Construction and Functional Assignment of a Manually Annotated Expressed Sequence Tag (EST) Library from the Pathogenic Fungus *Ophiostoma novo-ulmi*

by

Michael Pinchback
BSc, University of Victoria, 1999

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biology

Construction and Functional Assignment of a Manually Annotated Expressed Sequence Tag (EST) Library from the Pathogenic Fungus *Ophiostoma novo-ulmi*

by

Michael Pinchback
BSc, University of Victoria, 1999

Supervisory Committee

Supervisor: Dr. W.E. Hintz, (Biol)
Departmental Member: Dr. David Levin, Ph.D. (Biol)
Departmental Member: Dr. Nigel Livingston, Ph.D. (Biol)
Outside Member: Dr. Claire Cupples, Ph. D. (Bioc)

Supervisory Committee
Supervisor: Dr. W.E. Hintz, (Biol)
Departmental Member: Dr. David Levin, Ph.D. (Biol)
Departmental Member: Dr. Nigel Livingston, Ph.D. (Biol)
Outside Member: Dr. Claire Cupples, Ph. D. (Bioc)

Abstract

A genetic catalogue was generated from expressed sequence tags (ESTs) from the pathogenic filamentous fungus *Ophiostoma novo-ulmi*. Rather than full sequencing of the entire genome, fragments of each gene being actively expressed at a specific point in time were catalogued and annotated for identity and function. This catalogue represents a resource of considerable depth for the purposes of gene discovery, genetic regulation, protein expression, pathogenicity, and growth state studies. An online database was generated to serve as a powerful tool for downstream applications, facilitating and enhancing future research in all of these important areas of fungal biology.

The ascomycetous fungus *Ophiostoma novo-ulmi* represents an excellent model organism for genetic experimentation. A diversity of physiological functions, including dimorphism, pathogenicity, melanin biosynthesis, and glycoprotein secretion at high levels mean that principles elucidated from this fungus are likely of broad application. *Ophiostoma novo-ulmi* has been identified as the causative agent of Dutch elm disease, which has become an economic and horticultural pandemic in North America. As a result, the mechanisms of host-pathogen interaction of this fungus are of particular interest. Initial attempts at disruption of pathogenicity, most commonly by disruption of single genes identified as potential pathogenicity factors, have met with little success. As our understanding of the complexity and co-ordination of proteins involved with host-pathogen interaction deepens, the discovery of a single dominant pathogenicity gene is

becoming increasingly unlikely. As such, a broader genomics approach was employed to work towards identification of groups, or networks of genes that operate in a concerted manner, regulating pathogenicity or parasitic fitness.

A low redundancy library was constructed from *Ophiostoma novo-ulmi* complementary DNA, producing a total of 4386 readable expressed sequence tags (ESTs) from 5760 clones. Of these, 2093 sequences matched with sequences found in public databases while 2293 represented orphan sequences. Of the sequences in the former group, 1761 sequences matched with known proteins while 332 sequences matched with hypothetical/predicted proteins. Sequences matching known proteins included 880 singletons, corresponding to 49.97 % of the ESTs in this category. Extrapolating this proportion to the sequences matching hypothetical proteins estimated the number of singletons in this category to be 166. Similarly, 1835 orphan sequences were estimated to contain 917 unique sequences. Singletons matching entries in public databases (n=880) were manually annotated into functional categories as established by the Munich Information Centre for Protein Sequences (MIPS). Metabolism (21%), Protein Synthesis (10%), Subcellular Localization (10%), Biogenesis of Cell Components (8%), and Transcription (8%) categories were the most highly represented.

Supervisor: Dr. W.E. Hintz, (Biol)

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BT (toxin)..... *Bacillus thuringiensis* (toxin)

cDNA..... complementary deoxyribonucleic acid

DNA..... deoxyribonucleic acid

DED..... Dutch elm disease

EST..... expressed sequence tag

FunCat..... functional catalogue

HMW..... high molecular weight

LMW..... low molecular weight

MIPS..... Munich Information Centre for Protein Sequences

mt..... mitochondrial

nr..... nuclear

OCM..... *Ophiostoma* complete medium

RAPD..... random amplified polymorphism DNA

RFLP..... restriction fragment length polymorphism

RNA..... ribonucleic acid

UV..... ultraviolet (light)

Acknowledgments

I would like to thank my supervisor, Dr. William Hintz, for his patience, support, and guidance throughout my degree. Similarly, the members of the Hintz lab, past and present, were of tremendous help troubleshooting experimental difficulties; in particular, I am indebted to Steven Burgess, and Drs. Joshua Eades and Bradley Temple. Dr. Ben Koop, Matt Rise, and the remaining members of the Koop lab were invaluable during sequencing and BLAST analysis of the EST library.

I would like to thank the University of Victoria for fellowship support during my degree, and the National Sciences and Engineering Research Council of Canada for funding this research.

Above all, I would like to thank my wife Cari, whose patience and devotion made completion of this degree possible.

Introduction - *Ophiostoma novo-ulmi:* The Causal Agent of Dutch Elm Disease

Ophiostoma novo-ulmi has been proposed as an excellent model organism for the study of tree pathogenic fungi. This is due, in part, to biological characteristics including a haploid genome, heterothallism, and yeast/mycelium dimorphism (Bernier, 1993). In addition, methodological considerations such as the rapid and abundant production of biomass on a variety of solid and liquid media, inducible sexual matings, the availability of chemical, physical, and insertional mutagenesis protocols, and the development of elm calli for host-pathogen interactions make *Ophiostoma novo-ulmi* very attractive for laboratory experimentation.

Ophiostoma fungi have been studied in the laboratory for several decades, due in large part to the economic significance of many *Ophiostoma* species. Various saprobic *Ophiostoma* species have been extremely problematic within the lumber industry due to their association with sapstain in raw lumber (Kim et al., 1999). *Ophiostoma novo-ulmi* exists as a pathogenic fungus for American elms (*Ulmus americana*). First reported as the causative agent of Dutch elm disease (DED) in 1918, *O. novo-ulmi* has been responsible for the decimation of the majority of elm trees in eastern North America, at an enormous economic and horticultural cost.

Disease History

The American elm has been a favoured choice by North American city planners and landscape architects throughout the previous century, and exists now as a cultural and historic landmark. Chosen for its majesty and height, the tree also provides plentiful shade along innumerable streets and boulevards throughout North America. Abundant

foliage in a crown large enough to span a city street allow elms to cleanse city air, moderate temperature through summer months, and filter UV radiation. The elm is a particularly popular choice in harsh northern climates because of its resistance to frost, wind, salt, and extremes of temperature (Hubbes, 1999). Its beauty and robustness made the American elm the predominant shade tree in most North American urban centres; unfortunately, the speed with which DED has been transmitted is nothing short of remarkable.

Ophiostoma novo-ulmi exists in nature in both yeast-like and filamentous forms. Each of these growth states are utilized by the fungus for propagation and pathogenicity in host tissues (Agrios, 1997). Pathogenic attack of host tissues is accomplished by colonization of host vascular tissues by fungal mycelia. The fungus migrates vertically through the vascular tissues of the tree, usually killing it in 1-3 years. Simultaneous attack of a colonized tree by bark beetles for feeding and breeding purposes results in the formation of beetle galleries inside the tree. These breeding galleries also serve as incubation chambers, where thousands of *O. novo-ulmi* yeast-like spores are produced. These spores are hydrophobic and adhere to the bark beetles within the galleries. When an adult beetle leaves the tree, it carries thousands of fungal spores with it to neighbouring trees. Feeding activity by beetles on those trees create fresh wounds that offer fungal spores access to new host tissues for colonization, renewing the fungal life cycle.

Transmitted from tree to tree by insect vectors, primarily the European elm bark beetle *Scolytus multistriatus* (Parker et al., 1947) and the native elm bark beetle *Hylurgopinus rufipes* (Jin et al., 1996), the spores of the fungus can be dispersed through a tree population at a catastrophic rate. The characteristic symptoms of DED, yellowing and

wilting of leaves preceding rapid tree death, were first reported in Holland in 1918. Dutch elm disease spread rapidly through Europe, reaching Great Britain in 1927, and the United States in 1930 (Campana & Stipes, 1981). Losses in England in 1927 seemed significant, but not when compared to the decimation caused by a second wave of DED in the 1960s from a new strain of *O. novo-ulmi* (Gibbs, 1981). By 1980, 17 million of the 23 million elms in southern England were killed. Similar levels of devastation were suffered in North America.

Between 1930 and 1976, over 43 million elms in northeastern United States were lost to DED, cutting the population by more than half (Huntley, 1982). A second introduction of DED to North America via Sorell, Quebec, Canada, in 1945 initiated a particularly devastating series of events. The province of Quebec quickly lost more than 600 000 elms, and the elm population of the City of Toronto was rapidly reduced by 80% (Huntley, 1982). In short course the disease fronts of eastern Canada and northeastern United States met and began to migrate westwards through Ontario and Manitoba, and central United States. While the aforementioned records are admittedly dated, percentage losses observed in Winnipeg, Manitoba, Canada in the 21 years prior to 1996 continued to increase, illustrating the currency of this problem (Hubbes, 1999). Indeed, the disease continues to migrate westward threatening elm populations within Saskatchewan and Alberta. In 1999 the 700 000 remaining elms within Canada were valued at $2.5 billion dollars, making the economic significance of the disease and the need for a means of control abundantly clear.

Genetic Background

The current disease front threatens the remaining unaffected areas of Western Canada, prompting population genetic studies of this pathogen at the disease front. For many fungi, the analysis of mitochondrial (mt) and nuclear (nr) sequences by restriction site patterns and DNA footprinting has identified genetic markers useful for resolving phylogenetic and geographic relationships within closely related populations or species (Taylor, 1986). Recent molecular genetic evidence indicates that the population structure of *O. novo-ulmi* isolates sampled from 66 sites within Saskatchewan and Manitoba are distinguishable and repeatable (Hintz et al., 1993; Temple et al., (In press)). For comparative purposes, genetic analyses were performed concurrently with the closely related non-aggressive pathogen *O. ulmi*. Genetic fingerprinting of nuclear DNA demonstrated that all isolates of *O. novo-ulmi* in the sampled region are comprised of only two nuclear genotypes (A and B). Of these, nuclear genotype A is by far the most prevalent, found in 89% of isolates (Hintz et al., 1993). Isolates bearing this nuclear genotype were previously isolated from extremely diverse geographic locations (Hintz et al., 1991), illustrating the ability of the pathogen to migrate and adapt to new conditions, as well as the stability of the nuclear genome. The persistence of this genotype was further illustrated by random amplified polymorphism DNA (RAPD) marker analysis over a nine year interval.

Ophiostoma novo-ulmi isolates of the western Canada disease front subjected to RAPD marker analysis show that the homogeneity of the nuclear genotypes in the region, in fact, increased from 1993 to 2002 (Temple et al., (In press)). When isolates from the same 1993 collection were examined with RAPD markers alongside isolates sampled in 2002, the presence of only two nuclear genotypes were confirmed. The RAPD primer OPA-13

was used to delineate isolates within the species, and the OPA-13 pattern-1 genotype was found in 75% of the 1993 isolates. The dominance of this genotype increased to 81% for 2002 isolates. The OPA-13 pattern-2 genotype represented 25% and 19% of the 1993 and 2002 samples, respectively, along with a reduction in range when 2002 isolates are compared to those of 1993 (Temple et al., (In press)).

Restriction fragment length polymorphisms (RFLPs) have also been identified for the ribosomal DNA (rDNA) repeat to further elucidate the population structure of the western Canada disease front (Hintz et al., 1993). Ribosomal DNA restriction patterns were conserved for both *O. novo-ulmi* and *O. ulmi*, indicated by the absence of length mutations within each species. Restriction fragment length polymorphisms were observed, however, between the two species. This demonstrates a repeatable, definitive means of differentiating *O. novo-ulmi* and *O. ulmi*, which had historically relied upon morphological and physiological variations, and differences in pathogenicity to elm (Jeng et al., 1988). Restriction patterns of isolates from all sampled areas within the disease front were identical and consistent with those characteristic of *O. novo-ulmi* (Hintz et al., 1993). Repeated sampling nine years later again failed to detect the presence of a single less aggresive *O. ulmi* clone (Temple et al., (In press)). The absence of *O. ulmi* from the sampling area could indicate its displacement from the region by the more aggressive *O. novo-ulmi*, or simply a greater capacity for migration exhibited by *O. novo-ulmi*. The rDNA RFLP data presented here illustrates a clear genetic distinction between the two species, but fails to distinguish between isolates within each species. Mitochondrial genomes are widely accepted to exhibit increased mutation rates, as compared to nuclear genomes, and would therefore be expected to exhibit increased RFLP variation. As such,

the capacity of RFLP analysis of mitochondrial sequences to distinguish specific isolates has also been investigated.

Restriction fragment length polymorphisms reported in the mitochondrial genomes of isolates sampled from the western Canada disease front are comprised of four classes, types I – IV (Hintz et al., 1993). Of these, the type I mt genome was most prevalent (75%) and was ubiquitous throughout the sampling area. While impossible to definitively assign a progenitor mt genome, it was noted that only a single mutational event within the type III mt genome is required to generate the other three mt genotypes, while any other pairwise combination of mt genomes requires the fixation of at least two independent mutational events (Hintz et al., 1993). When associations of mt genotype with nr genotype were examined, it was noted that isolates bearing a type A nr genome in combination with a type I mt genome (A-I) were widely dispersed through the sampling area, and were by far the most abundant at 71.5%. All other pairwise combinations of nr genotype and mt genotype were present at less than 8% (Hintz et al., 1993).

Similarly to mitochondrial versus nuclear genome evolution, vegetative incompatibilities within isolates of the pathogen have been shown to develop much more quickly than observable genetic mutations (Brasier, 1996). Vegetative compatibility (vc) has been used previously to examine the genetic diversity of a pathogen population (Brasier, 1983; Brasier, 1996). Isolates of *O. novo-ulmi* sampled from the western Canada disease front in 1993 were shown to be fully compatible with those from 2002 (Temple et al., (In press)). The lack of vegetative incompatibility after nine years at the disease front indicates a very low inclination towards genetic diversification within this population.

The genotypic homogeneity exhibited by the *O. novo-ulmi* population at the western Canada disease front is uncharacteristic of similar pathogen populations previously examined. A greater diversity of vc types has been documented in the Eurasian aggressive races (EAN) of *O. novo-ulmi*, as compared to the North American aggressive (NAN) races (Brasier, 1996). In areas of Europe suffering a well established epidemic of high vc uniformity, vegetative incompatibilities have been reported within six to ten years (Brasier & Kirk, 2000). It is interesting to note that this time span mirrors the sampling interval of the western Canada disease front, which exhibited complete homogeneity of vc type. Recent phylogenetic evidence suggests that the NAN races are descendents of EAN races (Bates et al., 1993a; Bates et al., 1993b), which has led to the designation of EAN and NAN races as subspecies *novo-ulmi* and subsp. *americana*, respectively (Brasier & Kirk, 2001).

The extreme uniformity of nuclear genetic evidence, complete vegetative compatibility, RAPD marker analysis, and conservation of mt and rDNA repeat restriction patterns indicates that the vast majority of the current disease front affecting western Canada is comprised of a single, very large clone of *O. novo-ulmi*. The report of only two nuclear genotypes, and no transitional genotypes, suggests that sexual outcrossing is rare, and the transmission of *O. novo-ulmi* has predominantly been by asexual means (Hintz et al., 1993). The lack of vegetative incompatibilities within this population is uncharacteristic of established pathogen epidemics, although rapidly expanding pathogen populations have previously been reported to exhibit low genetic diversity (Frank, 1992), perhaps speaking to the urgency of this problem. It seems plausible that the clonal nature of this

highly successful pathogen could be exploited as a means of controlling Dutch elm disease in western Canada, and very likely through much of North America.

Disease Control

Numerous attempts have been made towards identification or development of an effective means of control of Dutch elm disease. The majority of these attempts have concentrated on exploitation of natural host resistance, reduction of insect vector populations, and application of chemical fungicides. Unfortunately, no attempt made thus far has provided a safe, specific, effective, economical solution.

Host Resistance

Although the mechanisms responsible have not yet been elucidated, it is clear that not all elms exhibit similar sensitivity to *O. novo-ulmi.* Early experiments consisting of standard breeding and selection programs of *U. americana* in hopes of developing trees with genetic resistance to DED were disappointing (Holmes, 1976; Oullette & Pomerleau, 1965). This outcome seems probable considering all North American species of elm (*U. rubra, U. thomasii, U. alata, U. serotina, U. erassifolia* and *U. americana*) are susceptible to DED, with *U. americana* exhibiting the highest level of susceptibility (Hubbes & Jeng, 1981). Subsequent work has focused on the development of genetic combinations from European and Asian gene pools (Sherald, 1993; Smalley & Guries, 1993; Smalley et al., 1993a; Townsend & Santamour Jr., 1993; Ware & Miller, 1997). From this work came a number of progeny trees with enhanced resistance to DED; of particular promise was the American Liberty elm. Unfortunately, the nature of these breeding experiments means that the specific mechanisms of increased resistance remain unknown, and as such, no estimate can be inferred regarding the permanence of the

phenotype. Small changes in the genetic background of the fungus could rapidly cause loss of resistance. This is particularly relevant when the generation time of trees is measured in decades while that of the pathogen in days. Unfortunately, formerly resistant Liberty elms have indeed been reported to exhibit susceptibility to DED (Hubbes, 1999). Conversely, the emergence of previously unobserved traits cannot be predicted for progeny from breeding programs, and a DED-resistant seedling could later exhibit an increased sensitivity to frost, for example. It is for these reasons that it has been suggested that modern molecular genetic techniques must be employed towards a deeper understanding of the molecular basis controlling pathogen virulence and host resistance.

Recent work involving host resistance has focused on Asian elm species, particularly the Siberian elm (*U. pumila*), due its native resistance to *O. novo-ulmi* (De Rafael et al., 2001; Solla et al., 2005). While its resistance to DED makes the Siberian elm of particular interest within the laboratory, many of its aesthetic characteristics have been regarded as shortcomings, making this tree an undesirable choice for horticulturists. As such, it serves as an excellent organism for molecular research, potentially allowing researchers to mimic in the American elm the resistance the Siberian elm has acquired through natural selection. Given the fact that DED resistance is specific to Asian elm species, it has been suggested that DED originated in Asia. More specifically, its origin has recently been placed in the Himalayas (Brasier & Mehrotra, 1995). While development of long-term resistance in elms has proven to be challenging, this line of research is actively being pursued. To the same end, albeit from the opposite direction,

population control of both the fungus and its insect vector has also been examined in hopes of reducing continued decimation of native elms.

Vector Control

Historically, the preferred choice for control of elm bark beetle species (predominantly *S. multistriatus* and *H. rufipes*) has been via chemical pesticides (Hubbes, 1999). In areas of high beetle population, it is anticipated that by reducing the beetle population by pesticides, the transmission of fungal spores could be quickly and dramatically stopped. Emerging concerns regarding the negative environmental impact of chemical pesticides has meant that the utility of this approach has become limited, and it is used in very specific situations. Particular attention must be given to appropriate spraying equipment and technique to ensure effectiveness of the pesticide (Roy et al., 1988). It seems clear that chemical control of bark beetle populations is no longer a viable option, and alternative techniques must be examined.

The use of biological control agents for the control of bark beetle holds much potential, but work in this area remains in the preliminary stages. Early work by Tomalak et al. (Tomalak et al., 1989) looking at insect parasitic (entomopathogenic) nematodes to control bark beetle awaits further development. The potential for Lepidopteran BT toxins as a means of control is also awaiting further examination (Sticklen et al., 1991). From an environmental perspective the use of pheromone traps for vector control holds a great deal of appeal. Unfortunately, the effectiveness of such techniques has been inconsistent (Birch et al., 1981; Sticklen et al., 1991). Similarly, the use of elm bark beetle trap trees (Lanier, 1989) has proven impractical since most cities and towns cannot spare the elm trees to serve as traps. Infected trees serve as incubators and micro-habitats for pathogen

and vector alike, and as such, must be effectively removed and destroyed. The benefits of a thorough sanitation program far outweigh the considerable costs; removal of major inoculum sources is an imperative component of DED control. It is worth noting that a thorough sanitation program also often includes removal of tree species other than American elm, since bark beetles are not restricted to elm trees alone and often migrate to numerous species of tree. Unfortunately, sanitation alone is insufficient to control the spread of this pathogen (Pomerleau, 1981).

Chemical Control of *O. novo-ulmi*

While a variety of chemical fungicides have shown promise for control of *O. novo-ulmi*, each has subsequently been shown to possess shortcomings preventing its broad application. The majority of work has focused on chemical control of the fungus post-infection; treating the tree in hopes of killing the colonized fungus. Benzimidazole systemic fungicides represent the class of chemical most thoroughly explored for DED control. This group of fungicides includes products such as benomyl (Kondo et al., 1973) and Arbotect 20-S (Prosser, 1998; Smalley, 1978). A triazole derivative, propiconazole (trade name "Alamo"), has been proposed as a potential fungicide for *O. novo-ulmi*, although the evidence of its effectiveness is yet to be well established (Prosser, 1998). The effectiveness of these chemicals has been debated for several years, in large part due to inconsistent results following spraying. These inconsistencies are likely consequences of response mechanisms within the tree following treatment.

Wounding of the tree during chemical injection has been associated with increased compartmentalization (Shigo & Campana, 1977). This response sequesters the fungicide within the tree, physically limiting its ability to reach and act on the fungus. Chemical

solubility has also been a problem with some fungicides, including carbendazim phosphate. This problem was addressed by direct injection of the fungicide into the tree's stem (Kondo, 1978; Stennes & French, 1987). Similarly, uptake experiments through the root system of the tree were employed to enhance the delivery of methyl-benzimidazol-2-yl carbamate phosphate (Roy et al., 1980). Physical and economic barriers have limited the effectiveness of chemical fungicide application for the purpose of DED control, and the emergence of newly developed resistance among fungal strains (Bernier & Hubbes, 1990; Schreiber, 1993) has emphasized the need for alternate means of control.

While the specific molecular mechanisms by which tree death occurs have not yet been completely clarified, it has been clear from the outset that DED is a vascular disease (Banfield, 1941). As such, the fungus must invade a large number of vessels in order to effectively colonize the tree. This occurs following spore germination, when fungal hyphae penetrate pit membranes in vessel walls and move from vessel to vessel (Hubbes, 1999). Naturally occurring spore deficient mutants of *O. novo-ulmi* lack the ability to produce conidiospores, blastospores, and ascospores. These mutants have been shown to be incapable of causing internal or external disease symptoms in *U. americana* (Richards et al., 1982; Richards, 1998). Similarly, sterol biosynthesis inhibitors have been shown to interfere with hyphae formation, and associated with suppression of disease development in elm (Scheffer et al., 1988). Unfortunately, the particular derivative inducing the highest level of disease suppression, fenpropimorph, rendered the tree frost sensitive.

Attempts at controlling the spread of DED through breeding programs and chemical spraying have been unsuccessful. Work thus far has been at arm's reach, watching for

changes in phenotype or physiology in response to a particular treatment. It seems evident that the development of an effective means of control for DED can only be attained through a deeper understanding of the molecular mechanisms of fungal pathogenicity and host defense response.

Elm Inoculation and Induced Resistance

Cross-protection of elms following inoculation by a non-aggressive strain of *O. novo-ulmi* was first reported in *U. hollandica* and *U. americana* (Hubbes & Jeng, 1981; Scheffer et al., 1980). Subsequent work supported that seedlings of *U. americana* could be protected from attack by aggressive strains of *O. novo-ulmi* by pre-inoculation by a non-aggressive strain (Duchesne et al., 1986; Jeng et al., 1983; Sutherland et al., 1995). Clearly, the tree is mounting a defense response following inoculation. The chemical nature of this defense has yet to be resolved.

A number of chemicals have been isolated from fungal inhibitory sapwood extract of elm seedlings following inoculation, including a class of molecules known as mansonones (Dumas et al., 1983; Dumas et al., 1986; Jeng et al., 1983). An accumulation of mansonones in elm inoculated with *O. novo-ulmi* was independently observed by Procter and Smalley (Procter & Smalley, 1988), and mansonones were shown to have toxic effects on the physiology and ultra-structure of *O. novo-ulmi* (Wu et al., 1989). Chemically induced mutants exhibiting increased sensitivity to mansonones have been shown to retain pathogenicity (Proctor et al., 1994; Smalley et al., 1993b), suggesting that mansonone production alone does not play a major role in resistance *in vivo*. Duchesne (Duchesne, 1993) suggests that the timing of mansosone production is critical for resistance, rather than the subsequent levels to which mansonones accumulate.

Mansonone levels rise at a much faster rate in the naturally resistant *U. pumila*, and accumulate faster in *U. americana* when inoculated with an aggressive strain of *O. novo-ulmi* versus a non-aggressive strain (Duchesne et al., 1985). While the specific mechanisms of mansonone production are not yet completely understood, it is clear that mansosones represent an inducible defense response in elm.

Mansonone production was induced in tissue cultures of *U. americana* and *U. pumila* following inoculation by *O. novo-ulmi* (Szczegola-Derkacz, 1988). Further, inoculations with autoclaved spores induced the same production as non-autoclaved spores, indicating that the eliciting compounds or structures are heat stable, and not dependent on living organisms. Fungal culture filtrates, cytoplasm, and cell walls of *O. novo-ulmi* have also been shown to elicit mansonone production in elm calli (Yang et al., 1989). The elicitor has now been purified (Yang, 1991), its structure identified (Hubbes, 1999), and has been employed as the basis of a novel biological control technology (http://elmcare.com/community/research/hubbes1.htm).

Injection of purified elicitor was shown to significantly reduce disease development in elm seedlings and trees (10 cm diameter) when challenged with 8000 to 1 million spores of *O. novo-ulmi*. A United States patent for the control of DED has been filed based on the structure of the elicitor (Hubbes, 1999). Both liquid and pellet forms of the elicitor provide similar reductions in disease development. The elicitor is heat stable, has an indefinite shelf life in pellet form, appears at this stage to be environmentally safe, and is easily administered. Field trials have been conducted in Alberta and eastward through to central Ontario. The mechanisms of resistance in elm induced by the elicitor appear to mimic defense cascades seen in agricultural crops (Somssich & Hahlbrock, 1998). It

should be noted though, that the genetic variability in wild type elm trees is considerably greater than that for agricultural crops. This will manifest itself as increased variability in the defense responses in elm.

Inoculation of healthy elm trees with purified elicitor represents a novel approach to DED control, in an arena that has historically focused on breeding programs and chemical spraying. The specific mechanisms of elm defense response to *O. novo-ulmi* are not yet fully understood, but it appears that a molecular approach to resistance, pathogenicity, and virulence has become the avenue of greatest potential.

Virulence and Pathogenicity Factors

A number of candidate pathogenicity factors and parasitic fitness factors have been examined from *O. novo-ulmi* including peptidorhamno-mannan (Claydon et al., 1980; Scheffer et al., 1987), glycopeptides and glycol protein elicitors (Hubbes, 1993; Yang et al., 1989), as well as an extra-cellular laccase thought to be important for survival of the fungus in its host (Binz & Canevascini, 1996).

