSGKGGLKGNEQWVNDMARTEAGYLYTQEENRLPRHFTQRKPES 713

Query: 184 ESESKDLPIHEKKAEVTA 131
           ESESKDLPIHEKK EVTA
Sbjct: 714 ESESKDLPIHEKKPEVTA 731

>gi|11359598|pir||T47213 probable trehalose synthase (EC 2.4.1.-) [imported] - Neurospora
 crassa    >gi|1184788|gb|AAA98472.1| (U46088) similar to sucrose-phosphate synthase;
Method: conceptual translation supplied by author [Neurospora crassa]
          Length = 384

 Score =  155 bits (392), Expect = 4e-37
 Identities = 74/78 (94%), Positives = 74/78 (94%)
 Frame = -1

Query: 364 LAWFYLAAKWTEVGVETSGKGGLKGNEQWVNDMARTEAGYLYTQEENRLPRHFTQRKPES 185
           LAWFYLAAKWT VGVETSGKGGLKGNEQWVNDMA TEAGYL TQEENRLPRHFTQRKPES
Sbjct: 307 LAWFYLAAKWTXVGVETSGKGGLKGNEQWVNDMAXTEAGYLXTQEENRLPRHFTQRKPES 366

Query: 184 ESESKDLPIHEKKAEVTA 131
           ESESKDLPIHEKK EVTA
Sbjct: 367 ESESKDLPIHEKKPEVTA 384


newccg9ctg2:

>gi|3746895|gb|AAC64285.1| (AF088906) clock-controlled gene-9 protein [Neurospora crassa]
          Length = 731

Score =  335 bits (860), Expect(2) = 0.0
 Identities = 166/167 (99%), Positives = 166/167 (99%)
 Frame = +2

Query: 2   FEKARKFSTGTSVHRKRQMSTLVEKEGHFGPALTTLYLGISAVFADDHTAVVALAIHDTV 181
           FEKARKFSTGTSVHRKRQMSTLVEKEGHFGPALTTLYLGISAVFADDHTAVVALAIHDTV
Sbjct: 3   FEKARKFSTGTSVHRKRQMSTLVEKEGHFGPALTTLYLGISAVFADDHTAVVALAIHDTV 62

Query: 182 YLVDFSVKHIELDDALKMGEDLIAEYVISEVQKYEHENFSKFVGAGLPTTLKYMSPTLCS 361
           YLVDFSVKHIELDDALKMGEDLIAEYVISEVQKYEHENFSKFVGAGLPTTLKYMSPTLCS
Sbjct: 63  YLVDFSVKHIELDDALKMGEDLIAEYVISEVQKYEHENFSKFVGAGLPTTLKYMSPTLCS 122

Query: 362 RLWLEVDIVPIVMRPDDEHKEATFWDVKRVDEQADSMARKCIM*GPP 502
           RLWLEVDIVPIVMRPDDEHKEATFWDVKRVDEQADSMARKCIM GPP
Sbjct: 123 RLWLEVDIVPIVMRPDDEHKEATFWDVKRVDEQADSMARKCIMKGPP 169


 Score = 1165 bits (3013), Expect(2) = 0.0
 Identities = 565/566 (99%), Positives = 565/566 (99%)
 Frame = +3

Query: 492  KGPPCPVSSALKSILTALVRHFGPSLVPLLQVGFRGIVQTDAGFRAHLTTVQNHKDTCGP 671
```

```
            KGPPCPVSSALKSILTALVRHFGPSLVPLLQVGFRGIVQTDAGFRAHLTTVQNHKDTCGP
Sbjct: 166  KGPPCPVSSALKSILTALVRHFGPSLVPLLQVGFRGIVQTDAGFRAHLTTVQNHKDTCGP 225

Query: 672  ATWETTLTFAKKLRANKLKMAFFSSTPQGGGVALMRHALVRFARLLGVDLTWYVPKPRPG 851
            ATWETTLTFAKKLRANKLKMAFFSSTPQGGGVALMRHALVRFARLLGVDLTWYVPKPRPG
Sbjct: 226  ATWETTLTFAKKLRANKLKMAFFSSTPQGGGVALMRHALVRFARLLGVDLTWYVPKPRPG 285

Query: 852  VFRITKNIHNILQGVSHPDQRVSAEEKQAIIDWINENASRYWFSEGGPLRAPEEGGADIV 1031
            VFRITKNIHNILQGVSHPDQRVSAEEKQAIIDWINENASRYWFSEGGPLRAPEEGGADIV
Sbjct: 286  VFRITKNIHNILQGVSHPDQRVSAEEKQAIIDWINENASRYWFSEGGPLRAPEEGGADIV 345

Query: 1032 VIDDPQMPGLIPLIKKYTPNRPVLYRSHIQIRSDLVAKAGSPQADIWDFLWGNIQGADMF 1211
            VIDDPQMPGLIPLIKKYTPNRPVLYRSHIQIRSDLVAKAGSPQADIWDFLWGNIQGADMF
Sbjct: 346  VIDDPQMPGLIPLIKKYTPNRPVLYRSHIQIRSDLVAKAGSPQADIWDFLWGNIQGADMF 405

Query: 1212 ISHPIPSFVPHNVPREKVVYLPATTDWLDGLNKHLNHWDSGYYGNLYNNACHSQRMTELN 1391
            ISHPIPSFVPHNVPREKVVYLPATTDWLDGLNKHLNHWDSGYYGNLYNNACHSQRMTELN
Sbjct: 406  ISHPIPSFVPHNVPREKVVYLPATTDWLDGLNKHLNHWDSGYYGNLYNNACHSQRMTELN 465

Query: 1392 WPARKYIIQVARFDPSKGIPTVIDSYAEFRRRCDKAGITDVPQLVVCGNGSVDDPDASLI 1571
            WPARKYIIQVARFDPSKGIPTVIDSYAEFRRRCDKAGITDVPQLVVCGNGSVDDPDASLI
Sbjct: 466  WPARKYIIQVARFDPSKGIPTVIDSYAEFRRRCDKAGITDVPQLVVCGNGSVDDPDASLI 525

Query: 1572 YDQTMAQLETYYPDLIRDVSVMRLEPNDQVINTLLSNAHVALQLSTREGFEVKVSEALHA 1751
            YDQTMAQLETYYPDLIRDVSVMRLEPNDQVINTLLSNAHVALQLSTREGFEVKVSEALHA
Sbjct: 526  YDQTMAQLETYYPDLIRDVSVMRLEPNDQVINTLLSNAHVALQLSTREGFEVKVSEALHA 585

Query: 1752 GRPVIVTNVGGIPLQVKDKVNGFLVAPGDWRAVAGHLMDLFTDDELWKRMHHAARTGVSD 1931
            GRPVIVTNVGGIPLQVKDKVNGFLVAPGDWRAVAGHLMDLFTDDELWKRMHHAARTGVSD
Sbjct: 586  GRPVIVTNVGGIPLQVKDKVNGFLVAPGDWRAVAGHLMDLFTDDELWKRMHHAARTGVSD 645

Query: 1932 EVGTVGNALAWFYLAAKWTEVGVETSGKGGLKGNEQWVNDMARTEAGYLYTQEENRLPRH 2111
            EVGTVGNALAWFYLAAKWTEVGVETSGKGGLKGNEQWVNDMARTEAGYLYTQEENRLPRH
Sbjct: 646  EVGTVGNALAWFYLAAKWTEVGVETSGKGGLKGNEQWVNDMARTEAGYLYTQEENRLPRH 705

Query: 2112 FTQRKPESESESKDLPIHEKKAEVTA 2189
            FTQRKPESESESKDLPIHEKK EVTA
Sbjct: 706  FTQRKPESESESKDLPIHEKKPEVTA 731

>gi|11359598|pir||T47213 probable trehalose synthase (EC 2.4.1.-) [imported] - Neurospora
            crassa
 gi|1184788|gb|AAA98472.1|  (U46088)  similar  to  sucrose-phosphate  synthase;  Method:
conceptual translation supplied by author [Neurospora crassa]
          Length = 384

 Score =  776 bits (2005), Expect = 0.0
 Identities = 375/379 (98%), Positives = 375/379 (98%)
 Frame = +3

Query: 1053 PGLIPLIKKYTPNRPVLYRSHIQIRSDLVAKAGSPQADIWDFLWGNIQGADMFISHPIPS 1232
            PGLIPLIKKYTPNRPVLYRSHIQIRSDLVAKAGSPQADIWDFLWGNIQGADMFISHPIPS
Sbjct: 6    PGLIPLIKKYTPNRPVLYRSHIQIRSDLVAKAGSPQADIWDFLWGNIQGADMFISHPIPS 65

Query: 1233 FVPHNVPREKVVYLPATTDWLDGLNKHLNHWDSGYYGNLYNNACHSQRMTELNWPARKYI 1412
            FVPHNVPREKVVYLPATTDWLDGLNKHLNHWDSGYYGNLYNNACHSQRMTELNWPARKYI
Sbjct: 66   FVPHNVPREKVVYLPATTDWLDGLNKHLNHWDSGYYGNLYNNACHSQRMTELNWPARKYI 125

Query: 1413 IQVARFDPSKGIPTVIDSYAEFRRRCDKAGITDVPQLVVCGNGSVDDPDASLIYDQTMAQ 1592
            IQVARFDPSKGIPTVIDSYAEFRRRCDKAGITDVPQLVVCGNGSVDDPDASLIYDQTMAQ
Sbjct: 126  IQVARFDPSKGIPTVIDSYAEFRRRCDKAGITDVPQLVVCGNGSVDDPDASLIYDQTMAQ 185

Query: 1593 LETYYPDLIRDVSVMRLEPNDQVINTLLSNAHVALQLSTREGFEVKVSEALHAGRPVIVT 1772
            LETYYPDLIRDVSVMRLEPNDQVINTLLSNAHVALQLSTREGFEVKVSEALHAGRPVIVT
Sbjct: 186  LETYYPDLIRDVSVMRLEPNDQVINTLLSNAHVALQLSTREGFEVKVSEALHAGRPVIVT 245

Query: 1773 NVGGIPLQVKDKVNGFLVAPGDWRAVAGHLMDLFTDDELWKRMHHAARTGVSDEVGTVGN 1952
            NVGGIPLQVKDKVNGFLVAPGDWRAVAGHLMDLFTDDELWKRMHHAARTGVSDEVGTVGN
Sbjct: 246  NVGGIPLQVKDKVNGFLVAPGDWRAVAGHLMDLFTDDELWKRMHHAARTGVSDEVGTVGN 305
```

```
Query: 1953 ALAWFYLAAKWTEVGVETSGKGGLKGNEQWVNDMARTEAGYLYTQEENRLPRHFTQRKPE 2132
          ALAWFYLAAKWT VGVETSGKGGLKGNEQWVNDMA TEAGYL TQEENRLPRHFTQRKPE
Sbjct: 306  ALAWFYLAAKWTXVGVETSGKGGLKGNEQWVNDMAXTEAGYLXTQEENRLPRHFTQRKPE 365

Query: 2133 SESESKDLPIHEKKAEVTA 2189
          SESESKDLPIHEKK EVTA
Sbjct: 366  SESESKDLPIHEKKPEVTA 384
```

Figure 3.40 The BlastX results for two new ccg9 contigs. The frame shift was showed in this Figure.

The DNA sequence of newccg9contig2 was compared with both the genomic DNA sequence of the *ccg-9* gene (AF088906) and the sequence of the *ccg-9* mRNA (U46088) (Figure3.41). The sequence of the newccg9contig2 is 99% similar to that of the *ccg-9* gene genomic DNA sequence except for two introns in the *ccg-9* gene sequence. It aligned with the *ccg-9* gene from 2 bp to 2189 bp and aligned with the *ccg-9* mRNA from 81 bp to 2189 bp (Figure 3.42). The sequence of the newccg9contig2 is also 99% similar to that of the *ccg-9* mRNA. When the sequences around the 491bp "T" of the newccg9contig2 were examined, a base "T" at the position 491bp was presented in both sequences of the *ccg-9* mRNA and the newccg9contig2 but missed in the sequence of the *ccg-9* gene sequences. Therefore, the sequence of the *ccg-9* gene at this region became $_{1139}$ATG AAG GGC CCC CCT TGT CCT$_{1159}$. Then, the translation continues after here in the sequence of the *ccg-9* gene and yield a 731 residue product. However, after check

---

newccg9contig2: $_{488}$ATG **TAA** GGG CCC CCC TTG TCC T$_{509}$
M stop*

ccg-9 gene: $_{1139}$ATG AAG GGC CCC CCT TGT CCT$_{1159}$....
M K G P P C P....

---

the trace files of the sequences of the newccg9contig2 which made of more than 80 high

quality ESTs, the base "T" indeed was presented at this position (Figure3.43). Therefore, the reported 731 residues of the *ccg-9* gene product are incorrect and the correct *ccg-9* gene product has only 165 amino acid residues.

```
   Ccg9mrna  ---------- ---------- ---------- ---------- ----------
Newccg9ctg2  ---------- ---------- ---------- ---------- ----------
    Ccg9dna  GCGACTTCCG AATACTTGAT CAAAAAGTGG AAGTAACACT GCAGAACCCG 50

   Ccg9mrna  ---------- ---------- ---------- ---------- ----------
Newccg9ctg2  ---------- ---------- ---------- ---------- ----------
    Ccg9dna  CACACCACCC ACCCAAGGCC ACTACAAGTA CTTGCCGTGC AACGTGAAGC 100

   Ccg9mrna  ---------- ---------- ---------- ---------- ----------
Newccg9ctg2  ---------- ---------- ---------- ---------- ----------
    Ccg9dna  CATCAATTGG ATGCAATGGT TGATAGGGGG GAACCCAAGC TGNCCGAGAC 150

   Ccg9mrna  ---------- ---------- ---------- ---------- ----------
Newccg9ctg2  ---------- ---------- ---------- ---------- ----------
    Ccg9dna  CGACGAGTCC GCGAGCGCCG GCATCGCCTT CAAAAACCCA TTCGCAGATA 200

   Ccg9mrna  ---------- ---------- ---------- ---------- ----------
Newccg9ctg2  ---------- ---------- ---------- ---------- ----------
    Ccg9dna  CCACCCCACA CCACCCGGCC TCGAATGACA ACGACTCGTC CCGCTCCGAC 250

   Ccg9mrna  ---------- ---------- ---------- ---------- ----------
Newccg9ctg2  ---------- ---------- ---------- ---------- ----------
    Ccg9dna  ACAAAAAGCC CTACAAGTCC GTTGAATGAT ACTTCCGGGC CCTACAGCTC 300

   Ccg9mrna  ---------- ---------- ---------- ---------- ----------
Newccg9ctg2  ---------- ---------- ---------- ---------- ----------
    Ccg9dna  GCGTCGAATT ACCCCTTCGC TTGGCGTCCA GCAGCTTCTC AGTAATCTTC 350

   Ccg9mrna  ---------- ---------- ---------- ---------- ----------
Newccg9ctg2  ---------- ---------- ---------- ---------- ----------
    Ccg9dna  AGCCCAGAGA TCCATTGGCC GACCAGACGC CTGCCCCCAC CAGCGACGCT 400

   Ccg9mrna  ---------- ---------- ---------- ---------- ----------
Newccg9ctg2  ---------- ---------- ---------- ---------- ----------
    Ccg9dna  GTTGCGTATA ATTCCAGACC TTCCCTTCCA CATCAGCTCC AGAAGCAACA 450

   Ccg9mrna  ---------- ---------- ---------- ---------- ----------
Newccg9ctg2  ---------- ---------- ---------- ---------- ----------
    Ccg9dna  ACACCTCGAC CACCTCCAGC CAACACCGGC CACTACACGT CTCCTAGCGG 500

   Ccg9mrna  ---------- ---------- ---------- ---------- ----------
Newccg9ctg2  ---------- ---------- ---------- ---------- ----------
    Ccg9dna  CGACAACACC CAGACCTACG ATACCCGGTC GGCTGCCTCC CATCATCGCA 550

   Ccg9mrna  ---------- ---------- ---------- ---------- ----------
Newccg9ctg2  ---------- ---------- ---------- -GTTTGAAAA GGCCCGCAAG 19
    Ccg9dna  GCATCAAGCC CATCGCAACC GCAGTCATGG CGTTTGAAAA GGCCCGCAAG 600

   Ccg9mrna  ---------- ---------- ---------- ---------- ----------
Newccg9ctg2  TTCTCCACGG GCACTTCGGT GCACCGTAAG CGCCAGATGA GCACTCTGGT 69
    Ccg9dna  TTCTCCACGG GCACTTCGGT GCACCGTAAG CGCCAGATGA GCACTCTGGT 650

   Ccg9mrna  ---------- -GTCACTTTG GTCCTGCTCT AACT...... .......... 23   Intron 1
Newccg9ctg2  GGAGAAAGAG GGTCACTTTG GTCCTGCTCT AACT...... .......... 103
    Ccg9dna  GGAGAAAGAG GGTCACTTTG GTCCTGCTCT AACTGTGAGT ACCTTGCCCA 700

   Ccg9mrna  .......... .......... .......... .......... ..........
Newccg9ctg2  .......... .......... .......... .......... ..........
```

```
    Ccg9dna  CCACGGCTTG CTTGACTGCG TCGACGTTTC ATGCTAATAT TGGATGTCCC 750

   Ccg9mrna  ....ACGCTC TATCTTGGTA TCTCTGCCGT CTTTGCTGAC GACCATACGG 69
Newccg9ctg2  ....ACGCTC TATCTTGGTA TCTCTGCCGT CTTTGCTGAC GACCATACGG 149
    Ccg9dna  ATAGACGCTC TATCTTGGTA TCTCTGCCGT CTTTGCTGAC GACCATACGG 800

   Ccg9mrna  CCGTTGTGGC CCTGGCCATT CATGACACAG TCTACCTGGT GGACTTCTCA 119
Newccg9ctg2  CCGTTGTGGC CCTGGCCATT CATGACACAG TCTACCTGGT GGACTTCTCA 199
    Ccg9dna  CCGTTGTGGC CCTGGCCATT CATGACACAG TCTACCTGGT GGACTTCTCA 850

   Ccg9mrna  GTTAAACACA TTGAGCTCGA TGACGCCCTG AAAATGGGCG AGGACCTCAT 169
Newccg9ctg2  GTTAAACACA TTGAGCTCGA TGACGCCCTG AAAATGGGCG AGGACCTCAT 249
    Ccg9dna  GTTAAACACA TTGAGCTCGA TGACGCCCTG AAAATGGGCG AGGACCTCAT 900

   Ccg9mrna  CGCAGAATAT GTCATCTCCG AAGTGCAAAA GTATGAGCAC GAGAACTTTT 219
Newccg9ctg2  CGCAGAATAT GTCATCTCCG AAGTGCAAAA GTATGAGCAC GAGAACTTTT 299
    Ccg9dna  CGCAGAATAT GTCATCTCCG AAGTGCAAAA GTATGAGCAC GAGAACTTTT 950

   Ccg9mrna  CCAAGTTCGT CGGTGCTGGT CTCCCCACGA CACTCAAGTA CATGAGCCCA 269
Newccg9ctg2  CCAAGTTCGT CGGTGCTGGT CTCCCCACGA CACTCAAGTA CATGAGCCCA 349
    Ccg9dna  CCAAGTTCGT CGGTGCTGGT CTCCCCACGA CACTCAAGTA CATGAGCCCA 1000

   Ccg9mrna  ACCCTCTGTT CGCGGCTCTG GCTCGAGGTG GATATTGTGC CCATCGTTAT 319
Newccg9ctg2  ACCCTCTGTT CGCGGCTCTG GCTCGAGGTG GATATTGTGC CCATCGTTAT 399
    Ccg9dna  ACCCTCTGTT CGCGGCTCTG GCTCGAGGTG GATATTGTGC CCATCGTTAT 1050

   Ccg9mrna  GCGCCCAGAT GATGAGCACA AGGAAGCGAC CTTCTGGGAT GTCAAGCGAG 369
Newccg9ctg2  GCGCCCAGAT GATGAGCACA AGGAAGCGAC CTTCTGGGAT GTCAAGCGAG 449
    Ccg9dna  GCGCCCAGAT GATGAGCACA AGGAAGCGAC CTTCTGGGAT GTCAAGCGAG 1100

   Ccg9mrna  TCGATGAACA AGCCGATTCC ATGGCCCGAA AGTGCATCAT GTAAGGGCCC 419
Newccg9ctg2  TCGATGAACA AGCCGATTCC ATGGCCCGAA AGTGCATCAT GTAAGGGCCC 499
    Ccg9dna  TCGATGAACA AGCCGATTCC ATGGCCCGAA AGTGCATCAT G.AAGGGCCC 1149

   Ccg9mrna  CCCTTGTCCT GTCAGTTCCG CCCTGAAGAG CATACTGACA GCCCTTGTTA 469
Newccg9ctg2  CCCTTGTCCT GTCAGTTCCG CCCTGAAGAG CATACTGACA GCCCTTGTTA 549
    Ccg9dna  CCCTTGTCCT GTCAGTTCCG CCCTGAAGAG CATACTGACA GCCCTTGTTA 1199

   Ccg9mrna  GGCACTTTGG NCCTTCCCTC GTTCCCCTCC TCCAGGTCGG ATTCCGAGGC 519
Newccg9ctg2  GGCACTTTGG CCCTTCCCTC GTTCCCCTCC TCCAGGTCGG ATTCCGAGGC 599
    Ccg9dna  GGCACTTTGG NCCTTCCCTC GTTCCCCTCC TCCAGGTCGG ATTCCGAGGC 1249

   Ccg9mrna  ATTGTTCAAA CCGATGCTGG ATTCCGCGCC CACCTTACCA CCGTACAGAA 569
Newccg9ctg2  ATTGTTCAAA CCGATGCTGG ATTCCGCGCC CACCTTACCA CCGTACAGAA 649
    Ccg9dna  ATTGTTCAAA CCGATGCTGG ATTCCGCGCC CACCTTACCA CCGTACAGAA 1299

   Ccg9mrna  CCACAAAGAC ACATGCGGTC CGGCAACGTG GGAAACCACT CTGACATTTG 619
Newccg9ctg2  CCACAAAGAC ACATGCGGTC CGGCAACGTG GGAAACCACT CTGACATTTG 699
    Ccg9dna  CCACAAAGAC ACATGCGGTC CGGCAACGTG GGAAACCACT CTGACATTTG 1349

   Ccg9mrna  CCAAGAAGCT TCGTGCCAAC AAGCTCAAGA TGGCCTTCTT CAGCTCAACG 669
Newccg9ctg2  CCAAGAAGCT TCGTGCCAAC AAGCTCAAGA TGGCCTTCTT CAGCTCAACG 749
    Ccg9dna  CCAAGAAGCT TCGTGCCAAC AAGCTCAAGA TGGCCTTCTT CAGCTCAACG 1399

   Ccg9mrna  CCCCAAGGCG GTGGTGTTGC ACTTATGCGG CACGCTCTTG TACGCTTTGC 719
Newccg9ctg2  CCCCAAGGCG GTGGTGTTGC ACTTATGCGG CACGCTCTTG TACGCTTTGC 799
    Ccg9dna  CCCCAAGGCG GTGGTGTTGC ACTTATGCGG CACGCTCTTG TACGCTTTGC 1449

   Ccg9mrna  TCGTCTTTTG GGTGTCGACC TCACCTGGTA C..... ..... 750   Intron2
Newccg9ctg2  TCGTCTTTTG GGTGTCGACC TCACCTGGTA C..... ..... 830
    Ccg9dna  TCGTCTTTTG GGTGTCGACC TCACCTGGTA CGGTAAGTAA AGCAGTTTGC 1499

   Ccg9mrna  ..... ..... ..... .....GTG CCAAAACCAC 763
Newccg9ctg2  ..... ..... ..... .....GTG CCAAAACCAC 843
    Ccg9dna  TGTCCGGAAC GCTAGACGGT CAACTAACAC TGTTTCAGTG CCAAAACCAC 1499

   Ccg9mrna  GACCCGGTGT CTTCCGCATC ACCAAAAACA TCCACAACAT CCTGCAGGGT 813
```

```
Newccg9ctg2  GACCCGGTGT CTTCCGCATC ACCAAAAACA TCCACAACAT CCTGCAGGGT 893
    Ccg9dna  GACCCGGTGT CTTCCGCATC ACCAAAAACA TCCACAACAT CCTGCAGGGT 1599

   Ccg9mrna  GTCAGCCATC CCGATCAGCG CGTATCTGCC GAAGAGAAGC AGGCCATCAT 863
Newccg9ctg2  GTCAGCCATC CCGATCAGCG CGTATCTGCC GAAGAGAAGC AGGCCATCAT 943
    Ccg9dna  GTCAGCCATC CCGATCAGCG CGTATCTGCC GAAGAGAAGC AGGCCATCAT 1649

   Ccg9mrna  TGATTGGATT AACGAGAATG CCAGTCGATA CTGGTTCTCC GAGGGCGGAC 913
Newccg9ctg2  TGATTGGATT AACGAGAATG CCAGTCGATA CTGGTTCTCC GAGGGCGGAC 993
    Ccg9dna  TGATTGGATT AACGAGAATG CCAGTCGATA CTGGTTCTCC GAGGGCGGAC 1699

   Ccg9mrna  CTTTGCGTGC ACCCGAGGAA GGCGGTGCTG ATATTGTCGT GATTGATGAC 963
Newccg9ctg2  CTTTGCGTGC ACCCGAGGAA GGCGGTGCTG ATATTGTCGT GATTGATGAC 1043
    Ccg9dna  CTTTGCGTGC ACCCGAGGAA GGCGGTGCTG ATATTGTCGT GATTGATGAC 1749

   Ccg9mrna  CCTCAAATGT CCCGGCTTGA TCCCCCTGAT CAAAAAGTAT ACCCCTAACC 1013
Newccg9ctg2  CCTCAAATG. CCCGGCTTGA TCCCCCTGAT CAAAAAGTAT ACCCCTAACC 1092
    Ccg9dna  CCTCAAATG. CCCGGCTTGA TCCCCCTGAT CAAAAAGTAT ACCCCTAACC 1798

   Ccg9mrna  GCCCCGTTCT CTACCGCTCG CACATCCAGA TTCGCAGCGA CCTCGTCGCC 1063
Newccg9ctg2  GCCCCGTTCT CTACCGCTCG CACATCCAGA TTCGCAGCGA CCTCGTCGCC 1142
    Ccg9dna  GCCCCGTTCT CTACCGCTCG CACATCCAGA TTCGCAGCGA CCTCGTCGCC 1848

   Ccg9mrna  AAGGCCGGTT CGCCTCAAGC CGACATCTGG GATTTCCTGT GGGGCAACAT 1113
Newccg9ctg2  AAGGCCGGTT CGCCTCAAGC CGACATCTGG GATTTCCTGT GGGGCAACAT 1192
    Ccg9dna  AAGGCCGGTT CGCCTCAAGC CGACATCTGG GATTTCCTGT GGGGCAACAT 1898

   Ccg9mrna  TCAGGGCGCC GACATGTTTA TCAGCCACCC GATTCCCAGT TTCGTCCCTC 1163
Newccg9ctg2  TCAGGGCGCC GACATGTTTA TCAGCCACCC GATTCCCAGT TTCGTCCCTC 1242
    Ccg9dna  TCAGGGCGCC GACATGTTTA TCAGCCACCC GATTCCCAGT TTCGTCCCTC 1948

   Ccg9mrna  ACAACGTCCC GCGCGAGAAG GTCGTCTACT TGCCAGCCAC CACCGACTGG 1213
Newccg9ctg2  ACAACGTCCC GCGCGAGAAG GTCGTCTACT TGCCAGCCAC CACCGACTGG 1292
    Ccg9dna  ACAACGTCCC GCGCGAGAAG GTCGTCTACT TGCCAGCCAC CACCGACTGG 1998

   Ccg9mrna  CTTGACGGGT TAAACAAGCA CCTCAACCAC TGGGACTCGG GTTATTACGG 1263
Newccg9ctg2  CTTGACGGCT TAAACAAGCA CCTCAACCAC TGGGACTCGG GTTATTACGG 1342
    Ccg9dna  CTTGACGGGT TAAACAAGCA CCTCAACCAC TGGGACTCGG GTTATTACGG 2048

   Ccg9mrna  AAATCTGTAC AACAACGCCT GCCACTCGCA GCGCATGACG GAGCTCAACT 1313
Newccg9ctg2  AAATCTGTAC AACAACGCCT GCCACTCGCA GCGCATGACG GAGCTCAACT 1392
    Ccg9dna  AAATCTGTAC AACAACGCCT GCCACTCGCA GCGCATGACG GAGCTCAACT 2098

   Ccg9mrna  GGCCCGCCCG CAAGTACATC ATCCAGGTGG CCCGCTTCGA CCCATCCAAG 1363
Newccg9ctg2  GGCCCGCCCG CAAGTACATC ATCCAGGTGG CCCGCTTCGA CCCATCCAAG 1442
    Ccg9dna  GGCCCGCCCG CAAGTACATC ATCCAGGTGG CCCGCTTCGA CCCATCCAAG 2148

   Ccg9mrna  GGCATCCCCA CCGTTATCGA CTCGTATGCG GAGTTCCGTC GCCGCTGCGA 1413
Newccg9ctg2  GGCATCCCCA CCGTTATCGA CTCGTATGCG GAGTTCCGTC GCCGCTGCGA 1492
    Ccg9dna  GGCATCCCCA CCGTTATCGA CTCGTATGCG GAGTTCCGTC GCCGCTGCGA 2198

   Ccg9mrna  CAAAGCTGGC ATCACCGACG TGCCGCAGCT CGTTGTGTGC GGCAATGGTT 1463
Newccg9ctg2  CAAAGCTGGC ATCACCGACG TGCCGCAGCT CGTTGTGTGC GGCAATGGTT 1542
    Ccg9dna  CAAAGCTGGC ATCACCGACG TGCCGCAGCT CGTTGTGTGC GGCAATGGTT 2248

   Ccg9mrna  CCGTCGATGA CCCCGACGCC TCCCTCATCT ACGATCAGAC CATGGCCCAG 1513
Newccg9ctg2  CCGTCGATGA CCCCGACGCC TCCCTCATCT ACGATCAGAC CATGGCCCAG 1592
    Ccg9dna  CCGTCGATGA CCCCGACGCC TCCCTCATCT ACGATCAGAC CATGGCCCAG 2298

   Ccg9mrna  CTTGAGACAT ACTACCCCGA CTTGATTCGG GACGTCAGCG TCATGCGTTT 1563
Newccg9ctg2  CTTGAGACAT ACTACCCCGA CTTGATTCGG GACGTCAGCG TCATGCGTTT 1642
    Ccg9dna  CTTGAGACAT ACTACCCCGA CTTGATTCGG GACGTCAGCG TCATGCGTTT 2348

   Ccg9mrna  GGAGCCCAAC GACCAGGTCA TCAACACCTT GCTCTCCAAT GCCCACGTTG 1613
Newccg9ctg2  GGAGCCCAAC GACCAGGTCA TCAACACCTT GCTCTCCAAT GCCCACGTTG 1692
    Ccg9dna  GGAGCCCAAC GACCAGGTCA TCAACACCTT GCTCTCCAAT GCCCACGTTG 2398
```

```
   Ccg9mrna  CGTTGCAGCT TTCCACCCGC GAAGGCTTCG AGGTCAAGGT TTCCGAGGCC 1663
Newccg9ctg2  CGTTGCAGCT TTCCACCCGC GAAGGCTTCG AGGTCAAGGT TTCCGAGGCC 1742
    Ccg9dna  CGTTGCAGCT TTCCACCCGC GAAGGCTTCG AGGTCAAGGT TTCCGAGGCC 2448

   Ccg9mrna  TTGCACGCCG GCAGACCCGT CATCGTCACC AACGTTGGCG GTATCCCGCT 1713
Newccg9ctg2  TTGCACGCCG GCAGACCCGT CATCGTCACC AACGTTGGCG GTATCCCGCT 1792
    Ccg9dna  TTGCACGCCG GCAGACCCGT CATCGTCACC AACGTTGGCG GTATCCCGCT 2498

   Ccg9mrna  GCAGGTCAAG GACAAGGTCA ATGGTTTCCT TGTCGCCCCT GGTGATTGGC 1763
Newccg9ctg2  GCAGGTCAAG GACAAGGTCA ATGGTTTCCT TGTCGCCCCT GGTGATTGGC 1842
    Ccg9dna  GCAGGTCAAG GACAAGGTCA ATGGTTTCCT TGTCGCCCCT GGTGATTGGC 2548

   Ccg9mrna  GTGCCGTGGC CGGTCATCTG ATGGACCTGT TCACCGATGA CGAGCTCTGG 1813
Newccg9ctg2  GTGCCGTGGC CGGTCATCTG ATGGACCTGT TCACCGATGA CGAGCTCTGG 1892
    Ccg9dna  GTGCCGTGGC CGGTCATCTG ATGGACCTGT TCACCGATGA CGAGCTCTGG 2598

   Ccg9mrna  AAGAGAATGC ATCATGCGGC ACGGACGGGC GTCTCTGATG AGGTCGGCAC 1863
Newccg9ctg2  AAGAGAATGC ATCATGCGGC ACGGACGGGC GTCTCTGATG AGGTCGGCAC 1942
    Ccg9dna  AAGAGAATGC ATCATGCGGC ACGGACGGGC GTCTCTGATG AGGTCGGCAC 2648

   Ccg9mrna  TGTCGGTAAT GCGCTTGCCT GGTTCTACTT GGCTGCCAAG TGGACCNAAG 1913
Newccg9ctg2  TGTCGGTAAT GCGCTTGCCT GGTTCTACTT GGCTGCCAAG TGGACCGAAG 1992
    Ccg9dna  TGTCGGTAAT GCGCTTGCCT GGTTCTACTT GGCTGCCAAG TGGACCGAAG 2698

   Ccg9mrna  TCGGCGTAGA GACGAGTGGC AAGGGTGGAT TGAAGGGCAA TGAGCAGTGG 1963
Newccg9ctg2  TCGGCGTAGA GACGAGTGGC AAGGGTGGAT TGAAGGGCAA TGAGCAGTGG 2042
    Ccg9dna  TCGGCGTAGA GACGAGTGGC AAGGGTGGAT TGAAGGGCAA TGAGCAGTGG 2748

   Ccg9mrna  GTGAACGACA TGGCGANGAC GGAGGCCGGA TACTTGTNCA CCCAAGAGGA 2013
Newccg9ctg2  GTGAACGACA TGGCGAGGAC GGAGGCCGGA TACTTGTACA CCCAAGAGGA 2092
    Ccg9dna  GTGAACGACA TGGCGAGGAC GGAGGCCGGA TACTTGTACA CCCAAGAGGA 2798

   Ccg9mrna  AAACCGGTTG CCGAGACACT TTACGCAGAG GAAGCCGGAG TCCGAGTCGG 2063
Newccg9ctg2  AAACCGGTTG CCGAGACACT TTACGCAGAG GAAGCCGGAG TCCGAGTCGG 2142
    Ccg9dna  AAACCGGTTG CCGAGACACT TTACGCAGAG GAAGCCGGAG TCCGAGTCGG 2848

   Ccg9mrna  AGTCTAAGGA CTTGCCAATT CATGAGAAGA AGCCAGAGGT TACTGCTTGA 2113
Newccg9ctg2  AGTCTAAGGA CTTGCCAATT CATGAGAAGA AGGCAGAGGT TACTGCTTGA 2192
    Ccg9dna  AGTCTAAGGA CTTGCCAATT CATGAGAAGA AGCCAGAGGT TACTGCTTGA 2898

   Ccg9mrna  TTTTATGTTG GCGGATCGCT TGTTTTTGGT CGAGTTTAGT TTGTTAGTTT 2163
Newccg9ctg2  TTTTATGTTG GCGGATCGCT TGTTTTTGGT CGAGTTTAGT TTGTTAGTTT 2242
    Ccg9dna  TTTTATGTTG GCGGATCGCT TGTTTTTGGT CGAGTTTAGT TTGTTAGTTT 2948

   Ccg9mrna  AAGATATGAG CTCATCCATG GAGGGAAAGA AGCCAGAGGT 2203
Newccg9ctg2  AAGATATGAG CTCATCCATG GAGAAAAAAA AAAAAAAAAA 2282
    Ccg9dna  AAGATATGAG CTCATCCATG GAGGG----- ---------- 2973
```

Figure 3.41 The sequence comparison of the *ccg-9* gene genomic DNA (Af088906), *ccg-9* mRNA (u46088), and the newccg9contig2. 2 introns and the deletion of base "T" at the position 1142 in the sequence of the *ccg-9* genomic sequence and the addition of base "T" at the position of 973 in the sequence of the *ccg-9* mRNA were marked with the underlined larger fonts.

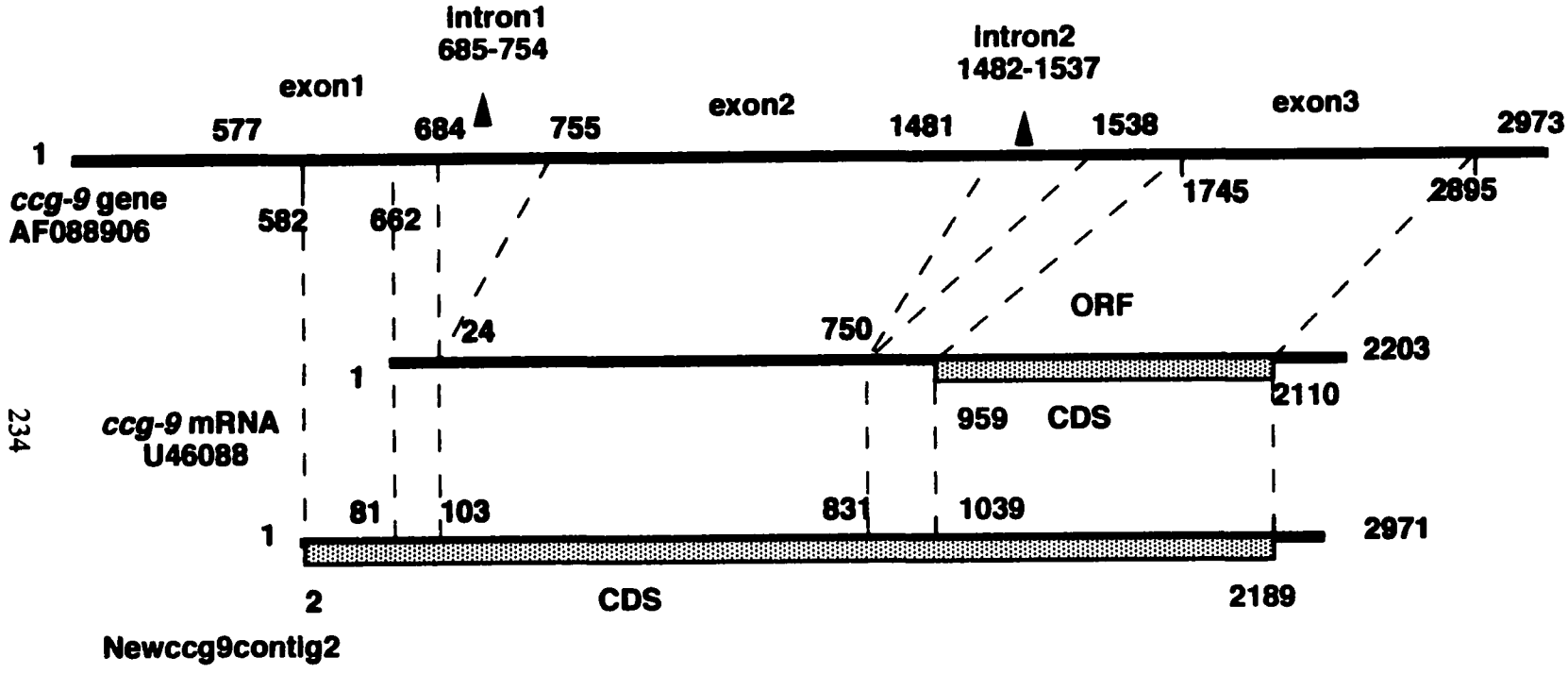


Figure 3.42 The sequence comparison of the *ccg-9* gene, the *ccg-9* mRNA and the newccg9contig2

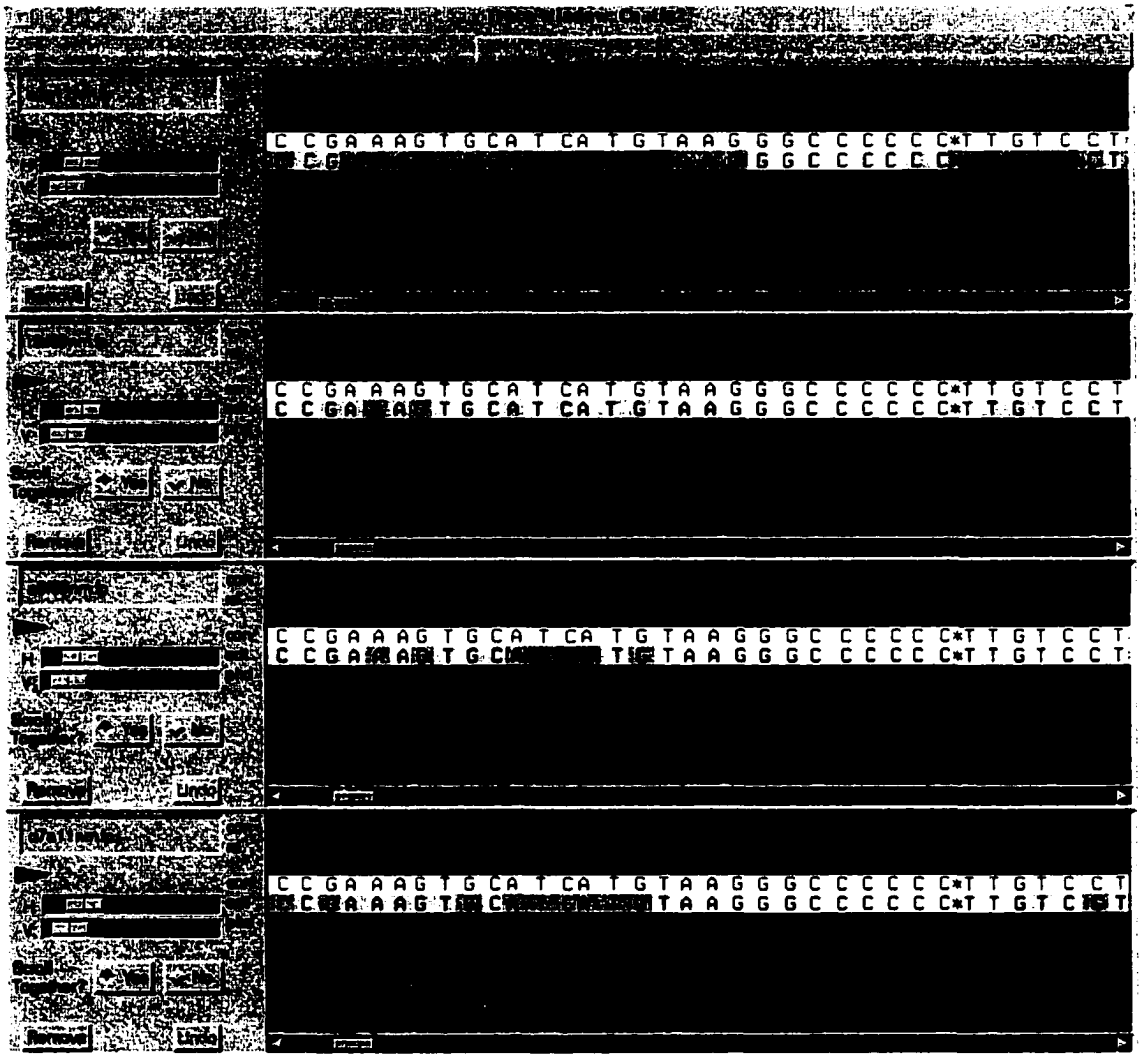


Figure3.43 The trace files in the newccg9contig2 represented the *ccg-9* gene. The base pair "T" was marked at the position 491 in the consensus sequence of the contig.

Both the newccg9contig1 and 2 were identical to the *Neurospora crassa* trehalose synthtase as revealed by a the BlastX search against the nr protein GenBank (Figure3.40) as both the ORF of the *ccg-9* mRNA (U46088) and this final newccg9contig2 sequences encode exactly the same 384 amino acid residues except for that the first five amino acids of the protein translated from the ORF of the *ccg-9* mRNA did not match with the

translated protein of the newccg9contig2 (Figure3.44) even though their DNA sequences have 99% similarity. From the sequence comparison in Figure3.41, an extra base "T" was found in the *ccg-9* mRNA sequence at the position 973 where it is at the position 1054 in

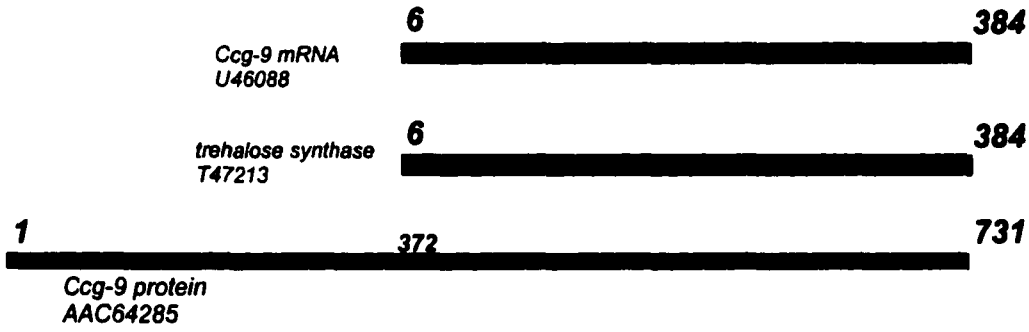


Figure 3.44 The sequence comparisons between the translation product(AAC64285) of *ccg-9* gene and the translated proteins of the *ccg-9* mRNA (U46088), contig1358, contig1332 and the possible *Neurospora* trehalose synthase (T47213).

the sequence of the newccg9contig2. This base "T" was absent in the sequences of both the *ccg-9* genomic DNA and the cDNA represented by newccg9contig2. When the trace files of the newccg9contig2 were carefully checked at position1054 in the sequence of the newccg9contig2, no "T" was present at this position (Figure3.45). The sequence around this position for the *ccg-9* mRNA (U46088) is: $_{959}$ATG ACC CTC AAA TGT CCC GGC TTG ATC$_{985}$...It is M T L K C P G L I, and the sequence of the newccg9conitg2, it is $_{1039}$ATG ACC CTC AAA TGC CCG GCT TGA TC$_{1064}$. It can only translate 7 amino acid residues. This is the reason why the match of the newccg9contig2 with the ccg-9 mRNA or the probable *Neurospora crassa* trehalose synthtase starts at the position $_{1053}$CCC GGC TTG ATC. This is the same situation as for the ccg-9 gene protein. Either the entry deposited in the GenaBank database for the protein (AAA98472) of the ccg-9 mRNA is incorrect because of a sequencing error.



Figure 3.45 The trace files in the newccg9contig2. No base "T" in front of base "C" at the position 1053 bp in its consensus sequence.

The sequence comparison among the two newccg9 contigs and the *ccg-9* mRNA were performed as well (Figire3.46). The cDNAs represented by both of two newccg9contigs indicate that the *ccg-9* mRNA is short but contains a complete CDS, and all three have different polyadenylation sites. Except the region before the poly A addition site, all the remainder of the sequences were similar and *ccg-9* most likely is a trehalose synthtase with alternative polyA addition sites.

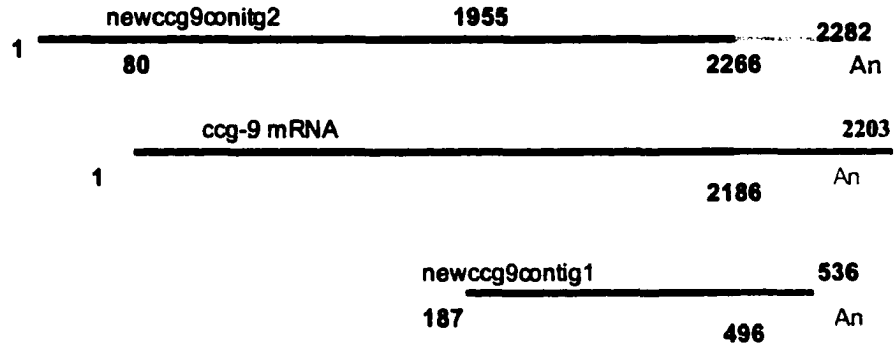


Figure 3.46 The relationship of the newccg9contig1, 2 and the *ccg-9* mRNA. The blue color represents the same DNA sequences and the different color indicates the different sequences in the polyadenylation region.

3.2.8 The identification of novel clock-controlled genes

The majority of this section was included in a paper that has appeared in the journal *Genetics* (zhu et al., 2001). The biochemical experiment and electronic transcriptional profile analysis were performed in Dr. Dunlap and Dr. Loros's lab, the Department of Biochemistry, Dartmouth Medical School.

In the identification of novel clock-controlled genes, 26 cDNA clones were used to probe a Northern blot. These 26 clones represent 26 different contigs (table3.37).

The RNA used in the Northern blots was prepared at different time points. The RNA was prepared according to the method as described before (Yarden et al., 1992). The analysis of the expression redundancy of the 26 selected cDNA clones was performed using standard Northern blots techniques (Sambrook et al.,1989). The DNA inserts from 26 cDNA clones were hybridized to RNA from both frq^{+} strain and frq^{7}

strain cultures. The intensity of the cDNA probe hybridization signal was examined using

Table 3.37 The 26 cDNA clones used for RNA hybridization (from Zhu et al., 2001)

Contig No#	Clone identity	Homology	Morning cDNAs	Evening cDNAs	Rhythmic expression
1314	a8d02ne	eIFA5	2	42	no
1339	a6h07ne		0	32	no
1378	c8a03ne	Histone H4	50	8	no
1382	c7h02ne	PEPCK	1	48	no
1387	c8f10ne	ADH1	1	74	no
1404	b7h07ne	Rib.prot. L6	49	13	no
1405	a9d08nm	Rib.prot./Ubiquitin	51	11	no
1411	a8d01ne	SPS2 homolog	140	30	yes *(ccg-15)*
1416	a1g07nm	Rib.prot. L14	86	4	no
1419	a2f01nm		222	38	no
1422	a8c05nm	HSP70	108	6	no
1423	b8d08ne	PGK	29	93	no
1427	a8a11nm		106	3	no
1428	a1f11nm	Ubiquitin	238	42	no
1432	a5f07nm	Phase sepc. protein	117	6	yes *(ccg-13)*
1434	d3f10nm	DnaJ	134	0	no
1442	a2f10nm	Lysozyme	202	38	yes *(lyz)*
1445	a3d04nm		108	181	no
1429c	a8g09nm	Thioredoxin	39	0	no
1334	a8d07nm	*RCO-3*	76	38	no
1415	a2a12nm		56	30	no
429	b8g08ne		0	3	no
814	a7h07ne		1	5	no
1401	a1d07nm		37	23	no
1412	a2h10nm	V-ATPase	38	41	no
1421	a3e01nm	Snodprot1	66	41	yes *(ccg-14)*

densitometric analysis (Loros, et al., 1989). If a gene is under control of the *Neurospora* biological clock, the change curve of the expression of this gene at different time points will be in a rhythmic way.

4 new clock-controlled genes were identified in this way. Three of these 4 new

*ccg*s have no known function. They were called *ccg-13*, *ccg-14*, *ccg-15*. The contigs representing these three clock-controlled genes in the combined EST database are contig1432, 1421, and 1417. The fourth newly identified clock-controlled gene is a lysozyme. It was represented by contig1442 of the combined EST database.

The consensus sequences of the contigs representing *ccg-13*, *ccg-14*, *ccg-15* and *lyz* had been used in a BlastX search against the nr protein database of Genbank (Table 3.38). The results of BlastX research revealed that the translated protein sequence of the *ccg-13* has homology to a phase specific protein of the fungus *Ajellomyces dermatitidis*. The function of this phase specific protein in *Ajellomyces* is not clear. The *ccg-14* showed good sequence similarity to the snodprot1 of *Phaeosphaeria nodorum*. And the translated protein sequence of the *ccg-15* is strongly similar to the sequence of a sporulation specific protein, an SPS2 homologue in *Saccharomyces cerevisiae*.

Table 3.38 BlastX sequence homologs of four novel clock-controlled genes (from Zhu et al., 2001)

Gene	Homolog, organism, and accession number	P/E value
ccg-13	Phase specific protein, *Ajellomyces dermatitidis* AF277086	3e-10
ccg-14	Snodprotein1, *Phasosphasria nodoroum* AF074941	7e-99
ccg-15	SPS2 homologue, *Saccharomyces cerevisiae* NP_009634	9e-45
lyz	Lysozyme, *Chalaropsis* sp, P00721	4e-85

3.2.9 Several common features of the *ccg* genes identified in *Neurospora crassa*

Totally, 11 clock-controlled genes were detected in this research (Table 3.39). 4 of them are newly detected *ccg* genes. Among the remaining 7 *ccg* genes, some other common features are shared. For example, 6 of these 7 *ccg* genes have one or two

Table 3.39 The expression of clock-controlled genes in two cDNA libraries of *Neurospora crassa*

#contigs in NMNE	ccg	Gene description	evening counts	morning counts	RatioM/E
1399, 1410 1446,480 634	ccg-1	Glucose-repressible gene protein (grg-1)	191	351	1.8
1447	ccg-2	Hydrophobin precursor(rodlet protein)(Blue light induced protein 7)	248	457	1.8
1002, 1262 605	ccg-4	Pheromone precursor homolog	42	0	0
1122,1165 1435c,d	ccg-6	No identity	179	33	0.2
1149,1200, 1394,1425, 1448,238, 273,53,975	ccg-7	Glyceraldehyde 3-phosphate dehydrogenase	214	692	3.2
1356	ccg-8	No identity	0	37	∞
1094,1332 1358,529 739	ccg-9	Possible *Neurospora* trehalose synthase	26	66	2.5
1432	ccg-13	New identified *ccg*	7	117	16.7
1421	ccg-14	New identified *ccg*	49	56	1.1
1411, 1417	ccg-15	New identified *ccg*	36	140	3.9
1442	lyz	lysozyme, New identified *ccg*	41	152	3.7

introns. But the size and the position of these introns are different. Some features of their promoters are also similar. Brief summaries of the features for these 7 clock-controlled genes and some related *ccg* genes are addressed here.

3.2.9.1 The introns in the clock-controlled genes of *Neurospora crassa*

Six of the Neurospora clock-controlled genes have intron(s) in their genomic DNA sequences (Table 3.40). The average intron size in fungi is < 500 bp (Dunlap, 1996). Most of introns in these Neurospora clock-controlled genes are <100 bp. The sequences of the introns in ccg genes of Neurospora crass are listed as follow. The 5' and 3' splice junctions and the branch sites are underlined and bolded. The sequences of all these introns follow the classic pattern of "5' GT ...AG 3' ".

ccg-1(grg-1):

976

GTAGGTTTCCTCCAGCTCTCGCCTCCAGCAC*CCG*AGGCACATCTCGGGCATCTTC ACAACAACAGACACTGACATCTCATTCTCAC**AG**

1063

1137

GTAAGCTCCATCACTTTTCCACTCTCACAGCCATCATCAGA**CACTAAC**ATTCATCT
TCCTCTC**AG** 1201

ccg-2 (bli-7):

1893

GTAAGTTTTTCCTCCTTCTCCCTCCACACTACAGCGCGCTCTCTTCAGAAACAGTTCCCCT**TGC**
TAACCTTCGCTTTTGTG*CCG*ACGAAAAAA**G** 1986

ccg-4 (ppg-1):

1417

GTTAGTAACAGCTTCCCTCTCTCGAGATGCC*CCG*AGCAGCCCAAA**TGCTAAC**TTTCATCAC
TTCGGC**AG** 1485

ccg-6 gene:

976
GTACGACACGGTCTCAC*CCG*AGACACGGGGAAGGGAACGCTGGGATGAAAGACGGGAACGGGAACTATAGCCTGGGGCTGGGTTGTGTG
AGGGAACA*CCG*TGTC*CCG*GTGTC*CCG*GTGTCC*CCG*GTGATA*CCG*CAACATAAAGCCCCATGGAACCTGGTC*CCG*GCC*CCG*GATTCGTCC
CACCTTCCATCTGTTCCTG*CCG*CAGTCTGGT*CCG*CTTAGAACGAGATTATCAT**TGCTAAC**GCGTCCTTTCCTTCTCT**AG** 1232

1316

GTGCGTCCCGTCTCCTCTTGGGAGCCGTTTAGAACATGAAATAACTAACACTTATTCACAG 1376

ccg-7:

272
GTATGTTGTGACTGCCCTCGCATTTACAGAAACCGAGCTTCCTTCCTCAACACTTCCAATCATCGTCACTTCCCTTGTCAGCGGCGGCG
GCAGCAGCAGCAGTAGCAGAAGCAGAAGCAGAAGCAGCAGCTACCCCGCACCTTCCTGACCCCGTCCCCGACCCCGTCCCATCTCATCC
TCAAGTCAAGTTCCTCCCGCCTCGCTGCCAAGCTGCGCACAGCATCTGGTGTCTGCGTCTGTTTCCCCCCAAGAGGAAGTGGACGAGAC
TCAGATCGGACTGGCATGGATGCTGGTGGTGGTGGCGGCATTGGAAGGGTTCCTCGGAATCGCTCCTCCCCGATCCTACCTGCAGTCGG
TCCCTCCGTGTTTTGGGCGCTCCTCGTGTCCAATTGTTCTGCCACGCAAACATGTGAACAGACGAGACCGAACAGGATAAGGAAGGGCA
GGCAGACGAGTCCGGCTTTTAAAACCCAGACTTTCCTTCATCCTACCACTCATCATCATCTTACAACCTTCAACAACTTGCTTCACAAG
GTCTTGATACTTACTCGTCTTCACTCCAACAAG 838 (567)

970 GTAAGTTGGCCTCGCTCACATAGATCCCTTGTCTCATATGACAACTCAGACTCTGACCATCATCCCTCTTACAG 1043 (74)

ccg-9 gene:

685

GTGAGTACCTTGCCCACCACGGCTTGCTTGACTGCGTCGACGTTTCATGCTAATATTGGATGTCCCATAG 754

1482

GTAAGTAAAGCAGTTTGCTGTCCGGAACGCTAGACGGTCAACTAACACTGTTTCAG 1537

Table 3.40 The size of introns of different *ccg* genes of *Neurospora crassa* as verified by comparison to the corresponding cDNA

	ccg-1	*ccg-2*	*ccg-4*	*ccg-6*	*ccg-7*	*ccg-9*
intron1	88	94	69	257	567	70
intron2	65			61	74	56

3.2.9.2 The promoters of several clock-controlled genes

The promoter of several clock-controlled genes detected in this research was mentioned for individual ccg's above, and is summarized in table 3.41.

Table3.41 Summary of the promoters of *ccg* genes in *Neurospora crassa*

	TATATAA		CAAT		GATA		GTTGGGAT		others	
	+/-	position	+/-	position	+/-	position	+/-	position	+/-	position
ccg-1	+	1523-	+	1549-			+	1346-1353	CRE	1077-1086
L14464		1529		1552					NRS	1122-1131
ccg-2	+	1454-	+	1242-	+	1359-1362	+	1431-1438	ACE	1401-1445
X62170		1460		1245					LRE	1285-1400
									PLE	878-1042

ccg-7 *U67457*			+	3-6	+	98-101 148-151 155-158			CT box	186-205
ccg-6 *AF098908*	+	603-609	+	591-594	+	494-497 505-508				
ccg-4 *AF088909*	+	359-365	+	426-429 152-155		350-353			CT box	394-408
ccg-8 *AF088907*			+	2094- 2097	+	2167-2170				
ccg-9 *AF088906*	+	407-410	+	114-117		520-523 278-281 197-120 122-125				

3.2.9.3 The poly A signal of *ccg* genes in *Neurospora crassa*

The 3' end poly (A) tail is required for transport of the mRNA from the nucleus and then serves to protect the 3' end of the mature mRNA from degradation (Sachs and Wahle, 1993). In addition, poly (A) on 3' tail may increase the formation of a translation initiation complex at the 5' end with the help of poly (A)-binding proteins (PABP). In yeast, polyadenylation is also involved in the degradation of mRNA (O'Hara et al, 1995). Since the position of the 3' UTR sequence of a gene is also a unique signature of a gene, it can provide us information of the identity of gene. The 3' end of a transcript is generated by cleavage, and then up to 200 polyadenylic acid nucleotides are added to the 3' end of the mRNA (Sachs and Wahle, 1993). Two enzymes are required to do this task: an endonuclease which cleaves the transcript and a poly (A) polymerase (PAP) which joins adenosine to the 3' end of the mRNA (Mitchell et al., 1997). Both enzymes require a signal sequence which is AAUAAA (AATAAA). This signal sequence is highly conserved in higher eukaryotes but not as well in yeast. The following list is the possible

poly (A) signal sequences identified from the *ccg* genes of *Neurospora crassa*. In addition, the polyadenylation signal sequence of the *Neurospora crassa Asm-1* gene is listed here as well because this was predicted from the sequence of the *Neurospora crassa* cosmids in this research. These sequences were located in the region from 11 to 30 bases upstream of the poly (A) site.

ccg-1 (L14464): ATTAAT , 2241-2246
ccg-2 (X62170): ATTAAT, 2348-2353
ccg-4 ((AF088909): AATACT, 1841-1846
ccg-6 (AF088908): AATACT, 2078-2083
ccg-7 (U67457): AATAAT, 2062-2067
ccg-8 (AF088907): ATTGAA, 3506-3511
ccg-9 (AF088906): AAGAAG, 2190-2195
Asm-1 (U51117): AATACG, 3747-3752

In summary, transcription of the clock-controlled genes (*ccgs*) is regulated and under control of the biological clock. Therefore the transcription levels or mRNA abundance changes in a rhythmic way. Although the clock-controlled genes are different from the clock gene, expression of the clock gene itself seems to be rhythmic. Therefore, the clock gene likely is self-regulated, by an as yet undescribed mechanism that controls its own transcription and likely its translation. Such a mechanism might entail product inhibition through an as yet undescribed feedback pathway.

3.3 Analysis and annotation of the sequences of *Neurospora* cosmids

Cosmids nc14, nc17, and nc19 contain inserts cloned from *Neurospora crassa* chromosome five and were provided by Dr. Rodolfo Aramayo, the Texas A & M University. The genomic insert in nc14 was from strain x14:G1, while the genomic insert

in nc17 was from strain x17:A7, and that in nc19 was from strain x19:C5. The initial sequencing of nc14 and nc17 was done by Jennifer Gray and the sequencing of nc19 was performed by Sara Downard.

A multiple sequence comparison revealed that more than 95% of the sequences of these three cosmids overlapped. However, they shared one gap in a GC rich region of approximately 300 base pairs which presently is being attempted to be closed. Since cosmid clone nc17 contained the largest insert, only the consensus sequence of the cosmid nc17 was used as query in the computational prediction of exons/genes in this research.

Both GENSCAN and powerBLAST were performed to predict genes and their possible structures in this research. GENSCAN is a program that was developed by Chris Burge of the Department of Mathematics, Stanford University, that is designed to predict the locations and the exon-intron structures of genes in genomic DNA sequences from various organisms (Burge and karlin, 1997, 1998). The GENSCAN web server at MIT and its web address is at *http://genes.mit.edu/GENSCAN*, allows for up to 1 Mbp sequence length in fasta format to be searched in 10 seconds. PowerBLAST is a database similarity search program developed at the NCBI (Zhang and Madden, 1997). In this program, several types of queries are masked or eliminated to reduce spurious or misleading results. In addition, powerBLAST organizes the BLAST search results to yield organism-specific results, and it performs a sensitive gapped alignment. PowerBLAST processes the sequence by dividing the sequence into small overlapping fragments, which are merged after searching. Several options are available for users to

view the BLAST result: graphic view, textual view, and/or an HTML page that links to GenBank and Entrez in this very convenient and user-friendly program. The address to access to NCBI powerBLAST is at *ftp://ncbi.nlm.nih.gov/blast/network/blast2/powerblast*.

3.3.1 The annotation of the nc17 sequence with GENSCAN 1.0

Since one gap still existed in the sequence of cosmid nc17, the sequences of the two biggest contigs of this cosmid were catenated, separated by 50 X's. Since the first contig contained 11781 base pairs and the second contig contained 25562 base pairs, the concatenated contigs contained a total of 37393 bp (11781+50+22562).

Ten genes were predicted in nc17 by GENSCAN using yeast as the model (**Table 3.42**). The symbol "+" indicates this gene was predicted on the input DNA strand and the symbol "-" indicates the gene was indicated from the opposite DNA strand.

Table 3.42 Ten Predicted genes from NC17 by GENSCAN

Gn.Ex	Type	S	.Begin	...End	.Len	Fr	Ph	I/Ac	Do/T	CodRg	P....	Tscr..
1.10	Intr	-	343	85	259	0	1	22	80	284	0.748	19.00
1.09	Intr	-	762	718	45	2	0	94	47	75	0.185	2.01
1.08	Intr	-	1784	990	795	1	0	50	115	695	0.562	59.64
1.07	Intr	-	4480	4439	42	0	0	141	76	-9	0.221	1.04
1.06	Intr	-	5702	5619	84	1	0	103	35	98	0.319	5.43
1.05	Intr	-	6008	5789	220	0	1	48	21	190	0.073	6.19
1.04	Intr	-	6753	6658	96	1	0	84	46	55	0.055	1.38
1.03	Intr	-	8593	8499	95	0	2	59	115	67	0.607	6.61
1.02	Intr	-	10672	10591	82	2	1	38	96	73	0.661	2.29
1.01	Init	-	11622	11262	361	0	1	43	105	111	0.893	5.56
1.00	Prom	-	12208	12169	40							-3.51
2.00	Prom	+	12826	12865	40							-0.61
2.01	Init	+	13042	13286	245	0	2	54	23	245	0.872	11.80
2.02	Intr	+	13445	13471	27	2	0	90	77	76	0.702	4.41
2.03	Intr	+	13816	13979	164	1	2	-2	75	117	0.278	1.53
2.04	Term	+	14981	15150	170	0	2	-32	52	194	0.329	2.16
2.05	PlyA	+	15525	15530	6							-0.45
3.08	PlyA	-	15538	15533	6							-10.57
3.07	Term	-	15922	15540	383	2	2	67	38	251	0.881	13.37
3.06	Intr	-	16737	16592	146	2	2	83	77	127	0.968	11.54

3.05	Intr	-	17425	17327	99	0	0	70	8	138	0.906	3.72
3.04	Intr	-	17718	17681	38	0	2	61	91	38	0.442	-1.05
3.03	Intr	-	17904	17803	102	0	0	66	97	27	0.340	2.27
3.02	Intr	-	18442	18297	146	2	2	30	87	72	0.595	1.81
3.01	Init	-	19838	19703	136	2	1	58	11	147	0.412	4.27
3.00	Prom	-	19960	19921	40							-12.33
4.00	Prom	+	19990	20029	40							-10.67
4.01	Sngl	+	20261	22192	1932	1	0	85	41	1841	0.994	173.70
4.02	PlyA	+	22342	22347	6							-0.45
5.00	Prom	+	22474	22513	40							-8.38
5.01	Init	+	22792	22844	53	0	2	98	103	-9	0.331	2.18
5.02	Intr	+	23430	25770	2341	0	1	52	69	2673	0.164	250.15
5.03	Term	+	25939	26028	90	0	0	29	48	95	0.207	-2.38
5.04	PlyA	+	26141	26146	6							-3.74
6.02	PlyA	-	26295	26290	6							-1.95
6.01	Sngl	-	28021	26684	1338	1	0	86	39	2523	0.997	243.08
6.00	Prom	-	29342	29303	40							-6.80
7.00	Prom	+	29487	29526	40							-7.20
7.01	Init	+	29692	30115	424	0	1	105	62	379	0.637	33.54
7.02	Intr	+	30183	30433	251	1	2	13	37	104	0.510	-4.51
7.03	Term	+	30493	31545	1053	0	0	8	49	1292	0.675	109.95
7.04	PlyA	+	31680	31685	6							-4.73
8.04	PlyA	-	32053	32048	6							1.05
8.03	Term	-	33175	33069	107	2	2	-1	44	140	0.531	-0.43
8.02	Intr	-	33518	33232	287	1	2	86	23	252	0.896	16.03
8.01	Init	-	34274	34102	173	2	2	42	63	141	0.672	5.88
8.00	Prom	-	34331	34292	40							-14.09
9.00	Prom	+	34558	34597	40							-10.49
9.01	Init	+	34721	35151	431	1	2	79	68	499	0.984	42.76
9.02	Term	+	35213	35657	445	2	1	-4	43	497	0.999	31.09
9.03	PlyA	+	35691	35696	6							-3.74
10.02	PlyA	-	35897	35892	6							-0.45
10.01	Term	-	37071	36918	154	2	1	100	44	112	0.330	5.71

The graphic view of the gene prediction output by GENSCAN is presented in **Figure 3.47.** Gene No# 2, 4, 5, 7, and 9 are predicted from the positive DNA strand and gene No# 1, 3, 6, 8, 10 are predicted from the negative DNA strand. Interestingly, the predicted genes #4 and #6 are two single-exon genes. The peptide sequences and their amino acid length of the above 10 predicted genes are given in Table 3.43.

GENSCAN predicted N.crassa genes in cosmid nc17

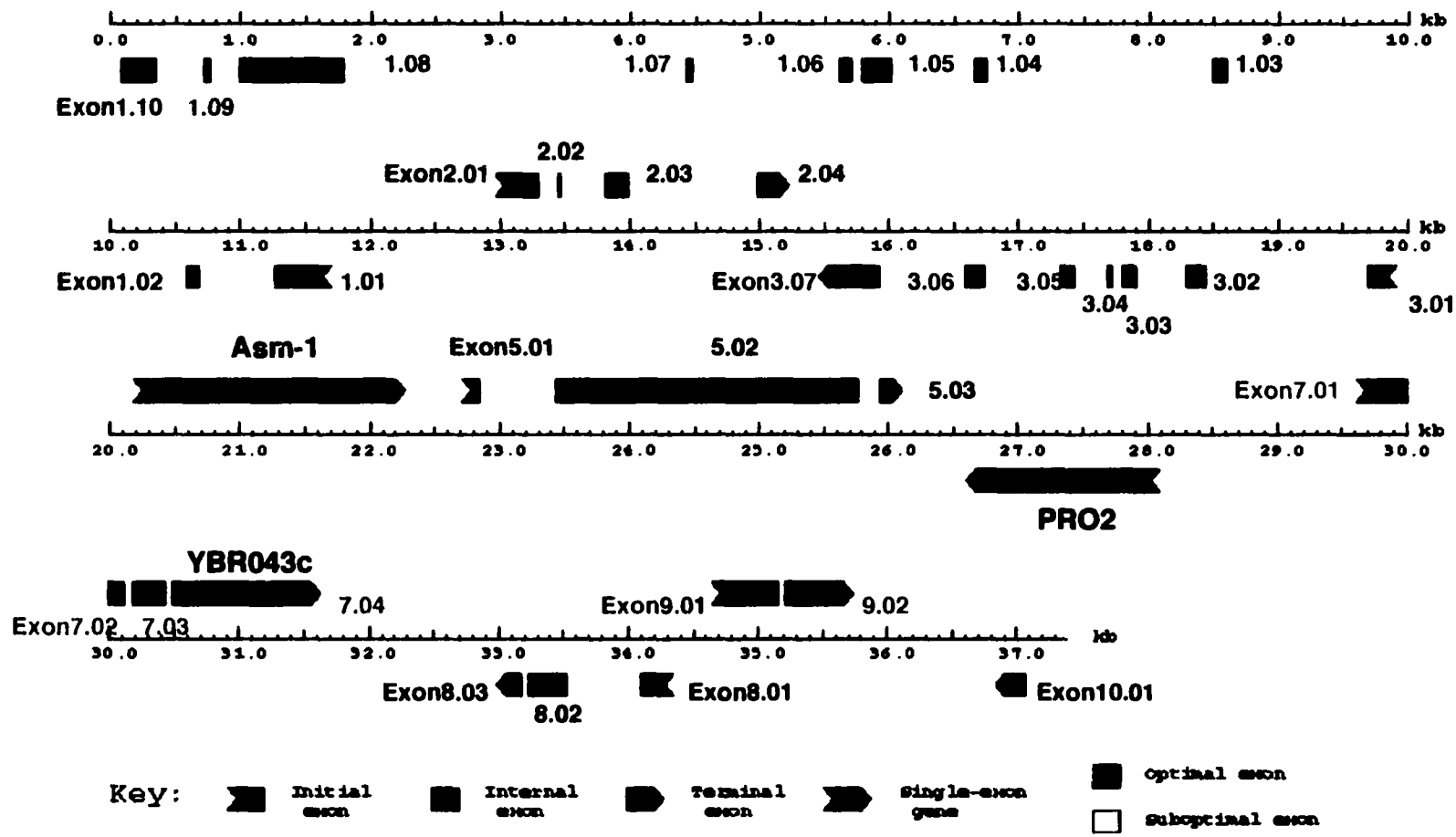


Figure 3.47 Ten genes predicted in cosmid nc17 by GENSCAN. Predicted #4 and #6 gene are two single-exon genes

Table 3.43 Ten Predicted peptides in cosmid nc17 by GENSCAN.

# peptide	Sequence of predicted peptides	length
GENSCAN_ predicted_ peptide_1	MLVVVIEEVFCFLKDVVATKILKPLGWDKRDCTNSARRYANHKAPRRAYTPSIPSYSPGT SCMRGPFSALHGPGMQQATVTVERGRTQVRSRETHKFTRWVPSCVESSTGLPRASPKYAG ERFCTEDFEIRSIGFTCHAYTHHPWSVRGNVEGRDEFRKHSPTSIYSGKKSITICWLLYG ISNFELANRFTDRVKVPLNPTPPPPAPYISDGFFCSCVSQTSLVVNQPVKRDGKDGRLAS HQFSELCKEPEPKTRKFYGDDGCLFDVDISMADGRWRPGGAATASPSHLYPAIATGCKQE HVNGLAYCHYLRRLSSVHSVVRGEAKSPEDAIQIQPTTTKQTGCAMAPPIHPKQPPPAEP IHAGRAFDPWNSVAAGHQRAETRVPSGWRESRARKLQSQFRASSSGGDRISDSVGAGAED FDEKRGVLTPKEVRARALNSVADMLRNPGTMATMSQTTSQSRARLPEKGASQNGSQPVEE EETWESGPTQHGQRKEEEGGGSSNIRKIFDGLNIYVNGSTHPLISDHKLKQILAENGARM SIHLGRRQVTHVILGKPTGRNTGAGGGLAGGKLEKEIRRVGGCGIKYVGVEWVLESIKAG KRLPEARTQFSKAKQEKHIPGTMIQYTVPLSTMEIQQREQHDTTLARIVSVMPYVRHLEK VEYRPDETRFATTTRPPGVRKQAIVSHSYPLKL	693_aa
GENSCAN_ predicted_ peptide_2	MLKGPIRFNFRKCIYLVEPVTLLRRRRGPEPYRDAGRAHGRASWNVKLVIGFDEWDLPKL SISVQDPGPSADELGEHLEETMPTHVNGGVREGRGTCVVALVCSSTSRYHNCEMQQRQPM MQGLGTLLPPNLLFSDSRIQNRVEIDQRSDRETSQSPRTTNLLRDTYQCNPREPWGTLSK VHEFQTATRRRLPRTPIMEYS	201_aa
GENSCAN_ predicted_ peptide_3	MVATVGIEVFQALSYAEPKTQCSFDAWSDRGRAHSCTKNRNTERGAQGLESDAASYTSQH FCILPTLSIVEDPKPPWSHMRRNFLDALRKDEKQVRSRAISLCTGTVNVHSVCDTSISFL TALVMVFKGCQREQTASVQVRPGVEILGWPPAPGSEKSSGEDECPSGNFARLPRSHAGPL AIVDKYSLRGFFRRSPSKRTTTCLHTHHVAEIPWNRRKPVRPALSECRELTSTSSRHAKQ PRRHYTILLHFTSGIAIAFGSVPLWRPAVPVLVCVSTHAHRKPFGKVTPILVNRQTRAVL AGRLQPEKNLIGYPAQPAASPTISDRRPHSAAPCVRCRNLPVITGSDQR	349_aa
GENSCAN_ predicted_ peptide_4	MNPNTPADVYYGQMSQGSSMPVTTVPSHSHYASQQPPPLLQPGSTYAHQYGTPQYGYANA LSSPASIPPSLPPSMNSMAGQSVLPLPGSGSMNPAVYASGGFDTTGQVAPPGMKPRVTAT LWEDEGSLCFQVEARGICVARREDNAMINGTKLLNVAGMTRGRRDGILKSEKVRHVVKIG PMHLKGVWIPFERALDFANKEKITELLYPLFVHNIGALLYHPTNQSRTSQVMAAAEQRRK DSHGQLRGPPGLPSLQQHHHHHSMLPGPPSLPSHPSMGRPALDRAHTFPTPPTSASSVMG PMGNSDGYQWSQQSMSGTQGNSSLSLDTSLGSNARSMPSTPATTPPGSTIQSMQNYPPVS QSYESSRQMYQGQSAQQAQYQSQQHYSSQPQHQERPVYSQSSYIKNDMGPPSGRPTGQSN DASDSKPPTGMIHQGQGQSDPGTHAGSEEDDDANNEAEYTHDSGGYDANRGSYNYNTQAV NSLPHDHGLAPEIGGSPHQAGSGRATPRTAAAPSSYYSAQGYHTPPRGQPSSSLYNVMSN ERTGSNGTQGNEMYAGQADMPSSLPNGYSAQPSVMNGSSGGLKRGRDDDDDGGRPTTSAP NLGPGMDMKRRKTMMDGGSLPSPTYTATIAQAAPSAIAAHRRR	643_aa
GENSCAN_ predicted_ peptide_5	MGTLDRCKNRDFGRYILCNRRQEISDIMDMDLDDAPSRRNNTQQTTAGDNADNSKSSEAD NTTSPPIGRDPSPPKSARPMSRIISGNELSPLKILQHQAQPQPLPESADSPLSPVAKSMA PPPLPQSARKTPIKRFPVKVNPSLPGSETTRSSSRDSQDRRISLQDVIRENGGIKQAIDI FEDDDVDMTGIEGDSGSKDTLLQNTSGNSTGSEHEQATREESVCEGNDTIVSTFSTFSAV PDLTKFSHMRSESASRFSVLGGGVSPAKQTPRVNGLVPDNKTTTGKKRDFDSGNASNLLD FTDSSRYGGYGAPQASPTRRTQFASPRPGSTDLATTTPQRPGSNLSNILDFDIPPMPTPR SVPTITPRELETLKSGFLSEISSLKASLSGKEAEVTSLKTAVSDAEKRVGECMERLHEVE SMHESLEAEKNTWERRGREMEVVLRQVKEEIVLGRRREEELEFKLDEAEKRREAAEMMAQ DAESKMAGMRAGKASAEAAAHESPDKARGQPSSNKEVEMAVERVARELHALYKSKHETKV AALKKSYESRWEKKVRDLQTQVDELSRENEELRKQRDQDKTALATLNPARLTELEEERKA DRARNAVQIREFEAEVEKLVAILQTVKDDNDELRALLEQERVEKGELVRLAEEMMSMQQE PPAPAPAPAPTHLSTSNGRGRADRRGSDVSATSNGDSGHHAPQPASMTSKIGAKPRLGLG GGPGAVRAPIKTQSASGIAPPSGMEGAFRAGGVGRPTSMRAPGGSGLGLKVSAGTRIGLG GGSGHGRAQSATTSGPNGVLKQQDKGGGKEGSNGDTTKTTAKRTDAK	827_aa
GENSCAN_ predicted_ peptide_6	MSLTNASPADAARAAKSASHVLATLSAEARNDALTAIHAGLTAARDEILAANARDLELAR QAAADGKLSASLVSRLDLGKTGKWEDMLKGIWDVRDLEDPVGRVTLRTKLDDGLELERVT CPIGVLLIIFEARPEVIANIAALAVKSGNAAILKGGKESTESFVAISTAISNALSQTQVP NDAIQLVTTRESIPQLLALDRDIDLVIPRGSNELVRYIKENTKIPVLGHADGLCSIYLAA SADAKLAADVVVDAKTSYPAACNSVETLLVQESALTTVFPAVASALAAKGVELRCDAASK AALSPEITTNIKDATAKDYDTEFLSLTLAVKTVPDLSAAIAHINTHGSHHTEAILTADQA EAERFMSAVDASGAYWNASTRFADGMRYGFGTEVGISTNKIHSRGPVGLEGLMIYKYKIR GQGHVSAVYGEGEGKKRFKHERLAI	445_aa

GENSCAN_ predicted_ peptide_7	MDERHPQKESPRPVALTNAQETLTGGVEEKKGVKEQVTSSSSDTDGPEVVDSSDHEPGPE QDKPHPGGDSALSQPYRTLSSSRVHTACVVARADRRGLLASFAFIPEVERPVEYKNKTKW MITTIVALAAAGAPFGSGIFLPALELMADDLHTSPTITNLAVALYMLAMSIFPLWWSSFS ETLGRRTIYIASFTLFVVFSVLSAVSVNTAMLVIMRLLAGGASASVQAVGAGTIADIWKP AERGRAMGIFYLGPLIGPLLSPIIGGALSQGFGWRSTMWFLAIYGSCTLILIIFALPETL TKRKPLPPPPAENNELKRVSTRQSVAEHSKRVGAITRRFIIEPLEVLLYLRYPPVLVSVY SAAIAFGALFILNISIQTTFHSAPYGYSTIIVGLLYLPSSLGYIVASLFGGRWTDKIMMR EARNAGRYDADGNPIFLPEDRMRENIWLAASMYPAAMVWFGWTAQHGVYWIVPCVANFFF GCGSMLVFGAVTTMLTEFMPHRSSSGVAVNNFVRNIFSCVGTIVAQPLINAMGIGWLCTM IGLFAWVTGNLAIWLLKRNAAKWRESMDNALNTKT	575_aa
GENSCAN_ predicted_ peptide_8	MKEETKNEHGLEKFGHPAALRRRASTCFAFMLAKNIGVGESMRSNERYSKVRQGQKQRAS VPSFAGCLHLPTIIIIVVVVVHQPTHSVLLIQDTQRARSNSFGFLDVVKIAPPSLNLSPS DKSDGIEATDPVSKLKHSSCSVAPYCDGKDVNTARTSEWRGSDGTDWPAGQILKVTDRPG AKKKTVNV	188_aa
GENSCAN_ predicted_ peptide_9	MAPHVKSSATPKDRRKSNGVGVSVASSNGASSKVVTLAVTPKNLRAIVDPGYVKEDTPVL KETNDSPVDSAVIPAVTNSAAENASDSNANTPAAGTPAPQSGPMGPPTDGLKKKGVKRAA GAGVNGNGEAKVRGKPGPKKKQRLDDGTIEGGRGGLAAHKLGPKANMGAINAGLRALDRS GKPCRKWTRGGFTLKSFTGVVWELPRWTAPPKPRPELTAEEPTPVSASAAGSSKENIVPG ENAQVKSESSNNGVDVEMQNAPSFAAANSAAPSPGPTPAPAPASVPTAIAV	291_aa
GENSCAN_ predicted_ peptide_10	XYRTEHIGTVKEWSVGIRTGLLVVVSINIEHRGCRGSNEPEGEQVRWLYDG	51_aa

The peptide sequences of these ten proteins were used in a BlastP search against the GenBank nr protein database and the results are listed in Table3.44. Only three predicted genes had significant sequence homologs in the GenBank nr protein database **(Figure 3.47, 3.48).**

Table 3.44 The BlastP results of ten predicted genes of nc17

Peptide #	aa	BlastP result description	HSP	E-value
1	693	none		
2	201	none		
3	349	none		
4	643	U51117, *Neurospora* ASM1 protein	827	0.0
5	827	none		
6	473	Gb|AAB06995, pir|T41722 probable gamma-glutamyl phosphate reductase	453	e^{-126}
7	575	NP_009599, S45901 probable membrane protein YBR043c, yeast	137	$5e^{-31}$
8	188	none		
9	291	none		
10	51	none		

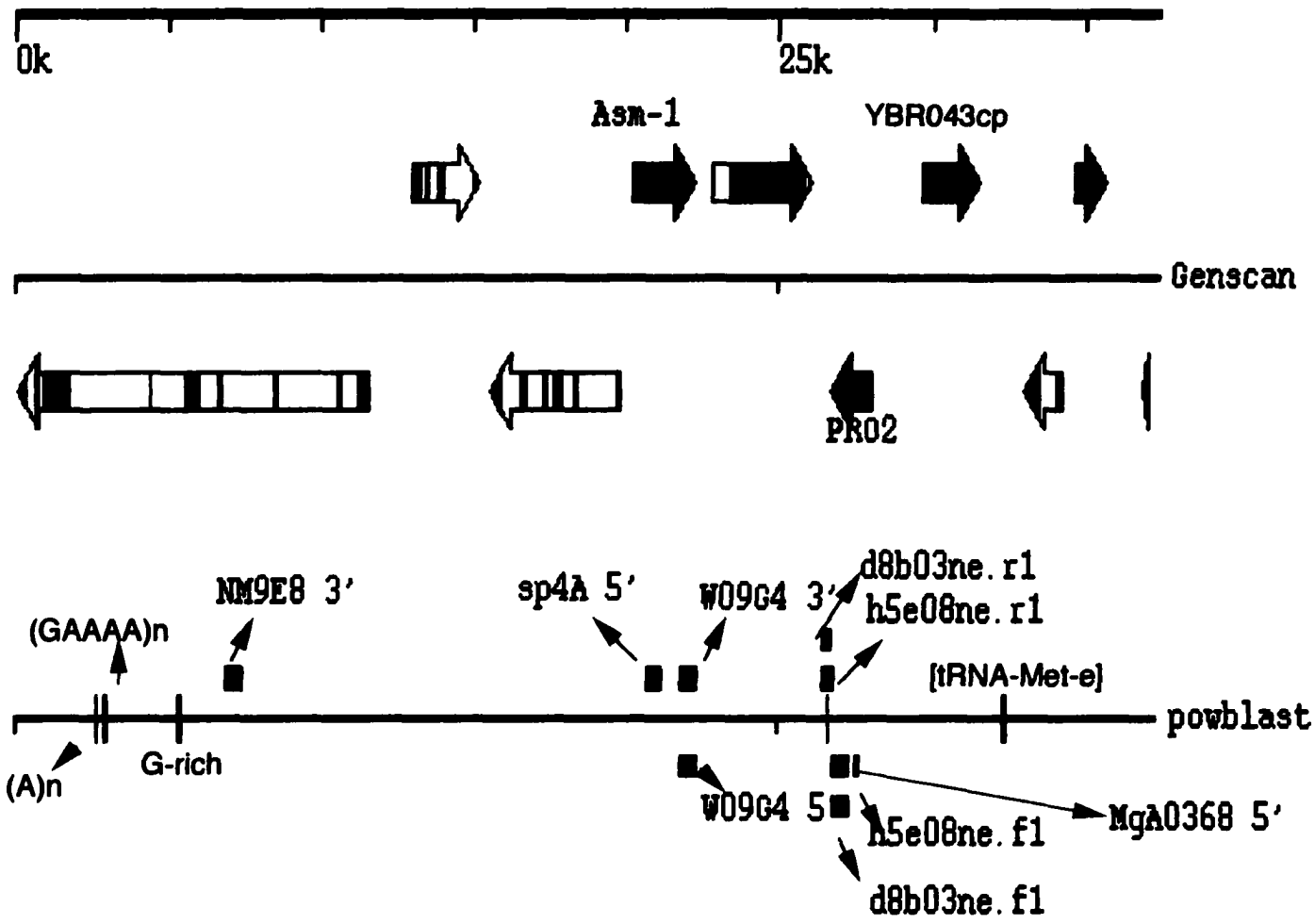


Figure 3.48 powerBLAST and GENESCAN prediction from nc17

A *Neurospora crassa* 3' EST, NM9E8 (AA897819), matched with the nc17 as is shown in the powerBLAST output in Figure 3.48. The sequence comparison between this EST and the nc17 sequence was performed using the GCG Dotter program. The region between 1 to 492 of this EST aligned with the nc17 from 6889 to 7379. The BlastX search against the GenBank nr protein database was performed using this EST sequence as query but no homology was detected. Since GENSCAN did not predict exon in this region from nc17 as well, it is likely that there is no gene around the region from 6889 to 7379 in the sequence of nc17.

3.3.1.1 *Neurospora Asm-1* gene, the #4 GENSCAN predicted gene on nc17

The powerBLAST search result showed that several ESTs and one cDNA had significant homology to ORF #4 (**Figure3.49**), which starts at position 20261 and ends at position 22192, indicating that this ORF encodes the *Neurospora crassa* ascospore maturation 1 protein. When the genomic DNA sequence of ASM1 (U51117) and the amino acid sequences of this protein were compared to the nc17 sequence, the alignment of nc17 DNA sequences with the ASM1 gene spanned from position 18748 to position 22747, matching the DNA sequence of the *Neurospora Asm-1* gene from 1 bp to 4000 bp. The predicted peptide#4 amino acid sequence also was compared with the amino acid sequence of *Neurospora* ascospore maturation 1 protein revealing that they had a 100% identity. The GENSCAN predicted ORF of the No#4 gene matched exactly the ORF of the *Neurospora Asm-1* gene (Figure 3.49). The coding region of this of this predicted #4 gene starts at 20261 bp (ATG=AUG="M", start codon) and ends at 22192 bp

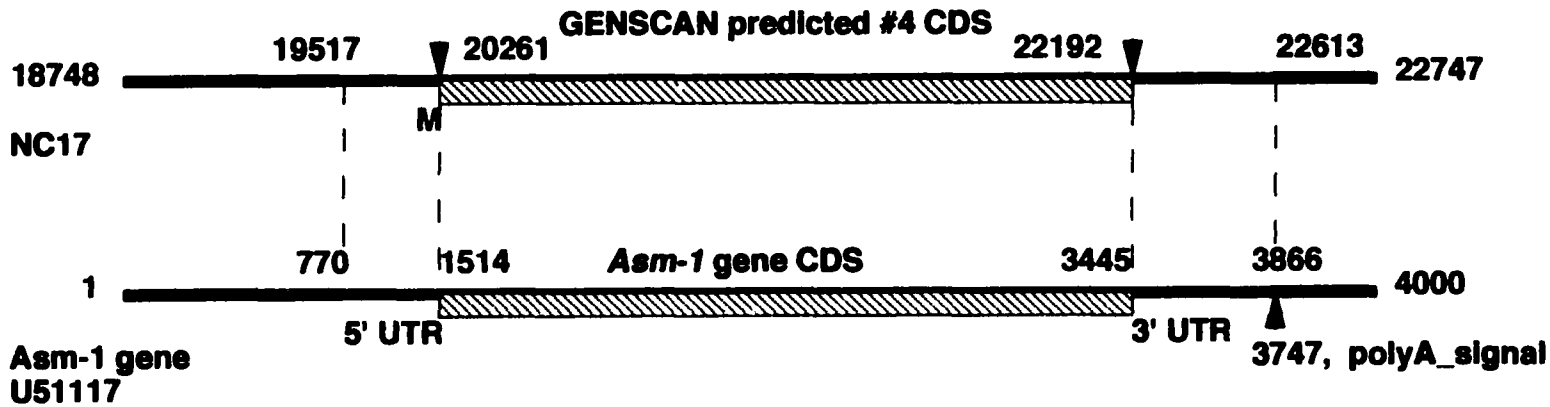


Figure 3.49 The annotation of predicted #4 gene and its comparison with *Asm-1* gene

(TAA=UAA=stop codon) on nc17. The transcription start is at 19517 bp and ends at 22584 bp. Therefore, the region between 19517 bp to 20260 bp is the 5 'UTR and the region between 22193 bp to 22584 bp is the 3' UTR. The promoter predicted by NNPP (Reese and Eeckman, 1995) and transcription start site is slightly different from that determined by GENSCAN. In the 500 bps upstream of 19517 bp of nc17, the NNPP predicted promoter sequence is CGCCG TGACC TAAAT CCAGC CGGGC CCCGT CCTTC CTCTC CCCAG CTCTT. The larger font indicates the transcription start site. The GENSCAN predicted promoter was the region between 19990 and 20029. The sequence of the GENSCAN predicted promoter is $_{19990}$CGGTT TCTTG AAAAG GCACT GAGTT GCAAA AACGC GAACA$_{20029}$. The Proscan and Promoter 2.0 promoter prediction programs failed to predict a promoter from the region between 19017 bp and 19517 bp because their prediction is based on the presence of a TATA box, CAAT box, GC island and/or other RNA polymerase binding sites.

Another three ESTs showed homology to this gene when Powerblast was used to search against Genbank. These three ESTs are SP4A1 (AI392229, 5' EST), W09G4 (AI398870, 3' EST) and W09E4 (AI3992421, 5' EST). The BlastX of these ESTs revealed homology to the ASM-1 protein of *Neurospora crassa* as expected but no ESTs from the *Neurospora crassa* NM or the NE EST databases were observed with homology to this gene indicating that this gene is not highly expressed in these two libraries.

ASM-1 is a 68.5-kD product of *Neurospora crassa* Asm-1^{+} gene (Ascospore maturation 1). It is a homologue of the *Aspergillus nidulans* stuA (stunted A) gene (Miller et al., 1992), a nuclear-localized protein that is abundant and important for

protoperithecial formation in *Neurospora crassa.* There is one region of the protein sequence that is well conserved between ASM-1 and its homologs from several other fungi **(Figure 3.50)**, that include *Aspergillus nidulans* StuA, *Saccharomyces cerevisiae* Phd1, *Saccharomyces cerevisiae* Sok2, and *Candida albicans* EFGTF-1. This conserved region is a DNA binding domain that is called the APSES motif (ASM-1, Phd1, StuA, EFGTF-1, and Sok2) (Aramayo et al., 1996). Because these proteins are required for fungal development, they may constitute a family of transcription factors that may be involved in the control of fungal development.

