(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 November 2010 (25.11.2010)

PCT

(10) International Publication Number
WO 2010/135678 A1

(51) **International Patent Classification**:
C12N 1/19 (2006.01)

(21) **International Application Number**:
PCT/US2010/035825

(22) **International Filing Date**:
21 May 2010 (21.05.2010)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
61/180,502 22 May 2009 (22.05.2009) US

(71) **Applicants** *(for all designated States except US)*: **RESEARCH CORPORATION TECHNOLOGIES, INC.** [US/US]; 5210 East Williams Circle, Suite 240, Tucson, Arizona 85711-4410 (US). **VIB,VZW** [BE/BE]; Rijvisschestraat 120, B-9052 Zwijnaarde (BE). **UNIVERSITEIT GENT** [BE/BE]; Sint-Petersnieuwstraat 25, B-9000 Gent (BE).

(72) **Inventors; and**

(75) **Inventors/Applicants** *(for US only)*: **CALLEWAERT, Nico L. M.** [BE/BE]; Begijnhoflaan 15, B-3850 Nevele-Hansbeke (BE). **DE SCHUTTER, Kristof** [BE/BE]; Kantonstraat 30, B-9320 Nieuwerkerken (BE). **TIELS, Petra** [BE/BE]; J. De Saint-Genoisstraat 54, B-9050 Gentbrugge (BE). **LIN, Yao-Cheng** [—/BE]; Brusselsesteenweg 244, B-9050 Ledeberg (BE).

(74) **Agent**: **GROLZ, Edward W.**; Scully, Scott, Murphy & Presser, 400 Garden City Plaza, Suite 300, Garden City, New York 11530 (US).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

WO 2010/135678 A1

(54) **Title**: NUCLEIC ACIDS OF *PICHIA PASTORIS* AND USE THEREOF FOR RECOMBINANT PRODUCTION OF PROTEINS

(57) **Abstract**: The present invention provides the genome sequence of *Pichia pastoris* and manually curated annotation of protein-coding genes. The invention provides novel nucleic acids, proteins, and related expression vectors useful for genetic engineering of methylo trophic yeast strains, as well as engineered methylotrophic yeast strains particularly *Pichia pastoris*, and use thereof for recombinant production of heterologous proteins including glycoproteins suitable for use in mammals including humans.

Nucleic Acids Of *Pichia pastoris* And Use Thereof For Recombinant Production Of Proteins

CROSS REFERENCE TO RELATED APPLICATION

[001] This application claims the benefit of priority from U.S. Provisional Application 61/180,502, filed May 22, 2009.

FIELD OF THE INVENTION

[002] This invention relates generally to novel nucleic acids and recombinant expression technology. In particular, the invention relates to determination and assembly of the genome sequence of *Pichia pastoris*. The invention provides novel nucleic acids, proteins, and related expression vectors useful for genetic engineering of methylotrophic yeast strains, as well as engineered methylotrophic yeast strains particularly *Pichia pastoris* and use thereof for recombinant production of heterologous proteins.

BACKGROUND OF THE INVENTION

[003] The methylotrophic yeast *Pichia pastoris* is by far the most often used yeast species in the production of recombinant proteins and is used in thousands of laboratories worldwide for the production of proteins for fundamental studies, as drug targets and for therapeutic use, and as a model for peroxisomal proliferation and methanol assimilation. The *P. pastoris* expression technology is available from Invitrogen (Carlsbad, CA) and from Research Corporation Technologies (Tucson, AZ), making it accessible for academic and commercial purposes alike. *P. pastoris* grows to high cell density, provides tightly controlled methanol-inducible transgene expression and efficiently secretes heterologous proteins in defined media. Indeed, several *P. pastoris*-produced biopharmaceuticals that are either not glycosylated (such as human serum albumin) or for which glycosylation is only needed for proper folding (such as several vaccines) are already on the market. *P. pastoris* strains with small, homogenous N-glycans have also been generated, which were then further engineered

into human-type N-glycosylation[1,2]. Glyco-engineered products are now moving to clinical development[3]. Moreover, monoclonal antibodies can be made at gram per liter scale in glycosylation-homogeneous strains[4]. For further strain engineering, a better understanding of all aspects of the yeast's protein production machinery is desired, and a number of studies relating to *Pichia*'s secretory system and engineered promoters have been forthcoming[5,6]. However, although *P. pastoris* is widely used for protein production, relatively few genetic tools, engineered strains and data on the biology of this organism are available.

SUMMARY OF THE INVENTION

[004] The present invention provides the determination, assembly and manually curated annotation of the 9.43 Mbp genomic sequence of the GS115 strain of *Pichia pastoris*.

[005] In one embodiment, the invention provides isolated nucleic acid molecules that encode a protein as set forth in any one of SEQ0001-0025, 0027-0126, 0128-0165, 0172-0174, 0176-0200, or 0202-0212, or a protein substantially homologous thereto. In specific embodiments, the nucleic acid molecules contain the coding sequence of an ORF as set forth in any one of SEQ0001-0025, 0027-0126, 0128-0165, 0172-0174, 0176-0200, or 0202-0212, or contain a nucleotide sequence that is substantially homologous to the coding sequence of an ORF as set forth in any one of SEQ0001-0025, 0027-0126, 0128-0165, 0172-0174, 0176-0200, or 0202-0212.

[006] In another embodiment, the present invention provides isolated proteins that have an amino acid sequence as set forth in any one of SEQ0001-0025, 0027-0126, 0128-0165, 0172-0174, 0176-0200, or 0202-0209, or an amino acid sequence substantially homologous thereto.

[007] In still another embodiment, the present invention provides 53 peptides, shown in SEQ0001-0025 and 0027-0054, which are signal peptides of secreted proteins of *Pichia pastoris*. Nucleic acids that encode any of these signal peptides are also parts of the present invention.

[008] In a further embodiment, the present invention provides a set of vectors useful for identification of the most effective choice of signal peptide for any given heterologous protein. Each vector contains a promoter and the coding sequence of one of the signal peptides identified in SEQ0001-0025 and 0027-0054, and a linker sequence or cloning site for inserting or receiving the coding sequence of the heterologous protein. In a specific embodiment, the linker sequence or cloning site simply includes a restriction endonuclease site. The heterologous coding sequence, with the same restriction site at its 5' end, can be joined to the signal peptide coding sequence via restriction enzyme digestion and subsequent ligation. In an alternative embodiment, the linker sequence is an intron sequence functional in *Pichia pastoris* and includes a restriction endonuclease site. The heterologous coding sequence, with the same intron sequence and restriction site at its 5' end, can be joined to the signal peptide coding sequence via the intron sequence. Transcription from the expression vector and subsequent RNA processing in a recipient cell lead to the generation of mRNAs without the intron, from which signal peptide-heterologous protein fusions are produced upon translation. A library of such fusion clones can be transformed into *Pichia pastoris* to select for the most effective choice of signal peptide for any given heterologous protein.

[009] In one embodiment, the invention provides isolated nucleic acid molecules composed of a promoter sequence of any one of the *Pichia pastoris* genes set forth in SEQ0055-0165, 0169-0200, or 0202-0212.

[010] Specific promoters of the present invention include those identified for SEQ0060-0085 (genes involved in glycolysis pathway), SEQ0096-0124 (genes showing high expression levels), SEQ0125-0128 (homologs of *S. cerevisiae* genes whose promoters are frequently used for recombinant expression), SEQ0169-0178 (methanol metabolism genes), and SEQ0202-0208 (genes involved in xylose, arabinose or threhalose metabolism). The promoters of these genes are located within the 1000 bp of the 5' region provided herein, and generally are located within the 500 bp immediately before the start codon of the gene, in some embodiments within 250 bp, 200 bp, 150 bp, 125 bp, 100 bp, 75 bp, 50 bp, 40 bp, or even 25 bp immediately before the start codon of the gene. In certain embodiments, the promoters include a TATA element identified herein in **Table 6.**

[011] Use of any of the newly identified promoters for expression of a heterologous gene is encompassed by the present invention, including the related expression vectors and cells transformed with any such expression vectors.

[012] In one embodiment, the present invention is directed to expression vectors and engineered methylotrophic yeast strains for increased expression (or overexpression) of one or more *Pichia* proteins involved in the secretory pathway, e.g., those as set forth in SEQ0055-0059, in order to achieve increased protein secretion.

[013] In another embodiment, the present invention is directed to expression vectors and engineered methylotrophic yeast strains for overexpression of *Pichia* glycosylation precursor synthesis enzymes or transporters, e.g. UDP-Gal or UDP-GlcNAc transporters and UDP-Glc-4-epimerase for ER or Golgi localization. Such expression vectors are constructed to contain, from 5' to 3', a promoter functional in the recipient strain, operably linked to a coding sequence as set forth in any one of SEQ0129-0132, and a transcription termination sequence. The encoded protein is preferably also designed to include an ER or Golgi localization signal. Such an expression vector can be introduced into a methylotrophic yeast strain by transformation. Thus, the resulting engineered strains capable of increased expression of a protein encoded by any one of SEQ0129-0132 constitute another embodiment of the invention.

[014] In a further embodiment, the present invention provides a methylotrophic yeast strain in which at least one (i.e., one or more) native gene encoding an O-mannosyl transferase or a beta- mannosyl transferase has been inactivated. In a specific embodiment, the strain is a *Pichia* strain, preferably, a *P. pastoris* strain. An O- or beta-mannosyl transferase knockout *P. pastoris* strain can be generated by inactivating at least one gene as set forth in SEQ0133-0137 or SEQ0210-0212, which can reduce or eliminate unwanted O- or beta-glycosylation of a heterologous protein.

[015] In one embodiment, the present invention provides a methylotrophic yeast strain in which one or more native genes encoding enzymes involved in the ER glycosylation pathway have been inactivated. In a specific embodiment, the native yeast STT3 gene has been

inactivated, and optionally the *Leishmania* STT3 gene has been introduced in place thereof. In another specific embodiment, the methylotrophic yeast strain is *P. pastoris* in which the native *Pichia* gene as set forth in SEQ0163 has been inactivated, and the *Leishmania* STT3 gene has been introduced in place thereof.

[016] In another embodiment, the present invention is directed to methylotrophic yeast strains which overexpress one or more of the MNN4 homologs as set forth in SEQ0166-0168, and related expression vectors.

[017] In still another embodiment, the present invention provides a protease-deficient methylotrophic yeast strain. In a specific embodiment, the strain is a *Pichia* strain, e.g., a *P. pastoris* strain. The protease-deficient *Pichia* strain can be generated by inactivating at least one (i.e., one or more) protease-encoding genes as set forth in SEQ0179 and SEQ0181-0186, such that there is no functional protease produced from a disrupted gene. Protease deficient strains allow more stable accumulation of heterologous proteins.

[018] In yet another embodiment, the present invention provides a methylotrophic yeast strain engineered to overexpress a protease inhibitor protein, encoded by the gene set forth in SEQ0180, at an elevated level compared to an unmodified strain. In a specific embodiment, the strain is a *Pichia* strain, e.g., a *P. pastoris* strain. Expression vectors created for making such strains form another embodiment of the invention.

[019] In one embodiment, the present invention is directed to expression vectors and engineered methylotrophic yeast strains for overexpression of at least one *Pichia* chaperones involved in secreted protein folding in the ER. Such expression vectors are constructed to contain, from 5' to 3', a promoter functional in the recipient strain, operably linked to a coding sequence as set forth in any one of SEQ0187-SEQ0200, and a transcription termination sequence. Such an expression vector can be introduced into a methylotrophic yeast strain by transformation. Another embodiment of the invention is directed to engineered methylotrophic yeast strains capable of overexpressing a protein encoded by any one of SEQ0187-SEQ0200. In still another embodiment, methylotrophic yeast strains capable of

overexpressing a combination of multiple chaperones are provided, which combination can be selected as most effective for recombination production of a particular heterologous protein.

[020] In a further embodiment, the present invention provides an isolated nucleic acid molecule containing the nucleotide sequence as set forth in SEQ0200, which encodes the 5S rRNA. Use of this nucleic acid in creating vectors to achieve multi-copy integration of a heterologous gene and generate strains having a heterologous gene stably integrated in the genome in multiple copies is also contemplated by the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[021] **Figures 1A-1C.** *Pichia pastoris* genome sequencing and overview. **1A.** Genome sequencing and assembly strategy. **1B.** *P. pastoris* chromosomes and known markers position. Genes that had been previously mapped to the chromosomes through PFGE are indicated in blue, and rDNA repeats in yellow, the 5S rRNA are indicated in yellow with the red arrow. **1C.** Phylogenetic tree. The phylogenetic tree was built on the concatenated sequence of 200 single-copy ortholog genes in all of the 6 species. Numbers next to each branch correspond to the number of Pfam domains uniquely present in the corresponding lineage.

[022] **Figures 2A-2B.** *Pichia pastoris* codon usage. **2A**. Codon usage in the *P. pastoris* ORFeome. The relative abundance of a codon is represented as a percentage of the total codon usage for the amino acid. **2B.** Correlation of tRNA genes and codon usage. Graph shows correlation between the codon usage in relation to the number of genes coding for tRNAs recognising this codon (Spearman rho=0.88, $P<0.0001$).

[023] **Figures 3A-3B.** *Pichia pastoris* pathways. **3A.** Methanol utilisation pathway in *Pichia pastoris*. A detailed table with the genes coding for the respective enzymes is shown in **Table 5A**. Abbreviations: [1]AOX: alcohol oxidase, [2]FLD: formaldehyde dehydrogenase, [3]FGH: S-formylglutathione hydrolase, [4]FDH: formate dehydrogenase, [5]CAT: catalase, [6]DAS: dihydroxyacetone synthase, [7]DAK: dihydroxyacetone kinase, [8]TPI: triosephosphate isomerase, [9]FBA: fructose-1,6-bisphosphate aldolase, [10]FBP: fructose-1,6-bisphosphatase;

DHA: dihydroxyacetone, GAP: glyceraldehyde-3-phosphate, DHAP: dihydroxyacetone phosphate, $F_{1,6}BP$: fructose-1,6-bisphosphate, F_6P: fructose-6-phosphate, P_i: phosphate, Xu_5P: xylulose-5-phosphate, GSH: glutathione. **3B.** Protein secretion pathway. Schematic representation of the secretion pathway in *P. pastoris*. A detailed table with the genes coding for the components involved in the represented complexes or processes is shown in **Table 5B**. The nascent protein is translocated to the ER by the Sec61 complex and N-glycosylation sites are glycosylated with the dolichol-linked $Glc_3Man_9GlcNAc_2$ oligosaccharide precursor by the OST complex. After processing of the signal peptide, the protein is folded by the aid of chaperones. ER N-glycan processing results in $Man_8GlcNAc_2$ type glycan. O-glycosylation is also initiated in the ER by the protein-O-mannosyltransferases. After transport to the Golgi apparatus, the N-glycans are further processed to the yeast-typical hypermannosyl-type glycans. In GlycoSwitch-engineered strains, the hypermannosylation is abolished and the glycans are processed to $Gal_2GlcNAc_2Man_3GlcNAc_2$. After processing of the pro-domain, the protein is secreted in the growth medium, where it may be a substrate for yeast proteases.

[024] Figure 4. Contig overlap processing. Conflicts at contig overlaps, due to low coverage of the sequence represented by "XXX", are resolved by trimming the extreme ends of the contig. Contig coverage is represented by heatmap. Low fidelity overlap regions (P-value $>e^{-50}$) were PCR amplified and Sanger sequenced.

[025] Figures 5A-5C. Chromosome assembly. **5A.** By PFGE and Southern blot detection, 2 supercontigs (FragB and FragD), 4 contigs (c121, c34, c131, c157, c159) and the contig containing the rDNA repeats (c2) were located on the different chromosomes. Every lane of the blot was incubated with a probe on an open reading frame of the indicated genome fragment. A probe on HIS4, GAP, URA3 and AOX1 was chosen to detect chromosome 1, 2, 3 and 4, respectively. The *H. wingeii* chromosomes were used as marker for the PFGE, but they also gave a signal on the blot with the conserved c2 probe. The rightmost 2 lanes derived from a different gel than the rest of the figure, and chromosomes 2 and 3 were not well resolved on this gel. Presence of an rDNA locus (corresponding to contig c2) on both of these chromosomes was ascertained through PCR. **5B.** Result of the PCRs performed to join the supercontigs and contigs. Lanes 1-8 are PCR with primers 1&4, 2&5, 6&25, 12&26, 8&9,

23&62, 15&60 and 17&21, respectively. **5C.** Representation of the chromosomes assembled by the supercontigs and contigs. The numbers in blue represent PCR primers that were chosen on each end of the supercontigs and contigs (~200 bp from the end). The size of the gap is depicted between each supercontig and contig.

[026] Figure 6. Distribution of gene ontology terms assigned to *P. pastoris*. A total of 4,262 *P. pastoris* genes were assigned with gene ontology (GO) terms: 3,142 genes with molecular function assignment, 3,647 genes with cellular component assignment and 3,182 genes with biological process assignment.

[027] Figure 7. *P. pastoris* secretion signals. 53 SignalP predicted signal peptides were manually curated to be secretion signals based on the function of orthologs. The predicted site of signal peptidase cleavage is indicated by the red triangle. Alignment of these peptides shows a hydrophobic consensus sequence (poly Leu), and a small amino acid residue at position -1 and -3 from the cleavage site.

[028] Figure 8. Protease Gene Insertional Inactivation Strategy. A drug resistance marker or other auxotrophic marker can be used for this method.

DETAILED DESCRIPTION OF THE INVENTION

[029] The present inventors have determined and assembled the 9.43 Mbp genomic sequence of the GS115 strain of *P. pastoris*, and manually curated annotation of 5,313 protein-coding genes. On this basis, the invention provides novel protein-encoding genes from *Pichia pastoris*, including identification of the 5' upstream region (including promoter), open reading frame (ORF) and 3' downstream region of these genes. The present invention also provides novel *Pichia pastoris* proteins, and certain signal peptides of some of these *P. pastoris* proteins. The nucleic acids and encoded proteins identified herein, as well as the promoters and signal peptides, can be used in engineering methylotrophic yeast strains, particularly *Pichia* strains, for recombinant production of proteins, including but not limited to glycoproteins having glycoforms suitable for therapeutic use in mammals especially humans. In addition to the values and utilities provided by each of the novel molecules

individually, the determination and annotation of the genome sequence of *Pichia pastoris* also permit a more complete, overall understanding of *Pichia pastoris* in respect to its protein modification and secretion system as well as methanol metabolism, by providing a complete set of *Pichia pastoris* genes coding for enzymes involved in methanol assimilation, a complete catalog of *Pichia pastoris* orthologs to the *S. cerevisiae* endoplasmic reticulum (ER) folding machinery, and a full collection of genes involved in the ER glycosylation pathway of *Pichia pastoris*. These findings of the present invention enable more efficacious overall design and engineering for recombinant production of heterologous proteins in methylotrophic yeasts.

[030] The various aspects and embodiments of the present invention are described in details below.

General Definitions:

[031] Open Reading Frame (ORF) – An ORF refers to the portion of a gene that begins with the start codon and ends with a stop codon and that encodes a protein. An ORF may include one or more intron sequences.

[032] Coding sequence – This term is used herein to refer to a contiguous sequence of codons of the protein encoded by the ORF and does not include intron.

[033] 5' upstream region – This term (or "5' region" or "upstream region" in abbreviation) is used herein to refer to the genomic region 5' relative to the ORF of a gene. Generally speaking, the Sequence file provided herein has set forth a 5' upstream region of approximately 1000 nucleotides, which, in some cases, includes the start and/or stop codons of the previous gene. The 5' upstream region of a gene includes the promoter of the gene. The extent of the 5' region provided herein for each gene is sufficient for targeting recombination to this site, e.g. to delete the entire ORF when used in combination with the 3' downstream region sequence, or to insert a different regulatory region or promoter immediately upstream of the ORF when used in combination with the ORF sequence itself.

[034] Promoter – This term refers to a portion of the 5' upstream region of a gene that directs the transcription of the gene. Promoters are located within the 1000 bp of the 5' upstream region of yeast genes, with a "TATA box" sequence most commonly located at 10-120 bp upstream from the start codon of a gene. A "TATA box" is a DNA sequence (cis-regulatory element) found in the promoter region, and has the core DNA sequence 5'-TATAAA-3' or a variant. A TATA box is usually located 25 base pairs upstream to the transcription site. In about 76% percent of the *Pichia pastoris* genes, at least one of the following TATA-elements has been located – TATAA: 50% of the elements are found 60 to 90 bp upstream of ATG (75% 40 to 110 bp upstream of ATG); TACAA: 78% of the elements are found 50 to 70bp upstream of ATG; TATA: 50% of the elements are found 10 to 40 bp upstream of ATG; TATATA: 50% of the elements are found 80 to 90bp upstream of ATG; and TATATATA: 50% of the elements are found 50 to 60bp of ATG. Details of the TATA boxes of the promoters provided by the present invention are set forth in **Table 6**. Accordingly, in certain specific embodiments, the promoter of a *Pichia pastoris* gene provided herein is located within the 500 bp immediately before the start codon of the gene; prepferably within 250 bp, 200 bp, 150 bp, 125 bp, 100 bp, 75 bp, 50 bp, 40 bp, or even 25 bp immediately before the start codon of the gene; and in particular embodiments, the promoter includes a TATA element identified herein in **Table 6**. The precise location and composition of a promoter can be determined by using well known techniques including deletion mapping and site-directed mutagenesis, as further described below.

[035] 3' downstream region - This term (sometimes "downstream region" or "3' region" in abbreviation) is used herein to refer to the genomic region 3' relative to the ORF of a gene. Generally speaking, the Sequence file provided herein has set forth a 3' downstream region of approximately 1000 nucleotides, which, in some cases, includes the start and/or stop codons of the next gene. The 3' down region of a gene includes the transcription termination sequence (or 3' termination sequence) of the gene. The extent of the 3' region provided herein is sufficient to target recombination to this site in the chromosome.

[036] Selectable marker – This term refers either to a dominant drug resistance marker or similar dominant marker, or to a more limited prototrphic selection such as HIS, applicable to a host that is defective for the key enzyme supplied by a selectable marker gene.

[037] "Signal peptide" and "mature proteins" – The term "signal peptide" or "signal sequence" refers to the short peptide sequence within a protein precursor synthesized in the cytoplasm that targets the precursor form to the endoplasmic reticulum. Signal peptides are typically cleaved from the precursor form by signal peptidase after the proteins are transported to the ER, and the resulting proteins move along the secretory pathway to their intracellular or extracellular location. For some proteins, cleavage of the signal peptide results in the mature form (i.e., the final, biologically active form) of the protein, while for other proteins, additional proteolytic processing may be required in order to generate the mature form of the protein.

[038] Substantially homologous amino acid sequences – When two or more amino acid sequences are said to be substantially homologous, it is meant that the sequences share a significant degree of similarity, for example, at least 85%, 90%, 95%, 98% or even 99% similarity. The degree of similarity can be determined, for example, as the index calculated using the BLAST or the Lipman-Pearson Protein Alignment program with the following choice of parameters for the latter program: Ktuple = 2, Gap Penalty = 4, and Gap Length Penalty = 12. The term "similarity" includes identity. Substantially homologous proteins can perform or possess substantially the same function; i.e., the enzymatic activities of the proteins differ by not more than 20%, 15%, 10%, or even 5% under a same set of conditions applicable for measuring enzymatic activity. A protein that is substantially homologous to a *Pichia pastoris* protein identified herein is, in some embodiments, a protein of methylotrophic yeast; for example, a protein of *Pichia.*

[039] Substantially homologous nucleotide sequences – When two or more nucleotide sequences are said to be substantially homologous, it is meant that the sequences share a significant degree of identity, for example, at least 85%, 90%, 95%, 98% or even 99% identity. The degree of homology is also reflected by hybridization characteristics. As

defined herein, a first nucleic acid sequence that is substantially homologous to a second nucleic acid sequence molecule also hybridizes to the complement of the second nucleic acid sequence under high stringency conditions. "High stringency conditions", as defined herein, include, for example, hybridization at 42°C in 50% v/v formamide, 1M NaCl, and 1% w/v SDS, and washing at 65°C in 0.1-2X SSC (e.g., 0.1, 0.2, 0.5, 1 or 2X SCC) and 1% w/v SDS. A nucleic acid that is substantially homologous to a *Pichia pastoris* nucleic acid identified herein is, in some embodiments, a nucleic acid of methylotrophic yeast; for example, a nucleic acid of *Pichia.*

[040] "Heterologous" versus "native"/"endogenous" – The term "heterologous" is used herein in several different contexts to reflect the fact that a molecule is placed in a genetic, molecular or cellular environment that is different than its native environment. For example, when the promoter of a gene is being utilized to drive the expression of a different gene, the promoter will be taken out of its native genetic context and placed in an operable linkage to a heterologous gene. As another example, the signal peptide of a protein can be used to direct the localization of a different protein, i.e., a heterologous protein. Additionally, a methylotrophic yeast strain such as *Pichia* can be transformed with a heterologous nucleic acid, i.e., a nucleic acid which the non-engineered *Pichia strain* does not have. The resulting engineered strain will express the protein encoded by the heterologous nucleic acid, i.e., a heterologous protein.

[041] Gene overexpression – Overexpression of a gene in a methylotrophic yeast is achieved by genetic modification of the yeast such that the expression of the gene is increased (measurable either at the mRNA level or the protein level), as compared to the unmodified yeast. The extent of increase in expression is at least 35%, or at least 50%, or preferably at least 150%, 200%, 250%, 300%, 400% or more. Overexpression of a gene can be achieved by introducing additional expression cassette carrying the gene (as a plasmid vector or integrated into the chromosome), or by replacing the native promoter of the gene with a stronger constitutive or inducible promoter.

[042] Gene inactivation – Inactivation of a gene in a methylotrophic yeast is achieved by genentic modification of the yeast such that substantially no functional protein is produced from the strain. By "substantially", it is meant that the level of functional protein produced from the modified strain is not more than 20%, or 15%, or 10% or even 5% or less of the level of functional protein produced from an unmodified strain. Inactivation of a gene can be achieved by disrupting the genomic ORF of the gene in the strain, or disrupting the native promoter, or replacing the native promoter with a repressible promoter (e.g., repressible by methanol).

[043] Methylotrophic yeast – Methylotrophic yeasts are those capable of growth on methanol, and include yeasts of the genera *Candida, Hansenula* (such as *H. polymorpha*, now classified as *Pichia angusta*), *Torulopsis,* and *Pichia* (e.g., *Pichia pastoris, Pichia methanolica*, *Pichia angusta* (formerly *Hansenula polymorpha*), *Pichia stipitis*, and *Pichia anomala*).

[044] Chaperones - Chaperones are proteins that assist in the non-covalent folding or unfolding of proteins, mediate the redox potential to assist the formation of disulphide bonds within or between protein subunits, assist in the assembly or disassembly of other macromolecular structures or complexes, and/or translocation of proteins across membranes.

[045] For convenience of discussion, the novel protein-encoding genes identified by the present invention have been grouped based on the functions of the encode proteins and are discussed in details below. Where the utility of a particular group or a particular gene is not specifically discussed, its utilities are apparent based on the function and utility of its homolog(s) from other yeast species such as *S. cerevisiae* which have been characterized.

Secreted proteins with a signal peptide

[046] In accordance with the present invention, 53 genes have been identified as encoding secreted proteins with a signal peptide. In the annotated Sequence file provided herein, the ORF and the amino acid sequence (with the signal peptide portion shown in bold) of each of the 53 genes are set forth in SEQ0001-0025 and 0027-0054.

[047] In one embodiment, the present invention is directed to isolated nucleic acid molecules comprising a nucleotide sequence that encodes a protein (e.g., a full-length protein) as set forth in any of SEQ0001-0025 or 0027-0054, or encodes a protein that is substantially homologous to a full-length protein as set forth in any of SEQ0001-0025 or 0027-0054.

[048] In a specific embodiment, the present invention is directed to isolated nucleic acid molecules comprising the coding sequence of an ORF as set forth in any of SEQ0001-0025 or 0027-0054, or a nucleotide sequence that is substantially homologous to the coding sequence of an ORF as set forth in any of SEQ0001-0025 or 0027-0054.

[049] In another embodiment, the invention is drawn to an isolated protein comprising an amino acid sequence (e.g., a full-length protein sequence) as set forth in any of SEQ0001-0025 or 0027-0054, or an amino acid sequence substantially homologous thereto.

[050] In still another embodiment, the invention is directed to the signal peptides of these 53 newly identified genes and uses thereof for recombinant expression of heterologous proteins. These signal peptides are summarized and aligned in **Figure 7**. Alignment of these peptides shows a hydrophobic consensus sequence (poly Leu), and a small amino acid residue (such as A, C, G and S) at position -1 and -3 from the cleavage site.

[051] Proteins destined for translocation into the endoplasmic reticulum carry a signal peptide that is recognized upon ribosomal translation by the signal recognition particle, which docks to the translocon, upon which translation continues and the protein is delivered into the ER lumen. The signal peptide is removed by signal peptidase.

[052] When over-expressing a protein destined for secretion, one can either use the protein's native signal peptide if the translocation and signal peptidase machinery of the host cell can efficiently recognize and then process it. Alternatively, the coding sequence of a heterologous signal peptide can be fused to the coding sequence for the mature form of the protein. In *Pichia pastoris*, the signal sequence which has been frequently used is the prepro signal of *S. cerevisiae* alpha mating factor. Whereas this signal works in many cases, processing of the propeptide can be problematic as it requires Kex2p protease cleavage followed by polishing

of the newly created N-terminus by Ste13p diaminopeptidase. Therefore, having a large library of *Pichia* signal peptides available is particularly useful. For example, one can screen for effective signal peptide(s) for a given protein desired to be expressed in *Pichia*. In addition, based on the library of signal peptides provided herein, a consensus artificial signal peptide can be designed which can be used to efficiently secrete multiple different heterologous proteins.

[053] To screen the library for a suitable *Pichia* signal peptide for a given target protein desired to be expressed in *Pichia*, an intron functional in *Pichia* (which can be selected from those shown in the Sequence File provided herein) can be cloned before the ORF or coding sequence for the target protein, wherein the intron contains either a unique restriction site or a recognition site for a recombinase. The same intron can be cloned behind (i.e., 3' of) each signal peptide coding sequence in the library, enabling rapid generation of an expression library of signal peptide-target ORF fusions through classical cloning using the unique restriction site or through recombinational cloning. Upon transcription, the intron is removed by the *Pichia* splicing machinery, resulting in an expressed library of in-frame fusion between the coding sequence for one of the signal peptides and the target ORF.

[054] Subsequently, secretion of the target protein by individual members of the expressed library can be evaluated by a suitable technique such as SDS-PAGE analysis, coomassie blue staining, or Western blot. In this way, the suitability of a given *Pichia* signal peptide can be evaluated for secretion of any target protein.

[055] Accordingly, in another embodiment, the present invention is directed to a library of expression vectors, wherein each of the 54 signal peptides is represented in the library. Each expression vector contains, from 5' to 3', a promoter functional in *Pichia*, operably linked to a coding sequence for a signal peptide, and an intron sequence containing a restriction endonuclease recognition site. The expression vector is designed to accommodate insertion of the coding sequence of a target protein, linked in its 5' to the same intron sequence containing the same restriction endonuclease recognition site on the expression vector.

[056] In a further embodiment, the present invention provides an expression vector capable of directing the expression and secretion of a heterologous protein in *Pichia pastoris* or another methylotrophic yeast. The expression vector contains, from 5' to 3', a promoter functional in the recipient strain, operably linked to a coding sequence for the fusion of a signal peptide (identified in SEQ0001-0025 and 0027-0054) and the heterologous protein. Host cells transformed with such an expression vector constitute another embodiment of the present invention.

[057] In another embodiment, the signal sequences from some of the secreted proteins are not cleaved off in the ER, but remain linked to the protein. This is usual for a number of those enzymes involved in the glycosylation pathway (for example those for mannosyl transferases), for secretion chaperones, and for other factors that help secretion in a particular compartment, for example. These signal sequences are useful for the localization of heterologous proteins to target them to particular compartments of the secretory pathway, which may be advantageous to various aspects of posttranslational modification. This may correspond to the same or equivalent compartment in which a heterologous enzyme or protein normally acts in its native cell. Alternatively, it may be a different compartment, but, is effective because it is the same as that in which another enzyme acts or is immediately downstream or upstream of the compartment in which a second enzyme acts, effectively channeling and coordinating the sequence of a metabolic process.

Proteins potentially involved in secretion

[058] In accordance with the present invention, 5 genes have been identified as encoding proteins potentially involved in secretion, including *P. pastoris* homologs of *S. cerevisiae* SEC1, SEC 11, and subunits SPC1, SPC2 and SPC3 of signal peptidase complex. In the annotated Sequence file provided herein, the nucleotide sequences of the 5' upstream region (including promoter), the ORF, and the 3' downstream region, as well as the amino acid sequence, of each of these 5 genes are set forth in SEQ0055 to SEQ0059.

[059] In one embodiment, the present invention is directed to isolated nucleic acid molecules comprising a nucleotide sequence that encodes a protein (e.g., a full-length protein)

as set forth in any of SEQ0055 to SEQ0059, or encodes a protein that is substantially homologous to a full-length protein as set forth in any of SEQ0055 to SEQ0059.

[060] In a specific embodiment, the present invention is directed to isolated nucleic acid molecules comprising the coding sequence sequence of an ORF as set forth in any of SEQ0055 to SEQ0059, or a nucleotide sequence that is substantially homologous to the coding sequence of an ORF as set forth in any of SEQ0055 to SEQ0059.

[061] In another embodiment, the invention is drawn to an isolated protein comprising an amino acid sequence (e.g., a full-length protein sequence) as set forth in any of SEQ0055 to SEQ0059, or an amino acid sequence substantially homologous thereto.

[062] In one embodiment, the invention is directed to a methylotrophic yeast strain, for example a *Pichia* strain such as *Pichia pastoris*, which overexpresses the *Pichia pastoris* SEC1 gene as set forth in SEQ0055. Expression vectors for achieving such overexpression are also contemplated by the present invention.

[063] Analysis of mutants of *S. cerevisiae* shows SEC1 to be required for the SNARE mediated docking and fusion of exocytic vesicles. Therefore, overexpression of the native SEC1 gene in *P. pastoris* may facilitate the exocytosis of secretory vesicles. The coding sequence of the *P. pastoris* SEC1 gene can be cloned into an expression vector and is introduced into a methylotrophic yeasr strain. To evaluate whether secretion levels are increased as a result of such overexpression, the quantity of secreted (glyco)proteins per cell can be analyzed.

[064] In another embodiment, the invention is directed to a methylotrophic yeast strain, for example a *Pichia* strain such as *Pichia pastoris*, which overexpresses at least one of the *Pichia pastoris* SEC11, SPC1, SPC2 or SPC3 genes as set forth in SEQ0056-0059. Expression vectors for achieving such overexpression are also provided by the present invention.

[065] Efficient processing of the signal peptide of a secreted protein is essential for high yield and to eliminate the presence of additional amino acids in the secreted protein. Overexpression of one or more subunits of the signal peptidase complex may increase the efficiency and the quality of the processing. To evaluate whether the efficiency of the signal peptide cleavage is improved as a result of overexpression of a subunit of signal peptidase complex, the quantity and the amino acid sequence of a heterologous secreted glycoproteins can be analyzed.

P. pastoris homologues of genes involved in the glycolysis pathway

[066] In accordance with the present invention, 26 genes have been identified as encoding proteins involved in the glycolysis pathway. In the annotated Sequence file provided herein, the nucleotide sequences of the 5' upstream region (including promoter), the ORF and the 3' downstream region, as well as the amino acid sequence, of each of these 26 genes are set forth in SEQ0060 to SEQ0085.

[067] In one embodiment, the present invention is directed to isolated nucleic acid molecules comprising a nucleotide sequence that encodes a protein (e.g., a full-length protein) as set forth in any of SEQ0060 to SEQ0085, or encodes a protein that is substantially homologous to a full-length protein as set forth in any of SEQ0060 to SEQ0085.

[068] In a specific embodiment, the present invention is directed to isolated nucleic acid molecules comprising the coding sequence of an ORF as set forth in any of SEQ0060 to SEQ0085, or a nucleotide sequence that is substantially homologous thereto.

[069] In another embodiment, the invention is drawn to an isolated protein comprising an amino acid sequence (e.g., a full-length protein sequence) as set forth in any of SEQ0060 to SEQ0085, or an amino acid sequence substantially homologous thereto.

[070] In a further embodiment, the present invention is directed to the promoters of each of the genes as set forth in SEQ0060 to SEQ0085. Because glycolysis is a central metabolic pathway, the promoters of the genes encoding proteins involved in this pathway are believed

to be strong promoters, which can be used for driving overexpression of heterologous genes in methylotrophic yeast such as *P. pastoris*. Accordingly, the present invention provides expression vectors containing any one of the promoters of the genes set forth in SEQ0060 to SEQ0085, operably linked to the coding sequence of a heterologous protein. Methylotrophic yeast strains transformed with such an expression vector are also provided by the invention.

P. pastoris homologues of genes involved in homologous recombination

[071] In accordance with the present invention, 10 genes have been identified as encoding proteins involved in the homologous recombination. In the annotated Sequence file provided herein, the nucleotide sequences of the 5' upstream region (including promoter), the ORF and the 3' downstream region, as well as the amino acid sequence, of each of these 10 genes are set forth in SEQ0086 to SEQ0095.

[072] In one embodiment, the present invention is directed to isolated nucleic acid molecules comprising a nucleotide sequence that encodes a protein (e.g., a full-length protein) as set forth in any of SEQ0086 to SEQ0095, or encodes a protein that is substantially homologous to a full-length protein as set forth in any of SEQ0086 to SEQ0095.

[073] In a specific embodiment, the present invention is directed to isolated nucleic acid molecules comprising the coding sequence of an ORF as set forth in any of SEQ0086 to SEQ0095, or a nucleotide sequence that is substantially homologous thereto.

[074] In another embodiment, the invention is drawn to an isolated protein comprising an amino acid sequence (e.g., a full-length protein sequence) as set forth in any of SEQ0086 to SEQ0095, or an amino acid sequence substantially homologous thereto.

[075] In a further embodiment, the invention is directed to methylotrophic yeast strains, especially *Pichia pastoris* strains, wherein one or more of the genes as set forth in SEQ0086 to SEQ0095 have been inactivated. Inactivation of these genes involved in recombination may prevent out-recombination of an expression unit or cassette containing a heterologous

gene integrated in the chromosome, therefore potentially "lock" or stabilize the insertion, especially multicopy insertions as further discussed hereinbelow.

Genes with high expression levels

[076] In accordance with the present invention, 29 genes have been identified as encoding proteins and showing expression levels 20x higher than GAP1 based on microarray analysis. In the annotated Sequence file provided herein, the nucleotide sequences of the 5' upstream region (including promoter), the ORF and the 3' downstream region, as well as the amino acid sequence, of each of these 10 genes are set forth in SEQ0096 to SEQ0124.

[077] In one embodiment, the present invention is directed to isolated nucleic acid molecules comprising a nucleotide sequence that encodes a protein (e.g., a full-length protein) as set forth in any of SEQ0096 to SEQ0124, or encodes a protein that is substantially homologous to a full-length protein as set forth in any of SEQ0096 to SEQ0124.

[078] In a specific embodiment, the present invention is directed to isolated nucleic acid molecules comprising the coding sequence of an ORF as set forth in any of SEQ0096 to SEQ0124, or a nucleotide sequence that is substantially homologous thereto.

[079] In another embodiment, the invention is drawn to an isolated protein comprising an amino acid sequence (e.g., a full-length protein sequence) as set forth in any of SEQ0096 to SEQ0124, or an amino acid sequence substantially homologous thereto.

[080] In a further embodiment, the present invention is directed to the promoters of each of these 29 genes showing high expression levels. These promoters can be used for driving overexpression of heterologous genes in methylotrophic yeast such as *P. pastoris*. Accordingly, the present invention provides expression vectors containing any one of the promoters of the 29 genes set forth in SEQ0096 to SEQ0124, operably linked to the coding sequence of a heterologous protein. Methylotrophic yeast strains transformed with such an expression vector are also provided by the invention.

Homologs of promoters used for expression of proteins in S. cerevisiae

[081] In accordance with the present invention, 4 genes have been identified as encoding *P. pastoris* protein homologs of *S. cerevisiae* whose promoters have been used for recombinant expression in *S. cerevisiae*, including SEQ0125 (homolog of *S. cerevisiae* glycerol-3-phosphate dehydrogenase 1 (GPD1) and GPD2), SEQ0126 (homolog of *S. cerevisiae* alcohol dehydrogenase 1 (ADH1) and ADH2), SEQ0127 (homolog of *S. cerevisiae* PHO5, although partial promoter and the ORF of the *Pichia pastoris* gene have been reported in the art), an SEQ0128 (homolog of *S. cerevisiae* sulfite reductase beta subunit ECM17). In the annotated Sequence file provided herein, the nucleotide sequences of the 5' upstream region (including promoter), the ORF and the 3' downstream region, as well as the amino acid sequence, of each of these 4 genes are set forth in SEQ0125 to SEQ0128.

[082] In one embodiment, the present invention is directed to isolated nucleic acid molecules comprising a nucleotide sequence that encodes a protein (e.g., a full-length protein) as set forth in any of SEQ0125, SEQ0126 or SEQ0128, or encodes a protein that is substantially homologous to a full-length protein as set forth in any of SEQ0125, SEQ0126 or SEQ0128.

[083] In a specific embodiment, the present invention is directed to isolated nucleic acid molecules comprising the coding sequence of an ORF as set forth in any of SEQ0125, SEQ0126 or SEQ0128, or a nucleotide sequence that is substantially homologous thereto.

[084] In another embodiment, the invention is drawn to an isolated protein comprising an amino acid sequence (e.g., a full-length protein sequence) as set forth in any of SEQ0125, SEQ0126 or SEQ0128, or an amino acid sequence substantially homologous thereto.

[085] In a further embodiment, the present invention is directed to the promoters of each of these 4 genes. These promoters can be used for driving expression of heterologous genes in *Pichia pastoris* or another methylotrophic yeast species. Expression vectors containing any of these promoters, operably linked to a heterologous coding sequence, and methylotrophic yeast strains transformed with any such expression vectors, are contemplated by the invention.

Genes involved in nucleotide sugar synthesis and transport

[086] In accordance with the present invention, 4 genes have been identified as encoding proteins involved in nucleotide sugar synthesis and transport, including UDP-GlcNAc transporter (SEQ0129), UDP-glucose-4-epimerase (SEQ0130), HUT1 (putative role of transporting UDP-galactose into Golgi) (SEQ0131), and a putative UDP-galactose transporter (SEQ0132). In the annotated Sequence file provided herein, the nucleotide sequences of the 5' upstream region (including promoter), the ORF and the 3' downstream region, as well as the amino acid sequence, of each of these 4 genes are set forth in SEQ0129 to SEQ0132.

[087] Efforts have been made to re-engineer the entire glycosylation pathway of *Pichia pastoris* in order to produce complex or hybrid type N-glycans (N-glycosylation humanization technology[1,2]; see also **Fig. 3b**). The heterologous glycosyltransferases needed for this engineered glycosylation processing utilize the sugar-nucleotides UDP-GlcNAc and UDP-Gal as monosaccharide donors. The identification of a UDP-GlcNAc transporter in the *Pichia* genome is consistent with the fact that UDP-GlcNAc is known to be synthesized in yeasts for the synthesis of cell wall chitin. However, no galactosylated glycoconjugates in *P. pastoris* have been previously described. It has been shown that the mere overexpression of a *Pichia* Golgi-targeted version of human beta-1,4-galactosyltransferase I is sufficient to achieve galactosylation of secreted glycoproteins[7]. While this finding may suggest that *Pichia* produces UDP-Gal and transports it into the Golgi apparatus, the present invention provides the molecular basis for the first time by identifying an endogenous cytoplasmic UDP-Glc-4-epimerase, and clear homologs of Golgi UDP-Galactose transporters. See also **Table 5B**. Researchers have previously overexpressed a heterologous UDP-Glc-4-epimerase in fusion to the galactosyltransferase to achieve higher levels of UDP-Gal in the yeast Golgi apparatus[2,8]. The identification of the *Pichia pastoris* proteins involved in nucleotide sugar synthesis and transport permits more effective glycan engineering in this and other methylotrophic yeasts.

[088] In one embodiment, the present invention is directed to isolated nucleic acid molecules comprising a nucleotide sequence that encodes a protein (e.g., a full-length protein) as set

forth in any of SEQ0129 to SEQ0132, or encodes a protein that is substantially homologous to a full-length protein as set forth in any of SEQ0129 to SEQ0132.

[089] In a specific embodiment, the present invention is directed to isolated nucleic acid molecules comprising the coding sequence of an ORF as set forth in any of SEQ0129 to SEQ0132, or a nucleotide sequence that is substantially homologous thereto.

[090] In another embodiment, the invention is drawn to an isolated protein comprising an amino acid sequence (e.g., a full-length protein sequence) as set forth in any of SEQ0129 to SEQ0132, or an amino acid sequence substantially homologous thereto.

[091] In a further embodiment, the present invention is directed to expression vectors to achieve increased expression in *Pichia* of native glycosylation precursor synthesis enzymes or transporters, e.g. UDP-Gal or UDP-GlcNAc transporters and UDP-Glc-4-epimerase for ER or Golgi localization. It is believed that increased expression of these proteins can increase homogeneity of final hybrid or complex type glycoform on the heterologous protein designed to be expressed in the same strain. The expression vectors contain the coding sequence of one of the desirable glycosylation precursor synthesis enzymes or transporters, which is placed in operable linkage to a promoter functional in recipient host cells. The promoters that control the expression of the precursor synthesis enzymes or transporters can be selected to be induced by similar conditions to those directing the expression of the heterologous protein.

[092] In another embodiment, the invention provides methylotrophic yeast strains especially *Pichia* strains which overexpress one or more native or *Pichia pastoris* glycosylation precursor synthesis enzymes or transporters, e.g. UDP-Gal or UDP-GlcNAc transporters and UDP-Glc-4-epimerase for ER or Golgi localization

Genes involved in O-glycosylation

[093] In accordance with the present invention, 5 genes have been identified as encoding proteins involved in O-glycosylation. In the annotated Sequence file provided herein, the nucleotide sequences of the 5' upstream region (including promoter), the ORF and the 3'

downstream region, as well as the amino acid sequence, of each of these 5 genes are set forth in SEQ0133 to SEQ0137.

[094] In one embodiment, the present invention is directed to isolated nucleic acid molecules comprising a nucleotide sequence that encodes a protein (e.g., a full-length protein) as set forth in any of SEQ0133 to SEQ0137, or encodes a protein that is substantially homologous to a full-length protein as set forth in any of SEQ0133 to SEQ0137.

[095] In a specific embodiment, the present invention is directed to isolated nucleic acid molecules comprising the coding sequence of an ORF as set forth in any of SEQ0133 to SEQ0137, or a nucleotide sequence that is substantially homologous thereto.

[096] In another embodiment, the invention is drawn to an isolated protein comprising an amino acid sequence (e.g., a full-length protein sequence) as set forth in any of SEQ0133 to SEQ0137, or an amino acid sequence substantially homologous thereto.

[097] In addition to N-glycosylation, yeasts also O-glycosylate secreted proteins with oligomannosyl-glycans that differ from the mucin-type O-glycosylation in humans[9]. No robust engineering approach has yet been developed to overcome this issue prior to this invention. The identification herein of the *Pichia* protein-O-mannosyltransferases that initiate this modification in the ER permits genetic modification of *Pichia* to reduce O-glycosylation.

[098] Specifically, in a further embodiment, the present invention is directed to methylotrophic yeast strains, e.g., *Pichia* such as *P. pastoris* strains, in which one of more of the identified O-mannosyl transferase genes are inactivated to reduce or eliminate unwanted O-glycosylation of a heterologous protein.

Genes encoding mannosyltransferases

[099] In accordance with the present invention, 18 genes have been identified as encoding mannosyltransferases, including 14 α- mannosyltransferases. In the annotated Sequence file provided herein, the nucleotide sequences of the 5' upstream region (including promoter), the

ORF and the 3' downstream region, as well as the amino acid sequence, of each of these 18 genes are set forth in SEQ0138-0152 and 0210-0212.

[0100] In one embodiment, the present invention is directed to isolated nucleic acid molecules comprising a nucleotide sequence that encodes a protein (e.g., a full-length protein) as set forth in any of SEQ0138-0152 or 0210-0212, or encodes a protein that is substantially homologous to a full-length protein as set forth in any of SEQ0138-0152 or 0210-0212.

[0101] In a specific embodiment, the present invention is directed to isolated nucleic acid molecules comprising the coding sequence of an ORF as set forth in any of SEQ0138-0152 or 0210-0212, or a nucleotide sequence that is substantially homologous thereto.

[0102] In another embodiment, the invention is drawn to an isolated protein comprising an amino acid sequence (e.g., a full-length protein sequence) as set forth in any of SEQ0138 to SEQ0152, or an amino acid sequence substantially homologous thereto.

[0103] In terms of N-glycosylation, methylotrophic yeasts such as *P. pastoris* modify proteins with a range of heterogeneous high-mannose glycans, which introduce a large amount of heterogeneity in the protein (reducing downstream processing efficiency and complicating product characterization) and induce fast clearance from the bloodstream. The highly immunogenic terminal alpha-1,3-mannosyl glycotypes that are abundantly produced by *S. cerevisiae* are not detected on *Pichia*-produced glycoproteins. Consistently, no ortholog of the *S. cerevisiae* MNN1 gene (encoding the alpha-1,3-mannosyltransferase) has been found in the *Pichia* genome. However, *Pichia* glycoproteins can in some cases be modified with beta-1,2-mannose residues[10], reminiscent of antigenic epitopes on the *Candida albicans* cell wall[11]. *P. pastoris* AMR2 beta-mannosyltransferase, which has been documented in the art, has been identified in the genome. 3 homologs of AMR2 beta-mannosyltransferase have also been identified (SEQ0210-0212), thus providing the basis for reducing the levels of undesired beta-mannosylation.

[0104] Specifically, in a further embodiment, the present invention is directed to methylotrophic yeast strains, particularly *Pichia* strains, in which one of more of the identified

P. pastoris mannosyltransferase genes are inactivated, for example, the homolog genes of AMR2 beta-mannosyltransferase as set forth in SEQ0210-0212, to reduce or eliminate unwanted beta-mannosylation of a heterologous protein.

Genes encoding proteins involved in ER glycosylation pathway

[0105] In accordance with the present invention, 13 genes have been identified as encoding proteins involved in the ER glycosylation pathway. In the annotated Sequence file provided herein, the nucleotide sequences of the 5' upstream region (including promoter), the ORF and the 3' downstream region, as well as the amino acid sequence, of each of these 13 genes are set forth in SEQ0153 to SEQ0165.

[0106] In one embodiment, the present invention is directed to isolated nucleic acid molecules comprising a nucleotide sequence that encodes a protein (e.g., a full-length protein) as set forth in any of SEQ0153 to SEQ0165, or encodes a protein that is substantially homologous to a full-length protein as set forth in any of SEQ0153 to SEQ0165.

[0107] In a specific embodiment, the present invention is directed to isolated nucleic acid molecules comprising the coding sequence of an ORF as set forth in any of SEQ0153 to SEQ0165, or a nucleotide sequence that is substantially homologous thereto.

[0108] In another embodiment, the invention is drawn to an isolated protein comprising an amino acid sequence (e.g., a full-length protein sequence) as set forth in any of SEQ0153 to SEQ0165, or an amino acid sequence substantially homologous thereto.

[0109] In a further embodiment, the invention is directed to methylotrophic yeast strains, especially *Pichia* strains such as *P. pastoris*, in which one or more of the genes as set forth in SEQ0153-0165 have been inactivated. An alternative to modification of glycans after transfer to a protein is the modification of the glycan precursor before transfer to the protein. Inactivation of enzyme activities in the synthesis of the glycan precursor in the ER can result in glycolysation of proteins with modified glycan structures. In the event that inactivation of one of the genes as set forth in SEQ0153-0165 results in a lethal phenotype as the generated

glycan-precursor is a non-optimal substrate for further processing steps, overexpression of a downstream enzyme, or expression of a modified or alternative downstream enzyme (e.g., LmSTT3 as further discussed below), may overcome this defect.

[0110] In still another embodiment, the invention is directed to the use of the *Pichia pastoris* STT3 gene sequence as set forth in SEQ0163 to disrupt the chromosomal STT3 gene, and optionally, to further insert a heterologous STT3 gene, such as the *Leishmania* STT3 gene into the *Pichia pastoris* STT3 locus.

[0111] In yeast, STT3 is a part of the oligosaccharyl-transferase (OT) complex which transfers the lipid linked oligosaccharide to the protein. *Leishmania major* has four STT3 paralogues, of which 3 could complement a yeast stt3 deletion. Furthermore, it has been reported that LmSTT3 does not work in the OT-complex but is active as a dimeric complex. It is suggested that the various LmSTT3 dimeric complexes display different protein substrate specificities at the level of individual glycosylation sequences. In addition, the LmSTT3D dimeric complex has a relaxed specificity with respect to the lipid linked oligosaccharide substrate. In contrast to the homogenous OT-complex, LmSTT3 OTase has no reduced transfer efficiency of glycans lacking α-1,2-linked mannoses on the B and C branch[12]. Replacing the *Pichia pastoris* STT3 by LmSTT3D can provide glycosylation flexibility when the native OT-complex of *Pichia pastoris* is unable to transfer modified lipid-linked oligosaccharides to the protein.

[0112] "Knock-in" of LmSTT3 - The coding sequence of the *LmSTT3D* gene is amplified by PCR using specific primers. The coding sequence is subsequently cloned into the expression plasmid using unique restriction sites which places the coding sequence under control of a promoter of the expression plasmid, these restriction sites are incorporated in the primers. The expression vector contains a sequence of the *Pichia pastoris* STT3 gene ("PpSTT3"). Prior to transformation in the *P. pastoris* strain, the plasmid containing the LmSTT3 expression cassette was digested with a unique site in the PpSTT3 sequence to facilitate integration. Transformants are selected for by using the selection marker present on the

expression plasmid. To evaluate whether glycosylation is increased, N-glycans derived from secreted glycoproteins are analyzed by DSA-FACE capillary electrophoresis.

[0113] "Knock-out" of PpSTT3 - The coding sequences of the *LmSTT3D* gene is amplified by PCR using specific primers. The coding sequence is subsequently cloned into the expression plasmid using unique restriction sites which place the coding sequence under control of a promoter of the expression plasmid, these restriction sites are incorporated in the primers. The expression vector contains two *Pichia pastoris* sequences flanking the LmSTT3 expression cassette, these sequences are upstream and downstream of the PpSTT3 gene. Prior to transformation into the *P. pastoris* strain, the plasmid containing the LmSTT3 expression cassette was digested with restriction endonuclease(s), excising the LmSTT3 cassette containing the PpSTT3 sequences. Transformants are selected by using the selection marker present in the expression cassette. To evaluate whether glycosylation is increased, N-glycans derived from secreted glycoproteins are analyzed by DSA-FACE capillary electrophoresis.

Other genes encoding proteins involved in the glycosylation pathway

[0114] In accordance with the present invention, 3 genes have been identified as encoding additional proteins involved in the glycosylation pathway (*S. cerevisiae* MNN4 homologs). In the annotated Sequence file provided herein, the nucleotide sequences of the 5' upstream region (including promoter), the ORF and the 3' downstream region, as well as the amino acid sequence, of each of these 3 genes are set forth in SEQ0166 to SEQ0168.

[0115] In one embodiment, the invention is directed to methylotrophic yeast strains which overexpress one or more of the MNN4 homologs to promote the core type phosphorylation of N-glycans. Increased phosphorylation of recombinant proteins can be useful in directing the protein for uptake throught the mannose-6-phosphate receptor[13]. Expression vectors for achieving such elevated expression are also part of the present invention. To evaluate whether manno-phosphorylation is increased from such modified strains, N-glycans derived from secreted glycoproteins after 48 hours culture in YPD medium are analyzed by DSA-FACE capillary electrophoresis. The amount of $Man_8GlcNAc_2$ will be drastically reduced in favor of two structures that migrate faster (compared to $Man_8GlcNAc_2$) and that are likely to

contain one (P) and two (PP) phosphate residues, respectively. Assuming that both peaks derive from the $Man_8GlcNAc_2$ peak, the amount of $Man_8GlcNAc_2$ converted to phosphorylated glycans can be quantitated.

Genes encoding proteins involved methanol metabolism

[0116] In accordance with the present invention, 10 genes have been identified as encoding proteins involved in methanol metabolism. In the annotated Sequence file provided herein, the nucleotide sequences of the 5' upstream region (including promoter), the ORF and the 3' downstream region, as well as the amino acid sequence, of each of these 10 genes are set forth in SEQ0169 to SEQ0178. While some of these genes have been documented in the art, the promoters have not been identified prior to the present application.

[0117] The commonly used methanol-inducible promoters in *P. pastoris*, the alcohol oxidase I (AOXI) promoter and the formaldehyde dehydrogenase (FLD) promoter, drive the production of enzymes needed for methanol assimilation and therefore produce extremely high levels of these transcripts upon switching the carbon source to methanol. The *P. pastoris* genome sequence has now allowed identification of all genes coding for enzymes involved in methanol assimilation (**Fig. 3A** and **Table 5A**) and their promoters, which are useful for driving transgene expression in *P. pastoris*.

[0118] In one embodiment, the present invention is directed to isolated nucleic acid molecules comprising a nucleotide sequence that encodes a protein (e.g., a full-length protein) as set forth in any of SEQ0172- SEQ 0174 or SEQ0176 to SEQ0178, or encodes a protein that is substantially homologous to a full-length protein as set forth in any of SEQ0172- SEQ 0174 or SEQ0176 to SEQ0178.

[0119] In a specific embodiment, the present invention is directed to isolated nucleic acid molecules comprising the coding sequence of an ORF as set forth in any of SEQ0172- SEQ 0174 or SEQ0176 to SEQ0178, or a nucleotide sequence that is substantially homologous thereto.

[0120] In another embodiment, the invention is drawn to an isolated protein comprising an amino acid sequence (e.g., a full-length protein sequence) as set forth in any of SEQ0172-SEQ 0174 or SEQ0176 to SEQ0178, or an amino acid sequence substantially homologous thereto.

[0121] In a further embodiment, the present invention is directed to the promoters of the genes disclosed in SEQ0169 to SEQ0178. These promoters can be placed in an operable linkage to a heterologous gene for methanol-inducible recombinant expression in *Pichia pastoris*. Therefore, expression vectors, host cells and methods of recombinant expression by utilizing any of the promoters disclosed in SEQ0169 to SEQ0178 are also embodiments of the present invention.

Genes encoding protein homologs of S. cerevisiae proteases and protease inhibitor

[0122] In accordance with the present invention, 8 genes have been identified as encoding protein homologs of *S. cerevisiae* proteases or protease inhibitors. In the annotated Sequence file provided herein, the nucleotide sequences of the 5' upstream region (including promoter), the ORF and the 3' downstream region, as well as the amino acid sequence, of each of these 8 genes are set forth in SEQ0179 to SEQ0186. SEQ0179 sets forth a *Pichia pastoris* gene encoding a serine-type peptidase. SEQ0180 sets forth a *Pichia pastoris* gene encoding a serine-type endopeptidase inhibitor. SEQ0181-0186 sets forth *Pichia pastoris* genes coding for aspartic-type endopeptidases.

[0123] In one embodiment, the present invention is directed to isolated nucleic acid molecules comprising a nucleotide sequence that encodes a protein (e.g., a full-length protein) as set forth in any of SEQ0179 to SEQ0186, or encodes a protein that is substantially homologous to a full-length protein as set forth in any of SEQ0179 to SEQ0186.

[0124] In a specific embodiment, the present invention is directed to isolated nucleic acid molecules comprising the coding sequence of an ORF as set forth in any of SEQ0179 to SEQ0186, or a nucleotide sequence that is substantially homologous thereto.

[0125] In another embodiment, the invention is drawn to an isolated protein comprising an amino acid sequence (e.g., a full-length protein sequence) as set forth in any of SEQ0179 to SEQ0186, or an amino acid sequence substantially homologous thereto.

[0126] In a further embodiment, the present invention is directed to vectors for inactivating one or more of the *Pichia* proteases, and to *Pichia* strains having one or more of the protease genes inactivated.

[0127] In another embodiment, the present invention is directed to expression vectors capable of expressing the serine-type endopeptidase inhibitor (SEQ0180) in a methylotrophic yeast strain such as *P. pastoris*, and methylotrophic yeast strains such as *P. pastoris* engineered transformed with such expression vector to produce the endopeptidase inhibitor at an elevated level.

[0128] The protease-deficient *Pichia* strains and strains that produce the endopeptidase inhibitor at an elevated level (i.e., overpression), are believed to allow more stable accumulation of a heterologous protein expressed in these strains. Such strains may be useful especially for producing recombinant immunoglobulins.

Genes encoding chaperones

[0129] In accordance with the present invention, 14 genes have been identified as encoding chaperones. In the annotated Sequence file provided herein, the nucleotide sequences of the 5' upstream region (including promoter), the ORF and the 3' downstream region, as well as the amino acid sequence, of each of these 14 genes are set forth in SEQ0187 to SEQ0200.

[0130] In one embodiment, the present invention is directed to isolated nucleic acid molecules comprising a nucleotide sequence that encodes a protein (e.g., a full-length protein) as set forth in any of SEQ0187 to SEQ0200, or encodes a protein that is substantially homologous to a full-length protein as set forth in any of SEQ0187 to SEQ0200.

[0131] In a specific embodiment, the present invention is directed to isolated nucleic acid molecules comprising the coding sequence of an ORF as set forth in any of SEQ0187 to SEQ0200, or a nucleotide sequence that is substantially homologous thereto.

[0132] In another embodiment, the invention is drawn to an isolated protein comprising an amino acid sequence (e.g., a full-length protein sequence) as set forth in any of SEQ0187 to SEQ0200, or an amino acid sequence substantially homologous thereto.

[0133] The present invention has now provided a complete catalog of orthologs of the *S. cerevisiae* ER folding machinery. This information is especially useful for design of an efficacious folding system. In one embodiment, the present invention is directed to expression vectors and engineered *Pichia* strains for increased expression of one or more native *Pichia* chaperones. The chaperone coding sequences can be placed under the control of a promoter selected to be induced by similar conditions to those for heterologous protein expression. Expression libraries of multiple chaperones can be screened to identify the most effective combination of chaperon expression for a particular heterologous protein. For example, a series of libraries of chaperones can be created, with each library having a different drug resistance marker incorporated for selection, such that either successive or combinatorial introduction of members of these libraries into a strain expressing a heterologous protein may be selected.

5S ribosomal RNA gene

[0134] In accordance with the present invention, the *P. pastoris* 5S ribosomal RNA gene has been identified and set forth in SEQ0201.

[0135] One strategy for optimizing protein expression levels in a recipient cell is to increase copy number of the expression cassette. This effort has been hampered by the absence of knowledge regarding sequences which occur multiple times in the *Pichia* genome and could be used for stable multi-copy strain generation through homologous recombination-mediated targeting of such multi-copy sequences. The 5S rRNA coding-sequence of *Pichia pastoris* provides the basis for multi-copy targeting. Contrary to the situation in *S. cerevisiase*, the 5S

rRNA-coding sequence is not a part of the rDNA repeat locus in *Pichia pastoris*, and many copies of the 5S rRNA coding-sequence are spread over the 4 chromosomes of *Pichia pastoris* (**Fig. 1B**), thus providing an ideal targeting site for multi-copy integration.

[0136] To practice this aspect of the invention, the 5S rRNA coding sequence can be placed in a vector which also carries an expression cassette of interest and a selectable marker. The selectable marker can be a dominant selection marker that confers drug resistance, or a marker that confers phenotype selectable based on prototrophy. A unique restriction site is made or designed to be available within the 5S rRNA coding sequence. The vector is linearized by using the restriction enzyme that cleaves in the 5S rRNA coding sequence. The linearized vector is transformed into *Pichia pastoris*, and drug-resistant clones are isolated at increasing drug concentrations. Those clones that are resistant against the highest drug concentrations are expectedly those that have taken up the largest number of expression cassettes. Alternatively, when the selectable marker supplies an enzyme (e.g., HIS4), the clones that grow faster in the appropriate seletable media are identified as having multicopy integration. This procedure can be repeated until an optimal number of expression cassettes in the strain is obtained. Protein production in the selected strains is evaluated using methods known in the art.

[0137] Accordingly, the present invention also provides expression vectors capable of mediating multi-copy integration of an expression cassette onto the chromosomes of *Pichia pastoris*, as well as *Pichia pastoris* containing multiple copies of the expression cassette, stably integrated into the chromosomes. An expression cassette, as used in the present context, refers to a nucleic acid that includes, from 5' to 3', a promoter, the coding sequence of a heterologous protein of interest, and a 3' downstream sequence including a transcription termination sequence.

Genes encoding proteins involved in xylose, arabinose and threhalose metabolism

[0138] In accordance with the present invention, 8 genes have been identified as encoding proteins involved in xylose, arabinose or threhalose metabolism. In the annotated Sequence

file provided herein, the nucleotide sequences of the 5' upstream region (including promoter), the ORF and the 3' downstream region, as well as the amino acid sequence, of each of these 7 genes are set forth in SEQ0202 to SEQ0209.

[0139] In one embodiment, the present invention is directed to isolated nucleic acid molecules comprising a nucleotide sequence that encodes a protein (e.g., a full-length protein) as set forth in any of SEQ0202 to SEQ0209, or encodes a protein that is substantially homologous to a full-length protein as set forth in any of SEQ0202 to SEQ0209.

[0140] In a specific embodiment, the present invention is directed to isolated nucleic acid molecules comprising the coding sequence of an ORF as set forth in any of SEQ0202 to SEQ0209, or a nucleotide sequence that is substantially homologous thereto.

[0141] In another embodiment, the invention is drawn to an isolated protein comprising an amino acid sequence (e.g., a full-length protein sequence) as set forth in any of SEQ0202 to SEQ0209, or an amino acid sequence substantially homologous thereto.

[0142] In still another embodiment, the present invention is directed to the promoters of the genes disclosed in SEQ0202 to SEQ0209. These promoters are expected by the induced by specific sugars (xylose and arabinose metabolism pathway by C5 sugars, and threhalose pathway by α-1,1-disaccharides) can be placed in an operable linkage to a heterologous gene for inducible recombinant expression in *Pichia pastoris*. Therefore, expression vectors, host cells and methods of recombinant expression by utilizing any of the promoters disclosed in SEQ0202 to SEQ0209 are also embodiments of the present invention.

General Methodology

[0143] Gene knockouts

[0144] To reduce the protease, protein-O-mannosyltransferase or beta-mannosyltransferase activity in *Pichia pastoris*, the genes encoding these enzymes are inactivated. This is achieved through standard yeast genetics techniques. Examples of such techniques include gene replacement through double homologous recombination, in which homologous regions

flanking the gene to be inactivated are cloned in a vector flanking a selectable marker gene (such as an antibiotic resistance gene or a gene complementing an auxotrophy of the yeast strain). Alternatively, the homologous regions can be PCR-amplified and linked through overlapping PCR to the selectable marker gene. Subsequently, such DNA fragments are transformed into *Pichia pastoris* through methods known in the art, e.g., electroporation. Transformants that then grow under selective conditions are analyzed for the gene disruption event through standard techniques, e.g. PCR on genomic DNA or Southern blot. In an alternative experiment, gene inactivation can be achieved through single homologous recombination, in which case, e.g. the 5' end of the gene's ORF is cloned on a promoterless vector also containing a selectable marker gene. Upon linearization of such vector through digestion with a restriction enzyme only cutting the vector in the target-gene homologous fragment, such vector is transformed into *Pichia pastoris*. Integration at the target gene site is confirmed through PCR on genomic DNA or Southern blot. In this way, a duplication of the gene fragment cloned on the vector is achieved in the genome, resulting in two copies of the target gene locus: a first copy in which the ORF is incomplete, thus resulting in the expression (if at all) of a shortened, inactive protein, and a second copy which has no promoter to drive transcription.

[0145] Alternatively, transposon mutagenesis is used to inactivate the target gene. A library of such mutants can be screened through PCR for insertion events in the target gene.

[0146] The functional phenotype (i.e., deficiencies) of an engineered/knockout strain can be assessed using techniques known in the art. For example, a deficiency of an engineered strain in protease activity can be ascertained using any of a variety of methods known in the art, such as an assay of hydrolytic activity of chromogenic protease substrates, band shifts of substrate proteins for the select protease, among others. A deficiency in protein O-mannosylation can be detected by mass changes of expressed O-glycosylated proteins and through mass spectrometrical techniques designed to detect the sites of O-glycosylation (such as beta-elimination in 18O-H2O), used in a comparative experiment with the same protein expressed in a non-engineered strain. Beta-mannosyltransferase deficiency can be detected

through glycan analysis of expressed proteins that are beta-mannosylated in non-engineered strains or through loss of signal with beta-mannosyl specific antibodies.

[0147] Gene over-expression

[0148] To increase the expression level in *Pichia pastoris* of desirable genes such as polypeptide-folding-promoting proteins, UDP-Gal or UDP-GlcNAc transporters and UDP-Glc-4-epimerase, the ORFs or coding sequences of such genes are cloned under the control of a promoter of desired strength and regulation (such as the methanol-inducible AOXI, AOXII or FLD promoters, or the constitutively expressed GAP or TEF1alpha promoter), in a vector also containing a selectable marker gene functional in *Pichia pastoris*. To increase transformation efficiency and/or target the vector to a selected genomic locus, the vector may be linearized in a genome-homologous sequence, although this is not essential. Subsequently, such vector is transformed into *Pichia pastoris* using techniques known in the art.

[0149] Promoter analysis and usage

[0150] Cloning of the promoters and generating engineered yeast strains - The promoter sequence is amplified from *Pichia pastoris* genomic DNA with primers designed to the 5' and 3' regions of the sequences provided herein. The corresponding PCR fragment is cloned in an appropriate vector by blunt or TA ligation or by restriction-ligation cloning using restriction sites, not present in the promoter, adapted to the primer. The promoter is subsequently cloned upstream of a gene of interest in a yeast expression vector. The expression vector is transferred into the *Pichia pastoris* strain according to protocols described by Cregg and Russel[14]. To generate stable transformants via homologous recombination, the vector is cleaved in a *P. pastoris* DNA segment (e.g., the *P. pastoris* promoter) with unique restriction sites. This allows integration into the genome by single crossover type insertion[15]. Transgenic yeast can be obtained on medium containing the selection marker or on medium lacking the complemented amino acid.

[0151] Analyzing promoter activity - To analyze the promoter activity, stable strains are generated as an example: a strain containing a marker gene under control of the promoter of

interest and a reference strain with the marker gene controlled by a reference promoter. Quantitative analysis of expression levels of this marker can determine the relative activity of the promoter.

[0152] Identifying promoter sequence and cis-acting elements - For the identification of the promoter sequence, a 5'-deletion series is generated by PCR on *P. pastoris* genomic DNA as template. For this deletion series, a set of forward primers is designed which hybridize at different distance from the start. As reverse primer, a primer is used that hybridizes to the 3' end of the suggested promoter sequence. In a first screen, the promoter deletions can be done in steps of 100 bp. By this a rough estimation of the promoter size can be made. Later a second screen in this region can be perfomed with smaller deletion steps. The promoter deletions are cloned in an expression vector, in front of a marker gene which allows to quantify the expression. Analysis of expression levels compared to the whole sequence will allow identification of the promoter.

[0153] To identify cis-acting elements, a series of 3'-deletion fragments of the promoter are obtained next to the 5'-deletions in a similar fashion. In addition, a minimal promoter is fused to the 3'-deletion fragments to compensate for the loss of the TATA-box. Further, combinations of the 3' deletions (without minimal promoter) and 5' deletions are made by overlapping PCR. Expression vectors are generated containing the promoter fragments controlling the expression of a marker gene. Analysis of the activity of these promoter fragments can identify regulatory elements in the promoter.

[0154] In addition to the methanol-inducible promoters of the Alcohol Oxidase I gene and the formaldehyde dehydrogenase I gene, and the constitutively active promoter of the glyceraldehyde-3-phosphate dehydrogenase gene, the present invention provides additional *Pichia pastoris* genes involved in the glycolysis pathway, methanol assimilation, xylose, arabinose and threhalose metabolisms. The promoters of these genes can be used for driving overexpression of heterologous genes.

[0155] For those promoters of the genes in the methanol utilization pathway, the xylose metabolism, the arabinose metabolism and the threhalose metabolism, these promoters can be used for inducible expression of a heterologous gene. To assess the inducibility of these promoters by methanol, xylose, arabinose or threhalose, the expression of their endogenous gene can be compared by qPCR of cells grown in medium with or without addition of the rsepective carbon sources. For example, *P. pastoris* wild type cells are grown in BMGY (uninduced condition), centrifuged and induced in BMMY (for methanol induction), in media containing xylose, in media containing arabinose or in media containing threhalose, respectively. At different time points, samples will be taken. At every time point, RNA is isolated from the yeast cells, treated with DNAse to remove all genomic DNA and then the mRNA converted into cDNA. On all genes to be tested, primers are designed (by Primer3Plus Software) and tested for amplification of the gene. To be able to compare the expression of the different genes, a good reference gene (e.g., actin; GAPDH or PDA1) is included. All reference genes are tested in uninduced and induced conditions by qPCR and the best reference gene is chosen by the geNorm software. After selecting the best reference gene, qPCR with primers on the reference gene and primers on the gene of interest can be performed on every sample. Thereafter, the relative quantity ($RQ= 2^{\Delta Cq}$, with ΔCq the difference in PCR cycles between induced and non-induced condition) of every gene can be calculated (qBASE*Plus* software). From these RQs, the ratio of the gene of interest (GOI) relative to to the reference gene ($NRQ= RQ_{GOI}/RQ_{REF)}$ is determined. When different genes are investigated, the gene that is most upregulated in the induced condition can be identified. The promoter driving this gene can then be used to drive expression of a heterologous test gene.

[0156] Counterpart genes from other methylotrophic yeast species

[0157] Given the availability of the *Pichia pastoris* genome and annotation, counterpart genes from other methylotrophic yeast species, e.g., other *Pichia* species, can be isolated. Isolation of such counterpart genes can be useful in generating engineered methylotrophic yeast strains other than *Pichia pastoris* for recombinant production of heterologous glycoproteins.

[0158] Other methylotrophic yeast species as genetic engineering host

[0159] The present invention contemplates the use of a methylotrophic yeast strain, including but not limited to *Pichia* strains such as *Pichia pastoris*, as the host for genetic engineering and recombinant production of proteins including glycoproteins. While the identification of native *Pichia pastoris* genes and their constituents (such as promoters, signal peptides, proteins including enzymes and chaperons) are preferred choices for engineering a *Pichia pastoris* strain, these choices are believed to also work for other methylotrophic yeast strains, especially other *Pichia* strains closely related to *Pichia pastoris*.

[0160] Combination of genetic modifications

[0161] Embodiments described herein can be combined as appropriate. For example, genetic modifications of a strain, such as usage of inducible promoters, usage of signal peptides, inactivation of one or more protease genes, overexpression of nucleotide sugar synthesis and transport, inactivation of O- or beta-mannosylation, expression of chaperones, and multicopy integration, for example, can be combined in any manner desirable.

[0162] The present invention is further supported and illustrated by the following examples.

EXAMPLE-1

[0163] This Example describes the methods employed to identify, assemble and annotate the full genomic sequence of *P. pastoris* GS115 strain.

DNA preparation

[0164] *P. pastoris* GS115 (Invitrogen, Carlsbad, CA) is a strain derived from the wild type strain NRRL-Y 11430 (Northern Regional Research Laboratories, Peoria, IL). It has a mutation in the histinol dehydrogenase gene (HIS4) and was generated by nitrosoguanidine mutagenesis at Phillips Petroleum Co. It is the most frequently used *Pichia* strain for heterologous protein production.

[0165] *P. pastoris* genomic DNA was prepared according to a published protocol with minor modifications. Instead of vortexing, the samples were shaked in a Mixer Mill (Retsch) for 2 minutes.

Sample preparation and sequencing with Roche/454 Genome Sequencer FLX

[0166] The shotgun library of *P. pastoris* for sequencing on the Genome Sequencer FLX (GS FLX) was prepared from five micrograms of intact genomic DNA. Based on random cleavage of the genomic DNA[16] with subsequent removal of small fragments with AMPure™ SPRI beads (Agencourt, Beverly, MA), the resulting single-stranded (sst) DNA library showed a fragment distribution between 300 and 900 bp with a maximum of 574 bp. The optimal amount of sstDNA library input for the emulsion PCR[16] (emPCR) was determined empirically through two small-scale titrations leading to 1.5 molecules per bead used for the large-scale approach. A total of 64 individual emPCRs were performed to generate 3,974,400 DNA carrying beads for two two-region-sized 70x75 PicoTiterPlates (PTP) and each region was loaded with 850.000 DNA carrying beads. Each of the two sequencing runs was performed for a total of 100 cycles of nucleotide flows[16] (flow order TACG) and the 454 Life Sciences/Roche Diagnostics software Version 1.1.03 was used to perform the image and signal processing. The information about read flowgram (trace) data, basecalls and quality scores of all high quality shotgun library reads was stored in a Standard Flowgram Format (SFF) file which was used by the subsequent computational analysis (see below).

[0167] Within this sequencing project, a paired end library of *P. pastoris* (strain GS115) was prepared for subsequent ordering and orienting of contigs (see computational analysis below). Six micrograms of intact genomic DNA was sheared hydrodynamically (Hydroshear - Genomic Solutions, Ann Arbor, MI) and purified with AMPure™ SPRI beads (Agencourt, Beverly, MA) into DNA fragments of about 3 kbp in length. After methylation of *Eco* RI restriction sites, a biotinylated hairpin adaptor was ligated to the ends of the *P. pastoris* DNA fragments, followed by *Eco* RI digestion with a subsequent circularization. The restriction of the circularized DNA fragments with *Mme* I, the subsequent ligation of paired end adaptors and the amplification of the remaining DNA fragments resulted in a double-stranded paired

end library with 130 bp in length. For the following eight individual emPCRs of the paired end library, 1.5 molecules per bead were used to generate 339,480 DNA-carrying beads of which 280,000 were loaded onto a region of a four-region sized 70 x75 PTP. The subsequent sequencing run with the GS FLX was performed for a total of 42 cycles of nucleotide flow (see above) and the 454 Life Sciences/Roche Diagnostics software Version 1.1.03 was used to perform the image and signal processing. The information about read flowgram (trace) data, basecalls and quality scores of all high quality shotgun library reads was also stored in a SFF file which was used by the subsequent computational analysis.

Computational analysis of GS FLX shotgun and paired end reads

[0168] An automatic assembly pipeline (in-house software, Eurofins MWG Operon) was used to *de novo* assemble the generated shotgun and paired end reads.

[0169] For *de novo* assembly of the *P. pastoris* genome sequence, a total of 897,197 good quality base called, clipped shotgun reads with an average read length of 243 bp and a total of 70,500 good quality base called, clipped 20 bp-Paired End tag reads were used.

[0170] Within this pipeline, the information about all sequences and their quality was extracted from the SFF- into a FASTA-file and subsequently converted into CAF format, the input format of choice of the used assembler mira (version 2.9 26x3) for contig creation. The provided mate and size information (*i.e.*, forward and reverse read and the 3 kbp of length) of the paired end reads was used to scaffold the resulting contigs from the *de novo* assembly[17].

*Assembly (**Fig. 1A** and **Fig. 4**)*

[0171] The initial assembly contained 1,154 contigs with 9.6 Mbp sequence and 20x sequencing depth. The contig N/L50 was 40/77 kbp. Assembly of the contigs was performed manually, based on homology between the contig ends. 13 contigs were assigned to chromosomes by identification of the chromosomal markers previously described[10] (Chromosome 1: HIS4, ARG4, OCH1, PAS5, PRB1, PRC1; Chromosome 2: PAS8, GAP; Chromosome 3: DAS1, URA3, PEP4; Chromosome 4: AOX1, AOX2). Starting from these

contigs, contigs with homologous contig ends were identified by BLASTN search with 500-1000 bp of the contig ends to a database with the contig sequences. Contigs sharing homology with a P-value < e-20 were assumed to be linked. Pools of potentially linked contigs were assembled to supercontigs by the SeqMan assembly software (DNASTAR inc, Madison, WI, USA). The resulting contig junctions were curated by removing the low-coverage ends of either joined contig. In the cases where the BLASTN P-value was >e-50, the junction was PCR-amplified and Sanger-sequenced (primer sequences: **Table 1**). This resulted in 10 supercontigs, with 9,1 Mbp of sequence, and a remaining 7 unassembled contigs. The supercontig N/L 50 was 3/1.544 Mbp.

[0172] The mitochondrial genome was also assembled and had extremely high coverage (859.9 fold), indicating the presence of approx. 43 mitochondrial genomes per cell in *P. pastoris* when grown on glucose as carbon source.

Gap joining and finishing

[0173] Supercontigs were linked by mapping contigs to paired-end scaffolds (n=1) and automated prediction of protein-coding sequences revealed a partial ORF at the end of a supercontig, homologous to a WD40 domain protein in other yeasts (including *Pichia guillermondii* homolog PGUG 04385). Finding the other part of this ORF on one of the unassembled contigs allowed joining of this supercontig to one of the as yet unassembled contigs. This was confirmed by PCR and Sanger sequencing.

[0174] Seven of the 9 thus generated supercontigs could be assigned to a specific chromosome when they contained one or more of the 13 genes for which chromosomal location had been previously established[10] (**Fig. 1B** and **Fig. 5C**). For those two supercontigs and the 6 unassembled contigs where this was not the case, Southern blot analysis of pulsed-field gel electrophoresis-separated *Pichia pastoris* chromosomes (see below) was used for the assignment (**Figs. 5A**). After assignment to the chromosomes, orientation of the supercontigs and contigs on the chromosomes was determined by PCR analysis with primers on the contig ends (**Table 1**). Gaps were PCR-amplified using primers flanking these regions (**Table 1**) and sequenced by Sanger sequencing for finishing.

[0175] rDNA repeat regions were detected by Southern blot on all four PFGE-separated chromosomes (**Fig. 5A**). The Southern signal on chromosomes 1 and 4 was as strong as that on chromosomes 2 and 3 combined. Through PCR, the location and orientation of the rDNA locus was determined to be at one end of Chromosomes 2 and 3 (**Fig. 1B**). The attempts at verification of the rDNA locus position on chromosomes 1 and 4 (still containing 1 gap) were inconclusive.

Pulsed-field Gel electrophoresis (PFGE)

[0176] A BioRad contour-clamped homogeneous electric field CHEF DRIII system was used for PFGE. Chromosomal DNA was prepared in agarose plugs with the CHEF Genomic DNA Plug kit (BioRad) following the instructions of the manufacturer. A 0.8% agarose gel in 1 x modified TBE (0.1 M Tris, 0.1 M Boric Acid, 0.2 mM EDTA) was used to separate the chromosomes. The gel was electrophoresed with a 106° angle at 14°C at 3 V/cm for 32 h, with a switch interval of 300 s, followed by 32 h with a switch interval of 600 s and 24 h with a switch interval of 900 s. After separation, the chromosomes were visualized with ethidium bromide and the different contigs were mapped onto the chromosomes by Southern blot. Therefore, the gel was incubated in 0.25 M HCl for 30 minutes, followed by capillary alkali transfer of the DNA onto a Hybond N^+ membrane (Amersham). The probes were prepared by PCR on an open reading frame. For chromosome specific probes, a part of the coding sequence of HIS4 (chromosome 1), GAP (Chromosome 2), URA3 (chromosome 3) and AOX1 (chromosome 4) was used. The probes were random labelled with $\alpha\ ^{32}$P dCTP, using the High Prime kit (Roche).

Automatic gene structure prediction & functional annotation

[0177] Protein-coding genes were predicted by the integrative gene prediction platform EuGene[18] (**Fig. 6**). A specific EuGene version was trained based on 108 manually checked *P. pastoris* genes. Documented genes from *P. stipitis* and *S. cerevisiae* were used to build *P. pastoris* orthologous gene models allowing the training of *P. pastoris*-specific Interpolated

Markov Models for coding sequences and introns. Splice sites were predicted by NetAspGene[19] and gene prediction from GeneMarkHMM-ES[20] trained for *P. pastoris* and AUGUSTUS[21] (*Pichia stipitis* model) were used to provide alternative gene models for EuGene prediction. The UniProt and the fungi RefSeq protein database were searched against the supercontig sequence by BLASTX to identify the coding area. The DeCypher-TBLASTX program was used to search the conserved sequence area between the *P. pastoris, P. stipitis* and *Candida guilliermondii* genomes.

[0178] All predicted protein-coding genes were searched against the yeast protein database, UniProt[22] and RefSeq[23] fungi protein database by BLASTP. Protein domains were detected by InterProScan[24] with various databases (BlastProDom, FPrintScan, PIR, Pfam, Smart, HMMTigr, SuperFamily, Panther and Gene3D) through the European Bioinformatics Institute Web Services SOAP-based web tools. Signal peptide and transmembrane helices were predicted by SignalP and TMHMM respectively. GO (Gene Ontology) terms[25] were derived from the InterProScan result and the KEGG (Kyoto Encycolopedia for Genes and Genomes) pathway and EC (Enzyme Commission) numbers were annotated by the annot8r pipeline.

Expert gene structure/functional annotation

[0179] The gene structure prediction and the database search results from various databases were formatted and stored in a MySQL relational database. A multiple alignment of each protein-coding gene with the top 10 best hits against the UniProt, RefSeq fungi and yeast protein database was built by MUSCLE[26]. A BOGAS (Bioinformatics Online Genome Annotation System) *P. pastoris* annotation website was setup as the workspace for expert annotators. The initial aim of BOGAS is to provide a workspace for gene structure and functional annotation. The editing of gene structure or gene function assignment is directly updated to the MySQL relational database through the web interface. All of the modification from expert annotators is traceable and reversible by the database system. Once the expert annotator modified the gene structure and change the translated protein product, the system will automatically trigger the update function to check the protein domain and protein database. BOGAS also provides a search function where users can search for genes by

sequence similarity (BLAST), gene id, gene name or InterPro domain. Each predicted *Pichia* gene's structure and the similarity search result were visually inspected through an embedded strip-down version of ARTEMIS. The splice sites of each gene were carefully checked and compared with *S. cerevisiae* and *P. stipitis* loci. A functional description of each gene was added to the gene annotation when a closely related homologous gene was available.

Estimate of the gene space completeness

[0180] Parra et al.[27] proposed a set of core eukaryotic genes (CEGs) to estimate the completeness of genome sequencing and assembly programmes. The CEGs contains 248 genes across six model organisms (*H. sapiens, D. melanogaster, C. elegans, A. thaliana, S. cerevisiae and S. pombe*) of which ~90% are single copy in *D. melanogaster, C. elegans, S. cerevisiae* and *S. pombe*. The protein-coding genes identified in this invention were checked with the HMM profile from the CEGs dataset by the HMMER package. All of the 248 CEGs were present in our curated gene set with full HMM domain coverage. Further, FUNYBASE (FUNgal phYlogenomic dataBASE)[28] provides 246 single-copy ortholog clusters in 21 sequenced fungal genomes. These single-copy protein sequences were extracted from the FUNYBASE website and built the HMM model for each cluster. The corrected *P. pastoris* protein sequences were searched with the FUNYBASE HMM database. All of the FUNYBASE models were presented in our gene catalog with complete domain coverage.

Detection of rRNA and tRNA loci

[0181] Ribosomal RNAs were detected automatically by INFERNAL 1.0 (INFERence of RNA ALignment) against the Rfam[29] database and manually confirmed by BLASTN search with *S. cerevisiae* homologs to the *P. pastoris* genome sequence. Localization of the ribosomal DNA locus was assayed by PFGE and PCR.

[0182] Transfer RNAs were automatically predicted by tRNA Scan-SE 1.21[30] and manually confirmed by BLASTN search with the *S. cerevisiae* homologs to the *P. pastoris* genome sequence.

Codon usage

[0183] Nucleotide sequences of the predicted *P. pastoris* ORFeome were analyzed with ANACONDA 1.5[31]. In addition to calculation of the codon use, the analysis by ANACONDA generates a codon-pair context map for the ORFeome. This map shows one colored square for each codon-pair, the first codon corresponds to rows and the second corresponds to columns in the map. Favored codon pairs are shown in green, underrepresented ones are shown in red.

Phylogenetic tree reconstruction of fungal genomes

[0184] The phylogenetic tree was based on 200 single-copy genes which were present in 12 sequenced fungal genomes. A multiple sequence alignment was constructed using the MUSCLE program and gap removal by in-house script based on the BLOSUM62 scoring matrix. The maximum likelihood tree reconstruction program TREE-PUZZLE[32] (quartet puzzling, WAG model, estimated gama distribution rate with 1000 puzzling step) was used for phylogenetic tree reconstruction. The tree was well supported by 1000 bootstraps in each node.

Comparative analysis of gene family and protein domain

[0185] The predicted proteomes used in this study were those of six hemiascomycetes (*P. pastoris, S. cerevisiae, K. lactis, P. stipitis, C. lustianiae and Y. lipolytica*)[33,34]. In order to obtain the gene families, a similarity search of all protein sequences from the 6 fungi (all-against-all BLASTP, e-value 1e-10) was performed. Gene families were constructed by Markov clustering[35] based on the BLASTP result. All predicted protein sequences from the six genomes were searched against the Pfam[36] database to obtain the protein domain

occurrence in each species. The protein domain loss and acquisition was counted based on the Dollo parsimony principle by the DOLLOP program from the PHYLIP package[37].

EXAMPLE-2

[0186] This Example describes the results of assembly and annotation of the full genomic sequence of *P. pastoris* GS115 strain.

Genome sequencing and assembly

[0187] Prior to this invention, very little was known about the genome features of *P. pastoris*. Ohi et al.[38] reported that the *P. pastoris* genome was organized in 4 chromosomes with a total estimated size of 9.7 Mbp (+/- a few hundred kbp, as the pulsed field gel electrophoresis technique has a relatively poor accuracy). In addition, Ohi et al. assigned 13 *P. pastoris* genes to the different chromosomes.