The most contentious pathogenicity factor examined to date is cerato-ulmin, a secreted hydrophobin first presented by Takai as a virulence factor (Takai, 1980). Experiments by Bernier (Bernier, 1988) failed to confirm a correlation between high levels of cerato-ulmin production and virulence. Production of UV-irradiated mutants deficient in CU production showed inconsistent levels of virulence, perhaps due to genetic changes at unknown loci through random mutagenesis (Tegli & Scala, 1996). Targeted gene disruption of the CU gene locus of an aggressive strain of *O. novo-ulmi* (Bowden et al., 1994; Jeng et al., 1996) revealed no significant changes in virulence. Cerato-ulmin therefore does not appear to be a major pathogenicity factor. This view is supported by

the observation that naturally occurring CU⁻ deficient mutants are pathogenic (Brasier et al., 1994). Over-expression of cerato-ulmin in transformed non-aggressive strains of *O. novo-ulmi* did not lead to an increase in virulence (Temple et al., 1997). Temple et al. (1997) suggest that while CU in itself does not increase virulence, it could increase the parasitic fitness of the fungus by providing a more hydrophobic and vector-adherent phenotype, while protecting spores from desiccation during transmission.

More recently, symptoms consistent with DED followed expression of cerato-ulmin in the normally non-pathogenic *O. quercus* (Del Sorbo et al., 2000). Scala et al. (Scala et al., 1997) also report higher levels of CU in wilting leaves of elm infected with aggressive strains of *O. novo-ulmi* compared to levels found in leaves of elm infected with non-aggressive strains. Care must be taken to recognize that the correlation in this case exists between elevated levels of CU and the presence of an aggressive strain, but not necessarily between CU and increased virulence.

The contradictory evidence regarding the effect of cerato-ulmin on virulence is indicative of the multi-factor mechanisms at play in host-pathogen interactions, which have prevented the complete suppression of pathogenicity by knockout of a single gene. This illustrates the need for genetic studies performed at the level of gene networks, facilitating the identification of genetic cascades, as changes in expression of single genes induce expression changes in related primary or secondary metabolic pathways. At this point, little work has been done examining the genome of *Ophiostoma* species. Depositions to public databases remain very limited (less than 100 total, of which many are redundant). The genome of *O. novo-ulmi* therefore represents an opportunity for

extensive gene discovery. The diversity of physiological function displayed by *O. novo-ulmi* adds further richness to the opportunity.

Expressed sequence tags provide a means of gathering large amounts of genetic information and expression data, due to advances in high-throughput sequencing. A genomic approach to pathogenicity could potentially bear fruit by elucidating sets of related pathogenicity genes and parasitic fitness factors which, controlled in a concerted manner, may offer a deeper suppression of pathogenicity. The recent emergence of RNA interference as an effective means of genetic interference seems a powerful technology when mated with the large amounts of genomic sequence data produced by EST analysis.

Scientific Objectives

i). Isolation of representative high-quality poly(A) RNA for complementary DNA synthesis and construction of low redundancy EST libraries for *Ophiostoma novo-ulmi* H327 cultured to induce growth in yeast-like and filamentous forms.

ii). Manual annotation of EST sequence information, indicating putative identity, closest matches to known or theoretical sequences in public databases, loci in physical library, and functional category.

iii). Functional assignment of known ESTs to illustrate functional profile of genome under specific growth conditions (yeast-like, filamentous), and comparison of the representation of various classes of ESTs to those described for another filamentous fungus, *Neurospora crassa*.

Methods and Materials

Construction of cDNA Library and EST Database

Examination of molecular pathways has traditionally been limited to a single gene. Conversely, metabolism *in vivo* is known to be a complex system where components of pathways are in constant flux as an organism responds to changing conditions within the cellular and systemic environments. A change in expression within a pathway induces subsequent changes in related secondary pathways.

The recent advent of high-throughput DNA sequencing techniques has facilitated the emergence of genomics, bioinformatics, and proteomics. Large amounts of genetic information can now be obtained in a relatively short time period, and at a fraction of the cost compared to even five years ago. Genome sequencing and expressed sequence tag (EST) projects are no longer regarded as an end; rather, they are a means of discovering not only novel genes and proteins, but also developing new questions regarding the complexity of gene regulation and protein function under defined growth conditions. To this end, an EST library was constructed for *O. novo-ulmi* H327 for the purposes of gene discovery and functional assignment of known ESTs, and will serve as a comparative database for future gene regulation experiments for filamentous fungi.

Fungal Strains and Culture Conditions

The wild-type, highly aggressive *Ophiostoma novo-ulmi* strain H327 (Bernier & Hubbes, 1990), was used for all experimentation including DNA and RNA extractions. *Ophiostoma novo-ulmi* is a dimorphic fungus and grows in mycelial and yeast-like forms. These growth states can generally be induced by culture conditions. Stock cultures were maintained on solid OCM plates at 24°C. For generation of yeast-like cultures, $1mm^2$

agar plugs cut from edges of actively growing stock culture plates were inoculated into 50 ml liquid *Ophiostoma* complete medium (OCM) in 125 ml or 250 ml Erlenmeyer flasks (Bernier & Hubbes, 1990) and incubated for 3 – 5 days at room temperature with agitation on an orbit shaker. Yeast cells were then seperated by filtration through 3 layers of sterile miracloth (Calbiotech, La Jolla, CA) and pelleted by centrifugation for 15 minutes at 700g. Mycelial cultures were initiated by generation of yeast-like cultures as described above, and inoculation of 75 µl dense liquid culture onto solid OCM plates overlaid with a single layer of sterile cellophane (Bio-Rad), followed by stationary incubation at 24°C for 5 – 7 days. Mycelial growth covered the layer of cellophane and was removed by peeling the cellophane from the plate. The mycelial mat was scraped from the surface of the cellophance and freeze-dried overnight.

Poly(A) mRNA Extraction

Yeast-like cDNA Library

For construction of the yeast cDNA library, a yeast-like culture of *O. novo-ulmi* was initiated as described above and shaken at 120 rpm for 4 days at room temperature. Prior to the appearance of any colour change or accumulation of cellular debris, the culture was vacuum-filtered through 3 layers of sterile miracloth and weighed. Extraction of poly(A) RNA was performed using a MicroPoly(A) Pure mRNA Isolation kit (Ambion), as per manufacturer's instructions. Moist cells weighing 210 mg were resuspended in 2.1 ml Lysis Solution and subjected to homogenization on ice for 20 seconds prior to storage at -20°C for 2 days. After thawing, 4.2 ml Dilution Buffer was added, and the sample was inverted several times for mixing. Particulate debris was then pelleted by centrifugation (Fisher Scientific) at 12,000 x g for 15 minutes. The supernatant was

transferred to a fresh tube to which one vial of oligo(dT) cellulose was added. The RNA mixture was incubated with the oligo(dT) cellulose for 50 minutes at room temperature with agitation at 60 rpm on a platform shaker (New Brunswick Scientific). The oligo(dT)-RNA was pelleted by centrifugation at 4000 x g for 3 minutes at room temperature. The supernatant was removed and the cellulose pellet was washed three times with 1.0 ml Binding Buffer followed by centrifugation at 4000 x g for 3 minutes. The washing procedure was repeated a further three times with 1.0 ml Washing Buffer. The final oligo(dT)-RNA pellet was resuspended in 400 µl Wash Buffer and the slurry was transferred to a spin column. The column was spun at 4000 x g for 20 seconds to drive the wash buffer through the cellulose into the microcentrifuge tube below. The eluant was discarded and the washes repeated twice in succession with 500 µl Wash Buffer, each time gently stirring the resin with a pipette tip. The column was transferred to a fresh microcentrifuge tube and 100 µl Elution Buffer pre-warmed to 65°C was applied to the column. After 30 seconds at room temperature, the column was spun at 4000 x g to elute the poly(A) RNA. An additional aliquot of 100 µl Elution Buffer was applied to the column and eluted into the same microfuge tube as described above.

The eluted RNA was precipitated by the addition of 20 µl 5M ammonium acetate, 1 µl glycogen, and 550 µl 90% ethanol. The precipitation mixture was incubated overnight at -20°C before being spun at 13,000 x g for 20 minutes at 4°C. The supernatant was aspirated with a P200 automatic pippetor (Brinkmann). The RNA pellet was dried by spinning in a speedvac (Eppendorf) at room temperature for 5 minutes. The RNA pellet was then resuspended in 20 µl RNase-free sterile distilled water (Gibco BRL) and stored

at -80°C. Spectrophotometric analysis indicated a concentration of 853 ng/µl, with an A_{260}/A_{280} ratio of 1.452.

Mycelium cDNA Library

Poly(A) RNA was also isolated from filamentous mycelial mats of *O. novo-ulmi* grown on solid OCM medium overlaid with cellophane. The mycelial culture was harvested after 5 days incubation at room temperature, yielding a mycelial mat of high structural integrity. Two 1 cm^2 samples were excised from the mycelial mat for poly(A) RNA isolation, yielding 208 mg and 225 mg of ribonucleic acid. Poly(A) RNA was isolated as described earlier using a MicroPoly(A) Pure mRNA Isolation kit (Ambion) as per the manufacturer's instructions. Each mycelial sample was mechanically homogenized on ice for 30 seconds in 2.1 ml Lysis Buffer. Following homogenization, samples were diluted three-fold by the addition of 4.2 ml Dilution Buffer to each sample. The samples were inverted rapidly eight times to ensure complete mixing, and particulate debris were pelleted by centrifugation at room temperature at 12,000 x g for 15 minutes. The supernatants of the two samples were each transferred to fresh 1.5 ml eppendorf tubes to which one vial of oligo(dT) cellulose was added for RNA-binding. The samples were incubated in the presence of oligo(dT) cellulose for 65 minutes at room temperature with agitation at 60 rpm on a platform shaker (New Brunswick Scientific). Oligo(dT)-bound RNA was pelleted by centrifugation (Eppendorf) at 4000 x g for 3 minutes at room temperature. The supernatants were removed and the cellulose pellets were washed three times in succession with 1.0 ml Binding Buffer followed by centrifugation at 4000 x g for 3 minutes. This washing procedure was repeated a further three times with 1.0 ml Wash Buffer. The final oligo(dT)-RNA pellets were resuspended in two 400 µl volumes of

Wash Buffer and the slurries were transferred to two spin columns. The columns were spun at 4000 x g for 20 seconds to drive the wash buffer through the cellulose into the microcentrifuge tubes below. The eluant was discarded and the washes repeated twice in succession with two 500 µl volumes of Wash Buffer, each time gently stirring the resin with a pipette tip. The columns were transferred to a fresh microcentrifuge tube and 100 µl Elution Buffer pre-warmed to 65°C was applied to each column. After 30 seconds at room temperature, the columns were spun at 4000 x g to elute the poly(A) RNA. An additional aliquot of 100 µl Elution Buffer was applied to each column and eluted into the same tube by centrifugation at 4000 x g.

The eluted RNA samples were precipitated by the addition of 20 µl 5M ammonium acetate, 1 µl glycogen, and 550 µl 90% ethanol to each of the two samples. The precipitation mixtures were incubated overnight at -20°C and pelleted by centrifugation (Fisher Scientific) at 13,000 x g for 20 minutes at 4°C. The supernatants were aspirated with a P200 automatic pippetor and the RNA pellets were dried by spinning in a speedvac (Eppendorf) at room temperature for 5 minutes. The two pellets were each resuspended in 10 µl DEPC-treated water. Spectrophotometric analysis indicated concentrations of 928 ng/µl and 1156 ng/µl, with A_{260}/A_{280} ratios of 1.131 and 1.145, respectively.

Complementary DNA Synthesis

For construction of both yeast-like and mycelium *O. novo-ulmi* cDNA libraries, the pBluescriptII XR cDNA Library Construction kit (Stratagene) was used through first and second round cDNA synthesis, cDNA termini blunting, *Eco*RI adapter ligation, and adapter phosphorylation. First-strand synthesis was performed at 42°C for one hour with

10.40 µg mycelium mRNA, and 9.20 µg yeast-like mRNA. Samples were cooled on ice five minutes prior to second strand synthesis at 16°C for 2.5 hours. The termini blunting reaction was stopped after exactly 30 minutes by extraction with 200 µl phenol-chloroform [1:1(v/v)]. Blunt cDNA termini were precipitated overnight at -20°C following addition of two volumes of 95% ethanol and 1/10th volume 3M sodium acetate. The subsequent pellet was dried by lyophilization and resuspended in 9 µl *Eco*RI adapters (Stratagene). Adapters were ligated to blunt cDNA termini after addition of 1 µl 10X ligase buffer, 1 µl 10 mM rATP, 4 units T4 DNA ligase and incubation at 8°C overnight. *Eco*RI ends were phosphorylated with 10 units T4 polynucleotide kinase and digested with 120 units *Xho*I as outlined by the manufacturer. Following digestion at 37°C for 2 hours, cDNA was ethanol precipitated overnight at -20°C, microcentrifuged at maximum speed for 1 hour 5 minutes at 4°C, and resuspended in 10 µl elution buffer (Qiagen).

cDNA Fractionation, Ligation and Transformation

The following procedure replaces the column fractionation protocol recommended by the manufacturer (Stratagene). Size fractionation of the synthesized cDNA was performed via gel electrophoresis through 1% agarose in nuclease-free TAE buffer (Gibco-BRL). Electrophoresis was conducted at 80V for 1 hour using a Bio-Rad power supply. The gel was immersed in 0.5 µg/ml ethidium bromide staining solution for one minute, then transferred to clean water to destain for 15 minutes. Smears of ethidium bromide-stained cDNA were visualized on a UV transilluminator (Bio-Rad) at 365 nm. Complementary DNA was fractionated by size into low molecular weight (LMW – 400 bp to 2000 bp) and high molecular weight (HMW – 2000 bp to 5000 bp) categories, for

both yeast-like and mycelium libraries. Ethidium bromide-visualized smears corresponding to each fraction were excised from the gel and cDNA isolated using a Qiaquick Gel Extraction kit (Qiagen). Fractionated cDNA was eluted with 50 µl elution buffer. To concentrate samples, each was ethanol precipitated overnight at -20°C, and the subsequent pellets resuspended in 5.5 µl nuclease-free dH_20 (Gibco-BRL). Spectrophotometric analysis of yeast HMW and LMW cDNAs indicated concentrations of 2.8 ng/µl and 3.9 ng/µl, respectively. The concentrations of Mycelium HMW and LMW cDNAs were estimated in a similar manner to be 1.1 ng/µl and 2.2 ng/ul, respectively.

Fractionated yeast-like cDNA and mycelium cDNA were each ligated independently into pBluescript II SK^- vector (pBluescript II XR cDNA Library Construction kit, Stratagene). Ligation reactions contained 10 ng fractionated cDNA (with the exception of Mycelium HMW where only 6 ng was isolated), 20 ng vector, and 2 units T4 DNA ligase in 1x ligase buffer/1 mM rATP (pH 7.5) in a final volume of 5.0 µl. Ligation reactions were incubated at 12°C for 24 hours.

Transformation of entire 5 µl of each ligation reaction into 50 µl Escherichia coli DH12S ultracompetent cells was performed using a BTX Electro Cell Manipulator 600 (1.30V; 2.5 kV/resistance; capacitance timing = out; 129 Ω) in 0.1 mm gap cuvettes (Bio-Rad). Electroporated bacteria were allowed to recover at room temperature for 30 minutes prior to plating 1 µl and 10 µl aliquots (for titering purposes) on 2YT plates (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, pH adjusted to 7.0 with 2N NaOH) amended with 50 µg/ml ampicillin (Sigma), 100 µg/ml X-galactose (Sigma), and 31 mg/ml isopropyl β-D-1-thiogalactopyranoside (Sigma). Transformant bacterial plates

were incubated overnight at 37°C, counted, and stored at 4°C for subculturing. Plate counts indicated the mycelium LMW library contained a total of 10,300 clones and the yeast LMW library contained a total of 22,000 clones. The primary stock culture of each library was stored at -80°C in 50 µl aliquots to avoid freeze-thaw cycling during subculturing.

DNA Sequencing and EST BLASTX Analysis

Insert integrity was verified by small-batch sequencing of discrete colonies from titer plates for mycelium LMW and yeast LMW libraries. A total of 12 discrete colonies were picked from the yeast LMW titer plates, along with 16 colonies picked from mycelium LMW, to 5 ml 2YT amended with 50 µg/ml ampicillin (Sigma) and incubated overnight at 37°C with shaking. Plasmid DNA was isolated using a Qiaprep Spin Miniprep kit (Qiagen) as per manufacturer's instructions using a microcentrifuge. Plasmid DNA was eluted with 20 µl dH_20 and presence of insert was verified by double digestion with *Sac*I and *Kpn*I. Preliminary sequencing of 16 mycelium LMW clones and 12 yeast LMW clones was performed in the laboratory of Dr. B. Koop (University of Victoria, Victoria, BC). While preliminary sequencing indicated that the complexity of the mycelium LMW library was of comparable quality, it was decided that deeper sequencing of a single library would be more cost-effective. The yeast LMW library was chosen because it is this growth state that plays a critical role in the transmission of the disease. A better understanding of gene expression during this growth state could contribute to the generation of a spore-deficient mutant with reduced vector transmittability.. The confirmation of authentic inserts of appropriate size at low levels of redundancy justified sequencing deeper into the yeast LMW library on a much larger scale.

A mid-level sequencing assessment was conducted for 384 yeast LMW clones. Clones from the primary yeast LMW cDNA library were prepared for sequencing by plating on 2YT amended with 50 μg/ml ampicillin (Sigma) at a density of 200 colonies/plate. Discrete colonies were transferred to 96-well cell culture plates (Corning) containing 200 μl 2YT amended with 50 μg/ml ampicillin. Cell culture plates were sealed with foil tape (Corning) and incubated overnight at 37°C without shaking, and delivered immediately to the sequencing facility for processing. The mid-level sequencing assessment indicated levels of insert authenticity and redundancy consistent with preliminary analysis. Large-batch samples of yeast LMW cDNA library clones were submitted for sequencing and BLASTX analysis, resulting in a total subsmission of 5760 putative EST clones.

Quality of Sequence Information

Recent advances in biotechnology have vastly accelerated the speed of nucleotide sequencing, facilitating the sequencing of large numbers of clones, or large genomes in relatively short periods of time. While large-scale sequencing has quickly come to the forefront of modern biology, the data generated from these projects never stand alone as a conclusion. Rather, large sequencing projects often serve to generate further questions, in the form of genetic data in need of further processing. The field of bioinformatics has advanced at an equal pace out of necessity for advanced sequence analysis and application tools. Downstream processing of the low-molecular weight yeast-like *O. novo-ulmi* cDNA library began with comparison to existing sequences already submitted to public databases.

In preparation for sequence comparisons the vector DNA were edited from authentic *O. novo-ulmi* sequences. Putative identities were assigned to each clone using the

BLASTX algorithm (Altschul et al., 1997), which compares nucleotide query sequence translated in all 6 reading frames against a public protein database (GenBank NCBI). This algorithm was chosen for its increased sensitivity, due to increased levels of conservation seen in protein sequences as compared to nucleotide sequences as a result of codon redundancy (Dufton, 1983). A low-complexity filter was applied to query sequences to remove regions of low-complexity, such as proline-rich regions or repeats of common acidic or basic residues. The removal of these low-complexity regions increased the fidelity of alignments, and enriched the data for biological significance (Wootton & Federhen, 1993), rather than statistical significance alone.

Incorporated into the BLAST algorithm are variables and parameters researchers can adjust to influence the output of the analysis. One such parameter is the Expected (E) value, defined as the statistical significance threshold for reporting matches against database sequences. An (E) value of 20 indicates that a given sequence of nucleotides is expected to be found in the NCBI GenBank database twenty times by random chance alone, according to the stochastic model of Karlin and Altschul (1990). If the statistical significance assigned to a match exceeds the threshold, the match is considered a chance event and is not reported. As such, lower (E) value thresholds are more stringent, leading to fewer chance matches being mistakenly reported. For the analysis of the *O. novo-ulmi* yeast-like cDNA library, the (E) value threshold was set at 0.1.

EST Database Construction

An online database (http://woodstock.ceh.uvic.ca/sdr/project.cgi/4; username and password available upon request) was constructed to store and analyze the data produced by sequencing 5,760 clones. Information on this database is accessible at the level of

clone, plate (n=60), and library. Included is raw and trimmed sequence information in Fasta format, chromatographic displays for sequencing reactions of each clone, and a summary report illustrating insert length and averages, identity and number of clones containing authentic insert. This database is continually updated by on-going BLASTX (Altschul et al., 1997) analysis in real-time and a BLAST report displays the three highest scoring alignments in public databases (NCBI GenBank). For ease of access, each sequence retrieved from public databases is accompanied by its accession number and hyperlink. For a list of unique ESTs retrieving hits from public databases, see Appendix I – Alphabetized Index of ESTs Matching Entries in Public Databases (GenBank).

Results

The titre of the low-molecular weight yeast *O. novo-ulmi* cDNA library was estimated by sequential dilution plating to be 22,000 cfu. When sequencing was terminated after 5,760 clones, just over 30% of readable sequences were still unique, illustrating the high level of complexity within the library.

Yeast LMW Library Characterization

Insert authenticity and redundancy were assessed by initial sequence analysis of 12 Yeast LMW clones and 16 Mycelium LMW clones. Vector trimming and BLASTX (Altschul et al., 1997) analysis showed that all 28 clones initially sequenced from the *O. novo-ulmi* LMW libraries contained authentic insert ranging in size from 621 bp to 1,156 bp (average 770.7 bp), and that they exhibited zero redundancy. BLASTX analysis showed that the Yeast LMW library contained ESTs exhibiting homology to characterized proteins from fungi and other organisms in public databases. These BLAST hits included a stress-responsive protein from *Fusarium oxysporum*, a methionine synthase from *Aspergillus nidulans*, and a cosmid contig ribosomal protein from *Neurospora crassa*, as well as several unidentified sequences derived from the Human Genome Sequencing Project (International Human Genome Sequencing Consortium., 2004). Following confirmation of authentic insert at low levels of redundancy, deep sequencing of the Yeast LMW library was performed.

High-quality Yeast LMW plasmid DNA was subjected to high-throughput sequencing. In total, 5,760 plasmid EST clones were sequenced, of which 4,386 gave readable sequence information from inserts ranging from 133 base pairs to 690 base pairs, with an average of 498 base pairs. Of these, BLASTX analysis (Altschul et al., 1997) matched

2,093 sequences with entries already found in public databases (NCBI GenBank). These included 1,761 matches with known proteins, and 332 matches with unknown (hypothetical/predicted) proteins. Matches with known proteins included 880 unique sequences (singletons), corresponding to 49.97 % of the EST sequences in this category. Extrapolation of this proportion to matches with unknown (hypothetical/predicted) proteins leads to an estimation of 166 singletons. When considering the 2293 EST sequences drawing no matches by BLAST analysis, 20% of the clones were estimated to contain non-authentic sequences, reducing the total to 1,835 sequences without a match. These were assumed to include a subset of 917 singletons as estimated above. Unique sequences from all categories totalled 1,963, which was estimated to represent 22% of the total genetic load, given that the *O. novo-ulmi* genome is estimated to contain 8K – 10K genes (Galagan et al., 2003; Kupfer et al., 1997).

Assignment of Functional Categories to ESTs

Acknowledging the need for a protein function management system, the Munich Information Centre for Protein Sequences (MIPS, now the Institute for Bioinformatics, Neuherberg, Germany) developed The Functional Catalogue (FunCat) (Ruepp et al., 2004), which has become a standard for the bioinformatics industry. Designed to serve as a stand-alone information management framework, FunCat is a hierarchically structured, organism-independent, flexible, scalable classification system enabling the functional assignment of proteins from any organism. The application of prokaryotic (Ruepp et al., 2000), fungal (Galagan et al., 2003), plant (Salanoubat et al., 2000), and animal (Mi et al., 2003) proteomes has demonstrated the flexibility and utility of this system.

The hierarchical design of FunCat also facilitates subsequent downstream bioinformatics applications, such as the integration of multiple large-scale experiments, prediction of experimental data in functional units, and the comparison of experimental data obtained by different methods. The utility of hierarchically-based classification systems is best demonstrated by the Enzyme Classification (EC) nomenclature first developed in 1956 based on the chemistry of the reaction an enzyme catalyses (Barrett, 1997). While each system is hierarchically-based, EC is based on the physical chemistry of specific reactions while FunCat categorizes proteins according to the physiological role or metabolic pathway in which that protein is found. In this sense, FunCat is a more intuitive, user-friendly system, refined with the end user in mind.

Introduced originally in 1997 (Mewes et al., 1997), FunCat has undergone four revisions, each time evolving to include additional physiological and metabolic functions previously unrepresented by the simple organisms first categorized. While these revisions have added multicellular functionality to facilitate annotation of plant and animal genomes, the fundamental design has not needed alteration. The FunCat scheme is comprised of 28 primary level categories including but not limited to general features such as cellular transport, energy, metabolism, and protein regulation (Appendix I). Each of these primary categories is organized as hierarchical branches, with each sub-branch representing functions of increasing specificity. The FunCat has been likened to a high-level textbook comprised of major sections, chapters, paragraphs, and sentences (Ruepp et al., 2004).

The superset of FunCat currently contains 1307 categories representing six levels of specificity. The nomenclature design reflects the hierarchical design. For example,

secondary FunCat 01.02 represents nitrogen and sulphur metabolism, and includes all categories noted in the manner 01.02.xx.xx.xx.xx, where each level of specificity is delineated by a period. Each category is species-independent, with the exception of FunCat 38 representing transposable elements, and viral and plasmid proteins. The lack of species specificity facilitates the application of FunCat to any organism. While independent of species, FunCat does include subcategories concerned with functional peculiarities that are specific to groups of organisms (Ruepp et al., 2004).

Application of the Functional Catalogue to *O. novo-ulmi*

To create a protein expression profile for an *O. novo-ulmi* yeast-like culture, 4,386 ESTs were BLASTX analyzed, of which 2,093 generated matches with sequences already found in public databases. Expressed sequence tags generating hits were automatically annotated with the three highest scoring alignments. These data were manually scrutinized and each EST annotated using the most meaningful alignment. Within this subset, 1,761 ESTs were annotated with known proteins while 332 matched with hypothetical/predicted proteins. Sequences matching known proteins included 880 singletons, corresponding to 49.97 % of the ESTs in this category. Extrapolating this proportion to the sequences matching hypothetical proteins estimated the number of singletons in this category to be 166. Singletons matching entries in public databases (n=880) were manually annotated using the Functional Catalogue as established by the Munich Information Centre for Protein Sequences (MIPS).

Many proteins are associated with more than one metabolic pathway, and many pathways influence more than one aspect of metabolism. As such, the assignment of a single functional category to a protein would often result in a loss of information, if not

an inaccurate functional assignment. Many multifunctional proteins are justifiably included in numerous functional categories. This results in a small number of proteins generating a massive number of functional assignments, with the absolute occurrence of functional categories being over-represented. To address this issue within this study, a given protein was assumed to assign 1 unit of metabolic function, such that multifunctional proteins would assign a proportional value less than one to each functional category with which it was associated. A protein associated with four FunCat categories would contribute 0.25 points to each category. In this sense, FunCat scores within the Yeast LMW EST library are standardized.

Functional Assignment of *O. novo-ulmi* Yeast LMW ESTs at Primary Level

Application of the FunCat functional catalogue to the *O. novo-ulmi* Yeast LMW library produced functional assignments in 20 primary functional categories (Table 1). FunCat 99 (Unknown) represented ESTs for which an identity could be assigned by homology to proteins in public databases, but not a metabolic function. Expressed sequence tags without an identity were not assigned to a category. Only ESTs with an assigned identity were included in the percent representation, thus this value represents the proportional prevalence of a functional category within ESTs identified through BLASTX analysis. FunCat scores for primary functional categories ranged from 0.70% to 20.73%, with a mean of 6.20%. FunCat 01 (Metabolism) scored with the highest frequency (20.73%), followed by FunCat 12 (Protein Synthesis, 10%), FunCat 70 (Subcellular Localization,

Table 1. Representation and Percent Representation of Functional Categories of *O.novo-ulmi* Yeast LMW ESTs at the Primary Level. A total of 880 assignable ESTs were standardized to 861.25 functional assignments due to multifunctionality. This table represents 16 of 20 primary functional categories annotated; those comprising less than 0.5% representation were omitted for clarity.