```
 Anstuaaa 123 PP GAKPRVTATL WEDEGSLCYQ VEAKGVCVAR 154
      Asm 110 PP GMKPRVTATL WEDEGSLCFQ VEARGICVAR 141
Scphd1paa 180 TS VLKPRVITTM WEDENTICYQ VEANGISVVR 211
 Scsok2aa 408 AS IIRPRVTTTM WEDEKTLCYQ VEANGISVVR 439
Efg1_Can1 198 PP GIRPRVTTTM WEDEKTLCYQ VDANNVSVVR 229

 Anstuaaa  155 REDNGMINGT KLLNVAGMTR GRRDGILKSE KVRNVVKIGP MHLKGVWIPF 204
      Asm  142 REDNAMINGT KLLNVAGMTR GRRDGILKSE KVRHVVKIGP MHLKGVWIPF 191
Scphd1paa  212 RADNNMINGT KLLNVTKMTR GRRDGILRSE KVREVVKIGS MHLKGVWIPF 261
 Scsok2aa  440 RADNDMVNGT KLLNVTKMTR GRRDGILKAE KIRHVVKIGS MHLKGVWIPF 489
Efg1_Can1  230 RADNNMINGT KLLNVAQMTR GRRDGILKSE KVRHVVKIGS MHLKGVWIPF 279

 Anstuaaa  205 DRALEFANKE KITDLLYPLF VQHISNLLYH PANQN 239
      Asm  192 ERALDFANKE KITELLYPLF VHNIGALLYH PTNQS 226
Scphd1paa  262 ERAYILAQRE QILDHLYPLF VKDIESIVDA RKPSN 296
 Scsok2aa  490 ERALAIAQRE KIADYLYPLF IRDIQSVLKQ NNPSN 524
Efg1_Can1  280 ERALAMAQRE QIVDMLYPLF VRDIKRVIQT GVTPN 314
```

Figure 3.50 Comparison of ASM-1 of *Neurospora crassa* with its homologs. This region contains a DNA-binding motif APSES(Adapted from Aramayo, 1996).

3.3.1.2. The #6 GENSCAN predicted gene on nc17

The predicted #6 gene is encoded on the minus strand of cosmid DNA. Like the predicted gene #4, it also is a single exon **gene (Figure 3.47, 48).** The coding region predicted by GENSCAN is from position 28021 to position 26684.

The powerBLAST search revealed that two pairs of ESTs from the NM and NE EST databases and another EST (MgA0358, 5' EST) from the dbEST of Genbank

matched with the predicted gene #6 **(Figure 3.48)**. These two pairs of ESTs are h5e08ne.r1, d8b03ne.r1, h5e08ne.f1, and d8b03ne.f1 and were assembled into the same contig: the NE.contig510 of the NE assembled EST database. The BlastX searches of these 5 ESTs revealed that this predicted gene most likely encodes gamma-glutamyl phosphate reductase.

The DNA sequence of NE.contig510 was compared with that of the nc17 using both crossmatch and Dotter. Both results revealed that except for the 14 poly (T) tail on the end of the cDNA, the sequence matched with the nc17 cosmid sequence completely. The sequence comparison also was performed on the protein level. However, the translated protein product of NE.contig 510 only matched with a portion of the 3' end of the GENSCAN predicted peptide #6 **(Figure 3.51)**. Since the length of the predicted peptide #6 was 445 amino acids, and the translated peptide product of NE.Contig510 was only 264 amino acids, only 191 of 264 amino acids match with 191 amino acids of the predicted peptide #6. The first amino acid of the translated product of NE.contig510 aligned with the 255th amino acid of the predicted peptide #6. The output of GENSCAN indicated that the coding region of the ORF for the predicted gene #6 started at position 28021 and ended at position 26684. However, the alignment of the NE.contig510 to nc17 begins at position 26478 and ends at position 27260. Therefore, the cDNAs represented by NE.contig510 only includes the complete 3' UTR region and 60% of 3' coding region but lacks the complete 5' end of its mRNA (Figure 3.52).

```
Ctg510.Pep2       1 KTSYPAACNSVETLLVQESALTTVFPAVASALAAKGVELRCDAASKAALS 50
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
Peptide6.Gcg    255 KTSYPAACNSVETLLVQESALTTVFPAVASALAAKGVELRCDAASKAALS 304

                 51 PEITTNIKDATAKDYDTEFLSLTLAVKTVPDLSAAIAHINTHGSHHTEAI 100
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
                305 PEITTNIKDATAKDYDTEFLSLTLAVKTVPDLSAAIAHINTHGSHHTEAI 354
```

```
101 LTADQAEAERFMSAVDASGAYWNASTRFADGMRYGFGTEVGISTNKIHSR 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
355 LTADQAEAERFMSAVDASGAYWNASTRFADGMRYGFGTEVGISTNKIHSR 404