[0188] During the course of this invention, chromosome assembly was completed according to the strategy shown in **Fig. 1A**. 454/Roche sequencing[16] (GS-FLX version) was utilized to highly oversample the genome (20x coverage) and generated 70,500 paired-end sequence tags, to enable the assembly of all but 7 contigs into only 9 "supercontigs" (plus the mitochondrial genome) using automated shotgun assembly and BLASTN-based contig end joining (see Example 1 and **Fig. 4**). Upon assigning these supercontigs to the 4 chromosomes (see Example 1 and **Figs. 5A-5C**), the order of the supercontigs was determined through PCR and Sanger sequencing of the amplificates. Through these finishing experiments, the 4 chromosomal sequences were reconstructed (**Fig. 1B**; results summarized in **Table 2**), with only 2 gaps remaining (one on chromosomes 1 and 4 each). A ribosomal DNA (rDNA) repeat sequence was present in the assembly as a separate contig of 7450 bp, with exceptionally high coverage (328.8 fold). Given that sequence coverage all over the assembly very closely approximates 20x, it has been interpreted that there are approximately 16 copies of the rDNA repeat region, thus accounting for about 119 kbp in sequence. These rDNA loci were detected on all chromosomes (Example 1, **Fig. 1B** and **Fig. 5A**). The rDNA locus contains the 18S, 5.8S and 26S rRNA coding sequences. Unlike the *S. cerevisiae* 5S rRNA,

which is present in the repeated rDNA locus, the 21 copies of the *P. pastoris* 5S ribosomal RNA (rRNA) were found to spread across the chromosomes. While the chromosomes of *P. pastoris* GS115 were estimated to be (from chromosome 1 to 4): 2.9, 2.6, 2.3 and 1.9 Mbp based on pulsed-field gel electrophoresis (PFGE)[38], it was estimated after assembly in this invention to be 2.88 (2.8 + 0.08), 2.39, 2.24 and 1.8 (1.78+0.017) Mbp. Including the estimated 0.12 Mbp of rRNA repeats, the genome size of *P. pastoris* was determined to be 9.43 Mbp. The *P. pastoris* genomic sequences are deposited in Genbank.

Genome sequence accuracy estimation

[0189] A concern with genome sequences largely generated through 454 sequencing was the potential for "indel errors" at homopolymeric sequences[39]. An analysis of the occurrence of such sequences in the *P. pastoris* genome was conducted. Two approaches were followed to estimate the accuracy of the genome sequence. First, 39 peer-reviewed Genbank coding sequences of *P. pastoris* strain GS115 were retrieved (**Table 3**; a total sequence length of 70,295 bp). These sequences were compared to the genome sequence of this invention, and 84 differences were encountered. To determine which sequence was correct, PCR was performed on GS115 genomic DNA and the amplificates were Sanger-sequenced. In all but 2 cases, the Sanger sequences confirmed the genome sequence of this invention, and thus the error rate was estimated to be 1 in 35,147 bp. In an alternative approach, all open reading frames encoding proteins with at least one clear homolog in the databases were analyzed. If an interrupted ORF was found to have clear homology to the 5' part of the homologs, immediately followed by a coding sequence with clear homology to the 3' part, the most logical interpretation would be that there was a frame-shift error mutation in the genome sequence (i.e. both coding sequences are extremely likely to be linked into one open reading frame (ORF)). On this premise, frameshift errors were found in 2.7% (106) of the 3,997 genes for which such analysis could be made, totaling 6.11 Mbp of coding sequence. Assuming (fairly conservatively) that such error would have been detected if it has occurred in the first 2/3 of the ORF, a frame-shift error rate in the coding sequences was determined to be 1 in 37,716 bp. Both of the approaches above show that high-coverage 454 sequencing indeed yielded highly accurate genome sequences.

Pichia pastoris phylogenetic position

[0190] Phylogenetic analysis (**Fig. 1C**; Example 1) shows that *P. pastoris* diverged before the formation of the CTG clade (yeasts which translate the CUG codon into serine instead of leucine[13]).

Genome sequence annotation: protein-coding genes

[0191] Protein-coding genes were automatically predicted using the EuGène[18] prediction platform (Example 1) and these gene models were manually curated for functional annotation, accurate translational start and stop assignment, and intron location. This resulted in a 5,313 protein-coding gene set of which 3,997 (75.2%) were found to have at least one homolog in the National Center for Biotechnology Information (NCBI) protein database (BLASTP e-value 1e-5, sequence length ≤20 % difference and sequence similarity ≥50 %). The protein-coding genes were found to occupy 80% of the genome sequence. According to recently proposed measures for genome completeness, the genome was searched for highly conserved single (or low) copy gene sets: CEG with 248 genes across six model organisms[15], and FUNYBASE[16] with 246 genes with orthologs in 21 fungi. All genes from both gene-sets were found to be present in our proteome with full domain coverage.

[0192] In accordance with this invention, 1,285 genes were assigned to the Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathways[40], and 4,262 of genes were annotated with Gene Ontology (GO) terms[25,41]. The GO slim categories of *P. pastoris* are presented in **Fig. 6**. A secretion signal peptide was predicted in 9% of the genes[42] and 4,274 of proteins were predicted to contain InterPro domains[23]. These include 2,320 distinct Pfam domains. In comparing the presence and absence of protein domains with five other yeasts proteomes, 32 domains in 32 genes were identified as specific to *P. pastoris*. The two fungi in the CTG clade of which the genomes have been sequenced (*P. stipitis* and *C. lusitaniae*) share 71 gene families which were found to be absent in *P. pastoris*.

[0193] Codon (pair) optimization of transgenes to the expression host organism often yields substantial improvements in recombinant protein yield[43]. *P. pastoris'* codon usage is shown

in **Fig. 3A**. Overall, the codon usage is similar to the one for *S. cerevisiae* (the same codons being preferred by both organisms for all amino acids). Some synonymous codon pairs are also more or less frequently used than expected (the "codon pair bias"). As previously reported for *S. cerevisiae*[44], underrepresented and overrepresented codon pair clusters were observed.

Genome sequence annotation: tRNA genes

[0194] tRNA coding genes were automatically predicted and manually confirmed by BLASTN with *S. cerevisiae* homologs, which identified 123 nuclear tRNA genes (**Table 4**), compared to 274 in the *S. cerevisiae* genome[45]. *P. pastoris* has three tRNA families not present in *S. cerevisiae* (tR(UCG), tL(CAG) and tP(CGG)), but also lacks one tRNA family (tL(GAG)).

[0195] Interestingly, a positive correlation was found between the number of tRNA genes for a given codon and the frequency of use of this codon (Spearman rho=0.88; P<0.0001, **Fig. 3C**).

EXAMPLE-3

[0196] This Example describes experiments conducted to test whether overexpression in engineered Glycoswitch *Pichia pastoris* an endogenous UDP-GlcNAc transporter, UDP-Gal transporter and UDP-Glc-4-epimerase can increase glycan conversion efficiency of these strains, thus resulting in more uniform glycosylation patterns.

Results

[0197] Expression vectors for these three genes as described in the experimental procedures below. In these vectors, the transporter or epimerase gene was placed under control of the AOX1 promoter. Selection of these vectors was based on G418 (for UDP-GlcNAc transporter) or hygromycin (for UDP-Gal transporter and UDP-Glc-4-epimerase).

[0198] These vector were transformed into GSGnM5 (UDP-GlcNAc transporter) or GSGalGnM5 (UDP-Gal transporter and UDP-Glc-4-epimerase). Clones for all three strains were methanol induced and relative contribution of the glycans was assessed (**Table 7**). Similar results were obtained using clones from a separate transformation.

[0199] It is possible that overexpression of these genes can have a beneficial influence on the glycan conversion efficiency in fermentation conditions.

Experimental procedures:

[0200] Strains and growth conditions - *Escherichia coli* MC1061 was used as the host strain for all molecular experiments. *E. coli* was cultivated in LB medium containing the appropriate antibiotics for selection. The *Pichia pastoris* GlycoSwitch GnMan5 and GalGnMan5 strains expressing recombinant human alfa-antitrypsin were used in this analysis. TheGnMan5 strain is a GS115 (Invitrogen, Carlsbad, CA, USA)–derived strain that modifies its glycoproteins predominantly with GlcNAcMan5GlcNAc2 N-glycans due to inactivation of Och1 and overexpression of an ER-localized *T. reesei* α-1,2-mannosidase and a Golgi-localized human GlcNAc transferase. The GalGnM5 strain was derived from the GnM5 strain by transformation of the latter with a Golgi-localized human Galactosyl transferase fused to the *S. cerevisiae* UDP-Glc-4-epimerase Gal10. This strain predominantly modifies its glycoproteins with GalGlcNAcMan5GlcNAc2 N-glycans. *Pichia pastoris* cultures were grown in YPD medium, minimal medium or BMGY and induced in BMMY medium. All *Pichia pastoris* media were prepared as described in the *Pichia* instruction manual (Invitrogen, Carlsbad, CA, USA).

[0201] Vector construction – All molecular experiments were carried out according to standard procedures (Sambrook et al., 1989). Phusion polymerase was used in PCR reactions according to manufacturer's instructions (Finnzymes, Espoo, Finland). To construct the expression vectors for the activated sugar transporters (UDP-GlcNAc transporter and UDP-Gal transporter) and the UDP-Glucose-4-epimerase, the genes were amplified by PCR on genomic DNA. The primers for the PCR reaction were designed to amplify the complete gene with the stop codon and to incorporate a *Sal*I site downstream of UDP-GlcNAcT and a

*Bst*BI restriction site upstream and a *Not*I site downstream of the UDP-GalT and UDP-Glc-4-epimerase genes to allow subcloning. After amplification, the genes were cloned into pCR®4Blunt-TOPO® (Invitrogen, Carlsbad, CA, USA) and sequenced. After *Bst*BI/*Not*I digestion the UDP-GalT and UDP- Glc-4-epimerase genes were ligated into the pPICHygMnn2DmManII vector in which the Mnn2-ManII fusiongene was removed by *Bst*BI/*Not*I digest. Similarly, UDP-GlcNAcT was cut from the TOPO clone by *Bst*BI/*Sal*I digest and ligated into the pPICKanMnn2SpGal10hGalT vector in which the Mnn2-Gal10-GalT fusiongene was removed by *Bst*BI/*Sal*I digest. After selection of good clones, the expression vectors are opened by *Pme*I or *Sal*I digest for transformation of *Pichia pastoris* strains. Insertion of the vector in the *Pichia pastoris* genome of *Pme*I opened vectors were expected to be targeted to the AOX1 promoter locus while *Sal*I opened vectors were expected to insert in the gene of interest.

[0202] Yeast treatments – Cells were grown in 5 ml BMGY medium at 30°C. After 48h of incubation, the medium of the cultures was replaced by BMMY. The induction was performed for 48h by spiking the cultures twice a day with 1% methanol. The cultures were harvested by centrifugation for 5 minutes at 3000g. Medium and cell pellet were frozen at -20°C.

[0203] Glycosylation analysis – N-linked glycans were analysed using DNA sequencer-assisted fluorophore-assisted carbohydrate electrophoresis (DSA-FACE) as described previously (Laroy et al., 2006).

EXAMPLE-4

[0204] This Example describes experiments conducted to test the effect of overexpression of chaperone genes identified hereinabove on yield of heterologous proteins.

Results

[0205] Of the endogenous chaperone genes in the genome of *Pichia pastoris*, three ER-resident chaperones were selected for testing: ROT1, SHR3 and SIL1. Expression vectors for

these genes were generated as described in the experimental procedures. In these vectors, the chaperone genes are under control of the GAP promoter.

[0206] For transformation of these vectors into the GlycoSwitch Man5 strain, two approaches were used. In the first approach, the construct were targeted to the GAP promoter by linearization of the vectors with *Avr*II. In the second approach, random integration was achieved by linearization of the vectors with *EcoR*V. Transformants were selected on medium containing 100 µg/ml nourseothricin. Clones for all three chaperone constructs, in both transformations, were grown and intra-cellular proteins were extracted and run on SDS-PAGE gel. For detection of chaperone expression, a Western blot analysis was performed using an antibody against the His-tag. Expression was shown for ROT1 (31 kDa) and SHR3 (28 kDa). No expression could be observed for SIL1 (44 kDa).

[0207] To analyse the effect of the chaperones on heterologous expression and secretion of IL10 and IFN-β, two approaches were followed. In the first approach, the GSM5 strains expressing the chaperone were transformed with an IL10- or IFNβ-expression plasmid. In the second approach, GSM5 strains expressing IL10 or IFN-b were transformed with expression vectors for ROT1 or SHR3. ROT1- or SHR3-expressing GSM5 strains were transformed with pPIC92mIL10, linearized with *Stu*I or with pPIC9MFhIFNb2, linearized with *BstB*I. Individual clones were isolated on minimal medium and used for expression experiments. Secretion of IL10 or IFN-β in the medium upon methanol induction was assessed by SDS-PAGE and Coomassie Brilliant Blue staining.

[0208] IL10 or IFN-β expressing GSM5 strains were transformed with pGAPNORPpROT1 or with pGAPNORPpSHR3, both linearized with *EcoR*V. Individual clones were selected on nourseothricine plates and used for induction experiments.

[0209] Upon overexpression of ROT1 or SHR3, no significant enhancement of the secretion of IL10 or IFN-β was observed, in comparison with control clones expressing IL10 or IFN-β.

[0210] In the ROT1-expressing strains, somewhat more of a 31 kDa protein was detected in the medium. To test whether some of the tagged ROT1 proteins were secreted upon

overexpression, Western blot-immunodetection was performed using anti-His antibodies. In both the media of the control strain and the ROT1 or SHR3- expressing strains, a faint band of approximately 31 kDa could be detected. Moreover, upon deglycosylation with PNGaseF, part of this band shifted to a faster migrating band. It was therefore concluded that this band is not related to ROT1.

[0211] Generation of expression vectors for the *P. pastoris* homologues of LHS1, CNE1 and EPS1 was also underway. Upon PCR amplification of the chaperone genes on *P. pastoris* genomic DNA, the PCR product for EPS1 (1945 bp) was cut with *EcoR*I and *Not*I and ligated into the vector fragment of pGAPNORPpROT1, cut with the same enzymes and transformed to *E. coli*. Because the genes for CNE1 and LHS1 contain *EcoR*I sites, the CNE1 (1753 bp)- and LHS1 (2700 bp)-PCR products were fused by PCR with the GAP promoter DNA of pGAPNORPpROT1 (500 bp). The resulting PCR-DNA was then cut with *Not*I and *Nsi*I (CNE1) or with NotI and AvrII (LHS1) and ligated with the vector fragment cut with the same enzymes and transformed to *E. coli*. Clones were checked by colony-PCR for the presence of chaperone containing plasmids. Once completed, the coding sequences for these chaperones would be fused with a C-terminal myc/His6 tag and placed under control of the GAP promoter. The resulting plasmids would also confer resistance to nourseothricine.

Experimental procedures

[0212] Strains and growth conditions - *Escherichia coli* MC1061 was used as the host strain for all molecular experiments. *E. coli* was cultivated in LB medium containing the appropriate antibiotics for selection. The *Pichia pastoris* GlycoSwitch GnMan5 and GalGnMan5 strains expressing recombinant human alfa-antitrypsin were used in this analysis. TheGnMan5 strain is a GS115 (Invitrogen, Carlsbad, CA, USA) –derived strain that modifies its glycoproteins predominantly with GlcNAcMan5GlcNAc2 N-glycans due to inactivation of Och1 and overexpression of an ER-localized *T. reesei* α-1,2-mannosidase and a Golgi-localised human GlcNAc transferase. The GalGnM5 strain was derived from the GnM5 strain by transformation of the latter with a Golgi-localized human Galactosyl transferase fused to the *S. cerevisiae* UDP-Glc-4-epimerase Gal10. This strain predominantly modifies

its glycoproteins with GalGlcNAcMan5GlcNAc2 N-glycans. *Pichia pastoris* cultures were grown in YPD medium, minimal medium or BMGY and induced in BMMY medium. All *Pichia pastoris* media were prepared as described in the Pichia instruction manual (Invitrogen, Carlsbad, CA, USA).

[0213] Vector construction – All molecular experiments were carried out according to standard procedures (Sambrook et al., 1989). Phusion polymerase was used in PCR reactions according to manufacturer's instructions (Finnzymes, Espoo, Finland). To construct the expression vectors for the chaperone genes *Sil1*, *Rot1* and *Shr3*, the chaperone genes were amplified by PCR on genomic DNA. The primers for the PCR reaction were designed to amplify the complete gene without the stop codon and to incorporate an *Eco*RI restriction site upstream and a *Not*I site downstream of the gene to allow subcloning. Ater amplification, the genes were cloned into pCR®4Blunt-TOPO® (Invitrogen, Carlsbad, CA, USA) and sequenced. After *Eco*RI/*Not*I digestion the chaperon genes were inserted into the pGAPNOURCre vector in which the Cre recombinase gene was removed by *Eco*RI/*Not*I digest. This created a transcriptional fusion of the chaperon gene with a myc and His6 tag. The pGAPNOURCre vector was generated by deletion of the *Pichia* autosomal replication sequence (PARS) from the pGAPNorCre1PARS1 vector by *Nsi*I digest and self ligation. After selection of good clones, the chaperone expression vectors were opened by *Avr*II or *Eco*RV digest for transformation of *Pichia pastoris* strains. Insertion of the vector in the *Pichia pastoris* genome of *Avr*II opened vectors was targeted to the Gap promoter locus while *Eco*RV opened vectors was inserted randomly.

[0214] Yeast treatments – Cells were grown in 5 ml BMGY medium at 30°C. After 48h of incubation, the medium of the cultures was replaced by BMMY. The induction was performed for 48h by spiking the cultures twice a day with 1% methanol. The cultures were harvested by centrifugation for 5 minutes at 3000g. Medium and cell pellet were frozen at -20°C.

EXAMPLE-5

[0215] A set of 54 *P. pastoris* genes was identified that contain a signal sequence, as discussed hereinabove. The expression of these genes in *P. pastoris* was assessed using microarray data published by Graf et al. (2008)[48], the presence of these proteins in *P. pastoris* grown in glucose containing medium (Mattanovich et al., 2009)[50] and for the protein abundance of their homologues in *S. cerevisiae* (Brockmann et al., 2007[46]; Ghaemmaghami et al., 2003[47]; Liu et al., 2004[49]; Newman et al., 2006[51]) (**Table 8**).

[0216] Based on these data ten genes (13 *P. pastoris* sequences) were selected for further analysis. The efficiency of these signal sequences is assessed in comparison with the prepro sequence of α-mating factor (α-MF) to drive the secretion of N-terminally tagged human IL10 upon methanol induction. The human IL10 expression plasmid pKai61EA-hIL10 is used. In this plasmid the mature hIL10 sequence is preceded by the prepro sequence from α-MF, followed by a His6-tag and a DEVD cleavage site. The His6-tag facilitates the purification of the acid-labile IL10 protein and the DEVD cleavage site makes it possible to remove this tag by incubation with purified caspase-3.

[0217] To facilitate the exchange of the α-MF prepro sequence with the signal sequences from the candidate genes, 13 synthetic genes are obtained that can be cut out using *BstB*I and *Kpn*I and that contain the signal sequence of the candidate gene followed by His6-tag, DEVD-site and part of the mature hIL10.

[0218] In a first attempt the *BstB*I-*Kpn*I fragment of the synthetic genes was PCR amplified with specific primers and subsequently cut with BstBI and KpnI and ligated into the vector fragment of pKai61EA-hIL10, cut with the same enzymes. The ligation mixture was then used for transformation of *E. coli*. Individual clones selected on zeocine containing plates were checked by colony PCR. Candidate clones were already identified for chr1-4_0611 and chr1-4_0426.

EXAMPLE-6

[0219] The initializing step in yeast O-glycosylation is known to be catalysed by a family of protein mannosyltransferases (PMT). As described hereinabove, in the *Pichia pastoris* genome, 5 orthologs of the PMT genes were annotated, with representatives in the 3 subfamilies. In *S. cerevisiae*, deletion of only one of these genes was found to be insufficient to abolish O-glycosylation. The double and triple knockouts resulted in a loss of O-glycosylation but showed a severe defect in growth. In this Example, two approaches are described for generating PMT deficient *Pichia pastoris* strains which could result in an O-glycosylation deficient strain.

[0220] In one approach, disruptions of the PMT ORFs are made through the use of a knock-in vector by single homologous recombination. In the second approach, PMT knock-outs are made by double homologous recombination.

Experimental procedures

[0221] Strains and growth conditions - *Escherichia coli* competent strains are used as the host strain for all molecular experiments. *E. coli* is cultivated in LB medium containing the appropriate antibiotics for selection. The *Pichia pastoris* GS115 strain (Invitrogen, Carlsbad, CA, USA) and glycoengineered strains are used in this analysis, which can express a protein of interest. *Pichia pastoris* cultures are grown in YPD medium, minimal medium or BMGY and induced in BMMY medium. All *Pichia pastoris* media are prepared as described in the *Pichia* instruction manual (Invitrogen, Carlsbad, CA, USA).

[0222] Vector construction – All molecular experiments are carried out according to standard procedures (Sambrook et al., 1989). Phusion polymerase is used in PCR reactions according to manufacturer's instructions (Finnzymes, Espoo, Finland).

[0223] To construct knock-in vectors for the *PMT* genes, a fragment of these genes is amplified by PCR. The primers for the PCR reaction are designed to amplify a fragment that

has low similarity to other regions in the genome, is not the full length gene, and is located so the protein fragment preceding the fragment is not functional. In addition, the primers contain restriction sites for subcloning of the fragment. In this fragment a restriction site, unique in the final vector, is incorporated to allow later linearization. After amplification, the genes are cloned into pCR®4Blunt-TOPO® (Invitrogen, Carlsbad, CA, USA) and sequenced. This fragment is cloned into a *Pichia* vector containing a selectable marker. After selection of good clones, the *PMT* knock-in vectors are opened by restriction digest using the unique site in the PMT fragment for transformation of *Pichia pastoris* strains. Insertion of the vectors in the Pichia pastoris genome will be targeted to the respective PMT loci.

[0224] To construct knock-out vectors for the *PMT* genes, two fragments of the PMT loci are amplified by PCR, cloned into the pCR®4Blunt-TOPO® vector (Invitrogen, Carlsbad, CA, USA) and sequenced. The primers for the PCR reaction are designed to amplify fragments within the PMT ORF, promoter or 5'UTR in a way that when these two fragments recombine with the PMT allele it will create an inactive allele. The primers contain restriction sites for subcloning of the fragments and a unique restriction site upstream of the 5' fragment and downstream of the 3' fragment. These fragments will be cloned into a *P. pastoris* vector up and downstream of a selectable marker. After selection of good clones, the *PMT* knock-out vectors are cut by restriction digest using the unique site incorporated into the PMT fragment, after which the excised fragment will be used for transformation of *Pichia pastoris* strains.

EXAMPLE-7

[0225] The stability of certain proteins expressed in *Pichia pastoris* has been observed to be influenced by the action of proteases such as protease A and B. As described above, in the *Pichia pastoris* genome, a series of orthologs of novel protease genes were identified and annotated, with representatives in the serine protease, aspartyl protease and cysteine protease subfamilies. One strategy is described in this example to generate a series of strains of *Pichia pastoris* deficient in one or more of the identified protease activities. This strategy can be applied to any strain of *Pichia pastoris* that expresses a heterologous protein to compare the stability of that heterologous protein with or without the particular protease being active. One

application of this strategy is to take a strain that expresses a protein of interest and use a set or "kit" of the insertional inactivation ("knock-in") vectors to generate a series of derivative strains that each lack activity of one of the endogenous proteases of *Pichia.*

[0226] To generate protease deficient *P. pastoris* strains, two we describe two approaches, one generates a disruption of the individual protease ORFs through knock-in of the vector by single homologous recombination. The second approach generates true protease knock-outs by double homologous recombination.

Experimental procedures

[0227] Strains and growth conditions – *Escherichia coli* competent strains are used as the host strain for all molecular experiments. *E. coli* is cultivated in LB medium containing the appropriate antibiotics for selection. The *Pichia pastoris* GS115 strain (Invitrogen, Carlsbad, CA, USA) or glycoengineered strains are used in this analysis, these strains can express a protein of interest. *Pichia pastoris* cultures are grown in YPD medium, minimal medium or BMGY and induced in BMMY medium. All *Pichia pastoris* media are prepared as described in the *Pichia* instruction manual (Invitrogen, Carlsbad, CA, USA).

[0228] Vector construction – All molecular experiments are carried out according to standard procedures (Sambrook et al., 1989). Phusion polymerase is used in PCR reactions according to manufacturer's instructions (Finnzymes, Espoo, Finland). To construct the knock-in vectors for the protease genes, a fragment of the corresponding gene is amplified by PCR. The primers for the PCR reaction are designed to amplify a fragment that has low similarity to other regions in the genome, is not the full length gene, and is located so the protein fragment preceding the fragment is not functional. This is readily achieved by using segments of the gene that lie within the highly conserved DNA/protein motifs that correspond to essential amino acid residues in the active site of that particular class of protease. In addition, the primers contain restriction sites for subcloning of the fragment. In this fragment a restriction site, unique in the final vector, is incorporated to allow later linearization. After amplification, the genes are cloned into pCR®4Blunt-TOPO® (Invitrogen, Carlsbad, CA, USA) and

sequenced. This fragment is cloned into a *Pichia* vector containing a selectable marker. After selection of good clones, the protease knock-in vectors are opened by restriction digest using the unique site in the protease fragment for transformation of *Pichia pastoris* strains. Insertion of the vectors in the *Pichia pastoris* genome are targeted to the respective protease gene loci. **Figure 8** shows the DNA sequences for one strategy of protease inactivation by knock-in.

[0229] To construct the knock-out vectors for the protease genes, two fragments of the protease loci are amplified by PCR, cloned into the pCR®4Blunt-TOPO® vector (Invitrogen, Carlsbad, CA, USA) and sequenced. The primers for the PCR reaction are designed to amplify fragments within the protease ORF, promoter or 5'UTR in such a way that when these two fragments recombine with the protease allele it creates an inactive allele. The primers contain restriction sites for subcloning of the fragments and a unique restriction site upstream of the 5' fragment and downstream of the 3' fragment. These fragments are cloned into a *P. pastoris* vector up and downstream of a selectable marker. After selection of good clones, the protease knock-out vectors are cut by restriction digest using the unique site incorporated into the protease fragment, after which the excised fragment is used for transformation of *Pichia pastoris* strains. In this way, more stable gene knock-out constructs are generated.

References:

1. Hamilton, S.R. & Gerngross, T.U. Glycosylation engineering in yeast: the advent of fully humanized yeast. *Curr. Opin. Biotechnol.* **18**, 387-392 (2007).
2. Jacobs, P.P., Geysens, S., Vervecken, W., Contreras, R. & Callewaert, N. Engineering complex-type N-glycosylation in *Pichia pastoris* using GlycoSwitch technology. *Nat. Protoc.* **4**, 58-70 (2009).
3. Ratner, M. Pharma swept up in biogenerics gold rush. *Nat. Biotechnol.* **27**, 299-301 (2009).
4. Potgieter, T.I. et al. Production of monoclonal antibodies by glycoengineered *Pichia pastoris*. *J. Biotechnol.* **139**, 318-325 (2009).
5. Mogelsvang, S., Gomez-Ospina, N., Soderholm, J., Glick, B.S., Stachelin, L.A. Tomographic evidence for continuous turnover of Golgi cisternae in *Pichia pastoris*. *Mol. Biol. Cell* **14**, 2277-2291 (2003).
6. Hartner, F.S. et al. Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. *Nucleic Acids Res.* **36**, e76 (2008).
7. Vervecken, W. et al. *In vivo* synthesis of mammalian-like, hybrid-type N-glycans in *Pichia pastoris*. *Appl. Environ. Microbiol.* **70**, 2639-2646 (2004).
8. Bobrowicz, P. Engineering of an artificial glycosylation pathway blocked in core oligosaccharide assembly in the yeast *Pichia pastoris*: production of complex humanized glycoproteins with terminal galactose. *Glycobiology* **14**, 757-766 (2004).
9. Trimble, R.B. et al. Characterization of N- and O-linked glycosylation of recombinant human bile salt-stimulated lipase secreted by *Pichia pastoris*. *Glycobiology* **14**, 265-274 (2004).
10. Mille, C., et al. Identification of a new family of genes involved in beta-1,2-mannosylation of glycans in *Pichia pastoris* and *Candida albicans*. *J. Biol. Chem.* **283**, 9724-9736 (2008).
11. Dalle, F. et al. Beta-1,2- and alpha-1,2-linked oligomannosides mediate adherence of *Candida albicans* blastospores to human enterocytes in vitro. *Infect. Immun.* **71**, 7061-7068 (2003).
12. Nasab et al., *Mol Biol Cell.* 19:3758-68, 2008.
13. *Glycobiology* 12(12): 821-828, 2002.
14. *Pichia Protocols. Methods in Molecular Biology* 103, 27-39, 1998.
15. Higgins and Cregg, *Pichia Protocols. Methods in Molecular Biology* 103, 1-16, 1998.
16. Margulies, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380 (2005).
17. Pop, M., Kosack, D.S. & Salzberg, S.L. Hierarchical scaffolding with Bambus. *Genome Res.* 14, 149-159 (2004).
18. Foissac, S. et al. Genome Annotation in Plants and Fungi: EuGene as a Model Platform. *Current Bioinformatics* **3**, 89-97 (2008).
19. Wang, K., Ussery, D.W. & Brunak, S. Analysis and prediction of gene splice sites in four *Aspergillus* genomes. *Fungal Genet. Biol.* 46 Suppl 1, S14-18 (2009).
20. Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. & Borodovsky, M. Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. *Genome Res.*, 13 (2008).

21. Stanke, M. et al. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7, 62 (2006).
22. The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res.* 37, D169-174 (2009).
23. Pruitt, K.D., Tatusova, T. & Maglott, D.R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61-65 (2007).
24. Mulder, N.J. et al. New developments in the InterPro database. *Nucleic Acids Res.* 35, D224 (2007).
25. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25-29 (2000).
26. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797 (2004).
27. Parra, G., Bradnam, K., Ning, Z., Keane, T. & Korf, I. Assessing the gene space in draft genomes. *Nucleic Acids Res.* **37**, 289-297 (2009).
28. Marthey, S. et al. FUNYBASE: a FUNgal phYlogenomic dataBASE. *BMC Bioinformatics* **9**, 456 (2008).
29. Griffiths-Jones, S. et al. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121-124 (2005).
30. Lowe, T.M. & Eddy, S.R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955-964 (1997).
31. Pinheiro, M. et al. Statistical, computational and visualization methodologies to unveil gene primary structure features. *Methods Inf. Med.* **45**, 163-168 (2006).
32. Schmidt, H.A., Strimmer, K., Vingron, M. & von Haeseler, A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502-504 (2002).
33. Rossignol, T. et al. CandidaDB: a multi-genome database for *Candida* species and related Saccharomycotina. *Nucleic Acids Res.* **36**, D557-D561 (2007).
34. Jeffries, T. et al. Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*. *Nat. Biotechnol.* **25**, 319-326 (2007).
35. Enright, A.J., Van Dongen, S. & Ouzounis, C.A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575 (2002).
36. Finn, R. et al. The Pfam protein families database. *Nucleic Acids Res.* **36**, D281 (2008).
37. Felsenstein, J. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol.* **266**, 418-427 (1996).
38. Ohi, H., Okazaki, N., Uno, S., Miura, M. & Hiramatsu, R. Chromosomal DNA patterns and gene stability of *Pichia pastoris*. *Yeast* **14**, 895-903 (1998).
39. Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L. & Welch, D.M. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* **8**, R143 (2007).
40. Kanehisa, M. et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**, D480 (2008).
41. Schmid, R. & Blaxter, M. annot8r: GO, EC and KEGG annotation of EST datasets. *BMC Bioinformatics* **9**, 180 (2008).

42. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2**, 953-971 (2007).
43. Hu, S. et al. Codon optimization, expression, and characterization of an internalizing anti-ErbB2 single-chain antibody in *Pichia pastoris*. *Protein Expr. Purif.* **47**, 249-257 (2006).
44. Friberg, M., von Rohr, P. & Gonnet, G. Limitations of codon adaptation index and other coding DNA-based features for prediction of protein expression in *Saccharomyces cerevisiae*. *Yeast* **21**, 1083-1093 (2004).
45. Hani, J. & Feldmann, H. tRNA genes and retroelements in the yeast genome. *Nucleic Acids Res.* **26**, 689-696 (1998).
46. Brockmann et al. (2007) *PLoS Computational Biology* 3, e57.
47. Ghaemmaghami et al. (2003) *Nature* 425, 737.
48. Graf et al. (2008) *BMC Genomics* 9, 390.
49. Liu et al. (2004) *Anal Chem* 76, 4193.
50. Mattanovich et al. (2009) *Microbial Cell Factories* 8, 29.
51. Newman et al. (2006) *Nature* 441, 840.

Table 1: Primers used in the genome assemblyprocess.

5'	Primers used for sequencing	3'
13	CCTATTCTCTCTTATCCCGACG	
16	GAAGCTTTATCGGCAACATC	
29	GGAAGATGCAGAGGATGTTAGG	
30	AGGTTATACTGTGTGTAGTACAGATGG	
33	CATCTATAACGAGTCAAAGGATGA	
35	TGGAGTGAATGGAAGTTCAGAG	
36	GATTAGCTAAGTTGCCCTATTAGG	
37	AAATTGTCGTTCTCACAACAGG	
38	CATCTTAAGAATACATCATATGCACC	
39	CCATGTTGCACTCACACTTGG	
40	CGTTGAGTAGCTTTCTCCTGTCC	
41	TATCACTCCATATGGCTCTTGG	
42	GTAGAGTAGCTGACCTAAAGTCTTGG	
43	AAGCCGTAGAGGAGTGTAAACG	
48	GAACTTCAACTGTCATTATTGATGG	
49	CAGGCATCTTCAGAGGGTACTC	
50	TTCCTGTGGCAGATATCAGCT	
51	GAGAGTACTAACGATTCTTCACTCACT	
52	TTACCAATCACGTAGTTCGCTC	
53	GACTTCTAACGGGTGAGATATCC	
54	TAGGTCAGTGCTTGATTGTAATCTC	
55	AATCGACAGGGATTGCTCAA	
56	ACACAAGGAGTTAGGGTACTAAGC	
57	CCACTTTGAGTTCCTTGATCTCAC	
58	CATTGACACTCCTTACAATGTGC	
63	TTTGCGAGTTCATGTTTGAAGAGG	
64	GAGTGGTATCAGAAACTTCAGATG	
65	GACCCTGGTAGTCTCAATCTAAG	
66	CTCACCATATAAGATTTATTCTGATAAAGC	
67	GCTCAACTTTCTGAACTTGCTCA	
68	CATCAAAGTCCAGTAAGCAACCC	
69	GGAGCTGTGTTATTTACTATTGAGG	
70	ATGATTTCGAAAGGTTCTACCCC	
71	TACTAACTTTGAGTTTAGAGGATGG	
72	AGGGAAGTCTAAGTTCTTTCGC	
73	GATGATAACGGTGCCTGGTTCA	
74	GTATACCAGTTTCCCAACAATCTTAG	
75	GGGGTTGGAATTTGCTCGATAA	
76	TACAGATGAATCTGGGTCGGTG	
77	CTGGTACTTACGAAACCACGTA	
78	TTTCGTAAGTACCAGTCCATGG	
79	GTGGTAGCTGGGGTGGTT	
80	TCTGTCTAATGCTTACCATCAGC	
81	CCCTCTGTCGATCTCAGGTTT	
82	CTGAACAGCATTGAAGATAAAAAAACG	
83	CAGAAGTTTATGGCTGATATGATGC	
84	GATAGTGGTGTTGGGAGGATATG	
85	AGAATGATTTGGGAGATACGAGC	
86	GCTGCCCTTTGAGTATCAGATC	
87	GCTGGCCAATCTCATCTCATG	
88	GGAAGGATTTGGGCTAATAAACTG	
115	CAACTATCCACTTGCTCATAACAGAC	
116	CCTACAGGGTTTCCATGTGATG	
117	GGTAAGATGAACTAATTCACAAGGTG	

5'	Primers used for probes	3'
Chr1	ATCGACTGAGCTTCAAGTTAATGTGC	
	TTCAAGCAGTCTCCATCACAATCC	
Chr2	ACTCTGGTGGAGTAACCGTACTCG	
	ATTGTCTCCAATGCTTCTTGTACTACC	
Chr3	CTACGATTATGATGTCAGTGCCAGTG	
	GTGAGAACAACTAAAGAATTATTGGAGC	
Chr4	TAGCACGGTAGATCTTGTGACTTGG	
	GTAGAAGAGCCATTGTTCCATGTG	
FragB	GCTTGCAGAAGATGATTACAAACTTGC	
	GACTGAACTTGTTGGATGTCTAAGGAG	
FragD	TCTATACAATGTCCAAGTGGACCAAG	
	GTTTCTTCGACGTTTCGATGATCC	
C2	AGCACGACGGAGTCTAAGATTCC	
	GTCAGAGGTGAAATTCTTGGATCG	
C34	TATGTCTTCTCCTGTTGTTCCTACCTCC	
	CGGAATCAATGATTAGCTCATTGG	
C121	GAATTCAGATGCTGATTTACTTGCAC	
	TGGATAAGCAATGGAGGAAGTTCTG	
C131	CATTAGTAGGAGCTTCCAAGGTACTGT	
	AGTCCGTGGCAATAGAGCGATTG	
C157	TCACCAAGAAAGCTCTTATTTGGAGAG	
	ATTTCAGTCTCGCATTGTCCGACT	
C159	CAAAGAAGCTAGCTGTGGATGGAG	
	AACTGTAGGAACTCCATTCATCGGTG	

5'	Primers used for fragment joining	3'
1	CGTTTAACGTCGATAGGCTAG	
2	CAGCAGTCAGACCCTTAATGTC	
4	GCAGAAAAGTTACCAAGTGAGC	
5	TCGTGCCATGGTAGTATTCCT	
6	TCTCAAACTGTTCAAAGACTTGTAC	
8	TAATGAAGGACAAATAATTTCAGC	
9	TGTTCACATAGCCAACTGATACTG	
62	GGCCGGGAATGCTATTTGGT	
60	TTTGGCATTATGCGAAAAGAATAGGAC	
12	GCAAGCTATTGGAAACAACC	
15	ATTTTGATAGGAAGCTTAGATTGC	
17	CCAACTCGTGACAACAACCTC	
21	GCCACCAACTTCAGTAGTCCAC	
23	CCTTAGAAGTCGTTATATGGAAGTG	
25	ATACGCATGTGGAGATTCCTC	
26	CTCTCATCCAACCAGTTGACC	

Table 2

A. Genome sequencing and assembly statistics

454 Sequencing

Sequenced Reads	Sequenced Length (bp)	Paired-end reads
897,197	218,602,026	11,538

MIRA assembly

Assembled reads	Assembled Contigs	Contigs (> 500 bp)	Length (bp)	N50	L50 (Kbp)	Average coverage
885,659	1,154	230	9,658,092	40	77	20

Contig joning			**Chromosomes**
Joined contigs	Supercontigs	Length (Mbp)	
203	10	9.3	4

N50: number of contigs that collectively cover at least 50% of the assembly.

L50: length of the shortest contig among those that collectively cover 50% of the assembly

B. Genome Contents Overview

General information

Size (Mbp): 9.3 (not including rDNA loci, estimated at 0.12 Mbp)

Genome GC content (%): 41.1

Assembled chromosomes: 4

Coding genes

Coding genes: 5,313

Coding %: 79.6

Coding GC (%) 41.6

Mean gene length (bp) 1,442

Single exon genes: 4,680

RNA genes

tRNA genes: 123

5s rRNA genes: 21

Mitochondrial genome

Size (bp) : 36,119

Genome GC content (%): 22

Coding genes: 16

tRNA genes: 31

Table 3. Resequencing of selected known *P. Pastoris* ORFs.
Comparisonof sequence differences between our 454 sequence and the ORF of genes present in Genbank. The genes containing sequence difference and our 454 sequence are given in the last column.

	Gene Name	Locus id	CDS (bp)	Differences Genbank	Differences Sanger
1	Cyc	chr4_0018	333	0	
2	Delta 15-fatty acid desaturase	chr4_0743	1248	5	0
3	YPS2	chr3_0299	1584	0	
4	PFK3	chr4_0943	1056	1	0
5	FET3	chr2-1_0787	1890	5	0
6	EF-1	FragB_0052	1380	5	0
7	GAS1	chr1-3_0226	1617	0	
8	PMR1	chr1-4_0325	2775	0	
9	Methionine synthase	chr2-1_0160	2307	10	0
10	Sphingolipid C9-methyltransferase	chr4_0465	1471	0	
11	DES	chr3_0939	1083	2	1
12	SLD1	chr1-1_0013	1629	0	
13	VAC8	chr1-4_0101	1671	0	
14	FTR1	chr1-4_0040	1098	0	
15	KEX2	chr2-1_0304	2334	0	
16	YPS1	chr4_0584	1800	2	0
17	PRC1	chr1-4_0013	1572	0	
18	PDI	chr4_0844	1554	2	1
19	YPT-like	chr1-4_0627	919	0	
20	SEC4	chr3_0143	615	0	
21	FAD2	chr4_0052	1263	1	0
22	ALG3	chr4_0712	1398	6	0
23	EF-2	chr2-1_0812	2605	0	
24	PNO1	chr1-4_0410	2334	0	
25	GSA10	chr2-1_0641	2391	1	0
26	CBS	chr2-2_0137	1506	2	0
27	ceramide glucosyltransferase	chr3_0357	1530	0	
28	GDI1	chr3_0531	1344	0	
29	GSA9	chr1-4_0555	3941	0	
30	SEC12	chr4_0606	3117	0	
31	SAR1	chr1-1_0180	648	0	
32	SEC18	chr3_0342	2277	7	0
33	ACT1	chr3_1169	1728	0	
34	UGT51B1	chr4_0167	3636	0	
35	GSA12	chr3_0931	1632	1	0
36	GSA11	chr1-3_0168	5580	24	0
37	SEC17	chr2-1_0644	936	7	0
38	SEC13	chr1-3_0057	943	3	0
39	AOX2 promoter	chr4_0821	1550	0	
	Total length		**70295**		
	Total differences			**84**	**2**
		Accuracy			**1/35147 bp**

Table 4. Nuclear tRNA genes identified in the *P. pastoris* genome
Overview of *P. pastoris* tRNA genes

tRNA species	Anticodon	*P. pastoris*	intron	*S. cerevisiae*[1]	intron
tRNA- Ala	AGC	5		11	
tRNA- Ala	UGC	2		5	
tRNA- Arg	ACG	2	i	6	
tRNA- Arg	CCG	1		1	
tRNA- Arg	CCU	1		1	
tRNA- Arg	UCG	1		0	
tRNA- Arg	UCU	4		11	
tRNA- Ans	GUU	4		10	
tRNA- Asp	GUC	6		15	
tRNA- Cys	GCA	2		4	
tRNA- Gln	CUG	2		1	
tRNA- Gln	UUG	3		9	
tRNA- Glu	CUC	4		2	
tRNA- Glu	UUC	5		14	
tRNA- Gly	CCC	1		2	
tRNA- Gly	GCC	5		16	
tRNA- Gly	UCC	3		3	
tRNA- His	GUG	3		7	
tRNA- Ile	AAU	5		13	
tRNA- Ile	UAU	1	i	2	i
tRNA- Leu	AAG	2		0	
tRNA- Leu	CAA	3	i	10	i
tRNA- Leu	CAG	1/1	-/i	0	
tRNA- Leu	GAG	0		1	
tRNA- Leu	UAA	1		7	
tRNA- Leu	UAG	1	i	3	i
tRNA- Lys	CUU	5		14	
tRNA- Lys	UUU	3	i	7	i
tRNA- $Met_{initiator}$	CAU	2		5	
tRNA- $Met_{elongator}$	CAU	2	i	5	
tRNA- Phe	GAA	5	i	10	i
tRNA- Pro	AGG	2		2	
tRNA- Pro	CGG	1		0	
tRNA- Pro	UGG	4		10	i
tRNA- Ser	AGA	4		11	
tRNA- Ser	CGA	1		1	i
tRNA- Ser	GCU	2	i	4	i
tRNA- Ser	UGA	2	i	3	
tRNA- Thr	AGU	5		11	
tRNA- Thr	CGU	1		1	
tRNA- Thr	UGU	1		4	
tRNA- Trp	CCA	3	i	6	i
tRNA- Tyr	GUA	4	i	8	i
tRNA- Val	AAC	5		14	
tRNA- Val	UAC	1		2	
tRNA- Val	CAC	1		2	
total		123		274	
different tRNAs		45		42	

[1] Hani and Feldmann (1998) Nucleic acids Research, 26 (3) 689

Table 5A. Methanol pathway genes in *P. pastoris*
Overview of *P. pastoris* genes involved in methanol metabolism (shown in Fig. 3A)

Reference	Gene	EC code	Locus id
1	AOX	1.1.3.13	chr4_0152 chr4_0821
2	FLD	1.2.1.1	chr3_1028
3	FGH	3.1.2.12	chr3_0867
4	FDH	1.2.1.2	chr3_0932
5	CAT	1.11.1.6	chr2-2_0131
6	DAS	2.2.1.3	chr3_0832 chr3_0834
7	DAK	2.7.1.29	chr3_0841
8	TPI	5.3.1.1	chr3_0951
9	FBA	4.1.21.13	chr1-1_0072 chr1-1_0319
10	FBP	3.1.3.11	chr3_0868

Table 5B. Protein secretion pathway in *P. pastoris*

Overview of *P. pastoris* genes involved in secreted (glyco)protein post-translational processes (shown in Fig. 3B)

Reference	Process or complex	Gene	locus id
[1]	Sec61 complex	SEC61	chr1-3_0202
		SBH1	chr2-2_0210
		SSS1	chr1-1_0023
[2]	OST complex	STT3	chr1-4_0685
		"	chr1-4_0496
		SWP1	chr1-3_0248
		WBP1	chr2-1_0423
		OST1	chr3_0741
		OST2	chr2-2_0346
		OST3	chr4_0610
		OST4	chr2-2 62421-62516
		OST6	chr3_1142
[3]	Signal Peptidase complex	SPC1	chr1-1_0491
		SPC2	chr2-1_0589
		SPC3	chr4_0874
		SEC11	chr1-4_0187
[4]	Quality control		
	Chaperones	ROT1	fragB_0048
		LHS1	chr1-3_0063
		CNE1	chr2-1_0322
		YDJ1	chr2-2_0066
		EPS1	chr2-1_0421
		SHR3	chr1-3_0116
		KAR2	chr2-1_0140
		SIL1	chr1-1_0237
	Folding sensors	HTM1(MNL1)	chr3_0891
		UGGT	chr1-3_0114
		"	chr3_0929
[5]	Early N-glycan processing	GLS1	chr1-1_0215
		GLS2	chr2-1_0778
		MNS1	chr2-1_0753
		GTB1	chr3_0179
[6]	N-glycan precursor synthesis	SEC59	chr2-1_0498
		ALG7	chr2-1_0727
		ALG13	chr1-4_0448
		ALG14	chr3_0944
		ALG1	chr2-1_0759
		ALG2	c121_0002
		ALG12	chr4_0544
		ALG11	chr1-4_0417
		ALG3	chr4_0712
		ALG9	chr2-2_0036
		ALG6	chr2-1_0549

		ALG10	Chr1-4_0475
		ALG8	chr3_0999
7	O-Glycosylation	PMT1	chr2-1_0212
		PMT2	chr2-1_0256
		PMT3	"
		PMT4	chr1-4_0033
		PMT5	chr1-1_0286
		PMT6	chr4_0777
8	Nucleotide sugar synthesis		
	GDP-Mannose synthesis	PMI40	chr3_1115
		MPG1(PSA1)	chr3_0870
		"	chr2-1_0093
		ALG4	chr2-2_0053
	UDP-Glc synthesis	UGP1	chr1-3_0122
		(QRI1)	chr3_0676
	UDP-GlcNAc synthesis	GFA1	chr2-1_0626
		GNA1	chr4_0066
		PCM1	chr1-1_0067
		QRI1/UAP1	chr3_0676
	UDP-Gal synthesis	Gal10	chr4_0839
9	Nucleotide sugar transport		
	GDP-Mannose transport	VRG4	chr3_0916
		GDA1	chr4_0021
	UDP-GlcNAc transport	YEA4	chr1-3_0163
	UDP-Gal transport	HUT1	chr2-1_0692
		Pisti UDP-GalT	chr4_0810
		(VRG4)	chr3_0916
		(YEA4)	chr1-3_0163
10	Golgi N-glycan processing		
	Hyperglycosyl and core type	OCH1	chr1-3_0251
		MNN9	chr4_0103
		VAN1	chr2-1_0772
		MNN10	chr2-2_0185
		MNN11	chr2-2_0125
		HOC1	chr3_0620
		ANP1	chr3_0515
		MNN2	chr1-4_0037
		MNN5	chr3_0370
		MNN6	chr3_1162
		"	chr3_0215
		MNN4	chr1-4_0409
		"	chr2-1_0718
		"	chr2-1_0706
		PNO1	chr1-4_0410
	β-Mannosyltransferase		chr1-4_0696
			chr4_0471
			chr4_0450
			chr4_0451

11	Pro-peptidase	KEX2	chr2-1_0304
		STE13	chr2-2_0310
		DAP2	chr3_0896
12	Protease		
	Aspartic-type endopeptidase	YPS1	chr4_0584
		YPS2	chr3_0299
		"	chr3_1157
		YPS3	chr3_0303
		"	chr3_0866
		YPS7	chr3_0394
		MKC7	chr1-1_0379
		PEP4	chr3_1087
	Cysteine-type peptidase	ATG4	chr1-4_0522
		GPI8	chr4_0261
		HSP31	chr3_0691
		HSP32	"
		HSP33	"
		SNO4	"

Table 6. Location of TATA elements in certain *Pichia pastoris* promoters.

class	SEQ#	Site (-bp vs. ATG)	tata-type	Site (-bp vs. ATG)	tata-type
glycolyse	60	23	TACAA		
	61	135	TATAA		
	62	143	TATAA		
	63	55	TATA		
	64	106	TATAA		
	65	229	TATAA		
	66	53	TACAA		
	67	205	TACAA	81	TATATA
	68	89	TATAA		
	69				
	70	61	TACAA	24	TATA
	71	122	TATAA		
	72	71	TATAA		
	73	74/21	TATAA		
	74	70	TATAA		
	75	136	TATAA		
	76	52	TATATATA		
	77	66	TATAA		
	78	68/108	TACAA		
	79	117	TATAA		
	80	47/106	TATAA		
	81	25	TATAA		
	82	82	TATAA		
	83	32	TATA		
	84	78	TATAA	72	
	85	112	TATATA	145	TATA
high expression	96	75	TATAA		
	97	84	TATAA		
	98				
	99	14/78	TATAA	22	TACAA
	100				
	101	87	TATAA		
	102	95	TATAA		
	103	29/45	TATAA	19	TATA
	104	21	TATA		
	105				
	106				
	107	38	TATAA		
	108				
	109				
	110				
	111	21	TATA		
	112				
	113	76	TATA		
	114	70	TACAA		
	115	8	TATATATA		
	116	120	TATA		
	117				
	118				
	119	66	TATAA		
	120				
	121	54	TACAA		

	122				
	123				
	124				
S. cerv	125				
	126	82	TATAA		
	127	68	TATAA		
	128	67	TATA		
MeOH	169	76	TATAA		
	170	50	TATATATA		
	171	48	TATAA		
	172	54	TATAA		
	173	102	TATAA	45	TACAA
	174	61	TACAA	24	TATA
	175	34/99	TATAA	12	TATA
	176/65	229	TATAA		
	177/66	53	TACAA		
	178	81	TATATA		
Xyl- and Arabin-ose	XYL1	63	TATAA		
	XYL2				
	XYL3	82/91	TATA		
	RPE1	55	TATATA	48	TATA
	LXR1				

Table 7. Relative glycan contribution

GlycoSwitch strain	relative glycan contribution		
	Man5	GnMan5	GalGnMan5
GnMan5	21,26 ± 0,52	78,74 ± 0,52	
GnMan5::UDP-GlcNAcT	30,52 ± 2,27	69,48 ± 2,27	
GalGnMan5	19,62 ± 0,64	8,30 ± 0,38	72,08 ± 1,01
GalGnMan5::UDP-GalT	21,53 ± 1,91	9,49 ± 3,75	68,98 ± 4,69
GalGnMan5::UDP-Glc-4-epi	20,69 ± 1,72	9,18 ± 0,41	70,13 ± 1,66

Table shows average ± standard deviation (n=12).

Table 8

	Pichia pastoris			*Saccharomyces cerevisiae*				
To check	Pipas annotation	Graf wild-type	Mattanovich Secreted *	Name	Newman	Ghaemmaghami	Liu (relative)	Brockmann
*	chr2-1_0140	54500		KAR2	197,0	336941,9	156,0	48808,8
**	chr1-1_0160/chr4_0844	249		PDI1	407,0	#N/A	195,0	32771,3
*	chr4_0559			CRH1	163,5	29521,4	#N/A	29521,4
*	chr1-4_0426	120000	*	BGL2		45010,4	27,0	28279,0
*	chr1-4_0013	16700		PRC1		44049,8	21,0	27096,5
*	chr2-1_0052	12100	*	SCW11	127,5	22621,3	#N/A	22621,3
**	chr1-1_0267/chr4_0545	3260		CPR5	112,5	#N/A	37,0	13586,4
**	chr1-3_0226/chr1-3_0227	51600	*	GAS1	6241,5	12359,8	27,0	11953,7
*	chr1-4_0611	16000		APE3		4738,8	70,0	11809,0
*	chr2-2_0148	186000		PST1	181,0	11676,9	#N/A	11676,9

chr3_0179				64,5	9755,8	#N/A	9755,8
chr3_1003	1750	*	CTS1	194,0	6348,4	#N/A	6348,4
chr1-3_0229		*	SCW4	575,0	6185,6	8,0	6175,1
chr3_0419	329		ECM14		468,4	26,0	5896,7
chr3_0299/chr4_0584	14900		YPS1		5436,2	#N/A	5436,2
chr2-1_0454	46000	*	EXG1	171,5	4277,0	8,0	5220,8
chr1-4_0037 /chr3_0370/chr3_0767	2300		MNN2	129,5	6725,3	1,0	4416,7
chr3_0107	977		CTS2		3049,1	#N/A	3049,1
chr1-1_0147			DFG5	343,0	2948,6	#N/A	2948,6
chr3_0120	1760		BIG1	64,0	3464,9	1,0	2786,6
FragB_0048			ROT1	635,5	2360,9	2,0	2687,8
chr1-4_0242			DCW1	403,0	2583,7	#N/A	2583,7
chr1-1_0293	21800		UTR2	684,0	#N/A	1,0	2108,2
chr3_0633	7380				1417,9	1,0	1763,1
chr1-1_0130		*			64,3	2,0	1539,5
chr1-4_0017	20,4		PPN1		319,3	1,0	1213,8
chr1-1_0379			MKC7		538,4	#N/A	538,4
chr2-1_0156	1870		NHX1	121,0	521,1	#N/A	521,1
chr3_0394	9960		YPS7	47,0	148,8	#N/A	148,8
chr4_0305	4810		PIR1	538,0			
chr1-4_0164/chr3_0076	61300	*	PRY2		#N/A	#N/A	#N/A
chr3_0306	985		GAS2		#N/A	#N/A	#N/A
chr3_0303/chr3_0866			YPS3	50,5	#N/A	#N/A	#N/A
chr3_0517			ZPS1		#N/A	#N/A	#N/A

P. pastoris Sequences

1 Potentially secreted P. pastoris genes and genes involved in secretion 76
1.1 Putative secreted proteins with a predicted signaling peptide 76
1.2 Proteins potentially involved in secretion 116
2 P. pastoris homologues of genes involved in the glycolysis pathway 123
3 P. pastoris homologues of genes involved in homologous recombination 157
4 P. pastoris genes with high expression levels 174
5 P. pastoris homologs of promoters used for expression of proteins in S. cerevisiae 208
6 P. pastoris genes involved in glycosylation 214
6.1 Nucleotide sugar synthesis and transport 214
6.2 O-glycosylation 219
6.3 Mannosyltransferases 227
6.3.1 α-Mannosyltransferases 227
6.3.2 x-mannosyltransferases 245
6.4 ER glycosylation pathway 247
6.5 Remaining genes of the glycosylation pathway 264
7 Methanol metabolism 268
8 Annotation of homologues of S. cerevisiae proteases 278
8.1 Serine-type peptidases 278
8.2 Serine-type endopeptidase inhibitor 280
8.3 Aspartic-type endopeptidases 282
9 Chaperones 291
10 5S ribosomal RNA gene 309
11 Xylose metabolism 310
12 Arabinose metabolism 314
13 Threhalose metabolism 316
14 β-mannosyltransferases 319

1 Potentially secreted *P. pastoris* genes and genes involved in secretion

1.1 Putative secreted proteins with a predicted signaling peptide

SEQ0001

P. pastoris homolog of *S. cerevisiae* **KRE9 (YJL174W)**

Glycoprotein involved in cell wall beta-glucan assembly; null mutation leads to severe growth defects, aberrant multibudded morphology, and mating defects

Annotated localization: extracellular region (IDA)

S. cerevisiae null mutant: inviable

Chr3, 0960

5' region (in bold start and stop codon previous genes)

agcagggaccctaacagcactatttggaaagtttgcaatgttgttcaatctggcaaactg
ccttgtgaatctgcccgtgggcagtttgggaaacatctggga**cat**gatgatttgagtcgg
tgatgaaaagttttgaggttggtagcgaaatctgactacaatgcgatggtttccctccta
agagcgaatccgcccagtaaaccaacggtagaaccacctatattacttcctatttacagc
ttatatactcatcacaacgtc**cta**cacgaccttgggcgccaacgatactattttgttaaa
gttcttgtcacgcagctccttggctacaacgctactgattctggtggctatgggttcatc
attcttgttcaacagaacacatgcgttat**cat**caaatctgatcactgaaccatctggtct
ctgagtttccttctttgttctcacaataactgcccggactatgtctccccttttgactct
gttagaagccgaagcaccagtgatcgtggaagtggaaggtctggccttctgcaccacaca
gacaatcttgtctccaatggaagcaaagttcttcgggttctttctcaacaccttgatgca
ttcaaccatctgcgcaccagaattgtcgatgcatttaagcatcgtcttaaggtatatcat
tgtgggtgtgaatgttttccagtgaggtaatgaaaaaatgagaaattcgatacctgaaat
cttagaactggaggagcagtagagggagaggactagtgggcggaggtgcctccgcgctcc
ttccacggtcatgtaggtttcttttcttctttcgctaacattctttgccaagttctattc
ttaacaccatttgttttgggcgtgtactaagcagcatcgcctcacgctgctgaatatgaa
atctatcctgtgctattctgtctttcactttgtcaacaactgcaaccagatcgaacacga
acgtgattccattcgtttactttctccattcgacacttcaa

ORF

ATGTTTTGGTTATTGGTTTTGTCTCTTATCTCACAAGCTTTAGCTGTCGTGGAGTTCACTAGCCCTGAAGGTGGT
GAAACCTTCAGCGTCAGTGGTTCGACCGTAACTATCCCTATTGAGTTCAAAGATGATGAAAGCTTCCCCAGCATT
GCAGATGCCAGTAACTTGGTTATCTCCTTATGTACTGGTCCAAATGGTGACATCAGCTGTTCCGAAATTGAGACC
ACTCCCCCCGCTGACCTTCTGTCGGGCTCCACGTATGAATACGATGCAAGCGTACTGGCAACATTTGGAAGAAGC
GGAAGCTACTATTTACAGATATATTCCAAATACTCGGGTGGTTATGCCATCGTGTATTCTGATAGATTTTCCCTG
CAGAGCATGACAGGTACTTTGACTCCTTCAGGTTCTGGTTCTCCTCCAGCAAACGACATTTCCATGAGCAACGCA
CAAAATGCCGAGATATCCAAATCTTTTACCGTACCATACACCTTACAAACTGGTCGTACGCGGTATGCTCCCATG
CAAACTCAACCATCTGGCACTGTGTCCGCAACCGGTTGGACTAGAAGATACCCCACCAGCTCTGTTAGTTACTAC
TCTACACTAGTCCCTAGCCCAGCTGTGTATTCGACCATAACTCCGGGGTGGGATTACTCATACACTTCTGCAGTT
AACTTTGCCACACCCGCTCCTTACCCTTCACAGAACGGAGGATGGTATGCAGCCAGCAGACGTATTAGCAGACCT
TTCATTCAAGCTACTACTGCTCTACAACGCCGATGGGCCTACTGA

Downstream (in bold start and stop codon next genes)

gaagttctgcctcactcttaaaaaccatagtaacaatgtatattttatttactctacca
caatatttttttccctgtttcctctgcggtactgtacctgctgtgtttcgaata**tca**ata
aaacagctcactgtgagcatggacttgttgcaccagtactttcaaggcccgctcatcaag
tgcgttacccatggaagttctgctgaactcttgcacactgccatcgtctttgaccttgcc
gacaaactcgtaaaagtagccattttgaagatcctcgtctctctcacttgctaggttaac
agtcacctttccattcgactctacttgcgctgttcgggaatcatgatgtacaactttacc
tataagtttaattggagaggttgtaaaccgtcccagcacggaagcatctactctgatggc
gtt**cat**atcagttctggttgtgaacgcgttagctgtaattgatttgaagaagggaatacg

cgacagacataatttttagaagacacttagtagcgagcccttccactgatgttttctct
tcgtctgttagactcatctatcaaaaccccttcgcgaaatagtaatacgtagcaccaagtt
ccctgactttcttttcagtaattttcact**atg**actaaacgagtcctttgattgataact
acgattccttcacttggaacttgtaccagtatctgtgccaggagggtgcccatgttgatg
tgtataggaatgacgagattacattggagcaagtctctcaactgaatcccgatgttgtcg
tcatctccccaggccctggtcatccaacaactgatagtggaatttcagtagacgtcatca
agacattcaagggtaagatacctatatttggtgtctgcatgggccagcaatgtatgattg
cagctttcggaggtgtggtagagtacgctggcgagattgttcatgggaagaccagcccta
tcgctcatgacggcaaagggctgttccaacatattccggac

AA (bold underlined: signal peptide)
MFWLLVLSLISQALAVVEFTSPEGGETFSVSGSIVTIPIEFKDDESFPSIADASNLVISLCTGPNGDISCSEIET
TPPADLLSGSTYEYDASVLATFGRSGSYYLQIYSKYSGGYAIVYSDRFSLQSMTGTLTPSGSGSPPANDISMSNA
QNAEISKSFTVPYTLQTGRTRYAPMQTQPSGTVSATGWTRRYPTSSVSYYSTLVPSPAVYSTITPGWDYSYTSAV
NFATPAPYPSQNGGWYAASRRISRPFIQATTALQRRWAY

P. pastoris homolog of *S. cerevisiae* **PRY2 (YKR013W)**
Protein of unknown function, has similarity to Pry1p and Pry3p and to the plant PR-1 class of pathogen related proteins
Annotated localization: extracellular region (IDA)
S. cerevisiae null mutant: viable, decreased osmotic stress resistance

2 homologues

SEQ0002
Chr3, 0076

5' region (in bold stop codon previous gene)
gcatatctcaaggcatcaaaggttctggatctctcgttgatacattccgcgatgggttgc
tctattcaaaggaacttgctgacctgaaaaggtttgaatatgctgaggatctggagtttt
atgatgctgaggaagaagaacagaactttaagcaagggatcactggcaacggaaaaacta
gcaaacaccagagagtcttatcccaagatatctccaataattccgtgaaaatgaccccag
taaatggtatcaacagtgattcgtatgtatcggtgaatac**atag**taccctcctatttagg
caaatgtcatgcttcaaatataatacccacgtttgttattgttacccctatccttttgtc
tagaacacgtttgtcgcacaccaactatgcctagaaccctgtagatcactctcattggct
cgaacgatggggatacagaatgaaacctggtggaaagttgtttacatttggaattgagtt
agcccctgtgattatgacaatctgcatgctgatgcatttgatgcatgtgcattctaatct
gagaattccactcagtagatcgcaacaaagaagacaaaaaagtaggacaatttctctggt
aaagcgagttgaatgatcaaagttaagataagctagcctatttacacctgttgtttctct
gtttgaaaaggccttttttgtgatgatttttttgactcatcacgttattccaaatttca
ctcgaaacaataggtcgtgcagatgcattttcatacacttaaaatcagtcttcttgtcta
atattaagaaacgcgttatttacgtctccccctttttggaaccctggcactcagggtttg
ttagagaagttataagactcctgtggttccctgaaatcgtgttgtttgtgattctttttg
atacatcaagtctttttcaaaaaacaaaagaagatactcttgagccgttttcttcaaatc
ccaactatctgtcactcctttaaaacacgttcactattcca

ORF
ATGAAGCTCTCCACCAATTTGATTCTAGCTATTGCAGCAGCTTCCGCCGTTGTCTCAGCTGCTCCAGTTGCTCCA
GCCGAAGAGGCAGCAAACCACTTGCACAAGCGTGCTTACTACACCGACACAACCAAGACTCACACTTTCACTGAG
GTTGTTACTGTCTACCGAACTTTGAAACCGGGCGAAAGTATCCCAACTGACTCTCCAAGCCACGGTGGTAAAAGT
ACTAAAAAGGGTAAGGGTAGTACCACTCACTCTGGTGCTCCAGGAGCTACCTCTGGTGCTCCAACTGACGACACC
ACTTCGACTAGTGGCTCAGTAGGGTTACCAACTAGCGCAACTTCAGTTACCTCTTCTACCTCCTCTGCAAGTACA
ACAAGCAGTGGAACTTCAGCCACTAGCACTGGTACCGGTACTAGCACTAGCACTAGCACTGGTACTGGTACTGGT
ACTACAGGCACAGGAACCACTAGTTCCAGCACTAGCTCTTCTGCTACTTCGACTCCAACCGGTTCTATCGACGCT
ATCAGCCAGACACTTCTGGATACTCACAATGATAAGCGTGCTTTGCACGGCGTCCCAGACCTTACTTGGTCTACC
GAACTCGCTGACTACGCCCAAGGTTACGCCGATTCATACACTTGTGGCTCTTCATTAGAACACACAGGTGGACCA
TACGGTGAAAATTTGGCCTCTGGATACTCTCCTGCTGGCAGTGTAGAAGCATGGTACAACGAGATCAGCGACTAC
GATTTCTCTAACCCAGGTTATTCTGCTGGTACCGGTCACTTCACCCAAGTTGTCTGGAAATCAACTACACAGCTG

GGCTGTGGATACAAGGAGTGCAGTACCGACAGATACTACATCATCTGCGAATACGCACCTCGTGGAAATATTGTT
TCTGCCGGCTACTTCGAAGACAACGTCCTGCCTCCTGTTTGA

Downstream (in bold start codon next gene)
tcggttcgttcttgcttcaacgtttacgtaatatttcatgtattccattctttaaattgt
ttgattatgtaagttattgtttcaacaccagtagtgtcttcgctgaaaatgtttccagtt
gggaaatttttgactggcagaagaatctcaggatccaactgactcaaaatcttagactag
gcaaacatactaaagaaccctaccgaatctactaata**atg**agcacatcaacttgccggtt
ctacgaaaacaaatatcctgaggtggatgatgttgtcatggtgaatgtacaacagatcgc
cgagatgggtgcctatgtgaaactattggagtatgacaatattgaaggaatggtgttact
ctctgagttatccagaagacgtatcagatctatccagaagcttatcagagtgggtaagaa
tgaagtagtggtggtactgagagttgataaggaaaaaggatacattgacctgtccaagag
acgagtctcgtcggaggatatagcaaaatgtgaggagagatacacaaagtccaagtctgt
ccattctatcttgaggcattgtgctgaaaagtttaacatgcctctggaggagctgtaccg
tactataggatggccattaagcaaggaatttggccacgcctacgatgccttcaagatatc
aatcactgacagcactgtttttgacaaggtccaacccccatcacaagaggtcttagaaga
actgaaagtttacatctcccggaggctaacgccgcaggccattaaatgtcgtgctgacgt
tgaggtctcatgcttcagctatgagggaattgaagccatcaagagttcactaaaagctgc
tgaggatatctccaccgaggctaatcaagtcaaggtcaagctagtagccgctccattgta
cgtcattacaacccaatccttagacaagacgcaaggcattgaaatcttagaacgtgcaat
caaaatcatcgaagactccatcaccaaacagggcggatcct

AA (bold underlined: signal peptide)
MKLSTNLILAIAAASAVVSAAPVAPAEEAANHLHKRAYYTDTTKTHTFTEVVTVYRTLKPGESIPTDSPSHGGKS
TKKGKGSTTHSGAPGATSGAPTDDTTSTSGSVGLPTSATSVTSSTSSASTTSSGTSATSTGTGTSTSTSTGTGTG
TTGTGTTSSSTSSSATSTPTGSIDAISQTLLDTHNDKRALHGVPDLTWSTELADYAQGYADSYTCGSSLEHTGGP
YGENLASGYSPAGSVEAWYNEISDYDFSNPGYSAGTGHFTQVVWKSTTQLGCGYKECSTDRYYIICEYAPRGNIV
SAGYFEDNVLPPV

SEQ0003
chr1-4_0164

ORF
ATGAGATTACTTCACATTTCATTGCTATCAATTATCTCAGTATTGACCAAGGCCAACGCTGAATGTTGTTACACC
AACACACATACTACCACTGAAGTCTGGTATACTACAGTATATGCTCGAGATGTTAGTGAAGAGACTTCTTCCACA
CTGGCTGGTGGAAGTGCAACTGTCAGCTCAGAAGTGAGTTCGACAATTGAATCTAGCGTTGCCACTTCCGCTACC
ACCGAATCTTCAAGTGAGACATCAGGGTCCACATCTGGGTCCACATCTGCCACTGAATCATCAACTGGTAGTAGC
TCGCTAGCAACCAGTTCATCGATAACCAGTTCAGAGTCTTCCACCATTACACAAACCACAGGACAAGAGTCAACA
AGCCCAACCCCATCGTCCTCAGAGACAGGTTCTTCTACTACTACTCCCTACGATATAAGTCCAACGGCAAGTTCC
GACTTTGATGCTTTTAAATATCAAATTCTTGATGAACACAACATAAAAAGAGCTCTACATGGAGTTGACGGATTA
GAGTGGGATGAAGAAGTATATGCTGCCGCCCAAGCATATGCTGACGCATACACTTGTGACGGAACCTTGGTTCAC
TCTGGAAATAGTCTGTACGGAGAAAACTTAGCGTATGGTTACTCAACCAGAGGGACTGTTGATGCCTGGTACAGT
GAAATTGAATATTATGACTTTAATAACCCAGGTTATACCCCAGGTGTTGGACATTTCACTCAAGTAGTTTGGAAA
AGCACCACAAAGCTCGGCTGCGCTTTCAAGTACTGCAATGACTATTACGGAGCCTACGTGGTATGCAACTACTCA
CCACCAGGAAATTATGTCAACGAGGGATACTTCGAAGCCAATGTGTTACCACTGGTAGATTAA

AA (bold underlined: signal peptide)
MRLLHISLLSIISVLTKANAECCYTNTHTTTEVWYTTVYARDVSEETSSTLAGGSATVSSEVSSTIESSVATSAT
TESSSETSGSTSGSTSATESSTGSSSLATSSSITSSESSTITQTTGQESTSPTPSSSETGSSTTTPYDISPTASS
DFDAFKYQILDEHNIKRALHGVDGLEWDEEVYAAAQAYADAYTCDGTLVHSGNSLYGENLAYGYSTRGTVDAWYS
EIEYYDFNNPGYTPGVGHFTQVVWKSTTKLGCAFKYCNDYYGAYVVCNYSPPGNYVNEGYFEANVLPLVD

SEQ0004

P. pastoris homolog of *Phanerochaete chrysosporium* ABB73028
Mannose-6-phosphatase
Ref: Computational analysis of the Phanerochaete chrysosporium v2.0 genome database and mass spectrometry identification of peptides in ligninolytic cultures reveal complex mixtures of secreted proteins (Vanden Wymelenberg et al., 2006)
Chr1-1, 0071

5' region

tgtgtaaaaaatatgttcaattgatatgtggggaactaaggggagaagaaaaaggatatt
gtcagcgggaactggcaccagttaaatagtacgcagccattatttcgaagcgctaggagg
gttccaatgagaggggggaagaggcctggtgagctccggggtaatcgcgtgttaagtgtag
cacaactgggacatagttgctggttcaagtcacgttatttcactttatacagcccggacc
cctatctgttcaggattgctctggatgccattggtccgttcacgaccttgatctggtgat
tgttgtatgagaaatatctggcaatagcggatccatgtacaccggagaccggtgagagta
cgtaagcaacggaggtgtgaaatgaattcaagtaattcggcaaggacgggatgagttgaa
gactcaattagtcgagaaaaagagcgtttttttgacgggggtagtctggacagtgtgcta
gaggtaacggttttcaattacgaagtgtgctttggggatgtcaggctagcatcagcttgg
tccctctatagtgttctctggtgggaattcccaatcgaaaccttattatctcccgaaaat
gataccaccagtgaaggctgaaaaagctctcaatttggtgggggaccatgctttccggat
tatgtaagggccttctactgtagcccgcacgaattgtattgaactaggcaataaacctag
gcatccaatccaggcagctcgaatgtcatctaggaggtgtggactagaagaggatcaaag
aaaaccttaattaggaacaatcgagataacagataaactacattgtatccccactatggt
agatactccaggttcgactgtattactgatccttgcacccttccccatacgagcacacaa
tggtaaactcttagccgaataaatcgaacagatggaaagctgcacccaagaccaataact
tcattcttcagacaaacttgtagaacttggacaactgtaccaccatggtcaacggccaag
gacgtctgtaaagagtccaacctatatggtgattctttcctatgttgttgtctgcctgtt
aaatttgcgttaagaagagaaagagtctggtctgaaaaattttatagctagtcagtcgtg
tcctaaccatatagcaggcgttccaaactttctagggatagaagaccgtcaagaaatgca
aattaaaggatgaaagaggggagaaaggggacaggggatctaggaggagcatgtgaaga
atggtagttcaacattaacttgaggtataagcaaaatggtagtttgatcaactcaatggg
ggtgcattccgctatccaggtggctattgtggatgggcaccttaattggtcagctcaaat
gtcaataatatcaggaaatgacgtcaggtttgtacccgccaagtgcggttctttgtccta
gcacgcatcgctggatgcacgtcctattctcttcttaaatgaggtagcaaaaccactcct
ttaggcatcgactttgctcatcgcgctatttcttcttccccggcataaccttaaatcagc
ccctttaaggtcaaagtgca

ORF

ATGAAGAATTTAGCTAAACCAAAGATTAATTATCCTTTTCTCATTCTTCTTCCATTCCTTTTATTTTTCATCTGG
TTTCCAAATTTCAAGCCGTTCAATAGAGAAATGCCGGAAGTTCAAGCTCCCTCTAAGAGCCTGATCAGACTTCAA
ACCTGGAATCTTCGATTCGATTCGCAACCCAATGGCGAAAGTGTTCAAGACACAATTAGTAGAATGGAACGAAAA
GTTCCTCAAGATAACAAAATCTCTTATTACGAGACTTCTGGTGAAGTTCCATGGTCAAAGAGAAGGATAAGCATT
GCCAACCACGTGGTCTTCAATGGAGCAGGTATCTTTGCAGTCCAGGAAGCATTGTATAGACAAGTTCAAGATCTC
AATGAACTTTTGAACCTTTTTGATTCCTCCTCCTGGGACTGGATTGGTCTTGGGAGAGACGATGGAAAATTAGGC
GGTGAGTTTGAGGCTATCTTCTATGATACAGAAAGATTTGAGCTGATAGAATGGGATAATTTCTGGCTTTCAAAA
ACTCCGTTTGAGCCTTCTAGGTATCCTGGAGCCGGTTCCTTGAGAACTGTAACAATTGGACTATTTAAGTTTAAA
GGAGAAGAATCAAAAAACCCCTTTATTATCATGAATACACACTACGACGAACAGAGTGAAAATCAAAGACGGTTG
GCTTCCTCTCTGATCAAATATAGAGCAAGCTATGAGTTCGAACGTTTAAAAAATGAATATGGTGTTGAAAATCCA
AGTATAATACTGTGTGGTGATTTGAATTCTCATAGTTTCAATGAAAATTCTGGAGGTTATAGAATCATTACAGGA
AACCAGCCCCTAGAGACTGATGAGATTTCTTCAGAATTTGCAGAGAGATTTCGAAGTGCCAACCTGAGCAAATTC
AACTTTAGAGATTTAATGGAAGAGACCCCTCCTCAATTTCGTTTTGGTCACTTATCCACTTTTACTGGGTTCAGC
AAAATTGGCGACACTTCCAAGTTTTCCCGTATTGATTTCCAGTTTGGAGCAACATGGGAAGAAGATGGCAAGACC
AATCATGGATGGAAGGCTTTAAAGTACAAAGTCGATGAGATATTCTATGACAACGAGTACCCATTGTCAGATCAC
CGTCCAGTAGTTACAGAACTTCTATTATAA

Downstream (in bold stop codon previous gene)

tctgtca**cta**gcttatctgtatcgttgcatatctttgaggtccttctatttcaaagacag
gctgtttcatgtatggggctaacaactttgaaaatcgcttttctacagtctccacctgtt
ctgcgagtggcacttgttcagctaatgtcttaagaatggttgggataactagtttatcaa
cattcttaccccatagtagatccatgtgttcataatcttgcactgggacgatatccagga
cattttcatgaggcagatatggtagcatcgtaggtatgttcattagagtgtcattggtac
cgtacatcaatagaagcttagtagaaatattattagttggaaaaggagttggcaaatgcc
ccttgagaccaaagataacattgagaggggagctatcgtggtccttctcgtcaaacattt
caaaacgttgatttttgattatctggaaccaatgcagaatcgatttaactgaactattgg
aaaagagatggggatatccgatttgcttctgctgtgatgagatgttatcgccttgccaat
tgaacaaataaatcaatgagagatcaacaacacgctgatagtattcctgtcccattgccc
atttccaaaaatgaactgagggcattatagccctatttccaagaaccagtggaagatagt
ggtacgcattcactgcaaacgaggcaactgggtgcctcaaattaggtggaatgaatgcag

gtgaaagacctatcattaactgaattttgtcgttcaaatatggattcaaagacagactgg
ctagtccttgagcagttccctgactgaacccaatataggccagtattttgtgtttattgg
taatgtgtagaatggtacggatggaatctggtatgttgtaaatggcatactcattcaacg
aaaagttccaaaatcgttttgatttagtagatattgatgtgtggtttctggagtatttgt
tacctctattatttcccagccatacgtcatatccgttatct

AA (in bold underlined: signal peptide)
MKNLAKPKINYPFLILLPFLLFFIWFPNFKPFNREMPEVQAPSKSLIRLQTWNLRFDSQPNGESVQDTISRMERK
VPQDNKISYYETSGEVPWSKRRISIANHVVFNGAGIFAVQEALYRQVQDLNELLNLFDSSSWDWIGLGRDDGKLG
GEFEAIFYDTERFELIEWDNFWLSKTPFEPSRYPGAGSLRTVTIGLFKFKGEESKNPFIIMNTHYDEQSENQRRL
ASSLIKYRASYEFERLKNEYGVENPSIILCGDLNSHSFNENSGGYRIITGNQPLETDEISSEFAERFRSANLSKF
NFRDLMEETPPQFRFGHLSTFTGFSKIGDTSKFSRIDFQFGATWEEDGKTNHGWKALKYKVDEIFYDNEYPLSDH
RPVVTELLL

SEQ0005

P. pastoris homolog of *C. albicans* SC5314
potential secreted Cu/Zn superoxide dismutase
Chr1-1, 0109

5' region (in bold start and stop codon next gene)
agaaatc**atg**tacaggttacttccaagatcatataacagcgctttgtctaggactttgtg
ttctagctccattgctcgctctacttataagacgttaagacccaccgttgctggcgacag
gcaggatgacttttttcaattcattgcagtttggagaagcctctaagtggtcgaaacattc
actaccacgcagtaggaggttgaaggacaagaattccaacgaggagaaagacgtttctgc
tgcttctgaagatagttccaaaccagcttcagtcacctccaattctacatcgtctcccaa
actatccacatccacagctacatccacattgacaacttcgaacaacactgttccacctac
acagaaaacacagaagaaacgccgtgcgttgcgcccccgcaaggctttgatcactctatc
gccaagcgcagtatcacatctacgagcattatcagatctacccgaaccaaagttgactag
aatcggtgtacgtaacagagggtgctcaggactcacctaccacttggaatacgtatccga
accgggtaagtttgatgagctggtggaacaagacggtgtaaaagttctgatagactctaa
ggctttgtttagtattgttgggtcagagatggactgggttgatgacaagctcagctccaa
gtttgtgtttaagaaccccaattccaaaggaacttgcggatgtggggaatcgttcatggt
t**tag**atgataaaccgaactgactcacattcctctccgagatcggatttgagacaacaaaa
gaaaagggttctttttttccttcttcgcgaaaggtggaatgccttgctatgcgacatcaaa
catctcacaaccaacatcatttctgcatcaaactcaaactctgcggcaccacctttgtgc
cttatagacacaaattcagcatatttaaactagcaagactgttatacacttcagaattgt
actcactcaatcgacttttttacacattctatcacatttata

ORF
ATGAAACTTCAGTTATTTACCCTAACTCTCTCCTTGATGGCATCAAGCACTCTTGCTGCAAAAGCGCCTAAGAAT
AAGGATTCCCCTTCAGGTGCCGTAGCTCGTGCCGACTTTGATGACGACAAACTACATGGGATCATTGAATTCTCT
ACTGCAAAGAACAAATCTGTAGAAGTTCATGTTGATATTACCGGTTTGCCTGAAACTGGTGGGCCGTTCTATTAC
CACATTCATAAGAAACCACTCAAGTCTGGAGCTTCTTGTGAGTCCACCAGTACTCATTTCAACCCGTATAATGCA
CCAGCTGATTGTCTCAAGACCAAAGATGACAGCTACTGTCAAGTCGGGGACTTATCAGGCAAGCACGGCTACATC
AACACCACCTGCTTTGAAGATTCCTATTACGACCCTTATATCTCCTTGAATCCTAAGAATAAGGCCTATCCAGTA
GGGTTGGCGATCAACCTTCACTATGCAAACCAGGAAAGGTTTGCCTGCGCTAACATTGAGTTAGTTGACAAGAAA
GTCAAGCGTGAATTTGATGATGTTGATGCTGAGTTTGTTGGTCAGATCCCAGAATCAGACAATGTAGTTTCAGCT
GAGTCTGTGAAGGAAGACACGAAGGTGAAATCTTCAAAGAAGGATGGATCGTCAACCTCAAAACCTCATGTGAGT
AAGCTTTCTTCCAAAAAGAATGTTACTAATGGTACGTCACCATTATTATCTAAGGGCAAATATTACACGAACTCT
ACATCGAAATATGGAAACTTTACCAATGGTACCAGTTATCCAACCGATGATGACTATGACGACGACGAAACCATA
GAAGTGTCATCCTCGTATGCCAATGCCGCAGGATATTTATTGAACGGAAACACCGTTTATGGAGCTACTCTTGCA
GCCATTATTTCCTACCTGGTCTAG

Downstream (in bold start and stop codon previous gene)
actttcattaactgagctaaatacgcactaaatatatgactaaagaaaagattaagagta
gatgaactaatagtgagccgcatacaagtcgaatgcacgacaaagagtttgtcgcgacgc
atcatctacagacagaagctcctcactaacagagaattcggtgccatgtgaaactgattg
ctgactaggctatcacatggcactaaactgtatataagccagccagtggccaattttatt
ctctcttctagtcaagaaa**atg**tcattcaaatcagagccattgttgcagggaatagctga
caccctcagcaaaaatgaggcgctaaaagaaagaacccttcaaaagaccagagcaatttt

tgtctttaatgtaaaaaataaggagggtaaagccaaggcatgggccttagatctgaaaaa
ggcaggaactttaacagaactgaagagtactgaggaattggaggaattgaagcctgacgc
tactatcacgatcaaagatggcgatcttcgtagtttgatccggggaaagtccaaccctca
aaaacttttatgagtgggaagctgaaaatcaagggaaatgtggctaaagccgcttcaat
cgaaacagtgttgaagagcactaggccagccaagtccaaacta**tga**gattcttatttaca
aaaaaatgaataaattaactaactattttcagtaagtgtctcccttcgaactcaagaatt
gatactacccttatacctgatcttaaaatttttaacctttttctcctcacgtggattctt
tccccaaaggaggatccccacggagattctgttcactagcgtcttttgctgaaccccggt
gtgtactgcgaaatggcctcaccgaattctgtggtgcccaagcttctccgataagtgggt
gttctagcgactgctgcatcataggccctacttgagaataaaaagaaccgagaaaaagaa
acgcgcaaacacagtgtggttgcgccaaatccgaagagact

AA (bold underlined: signal peptide)
MKLQLFTLTLSLMASSTLAAKAPKNKDSPSGAVARADFDDDKLHGIIEFSTAKNKSVEVHVDITGLPETGGPFYY
HIHKKPLKSGASCESTSTHFNPYNAPADCLKTKDDSYCQVGDLSGKHGYINTTCFEDSYYDPYISLNPKNKAYPV
GLAINLHYANQERFACANIELVDKKVKREFDDVDAEFVGQIPESDNVVSAESVKEDTKVKSSKKDGSSTSKPHVS
KLSSKKNVTNGTSPLLSKGKYYTNSTSKYGNFTNGTSYPTDDDYDDDETIEVSSSYANAAGYLLNGNTVYGATLA
AIISYLV

SEQ0006

P. pastoris homolog of *S. cerevisiae* **DSE4 (YNR067C)**
Daughter cell-specific secreted protein with similarity to glucanases, endo-1,3-beta-glucanase, degrades cell wall from the daughter side causing daughter to separate from mother
Annotated localization: extracellular region (ISS)
S. cerevisiae null mutant: viable, no cytokinesis
Chr1-1, 0130

5' region (in bold stop codon previous gene)
tcgaccctacactttacgccgcctttgaagccggtgagcagagcaaactcaatggaccct
atcccacaaccagtaaaagtccccgtggaggatttttttatcccaccaccggggctagtg
cggaggtcaacatcaaattacaagaacgagatgtccaaggcaaagcttcgcaggcaagga
agcgcaggatccaaggaagaggagaaagcaatatttgaccacatgtcgcctcaaaatctc
aaaagaaggtgctccaccgagctgaacagtgcgctcccaaatatggacccaagctcaacc
acaattgcagagcgcaatggggttgagtttacaggtcatacaaagaataattctctgttg
gatcaggttaaggaaaagggagaacacatgtacaataaaattcacgacaaatgggaaaac
cgatttgaatttggattgaagagttgcgatgagtcttccgcatctgtcaacgcacccaat
tcagatgaagaagaagatccgttgacctcaaggaagggcagttttgattttaacgagtat
aagggagtaatgtacaaaaaggggaaa**taa**ataggcgaaataacagtgcatatcatagtt
tacgatcgtattgattgcgatcagatcagattgaatcaatagtaatgaaattaacaccct
tctcagagtggactgaatcgacctaaaaatggtttggaagggcatggctggtacgtgatt
gctcgatttggcaataaccttgcatttcaacggtttttttcgagctttttattattagt
ttttgctgcatccgtgtcgctgtgcatcagtagatttctaaagtagaaatgaaaccaacc
ttctttgttaatgtgcctgattcacgggtatttgaggtaaaaaacaggacctaatgatca
atggctggtgtcatccaccatccaacggttatatatatgagtagcatagcattcaggttg
ttccgtaatcatatcaaacctttctctcgttagcttcccaa

ORF
ATGTCATTCTCTTCCAACGTGCCACAACTTTTCTTGTTGTTGGTTCTGTTGACCAATATAGTCAGTGGAGCAGTC
ATTTCAGTATGGTCCACCTCAAAGGTTACGAAGACTTTTACTGCCGCTCCAAGAACTGTCACTGTTTATGCTGAT
GGTGACGGAGTTGTACCATCCAAGTCGGCCGATGGGGACTCAACGAAAGCAGCCACTGGAGAAATTTCAAATGGC
AATTCGAACGAAGGATCAAACCCAAATGAACTTCCCGCATTTTCCACTATCTACGGCCCCAGTGGACGGTACGAA
CTATACCTCTCCACAAGCTATAGTACTTATGCTACCACTCTCTACTACACTAGCACAGCAACTTATGGTACCATT
CGTTCTTCAGCCGCACCTGCAAAGGAGACTCAATCGGTAAGTGAAACTGAAAAAATCATCTCAGTGAATCCAATT
GAGGCCTCCACGGCTCCGACGACTACCAGCACGGTTTTTTCGGCTACAACATCAATCCCAACAAGTTCTTTGGGA
CGAATCAGCACAGGATTAGAGACTTCAATTTCTAGCAACCCTCTGGTCACAGAGCCTGTGGCAACTTCGTCGGCC
TCAAGAACTATAGAAACTGCAACCTTGACTGAAAGTGTGCAGATTAGCACCGGTCTAGTGAACGGAACTCGGTCA
ACTTCAACTTATGAAACTTTGAAGACAGTCTACAGTAGCTTCAATTCGTCTACAATAATTGGAACTGCGACTGCT
ACGCAAACCTCACTTGGCTCGACTTTCACAACTTTGTCAACGCTCTATAGTAGCTTCAATTCCTCCAGCGCCATT
GGGACTGAAACTGCAACAGAAACTTCAAAGGTCTCTCCAACTCAGACTACCGTGCCAAGTCCAGATGTTGGTGTC