FunCat #	Representation	% Total Representation
01 - Metabolism	178.5	20.73
02 - Energy	39.5	4.59
10 - Cell Cycle & DNA Processing	6	0.70
11 - Transcription	72.75	8.45
12 - Protein Synthesis	89.25	10.36
14 - Protein Fate	7.5	0.87
16 - Binding Proteins & Cofactor Dependent	12.25	1.42
18 - Protein Activity Regulation	7.25	0.84
20 - Cell Transport, Facilitation, Route	53.5	6.21
32 - Cell Rescue, Defense, Virulence	46.75	5.43
34 - Cellular Environment Interaction	32	3.72
42 - Biogenesis of Cell Components	71.75	8.33
70 - Subcellular Localization	90	10.45
73 - Cell Type Localization	30	3.48
98 - Classification Unresolved	51	5.92
99 - Unknown	66	7.66
Sum	854	99.16
FunCat's 36, 38, 40, 43	7.25	0.85
Total	861.25	100.01

10%), FunCat 42 (Biogenesis of Cell Components, 8%), and FunCat 11 (Transcription, 8%). FunCat 99 (Unknown) represented 7.66% of identified ESTs. Figure 1 illustrates the relative representation within the *O. novo-ulmi* Yeast LMW library of each primary functional category.

Least represented was FunCat 43 (Cell Type Differentiation), representing 0.12% of identified ESTs. FunCat's 36 (Systemic Environment Interaction), 38 (Transposable Elements, Viral, Plasmid), and 40 (Cell Fate) scored similarly low, representing 0.14%, 0.35%, and 0.53% respectively, of ESTs. Table 1 shows that FunCat 10 (Cell Cycle & DNA Processing) was also found to contribute minimally to functional assignments, representing 0.70% of identified ESTs, as were FunCat's 14 (Protein Fate, 0.84%), and 18 (Protein Activity Regulation, 0.84%).

Functional Assignment of *O. novo-ulmi* Yeast LMW ESTs at Secondary Level

The MIPS FunCat functional catalogue is of a scalable, hierarchical design, allowing assignments of increasing specificities as deeper subcategories of the catalogue are examined. The ESTs within the eight primary categories representing greater than 4.50% of identified ESTs were assigned functional categories at the secondary FunCat level (Figs. 2 through 9). Secondary functional categories represented in this group exhibited wide variation both in the number of subcategories within them, and the proportional distribution of their subcategories.

Assignment of secondary functional categories to FunCat 01 (Metabolism) was comprised of 178.5 standardized functional assignments of identified ESTs (Fig. 2). FunCat 01 was the most highly represented functional category, comprising 20.73% of the assignable ESTs within the *O. novo-ulmi* Yeast LMW library. Within this primary

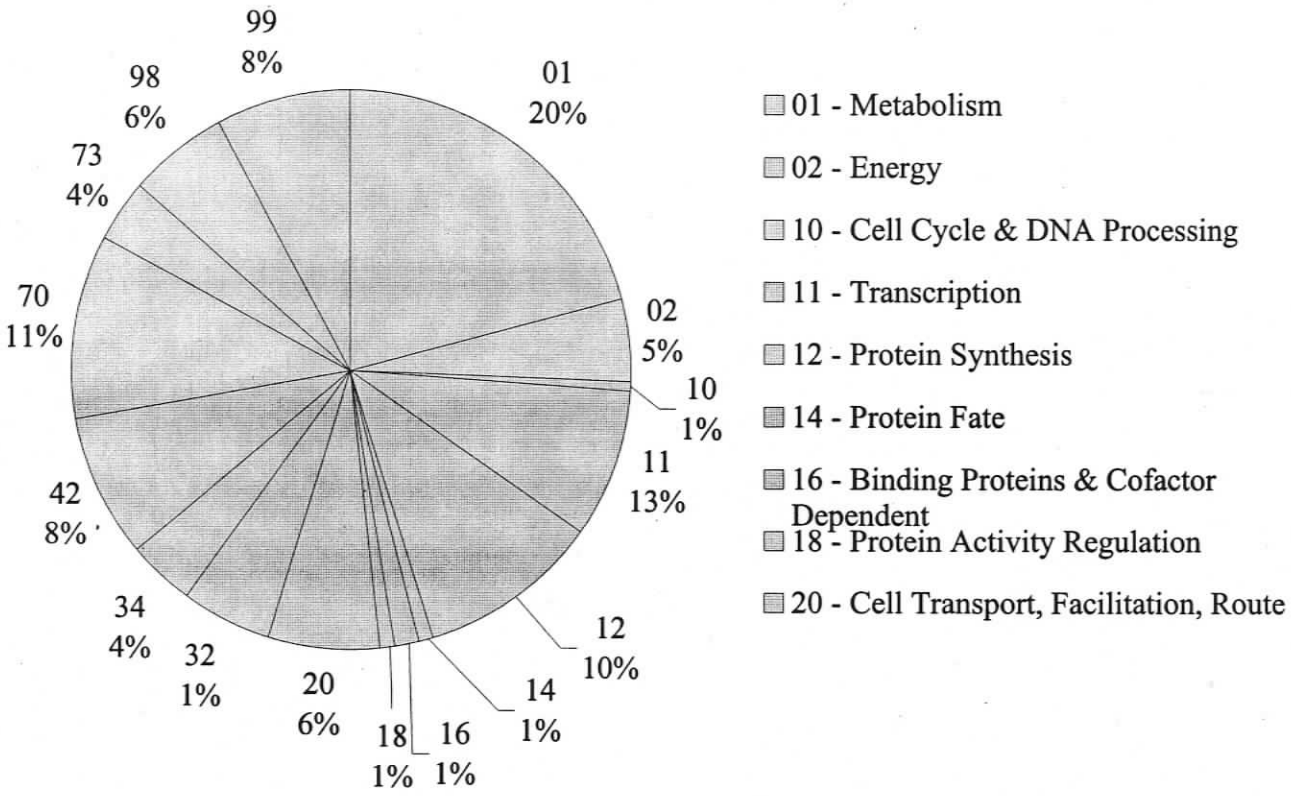


Figure 1. Representation of Functional Categories at the Primary Level. A total of 880 assignable ESTs were standardized to 861.25 functional assignments due to multifunctionality. This chart represents 16 of 20 primary functional categories annotated; those comprising less than 0.5% representation werer omitted for clarity.

Figure 2. Percent Representation of Functional Category 01 (Metabolism) at the Secondary Level. Functional Category 01 was comprised of 178.5 assignable ESTs, representing 20.7% of functional assignments overall. This category demonstrated the second greatest level of complexity at the secondary level.

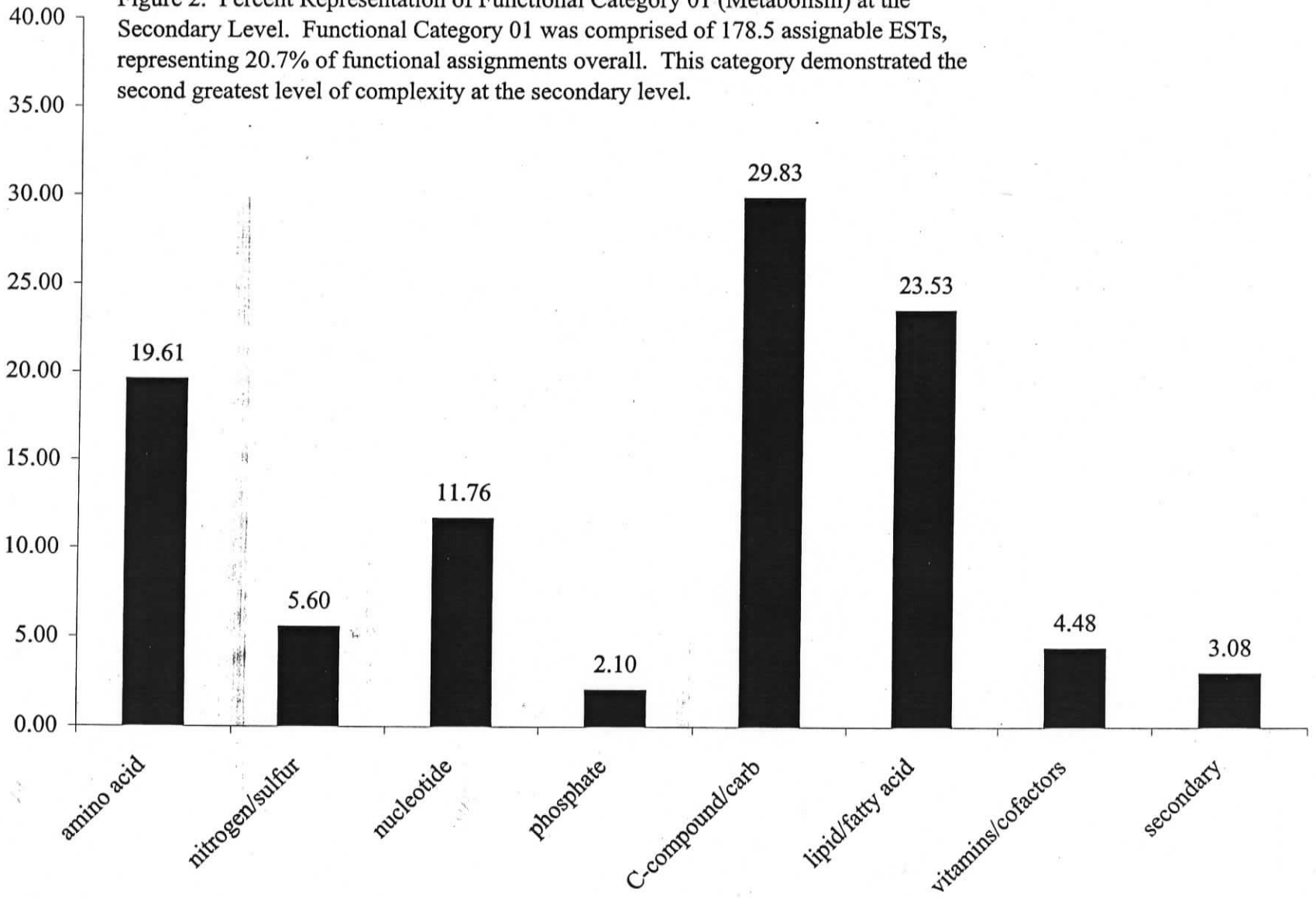


Figure 3. Percent Representation of Primary Functional Category 02 (Energy) at the Secondary Level. Functional Category 02 included of 38.75 standardized functional assignments, constituting 4.61% of total functional assignments.

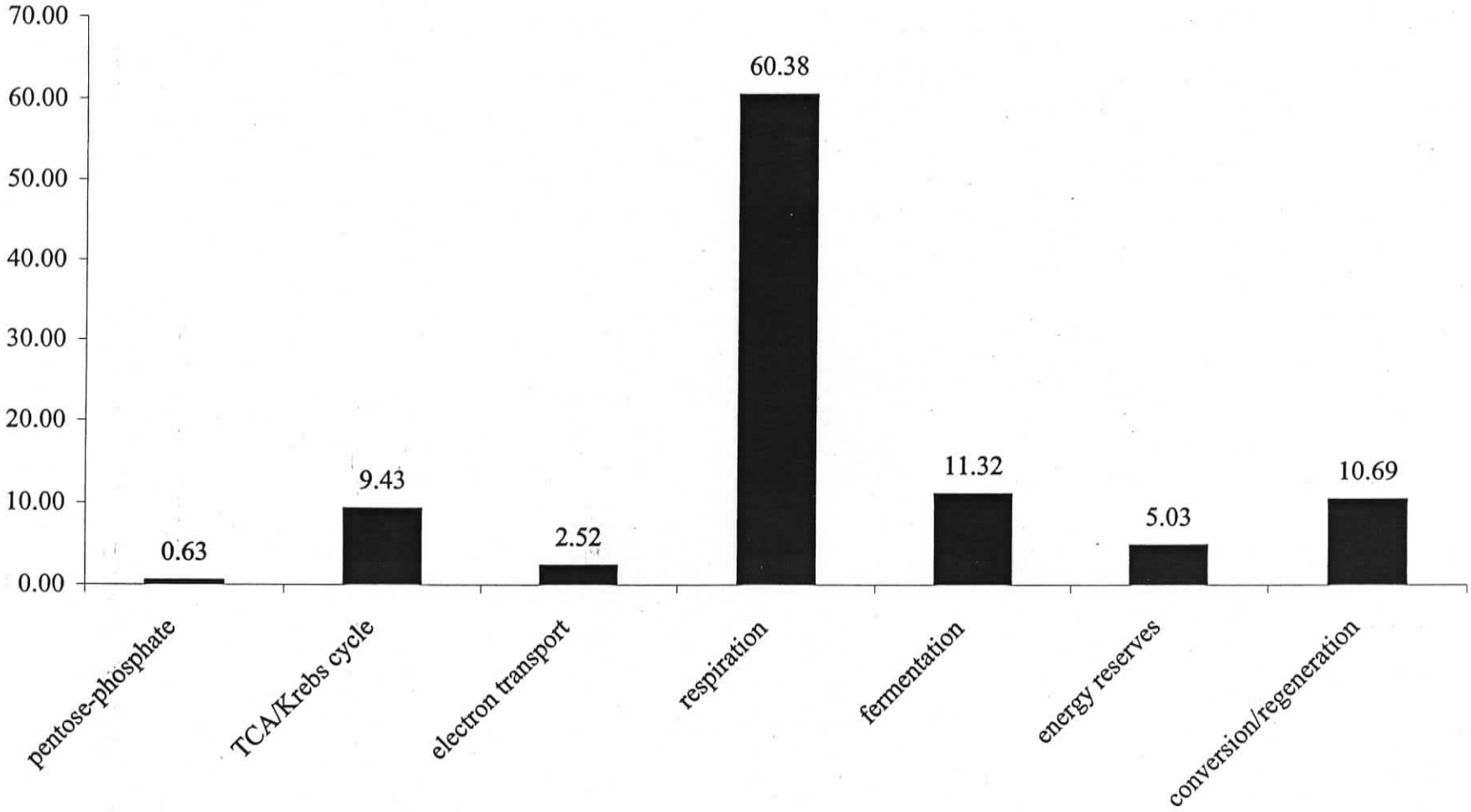


Figure 4. Percent Representation of Functional Category 11 (Transcription) at the Secondary Level. Functional Category 11 was comprised of 72.75 assignable ESTs, representing 8.4% of functional assignments overall. This category demonstrated the lowest level of complexity at the secondary level.

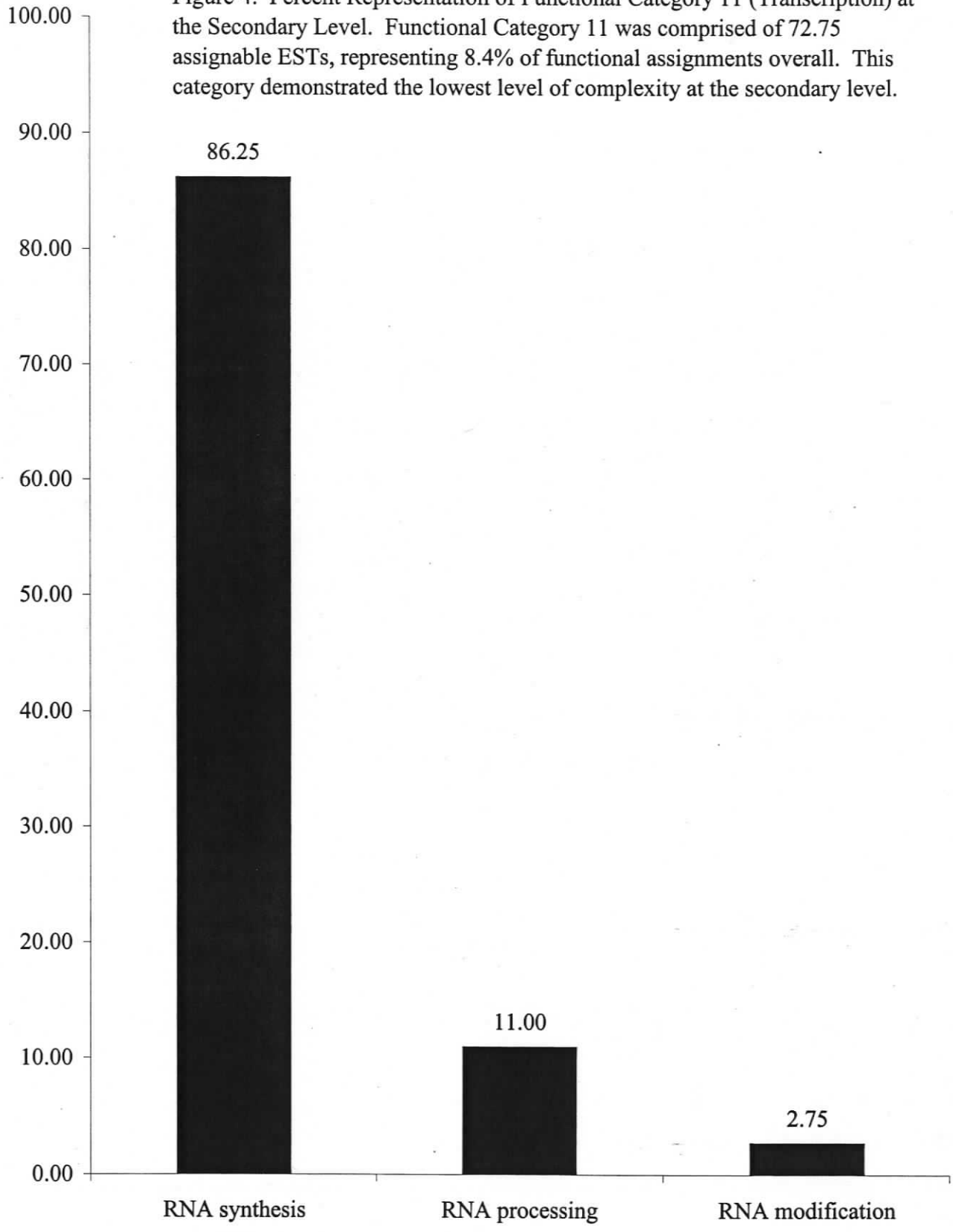


Figure 5. Percent Representation of Functional Category 12 (Protein Synthesis) at the Secondary Level. Functional Category 12 was comprised of 89.25 assignable ESTs, representing 10.4% of functional assignments overall.

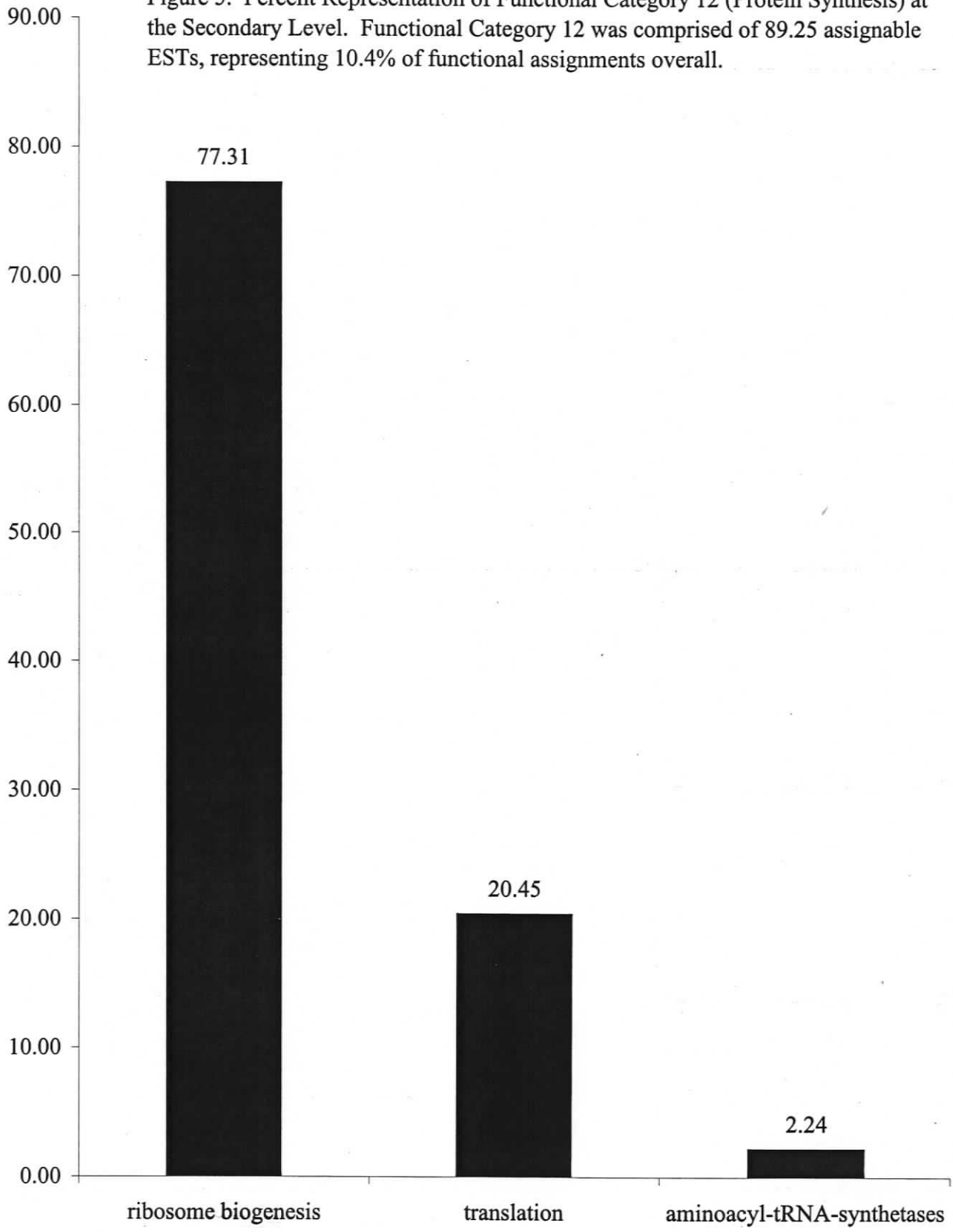
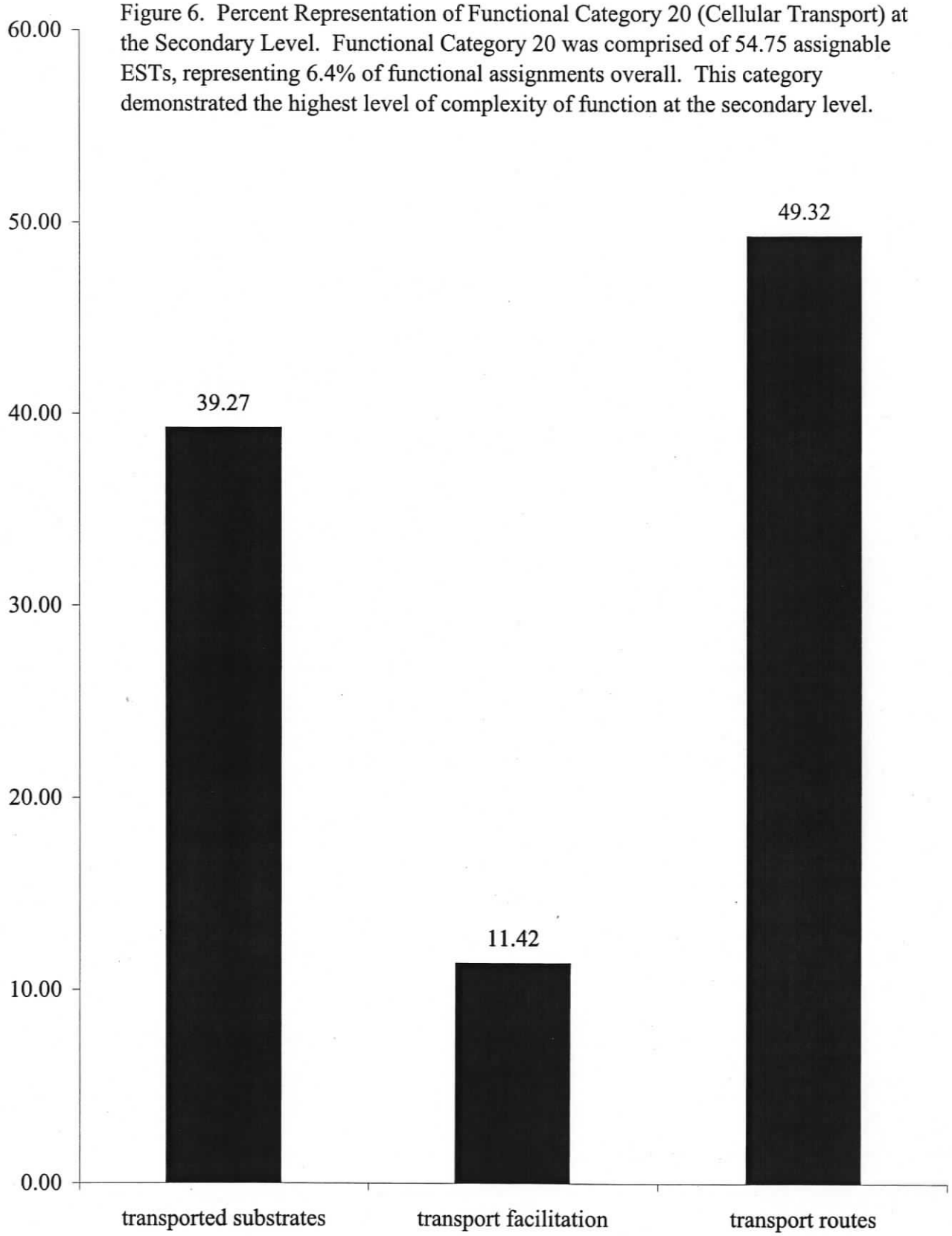


Figure 6. Percent Representation of Functional Category 20 (Cellular Transport) at the Secondary Level. Functional Category 20 was comprised of 54.75 assignable ESTs, representing 6.4% of functional assignments overall. This category demonstrated the highest level of complexity of function at the secondary level.