151 GPVGLEGLMIYKYKIRGQGHVSAVYGEGEGKKRFKHERLAI 191
    |||||||||||||||||||||||||||||||||||||||||
405 GPVGLEGLMIYKYKIRGQGHVSAVYGEGEGKKRFKHERLAI 445
```

Figure 3.51 The comparison between predicted peptide#6 and the translated product of Contig510.

Therefore, the transcription start site could not be determined exactly. According to GENSCAN, the predicted transcription start site is at position 29394 in nc17, while the coding region of this gene starts at position 28021 and ends at position 26684. By aligning the cDNA sequences, we can deduce that the 5' UTR region is from position 29394 to position 28022 and the 3' UTR region is from position 26683 to position 26482 (Figure 3.52). Since the initiation signal score (I/AC) is 86 for predicted gene 6, which is lower than the high probability score of I/AC >= 100, the probability of this site as an initiation signal is only marginal.

To discover if peptide 6 contained any known domain region, it was searched using a variety of programs. Profile (Gribskov, McLachlan and Eisenberg, 1987; Bucher, Karplus, Moeri and Hofmann, 1996) is a quantitative description of a motif. It is a matrix of probabilities for the occurrence of a particular amino acid at each position. It can be used to describe very divergent protein motifs. BLOCKS is a profile database (Heikoff and Heikoff, 1994). BLOCKS contains multiple alignments of conserved regions in protein families. It is based on PROSITE patterns. Each block family consists of one or more ungapped blocks in a discrete order and a database search employs a quantitative profile for each block. A pattern is a qualitative description of a motif (Zhang et al., 1998). A motif is a biological object that is approximated by a pattern or profile such as a

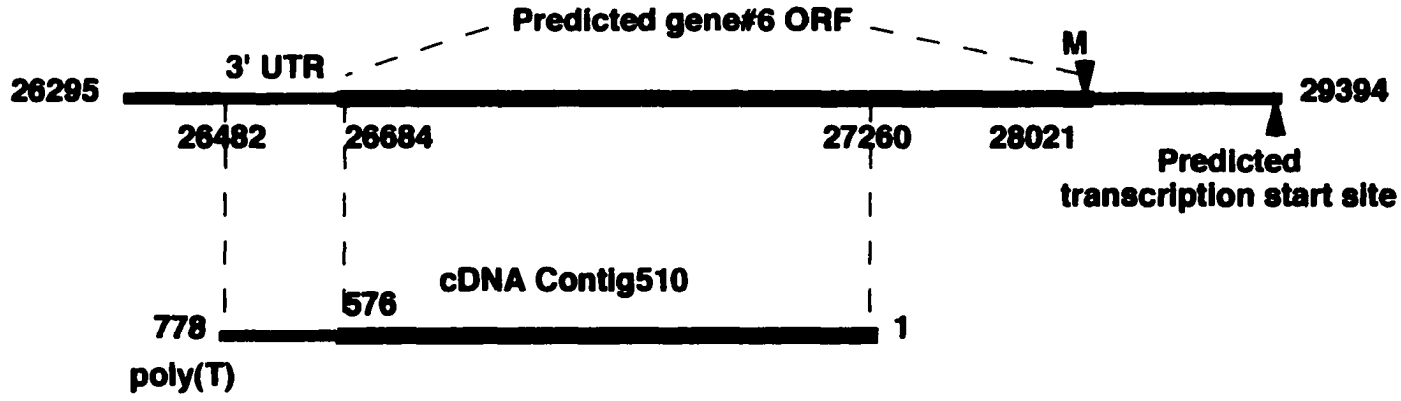
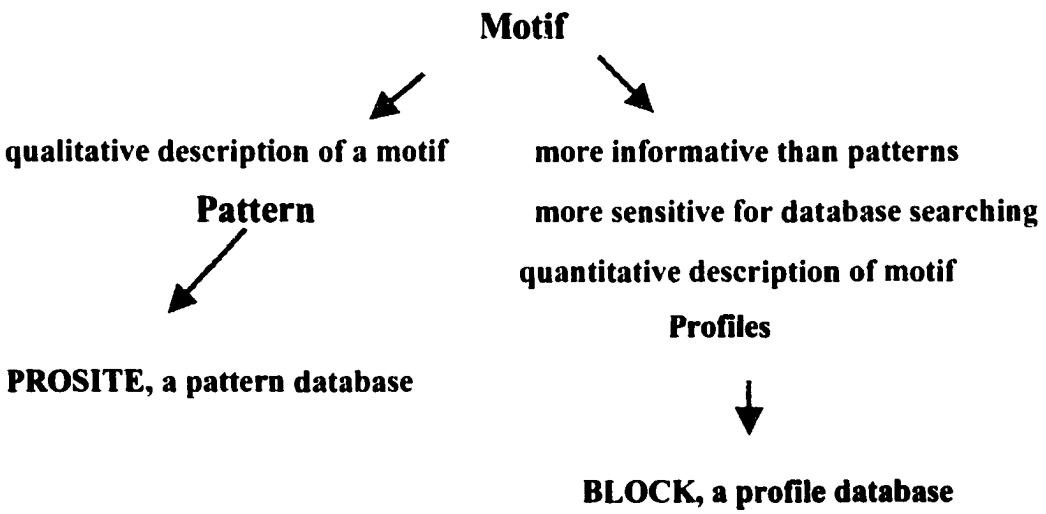


Figure 3.52 The annotation of predicted gene #6 from cosmid nc17 and the comparison with NE.Contig510 of the NE database of *Neurospora crassa*

portion of sequence of amino acid residues that might function as enzyme catalytic sites, prosthetic group attachment sites, icon binding residues, cysteine that is involved in disulfide bonds, or small molecular or protein binding region. Prosite, a database of biological significant sites, patterns, and profiles, that contains a collection of protein patterns and profiles. The search of prosite is based on multiple alignments of SWISS-PROT sequences and linked to SWISS-PROT entries. The following figure is a graphic explanation of above items.



The online program profileScan at http://www.isrel.isb-sib.ch/cgi-bin reveals two possible domains in peptide 6 and these are listed as in table 3.45.

Table 3.45 Two domains searched in peptide 6 using profileScan

Profile ID	Alignment	Domain
PS50310	30-95	ALA_RICH Alanine_rich region
PS50007	366-390	PIPLC_X_Domain Phosphatidylinosital-specific phospholipase x-box domain profile

Seven possible prosite patterns also were revealed in the peptide 6 sequence using

the online program ScanProsite at *http://www.expasy.cbr.nrc.ca/tools/scnpsitl.html*. Amino acid residues 333-354 VPDLSAAIAHINTHGSHHTEAI hits prosite pattern PDOC00940 PS01223 PROA, gamma-glutamyl phosphate reductase signature. The PROSITE pattern search results are given in table 3.46.

Table 3.46 The PROSITE pattern searched in peptide 6 using ScanProsite

	Pattern ID	Name of site	Number of matches
1	PDOC00001 PS00001	ASN_GLYCOSYLATION N-glycosylation site	377-380 NAST
2	PDOC00005 PS00005	PKC_PHOSPHO_SITE Protein kinase C phosphorylation site	81-83 TGK 105-107 TLR 188-190 TTR 315-317 TAK 379-381 STR 398-400 TNK
3	PDOC00006 PS00006	CK2_PHOSPHO_SITE Casein Kinase II phosphorylation site	7-10 SPAD 74-77 SRLD 108-111 TKLD 188-191 TTRE 315-318 TAKD 332-335 TVPD 367-370 SAVD
4	PDOC00007 PS00007	TYR_PHOSPHO_SITE Tyrosine kinase phosphorylation site	209-217 RGSNELVRY 381-370 RFADGMRY

5	PDOC00008 PS00008	MYRISTYL N-myristoylation site	40-45 155-160 347-352 373-378 385-390 391-396 395-400	GLTAAR GGKEST GSHHTE GAYWNA GMRYGF GTEVGI GISTNK
6	PDOC00009 PS00009	AMIDATION Amidation site	433-436	EGKK
7	PDOC00940 PS01223	PROA Gamma-glutamyl phosphate reductase signature	33-354 VPDLSAAIAHINTHGSHH TEAI	

The sequence of peptide 6 also was used as a query to search against the BLOCKS database at *http://www.blocks.fhcrc.org/blocks/blocks_search.html*. These results indicate that peptide 6 likely belongs to "family IPB000965 Gamma-glutamyl phosphate reductase" as it had a combined E-value of $1.2e^{-106}$.

The predicted amino acid sequence of peptide#6 also was used in a BlastP search against the GenBank nr protein database and revealed 59 hits that scored higher than 100. The DNA sequence of NE.contig510 also was used for a BlastX search against the GenBank nr protein database and revealed 41 hits that scored higher than 100. The five most significant homologs were identical in both blast searches and shown in **table 3.47**.

Table 3.47 Top 5 significant homologs of peptide 6 and NE.contig510 searched in Genbank

Pir|T41722	Probable gamma-glutamyl phosphate reductase	472, e^{-132}	*Schizosaccharomyces pombe*
Ref|NP_014968.1	Gamma-glutamyl phosphate reductase (GPR), Pro2R	386, e^{-106}	*Saccharomyces cerevisiae*

Sp|P54902	Probable gamma-glutamyl phosphate reductase	382, e^{-105}	*Synechocystis* sp
Gb|AAG22031.1 AF305580_1	ProA	355, $4e^{-92}$	*Mathanosarcina acetivorans*
Sp|P54886	Delta 1-pyrroline-5-carboxylate synthase (p5cs)	302, 6^{e-81}	*Homo sapiens*

Note: gamma-glutamyl phosphate reductase (GPR)
= Glutamate-5-semiladehyde dehydrogenase
=Glutamyl-Gamma-semialdehyde dehydrogenase
P5CS includes: Glutamate 5-kinase (Gamma-glutamyl kinase) (GK), gamma-glutamyl phosphate reductase (Glutamate-5-semialdehyde dehydrogenase) (Glutamyl-gamma-semialdehyde dehydrogenase)

Another EST with homology to the predicted #6 gene was EST AW180276 a *Mycosphaerella graminicola* cDNA clone 5'-end sequence, which also is a probable gamma-glutamyl phosphate reductase. Taken together, the above results are consistent with the conclusion that the predicted gene#4 encodes a probable gamma-glutamyl phosphate reductase of *Neurospora*.

Gamma-glutamyl phosphate reductase (GPR) is the second enzyme in the proline biosynthesis. The first enzyme of the proline biosynthetic pathway is gamma-glutamyl kinase (GK). In *E. coli* (Deutch et al., 1984) and in *Serratia marcescens* (Omori et al., 1991), these two enzymes are encoded by two genes of the proBA operon. The ProA gene encodes gamma-glutamyl phosphate reductase and the proB gene encodes gamma-glutamyl kinase. In humans, both genes are part of the delta 1-pyrroline-5-carboxylate synthtase (P5CS), a bifunctional enzyme that has both gamma-glutamyl kinase (GK) and gamma-glutamyl phosphate reductase (GPR) activities and it catalyzes the conversion of L-glutamate to glutamic gamma-semialdehyde (GSA) in the proline biosynthesis (Aral, B. et al., 1996). In *Saccharomyces cerevisiae*, three genes encode the above three

enzymes. The PRO1 gene encodes gamma-glutamyl kinase, the PRO2 gene encodes gamma-glutamyl phosphate reductase, and the PRO3 gene encodes delta 1-pyrroline-5-carboxylate reductase (Tomenchok and Brandriss, 1987).

In summary, 10 genes were predicted in the nc17 genomic sequence by GENSCAN. Predicted #4 and #6 genes were single-exon genes. Three have sequence homolog in the GenBank nr protein database and the sequence of predicted #6 gene matches that of two cDNAs of the NE cDNA library of *Neurospora crassa* and one *Neurospora* EST in GenBank. Three ESTs from two *Neurospora* cDNAs deposited in GenBank aligned with the sequence of the predicted #4 gene, which encodes ASM-1 protein. The genome size of *Neurospora crassa* is approximately 42.5 Mb (Radford and Parish, 1997), and the estimated gene number of *Neurospora crassa* is about 8000 to 10000 (Kupfer et al., 1997). If clone nc17 is representative of the *N. crassa* genome, then the presence of the 10 predicted genes in this 37 Kb of genomic sequence, 11594 genes could be expected in the genome of *Neurospora crassa*, a number in agreement with the earlier estimation.

Chapter IV

Conclusion

Two time-of-day-specific cDNA libraries from vegetative mycelium of *Neurospora crassa* were constructed and approximately 20,000 expressed sequence tags were generated to identify and compare the genes transcribed and expressed in these two growth stages, thereby discovering additional genes whose expression is regulated by the *Neurospora* clock.

In total, 10871 ESTs were sequenced and 377 genes were identified from the morning cDNA library while 9148 ESTs were sequenced and 1302 genes were identified from the evening cDNA library (Table 3.08). Although more ESTs were assembled but fewer genes were detected in the NM cDNA library, when compared to the NE library, the redundancy of an individual gene was higher and the variety of gene species was limited in the morning cDNA library.

The cDNA libraries chosen for sequencing in this research represent the expressed genes present within *Neurospora crassa* under conditions which represent the biological clock morning and evening time points. Through results of this study, it is clear that there are almost 4 times of numbers of genes expressed in the evening cDNA library than in the morning library. The expression profiles of two cDNA libraries also were quite different (Figure4.01). In particular, I observed that enzymes in intermediate metabolisms, bioenergetic pathways, protein synthesis, and DNA replication are much higher expressed in NE than in NM library. For example, from table 3.17, it can be seen

that enzymes involved metabolism of lipids, fatty acids, sterols, the metabolism of aromatic compound, sulfur metabolism, phosphate metabolism, nitrogen metabolism were not detected in the NM cDNA library while 30 cDNAs from this group were detected in the NE cDNA library (Figure 4.01). None of the glyoxylate cycle enzymes were detected in the NM library while five enzymes of this pathway were detected in the NE library. Likewise, enzymes in pathways that use fatty acids as an energy source were not detected in the NM library while 3 such enzymes were detected in the NE library. 10 genes for nuclear membrane transport and 8 genes for mitochondrial transport were identified in the NE library while none were observed in the NM library. The expression of genes involved in metabolism (Figure 4.02), bioenergetic pathways (Figure 4.03) and genetic information processes also were very different for these two cDNA libraries. In total, 73 expressed genes for metabolism and bioenergetic pathways and 91 expressed genes for genetic information processes were detected in the NM library while 216 and 248 expressed genes for these processes, respectively, were observed in the NE library. 32 expressed ribosomal protein genes were detected in the NM library while 102 were detected in the NE cDNA library. 8 expressed genes were detected for electron transport in the NM library while 51 were detected in the NE library. The three major pathways using carbohydrate as energy sources, also showed differing gene expressions levels in the two cDNA libraries as shown in table 4.01. Interestingly, although ESTs encoding enzymes for the pentose phosphate pathway (PPP) and the citric acid cycle (TCA) are more highly expressed in the NE cDNA library, more ESTs encoding proteins for the gluconeogenesis pathway were detected in the NM library than in the NE library even

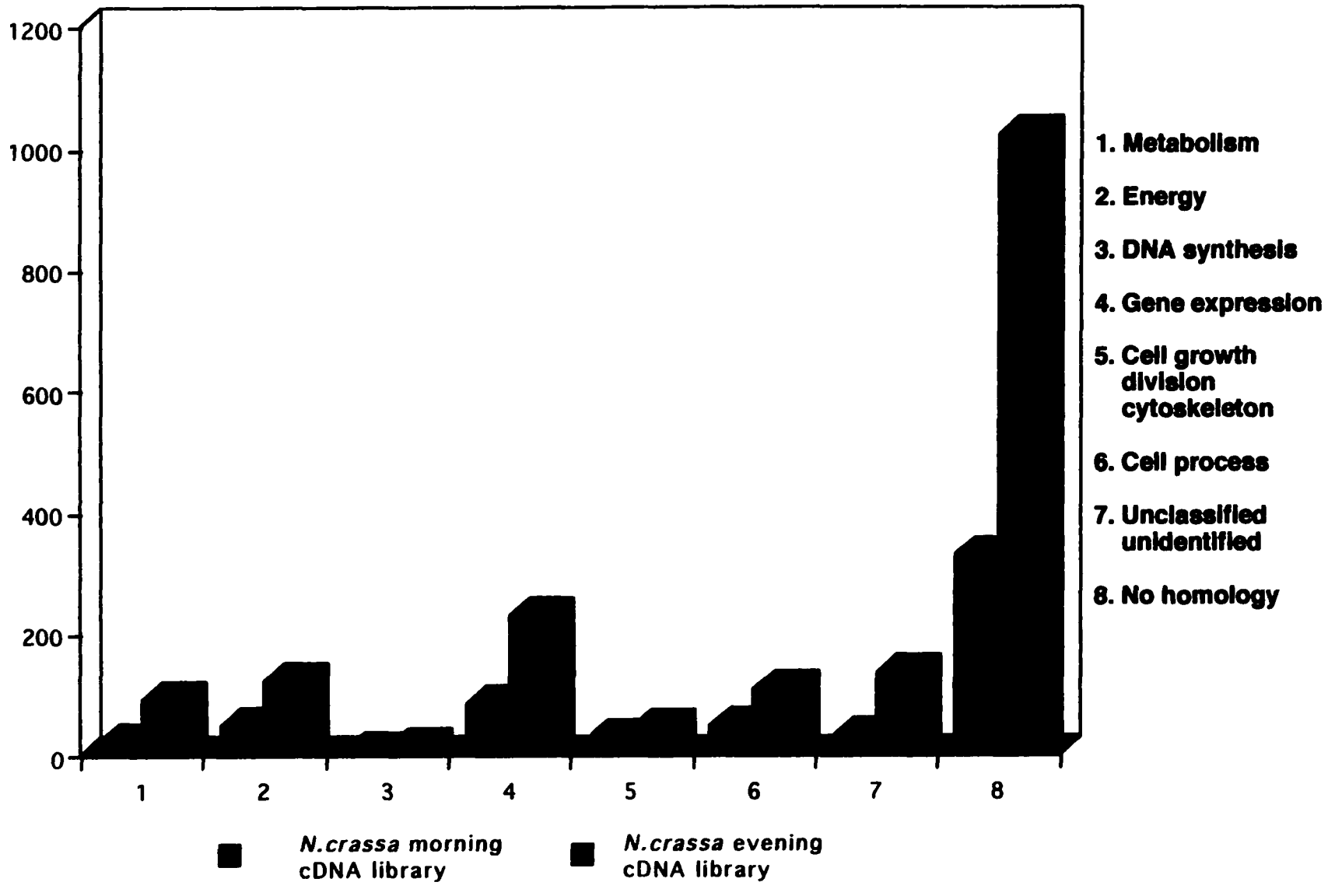


Figure 4.01 Gene expression profiles of the two cDNA libraries of *Neurospora crassa*

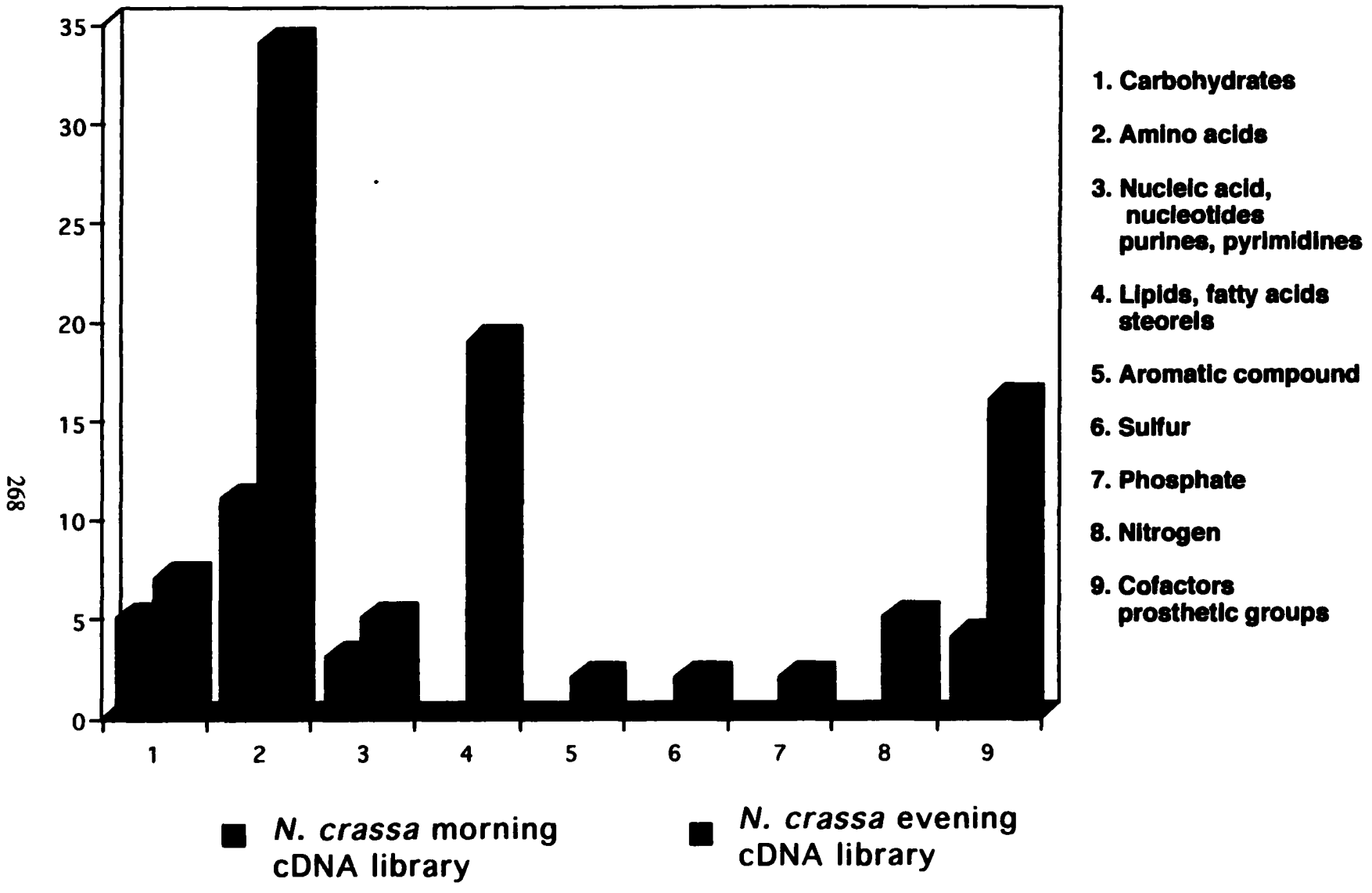


Figure 4.02 Gene expression for the metabolism pathways of the two cDNA libraries of *Neurospora crassa*

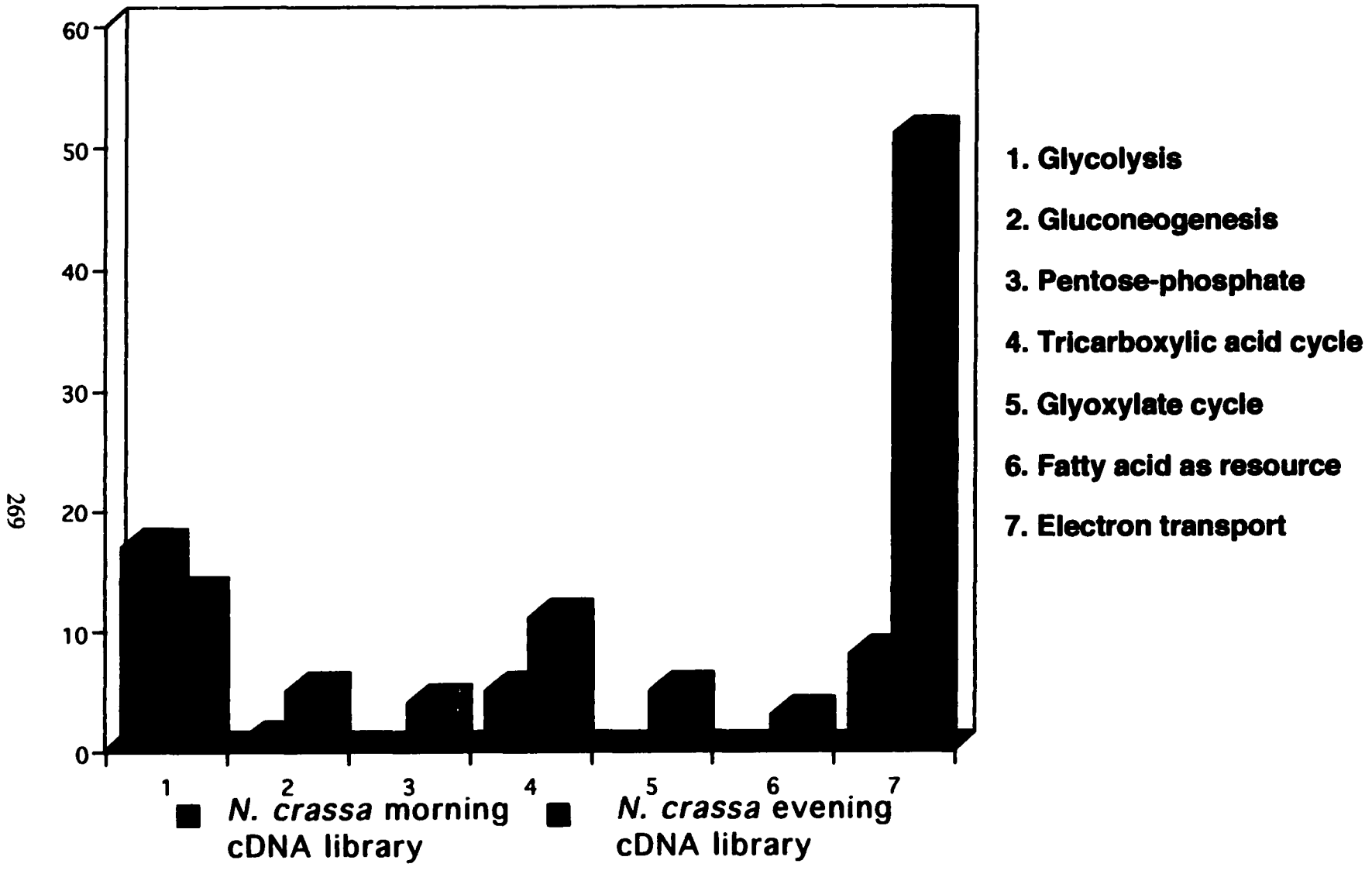


Figure 4.03 Gene expression for the bioenergetic pathways of the two *Neurospora crassa* cDNA libraries

though the total number of expressed genes detected in these two pathways was same for the two cDNA libraries. The expressed genes representing enzymes in these three pathways that were not detected are listed in table 4.02. As can be seen, no expressed genes were identified for enzymes in the pentose phosphate pathway in the NM cDNA library.

These results are consistent with known biological properties of this organism at these two points in circadian rhythms. Specifically, Loros reported that the genes for conidiation are activated in the evening portion of circadian cycle (Loros, 1998). It also is known (Bell-Pedersen, Dunlap and Loros, 1992) that the asexual developmental pathway is initiated in the evening, resulting in spores being produced in the morning. Bell-Pedersen reported most of clock-controlled genes expressed more in the morning than in the evening (Bell-Pedersen et al., 1996c). Clearly, increased gene expression in intermediate metabolism, protein synthesis, DNA replication in the NE library is advantageous. Conidiation controlled by the biological clock may represent an evolutionary adaptation of this fungus that ensures that the spores are produced and disseminated at the proper time of day. Also, since *N. crassa* lives in warm subtropical environment, spore release and viability will depend on both wind velocity and humidity. In a subtropical area, where there is a large temperature variation between morning and evening, any conidia that are elaborated in the day, will be released during the heat of the day and the high temperature would cause them to desiccate. Therefore it would be advantageous if conidial development were initiated at night resulting in spores being produced in the cooler morning, and then being dispersed by the morning winds (Bell-

Pedersen, Dunlap and Loros, 1992).

In contrast, increased clock-controlled gene expression in the NM library also is reasonable because the transcription of the clock-controlled genes are directly regulated and controlled by clock gene *frq*. Oscillation of the *frq* gene transcripts determines the expression of the *ccg* genes. Conidiation and all other metabolic pathways are secondary target pathways whose genes regulated and determined by expression of the clock-controlled genes. Thus, for example, the clock-controlled genes do not necessarily have the same expression pattern as the genes involved in the asexual conidiation development. Therefore, although 894 genes were detected only in the NE library and 281 genes were detected only in the NM library not all of these 1175 genes are clock-controlled genes as only a few have their expression directly modulated by the clock gene product. These primary targets are relatively easy to detect, as they are the genes whose expression shows rhythmic changes in overall 24 hour period and a total 11 *ccg* genes were detected in this research. When these *ccg* genes were expressed, they interact with circadian clock responsive element (CCRE) (Figure 1.05) and then the clock-controlled processes are activated. For example, the assembly of the components for the aerial hyphae and the preparation of conidia at the night time is one of these *ccg* gene regulated processes (Loros, 1998). In addition, the sexual developmental cycle might also be under direct clock control (Lakin-Thomas et al., 1990). Thus, although a large number of genes and gene products, and there are four-times more genes expressed in the NE than in the NM library involved in necessary cellular activities, only a few are under direct control of the biological clock.

Table 4.01 Enzymes detected in three major energy pathways in two cDNA libraries

Pathway	NE	NM
Glycolysis /Gluconeogenase	Glucokinase Fructose-bisphosphate aldolase Glyceraldehyde-3-phosphate dehydrogenase Phosphoglycerate kinase Pyruvate kinase Phosphoenolpyruvate carboxykinase Enolase Phosphoenolpyruvate carboxylase Triose-phosphate isomerase Pyruvate carboxylase	Glucokinase Fructose-bisphosphate aldolase Glyceraldehyde-3-phosphate dehydrogenase Phosphoglycerate kinase Pyruvate kinase Phosphoenolpyruvate carboxykinase Enolase Phosphoenolpyruvate carboxylase Hexokinase Phosphoglycerate mutase Phosphopyruvate hydratase
Pentose-phosphate pathway	Phosphogluconate dehydrogenase Transketolase Transaldolase	
TCA	Isocitrate dehydrogenase (NADP+) succinyl-CoA ligase (GDP-forming) Aconitase 2-oxoglutarate dehydrogenase alpha-ketoglutarate dehydrogenase dihydrolipoamide succinyltransferase fumarase malate dehydrogenase	Isocitrate dehydrogenase (NAD+) succinyl-CoA ligase (GDP-forming) ATP-specific succinyl-CoA synthetase

Table 4.02 The enzymes missed in three major energy pathways in two cDNA libraries

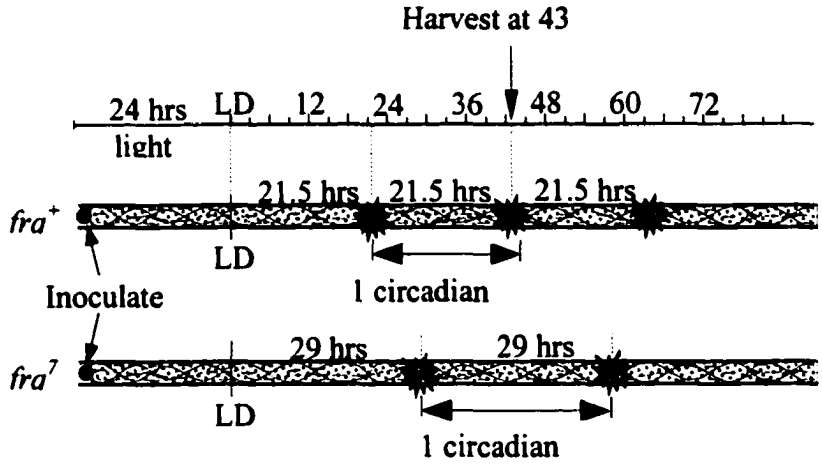
Pathways	NE	NM
Glycolysis /Gluconeogenase	Hexokinase 6-phosphofructokinase Fructose-biphosphatase Phosphoglycerate mutase Phosphoglucose isomerase	6-phosphofructokinase Fructose-biphosphatase Triose-phosphate isomerase Phosphoglucose isomerase Pyruvate carboxylase
Pentose-phosphate pathway	Glucose 6-phosphate dehydrogenase Lactonase Phosphopentose epimerase Phosphopentose isomerase	Phosphogluconate dehydrogenase Transketolase Transaldolase Glucose 6-phosphate dehydrogenase Lactonase Phosphopentose epimerase Phosphopentose isomerase

TCA	Citrate synthase Succinate dehydrogenase	Citrate synthase Aconitase 2-oxoglutarate dehydrogenase Succinate dehydrogenase fumarase, malate dehydrogenase

Several possible factors could cause the difference of the gene expression between these two cDNA libraries. First, the two strains used for the construction of the studied cDNA libraries of *Neurospora* were different. The *frq*^{7} mutation strain that has a 29 hours of long period of circadian rhythm was used for the construction of the NM cDNA library while the wild type *frq*+ strain that has a 21.5 hours circadian rhythm was used for the construction of the NE cDNA library. Even though the *frq* allele mutation only was expected to change the rhythm of *Neurospora* clock but have no other effect on phenotype (Bell-Pedersen et al., 1996b), this mutation may alter the redundancy and the variety of gene expression in this cDNA library. Second, an EST project has its own limit for the representative of the genes expressed in the studied cDNA library, Since even though the sampling of ESTs continued until the redundancy reached above 99% (table 3.05, Figure 3.01), this approach can not guarantee that all rarely expressed genes are represented in the cDNA library. Therefore, although it is obvious that when an expressed gene is detected in a cDNA library, if a gene is not detected, it can not be concluded that this gene is not expressed in the studied cDNA library.

Even with above taken into account different expression profiles were observed between these two cDNA libraries and these differences may relate to the different biological clock phases for the two strains used for the cDNA construction. For the wild

type *frq*+ strain used for the NE cDNA library, at the time of harvest at DD 43 hours, a time point at which sporulation is beginning since conidia would be produced late from this mycelium. In contrast, the cDNA were collected from strain frq^7 while it was in the middle of two cycles of conidiation and the mycelium of this stage do not produce conidia (Figure 4.04). Thus, for sporulation, the expression of many genes needs to be activated as cell growth, cell division is much more active during the sporulation phase than in the growth stage between two sporulation cycles. Therefore, expression of genes producing components for the protein synthesis machinery and those involved in energy metabolism clearly are much more active in the NE library than in the NM library.



LD: *Light to dark transfer*

Figure 4.04 Diagram that shows the different growth stages of two *Neurospora crassa* strains at the harvest time in constant darkness 43 hours after LD transfer. Growing for 24 hours in light after inoculation allows the same circadian time for these two strains.

This may explain why there were more 1000 different expressed genes detected in the NE cDNA library while only 377 different expressed genes were detected in the NM cDNA

library and many metabolic pathways related expressed genes were absent in the NM cDNA library when compared to in the NE cDNA library. In addition, two conidiation-specific genes, *con-6* and *con-10* gene, also were detected only in the NE cDNA library, an observation in agreement with above conclusion.

Eleven clock-controlled genes were detected in this research. Four of them (*ccg-13, ccg-14, ccg-15* and *lysozyme*) were newly identified based on this research, and 7 of them were identified previously (Bell-Pedersen et a., 1996c). Except for the *ccg-4* gene and *ccg-6* gene, all the *ccgs* detected in this research had an expected higher expression level in the NM cDNA library. This observation indicates that the clock set in the strain for the NM cDNA library may prefer the transcription of genes that are under control of the *Neurospora* clock. Interestingly, the *ccg-7* gene, which encodes the glyceraldehyde 3-phosphate dehydrogenase, and is a housekeeping gene involved in the glycolysis pathway, has higher expression level in the NM library than in the NE library (table 3.17 and 3.39).

The identity of the *ccg-4* gene now has been determined as part of this research. In previous studies, this clock-controlled gene was reported to have two putative translation products, *ccg-4* protein 1 and *ccg-4* protein2. Based on this research, the identity of the *ccg-4* gene is a *Neurospora crassa* homolog of *Sordaria macrospora* pheromone precursor 1 gene (*ppg-1*). This finding is very significant in the studies of circadian rhythms. Pheromones are proteins involved in the mating activities and sexual cycle. Therefore, since the *ccg-4* gene is a mating type specific pheromone, it is likely that pheromone biosynthesis is under control of the clock, and both the sexual cycle and

meiosis also may be regulated by the biological clock (personal communication with Dr.Bell-Pedersen). The observation that many more clock-controlled genes are involved in different metabolic pathways than could be identified in earlier studies also is consistent with the hypothesis that the regulation and control of gene expression by the clock is much more pervasive in the cell than expected (Dunlap, 1996).

In addition, since the *ccg-6* gene had a higher expression level in the NE cDNA library, its transcription product might be involved in the activities related to sporulation.

The sequence comparison between the genomic DNA and the cDNA of the clock-controlled genes can help to discover the control and regulation elements in the promoter sequence of *ccg* genes. The features and the functions of these elements in the promoter region of *ccgs* can aid in to detecting and identifying additional clock-controlled genes. This information also provides insight into understanding the interactions between the clock gene and the clock-controlled genes, and to uncover how the clock gene affects the transcription and expression of the *ccgs* through interactions between the protein-binding sites on the regulation sequence and proteins such as white-collar protein 1 and 2. For example, the GATA repeat motif is an important element on the promoter sequence of the *Arabidoposis* clock-controlled gene CAB2 (Anderson et al., 1994). Two tobacco nuclear proteins can bind to the protein-binding site that contains the GATA repeat motif (Anderson and Kay, 1995). Thus, this sequence has important function in the light signal transduction pathway and in mediating the Cab2 expression in response to phytochrome activation, is as well as in the light-induced high-amplitude circadian oscillation in Cab2 expression (Anderson and Kay, 1995; 1997). In *Neurospora*, the clock gene product,

FRQ protein, interacts with the GATA motif and a the putative GATA motif was also detected in several clock-controlled genes identified in this research. Therefore, the GATA motif might link clock gene product expression and clock-controlled gene expression in *Neurospora* as well.

Through the sequence comparisons between the genomic DNA and the cDNA of the expressed *ccg* genes identified in this research, both the alternative splicing and the alternative polyadenylation were observed. With the exception of the *ccg-8* gene transcript, other *ccg* genes transcripts, *ccg-1*, *2*, *4*, *7*, *6*, *9* had instances of alternative polyadenylation in their RNA processing and the *ccg-1* gene transcript was shown to be alternatively spliced. Therefore more than one gene product is produced which may result in differential translation levels.

During the sequence comparisons between the *ccg-6* mRNA and the cDNA represented by the contig1435c, a base "C" deletion was found in the *ccg-6* mRNA sequence (U46086). Through the sequence comparisons among the *ccg-9* gene genomic DNA (Af088906), *ccg-9* mRNA (U46088) and the cDNA represented by the newccg9contig2, a base "T" deletion was found in the genomic DNA sequence of the *ccg-9* gene while a base "T" addition was found in the *ccg-9* mRNA sequence. The cDNAs represented by these two contigs of these two *ccg* genes may indicate that these two *ccg* genes also may result in more than one gene product.

In the genomic cosmid sequencing portion of this research, 10 genes were predicted in the 37 Kbp nc17 cosmid by GENSCAN, only three of which had homologs in the GenBank nr protein database. Accordingly and consistent with earlier predictions,

it is likely that as many as 11,000 genes will be present in the 42.5 Mb genome of *Neurospora crassa*.

More than 50% of genes identified in this research are novel, previously unknown genes. The *Neurospora crassa* genome is being sequenced at the Whitehead Institute at the Massachusetts Institute of Technology and the initial working draft sequence only recently has become available. The EST sequences determined in this dissertation research and made publicly available will aid greatly in the annotation of this sequence as well as permit identification of possible protein-binding sites on the promoter sequences of these genes.

Chapter V

Literature cited

ABI prism®BigDye™ terminator cycle sequencing ready reaction kits protocol, 2000 copyright.

Abola et al.(1987). In Crystallographic databases. Allen F.H. et al.(eds). Cambridge, pp.107-132.

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

Adams, M. D. *et al.* (1995). Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. ***Nature 377***: 3-21.

Adams, M. D., Dubnick, M., Kerlarage, A. R., Moreno, R., Kelley, J. M., Utterback, T. R., Nagle, J. W., Field, C. and Venter, J. C. (1992). Sequence identification of 2375 human brain genes. ***Nature 355***: 632-634.

Adams, M. D., Kerlavage, A. R., Field, C., Venter, J. C. (1993a). 3400 new expressed sequence tags identify diversity of transcripts in human brain. ***Nature Genet 4***: 256-267.

Adams, M. D., Soares, M. B., Kerlavage, A. R., Field, C., Venter, J. C. (1993b). Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. ***Nature Genet 4***: 373-380.

Allen, R.W., Trach, K.A. and Hoch, J.A.(1987). Identification of the 37-kDa protein displaying a variable interaction with the erythroid cell membrane as glyceraldehyde-3-phosphate dehydrogenase. ***J. Biol. Chem. 262***: 649-53.

Altchul, 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.

Amplitaq-FS Polymerase Protocol for use in double-stranded cycle sequencing, 1998. In Protocol: PE Applied Biosystems, pp. 45.

Anderson, S.L. and Kay, S.A. (1995). Functional dissection of a circadian clock- and phytochrome- regulated transcription of the Arabidopsis CAB2 gene. ***Proc. Natl. Acad. USA 92***: 1500-1504.

Anderson, S.L. and Kay, S.A. (1997). Phototransduction and circadian clock pathways regulating gene transcription in higher plants. ***Adv Genet 35***: 1-34.

Anderson, S.L., Teakle, G.R., Martino-Catt, S.J. and Kay, S.A. (1994). Circadian clock- and phytochrome-regulated transcription is conferred by a 78-bp cis-acting domain of the

Arabidopsis CAB2 promoter. ***Plant J. 6***: 457-470.

Applied Biosystems Incorporated Users Manual ABI 373, 1994.

Applied Biosystems Incorporated Users Manual ABI 377, 1998.

Aral, B., Schlenzig, J.S. and Kamoun, P. (1996). Database cloning human delta 1-pyrroline-5-carboxylase synthtase (P5CS) cDNA: a bifunctional enzyme catalyzing the first 2 steps in proline biosynthesis. ***C R Acad Sci III 319***: 171-178.

Aramayo, R., Peleg, Y., Addison, R. and Metzenberg, R.(1996). *Asm-1*$^{+}$, a *Neurospora crassa* gene related to transcriptional regulators of fungal development. ***Genetics 144***: 991-1003.

Aronson, B. D., Johnson, K. A., and Dunlap, J. C. (1994). Circadian clock locus frequency: protein encoded by a single open reading frame defines period length and temperature compensation. ***Proc. Natl. Acad. Sci. USA 91***: 7683-7687.

Arpaia, G., Loros, J.J., Dunlap, J.C. and Morelli, G.(1995). Light induction of the clock-controlled gene *ccg-1* is not transduced through the circadian clock in *Neurospora crassa*. ***Mol Gen Genet 247***: 157-163.

Arpaia, G., Loros, J.J., Dunlap, J.C., Morelli, G. and Macino, G. (1993). The interplay of light and the circadian clock: independent dual regulation of clock-controlled gene *ccg-2* (*eas*). ***Plant physiol 102***: 1299-1305.

Audic, S. and Clasverie, J.M (1997). The significance of digital gene expression profiles. ***Genome Research*** *7*: 986-995.

Bailey, 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.

Bairoch, A. and Apweiler, R. (1997). The SWISS-PROT protein sequence databank and its supplement TrEMBL. ***Nucleic Acids Res. 25***: 31-36.

Ballabio, A. (1993). The rise and fall of positional cloning? ***Nature Genetics 3***: 277-279.

Ballario, P. and Macina, G. (1997). While collar proteins: PASing the light signal in *Neurospora crassa*. ***Trends in Microbiology 5***: 458-462.

Ballario, P., Vittorioso, P., Magrelli, A., Talora, C., Cabibbo, A. and Macino,G.(1996). White collar-1, a central regulator of blue light response in *Neurospora*, is a zinc finger protein. ***EMBO J 15***: 1650-1657.

Banfi, S., Guffanti, A. and Borsani, Giuseppe. (1998). How to get the best of dbEST. ***TIG 14***: 80-81.

Bargmann, C. I. (1992). cDNA Sequencing: A report from the worm front. ***Nature Genetics 1***: 79-80.

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

Beadle, G. W. and Tatum, E. L. (1941). Genetic control of chemical reactions in *Neurospora*. ***Proc. Natl. Acad. Sci. USA 27***: 499-506.

Beaudoing, E., Freier, S., Wyatt, J.R., Claverie, J.M. and Gautheret, D.(2000), Patterns of variant polyadenylation signal usage human genes. ***Genome Research 10***: 1001-1010.

Beever, R.E. and Dempsey,G.P. (1978). Function of rodlets on the surface of fungalspores. ***Nature 272***: 608-610.

Begum, F. and Samajpati, N. (2000). Mycotoxin production on rice, pulses and oilseeds. ***Naturwissenschaften 87*** :275-7.

Bell-Pedersen, D.(1998). Keeping pace with *Neurospora* circadian rhythms. ***Microbiology 144***: 1699-1711.

Bell-Pedersen, D., Dunlap, J. C., and Loros, J. J. (1992). The *Neurospora* circadian clock-controlled gene, *ccg-2*, is allelic to *eas* and encodes a fungal hydrophobin required for formation of the conidial rodlet layer. ***Genes & Development 6***: 2382-2394.

Bell-Pedersen, D., Dunlap, J.C., Loros, J.J. (1996a). Distinct cis-acting elements mediate clock, light, and developmental regulation of the *Neurospora crassa eas* (*ccg-2*) gene. ***Molecular and Cellular Biology 16***: 513-521.

Bell-Pedersen, D., Garceau, N. and Loros, J.J. (1996b). Circadian rhythms in fungi. ***J. Genet. 75***: 387-401.

Bell-Pedersen, D., Shinohara, M. L., Loros, J. J., and Dunlap, J. C. (1996c). Circadian clock-controlled genes isolated from *Neurospora crassa* are late night to early morning-specific. ***Proc. Natl. Acad. Sci. USA 93***: 13096-13101.

Bensen, D.A., Boguski, M.S., Lipman, D.J., Ostell, J. and Ouellette, B.F.F. (1998). GenBank. ***Nucleic Acids Res. 26***: 1-7.

Berger, S. L. (1987a). In Methods in enzymology, guide to molecular cloning techniques, S. L. Berger & A. R. Kimmel (eds). Academic Press, San Diego, pp. 215-219.

Berger, S. L. (1987b). In Methods in Enzymology, Guide to Molecular Cloning Techniques, S. L. Berger & A. R. Kimmel (eds). Academic Press, San Diego, pp. 227-234.

Berlin, V., and Yanofsku\y, C. (1985b). Isolation and characterization of genes

differentially expressed during conidiation of *Neurospora crassa*. ***Molecular & Cellular Biology 5***: 849-855.

Berlin, V., and Yanofsky, C. (1985a). Protein changes during the asexual cycle of *Neurospora crassa*. ***Molecular & Cellular Biology 5***: 839-848.

Berry, R., Stevens, T. J., Walter, N. A. R. et al. (1995). Gene-based sequence-tagged-sites (STSs) as the basis for a human gene map. ***Nature Genetics 10***: 415-423.

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

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

Boguski, M. S. and Tolstoshev, C.(1994). Gene discovery in dbEST. ***Science 265***: 1993-1994.

Boguski, M.S. (1995). The turning point in genome research. ***Trends in Biochemical Science 20***: 295-6.

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

Brann, M.R. and Cohen, L.V. (1987). Diurnal expression of transducin mRNA and translocation of transducin in rods of rat retina. ***Science 235***: 585-587.

Brody, L. C., Abel, K.J., Castilla, L.H., Cauch, F.J., McKinley, D.R., Yin, G., Ho, P.P., Merajver, S., Chandrasekharappa, S.C., Xu, J. et al. 1995. Construction of a transcription map surrounding the BRCA 1 locus of human chromosome 17. ***Genomes 25***: 238-247.

Brown, S.M.(2000). Bioinformatics: a biologist's guide to biocomputing and the internet. Unpublished.

Brown, T.A. (1999). Genomes. John Wiley & Sons Inc., New York, NY10158.

Bucher, P., Karplus, K., Moeri, N. and Hofmann, K. (1996) A flexible search technique based on generalized profiles. ***Computers and Chemistry 20***: 3-24.

Bucher,P. (1999). Gene feature identification. In: Genetics database, edited by Bishop, M.J., pp135-164. Academic Press, CA92101.

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

Burge, C. and Karlin, S. (1998). Finding the genes in genomic DNA. ***Curr. Opin. Struct.***

Biol. 8:340-354.

Carlile, M.J. and Watkinson, S.C.(1996). The Fungi. Academic Press, San Diego, CA 92101.

Carre, I.A. and Kay,S.A.(1995). Multiple DNA-protein complexes at a circadian-regulated promoter element. ***Plant Cell*** 7: 2039-2051.

Chen, F. (1997). Sequence and analysis of approximately 0.5 mega-base pair of the DiGeorge Syndrome critical region in human chromosome 22 band 11 and a syntenic mouse Bac. Ph.D dissertation, the University of Oklahoma, Norman, USA.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. ***Biochemistry 18***: 5294-5299.

Christopher, D.A, Kim, M. and Mullet, J.E. (1992). A novel light-regulated promoter is conserved in cereal and dicot chloroplasts. ***Plant Cell 4***: 785-798.

Cirera, S., Winterϕ, A.K. and Fredholm, M. (2000). Why do we still find anonymous ESTs? ***Mammlian Genome 11***: 689-693.

Cohen, S. M., Chang, A. C. Y., and HSO, L. (1972). Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. ***Proc. Natl. Acad. Sci. USA 69***: 2110-2114.

Collins, F. S. (1992). Positional cloning: let's not call it reverse anymore. ***Nature Genetics 1***: 3-6.

Collins, F. S. (1995). Positional cloning moves from perditional to traditional. ***Nature Genetics 9***: 347-350.

Corrochano, L.M., Lauter, F.R., Ebbole, D.J. and Yanofsky, C. (1995). Light and developmental regulation of the gene *con-10* of *Neurospora crassa*. ***Dev. Biol. 176***: 190-200.

Crosthwaite, S. K., Dunlap, J. C., and Loros, J. J. (1997). *Neurospora* WC-s and WC-2: transcription, photoresponses, and the origins of circadian rhythmicity. ***Science 276***: 763-769.

Denault, D.L., Loros, J.J., Dunlap, J.C. (2001). WC-2 mediates WC-1-FRQ interaction within the PAS protein-linked circadian feedback loop of *Neurospora*. ***EMBO J 20***: 109-117.

Deutch, A.H., Rushow, K.E. and Smith, C.J. (1984). Analysis of the *Escherichia coli*

proBA locus by DNA and protein sequencing. ***Nucleic Acids Res. 12***: 6337-6355.

Duggan, D.J., Bittner, M., Chen, Y.D., Meltzer, P. and Trent, J.M. (1999). Expression profiling using cDNA microarrays. ***Nature Genetics 21***: 10-14.

Dunlap, J. C. (1990). Closely watched clocks: molecular analysis of circadian rhythms in *Neurospora* and Drosophila. ***TIG 6***: 159-165.

Dunlap, J. C. (1993). Genetic analysis of circadian clocks. ***Annu. Rev. Physiol. 55***: 683-728.

Dunlap, J. C. (1996). Genetic and molecular analysis of circadian rhythms. ***Annu. Rev. Genet 30***: 579-601.

Dunlap, J. C. (1999). Molecular bases for circadian clocks. ***Cell 96***: 271-290.

Dunlap, J. C., and Feldman, J. F. (1988). On the role of protein synthesis in the circadian clock of *Neurospora crassa*. Proc. ***Natl. Acad. Sci. USA 85***: 1096-1100.

Dunlap, J. C., Loros, J. J., Merrow,M., Crosthwaite, S., Bell-Pedersen, D., Garceau, N., Shinohara, M., Cho, H., Luo, C. (1996). The genetic and molecular dissection of a prototypic circadian system. ***Progress in Brain Research 111:***11-27.

Dunlap, J. C., Loros, J.J., Aronson, B.D., Merrow, M., Crosthwaite, S., Bell-Pedersen, D., Johnson, K., Lindgren, K. and Garceau, N.Y.(1995). The Genetic basis of the circadian clock: identification of *frq* and *FRQ* as clock components in *Neurospora*. ***Ciba Foundation Symposium 183***: 3-25.

Dunlap, J.C. (1998). Common threads in eukaryotic circadian systems. ***Curr. opin. Genet. Dev*** 8: 400-406.

Dynan, W.S. (1989). Modularity in promoters and enhancers. ***Cell 58***: 1-4.

Ebbole, D.J. (1998). Carbon catabolite repression of gene expression and conidiation in *Neurospora crassa*. ***Fungal Genet. Biol 25***: 15-21.

Eberle, J and Russo, V.E.A. (1992). *Neurospora crassa* blue-light inducible gene *bli-7* encodes a short hydrophobic protein. ***DNA Sequencing 3***: 131-141.

Edmunds, L.N. (1988). Cellular and Molecular Bases of Biological Clocks. Springer-Verlag, New York.

Edwards-Gilbert, G., Veraldi, K.L. and Milcarek, C. (1997). Alternative poly(A) site selection in complex transcription units: means to an end? ***Nucleic Acid Res. 25***: 2547-2561.

Eizinger, A., Jungblut, B. and Summer, R.J. (1999). Evolutionary change in the

functional specificity genes. ***Trends Genet 15***: 197-202.

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

Ewing, B., Hillier, L., Wendl, M. C. and Green, P. (1998). Base-Calling of Automated Sequencer Traces Using Phred I. Accuracy Assessment. ***Genome Research 8***: 175-185.

Frederick C. Neidhardt, *et al.* Escherichia coli and Salmonella., cellular and molecular biology. 2nd Edition, Vol. 2, 1996. ASM Press, Washington, D. C. (Monica Riley).

Gallie, D.R. (1991). The cap and poly (A) tail function synergistically to regulate mRNA translational efficiency. ***Genes Dev 5***: 2108-2116.

George, D.G., Barker, W.C., Mewes, H.W., Pfeiffer, F. and Tsugita, A. (1996). The PIR-International Protein Sequence Database. ***Nucleic Acids Res. 24***: 17-20.

Gerhold, D., Rushmore, T. and Caskey, C.T.(1999). DNA chips: promising toys have become powerful tools. ***TIBS 24***: 168-173.

Gidoni, D., Brosio, P., Bond-Nutter, D., Bedbrook, J. and Dunsmuir, P.(1989). Novel cis-acting elements in Petunia *Cab* gene promoters. ***Mol Gen Genet 215***: 337-344.

Gillette, M.U.(1997). Cellular and biochemical mechanisms underlying circadian rhythms in vertebrates. ***Current Opinion in Neurobiology*** *7*: 797-804.

Giuliano, G., Hoffman, N. E., Ko, K., Scolnik, P. A., and Cashmore, A. R. (1988). A light-entrained circadian clock controls transcription of several plant genes. ***The EMBO Journal*** *7*: 3635-3642.

Gleeson, T.J. and Staden, R. (1991) An x window and unix implementation of our sequence analysis package. ***Compt. Appl. Biosci*** *7*: 398-401.

Golden, S.S., Ishiura, M., Johnson, C.H. and Kondo, T.(1997). *Cyanobacterial* circadian rhythms. ***Annu Rev Plant Physiol Plant Mol Biol 48***: 327-354.

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

Green, C.B. and Besharse, J.C.(1996). Identification of a novel vertebrate circadian clock-regulated gene encoding the protein nocturnin. ***Proc. Natl. Acad. Sci. USA 93***: 14884-14888.

Green, P., and Ewing, B. (Copyright 1994-1999) Phrap Documentation.

Gribskov, M.(1999). Sequence comparison. In: Genetics database, edited by Bishop,

M.J., pp105-120. Academic Press, CA92101.

Gribskov, M., McLachlan, A.D. and Eisenbergm D. (1987) Profile analysis: detection of distantly related protein. ***Proc. Natl. Acad. Sci.USA 84***: 4355-4358.

Gurr, S.J., Unkles, S.E. and Kinghorn, J.R. (1987). The structure and organization of nuclear genes of filamentous fungi. In: Gene structure in eukaryotic microbes. Edited by Kinghorn, J.R. Andrews, Fife KY16, 9AL, UK.

Haley, J. and Bogorad, L. (1990). Alternative promoters are used for genes within maize chloroplast polycistronic transcription units. ***Plant Cell 2***: 323-333.

Hastie, N.D. and Bishop, J.O.(1976). The expression of the three abundance classes of messenger RNA in mouse tissue. ***Cell 9***: 761-774.

Henikoff, S and Henikoff, J.G. (1994). Protein family classification based on searching a database of blocks. ***Genomics 19***: 97-107.

Hillier, L et al., (1996). Generation and analysis of 280,000 Human Expressed Sequence Tags. ***Genome Research 6***: 807-838.

Hua, A.(1999). Comparative sequencing of human and mouse syntenic regions from DGCR and MND2. Ph.D dissertation, the University of Oklahoma, Norman, USA.

Hunt, A.E., Al-Ghoul, W.M., Gillette, M.U., Dubocovich, M.L.(2001). Activation of MT(2) melatonin receptors in rat suprachiasmatic nucleus phase advances the circadian clock. ***Am. J. Physiol. Cell. Physiol. 280*** :C110-8.

Hwung, D.M., Fung, Y.W., Wang, R.X., Laurenssen, C.M and Ng SH et al.(1995). Analysis of expressed sequence tags from fetal human heart cDNA library. ***Genomics 30***: 293-298.

Innis, M. A., Myambo, K. B., Gelfand, D. H., and Drow, M. A. D. (1988), DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase and direct sequencing of polymerase chain reaction-amplified DNA. ***Proc. Natl. Acad. Sci. USA 85***: 9436-9440.

Jackson, R. and Standart, N. (1990). Do the poly(A) tail and 3' untranslated region control mRNA translation? ***Cell 62***: 15-24.

Jakobsen, K. S., Breivold, E. & Hornes, E. (1990). Purification of mRNA directly from crude plant tissues in 15 minutes using magnetic oligo dT microspheres. ***Nucleic Acids Res 18***: 3669.

Jiang, J. and Jacob, H. (1998). EbEST: An automated gene structure. ***Genome Research 8***: 268-275.

Jungehulsing, U., Arntz, C., Smit, R. and Tudzynski, P. (1994). The claviceps purpurea glyceraldehyde-3-phosphate dehydrogenase gene: cloning, characterization, and use for the improvement of a dominant selection system. ***Curr. Genet.25***: 101-106.

Kanehisa, M. (1997). Kegg: Kyoto "Encyclopedia of Genes and Genomes". Release 4.0. ***Trends Genet. 13***: 375-376.

Katagiri, S., Onai, K. and Nakashima, H. (1998). Spermidine determines the sensitivity to the calmodulin antagonist, chlorpromazine, for the circadian conidiation rhythm but not for the mycelial growth in *Neurospora crassa*. ***J Biol Rhythms 13***: 452-60.

Kawamoto, R.M. and Caswell, A.H.(1986). Autophosphorylation of glyceraldehyde phosphate dehydrogenase and phosphorylation of protein from skeletel muscle microsomes. ***Biochemistry 25***: 657-661.

Kershaw, M.J. and Talbot, N.J.(1998). Hydrophobins and repellents: proteins with fundamental roles in fungal morphogenesis. ***Fungal Genetics and Biology 23***: 18-33.

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

Khan, J., Saal, L.H., Chen, Y.D., Trent, J.M. and Meltzer, P.S. (1999). Expression profiling in cancer using cDNA microarrays. ***Eletrophoresis 20***: 223-229.

Klenow, H., Overgaard-Hansen, K., Patkar, S.A.(1971). Proteolytic cleavage of native DNA polymerase into two different catalytic fragments. Influence of assay conditions on the change of exonuclease activity and polymerase activity accompanying cleavage. ***Eur. J. Biochem. 22***: 371-381.

Klich, M.A., Mullaney, E.J., Daly, C.B. and Cary, J.W. (2000). Molecular and physiological aspects of aflatoxin and sterigmatocystin biosynthesis by *Aspergillus tamarii* and *A. ochraceoroseus*. ***Appl Microbiol Biotechnol 53***: 605-9.

Ko, M. S. H. (1990). An equalized cDNA library by reassociation of short double stranded cDNAs. ***Nucleic Acids Research 18***: 5705-5711.

Krug, M. S. & Berger, S. L. (1987). In Methods in Enzymology, Guide to Molecular Cloning Techniques, S. L. Berger & A. R. Kimmel (eds) Academic Press, San Diego, pp. 316-325.

Kupfer, D. M. (1999). Development, analysis and use of an expressed sequence tag database from the multicellular ascomycete, *Aspergillus nidulans*. Ph.D dissertation, the University of Oklahoma, Norman, OK, USA.

Kupfer, D. M., Reece, C. A., Clifton, S. W. and Roe, B. A. (1997). Multicellular

ascomycetous fungal genomes contain more than 8000 Genes. ***Fungal Genetics and Biology 21***: 364-372.

Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. ***J. Mol. Biol. 157***: 105-132.

Lakin-Thomas, P., Coté, G. G., and Brody, S. (1990). Circadian rhythms in *Neurospora crassa*: biochemistry and genetics. ***Critical Reviews in Microbiology*** *7*: 365-416.

Lauter, F.R. and Yanofsky, C. (1993). Day/night and circadian rhythm control of *con* gene expression in *Neurospora*. ***Proc. Natl. Sci. USA 90***: 8249-8253.

Lauter, F-R., Russo, V. E. A., and Yanofsky, C. (1992). Developmental and Light Regulation of Eas, the Structural Gene for the Rodlet Protein of *Neurospora*. ***Genes & Development 6***: 2373-2381.

Lee, F-J.S., Lin, L.W and Smith, J.A. (1990). A glucose-repressible gene encodes acetyl-CoA hydrolase from *Saccharomyces cerevisiae*. ***J. of Biological Chemistry 265***: 7413-7418.

Lee, K, and Ebbole, D.J. (1998). Tissue-specific repression of starvation and stress responses of the *Neurospora crassa con-10* gene is mediated by RCO1. ***Fungal Genetics and Biology 23***: 269-278.

Lee, N. H. et al.(1995). Comparative expressed-tags analysis of differential gene expression profiles in PC-12 cells before and after nerve growth factor treatment. ***Proc. Natl. Acad. Sci. USA 92***: 8303-8307.

Lewin, B. (2000). Genes VII. Oxford University Press, New York.

Lewis, M. T., and Feldman, J. F. (1993). The Putative frequency (*frq*) clock protein of *Neurospora crassa* contains sequence elements that suggest a nuclear transcriptional regulatory role. ***Protein Seq. Data Anal. 5***: 315-323.

Lewis, M. T., Morgan, L. W., and Feldman, J. F. (1997). Analysis of frequency (*frq*) clock gene homologs: evidence for a helix-turn-helix transcription factor. ***Mol. Gen. Genetic 253***: 401-414.

Linden, H. and Macino, G. (1997). White collar 2, a partner in blue light signal transduction, controlling expression of light regulated genes in *Neurospora crassa*. ***EMBO J 16***: 98-109.

Link, A. J. and Olson, M. V. (1991). Physical map of the *Saccharomyces cerevisiae* genome at 110-kilobase resolution. ***Genetic 127***: 681-698.

Liu, B.H., Bhatnagar, D. and Chu, F.S. (1999). Purification and characterization of 40-kDa sterigmatocystin O-methyltransferase involved in aflatoxin biosynthesis. ***Nat.***

Toxins *7*: 63-9.

Loros, J. J., Denome, S. A., and Dunlap, J. C. (1989). Molecular cloning of genes under control of the circadian clock in *Neurospora*. ***Science 243***: 385-388.

Loros, J.J. (1998). Time at the end of the millennium: the Neurospora clock. ***Current Opinion in Microbiology*** *1*: 698-706.

Luo, C., Loros, J. J. and Dunlap, J. C. (1998). Nuclear localization is required for function of the essential clock protein *FRQ*. ***The EMBO Journal*** *17*: 1228-1235.

Madi, L., McBride, S.A., Baily, L.A. and Ebbole, D.J. (1997). *rco-1*, a gene involved in glucose transport and conidiation in *N. crassa*. ***Genetics 146***: 499-508.

Mandel, M. and Higa, A. (1970). Calcium dependent bacteriophage DNA infection. ***J. Mol. Biol. 53***: 154-159.

Martinez, P., Martin, W. and Cerff, R. (1989). Structure, evolution and anaerobic regulation of a nuclear gene encoding cytosolic glyceraldehyde-3-phosphate dehydrogenase from maize. ***J. Mol. Biol. 208***: 551-565.

Martin-Gallardo, A., McCombie, W.R., Gocayne, J.D., Fitz Gerald, M.G. et al.(1992). Automated DNA sequencing and analysis of 106 kilobases from human chromosome 19q13.3. ***Nat Genet.*** *1*: 34-39.

Matoba, R. et al. (1994). The addition of 5'-codingInformation to a 3'-directed cDNA library improves analysis of gene expression. ***Gene 146***: 199-207.

Maxam, A.M. and Gilbert, W. (1977). A new method for sequencing DNA. ***Proc. Natl. Acad. Sci. USA 74***: 560-564.

McCombie, W. R. et al. (1992). *Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease gene homologues. ***Nature Genetics*** *1*: 124-131.

McDonald, R. J., Swift, G., Przybyla, A. E., and Chirgwin, J. M. (1987). In Methods in Enzymology, Guide to Molecular Cloning Techniques, S. L. Berger & A. R. Kimmel (eds). Academic Press, San Diego, pp. 219-227.

McNally, M. and Free, S. (1988). Isolation and characterization of a *Neurospora* glucose repressible gene. ***Curr. Genet.14***: 545-551.

McWatters, H., Dunlap, J. C. and Millar, A. (1999). Circadian biology: clocks for the real world. ***Current Biology 9***: R633-R635.

Mechler, B. M. (1987). In Methods in Enzymology, Guide to Molecular Cloning Techniques, S. L. Berger, and A. R. Kimmel (eds). Academic Press, San Diego, pp. 241-

248.

Merrow, M. W., and Dunlap, J. C. (1994). Intergeneric complementation of a circadian rhythmicity defect: phylogenetic conservation of structure and function of the clock gene frequency. ***The EMBO Journal 13***: 2257-2266.

Meyer-Siegler,K., Mauro, D.J., Seal, G., Wurzer, J., deRiel, J.K. and Sirover, M.A. (1991). A human nuclear urail DNA glycosylase is the 37-kDa subunit of glyceraldehyde-3-phosphate dehydrogenase. ***Proc. Natl. Acad. Sci. USA 88***:8460-8464.

Millar, A.J. and Kay, S.A. (1997). The genetics of phototransduction and circadian rhythms in Arabidopsis. ***BioEssays 19***: 209-214.

Miller, K.Y., Wu, J. and Miller, B.L (1992). StuA is required for cell pattern formation in *Aspergillus*. ***Genes Dev 6***: 1770-1782.

Mishra, N.C. (1991). Genetics and molecular biology of *Neurospora crassa*. ***Advanced in Genetics 29***: 1-62.

Mitchell, P. et al. (1997). The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-5' exoribonuclease activities. ***Cell 91***: 457-466.

Moreira, A., Takagaki, Y., Brackenridge, S., Wollerton, M., Manley, J.L. and Proudfoot, N.J.(1998). The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. ***Genes Dev 12***: 2522-2534.

Moreno-Palanques, R.F. and Fuldner, R.A.(1993). Construction of cDNA libraries. In: Automated DNA sequencing and analysis techniques, edited by Venter, C., pp.102-109. Academic Press, London.

Mullis, K.B. and Faloon, F.A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. ***Method Enzymol.155***: 335-350.

Murtas, G. and Millar, A.J.(2000). How plants tell the time. ***Current Opinion in Plant Biology 3***: 43-46.

Nagy, E. and Rigby, W.F. (1995). Glyceraldehyde-3-phosphate dehydrogenase selectively binds AU-rich RNA in the NAD(+)-binding region (Rossmann fold). ***J.Biol.Chem.270***: 2755-2763.

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

Nelson, P.S., Ng WL., Schummer, M., True, L.D. and Liu, A.Y.(1998). An expressed-sequence-tag database of the human prostate: sequence analysis of 1168 cDNA clones.

Genomics 47: 12-25.

O'Hara, E.B. et al.(1995). Polyadenylation helps regulate mRNA decay in *E. coli*. ***Proc. Natl. Acad. Sci. USA 92*** : 1807-1811.

Ohlrogge, J. and Benning, C. (2000). Unraveling plant metabolism by EST analysis. ***Current opinion in Plant Biology 3***: 224-228.

Okubo, K. and Matsubara, K. (1993). Body mapping of expressing genes. In: Meyers, R. A. (Ed). ***Encyclopedia of Molecular Biology***. VCH, NY, NY, 1993.

Okubo, K., *et al.* (1992). Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. ***Nature Genetics 2***: 173-179.

Okubo, K., Hori, N., Matoba, R., Niiyama, T. and Matsubara, K. (1991). A novel system for large-scale sequencing of cDNA by PVR amplification. ***DNA Sequencing 2***: 137-144.

Okubo, K., Itoh, A., Fukushima, A., Yoshii, J. and Matsubara, K. (1995). Monitoring cell physiology by expression profiles and discovering cell type-specific genes by compiled expression profiles. ***Genomics 30***: 178-186.

Omori, K., Suzuki, S., Imai, Y and Komatsubara, S. (1991). Analysis of the *Serratia marcescens* proBA operon and feedback control of proline biosynthesis. ***J GEN Microbiol 137***: 509-17.

Pan, H.Q. (1996) Half million bases of DNA sequence in two meningioma regions in chromosome 22. Ph.D dissertation, the University of Oklahoma, Norman, Ok 73072.

Pandey, A and LeWitter, F (1999). Nucleotide sequence databases: a gold mine for biologists. ***TIBS 24***: 276-280.

Park, Y.S., Kwak, J.M., Kwon, O.Y., Kim, Y.S., Lee, D.S. et al.(1993). Generation of expressed sequence tags of random root cDNA clones of *Brassica napus* by single-run partial sequencing. ***Plant Physiol 103***: 359-370.

Patanjali, S. R., Parimoo, S., and Weissman, S. M. (1991). Construction of a Uniform-abundance (normalized) c DNA Library. ***Proc. Natl. Acad. Sci. USA 88***: 1943-1947.

Pauws, E., van Kampen, A.H.C., van de Graaf, S.A.R., de Vijlder, J.J.M. and Ris-Stalpers. (2001). Heterogeneity in polyadenylation cleavage sites in mammalian mRNA sequences: implications for SAGE analysis. ***Nucleic Acids Research 29***: 1690-1694.

Perkins, D.D. (1997). Chromosome rearrangements in *Neurospora crassa* and other filamentous fungi. ***Advanced in Genetics 36***: 239-398.

Pevet, P.(2000). Melatonin and biological rhythms. ***Biol Signals Recept 9(3-4)***: 203-12.

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.

Radford, A. and Parish, J. H. (1997). The genome and genes of *Neurospora crassa*. ***Fungal Genetics and Biology 21***: 258-266.

Reese, M.G. and Eeckman,F.H. (1995). Novel neural network algorithms for improved Eukaryotic promoter site recognition. In: The seventh international genome sequencing and analysis conference, Hilton Head Island, South Carolina.

Retief, J.D., Lynch, K.R. and Pearson, W.R.(1999). Panning for genes-a novel visual strategy for identifying novel gene orthology and paralogs. ***Genome Res 9***: 373-382.

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.

Roe, B.A. (1997). Protocols for recombinant DNA isolation, cloning, and sequencing. The University of Oklahoma, Norman, OK 73019.

Roenneberg, T. and Merrow, M. (1999) Circadian system and metabolism. ***J of Biological Rhythms 14***: 449-459.

Rosbash, M. (1995). Molecular control of circadian rhythms. ***Current Opinion in Genetics & Development 5***: 662-668.

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

Rowland, H.D. (2000). *Neurospora*: contributions of a model organism. Oxford University Press, new York, NY 10016.

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

Sambrook, J., Fritsch, E. F., and Maniatis, T.: Molecular Cloning. A Laboratory Press, Cold Spring Harbor, NY, 1989.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. ***Proc. Natl. Acad. Sci. USA 74***: 5463-5467.

Sargent, M.L., Briggs, W.R. and Woodward, D.O. (1966). Circadian nature of a rhythm

expressed by an invertaseless strain of *Neurospora crassa*. ***Plant Physiol.41***: 1343-1349.

Schmitt, A.O., Specht, T., Beckmann, G., Dahl, E., Pilarsky, C.P., Hinzmann, B. and Rosenthal, A. (1999). Exhaustive mining of EST libraries for genes differentially expressed in normal and tumor tissues. ***Nucleic Acids Res. 27***: 4251-4260.

Schwerdtfeger, C and Linden, H. (2000). Localization and light-dependent phosphorylation of white collar 1 and 2, the two central components of blue light signaling in *Neurospora crassa*. ***Eur. J. Biochem 267***: 414-421.

Selker, E.U.(1990). Premeiotic instability of repeated sequences in *Neurospora crassa*. ***Annu. Rev. Genet. 24***: 579-613.

Selkov, E. Jr., Grechkin, Y., Mikhailova, N. and Selkov, E. (1998). MPW: the metabolic pathways database. ***Nucleic Acids Res. 26***: 43-45.

Sharma, R.C. and Schimke, R.T.(1996) Preparation of electro-competent *E. coli* using salt-free growth medium. ***BioTechniques 20***:42-44.

Sharp, P.A.(1985). On the origin of RNA splicing and introns. ***Cell 42***: 397-400.

Shear, C.L. and Dodge, B.O. (1927). Life histories and heterothallism of the red bread-mold fungi of the *Monilia sitophila* group. ***J. Agric. Res. 34***: 1019-1042.

Shinohara, M. L., Loros, J. J. and Dunlap, J. C. (1998). Glyceraldehyde-3-phosphate dehydrogenase is regulated on a daily basis by the circadian clock. ***The Journal of Biological Chemistry 273***: 446-452.

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.

Sikela, J. M. & Ayffray, C. (1993). Finding new genes faster than ever. ***Nature Genetics 3***: 189-191.

Silver, R., Sookhoo, A.L., LeSauter, J., Stevens, P., Jansen, H.T. and Lehman, M.N. (1999). Multiple regulatory elements result in regional specificity in circadian rhythms of neuropeptide expression in mouse SCN. ***NeuroReport 10***: 3165-3174.

Singer, M. and Berg, P. (1991) Genes and genomes: a changing perspective. University Science Books press, Mill Valley, California.

Smith, L. M., Sander, J. Z., Kaiser, R. J., Huges, P., Dold, C., Connell, C. R., Heiner, C., Kemt, S. B. H., and Hood, L. E. (1986). Florescence detection in automated DNA sequence analysis. ***Nature 321***: 674-679.

Solovyev, V. and Salamov, A. (1997). The gene-finder computer tools for analysis of

human and model organisms genome sequences. ***Ismb 5***: 294-302.

Sonnhammer, E.L. and Durbin, R. (1995). A dot-matrix program with dynamic threshold control suited for genomic DNA protein sequence analysis. ***Gene 167***: GC1-10.

Springer, M. L. (1993). Genetic Control of Fungal Differentiation: The three sporulation pathways of *Neurospora crassa*. ***BioEssays 15***: 365-374.

Springer, M.L. and Yanofsky, C.(1989). A morphological and genetic analysis of conidiophore development in *Neurospora crassa*. ***Genes Dev. 4***: 559-71.

Stenesh, J. and Roe, B.A. (1972). DNA polymerase from mesophilic and thermophilic bacteria. I. purification and properties of DNA polymerase from *Bacillus licheniformis* and *Bacillus stearothermophilus*. ***Biochem.Biophys.Acta.272***: 156-166.

Stoesser, G., Moseley, M.A., Sleep, J. and McGowarn, M., Garcia-Pastor, M. and Sterk, P. (1998). The EMBL nucleotide sequence database. ***Nucleic Acids Res 26***: 8-15.

Stratagene instruction manual: cDNA synthesis kit, ZAP-cDNA synthesis kit, ZAP-cDNA Gigapack II Gold cloning kit and ZAP-cDNA gigapackIII gold cloning kit, 1996 copyright.

Stringer, M.A., Dean, R.A., Sewall, T.C. and Timberlake, W.E.(1991). Rodletless, a new *Aspergillus* developmental mutant induced by directed gene inactivation. ***Gene & Dev 5***: 1161-1171.

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

Sugano, S., Andronis, C., Green, R.M., Wang, Z.Y., Tobin, E.M.(1998). Protein kinase CK2 interacts with and phosphorylates the *Arabidopsis* circadian clock-associated 1 protein. ***Proc Natl Acad Sci USA 95***: 11020-11025.

Suzuki, S., Katagiri, S. and Nakashima, H.(1996). Mutants with altered sensitivity to a calmodulin antagonist affect the circadian clock in *Neurospora crassa*. ***Genetics 143***: 1175-80.

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***: 6447-58.

Tabor, S. and Richarson, C.C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. ***Proc. Natl. Acad. USA 84***: 4767-4771.

Tabor, S., Richardson, C. C. (1995). A Single residue in DNA polymerases of the Escherichia coli DNA polymerase 1 family is critical for distinguishing between deoxy- and dideoxyribonucleotides. ***Proc. Natl. Acad. Sci. USA 92***: 6339-6343.

Takahashi, J. S. and Kornhauser, J. M. (1993). Molecular approaches to understanding circadian oscillations. ***Annu. Rev. Physiol. 55***: 729-753.

Talora, C., Franchi, L., Linden, H., Ballario, P. and Macino, G. (1999). Role of a white collar –1-white collar-2 complex in blue-light signal transduction. ***EMBO J. 18***: 4961-4968.

Temin, H.M. and Baltimore, D. (1972). RNA-directed DNA synthesis and RNA tumor viruses. ***Adv Virus Res 17***: 129-86.

Templeton, M.D., Greenwood,D.R. and Beever, R.E.(1995). Solubilization of *Neurospora crassa* rodlet proteins and identification of the predominant protein as the proteolytically processed *eas (ccg-2)* gene product. ***Experimental Mycology 19***:166-169.

Templeton, M.D., Rikkerink, E.H., Solon, S.L. and Crowhurst, R.N.(1992). Cloning and molecular characterization of the glyceraldehyde-3-phosphate dehydrogenase-encoding gene and cDNA from the plant pathogenic fungus *Glomerella cingulata*. ***Gene 122***: 225-230.

Tomenchok, D.M. and Brandriss, M.C. (1987). Gene-enzyme relationships in the proline biosynthetic pathway of *Saccharomyces cerevisiae*. ***J Bacteriol 169***: 5364-5372.

Tracy, T. E. and Mulcachy, L. S. (1991). A simple method for direct automated sequencing of PCR fragments. ***Biotechniques 11***: 68-75.

Turian, G. and Bianchi, D.E. (1972). Conidiation in *Neurospora*. ***The Botanical Review 38***:119-154.

Van Gelder, R.N. and Krasnow, M.A.(1996). A novel circadianly expressed *Drosophila melanogaster* gene dependent on the period gene for its rhythmic expression. ***EMBO J 15***: 1625-1631.

Vanecek, J. and Watanabe, K. (1999). Mechanisms of melatonin action in the pituitary and SCN. ***Adv Exp Med Biol 460***: 191-8.

Wagner, A.(1997). A computational genomics approach to the identification of gene networks. ***Nucleic Acids Res. 25***: 3594-3604.

Wang, Z., Deck, M. and Free, S.J. (1994). A cis-acting region required for the regulated expression of *grg-1*, a *Neurospora* glucose-repressible gene. ***J. Mol. Biol. 237***: 65-74.

Waterston, R. et al. (1992). A survey of expressed genes in *Caenorhabditis elegans*. ***Nature Genetics 1***: 114-123.

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.

Welsh, J., Liu, J. P. & Efstratiadis, A. (1990). Cloning of PCR-amplified total cDNA: construction of a mouse oocyte cDNA library. ***Genet. Anal. Technol. Appl.*** *7*:5-17.

Wilcox, A. S., Khan, A.S., Hopkins, J.A. and Sikela, J.M.(1991). Use of 3' untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: implications for an expression map of the genome. ***Nucleic Acids Res. 19***: 1837-1843.

Winship, P.R. (1989). An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. ***Nucleic Acids Res. 17***: 1266.

Wolin, S.L. and Walther, P. (1991). Small ribonucleoproteins. ***Curr. Opin. Struct. Biol. 1***: 251-257.

Wong, L.J. and Marzluf, G.A.(1980). Sequence complexity and abundance classes of nuclear and polysomal polyadenylated RNA in *Neurospora crassa*. ***Biochem. Biophys. Acta 607***: 122-135.

Xu, Y., Mural, R.J., Shah, M. and Uberbacher, E.C. (1994). Recognizing exons in genomic sequence using GRAILII. ***Genetic Engneering 16***: 241-53.

Yamamoto, K. and Sasaki, T.(1997). Large-scale EST sequencing in rice. ***Plant Mo. Bio. 35***: 135-144.

Yamashiro, C.T., Ebble, D.J., Lee, B., Brown, R.E., Bourland, C., Madi, L. and Yanofsky, C. (1996). Characterization of *rco-1* of *N. crassa*, a pleiotropic gene affecting growth and development that encodes a homolog of TUP1 of *Saccharomyces cerevisiae*. ***Molecular and Cellular Biology 16***: 6218-6228.

Zate, M. (1993). Convergence and divergence in chick pineal regulation. ***Brain Res. 18***: 326-327.

Zhang, J. and Madden, T. L. (1997). PowerBLAST: a new application for interactive or automated sequence analysis and annotation. ***Genome Research*** *7*: 649-656.

Zhang, Z., Schaffer, A.A., Miller, W., Madden, T.L., Lipman, D.J., Koonin, E.V. and Altschul, S.F.(1998). Protein sequence similarity searches using patterns as seeds. ***Nucleic Acids Res. 26***: 3986-3990.

Zhu, H., Nowrousian, M., Kupfer, D., Colot, H.V., Berrocal-Tito, G., Lai, H., Bell-Pedersen, D., Roe, B.A., Loros, J.J. and Dunlap, J.C. (2001). Analysis of expressed sequence tags from two starvation, time-of-day-specific libraries of *Neurospora crassa* novel clock-controlled genes. ***Genetic 157***: 1057-1065.

Appendix I. *Neurospora crassa* morning library categories of cellular functions

I. Bioenergetics and Metabolisms

Part one: Metabolisms

A. Metabolism of carbohydrates(for glucose see energy)

1. Chitin metabolism
CHITIN SYNTHASE 3
&CHITIN-UDP ACETYL-GLUCOSAMINYLTRANSFERASE 3

2. Galactose metabolism
alpha-1,4 polygalactosaminidase

3. Mannitol metabolism
mannosyl-oligosaccharide 1,2-alpha-mannosidase
MANNOSYL-OLIGOSACCHARIDE ALPHA-1,2-MANNOSIDASE PRECURSOR

4. Sorbitol metabolism
SORBITOL DEHYDROGENASE
&sorbitol dehydrogenase

B. Metabolism of amino acids and related molecules

1. Glutamine metabolism
GLUTAMINE SYNTHETASE
glutaminase A—convert glutamine to glutamate by hydrolysis

2. Isoleucine metabolism
methylcrotonyl-CoA carboxylase
acyl-CoA dehydrogenase

3. Lysine metabolism
SACCHAROPINE DEHYDROGENASE

4. Serine metabolism
L-SERINE DEHYDRATASE--generates pyruvate for gluconeogenesis
&L-SERINE DEAMINASE

5. Tyrosine metabolism
TYROSINE DECARBOXYLASE 4

6. Beta-alanine metabolism
METHYLMALONATE-SEMIALDEHYDE DEHYDROGENASE
&ACYLATING
&MMSDH

7. Branch-amino acid metabolism
KETOL-ACID REDUCTOISOMERASE PRECURSOR--pathway for biosynthesis of branched-chain amino acids
&ALPHA-KETO-BETA-HYDROXYLACIL REDUCTOISOMERASE--SECOND STEP IN VALINE AND ISOLEUCINE BIOSYNTHESIS
&ACETOHYDROXY-ACIDREDUCTOISOMERASE

C. Metabolism of nucleotides and nucleic acids, purines, pyrimidines

1. Nitrogen metabolism (see also amino acid metabolism)
nitrite reductase --NEGATIVE REGULATORY PROTEIN IN THE NITROGEN CONTROL CIRCUIT PROTEIN
&NITRITE REDUCTASE
&NITROGEN METABOLIC REGULATION PROTEIN
&nitrogen metabolic regulation protein
&NMR PROTEIN
Nitrilase
NITRILASE 3
URICASE
&URATE OXIDASE
& urate oxidase
& EC1.7.3.3

D. Metabolism of cofactors, prosthetic groups

1. Thiamine
THIAMIN BIOSYNTHETIC ENZYME
&STRESS-INDUCIBLE PROTEIN STI35
&stress-inducible protein sti35
&THIAZOLE BIOSYNTHETIC ENZYME PRECURSOR
&thiazole biosynthetic enzyme

Part two: Energy

A. Carbohydrate as energy source

1. Glycolysis
1.1. Hexokinase
hexokinase
&HEXOKINASE

1.2. GLUCOKINASE
GLUCOKINASE
&GLUCOSE KINASE
&GLK

1.3. Fructose-bisphosphate aldolase
fructose-bisphosphate aldolase--also gluconeogenesis, PP cycle, carbon fixation, fructose and mannose metabolism
&FRUCTOSE-BISPHOSPHATE ALDOLASE
&fructose-1,6-bisphosphate aldolase

1.4. Glyceraldehyde-3-phosphate dehydrogenase
glyceraldehyde-3-phosphate dehydrogenase
&GLYCERALDEHYDE-3-PHOSPHATE DEGYDROGENASE
&GAPDH
&CLOCK-CONTROLLED PROTEIN 7
&clock-controlled gene 7 protein
&ccg-7 gene protein

1.5. Phosphoglycerate kinase
phosphoglycerate kinase

&PHOSPHOGLYCERATE KINASE

1.6. Phosphoglycerate mutase
phosphoglycerate mutase
&PHOSPHOGLYCERATE DEHYDRATASE

1.7. Phosphopyruvate hydratase
phosphopyruvate hydratase

1.8. Pyruvate kinase
pyruvate kinase
&PYRUVATE KINASE

1.9. ENOLASE
ENOLASE
&2-PHOSPHOGLYCERATE DEHYDRATASE
&2-PHOSPHO-D-GLYCERATE HYDRO-LYASE
&ALLERGEN CLA H 6
&CLA H VI

2. Gluconeogenesis
2.1. Phosphoenolpyruvate carboxykinase
phosphoenolpyruvate carboxykinase
&PEP carboxykinase
&PHOSPHOENOLPYRUVATE CARBOXYKINASE

3. Pyruvate metabolism
PYRUVATE DEHYDROGENASE E1 COMPONENT, ALPHA SUBUNIT PRECURSOR
&PDHE1-A
&PYRUVATE DEHYDROGENASE E1 COMPONENT, ALPHA SUBUNIT, MITOCHONDRIAL PRECURSOR

4. Tricarboxylic acid pathway
4.1. Isocitrate dehydrogenase
isocitrate dehydrogenase
&ISOCITRATE DEHYDROGENASE [NAD], MITOCHONDRIAL SUBUNIT 2 PRECURSOR
&ISOCITRIC DEHYDROGENASE
&NAD+-SPECIFIC ICDH
NAD-dependent isocitrate dehydrogenase subunit 2
NAD(+) -isocitrate dehydrogenase subunit I

4.2. SUCCINYL-COA LIGASE
SUCCINYL-COA LIGASE
&SUCCINYL-COA LIGASE, ALPHA-CHAIN PERCURSOR
&alpha subunit of succinyl-CoA ligase
&LSC1p
ATP-specific succinyl-CoA synthetase beta subunit

5. Fermentation, alcoholic
5.1. Alcohol dehydrogenase
alcohol dehydrogenase
ALCOHOL DEHYDROGENASE I

6. Metabolism of energy reserves (glycogen, starch, trehalose)
6.1. Glycogen degradation
glycogen phosphorylase

&AA 1-891
&EC 2.4.1.1
&Glycogen phosphorylase
&Gph1p
GLYCOSYLTRANSFERASE HOC1 PRECURSOR
PHOSPHOGLUCOMUTASE 1
&phosphoglucomutase (EC 5.4.2.2)
&GLUCOSE PHOSPHOMUTASE 1
&PGM 1

6.2 Starch degradation
glucan 1,4-alpha-glucosidase
&glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor
&GLUCAN 1,4-ALPHA-GLUCOSIDASE
&GLUCOAMYLASE PRECURSOR
&1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE

6.3. Trehalose degradation
NEUTRAL TREHALASE
&ALPHA,ALPHA-TREHALASE
&ALPHA,ALPHA-TREHALOSE GLUCOHYDROLASE

7. Related reactions
ATP citrate lyase

B. Metabolism of other energy sources
Nitrilase--catalyzing the terminal activation step in indole-acetic acid biosynthesis

C. Electron transport

1. Complex III-Ubiquinone to cytochrome C
ubiquinol-cytochrome c
& ubiquinol-cytochrome c reductase complex subunit
CYTOCHROME C OXIDASE POLYPEPTIDE II
NADH-UBIQUINONE OXIDOREDUCTASE 14.8 KD SUBUNIT
&COMPLEXI-14.8KD
&CI-14.8KD

2. ATP synthase
ATPase subunit 6
ATPase
&Rpt6p/26s PROTEASE REGULATORY SUBUNIT 8
PLASMA MEMBRANE ATPASE
& PROTON PUMP
&H+-transportingATPase
&H+-transporting ATPase
&V-ATPase
&p-ATPase
ATP SYNTHASE PROTEIN 9, MITOCHONDRIAL PRECURSOR
&LIPID-BINDING PROTRIN
ATP SYNTHASE DELTA CHAIN, MITOCHONDRIAL PRECURSOR

D. Reducing carriers

1. Gluathione
glutathione S-transferase 3

&microsomal glutathione S-transferase 3

2. Thioredoxin
thioredoxin
&THIOREDOXIN

II: GENE EXPRESSION AND GENETIC INFORMATION PROCESSING

A. DNA synthesis

1. DNA replication
minichromosome maintenance protein Mcm7p--The MCM proteins are essential replication initiation factors originally identified as proteins required for minichromosome maintenance in Saccharomyces cerevisiae

2. DNA modification and DNA repair
EXCINUCLEASE ABC SUBUNIT A--excision nuclease is an ABC complex containing all three Uvr protein
& UvrA protein
Hmp1
&mismatched base pair and cruciform DNA recognition protein (HMP1)

3. DNA packaging
3.1.Histone
Histones
histones H2A
HISTONE H4
&histone H4
HISTONE H3
&histone H3
HISTONE H2B
&histone H2B

B. Gene Expression

1. Transcription
1.1. Regulation
CROSS-PATHWAY CONTROL PROTEIN 1--contains segments similar to the DNA-binding and transcriptional activation domains of GCN4
&CPC1
homeodomain DNA-binding transcription factor
CCAAT/enhancer-binding protein
&C/EBP
ATP-DEPENDENT RNA HELICASE P47
&ATP-DEPENDENT RNA HELICASE P47 HOMOLOG
transcription factor
AP-1-like TRANSCRIPTION FACTOR
&transcription factor K1YAP1
TRANSCRIPTION FACTOR BTF3 HOMOLOG
RNA helicase
&Sub2p

1.2. RNA Processing
a. SPLICEOSOME
pre-mrna splicing factor atp-dependent rna helicase
&PRE-MRNA SPLICING FACTOR ATP-DEPENDENT RNA HELICASE KIAA0224

&HA4657
Lsm5 protein--associated with the U6 snRNA

1.3. tRNA synthesis and modifications
phenylalanyl-trna synthetase
& phenylalanyl-trna synthetase beta-subunit
&pheHB
ASPARTYL-TRNA SYNTHETASE
ASPARTYL-TRNA SYNTHETASE, CYTOPLASMIC
&ASPARTATE--TRNALIGASE
&ASPRS
&aspartyl-trna synthetase, cytoplasmic

1.4. RNA replication
DKA1 PROTEIN--Membrane association of nsP1, and its affinity to endosomes and lysosomes, suggest a role of this protein in the biogenesis of the alphavirus-specific RNA replication complex
&NSP1 PROTEIN
&TFS1 PROTEIN

2..Protein Biosynthesis
2.1. Translation
TRANSLATION FACTOR SUI1
&GOS2 PROTEIN
&PROTEIN TRANSLATION FACTOR SUI1

a. Initiation
INITIATION FACTOR 5A
&EIF-5A
&inition factor eIF-5A
&EIF-4D
PSI PROTEIN
&DNAJ-like protein homolog
&psi protein

b. Elongation
ELONGATION FACTOR 1-ALPHA
&EF-1-ALPHA
&elongation factor 1-alpha
elongation factor 2

c. Ribosomal proteins
MITOCHONDRIAL RIBOSOMAL PROTEIN S5
acidic ribosomal protein P0.e
50S RIBOSOMAL PROTEIN
5S rRNA binding ribosomal protein

1). 40S ribosomal protein
40S ribosomal protein S24E
40S RIBOSOMAL PROTEIN S17
40S ribosomal protein S12
40S RIBOSOMAL PROTEIN S15
40S RIBOSOMAL PROTEIN S26E
40s ribosomal protein s2
40S RIBOSOMAL PROTEIN S22
40S RIBOSOMAL PROTEIN S30
40s ribosomal protein s3

40s ribosomal protein s27
40S RIBOSOMAL PROTEIN S9
40S RIBOSOMAL PROTEIN S24E
40S RIBOSOMAL PROTEIN S28

2). 60S ribosomal protein
ribosomal protein L22
ribosomal protein L23
ribosomal protein l12
60S ribosomal protein
60S RIBOSOMAL PROTEIN YL6
60S RIBOSOMAL PROTEIN L32
60S RIBOSOMAL PROTEIN L35
60S RIBOSOMAL PROTEIN L15
60S ACIDIC RIBOSOMAL PROTEIN P1
&ALLERGEN ALT A 12
&ALTA XII
60S RIBOSOMAL PROTEIN L37E A
60S RIBOSOMAL PROTEIN YL16A
60S RIBOSOMAL PROTEIN L17
60s ribosomal protein l36
60S RIBOSOMAL PROTEIN L18
60s ribosomal protein L46
60S RIBOSOMAL PROTEIN L14EB

2.2 . Post-translational modifications and regulation
a. Glycosylation and addition of other sugars
DOLICHYL-PHOSPHATE-MANNOSE--PROTEIN MANNOSYLTRANSFERASE 2

2.3. Folding and Targeting
a. Folding
cyclophilin--Cyclophilins are a family of cyclosporin-A-binding proteins which catalyse rotation about prolyl peptide bonds
&CYCLOPHILIN
&PEPTIDYL-PROLYL CIS-TRANS ISOMERASE PRECURSOR
&CYCLOSPORIN A-BINDING PROTEIN
&PPIASE
&ROTAMASE
&CPH
&SCYLP
protein disulfide-isomerase
PROTEIN DISULFIDE ISOMERASE PRECURSOR
&PDI
FK506-BINDING PROTEIN--protein folding inhibitor
&PEPTIDYL-PROLYL CIS-TRANSISOMERASE
&FKBP
&FK506-binding protein
endo alpha-1,4 polygalactosaminidase