GTAGATTTGTTCGAAACTATCGACACTGCACCACCTCCAAGTGTCTTCTCTAGACAGGATTTACCTCTTCCATTG
TTGAGCTCAGTGTCTAACAACGGCAAACCCTACCAGACCAACAAGTTTTACCAAAACATTGTTATTGGTGATCAA
AAGGGGGCGGTAATCACCTATCCATACAGTGTATTCTACACTTCTGAAAACCATCTCGGGCTGGGAATCTCTCAC
TACACAGAGAGTGATAAGGTGTTTGGGCCCACCAATTCAAACGGAGCCTCCAGTTACTTCCTTAATCCTACCCTG
GTTGGTCACTTAGTTTTATCGGCCAAAGAGTTCACATCTTCAGGATTGAGTTTGCAAGTTTCAGAGTTGAAAGAA
TTGTCAGTATTGGCAACTTTGCAATCCTCCAGCGGATCTGTTCACGCTCCATTGGTGGAAGGAATGGGATTTGTC
TCTGGGGTTTACGAGGGATCTCTGACACTCAGACTGAGCTCTCAAGTTGGTTTTAAGACTCTAACTCAGGAGACT
TCTAAAACTCTTCCTTCCGGAGTTCTAAAGTTCAGAGTAACTCTTTTCAGTGATATTGAATGGTTGGTTTATGTG
ACACTCCCAGAGGGCACTTCGTCAGACTACAAGTTAGAAATCACAGACAATTACACCCTGCAAGGAACCGAGTCG
GTTGACGGGTTAATTGTGCAAGTTGCCAAGGCAGCTGAGAAAGAAGAGTCTGAAATTTACTACGACTATGCGGCT
GGTATGTACCCTGAGTCTGCAGTGGTTACCGGTTCGACAGATATTTCCACCAATGTTGCCTCATATAGTATCGTA
TATGACACTAAGGGGAAGTCATCTTCAGGAAGCACCCTAGTGTTTGCCCTACCCCACCAAGTCGACTCCATCGAC
GAATCAACCAATGCTCGTGACACGGGAGTTCAATTGGAATCAACCACCAAGGGCAAACTAAGGGCTTACCTGACC
AATGTCCTAGGACTCCAAGAGACCGTAACCTCTCAAATCCAATTTCTGCCTTGGGCCTCTTTCATGAGTAAAACT
CTCTCATACACAGCACAACAACTACAAGAGCTTGCCCAAATTGCCAATGACGAAATCTCGCAGGACATCAAAAAT
GCAGTGACGGGGCTGGACTCCAACTACTTTTCAGGAAAAGTCCTGGACAAGTATGCTTACATTTTGCTGGTCATT
TCAGACATCATCCAGGACGATGAGGTTGCTCTGAGTCTTTTAGAGACCTTGAAAGAGACTATTGGAGTGTTCACC
TCAAACCAGCAGTACTATCCACTAATGTACGACACAAGGTATGGAGGAGTGACATCCACTGGGTCTCAAAGTGGA
GACACCGGTGTTGACTTTGGATCTGCATATTACAACGATCACCATTTCCATTACGGTTACTTCATCCACGCAGCA
GCTGTAATCGGACATGTCGACAACAAGATAGGCAATGGGACCTGGGCTCAAGCCAATAAAGACTGGGTCAACTCT
TTGGTCAGAGATGTTGCCAACCCTTCAGATGGGGATTCTTACTTCCCTGTTTCCAGGTCTTTCGATTGGTACCAG
GGCCATTCGTGGGCCAGCGGGCTGTTCTCCGCATACGATGGACGAAACCAAGAGTCTAGTTCGGAAGATTACAAC
TTTGCCTATGGAATGAAGTTATGGGGAGCAGTAATTGGCGATACCTCCATGGAGCAGAGAGGAAACCTGATGCTT
TCTATAATGAGTCGAGCTATAGGAACGTACTTCCTCTTTGAGGACAGCAACGCTGTGGTCCCATCGGAGATCATC
CCCAACAGGGTTGCGGGAATTCTGTTTGAAAACAAAATGGATTACACCACTTATTTTGGTACTAATAAGGAATAC
ATCCACGGTATCCACATGCTGCCTATCACACCAGCCTCCTCGTTGATCAGGACGCCCCAATTTGTCTCTCAGGAG
TGGTCCTCTGTGATTTCAAACATACTGCCAGATGTGAACTCAGGATGGACGGGAATTCTGAGATTAAACCAGGCA
CTTTCTGATCCTGTGTCTGCATTTTCATTTTTTTCAAATGATGATGTATGGAACAATAACTATCTAGATAATGGC
CAAAGTCGTACATGGGCATACGCATTTTCTGCAGGACTGGCTAATTCCGTTTAA

Downstream

agaggtgtaatttcatatataaaatttggcgatatttttaataaggtaccacacatagtt
gacgttgattgtgctgttcttttcgcgataggacaggtgcacgggagatcatccttgtat
cttttctttttgcatgtgggtctttttttttttctcttgtcgcctagaattgttatttca
ttccattccatttcattttgacttctgacagtccctccccactggtaacaatctgccagt
aaaagtttcgtcagtttctttcttggtccaactgggaatggaagggtgcagcctctcatt
tctgttgtgccctttcaaggcgtgtttaccattttcggctgatgggagtggactaaaagg
aaattaccctattgacacaatccgactttcgtggatcttgttttagtaaacccatttttg
ggttgcttttgaacaaaaaaaaaagacacaagtctgcatattactaagaattaccatggt
aacatattgcacttgcatgcgcagaatgcattaacaacaagcagctcagaaaggctgaaa
ttgaggatacgcgcttcattaaaaaatgctgcgtttgaaaatgtgtgcaacaactttgca
tcagcacagaaaacacgacgcagtagtcctagtagggtgatatttggccgccatcttgct
ctggagtgtcgagttaattattttgccccttgatttcaaaacgttactttgtttccttttt
tttctagttcacctttgtacgggttagtaactttgcatatctgattgttacaaaagtgg
atggcaatgcaagctggatgcgcatgaacagtgttccaggcactaattttaacgtgatgt
gtctcttctcgaagatgcattatttttatacacagcatcttggttgcattacattatgtat
gcacctatcatgtatatatgtacgcccgtaagtaatcgtaggtctaataagcaaacgtat
gcagaggttgtaaatttttgaaacgctcagtcggagcatca

AA (bold underlined: signal peptide)

MSFSSNVPQLFLLLVLLTNIVSGAVISVWSTSKVTKTFTAAPRTVTVYADGDGVVPSKSADGDSTKAATGEISNG
NSNEGSNPNELPAFSTIYGPSGRYELYLSTSYSTYATTLYYTSTATYGTIRSSAAPAKETQSVSETEKIISVNPI
EASTAPTTTSTVFSATTSIPTSSLGRISTGLETSISSNPLVTEPVATSSASRTIETATLTESVQISTGLVNGTRS
TSTYETLKTVYSSFNSSTIIGTATATQTSLGSTFTTLSTLYSSFNSSSAIGTETATETSKVSPTQTTVPSPDVGV
VDLFETIDTAPPPSVFSRQDLPLPLLSSVSNNGKPYQTNKFYQNIVIGDQKGAVITYPYSVFYTSENHLGLGISH
YTESDKVFGPTNSNGASSYFLNPTLVGHLVLSAKEFTSSGLSLQVSELKELSVLATLQSSSGSVHAPLVEGMGFV
SGVYEGSLTLRLSSQVGFKTLTQETSKTLPSGVLKFRVTLFSDIEWLVYVTLPEGTSSDYKLEITDNYTLQGTES
VDGLIVQVAKAAEKEESEIYYDYAAGMYPESAVVTGSTDISTNVASYSIVYDTKGKSSSGSTLVFALPHQVDSID
ESTNARDTGVQLESTTKGKLRAYLTNVLGLQETVTSQIQFLPWASFMSKTLSYTAQQLQELAQIANDEISQDIKN
AVTGLDSNYFSGKVLDKYAYILLVISDIIQDDEVALSLLETLKETIGVFTSNQQYYPLMYDTRYGGVTSTGSQSG
DTGVDFGSAYYNDHHFHYGYFIHAAAVIGHVDNKIGNGTWAQANKDWVNSLVRDVANPSDGDSYFPVSRSFDWYQ

GHSWASGLFSAYDGRNQESSSEDYNFAYGMKLWGAVIGDTSMEQRGNLMLSIMSRAIGTYFLFEDSNAVVPSEII
PNRVAGILFENKMDYTTYFGTNKEYIHGIHMLPITPASSLIRTPQFVSQEWSSVISNILPDVNSGWTGILRLNQA
LSDPVSAFSFFSNDDVWNNNYLDNGQSRTWAYAFSAGLANSV

SEQ0007

Homolog of S. cervisiae **YHR202W**

Putative protein of unknown function; green fluorescent protein (GFP)-fusion protein localizes to the vacuole, while HA-tagged protein is found in the soluble fraction, suggesting cytoplasmic localization
S. cerevisiae null mutant: viable

Chr4_0932

ORF

ATGGGATTGTTGGAGAAGAGGGTGATAACCCTCTCAATAGTCACATTGCTTGTGCTGGTTAACCTTCTAGACTTT
AGATCAGTATCTAATGGGTCAACGTCACAGGATTATAGGGCGAACAACATTTTATTTAAACTGAATCTGCGCCAT
GCAGATGCTATTGATTCAAGACAACAATTGCAATCAAGGTTCATTAATTACTACGACCAACTCCGCTCTTTCAAC
ACAGGAGAGCTCAATTTCTTACACACTACTGATACTCATGGCTGGTATATGGGTCATCCCACTCAACCCCAATAT
TCCTCGGATTGGGGAGACTTTTTGTCATTCGTTCAGAGGTTAAAAGATTTACTCTTTCAAGAAGGAAAAGATTTA
ATTTTGGTCGATTCTGGTGATCGACACGATGGAAATGGATTGAGTGACTGCACTTCCCCCAATGGACTACGATCC
AACAATATCTTCTCGGCTATTGATTTCGACTTACTTACTGTCGGTAATCATGAGCTTTACTCTGAAGCTGTGAGT
GAGTTAGAATTTTCCACAGTAATTCCGCATTATGGTGATAGATATATTTCCACAAATGTGGAATACCTTACAGGT
GACGGTGAATGGGTTCCATTCGGGAACAGCCAATATCGTGTCTGGAAGTCCAAAAACACCAGGAGAACTATTCTT
GGATTGTCTTTCATGTTCGACTTTAGATATACTGGAAATGAGAGGGTGAGAGTAACTTCAATTTCTACAATGATC
CAAACTGGTTCTCTTGATGCAATCATTGAGACTATCAAGAGCAATCACAGGCCAGAAATCATCGTAGTTGTGGGC
CACCTTCCAGTCCATCATGTCTGGGCTGAGTCTCGTATGTTACATGACTATCTCCGTTCTAAGTTTCCGGATAAG
ATAATTCAATATTTTGGAGGTCATTCACATATCAGAGACTTTGTTGTTCATGATGACCTTTCAACAGGGTTACAG
AGTGGTAGATTTTGTGAAACCATCGGTTTCTTGTCAATGAACAATATAGACAGTGCCACCTCAAAGGACACAAGC
AATTGGGCTTGGGATAATATTGACAGGAAGTATATCGATTTCAACTTACATTCTATGCTTTTTCATACAAATATC
TCCAAAGTTAGTGAATTCAACACGGAAGAGGGAGTCACACTTTCAGACAAGATTATATCAACAGTAAGTGAACTA
AACCTAACCAAAGAATACGGTAGGGTTCCAAGAAATTACTACCTTGCAGGAAGACCATATCCACACCACCAGAAT
CTTTACAGTCTCATTTCTGATGAAATCCTCCCAACGCTGCCTTATAGAGTTAAGTTAGGAGATGAACCGTTTGAA
CCCAGTAATGCCCGAGTCGTTATCATCAACACAGGTGCTATACGGTATGATCTTTACAAAGGTATTTTTTCCGAA
AACACCAAATATACAATCAGCCCATTCTTGAACCAATGGAGGGTTATCCCAGCAGTCCCTAAGAAAATTGCGTTG
CAGATACTTCCAACCTTAAACCAGTTAGATTATATCATTAGCAATTTGGATAAGAATGCAGACCAGGACTTCTCT
TTGGACATTATGAGCGATTACTATAATGGAGACCATTCTTCCAGAGATCGACTCTTTGGCCAGGATATTGTTGAT
CTCACAACTTTGATGTCCCCTTTTGCACGTTCAAATTTGATTGAAAAGTTGATCATGATGGGCAAGTTCCAATCG
GATGTTAGAAACGCCAAAGAGAAAGAGCAACAGCCCTTCTTACCTCATCATGAGCAGCATGTTTTGGTCAAGCTG
CTAGAATTTTTTAATTGGAACAAGCTTACCCACGGATACGTAACAAATGACGATTTTGGAGATGAAGGAGACGAT
ACTAAACATAAGCAAATCAGTTTTTATGAACCTCCATGCGTTATTCAAAGCTATCAAAGACACTACAATGATAGC
ATATCCAATATAATATTTGACTTGGTGATACCGGATGATGAAGATGAAGTGATTGATGTCGTGTTTTACGACTTT
ATAGAGCCATACATTTTGTACGCTTTGCGACAAATCCTCAATGTCGATTTAGACAAGAGATCATGGGTGCAACCA
AGGGAACCCGTATACTTACAAAGTGCTACCAAAGTTGGCAACACTACATTATATCAGAGTTTCTTGCGATCAATT
CAAGTTTACAATGATGTTGAAAATGAATGGAACGTTGGAGAGATGCTGAAGAGATATGTCACAGAGAACTGGGCC
TAA

AA

MGLLEKRVITLSIVTLLVLVNLLDFRSVSNGSTSQDYRANNILFKLNLRHADAIDSRQQLQSRFINYYDQLRSFN
TGELNFLHTTDTHGWYMGHPTQPQYSSDWGDFLSFVQRLKDLLFQEGKDLILVDSGDRHDGNGLSDCTSPNGLRS
NNIFSAIDFDLLTVGNHELYSEAVSELEFSTVIPHYGDRYISTNVEYLTGDGEWVPFGNSQYRVWKSKNTRRTIL
GLSFMFDFRYTGNERVRVTSISTMIQTGSLDAIIETIKSNHRPEIIVVVGHLPVHHVWAESRMLHDYLRSKFPDK
IIQYFGGHSHIRDFVVHDDLSTGLQSGRFCETIGFLSMNNIDSATSKDTSNWAWDNIDRKYIDFNLHSMLFHTNI
SKVSEFNTEEGVTLSDKIISTVSELNLTKEYGRVPRNYYLAGRPYPHHQNLYSLISDEILPTLPYRVKLGDEPFE
PSNARVVIINTGAIRYDLYKGIFSENTKYTISPFLNQWRVIPAVPKKIALQILPTLNQLDYIISNLDKNADQDFS
LDIMSDYYNGDHSSRDRLFGQDIVDLTTLMSPFARSNLIEKLIMMGKFQSDVRNAKEKEQQPFLPHHEQHVLVKL
LEFFNWNKLTHGYVTNDDFGDEGDDTKHKQISFYEPPCVIQSYQRHYNDSISNIIFDLVIPDDEDEVIDVVFYDF
IEPYILYALRQILNVDLDKRSWVQPREPVYLQSATKVGNTTLYQSFLRSIQVYNDVENEWNVGEMLKRYVTENWA

Peptide

MGLLEKRVITLSIVTLLVLVNLLDFRSVSNGS

SEQ0008

Homolog of YDR452W; **PPN1**

Vacuolar endopolyphosphatase with a role in phosphate metabolism; functions as a homodimer
S. cerevisiae null mutant: viable
Chr1-4_0017

ORF
ATGATGGTGTTTGGATGCTTATTTTATCCTGAAACGCTCACGTTGAACTTTCAAACGACAGTCAACATGTCCCAT
CCACTATTTATAGCAATTGTAAGCCTTGTTATATTTCTAACAGCTTACGGAAACAGGGAATTTATATTTGGCAAG
TTTGAAGGAAATGTTGTATTTGACTACAGCAAGAACTCTGATATTTTATCACAATTAGGACTAGTAGGCTCATCT
TATGCTAGCAAATGCCATGGTAAGGTAGTGAAGAACAAGAGCCAGCAAAATCATACTGCTTCACGTTATGGAGAT
GCACTCAGTGGTTGCGACAGTCCCATTGAATTGATGCAATCTACCCTCGATTGGATTAAAGATAACCTATTGGAT
AAAATTGATTTTGTTGTCTGGACAGGTGACAACGTTAGACACGATAATGACAGGGATCATCCTCGTTTGGAATCG
CAAATATTCGAAATGAATCAAGTTGTAGCTGATCTTATTCACAGTACATTTCTTACTGATAAAGATAGAGAAAAC
CAAGAGAATATAATCGATAATTCGAGCCGTGATATCAAAATAATTCCTTCCTTGGGAAATAACGATGTATATCCT
CATAATATGTTCTCCCCAGGACCCACTCTTCAAATCAGAGAATTCTACAGGATCTGGAGAAACTTTGTTCCAGAG
GAGCAATTACACGTATTTGGCCGAGGTGCATACTTTTTTGTGGAAGTCATAAAAGGAAAACTCGCTGTATTATCA
TTCAATACTTTGTATTTGTACACAGCAAACCCTCTAGTGGATAACTGTGATAGTCGTAAACAGCCAGGCTACCAG
CTTTTTCGTTGGCTTGGTGTTGTACTCGAAGAAATGAGGAAAAGGGATATGAAAATATGGCTGACTGGCCACGTT
CCTCCGGTACCGAAAAATTATGATGAATCTTGCTTGAGAAAGTATATTATCTGGACTCATGAGTATAGAGACGTC
ATTATTGGTGGATTGTATGGACATATGAACATTGATCACTGGATCCCTCTAGATTCTGAGTGGGCATATAATTCT
CTAGCAGCTGAAGTGTTTCCATCTGAAAGCATACCATTTGATTTATACACTGGATTCCCAGAACTTGATATTAAA
GATGCAGACGGCATGTCACATCGTTATGACGGAGTTCCTGTTGGAAAGGTTAGGTACATGAACTCGGTGCGTGAA
GATCTCTACGCCCGAATAGCAGAGAGTAATGATTCGGGTATATTCTCTGATCGATATTCCATTGCACATGTATCT
ACCAGCGTGATACCAACCTACAACCCAGGAATCAGAGTTTGGGAATACAACATTACAGGCCTGGTTGATGAAGAT
GAAAAGAGAGCTGAGATTCAAAGGCATCCTCCATGGTCGACTTTTTTTGAAGATTTGGAGGTGGAAATGGCGGAA
GATGAGGAGACCGACGAATTTGCCTTTCCTACAAAGGACCCTTCGTTACCCGTGAAAATGCCATCAAAGACACCT
CTTGGTCCTGCCTATGTTTCTCAGACATTCTCTCCTTTGAGGTACGTACAGCTCTTTGCCAATTTGACTGCCATC
AATTCAGGTCAGTCCCCGTTCCAGTATGAACTTGAATATACCACTGATGACTCACCTTACCAGATGAACAGTCTG
TTGGTTGATGATTGGATCAAGCTTGGTCGAAAACTAGGTGCCTCCACGGATGATCAATGTCCCTCGCAATATAAG
CACCTTTGGAAGAGTTTTCTTTCCCGGGCTTTTGTCCAGACAGACTACGAAGATACAGGAATATAG
AA
MMVFGCLFYPETLTLNFQTTVNMSHPLFIAIVSLVIFLTAYGNREFIFGKFEGNVVFDYSKNSDILSQLGLVGSS
YASKCHGKVVKNKSQQNHTASRYGDALSGCDSPIELMQSTLDWIKDNLLDKIDFVVWTGDNVRHDNDRDHPRLES
QIFEMNQVVADLIHSTFLTDKDRENQENIIDNSSRDIKIIPSLGNNDVYPHNMFSPGPTLQIREFYRIWRNFVPE
EQLHVFGRGAYFFVEVIKGKLAVLSFNTLYLYTANPLVDNCDSRKQPGYQLFRWLGVVLEEMRKRDMKIWLTGHV
PPVPKNYDESCLRKYIIWTHEYRDVIIGGLYGHMNIDHWIPLDSEWAYNSLAAEVFPSESIPFDLYTGFPELDIK
DADGMSHRYDGVPVGKVRYMNSVREDLYARIAESNDSGIFSDRYSIAHVSTSVIPTYNPGIRVWEYNITGLVDED
EKRAEIQRHPPWSTFFEDLEVEMAEDEETDEFAFPTKDPSLPVKMPSKTPLGPAYVSQTFSPLRYVQLFANLTAI
NSGQSPFQYELEYTTDDSPYQMNSLLVDDWIKLGRKLGASTDDQCPSQYKHLWKSFLSRAFVQTDYEDTGI
Peptide
MMVFGCLFYPETLTLNFQTTVNMSHPLFIAIVSLVIFLTAYGN

SEQ0009
Homolog of YDR371W; **CTS2**
Sporulation-specific chitinase
S. cerevisiae null mutant: viable
Chr3_0107

ORF
ATGAAAATTGCCCAGTTGTTTTTGCGCAACCATCTCGATTTCGTACTAATCTCTGGCACCGGCGTAGTACTTTTT
ACGTTAATTGCCCTGTTTGGCGACATTTCCACGTTAGACTCAAATAATCGACAAACTCAAAAGGTCTTGAAAGAT
TCAATGAAAGCATTAGATAAAGAACAGAAGGAACAGCAGCAAGTAATCCACGGTTCTCCTCACGCGCCGTTGTCA
TACAGAACGTGTGCCTATTTTTCATATTGGTCTGTTTACGAACCTCGTAATTTTACTCCCTCCGATCTCCCAATT
AGTCAACTGACAAACGTGTTTTATGCTTTCTTCGACATCGACTGTGAAACGGGGAAGGTGGTTGTGATTGATAAG
TGGTCATCTCTGGAGATGCCACTAAAGTTTTCACCTCGTAGTGACGATATCAAGACCCTTTACGCCGATGGTACT
GCAGAAGCCAAAAAACTGAAACAGCTTCATTTCAGTAAGAAAACCGTTGGAGCCCTCAAACAATTGTATCAAATG
AAATTGATCAAACCTTCCCTTCAAACTTCAATGTCCATTGGAGGATGGGGTCGCACCGACGGGTTCAAGGCTTTA
ATGAAAGACGAATCTAAAGTTGAGCAGTTCGTGTCTTCATGTGTCTCTGCCATGCAGGAGTTTGGGTTTGATGGT

GTTGATCTAGATTGGGAATACCCGTCAAATGATGAAGAAGGGCAATTCTTACTGCACTTAGTTTCCTCTCTAAAA
AGGAAGCTATCCCAATTGGAAGCTCATTACGAATTGCCTCCAGACACATTTCTTATCACGTTGGCTACTCCAGCA
TCTCTGTATACCTTACAGCACTACCCCTTATCAAGATTGGATTTGTTTGTTTCATTTTGGAACGTGATGACTTAC
GACTTTGCTGGCTCTTGGTCCCCACAAACTGAATATCACTCTAATCTTTACTCGCCAAAAGGTGAAATGTTGAGC
GTGAAAGATGCGATGGCTTACTACGTGAGAAACGCTATTCCGCCCAATAAATTGATCATGGGATTACCTAACTAT
GGCCGTGGATTCGGCAACACGGATGGATACAATAAGAAGTTCAAAGGAGTTGGTAAAGGAACCTCCGATGAGGAA
GGCATATACAATTACAATATGCTACCTCAAAATTTCAAGGATAATGAAAAGTACGACCCTGAGGCTGTCGCTGCG
TACAGTTATGAACCCATATCTAGAACGTTTGTTTCATATGATACTCCAAGAGTTGTACGAGAAAAAGCTAAGTTT
GTGAAAAATAATGGTTATGGAGGAGCTATGTGGTGGGAAGCCTGCGGAGATTTTTACTTTGATAATCCCGAACAG
TCTCTACTTTTGAACTTTGTGGATGAACTGGGTGGTACTCAAACACTGCTCAAAGTGCCATCCTCCCAGCACCAA
GATAATAATACCAAGAGTGCATTCTACAAAAAAAGTAACACCCTGAAAGGACTGGTTTAG

AA

MKIAQLFLRNHLDFVLISGTGVVLFTLIALFGDISTLDSNNRQTQKVLKDSMKALDKEQKEQQQVIHGSPHAPLS
YRTCAYFSYWSVYEPRNFTPSDLPISQLTNVFYAFFDIDCETGKVVVIDKWSSLEMPLKFSPRSDDIKTLYADGT
AEAKKLKQLHFSKKTVGALKQLYQMKLIKPSLQTSMSIGGWGRTDGFKALMKDESKVEQFVSSCVSAMQEFGFDG
VDLDWEYPSNDEEGQFLLHLVSSLKRKLSQLEAHYELPPDTFLITLATPASLYTLQHYPLSRLDLFVSFWNVMTY
DFAGSWSPQTEYHSNLYSPKGEMLSVKDAMAYYVRNAIPPNKLIMGLPNYGRGFGNTDGYNKKFKGVGKGTSDEE
GIYNYNMLPQNFKDNEKYDPEAVAAYSYEPISRTFVSYDTPRVVREKAKFVKNNGYGGAMWWEACGDFYFDNPEQ
SLLLNFVDELGGTQTLLKVPSSQHQDNNTKSAFYKKSNTLKGLV

Peptide

MKIAQLFLRNHLDFVLISGTGVVLFTLIALFGDISTLDSN

SEQ0010

Homolog of YJL034W; **KAR2**

ATPase involved in protein import into the ER, also acts as a chaperone to mediate protein folding in the ER and may play a role in ER export of soluble proteins; regulates the unfolded protein response via interaction with Ire1p

S. cerevisiae null mutant: decreased metal resistance; abnormal nuclear fusion during mating; inviable

Chr2-1_0140

ORF

ATGCTGTCGTTAAAACCATCTTGGCTGACTTTGGCGGCATTAATGTATGCCATGCTATTGGTCGTAGTGCCATTT
GCTAAACCTGTTAGAGCTGACGATGTCGAATCTTATGGAACAGTGATTGGTATCGATTTGGGTACCACGTACTCT
TGTGTCGGTGTGATGAAGTCGGGTCGTGTAGAAATTCTTGCTAATGACCAAGGTAACAGAATCACTCCTTCCTAC
GTTAGTTTCACTGAAGACGAGAGACTGGTTGGTGATGCTGCTAAGAACTTAGCTGCTTCTAACCCAAAAAACACC
ATCTTTGATATTAAGAGATTGATCGGTATGAAGTATGATGCCCCAGAGGTCCAAAGAGACTTGAAGCGTCTTCCT
TACACTGTCAAGAGCAAGAACGGCCAACCTGTCGTTTCTGTCGAGTACAAGGGTGAGGAGAAGTCTTTCACTCCT
GAGGAGATTTCCGCCATGGTCTTGGGTAAGATGAAGTTGATCGCTGAGGACTACTTAGGAAAGAAAGTCACTCAT
GCTGTCGTTACCGTTCCAGCCTACTTCAACGACGCTCAACGTCAAGCCACTAAGGATGCCGGTCTCATCGCCGGT
TTGACTGTTCTGAGAATTGTGAACGAGCCTACCGCCGCTGCCCTTGCTTACGGTTTGGACAAGACTGGTGAGGAA
AGACAGATCATCGTCTACGACTTGGGTGGAGGAACCTTCGATGTTTCTCTGCTTTCTATTGAGGGTGGTGCTTTC
GAGGTTCTTGCTACCGCCGGTGACACCCACTTGGGTGGTGAGGACTTTGACTACAGAGTTGTTCGCCACTTCGTT
AAGATTTTCAAGAAGAAGCATAACATTGACATCAGCAACAATGATAAGGCTTTAGGTAAGCTGAAGAGAGAGGTC
GAAAAGGCCAAGCGTACTTTGTCTTCCCAGATGACTACCAGAATTGAGATTGACTCTTTCGTCGACGGTATCGAC
TTCTCTGAGCAACTGTCTAGAGCTAAGTTTGAGGAGATCAACATTGAATTATTCAAGAAGACACTGAAACCAGTT
GAACAAGTCCTCAAAGACGCTGGTGTCAAGAAATCTGAAATTGATGACATTGTCTTGGTTGGTGGTTCTACCAGA
ATTCCAAAGGTTCAACAATTATTGGAGGATTACTTTGACGGAAAGAAGGCTTCTAAGGGAATTAACCCAGATGAA
GCTGTCGCATACGGTGCTGCTGTTCAGGCTGGTGTTTTGTCTGGTGAGGAAGGTGTCGATGACATCGTCTTGCTT
GATGTGAACCCCCTAACTCTGGGTATCGAGACTACTGGTGGCGTTATGACTACCTTAATCAACAGAAACACTGCT
ATCCCAACTAAGAAATCTCAAATTTTCTCCACTGCTGCTGACAACCAGCCAACTGTGTTGATTCAAGTTTATGAG
GGTGAGAGAGCCTTGGCTAAGGACAACAACTTGCTTGGTAAATTCGAGCTGACTGGTATTCCACCAGCTCCAAGA
GGTACTCCTCAAGTTGAGGTTACTTTTGTTTTAGACGCTAACGGAATTTTGAAGGTCTCTGCCACCGATAAGGGA
ACTGGAAAATCCGAGTCCATCACCATCAACAATGATCGTGGTAGATTGTCCAAGGAGGAGGTTGACCGTATGGTT
GAAGAGGCCGAGAAGTACGCCGCTGAGGATGCTGCACTAAGAGAAAAGATTGAGGCTAGAAACGCTCTGGAGAAC
TACGCTCATTCCCTTAGGAACCAAGTTACTGATGACTCTGAAACCGGGCTTGGTTCTAAATTGGACGAGGACGAC
AAAGAGACATTGACAGATGCCATCAAAGATACCCTAGAGTTCTTGGAAGACAACTTCGACACCGCAACCAAGGAA
GAATTAGACGAACAAAGAGAAAAGCTTTCCAAGATTGCTTACCCAATCACTTCTAAGCTATACGGTGCTCCAGAG

GGTGGTACTCCACCTGGTGGTCAAGGTTTTGACGATGATGATGGAGACTTTGACTACGACTATGACTATGATCAT
GATGAGTTGTAG

AA

MLSLKPSWLTLAALMYAMLLVVVPFAKPVRADDVESYGTVIGIDLGTTYSCVGVMKSGRVEILANDQGNRITPSY
VSFTEDERLVGDAAKNLAASNPKNTIFDIKRLIGMKYDAPEVQRDLKRLPYTVKSKNGQPVVSVEYKGEEKSFTP
EEISAMVLGKMKLIAEDYLGKKVTHAVVTVPAYFNDAQRQATKDAGLIAGLTVLRIVNEPTAAALAYGLDKTGEE
RQIIVYDLGGGTFDVSLLSIEGGAFEVLATAGDTHLGGEDFDYRVVRHFVKIFKKKHNIDISNNDKALGKLKREV
EKAKRTLSSQMTTRIEIDSFVDGIDFSEQLSRAKFEEINIELFKKTLKPVEQVLKDAGVKKSEIDDIVLVGGSTR
IPKVQQLLEDYFDGKKASKGINPDEAVAYGAAVQAGVLSGEEGVDDIVLLDVNPLTLGIETTGGVMTTLINRNTA
IPTKKSQIFSTAADNQPTVLIQVYEGERALAKDNNLLGKFELTGIPPAPRGTPQVEVTFVLDANGILKVSATDKG
TGKSESITINNDRGRLSKEEVDRMVEEAEKYAAEDAALREKIEARNALENYAHSLRNQVTDDSETGLGSKLDEDD
KETLTDAIKDTLEFLEDNFDTATKEELDEQREKLSKIAYPITSKLYGAPEGGTPPGGQGFDDDDGDFDYDYDYDH
DEL

Peptide

MLSLKPSWLTLAALMYAMLLVVVPFAKPVRAD

Homolog of YBR015C; **MNN2**

Alpha-1,2-mannosyltransferase, responsible for addition of the first alpha-1,2-linked mannose to form the branches on the mannan backbone of oligosaccharides, localizes to an early Golgi compartment

S. cerevisiae null mutant: budding pattern: abnormal; cell shape: abnormal; cell size: decreased; glycogen accumulation: increased; chitin deposition: increased; competitive fitness: decreased; resistance to hygromycin B: decreased; resistance to Calcofluor White: decreased; viable

3 homologs

SEQ0011

Chr3_0370

ORF

ATGTTTGGAAAGAGGAGGCAGGTAAGAAAATTACTTATATGGGTTGTCTTACTGTTGATAGTTTACTTTTTTGGT
TTGCAATTTAGAGCCAAAAACTCAGCTCATCAGTCTTCCATCAGGAGCTTTTATGCGGATAACAAAGAGTTTTTT
GATCGGCAATACAGCAGATATGACGAGTATGATATCATAGACAACATGAATAGCCATAACGAGTTACTACAGGAA
CAGTTTCGCAATGGAAAGCTCGCTGCGGGACTCAGAGGTGTAGCCGAAGAACCTAATTCCGATGAGGTTACAGAT
GACACTGCCATCGAAGAAGACGAGCAAGCTGCAATGATTAATTTCCCGAAGAGATCTCCCCAGAGAGAGAAGAGT
CTAGTTGAGTTACGCAAGTTTTACAAGAACGTGCTTTCGATTATAATCAACAACAAACCTGCTATGCCAATTGAA
AATCCTCGTGATCCCACACCTAACGAAAACGCACTCAAGAGAAAGTTCGGTAAAAGTGGAATCATCAACATTGCT
TTACATGACACCGATCCTTCGCTGCCTATTCTTTCTGAAGCGTATCTCCGGGACTCGTTGCAGTTGAGCCCCTCT
TTTATTGCCTCATTGTCCAAATCCCACAGCGCCGTGGTAAAGGCGTTTCCCCCTTCCTTTCCTGCGAATGCATAT
AATGGGACAGGAATTGTATTTATTGGTGGTCAAAAGTTTTCCTGGTTATCATTACTTTCTATCGAAAACCTTAGG
AAAACAGGATCCAAAGTACCTGTCGAGTTGATAATCCCATTTGCACATGAATATGAACCTCAATTATGTGAAGAG
ATACTACCAAAGTTGAATGCTACTTGCGTTCTTTTGCAGGAGACTGTGGGAATTGATCTGCTTAAATCTGGACAT
CTCAAAGGGTACCAGTTTAAGTCATTAGCGCTTTTGGCCTCCTCTTTTGAGCAGGTTCTACTAGTAGACTCTGAT
AACATCATTGTTGAAAATCCAGACCCCATATTTGATTCAGAGGTTTTTCAACGGACGGGATTAGTGTTGTGGCCT
GATTTCTGGAGGCGCGTTACCCATCCTGATTACTATAAGATTGCTGGCATCAAGTTGGGCTCTGAACGAGTAAGG
CATGTTGTTGACTCCTACACTGACCCCTCGTTGTATACCTCTAGTTCGGAAGATCCCTTTACCGATATTCCTTTG
CATGATAGGGAAGGGGCAATTCCAGATGGTTCCACAGAATCTGGCCAAATACTAATATCCAAGACCAAACATTGC
CAAACGATCCTTCTATCTTTGTATTATAACTTTTTTGGACCAGATTACTACTACCCTCTGTTCACCCAAGGAGCA
AGCGGAGAAGGAGATAAAGAGACCTTTTTAGCTGCTGCTAACTACTATAAACTACCATTTTACAACATCAAGAAA
GGTGTGGACGTCATTGGATATTGGAAGCCTGACCAATCTGCATATCAGGGTTGCGGTATGCTGCAATACGACCCA
ATCGTAGATTATCAAAACTTACAAACATTTTTAAAGACTCATAAAGGCTCAAGGGTCAACAAACTTGAACAGTCG
GAGCTAGACAAACCCGGATTACTATCCAGATTAATACCAAAATTTTTCTTTCGCAAAACTTTTGATGAGCATCAG
CTTCAGAGTCATTTCACCAAGGACAGATCTAAGATCATGTTTATACACTCAAACTTTCCAAAACTAGACCCATTT
GGATTGAAGCTTCACAACTATCTTTTTGTAGATCAAGACACTCATAAACCAAGGATAAGAATGTATGCGGACCAA
ACAGGACTTTCATTTGACTTTGAGCTCCGTCAATGGATTATCATTCATGAATACTTTTGTGAGTACCCAGATTTC
AACTTGAAATATTTGGAGAACGCTAACGTGAAGCCTCAAGATTTATGTATGTTCATCAAGGAAGAGCTAAACTTT
TTACAGAATAACCCAATACAATTGACGTGA

AA

MFGKRRQVRKLLIWVVLLLIVYFFGLQFRAKNSAHQSSIRSFYADNKEFFDRQYSRYDEYDIIDNMNSHNELLQE
QFRNGKLAAGLRGVAEEPNSDEVTDDTAIEEDEQAAMINFPKRSPQREKSLVELRKFYKNVLSIIINNKPAMPIE
NPRDPTPNENALKRKFGKSGIINIALHDTDPSLPILSEAYLRDSLQLSPSFIASLSKSHSAVVKAFPPSFPANAY
NGTGIVFIGGQKFSWLSLLSIENLRKTGSKVPVELIIPFAHEYEPQLCEEILPKLNATCVLLQETVGIDLLKSGH
LKGYQFKSLALLASSFEQVLLVDSDNIIVENPDPIFDSEVFQRTGLVLWPDFWRRVTHPDYYKIAGIKLGSERVR
HVVDSYTDPSLYTSSSEDPFTDIPLHDREGAIPDGSTESGQILISKTKHCQTILLSLYYNFFGPDYYYPLFTQGA
SGEGDKETFLAAANYYKLPFYNIKKGVDVIGYWKPDQSAYQGCGMLQYDPIVDYQNLQTFLKTHKGSRVNKLEQS
ELDKPGLLSRLIPKFFFRKTFDEHQLQSHFTKDRSKIMFIHSNFPKLDPFGLKLHNYLFVDQDTHKPRIRMYADQ
TGLSFDFELRQWIIIHEYFCEYPDFNLKYLENANVKPQDLCMFIKEELNFLQNNPIQLT

Peptide

MFGKRRQVRKLLIWVVLLLIVYFFGLQF

SEQ0012

Chr1-4_0037

ORF

ATGCTCTTCGGACTAATTAGGCACTCACGTAGACAGTTGTTGTTTTTGGGTGCGCTGGTAACCGTGATAGTCTTA
ATTTTTACCTTGCCCAACACATCACCAATAGAAGCAAATGGAGTAAAATCAGAAGAAGGGTCCATTACACCTATT
ATTCCAGTTTTGGAAAGTCCAGCAAACTCATTGGAAAAAATTGTTGATACTGCCTCAGAGGAACGAATTGGAGGT
GCAACTTTGGAAGAAGGACATGAAAACAATAAGGAAGAACAAGCTTTGGAAAATGCTGAACGAGCTAAAGAGAAA
GAAAAAACTGAGGCGATTGCGGCAGAAGAGGAAAAACTGAAAGCAGCTGAGTTATTGAGACAGCAAGAAACTACG
AGGGAGAAGGAAGCGGCGAAAGAGGACGATAGCAAGAAACCAAACCAGGAACTTGTGGAACAAGACACTTACCTT
GACGATATTCCAGATGACGTAGAAGATAATATTATAATCTCAGAACAAGATCGCAAGAAAATTATATTACCGTCC
TACACTCCTAAAACAGACCCTGCCTATTCCAAGCGGGCCACTGCTTTAAAAATTTTTTATAATGATTTTTCATA
AAGGTGGCTGATTCAGGACCTAATACAGCGCCCATAACCAAGAAGACAAGAAAGAAAGGAAAAAGTAAGTTAAAG
GGTGACGTTTCCTCAGGTGATAAATATGAAGGTCCTGTATTGACGGAAGATTTTCTACGTTTCATGGAGATCTAT
TCTGACGAGTTCATCGATGCTGTATCAGAATCCCACTCAAAAATTGTCAATTTGATGCCCGAATCTTTCCCCAAA
GGAATGTATCAAGGCGACGGTATTGTCATTATTGGAGGAGGGGTGTACAGTTGGTATGGACTGTTAGCAATACGC
AACTTAAGAGATGGTGGGAATACTCTGCCAGTGGAGTTAATGTTGCCAAGCGACAATGAATATGAGCCTCAATTA
TGTGAGCAGATTTTACCGTCTTTGAATGCCAAATGTATTATGCTCTCTGATATTGTTGATCAAGATGTGTTAAAG
AAGCTTGATTTCAAGGGATATCAGTTTAAGGCATTATCTTTGTTGGCCTCAAGCTTTGAGAATGTTCTTTCATTA
GACTCTGATAATATCCCAGTTGCAAATGTTAGCCATTTATTTGATCATGAACCATTCAGCGAAACTGGGCTTGTA
TCATGGCCTGATTTTTGGAGAAGAACTACAAACCCTCGGTATTATGAAGCGGCAGGAATCAAGATAGGAGAATAC
CAAGTCAGAAATTGCCTTGATGGCTTTGTTCCTGAGTCTGATTTTGTTCATATTGGACTGAAAGATATTCCTCTT
CATGATCGGAATGGGACAATACCTGATGCGTCAACTGAGTCTGGACAATTATTGGTAAACAAGAATAAGCATGCA
AAAACACTTATGCTAATGTTCTACTATAATTTTTACGGCCCAGGTTACTATTATCCATTATTAAGTCAGGGAATG
GCTGGTGAAGGAGACAAAGAAACTTTTTTGGCTGCGGCAAATTTTTTTGGGCTTCCCTTCTACCAAGTCAAGGCA
GGTCCAGGGATTTTGGGCCATCATGACTCAACGGGTGCGTTTACAGGCGTCGCAATTGTGCAATACGATCCAATC
GCTGATTACGAGCTTACAAAAGAAAATTTTGTGGGTGAAAAGAGAAAGGGAATTGAGGCACCAAAGGCATTTTAT
GGTAACAATAACAAATCTCCATTGTTCCATCATTGTAATTTCCCAAAATTGGACCCTGTAAAACTAATCAAAGAG
AAGAAGTTGATCGACAACAAAACTCACAAATTCAACAGAATGTATGGCCCAAACACCAAATTGAAGTACGATTTT
GAAGAAAGACAATGGAAGTATACCAAAGAGTACTTGTGTGAAAAAAAGTACAACCTACTATATTTTACTGAACAA
TACAAAAATTACGGACAGGGATACTCCCAGGAGCGTATTTGTAAGTTTTCTGACCGATTTTTGAAATTTTTGAGT
GACAATCCAATTAGGATCGAGGGCTAA

AA

MLFGLIRHSRRQLLFLGALVTVIVLIFTLPNTSPIEANGVKSEEGSITPIIPVLESPANSLEKIVDTASEERIGG
ATLEEGHENNKEEQALENAERAKEKEKTEAIAAEEEKLKAAELLRQQETTREKEAAKEDDSKKPNQELVEQDTYL
DDIPDDVEDNIIISEQDRKKIILPSYTPKTDPAYSKRATALKIFYNDFFIKVADSGPNTAPITKKTRKKGKSKLK
GDVSSGDKYEGPVLTEDFLRFMEIYSDEFIDAVSESHSKIVNLMPESFPKGMYQGDGIVIIGGGVYSWYGLLAIR
NLRDGGNTLPVELMLPSDNEYEPQLCEQILPSLNAKCIMLSDIVDQDVLKKLDFKGYQFKALSLLASSFENVLSL
DSDNIPVANVSHLFDHEPFSETGLVSWPDFWRRTTNPRYYEAAGIKIGEYQVRNCLDGFVPESDFVHIGLKDIPL
HDRNGTIPDASTESGQLLVNKNKHAKTLMLMFYYNFYGPGYYYPLLSQGMAGEGDKETFLAAANFFGLPFYQVKA
GPGILGHHDSTGAFTGVAIVQYDPIADYELTKENFVGEKRKGIEAPKAFYGNNNKSPLFHHCNFPKLDPVKLIKE
KKLIDNKTHKFNRMYGPNTKLKYDFEERQWKYTKEYLCEKKYNLLYFTEQYKNYGQGYSQERICKFSDRFLKFLS
DNPIRIEG

Peptide

MLFGLIRHSRRQLLFLGALVTVIVLIFTL

SEQ0013

Chr3_0787

ORF

ATGTTTAATAGCTTGGCACCCATGAGGTTGAAGAAACTACTTAAGGTGTTTTGTGCCTCAGTAGTGCTGTTAGCG
GCGACATCTGTGGTGTTATTTTTCCACTTTGGAGGTCAGATCATAATCCCCATACCGGAACGCACTGTGACCTTA
AGTACTCCTCCCGCAAACGATACTTGGCAGTTTCAACAGTTCTTCAACGGCTATTTAGACGCCCTGTTAGAGAAT
AACCTGTCGTATCCGATACCAGAAAGGTGGAATCATGAAGTTACAAATGTAAGATTCTTCAATCGCATAGGTGAA
TTGCTCTCGGAGAGTAGGCTACAGGAGCTGATTCATTTTAGTCCTGAGTTCATAGAGGATACCAGTGACAAATTC
GACAATATTGTTGAACAAATTCCAGCAAAATGGCCTTACGAAAACATGTACAGAGGAGATGGATACGTTATTGTT
GGTGGTGGCAGACACACCTTTTTGGCACTGCTGAATATCAACGCTTTGAGAAGAGCAGGCAATAAACTGCCAGTT
GAGGTCGTGTTGCCAACTTACGACGACTATGAGGAAGATTTCTGTGAAAATCATTTTCCACTTTTGAATGCAAGA
TGCGTAATCTTAGAAGAACGATTTGGTGACCAAGTTTATCCCCGGTTACAACTAGGAGGCTACCAGTTTAAAATA
TTTGCGATAGCAGCAAGTTCATTCAAAAACTGCTTTTTGTTAGATTCAGATAATATACCCTTGCGAAAGATGGAT
AAGATATTCTCAAGCGAACTATACAAGAATAAGACAATGATTACTTGGCCAGACTTTTGGCTTCGAAGTACCAGC
CCGCACTATTATCATAACATCACCAAGACCCCTATTGGTGATAAAAGAGTACGATATTTTAACGATTTCTACACG
AATCCAAACGAGTATTACTACGGAGATGAGGATCCCAGATCAGAAATACCGTTTCACGATAGGGAGGGCACGATA
CCTGACTGGACCACAGAAAGTGGACAATTGGTTATAAATAAAGAAGTTCATTTTCCTGCAATTCTTCTAGGTCTT
TTCTACAACTTTAACGGGCCGATGGGCTTCTACCCTCTTCTCTCCCAAGGTGGAGCCGGAGAGGGAGATAAAGAC
ACATTCGTTGCAGCTTCCCATTACTACAATCTCCCGTATTACCAAGTTTATAAGAACTGTGAGATGTTGTATGGC
TGGGTTGATCATGCCAATAGTGGAAGAATTGAACACTCTGCAATCGTCCAGTACAATCCCATCGTGGACTATGAG
AATCTCCAAAGTGTCAAGGCGAAAGCGGAGATTATACTGAAAAACCACGAGCCAGATTCAAGGAAGAAAAGTTCG
AAACCAAAGTCTTACAGCAAAACAAGGCTTTCCACTCATGTCAAGGGTTCAATATACTCATATAGACGGCTCTTC
AGAGACTCGTTCAATAAAGCCAACTCTGATGAAATGTTCCTACATTGTCATACCCCTAAGATTGAACCTTACAGG
ATAATGGAAGATGATTTGACACTTGGGAGGAACAAAGAAGCTAAACAACGATGGTATGGAGGACGAAAGAATAGA
GTGCGGTTTGGATATGATGTGGAACTTTACATCTGGGAGCTAATTGACCAATATATCTGTGACAAGAATATCCAA
TACAAAATTTTTGAAGGTAAGGACAGAGATGCTCTTTGTGGGAGTTTCATGAGGGAACAGTTGGGATTTTTACGC
TCCACGGGGGATTAG

AA

MFNSLAPMRLKKLLKVFCASVVLLAATSVVLFFHFGGQIIIPIPERTVTLSTPPANDTWQFQQFFNGYLDALLEN
NLSYPIPERWNHEVTNVRFFNRIGELLSESRLQELIHFSPEFIEDTSDKFDNIVEQIPAKWPYENMYRGDGYVIV
GGGRHTFLALLNINALRRAGNKLPVEVVLPTYDDYEEDFCENHFPLLNARCVILEERFGDQVYPRLQLGGYQFKI
FAIAASSFKNCFLLDSDNIPLRKMDKIFSSELYKNKTMITWPDFWLRSTSPHYYHNITKTPIGDKRVRYFNDFYT
NPNEYYYGDEDPRSEIPFHDREGTIPDWTTESGQLVINKEVHFPAILLGLFYNFNGPMGFYPLLSQGGAGEGDKD
TFVAASHYYNLPYYQVYKNCEMLYGWVDHANSGRIEHSAIVQYNPIVDYENLQSVKAKAEIILKNHEPDSRKKSS
KPKSYSKTRLSTHVKGSIYSYRRLFRDSFNKANSDEMFLHCHTPKIEPYRIMEDDLTLGRNKEAKQRWYGGRKNR
VRFGYDVELYIWELIDQYICDKNIQYKIFEGKDRDALCGSFMREQLGFLRSTGD

Peptide

MFNSLAPMRLKKLLKVFCASVVLLAATSV

SEQ0014

Homolog of YAL063C; **FLO9**

Lectin-like protein with similarity to Flo1p, thought to be expressed and involved in flocculation

S. cerevisiae null mutant: filamentous growth: decreased; haploid invasive growth: absent

chr1-4_0584

ORF

ATGTTTGAGAAGAGTAAATTTGTGGTTTCGTTTCTGCTTTTACTGCAGCTATTTTGTGTCCTTGGTGTACATGGA
CAAGAAAGTGGTAACGGTACCACTAGTGACACTGCCTACGCATGCGACATCGGTGCAACACCTTTTGACGGATTC
AATGCTACTATTTATCAGTACCAGGCAAGTGATGACAATTCAATACAGGACCCGGTTTTTATGAGCACTGGGTAC
TTGCAACGTAACCAGTTGCATAGTACAACTGGAGTGACTAATCCGGGGTTCAACATCTTCACTGCAGGCGTTGCC
ACCACAACACTTTATGGGATACCAAACGTCAACTACCAAAATATGCTTTTAGAATTGAAAGGCTATTTCAGAGCT
GACGCCTCTGGAAATTATGGACTATCACTAAGGAACATCGATGATAGTGCGATTCTTTTCTTTGGGAGGGAGACT
GCGTTTGAATGTTGCAATGAAAATTTGATTCCCTTGGACGAAGCCCCAACTGATTACTCACTTTTTACCATTAAA
GAAGGGGAAGCTTCAACAAACCCCGACTCTTATACGTACACTCAATACTTGGAGGCTGGGAGATACTATCCCGTA
AGAACTTTCTTTGCCAACATAAGAACGCGTGCCGTGTTCAACTTTACGATGACACTGCCAGACGGGAGTGAACTA
ACAGATTTTCAAAACTACATTTTTCAATTTGGAGCTCTTAACCAACAACAATGCCAAGCAGAAATAGTTACCAGG
GAAAATTACACAACCACCACCGAACCATGGACTGGTACCTTTGAAGCCACGACCACGGTGATTCCAAGTGGCACT
GAGCCAGGTACCGTGATTGTCCAGACTCCTTATTCTACTATCGACTCCACATCTACTTGGACCGGTACTTTCACC
ACTTTCACCACCGATGCTGATGGAAGCACCATCGCCGTTGTGCCATCGTCGACTATTGATGATCACTTTGCGTCG
ACTGAAACTGTTTTGACAGACACTGCTATTTCAACTACTGTTATTACCGTGACCAGTTGTGGTACCAGTAAGTGT
ACCAAGACTACAGCACTAACAGGGGTGACTCAGCGAACTCTCACAATCGATGATAGAACTACTGTCGTGACGACT
TACTGTCCTTTGCCAACTGACGTTGCCACTATAAAAACAGCAAGTGTTAGCGGAAGTGAGGTTGTGCAAACAATC

TACACTGCGAAACACAGCCAGGCCGTTTCTTATGTTCATCCATCAACTGTTACAATCACAAGAGAGGTATGTGAT
GCCCAAACTTGCACGCAGGCTACTATTGTGACTGGAGAAATTCTGCAAACTACTGTTGTGGACTCTGGTTCTACA
ACTGTTGTTCCAAAGTATGTTCCGGTGGAGACGCATGAACCAACATTTGAATTGAGTACTCTTTAA

AA

MFEKSKFVVSFLLLLQLFCVLGVHGQESGNGTTSDTAYACDIGATPFDGFNATIYQYQASDDNSIQDPVFMSTGY
LQRNQLHSTTGVTNPGFNIFTAGVATTTLYGIPNVNYQNMLLELKGYFRADASGNYGLSLRNIDDSAILFFGRET
AFECCNENLIPLDEAPTDYSLFTIKEGEASTNPDSYTYTQYLEAGRYYPVRTFFANIRTRAVFNFTMTLPDGSEL
TDFQNYIFQFGALNQQQCQAEIVTRENYTTTTEPWTGTFEATTTVIPSGTEPGTVIVQTPYSTIDSTSTWTGTFT
TFTTDADGSTIAVVPSSTIDDHFASTETVLTDTAISTTVITVTSCGTSKCTKTTALTGVTQRTLTIDDRTTVVTT
YCPLPTDVATIKTASVSGSEVVQTIYTAKHSQAVSYVHPSTVTITREVCDAQTCTQATIVTGEILQTTVVDSGST
TVVPKYVPVETHEPTFELSTL

Peptide

MFEKSKFVVSFLLLLQLFCVLGVHGQ

SEQ0015

Homolog of **FLO1** (YAR050W)

Lectin-like protein involved in flocculation, cell wall protein that binds to mannose chains on the surface of other cells, confers floc-forming ability that is chymotrypsin sensitive and heat resistant; similar to Flo5p

S. cerevisiae null mutant: flocculation: absent; oxidative stress resistance: absent; resistance to ethanol: decreased; toxin resistance: absent

S. cerevisiae over expression: increased flocculation

chr3_1145

ORF

ATGAAATTTCCTGTGCCACTTTTGTTTCTACTGCAGCTGTTCTTTATTATTGCAACACAAGGAGATGAGAGTGGA
AATGGTGATGAAAGTGATACTGCTTATGGATGCGATATCACTAGCAATGCTTTTGACGGATTTGATGCTACCATC
TATGAGTACAATGCAAATGATTTGAAACTTATACGCGACCCAGTTTTTATGAGTACTGGGTACCTGGGACGCAAC
GTACTGAATAAAATATCTGGGGTTACTGTTCCTGGGTTTAACATCTGGAATCCACGTTCTAGGACTGCAACGGTA
TATGGGGTCCAGAACGTAAACTATTATAACATGGTCCTTGAATTAAAAGGCTATTTCAAAGCTGCTGTTTCAGGA
GACTACAAACTAACATTGAGCAACATTGATGATAGTTCGATGCTTTTCTTTGGGAAAAATACTGCCTTTCAATGT
TGTGATACAGGCTCAATTCCCGTGGATCAAGCACCAACAGACTATTCGCTGTTTACTATCAAACCATCCAATCAG
GTAAACTCTGAGGTGATTTCATCCACGCAGTACTTAGAGGCTGGAAAGTACTATCCTGTAAGAATCGTCTTTGTA
AATGCTTTGGAACGAGCGCTTTTCAATTTCAAGCTTACAATTCCAAGCGGAACAGTATTGGATGACTTTCAAGAT
TACATCTACCAATTTGGAGCGCTAGATGAAAACTCCTGTTACGAAACAACTGTTTCAAAAATTACAGAATGGACA
ACGTATACTACGCCATGGACTGGAACATTCGAAACAACAAGAACAATCACTCCAACAGGAACAGAAGGAACAGTT
GTGATCGAGACTCCAGAGAGTTACGTGACTACCACCCAACCATGGACGGGTACTTACGAGACCACTTACACTGTC
CCACCCACCGGCACTGAGCCAGGCACCGTCATTATTGAGACTCCGGAGATTATCGATTGTGAAGCTGTTTGCTGT
GGACCATTTCTAACAGCATTCTCGTTCAGAAAAAGAGAAGAGTGCCAGTGTGAGAACATTTGCTGTCCAGGAGAT
ACAAACTGTGAAACATATGTAACCACAACCCAACCATGGACCGGTACTTATGAGACTACTTACACCGTCCCACCC
ACCGGCACTGAACCAGGCACCGTCATTATTGAGACTCCGGAGAGTTATGTAACCACAACCCAACCATGGACCGGT
ACTTACGAGACCACTTATACCGTCCCACCCACCGGCACTGAGCCAGGCACCGTCATTATTGAGACTCCAGAGAGT
TACGTGACTACCACCCAGCCATGGACCGGTACTTACGAGACCACTTACAGTGTCCCACCAAGTGGCACTGAGCCA
GGTACTGTTGTGATCGAGACTCCAGAGAGTTATGTGACTACCACCCAGCCATGGACGGGTACTTACGAGACCACT
TACAGTGTCCCACCAAGTGGCACTGAGCCAGGCACCGTTGTCATTGAGACTCCTGAGGCGTCCACTGCCCGCACA
AAGTTTACCACGGTTACATCTTCTTGGACTGGGGTATTCACAACTACCAAGACTCTACCAGCTTCTGGTACCGAG
CCAGCGACCATAGTAATTCAGACCCCAACTGGATACTTCAATACTTCAAGCTTAGTATCCACGAGAACAAAAACT
AATGTGGACACTGTTACGCGTGTTATCCCATGTCCTATTTGTACGGCACCTAAAACAATTACAGTTGTTCCTGAA
GAACCAAACGAAAGTGTGAGTGTCATTATTTCACAACCTCAAAGTTCCTCAACCGATACAACTTTATCCAAGCCT
GATAGTGTGAGAGTCATTTCCCAGCCCGAAACTGCTTCTCAAATGGATACTTCTTTATCCAAGACAGACAGTGCA
GTAATTTCTACCGAGACAGCCGGAAATAATATTATTCCTCTTGCAGGTAGTCATAGCTACAACACGATTGTGACT
ACAGTTACTGACTCACCTCAAGTTGCTCAGAGCACTACTGCTACAAGCAGCTCTAATGTTCATCTTACTATTAGC
ACCCAAACAACAACGCCCAGTTTGGTGTATTCCAGCTCACTTTCAACTGTTCACCAAGTCTCACCTTCAAATGGA
GGGTTCCGGTCAAGTATAACAGTGCACCCTCTATTGAGTGTGATTGGTGCCATCTTTGGTGCCCTGTTCATGTAA

AA

MKFPVPLLFLLQLFFIIATQGDESGNGDESDTAYGCDITSNAFDGFDATIYEYNANDLKLIRDPVFMSTGYLGRN
VLNKISGVTVPGFNIWNPRSRTATVYGVQNVNYYNMVLELKGYFKAAVSGDYKLTLSNIDDSSMLFFGKNTAFQC
CDTGSIPVDQAPTDYSLFTIKPSNQVNSEVISSTQYLEAGKYYPVRIVFVNALERALFNFKLTIPSGTVLDDFQD
YIYQFGALDENSCYETTVSKITEWTTYTTPWTGTFETTRTITPTGTEGTVVIETPESYVTTTQPWTGTYETTYTV

PPTGTEPGTVIIETPEIIDCEAVCCGPFLTAFSFRKREECQCENICCPGDTNCETYVTTTQPWTGTYETTYTVPP
TGTEPGTVIIETPESYVTTTQPWTGTYETTYTVPPTGTEPGTVIIETPESYVTTTQPWTGTYETTYSVPPSGTEP
GTVVIETPESYVTTTQPWTGTYETTYSVPPSGTEPGTVVIETPEASTARTKFTTVTSSWTGVFTTTKTLPASGTE
PATIVIQTPTGYFNTSSLVSTRTKTNVDTVTRVIPCPICTAPKTITVVPEEPNESVSVIISQPQSSSTDTTLSKP
DSVRVISQPETASQMDTSLSKTDSAVISTETAGNNIIPLAGSHSYNTIVTTVTDSPQVAQSTTATSSSNVHLTIS
TQTTTPSLVYSSSLSTVHQVSPSNGGFRSSITVHPLLSVIGAIFGALFM

Peptide

MKFPVPLLFLLQLFFIIATQGD

SEQ0016

Homolog of YHR132C; **ECM14**

Putative metalloprotease with similarity to the zinc carboxypeptidase family, required for normal cell wall assembly

S. cerevisiae null mutant: abnormal polyphosphate accumulation

chr3_0419

ORF

ATGCTCTATAAGACCACCTTGTCAATAGCACACACGAGTGTGATATTGTTGTCATTGATAACCGCCATAAGTTGC
TTTGAGTTGCATCTTCCTCAGAAGGTTTCTCATATAGTAGACAGTTTACAATATACTTGCGGCCAATTTTTGCAA
AAGCAGCAGATCTTTGCACTCTATAACAAGCAAAATTTCACCGAAATAGTGAACCAGAATATCAAGGGAATAGAG
GAGAGAGTTTTGTCTGAGTTGCTTGAAGAAAGATTAGAGAATGAATCCCAGAATGATTATTATACCGCCAATTCT
CAAAATTGGCCTATCGACTTGGATCAGTACTCAGAATCATTTGTAATAAGGATCACATCTGAAGATGAGTTTATC
AAGTACTTGATCTTCAAGGAAGCTAAAGCTTTGCATATTTCCATATGGGAGCAATCTGTTGGTTTGATAGATTTG
AAGGTTGACCGTGATCAGATGCACCGCCTACTTTACAACGTGGAGTCACGCATACTGGAACGAAGAACGAGAAGT
GTTGACAGTCCAGTTTCTGAATATAAAGTACAATTGATGATTGGAGATCTTCCACAGCGAATCTACGAAACATAT
CCTTCGACAAAAGTGACATCTTTGCAAGCCCTAGGAGAGTTCCCTTCTTTCCAGAACCTAAGTAATGCTTTTTTT
GAGGATTTTAGAACGCTGGAAACTATATACGACTGGTTCGAAGAAATACAGAAGGAATTTCCTAAGCTAGTGTCG
ATCAACTGGATTGGGCAAACTTATGAAGGTCGTGATCTGAAGGCTCTTCACGTTAGAGGGAAGCACTCTGGCAAC
AAAACAGTAGTCGTTACAGGTGGAATGCATGCGCGTGAATGGATATCAGTAACCAGTGCATGCTATGCCGTTCAC
AAACTGCTCCAAAACTATGCTGACGGACACCACAAGGAAGCGAAATACCTGGACAAGTTGGACTTTTTGTTTGTT
CCAGTTTTGAATCCTGATGGATACGAATATAGCTTTAACGAAGACAGGTTGTGGAGGAAGAACAGACAAGAAACT
TATATGCCCCGATGTTTTGGTATAGACATTGACCATTCATTTGATTATCATTTCGTGAAATCAGAAGACTTACCC
TGTGGAGAGGAATATTCGGGTGAGTCCCCTTTCGAAAGTATAGAAAGTGAAGTGTGGAATAATTTCCTGAACAGA
ACCAAAGAAGAACATAAGATCTACGGCTATATCGACTTACACTCGTATTCGCAAACGGTGCTGTATCCCTATGCG
TACTCATGCGAAATCTTACCAAGGGACGAGGAAAACCTGATTGAGCTAGGTTACGGTATTGCAAGGGCCATAAGA
AAGAGTACAGGGAAAAAATATCAAGTGTTGAAGGCATGCGAAGACAGGGATGCAGATCTATTGCCTGATTTGGGA
GGAGGAACCGCTTTAGATTATATGTACCACAACCGTGCATACTGGGCGTTTCAGATCAAATTGAGGGATTCCGGT
AATCATGGCTTTCTCCTTCCCAAAAAGTTTATATACCCAGTTGGAACAGAGGTTTATGCCTCAATTCAGTACTTT
TGTTCTTTTGTGCTGAATTTAGAAGGCTAA

AA

MLYKTTLSIAHTSVILLSLITAISCFELHLPQKVSHIVDSLQYTCGQFLQKQQIFALYNKQNFTEIVNQNIKGIE
ERVLSELLEERLENESQNDYYTANSQNWPIDLDQYSESFVIRITSEDEFIKYLIFKEAKALHISIWEQSVGLIDL
KVDRDQMHRLLYNVESRILERRTRSVDSPVSEYKVQLMIGDLPQRIYETYPSTKVTSLQALGEFPSFQNLSNAFF
EDFRTLETIYDWFEEIQKEFPKLVSINWIGQTYEGRDLKALHVRGKHSGNKTVVVTGGMHAREWISVTSACYAVH
KLLQNYADGHHKEAKYLDKLDFLFVPVLNPDGYEYSFNEDRLWRKNRQETYMPRCFGIDIDHSFDYHFVKSEDLP
CGEEYSGESPFESIESEVWNNFLNRTKEEHKIYGYIDLHSYSQTVLYPYAYSCEILPRDEENLIELGYGIARAIR
KSTGKKYQVLKACEDRDADLLPDLGGGTALDYMYHNRAYWAFQIKLRDSGNHGFLLPKKFIYPVGTEVYASIQYF
CSFVLNLEG

Peptide

MLYKTTLSIAHTSVILLSLITAISCF

SEQ0017

Homolog of YMR200W; **ROT1**

Essential ER membrane protein; may be involved in protein folding; mutation causes defects in cell wall synthesis and in lysis of autophagic bodies, suppresses tor2 mutations, and is synthetically lethal with kar2-1 and with rot2 mutations

S. cerevisiae null mutant: cell shape: abnormal; chitin deposition: increased; inviable; killer toxin resistance: increased; liquid culture appearance: abnormal; resistance to hygromycin B: decreased; resistance to sodium dodecyl sulfate: decreased
S. cerevisiae conditional mutant: increased heat sensitivity
S. cerevisiae reduction of function mutant: growth rate in exponential phase: decreased; resistance to tunicamycin: decreased; vegetative growth: decreased

FragB_0048

ORF

ATGGTTTTGATACAAAATTTTTTGCCGCTGTTTGCGTACACTTTGTTCTTCAACCAGCGGGCCGCCCTTGCTGAT
GACAGCGTCCCAGAATCCGAGCTTACTATCGTCGGAACATGGTCATCCAAATCCAACACAGTGTTCACAGGATCA
GGCTTTTATGATCCAGTGGACGAGCTATTGATTGAACCAGATTTGCCTGGGATCTCCTACTCTTTCACTGATGAC
GGATACTTTGAAGAGGCCCTTTACCAGGTTGCGGGAAATGCGAAGGATCATCATTGTCCCACAGCTGTTTTAATT
TTCCAACACGGTACCTACAGAGAGTTAGATAACGGAACTCTCGTTTTGGAGCCATATGATGTCGACGGAAGACAG
CTTTTATCGCAACCATGCGAGGATAAAGGGATTTCAACCTATTCTCGTTACAACCAAACAGAGGTGTTCCGTAAC
TATGAAGTGTCCCTGAATACGTACCACGGTCGTATCCAGCTGCAGCTATATGCCAGTGATGGTGTGAAACAGCCT
CCCCTCTACTTAGCGTACAGACCTCCCAACATGTTGCCTACGATCGTACTAAACCCCACTGCAGCAAGTGATGAG
GCCCAAGCAACCGCCACTGGTGCCTCTGCCAAGATAAAAAGATCACTAGAGAACAGATACAGGACCAACGCTGTC
AAGGAATCTTCTTTGAATTACACTCTATGGTGGTGGCTTGGGGCAATTCTGATGGGTGCTGGATCCACCATCTAC
TTCCTGTATTAA

AA

MVLIQNFLPLFAYTLFFNQRAALADDSVPESELTIVGTWSSKSNTVFTGSGFYDPVDELLIEPDLPGISYSFTDD
GYFEEALYQVAGNAKDHHCPTAVLIFQHGTYRELDNGTLVLEPYDVDGRQLLSQPCEDKGISTYSRYNQTEVFRN
YEVSLNTYHGRIQLQLYASDGVKQPPLYLAYRPPNMLPTIVLNPTAASDEAQATATGASAKIKRSLENRYRTNAV
KESSLNYTLWWWLGAILMGAGSTIYFLY

Peptide

MVLIQNFLPLFAYTLFFNQRAALAD

Homolog of YDR304C; **CPR5**

Peptidyl-prolyl cis-trans isomerase (cyclophilin) of the endoplasmic reticulum, catalyzes the cis-trans isomerization of peptide bonds N-terminal to proline residues; transcriptionally induced in response to unfolded proteins in the ER
S. cerevisiae null mutant: viable
S. cerevisiae overexpression mutant: decreased vegetative growth, viable

2 homologues

SEQ0018

chr1-1_0267

ORF

ATGAAATTGTTGAACTTTCTGCTTAGCTTCGTAACTCTGTTCGGACTATTATCAGGTTCTGTGTTTGCACAAGAA
CTGAAAGATCTGCCAGAAATTACCCAAAAGGTCTATTTTGATATTAACGAGGGAGACAAGTTTTTAGGAAGAATT
GTCATTGGTCTATTCGGAGATGTCGTTCCAAGAACTGTGGAGAACTTCTACCAATTGGCCATCAGTGAAGATCCT
GACTTCGGTTATAAGGGCTCCAAATTCCACCGTGTCATCAAGGACTTTATGATCCAAGGTGGTGATTTTACTTCA
GGAGACGGCAGAGGAGGAAAATCTATCTACGGTAGCAGATTCAAGGACGAGAACTTCAAACTCAAGCACGATCGC
CCTGGGCTGTTATCCATGGCCAATGCCGGAAAGGATACCAATGGATCTCAGTTTTTCATCACCACCGTGGTGACT
TCGTGGTTGGACAACAAACATGTTGTCTTTGGCTATGTCGTTGATGGGTTCGATGTTGTCAAGAAGATTGAGTCC
ACCAAGACTTCCGGTAGTGCTCCAAAGGAAGACATTGTCATCGTTGACTGTGGAGAAGTCAAAGAGTCTGTTGAG
GATGTCATCAAGGACGATGCTGAAGATTACGCCGAGGCCATCGAGGAGGTCGTGAAAGATGAGTTGTAG

AA

MKLLNFLLSFVTLFGLLSGSVFAQELKDLPEITQKVYFDINEGDKFLGRIVIGLFGDVVPRTVENFYQLAISEDP
DFGYKGSKFHRVIKDFMIQGGDFTSGDGRGGKSIYGSRFKDENFKLKHDRPGLLSMANAGKDTNGSQFFITTVVT
SWLDNKHVVFGYVVDGFDVVKKIESTKTSGSAPKEDIVIVDCGEVKESVEDVIKDDAEDYAEAIEEVVKDEL

Peptide

MKLLNFLLSFVTLFGLLSGSVFAQ

SEQ0019

chr4_0545

ORF

ATGAATTTAACTTTGATCTTCACCCTGATCAGCCTTTTGCTTGGTGTTTGGAGTGCACCGATAGAAGAACCTCCT
TATACGCATTTCGTGACGTTTCACATCACCCAGGACGATCAGCCTTTGGGTGATCTGGTGCTGGGACTGTACGGT
ACTGTCGTTCCTAAGACCGTGAAGAACTTCTACCAGCTCGCTGCTATGACACCCGGGTTTGGATACGATGGATCC
TTATTCCATAGAGTCATTGAGGATTTTATGATTCAAGGAGGAGATTTTACATCTGGAGACGGGAAAGGTGGTCGT
TCGATATATGGTGAGAACAAAGGAGATTTTCCTGATGAGAATTTTGTTTTGAAGCATGACAGATTAGGAAGAGTA
AGCATGGCAAACTCTGGAAAGGATACCAATGGATCTCAATTCTTTATTACTGCCAAAGCTACTTCTTGGTTAGAT
GGACATCACGTTGTGTTTGGCCAATTGGTTGATGGGTTTGAAGTGTTTAGTAGCATCATCAAAACTCCTACCAAC
TCAAGGGATAATAAGCCAACGGCTAAAATTATGATTAAAAACGCAACCGTAAAGGAAGTTGATGAGTCAGATATA
AAGTCGAAACAAACTGACTCACCAAAGGCAGAAGAATCAAACGACAAAGGGTCGGAAAGTAAGCAATCCGTCGAA
GGCACCGCTCATACTGGTTTATGGGCATTGGTATTTTTTGCAGCAATTGGAATTGTAGCTTTCCTGATTTACAAG
AAGTCTATCCGCCACACTCTTTACACCAAATTGAACAAATGA

AA

MNLTLIFTLISLLLGVWSAPIEEPPYTHFVTFHITQDDQPLGDLVLGLYGTVVPKTVKNFYQLAAMTPGFGYDGS
LFHRVIEDFMIQGGDFTSGDGKGGRSIYGENKGDFPDENFVLKHDRLGRVSMANSGKDTNGSQFFITAKATSWLD
GHHVVFGQLVDGFEVFSSIIKTPTNSRDNKPTAKIMIKNATVKEVDESDIKSKQTDSPKAEESNDKGSESKQSVE
GTAHTGLWALVFFAAIGIVAFLIYKKSIRHTLYTKLNK

Peptide

MNLTLIFTLISLLLGVWSA

Homolog of YLR120C; **YPS1**

Aspartic protease, attached to the plasma membrane via a glycosylphosphatidylinositol (GPI) anchor

S. cerevisiae null mutant: viable; decreased competitive fitness

2 homologues

SEQ0020

chr4_0584

ORF

ATGTTGAAGGATCAGTTCTTGTTATGGGTTGCTTTGATAGCGAGCGTACCGGTTTCCGGCGTGATGGCAGCTCCT
AGCGAGTCCGGGCATAACACGGTTGAAAAACGAGATGCCAAAAACGTTGTTGGCGTTCAACAGTTGGACTTCAGC
GTTCTGAGGGGTGATTCCTTCGAAAGTGCCTCTTCAGAGAACGTGCCTCGGCTTGTGAGGAGAGATGACACGCTA
GAAGCTGAGCTAATCAACCAGCAATCATTCTACTTGTCACGACTGAAAGTTGGATCACATCAAGCGGATATTGGA
ATCCTAGTGGACACAGGATCCTCTGATTTATGGGTAATGGACTCGGTAAACCCATACTGCAGTAGCCGTTCCCGC
GTGAAGAGAGATATACACGATGAGAAGATCGCCGAATGGGATCCCATCAATCTCAAGAAAAATGAAACTTCTCAG
AATAAAAATTTTTGGGATTGGCTCGTTGGAACTAGCACTAGTTCTCCTTCCACCGCCACGGCAACTGGTAGTGGT
AGTGGTAGTGGTAGTGGTAGTGGTAGTGCTGCCACAGCCGTATCGGTAAGTTCTGCACAGGCAACATTG
GATTGCTCTACGTATGGAACGTTTGATCACGCTGATTCCTCGACGTTCCATGACAATAATACAGACTTTTTCATC
TCATACGCTGATACCACTTTTGCTTCAGGAATCTGGGGTTATGACGACGTCATTATCGACGGCATAGAGGTGAAA
GAACTTTCCTTCGCCGTTGCAGACATGACCAATTCCTCTATTGGTGTGTTAGGTATTGGACTGAAAGGCCTAGAA
TCCACATATGCTAGTGCATCTTCGGTCAGTGAAATGTATCAGTATGACAATTTGCCAGCCAAGATGGTCACCGAT
GGGTTGATCAACAAAAATGCATACTCCTTGTACTTGAACTCCAAGGACGCCTCAAGTGGTTCCATCCTCTTTGGA
GGTGTGGATCATGAAAAATATTCGGGACAATTGTTGACAGTTCCAGTCATCAACACACTCGCTTCCAGTGGTTAC
AGAGAGGCAATTCGTTTACAAATTACTTTAAATGGAATAGATGTGAAAAAGGGTTCTGACCAGGGAACTCTTTTA
CAAGGGAGATTTGCTGCATTATTGGACTCTGGAGCTACGCTAACGTATGCTCCTTCTTCTGTTTTAAATTCAATT
GGCCGGAACCTGGGCGGCTCCTATGATTCGTCAAGACAAGCTTATACCATTCGTTGTGTTTCTGCATCAGATACC
ACTTCTCTGGTATTCAATTTTGGGGGTGCTACAGTGGAAGTTTCCCTGTACGATCTACAGATTGCAACATATTAC
ACCGGGGGGAAGTGCCACGCAATGTCTTATTGGAATATTCAGCTCTGGAAGTGATGAGTTTGTGCTCGGTGATACC
TTCTTGAGGTCAGCCTACGTGGTTTACGATCTTGATGGGCTTGAAGTGTCGCTTGCCCAAGCCAACTTCAACGAA
ACCGATTCTGATGTTGAGGCTATTACCTCCAGTGTACCTTCCGCTACTCGTGCATCCGGATACAGTTCTACATGG
TCTGGTTCTGCCAGCGGTACAGTTTACACTTCGGTTCAGATGGAATCCGGTGCTGCTTCCAGCTCCAACTCTTCT
GGATCGAATATGGGTTCCTCTTCCTCATCGTCCTCTTCATCGTCCTCGACTTCCAGTGGAGACGAAGAAGGAGGG
AGCTCCGCCAACAGGGTCCCCTTCAGCTACCTTTCTCTCTGTTTGGTAGTTATTCTCGGCGTGTGTATAGTATAG

AA

MLKDQFLLWVALIASVPVSGVMAAPSESGHNTVEKRDAKNVVGVQQLDFSVLRGDSFESASSENVPRLVRRDDTL
EAELINQQSFYLSRLKVGSHQADIGILVDTGSSDLWVMDSVNPYCSSRSRVKRDIHDEKIAEWDPINLKKNETSQ

NKNFWDWLVGTSTSSPSTATATGSGSGSGSGSGSGSAATAVSVSSAQATLDCSTYGTFDHADSSTFHDNNTDFFI
SYADTTFASGIWGYDDVIIDGIEVKELSFAVADMTNSSIGVLGIGLKGLESTYASASSVSEMYQYDNLPAKMVTD
GLINKNAYSLYLNSKDASSGSILFGGVDHEKYSGQLLTVPVINTLASSGYREAIRLQITLNGIDVKKGSDQGTLL
QGRFAALLDSGATLTYAPSSVLNSIGRNLGGSYDSSRQAYTIRCVSASDTTSLVFNFGGATVEVSLYDLQIATYY
TGGSATQCLIGIFSSGSDEFVLGDTFLRSAYVVYDLDGLEVSLAQANFNETDSDVEAITSSVPSATRASGYSSTW
SGSASGTVYTSVQMESGAASSSNSSGSNMGSSSSSSSSSSTSSGDEEGGSSANRVPFSYLSLCLVVILGVCIV

Peptide

MLKDQFLLWVALIASVPVSGVMAA

SEQ0021

chr3_0299

ORF

ATGAACCCTAGCAGCTTAATTCTACTTGCACTCAGCATTGGCTACTCCATTGCTGAGTCAAATTTCTCTTTCAAA
CCCAGCAAGTTACCTCTCAAAAAACATCGTGATTCTTCTTCCCCGCATGAACGATTTCTTAAACGAGATGGACCC
TATCATCCGCTAGAAGCCGACGCTTACTTTTACTACACTACGTCTATATTGGTTGGATCAGAAGAAGAAAAAGTT
GAAGTAACAGTTGATTTAGGAACCTCTGATTTATGGGTCGTCGATTACAACACCGGTTTATGTGATAGATCCTTT
GACGAAACCTATCTTAAACGTAGTCTGGATACTTCTGAGGAAGATTATTCTGCTGGAGATCTTGGCTCCTCAGTC
GGTGTACGCAGCGCTAGAAAATTCTTGCGCAAAAGGGACACCAATCAAACTGAGGTTAATGAAGCTAACTATGGT
GCTTGTCCAAATTCGATTACCTTCAATCCAGAAAACTCGTCTTCTTTCCAGAGTAATGATACTGCTTTCAATATC
AGCTACTTTGATGGAACCAGTGCTAGTGGTTTTTGGGCTACTGATACAATTTACTTTGGTGACCTTGAGGTCAGC
GAGCAATTTTTTGGGCTGGCAAACTTAACAATAAGTTATGGAGGAGTCTTAGGTCTTGGCCCTTCCAACCTACAA
ACAACCAATGCTAACCCCAACGGTGAGGAATTCATTTACAGCGGAGTCTTAGATTCCATGCGTGATCAAGGGCTT
ATCAACTCGGCTTCTTTCTCAATCTATCTCAATCCAGAGAATTTCAGAGATGAAGATAACTATTCTAATGAAGGA
GCGATTTTGTTCGGAGCAATTGATAATGCGAAGATTGACGGGTCATTGAAGCTGTTACCATACGTGACTTCAGGT
GGACACTCTCAGATTGATGCTAATTTCACTTACATCACCTTGAATAATATTGCCGTGGCTGACAATGATACAGCC
CTGATCGTTGAGACCAACCCCCAATTGGCAATGTTGAATCCAAAGTTTATATACACCTATTTTCCAAACGAAGTA
TTGACCCGGCTGGTAAACTCTATTGACAATCTAGAATATGATCCTGTTGAGGGGTTATATAGGATAAGGAGAACA
AACATTAGGGATATTAACAAAAAAATCATAGAGTTTCAATTTGGTGACGAGATTGTGATACATTCTCCCTTATCA
AATTATCTGTCTGATACATGGGTTCCAAGCACAAACTACACCTATTTGGAGATTCAGGATAGCAGAGAGGATTTC
TTTATCCTTGGTAATGCATTTTTCAAGTCTGCGTATTTGTTTTTTGACAATGATAACAGTGAAGTCGGTATTGGC
CAACTAAAGGTTACCGATAAGGAGGACATCGTTCCAGTTGGTGAATTTTCTTTGGATCAAGATTCAGGGTACTCG
TCAACCTGGTCAACGTTCTCCTATGAAACTGGTTCAGCTCCCTTGGGTACGTCAACTTTCGAAACGAGTACAAAA
ACTAGTTCAGATGGAGCTGCCCCGTCGGTGTCTCACATTAACACTAGTTCCTACTTATTTGCGTTTGTACTACTT
TTCCTTTAG

AA

MNPSSLILLALSIGYSIAESNFSFKPSKLPLKKHRDSSSPHERFLKRDGPYHPLEADAYFYYTTSILVGSEEEKV
EVTVDLGTSDLWVVDYNTGLCDRSFDETYLKRSLDTSEEDYSAGDLGSSVGVRSARKFLRKRDTNQTEVNEANYG
ACPNSITFNPENSSSFQSNDTAFNISYFDGTSASGFWATDTIYFGDLEVSEQFFGLANLTISYGGVLGLGPSNLQ
TTNANPNGEEFIYSGVLDSMRDQGLINSASFSIYLNPENFRDEDNYSNEGAILFGAIDNAKIDGSLKLLPYVTSG
GHSQIDANFTYITLNNIAVADNDTALIVETNPQLAMLNPKFIYTYFPNEVLTRLVNSIDNLEYDPVEGLYRIRRT
NIRDINKKIIEFQFGDEIVIHSPLSNYLSDTWVPSTNYTYLEIQDSREDFFILGNAFFKSAYLFFDNDNSEVGIG
QLKVTDKEDIVPVGEFSLDQDSGYSSTWSTFSYETGSAPLGTSTFETSTKTSSDGAAPSVSHINTSSYLFAFVLL
FL

Peptide

MNPSSLILLALSIGYSIAE

Homolog of YLR121C; **YPS3**

Aspartic protease, attached to the plasma membrane via a glycosylphosphatidylinositol (GPI) anchor

S. cerevisiae null mutant: viable

2 homologues

SEQ0022

chr3_0303

ORF

ATGTTGCCCATCCGCTTATCCAAACTTCTGCTTTTGCTCTCCTTAAAGTTGAAATTGGGTACAGCTGAAGAAAAA
TACCAAAAGTTGGATTTAAAAAGAATTGACAAAGACTATTATGCCGTCGATGTCAAAGTCGGCTCCGATGAGCAG
GAGATCAAAGAGGTACTAATAGATACGGGTTCATCTGATTTCTGGATCTTGGACAAATCGTTCTGTAATTCTCCA
ACATCAGAGGAAGAAGAGAACAGTAACGGGCGTAGCAACAAGGAAAGCTGTGGAGTCTATGGCTCGTTCGACTCC
AACAAGTCAGAGACATTTCAGGCAACTGGCCAAGTATTTGACGCTGCTTACGGTGACACCACAGCCGAGTCGACA
GGATCTTCAGGAGTTCGAGGAATTGATCAGCTACGGGTAGGAGATATTCATATAGAAGAACTCTATTTTGGACTA
GTGACAAACACTACAAGTTTACCACCCGTTTTAGGAATTGCCCAGCTTTCCGAAGAGTTCAGCAACAACTCTTAT
CCTAACTTTCCATACCAGATGAAAGAGGAAGGTCTGATTGATGTTGTTGCATACTCTCTCTCCTTGGGCCAAAGT
AAAGGTGAACTACTGTTCGGGGCTATGGACCACTCAAAATATAATGGAACACTATTGAAAGCCCCTATATTGCAG
GCGGGCACACCAGGAATGCAAGTTCTTTTAACTGGAGTGGCCCTTACAAATGGTTCATCAAGCGTCTTCAATGAG
ACAGACAATAAAGGTTTTATCTACTTTGACAGTGGGACTACTGCTTCCACTCTGCCATCAGAGCACTTTGATGAT
CTTTTCAACCATCACGGATGGGCGTACGATGGTGATACATTGACATATTCGATTCAATGCGATAGTGAGGGAGAA
AAATCTTTACTTGACTTCACTTTAGAATATACCATTGCTGGTAATATTGTCATCAAAGTACCATTTGAAGACATT
ATTATGAAGAATGAAAATGATGGAGAATGCCTCTCAACCGTAATGGTGTCGAACCAGACTTCTTTTTCATATTCC
GATGACACACCCTTTTTCGTTGCTGGAGACGAAGTTCTGTTGAACGCTTATGTTGTTTACAACCTAGAAACACAA
GAGCTGGCCATTGCTCCAGCAGTGGATAATCCAGAAGATACTGAAGAAGATATTGAGATTATCTCCGCAGACTTT
GATATTTCAGAAGCCAGAGATTATAGCGTTGGATTAGAGTTCAGAAATACCACAATTCCAGCTACAACTGATTAC
TTGCCTTCCTCGATGTCGTCAGGTTCAGTCAGCGAAGAGACTGGTTCCAAGTCTGAGAGCTCTACTTCTGAGGAC
TTTGCTGCAGCCACGTTGAAACCATTTACATTTTGGGGTTTCGTCCTTTTTTTCTTTCACTTTTTGATTTGA

AA

MLPIRLSKLLLLLSLKLKLGTAEEKYQKLDLKRIDKDYYAVDVKVGSDEQEIKEVLIDTGSSDFWILDKSFCNSP
TSEEEENSNGRSNKESCGVYGSFDSNKSETFQATGQVFDAAYGDTTAESTGSSGVRGIDQLRVGDIHIEELYFGL
VTNTTSLPPVLGIAQLSEEFSNNSYPNFPYQMKEEGLIDVVAYSLSLGQSKGELLFGAMDHSKYNGTLLKAPILQ
AGTPGMQVLLTGVALTNGSSSVFNETDNKGFIYFDSGTTASTLPSEHFDDLFNHHGWAYDGDTLTYSIQCDSEGE
KSLLDFTLEYTIAGNIVIKVPFEDIIMKNENDGECLSTVMVSNQTSFSYSDDTPFFVAGDEVLLNAYVVYNLETQ
ELAIAPAVDNPEDTEEDIEIISADFDISEARDYSVGLEFRNTTIPATTDYLPSSMSSGSVSEETGSKSESSTSED
FAAATLKPFTFWGFVLFFFHFLI

Peptide

MLPIRLSKLLLLLSLKLKLGTAE

SEQ0023

chr3_0866

ORF

ATGTTAGTTGCTGTTGCCCTAGTGTTGTTACTGTCTACAGGCTATGCTGGAATCGTCGCCATTGATACCGAATAT
GAGTTCACCATTGGTTTTCTTAGTACGATAGAAATAGGGTTTCCCCCACAAAGCATAACGGCTCAATGGGATACA
GGATCGTCTGACCTCTTGGTCAATTCCGTGACAAATTCACAGTGTGCTCAGGACGGATGTAGCTTTGGTGCGTTC
GCCTTCAACAAATCCACCACTTATTCCAATATAACAAACCCTAACAACCTTCATGTTCAGTTCTCCTTTGCAAGC
GGCAGCGTGGTTGATGACAAACTTGTGAGTGACACTATTTTTGTAGATTCCAAGGTAATCCCACGGTTCAACTTT
GCACTGGTATCGAAGGGAGACCTGTATGGTGATAATATTTTTGGTATTGGACCGAGAGGGAACCAGGGAACATTC
GATTCCAATGGAACTCCAGCTTTCTATGATAGCTTTCCTTATCACTTGAAGGCCCTCGGTTTAATCAAACGACTG
GCTTACTCATTTTACACTGGGCCCACCCAGGGAAAGGTAGTATTTGGAGGGGTGGATCATGGAAAGTACGATGGG
TGCCTGGAGAAACTCGAGATTGTCCATGACAGTGCTTTTTACACACTGCTTGAGGCAATTGATGCTGATGATACT
TCCGTCTTGGATGAGCAAATTCATGTTTTGTTTGATACTGGTACCGCCTTGACACTTTTTCCCAGCTTTATTGCT
GAACAACTGGCTGATTTTTTGAAAGCTACATATTCGGACGAATACAATACGTTTGTAGTTCCCTGCGACCAAGAT
TTTGATTTTGAATACCTTCATTTTGGTTTTCGAAACATTAAGTTGTCGGTGCGCTTTAAGGATCTGTTTTTAGTC
ATTGACGATAGTGTTTGTGCTGTGGGGTTTGATCAAGGGGCAGATGCAAACAAGATAACCTTTGGGTCTTCACTT
TTAAGAAACTACTACACGCTTTATGATCTAGATTCCAAAGAAATTTTGATTGCTGACGTCAAGCCTGATGGTCCA
GACGATATTGAAATATTATCGGGTCCAGTTCAACGAATTTGTGATGAAAAGGGTGTCAGTAGCACTTCATTATGG
AGTAGTCTGAGTATAGAGTCCACGATAGAACCAGACACTTTTACCACTAAGCCTTCTATTTCCCAGACACGGTAT
TCGACTAGCTCCATTGGACCTCAAAACATTTCTAACTCTTTAGGTGAATATCCTTCAGTTTCCGTCACTCTTTCT
GAACACCATAACACTACTTCCATAGCCTCAAATTCCTCATTAGAAGGGAAACCAGCAACTCCAACTGTTACAGAC
CAGTCGTACCAGAATAATAAGACTACCTCTACCGTAATTGCTGTGAATTTGATTACCCATTCAACCACTCATTCA
ACCACTCATTCACCCACCTATTCAACCACTCATTCTAGTAATGGATCACGCTCAACTTTAGAGTACACTTCAACC
AAGGAATCCTCGGTGAAAATGCCCTGTGCGTTGATCATCTCCGACACAATTCCGTACAATGCTTCCGGTGGGAAT
AGTAGTTATGGATCGTTAATTTCAACATCTACGGTTAACAATGTTGAAGAGAATAATTCAAACACTGTTAGACCA
AGAAAAAGACAGACCTTCGTTTCGGGAACCACTTCCACGATACTACTCTATTCCTCAACTACGACCCAAGCATAT
CAGATGTTGTCCTCAACTTCAATCCCCCGACCATCCATAAAAGCCAGTTCAAATGCTGGTAGCCGCAAAACTTCA
AAGACATTATTAACATTTATCATATTGTATATTTTTTAG