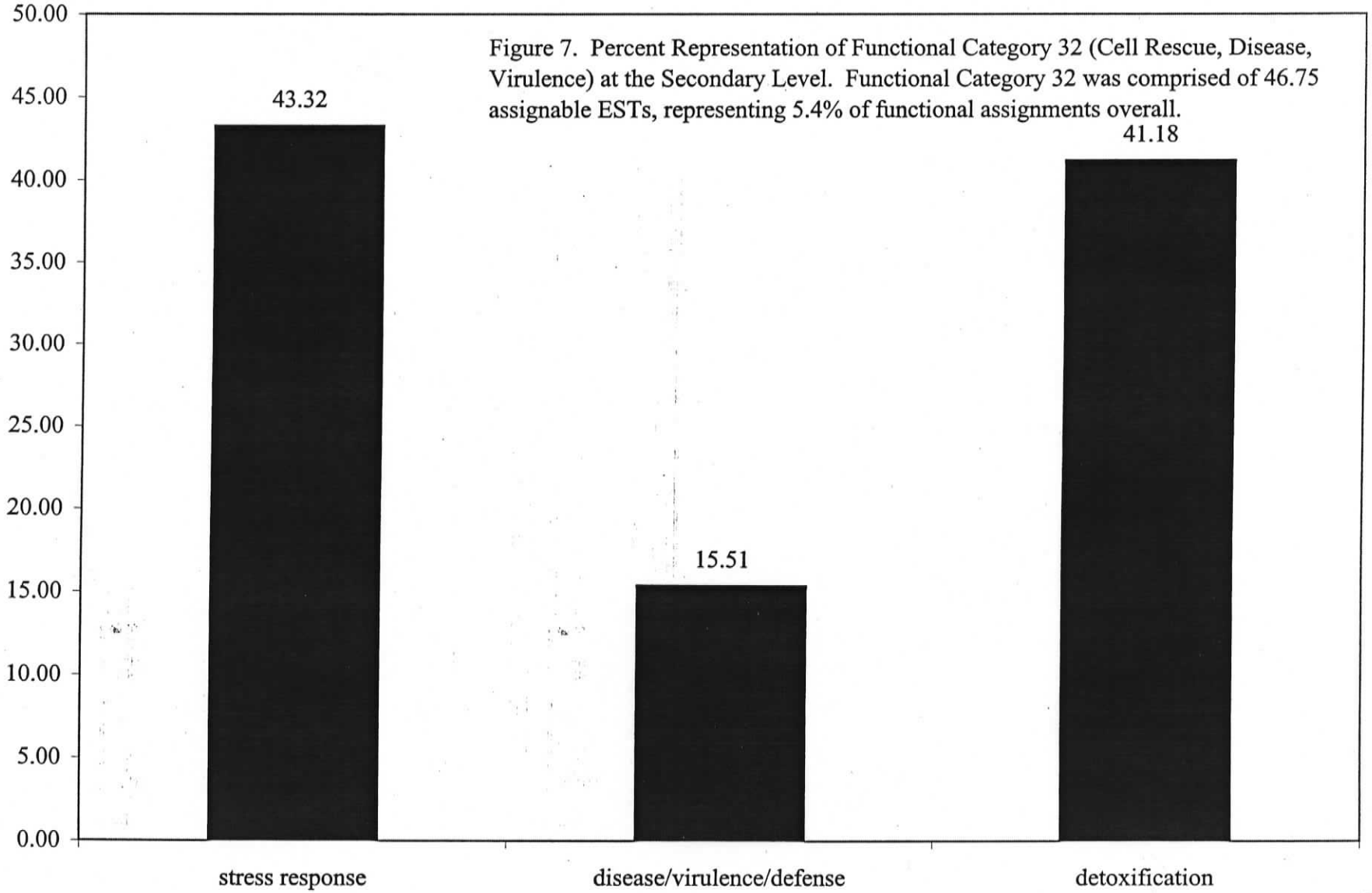


Figure 7. Percent Representation of Functional Category 32 (Cell Rescue, Disease, Virulence) at the Secondary Level. Functional Category 32 was comprised of 46.75 assignable ESTs, representing 5.4% of functional assignments overall.

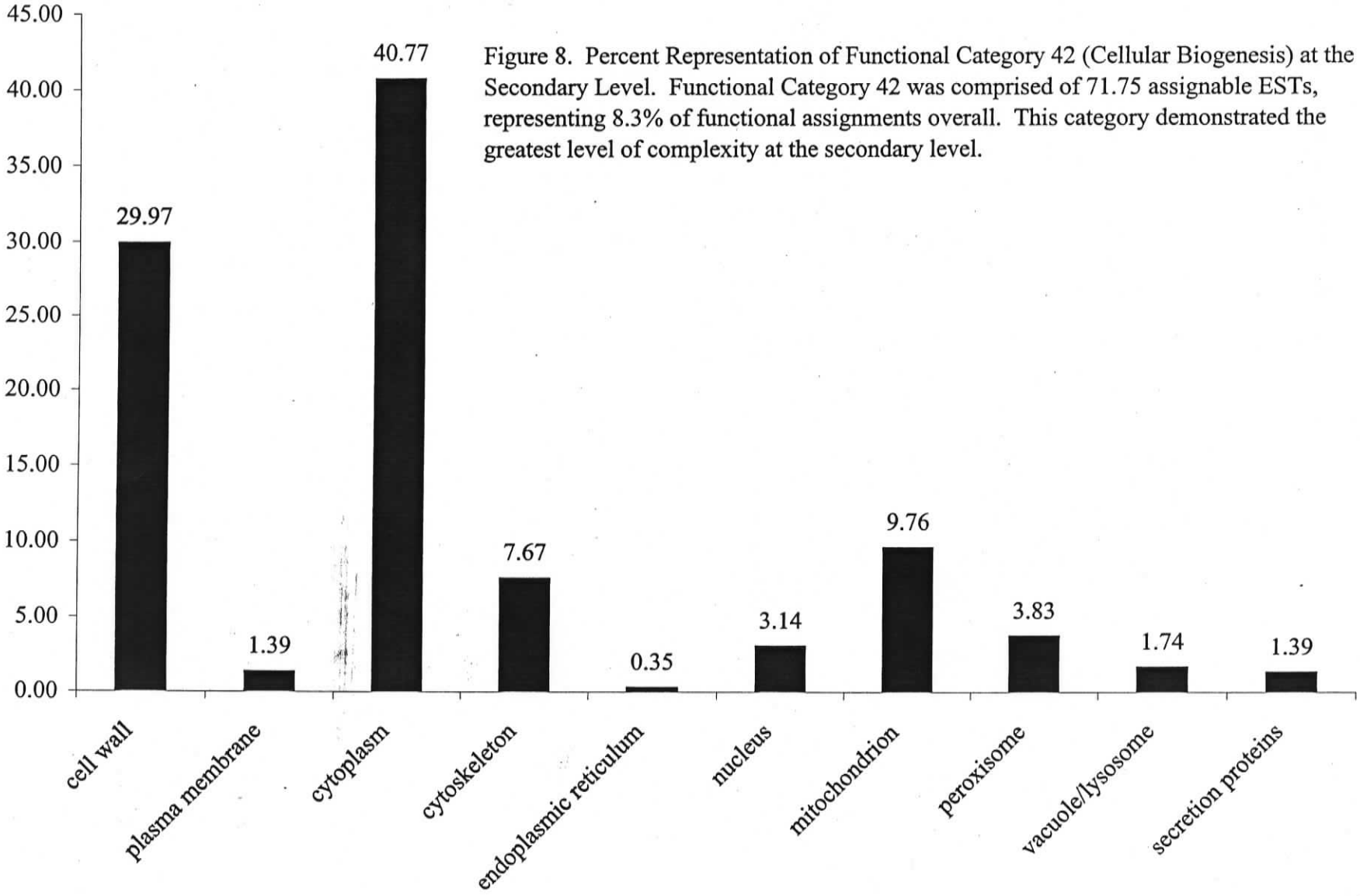
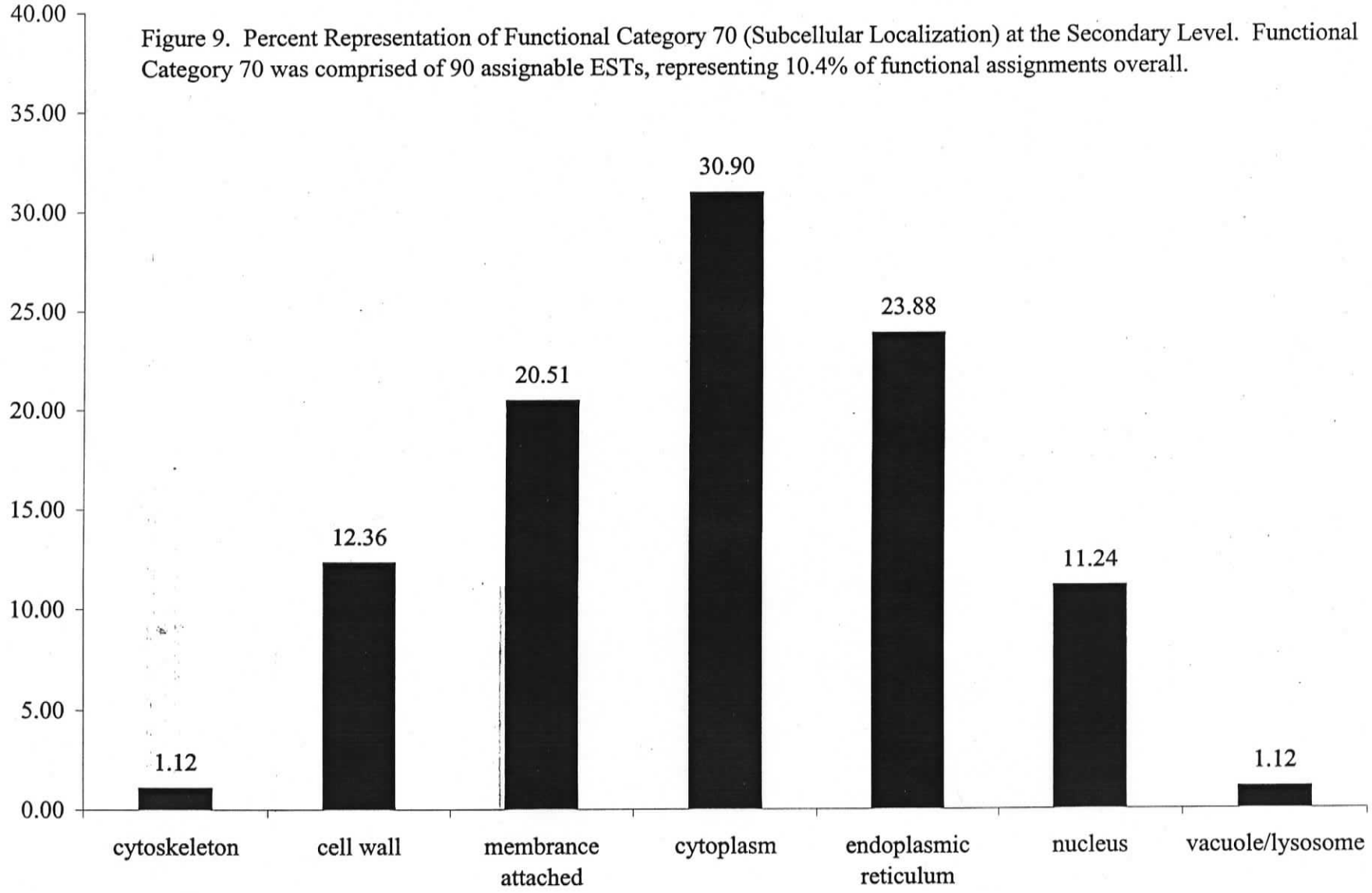


Figure 8. Percent Representation of Functional Category 42 (Cellular Biogenesis) at the Secondary Level. Functional Category 42 was comprised of 71.75 assignable ESTs, representing 8.3% of functional assignments overall. This category demonstrated the greatest level of complexity at the secondary level.

Figure 9. Percent Representation of Functional Category 70 (Subcellular Localization) at the Secondary Level. Functional Category 70 was comprised of 90 assignable ESTs, representing 10.4% of functional assignments overall.



category, eight subcategories relating to fungal metabolism were represented to varying degrees. Expressed sequence tags associated with carbon compound metabolism were most highly represented, comprising 29.83% of FunCat 01 (Metabolism). Included in this subcategory were glucoamylase I and an alphamannosidase, each of which have been previously characterized in filamentous fungi. Glucoamylase is often expressed at high levels in the presence of glucose, starch, and a variety of hexose sugars (Nunberg et al., 1984). Fusion proteins of glucoamylase have been utlized to examine secretion in filamentous fungi (Jeenes et al., 1993). Many Class I alpha-mannosidases exist as Type II transmembrane proteins (Eades & Hintz, 2000) and exhibit high activity towards high-mannose oligosaccharides (Yoshida et al., 2000). Alpha-mannosidases are critical proteins for high fidelity glycosylation of proteins, and as such, are of significant interest within therapeutics. The metabolism of fatty acids and amino acids were also highly represented, comprising 23.53% and 19.61 %, respectively (Fig. 2). These three subcategories alone accounted for 72.97% of FunCat 01. Tags associated with nucleotide metabolism were also well represented at 11.76%. Guanylate kinase, a nucleotide binding protein involved in the transfer of HPO_3 groups between nucleotides (Prinz et al., 1999), was among several nucleotide metabolism genes identified in the Yeast LMW EST library. The remaining subcategories relating to nitrogen/sulphur, phosphate, vitamins/cofactors, and secondary metabolism each comprised less than 5.6% of FunCat 01 (Metabolism).

While a total of seven subcategories were represented within FunCat 02 (Energy), the second most highly represented subcategory (fermentation, 11.32%) was exceeded almost 6-fold by respiration genes. High copy numbers of cytochrome genes contributed

to elevated representation of respiration genes. Cytochrome genes are generally expressed at high levels, as respiration is a primary function necessary for survival. One example of this is the Cox5 gene, which codes for one peptide of the cytochrome C oxidase complex. Interestingly, recent studies in the filamentous fungus *Podospora anserina* have correlated loss of cytochrome C oxidase function with an increase in longevity (Dufour et al., 2000). In these mutants, loss of cytochrome C oxidase function forces the up-regulation of alternate energy-generating pathways with a lower free-radical production rate. Fewer free-radical species within the organism reduces the accumulation of damage to mitochondrial DNA, extending the longevity as much as 20-fold. Aconitase was also identified in the *O. nov-ulmi* LMW Yeast EST library, and represents a second well characterized respiration gene garnishing new interest.

Aconitase is active within the citric acid cycle to isomerize citrate to isocitrate (Hausladen & Fridovich, 1994), and is a critical component of cellular respiration. Aconitase has also been examined recently in *Cryptococcus neoformans*, a mammalian pathogen, where it was shown to be subject to post-translational modifications in response to nitric oxide stress (Missall et al., 2006). Mutants defective in similarly modified proteins were shown to exhibit increased susceptibility to phagocytosis and reduced virulence (Missall et al., 2006). Given the importance of this gene in the TCA cycle, it is unlikely to serve directly as a target for virulence control on *O. novo-ulmi*, but further examination of this gene product could elucidate similarly modified proteins to serve as potential targets for control of this tree pathogen. Figure 3 shows that subcategories within FunCat 02 (Energy) existed in three tiers, where respiration (60.38%) genes dominated the highest level, followed by fermentation (11.32%), energy

conversion/regeneration (10.69%), and Krebs cycle (9.43%) genes comprising a middle tier, with genes associated with energy reserves (5.03%), electron transport (2.52%), and the pentose-phosphate pathway (0.63%) exhibiting the least representation within FunCat 02 (Energy).

Of the eight primary functional categories examined at the secondary level, FunCat 11 (Transcription) exhibited the least amount of complexity within its subcategories, with ESTs associated with RNA synthesis comprising 86.25% of this primary functional category (Fig. 4). In the *O. novo-ulmi* Yeast LMW library, this subcategory was comprised not only of multiple histone proteins including H2A, H3 and H4, but also splicing and transcription factors. Zinc finger proteins and ATP-dependent RNA helicases also contributed to the RNA synthesis cubcategory. The subcategories within FunCat 11 were represented by a total of 110 standardized functional assignments. Not only was FunCat 11 dominated by a single subcategory, it was comprised of only three subcategories. Expressed sequence tags associated with RNA processing comprised 11.00% of FunCat 11, while those associated with RNA modification only contributed 2.75%.

The expression profile for FunCat 12 (Protein Synthesis) closely resembled that of FunCat 11, as can be seen in Figures 5 and 4, respectively. Expressed sequence tags associated with protein synthesis were assigned to one of three subcategories, with those associated with ribosome biogenesis representing 77.31% of FunCat 12. As is often the case when examining protein synthesis, genes in this subcategory were dominated by 40S and 60S ribosomal proteins. While ribosome biogenesis ESTs comprised the bulk of this primary functional category, ESTs associated with translation were represented at

20.45% of FunCat 12 (Fig. 5). Tags associated with aminoacyl-tRNA synthetases comprised only 2.24% of the protein synthesis primary category.

Assignment of Yeast LMW ESTs to secondary functional categories within FunCat 20 (Cellular Transport) included 54.75 standardized functional assignments. This primary functional category was comprised of three subcategories, two of which were highly represented (Fig. 6). Expressed tags associated with cellular transport routes were most highly represented, comprising 49.32% of FunCat 20, and included nuclear transport factor 2 and carbonic anhydrase. The conversion of CO_2 to HCO_3 is catalysed a million fold by carbonic anhydrase. The expression of this important gene product has recently been induced in *Saccharomyces cerevisiae* by controlling the concentration of inorganic carbon in culture medium (Amoroso et al., 2005). Tags associated with transported substrates were also highly represented at 39.27%. Substrate-transport proteins in *O. novo-ulmi* included the arginine/ornithine antiporter, which facilitates the stoichiometric exchange of arginine for ornithine across a semi-permeable membrane (Arena et al., 2002). Also present in *O. novo-ulmi* was the amino acid transporter, responsible for the transport of amino acids across a plasma membrane, with or without efflux. Many amino acid transporters are Na+ or pH dependent; confidently recording the dependencies of the amino acid transporter found in the *O. novo-ulmi* Yeast LMW library would be difficult without further examination. The two subcategories representing cellular transport routes and transported substrates comprised 88.59% of FunCat 20 (Cellular Transport), with ESTs associated with transport facilitation comprising the remaining 11.41%.

The greatest amount of parity between subcategories was observed within FunCat 32 (Cell Rescue, Disease, Virulence), as shown in Figure 7. A total of 46.75 standardized

functional assignments comprised three subcategories within this primary functional category, which represented 5.4% of total functional assignments. Expressed sequence tags associated with stress response and detoxification were almost equally represented, at 43.32% and 41.18%, respectively. As expected, the ESTs associated with stress response represented inducible gene products sensitive to environmental stimuli such as UV irradiation, desiccation, and heat shock. Included in this subcategory was the UV-inducible gene uvi31, which has been previously characterized and induced in *Saccharomyces pombe* (Kim et al., 2002). Null mutants at *uvi31* were shown to exhibit an elevated vegetative growth rate with a reduced cell size. Not only was ultraviolet sensitivity elevated in *uvi31* mutants, but defects in cytokinesis and septum formation were also observed following cessation of UV treatment. This represents a locus with an inducible lethal mutation, which is of significant interest when studying a pathogenic fungus such as *O. novo-ulmi.* It is believed that uvi31 is involved with control of cell division, particularly in the resumption of division following cell cycle arrest. What will be of critical importance when assessing the practicality for this locus as a mode of control is the specificity to this organism. A pathogen gene target sharing sufficient similarity with a host gene to prevent species-specific discrimination is of no use. It should be noted that while *uvi31* has been shown to exhibit homology to ORFs from bacteria, *Arabidopsis thaliana, Mus musculla,* and *Homo sapiens,* it remains unclear if sufficient specificity exists for a potential target of control. With only three subcategories represented within FunCat 32, the remaining 15.51% of tags were associated with disease/virulence/defense.

The greatest number of subcategories represented was observed within FunCat 42 (Cellular Biogenesis), as shown in Figure 8. A total of 71.75 standardized functional assignments were distributed through ten subcategories, although four of these subcategories accounted for 88.17% of this primary functional category. Of these functional assignments, ESTs associated with cytoplasm biogenesis represented the greatest proportion of FunCat 42, at 40.77% and included BNS1 protein and DC13 protein. Cell wall biogenesis genes were also highly represented and included stomatin, mucin, and cell wall surface anchor family protein. Cell wall biogenesis genes comprised 29.97% of the Cellular Biogenesis primary functional category. Expressed sequence tags associated with mitochondrion biogenesis comprised 9.76% of FunCat 42, followed by ESTs associated with cytoskeleton biogenesis, comprising 7.67% of total cellular biogenesis functional assignments (Fig. 8). The six subcategories representing peroxisome (3.83%), nucleus (3.14%), vacuole/lysosome (1.74%), secretion proteins (1.39%), plasma membrane (1.39%), and endoplasmic reticulum (0.35%) biogenesis accounted for the remaining 19.83% of FunCat 42 functional assignments.

Assignment of secondary functional categories to FunCat 70 (Subcellular Localization) included 90 standardized functional assignments, distributed among seven subcategories (Fig. 9). The greatest proportion of ESTs were associated with cytoplasm localization (30.90%), and included chromosome segregation protein pcs1, ubiquitin-protein ligase, and N-acetylmuramoyl-L-alanine amidase AMIC precursor. Beyond this single dominant subcategory, ESTs associated with endoplasmic reticulum localization (23.88%), and the localization of membrane attached subcellular structures (20.51%) were most highly represented. Expressed sequence tags associated with the cell wall

comprised 12.36% of subcellular localization functional assignments, while nucleus localization ESTs comprised 11.24%. Subcategories representing genes associated with vacuole/lysosome and cytoskeleton localization each comprised 1.12% of assignments to FunCat 70 (Subcellular Localization).

Comparative Analysis of *O. novo-ulmi* and *N. crassa* Functional Catalogues

An ideal reference set for comparative analysis of the *O. novo-ulmi* EST functional catalogue would be a functional catalogue generated from the ESTs of a related filamentous fungus. This would allow a direct comparison of expression profiles in related fungi under similar growth conditions. As of the time of writing, no such EST-based catalogue exists within the MIPS database for direct comparison. Further, full genome sequencing of *O. novo-ulmi* has not been completed and a genomic dataset from *O. novo-ulmi* is not available for comparative analysis at this time. However, the completion of the *Neurospora crassa* genome sequencing project has facilitated the construction of a complete FunCat functional catalogue for all open reading frames. In the absence of a complete *O. novo-ulmi* genomic database or EST-based dataset in a related filamentous fungus for reference purposes, the MIPS *Neurospora crassa* genomic database (http://mips.gsf.de/genre/proj/ncrassa/) was utilized for an initial comparative analysis of the *O. novo-ulmi* Yeast LMW EST functional catalogue. The *N. crassa* genomic database provides an approximation of what the related *O. novo-ulmi* genome is likely to comprise, and thus provides a representation of total potential gene expression versus actual gene expression for the *O. novo-ulmi* EST library. Figure 10 illustrates a comparative analysis of primary functional categories from each dataset.

Figure 10. Comparative analysis of *N. crassa* and *O. novo-ulmi* Yeast LMW EST functional catalogues. A total of 7,350 functional assignments were included for N. crassa, while the O. novo-ulmi functional catalogue included 861.25 functional assignments.

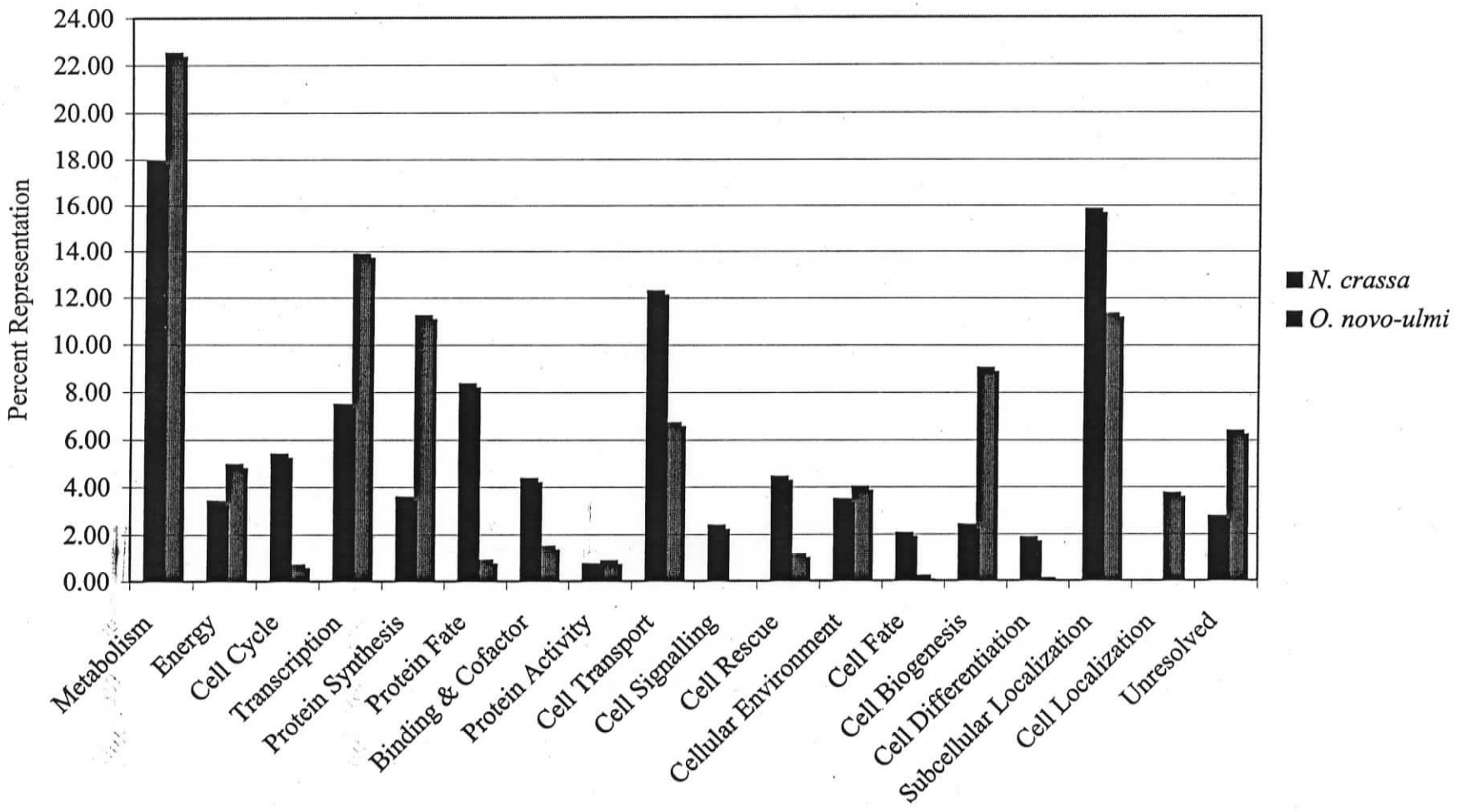


Figure 10 shows that similarity between *O. novo-ulmi* and *N. crassa* functional datasets was inconsistent across categories. It is useful to consider an increase in percent representation as a 'focus' within the *O. novo-ulmi* functional catalogue. In comparison to the *N. crassa*genome database, the *O. novo-ulmi* functional catalogue exhibited a greater percent representation of FunCat 01 (Metabolism), and as such, the expression profile of *O. novo-ulmi* had a greater focus on metabolism. Out of 18 primary functional categories, the *O. novo-ulmi* catalogue exhibited a greater percent representation in eight categories. In some cases the difference between datasets was minor, in other cases representation of a primary category in the *O. novo-ulmi* functional catalogue was eight-fold less than the represenation in the *N. crassa* genomic database (Fig. 10, Protein Fate). Since the *O. novo-ulmi* represented ESTs as opposed to a genomic database for *N. crassa* this type of discrepancy is not unexpected. The EST library shows a bias for highly expressed genes which might shift the percent representation in the various functional categories.

Other categories that exhibited significant discrepancies in representation between the *O. novo-ulmi* EST functional catalogue and the *N. crassa* genomic database included Cellular Biogenesis (*O. novo-ulmi* 9.07%, *N. crassa* 2.44%), Protein Synthesis (*O. novo-ulmi* 11.28%, *N. crassa* 3.62%), and Transcription (*O. novo-ulmi* 13.91%, *N. crassa* 9.44%). It was interesting to note that the Cell Localization functional category was not represented in the *N. crassa* genomic database, while it comprised 3.79% of the *O. novo-ulmi* functional catalogue.

Conversely, some categories were represented at much higher levels in the *N. crassa* genomic database than in the *O. novo-ulmi* functional catalogue. These primary

categories included Protein Fate (*N. crassa* 8.39%, *O. novo-ulmi* 0.95%), Cell Transport (*N. crassa* 12.33%, *O. novo-ulmi 6.76%),* and Subcellular Localization (*N. crassa* 15.88%, *O. novo-ulmi* 11.38%). Primary categories roughly equally represented in each functional database included Metabolism, Energy, and Cellular Environment (Figure 10).