b. Chaperones
heat-shock protein
heat-shock protein30
&30 KD HEAT SHOCK PROTEIN
&HSP30
&heat shock protein 30
Chaperonin hsp78p

heat shock protein 70
&HEAT SHOCK 70 KD PROTEIN
&DNAK Protein
&HEAT SHOCK 70
&HSP70
MOD-E--A mutation in an HSP90 gene
suppressor of vegetative incompatibility MOD-E (mod-E) gene
activator of Hsp70 and Hsp90 chaperones

c. Protein sorting
VACUOLAR PROTEASE A PRECURSOR
Golgi membrane protein

2.4.Turnover-protein degradation-including vacuolar
protease subunit
PROTEASOME COMPONENT SUN4
26S PROTEASE REGULATORY SUBUNIT 8
&SUG1 PROTEIN
&CIM3 PROTEIN
&TAT-BINDING PROTEIN TBY1
ubiquitin precursor
ubiquitin conjugating enzyme UBC1
ubiquitin / ribosomal protein S27a
methionine aminopeptidase
progesterone-binding protein
&progesterone-binding protein homolog--endomembranes protein

III: CELL GROWTH, CELL DIVISION AND CELL PROCESS

A. Cell Growth, Cell Division

1. Cell walls
N,O-DIACETYLMURAMIDASE--THIS EXTRACELLULAR ENZYME HAS BOTH LYSOZYME (ACETYLMURAMIDASE) AND DIACETYLMURAMIDASE ACTIVITIES
&DIACETYLMURAMIDASE
&LYSOZYME CH
RODLET PROTEIN--spore-wall fungal hydrophobin
&HYDROPHOBIN PRECURSOR
&CLOCK-CONTROLLED PROTEIN 2
&BLUE LIGHT INDUCED PROTEIN 7
cell wall protein
EPD1 PROTEIN PRECURSOR

2. Biomembranes
membrane protein

3. Cytoskeleton, organelle biogenesis
TUBULIN ALPHA CHAIN
TUBULIN ALPHA-B CHAIN
actin
&ACTIN
ARP2/3 COMPLEX--actin polymerization
&P34-ARC
PEROXISOMAL MEMBRANE PROTEIN PMP20
OLEATE-INDUCED PEROXISOMAL PROTEIN POX18
&oleate-inducible peroxisomal protein

&LIPID-TRANSFERPROTEIN
&PXP-18
3-KETOACYL-COA THIOLASE PEROXISOMAL PRECURSOR
&BETA-KETOTHIOLASE
&ACETYL-COA ACYLTRANSFERASE
&PEROXISOMAL 3-OXOACYL-COA THIOLASE

4. Cell cycle control
CELL DIVISION CONTROL PROTEIN 11
Trp-Asp repeat protein--WD repeat proteins are components of multiprotein complexes that are involved in a wide spectrum of cellular activities, such as cell cycle progression, signal transduction, apoptosis, and gene regulation
&WD repeat protein

5. Mitosis/cytokinesis
PSI PROTEIN

6. Other
keratin
&keratin 2 epidermis

B. Cell Processes

1. Cell rescue, defense, osmotic adaptation, starvation response, development (asexual, sexual) (includes antibiotics, toxins)see also B.cell signalling, signal transduction and C. transmembrane transport

1.1. Development

a. Asexual
CONIDIATION-SPECIFIC PROTEIN 8
Modin--MOD-A-encoded polypeptid and is involved in differentiation is a key regulator of growth arrest associated with vegetative incompatibility
COPROPORPHYRINOGEN III OXIDASE PRECURSOR
&COPROPORPHYRINOGENASE
&COPROGEN OXIDASE
&COX

b. Morphology, sporulation, conidiation, growth of fungi
EPD2--Epd2p is required for efficient pseudohyphal formation
SPS2 protein
&sporulation-specific protein 2

c Fungi pathogenicity (cause disease)
snodprot1--belong to cerato-platanin, A new phytotoxic protein

1.2. Detoxification
CATALASE A
super oxide dismutase
&SUPEROXIDE DISMUTASE
&CU-ZN
SUPEROXIDE DISMUTASE PRECURSOR
&MN
catalase-peroxidase
&catalase/peroxidase

1.3. Oxidative stress
manganese superoxide dismutase
& manganese superoxide dismutase precursor
flavohemoglobin--function in storage or as sensors for O2, and in defense against oxidative stress and/or nitric oxide (NO)
&FLAVOHEMOGLOBIN
&DIHYDROPTERIDINE REDUCTASE
&FERRISIDEROPHORE REDUCTASE B
&NITRIC OXIDE DIOXYGENASE
&NOD
&HAEMOGLOBIN-LIKE PROTEIN

1.4. Stress-inducible protein
78 KD GLUCOSE REGULATED PROTEIN HOMOLOG PRECURSOR
&GRP78
&78-kD glucose-regulated protein GRP78
& (GRP78)
cyclophilin B--CypB is an endoplasmic reticulum-localized prolyl-isomerase that interacts with elongation initiation factor 2-beta, an important regulator of protein translation and a central component of the endoplasmic reticulum stress response to hypoxia or ATP depletion
&CYPB

1.5. Clock-controlled genes (circadium rhythm--biological clock)—the ccg-7 gene protein and ccg-2 gene protein were placed under the pathways that they are involved in the cells.
clock-controlled gene-6 protein
clock-controlled gene-8 protein
clock-controlled gene-9 protein
GLUCOSE-REPRESSIBLE GENE PROTEIN
&grg 1 protein
&ccg-1 gene protein
&clock-controlled gene 1 protein
&CLOCK-CONTROLLED GENE PROTEIN 1

2. Cell signalling, signal transduction
2.1. Kinases
protein kinase C
SERINE/THREONINE-PROTEIN KINASE
&serine/threonine protein kinase
&SERINE/THREONINE PROTEIN KINASE
&serine threonine-protein kinase

2.2. Calmodulin
calmodulin
&calcium-modulating protein
VU91D calmodulin

2.3. G protein
GTP-BINDING NUCLEAR PROTEIN GSP2/CNR2
DEVELOPMENTAL REGULATOR FLBA
&regulator of G protein signaling

2.4. Membrane receptor
peroxisomal receptor for PTS2-containing proteins Pex7p--(Pex7p functions as a mobile receptor, shuttling PTS2-containing proteins from the cytosol to the peroxisomes)
&PEROXISOMAL TARGETING SIGNAL 2 RECEPTOR

&PTS2 RECEPTOR
&PEROXISOME IMPORT PROTEIN PAS7
PEROXIN-7

2.5. Hormone
AA-NAT enzyme--the rate-limiting enzyme in the conversion of serotonin to melatonin.
melatonin plays an important physiological role in synchronizing the biological clock
&arylalkylamine N-acetyltransferase

3. Transmembrane transport

3.1. Transport
a. Sugar transport
sugar transport
&sugar transporter like protein
monosaccharide transporter
&AmMst-1
GLUCOSE TRANSPORTER RCO-3
GLUCOSE/GALACTOSE TRANSPORTER

b. Cation transport-ATPase, or major facilitator superfamily
ZRT1 PROTEIN--The yeast ZRT1 zinc transporter is regulated by zinc at both transcriptional and post-translational levels
&high-affinity zinc transport protein
&Zrt1p

c. Anion transport
chloride channel protein
&CHLORIDE CHANNEL PROTEIN

d. Protein, amino acid transport
NUCLEAR SEGREGATION PROTEIN BFR1
autophagy protein
IMPORTIN BETA-1 SUBUNIT
&KARYOPHERIN BETA-1 SUBUNIT
&IMPORTIN 95

IV: UNCLASSIFIED, UNIDENTIFIED, NO SIGNIFICANT HOMOLOGY

A. Classes of enzymes (from M. Riley and KEGG; no pathway specified)

1. Oxidoreductases
cytochrome P450 monooxygenase
&cytochrome P450
2-HYDROXYACID DEHYDROGENASE
&2-hydroxyacid dehydrogenase homolog
&ddh
GMC oxidoreductase

2. Transferases
tyrosine aminotransferase

3. Hydrolases
fatty acid omega-hydroxylase
&P450foxy

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

1. Leucine zipper motif
Ern4p--basic-leucine zipper proteinHac1p/Ern4p functions as a trans-acting factor responsible for the unfolded protein response(UPR)

C. Unclassified (significant homolog but function uncertain in *Neurospora crassa*)
uncertain function
&unclassified
fadE2
PAB1324
dehydrogenase
alpha NAC/1.9.2 protein
IgE-binding protein

D. Unidentified (includes significant match with ORFs)
unknown function
&ORF YKL081w
&unclear function
&hypothetical protein
&Hypothetical protein
&hypothetical protein YLR019w
&BcDNA.GM14838
&HYPOTHETICAL 29.4 KD PROTEIN IN STE6-LOS1 INTERSECTION
&hypothetical protein L
&HYPOTHETICAL 56.4 KD PROTEIN IN RPL32-CWH41 INTERGENICREGION PRECURSOR
&HYPOTHETICAL 25.1 KD PROTEIN IN PMI40-PAC2 INTERGENICREGION
&BcDNA.GM14838
&HYPOTHETICAL 49.6 KD PROTEIN IN ELM1-PRI2 INTERGENICREGION PRECURSOR
&hypothetical protein YOR052c
&&hypothetical protein YOR297c
&orf353
&HYPOTHETICAL PROTEIN 32.6 KD PROTEIN IN VPS15-YMC INTERSECTION
& HYPOTHETICAL PROTEIN 64.3 KD PROTEIN IN CDC12-ERP5 INTERSECTION
&HYPOTHETICAL 40.5 KD PROTEIN IN UBP15-GAS1 INTERGENICREGION PRECURSOR
&HYPOTHETICAL 8.3 KD PROTEIN IN ANSP-RHSE INTERGENICREGION
& HYPOTHETICAL PROTEIN 56.8 KD PROTEIN IN SCJ1-GUA1 INTERSECTION
& HYPOTHETICAL PROTEIN 54.3 KD PROTEIN IN C23D3.03C IN CHROMOSOME I
&hypothetical protein YOR052c

E. No significant homolog
NONE

Appendix II. *Neurospora crassa* evening library categories of cellular functions

I. Bioenergetics and Metabolisms

Part one: Metabolisms

A. Metabolism of carbohydrates (for glucose see energy)

1. Chitin metabolism
chitin synthase
CHITIN BIOSYNTHESIS PROTEIN

2. Cellulose degradation
secreted glucosidase

3. Galactose metabolism
UDP-GLUCOSE 4-EPIMERASE
&GALACTOWALDENASE
&ALDOSE1-EPIMERASE
&MUTAROTASE
GALACTOSE-1-PHOSPHATE URIDYLYLTRANSFERASE

4. Mannitol metabolism
MANNOSYLTRANSFERASE
&MANNOSYLTRANSFERASE KTR4
&alpha-1,2-mannosyltransferase
&ktr4p

5. Calvin cycle
RIBULOSE-PHOSPHATE 3-EPIMERASE--ribulose-5 PO4 to xylulose-5 PO4
&PENTOSE-5-PHOSPHATE3-EPIMERASE
&PPE
&RPE

B. Metabolism of amino acids and related molecules

1. Arginine metabolism
1.1 Arginine anabolism-glutamine, CO2 to arginine
ARGININOSUCCINATE LYASE
&ARGINOSUCCINASE
&argininosuccinate lyase
&ASAL

1.2. Arginine catabolism-arginine to proline
ornithine decarboxylase
&ORNITHINE DECARBOXYLASE
&ornithinedecarboxylase
&ODC
&EC 4.1.1.17
ORNITHINE AMINOTRANSFERASE—convert ornithine to arginine in the biosynthesis of the glutamate family amino acids ornithine, arginine, and proline
&ORNITHINE--OXO-ACIDAMINOTRANSFERASE
&ornithine aminotransferase

2. Aspartic acid metabolism--aspartate anabolism-oxaloacetate, glutamate to aspartate

ASPARTATE AMINOTRANSFERASE
&mitochondrial aspartate aminotransferase precursor
Aspartase

3. Glutamine metabolism
GLUTAMINE SYNTHETASE
&GLUTAMATE--AMMONIA LIGASE
&glutamine synthetase

4. Histidine metabolism
histidine-3 protein
&histidinol dehydrogenase

5. Isoleucine metabolism
ilvX
Acetolactate synthase--see also leucine, isoleucine biosynthesis
&acetolactate synthase
&ALS
&(EC 4.1.3.18)
ketol-acid reductoisomerase
3-methylcrotonyl-CoA carboxylase precursor

6. Leucine metabolism
2-isopropylmalate synthase
acetohydroxy acid synthase (AHS)
Acetolactate synthase (ALS)

7. Lysine metabolism
SACCHAROPINE DEHYDROGENASE[NAD+, L-LYSINE FORMING]
&LYSINE--2-OXOGLUTARATE REDUCTASE
&SDH

8. Phenylalanine metabolism
isobutene-forming enzyme and benzoate 4-hydroxylase
&Cytochrome P450rm
ALANINE AMINOTRANSFERASE, MITOCHONDRIALPRECURSOR
&GLUTAMIC--PYRUVIC TRANSAMINASE

9. Proline metabolism
delta-1-pyroline-5-carboxylatedehydrogenase
& delta-1-pyroline-5-carboxylatedehydrogenase precursor
proline oxidase
SPERMIDINE SYNTHASE
&PUTRESCINE AMINOPROPYLTRANSFERASE
&SPDSY
&spermidine synthase
GAMMA-GLUTAMYL PHOSPHATE REDUCTASE
&GAMMA-GLUTAMYL PHOSPHATE REDUCTASE--EC 1.2.1.41--proline biosynthesis
&GLUTAMATE-5- SEMIALDEHYDE DEHYDROGENASE
&GPR
&gamma-glutamyl phosphate reductase

10. Serine metabolism
L-serine dehydratase--generates pyruvate for gluconeogenesis
&L-SERINE DEHYDRATASE
&L-SERINE DEAMINASE

11. Tryptophan metabolism
ANTHRANILATE SYNTHASE COMPONENT I
&anthranilate synthase component i
&anthranilatesynthase component I

12. Tyrosine metabolism
fumarylacetoacetate hydrolase--prevents degradation of the toxic metabolite fumarylacetoacetate (FAA) in the tyrosine catabolic pathway
&FAH

13. Valine metabolism
METHYLMALONATE-SEMIALDEHYDE DEHYDROGENASE PRECURSOR
&ACYLATING
&MMSDH
beta isopropylmalate dehydrogenase
&IPMDH
&3-isopropylmalate dehydrogenase-see aslo leucine biosynthesis
ketol-acid reductoisomerase precursor
&KETOL-ACID REDUCTOISOMERASE PRECURSOR
&ACETOHYDROXY-ACIDREDUCTOISOMERASE
&ALPHA-KETO-BETA-HYDROXYLACIL REDUCTOISOMERASE
ilvX

14. Aromatic amino acid metabolism
phenylacetyl-CoA ligase

15. Glutamate metabolism
glutamate synthase

16. Threonine biosynthesis
L-serine/L-threonine dehydratase

17. Butanoate metabolism
SUCCINATE SEMIALDEHYDE DEHYDROGENASE
&NAD(+)-DEPENDENTSUCCINIC SEMIALDEHYDE DEHYDROGENASE

C. Metabolism of nucleotides and nucleic acids,purines, pyrimidines

1. Purine metabolism
1.1. inosine mono phosphate de novo biosynthesis
PHOSPHORIBOSYLAMINOIMIDAZOLECARBOXAMIDE FORMYLTRANSFERASE
&IMP cyclohydrolase
&Ade17p
PHOSPHORIBOSYLAMINOIMIDAZOLECARBOXAMIDE FORMYLTRANSFERASE 2
&5-aminoimidazole-4-carboxamide ribonucleotide (AICAR) transformylase
&AICAR TRANSFORMYLASE
URICASE
&uricase

1.2. Other purine metabolic enzymes
ADENYLATE KINASE 2
&ATP-AMP TRANSPHOSPHORYLASE
&adenylate kinase

2. Pyrimidine metabolism

2.1. De novo pyrimidine biosynthesis
carbamyl phosphate synthetase--also arginine biosynthesis and urea cycle
&CARBAMOYL-PHOSPHATE SYNTHASE
URIDINE KINASE
&URIDINE MONOPHOSPHOKINASE
&uridine kinase
& EC 2.7.1.48
dihydroorotate dehydrogenase

D. Metabolism of lipids, fatty acids,sterols--see also fatty acid degradation

1. Fatty acid biosynthesis
1.1ACYL-CARRIER PROTEINS
3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE
&3-KETOACYL-ACYL CARRIER PROTEIN REDUCTASE
&3-ketoacyl-acyl carrier protein reductase

1.2. FATTY ACID SYNTHASE
acyl-CoA dehydrogenases
&acyl dehydrogenase

1.3. BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE
BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE DEHYDROGENASE--keto acids to short branch-chain fatty acids
&branched-chain alpha keto-acid dehydrogenase E1 alpha subunit

1.4. Other
stearoyl-CoA desaturase--adds double bonds to fatty acyl coA
&EC 1.14.99.5

2. Sterols
2.1. General
sterol transmethylase
HYDROXYMETHYLGLUTARYL-COA SYNTHASE--condenses Acetyl-CoA w/Acetoacetyl-CoA to form HMG-CoA which is the substrate for HMG-CoA reductase
&HMG-COA SYNTHASE
&3-HYDROXY-3-METHYLGLUTARYL COENZYME A SYNTHASE
&3-hydroxy-3-methylglutaryl-CoA-synthase
DIPHOSPHOMEVALONATE DECARBOXYLASE--catalyses the synthesis of isopentenyl diphosphate, the building block of sterol
&diphosphomevalonate decarboxylase
&MEVALONATEPYROPHOSPHATE DECARBOXYLASE
&MVD

2.2. Cholesterol metabolism
C-4 METHYL STEROL OXIDASE--cholesterol biosynthesis
eburicol 14alpha demethylase
&CYP51
&cytochrome P450 sterol 14-demethylase--housekeeping gene of the cytochrome P450 superfamily which is involved in cholesterol biosynthesis in animals

3. Lipids
3.1. SPHINGOLIPIDS
serine palmitoyltransferase

3.2. Lipopolysaccharide biosynthesis--biomembrane precursors

UDP-glucose:sterol glucosyltransferase
UDP-GLUCOSE PYROPHOSPHORYLASE
&udp-glucose pyrophosphorylase
&UTP--GLUCOSE-1-PHOSPHATEURIDYLYLTRANSFERASE
&utp--glucose-1-phosphateuridylyltransferase
&EC 2.7.7.9
&UDPGP

E. Aromatic compound metabolism
4-AMINOBUTYRATE AMINOTRANSFERASE
&GAMMA-AMINO-N-BUTYRATETRANSAMINASE
&GABA TRANSAMINASE

F. Sulfur metabolism
adenosine-5'phosphosulphate kinase
sconCp--sulphur metabolite repression gene, 4 genes, met down, no S up

G. Phosphate metabolism
INORGANIC PYROPHOSPHATASE
&PYROPHOSPHATEPHOSPHO-HYDROLASE

H. Nitrogen metabolism (see also amino acid metabolism)
NITROGEN METABOLIC REGULATION PROTEIN --NEGATIVE REGULATORY PROTEIN IN THE NITROGEN CONTROL CIRCUIT
&NMR PROTEIN
ALIPHATIC NITRILASE
&aliphatic nitrilase
NAD(+)-specific glutamate dehydrogenase
&NAD-GDH
&EC 1.4.1.2

I. Metabolism of cofactors, prosthetic groups

1. Thiamine
thiamine synthase
&THIAMIN BIOSYNTHESIS
&THIAMIN-PHOSPHATE PYROPHOSPHORYLASE
&THIAMIN BIOSYNTHETIC BIFUNCTIONAL ENZYME
&nmt1 protein
&NMT1 PROTEIN
THIAZOLE BIOSYNTHETIC ENZYME
&thiazole biosynthetic enzyme
&STRESS-INDUCIBLE PROTEINSTI35
&stress-inducible protein STI35
&THIAZOLE BIOSYNTHETIC ENZYME PRECURSOR
NMT1 PROTEIN HOMOLOG

2. Coenzyme A
acetyl-coenzyme A synthetase
&acetyl-CoA synthetase
&acoE
acetyl coenzyme A acetyltransferase
&acetyl-coa acetyltransferase

3. Vitamin

pyridoxal reductase-- catalyzes the reversible oxidation of pyridoxine by NADP to yield pyridoxal and NADPH
&Pyridoxine 4-dehydrogenase
&Pyridoxin dehydrogenase

4. Flavins
6,7-DIMETHYL-8-RIBITYLLUMAZINE SYNTHASE
&LUMAZINE SYNTHASE
&DMRL SYNTHASe
&RIBOFLAVIN SYNTHASE BETA CHAIN
&riboflavin synthase beta subunit

5. Folate-methyl donor
non-functional folate binding protein

6. Heme
cytochrome P450 monooxygenase
&cytochrome P450s--Cytochromes P450 are important heme-containing enzymes that catalyze the oxidation of a vast array of endogenous and exogenous compounds, including drugs and carcinogen
iucB protein--major gene involved in Fe(III) uptake

Part two: Energy

A. Carbohydrate as energy source

1. Glycolysis
1.1. GLUCOKINASE
GLUCOKINASE
&GLUCOSE KINASE
&GLK

1.2. Fructose-bisphosphate aldolase
&fructose-bisphosphate aldolase--also gluconeogenesis, PP cycle, carbon fixation, fructose and mannose metabolism
&FRUCTOSE-BISPHOSPHATE ALDOLASE

1.3. Triose-phosphate isomerase
triose-phosphate isomerase
&TRIOSEPHOSPHATE ISOMERASE
&TIM
&triose phosphate isomerase

1.4. Glyceraldehyde-3-phosphate dehydrogenase
GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE
&glyceraldehyde-3-phosphate dehydrogenase
&GAPDH
&CLOCK-CONTROLLED PROTEIN 7
&clock-controlled protein 7

1.5. Phosphoglycerate kinase
phosphoglycerate kinase
&PHOSPHOGLYCERATE KINASE

1.6. Pyruvate kinase
pyruvate kinase

&PYRUVATE KINASE

1.7. ENOLASE
ENOLASE
&2-PHOSPHOGLYCERATE DEHYDRATASE
&2-PHOSPHO-D-GLYCERATE HYDRO-LYASE

2. Gluconeogenesis
2.1.LACTATE DEHYDROGENASE
D-LACTATE DEHYDROGENASE (CYTOCHROME) PRECURSOR--CATALYZE THE STEREOSPECIFIC OXIDATION OF D-LACTATE TO PYRUVATE
&D-LACTATEFERRICYTOCHROME C OXIDOREDUCTASE

2.2. Pyruvate carboxylase
pyruvate carboxylase

2.3. Phosphoenolpyruvate carboxykinase
phosphoenolpyruvate carboxykinase
&PHOSPHOENOLPYRUVATE CARBOXYKINASE

3. Pentose-phosphate pathway
3.1. Phosphogluconate dehydrogenase
6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING 1

3.2. Transketolase
transketolase I
transketolase2
&TK2

3.3. Transaldolase
Tal1p transaldolase

4. Pyruvate metabolism
DIHYDROLIPOAMIDE DEHYDROGENASE PRECURSOR
pyruvate dehydrogenase E1-beta subunit
pyruvate dehydrogenase precursor

5. Tricarboxylic acid pathway
5.1. Aconitate hydratase
aconitase
&ACONITASE

5.2. Isocitrate dehydrogenase
ISOCITRATE DEHYDROGENASE (NADP), MITOCHONDRIAL PRECURSOR
&NADP+-SPECIFIC ICDH
&IDP
&OXALOSUCCINATE DECARBOXYLASE
&IDH

5.3. Alpha-ketogiutarate dehydrogenase
2-oxoglutarate dehydrogenase e1 component
2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT PRECURSOR
&ALPHA-KETOGLUTARATE DEHYDROGENASE
&alpha-ketoglutarate dehydrogenase
&kgd1p

DIHYDROLIPOAMIDE SUCCINYLTRANSFERASE--part of 2-oxoglutarate DH complex (e2) which catalyzes the overall conversion of 2-oxoglutarate to succinyl-CoA + CO2
&dihydrolipoamide succinyltransferase

5.4. SUCCINYL-COA LIGASE
SUCCINYL-COA LIGASE[GDP-FORMING] ALPHA-CHAIN, MITOCHONDRIAL PRECURSOR
&SUCCINYL-COA SYNTHETASE, ALPHA CHAIN
&SCS-ALPHA

5.5. Fumarase
fumarase

5.6. Malate dehydrogenase
malate dehydrogenase
MALATE DEHYDROGENASE, MITOCHONDRIAL PRECURSOR

6. Related reactions
citrate lyase-citrate to oxaloacetate+acetylcoA

7. Glyoxylate cycle
malate synthase
&MALATE SYNTHASE, GLYOXYSOMAL
&EC4.1.3.2
isocitrate lyase
&ISOCITRASE
&ISOCITRATASE
&ICL

8. Fermentation, alcoholic
8.1. Alcohol dehydrogenase
alcohol dehydrogenase
ALCOHOL DEHYDROGENASE I
&ADH 2
type III alcohol dehydrogenase
ZINC-TYPE ALCOHOL DEHYDROGENASE

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

9.1. Glycogen degradation
glycogen phosphorylase
&EC 2.4.1.1
alpha-amylase
& EC 3.2.1.1
PHOSPHOGLUCOMUTASE 2
&GLUCOSE PHOSPHOMUTASE 2
&PGM 2

9.2. Starch degradation
ALPHA-GLUCOSIDASE PRECURSOR
&alpha-glucosidase
&MALTASE
alpha-glucosidase AgdA

B. Fatty acid as energy source

1. Lipase-triacylglycerols to glycerol+FA
LIPASE 5 PRECURSOR

2. Beta-oxydation of fatty acids
2.1. Carnitine acetyl transferase
carnitine racemase-d to l form

3. Ketone body metabolism
ACETYL-COA HYDROLASE
&ACETYL-COA DEACYLASE
&ACETYL-COAACYLASE
&ACETATE UTILIZATION PROTEIN

C. Metabolism of other energy sources
acetate kinase
&ACETATE KINASE
&ackA
ALDEHYDE DEHYDROGENASE--broad substrate specificity
&ALLERGEN CLA H 3
&CLA HIII
&ALDDH
GLUTATHIONE-DEPENDENT FORMALDEHYDE DEHYDROGENASE
&FDH
&FALDH
GLYCEROL KINASE
&ATP:GLYCEROL 3-PHOSPHOTRANSFERASE
&GLYCEROKINASE
&ATP-STIMULATED GLUCOCORTICOID-RECEPTOR TRANSLOCATION PROMOTER
&GK
PRPD PROTEIN

D. Electron transport

1. Complex I-NADH-ubiquinone
mitochondrialcomplex I
&19.3kD iron-sulfur subunit of mitochondrialcomplex I
&iron-sulfur subunit of mitochondrialcomplex I
NADH-UBIQUINONE DEHYDROGENASE
&NADH-UBIQUINONE DEHYDROGENASE 24 KD SUBUNIT PRECURSOR

2. Complex II-Succinate-ubiquinone
succinate dehydrogenase
&SUCCINATE DEHYDROGENASE (UBIQUINONE) IRON-SULFUR PROTEIN PRECURSOR

3. Complex III-Ubiquinone to cytochrome C
NADH-CYTOCHROME B5 REDUCTASE PRECURSOR
&P34/P32
CYTOCHROME B5
&CYB5
nadh-cytochrome b5 reductase
&CYTOCHROME B5 REDUCTASE
&NADH-cytochrome b5 reductase
&cytochrome-b5 reductase
cytochrome c
&CYTOCHROME C
&core protein II
&cytochrome C reductase complex core protein 2
CYTOCHROME C OXIDASE POLYPEPTIDE VI PRECURSOR

ACYL CARRIER PROTEIN, MITOCHONDRIAL PRECURSOR
&ACP
&NADH-UBIQUINONE OXIDOREDUCTASE
NADH-UBIQUINONE OXIDOREDUCTASE B22 SUBUNIT
&COMPLEXI-B22
&CI-B22
ubiquinol-cytochrome c reductase complex subunit
UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX UBIQUINONE-BINDING PROTEIN QP-C
&UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX UBIQUINONE-BINDING PROTEIN QP-C PRECURSOR
&UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX 11 KD PROTEIN
&COMPLEX III SUBUNIT VII
&ubiquinone-binding protein (QP-C)
BETA-MPP
&MITOCHONDRIAL PROCESSING PEPTIDASE BETA SUBUNIT PRECURSOR
&UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX CORE PROTEIN I
&EC 3.4.99.41
NADH-UBIQUINONE OXIDOREDUCTASE 40 KD SUBUNIT PRECURSOR
&COMPLEX I-40KD
&CI-40KD
NADH-UBIQUINONE OXIDOREDUCTASE 21 KD SUBUNIT
&COMPLEXI-21KD
&CI-21KD
Cytochrome C oxidase subunit
CYTOCHROME C OXIDASE POLYPEPTIDE VIIA
CYTOCHROME C OXIDASE POLYPEPTIDE V PRECURSOR
CYTOCHROME C OXIDASE POLYPEPTIDE VIB
&AED
Mitochondrial Cytochrome Bc1 Complex

4. Electron carriers
flavoprotein
QUINONE OXIDOREDUCTASE
&NADPH:QUINONE REDUCTASE

5. Component enzymes and molecules
MITOCHONDRIAL CARRIER PROTEIN

6. ATP synthsis and degradation
ATP SYNTHASE OLIGOMYCIN SENSITIVITY CONFERRAL PROTEIN
PLASMA MEMBRANE ATPASE (PROTON PUMP)
&H+-transporting ATPase
&V-ATPase
&p-ATPase
&PROTEIN PUMP
oligomycin sensitivity conferring protein--ATP5-subunits d of Saccharomyces cerevisiae mitochondrial ATP synthase
&mitochondrial F0F1-ATP synthase/ATPase
&OSCP--a subunit of mitochondrial F0F1-ATP synthase/ATPase
mitochondrial F0F1-ATP synthase
&Tim11p
ATP SYNTHASE PROTEIN 9, MITOCHONDRIAL PRECURSOR
&LIPID-BINDING PROTEIN
14-3-3 PROTEIN HOMOLOG
&TH1433

&14-3-3 protein--14-3-3 protein is a natural ligand of the plasma membrane H(+)-ATPase, regulating proton pumping by displacing the C-terminal autoinhibitory domain of the H(+)-ATPase.
ATP SYNTHASE ALPHA CHAIN, MITOCHONDRIAL PRECURSOR
ATP SYNTHASE E CHAIN, MITOCHONDRIAL
VACUOLAR ATP SYNTHASE SUBUNIT G
&V-ATPASE 13 KD SUBUNIT
&VACUOLAR H(+)-ATPASE SUBUNIT G
PROTEOLIPID PROTEIN PPA1
&VACUOLAR ATP SYNTHASE 22 KD PROTEOLIPID SUBUNIT
VACUOLAR ATP SYNTHASE 16 KD PROTEOLIPID SUBUNIT
ATP SYNTHASE D CHAIN, MITOCHONDRIAL
VACUOLAR ATP SYNTHASE 98 KD SUBUNIT
&VACUOLAR ATPASE 98KD SUBUNIT
ATP SYNTHASE J CHAIN, MITOCHONDRIAL
VACUOLAR ATP SYNTHASE SUBUNIT AC39
&V-ATPASE AC39 SUBUNIT
&V-ATPASE 41 KD SUBUNIT
ATP SYNTHASE DELTA CHAIN, MITOCHONDRIAL PRECURSOR
ATP SYNTHASE SUBUNIT 4, MITOCHONDRIAL PRECURSOR
ATP SYNTHASE BETA CHAIN, MITOCHONDRIAL PRECURSOR

7. Alternative respiratory path
ALTERNATIVE OXIDASE PRECURSOR
&ALTOX

E. Reducing carriers

1.glutaredoxin
glutaredoxin
&GLUTAREDOXIN

II: Gene expression and genetic information processing

A. DNA synthesis

1. DNA replication
PROLIFERATING CELL NUCLEAR ANTIGEN--auxillary protein of DNA pol sigma, involved in control of eukaryotic DNA replication by increasing processibility
&PCNA
DNA replication licensing factor
minichromosome maintenance protein Mcm7p

2. DNA modification and DNA repair
Hmp1
&mismatched base pair and cruciform DNA recognition protein (HMP1)

3. DNA packaging
3.1.Histone
HISTONE H4
&histone H4
HISTONE H3
&histone H3
HISTONE H2B
&histone H2B
histone H2A

HISTONE H1

3.2. DNA-binding
CURVED DNA-BINDING PROTEIN
&42 KD PROTEIN
CELLULAR NUCLEIC ACID BINDING PROTEIN
SAP1 PROTEIN
&switch-activating-protein

B. Gene Expression

1. Transcription
1.1. RNA Polymerase
DNA-DIRECTED RNA POLYMERASE II 13.3 KD POLYPEPTIDE
&RPB11
&RPB14
auxin-induced protein--an auxin-induced protein that modulates the specific activity of the nucleolar RNA polymerase
&aldo/ketoreductase family

1.2. Regulation
CROSS-PATHWAY CONTROL PROTEIN 1--contains segments similar to the DNA-binding and transcriptional activation domains of GCN4
transcription factor
&cross-pathway control protein 1
&CPC1
CCAAT/enhancer-binding protein
&C/EBP
TRANSCRIPTIONAL REPRESSOR RCO-1
TRANSCRIPTIONAL ACTIVATOR PROTEIN ACU-15
TRANSCRIPTIONAL ADAPTOR
MBF1--a transcriptional coactivator
alpha NAC/1.9.2. protein
Cad1 protein
TRANSCRIPTION ELONGATION FACTOR S-II
&TFIIS

1.3. Processing
a. SPLICEOSOME
spliceosomal protein
splicing factor
U5 snRNP-specific 40 kDa protein
snRNP core Sm protein
small nuclear ribonucleoprotein
Sm protein F
MSS51 PROTEIN--be involved in the splicing of the mitochondrial pre-mRNA of cytochrome oxidase subunit I (COX1)
&MSS51 protein
b. Other
NUCLEOLAR PROTEIN NOP5--a small nucleolar ribonucleoprotein component required for pre-18 S rRNA processing in yeast
RNA-binding protein
RNA binding domain
RNA12 PROTEIN--rna12+, a gene of Saccharomyces cerevisiae involved in pre-rRNA maturation
&RNA12 protein

1.4. tRNA synthesis and modifications
ISOLEUCYL-TRNA SYNTHETASE, CYTOPLASMIC
&ISOLEUCINE--TRNALIGASE
&ILERS
LYSYL-TRNA SYNTHETASE
&LYSINE--TRNA LIGASE
&LYSRS
&KIAA0070
&lysyl tRNA synthetase
&lysyl-trna synthetase
ALANYL-TRNA SYNTHETASE, CYTOPLASMIC
&ALANINE—TRNALIGASE
&ALARS
VALYL-TRNA SYNTHETASE, MITOCHONDRIAL PRECURSOR
&VALINE--TRNA LIGASE
&VALRS
THREONYL-TRNA SYNTHETASE, MITOCHONDRIALPRECURSOR
&THREONINE--TRNA LIGASE
&THRRS
PSEUDOURIDYLATE SYNTHASE 1
&pseudouridylate synthase
&PSEUDOURIDINE SYNTHASE I

1.5. RNA replication
DKA1 PROTEIN--Membrane association of nsP1, and its affinity to endosomes and lysosomes, suggest a role of this protein in the biogenesis of the alphavirus-specific RNA replication complex
&NSP1 PROTEIN
&TFS1 PROTEIN

2.Protein biosynthesis
2.1. Translation
TRANSLATION FACTOR SUI1
&GOS2 PROTEIN
&PROTEIN TRANSLATION FACTOR SUI1
TRANSLOCATION PROTEIN SEC66
&HSS1 PROTEIN

a. Initiation
EUKARYOTIC INITIATION FACTOR 4A
&EIF-4A
INITIATION FACTOR
INITIATION FACTOR 5A
&EIF-5A
&EIF-4D
translation initiation factor 4e
CPC3 protein--homologue of yeast GCN2 (eIF-2 alpha protein kinase GCN2=eukaryotic translation initiation factor 2)

b. Elongation
elongation factor 1 beta
ELONGATION FACTOR 1-ALPHA
&EF-1-ALPHA
EF-TU
elongation factor 2
&ELONGATION FACTOR 2
&EF-2

ELONGATION FACTOR 3
&EF-3
ELONGATION FACTOR 1-GAMMA 2
&EF-1-GAMMA 2

c. Termination
translation release factor subunit 1

d. Ribosomal proteins
5S rRNA binding ribosomal protein
MITOCHONDRIAL RIBOSOMAL PROTEIN S5
acidic ribosomal protein P0.e, cytosolic
ribosomal protein S14.e
ribosomal protein L21
ribosomal protein CRP7
ribosomal protein L31.e.B, cytosolic
ribosomal protein S7
ribosomal protein L27
ribosomal protein L13E

1). 40S ribosomal protein
40S ribosomal protein
40S ribosomal protein S12
40S RIBOSOMAL PROTEIN S13
40S RIBOSOMAL PROTEIN S15
40S RIBOSOMAL PROTEIN S26
40s ribosomal protein s2
40S RIBOSOMAL PROTEIN S30
40s ribosomal protein s27
40S RIBOSOMAL PROTEIN S9
40S RIBOSOMAL PROTEIN S28
40S RIBOSOMAL PROTEIN SA HOMOLOG
&RIBOSOME-ASSOCIATEDPROTEIN 1
40S ribosomal protein S5
40S RIBOSOMAL PROTEIN S11
40S RIBOSOMAL PROTEIN RP10
40S RIBOSOMAL PROTEIN S6
40S RIBOSOMAL PROTEIN S16
40S RIBOSOMAL PROTEIN S18E
40S RIBOSOMAL PROTEIN S13
40S RIBOSOMAL PROTEIN S19
40S RIBOSOMAL PROTEIN YS29A
40S RIBOSOMAL PROTEIN S31

2). 60S ribosomal protein
60S ribosomal protein
ribosomal protein L23
ribosomal protein
ribosomal protein l12
60S RIBOSOMAL PROTEIN YEL050C
60S RIBOSOMAL PROTEIN YL6
60S RIBOSOMAL PROTEIN L32
&60s ribosomal protein L32
60S RIBOSOMAL PROTEIN L35
60S ACIDIC RIBOSOMAL PROTEIN P1
60S ACIDIC RIBOSOMAL PROTEIN P2

&MINOR ALLERGEN ALT A 6
&ALT A VI
60S RIBOSOMAL PROTEIN L37E A
60S RIBOSOMAL PROTEIN L37B
60S RIBOSOMAL PROTEIN L17
60S RIBOSOMAL PROTEIN L18
60S RIBOSOMAL PROTEIN L18A
60s ribosomal protein L46
60S RIBOSOMAL PROTEIN L14EB
60S RIBOSOMAL PROTEIN YL35
60S RIBOSOMAL PROTEIN YL16B
60S RIBOSOMAL PROTEIN YL43
60S RIBOSOMAL PROTEIN YL39
60S RIBOSOMAL PROTEIN L26
60S RIBOSOMAL PROTEIN YL17-A
60S RIBOSOMAL PROTEIN L9 B
60S RIBOSOMAL PROTEIN L38
MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L2
60S RIBOSOMAL PROTEIN L11
60S ribosomal protein L24
60s ribosomal protein l27
60S RIBOSOMAL PROTEIN L22
MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L33
60s ribosomal protein l2

2. 2. Post-translational modifications and regulation
a. Methylation
serine hydroxymethyltransferase
&SERINE HYDROXYMETHYLTRANSFERASE, CYTOSOLIC
&SERINEMETHYLASE
&GLYCINE HYDROXYMETHYLTRANSFERASE
&SHMT

b. Glycosylation and addition of other sugars
glycosyl transferases
GPI-ANCHOR TRANSMIDASE
DOLICHYL-PHOSPHATE-MANNOSE--PROTEIN MANNOSYLTRANSFERASE 2
dolichol-phosphate-mannose synthase

c. Other
26S PROTEASE REGULATORY SUBUNIT 7
&CIM5 PROTEIN
&TAT-BINDING HOMOLOG 3
26S PROTEASE REGULATORY SUBUNIT 4
&MTS2 PROTEIN
NEUTRAL PROTEASE II
MITOCHONDRIAL RESPIRATORY CHAIN COMPLEXES ASSEMBLYPROTEIN RCA1
&(TAT-BINDING HOMOLOG 12)

2.3. Folding and targeting
a. Folding
PEPTIDYL-PROLYL CIS-TRANS ISOMERASE
&&Peptidyl Prolyl cis-trans isomerase (catalyzes folding)
PEPTIDYL-PROLYL CIS-TRANS ISOMERASE B PRECURSOR
&CYCLOPHILIN B
&S-CYCLOPHILIN

&ROTAMASE
&PPIASE
&SCYLP
cyclophilin--Cyclophilins are a family of cyclosporin-A-binding proteins which catalyse rotation about prolyl peptide bonds
&CYCLOPHILIN
&PEPTIDYL-PROLYL CIS-TRANS ISOMERASE PRECURSOR
&CYCLOSPORIN A-BINDING PROTEIN
&PPIASE
&ROTAMASE
&CPH
PEPTIDYL-PROLYL CIS-TRANS ISOMERASE
CALNEXIN HOMOLOG PRECURSOR
CALNEXIN HOMOLOG--folding of glycoproteins
FK506-BINDING PROTEIN--protein folding inhibitor
&&PEPTIDYL-PROLYL CIS-TRANSISOMERASE
&FKBP
&PPIASE
FK506-BINDING PROTEIN PRECURSOR
&PEPTIDYL-PROLYL CIS-TRANSISOMERASE
&PPIASE
&FKBP-21
DISULFIDE ISOMERASE ERP38 PRECURSOR

b. Chaperones
chaperone
&Chaperonins
prefoldin--chaperone which delivers unfolded proteins to another chaperonin
heat-shock protein30
&30 KD HEAT SHOCK PROTEIN
&HSP30
&heat shock protein 30
Chaperonin hsp78p
heat shock protein 70
&DNAK Protein
&HEAT SHOCK 70
MOD-E--a mutation in an HSP90 gene
suppressor of vegetative incompatibility MOD-E (mod-E) gene
&HEAT SHOCK PROTEIN 90 HOMOLOG
HEAT SHOCK PROTEIN HSP1
&65 KD IGE-BINDING PROTEIN
T-COMPLEX PROTEIN 1, BETA SUBUNIT-chaperone of actin, tubulin
&TCP-1-BETA
&CT-BETA
chaperonin TCP1 epsilon
heat shock protein
zuotin --putative Z-DNA binding protein
&ZUOTIN
&Zuo1p

c. Protein sorting and targeting
vacuolar protein sorting
SERINE CARBOXYPEPTIDASE PRECURSOR
CARBOXYPEPTIDASE Y-sorting of vacuolar protein
GAMMA-ADAPTIN--associated with clathrin-coated vesicles or with any of the components of the AP-1 comple

&gamma-adaptin
&GOLGI ADAPTOR HA1/AP1 ADAPTIN GAMMA SUBUNIT
&CLATHRIN ASSEMBLY PROTEIN COMPLEX 1 GAMMA LARGE CHAIN
&GAMMA-ADA
ER lumen protein retaining receptor protein
edoplasmic reticulum associated protein
COATOMER ZETA SUBUNIT--trafficing to golgi, nonclathrin vesicles
snare protein--the SNARE complex, synaptotagmin III, nSec1, domains of NSF and its adaptor SNAP, along with Rab3 and some of its effector related to vesicle fusion protein mechnisms.--exocytosis
Ca+2-binding EF hand protein--EF-hand protein which undergoes a Ca(2+)-induced translocation from cytoplasm to membranes
VACUOLAR PROTEIN SORTING-ASSOCIATED PROTEIN VPS28
VACUOLAR PROTEIN SORTING/TARGETING PROTEIN PEP1 PRECURSOR
&TARGETING PROTEIN PEP1 PRECURSOR
&VACUOLAR CARBOXYPEPTIDASE SORTING RECEPTOR VPS10
&CARBOXYPEPTIDASE Y RECEPTOR
&CPY RECEPTOR
CLATHRIN COAT ASSEMBLY PROTEIN AP19
&CLATHRIN COAT ASSOCIATED PROTEIN AP19
&GOLGI ADAPTOR -1 19KD
&HA119KD SUBUNIT
&CLATHRIN ASSEMBLY PROTEIN COMPLEX CHAIN
CARBOXYPEPTIDASE Y PRECURSOR
&CARBOXYPEPTIDASE YSCY
SSO1 PROTEIN
&syntaxin
&T-SNARE
Rer1 protein--ts potential role in the endoplasmic reticulum localization of membrane proteins
MITOCHONDRIAL PROCESSING PEPTIDASE ALPHA SUBUNIT PRECURSOR
&ALPHA-MPP
prohibitin--Prohibitins act as a membrane-bound chaperone for the stabilization of mitochondrial proteins, Prohibitins are ubiquitous, abundant and evolutionarily strongly conserved proteins that play a role in important cellular processes, suggests a functional homology with protein chaperones with respect to their ability to hold and prevent misfolding of newly synthesized proteins.

2. 4. Turnover-protein degradation-including vacuolar
proteosome--THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE, PEPTIDES WITH ARG, PHE,
TYR, LEU, AND GLU ADJACENT TO THE LEAVING GROUP AT NEUTRAL OR SLIGHTLY BASIC PH. THE PROTEASOME HAS AN ATP-DEPENDENT
PROTEOLYTI ACTIVITY.[PATHWAY] IS INVOLVED IN AN ATP/UBIQUITIN-DEPENDENT NON-LYSOSOMAL PROTEOLYTIC PATHWAY
& PROTEOSOME
POTENTIAL PROTEASOME COMPONENT C5
&MULTICATALYTICENDOPEPTIDASE COMPLEX SUBUNIT C5
26S proteasome regulatory subunit mts3
PROTEASOME COMPONENT C9/Y13
&MACROPAIN SUBUNIT
&MULTICATALYTIC ENDOPEPTIDASE COMPLEX SUBUNIT
PROTEASOME COMPONENT PRE3 PRECURSOR
&MACROPAIN SUBUNITPRE3
26S proteasome subunit 9
ubiquitin precursor
ubiquitin conjugating enzyme UBC1
ubiquitin conjugating enzyme
&UBIQUITIN-CONJUGATING ENZYME

&ubiquitin-conjugating-enzyme-like protein
UBIQUITIN-CONJUGATING ENZYME E2-24 KD
&UBIQUITIN-PROTEINLIGASE
&UBIQUITIN CARRIER PROTEIN
&ubiquitin-protein ligase
UBIQUITIN-CONJUGATING ENZYME E2-17 KD
&UBIQUITIN-PROTEINLIGASE 2
&UBIQUITIN CARRIER PROTEIN
ubiquitin-conjugating enzyme protein E2
ubiquitin/S27a fusion protein
&ubiquitin/ribosomal protein S27a fusion protein
ubiquitin fusion protein
CAAX PRENYL PROTEASE--cleavage of alpha factor for activatio
Lon serine protease
&MITOCHONDRIAL ATP-DEPENDENT PROTEASE
ATP-dependent Clp protease proteolytic subunit
Lon protease-like protein
PROTEASOME COMPONENT Y7
&MACROPAIN SUBUNIT Y7
&PROTEINASE YSCE SUBUNIT 7
&MULTICATALYTIC ENDOPEPTIDASE COMPLEX SUBUNIT Y7
UBIQUITIN-LIKE PROTEIN SMT3
REXB PROTEIN--rexB function: it prevents degradation of the short-lived protein lambda O known to be involved in lambda DNA replication
&rexB protein
ROTEASOME COMPONENT C7-ALPHA
&MACROPAIN SUBUNITC7-ALPHA
&PROTEINASE YSCE SUBUNIT 7
&COMPONENT Y8
&SCL1 SUPPRESSOR PROTEIN
&MULTICATALYTIC ENDOPEPTIDASE COMPLEX C7

III: Cell growth, cell division and cell process

A. Cell walls, biomembranes and cytoskeleton

1. Cell walls
septin B
N,O-DIACETYLMURAMIDASE--THIS EXTRACELLULAR ENZYME HAS BOTH LYSOZYME (ACETYLMURAMIDASE) AND DIACETYLMURAMIDASE ACTIVITIES
&DIACETYLMURAMIDASE
&LYSOZYME CH
RODLET PROTEIN--spore-wall fungal hydrophobin
&HYDROPHOBIN PRECURSOR
&CLOCK-CONTROLLEDGENE PROTEIN 2
&BLUE LIGHT INDUCED PROTEIN 7
cell wall alpha-glucan synthase
ENDOLYSIN
&LYSIS PROTEIN
&LYSOZYME
cell wall biogenesis protein
glycine rich protein
proline-rich protein 15--cell envolupe
SEPTIN HOMOLOG SPN4

2. Biomembranes

OUTER MEMBRANE USHER PROTEIN
membrane protein
erythrocyte membrane antigen 1

3. Cytoskeleton, organelle biogenesis
kinesin related protein 1
TUBULIN ALPHA CHAIN
&alpha-tubulin chain
ankyrin
actin
&ACTIN
PROFILIN-assembly of actin monomers
COFILIN--actin binding protein ADF family
&cofilin
myosin-II
&myp2+
PEROXISOMAL MEMBRANE PROTEIN PMP20
&ALLERGEN ASPF 3
OLEATE-INDUCED PEROXISOMAL PROTEIN POX18
&oleate-inducible peroxisomal protein
&LIPID-TRANSFERPROTEIN
&PXP-18
phosphatidylethanolamine methyltransferase
PEROXISOMAL MEMBRANE PROTEIN PMP27
&PEROXIN-11
PEROXISOMAL-COENZYME A SYNTHETASE

4. Cell cycle control
cell division cycle CDC48 homolog
CELL DIVISION CONTROL PROTEIN 11
SCH9 protein--cell progress through G1
BARRIERPEPSIN PRECURSOR--a scaffold protein capable of bridging two major apoptosis pathways
&EXTRACELLULAR "BARRIER" PROTEIN
&BAR PROTEINASE
&BAR PROTEINASE

5. Mitosis/cytokinesis
5.1. MITOSIS
CENTROMERE/MICROTUBULE BINDING PROTEIN CBF5
&NUCLEOLAR PROTEIN CBF5
&CENTROMERE-BINDING FACTOR 5
&CENTROMERE/MICROTUBULE BINDING PROTEIN
SMC PROTEIN—functions in hromosome ondensation, segregation, and global gene regulation
&extragenic suppressor of the bimD6 mutation

5.2. Cytokinesis
TROPOMYOSIN--component of contractile ring

6. Other
PH RESPONSIVE PROTEIN 1 PRECURSOR
&PH-REGULATED PROTEIN1
silk fibroin heavy chain

B. Cell processes

1. Cell rescue, defense, osmotic adaptation, starvation response, development (asexual, sexual) (includes antibiotics, toxins) see also cell signalling, signal transduction and transmembrane transport

1.1. Development
a. Asexual
CONIDIATION-SPECIFIC PROTEIN 6
CONIDIATION-SPECIFIC PROTEIN 8
CONIDIATION-SPECIFIC PROTEIN 10
COPROPORPHYRINOGEN III OXIDASE PRECURSOR
&COPROPORPHYRINOGENASE
&COPROGEN OXIDASE
&COX
UROPORPHYRINOGEN DECARBOXYLASE
&UPD

b. Sexual cycle
krev-1
GUANINE NUCLEOTIDE-BINDING PROTEIN BETA SUBUNIT-LIKEPROTEIN--The cpc-2 gene of *Neurospora crassa* encodes a protein entirely composed of WD-repeat segments that is involved in general amino acid control and female fertility
&CROSS-PATHWAY CONTROL WD-REPEAT PROTEIN CPC-2
&cpc-2 gene
&WD-repeat protein

c. Morphology, sporulation, growth of fungi
SPS2 protein
&sporulation-specific protein 2

d. Fungi pathogenicity (cause disease)
snodprot1--belong to cerato-platanin, A new phytotoxic protein

1.2. Defense
a. Defense protein
L-AMINO ACID OXIDASE PRECURSOR
&LAO
cytosolic NADPH oxidase p67-phox--a tightly regulated multicomponent enzyme complex, the NADPH oxidase, which produces superoxide, a reactive oxygen molecule that is an essential component of host defense against infection
NADPH oxidase

b. Sterigmatocystin biosynthesis
sterigmatocystin
&norsolorinic acid reductase
&(U34740)
&versicolorin B synthase
STERIGMATOCYSTIN BIOSYNTHESIS PROTEIN

1.3. Detoxification
singlet oxygen resistance protein
CATALASE A
superoxide dismutase
&super oxide dismutase
&SUPEROXIDE DISMUTASE
&CU-ZN

cytochrome P450 monooxygenase--Microsomal cytochrome P450s participate in xenobiotic detoxification, procarcinogen activation, and steroid hormone synthesis
CYTOCHROME P450 55A2
&CYTOHROME P450NOR1

1.4. Dessication tolerance
rehydrin--stress protein

1.5. Oxidative stress
flavohemoglobin-function in storage or as sensors for O2, and in defense against oxidative stress and/or NO toxicity
&FLAVOHEMOGLOBIN
&DIHYDROPTERIDINE REDUCTASE
&FERRISIDEROPHORE REDUCTASE B
&NITRIC OXIDE DIOXYGENASE
&NOD
&NO oxygenase
&HAEMOGLOBIN-LIKE PROTEIN

1.6. Night/day rhythm (circadium rhythm--biological clock)—the ccg-2 gene product and the ccg-7 gene product were placed under the pathways they are involved.
ccg-4 putative polypeptide 2
ccg-4 putative polypeptide 1
clock-controlled gene-6 protein
clock-controlled gene-8 protein
clock-controlled gene-9 protein
GLUCOSE-REPRESSIBLE GENE PROTEIN
&grg1 protein
&CLOCK-CONTROLLED GENE 1 PROTEIN
&clock-controlled gene 1 protein

1.7. Tumor protein and tumor suppressor
TRANSLATIONALLY CONTROLLED TUMOR PROTEIN
& TCTP
tumor metastasis inhibitor nm23-H2--The nm23 gene is a potential metastasis-suppressor gene originally identified in a murine melanoma line, Reduced expression of nm23-H1, but not of nm23-H2, is concordant with the frequency of lymph-node metastasis of human breast cancer

1.8. Multidrug resistance
CYANIDE HYDRATASE
&FORMAMIDE HYDROLYASE

1.9. Other
heavy metal tolerance protein precursor
carboxyphosphonoenolpyruvate mutase--catalyses the formation of one of the two C-P bonds in bialaphos, a potent herbicide isolated from Streptomyces hygroscopicus.

2. Cell signalling, signal transduction and second messengers

2.1. PHOSPHATASES
PROTEIN-TYROSINE PHOSPHATASE

2.2. Kinases
protein kinase skp1p
protein kinase CK2 beta subunit
protein kinase kin1

&PROTEIN KINASE KIN1
protein kinase C
mitogen-activated protein kinase--MAP kinases p42mapk and p44mapk participate in a protein kinase cascade(s) important for signaling in many ##cell types and contexts. Both MAP kinases are activated in vitro by MAP kinase kinase, a protein-tyrosine and threonine kinase
&MAPK
&MAP kinase
mitogen-activated protein kinase CPK1
&mitogen-activated protein kinase kinase CPK1--two protein kinases, designated CPK1 (25 kDa) and CPK2 (38 kDa), are present in spinach thylakoid membranes

2.3. cAMP
amiB--plays a role at the start of Dictyostelium differentiation through induction of the ACA expression which is essential for cAMP signalling
&aggregation minus B

2.4. Calmodulin
calmodulin
&CALMODULIN
&calcium-modulating protein
CALCIUM-BINDING PROTEIN
&calcium-binding protein

2.5. G protein
ADP-RIBOSYLATION FACTOR
GTP-BINDING NUCLEAR PROTEIN GSP2/CNR2
rho-gdp dissociation inhibitor--prevents cycling of GDP with GTP of rho protein family
GTPase
RAN-SPECIFIC GTPASE-ACTIVATING PROTEIN
&RAN BINDINGPROTEIN 1 HOMOLOG
&RANBP1
&PERINUCLEAR ARRAY-LOCALISED PROTEIN
YPT1-RELATED PROTEIN 2--an essential ras-related gene in the fission yeast Schizosaccharomyces pombe
YPT1-RELATED PROTEIN 5
GTP-BINDING PROTEIN YPT51/VPS21
&GTP-binding protein VPS21

2.6. Membrane receptor
trk-1--trk-mediated intracellular signal transduction pathway

3. Transmembrane transport
3.1. secretion
SEC14 CYTOSOLIC FACTOR
&PHOSPHATIDYLINOSITOL/PHOSPHATIDYLCHOLINE TRANSFER PROTEIN
SECRETORY PATHWAY GDP DISSOCIATION INHIBITOR--regulates GDP/GTP exchange rxn of Sec4 by inhibiting dissociation of GDP from it, plays essential role in yeast secretory pathway

3.2. Transport
a. Sugar transport
sugar transport protein
AmMst-1
&monosaccharide transporter
GLUCOSE TRANSPORTER
GLUCOSE TRANSPORTER RCO-3
hexose transporter

GLUCOSE/GALACTOSE TRANSPORTER
&glucose/galactose transporter

b. Cation transport-ATPase, or major facilitator superfamily
E1-E2 ATPases
CALCIUM-TRANSPORTING ATPASE
&Ca2+-transporting ATPase
CCC1 PROTEIN--putative transmembrane Ca2+ transporter
COPPER TRANSPORT PROTEIN CTR3
&COPPER TRANSPORTER 3
VITAMIN D3 HYDROXYLASE-ASSOCIATED PROTEIN
&VDHAP
potassium channel subunit
oxaloacetate decarboxylase--a membrane-bound, Na+-activated, biotin-containing enzyme that functions as a Na+ pump
&OXALOACETATE DECARBOXYLASE ALPHA CHAIN
&oxaloacetate decarboxylase alpha chain

c. Anion transport
tartrate transport
&TARTRATE TRANSPORTER
sulfate permease II--encoded by the cys-14 gene,CYS-14 protein appears to be localized in the plasma membrane, suggesting that it functions ##as a sulfate ion transporter
&SULFATE PERMEASE II

d. Protein, amino acid transport
PROTEIN TRANSPORT PROTEIN
&protein transport protein
&PROTEIN TRANSPORT PROTEIN SEC 61 GAMMA SUBUNIT
ERV25 PROTEIN PRECURSOR--CONSTITUENT OF OF COPII-COATED ENDOPLASMIC RETICULUM-DERIVED TRANSPORT VESICLES
PROLINE-SPECIFIC PERMEASE
&PROLINE TRANSPORT PROTEIN
SEC61 protein
MITOCHONDRIAL PHOSPHATE CARRIER PROTEIN
&PHOSPHATETRANSPORT PROTEIN
&MITOCHONDRIAL PHOSPHATE CARRIER PROTEIN PRECURSOR (PTP)
AUTOPHAGOCYTOSIS PROTEIN AUT1--Autophagocytosis is a starvation-induced process responsible for transport of cytoplasmic proteins to the vacuole
Opt1p
&An oligopeptide transporter
AMINO-ACID PERMEASE
&amino acid permease

e. Mitochondrial transport
2-oxoglutarate/malate translocator
&MITOCHONDRIAL 2-OXOGLUTARATE/MALATE CARRIER
oxoglutarate malate translocator
ADP, ATP CARRIER PROTEIN
&ADP/ATP TRANSLOCASE
&ADENINENUCLEOTIDE TRANSLOCATOR
&ANT
OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN
frataxin--human frataxin co-localizes with a mitochondrial protein, mouse and yeast frataxin homologues contain a potential mitochondrial targeting sequence in their N-terminal domains and that disruption of the

yeast gene results in mitochondrial dysfunction; mitochondrial protein-required for mitochondrial iron efflux

f. Other
transport protein
ATP-DEPENDENT BILE ACID PERMEASE

IV: Unclassified, unidentified, no significant homology

A. Classes of enzymes (from M. Riley and KEGG; no specific pathway)

1. Oxidoreductases
OXIDOREDUCTASE
SQUALENE MONOOXYGENASE
&SQUALENE EPOXIDASE

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

1. Zinc finger motif-DNA binding
zinc-finger protein

C. Unclassified (significant homolog but function uncertain in *Neurospora crassa*)
uncertain function
&unclassified
Ran/spi1 binding protein
het-c2 protein
YSA1 PROTEIN--similarity to proteins w/core mutt domain
23S rRNA intron 2 protein
Pmt3p
SONA
rAsp f 7
Vip1 protein
AX110P
PDH1P--a novel trans-membrane protein
Ser/Arg-related nuclear matrix protein
secretory protein
PROTEIN SSP120 PRECURSOR
phosphoribosylaminoimidazolesuccinocarboxamide synthase
endosomal P24A protein
subunit of the final step of the secretory pathway
YnaD
RDS1 PROTEIN--an adenine-repressible gene, rds1 protein, regulated by glucose, ammonium, phosphate, carbon dioxide and temperature.
rds1 protein
2-hydroxyhepta-2,4-diene-1,7-dioate isomerase
&hpcE-2
NIPSNAP1 protein
acetyltransferase
Flavin-binding monooxygenase
symbiosis-related protein
IgE-binding protein
ferric leghemoblobin reductase-2 precursor
methylumbelliferyl-acetate deacetylase
&EC 3.1.1.56
BLI-3 PROTEIN

D. Unidentified (includes significant match with ORFs)
unknown function
&HYPOTHETICAL PROTEIN
&HYPOTHETICAL
&unclear function
&hypothetical protein
&hypothetic 70K protein
&unknown
&putative protein
&PUTATIVE PROTEIN
&HYPOTHETICAL 54.5 KD PROTEIN IN CBF2-SKN1 INTERGENICREGION
& HYPOTHETICAL 89.4 KD TRP-ASP REPEATS CONTAINING PROTEININ PMT6-PCT1 INTERGENIC REGION
&ORF YOL057w
&HYPOTHETICAL 161.2 KD PROTEIN IN NMD5-HOM6 INTERGENICREGION
&HYPOTHETICAL 13.9 KD PROTEIN IN FCY2-PET117 INTERGENICREGION
&HYPOTHETICAL 29.1 KD PROTEIN IN URA7-POL12 INTERGENICREGION
&cDNA EST yk302b12.3
&HYPOTHETICAL 77.8 KD PROTEIN IN MRPS28-HXT7 INTERGENICREGION
&HYPOTHETICAL 38.1 KD PROTEIN IN RCK1-AMS1 INTERGENICREGION
&Saccharomyces serevisiae hypothetical52.9KD protein in CDC26-YMR31 int
&HYPOTHETICAL 41.7 KD PROTEIN IN SFP1-CTR3 INTERGENICREGION
&HYPOTHETICAL 22.7 KD PROTEIN IN PAS1-MST1 INTERGENICREGION
&HYPOTHETICAL 54.2 KD PROTEIN IN ERP5-ORC6 INTERGENICREGION
&HYPOTHETICAL 24.3 KD PROTEIN IN PEM2-HOC1 INTERGENICREGION
& HYPOTHETICAL 42.4 KD PROTEIN IN CDC12-ORC6 INTERGENICREGIO
&HYPOTHETICAL 11.9 KD PROTEIN IN TGT-SECD INTERGENICREGION
& HYPOTHETICAL 124.0 KD PROTEIN IN PCS60-ABD1 INTERGENICREGION
&C34D4.12 gene product
&J1590 gene product
&YGL010w-like proein
&similar to C.elegans F38E1.9 gene product
&alternate gene name: yeeM, yfxB; similar to hypothetical protein
&HYPOTHETICAL 41.5 KD PROTEIN C1F5.03C IN CHROMOSOME I
&HYPOTHETICAL
&HYPOTHETICAL 50.8 KD PROTEIN IN MIR1-STE18 INTERGENICREGION
&hypothetical conserved protein
&100 kDa protein
&hypothetic protein 2 (cpc-1 5' region)
&HYPOTHETICAL 187.1 KD PROTEIN IN OGG1-CNA2 INTERGENICREGION
&HYPOTHETICAL 26.6 KD PROTEIN T19C3.4 IN CHROMOSOME III
&HYPOTHETICAL 20.8 KD PROTEIN IN CNA2-CYB2 INTERGENICREGION
& hypothetical 70K protein
&putative protein
&YKL117w
&hypothetic protein L
&HYPOTHETICAL 27.7 KD PROTEIN IN PRP19-HSP104 INTERGENICREGION
&HYPOTHETICAL 97.1 KD PROTEIN C32A11.02C IN CHROMOSOME I
&HYPOTHETICAL 34.2 KD PROTEIN IN CUS1-RPL18A1 INTERGENICREGION
&HYPOTHETICAL 42.4 KD PROTEIN IN CDC12-ORC6 INTERGENICREGION
&HYPOTHETICAL 89.4 KD TRP-ASP REPEATS CONTAINING PROTEIN IN PMT6-PCT1 INTERGENIC REGION
&36.7 KD PROTEIN IN BR-NOT3 INTERGENIC REGION
&F38E1.9 gene product
&Saccharomyces cerevisiae SCD6 protein

E. No significant homolog

NONE
Contig--582
Singlets--428

Appendix III. *Neurospora crassa* morning library categories of cellular functions

I: Bioenergetics and metabolisms (73)

Part one. Metabolisms (25)

A. Metabolism of carbohydrates(for glucose see energy) (5)

1. Chitin metabolism (1)

<CHITIN SYNTHASE 3>

Contig317	284	6.5e-23	368	622	sp|P29070|CHS3_NE CHITIN SYNTHASE 3 (CHITIN-UDP ACETYL-GLUCOSAMINYLTRANSFERASE 3) >pir||A41638 chitin synthase (EC 2.4

2. Galactose metabolism (1)

<alpha-1,4 polygalactosaminidase>

Contig465	217	4.5e-17	235	594	emb|CAB51262.1| (AL096872) putative endo alpha-1,4 polygalactosaminidase[*Streptomyces* coelicolor A3(2)]

3. Mannitol metabolism (2)

<mannosyl-oligosaccharide 1,2-alpha-mannosidase>fs

Contig157	399	2.3e-36	34	402	pir||S63701 mannosyl-oligosaccharide 1,2-alpha-mannosidase (EC 3.2.1.113)precursor - Aspergillus phoenicis

<MANNOSYL-OLIGOSACCHARIDE ALPHA-1,2-MANNOSIDASE PRECURSOR>

Contig155	200	1.8e-14	291	509	sp|P31723|MA12_PE MANNOSYL-OLIGOSACCHARIDE ALPHA-1,2-MANNOSIDASE PRECURSOR(MAN(9)-ALPHA-MANNOSIDASE) >pir||S58766 mann

4. Sorbitol metabolism (1)

<SORBITOL DEHYDROGENASE>

Contig145	428	2e-39	13	471	emb|CAA94841| (Z70782) similar to sorbitol dehydrogenase; cDNA EST EMBL:T00701comes from this gene [Caenorhabditis

B. Metabolism of amino acids and related molecules (11)

1. Glutamine metabolism (2)

<GLUTAMINE SYNTHETASE>

Contig132	559	2.4e-53	214	618	sp|Q12613|GLNA_CO GLUTAMINE SYNTHETASE (GLUTAMATE—AMMONIA LIGASE)>gi|1322275 (L78067) glutamine synthetase [Glomerell

Contig71	242	9.8e-20	332	493	sp|Q12613|GLNA_CO GLUTAMINE SYNTHETASE (GLUTAMATE–AMMONIA LIGASE)>gi|1322275 (L78067) glutamine synthetase [Glomerell

2. Isoleucine metabolism (3)

<methylcrotonyl-CoA carboxylase>					
c1d08nm.f1	421	4.5e-38	33	482	gb|AAD25800.1|AC0 (AC006550) Identical to gb|U125363-methylcrotonyl-CoA carboxylase precursor protein from Arabidopsis
<acyl-CoA dehydrogenase>					
Contig401	347	1.5e-39	92	553	gb|AAB52261.2| (U97002) similar to acyl-CoA dehydrogenases and epoxidehydrolases; Pfam domain PF00441 (Acyl-CoA_dh)
Contig441	297	1.2e-25	6	536	gi|2649568 (AE001032) acyl-CoA dehydrogenase (acd-7) [Archaeoglobus fulgidus]

3. Lysine metabolism (1)

<SACCHAROPINE DEHYDROGENASE>					
c1f04nm.r1	99	0.03	399	494	sp|P38999|LYS9_YE SACCHAROPINE DEHYDROGENASE [NADP+, L-GLUTAMATE FORMING]>pir||S41937 saccharopine dehydrogenase (NADP

4. Serine metabolism (1)

<L-SERINE DEHYDRATASE>					
Contig168	160	1.8e-10	222	623	sp|P17324|SDHL_YE L-SERINE DEHYDRATASE (L-SERINE DEAMINASE)>pir||S12731L-serine dehydratase (EC 4.2.1.13) SDL1 - yeas

5. Tyrosine metabolism(1)

<TYROSINE DECARBOXYLASE 4>					
Contig389	135	2e-07	1	231	sp|Q06088|TYD4_PE TYROSINE DECARBOXYLASE 4 >gi|169677 (M95685) tyrosinedecarboxylase [Petroselinum crispum]

6. Beta-alanine metabolism (1)

<METHYLMALONATE-SEMIALDEHYDE DEHYDROGENASE>					
c1d01nm.r1	441	7.3e-41	71	487	sp|Q02252|MMSA_HU METHYLMALONATE-SEMIALDEHYDE DEHYDROGENASE (ACYLATING) (MMSDH) >gi|188696 (M93405) methylmalonate semi

7. Branch-amino acid metabolism (2)

<KETOL-ACID REDUCTOISOMERASE PRECURSOR>					
c1e10nm.f1	688	4.8e-67	15	443	sp|P38674|ILV5_NE KETOL-ACID REDUCTOISOMERASE PRECURSOR (ACETOHYDROXY-ACIDREDUCTOISOMERASE) (ALPHA-KETO-BETA-HYDROXYLA
c1e10nm.r2	210	8.5e-16	362	493	sp|P38674|ILV5_NE KETOL-ACID REDUCTOISOMERASE PRECURSOR (ACETOHYDROXY-ACIDREDUCTOISOMERASE) (ALPHA-KETO-BETA-HYDROXYLA

C. Nitrogen metabolism (see also amino acid metabolism) (5)

<nitrite reductase>					
Contig256	442	7.4e-62	349	594	sp|P38681|NIR_NEU NITRITE REDUCTASE (NAD(P)H) >pir||A49848 nitrite reductase-

					Neurospora crassa<NITROGEN METABOLIC REGULATION PROTEIN>
Contig99	371	2.3e-33	2	214	sp|P23762|NMR_NEU NITROGEN METABOLIC REGULATION PROTEIN (NMR PROTEIN)>pir||S11910 nitrogen metabolic regulation protei
<URICASE>					
Contig331	571	1.3e-54	267	755	sp|Q00511|URIC_AS URICASE (URATE OXIDASE) >pir||A38097 urate oxidase (EC1.7.3.3) - Aspergillus flavus >emb|CAA43895| (
Contig225	359	4.1e-32	276	536	sp|P33282|URIC_EM URICASE (URATE OXIDASE) >pir||A48879 urate oxidase (EC1.7.3.3) - Emericella nidulans >emb|CAA51009|
<cyanate lyase-cyanate, bicarbonate substrates>					
<NITRILASE 3>					
Contig370	159	2.6e-10	242	469	sp|P46010|NRL3_AR NITRILASE 3 >gi|508735 (U09959) nitrilase [Arabidopsisthaliana]

D. Metabolism of cofactors, prosthetic groups (4)

1. Thiamine (3)

<THIAZOLE BIOSYNTHETIC ENZYME>					
Contig440	1072	1e-107	226	1104	sp|P23618|THI4_FU THIAZOLE BIOSYNTHETIC ENZYME (STRESS-INDUCIBLE PROTEINSTI35) >pir||B37767 stress-inducible protein s
Contig316	591	1.1e-56	150	545	sp|P23618|THI4_FU THIAZOLE BIOSYNTHETIC ENZYME (STRESS-INDUCIBLE PROTEINSTI35) >pir||B37767 stress-inducible protein s
Contig184	213	3e-34	493	669	sp|P40998|THI2_SC THIAZOLE BIOSYNTHETIC ENZYME >pir||S45597 nmt2 protein fission yeast (Schizosaccharomyces pombe) >e 2.folate-methyl donor (1) <folate>
c1d05nm.r1	206	6.7e-16	137	412	gi|2565196 (AF000381) non-functional folate binding protein [Homosapiens]

Part two. Energy (48)

A. Carbohydrate as energy source (37)

1. Glycolysis (17)

1.1. Hexokinase (1)

<hexokinase>					
Contig228	285	9e-24	210	473	emb|CAA08922|(AJ009973) hexokinase [Aspergillus niger]

1.2. Glucokinase (1)

<GLUCOKINASE>					
Contig419	168	1.1e-09	9	287	sp|Q92407|HXKG_AS GLUCOKINASE (GLUCOSE KINASE) (GLK)>emb|CAA67949|(X99626) glucokinase [Aspergillus niger]

1.3. Fructose-bisphosphate aldolase (3)

<fructose-bisphosphate aldolase>					
Contig282	687	6.2e-67	130	654	sp|P53444|ALF_NEU FRUCTOSE-BISPHOSPHATE ALDOLASE >gi|1334980 (L42380)fructose 1,6

					bisphosphate-aldolase [Neurospora cr
g7b02nm.f1	339	5.3e-30	2	232	sp|P53444|ALF_NEU FRUCTOSE-BISPHOSPHATE ALDOLASE >gi|1334980 (L42380)fructose 1,6 bisphosphate-aldolase [Neurospora cr
h7d11nm.r1	105	0.00022	331	444	sp|P53444|ALF_NEU FRUCTOSE-BISPHOSPHATE ALDOLASE >gi|1334980 (L42380)fructose 1,6 bisphosphate-aldolase [Neurospora cr

1.4. Glyceraldehyde 3-phosphatedehydrogenase (5)

< glyceraldehyde	3-phosphate dehydrogenase >				
Contig523	1733	8e-178	213	1226	gi|1532189 (U67457) glyceraldehyde 3-phosphate dehydrogenase [Neurosporacrassa]
Contig254	768	1.6e-75	216	674	sp|P54118|G3P_NEU GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (GAPDH) (CLOCK-CONTROLLED PROTEIN 7) >gi|1326237 (U56397)gly
Contig448	654	1.8e-63	197	589	sp|P54118|G3P_NEU GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (GAPDH) (CLOCK-CONTROLLED PROTEIN 7) >gi|1326237 (U56397)gly
Contig310	302	4.2e-26	190	366	sp|P54118|G3P_NEU GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (GAPDH) (CLOCK-CONTROLLED PROTEIN 7) >gi|1326237 (U56397)gly
i7e03nm.f1	128	5.3e-07	5	85	gi|1532189 (U67457) glyceraldehyde 3-phosphate dehydrogenase [Neurosporacrassa]

1.5. Phosphoglycerate kinase (2)

<phosphoglycerate kinase>					
Contig265	932	6.1e-93	34	612	sp|P38667|PGK_NEU PHOSPHOGLYCERATE KINASE >emb|CAA39865| (X56512)phosphoglycerate kinase [Neurospora crassa]
Contig388	683	1.6e-66	218	613	sp|P38667|PGK_NEU PHOSPHOGLYCERATE KINASE >emb|CAA39865| (X56512)phosphoglycerate kinase [Neurospora crassa]

1.6. Phosphoglycerate mutase (1)

<phosphoglycerate mutase>					
Contig281	338	1.5e-29	1	351	gi|2773203 (AF039713) Similar to phosphoglycerate mutase; coded for by C.elegans cDNA yk357d11.5; coded for by

1.7. Phosphopyruvate hydratase (1)

<phosphopyruvate hydratase>					
i8h02nm.r1	201	1.2e-14	304	486	gi|3885968 (AF100985) phosphopyruvate hydratase [Penaeus monodon]

1.8. Pyruvate kinase (1)

<pyruvate kinase>					
Contig313	686	7.9e-67	210	707	sp|P31865|KPYK_TR PYRUVATE KINASE >pir||JN0780 pyruvate kinase (EC2.7.1.40) - fungus (Trichoderma reesei) >gi|170553 (

1.9. ENOLASE (2)

<ENOLASE>					
Contig50	584	4.9e-56	6	443	sp|P42040|ENO_CLA ENOLASE (2-PHOSPHOGLYCERATE DEHYDRATASE) (2-PHOSPHO-D-GLYCERATE HYDRO-LYASE) (ALLERGEN CLA H 6) (CLA

Contig226 582 8.3e-56 3 428 sp|Q12560|ENO_ASP ENOLASE (2-PHOSPHOGLYCERATE DEHYDRATASE)(2-PHOSPHO-D-GLYCERATE HYDRO-LYASE) >pir||JC45426beta-hydrox

2. Gluconeogenesis (1)

2.1. Phosphoenolpyruvate carboxykinase (1)

<phosphoenolpyruvate carboxykinase>
c1d10nm.f1 640 5.7e-62 11 505 sp|O13434|PPCK_CA PHOSPHOENOLPYRUVATE CARBOXYKINASE (ATP) >gi|2267237(U70473) PEP carboxykinase [Candida albicans]

3. Pyruvate metabolism (2)

<PYRUVATE DEHYDROGENASE E1 COMPONENT, ALPHA SUBUNITPRECURSOR>
Contig359 350 3.2e-31 114 488 sp|P16387|ODPA_YE PYRUVATE DEHYDROGENASE E1 COMPONENT, ALPHA SUBUNITPRECURSOR (PDHE1-A) >emb|CAA50657| (X71664) PDA1 [
Contig397 235 1.3e-18 223 495 sp|Q10489|ODPA_SC PYRUVATE DEHYDROGENASE E1 COMPONENT, ALPHA SUBUNITPRECURSOR (PDHE1-A) >emb|CAA97360.1| (Z73100) pyru

4. Tricarboxylic acid pathway (5)

4.1. Isocitrate dehydrogenase (3)

< isocitrate dehydrogenase >
c6h06nm.f3 346 9e-31 174 488 sp|P28241|IDH2_YE ISOCITRATE DEHYDROGENASE [NAD], MITOCHONDRIAL SUBUNIT 2 PRECURSOR (ISOCITRIC DEHYDROGENASE) (NAD+-SPE

<NAD-dependent isocitrate dehydrogenase subunit 2>
Contig165 655 1.7e-63 3 629 gi|3820488 (AF045154) NAD-dependent isocitrate dehydrogenase subunit 2[Kluyveromyces lactis]

<NAD(+)-isocitrate dehydrogenase subunit I>
Contig36 214 2.7e-16 384 536 gb|AAB63461.1| (AF009036) NAD(+)-isocitrate dehydrogenase subunit I[Ajellomyces capsulatus]

4.2. *SUCCINYL-COA LIGASE* (2)

<SUCCINYL-COA LIGASE>
Contig150 362 1.9e-32 120 476 sp|O13750|SUCA_SC PROBABLE SUCCINYL-COA LIGASE (GDP-FORMING), ALPHA-CHAINPRECURSOR (SUCCINYL-COA SYNTHETASE, ALPHA CHA
<ATP-specific succinyl-CoA synthetase beta subunit>
Contig19 431 8.3e-40 117 497 gi|3766201 (AF058955) ATP-specific succinyl-CoA synthetase beta subunit [Musmusculus]

5. Fermentation, alcoholic (4)

5.1.Alcohol dehydrogenase (4)
<alcohol dehydrogenase>
Contig435 363 1.5e-32 116 628 emb|CAA21911.1| (AL033389) alcohol dehydrogenase [Schizosaccharomyces pombe]

Contig121	359	4e-32	130	597	emb|CAA21911.1| (AL033389) alcohol dehydrogenase [Schizosaccharomyces pombe]
Contig324	324	1.4e-28	130	621	emb|CAA21911.1| (AL033389) alcohol dehydrogenase [Schizosaccharomyces pombe]
<ALCOHOL DEHYDROGENASE I>					
c1c12nm.f1	223	8e-31	111	254	sp|P41747|ADH1_AS ALCOHOL DEHYDROGENASE I >gi|439867 (L27434) alcoholdehydrogenase [Aspergillus flavus]

6. Metabolism of energy reserves (glycogen, starch, trehalose) (7)

6.1. Glycogen degradation (5)

<glycogen phosphorylase>					
Contig328	628	1.3e-60	2	616	emb|CAA28273| (X04604) glycogen phosphorylase (AA 1-891) [Saccharomycescerevisiae] >prf||1212353A phosphorylase,gl
Contig104	451	6.2e-41	196	654	pir||S61144 glycogen phosphorylase (EC 2.4.1.1) - yeast (Saccharomycescerevisiae) >gi|849168 (U28371) Glycogen p
Contig238	256	5.6e-20	276	506	pir||S61144 glycogen phosphorylase (EC 2.4.1.1) - yeast (Saccharomycescerevisiae) >gi|849168 (U28371) Glycogen p
<GLYCOSYLTRANSFERASE HOC1 PRECURSOR>					
Contig424	376	6.5e-34	2	478	sp|P47124|HOC1_YE PUTATIVE GLYCOSYLTRANSFERASE HOC1 PRECURSOR >pir||S57094probable membrane protein YJR075w - yeast (S
<PHOSPHOGLUCOMUTASE 1>					
Contig80	603	4e-88	443	985	sp|P33401|PGM1_YE PHOSPHOGLUCOMUTASE 1 (GLUCOSE PHOSPHOMUTASE 1) (PGM 1)>pir||S41199 phosphoglucomutase (EC 5.4.2.2) P

6.2 Starch degradation (1)

<glucan 1,4-alpha-glucosidase>					
Contig253	757	2.1e-74	285	722	pir||S36364 glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - Neurosporacrassa

6.3. Trehalose degradation (1)

<NEUTRAL TREHALASE>					
Contig212	272	7.4e-22	2	175	sp|O42622|TREB_MA NEUTRAL TREHALASE (ALPHA,ALPHA-TREHALASE)(ALPHA,ALPHA-TREHALOSE GLUCOHYDROLASE) >gi|2688970 (AF02798

7. Related reactions (1)

<ATP citrate lyase>					
Contig114	1267	2.3e-128	3	797	emb|CAA12224.1| (AJ224922) ATP citrate lyase [Sordaria macrospora]

B. Metabolism of other energy sources (1)

<Nitrilase>					
Contig347	352	2.2e-31	20	409	emb|CAA84681.1| (Z35604) similar to Nitrilase [Caenorhabditis elegans]

C. Electron transport (8)

1. Complex III-Ubiquinone to cytochrome C (3)

<ubiquinol-cytochrome c>					
Contig491	321	3e-28	47	412	emb|CAA20859.1| (AL031546) ubiquinol-cytochrome c reductase complex

					subunit[Schizosaccharomyces pombe]
<CYTOCHROME C OXIDASE POLYPEPTIDE II>					
b2a10nm.f1	847	7.1e-84	66	656	sp|P00411|COX2_NE CYTOCHROME C OXIDASE POLYPEPTIDE II >pir||OBNC2cytochrome-c oxidase (EC 1.9.3.1) chain II - Neurospo
<NADH-UBIQUINONE OXIDOREDUCTASE 14.8 KD SUBUNIT>					
Contig44	653	3e-63	19	390	sp|P42114|NB4M_NE NADH-UBIQUINONE OXIDOREDUCTASE 14.8 KD SUBUNIT (COMPLEXI-14.8KD) (CI-14.8KD) >pir||S43840 NADH dehyd

2. ATP synthase (5)

<ATPase subunit 6>					
i4h12nm.f1	542	1.5e-51	57	464	gi|805071 (L14642) ATPase subunit 6 [Neurospora crassa]
<PLASMA MEMBRANE ATPASE>					
Contig81	211	2.2e-34	507	668	gi|2197050 (AF001033) putative 20kDa subunit of the V-ATPase [Neurosporacrassa]
Contig457	331	3.4e-29	285	518	gi|2197050 (AF001033) putative 20kDa subunit of the V-ATPase [Neurosporacrassa]
<ATP SYNTHASE PROTEIN 9, MITOCHONDRIAL PRECURSOR>					
Contig82	613	4.2e-59	80	520	sp|P00842|ATP9_NE ATP SYNTHASE PROTEIN 9, MITOCHONDRIAL PRECURSOR(LIPID-BINDING PROTEIN) >pir||LWNCA H+-transporting A
<ATP SYNTHASE DELTA CHAIN, MITOCHONDRIAL PRECURSOR>					
Contig87	671	2.8e-65	115	531	sp|P56525|ATPD_NE ATP SYNTHASE DELTA CHAIN, MITOCHONDRIAL PRECURSOR

D. Reducing carriers (2)

1.Gluathione (1)

<glutathione S-transferase 3>					
Contig241	180	3e-13	110	457	gi|4758714 ref|NP_004519.1|pMGST3| microsomal glutathione S-transferase 3>gi|2583081 (AF026977) microsomal glut

2.Thioredoxin (1)

<thioredoxin>					
Contig468	256	3.2e-21	113	406	sp|P29429|THIO_EM THIOREDOXIN >pir||S27053 thioredoxin - Emericellanidulans >bbs|120057 thioredoxin [Aspergillus nidul

II: Gene expression and genetic information processing (91)

A. DNA synthesis (7)

1. DNA replication (1)

<minichromosome maintenance protein Mcm7p>

Contig284	476	5.6e-44	273 677	gi|3236468 (AF070481) minichromosome maintenance protein Mcm7p[*Schizosaccharomyces* pombe] >emb|CAA20099| (AL031

2. DNA modification and DNA repair (2)

<EXCINUCLEASE ABC SUBUNIT A>				
h2a05nm.f1	820	6.2e-81	2 472	sp|P07671|UVRA_EC EXCINUCLEASE ABC SUBUNIT A >pir||BVECUA uvrA protein-Escherichia coli >gi|148165 (M13495) UvrA protein
<Hmp1>				
Contig398	134	2.6e-08	299 571	gi|1176420 (U39049) Hmp1 [Ustilago maydis]

3. DNA packaging (4)

3.1. Histone (4)

<histone>				
Contig100	317	1.1e-27	202 465	emb|CAA07351| (AJ006959) histone H2A [Botryotinia fuckeliana]
<HISTONE H4>				
Contig251	409	1.9e-37	76 321	sp|P04914|H4_NEUC HISTONE H4 >pir||S07913 histone H4 - Neurospora crassa>emb|CAA25760| (X01611) histone H4 [Neurospora
<HISTONE H3>				
Contig483	666	1.1e-64	41 448	sp|P07041|H3_NEUC HISTONE H3 >pir||S07350 histone H3 - Neurospora crassa>emb|CAA25761| (X01612) histone H3 [Neurospora
<HISTONE H2B>				
a4f06nm.r1	282	5.7e-24	303 482	sp|P23754|H2B_EME HISTONE H2B >pir||S11937 histone H2B - Emericella nidulans>emb|CAA39153| (X55547) H2B [Emericella ni

B. Gene expression (84)

1. Transcription (14)

1.1 Regulation (6)

<CROSS-PATHWAY CONTROL PROTEIN 1>				
Contig299	284	3.8e-24	265 528	sp|P11115|CPC1_NE CROSS-PATHWAY CONTROL PROTEIN 1 >gi|168793 (J03262)cross-pathway control protein 1 [Neurospora crass
<transcription factor>				
Contig92	391	1.5e-35	473 901	emb|CAB11717| (Z98980) transcription factor [Schizosaccharomyces pombe]
<homeodomain DNA-binding transcription factor>				
Contig244	444	1.1e-40	30 551	gi|3411264 (AF080600) homeodomain DNA-binding transcription factor [Emericellanidulans]
<AP-1-LIKE TRANSCRIPTION FACTOR>				
Contig449	143	1.7e-06	78 452	sp|P56095|AP1_KLU AP-1-LIKE TRANSCRIPTION FACTOR >gi|2245654 (AF006499)transcription factor KlYAP1 [Kluyveromyces lact
<CCAAT/enhancer-binding protein>				
Contig167	120	5.7e-06	240 455	gi|1947129 (AF000262) similar to CCAAT/enhancer-binding protein[Caenorhabditis

					elegans]
<ATP-DEPENDENT RNA HELICASE P47>					
Contig142	479	7.2e-45	2	388	sp|Q07478|HE47_YE PROBABLE ATP-DEPENDENT RNA HELICASE P47 HOMOLOG>pir||S67620 hypothetical protein YDL084w - yeast (Sa

1.2. RNA Processing (3)

a. SPLICEOSOME (3)

< pre-mrna splicing factor atp-dependent rnahelicase >					
Contig470	803	3.7e-79	1	633	emb|CAB52799.1| (AL109846) putative pre-mrna splicing factor atp-dependent rnahelicase [Schizosaccharomyces pombe]
Contig471	442	1.2e-39	124	528	sp|Q92620|Y224_HU PUTATIVE PRE-MRNA SPLICING FACTOR ATP-DEPENDENT RNAHELICASE KIAA0224 (HA4657) >dbj|BAA13213| (D86977
<Lsm5 protein>					
a7f08nm.r2	300	6.9e-26	144	374	emb|CAB45868.1| (AJ238097) Lsm5 protein [Homo sapiens]

1.3. *tRNA synthesis and modifications* (3)

<phenylalanyl-trna synthetase>					
Contig278	548	2e-60	4	528	emb|CAA16986.1| (AL021813) probable phenylalanyl-trna synthetase[Schizosaccharomyces pombe]
<ASPARTYL-TRNA SYNTHETASE>					
Contig137	299	4.7e-25	161	505	emb|CAA20876| (AL031579) ASPARTYL-TRNA SYNTHETASE [Schizosaccharomyces pombe]
<ASPARTYL-TRNA SYNTHETASE, CYTOPLASMIC>					
b4b04nm.f1	158	7.7e-10	16	162	sp|P04802|SYDC_YE ASPARTYL-TRNA SYNTHETASE, CYTOPLASMIC (ASPARTATE—TRNALIGASE) (ASPRS) >pir||SYBYDC aspartate—tRNA l

1.4. RNA replication (2)

<DKA1 PROTEIN>					
Contig305	121	3.8e-05	282	593	sp|P14306|DKA1_YE DKA1 PROTEIN (NSP1 PROTEIN) (TFS1 PROTEIN) >pir||S18843DKA1 protein - yeast (Saccharomyces cerevisia
Contig240	109	0.0003	398	496	sp|P14306|DKA1_YE DKA1 PROTEIN (NSP1 PROTEIN) (TFS1 PROTEIN) >pir||S18843DKA1 protein - yeast (Saccharomyces cerevisia

2.Protein biosynthesis (70)

2.1. Translation (38)

<TRANSLATION FACTOR SUI1>					
Contig445	365	9.3e-33	117	467	sp|P32911|SUI1_YE PROTEIN TRANSLATION FACTOR SUI1 >pir||S31245 translationinitiation factor eIF-2A - yeast (Saccharomy

a. Initiation (2)

<INITIATION FACTOR 5A>					
Contig60	366	7.3e-33	248	547	sp|P38672|IF5A_NE INITIATION FACTOR 5A (EIF-5A) (EIF-4D) >pir||S55278translation initiation factor eIF-5A - Neurospora
<PSI PROTEIN>					
Contig500	314	2.3e-27	4	357	sp|Q09912|PSI_SCH PSI PROTEIN >pir||S55900 DNAJ-like protein homolog fission yeast (Schizosaccharomyces pombe) >gi|95

b. Elongation (3)

<ELONGATION FACTOR 1-ALPHA>					
Contig98	999	5.4e-100	52	636	sp|Q01372|EF1A_NE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) >dbj|BAA08274|(D45837) elongation factor 1-alpha [Neurospora
Contig200	929	1.4e-92	265	810	sp|Q09069|EF1A_SO ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) >emb|CAA65435|(X96615) EF1-alpha translation elongation facto
<elongation factor 2>					
Contig192	107	0.00039	239	331	emb|CAB52147.1| (AL109734) elongation factor 2 [Schizosaccharomyces pombe]

c.Ribosomal proteins (32)

<MITOCHONDRIAL RIBOSOMAL PROTEIN S5>					
Contig349	338	6.6e-30	1	399	sp|P23351|RMS5_NE MITOCHONDRIAL RIBOSOMAL PROTEIN S5 >pir||A19079 23S rRNAintron protein - Neurospora crassa mitochondrial
<5S rRNA ribosomal protein>					
Contig415	926	4e-145	46	597	gi|3003044 (AF054907) putative 5S rRNA binding ribosomal protein [Neurosporacrassa]
<acidic ribosomal protein P0.e>					
c1d04nm.f1	271	8.4e-23	60	317	pir||R5BY0E acidic ribosomal protein P0.e, cytosolic - yeast (Saccharomycescerevisiae) >gi|171806 (M37326) ribos
<50S RIBOSOMAL PROTEIN>					
Contig452	282	5.8e-24	117	500	emb|CAA21892.1| (AL033388) putative 50s ribosomal protein l14[Schizosaccharomyces pombe] pombe]>emb|CAA21221| (AL031824) 60s ribosomal

1). 40S ribosomal protein (12)

<40S RIBOSOMAL PROTEIN S17 (CRP3)>					
Contig126	236	3.3e-19	1	141	sp|P27770|RS17_NE 40S RIBOSOMAL PROTEIN S17 (CRP3) >pir||S34441 ribosomalprotein L17.e - Neurospora crassa >gi|168796
<40S ribosomal protein S12>					
Contig258	527	6.2e-50	129	500	gi|3114615 (AF052483) 40S ribosomal protein S12 [Erysiphe graminis f. sp.hordei]
<40S RIBOSOMAL PROTEIN S15 (S12)>					
Contig361	757	2.2e-74	206	661	sp|P34737|RS15_PO 40S RIBOSOMAL PROTEIN S15 (S12) >pir||A53793 ribosomalprotein S12, cytosolic - Podospora anserina >e
<40S RIBOSOMAL PROTEIN S26E>					
Contig394	479	7.9e-45	122	430	sp|P21772|RS26_NE 40S RIBOSOMAL PROTEIN S26E (CRP5) (13.6 KD RIBOSOMALPROTEIN) >pir||R4NC26 ribosomal protein S26.e -
<40s ribosomal protein s2>					
Contig31	307	1.1e-26	202	435	emb|CAA21187| (AL031798) 40s ribosomal protein s2 [Schizosaccharomyces pombe]
<40S RIBOSOMAL PROTEIN S22>					
Contig431	588	2.1e-56	678	1067	sp|P04648|RS22_YE 40S RIBOSOMAL PROTEIN S22 (YS24) (YP58)>pir||R4BY24ribosomal protein S15a.e.c10 - yeast (Saccharomy
<40S RIBOSOMAL PROTEIN S30>					
Contig486	153	2.8e-10	280	459	sp|Q12087|RS30_YE 40S RIBOSOMAL PROTEIN S30 >pir||S67074 ribosomal proteinS30.e,

					cytosolic - yeast (Saccharomyces cere
<40s ribosomal protein s3>					
Contig55	615	2.1e-59	51	488	emb|CAA19033| (AL023554) 40s ribosomal protein s3. [Schizosaccharomyces pombe]
<40s ribosomal protein s27>					
Contig66	367	5.7e-33	80	325	emb|CAA20058| (AL031154) 40s ribosomal protein s27 type [Schizosaccharomycespombe]
<40S RIBOSOMAL PROTEIN S9>					
Contig7	515	1.1e-48	3	317	sp|P52810|RS9_POD 40S RIBOSOMAL PROTEIN S9 (S7) >emb|CAA65433| (X96613)cytoplasmic ribosomal protein S7 [Podospora ans
<40S RIBOSOMAL PROTEIN S24E (RP50)>					
Contig8	418	2.1e-38	253	603	sp|P26782|RS24_YE 40S RIBOSOMAL PROTEIN S24E (RP50) >pir||S48410 ribosomalprotein S24.e - yeast (Saccharomyces cerevis
<40S RIBOSOMAL PROTEIN S28>					
Contig95	608	1.6e-58	314	748	sp|P32827|RS28_YE 40S RIBOSOMAL PROTEIN S28 >pir||A46703 ribosomal proteinS23.e - yeast (Saccharomyces cerevisiae) >gi
2). 60S ribosomal protein (16)					
<ribosomal protein L22>					
Contig12	292	4.4e-25	174	536	dbj|BAA13074| (D86349) ribosomal protein L22 homolog [Schizosaccharomycespombe]
<ribosomal protein l12>					
Contig9	234	6e-19	6	203	emb|CAA20752.1| (AL031535) ribosomal protein l12. [Schizosaccharomyces
<ribosomal protein L23>					
Contig307	440	9.6e-41	121	498	gi|306549 (L13799) homology to rat ribosomal protein L23 [Homo sapiens]
<60S ribosomal protein>					
Contig338	400	1.8e-36	45	356	sp|P40525|YIF2_YE PROBABLE 60S RIBOSOMAL PROTEIN YIL052C >pir||S48427ribosomal protein L34.e.B, cytosolic - yeast (Sac
Contig11	233	8.8e-19	7	159	emb|CAA20364| (AL031307) 60s ribosomal protein L46 [Schizosaccharomyces pombe]
<60S RIBOSOMAL PROTEIN YL6 (L5)>					
Contig169	1080	1.4e-108	7	768	sp|P05736|RL6_YEA 60S RIBOSOMAL PROTEIN YL6 (L5) (RP8) >pir||S50243ribosomal protein L8.e - yeast (Saccharomyces cerev
<60S RIBOSOMAL PROTEIN L32>					
Contig17	406	4e-37	190	555	sp|P79015|RL32_SC 60S RIBOSOMAL PROTEIN L32 >dbj|BAA19212| (AB000914)ribosomal protein L32 homolog [Schizosaccharomyce
<60S RIBOSOMAL PROTEIN L35>					
Contig195	315	1.8e-27	75	434	sp|P42766|RL35_HU 60S RIBOSOMAL PROTEIN L35 >gi|562074 (U12465) ribosomalprotein L35 [Homo sapiens]
<60S RIBOSOMAL PROTEIN L15>					
Contig334	839	3.8e-83	2	517	sp|O13418|RL15_AS 60S RIBOSOMAL PROTEIN L15 >emb|CAA75582| (Y15321)putative ribosomal protein L15 [Aspergillus niger]
<60S ACIDIC RIBOSOMAL PROTEIN P1>					
Contig280	305	1.9e-26	212	538	sp|P49148|RLA1_AL 60S ACIDIC RIBOSOMAL PROTEIN P1 (ALLERGEN ALT A 12) (ALTA XII) >emb|CAA58998| (X84216) ribosomal pro
<60S RIBOSOMAL PROTEIN L37E A>					

Contig393	202	3.4e-24	556	744	sp|P49166|R7EA_YE 60S RIBOSOMAL PROTEIN L37E A (YP55) >pir||S51430ribosomal protein L37.e.A, cytosolic - yeast (Saccha
<60S RIBOSOMAL PROTEIN YL16A>					
Contig479	157	9.6e-11	62	214	sp|Q02326|R16A_YE 60S RIBOSOMAL PROTEIN YL16A >pir||S28944 ribosomalprotein L6.e.A, cytosolic - yeast (Saccharomyces c
<60S RIBOSOMAL PROTEIN L17>					
Contig504	580	1.5e-55	51	449	sp|P04451|RL1A_YE 60S RIBOSOMAL PROTEIN L17 >pir||R5BY17 ribosomal proteinL23.e, cytosolic - yeast (Saccharomyces cere
<60s ribosomal protein l36>					
Contig53	256	3.3e-21	81	365	emb|CAB38606.1| (AL035655) 60s ribosomal protein l36 [Schizosaccharomycespombe]
<60S RIBOSOMAL PROTEIN L18>					
c1g07nm.r1	234	6.6e-19	253	495	sp|P42791|RL18_AR 60S RIBOSOMAL PROTEIN L18 >gi|606970 (U15741) cytoplasmicribosomal protein L18 [Arabidopsis thaliana
<60S RIBOSOMAL PROTEIN L14EB>					
Contig252	208	3.6e-16	6	206	sp|P38754|R14B_YE PROBABLE 60S RIBOSOMAL PROTEIN L14EB >pir||S46797ribosomal protein L14.e.B, cytosolic - yeast (Sacch

2.2. Post-translational modifications and regulation (2)

a. Glycosylation and addition of other sugars (2)

<DOLICHYL-PHOSPHATE-MANNOSE-PROTEIN MANNOSYLTRANSFERASE 2>					
Contig211	567	3.5e-54	1	783	sp|P31382|PMT2_YE DOLICHYL-PHOSPHATE-MANNOSE-PROTEIN MANNOSYLTRANSFERASE 2>pir||S36711 hypothetical protein YAL023 -
Contig206	284	3.6e-23	220	594	sp|P31382|PMT2_YE DOLICHYL-PHOSPHATE-MANNOSE-PROTEIN MANNOSYLTRANSFERASE 2>pir||S36711 hypothetical protein YAL023 -

2.3. Folding and targeting (19)

a. Folding(4)

<cyclophilin>					
Contig262	939	1.3e-93	155	733	sp|P10255|CYPH_NE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE PRECURSOR (PPIASE)(ROTAMASE) (CYCLOPHILIN) (CYCLOSPORIN A-BINDING PROTEIN) (CPH)
Contig488	669	5.6e-65	121	657	dbj|BAA34384| (AB019518) cyclophilin [Trichophyton mentagrophytes]
<PROTEIN DISULFIDE ISOMERASE PRECURSOR>					
Contig103	267	8.5e-22	225	509	sp|P55059|PDI_HUM PROTEIN DISULFIDE ISOMERASE PRECURSOR (PDI) >pir||JC2291protein disulfide-isomerase (EC 5.3.4.1) - H
<FK506-BINDING PROTEIN>					
Contig2	452	5.3e-42	236	553	sp|P20080|FKBP_NE FK506-BINDING PROTEIN (FKBP) (PEPTIDYL-PROLYL CIS-TRANSISOMERASE) (PPIASE) >pir||S11090 FK506-bindin

b. Chaperones (12)

<heat-shock protein>					
Contig116	239	4.1e-39	338	505	gi|4099014 (U81786) heat-shock protein [Coccidioides immitis]
<heat-shock protein30>					
Contig106	640	6.4e-62	113	772	sp|P19752|HS30_NE 30 KD HEAT SHOCK PROTEIN >pir||A38360 heat shock protein30 -

					Neurospora crassa >gi|168820 (M55672) h
Contig141	178	6e-13	390	497	sp|P19752|HS30_NE 30 KD HEAT SHOCK PROTEIN >pir||A38360 heat shock protein30 - Neurospora crassa >gi|168820 (M55672) h
Contig459	186	1.4e-12	597	956	dbj|BAA33053| (AB003518) heat shock protein [Coriolus versicolor]>dbj|BAA76590.1| (AB018406) heat shock protein 30
Contig46	147	2.3e-08	312	623	dbj|BAA76589.1| (AB018405) fdd123b [Coriolus versicolor] >dbj|BAA76591.1|(AB018407) heat shock protein 30 [Coriolus
<Chaperonin hsp78p>					
Contig133	623	3.7e-60	1	516	emb|CAA20737.1| (AL031534) Chaperonin hsp78p [Schizosaccharomyces pombe]
Contig163	191	4.2e-13	318	506	emb|CAA20737.1| (AL031534) Chaperonin hsp78p [Schizosaccharomyces pombe]
<heat shock protein 70>					
Contig318	894	7e-89	2	646	sp|Q01233|HS70_NE HEAT SHOCK 70 KD PROTEIN (HSP70) >gi|607818 (U10443) 70kDa heat shock protein [Neurospora crassa]
Contig505	786	2.1e-77	307	945	sp|Q01233|HS70_NE HEAT SHOCK 70 KD PROTEIN (HSP70) >gi|607818 (U10443) 70kDa heat shock protein [Neurospora crassa]
Contig164	645	1.8e-62	186	560	sp|Q01233|HS70_NE HEAT SHOCK 70 KD PROTEIN (HSP70) >gi|607818 (U10443) 70kDa heat shock protein [Neurospora crassa]
<MOD-E>					
Contig450	740	1.4e-72	2	664	gi|2804612 (U81165) MOD-E [Podospora anserina]
<activator of Hsp70 and Hsp90 chaperones>					
Contig139	199	3.2e-14	211	444	emb|CAB39910.1 (AL049498) activator of Hsp70 and Hsp90
c. Protein sorting (3)					
<VACUOLAR PROTEASE A PRECURSOR>					
Contig79	641	5e-62	102	476	sp|Q01294|CARP_NE VACUOLAR PROTEASE A PRECURSOR >gi|1039445 (U36471)vacuolar protease A [Neurospora crassa]
Contig229	504	1.8e-47	243	539	sp|Q01294|CARP_NE VACUOLAR PROTEASE A PRECURSOR >gi|1039445 (U36471)vacuolar protease A [Neurospora crassa]
<Golgi membrane protein>					
Contig402	175	1.2e-12	1	240	emb|CAA22273| (AL034381) putative Golgi membrane protein [Schizosaccharomycespombe]