AA

MLVAVALVLLLSTGYAGIVAIDTEYEFTIGFLSTIEIGFPPQSITAQWDTGSSDLLVNSVTNSQCAQDGCSFGAF
AFNKSTTYSNITNPNNLHVQFSFASGSVVDDKLVSDTIFVDSKVIPRFNFALVSKGDLYGDNIFGIGPRGNQGTF
DSNGTPAFYDSFPYHLKALGLIKRLAYSFYTGPTQGKVVFGGVDHGKYDGCLEKLEIVHDSAFYTLLEAIDADDT
SVLDEQIHVLFDTGTALTLFPSFIAEQLADFLKATYSDEYNTFVVPCDQDFDFEYLHFGFRNIKLSVRFKDLFLV
IDDSVCAVGFDQGADANKITFGSSLLRNYYTLYDLDSKEILIADVKPDGPDDIEILSGPVQRICDEKGVSSTSLW
SSLSIESTIEPDTFTTKPSISQTRYSTSSIGPQNISNSLGEYPSVSVTLSEHHNTTSIASNSSLEGKPATPTVTD
QSYQNNKTTSTVIAVNLITHSTTHSTTHSPTYSTTHSSNGSRSTLEYTSTKESSVKMPCALIISDTIPYNASGGN
SSYGSLISTSTVNNVEENNSNTVRPRKRQTFVSGTTSTILLYSSTTTQAYQMLSSTSIPRPSIKASSNAGSRKTS
KTLLTFIILYIF

Peptide

MLVAVALVLLLSTGYAG

Homolog of YCL043C; **PDI1**

Protein disulfide isomerase, multifunctional protein resident in the endoplasmic reticulum lumen, essential for the formation of disulfide bonds in secretory and cell-surface proteins, unscrambles non-native disulfide bonds

S. cerevisiae null mutant: inviable

S. cerevisiae conditional mutant: increased heat sensitivity; delayed endocytosis

2 homologues

SEQ0024

chr4_0844

ORF

ATGCAATTCAACTGGAATATTAAAACTGTGGCAAGTATTTTGTCCGCTCTCACACTAGCACAAGCAAGTGATCAG
GAGGCTATTGCTCCAGAGGACTCTCATGTCGTCAAATTGACTGAAGCCACTTTTGAGTCTTTCATCACCAGTAAT
CCTCACGTTTTGGCAGAGTTTTTTGCCCCTTGGTGTGGTCACTGTAAGAAGTTGGGCCCTGAACTTGTTTCTGCT
GCCGAGATCTTAAAGGACAATGAGCAGGTTAAGATTGCTCAAATTGATTGTACGGAGGAGAAGGAATTATGTCAA
GGCTACGAAATTAAAGGGTATCCTACTTTGAAGGTGTTCCATGGTGAGGTTGAGGTCCCAAGTGACTATCAAGGT
CAAAGACAGAGCCAAAGCATTGTCAGCTATATGCTAAAGCAGAGTTTACCCCCTGTCAGTGAAATCAATGCAACC
AAAGATTTAGACGACACAATCGCCGAGGCAAAAGAGCCCGTGATTGTGCAAGTACTACCGGAAGATGCATCCAAC
TTGGAATCTAACACCACATTTTACGGAGTTGCCGATGAGACTCATTTGGTGCACTGGATTGATATTGAGTCCAAA
CCTCTATTTGGAGACATTGACGGATCCACCTTCAAATCATATGCTGAAGCTAACATCCCTTTAGCCTACTATTTC
TATGAGAACGAAGAACAACGTGCTGCTGCTGCCGATATTATTAAACCTTTTGCTAAAGAGCAACGTGGCAAAATT
AACTTTGTTGGCTTAGATGCCGTTAAATTCGGTAAGCATGCCAAGAACTTAAACATGGATGAAGAGAAACTCCCT
CTATTTGTCATTCATGATTTGGTGAGCAACAAGAAGTTTGGAGTTCCTCAAGACCAAGAATTGACGAACAAAGAT
GTGACCGAGCTGATTGAGAAATTCATCGCAGGAGAGGCAGAACCAATTGTGAAATCAGAGCCAATTCCAGAAATT
CAAGAAGAGAAAGTCTTCAAGCTAGTCGGAAAGGCCCACGATGAAGTTGTCTTCGATGAATCTAAAGATGTTCTA
GTCAAGTACTACGCCCCTTGGTGTGGTCACTGTAAGAGAATGGCTCCTGCTTATGAGGAATTGGCTACTCTTTAC
GCCAATGATGAGGATGCCTCTTCAAAGGTTGTGATTGCAAAACTTGATCACACTTTGAACGATGTCGACAACGTT
GATATTCAAGGTTATCCTACTTTGATCCTTTATCCAGCTGGTGATAAATCCAATCCTCAACTGTATGATGGATCT
CGTGACCTAGAATCATTGGCTGAGTTTGTAAAGGAGAGAGGAACCCACAAAGTGGATGCCCTAGCACTCAGACCA
GTCGAGGAAGAAAAGGAAGCTGAAGAAGAAGCTGAAAGTGAGGCAGACGCTCACGACGAGCTTTAA

AA

MQFNWNIKTVASILSALTLAQASDQEAIAPEDSHVVKLTEATFESFITSNPHVLAEFFAPWCGHCKKLGPELVSA
AEILKDNEQVKIAQIDCTEEKELCQGYEIKGYPTLKVFHGEVEVPSDYQGQRQSQSIVSYMLKQSLPPVSEINAT
KDLDDTIAEAKEPVIVQVLPEDASNLESNTTFYGVADETHLVHWIDIESKPLFGDIDGSTFKSYAEANIPLAYYF
YENEEQRAAAADIIKPFAKEQRGKINFVGLDAVKFGKHAKNLNMDEEKLPLFVIHDLVSNKKFGVPQDQELTNKD
VTELIEKFIAGEAEPIVKSEPIPEIQEEKVFKLVGKAHDEVVFDESKDVLVKYYAPWCGHCKRMAPAYEELATLY
ANDEDASSKVVIAKLDHTLNDVDNVDIQGYPTLILYPAGDKSNPQLYDGSRDLESLAEFVKERGTHKVDALALRP
VEEEKEAEEEAESEADAHDEL

Peptide

MQFNWNIKTVASILSALTLAQAS

SEQ0025

chr1-1_0160

ORF

ATGAAAATATTAAGTGCATTGCTTCTTCTTTTTACGTTGGCCTTTGCTGAGGTCATTGAGCTGACCAACAAGAAC
TTTGATGACGTGGTTCTAAAGTCCGGAAAGTACACCTTAGTGAAGTTTTATGCCGATTGGTGTTCGCATTGCAAG

CGAATGAATCCAGAGTATGAAAAGCTGGCCGAAGAACTGAAGCCAAAGAGTGATCTGATCCAGATTGCCGCCATT
GATGCTAACAAATACTCAAAGTACATGAAGGTGTACGATATTGATGGATTTCCGACGATGAAATTGTTCACACCC
AAGGACATATCTCATCCGATCGAATTTTCTGGATCAAGAGACAGTGAAAGCTTTTTGAACTTTTTGGAGTCAACT
ACTGGTTTGAAGTTGAAGAAGAAGGCGGAAGTAAATGAGCCTTCGTTAGTTCAATCAATTGATGATTCAACAATA
GATGACCTTGTTGGGAAGGACAGGTTTATTGCAGTTACTGCTTCGTGGTGTGGATATTGCAAAAGATTGCATCCT
GAATGGGAGAAGTTAGCCAAAGCTTTTGGCAATGACGATATTGTCATCGGAAACGTTGTTACCGATGTTGTGGAA
GGTGAGAATATTAAGGCGAAGTATAAAGTTCAATCTTTCCCGACTATCCTGTACTTCACAGCAGGCTCAGATGAA
CCAATAAGATATGAATCTCCAGATAGAACTGTTGAAGGTTTGGTTAAATTTGTCAATGAACAAGCTGGCTTATTT
CGTGATCCAGATGGAACTTTGAATTTCAACGCCGGTCTAATTCCAGGAGTGAGTGATAAACTTACCAATTACATT
AAGGAAAAAGACCAAAGTTTATTGGAGTCAACGTTAGACTTGCTAAGCAACCATGAACATATCAAGGACAAATTC
AGTGTCAAATACCACAAGAAGGTCATAGAAAAGTTGTTGAAGGGAGAGAATGAATTCCTCAACAATGAAGTTGAG
AGGCTATCAAAAATGCTGAATACAAAGCTATCGGCAAACAATTCAGACTCTGTGATTAAAAGACTGAATATCTTG
CGCAATTTTATTGAGGCCAAAACTGAGTCAAAACCCCAGTTATTACACCAAGAGCTATAA

AA

MKILSALLLLFTLAFAEVIELTNKNFDDVVLKSGKYTLVKFYADWCSHCKRMNPEYEKLAEELKPKSDLIQIAAI
DANKYSKYMKVYDIDGFPTMKLFTPKDISHPIEFSGSRDSESFLNFLESTTGLKLKKKAEVNEPSLVQSIDDSTI
DDLVGKDRFIAVTASWCGYCKRLHPEWEKLAKAFGNDDIVIGNVVTDVVEGENIKAKYKVQSFPTILYFTAGSDE
PIRYESPDRTVEGLVKFVNEQAGLFRDPDGTLNFNAGLIPGVSDKLTNYIKEKDQSLLESTLDLLSNHEHIKDKF
SVKYHKKVIEKLLKGENEFLNNEVERLSKMLNTKLSANNSDSVIKRLNILRNFIEAKTESKPQLLHQEL

Peptide

MKILSALLLLFTLAFAE

SEQ0027

Homolog of YPR026W; **ATH1**

Acid trehalase required for utilization of extracellular trehalose

S. cerevisiae null mutant: viable

chr4_0342

ORF

ATGCTGAATAGAGTGTTGTTGGTTGCCCTTTCTTGTGTGGTTTTCTTTCACTTGGTTACAACCTTTCCTGTTGGC
ACTTCAAGTGATTCCTTGCAGATTAGGAATCTTCTGAGTCACAACTTCACCAGAGCTAATATCTCTGAAGGTCTC
AGTTCAGGAGCCACTTATTTTGTCGATGAAGACACTGAGACATATTACGACAAAGAGCTAAAGGTCCTGAGAACT
ACAAGATTCCCTCGCTATAACAATTACCAGCTGCAACCTTATGTTGCCAACGGTTATATTGGTTCCCGAATCCCC
CGAGTTGGGTCAGGATTTACCTACGACACTAGCGATAATAAGACTTCCGAGAATCTGAAAAACGGCTGGCCCTTA
TTCAACAAGAGGTACTCAGGGGCCTTTATTGCTGGATTTTTCAACTCTCAACCGACCGTACCTGAGACCAACTTC
GAAGAGTTGGAAAAAGACGGTTATGAGAGTATTATTGCCTCTATTCCTCAATGGACTAGCCTAGAGTTAACCGTT
AATGTGAATGGAACCAACCAAACTCTCAAAGCAGATGATGTCGATATAACCCACATCAGTGACTATTCTCAACAG
TTATCGTTGTTAGATGGAATTGTCACTACTAATTACACATGGTTGGGACTGGTCAATGTTTCAATCTCTGTCCTA
GCCCATAGAGACATTGTAAGTTTGGGATTCGTGTCTCTGGAGCTATCATCGCAGAAGAATATAACTGTTTCCGTC
ACTGATATTCTTGATTTCGCAACCAGCACCAGGTGTTCTTATTTGGACTCAGGAGTAAATGAACAGAGTATTTTC
ATGAAGGTTCAACCGTCAAATGTTCCGACGAATGCTACGATTTACTCTTCATTGATGAGCTCTAATTCGACCTCG
TCCCTTTTAAAGCAAAATCAAACGGTATCTCAAACACTTCGGGTAAATCTATCGAAGAACCAAGCTGCATCGTTT
CAAAAATACGTCGGCGTTGTCTCTGACGACTATCTAGATAGCATTGAAACTAACTTGACCTCTTACCAATTTGCT
AGAGAAACTGCTAAGTTTGCTGAAATAAAAGGGCGTTCGTGGATCTTGAAATCTCATAAAGAGGCATGGAACGAA
TTGTTAAATGGAAAGTCTATTGTCTTTCACGACAATGATTTTTTGACATTGGCTTCCGATTCAAGCATTTACCAC
TTGATGGCCAATACAAGATCAGAAGCCAATGGAGGCACTTCTGCTTTAGGGGTTTCCGGGCTGTCTTCGGATTCA
TACGGAGGAATGGTGTTTTGGGACACTGACTTTTGGATGCTACCGAGTGTCCAAGCCTTTAGTCCCCGACATGCC
GTTTCTTTATCCAAATTTAGAGACCATACCCATGATCAAGCTAAGAAGAATGCCCAAACTCGTGACATGAATGGG
GCAGTCTACCCCTGGACTAGTGGTAGATTTGGAAATTGTACCTCAACTGGTCCGTGTTATGATTACGAGTATCAT
ATCAATATAGACATTGCCTTCATGTTCTGGAAATTATATTTGGGAGGTGCAATTGATGATGATTATATGAAGGAA
TTTGGATACCCAATAATTGAAGATGTTGCCTCTTTTTTCGTGGATTACGTTGACTACAATTCAACTTTAGATAAA
TATACGACCAGAAATCTGACAGACCCAGATGAGTATGCTGAGTTCAAAAATAACGCAGCATTTACCAACGTCGGT
ATTTCTCAGCTCATGAAATGGGCTCTCATCCTTGGAAAGCATCTTAAAGTCGGAAATGAAAGGAGTTACGACAAA
TGGGAAGATATCATGACCAAAATGTATTTACCTGTCAACCATGCAGGAGATGTTACGTTGGAATATACAGGGATG
AACAACTCTATCGAAGTCAAACAAGCTGATGTGGTTCTTATTTCTTATCCTTTAGATGATGAGGATGGTGCCTTA
CAGGAATATTTCGATTACGATGAGGATAGGGCTATCAGCGACGTTAGATATTACTCTGATAAACAAACAGATGAA
GGCCCGGCTATGACTTTCTCTGTTTACTCTGCGGTCAATGCGAAGTTTAACAAAGAAGGATGCTCTTCTCAAACC
TATCTCTTGAAATCTGTAGAACCTTATTTCAGGTTTCCATTTGGACAGATGAGTGAACAGTCAACGGATCAGTAT
GATACCAATGGAGGAACTCACCCTGCCTTCCCATTTTTAACCGGTCATGGAGCTTTCTTGCAAAGTTCTATTTAT
GGTTTGACCGGATTAAGATTTAGCTATATTTACAACGATACAGACAAATCTATCAAGAGAAGGCTGGCCTTCGAT

CCCCTTCAACTACCGTGTTTACCGGGCGGATTCTCTATAAATGGATTTGTCTATATGAACCAGACGTTGGACATC
ACTGTGAACGATACGTATGCTACAATTGCACATCGGGGAAATGCTACCACCATCAATGTTTATGTGGATTCTAGA
AATGAAATGGGAGGTAAGGAACACAAGATCCAACCAGGAAAGAGCCTTTCCATCCCATTATACCAAACAGAGCAG
AATATTCCAGGATCATTCATTGAATGCACGGTAAAGAATGTCACTGCATTGCAACCTGGCGTGGTAGGAGACCCT
ATCCAGGCAGTTGCTGACGGAGATAACTCCACTATCTGGAAAATTGAATCTCGTGAAGAACCGACTCATCTGATA
TTTGATCTTGGAGATGAGCTTGACATTGAAGGAGGCTTGGTTGTATGGGGAACATATCCTGCAGAGAGCTTTTCA
GTCTCGGTTTTGAGAGATTTTAATTCAACGAACTACAGAGTCATCAACAACGTTGAAAACTACGATTTGATCTAC
GAATCTGGGAATGTCACAGCATCCAGCCCTTTTGACGAATCTCACATCAAGAAAGTCCAAATCTTGCCTCACAAT
TGTACGAATTTTACGTTTTCCGAACTGACAGCCTCTAGGTATGTCTTATTTGAGTTCACAGATGTTCTAGGTTAC
CCCCAAGATTATTCATATGGAGCCCAAGTAGCCGAAGTTGTTCTCTACTAA

AA

MLNRVLLVALSCVVFFHLVTTFPVGTSSDSLQIRNLLSHNFTRANISEGLSSGATYFVDEDTETYYDKELKVLRT
TRFPRYNNYQLQPYVANGYIGSRIPRVGSGFTYDTSDNKTSENLKNGWPLFNKRYSGAFIAGFFNSQPTVPETNF
EELEKDGYESIIASIPQWTSLELTVNVNGTNQTLKADDVDITHISDYSQQLSLLDGIVTTNYTWLGLVNVSISVL
AHRDIVSLGFVSLELSSQKNITVSVTDILDFATSTRCSYLDSGVNEQSIFMKVQPSNVPTNATIYSSLMSSNSTS
SLLKQNQTVSQTLRVNLSKNQAASFQKYVGVVSDDYLDSIETNLTSYQFARETAKFAEIKGRSWILKSHKEAWNE
LLNGKSIVFHDNDFLTLASDSSIYHLMANTRSEANGGTSALGVSGLSSDSYGGMVFWDTDFWMLPSVQAFSPRHA
VSLSKFRDHTHDQAKKNAQTRDMNGAVYPWTSGRFGNCTSTGPCYDYEYHINIDIAFMFWKLYLGGAIDDDYMKE
FGYPIIEDVASFFVDYVDYNSTLDKYTTRNLTDPDEYAEFKNNAAFTNVGISQLMKWALILGKHLKVGNERSYDK
WEDIMTKMYLPVNHAGDVTLEYTGMNNSIEVKQADVVLISYPLDDEDGALQEYFDYDEDRAISDVRYYSDKQTDE
GPAMTFSVYSAVNAKFNKEGCSSQTYLLKSVEPYFRFPFGQMSEQSTDQYDTNGGTHPAFPFLTGHGAFLQSSIY
GLTGLRFSYIYNDTDKSIKRRLAFDPLQLPCLPGGFSINGFVYMNQTLDITVNDTYATIAHRGNATTINVYVDSR
NEMGGKEHKIQPGKSLSIPLYQTEQNIPGSFIECTVKNVTALQPGVVGDPIQAVADGDNSTIWKIESREEPTHLI
FDLGDELDIEGGLVVWGTYPAESFSVSVLRDFNSTNYRVINNVENYDLIYESGNVTASSPFDESHIKKVQILPHN
CTNFTFSELTASRYVLFEFTDVLGYPQDYSYGAQVAEVVLY

Peptide

MLNRVLLVALSCVVFFHLVTTF

SEQ0028

Homolog of YOR190W; **SPR1**

Sporulation-specific exo-1,3-beta-glucanase; contributes to ascospore thermoresistance

S. cerevisiae null mutant: viable

chr4_0692

ORF

ATGATGATTAGTGCGTTTGTTTGGTCAAGTCTAATTGTTGGTCTAATTTCTGGCTTGTTGGCCCTTGAACATCCT
CAGTTCAAGGAATTTTCTAACCGCACAATTATTCAACATAATGATTTACTTGAAAGAAACCGGTCAACAAGTAGA
TTTGATAAACCACCTAAATTATACGGAGTTGCATTGGGTGGCTGGCTTGTTCTTGAACCATATATCACCCCCAGT
CTTTTCAATGATACCGTGGAAGAAACTGTGGACGAGTACACTCTATGTCATAAACTAGGCAAGCAAAAAGCTACT
GAAGTGCTAAAAAAGCATTGGTCGACTTTTATTACGGAAAGTGACATTATCAAGATTAAGAACGTTGGTTTGAAC
AGTGTAAGAATTCCCATTGGCTACTGGGCGTATGATCTTCTTGAAGATGATCCTTATATTCAGGGTCAAGATGAA
TTTTTATCCCAATGTATAGACTGGTGTGCCAAACATGGTCTGAGTGTTTGGATAGATTTGCATGGTGCTCCTAGT
TCACAAAACGGATTTGACAATTCTGGGCGTAGAGGAAGAGCTGGATGGCAGGATGAGCAGAGGTATATTGACAAG
ACATTATATGTGCTAGAAACAATTGCCAAACGGCATGGAAATAAATCCAATGTTATTGGAATCGAAATTCTAAAC
GAGCCTTTTGGCCCAGTTTTGAACATTGAAAAGTTAAAACAGTTCTATCAAAAGGGGATAACGGTGATTAGGAAT
ACAGGCTACTCCAAAGATATAGTGATATCTGATGCATTCCAAGGAATATTTTATTGGAACGATTTTCAGCCCTCA
GATTCAAATCTTATTCTAGATAGACACCATTACGAAGTTTTCTCCGATGGACAGCTAAGATCCAGTTTCGAAGGT
CATCTGAGGGGTATTGAGGCTTTTGGCAGGGCCATTGCCATAGAGAAGCCTACAGTAGTAGTTGGAGAATGGTCT
GCAGCTATTACAGATTGTGCTCCCTGGGTCAATGGTGCCGGAAGACCTTCAAGATATCATGGAATGGTTCTTGAA
GATGGAACGGTTGGAGATTGTTCTGGTGCTACCAACATTGAACAGTGGGCAGGGAAAAGACGTGAAGAATTTTTA
AAGATAATCGAGACTTCATTAAAGGCCTACAATGCAGCTGACGGGTGGTTTTTCTGGTGTTGGAAAACCGAATCC
GCTTTGGAATGGGATATGGAGAAGCTCTTACATTATACTCTTTTCCCTCGGGTTGGGTTTAA

AA

MMISAFVWSSLIVGLISGLLALEHPQFKEFSNRTIIQHNDLLERNRSTSRFDKPPKLYGVALGGWLVLEPYITPS
LFNDTVEETVDEYTLCHKLGKQKATEVLKKHWSTFITESDIIKIKNVGLNSVRIPIGYWAYDLLEDDPYIQGQDE
FLSQCIDWCAKHGLSVWIDLHGAPSSQNGFDNSGRRGRAGWQDEQRYIDKTLYVLETIAKRHGNKSNVIGIEILN
EPFGPVLNIEKLKQFYQKGITVIRNTGYSKDIVISDAFQGIFYWNDFQPSDSNLILDRHHYEVFSDGQLRSSFEG
HLRGIEAFGRAIAIEKPTVVVGEWSAAITDCAPWVNGAGRPSRYHGMVLEDGTVGDCSGATNIEQWAGKRREEFL
KIIETSLKAYNAADGWFFWCWKTESALEWDMEKLLHYTLFPSGWV

Peptide
MMISAFVWSSLIVGLISGLLAL

SEQ0029
Homolog of YML130C; **ERO1**
Thiol oxidase required for oxidative protein folding in the endoplasmic reticulum
S. cerevisiae null mutant: inviable
S. cerevisiae overexpression mutant: decreased vegetative growth
S. cerevisiae misexpression mutant: decreased resistance to 1,4-dithiothreitol
chr1-1_0011

ORF
ATGAGGATAGTAAGGAGCGTAGCTATCGCAATAGCCTGTCATTGTATAACAGCGTTAGCAAACCCTCAAATCCCT
TTTGACGGCAACTACACCGAGATCATCGTGCCAGATACCGAAGTTAACATCGGACAGATTGTAGATATTAACCAC
GAAATAAAACCCAAACTGGTGGAACTGGTCAACACAGACTTCTTCAAATATTACAAATTAAACCTATGGAAACCA
TGTCCGTTTTGGAATGGTGATGAGGGATTCTGCAAGTATAAGGATTGCTCTGTCGACTTTATCACTGATTGGTCT
CAGGTGCCTGATATCTGGCAACCAGACCAATTGGGTAAGCTTGGAGATAACACGGTACATAAGGATAAGGGCCAA
GATGAAAATGAGCTGTCCTCAAATGATTATTGCGCTTTGGATAAAGACGACGATGAAGATTTAGTATATGTCAAT
TTGATTGATAACCCTGAAAGATTCACCGGTTATGGTGGTCAGCAATCTGAATCTATTTGGACTGCGGTCTATGAT
GAGAACTGTTTCCAGCCGAATGAAGGATCACAATTGGGTCAAGTTGAAGACCTCTGTTTGGAGAAACAGATCTTT
TACCGATTGGTTTCTGGTTTGCATTCTAGTATCTCCACCCACCTCACAAACGAATATCTGAATTTGAAAAATGGA
GCATACGAACCAAATTTGAAACAGTTCATGATCAAAGTTGGGTATTTTACTGAAAGAATTCAAAACTTACATCTC
AATTATGTCCTTGTATTGAAGTCACTAATAAAGCTACAAGAATACAATGTTATCGACAATCTACCTCTCGATGAC
TCTTTGAAAGCTGGTCTTAGCGGTTTAATATCTCAAGGAGCACAGGGTATTAACCAGAGCTCTGATGATTATCTA
TTTAACGAGAAGGTTCTTTTCCAAAATGACCAAAATGATGATTTGAAAAATGAATTCCGTGACAAATTCCGCAAC
GTGACTAGATTAATGGATTGTGTCCATTGCGAGAGATGCAAATTATGGGGAAAATTGCAAACTACAGGGTACGGG
ACTGCATTGAAGATTCTATTTGATTTGAAGAATCCTAATGACTCCATCAATTTAAAGAGAGTTGAGTTAGTTGCT
CTAGTCAACACATTCCATAGATTGTCCAAATCTGTTGAAAGCATTGAAAACTTTGAAAAACTATATAAGATTCAA
CCGCCAACGCAGGATCGTGCATCAGCGTCGTCCGAATCCTTAGGCCTTTTCGATAACGAAGATGAACAAAATCTC
CTCAACTCGTTTTCGGTTGATCAGGCAGTCATTTCATCGAAAGAGGCACCAGAAGAAATCAAAAGCAAACCTGTT
GGAAAAGCCGCATATAAACAAAACAGTTGTCCATCATTGGGTTCAAAATCTATCAAAGAAGCATTCCATGAAGAA
CTTCACGCATTTATTGATGCAATTGGATTTATATTGAACTCTTACAGGACTTTGCCCAAGCTGTTGTACACACTT
TTCCTCGTTAAATCATCTGAATTATGGGACATTTTCATTGGCACTCAAAGGCACCGAGATACCACATATAGAGTA
GACTTGTAA
AA
MRIVRSVAIAIACHCITALANPQIPFDGNYTEIIVPDTEVNIGQIVDINHEIKPKLVELVNTDFFKYYKLNLWKP
CPFWNGDEGFCKYKDCSVDFITDWSQVPDIWQPDQLGKLGDNTVHKDKGQDENELSSNDYCALDKDDDEDLVYVN
LIDNPERFTGYGGQQSESIWTAVYDENCFQPNEGSQLGQVEDLCLEKQIFYRLVSGLHSSISTHLTNEYLNLKNG
AYEPNLKQFMIKVGYFTERIQNLHLNYVLVLKSLIKLQEYNVIDNLPLDDSLKAGLSGLISQGAQGINQSSDDYL
FNEKVLFQNDQNDDLKNEFRDKFRNVTRLMDCVHCERCKLWGKLQTTGYGTALKILFDLKNPNDSINLKRVELVA
LVNTFHRLSKSVESIENFEKLYKIQPPTQDRASASSESLGLFDNEDEQNLLNSFSVDQAVISSKEAPEEIKSKPV
GKAAYKQNSCPSLGSKSIKEAFHEELHAFIDAIGFILNSYRTLPKLLYTLFLVKSSELWDIFIGTQRHRDTTYRV
DL
Peptide
MRIVRSVAIAIACHCITALAN

SEQ0030
Homolog of YMR297W; **PRC1**
Vacuolar carboxypeptidase Y (proteinase C), broad-specificity C-terminal exopeptidase involved in non-specific protein degradation in the vacuole; member of the serine carboxypeptidase family
S. cerevisiae null mutant: viable; decreased phytochelatin accumulation
chr1-4_0013

ORF
ATGATATTACACACCTATATTATTCTCTCGTTATTGACTATATTTCCTAAAGCTATTGGTCTGTCCTTGCAGATG
CCAATGGCCTTGGAAGCTAGTTATGCCTCATTAGTGGAGAAAGCAACCCTCGCTGTTGGACAAGAAATTGATGCC
ATACAAAAGGGTATTCAGCAAGGTTGGTTGGAAGTAGAGACAAGATTTCCAACTATAGTGTCACAGTTATCCTAT

AGTACTGGCCCAAAATTTGCGATCAAGAAGAAAGATGCAACTTTTTGGGATTTCTATGTTGAAAGTCAAGAGTTG
CCAAACTACCGGTTACGCGTGAAAAGGAACAATCCCGAAGTTTTGAAAGTTGATTTTACTAAACAGTATAGCGGG
TATTTGGACGTAGAGGCTGATGACAAGCATTTTTTTTACTGGTTTTTTGAATCGAGGAATGACCCACAGAATGAC
CCAATTATTTTATGGTTGAATGGTGGCCCTGGCTGCTCTTCCTTGACTGGTCTCTTTTTCGAATTAGGATCGTCT
AGAATTAATGAAAATCTGAAACCAATTTTCAACCCCTATTCGTGGAATGGTAATGCTTCAATCATCTACTTAGAT
CAACCGGTCAATGTTGGGTTTTCTTATTCTTCATCATCGGTGAGTAACACTGTTGTTGCGGGAGAAGATGTGTAT
GCATTTCTTCAGCTTTTTTTTCAACACTTCCCGGAATATCAAACTAATGACTTTCATATTGCCGGTGAATCTTAT
GCAGGACATTACATTCCGGTGTTTGCAGACGAAATTTTGAGTCAAAAGAACAGAAATTTCAATCTTACTTCAGTC
TTGATCGGAAATGGATTAACTGACCCTTTGACTCAATACCGATATTACGAGCCAATGGCTTGTGGTGAAGGTGGT
GCCCCGTCAGTACTGCCTGCCGATGAGTGCGAAAATATGCTAGTTACCCAAGATAAATGTTTGTCTTTAATTCAA
GCATGCTACGACTCACAGTCGGCATTCACATGCGCACCGGCTGCCATTTATTGTAATAACGCTCAGATGGGACCC
TATCAGAGAACTGGGAAGAATGTGTATGATATTCGTAAGGAATGTGATGGTGGATCCTTGTGCTATAAGGACCTT
GAATTCATCGATACCTACTTAAATCAAAAGTTTGTTCAAGATGCTTTGGGCGCCGAGGTCGATACCTATGAATCT
TGCAATTTTGAAATCAACAGAAACTTTTTATTTGCTGGAGATTGGATGAAACCTTATCATGAACATGTCAGCAGT
CTCTTGAACAAAGGTTTGCCCGTTTTGATTTACGCAGGGGACAAAGATTTCATTTGCAACTGGTTGGGTAATCGA
GCATGGACTGATGTCTTGCCGTGGGTTGATGCTGATGGTTTTGAAAAAGCCGAAGTCCAAGATTGGTTGGTTAAT
GGAAGGAAGGCTGGTGAATTTAAGAACTATAGCAACTTCACCTACCTAAGGGTTTATGATGCTGGTCATATGGCC
CCATATGATCAGCCAGAGAATTCTCATGAAATGGTCAATAGATGGATATCCGGAGACTTTAGCTTTCACTAG

AA

MILHTYIILSLLTIFPKAIGLSLQMPMALEASYASLVEKATLAVGQEIDAIQKGIQQGWLEVETRFPTIVSQLSY
STGPKFAIKKKDATFWDFYVESQELPNYRLRVKRNNPEVLKVDFTKQYSGYLDVEADDKHFFYWFFESRNDPQND
PIILWLNGGPGCSSLTGLFFELGSSRINENLKPIFNPYSWNGNASIIYLDQPVNVGFSYSSSSVSNTVVAGEDVY
AFLQLFFQHFPEYQTNDFHIAGESYAGHYIPVFADEILSQKNRNFNLTSVLIGNGLTDPLTQYRYYEPMACGEGG
APSVLPADECENMLVTQDKCLSLIQACYDSQSAFTCAPAAIYCNNAQMGPYQRTGKNVYDIRKECDGGSLCYKDL
EFIDTYLNQKFVQDALGAEVDTYESCNFEINRNFLFAGDWMKPYHEHVSSLLNKGLPVLIYAGDKDFICNWLGNR
AWIDVLPWVDADGFEKAEVQDWLVNGRKAGEFKNYSNFTYLRVYDAGHMAPYDQPENSHEMVNRWISGDFSFH

Peptide

MILHTYIILSLLTIFPKAIGL

SEQ0031

Homolog of YKL046C; **DCW1**

Putative mannosidase, GPI-anchored membrane protein required for cell wall biosynthesis in bud formation;homologous to Dfg5p

S. cerevisiae null mutant: viable

S. cerevisiae overexpression mutant: decreased vegetative growth

chr1-4_0242

ORF

ATGTGGAGACCTCTTGTTCTTGTACTGGCTCTTCTGCGGGTGCTAAGCCCTGCCAAAGGCCTGGAGTTGGAGCTT
TCCGATTTAAGCTCGCTACAACACGCAACCTCGTTGGTTGCAGATGGTTTAATGGACTATTACGAAGGATTTCAT
TTAGGAGGAACAATAGGTATGTTCACCAATCCTTACTATTGGTGGCAGTCTGGAGCTGCATTTGGTTCGATGCTG
GACTATTGGTGGTACATGGAAAATGATACCTATCACGATGCCATTATGCAGGCTATAGTTTACCAGGCTGGAGAC
AACGCGGACTTTATGCCTTTAAATCAAACAACTACAGAGGGAAACGATGATCAAGGATTTTGGGGTATCACTGCG
ATGGCTGCCGTTGAAAGAAACTTCACCAACCCCCCCAGAAGATGAACCACAATGGTTGTATCTGGTTCAAGCTACT
GTTAACACCATGTGGGAAAGATGGGATCTTGAGCATTGTAATGGAGGCCTAAGGTGGCAGATTTTCCAGTGGAAT
GCTGGTTACAATTACAAAAACACCGTTTCAAATGCCTGTCTTTTCCAATTGAGTGCCCGTTTGGCACGATACACA
GCAAATGATACTTACATCACACTTGCTGAAGAGGCCTTTGACTGGATGTATGGGGCTGGGTTCCTAACTGAAGGT
GATTGGTGGTTCGTTTACGATGGAGCATTTGTGGATGATAATTGTACAGAAATTGTCATGTTACAGTGGACATAT
AATGCAGGGCTGATGGTGTCTGGGTGTGCATACCTTGCTAACTATACAGGCGACGAGATGTGGTTAGACCGTACC
GAAAACTTCCTACATGGAATTCAAGTTTTCACCAATCAAAGTGTTTTCTTTGAAGCAGCATGTCAAGGATCTGGA
AACTGTAACACGGATCAAAGATCTTTTAAAGCTTATCTAGCGCGTTTCCTTGGTCTTACTGCACAAATGGTTCCA
TCCACTGCAGAAACTATTATGAACTGGATGAATACATCTGCGGTGGCAGTAGCTCAATCATGTTCTGGAGGTACA
GATGGACACACTTGTGGTATGAACTGGCTGGCCGATGGTTGGGACGGCTTTTATGGATTAGGAGAACAAATGTCG
GCTCTGGAGACCTTACAAAACACTCGTGCACTTGTCAGACCTGCTCCTTACACCGCACAAACTGGTGGTTCTTCG
CAGGGAGATCCTGCTGCTGGATTAGGAGTGAAGACGGAGGCTGTGCCACCGTTGAAATTAACTAATGCGGACGTT
GCTGGTGCCGCCATTATCACAGCCATTATAGGACTATCAGTTATTGCCGGTGCCATTTGGTTACTCTTATAG

AA

MWRPLVLVLALLRVLSPAKGLELELSDLSSLQHATSLVADGLMDYYEGFHLGGTIGMFTNPYYWWQSGAAFGSML
DYWWYMENDTYHDAIMQAIVYQAGDNADFMPLNQTTTEGNDDQGFWGITAMAAVERNFTNPPEDEPQWLYLVQAT

VNTMWERWDLEHCNGGLRWQIFQWNAGYNYKNTVSNACLFQLSARLARYTANDTYITLAEEAFDWMYGAGFLTEG
DWWFVYDGAFVDDNCTEIVMLQWTYNAGLMVSGCAYLANYTGDEMWLDRTENFLHGIQVFTNQSVFFEAACQGSG
NCNTDQRSFKAYLARFLGLTAQMVPSTAETIMNWMNTSAVAVAQSCSGGTDGHTCGMNWLADGWDGFYGLGEQMS
ALETLQNTRALVRPAPYTAQTGGSSQGDPAAGLGVKTEAVPPLKLTNADVAGAAIITAIIGLSVIAGAIWLLL

Peptide

MWRPLVLVLALLRVLSPAKGL

SEQ0032

Homolog of YGR282C; **BGL2**

Endo-beta-1,3-glucanase, major protein of the cell wall, involved in cell wall maintenance

S. cerevisiae null mutant: viable

S. cerevisiae overexpression mutant: decreased vegetative growth

chr1-4_0426

ORF

ATGATCTTTAATCTTAAAACACTGGCTGCGGTTGCAATCTCCATTTCACAAGTGTCTGCAGTTTCCTCTCTGGGT
TTTGCTCTCGGAAACAAGAACGTTGACGGAACTTGTAAGTACTTGGCCGACTACGAGGCCGACTTGGATACTATT
AGAGGCGGCTCTGAAGCCGTTGCTATTAGAGCTTATTCCGCTGAAGACTGTAACACTTTACAATACCTCGGTCCT
GCTGTTGAAGAGAAGGGCTTCAAATTAGTTCTATCAGTCAGACCACTGGATGAGAGCTACTACCAGGCAGAAAAG
AATGCACTAAGTGAATACCTTCCCCAATTATCTGTTTCGACTTTGCAATTTTTGTCAGTTGGATCCGAAGCTTTG
TACAGAGACGACTTGCCAGCTTCAGATCTGGCTGATAAAATTAGAGATATGAAGGAGTTTTTGGCTGGCTTGACT
GACAAGAATGGGGACTCCTACTCTTCCGTCCCAGTCGGAACCATCGATTCATGGAACGTCCTTGTAGACGCCTCC
GCTGCACCAGCTATTGAAGCATCTGATGCCGTTTACGCCAACGCTTTCTCATACTGGCAAGGTCAAGGTCCTTCC
AACTCTACCTATTCCTTCTTTGACGATATCATGCAAGCATTGCAAGTAATCCAGACCATCAAGGGATCTACTGAT
ATCGATTTCTGGGTTGGTGAGACCGGATGGCCTACCGACGGTGACAACTTTGGTGATGCTGTTCCATCTATTGAG
AACGCTGACAACTTTTGGAAAGAAGCTATCTGCGGTATCAGAGGTTGGGGCATTAACACATTCGTTTTCCAGGCA
TTTGACGAAGACTGGAAGGAAGAGGACGACGCTGTTGAAAACCACTTCGGTGTTTGGGACAGTTCCAGACAGTTA
AAGCTGGACTCATTAGGTTGCGACTTTTCTTCTTGA

AA

MIFNLKTLAAVAISISQVSAVSSLGFALGNKNVDGTCKYLADYEADLDTIRGGSEAVAIRAYSAEDCNTLQYLGP
AVEEKGFKLVLSVRPLDESYYQAEKNALSEYLPQLSVSTLQFLSVGSEALYRDDLPASDLADKIRDMKEFLAGLT
DKNGDSYSSVPVGTIDSWNVLVDASAAPAIEASDAVYANAFSYWQGQGPSNSTYSFFDDIMQALQVIQTIKGSTD
IDFWVGETGWPTDGDNFGDAVPSIENADNFWKEAICGIRGWGINTFVFQAFDEDWKEEDDAVENHFGVWDSSRQL
KLDSLGCDFSS

Peptide

MIFNLKTLAAVAISISQVSAV

SEQ0033

Homolog of YGL028C; **SCW11**

Cell wall protein with similarity to glucanases; may play a role in conjugation during mating based on its regulation by Ste12p

chr2-1_0052

ORF

ATGCTATCAACTATCTTAAATATCTTTATCCTGTTGCTCTTCATACAGGCATCCCTACAGGCTCCAATACCTGTA
GTTACCAAATATGTTACCGAAGGTATTGCCGTTGTTACGGAGACCAACGTCCGCGTTGTAACTAAAACAATTCCA
ATCGTTCAAGTCCTAATCTCCGATGGGGCAACTTACACTCACACACTCACTACCGTTAGCACCGCAGAAGAGAAC
GGAAATTTTCAACCGATCACAACGACATCCATCGTGAATAAAGAGGTTGTTGTGCCAACTTCTGTTACCCCGAAC
ACCCAACAAACACGACCAACGCAGGTGGATACAACTCAAAACAACGCTGACACTCCTGCCGCTCCTACTCCGTCC
CCTACTACGTCCAGCAATAACGGCGTATTCACTACCTACAGTACTACTCGTAGTGTAGTGACGTCTGTAGTGGTA
GTTGGTCCTGATGGGTCCCCTATTGAAAACACTGGACAAACTGCAAATCCAACAACAACCGCCCCTACTACTTCG
ACTACGGCCGCTCGTACTACAAGTTCAACATCAACCACTCCTACCGCATCTAGTACCCCTGGTGGTAACCATCCA
CGAAGCATTGTTTACAGTCCATATTCCGACTCAAGTCAATGCAAAGATGCCACGACTATTGAAACTGATTTGGAA
TTTATTGCCTCTAAAGGAATTTCTGCTGTAAGAATTTACGGTAATGATTGCAACTATCTCACGGTTGTGCTTCCC
AAGTGTGCTTCTTTGGGTCTAAAGGTTAACCAAGGTTTCTGGATTGGTCCATCTGGTGTTGACTCTATAGATGAT
GCTGTCCAGGAATTTATTCAAGCTGTGAATGGAAACAATGGGTTTAATTGGGATCTCTTTGAGCTCATTACTGTG
GGCAATGAAGCTATCAGCGCAGGTTACGTTTCAGCTTCTAGTCTAATTTCAAAGATCAAAGAAGTCTCGTCCATA
TTATCAAGTGCAGGATACACTGGTCCAATTACTACTGCTGAGCCACCAAATGTTTATGAAGATTACGGAGATCTA

TGTTCCACGGATGTGATGTCTATCGTTGGAGTTAATGCTCATTCCTACTTCAACACCTTGTTTGCTGCTTCTGAT
TCCGGTTCTTTTGTCAAAAGCCAGATTGAGGTCGTCCAAAAAGCATGTTCCCGATCAGACATCACCATTATAGAA
ACAGGATATCCAAGCCAAGGAGCAACTAACGGAAAGAATGTTCCTAGCAAAGAAAATCAAAAAACTGCAATATTT
TCCATTTTTGAGGTAGTGGGAACAGATGTCACAATCCTGAGTACCTACGATGACCTATGGAAAGACCCAGGCCCG
TATGGCATTGAGCAGTTTTTCGGAGCTATCGATCTGTTTAGTTAA

AA

MLSTILNIFILLLFIQASLQAPIPVVTKYVTEGIAVVTETNVRVVTKTIPIVQVLISDGATYTHTLTTVSTAEEN
GNFQPITTTSIVNKEVVVPTSVTPNTQQTRPTQVDTTQNNADTPAAPTPSPTTSSNNGVFTTYSTTRSVVTSVVV
VGPDGSPIENTGQTANPTTTAPTTSTTAARTTSSTSTTPTASSTPGGNHPRSIVYSPYSDSSQCKDATTIETDLE
FIASKGISAVRIYGNDCNYLTVVLPKCASLGLKVNQGFWIGPSGVDSIDDAVQEFIQAVNGNNGFNWDLFELITV
GNEAISAGYVSASSLISKIKEVSSILSSAGYTGPITTAEPPNVYEDYGDLCSTDVMSIVGVNAHSYFNTLFAASD
SGSFVKSQIEVVQKACSRSDITIIETGYPSQGATNGKNVPSKENQKTAIFSIFEVVGTDVTILSTYDDLWKDPGP
YGIEQFFGAIDLFS

Peptide

MLSTILNIFILLLFIQASLQA

SEQ0034

Homolog of YOL154W; **ZPS1**

Putative GPI-anchored protein; transcription is induced under low-zinc conditions, as mediated by the Zap1p transcription factor, and at alkaline pH

chr3_0517

ORF

ATGCGATTCATTAACTTGACTATTACAAGTTTACTTGCACTGGCTTCCAGAACAACAGCTGAAGATGTCACCGTC
ACACAATTGGTTTCTAGCCCAACCTCTGCGTCTGAAGCAGGGGACTGGTCTGCTAATTGGGTTAAGGGGTTTCCA
ATACATTCTTCCTGTAACCAGACTAAGTTCAACCAACTGTCCCTGGGTCTAGAAGAGGCTCAGATCATGGCCGCT
CATGCCAGAGATCACACCTTGAGATATGGTAACGAGTCTGAGTTCTTCACTCAATACTTCGGTAATGCTAGTACC
GCAGAAGTCATCGGATGGTTCTCAGTGGTCGTTGATGCCGACAAATCCAACCTGCTATTCAGATGCGATGACATC
GACGGAAACTGTAAATTTGATGGCTGGGCCGGTCACTGGAGAGGTGAAAATGGATCCGACGAAACTGTCATTTGT
GATCTTTCCTTCCAGACCAGGAAGTTCCTTGCTCAAATGTGCTCCAACGGTTATAACGTTGCCAACTATCCAAAC
AACCTCTACTGGGCTTCTGACTTGCTCCACAGACTCTATCACACTACCACCATTGGTCAATTGACAGTCGATCAC
TACGCTGATACGTATGAAGAATGCTTAGAATTGGCTATTGAGAGTCCAGATGAAGCAGTTAGAAACACCGCTTCC
CTAAGATTTTATGCTCTAGATGTGTATGCATACGACATCGCCGTCCCAGGCGAGGGATGCAGTGGAGAAATTGGT
GAGGATGATGAAGAGTCTACATCTTCTGGTAGTGCTACCGTCACTACAACGTCTACTTCTGCTGGAACAGAGTGC
CACACACACTCAGATGGCGAAGTCCACTGTGCTTAA

AA

MRFINLTITSLLALASRTTAEDVTVTQLVSSPTSASEAGDWSANWVKGFPIHSSCNQTKFNQLSLGLEEAQIMAA
HARDHTLRYGNESEFFTQYFGNASTAEVIGWFSVVVDADKSNLLFRCDDIDGNCKFDGWAGHWRGENGSDETVIC
DLSFQTRKFLAQMCSNGYNVANYPNNLYWASDLLHRLYHTTTIGQLTVDHYADTYEECLELAIESPDEAVRNTAS
LRFYALDVYAYDIAVPGEGCSGEIGEDDEESTSSGSATVTTTSTSAGTECHTHSDGEVHCA

Peptide

MRFINLTITSLLALASRTTAE

SEQ0035

Homolog of YGR189C; **CRH1**

Putative chitin transglycosidase, cell wall protein that functions in the transfer of chitin to beta(1-6)glucan; localizes to sites of polarized growth; expression is induced under cell wall stress conditions

S. cerevisiae null mutant: resistance to ketoconazole: normal; resistance to miconazole: normal; viable

S. cerevisiae overexpression mutant: resistance to ketoconazole: increased; resistance to geneticin: normal; resistance to miconazole: increased; resistance to benomyl: increased

chr4_0559

ORF

ATGGTTTCTTTAACAAGACTACTAATTACCGGAATCGCCACCGCTTTGCAGGTGAATGCCTTGCCCGTGGCCGAT
TCAAACCCAGCTGGAACACCAGTCGTAAAGAGGGAAACTGCTACTTCAGCAGAGTGTAACCCTCTGTTGACAGCC

GATTGTCCTCCCGTCAAGGCTCTTGCCACGTCAATTTCCGATGATTTCACCTCAGCATCAGACAACTATCACATT
GTTTCCTTCGAAGACCAAGTCACATATGACGAAGATGGTCTTACGTTGACTCTTGCTGAACGTTGGGACAACCCC
TCTTTGAAATCGAACTTCTACATTATGTTTGGTAAAGTTGAGGTGGTTTTGAAGGCCGCTACAGGACAGGGAATT
ATCTCCTCCTTCTATCTCCAAAGTGACGACCTAGACGAAATTGATCTAGAATGGTTTGGAGGTGATCCATATGAA
GTGCAAACCAACTTTTTTAGCAAGGGTAACGTCACCACTTATGACCGTGGTGGATATCATAACATCGCAGATCCA
ACTGCTAACTTCCACAACTACACTATCGACTGGAATTACGACAAGTTAACTTTCTACTTTGATGGAAACGTTATT
CGTGAATTGGCCAACGATACTAGTAGTGGCTACCCTCAGAGTCCAATGTACCTGAGAACTGGAATCTGGGCAGGT
GGTGATCCATCCAATCAGCCAGGAACTATCGAATGGGCTGGAGGATTGACTGACTATTCTCAGGCTCCATTCTCC
ATGCACATCAAACATTTGGTTGTTTCTGACTACTCCTCTGGTACTGAGTACAAATACACTGATCAATCAGGTGCG
TGGACTTCTATTGAGGCTGAAGGTGGAGAGGTTTTGGGCAGAGTTGACCAAGCCAACGAGGAGTTTGCTCTTCTT
GTCTCTGGTCAAGAGTTAGACGACACAAGTTCAAGTGAGAGCTCCTCGTCTTCTTCGTCGTCTTCTTCGTCATCA
TCATCATCATCATCATCTTCCTCATCAACTTCCTCCTCATCCTCTTCCACCACTGCGTCCTCCTCCTCTGAGTCT
TCGTCCACATCAGAATCGTCTTCTTCAGTAATATCTTCCACTACAGCAAGTTCCTCCACAACCACAAGTGCTAAT
TCTACAACACATGAGTCCACGACCACAAGACCAACTTCGGAATCAACAACGCAAAGCACAGCAACATCATCTCAT
GTATCAGAGACGTCTCAGTCCGCACATTCAACCACGACAGGGCCATCTAGGAACTCATCTATTGCAGCCGAATCA
GCCGCAACGTCTAGTGTTTCCTCCCACTCCCCAGAGGTTTCTGCATCCAACGGAGCCGACAGGAACATTTTACAA
ATCAGAGCTGCGTACTTGGTAGGAATACCGTTCTCAGTTCTTTCATTGTTAATGATATGA

AA

MVSLTRLLITGIATALQVNALPVADSNPAGTPVVKRETATSAECNPLLTADCPPVKALATSISDDFTSASDNYHI
VSFEDQVTYDEDGLTLTLAERWDNPSLKSNFYIMFGKVEVVLKAATGQGIISSFYLQSDDLDEIDLEWFGGDPYE
VQINFFSKGNVTTYDRGGYHNIADPTANFHNYTIDWNYDKLTFYFDGNVIRELANDTSSGYPQSPMYLRTGIWAG
GDPSNQPGTIEWAGGLTDYSQAPFSMHIKHLVVSDYSSGTEYKYTDQSGAWTSIEAEGGEVLGRVDQANEEFALL
VSGQELDDTSSSESSSSSSSSSSSSSSSSSSTSSSSSTTASSSSESSSTSESSSSVISSTTASSSTTTSAN
STTHESTTTRPTSESTTQSTATSSHVSETSQSAHSTTTGPSRNSSIAAESAATSSVSSHSPEVSASNGADRNILQ
IRAAYLVGIPFSVLSLLMI

Peptide

MVSLTRLLITGIATALQVNAL

Homolog of YMR307W; **GAS1**

Beta-1,3-glucanosyltransferase, required for cell wall assembly; localizes to the cell surface via a glycosylphosphatidylinositol (GPI) anchor

S. cerevisiae null mutant: alkaline pH resistance: decreased; budding pattern: abnormal; cell shape: abnormal; chitin deposition: increased; competitive fitness: decreased; freeze-thaw resistance: decreased; resistance to doxorubicin: decreased; resistance to Calcofluor White: decreased; resistance to calcium dichloride: decreased; resistance to caffeine: decreased; resistance to hygromycin B: decreased; resistance to mycophenolic acid: decreased; resistance to rapamycin: decreased; viable; oxidative stress resistance: decreased; pheromone sensitivity: abnormal; Bud8p-GFP distribution: abnormal; resistance to FK506: decreased; resistance to Congo Red: decreased

2 homologues

SEQ0036
chr1-3_0226

ORF

ATGTTTAAATCTCTGTGCATGTTAATAGGATCCTGCCTATTGAGTTCAGTCTTGGCAGCAGATTTCCCCACTATA
GAAGTAACTGGAAACAAATTTTTCTACTCCAACAATGGATCCCAATTCTACATCAAGGGTGTTGCGTACCAAAAG
GACACATCTGGGCTCTCTTCAGACGCCACTTTCGTTGATCCTTTAGCTGACAAGAGTACCTGTGAGAGAGATATT
CCATACTTGGAAGAACTTGGAACCAACGTCATCCGTGTTTATGCTGTCGATGCAGACGCCGATCACGATGATTGT
ATGCAAATGTTGCAAGATGCCGGTATATATGTCATTGCCGACCTTTCCCAGCCAAACAACTCTATCATTACTACT
GATCCTGAATGGACTGTTGATCTTTACGACGGTTACACTGCTGTGCTTGACAATTTGCAAAAATATGACAACATT
CTTGGATTCTTTGCCGGCAATGAAGTTATCACCAACAAGTCCAATACTGACACAGCTCCATTCGTTAAGGCTGCG
ATCCGTGATATGAAAACTTACATGGAAGACAAGGGTTATAGATCGATCCCTGTGGGATATTCTGCTAACGATGAC
GAACTTACCAGGGTTGCCTCGGCCGACTACTTTGCTTGTGGAGACTCCGATGTTAAAGCCGACTTCTATGGTATC
AACATGTACGAATGGTGTGGAAAGGCTACTTTTTCCAACTCTGGATATAAGGACAGAACCGCAGAGTTCAAAAAC
TTGTCTATTCCCGTTTTCTTCTCTGAGTATGGTTGTAACGAAGTTCAACCACGTCTGTTCACTGAAGTTCAATCC
CTTTACGGTGACGACATGACAGATGTCTGGTCTGGAGGTATTGTCTACATGTACTTTGAAGAGACTAACAACTAC
GGTTTAGTCACCATCAAGTCAGATGGAGATGTTTCTACCTTGGAAGATTTCAACAATCTCAAAACTGAATTGGCC

AGCATTTCTCCATCAATCGCTACCCAGTCGGAAGTTTCTGCTACTGCTACCGAAATTGACTGTCCTGCTACTGGA
AGCAACTGGAAGGCCTCTACTGATCTGCCACCAGTCCCAGAACAGGCTGCTTGCCAGTGCATGGCTGATGCGCTA
TCCTGCGTTGTGTCTGAAGATGTCGACACTGATGATTACTCTGACCTATTTAGCTACGTCTGTGAGAATGTCTCC
AGTTGTGACGGTGTCAGTGCGGATTCAGAATCCGGAGAGTATGGTTCTTACTCTTTCTGCTCATCCAAAGAAAAG
CTATCGTTCCTGCTCAACCTTTACTACTCTGAGAATGGGGCCAAGAGCTCGGCCTGTGACTTCAGTGGGTCTGCT
ACCTTAGTCTCGGGTACAACTGCATCCGAATGCTCCTCTATCCTTAGCGCTGCAGGAACCGCTGGAACAGGATCC
ATTACTGGTATCACTGGATCGGTTGAAGCTGCCACCCAGAGCGGATCCAATTCTGGATCTTCAAAGTCCAGCTCC
GCTTCTCAATCTAGTTCAAGCAATGCTGGAGTTGGTGGTGGTGCCTCAGGCTCCAGCTGGGCAATGACCGGGCTT
GTGTCAATTTCCGTAGCTCTTGGAATGATTATGTCATTCTAA

AA

MFKSLCMLIGSCLLSSVLAADFPTIEVTGNKFFYSNNGSQFYIKGVAYQKDTSGLSSDATFVDPLADKSTCERDI
PYLEELGTNVIRVYAVDADADHDDCMQMLQDAGIYVIADLSQPNNSIITTDPEWTVDLYDGYTAVLDNLQKYDNI
LGFFAGNEVITNKSNTDTAPFVKAAIRDMKTYMEDKGYRSIPVGYSANDDELTRVASADYFACGDSDVKADFYGI
NMYEWCGKATFSNSGYKDRTAEFKNLSIPVFFSEYGCNEVQPRLFTEVQSLYGDDMTDVWSGGIVYMYFEETNNY
GLVTIKSDGDVSTLEDFNNLKTELASISPSIATQSEVSATATEIDCPATGSNWKASTDLPPVPEQAACQCMADAL
SCVVSEDVDTDDYSDLFSYVCENVSSCDGVSADSESGEYGSYSFCSSKEKLSFLLNLYYSENGAKSSACDFSGSA
TLVSGTTASECSSILSAAGTAGTGSITGITGSVEAATQSGSNSGSSKSSSASQSSSSNAGVGGGASGSSWAMTGL
VSISVALGMIMSF

Peptide

MFKSLCMLIGSCLLSSVLAA

SEQ0037

chr1-3_0227

ORF

ATGTTGTCCATTTTAAGTGCATTAACTCTGCTGGGCCTGTCTTGTGCTTCAGATCTGACCCCACCTATAGAGGTC
ACAGGAAACAAGTTCTTCTTCTCCAATAATGGATCCCAGTTCTACATGAAGGGTATCGCCTACCAACAGGATACA
TCCAACTCCACCACAGGCGACAACTTCATTGATCCGTTGGCCAATGTCGAGGCCTGCAGGCGTGATTTCCCTTAT
TTAGAGGAGTTACACACAAACGTCATCCGTGTCTACGCCATCGACGTAGACGCTGATCACGACGAGTGTATGCAA
TTGCTTGAAGAAGCAGGTATCTACGTTATTGCTGACCTTTCTGAACCAGGTGATTCCATCATCACCACTGATCCG
GCCTGGGATGTTCACTTGTATGACCGTTACACCTCAGTTATTGACGAACTGCAACAGTACGACAACGTTCTAGGT
TTTTTCGCCGGTAACGAAGTTATCACAAACAGATCAAACACCGATGCTGCTCCATTCGTCAAAGCTGCAATCCGT
GATATGAAGGCTTACATGGAAGACCAAGGCTACAGACAAATTCCAATTGGTTATTCTGCCAATGATGACGCGCTG
ACAAGAATTTCTTCCGCTAACTACTTTGCTTGTGGTGACTCGGACGTTAGGGCCGACTTCTACGGTTACAACATC
TACTCCTGGTGTGGTAGATCTTCTTTCACAACTTCCGGATACGACGATAGAACCGAAGACTTCAGAAACCTTTCA
GTTCCTCTCTTCTTCTCCGAGTATGGTTGTAACGAAGTTCAACCACGTCTGTTTACTGAAGTTCAAGCCCTTTAC
GGTGAAAACATGACTGATGTCTGGTCCGGAGGTATTGTTTACATGTATTTCGAAGAGGAAAACAATTACGGTTTG
GTTACTATTGAACCCGATGGAGATGTTTCAACCTATTCGGACTTTAACTACCTTAGATCGGAATTGGCCACTATT
TCTCCATCGCTCGCTACTCAATCTGAAGTTTCTGCCACCGCCACCGAATTGGAATGTCCACCAACCGGAAACAAC
TGGAGAGCATCTACCGATCTTCCTCCAACTCCAAACAAGGAGGTATGTGAGTGTTTGCAAAATACACTTGCATGT
GTTGTTGCTGACGACGTTGAACCTGAAGACTACGCAGACATTTTTAACTACGTTTGTTCGCGTATTGAGTGTGGA
GACATTGGCACCGATGCAGAGATTGGTGAGTACGGTCCTTACTCTTTCTGTGATGCCAAGGAGAAACTTTCTTTC
GTTTTGAACTTATATTACGAGGACCAGAATGGTGCCTCTTCTGCTTGTGACTTTGACGGCTCTGCCTCGATCGTA
AGTACCACTGCCGTTTCCTCTTGTGTCTCCATTTTGCAGTCTGCCTCCACCGTTGGTACCGCGCCAAGCTCCAGT
AGAACTGCCACTGGTTCAGGCAGCTCAGGCACCGGATCTGGATCTGGATCATCTGCCGACGCCACTTCCACTGAA
TCCGATGATGCTGCATTTATTCACAAGCCAAAGTCTGGATCAGAATTATTAATGGCCGCATTCTTTGTAACATTT
GTGACCTCAGGATTCACTTACACTCTTTTATAG

AA

MLSILSALTLLGLSCASDLTPPIEVTGNKFFFSNNGSQFYMKGIAYQQDTSNSTTGDNFIDPLANVEACRRDFPY
LEELHTNVIRVYAIDVDADHDECMQLLEEAGIYVIADLSEPGDSIITTDPAWDVHLYDRYTSVIDELQQYDNVLG
FFAGNEVITNRSNTDAAPFVKAAIRDMKAYMEDQGYRQIPIGYSANDDALTRISSANYFACGDSDVRADFYGYNI
YSWCGRSSFTTSGYDDRTEDFRNLSVPLFFSEYGCNEVQPRLFTEVQALYGENMTDVWSGGIVYMYFEEENNYGL
VTIEPDGDVSTYSDFNYLRSELATISPSLATQSEVSATATELECPPTGNNWRASTDLPPTPNKEVCECLQNTLAC
VVADDVEPEDYADIFNYVCSRIECGDIGTDAEIGEYGPYSFCDAKEKLSFVLNLYYEDQNGASSACDFDGSASIV
STTAVSSCVSILQSASTVGTAPSSSRTATGSGSSGTGSGSGSSADATSTESDDAAFIHKPKSGSELLMAAFFVTF
VTSGFTYTLL

Peptide

MLSILSALTLLGLSCAS

SEQ0038

Homolog of YDR055W; **PST1**

Cell wall protein that contains a putative GPI-attachment site; secreted by regenerating protoplasts; up-regulated by activation of the cell integrity pathway, as mediated by Rlm1p; upregulated by cell wall damage via disruption of FKS1

S. cerevisiae null mutant: viable; increased budding index

chr2-2_0148

ORF

ATGCAGTTTGGAAAGGTTCTATTTGCTATTTCTGCCCTGGCTGTCACAGCTCTGGGAGAAACAACTTCTTCATTG
AGTGCCACATTGACTTCAACCACACGAATTTCAATTGCATCTGGATGTAGTCTTGAAGACTTCACTGCTACTGCT
CAGTCTAATCTTGATGAGTTGTCCGATTGTGAAGCTGTCAGAGGTGATATCCACATTGCCGGAAGTCTGGGATCA
GCTGCTATCGCTAATGTCAAGGCCATTTACGGTTCTTTAATTGTCAAGAATGCCACCTCTTTGGTTTCATTGACC
GCTGATTCTTTGACCACCATTACCGAACAATTGGCTCTTTCTGAGTTGACCATTTTGGACACACTCTCTTTTGCT
CAATTGACTTCCGTTGGTTCTATTTACTTCGTCACCTTGCCAGCTTTGGAAGAGACCGGTTTCCACACTGGTGTC
TCTGACACTGAGTCTGTTTACATTTCTGACACTGCTTTAACCAACTTGGACGGTATCGTTGCTACCGATGTCGAT
GTTTTCAATGTCAACAACAACTTCAATTTGGACACTGTTGACTCGCACTTGGAAACTGTTTCCTCTGCTTTGGAG
GTTTCCTTCAACTCAGACGATGTTGAAGTATCCTTTGACAATTTGTTGTGGGCTAACAACATTACCTTCCGTAAC
GTTGCTTCCGTCTCTTTGAACAACCTGACCACTGTTAACGCTTCTTTGGGTTTCATTGAGTCTGGTTTCCAAAAG
TTGAGTTTCCCAGGAATCACCAGGGTTGGTGGATCTTTCTCCATTGTTGACAACGACGATTTAGAAGAAATAGAC
TTTTCTAACGTTCAGTCGATTGGTGGTGGTTTAATCATCGCTAACAACTCTAAGTTGACTGACTTCAGTGAATGG
GAGGACTTGCAAACTGTTGGTGGTGCTCTTGTTTTGGAGGGTTCTTTCGACAACGGTTCCTTCCCATCTTTGAGA
GCTGTTAGAGGTGCCTTCTCTTTGGAGTCTGATGGAGATATCTCTTGTGATGACTTCTCTGATATCAGGGGTGAC
ACTGCCGGTGAGTATCAGTGCTCTGCTGCTTCGTTCTCCACTTCTGCTAGTGCTCAGTCTTCTTCTAGTTCAACC
AGTACTGGCGGATCCTCTACCCACACCGGTAGTTCCACTGCTACCAGCTCTTCTAGTGAGGATGCTGGTGTAGCT
TTGGCTCCTGCTTCTCTCTTCACTTTGTTAGCCTCAATTGTTCTCGGATTCTTGTAG

AA

MQFGKVLFAISALAVTALGETTSSLSATLTSTTRISIASGCSLEDFTATAQSNLDELSDCEAVRGDIHIAGSLGS
AAIANVKAIYGSLIVKNATSLVSLTADSLTTITEQLALSELTILDTLSFAQLTSVGSIYFVTLPALEETGFHTGV
SDTESVYISDTALTNLDGIVATDVDVFNVNNNFNLDTVDSHLETVSSALEVSFNSDDVEVSFDNLLWANNITFRN
VASVSLNNLTTVNASLGFIESGFQKLSFPGITRVGGSFSIVDNDDLEEIDFSNVQSIGGGLIIANNSKLTDFSEW
EDLQTVGGALVLEGSFDNGSFPSLRAVRGAFSLESDGDISCDDFSDIRGDTAGEYQCSAASFSTSASAQSSSSST
STGGSSTHTGSSTATSSSSEDAGVALAPASLFTLLASIVLGFL

Peptide

MQFGKVLFAISALAVTALGE

SEQ0039

Homolog of YDR221W; **GTB1**

Glucosidase II beta subunit, forms a complex with alpha subunit Rot2p, involved in removal of two glucose residues from N-linked glycans during glycoprotein biogenesis in the ER

S. cerevisiae null mutant: viable

chr3_0179

ORF

ATGTTGAGACTTCTGACCATTGGAAGTATAGCAGTGTCCCTTTTCCCTGCTTCTGCTGAGATTCCACCTTTAAGA
GGGGTAGCTCCTGACCTTTTGGAGAAGTACGTTCCTGATAAGGATGGAAATTGGAAGTGTCTCGGTCATCCTGAG
ATCGTTCTGCATTTCGATCAAGTAAATGACGACTACTGTGATTGTCCAGATGGATCAGATGAGCCAGGAACTGCA
GCGTGTGAAAATGGAAAGTTTTACTGTGCAAATGAAGGCTTTGAACCGAATTTCATACCTACCTTCTTGGTGGAT
GATGGAGTGTGCGATTACAAGGTGTGCTGTGATGGTTCAGATGAAAAATCTGGTAAATGCCCTAACAGGTGCCTT
GAACTAGCTGAGAAGGCGGAATTGCTGCGGTTGGAAAGAAAGGCACGATTGGAAAACGGATTGAAAGCAAAGAGA
GATTTAATTGTCCAATCCCAGAATAAATTAAGAGAAATCTCCCAAAGACGCGCCGAGCTGGAGAAAACAATCCAA
TTGGAACAACAGCAACTAAACTATCTTAATGAGCAAGATCCAAATGATTCTGCCTTGCGAAAGGAAGTGGATCCC
CAATTAGAAAAAGTCGTACAGATTCTGAAAGTTCAAAAAGATGAAATTTTACTTCTCAAAACCAAAGCAGATTCG
TTAGAGAGATTACTCAAGGAGCTAGAGAAAAACTACAACCCAAACTTTAATGACGCCGCTGTGAAGGCAGCAGTT
CAAGGGTTCAGAGATTATTCAAGTAATTCTAACGATGACTCTTATTCTGACTTGGATTCTGTTGAACCGGAAAAT
ATCGTCAAAGATATCCAAAAGACCATTAACACATTTGCTTTCGACACTTCAAACGATTCTAAATACGAATCGATT
TGGAGCTTCTGGAAAGAGGTATTAGCAGACAAAATCGTTAAGATTTATCAAGAGTTTACCGCGGTGCCGCTGAAA

ACAGCTTCAAAGCCTTTGAAACCCATATCTGGAACAGCTCAAGAGATACAGAAAAGCATTGATTCGCATAATAAG
GAAATAGTGAAAATTGATTCTGATTTAGAGAATTCAGAAAGCCGATACGGACTTGAAGACATTTTCAGAGCATAT
GACGGGAGATGTTTCGTTGAAAAAATTGGAGATTATGATTATGAATACTGCTTTACGGGTTCCTTGACCCAAATT
TCTTCTAATGGCCAAAGGGTTTCGATTGGGACTAGAGATAAGATAGAGGTGTTGGAAGACGAACAAAGTGTTGGT
GGTTATTCCTATAGAGTTTACTATGAAAAGGGAGCTAAATGTTGGAACGGCCCTGTTAGAAAGGCAATAGCCGTT
GTGCAATGTGGTGATGTTGAACAATTGGTTTCTGTGTCGGAACCAGAGAAGTGTGAATACCATCTCGTGGTGCGG
AGCCCAGTTGGATGTTTCAGTGAAGACAATGTTTATCTGGTAAAGGAGATTGCTGAAGGAGAGTCACATGCCCCT
TTCAAGGTGATACCTTTCAAAGATGAACTCTAA

AA

MLRLLTIGSIAVSLFPASAEIPPLRGVAPDLLEKYVPDKDGNWKCLGHPEIVLHFDQVNDDYCDCPDGSDEPGTA
ACENGKFYCANEGFEPNFIPTFLVDDGVCDYKVCCDGSDEKSGKCPNRCLELAEKAELLRLERKARLENGLKAKR
DLIVQSQNKLREISQRRAELEKTIQLEQQQLNYLNEQDPNDSALRKEVDPQLEKVVQILKVQKDEILLLKTKADS
LERLLKELEKNYNPNFNDAAVKAAVQGFRDYSSNSNDDSYSDLDSVEPENIVKDIQKTINTFAFDTSNDSKYESI
WSFWKEVLADKIVKIYQEFTAVPLKTASKPLKPISGTAQEIQKSIDSHNKEIVKIDSDLENSESRYGLEDIFRAY
DGRCFVEKIGDYDYEYCFTGSLTQISSNGQRVSIGTRDKIEVLEDEQSVGGYSYRVYYEKGAKCWNGPVRKAIAV
VQCGDVEQLVSVSEPEKCEYHLVVRSPVGCFSEDNVYLVKEIAEGESHAPFKVIPFKDEL

Peptide

MLRLLTIGSIAVSLFPASAE

SEQ0040

Homolog of YLR343W; **GAS2**

1,3-beta-glucanosyltransferase, involved with Gas4p in spore wall assembly; has similarity to Gas1p

S. cerevisiae null mutant: viable

chr3_0306

ORF

ATGCTGGGTTTCAAGGACTTCTTTCGCCTATTAATTCTCTTTGTTACTGCATTTGCCTCCGACTCCTATGAGTTT
TCGTCTTCCGTCAAACATATAGCTTCACCCACAACAGTTGAAAGAATCCGAAAGTTGACTCCCACAATTGAAGCG
TACGGATCTAAGTTTTTTTATTCAAACAACGGATCTCAATTCTTCATCAAAGGAATAGCCTATCAGCAAGATTTC
ACTGAATACTTGCGTCGTCCTGAATCCCTTGTCGACGTTGACGCTCCCAAAACCCAAACTCTATGTGAATTGAAC
TTTAGGAACGGCCGACATTTTGTTGACCCTCTAGCTTGGCCTGATATCTGTACCAGAGATCTGCCATATCTTCAA
GAGCTAGGAGTCAACACTCTCAGGGTCTACTCTATAGATCCAAGGGCTAATCATGACGTCTGTATGGAAAGATTA
GCTGAGGCTGGGATTTATGTTATATTGGACCTGTCTAACGTTGATTGCTCTATTAATAGGGACGATCCCTCATGG
GATCTTGAAATTATGGACTGCTACACTTCTGTGATCGATACGATGGCTCCTTATGACAATGTATTAGGTTTTTTT
GCTGGCAATGAGGTCAACACGGATGTTGCAAACACTTACACAATGCCGTTTGTAAAAAGTTCTATTAGAGATATA
AAGAGTTACATTAAACGAAAGAATATTAGGAAAATTCCAGTTGGCTACTCCTCCAATGATGATTCCCTTACTAGA
AATAATTTGGCGGACTATCTAGTGTGTGGGGAAAATGATTTGGCCAAGATTGACTTTTTCGGTATCAATATGTAC
GAATGGTGTGGTTACTCATCTTATGCCACATCAGGATATAAGGAGAGAACTTTAGAGATGGCTGGCTATCCTGTT
CCCGTATTTTTAAGTGAATTTGGTTGTAATCTGAAACGTCCAAGGCCGTTCACAGAAGTGGAAGCTATATTCAAC
AAACCCATGACAAATGTTTGGTCAGGTGCAATAATGTACATGTATTTTGAAGAAAGAAATGAGTACGGAGTGGTC
AAAGTCAAGGGCGATTACATCACCGACCCCGTAGAAAAACTGCCGGATTTTAATCTTTTGAAGGCACAGTACAGG
AATGCATCTCCTAAGGGTGTGAACAAACAAGATTATTCATCTCATGGATCTAAAGACGTTCAATGTCCTGTGGCT
AGTGAAGATTGGAAAGTATCCGAAGAGCTACCTTCTACACCTAGAGCACCAAAATGTGATTGCATGATGGCAGCT
CTGAAATGCTCAATTGAAGAGGTTCCGAATAGATCAGAGCTGGCAGAAATAGTGGGGTACATTTGTGGAGAAATC
AGTTGTCACAATATTGTTGCTAATGGAACCACTGGCTGGTACGGAGAATTTTCCGATTGTTCTTTCCACCAGAAG
CTATCCTATGTGGCCCAATTATATTATAATTACCATAACTCCACTCCTGATGCATGCTATTTTGAAGGGTATGGT
CGTATCAACGCTTATCCAAACACATTAGAGGCATTAGCTATGTTTAAGACATTGAAAGGGACTAGATGCAAGGAA
GAATTAGGACCAGTCAAGGAGATTAGTCGTTTGATAACTGGGGAACCGTTTGGTTCGCAGAATAGAGATTCCCCT
CGGAGAATTCTGGTAGATAGTAGGTTGAAAAATACTAACTTGGACACGTCCGAAAGCAAAGCCATCACTGCAAAG
ATTATTTCCAAATCCAGCCTTTTGACCAATCTTATTGCTATGTTCTTAGTAATATTATTTACTTAG

AA

MLGFKDFFRLLILFVTAFASDSYEFSSSVKHIASPTTVERIRKLTPTIEAYGSKFFYSNNGSQFFIKGIAYQQDF
TEYLRRPESLVDVDAPKTQTLCELNFRNGRHFVDPLAWPDICTRDLPYLQELGVNTLRVYSIDPRANHDVCMERL
AEAGIYVILDLSNVDCSINRDDPSWDLEIMDCYTSVIDTMAPYDNVLGFFAGNEVNTDVANTYTMPFVKSSIRDI
KSYIKRKNIRKIPVGYSSNDDSLIRNNLADYLVCGENDLAKIDFFGINMYEWCGYSSYATSGYKERTLEMAGYPV
PVFLSEFGCNLKRPRPFTEVEAIFNKPMTNVWSGAIMYMYFEERNEYGVVKVKGDYITDPVEKLPDFNLLKAQYR
NASPKGVNKQDYSSHGSKDVQCPVASEDWKVSEELPSTPRAPKCDCMMAALKCSIEEVPNRSELAEIVGYICGEI

SCHNIVANGTTGWYGEFSDCSFHQKLSYVAQLYYNYHNSTPDACYFEGYGRINAYPNTLEALAMFKTLKGTRCKE
ELGPVKEISRLITGEPFGSQNRDSPRRILVDSRLKNTNLDTSESKAITAKIISKSSLLTNLIAMFLVILFT

Peptide

MLGFKDFFRLLILFVTAFAS

SEQ0041

Homolog of **YBR139W**

Putative serine type carboxypeptidase with a role in phytochelatin synthesis; green fluorescent protein (GFP)-fusion protein localizes to the vacuole; expression induced by nitrogen limitation in a GLN3, GAT1-independent manner

S. cerevisiae null mutant: viable

chr3_0633

ORF

ATGAAATCGGTTATTTGGAGCCTTCTATCTTTGCTAGCATTGTCGCAGGCATTGACTATTCCATTGCTGGAAGAG
CTTCAACAGCAAACATTTTTTAGCAAGAAAACCGTTCCTCAACAAGTTGCTGAATTGGTGGGCACCCATTACTCT
AAGGATGAGATAATCAGTCTATGGAAGGACATTGAGCTGGATGTACCCAGGGAAAAGATCCAAGAGGCCTTCGAT
AAGTTCGTAAAACAATCAACTGCCACTTCCCCCGTTAGAAATGAATTTCCCTTGTCTCAGCAAGATTGGGTGACA
GTGACCAACACCAAGTTTGATAATTATCAATTGAGGGTTAAAAAATCCCACCCTGAAAAGCTAAACATTGATAAG
GTAAAGCAATCTTCGGGATACCTGGATATCATTGATCAAGATAAGCATCTTTTCTATTGGTTTTTTGAATCCCGA
AATGATCCGTCCACAGACCCAATCATCCTATGGTTGAATGGTGGACCCGGCTGCTCTTCTATTACAGGGTTGCTA
TTCGAAAAGATTGGCCCCAGTTACATCACCAAAGAGATTAAGCCGGAACATAATCCTTATTCATGGAACAACAAT
GCTAGTGTTATCTTCCTTGAGCAACCGGTTGGAGTAGGATTTTCTTACTCTTCTAAGAAAGTCGGTGATACTGCA
ACTGCTGCCAAAGATACATATGTGTTTTTGGAGCTTTTCTTCCAAAAGTTTCCTCAGTTCCTGACCTCTAATCTG
CACATTGCTGGGGAATCGTATGCTGGCCATTATTTGCCCAAGATTGCTTCTGAGATTGTGTCTCACGCAGACAAG
ACGTTTGACCTTTCAGGAGTCATGATCGGTAATGGTCTTACTGATCCTCTAATTCAGTATAAGTACTATCAGCCA
ATGGCCTGTGGAAAAGGTGGCTACAAGCAGGTCATTTCGGACGAGGAATGTGATGAATTGGATAGGGTCTATCCA
AGATGTGAACGTTTAACGCGGGCATGTTATGAGTTCCAAAATTCAGTTACTTGTGTTCCGGCAACACTTTATTGC
GACCAAAAGCTACTGAAGCCGTACACTGACACTGGCTTGAATGTCTATGATATTCGTACAATGTGCGATGAAGGG
ACTGATTTGTGTTACAAAGAACTGGAATACGTGGAGAAGTACATGAACCAGCCTGAAGTGCAGGAAGCCGTGGGC
TCTGAAGTCAGTTCTTACAAAGGTTGTGACGATGATGTCTTCTTAAGATTTTTGTACTCTGGCGATGGATCTAAG
CCTTTCCACCAGTATATCACGGATGTTCTCAATGCAAGTATTCCGGTTCTGATTTACGCAGGTGATAAAGATTAT
ATCTGTAATTGGCTAGGAAACCAAGCTTGGGTCAATGAGCTAGAATGGAACTTGTCTGAGGAATTCCAGGCAACT
CCGATTCGACCGTGGTTCACTTTGGACAATAACGATTATGCAGGAAACGTACAAACTTATGGAAACTTTTCCTTT
CTAAGAGTATTTGATGCTGGTCACATGGTTCCTTACAATCAACCAGTCAACGCACTTGACATGGTTGTCAGATGG
ACACACGGTGATTTCTCATTTGGTTATTAA

AA

MKSVIWSLLSLLALSQALTIPLLEELQQQTFFSKKTVPQQVAELVGTHYSKDEIISLWKDIELDVPREKIQEAFD
KFVKQSTATSPVRNEFPLSQQDWVTVTNTKFDNYQLRVKKSHPEKLNIDKVKQSSGYLDIIDQDKHLFYWFFESR
NDPSTDPIILWLNGGPGCSSITGLLFEKIGPSYITKEIKPEHNPYSWNNNASVIFLEQPVGVGFSYSSKKVGDTA
TAAKDTYVFLELFFQKFPQFLTSNLHIAGESYAGHYLPKIASEIVSHADKTFDLSGVMIGNGLTDPLIQYKYYQP
MACGKGGYKQVISDEECDELDRVYPRCERLTRACYEFQNSVTCVPATLYCDQKLLKPYTDTGLNVYDIRTMCDEG
TDLCYKELEYVEKYMNQPEVQEAVGSEVSSYKGCDDDVFLRFLYSGDGSKPFHQYITDVLNASIPVLIYAGDKDY
ICNWLGNQAWVNELEWNLSEEFQATPIRPWFTLDNNDYAGNVQTYGNFSFLRVFDAGHMVPYNQPVNALDMVVRW
THGDFSFGY

Peptide

MKSVIWSLLSLLALSQALTI

SEQ0042

Homolog of YKL164C; **PIR1**

O-glycosylated protein required for cell wall stability; attached to the cell wall via beta-1,3-glucan; mediates mitochondrial translocation of Apn1p; expression regulated by the cell integrity pathway and by Swi5p during the cell cycle

S. cerevisiae null mutant: viable; increased budding index

chr4_0305

ORF

ATGAAGCTCGCTGCACTCTCCACTATTGCATTAACTATTTTACCCGTTGCCTTGGCTGGCTATGCTCCACCCGAT
GATTGGTCGACCCTTACTGCCAAAGGAGTGTACCCAGGTGCCTTCTCATCCTATTCTAATACATTCGGTATAATA
GTTGAACCTTTGACGAGCTCAGTGATTTTAACTCCTGCTACTACGACTCATGTTGTTTCACAGATTGATGACGGT
CAAATTCAGCACACTAATACAGCCTATGTAGGTACAGCTCACCAAGTTGTGTCTCAGATTGGTGACGGGCAGATT
CAGGCTACGGCTTCTGCTGTCCCATTGCCTACTGAGCTTGCTTCTCAAATTGCTGATGGACAAATCCAAGCAACT
ACTCCTGCTGGAGCTCCCGCCACTCCAGCTTCTCAGATACAAGACGGTCAGGTTCAAGCTACTTCTTCGGCTGAT
GCTCACCCTACTGCACACTCACAAGCTGAAGACATTGGGGCCCATTCTTTAAGCAGCACTGGTTTGATTCCTGGT
ACATTGACTACTGTCTTGACTTCAACTGGAAGTGACACCACACTGACCCTTGTAACTGTCGAGACTGAAGTTGTC
ACTTACACACCAGAAGTTACAGTTACTGTGAATCGGAATGCTGCCAAAGTGAAGAGGGATAACATTGAATCTGCT
TGCTTAACCCCACAGGCCCTTGGCCTAACCCTAAAGGATTCTGTTTTATTAGATCTCCAAGGCAGAGTCGGCTCC
ATCGTCGCCAATAGACAATTCCAATTCGATGGACCTCCTCCGCAAGCCGGTACTATTTATGCTGTTGGTTGGTCC
ATTACGCCTAATGGATATCTGGCATTAGGTGACTCAGAAGTCTTCTATCAATGCCTGTCTGGTTCGTTTTACAAT
CTATATGATCAGCACATTGCTGAGCAGTGTGAAGCTGTCCACTTGAAGGCTGTTGATCTCATTTCATGTTAG

AA

MKLAALSTIALTILPVALAGYAPPDDWSTLTAKGVYPGAFSSYSNTFGIIVEPLTSSVILTPATTHVVSQIDDG
QIQHTNTAYVGTAHQVVSQIGDGQIQATASAVPLPTELASQIADGQIQATTPAGAPATPASQIQDGQVQATSSAD
AHPTAHSQAEDIGAHSLSSTGLIPGTLTTVLTSTGSDTTLTLVTVETEVVTYTPEVTVTVNRNAAKVKRDNIESA
CLTPQALGLTLKDSVLLDLQGRVGSIVANRQFQFDGPPPQAGTIYAVGWSITPNGYLALGDSEVFYQCLSGSFYN
LYDQHIAEQCEAVHLKAVDLISC

Peptide

MKLAALSTIALTILPVALAG

SEQ0043

Homolog of YDL037C; **BSC1**

Protein of unconfirmed function, similar to cell surface flocculin Muc1p; ORF exhibits genomic organization compatible with a translational readthrough-dependent mode of expression

S. cerevisiae null mutant: viable; increased glycogen accumulation

FragB_0067

ORF

ATGTTATCCCTAACAAAACTTTTCATTGGTATAATATTAGGTGCCTTTGTCGCCGCAGAGCAGTCTGATGTCGAA
CCAGCATGTTATGCCAGACAGTGGGAGAACACATTCCCTCCCTCAAATATATACGTGTCTAGTGCAACTTGGGTT
CGTGATAATGTCTACGATGTCACTCTTCGCTATGAAGCCGAGTCCTTGCATTTGAGGAGTTTGGGAGAGCTCAAA
GTTATCGGCCTTAACTCTCCAGTATCTGGAACTAAAGTGGTGTACTCCCGAAACTCCAAGATTTATGAAATCAAT
GACCCTTCTAGTTGGTCAACTACAATCAGAGTGAATGCTAAAGATTCTGCTGACGGCTGCCATGTTGAGATGTAT
CCTTTCCAAATCCAAGTCGAATGGTGTCAAGCAGGCGCTGACACTGATGAATGCAGTTCCTGGCCATACCCTAGT
GTCTACGATTATGACATTGGCTGTGATAACATGCAGGATGGAGTTTCTAGAAAGCACCATGCCGTTTACAGATGG
CTCAAGCACTCTGTTTCTGGATGCTCAGCCGCCTCCACAGTTGCCGCTATCAGTTCTCAAACTACATCAAGCTCT
CAAACTTCAACGAGTTCTCAGTCATCATCCAGTACTGCTATAAGCTCAGTGACTACCCAAACCACACAAACAACT
ACTAGTGGGGTCCCGATTTGGGGACAATGTGGTGGAAATGGCCACTCTGGGTCAACTAGTTGCGCCGAAGGAGTC
TGCGTGTACGTGAACGACTGGTACTACCAATGCCAACCAGGTGCATCTACCACCACAACTTCCCAAGCATCTAGT
AGCCAAGTATTCTCACAGTTAACTACCACCGAGGAGTCTATTTCCTCCGAGGAATCCATTGGACTGACCTCTACT
GTGGAAACAACTTCCTCCGCAGAGGTCGTTGAATCGACTACTTCTGGTAACCAAGCAATCCCCACTATCGAGTTT
GACGTCGAAATTGAATTTCCAGCGGCAGAGATAACAGAGGAGGAAACTACAGAAACAGAAGAACAAGTGCCACTT
CATGGAAAATGTGGCGGTATTGGTCACACTGGCTCAACTACCTGCGCCAATGGTATCTGTACATACGTGAATGAT
TGGTATTACGAATGTCTTCCAAGAGAGGCTGAGTCCGAAGAAGAAGTCCCACTTTATGGTCAATGCGGTGGTATA
AATTATTCTGGAGCAACTAAATGTGCAGAAGGAGTTTGTACCTATGTGAACGATTGGTATTATCAATGTCTCATA
GGTCCTGCAACTGGTACACCGCAATCAACTTCCAGCACCGACAGTCAAAGTGTTTCTGAGTCATCCGCTGCTGAA
GACTCTACGCCCACTGAGGAGGCTGAAGAATCAGCTGAGGAGTCTACCTCTACTGAAGATGCTGAGGAATCCACT
GAAGATTTTGCTACTACTGAGGAGGTTGAGGAATCTACCTCTACTGAGGATGCTGAGGAATCCACCTCCACCGAG
GAGGCTGAGGAATCTACCTCTACTGAGGATGCTGAGGAATCCACCTCCACCGAGGAGGCTGAGGAATCCACAGAG
GAATCCACATCTACTGATGAGGTTGAGGAGTCTACTTCCACTGAGGAGGTTGAGGAATCTACTGAGGAATCCACT
TCTACTGAGGATGCTGAGGAATCCACCTCCACCGAGGAGGCTGAGGAATCCACAGAGGAATCCACATCTACTGAT
GAGGTTGAGGAGTCTACTTCCACTGAGGAGGTTGAGGAATCTACTGAGGAATCCACATCTACTGATGAGGTTGAG
GAGTCTACTTCCACTGAGGAGGTTGAGGAATCTACTGAGGAATCCACATCTACTGATGAGGTTGAGGAGTCTACT
TCCACTGAGGAGGTTGAGGAATCTACTGAGGAATCCACTTCTACTGAGGATGCTGAGGAATCCACTTCCACCGAG
GAAGCTGAAGAATCCACAGAGGAATCCACATCTACTGATGAGGTTGATGAATCTACCTCCACCGAGGAGGCTGAG
GAATCCACTGAGGAATCTACCTCTACTGAGGATGCTGAGGAATCCACCTCCACCGAGGAGGCTGAGGAATCCACT