Discussion

High throughput sequencing projects can generate information which can be used to relate gene expression to life cycle strategies for organisms. Sequence tags need to be assigned identifiers so that comparisons to other data sets can be made to infer biological significance. Identifiers are most often assigned on the basis of pairwise sequence comparisons to published sequences already existing in public databases. With advances in sequencing throughput have come automated techniques for the assignment of identifiers to novel sequences. While automated annotation systems offer significant reductions in processing time as compared to manual annotation of datasets, the algorithms existing today are in need of further refinement before they offer a product incapable of improvement by manual review.

Limitations of Automated Functional Category Annotation

Genetic analysis of novel sequence data often begins with BLASTX (Altschul et al., 1997) analysis to assign putative identities based on alignments with derived proteins in public databases. The alignment matches must often be improved by manual scrutiny to increase the probability of a correct identification, or annotation. Automated annotation algorithms have not yet been sufficiently refined to ensure the exclusion of erroneous false positive matches. Even more challenging is the development of an algorithm capable of recognizing when the highest alignment score may not reflect the most reasonable or meaningful match. For example, recent genome sequencing projects have resulted in the deposition of hundreds of thousands of theoretical proteins, predicted by computer analysis from sequenced genomes. Theoretical proteins frequently match with novel ESTs at a high alignment score, but are of little consequence if they do not assign

function or identity to the EST. It is expected that these assignments will improve with a further decoration of the database. In many cases a protein of known function or identity provides more meaningful information, even when there is a lower alignment score. These limitations mean that automated alignment and annotation algorithms serve to provide a good approximation of most EST identities, but manual scrutiny and annotation is necessary to improve fidelity.

Annotation of EST libraries often not only assigns putative identity, but also assignment or categorization of function as well, whenever possible. In some instances the presence of a characteristic functional group or structural domain indicates the probable molecular mechanism of a protein, but offers no insight to the physiological function that protein serves (Ruepp et al., 2004). Similarly, the catalytic nature of biological enzymes dictates that chemical reactions are facilitated equally in both directions.. Biochemical reactions *in vivo* are generally driven preferentially in only one direction by cellular physiology, and this preference can not be predicted from structural analysis of a protein. Glutamate dehydrogenase has been shown to exhibit both anabolic and catabolic activities within liver mitochondria under different metabolic states (Michal, 1998). Similar instances of selective bidirectional activity have been described for proteins found in multiple tissues.

Functional assignment requires consideration of the metabolic context each protein functions within. The specific molecular mechanism of a protein is not always known. In some instances, however, inferences regarding the pathway affected or physiological role can be made based on conservation of concensus sequences among related proteins (Ruepp et al., 2004). Recognition sequences and cellular transport signals are often

shared among related proteins and provide a means of deducing a putative physiological role, when compared to previously categorized proteins bearing similar concensus sequences. In this way, biological function can be assigned without knowledge of molecular mechanisms.

Automated annotation alogorithms utilizing BLASTX analysis identify proteins of similar sequence or those containing shared concensus sequences, of known or unknown biological function, suggesting putative functional assignments. Manual scrutiny of retrieved protein sequences, identities, and functions greatly enhance the fidelity of putative assignments by reduction of false positive matches and recognition of lesser scoring alignments that ultimately provide more meaningful information.

An *O. novo-ulmi* Functional Catalogue

Plasmid purified from an expressed sequence tag library constructed from poly-A mRNA from *Ophiostoma novo-ulmi* cultured to induce the yeast-like isoform was subjected to high-throughput sequencing and functional assignment in accordance with the MIPS FunCat functional catalogue. Multifunctional proteins were assumed to contribute one unit of functional assignment, proportionally distributed among the functional categories with which they were associated. This assumption generated 861.75 standardized functional assignments from a total of 880 ESTs.

The discrepancy representative of 18.75 missing functional assignments corresponds to 2.13% of total ESTs, and can be attributed to error introduced during the standardization of multifunctional assignments. In instances where an EST was associated with more than four secondary functional categories, each category was allotted 0.1 units of functional assignment. This was done after visual inspection indicated that

multifunctional assignments in many cases involved between six and twenty secondary functional categories. Accurate tallying, averaging, and scoring of functional assignments by manual means of up to 20 secondary categories for 880 clones became unfeasible and, as such, multifunctional assignments were assumed to allot 0.1 units to each subcategory when ESTs were associated with more than four subcategories. This resulted in allotment of less than 1 total units for ESTs associated with more than four and less than ten secondary functional categories. Considering that less than 880 units of function were assigned, it appears that of the multifunctional ESTs associated with more than four subcategories, the proportion associated with less than ten subcategories exceeds the proportion associated with more than ten subcategories. As automated algorithms continue to improve, the accurate tallying and scoring of functional assignments from increasingly large datasets will become increasingly feasible. This will drastically reduce the discrepancies in clone numbers versus functional assignments introduced by manual standardization of multifunctional assignments.

When discussing the analysis of biological pathways downstream of RNA synthesis, it is imperative to consider the cellular context the samples were drawn from. Since the transcription of the majority of genes is non-constitutive, an RNA pool serves as a physiological snapshot from within a population of cells at a specific point in time. The population of RNA molecules contained within a single cell will change over time, often rapidly, and will change in response to environmental and molecular stimuli. It should be noted that higher representations of functional assignments within a category indicate elevated levels of transcription complexity for that group, and not transcription

frequency. Since only singleton ESTs were included in FunCat analysis, high values indicate a larger number of different tags within a category.

Secondary Functional Category Assignments

Expressed sequence tags associated with FunCat 01 (Metabolism) represented the greatest proportion of functional assignments within the *O. novo-ulmi* Yeast LMW EST library (Fig. 1). FunCat 01 was comprised of 178.5 standardized functional assignments corresponding to 20.73% of total functional assignments. Of the functional assignments within FunCat 01 (Metabolism), 72.97% were associated with only three subcategories (Fig. 2). Carbon compound, fatty acid, and amino acid metabolism were all highly represented. This is not suprising considering the metabolic demands of a laboratory culture actively growing in exponential phase. Physiological demands during exponential growth are high, and it appears that *O. novo-ulmi* focuses its energy on metabolic building blocks while actively growing. The metabolism of vitamins, cofactors, nitrogen, sulphur, and phosphate compounds, along with secondary metabolism, are often associated with stationary growth. This is in agreement with the limited representation of each of these groups within FunCat 01 (Fig. 2). This primary functional category would be expected to be particularly senstitive to culture conditions and growth phases, and expression profiles from older, stationary cultures might exhibit a significantly higher representation of ESTs associated with secondary metabolism.

A single secondary subcategory within FunCat 02 (Energy) comprised 60.38% of energetic functional assignments (Fig. 3). Expressed sequence tags associated with respiration comprised the vast majority of energy functional assignments, exceeding the second most highly represented group nearly 6-fold. This does not necessarily indicate

that expression levels for respiration genes were exceedingly high; rather, it indicates that a high number of different respiration genes had been transcribed at some level. Respiration functional assignments were comprised of a highly rich set of genes, as compared to ESTs associated with energy reserves, electron transport, or even TCA/Krebs cycle.

As one might expect, ESTs associated with FunCat 11 (Transcription) were almost entirely associated with RNA synthesis (Fig. 4). Of the 72.75 standardized functional assignments within FunCat 11, 62.75 were associated with RNA synthesis. This would indicate that RNA synthesis in *O. novo-ulmi* requires a considerable number of proteins and cofactors. During late exponential growth, the expression of genes associated with RNA processing and RNA modification was very limited. It seems reasonable to assume that RNA processing and modification is not a functional priority to the fungus at this time. Further, complex RNA processing and modification could not be accomplished by the extremely limited number of ESTs associated with these secondary functional categories.

Comparative Analysis of *O. novo-ulmi* and *N. crassa* Functional Catalogues

As is shown in Figure 10, FunCat 01 (Metabolism) is the most highly represented in each dataset. This indicates considerable genetic depth in filamentous fungi with regard to metabolic function. As each dataset has been standardized to eliminate redundancy in gene identities, greater levels of representation within a functional category indicate greater numbers of unique gene products associated with that biological function. Increased representation indicates increased genetic complexity within that functional

category. That said, one must bear in mind that an absence of complexity at the EST level does not mean that genetic complexity is necessarily absent at the nucleotide level. The *O. novo-ulmi* functional catalogue presented here was built from transcribed RNA, and it is reasonable to assume that only a portion of the genome was actively being transcribed at the time of harvest. Table 2 shows a 4-fold disparity between representation of gene products associated with cellular biogenesis. It is likely that a portion of these genes in *O. novo-ulmi* were not being expressed at the time of harvest, and these genes are simply not contributing to this category of the functional catalogue. Conversely, the *N. crassa* genomic database represents genes, rather than transcribed tags, and is independent of expression levels. Genes within a functional category in the *N. crassa* database represent the total potential expression for that category, which may or may not be realized under different cellular conditions.

Gene products associated with subcellular localization were also highly represented in both datasets (Figure 10). It is interesting to note that with respect to both genes (*N. crassa)* and ESTs *(O. novo-ulmi)* the level of genetic complexity exhibited within the Subcellular Localization category exceeded that of the Energy category nearly four-fold. This comparison illustrates well that percent representation of the functional catalogue by unique ESTs is not indicative of expression levels. It is reasonable to assume that several genes associated with energy production were being expressed at high levels at the time of harvest, but a great number of unique subcellular localization genes were being expressed at a low level, resulting in a greater percent representation of subcellular localization genes within that functional category.

Table 2. Comparative analysis of *N. crassa* and *O. novo-ulmi* Yeast LMW EST functional catalogues in accordance with MIPS FunCat. A total of 7,350 functional assignments were included for *N. crassa*, while the *O. novo-ulmi* functional catalogue included 861.25 functional assignments.

Functional Category		*O. novo-ulmi* Representation	*O. novo-ulmi* Percent Representation	*N. crassa* Representation	*N. crassa* Standardized Representation	*N. crassa* Percent Representation
01	Metabolism	178.5	22.57	1322	660.34	17.99
02	Energy	39.5	4.99	254	126.87	3.46
10	Cell Cycle	6	0.76	399	199.30	5.43
11	Transcription	110	13.91	553	276.22	7.52
12	Protein Synthesis	89.25	11.28	266	132.87	3.62
14	Protein Fate	7.5	0.95	617	308.19	8.39
16	Binding & Cofactor	12.25	1.55	323	161.34	4.39
18	Protein Activity	7.25	0.92	58	28.97	0.79
20	Cell Transport	53.5	6.76	906	452.55	12.33
30	Cell Signalling	0	0.00	177	88.41	2.41
32	Cell Rescue	9.5	1.20	329	164.34	4.48
34	Cellular Environment	32	4.05	260	129.87	3.54
40	Cell Fate	2	0.25	153	76.42	2.08
42	Cell Biogenesis	71.75	9.07	179	89.41	2.44
43	Cell Differentiation	1	0.13	139	69.43	1.89
70	Subcellular Localization	90	11.38	1167	582.92	15.88
73	Cell Localization	30	3.79	0	0.00	0.00
98	Unresolved	51	6.45	206	102.90	2.80
Subtotal		791	100.00	7350	3671.33	99.43
99	99 - Unknown	66		5354	4818.60	
	Total Functional Assignments	**861.25**			**8468.95**	

Standardization involved the reduction of absolute Functional Category scores to account for multifunctional proteins by the multiplication factor 0.555 (4081 sequences generating a functional category divided by 7350 functional categories All totals were also reduced by 10% to account for redundancy observed within the dataset as a whole.

A much greater proportion of gene products associated with cell biogenesis were represented in the *O. novo-ulmi* Yeast LMW library, as compared to the *N. crassa* genomic database (Fig. 10). This is indicative of the growth state in which the *O. novo-ulmi* culture was harvested. It seems reasonable to assume, and Figure 10 appears to confirm that actively growing yeast-like cultures of *O. novo-ulmi* require the expression of a high number of unique Cellular Biogenesis genes. The proportion of unique actively expressed *O. novo-ulmi* genes associated with cellular biogenesis is four-fold higher than that of the *N. crassa*genomic database for this category. The functional catalogues derived from an exponential phase *O. novo-ulmi* could be described as having a biogenesis focus. Conversely, genes associated with decay or senescence constitute a greater proportion of the genome in *N. crassa.*

Biogenesis of cellular protein is only one of many active mechanisms utilized by organisms for control of biological pathways. Protein Fate is an important primary functional category for the control and degradation of enzymes and cofactors. These pathways could be described as secondary pathways, and as can be seen in Figure 10, are not constituitively expressed. Gene products associated with protein fate comprise a considerably larger component of the *N. crassa* genomic database, as compared to the *O. novo-ulmi*functional catalogue. This is not surprising when one considers that an actively growing culture of *O. novo-ulmi* has a focus on biogenesis, rather than on degradation of newly synthesized proteins.

Forward Value

Described here was the construction of a low redundancy library from *Ophiostoma novo-ulmi* complementary DNA, producing a total of 4386 readable expressed sequence

tags (ESTs). Annotation and functional categorization in accordance with the FunCat scheme (MIPS) produced a functional catalogue with a focus on metabolism and subcellular localization. Comparative analysis with the *Neurospora crassa* functional catalogue showed that in some instances, primary functional categories associated with biogenesis of exponential phase fungal cultures were favoured in the *O. novo-ulmi* functional catalogue, while those gene products associated with degradation and senescence comprised a greater majority of the *N. crassa* functional catalogue, as compared to that of *O. novo-ulmi*.

In addition to an early glimpse into the expression profile of an actively growing culture of *Ophiostoma*, which could be representative of many additional filamentous fungi, an online database was established for downstream gene discovery by private and public organizations. The pathogenic nature of *O. novo-ulmi* makes this organism both scientifically and economically significant. Any potential means of virulence control or pathogenicity reduction could prove extremey useful, and profitable. Among the 4,386 ESTs in this database exist a cohort of pathogenicity and virulence genes. Until recently, pathogenicity studies were limited to a single locus, which made the discovery of an effective means of control exceedingly difficult. The large dataset presented here, along with recent advances in genomics, proteomics, and bioinformatics, present a new collection of tools to the biotechnologist studying tree pathogens. A wide-angle approach examining interactions of gene networks, rather than the actions of single gene products, could provide a solution to the multifactor nature of this host-pathogen interaction.

Future Experiments

The fullest potential of the data generated in an EST library can be realized through high-throughput screening via microarray analysis. This technique involves the spotting of microvolumes of thousands of DNA sequences onto a glass slide, which then undergoes a series of hybridizations with dye-conjugated mRNAs. The specificity of mRNA for its original template DNA sequence ensures that only a single spot on a microarray slide is bound by each kind of mRNA molecule. Microarray plate-reading instruments are able to detect and quantify the amount of bound mRNA at each site on the microarray. Pair-wise hybridizations of duplicate microarray slides with mRNA populations from different treatments or cell types allows a researcher to compare the expression levels of thousands of genes simultaneously across multiple treatments or cell types.

With regard to the ESTs discussed here, interesting cell types and treatments would include yeast versus mycelium growth states, and presence versus absence of elm host tissues. Genes exhibiting changes in expression during these key transition states are potential pathogenicity factors. A fungal isolate incapable of growth in a yeast-like form would be non-transmittable by vector transmission. Conversely, an isolate incapable of mycelial growth would be non-pathogenic since colonization of host vascular tissues requires mycelial hyphae. Induction of gene expression upon exposure to elm tissues nominates those genes as critical components in host-pathogen interaction. Before a gene or protein network can be understood or manipulated, it must first be identified. Microarray analysis of the *O. novo-ulmi* EST library constructed here would likely identify many key pathogenicity factors.

A narrower approach might focus on full characterization of specific fragments represented within the catalogue. Average insert size of ESTs within the *O. novo-ulmi* library was 498 bp. Specific sequence information of this length is suitable for designing further cloning experiments to capture full length genes. This can be accomplished by probing of a genomic library, or by use of genome walking techniques that generate stepwise lengths of sequence that can be concatamerized to generate full length genes. In fact, genome walking experiments have already been initiated to fully characterize the alcohol dehydrogenase I gene and its promoter, based on the sequence information from this EST catalogue.

Bibliography

Agrios, George N., 1997. *Plant Pathology* fourth edition, Academic Press. New York.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, & Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389.

Amoroso G, Morell-Avrahov L, Muller D, Klug K, & Sultemeyer D. 2005. The gene NCE103 (YNL036w) from saccharomyces cerevisiae encodes a functional carbonic anhydrase and its transcription is regulated by the concentration of inorganic carbon in the medium. Mol Microbiol 56:549.

Arena ME, Manca de Nadra MC, & Munoz R. 2002. The arginine deiminase pathway in the wine lactic acid bacterium lactobacillus hilgardii X1B: Structural and functional study of the arcABC genes. Gene 301:61.

Banfield W. 1941. Distribution by the sap stream of spores of three fungi that induce vascular wilt disease in elm. J Agric Res 65:637.

Barrett AJ. 1997. Nomenclature committee of the international union of biochemistry and molecular biology (NC-IUBMB). enzyme nomenclature. recommendations 1992. supplement 4: Corrections and additions (1997). Eur J Biochem 250:1.

Bates M, Buck K, & Brasier C. 1993a. Molecular relationships between *ophiostoma ulmi* and the NAN and EAN races of *O. novo-ulmi* determined by restriction fragment length polymorphisms of nuclear DNA. Mycological Research 97:449.

Bates M, Buck K, & Brasier C. 1993b. Molecular relationships of the mitochondrial DNA of *ophiostoma ulmi* and the NAN and EAN races of *O. novo-ulmi* determined by restriction fragment length polymorphisms . Mycological Research 97:1093.

Bernier L. 1993. Conventional and molecular genetic approaches to the study of pathogenicity in *ophiostoma ulmi sensu lato*. In: Sticklen M and Sherald J, editors. Dutch elm disease Research: Cellular and Molecular Approaches. New York: Springer-Verlag. p 293

Bernier L. 1988. Induction, characterization, and mapping of mutations in *ophiostoma ulmi*, the causal agent of dutch elm disease .

Bernier L, & Hubbes M. 1990. Mutations in *ophiostoma ulmi* induced by N-methyl-N'-nitro-N-nitrosoguanidine . Can J Bot 68:225.

Binz t, & Canevascini G. 1996. Differential production of extracellular laccase in the dutch elm disease pathogens *ophiostoma ulmi* and *O. novo-ulmi* . Mycol Res 100:1060.

Birch M, Paine T, & Miller J. 1981. Effectiveness of pheromone mass-trapped of the smaller european elm bark beetle. Calif Agric 35:6.

Bowden CG, Hintz WE, Jeng R, Hubbes M, & Horgen PA. 1994. Isolation and characterization of the cerato-ulmin toxin gene of the dutch elm disease pathogen, ophiostoma ulmi. Curr Genet 25:323.

Brasier C. 1996. Low genetic diversity of the *ophiostoma novo-ulmi* population in north america. Mycologia 88:951.

Brasier C. 1983. A cytoplasmically transmitted disease of *ceratocystis ulmi*. Nature 305:220.

Brasier C, Kirk S, & Tegli S. 1994. Naturally occurring non-cerato-ulmin producing mutants of *ophiostoma novo-ulmi* are pathogenic but lack of aerial mycelium Mycol Res 99:436.

Brasier C, & Kirk S. 2001. Designation of the EAN and NAN races of *ophiostoma novo-ulmi* as subspecies . Mycological Research 105:547.

Brasier C, & Kirk S. 2000. Survival of clones of NAN *ophiostoma novo-ulmi* around its probable centre of appearance in north america. Mycological Research 104:1322.

Brasier C, & Mehrotra M. 1995. *Ophiostoma himal-ulmi* sp. nov., a new species of dutch elm disease fungus endemic to the himalayas. Mycol Res 99:205.

Campana R, & Stipes R. 1981. Dutch elm disease in north america with particular reference to canada: Success or failure of conventional control methods. Can J Plant Pathol 31:252.

Claydon N, Elgersma D, & Grove J. 1980. The phytotoxicity of some phenolic metabolic products of *ophiostoma ulmi* to *Ulmus* species . Nether J Plant Pathol 86:229.

De Rafael MA, Valle T, Babiano MJ, & Corchete P. 2001. Correlation of resistance and H2O2 production in ulmus pumila and ulmus campestris cell suspension cultures inoculated with ophiostoma novo-ulmi. Physiol Plant 111:512.

Del Sorbo G, Scala F, Parrella G, Lorito M, Comparini C, Ruocco M, & Scala A. 2000. Functional expression of the gene cu, encoding the phytotoxic hydrophobin cerato-ulmin, enables ophiostoma quercus, a nonpathogen on elm, to cause symptoms of dutch elm disease. Mol Plant Microbe Interact 13:43.

Duchesne L. 1993. Mechanisms of resistance: Can they help save susceptible elms? In: Sticklen M and Sherald J, editors. Dutch Elm Disease Research: Cellular and Molecular Approaches. Berlin, New York: Springer-Verlag. p 139

Duchesne L, Hubbes M, & Jeng R. 1986. Mansonone E and F accumulation in *ulmus pumila* resistant to dutch elm disease . Can J for Res 16:410.

Duchesne L, Jeng R, & Hubbes M. 1985. Accumulation of phytoalexins in *ulmus americana* in response to injection of a non-aggressive and an aggressive strain of *ophiostoma ulmi* . Can J Bot 63:678.

Dufour E, Boulay J, Rincheval V, & Sainsard-Chanet A. 2000. A causal link between respiration and senescence in podospora anserina. Proc Natl Acad Sci U S A 97:4138.

Dufton MJ. 1983. The significance of redundancy in the genetic code. J Theor Biol 102:521.

Dumas M, Strunz G, Hubbes M, & Jeng R. 1986. Inhibition of *ceratocystis ulmi* by mansonones A, C, D, F and G isolated from *ulmus americana* . Europ J for Pathol 16:217.

Dumas M, Strunz G, Hubbes M, & Jeng R. 1983. Isolation and identification of six mansonones from *ulmus americana* infected with *ceratocystis ulmi* . Experentiaa 39:1089.

Eades CJ, & Hintz WE. 2000. Characterization of the class I alpha-mannosidase gene family in the filamentous fungus aspergillus nidulans. Gene 255:25.

Frank S. 1992. Models of plant-pathogen co-evolution Trends in Genetics 88:213.

Galagan JE, Calvo SE, & Borkovich KA *et al.* 2003. The genome sequence of the filamentous fungus neurospora crassa. Nature 422:859.

Gibbs J. 1981. Dutch elm disease. In: Campana R and Stipes R, editors. Compendium of elm diseases. St. Paul, MN: American Phytopathological Society. p 7

Hausladen A, & Fridovich I. 1994. Superoxide and peroxynitrite inactivate aconitases, but nitric oxide does not. J Biol Chem 269:29405.

Hintz W, Jeng R, Hubbes M, & Horgen P. 1991. Identification of three populations of *ophiostoma ulmi* (aggressive subgroup) by mitochondrial restriction-site mapping and nuclear DNA fingerprinting . Experimental Mycology 15:316.

Hintz WE, Jeng RS, Yang DQ, Hubbes MM, & Horgen PA. 1993. A genetic survey of the pathogenic fungus ophiostoma ulmi across a dutch elm disease front in western canada. Genome 36:418.

Holmes F. 1976. While dutch elm disease continues to spread, natural selection in wild elms may save the species from extinction. Horticulture 54:72.

Hubbes M. 1999. The american elm and dutch elm disease. Forest Chron 75:265.

Hubbes M. 1993. Mansonones, elicitors, and virulence. In: Sticklen M and Sherald J, editors. Dutch Elm Disease Research: Cellular and Molecular Approaches. Berlin, New York: Springer-Verlag. p 208

Hubbes M, & Jeng R. 1981. Aggressiveness of *ceratocystis ulmi* and induction of resistance in *ulmus americana* . Europ J for Pathol 11:257.

Huntley G. 1982. The elm - a resurgent resource or a persistent problem?103.

International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the human genome. Nature 431:931.

Jeenes DJ, Marczinke B, MacKenzie DA, & Archer DB. 1993. A truncated glucoamylase gene fusion for heterologous protein secretion from aspergillus niger. FEMS Microbiol Lett 107:267.

Jeng R, Alfenas A, Hubbes M, & Dumas M. 1983. Presence and accumulation of fungitoxic substances against *ceratocystis ulmi* in *ulmus americana*: Possible relation to induced resistance . Europ J for Pathol 13:267.

Jeng R, Bernier L, & Brasier C. 1988. A comparative study of cultural and electrophoretic characeristics or eurasian and north american races of *ophiostoma ulmi* . Canadian Journal of Botany 66:1325.

Jeng R, Hintz WE, Bowden CG, Horgen PA, & Hubbes M. 1996. A comparison of the nucleotide sequence of the cerato-ulmin gene and the rDNA ITS between aggressive and non-aggressive isolates of ophiostoma ulmi sensu lato, the causal agent of dutch elm disease. Curr Genet 29:168.

Jin H, Webster G, Holliday N, Pines P, & Westwood A. 1996. An elm bark beetle bioassay for residual efficacy of chlorpyrifos and cypermethrin used for the control of dutch elm disease in manitoba. J Environ Sci Health Part B 31:751.

Kim MJ, Kim HS, Lee JK, Lee CB, & Park SD. 2002. Regulation of septation and cytokinesis during resumption of cell division requires uvi31+, a UV-inducible gene of fission yeast. Mol Cells 14:425.

Kim SH, Uzunovic A, & Breuil C. 1999. Rapid detection of ophiostoma piceae and O. quercus in stained wood by PCR. Appl Environ Microbiol 65:287.

Kondo E. 1978. Scope and limitations of carbendazim phosphate injections in dutch elm disease control. J Arboric 4:86.

Kondo E, Roy D, & Jorgensen E. 1973. Salts of methyl-2-benzimidazole carbamate (MBC) and assessment of their potential in dutch elm disease control. Can J for Res 3:548.

Kupfer DM, Reece CA, Clifton SW, Roe BA, & Prade RA. 1997. Multicellular ascomycetous fungal genomes contain more than 8000 genes. Fungal Genet Biol 21:364.

Lanier G. 1989. Trap trees for the control of dutch elm disease. J Arboric 15:105.

Mewes H, Albermann K, & Bahr M *et al.* 1997. Overview of the yeast genome. Nature 387:7.

Mi H, Vandergriff J, Campbell M, Narechania A, Majoros W, Lewis S, Thomas PD, & Ashburner M. 2003. Assessment of genome-wide protein function classification for drosophila melanogaster. Genome Res 13:2118.

Michal G. 1998. Biochemical pathways: An atlas of biochemistry and molecular biology. Hoboken, NJ: Wiley and Sons, Inc.

Missall TA, Pusateri ME, Donlin MJ, Chambers KT, Corbett JA, & Lodge JK. 2006. Posttranslational, translational, and transcriptional responses to nitric oxide stress in cryptococcus neoformans: Implications for virulence. Eukaryot Cell 5:518.

Nunberg JH, Meade JH, Cole G, Lawyer FC, McCabe P, Schweickart V, Tal R, Wittman VP, Flatgaard JE, & Innis MA. 1984. Molecular cloning and characterization of the glucoamylase gene of aspergillus awamori. Mol Cell Biol 4:2306.

Oullette C, & Pomerleau R. 1965. Recherches sur la resistance de L'orme D'amerique au *ceratocystis ulmi* Can J Bot 43:85.

Parker K, Collins D, Tyler L, Connola D, Ozard W, & Dietrich H. 1947. The dutch elm disease - association of *ceratostomella ulmi* with *scolytus multistriatus*, its advance into new areas, and control of the disease Cornell Univ Agr Exp Stn 275:1.