2.4.Turnover-protein degradation-including vacuolar (11)					
<protease subunit>					
b8h08nm.f1	391	1.8e-34	8	298	emb|CAB38512.1| (AL035637) putative protease subunit; chaperonin[Schizosaccharomyces pombe]
<PROTEASOME COMPONENT SUN4>					
Contig214	331	3.7e-29	205	696	sp|P53616|SUN4_YE PROTEASOME COMPONENT SUN4 >pir||S53916 SUN4 proteinprecursor - yeast (Saccharomyces cerevisiae) >emb
<26S PROTEASE REGULATORY SUBUNIT 8>					
Contig255	603	5e-58	101	523	sp|Q01939|PRS8_YE 26S PROTEASE REGULATORY SUBUNIT 8 HOMOLOG (SUG1 PROTEIN)(CIM3 PROTEIN) (TAT-BINDING PROTEIN TBY1) >p
<ubiquitin precursor>					
Contig513	1145	1.9e-115	239	925	pir||UQNC ubiquitin precursor - Neurospora crassa >emb|CAA31530| (X13140)ubiquitin [Neurospora crassa]

Contig508	1106	2.5e-111	111	773	pir||UQNC ubiquitin precursor - Neurospora crassa >emb|CAA31530| (X13140)ubiquitin [Neurospora crassa]
Contig74	765	3.7e-75	209	667	pir||UQNC ubiquitin precursor - Neurospora crassa >emb|CAA31530 (X13140)ubiquitin [Neurospora crassa]
Contig181	385	7.2e-35	239	469	pir||UQNC ubiquitin precursor - Neurospora crassa >emb|CAA31530| (X13140)ubiquitin [Neurospora crassa]
<ubiquitin conjugating enzyme UBC1>					
Contig162	753	6.2e-74	100	540	gi|3323498 (AF030296) ubiquitin conjugating enzyme UBC1 [Glomerella cingulata]
<ubiquitin / ribosomal protein S27a>					
Contig245	318	4.2e-37	180	434	pir||UQNCR ubiquitin / ribosomal protein S27a - Neurospora crassa (fragment)>emb|CAA33390| (X15338) UBI 3 fusio
<methionine aminopeptidase>					
Contig403	685	9.2e-67	3	881	emb|CAA19013| (AL023534) putative methionine aminopeptidase 1[Schizosaccharomyces pombe]
<progesterone-binding protein>					
Contig330	191	2e-14	134	475	gb|AAD34615.1|AF1 (AF153283) putative progesterone-binding proteinhomolog [Arabidopsis thaliana]

III:Cell growth, cell division and cell process (77)

A. Cell growth, cell division (28)

1. Cell walls (10)

<N,O-DIACETYLMURAMIDASE>					
Contig518	805	1.9e-79	103	705	sp|P00721|LYCH_CH N,O-DIACETYLMURAMIDASE (LYSOZYME CH) >pir||MUKAD lysozyme(EC 3.2.1.17) - fungus (Chalara sp.)
Contig482	376	6.5e-34	291	563	sp|P00721|LYCH_CH N,O-DIACETYLMURAMIDASE (LYSOZYME CH) >pir||MUKAD lysozyme(EC 3.2.1.17) - fungus (Chalara sp.)
<cell wall protein>					
Contig274	302	3.4e-26	157	669	emb|CAA09585.1| (AJ011296) putative cell wall protein [Emericella nidulans]
<EPD1 PROTEIN PRECURSOR>					
b5d02nm.f1	373	1.4e-33	145	576	sp|P56092|EPD1_CA EPD1 PROTEIN PRECURSOR >dbj|BAA21103| (AB005130) EPD1[Candida maltosa]
<RODLET PROTEIN>					
Contig522a	205	8.3e-16	2	142	>sp|Q04571|RODL_NEUCR HYDROPHOBIN PRECURSOR (RODLET PROTEIN) (CLOCK-CONTROLLED GENE PROTEIN 2) (BLUE LIGHT INDUCED PROTEIN 7)
Contig522b	451	6.5e-42	377	700	>sp|Q04571|*RODL_NEUCR HYDROPHOBIN* PRECURSOR (RODLET PROTEIN) (CLOCK-CONTROLLED GENE PROTEIN 2) (BLUE LIGHT INDUCED PROTEIN 7)
Contig522c	451	6.5e-42	101	424	>sp|Q04571|RODL_NEUCR HYDROPHOBIN PRECURSOR (RODLET PROTEIN) (CLOCK-CONTROLLED GENE PROTEIN 2) (BLUE LIGHT INDUCED PROTEIN 7)
Contig421	275	2.7e-23	400	558	sp|Q04571|RODL_NE HYDROPHOBIN PRECURSOR (RODLET PROTEIN) (CLOCK-CONTROLLEDGENE PROTEIN 2) (BLUE LIGHT INDUCED PROTEIN
Contig357	205	7.7e-16	358	498	sp|Q04571|RODL_NE HYDROPHOBIN PRECURSOR (RODLET PROTEIN) (CLOCK-CONTROLLEDGENE PROTEIN 2) (BLUE LIGHT INDUCED PROTEIN

Contig301	188	4.7e-14	378	500	sp|Q04571|RODL_NE HYDROPHOBIN PRECURSOR (RODLET PROTEIN) (CLOCK-CONTROLLEDGENE PROTEIN 2) (BLUE LIGHT INDUCED PROTEIN
2.Biomembranes (7)					
<membrane protein>					
Contig61	387	4.1e-35	4	534	pir||S67049 probable membrane protein YOR161c - yeast (Saccharomycescerevisiae) >gi|1293722 (U55021) O3568p [Sac
i5g01nm.r1	253	7e-21	52	330	sp|P38301|YB37_YE HYPOTHETICAL 30.3 KD PROTEIN IN MBA1-RPS13 INTERGENICREGION >pir||S46059 probable membrane protein Y
Contig365	250	1.4e-20	310	621	sp|P40515|YIG5_YE HYPOTHETICAL 17.7 KD PROTEIN IN RNR3-ARC15 INTERGENICREGION >pir||S48414 probable membrane protein Y
Contig286	235	1.1e-17	341	562	sp|P53920|YNM3_YE HYPOTHETICAL 110.9 KD PROTEIN IN SPC98-TOM70 INTERGENICREGION >pir||S63064 probable membrane protein
Contig201	136	1.7e-08	328	465	sp|P47111|YJ14_YE HYPOTHETICAL 15.7 KD PROTEIN IN NUP85- INTERGENICREGION >pir||S57063 probable membrane protein Y
Contig227	126	2e-07	329	496	sp|P47111|YJ14_YE HYPOTHETICAL 15.7 KD PROTEIN IN NUP85-SSC1 INTERGENICREGION >pir||S57063 probable membrane protein Y
Contig39	124	9.2e-05	100	501	pir||S61117 probable membrane protein YDR262w - yeast (Saccharomycescerevisiae) >emb|CAA92580| (Z68290) unknown
3. Cytoskeleton, organelle biogenesis (7)					
<TUBULIN ALPHA CHAIN>					
Contig219	752	8.1e-74	44	460	sp|Q92335|TBA_SOR TUBULIN ALPHA CHAIN >emb|CAA94304| (Z70290) alpha-tubulin[Sordaria macrospora]
Contig156	294	4.3e-25	346	534	sp|Q92335|TBA_SOR TUBULIN ALPHA CHAIN >emb|CAA94304| (Z70290) alpha-tubulin[Sordaria macrospora]
<actin>					
Contig384	867	5.9e-86	123	785	dbj|BAA74960| (AB003111) actin [Humicola grisea var. thermoidea]
<ARP2/3 COMPLEX>					
Contig298	491	4.1e-46	170	598	sp|O14241|AR34_SC PROBABLE ARP2/3 COMPLEX 34 KD SUBUNIT (P34-ARC)>emb|CAB11733| (Z98981) probable Arp2-3 complex subun
<PEROXISOMAL MEMBRANE PROTEIN PMP20>					
Contig352	300	7e-26	9	416	sp|O14313|PM20_SC PEROXISOMAL MEMBRANE PROTEIN PMP20 >emb|CAA05528|(AJ002536) PMP20 [Schizosaccharomyces pombe]
<OLEATE-INDUCED PEROXISOMAL PROTEIN POX18>					
Contig372	243	7.4e-20	284	661	sp|P22009|PX18_CA OLEATE-INDUCED PEROXISOMAL PROTEIN POX18 (LIPID-TRANSFERPROTEIN) (PXP-18) >pir||JS0155 POX18 protein
<3-KETOACYL-COA THIOLASE PEROXISOMAL PRECURSOR>					
a3b07nm.f1	349	3.3e-31	26	496	sp|Q05493|THIK_YA 3-KETOACYL-COA THIOLASE PEROXISOMAL PRECURSOR(BETA-KETOTHIOLASE) (ACETYL-COA ACYLTRANSFERASE) (PEROX
4. Cell cycle control (2)					
<CELL DIVISION CONTROL PROTEIN 11>					
Contig77	158	4.7e-10	265	489	sp|P32458|CC11_YE CELL DIVISION CONTROL PROTEIN 11 >pir|| celldivision control protein CDC11 - yeast (Saccharomy
<Trp-Asp repeat protein>					
Contig190	116	0.00031	24	350	emb|CAB52267.1| (AL109739) trp-asp repeat protein [Schizosaccharomyces pombe]
6. Mitosis/cytokinesis (1)					

<PSI protein>
Contig500 314 2.3e-27 4 357 sp|Q09912|PSI_SCH PSI PROTEIN >pir||S55900 DNAJ-like protein homolog fission yeast (Schizosaccharomyces pombe) >gi|95
REQUIRED FOR NUCLEAR MIGRATION DURING MITOSIS. IT IS REQUIRED FOR THE NORMAL INITIATION OF TRANSLATION

6. Other (1)
<keratin>
Contig490 129 2.9e-05 91 483 pir||A60830 keratin, 70k type II, epidermal - mouse (fragment >gi|387392(M24151) keratin [Mus musculus]

B. Cell processes (49)

1. Cell rescue, defense, osmotic adaptation, starvation response, development (asexual, sexual) (includes antibiotics, toxins)see also B.cell signalling, signal transduction and C. transmembrane transport (28)

1.1 Development (6)

a. Asexual (3)
<CONIDIATION-SPECIFIC PROTEIN 8>
Contig45 312 3.1e-27 190 417 sp|P10169|CON8_NE CONIDIATION-SPECIFIC PROTEIN 8 >pir||S02210 con-8 protein-Neurospora crassa >emb|CAA30092| (X07040)

<Modin>
Contig477 192 2.6e-13 315 482 gi|3115381 (AF025289) Modin [Podospora anserina]

<COPROPORPHYRINOGEN III OXIDASE PRECURSOR>
Contig178 368 3.8e-33 154 474 sp|P36551|HEM6_HU COPROPORPHYRINOGEN III OXIDASE PRECURSOR(COPROPORPHYRINOGENASE) (COPROGEN OXIDASE) (COX) >pir||I3725

b. Morphology, sporulation, conidiation and growth of Fungi (2)
<EPD2>
Contig321 114 0.0015 2 145 dbj|BAA32730| (AB011286) EPD2 [Candida maltosa]

<SPS2 protein>
Contig514 482 3.6e-45 306 1301 pir||S70297 SPS2 protein homolog YBR078w - yeast (Saccharomyces cerevisiae)>emb|CAA85022| (Z35947) ORF YBR078w [

c. Fungi pathgen(cause disease) (1)
<snodprot1>
Contig496 405 4.4e-37 429 776 gi|3329509 (AF074941) snodprot1 [Phaeosphaeria nodorum]

1.2.Detoxification (6)
<CATALASE A>
Contig218 736 3.1e-81 188 769 sp|P78574|CATA_AS CATALASE A >gi|1843578 (U87630) catalase [Aspergillusfumigatus]
Contig404 411 6.7e-37 91 624 sp|P78574|CATA_AS CATALASE A >gi|1843578 (U87630) catalase [Aspergillusfumigatus]

<super oxide dismutase>
c1d10nm.r2 449 1e-41 118 432 sp|P07509|SODC_NE SUPEROXIDE DISMUTASE (CU-ZN) >pir||A36591 superoxidedismutase (EC 1.15.1.1) (Cu-Zn) - Neurospora cra

<SUPEROXIDE DISMUTASE PRECURSOR>

Contig136	622	4.5e-60	87	698	sp|P00447|SODM_YE SUPEROXIDE DISMUTASE PRECURSOR (MN) >pir||DSBYNsuperoxide dismutase (EC 1.15.1.1) (Mn) precursor - y

<catalase-peroxidase>

Contig14	672	2.5e-65	52	672	gi|2605730 (AF027168) catalase-peroxidase [Caulobacter crescentus]

<catalase/peroxidase>

b3e01nm.r1	162	5e-10	380	508	emb|CAA74698| (Y14317) catalase/peroxidase [Streptomyces reticuli]

1.3. Oxidative stress (3)

<flavohemoglobin>

Contig300	479	6.6e-45	106	657	dbj|BAA33011| (AB016807) flavohemoglobin [Fusarium oxysporum]
Contig315	365	9e-33	93	698	dbj|BAA33011| (AB016807) flavohemoglobin [Fusarium oxysporum]

<manganese superoxide dismutase>

Contig436	1107	2.1e-111	198	869	gb|AAD28503.1|AF1 (AF118809) manganese superoxide precursor[Neurospora crassa]

1.4.Stress-inducible protein (3)

<78 KD GLUCOSE REGULATED PROTEIN HOMOLOG PRECURSOR>

Contig308	821	4e-81	3	494	sp|P78695|GR78_NE 78 KD GLUCOSE REGULATED PROTEIN HOMOLOG PRECURSOR (GRP78) (IMMUNOGLOBULIN HEAVY CHAIN BINDING PROTEI
Contig275	565	5.3e-54	261	608	sp|P78695|GR78_NE 78 KD GLUCOSE REGULATED PROTEIN HOMOLOG PRECURSOR (GRP78) (IMMUNOGLOBULIN HEAVY CHAIN BINDING PROTEI

<cyclophilin B>

Contig291	409	1.9e-37	216	512	gb|AAD17998| (AF107254) cyclophilin B; CYPB [Emericella nidulans]

1.5.Night/day rhythm (circadium rhythm-biological clock) (10)

<GLUCOSE-REPRESSIBLE GENE PROTEIN>

Contig494	359	3.5e-32	73	285	sp|P22151|GRG1_NE GLUCOSE-REPRESSIBLE GENE PROTEIN >emb|CAA32907| (X14801)grg1 [Neurospora crassa]
Contig521	359	3.5e-32	392	604	sp|P22151|GRG1_NE GLUCOSE-REPRESSIBLE GENE PROTEIN >emb|CAA32907| (X14801)grg1 [Neurospora crassa]
Contig411	359	3.6e-32	263	475	sp|P22151|GRG1_NE GLUCOSE-REPRESSIBLE GENE PROTEIN >emb|CAA32907| (X14801)grg1 [Neurospora crassa]
Contig129	354	1.2e-31	255	464	sp|P22151|GRG1_NE GLUCOSE-REPRESSIBLE GENE PROTEIN >emb|CAA32907| (X14801)grg1 [Neurospora crassa]
Contig62	330	4.5e-29	245	457	sp|P22151|GRG1_NE GLUCOSE-REPRESSIBLE GENE PROTEIN >emb|CAA32907| (X14801)grg1 [Neurospora crassa]

<clock-controlled gene-6 protein>

Contig423	398	2.6e-36	144	443	gi|3746899 (AF088908) clock-controlled gene-6 protein [Neurospora crassa]

<clock-controlled gene-8 protein>

Contig462	827	1.1e-81	3	497	gi|3746897 (AF088907) clock-controlled gene-8 protein [Neurospora crassa]

<clock-controlled gene-9 protein>

Contig444	725	7.3e-104	3	422	gi|3746895 (AF088906) clock-controlled gene-9 protein [Neurospora crassa]
Contig443	896	5e-89	94	606	gi|3746895 (AF088906) clock-controlled gene-9 protein [Neurospora crassa]

Contig215	415	2.4e-37	131	391	gi|3746895 (AF088906) clock-controlled gene-9 protein [Neurospora crassa]
2. Cell signalling, signal transduction (9)					
2.1. Kinases (2)					
<protein kinase C>					
c1d05nm.f1	638	9.5e-61	28	450	emb|CAA72731| (Y12002) protein kinase C homologue [Neurospora crassa]
<SERINE/THREONINE-PROTEIN KINASE>					
Contig124	227	8.1e-17	107	352	emb|CAA22846| (AL035248) serine threonine-protein kinase [Schizosaccharomycespombe]
2.2 Calmodulin (2)					
<calcium binding protein>					
Contig287	228	3e-18	319	597	gi|3342794 (AF035606) calcium binding protein [Homo sapiens]
<VU91D calmodulin>					
Contig93	97	0.00081	327	431	gi|3800851 (AF081671) VU91D calmodulin [synthetic construct]
2.3. G protein (2)					
<GTP-BINDING NUCLEAR PROTEIN GSP2/CNR2>					
Contig323	607	1.9e-58	189	674	sp|P32836|GSP2_YE GTP-BINDING NUCLEAR PROTEIN GSP2/CNR2 >pir||S35505GTP-binding protein GSP2 - yeast (Saccharomyces ce
<DEVELOPMENTAL REGULATOR FLBA>					
Contig161	363	1.1e-31	77	430	sp|P38093|FLBA_EM DEVELOPMENTAL REGULATOR FLBA >pir|| developmentalregulator flbA - Emericella nidulans >gi|4023
2.4. Membrane receptor (2)					
<peroxisomal receptor for PTS2-containing proteins Pex7p>					
Contig456	231	2.3e-28	230	475	gi|2992543 (AF021797) peroxisomal receptor for PTS2-containing proteins Pex7p[Pichia pastoris]
Contig369	233	1.7e-18	230	601	gi|2992543 (AF021797) peroxisomal receptor for PTS2-containing proteins Pex7p[Pichia pastoris]
2.5.Hormone (1)					
<AA-NAT enzyme>					
h6allnm.f1	102	0.0003	43	195	gb|AAD21316.1| (AF034081) arylalkylamine N-acetyltransferase [Esox lucius] melatonin-secreted primarily from the pineal gland, plays an important physiological role in synchronizing biological rhythms and neuroendocrine
3. Transmembrane transport (12)					
3.1. Transport (12)					
a. Sugar transport (6)					
<sugar transport>					
Contig264	176	6.5e-12	388	669	emb|CAB16808.1| (Z99708) sugar transporter like protein [Arabidopsis thaliana]
<monosaccharide transporter>					
Contig473	672	2.7e-65	1	723	emb|CAB06078| (Z83828) AmMst-1 [Amanita muscaria]

Contig217	237	1.9e-18	301	504	emb|CAB06078| (Z83828) AmMst-1 [Amanita muscaria]
<GLUCOSE TRANSPORTER RCO-3>					
Contig339	510	4e-48	43	699	sp|Q92253|RCO3_NE PROBABLE GLUCOSE TRANSPORTER RCO-3 >gi|1314738 (U54768)RCO3 [Neurospora crassa]
Contig233	317	5.1e-27	175	507	sp|Q92253|RCO3_NE PROBABLE GLUCOSE TRANSPORTER RCO-3 >gi| (U54768)RCO3 [Neurospora crassa]
<GLUCOSE/GALACTOSE TRANSPORTER>					
Contig512b	434	4.6e-40	231	1376	sp|Q44623|GLUP_BR GLUCOSE/GALACTOSE TRANSPORTER >gi|1171339 (U43785) glucose/galactose transporter [Brucella abortus]
b. Cation transport-ATPase, or major facilitator superfamily (1)					
<ZRT1 PROTEIN>					
Contig351	364	1.1e-32	206	622	sp|P32804|ZRT1_YE ZRT1 PROTEIN >pir||S33654 zinc transport protein, highaffinity - yeast (Saccharomyces cerevisiae) >e
c. Anion transport (2)					
<chloride channel protein>					
c2b04nm.f1	137	2.1e-07	3	356	emb|CAA92728.1| (Z68334) Similarity to Rat chloride channel protein (PIR Acc.No. S47327); cDNA EST yk354e5.3 comes f
Contig250	104	0.045	460	636	emb|CAA92728.1| (Z68334) Similarity to Rat chloride channel protein (PIR Acc.No. S47327); cDNA EST yk354e5.3 comes f
d. Protein, amino acid transport (3)					
<NUCLEAR SEGREGATION PROTEIN BFR1>					
f5f09nm.f1	180	1.5e-12	19	468	sp|P38934|BFR1_YE NUCLEAR SEGREGATION PROTEIN BFR1 >pir||S47887 BFR1protein - yeast (Saccharomyces cerevisiae) >gi|458
<autophagy protein>					
Contig420	493	2.5e-46	2	316	emb|CAA21809| (AL032684) putative autophagy protein [Schizosaccharomycespombe]
<IMPORTIN BETA-1 SUBUNIT>					
Contig463	193	2.9e-13	10	228	sp|Q06142|IMB1_YE IMPORTIN BETA-1 SUBUNIT (KARYOPHERIN BETA-1 SUBUNIT)(IMPORTIN 95) >pir||S51350 KAP95 protein - yeast

IV. Unclassified, unidentified, no significant homology (365)

A. Classes of enzymes (from M. Riley and KEGG; no pathway specified) (4)

1. Oxidoreductases (2)					
<cytochrome P450 monooxygenase>					
Contig33	188	5.7e-14	130	261	emb|CAA06156| (AJ004810) cytochrome P450 monooxygenase [Zea mays]
<2-HYDROXYACID DEHYDROGENASE>					
Contig20	195	2e-14	202	465	sp|P44501|DDH_HAE 2-HYDROXYACID DEHYDROGENASE HOMOLOG >pir||F640472-hydroxyaciddehydrogenase homolog (ddh) homolog - H
2. Transferases (1)					
<tyrosine aminotransferase>					

Contig247	148	5e-09	74	568	dbj|BAA81592.1| (AP000064) 397aa long hypothetical tyrosine aminotransferase [Aeropyrum pernix]
3.Hydrolases (1)					
<fatty acid omega-hydroxylase>					
c1c02nm.r1	230	3.8e-17	357	587	dbj|BAA82526.1| (AB030037) fatty acid omega-hydroxylase (P450foxy) [Fusariumoxysporum]

B. Non-enzymatic classes (not in defined pathways) (1)

1. Leucine zipper motif

<Ern4p>					
Contig76	143	2.2e-07	2	547	dbj|BAA19565| (D86413) Ern4p [Saccharomyces cerevisiae >dbj|BAA24425.1|(D50617) bZIP protein binding to the CRE m

C. Unclassified (significant homolog but function uncertain in Neurospora crassa) (5)

<fadE2>					
Contig360	154	1.3e-09	334	510	emb|CAB07030| (Z92770) fadE2 [Mycobacterium tuberculosis]
<alpha NAC/1.9.2. protein>					
Contig16	234	1.3e-31	166	354	gi|1142653 (U22151) alpha NAC/1.9.2. protein [Mus musculus] >gi|1666690(U48363) alpha-NAC, non-muscle form [Mus
<IgE-binding protein>					
Contig515	218	3.6e-17	243	776	emb|CAA07186| (AJ006688) IgE-binding protein [Aspergillus fumigatus]
<PAB1324>					
b3c01nm.f1	109	0.00088	317	598	emb|CAB50468.1| (AJ248288) PAB1324 [Pyrococcus abyssi]
<dehydrogenase>					
Contig493	284	3.3e-24	196	906	emb|CAB02087.1| (Z79753) similar to dehydrogenase [Caenorhabditis elegans]

D. Unidentified (includes significant match with ORFs) (25)

<unknown function>					
Contig273	662	3.4e-64	17	493	pir||S64841 hypothetical protein YLR019w - yeast (Saccharomyces cerevisiae)>emb|CAA97541| (Z73191) ORF YLR019w [
Contig34	454	2.2e-42	2	490	sp|Q03655|YM64_YE HYPOTHETICAL 56.8 KD PROTEIN IN SCJ1-GUA1 INTERGENICREGION PRECURSOR >pir||S55097 probable membrane
Contig122	211	2.4e-36	335	496	emb|CAA81918| (Z28080) ORF YKL081w [Saccharomyces cerevisiae]
Contig188	315	1.6e-27	233	589	gb|AAD46837.1|AF1 (AF160897) BcDNA.GM14838 [Drosophila melanogaster]
Contig209	292	5.1e-25	15	395	sp|P36039|YKU7_YE HYPOTHETICAL 29.4 KD PROTEIN IN STE6-LOS1 INTERGENICREGION >pir||S38045 hypothetical protein YKL207w
Contig152	285	2.8e-24	245	466	pir||S04556 hypothetical protein L - Neurospora crassa mitochondrion (SGC3)>emb|CAA31720| (X13337) ORF L [Neuros
Contig138	274	9.3e-23	194	505	sp|P36091|YKE6_YE HYPOTHETICAL 49.6 KD PROTEIN IN ELM1-PRI2 INTERGENICREGION >pir||S37867 hypothetical protein YKL046c
Contig21	266	1e-21	4	411	sp|Q09844|YAE3_SC HYPOTHETICAL 54.3 KD PROTEIN C23D3.03C IN CHROMOSOME I>pir||S62494

				hypothetical protein SPAC23D3.03c
Contig400	232	2.5e-18	213 533	sp|Q04951|YM8Z_YE HYPOTHETICAL 40.5 KD PROTEIN IN UBP15-GAS1 INTERGENICREGION PRECURSOR >pir||S53975 probable membrane
Contig171	216	5.5e-17	329 604	sp|P40008|YEJ4_YE HYPOTHETICAL 25.1 KD PROTEIN IN PMI40-PAC2 INTERGENICREGION >pir||S50462 hypothetical protein YER004
Contig311	203	1.4e-15	490 633	sp|P76118|YNCH_EC HYPOTHETICAL 8.3 KD PROTEIN IN ANSP-RHSE INTERGENICREGION >gi|1787727 (AE000242) orf, hypothetical p
Contig166	198	3.3e-14	155 475	sp|P53189|YGC8_YE HYPOTHETICAL 56.4 KD PROTEIN IN RPL32-CWH41 INTERGENICREGION PRECURSOR >pir||S64030 probable membran
Contig246	197	5.5e-14	21 479	sp|P38817|YHQ8_YE HYPOTHETICAL 64.3 KD PROTEIN IN CDC12-ERP5 INTERGENICREGION pir||S48950 hypothetical protein YHR108
Contig395	178	6.3e-13	112 396	pir||S66926 hypothetical protein YOR052c - yeast (Saccharomyces cerevisiae)>emb|CAA99244| (Z74960) ORF YOR052c [
Contig373	165	5.9e-11	2 331	gb|AAD18296| (AE001600) Hypothetical Protein [Chlamydia pneumoniae]
Contig249	177	1.4e-10	154 390	emb|CAB43253.1| (AL050060) hypothetical protein [Homo sapiens]
Contig378	143	2.9e-09	3 239	pir||S67201 hypothetical protein YOR297c - yeast (Saccharomyces cerevisiae)>emb|CAA99525| (Z75205) ORF YOR297c [
Contig10	137	1.4e-06	128 406	gi|1240076 (M32346) orf353; putative [Methylobacterium sp.]
Contig198	134	4.2e-06	387 620	sp|P38260|YBV1_YE HYPOTHETICAL 32.6 KD PROTEIN IN VPS15-YMC2 INTERGENICREGION >pir||S48266 hypothetical protein YBR101
Contig27	121	4.3e-06	288 479	sp|Q04951|YM8Z_YE HYPOTHETICAL 40.5 KD PROTEIN IN UBP15-GAS1 INTERGENICREGION PRECURSOR >pir||S53975 probable membrane
Contig159	118	8.4e-06	238 360	gb|AAD18296| (AE001600) Hypothetical Protein [Chlamydia pneumoniae]
Contig261	133	2.5e-05	389 676	emb|CAB37070| (AJ133034) hypothetical protein [Chlamydia pneumoniae]
Contig461	131	4.1e-05	402 1097	emb|CAA22582| (AL034583) hypothetical protein [Schizosaccharomyces pombe]
Contig243	128	8e-05	270 455	gb|AAD18296| (AE001600) Hypothetical Protein [Chlamydia pneumoniae]
Contig235	99	0.085	29 148	emb|CAB16400.1| (Z99261) hypothetical protein [Schizosaccharomyces pombe]

E. No significant homolog (330)

<NONE>

1. **Contigs (285)**
2. **Singlets (45)**

Appendix IV. *Neurospora crassa* evening library Categories of Cellular Functions

I: Metabolisms and bioenergetics (216)

Part one: Metabolism (93)

A. Metabolism of carbohydrates (for glucose see energy) (7)

1. Chitin metabolism (2)

<chitin synthase>

NE.Contig447	752	7e-74	1	717	pir||JC2408 chitin synthase (EC 2.4.1.16), class I - Emericella nidulans>prf||2102237A chitin synthase

<CHITIN BIOSYNTHESIS PROTEIN>

NE.Contig115	489	5.3e-46	83	712	sp|Q92357|CHS5_SCHPO PROBABLE CHITIN BIOSYNTHESIS PROTEIN C6G9.12 (CHS5HOMOLOG) >gnl|PID|e276616 (Z81317) yeast

2. Cellulose degradation (1)

<secreted glucosidase>

c1c06ne.r1	213	5.1e-14	2	232	gnl|PID|e1311957 (AL031107) putative secreted glucosidase [Streptomycescoelicolor]

3. Galactose metabolism (2)

<UDP-GLUCOSE 4-EPIMERASE>

e8b09ne.f1	446	5.9e-41	2	445	sp|P04397|GALX_YEAST UDP- GLUCOSE 4-EPIMERASE (GALACTOWALDENASE) / ALDOSE1-EPIMERASE (MUTAROTASE) >pir||XEBYUG U

<GALACTOSE-1-PHOSPHATE URIDYLYLTRANSFERASE>

NE.Contig60	216	1.1e-16	643	795	sp|P08431|GAL7_YEAST GALACTOSE-1-PHOSPHATE URIDYLYLTRANSFERASE >pir||XNBYUGUDPglucose--hexose-1-phosphate uridy

4. Mannitol metabolism (1)

<MANNOSYLTRANSFERASE>

NE.Contig49	467	1.3e-43	1	465	sp|P38131|KTR4_YEAST PROBABLE MANNOSYLTRANSFERASE KTR4 >pir||S34024alpha-1,2-mannosyltransferase homolog YBR199

5. Calvin cycle (1)

<RIBULOSE-PHOSPHATE 3-EPIMERASE>

h3a05ne.f1	333	2e-29	126	425	sp|P46969|RPE_YEAST RIBULOSE-PHOSPHATE 3-EPIMERASE (PENTOSE-5-PHOSPHATE3-EPIMERASE) (PPE) (RPE) >pir||S51587 P

B. Metabolism of amino acids and related molecules (35)

1.Arginine metabolism (3)

a. Arginine anabolism-glutamine, CO2 to arginine (1)

<ARGININOSUCCINATE LYASE>

NE.Contig498	448	1.3e-41	6	440	sp|P50514|ARLZ_SCHPO PROBABLE ARGININOSUCCINATE LYASE (ARGINOSUCCINASE) (ASAL)>gi|619217 (U13259) L-argininosuc

b. Arginine catabolism-arginine to proline (2)

<ornithine decarboxylase>

NE.Contig486	732	1e-71	316	801	sp|P27121|DCOR_NEUCR ORNITHINE DECARBOXYLASE (ODC) >pir||A42065 ornithinedecarboxylase

(EC 4.1.1.17) - Neurospo

<ORNITHINE AMINOTRANSFERASE>

NE.Contig120	617	2.2e-78	412	996	sp|Q92413|OAT_EMENI ORNITHINE AMINOTRANSFERASE (ORNITHINE--OXO-ACIDAMINOTRANSFERASE) >gi|1658173 (U74303) orni

2.Aspartic acid metabolism (1)

aspartate anabolism-oxaloacetate, glutamate to aspartate

<ASPARTATE AMINOTRANSFERASE>

a3h08ne.r1	299	8.2e-26	248	544	gi|2690302 (U82470) mitochondrial aspartate aminotransferase precursor [Musmusculus]

3.Glutamine metabolism (2)

<GLUTAMINE SYNTHETASE>

NE.Contig370	734	6e-72	8	499	sp|Q12613|GLNA_COLGL GLUTAMINE SYNTHETASE (GLUTAMATE--AMMONIA LIGASE)>gi|1322275 (L78067) glutamine synthetase
NE.Contig981	284	3e-24	313	501	sp|Q12613|GLNA_COLGL GLUTAMINE SYNTHETASE (GLUTAMATE--AMMONIA LIGASE)>gi|1322275 (L78067) glutamine synthetase

4.Histidine metabolism (3)

<histidine-3 protein>

NE.Contig893	788	1e-77	2	505	gi|2853023 (AF045455) histidine-3 protein [Neurospora crassa]
NE.Contig899	750	1.1e-73	42	497	gi|2853023 (AF045455) histidine-3 protein [Neurospora crassa]
NE.Contig563	681	2.4e-66	1	480	gi|2853025 (AF045456) histidine-3 protein [Neurospora crassa]

5.Isoleucine metabolism (3)

<ketol-acid reductoisomerase>

NE.Contig465	108	1.4e-05	353	430	gnl|PID|d1024906 (AB009603) ketol-acid reductoisomerase [Schizosaccharomycespombe]

<3-methylcrotonyl-CoA carboxylase precursor>

NE.Contig885	455	5.4e-42	1	471	gi|533707 (U12536) 3-methylcrotonyl-CoA carboxylase precursor [Arabidopsisthaliana]

<Acetolactate synthase>

c6e12ne.f1	483	2.2e-45	1	399	gi|2547090 (AF013601) acetolactate synthase [Magnaporthe grisea]

6.Leucine metabolism (1)

<2-isopropylmalate synthase>

NE.Contig865	359	1e-31	200	469	gnl|PID|e1319463 (AL031534) 2-isopropylmalate synthase. [Schizosaccharomycespombe]

7.Lysine metabolism (2)

<SACCHAROPINE DEHYDROGENASE>

NE.Contig570	830	3.7e-82	2	911	sp|P38997|LYS1_YARLI SACCHAROPINE DEHYDROGENASE [NAD+, L-LYSINE FORMING](LYSINE--2-OXOGLUTARATE REDUCTASE) (SDH
NE.Contig685	262	7e-22	323	562	sp|P38997|LYS1_YARLI SACCHAROPINE DEHYDROGENASE [NAD+, L-LYSINE FORMING](LYSINE--2-OXOGLUTARATE REDUCTASE) (SDH

8.Phenylalanine metabolism (2)

<isobutene-forming enzyme and benzoate 4-hydroxylase>

NE.Contig751	146	1.3e-08	364	618	gnl|PID|d1010478 (D63703) isobutene-forming enzyme and benzoate 4-hydroxylase[Rhodotorula minuta]

<ALANINE AMINOTRANSFERASE, MITOCHONDRIALPRECURSOR>

a8g05ne.f1	431	8.1e-40	67	639	sp|P52893|ALAM_YEAST PUTATIVE ALANINE AMINOTRANSFERASE, MITOCHONDRIALPRECURSOR (GLUTAMIC--PYRUVIC TRANSAMINASE)
9.Proline metabolism (5)					
<delta-1-pyroline-5-carboxylatedehydrogenase>					
NE.Contig1038	764	4.3e-75	2	775	gnl|PID|e1326233 (AL031786) putative delta-1-pyroline-5-carboxylatedehydrogenase [Schizosaccharomyces pombe
<proline oxidase>					
NE.Contig727	223	7.4e-15	883	1170	gnl|PID|e1296596 (AL023794) putative proline oxidase precursor[Schizosaccharomyces pombe]
<SPERMIDINE SYNTHASE>					
NE.Contig303	580	1.1e-55	49	492	sp|Q12074|SPEE_YEAST SPERMIDINE SYNTHASE (PUTRESCINE AMINOPROPYLTRANSFERASE)(SPDSY) >pir||S54090 SPE3 protein -
<GAMMA-GLUTAMYL PHOSPHATE REDUCTASE>					
NE.Contig510	470	5.5e-44	2	559	sp|P54902|PROA_SYNY3 PROBABLE GAMMA-GLUTAMYL PHOSPHATE REDUCTASE (GPR)(GLUTAMATE-5-SEMIALDEHYDE DEHYDROGENASE)
NE.Contig296	131	1.5e-06	16	204	gnl|PID|e1351641 (Z54342) 951003: Homology with human lipoprotein-bindingprotein (PIR Acc. No. A44125); cDN
10.Serine metabolism (1)					
<L-serine dehydratase>					
NE.Contig372	112	1.6e-05	223	342	pir||A37817 L-serine dehydratase inactive splice form 2 - rat
11.Tryptophan metabolism (1)					
<ANTHRANILATE SYNTHASE COMPONENT I>					
b9g02ne.r1	195	2.4e-14	194	394	gnl|PID|d1014614 (D89256) similar to *Saccharomyces* cerevisiae anthranilatesynthase component I, SWISS-PROT
12.Tyrosine metabolism (2)					
<fumarylacetoacetate hydrolase>					
NE.Contig381	344	1.4e-30	181	591	gi|1130507 (L41670) fumarylacetoacetate hydrolase [Emericella nidulans]
b9h09ne.f1	330	4.4e-29	26	445	gi|1130507 (L41670) fumarylacetoacetate hydrolase [Emericella nidulans]
13.Valine metabolism (4)					
<METHYLMALONATE-SEMIALDEHYDE DEHYDROGENASE PRECURSOR>					
NE.Contig266	553	7.7e-53	2	478	sp|P52713|MMSA_CAEEL PROBABLE METHYLMALONATE-SEMIALDEHYDE DEHYDROGENASEPRECURSOR (ACYLATING) (MMSDH) >gnl|PID|e
<beta isopropylmalate dehydrogenase>					
NE.Contig267	220	3.8e-17	3	188	prf||2004294A beta isopropylmalate dehydrogenase [Neurospora crassa]
<ketol-acid reductoisomerase precursor>					
h7b06ne.f1	739	1.3e-72	13	441	sp|P38674|ILV5_NEUCR KETOL-ACID REDUCTOISOMERASE PRECURSOR (ACETOHYDROXY-ACIDREDUCTOISOMERASE) (ALPHA-KETO-BETA
<ilvX>					
b3b10ne.f1	116	2e-05	99	392	gnl|PID|e1254637 (AL022022) ilvX [Mycobacterium tuberculosis]
14.Aromatic amino acid metabolism (2)					
<phenylacetyl-CoA ligase>					
a6a04ne.f1	428	1.8e-39	20	658	gnl|PID|e1321225 (AJ001540) phenylacetyl-CoA ligase [Penicillium chrysogenum]

a6h07ne.r1 285 1.4e-23 231 479 gnl|PID|e1321225 (AJ001540) phenylacetyl-CoA ligase [Penicillium chrysogenum]

15.Glutamate metabolism (1)
<glutamate synthase>
NE.Contig546 194 6e-13 253 444 gnl|PID|e1349111 (Z49889) similar to glutamate synthase; cDNA EST EMBL:D27720comes from this gene; cDNA EST

16.Threonine biosynthesis (1)
<L-serine/L-threonine dehydratase>
NE.Contig584 173 2.2e-10 234 653 gi|172589 (M85194) L-serine/L-threonine dehydratase [Saccharomyces cerevisiae]

17. Butanoate metabolism (1)
<SUCCINATE SEMIALDEHYDE DEHYDROGENASE>
b6h07ne.r1 661 2.8e-64 3 620 sp|P51650|SSDH_RAT SUCCINATE SEMIALDEHYDE DEHYDROGENASE (NAD(+)-DEPENDENTSUCCINIC SEMIALDEHYDE DEHYDROGENASE)

C. Metabolism of nucleotides and nucleic acids, purines, pyrimidines (5)

1. Purine metabolism (2)

a. Inosine mono phosphate de novo biosynthesis (1)
<PHOSPHORIBOSYLAMINOIMIDAZOLECARBOXAMIDE FORMYLTRANSFERASE2>
NE.Contig657 161 3.9e-10 2 151 sp|P38009|PU92_YEAST PHOSPHORIBOSYLAMINOIMIDAZOLECARBOXAMIDE FORMYLTRANSFERASE2 (AICAR TRANSFORMYLASE) / IMP CY

b. Other purine metabolic enzymes (1)
<ADENYLATE KINASE 2>
NE.Contig749 208 3.5e-16 74 322 sp|P26364|KAD2_YEAST ADENYLATE KINASE 2 (ATP-AMP TRANSPHOSPHORYLASE)>pir||S23568 adenylate kinase (EC 2.7.4.3)

2. Pyrimidine metabolism (3)

a. De novo pyrimidine biosynthesis
<carbamyl phosphate synthetase>
NE.Contig106 565 2e-52 1 516 gi|854566 (X87371) carbamyl phosphate synthetase [Saccharomyces cerevisiae]

<dihydroorotate dehydrogenase>
NE.Contig515 178 4.6e-12 175 342 gi|1181887 (U47318) dihydroorotate dehydrogenase [Emericella nidulans]

<URIDINE KINASE>
NE.Contig568 126 1.5e-06 257 472 sp|P27515|URK1_YEAST URIDINE KINASE (URIDINE MONOPHOSPHOKINASE) >pir||S29374uridine kinase (EC 2.7.1.48) - yeas

D. Metabolism of lipids, fatty acids, sterols-see also fatty acid degradation (19)

1. Fatty acid biosynthesis (8)

a. ACYL-CARRIER PROTEINS (1)

<3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE>
NE.Contig537 185 9.2e-14 252 605 sp|Q56318|FABG_THEMA 3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE(3-KETOACYL-ACYL CARRIER PROTEIN REDUCTASE) >gnl

b. FATTY ACID SYNTHASE (2)
<acyl-CoA dehydrogenases>
NE.Contig615 353 2.3e-30 40 501 gi|1938424 (U97002) similar to acyl-CoA dehydrogenases and epoxide hydrolases[Caenorhabditis elegans]
NE.Contig606 219 5e-16 205 486 gi|1938424 (U97002) similar to acyl-CoA dehydrogenases and epoxide hydrolases[Caenorhabditis elegans]

c. BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE--keto acids to short branch-chain fatty acids (2)
<BRANCHED-CHAIN ALPHA-KETO ACID DEHYDROGENASE>
NE.Contig88 331 3.2e-29 78 542 bbs|62654 (S62652) branched-chain alpha-keto acid dehydrogenase E1 alphasubunit [human, Peptide, 443
<branched-chain alpha keto-acid dehydrogenase E1 alphasubunit>
NE.Contig108 314 3.1e-27 142 621 gi|3822223 (AF077955) branched-chain alpha keto-acid dehydrogenase E1 alphasubunit [Arabidopsis thali

d. Other (3)
<stearoyl-CoA desaturase>
NE.Contig349 797 1e-78 23 499 pir||S52745 stearoyl-CoA desaturase (EC 1.14.99.5) - Ajellomyces capsulata>gi|757860 (X85963) delta-9
NE.Contig724 287 3.4e-24 425 745 pir||S52746 stearoyl-CoA desaturase (EC 1.14.99.5) - Ajellomyces capsulata>gi|757862 (X85962) delta-9
NE.Contig1024 225 2.9e-17 454 636 pir||S52746 stearoyl-CoA desaturase (EC 1.14.99.5) - Ajellomyces capsulata>gi|757862 (X85962) delta-9

2. Sterols (7)

a. General (5)
<sterol transmethylase>
NE.Contig330 316 1.2e-27 92 493 gi|3323500 (AF031941) sterol transmethylase [Candida albicans]
NE.Contig317 306 1e-26 35 424 gi|3323500 (AF031941) sterol transmethylase [Candida albicans]

<HYDROXYMETHYLGLUTARYL-COA SYNTHASE>
b6c02ne.f1 685 8.7e-67 102 587 sp|P54839|HMCS_YEAST HYDROXYMETHYLGLUTARYL-COA SYNTHASE (HMG-COA SYNTHASE)(3-HYDROXY-3-METHYLGLUTARYL COENZYME

<DIPHOSPHOMEVALONATE DECARBOXYLASE>
c5a02ne.f1 336 8.6e-30 124 471 sp|P32377|ER19_YEAST DIPHOSPHOMEVALONATE DECARBOXYLASE (MEVALONATEPYROPHOSPHATE DECARBOXYLASE) >pir||S63374 dip
NE.Contig792 199 1.1e-14 3 278 gnl|PID|e334372 (Z98601) diphosphomevalonate decarboxylase[Schizosaccharomyces pombe]

b. Cholesterol metabolism (2)
<C-4 METHYL STEROL OXIDASE>
d5h09ne.f1 300 6e-26 132 464 sp|P53045|ER25_YEAST C-4 METHYL STEROL OXIDASE >pir||S64354 ERG25 protein -yeast (Saccharomyces cerevisiae) >gi

<eburicol 14alpha demethylase>
NE.Contig25 247 1.4e-19 3 233 gi|4049645 (AF052515) eburicol 14alpha demethylase; CYP51; cytochrome P450sterol 14-demethylase [Erys

3. Lipids (4)

a. SPHINGOLIPIDS (1)

<serine palmitoyltransferase>

NE.Contig614	292	1.5e-24	290	736	gnl|PID|e1285366 (AL022299) putative serine palmitoyltransferase[Schizosaccharomyces pombe] >gnl|PID|e13627

b. Lipopolysaccharide biosyn-biomembrane precursors (3)

<UDP-glucose:sterol glucosyltransferase>

a9b10ne.f1	499	5.2e-47	41	634	gnl|PID|e1169031 (Z83832) UDP-glucose:sterol glucosyltransferase [Avenasativa]

<UDP-GLUCOSE PYROPHOSPHORYLASE>

c6f01ne.f1	568	1.8e-54	5	469	gnl|PID|e1371798 (AL035259) putative utp--glucose-1-phosphateuridylyltransferase [Schizosaccharomyces pombe
c6f01ne.r1	146	1.3e-08	328	444	gnl|PID|e1348003 (Z81568) similar to UTP--GLUCOSE-1-PHOSPHATEURIDYLYLTRANSFERASE (EC 2.7.7.9) (UDP-GLUCOSE

E. Aromatic compound metabolism (2)

<4-AMINOBUTYRATE AMINOTRANSFERASE>

NE.Contig18	615	2.7e-59	161	634	sp|P14010|GATA_EMENI 4-AMINOBUTYRATE AMINOTRANSFERASE (GAMMA-AMINO-N-BUTYRATETRANSAMINASE) (GABA TRANSAMINASE)
NE.Contig472	424	4e-39	235	621	sp|P14010|GATA_EMENI 4-AMINOBUTYRATE AMINOTRANSFERASE (GAMMA-AMINO-N-BUTYRATETRANSAMINASE) (GABA TRANSAMINASE)

F. Sulfur Metabolism (2)

<adenosine-5'phosphosulphate kinase>

c1b05ne.r1	557	3.6e-53	25	555	gnl|PID|e275051 (Y08866) adenosine-5'phosphosulphate kinase [Emericellanidulans]

<sconCp>

NE.Contig400	130	6.5e-08	426	509	gi|1658298 (U75874) sconCp [Emericella nidulans]

G. Phosphate metabolism (2)

<INORGANIC PYROPHOSPHATASE>

c6b09ne.f1	540	1.9e-51	62	505	sp|P19117|IPYR_SCHPO INORGANIC PYROPHOSPHATASE (PYROPHOSPHATEPHOSPHO-HYDROLASE) (PPASE) >pir||S11496 inorganic
h6e06ne.r1	349	3.5e-31	78	398	sp|O13505|IPYR_PICPA INORGANIC PYROPHOSPHATASE (PYROPHOSPHATEPHOSPHO-HYDROLASE) (PPASE) >gnl|PID|e1180018 (AJ00

H. Nitrogen metabolism (see also amino acid metabolism) (5)

<NITROGEN METABOLIC REGULATION PROTEIN>

c3c08ne.r1	818	8.4e-81	10	540	sp|P23762|NMR_NEUCR NITROGEN METABOLIC REGULATION PROTEIN (NMR PROTEIN)>pir||S11910 nitrogen metabolic regulat

<ALIPHATIC NITRILASE>

b4a10ne.f1	417	2.6e-38	7	567	sp|Q02068|NRL1_RHORH ALIPHATIC NITRILASE >pir||A43470 aliphatic nitrilase -Rhodococcus rhodochrous >gnl|PID|d10
NE.Contig232	182	5.9e-13	84	446	sp|Q03217|NRL2_RHORH ALIPHATIC NITRILASE >pir||A45070 nitrilase - Rhodococcusrhodochrous >gnl|PID|d1002476 (D11 urea cycle

<NAD(+)-specific glutamate dehydrogenase>

NE.Contig593	979	7.1e-98	3	566	bbs|138429 (S66039) NAD(+)-specific glutamate dehydrogenase, NAD-GDH (EC1.4.1.2) [Neurospora crassa,
NE.Contig895	645	9.3e-62	269	658	bbs|138429 (S66039) NAD(+)-specific glutamate dehydrogenase, NAD-GDH (EC1.4.1.2) [Neurospora crassa,

I. Metabolism of cofactors, prosthetic groups (16)

1.Thiamine (9)

<thiamine synthase>

NE.Contig989	242	9.1e-20	290	490	gi|3282220 (U68718) thiamine synthase homolog [Botryotinia fuckeliana]

<THIAZOLE BIOSYNTHETIC ENZYME>

NE.Contig1061	1072	9.5e-108	238	1116	sp|P23618|THI4_FUSOX THIAZOLE BIOSYNTHETIC ENZYME (STRESS-INDUCIBLE PROTEINSTI35) >pir||B37767 stress-inducible
NE.Contig1004	533	1.3e-50	149	511	sp|P23618|THI4_FUSOX THIAZOLE BIOSYNTHETIC ENZYME (STRESS-INDUCIBLE PROTEINSTI35) >pir||B37767 stress-inducible
NE.Contig226	420	1.2e-38	116	418	sp|P23618|THI4_FUSOX THIAZOLE BIOSYNTHETIC ENZYME (STRESS-INDUCIBLE PROTEINSTI35) >pir||B37767 stress-inducible
NE.Contig826	330	4.2e-29	146	484	sp|P23618|THI4_FUSOX THIAZOLE BIOSYNTHETIC ENZYME (STRESS-INDUCIBLE PROTEINSTI35) >pir||B37767 stress-inducible
NE.Contig585	267	1.9e-22	293	613	gnl|PID|e1325578 (AL031743) thiazole biosynthetic enzyme. [Schizosaccharomycespombe]

<NMT1 PROTEIN HOMOLOG>

NE.Contig1021	1034	1e-103	65	763	sp|P42882|NMT1_ASPPA NMT1 PROTEIN HOMOLOG >pir||S53697 nmt1 protein -Aspergillus parasiticus >gi|557050 (U15196
NE.Contig1019	368	4e-33	273	563	sp|P42882|NMT1_ASPPA NMT1 PROTEIN HOMOLOG >pir||S53697 nmt1 -Aspergillus parasiticus >gi|557050 (U15196
NE.Contig469	185	2.8e-13	350	463	sp|P42882|NMT1_ASPPA NMT1 PROTEIN HOMOLOG >pir||S53697 nmt1 protein -Aspergillus parasiticus >gi|557050 (U15196

2.Coenzyme A (2)

<acetyl-coenzyme A synthetase>

NE.Contig467	343	8.8e-30	227	799	gi|2314192 (AE000612) acetyl-CoA synthetase (acoE) [Helicobacter pylori 26695]

<acetyl coenzyme A acetyltransferase>

NE.Contig1017	560	1.5e-53	168	686	gnl|PID|e1343358 (AL033534) acetyl-coa acetyltransferase [Schizosaccharomycespombe]

3. Vitamin (1)

<pyridoxal reductase>

b1h01ne.f1	436	2.4e-40	27	596	gnl|PID|e349610 (Z99262) pyridoxal reductase. [Schizosaccharomyces pombe]

4.Flavins (2)

<6,7-DIMETHYL-8-RIBITYLLUMAZINE SYNTHASE>

c4d09ne.r1	243	6.7e-20	161	403	sp|P50861|RIB4_YEAST 6,7-DIMETHYL-8-RIBITYLLUMAZINE SYNTHASE (DMRL SYNTHASE)(LUMAZINE SYNTHASE) (RIBOFLAVIN SYN
c4d09ne.f1	176	8.7e-13	133	444	sp|P50861|RIB4_YEAST 6,7-DIMETHYL-8-RIBITYLLUMAZINE SYNTHASE (DMRL SYNTHASE)(LUMAZINE SYNTHASE) (RIBOFLAVIN SYN

5.Folate-methyl donor (1)

<non-functional folate binding protein>

NE.Contig631	208	3.6e-16	128	493	gi|2565196 (AF000381) non-functional folate binding protein [Homo sapiens]

6. Heme (1)

<iucB protein>

NE.Contig351	236	4e-19	123	437	pir||S44019 iucB protein - Escherichia coli >gi|474191 (X76100) iucB geneproduct [Escherichia coli]

Part two: Energy (123)

A. Carbohydrate as energy source (55)

1. Glycolysis (13)

1.1 GLUCOKINASE (1)

<GLUCOKINASE>

NE.Contig344	190	1.9e-13	12	356	sp|Q92407|HXKG_ASPNG GLUCOKINASE (GLUCOSE KINASE) (GLK) >gnl|PID|e274340(X99626) glucokinase [Aspergillus niger

1.2 Fructose-bisphosphate aldolase (2)

<fructose-bisphosphate aldolase>

NE.Contig282	326	9.8e-29	169	465	sp|P53444|ALF_NEUCR FRUCTOSE-BISPHOSPHATE ALDOLASE >gi|1334980 (L42380)fructose 1,6 bisphosphate-aldolase [Neu
NE.Contig638	216	1e-16	315	494	sp|P53444|ALF_NEUCR FRUCTOSE-BISPHOSPHATE ALDOLASE >gi|1334980 (L42380)fructose 1,6 bisphosphate-aldolase [Neu

1.3 Triose-phosphate isomerase (2)

<triose-phosphate isomerase>

NE.Contig262	433	4.5e-40	152	508	gnl|PID|e329478 (X64537) triosephosphate isomerase +glyceraldehyde-3-phosphate dehydrogenase [Phytophthora
NE.Contig416	394	5e-36	94	483	sp|P00942|TPIS_YEAST TRIOSEPHOSPHATE ISOMERASE (TIM) >pir||ISBYTtriose-phosphate isomerase (EC 5.3.1.1) - yeast

1.4 Glyceraldehyde-3-phosphate dehydrogenase (3)

<GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE>

NE.Contig1109b	1733	1.0e-177	203	1216	gb|AAB95425.1| (U67457) glyceraldehyde 3-phosphate dehydrogenase [Neurospora crassa] sp|P54118|G3P_NEUCR GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (GAPDH)(CLOCK-CONTROLLED PROTEIN 7)
NE.Contig516	457	1.4e-42	199	468	gi|1532189 (U67457) glyceraldehyde 3-phosphate dehydrogenase [Neurosporacrassa]
NE.Contig240	427	1.9e-39	109	375	sp|P54118|G3P_NEUCR GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (GAPDH)(CLOCK-CONTROLLED PROTEIN 7) >gi|1326237 (

1.5 Phosphoglycerate kinase (2)

<phosphoglycerate kinase>

NE.Contig1102	2127	1.2e-219	199	1452	sp|P38667|PGK_NEUCR PHOSPHOGLYCERATE KINASE >gi|3052 (X56512) phosphoglyceratekinase [Neurospora crassa]
NE.Contig1054	468	8.8e-44	207	485	sp|P38667|PGK_NEUCR PHOSPHOGLYCERATE KINASE >gi|3052 (X56512) phosphoglyceratekinase [Neurospora crassa]

1.6 Pyruvate kinase (1)

<pyruvate kinase>
NE.Contig85 355 1.5e-31 3 254 sp|P31865|KPYK_TRIRE PYRUVATE KINASE >pir||JN0780 pyruvate kinase (EC2.7.1.40) - fungus (Trichoderma reesei) >g

1.7 ENOLASE (2)
<ENOLASE>
g2f02ne.f1 557 3.2e-53 45 467 sp|P42040|ENO_CLAHE ENOLASE (2-PHOSPHOGLYCERATE DEHYDRATASE)(2-PHOSPHO-D-GLYCERATE HYDRO-LYASE) (ALLERGEN CLA
NE.Contig1047 554 6.9e-53 213 623 sp|Q12560|ENO_ASPOR ENOLASE (2-PHOSPHOGLYCERATE DEHYDRATASE)(2-PHOSPHO-D-GLYCERATE HYDRO-LYASE) >pir||JC45426b

2. Gluconeogenesis (5)

2.1.LACTATE DEHYDROGENASE (3)
<D-LACTATE DEHYDROGENASE (CYTOCHROME) PRECURSOR>
c4g02ne.f1 487 9e-46 2 487 sp|P32891|DLD1_YEAST D-LACTATE DEHYDROGENASE (CYTOCHROME) PRECURSOR (D-LACTATEFERRICYTOCHROME C OXIDOREDUCTASE)
NE.Contig499 421 9.7e-39 10 555 sp|Q12627|DLD1_KLULA D-LACTATE DEHYDROGENASE [CYTOCHROME] PRECURSOR (D-LACTATEFERRICYTOCHROME C OXIDOREDUCTASE)
c4g02ne.r1 123 0.0005 372 653 sp|Q12627|DLD1_KLULA D-LACTATE DEHYDROGENASE [CYTOCHROME] PRECURSOR (D-LACTATEFERRICYTOCHROME C OXIDOREDUCTASE)

2.2 Pyruvate carboxylase (1)
<pyruvate carboxylase>
c4c01ne.f1 711 1.1e-68 41 556 gi|3806120 (AF097728) pyruvate carboxylase [Aspergillus terreus]

2.3 Phosphoenolpyruvate carboxykinase (1)
<phosphoenolpyruvate carboxykinase>
NE.Contig1099 639 6.7e-62 466 957 sp|O13434|PPCK_CANAL PHOSPHOENOLPYRUVATE CARBOXYKINASE (ATP) >gi|2267237(U70473) PEP carboxykinase [Candida alb

3. Pentose-phosphate pathway (4)

3.1 Phosphogluconate dehydrogenase (1)
<6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING 1>
a5f03ne.f1 535 7.6e-51 74 583 sp|P38720|6PG1_YEAST 6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING 1>pir||S46671 phosphogluconate dehydroge

3.2 Transketolase (2)
<transketolase I>
NE.Contig329 157 1.2e-09 278 433 gnl|PID|e1341027 (AL033501) transketolase I [Candida albicans]
NE.Contig336 633 2.9e-61 9 497 gnl|PID|d1014531 (D89172) similar to Saccharomyces serevisiae transketolase2(TK2), SWISS-PROT Accession Num

3.3 Transaldolase (1)
<Tal1p transaldolase>
NE.Contig735 413 5.7e-38 1 363 gnl|PID|e1292580 (AL023518) Tal1p transaldolase [Schizosaccharomyces pombe]

4. Pyruvate metabolism (3)

<DIHYDROLIPOAMIDE DEHYDROGENASE PRECURSOR>

c7c02ne.r1	233	4.3e-18	289	450	sp|P09624|DLDH_YEAST DIHYDROLIPOAMIDE DEHYDROGENASE PRECURSOR >pir||A30151dihydrolipoamide dehydrogenase (EC 1.

<pyruvate dehydrogenase E1-beta subunit>

c1f09ne.r1	342	2.1e-30	66	458	gi|171429 (M98476) pyruvate dehydrogenase E1-beta subunit [Saccharomycescerevisiae]

<pyruvate dehydrogenase precursor>

a3d04ne.r1	157	1e-08	291	539	gi|172108 (M29582) pyruvate dehydrogenase precursor (EC 1.2.4.1) [Saccharomyces cerevisiae]

5. Tricarboxylic acid pathway (11)

5.1 Aconitate hydratase (2)

<aconitase>

NE.Contig883	928	1.6e-92	250	924	gi|3661614 (AF093142) aconitase [Aspergillus terreus]
NE.Contig868	453	1.8e-41	250	621	gi|3661614 (AF093142) aconitase [Aspergillus terreus]

5.2 Isocitrate dehydrogenase (1)

<ISOCITRATE DEHYDROGENASE (NADP), MITOCHONDRIAL PRECURSOR>

NE.Contig276	665	1e-64	1	447	sp|P79089|IDHP_ASPNG ISOCITRATE DEHYDROGENASE (NADP), MITOCHONDRIAL PRECURSOR(OXALOSUCCINATE DECARBOXYLASE) (ID

5.3 Alpha-ketoglutarate dehydrogenase (4)

<2-oxoglutarate dehydrogenase e1 component>

NE.Contig931	244	1e-18	427	684	gnl|PID|e1315370 (AL031261) 2-oxoglutarate dehydrogenase e1 component [Schizosaccharomyces pombe]

<2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT PRECURSOR>

NE.Contig812	606	1.3e-57	7	459	gi|171785 (M26390) alpha-ketoglutarate dehydrogenase [Saccharomycescerevisiae]

<DIHYDROLIPOAMIDE SUCCINYLTRANSFERASE>

NE.Contig717	580	1.4e-55	66	503	gnl|PID|e1371871 (AL035263) dihydrolipoamide succinyltransferase component [Schizosaccharomyces pombe]
NE.Contig884	259	3.8e-21	165	425	gnl|PID|e1371871 (AL035263) dihydrolipoamide succinyltransferase component [Schizosaccharomyces pombe]

5.4 SUCCINYL-COA LIGASE (1)

<SUCCINYL-COA LIGASE [GDP-FORMING] ALPHA-CHAIN, MITOCHONDRIAL PRECURSOR>

NE.Contig37	687	5.6e-67	27	578	sp|O13750|SUCA_SCHPO PROBABLE SUCCINYL-COA LIGASE (GDP-FORMING), ALPHA-CHAINPRECURSOR (SUCCINYL-COA SYNTHETASE,

5.5 Fumarase (1)

<fumarase>

NE.Contig333	391	1.2e-35	138	479	gi|2443751 (AF020303) fumarase [Arabidopsis thaliana] >gi|2529676 (AC002535)putative fumarase [Arabid

5.6 Malate dehydrogenase (2)

<malate dehydrogenase>

NE.Contig911	487	8.4e-46	2	469	gnl|PID|e1360271 (Y16748) malate dehydrogenase [Piromyces sp. E2]

<MALATE DEHYDROGENASE, MITOCHONDRIAL PRECURSOR>
NE.Contig1011 1087 1.9e-109 45 1028 sp|P17505|MDHM_YEAST MALATE DEHYDROGENASE, MITOCHONDRIAL PRECURSOR>pir||DEBYMM malate dehydrogenase (EC 1.1.1.3

6. Related reactions (1)

<citrate lyase-citrate to oxaloacetate+acetylcoA>
a5b02ne.f1 305 1.9e-26 7 186 gi|1778535 (U82598) HI0025 homolog [Escherichia coli] >gi|1786835 (AE000166)citrate lyase synthetase

7. Glyoxylate cycle (5)

<malate synthase>
d1h03ne.f1 804 1.8e-79 2 499 sp|P28344|MASY_EMENI MALATE SYNTHASE, GLYOXYSOMAL >pir||S17773 malate synthase(EC 4.1.3.2) - Emericella nidulan
NE.Contig815 714 7e-70 255 734 sp|P28345|MASY_NEUCR MALATE SYNTHASE, GLYOXYSOMAL >pir||S17774 malate synthase(EC 4.1.3.2) - Neurospora crassa
NE.Contig969 542 1.1e-51 167 508 sp|P28345|MASY_NEUCR MALATE SYNTHASE, GLYOXYSOMAL >pir||S17774 malate synthase(EC 4.1.3.2) - Neurospora crassa
NE.Contig681 293 1.4e-24 314 526 sp|P28345|MASY_NEUCR MALATE SYNTHASE, GLYOXYSOMAL >pir||S17774 malate synthase(EC 4.1.3.2) - Neurospora crassa

<isocitrate lyase>
NE.Contig791 688 4.4e-67 330 734 sp|P28299|ACEA_NEUCR ISOCITRATE LYASE (ISOCITRASE) (ISOCITRATASE) (ICL)>pir||S26858 isocitrate lyase (EC 4.1.3.

8. Fermentation, alcoholic (8)

8.1 Alcohol dehydrogenase (8)

<alcohol dehydrogenase>
NE.Contig920 373 1.1e-33 118 552 gnl|PID|e1339990 (AL033389) alcohol dehydrogenase [Schizosaccharomyces pombe]
NE.Contig221 197 1e-14 180 713 gnl|PID|e1339990 (AL033389) alcohol dehydrogenase [Schizosaccharomyces pombe]

<ALCOHOL DEHYDROGENASE I>
NE.Contig1056 1044 7.3e-105 119 919 sp|P41747|ADH1_ASPFL ALCOHOL DEHYDROGENASE I >gi|439867 (L27434) alcoholdehydrogenase [Aspergillus flavus]
NE.Contig1015 223 6.6e-31 96 239 sp|P41747|ADH1_ASPFL ALCOHOL DEHYDROGENASE I >gi|439867 (L27434 alcoholdehydrogenase [Aspergillus flavus]
NE.Contig1100 305 1.8e-26 324 563 sp|O00097|ADH1_PICST ALCOHOL DEHYDROGENASE I (ADH 2) >gnl|PID|e1191432(Y13397) alcohol dehydrogenase [Pichia st

<type III alcohol dehydrogenase>
NE.Contig850 285 2.3e-24 5 403 gi|2431772 (U66411) putative type III alcohol dehydrogenase [Drosophilamelanogaster]

<ZINC-TYPE ALCOHOL DEHYDROGENASE-LIKE PROTEININ>
b7a06ne.f1 377 3.9e-34 178 558 sp|Q04894|YM97_YEAST HYPOTHETICAL ZINC-TYPE ALCOHOL DEHYDROGENASE-LIKE PROTEININ PRE5-FET4 INTERGENIC REGION >p
b7a06ne.r1 215 9.8e-17 210 719 sp|P25377|YCZ5_YEAST HYPOTHETICAL ZINC-TYPE ALCOHOL DEHYDROGENASE-LIKE PROTEININ PAU3 3'REGION >pir||S19417 pro

9. Metabolism of energy reserves (glycogen, starch, trehalose) (5)

9.1 Glycogen degradation (4)

<glycogen phosphorylase>
b6b02ne.f1 376 6.5e-33 167 589 pir||S61144 glycogen phosphorylase (EC 2.4.1.1) - yeast (Saccharomycescerevisiae) >gi|849168 (U28371)

<PHOSPHOGLUCOMUTASE 1>
g9g04ne.r1 317 4.4e-27 145 516 sp|P33401|PGM1_YEAST PHOSPHOGLUCOMUTASE 1 (GLUCOSE PHOSPHOMUTASE 1) (PGM 1)>pir||S41199 phosphoglucomutase (EC

<alpha-amylase>
NE.Contig535 131 4.4e-07 323 442 pir||A45738 alpha-amylase (EC 3.2.1.1), cytoplasmic - Escherichia coli>gi|146023 (L01642) alpha-amylas

<PHOSPHOGLUCOMUTASE 2>
NE.Contig319 661 3.7e-64 6 491 sp|P37012|PGM2_YEAST PHOSPHOGLUCOMUTASE 2 (GLUCOSE PHOSPHOMUTASE 2) (PGM 2)>pir||S41200 phosphoglucomutase (EC

9.2. Starch degradation (1)
<ALPHA-GLUCOSIDASE PRECURSOR>
NE.Contig608 615 1.2e-58 4 513 sp|P56526|AGLU_ASPNG ALPHA-GLUCOSIDASE PRECURSOR (MALTASE) >gnl|PID|d1024495(D45356) alpha-glucosidase [Aspergi

B. Fatty acid as energy source (3)

1. Lipase-triacylglycerols to glycerol+FA (1)
<LIPASE 5 PRECURSOR>
a8e01ne.f1 284 1.4e-23 71 643 sp|P32949|LIP5_CANRU LIPASE 5 PRECURSOR >pir||JN0553 triacylglycerol lipase(EC 3.1.1.3) 5 precursor - yeast (Ca

2. Beta-oxydation of fatty acids (1)
<carnitine racemase>
d4f06ne.f1 166 9.9e-12 6 266 gnl|PID|e327452 (Z97336) carnitine racemase homolog [Arabidopsis thaliana]

3. Ketone body metabolism (1)
<ACETYL-COA HYDROLASE>
f9b06ne.r1 167 6.8e-11 451 549 sp|P15937|ACU8_NEUCR ACETYL-COA HYDROLASE (ACETYL-COA DEACYLASE) (ACETYL-COAACYLASE) (ACETATE UTILIZATION PROTE

C. Metabolism of other energy sources (13)

<acetate kinase>
NE.Contig527 237 7.7e-19 256 474 gi|2688543 (AE001163) acetate kinase (ackA) [Borrelia burgdorferi]
NE.Contig830 126 1.2e-06 262 444 gnl|PID|d1005027 (D17576) acetate kinase [Escherichia coli]

<GLUTATHIONE-DEPENDENT FORMALDEHYDE DEHYDROGENASE>
NE.Contig237 672 1.9e-65 33 479 sp|Q06099|FADH_CANMA GLUTATHIONE-DEPENDENT FORMALDEHYDE DEHYDROGENASE (FDH)(FALDH) >pir||JN0447 FDH1 protein -
e5d11ne.r1 571 1.2e-54 74 469 sp|Q06099|FADH_CANMA GLUTATHIONE-DEPENDENT FORMALDEHYDE DEHYDROGENASE (FDH)(FALDH) >pir||JN0447 FDH1 protein -

<GLYCEROL KINASE>

b9e11ne.f1	321	9.9e-28	1	330	sp|Q64516|GLPK_MOUSE GLYCEROL KINASE (ATP:GLYCEROL 3-PHOSPHOTRANSFERASE)(GLYCEROKINASE) (GK) >gi|1480469 (U4840
<PRPD PROTEIN>					
NE.Contig959	676	6.5e-66	2	679	sp|P77243|PRPD_ECOLI PRPD PROTEIN >gi|1657530 (U73857) similar to yqiP of B.subtilis [Escherichia coli] >gi|178
d6b05ne.f1	486	8.3e-46	11	457	sp|P77243|PRPD_ECOLI PRPD PROTEIN >gi|1657530 (U73857) similar to yqiP of B.subtilis [Escherichia coli] >gi|178
<ACETAMIDASE>					
NE.Contig643	218	2.1e-16	153	551	pir||JS0633 amidase (EC 3.5.1.4) - Aspergillus oryzae >gnl|PID|d1001845(D10492) acetamidase [Aspergill
<ALDEHYDE DEHYDROGENASE>					
NE.Contig1039	817	8.1e-81	242	823	sp|P40108|DHAL_CLAHE ALDEHYDE DEHYDROGENASE (ALDDH) (ALLERGEN CLA H 3) (CLA HIII) >pir||S43114 aldehyde dehydro
NE.Contig440	327	1.1e-28	152	469	sp|P42041|DHAL_ALTAL ALDEHYDE DEHYDROGENASE (ALDDH) (ALLERGEN ALT A 10) (ALT AX) >pir||S43108 aldehyde dehydrog
d5c02ne.f1	326	1.7e-28	89	481	sp|P42041|DHAL_ALTAL ALDEHYDE DEHYDROGENASE (ALDDH) (ALLERGEN ALT A 10) (ALT AX) >pir||S43108 aldehyde dehydrog
d5c02ne.r1	305	5.1e-26	164	457	sp|P47771|DHA5_YEAST ALDEHYDE DEHYDROGENASE (NAD(P)+) 2 >pir||S54615 aldehydedehydrogenase (NAD(P)+) (EC 1.2.1.
g9c12ne.r1	174	1.1e-11	336	464	sp|P40108|DHAL_CLAHE ALDEHYDE DEHYDROGENASE (ALDDH) (ALLERGEN CLA H 3) (CLA HIII) >pir||S43114 aldehyde dehydro

D. Electron transport (51)

1. Complex I-NADH-ubiquinone (3)

<NADH-UBIQUINONE DEHYDROGENASE>					
NE.Contig566	430	9.8e-40	271	519	sp|P40915|NUHM_NEUCR NADH-UBIQUINONE DEHYDROGENASE 24 KD SUBUNIT PRECURSOR>gi|577595 (X78083) NUO-24 gene produ
<mitochondrialcomplex I>					
NE.Contig578	555	5.2e-53	80	436	gnl|PID|e1227831 (AJ001520) 19.3kD iron-sulfur subunit of mitochondrialcomplex I [Neurospora crassa]
NE.Contig687	306	1.3e-26	326	538	gnl|PID|e1227831 (AJ001520) 19.3kD iron-sulfur subunit of mitochondrialcomplex I [Neurospora crassa]

2. Complex II-Succinate-ubiquinone (1)

<succinate dehydrogenase>					
NE.Contig589	593	5.2e-57	2	361	sp|O42772|DHSB_MYCGR SUCCINATE DEHYDROGENASE (UBIQUINONE) IRON-SULFUR PROTEINPRECURSOR (IP) >gi|2801670 (AF0420

3. Complex III-Ubiquinone to cytochrome C (20)

<CYTOCHROME B5>					
a6c07ne.f1	183	1.1e-13	152	460	sp|P49096|CYB5_MUSDO CYTOCHROME B5 (CYTB5) >gi|600524 (L38464) cytochrome b5[Musca domestica]
<cytochrome c>					
NE.Contig903	583	5.7e-56	263	586	sp|P00048|CYC_NEUCR CYTOCHROME C >pir||CCNC cytochrome c - Neurospora crassa>gi|3010 (X05506) cytochrome c [Ne

<CYTOCHROME C OXIDASE POLYPEPTIDE VI PRECURSOR>
NE.Contig1023 270 8.2e-23 48 473 sp|P00427|COX6_YEAST CYTOCHROME C OXIDASE POLYPEPTIDE VI PRECURSOR >pir||OTBY6cytochrome-c oxidase (EC 1.9.3.1)

<ACYL CARRIER PROTEIN, MITOCHONDRIAL PRECURSOR>
NE.Contig849 489 4.8e-46 152 448 sp|P11943|ACPM_NEUCR ACYL CARRIER PROTEIN, MITOCHONDRIAL PRECURSOR (ACP)(NADH-UBIQUINONE OXIDOREDUCTASE 9.6 KD
NE.Contig649 106 2.2e-05 425 487 sp|P11943|ACPM_NEUCR ACYL CARRIER PROTEIN, MITOCHONDRIAL PRECURSOR (ACP)(NADH-UBIQUINONE OXIDOREDUCTASE 9.6 KD

<NADH-UBIQUINONE OXIDOREDUCTASE B22 SUBUNIT>
NE.Contig132 129 1.2e-06 337 528 sp|Q02369|NI2M_BOVIN NADH-UBIQUINONE OXIDOREDUCTASE B22 SUBUNIT (COMPLEXI-B22) (CI-B22) >pir||S28256 NADH dehyd

<ubiquinol-cytochrome c reductase complex subunit>
NE.Contig594 321 3e-28 584 949 gnl|PID|e1319808 (AL031546) ubiquinol-cytochrome c reductase complex subunit[Schizosaccharomyces pombe]

<UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX UBIQUINONE-BINDING PROTEIN QP-C>
NE.Contig264 433 4.9e-40 275 571 sp|P48503|UCRQ_NEUCR UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX UBIQUINONE-BINDING PROTEIN QP-C (UBIQUINOL-CYTOCHR

<nadh-cytochrome b5 reductase>
NE.Contig764 285 2.1e-24 78 476 gnl|PID|e1319411 (AL031530) putative nadh-cytochrome b5 reductase[Schizosaccharomyces pombe]
NE.Contig639 198 3.6e-15 324 503 gnl|PID|e1319411 (AL031530) putative nadh-cytochrome b5 reductase[Schizosaccharomyces pombe]

<NADH-CYTOCHROME B5 REDUCTASE PRECURSOR>
a1c12ne.r1 287 1.4e-24 225 563 sp|P36060|MCR1_YEAST NADH-CYTOCHROME B5 REDUCTASE PRECURSOR (P34/P32)>pir||S37800 cytochrome-b5 reductase (EC 1

<BETA-MPP>
NE.Contig283 752 6.8e-74 12 452 sp|P11913|MPP2_NEUCR MITOCHONDRIAL PROCESSING PEPTIDASE BETA SUBUNIT PRECURSOR(BETA-MPP) (UBIQUINOL-CYTOCHROME
d6d07ne.r1 485 1.3e-45 138 437 sp|P11913|MPP2_NEUCR MITOCHONDRIAL PROCESSING PEPTIDASE BETA SUBUNIT PRECURSOR(BETA-MPP) (UBIQUINOL-CYTOCHROME

<NADH-UBIQUINONE OXIDOREDUCTASE 40 KD SUBUNIT PRECURSOR>
e2d09ne.f1 685 1e-66 51 455 sp|P25284|NUEM_NEUCR NADH-UBIQUINONE OXIDOREDUCTASE 40 KD SUBUNIT PRECURSOR(COMPLEX I-40KD) (CI-40KD) >pir||S13

<NADH-UBIQUINONE OXIDOREDUCTASE 21 KD SUBUNIT>
f9e02ne.r1 389 2.3e-35 278 502 sp|Q02854|NUXM_NEUCR NADH-UBIQUINONE OXIDOREDUCTASE 21 KD SUBUNIT (COMPLEXI-21KD) (CI-21KD) >pir||S27171 NADH d

<Cytochrome C oxidase subunit>
NE.Contig874 214 6.8e-17 4 255 gnl|PID|e1319667 (AL031540) putative Cytochrome C oxidase subunit via[Schizosaccharomyces pombe]

<CYTOCHROME C OXIDASE POLYPEPTIDE VIIA>
NE.Contig986 130 5.6e-08 94 252 sp|P07255|COX9_YEAST CYTOCHROME C OXIDASE POLYPEPTIDE VIIA >pir||OBBY7Acytochrome-c oxidase (EC 1.9.3.1) chain

<CYTOCHROME C OXIDASE POLYPEPTIDE V PRECURSOR>
b9b12ne.f1 721 1.3e-70 64 564 sp|P06810|COX5_NEUCR CYTOCHROME C OXIDASE POLYPEPTIDE V PRECURSOR >pir||OTNCVcytochrome-c oxidase (EC 1.9.3.1)

<CYTOCHROME C OXIDASE POLYPEPTIDE VIB>
g9e10ne.f1 281 6.1e-24 73 312 sp|Q01519|COXG_YEAST CYTOCHROME C OXIDASE POLYPEPTIDE VIB (AED) >pir||S31256cytochrome-c oxidase (EC 1.9.3.1) c

<Mitochondrial Cytochrome Bc1 Complex>
NE.Contig564 117 1.6e-06 70 222 pdb|1QCR|J Chain J, Crystal Structure Of Bovine Mitochondrial Cytochrome Bc1Complex, Alpha Carbon Ato

4. Electron carriers (2)
<flavoprotein>
g7a06ne.f1 352 1.7e-31 6 452 gnl|PID|e349663 (Z99292) flavoprotein [Schizosaccharomyces pombe]

<QUINONE OXIDOREDUCTASE>
g9c08ne.f1 341 2.3e-30 18 434 sp|P38230|QOR_YEAST PROBABLE QUINONE OXIDOREDUCTASE (NADPH:QUINONE REDUCTASE)>pir||S45904 quinone oxidoreducta

5. Component enzymes and molecules (2)
<MITOCHONDRIAL CARRIER PROTEIN>
NE.Contig669 201 2e-15 325 480 sp|P38988|SHM1_YEAST PUTATIVE MITOCHONDRIAL CARRIER PROTEIN YHM1/SHM1>pir||S58779 probable carrier protein SHM1
NE.Contig825 148 2.3e-09 405 503 gnl|PID|e1319387 (AL031525) mitochondrial carrier protein [Schizosaccharomycespombe]

6. ATP synthase (21)
<ATP SYNTHASE DELTA CHAIN, MITOCHONDRIAL PRECURSOR>
NE.Contig699 659 4.7e-64 119 535 sp|P56525|ATPD_NEUCR ATP SYNTHASE DELTA CHAIN, MITOCHONDRIAL PRECURSOR

<ATP SYNTHASE OLIGOMYCIN SENSITIVITY CONFERRAL PROTEIN>
NE.Contig660 153 2e-10 320 487 sp|P09457|ATPO_YEAST ATP SYNTHASE OLIGOMYCIN SENSITIVITY CONFERRAL PROTEINPRECURSOR, MITOCHONDRIAL (OSCP) (ATP

<PLASMA MEMBRANE ATPASE (PROTON PUMP)>
NE.Contig1094 803 2.8e-79 399 926 gi|2197050 (AF001033) putative 20kDa subunit of the V-ATPase [Neurosporacrassa]

<oligomycin sensitivity conferring protein>
NE.Contig445 277 1.3e-23 81 434 gi|3273480 (AF008185) oligomycin sensitivity conferring protein [Kluyveromyceslactis]

<ATP SYNTHASE PROTEIN 9, MITOCHONDRIAL PRECURSOR>
NE.Contig1053 613 3.8e-59 57 497 sp|P00842|ATP9_NEUCR ATP SYNTHASE PROTEIN 9, MITOCHONDRIAL PRECURSOR(LIPID-BINDING PROTEIN) >pir||LWNCA H+-tran

<14-3-3 PROTEIN HOMOLOG>
NE.Contig1026 775 2.5e-76 144 752 sp|Q99002|1433_TRIHA 14-3-3 PROTEIN HOMOLOG (TH1433) >gi|806859 (U24158)14.3.3. protein [Trichoderma harzianum]

<ATP SYNTHASE ALPHA CHAIN, MITOCHONDRIAL PRECURSOR>
NE.Contig1093a 1053 1.2e-105 84 725 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]

NE.Contig1093c	1050	2.8e-105	3	668	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]
h6g01ne.f1	596	2.2e-57	64	432	sp|P37211|ATPA_NEUCR ATP SYNTHASE ALPHA CHAIN, MITOCHONDRIAL PRECURSOR>pir||JC1111 H+-transporting ATP synthase
<ATP SYNTHASE E CHAIN, MITOCHONDRIAL>					
NE.Contig1096	133	2.9e-08	243	473	sp|P81449|ATPJ_YEAST ATP SYNTHASE E CHAIN, MITOCHONDRIAL
<VACUOLAR ATP SYNTHASE SUBUNIT G>					
NE.Contig128	159	1.3e-19	340	522	sp|P78713|VATG_NEUCR VACUOLAR ATP SYNTHASE SUBUNIT G (V-ATPASE 13 KD SUBUNIT)(VACUOLAR H(+)-ATPASE SUBUNIT G) >
<PROTEOLIPID PROTEIN PPA1>					
NE.Contig208	332	2.4e-29	198	449	sp|P23968|PPA1_YEAST PROTEOLIPID PROTEIN PPA1 >pir||A34633 probableH+-transporting ATPase (EC 3.6.1.35) lipid-b
<VACUOLAR ATP SYNTHASE 16 KD PROTEOLIPID SUBUNIT>					
NE.Contig233	514	1.3e-48	85	462	sp|P31413|VATL_NEUCR VACUOLAR ATP SYNTHASE 16 KD PROTEOLIPID SUBUNIT>pir||S43893 H+-transporting ATPase (EC 3.6
<ATP SYNTHASE D CHAIN, MITOCHONDRIAL>					
NE.Contig288	340	2.7e-30	2	334	sp|O13350|ATP7_KLULA ATP SYNTHASE D CHAIN, MITOCHONDRIAL >gi|2425073(AF019223) F1Fo-ATP synthase subunit 7 [Klu
<VACUOLAR ATP SYNTHASE 98 KD SUBUNIT>					
h6g06ne.f1	577	5.2e-55	2	448	sp|Q01290|VPH1_NEUCR VACUOLAR ATP SYNTHASE 98 KD SUBUNIT (VACUOLAR ATPASE 98KD SUBUNIT) >gi|1237128 (U36396) va
<ATP SYNTHASE J CHAIN, MITOCHONDRIAL>					
NE.Contig429	140	2.3e-08	757	945	sp|O13931|AT18_SCHPO PUTATIVE ATP SYNTHASE J CHAIN, MITOCHONDRIAL>gnl|PID|e353257 (Z99753) hypothetical protein
<VACUOLAR ATP SYNTHASE SUBUNIT AC39>					
g2d08ne.f1	819	5.3e-81	1	471	sp|P53659|VATX_NEUCR VACUOLAR ATP SYNTHASE SUBUNIT AC39 (V-ATPASE AC39SUBUNIT) (V-ATPASE 41 KD SUBUNIT) >gi|103
NE.Contig637	190	8.8e-14	385	498	sp|P53659|VATX_NEUCR VACUOLAR ATP SYNTHASE SUBUNIT AC39 (V-ATPASE AC39SUBUNIT) (V-ATPASE 41 KD SUBUNIT) >gi|103
<ATP SYNTHASE SUBUNIT 4, MITOCHONDRIAL PRECURSOR>					
NE.Contig886	492	2.2e-46	76	756	sp|O13349|ATPF_KLULA ATP SYNTHASE SUBUNIT 4, MITOCHONDRIAL PRECURSOR>gi|2425071 (AF019222) F1Fo-ATP synthase su
NE.Contig635	179	4e-13	347	508	sp|O13349|ATPF_KLULA ATP SYNTHASE SUBUNIT 4, MITOCHONDRIAL PRECURSOR>gi|2425071 (AF019222) F1Fo-ATP synthase su
<ATP SYNTHASE BETA CHAIN, MITOCHONDRIAL PRECURSOR>					
NE.Contig891	987	8.6e-99	217	813	sp|P23704|ATPB_NEUCR ATP SYNTHASE BETA CHAIN, MITOCHONDRIAL PRECURSOR>pir||JC1112 H+-transporting ATP synthase
7. Alternative respiratory path (2)					
<ALTERNATIVE OXIDASE PRECURSOR>					
NE.Contig211	778	1.5e-76	29	463	sp|Q01355|AOX_NEUCR ALTERNATIVE OXIDASE PRECURSOR (ALTOX) >pir||S65752alternative oxidase precursor - Neurospo
NE.Contig320	528	4.1e-50	181	471	sp|Q01355|AOX_NEUCR ALTERNATIVE OXIDASE PRECURSOR (ALTOX >pir||S65752alternative

oxidase precursor - Neurospo

E. Reducing carriers (1)

<glutaredoxin>

NE.Contig894	281	5.7e-24	311	625	sp|P55143|GLRX_RICCO GLUTAREDOXIN >pir||S54825 glutaredoxin - castor bean

II: Gene expression and genetic information processing (248)

A. DNA synthesis (16)

1. DNA replication (4)

<PROLIFERATING CELL NUCLEAR ANTIGEN>

h3c03ne.fl	457	1.1e-42	131	475	sp|Q03392|PCNA_SCHPO PROLIFERATING CELL NUCLEAR ANTIGEN (PCNA) >pir||WMZPETproliferating cell nuclear antigen -

<DNA replication licensing factor>

NE.Contig250	674	1.1e-65	1	465	gi|2735931 (AF014813) DNA replication licensing factor [Emericella nidulans]
NE.Contig253	272	8.6e-22	304	486	gi|2735931 (AF014813) DNA replication licensing factor [Emericella nidulans]

<minichromosome maintenance protein Mcm7p>

NE.Contig252	476	4.7e-44	257	661	gi|3236468 (AF070481) minichromosome maintenance protein Mcm7p[Schizosaccharomyces pombe] >gnl|PID|e1

2. DNA modification and DNA repair (1)

<Hmp1>

NE.Contig620	134	2.7e-08	299	571	gi|1176420 (U39049) Hmp1 [Ustilago maydis]

3 DNA packaging (11)

3.1 Histone (7)

<HISTONE H4>

NE.Contig950	409	1.5e-37	118	363	sp|P04914|H4_NEUCR HISTONE H4 >pir||S07913 histone H4 - Neurospora crassa>gi|3018 (X01611) histone H4 [Neuros
NE.Contig1090	409	1.7e-37	165	410	sp|P23750|H41_EMENI HISTONE H4.1 >pir||S11939 histone H4.1 - Emericellanidulans >gi|296341 (X55549) H4.1 [Emer

<HISTONE H3>

NE.Contig1049	666	1e-64	100	507	sp|P07041|H3_NEUCR HISTONE H3 >pir||S07350 histone H3 - Neurospora crassa>gi|3016 (X01612) histone H3 [Neuros
NE.Contig324	140	5.7e-09	383	466	sp|P07041|H3_NEUCR HISTONE H3 >pir||S07350 histone H3 - Neurospora crassa>gi|3016 (X01612) histone H3 [Neuros

<HISTONE H2B>

NE.Contig642	507	6.4e-48	302	637	sp|P23754|H2B_EMENI HISTONE H2B >pir||S11937 histone H2B - Emericella nidulans>gi|296335 (X55547) H2B [Emerice

<histone H2A>
NE.Contig1055 467 1.2e-43 146 499 gnl|PID|e1181728 (Y15320) histone H2A [Aspergillus niger]

<HISTONE H1>
NE.Contig164 137 1.7e-08 223 423 sp|P53551|H1_YEAST HISTONE H1 >pir||S69056 histone H1 - yeast (Saccharomycescerevisiae) >gi|1244786 (U43703)

3.2 DNA-binding (4)
<CURVED DNA-BINDING PROTEIN>
NE.Contig697 231 1.9e-18 3 374 sp|Q09184|CDB4_SCHPO CURVED DNA-BINDING PROTEIN (42 KD PROTEIN) >pir||S4658342K protein - fission yeast (Schizo
NE.Contig737 226 8.2e-18 189 575 sp|Q09184|CDB4_SCHPO CURVED DNA-BINDING PROTEIN (42 KD PROTEIN) >pir||S4658342K protein - fission yeast (Schizo

<CELLULAR NUCLEIC ACID BINDING PROTEIN>
a5h06ne.f1 334 1.1e-29 133 594 sp|P36627|BYR3_SCHPO CELLULAR NUCLEIC ACID BINDING PROTEIN HOMOLOG >bbs|112403(S45038) cellular nucleic acid bi

<SAP1 PROTEIN>
c6f09ne.f1 369 3.6e-32 200 502 sp|P39955|SAP1_YEAST SAP1 PROTEIN >pir||S50550 hypothetical protein YER047c -yeast (Saccharomyces cerevisiae) >

B. Gene expression (232)

1. Transcription (34)

1.1 RNA polymerase (2)
<DNA-DIRECTED RNA POLYMERASE II 13.3 KD POLYPEPTIDE>
NE.Contig924 214 3.1e-16 646 939 sp|O08740|RPBY_MOUSE DNA-DIRECTED RNA POLYMERASE II 13.3 KD POLYPEPTIDE(RPB11) (RPB14) >gnl|PID|d1020707 (D8581

<auxin-induced protein>
b1h03ne.r1 211 7e-15 131 562 gi|2462762 (AC002292) Highly similar to auxin-induced protein (aldo/ketoreductase family) [Arabidopsi

1.2 Regulation (10)
<CROSS-PATHWAY CONTROL PROTEIN 1>
NE.Contig1044 399 5.3e-60 269 613 sp|P11115|CPC1_NEUCR CROSS-PATHWAY CONTROL PROTEIN 1 >gi|168793 (J03262)cross-pathway control protein 1 [Neuros

<TRANSCRIPTIONAL REPRESSOR RCO-1>
b5c06ne.r1 564 7.1e-54 63 389 sp|P78706|RCO1_NEUCR TRANSCRIPTIONAL REPRESSOR RCO-1 >gi|1698504 (U57061)rco-1 gene product [Neurospora crassa]

<TRANSCRIPTIONAL ACTIVATOR PROTEIN ACU-15>
NE.Contig135 353 1.8e-30 66 299 sp|P87000|AC15_NEUCR TRANSCRIPTIONAL ACTIVATOR PROTEIN ACU-15 >gnl|PID|e308394(Y11565) transcriptional activato

<TRANSCRIPTIONAL ADAPTOR>
NE.Contig356 350 3.2e-31 269 622 sp|Q02336¦ADA2_YEAST POTENTIAL TRANSCRIPTIONAL ADAPTOR >pir||A43252 probabletranscriptional adaptor ADA2 - yeas

<MBF1>
NE.Contig460 139 6.8e-09 250 441 gnl|PID|d1034183 (AB017593) MBF1 [Saccharomyces cerevisiae]

<CCAAT/enhancer-binding protein>					
NE.Contig689	120	4.7e-06	224	439	gi|1947129 (AF000262) similar to CCAAT/enhancer-binding protein[Caenorhabditis elegans]
<alpha NAC/1.9.2. protein>					
NE.Contig961	234	1.2e-31	143	331	gi|1142653 (U22151) alpha NAC/1.9.2. protein [Mus musculus] >gi|1666690(U48363) alpha-NAC, non-muscle
<Cad1 protein>					
d8a09ne.f1	200	1.2e-14	213	497	dbj||AB007905_1 (AB007905) Cad1 protein [Schizosaccharomyces pombe]
<TRANSCRIPTION ELONGATION FACTOR S-II>					
a9c03ne.r1	188	3.2e-12	374	550	sp|P49373|TFS2_SCHPO TRANSCRIPTION ELONGATION FACTOR S-II (TFIIS) >pir||S63845transcription elongation factor T
a9c03ne.f1	151	1.9e-08	196	420	sp|P49373|TFS2_SCHPO TRANSCRIPTION ELONGATION FACTOR S-II (TFIIS) >pir||S63845transcription elongation factor T

1.3 RNA processing (13)

a. SPLICEOSOME (8)

<spliceosomal protein>					
c4h01ne.f1	258	1.7e-21	177	446	gnl|PID|d1022283 (AB004538) spliceosomal protein [Schizosaccharomyces pombe]
<splicing factor>					
NE.Contig785	143	9.9e-09	413	619	gnl|PID|e1314261 (AL031179) splicing factor [Schizosaccharomyces pombe]
<U5 snRNP-specific 40 kDa protein>					
NE.Contig956	295	2.1e-25	128	469	gi|3820594 (AF090988) U5 snRNP-specific 40 kDa protein [Homo sapiens]
<snRNP core Sm protein>					
a7f11ne.r1	157	8.9e-11	324	455	gi|2463648 (U85207) snRNP core Sm protein homolog Sm-X5 [Mus musculus]
<small nuclear ribonucleoprotein>					
NE.Contig118	411	1.1e-37	453	776	gnl|PID|e349593 (Z99259) small nuclear ribonucleoprotein [Schizosaccharomycespombe]
<Sm protein F>					
NE.Contig789	244	5.3e-20	261	512	pir||S55053 Sm protein F - human >gi|806564 (X85372) Sm protein F [Homosapiens]
<MSS51 PROTEIN>					
a3a10ne.r1	301	4.8e-26	158	577	sp|P32335|MS51_YEAST MSS51 PROTEIN >gi|172010 (J01487) MSS51 [Saccharomycescerevisiae] >bbs|113259 (S43721) MSS
a3a10ne.r1	301	4.8e-26	158	577	pir||S42160 MSS51 protein - yeast (Saccharomyces cerevisiae) >gi|544513(U14913) Mss51p: Probable initi

b. Other (5)

<NUCLEOLAR PROTEIN NOP5>					
NE.Contig290	568	2.1e-54	3	464	sp|Q12499|NOP5_YEAST NUCLEOLAR PROTEIN NOP5 >pir||S58322 hypothetical proteinYOR310c - yeast (Saccharomyces cer
<RNA binding protein>					

b3d01ne.f1	203	1.1e-15	231	470	dbj||D28862_1 (D28862) RNA binding protein [Nicotiana sylvestris]
NE.Contig301	175	9.9e-13	302	472	gnl|PID|e1291624 (AL023287) RNA binding protein [Schizosaccharomyces pombe]
<RNA binding domain>					
c4b09ne.f1	284	3.1e-24	221	481	gnl|PID|e1348515 (Z32683) similar to RNA binding domain; cDNA EST EMBL:D74891comes from this gene; cDNA EST
<RNA12 PROTEIN>					
NE.Contig596	321	3.8e-27	240	1118	sp|P32843|RN12_YEAST RNA12 PROTEIN >pir||S20462 RNA12 protein - yeast(Saccharomyces cerevisiae) >bbs|92206 (S92

1.4 tRNA synthesis and modifications (8)

<ISOLEUCYL-TRNA SYNTHETASE, CYTOPLASMIC>					
NE.Contig754	145	3.7e-08	342	506	sp|O13651|SYIC_SCHPO ISOLEUCYL-TRNA SYNTHETASE, CYTOPLASMIC (ISOLEUCINE--TRNALIGASE) (ILERS) >gnl|PID|d1022285
<lysyl tRNA synthetase>					
NE.Contig914	294	1.2e-24	1	492	sp|Q15046|SYK_HUMAN LYSYL-TRNA SYNTHETASE (LYSINE--TRNA LIGASE)
NE.Contig880	154	1.6e-09	289	429	gnl|PID|e1332819 (AL031907) lysyl-trna synthetase [Schizosaccharomyces
<ALANYL-TRNA SYNTHETASE, CYTOPLASMIC>					
a2b07ne.f1	356	1.1e-30	64	585	sp|P40825|SYAC_YEAST ALANYL-TRNA SYNTHETASE, CYTOPLASMIC (ALANINE--TRNALIGASE) (ALARS) >pir||S62065 alanine--tR
a3b07ne.r1	273	7.3e-22	256	687	sp|P40825|SYAC_YEAST ALANYL-TRNA SYNTHETASE, CYTOPLASMIC (ALANINE--TRNALIGASE) (ALARS) >pir||S62065 alanine--tR
<VALYL-TRNA SYNTHETASE, MITOCHONDRIAL PRECURSOR>					
e2c07ne.f1	749	3.2e-73	1	462	sp|P28350|SYV_NEUCR VALYL-TRNA SYNTHETASE, MITOCHONDRIAL PRECURSOR(VALINE--TRNA LIGASE) (VALRS) >pir||A41251 v
<THREONYL-TRNA SYNTHETASE, MITOCHONDRIALPRECURSOR>					
g9b12ne.f1	239	8.7e-19	216	428	sp|O13969|SYTM_SCHPO PROBABLE THREONYL-TRNA SYNTHETASE, MITOCHONDRIALPRECURSOR (THREONINE--TRNA LIGASE) (THRRS)
<PSEUDOURIDYLATE SYNTHASE 1>					
NE.Contig322	227	2.1e-17	229	495	sp|Q12211|PUS1_YEAST PSEUDOURIDYLATE SYNTHASE 1 (PSEUDOURIDINE SYNTHASE 1)>pir||S65231 tRNA-pseudouridine synth

1.5 RNA replication (1)

<DKA1 PROTEIN>					
NE.Contig1036	121	0.00017	649	960	sp|P14306|DKA1_YEAST DKA1 PROTEIN (NSP1 PROTEIN) (TFS1 PROTEIN) >pir||S18843DKA1 protein - yeast (Saccharomyces

2.Protein biosynthesis (198)

2.1 Translation (120)

<PROTEIN TRANSLATION FACTOR SUI1>					
NE.Contig547	286	1.9e-24	467	730	sp|P32911|SUI1_YEAST PROTEIN TRANSLATION FACTOR SUI1 >pir||S31245 translationinitiation factor eIF-2A - yeast (
<TRANSLOCATION PROTEIN SEC66>					
a7a05ne.f1	178	5.5e-13	132	602	sp|P33754|SC66_YEAST TRANSLOCATION PROTEIN SEC66 (HSS1 PROTEIN)

					>pir||A47735endoplasmic reticulum translocation
a. Initiation (6)					
<EUKARYOTIC INITIATION FACTOR 4A>					
NE.Contig206	313	2.4e-27	229	462	sp|P47943|IF4A_SCHPO EUKARYOTIC INITIATION FACTOR 4A (EIF-4A) >gnl|PID|e114182(X80796) translation initiation f
NE.Contig827	227	8.4e-18	2	187	sp|P47943|IF4A_SCHPO EUKARYOTIC INITIATION FACTOR 4A (EIF-4A) >gnl|PID|e114182(X80796) translation initiation f
<INITIATION FACTOR>					
NE.Contig976	367	4.3e-33	269	490	sp|P38672|IF5A_NEUCR INITIATION FACTOR 5A (EIF-5A) (EIF-4D) >pir||S55278translation initiation factor eIF-5A -
<INITIATION FACTOR 5A>					
NE.Contig1070	840	3.4e-83	126	614	sp|P38672|IF5A_NEUCR INITIATION FACTOR 5A (EIF-5A) (EIF-4D) >pir||S55278translation initiation factor eIF-5A -
<translation initiation factor 4e>					
NE.Contig51	232	1e-18	150	677	gi|3329384 (AF038957) translation initiation factor 4e [Homo sapiens]
<CPC3 protein>					
d9h02ne.r1	499	1.5e-45	148	429	gnl|PID|e299527 (X91867) CPC3 protein [Neurospora crassa]
b. Elongation (9)					
<elongation factor 1 beta>					
NE.Contig99	336	8.2e-30	77	457	gnl|PID|d1012239 (D82574) elongation factor 1 beta [Schizosaccharomyces pombe]
<ELONGATION FACTOR 1-ALPHA>					
NE.Contig1062	1136	1.5e-114	279	944	sp|Q01372|EF1A_NEUCR ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) >gnl|PID|d1008869(D45837) elongation factor 1-alpha
NE.Contig12	711	1.7e-69	589	996	sp|Q01520|EF1A_PODAN ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA) >pir||S43861translation elongation factor eEF-1 alp
<elongation factor 2>					
NE.Contig856	700	2.3e-68	357	857	sp|P32324|EF2_YEAST ELONGATION FACTOR 2 (EF-2) >pir||A41778 translationelongation factor eEF-2 - yeast (Saccha
e3a03ne.f1	621	4.8e-60	30	485	sp|P32324|EF2_YEAST ELONGATION FACTOR 2 (EF-2) >pir||A41778 translationelongation factor eEF-2 - yeast (Saccha
i1a01ne.f1	608	1.2e-58	6	446	sp|P32324|EF2_YEAST ELONGATION FACTOR 2 (EF-2) >pir||A41778 translationelongation factor eEF-2 - yeast (Saccha
NE.Contig946	323	2.5e-27	233	502	sp|P32324|EF2_YEAST ELONGATION FACTOR 2 (EF-2) >pir||A41778 translationelongation factor eEF-2 - yeast (Saccha
<ELONGATION FACTOR 3>					
c2c02ne.r1	148	1.4e-06	412	552	sp|P25997|EF3_CANAL ELONGATION FACTOR 3 (EF-3l) >gi|2498 (Z11484) elongationfactor 3 [Candida albicans]
<ELONGATION FACTOR 1-GAMMA 2>					
c9e10ne.f1	473	2.4e-44	2	442	sp|P36008|EF1H_YEAST ELONGATION FACTOR 1-GAMMA 2 (EF-1-GAMMA 2) >pir||S37906translation elongation factor eEF-1
c. Termination (1)					
<translation release factor subunit 1>					
a4b11ne.f1	939	1.2e-93	87	638	gi|2996008 (AF053983) translation release factor subunit 1 [Podosporaanserina]

d. Ribosomal proteins (102)

<MITOCHONDRIAL RIBOSOMAL PROTEIN S5>					
NE.Contig468	642	3e-62	26	424	sp|P23351|RMS5_NEUCR MITOCHONDRIAL RIBOSOMAL PROTEIN S5 >pir||A19079 23S rRNAintron protein - Neurospora crassa
<acidic ribosomal protein P0.e>					
NE.Contig1066	516	7.4e-49	80	532	pir||R5BY0E acidic ribosomal protein P0.e, cytosolic - yeast (Saccharomycescerevisiae) >gi|171806 (M37
NE.Contig1041	176	1.5e-12	294	488	pir||R5BY0E acidic ribosomal protein P0.e, cytosolic - yeast (Saccharomycescerevisiae) >gi|171806 (M37
NE.Contig1078b	396	5.8e-36	217	495	gb|AAC15802.1| (AF061259) ribosomal protein rpS12 [Blumeria graminis f. sp.hordei]
<ribosomal protein S14.e>					
NE.Contig1005	597	2e-57	47	463	pir||S11667 ribosomal protein S14.e - Neurospora crassa
<ribosomal protein L21>					
NE.Contig1035	544	7e-52	187	660	gnl|PID|d1025723 (AB010901) ribosomal protein L21 homolog [Schizosaccharomycespombe]
<ribosomal protein CRP7>					
NE.Contig1068	449	9.3e-42	52	312	gnl|PID|d1036047 (AB015207) ribosomal protein CRP7 [Neurospora crassa]
<ribosomal protein L31.e.B, cytosolic>					
NE.Contig423	339	4.2e-30	169	471	pir||S55962 ribosomal protein L31.e.B, cytosolic - yeast (Saccharomycescerevisiae) >gi|625117 (U19729)
NE.Contig684	236	3.5e-19	275	499	pir||S55962 ribosomal protein L31.e.B, cytosolic - yeast (Saccharomycescerevisiae) >gi|625117 (U19729)
NE.Contig289	166	9.5e-12	300	449	pir||S55962 ribosomal protein L31.e.B, cytosolic - yeast (Saccharomycescerevisiae) >gi|625117 (U19729)
<ribosomal protein S7>					
NE.Contig1002	795	2.1e-78	27	512	sp|P52810|RS9_PODAN 40S RIBOSOMAL PROTEIN S9 (S7) >gnl|PID|e242707 (X96613)cytoplasmic ribosomal protein S7 [P
NE.Contig554	742	6.7e-73	7	513	gi|2997729 (AF054511) ribosomal protein S7 [Yarrowia lipolytica]
<ribosomal protein L27>					
NE.Contig574	113	4.1e-06	398	478	gnl|PID|d1029789 (AB015354) ribosomal protein L27 homolog [Schizosaccharomycespombe]
<ribosomal protein L13E>					
NE.Contig949	414	4.9e-38	327	920	gi|2981202 (AF050672) ribosomal protein L13E [Candida albicans]>gnl|PID|e1340956 (AL033497) ribosomal
<5S rRNA binding ribosomal protein>					
NE.Contig982	760	8.3e-75	43	486	gi|3003044 (AF054907) putative 5S rRNA binding ribosomal protein [Neurosporacrassa]
NE.Contig1064	670	3.5e-65	196	642	gi|3003044 (AF054907) putative 5S rRNA binding ribosomal protein [Neurosporacrassa]
1.) *40S* ribosomal protein (34)					
<40S ribosomal protein>					
NE.Contig966	677	5.5e-66	248	649	sp|P27770|RS17_NEUCR 40S RIBOSOMAL PROTEIN S17 (CRP3) >pir||S34441 ribosomalprotein

				L17.e - Neurospora crassa >
a2c05ne.f1	607	1.5e-58	115 588	gnl|PID|e1292706 (AL023554) 40S ribosomal protein [Schizosaccharomyces pombe]
NE.Contig369	364	9.2e-33	256 495	gnl|PID|e1168591 (Z99168) 40s ribosomal protein [Schizosaccharomyces pombe]
NE.Contig755	329	5.4e-29	697 975	sp|Q08745|YO93_YEAST PUTATIVE 40S RIBOSOMAL PROTEIN IN SNF2-CPA1 INTERGENICREGION >pir||S67197 ribosomal protei
h8h10ne.f1	291	5.2e-25	163 417	sp|P53733|YN8L_YEAST PUTATIVE 40S RIBOSOMAL PROTEIN YNR037C >pir||S63368probable ribosomal protein S19, mitocho
NE.Contig902	276	2.1e-23	304 501	sp|P05753|RS4E_YEAST 40S RIBOSOMAL PROTEIN S4 (S7) (YS6) (RP5) >pir||S20054ribosomal protein S4.e, cytosolic -
NE.Contig509	258	1.8e-21	40 357	sp|P10663|RT02_YEAST MITOCHONDRIAL 40S RIBOSOMAL PROTEIN MRP2 >pir||R3BY14ribosomal protein S14 precursor, mito
<40S ribosomal protein S12>				
NE.Contig1078c	527	7.0e-50	153 524	gi|3114615 (AF052483) 40S ribosomal protein S12 [Erysiphe graminis f. sp.hordei]
<40S RIBOSOMAL PROTEIN S13>				
NE.Contig786	307	1.2e-26	398 601	sp|P33192|RS13_CANMA 40S RIBOSOMAL PROTEIN S13 (S15) >pir||S25374 ribosomalprotein S13.e - yeast (Candida malto
<40S RIBOSOMAL PROTEIN S15>				
NE.Contig944	757	1.9e-74	69 524	sp|P34737|RS15_PODAN 40S RIBOSOMAL PROTEIN S15 (S12) >pir||A53793 ribosomalprotein S12, cytosolic - Podospora a
NE.Contig278	109	1e-05	351 479	sp|P34737|RS15_PODAN 40S RIBOSOMAL PROTEIN S15 (S12) >pir||A53793 ribosomalprotein S12, cytosolic - Podospora a
40S RIBOSOMAL PROTEIN S26E>				
NE.Contig1031	479	6.9e-45	295 603	sp|P21772|RS26_NEUCR 40S RIBOSOMAL PROTEIN S26E (CRP5) (13.6 KD RIBOSOMALPROTEIN) >pir||R4NC26 ribosomal protei
<40s ribosomal protein s2>				
NE.Contig1052	828	5.7e-82	188 862	gnl|PID|e1326284 (AL031798) 40s ribosomal protein s2 [Schizosaccharomycespombe]
<40S RIBOSOMAL PROTEIN S30>				
a6b01ne.f1	226	4.6e-18	78 260	sp|Q12087|RS30_YEAST 40S RIBOSOMAL PROTEIN S30 >pir||S67074 ribosomal proteinS30.e, cytosolic - yeast (Saccharo
NE.Contig877	153	2.3e-10	495 674	sp|Q12087|RS30_YEAST 40S RIBOSOMAL PROTEIN S30 >pir||S67074 ribosomal proteinS30.e, cytosolic - yeast (Saccharo
NE.Contig942	153	2.4e-10	446 625	sp|Q12087|RS30_YEAST 40S RIBOSOMAL PROTEIN S30 >pir||S67074 ribosomal proteinS30.e, cytosolic - yeast (Saccharo
<40s ribosomal protein s27>				
NE.Contig1095	367	5.1e-33	100 345	gnl|PID|e1313483 (AL031154) 40s ribosomal protein s27 type[Schizosaccharomyces pombe]
<40S RIBOSOMAL PROTEIN S9>				
NE.Contig270	266	2.1e-22	190 384	sp|P52810|RS9_PODAN 40S RIBOSOMAL PROTEIN S9 (S7) >gnl|PID|e242707 (X96613)cytoplasmic ribosomal protein S7 [P
<40S RIBOSOMAL PROTEIN S28>				
NE.Contig709	608	1.3e-58	16 450	sp|P32827|RS28_YEAST 40S RIBOSOMAL PROTEIN S28 >pir||A46703 ribosomal proteinS23.e - yeast (Saccharomyces cerev
<40S RIBOSOMAL PROTEIN SA HOMOLOG>				
NE.Contig624	1052	1.3e-105	371 997	sp|Q01291|RSP4_NEUCR 40S RIBOSOMAL PROTEIN SA HOMOLOG (RIBOSOME-ASSOCIATEDPROTEIN 1)