GAGGAATCCACCTCCACCGAGGAGACTGAAGAATCCACTGAGGAATTGACTTCTACTGAAGAGGCTGAAGAATCC
ACTGAGGAACCCACATCTACTGATGAGGTTGATGAGTCTACGTCCACTGATGACGTTGATGAATCCACCTCCACT
GAGGGCACTGAACAATTCAGTTCCACTGATGTTCCTCAGGGAAGACCAGGATTCGAGAATCCAACTGAAGAGGTT
GAATCTTCTTCTACAGAGGAGTTCGAGGAACCAACTTCCACCGACGAAACTGACGAATCGACTGAGGAAGCCACT
TCCACTGAGGAAGCAGAAGAATCTATCTCCACTGACGATGTTGAGCAATCTACCTCCGTAGAAGAAGCTGAGGAA
TCTACAGAAGAATCTACCTCCACCGAGGCTCTTGAGGAGTCTACTTCCACTGGCGACTTTGAGAATATTTCTGCC
GTTGACGAAGAACTCGAGGAGTCTACTGAAGAATCTACGTCCACTGAAGAGGTTGAGGAATCTACCTCTACTGAG
GATGCTGAGGAATCCACCTCCACCGAGGAGGCTGAAGAATCCACTGAGGAATCCACTTCTACTGAGGATGCTGAG
GAATCCACCTCCACCGAAGAGGCTGAAGAATCCACTGAGGAATCCACTTCTACTGAGGATGCTGAGGAATCTACT
TCCTCTGATGAGGTTGAGGAATCCACTTCCACCGAAGAGGTTGAAGAATCTACTGAGGAATCCACTTCTACTGAG
GATGCTGAGGAATCCACCTCCACCGAGGAGGCTGAAGAATCCACTGAGGAATCCACTTCTACTGAGGATGCTGAG
GAATCCACCTCCACCGAGGATGCTGAAGAATCTACCTCTACTGAATATGTTGAGGAATCAACTTCAATCCCTTCT
TCAACATCCGAGCCATCTCTTTCAACTTCAATTTCTCCATCTTCCTCTAGTTCATTTGAGACTGCCACTGTTACC
ACCGATGTGTCAACAACAGTACTTACAACCACCGATTGTGGCGAGCAGACTTGTTACAAATCTTTAATCATAACC
GGTGTTACTGAGAAAACCATTACAACCAGTGGAAGAACCACAGTCATCACCACATACTGTCCTTTGCCCACTGAG
ACTGTCATTCCAACTCCTGTCGTTGTCACATCCACTATCTATGCGGATGAGACAATTACTGAGACTACATGCTAT
GTCACTGGTGCTATCGAAAAGACTGTTACTGTCGGTGGATCCCCAACTGTCGTGACTGTTCACACTACTCTACCA
ATCGAAGCCATAAAACCAACTCCAGTCACTGTTACTTCTACCATTTATGCAGATGAGACTGTTACAAAGACTACA
TGTTATGAAACTGGTGCCATAGAAAAGACGTACACAATTGGAGGATCATCTACTATTGTTACAATTCACACTCCT
CTTTCCACTGTGCAGCCAACTCCAGTCACTGTTACTTCTACCATCTATGCAGATGAGACTGTTACAAAGACTACA
TGTTATGTAACGGGTGCTATAGAGAAGACGTACACAGTTGGAGGATCTGCTTCTGTCGCCACAGTTTACACCCCT
CTGCCAACTGAACCAAGCAAGACATCAATTGTAACTTTAACATCTGTCATCTCCAACAGTGGATCCCACCCGTTG
ACAACTTCCATTGTGACCGGAGCCGTAGAAAAGACAATCACAGTTGATGGAACAGCCTCCATAGTCACAGTTCAC
ACTCCTATTGCTTCTGTTTACTCAAGCGCACATGAACATTCAAAGGTTACTAAAGCTATTACACCATCAACAACT
ACTATCATTAGAGACGTTTGCGATTCTCACTCCTGCACTCAAGCTACTATTGAGACTGGTGTAGTTTCCAAAACT
ATCACCATCGGACAATCTGTTACAACTGTGGTACCATCATATGTTCCTTTGGAACATGAACAACAGGAAGTACAC
ACTACCGCTGCCACTCCAGTTGATTCAAGTGTCCACTCTCAATCAGGAAGCCCAGTTCCACCTGCCATTACAACA
CAATCTTACAGCGCTCAAGCGTCATCAGTTCCAGTTGCTTCTTATTACGAGGGCTCTGCCAATAATCTGGTGTTT
TCAGTGGCTAGTGGAGTTGCGGCATTGATTGTTAATGCATTGTTCTTATGA

AA

MLSLTKLFIGIILGAFVAAEQSDVEPACYARQWENTFPPSNIYVSSATWVRDNVYDVTLRYEAESLHLRSLGELK
VIGLNSPVSGTKVVYSRNSKIYEINDPSSWSTTIRVNAKDSADGCHVEMYPFQIQVEWCQAGADTDECSSWPYPS
VYDYDIGCDNMQDGVSRKHHAVYRWLKHSVSGCSAASTVAAISSQTTSSSQTSTSSQSSSSTAISSVTTQTTQTT
TSGVPIWGQCGGNGHSGSTSCAEGVCVYVNDWYYQCQPGASTTTTSQASSSQVFSQLTTTEESISSEESIGLTST
VETTSSAEVVESTTSGNQAIPTIEFDVEIEFPAAEITEEETTETEEQVPLHGKCGGIGHTGSTTCANGICTYVND
WYYECLPREAESEEEVPLYGQCGGINYSGATKCAEGVCTYVNDWYYQCLIGPATGTPQSTSSTDSQSVSESSAAE
DSTPTEEAEESAEESTSTEDAEESTEDFATTEEVEESTSTEDAEESTSTEEAEESTSTEDAEESTSTEEAEESTE
ESTSTDEVEESTSTEEVEESTEESTSTEDAEESTSTEEAEESTEESTSTDEVEESTSTEEVEESTEESTSTDEVE
ESTSTEEVEESTEESTSTDEVEESTSTEEVEESTEESTSTEDAEESTSTEEAEESTEESTSTDEVDESTSTEEAE
ESTEESTSTEDAEESTSTEEAEESTEESTSTEETEESTEELTSTEEAEESTEEPTSTDEVDESTSTDDVDESTST
EGTEQFSSTDVPQGRPGFENPTEEVESSSTEEFEEPTSTDETDESTEEATSTEEAEESISTDDVEQSTSVEEAEE
STEESTSTEALEESTSTGDFENISAVDEELEESTEESTSTEEVEESTSTEDAEESTSTEEAEESTEESTSTEDAE
ESTSTEEAEESTEESTSTEDAEESTSSDEVEESTSTEEVEESTEESTSTEDAEESTSTEEAEESTEESTSTEDAE
ESTSTEDAEESTSTEYVEESTSIPSSTSEPSLSTSISPSSSSSFETATVTTDVSTTVLTTTDCGEQTCYKSLIIT
GVTEKTITTSGRTTVITTYCPLPTETVIPTPVVVTSTIYADETITETTCYVTGAIEKTVTVGGSPTVVTVHTTLP
IEAIKPTPVTVTSTIYADETVTKTTCYETGAIEKTYTIGGSSTIVTIHTPLSTVQPTPVTVTSTIYADETVTKTT
CYVTGAIEKTYTVGGSASVATVYTPLPTEPSKTSIVTLTSVISNSGSHPLTTSIVTGAVEKTITVDGTASIVTVH
TPIASVYSSAHEHSKVTKAITPSTTTIIRDVCDSHSCTQATIETGVVSKTITIGQSVTTVVPSYVPLEHEQQEVH
TTAATPVDSSVHSQSGSPVPPAITTQSYSAQASSVPVASYYEGSANNLVFSVASGVAALIVNALFL

Peptide

MLSLTKLFIGIILGAFVAAE

SEQ0044

Homolog of YDR144C; **MKC7**

GPI-anchored aspartyl protease (yapsin) involved in protein processing; shares functions with Yap3p and Kex2p

S. cerevisiae null mutant: viable

chr1-1_0379

ORF

ATGTTTGTGATCCAGCTGGCATTCCTATGTCTAGGCGTCAGCCTAACCACTGCACAACCTAGTTCACCTTTCAAG
GCAAATAAGTTTCCTTTTAAAAAGGTTCACTACTCATCAAACCCTAGCGATCGCCTTATTAAGCGAGACAACTAT
AAGAAGCTTGACTTGAGACATCTTGGCGTCTTGTATACTGCGGAAATTGAAATTGGTTCAGGCAAAACTGAAATC
GAAGTTATTGTTGACACCGGATCTGCAGATTTGTGGGTAATTGACTCAAATGCAGCCGTATGCGATTGTCCTATC
TTGAGATACAAGTTTCAGGAAAACCATCTTTTGAAGAGAAACGAGGCCTTGAATTTTGATGTCGATCTGAATAAG
CCCATTTGTGATCAATTTGGATCCTTCAATCCACAGTCATCAAGAACTTTTCAAAGCAACGACACAGCATTTAGT
ATCAGATATCTGGACAACTCTTTTGCCAATGGATCGTGGGTGAGGGATACGGTTTATGTTGGTGATTTTGAAATT
GACCAGCAAAGTTTTGCATTGGTTGATATCACAAATAACTACATGGGAATTCTGGGCCTTGGTCCTTCTAGTCAG
CAGACAACCAATAGTGATCCTACAGATAACAGTTTCACTTATCTTGGTATTCTGGATTCTTTGCGGGCCCAAGGA
TTCATTAATTCAGCCTCGTACTCGGTTTATCTGGCCCCAGATGGTAAGACTGATGATACTGATCACGATGATGGT
GAGATCCTGTTTGGTGCTATCGACGAGGCTAAAATTAATGGACAGTTGAAGTTGTTTCCATATGTCAATCCTTAT
AAATCGGTATACCCTGACCAATACGCTTCATACATCACCGTTTCCAGTATTACTGTAGCCAGTTATTTTAGTAGC
CGCTTGGTTGAAAGAATCCCTCAATTAGCTCTTTTAGACACTGGTGCCACATTTTCTTACTTGCCAACTTATACG
CTGATACGTCTCGCCTATGCCATCCATCCTGGTTTTGAGTATGTCCGACAACTGGGTTTATTTATTATAGAGTCA
AACGTACTCTCCAGTGCGAGACAAAGTACCATTGACTTCCGGTTTGGCAAAGACGTAGTAATTCGATCCAATGTT
TCAGACCATCTACTCGACGTATCACAATACTTCACATCTGGACATTATCTTGCACTTACCATCCATGAAAGTGTC
GATGGGCTTCTCATTTTGGGTGACACGTTTATCAAGTCCACCTACTTATTTTTCGACAATGATAACAGTGAATTG
GGTATTGGTCAGATCAAAATTACCAATGACGAGGATATTCAAGAAGTTGGTGAATTCACCTTAGAACGCGATTCA
GACTATTCTTCTACATGGTCCATTTACTCTTATGAAACTTCTTTGGATCCCTTAAGCACTGGCACTGGTACGGGG
TCAACCTATTCTCCTACTCGCAGTACTACAGCTAGAAGCGAACCGACTACGTCTCGACGCTCCACCACCCTTCAA
CCCAGAACAACTGTGATTCCTTCTATTGACAGGCTTTCATTGAACAGCATAACTAGTCATGGTTCCTCTACTAAC
GGAACCTCCCCAACTAATGAGACTTCTTTTGCTGAGGATGGAGGAACTTTGACACCCGAAGAAGCTTCTTTGACA
ACTTCACTAAATTCTGCTACTATTTCTGAGACTACTTTTGTCGATGTTGAAACTTCTACTACCAATGGTGCTTCA
GTTGTATCTTTGAGTGTTGGTCCCTGCATTATTGCCTTCCTACTACTCATCTCTTAA

AA

MFVIQLAFLCLGVSLTTAQPSSPFKANKFPFKKVHYSSNPSDRLIKRDNYKKLDLRHLGVLYTAEIEIGSGKTEI
EVIVDTGSADLWVIDSNAAVCDCPILRYKFQENHLLKRNEALNFDVDLNKPICDQFGSFNPQSSRTFQSNDTAFS
IRYLDNSFANGSWVRDTVYVGDFEIDQQSFALVDITNNYMGILGLGPSSQQTTNSDPTDNSFTYLGILDSLRAQG
FINSASYSVYLAPDGKTDDTDHDDGEILFGAIDEAKINGQLKLFPYVNPYKSVYPDQYASYITVSSITVASYFSS
RLVERIPQLALLDTGATFSYLPTYTLIRLAYAIHPGFEYVRQLGLFIIESNVLSSARQSTIDFRFGKDVVIRSNV
SDHLLDVSQYFTSGHYLALTIHESVDGLLILGDTFIKSTYLFFDNDNSELGIGQIKITNDEDIQEVGEFTLERDS
DYSSTWSIYSYETSLDPLSTGTGTGSTYSPTRSTTARSEPTTSRRSTTLQPRTTVIPSIDRLSLNSITSHGSSTN
GTSPTNETSFAEDGGTLTPEEASLTTSLNSATISETTFVDVETSTTNGASVVSLSVGPCIIAFLLLIS

Peptide

MFVIQLAFLCLGVSLTTAQ

SEQ0045

Homolog of YGR279C; **SCW4**

Cell wall protein with similarity to glucanases; scw4 scw10 double mutants exhibit defects in mating

S. cerevisiae null mutant: viable

chr1-3_0229

ORF

ATGCAAGTTAAATCTATCGTTAACCTACTGTTGGCATGTTCGTTGGCCGTGGCCAGACCTTTGGAGCACGCCCAC
CATCAACATGACAAGAGAGGTGTTGTCGTCGTGACCAAGACCATCGTTGTCGATGGTTCCACCGTTGAGGCTACC
GCAGCAGCCCAGGTCCAAGAGCACGCTGAAACTTTCGCAGAGTCGACTCCTTCAGCTGTTGTTTCCAGTTCATCA
GCTCCTTCCTCTGCTTCTTCTGCTTCCGCTCCTGCCTCTTCCGGTAGCTTCTCCGCTGGAACCAAAGGTGTCACC
TACTCTCCTTACCAAGCCGGTGGTGGCTGTAAGACTGCAGAGGAGGTCGCTTCCGATCTTTCCCAGTTGACCGGC
TACGAGATCATCAGATTGTACGGTGTTGACTGTAACCAAGTCGAGAACGTCTTCAAGGCCAAGGCCCCAGGTCAA
AAGTTGTTTTTGGGAATCTTCTTTGTTGATGCTATCGAGAGCGGTGTCAGTGCTATTGCCTCCGCTGTCAAGTCA
TACGGTTCTTGGGACGATGTCCACACTGTCTCTGTTGGTAACGAGTTGGTCAACAACGGTGAGGCCACTGTTTCC
CAAATCGGTCAATACGTTTCTACCGCTAAATCTGCCTTGAGATCTGCTGGTTTCACCGGTCCAGTTCTTTCTGTC
GACACTTTCATTGCTGTCATCAACAACCCAGGGTTGTGTGACTTTGCTGACGAGTACGTTGCCGTTAACGCTCAC
GCTTTCTTTGACGGTGGCATTGCTGCCTCTGGAGCTGGTGATTGGGCTGCTGAGCAGATCCAAAGAGTTTCTTCC
GCTTGTGGTGGCAAGGATGTCCTCATTGTCGAGTCCGGATGGCCTTCAAAGGGAGACACCAACGGTGCTGCCGTT
CCTTCCAAGTCAAACCAACAGGCTGCTGTCCAATCTCTAGGTCAGAAGATCGGTTCTTCTTGTATTGCTTTCAAC
GCTTTCAACGATTACTGGAAGGCCGATGGTCCTTTCAACGCTGAGAAGTACTGGGGTATCCTAGACTCCTAA

AA

MQVKSIVNLLLACSLAVARPLEHAHHQHDKRGVVVVTKTIVVDGSTVEATAAAQVQEHAETFAESTPSAVVSSSS
APSSASSASAPASSGSFSAGTKGVTYSPYQAGGGCKTAEEVASDLSQLTGYEIIRLYGVDCNQVENVFKAKAPGQ
KLFLGIFFVDAIESGVSAIASAVKSYGSWDDVHTVSVGNELVNNGEATVSQIGQYVSTAKSALRSAGFTGPVLSV
DTFIAVINNPGLCDFADEYVAVNAHAFFDGGIAASGAGDWAAEQIQRVSSACGGKDVLIVESGWPSKGDTNGAAV
PSKSNQQAAVQSLGQKIGSSCIAFNAFNDYWKADGPFNAEKYWGILDS

Peptide

MQVKSIVNLLLACSLAVAR

SEQ0046

Homolog of YLR286C; **CTS1**

Endochitinase, required for cell separation after mitosis; transcriptional activation during late G and early M cell cycle phases is mediated by transcription factor Ace2p

S. cerevisiae null mutant: viable; decreased competitive fitness

S. cerevisiae overexpression mutant: decreased vegetative growth

chr3_1003

ORF

ATGAAATTTTTTTACTTTGCGGGGTTCATATCTCTGTTACAGCTGATATTCGCCTTTGACAATTCAGCTAAAAAC
AACGTTGCTCTCTATTGGGGGCAAAACTCTGCTGGTTCCCAGGAGAGATTGTCTTACTACTGCCAGAGTGACTCT
GTCGATATAGTGTTACTGTCTTTCCTCTACATTTTCCCCGCTAACCCACTTGGATTGGACTTTTCCAATGCCTGT
GGAGACCAATTTCCCAGCGGACTTTTGAAATGTGATACCATTGCTGAAGACATTCAAACTTGTCAGTCACTGGGC
AAAAAAGTTCTTCTTTCTCTAGGTGGTGCAACAGGAACCTATGGTTTCTCCTCTGATTCCGAAGCTGAAGACTTT
GCTGAGGTCTTGTGGGACACTTTCTTGGGTGGTTCTACAGATGAAAGACCATTCGGTGATTCCATTCTGGATGGT
ATTGACTATGACGCGGAAAACAACAATCCTACTGGATACACAGCTCTTAGTGCTAAACTTAGAGAGTTCTATGCA
TCAGATCCATCGCGTACCTATTACATTGCAGCAGCTCCCCAGTGTCCTTACCCTGATGCTAGTGTTGGCGATGTG
TTGGCCAATGCTGATGTTGACTTTGTCTTTATCCAGTTTTACAACAATTACTGCGCTTTGGCAAGTACCAGCTTC
AACTGGGCTACTTGGCTCGATTACGCACAGAATACCTCACCAAATCCGAATGTTAAACTGTATGTCGGTTTACCT
GGTGGCCCTACTGGTGCATCCTCTGGATATGTCGGAACTGATGTCGTCAAACAAAGAATTGATGAGATTGGTGCT
AGCAGTAGTTTAGGTGGCATCATGTTATGGGACGCTTCTCAAGGATTTTCAAATCAAGTTGACGGAGGAAACTAC
GTGGATGCAATGAAGAGTATTCTAAATGGTTTAGGGTCTGTCGATGCATCTACCACTTCTAGTTCTCAAGCTGCT
GCTACGTCGCAGACTACTTCAACTTTGGCTACATCCATCTCGTCCACTCCAGGATCCTCATCAACTGTGTCTTCA
TCCTCATCCTTATCCTCATCATCCCTCCCCCTCATCTTCATCATCCTCTACCACGATTTGAATTCACATGGTGAA
GTTGATTCCGAAGGAACAACTTTGACTGGCACTTCGACAATCTGGTGGACTCCATCCGAAGCTCAGTCGTATGAA
ACAAGCTCGTTGTCGTCTGTCTCATCCATTCCTACTGGAAATAAAGATGTCAGTTCAATCTTAGTGATCACTGAT
GTCACTGATTCTTTAACCTCCACGAAGGAAAGTTCGGATTCTGCTTTAACTATTTCTACTTCACTCTCTTCATCA
CCCTCCCTTGCTGACAGTAGCAGGGATGGAGAAACATCTACAGTGGTCCAGGTAACTAGTTCGACAACAACAGTA
GTCGGAGGAAACACTGGCAACACTGCTAGCCCTAGCTCAACGGCTTCAGGTCTTTCAAGCTTCGTTTTCGATCCA
ACCACATTGTCTACAAAGATCACTTCTTCCGAATTAGTCGTATCTTCTACACTAGAAAACTCGTTGGGAACAACT
AAAATCGTGTTGACGCCGGTTTCTACAGTACCATCGTCAACAGACTTAGGAACCACAACAGATTCATCACCAGCA
CCAACAGCATCAACTACTGGAACCTACTTGGATTGCTCTTCGTTAAGTGGCAAAGCCAAGGCTGCTTGCTTGAAT
AAGAATTTTGCTAACGGTTTCTTCCTTTCTGGTTTACAAGGTTGCAAGGATGGTGATCATGCTTGTTCGAGCGAT
GGATACTTTTCAGTGTGTGACCACGGTCAGTGGGTATTCTTCAACTGCCCAGCTGGAACGGCTTGCTATGCCTCC
AACCAAGATGATGAAACATTTGTTGGTTGTAACTTCTCTGAATTAAAGGATAAGTTCACGAAACGAAGTTGGCTG
GACCGTTTCTTCTAA

AA

MKFFYFAGFISLLQLIFAFDNSAKNNVALYWGQNSAGSQERLSYYCQSDSVDIVLLSFLYIFPANPLGLDFSNAC
GDQFPSGLLKCDTIAEDIQTCQSLGKKVLLSLGGATGTYGFSSDSEAEDFAEVLWDTFLGGSTDERPFGDSILDG
IDYDAENNNPTGYTALSAKLREFYASDPSRTYYIAAAPQCPYPDASVGDVLANADVDFVFIQFYNNYCALASTSF
NWATWLDYAQNTSPNPNVKLYVGLPGGPTGASSGYVGTDVVKQRIDEIGASSSLGGIMLWDASQGFSNQVDGGNY
VDAMKSILNGLGSVDASTTSSSQAAATSQTTSTLATSISSTPGSSSTVSSSSSLSSSSLPLIFIILYHDLNSHGE
VDSEGTTLTGTSTIWWTPSEAQSYETSSLSSVSSIPTGNKDVSSILVITDVTDSLTSTKESSDSALTISTSLSSS
PSLADSSRDGETSTVVQVTSSTTTVVGGNTGNTASPSSTASGLSSFVFDPTTLSTKITSSELVVSSTLENSLGTT
KIVLTPVSTVPSSTDLGTTTDSSPAPTASTTGTYLDCSSLSGKAKAACLNKNFANGFFLSGLQGCKDGDHACSSD
GYFSVCDHGQWVFFNCPAGTACYASNQDDETFVGCNFSELKDKFTKRSWLDRFF

Peptide

MKFFYFAGFISLLQLIFAF

SEQ0047

Homolog of YIL099W; **SGA1**

Intracellular sporulation-specific glucoamylase involved in glycogen degradation; induced during starvation of a/a diploids late in sporulation, but dispensable for sporulation
S. cerevisiae null mutant: viable; decreased resistance to sulfanilamide
chr4_0579

ORF
ATGAAGCTATTAGACGGCTTGACAATATCTTTGTGCATAAGCATGGCCACCTCGCTGGTGTTGCCTGGCTTTGAC
CAACTTAGATTACAGCAGCAAAATGTATTATCGTTACTCCCAAAGGAAAAAATCCAGATCCAAACACTCTCAGAG
ATCTGGTCTTATGTGGATTGCGCTATTGCCTCCTTGGGACCAAAGGAGATCGTTATCGACTCAGACTTGTCGATT
TTGAAGCCTGTCCAGGCTTCTTCTATCCCTAGATCAGAATTTAATCAATGGTTGCTCTACCAACGAAATGTCTCA
TTTCATGGTGTGCTCAACAACATCGGAGGATACGGCTATAACGCAGACAACGTTTCTGTGGGATGCATTATCGCC
TCTCCTTCGAAGTCTTCTCCAAACTACTTCTATCAGTGGGTTCGTGATAGTGCAATCACTATCAACACAGTGATA
GAGTATTTGTATGACGACAACATAGACGATTCTATAGACAAAAGTGAGCTTTTAGAAGCCATAGACGGCTACATC
AACAACATTCATCACTTACAACGTCAGGACAATAGATCTGGTAAGTTCAAAGATGGCTACGCCAGTCTGGGAGAG
CCAAAGTTTATGGTCAACGGTGACCCCTTCAATGACTCTTGGGGACGTCCCCAGAGGGATGGACCAGGACTTCGT
GCACTATCCGTGGCCAACTATATTGATCTGCTGGATAAGCTGGGAATTCAGAAAGATTCAGAAAGACTCTCATTC
ATCTACAATGAAGTATTGAAGCCAGATCTCACTTATGTCACCCTCTATTGGGACCACGATGGATTTGACCTGTGG
GAAGAGACTAATGGATTGCATTTTTTCACTAGTTTGACTCAACTGAAGGCTCTTAGAAGAGGTATTGAACTGGCT
CTCAGGTTTGAGGACTACGAATTCCATCGTGACCTGAAGAATGCTTATGCCCAACTAAGAAACTTTATTGTTGGT
AGAGACTCTGGATTCCAAGATCCTCGTTTCCCTCATATCATTGAGCACCCTAATTTGGTAGACGCTCCAGCCCAT
ATTAGAAGAAATGGTTTAGATATCGCAACCATCATTGCTGCGCTGAGATCTCACGATGTTGATGATGCAGGCGAT
ATAGTCAACATCCCGTTTGGCGTGGACAATGCAAGGGTCCTCAATTCCCTGACCTATTTGGTTAATGACATGAAG
TTCCGGTATCCACTCAACCAGGGTCGTATCGTCCCTGGTTTAACTCTAGGAATGGCCTTAGGAAGATATCCTGAA
GACACCTACAACGGTGTTGGTACCAGTGAAGGAAATCCATGGTTCATTTCAACTTCCAGTGCAGGTGAACTCATC
TACAAGCTGTTGTATTTGCAATACAAATATCAACAAGACTTCGTGATAGATTCTTCCAATAGGGAATTCTACGAA
ATGTTTTTGGATCTGCCCAAAGACTCTGCTCGTGGAATTCAAATTCGCATACCTTTCGGAAGCGACACCTATCGT
GCACTGTCCGTTTCTTTGATCCAATATGCGGATTCCTTCTTGGACGTGATCAGGGAACATGTGGATAACGAAGGT
CAAATGAGTGAACAATTTAACAGGTACAATGGTTATATGCAAGGAGCCGAGAAATTAACTTGGAGTTTTGGTTCT
GTTTGGAGTTGTCTCCGTTGGAGGAACAGAATACTAAACTTACTTGCAACAATTAATGAATAG
AA
MKLLDGLTISLCISMATSLVLPGFDQLRLQQQNVLSLLPKEKIQIQTLSEIWSYVDCAIASLGPKEIVIDSDLSI
LKPVQASSIPRSEFNQWLLYQRNVSFHGVLNNIGGYGYNADNVSVGCIIASPSKSSPNYFYQWVRDSAITINTVI
EYLYDDNIDDSIDKSELLEAIDGYINNIHHLQRQDNRSGKFKDGYASLGEPKFMVNGDPFNDSWGRPQRDGPGLR
ALSVANYIDLLDKLGIQKDSERLSFIYNEVLKPDLTYVTLYWDHDGFDLWEETNGLHFFTSLTQLKALRRGIELA
LRFEDYEFHRDLKNAYAQLRNFIVGRDSGFQDPRFPHIIEHPNLVDAPAHIRRNGLDIATIIAALRSHDVDDAGD
IVNIPFGVDNARVLNSLTYLVNDMKFRYPLNQGRIVPGLTLGMALGRYPEDTYNGVGTSEGNPWFISTSSAGELI
YKLLYLQYKYQQDFVIDSSNREFYEMFLDLPKDSARGIQIRIPFGSDTYRALSVSLIQYADSFLDVIREHVDNEG
QMSEQFNRYNGYMQGAEKLTWSFGSVWSCLRWRNRILNLLATINE
Peptide
MKLLDGLTISLCISMATSL

SEQ0048
Homolog of YEL040W; **UTR2**
Cell wall protein that functions in the transfer of chitin to beta(1-6)glucan; putative chitin transglycosidase; glycosylphosphatidylinositol (GPI)-anchored protein localized to the bud neck; has a role in cell wall maintenance
S. cerevisiae null mutant: viable
chr1-1_0293

ORF
ATGAGACCAGTGCTTTCGTTATTACTCTTGCTGGCTTCTTCGGTACTCGCTGACGAGGTCATTGAATGCGATGCT
GACAACAAGTGCCCTGAAGATAAACCATGTTGTAGTCAATATGGAGTATGTGGGACTGGTGTGAATTGTCTAGGA
GGATGTGACCCTAGACATTCTTTCAACGCGTCCGCATGTTTACCCATGCCAGTTTGTCGTGACGTGGATTTGAAG
GCTAGCACTGATGCTTTCGAAATTGATACCAACTACCTTGGTGATGCTAATGAAACCGACTGGGTTTACAACGGT
TATCTAATAGATTATGACGACTCAGTATTGTTGGCCATGCCCAAGGAGTCTTATGGTACCGTTGTGTCTTCCACC
TTCTATGTGTGGTACGGAAAAATCACAGCTACTCTTAAGACTTCCAGAGGTGCCGGTGTTGTCACGTCTTTCATT
TTATTCTCCAATGTGCATGATGAAATCGATTGGGAATTTGTTGGTTATAATCTAAGTCAAGTCGAAACGAACTAC
TACTACCAGGGTGTCTTGAACTATACCAACGGTCGTAATGTTTCCCTAGAGGAGGACGTCAATTCCTTTGAATAC

TTCCACGATTATGAAATTGACTGGAAGGAAGACGTCATAACCTGGTCTATTGACGGCGACGTTGTCAGGACACTG
AAAAAGGAAGACACTTATAACGAAACTACAGACAAATACATGTTTCCTCAAACTCCATCGAGGGTTCAGTTATCC
ATCTGGCCGGCTGGTGCTGAATCCAATGCGATAGGTACAGTTTCTTGGGCTGGTGGAAACGTTGACTGGGATTCC
GAAGACATCCAAGACCCAGGATACTTTTACTACACGTTGAAAGAGCTGACTGTTGAGTGTTATGACGTCCCAGAT
GGTACCGAGGAAGATGGTGAATTGGCATATTACTTCAAAGAGTCCGATGCTTTTGACCAAGGGGATATCATAATT
ACTAATAACAGCACCAAAATAAAGTCTTTAGACGACACTGGCTTTGACCCCGATGAAGACGATGACGATGACGAA
TCCAGCAGTAGCAGCAGCAGCAGCAGATCGTCTTCAAGTTCTTCAAGGACCGGCAGTTCATCTACATCCAGT
GCTACTTCCACTTCCACTTCTAACTCAAATGATGATGATGATAACAATAATTCGTCTCCTACTACTTCGTCTGGC
ACTTCCTCCGCAGCTTCCGGATTCGTTCAAAATATGAGCCAAACTTCAGGCTCTTCATCAGCAACGTCAAATAAT
GCCGCTGCCTCTCTTTCAGCTGGTTTCCTTACTACTATTTCCTTTTTTGCATCTGTATTAGGATTCCTTTAG

AA

MRPVLSLLLLLASSVLADEVIECDADNKCPEDKPCCSQYGVCGTGVNCLGGCDPRHSFNASACLPMPVCRDVDLK
ASTDAFEIDTNYLGDANETDWVYNGYLIDYDDSVLLAMPKESYGTVVSSTFYVWYGKITATLKTSRGAGVVTSFI
LFSNVHDEIDWEFVGYNLSQVETNYYYQGVLNYTNGRNVSLEEDVNSFEYFHDYEIDWKEDVITWSIDGDVVRTL
KKEDTYNETTDKYMFPQTPSRVQLSIWPAGAESNAIGTVSWAGGNVDWDSEDIQDPGYFYYTLKELTVECYDVPD
GTEEDGELAYYFKESDAFDQGDIIITNNSTKIKSLDDTGFDPDEDDDDDESSSSSSSSSRSSSSSRTGSSSTSS
ATSTSTSNSNDDDDNNNSSPTTSSGTSSAASGFVQNMSQTSGSSSATSNNAAASLSAGFLTTISFFASVLGFL

Peptide

MRPVLSLLLLLASSVLAD

SEQ0049

Homolog of YOL132W; **GAS4**

1,3-beta-glucanosyltransferase, involved with Gas2p in spore wall assembly; has similarity to Gas1p; localizes to the cell wall

S. cerevisiae null mutant: viable; normal sporulation

chr3_0184

ORF

ATGCTATATTTGGTGACAGTTCTTTTGTTTCTCGTGCATGTGGTCCTTGGCTATATCCATCCCATCACCATTAAG
GGGAAGCACTTTTATGACTCTCTTACGGGTGAGCTTTTCATGGTCAAGGGAGTTGATTATCAGCCAGGAGGTGCC
TCTGGAGTTTCAACGGACAAAGATCCTTTATCAGAAATAGAGGAATGTGCTAGAGATATTCTCTTGTTCCAGGAA
CTTGGAATAAATACTATTAGAGTTTATTCGGTTAATCCTGACTTGAATCATGACAAATGCATGTCGCTTTTGGCT
TCTGCTGGGATATACTTGATCTTGGATGTTAATTCACCTTTGCCAAATCAAAGCATTAATAGATATGAGCCTTGG
AGCTCCTATAATCATGATTATCTATACCATATCTTCAAAGTGGTTGAACAATTTTCTCATTATAACAATACCTTG
GCCTTTTTTGCTGGTAATGAAGTCGTGAATGATAAGGTCTCCGCCACACATTCTTCTAACTATATGAAAGCAGTT
GTAAGAGATTTGAAAACATACATCACACATCATAGCGATAGACCTATCCCTGTTGGATATTCTGCAGCCGATGAT
CTTTCTTTTCGAATTCCATTGGCAAAGTACCTAGAATGCTCAAATAGTACAGCTGAATTGGATAGCGTTGACTTT
TATGGAGTCAACAGCTACCAATGGTGTGGCTATCAAAACATGCAAACGAGCGGCTATGACCAGTTGGTAGCTCAT
TATGCGGATTATACCAAACCAATCATGTTCTCAGAATACGGATGTAACGAAGTGACTCCAAGAATTTTCCAAGAG
GTGGAGGCTATTTATTCTTCTAAGATGAGCTCTGTTTTCAACGGAGGTCTTGCTTATGAGTTTGCCCAAGAACCG
AATAATTATGGAATGGTGGAATATCTTTCAGATGGGAGAGTTAAACTATTGAAGGATTTTGAAACTTTCAAGAAC
CAATTAGCTAAGGCCTCAGAAACTTACCACATAAAGCAGATGAATGAAAATGTTAGATCCGCACCGCTTAAATGT
CAGGGCACTTATGAAAATTTGGGGGACAAATTGGTTGTTCCTCAAAGTCTAGGAATGGCCTTCATTACGCAGGGA
GTCAAAGGTGAGAAAGGCTCCTACGTAGAGTTGAGTGATGATGATTTTGTTGTAAAGAAAGAAATCTTTCTATCA
GATGGAACAAAATTCCAGAAGAGCAAAATTGTTGCTACCCATAATTTGAATGGACCAGATACACAGCAAACAGTT
CCAGAAAAATCCCAGGGCTCAAAAAATGGACAAATAACAAACCCAAATCCCAAACCCAAACCCCAAACCAGGAAG
ATTCCAAACAAAAAGACCAAACCAAAGCAGAATAAACAGGAAAAGTCAAAGGTACAGAAGCAAATTGCAAGAGAA
AAAAAAATCAGAGCAATCAGGCAAAGGCAACAGGAATGGAACTTTAAGGTTTGA

AA

MLYLVTVLLFLVHVVLGYIHPITIKGKHFYDSLTGELFMVKGVDYQPGGASGVSTDKDPLSEIEECARDILLFQE
LGINTIRVYSVNPDLNHDKCMSLLASAGIYLILDVNSPLPNQSINRYEPWSSYNHDYLYHIFKVVEQFSHYNNTL
AFFAGNEVVNDKVSATHSSNYMKAVVRDLKTYITHHSDRPIPVGYSAADDLSFRIPLAKYLECSNSTAELDSVDF
YGVNSYQWCGYQNMQTSGYDQLVAHYADYTKPIMFSEYGCNEVTPRIFQEVEAIYSSKMSSVFNGGLAYEFAQEP
NNYGMVEYLSDGRVKLLKDFETFKNQLAKASETYHIKQMNENVRSAPLKCQGTYENLGDKLVVPQSLGMAFITQG
VKGEKGSYVELSDDDFVVKKEIFLSDGTKFQKSKIVATHNLNGPDTQQTVPEKSQGSKNGQITNPNPKPKPQTRK
IPNKKTKPKQNKQEKSKVQKQIAREKKIRAIRQRQQEWNFKV

Peptide

MLYLVTVLLFLVHVVLGY

SEQ0050

Homolog of YBR286W; **APE3**

Vacuolar aminopeptidase Y, processed to mature form by Prb1p

S. cerevisiae null mutant: viable; decreased telomere length; decreased acid pH resistance

chr1-4_0611

ORF

ATGAAATATTTGCCACTCGTTGCTACCCTGGCCTCTTCGGCCCTCGCTGCTGGCATCAACTTCGCCCAATTACTG
GACCAGAAGCCACTGGACATTGCCGATAATGTTAAATGGGAATTGAAGCCTGAGGTCGACTCTGCTGCTCTTCAA
AGTGCAGTCAATGAGCTAGACTTGAAAATCGAAGCCAGCTATTTGTTTAAAGTTGCACATGGTTCCGTCTTTGAA
TACGGACATCCTACCAGAGTCATCGGTTCTCCTGGTCACTGGTCCACAATCAACCATGTCCTCGACACATTACAT
AACTTCAAACACTACTACGACGTTGACGTTCAGCCATTTGAAGCCTTTACCGGTATCCTTAAGTCTTTCTCATTG
ACCATTAACGGAGTTGCACCAAAGTCTGCAGAAGCTTTAGATTTAACTCCTCCTACTCCTGGCGGTTTTCCAGTG
ACCGGTCCAGTCGTTTTAGTTGATAATTATGGTTGTCAAGCTTCTGACTATCCATTCAACGTGACTAACGGAATT
GCCTTAATTCAAAGGGGTTCTTGTTCATTCGGTCAAAAATCAGAACTTGCTGGTCTCCGTGGAGCCAAAGCCGCT
CTCATTTACAACAACGTGCCAGGTAGTGCTAAGGGAACCTTAGGTGCCCCAACTCCTCATCAGGTACCATCGTTG
TCACTTTCTCAGGAAGATGGAGAGGCCGTCAAGCGTCAGCTTCTGACTTCTGGAAGCGTAATTGCAACTGTCGCT
GTCGATTCCTACGTTAAGAAGTTCAAAACCAAGAATGTGATTGCTACCACTCGTTACGGTAATGATAGCAACATT
GTGATGCTAGGTGCACATTCAGACTCTGTTGCTGCTGGACCAGGTATCAATGACGATGGTTCTGGTACCATCTCT
CTTTTGAACGTGGCCAAATACCTAACTAAATTCAAAGTTAATAACAAGGTTCGTTTCGCTTGGTGGGCAGCTGAA
GAAGAAGGATTACTTGGATCCGACTACTACGTTTCAAAGTTAACCCCCAAGGAGAAATCTCAGATTCGTTTGTTT
ATGGACTACGATATGATGGCTTCCCCTAACTACGCCTACCAGGTCTATAATGCCACTAACAGCGAGAACCCAGTT
GGATCTGAGGAGCTTAAGAATTTATACATTGACTGGTACGTTGAACAGGGTCTGAACTACACTCTAGTTCCATTT
GATGGCCGATCCGACTATGATGGATTCATCAAGAGCGGTATTCCCGGAGGTGGTATTGCTACCGGAGCAGAAGGT
TTGAAGACCGAAGAGGAGGCTGAACTATTTGGTGGTGAAGCTGGAGTTGCATATGACCCATGTTACCACTCTCTT
TGTGACGATTTGGCCAACCCTGACTATGTTCCATGGGTTGTCAATACTAAATTAATTGCCCACAGTGTCGCCACT
TATGCAAAGAGCTTGGACGGATTCCCATTGCGTGAGGAGCCTAGCCCATTCAAGATGACTGCCCAGTCAAACTTC
AAGTACCACGGTCCAAAACTTGTCCTTTAG

AA

MKYLPLVATLASSALAAGINFAQLLDQKPLDIADNVKWELKPEVDSAALQSAVNELDLKIEASYLFKVAHGSVFE
YGHPTRVIGSPGHWSTINHVLDTLHNFKHYYDVDVQPFEAFTGILKSFSLTINGVAPKSAEALDLTPPTPGGFPV
TGPVVLVDNYGCQASDYPFNVTNGIALIQRGSCSFGQKSELAGLRGAKAALIYNNVPGSAKGTLGAPTPHQVPSL
SLSQEDGEAVKRQLLTSGSVIATVAVDSYVKKFKTKNVIATTRYGNDSNIVMLGAHSDSVAAGPGINDDGSGTIS
LLNVAKYLTKFKVNNKVRFAWWAAEEEGLLGSDYYVSKLTPKEKSQIRLFMDYDMMASPNYAYQVYNATNSENPV
GSEELKNLYIDWYVEQGLNYTLVPFDGRSDYDGFIKSGIPGGGIATGAEGLKTEEEAELFGGEAGVAYDPCYHSL
CDDLANPDYVPWVVNTKLIAHSVATYAKSLDGFPLREEPSPFKMTAQSNFKYHGPKLVL

Peptide

MKYLPLVATLASSALAA

SEQ0051

Homolog of YLR300W; **EXG1**

Major exo-1,3-beta-glucanase of the cell wall, involved in cell wall beta-glucan assembly; exists as three differentially glycosylated isoenzymes

S. cerevisiae null mutant: viable

chr2-1_0454

ORF

ATGAACTTGTACCTAATTACATTACTATTCGCCAGTCTATGCAGCGCAATTACTCTCCCAAAGAGAGATATAATC
TGGGATTACTCCAGTGAAAAAATCATGGGTGTCAACCTCGGTGGATGGCTGGTTTTGGAGCCCTATATCACCCCT
TCGCTTTTTGAAGCTTTCGGCGATGATGTACCAGTAGATGAATATCGTTACACTGAACGACTAGGTAAATCGCTG
GCTTTGGATAGACTTCAACAGCATTGGTCAACGTTCTACGACGAAAAAGATTTTCAAGACATCGCAGCTTACGGA
CTTAACTTTGTTCGTATACCAATTGGCTACTGGGCCTTCCAATTGCTGGATGATGATCCTTATGTTCAAGGCCAG
GAAGAGTATCTTGACAAGGCTTTGGAGTGGAGTAGAAAGCATGGACTGAAAGTTTGGATTGACCTTCATGGAGCT
CCAGGATCTCAGAACGGCTTTGACAACTCCGGTAAGCGAGACAGTTGGGATTTCCAAAATGGTAACAACGTTCAA
GTAACTTTGGATGTTCTGAAATATATTTCAAAAAAATATGGAACCACGGACTACTACGATGTGGTCATTGGTATT
CAACTCTTAAACGAACCATTGGGACCTATTTTAGACATGGATAATCTAAGACAGTTCTATGCGGATGGTTATGAT
CTAGTTAGAGATGTTGGGAACAACTTTGTTGTAATCCACGATGCATTTTACCAGGCGCCAGAGTACTGGGGGGAC
GATTTCACCTCAGCGGAAGGTTACTGGAACGTGGTGCTTGATCACCACCATTATCAAGTCTTCGATGCAGATGAA

TTGCAAAGAAGTATCGATGAACATATAGAAGCCGCCTGTGATTGGGGTAGAGATGCAAATAAAGAGTACCACTGG
AACCTCTGTGGTGAATGGTCGGCAGCACTTACTGATTGTACTCCTTGGTTAAATGGTGTCGGAAAAGGCACGAGA
TATGAAGGTCAACTTGATAACTCCCCTTGGATCGGATCTTGTGAGAATAGCCAGGATCCTTCGAAATTGAGCTCT
GAACGTATCTGTGAGTACAGAAGGTACGTAGAAGCCCAGCTAGATGCTTTCTTACACGGGAAAAGCGCAGGTTTT
ATTTTCTGGTGTTTCAAGACAGAGGCCAGTTTGGAGTGGGATTTTAAAAGGTTGGTTAATGCCGGTATCATGCCT
CAGCCATTGGACGACAGACAGTATCCAAATCAATGTGGGTTCTAA

AA

MNLYLITLLFASLCSAITLPKRDIIWDYSSEKIMGVNLGGWLVLEPYITPSLFEAFGDDVPVDEYRYTERLGKSL
ALDRLQQHWSTFYDEKDFQDIAAYGLNFVRIPIGYWAFQLLDDDPYVQGQEEYLDKALEWSRKHGLKVWIDLHGA
PGSQNGFDNSGKRDSWDFQNGNNVQVTLDVLKYISKKYGTTDYYDVVIGIQLLNEPLGPILDMDNLRQFYADGYD
LVRDVGNNFVVIHDAFYQAPEYWGDDFTSAEGYWNVVLDHHHYQVFDADELQRSIDEHIEAACDWGRDANKEYHW
NLCGEWSAALTDCTPWLNGVGKGTRYEGQLDNSPWIGSCENSQDPSKLSSERICEYRRYVEAQLDAFLHGKSAGF
IFWCFKTEASLEWDFKRLVNAGIMPQPLDDRQYPNQCGF

Peptide

MNLYLITLLFASLCSAI

SEQ0052

Homolog of YDR349C; **YPS7**

Putative GPI-anchored aspartic protease, located in the cytoplasm and endoplasmic reticulum *S. cerevisiae* null mutant: viable; abnormal budding pattern; abnormal cell shape; decreased cell size; decreased resistance to hygromycin B

chr3_0394

ORF

ATGTACCAGGCGTTGTTGGTTTTGTCTCTGATATGCTTTTCGTCGGCTAATTTTGTTAAGCTGCGAAGCAACGCT
GGTATGTTTTATGATACTATGGCTGGAGTTCCACGTTCAGATGAAGAGTTCTGGTTGCGTTTGGATATTAACCAA
GGTCTCTCTTGGACTCTGGATAGTAGCTACTACTCCTGTAATGGCTCAAATGTTTCGTCTTCCCTGTGTTTCAAT
TCTGCTCAAAACGTTTACGATGCTTCCAATAGTCCAACTGCAGATTTCGTTGATGTCTACGCAAACACAACTGTA
AACAATACAGATGAGGCATCGGCCGAGAGAGTAAATCTTACAAACAACTTATTTGCTGATGGCGTTTATATGGAA
GACAATTTTTACGTCACATTGAATAATGGAGCAAGAATGACTGCTACAGATCTGAAATTTTTGAATGCCCACAAT
AGTAGCGCCGCTGTGGGGTCTTTGGCGTTGGGGAGTTACACCTCACAGGACGTGCCAACTTTCTTACAAAGACTC
CAAAGCGGTGGTCTTATTGAATCCAACTCGTTTTCATTGGCATTAAACGAAATCGATTCTTCATATGGAGAGCTC
TATTTGGGGACAATAAACTCTACCAAGTATGTCGAGCCTCTGGTAGAATTCGATTTTATTCCGGTGTCAGATCCC
AATGGAGTTTTTGGATTCGATTGGGAAGATACATTCCCTACAGTTCCGATCAGCGGATTAAGCATGTCTTCGAAT
GACAAACAGAGAACTGTCTTTTTCCCCAATGAGTGGAACAACACGGTCTTAACGGGAACATACCCACTTCCAATG
ATGTTAGATTCAAGAAACATCTTTATCCATCTTCCATTCTCTTCAATCATACATATAGCAGTGCAGCTTAATGCA
CTGTATCTTGATACACTTCATAAATGGGCCGTGAACTGTTCTGTTGGTCAACTGGACGCAACTTTAAACTTTCAC
ATGGGTAACCTTACCGTTCATGCTCCTATCAAGGAGTTGATTTATCCAGCATACCAAGGAGACAAAAGGCTGAGC
TTTGCTAATGGAGAAGATGTTTGTATTCTTGCCATGGCTCCTGATGTTTACATTGGTTATCCACTGCTAGGAACC
CCCTTTTTAAGGAATGCAGTGGTTGCCGTTAATCATGATTCAAAAAAGGTCGCCGTTGCCAATCTTAATAGAGAT
AGCATTCCTCCCGCTTCGAACGTTTCTGTTTCGGAATCAATGGGAGTTTATGTTCCTCCACCTGTTTCAACTTCA
AGAACATCGGAGAGACCGTCCACACTAGATGAGACTAGTACAGCCAATTTTGACAAAAGGGAAGAGTCTGCAATA
TCATCAAGTTCAGTCACTAACAGCTCGTCTAGAAATTCTTCAACCATAACTTCTTCAGGAACTCAAACCGAGCAA
ACATCAGGCATAGCTACCATCGAAACAGATAGCATACCAGGAGCTCTAGGGAATAATTTAACTGATTATTCAACG
CTGACTCTAACAATATACACCAATTCCGAAGTGGACGAACTCAATCCTAACATAGCAACAGCATTCATTTCCAAT
GGTTCTATTTATTCAGAGCCTTACCCCTTTTCCGGAACTGCAGTTGCTGAATCATTCAGTGCATCACCTTCACAG
GCTGAAGGATCGAACTCATCGTCCTCAGGATCTTCTTTAGTTTTGTGTTTCTTTACATCATTGGCCAGTCTGTTG
ACTGTGAGCTGTCTACTACTGTAA

AA

MYQALLVLSLICFSSANFVKLRSNAGMFYDTMAGVPRSDEEFWLRLDINQGLSWTLDSSYYSCNGSNVSSSLCFN
SAQNVYDASNSPTADFVDVYANTTVNNTDEASAERVNLTNNLFADGVYMEDNFYVTLNNGARMTATDLKFLNAHN
SSAAVGSLALGSYTSQDVPTFLQRLQSGGLIESNSFSLALNEIDSSYGELYLGTINSTKYVEPLVEFDFIPVSDP
NGVFGFDWEDTFPTVPISGLSMSSNDKQRTVFFPNEWNNTVLTGTYPLPMMLDSRNIFIHLPFSSIIHIAVQLNA
LYLDTLHKWAVNCSVGQLDATLNFHMGNLTVHAPIKELIYPAYQGDKRLSFANGEDVCILAMAPDVYIGYPLLGT
PFLRNAVVAVNHDSKKVAVANLNRDSIPPASNVSVSESMGVYVPPPVSTSRTSERPSTLDETSTANFDKREESAI
SSSSVTNSSSRNSSTITSSGTQTEQTSGIATIETDSIPGALGNNLTDYSTLTLTIYTNSEVDELNPNIATAFISN
GSIYSEPYPFSGTAVAESFSASPSQAEGSNSSSSGSSLVLCFFTSLASLLTVSCLLL

Peptide

MYQALLVLSLICFSSAN

SEQ0053

Homolog of YDR456W; **NHX1**

Endosomal Na+/H+ exchanger, required for intracellular sequestration of Na+; required for osmotolerance to acute hypertonic shock

S. cerevisiae null mutant: viable; decreased hyperosmotic stress resistance; decreased respiratory growth rate; decreased chloride accumulation; decreased resistance to gentamycin, hygromycin B and 1,2-bis(2-aminophenoxy)ethane-N,N,N′,N′-tetraacetic acid

S. cerevisiae overexpression mutant: decreased vegetative growth

chr2-1_0156

ORF

ATGATCCGATTGTTGGCTTTATTCTTTGCGAGACAGATACTTGCCAATGAAATAACTGATCCCACAGATGAGAAT
CCGGTGCTCGTCGGGCCCGAAGCTCCTGAAGAGGAAACCAATCCGCTGACTGAAGAGATCTTCTCGTCTTGGGCT
TTGTTCATTGTCTTGTTGCTCGTCGTCTCCGCTCTATGGTCGTCTTACTATTTGCAGCAAAGGAGGGTCAAGTCA
ATTCACGAAACAGTATTATCGATCTTCTATGGAATGTTTGTGGGGTTGATACTGAGGGTCACTCCTGGCCACTAC
ATTCAAGACGCGGTTAAGTTCAATTCAGGTTATTTCTTTAATTTCTTGCTGCCTCCTATCATCTTGAATAGTGGT
TATGAGTTGCATCAGGCTAATTTTTTTCGAAACATCGGCTCAATTCTGACTTTTGCTATTCCAGGTACGTTTATC
AGTGCGATTGTGCTGGGTGTCATCTTGTTTATTTGGACCAAGTTGGGTCTGGATGGAATTGATGTATCCTTGGTT
GATGCTTTATCTGTTGGCGCTACCCTTTCAGCCACGGATCCAGTAACTATTTTGTCTATTTTCAATTCCTACAAG
GTCGACCCAAAGTTGTACACTATCATTTTCGGTGAATCATTGCTAAACGACGCTATCTGTATCGTGATGTTTGAA
ACGTGCCAGAAGTTCCACGGTCAGGCAGTCAGCGTTAGCAGCGTGCTCAAAGGTATTGGTCTCTTTCTTATGACA
TTCACTGTTTCATTGTTGATTGGGGTTGTCGTCGGTGTATTTATTGCCTTGGTCCTAAAACATTCCCTCATCAGA
AGGTACCCTCAAATCGAAACATGTCTCGTGTTGCTTTTCGCATACGAATCCTATTTCTTCTCCAATGGTTGCCAC
ATGTCTGGAATTGTGTCTCTTCTATTCTGCGGTATTACGATGAAACATTATGCATACTTCAACATGTCCAGAAGG
ACTCAGATTGCTACCAAGTACATTTTCCAGTTGTTGGCCCAGCTGAGTGAAAATTTCATTTTCATTTACTTAGGA
TTGTCCCTGTTTACAGAAGTTGAACTAGTATTCAGACCCATGCTGATTATTGTGACCACAATCTCTATCTGCATT
TCCAGATGGTGTGCCGTTTTCCCACTTTCACGACTGATCAATTGGACTACAAGAGCTAAACACAAAGGTGGTTCC
TCAGCAATCAATTATACCCAGGATGAGATCCCGCCAAATTACCAAATGATGATCTTTTGGGCCGGCTTGAGAGGT
GCAGTCGGTGTCGCACTTGCTATGGGATTGCAAGGTGAAGCTAAGTCTTCTCTACTAGCCACTGTACTAGTGGTT
GTGGTATTGACCGTTATTTTGTTTGGTGGGACAACTGCTGGTATGTTGGAAACCCTGAATATTCGAGTTGGTGTT
ATCGACGAACAGGAAAGTGATGATGAGTTTGATATAGAAGCTCCAAAACCAATGCAGCTGAACCAAATTATCCCT
GGGGCAACAACTCCTGTCTATTCCATATATTCTGATGCAGCAGGTTCAAGGTCCCGAACAGGTTCTCAGCAGAGC
TTTTACAACAATGAGGATGAGGATGCGGCTGATGCTCCACCAGATATGAACGATGATGAGATGACCAGTGACTTG
GATGACATCCCACCTATGGCGAACCAGGCACAATCCAGTAAAATAAACCTCCAAAGCATGAGTTTCAACAATCTA
TTGAGTATGGACGACCATGCCAAATGGTTTACACACTTTGACGAACAAGTGTTAAAGCCAGTTTTGTTAGATAAT
CTAACTGAAAACAACAACAGCAAGGACAAAATAGAGGACTAA

AA

MIRLLALFFARQILANEITDPTDENPVLVGPEAPEEETNPLTEEIFSSWALFIVLLLVVSALWSSYYLQQRRVKS
IHETVLSIFYGMFVGLILRVTPGHYIQDAVKFNSGYFFNFLLPPIILNSGYELHQANFFRNIGSILTFAIPGTFI
SAIVLGVILFIWTKLGLDGIDVSLVDALSVGATLSATDPVTILSIFNSYKVDPKLYTIIFGESLLNDAICIVMFE
TCQKFHGQAVSVSSVLKGIGLFLMTFTVSLLIGVVVGVFIALVLKHSLIRRYPQIETCLVLLFAYESYFFSNGCH
MSGIVSLLFCGITMKHYAYFNMSRRTQIATKYIFQLLAQLSENFIFIYLGLSLFTEVELVFRPMLIIVTTISICI
SRWCAVFPLSRLINWTTRAKHKGGSSAINYTQDEIPPNYQMMIFWAGLRGAVGVALAMGLQGEAKSSLLATVLVV
VVLTVILFGGTTAGMLETLNIRVGVIDEQESDDEFDIEAPKPMQLNQIIPGATTPVYSIYSDAAGSRSRTGSQQS
FYNNEDEDAADAPPDMNDDEMTSDLDDIPPMANQAQSSKINLQSMSFNNLLSMDDHAKWFTHFDEQVLKPVLLDN
LTENNNSKDKIED

Peptide

MIRLLALFFARQILAN

SEQ0054

Homolog of YMR238W; **DFG5**

Putative mannosidase, essential glycosylphosphatidylinositol (GPI)-anchored membrane protein required for cell wall biogenesis in bud formation, involved in filamentous growth, homologous to Dcw1p

S. cerevisiae null mutant: viable; competitive fitness: decreased; resistance to rapamycin: decreased; resistance to wortmannin: increased; resistance to oleate: decreased; sporulation: decreased

chr1-1_0147

ORF

ATGTGGATCTTGTGGCTTTTGACGACAGTCTCAGCTCTATCGCTGGACGTTGACTCTAAGGACTCCATCTGCGAG
GCCACTACACATATTATAAATGGTATGATGGACTACTATGAAGGAACGAGGTATGGAGGTACTGTGGGAATGTTC
CAGACACCATATTACTGGTGGCAGGCGGGTGAAGCATTTGGAGGAATGATCGAAAACTGGTTCATGTGTGACAAC
AACACGTACGAAGAGATCATTTATGATGCTCTACTGCACCAAACAGGTGACAACTTTAACTATATCCCTGCCAAT
CAATCCACCACTGAAGGAAATGATGATCAGGGTTTCTGGGGATTTGCAGCAATGGGTGCCGCTGAGAGAAACTTT
ACTGATCCCCCAGAAGATTATCCAAGTTGGGTTGCGCTGACTCAAGCAGTCTACAACACAATGTGGGCTCGTTGG
GATAATGCCACCTGTTCTGGAGGTTTACGTTGGCAAATTTTCACGTGGAATTCAGGCTACGATTACAAGAATACT
ATCAGTAATGGGTGTCTTTTGCACATTGCCGCGCGTCTGGCCAGGTTTACTCAAAATGAGACTTATGGCGAGACG
GCGGAACAAGTCTGGCAATGGCTAGATGATGTTGGGTTTATCAACGACCAAGGTGGAACATATCTTATCTACGAT
GGTGCGAATATCGTAAGTGGAGAGTGTACTGATATTACAACTATCGAGTGGACCTATAACTACGGTGTCTTAATG
GCTGGCTGTGCTTATATGTACGATTTCACCCATGATGAAGTGTGGCTAACTCGAACGACCACGTTGCTTTCTTCA
ATATCCATTTTCCTGAATGACAATGTCATTTATGAACAACAATGCCAGGCTGCTATGAATTGTAACAACGATCAG
CGGTCTTTCAAGAGTATTTTCAGTCGATGCTTGGGCTACACGGCCAAGCTAGTACCCTCTTTGACCGACCAAATA
ATGACAATCCTGACAGCTTCGGCTGAAGGGGCTGCCAAATCGTGTTCTGGTGGTACTGATGGAGTTACCTGTGGT
CTCAATTGGAATATAGGTGCTTATGATAACATGTATGGGCTAGGTGAGCAGATGAGTGCTTTGGAGGTAATTACA
CAGTTATTGGTGCTTGATTCCCCGGCTCCTTACTCGCAGAACGAGTCGTCGTCTAAGAGTGACCCTGAAGCTGGA
ATAGGGTCTTCCACCCGAATTAATATGAATGATTTGGACATCACAGGAAAGGATAAAGCTGGTGCTGCTATCTTG
ACAGTGTTGGTACTCGGCGTTTTCATTGGAGGTCTGGTATGGATAGTGCTCTAG

AA

MWILWLLTTVSALSLDVDSKDSICEATTHIINGMMDYYEGTRYGGTVGMFQTPYYWWQAGEAFGGMIENWFMCDN
NTYEEIIYDALLHQTGDNFNYIPANQSTTEGNDDQGFWGFAAMGAAERNFTDPPEDYPSWVALTQAVYNTMWARW
DNATCSGGLRWQIFTWNSGYDYKNTISNGCLLHIAARLARFTQNETYGETAEQVWQWLDDVGFINDQGGTYLIYD
GANIVSGECTDITTIEWTYNYGVLMAGCAYMYDFTHDEVWLTRTTTLLSSISIFLNDNVIYEQQCQAAMNCNNDQ
RSFKSIFSRCLGYTAKLVPSLTDQIMTILTASAEGAAKSCSGGTDGVTCGLNWNIGAYDNMYGLGEQMSALEVIT
QLLVLDSPAPYSQNESSSKSDPEAGIGSSTRINMNDLLDITGKDKAGAAILTVLVLGVFIGGLVWIVL

Peptide

MWILWLLTTVSAL

1.2 Proteins potentially involved in secretion

SEQ0055

P. pastoris homolog of *S. cerevisiae* **SEC1 (YDR164C)**

Sm-like protein involved in docking and fusion of exocytic vesicles through binding to assembled SNARE complexes at the membrane; localization to sites of secretion (bud neck and bud tip) is dependent on SNARE function

S. cerevisiae null mutant: inviable

S. cerevisiae conditional mutant: increased heat sensitivity; abnormal subcellular morphology

Chr4, 0134

5' region (in bold start codon next gene)

ctccatggacagaatttgaaccaatgatagaattggaacctcttcttgcaaactgaattc
tttggggactattgacgtataaagtccaaacagagtcgtcaaagagctcaattctgtatt
tgtcaataccatttcttggggcgatgtctccagcgaccgatctaatacccacagcagagt
gaaagtaagggatcaagttatagtatacaccagtgatgaatcccattccttgtacaacag
ggaattgtataaacgaggaccccgaggtgagattaacattgatagagaactttcgtggag
tatccagatcgatatcaatgttggattcaaactctcttgcactcattacgaaatacttga
ttccaactggagaatagaagaactgtggaggattaccatcagcgaaatgacgttggctga
tgctggtttgggagatcgccaatccattgaatccgtccccagtatttttggaataccaca
atgcatacggattggtgtagattggagttgtctgatcaccgaagaagaatattcccgtta
agaaattattagtctgcaccgcagtcaaaaggttctcttaagttacgtgatagacctgca

ggaagacctacaggggaaggagctaaagggaacctagttggcgtggaagttgcaactggt
tggaaaagattaccggaggtgtcaga**cat**gtgaattcactgtggcacaagacgatgactg
gagatgaggaagagaaaccgctggagaaaaaagaagaaggaagaggaagtcggctttgta
tctgaatccatccgattagtccctgttgcacagaactataatacgcatgcaacagcgcat
ttgttgttctcctgttattctttgctggtgccttgcagggcaacgttttcctcgaatttt
ctgtatcttatgtaatctttttgaagacccgctcctgtaatcccaatatttaccgctag
aaaagacaaccagggtacttgaagcaaaactatgagcaata

ORF
ATGGCTTCTGATCTGATTAACCTCCAAAGAGATTATTTACTGAAACTAATAGGGAGTGTGGAAACGTCCAATGGC
CTCAAGTGTCTTGTCCTGGACGCCAATTCAGAGAGGTTAGTCAATAGCTTAATCGACAGCAACACTTTGCTGAGG
TATGTTACCACGGTGGAAAGAATAGATAAGAAGCGGAAAATACGGCTCTCAATGGAAGGAGTCTATTTAATTGGA
CCTACAAAGTTTTCAGTTAATTGCCTTCTGGCAGACTTCCAGATCAATCCCACTCGATACAAAAAAGCTCATTTG
TTGTTTTTATCTCCGTTAGCAAGAGAGTTGACGAATCTAATTATGGGAAACAAACAGTTGGAGGCAAACACAATT
ACGCGAAGGACCGTAGATTTTACTTTGCTGCCTTTGGAAAGTCATGTCTTCCTTTCGGATGCTCCCGATTCTCTC
CCAACCTTGTACAATGAGAACTGTTTGGACTTGATTCGATATCAAGCGTCTCGAGCTGTGCAAACTTTAATGAAT
CTTTGCATTATAACTGGAGAGTATCCTTTAGTTAGGTACTATTCTCCCCAGAATCCCATTAATAAGTCCTCGGTA
TTGCCCCGAATGATTGCACAGGAGTTTCAAAGCACACTGGATGATTATTGCAGAATTAAGCAAGATTTCCCTGGT
GATAACCCAAGGCCCCGGTCAATTTTTATCATCACCGATAGAACAATGGACCTGCTGGCACCATTGATGCATGAT
TTTACCTATGAGGCAATGTGTTTTGATCTACTTGAGTTCGCTGAAAATGTTGACGGTGATTATCCCAATACGTAT
CGTTATTCTGTGGAGAATGAAAATGGAGAGTTATTGGATAGAGAAGCTTCATTAAAACCCCCCATTGATGACTAC
TGGGAAGAGTTGAGAAATATGCATATTTTAGATGCCAGCAATCAATTAGATGTGAAATTGAATAAGCTGATAACT
AATAACCCTATGATGGTTGATAGAGATAAAGCCTCAGGAACAAGAGATTTTCTTTTCATTGTTGCTCACTTGCAT
GGATTTGACGAAGAGCGCAGGAAGATAATGTTACACAAGAAATTGACTGAAGAGCTGTTGGTAATAAATAATGAA
CGTCATTTGGCAGAGTGTGCTGATTTTGAGCAAAACTGTGCTGCATTTGGAGTTTCTTATGATGGCGAAAAAATC
AAAGACATGGCATCTTTTTTACTTTCATGGATCTCTTTAGACTATTTTACCACTTCCGATAAGATACGTTTAATC
CTGATTTACGCCATCTACAGAGGTGGACTGATTAGAGCAGATGTATCAAAGCTGGTAAAATTTGCGGGATTAGCT
TCTGCTGAGGAACACGTCATGACATTGTTTGAGAATTTTTCTTTACTTGGATTTCAGCTTTTGAAAGCTCATCCA
AAAGACAAAAGTTTCAAGAAACAGTTTTGGCACAAGATAGATAGTAATGCTGTTCTAAACACATCCAGATATAAA
CCAGCAATCCAGGCTATCGTAGAACTAGCATCAAAGGGAATCCTGGATGAAGCTTCATTTCCTTACATAAAGGAT
AAGCCTTTGGAAGTTAGTGAAACAAACCCTGATTCGGCAACTTCTCTCAAAAATCCAAGGTATAGAGCTGCTTGG
AGCAGAAAGGGTAGCAGTTATAGTCCGCCTAAGCAGAGAATAGTAGTATATTCTGCAGGCGGTATAACATACTCT
GAAATGAAGGCTGGATATGACGCTGGATGTTTATTGAATAAAGATGTTTTTATTGGAAGTGATGAGGTCATCACT
CCGAGGATGTTTGTGAATAATGTTATTGACTTGACGTCTGATAGAGCCTCATTATCGCTATTCTATGATCGAAGA
AGAGCAGCGGGAGAGTCTGCTCCAAAAGTTTTGTTTGAACAGGAGAGTCATCATCGTCCTTCTATAGGTGGTCCT
GTTGATTCTTCTGCTAGTTTGGCGTCCACCACATCGCAATCCCACGAGCCTCCTACAAATGATAAAGAAAAGCAT
CGAAAAAGGGATAAGCTGAAGAAATTTTGGAAATAG

Downstream (in bold start and stop codons previous genes)
aatgtactcttatcgaaa**tta**atattcaatcttgggtcttttcctagcctttctacggtt
atttttaggctgtatcactcctacagccacgggttgcggcggtggtggtaatggcggagg
aggaggtggtggtggtgctgaaaggggaaatgcaaatgacaactggaatgggggtacgaa
ggtttggttaggaccagaattagattggattgtaggattaatagaagaaaaatgctgtga
tgttggtggtggtggtggtgccaacgaatgatggcccggtgtgggcaaatgttgtac
agtagattcagaattctcaggtttctggatgacaggggatattggttcgatttctaagtt
ctgtccctcgttggataatttctgaaacccatatggagttaccgaatccttgaatccccc
ggatcccctaattttctttaattgtaagacgtatcgggaccttgcattatgttgtctagt
gatgtctgaatttccagttttcgcctgtgctaatcggtttgccattctaccaaaattatc
atccttcggtataagatctacccttccctttttctttccattgtgctatttggtcctgctt
atatctttcacttctgaggattctatcatttctaagacgttcatatcgtcgtattgtatc
tgggtctgaggaagtggcagctttgacgagagccttatttctcaacctaactaacttcct
agtttctttttccacgtctattaacaaacctgagtcttcatgggttggtgtttccgctgg
cgttcctgtgctact**cat**attatcatgccataggatgggcttaagagtgatgatcagata
tcaaaggtaggatcaaaaaaaattcgagagatatgagcattctgc**atg**aagatagacctt
taagtaacgattcatgaagtttgtggctctggtttcaggaggaaaagattcatgttttaa
catatggcattgcttgagccagggacatgaactcatagcct

AA
MASDLINLQRDYLLKLIGSVETSNGLKCLVLDANSERLVNSLIDSNTLLRYVTTVERIDKKRKIRLSMEGVYLIG
PTKFSVNCLLADFQINPTRYKKAHLLFLSPLARELTNLIMGNKQLEANTIIRRTVDFTLLPLESHVFLSDAPDSL
PTLYNENCLDLIRYQASRAVQTLMNLCIITGEYPLVRYYSPQNPINKSSVLPRMIAQEFQSTLDDYCRIKQDFPG

DNPRPRSIFIITDRTMDLLAPLMHDFTYEAMCFDLLEFAENVDGDYPNTYRYSVENENGELLDREASLKPPIDDY
WEELRNMHILDASNQLDVKLNKLITNNPMMVDRDKASGTRDFLFIVAHLHGFDEERRKIMLHKKLTEELLVINNE
RHLAECADFEQNCAAFGVSYDGEKIKDMASFLLSWISLDYFTTSDKIRLILIYAIYRGGLIRADVSKLVKFAGLA
SAEEHVMTLFENFSLLGFQLLKAHPKDKSFKKQFWHKIDSNAVLNTSRYKPAIQAIVELASKGILDEASFPYIKD
KPLEVSETNPDSATSLKNPRYRAAWSRKGSSYSPPKQRIVVYSAGGITYSEMKAGYDAGCLLNKDVFIGSDEVIT
PRMFVNNVIDLTSDRASLSLFYDRRRAAGESAPKVLFEQESHHRPSIGGPVDSSASLASTTSQSHEPPTNDKEKH
RKRDKLKKFWK

SEQ0056

P. pastoris homolog of *S. cerevisiae* **SPC3** (**YLR066W**)

Subunit of signal peptidase complex (Spc1p, Spc2p, Spc3p, Sec11p), which catalyzes cleavage of N-terminal signal sequences of proteins targeted to the secretory pathway; homologous to mammalian SPC22/23

S. cerevisiae null mutant: inviable

Chr4, 0874

5' region

catggattggaatctatgcaaggtactacacaaaactcgagacactgctcggaatccgga
tgtatatgaaattcctgtcaatgctttgaatttttgtaatgttgctactcgagctgaaat
gattgcaactccagcatctttgtcatcggaaagttttgatatttatgttgtcagtaacca
ccggttagagcgcttgcattcgaacgtagatatatgtcaaatcttacagcaaaatcagcg
ttcaaatgaaaccgacgaaaaccgaaaatttggtgctatcatgcaaatcacttggacggg
agatttccagttagctgtaggttacgagagtggaattgttgcttgtattcagttgcaaga
gagacatactcaagttgtctggatagacagttctcatctcacacaccctatattagggat
tgtgtcaaatttggatagggtctttacctgttctgcaagcgatagaattgtcgtcaatag
ccttcaaactggaaagactgttgcaatctctaagatcaaacataaaggtatttcaagcct
agatgtaactattaataacagcatttgccttgtctccgtgactacttgggatgggttttc
aaggatttacagatatgatgaagattcgtcagctgctgtggatgaggctctaaaaggtgt
tttcattccctggactaaattgagaagaattccaccaaaagttattcctaacttcaaaag
ctcctccaatcatcaaacgattgatcttcctgatttaaagggtagagtcatcaagttgtc
aaaaactcaaaatggatgtaataatgcactagtggagagaagtgatggtcagagcaaaat
gatccttcgcagttcagagaagcgtgcattaacacgatacgcgtttgtaggttatcaaga
tggtcgtgtagcaacctatattctttgtgattca**tag**ttttttttctacgattgttcggta
tgaacaaaattcaaggtatcatcatatcacgatcgccaggc

ORF

ATGTTCAACATTATCCAACGGATACAGAGTTTGAGCAATTTTTATTTAACGGTTTCCATTCTATTATGTATTGTT
ACAACAGTTGTCTCAATTATTAGTATGTTCTTGGATGAAACGTCCAGTATTCCTGCCCAATTAAGCAATGTTGTA
ATATCAACAAATTTAAAGTATAGCAGATCGTTTGGTTCAGTCGGTGGTAGACCTAAAGAAAACTCCAAGATTTTA
TTTGATCTTGATATGGATCTGGCTCCATTATTCAATTGGAATACTAAACAACTGTTTGTACAATTGGTAGCAGAG
TACCCTACCTCTGTTGCCGATGATGGTGCGAAGGTGACCTATTGGGATAGCATAATTACTGAGAAAAAGTACGCA
AGAGTGCATGTCAATAAGCAGAGGGGAAAATACTCAGTTTGGGACGTGTCGGACTCCTTTCAAGGCCGCAATGCT
ACGGTTAAACTGAAATGGAACTTACAGCCCTATGTCGGCTTTCTATTCTTTGGACAAACTAAGGGAGAGATTGAG
GTGGCCTATCCTGCAACATAA

Downstream

gagcattgcgatgcatccaggtcta**tta**attgtgtctagcttctatatgtacaggtgata
caggctccccatcatcgtttgcagtattcttctttgtctgcaacacatgtggtaccaatc
gttctggtctaactgcagggtttttcttttttgattttgaaatcaagcgacctgttctcct
gcgttatccgttcccattcgtcaagttgcgatatattatcatcatatcgggccatagtca
tgggatataaatgtttaaacctcatgagaacttgagcacgctccttaccttcaacgcgct
gagatataaatttctctatcgcacgcagtgtctgaaagtatgttgctccgctcaagtaat
aggaagtatgcggcctgaggcgcaacatagtttggagagctgacctgtccaactcggatt
gtttttctcttaccaattttctcttagatttcagaaaaagataacctttgatcaaacgac
cttcaataaactctaaggcatcttccaaagtggtgtgatgtatcaaaagcgatctagctg
tttcattcatggcatagttcgaaagcttgaagccgattgattctaaatgacgtaagagag
gaatgatagatttgtagttcttaacctgtatcatatgcctcagcttaacggcccacacag
tagacaacctatttgccacaagataagctgcacgaacattcgaatctgaaaagtccgaat
ctttttcaacggggcggtttaactgtgtcgccaaacttttgacatacagatcgaaaatta

gctcaaaaatttgccaattggcttcttctgcatagattgcaagtagagcttctaagaata
aataactggaaatgctatcaggggggcagcatctttataaactttcagcaaatgtctttgc
ctcacctgttttttctacggaggccaacttttttaataaaaggtaccaccaactgcaaagg
aagatatttaataggctgagacagagtctctttcttctttt

AA
MFNIIQRIQSLSNFYLTVSILLCIVTTVVSIISMFLDETSSIPAQLSNVVISTNLKYSRSFGSVGGRPKENSKIL
FDLDMDLAPLFNWNTKQLFVQLVAEYPTSVADDGAKVTYWDSIITEKKYARVHVNKQRGKYSVWDVSDSFQGRNA
TVKLKWNLQPYVGFLFFGQTKGEIEVAYPAT

SEQ0057
P. pastoris homolog of *S. cerevisiae* **SEC11** (**YIR022W**)
18kDa catalytic subunit of the Signal Peptidase Complex (SPC; Spc1p, Spc2p, Spc3p, and Sec11p) which cleaves the signal sequence of proteins targeted to the endoplasmic reticulum
S. cerevisiae null mutant: inviable
chr1-4_0187

5' region
ctgcgtttctcaaggcatgggcaaccgcttacggaaatcttgctcaaatcctgataaacg
ctgaatgggcagagttccagaagcagaagtggcatgatttcaaagagtttaaagtgacca
aaatcgaaaatgaatgtgacgatgtgaagtctgtttatttcactccagtagaaggagaga
tagcaaaaccttggatgggcaatacgtctgcatcagatggaaattacctggggaaaagt
ttgagaaatctagggagtactctctttcttccaggccaaacaacaatacctacagaatct
ctgtcaggttgttggaaaatggtaaaatttcgacatttgttcacaatcaactgaaggtgg
gagacataatcactgtagcaccacccgctggacaattgttatacgaagaatctcaaaagg
atgctgtttttttttatcggaggtattggtatcactccagtagtatcaatcatggaaactg
cccttgaaaggggtcagagagttacattatttttactccaatagaacctcaaagagtactg
cctttagaggttggttgaaagaactaaagagtaaattcaatttgcaactgacagtgaagg
aattcgtctctgaagaacaagttactgaaggtgtagaccaaatcaattctgctcaattac
agagatcagacatccaaacggtgtcgccggagaatgaagtttatctggttggacctgttc
catacatgcagtttgttagctctgagctcaacaaactgggagtccaaaatattcattccg
agttctttggaccaactgtggttgcttgaacgagaaaacgatatttattatatagtgaat
taatgcttacgatatgcagtgtaagtcactggatcccttctcttggtagctagttagtta
cttcgtacctttacatccaacatcttacaaaaatatttcgcggctacgagtatcattaaa
ccatcattttacgtggatcatccaactataatttcaccacg

ORF
ATGAACATCAGACAACAACTAGTTCAATTGCTCAACCTAGCCATGGTTCTGAGTACCGCTTTTATGTTTTGGAAG
GGGCTGGGTCTGGTTACCAACAGTAACTCTCCTATAGTTGTCGTTCTGTCTGGATCTATGGAGCCCGCTTTTCAA
CGAGGTGATATTCTCTTCTTGTGGAATAGAGACAAATATGTTGATATTGGGGATGTAGTGGTTTACGAGGTTAAA
GGAAAGCCAATTCCAATTGTCCATAGAGTCCTCAGAGAACACAAAGTTACGAATAAAGATCGTAAAGTCAGACAA
TTACTTCTCACCAAGGGAGACAACAACCCTACTGATGACCTGTCCTTATATGCACACAAATCGAATTACTTGGAT
AGAGATGAAGATGTTTTGGGCACTGTCAAAGCTTATCTTCCCAAAGTGGGATACGTAACAATTTTGATTACAGAG
AATAAGTATGCTAAATTGGGACTATTAGGATTAATGGCGTTGAGCACGTTATTGACTCGAGAATGA

downstream
atatgacgattattaaaaagatccatagccaaaagtattgattttctttttgggtcgtga
aacttcaggggtaattttccgtttttagattgcaactctttttccttctccacttggaa
cattccaaagagaatatttgagttcttatcagagacttgataatcgaggtcagcaatctt
tcttttgaatgattcaagatcagatatattttccatttccggctttattaataactgaat
caatatttgttgcagtccataattttcactatccagcaagtcccaaaatacatcaatagg
ggaggatccatttcgacccaaaatttctttaaactctggtaagtctttgatttgctggat
gtactcataatatgttaaatttgcattctgaatctcttttagtttcaaagaaattgcctt
tttaaacgaatctcttgcttgcctgtctttcctgtagttaagcttttgattgagtttgat
agcttggtctaattgctttacatggaggtccatttcttgttcatagcagcccaactgctc
catagatgaaagatgagggaacttttccatcaattgacctgatatatctttccattttga

ttcaatagttaagcttgattttaaatatgtcttcagatcttgcagttcgttagatttaag
ttcattagtttgtctttcgtgttcctctctgagactgctagtgtactcttcaaacaatct
tcttttcaatttctcggatccaagagagtatataggctcgtcaatgattaatctggaaca
ggttttccaccttgtataatattttatctctgggtaattagataagactttgagaaatgt
gtttttaaatgactgctctgcatcctgttgctctttaaagacttcctgctgtttggatac
ttgatatgactggtataggtcatttctcttcagggaatcctccacgttccaatatctgga
atcttgaatgattaatgaaatagctttggtaaatggtaact

AA
MNIRQQLVQLLNLAMVLSTAFMFWKGLGLVTNSNSPIVVVLSGSMEPAFQRGDILFLWNRDKYVDIGDVVVYEVK
GKPIPIVHRVLREHKVTNKDRKVRQLLLTKGDNNPTDDLSLYAHKSNYLDRDEDVLGTVKAYLPKVGYVTILITE
NKYAKLGLLGLMALSTLLTRE

SEQ0058

P. pastoris homolog of *S. cerevisiae* **SPC1** (YJR010C)

Subunit of the signal peptidase complex (SPC), which cleaves the signal sequence from proteins targeted to the endoplasmic reticulum (ER), homolog of the SPC12 subunit of mammalian signal peptidase complex

S. cerevisiae null mutant: viable

chr1-1_0491

5' region (in bold stop codon previous gene)

ggtagtgatgtccatttctacacagatggatgctgtggatcattcaaacttgagtcaccg
atggttcttgggcacgagtctgcgggaattgtcgttgaagttggctccgaggttaagtcg
cttagggtgggtgacaaagttgcttgtgaaccgggtatcccgtctcgttacagcaatgcc
tataaatcaggacactacaatttgtgtccagagatggcgtttgctgctactccacctatt
gatggtaccctatgtcgctacttttttacttcctgaagatttctgcgtcaagcttcctgaa
catgtgtcattagaggaaggtgcattagtagaacctttgagtgttgctgtccatgctgca
agacttgcgaagattacctttggagacagtgtagtagtttttggagctggncctgttggt
ctgctggttgctgctacggctagagcttacggcgcaaccaacgttctcatcgtggatatt
ttcgatgataaactgacactcgcaaaggacaccttacaagtggccacgcatagtttcaac
tcaaagaatggtatggataatcttttggaatcatttgaaggaaagcatccaaacgtttcc
attgattgtactggagttgaatcgtgtattgcggcaggtatcaatgcactggctccaagg
ggagtgcacgttcaagtcggaatgggaaaatccgaatataacaacttcccattgggactt
atatgtgagaaggaatgtatcgtaaagggtgttttcagatactgttacaacgattacaac
ttagcagttgaactgatagcttcaggaaaagtcgaagtgaaaggattagtaacccacagg
ttcaaatttaccgaagcagtagatgcctatgatactgttaggcaaggtaaggctatcaag
gctatcattgacggcccagag**taa**acggagacatatataataataaaattgtaaggttct
cttagcattgcaccccaccctaggtttctttttagactgtt

ORF
ATGCACTCGAAATTTAGGTGGGTATGTGTCGATACTCAATTCTGCACACACCACCAAAATCTGTCGCCTTTCTCT
TATATCTCCAACCCGAGTCCAATGTCATTTTCTTACCTTGAAGGCAACATCGATTTTAAAGGACAGGAACTTGCA
AACAGGATCACTAAAAAACTAATCACATTTGGTGCAATTATTAGTTTTCTGGTAGGATTTTTGAGTGACAACATC
TTATACACTGTATACACTTTCGCAGCTTTTGGTTTATTGACTGCTTCTTTGGTTATTCCCCCTTTTAGCTTCTAC
AAAAAGAACCCTGTAACATGGTTACCAAAGAAATCCAAAATAGAGATTCAGCATTGA

downstream
ccgaatcagttccgaatgacctttaccttgggtttgatctaagtacccaacaattgaaaa
tcacctctttcgaagggcgttcattgacacatttcaagacttatagagttgatttcgatg
aggagttgagtgtttatggaatcaataatggagtttatgtcaatgaagagactggagaga
tcaacgctccagtggcaatgtgggtcgaagcgttagatctgatcttttctaagatgcaga
aggacaagtttccatttggcatcgtgaaagggatgtctggatcttgtcaacagcatggct
cagtttattggtctaaagatgccccagatctcttatcatctctctctccatcgaaagatt
tgaagtcccaactgtgtccgaaggctttcacattcgaaaagtctcccaattggcaagatc
actcaaccggagaagagctggaaatatttgaaagaaaagcaggatctcctgagaatctgt
ctaagattacaggatccagggctcattatcgatttacgggatctcagattagaaaattgg
ccaaacgcgtaaatccagaactttataaggaaacatacagaatttctttaatcagtagtt

ttctaagttctctactatgcggaagaataactaaaattgaagagagtgatggatgtggaa
tgaatatctacgatattcagaactcgaggtatgatgaagacttgttagctgttacagcag
ctgttgatccagagatcgacggggccactgaacatgaacgtcaagaaggcgtagctaggc
tgaaagataaactgcaagatttggaaccagtcgggtatcgttcaattggcactatagctg
catattttgtggaaaaatatggtttcagtgaagatagcaaagtattctcctttacaggtg
acaacttggctaccattctttccttaccactgcataacgatgacattcttgtttctttgg
gtacatctacaacagtccttttggttacagagacttattgg

AA
MHSKFRWVCVDTQFCTHHQNLSPFSYISNPSPMSFSYLEGNIDFKGQELANRITKKLITFGAIISFLVGFLSDNI
LYTVYTFAAFGLLTASLVIPPFSFYKKNPVTWLPKKSKIEIQH

SEQ0059

P. pastoris homolog of *S. cerevisiae* **SPC2** (**YML055W**)

Subunit of signal peptidase complex (Spc1p, Spc2p, Spc3p, Sec11p), which catalyzes cleavage of N-terminal signal sequences of proteins targeted to the secretory pathway; homologous to mammalian SPC25

S. cerevisiae null mutant: viable

S. cerevisiae overexpression mutant: abnormal cell cycle progression; abnormal nuclear morphology

chr2-1_0589

5' region (in bold start codon next gene)

cagctctaggaacctcaacaaacatgaaaaacaaagaacatgtccacatttgatcatcct
cggagcaattggttcttcattcaagcaaataggacattgagcatcgtttttggagatgat
gactcgtaaaatatctctttgcttcaaaggaacattaggatctaaaacttgagctctgaa
atctcccgtagaattcactataaatttgaaattcacattgatatattccatattgcttag
atgaaccgaatcggatcttctccttggcctaggtagtcgtttgttaagtggtctgttatt
gctggaataatcattaggtagtgtgtactcgactaaatggcttatattaaccaagttctt
cctgttgctgggaaactttccaacatgatcatctaataaaagatcatcagctgatacctt
agagtgttttgaccgacctttcctggacatcctgatggtttggacggtcccctgatttat
ccagtacgcagagtagagatggagaaaggagacacatgcgtc**cat**attggttggtgcact
ttgttatcgggaactgtacagattaggatggtcaaagttcagaaggttaagaatagtaag
cagtccaaacaagagaagacgaggactcttggttttgtggtacttgaccagaaaatgact
tctagagctgtttgagacgtttaggtgtaaattacatacgatttgcccaagacttgagca
ttatcgggcttacagcgcagtacacagttaatggacagaagagctaaaaaagcagcaaaa
agagctacactagatggtagttctactgaacctcgttctcctcatgtcttctaacccttc
atcaccagcgctgaacgttatatacacacttactggatgaatgatgaaactgtcatttt
tgtcacgtgatcatctgaatatgtgatttcctaactcacgcttcttgatcctctaacctg
cttcttctcaacacatcatcagtcacggacagagtttaaaa

ORF
ATGTCAGGAATTACTAAACCTGTCAACCTGTACTCGGTTGCCGAACTCAGAAATGCCACTGATGATGCTCTACCA
TTGGCGCTATCCAAACTGGGATACGAACAGAGTTTTTCACTGATTGATACGAAACTGGCTATCGGTTACGTTGCC
ACAATTTTGGCTGGATCACTCTATTATTTGGAGAAAAAATATAACAACGATTTCTCCAACTTGACCTACTACTAC
TCAATGGTTGCACTGGTTGTGGGATATTTCGCTTTGAATGGATTGTTGTGGCTACACGGAAAATATAGGGAGAAA
GACATTAAATACGTTGGAATCAACAAGAATAACCAAAAGAAGATTTCAATTGAGGCCCGGGTTGAAAATAATACC
ACGCCAATCTACAAAGTCACTGTGATTGAAAATAATAGATTACTTGGGAAGCAGGATATTCCGTTCACAGGCATA
TTTGACGAGGATGGGTTTATTCATATTGATCAGTTGGTTGAGATCTTTACTAGACTATTACAAGAGAAAGAGAAA
TGA

downstream (in bold stop codon previous gene)

aa**tta**ttgtattatctgattgtccatgccactgaactgtctcgcccattcgacaatagta
aaaggcagtatacgtagaatgggaatggcccatgtatacataggccacatgaatgtgcct
tctcgagtgtataaacaggcatctactattctagtagcaagcttttcagcctctaccact
ggagccagaaagttatgtgtaacagaaacatccgaaaacatggctgtatctaattgtcca
ggtaagaatgttgcaaaactgacgtcttgtgagcgaacttcatgtgataaactttcaata
aaggctagaagtgcactcttggtggcagaatacatggataaattcctaggacccaccaat
cccagtatagaacttatagttatgatatgaaggtttttccgtggtatatttggattcata

tagttactgctgtgcattttgatcatctttctcatcaatccaacatgtgacaaaaagttg
acattaaatccatgcaggaaatgttcttgactggaattcataaaagcacctgaacctctt
attgcagcattatttatcagtatatccaactggcctgcgtctcgtattatcttatcaaaa
cttatgctcaactcatgagtatctgctatgttacatttatggtattccaatttttcatgg
tgaaacttgggcttttcgacgtctaagatgtacaccttatcggatactgggactaatttt
ttgactacttccattcctaaaccttgcgagccaccggttattagcactttgatcttatcc
ctcttgaccttctttcttccactagcccatataatcagccatattatatctactgtcaaa
cagaaacatgctggtaatccaagcagccagaacacccaggtcagcgcaggatgtatcagt
gtcaattgagtccaatatatgacaagatcaatactggaccgacttgaactcaccagtggt
tctggaacaccgagcattgacgaatatgcttttcgggccta