Pomerleau R. 1981. Dutch elm disease in canada. In: Stipes R and Campana R, editors. Compendium of elm diseases. St. Paul, MN: Am. Phytopathol. Soc. p 8

Prinz H, Lavie A, Scheidig AJ, Spangenberg O, & Konrad M. 1999. Binding of nucleotides to guanylate kinase, p21(ras), and nucleoside-diphosphate kinase studied by nano-electrospray mass spectrometry. J Biol Chem 274:35337.

Procter R, & Smalley E. 1988. Localized accumulation of mansonones E and F in elms following inoculation with *ophiostoma ulmi* Can J Plant Pathol 10:371.

Proctor R, Guries R, & Smalley E. 1994. Lack of association between tolerance to the elm phytoalexin mansonone E and virulence in *ophiostoma novo-ulmi* . Can J Bot 72:1355.

Prosser T. 1998. One company's success against dutch elm disease. Tree Care Ind June:60.

Richards W. 1998. Novel spore deficient mutants of the dimorphic fungal plant pathogen *ophiostoma ulmi* . Phytopathol Supp 88:113.

Richards W, Takai S, Lin D, Hiratsuka Y, & Asina S. 1982. An abnormal strain of *ceratocystis ulmi* incapable of producing external symptoms of dutch elm disease . Europ J for Pathol 12:193.

Roy D, Purdy J, & Ayyamperumal P. 1980. Distribution of methyl benzimidazol-2-yl carbamate phosphate in elm: Effects of chemical properties and formulation variables. Can J for Res 10:143.

Roy D, Purdy J, Perumal P, & Grace J. 1988. Effect of application equipment on the distribution of chlorpyrifos applied for dutch elm disease vector control. Proc Entomol Soc Ont 119:63.

Ruepp A, Graml W, Santos-Martinez ML, Koretke KK, Volker C, Mewes HW, Frishman D, Stocker S, Lupas AN, & Baumeister W. 2000. The genome sequence of the thermoacidophilic scavenger thermoplasma acidophilum. Nature 407:508.

Ruepp A, Zollner A, & Maier D *et al.* 2004. The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. Nucleic Acids Res 32:5539.

Salanoubat M, Lemcke K, & Rieger M *et al.* 2000. Sequence and analysis of chromosome 3 of the plant arabidopsis thaliana. Nature 408:820.

Scala A, Pettuelli M, Coppola L, Guastini M, Tegli S, Del Sorbo G, Mittempherger L, & Scala F. 1997. Dutch elm disease and progression and quantitative determination of cerato-ulmin in leaves, stems, and branches of elms inoculated with *ophiostoma novo-ulmi* and *O. ulmi* . Physiol Mol Plant Pathol 50:349.

Scheffer R, Brakenhoff A, Kerkenaar A, & Elgersma D. 1988. Control of dutch elm disease by the sterol biosynthesis inhibitors fenpropimorph and fenpropidin. Neth J Plant Pathol 94:161.

Scheffer R, Heybroek H, & Elgersma D. 1980. Symptom expression in elms after inoculation with combinations of an aggressive and non-aggressive strain of *ophistoma novo-ulmi* . Neth J Plant Pathol 86:315.

Scheffer R, Liem J, & Elgersma D. 1987. Production *in vitro* of phytotoxic compounds by non-aggressive and aggressive isolates of *ophiostoma ulmi*, the dutch elm disease pathogen . Physiol Molec Plant Pathol 30:321.

Schreiber L. 1993. An old problem; a new approach. In: Sticklen M and Sherald J, editors. Dutch Elm Disease Research: Cellular and Molecular Approaches. Berlin, New York: Springer-Verlag. p 51

Sherald J. 1993. Demands and opportunities for selecting disease resistant elms. In: Sticklen M and Sherald J, editors. Dutch Elm Disease Research: Cellular and Molecular Approach. Berlin, New York: Springer-Verlag. p 60

Shigo A, & Campana R. 1977. Discoloured and decayed wood associated with injection wounds in american elm. J Arboric 3:230.

Smalley E. 1978. Control tractics in research and practices: IV. systemic chemical treatments of trees for protection and therapy. In: Sinclair W and Campana R, editors. Dutch elm disease; perspectives after 60 years. New York: Cornell Univ. Agric. Exp. Stn. p 34

Smalley E, & Guries R. 1993. Breeding elms for resistance to dutch elm disease. Annu Rev Phytopathol 31:325.

Smalley E, Guries R, & Lester G. 1993a. American liberty elms and beyond: Going from the impossible to the difficult . In: Sticklen M and Sherald J, editors. Dutch Elm Disease Research: Cellular and Molecular Approaches. Berlin, New York: Springer-Verlag. p 26

Smalley E, Raffa K, Proctor R, & Klepzig K. 1993b. Tree responses to infection by species of *ophiostoma* and *ceratocystis* . In: Wingfield M, Seifert K and Webber J, editors. *Ceratocystis* and *Ophistoma*: Taxonomy, ecology, and pathogenicity. St. Paul, MN: APS. p 207

Solla A, Martin JA, Corral P, & Gil L. 2005. Seasonal changes in wood formation of ulmus pumila and U. minor and its relation with dutch elm disease. New Phytol 166:1025.

Somssich I, & Hahlbrock K. 1998. Pathogen defence in plants - a paradigm of biological complexity. Trends Plant Sci 3:86.

Stennes M, & French D. 1987. Distribution and retention of thiabendazoic hypophosphite and carbendazim phosphate injected into mature american elms. Phytopathology 77:707.

Sticklen M, Bolyard M, Hajela R, & Duchesne L. 1991. Molecular and cellular aspects of dutch elm disease. Phytoprotection 72:1.

Sutherland M, Mittempergher L, & Brasier C. 1995. Control of dutch elm disease by induced host resistance. Eur J for Pathol 25:307.

Szczegola-Derkacz M. 1988. Induction of mansonones by *ophiostoma ulmi* in tissue cultures of *ulmus americana* and *ulmus pumila* .

Takai S. 1980. Relationship of the production of the toxin, ceratoulmin, to synnemata formation, pathogenicity, mycelial habit, and growth of *ceratocystis ulmi* isolates Can J Bot 58:658.

Taylor J. 1986. Fungal evolutionary biology and mitochondrial DNA. Experimental Mycology 10:259.

Tegli S, & Scala A. 1996. Isolation and characterization of non-cerato-ulmin producing laboratory induced mutants of *ophiostoma novo-ulmi* Mycol Res 100:661.

Temple B, Hintz WE, & Pines PA. (In press). A nine-year genetic survey of the causal agent of dutch elm disease, *ophiostoma novo-ulmi*; in winnipeg, canada .

Temple B, Horgen P, Bernier L, & Hintz W. 1997. Cerato-ulmin, a hydrophobin secreted by the causal agents of dutch elm disease, is a parasitic fitness factor. Fungal Genet Biol 22:39.

Tomalak M, Welch H, & Galloway T. 1989. Nematode parasites of bark beetles (*scolytidae*) in southern manitoba, with descriptions of three new species of *sulphuretylenchus* Ruhm (nematoda: Allantonematidae) . Can J Zool 67:2497.

Townsend A, & Santamour Jr. F. 1993. Progress in the development of disease-resistant elms. In: Sticklen M and Sherald J, editors. Dutch Elm Disease Research: Cellular and Molecular Approach. berlin, New York: Springer-Verlag. p 46

Ware G, & Miller F. 1997. Developing better elms. Amer Nurseryman Aug 15:44.

Wootton J, & Federhen S. 1993. Statistics of local complexity in amino acid sequences and sequence databases. Computational Chemistry 17:149-63.

Wu W, Jeng R, & Hubbes M. 1989. Toxic effect of elm phytoalexin mansonones on *ophiostoma ulmi*, the causal agent of dutch elm disease . Europ J for Pathol 19:343.

Yang D. 1991. Isolation, identification, and characterization of phytoalexin elicitors from *ophiostoma ulmi* .

Yang D, Jeng R, & Hubbes M. 1989. Mansonone accumulation in elm callus induced by elicitors of *ophiostoma ulmi*, and general properties of elicitors . Can J Bot 67:3490.

Yoshida T, Kato Y, Asada Y, & Nakajima T. 2000. Filamentous fungus aspergillus oryzae has two types of alpha-1,2-mannosidases, one of which is a microsomal enzyme that removes a single mannose residue from Man9GlcNAc2. Glycoconj J 17:745.

Appendix I
MIPS Functional Catalogue Condensed to Two Levels

01 METABOLISM
01.01 amino acid metabolism
01.02 nitrogen and sulfur metabolism
01.03 nucleotide metabolism
01.04 phosphate metabolism
01.05 C-compound and carbohydrate metabolism
01.06 lipid, fatty acid and isoprenoid metabolism
01.07 metabolism of vitamins, cofactors, and prosthetic groups
01.20 secondary metabolism
01.25 extracellular metabolism
02 ENERGY
02.01 glycolysis and gluconeogenesis
02.04 glyoxylate cycle
02.05 Entner-Doudoroff pathway
02.07 pentose-phosphate pathway
02.08 pyruvate dehydrogenase complex
02.09 anaplerotic reactions
02.10 tricarboxylic-acid pathway (citrate cycle, Krebs cycle, TCA cycle)
02.11 electron transport and membrane-associated energy conservation
02.13 respiration
02.16 fermentation
02.17 chemolithotrophie (e.g. sulfide, nitrogenous compounds)
02.19 metabolism of energy reserves (e.g. glycogen, trehalose)
02.25 oxidation of fatty acids
02.30 photosynthesis
02.45 energy conversion and regeneration
04 STORAGE PROTEIN
04.01 storage facilitating proteins
04.02 stored proteins
10 CELL CYCLE AND DNA PROCESSING
10.01 DNA processing
10.03 cell cycle
11 TRANSCRIPTION
11.02 RNA synthesis
11.04 RNA processing
11.06 RNA modification
12 PROTEIN SYNTHESIS
12.01 ribosome biogenesis
12.04 translation
12.07 translational control
12.10 aminoacyl-tRNA-synthetases
14 PROTEIN FATE (folding, modification, destination)
14.01 protein folding and stabilization
14.04 protein targeting, sorting and translocation
14.07 protein modification
14.10 assembly of protein complexes
14.13 protein degradation
16 PROTEIN WITH BINDING FUNCTION OR COFACTOR REQUIREMENT (structural or catalytic)
16.01 protein binding
16.02 peptide binding
16.03 nucleic acid binding
16.05 polysaccharide binding
16.06 motor protein
16.07 structural protein
16.09 lipid binding
16.11 amino acid binding
16.12 sulfate binding
16.13 C-compound binding
16.17 metal binding
16.19 nucleotide binding
16.21 complex cofactor/cosubstrate binding
18 PROTEIN ACTIVITY REGULATION
18.01 mechanism of regulation
18.02 target of regulation
20 CELLULAR TRANSPORT, TRANSPORT FACILITATION AND TRANSPORT ROUTES
20.01 transported compounds (substrates)
20.03 transport facilitation
20.09 transport routes
30 CELLULAR COMMUNICATION/SIGNAL TRANSDUCTION MECHANISM
30.01 intracellular signalling
30.05 transmembrane signal transduction
32 CELL RESCUE, DEFENSE AND VIRULENCE
32.01 stress response
32.05 disease, virulence and defense
32.07 detoxification
32.10 degradation of foreign (exogenous) compounds
34 INTERACTION WITH THE CELLULAR ENVIRONMENT
34.01 ionic homeostasis
34.03 membrane excitability
34.05 cell motility

34.07 cell adhesion
34.11 cellular sensing and response
36 INTERACTION WITH THE ENVIRONMENT (Systemic)
36.03 nutrients uptake and absorption (e.g. digestion)
36.05 osmoregulation and excretion
36.07 gas and metabolite distribution
36.09 systemic temperature regulation
36.11 systemic rhythm control
36.20 plant / fungal specific systemic sensing and response
36.25 animal specific systemic sensing and response
38 TRANSPOSABLE ELEMENTS, VIRAL AND PLASMID PROTEINS
38.01 LTR retroelements (retroviral)
38.02 non-LTR retroelements
38.03 transposons
38.05 viral proteins
38.06 phage proteins
38.07 proteins necessary for the integration or inhibition of transposon movement
40 CELL FATE
40.01 cell growth / morphogenesis
40.02 cell differentiation
40.05 dedifferentiation
40.10 cell death
40.20 cell aging
41 DEVELOPMENT (Systemic)
41.01 fungal/microorganismic development
41.03 plant development
41.05 animal development
42 BIOGENESIS OF CELLULAR COMPONENTS
42.01 cell wall
42.02 eukaryotic plasma membrane
42.03 cytoplasm
42.04 cytoskeleton
42.05 centrosome
42.06 cell junction
42.07 endoplasmic reticulum
42.08 Golgi
42.09 intracellular transport vesicles
42.10 nucleus
42.16 mitochondrion
42.19 peroxisome
42.22 endosome
42.25 vacuole or lysosome
42.26 plastid
42.27 extracellular / secretion proteins
42.28 periplasmatic space
42.29 bud / growth tip
42.30 prokaryotic cytoplasmic membrane
42.32 flagellum
42.33 pilus/fimbria
42.34 prokaryotic cell envelope structures
42.35 prokaryotic intracytoplasmic membrane
42.36 prokaryotic cell inclusions
42.37 prokaryotic nucleoid
43 CELL TYPE DIFFERENTIATION
43.01 fungal/microorganismic cell type differentiation
43.02 plant cell type differentiation
43.03 animal cell type differentiation
45 TISSUE DIFFERENTIATION
45.01 fungal/microorganismic tissue
45.02 plant tissue
45.03 animal tissue
47 ORGAN DIFFERENTIATION
47.01 fungal organ
47.02 plant organ
47.03 animal organ
70 SUBCELLULAR LOCALIZATION
70.01 cell wall
70.02 eukaryotic plasma membrane / membrane attached
70.03 cytoplasm
70.04 cytoskeleton
70.05 centrosome
70.06 cell junction
70.07 endoplasmic reticulum
70.08 Golgi
70.09 intracellular transport vesicles
70.10 nucleus
70.16 mitochondrion
70.19 peroxisome
70.22 endosome
70.25 vacuole or lysosome
70.26 plastid
70.27 extracellular / secretion proteins
70.28 periplasmatic space
70.29 bud / growth tip
70.30 prokaryotic cytoplasmic membrane
70.32 flagellum
70.33 pilus/fimbria
70.34 prokaryotic cell envelope component
70.35 prokaryotic intracytoplasmic membrane
70.36 prokaryotic cell inclusions
70.37 prokaryotic nucleoid
73 CELL TYPE LOCALIZATION
73.01 fungal / microorganismic cell type
73.02 plant cell type
73.03 animal cell type
75 TISSUE LOCALIZATION
75.01 fungal/microorganismic tissue
75.02 plant tissue
75.03 animal tissue
77 ORGAN LOCALIZATION
77.01 fungal organ
77.02 plant organ
77.03 animal organ

78 ubiquitous expression

98 CLASSIFICATION NOT YET CLEAR-CUT

99 UNCLASSIFIED PROTEINS

Appendix II

Index of Assignable *O. novo-ulmi* Yeast LMW ESTs With Functional Assignments

Protein Identity	Copy #	FunCat #	Loci in library
10 KD chaperonin (protein CPN10) (protein groes)	1	42.16; 70.02	35B02
100 kDa protein	1	99	24E02
20beta-hydroxysteroid dehydrogenase	1	01.06	05D10
20S proteasome beta-type subunit, Epoxomicin	1	70.13	58D07
26S ATP/ubiquitin-dependent proteinase chain S4	2	70.13; 42.04; 73.01	29F01; 06H10
26S protease regulatory subunit 6B homolog	4	70.13	34B08; 43A05; 48D02; 55F08
26S proteasome regulatory particle chain RPN11	1	70.13	43G03
26S proteasome regulatory particle chain RPN7	1	70.13	34H03
2-heptaprenyl-1,4-naphthoquinone methyltransferase	1	02.11	59D12
30 KD heat shock protein	10	11.01	29C05; 31E03; 34D11; 34G05; 38E04; 39C03;
3-beta-hydroxysteroid-delta(8),delta(7)-isomerase	1	01.06	05A02
3-hydroxyisobutyrate dehydrogenase and related beta-hydroxyacid dehydrogenases	1	01.05	08B08
3-ketoacyl-CoA thiolase	2	01.06; 02.13	30F04; 39H04
3-oxoadipate enol-lactonase II	1	02.10	07D10
3-oxoadipate enol-lactone hydrolase/4-carboxymuconolactone decarboxylase	2	98	47G03; 56E03
3-phosphoserine phosphatase	2	01.01	28C12; 43G08
3-phytase	1	01.04; 01.07	29G09
40S cytoplasmic ribosomal protein S7	5	12.01	29A12; 40E04; 20D03; 50E03; 57H08
40S ribosomal protein S11	1	12.01	17E05
40S ribosomal protein S12	19	12.01	27C01; 28H01; 29A11; 33B10; 34E02; 35F08; 43C08; 47A01; 01D02; 19C07; 19F04; 24D02; 53E07; 56F06; 57G08; 58D06; 58G07; 60D01;
40S ribosomal protein S13	4	12.01	38A04; 12H05; 17H11; 49F11
40S ribosomal protein S15	2	12.01	30C01; 30F09
40S ribosomal protein S16	5	12.01	06H03; 12B02; 15D04; 22F02; 52C01
40S ribosomal protein S17	1	12.01	28E11
40s ribosomal protein S2	1	12.01	44H10
40S ribosomal protein S20	5	12.01	29D01; 35H01; 43E09; 59H10; 60A12
40S ribosomal protein S22	3	12.01	39D05; 48F08; 53G01
40S ribosomal protein S23	5	12.01	27G10; 33F11; 40D06; 18G06; 57H12
40s ribosomal protein s24b	1	12.01	58F04
40S ribosomal protein S25	15	12.01	32F06; 35F04; 48A06; 03C11; 04G07; 05A03; 07F07; 12F05; 13E04; 18F08; 20C02; 24A02; 54G08; 55A05; 57D01
40S ribosomal protein S26E	5	12.01	27G02; 35F09; 02D04; 07A11; 09B06
40S ribosomal protein S27	3	12.01	32D07; 32H11; 39E02
40S ribosomal protein S28	2	12.01	36C06; 54A01
40s ribosomal protein s30	1	12.01	22H04
40S ribosomal protein S3AE (S1)	2	12.01	45G07; 11B10
40S ribosomal protein S4	4	12.01	31B11; 46B05; 06B04; 10B01
40S ribosomal protein S5	3	12.01	48E02; 10B10; 21F07
40S ribosomal protein S6	1	12.01	02B04
40S ribosomal protein S8	1	12.01	08C01
40S ribosomal protein S9	1	12.01	07E03
4-aminobutyrate aminotransferase protein	2	01.01; 01.02	33H08; 23A08
5'-amp-activated protein kinase	1	42.03	49H03
5-epimerase	1	01.05	44C05
60S acidic ribosomal protein P1-beta	1	12.01	25C01
60S mitochondrial ribosomal protein L2	1	12.01	18D08
60s ribosomal	3	12.01	27C07; 41F01; 45B12
60s ribosomal protein	1	12.01	56F03

60s ribosomal protein L10	8	12.01	27G11; 29B08; 30D05; 10D02; 16H11; 19F08; 21E11; 60D10
60s ribosomal protein L11	3	12.01	25A02; 58D12; 59D04
60s ribosomal protein L12	6	12.01	33C11; 05D12; 24A07; 24A08; 56D09; 58A01
60S ribosomal protein L13	1	12.01	15B04
60S RIBOSOMAL PROTEIN L15	1	12.01	52F10
60S ribosomal protein L17	1	12.01	37C02
60S ribosomal protein L19B	7	12.01	39E11; 42F01; 43F10; 46F08; 03A03; 52H10; 53H01
60S ribosomal protein L20 (L18A)	3	12.01	33F09; 42H11; 20G02
60S ribosomal protein L21	3	12.01	19F07; 24A01; 24B11
60s ribosomal protein L22	5	12.01	30E07; 41E02; 44C04; 52G11; 54E07
60S ribosomal protein L23a	2	12.01	45D01; 25D05
60s ribosomal protein L27-a	4	12.01	32C09; 05D11; 07B07; 23B07
60S ribosomal protein L28	2	12.01	46H01; 17F07
60S ribosomal protein L3	2	12.01	36A02; 52B05
60S ribosomal protein L31	4	12.01	01B07; 04H07; 08B11; 59B05
60s ribosomal protein L32	3	12.01	41H01; 48E03; 19F01
60S ribosomal protein L33-A	1	12.01	16E09
60S ribosomal protein L35a	2	12.01	28G01; 06G06
60S ribosomal protein L37	7	12.01	32F10; 42D12; 12B10; 15F12; 20D05; 49E06; 56F07
60S ribosomal protein L41	3	12.01	31A10; 38D06; 54F05
60S ribosomal protein L44	6	12.01	35H05; 37D01; 40F08; 48H06; 08H03; 57G09
60S ribosomal protein L5 (CPR4)	1	12.01	23D06
60s ribosomal protein L7 subunit	5	12.01	30C09; 36B08; 06E12; 20F11; 24F01
6-phosphofructokinase alpha subunit	1	01.05: 02.01	50H09
6-phosphogluconate dehydrogenase	4	01.05; 02.07	37H04; 03F11; 56C10; 58H09
78 kda glucose-regulated protein homolog precursor	2	99	31F12; 31G01
7alpha-cephem-methoxylase P8 chain	1	01.20	32H08
7-dehydrocholesterol reductase family member	1	01.06	27F09
aarF domain containing kinase 1	1	98	49D07
ABC transporter, ATP-binding protein	1	11.02	57D04
ABC1 transporter; ABC-type ATPase	3	20.04; 20.13; 11.07; 34.01	27H04; 12D11; 12E11
Acetate kinase	4	01.20	01A04; 17F10; 24F06; 49G06;
Acetoacetyl-CoA reductase	1	01.06; 02.13	04B03
Acetylcholinesterase collagenic tail peptide	1	98	25G06
acetyl-CoA hydrolase (acetyl-CoA deacylase)	1	01.06: 73.01	59C10
acetyl-CoA synthetase	1	01.05; 02.45; 42.01	29E05
acid sphingomyelinase	1	01.06; 42.01	28D01
acidic ribosomal protein P1	1	12.01	40B11
aconitase	1	01.01; 01.05; 02.10; 02.16	25F12
aconitate hydratase	1	01.01; 01.05; 02.10; 02.16	30A11
acriflavine resistance protein B	1	11.07	56B05
actin filament binding protein; Abp140p	1	42.04	19C11
actin related protein 2/3 complex subunit 4	1	16.01; 20.09; 42.04	13E08
Actin, gamma	1	42.03; 08.19; 42.04; 73.01	51B11
actin-related protein	1	42.03	38A06
activator of Hsp70 and Hsp90 chaperones	1	11.01	56F02
acyl carrier protein	1	01.06; 02.13	44A10
acyl-CoA dehydrogenase	1	01.06	41B06
Acyl-CoA thioesterase	1	01.06	20C08
Acyl-CoA-binding protein like	1	01.06; 20.13	22H11
acyl-Coenzyme A dehydrogenase short branched chain	1	01.06	52F12
Acyltransferase	2	98	27F05; 57C06

adenylosuccinate lyase	2	01.03	29F06; 31B07
ADP-ATP translocase	4	20.16; 08.04	32B03; 36E10; 40E01; 50G09
ADP-ribosylation factor (ARF)	8	18.02; 20.09	29B04; 29C04; 30D07; 39F02; 43E10; 04E03; 05C07; 19G12
ADP-ribosylation factor 1	1	14.07; 16.19	55C01
ADP-ribosylation factor 2	1	14.07; 16.19	58B09
alanine transaminase	1	01.01: 01.02	50H10
Alcohol dehydrogenase I (ADH 2)	5	01.05; 02.16	31B09; 43E01; 47C02; 02E07; 24G02
aldehyde dehydrogenase	2	02.16; 42.16	29G03; 37E07
aldehyde oxidase	1	01.05	30H10
Alkaline phytoceramidase (Alkaline ceramidase)	1	01.06	52C11
alkaline serine protease	3	70.13	42G05; 44C06; 55C02
allantoate transporter	1	20.19	53D09
alpha-glucoside transport protein	1	20.07	30C06
alpha-mannosidase	2	01.05	34A08; 43C05
Alpha-tubulin	4	42.03	32C11; 01B06; 03B08; 16A08
alpha-tubulin supressor	1	42.03; 42.04	37F05
amino acid permease 2	1	20.01	44G02
amino acid transpoter	1	20.01	43B09
aminopeptidase B	1	70.13	44B04
aminopeptidase yscII	1	01.01; 70.07	59C04
annexin XIV	1	70.02	59F09
Anthranilate synthase component II	1	01.01	44A11
antifreeze glycoprotein	1	11.01	53B04
aorsin	1	14.13; 18.01; 42.25	24F11
AP2 domain transcription factor	1	11.02	27B10
Apo Saccharopine Reductase Chain A	2	01.01	27E08; 19A09
apoptosis-associated tyrosine kinase	1	70.1	28C04
arabinosyl transferase	1	98	26B02
archain/delta-COP	1	98	44D09
Arginine/Ornithine Antiporter	1	20.01	26A08
Argininosuccinate synthetase	1	01.01; 01.02	07C05
Arginosuccinate synthetase	1	01.01; 01.02	01E06
argonaute (plant)-Like Gene	1	99	28E12
Arp2/3 complex subunit homolog ARC15	1	42.04; 70.04	35D07
arsA arsenite transporter, ATP-binding	2	11.07	03C03; 58D08
arylalcohol dehydrogenase	1	01.05; 02.16	43B08
ascorbate peroxidase	1	11.07; 42.16	57E08
aspartate aminotransferase precursor	2	01.01; 01.02	13A05; 25D08
aspartate kinase	1	01.01	26F08
Aspartic protease	2	70.07; 70.13	02G08; 04A04
aspartyl protease	1	70.07; 70.13	52F06
aspartyl proteinase	3	70.07; 70.13	27A02; 29A01; 52F03
ATP citrate lyase	2	01.05;,02.01; 02.22	32F07; 11G03
ATP synthase alpha chain	4	02.13; 34.01; 08.04	44D05; 57A07; 57H02; 60D04
ATP synthase beta chain	2	02.13; 08.04; 34.01	02B05; 57G07
ATP synthase D chain	2	02.13; 34.01; 08.04	46H02; 52B02
ATP synthase oligomycin sensitivity conferral protein	1	02.13; 08.04; 34.01	22C06
ATP synthase protein 9 (Lipid-binding protein)	3	02.13; 34.01; 08.04	44E08; 48H01; 52C03
ATP synthase subunit 4	3	02.13	26A09; 34E11; 55H10
ATP(GTP)-binding protein	1	99	42E02
ATPase inhibitor	11	02.45	31A04; 37A02; 38B08; 39F07; 41H08; 47D10; 47E05; 47F06; 22A11; 24H09; 56A09
ATP-dependent protease ATP-binding subunit	1	70.10: 70.13	26G05
ATP-dependent RNA helicase	2	11.02; 11.02	40E09; 52E08
Atrophin-1	2	98	27D10; 40D04
a-type carbonic anhydrase	1	20.09	46E08
autophagocytosis protein Atg3p	1	70.02; 36.09	50E12
autophagy protein AUT7	1	70.13: 08.22	21G04
Avenacinase	1	11.05 *	07G04