```
                                                                >gi|1039443 (U36470) putat
NE.Contig1034      163  4.4e-11   531 695       sp|Q01291|RSP4_NEUCR 40S RIBOSOMAL PROTEIN SA HOMOLOG (RIBOSOME-ASSOCIATEDPROTEIN 1)
                                                                >gi|1039443 (U36470) putat

<40S ribosomal protein S5>
NE.Contig1043      791  5.2e-78   209 766       gnl|PID|e1286357 (AJ005346) 40S ribosomal protein S5 [Cicer arietinum]

<40S RIBOSOMAL PROTEIN S11>
NE.Contig154       653  2.1e-63    21 503       sp|P79013|RS11_SCHPO 40S RIBOSOMAL PROTEIN S11 >gnl|PID|e339937 (Z98979) 40sribosomal
                                                protein s11 [Schizosaccha

<40S RIBOSOMAL PROTEIN RP10>
NE.Contig773       698  3.4e-68   147 734       sp|P40910|RS3A_CANAL 40S RIBOSOMAL PROTEIN RP10 >pir||S49366 ribosomal proteinS0.e.B,
                                                cytosolic - yeast (Candid
NE.Contig172       280  7.3e-24   189 407       sp|P40910|RS3A_CANAL 40S RIBOSOMAL PROTEIN RP10 >pir||S49366 ribosomal proteinS0.e.B,
                                                cytosolic - yeast (Candid
NE.Contig984       166  7.2e-11   581 721       sp|P40910|RS3A_CANAL 40S RIBOSOMAL PROTEIN RP10 >pir||S49366 ribosomal proteinS0.e.B,
                                                cytosolic - yeast (Candid

<40S RIBOSOMAL PROTEIN S6>
NE.Contig348       603  3.4e-58    24 464       sp|P05752|RS6_SCHPO 40S RIBOSOMAL PROTEIN S6 >pir||R3ZP6E ribosomal proteinS6.e,
                                                cytosolic - fission yeast (Sc
NE.Contig247       307    1e-26   257 475       sp|P05752|RS6_SCHPO 40S RIBOSOMAL PROTEIN S6 >pir||R3ZP6E ribosomal proteinS6.e,
                                                cytosolic - fission yeast (Sc

<40S RIBOSOMAL PROTEIN S16>
NE.Contig427       222    1e-17   294 449       sp|Q42340|RS16_ARATH 40S RIBOSOMAL PROTEIN S16

<40S RIBOSOMAL PROTEIN S18E>
NE.Contig758       586  3.1e-56   355 765       sp|P35271|RS18_YEAST 40S RIBOSOMAL PROTEIN S18E >pir||S50886 ribosomal proteinS18.e,
                                                cytosolic - yeast (Sacchar

<40S RIBOSOMAL PROTEIN S13>
NE.Contig203       622  4.1e-60    42 467       sp|P33192|RS13_CANMA 40S RIBOSOMAL PROTEIN S13 (S15) >pir||S25374 ribosomalprotein
                                                S13.e - yeast (Candida malto

<40S RIBOSOMAL PROTEIN S19>
NE.Contig947       615  2.4e-59    47 466       sp|P27073|RS19_EMENI 40S RIBOSOMAL PROTEIN S19 (S16) >pir||JQ1349 ribosomalprotein
                                                S19.e, cytosolic - Emericell

<40S RIBOSOMAL PROTEIN YS29A>
NE.Contig996       247  2.5e-20   567 734       sp|P41057|R29A_YEAST 40S RIBOSOMAL PROTEIN YS29A >pir||S48503 ribosomalprotein
                                                S29.e.A, cytosolic - yeast (Sacc

<40S RIBOSOMAL PROTEIN S31>
NE.Contig1083      195  8.3e-15   208 423       sp|P07282|RS31_YEAST  40S RIBOSOMAL PROTEIN S31 PRECURSOR (YS23) >pir||S51338ribosomal
                                                protein S25.e.B, cytosoli

2.)  60S ribosomal protein (52)
<60S RIBOSOMAL PROTEIN YEL050C>
g7d06ne.r1     336  9.6e-30   137 454  sp|P32611|YEG0_YEAST  PUTATIVE 60S RIBOSOMAL PROTEIN YEL050C >pir||S30827hypothetical
                                       protein YEL050c - yeast (S

<ribosomal protein l12>
NE.Contig656       660  3.8e-64    46 540       gnl|PID|e1319493 (AL031535) ribosomal protein l12. [Schizosaccharomyces
                                                pombe]>gnl|PID|e1330136 (AL031824)