AA
MSGITKPVNLYSVAELRNATDDALPLALSKLGYEQSFSLIDTKLAIGYVATILAGSLYYLEKKYNNDFSNLTYYY
SMVALVVGYFALNGLLWLHGKYREKDIKYVGINKNNQKKISIEARVENNTTPIYKVTVIENNRLLGKQDIPFTGI
FDEDGFIHIDQLVEIFTRLLQEKEK

2 *P. pastoris* homologues of genes involved in the glycolysis pathway

SEQ0060

P. pastoris homolog of *S. cerevisiae* **PGM1 (YKL127W)**

catalyzes the conversion from glucose-1-phosphate to glucose-6-phosphate, which is a key step in hexose metabolism; functions as the acceptor for a Glc-phosphotransferase

chromosome 1-4, 0264

E.C.5.4.2.2

5' region

catctatccaggtacatcgtagatgatcatttggatttttcaccacgtcactcacttggt
attaatagttgaactcatattgaagacgttcacaagttctaatataatctatgagatcta
aactagttgataaagtaactttggattcttgtggtgatcttatgaaaccattgcacagcg
actactagggatttgcatcgaataacagtggatgttgacgacaaagagaaatagacggag
cagatttatttattattccagttgaaaagcttatttaaatgattcaagccttgaccaca
aatcaaatcattcaagtcatatcagaaagtcccttctgatgctggaaagcagcctttgaa
ggacatttggcaactgatgtgatactctacctgcctaagagctctgtataaggattatag
gctcctcctcatggggaaaaccaaggctgaacttgcttatctcctcacattcagaggtta
ctaaaggcgtcaagaatccatctgaaaacaatcgggggcatcttgcaccagttggcgaac
caggggaccgataactttcagattaaatgtcaggaaataccgcatgctgagatctgtgga
aatatgcaagtcaagcgatatgcaggtgatgagatcgagaaagtgaataatcttaaatct
gtttgccttgccaacacagccaatccttgaaccaaatgaaccaagtgagggcatgaaatt
cctcagtaagcaagaattagcaggatttctgtttcttccctctcaaacgcctaatcttga
gagcttggactaacaacgatcaagtaccgaaaaactactctaaacactgactcccactag
tgacagaatcaatgaggagtatgggcattgttcttgatttataaatcattctttccaaat
ctctgtgatgctactaaaaccaactatatggaaatagaaagtatattcagtcatttttgt
ttgaaaacactgggcattttcattgcctaagctgacaatacacgattcttgttcattgcc
gtgcaaaaaatgattacttcctaacgctaatctgtaatcagtctaagcccattccatgta
gacgtttctctttatggttggagttcagaaaacaaccgattcaagccatacgttttgcga
acaatcaggatttactccacagcaatccttcaaatctcatgcagctttgcaaaccactt
gaagtaacaacctccaaaattccaccaaacagacaatctagttttaatagttctaaaaca
tcatgtggatctcgagatctgactattattcacgcaattacctaatgacggctcatgaag
tcgcccccttccattcaagaggggataggaagtagccgtaccaacaaacttagtctgtagc
atgcggtctattcattctcaagttaaagacaacgtaaaagcatgcttagcacccttttaa
tctctttgtcacctggaagagcagcagaaatgggggggtctccatctgcaaagtacagaac
aatcacctggttgcttcctttcgcgggagcactccacctttggggatgtcttattgtctc
atctcttctaccccaaagttacctacctacccagcatacaaaactgttaagagtgctca

ORF

ATGTCATTTAGCCCTAGAATTGTTCCAACCACTGCCTTCTCGGATCAGAAGCCAGGTACTTCCGGT
CTCAGAAAGAAGGTGACGGTCTTTCAGCAGCCAAATTACACCGAGAATTTCCTCCAAGCAATCTTT
GATTCCATCCCCGAGGGTGCGCAGGATTCGGTCCTTGTAATTGGTGGCGATGGAAGATATTACAAC
GATACAGTCGTTCAATTGATTGCCAAGATTGGCTTGGCCAACGGTGTCAAACGTATCATTGTAGGA
CAGAATGGTATTCTATCCACTCCGGCTACTTCCCATATTATCAGGACGTACAAGGATGTCAAACCT
ACAGGTGGAATCATCCTGACCGCTTCTCATAATCCTGGTGGACCAACTAATGATCTGGGAATCAAA
TACAATCTCAGTAATGGTGGTCCGGCACCAGAAACTGTTACTGACAAAATGTTTGCCAAATCATTA
GCCTTGACGGAATACAAGATTATTGAGAATCTACCCACCATTGACCTTTCTAAACTGGGACTCAGC
AAATGTGGACCATTGGAAGTTGATATTATTCATTCAACAGAGGCTTATGTTGAAATGCTAAAGGAT
ATTTTTGATTTCCCACTGATCAAGTCCTTCATCAAACGTAGATCCCCAGAGGGGTTTAAGGTTCTG
TTTGATGCATTGAATGGTGTAACTGGTCCTTATGGAAAGTCGATCTTCATTGATGAATTAGGACTG
CCGGAGTCTTCAATTCAAAATTGTGTTCCCAAGGCTGATTTTGGTGGGCTGCACCCGGATCCAAAC
CTTACCTATGCAAAGACCTTAGTTGACAGAGTAGACAAGGAAAACATCGCGTTTGGTGCTGCCTCT
GATGGTGATGGTGACAGAAATATGATTTATGGTGCCTCTACTTTTGTATCTCCTGGTGATTCCGTC
GCAATCATTGCGGAGCACGCTGAATCTATTCCATATTTCAAGAAACTTGGAGTTCACGGCTTAGCA
AGATCTATGCCAACAAGTGGAGCTTTGGATTTGGTTGCCAAAGCAAAGGGCCTCAATGTCTATGAA
GTCCCTACAGGATGGAAGTTTTTCTGTGCTCTTTTTGATGCCAAAAAGCTGTCCATTTGTGGGGAA
GAGAGTTTCGGAACTGGTTCTGACCATATCAGGGAAAAAGACGGTTTATGGGCCATTGTTGCATGG

CTGAATGTCCTTGCTGCATTTGATGCTCAGCATCCAGAAATTGAAGGAGGTGCCACTATTGCATTA
GTCCAGAAAAACTTCTGGGAAACATATGGTAGAACTTTCTTTACCAGATACGATTATGAAGGATGC
GAGTCGATCCCAGCCAACAAATTGATTGAATTCCTACAAGAGAAGGTTGATGATACCTCTTTTGTT
GGTTCAGAGCTAGCTCCAGGCTACACAGTCAAGGAAGCTGCTAACTTCTCATACACCGACTTGGAT
GGTTCTGTCTCATCAAAGCAAGGTTTGTTTGTTAAGTTTACTTCCGGCTTGAGATTTATTGTCAGA
CTGTCTGGTACCGGATCTTCTGGTGCCACAATTCGTTTGTACTTGGAGAAACATACATCTGACAAA
TCTAAGAACTCTTTATCTGCATCAGAGTTTTTAGCCGATGATGTCCGTTTTGTGTTGAACTTTTTG
CAATTCCAGAAGTTTGTAGGTAGAGAGGAGCCTGATGTTCGTACATGA

Downstream (in bold stop codon previous gene)
taattgaaagaaaattctattgttattatacatt**cta**tgtagtccaaagacgggcagtag
catctgctccagatgtgaagacccatgcttctttgggatgccattccatattcaatactc
ctagactctggattacgcgatgaccagttagcttcttcaaaggaactaacaatggattgg
acatcagatcatcataaacggtaccgtggaaaatatgaaccgttccatcgtcggaagctg
aaccgaataatggtagccctcctttgtggaaagcaacagctctgacagctttctggtgat
acctcattgtcttgtatggctctgaggataaatctaaatcatgccatagcactctcttgt
cgtaagaggcagtgataacattgtcacctcttggatgtatatcaatagaagccaaccatc
ttgcctgagatactaatttcttcactaggacctgtttctgtagatcataaattctaatat
atctttgtgatacaacaaatagctgaggcttgaaaggatggaacttagcgtccatgatga
ctcctcttgatttgttgaatggggcctgcgaaatatgtcttgaaagctggtgcaccaaaa
cagatgtttttccagcattaggcgaggaggtgacaaagtagtcacctttacggtgccatg
aaatcttcttgatgacagagaccttcttcatagagacaaccaaggatattcctcgatcag
cttgttctggggttggcttaaaccaggaagtgtgttgtttagtaacggcttcggtggtgt
cttcgtcgtcagatgcgttggatggacccttggcagccttgcgaatatttccgtgtttag
cataaccaaagccatgctcaatcttagacttcgcagcattttcaatatcaaatccaaaaa
taggaggcacgattaaataaacgtgctctcctgcggcaattgctaaaattcctgagtttt
cagctgggttccatgctatgcattcgataaaatcttctccc

AA
MSFSPRIVPTTAFSDQKPGTSGLRKKVTVFQQPNYTENFLQAIFDSIPEGAQDSVLVIGGDGRYYNDTVVQLIAK
IGLANGVKRIIVGQNGILSTPATSHIIRTYKDVKPTGGIILTASHNPGGPTNDLGIKYNLSNGGPAPETVTDKMF
AKSLALTEYKIIENLPTIDLSKLGLSKCGPLEVDIIHSTEAYVEMLKDIFDFPLIKSFIKRRSPEGFKVLFDALN
GVTGPYGKSIFIDELGLPESSIQNCVPKADFGGLHPDPNLTYAKTLVDRVDKENIAFGAASDGDGDRNMIYGAST
FVSPGDSVAIIAEHAESIPYFKKLGVHGLARSMPTSGALDLVAKAKGLNVYEVPTGWKFFCALFDAKKLSICGEE
SFGTGSDHIREKDGLWAIVAWLNVLAAFDAQHPEIEGGATIALVQKNFWETYGRTFFTRYDYEGCESIPANKLIE
FLQEKVDDTSFVGSELAPGYTVKEAANFSYTDLDGSVSSKQGLFVKFTSGLRFIVRLSGTGSSGATIRLYLEKHT
SDKSKNSLSASEFLADDVRFVLNFLQFQKFVGREEPDVRT

SEQ0061
P. pastoris homolog of *S. cerevisiae* **PGI1 (YBR196C)**
Glycolytic enzyme phosphoglucose isomerase, catalyzes the interconversion of glucose-6-phosphate and fructose-6-phosphate; required for cell cycle progression and completion of the gluconeogenic events of sporulation
chromosome 3, 0456
E.C.5.3.1.9

5' region (in bold start and stop codon of previous gene)
agttgagataagaaggatgaatctgtgtagttatccgcataatcctgtttcaaccataat
aacttcttccatggatgacgagccatttgagtggactcggcgcaaaagtgacaacgtttc
tagagggctactcgaaagcactttctctggagattagtaatcgatggtcaggatgggaac
cgaatgtctttggctggcctttccccaaatttgagaacaaaattcaattgctgaggtaat
actttttttccccatacaagaaaggccatcgtgtaattatctcttaccttctactattcga
gaagttgtaaccaa**atg**gctgataacatcgaaaccatccgtcctggaatcaactaccagg
acaaactattgaccgaaatagacattcttgatgatgttaaacaactgtcaacgcagatgg
aacaggaagggaaaggattcgtcttccctaaagatcattactccaagataaaatcactaa
aggggttgcaagtgaagctgttgaaagaaatgcaagagttggctgaaattcagcagtcta
ggcatcaagacgaaactgcatacaggcagaagcttcaagacttggaagatactacatcag
aacttagaaacgttgctttgacg**taa**taggtcacctgtgttactagatgtcggatgtctg
tgattgcagtaacacccggcattcggtcagctgtctacctttcaggatatcctcccccca
ccgagctcggacgcctttctcgcccacccaccacgcaaagttccgcccacctttatacttt
gtggaagtcttttcgccgcgcaccacacttgcccgttcgcagaaaaagaaattccaggta

gcccttattccgtcaggcaataaata*tataa*gcgattgcagacaatctgacgtcccatcc
ccttgtctatttaaacctcctaggttgcttaaatttaaaatctattcgagtcccaacctt
tccattcttccggaataattcaactccaaccaattgataaa

ORF

ATGCCGTCTCTATTGCAAGAGGACAATGCTACTTTCAAGCTCGCATCCGAACTACCAGCTTTCGAA
GAGCTAAAAGAGCTTTATAAGTCAAAGGGAAAGAACTTTTCTGCCAAACAGGCTTTCCAAAAGGAT
CCAGCCAGATCTTCCAAGTTCAGCCACACTTTCAAGAACTTCGACGGGACTGAGGTGTTTTTCGAC
TTTTCCAAGAACTTGATCGATGATGAGATTCTCGCCAAACTGTTCGACTTGGCCAGACAGGCAAAC
GTCGAGAAACTCCGAAACGAGATGTTTGCCGGAGAACATATTAATGTCACAGAGGACAGGGCTGTT
TTCCACGTCGCTCTGAGAAACAGAGCTAACCGCCCAATGTACGTTGACGGCAAGAACGTCGCCCCA
GAAGTTGATAGTGTTTTGCAACATATGAAGGAGTTCTCTACGCAGGTTCGCGATGGTACCTGGAAG
GGATACACTGGTAAGCAGATCACTGATGTGGTCAACATTGGTATCGGAGGCTCTGACTTGGGTCCA
GTCATGGTGACAGAGGCATTGAAGCCTTACGCCCAGGAAGGACTGCATGTTCACTTCGTATCCAAC
GTGGACGGTACCCATATTGCTGAGACTCTAAAATACTTGGATCCTGAGTCTACTCTTTTCTTGATT
GCATCCAAGACTTTCACAACCGCTGAAACCATCCGTAACGCCAATACTGCTAAGGACTGGTTCCTT
TCGAAAACTGGTAACAAAAGTGAGGCAATTGCCAAGCATTTTGCTGCTTTATCCACAAATGCCGAG
GAGGTCGCAAAGTTCGGTATCGACACTAAGAATATGTTCGGTTTTGAAAACTGGGTTGGTGGACGT
TACTCTGTGTGGTCTGCTATCGGTCTTTCAGTTGCCATCTACATTGGTTTTGACAACTTTGAGGAC
TTCTTGAAGGGTGCCGAAGCCGTGGACAGACATTTCCTGGAAACTCCTCTGGAGCAAAACATCCCA
GTTATTGGTGGACTACTCTCCGTTTGGTATACTAACTTCTTTGGAAGTCAGACACATTTGGTCACT
CCATTTGACCAATATATGCACAGATTCCCTGCCTACTTACAACAATTGTCCATGGAATCCAACGGT
AAATCTGTTACCAAGGGCAATGTTTTCGCCAACTACAGCACCGGCCCTGTCGTCTTTGGTGAGCCA
ACAACAAATGCTCAACATTCATTCTTCCAATTGGTGCATCAAGGTACTCATTTGATCCCTGCCGAT
TTCATTTTGGCTGCAAAATCCCACAACCCTGTTGCAAACAACGCTCACCAAATCTTGTTGGCATCT
AACTTCTTGGCTCAAGCCGAGTCTTTATTGCTAGGAAAGACTGAAGAGGAAGTAGCTGCTGCTGGT
GCTACTGGTGGTCTAATTCCACACAAAGTATTTTCAGGTAACAGACCAACTACATCTATTCTGACA
CAGAAAATCACTCCCGCAACCTTAGGTTCTTTGATCGCTTATTATGAGCACGTCACATTCACCGAA
GGAGCTATATGGAACATCAACTCATTTGATCAATGGGGTGTTGAGCTAGGAAAGGTTCTAGCCAAG
GCTGTCCAGAAAGATCTGCAGGATGACAGTGCCAACGTTGAAGAAAGCCACGACTCATCCACTGCT
CAATTGATCAAGAAGTTCAAAGCTTGGGCTTAA

Downstream (in bold start codon next gene)

ggcgttctagttatagagatgtatattgtaatatggtatatgagtactaaagaagtgatt
atctaataaaaatttgagtaatgcactgacattttcactatcagttttggaagagggtat
gggacgcgttcattcacacgtctttggtaatt**atg**gccgaagaagttgaccccgttagtc
tgggtgaccagacgattctccgctataaaatatggaagaagaactctccatacttgtatg
attatttccaaagcaagtctctgctgtggccctctttatctgttgagtttttgccagaca
ttgaacgaaatgacgaagatgagttcgattaccaaaggcttatttttggaacatttacgt
cgggagccagcaatgagtttttgaactttgggatgtttagtagacacaacgaagtctctt
tgagagagtcactgaggaactctctggacaattttgacagcgtcaaaggagaaatatcac
cactagtattaccatcttccaaagactccaaaaactctaatcgcagctgcgaaaagttga
gcatcatccaacgaatagcacataatggagaagttaataaatgcaaatatcttcctcaaa
atcccgacatcatagcgacaattaataattatgggagtgtttcgatttttgatcgaacaa
aacatccttctcaaccactaagcggcacaattaaaccagatatttactgtacatatcata
aggatgaaggttcctgtttgagttggaatcctagcgttgaaggggaactgttgtcaggct
caatggacggaacggttgtattatgggatatcaaaaagtacacgagggacaaagattctc
ttgatccatacaagatattcattgctcatgacaatggctgcaatgaccttaaattcatcc
ctagacacacatcaatttttggttctgttggagaagatggctttttttaaactttgggata
ccagacagggactggatcccgttaagtcaacacgacttcat

AA

MPSLLQEDNATFKLASELPAFEELKELYKSKGKNFSAKQAFQKDPARSSKFSHTFKNFDGTEVFFDFSKNLIDDE
ILAKLFDLARQANVEKLRNEMFAGEHINVTEDRAVFHVALRNRANRPMYVDGKNVAPEVDSVLQHMKEFSTQVRD
GTWKGYIGKQITDVVNIGIGGSDLGPVMVTEALKPYAQEGLHVHFVSNVDGTHIAETLKYLDPESTLFLIASKTF
TTAETIRNANTAKDWFLSKTGNKSEAIAKHFAALSTNAEEVAKFGIDTKNMFGFENWVGGRYSVWSAIGLSVAIY
IGFDNFEDFLKGAEAVDRHFLETPLEQNIPVIGGLLSVWYTNFFGSQTHLVTPFDQYMHRFPAYLQQLSMESNGK
SVTKGNVFANYSTGPVVFGEPTTNAQHSFFQLVHQGTHLIPADFILAAKSHNPVANNAHQILLASNFLAQAESLL
LGKTEEEVAAAGATGGLIPHKVFSGNRPTTSILTQKITPATLGSLIAYYEHVTFTEGAIWNINSFDQWGVELGKV
LAKAVQKDLQDDSANVEESHDSSTAQLIKKFKAWA

SEQ0062

P. pastoris homolog of *S. cerevisiae* **PFK1 (YGR240C)**

Alpha subunit of heterooctameric phosphofructokinase involved in glycolysis, indispensable for anaerobic growth, activated by fructose-2,6-bisphosphate and AMP, mutation inhibits glucose induction of cell cycle-related genes
chromosome 2-1, 0402
E.C.2.7.1.11

5' Region (in bold start codon previous gene)

gctatgttcattcttcacgttattatcggtttggtcaacctggactgtggtattttcagt
agggtcgttatggccattttcattgacattttcattagtcaacttgtcgccgccatcatc
gtcgtcctgatcctggtatgcctggtcaaagatagcatcctcgttataatttacctttg
ctttttacgcctgggaggtaaaatctcatcttcttcgttggaatgtttctcctcggtgag
aatatcactcgtttgacgacgatgcctagctctaactctgggggacgattccacggaatt
gttttgggacggcgtccttctgcgaaccat**cta**ccttctgaatgtatcaaaagtactcaa
ttgaagaccagtaattcagatatgatactttgaggacaaattaatccgactgtgctgtct
tcagtggttcgaagatctagttactaggcaccaaagtgaaaggtgatgaaagagacgtac
gaaaatggtaactgatagaccatacgcgaagtcagcgcgctttttcattagccttgaaat
ggtcgtctgggaaccgctcgggacaccgtctcacgtactaagatcgaggctgacagagtg
aactacaacttttagcttttagaacgttgaaaaatcatcgattttggaaaagattcagct
gctcttgagcaattaataaaaactagataatcgcaagcatatattatgctttaaacgcac
cagttaggcccgaatttgttcctgaatttctctgctccggcgatgagctattatcgtaag
gtgattttcttagtctcgaaggtgtcaaatccataaaaaagagcgatgcagccagggaga
agatctttcttctcttgc*tataa*atcaaacattattcatacagagtttaatcgattcaac
aaagcataacattgattgcaattggttccgtcttgaactgctaaggagaacaatcataaa
ttagtattttgtttgcttgttagactcaaatcgaattacag

ORF

ATGCCAGAACCATCTATAAGTGCACTTTCCTTCACTTCGTTTGTCACTAATGATGACAAACTGTTTGAAGAGACT
TTCAATTTTTACACGAAGTTGGGCTTCCACGCAACACGCTCATATGTTAAAGACAACCGGTCAGACTTTGAATTG
ACGGGGATTTCCACGGATTCAATCAAGGAAATCTGGCTGGAAAGTTTCCCACTATCTGAAGTGGTCGAAACGTCA
GCTGGTAGAGAGTTGAGAAAACCACTGCAAGAATCTGTGGGCTACCAATCTGAAGCTCTTCTGGGATATTCTCCC
TACCAGAGTGACGGTGTTGTTATAAAATTAAGGTTATCAAATCATGACCTTCAGAAAAACAAAGACTTGCCCGGT
GAAGTTACGTTTTTCACCGCTAGTATCGACAAATTAAGAGCTAAACTCATTGAAATTGGTGCTGAGATAATTCCC
TCAGAAATAGACCTTGTTGAATTTTCAACCAAGGATCCTATGGGCGACGTCATTAGCTTTTCTTCTTATCCCTCT
TTGAGTTCCAAGAAGATTACCTCTCCAGACTTTTTCCTCCACCCTAAGAAGGAAGTACGCTCCCAAGAATCAATA
GTTGAGCAGGTTAAATCTGAAGAAGGTAAGAAGAAGATTGCCATCATAACTTCAGGTGGAGACGCACCGGGAATG
AATGCTGCAGTAAGGGCTGTGACAAGAGCCGGTATTTTCTATGGCTGTAAAGTTTACGCTTGTTATGAAGGTTAC
ACTGGACTGGTTAAGGGTGGTGATATGTTAAAGGAACTGCAGTGGCAAGATGTCCGTGGTTTACTTTCCATTGGT
GGTACCATAATTGGTACTGCAAGAAGTAAGGAATTCAGAGAACGATGGGGCCGTCTTCAAGCTTGCTACAATATG
GTCAGCAATGGTATTGATGCGTTAGTTGTTTGTGGAGGTGACGGATCTCTTACAGGTGCCGATCTATTTCGAAAT
GAATGGCCTGAACTGATAAAGGAACTTTGGGTGAGGGCAAAATTACAAAAGAACAATATGAAACACACAGAAAC
TTGACAATCGTAGGTCTCGTTGGTTCTATCGATAACGATATGTGCGGAACTGATTCCACAATTGGTGCTTATTCC
TCATTGGAGAGAATCATAGAGCTGGTAGACTACATCGATGCTACTGCCGCCTCCCATTCACGAGCCTTCGTGGTG
GAAGTCATGGGTAGACATTGTGGATGGTTAGGTTTAATGTCCGGAATTGCTACTGGAGCTGATTACATTTTCATC
CCTGAAAGACCTCCAAGTGAAACAAACTGGAAGGACGACTTGAAGAAAGTCTGTTTGAGACATAGAGAGAAAGGA
CGCAGGAAGACCACCGTTATTGTTGCTGAAGGTGCTATTGATGACCAACTGAACCCTATCACTTCTGAAGAGGTG
AAAGATGTACTAGTGGAGATTGGTTTGGACACTCGTATTACCCGTCTAGGACATGTCCAAAGAGGTGGAGCTCCG
TGTGCTTTTGATAGATTCTTGGCCACTGTTCAAGGTGTTGATGCTGTTAGGGCTGTTTTAGAAAGCACCCCAGCA
ATTCCTTCTCCTGTCATCAGCATTTTGGAGAACAAAATTGTTCGCCAGCCGTTGGTGGAATCTGTTGCTCAAACA
AAGACTGTCAGTGATGCTATCGAGGCCAAGGATTTCGATAAAGCTTTGAAATTAAGAGACCAAGAGTTTGCCACA
TCATATGAGAGCTTCCTGTCCGTTTCCAAGTATGACGATGGATCATATCTAGTACCAGAGAGCTCAAGATTAAAT
ATTGCCATCATCCATGTGGGAGCTCCAACATCTGCGTTGAATCCTGCCACAAGAGTTGCTACTTTGAACTCGTTG
GCAAAAGGACACAGAGTTTTTGCTATTCGAAACGGATTTGCAGGATTAATTCGCCACGGCGCTGTACGAGAGCTC
AACTGGATTGATGTTGAGGACTGGCACAACACAGGTGGGTCGGAGATTGGTACCAACAGAAGTCTTCCTAGTGAT
GATATGGGCACTGTGGCGTACTACTTCCAGCAATACAAGTTTGATGGTCTTATTATTATCGGTGGATTTGAAGCT
TTCACAGCTCTGTACCAGCTGGACGCAGCTCGTGCTCAGTATCCTATCTTCAATATTCCAATGTGTTGTCTTCCA
GCTACTGTTTCTAATAACGTTCCTGGTACCGAGTATTCCTTAGGGTCTGACACATGTCTAAACACCTTGTCTGGA
TACTGTGATGCTGTGAAACAATCTGCTTCTGCTAGTAGAAGAAGAACATTTGTTGTGGAAGTTCAAGGTGGATAC
TCAGGATATCTTGCCAGCTACGCTGGTCTGATCACAGGAGCTTTGGCTGTTTATACTCCTGAAAACCCAATCAAC
CTTCAAACAGTGCAGGAAGACATTGAATTGTTGACTCGAACATACGAGGAAGACGATGGTAAGAACAGATCGGGT

AAAATCTTTATTCATAATGAAAAGGCTTCAAAGGTTTACACCACGGATCTGATTGCTGCTATCATAGGTGAAGCT
GGAAAGGGTAGGTTTGAGAGCCGTACTGCCGTGCCTGGTCATGTACAACAGGGTAAATCTCCCTCATCTATTGAC
CGGGTTAATGCCTGCAGACTGGCTATCAAATGTTGTAACTTCATCGAGGACGCCAATTTCCAGGTGAAACACAAT
GCCAATTTGAGCGCCGACGAACGTCATTTGAGATTCTTTTACGATGACGGAGTTAAGACATCTGCAGTGAGCGGC
AAATCTTCCGTGATAGATGATAACACGTCAGTGGTCATTGGAATCCAAGGTTCCGAGGTTACATTCACTCCTGTA
AAACAGCTATGGGAGAAGGAAACTCATCATAAGTGGCGAAAGGGTAAGAACGTTCATTGGGAGCAGTTGAACATT
GTCTCTGACCTCTTGAGTGGTCGTTTGTCTATTCGTACCACGTAA

Downstream (in bold start and stop codon next gene)
aagacggatcaaatcggttgtttgggtactaaagacaatccatttttttctttctctcg
agctggatgaaactagtgcatgtacgaatccgtgtgtaatctactgggatgcttatttta
cgcatttttgttgataaaaatagagacctactactactcctgatttcaagcctttctacc
tgtaagtttctttttttttgctggtgacaatagccttttttttacctttttgccatc
gttgcgtcctgtatagcttcttaaatgtctccgcaacaatttttatgggagcaatagatt
acctctggttttactcgtaacctctagggttaataggcccattaaattggctatgggaat
gcgagtgat**atg**gcttcaactatagagagtgcggattattgttcattttctactgtatgc
aagatgtggcatttttcttattgtggtgaaaggtaccgctgcccctcaattggttgcata
ctgaaacgacaagttgagtcgagtccctggaaaatcctaccagttgcaggcaaaataagt
caaccacgccgatcatcacagcgaaaagagttgcatttggttttgagagacggtgttttg
gtaaacagcaagccccaatgcaaacgatcataccgatctgccatggcagcctgttggtca
catgaaacacacaggagcacaggggaagttttattatcatgcgcggtgcaa**taa**atgtgc
atgagttacgcatttcggatatctgctccttatataaaagtgtttacagagaccatagta
tgatgcgcgtatagtctctgtatctaccttgcttttgttctgcacaacaccacatttttc
ttggaccacttctagcataatccgtctgtgaaaaatgcacaccgcacttatctttcacgt
atttgaagaaaagccttttttcatcgatcgaatgatattgcacttggacgcacgaacacac
ccattaccaataagaatgttcaaaaaaaaacaaccatcatg

AA
MPEPSISALSFTSFVTNDDKLFEETFNFYTKLGFHATRSYVKDNRSDFELTGISTDSIKEIWLESFPLSEVVETS
AGRELRKPLQESVGYQSEALLGYSPYQSDGVVIKLRLSNHDLQKNKDLPGEVTFFTASIDKLRAKLIEIGAEIIP
SEIDLVEFSTKDPMGDVISFSSYPSLSSKKITSPDFFLHPKKEVRSQESIVEQVKSEEGKKKIAIITSGGDAPGM
NAAVRAVTRAGIFYGCKVYACYEGYTGLVKGGDMLKELQWQDVRGLLSIGGTIIGTARSKEFRERWGRLQACYNM
VSNGIDALVVCGGDGSLTGADLFRNEWPELIKELLGEGKITKEQYETHRNLTIVGLVGSIDNDMCGTDSTIGAYS
SLERIIELVDYIDATAASHSRAFVVEVMGRHCGWLGLMSGIATGADYIFIPERPPSETNWKDDLKKVCLRHREKG
RRKTTVIVAEGAIDDQLNPITSEEVKDVLVEIGLDTRITRLGHVQRGGAPCAFDRFLATVQGVDAVRAVLESTPA
IPSPVISILENKIVRQPLVESVAQTKTVSDAIEAKDFDKALKLRDQEFATSYESFLSVSKYDDGSYLVPESSRLN
IAIIHVGAPTSALNPATRVATLNSLAKGHRVFAIRNGFAGLIRHGAVRELNWIDVEDWHNTGGSEIGTNRSLPSD
DMGTVAYYFQQYKFDGLIIIGGFEAFTALYQLDAARAQYPIFNIPMCCLPATVSNNVPGTEYSLGSDTCLNTLSG
YCDAVKQSASASRRRTFVVEVQGGYSGYLASYAGLITGALAVYTPENPINLQTVQEDIELLTRTYEEDDGKNRSG
KIFIHNEKASKVYTTDLIAAIIGEAGKGRFESRTAVPGHVQQGKSPSSIDRVNACRLAIKCCNFIEDANFQVKHN
ANLSADERHLRFFYDDGVKTSAVSGKSSVIDDNISVVIGIQGSEVTFTPVKQLWEKETHHKWRKGKNVHWEQLNI
VSDLLSGRLSIRTT

SEQ0063

P. pastoris homolog of *S. cerevisiae* **PFK2 (YMR205C)**

Beta subunit of heterooctameric phosphofructokinase involved in glycolysis, indispensable for anaerobic growth, activated by fructose-2,6-bisphosphate and AMP, mutation inhibits glucose induction of cell cycle-related genes
chromosome 1-4, 0047
E.C.2.7.1.11

5' region (in bold stop codon previous gene)
attctagcaccactcaagttacaaacgattattatgatcaacatctaatgctaaaggatt
tctttgaaagaaaagtgattcgaaattcacaattcatccaaagcgtatcgtatgccccta
aagtgttccattctattcgaatgtttcaaagttcagatgaaatgtatatcaatcagtatc
tccaacaattcaaaagatgtgatttacggatctccgatacactgtacttcatcaaagacc
attacatgctgaaaggtattcgagagatcaatttagatcttttcttacaggcaaaattac
atcctgaactgttaggtggccggcttattggcttttatgagataagggcaccctggggaa
aatttgagtttatggctattgaatgtgattgtcagcctcaccaatctatggttaatatgt
tcctgcagcgcgac**taa**tactacgaatttgcgcgcctgcaactctcctctctcagtctgg
ttcttgaagatgtgtataagtctcgcaatgaaacaaatgttctgggagatggatcgttgc

tcaatatgaatatcggattagtaaagtataaacttataagactttcagtggttttcacat
tgacctaggcttcttataacaaggcgactccgagaacaaaagaaaaatagaatggccctc
aaagtagcttatgtgagatctgggcttgattctagatgtggtaatacttgatattttctt
ttttgattattcttccatgctatcaataagtgaaaaaacagtctgtctttctttggcggt
gaatatgtgtccattatgcacaaattccgatgtccgtatcggatatcggacctccataat
ccataatcgaaccataagtgtgcttatgtaatcaatttttttaccctgctgcgaatatctg
acagcttttaattacggctccctccaacttatcaacctttcctagttatattccatttca
agatttacaaaggtaaataacatctcaagcttttgttaaag

ORF (in bold intron)

ATGCCAGACGCTTCACTCTTTAACGGAACTTCGTTTATAACGTTATTTGCGCCCAACATCTCCTTATTCCAGGCC
TCCATTGATTTCTACACAGATCGCTTGGGGTTTGCTATTAAGGAGACCTCAAATCAGAAGTTAGTTTGGTTACAG
CTTGAAGAAGATTCAAATAATGTTTCAATTCAATTGCTTCTAGATCCAGAACATGCTGCTTCCGTGTCTCAGATT
GATCAGAATATTCGCAATTTGACCAGATCGTTATACCGGAAAGACTGGAGATCAATTCAGTCTAATATTGCGTTC
AAGAGCTCTTCACTAAGTAAGCTCGTCAAGTTGCTGAAGGATGGAGGCCATCCTGTACAGCAAAGTCCTAATGAG
ATTTCACCTTTTGAAGTTTACACGTTAGATCCCCTAGGCTCATTGATAGGTTTCAGCGGATTCAAGAATCCATTT
GCTGTCAACGAACGTTCTTTGCTTCCAAAGGTTTCCGAAGAGAAAGCTTACCGTACTGAAGATGACTCAGAAAAG
CTTTTCACTCCCATCCGTAAAACCATCGGAGTTATGACATCCGGAGGTGACTCTCCAGGGATGAACCCTTTCGTT
AGAGCGGTGGTTCGGGCAGGAATCTACAAGGGTTGCAAAGTATTTTGTATTCACGAAGGATATGAGGGACTTGTA
CGTGGGGGTGAAAAATACATCAAGGAGACGCAATGGCACGACGTTCGTGGATGGCTTGTAGAAGGCGGAACCAAT
ATAGGAACTGCTAGGTGCAAAGAATTCAGAGAACGCTCTGGCCGACTTAAAGCTTGCAAGAATATGATTGATATG
GGTATTGATGCGCTAATCGTTTGTGGGGGTGATGGTTCGCTAACAGGTGCGGACCGATTTCGTTCAGAATGGCCA
AGTTTGATTGAAGAGTTATTGCAAACTGAACGAATCTCCCAGCAACAGTTTGAAACTTATCAAAATTTGAATATA
TGCGGAGCCGTCGGCTCTATCGACAATGATATGTCATCAACAGATGCCACGATCGGAGCCTTCTCCTCTTTGGAT
AGAATCTGCCGGGCTATTGACTACATTGATGCTACCGCCAATTCTCATTCTCGAGCCTTCATCGTTGAAGTCATG
GGCCGCCATTGTGGTTGGTTAGGTCTTTTGGCTGGTCTAGCGACAAGTGCAGACTACATCTTGATCCCAGAGAAA
CCTGCATCTTCAAGAGAGTGGCAAGATCAAATGTGTGACATAGTCAGCAAACATAGGGCAAGGGGAAAACGAAAG
ACCATTGTTATTGTTGCTGAAGGTGCTATCAGCAATGATTTGAGTCCAATTTCCTGCGATCAAGTCAAGGACGTT
CTGGTCAATAGATTAGGGCTTGATACCAGAGTTACCACTCTTGGTCACGTACAAAGGGGCGGTACTGCGGTTGCT
TTTGACAGAATATACGCTACTTTGCAAGGGGTCGAAGCCGTTAATGCTGTTTTGGAATGCAATGCGGACACGCCG
TCACCCATGATTGCTATTAAGGAAGATCAAATTACGAGAGTTCCTCTTGTTGACGCTGTGGAGCTTACACAGCAA
GTTGCAAAGTCCATCGAATCTCGTAACTTTAAGAGGGCTATTTCATTGCGAGATTCTGAGTTTGTCGAACACATG
AAGAACTTTATCTCAACTAACTCCGCGGATCATGTTCCTCCTTCTTTACCTCTAGAAAAGCGAAAGAAAGTTGCA
ATCATAAACGTTGGAGCTCCAGCGGGAGGAATGAACAGTGCGGTTTACAGTATGGCCACCTATTGCATGTCCAGA
GGGCATGTTCCATATGCTATCCACAACGGGTTCAGTGGTTTGGCTCGTCATGAATCTGTTAGATCTATAAACTGG
TTGGACATTGAAGGTTGGGGTTCTCTAGGTGGCTCAGAGATAGGAACCAACAGAACTCTTCCAAATGATGCCGAT
ATTGGGATGATTGCCTACTTCTTTGAAAAGTATGGTTTTGATGGCCTGATTCTTGTTGGTGGATTTGAAGCTTTT
ATCTCACTTCATCAATTGGAAAGGGCCCGAATCAATTATCCGTCTCTAAGAATACCCTTGGTTCTTATTCCTGCT
ACCATTTCTAACAATGTGCCTGGAACCGAATATTCCTTGGGTTCAGATACGTGTCTTAACTCTTTTATGGAATAC
TGTGATGTAATCAAGCAATCTGCAGCTGCAACTAGAAACAGAGTGTTTGTTGTTGAGGTCCAGGGTGGAAATTCA
GGATACATTGCAACCCATGCGCAACTCGCTTGTGGTGCTCAGATTTCTTACGTCCCTGAGGAAGGTATTTCGCTG
GCTCAGCTTGAAATGGATATCAATTCTTTGAAAGAGTCCTTTGCCAATGACCAAGGAAAGACAAAATTCAGGCAG
ACTGATTTTGAAGTCGGAGAATGCATCGAAG**gtactcactacggaagtaatctccactatcatagatgatgaagc**
tagtggcagatttgactcaaaaacagCTATTCCGGGGCATGTTCAGCAGGGTGGTATTCCTTCTCCGATGGATCG
TGTGAGGGCCTCTAGATTTGCAATTCGTGCTGTCTCTTTTATTGAGAAGCACTCTGATAAATGTCAAGCTTTCAA
AAATTCGATTAGTTTCCGACAAACAGATGAGATCACTTCCACGGCTGTTGTCCTTGGTATTCACAAGTCCCAACT
GAGATTCACTCCAATTCGACAATTATACGATTTTGAGTCTGATGTACCAAGGCGTATGAGAAAAAATATTTTTTG
GAGTAACGTACGAGAAATCAGTGATATGTTAAGTGGAAGAACCTCGTTATAG

Downstream (in bold start codon next gene)

gcattcttacataaggaacatactgaaagtatggcaatttttgttagcaatctcatgaac
actttttttctatttaatttctatctactttgttacatgagcgctatccaacaataaatgc
cgcaagctacaaaattagttcaggaatataacttgtccatcttgtaacttaaataaatct
tgcttcaacagtaacaacatggttctgtggtctagtggtatgatacctgcttcacacgca
gggggtctccagttcgaacctgggcagaatcatctttttgaagaataagcgatagatgtt
tgtcgccactcaaaggctttattatgaaaggcaaactgaagtccttcagacttcaaggta
aatttttgaatttagtgcgtaatcttagggcgtatgtttgatacttaagtataaacttta
aagtacacttgaagtatctac**atg**accaatgattacataaccacatcgctatgttggtca
aagaaaatattaacaacaaaagctcccatttcaagtaaccactcgatgtctctttcatat
cccgttctcctctccaagttttctgagctgtttcacttgggaggtaaatacccaccagcc
cttgtttccggcctcgttatatttttgatatctttttatatcgttggagttttcacttct

ttgagatcaagtttggcacttgatccttgggctctctatcatttgaatcttaacaagata
tcactatacccttttggttcactcatccttcctccacctttttcttcaacatttttgctttg
atctctccactgtcgttatatgaaagatcgaatggtacagttcatacaggagtggtcctc
aacgttttggcagttgttactgcattgccttactgtgttttaggaatggtatttttcccc
aaagttagtgtagtaggagcatctgcttggtgcttttcattctttggttattattcttat
ttgcagagtctcagttatcccactttcaaagtacaagatta

AA

MPDASLFNGTSFITLFAPNISLFQASIDFYTDRLGFAIKETSNQKLVWLQLEEDSNNVSIQLLLDPEHAASVSQI
DQNIRNLTRSLYRKDWRSIQSNIAFKSSSLSKLVKLLKDGGHPVQQSPNEISPFEVYTLDPLGSLIGFSGFKNPF
AVNERSLLPKVSEEKAYRTEDDSEKLFTPIRKTIGVMTSGGDSPGMNPFVRAVVRAGIYKGCKVFCIHEGYEGLV
RGGEKYIKETQWHDVRGWLVEGGTNIGTARCKEFRERSGRLKACKNMIDMGIDALIVCGGDGSLTGADRFRSEWP
SLIEELLQTERISQQQFETYQNLNICGAVGSIDNDMSSTDATIGAFSSLDRICRAIDYIDATANSHSRAFIVEVM
GRHCGWLGLLAGLATSADYILIPEKPASSREWQDQMCDIVSKHRARGKRKTIVIVAEGAISNDLSPISCDQVKDV
LVNRLGLDTRVTTLGHVQRGGTAVAFDRIYATLQGVEAVNAVLECNADTPSPMIAIKEDQITRVPLVDAVELTQQ
VAKSIESRNFKRAISLRDSEFVEHMKNFISTNSADHVPPSLPLEKRKKVAIINVGAPAGGMNSAVYSMATYCMSR
GHVPYAIHNGFSGLARHESVRSINWLDIEGWGSLGGSEIGTNRTLPNDADIGMIAYFFEKYGFDGLILVGGFEAF
ISLHQLERARINYPSLRIPLVLIPATISNNVPGTEYSLGSDTCLNSFMEYCDVIKQSAAATRNRVFVVEVQGGNS
GYIATHAQLACGAQISYVPEEGISLAQLEMDINSLKESFANDQGKTKFRQTDFEVGECIEAIPGHVQQGGIPSPM
DRVRASRFAIRAVSFIEKHSDKCQAFKNSISFRQTDEITSTAVVLGIHKSQLRFTPIRQLYDFESDVPRRMRKNI
FWSNVREISDMLSGRTSL

SEQ0064

P. pastoris homolog of *S. cerevisiae* **TDH3 (YGR192C)**

Glyceraldehyde-3-phosphate dehydrogenase, isozyme 3, involved in glycolysis and gluconeogenesis; tetramer that catalyzes the reaction of glyceraldehyde-3-phosphate to 1,3 bis-phosphoglycerate; detected in the cytoplasm and cell-wall
chromosome 2-1, 0437
E.C.1.2.1.12

5' region (in bold start and stop codon next gene)

tctgctactctggtcccaagtgaaccacctttttggaccctattgaccggaccttaacttg
ccaaacctaaacgcttaatgcctcagacgttttaatgcctctcaacacctccaaggttgc
tttcttgagcatgcctactaggaactttaacgaactgtggggttgcagacagtttcaggc
gtgtcccgaccaatatggcctactagactctctgaaaaatcacagttttccagtagttcc
gatcaaattaccatcgaaatggtcccataaacggacatttgacatccgttcctgaattat
agtcttccaccgtggatcatggtgttccttttttttcccaaagaatatcagcatcccttaa
ctacgttaggtcagtgatgacaatggaccaaattgttgcaaggtttttcttttttctttca
tcggcacatttcagcctcacatgcgactattatcgatcaatgaaatccatcaagattgaa
atcttaaaattgccccttttcacttgacaggatccttttttgtagaaatgtcttggtgtcc
tcgtccaatcaggtagccatctctgaaatatctggctccgttgcaactccgaacgacctg
ctggcaacgtaaaattctccggggtaaaacttaaatgtggagtaatggaaccagaaacgt
ctcttcccttctctctccttccaccgcccgttaccgtccctaggaaattttactctgctg
gagagcttcttctacggccccccttgcagcaatgctcttcccagcattacgttgcgggtaa
aacggaggtcgtgtacccgacctagcagcccagggatggaaaagtcccggccgtcgctgg
caataatagcgggcggacgcatgtcatgagattattggaaaccaccagaatcgaatataa
aaggcgaacacctttcccaattttggtttctcctgacccaaagactttaaatttaattta
tttgtccctatttcaatcaattgaacaactatcaaaacaca

ORF

ATGGCTATCACTGTCGGTATTAACGGTTTCGGACGTATTGGACGTCTCGTTCTGAGAGTCGCTCTTTCCAGAGCT
GACATCAAGGTTGTTGCTATCAACGACCCATTCATTGCTCCAGAATACGCTGCTTACATGTTCAAGTACGACTCT
ACCCACAAGGCTTACAAGGGTGAGGTTTCTGCCAGCGGCAACAAGATCAACATTGACGGTAACGAAATCACCGTT
TTCCAAGAGAGAGACCCTGTCAACATCCCATGGGGTAAGGCTGGTGTCGACTACGTCATTGAGTCCACCGGTGTT
TTCACCACTTTGGAGGGTGCCCAAAAGCACATCGACGCCGGTGCCAAGAAGGTCGTCATCACTGCTCCATCCAAG
GATGCTCCAATGTTCGTTGTCGGTGTCAACGAGGAGAAATACACTTCTGACTTGAACATTGTCTCCAATGCTTCT
TGTACTACCAACTGTTTGGCTCCATTGGCCAAGGTTGTCAACGACACTTTCGGAATTGAGTCCGGTTTGATGACC
ACCGTCCACTCCATGACCGCCACTCAAAAGACCGTTGACGGTCCATCCCACAAGGACTGGAGAGGTGGTAGAACC
GCTTCTGGTAACATCATTCCATCTTCCACTGGTGCTGCTAAGGCCGTCGGTAAGGTTATTCCAGAATTGAACGGT
AAGCTGACCGGTTTGGCTTTCCGTGTCCCAACCGTCGATGTCTCCGTTGTTGACTTGACCGTCAACTTGAAGAAG
GAGACTACCTACGAGGAGATCAAGTCTGTTATCAAGGCTGCTTCCGAGGGTAAGCTCAAGGGTGTTTTGGGTTAC

ACTGAAGATGCCGTTGTCTCTTCTGACTTCTTGGGTGACGAGAGATCCTCCATCTTCGACGCTTCTGCCGGTATT
CAATTGACTCCATCTTTCGTCAAGCTGATCTCTTGGTACGACAACGAGTACGGTTACTCCACCAGAGTCGTCGAC
TTGTTGCAACACGTTGCTAAGGCTTAA

Downstream (in bold start codon previous gene)
atcgatttgtatgtgaaatagctgaaattcgaaaatttcattatggctgtatctacttta
gcgtattaggcatttgagcattggcttgaacaatgcgggctgtagtgtgtcaccaaagaa
accattcgggttcggatctggaagtcctcatcacgtgatgccgatctcgtgtattttatt
ttcagataacacctgaagacttttgggtcggaggactggctctttccgatcaaattggaa
tggaaaattgctcctctaagaaagggtgccaacactctttgtaacacaggacaccgttta
ttgctaactcgattgcattctttcctttcccacaccgggatctggtcttggtgaacaatc
tctcctgtccttatctaaatatatcatcgcactgtaaccttccttattacttttcgagcg
tccgtcctgtattatcttcaacctgaaaccaaactctaaccaggcttcactcgtggatct
ataattgaac**atg**aaaaacttctcttataccagtcaaccgggaactccgttggactacaa
ctatggtcggtccatgctgccgtcacatttgttgacacccttctagctacaccagttct
tccatcccagccaagcacgccatatattgatcaacacatgagcaatctggatcagaaata
tttaacggagcactaccagttgctgcatgaggtgaatcactctagatccttgttatttga
gaaattgcccaaggagttgtctctgaaagagtttttggacgcgttcaaatcgagtcagtt
cttggagaatgtgaaaataaaagaggatagtggtgaagggacatccgtgttggtgcagtt
cgtcaacgatgagtctgctttagtctttctgcaaaattttgactggagggactttgtttc
ttcaacaggctttgacaatctcaagatctcatggtgtttcaataactcatccgatgacta
tttggtagatgggacttatatggataggattctgaatctga

AA
MAITVGINGFGRIGRLVLRVALSRADIKVVAINDPFIAPEYAAYMFKYDSTHKAYKGEVSASGNKINIDGNEITV
FQERDPVNIPWGKAGVDYVIESTGVFTTLEGAQKHIDAGAKKVVITAPSKDAPMFVVGVNEEKYTSDLNIVSNAS
CTTNCLAPLAKVVNDTFGIESGLMTTVHSMTATQKTVDGPSHKDWRGGRTASGNIIPSSTGAAKAVGKVIPELNG
KLTGLAFRVPTVDVSVVDLTVNLKKETTYEEIKSVIKAASEGKLKGVLGYTEDAVVSSDFLGDERSSIFDASAGI
QLTPSFVKLISWYDNEYGYSTRVVDLLQHVAKA

P. pastoris homolog of *S. cerevisiae* FBA1 (YKL060C)

Fructose 1,6-bisphosphate aldolase, required for glycolysis and gluconeogenesis; catalyzes conversion of fructose 1,6 bisphosphate to glyceraldehyde-3-P and dihydroxyacetone-P; locates to mitochondrial outer surface upon oxidative stress
E.C.4.1.2.13

2 homologs

SEQ0065

chromosome 1-1, 0072

5' region
tgcactttgaccttaaaggggctgatttaaggttatgccggggaagaagaaatagcgcga
tgagcaaagtcgatgcctaaaggagtggttttgctacctcatttaagaagagaataggac
gtgcatccagcgatgcgtgctaggacaaagaaccgcacttggcgggtacaaacctgacgt
catttcctgatattattgacatttgagctgaccaattaaggtgcccatccacaatagcca
cctggatagcggaatgcacccccattgagttgatcaaactaccattttgcttatacctca
agttaatgttgaactaccattcttcacatgctcctcctagatcccctgtccccttttctcc
ccctctttcatcctttaatttgcatttcttgacggtcttctatccctagaaagtttggaa
cgcctgctatatggttaggacacgactgactagctataaaatttttcagaccagactctt
tctcttcttaacgcaaatttaacaggcagacaacaacataggaaagaatcaccatatagg
ttggactctttacagacgtccttggccgttgaccatggtggtacagttgtccaagttcta
caagtttgtctgaagaatgaagttattggtcttgggtgcagctttccatctgttcgattt
attcggctaagagtttaccattgtgtgctcgtatggggaagggtgcaaggatcagtaata
cagtcgaacctggagtatctaccatagtggggatacaatgtagtttatctgttatctcga
ttgttcctaattaaggttttctttgatcctcttctagtccacacctcctagatgacattc
gagctgcctggattggatgcctaggtttattgcctagttcaatacaattcgtgcgggcta
cagtagaaggcccttacataatccggaaagcatggtcccccaccaaattgagagcttttt
cagccttcactggtggtatcattttcgggagataataaggtttcgattgggaattcccac
cagagaacactatagagggaccaagctgatgctagcctgacatccccaaagcacacttcg

taattgaaaaccgttacctctagcacactgtccagactacccccgtcaaaaaaacgctct
ttttctcgactaattgagtcttcaactcatcccgtccttgccgaattacttgaattcatt
tcacacctccgttgcttacgtactctcaccggtctccggtgtacatggatccgctattgc
cagatatttctcatacaacaatcaccagatcaaggtcgtgaacggaccaatggcatccag
agcaatcctgaacagataggggtccgggctgtataaagtgaaataacgtgacttgaacca
gcaactatgtcccagttgtgctacacttaacacgcgattaccccggagctcaccaggcct
cttccccctctcattggaaccctcctagcgcttcgaaataatggctgcgtactatttaac
tggtgccagttcccgctgacaatatccttttttcttctcccccttagttccccacatatcaa
ttgaacatatttttttacaca

ORF

ATGCCAGATCAATTATCATTTTTAAAGAGAAAGACCGGTGTCATCGTTGGTGACGATGTCAGAAAC
CTGTTCCTTCACGCACAGAAGAAGGGCTTCGCCATTCCAGCTATCAACGTCACCTCTTCCTCCACC
GCCGTGGCTGCTTTGGAAGCTGCTAGGGACAACAACTCCCCAATTATTTTGCAAACTTCCCAAGGT
GGTGCCGCCTACTTTGCTGGTAAGGGTGTCAAGAACTCTAACCAGGAGGCTTCCATTGGTGGTGCC
GTTGCTGCTGCTCATTACATTCGTTCCGTAGCTCCATTGTATGGAATTCCAGTTATTTTGCACACT
GACCACTGTGCCAAGAAGCTTTTGCCTTGGTTCGAAGGTTTCTTAGCTGCCGATGAGGCTTACTAC
AAGGAGCACGGTGAACCTTTGTTCTCTTCCCACATGTTGGATCTGTCTGAGGAGTCTGATGACGAG
AACATTGCCACTTGTGTCAAGTACTTCAAGAGAATGGCTAAGATGAACCAATGGTTAGAGATGGAG
ATTGGTATCACCGGTGGTGAGGAAGATGGTGTCAACAACGAGAATGCCGACAAGGAATTGCTGTAC
ACCAAGCCAGAGACCGTCTTTGCTGTTCACAAAGCTCTTGCTCCAATTTCTCCAAACTTCTCCATT
GCTGCTGCCTTCGGTAACGTGCACGGTGTTTACAAGCCAGGTAACGTCGTTCTTCGTCCTTCCTTG
TTGGGTGACCACCAGAAATACGCCAAGGAGCAATTGAAGTCTGAGAGTGACAAACCATTGTTCTTG
GTCTTCCACGGTGGATCTGGATCTACCCAAGAAGAGTTTGACACCGGTATCCGTAACGGTGTCGTC
AAGGTCAACTTGGACACTGACTGTCAGTACGCTTACCTCATTGGTATCCGTGACTACGTCTTGAAG
AAGAAGGACTACATCATGTCCCAAGTCGGAAACCCAGAGGGTGACGACAAGCCAAACAAGAAGTAC
TTCGACCCAAGAGTTTGGGTTCGTGAGGGTGAAAAGACCATGAGTGCTAGAGTCACCGAGGCTTTG
AACATTTTCCACACTGCAAACCAGTTGTAA

Downstream (in bold start codon next gene)

gtagtagtagtagtagcagagtatctacagtggtgtgtataatgtatatgagtgtactta
ccaaccaaattccgtttagtatttcgtcgacgatgatgtagtacgagtccttcgcgaatc
cgttactctcaagacggggaaaaaaaacgacgaaaatgaccaacttactcaactaagcaa
acctcaagaaacataacactttgttgtgagacagtaataaaaagctcacagcgtacacat
caccactcatata**atg**tcagagagaagcagcaagaaaggccctaaaggcggcgcgaagcg
ttcgtcacagggctcctctcaggggctggaaagcactaaactggccactttgaccgaatt
gttcccagattggacggcacaagacttggagcctgtgctggaggaatatccagatgaaga
cctcaatgtgattatagaaaacattatcagcggaaaaataaacaaatggactgatccatc
agctaagaaggagaagaaaaagagagaagaatcctttaatgcaagtgaagaattatcaac
tccctcttatcaccaaacacctaacagcgcaaagaaagagtatcctaagaaagaagttaa
ggctaaatccaagaagtctcaaccgcgttctacgacatccacgactactgcatctactaa
agctcaactgacgccatcgtctaatccaagcacaaaaagttcatgggcggctgctctaca
tcagaaacaagaggataaaccttcttcaaccgtaactcccactactgaaaccgaaactcc
aaatggcgaaaacgcatctcagtctccagttgccgagacaaagtctgaacaagaagagtc
ttttgcccccgccgcagttgttgaaacttccgctaaaccaaagtcttgggctgctatggt
cgctcaatctgctaaacccaagaagaagattttgaaaagacctgaacaagctgcaaagcc
ctctagcaacgaggaattgtcgcaacaaaatggggaaattc

AA

MPDQLSFLKRKTGVIVGDDVRNLFLHAQKKGFAIPAINVTSSSTAVAALEAARDNNSPIILQTSQGGAAYFAGKG
VKNSNQEASIGGAVAAAHYIRSVAPLYGIPVILHTDHCAKKLLPWFEGFLAADEAYYKEHGEPLFSSHMLDLSEE
SDDENIATCVKYFKRMAKMNQWLEMEIGITGGEEDGVNNENADKELLYTKPETVFAVHKALAPISPNFSIAAAFG
NVHGVYKPGNVVLRPSLLGDHQKYAKEQLKSESDKPLFLVFHGGSGSTQEEFDTGIRNGVVKVNLDTDCQYAYLI
GIRDYVLKKKDYIMSQVGNPEGDDKPNKKYFDPRVWVREGEKTMSARVTEALNIFHTANQL

SEQ0066

chromosome 1-1, 0319

5' region (in bold start codon next gene)

aactctacccaggattatttttcttctgcgaatacaaaactgcttatatgtcacacggat

aactcctcttttaacgagatagttgacttctattaaaaagtccgcatagttagatttacc
tccatcttgagttagaagatgaacctttcattatagggtggatcataccaattccaacc
atctggggccaaggttgaaaaactgggggttcttggaaagttatagaagaacaaagttcc
tattctggcaaaattatgaaggtcattggtgacattggtaacacttgtcagatc**cat**aaa
ttaatccataagataaggcaaatgtgcttaagtaattgaaaacagtgttgtgattatata
agcatggtatttgaatagaactactggggttaacttatctagtaggatggaagttgaggg
agatcaagatgcttaaagaaaaggattggccaatatgaaagccataattagcaatactta
tttaatcagataattgtggggcattgtgacttgacttttaccaggacttcaaacctcaac
catttaaacagttatagaagacgtaccgtcacttttgcttttaatgtgatctaaatgtga
tcacatgaactcaaactaaaatgatatcttttactggacaaaaatgttatcctgcaaaca
gaaagctttcttctattctaagaagaacatttacattggtgggaaacctgaaaacagaaa
ataaatactccccagtgaccctatgagcaggatttttgcatccctattgtaggcctttca
aactcacacctaatatttcccgccactcacactatcaatgatcacttcccagttctcttc
ttcccctattcgtaccatgcaacccttacacgccttttccatttcggttcggatgcgact
tccagtctgtggggtacgtagcctattctcttagccggtatttaaacatacaaattcacc
caaattctaccttgataaggtaattgattaatttcataaat

ORF

ATGTCTACATTTGATTTCCTTTCCAGAAAAAGCGGTGTCATCGTTGGAGATGACGTTAGAAAGCTATTCGAATAT
GCAAGAGAAAGGAAATTTGCCATTCCTTCCATCAATGTAACGTCCTCTTCTACAGCGGTTGCTGTATTGGAGGCT
GCTAGAGACAACAAATCACCAGTTATGCTGCAAGTATCTCAGGGAGGTGCTGCTTTTTTCCTAGGGAAAGGCGTC
AATAATAAAGACCTTAGCGCTTCAGTGACTGGATCTATTGCTGCTGCTCTATTGATCAGAACTATTGCACCTTCT
TATGGCATACCTGTCATTCTGCACACTGACCACTGTCAAAAAAAAATGGCTCCCTTGGTTTGACGGAATGTTAGAT
GCAGATGAAGAATATTTCAAGACCCATGGAGAGCCTTTGTTCTCCTCCCACATGTTAGACCTGTCGGAGGAAACA
GACGATGAAAACATTGCTATTTGTGTGAAATATTTCAAGAGAATGACAAAAATGAACCAGTGGTTAGAAATGGAG
ATTGGGATAACTGGTGGGGAGGAAGATGGAGTGAATAACGAAAATGCTGATAAAGACTCACTCTATACCAGTCCC
GAAACTGTTTTTGCAGTTCATAAAGCACTGGCTCCTATTTCTCCAAACTTCGCCATTGCTGCAGCCTTTGGCAAT
GTTCATGGGGTCTACAAACCTGGCAATGTGGAGTTGAGACCATCAATTTTGGGTGAACATCAAGCTTACGCAGCT
CAACAGTTAGGCCTTAAAAATGGATCAAAACCACTTTTCCTAGTTTTCCATGGCGGTTCGGGATCTTCTCAACAA
GAGTTCAACACTGCTATTAACCATGGAGTTGTTAAGGTTAATTTAGACACTGATTGTCAATATGCATACACCATT
GGTTCAAGAGACTATATCCTGAAAAACAAAGACTATCTTCAATCCATGGTTGGAAATCCGCAGGGAGCTGATAAA
CCCAACAAGAAATACTTTGATCCAAGAGTGTGGATTAGAGAGAGTGAGAAAACCATGAGTGGCCGTGTTAAAGAG
GCTCTGGAGGTGTTCCATGCTGCTGGTACCTTCAAGTCTGAGTCAAAACTGTGA

Downstream (in bold stop codon next gene)

ataaaatttctctttattctttaagt**cta**taggctattatacttatccaggtagttact
gtagtaagcagaaaggagacctcttggagttacaagaaaaactcgcagtggttcctttga
actcgttcctgttgtcgagatacccatacccgtaagcttattaagctttgacgcaataga
tgattgtttacgtcttgcaggctttaccagtagctttccggaccatgtttcgttggaagg
gagatgctgaatacacttgatagagtctcggattattcttctcctggaaggtatttttc
atttatgggcctccacttaatttcggtttctgttgagctcgctgtggttatagaagatcc
tgtgacattagaagatgaaaatgttaattgagtggaatcggctatggaaggtgaagtact
ggcggaaagatgaaccatgtctagatcgtgatttaaatcttccactctaagaaactgttc
tataggttccagtctttcttcctcttcttcatcttcatcttcatcgtcgtcgtcttcatc
ctgaggaggatactcatcaagatcatcctgatcgtcatgattagtattttgattttgggc
ttcatcgactgtgtcatcacaggaacaaccctcaggaatatcattttccatcaacactgg
agcaaggcagtcaatttcatcaagagtatcgtcattactttggcccatgagaaattttgg
tttctgccccggacgcaatggagagtgttcaactgaactggaaatagttggggtgacagt
aggtgtaagaggtaaatcgttggaactggatatagtgagattaagtcgttcgttcaagga
agacggcaaagtgggagttgttgcaccagattttgatctcttcctaaagaaatcagacat
agatctcgatctatatctagaggcagcgtctgatgatcttgtctccgtagagcttagtat
ccttgggttggtgttccttcgggtccttcgaggaactgttct

AA

MSTFDFLSRKSGVIVGDDVRKLFEYARERKFAIPSINVTSSSTAVAVLEAARDNKSPVMLQVSQGGAAFFLGKGV
NNKDLSASVTGSIAAALLIRTIAPSYGIPVILHTDHCQKKWLPWFDGMLDADEEYFKTHGEPLFSSHMLDLSEET
DDENIAICVKYFKRMTKMNQWLEMEIGITGGEEDGVNNENADKDSLYTSPETVFAVHKALAPISPNFAIAAAFGN
VHGVYKPGNVELRPSILGEHQAYAAQQLGLKNGSKPLFLVFHGGSGSSQQEFNTAINHGVVKVNLDTDCQYAYTI
GSRDYILKNKDYLQSMVGNPQGADKPNKKYFDPRVWIRESEKTMSGRVKEALEVFHAAGTFKSESKL

SEQ0067

P. pastoris homolog of *S. cerevisiae* **FBP1 (YLR377C)**

Fructose-1,6-bisphosphatase, key regulatory enzyme in the gluconeogenesis pathway, required for glucose metabolism
chromosome 3, 0868
E.C.3.1.3.11

5' region (in bold start codon next gene)

ctttgatgatgatttgttgtgaacgttccatgtgtttgaatctttggacggccggtaaaa
accgactgtttttaatgattgctggttagccttgtcgacaaatgatttgtatgaagctct
gaccactgaaaggatttcagatcggttgaaatgatcgacttttgctgtttgagtgttggt
attaggagttgtagtggcaccagtcggggtgctacttcgtgatgagctgggattctttgc
gttcacagccttcgaagtggcatgcttgtcatttcttatcttatccttaccagcagccag
gttcgcccactcaccttgaactggtacaaggatatttctctgtttttgctcggtggttcc
cattgctcaaacgagtggagagggaaatcgattcagcagttaaatcaatgctggaaaata
ttcgagattacctaatcggatctggaacttacttcgacctgacattttcttgcctgggga
gccacgatcgattatgtaatcaagaatatggacagagggaaacagatttagctgtcaaaa
gcccaagagaagctaccgatcaatggatgcggatagataaagaaaagtccttttttttttc
attagccatccgagttgtccaatcaaatgtctgcctgctacgctggagaggaatcacgcg
tgtttaacattcggattgtcgcctaaaataagcctattacctacacagtaaaacccgggg
ggtgctttggtatcaatgaccccgggattttatccaccagtttttttctttctggcaaga
gtgcattgcatccccgtacaaatagtagcaacctccacaagaggaatcccctatgagcga
gaagtccatagtaataccccgcggaaaagagatattttgtttccgtgttgcccttgaac
ttcagtttcccccatcagtttatatagtagccgggttcccaatctctagcccttctttcc
tcctatttcattcctctcttcttacgttatcttacattagc

ORF

ATGTCCAATAACACCACCCAAAACTTGGCTGAACAGAAGGGTATTCAGACCGACTTGGTCACTCTGACTCGTTTC
ATCTTGGACGAGCAGAAGAAATCTGCCCCAAACGCTACCGGTGAGCTGACTTTGCTTCTGAACTCCCTGCAATTT
GCCTTCAAATTCATTGCTCACACCATCAGAAGATCCGAATTGGTCAACTTAATTGGATTGGCCGGTGTGACCAAC
GCCACTGGTGACGATCAGAAAAAGCTGGATGTTATTGGAGATGAGATCTTCATCAATGCCATGAAGGGATCCGGT
AACGTTAAGTTACTTGTTTCCGAAGAGCAAGAAGACCTTATTGTGTTTGAGTCCTCCAAGGGAAACTATGCTGTT
GTCTGCGACCCAATTGACGGTTCATCCAACTTGGACGCTGGTGTCTCAGTTGGTACCATTTTTGGTGTATACAAA
CTGTTGCCAGGCTCAGCTGGTTCTATTAAGGATGTTTTGAGATCAGGAACAGAGATGGTGGCTGCCGGTTACACC
ATGTACGGTGCATCTTCTCACTTGATGCTTACTACCGGAAATGGAGTCAACGGTTTCACCTTGGACACCGACTTG
GGAGAGTTCATTCTCACGTATCCATCCTTAAAAATTCCGCACACCAGGGCTATCTACTCTATCAATGAAGGTAAC
TCGCACTACTGGACCGATGGAGTCAATGAGTACATTGCTTCTTTGAAGAAACCCCAAGCAAACGGAAAACCATAC
AGTGCCCGTTATATCGGATCAATGGTTGCTGATGTCCACAGAACTCTGTTATATGGAGGTATCTTTGGCTATCCA
GCAGACTCCAAGTCTAAGTCCGGTAAACTGAGAGTCTTGTACGAGTGTTTCCCAATGGCCCTCCTGTTGGAGCAG
GCTGGTGGTGAGGCAGTCAACGATAAGGGTGAGAGAATTCTAAACTTGGAGCCAAAACAAGTGCATGAGAGATCT
GGTATCTGGTTGGGTTCCAAGGGAGAAGTAGAGAGATTACTACCCTATCTAACAAAAAAAATTAAGATCCAATCT
GTGAATCTCTAG

Downstream (in bold stop codon previous gene)

tgcgttcgattggcactgtttccgagatttgatactttgtacacgatgttatatgaggtt
atatacttgataagagggtttttacgtttgcaattagcacaatttcggagtagcactggc
gggagtgaaccttgagtagtctggatcaatgtaatcttcgtataggctagacaccccgga
ttgggagtgctgacgaaatgatgttgggatgtgatgacatcatagggataatagaaaagt
aaggttccgcgtgagccgttgaacgcgcactggaatggatggtctgtgacgtagccagac
tgaacttgaaattccttccaagaaagtacatttttatttcattcattcattcgaaaggga
ggcttgtgggggaacccccaatcaaatacctaactactacttacaatatccaaacctaac
aacgagctcctcattcccga**tca**ctccttctttctaccaaataatcctctcttcttcttg
cccttaccaccacttcccgttggtgaattaacatcctgttcaaaagtattctgatctcct
gaaaacgagatcacggggttttgaccagcaatgggggacgaggtctcattctgtcgtga
tattcgtagtttctgccggactcataacctggaaaatcaggacgtaccctccaacctaaa
gtcggggtctccagtagttgtttataagtagtatctccagtgatggcatattcaaaggaa
cggatcgtatctaaaggcctctcgtctctcgatctggcaggattggaaacgtcggggtta
ttgattggttgtccaaaaatgtcacggatgtactctccatcgcccgacttattactaaac
gattgttgtctggcatgagtgtagtccgttgattgttcgaacggctgagcctcgttgaca
gccatcaagatggggtcgtgaatgttgtttgatctgtacttgagtctgtccgtgtcgtct
ttggtgacgtgaacttttggtactttaccaacatcagtttt

AA

MSNNTTQNLAEQKGIQTDLVTLTRFILDEQKKSAPNATGELTLLLNSLQFAFKFIAHTIRRSELVNLIGLAGVTN
ATGDDQKKLDVIGDEIFINAMKGSGNVKLLVSEEQEDLIVFESSKGNYAVVCDPIDGSSNLDAGVSVGTIFGVYK
LLPGSAGSIKDVLRSGTEMVAAGYTMYGASSHLMLTTGNGVNGFTLDTDLGEFILTYPSLKIPHTRAIYSINEGN
SHYWTDGVNEYIASLKKPQANGKPYSARYIGSMVADVHRTLLYGGIFGYPADSKSKSGKLRVLYECFPMALLLEQ
AGGEAVNDKGERILNLEPKQVHERSGIWLGSKGEVERLLPYLTKKIKIQSVNL

P. pastoris homolog of *S. cerevisiae* **HXK2 (YGL253W)**

Hexokinase isoenzyme 2 that catalyzes phosphorylation of glucose in the cytosol; predominant hexokinase during growth on glucose; functions in the nucleus to repress expression of HXK1 and GLK1 and to induce expression of its own gene
E.C.2.7.1.1
E.C.2.7.1.2

2 homologues

SEQ0068

chromosome 1-4, 0561

5' region (in bold start codon next gene)

cattttcaccagttccatgttcttttgctttcactggaccgctttcatctggggaaacct
ctttcgggaaatcctcatccggggaaatcttagtgggggaagcgatttctgggggatatt
tttcatcgctcctttgactcgcccatctttaaacctattctggtattcgatcccacggc
tcgtatcaaatgagtctcttccgaaatccaacgtaggccttcctgattctgcatacttgg
gtgaattatcagcagacgacttttccttcctttttctgacgttgtgcggctggtctct
tcttttcactacagatcctttcctagatacagttttggaggacggataagaattccccg
gattacgacgggaagcagagtagtggttagaagataccctctgtctacggtttttagacc
taatctttttctctctgatcctctgggccttgaacttcctggagtcatttctcttactct
gtgaccactgggaattatcctcatctaccaggttccacggaacatttaccacctgaagga
aaatttgatcagaatcgagtggctccactccagaattttccattttctccgtggggcttg
gcgtagaaacggacacagccacgagagagagaaatacttcagtaagaattatcaagtt**ca**
tatctagcaactaatagtctctgtacttctacagtagtaactaaggaggagtcacgcgtg
ttgtctcctctgtgtttttaacccgaaagggcaggggggggcagcaaaatgtttcacaaaa
gcagcgcagactcggatgcggataatactcaaatttgtcacgtcagagtttttccgaaag
tgagatgcaagaagcgcgcgggggatgagagtgccttaaaccaacatgcaaaatacccccc
ctcaaaaatctatataaagtgtgcccatttttctccaaagtatttttcacacgtttttttct
ttcactctacactagagcttctgctacttcatcacaccata

ORF

ATGCCTATTGCTAAACCAGCTAACCAGGTTGCCTCCACCATAACTAGCCAACACCTTTCTGATAGGGTGTTGGAG
ATTCAAGATGCCTTCGAAGTATCTCCCACCAAATTGCAGCAGATCGTTGCTCACTTTGTGGAAGAACTCAAGAAG
GGACTGTCTGCAAAGGGCGGTAATATTCCCATGATCCCTGTCTGGGTCATGGACTACCCCAACGGAACCGAAACC
GGTGACTATCTGGCAATCGACTTGGGTGGAACCAACCTGAGAGTTGTCCTCGTTCACCTGTTGGGAGGCCAAAAG
TTCGAAACTGAACAGGAGAAATACCATCTTCCCAAAGGAATGAGAACTACGAGGAATAGAGACGAACTGTTTGAG
TTTATAGCTGATTGCTTAGAGAAGTTTTTCTACAAGCTGCATCCGAACGGAATCGAGAAAGGTACTCTCCTTCCA
TTGGGATTCACATTTTCTTACCCTGCCTCTCAAACAAGAATCGACACCGGGGTCCTTCAAAGATGGACTAAAGGT
TTTGATATCCCTAACGTCGAAGGTGAGGACGTTGTCCCATTATTGATGGACAAAATTAACGAAAAAAAGCTGCCA
ATCAAAGTGGTTGCCTTGATTAACGACACCGCTGGAGCCCTAGTTGCTTCCAGATACACCGACCCCACAACTGAG
ATGGGCCTGATTTTTGGAACCGGTGTCAACGGTGCTTACTATGATAGGGTCGGAAACATCGAAAAGTTGAAAGGA
AAGCTTTTACCGGACATTACCGATGAATCTCCAATGTTGATCAACTGCGAGTACGGATCCTTTGACAACGAACAT
GAGTCTCTCCCACGTACAAAGTACGATATTCTTATCGACGAACAATCCCCAAGACCAGGGCAACAAGCCTTCGAG
AAGATGACTGCTGGTTACTATTTGGGTGAACTCATCAGACTAATCCTTGTTGAACTTTACGAAGAGAAACAGGTG
TTCCAAAAAATATTCCAAGGATTCTGAACAGATTAAGTTACTATATACCCCATACCTCCTAGACACATCTTTCCTG
GCCGAAATTGAAGGTGATCAGGACTTGGAGAATTTCCCAGAGGTTGTCCGTCTTTTCCAAGAGTTCTTGAAGATT
GAGCCCACATTGGATGAAAGAAGGTTATCCAGAGCACTATCTGAGATCATCGGAAACAGATCAGCCAGACTTTCT
GTCTGCGGAATTGGAGCTGCTTGCACCAAGATGAATATCAAGAAATGCCACTGTGCCGCTGACGGTTCAGTCTTC
CACAAGTATCCAAAGTTCCCAGAACGTGCTGCAGACTCCTTGGCTGATATTTTTGGATGGAAGGAACAGAATATC
CAACCAAAGAACTACCCAATTCAAATTGTACCGTCTCAGGATGGATCAGGAGTCGGGGCTGCCGTTATTGCTGCT
CTTGCCCATGCTAGACAGGCAAAGGGTCTGTCTCTGGGATTGGCTGAAGCTTACAAAAAAGCCTAG

Downstream
atgttaaatgttttactctacttcctttctatgtataaagcggaaatgacaacgctttag
gaacttcc**tta**gattccttcatgagtttgatttatatcttcgatgaacataattatttta
tcatttggatcggggatttgaacaccacgtgaagctagaccgtgcatccaattcaacccc
tgcagctctttaactgtgcgtaaaaatgggattttgtctctctgtgctaattgtactatt
attttactatctcctctcaaatcatcttttaaatcgacttgttgctgggccagtttcaat
tgtggaactgccgccgagtgagtcaagtcaattaggaacccaaatagaaggatatgcaca
atcttgtaacgtaggaacattcttgatactgtccaagggagagtttacgcccttgcgatg
cttagccttttgcacgatgcactgacttgcggctttagcagcggtgagatcctgaaattc
tcggaaaaaacctagttcccttcggaaaaactttgaataaaaatccacttaccacgtctc
ctggatgtatctcccctttattctctagctctaaacataaactcttcgactcttcgttcga
aattcccaaatgatcttctaatatggtttctatggtgattctaacctgaatgggcatttt
tccggatgatccacataggtagaaagtggcatttttattataaataaggtcaaagatcag
gttactctgtgcatacaatgtgtcttgaacatatgatttttctcatctctggaaaatga
gggaaagagttggagttggctctttgatgcgagagattcccataaatcaccgtaaagaaa
atccttatcatgaaatctgtttccagtaaagagatacatgtcttgcagtcctagttctag
tctttgctcaatgattgatttgactggtgctattccagttcctggggcaaccattatcaa
aggtccagaggaaaacttgacattgttcgggtgaattgaaa

AA
MPIAKPANQVASTITSQHLSDRVLEIQDAFEVSPTKLQQIVAHFVEELKKGLSAKGGNIPMIPVWV
MDYPNGTETGDYLAIDLGGTNLRVVLVHLLGGQKFETEQEKYHLPKGMRTTRNRDELFEFIADCLE
KFFYKLHPNGIEKGTLLPLGFTFSYPASQTRIDTGVLQRWTKGFDIPNVEGEDVVPLLMDKINEKK
LPIKVVALINDTAGALVASRYTDPTTEMGLIFGTGVNGAYYDRVGNIEKLKGKLLPDITDESPMLI
NCEYGSFDNEHESLPRTKYDILIDEQSPRPGQQAFEKMTAGYYLGELIRLILVELYEEKQVFQKYS
KDSEQIKLLYTPYLLDTSFLAEIEGDQDLENFPEVVRLFQEFLKIEPTLDERRLSRALSEIIGNRS
ARLSVCGIGAACTKMNIKKCHCAADGSVFHKYPKFPERAADSLADIFGWKEQNIQPKNYPIQIVPS
QDGSGVGAAVIAALAHARQAKGLSLGLAEAYKKA

SEQ0069
chromosome 3, 1192

5' region (in bold start codon previous gene)
aaagttggcaaggccccaatcaaaagatcactggcttctttgaattttctcattgaaaga
tcatgaattcctgtgtataccttataacgatttttcctctcccaatcacccccccttctcg
attaaagcatttgcttcatccaagtacttcttggtgaattgtttgtcatcataaaagaat
ccaatacgagataaggtgagtttgatatcgatctttgcaccgatacttggagctaaatcg
gttgcctttccaagtgtttctattgctttattcttgtctccaatttgagcgtagtagttt
ccaagttcggcccaacattcagcgttctctacttcaccttcatcgtcttcatctacgact
tttattttgttggtcaattcctctaccttcttttcattttgcaagactagtttgtcatac
aacccttgatcccactccaagccgggcaatgctatttggtggtagagataatgatagtag
ggagccatctcattcgtctcaacgccctgaatgactttgtcaaccaatgcagcgtctcca
tcgccatcggctatttgtgcaagaaaactaagctttgatacgtcaaagtctggaattttg
ggaattcctgatgcataggttgc**cat**cttgaggtttcgtttagcgtaataggattggaag
gcaaatgattaacatctgggattgatccactgttggagtgtgggtagtacacgatttgaa
ttttgacggttgtcagatggaagttattctaatctataaattggaattcttagggttagt
caaacagctgacaagatctgtcaaggagtgcacaatttaccattcagtgttttcttattc
ctatctgagggctatttcggtattctcatcttaaaaaaaaagtcagagaattttggccga
ttttagctgaaatctgaatttgattctcctaatccacaaaagcacttaaagcatatcgac
accgacctttgcctgcaaagactgttgacaagctagtaaaa

ORF (in bold intron)
ATGGTTCACTTAGGGGCGAAGAAGCCTCAGCATAGAAAA**gtgggtattgacctctgaatcaggatttgacatgtc**
attaacgataagggatatctatttaatcagCTTAGTCCAGAGTTACGAAAAGCTTATAAAGAAGTAGAGGCACAG
TTCGTTGTATCAACTCCCAGGTTAAAGCAGATAGTGGATCAATTTGTTGCTGAATTGAAGGAAGGCTTGAAATCA
AGCAGCTCTAACATCCCAATGCTACCTACCTGGGTGATGGATTTCCCCACAGGAGAGGAAACTGGAGACTATCTC
GCAATTGACCTGGGCGGGACCAATATAAGGGTTATATTGGTCAGACTGTTGGGAAATAGGAAGTTTGATACCATT
CAATCAAAATATGTTTTGCCTAAATGGATCAGAACCTCCACATCAAATGAACTTTGGCTTTTTATTGCTCAATGT
GTGAAGACTTTCATTGATGAAGAGTTTGATTATAGAGAAAGTCCGGAAGACCCAATCCCCTTAGGGTTCACATTT
TCTTACCCTGCATTCCAAAGTAGGATTAATTCGGGTGTCTTACAACGTTGGACGAAAGGATTTGATATTCCGGAC
GTTGAAGGCCATGACGTTGTCCCTATGCTGCAAGACGCATTGGAATCATTGGGATTGTCCGTTGTGGTAGTGGCT

TTAATTAATGACACTACAGGCACTTTGGTAGCTTCTACGTATACGGATCCAGAGACTAAAATGGGACTGATATTT
GGAACTGGTGTTAATGGTGCATACTATGATACGATAAGTTCGGTCTCAAAGATATCAAATGCTCTTCCACCAGAT
ATTCAAGAGGATGCAAGTATGGCCATCAACTGCGAATACGGTGCTTTTGACAATAACATCTCAGTTCTACCTAGG
ACCAAATACGATGATACAATTGATTTAGAATCGCCCAGACCAGGGCAACAGTCGTATGAAAAAATGATATCAGGC
TATTATCTTGGAGAATTGTTACGATTGGTGCTTGTAGATTTGCACCATCAAGGACACATTTTCAAGGGACAAACA
ATCGGGAAACTAAATGAACCGTTCATTATGGATACATCCTTTCCTGCGAGAATTGAAGAGGATCCGTTTGAGAAT
CTATGTGAGACTGGAGAACTTTTTAACAGCCTAGGAATTGAAACTACGGTTCCCGAAAGAGAACTGATTAGACGA
ATTTGCGAACTCATAGGAACAAGAGCAGCCAGACTGTCAGTATGTAGCATTGCCGCAATTTGCAAAAAACGAGGC
TACAAGAAAGCCCATTGTGCTGCTGATGGCTCTGTCTTCACCCGCTATCCTTATTTCCCTGACAGAGCTGCAAGA
GCGCTCCGAGATATATTCCAATGGGGCCATTCAACTCCGGACTTAGTTACTGTAGTTCCAGCAGAAGATGGCTCG
GGAGTTGGGGCAGCTATTATTGCTGCACTGACCAAACAACGGATGGCAAATGGTGAATCCGTGGGCCTTGACGAA
TATCACCCTCAAGATGGAAAGGATAGTAACTAA

Downstream (in bold start codon next gene)

gtaactaatcccagatctgcttatgtaacgaaatatttacttctcgcttttactcccacc
aacgaactcgaacgctcatacagtccaagcctaccatctacacacaacatcaccaaataa
atggaacaaccaattggtaagttttaagaacttttttgggccgcgtaggctccagaatg
ggccaatggaagtttgtattattcacagggcagaatcaatgctcttgtcgcagtattgca
acccgttgtccacttctgtatctcaagaattcaaccactaaccgcattttaagttctgga
tcaaggaacaggattcgttaagatcggccgtggaggtaccaacttcccggaccacacgtt
cccttctattgttggaagacctatactgcgtgctgaagaaagaactgggactgtggaaat
taaagatatcatgtgtggagacgaagccagtagtgttcgtagtgctctgcaaatatcgta
tcccatggagaacggaatcatcaaaaactgggaagacatggaacacctctgggattacgc
cttttatgagcgtatgaagatagataccacggatcgtaaagtcttactcacagaaccccc
gatgaacccgctcaagaatagagaaaagatgtgtgatgtcatgtttgaaaaatatcactt
cggacacgtttatgtggccattcaggccgttttagcgctttatgctcaaggattatcatc
aggagttgtcgttgatagtggagacggtgtcactcatattgtccctgtttatgaatctgt
ggttctgagccacctgacaaagcgccttgacgtcgccggtagagatgttacaagaaatct
catcaacttgcttcttagaagaggctatgcgtttaacagaacagctgatttcgaaacagt
tcgtcagattaaggagaaactatgttacgtttcatacgacttggattttgacaccaagct
ggccaatgatactaccacactagtcgaaagctatgaacttc

AA

MVHLGAKKPQHRKLSPELRKAYKEVEAQFVVSTPRLKQIVDQFVAELKEGLKSSSSNIPMLPTWVMDFPTGEETG
DYLAIDLGGTNIRVILVRLLGNRKFDTIQSKYVLPKWIRTSTSNELWLFIAQCVKTFIDEEFDYRESPEDPIPLG
FTFSYPAFQSRINSGVLQRWTKGFDIPDVEGHDVVPMLQDALESLGLSVVVVALINDTTGTLVASTYTDPETKMG
LIFGTGVNGAYYDTISSVSKISNALPPDIQEDASMAINCEYGAFDNNISVLPRTKYDDTIDLESPRPGQQSYEKM
ISGYYLGELLRLVLVDLHHQGHIFKGQTIGKLNEPFIMDTSFPARIEEDPFENLCETGELFNSLGIETTVPEREL
IRRICELIGTRAARLSVCSIAAICKKRGYKKAHCAADGSVFTRYPYFPDRAARALRDIFQWGHSIPDLVTVVPAE
DGSGVGAAIIAALTKQRMANGESVGLDEYHPQDGKD

SEQ0070

P. pastoris homolog of *S. cerevisiae* **TPI1 (YDR050C)**

Triose phosphate isomerase, abundant glycolytic enzyme; mRNA half-life is regulated by iron availability; transcription is controlled by activators Reb1p, Gcr1p, and Rap1p through binding sites in the 5' non-coding region
chromosome 3, 0951
E.C.5.3.1.1

5' region (in bold stop codon previous gene)

tcaacgagacactcttccgtcagttccaaaaccataagtttgccgatgtgttggtcctt
gtaacgcatggaatttgggccagggtatttttgatgaaatggttcagatggtctgtggag
gagtttgaaggcttacgaaatataccacattgccagtttatacagatggttaagggtgaa
aatcaacgttacaccttgacgaccccattattacgatggcgtgaaggagatgaagaccgg
gtagaagaaataagaaaagcggtacagtttaggtccggagatctagggaaggaggcctta
gcttatattgtagctgctgagagagaggcagctgctggaagatctgaaggccctatcacg
tatgatgatggtgatgaccat**tag**agaacgcccagagattgatagccagttcttggacaa
caattcggaactttattcacggtgcaaacatgatttgtgtggatagcttcaagtcagaca
tttcatctcatccccctttttactgctgctaatcaccgttagtccgacagttactctaat
caatatttattagtgttttagttgcgcaaaactcgagcctcttttccttatctcttgaca
cttcctggagtcgaagtttttcagcgcaaattcactctacaatgtctaccgatactagac

cgcctatcttccccctctaaatagcctattggaagggtgcaataaggtatataaatctgg
cgcgattcccccggacttttatgatccacatcacctcatcttactgccctcactctctttt
cctgatcctcccaggtccaccgatttcctcactatcgtcggatttctccttccagcgccc
tagagaattccgtaaccaccgcaaaaatagcagcccccccctcacccatttttttattta
aaagaacaccttactggcccgttttcgtttctcctttactacaattgatttttaattttc
agtttttttttcattgatatacaagatctatcacaaacaca

ORF (in bold intron)
ATG**gtacgtgcattgagtctcataagtgccatccagtcacacctggcaggtggaggcaatgggaggtcggagaaa**
cagaacccacatcagaaggaccatactaactcttcccagGCTAGAACATTTTTCGTAGGAGGAAACTTCAAAATG
AATGGCTCCAAGAAGTCAATTCACGAGATTATCGAGAGATTGAACAACACCAAGTTACCAGAGAACGTAGAGGTT
GTGATTGCTCCTCCTGCTCCCTACTTGCAGCAGGCTGTCACCGAGAACAAGCAGAAGACTGTTTACGTCTCCGCA
CAAAACAGTTTCGACAAAGCTTCTGGTGCCTACACAGGTGAAGTTTCTGTCGAGGCTTTGAAGGACTTGGGTGTC
CCTTACGTTATCTTGGGTCACTCCGAACGTAGAACTATTAACAAGGAAGACGATGCTTTCATTGCTTCCAAGACC
AAATTTGCTCTTGACCAGGGTTTGAAGGTTATCCTCTGTATTGGTGAGACACTCGAGGAGAAGCAAGCCAACATC
ACGCTAGACGTTGTTAAAAGACAACTGCAAGCTGTTGTCGATGTTGTTTCCGACTGGACTAACATTGTCGTTGCT
TACGAGCCTGTCTGGGCTATCGGTACTGGTCTTGCTGCCACTCCAAGCGATGCCCAAGACGTTCACAAGCAAATT
AGAGATTTCCTTGCTACTGTCATTGGTAAGGACCAAGCTGAAAAAGTCCGTATCTTGTACGGTGGATCCGTCAAC
GGCAAGAACGCTGTCGAGTTCCGTGACAAGGCCGATGTTGATGGATTCTTGGTTGGTGGAGCTTCTCTGAAACCT
GAATTCGTGGACATTATTAATTCCAGAAACTAG

Downstream (in bold stop codon next gene)
tttacatatgaacatattactactctatattcgggacagcctcgattatttctctttct
cttcgtctcttgtttaaagtcttctt**tca**tatcgttccttttttcatcctctcgttccgct
cgatccttcaacgttgaaagagccagaggtgtcatattgggaagagaaccaccaacggga
tctaagtttggagattctgaccctatatgctgtatgacggattgttgctgtatctgtaac
ccaaggggtggagtttgtagccacatctttgatggtgacccgacgactgagactggcatc
aagccaggtgagtcaggggtgcggaccgtacttccagagccgtgttgcaccagtgttatt
gcaggaccatgagtgggcaccgccgtataatgctgatgcctttggagcatgagatggttg
ctggaattagctcttagatgctgtggcggtagccggatagcctttacaacttgttcctct
tcatcgtccactactgcactagaagtgctatcattactattgttctgttggttgtgtata
ttcaatgaaagtaaactaatagtagaggagctgggtactagacggccctccacatcttta
ccaacagacgaggccagaagtggtggtgacggatgttcatcagtgttgttagtccatcta
cttgaatttgtattgagcccgcttgctggaatttgaatgggcatgccctgcatgctttcc
attagctcaggctctgaaagtactgagcctctaaaagatggaggcatgtctaagtcgaag
tggattggttttgataagtttgctgtaaactaatttttcgtatagagttcgcgaaaattc
tcaccaccaccaagggtcaagtcactgatcgcttcaacccacaaaacgacgatattaaag
ttcaacatattgttctaggtacgccagtccgcgatccaagtttgaaatctccacatctac
aggtgcgtactgtagtagttcatccaaagctgggatcttg

AA
MARTFFVGGNFKMNGSKKSIHEIIERLNNTKLPENVEVVIAPPAPYLQQAVTENKQKTVYVSAQNSFDKASGAYT
GEVSVEALKDLGVPYVILGHSERRTINKEDDAFIASKTKFALDQGLKVILCIGETLEEKQANITLDVVKRQLQAV
VDVVSDWTNIVVAYEPVWAIGTGLAATPSDAQDVHKQIRDFLATVIGKDQAEKVRILYGGSVNGKNAVEFRDKAD
VDGFLVGGASLKPEFVDIINSRN

SEQ0071
P. pastoris homolog of *S. cerevisiae* **ALD4 (YOR374W)**
Mitochondrial aldehyde dehydrogenase, required for growth on ethanol and conversion of acetaldehyde to acetate; phosphorylated; activity is K+ dependent; utilizes NADP+ or NAD+ equally as coenzymes; expression is glucose repressed
chromosome 4, 0043
E.C.1.2.1.9.