basic proline-rich protein	1	99	27D01
BET1 protein	1	20.09; 70.08	38H06
beta-galactosidase	1	01.05	49H09
Beta-glucanase/Beta-glucan synthetase	1	01.05; 42.01	57H03
Beta-glucosidase	1	42.03; 73.01	08D11
beta-glucosidase 5	2	01.05	43B05; 57B10
bll6449	1	99	09F04
BNS1 protein	2	42.03	10H10; 20H07
BPLF1	1	99	20B09
branched-chain alpha-ketoacid dehydrogenase kinase	1	01.01	57E04
branched-chain amino acids aminotransferase	2	01.01	12F08: 24C03
BRR5 (component of pre-mRNA polyadenylation factor PF I)	1	11.02	14C09
Brt1	2	34.11	22E11: 56C03
Btn1p	1	34.01	31C10
Bv8/prokineticin 2-like protein	1	99	39C05
bZIP transcription factor family protein	2	11.02	26C01; 32C10
C1 tetrahydrofolate synthase C1-THFS	1	01.01; 01.42.; 01.07; 11.02	10E06
C-4 methyl sterol oxidase	28	01.06	27H05; 29H04; 30D09; 31D10; 31G08; 33A12; 35G07; 38B02; 38C12; 40C10; 44C11; 45D11; 47E04; 48F04; 04A10; 04C09; 11B02; 18E03; 25A03; 49E12; 51H02; 52A04; 55E02; 55F05; 58C10; 58F05; 59C02; 60G04
C-8 sterol isomerase (Delta-8--delta-7 sterol isomerase)	1	01.06	22C09
Ca/CaM-dependent kinase-1	2	42.01	48C02; 21D10
calcium channel	1	20.01; 34.01; 34.11	18H09
Calmodulin (CaM)	3	42.03; 08.19; 34.11; 73.01	37G10; 40H10; 45E12
cAMP-dependent protein kinase regulatory chain	1	01.05; 11.02; 11.01; 42.04; 73.01	40E07
Capsular associated protein	1	98	09G11
carnitine acetyl transferase FacC	2	20.13; 08.04	35C09; 36H04
CAS/CSE1 segregation protein CG13281-PA	1	70.02; 08.01; 42.10	52B07
catalase/peroxidase HPI	3	11.07	38B04; 43A07; 06H04
cation efflux family protein	1	20.04; 34.01	57A01
Cdc45	1	42.01; 42.03	32D04
Cdc48p	1	42.03; 70.13	58C03
Cell cycle control protein cwf18	1	42.03	57H01
cell cycle inhibitor nif1	1	42.03	41B11
Cell division control protein 4	1	42.01; 42.03; 73.01; 70.10	45H04
cell division cycle 2 homolog -like	2	42.03	28F02; 50C04
cell wall surface anchor family protein	1	42.01	39H07
cellular aspartic protease	1	70.07	51C03
cerato-ulmin precursor	36	11.05 *	28F12; 33G08; 34G09; 38H08; 40A01; 40H12; 43E05; 45G05; 46F05; 47C10; 48A10; 01A12; 02A04; 02H05; 04G02; 04G03; 08F12; 08H06; 13B09; 20G03; 21A04; 2104; 23B03; 25D09; 25H03; 51G02; 52G06; 54H02; 54H11; 55A07; 55E06; 56D05; 56D12; 58G04; 59G03; 60E03
CG5161-PA	1	99	07D08
Chain A, 4ank: A Designed Ankyrin Repeat Protein With Four Identical Consensus Repeats	1	34.11	23B06
Chain A, Translationally Controlled Tumor-Associated Protein P23fyp	1	99	56E06
Chain E, Endothia Aspartic Proteinase (Endothiapepsin)	2	70.07	54E02; 60G09
Chain E, Proteinase K	1	70.13	51G05

chaperonin CCT4	2	70.01	28F04; 47F03
chaperonin, t-complex-type	1	70.01	59A06
chitin synthase	2	01.05; 42.03; 34.11; 73.01	39F12; 48F03
Chitin synthase 3	1	01.05; 34.11; 73.01	08G03
chitin synthase C	3	01.05; 73.01	31F07; 43A11; 59F03
Chitinase	1	01.05; 73.01	04D05
chitinase A	2	01.05; 42.03	38D08; 46E07
Chorion protein 38	1	42.02	12D07
chorismate mutase (CM)/prephenate dehydratase	1	01.01	18H03
chromosome partitioning protein ParB	1	42.03; 70.10	42E04
chromosome segregation protein cut3	1	42.10	55E01
Chromosome segregation protein pcs1	1	11.02; 70.13	55G03
CipA protein	2	01.20	11F09; 51G04
CipC protein	1	70.13	46G08
circumsporozoite protein	1	98	29F12
Cisplatin resistance protein CRR9p	1	98	03E01
cis-prenyltransferase	1	42.01; 70.07	50B11
Class 2 transcription repressor (NC2)	1	11.02	46C05
class IV chitin synthase	1	01.05; 42.03; 34.11; 73.01	22B10
class V chitin synthase	3	01.05; 34.11; 73.01	30G09; 41G01; 20C11
cleavage and polyadenylation specificity factor 30 kDa subunit	1	11.02	44G01
cleavage stimulation factor 64	1	11.04	48B07
CLL-associated antigen KW-12	2	99	26A01; 07D01
clock-controlled protein 6	1	70.2	34B10
coatomer alpha subunit	2	20.09; 70.03	13E10; 51F02
coatomer beta subunit	2	20.09; 70.03	26H04; 47H05
coatomer complex COPI delta-COP	1	20.09; 42.09	48E11
coatomer gamma subunit	1	20.09; 70.03	26C05
cobalamin-independent methionine synthase	1	01.01	38G01
Cobyric acid synthase	1	98	50E06
Cofilin	3	11.01; 42.01; 70.04	31G11; 45E03; 18C05
Collagen alpha 1(XII) chain precursor	1	98	26E12
Colony 1	1	70.01	06E02
component Tra1 of the SAGA complex	1	11.02	33C12
Conidiophore development protein hymA	1	73.01	27B11
COPII-coated vesicle proteins	1	20.09; 70.09	37D11
copper-zinc superoxide dismutase	5	11.01; 11.07	33E10; 14D09; 21D05; 24F02; 58B03
coproporphyrinogen oxidase precursor	3	01.07	37E10; 51D10; 51G11
cross-pathway control protein 1	3	01.01; 11.02; 16.03; 32.01; 70.10	34H12; 38G10; 02E09
CsgA	2	01.05; 73.01	08B09; 08H09
cullin-like protein1	1	01.01; 01.02; 42.01; 70.07; 16.01; 34.11	25F09
curved dna-binding protein	2	10.03; 16.03	47C09; 07F05
cyclase	1	42.03; 42.01; 34.11; 73.01	51D12
cyclase SCIF3.09c	1	11.04	36D09
cyclin-dependent kinase 4	1	42.03	29C08
Cycloinulo-oligosaccharide fructanotransferase	1	01.05	09C09
cyclopentanone 1,2-monooxygenase	1	01.07	26G12
Cys2/His2 zinc finger protein (rKr1)	1	01.05; 11.02; 16.17	42D08
cystathionine beta synthase	1	01.01	52D07
cysteine dioxygenase	7	01.01	27C02; 30B10; 45C04; 07D09; 25E08; 53C09; 56F12
cytidine deaminase	1	01.03	47H06
cytochrome b5	1	01.06	37E12
Cytochrome C	6	02.13	30E01; 33A11; 22D10; 24G10; 54A07; 54D01

Cytochrome C oxidase	2	02.13	04C04; 05B12
Cytochrome c oxidase polypeptide I	2	02.13	26D07; 23F01
cytochrome c oxidase subunit V	1	02.13	59B09
cytochrome P450	1	11.07; 42.01; 73.01	29E06
cytoplasmic dynein light chain Dlc1	8	10.03	29G04; 31H07; 06B07; 12B08; 20B04; 22E12; 23H03; 24A11
cytoskeleton assembly control protein homolog Sla2	1	01.03; 70.10; 08.19; 34.11; 42.04; 73.01	40E03
cytosolic epoxide hydrolase	4	11.07	47D11; 09C07; 24D05; 52H04
DC13 protein	1	42.03	27A10
dCMP deaminase	1	01.03	25D12
dehydrin	1	99	60A10
dehydrogenase	2	99	34B03; 43H08
delta(24)-sterol C-methyltransferase (ERG6)	1	01.06	25F07
delta-9 fatty acid desaturase; stearoyl-CoA desaturase	2	01.06	33H12; 56D10
Demethylmenaquinone methyltransferase	1	98	28D04
DFG5 protein	1	42.04; 73.01	50A08
D-fructose-6-phosphate amidotransferase	1	01.05; 34.11	51E01
D-hydroxyacid dehydrogenase	1	01.01	49D04
diacylglycerol acyltransferase	1	01.06; 73.01	39C11
dihydrofolate reductase	1	01.01; 01.03; 01.07	35H03
dihydrolipoamide dehydrogenase	1	01.01; 02.10	38G12
Dihydroorotase and related cyclic amidohydrolases	1	01.03	60E08
dioxygenase	1	11.07	42H05
D-lactate dehydrogenase	5	02.16; 42.16	31A08; 31H09; 43E08; 07F03; 18C03
D-mandelate dehydrogenase	1	01.05; 02.13	43A03
DNA j domain containing protein	2	16.03	46C09; 25B01
DNA lyase, endonuclease	1	01.03	44A06
DNA mismatch repair protein MutS	1	42.01	27C03
DNA photolyase	1	42.01	52F09
dna polymerase IIII (alpha chain) protein	1	42.01	38E10
DNA-binding protein creA	1	11.02	33E12
DNA-dependent RNA polymerase II RPB140	2	11.02	28F07; 56C09
DNA-directed DNA polymerase III	1	42.01	24H04
dna-directed rna polymerase i ii and iii 24 kd polypeptide	1	11.02; 11.02; 11.02	51D01
DNA-directed RNA polymerase II chain Rpb7	1	11.02; 11.01	60A09
DNAJ-like protein homolog	4	70.01	27D07; 04E09; 09A09; 14B11
dnaK-type molecular chaperone BiP	19	98	27G04; 31B05; 32B05; 32H04; 36D01; 38A11; 38F04; 39F09; 40C05; 03C04; 16B03; 56G11; 56G12
Dolichol phosphate-mannose biosynthesis regulatory protein	1	01.05; 70.07	44D12
Dolichyl pyrophosphate Man9GlcNAc2 alpha-1,3-glucosyltransferase	1	01.05; 70.07	01G10
dolichyl-phosphate-mannose--protein mannosyltransferase	1	01.05; 70.07	29G07
double homeobox protein	1	98	54B02
dpm2-like protein	1	01.05; 16.01	15D09
DREB1B	1	16.03	53G07
Dynein heavy chain	1	42.03; 08.07; 08.22	28H12
Dynenin heavy chain, cytosolic (DYHC)	1	10.03, 20.09	01B02
ECM14 protein	1	42.01	46D03
efflux pump	1	20.01; 20.09; 32.05; 32.07	59G05
eIF-1A	4	11.06	37G07; 45C01; 07G02; 11C01
elongation factor 1 beta subunit	2	12.04	39B03; 56D06
Elongation factor 1-alpha	19	12.04	27F10; 30E03; 34F04; 34H01; 38E01; 39E06; 41C08; 41H10; 42A04; 42D10; 45E07; 06C11; 13G10; 20E08; 25D11; 50F12; 51F08; 55F03; 56C11
elongation factor 1-gamma	1	12.04	51F01

elongation factor 2 (EF-2)	1	12.04	58B06
endo-1,4-beta-glucanase	6	01.05	30E09; 39A05; 40F09; 01F04; 51B10; 58B12
endo-alpha-1,5-arabinanase precursor	1	98	19E04
endochitinase	1	01.05: 42.03	48C09
Endoglucanase	1	70.10; 42.04; 73.01	32G04
endonuclease III	1	42.01	28A07
Endothiapepsin precursor	1	70.13	03E11
enolase	8	01.05; 02.01	33D10; 35H02; 37C04; 37H06; 06E05; 11F12; 11G12; 52D05
enzyme involved in pigment biosynthesis	1	01.20	40E05
epithelial zinc-finger ezf protein	1	11.02	41H06
EpsE glycosyltransferase	1	98	16H10
Erwinia chrysanthemi IndA protein	1	01.01	11B04
Esterase	2	01.06	26D04; 32G05
Esterase D	1	01.06	40E06
ethionine resistance gene	1	11.07	29E09
eukaryotic initiation factor 4a (EIF4A)	3	12.04	38B11; 04B02; 53B02
Eukaryotic translation initiation factor 2	1	12.04	35E06
Eukaryotic translation initiation factor 3 subunit 7 homolog (Microtubule destabilizing protein moe1)	2	11.06	41E10; 44H12
eukaryotic translation initiation factor 3, subunit 1 alpha, 35kDa	1	11.02	57D02
eukaryotic translation initiation factor 5	3	12.04	30H09; 39B06; 48G07
Eukaryotic translation initiation factor 5A (eIF-5A)	7	01.03; 12.04	03F06; 06H05; 10F05; 15C03; 19G03; 24C08; 54D05
Exportin 1	1	11.04; 08.01	57G02
extensin-like protein	1	98	26A10
F-actin capping protein beta subunit	1	34.11; 42.04; 73.01	30A04
farnesyl pyrophosphate synthase	1	01.06	28A08
fatty acid hydroxylase	1	01.06	36F11
Fatty acid synthase alpha-subunit	1	01.06; 70.10	06C09
fatty acid synthase, beta subunit	1	01.06	56A12
fatty acid transporter FAT2	2	20.13	22C05; 22H08
ferredoxin-like iron-sulfur protein	1	01.07; 36.01	44E10
fibrillin-1 precursor	1	98	28B03
fibroin	1	98	47E03
flavohemoglobin	5	11.01	30A08; 32C08; 42E09; 06F03; 53D07
flavoprotein subunit	1	20.03	46F07
fructose-bisphosphatase	1	01.05; 02.01	40A12
fructosyl amino acid oxidase	1	01.01	48A09
fumarate reductase	1	02.10	54H04
G10 protein homolog	2	98	07H02; 22H05
G2 allele of skp1 suppressor; subunit of SCF ubiquitin ligase	1	42.03; 70.13; 42.01	53H09
GAF domain-containing protein	1	99	26A06
gamma-adaptin precursor	1	70.10; 08.19	37B01
Gbeta like protein	3	16.19; 42.01; 73.01	35F07; 39C09; 46E04
GCN5 acetylase	1	42.01; 11.02; 70.07	20E02
gcn5-related N-acetyltransferase (GNAT)	3	42.03; 70.07	43F02; 04B06; 54D10
GDP/GTP exchange factor Rom2p	2	01.05; 18.02; 42.05; 34.11; 42.01; 42.04; 73.01	04A05; 20H12
GDP-fucose transporter 1	2	98	20D04; 55C09
gephyrin-like protein	1	70.02; 16.01; 70.04	41G06
geranylgeranyl pyrophosphate synthetase	2	01.06	29A03; 24D03
gibberellin 3-oxidase	1	01.06; 36.20	37E09
Gim complex component GIM3	1	42.03; 70.01; 42.04	39B11
Glc8 protein	1	01.05	09H08
glucan 1,3 beta-glucosidase	4	42.03; 42.01	29C12; 29F05; 34C09; 58C04
glucan synthase	1	01.05	36G03

glucoamylase 1 (glucan 1,4-alpha-glucosidase)	1	01.05	12A12
glucoamylase I	1	01.05	54D08
Glucokinase	3	01.05; 02.01	46F11; 47D05; 07G11
Glucosamine--fructose-6-phosphate aminotransferase	1	01.05; 34.11	21H11
glucose repressor	1	01.05; 11.02	37B02
glucose-6-phosphate 1-dehydrogenase	1	01.05; 02.07; 11.01; 11.07	23C09
Glucose-6-phosphate isomerase	1	01.05; 02.01	53C07
glutamate carboxypeptidase	3	70.07	30C08; 24E04; 60F07
Glutamate decarboxylase	4	01.01	06A02; 19F10; 49H07; 53E03
glutamate synthase	1	01.01; 01.02	47G05
glutaminase A	3	01.01	28B04; 37E06; 39H05
glutamine amidotransferase, class I	1	01.01	29F11
glutamine synthase	1	01.01	27C04
Glutamine synthetase	2	01.01	36A11; 02F06
glutamyl-trna synthetase	1	12.10	43A04
Glutaredoxin	3	01.03; 11.01; 11.07	32A07; 47F01; 51E04
Glutathione peroxidase paralogue	1	11.07	10C10
Glyceraldehyde 3-phosphate dehydrogenase	41	01.05; 02.01	26H10; 27H01; 29A04; 31G07; 32E01; 32E12; 33B12; 35B09; 37E01; 39F06; 40C01; 42H01; 44F10; 45G11; 01D03; 01G01; 02B01; 03B03; 04H10; 05E01; 06E07; 12A07; 12B07; 14B12; 15B09; 20B03; 20D10; 20E10; 21H02; 23C03; 23F08; 50E04; 51H03; 53B06; 53C04; 54G12
glycerol dehydrogenase	1	01.05; 11.01; 34.11	53D01
glycine dehydrogenase (decarboxylating)	1	01.01	60H09
glycine-rich RNA-binding protein grp1a	1	01.05	15C06
glycogen branching enzyme	1	01.05; 02.19	45F12
glycogen phosphorylase 1	4	01.05; 02.19	47F12; 48D05; 09B04; 09C04
Glycolipid 2-alpha-mannosyltransferase (Alpha-1,2-mannosyltransferase)	2	01.05; 70.07	44H08; 02C06
glycoprotein gp2	1	70.02	47G11
Glycosyltransferase	1	01.05	18F05
glyoxalase I	1	01.01	26B11
Golgi-specific brefeldin A-resistance guanine nucleotide exchange factor 1	1	40.01	28A01
GPI-anchored wall protein transfer 1	1	20.11	45G06
GTPase activator protein	2	18.02	33D01; 49F02
GTP-binding protein beta subunit-like protein	1	42.01; 34.11; 73.01	27E04
GTP-binding protein SAR1	2	16.19; 20.09	43G05; 59C03
Guanine nucleotide-binding protein beta subunit-like protein	1	42.01; 34.11; 73.01	55G05
Guanylate kinase; Guk1p	1	01.03	05F11
H+-transporting ATP synthase protein 6 homolog	1	20.04; 34.01	27C05
haloacid dehalogenase-like hydrolase family	1	11.07	46B08
heat shock protein 30	2	11.01; 34.01	03B10; 19C10
heat shock protein 70	14	14.01; 32.01; 70.03	26G11; 31C08; 34G06; 44C01; 46B06; 10C08; 19C12; 23D01; 51F07; 52B09; 54D03; 57C09; 57G12; 58G06
heat shock protein 78	11	11.01	31F09; 32A05; 33A07; 37F01; 42F12; 45F01; 24B03; 54H06; 58B08; 58H05; 59H11
heat shock protein 80	10	70.01; 32.01	27F11; 30A06; 37D04; 08F08; 09B08; 13A11; 14D02; 15E12; 18B04; 25C10
heat shock protein 90	5	02.19; 11.01; 34.11; 73.01	30E02; 39C06; 40C03; 40D11; 45A12
heat shock protein CLPA	7	11.01	42E10; 44A02; 12E05; 13B05; 13C05; 19G11; 22G03
Heat shock protein HSP1	4	14.13; 16.19; 32.01	28D05; 38B12; 51B04; 51C05

heat shock protein; hsp48	1	01.05; 02.01	27D05
helix-loop-helix DNA binding protein	1	16.03	57H05
Hemagglutinin	1	98	55B07
hemolysin	1	01.04; 01.06	51B01
HEP200 protein	1	42.01	27C06
hepatocellular carcinoma-associated protein HCA1	1	99	45H03
het-c	1	43.01	17A10
het-c2 protein	1	34.11	30B04
HEXOKINASE	3	01.05; 02.01	35D01; 41F12; 57E07
hexose transporter	1	20.07	24F12
high affinity methionine permease; Mup1p	1	01.01; 20.10; 08.19	04C10
high molecular weight glutenin subunit x	1	99	60H03
high-affinity branched-chain amino acid transport system, permease protein	1	20.10; 08.19	55D02
histidine kinase	1	14.07; 30.05	43H06
histidine triad superfamily, third branch; Hnt3p	1	99	22G12
Histone H2A	16	11.02	26E08; 28A10; 31D01; 35E11; 41D05; 47F08; 02D08; 05A10; 12E09; 12E12; 18H05; 52C06; 56F10; 57A08; 58E07; 59A10
Histone H2A F/Z family member HTZ1	1	11.02	08E07
histone H2B	10	11.02	29A08; 35B11; 45E09; 01F09; 03H07; 51D06; 52A10; 53G08; 54C03; 60C06
histone H3	14	11.02	28D12; 30F10; 31F02; 35D12; 37C05; 48D01; 03A09; 08B04; 20C10; 23B10; 24G11; 52A03; 56C05; 56D04
histone H4	7	11.02	33E02; 34E01; 42A12; 09F11; 10D09; 53D03; 59H12
Histone H4.1	5	11.02	28H02; 35F02; 55C03; 56C08; 60B09
histone H4.2	1	11.02	26C06
HMW glutenin subunit	1	98	30D06
HNH endonuclease family protein	1	01.03; 11.02	52G10
Homeobox protein PKNOX1	1	99	36G07
homocitrate synthase	1	01.01; 01.05	50G10
homolog of mammalian eIF2A	1	12.04	26F10
host-specific AK-toxin Akt2	4	32.05	33A06; 35G10; 10E04; 25C11
Hsp70 (Ssa1p) nucleotide exchange factor	1	18.02	53E01
human downs syndrome critical region homolog	1	99	47A02
hydrolase, isochorismatase family	2	99	18C02; 50D10
Hydroxypyruvate isomerase	1	01.05; 02.01	02G06
IclR-type transcriptional regulator	1	11.02	35A10
IgE-binding protein	1	98	58F07
IgE-dependent histamine-releasing factor homolog	2	12.01	29A05; 37H12
import receptor MOM19	1	70.02; 08.04	23F09
infection structure specific protein	1	11.05 *	16H05
Inorganic pyrophosphatase (Pyrophosphate phospho-hydrolase)	2	01.04	32F05; 51C11
integrase/recombinase	1	42.01	28D08
involved in Chs3p export from the ER; Chs7p	1	20.09	58E05
Involved in copper metabolism and assembly of cytochrome oxidase; Cox17p	1	02.13; 70.10; 34.01	20A04
Involved in ubiquinone biosynthesis.; Coq4p	1	01.07	23H02
involvement in de-repression of telomeric silencing	1	11.02; 14.07	52C05
Iron inhibited ABC transporter 2	2	01.04;12.04; 16.19; 18.02; 36.03; 32.01	06C02; 15A05
Isoamyl alcohol oxidase	1	01.05; 02.16	07F06
JUN kinase activator protein	1	11.02; 11.07	53D08
Ke3	5	12.01	27G05; 03E02; 06D08; 18G02; 59D11
keratin associated protein 4.15	2	99	30B06; 11C07
ketoacyl reductase	1	01.06; 02.13	60F04

Ketol-acid reductoisomerase	1	01.01	55C12
kinase	10	14.07; 30.01	35H10; 14C10; 21E01; 51E12; 52E04; 54E01; 54H10; 55G09; 58E04; 59F12
kinase substrate HASPP28	1	99	53H12
kinesin-related protein 2	1	42.03	26B09
Klu	1	99	28B02
Large exoproteins involved in heme utilization or adhesion	1	16.17	40A03
Lipase (lipP-1)	2	01.06	19B11; 60E05
Loc55831-prov protein	1	99	57D10
lon proteinase	1	14.13; 32.01; 70.16	27A04
long-chain-fatty-acid--CoA ligase	2	01.06; 20.13; 08.10	35B03; 55H01
low density lipoprotein receptor protein LRP1B/LRP-DIT	1	18.01; 18.02	60F12
lysophospholipase A	1	01.06	39C07
magnesium-dependent phosphatase-1	1	01.05; 02.19; 42.03	21E09
major facilitator MIRB	1	20.01; 34.01	34G03
malate dehydrogenase	1	01.05; 01.06; 02.22	55F06
mandelate racemase	1	98	50A03
mannose-1-phosphate guanyltransferase	4	01.05; 42.01	37B11; 43G01; 02E11; 11B09
MAP kinase kinase kinase	2	11.01; 34.11	26D10; 26E07
maturase	1	11.02	17D04
maturase K	1	11.02	27D12
MBF1	2	11.02	27G03; 33D08
melanocyte proliferating gene 1	1	99	26A12
membrane protein YGR149w	1	99	33H11
Membrane protein YNL115c	1	99	01A06
membrane protein YNL266w	1	99	26A04
membrane-bound non-heme di-iron oxygenase	1	11.07	30G01
metacaspase	1	70.10 *	58A08
methionine aminopeptidase; Map1p	1	70.07	34C07
methionyl-tRNA formyltransferase protein	1	11.02	29F09
methyl-accepting chemotaxis I (serine chemoreceptor)	2	98	25C07; 59D03
Methyltransferase	4	99	29E12; 43E12; 20G11; 54C02
MEVALONATE KINASE	1	01.06	29D07
MGC14151 protein	1	99	24B09
Micronuclear linker histone polyprotein	2	42.03	09A11; 14F01
microtubule-associated protein EB1	1	42.03	52F07
MIT family metal ion transporter	1	20.04; 34.01	01H12
mitochondria-associated granulocyte macrophage CSF signaling molecule	1	32.05	53D04
mitochondrial citrate synthase	1	01.05; 02.10; 42.16	53A01
Mitochondrial Distribution and Morphology; Mdm35p	1	42.16	35B01
mitochondrial elongation factor G	2	12.04	51G07; 54G02
Mitochondrial import inner membrane translocase subunit TIM8	1	14.04; 20.09; 42.16; 70.16	08G11
mitochondrial inner membrane protease subunit	1	42.01; 70.02; 70.07	24A05
Mitochondrial processing peptidase beta subunit	1	70.07	13D10
mitochondrial ribosomal protein L23	1	12.01	60E06
Mitochondrial ribosomal protein MRPL24	1	12.01	27B03
mitochondrial rRNA processing protein PRP12	2	11.02	37H10; 16A12
mitogen activated protein kinase	1	42.03; 70.07; 16.01; 18.02; 42.01; 34.11	44B11
MKT1 protein	1	38.03	56G10
Mlx interactor beta	1	99	39D04
Modulates cytochrome c oxidase activity; Cox13p	5	02.13	38G11; 07E10; 12F09; 12G09; 25G02
molybdopterin synthase large subunit CnxH	1	01.07	57F03
monosaccharide transporter	5	20.07	29A02; 37F02; 10G12; 49D06; 57F05
mportin-alpha export receptor	1	70.02; 08.01; 42.10	52C07
Muc5b protein	1	42.27	59C09
mucin	1	42.01	55F07
Multidrug resistance protein 2 (P-glycoprotein 2)	1	20.28; 11.07	36G01