<ribosomal protein L23>
```

NE.Contig215	440	8.8e-41	120	497	gi|306549 (L13799) homology to rat ribosomal protein L23 [Homo sapiens]
< 60S ribosomal protein>					
NE.Contig794	902	9.7e-90	41	688	gnl|PID|e1362744 (AL035077) 60s ribosomal protein l10 [Schizosaccharomycespombe]
g9d08ne.f1	511	2.1e-48	26	436	gnl|PID|e1325545 (AL031740) 60s ribosomal protein l10a. [Schizosaccharomycespombe]
NE.Contig256	509	3.3e-48	30	467	sp|P29453|RL4B_YEAST 60S RIBOSOMAL PROTEIN L7A-1 (L4-1) (YL5) (RP6)>pir||S16810 ribosomal protein L7a.e.B - yea
NE.Contig1107c	200	3.2e-15	40	369	sp|O14069|YEA4_SCHPO PROBABLE 60S RIBOSOMAL PROTEIN C2E11.04 >pir||T37749 60sribosomal protein l28 - fission yeast (Schizosaccharomyces pombe)>emb|CAA20151.1| (AL031181) 60s ribosomal protein L28[Schizosaccharomyces pombe]>emb|CAA22600.1| (AL035064) 60sribosomal protein L28/L44 [Schizosaccharomyces pombe]
NE.Contig808	478	7.1e-45	150	488	gnl|PID|e1285382 (AL022304) 60s ribosomal protein [Schizosaccharomyces pombe]
NE.Contig965	435	2.8e-40	1	264	sp|P54780|R15B_YEAST 60S RIBOSOMAL PROTEIN YL10 B (L13) (RP15R) (YP18)>pir||S54490 ribosomal protein L15.e.B, c
NE.Contig661	400	1.6e-36	372	683	sp|P40525|YIF2_YEAST PROBABLE 60S RIBOSOMAL PROTEIN YIL052C >pir||S48427ribosomal protein L34.e.B, cytosolic -
NE.Contig952	380	1.9e-34	254	520	gnl|PID|e1285382 (AL022304) 60s ribosomal protein [Schizosaccharomyces pombe]
NE.Contig407	371	1.8e-33	223	492	gnl|PID|e1325545 (AL031740) 60s ribosomal protein l10a. [Schizosaccharomycespombe]
NE.Contig340	355	8.6e-32	308	802	sp|P38064|RM16_YEAST PROBABLE MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L16PRECURSOR >pir||S50292 ribosomal protein L
NE.Contig588	304	1.8e-26	43	462	gnl|PID|e1285382 (AL022304) 60s ribosomal protein [Schizosaccharomyces pombe]
<ribosomal protein>					
NE.Contig839	532	1.6e-50	330	686	gi|596086 (U17360) ribosomal protein YL6b (L5) [Saccharomyces cerevisiae]
NE.Contig734	491	3.2e-46	14	331	gi|3128243 (AF004672) ribosomal protein L41 [Phaffia rhodozyma]
NE.Contig772	438	1.3e-40	157	501	pir||S56056 ribosomal protein L18a.e.c13 - yeast (Saccharomyces cerevisiae)>gi|736308 (Z48756) unknown
d8a09ne.r1	288	1.1e-24	198	431	pir||S51385 ribosomal protein S15a.e.c12 - yeast (Saccharomyces cerevisiae)>gi|609410 (U19103) Rps24bp
NE.Contig599	277	1.5e-23	152	409	gnl|PID|e1285350 (AL022299) ribosomal protein [Schizosaccharomyces pombe]>gnl|PID|e1291883 (AJ001133) ribos
NE.Contig3	239	1.8e-19	222	422	pir||S56056 ribosomal protein L18a.e.c13 - yeast (Saccharomyces cerevisiae)>gi|736308 (Z48756) unknown
<60S RIBOSOMAL PROTEIN YL6>					
NE.Contig788	642	3.1e-62	47	520	sp|P05736|RL6_YEAST 60S RIBOSOMAL PROTEIN YL6 (L5) (RP8) >pir||S50243ribosomal protein L8.e - yeast (Saccharom
<60S RIBOSOMAL PROTEIN L35>					
NE.Contig1040	315	1.4e-27	233	592	sp|P42766|RL35_HUMAN 60S RIBOSOMAL PROTEIN L35 >gi|562074 (U12465) ribosomalprotein L35 [Homo sapiens]
<60S ACIDIC RIBOSOMAL PROTEIN P1>					
NE.Contig609	305	1.6e-26	212	538	sp|P49148|RLA1_ALTAL 60S ACIDIC RIBOSOMAL PROTEIN P1 (ALLERGEN ALT A 12) (ALTA XII) >gi|1006626 (X84216) riboso
<60S ACIDIC RIBOSOMAL PROTEIN P2>					
NE.Contig545	299	7.5e-26	263	592	sp|P42037|RLA2_ALTAL 60S ACIDIC RIBOSOMAL PROTEIN P2 (MINOR ALLERGEN ALT A 6)(ALT A VI) >pir||S43109 acidic rib
<60S RIBOSOMAL PROTEIN L37E A >					

NE.Contig744	196	6.2e-15	3	173	sp|P49166|R7EA_YEAST 60S RIBOSOMAL PROTEIN L37E A (YP55) >pir||S51430ribosomal protein L37.e.A, cytosolic - yea
<60S RIBOSOMAL PROTEIN L37B >					
NE.Contig1059	383	9.7e-35	92	400	sp|P41056|R372_YEAST 60S RIBOSOMAL PROTEIN L37B (YL37) (RP47) >pir||S44069ribosomal protein L35a.e.c15 - yeast
<60S RIBOSOMAL PROTEIN L17>					
NE.Contig714	580	1.3e-55	80	478	sp|P04451|RL1A_YEAST 60S RIBOSOMAL PROTEIN L17 >pir||R5BY17 ribosomal proteinL23.e, cytosolic - yeast (Saccharo
NE.Contig346	222	9.5e-18	293	478	sp|P18621|RL17_HUMAN 60S RIBOSOMAL PROTEIN L17 (L23) >pir||R5HU22 ribosomalprotein L17 - human >gi|34199 (X5377
<60S RIBOSOMAL PROTEIN L18>					
NE.Contig297	234	5.8e-19	246	488	sp|P42791|RL18_ARATH 60S RIBOSOMAL PROTEIN L18 >gi|606970 (U15741) cytoplasmicribosomal protein L18 [Arabidopsi
<60S RIBOSOMAL PROTEIN L18A>					
NE.Contig941	647	9.7e-63	52	558	sp|P47913|RL1X_YEAST 60S RIBOSOMAL PROTEIN L18A >pir||S59848 ribosomal proteinL18a.e.c15 - yeast (Saccharomyces
<60s ribosomal protein L46>					
NE.Contig1045	233	7.7e-19	34	186	gnl|PID|e1316129 (AL031307) *60s ribosomal protein L46 [Schizosaccharomycespombe]*
<60S RIBOSOMAL PROTEIN YL35>					
NE.Contig1003	362	1.5e-32	53	325	sp|P49631|R37A_YEAST 60S RIBOSOMAL PROTEIN YL35 (L37A) >pir||S54068 ribosomalprotein L37a.e - yeast (Saccharomy
NE.Contig692	355	8.5e-32	157	429	sp|P49631|R37A_YEAST 60S RIBOSOMAL PROTEIN YL35 (L37A) >pir||S54068 ribosomalprotein L37a.e - yeast (Saccharomy
<60S RIBOSOMAL PROTEIN YL16B>					
NE.Contig1006	478	7.8e-45	108	650	sp|P05739|R16B_YEAST 60S RIBOSOMAL PROTEIN YL16B >pir||S55970 ribosomalprotein L6.e.B, cytosolic - yeast (Sacch
<60S RIBOSOMAL PROTEIN L14EB>					
NE.Contig1007	333	1.7e-29	53	433	sp|P38754|R14B_YEAST PROBABLE 60S RIBOSOMAL PROTEIN L14EB >pir||S46797ribosomal protein L14.e.B, cytosolic - ye
<60S RIBOSOMAL PROTEIN YL43>					
NE.Contig166	255	3.5e-21	41	217	sp|P05747|RL43_YEAST 60S RIBOSOMAL PROTEIN YL43 >pir||S71066 ribosomal proteinL29.e, cytosolic - yeast (Sacchar
<60S RIBOSOMAL PROTEIN YL39>					
NE.Contig304	251	9e-21	50	337	sp|P05745|RL39_YEAST 60S RIBOSOMAL PROTEIN YL39 >pir||S50922 ribosomal proteinL36.e.A, cytosolic - yeast (Sacch
<60S RIBOSOMAL PROTEIN L26>					
NE.Contig374	386	3.4e-35	70	405	sp|Q02877|RL26_HUMAN 60S RIBOSOMAL PROTEIN L26 >pir||S33713 ribosomal proteinL26 - human >pir||S48864 gene L26
<60S RIBOSOMAL PROTEIN YL17-A>					
NE.Contig652	601	6.8e-58	402	929	sp|P05740|RL7A_YEAST 60S RIBOSOMAL PROTEIN YL17-A >pir||S38012 ribosomalprotein L17.e.A, cytosolic - yeast (Sac
<60S RIBOSOMAL PROTEIN L9 B>					
NE.Contig665	620	6.8e-60	158	733	sp|P51401|RL9B_YEAST 60S RIBOSOMAL PROTEIN L9 B (YL11) (RP25) >pir||S53915ribosomal protein L9.e.B, cytosolic -
NE.Contig824	233	8.1e-19	315	494	sp|P51401|RL9B_YEAST 60S RIBOSOMAL PROTEIN L9 B (YL11) (RP25) >pir||S53915ribosomal protein L9.e.B, cytosolic -

<60S RIBOSOMAL PROTEIN L38>					
NE.Contig668	221	1.4e-17	316	537	sp|Q09900|RL38_SCHPO PROBABLE 60S RIBOSOMAL PROTEIN L38 >pir||S62570 ribosomalprotein L38.e, cytosolic - fissio
<MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L2>					
NE.Contig450	199	6.5e-15	190	474	sp|P12687|RM02_YEAST MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L2 PRECURSOR (YML2)(YMR6) >pir||R6BYM7 ribosomal prote
<60S RIBOSOMAL PROTEIN L11>					
NE.Contig940	675	9.6e-66	56	562	sp|Q10157|RL11_SCHPO PROBABLE 60S RIBOSOMAL PROTEIN L11 >gnl|PID|e220679(Z69240) 60s ribosomal protein L11 [Sch
NE.Contig672	157	8.1e-11	317	484	sp|Q10157|RL11_SCHPO PROBABLE 60S RIBOSOMAL PROTEIN L11 >gnl|PID|e220679(Z69240) 60s ribosomal protein L11 [Sch
<60s ribosomal protein L32>					
NE.Contig695	457	1.3e-42	190	558	sp|P79015|RL32_SCHPO 60S RIBOSOMAL PROTEIN L32 >gnl|PID|d1019962 (AB000914)ribosomal protein L32 homolog [Schiz
<60s ribosomal protein l27>					
NE.Contig798	796	1.6e-78	34	480	sp|P08978|RL2A_NEUCR 60S RIBOSOMAL PROTEIN L27A (L29) (CRP1) >pir||R6NC7Aribosomal protein L27a.e - Neurospora
NE.Contig876	491	3.5e-46	397	801	gnl|PID|e1319770 (AL031543) 60s ribosomal protein l27 [*Schizosaccharomycespombe*]
<60S RIBOSOMAL PROTEIN L22>					
NE.Contig999	280	7.3e-24	174	482	sp|Q09668|RL22_SCHPO 60S RIBOSOMAL PROTEIN L22 >gnl|PID|e334258 (Z98595) 60sribosomal protein l22 [Schizosaccha
<MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L33>					
b1b08ne.r1	117	3.8e-06	489	635	sp|P20084|RM33_YEAST MITOCHONDRIAL 60S RIBOSOMAL PROTEIN L33 (YML33)>pir||S54593 ribosomal protein L30, mitocho
<60s ribosomal protein l2>					
b9f04ne.r1	500	3.7e-47	8	358	gnl|PID|e1335782 (AL032684) 60s ribosomal protein l2 [Schizosaccharomycespombe]
<60S ribosomal protein L24>					
NE.Contig820	413	5.9e-38	453	767	gnl|PID|e276614 (Z81317) 60S ribosomal protein L24 [Schizosaccharomyces pombe]

2.2. Post-translational modifications and regulation (10)

a. Methylation (2)

<serine hydroxymethyltransferase>					
d2f10ne.f1	798	9e-79	2	475	sp|P34898|GLYC_NEUCR SERINE HYDROXYMETHYLTRANSFERASE, CYTOSOLIC (SERINEMETHYLASE) (GLYCINE HYDROXYMETHYLTRANSFE
NE.Contig511	723	6.6e-71	1	498	sp|P34898|GLYC_NEUCR SERINE HYDROXYMETHYLTRANSFERASE, CYTOSOLIC (SERINEMETHYLASE) (GLYCINE HYDROXYMETHYLTRANSFE

b. Glycosylation and addition of other sugars (5)

<glycosyl transferases>					
NE.Contig575	274	3.6e-23	219	488	gi|4226150 (AF125462) similar to glycosyl transferases (Pfam:PF00535,Score=80.8, E=2.9e-20, N=1); str
<GPI-ANCHOR TRANSMIDASE>					

a2c12ne.r1	636	1.5e-61	16	570	sp|P49018|GPI8_YEAST GPI-ANCHOR TRANSMIDASE >pir||S59796 probable membraneprotein YDR331w - yeast (Saccharomyce
<DOLICHYL-PHOSPHATE-MANNOSE--PROTEIN MANNOSYLTRANSFERASE 2>					
NE.Contig474	581	1e-55	430	1086	sp|P31382|PMT2_YEAST DOLICHYL-PHOSPHATE-MANNOSE--PROTEIN MANNOSYLTRANSFERASE 2>pir||S36711 hypothetical protein
b2g02ne.f1	538	3.4e-51	2	547	sp|P31382|PMT2_YEAST DOLICHYL-PHOSPHATE-MANNOSE--PROTEIN MANNOSYLTRANSFERASE 2>pir||S36711 hypothetical protein
<dolichol-phosphate-mannose synthase>					
NE.Contig415	417	2.1e-38	114	473	gnl|PID|d1026693 (AB004789) dolichol-phosphate-mannose synthase [Mus musculus]
c. Other (3)					
<26S PROTEASE REGULATORY SUBUNIT 7>					
b8a01ne.r1	347	6.8e-31	278	763	sp|P33299|PRS7_YEAST 26S PROTEASE REGULATORY SUBUNIT 7 HOMOLOG (CIM5 PROTEIN)(TAT-BINDING HOMOLOG 3) >pir||S343
<26S PROTEASE REGULATORY SUBUNIT 4>					
NE.Contig604	273	1.1e-22	2	229	sp|P36612|PRS4_SCHPO 26S PROTEASE REGULATORY SUBUNIT 4 HOMOLOG (MTS2 PROTEIN)>pir||S39348 26S ATP/ubiquitin-dep
<MITOCHONDRIAL RESPIRATORY CHAIN COMPLEXES ASSEMBLYPROTEIN RCA1>					
NE.Contig636	514	4.7e-48	2	499	sp|P40341|RCA1_YEAST MITOCHONDRIAL RESPIRATORY CHAIN COMPLEXES ASSEMBLYPROTEIN RCA1 (TAT-BINDING HOMOLOG 12) >p

2.3 Folding and targeting (41)

a. Folding (10)					
<PEPTIDYL-PROLYL CIS-TRANS ISOMERASE B PRECURSOR>					
NE.Contig834	969	7.4e-97	85	681	sp|P10255|CYPH_NEUCR PEPTIDYL-PROLYL CIS-TRANS ISOMERASE PRECURSOR (PPIASE)(ROTAMASE) (CYCLOPHILIN) (CYCLOSPORI
NE.Contig704	440	8.4e-41	140	388	sp|P10255|CYPH_NEUCR PEPTIDYL-PROLYL CIS-TRANS ISOMERASE PRECURSOR (PPIASE)(ROTAMASE) (CYCLOPHILIN) (CYCLOSPORI
<cyclophilin>					
NE.Contig217	140	5.7e-09	140	238	sp|P10255|CYPH_NEUCR PEPTIDYL-PROLYL CIS-TRANS ISOMERASE PRECURSOR (PPIASE)(ROTAMASE) (CYCLOPHILIN) (CYCLOSPORI
<PEPTIDYL-PROLYL CIS-TRANS ISOMERASE>					
h5d04ne.r1	224	6.5e-18	311	484	sp|P87051|YDJ3_SCHPO PROBABLE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE C57A10.03>gnl|PID|e313994 (Z94864) peptidyl-p
<peptidylprolyl isomerase>					
NE.Contig612	190	2.9e-14	451	567	gnl|PID|e1292430 (AJ006023) peptidylprolyl isomerase [Neurospora crassa]
<CALNEXIN HOMOLOG PRECURSOR>					
NE.Contig248	257	1.3e-20	159	419	sp|P36581|CALX_SCHPO CALNEXIN HOMOLOG PRECURSOR >pir||A56106 calnexin homologcnx1 - fission yeast (Schizosaccha
<CALNEXIN HOMOLOG>					
NE.Contig248	257	1.3e-20	159	419	sp|P36581|CALX_SCHPO CALNEXIN HOMOLOG PRECURSOR >pir||A56106 calnexin homologcnx1 - fission yeast (Schizosaccha
<FK506-BINDING PROTEIN>					
NE.Contig1084	622	4.3e-60	242	601	sp|P20080|FKBP_NEUCR FK506-BINDING PROTEIN (FKBP) (PEPTIDYL-PROLYL CIS-TRANSISOMERASE)

					(PPIASE) >pir||S11090 FK
c1d11ne.r1	447	1.7e-41	96	641	sp|O60046|FK21_NEUCR FK506-BINDING PROTEIN PRECURSOR (FKBP-21)(PEPTIDYL-PROLYL CIS-TRANS ISOMERASE) (PPIASE) >g
<DISULFIDE ISOMERASE ERP38 PRECURSOR>					
NE.Contig531	378	3e-34	145	432	sp|Q92249|ER38_NEUCR PUTATIVE DISULFIDE ISOMERASE ERP38 PRECURSOR>gnl|PID|e259414 (Y07562) ERp38 [Neurospora cr
b. Chaperones (14)					
<chaperone>					
h4g11ne.r1	311	3.8e-27	76	375	gnl|PID|e1293296 (AL023592) Chaperonins 10 Kd subunit [Schizosaccharomycespombe]
<prefoldin>					
NE.Contig767	314	1.9e-27	63	488	gnl|PID|e1343351 (AL033534) putative prefoldin subunit, molecular chaperoneputative non-native actin bindin
<heat-shock protein30>					
NE.Contig204	162	2.6e-11	20	205	sp|P19752|HS30_NEUCR 30 KD HEAT SHOCK PROTEIN >pir||A38360 heat shock protein30 - Neurospora crassa >gi|168820
<Chaperonin hsp78p>					
NE.Contig347	581	1.1e-55	330	809	gnl|PID|e1319477 (AL031534) Chaperonin hsp78p [Schizosaccharomyces pombe]
<heat shock protein 70>					
NE.Contig561	703	8.9e-69	2	472	sp|Q01233|HS70_NEUCR HEAT SHOCK 70 KD PROTEIN (HSP70) >gi|607818 (U10443) 70kDa heat shock protein [Neurospora
NE.Contig691	438	1.4e-40	329	703	gnl|PID|e267541 (X98931) heat shock protein 70 [Emericella nidulans]
<MOD-E>					
b7e12ne.f1	803	2.4e-79	5	565	gi|2804612 (U81165) MOD-E [Podospora anserina]
NE.Contig238	500	3.8e-47	148	468	gi|2804612 (U81165) MOD-E [Podospora anserina]
<HEAT SHOCK PROTEIN HSP1>					
NE.Contig1001	450	6.8e-42	223	630	sp|P40292|HS82_ASPFU HEAT SHOCK PROTEIN HSP1 (65 KD IGE-BINDING PROTEIN)>gi|1930153 (U92465) heat shock protein
<T-COMPLEX PROTEIN 1, BETA SUBUNIT>					
b9c05ne.f1	450	8e-42	129	482	sp|Q10147|TCPB_SCHPO PROBABLE T-COMPLEX PROTEIN 1, BETA SUBUNIT (TCP-1-BETA)(CCT-BETA) >gnl|PID|e1188807 (Z6923
<chaperonin TCP1 epsilon>					
h6b03ne.f1	352	4.9e-31	152	439	pir||S57083 chaperonin TCP1 epsilon - yeast (Saccharomyces cerevisiae)>gi|1015739 (Z49564) ORF YJR064w
<heat shock protein>					
NE.Contig766	248	2.2e-20	166	867	gnl|PID|d1034019 (AB003518) heat shock protein [Coriolus versicolor]
<zuotin>					
NE.Contig185	402	6.7e-37	49	558	sp|P32527|ZUO1_YEAST ZUOTIN >pir||S25194 zuotin - yeast (Saccharomycescerevisiae) >gi|4837 (X63612) put. zuotin
NE.Contig1020	227	4.1e-18	519	746	gnl|PID|e330325 (Z97992) zuotin like protein [Schizosaccharomyces pombe]

c. Protein sorting and targeting (17)

<vacuolar protein sorting>
NE.Contig388 229 7.6e-17 1 492 gnl|PID|e1250327 (AL021767) vacuolar protein sorting [Schizosaccharomycespombe]

<SERINE CARBOXYPEPTIDASE PRECURSOR>
b3g10ne.f1 122 5.1e-06 16 174 sp|P32826|CBPX_ARATH SERINE CARBOXYPEPTIDASE PRECURSOR >gi|166674 (M81130)carboxypeptidase Y-like protein [Arab

<clathrin-GOLGI ADAPTOR HA1/AP1 ADAPTIN GAMMASUBUNIT>
NE.Contig493 430 8.6e-39 2 433 pir||S49876 gamma-adaptin - smut fungus (Ustilago maydis) >gi|600100 (Z46804)gamma-adaptin [Ustilago m

<ER lumen protein retaining receptor protein>
a5h12ne.f1 270 7.7e-23 295 606 gnl|PID|e1360887 (AL032684) ER lumen protein retaining receptor protein[Schizosaccharomyces pombe]

<edoplasmic reticulum associated protein>
NE.Contig55 314 1.9e-27 176 625 gnl|PID|e1319414 (AL031530) putative edoplasmic reticulum associated protein[Schizosaccharomyces pombe]

<COATOMER ZETA SUBUNIT>
NE.Contig974 372 1.5e-33 308 916 gnl|PID|e1326283 (AL031798) coatomer zeta subunit [Schizosaccharomyces pombe]

<snare protein>
a9b11ne.f1 270 9.4e-23 338 643 gnl|PID|e1340972 (AL033497) probable snare protein [Candida albicans]

<VACUOLAR PROTEIN SORTING-ASSOCIATED PROTEIN VPS28>
NE.Contig795 339 4.7e-30 122 748 sp|Q02767|VP28_YEAST VACUOLAR PROTEIN SORTING-ASSOCIATED PROTEIN VPS28>pir||S60925 hypothetical protein YPL065w

<VACUOLAR PROTEIN SORTING/TARGETING PROTEIN PEP1 PRECURSOR>
c4b11ne.f1 306 4.1e-25 47 436 sp|P32319|PEP1_YEAST VACUOLAR PROTEIN SORTING/TARGETING PROTEIN PEP1 PRECURSOR(VACUOLAR CARBOXYPEPTIDASE SORTIN

<CLATHRIN COAT ASSEMBLY PROTEIN AP19>
NE.Contig373 169 4.5e-12 299 478 sp|P35181|AP19_YEAST CLATHRIN COAT ASSEMBLY PROTEIN AP19 (CLATHRIN COATASSOCIATED PROTEIN AP19) (GOLGI ADAPTOR

<Ca+2-binding EF hand protein>
NE.Contig131 461 4.7e-43 81 758 gi|2270994 (AF004809) Ca+2-binding EF hand protein [Glycine max]

<CARBOXYPEPTIDASE Y PRECURSOR>
NE.Contig261 594 4.3e-57 1 471 sp|P30574|CBPY_CANAL CARBOXYPEPTIDASE Y PRECURSOR (CARBOXYPEPTIDASE YSCY)

<SSO1 PROTEIN>
NE.Contig61 331 3.2e-29 5 667 sp|P32867|SSO1_YEAST SSO1 PROTEIN >pir||S39569 syntaxin-related protein SSO1 -yeast (Saccharomyces cerevisiae)

<Rer1 protein>
NE.Contig813 265 2.7e-22 170 469 gnl|PID|e339854 (AJ001421) Rer1 protein [Homo sapiens]

<MITOCHONDRIAL PROCESSING PEPTIDASE ALPHA SUBUNIT PRECURSOR>
NE.Contig231 415 4e-38 168 416 sp|P23955|MPP1_NEUCR MITOCHONDRIAL PROCESSING PEPTIDASE ALPHA SUBUNITPRECURSOR (ALPHA-

					MPP) >pir||A36442 mitocho
<prohibitin>					
a1b07ne.f1	406	4e-37	92	625	gi|2582388 (AF022225) prohibitin [Pneumocystis carinii]
a1b07ne.r1	393	8e-36	228	575	sp|P40961|PHB_YEAST PROHIBITIN >pir||S64441 prohibitin - yeast (Saccharomycescerevisiae) >gnl|PID|e243702 (Z72

2.4.Turnover-protein degradation-including vacuolar (27)

<proteosome>					
g1d08ne.f1	535	6.6e-51	15	458	gnl|PID|e334324 (Z98603) putative proteosome component [Schizosaccharomycespombe]
g1d08ne.r1	243	6.3e-20	172	492	gnl|PID|e334324 (Z98603) putative proteosome component [Schizosaccharomycespombe]
<POTENTIAL PROTEASOME COMPONENT C5>					
b8b03ne.f1	163	1.1e-09	227	463	sp|P23724|PRC5_YEAST POTENTIAL PROTEASOME COMPONENT C5 (MULTICATALYTICENDOPEPTIDASE COMPLEX SUBUNIT C5) >pir||S
<26S proteasome regulatory subunit mts3>					
a7a06ne.r1	141	3.6e-09	223	486	gnl|PID|e1292693 (AL023554) 26S proteasome regulatory subunit mts3[Schizosaccharomyces pombe]
<PROTEASOME COMPONENT C9/Y13>					
NE.Contig769	891	1.3e-88	100	819	sp|Q09682|PRC9_SCHPO PUTATIVE PROTEASOME COMPONENT C9/Y13 (MACROPAIN SUBUNIT)(MULTICATALYTIC ENDOPEPTIDASE COMP
<PROTEASOME COMPONENT PRE3 PRECURSOR>					
NE.Contig866	676	9.1e-66	174	815	sp|P38624|PRCD_YEAST PROTEASOME COMPONENT PRE3 PRECURSOR (MACROPAIN SUBUNITPRE3) (PROTEINASE YSCE SUBUNIT PRE3)
<26S proteasome subunit 9>					
NE.Contig743	297	1.3e-25	201	503	gi|2150046 (AF001212) *26S proteasome subunit 9* [Homo sapiens]
<ubiquitin precursor>					
NE.Contig922	1015	9.4e-102	88	699	pir||UQNC ubiquitin precursor - Neurospora crassa >gi|295930 (X13140)ubiquitin [Neurospora crassa]
NE.Contig1081	555	5.9e-53	240	572	pir||UQNC ubiquitin precursor - Neurospora crassa >gi|295930 (X13140)ubiquitin [Neurospora crassa]
<ubiquitin conjugating enzyme UBC1>					
NE.Contig572	749	1.4e-73	165	602	gi|3323498 (AF030296) ubiquitin conjugating enzyme UBC1 [Glomerella cingulata]
<UBIQUITIN-CONJUGATING ENZYME E2-24 KD>					
NE.Contig122	345	1.1e-30	3	254	sp|P28263|UBC8_YEAST UBIQUITIN-CONJUGATING ENZYME E2-24 KD (UBIQUITIN-PROTEINLIGASE) (UBIQUITIN CARRIER PROTEIN
<UBIQUITIN-CONJUGATING ENZYME E2-17 KD>					
a1d01ne.f1	664	1.8e-64	131	535	sp|P52493|UBC2_NEUCR UBIQUITIN-CONJUGATING ENZYME E2-17 KD (UBIQUITIN-PROTEINLIGASE 2) (UBIQUITIN CARRIER PROTE
NE.Contig644	604	3.5e-58	90	530	sp|P35128|UBC3_DROME UBIQUITIN-CONJUGATING ENZYME E2-17 KD (UBIQUITIN-PROTEINLIGASE) (UBIQUITIN CARRIER PROTEIN
<ubiquitin-conjugating enzyme protein E2>					
NE.Contig517	303	2.3e-26	6	320	gi|2641619 (AF032468) ubiquitin-conjugating enzyme protein E2 [Zea mays]
<ubiquitin/S27a fusion protein>					
NE.Contig1009	616	1.8e-59	48	509	gi|402242 (U01220) ubiquitin/S27a fusion protein [Neurospora crassa]>gi|402244

					(U01221) ubiquitin/ri
<ubiquitin fusion protein>					
NE.Contig2	655	1.4e-63	45	428	gi|3047314 (AF056623) ubiquitin fusion protein [Magnaporthe grisea]
NE.Contig47	277	1.6e-23	68	223	gi|3047314 (AF056623) ubiquitin fusion protein [Magnaporthe grisea]
					<SERINE-TYPE CARBOXYPEPTIDASE F PRECURSOR>
NE.Contig170	386	4.8e-35	1	471	sp|P52718|PEPF_ASPNG SERINE-TYPE CARBOXYPEPTIDASE F PRECURSOR (PROTEINASE F)(CPD-II)
					>gnl|PID|e1186690 (X79541)
<subtilisin-like serine protease>					
b1d02ne.r1	164	4e-09	434	682	gi|2905804 (AF047689) subtilisin-like serine protease [Podospora anserina]
<CAAX PRENYL PROTEASE 1>					
NE.Contig45	307	1.2e-26	114	551	sp|P47154|ST24_YEAST CAAX PRENYL PROTEASE 1 (PRENYL PROTEIN-SPECIFICENDOPROTEASE 1)
					(PPSEP 1) (A-FACTOR CONVERT
c4d10ne.r1	254	1.6e-20	252	521	sp|P47154|ST24_YEAST CAAX PRENYL PROTEASE 1 (PRENYL PROTEIN-SPECIFICENDOPROTEASE 1)
					(PPSEP 1) (A-FACTOR CONVERT
<ATP-dependent Clp protease proteolytic subunit>					
a3a09ne.r1	353	1.3e-31	187	615	gi|2983755 (AE000735) ATP-dependent Clp protease proteolytic subunit [Aquifexaeolicus]
<Lon protease-like protein>					
a9c06ne.r1	648	1.5e-75	6	464	gi|414046 (X74215) Lon protease-like protein [Homo sapiens]
<PROTEASOME COMPONENT Y7>					
NE.Contig275	370	2.2e-33	94	405	sp|P23639|PRC3_YEAST PROTEASOME COMPONENT Y7 (MACROPAIN SUBUNIT Y7)(PROTEINASE YSCE
					SUBUNIT 7) (MULTICATALYTIC
<UBIQUITIN-LIKE PROTEIN SMT3>					
a2b02ne.r1	128	1.1e-07	439	567	sp|Q12306|SMT3_YEAST UBIQUITIN-LIKE PROTEIN SMT3 >pir||S63999 SMT3 protein -yeast
					(Saccharomyces cerevisiae) >g
<REXB PROTEIN>					
b9e02ne.r1	389	2.3e-35	28	333	sp|P03759|VRXB_LAMBD REXB PROTEIN >pir||IMBPBL rexB protein - phage lambda>gi|215145
					(J02459) rexb (exclusion;1
<PROTEASOME COMPONENT C7-ALPHA>					
a8g09ne.f1	551	1.6e-52	140	631	sp|P21243|PRCI_YEAST PROTEASOME COMPONENT C7-ALPHA (MACROPAIN SUBUNITC7-ALPHA)
					(PROTEINASE YSCE SUBUNIT 7) (MUL

III: Cell growth, cell division and cell process (153)

A. Cell walls, biomembranes and cytoskeleton (45)

1. Cell walls (13)

<septin B>					
NE.Contig448	633	3.1e-61	2	424	gi|1791305 (U83489) septin B [Emericella nidulans]
NE.Contig184	132	3.5e-05	794	964	gi|1791305 (U83489) septin B [Emericella nidulans]
<N,O-DIACETYLMURAMIDASE>					
NE.Contig1088	805	1.7e-79	108	710	sp|P00721|LYCH_CHASP N,O-DIACETYLMURAMIDASE (LYSOZYME CH) >pir||MUKAD lysozyme(EC

					3.2.1.17) - fungus (Chalara s
NE.Contig227	275	2.9e-23	271	477	sp|P00721|LYCH_CHASP N,O-DIACETYLMURAMIDASE (LYSOZYME CH) >pir||MUKAD lysozyme(EC 3.2.1.17) - fungus (Chalara s
<cell wall alpha-glucan synthase>					
NE.Contig943	209	6e-15	4	471	gnl|PID|e1371007 (AL035218) putative cell wall alpha-glucan synthase[Schizosaccharomyces pombe]
<ENDOLYSIN>					
NE.Contig746	507	7.8e-48	1	285	sp|P03706|LYCV_LAMBD ENDOLYSIN (LYSIS PROTEIN) (LYSOZYME) >pir||EYBPLendolysin - phage lambda >gi|215164 (J0245
<cell wall biogenesis protein>					
NE.Contig819	153	7.4e-08	328	684	gnl|PID|e1320954 (AL031579) putative cell wall biogenesis protein, yeast ECM33homolog [Schizosaccharomyces
<glycine rich protein>					
a9d08ne.f1	720	2e-70	1	774	gi|1261823 (L77234) glycine rich protein [Neurospora crassa]
NE.Contig285	174	2.6e-11	9	293	gi|1261823 (L77234) glycine rich protein [Neurospora crassa]
<proline-rich protein 15>					
b8d02ne.r1	122	*0.00024*	564	923	pir||*B39066* proline-rich protein 15 - rat
<SEPTIN HOMOLOG SPN4>					
b7e09ne.f1	376	5.1e-34	233	565	sp|P48009|SPN4_SCHPO SEPTIN HOMOLOG SPN4 >gi|987283 (U29890) septin homolog[Schizosaccharomyces pombe] >gnl|PID
b7e09ne.r1	190	1.1e-13	500	679	sp|P48009|SPN4_SCHPO SEPTIN HOMOLOG SPN4 >gi|987283 (U29890) septin homolog[Schizosaccharomyces pombe] >gnl|PID
<RODLET PROTEIN>					
NE.Contig1110	451	5.2e-42	376	699	sp|Q04571|RODL_NEUCR HYDROPHOBIN PRECURSOR (RODLET PROTEIN) (CLOCK-CONTROLLEDGENE PROTEIN 2) (BLUE LIGHT INDUCE
2. Biomembranes (8)					
<OUTER MEMBRANE USHER PROTEIN>					
NE.Contig926	311	5.3e-26	432	605	sp|P42915|YRAJ_ECOLI HYPOTHETICAL OUTER MEMBRANE USHER PROTEIN IN AGAI-MTRINTERGENIC REGION PRECURSOR >gi|17895
<membrane protein>					
NE.Contig271	338	5.7e-30	197	505	sp|P53224|ORM1_YEAST ORM1 PROTEIN >pir||S64329 probable membrane proteinYGR038w - yeast (Saccharomyces cerevisi
NE.Contig678	319	6.2e-28	91	489	pir||S67579 probable membrane protein YDL046w - yeast (Saccharomycescerevisiae) >gnl|PID|e237287 (Z717
b9e09ne.f1	279	1.1e-23	14	400	pir||S66972 probable membrane protein YOR087w - yeast (Saccharomycescerevisiae) >gnl|PID|e251996 (Z749
NE.Contig867	221	8e-16	304	618	pir||S58824 probable membrane protein YPR194c - yeast (Saccharomycescerevisiae) >gi|786296 (U25841) Si
NE.Contig1014	194	2e-14	1	531	pir||S64936 probable membrane protein YLR100w - yeast (Saccharomycescerevisiae) >gi|1256850 (U53876) Y
NE.Contig72	142	1.2e-08	666	830	pir||S66709 probable membrane protein YOL026c(Saccharomycescerevisiae) >gnl|PID|e252264 (Z747
<erythrocyte membrane antigen 1>					
NE.Contig177	166	5.9e-11	129	593	gi|535699 (L27591) erythrocyte membrane antigen 1 [Plasmodium chabaudi]

3. Cytoskeleton, organelle biogenesis (15)

<kinesin related protein 1>					
NE.Contig958	656	1.1e-63	1	420	gi|4092763 (AF102992) kinesin related protein 1 [Nectria haematococca]
<TUBULIN ALPHA CHAIN>					
NE.Contig214	593	5.2e-57	116	439	sp|Q92335|TBA_SORMA TUBULIN ALPHA CHAIN >gnl|PID|e233852 (Z70290)alpha-tubulin [Sordaria macrospora]
NE.Contig212	455	2.2e-42	221	493	sp|Q92335|TBA_SORMA TUBULIN ALPHA CHAIN >gnl|PID|e233852 (Z70290)alpha-tubulin [Sordaria macrospora]
<ankyrin>					
g4c01ne.r1	127	1.6e-05	177	491	pir||A55575 ankyrin 3, long form - human >gi|608025 (U13616) ankyrin G [Homosapiens]
<nucleosome assembly protein>					
NE.Contig616	141	2.5e-08	666	773	gnl|PID|e1284405 (AL022243) nucleosome assembly protein. [Schizosaccharomycespombe]
<actin>					
NE.Contig366	285	2.3e-24	1	162	sp|O13419|ACT_BOTCI ACTIN >gnl|PID|e328250 (AJ000335) actin [Botryotiniafuckeliana]
<PROFILIN>					
b9f02ne.f1	180	1.3e-12	61	282	sp|P39825|PROF_SCHPO PROFILIN >pir||A53952 profilin - fission yeast(Schizosaccharomyces pombe)
b9f02ne.r1	154	1.8e-10	322	513	sp|P39825|PROF_SCHPO PROFILIN >pir||A53952 profilin - fission yeast(Schizosaccharomyces pombe)
<COFILIN>					
NE.Contig983	307	1.1e-26	177	626	sp|P78929|COFI_SCHPO COFILIN >gnl|PID|d1014741 (D89939) actin depolymerazingfactor [Schizosaccharomyces pombe]
<myosin-II>					
NE.Contig146	294	1.4e-23	3	641	gi|2731818 (AF029788) myosin-II; Myp2p [Schizosaccharomyces pombe]
<PEROXISOMAL MEMBRANE PROTEIN PMP20>					
NE.Contig739	361	1.9e-32	2	364	sp|O43099|PM20_ASPFU PROBABLE PEROXISOMAL MEMBRANE PROTEIN PMP20 (ALLERGEN ASPF 3) >gi|2769700 (U58050) peroxis
<OLEATE-INDUCED PEROXISOMAL PROTEIN POX18>					
NE.Contig975	243	6.4e-20	91	468	sp|P22009|PX18_CANTR OLEATE-INDUCED PEROXISOMAL PROTEIN POX18 (LIPID-TRANSFERPROTEIN) (PXP-18) >pir||JS0155 POX
<phosphatidylethanolamine methyltransferase>					
NE.Contig242	414	3.2e-37	1	435	gi|2209107 (AF004113) phosphatidylethanolamine methyltransferase[Schizosaccharomyces pombe]
<PEROXISOMAL MEMBRANE PROTEIN PMP27>					
NE.Contig64	188	8.6e-14	190	543	sp|Q12462|PEXB_YEAST PEROXISOMAL MEMBRANE PROTEIN PMP27 (PEROXIN-11)>pir||A56509 peroxisomal membrane protein P
<PEROXISOMAL-COENZYME A SYNTHETASE>					
NE.Contig1028	802	3.2e-79	1	714	sp|P38137|FAT2_YEAST PEROXISOMAL-COENZYME A SYNTHETASE >pir||S46098 probableAMP-binding protein - yeast (Saccha

4. Cell cycle control (4)

<cell division cycle CDC48 homolog>					

NE.Contig139	125	1.5e-06	160	420	gi|2612914 (AF015825) cell division cycle CDC48 homolog [Bacillus subtilis]>gnl|PID|e1183262 (Z99110)
<CELL DIVISION CONTROL PROTEIN 11>					
a3g10ne.f1	207	1.6e-15	273	605	sp|P32458|CC11_YEAST CELL DIVISION CONTROL PROTEIN 11 >pir||S40911 celldivision control protein CDC11 - yeast (
<SCH9 protein>					
c3f08ne.r1	649	6.9e-63	9	443	gi|4426 (X12560) SCH9 protein (AA 1-824) [Saccharomyces cerevisiae]
<BARRIERPEPSIN PRECURSOR>					
NE.Contig613	327	3.8e-28	2	856	sp|P12630|BAR1_YEAST BARRIERPEPSIN PRECURSOR (EXTRACELLULAR "BARRIER" PROTEIN)(BAR PROTEINASE) >pir||A34084 ext

5. Mitosis/cytokinesis (3)

5.1 MITOSIS (2)

<CENTROMERE/MICROTUBULE BINDING PROTEIN CBF5>					
NE.Contig316	659	5e-64	1	465	sp|O43102|CBF5_ASPFU CENTROMERE/MICROTUBULE BINDING PROTEIN CBF5(CENTROMERE-BINDING FACTOR 5) (NUCLEOLAR PROTEI
<extragenic suppressor of the bimD6 mutation>					
NE.Contig863	163	1.8e-10	138	482	gi|2338556 (AF013590) extragenic suppressor of the bimD6 mutation [Emericellanidulans]

5.2 Cytokinesis (1)

<TROPOMYOSIN>					
NE.Contig750	182	2e-13	271	435	sp|Q02088|TPM_SCHPO TROPOMYOSIN >pir||S27127 tropomysin - yeast(Schizosaccharomyces pombe) >gi|173517 (L04126)

6. Other (2)

<PH RESPONSIVE PROTEIN 1 PRECURSOR>					
b3f04ne.f1	419	1.1e-38	5	550	sp|P43076|PHR1_CANAL PH RESPONSIVE PROTEIN 1 PRECURSOR (PH-REGULATED PROTEIN1) >gi|857672 (M90812) pH responsiv
<silk fibroin heavy chain>					
NE.Contig641	211	1.3e-15	425	751	bbs|157676 (S74439) silk fibroin heavy chain (C-terminal) [Bombyxmori=silkworms, Peptide Partial, 633

B. Cell processes (108)

1. Cell rescue, defense, osmotic adaptation, starvation response, development (asexual, sexual)(includes antibiotics,toxins) see also B.cell signalling, signal transduction and C. transmembrane transport (49)

1.1 Development (10)

a. Asexual (6)

<CONIDIATION-SPECIFIC PROTEIN 6>					
NE.Contig954	486	1.1e-45	90	368	sp|P34762|CON6_NEUCR CONIDIATION-SPECIFIC PROTEIN 6 >gi|415714 (L26036)conidiation protein [Neurospora crassa]
<CONIDIATION-SPECIFIC PROTEIN 8>					
NE.Contig157	419	1.4e-38	174	629	sp|P10169|CON8_NEUCR CONIDIATION-SPECIFIC PROTEIN 8 >pir||S02210 con-8 protein-Neurospora crassa >gi|2989 (X07
<CONIDIATION-SPECIFIC PROTEIN 10>					
NE.Contig951	279	1e-23	257	514	sp|P10713|CONX_NEUCR CONIDIATION-SPECIFIC PROTEIN 10 >pir||A31849conidiation-specific

					protein - Neurospora cras
<COPROPORPHYRINOGEN III OXIDASE PRECURSOR>					
NE.Contig916	414	4.1e-38	153	514	sp|P36551|HEM6_HUMAN COPROPORPHYRINOGEN III OXIDASE PRECURSOR(COPROPORPHYRINOGENASE) (COPROGEN OXIDASE) (COX) >
NE.Contig985	237	2.6e-19	251	505	sp|P36552|HEM6_MOUSE COPROPORPHYRINOGEN III OXIDASE PRECURSOR (COPROPORPHYRINOGENASE) (COPROGEN OXIDASE) (COX) >
<UROPORPHYRINOGEN DECARBOXYLASE>					
a6d11ne.f1	551	1.3e-52	5	586	sp|P32347|DCUP_YEAST UROPORPHYRINOGEN DECARBOXYLASE (UPD) >pir||S23471uroporphyrinogen decarboxylase (EC 4.1.1.
b. Sexual cycle (2)					
<krev-1>					
h7b01ne.f1	537	5e-51	122	433	gnl|PID|d1033372 (AB000281) krev-1 [Neurospora crassa]
<GUANINE NUCLEOTIDE-BINDING PROTEIN BETA SUBUNIT-LIKEPROTEIN>					
NE.Contig494	1398	2.3e-142	61	915	sp|Q01369|GBLP_NEUCR GUANINE NUCLEOTIDE-BINDING PROTEIN BETA SUBUNIT-LIKEPROTEIN (CROSS-PATHWAY CONTROL WD-REPE
c. Morphology, sporulation, growth of Fungi (1)					
<SPS2 protein>					
NE.Contig1050	358	4.3e-32	506	1135	pir||S70297 SPS2 protein homolog YBR078w - yeast (Saccharomyces cerevisiae)>gnl|PID|e304680 (Z35947) O
d Fungi pathgen(cause disease) (1)					
<snodprot1>					
NE.Contig1077	405	3.8e-37	185	532	gi|3329509 (AF074941) snodprot1 [Phaeosphaeria nodorum]
1.2. Defense (5)					
a. Defense protein (3)					
<L-AMINO ACID OXIDASE PRECURSOR>					
NE.Contig896	803	2.9e-79	41	496	sp|P23623|OXLA_NEUCR L-AMINO ACID OXIDASE PRECURSOR (LAO) >pir||A38314L-amino-acid oxidase (EC 1.4.3.2) precurs
f9d04ne.f1	785	2.6e-77	6	452	sp|P23623|OXLA_NEUCR L-AMINO ACID OXIDASE PRECURSOR (LAO) >pir||A38314L-amino-acid oxidase (EC 1.4.3.2) precurs
<cytosolic NADPH oxidase p67-phox>					
e3e12ne.f1	161	3.3e-10	169	474	gi|3687891 (AF079303) cytosolic NADPH oxidase p67-phox [Bos taurus]
b. Sterigmatocystin biosynthesis (2)					
<sterigmatocystin>					
b8e06ne.r1	190	1.3e-12	200	532	sp|P55790|OMTA_ASPFL STERIGMATOCYSTIN 7-O-METHYLTRANSFERASE PRECURSOR>gi|413844 (L25836) O-methyltransferase [A
<STERIGMATOCYSTIN BIOSYNTHESIS PROTEIN>					
g5h05ne.f1	313	2.1e-27	2	421	sp|Q00717|STCT_EMENI PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS PROTEIN STCT>gi|1235632 (U34740) putative translati
1.3. Detoxification (8)					
<singlet oxygen resistance protein>					
NE.Contig995	903	6.4e-90	192	1073	gi|2979688 (AF035619) singlet oxygen resistance protein [Cercosporanicotianae]
NE.Contig851	212	2.1e-16	189	491	gi|2979688 (AF035619) singlet oxygen resistance protein [Cercosporanicotianae]

NE.Contig690	211	2.8e-16	210	509	gi|2979688 (AF035619) singlet oxygen resistance protein [Cercosporanicotianae]
<CATALASE A>					
NE.Contig870	372	9.9e-33	1	438	sp|P78574|CATA_ASPFU CATALASE A >gi|1843578 (U87630) catalase [Aspergillusfumigatus]
<SUPEROXIDE DISMUTASE PRECURSOR>					
NE.Contig935	609	9.1e-59	1	669	sp|O13401|SODM_CANAL SUPEROXIDE DISMUTASE PRECURSOR (MN) >gi|2623885(AF031478) manganese-superoxide dismutase p
<SUPEROXIDE DISMUTASE>					
NE.Contig225	461	4.5e-43	118	390	sp|P07509|SODC_NEUCR SUPEROXIDE DISMUTASE (CU-ZN) >pir||A36591
<cytochrome P450 monooxygenase>					
NE.Contig970	140	1.4e-08	389	499	gnl|PID|e1284329 (AJ004810) cytochrome P450 monooxygenase [Zea mays]
<CYTOCHROME P450 55A2>					
h8b02ne.f1	374	8.6e-34	148	453	sp|Q00616|NOR1_CYLTO CYTOCHROME P450 55A2 (CYTOCHROME P450NOR1)>gnl|PID|d1012073 (D78511) cytochrome P450nor1 [superoxidedismutase (EC 1.15.1.1) (Cu-Zn) - Neur
1.4. Dessication tolerance (1)					
<rehydrin>					
NE.Contig730	716	4.1e-70	3	611	gnl|PID|e1340097 (AL033396) rehydrin-like protein [Candida albicans]
1.5. Oxidative stress (4)					
<flavohemoglobin>					
NE.Contig840	437	1.5e-40	99	503	gnl|PID|d1033977 (AB016807) flavohemoglobin [Fusarium oxysporum]
NE.Contig807	211	6.4e-16	104	472	gnl|PID|d1033977 (AB016807) flavohemoglobin [Fusarium oxysporum]
NE.Contig846	159	3.4e-10	220	468	gnl|PID|d1033977 (AB016807) flavohemoglobin [Fusarium oxysporum]
NE.Contig1092	134	3.5e-05	165	344	gnl|PID|d1033977 (AB016807) flavohemoglobin [Fusarium oxysporum]
1.6. Night/day rhythm (circadium rhythem--biological clock) (13)					
<ccg-4 putative polypeptide 2>					
NE.Contig1042	456	1.6e-42	731	991	gi|1184781 (U46085) ccg-4 putative polypeptide 2; Method: conceptualtranslation supplied by author. [
<ccg-4 putative polypeptide 1>					
NE.Contig872	381	1.7e-34	270	482	gi|1184782 (U46085) ccg-4 putative polypeptide 1; Method: conceptualtranslation supplied by author. [
NE.Contig512	344	1.4e-30	50	238	gi|1184782 (U46085) ccg-4 putative polypeptide 1; Method: conceptualtranslation supplied by author. [
<clock-controlled gene-6 protein>					
NE.Contig1104	398	2.5e-36	151	450	gi|3746899 (AF088908) clock-controlled gene-6 protein [Neurospora crassa]
NE.Contig1103d	163	2.3e-11	3	95	gi|3746899 (AF088908) clock-controlled gene-6 protein [Neurospora crassa]
NE.Contig992	130	6.5e-08	617	697	gi|3746899 (AF088908) clock-controlled gene-6 protein [Neurospora crassa]
<clock-controlled gene-9 protein>					
NE.Contig955	919	1.6e-91	119	646	gi|3746895 (AF088906) clock-controlled gene-9 protein [Neurospora crassa]

NE.Contig485	835	1.2e-82	1	459	gi|1184788 (U46088) similar to sucrose-phosphate synthase; Method: conceptualtranslation supplied by
c5b02ne.f1	798	1.1e-78	3	470	gi|3746895 (AF088906) clock-controlled gene-9 protein [Neurospora crassa]
NE.Contig761	391	5.7e-66	237	455	gi|3746895 (AF088906) clock-controlled gene-9 protein [Neurospora crassa]
<GLUCOSE-REPRESSIBLE GENE PROTEIN>					
NE.Contig1108	359	3.1e-32	73	285	sp|P22151|GRG1_NEUCR GLUCOSE-REPRESSIBLE GENE PROTEIN >gi|3014 (X14801) grg1gene product [Neurospora crassa]
NE.Contig352	352	2e-31	265	477	sp|P22151|GRG1_NEUCR GLUCOSE-REPRESSIBLE GENE PROTEIN >gi|3014 (X14801) grg1gene product [Neurospora crassa]
NE.Contig1	313	2.4e-27	266	466	sp|P22151|GRG1_NEUCR GLUCOSE-REPRESSIBLE GENE PROTEIN >gi|3014 (X14801) grg1gene product [Neurospora crassa]
1.7. Tumor protein and tumor suppressor (2)					
<TRANSLATIONALLY CONTROLLED TUMOR PROTEIN>					
NE.Contig666	495	1.2e-46	193	702	sp|P35691|TCTP_YEAST TRANSLATIONALLY CONTROLLED TUMOR PROTEIN HOMOLOG (TCTP)>pir||S37878 IgE-dependent histamin
<tumor metastasis inhibitor nm23-H2>					
NE.Contig980	546	3.7e-52	164	667	pir||A49798 tumor metastasis inhibitor nm23-H2 - human (fragment)
1.8. Multidrug resistance (3)					
<CYANIDE HYDRATASE>					
NE.Contig972	654	1.8e-63	62	532	sp|P32964|CYHY_GLOSO CYANIDE HYDRATASE (FORMAMIDE HYDROLYASE) >pir||JQ1613cyanide hydratase (EC 4.2.1.66) - imp
NE.Contig801	351	2.5e-31	235	492	sp|P32963|CYHY_FUSLA CYANIDE HYDRATASE (FORMAMIDE HYDROLYASE) >gi|168151(M99046) cyanide hydratase [Fusarium la
NE.Contig328	289	9.4e-25	216	455	sp|P32963|CYHY_FUSLA CYANIDE HYDRATASE (FORMAMIDE HYDROLYASE) >gi|168151(M99046) cyanide hydratase [Fusarium la
1.9. Other (3)					
<heavy metal tolerance protein precursor>					
a7c03ne.f1	362	1.8e-32	26	541	gnl|PID|e1375219 (AL031546) heavy metal tolerance protein precursor[Schizosaccharomyces pombe]
<CAP20>					
b9d03ne.f1	197	5.5e-15	172	435	*gi|603050 (U18061) CAP20 [Glomerella cingulata]*
<carboxyphosphonoenolpyruvate mutase>					
e3e10ne.f1	301	5e-26	1	471	gi|47149 (X67953) carboxyphosphonoenolpyruvate mutase [Streptomyceshygroscopicus]
2. Cell signalling, signal transduction and second messengers (22)					
2.1.PHOSPHATASES (1)					
<<PROTEIN-TYROSINE PHOSPHATASE>					
a5c10ne.f1	180	5.5e-12	3	191	sp|P38148|YB9T_YEAST PROBABLE PROTEIN-TYROSINE PHOSPHATASE YBR276C>pir||S44538 probable protein-tyrosine-phosph
2.2. Kinases (8)					
<protein kinase skp1p>					
c1g06ne.f1	190	8.9e-12	159	377	gnl|PID|e1362542 (AL035064) protein kinase skp1p [Schizosaccharomyces pombe]

<protein kinase CK2 beta subunit>
c2b07ne.f1 173 2.9e-10 296 643 gi|3093417 (AF036546) protein kinase CK2 beta subunit [Candida albicans]

<protein kinase kin1>
a3b11ne.r1 150 2.4e-21 390 617 gnl|PID|e1319466 (AL031534) protein kinase kin1 [Schizosaccharomyces pombe]
a2b11ne.f1 146 7.4e-07 28 378 gnl|PID|e1319466 (AL031534) protein kinase kin1 [Schizosaccharomyces pombe]

<protein kinase C>
NE.Contig618 823 2.4e-81 3 485 gnl|PID|e311856 (Y12002) protein kinase C homologue [Neurospora crassa]
NE.Contig655 356 1.4e-30 270 476 gnl|PID|e311856 (Y12002) protein kinase C homologue [Neurospora crassa]

<mitogen-activated protein kinase>
a1e12ne.r1 408 2.1e-37 283 567 gi|4106374 (AF069777) mitogen-activated protein kinase kinase CPK1 [Cryphonectria parasitica]
a3d09ne.r1 399 1.8e-36 353 619 gi|1279911 (U52963) mitogen-activated protein kinase [Nectria haematococcavar. brevicona]

2.3.cAMP (1)
<amiB>
a5g09ne.f1 225 1.2e-17 173 604 gnl|PID|e1251137 (AL021841) amiB [Mycobacterium tuberculosis]

2.4. Calmodulin (3)
<calmodulin>
NE.Contig1073 757 2.1e-74 76 522 sp|Q02052|CALM_NEUCR CALMODULIN >pir||S58709 calmodulin - Neurospora crassa>gi|5542 (X70923) calmodulin [Neuros
NE.Contig898 283 3.5e-24 213 476 gnl|PID|e321945 (Y13784) Calmodulin [Mougeotia scalaris]

<CALCIUM-BINDING PROTEIN>
NE.Contig32 137 1.3e-08 383 505 sp|Q09711|NCS1_SCHPO HYPOTHETICAL CALCIUM-BINDING PROTEIN C18B11.04 INCHROMOSOME I >pir||S58303 hypothetical pr

2.5. G protein(8)
<ADP-RIBOSYLATION FACTOR>
NE.Contig667 570 1.4e-54 149 490 sp|P34727|ARF_AJECA ADP-RIBOSYLATION FACTOR >pir||D49993 ADP-ribosylationfactor - Ajellomyces capsulata >gi|40

<GTP-BINDING NUCLEAR PROTEIN GSP2/CNR2>
NE.Contig1030 915 4.1e-91 578 1201 sp|P32836|GSP2_YEAST GTP-BINDING NUCLEAR PROTEIN GSP2/CNR2 >pir||S35505GTP-binding protein GSP2 - yeast (Saccha

<rho-gdp dissociation inhibitor>
NE.Contig897 153 2.4e-10 18 158 gnl|PID|e334111 (Z98533) rho-gdp dissociation inhibitor [Schizosaccharomycespombe]

<GTPase>
NE.Contig718 119 4.4e-05 2 85 gnl|PID|e311352 (Y12314) GTPase [Schizosaccharomyces pombe]

<RAN-SPECIFIC GTPASE-ACTIVATING PROTEIN>
NE.Contig854 185 8.1e-14 363 512 sp|P41920|RANG_YEAST RAN-SPECIFIC GTPASE-ACTIVATING PROTEIN (RAN BINDINGPROTEIN 1 HOMOLOG) (RANBP1) (PERINUCLEA

<YPT1-RELATED PROTEIN 2>

a1e04ne.r1	421	9.6e-39	157	495	sp|P17609|YPT2_SCHPO YPT1-RELATED PROTEIN 2 >pir||S12790 GTP-binding proteinypt2 - fission yeast (Schizosacchar
<YPT1-RELATED PROTEIN 5>					
NE.Contig96	389	2.2e-35	239	565	sp|P36586|YPT5_SCHPO YPT1-RELATED PROTEIN 5 >pir||S34729 GTP-binding proteinypt5 - fission yeast (Schizosacchar
<GTP-BINDING PROTEIN YPT51/VPS21>					
e4c08ne.f1	441	7e-41	152	466	sp|P36017|YP51_YEAST GTP-BINDING PROTEIN YPT51/VPS21 >pir||S43399 GTP-bindingprotein VPS21 - yeast (Saccharomyc
2.6. Membrane receptor(1)					
<trk-1>					
NE.Contig35	249	3.1e-19	425	637	gnl|PID|e1331352 (AJ009758) trk-1 [Neurospora crassa]
3. Transmembrane transport (37)					
3.1. Secretion (2)					
<SEC14 CYTOSOLIC FACTOR>					
a5d01ne.f1	336	8.8e-30	265	618	sp|Q10137|SC14_SCHPO PUTATIVE SEC14 CYTOSOLIC FACTOR(PHOSPHATIDYLINOSITOL/PHOSPHATIDYL-CHOLINE TRANSFER PROTEIN
<SECRETORY PATHWAY GDP DISSOCIATION INHIBITOR>					
NE.Contig741	404	5.7e-37	3	374	sp|P39958|GDI1_YEAST SECRETORY PATHWAY GDP DISSOCIATION INHIBITOR >pir||S44446GDP dissociation inhibitor GDI1 -
3.2. Transport (35)					
a. Sugar transport (6)					
<sugar transport protein>					
c4c09ne.f1	252	3.5e-20	5	280	gi|409547 (L07492) sugar transport protein [Saccharomyces cerevisiae]
<AmMst-1>					
NE.Contig1067	349	5.3e-31	39	386	*gnl|PID|e290368 (Z83828) AmMst-1 [Amanita muscaria]*
<GLUCOSE TRANSPORTER>					
NE.Contig402	396	7.2e-36	52	573	sp|Q92253|RCO3_NEUCR PROBABLE GLUCOSE TRANSPORTER RCO-3 >gi|1314738 (U54768)RCO3 [Neurospora crassa]
<GLUCOSE TRANSPORTER RCO-3>					
NE.Contig836	359	1e-31	132	4d5	sp|Q92253|RCO3_NEUCR PROBABLE GLUCOSE TRANSPORTER RCO-3 >gi|1314738 (U54768)RCO3 [Neurospora crassa]
<hexose transporter>					
c4a06ne.f1	303	1.1e-25	64	525	gnl|PID|e1312137 (Y16834) hexose transporter [Candida albicans]
<GLUCOSE/GALACTOSE TRANSPORTER>					
NE.Contig1098	434	3.7e-40	340	1485	sp|Q44623|GLUP_BRUAB GLUCOSE/GALACTOSE TRANSPORTER >gi|1171339 (U43785)glucose/galactose transporter [Brucella
b. Cation transport-ATPase, or major facilitator superfamily (7)					
<E1-E2 ATPases>					
a3c11ne.r1	508	5e-47	103	612	gnl|PID|e1348085 (Z92807) similar to E1-E2 ATPases; cDNA EST EMBL:C11117 comesfrom this gene; cDNA EST EMBL
<CALCIUM-TRANSPORTING ATPASE>					
a2c11ne.f1	195	2.1e-13	104	457	pir||S27763 Ca2+-transporting ATPase (EC 3.6.1.38) - tomato >pir||A46284Ca(2+)-ATPase, LCA1 - Lycopers

<calcium/proton exchanger>					
<CCC1 PROTEIN>					
a5a12ne.r1	163	5.6e-11	242	598	sp|P47818|CCC1_YEAST CCC1 PROTEIN >pir||S43453 CCC1 protein - yeast(Saccharomyces cerevisiae) >gi|609369 (U1902
<COPPER TRANSPORT PROTEIN CTR3>					
NE.Contig557	134	1.7e-06	279	413	sp|Q06686|CTR3_YEAST COPPER TRANSPORT PROTEIN CTR3 (COPPER TRANSPORTER 3)>pir||S59377 probable membrane protein
<VITAMIN D3 HYDROXYLASE-ASSOCIATED PROTEIN>					
a3b11ne.f1	317	9.2e-28	59	655	sp|Q90578|VDHA_CHICK VITAMIN D3 HYDROXYLASE-ASSOCIATED PROTEIN (VDHAP)>pir||A53101 vitamin D3 hydroxylase-assoc
<potassium channel subunit>					
c5d12ne.f1	441	7.1e-41	33	410	gnl|PID|e1293250 (AL023590) putative potassium channel subunit[Schizosaccharomyces pombe]
<oxaloacetate decarboxylase>					
NE.Contig796	143	2.9e-08	174	392	gnl|PID|d1030871 (AP000003) 571aa long hypothecical oxaloacetate decarboxylasealpha chain [Pyrococcus horik
c. Anion transport (2)					
<tartrate transport>					
d7g06ne.f1	182	1.1e-12	22	489	sp|P70786|TUB3_AGRVI PUTATIVE TARTRATE TRANSPORTER >gi|984367 (U32375)membrane protein [Agrobacterium vitis]
<sulfate permease II>					
NE.Contig844	365	6.4e-32	248	490	gi|168911 (M59167) sulfate permease II [Neurospora crassa]
d.Protein, amino acid transport (10)					
<PROTEIN TRANSPORT PROTEIN>					
NE.Contig938	245	4.2e-20	440	649	sp|Q09827|S61G_SCHPO PUTATIVE PROTEIN TRANSPORT PROTEIN SEC61 GAMMA SUBUNIT>pir||S62479 hypothetical protein SP
<ERV25 PROTEIN PRECURSOR>					
c2b10ne.r1	288	1.2e-24	228	578	sp|P54837|ERV5_YEAST ERV25 PROTEIN PRECURSOR >pir||S55107 probable membraneprotein YML012w - yeast (Saccharomyc
c2b10ne.f1	161	3.2e-11	234	437	sp|P54837|ERV5_YEAST ERV25 PROTEIN PRECURSOR >pir||S55107 probable membraneprotein YML012w - yeast (Saccharomyc
<AMINO-ACID PERMEASE>					
NE.Contig79	182	1.5e-24	1	249	gnl|PID|d1014066 (D87432) Similar to Schistosoma mansoni amino acid permease(L25068). [Homo sapiens]
<PROLINE-SPECIFIC PERMEASE>					
NE.Contig175	609	1.2e-58	2	595	sp|P18696|PUTX_EMENI PROLINE-SPECIFIC PERMEASE (PROLINE TRANSPORT PROTEIN)>gnl|PID|e264747 (X79797) proline per
<SEC61 protein>					
b8e07ne.r1	217	2.1e-16	413	625	gnl|PID|e1188753 (Y11322) SEC61 protein [Yarrowia lipolytica]
<MITOCHONDRIAL PHOSPHATE CARRIER PROTEIN>					
d5g09ne.f1	498	4e-47	1	465	sp|P23641|MPCP_YEAST MITOCHONDRIAL PHOSPHATE CARRIER PROTEIN (PHOSPHATETRANSPORT PROTEIN) (MITOCHONDRIAL IMPORT
<AUTOPHAGOCYTOSIS PROTEIN AUT1>					
NE.Contig501	223	8.7e-18	245	556	sp|P40344|AUT1_YEAST AUTOPHAGOCYTOSIS PROTEIN AUT1 >pir||S45130 hypotheticalprotein YNR007c - yeast (Saccharomy

NE.Contig110	138	1.2e-06	268	393	sp|P40344|AUT1_YEAST AUTOPHAGOCYTOSIS PROTEIN AUT1 >pir||S45130 hypotheticalprotein YNR007c - yeast (Saccharomy
<Opt1p>					
NE.Contig658	291	6.6e-24	237	614	gi|2367386 (U60973) Opt1p [Candida albicans]
e. Mitochondrial transport (8)					
<2-oxoglutarate/malate translocator>					
NE.Contig199	208	6.4e-29	2	355	pir||S65042 2-oxoglutarate/malate translocator (clone OMT103), mitochondrialmembrane - proso millet >g
NE.Contig371	179	7.1e-13	164	484	pir||S65042 2-oxoglutarate/malate translocator (clone OMT103), mitochondrialmembrane - proso millet >g
<oxoglutarate malate translocator>					
NE.Contig712	156	2.8e-10	349	513	gnl|PID|e257809 (X99853) oxoglutarate malate translocator [Solanum tuberosum]
<ADP,ATP CARRIER PROTEIN>					
NE.Contig701	698	3.2e-68	40	480	sp|P02723|ADT_NEUCR ADP,ATP CARRIER PROTEIN (ADP/ATP TRANSLOCASE) (ADENINENUCLEOTIDE TRANSLOCATOR) (ANT) >pir|
NE.Contig991	504	1.4e-47	364	666	sp|P02723|ADT_NEUCR ADP,ATP CARRIER PROTEIN (ADP/ATP TRANSLOCASE) (ADENINENUCLEOTIDE TRANSLOCATOR) (ANT) >pir|
<OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN>					
NE.Contig379	764	3.4e-75	1	447	sp|P07144|PORI_NEUCR OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN >pir||MMNCPporin - Neurospora crassa >gi|3057 (
NE.Contig601	319	5.8e-28	398	616	sp|P07144|PORI_NEUCR OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN >pir||MMNCPporin - Neurospora crassa >gi|3057 (
<frataxin>					
NE.Contig357	226	4.3e-18	336	674	gi|2344992 (U95736) frataxin [Mus musculus]
f. Other(2)					
<transport protein>					
NE.Contig640	118	2.1e-05	3	176	sp|P33897|ALD_HUMAN ADRENOLEUKODYSTROPHY PROTEIN (ALDP) >pir||S30059 probabletransport protein ALD - human >gi
<ATP-DEPENDENT BILE ACID PERMEASE>					
c6a07ne.f1	366	2e-31	30	380	sp|P32386|YBT1_YEAST ATP-DEPENDENT BILE ACID PERMEASE >pir||S64800 probablemembrane protein YLL048c - yeast (Sa

IV: UNCLASSIFIED, UNIDENTIFIED, NO SIGNIFICANT HOMOLOGY (1161)

A. Classes of enzymes (from M. Reily and KEGG; no pathway specified) (3)

1. Oxidoreductases (3)					
<OXIDOREDUCTASE>					
NE.Contig195	184	7e-12	393	575	sp|O07575|YHDF_BACSU HYPOTHETICAL OXIDOREDUCTASE IN CITA-SSPB INTERGENICREGION >gnl|PID|e1191878 (Y14082) hypot
NE.Contig109	456	1.8e-42	100	738	sp|Q10216|YAY8_SCHPO HYPOTHETICAL OXIDOREDUCTASE C4H3.08 IN CHROMOSOME I>gi|1184021 (Z69380) unknown [Schizosac
<SQUALENE MONOOXYGENASE>					

NE.Contig20	496	1.1e-46	19	708	sp|Q92206|ERG1_CANAL SQUALENE MONOOXYGENASE (SQUALENE EPOXIDASE) (SE)>gnl|PID|d1014259 (D88252) squalene epoxid

B. Non-enzymatic classes (not in defined pathways) (1)

1. Zinc finger motif-DNA binding (1)

<zinc finger protein>					
d3b11ne.r1	139	4.3e-08	339	473	gnl|PID|e1291640 (AL023288) Zinc finger protein [Schizosaccharomyces pombe]

C. Unclassified (significant homolog but function uncertain in Neurospora crassa) (30)

<Ran/spi1 binding protein>					
NE.Contig395	338	5.6e-30	395	670	gnl|PID|d1013767 (D86381) Ran/spi1 binding protein [Schizosaccharomyces pombe]>gnl|PID|e1339991 (AL033389)
<het-c2 protein>					
NE.Contig576	510	3.1e-48	3	437	pir||S59950 het-c2 protein - Podospora anserina >gi|523338 (U05236) HET-C2[Podospora anserina]
<YSA1 PROTEIN>					
a1d08ne.r1	143	2.5e-09	238	591	sp|Q01976|YSA1_YEAST YSA1 PROTEIN >pir||S48276 YSA1 protein - yeast(Saccharomyces cerevisiae) >gi|476067 (X7899
<23S rRNA intron 2 protein>					
NE.Contig703	178	3.4e-12	142	261	pir||S06607 23S rRNA intron 2 protein - Podospora anserina mitochondrion(SGC3) >gi|13307 (X14735) unid
NE.Contig831	166	6.9e-11	2	124	pir||S06607 23S rRNA intron 2 protein - Podospora anserina mitochondrion(SGC3) >gi|13307 (X14735) unid
NE.Contig107	145	1.2e-08	27	119	pir||S06607 23S rRNA intron 2 protein - Podospora anserina mitochondrion(SGC3) >gi|13307 (X14735) unid
<Pmt3p>					
a3b02ne.f1	144	2.1e-09	297	593	gnl|PID|d1033560 (AB017187) Pmt3p [Schizosaccharomyces pombe]
<SONA>					
b4h07ne.r1	580	1.3e-55	104	619	gi|3202044 (AF069492) SONA [Emericella nidulans]
<ferric leghemoglobin reductase-2 precursor>					
c7c02ne.f1	188	3.3e-13	79	423	gi|3309269 (AF074940) ferric leghemoglobin reductase-2 precursor [Glycine max]
<methylumbelliferyl-acetate deacetylase (EC 3.1.1.56)>					
d2h03ne.r1	275	2.6e-23	123	473	pir||A23543 methylumbelliferyl-acetate deacetylase (EC 3.1.1.56) - human(fragment) >gi|182265 (M13450)
<IgE-binding protein>					
NE.Contig1085	218	3.1e-17	124	657	gnl|PID|e1299119 (AJ006688) IgE-binding protein [Aspergillus fumigatus]
<symbiosis-related protein>					
NE.Contig1008	553	9.5e-53	194	565	gi|2072023 (U93506) symbiosis-related protein [Laccaria bicolor]
<rAsp f 7>					
NE.Contig1025	148	8e-10	131	427	gnl|PID|e1250610 (AJ223315) rAsp f 7 [Aspergillus fumigatus

<phosphoribosylaminoimidazolesuccinocarboxamide synthase>					
NE.Contig6	525	8e-50	10	471	pir||JQ1395 phosphoribosylaminoimidazolesuccinocarboxamide synthase (EC6.3.2.6) - yeast (Saccharomyces
i2c04ne.r1	155	3.4e-10	220	381	pir||S55292 phosphoribosylaminoimidazolesuccinocarboxamide synthase (EC6.3.2.6) (clone CipA400) - yeas
<Vip1 protein>					
NE.Contig925	249	1.4e-20	82	480	gnl|PID|e321532 (Y13635) Vip1 protein [Schizosaccharomyces pombe]>gnl|PID|e1202248 (AL009197) hypothetical
<AX110P>					
a7d10ne.r1	115	1.5e-05	278	466	gnl|PID|d1003966 (D14605) AX110P [Daucus carota] >prf||2004427Aembryogenesis-associated protein [Daucus car
<PDH1P>					
b2a09ne.f1	133	3.9e-08	149	439	gnl|PID|d1025871 (D88384) PDH1P [Schizosaccharomyces pombe] >gnl|PID|e1325973(AL031764) transembrane protei
<Ser/Arg-related nuclear matrix protein>					
b8a09ne.f1	203	8.7e-13	108	431	gi|3005587 (AF048977) Ser/Arg-related nuclear matrix protein [Homo sapiens]
<secretory protein>					
c6b12ne.f1	122	4.2e-07	317	451	gnl|PID|e1323863 (AL031644) putative secretory protein [Schizosaccharomycespombe]
<PROTEIN SSP120 PRECURSOR>					
c6b12ne.r1	189	2.9e-14	220	441	sp|P39931|S120_YEAST PROTEIN SSP120 PRECURSOR >pir||JH0483 secretory proteinSSP120 precursor - yeast (Saccharom
<endosomal P24A protein>					
NE.Contig323	270	7.6e-22	2	457	gnl|PID|e1351322 (Z79759) Similarity to Yeast endosomal P24A protein(SW:EM70_YEAST); cDNA EST CEMSB40F come
<subunit of the final step of the secretory pathway>					
c5e10ne.f1	128	2.1e-06	149	451	gnl|PID|e1316739 (AL031324) subunit of the final step of the secretory pathway[Schizosaccharomyces pombe] >
<YnaD>					
NE.Contig968	138	7.3e-09	225	476	gi|1750115 (U66480) YnaD (Bacillus subtilis] >gnl|PID|e1183411 (Z99113)similar to hypothetical protei
<RDS1 PROTEIN>					
NE.Contig200	222	3.6e-17	285	821	sp|P53693|RDS1_SCHPO RDS1 PROTEIN >pir||S58477 rds1 protein - fission yeast(Schizosaccharomyces pombe) >gnl|PID
<2-hydroxyhepta-2,4-diene-1,7-dioate isomerase>					
NE.Contig963	467	1.1e-43	105	809	gi|2648302 (AE000952) 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase (hpcE-2)[Archaeoglobus fulgidus]
<NIPSNAP1 protein>					
d3g10ne.f1	327	7.7e-29	3	377	gnl|PID|e1231231 (AJ001258) NIPSNAP1 protein [Homo sapiens]
<Flavin-binding monooxygenase>					
NE.Contig150	142	2.6e-08	3	293	gnl|PID|e1343666 (Z71258) similar to Flavin-binding monooxygenase-like[Caenorhabditis elegans]
<acetyltransferase>					
NE.Contig848	246	2.6e-20	88	402	gi|1766060 (U82217) putative acetyltransferase [Schizosaccharomyces pombe]

<BLI-3 PROTEIN>

NE.Contig670	1089	1.3e-109	179	805	sp|Q01358|BLI3_NEUCR BLI-3 PROTEIN >gi|602075 (X81318) bli-3 gene product [Neurospora crassa] >prf||2023159A bli

D. Unidentified (includes significant match with ORFs) (105)

<unknown function>

a4a02ne.f1	878	3.4e-87	3	641	gnl|PID|d1019541 (D90917) hypothetical protein [Synechocystis sp.]
NE.Contig548	786	1.9e-77	20	493	pir||A25784 hypothetical 70K protein - Neurospora crassa mitochondrion (SGC3)>gi|13144 (X04512) pot. U
NE.Contig386	638	1.6e-60	424	975	sp|P47169|YJ9F_YEAST HYPOTHETICAL 161.2 KD PROTEIN IN NMD5-HOM6 INTERGENICREGION >pir||S57160 sulfite reductase
NE.Contig771	614	2.5e-59	5	478	gnl|PID|d1014509 (D89150) similar to Saccharomyces serevisiae hypothetical52.9KD protein in CDC26-YMR31 int
b7b05ne.f1	570	5.6e-54	14	574	gnl|PID|e1292820 (AJ005963) 100 kDa protein [Ajellomyces capsulatus]
NE.Contig44	432	1.2e-53	757	1275	sp|Q04958|YMF9_YEAST HYPOTHETICAL 187.1 KD PROTEIN IN OGG1-CNA2 INTERGENICREGION >pir||S49802 probable membrane
NE.Contig864	538	3.2e-51	416	898	pir||S69049 hypothetical protein YPL135w - yeast (Saccharomyces cerevisiae)>gi|1244779 (U43703) Lpi10p
f6d11ne.f1	493	2.3e-46	14	343	sp|P19677|YAJC_ECOLI HYPOTHETICAL 11.9 KD PROTEIN IN TGT-SECD INTERGENICREGION (ORF12) >pir||D38530 yajC protei
NE.Contig573	483	2.3e-45	1	480	pir||A25784 hypothetical 70K protein - Neurospora crassa mitochondrion (SGC3)>gi|13144 (X04512) pot. U
NE.Contig805	477	1.1e-44	559	924	sp|Q04013|YM78_YEAST HYPOTHETICAL 34.2 KD PROTEIN IN CUS1-RPL18A1 INTERGENICREGION >pir||S56055 hypothetical pr
NE.Contig477	471	4.1e-44	2	778	sp|P38197|YBD6_YEAST HYPOTHETICAL 29.1 KD PROTEIN IN URA7-POL12 INTERGENICREGION >pir||S50294 hypothetical prot
NE.Contig496	430	1.1e-39	46	309	pir||S04556 hypothetical protein L - Neurospora crassa mitochondrion (SGC3)>gnl|PID|e14054 (X13337) OR
NE.Contig42	432	5.9e-39	177	719	gnl|PID|e1292820 (AJ005963) 100 kDa protein [Ajellomyces capsulatus]
NE.Contig869	428	1.2e-38	1	444	gnl|PID|d1019541 (D90917) hypothetical protein [Synechocystis sp.]
c4c12ne.r1	419	1.6e-38	6	497	sp|Q09885|YAH9_SCHPO HYPOTHETICAL 43.0 KD PROTEIN C8A4.09C IN CHROMOSOME I>pir||S62525 hypothetical protein SPA
a9a12ne.f1	405	4.8e-37	269	763	gi|3676056 (M59935) unknown [Emericella nidulans]
NE.Contig123	410	1.4e-36	222	725	gnl|PID|e1284411 (AL022244) hypothetical protein [Schizosaccharomyces pombe]
c4g10ne.f1	391	1.4e-35	139	483	sp|Q04013|YM78_YEAST HYPOTHETICAL 34.2 KD PROTEIN IN CUS1-RPL18A1 INTERGENICREGION >pir||S56055 hypothetical pr
NE.Contig847	389	2.1e-35	214	891	sp|P53189|YGC8_YEAST HYPOTHETICAL 56.4 KD PROTEIN IN RPL32-CWH41 INTERGENICREGION PRECURSOR >pir||S64030 probab
NE.Contig622	386	4.7e-35	173	688	gnl|PID|e1250326 (AL021767) hypothetical protein [Schizosaccharomyces pombe]
NE.Contig148	378	1.2e-33	8	448	gi|1129167 (X87297) J1590 gene product [Saccharomyces cerevisiae]
NE.Contig800	284	2.3e-33	462	947	sp|Q09875|YAGC_SCHPO HYPOTHETICAL 35.8 KD PROTEIN C12G12.12 IN CHROMOSOME I>pir||S62543 hypothetical protein SP
NE.Contig932	277	1.1e-32	332	640	gnl|PID|e1359019 (AL034433) hypothetical protein [Schizosaccharomyces pombe]
NE.Contig562	355	6.9e-32	27	458	sp|P36160|YK61_YEAST HYPOTHETICAL 39.6 KD PROTEIN IN MTD1-NUP133 INTERGENICREGION >pir||S38159 hypothetical pro

NE.Contig197	353	1.5e-31	163	456	sp|Q09896|YAI9_SCHPO HYPOTHETICAL 13.5 KD PROTEIN C24B11.09 IN CHROMOSOME I>pir||S62554 hypothetical protein SP
b2d12ne.f1	346	8.4e-31	115	441	pir||S61029 hypothetical protein YPL235w - yeast (Saccharomyces cerevisiae)>gi|1061254 (Z67751) putati
NE.Contig977	332	2.5e-29	6	485	gnl|PID|e1263908 (AL022103) hypothetical protein [Schizosaccharomyces pombe]
h5d04ne.f1	327	7.5e-29	170	478	gi|1330343 (U58755) C34D4.12 gene product [Caenorhabditis elegans]
NE.Contig997	322	3.8e-28	169	507	sp|P43616|YFL4_YEAST HYPOTHETICAL 52.9 KD PROTEIN IN SAP155-YMR31 INTERGENICREGION >pir||S56299 hypothetical pr
NE.Contig394	317	8.8e-28	42	398	sp|P40037|YEO7_YEAST HYPOTHETICAL 13.9 KD PROTEIN IN FCY2-PET117 INTERGENICREGION >pir||S50560 hypothetical pro
NE.Contig397	316	1.2e-27	17	469	gnl|PID|e1330143 (AL031825) hypothetical protein [Schizosaccharomyces pombe]
d4g05ne.f1	313	2.3e-27	1	498	sp|P38716|YHR2_YEAST HYPOTHETICAL 42.4 KD PROTEIN IN CDC12-ORC6 INTERGENICREGION >pir||S48954 hypothetical prot
c3a05ne.f1	304	3e-25	11	694	sp|Q10327|YD72_SCHPO HYPOTHETICAL 97.1 KD PROTEIN C32A11.02C IN CHROMOSOME I>gi|1213266 (Z69796) unknown [Schiz
b9e04ne.f1	291	5.9e-25	54	440	pir||S61091 hypothetical protein YPL051w - yeast (Saccharomyces cerevisiae)>gi|1079691 (U39205) Lpe21p
NE.Contig186	295	2.4e-24	3	446	sp|P42935|YG4C_YEAST HYPOTHETICAL 89.4 KD TRP-ASP REPEATS CONTAINING PROTEININ PMT6-PCT1 INTERGENIC REGION >pir
NE.Contig418	281	5.4e-24	17	412	pir||S59765 hypothetical protein YPR100w - yeast (Saccharomyces cerevisiae)>gi|914971 (U32445) Note th
NE.Contig449	290	6.7e-24	183	551	pir||S66749 hypothetical protein YOL057w - yeast (Saccharomyces cerevisiae)>gnl|PID|e251865 (Z74799) O
a4b08ne.f1	279	1e-23	292	603	sp|P87132|YDM1_SCHPO HYPOTHETICAL PROTEIN C57A7.01 IN CHROMOSOME I>gnl|PID|e316110 (Z95396) unknown [Schizosacc
NE.Contig648	267	1.8e-22	115	804	sp|P53111|YGP7_YEAST HYPOTHETICAL 38.1 KD PROTEIN IN RCK1-AMS1 INTERGENICREGION >pir||S60428 hypothetical prote
d4g05ne.r1	257	2.1e-21	171	428	sp|P38716|YHR2_YEAST HYPOTHETICAL 42.4 KD PROTEIN IN CDC12-ORC6 INTERGENICREGION >pir||S48954 hypothetical prot
c4c12ne.f1	255	3.2e-21	104	478	sp|Q09885|YAH9_SCHPO HYPOTHETICAL 43.0 KD PROTEIN C8A4.09C IN CHROMOSOME I>pir||S62525 hypothetical protein SPA
NE.Contig48	255	3.6e-21	193	687	sp|P40531|YIE1_YEAST 36.7 KD PROTEIN IN CBR5-NOT3 INTERGENIC REGION>pir||S49937 hypothetical protein YIL041w -
a9g08ne.f1	260	4.3e-21	385	597	sp|P38821|YHR3_YEAST HYPOTHETICAL 54.2 KD PROTEIN IN ERP5-ORC6 INTERGENICREGION >pir||S48955 hypothetical prote
NE.Contig398	252	1.2e-20	109	420	gnl|PID|e1330143 (AL031825) hypothetical protein [Schizosaccharomyces pombe]
a8a08ne.r1	250	1.3e-20	336	533	gi|3676056 (M59935) unknown [Emericella nidulans]
NE.Contig492	248	2e-20	622	819	pir||S64830 hypothetical protein YLR008c - yeast (Saccharomyces cerevisiae)>gnl|PID|e245490 (Z73180) O
NE.Contig129	246	3.4e-20	129	443	gnl|PID|e332509 (Z74798) ORF YOL057w [Saccharomyces cerevisiae]
g2b12ne.r1	243	6.8e-20	186	419	sp|P50085|YG4W_YEAST HYPOTHETICAL 34.9 KD PROTEIN IN SMI1-PHO81 INTERGENICREGION >pir||S57696 prohibitin homolo
a2h07ne.f1	238	2.4e-19	124	570	gnl|PID|e1319374 (AL031523) hypothetical protein [Schizosaccharomyces pombe]
a3a12ne.r1	235	5e-19	236	616	gnl|PID|e1182667 (Z99107) alternate gene name: yeeM, yfxB; similar tohypothetical proteins [Bacillus subtil
					sp|P53823|YN74_YEAST HYPOTHETICAL 25.2 KD PROTEIN IN THI12 5'REGION>pir||S63320 probable

NE.Contig901	234	5.5e-19	81	473	membrane protein YNL33
g2b12ne.f1	229	2.1e-18	223	465	sp|P50085|YG4W_YEAST HYPOTHETICAL 34.9 KD PROTEIN IN SMI1-PHO81 INTERGENICREGION >pir||S57696 prohibitin homolo
NE.Contig560	223	9.5e-18	41	478	gnl|PID|d1018544 (D90909) hypothetical protein [Synechocystis sp.]
a8b07ne.r1	221	2.3e-17	338	535	sp|P36015|YKT6_YEAST HYPOTHETICAL 22.7 KD PROTEIN IN PAS1-MST1 INTERGENICREGION >pir||S38033 cell division cont
NE.Contig444	225	2.4e-17	314	610	gnl|PID|e351278 (Z99532) hypothetical protein [Schizosaccharomyces pombe]
NE.Contig677	216	4e-17	73	330	pir||S72314 hypothetical protein YHR004c-a - yeast (Saccharomyces cerevisiae)>gnl|PID|e273884 (Z80875)
b8h05ne.r1	214	8.3e-17	259	642	gnl|PID|e315884 (Z95397) unknown [Schizosaccharomyces pombe]
NE.Contig605	212	7.5e-16	81	245	gnl|PID|e1293418 (AL023595) hypothetical protein [Schizosaccharomyces pombe]
NE.Contig101	203	1.2e-15	359	724	gnl|PID|e1340002 (Z98533) hypothetical protein [Schizosaccharomyces pombe]
f9f10ne.f1	193	1e-14	122	445	pir||S18542 hypothetical protein 4 - Streptomyces coelicolor >gi|46816(X58833) actVA 4 gene product [S
NE.Contig311	200	1.3e-14	1	372	sp|P47125|YJ48_YEAST HYPOTHETICAL 50.8 KD PROTEIN IN MIR1-STE18 INTERGENICREGION >pir||S57097 indoleamine-pyrro
NE.Contig389	191	2e-14	126	293	sp|O14155|YE72_SCHPO HYPOTHETICAL 15.9 KD PROTEIN C4A8.02C IN CHROMOSOME I>gnl|PID|e338958 (Z98762) hypothetica
g6b09ne.f1	203	3.1e-14	89	322	sp|P38329|YB85_YEAST HYPOTHETICAL 124.0 KD PROTEIN IN PCS60-ABD1
NE.Contig770	189	3.6e-14	1	162	gnl|PID|e221782 (Z69379) unknown [Schizosaccharomyces pombe]
NE.Contig882	186	7.6e-14	264	416	sp|O14171|YE54_SCHPO HYPOTHETICAL 30.2 KD PROTEIN C4D7.04C IN CHROMOSOME I
a4a04ne.f1	184	6.6e-13	176	427	sp|Q09849|YAE9_SCHPO HYPOTHETICAL 47.8 KD PROTEIN C23D3.09 IN CHROMOSOME I>pir||S62500 hypothetical protein SPA
NE.Contig59	179	7.8e-13	104	595	gnl|PID|e1349781 (Z82053) cDNA EST yk302b12.3 comes from this gene[Caenorhabditis elegans]
NE.Contig136	182	8.6e-13	110	412	pir||S59404 hypothetical protein YLR435w - yeast (*Saccharomyces cerevisiae*)>gi|665970 (U21094) Ylr435w
a2c08ne.r1	172	2.4e-12	448	585	pir||S61012 hypothetical protein YPL252c - yeast (Saccharomyces cerevisiae)>gi|1061237 (Z67751) putati
NE.Contig603	169	4.6e-12	450	626	gnl|PID|e1198269 (Z97210) hypothetical protein [Schizosaccharomyces pombe]
b4g11ne.r1	179	5.3e-12	303	467	gnl|PID|e1287629 (AL022580) putative protein [Arabidopsis thaliana]
NE.Contig57	168	5.5e-12	80	379	sp|P47131|YJ55_YEAST HYPOTHETICAL 11.3 KD PROTEIN IN MIR1-STE18 INTERGENICREGION >pir||S57104 probable membrane
NE.Contig504	166	9.8e-12	35	250	gnl|PID|e1341122 (AJ012688) hypothetical protein [Cicer arietinum]
a1h12ne.r1	167	3e-11	217	513	sp|Q06063|YL05_YEAST HYPOTHETICAL 41.7 KD PROTEIN IN SFP1-CTR3 INTERGENICREGION >pir||S55961 hypothetical prote
NE.Contig63	171	4.1e-11	1	321	sp|Q05497|YD38_YEAST HYPOTHETICAL 77.8 KD PROTEIN IN MRPS28-HXT7 INTERGENICREGION >pir||S70103 probable membran
NE.Contig355	154	1.7e-10	151	420	gnl|PID|e1132726 (Z99167) hypothetical conserved protein [Schizosaccharomycespombe]
NE.Contig708	154	1.8e-10	255	377	pir||C30208 hypothetical protein 2 (cpc-1 5' region) - Neurospora crassa>gi|168794 (J03262) open readi
d9b04ne.f1	163	2.9e-10	13	480	gnl|PID|e1371866 (AL035263) hypothetical protein [Schizosaccharomyces pombe]
NE.Contig625	161	3.3e-10	287	553	sp|P53189|YGC8_YEAST HYPOTHETICAL 56.4 KD PROTEIN IN RPL32-CWH41 INTERGENICREGION PRECURSOR >pir||S64030 probab
NE.Contig268	149	6.1e-10	198	446	gnl|PID|e1295831 (AL023781) hypothetical protein [Schizosaccharomyces pombe]
NE.Contig1029	150	6.4e-10	234	641	gi|1323704 (U55387) similar to C. elegans F38E1.9 gene product encoded byGenBank Accession Number U41

NE.Contig532	147	8.9e-10	331	495	sp|Q10010|YSV4_CAEEL HYPOTHETICAL 26.6 KD PROTEIN T19C3.4 IN CHROMOSOME III>gi|849238 (U28412) similar to polyp
NE.Contig777	152	1.9e-09	167	472	pir||S67639 hypothetical protein YDL097c - yeast (Saccharomyces cerevisiae)>gnl|PID|e223236 (X95644) O
NE.Contig305	138	9.5e-09	46	183	sp|P34655|YOT0_CAEEL HYPOTHETICAL 8.7 KD PROTEIN ZK632.10 IN CHROMOSOME III>gnl|PID|e1351197 (Z22181) ZK632.10
NE.Contig325	138	9.5e-09	181	441	gi|2982301 (AF051235) YGL010w-like protein [Picea mariana]
a1c10ne.r1	151	9.6e-09	70	555	gnl|PID|e1256460 (AL022070) hypothetical protein [Schizosaccharomyces pombe]
c1g08ne.f1	155	1.1e-08	345	494	sp|Q07821|YL27_YEAST HYPOTHETICAL 27.7 KD PROTEIN IN PRP19-HSP104 INTERGENICREGION >pir||S64778 hypothetical pr
NE.Contig119	143	1.3e-08	220	555	gnl|PID|e311459 (Z93938) unknown [Bacillus subtilis] >gnl|PID|e1185992(Z99119) similar to hypothetical pro
h3e04ne.f1	144	1.4e-08	177	419	gnl|PID|e1359036 (AL034433) hypothetical protein [Schizosaccharomyces pombe]
NE.Contig533	131	4.5e-08	78	461	sp|Q04969|YMF5_YEAST HYPOTHETICAL 20.8 KD PROTEIN IN CNA2-CYB2 INTERGENICREGION >pir||S49805 probable membrane
b4c06ne.f1	143	1.1e-07	60	560	sp|P47123|YJ44_YEAST HYPOTHETICAL 24.3 KD PROTEIN IN PEM2-HOC1 INTERGENICREGION >pir||S57093 hypothetical prote
c8g11ne.f1	127	1.6e-07	234	503	sp|P28707|YKL7_YEAST 24.1 KD PROTEIN IN VMA12-APN1 INTERGENIC REGION>pir||S27382 hypothetical protein YKL117w -
c5c06ne.f1	137	2.3e-07	57	425	sp|Q10327|YD72_SCHPO HYPOTHETICAL 97.1 KD PROTEIN C32A11.02C IN CHROMOSOME I>gi|1213266 (Z69796) unknown [Schiz
NE.Contig213	124	3.2e-07	38	367	gi|2924772 (AC002334) unknown protein [Arabidopsis thaliana]