5' region (in bold stop codon next gene)
atggactgttcaatttgaagtcgatgctgacgatgtcaagagagatgctcaattatattt
gtcatttgctggttacactggaaacgctacttttgttggcggaaactctaccagtttggc
cgtccatgtaaacgatgtcgttctgggccgtgaccgtttcaacacgaacataaccaatga
caaatccacttacaggtctagttcatatggaggcaattggtaccttacttctttggatgt
cccaagtggggctttaacgtctggtactaacaatgtctcgtttgtcactacaaactccga
ggtaaataaaggattcttgtgggattctctcaagtttgtttggaagttg**taa**caggttta

taagcatatcgtgcgcttgtccacaattgaatcatttattgttgcgagatacatgaacaa
agtgtgaactgggacccattactacaattcccacgcaaccgttgtttcaaagcccatatt
ttttgacaattgtttcgttacacccccagtttgatgtacatcgcttgcaatgatgtgtgt
cccggagtattttccatattcagcttgaattcgtatactcaaccaatatctgggggtata
cttttatgtaacctatacaaatcaactatactatttcacctttcgaccaatcatctccca
tcttgttaagttttgcttcctatatccctgaccctgacatcacccatgattccgctcaac
ggttctcctctacatcgtccctcttttggagagggtgttcagtttgacattcaaattacc
ccccgccatcacgcgcaaccgagaccgcacccccgaattttcacaaattaccccacaccc
tatactccaccactatgagggttattagaactgatcacgtataaataccaccgcaagttc
ccaagggatcgtgttcttcttctccaattgcaatcatatttctgactctttctagttcag
attaattcctttacacttgctttttttcccttacctttatcc

ORF

ATGACATTTGCTCCTCCCTTAGAATTCGAGATTGACCTTCCTAACGGATTGAAGTACACT
CAACCATTGGGACTCTTCATCAACAATGAGTTTGTTGAAGGTGTAGAGGGAAAGCTCTTA
CCAGTGATCAATCCTTGTGATGAGACTAAAATAACCCAAGTTTGGGAAGCTTCTGCAGCG
GATGTTGACCGTGCTGTTGATGCCGCTGAAGATGCTTTCAACAACTCCGTATGGGCTACT
CAGGACCCATTAGAGAGGGGAAAGCTGATGAACAAATTGGCAGACCTTATCGATCGTGAC
TTCAACATCTTGGCTGGTATCGAATCCATCGACAATGGTAAGGCCTATACCTCTGCCCAG
GGTGATGTTACTCTTGCTGTCAACTACATCAGATCCTGTGCTGGATGGGCCGACAAGATT
TTGGGAAACGTTGTTGATTCCGGAAACACCCACCTTAACTTGGTTAAAAGAGAGCCATTG
GGTGTTGTGGGACAAATTATCCCATGGAACTTTCCTCTCCTGATGTTGGCTTGGAAGTTG
GGACCTGCGCTGGCCACAGGTAACACTGTTGTTTTGAAGACTGCCGAGTCTACCCCTCTG
TCGGGTTTATACGTTGCCAAATTGATCAAGGAGGCCGGTTTCCCACCTGGTGTGGTTAAC
ATTCTCAGTGGTTTCGGTAACCCAGCTGGAGCTGCCATCGCTGCTCATCCCAGAATCAAG
AAGATTGCTTTCACCGGATCCACTGCAACAGGCCGTAAGATCATGGAAGCAGCCGCTAAA
TCTAACCTGAAAAAAGTCACTTTGGAACTAGGTGGTAAATCTCCAAACATTGTGTTTGAA
GATGCTGATATCCAGAAGACTATCCATAACATTATTTTGGGAATCTTCTTCAATTCTGGT
GAAGTCTGTTGTGCAGGTTCCAGAGTCTACATTCAAGACACTGTGTATGAAGAAGTGCTT
GAAGCCTTCAAGAAGGAGACTGATAACGTTAAGGTTGGTGGACCATTCGAAGAAGGTGTC
TTCCAAGGGCCTCAGACCTCTGAGTTGCAACTTAACAGAATCCTTAGTTACATCAAACAC
GGTAAGGATGAAGGTGCTCGTGTAATTACCGGTGGTTCAAGATACCGTAACCGAGGTTAC
TACATTAAGCCCACAATTTTTGCTGACGTTACTGAAGACATGAAGATTGTCAAGGAGGAG
ATTTTTGGTCCTGTGGTTACTATCACTAAGTTCTCTACCGTGGATGAGGTTGTTGGATAT
GCCAACAACACCAACTATGGTCTAGCTGCTGGTATTCACACAAACAACTTGAACAAAGCC
ATTGATGTTGCCAGTAGAATCAAGGCGGGTGTCGTTTGGATTAACACCTACAACGATTTC
CACCACATGGTTCCTTTCGGAGGTTATGGAGAATCTGGTATTGGCAGAGAGCTTGGTGCT
GAGGCTTTGGATAACTACACTCAAGCCAAGGCTATCAGAATTGCTTACACTCCTGAACAT
AAGTAG

Downstream (in bold start codon previous gene)

tagttatgtatgtttatgtgatgttaatgatctaattgtgaacttcattcaactggatga
cgctgtaaaaggaatgccttggtggctgatgtcgatgggggcttgcaagaaacctttgct
gaatggttggtggtgactcatataagcaagatcatgcagtcatggagccccatcagttat
cagtaaattgtcctttcaaggcacaagtgtttgcgtagaaagcttaatcattcagttaca
agcagttactataaaggggttcaatcggatccagggcgaggctgaatcatcaatgctgca
ctactaagcacattcaagtacacccgacaacactgaagactccaataccagcatcggctt
acccaaaaaaaataataaaacggggcggtgcttcaatgagtcatatcaatgcggagtatgc
aagtgtatagtcctagtttcccactcgtagatctgacacatctgaaaaattataaatagg
catcgagttttccctacctcgctttatcatttttcaaccatttcgttccattcacaatga
agttggcagaagccctcatcatcagaaaagacatatataagaacctgaagcaattggaga
ataggatacggcggaacatcacaattcaagaaggtacttccgcaccagagaatagcaatg
cattattggtagagtacaatgaacagcacaaggagtacgtcggccttgtcgttaaaatca
accttaccaatgccaagatctttttaaaattctatcaccccattgaagagaaagaggtag
aagcaacaatgaccgaggcgttagcattgagggactacctgaaagagagatcgtctgcat
taagagagtttgctgcggaagcctccagttccagagctgttttgacaagaacagaagtca
gatttaaaaccacaatgaaagctagtgatatacaaaaagaacaagatagaaccggtaaac
ttttgcgacagttggatctcaagatacaagaaaaaaattgg

AA

MTFAPPLEFEIDLPNGLKYTQPLGLFINNEFVEGVEGKLLPVINPCDETKITQVWEASAADVDRAV
DAAEDAFNNSVWATQDPLERGKLMNKLADLIDRDFNILAGIESIDNGKAYTSAQGDVTLAVNYIRS
CAGWADKILGNVVDSGNTHLNLVKREPLGVVGQIIPWNFPLLMLAWKLGPALATGNTVVLKTAEST
PLSGLYVAKLIKEAGFPPGVVNILSGFGNPAGAAIAAHPRIKKIAFTGSTATGRKIMEAAAKSNLK
KVTLELGGKSPNIVFEDADIQKTIHNIILGIFFNSGEVCCAGSRVYIQDTVYEEVLEAFKKETDNV
KVGGPFEEGVFQGPQTSELQLNRILSYIKHGKDEGARVITGGSRYRNRGYYIKPTIFADVTEDMKI
VKEEIFGPVVTITKFSTVDEVVGYANNTNYGLAAAGIHTNNLNKAIDVASRIKAGVVWINTYNDFHH
MVPFGGYGESGIGRELGAEALDNYTQAKAIRIAYTPEHK

SEQ0072

P. pastoris homolog of *S. cerevisiae* **PGK1 (YCR012W)**

3-phosphoglycerate kinase, catalyzes transfer of high-energy phosphoryl groups from the acyl phosphate of 1,3-bisphosphoglycerate to ADP to produce ATP; key enzyme in glycolysis and gluconeogenesis
chromosome 1-4, 0292
E.C.2.7.2.3

5' upstream region (in bold start codon previous gene)

aaagttggtacccagccgatcacgcctgctctgagtttggctggagcagcaaatctcatg
ataaccgaggtttaaattaaggtacataacaaaaattcaatgttcaaagacgcacatacc
aagacttactaatcgcagaatgttggtgcagtatttgtcgtaagccaaaac**cat**cgatgt
tgacttcctaattcagtctttaaaccgcaaaaggattctgattcgcagatggcctgatct
ccaaactcaggctggggctctaactcgagcaagtgtcctatgctgtaggccgcagccctt
ttggttcgacgacgtgcgtggttatgagacgctcggctgttttgcgctaagctggccgta
tcgagtaaattctacaggcacctgcgaggcaagcaatctactaatgtttatttttcgtcc
aacctaattgtggtttcaaagcgctatcaggtgggggtaagaggaatgtgagtggaaag
cgaaaataactggcagctggggtcagatcccgtgatgccacctcttgtggtattttgaaa
cgcgtgttgcgattggccgcgagaacggaaaggaatatatttactgccgatcgcattttg
gcctcaaataaatcttgagcttttggacatagattatatgttctttcttggaagctcttt
cagctaatagtgaagtgtttcctactaaggatcgcctccaaacgttccaactacgggcgg
aggttgcaaagaaaacgggtctctcagcgaattgttctcatccatgagtgagtcctctcc
gtcctttcctcgcgcctggcaataaagcctccttcggaggagctccgtctagagaataat
tgctgcctttctgactttcggactagcgccaaccgcgaaccacaccaccacaccatcact
gtcacccgtcatagttcatccctctctccttataaagcatctaataggttccacaattgt
ttgccaaaaaatctcttagcatagcccaattgattacgaaa

ORF

ATGTCTCTTTCAAATAAACTTTCAGTCAAAGATCTCGATGTTGCCGGAAAGCGTGTCTTC
ATCCGTGTCGACTTCAACGTTCCTCTGGATGGTGACAAGATCACCAACAACCAGCGTATC
GTTGCTGCTTTGCCAACTATCCAATATGTTTTGGATCACAAGCCAAAGGTCGTCGTTCTG
GCTTCTCATTTAGGCCGTCCAAACGGAGAGGTCAACCCAAAATTCTCTTTAAAACCAGTT
GCTGCTGAATTGTCCTCCCTACTAGGTAAGAAGGTGACTTTCTTGAACGATAGTGTTGGA
CCAGAGGTTGAGAAGGCTGTCAACTCTGCCTCCAATGGAGAGGTTATTCTTTTGGAGAAC
TTGCGTTTCCACATTGAAGAAGAAGGATCTCAAAAGAAAGATGGTCAAAAGATCAAGGCC
GACAAGGAGGCTGTTGCCAGGTTCAGAAAGCAATTGACCGCATTGGCCGATGTCTACGTT
AACGACGCCTTCGGTACCGCTCACAGAGCCCACTCCTCCATGGTTGGATTTGAATTGGAG
CAAAGAGCTGCTGGTTTCTTGATGGCTAAGGAGTTGACATACTTCGCTAAGGCCCTGGAA
AACCCTGTCAGACCATTCTTGGCCATCCTTGGTGGTGCTAAGGTTTCTGACAAGATTCAA
TTGATTGACAATTTGCTGGACAAGGTCGATTCCATCATCATTGGTGGAGGAATGGCTTTC
ACTTTTATCAAGGTTTTGGATAACGTTGCCATTGGTAACTCTTTGTTCGACGAGGCTGGT
GCCAAGTTAGTTCCCGGCTTAGTTGAGAAAGCCAAGAAGAACAATGTCAAACTGGTTCTT
CCAGTCGACTTCGTCACTGCCGACGCCTTCTCCAAGGATGCAAAGGTCGGTGAAGCCACG
GTTGAGTCTGGTATTCCAGACGGATTGCAAGGATTGGACGCTGGTCCAAAATCCAGAGAA
TTGTTCGCAGCTACCATCGCTGAGGCTAAGACAATCGTCTGGAACGGTCCTCCAGGTGTT
TTCGAGTTTGACAAGTTTGCTGAAGGTACCAAGTCTATGTTGGCAGCTGCCATCAAGAAC
GCTCAGAACGGTGGAACTGTCATCGTTGGTGGTGGTGACACGGCTACCGTTGCTAAGAAG
TTCGGTGGTGCTGACAAGCTATCCCACGTTTCCACTGGAGGAGGAGCTTCTTTGGAACTG
TTGGAGGGAAAGGAGCTTCCAGGTGTAGTTTACTTGTCCAACAAGGCTTAA

Downstream (in bold start codon next gene)

ttagttcatatagtttgaattctgattttgatgacgctcgcataaaccgtagagccacta

cggccacatgttagttgtccgtgaattcacaatttacatgatattattgcaatgctgctg
ttcgtcaattcgttgcagtcgtgataagagggattgtcatgtatgcaaaggtattcagca
actgtacaaggctgacgtacgtttagtggtcagatttgagacctcccaagttggcgcacg
gcggataacccacgttggggtcacggtgggggcatgaactttcgcgtcgatttgggtggg
ggtttttttatctgagaagttcccctgttgatgactgatggcctggaccgg**atg**gtgcatg
gcatggacgtgaacataaacagaaatatggacctctctaaatttgtgcctttaacacatc
aggctgcaggacacccggacttgctggagtctgaagactcaggcctattcgccaagttaa
caaataggaaggaggtcgaattttacagcaggctgaattctaatgtctctgaagataaac
cattgggaagcggtttgattgactgggttcctcagtttatgggagtcctaaccccaggaa
tttcacctgacttgaaatctcaaggcgctcctgtagctgctgagttggagaagaaggcct
ctgtgcaaccttcttcagataaacagtacatcttgttggagaacctattgtttggcttta
gccagccctcagtattggatatcaaattgggagtcaaactatatgatgatgatgccacag
atgataaaaaggagagactgggtaaagtcagtgattctactactagtggtagcctaggtt
ttcgaatatgtggaatggacatcaaaaagacccgtaaagaagtccacgagaaatggtccg
actacgtcacaacttaccaagacgcgcacaaggttgagtatctcaagttcgataaatggt
ttggaagagcactagacgtagactcgatccttgaagggctg

AA
MSLSNKLSVKDLDVAGKRVFIRVDFNVPLDGDKITNNQRIVAALPTIQYVLDHKPKVVVLASHLGRPNGEVNPKF
SLKPVAAELSSLLGKKVTFLNDSVGPEVEKAVNSASNGEVILLENLRFHIEEEGSQKKDGQKIKADKEAVARFRK
QLTALADVYVNDAFGTAHRAHSSMVGFELEQRAAGFLMAKELTYFAKALENPVRPFLAILGGAKVSDKIQLIDNL
LDKVDSIIIGGGMAFTFIKVLDNVAIGNSLFDEAGAKLVPGLVEKAKKNNVKLVLPVDFVTADAFSKDAKVGEAT
VESGIPDGLQGLDAGPKSRELFAATIAEAKTIVWNGPPGVFEFDKFAEGTKSMLAAAIKNAQNGGTVIVGGGDTA
TVAKKFGGADKLSHVSTGGGASLELLEGKELPGVVYLSNKA

SEQ0073

P. pastoris homolog of *S. cerevisiae* **ENO1 (YGR254W)**

Enolase I, a phosphopyruvate hydratase that catalyzes the conversion of 2-phosphoglycerate to phosphoenolpyruvate during glycolysis and the reverse reaction during gluconeogenesis; expression is repressed in response to glucose
chromosome 3, 0082
E.C.4.2.1.11

Upstream (in bold start and stop codon previous gene)
aatgaaagagtgagaggaaagtacctgggcaaaatcacacaattccaaaccatgctaaat
gagatttaaagaacaaacgatggcaaaaggcaaccgttataaatgtgatctttcttggca
gttatctgtcaatttttctaaggaacagtgaattcatcataggagagatgttatacgtta
cataatcatacatactgcatgtatctcacctactttacctcatcaactctaaaacagttc
tagtcccaaccccagattcctagtcatgacacaagtccgcaccggacaggactcacaacc
agcaagagaagctaacaaatttacgccccggtaaaacattctttaggggccgttcaatgg
taattttcctctcacccgtttaaacttacctccgggcggtatcttcaataacctctgttg
tccccgggtatcattggaaacagtgagggacgttgaacagaagagaggatcaccgtaaat
ttgccttgcaattggcc**cta**accacggatggttaacttcaagccatcacgacagcaattg
agtcggcgcatagctaccctcctcttcttgaccccatgcataggaccaaccttaaccgat
ggaacaggttcctccgctccgtcccctggtagtgtctctgcgcaagaaatagttaaggta
tgaagactgatctctcgcacccccctcacagtactgttatggtgaattgacaaagc**cat**t
ggctagattgaaacatgtaattcatatgtaatcttgttcaattaacgagcttcgtacagt
ctcaatctagacgtctgataatggcgtttgtgctcctaatcgatgagccatctcatgtga
cgtctatacgcttcgatggcttccgtcgcgaatatagaaccacttgaaatatgctgcaaa
ccacgatccaccctggtcctgaaaagatataaatacagcacatctagcaggcttttgtct
tcttggttgaaacacacaattataacaatctacatctaaaa

ORF
ATGGCTATTTCAAAAGTATTTGCCCGTTACGTCTACGACTCCAGAGGAAACCCAACCGTTGAGGTT
GACCTCTACACTGAAAAGGGACTTTTCAGAGCAATTGTTCCTTCCGGTGCTTCTACCGGTGTCCAC
GAGGCTCTGGAATTGAGAGACGGTGACAAGTCAAAGTGGCTCGGTAAGGGTGTCTTGAAGGCTGTT
GCCAATGTCAATGACATCATTGCTCCAGCTATCGTTAAGGCAAACATTGACGTTAAGGACCAAGAG
GCCATTGACGCTTTCTTGAACAAATTGGACGGAACTCCAAACAAGGGTAAGTTGGGTGCCAATGCT
GTCCTTGGTGTCTCTTTGGCTGCCGCTAAGGCCGGTGCCGCTGAGAAGAACGTTCCTCTTTACCAA

CACATTGCCGACCTGTCCGGTACTCCAAAGCCATACGTCTTGCCAGTTCCATTCCAGAATGTCTTG
AACGGTGGTTCTCACGCTGGTGGTGCTCTAGCTTTCCAAGAGTTCATGATTGTTCCAACCGATGCT
CCAACCTTTTCTGAGGCTCTGAGAATTGGTACCGAGGTCTACCACAACTTAAAGTCTTTGGCTAAG
AAGACTTACGGACAATCTGCTGGTAACGTCGGTGACGAAGGTGGTGTTGCTCCTGACATTTCTACT
CCAAAGGAGGCTTTGGACCTGATCTCTGCCGCCATTGAGAAGGCTGGCTACACCGGAAAGATTGGT
ATCGCTTTGGACGTTGCCTCTTCCGAGTTCTACAAAGACGGCAAGTACGACTTGGACTTCAAGAAC
CCTAACTCTGACCCATCTAAGTGGTTGAGTGGCCAGGAGTTGGCTGCCCTGTACAAGGAATTGATT
TCCCAGTACCCAATTGTATCCATTGAAGATCCATTTGCTGAAGATGACTGGGCTGCCTGGTCTCAC
TTTTTCTCTACTGTCGACATCCAAATCGTTGGTGATGACTTGACTGTTACCAACCCTATCAGAATC
AAGAGAGCTATTGAGGAGAAGTCTGCCAATGCTTTGTTGTTGAAGGTAAACCAAATTGGTACTTTG
ACCGAATCTATCAAGGCTGCTACTGACTCTTATGCTGCTGGCTGGGGTGTAATGGTCTCCCACAGA
TCTGGTGAGACTGAAGATACCACCATCGCTGACATTGCTGTTGGTCTGAGAGCCGGCCAGATCAAG
ACTGGTGCTCCAGCCAGATCTGAGAGATTGGCCAAGTTGAACCAAATTCTGAGAATTGAGGAAGAA
CTTGGTGACAAGGCCATCTACGCTGGTAAGAACTTCCACAAATCTGTTGCTATCTAA

Downstream (in bold stop codon next gene)
attggatattaatctatgtattaaggtaatgagtcaaacttacaattgagaactagtatt
ttacttt**tca**gaacctaagaataaggtcatctgattcattctgaggttctttaagcttat
ttaggagatttagagagtctaggaacccaagacattcatcttctgagtaaaatctgagct
caaaaataagtacccgcaatgggatttccgtgaatttctgttgaattactgtgatcgcct
tttcacgctccctgtctagaaaagttgctagagaaaaccggatggactgaagacgaaatt
gctttccagcagcatggtcctccgtatcgataatcgtcaccaattcgaaacaacttttaa
aaaggtcgtggtagttccgatcaagaaaggcataaagaatagacaaaccaatcttgcaca
gcaacattggcagtggctggttcgcttgactatcaagtttagaatccaatgcttcaatgc
gatcaacgatacattgcgcaaatttatccagttccggttccatgcgatggtaaatatagt
aacacgttaatagacagaggtgatattcataagaatcctgttcttgataaagatcaaatg
tacttctattttcggcaataaattgtagatatggtagaacttcagttgaaaaatatgatc
gtttaaggaacagtttaaccaaaacttgactgcaaactgttttgaagacaacccttttct
ccaaatctttgatagcagcagtttttttcaaatatactctctatctcttgttgcatggcta
gatcatctgagcggcttgaaatattaacgaaatgccttcgtagcagctcaacttgcttct
tcaaatttaatttgtctacttcattagtcactgtgttagaaagctgttgcgaattcggag
cactttctcgtccgaataaggaattatcaggattcaagtttggacctgatgatggaagtg
atgcagtcggtgaacctttggtgatgttggagagccatct

AA
MAISKVFARYVYDSRGNPTVEVDLYTEKGLFRAIVPSGASTGVHEALELRDGDKSKWLGKGVLKAVANVNDIIAP
AIVKANIDVKDQEAIDAFLNKLDGTPNKGKLGANAVLGVSLAAAKAGAAEKNVPLYQHIADLSGTPKPYVLPVPF
QNVLNGGSHAGGALAFQEFMIVPTDAPTFSEALRIGTEVYHNLKSLAKKTYGQSAGNVGDEGGVAPDISTPKEAL
DLISAAIEKAGYTGKIGIALDVASSEFYKDGKYDLDFKNPNSDPSKWLSGQELAALYKELISQYPIVSIEDPFAE
DDWAAWSHFFSTVDIQIVGDDLTVTNPIRIKRAIEEKSANALLLKVNQIGTLTESIKAATDSYAAGWGVMVSHRS
GETEDTTIADIAVGLRAGQIKTGAPARSERLAKLNQILRIEEELGDKAIYAGKNFHKSVAI

SEQ0074

P. pastoris homolog of *Saccharomyces cerevisiae* **CDC19 (YAL038W)**

Pyruvate kinase, functions as a homotetramer in glycolysis to convert phosphoenolpyruvate to pyruvate, the input for aerobic (TCA cycle) or anaerobic (glucose fermentation) respiration
chromosome 2-1, 0769
E.C.2.7.1.40

5' region (in bold stop codon next gene)
ctctcatgaaccggatagcacgtttgccgctaacaaatcagttaaaatgtataagtacca
ttctttttgacagagaaacttctcgtacagagttcatcttttactttgatcgggttgcta
acatgctgatccatctggcattggaacaggtagagttcggaccctcgcaagatgaggtat
tgaccccgcaataccattgcctaactgatgcgatacgaccgttacaatcggttgtcgttg
tgactatggtacggacaggtgatgtatttatgaattcaatcagaaaaactattccagatg
taagagttggtaagttgctaattcaatcagacctaattacaggcgaacctcaattgcata
caaagtcgctgcctccatgtgaacaaactaccaagctactattattcgatgcgcacatta
tatcgggggccgcagcaattatgggcattcaagtacttctggaccatggtattgaagaag
gtaatatcgtgattgtaagttatcttgcagaagaagctggcctacgtcgcatactgaacg
ctttccaaaaggttactattatcgtaggcttatcctctgggaggatgacctcattattga

aagagccaatgtttcgtacacggttcatcgacgattactacttcggcagtacgtagttca
gtgtgcgggatactgtattccgctcggggttctaaagaaattgtttaaactaaaccaaat
cggatcagaggttccgtacgtttttcacattcaaggatgagggttttccacgagtgaact
attactccggtctcccaccatcatttgcggaatgaaacctttgtgctgagattgtatag
ggcgtggggacggacgcttcttaaccgttcccctagaatgtcgtcccctgatcaaaattt
aatggcatccaactttgctgtaataggtatatataacctagcaggcgaccgttcatgtac
agtaaattgttttagacttttttttaactgaaatcaatcca

ORF (in bold stop codon previous gene)
ATGTCAACTTCTTCAAAACTTGGATGGCTATCCAAACTGGATGTATCATCCACTCCAGAG
CGCAATCTTCGCCGATCTTCCATTATTGGTACGATTGGTCCTAAGACTAACAGTCCAGAA
GTACTTGTCAGTCTGAGACAGGCTGGGCTTAATATTGTTAGAATGAACTTCTCTCATGGT
TCTTACGAGTACCATCAATCAGTTGTGGACAATGCCAGAAAATCAGAGGAAATATATCCT
GGTAGACCGCTTGCAATTGCTTTGGACACCAAGGGTCCTGAAATCAGAACAGGAACTACC
AAAGGTGAAACAGATTATGCAATTCCTATGGGACATGAAATGATATTCACCACAGATTTG
TCTTTCGCAAAGTCTAGCGATGACAAAGTGATGTTCATTGATTATAAGAACATTACAAAG
GTTATTGAACCTGGAAAGATTATTTATGTGGATGATGGTGTTCTTTCATTTGAAGTTTTA
GAAGTCGTTGATGAAAACACTCTCAAAGTCAGATCTATCAATGCTGGTGCAATTTCATCC
CATAAAGGTGTCAATTTACCTAACACAGACGTGGACTTGCCAGCTCTTAGTGAAAAAGAC
AAGCAAGATCTGAGGTTTGGTGTGAAAAACAAAGTCAACATGGTATTTGCATCCTTCATT
CGTTGTGCTAATGATATCAAGGAGATTCGCCATGTTCTCGGAGAAGATGGCAAACAGATC
CAGATTATTGCTAAGATTGAGAACCAGCAAGGAGTTAACAATTTTGATGAAATCCTGGAA
GTCACTGATGGTGTCATGGTCGCTAGAGGAGATCTAGGTATTGAAATCCCTGCACCTCAG
GTATTTGTTGTTCAGAAGCAATTGATTGCCAAGTGCAATCTGGCAGGTAAACCCGTAATC
TGTGCTACCCAAATGTTAGAATCTATGACTTACAATCCTAGGCCAACAAGAGCCGAGGTT
TCGGACGTTGGAAATGCTATTTTGGATGGCGCCGATTGTGTGATGTTATCCGGGGAAACT
GCCAAGGGAAACTACCCTCATGAGGCTGTTGCTATGATGCACCACACTGCTCTAATTGCA
GAGTCAGCTATTGCTTACCTTCCACACTACAACGAAATCAAGGATCTTGCTCGTGGTCTT
ATTAACACAGTTGAAACTATTGCTATTGCCGCTGTTTCTGCTCACTTTGAACAAAATGCC
AAGGCCATTGTTGTGCTTTCTACTTCAGGAACTTCAGCAAGAATGATTTCTAAGTATAGA
CCGAATTGCCCAATCCTTATGGTAACCAGAAATGATGAGGCAGCAAGATATTCTCATCTC
TATCGTGGAGTATATCCATTCATCTATAAACAGGAAGTTAATGATAACTGGCAACAAGAT
GTTGAAGAACGTTTACAATATGCCATCACTGAAGCCATTGGCATGGGAATATTGAAAAAA
GGTGATGCTATCGTTGCAGTACAGGGGTGGACCAAGGGACTGGGTCACACCAATACTATG
CGCGTTGTTTTTGCTTAA

Downstream
gtagtgtcaatacatgcgccctttggactaggttggttttgagcatatttaatagctct
cctctaatctttcgccaatcactctcattgttagtatcattgtcaactaactcgcttgag
tttaaactacaggttatgatatctgctagtgacaacacttcataaatcttttcattcagg
attaatttgtgaatgaaaccgtaaaagacatctaggtcattctcatgaataccgattggc
ttctcgttttcattttcaaaactgaaccgaacccaatctttcaagagatttttccaatct
tcgaactcatagaatgtctgaatgcctaaattgtagtgcttgttgaacgtcttcgttttc
tcggaagcccagtaggagttcacaagggttttctgtccctccatagttaaaccacttagt
tcagtcttctggtgaattacagaatcccctaaattttcaattggaatattcaacgttagt
ttgtttaatgttctttggttactaattagatcaatcctcttgttctgggtttctggccgc
actttaccaatctttttgagagatctattttgaagtagacaattgcgattgcgattcttt
tttggacttttctcaaataactgctcaacactacgtctctctagttcaagtttaatttct
tgctgaaaactcgttaggcaaagctttgaaggcgtcaagatctacaagatgcttctgcaga
agtttttgtaaagtttcttttgctggagagtagtgagggaaaacagaaccacctttgagt
acattctgttcgatgccttgcacttttttttctttaggcttttggcttcgaattgtaat
tgcttttggtccacagctgaagttgcgtcactatcacacttttgcagttcactaatggtc
acagaaacacctctgagatcagcaattgcaaccccaattatgcggtataatgatttcaac
tctgtggctattagaccaacctctcttgttggtgctcctaa

AA
MSTSSKLGWLSKLDVSSTPERNLRRSSIIGTIGPKTNSPEVLVSLRQAGLNIVRMNFSHGSYEYHQSVVDNARKS
EEIYPGRPLAIALDTKGPEIRTGTTKGETDYAIPMGHEMIFTTDLSFAKSSDDKVMFIDYKNITKVIEPGKIIYV
DDGVLSFEVLEVVDENTLKVRSINAGAISSHKGVNLPNTDVDLPALSEKDKQDLRFGVKNKVNMVFASFIRCAND
IKEIRHVLGEDGKQIQIIAKIENQQGVNNFDEILEVTDGVMVARGDLGIEIPAPQVFVVQKQLIAKCNLAGKPVI

CATQMLESMTYNPRPTRAEVSDVGNAILDGADCVMLSGETAKGNYPHEAVAMMHHTALIAESAIAYLPHYNEIKD
LARGLINTVETIAIAAVSAHFEQNAKAIVVLSTSGTSARMISKYRPNCPILMVTRNDEAARYSHLYRGVYPFIYK
QEVNDNWQQDVEERLQYAITEAIGMGILKKGDAIVAVQGWTKGLGHTNTMRVVFA

SEQ0075

P. pastoris homolog of *Saccharomyces cerevisiae* **PDB1 (YBR221C)**

E1 beta subunit of the pyruvate dehydrogenase (PDH) complex, which is an evolutionarily-conserved multi-protein complex found in mitochondria

chromosome 1-4, 0593

E.C.1.2.4.1

5' region (in bold stop codon of next gene)

atgggaattcggaggatatgtcccatttggtggacttgttgctgttttctcctggagtta
agaactatagtttctacagagctactccaagtcaacagcagttacagcagcagcagcaat
ctcagcaggtgcaacaggcacagcaagttcaacaggctcaacaagctcagcaggcccagc
aggctcagcaggttcaacaacaggctcagcaggcccaacagcagcaggcacagcaacagc
aacaacagcaacaattgcagcagcagcacatccaacaattgcaacaaatacagcaactac
aacttgcgcaacagcagcagcaacagcaacagcagcgtcaacaacaggaacaccttcttc
agcaattgcagcagcagcagcagcaacaatatcaccaacctcaaacacaaggccaaaatt
tcccacaacagtacttgatgcaacaggcccaagctcaggcccaagctcaagctcaagttt
tagcccatgctcagcatgacaactcatccaatccaatgttcactaacatcagagaaggat
ctggggtctcagctacacctccacccccaggattgtttagtggagtgaccaatcacacta
cggaaactcataacactccatcatctgaactattgaaccagctgatgaacggggagggta
gaaactcgatcaatgct**tag**tcggtgattatctgttatataagaagtacctattagttga
ataaagtaataatattggatgtctgatgttccgaggcttccctagtccgagtcgattgcg
cgcgtaaattggtgcttttcccccaccgaaacaataatgaggggatctccatatcacgtg
atgcattcggtgtaacttttagtggtataaaccgcggtcggatgcactccgcctaacaaa
cttctgtgaggtgcgaaacaaggaacccgtaaaggaaaacctcattgattacctgttagt
tcctactttctcttttacccacaaggttcactctcaccaca

ORF

ATGTTGCCATTTCAAGCAGTAGCCAAAGCAGGCCTCAAGCCCCAATTGACCCGCATGAGT
GTAAACTCATTGAGAGCAGCTTCTTCCTCGGCGGGTCCAACAAAGTTGTCTGTCAGAGAC
GCTTTGAATTCAGCCATGGCCGAAGAATTGGACAGAGACCCTGAGGTGTTCTTGATCGGT
GAGGAGGTCGCACAGTACAACGGTGCTTACAAGGTTTCCAGAGGACTGCTAGACAAATAC
GGGCCCAAACGAATCGTTGATACCCCAATTACCGAAATGGGTTTCACTGGTCTTGCTGTG
GGTGCTTCGTTGGCAGGCTTGAAGCCAATCTGCGAATTCATGACATTTAACTTTGCCATG
CAGTCAATCGATCACATTATCAATTCCGCTGCCAAGACCCTCTACATGTCTGGTGGTAAG
CAACCCTGTAACATCACTTTCCGTGGTCCTAACGGAGCTGCTGCTGGTGTTGCAGCCCAA
CATTCCCAGGACTACTCTGCTTGGTACGGATCTATCCCAGGTCTGAAAGTTATCTCTCCC
TACTCTGCCGAGGACTATAAGGGTCTGTTCAAGAGCGCCATCAGAGACCCAAACCCTACC
ATCTTTTTGGAAAATGAACTGTTGTACAACGAAGAGTTCGAAGTTTCTCCTGAGGTTCTG
TCCCCTGATTTCACTGTTCCAATTGGTAAAGCCAAGATCGAGCGTGAAGGTACCGATATC
ACGATTGTATCCCACAGCAGAAATTTGCAGTTCTGTTTGGAGGCAGCCACCATTTTGAAG
GAAAAGTATGGTGTCTCATCTGAGGTTCTCAACCTTCGTTCCATCAAGCCATTGGATGTT
CCTGCCATTGTTGAATCTGTCAAGAAGACCAACCATCTGATAACTGTTGAAGCCGGTTTC
CCAGCCTTTGGTGTTGGTTCCGAGATTTGCGCTCAGGTCATGGAATCCGAAGCTTTTGAC
TACCTAGATGCCCCTGTTGAAAGAGTGACCGGATGTGAAGTTCCAACCCCCTACGCCAAG
GAATTAGAAGACTTTGCCTTCCCAGACACCCCAACTATTATAAGAGCTGTCGAGAAGGTT
CTTTCGTTGAAAGAGTGA

Downstream (in bold stop codon previous gene)

ttatagagttcaataatgcccatgatatattataaactaattttgaaaagatatacagtt
gg**cta**aatagacctatccaccctctgtttgactctttcaagactcatcttgacgtctccc
ttagaaaagttttctagattgtaataaacgggagccaattctatcaaccactcaggtcta
attgatgttactgttcggatgtagttcttggtcgtcaagacgaattcattatagattacc
cactcgttctcctttgcaagaacggtacttgggtggattaaaacgtcttggttatctttc
actgtaatgtagcctttggccccagatcttttctttgcgacttgcatgaaaaatccagac
gtaagggctttccggatgttatcgtagtatcttggatcttcaaagtcagttgatacgagg
tgcaagtcatacctctccatcaacctttctaattgccttctaacgttttcggcgctcttt
agggatctataggagaggaaatgctctttacaccaagggttcaggccctttcgtaggct

tcctcagactggaaggcatgataaacgttcaaaagagttagatgatcaccatctgagtgg
gcaaacaaactcttcatttcatcggctttcttcttgttatttggtggacgcacaaatacc
gagggaactgacaacatagccacaatggtaagaatctcgttcgaacaattgaactctgaa
gaaccaatcaacataactgccaacattggatcaagagggaattgtgatgctaatcttccc
agagctgtcaacgaaccatcgtcgtctaaacaagccagatagttcagctcttccaaagca
cgcatcatggtttctggtgctggagggtccataaaatcgaagtgaaccaaatcatcaatt
cctagctttttaactctaacacagttgaagccaggttgcttctcaggatttctggataa
ctctgttcaattagttctttctggaatgcctcctcggtgta

AA

MLPFQAVAKAGLKPQLTRMSVNSLRAASSSAGPTKLSVRDALNSAMAEELDRDPEVFLIGEEVAQYNGAYKVSRG
LLDKYGPKRIVDTPITEMGFTGLAVGASLAGLKPICEFMTFNFAMQSIDHIINSAAKTLYMSGGKQPCNITFRGP
NGAAAGVAAQHSQDYSAWYGSIPGLKVISPYSAEDYKGLFKSAIRDPNPTIFLENELLYNEEFEVSPEVLSPDFT
VPIGKAKIEREGTDITIVSHSRNLQFCLEAATILKEKYGVSSEVLNLRSIKPLDVPAIVESVKKTNHLITVEAGF
PAFGVGSEICAQVMESEAFDYLDAPVERVTGCEVPTPYAKELEDFAFPDTPTIIRAVEKVLSLKE

SEQ0076

P. pastoris homolog of *Saccharomyces cerevisiae* **ARO10 (YDR380W)**

Phenylpyruvate decarboxylase, catalyzes decarboxylation of phenylpyruvate to phenylacetaldehyde, which is the first specific step in the Ehrlich pathway
chromosome 4, 0314
E.C.4.1.1.1

5' region (in bold start codon previous gene)

Agcatctgaagggaagtactctgtagcggagaaaga**cat**atggacttttgtgttttattt
Ttctgacgctattagtttctttgcagtaagaagtcagcactgttgtttcaaacttcagct
Tttcaaatggggagaatgaaggggggatatggaaaagaaggcaggttgtgaaggataaa
Gcagcagaactcacagtatactctcaattctagaaatcagataaaacttcaaactgaaac
Gtgtagctctccgactttacgtgcaaggaagagatgagggtaataaatttaaaatcggag
Ttgttgacgtctttaacccaatcgggctcgatacactatctggaaacgcggtttagcctg
Atactcgtgtggtggaaggggttattttaattttgcaggactaatcccaatgcggaaatt
Tcacctagactgcgtttgtttcaccgaaaacaaccttgatttgaaatttgaaaggttgtg
Gagttcgaaggagttgccttgcatgaaataaattggaagcatagttgcagcccaataggg
Atgtttgatgatacagacatcaatggtttgcaaacttatatgactagatgttggatggcc
Tgtatatggtgcctctacttttgcttactagactaagttagtgggttggttagagtgta
Catggtgtcagaacgatctatcagtttacagcagacactaagtaagggctaaggagcgtt
Tacctaaggcgtcaaagaacggaaggttattctaggaaggatctgtgtcctgtctaagct
Ttagattacacacggatatctattatgtgcaagaagtttccgccataaacctcgaataac
Cgaataaccgctcatatccgaaagtaaccgcacatttccgctatgggggggagaacattca
Cctagcttatccgtaacaagatcatgaggcattagcaaaatcttcatgtgtactgtggta
tatatagacctgtattatccttccgtaaagtcattcaaaccaatttcaag

ORF

ATGGCCCCAGTTAAACAAGACTTTAACATAGACGTTCAAACAATCGAGAACACTGACATC
TCTTTGTCTGAATACATATATCTTAGGATAGCCCAATTGGGTGTCAAGTCGATCTTTGGT
GTCCCAGGTGACTTCAACCTGAATCTTGTGGATGAACTAGACAAAGTTCCCCAATTGAAA
TGGATAGGATGTTGTAATGAGCTAAATGCCACTTATGCTGCTGACGGCTATGCAAAAGCA
TCAGGAACGATAGGCGTTGTGGTCACTACTTATGGTGTGGGTGAGCTAAGCGCCATCAAC
GGTATAGCAGGAGCATTCGCCGAATATGCTCCTGTTCTTCACATTGTAGGCACCTCCGCT
ATGGCAACAAAACGACTTGAACATGTGCACAACATTCATCATCTTGCAGGGTCTAAGAAT
TTCTTGGATAGACCAGACCATTATATATACGAAAAAATGGTTGATGACATCTGCATTGTC
AAAGAGTCTTTAAGCGAGATCGAAAATGCCTGTGGCCAGATAGATAATGCCATCGTACAG
ACTTATCTGCTTTCCAGACCAGGATACCTATTTTTACCTAGAAATATGGCCACTATGAAA
GTTCCAAGAGAAAGGCTATTCAACCAACCATTAGCCTTGGAAAGAGTTGATCTTCACCCC
GGTGAAACGCTACAAGTTGTTGAGAAGATCTTGGAAAAGTTCTATCATGCAAAAGAACCG
GCCTTGATTGTGGATTATTTAACCAGACCCTTTAGAATGATGGAAAACTGCTCCAAGTTG
ATCGGTGCTTTGGAGAATAAAGTCAACATTTTCAGCAGACCAATGTCCAAAGGCTTTGTC
GATGAGAGCCACCCAAGATATATTGGCTGTTATATTGGAAAACAATCCAAGCACCCAAGT
ACTAGTGATATTCTAGAAAAAAAAGAGCGATTTCATTCTGAGTGTGGGAACGTTTGATGTT
GAAACGAATAACGGAGGCTTCACATCCAAGCTTCCCCAAGAACATTTGGTGGAATTGAAC
CCTCATTTTACTCGTGTTGGAACTCAATGCTTTTCAAATGTTAACATGTGCCATGTCCTC

CCACTCCTGGCAAGTAAGCTGAGAGGTGATTTGATCAGCATGGCCACAGTTCATCCAAAT
GATTTCAGCCTCAGAAAGAAAGAAAAGGCACAGGATAAAATGAAGGCTCTCAACCAGAGC
CATCTTGTAAAATCTACAGAGCTACTCCTAAATGCAAACGACACTTTGATAGTTGAAACT
TGCTCATTTATGTTTGCAGTTCCAGATATCGCGTTCCCTAACAATACACAATTTATCAGT
CAGTCATTCTACAACTCCATAGGATACGCACTTCCCGCCACCTTGGGTGTCAGCATTGCA
AAGCGTGATTTCAGAAAACCAGGAAAAGTCGTACTTATTCAAGGGGATGGCTCTGCTCAA
ATGACCATTCAAGAACTGGCTACCATGGTCAGACAAAAGGTCAAACCCACCATTTTACTC
CTCAACAACGAAGGATACACAGTTGAACGAATGATTCTAGGTCCAACCAAAGAATATAAC
GATATAGCTCCTAATTGGGATTGGACTGGAATGCTGAGAGCATTCGGCGATATACGTGGC
CATTCCAAGAGCATCTCAATTGACACATGTGGTAGATTAGATAAGCTCGTTCAAACTCGA
GAGTTTCAAGAACCTACACACCTAAATTTTGTGGAACTTATCTTGGGCAGAATGGACGCC
CCGGAAAGGTTTGCCAATATGGTAAAAGAAATTGCTAACTTAGAGCACGCAAGTAAAAGT
ATTCATTAG

Downstream (in bold start codon next gene)

taattaaaataagactacagatcacgtgccgtaaagggtaaaacctggccgagggccagc
cttatcatgggtgtttttttatcttgtcaaatgcgtattatatttgctcctcgctcgcccg
tttggcttctcctctctaatctccacgtgacccaaaacctgacgtctaatagtaccccac
atttgacccccctgatgttgagttgagtcatccatcacagaaccaaccagctaagagtttg
gcttcccggaaaaatttcccgtccaccgtctccgtctcgtaaactcttctttaaagggtg
gtgtttatcgtcgattgactcagtgattagcctttgccctttttctctcgatagggaaaa
acctcgcgaccaaaaaaagctatcgatccctctttccttaatctctccttttttttttatt
gatagccccagtgtttactagcaaactctgactatttataaacttgaaacagaaacaca**a**
tgaagcaaacgtattacgatccaaaggtagctgagccaattggatccgagtcaatggatg
atagtaagctggacgacatggttcaatcgttgcccccagttgacccttgttagatactc
ccttcccgttcaggagtgcagacgtggaacaaggtgacacgtttgccacagagaacaacg
ctgggcctacgcagtccgcaaggactaacgtgaacttcttgcccccacatcaatcattgt
tcgatgccttacagagctctaacctgaacatggttgaggagattgacccttgagacgaa
gcaaccaaccggaagaatgttccccttcttgttggaaccaggagttctccgtgactgcgt
ggcaaatggtccaccgagacgatgtgaagttgcaaaatctttggacccacacaaaaatgt
acggaacaggcagtagacgtgttccaggtacggggaacggtgaatcagtgctgtccagct
cagagtacgaggagaccgaccttccagatacccctctcctc

AA

MAPVKQDFNIDVQTIENTDISLSEYIYLRIAQLGVKSIFGVPGDFNLNLVDELDKVPQLKWIGCCNELNATYAAD
GYAKASGTIGVVVTTYGVGELSAINGIAGAFAEYAPVLHIVGTSAMATKRLEHVHNIHHLAGSKNFLDRPDHYIY
EKMVDDICIVKESLSEIENACGQIDNAIVQTYLLSRPGYLFLPRNMATMKVPRERLFNQPLALERVDLHPGETLQ
VVEKILEKFYHAKEPALIVDYLTRPFRMMENCSKLIGALENKVNIFSRPMSKGFVDESHPRYIGCYIGKQSKHPS
TSDILEKKSDFILSVGTFDVETNNGGFTSKLPQEHLVELNPHFTRVGTQCFSNVNMCHVLPLLASKLRGDLISMA
TVHPNDFSLRKKEKAQDKMKALNQSHLVKSTELLLNANDTLIVETCSFMFAVPDIAFPNNTQFISQSFYNSIGYA
LPATLGVSIAKRDFRKPGKVVLIQGDGSAQMTIQELATMVRQKVKPTILLLNNEGYTVERMILGPTKEYNDIAPN
WDWTGMLRAFGDIRGHSKSISIDTCGRLDKLVQTREFQEPTHLNFVELILGRMDAPERFANMVKEIANLEHASKS
IH

SEQ0077

P. pastoris homolog of *Saccharomyces cerevisiae* **PCK1 (YKR097W)**

Phosphoenolpyruvate carboxykinase, key enzyme in gluconeogenesis, catalyzes early reaction in carbohydrate biosynthesis, glucose represses transcription and accelerates mRNA degradation, regulated by Mcm1p and Cat8p, located in the cytosol
chromosome fragB, 0061
E.C.4.1.1.49

5' region (in bold stop codon previous gene)

gctacatcggaaccaattttgaattcgccgatacgaaaccggacatcaagttccatcatc
atgcaagatgatcatcaggaagagccgtggtcagtcaacactgcaagtgtcaaagagaca
tcaattattgacattagtaccccggatgaggcttacaatgctcgcgggaaagacgatgcg
atacgggttgattcggagggcccttcaaaaagaacaactattctcagccgtctgatggag
acacaagattcggacaatgacggaacggagtctgacgtaggggagagttctagtctgata
attcttagt**tga**ccagccttgcatcgggttgaacaaacttttgtgggttgggccactttg
cgccaacaacacaaggttttcacaccacccggttgtcattaccgcccggaaaactcaatg
attgacggagcgtaggttcacggaagtcaagttagagtagtcggagagtttcgggtgata

agaatttcccgggtgctttggagtgccaaagtgactgaatcaggaactaaaaacccgggc
taacggctgaaggcctgtgttcagtgcactatattgtattctagtctggggtaggttgaa
gtttggtggattcacgactccggagtgatggagggacaataccgaagatgaggtcttgcg
agtggaaagataggttaagaataattaataacattgtatgaaacaagatggggagacaat
tttagctatcgatccgttatcattggtcgaatgatgacttggaaccttatttgcactttt
tttgtcaggtgttcgccgaaacttaaagatagttttattgatgtgtgccaaaaatgtggg
gacaaaaggtcctccacccgccatcccccggtctcatacaaaaaaaaagaaaaatacctc
tactagtccggggtcaaggcaagtcatgataggcatataaataggggcaggctggccgcc
attacgaattagattaaccttctcatagattattatccaca

ORF

ATGGCTCCTACTGCTATAGATTTACAACAAGTGTCTGATGAAGACTACCAAACCAAAATA
AAAAAGGCTACCTCGCCTTCATATTCTGTCATTGCCTCTCTGGGTTTAAACCCTGAGGCA
GTGATCCGCCACAACGCTGCCGTTCCTACTCTTTACGAGGACGGTTTGTTGGAGAAAGGA
ACTGCCATCAGTTCTACTGGTGCTTTGATGGCCTACTCCGGTAAGAAGACCGGAAGATCT
CCAAAGGACAAGAGAATTGTCGACGAGGAGACCTCCACCGAACACATCTGGTGGGGACCT
GTTAACAAGAAAGTTGACGAGCTGACCTGGAAGATTTCCAAGGCCAGAGCCATCGACTAT
CTGAGAACAAGAGAAAAGATCTACGTCATTGACGCTTTCGCTGGTTGGGATCCTCGTTAC
AGAATCAAGGTTCGTGTTGTTTGTGCCAGAGCTTACCACGCTCTGTTCATGAAGAACATG
TTGATTCGCCCTACTGAGGAAGAGTTGCAGAACTTTGGTGAACCGGACTTCACCATCTGG
AATGCAGGCCAGTTCCCAGCCAACGTCCACACCAAAGGTATGACCTCTTCCACATCCGTC
GAGATCAACTTCAAATCCATGGAGATGGTCATTTTGGGTACTGAGTACGCCGGTGAGATG
AAGAAGGGTATTTTCACTGTTATGTTCTACCTGATGCCAATTCGTCACAAGGTCTTGACT
TTGCACTCCAGTGCCAACCAGGGTGTCCAGAACAAGGACGTCACTTTGTTCTTCGGTCTT
TCTGGTACTGGTAAGACCACTCTGTCTGCTGACCCTAACAGAAAGCTGATCGGTGATGAT
GAGCACTGTTGGTCAGACAACGGTGTTTTCAACATCGAGGGTGGTTGTTACGCCAAGTGT
CTGGACCTGAGTGGTGAGAAGGAGCCTGAAATTTTCAATGCTATCAAGTTCGGTTCCGTT
TTGGAGAACGTTGGCTACGACCCAGTTACCAAGATTGTTGATTACTATGACAAGTCTATT
ACTGAGAACACCAGATGTGCTTACCCAATTGACTACATTCCATCTGCTCTTATCCCATGC
ACTGTGGACTCCCATCCTAAGAACATTGTGTTGTTAACTTGTGATGCTAGAGGTGTTTTG
CCTCCAATTTCCAAGCTAACCAACGCTCAAGTCATGTACCACTTCATTTCCGGTTACACC
TCTAAGATGGCCGGTACTGAAGACGGTATCACAGAGCCAGAGGCTACCTTCTCTGCTTGT
TTCGGTCAACCATTCTTGGTTCTGCACCCTATGAAGTATGCTCAACAACTGTCAGACAAG
ATGATTCAGCATGAATCTACCGCTTGGTTGTTGAACACTGGATGGACTGGATCTTCTGCC
ACCAAGGGTGGTAGCCGTTGCCCACTGAAGTACACCAGAAGAATTCTGGACAGCATCCAC
AGTGGTGAATTAGCCAAGGCTGAGTACAAGAACTACCCAACCTTCGGTCTACAAATTCCA
AAGGAAGTTCCAGGTGTTCCTACCGAGATTTTGGATCCTTCTAACAACTGGGTTGACGGT
AAAGACAACTTCAACACCGAGGTTAGAAACCTGGCCCAACTTTTCATGGAGAACTTTGAG
AAGTACGCAAGTGAGTGCACTACTGAGGTTATCCAGGCCGGCCCTCAGTTGTAA

Downstream (in bold stop codon next gene)

ttgcaacaggataatattatttgaactttgaagtgtatatatacatatttaactttgctc
attgagagattaacttattaactatgactggttgtagaggagagagtttctctgagttgt
ttcgtatacctattatttttcttttgatagt**cta**tcgtcttccttcggacacttgagcgct
caccacactttctgcgtcactgtctataacgaggtcgtcctcttcgtcctcactaatctc
attccaggacccgtagatcaatagaccaaatccaagtacgacgaccatgtcaccaaagat
ctgtgtccagctcagctgaataccgaagagtagccaatccgtgattcccacaaacaggat
tgtcactagggacgaaaccgaagagagaacaggagaagttaaagaagtaagggctaaaaa
tgaacaactgaataaaaggtttgataacactgacaagccaaccaggagccaaatggtccc
actgtagtccagattgaacttatggagtccggtgaaatgtgcaaataccagaggagccca
gaagagaccaaaattgaatattccaatcactgacatagcaaagtttgacatagatgcctg
tcttcgagaggagattaagcttgatggacagcagaggtacttcttgtagagaacctcgta
aaggccatagagtacggcacccactgatatgatgacatttcccacgaacctatatgactt
gtcatttccagattgatcgtcatcggaaggcacagcagggttgtaagtgacaataaagac
tccgaaaactgcaattattacgcttgaaagtttcaaaaaggaaaacttttcacctagaag
cggaattgcaaaggcataggcagtaaatgcggaacaattgtaaatggccgtaacgtcgtt
acttgtgctcaaggtcagggcaaaataccacgatgatccagccacattcaatatcagtgt
aatagggatactggtcagcaagatatactttatggaattgg

AA

MAPTAIDLQQVSDEDYQTKIKKATSPSYSVIASLGLNPEAVIRHNAAVPTLYEDGLLEKGTAISSTGALMAYSGK
KTGRSPKDKRIVDEETSTEHIWWGPVNKKVDELTWKISKARAIDYLRTREKIYVIDAFAGWDPRYRIKVRVVCAR
AYHALFMKNMLIRPTEEELQNFGEPDFTIWNAGQFPANVHTKGMTSSTSVEINFKSMEMVILGTEYAGEMKKGIF
TVMFYLMPIRHKVLTLHSSANQGVQNKDVTLFFGLSGTGKTTLSADPNRKLIGDDEHCWSDNGVFNIEGGCYAKC
LDLSGEKEPEIFNAIKFGSVLENVGYDPVTKIVDYYDKSITENTRCAYPIDYIPSALIPCTVDSHPKNIVLLTCD
ARGVLPPISKLTNAQVMYHFISGYTSKMAGTEDGITEPEATFSACFGQPFLVLHPMKYAQQLSDKMIQHESTAWL
LNTGWTGSSATKGGSRCPLKYTRRILDSIHSGELAKAEYKNYPTFGLQIPKEVPGVPTEILDPSNNWVDGKDNFN
TEVRNLAQLFMENFEKYASECTTEVIQAGPQL

SEQ0078

P. pastoris homolog of *Saccharomyces cerevisiae* **PDX1 (YGR193C)**

Dihydrolipoamide dehydrogenase (E3)-binding protein (E3BP) of the mitochondrial pyruvate dehydrogenase (PDH) complex, plays a structural role in the complex by binding and positioning E3 to the dihydrolipoamide acetyltransferase (E2) core
chromosome 1-4, 0254
E.C.2.3.1.12

5' region (in bold stop codon next gene)

gattcttgtggaagcatatggctctggttttacggatgaagcattcacaaagctagagaa
agaacttgagtacatttcagcccaagtggaaagtaataaaagttgtcccacgtttgaaaa
acatgcattctatagaagtcgacttttgatgttggaactatcatgcctgagggctcaaac
agttttacttctatatcgtccaaaattgactactggaaagtccctccgggctgtcaatgt
ggcaaagtcaataatcctcgatatatggagtcaatatacgaaacagtatcccagcaacaa
aaaagaaatgctggaacacctagattggaatttttgctatcccttgcgaactgcgacact
gacgttgtggatctcgtgtgctatcctccttgaatataaacacatgattgaattcatgag
aaccagtgatccatttgaatacattctggcattagaagtattggaaggactagttcaaat
acttccaattgagcaacagcttttagatctgctaaaatctcccaacgattcggttatagg
gacggacggctgtacaggcaatttttggacgactatgtctgatcaaaagctgcaaaataa
g**tag**gattcttcacttgtccggctatgtgatggtcattattgagggttttataagacgta
tcagtttagaaagtgaatctttaaactttaaagtcgagattggtaaatgatacttaatca
caaacttcaaaaatgccccacttttctttgcaaatcaatttctaggctgagcttcgttc
aacgttaacgagtagtctagtagccttccgagttgaagatcatagctcacaattcataac
gtcaaactcgaacctctaacttaaagtaagcttccaccaaatcgtgcacccactacaaat
cgcgacttgttccctccacttcggggtatcttatacaatcatcgaagccgctaaaacaat
agttggcgaaattgaacattaaatcaatctctttaaggcaa

ORF

ATGTTGAAACAAATAAGCCGTCAGTTCAAACAATTTCCTGTGAGCCGCAGATGTCTTCAC
GGATCTTCTTTCAGATTGGGAGCTACAGTGTTTGACATGCCAGCAATGTCCCCTACCATG
GAGAAGGGAGGTGTTGTCTCCTGGAAAATCAAAGAGGGAGAAAAATTTAGTGGTGGCGAT
GTGCTGTTGGAAGTTGAGACAGACAAAGCCCAGATTGAGGTAGAGGCACAAGACGATGGT
GTGCTCGCAAAGATACTGGTACCAGCGGGTACCAATGACATCCCAGTGGGAAAACCAATT
GCCTTTCTGGCCGAGCAGGATGACGACTTGAGTACTCTAGAGTATCCTAAACTGGAAGAA
ACTGCCTCTAAGAAGATAGAATCAAAACCAGAGAAAGCAGAGGAGAAAATTGAACCACCG
CAACCAAAGGAGGAGAAGAATACCAGCGGCAGTGACAGTGGTAAACTGGGTAATCCGAAA
CAGACTTTATTACCCTCGGTGGAACTGCTCCTACACGAAAATGGATTGAGCAAAGAAGAT
GCTCTTGCCAAAATTCAGGCAACAGGACCCTCCGGAAGAATTTTGAAGGGTGACGTATTG
GCATATCTGGGCAAAATTTCACCCGAATTGAATGTCAAAATTTCCGAGTACATTAATAAC
AAGTCTCACTTAGATCTGTCTAAGATACAAATCAGGGAAACCAAGGCAAGCGAACAATCT
CCATCATCTTCAGGAGATAAACCAGCCAAACCTGAAAAACCTGCTAAGAAAGAGCCTCTA
AAGATCGAGAAGGAATTGACACTATCTCAAGCTTACAGTAAGGAGGCCCTGCAAAGATTA
TTAACTATTGCAGAACAACATGCATACAGTGCCAAGCTGTACCAAGAGGATTCAGAAGTG
ATTGATCCACTGTTTGAGGAGATCGTTGCTCCACCTAGAAATGCCGAACGATTCAAAACA
CAATTTATCGTCACGCCCTCAGACAATAGTGTTTCCACATTGAAGCTTACCTTATTAGTC
AACGAGAGTATATTGGATGCCAAGCAAAGAGCTCAGTTATTCCTTGATGAAGTCAAGGAT
CAATTGACGCAAGATAATGGTGGCGTTTCGTCGTCTCCCCAACTTGAAGAGTTATTCTGA

Downstream (in bold start codon previous gene)

gttgcaacaagtctaagtagtaagtaattaaaccatcatgatcctatgatcgtgatcatt
cattaaagcacggtgtggcaattattgctagggagatcgtcactgtatggtggcagaatt
atctctacaagatgtctcaaagtccccacaaagcttggaccctctcatctgtaatgcatt

ttcctgtaactccccttagccacacgtcaagggctctgaatccgttgaaaagctgtggcg
tctgccacctttaacgtcttcatgagggatgtgcacgtgatattgtctttcccttctcta
aagcttcgaaaaaaacgcatctcaatgcgagaagcagatcgatatatataaagaactagt
ccattgaaagatctctcaatttcactggaaaccaactcagaaagaa**atg**ccttctcctca
cggtggtgtgctacaagaccttattaagcgtgacgcttctatcaaggaagatttgttgaa
ggaagtccctcagcttcaaagtattgtgctaactggtagacaactctgtgatttagagtt
aatcctaaatggaggtttcagtcctttgacaggatttctgaccgagaaggattatcgctc
cgttgttgacgatttgagactcgccagtggtgatgtttggtctattccaatcaccctgga
cgtcagcaagaccgaggctagtaagttccgtgtcggcgaaagagtggtgttgagagatct
tcgtaacgacaatgctctgagtattctgaccatcgaggatatatacgaacctgataagaa
cgttgaggctaagaaagtcttccgcggtgatccagaacacccagctgtcaagtacctctt
tgatgttgccggtgatgtgtatattggtggcgctttgcaagctctacaattgcctactca
ttacgactacaccgccctgagaaaaacgccagcccaattgaggtctgagtttgagagccg
taattgggaccgtgttgtcgctttccaaacccgtaacccaa

AA

MLKQISRQFKQFPVSRRCLHGSSFRLGATVFDMPAMSPTMEKGGVVSWKIKEGEKFSGGDVLLEVETDKAQIEVE
AQDDGVLAKILVPAGTNDIPVGKPIAFLAEQDDDLSTLEYPKLEETASKKIESKPEKAEEKIEPPQPKEEKNTSG
SDSGKLGNPKQTLLPSVELLLHENGLSKEDALAKIQATGPSGRILKGDVLAYLGKISPELNVKISEYINNKSHLD
LSKIQIRETKASEQSPSSSGDKPAKPEKPAKKEPLKIEKELTLSQAYSKEALQRLLTIAEQHAYSAKLYQEDSEV
IDPLFEEIVAPPRNAERFKTQFIVTPSDNSVSTLKLTLLVNESILDAKQRAQLFLDEVKDQLTQDNGGVSSSPQL
EELF

SEQ0079

P. pastoris homolog of *Saccharomyces cerevisiae* **ACS2 (YLR153C)**

Acetyl-coA synthetase isoform which, along with Acs1p, is the nuclear source of acetyl-coA for histone acetylation; mutants affect global transcription; required for growth on glucose; expressed under anaerobic conditions
chromosome 3, 0403
E.C.6.2.1.1

5' region

aacttgaaggagcatgtaaacctttctttattcgcagagtagtttttatgaaagggact
caatgttccaagatacgattgatacaaatttgcgtaatgtaatatcttctttaactgtga
ttaaatcttgcacatccagaatatctcttatcgtgtttctttgataaacgtcacttatat
atatggttggcatttattcactgtctgtgaacgtatatctggcaactaggaattaccatt
caacattcaaacattcagaggattttaggacgctgttaataagagtttaatcacctttcc
gagctgaaggtttccactagaccctcatttgacatagattgcaactacaaagttaccctg
ccaagcatgtcgattgactatcttatggagtctacacagttcttgttgaacgttgcttgc
caagcttgatgtacaagactaattgacttagatttggctcctatcaatgcaataaacgtt
agcaatgctattcgatgttcccgtataattaatactaggctagccgcggttagagatgcg
ccatcgctcataatgtctggtgttaggcttttgaaaatcacatgctcattaaaaagcgcg
caaatagtaccatactgtgcgcagcccgagaactgaacatgatcaacctcgtcgataagg
aagtgctttacagaatcgatcgccgttaggccgtgagccggtatcacggtatcagcttgt
atcgagatacaactgtacctatcaaagtgtcgtccgcccatgcgaccttgcagtacatta
tatatataagctggggtagattccagatacaaaaactctgtgcgattgttgcaactacgt
aaataaatacaagaaatctctcaaggtgcgttttatcccatactggcatacgaatatgca
aattaggtccacgttgagccggtcatcatcttttccccatttccatgcatatgacaattt
ggtgaggagggacgtgtaaaccgttttgaatattctgacatgctataatggcccgctttt
aggaattttgttgcatgaggtcgttcgggaacgacacccccccgccaatattttcgaggc
actggtcatttttttaacccaaatactaactatcatagta

ORF

ATGACTTTTCCAGAGCCAAGAGAACACAAAGTGGTGCACGAAGCCAACGGCGTAAGGGCT
ATCAAAACCCCTCAATCATTTTATGACAAGCAACCTGTTAAGTCATTGGAGGCATTGGAA
CATTATCAAGAGCTGTACCAGAAGTCCATCGAGGACCCAGAGGAATTCTTCGGCCAAATG
GCAAAGCAGTTTCTAGATTGGGACAAAGACTTTGGTAAGGTCTCCTCTGGATCTTTGAAA
GAAGGTGATGCTGCGTGGTTCCTTGGTGGAGAGCTGAATGCTTCGTACAACTGTGTTGAC
CGACATGCTTTTTCGCACCCTGATCGTCCCGCCGTAATTTTCGAAGCGGACGAGGAATCT
GAATCTCGAACAATAACTTATGCAGAACTTCTACGTGAGGTCTCTCGTGTTGCAGGAGTA

CTGCAGAGCTGGGGTGTACGCAAAGGTGACACTGTCGCAATCTACTTGCCCATGACTACC
GAGGCCATTGTGGCCATGCTGGCAGTGGCACGTCTGGGTGCAGTGCACTCCGTTATCTTT
TCTGGATTTTCGTCAGGATCTATCCGGGACAGAGTTAACGATGCTGGATCTAAGGCAATT
ATTACCTGTGATGAGGGACGCCGTGGGGGTCGTATTGTGAACAATAAGAAAATTGTCGAT
GCCGCTGTTGACAGCTGCCCCACAGTGGAAAAAATCCTGGTTTATAAGAGGACTGGTAAC
CCAGAAATCAAGATGGTAGAAGGAAGAGACTTCTGGTGGCAGGAAGAGGTTGAGAAATTC
CCTGGTTACATTGCCCCTGTCCCTGTAAACTCGGAGGACCCACTATTTCTTTTGTATACT
TCGGGATCTACTGGTTCTCCCAAAGGTGTGGTACACTCCACAGGTGGTTATTTGCTGGGA
GCAGCATTGACAACTCGTTATGTGTTTGATGTCCAGGATGAGGATATTATATTTACTGCT
GGTGACGTCGGATGGATTACTGGTCACACATACTCGTTGTATGGACCACTTGTTCTGGGT
GTTCCAACCATTGTTTTTGAGGGAACTCCTGTCTACCCTGACTACGGAAGATTGTGGAAG
ATTTGCGCCAAACATAAAGCCACACACTTTTACATCGCTCCTACTGCTCTTCGTCTTTTG
AAAAAGGCTGGTGAAGAAGAAATTAAAAAGTACGACTTGTCTAGACTTCGTACTTTAGGA
TCTGTTGGTGAACCAATTGCCCCCGAATTGTGGGAGTGGTACAATGAGAAAATCGGAAAC
GGAAACTGTCATATTGCTGATACTTACTGGCAGACTGAATCTGGTTCTCATTTGATTGCT
CCATTAGCAGGTGCCGTTCCCCAAAAGCCGGGTGCAGCTACTGTTCCTTTCTTTGGTATT
GATGCTTGTATCATTGACCCTGTTTCTGGTAAGGAACTTGAAGGCAACGATGTGGAAGGT
GTTTTAGCTGTCAAGTCCACTTGGCCATCAATGGCTCGTACAGTCTGGAGAAACCACGCT
AAATACCTCGACACATATATGCGTCCTTATCCAGGCTACTACTTTACTGGCGATGGTGCC
GGTAGAGATCACGATGGTTATTACTGGATCCGTGGTCGTGTTGACGATGTTGTCAATGTA
TCTGGCCACCGTTTATCCACTTCTGAAATTGAAAGTGCTTTACTGGAAAATGGCAAAGTT
GCTGAAGCTGCTGTGATTGGTATTTCCGATGAGCTAACTGGTCAAGCTGTTATTGCTTTT
GTCGCCTTGAAAGATGCCACTGACTCTGAGAATTTAGACGCTCTCAGACGTGCCTTAGTC
TTGCATGTTCGTGGAGAAATTGGTCCATTTGCAGCTCCTAAGTCCGTGATTGTGGTTGAT
GACTTGCCTAAGACCCGATCAGGTAAGATCATGCGTAGAGTTTTAAGAAAGATTTCTTGC
CATGAAGCTGATCAATTGGGTGATATGTCTACTTTGGCCAATCCTGAATCGGTAGACTCT
ATAATCGGAGCTGTTGATAACCAGTTCTTCAAGAAGTAG

Downstream (in bold stop codon previous gene)
aatttatgcatttttggatttatacatctttatatttaatagtacctttctctcccagc
taactatttacatggcaccgtcagcgcctat**cta**atcgttcttattctttgacaaggttc
tcattttctgagcagctttccccaccgccttcagtgtttctgtttgaagtagttgttgtg
acgattctaaaacactgatatcttctaatttttcatcccgggcctgccgaattacttccc
gtaggtcagaaatacgctttgtcaactcttgcctgtcagagtcgtccatatatccacccc
attgagataaagcattatcacagtcaaacagtagtagttccaaattgttggcttttgtgt
atatttccttcctaatacgatcagtttctgaattactctttgattcggctacaattcttt
caacctcttcagaagtcaaacctgtggatccagatacagttattgaagcgtcttgattag
tggtcgtctctttcgcatttaccttgatgattccatccgcatcaatttcgaaactgacct
tgatctgaggaacacccttgggttgaggttccaagtctgacagtttgaactgtccaatca
tcttattgtctctcaccaaagatctttcaccctgatatacgttaacttcgattgaattct
ggccatctacagcagttgagtacatctgttgtactttgcaaggcactttggtgtttcttg
ggatcaaaggagaaaaaatacctccaaaggtttcgataccgagagtcaaaggtgttacat
ctaaaagcaagacatctttaatctccccagataaaactgcaccttgaattgcagcaccaa
gagcgacggcttcatcgggatttacagaagtgtttggcttcttgccaaatgtttcttcca
caactttgcgaatcttgggcatacgagtcattccaccaacaaggattacttcatcaatgt
cggaatttgtaagatctgcatcacgaagacatttttaact

AA
MTFPEPREHKVVHEANGVRAIKTPQSFYDKQPVKSLEALEHYQELYQKSIEDPEEFFGQMAKQFLDWDKDFGKVS
SGSLKEGDAAWFLGGELNASYNCVDRHAFSHPDRPAVIFEADEESESRTITYAELLREVSRVAGVLQSWGVRKGD
TVAIYLPMTTEAIVAMLAVARLGAVHSVIFSGFSSGSIRDRVNDAGSKAIITCDEGRRGGRIVNNKKIVDAAVDS
CPTVEKILVYKRTGNPEIKMVEGRDFWWQEEVEKFPGYIAPVPVNSEDPLFLLYTSGSTGSPKGVVHSTGGYLLG
AALTTRYVFDVQDEDIIFTAGDVGWITGHTYSLYGPLVLGVPTIVFEGTPVYPDYGRLWKICAKHKATHFYIAPT
ALRLLKKAGEEEIKKYDLSRLRTLGSVGEPIAPELWEWYNEKIGNGNCHIADTYWQTESGSHLIAPLAGAVPQKP
GAATVPFFGIDACIIDPVSGKELEGNDVEGVLAVKSTWPSMARTVWRNHAKYLDTYMRPYPGYYFTGDGAGRDHD
GYYWIRGRVDDVVNVSGHRLSTSEIESALLENGKVAEAAVIGISDELTGQAVIAFVALKDATDSENLDALRRALV
LHVRGEIGPFAAPKSVIVVDDLPKTRSGKIMRRVLRKISCHEADQLGDMSTLANPESVDSIIGAVDNQFFKK

SEQ0080
P. pastoris homolog of *Saccharomyces cerevisiae* **ASC1 (YAL054C)**

Acetyl-coA synthetase isoform which, along with Acs2p, is the nuclear source of acetyl-coA for histone acetlyation; expressed during growth on nonfermentable carbon sources and under aerobic conditions
chromosome 2-1, 0767
E.C.6.2.1.1

5' region

ctggctcagcggttatgttggttagggtctaagttttgaccaatttgcttgtatgtgcgt
taaaaataggacagttgggggacaatagcgcgactggatttggtagacagggcttggttg
aaggttgacaattgaccaaccaagcaataccggtagtgggcaaatgaaaaaagaccacca
tccgccatccgacagcacactaccgtcatcccctcttccgtaagaccaaaacatcattct
ctcttcccctcctaggttggtgattcgcgacccattcctcaactccagagttgttcattt
tacaatggggaagtgggatatgatttcgttcctcaattgacttatttgtcatgctttttc
ggagcttgaagcatatcggtctctgtaccctgagcctttcgcattagtcggagtgcggat
atgttatgggtcagggtctgtatcttgaattattgtaaaattgaaagaaaggatctaagg
attgtttttccctttgaatgtgtatcttaagaaattgagtttcaccagcgtgttgtttac
taccagcgtctctcttctgtgtgatggccatactatactacataccactttactccattt
ccgataataagtttcacttcatcgtctttacctccttgactccgtgcggctccaactcta
aatctattagatgatcatgaggggcagcatttctcgtaatatcccacagagatacgtcaa
atcccattgttccgggtatatgccctcagtaaagacatggcgtccagcatcttgtactca
catctcgtttgagcgggtgacgggaggggtatatgcatctatttttcttgagtcttgact
ctgaagtaggtctcgtacaatatgtgacggagagcattcgtgttataaaaaggtaagaat
ggtcccccatttttcttctttctttctttctttctactatctttataaagtttttttttac
aaaggatacgacttagttgttgatcaatt

ORF

ATGCCATTAGATAACGAACACTTACTTCATGAAAATTCCATTGACCCACCAAAGGGATTC
TTTGAAAGACACCCTGGAACTCCTAATATACCAGGCGGTTGGGAAGAATACTTGAAGCTG
TACAATCAGTCCATCGAGAACCCCTCAAAGTTTTTTGGAGAAAAAGCAAAGGAATTCTTG
TCATGGGCTACTCCTTTCACTGACGCTCGTTACCCACCTGGTAATGGATTTCAGAATGGT
GACTCCGCCGCTTGGTTTCTGAATGGTGAGTTGAACGCGTCGTACAACTGTGTTGATAGA
CATGCTTTAAAGAATCCAGACAAACCTGCCATTATTTATGAGGCTGATGAACCTAATCAA
GGCCGTACGGTTACCTATGGAGAGTTGCTGAAGGATGTTTGTCGAATTGCCCAAGTATTG
ACTGACCTGGGTGTGAAAAAGGGTGACACTGTTGCTGTTTACCTGCCTATGGTTCCAGAA
GCTATCACCACTTTATTGGCTATCGTTAGAATCGGTGCTATCCACTCTGTTGTCTTCGCA
GGTTTTTCAGCTGGTTCTCTACGTGATCGTATATTGGATGCTGATTCTAGAATTGTTATC
ACTTCTGATGAATCTCTGAGAGGTGGGAAGATCATCGAGACTAAGAAGATTGTTGACGAG
GCTCTGAAGTCTTGCCCAGATGTTCGTAATGTGCTGGTCTTCAAAAGAACAGGTACACCA
CATCTTCCATGGGTTGAGGGTCGTGATCTTTGGTGGCACGAGGAAATCATTAAGCATGTT
CCGTACTCTCCCCCAGTGAATGTTAGATCTGAAGATACTTCATTTTTGCTTTACACTTCT
GGCTCTACCGGAAAGCCTAAAGGTATCCAGCATTCAACTGCTGGCTACTTACTGGGAGCT
CTTTTGACCACCAAGTATGTCTTTGATGTTCAGGGTGATGATATTTTATTCACTGCTGGT
GATGTGGGCTGGATCACAGGGCATTCTTATGTAGTTTACGGTCCACTTTTAAACGGGGCT
ACGACAGTTGTTTTTGAGGGCACCCCAGCTTACCCAGACTATTCACGTTATTGGGATATC
GTTGACAAACACAAAGTTACTCAGTTTTATGTAGCACCAACTGCTCTTAGGTTGCTGAAG
AGAGCTGGTAGCAAGTATGTCCAGAATCATGATTTGTCTTCAATCAGGGTTTTGGGTTCC
GTTGGTGAACCTATAGCCGCTGAAGTTTGGGAATGGTACAACGAGTATGTTGGAAGAGGA
AAAGCTCATATTTGTGATACGTATTGGCAAACAGAGACTGGTTCTCACATTATTGCTCCA
ATAGCTGGTGTGTCAAAGACCAAACCAGGTTCAGCATCTTTCCCCTTCTTCGGTATTGAT
CCGGTTATTCTAGATGCTACTACTGGAGAGGAACTCAAAGGTAATAATGTTGAAGGTGTT
TTGGCTATCAGAAATCCATGGCCATCTATGGCTAGAACAGTCTGGAAGGACTACAACCGT
TTCCTGGATACATATCTCAGGCCATATGAAGGTTATTACTTCACTGGTGATGGAGCTGCC
AGAGATCAGGAAGGATTTTATTGGGTTCTGGGTAGAGTTGATGATGTTGTTAATGTGTCA
GGTCACAGATTGTCTACTGCCGAGATTGAAAGCGCTCTAATCGAACACAATTTGGTAGGA
GAGTCTGCTGTCGTCGGATTCCCTGACGAGCTGACTGGTTCTGCTGTGGCCGCGTTTGTG
TCTTTGAAGAAGGACGTCGACAATCCAGCGGAAGTGAAAAAGGAGTTAATCCTTACTGTC
AGAAAAGAGATTGGACCATTCGCTGCACCTAAACTCATCATCTTGGTAAGTGATCTTCCA
AAGACCAGATCAGGTAAGATAATGAGACGTATTCTCAGAAAGGTTTTGGCTGGAGAGGAA
GACTCTCTGGGCGACATTTCAACTCTTTCAAACCCTTCGATTGTGGAAGAGATAATCTCT
ACCGTTAAAAGGGATGCCCGCAAATGA

Downstream (in bold start codon next gene)

gcatctgattaggacttacacttctatttcacatatctattttattatcgtgaagcagta
gtcaaatatgtagtatagttttattagtttaccttatcgctactattgctattttgggag
gaagcctcgaagacaccatacctaccaccttgggagaacttttttctcacacatagatagt
tcagtatttggtaaatgtgaactaggtttaccttcgtttcaacaaatcagaaaacgttgc
agtttcatagtgggactagtagcattttgtatgaggggtgatcatttatcaaagctattt
cacagcctaga**atg**aataagaattcaaattcacaacccaagtcaaacgagcaagaaatcg
gagattcaacatttgaagcctttaatggaacattttcctcaacacaattctctaatgata
atcttgttgaagatcccagttttgatgacttggccaaagatgagtttatggaatgtgacg
acatttcgccaccaagaaagacacctttttttaaacgatacaaaatcgcaaggcacattag
acgcatccctcactgaattagtcgaccaacaacctgatcataaagaaacaaaggaactga
aatttggagattatgaaacctatttctatcacaagaaattaaagcagcaggagaaagatg
atcagtatgtacagtggcgcacaaagaattcaaaccagtttcctcccatttttttctaatt
gtgtgatctatgttaacggaaggacagatccagatataacccagctccatagaatgatta
ttttgcatggaggagtattcttgaattatcttggtgctaagtctaatgccactcatgtga
ttgcctccaacattactccaagaaagataatagagttcaaaaacttcaaagtaattaaac
cagaatggattactgatagtataaaggagaagaagattctggattggaccaaatacactc
tattcagtccacagtatgatcaaaaggttttggaatcagat

AA

MPLDNEHLLHENSIDPPKGFFERHPGTPNIPGGWEEYLKLYNQSIENPSKFFGEKAKEFLSWATPFTDARYPPGN
GFQNGDSAAWFLNGELNASYNCVDRHALKNPDKPAIIYEADEPNQGRTVTYGELLKDVCRIAQVLTDLGVKKGDT
VAVYLPMVPEAITTLLAIVRIGAIHSVVFAGFSAGSLRDRILDADSRIVITSDESLRGGKIIETKKIVDEALKSC
PDVRNVLVFKRTGTPHLPWVEGRDLWWHEEIIKHVPYSPPVNVRSEDTSFLLYTSGSTGKPKGIQHSTAGYLLGA
LLTTKYVFDVQGDDILFTAGDVGWITGHSYVVYGPLLNGATTVVFEGTPAYPDYSRYWDIVDKHKVTQFYVAPTA
LRLLKRAGSKYVQNHDLSSIRVLGSVGEPIAAEVWEWYNEYVGRGKAHICDTYWQTETGSHIIAPIAGVSKTKPG
SASFPFFGIDPVILDATTGEELKGNNVEGVLAIRNPWPSMARTVWKDYNRFLDTYLRPYEGYYFTGDGAARDQEG
FYWVLGRVDDVVNVSGHRLSTAEIESALIEHNLVGESAVVGFPDELTGSAVAAFVSLKKDVDNPAEVKKELILTV
RKEIGPFAAPKLIILVSDLPKTRSGKIMRRILRKVLAGEEDSLGDISTLSNPSIVEEIISTVKRDARK

SEQ0081

P. pastoris homolog of *Saccharomyces cerevisiae* **LPD1 (YFL018C)**

Dihydrolipoamide dehydrogenase, the lipoamide dehydrogenase component (E3) of the pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase multi-enzyme complexes
chromosome 2-2, 0048
E.C.1.8.1.4

5' region (in bold start codon previous gene)

cgatagtaatggtttcctgactagctgcagatgtggccagagagtgaacactctcatcaa
ctgagttatcctttatgctcctattcaaggagatatcagtagatttttcagc**cat**ttagt
tagagccttcaaaagaatggcagaacattcacctatttttataccctgtgataagagaag
cttgaaaaaagcttgttcatagtcatctccacagttgacaggcgaccactgaaatggtgg
taccttgcgtcatcggatcgaatttttgaggatttcgttttaaaagtttggatcttaccc
gttattaccataggcatgggatttgaaaacgtttcaatgtggagtggcgggtgacgcaat
gtgtggtaagctttttcggcaatggcgaaattgggctcctgctaacttaaaaatactgag
tctaggcacaatatcttgtttgatgcctcaaaaatggcttcagcaaaccccgaatctcct
cccttaacgcttgaactgtatctggccctataatgtctcaattccatctaggcataagtt
cctgaaggcttttgtttcttgcacttaccgaagaaatagagttgatttcgggcaaaacga
atatccactcaaacatgagatacacttgaattgggctttcgttcccaatttcgtggaagt
tcaagtttagatggaataattatgacatctgtggatggcctaatataggcatgtccaaaa
agcttaattgtcgagatagtgcatgaagtttggataaacgtcccaaaggtcaactctcct
ctccctgccccccactaaagcatataccccccactaccccgggagatcccggtgtccgatgt
ctacaacaatactgcgatcgaccaatgtcacaacatcattacaccatatcggagaacccg
gactcgacctgccaacttccagctcgttctttctttgttagtccaaccagtcatagattt
agtagacctttttggaatataaagtaccaccaaagagacata

ORF

ATGCTCAAATTCGGACAAAGACGTTTTCTTAGCACTTCAAGAATCCTCGCTCAGGCCATC
AAACACGATGTCGTCGTGATTGGTGGTGGACCTGGTGGTTATGTCGCTGCCATTAAGGCT
GCACAATTGGGTCTAGATACTGCTTGCATTGAGAAAAGAGGGGCCCTTGGAGGAACCTGT
TTGAATGTTGGTTGTATCCCTTCTAAGTCTCTGTTGAACAACTCTCACCTTTATCACACC

ATCAAGCACGATACTAAAGAGAGAGGAATCAACGTTGCTGATGTTCAAATTGACATTGGC
CAATTGCAAAATGCCAAGGAGAAATCTGTGAAGCAATTGACTGGAGGTATCGAAATGCTG
TTTAAGAAAAACGGTGTCAAATATTACAAGGGTTCTGGTTCTTTCGTTGATGAACACACT
ATCAACGTTGATCCTGTTGAGGGAGGCGACAAAGTTGAACTCAAAGCTGATAATGTGATT
ATTGCTACTGGATCTGAACCTTCGCCATTCCCAGGAATTACTGTCGACGAGGAGCGTATC
GTGACTTCAACTGGAGCTTTGGACTTGAAGGAGGTTCCAAAGCGTATGGCTATTATTGGT
GGTGGTATCATTGGTTTGGAAATGGCTTCTGTTTGGTCTAGAGTAGGCTCTGAAGTCACT
ATCATTGAATACCGTGACAGTATTGGTGCTGGTATGGATGCCGAAGTTGCAAAGAGTACG
CAAAAGTTCTTGACCAAGCAAGGCTTGAAGTTCAAGTGTGGTGCCAAGGTCACTAAGGGT
GAACGCGTAGGCGAAGTTGTGAACATTGAGATTGAAACAACAAAGGATGGTAAAACTGAG
CAATTTGAGGCTGATGTCCTATTAGTTGCCGTTGGCCGTAGACCTTACACTGAAGGCTTG
AACGCCGAGGCAATTGGATTGGACTTTGACAACAGAGGAACTTTGGTCATCGACTCTGAA
TACCGTACCAAGCACCCTCATATTCGTGTAATTGGAGATGTTACTTTTGGTCCTATGTTG
GCTCACAAGGCTGAGGAAGAAGGTATTGCCGCCGCAGAGTTTATAAAGAAGGGCCATGGT
CACGTCAATTACGGAAACATTCCATCTGTTATGTACACTCACCCAGAAGTCGCTTGGGTT
GGTCAGAACGAACAACAACTTAAGGAGGCTGGTATCAAATACAAAGTTGGTAAATTCCCA
TTTATTGCTAACTCTAGAGCTAAGACCAACTTGGACACTGAAGGATTTGTTAAGTTCCTT
GCTGATGCCGAAACTCAACGTGTTCTGGGTGTTCACATCATCGGTCCCAACGCTGGTGAG
ATGATTGCTGAGGCTGGGCTTGCTTTAGAATACGGTGCTTCTACAGAAGATATCGCTCGT
GTCTGCCATGCTCATCCTACTTTGTCGGAGGCTTTCAAGGAAGCAGCTCTGGGTACCTTT
GACAAGACTATCAACTTTTAA

Downstream (in bold stop codon next gene)

gttttctaaataatgtacaatgaaatgtcttcaaacctatat**tta**ctctttgtcattttc
tactacttctgtgatatcccctattagtggctttaatgcctgtgggagcaactccgctgc
atctcccaggaacttccacccatcacacccttcaagatcttcttcatcagtattaaaaat
ggcaactttaccaccctgtagtttcactctgttcacatagcctgcggcgggccatacaga
ttgactggtgccgataactagtatgaggtccacctttgtgacataatgaattgatctgc
ttgatcaagcaccttaaagggtagggattctccaaaccagacaaccccaggacgtaataa
tccagtcttacattgagggcatgtgggaagcccacttctaggaatttccttaactggcgt
gaattctggagatccagttgtcgaaattacagcgtctttaccatcctcattctcattctc
gttctcattttcattctcagttcgtctctgtctttttctaggattgatccactcctcttc
acaaccttgcaattgaggggtcaaggggtgcttaaaattattattctctgtgtatgtgca
gtcaaaacttgtgcatttgagcgtgaacaagcttccatgcaactccaaaagctgttctgg
atcatggtgggcccttctagaaagaccgtccacattctgagtcaatgttaagaatttctt
acccttgaggctgccactttggaaagttgactcaaggcgtaatggccgttattcggttt
tgcctggatcgctttgtagcgtctataagagtagaattgccacactagccctggatcaac
gttgaaagcatctggggtagccagatccatcgaactgtaatttttccacaacccacctga
cccccggaaagttggtagtcctgaactcgccgaaagtccagcaccacaaagggccaatat
ggtacggcacttatctgtggttatgtattcatggaaactgg

AA

MLKFGQRRFLSTSRILAQAIKHDVVVIGGGPGGYVAAIKAAQLGLDTACIEKRGALGGTCLNVGCIPSKSLLNNS
HLYHTIKHDTKERGINVADVQIDIGQLQNAKEKSVKQLTGGIEMLFKKNGVKYYKGSGSFVDEHTINVDPVEGGD
KVELKADNVIIATGSEPSPFPGITVDEERIVTSTGALDLKEVPKRMAIIGGGIIGLEMASVWSRVGSEVTIIEYR
DSIGAGMDAEVAKSTQKFLTKQGLKFKCGAKVTKGERVGEVVNIEIETTKDGKTEQFEADVLLVAVGRRPYTEGL
NAEAIGLDFDNRGTLVIDSEYRTKHPHIRVIGDVTFGPMLAHKAEEEGIAAAEFIKKGHGHVNYGNIPSVMYTHP
EVAWVGQNEQQLKEAGIKYKVGKFPFIANSRAKTNLDTEGFVKFLADAETQRVLGVHIIGPNAGEMIAEAGLALE
YGASTEDIARVCHAHPTLSEAFKEAALGTFDKTINF