Multidrug resistance protein MDR	1	20.28; 11.07	31G10
MUM2	1	42.03: 73.01	30G03
MUS38	1	10.01; 70.10	44B02
myb family transcription factor	1	11.02; 11.02	28B11
myc-type bHLH transcription factor	2	11.02	54B12; 58A10
Myocardin-related transcription factor A	2	11.02	03B04; 20H08
Myocardin-related transcription factor B	6	11.02	32E09; 39C04; 39H08; 46B04; 04B09; 55D08
myo-inositol-1-phosphate synthase	6	01.05	41A02; 01E04; 03C01; 07A04; 52E12; 52H06
Myosin IC heavy chain	1	10.03; 42.25; 73.01	49F08
N1315	1	99	33B01
N-acetylmuramoyl-L-alanine amidase AMIC precursor	1	70.07; 70.13	45E01
N-acetyltransferase	1	01.06; 70.07	53F09
NAD-dependent formate dehydrogenase	4	01.05	35D04; 41A01; 15A12; 16D04
NAD-dependent protein deacetylases, SIR2	1	11.02; 34.11	05F10
NADH dehydrogenase (complex I)	3	02.13; 42.16	45G09; 07C01; 58C09
NADH dehydrogenase (ubiquinone) 1 beta subcomplex	1	02.13; 42.16	58H03
NADH dehydrogenase (ubiquinone) 29/21K chain precursor	1	02.13; 42.16	52C08
NADH dehydrogenase subunit 4	1	02.45	27A09
NADH dehydrogenase subunit 5	1	02.45	26E03
NADH2 dehydrogenase	3	02.13; 42.16	34D12; 50B12; 51D07
NADH-cytochrome b5 reductase	1	02.13	23E09
NADH-ubiquinone oxidoreductase 21.3kDa subunit	1	02.13	11E09
NADH-ubiquinone oxidoreductase 36.8 kDa	2	02.13	02F12; 22G10
NADH-ubiquinone oxidoreductase chain 3	1	02.13; 42.16	45G12
NADP-dependent mannitol dehydrogenase	2	01.05; 02.16	26G09; 27G06; 57B11
NADPH-cytochrome P450 reductase	1	01.06; 11.07	60E10
nascent polypeptide-associated complex alpha polypeptide	1	01.05; 11.02	57C10
Nca2p: Expression regulator of ATP synthase	1	02.13; 11.02; 42.16	59C06
negative regulator Moe1	1	10.03; 16.01; 70.10	46E06
Negative regulator sulfur controller-3; sconCp	1	01.02	01E09
neuronal gamma-aminobutyric acid-glycine vesicular transporters; Avt1p	2	20.10; 08.13	13C10; 24A12
Ni,Fe-hydrogenase III large subunit	1	98	27F08
NifS-like protein; Nfs1p	1	34.01	51G03
Nifu-like protein; Nfu1p	1	01.02; 34.01	33C01
NIPSNAP1 protein	1	99	40H01
Nitroreductase family protein	1	01.02	11C06
Nonhistone chromosomal protein 6A	7	73.01	02E06; 06C05; 08A07; 09D10; 15B01; 18G09; 19E12;
Nonhistone chromosomal protein 6B	6	73.01	37C03; 37G11; 45E11; 48A12; 18F09; 56H02
nonhistone protein 6	15	73.01	29C01; 34D09; 37C06; 37G01; 39E04; 40D01; 15C11; 52H05; 54F10; 57E02; 58F03; 58F12; 59G06; 60B05; 60F10
Norsolorinic acid reductase	1	01.05; 02.16	23C07
novel protein with ATPase domain	1	99	28A02
NST UDP-galactose transporter	1	01.05: 20.07	40A11
nuclear distribution gene C homolog	1	10.03	24F07
nuclear matrix transcription factor	1	11.02	44H06
nuclear pore complex protein sonA	1	11.04; 08.01	40D09
nuclear protein p30	2	70.10	30B12; 18B06
nuclear transport factor 2	1	20.09	40D02
nucleoside diphosphate kinase	1	01.03	35E12
nucleosome assembly protein I	1	42.03; 70.10; 42.10; 73.01	48F05
ochre suppressor tyr-tRNA	1	98	53F07
oleate delta-12 desaturase	1	01.06	30B11
Oligomycin sensitivity conferring protein	2	02.13; 08.04; 34.01	32B07; 40F07;
oligosaccharyl transferase STT3 subunit	1	01.05; 70.07	21D08
oligosaccharyltransferase	1	01.05; 70.07	43H02

Ornithine carbamoyltransferase	1	01.01	06H12
ornithine-N5-oxygenase	1	98	16B10
Ornithine--oxo-acid aminotransferase	3	01.01; 01.02	41A11; 20G06; 23C02
Orotidine 5'-phosphate decarboxylase	1	01.03	47A12
osmotic sensitivity MAP Kinase	2	01.05; 11.01; 34.11	37H02; 39G05
outer membrane autotransporter	1	99	12D08
outer mitochondrial membrane protein porin	1	20.03; 20.09; 70.16	59H09
oxidoreductase	3	99	34A02; 42D01; 60A04
p32INGL	1	99	41D09
PAK kinase	1	34.11; 42.04; 73.01	50B04
paraben-hydrolyzing esterase precursor	1	98	35C07
paraflagellar rod protein	1	98	27E03
parasitic phase-specific protein PSP-1	1	98	21H03
Partner of Nob1; Pno1p	2	98	13F05; 18C04
PAS domain-containing Serine/threonine Kinase; Psk2p	1	01.05	36C01
pathogenesis-related protein precursor	1	32.05	27F01
PDGFA associated protein 1	1	98	21C09
PE-PGRS FAMILY PROTEIN	2	99	54D12; 60G12
Pepsinogen	6	70.07; 70.13	36H10; 46E12; 06A09; 21A05; 21B05; 59B10
Peptide permease Ptr2	2	20.11	47F04; 10D04
Peptidyl-prolyl cis-trans isomerase (Cyclophilin)	10	11.01; 70.01	31A09; 33E11; 36E03; 48B09; 01A09; 03A11; 08G08; 08F10; 11B11; 22G04;
peptidylprolyl isomerase D (cyclophilin D)	1	70.01	30E08
peptidylprolyl isomerase-like 4	2	11.01; 70.01	26E09; 08E10
Peripheral-type benzodiazepine receptor	2	01.06; 70.01; 70.10	07C06; 53C02
Peritrophin	1	98	27G08
Permeases of the major facilitator superfamily	1	20.09	06F04
peroxiredoxin 5	1	11.07	46D05
Peroxisomal hydratase-dehydrogenase-epimerase	1	01.06; 02.25; 42.01	15G10
peroxisomal membrane protein 4	1	70.07; 70.10	55B12
peroxisomal protein POX18	4	70.02; 08.10; 42.19	28E02; 29H09; 32G08; 19C08
P-GlycoProtein related (pgp-5)	1	20.01	26C07
PHD finger	1	11.02	46A11
PHD zinc finger protein rhinoceros	1	98	54F11
PheA	1	99	48A05
phosphate ABC transporter, periplasmic phosphate-binding protein	1	01.04;12.04; 16.19; 18.02; 36.03; 32.01	12B01
phosphate translocator-related	1	20.09	34E10
phosphate transport protein MIR1	1	20.03	25E03
phosphatidylserine decarboxylase	1	01.06	58F06
Phosphinothricin N-acetyltransferase	1	99	01E12
Phosphoglucomutase (Glucose phosphomutase)	3	01.05; 02.19	36F01; 02C01; 03E04
Phosphoglycerate dehydrogenase; SerA2	2	01.01	31D06; 49H12
Phosphoglycerate kinase	15	01.05; 02.01	26F11; 27E10; 27F06; 32A01; 35E07; 39A10; 01G03; 04A07; 10B12; 16E01; 18C06; 24E05; 52E06; 56E08; 56H03
phosphoglyceromutase	1	01.05; 02.01	40G09
Phosphoketolase	31	01.05; 02.07;	26E10; 26G03; 27C08; 30D12; 31H10; 36B11; 36F12; 41F02; 42H06; 44H02; 44H11; 45H02; 47G08; 48F02; 01F05; 02E08; 02H02; 05B11; 10D11; 14C11; 22A04; 22E03; 23E03; 23H11; 25H07; 49C11; 49G07; 50E02; 51E05; 54G01; 60E01
phospholipase D	1	01.06; 42.03; 73.01	39D11
Phosphopantothenoylcysteine Synthetase	1	01.03	54H03
Pirin	2	11.02	02F02; 60E12
plasma membrane ATPase (proton pump)	4	20.04; 34.01	26H06; 40F05; 18D10; 49F03
polyA-binding protein	1	11.02; 12.04	45E10
Polyamine transport protein Tpo4p	1	20.28; 11.07	39E03

polyketide synthase	10	11.07	26G02; 31C02; 42H09; 02D07; 07B10; 07C10; 13G05; 16H06; 51C10; 54A09
polyphosphate synthetase	1	01.04	49G04
Polyprotein	1	11.04; 16.03	26B06
poly-ubiquitin	8	70.10; 70.13; 11.01; 73.01	30F07; 37B10; 48F10; 05D05; 09A04; 19H03; 23A05; 58D01
porphobilinogen synthase	1	01.07	17D12
potassium channel	1	20.01; 08.16; 34.01	41D04
prefoldin subunit 1	1	11.02; 34.11	59G02
prefoldin subunit 2	1	70.01	45D09
prefoldin subunit; molecular chaperone non-native actin binding complex subunit	1	70.01	59E12
Pre-mRNA splicing factor PRP8	1	42.03; 11.02	60F06
prephenate dehydrogenase	1	01.01	51E03
prepilin peptidase transmembrane protein	1	70.07; 70.13	60C01
preprotein translocase, SecA subunit	1	20.11	28G10
Preproteolipid	2	99	09D09; 17C03
Probable 26S proteasome regulatory subunit rpn3	1	70.13	60A01
probable cytosceletal binding protein	2	16.07	38H01; 60B02
probable helicase	1	11.04	42H10
probable membrane protein YPL183c	1	99	37A09
Profilin A	2	42.04; 73.01	18H11; 60B04
profilin P	1	42.04; 73.01	36G11
programmed cell death 6	2	70.1	41F03; 51A03
proliferating cell nuclear antigen (PCNA) (cyclin)	2	42.01	37B05; 20A12
proline-specific permease proY	1	20.10; 08.19	53C10
prolipoprotein signal peptidase	1	14.07	26G10
Proopiomelanocortin	1	98	28C09
proteasome protein-related	1	70.13	54A10
proteasome subunit alpha type 2	1	70.13	30A07
proteasome subunit; Pre7p	3	70.13	43D12; 43H07; 47G07
protein disulphide isomerase	12	70.01	35H06; 39G07; 43E06; 44C07; 45H10; 16H03; 18F02; 19C05; 52H09; 52H11; 55D12; 58B07
protein kinase	1	42.03; 42.01	21D04
protein kinase C substrate 80k-h	2	99	27F02; 45B01
protein of fungal metazoan origin	1	99	17D10
Protein phosphatase 2A at 29B	1	42.03; 73.01	12A10
protein phosphatase ssd1 homologue	1	42.03; 11.01; 73.01	53E06
protein phosphotase 2a 65kd regulatory sububit	1	42.03; 04; 73.01	29G08
protein serine threonine kinase Clk4	1	98	28C05
protein snodprot1 precursor	21	32.05	27A07; 31E06; 36C11; 37H09; 40B09; 42F07; 42G02; 44D11; 45F04; 07H03; 12F12; 12G11; 14F10; 14G10; 23D11; 50F05; 50G05; 53B08; 53C06; 57A09; 57F09;
Protein snwA	1	10.01; 70.10	27E01
PrpD	1	99	58G12
Psi protein	5	01.03	32E05; 41D12; 45B11; 48E08; 52H03
P-type ATPase	1	20.03	36E04
P-type calcium ATPase 3	1	20.04; 34.01	20A05
purine-cytosine permease	1	01.03; 20.16; 08.19	30H08
purine-cytosine permease; Fcy2p	2	01.03; 20.16; 08.19	27B02; 30H08
Putative GabA permease	2	01.02; 20.10; 08.19	45F05; 18H02
Putative GTPase	1	14.07; 30.01	04C08
putative hydroxyproline-rich protein	1	99	30B09
Putative membrane protein of ancient origin	1	99	09D11
Putative senescence-associated protein	1	40.20	05C11
putative z-protein	1	99	39F01
PYM protein	1	99	29E10

pyruvate decarboxylase	26	01.05; 02.16	27E02; 28F11; 33F07; 35A12; 36C08; 38A07; 39C08; 39H12; 40B10; 42A10; 42C03; 42H07; 43F08; 45F08; 48H08; 15D10; 15E08; 16C07; 22E04; 25A10; 52H01; 54E03; 56E10; 59C05; 59E02; 60G06
Pyruvate kinase	4	01.05; 02.01	43A10; 01G09; 13A12; 55G08
QRI2 protein	1	99	11B07
Quinone oxidoreductase homolog	1	02.13; 42.16	35C05
Ran-Binding Protein Mog1p	1	70.02; 08.01	60E09
rap55	1	99	51C04
RAS-related protein RAB5	1	70.02; 08.13; 08.19	45D05
regucalcin (senescence marker protein-30)	1	70.1	28G08
Regulator of (H+)-ATPase in vacuolar membrane	1	02.13; 42.01	50F03
REGULATORY PROTEIN WETA	1	11.02	37B09
rehydrin protein homolog	2	98	30A10; 51G08
related to aimless RasGEF (aleA)	2	01.02; 01.02; 01.05; 42.03; 18.02; 34.01; 34.11	35A11; 12G06
related to SSD1 protein	1	42.03; 11.01; 73.01	34C08
replication factor-a protein	1	42.01; 34.11	24B04
replication protein B	1	42.01	26B01
retinal short-chain dehydrogenase/reductase	1	02.45	29C09
rho GDP dissociation inhibitor	1	18.02; 73.01	18C09
rho GTPase	2	73.01	26F07; 45E04
Ribonucleoside-diphosphate reductase large chain	1	01.03; 42.01	32A02
ribophorin II	1	01.05; 70.07	51G10
ribose 5-phosphate isomerase	1	01.05; 02.07	38C10
Ribosomal protein L13a	1	12.01	32F08
ribosomal protein L14	1	12.01	23G03
ribosomal protein L26	1	12.01	23F12
ribosomal protein L30	2	12.01	38C01; 13H09
ribosomal protein L35	3	12.01	42E05; 48B01; 52D11
ribosomal protein L4	1	12.01	33F02
ribosomal protein L9	4	12.01	29B10; 45E06; 50F02; 59G01
ribosomal protein MRP49	1	12.01	57B01
ribosomal protein P0 gene	1	12.01	29B07
ribosomal protein S18	7	12.01	28C07; 35H04; 40D12; 46F02; 08A03; 11H10; 24F03
ribosomal protein Srp1	1	12.01	33A04
Ribosomal protein YL6b (L5)	1	12.01	36B04
Ribosomal S29-like protein	2	12.01	31C06; 52E05
RIC1 protein	1	11.02; 11.02	46E03
RIKEN cDNA 2700094L05	1	99	26A07
ring finger protein 13	1	18.02	56G06
RNA annealing protein YRA1	2	20.16	17E09; 20H10
RNA polymerase sigma factor	1	11.02	27B08
RNA:pseudouridine (psi)-synthases	1	01.03	53G10
RNA-binding protein	1	11.02	30B05
RNA-binding region RNP-1 family member	1	11.04	57B03
RRM3/PIF1 helicase homolog	1	42.01	04G09
RuvB-like DNA helicase (Reptin)	1	42.01	39B02
S164	1	99	50D08
S-adenosylmethionine-dependent methyltransferase of the seven beta-strand family	1	01.06	24G08
SAS family Acetyltransferase	1	42.01; 70.07	44B06
SEC61, gamma subunit	7	70.02; 08; 67	43G06; 46C01; 19E10; 25H05; 51C01; 54B05; 37A04
Sed5-Vesicle Protein of 26 kDa	1	99	
sedoheptulose-1,7-bisphosphatase	1	01.05	42G12
Selenoprotein W-related protein	2	32.01	39A02; 16C12
sensor histidine kinase	1	14.07; 30.05	27A05

Septin B	1	42.03; 73.01	05C10
Serine hydroxymethyltransferase (Serine methylase)	4	01.03; 01.06	43D11; 18E04; 18F10; 59D08
serine palmitoyl CoA transferase subunit LCBA	1	01.06	37G09
serine protein kinase SRPK1	1	98	17G11
serine proteinase inhibitor IA-2	4	70.13; 11.01	28E09; 31A11; 35D06; 09G02
Serine threonine protein kinase SNF1	1	01.05; 18.01; 18.02; 11.01	02G05
Serine/threonine protein kinase	2	14.07; 30.01	32A08; 43D09
serine/threonine protein phosphatase	2	42.03; 73.01	49E02; 54H05
serine-arginine-rich splicing regulatory protein 86	1	11.02	33G05
serologically defined breast cancer antigen 84	1	99	53D12
seryl-tRNA synthetase	2	12.10	28C10; 56H10
severin kinase	1	42.03; 73.01	56G04
Shikimate 5-dehydrogenase	1	01.01	52G07
SHK1 kinase-binding protein	1	70.1	43C10
short-chain dehydrogenase/reductase (SDR) family protein	1	01.06	57G11
sid3 protein; gtpase; inducer of septum formation	1	42.03; 42.01	58G10
sigma-54 interacting transcription regulator protein	1	11.02	49C07
signal peptidase	1	70.07	30A12
signal recognition particle 72 kda protein	1	70.02	30G06
signal recognition particle subunit	1	70.02	46D08
Similar to 30 kDa protein	1	11.01	38F09
similar to trichohyalin	1	99	28C01
Similarity to human amsh protein	1	98	32D01
similarity to yeast zrg17	1	20.03	48G02
Ski8	1	42.01; 42.03; 11.10	41B05
S-layer protein	1	98	20G07
small GTPase RAS2	1	01.03; 01.05; 42.03; 42.01; 11.01; 34.11; 73.01	34G04
small GTP-binding protein YPTI	3	98	41B04; 19E08; 24G09
small nuclear ribonucleoprotein E	1	11.02; 11.04; 70.02; 08.01	54C06
small subunit of yeast RNA polymerase	1	11.02	36A08
small zinc finger protein tim13	1	70.02	24E11
SMY2	1	20.09; 42.04	25H06
Solanesyl pyrophosphate synthase	1	01.06; 01.07	07H01
Sorting NeXin; Snx4p	1	70.13	46E01
Spermidine synthase	5	01.20; 42.01	31B06; 32D12; 38C05; 39D12; 20F04
splicing factor	1	11.02	25A11
Splicing factor U2AF 50 kDa subunit	1	11.02	37A06
splicing factor, arginine/serine-rich 4	1	11.02	28E03
spore coat protein SP96 precursor	1	73.01	22E09
Spt3	1	11.02; 34.11	59F05
STE11	1	34.11	26A03
Ste50p	1	42.01; 11.01; 34.11; 36.20; 42.04; 73.01	38H05
Steroid binding protein	1	99	08D05
stomatin	1	42.01; 42.03; 42.16	60H02
stress protein p66	1	11.01	49D09
subtilisin proteinase-like protein	1	70.13	22F10
subunit of COP9 signalosome; Csn12p	1	34.11	18F06
subunit VI of cytochrome c oxidase; Cox6p	2	02.13	34B09; 49F12
Subunit VIb of cytochrome c oxidase; Cox12p	2	02.13; 70.10	44E09; 59B11
Succinyl-CoA synthetase, beta chain	1	01.05; 02.10	02G03
sugar aldolase	1	01.05; 02.01	53E10
sulphur metabolism negative regulator SconC	4	01.02	44G11; 53F02; 58A07; 59D10
SUMO (small ubiquitin-like modifier) SMO-1	1	70.13	10F11
superoxide dismutase	3	11.07	34F11; 01A08; 18D06

Suppressor of mutant AC40 subunit of RNA polymerase I and III (high serine); Srp40p	1	11.02	33B03
Suppressor of Sulfoxyde Ethionine resistance; Seo1p	1	20.03	56E04
suppressor of Ty 4 homolog 1	1	42.01	48H09
Surfeit 4	1	99	52B08
SURF-family protein	1	02.13	02E01
Svf1p; Involved in the diauxic switch	2	99	28E06; 13H01
swiss cheese protein	2	98	33E09; 50F10
Synaptobrevin	1	14.04; 20.09; 70.09	09G10
syntaxin 18	2	20.09; 70.07	33F06; 25E05
syntaxin 2	1	14.04; 20.09	35C11
syntaxin family member TLG1	1	70.02; 08.07; 08.13; 08.19; 42.07; 42.25	56H09
tandem pore domain potassium channel	1	20.01	26C03
tapetum-specific zinc finger protein 1	1	11.02; 16.17	55G07
TATA binding protein	1	11.02; 11.02; 11.02	41B10
t-complex protein 1, theta subunit	1	70.01	40E10
TEF4	2	12.04	33F12; 39B12
tetraacyldisaccharide 4-kinase	1	01.05	52A01
Tetracycline transporter-like protein	1	32.07	36B09
tetraspanin	2	99	54B07; 58F10
Thermoresistant gluconokinase	1	01.05	11C09
thiamine biosynthesis protein NMT-1	2	01.03	36A09; 38H12
Thiazole biosynthetic enzyme	2	01.07; 42.01	33E07; 48E09
ThiJ/PfpI family protein	4	99	07A12; 21H08; 51F06; 56C07
thioesterase	1	01.06	30D10
thiophene and furan oxidation protein	1	11.02; 12.04; 12.07	16H09
Thioredoxin-like protein	5	01.03; 70.07; 11.07	36D12; 12C12; 12G10; 18E07; 60F03
TIM23	1	70.02; 08.04	55B11
TPA regulated locus	1	98	27D06
TPA: carboxypeptidase; kex1	1	70.07	38E02
transaldolase	2	01.05; 02.07	48G06; 08C12
transcription elongation factor S-II	1	11.02	33D11
transcription factor	1	11.02	52E03
transcription factor Fst12	2	11.02	16D10; 54B08
transcription factor hunchback	1	11.02	28B07
transcription factor-like protein 4	1	11.02	28H04
transcription initiation factor IIA large subunit	1	11.02	56B06
transcription initiation factor IID beta chain	2	11.02	38D01; 51F04
transcriptional receptor-like protein	1	11.02	26F05
transcriptional regulator protein SPT6	1	42.01; 11.02	30F02
transcriptional regulator, AraC family	1	11.02	57H04
Transcriptional regulatory protein pro1	1	11.02	13D12
Transketolase	8	01.01; 01.05; 02.07	29B09; 39H09; 45G01; 03G02; 13F11; 14A09; 14H10; 49C05
translation elongation factor 3	1	12.04	52H12
Translation elongation factor EF-1gamma	3	12.04	26D12; 32C05; 56D02
translation elongation factor EF-Tu precursor	1	12.04	29E02
translation initiation factor	2	12.04	26C12: 28H06
Translation initiation factor 3 (eIF-3 gamma)	1	12.04	11A10
translation initiation factor eIF-3 beta subunit	2	12.04	50D04; 53D06
translation release factor erf3	1	42.03; 12.04	23F04
translationally controlled tumor protein homolog	1	99	43B06
translocon subunit	1	70.02; 20.11	35C08
transmembrane protein with a GGDEF domain	2	99	14H04; 14H05
transport protein SEC6	1	70.02	27C10
transport protein SEC61 beta subunit	1	70.02	55H04
Transposase	1	38.03	08D07
trichothecene C-15 hydroxylase	1	01.06	41B09
triosephosphate isomerase (TIM)	1	01.05; 02.01	15H10

tropomyosin TPM1	12	34.11; 73.01	29H05; 38G07; 45C02; 05C03; 06A12; 06D07; 08E11; 09B11; 09G04; 21E03; 49H02; 54C05
Tryptophan 2,3-dioxygenase	1	01.01; 01.07	58D03
Tryptophanase	1	01.01	11F06
Tubulin beta chain	2	42.03	46G12; 09H01
tubulin folding cofactor C	1	42.03	35E05
Tubulin gamma chain (Gamma tubulin)	1	42.03	44B12
Type 2C Protein Phosphatase	1	10.03	59H04
type I transmembrane protein	1	98	27A06
Type I transmembrane protein, component of CopII-coated er-derived transport vesicles; Emp24p	1	20.09; 70.09	11F10
Type IV secretory pathway, VirD2 components (relaxase)	1	98	31E08
tyr-inhibited DAHP synthase	2	01.01	45F03; 06G07
Ubiquinol-cytochrome c oxidoreductase subunit 7	1	02.13; 70.10	06C08
Ubiquinol-cytochrome c reductase complex subunit	8	02.13	31C07; 34H02; 46C10; 02F10; 49C10; 52E10; 57C08; 59G12
ubiquitin / ribosomal protein S27a	11	12.01	35C10; 35G03; 41E06; 44A03; 45G02; 02D06; 09E08; 20B10; 55B09; 56H12; 59E09
ubiquitin carboxyl terminal hydrolase	1	70.13	27F07
ubiquitin conjugating enzyme	5	70.07; 70.13; 08.13; 11.01; 34.11; 73.01	34A09; 51B08; 52F11; 55H07; 56G05
ubiquitin fusion degradation protein 2	2	70.13	41A06; 18G03
Ubiquitin fusion protein	10	70.13	32G10; 35G06; 37C10; 03A10; 10F03; 18E10; 20D06; 51H04; 55E04; 56D11
ubiquitin transferase	1	70.13	33E06
Ubiquitin-activating enzyme E1 1	1	11.02; 70.07; 70.13	47B10
ubiquitin-conjugating enzyme E2	1	70.07	51B02
ubiquitin-like modifier	1	70.07	47D04
ubiquitin-protein ligase	2	70.07; 70.13; 11.01; 73.01	56A01; 56B07
UDP-glucose 4-epimerase Gal10	1	01.05	32A06
UDP-glucose 6-dehydrogenase	3	01.05	34A10; 34H11; 48D04
udp-n-acetylglucomamine peptide n-acetylglucosaminyltransferase	2	01.05; 01.06; 42.01	29E11; 60G10
UDP-N-acetylglucosamine pyrophosphorylase	3	42.01	06D05; 11F11; 24C12
UDP-N-acteylglucosamine pyrophosphorylase 1	1	42.01	10D07
Undecaprenyl pyrophosphate synthetase	1	01.06	55G06
uracil permease	1	01.03; 20.16; 08.19	19G04
Urease	1	01.02	07C11
uricase	2	01.02	33G12; 10G10
uridine kinase-like 1	1	01.03	60H08
uridine nucleosidase (uridine ribohydrolase)	1	01.03	53F01
uridine permease; Fui1p	1	01.03	50F11
uv-induced protein uvi31	2	32.01	47D01; 07A08
Vacuolar ATP synthase subunit d (V-ATPase d)	1	20.04; 34.01	08H01
vacuolar ATP synthase subunit F	1	20.04; 08.13; 34.01; 42.01	41C03
vacuolar ATPase V1 domain subunit D	1	20.04; 08.13; 34.01	36H11
Vacuolar-ATPase	12	20.04; 08.13; 11.07; 34.01	32B09; 32F09; 01A03; 01F11; 01G02; 03B09; 05C02; 05E11; 07E08; 10C06; 12G01; 60H06
Vacuole import and degradation; Vid27p	2	99	13A04; 13F09
vesicular transport and membrane fusion protein; sec18	1	20.09	59A01
vesicular-fusion protein sec17 homolog	1	01.04; 1.02; 16.19; 20.09	46D12
VHL binding protein-1	1	10.03; 14.01; 16.01; 42.04	25F11
vicilin-like embryo storage protein	1	98	55F04
vivid PAS protein VVD	2	16.01; 18.02; 34.11	28E08; 10F06

VpsB	1	70.02; 08.07; 08.13; 42.25; 73.01	60B03
V-type ATPase; ATP synthase j chain	1	20.04; 08.13; 34.01; 42.01	50A04
White collar 1 protein (WC1)	1	99	37G02
Woronin body major protein	3	42.19	43G07; 46C06; 55C05
Xanthine dehydrogenase	1	01.03; 01.05	09A03
XAP-5	1	38.03	04F09
XG glycoprotein precursor	1	98	26C02
Xylitol dehydrogenase	2	01.05	02C03; 24D06
YCL012Cp	1	98	42D09
Yippee-like protein 3	1	99	43E07
YRO2 protein	14	11.01	26B10; 33B07; 34D04; 36G11; 46A12; 02B08; 19A06; 19C06; 22B08; 25D07; 54B04; 55B05; 55F10; 59B12
zinc finger protein 353	2	10.03; 11.02; 70.10	27D04; 31C09