c1d01ne.f1	148	8.1e-07	329	595	gnl|PID|e1253936 (AL022002) hypothetical protein Rv1672c [Mycobacteriumtuberculosis]
NE.Contig736	116	2.3e-06	228	383	gnl|PID|d1031403 (AP000006) 196aa long hypothetical protein [Pyrococcushorikoshii]
NE.Contig939	121	1.3e-05	224	487	gnl|PID|e1249670 (AL021712) putative protein [Arabidopsis thaliana]
NE.Contig738	128	1.5e-05	765	947	gi|2984241 (AE000767) putative protein [Aquifex aeolicus]
NE.Contig205	129	1.6e-05	42	425	sp|Q10058|YAM3_SCHPO HYPOTHETICAL 41.5 KD PROTEIN C1F5.03C IN CHROMOSOME I>gi|1103730 (Z68136) unknown [Schizos
b3c12ne.r1	134	3.1e-05	323	517	pir||S51348 hypothetical protein YLR345w - yeast (Saccharomyces cerevisiae)>gi|609382 (U19028) Ylr345w
NE.Contig113	125	4e-05	329	523	sp|P53285|YG3H_YEAST HYPOTHETICAL 54.5 KD PROTEIN IN CBF2-SKN1 INTERGENICREGION >pir||S64450 probable membrane
c4g08ne.r1	124	5.5e-05	98	433	gi|2649315 (AE001017) conserved hypothetical protein [Archaeoglobus fulgidus]
a5d12ne.r1	123	7.9e-05	427	690	gnl|PID|e334110 (Z98533) hypothetical protein [Schizosaccharomyces pombe]
NE.Contig480	120	0.00015	181	561	pir||I39929 hypothetical protein orfM - Bacillus subtilis (fragment)>gi|551718 (L16808) Gene disrupted
<Saccharomyces cerevisiae SCD6 protein>					
a4b09ne.f1	234	8.1e-19	194	508	gnl|PID|d1014528 (D89169) similar to Saccharomyces cerevisiae SCD6 protein,SWISS-PROT Accession Number P459

E. No significant homolog (1022)

<NONE>

1. Contigs (594)

2. Singlets(428)

Appendix V. Genes only expressed in NM cDNA library that contain 3'EST only

	GENE NAME	HSP Score	P Value
Contig1380	None		
Contig1373	putative RNA helicase [Mus musculus]	438	1.9e-40
Contig1332	clock-controlled gene-9 protein [Neurospora crassa]	896	5.6e-89
Contig1293	None		
Contig1285	None		
Contig1264	None		
Contig1233	None		
Contig1230	None		
Contig1224	None		
Contig1222	peroxisomal receptor for PTS2-containing proteins Pex7p	233	1.8e-18
Contig1218	NITRILASE 3 >gi|508735 (U09959) nitrilase [Arabidopsisthaliana]	159	2.9e-10
Contig1180	acyl-CoA dehydrogenase, putative[Deinococcus radiodurans]	181	1.7e-12
Contig1179	None		
Contig1144	glutaminase A [Aspergillus oryzae]	237	3.6e-18
Contig1134	None		
Contig1128	CHITIN SYNTHASE 3 (CHITIN-UDP ACETYL-GLUCOSAMINYLTRANSFERASE 3)	284	7.3e-23
Contig1126	None		
Contig1095	HYPOTHETICAL 8.3 KD PROTEIN IN ANSP-RHSE INTERGENICREGION	203	1.6e-15
Contig1092	cyclophilin [Trichophyton mentagrophytes]	402	1.1e-36
Contig1090	None		
Contig1088	None		
Contig1076	calcium binding protein [Homo sapiens]	228	3.3e-18
Contig1058	78 KD GLUCOSE-REGULATED PROTEIN HOMOLOG PRECURSOR (GRP78)	565	5.9e-54
Contig1049	PROTEIN SNODPROT1 PRECURSOR >gi|3329509 (AF074941)snodprot1	172	2.6e-12
Contig1028	none		
Contig1014	Hypothetical Protein [Chlamydia pneumoniae]	133	2.8e-05
Contig993	None		
Contig986	None		
Contig980	None		
Contig956	ubiquitin/S27a fusion protein [Neurospora crassa]	315	9.6e-37
Contig955	none		
Contig943	Glycogen phosphorylase; Gph1p	256	6.2e-2
Contig903	None		
Contig897	HYPOTHETICAL 15.7 KD PROTEIN IN NUP	126	2.2e-07
Contig896	TFS1 [Saccharomyces cerevisiae]	109	2.9e-05
Contig887	uricase [Nilaparvata lugens yeast-like symbiont]	350	4.1e-31
Contig881	none		
Contig877	AmMst-1 [Amanita muscaria]	237	2.1e-18
Contig876	VACUOLAR PROTEASE A PRECURSOR	504	2.0e-47

Contig861	none		
Contig846	trp-asp repeat protein [Schizosaccharomyces pombe]	116	0.00034
Contig826	None		
Contig805	none		
Contig789	None		
Contig762	None		
Contig758	None		
Contig739	clock-controlled gene-9 protein [Neurospora crassa]	415	2.6e-37
Contig737	None		
Contig733	None		
Contig723	none		
Contig713	None		
Contig708	TUBULIN ALPHA-B CHAIN	288	2.5e-24
Contig702	yeast Chaperonin hsp78 homolog [Schizosaccharomycespombe]	191	4.7e-13
Contig693	MANNOSYL-OLIGOSACCHARIDE ALPHA-1,2-MANNOSIDASE PRECURSOR (MAN(9)-ALPHA-MANNOSIDASE)	200	2.0e-14
Contig685	None		
Contig680	None		
Contig677	Hypothetical Protein [Chlamydia pneumoniae]	118	9.3e-06
Contig671	None		
Contig668	alpha subunit of succinyl-CoA ligase; Lsc1p	358	5.7e-32
Contig652	serine threonine-protein kinase[Schizosaccharomyces pombe]	227	9.0e-17
Contig649	30 KD HEAT SHOCK PROTEIN >pir||A38360 heat shock protein30	178	6.7e-13
Contig622	none		
Contig541	aspartyl-trna synthetase, cytoplasmic[Schizosaccharomyces pombe]	299	5.3e-25
Contig537	None		
Contig504	soluble cell wall protein; Scw10p	121	4.8e-06
Contig498	None		
Contig493	None		
Contig471	PROTEIN DISULFIDE ISOMERASE PRECURSOR (PDI)	267	9.5e-22
Contig464	putative phenylalanyl-tRNA synthetasebeta-subunit; PheHB	182	2.6e-12
Contig461	None		
Contig420	None		
Contig414	None		
Contig353	None		
Contig351	None		
Contig349	None		
Contig318	None		
Contig302	None		
Contig274	None		
Contig254	None		
Contig234	None		

Contig215	None		
Contig169	None		
Contig139	None		
Contig134	None		
Contig96	None		
Contig84	None		
Contig57	NAD(+)-isocitrate dehydrogenase subunit I[Ajellomyces capsulatus	214	3.0e-1
Contig55	None		
Contig53	glyceraldehyde 3-phosphate dehydrogenase [Neurosporacrassa]	313	7.7e-54
Contig48	GLUCOSE-REPRESSIBLE GENE PROTEIN	330	5.1e-29
Contig10	THIAZOLE BIOSYNTHETIC ENZYME PRECURSOR (STRESS-INDUCIBLEPROTEIN STI35)	465	2.6e-43
Contig1	None		

Appendix VI. Genes only expressed in the NM cDNA library that contain 5' ESTs only

	Gene Name	HSP Score	P Value
Contig1391	None		
Contig1383	None		
Contig1370	putative pre-mrna_splicing factor atp-dependent rnahelicase	803	4.1e-79
Contig1345	peroxisomal receptor for PTS2-containing proteins Pex7p	231	2.6e-28
Contig1335	AP-1-LIKE TRANSCRIPTION FACTOR	143	1.9e-06
Contig1309	None		
Contig1308	None		
Contig1273	None		
Contig1268	putative methionine aminopeptidase 1	685	1.0e-66
Contig1252	PYRUVATE DEHYDROGENASE E1 COMPONENT ALPHA SUBUNIT, MITOCHONDRIAL PRECURSOR (PDHE1-A)	235	1.5e-18
Contig1235	None		
Contig1193	None		
Contig1183	high-affinity zinc transport protein; Zrt1p	364	1.2e-32
Contig1172	similar to Nitrilase	352	2.4e-31
Contig1170	none		
Contig1139	None		
Contig1135	GLYCOGEN PHOSPHORYLASE	628	1.9e-60
Contig1131	URICASE (URATE OXIDASE)	571	1.4e-54
Contig1130	alcohol dehydrogenase	324	1.6e-28
Contig1109	HEAT SHOCK 70 KD PROTEIN (HSP70)	894	7.8e-89
Contig1106	78 KD GLUCOSE-REGULATED PROTEIN HOMOLOG PRECURSOR (GRP78)	821	4.4e-81
Contig1100	None		
Contig1099	None		
Contig1077	glutaminase A	302	4.1e-25
Contig1067	None		
Contig1064	None		
Contig1062	None		
Contig1042	probable phenylalanyl-trna synthetase	548	2.3e-60
Contig1026	HYPOTHETICAL 110.9 KD PROTEIN IN SP	235	1.2e-17
Contig1024	None		
Contig1013	ATPase; Rpt6p/26S PROTEASE REGULATORY SUBUNIT 8	603	5.6e-58
Contig998	None		
Contig972	homeodomain DNA-binding transcription factor	444	1.2e-40
Contig947	None		
Contig944	Arf-binding protein; Gga2p	197	6.1e-14
Contig933	none		
Contig929	397aa long hypothetical tyrosine aminotransferase	148	5.5e-09
Contig920	None		
Contig911	None		
Contig892	YKL207W/Ykl207wp/HYPOTHETICAL 29.4 KD PROTEIN IN ST	292	5.7e-25
Contig888	3'phosphoadenylyl thiosulfotransferase	112	2.9e-05
Contig874	hexokinase	285	1.0e-23

Contig872	None		
Contig853	None		
Contig837	HYPOTHETICAL 32.6 KD PROTEIN IN VP	134	4.7e-06
Contig811	thiazole biosynthetic enzyme	213	4.6e-34
Contig798	None		
Contig786	None		
Contig743	HYPOTHETICAL 25.1 KD PROTEIN IN PMI	216	6.1e-17
Contig738	HEAT SHOCK 70 KD PROTEIN (HSP70)	645	2.0e-62
Contig734	hypothetical protein	156	1.5e-10
Contig718	GMC oxidoreductase	111	0.0016
Contig715	DEVELOPMENTAL REGULATOR FLBA	363	1.2e-31
Contig699	mannosyl-oligosaccharide 1,2-alpha-mannosidase	399	2.5e-36
Contig676	None		
Contig655	similar to sorbitol dehydrogenase; cDNA ESTEMBL:T00701 comes from this gene	428	2.2e-39
Contig642	None		
Contig638	None		
Contig603	HYPOTHETICAL 49.6 KD PROTEIN IN ELM	274	1.0e-22
Contig591	None		
Contig583	heat-shock protein	239	4.6e-39
Contig582	None		
Contig581	None		
Contig574	None		
Contig569	None		
Contig563	None		
Contig560	yeast Chaperonin hsp78 homolog	623	4.1e-60
Contig536	None		
Contig534	None		
Contig511	VU91D calmodulin	97	0.00090
Contig510	None		
Contig505	30 KD HEAT SHOCK PROTEIN	640	7.2e-62
Contig479	None		
Contig472	Glycogen phosphorylase; Gph1p	451	7.0e-41
Contig444	None		
Contig431	VACUOLAR PROTEASE A PRECURSOR	641	5.5e-62
Contig423	None		
Contig413	None		
Contig401	None		
Contig395	Ern4p	143	2.4e-07
Contig392	None		
Contig324	None		
Contig310	2-HYDROXYACID DEHYDROGENASE HOMOLOG	195	2.3e-14
Contig301	HYPOTHETICAL 54.3 KD PROTEIN C23D3.03C IN CHROMOSOME I	266	1.1e-21
Contig238	fdd123b [Coriolus versicolor]	147	2.5e-08
Contig219	None		
Contig198	None		
Contig191	cytochrome P450 monooxygenase	188	6.4e-14

Contig185	probable membraneprotein YOR161c - yeast	387	4.5e-35
Contig180	None		
Contig152	None		
Contig149	HYPOTHETICAL 56.8 KD PROTEIN IN SCJ	454	2.4e-42
Contig100	catalase-peroxidase	672	2.8e-65
Contig71	None		
Contig46	ATP-specific succinyl-CoA synthetase beta subunit	431	9.2e-40
Contig28	None		
Contig20	none		

Appendix VII. Genes only expressed in the NM cDNA library that contain both 3' and 5' ESTs

	Gene Name	HSP Score	P value
Contig1434	psi protein [Schizosaccharomyces pombe]/ DNAJ-like protein homolog	314	2.7e-2
Contig1402	similar to dehydrogenase [Caenorhabditis elegans]	284	3.7e-24
Contig1392	cyclophilin [Trichophyton mentagrophytes]	669	6.3e-65
Contig1390	keratin 2 epidermis [Mus musculus]	126	0.00039
Contig1388	None		
Contig1386	none		
Contig1366	None		
Contig1362	putative endo alpha-1,4 polygalactosaminidase	217	5.0e-17
Contig1360	None		
Contig1357	95 kDa structural and functional homolog ofvertebrate karyopherin b	193	3.3e-13
Contig1356	827 1.2e-81clock-controlled gene-8 protein [Neurospora crassa]	827	1.2e-81
Contig1355	None		
Contig1349	None		
Contig1348	None		
Contig1341	putative 50s ribosomal protein l14[Schizosaccharomyces pombe]	282	6.5e-24
Contig1318	manganese superoxide dismutase precursor[Neurospora crassa]	1107	2.4e-111
Contig1302	None		
Contig1288	None		
Contig1277	None		
Contig1276	None		
Contig1270	None		
Contig1266	putative Golgi membrane protein[Schizosaccharomyces pombe]	175	1.4e-12
Contig1263	None		
Contig1255	TYROSINE DECARBOXYLASE 3	135	2.2e-07
Contig1249	None		
Contig1243	None		
Contig1242	None		
Contig1226	putative succinate dehydrogenase membrane anchorsubunit precursor	203	1.4e-15
Contig1225	None		
Contig1217	None		
Contig1211	None		
Contig1203	None		
Contig1202	None		
Contig1200	HYPOTHETICAL 17.7 KD PROTEIN IN RNR	250	1.6e-20
Contig1197	None		
Contig1190	None		
Contig1188	None		
Contig1187	PEROXISOMAL MEMBRANE PROTEIN PMP20	300	7.8e-26
Contig1182	None		
Contig1174	None		
Contig1173	None		
Contig1169	None		
Contig1152	None		
Contig1125	None		

Contig1114	None		
Contig1111	None		
Contig1104	putative progesterone-binding proteinhomolog [Arabidopsis thaliana]	191	2.3e-14
Contig1084	None		
Contig1074	PROBABLE ARP2/3 COMPLEX 34 KD SUBUNIT (P34-ARC)	491	4.6e-46
Contig1071	None		
Contig1070	None		
Contig1066	putative cell wall protein [Emericella nidulans]	302	3.8e-26
Contig1056	Similar to phosphoglycerate mutase; coded for by C.elegans cDNA yk357d11.5; cod	338	1.7e-29
Contig1053	None		
Contig1030	None		
Contig1006	None		
Contig1000	None		
Contig990	NITRITE REDUCTASE Neurospora crassa	442	8.3e-62
Contig987	None		
Contig979	None		
Contig978	sugar transporter like protein [Arabidopsis thaliana]	176	7.2e-12
Contig964	F1E22.17 [Arabidopsis thaliana]	202	1.5e-15
Contig963	glucan 1,4-alpha-glucosidase (EC 3.2.1.3) precursor - Neurosporacrassa	757	2.3e-74
Contig960	None		
Contig958	RING zinc finger protein [Mus musculus]	177	2.7e-10
Contig954	hypothetical protein [Schizosaccharomyces pombe]	99	0.095
Contig931	Hypothetical Protein [Chlamydia pneumoniae]	128	8.9e-05
Contig918	CATALASE A >gi|1843578 (U87630) catalase [Aspergillusfumigatus]	736	3.4e-81
Contig901	NEUTRAL TREHALASE (ALPHA,ALPHA-TREHALASE) (ALPHA,ALPHA-TREHALOSE GLUCOHYDROLASE)	272	8.2e-22
Contig900	None		
Contig893	None		
Contig860	None		
Contig834	None		
Contig792	None		
Contig750	NAD-dependent isocitrate dehydrogenase subunit 2[Kluyveromyces lactis]	655	1.9e-63
Contig721	None		
Contig719	None		
Contig690	ubiquitin precursor - Neurospora crassa	765	4.1e-75
Contig669	None		
Contig631	ATP citrate lyase [Sordaria macrospora]	1267	2.6e-128
Contig609	None		
Contig608	None		
Contig602	RNA helicase; Sub2p >sp|Q07478|HE47_YEASTPROBABLE ATP-DEPENDENT RNA	479	8.1e-45
Contig600	None		
Contig523	none		
Contig443	None		
Contig437	None		
Contig385	TRANSCRIPTION FACTOR BTF3 HOMOLOG	391	1.7e-35
Contig362	None		

Contig350	None		
Contig277	probable membraneprotein YDR262w	124	0.00010
Contig148	NADH-UBIQUINONE OXIDOREDUCTASE 14.8 KD SUBUNIT (COMPLEXI-14.8KD) (CI-14.8KD)	653	3.4e-63

Appendix VIII. Genes expressed in the NE cDNA library that contain 3' EST only

	Gene Name	HSP Score	P Value
Contig1319	None		
Contig1284	PHOSPHOGLYCERATE KINASE	468	1.1e-43
Contig1267	60S ACIDIC RIBOSOMAL PROTEIN P0 (L10E)	176	2.0e-12
Contig1189	None		
Contig1167	None		
Contig1124	conserved hypothetical protein	150	5.3e-09
Contig1123	cytochrome P450 monooxygenase	140	1.8e-08
Contig1117	None		
Contig1108	None		
Contig1107	None		
Contig1082	None		
Contig1081	40S RIBOSOMAL PROTEIN S26E (CRP5)(13.6 KD RIBOSOMALPROTEIN)	287	I.9e-24
Contig1051	None		
Contig1050	ORF_o863; overlaps o231, other starts possible	311	7.2e-26
Contig1048	None		
Contig1036	None		
Contig1029	HISTIDINE BIOSYNTHESIS TRIFUNCTIONAL PROTEIN	750	1.5e-73
Contig1025	None		
Contig1021	extragenic suppressor of the bimD6 mutation	163	2.3e-10
Contig1018	probable membraneprotein YPR194c	221	1.0e-15
Contig1015	None		
Contig1002	ccg-4 putative polypeptide 1;	381	2.2e-34
Contig999	aconitase	453	2.3e-41
Contig996	None		
Contig991	lysyl-trna synthetase	54	2.1e-09
Contig988	HYPOTHETICAL 30.5 KD PROTEIN C4D7.04C IN CHROMOSOME I	186	9.7e-14
Contig968	SULFATE PERMEASE II	365	8.6e-32
Contig962	None		
Contig957	ATP SYNTHASE PROTEIN 9, MITOCHONDRIAL PRECURSOR (LIPID-BINDING PROTEIN	196	8.8e-15
Contig953	PYROA	223	1.1e-17
Contig952	none		
Contig949	flavohemoglobin	159	4.4e-10
Contig941	None		
Contig934	None		
Contig919	60S RIBOSOMAL PROTEIN L9-A	236	5.0e-19
Contig917	60S RIBOSOMAL PROTEIN L7-C	478	9.1e-45
Contig880	None		
Contig868	DNA REPAIR PROTEIN MMS21	108	0.00013
Contig866	CYANIDE HYDRATASE	351	3.3e-31
Contig863	HYPOTHETICAL 35.8 KD PROTEIN C12G12.12 IN CHROMOSOME I	284	2.9e-33
Contig858	conserved hypothetical protein	124	7.1e-07
Contig857	None		

Contig847	None		
Contig844	hypothetical protein	165	3.3e-11
Contig833	None		
Contig812	26S proteasome regulatory subunit	330	5.2e-29
Contig810	ISOLEUCYL-TRNA SYNTHETASE, CYTOPLASMIC (ISOLEUCINE--TRNALIGASE) (ILERS)	145	4.8e-08
Contig804	none		
Contig802	60S ribosomal protein L20B (L18B); Rpl20bp	438	1.6e-40
Contig796	dihydrolipoamide succinyltransferase	688	6.4e-67
Contig782	None		
Contig781	PYROA	222	1.5e-17
Contig779	MALATE SYNTHASE,GLYOXYSOMAL/malate synthase(EC 4.1.3.2)	293	1.8e-24
Contig776	Ribosomal protein L31B (L34B) (YL28);Rpl31bp	236	4.4e-19
Contig774	None		
Contig773	None		
Contig771	PROBABLE NADH-UBIQUINONE OXIDOREDUCTASE 19.3 KD SUBUNITPRECURSOR (COMPLEX I-19.3KD)	306	1.7e-26
Contig770	SACCHAROPINE DEHYDROGENASE [NAD+, L-LYSINE FORMING] (LYSINE--2-OXOGLUTARATE REDUCTASE)	262	9.0e-22
Contig765	Ribosomal protein L43B; Rpl43bp	355	1.1e-31
Contig763	None		
Contig749	(putative) mitochondrial carrier protein;Yhm1p	201	2.6e-15
Contig746	ACYL CARRIER PROTEIN, MITOCHONDRIAL PRECURSOR (ACP)	106	2.8e-05
Contig742	ATP SYNTHASE SUBUNIT 4, MITOCHONDRIAL PRECURSOR	179	5.1e-13
Contig741	PROTEIN KINASE C-LIKE/proteinkinase C homologue	356	1.8e-30
Contig735	None		
Contig725	ATP synthase subunit 5; oligomycinsensitivity-conferring protein;	153	2.6e-10
Contig712	VACUOLAR ATP SYNTHASE SUBUNIT AC39 (V-ATPASE AC39SUBUNIT) (V-ATPASE 41 KD SUBUNIT	190	1.1e-13
Contig706	GTP-BINDING NUCLEAR PROTEIN SPI1/spilhypothetical protein	146	3.7e-08
Contig703	None		
Contig692	60S RIBOSOMAL PROTEIN L11	157	1.0e-10
Contig688	putative nadh-cytochrome b5 reductase	198	4.7e-15
Contig678	None		
Contig673	GTP-BINDING NUCLEAR PROTEIN SPI1/spilhypothetical protein	286	2.6e-24
Contig667	None		
Contig666	PROBABLE PROLINE OXIDASE, MITOCHONDRIAL PRECURSOR (PROLINE DEHYDROGENASE)	98	0.27
Contig663	None		
Contig660	None		
Contig654	acyl-CoA dehydrogenase, putative	265	6.9e-22
Contig650	None		
Contig639	None		
Contig636	Uridine kinase; Urk1p	126	1.9e-06
Contig634	GLUCOSE-REPRESSIBLE GENE PROTEIN	352	2.6e-31
Contig630	None		

Contig626	rna-binding protein	95	0.041
Contig625	None		
Contig619	60S RIBOSOMAL PROTEIN L27-A	113	5.3e-06
Contig618	None		
Contig617	None		
Contig612	similar to glycosyl transferases (Pfam:PF00535,Score=80.8, E=2.9e-20, N=1)	274	4.6e-23
Contig599	None		
Contig595	None		
Contig584	None		
Contig575	None		
Contig566	None		
Contig565	hypothetical 70K protein-Neurospora crassa mitochondrion	786	2.5e-77
Contig564	Copper Transporter; Ctr3p	134	2.2e-06
Contig559	acetate kinase (ackA)	237	9.9e-19
Contig557	*Coproporphyrinogen III oxidase; Hem13p*	349	4.5e-31
Contig556	None		
Contig538	None		
Contig535	None		
Contig533	None		
Contig522	None		
Contig521	None		
Contig518	None		
Contig515	None		
Contig512	PUTATIVE125.2 KD MEMBRANE GLYCOPROTEIN IN BIO3-HXT17 INTERGENIC REGION	275	8.3e-22
Contig508	None		
Contig507	hypothetical proline-rich protein; possiblecoiled-coil region	109	0.0026
Contig503	None		
Contig502	None		
Contig497	None		
Contig496	fumarylacetoacetate hydrolase	344	1.8e-30
Contig490	None		
Contig488	None		
Contig487	None		
Contig485	none		
Contig477	putative vacuolar h(+)-atpase subunit	225	3.3e-17
Contig467	chitin synthase 3	221	3.2e-16
Contig466	None		
Contig442	None		
Contig438	60S RIBOSOMAL PROTEIN L1-B (L10A)	371	2.3e-33
Contig436	None		
Contig435	hypothetical protein	252	1.6e-20
Contig433	None		
Contig432	none		
Contig429	None		

Contig427	histone H4 (clone H4g)	99	0.00059
Contig425	None		
Contig418	None		
Contig417	None		
Contig416	Ribosomal protein L31B (L34B) (YL28);Rpl31bp	339	5.4e-30
Contig415	None		
Contig410	2-oxoglutarate/malate translocator (clone OMT103), mitochondrialmembrane - proso millet	179	9.1e-13
Contig404	None		
Contig403	40S RIBOSOMAL PROTEIN S16	222	1.3e-17
Contig399	None		
Contig398	sconCp	130	8.3e-08
Contig390	None		
Contig387	None		
Contig384	IucB	261	1.1e-21
Contig382	none		
Contig380	60S RIBOSOMAL PROTEIN L17 (L23) (AMINO ACIDSTARVATION-INDUCED PROTEIN) (ASI)	222	1.2e-17
Contig370	None		
Contig366	None		
Contig365	None		
Contig358	40S RIBOSOMAL PROTEIN S6	307	1.3e-26
Contig356	None		
Contig347	DNA replication licensing factor	272	1.1e-21
Contig346	proteasome component Y7; Pre8p	370	2.9e-33
Contig345	None		
Contig343	None		
Contig338	None		
Contig334	None		
Contig333	CYANIDE HYDRATASE (FORMAMIDE HYDROLYASE)	289	1.2e-24
Contig331	None		
Contig330	ntranuclear protein which exhibits a nucleotide-specific intron-dependent tRNA pseudouridine synthase activity; Pus1p	227	2.7e-17
Contig325	None		
Contig323	None		
Contig322	2-OXOISOVALERATE DEHYDROGENASE ALPHA SUBUNIT, MITOCHONDRIAL PRECURSOR	331	4.1e-29
Contig320	ALTERNATIVE OXIDASE PRECURSOR (ALTOX)	528	5.3e-50
Contig313	None		
Contig307	HYPOTHETICAL 54.5 KD PROTEIN IN CBF2-SKN1 INTERGENIC REGION	125	5.1e-05
Contig304	None		
Contig303	None		
Contig297	None		
Contig295	None		
Contig289	None		
Contig287	None		
Contig280	None		

Contig279	None		
Contig278	Ribosomal protein L23A (L17aA) (YL32);Rpl23ap	340	4.5e-30
Contig276	40S RIBOSOMAL PROTEIN S15 (S12)	109	1.3e-05
Contig275	None		
Contig272	None		
Contig268	None		
Contig261	triosephosphate isomerase + glyceraldehyde-3-phosphatedehydrogenase	433	5.8e-40
Contig260	None		
Contig256	None		
Contig253	None		
Contig251	MITOCHONDRIAL PROCESSING PEPTIDASE ALPHA SUBUNITPRECURSOR (ALPHA-MPP)	415	5.1e-38
Contig250	glutaminase A	265	3.3e-21
Contig245	transketolase I	157	1.5e-09
Contig244	None		
Contig243	None		
Contig241	None		
Contig239	None		
Contig237	None		
Contig236	DELTA(24)-STEROL C-METHYLTRANSFERASE	306	1.3e-26
Contig235	CALNEXIN HOMOLOG PRECURSOR	257	1.6e-20
Contig230	None		
Contig229	None		
Contig228	None		
Contig224	None		
Contig220	None		
Contig218	fumarase	391	1.6e-35
Contig217	None		
Contig209	40S RIBOSOMAL PROTEIN S26E (CRP5) (13.6 KD RIBOSOMALPROTEIN)	129	1.1e-07
Contig207	Involved in autophagocytosis.; Aut1p	138	1.6e-06
Contig206	None		
Contig197	None		
Contig196	None		
Contig194	None		
Contig193	TUBULIN ALPHA CHAIN	455	2.8e-42
Contig189	None		
Contig188	None		
Contig186	None		
Contig184	None		
Contig17	PROBABLE PROLINE OXIDASE, MITOCHONDRIAL PRECURSOR (PROLINE DEHYDROGENASE)	98	0.0025
Contig175	None		
Contig172	activator of Hsp70 and Hsp90 chaperones	113	7.1e-05
Contig164	None		
Contig161	NADH-UBIQUINONE OXIDOREDUCTASE B22 SUBUNIT (COMPLEXI-B22) (CI-B22)	129	1.6e-06
Contig156	INITIATION FACTOR 5A (EIF-5A) (EIF-4D)	690	3.5e-67

Contig154	hypothetical proteinYOL057w - yeast	246	5.0e-19
Contig153	VACUOLAR ATP SYNTHASE SUBUNIT G (V-ATPASE 13 KD SUBUNIT) (VACUOLAR H(+)-ATPASE SUBUNIT G)	159	1.7e-19
Contig144	None		
Contig141	None		
Contig131	None		
Contig127	None		
Contig126	None		
Contig12	oxidoreductase, short-chaindehydrogenase/reductase family	200	2.8e-13
Contig117	rds1 protein - fission yeast	222	4.6e-17
Contig107	None		
Contig102	None		
Contig97	None		
Contig95	40S RIBOSOMAL PROTEIN S3AE (S1)	280	9.4e-24
Contig91	Peroxisomal membrane protein; Pex11p	188	1.1e-13
Contig78	100 kDa protein	432	7.5e-39
Contig70	erythrocyte membrane antigen 1	166	7.6e-11
Contig68	None		
Contig67	trk-1	249	4.0e-19
Contig66	None		
Contig65	None		
Contig64	None		
Contig63	None		
Contig58	none		
Contig50	None		
Contig44	None		
Contig43	None		
Contig41	None		
Contig38	4-AMINOBUTYRATE AMINOTRANSFERASE (GAMMA-AMINO-N-BUTYRATETRANSAMINASE) (GABA TRANSAMINASE)	615	3.5e-59
Contig36	None		
Contig32	None		
Contig22	None		
Contig13	None		
Contig11	putative 20kDa subunit of the V-ATPase	348	7.0e-31
Contig8	None		
Contig7	thiazole biosynthetic enzyme	141	2.3e-08
Contig5	HISTONE H3	147	4.4e-09
Contig4	None		
Contig3	None		

Appendix IX. Genes only expressed in the NE cDNA library and containing 5' ESTs only

	GENE NAME	HSP Score	P Value
Contig1337	None		
Contig1283	ALCOHOL DEHYDROGENASE I	1044	9.3e-105
Contig1229	None		
Contig1220	None		
Contig1208	NMT1 PROTEIN HOMOLOG	1034	1.3e-103
Contig1194	None		
Contig1140	MALATE SYNTHASE, GLYOXYSOMA	542	1.4e-51
Contig1127	None		
Contig1119	CYANIDE HYDRATASE (FORMAMIDE HYDROLYASE)	654	2.4e-63
Contig1110	WD repeat protein; human U5 SNRNP-specific-like	329	6.7e-29
Contig1086	putative protein	121	1.7e-05
Contig1069	Mok13; Mok13 is homologous to Mok1 which is analpha-glucan synthase	222	9.2e-16
Contig1063	malate dehydrogenase	487	1.1e-45
Contig1061	Vip1 protein	249	1.9e-20
Contig1047	Induced in stationary phase; Sno2p	234	7.1e-19
Contig1045	LYSYL-TRNA SYNTHETASE (LYSINE--TRNA LIGASE) (LYSRS	294	1.5e-24
Contig1040	calmodulin - moss (Physcomitrella patens)	283	4.5e-24
Contig1039	None		
Contig1022	None		
Contig1017	None		
Contig1011	dihydrolipoamide succinyltransferase	375	8.3e-34
Contig1009	aconitas	928	2.0e-92
Contig1007	HISTIDINE BIOSYNTHESIS TRIFUNCTIONAL PROTEIN	788	1.3e-77
Contig981	2-isopropylmalate synthase	359	1.3e-31
Contig974	ACYL CARRIER PROTEIN, MITOCHONDRIAL PRECURSOR (ACP)	489	6.2e-46
Contig970	(AL132949) predicted using Genefinder	119	2.9e-06
Contig967	nuclear GTPase-activating protein for Ran;Yrb1p	185	1.0e-13
Contig924	similar to S. cerevisiae RER1	265	3.5e-22
Contig914	alpha-ketoglutarate dehydrogenase; Kgd1p	606	1.7e-57
Contig910	THIAZOLE BIOSYNTHETIC ENZYME PRECURSOR (STRESS-INDUCIBLEPROTEIN STI35)	330	5.3e-29
Contig908	mitochondrial carrier protein	148	3.0e-09
Contig899	EUKARYOTIC INITIATION FACTOR 4A (EIF-4A)	227	1.1e-17
Contig891	None		
Contig886	None		
Contig878	None		
Contig852	None		
Contig848	None		
Contig840	Adenylate kinase (mitochondrial GTP:AMPphosphotransferase); Adk2p	208	4.6e-16
Contig838	stearoyl-CoA desaturase (EC 1.14.99.5)	287	4.4e-24
Contig836	TROPOMYOSIN	182	2.5e-13
Contig830	similar to Saccharomyces serevisiae	614	3.2e-59

	hypothetical 52.9KDprotein in CDC26-YMR31 int		
Contig824	40S RIBOSOMAL PROTEIN S3AE (S1)	698	4.4e-68
Contig821	putative nadh-cytochrome b5 reductase	285	2.7e-24
Contig820	None		
Contig815	None		
Contig813	26S proteasome regulatory subunit	164	1.2e-10
Contig794	PROBABLE PROLINE OXIDASE, MITOCHONDRIAL PRECURSOR (PROLINE DEHYDROGENASE)	223	9.5e-15
Contig784	None		
Contig780	hypothetical protein 2 (cpc-1 5' region	154	2.3e-10
Contig777	ADP,ATP CARRIER PROTEIN (ADP/ATP TRANSLOCASE) (ADENINENUCLEOTIDE TRANSLOCATOR) (ANT)	698	4.1e-68
Contig764	CURVED DNA-BINDING PROTEIN (42 KD PROTEIN)	231	2.5e-18
Contig761	none		
Contig755	salivary gland growth factor-1 precursor	101	0.00084
Contig730	small zinc finger-like protein	237	3.1e-19
Contig722	Ytal2p/ITOCHONDRIALRESPIRATORY CHAIN COMPLEX	514	6.1e-48
Contig698	ADP-RIBOSYLATION FACTOR	570	1.8e-54
Contig694	Optlp	291	8.5e-24
Contig691	None		
Contig679	None		
Contig664	PROBABLE NADH-UBIQUINONE OXIDOREDUCTASE 19.3 KDA SUBUNIT PRECURSOR(COMPLEX I-19.3KD) (CI-19.3KD)	555	6.6e-53
Contig647	Orf2	188	5.7e-14
Contig645	NAD(+)-specific glutamate dehydrogenase, NAD-GDH (EC1.4.1.2)	979	9.1e-98
Contig644	hypothetical 70K protein - Neurospora crassa mitochondrion (SGC3)	483	3.0e-45
Contig637	none		
Contig628	60S RIBOSOMAL PROTEIN L7-C	304	2.3e-26
Contig624	None		
Contig620	None		
Contig616	SACCHAROPINE DEHYDROGENASE [NAD+, L-LYSINE FORMING] (LYSINE--2-OXOGLUTARATE REDUCTASE)	830	4.8e-82
Contig613	None		
Contig611	None		
Contig610	putative endocytosis and cytoskeleton protein	212	9.7e-16
Contig607	alpha-glucosidase AgdA	640	2.8e-61
Contig604	subunit of signal peptidase complex, homologous to mammalian protein	131	5.7e-08
Contig597	None		
Contig596	PUTATIVE DISULFIDE ISOMERASE ERP38 PRECURSOR	378	3.9e-34
Contig593	None		
Contig578	CYTOPLASMIC ALPHA-AMYLASE (1,4-ALPHA-D-GLUCANGLUCANOHYDROLASE)	131	5.7e-07
Contig567	None		
Contig558	hypothetical protein	378	3.3e-34

Contig554	HISTIDINE BIOSYNTHESIS TRIFUNCTIONAL PROTEIN [INCLUDES:PHOSPHORIBOSYL-AMP CYCLOHYDROLASE ; PHOSPHORIBOSYL-ATP PYROPHOSPHOHYDROLASE ; HISTIDINOL DEHYDROGENASE (HDH)]	681	3.1e-66
Contig553	None		
Contig552	40S RIBOSOMAL PROTEIN S4 (S7)	742	8.6e-73
Contig550	extracellular protease synthesized in a-cells that cleaves and inactivates alpha factor; Barlp	327	4.9e-28
Contig549	None		
Contig548	None		
Contig546	hypothetical protein	255	4.9e-21
Contig545	SERINE HYDROXYMETHYLTRANSFERASE, CYTOSOLIC (SERINE METHYLASE) (GLYCINE HYDROXYMETHYLTRANSFERASE) (SHMT)	723	8.5e-71
Contig543	putative serine palmitoyltransferase	292	1.9e-24
Contig532	MBF1	139	8.7e-09
Contig530	None		
Contig529	clock-controlled gene-9 protein	835	1.5e-82
Contig528	Yol057wp/hypothetical proteinYOL057w	290	8.6e-24
Contig524	GAMMA-ADAPTIN (GOLGI ADAPTOR HA1/AP1 ADAPTIN GAMMA SUBUNIT) (CLATHRIN ASSEMBLY PROTEIN COMPLEX 1 GAMMA LARGE CHAIN) (GAMMA-ADA)	430	1.1e-38
Contig514	THIAZOLE BIOSYNTHETIC ENZYME PRECURSOR	267	2.3e-22
Contig509	None		
Contig495	None		
Contig489	None		
Contig486	oligomycin sensitivity conferring protein	277	1.6e-23
Contig484	chitin synthase 3 [Neurospora crassa]	1106	2.8e-111
Contig483	F1E22.10	101	0.0015
Contig482	NMT1 PROTEIN HOMOLOG	185	3.7e-13
Contig481	none		
Contig478	predicted using Genefinder; cDNA EST yk524f8.5comes from this gene	162	5.0e-11
Contig475	GUANINE NUCLEOTIDE-BINDING PROTEIN BETA SUBUNIT-LIKE PROTEIN(CROSS-PATHWAY CONTROL WD-REPEAT PROTEIN CPC-2)	1398	2.9e-142
Contig474	None		
Contig469	None		
Contig462	None		
Contig458	MITOCHONDRIAL RIBOSOMAL PROTEIN S5	642	3.9e-62
Contig456	heat-regulated protein; Higlp	317	1.1e-27
Contig454	None		
Contig450	ALDEHYDE DEHYDROGENASE (ALDDH) (ALLERGEN ALT A 10) (ALT AX)	327	1.4e-28
Contig447	Mitochondrial ribosomal protein MRP7 (YmL2)(E. coli L27); Mrp7p	199	8.4e-15
Contig445	None		
Contig440	septin B	633	4.0e-61
Contig434	hypothetical protein	316	1.5e-27
Contig428	4-AMINOBUTYRATE AMINOTRANSFERASE (		

	GAMMA-AMINO-N-BUTYRATETRANSAMINASE) (GABA TRANSAMINASE)	424	5.2e-39
Contig426	OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN	764	4.3e-75
Contig421	None		
Contig412	clathrin-associated protein complex, smallsubunit; Aps1p	169	5.8e-12
Contig409	None		
Contig407	dolichol-phosphate-mannose synthase	423	6.4e-39
Contig406	None		
Contig400	None		
Contig391	None		
Contig388	triose phosphate isomerase	409	1.6e-37
Contig383	None		
Contig379	Involved in autophagocytosis.; Aut1p	223	1.1e-17
Contig378	40S RIBOSOMAL PROTEIN S6	603	4.3e-58
Contig375	none		
Contig373	conserved hypothetical protein	154	2.2e-10
Contig369	40S RIBOSOMAL PROTEIN S5	364	1.2e-32
Contig363	stearoyl-CoA desaturase (EC 1.14.99.5)	797	1.3e-78
Contig360	None		
Contig359	HISTONE H3	140	7.3e-09
Contig355	None		
Contig354	putative amino acid permease	109	0.00015
Contig344	Yjr078wp/HYPOTHETICAL 50.8 KD PROTEIN IN MIR	200	1.7e-14
Contig342	None		
Contig341	YGL010w-like protein	138	1.2e-08
Contig340	dolichyl-phosphatebeta-glucosyltransferase	168	3.1e-11
Contig339	MITOCHONDRIAL PROCESSING PEPTIDASE BETA SUBUNIT PRECURSOR (BETA-MPP)(UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX CORE PROTEIN I)	752	8.7e-74
Contig337	EUKARYOTIC INITIATION FACTOR 4A (EIF-4A)	313	3.1e-27
Contig336	DNA replication licensing factor	674	1.4e-65
Contig335	YPT1-RELATED PROTEIN 5	389	2.9e-35
Contig332	HET-C protein	184	2.0e-12
Contig329	DELTA(24)-STEROL C-METHYLTRANSFERASE	316	1.5e-27
Contig327	branched-chain alpha keto-acid dehydrogenase E1 alphasubunit	314	3.9e-27
Contig326	None		
Contig321	Similarity to Yeast endosomal P24A protein	270	9.7e-22
Contig319	PSEUDOURIDYLATE SYNTHASE 2 (PSEUDOURIDINE SYNTHASE 2)	117	9.2e-06
Contig314	Orm1p	338	7.3e-30
Contig312	ALTERNATIVE OXIDASE PRECURSOR (ALTOX)	778	2.0e-76
Contig311	Ygr200cp/HYPOTHETICAL 89.4 KD TRP-ASP REPEAT	295	3.1e-24
Contig309	Zuo1p/ZUOTIN	402	8.6e-37
Contig308	proteolipid protein of the proton ATPase;Ppa1p	332	3.0e-29
Contig306	40S RIBOSOMAL PROTEIN S9 (S7)	266	2.7e-22
Contig305	PROBABLE METHYLMALONATE-SEMIALDEHYDE DEHYDROGENASE[ACYLATING] PRECURSOR (MMSDH)	553	9.9e-53
Contig300	spermidine synthase	706	6.5e-69
Contig299	None		

Contig298	CARBOXYPEPTIDASE Y PRECURSOR (CARBOXYPEPTIDASE YSCY)	594	5.5e-57
Contig296	PROBABLE CHITIN BIOSYNTHESIS PROTEIN C6G9.12 (CHS5HOMOLOG)	489	6.8e-46
Contig290	putative protein	231	1.8e-18
Contig288	None		
Contig285	None		
Contig284	mitochondrial 60s ribosomal protein L2	161	1.3e-10
Contig281	ISOCITRATE DEHYDROGENASE [NADP], MITOCHONDRIAL PRECURSOR(OXALOSUCCINATE DECARBOXYLASE) (IDH) (NADP+-SPECIFIC ICDH) (IDP)	665	1.3e-64
Contig273	GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (GAPDH)(CLOCK-CONTROLLED PROTEIN 7)	427	2.5e-39
Contig271	CENTROMERE/MICROTUBULE BINDING PROTEIN CBF5 (CENTROMERE-BINDING FACTOR 5) (NUCLEOLAR PROTEIN CBF5)	659	6.5e-64
Contig269	ubiquitin-conjugating-enzyme-like protein	149	7.9e-10
Contig265	None		
Contig263	putative syntaxin	355	1.2e-31
Contig259	None		
Contig257	Ribosomal protein L8B (L4B) (rp6) (YL5);Rpl8bp	509	4.2e-48
Contig252	57 kDa nucleolar protein; Nop58p	568	2.7e-54
Contig249	Ribosomal protein L31B (L34B) (YL28);Rpl31bp	166	1.2e-11
Contig248	transketolase	633	3.8e-61
Contig247	hypothetical protein	257	2.3e-21
Contig246	None		
Contig242	None		
Contig240	None		
Contig233	None		
Contig232	None		
Contig222	40S RIBOSOMAL PROTEIN S13 (S15	622	5.2e-60
Contig221	phosphatidylethanolamine methyltransferase	414	6.5e-37
Contig214	2-oxoglutarate/malate translocator (clone OMT103), mitochondrialmembrane - proso millet	208	8.2e-29
Contig213	ALIPHATIC NITRILASE	182	7.5e-13
Contig210	None		
Contig202	Similar to gb|U90212 DNA binding protein ACBF from Nicotiana tabacum and contains 3 PF|00076 RNA recognition motif domains. ESTs gb|T44278, gb|R65195, gb|N65904, gb|H37499, gb|R90487, gb|N95952, gb|T44278, gb|Z20166, gb|N96891, gb|W43137, gb|F15504, gb|F15495 and gb|Z30868 come from this gene. [Arabidopsis thaliana	161	2.2e-10
Contig201	None		
Contig200	VACUOLAR ATP SYNTHASE 16 KD PROTEOLIPID SUBUNIT	514	1.6e-48
Contig199	putative GTP cyclohydrolase	194	3.0e-13
Contig192	None		
Contig187	None		
Contig179	None		
Contig177	None		

Contig176	ATP-binding protein (CDC48/PAS1/SEC18 family)	125	2.1e-06
Contig173	None		
Contig171	histone H1	222	9.7e-18
Contig166	Ribosomal protein L29 (YL43); Rpl29p	255	4.4e-21
Contig165	None		
Contig163	None		
Contig162	KIAA1223 protein	133	7.6e-07
Contig155	J1590	378	1.6e-33
Contig146	None		
Contig143	PROBABLE SUCCINYL-COA LIGASE [GDP-FORMING] ALPHA-CHAIN, MITOCHONDRIAL PRECURSOR (SUCCINYL-COA SYNTHETASE, ALPHA CHAIN) (SCS-ALPHA)	687	7.2e-67
Contig142	None		
Contig130	None		
Contig125	None		
Contig123	HYPOTHETICAL OXIDOREDUCTASE C4H3.08 IN CHROMOSOME I	456	2.4e-42
Contig114	None		
Contig106	SERINE-TYPE CARBOXYPEPTIDASE F PRECURSOR (PROTEINASE F)(CPD-II)	386	6.2e-35
Contig103	None		
Contig101	None		
Contig99	None		
Contig92	None		
Contig90	HEAT SHOCK PROTEIN 90 HOMOLOG (SUPPRESSOR OF VEGETATIVEINCOMPATIBILITY MOD-E)	500	4.9e-47
Contig89	Yjr085cp	168	7.0e-12
Contig88	GLUTATHIONE-DEPENDENT FORMALDEHYDE DEHYDROGENASE (FDH)(FALDH)	672	2.4e-65
Contig86	None		
Contig83	None		
Contig82	None		
Contig76	zinc metallo-protease that catalyzes the first step of N-terminal processing of the yeast a-factor precursor; Ste24p	307	1.5e-26
Contig73	SURFEIT LOCUS PROTEIN 4 HOMOLOG	314	2.5e-27
Contig72	None		
Contig69	Fat1p	195	1.6e-13
Contig62	None		
Contig61	None		
Contig60	None		
Contig52	Yil041wp /36.7 KDPROTEIN IN CBR5-NOT3 INTERG	255	4.6e-21
Contig42	hypothetical protein	168	1.3e-11
Contig37	ubiquitin fusion protei	277	2.1e-23
Contig34	Putative copper binding/homeostasis protein;Atx1p	124	3.4e-07
Contig33	none		
Contig29	high molecular mass nuclear antigen	96	0.010
Contig25	None		

Contig24	F3F19.15	140	7.5e-08
Contig21	elongation factor 1 beta	336	1.1e-29
Contig19	HYPOTHETICAL CALCIUM-BINDING PROTEIN C18B11.04 INCHROMOSOME I	137	1.6e-08
Contig14	phosphoribosyl aminoimidazolesuccinocarbozamide synthetase; Ade1p	525	1.0e-49
Contig12	Putative alpha-1,2-mannosyltransferase; Ktr4p	467	1.7e-43
Contig9	ubiquitin fusion protein	655	1.8e-63
Contig6	None		
Contig2	60S ribosomal protein L20B (L18B); Rpl20bp	239	2.3e-19

Appendix X. Genes expressed only in the NE cDNA library that contain both 3' ESTs and 5'ESTs

Contig1374	subunit e of mitochondrial F1F0-ATPase;Tim11p	133	3.7e-08
Contig1367	ALPHA CHAIN, MITOCHONDRIAL PRECURSOR	1064	7.2e-107
Contig1353	None		
Contig1342	None		
Contig1339	None		
Contig1338	Ribosomal protein S25B (S31B) (rp45)(YS23); Rps25bp	195	1.1e-14
Contig1317	calmodulin - Neurospora crassa	757	2.6e-74
Contig1316	ribosomal protein CRP7 [Neurospora crassa]	449	1.2e-41
Contig1315	None		
Contig1295	THIAZOLE BIOSYNTHETIC ENZYME PRECURSOR (STRESS-INDUCIBLEPROTEIN STI35)	1072	1.2e-107
Contig1294	ribosomal protein l37 homolog	387	4.7e-35
Contig1291	None		
Contig1269	ALDEHYDE DEHYDROGENASE (ALDDH) (ALLERGEN CLA H 3) (CLA HIII)	817	1.0e-80
Contig1265	40S RIBOSOMAL PROTEIN S5	791	6.6e-78
Contig1262	ccg-4 putative polypeptide 2	456	2.0e-42
Contig1258	60s ribosomal protein l21	549	2.6e-52
Contig1256	40S RIBOSOMAL PROTEIN S0 (RIBOSOME-ASSOCIATED PROTEIN 1)	163	5.7e-11
Contig1254	delta-1-pyrroline-5-carboxylate dehydrogenaseprecursor	764	5.6e-75
Contig1251	None		
Contig1247	None		
Contig1245	suppressor of Lec15 homolog (C.griseus)	159	7.0e-11
Contig1240	None		
Contig1238	rAsp f 7	148	1.0e-09
Contig1232	stearoyl-CoA desaturase (EC 1.14.99.5)	225	3.7e-17
Contig1231	Probable AMP-binding protein; Fat2p	802	4.1e-79
Contig1228	14-3-3 [Lentinula edodes]	796	1.9e-78
Contig1213	zuotin like; putative zdna binding; dnaj domaincontaining protein	227	5.3e-18
Contig1210	acetyl-coa acetyltransferase (EC 2.3.1.9)	560	1.9e-53
Contig1207	NMT1 PROTEIN HOMOLOG	368	5.2e-33
Contig1206	None		
Contig1204	mitochondrial malate dehydrogenase; Mdh1p	1087	2.4e-109
Contig1201	ubiquitin/S27a fusion protein	616	2.3e-59
Contig1196	None		
Contig1192	ribosomal protein S14.e	597	2.6e-57
Contig1191	Ribosomal protein L43B; Rpl43bp	362	2.0e-32
Contig1186	None		
Contig1175	thiamine synthase homolog	242	1.2e-19
Contig1165	clock-controlled gene-6 protein	145	2.1e-09
Contig1164	Yfr044cp/HYPOTHETICAL 52.9 KD PROTEIN IN SAP	322	4.9e-28
Contig1162	ADP,ATP CARRIER PROTEIN (ADP/ATP TRANSLOCASE) (ADENINENUCLEOTIDE TRANSLOCATOR) (ANT)	504	1.8e-47
Contig1161	PYROA	960	7.5e-96
Contig1155	Subunit VIIa of cytochrome c oxidase; Cox9p	130	7.1e-08

Contig1154	None		
Contig1153	None		
Contig1150	None		
Contig1148	40S RIBOSOMAL PROTEIN S3AE (S1)	166	9.2e-11
Contig1147	COFILIN	307	1.4e-26
Contig1146	None		
Contig1145	INITIATION FACTOR 5A (EIF-5A) (EIF-4D)	367	5.6e-33
Contig1143	nucleoside diphosphate kinase	799	7.3e-79
Contig1138	putative acetyltransferase	163	2.1e-11
Contig1137	None		
Contig1136	CGI-105 protein	507	8.2e-48
Contig1132	PROBABLE COATOMER ZETA SUBUNIT (ZETA-COAT PROTEIN)(ZETA-COP)	372	1.9e-33
Contig1122	clock-controlled gene-6 protein	130	8.4e-08
Contig1121	PRPD PROTEIN	676	8.3e-66
Contig1120	None		
Contig1118	None		
Contig1116	CONIDIATION-SPECIFIC PROTEIN 10	279	1.3e-23
Contig1113	None		
Contig1105	60S RIBOSOMAL PROTEIN L13	414	6.3e-38
Contig1103	60S RIBOSOMAL PROTEIN L7-C	380	2.5e-34
Contig1102	kinesin related protein 1	656	1.4e-63
Contig1101	40S RIBOSOMAL PROTEIN S19 (S16)	615	3.1e-59
Contig1098	CONIDIATION-SPECIFIC PROTEIN 6	486	1.5e-45
Contig1094	clock-controlled gene-9 protein	919	2.0e-91
Contig1093	HISTONE H4.1	409	1.9e-37
Contig1091	hypothetical protein	277	1.4e-32
Contig1087	60S RIBOSOMAL PROTEIN L11	675	1.2e-65
Contig1085	60S ribosomal protein L20B (L18B); Rpl20bp	647	1.2e-62
Contig1080	None		
Contig1079	None		
Contig1075	None		
Contig1072	2-oxoglutarate dehydrogenase e1 component	244	1.3e-18
Contig1068	None		
Contig1060	None		
Contig1059	None		
Contig1055	None		
Contig1054	None		
Contig1052	Ribosomal protein S4B (YS6) (rp5) (S7B);Rps4bp	276	2.7e-23
Contig1043	rho gdp dissociation inhibitor	153	3.0e-10
Contig1041	None		
Contig1037	L-AMINO ACID OXIDASE PRECURSOR (LAO)	803	3.7e-79
Contig1035	CYTOCHROME C	583	7.3e-56
Contig1034	NAD(+)-specific glutamate dehydrogenase, NAD-GDH (EC1.4.1.2)	645	1.2e-61
Contig1033	None		
Contig1032	None		
Contig1027	RNA polymerase II 4 (14 kDa subunit	214	3.9e-16
Contig1023	GLUTAREDOXIN	281	7.3e-24

Contig1019	PROBABLE CYTOCHROME C OXIDASE POLYPEPTIDE VIA PRECURSOR	214	8.7e-17
Contig1016	None		
Contig1012	hypothetical protein	428	1.5e-38
Contig1010	60S RIBOSOMAL PROTEIN L27-A	500	5.0e-47
Contig1008	Subunit of 20S proteasome; Pre3p	676	1.2e-65
Contig1003	ATP SYNTHASE SUBUNIT 4, MITOCHONDRIAL PRECURSOR	492	2.8e-46
Contig995	Iron-sulfur cluster nifU-like protein; Isu1p	538	4.1e-51
Contig994	None		
Contig989	None		
Contig985	None		
Contig984	ATP SYNTHASE BETA CHAIN, MITOCHONDRIAL PRECURSOR	987	1.1e-98
Contig983	None		
Contig977	Ribosomal protein S30A; Rps30ap	153	3.0e-10
Contig976	None		
Contig973	None		
Contig971	None		
Contig966	PROBABLE PROPIONATE KINASE/PduW(AF026270)	145	1.3e-08
Contig961	None		
Contig959	None		
Contig951	Ats1	246	3.3e-20
Contig950	None		
Contig942	None		
Contig939	putative type III alcohol dehydrogenase	285	3.0e-24
Contig938	23S rRNA intron 2 protein - Podospora anserina mitochondrion(SGC3)	166	8.8e-11
Contig937	None		
Contig935	NOne		
Contig932	membrane glycoprotein	102	0.22
Contig926	None		
Contig925	None		
Contig922	60S ribosomal protein L24	413	7.5e-38
Contig921	splicing factor	143	1.3e-08
Contig916	None		
Contig915	DNA-binding protein, mtDNA stabilizingprotein	477	1.4e-44
Contig913	None		
Contig912	None		
Contig907	40S RIBOSOMAL PROTEIN S13 (S15)	307	1.5e-26
Contig904	None		
Contig902	MALATE SYNTHASE, GLYOXYSOMAL	714	9.0e-70
Contig895	involved in vacuolar protein targeting;Vps28p	339	6.0e-30
Contig894	None		
Contig890	snRNP Sm protein F-like	248	2.6e-20
Contig884	None		
Contig883	translation elongation factor 2 (EF-2); Eft2p	700	3.0e-68
Contig882	ISOCITRATE LYASE (ISOCITRASE) (ISOCITRATASE) (ICL)	688	5.7e-67
Contig879	diphosphomevalonate decarboxylase	199	1.4e-14
Contig875	60S RIBOSOMAL PROTEIN L28 (L27A) (L29) (CRP1)	796	2.0e-78
Contig870	None		

Contig869	60s ribosomal protein l10	902	1.2e-89
Contig867	cytochrome P450 monooxygenase	101	0.033
Contig865	putative GTP cyclohydrolase	773	6.3e-76
Contig864	None		
Contig862	HYPOTHETICAL 35.8 KD PROTEIN C12G12.12 IN CHROMOSOME I	284	2.9e-33
Contig859	Ribosomal protein S18A; Rps18ap	586	4.0e-56
Contig855	None		
Contig854	60S RIBOSOMAL PROTEIN L44 (L41)	491	4.1e-46
Contig849	None		
Contig845	None		
Contig842	ENDOLYSIN (LYSIS PROTEIN) (LYSOZYME)	507	1.0e-47
Contig835	sid3	119	5.8e-05
Contig831	CURVED DNA-BINDING PROTEIN (42 KD PROTEIN)	226	1.0e-17
Contig828	None		
Contig827	GDP dissociation inhibitor; Gdi1p	404	7.3e-37
Contig825	SERUM PARAOXONASE/ARYLESTERASE 2 (PON 2) (SERUM ARYLDIAKYLPHOSPHATASE 2) (A-ESTERASE 2) (AROMATIC ESTERASE 2)	122	0.00035
Contig823	putative prefoldin subunit; molecular chaperonenon-native actin binding complex	314	2.5e-27
Contig822	None		
Contig819	Yer128wp	110	0.00039
Contig816	None		
Contig809	isobutene-forming enzyme and benzoate 4-hydroxyl891e	146	1.6e-08
Contig808	opsin-1	1378	5.1e-140
Contig807	rehydrin-like protein	716	5.3e-70
Contig803	None		
Contig801	None		
Contig800	Ribosomal protein S10A; Rps10ap	329	6.9e-29
Contig799	oxoglutarate malate translocator	156	3.6e-10
Contig797	None		
Contig795	None		
Contig793	None		
Contig791	PROBABLE PEROXISOMAL MEMBRANE PROTEIN PMP20 (ALLERGEN ASPF 3)	361	2.4e-32
Contig790	transaldolase	413	7.3e-38
Contig787	PUTATIVE PROTEASOME COMPONENT C9/Y13 (MACROPAIN SUBUNIT) (MULTICATALYTIC ENDOPEPTIDASE COMPLEX SUBUNIT)	891	1.6e-88
Contig785	None		
Contig783	SRP1 PROTEIN	189	4.6e-14
Contig778	None		
Contig775	dihydrolipoamide S-(2-methylpropanoyl)transferase (EC 2.3.1.-)precursor	122	4.8e-06
Contig772	None		
Contig769	heat shock protein 70	438	1.7e-40
Contig767	5-aminoimidazole-4-carboxamide ribonucleotide (AICAR) transformylase/IMP cyclohydrolase; Ade17p	161	5.0e-10
Contig760	None		

Contig759	None		
Contig756	None		
Contig754	None		
Contig753	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE PRECURSOR (PPIASE) (ROTAMASE) (CYCLOPHILIN) (CYCLOSPORIN A-BINDING PROTEIN)(CPH)	440	1.1e-40
Contig752	None		
Contig751	None		
Contig747	phosphatidylglycerol/phosphatidylinositol transferprotein	411	1.4e-37
Contig745	Ykl056cp/TRANSLATIONALLY CONTROLLED TUMOR PRotein	495	1.5e-46
Contig744	None		
Contig736	None		
Contig732	None		
Contig731	None		
Contig729	None		
Contig728	None		
Contig727	Ribosomal protein L9B (L8B) (rp24) (YL11);Rpl9bp	620	8.7e-60
Contig726	None		
Contig724	BLI-3 PROTEIN	1089	1.7e-109
Contig720	None		
Contig716	hypothetical protein	386	6.1e-35
Contig714	None		
Contig710	aminopeptidase	133	8.1e-07
Contig709	hypothetical protein	227	4.0e-18
Contig700	40S RIBOSOMAL PROTEIN S0 (RIBOSOME-ASSOCIATED PROTEIN 1)	1052	1.6e-105
Contig697	Ribosomal protein L17A (L20A) (YL17);Rpl17ap	601	8.7e-58
Contig696	none		
Contig695	Ygl157wp/HYPOTHETICAL 38.1 KD PROTEIN IN RCK	267	2.3e-22
Contig689	None		
Contig687	adrenoleukodystrophy protein	118	2.7e-05
Contig686	None		
Contig684	None		
Contig683	None		
Contig682	60s ribosomal protein l38	247	3.1e-20
Contig681	amidase (EC 3.5.1.4) - Aspergillus oryzae	218	2.7e-16
Contig670	none		
Contig662	nucleosome assembly protein	141	3.2e-08
Contig661	None		
Contig659	similar to cytochrome Bc1 J chain; similarto 1BGY	117	2.1e-06
Contig658	proline-rich protein - African clawed frog	93	0.40
Contig657	60S RIBOSOMAL PROTEIN L8 (L7A) (L4)	277	2.0e-23
Contig653	14 kDa mitochondrial ribosomal protein	258	2.4e-21
Contig651	None		
Contig646	None		
Contig643	het-c2 protein	510	4.0e-48
Contig641	SUCCINATE DEHYDROGENASE[UBIQUINONE] IRON-SULFUR PROTEIN, MITOCHONDRIAL PRECURSOR (IP)	593	6.6e-57

Contig640	None		
Contig635	26S PROTEASE REGULATORY SUBUNIT 4 HOMOLOG (MTS2 PROTEIN)	273	1.4e-22
Contig633	None		
Contig632	60S ACIDIC RIBOSOMAL PROTEIN P2 (MINOR ALLERGEN ALT A 6)(ALT A VI)	299	9.6e-26
Contig627	None		
Contig615	None		
Contig614	None		
Contig606	OUTER MITOCHONDRIAL MEMBRANE PROTEIN PORIN	319	7.5e-28
Contig605	ccg-4 putative polypeptide 1	344	1.8e-30
Contig598	None		
Contig594	None		
Contig592	None		
Contig589	None		
Contig587	None		
Contig580	dihydroorotate dehydrogenase	178	5.9e-12
Contig579	None		
Contig577	Integral membrane mitochondrial protein;Prp12p	321	4.9e-27
Contig576	None		
Contig573	None		
Contig572	None		
Contig571	probable ubiquitin-conjugating enzyme e2 (EC6.3.2.19)	329	5.3e-29
Contig568	None		
Contig562	similar to glutamate synthase; cDNA EST EMBL:D27720comes from this gene	194	7.7e-13
Contig561	putative flavoprotein	169	5.9e-12
Contig551	none		
Contig547	None		
Contig544	probable gamma-glutamyl phosphate reductase	598	1.9e-57
Contig542	NADH-UBIQUINONE OXIDOREDUCTASE 24 KD SUBUNIT PRECURSOR	430	1.3e-39
Contig539	L-SERINE DEHYDRATASE/L-THREONINE DEAMINASE	216	1.3e-16
Contig531	PUTATIVE ATP SYNTHASE J CHAIN, MITOCHONDRIAL	140	3.0e-08
Contig527	None		
Contig525	D-LACTATE DEHYDROGENASE [CYTOCHROME] PRECURSOR (D-LACTATEFERRICYTOCHROME C OXIDOREDUCTASE)	421	1.2e-38
Contig520	None		
Contig517	None		
Contig516	None		
Contig513	None		
Contig506	None		
Contig500	None		
Contig499	None		
Contig494	Hypothetical protein	166	1.3e-11
Contig492	None		
Contig491	GLYCOPROTEIN X PRECURSOR	105	0.084
Contig480	none		
Contig473	None		
Contig470	PROBABLE ARGININOSUCCINATE LYASE (ARGINOSUCCINASE) (ASAL)	448	1.6e-41

Contig468	None		
Contig465	PROBABLE ALKANESULFONATE MONOOXYGENASE	120	0.00020
Contig463	Mitochondrial ribosomal protein MRPL16;Mrpl16p	355	1.1e-31
Contig460	None		
Contig459	None		
Contig455	ACETYL-COENZYME A SYNTHETASE (ACETATE--COA LIGASE) (ACYL-ACTIVATING ENZYME)	343	1.1e-29
Contig453	Friedreich ataxia/FRATAXINPRECURSOR	226	5.5e-18
Contig451	None		
Contig449	None		
Contig448	none		
Contig446	Subunit g homolog of ATP synthase; Atp20p	162	3.0e-11
Contig441	None		
Contig439	None		
Contig424	None		
Contig422	yeast Chaperonin hsp78 homolog	581	1.4e-55
Contig419	None		
Contig411	Homolog to twitching motility protein (P.aeroginosa); Ybl036cp	471	5.3e-44
Contig408	ORNITHINE DECARBOXYLASE (ODC)	732	1.3e-71
Contig405	Ylr008cp/hypothetical proteinYLR008c - yeast	248	2.5e-20
Contig402	None		
Contig397	None		
Contig396	none		
Contig394	HYPOTHETICAL 15.9 KD PROTEIN C4A8.02C IN CHROMOSOME I	191	2.6e-14
Contig389	None		
Contig386	None		
Contig376	transcription factor, member of ADA and SAGA	350	4.1e-31
Contig374	None		
Contig372	None		
Contig371	hypothetical protein	638	2.1e-60
Contig367	silk fibroin heavy chain {C-terminal}	211	1.7e-15
Contig364	None		
Contig361	hypothetical protein	308	9.6e-27
Contig357	None		
Contig352	ATP SYNTHASE D CHAIN, MITOCHONDRIAL	340	3.5e-30
Contig348	None		
Contig328	AttT (U59485) [Agrobacterium tumefaciens]	145	2.2e-09
Contig317	none		
Contig316	CGI-39 protein	163	2.6e-11
Contig315	None		
Contig294	none		
Contig293	Yol026cp/probable membraneprotein YOL026c - yeast	142	1.6e-08
Contig292	None		
Contig291	Y50E8A.4	137	1.5e-05
Contig286	None		
Contig283	None		
Contig282	HYPOTHETICAL 8.7 KD PROTEIN ZK632.10 IN CHROMOSOME III	138	1.2e-08

Contig270	None		
Contig267	conserved hypothetical protein	181	2.9e-12
Contig266	3-ISOPROPYLMALATE DEHYDROGENASE (BETA-IPM DEHYDROGENASE) (IMDH) (3-IPM-DH)	220	4.8e-17
Contig264	UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX UBIQUINONE-BINDING PROTEIN QP-C PRECURSOR (UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX 11 KDA PROTEIN) (COMPLEX III SUBUNIT VII)	433	6.4e-40
Contig262	galactose-1-phosphate uridyl transferase;Gal7p	216	1.4e-16
Contig258	None		
Contig255	conserved hypothetical protein	244	6.8e-20
Contig231	None		
Contig227	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE PRECURSOR (PPIASE) (ROTAMASE) (CYCLOPHILIN) (CYCLOSPORIN A-BINDING PROTEIN)CPH)	140	7.3e-09
Contig226	None		
Contig225	None		
Contig223	alcohol dehydrogenase	197	1.3e-14
Contig216	None		
Contig212	None		
Contig211	None		
Contig208	None		
Contig205	None		
Contig204	30 KD HEAT SHOCK PROTEIN	162	3.4e-11
Contig203	None		
Contig195	951003: Homology with human lipoprotein-bindingprotein (PIR Acc. No. A44125);	131	1.9e-06
Contig190	None		
Contig183	None		
Contig182	None		
Contig181	myosin-II; Myp2p	294	1.8e-23
Contig174	Ylr435wp/hypothetical proteinYLR435w - yeast	182	1.1e-12
Contig170	TRANSCRIPTIONAL ACTIVATOR PROTEIN ACU-15	353	2.3e-30
Contig168	None		
Contig167	None		
Contig160	None		
Contig159	caleosin	501	3.5e-47
Contig158	None		
Contig151	None		
Contig150	YPT1-RELATED PROTEIN 5	115	5.4e-06
Contig147	None		
Contig145	None		
Contig140	putative MSF transporter	410	1.7e-36
Contig138	ubiquitin carrier protein 4	350	4.3e-31
Contig137	None		
Contig136	None		
Contig135	None		
Contig133	ORNITHINE AMINOTRANSFERASE (ORNITHINE--OXO-ACIDAMINOTRANSFERASE)	617	2.9e-78
Contig132	hypothetical protein Cj0488	218	6.3e-17

Contig129	putative pollen surface protein	104	0.038
Contig128	None		
Contig121	None		
Contig120	40S RIBOSOMAL PROTEIN S11	653	2.7e-63
Contig119	HYPOTHETICAL 13.5 KD PROTEIN C24B11.09 IN CHROMOSOME I	353	1.9e-31
Contig118	small nuclear ribonucleoprotein; sm type	411	1.4e-37
Contig115	Ydr100wp/probable membraneprotein YDR100w - yeast	203	1.6e-15
Contig113	kexin precursor	161	4.3e-08
Contig111	None		
Contig110	None		
Contig109	None		
Contig108	none		
Contig105	None		
Contig104	solute carrier family 7 (cationic amino acidtransporter, y+ system)	182	2.0e-24
Contig98	PyrABCN	565	3.0e-52
Contig94	None		
Contig93	None		
Contig87	None		
Contig85	None		
Contig81	None		
Contig80	similar to Flavin-binding monooxygenase-like	142	3.3e-08
Contig79	hypothetical UPF0028 family protein	677	1.1e-64
Contig77	None		
Contig75	None		
Contig74	PROLINE-SPECIFIC PERMEASE (PROLINE TRANSPORT PROTEIN)	609	1.5e-58
Contig59	forkhead nuclear signalling domain protein	124	0.00021
Contig56	None		
Contig54	None		
Contig51	eburicol 14 alpha-demethylase	267	267
Contig49	None		
Contig47	None		
Contig45	HYPOTHETICAL 41.5 KD PROTEIN C1F5.03C IN CHROMOSOME I	129	2.0e-05
Contig40	None		
Contig39	None		
Contig35	None		
Contig33	putative mutT protein	247	3.2e-20
Contig30	None		
Contig27	None		
Contig26	None		
Contig23	None		
Contig17	eIF4E-like cap-binding protein	232	1.3e-18

Appendix XI. Genes expressed in both NM and NE cDNA libraries

	Name	HSP Score	P Value	#EST
Contig1447	HYDROPHOBIN PRECURSOR (RODLET PROTEIN) (CLOCK-CONTROLLED GENE PROTEIN 2) (BLUE LIGHT INDUCED PROTEIN 7)	451	6.7e-42	664
Contig1446	GLUCOSE-REPRESSIBLE GENE PROTEIN	359	4.2e-32	371
Contig1445	None			289
Contig1444	None			252
Contig1443	None			194
Contig1442	N,O-DIACETYLMURAMIDASE (LYSOZYME CH)	805	2.1e-79	186
Contig1441	None			175
Contig1440	IgE-binding protein	218	4.0e-17	174
Contig1439	None			168
Contig1438	None			165
Contig1437	ubiquitin precursor - Neurospora crassa	1145	2.2e-115	161
Contig1436	putative sugar transporter	472	4.5e-44	156
Contig1435	problem assembly contig, see detail at the end of this table (clock-controlled gene-6 protein)			154
Contig1433	flavohemoglobin	134	4.3e-05	129
Contig1432	None			123
Contig1431	None			120
Contig1430	None			115
Contig1429	Assembly problem contig, see detail at the end of this table (GTP binding protein)			114
Contig1428	ubiquitin precursor - Neurospora crassa	1106	2.8e-111	109
Contig1427	None			109
Contig1426	None			101
Contig1425	GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (GAPDH CLOCK-CONTROLLED PROTEIN 7)	447	2.0e-41	102
Contig1424	None			101
Contig1423	PHOSPHOGLYCERATE KINASE	2122	5.3e-219	98
Contig1422	HEAT SHOCK 70 KD PROTEIN (HSP70)	786	2.4e-77	97
Contig1421	PROTEIN SNODPROT1 PRECURSOR	405	4.9e-37	97
Contig1420	None			97
Contig1419	None			95
Contig1418	None			95
Contig1417	HYPOTHETICAL 48.3 KD PROTEIN IN HSP26-TIF32 INTERGENICREGION	119	9.0e-05	92
Contig1416	Ribosomal protein L23A (L17aA) (YL32);Rpl23ap	580	1.7e-55	90
Contig1415	None			86
Contig1414	None			84
Contig1413	None			81
Contig1412	putative 20kDa subunit of the V-ATPas	803	3.5e-79	79
Contig1411	HYPOTHETICAL 48.3 KD PROTEIN IN HSP26-TIF32 INTERGENICREGION	482	4.0e-45	78

Contig1410	GLUCOSE-REPRESSIBLE GENE PROTEIN	205	9.1e-16	75
Contig1409	None			73
Contig1408	HISTONE H3	666	1.2e-64	71
Contig1407	None			70
Contig1406	60S RIBOSOMAL PROTEIN L5	1450	9.0e-148	64
Contig1405	Ribosomal protein S30A; Rps30ap	153	3.2e-10	66
Contig1404	60S ribosomal subunit protein L6B (L17B) (rp18) (YL16); Rpl6bp	479	7.8e-45	62
Contig1403	None			61
Contig1401	hypothetical protein	131	4.5e-05	60
Contig1400	None			59
Contig1399	GLUCOSE-REPRESSIBLE GENE PROTEIN	359	4.0e-32	58
Contig1398	None			57
Contig1397	ubiquinol-cytochrome c reductase complex subunit	321	3.5e-28	57
Contig1396	none			55
Contig1395	None			55
Contig1394	GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (GAPDH)(CLOCK-CONTROLLED PROTEIN 7)	645	1.8e-62	54
Contig1393	N,O-DIACETYLMURAMIDASE (LYSOZYME CH)	376	7.2e-34	54
Contig1389	None			53
Contig1387	ALCOHOL DEHYDROGENASE I (ADH 2)	305	2.3e-26	51
Contig1385	None			50
Contig1384	None			50
Contig1382	PHOSPHOENOLPYRUVATE CARBOXYKINASE [ATP]	639	8.6e-62	49
Contig1381	HEAT SHOCK PROTEIN 90 HOMOLOG (SUPPRESSOR OF VEGETATIVEINCOMPATIBILITY MOD-E)	1319	7.0e-13	49
Contig1379	(putative) lipid binding protein; supressorof a cdc25 mutation; Tfs1	121	0.00023	47
Contig1378	HISTONE H4.1	409	2.2e-37	46
Contig1377	40S RIBOSOMAL PROTEIN S27	367	6.5e-33	46
Contig1376	None			46
Contig1375	AmMst-1	672	3.0e-65	46
Contig1372	None			44
Contig1371	symbiosis-related protein	553	1.2e-52	42
Contig1369	None			40
Contig1368	soluble cell wall protein; Scw11p	461	6.4e-43	40
Contig1365	None			40
Contig1364	None			39
Contig1363	Yor052cp/hypothetical proteinYOR052c - yeast	178	2.6e-12	39
Contig1361	None			38
Contig1359	PROTEIN FDD123 (CVHSP30/1)	186	1.6e-12	38
Contig1358	clock-controlled gene-9 protein	860	3.0e-185	37
Contig1354	Protein involved in the aging process; Sun4p	582	1.0e-55	36
Contig1352	alcohol dehydrogenase	363	1.7e-32	36
Contig1351	translation initiation factor 3 (eIF3); Sui1p	365	1.0e-32	36
Contig1350	None			36
Contig1347	None			35

Contig1346	None			35
Contig1344	FK506-BINDING PROTEIN (FKBP) (PEPTIDYL-PROLYL CIS-TRANSISOMERASE) (PPIASE)	622	5.5e-60	35
Contig1343	CROSS-PATHWAY CONTROL PROTEIN 1	284	9.3e-48	34
Contig1340	activator of Hsp70 and Hsp90 chaperones	199	9.4e-13	33
Contig1336	ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA)	1136	1.9e-114	32
Contig1334	AmMst-1	291	2.5e-24	32
Contig1333	THIAZOLE BIOSYNTHETIC ENZYME PRECURSOR (STRESS-INDUCIBLEPROTEIN STI35)	1072	1.2e-107	31
Contig1331	CATALASE A	372	1.3e-32	31
Contig1330	Fox2 protein	246	8.1e-19	31
Contig1329	None			31
Contig1328	acyl-CoA dehydrogenase (acd-7)	297	3.8e-25	31
Contig1327	Ylr019wp/hypothetical proteinYLR019w - yeast	662	3.5e-64	30
Contig1326	40S RIBOSOMAL PROTEIN S26E (CRP5) (13.6 KD RIBOSOMALPROTEIN)	479	8.8e-45	30
Contig1325	None			30
Contig1324	40S ribosomal protein S12	527	6.9e-50	29
Contig1323	ENOLASE (2-PHOSPHOGLYCERATE DEHYDRATASE) (2-PHOSPHO-D-GLYCERATE HYDRO-LYASE)	582	9.3e-56	29
Contig1322	None			29
Contig1321	putative glycosyltransferase	422	9.7e-39	29
Contig1320	peptidylprolyl isomerase	190	3.7e-14	29
Contig1314	INITIATION FACTOR 5A (EIF-5A) (EIF-4D)	840	4.4e-83	28
Contig1313	60S ribosomal protein P0 (A0) (L10E); Rpp0p	516	9.5e-49	28
Contig1312	ribosomal protein L35	315	1.9e-27	28
Contig1311	40S RIBOSOMAL PROTEIN S15 (S12)	757	2.5e-74	28
Contig1310	GLUCOKINASE (GLUCOSE KINASE) (GLK)	190	3.2e-12	28
Contig1307	Ribosomal protein S22A (S24A) (rp50)(YS22); Rps22ap	588	2.3e-56	28
Contig1306	None			28
Contig1305	THIAZOLE BIOSYNTHETIC ENZYME PRECURSOR (STRESS-INDUCIBLEPROTEIN STI35)	591	1.2e-56	28
Contig1304	None			28
Contig1303	PUTATIVE PROTEIN TRANSPORT PROTEIN SEC61 GAMMA SUBUNIT	245	5.4e-20	27
Contig1301	HISTONE H3	316	1.7e-27	27
Contig1300	histone H2A	467	1.5e-43	27
Contig1299	None			27
Contig1298	60S RIBOSOMAL PROTEIN L15	920	1.2e-91	27
Contig1297	similar to acyl-CoA dehydrogenases and epoxidehydrolases; Pfam domain PF00441;	353	3.9e-40	26
Contig1296	None			26
Contig1292	Hmp1	134	3.0e-08	26
Contig1290	None			25
Contig1289	40S RIBOSOMAL PROTEIN S2	828	7.3e-82	25
Contig1287	None			25
Contig1286	None			25
Contig1282	Ribosomal protein L14B; Rpl14bp	333	2.1e-29	24
Contig1281	60S RIBOSOMAL PROTEIN L39 (YL36)	233	9.9e-19	24

Contig1280	60S ribosomal protein L37A (L43) (YL35);Rpl37ap	202	3.7e-24	24
Contig1279	None			24
Contig1278	Hypothetical Protein	165	6.6e-11	23
Contig1275	subunit VI of cytochrome c oxidase; Cox6p	270	1.1e-22	23
Contig1274	None			23
Contig1272	putative protein	125	4.5e-05	23
Contig1271	gamma-actin	868	5.2e-86	23
Contig1261	None			22
Contig1260	None			22
Contig1259	None			22
Contig1257	ATP SYNTHASE PROTEIN 9, MITOCHONDRIAL PRECURSOR (LIPID-BINDING PROTEIN)	613	4.8e-59	21
Contig1253	None			21
Contig1250	None			21
Contig1248	Ribosomal protein L2A (L5A) (rp8) (YL6);Rpl2ap	1080	1.6e-108	20
Contig1246	flavohemoglobin	479	7.3e-45	20
Contig1244	Ribosomal protein L34B; Rpl34bp>	400	2.0e-36	20
Contig1241	BcDNA.GM14838	315	1.8e-27	20
Contig1239	MITOCHONDRIAL RIBOSOMAL PROTEIN S5	338	7.4e-30	20
Contig1237	flavohemoglobin	365	1.0e-32	19
Contig1236	None			19
Contig1234	None			19
Contig1227	None			19
Contig1223	40S RIBOSOMAL PROTEIN S9 (S7)	795	2.7e-78	18
Contig1221	PROBABLE GLUCOSE TRANSPORTER RCO-3	510	4.5e-48	18
Contig1216	elongation factor 2	325	2.0e-27	18
Contig1215	none			18
Contig1214	None			18
Contig1212	60s ribosomal protein l22	284	3.5e-24	18
Contig1209	None			18
Contig1205	GLUTAMINE SYNTHETASE (GLUTAMATE--AMMONIA LIGASE)	358	5.6e-32	17
Contig1199	PYRUVATE DEHYDROGENASE E1 COMPONENT ALPHA SUBUNIT, MITOCHONDRIAL PRECURSOR (PDHE1-A)	350	3.7e-31	17
Contig1198	PEPTIDYL-PROLYL CIS-TRANS ISOMERASE PRECURSOR (PPIASE) (ROTAMASE) (CYCLOPHILIN) (CYCLOSPORIN A-BINDING PROTEIN)(CPH)	969	9.6e-97	17
Contig1195	40S RIBOSOMAL PROTEIN S17 (CRP3)	677	7.0e-66	17
Contig1185	Ribosomal protein S29A (S36A) (YS29);Rps29ap	247	3.2e-20	16
Contig1184	None			16
Contig1181	None			16
Contig1178	None			15
Contig1177	PH RESPONSIVE PROTEIN 1 PRECURSOR (PH-REGULATED PROTEIN1)	431	9.5e-40	15
Contig1176	Manganese-containing superoxide dismutase;Sod2p	622	5.0e-60	15
Contig1171	CLC chloride channel protein	104	0.061	15
Contig1168	alpha-NAC, muscle-specific form gp220	234	4.3e-28	15
Contig1166	None			15
Contig1163	None			15
Contig1160	None			15

Contig1159	None			15
Contig1158	60S ACIDIC RIBOSOMAL PROTEIN P1 (ALLERGEN ALT A 12) (ALTA XII)	305	2.1e-26	14
Contig1157	60S RIBOSOMAL PROTEIN L23A	440	1.1e-40	14
Contig1156	None			14
Contig1151	PYRUVATE KINASE	686	8.7e-67	14
Contig1149	GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (GAPDH) (CLOCK-CONTROLLED PROTEIN 7)	302	4.7e-26	14
Contig1142	None			14
Contig1141	None			14
Contig1133	None			13
Contig1129	None			13
Contig1115	FRUCTOSE-BISPHOSPHATE ALDOLASE	687	6.9e-67	12
Contig1112	None			12
Contig1097	minichromosome maintenance protein Mcm7p	476	6.2e-44	12
Contig1096	none			12
Contig1089	None			11
Contig1083	COPROPORPHYRINOGEN III OXIDASE PRECURSOR (COPROPORPHYRINOGENASE) (COPROGEN OXIDASE) (COX)	436	2.4e-40	11
Contig1078	PROBABLE GLUCOSE TRANSPORTER RCO-3	359	1.3e-31	11
Contig1073	None			11
Contig1057	40S ribosomal protein S24A; Rps24ap	372	1.9e-33	10
Contig1046	None			10
Contig1044	similar to CCAAT/enhancer-binding protein	120	6.4e-06	10
Contig1038	None			10
Contig1031	Identical to gb|U12536 3-methylcrotonyl-CoA carboxylase precursor protein from Arabidopsis thaliana. ESTs gb|H35836, gb|AA651295 and gb|AA721862 come from this gene.	462	1.3e-42	10
Contig1020	None			9
Contig1005	None			9
Contig1004	ubiquitin conjugating enzyme UBC1	753	7.1e-74	9
Contig1001	None			9
Contig997	None			9
Contig992	TUBULIN ALPHA CHAIN	752	8.8e-74	9
Contig982	dolichyl phosphate-D-mannose: proteinO-D-mannosyltransferase; Pmt2p	581	1.3e-55	9
Contig975	glyceraldehyde 3-phosphate dehydrogenase	591	1.1e-56	8
Contig969	None			8
Contig965	None			8
Contig948	ATP SYNTHASE DELTA CHAIN, MITOCHONDRIAL PRECURSOR	671	3.2e-65	8
Contig946	None			8
Contig945	ribosomal protein S28	630	8.4e-61	8
Contig940	None			8
Contig936	None			8
Contig930	None			8
Contig928	dolichyl phosphate-D-mannose: proteinO-D-mannosyltransferase; Pmt2p	646	1.6e-62	8

Contig927	hypothetical protein L - Neurospora crassa mitochondrion (SGC3)	553	1.3e-52	8
Contig923	None			7
Contig909	None			7
Contig906	L-SERINE DEHYDRATASE (L-SERINE DEAMINASE)	201	4.5e-15	7
Contig905	60S RIBOSOMAL PROTEIN L12	660	4.9e-64	7
Contig898	None			7
Contig889	GLUTAMINE SYNTHETASE (GLUTAMATE--AMMONIA LIGASE)	748	2.5e-73	7
Contig885	fatty acid omega-hydroxylase (P450foxy)	243	1.6e-18	7
Contig873	60S RIBOSOMAL PROTEIN L32-A	457	1.6e-42	7
Contig871	ubiquitin-conjugating enzyme E2	607	2.2e-58	7
Contig856	FRUCTOSE-BISPHOSPHATE ALDOLASE	216	1.3e-16	6
Contig851	None			6
Contig850	HISTONE H2B	507	8.2e-48	6
Contig843	None			6
Contig841	phosphoglucomutase, minor isoform; Pgm1p	973	4.1e-97	6
Contig839	PROTEIN SNODPROT1 PRECURSOR	104	4.0e-05	6
Contig832	None			6
Contig829	None			6
Contig818	alcohol dehydrogenase	373	1.4e-33	6
Contig817	None			6
Contig814	non-functional folate binding protein			6
Contig806	None			6
Contig788	None			6
Contig768	None			5
Contig766	None			5
Contig757	Translation elongation factor EF-1gamma;Tef4p	208	9.9e-35	5
Contig748	None			5
Contig740	ribosomal protein L26 homolog	387	4.3e-35	5
Contig717	None			5
Contig711	PROTEIN KINASE C-LIKE	823	3.1e-81	5
Contig707	None			5
Contig705	NITROGEN METABOLIC REGULATION PROTEIN (NMR PROTEIN)	371	2.5e-33	5
Contig704	None			5
Contig701	ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA)	999	6.6e-100	5
Contig675	None			5
Contig674	None			5
Contig672	None			5
Contig665	None			4
Contig656	None			4
Contig648	None			4
Contig629	None			4
Contig623	None			4
Contig621	NONE			4
Contig601	60s ribosomal protein l36	255	4.7e-21	4
Contig590	Component of 10 nm filaments of mother-budneck; Cdc11p	227	1.4e-17	4
Contig588	Ran/spi1 binding protein	338	7.2e-30	4
Contig586	KETOL-ACID REDUCTOISOMERASE PRECURSOR			

	(ACETOHYDROXY-ACIDREDUCTOISOMERASE)	202	7.4e-15	4
Contig585	ubiquitin precursor - Neurospora crassa	534	1.2e-50	4
Contig570	None			4
Contig555	None			4
Contig540	CONIDIATION-SPECIFIC PROTEIN 8	432	7.4e-40	4
Contig526	40S RIBOSOMAL PROTEIN S3	565	4.8e-54	3
Contig519	None			3
Contig501	ENOLASE (2-PHOSPHOGLYCERATE DEHYDRATASE) (2-PHOSPHO-D-GLYCERATE HYDRO-LYASE)	593	6.2e-57	3
Contig457	None			3
Contig452	CONIDIATION-SPECIFIC PROTEIN 8	168	8.2e-12	3
Contig430	None			3
Contig393	60S RIBOSOMAL PROTEIN L18	234	7.4e-19	3
Contig381	23S rRNA intron 2 protein	178	4.4e-12	3
Contig377	None			3
Contig368	SUPEROXIDE DISMUTASE [CU-ZN]	454	3.4e-42	3
Contig157	None			2
Contig18	ankyrin-like protein	99	0.11	2
Contig16	KETOL-ACID REDUCTOISOMERASE PRECURSOR (ACETOHYDROXY-ACIDREDUCTOISOMERASE)	1000	3.7e-100	2
Contig15	None			2
Contig1435a	None			1
Contig1435b	ALCOHOL DEHYDROGENASE I	223	8.3e-31	2
Contig1435c	CLOCK-CONTROLLED PROTEIN 6	398	3.1e-36	105
Contig1435d	CLOCK-CONTROLLED PROTEIN 6	381	2.2e-34	11
Contig1435e	None			17
Contig1435f	ALCOHOL DEHYDROGENASE I-ADH1 [Neurospora crassa]	285	7.1e-47	17
Contig1429a	GTP binding protein, almost identical to Gsp1p; Gsp2p	915	4.5e-91	23
Contig1429b	None			49
Contig1429c	THIOREDOXI	256	4.0e-21	39
Contig1429d	None			3