SEQ0082

P. pastoris homolog of *Saccharomyces cerevisiae* **ADH3 (YMR083W)**

Mitochondrial alcohol dehydrogenase isozyme III; involved in the shuttling of mitochondrial NADH to the cytosol under anaerobic conditions and ethanol production
chromosome 2-1, 0472
E.C.1.1.1.1

5' region

cgcagcgttttctgacggtactagaggactcttaggggaaggtagaatcaataaagatca
tattaggtaagcaaattttggatggaataggagactaggtgtggatgcgcgatctcgcca

aattgcacgaccagagtggatgccggatggtggtaaaccgtttcttccttttaccaccc
aagtgcgagtgaaacaccccatggctgctctccgattgcccctctacaggcataagggtg
tgactttgtgggcttgaattttacacccccctccaacttttctcgcatcaattgatcctgt
taccaatattgcatgcccggaggagacttgccccctaatttcgcggcgtcgtcccggatc
gcagggtgagactgtagagaccccacatagtgacaatgattatgtaagaagaggggggtg
attcggccggctatcgaactctaacaactaggggggtgaacaatgcccagcagtcctccc
cactctttgacaaatcagtatcaccgattaacaccccaaatcttattctcaacggtccct
catccttgcacccctctttggacaaatggcagttagcattggtgcactgactgactgccc
aaccttaaacccaaatttcttagaaggggcccatctagttagcgaggggtgaaaaattcc
tccatcggagatgtattgaccgtaagttgctgcttaaaaaaaatcagttcagatagcgag
actttttttgatttcgcaacgggagtgcctgttccattcgattgcaattctcacccccttct
gcccagtcctgccaattgcccatgaatctgctaatttcgttgattcccaccccccctttcc
aactccacaaattgtccaatctcgttttccatttgggagaatctgcatgtcgactacata
aagcgaccggtgtccgaaaagatctgtgtagttttcaacattttgtgctccccccgctgt
ttgaaaacgggggtgagcgctctccggggtgcgaattcgtgcccaattcctttcaccctg
cctattgtagacgtcaacccgcatctggtgcgaatatagcgcacccccaatgatcacacc
aacaattggtccacccctccccaatctctaatattcacaattcacctcactataaataacc
cctgtcctgctcccaaattctttttttccttcttccatcagctactagcttttatcttatt
tactttacgaaa

ORF

ATGTCTCCAACTATCCCAACTACACAAAAGGCTGTTATCTTCGAGACCAACGGCGGTCCC
CTAGAGTACAAGGACATTCCAGTCCCAAAGCCAAAGTCAAACGAACTTTTGATCAACGTT
AAGTACTCCGGTGTCTGTCACACTGATTTGCACGCCTGGAAGGGTGACTGGCCATTGGAC
AACAAGCTTCCTTTGGTTGGTGGTCACGAAGGTGCTGGTGTCGTTGTCGCTTACGGTGAG
AACGTCACTGGATGGGAGATCGGTGACTACGCTGGTATCAAATGGTTGAACGGTTCTTGT
TTGAACTGTGAGTACTGTATCCAAGGTGCTGAATCCAGTTGTGCCAAGGCTGACCTGTCT
GGTTTCACCCACGACGGATCTTTCCAGCAGTATGCTACTGCTGATGCCACCCAAGCCGCC
AGAATTCCAAAGGAGGCTGACTTGGCTGAAGTTGCCCCAATTCTGTGTGCTGGTATCACC
GTTTACAAGGCTCTTAAGACCGCTGACTTGCGTATTGGCCAATGGGTTGCCATTTCTGGT
GCTGGTGGAGGACTGGGTTCTCTTGCCGTTCAATACGCCAAGGCTCTGGGTTTGAGAGTT
TTGGGTATTGATGGTGGTGCCGACAAGGGTGAATTTGTCAAGTCCTTGGGTGCTGAGGTC
TTCGTCGACTTCACTAAGACTAAGGACGTCGTTGCTGAAGTCCAAAAGCTCACCAACGGT
GGTCCACACGGTGTTATTAACGTCTCCGTTTCCCCACATGCTATCAACCAATCTGTCCAA
TACGTTAGAACTTTGGGTAAGGTTGTTTTGGTTGGTCTGCCATCTGGTGCCGTTGTCAAC
TCTGACGTTTTCTGGCACGTTCTGAAGTCCATCGAGATCAAGGGATCTTACGTTGGAAAC
AGAGAGGACAGTGCCGAGGCCATCGACTTGTTCACCAGAGGTTTGGTCAAGGCTCCTATC
AAGATTATCGGTCTGTCTGAACTTGCTAAGGTCTACGAACAGATGGAGGCTGGTGCCATC
ATCGGTAGATACGTTGTGGACACTTCCAAATAA

Downstream

gccgaatagtttgtatacgtc**tta**tgtaatgagtttcaatgaattacttatttttacctc
tccttttggctcaattcaactagcctctgtagcaatctgtttgcgaagaagaacttatc
taatttttcatgggttttccccacgtttttgaaaagactttgtcttaactctctgtgatc
aaaccgttgtggtggaccgctgattgattgtgattcttcttccagatctagcgactctgg
cagataagagtccccggagtcaataattaccccaagaatataccttgaaacctccgagtt
ggccagtatttgagataacagcagctccttgtatttgtaagtgccttcgttatcaattga
tccgatgtccagcattattttgatatcaggtctcccaatcttgacaaacttttctacaat
gtgtttttgaagcagttgcacaagaggagcaatatcttcaaggtatgagtcttgctggtc
gccgttcaattttaaaataagaaatgtatctgagttccctgtcgtcccaataactgctac
actccccaactttgtgacagagagaaaatgatgctcttgtgaagtggaatatatggcgtc
aatgctactttgtaatttatggttgaacgcattctcagtatctccataaacttgaaagcc
cacggggtaagagacgccagatgctaattctctatgcaattgcgactctgtgtaagtgtg
actcaaaagtccaatacagtagagatcactcacatactgaggggtcagtgtgtcggagag
ttcacctaccaatggacagtactgggctaattcggttagcaagactcgacatatcgggat
cccatactgaatctcataggacattattgcctgatcattgtctccgtttccgtactcgct
caaattggttctcattgtcactagcaagtcttttcgatctgtagacgtcaaggggttgaa
taatggagttatctctttttccaaaacaccgtccatttctc

AA

MSPTIPTTQKAVIFETNGGPLEYKDIPVPKPKSNELLINVKYSGVCHTDLHAWKGDWPLD

NKLPLVGGHEGAGVVVAYGENVTGWEIGDYAGIKWLNGSCLNCEYCIQGAESSCAKADLS
GFTHDGSFQQYATADATQAARIPKEADLAEVAPILCAGITVYKALKTADLRIGQWVAISG
AGGGLGSLAVQYAKALGLRVLGIDGGADKGEFVKSLGAEVFVDFTKTKDVVAEVQKLTNG
GPHGVINVSVSPHAINQSVQYVRILGKVVLVGLPSGAVVNSDVFWHVLKSIEIKGSYVGN
REDSAEAIDLFTRGLVKAPIKIIGLSELAKVYEQMEAGAIIGRYVVDTSK

SEQ0083

P. pastoris homolog of *Saccharomyces cerevisiae* **ADH6 (YMR318C)**

NADPH-dependent medium chain alcohol dehydrogenase with broad substrate specificity; member of the cinnamyl family of alcohol dehydrogenases; may be involved in fusel alcohol synthesis or in aldehyde tolerance
chromosome 4, 0576
E.C.1.1.1.2

5' region (in bold start of next gene)

acttggggaacacagttgaaagaccgtcgatgactccctcaattcccacacttaaagctg
ggactaaaacttgccctacccctgcaaatggtgagaaattattatctctcaattccgtgc
ttttcccaactaaagtgtaaagagtaaaatcgtaatgggtcaacttacatccgacaatct
tttcatgcctagataattccttatagctgtcaaaagtgagctgaatcccattttgcactc
caggatagttatacagaataattggcaattctgctttatcggcaatggaattgaaccagc
tgataatccccttttgggataccaaacttgggccgtagtaccctggtacaagaatgactg
aatatgaagctccattctctttactaacggagatctgacgaataatatcctcaatacagt
agccaatcactccagcaataatggtaaattcggggacagccttatggagagtacgtatca
atgtcaaccgctcctcctctgttaaatgaatggcctctccaattgatcctccgactaaaa
ttcccttaatgccactatcatacaaatgtcttgcatgggccacttgagcctctagatcca
gagatttcttggcatcgttcttgaaaaatgtgggcagtggagaataaaccccgccagaaa
ggactttatatgacac**cat**cgtcgatggtttcaatggagggagcaagtacctttcatgtt
atatgatctggagctcctatgaaggtcagccttttgaccagtttcgagtgggataatata
cccgcagcaaccccactataagccatcccacgcaaattttccgattctagcacgaaaaag
cttggaccctactaataggatgggtagcaatcttaactcctgaagcaacgggtgtacttg
tcgatgttgcacccaccttgtgaaaaaaatccccacgcttcattacctatattaagcttg
aacatcaggtatatcaccagaaaactagtaaaaaccaaaca

ORF

ATGGCGTACCCAGACACCTTTGAAGGATTTGCCGTCACTGACACTGCAAAATGGTCCACA
ACCAAGAAGATAGAATTCACCCCAAAAAGGTTCCAGGAACATGATATCGATGTCAAGATC
CATGCCTGTGGTATCTGCGGGAGTGATGTTCACACTGTTTGCGGGGGATGGGCAAAACCA
GACCTTCCCGTGATCCCAGGACATGAGATCGTTGGTGAGGTTGTTAGAGTGGGCCCAAAA
GTGAAGGGATTTGAAATTGGGCAAAGAGTTGGTGTTGGAGCTCAAGTTTGGGCCTGTCTA
GAGTGCGACACATGCAAGGATAACAACGAAACGTACTGTCCTCAATGGGTGGACACTTAC
AATGCCACTTATCCTGATGGTGACAAGGCATGGGGTGGTTATTCCTCTCACATCAGAGTC
CACGATCACTTTGTATTCCCTATTCCTGATGAACTTCCAACTAATGCTGTGGCCCCAATG
TTGTGCGCTGGTATCACCACGTACTCTCCGTTGGTAAGAAATGGAGCTGGTCCAGGAAAG
AAGGTGGGTATCATCGGAATTGGAGGGTTGGGACATTTTGCCATCATGTGGGCTAGGGCT
CTTGGTTGCGAAGTGTACACGTTTTCTAGAACACATAGCAAGGAAGCTGATGCTAAGAAA
TTGGGAACTGACCATTTTATTGCGACGTGGGAGGACAAAGACTGGGCCAAGAAGATTGGC
AGAAAGCTGGACTTTATCATTTCGTGTGGAAATTCGGCCACGAACTTTGATATGGATGGT
TACCTCAGTGTGCTGAAGGTTCATGGTAAACTCATTTCCGTCGGCCTTCCAGAGGAGCCA
TTCACGCTGTCTGCTGGAAGCTTTATCAAGAACGGTTGCTACTTGGGATCGTCCCACTTG
GGGAACAGACAGGAGATGCTTGATATGCTGAAACTTGCTGCTGATAAGGGCATTGGTTCT
TGGTATGAGGAGCTCCCAATCTCTGAGGAAGGGCTGAAGGAAGGACTGGAGAGATGCCAC
AACAATGACGTTAAGTATAGGTTCACCCTGACCGGTTACGATAAGGCATTCAAATAG

Downstream

acctaccccatggactgaaacttcaaagtccagggtgaactatctttgactaaactgagc
atcttagttaacaaaaaatcgggcagagacccaagttagtcagcagacctaccccatgga
ctgaaacttcaaagtccagggtgaactatcttctactaacctgagcatctcagtcaataa
tacaaacatccgagcaggaggtgttcaattggttctcaaaatcacactctctagtcagaa
ggctaacatcgtgactgtcattctaagcgcatgcaacacttagccaagtcccaagatagt
aacgtatctaccacaatccactataagttagtaaaagaaacacagaaaacaggaagcaga
gtaaaatcagtgtcatgacttgatagtattccaggcacggtcatccagccaatcaattga

tccttccaggcttcacgacatctcatcctttagcaccgaggttcataccttcctgaaac
aaaatgtagcacccctcctacaagctcatgaaaggaagaacaagaccaaccatttgcgaa
aacttctatcttgcgaatgaatggttctgtgttcagaaagttgtcacaaacgctaaagga
ataatcgccttaccattggagcacggtttggagcagcacgagaagcaaacgtcttgcacc
atcaaaatagatgctagactcgttttacgccttggtggccgatatatacggcctatagat
atggccttcatagtggcaattccttgacagtgctgcccacatcgtgttaagaggaagaga
gcacaaatagagaacttcgctcctatgaccactcaatctggggtaagattacatcagcaa
ttagattccacatttccccctcccatctacacttatcgggaagcactaaaccgttgaacg
acttacacactctacaaagggcatatgcaaccaacgagcagacactacaacagacaatta
actccagacctcccattcactatctattatgcatcatctat

AA
MAYPDTFEGFAVTDTAKWSTTKKIEFTPKRFQEHDIDVKIHACGICGSDVHTVCGGWAKP
DLPVIPGHEIVGEVVRVGPKVKGFEIGQRVGVGAQVWACLECDTCKDNNETYCPQWVDTY
NATYPDGDKAWGGYSSHIRVHDHFVFPIPDELPTNAVAPMLCAGITTYSPLVRNGAGPGK
KVGIIGIGGLGHFAIMWARALGCEVYTFSRTHSKEADAKKLGTDHFIATWEDKDWAKKIG
RKLDFIISCGNSATNFDMDGYLSVLKVHGKLISVGLPEEPFTLSAGSFIKNGCYLGSSHL
GNRQEMLDMLKLAADKGIGSWYEELPISEEGLKEGLERCHNNDVKYRFTLTGYDKAFK

SEQ0084

P. pastoris homolog of *Saccharomyces cerevisiae* **ADH6 (YMR318C)**

NADPH-dependent medium chain alcohol dehydrogenase with broad substrate specificity; member of the cinnamyl family of alcohol dehydrogenases; may be involved in fusel alcohol synthesis or in aldehyde tolerance
chromosome 1-1, 0357
E.C.1.1.1.2

5' region (in bold start next gene)
ATCAAAAATTCACGGTAGAAACTGAAGAAAATCATGAAGTAAAACACAAAGCAGAAGTCT
TTCCATCCTTTGCCATACATGGGAGGGTTAGTTCCTTCGATTTTATATGAAAGATCAACG
AACTTGTGCAATGGATTGGATGGAGTGTGGTTATTAGATAACAGATACGAGCCATGTATG
GCAACCAAAGAAACCAAAAATGGCAACCATGTGTGACGCGAGCTCAGTTCGACCAAGGAA
AGTCCAATTTTTTTAACTAAACCCCAATCGGAACTTCCCTTCTTGTTGATTAAATTAATA
CGGGCGGAGCTAGACTTTCTCTGCTGAGCATTCGACTTTCGGGTAGTCAAAGAAGGAACT
GCATTGTCACCTAGGTTTATGTTTCCAATGGATGACGCTCTTCGACGTCTTGTCTTTTCC
TCTTTAGA**CAT**TTGTTGTGCTAACCTGGTATACCTGTGAGAGCTGCTTTACGGAAGCAGG
AAGCAGAATTCCAAAAACTTTGGGAAGACCCGCGCGAATTGACTCGAGAGTACTTCTACT
GCTTATGTAATGTAGATGTCAAATGTCTGGCAGGTGTAGATCCGATCGAGGAGTAAATAT
GTGAATGTTACTGTACATAAGGTCAAAATTTGAGTTGGAGAGGAGAGTAAACAAAGGCAA
GATTGCCAATGACCGCACTCAGAGTTGGTTATCTGCATAGTCAAAACTAGTCACCTTTGG
TTAGTCATTACATTTGAATACTTCAGAAATAATCGCCTTCCCCAGAATTTAATGCATCCC
CCATTCCCACCGAAATTCCACTAAACTCCCACCAGTTTTCAGATTCGGCTGTGAGTGCGC
ATCAGTTGCTCGGATAAACTCTTTCAAAACCATAACGTTCGCCCCTCCACCTCAACATCA
GTGGGGGCACTATTGGCTTCAAATATAATATATATAGATGGCGAAGTGAAATTCGTTTCT
AACCATCTTATTTTTAATTTTTTATTTCATCACTATAACAA

ORF
ATGGCTTACCCAGACACTTTTGAAGGATTTGCAGTTCACGACCCAAGTAAGTGGTCCGAAGTAAAG
AAAATTCAATTTACTCCTAGACCTTTCCAGGAAGATGATGTTGACATCAAGATCGAAGCATGTGGT
ATATGCTCCAGTGATATTCACACCATTTCCGGAGGGTGGGGTCAGCCAAAGCTTCCCTCCATTGTT
GGTCATGAGATTGTTGGAACTGTAGTTAGAGTTGGTTCAAAGGTCGACAATATCAAGGTTGGAGAC
ACTGTCGGTATGGGAGCAATGTGTTGGGCTGATTTGACGTGTGATGTCTGCAAGTCCAAAAACGAA
AACTACTGTCCTAACTGGATTGATACATATGATGACGCTTATCCCGATGGGTCAAGAACCTATGGT
GGTTACTCCAACTATGCTCGTTGCAACAAAGAGTTTGCATTCAAGATTCCAAAAGGATTATCCGTA
GAAGGAGTTGCTCCAATGTTGTGCGCTGGTATCACCACTTATTCTCCTTTAAAGAGAAACAATATT
GGACCAGGCAAGAAGGTGGGTGTTGTTGGTATCGGTGGTTTGGGGCATTTTGCTCTTCAATTCGCC
AAGGCTCTTGGAGCTGAAGTCTACGCAATTTCCAGGAATGACAAGAAGAAAGCAGACGCTCTTAAA
TTGGGAGCTGATTACTTCATTGAAACTGAGAAGGAAGGTTGGAACTTACCATACAAGTATAAGTTC
GACTTAATCATTAGTACAGCCAATTCCAGCCAGAACTTTGACCTTGATGCCTATGTATCCACACTT
AATATTGGGGCAAAGTTTGTATCTGTCGGACTTCCAGAGGACAAGATGGAAATGAATGCTGGATCT
TTCATCAAGAACGGCTGCTATTTCGGTTCTTCTCATTTGGGTAACCGTGAAGAAATGAAAGAGATG

CTTGAGTTAGCTGCTGAAAAGGGTATTGAAGCTTGGTATGAGCCAATTTCAATCTCTCAGGAGGGT
ATTAAAAACGGACTGGAGAAGTTGCACCGCAACGATGTCAAATACCGTTTCACCCTTACTAATTAC
GGAAAACAGTTTGAATAG

Downstream (in bold start codon previous gene)
ggtaaaatagatatgaatatataagttaggtgacgagatatgtaggaatggaaggtcgaa
ggattttattcaatgctaaaataccattttccgttgtatcccagcataattttcatgatg
gagtgaaactattcggtatttgatatctattatcctctctcctcggcttcgaagtttgga
agcatcgaaattgtttccacttgtgcggaggcgcccccacacttaatagtcatgcatgat
gtatcaattatgtactagatgcacccctgtgccccaccaaaaccgtgatgtaacccccac
tgtggagccatatggtcaggcgtcaaaataaacttcttctgttttcaactccactttcag
ttaatgataaatcgaaggtaatatcattattcggtggcctccccaatcatataaagctta
atgagtaagtattaagtttcatttcgctctccacttttacttgtaaacttttctctaac
atgacagttgactcttttgttttgccccgtggaatatacacccccatcccaacctacttc
aagaaggattactccttagacttagaaacccaagtagagcacgcaaagtttttacataat
gctggtatcaatgggctggtcgtcgctggatccatgggcgagtctacacatctgactaga
gatgaaagattgtcacttttccgtactttgagagaggctgtccctgatcccaatttcaag
atcattgctggagctccgtcaggacctgtggaagagattagagaggagatgaggttggca
aaggaagcaggtgctgatttttcaattatattggtcccgggatacttcggtccaaacttg
atcagccaagaggggataatagaatactttgaacttgtttccaaggactctccgttgcca
ataattatttataactacccaggtaactgtaatggtgtcgatatcaaaccggaaacattt
gaaagtttgggtaagcttcccgcaattgttgccgtcaagtt

AA
MAYPDTFEGFAVHDPSKWSEVKKIQFTPRPFQEDDVDIKIEACGICSSDIHTISGGWGQPKLPSIV
GHEIVGTVVRVGSKVDNIKVGDTVGMGAMCWADLTCDVCKSKNENYCPNWIDTYDDAYPDGSRTYG
GYSNYARCNKEFAFKIPKGLSVEGVAPMLCAGITTYSPLKRNNIGPGKKVGVVGIGGLGHFALQFA
KALGAEVYAISRNDKKKADALKLGADYFIETEKEGWNLPYKYKFDLIISTANSSQNFDLDAYVSTL
NIGAKFVSVGLPEDKMEMNAGSFIKNGCYFGSSHLGNREEMKEMLELAAEKGIEAWYEPISISQEG
IKNGLEKLHRNDVKYRFTLTNYGKQFE

SEQ0085

P. pastoris homolog of *Saccharomyces cerevisiae* **ALD2 (YMR170C)**

Cytoplasmic aldehyde dehydrogenase, involved in ethanol oxidation and beta-alanine biosynthesis; uses NAD+ as the preferred coenzyme; expression is stress induced and glucose repressed; very similar to Ald3p

chromosome 2-1, 0453
E.C.1.2.1.5

5' region (in bold stop codon next gene)
atttccaaaatggtaacaacgttcaagtaactttggatgttctgaaatatatttcaaaaa
aatatggaaccacggactactacgatgtggtcattggtattcaactcttaaacgaaccat
tgggacctattttagacatggataatctaagacagttctatgcggatggttatgatctag
ttagagatgttgggaacaactttgttgtaatccacgatgcattttaccaggcgccagagt
actgggggggacgatttcacctcagcggaaggttactggaacgtggtgcttgatcaccacc
attatcaagtcttcgatgcagatgaattgcaaagaagtatcgatgaacatatagaagccg
cctgtgattggggtagagatgcaaataaagagtaccactggaacctctgtggtgaatggt
cggcagcacttactgattgtactccttggttaaatggtgtcggaaaaggcacgagatatg
aaggtcaacttgataactcccccttggatcggatcttgtgagaatagccaggatccttcga
aattgagctctgaacgtatctgtgagtacagaaggtacgtagaagcccagctagatgctt
tcttacacgggaaaagcgcaggttttattttctggtgtttcaagacagaggccagtttgg
agtgggattttaaaaggttggttaatgccggtatcatgcctcagccattggacgacagac
agtatccaaatcaatgtgggttc**taa**gtttctccaatcaatgaaagcattaggatacttg
aaagcaattgacgattagttttctgtcaatggcccttgcagacccttgtcatatgcaccc
tttcctttcgttatgatatatagctgcattttatcgttcttttcttcattatatatgaaa
attaaaacttcttcaggtagagcgaagggttgtaccccgtagtaaagtattcgctgatca
ttttcatttgctcttatcttccccttcactttgaaaaaaaa

ORF

ATGGCAAGTCCCTTAAGCAAATCAATTGCCTTTCCTACCGGACAAAAATACGATCAGCCAATTGGC
CTCTATATCAATGGTGAATGGCGTGAGTCAAAAGACACCATTGATGTTATTAACCCATCCAATGGA
GAAGTGATTACCTCGGTCTACGCTGCCCAAGAGAGCGATGTTGACAATGCTGTTTCTAGTGCCCGA
AAAGCATTCAAAACTTGGAAAAAGCTAGCTGGAGAAGAGCGTGCAACCTTGATGAACCGTTTAGCA
GATCTTTTGGAGAAGAACGCGGAAACTGTAGCTGGAATTGAGGCCTTGGACGCAGGCAAACCCCAG
TTCTCTAACGCTCTCCCTGACATCGAAGGATCCGTTAGTATTTTGAGATACTGCGCAGGTTGGGCT
GACAAAATTTATGGCAATGTAATTCCGTCAGGCCCAGATAAACTGTTGACGTCGAAAAGGATTCCC
TACGGTGTTGTCAGCCAGATTGTTCCTTGGAACTATCCGTTGAATATGGCCATGTGGAAAATTGCT
CCCGCACTGTGTGCTGGAAACTGTATTGTCATCAAGTCCTCGGAATCTTCTCCACTTTCTTTGCTG
TACTTTGCTGAACTAGTAAATGACGCCGGGTTCCCACCTGGTGTGCTAAACATTATCAGTGGACTA
GGTTCAGTAGCTGGTGCCCGTATGGCTAGTCATCCAGATGTCGATAAAATAGCCTTTACTGGGTCC
ACAAAAACAGGCAAAGAGATTCAAAAATTGGCCTCCTCAAATTTAAAGACTGTTACTCTAGAATGT
GGAGGTAAGTCTCCGTTGATAGTCTTTGATGATGCCAAACTGGATCAGGCAATCTATTGGGCTGCG
TTTGGAATAATGTACAACACAGGTCAGATTTGCACGGCAAACTCTCGTGTCTTGGTGCAAGACACA
ATCTACGACGAGTTCATTCAAAAGTTCAAGGCTCATGTGCAGGAAAACTGGTTCATCGGGTCTCCA
TTTGATAAGAAATCCACCATGGGGCCTGTGATCAACAAATCTCAGTTTGAGAAGGTTAAAGGTTAT
ATCCAGAAGGGTAAGGATGAGGGTGCAAAGTTAGTTATTGGAGATGAACCAGTTACATTTGAGAGT
GGTTACTGGATCCATCCAACTATTTTTGTTGACTGTACTCAGGATATGTCGATTGTGAAGGACGAG
ATATTTGGGCCCGTAGTTGCCATCAGTAAATTCCACTCTCAGGAAGAGGCAATTGAGCTAGCCAAT
GATACGGAATACGGTTTGGCAGCTATGGTTTTTTCTAAGGATATTGTCACAGCCAATACCGTTGCC
AGCCAGTTGGAGGCTGGAACTGTGTACATCAATTCTTCAAATGATGATAATATCCGTGTTCCATTT
GGAGGGTTCAAGATGAGTGGTACTGGTAGCGAGTTGGGAATGGAAGGAGTGTTAGCGTACACTAAA
ATCCAGGCCATTCACACTAACCTGACCAGAGATTAA

Downstream (in bold stop codon previous gene)
agggaatagctctcctg**cta**tatgcctttcttcgttctgtatttttccacttctccatc
gtttcaatagcagccttccaattatgtgttacactttccagctcttttcgaagagattca
atctgtgtctccctttcatgtagttgttgtttgagaccgtttctggaggcaacttccgag
tctgcagatagttcattcagttcaaaagcatttcctaaagagttgtatagttcattcatt
tgggaagaaaactgtctgagaatattctctgattctgttttcgataattgtagagacttc
accaactcttcttttgtctcatgaactgtagaggcagacccaggaggggaatttatctcc
attgaagtcaattgggcatactcttctcttaggtcatacgatattttgccgtcgtagttt
gcaatctttgaaatgaaccttactaaattcttgatgaaattgttgtaaatagtcaagtac
ctggttttagagttattatcttgaatcagtgagtagattatatcatataaatcagacagc
aaagtcatggtttctttctgtgctacgggaaatctctccactggggttaacttttgagtc
tccgttgaagactcggctatttttgtctcttctttggttgatcatcaggcatgatagga
agactgtatagggagaatactttccctttaatacccaactgttttgcgactagcttttct
tggagttgatcaatctgaaggtctttcctcctgaattctgcattagatgattggagtaca
gtttgtagttgtgatttagtcttgttgtgttgtttaatgtttgtttgcactgaacttgtc
aattcatcgatctgtgctgttaactggttaatatgtctctcttgagaagttgacttggcc
tccaaagtgctaatacgcttcttcaagagagattggtcagttttcagggagtctatcact
tgtaaatgatctttttcccgtaagagaagatcgtctctgat

AA
MASPLSKSIAFPTGQKYDQPIGLYINGEWRESKDTIDVINPSNGEVITSVYAAQESDVDNAVSSAR
KAFKTWKKLAGEERATLMNRLADLLEKNAETVAGIEALDAGKPQFSNALPDIEGSVSILRYCAGWA
DKIYGNVIPSGPDKLLTSKRIPYGVVSQIVPWNYPLNMAMWKIAPALCAGNCIVIKSSESSPLSLL
YFAELVNDAGFPPGVLNIISGLGSVAGARMASHPDVDKIAFTGSTKTGKEIQKLASSNLKTVTLEC
GGKSPLIVFDDAKLDQAIYWAAFGIMYNTGQICTANSRVLVQDTIYDEFIQKFKAHVQENWFIGSP
FDKKSTMGPVINKSQFEKVKGYIQKGKDEGAKLVIGDEPVTFESGYWIHPTIFVDCTQDMSIVKDE
IFGPVVAISKFHSQEEAIELANDTEYGLAAMVFSKDIVTANTVASQLEAGTVYINSSNDDNIRVPF
GGFKMSGTGSELGMEGVLAYTKIQAIHTNLTRD

3 *P. pastoris* homologues of genes involved in homologous recombination

SEQ0086
P. pastoris homolog of *S. cerevisiae* **Rad50 (YNL250W)**

Subunit of MRX complex, with Mre11p and Xrs2p, involved in processing double-strand DNA breaks in vegetative cells, initiation of meiotic DSBs, telomere maintenance, and nonhomologous end joining
S. cerevisiae null mutant: UV resistance: decreased; chromosome/plasmid maintenance: abnormal; colony sectoring: increased; gamma ray resistance: decreased; resistance to cisplatin: decreased; resistance to N-Methyl-N'-Nitro-N-Nitrosoguanidine (MNNG): decreased; resistance to daunorubicin: decreased; resistance to 5-fluorouracil: decreased; resistance to menadione: decreased; resistance to camptothecin: decreased; resistance to tetrabutyl hydrogen peroxide: increased; resistance to chromium(6+): normal; resistance to hydroxyurea: decreased; resistance to chromium(6+): decreased; resistance to methyl methanesulfonate: decreased; resistance to ethylmethane sulfonate: decreased; spore germination: decreased; transposable element transposition: increased; UV resistance: decreased; competitive fitness: decreased; fermentative growth rate: decreased; resistance to doxorubicin: decreased; resistance to hydroxyurea: decreased; resistance to 5-fluorouracil: decreased; resistance to methyl methanesulfonate: decreased; resistance to cisplatin: decreased; resistance to L-1,4-dithiothreitol: decreased; sporulation: decreased; telomere length: decreased; transposable element transposition: increased; viable

S. cerevisiae reduction of function mutant: UV resistance: decreased; X ray resistance decreased; spore germination: absent

Chr 1-4, 0513

5' region (in bold start codon next gene)

tccaagtctgctttgttccaaggctgaagcttggcataatcgtgtttctcattagtgtcg
aattgagtggtgtcaaagtagacagatccatccttggtaacgtaggcataccctctatca
ataattccctgaacaaattcaacaatttcaggtacatactcggatactctagtcgtaaca
gtagctggtaaaatgttcaaacgttccatatcttcatcatatctctgttcccaataggaa
gaagtctgtttgaagatctctggatctgtcacggtagatccatacaatttatccagctcc
ggcacaacaacatccttgatattttctaaatattcattcaaatcagaacttgttttcagt
gccttttttgcagcagtgatatgcataggaattttgggattgtccactgccttttcagca
atgtttaaggattgagcccaactttcaaagtttgaagaatcacccttgaaatctgtcaga
ttcttagcaacatagctgttccagctccctaacgcaaactgcttcagctcattggttatc
tctgaatacttttcgaaaactctttgaataaatattcctgtctagctttcagtataatc
ttgtcgtcaatatcagtgacattctgaacaaacttgatattataaccaaaaaagtcctgc
aaaatacgacgattgatatcaatcgtgacgtagtttctagcgtgacccatgtgggaggaa
tcgtagactgtaggaccacaactgtaccacgtcacctcattggcgtttgtgggaataaaa
ggtacctttgtccttgtcaaagagttgaataatttaaggacgggccgttccacctttttgg
gtaggctgattccatgctggttgagacatttgtgtgtggaaatttcttattatataaaaa
gtgtttctcttccagggtctaatgtt**cat**ggaaacaaagtgagtgatcagacagtttgtt
ttttttagccggctggtctggcctttgggcggaggctttagatggccaactggaaggacc
tgcgcggaggaaaagttgttcatctctcccgttctcctaaacatatcgcgcctggatccc
cgtgtagttaattattcaccctcagggaaccaagtccggttactgaactgacctagcc

ORF (in bold intron)

AT**gtaagttttaacctcaattcctctattcaatgctctttgactctaacagaaag**GTCTTCAATTTATAAATTGG
CCATACAGGGAGTGCGGTCTTTTGACCCTCACACTCCGGAGACAATTCAGTTCAGCAAGCCTTTAACATTAATAG
TTGGACAGAACGGGTCTGGGAAGACAACCATCATCGAGTGCTTGAAGTATGCAACCACTGGAGATTTGCCGCCCA
ATTCTAAAGGAGGTGCCTTTGTTCACGATCCCAAGATCGCCGGAGATAACCAAGTTAAAGCGCAGGTCAAATTGG
CATTTCAGAACATAAGTGGAGTGAGTATGATTCTGACGAAAACTTTACAGGTCGTACAAAAACCCGGCAAGTCTT
TGCAATTTAAAACTTTAGAGAATCAATTGTCTGTGATAAACAACGGAGAAAGGACACGAGTGTCGCAAAAAGCCG
CGGAGATTGAATCGTTAATCCCTAATTACTTGGGTGTATCTAAAGCGGTGCTGAACTATGTCATTTTTTGCCATC
AGGAGGAAAATTTATGGCCAATTTCTGAACCAATGGCACTGAAAAAGAAATTTGATGAGATTTTTGATTCTGTCA
AGTTTATCAAGGTGTTGGAGGGGTTTAAAGGCATCACAAAAGATATGAGTGTCGATATCAAACTACTCACAAATA
ATGTGGAACATTTGAAAAATGACAAGAAAAGGGCCGACTTGAAGCGGGCGGAAATGGAGGCTCTGCAGAACACTG
TTGAGGATTACAACGCTGAAATTATTGATCTGAATACCCAAGTCGAGGAAGTCACTAGAAAGTTAGATGAGTTAT
TCAAATCAAATCAAGATTTTGAAAAAGTCTTATCAAAGTTAGATTTTCTGGCTACAGAAAAGCAGAGTACTCAAC

AACAGATTGATAGGCTTTTCAGTTCTCTGACAGTGCTACCTGATAGCACAGAAATCTTGGAGAATAACCTGTCGA
ATTACGGCAAACTTTTGACCGAAAAAAGAGTCAAGGTTAATGAGCAAATTCAACTCGCAAAAGAAGCCTCAGAAT
CTTTAAACGCTCTGAGAGACGAATACAGTGATACTATTAAAAAGGAAGGAGAACTTAGAGGTTTAGAAAGTACCT
ATAAAAATTATATTCAGGAAAGAATACAACTAATCCAAGAAAATGCGTCTCTACTTGGTATGACTATCTCCGAAA
CCACGATTAGTTCAGATGAGATTGAAGAAGCCTCCTCCAAGGCATCGGTCCTCTTTTCCACCTGTAAAAAGAAAC
TTGATCTACAAACAGAACACTATGACACCCGTATTCACGATCTTAACTTGGAAATAGGACAGGCCGAAAGTAAAC
TTGGAAAGGAAGAAGAAAGATCATCCTATTTGAAAAATGACATCAACAGTCTGAAGAAACGAAATCAGGCTCTTC
AAAAAAGCATCAATGATATCAATTCGAACGAGTCTGAGTTTAACGAAACTAAAGAAGATATCGAGAGGCTGACTA
AACAGCTTGAGGATCTTAGGTCAGAAAACAAACTGGCATCAATTAATAACGATCTGAAGCAAAATCAAGATAAAA
TTCTTGTTCTGGAAAACGAGTTGGACCAGATTAATAAACAAATAATAACTAGTAACCGACAAGGTGAAGTATTGG
CAAAGTTACACTTACTCAAGGAGAATACCAAAAAAGGAAACTCATCGATTTCAAAATTGGTTGAAAGCTATGGTG
AACAGTTTAAAGAGTTCACTGGTGAAGATCTTAACCCCGAAGACTGCTTACCTGTGTTTCTGGAGGTTCTGAAAA
AACGCCAGGAAGATACTGATTTAAAGAGGAAAGAAGTTGCATCATTCAAGCAGAACGAATACGAGAGTAATCATG
ATAGGTCTTTATTAGAAAAGAAACTGGAGCAAAGTAGAAGTCAGCTTCAGGAATGTCGTAGTCGAATTGTATCGA
TTCTTGAAGATGAACCAATAGAGGAGTACGAATCGATAGTCAAAGACTTGGAATCAGATTACGAGATTGCTTTGC
AAAATTCTAAATTGAATTGGGCCACCAAAAACTTTAACGAAACAGCACTAAAGATTGCCAAAGAGCATCAGTATT
GTATTTTGTGTAAGAGAGAGCTCAATCATGACGAACTTGGACCAGTAATGGTAACCATTTCAGAAAATATTGAAA
AAGCCAACGACGACCTTTATTCTAAAGAGAAAGATAGAATCAAGCAAGATCTGGATGACTTGAAGAGTATCAGGG
ATGATATTAGCAATTTTAGGAACTTAGAAGGTTCGATTGCTTCAGCTAATGACGAGTTGAATGAGTTGAAGACTT
TTGTTTCTGATGACTTGGCAAGAGTAAGTTCAGAACTCGAACAGCTTGAAAGTGATTTAGTCTCCCTGGAATCTC
TACGAAAACACGTAGTTGAAATTAGCAAACTTCAAGAGGATTTGAGTCATTATTATGTTGAAAACAAGTCCATTG
AATCTGAATTGAGTGCGTATGGTGTTCCAGCTAAGACTTTATCTGAACTTCAGGAGGATTTGAATGGTAAAAACT
TTCAATTGAAAGAATTGAGAAGGCAAGCTGATGATCTTAAAGAACAAAGAGAGTTTTCAAATCGAGAGTTAAGCA
TGCTTGAAGGCAATGTGAAGGATAAACGTCTATTGATTAGCAATTTTGAGAAATCGTTAATGATTAAACTTAATT
TGGAGAAGGGAATTGATGAGAATAAGGCTAGAATAGACGAACTTCAACAAACGGGAGAAACTGTACTTGAATCCA
TGAAACTGACCAAAGACAGATTGAATAAGCTTAAGGACAACGTGAAAACTCTTGAAGAGGAAAGAGATTCTACTT
TAACTCAATTGCAAAGCAACGTGGAACAGTTCAAAACAGTACAGGATAGCTTGGTGCGATTAAACTCTTTCGTCA
ACAAGTACGAGCTTGAAGATGAGCCAATTTTGAGGCAATGCGAAAAGAATAGTGAACACCTGAAAGCCTCCATCA
AAGATGCAGAGGGTCAGTTGAACAAACTTCATGAAAAAGTCAATGTGCTGGAAAAGCAGCTAAGCGAAGCAGATA
CCGAAGAACGAAACATTAAATTCAACCTGGATCTAAGAAGCCTAAACAAGGAGCTCGTTCAGATCGAAGAGAGTA
TTGAATCCCTTGATCGGCAGAACGCAACTAGTAAAAGGCAGGAGTTTCAAAAAGAAACCGCCATTCTTAGAGAAA
GTTATTCCCGGTTAAGTTCCGCTCATAGTAGTAAGATGGGGGAAGTGTCCCAATTGCAGAAACAGATTCAGAGCA
TAACAGAGGAAATCAAGAGGGATTATGAACATGTCGAAGATGAATACTACCAAGAATACTTGAAACTGCAAACTA
AAATGTTCATTTCAAATGATATCTCGGTTTATTCCAAGGGATTGGACAATGCTGTAATGAAGTATCATTCCATTA
AAATGGAGGAGATTAATAGAATTATTGACGAACTTTGGAAGAGAACGTACAGTGGGACAGATGTTGACACAATCA
TGATCAAAAGTGACATGAATACTCAAGTTAAAGGCAATAGATCTTACAATTACAGGGTAGTCATGATGAAAGAAG
ATGTGGAATTGGATATGAGAGGAAGATGTTCGGCAGGCCAGAAAGTACTCGCAAGCATTATCATCAGACTAGCAT
TGGCTGAATGTTTTGGAGTTGGGTGTGGCATGATAGCACTTGATGAACCAACGACAAACCTTGATGAAGAGAATA
TTGAAAGTCTTGCCAAGGCACTCAATTCAATCATTCATCTAAGAATGTCTCAAAAGAACTTTCAACTAATTGTCA
TTACTCACGATGAACACTTCTTAAGACACATGAATGCTACAGAGTTTTGTGATCATTTTTTTAGAATCAGCAGAA
ACGAGAGGCAAAAATCTCAGATATCAATGGTAAAGAACTTTGATTAA

Downstream (in bold stop codon previous gene)

agttgatcaagatttgctaaggaggatagagacggaatgtttcccattattcgacattca
agattcaacatttcaaatctcaaacttcctaatgggaaaataatttttctcagagacgat
catttaggtctaggaccagaaatggcaacgaacataacttggcaattaacattaattcgc
cgtggcaaatgtctcttgtacttttgattaaagagaagaagaaatattttgcgggagga
aaaatatatatatactactattcgcgattacaggagggtaatgcc**cta**taatctaatgtt
actcctcatgcccttaccatgatagatcaattgaagcttgaaagattgtaaaccaaactc
gacaacggcctgcgcgtcgaccttgagttcaccgttcacagctttccttaaaatatcagc
tcctattccggaagttgcatatcggataggaattcgaggagttgtgcttactcccggtgg
gatgataagactctgttctctcccaatatatcctaatacagtgccttcataagaggcttt
ggcgttctgtatcaagacaaatgcttcagagttagatattggattgtagactgtaaattc
tatctcggatgacatgagatgcatcgttgaatctaggataaaaccagattggtaatctgg
gtttgatgataggattttgtgttcttcgtagacaatttgtggaattggtaatgtcactga
taaatttttcaaaacttgagatagtaaaggtgaagtcggtatacttctatcgtggccttc
aatggttactaatgtttcctcgccagataggtatccgctgaccaattgttcgagtttctc
tattggggatacatctgatgatgggatattgtcaattgtgtatcgattcggatcaatgtc
aaaaagcgcagaagctgagaaccaggagcttcctgaactttcaatgttctcaaataagga
gaggtggcctagtgtcgattcattgtatgcaagttgtaaat

AA
MANWKDLRGGKVVHLSRSPKHIAPGSPSSIYKLAIQGVRSFDPHTPETIQFSKPLTLIVGQNGSGKTTIIECLKY
ATTGDLPPNSKGGAFVHDPKIAGDNQVKAQVKLAFQNISGVSMILTKTLQVVQKPGKSLQFKTLENQLSVINNGE
RTRVSQKAAEIESLIPNYLGVSKAVLNYVIFCHQEENLWPISEPMALKKKFDEIFDSVKFIKVLEGFKGITKDMS
VDIKLLTNNVEHLKNDKKRADLKRAEMEALQNTVEDYNAEIIDLNTQVEEVTRKLDELFKSNQDFEKVLSKLDFL
ATEKQSTQQQIDRLFSSLTVLPDSTEILENNLSNYGKLLTEKRVKVNEQIQLAKEASESLNALRDEYSDTIKKEG
ELRGLESTYKNYIQERIQLIQENASLLGMTISETTISSDEIEEASSKASVLFSTCKKKLDLQTEHYDTRIHDLNL
EIGQAESKLGKEEERSSYLKNDINSLKKRNQALQKSINDINSNESEFNETKEDIERLTKQLEDLRSENKLASINN
DLKQNQDKILVLENELDQINKQIITSNRQGEVLAKLHLLKENTKKGNSSISKLVESYGEQFKEFTGEDLNPEDCL
PVFLEVLKKRQEDTDLKRKEVASFKQNEYESNHDRSLLEKKLEQSRSQLQECRSRIVSILEDEPIEEYESIVKDL
ESDYEIALQNSKLNWATKNFNETALKIAKEHQYCILCKRELNHDELGPVMVTISENIEKANDDLYSKEKDRIKQD
LDDLKSIRDDISNFRNLEGSIASANDELNELKTFVSDDLARVSSELEQLESDLVSLESLRKHVVEISKLQEDLSH
YYVENKSIESELSAYGVPAKTLSELQEDLNGKNFQLKELRRQADDLKEQREFSNRELSMLEGNVKDKRLLISNFE
KSLMIKLNLEKGIDENKARIDELQQTGETVLESMKLTKDRLNKLKDNVKTLEEERDSTLTQLQSNVEQFKTVQDS
LVRLNSFVNKYELEDEPILRQCEKNSEHLKASIKDAEGQLNKLHEKVNVLEKQLSEADTEERNIKFNLDLRSLNK
ELVQIEESIESLDRQNATSKRQEFQKETAILRESYSRLSSAHSSKMGEVSQLQKQIQSITEEIKRDYEHVEDEYY
QEYLKLQTKMFISNDISVYSKGLDNAVMKYHSIKMEEINRIIDELWKRTYSGTDVDTIMIKSDMNTQVKGNRSYN
YRVVMMKEDVELDMRGRCSAGQKVLASIIIRLALAECFGVGCGMIALDEPTTNLDEENIESLAKALNSIIHLRMS
QKNFQLIVITHDEHFLRHMNATEFCDHFFRISRNERQKSQISMVKNFD

SEQ0087

P. pastoris homolog of *S. cerevisiae* **Mre11 (YMR224C)**

Subunit of a complex with Rad50p and Xrs2p (MRX complex) that functions in repair of DNA double-strand breaks and in telomere stability, exhibits nuclease activity that appears to be required for MRX function; widely conserved

S. cerevisiae null mutant: chromosome/plasmid maintenance: abnormal; gamma ray resistance: decreased; Rec104p accumulation: decreased; resistance to hydroxyurea: decreased; resistance to chromium(6+): decreased; resistance to 2-phenyl-3-nitroso-imidazo[1,2-a]pyrimidine: decreased; resistance to daunorubicin: decreased; resistance to camptothecin: decreased; resistance to N-Methyl-N'-Nitro-N-Nitrosoguanidine (MNNG): decreased; resistance to cisplatin: decreased chitin deposition: increased; competitive fitness: decreased; fermentative growth rate: decreased; resistance to doxorubicin: decreased; resistance to hydroxyurea: decreased; resistance to methyl methanesulfonate: decreased; resistance to Calcofluor White: decreased; sporulation: decreased; telomere length: decreased; transposable element transposition: increased; viable; resistance to methyl methanesulfonate: decreased; resistance to chromium(6+): normal; spore wall formation: abnormal; telomere length: decreased; transposable element transposition: increased

S. cerevisiae reduction of function mutant: gamma ray resistance: decreased; resistance to methyl methanesulfonate: decreased; resistance to hydroxyurea: decreased; telomere length: decreased

Chr 3, 0851

5' region (in bold start codon previous gene)
tgtcaaacacatagaagtcttttgagtttagctcacagtgaggtggaactaaatatttct
cttcaatttcataggtgttcaaccccagacttctcctcttttttcttgttactaccgtcgc
tactcgttttcgaattcttactgcggtctgtagttgaagacgatgaatcatcaaattggt
tagtctcattgccatcattatcgtaattgtcagacgtattcgttcttcgtgtgagctccc
tctggatgagatcattcagcatttcttcttcctttgcacggtcctcgttctcctttgcca
gccttgccctttcaattgcatccagttcttcttgtcttctctgaagtctgtgcactctct
cttcctccaatgattcatttctggcacttgattgccactcatcaagtttctcctggatct
gagttatgatttcgtagcacatttcttgacctttcaaatctttgaaagtatgcttaatca
tatcgttgagcttatctatctgagaagtaagcacattcttagatttgacgaactttatga
taggcattgtcttgggataggtcaaggtcatctcaaactggattgataaagacaagacgg
gttcttttgtatcgtctgaatgaagatgtacctcaaatttgggataaggatccttgtgcc

aggctgttgatgtctctgtcaggtcgatgagatcatccatatagatcgccttgagagcct
ccagctcctgaacttggagatcatgattttcctttaaagtagatgt**cat**ggtagtcaata
taccttgcctgtagatccagaagccgttttagtgaaaaaactactctgctccaaggttag
aaggtgtgttgatgtatgagcgtaagcaggaagccaggtatttttttaagagattctgat
cgaggccaactacaaatcgagattgtcaggaggggacagaggacaaaggatcacaccagc
ctacagaacagctagcccagcctctcaggaacttacagtaa

ORF
ATGCCACACGTTGACAGAATACTCCCGGGAAAGGACACGTTGCGGCTTCTGTTGACGACAGACAATCATGTCGGA
TACAATGAGTTGGATCCGATTGTTGGGGATGACAGTTGGAAGACCTTTGAAGAGATAATGCTGCTAGCCAAAGAC
AGGGATGTTGACATGGTGCTTCAATCGGGAGATCTGTTTCACGTTAACAAACCGACAAAGAAATCAATGTACCAT
GTGATGAGGATCCTGAGGTCCAATTGTTACGGAGAGAAACCCATAGAGTTTGAGCTACTCTCAGATCCTTCTCTC
TGTCTAGATAACAGGGGTTTCAACTACCCTAACTATGAAGACCCCAATATAAATGTTTCTGTTCCGTTTTTTGCC
ATTTCGGGCAATCATGACGATGCTACAGGTGACGACAACTTGTCTCCCTTGGACGTTCTAAGCGTGAGTGGGTTG
ATGAACTATTTTGGAAGAGTGGTGGACAATGACAATATTAATGTAAAGCCCTTGTTGTTTCAAAAAGGCAGAACC
AAACTGGCTCTTTATGGGATGTCCAATATCAGGGATGAAAGAATGTTCAAGACATTTAGAGATGGAAGGGTTACA
TTCTCTACCCCAGGTATTCAGACTGATTCATGGTTCAATCTAATGTGTGTGCATCAGAACCATGTCCAGCATGGT
GCTAGAACTGCATATTTGCCAGAGAACTTTTTGCCTACTTTCTTGGACCTTGTTGTGTGGGGTCACGAACATGAT
TGCATACCTTACCCTGTCCCCAACCCTGAGACTGGATTCGACACGCTACAACCGGGCTCCTCAGTGGCTACGTCT
TTGTCCAATGGGGAGACGCTGGAGAAAAATGTGTTCATTCTAAATATAAAAGGAAAAGACTATTCTTTGGAGAAG
ATTCCTCTGAAGACAGTTAGACCGTTTGTTATGAAAGATATCAGTCTAACTCAACTGGGACTAAACCCAAATTCC
AGGAATAAAAAGGAAGTTCTTGATTTTATGATTGACGAGATTAATGGGCTTATTGAAGAAGCTCAGAAGTCATGG
TTGGATAAACAAGCTGAAAACTCTTCAAGCGTAGATGATAGCGAGGTAGATACCCCATTACCTTTAGTTCGTCTA
AGGGTGGAATATTCAGGCGGCTTTGAAGTGGAAAACCCCAGAAGGTTTTCGAATCGGTTTGTTGGTAAAGTTGCC
AATGTCAACGACATAGTCATTTTCCATAGGAAGAAGGAGCATACAACGGGGGCAACCCGAACTAAACCTAATTTG
AAGAACGGCGAGGAACATTTAGAGTTAGATGAGTTGAATATTTCTAAACTTGTTGACACGTTTGTGGATGATAAT
CAGTTGAATCTACTAAATAAGAAGGACGTGGGATCTGTTGTCAAAGCCTTCGTTGAGAAAGATGACAAGGCGGCC
TTGAAAACATTCATAGATGAAGAATTATCCAAAGACCTCAAACTGTTGATGGGTTTAAGCCATGGAGAACACATA
GAAGATGAGTCAATTTCCCAAAAGAAATCCTTCAATAAGATATTGAAGGATATCAAAAAGGAAAAGTCTCAGGCA
TTGATGAGCAAAATTTGTACGGACTCCATTAAACACATCCCGGAAAGTTTGCCAGAAAGACCAGCATTCTTAAGA
TCAATTACTGGACCGGACAGAGACTCAGAGGATGCGCAGCCTAAACTTACAAACCGCAAAATATCGCCATCTGTC
ACAAAGAAACGCGTTTCCAAGCCTAGCATTCAATCGATAGTTTCTTCAGAGTCAGAGATCCTGTCCGAAGAATTA
GATGATTTCATAGATGATGATATTGATGAGGACATGGATATCAATTCGAATTCCGAGTCAGATATTGACGATTTC
ATTGATGTTCCTCCGCCAAAGAAAACTAGACAAACCAGAGCAAAAGCTAAGACTGCGCCCATAGCTAAACCAAGG
GCCACGCCCAAACCAAAAACAAAACCAAAACCAAAACCAAAACCGAAACCAAAAACTGCCAAGCCCAAAGAAACG
GGAGAGACTCTGGGATCTTTGATTGGCAACCTAAGTAGACGTTGA

Downstream (in bold stop codon next gene)
cgaagacaaacttccaattaatatacatatatgatgataaaaat**cta**taaattcgttaga
tactctcttttccaagtccaatattcttgcagtcgaagagcacaatactctgaatagtca
aactctatcaatgatattttactctggataccagcccagattccccagtaaaaccctgga
agcccaaaaaacaaagaaatttcatcgataagttgactaacgtcagcctcgtcaacttgc
ttgtccacagaggcagagacgtaagatcttacccagcttcttagaattgggttatcctgg
gatatatcgagaattctggattgatcgcattcaaaaccctgccattcaaccagatggttt
gcgatatcaaatgcccttggtcctggcagaacgtattcgtaatcaataaaaagaataggg
ttattttcaatagtgggcaacttgctactagttggggttcctctcagaatcacatttcca
cttagtaaatcacaatgacaggtaacagttggagatttgtttgagatttggctttttatc
cattccagctctttcctgaaaatggcttgtagatctaatttatcgtgggtgtcctgcaaa
atgtctttattttcaaggcatgatttttcaagctcttttataggaggcagaatattgatc
caagcctccaaaagctgccagatgtctgccgaaaacttctgtgaagcaccttgggcatta
gtgaacgatttcaacttcacaagggcgttttctatttcgtctttgtcaataatgttatgc
cactgtcctaggcgttgtgcaatcaaagggtataaggtggggtcagacagttcctctggg
gtcaacgagcgaccttccagatatccgtagaccaacccgttaccaaatcgactaaaaatt
tgggggggccaaacccagattattcaggagcaactgggacacaaactctctatctctgtct
atgatcatgtctgtgcccttaccatacgttctgatcaagac

AA
MPHVDRILPGKDTLRLLLTTDNHVGYNELDPIVGDDSWKTFEEIMLLAKDRDVDMVLQSGDLFHVNKPTKKSMYH
VMRILRSNCYGEKPIEFELLSDPSLCLDNRGFNYPNYEDPNINVSVPFFAISGNHDDATGDDNLSPLDVLSVSGL
MNYFGRVVDNDNINVKPLLFQKGRTKLALYGMSNIRDERMFKTFRDGRVTFSTPGIQTDSWFNLMCVHQNHVQHG

ARTAYLPENFLPTFLDLVVWGHEHDCIPYPVPNPETGFDTLQPGSSVATSLSNGETLEKNVFILNIKGKDYSLEK
IPLKTVRPFVMKDISLTQLGLNPNSRNKKEVLDFMIDEINGLIEEAQKSWLDKQAENSSSVDDSEVDTPLPLVRL
RVEYSGGFEVENPRRFSNRFVGKVANVNDIVIFHRKKEHTTGATRTKPNLKNGEEHLELDELNISKLVDTFVDDN
QLNLLNKKDVGSVVKAFVEKDDKAALKTFIDEELSKDLKLLMGLSHGEHIEDESISQKKSFNKILKDIKKEKSQA
LMSKICTDSIKHIPESLPERPAFLRSITGPDRDSEDAQPKLTNRKISPSVTKKRVSKPSIQSIVSSESEILSEEL
DDFIDDDIDEDMDINSNSESDIDDFIDVPPPKKTRQTRAKAKTAPIAKPRATPKPKTKPKPKPKPKPKTAKPKET
GETLGSLIGNLSRR

SEQ0088

P. pastoris homolog of *S. cerevisiae* **RAD51 (YER095W)**

Strand exchange protein, forms a helical filament with DNA that searches for homology; involved in the recombinational repair of double-strand breaks in DNA during vegetative growth and meiosis; homolog of Dmc1p and bacterial RecA protein

S. cerevisiae null mutant: chromosome/plasmid maintenance: abnormal; gamma ray resistance: decreased; metal resistance: normal; resistance to 5-fluorouracil: increased; resistance to 2-phenyl-3-nitroso-imidazo[1,2-a]pyrimidine: decreased; resistance to daunorubicin: decreased; resistance to camptothecin: decreased; resistance to N-Methyl-N'-Nitro-N-Nitrosoguanidine (MNNG): decreased; resistance to cisplatin: decreased UV resistance: decreased; competitive fitness: decreased; fermentative growth rate: decreased; Rad52-YFP distribution: increased; resistance to 5-fluorouracil: decreased; resistance to hydroxyurea: decreased; resistance to doxorubicin: decreased; resistance to methyl methanesulfonate: decreased; resistance to 2-phenyl-3-nitroso-imidazo[1,2-a]pyrimidine: decreased; resistance to cisplatin: decreased; resistance to lovastatin: decreased; sporulation: decreased; transposable element transposition: increased; viable; resistance to hydroxyurea: decreased; transposable element transposition: increased

S. cerevisiae reduction of function mutant: UV resistance: decreased; X ray resistance: decreased; sporulation: decreased

Chr 3, 0904

5' region (in bold start/stop codon of previous genes)

tccacagtgaagagatccatctcctcaggattaccaccgtattgtatatcattttctctg
gactcttctagacctttctcaacatcctccaagtcaatgtttttttctccaagctctcttt
cctttccttgaagaatgattcttcttgaaagtacctgcttcact**cat**agttatttttcaag
tctgaaatttttttaagataaaaaaattcaggtgcgaaaaagaaaaacagcgattagatc
atcaagcttgatgtgataacagctaatggggaaatatagtgtacaggagttcaaattggc
cctgctcaggttcagccgacaaagttttgagacaagacagatcttctctacttgaggtgg
ggttagactgatagtaaagtctgcttct**tta**tgcttcattcactgttgatgacaatgggt
tctgggtacctctaccatctttcaataacttgtccacctctaaaataacatccctaatac
tttcgtctgtaaagtttccggagatggagacatatgtaaccacttggtatttcttggtga
gcagataggataggcgtttcgtcacatctaatagctcttgatttccgttgagtaacgtct
gatggacatttcctgttctctaaaggcgaagttagttatttgcaacactttgcatcaaaa
aacaactacttacatctgggattcctaatacgtaacatcctaaattgtcatcggttccat
tctcactgaggtgaagagtcatggggattcgacatctcgagctcgttttgatttcttcca
atgatggcgtttctaaatacaataagaaaggaccgcccataaatgggcttaattcagctt
taaagactcttcttgt**cat**gtcaaacctcaatcacttcgtattcagaggcaatacaaaaa
gagaggagagaagatcgatgatcatgtaatatttcaaaagtcagtcgcgtaaacataaag
tatttatttacccacaacacgtaagttactctgaacgaaag

ORF

ATGTCGAACCATGAAGTCATAGAGGACTCGCAGCGGGTGATGAATCTAGAAGAGGAGCATCAGGTTAATTCTGGA
AATGGTCATGTAGAGCAAGAACAAGAGGAGGATGATGAAATGTTCGGCCCAATGCCTATATCCAAATTGGAAGGC
AACGGGATCAGTCCCGGAGATATTAGGAAACTTATGGAAGCAGGGTATAATACAGTTGAAGCAATCGCATATACT
CCTAAAAGAGCTCTGCTTACTGTTAAGGGAATTTCTGAGATAAAGGCTGACAAGTTGCTAGCAGAGGCCTCAAAG
TTTGTTCCCATGGGTTTCACCACAGCCTCAGAATTTCATCACAGAAGATCTGAGTTGATTTGCATTACCACAGGC
TCCAAGAAACTAGATACTCTTCTTGGTGGGGGTATAGAGACAGGTTCGATAACGGAAGTGTTTGGTGAATTCAGA

ACAGGAAAATCTCAACTATGCCACACACTAGCAGTCACTTGTCAACTGCCTATTGATATGGGAGGAGGAGAGGGT
AAATGCCTTTATATTGACACAGAGGGCACTTTTCGTCCAATTCGTTTGGTTTCGATTGCAAAACGTTATGGTCTG
AATGAAGATGATACCTTGGACAACGTTGCCTATGCTAGAGCTTATAATGCTGATCATCAACTTCAACTGCTGAAC
CAGGCAGCTGCAATGATGTCTGAATCAAGATTTTCCCTTCTTATCGTGGACTCCATAATGGCTTTATACAGAACC
GATTTCAGTGGAAGAGGTGAACTATCTGCAAGACAGATGCATGTAGCCAAATACATGAGAACTCTACAACGATTA
GCTGATGAGTTTGGAATTGCAGTGCTGATCACTAATCAAGTTGTAGCACAAGTTGATGCCTCCGCTGTATTCAAT
CCAGATCCAAAGAAACCAATTGGTGGAAACATCGTTGCTCACTCGTCAACAACCCGTTTATCATTCAAGAAGGGT
AAGGCTGAGCAAAGAATTTGTAAGATTTATGACTCCCCATGTCTTCCCGAGAGTGAATGTGTATTTGCTATCTAC
GAAGATGGAATTGGGGACCCTAGAGAGGATGATGACTAA

Downstream (in bold stop codon next gene)
atatgattgctatttacaccctaattaatattgtgagagaat**tca**aaactccggagccat
atttgtatccagccatcctaggaaatcatctcctttatcggaccagttgttctgttgttg
gaagattgaatcgataaaactttgatcattattcagatcattcatattgaactcttcgtt
ggccgacaagaacgattgaggattgttattcgaacttgtagtagacgcagtgctgtagac
atcattgacgtctaattgcgagttgaacaatgacgagttgaacggactctccgttactgg
aagtttttcaattttaggtacaacaggctctttactgttttgagaaactcttctaggtgt
gttcacatgggcttgagagttggaaacagctttggaaacatccttgctaggtttgtttct
cttttccttttggttgctggaggccaatgaacctgtggcttcaagcatcgccatgggaat
tccattaatcattgtaggttttgctaagcccgctgctgtagctgtagctgtggcatttaa
taaggcctgatccgtaggaaccaatgttgtaattgtggtaccattcggcgtcgttgaggt
gatcgtggtaaagtcctctttcgtaatttgattgtaaaatggtaatggaggaggtctact
agcgagcttatccttagatttatctttgcttttagtctgagtttcgttcaaagttcttct
gcgagcttcatgaatacaccaaacgaggtcgtagaataaagatcctgtcaaatgtgacct
catccgggaaattattccttcagatttgacaaataattctggatgtgtgattatcactag
gtttagcttctccaaaactgttgcagttcgactaatatcattctctacatctttccacgc
tgaaagcatgtttctgaagagtctgtggactgtgacaatactttgcctcgcactgtcaac
gtacctgggtaataaataaggtgttagatgcaatctgaata

AA
MSNHEVIEDSQRVMNLEEEHQVNSGNGHVEQEQEEDDEMFGPMPISKLEGNGISPGDIRKLMEAGYNTVEAIAYT
PKRALLTVKGISEIKADKLLAEASKFVPMGFTTASEFHHRRSELICITTGSKKLDTLLGGGIETGSITEVFGEFR
TGKSQLCHTLAVTCQLPIDMGGGEGKCLYIDTEGTFRPIRLVSIAKRYGLNEDDTLDNVAYARAYNADHQLQLLN
QAAAMMSESRFSLLIVDSIMALYRTDFSGRGELSARQMHVAKYMRTLQRLADEFGIAVLITNQVVAQVDASAVFN
PDPKKPIGGNIVAHSSTTRLSFKKGKAEQRICKIYDSPCLPESECVFAIYEDGIGDPREDDD

SEQ0089

P. pastoris homolog of *S. cerevisiae* **Rad52 (YML032C)**

Protein that stimulates strand exchange by facilitating Rad51p binding to single-stranded DNA; anneals complementary single-stranded DNA; involved in the repair of double-strand breaks in DNA during vegetative growth and meiosis

S. cerevisiae null mutant: UV resistance: decreased; colony sectoring: increased; gamma ray resistance: decreased; metal resistance: decreased; metal resistance: normal; resistance to cisplatin: decreased; resistance to hydroxyurea: decreased; resistance to N-Methyl-N'-Nitro-N-Nitrosoguanidine (MNNG): decreased; resistance to camptothecin: decreased; resistance to daunorubicin: decreased; resistance to bleomycin A2: decreased; resistance to tin(2+): decreased; resistance to methyl methanesulfonate: decreased; resistance to tetrabutyl hydrogen peroxide: increased; resistance to chromium(6+): normal; resistance to methyl methanesulfonate: absent; resistance to chromium(6+): decreased; resistance to menadione: decreased; resistance to ethylmethane sulfonate: decreased; transposable element transposition: increased; glycogen accumulation: increased; competitive fitness: decreased; fermentative growth rate: decreased; mitochondrial genome maintenance: abnormal; Rad52-YFP distribution: increased; resistance to sulfanilamide: decreased; resistance to doxorubicin: decreased; resistance to hydroxyurea: decreased; resistance to 5-fluorouracil: decreased; resistance to methyl methanesulfonate: decreased; resistance to cisplatin: decreased; transposable element transposition: increased; viable

S. cerevisiae reduction of function mutant: X ray resistance: decreased; mutation frequency: decreased; sporulation: decreased

Chr 2-1, 0153

5' region (in bold start codon next gene)

aagtcaacaggagagaatggtaaaacaaggaacagaataacccaaatccgatgttcaaat
gtggcctgttgtgattcagcccaagaacccaatagaaaggatataacaatacgttgtaaa
ctgaagatcttgacaaaacgtatagaattggtgttgaaagagagaccaaaaccgaagagg
tcaatgatccactcaagacagtggtgaagttgatcaaattgaattctgaacgtggctcaa
tgcggtacacaccaatttgaaagtttagtttgtttcgcaagaatactgcctgttgaattg
attgccatagcaccaaagtgccaaaaaagaaccaaaagtacacaaactggtcgttgatat
aaggtagcactgtccttgtgggagtgataatataatacgtaagatcgtcactcttctgtg
agaataagataacccagtaagtgaaacctgatattccatatgccaaaagatgggagacaa
aatctgcacaaaatatctggcccagaatttgtctgaaaaaggttttgaaaccaggatttt
ccaacgtttgagttgactccttgcaatgcctgaccaggatcaaagaaacacccagcaata
accatttcaaaggccaaaaccaaagtgatgaacctgcaaaaatgccaatgaaaagtgacg
agaaaaaggtcaacaaaacaacaattctgttcacaaagctcagacgttttctttggactt
gttcaaagatcgttttgtagtagattaatttgggtttggcatcttctatggt**cat**tgttt
gtagtgtaggagaggacaacttgagaagtaattaatttttgagaacacggcgaggctaaa
gggcagagcaatgtccgcctaacggagaaaccgtgcaaggggtgactttagtacttgtcc
cgtatgcgtggaggggcatcgaaaacgcgcaccgaacatcaaatatttactctatcgaac
gcgttccacctttgttctctctttcctgcaaaccacactcc

ORF

ATGTCTTTCGATGACGCTGAGCTCAAACGCATATCTAAGGAGCTCGACAAGCAACTTGGTCCAGAGTTTATTTGT
ACAAGACCTGGCCAAGGTGGTATGAAAGTGTCGTATCTCTCTGGCACTACTGCTATTAGTTTGGCCAATCACATC
TTTGGCTTTAATGGGTGGCATTCGGAAGTTAAGAGTACTACAGTGGATTTTGTAGATACTCAGCATGGCAAAATT
TCAATGGGACTTTCGTCTGTTATTCGGGTCACCTTGAAAGACGGTTCTTTTCATGAGGATATTGGTTATGGAAGC
GTGGAAAATGCAAAATCCAAGGCCATAGCTTTTGAGAAGTGCAGAAAAGAGGCTATCACCGACGGACTCAAGAGA
GTTTTGAGATGTTTTGGTAATGCTTTGGGTAATTGTCTGTACGACAAGGAATATTTACGAAAGATTTCAAACGTT
AAAACTCAAGCTATCACATTTAATGAAGGCGATCTTATGAGACACAATCAGCTTGAGGCACGAACTCTCAAACAG
GAAGCCAAGCTTCTGGAGATTAATCAAAAGAAAGCTAAAAACAGCAGCAAGTCAAGAATTAATGTCCCTCCCAAG
CATTTTGGTGACAATGAAGACGACGATTCGCATTTGTTCAGCGATGAAATCAACATCGATAGTGAGGACTTTATG
AATGAAATTGATGATTACGAGATGGATTTGCTGATGCAAAAGAACTCTCAGAGAGAAATTGAAAACGTAAACGAC
GATGAGATTGAAAATAAGGGTGACGCTGTCGACGCTGTTACCAAGAGTAATATTGAAAATCAGTTCATTGCCCCC
AATAGTCCCCAGATTATGGACAACATGGTACTGACTCCAGATAAAATCCCAGCCAAGGTTGAATTTGTTTCTGCC
AAGGTTGCCGAGAAGGTCCAAAACAATAGTCCATTGAACGTGGAAGAAAAGTTCAACCCATCATTCCAGTCTCCT
TCTTTAAGAAGAACTGTAGATCCTACGAAATCAATGCCCATTAGACGTACAATGGTTCAAACATCTGTGCTTAGT
CAGTCCAAGCAGGGTCCTAAACGAATGATCGGAATTCCTCCAGATACGGCAAAAAGGCATAAGCTGGGATCACAA
AACCAAAACACTGCCAACGATAAACCCGAGGAGAAGCAATTGAGTCGGTTCAATTCTACAAAGGCCCCAGGAAAA
GAAAACTCTCCAGCTTCGAATTAA

Downstream (in bold start codon previous gene)

gtaagcagttccaacttttggagatttgtatgcatacaatcactcatccctcgcatccac
cttattttattcatcgcgactcgagccaagcatgctttctttcctctttagcccccaacc
tacagtgaaccaaatac**atg**tccattaactcaatcaaaccatttacggagaagaagggtc
cattcacagacgatgtttacgagcctaatacggaagaaacattcaatgccatcagaagtt
caaaaatcctagtaattggggcaggaggattaggatgtgaaatcctcaaaaatctgagtc
tttcaggattccaggatattcatgttattgatatggatacaattgatttaaccaatttaa
atcgtcaatttctattccgaaataaagatattggaaaaagcaaggccaaggtagcatccc
agtttgtcatgaaccgtatacctaacgttcaaataactcctcatttttgtaggatccaag
ataaagacgatctgttctacaggcaattccaactggtaatctgtggtttggacagcacag
aggcaaggaggtggataaatcataaattagtcaccttgctagatccaaatgacttttcat
cgttaattccaatgatagatgggggaacggaggggtttcggggtcaatcaaggttgattt
tacctacgttatcatcatgttttgaatgctcactagacatgatcccaacaaatgtaacat
atccagtgtgtacgatagcgaacactccgagacttcccgaacactgtatcgaatgggcac
atcagttggaatggcccaagaaatttggtgacaaaccttttgacgcagacgatccgtcac
aagttgactggatgtacaaaacaagtctagaaagagcaaagcattttgatattgaagggg

tcactctgtccttgactttgggagtggtaaagaatatcattcctgctatttcatctacaa
acgccatcattgctgcttcctgttgcaatgaagccctgaaa

AA
MSFDDAELKRISKELDKQLGPEFICTRPGQGGMKVSYLSGTTAISLANHIFGFNGWHSEVKSTTVDFVDTQHGKI
SMGLSSVIRVTLKDGSFHEDIGYGSVENAKSKAIAFEKCRKEAITDGLKRVLRCFGNALGNCLYDKEYLRKISNV
KTQAITFNEGDLMRHNQLEARTLKQEAKLLEINQKKAKNSSKSRINVPPKHFGDNEDDDSHLFSDEINIDSEDFM
NEIDDYEMDLLMQKNSQREIENVNDDEIENKGDAVDAVTKSNIENQFIAPNSPQIMDNMVLTPDKIPAKVEFVSA
KVAEKVQNNSPLNVEEKFNPSFQSPSLRRTVDPTKSMPIRRTMVQTSVLSQSKQGPKRMIGIPPDTAKRHKLGSQ
NQNTANDKPEEKQLSRFNSTKAPGKENSPASN

SEQ0090

P. pastoris homolog of *S. cerevisiae* **Rad57 (YDR004W)**

Protein that stimulates strand exchange by stabilizing the binding of Rad51p to single-stranded DNA; involved in the recombinational repair of double-strand breaks in DNA during vegetative growth and meiosis; forms heterodimer with Rad55p

S. cerevisiae null mutant: X ray resistance: decreased; gamma ray resistance: decreased; resistance to 2-phenyl-3-nitroso-imidazo[1,2-a]pyrimidine: decreased; resistance to daunorubicin: decreased; resistance to N-Methyl-N'-Nitro-N-Nitrosoguanidine (MNNG): decreased; resistance to cisplatin: decreased; resistance to hydroxyurea: decreased; sporulation efficiency: decreased; transposable element transposition: increased; competitive fitness: decreased
Rad52-YFP distribution: increased; resistance to hydroxyurea: decreased; resistance to methyl methanesulfonate: decreased; resistance to doxorubicin: decreased; resistance to cisplatin: decreased; sporulation: decreased; transposable element transposition: increased; viable

S. cerevisiae reduction of function mutant: X ray resistance: decreased; spore germination: decreased

Chr 1-3, 0025

5' region (in bold stop codon previous gene)
aaaaagtgtccttcactggttctactggagttggaagtgtcttgatgactcaatccgcta
agactttgaagaagctttcgtttgagctcggtggtaacgctccattcatcgtttttgaag
atgcagacattgacaaggctgtagaaggggctttgatatctaaattgagacaatcagggc
aaacctgtgtgtgtgccaacagattgtacatccatgagtcaatttacgatgaatttacag
aaaaattggttgcaaaggtaaagaaatgcaaactgggcaatggtcttacagaaggcacta
cacacggacctctcgtccacagtagagctgttgataaagttcaacaacatgttaaggact
gtttggacaaaggagccactcttttgttaggtggtaaagtaagaaccgatttggggccaa
atttccatgatttgaccgtcattggcgatgtgacccaagagatggcagttgcacacgagg
aaacatttggaccattggttcctctcataaagttcagctccgaggagcaagtactagaat
gggcaaacgacacagaatttggtcttgctggctacttttcaccaacaactactcacgaa
tctttagattgggtgagcaaatcgaatgtggtatggttggtatcaatactggtgccatct
ctgaggccgctgtaccatttggtggtattaagcactccggttttggtagagagggttcca
agtacggtgtcaacgattacatcaacatcaagaccatggtgattggaggactc**tga**actc
gccgttcattttgtatagtttgtttatttaaggaggttgactaaaattacgatatttaat
atagaggataattattgctaaggaagatttcggaagttccattatctgtgctcgatttaa
ttcaccttcggatgtctactcgatatccttccaccaggaacctcttggtataacacgctt
tggtttcaccttccagtgtcttttcagatcagctgatagtt

ORF (in bold stop codon next gene)
ATGGATCTGTATGAGCAATGTCCGAGCTCAGAACTTTTGTTGAACGAACGATTTGAGCCTTTGCTGACGGTGCTG
ACTAAGGCGTGGATACTTACTTCTGATATAGTCATTTCCTCCCCACAAGACATACACAAACGTGTCTCCAATTTG
ACAGGGGGAAAACTGACTATATCGCTGAATGAACTGATAGTATTTAGAAGTCTCTTGATGAAGGAATTGAACTCG
CAAATTGAGTCCAACTTGGTGAACTTGAACACATCAATGCTCCAAGAGCAAACACTGAAGCTGAATGATTTATCT
TTTACCACGGGAGATGAATTATTGGATTACTATCTACGGGGAGAGTATGACGATAACGTCGTTTTGCCCTTGGGT
TATCTGATAGAAGTGACAGGAGGCAGTTCATGCGGCAAATCTCATTTCCTGATGCAATTTTGTCTTACAGTGCAG

TTACCGATTCAACTTGGAGGCTTAGGGGCCGACCTAGAGGCGCCAGAGTCTGAACGAACTAACGAGTGCAAAGCT
GTTTATGTTTCCACAGAGGGGTATATGGAAACTAGACGGTTAAGTGAAATGTGTGACTATTATGGAAATATCATG
ACAACCAATGGAATTGCTGATTCGGATCTGTATCCGTCAACCTCAAATGTCTTGCAAATTCAGTGTTCAGATTTA
GAAACGCAAGACCACATTATTTTTGTACAGCTTCCAGCTTTACTAGAGAAAGAATCAGGAAAGGTGAAACTGATC
TGCATAGATAGCATTTCGCATCATGTAAGAGTGGAATTCGAAAACCGATCATACCAAGAGTCAATGGCTCGGAAT
ACTTATCTTTACGACTTGAGCAGACATTTGAAACGGCTGGCAAAACAGTATGAGGTTTGCATTGTTTTTAGCAAC
CAAGTCAGTGACAAGCCTATTACTAAACTCATTAGTAATGATCGACTTGCCCTAGATTATCAACTGGGCTGGATG
AGCGGCTGGTCCAACGAAACTATCCTGTATAGACAAAACTATCAGGATGATGATGGAGACGCAGACGCTTCGTTC
ACGTCAAGTAGCAGAAAAAGAGATTTCTCAAGTATGACTGAAGAGGACTTTGCAGATGAGGAAGATGACGGCACC
AAAGAGGATACGGACACTGACGAGAGAAATTTCACTCTTCAAGATGAATTATCATTCACAGAACAGATCAAACTA
GACAAGCATGAGTATCTATTTGGGAATGAAGTTACTAAATTAAAAGAGACCAAAATTCCTGCTCTTGGGTATTTA
TGGAATACATTAATTGATGTGCGCATTGTGTTGTTCAAAAACAGCAGACCGATAATTGATGAAAATTTAATCGAT
AGCATGGCTTTTGATTTAGGATTATCAGACTTTAAAAACATCGATCAGCTAAAAGAAAACCGAATGATGAATGAA
GGGAATTGGGCAACAGAGAGATGGTGCAGACTTTGTTTCAGTCCATTCAACGGGTATTCAAGTGATAAGGTCCAA
AGATTTGAGATCTGGAAAGGTGGTATAAGGGGGGTGG**TTA**ACTAA

Downstream (in bold start codon next gene)

cctgcaggtcttccgcaacaatagtgttgttgtcatcctcatcatcatcctcttcaataa
atacatcaggatcttcttgattaggatcgtcatcgtcgtaagtgagatcatattcttcat
cgtttagattgaagttcccactaacttcgtctagagtgcttactttggaatttcttcgca
tgggaggggagtatttgttcttaatagtggccttgagatgcaggaacgcgacccgtttcc
gtttcttattaaagatgaaccacattaaacaccatagcgtattaggctctaagtcagcca
aaactccatccttgggttccaacgcgtaaatggagcagtcttgaatgtccatatgggaat
tgatagtctcccaaatccagttaatatccttatcaccatcgtctatctgatccttatagc
ccaaagacaatacaattgagttgaaacggttcatcatggaatttgcagagtcaactgctg
tgaagtagtccggcgtcaatgaactaaagtcatggtcgggatacgacgcgttcagtattc
ctattaggtacgaaaagatttttctgctatgtgtggaagttatgggaccaaatggggtct
cctcgtcatacagcgcgctttcgtttatgtcaatattctccaaagacttactcaatgtag
ttgcttggtggtggttgaaagactttcttttcaaagttggtttcacaatactagatgagg
aggccctccggctaaaccgcgtctgaaaatcaacattgtcggttatcctggaagctgtgg
gcggagagtgagacatgttccaattcaagctttcctgcaatatgtcatctagatgcttat
caatcgttttatatagtttcttgtccgaccccacaggcttcgtcgtgaacaaatcgcaac
cgccacgaaccgaatagtcaggagtttcaaaacagagagcctggttaaccagctctatat
caatttcatcaatgaactt**cat**tgtagtgaacgtagcatgc

AA

MDLYEQCPSSELLLNERFEPLLTVLTKAWILTSDIVISSPQDIHKRVSNLTGGKLTISLNELIVFRSLLMKELNS
QIESNLVNLNTSMLQEQTLKLNDLSFTTGDELLDYYLRGEYDDNVVLPLGYLIEVTGGSSCGKSHFLMQFCLTVQ
LPIQLGGLGADLEAPESERTNECKAVYVSTEGYMETRRLSEMCDYYGNIMTTNGIADSDLYPSTSNVLQIQCSDL
ETQDHIIFVQLPALLEKESGKVKLICIDSISHHVRVEFENRSYQESMARNTYLYDLSRHLKRLAKQYEVCIVFSN
QVSDKPITKLISNDRLALDYQLGWMSGWSNETILYRQNYQDDDGDADASFTSSSRKRDFSSMTEEDFADEEDDGT
KEDTDTDERNFTLQDELSFTEQIKLDKHEYLFGNEVTKLKETKIPALGYLWNTLIDVRIVLFKNSRPIIDENLID
SMAFDLGLSDFKNIDQLKENRMMNEGNWATERWCRLCFSPFNGYSSDKVQRFEIWKGGIRGVVN

SEQ0091

P. pastoris homolog of *S. cerevisiae* **RFA1 (YAR007C)**

Subunit of heterotrimeric Replication Protein A (RPA), which is a highly conserved single-stranded DNA binding protein involved in DNA replication, repair, and recombination

S. cerevisiae null mutant: UV resistance: decreased; resistance to methyl methanesulfonate: absent; resistance to hydroxyurea: decreased; inviable

S. cerevisiae reduction of function mutant: budding: abnormal; heat sensitivity: increased; radiation resistance: decreased; resistance to methyl methanesulfonate: decreased

S. cerevisiae overexpression: cell cycle progression in M phase: abnormal; chromosome segregation: abnormal; nuclear morphology: abnormal; position of spindle pole body: abnormal; resistance to hydroxyurea: decreased; resistance to nocodazole: decreased; spindle morphology: abnormal; cell cycle progression: abnormal

Chr 1-1, 0113

5' region (in bold start codon next gene)

tcttattacttgtatctagatgttttatgagagcttggtcagcaatcctccagaacttca
catcaccattctctaaaccgcagatggcaatttttccttcaacctcaaagaggttttcga
tacgagtattcaccactgaattggcctgactcctgtcttcatgtattttgacagcggtaa
ttccacacaaacaatccatcgaggcaacgctcatcccactagagcattcccacagtttca
attgtgcgtctaaggatgatgtcacaaagtttggcccctcccaaccagactgaagtccgt
aatcgatgcctgatgtcctgaaaaagttcttgcattgtctccagtctcacaatcccaaat
ttttagcatcatgtccaagccagcggaaagcaggaccttatcagatggaaaataaaggac
ctttacgatcccaccatcatgagccattacttctttcaactcatccctctttgatcccag
ctttccatacaacagtcggccattgcggtctccgacaacaaaattgcattttgccgtaac
attgacagacgtgagcttatgcgaaataaagtcatacttttctacggggaagacacattt
caatgttgactcttcatctgagagctgatatttccagtgatctagcttacgtaacgtcaa
acctggactgagtggttctagcttgtccttgtagactctaaacttgctaggtttggtttt
gccatcaattatttcagcccaaacatcttcatgattcacacggccagactccacgtcact
aataacgtccaaaaagctgtcctgaattctcacaagctt**cat**cgttgggagttgacacag
gaaagtcaagggcatctaacgaaggcgcgataaaacgaaatcacctagttctggcacccg
tcatcggggtgttccctgtgctgatcgcgcatcgcgtaaattgccatcagaacgtatgta
acgcgtctcacgctttctctactcactctacgactaagaca

ORF

ATGACAGAGTTCAGTAAAGGTAGTTTGGTTGAAATCTTCCAGAAGGGCTATAAAGGTGGACTAAAACCTTTGACG
GTTCAAGTTTTGAATTTGAAGGCAATTCCCAACAATACGGGCAAAAGGCTTCGACTCGCCCTTTGTGATGGTTTG
TATAATGCCAACGCTGTTATTAGGCCTGAATCGGTCGAAAAGGCCGAAGCTCAAGGAATCAAGAAGGGAAGCATT
GTCCAATTGCTCGAGTACAAAGCTTCAATGATAAGTCCTGTCAAACACGTCTTGATCATCGATAATCTTCAAGTT
TTGGGATTTCAAGAGGAGAAAATCAACCCATCCCCAACTAGTGTAGACCAATATTTCAGCAACCACTCTGGAGAA
AGCAACGAAGACTTGTTGGGCACCTCAATGAATTCTCCAGCGCCTCAAGAGCCTGCGCAGAAAGCGCAATCACAT
CACCAAGAAGATGCAAAGCCAAAACTGTCAGCACAGGTTACGTCCAAACCCCAACAGACCAATTCAAGTACCGCT
AAATTTCCAAATATCCATGCAATTGATCAACTAAACCCCTATCAGAATAATTGGACCATTAAGGCTCGGGTTTCT
TATAAATCTGATATGAGAAAATGGTCCAACCAGAGGGGTGAAGGTCAGCTCTTCAATGTGAATCTTTTAGATGAA
ACTAATGAAATCAGAGCAACTGCTTTCAACGATGTTGCAGATAAATATTACGACCTCCTACAAGAGGGCAAGGTT
TACTACATCAGCAAAGCACGTATTCAGCCAGCAAAACCTCAATTTTCCAACCTTACCCATACTTACGAATTAGCT
TTGGATCGGGACACTCAGATCATCGAAGCAGAGGATGCATCTGATGTCCCTTCTCTTCATTTTAATTTTGTTAAA
CTCAACAAGGTACAGGATCTTGACGCAAATGCCATTATTGACGTTATTGGTGTCATCAAAGTGGTGAATCCTGCT
TTCCAAATCGTGGCCAAGTCAACCGGAAGACCATTTGACAGGAGAGATATTGAAGTTGTAGATAATACTGGATTT
GCTATAACTGTAGGGTTGTGGAATAACACTGCTCTTGAGTTTGACATTCCTGTTGGATCAGTAGTTGCGTTTAAA
GGTGCTAAAGTTCAAGATTTTGGTGGCAGAAGTCTTTCATTGACTCAATCAGCTACCATAATCACTAATCCAGAC
TCTCCAGAGGCTTATCAGTTGAAAGCATGGTATGATCAACAAGGTGGCTCCAATCAGGAGTTTAAGTCCCTGAAA
AACGAAGTATCTTCCAACAGTGGTTTGAATACGAAACAAGACATTCAGTCCCGCAAAACCATTTTGCAAGCTCAG
TCTGAAGAACTCGGTAAGAATGACAAGCCGGACTATTTCTCCATTAAAGCTTACATCAGTTATATTAGGACAGAA
AACTTTTCGTATCCAGCTTGTGCATCCGAGGGTTGCAATAGAAAAGTCATCCAACAGAGTGACGATACTTGGAGG
TGTGAAAAGTGTGATGTCAACCACCCCAAACCGAACCATCGTTATATATTGACATTGTCTGTGGTAGACCACACA
GGACAACTATGGGTCACATTATTTGACGACCAAGCACAACAGCTCTTGGGACAATCAGCTGGTGAATTGATCGAC
TTGAAAGAGAATGACATGTCTGAAAATAACCATGCATTCCAGCAAGTATTCAACAGGATACAAATGAAGGAGTTT
TCGTTCAGAGTAAAGGCATCTCCGGATTCTTACAAGGGTCAAACCCGTATTCGGTATAATGCCGTGTCTCTGGCC
AAACTGGATTTCGCTTTGGAAGCAGATGCATTGGCAGATTATTTTGAGGGTGTCAAGGTTAATTAA

Downstream (in bold stop codon next gene)

gtttgagttctacgtacggtgtatttttcttttaatttcattaaaggcacttgttcct**t**
taagaatattctattagttcgtttacctgtagccagaacttttcagtaagtcattcacc
tttcttgaaccgcgagttaacatcagttgctgacctagcctaacggtatcattggtagga
atctgtttttcaagagctggttattgttgatagtttcaaatgctagattaaaccatttc
ttgaagttcaaaggatatttgacaagtatcaccttgaatagttccatgaagttttccaaa
tcagatctagatgccatcaagaacgaagcaaccaacttgaatatcaatgacggccctagt
tcggtatgctcaatcatgtctctgataataactgaatcttctttgtctccctttctgaga
gtgacaagtttgatccaaaacttacagcctgatctaagcacaaatttttctttagaaccc
aaagattccaaaacgaatggtaatatgttatgcaaaagatagccttttattagtatgcct
ggtttcttctcaagtatcactaccaaaatgttgaggaaacataatgctagatcaggatcc
tgttggatgatagccaaatactccacacatgattgctgaagcgtggtcgtcaacaaattt
gaacctattgacgagctctttattaaggattctagtaaagaatactgtaatggaacaacg
ttcaatgaatttttggaaaaacattgcaaaatcttcgcattcaacagatcgatgatagag
taggtgtcaaaaacaaaaggacctggaatagcttctgtcaaacctgctttcaaaattcca
gtacaaacttcgttcatcatataattatcagaagtaagaggattcaaaatcataaactgt

ttcactacttcagtgatttgctttttaatgtggaggggtcttcaatccagtaagcgtac
acggcatctctttcttcaggtgtgtaccagtcgttcacatc

AA
MTEFSKGSLVEIFQKGYKGGLKPLTVQVLNLKAIPNNTGKRLRLALCDGLYNANAVIRPESVEKAEAQGIKKGSI
VQLLEYKASMISPVKHVLIIDNLQVLGFQEEKINPSPTSVDQYFSNHSGESNEDLLGTSMNSPAPQEPAQKAQSH
HQEDAKPKLSAQVTSKPQQTNSSTAKFPNIHAIDQLNPYQNNWTIKARVSYKSDMRKWSNQRGEGQLFNVNLLDE
TNEIRATAFNDVADKYYDLLQEGKVYYISKARIQPAKPQFSNLTHTYELALDRDTQIIEAEDASDVPSLHFNFVK
LNKVQDLDANAIIDVIGVIKVVNPAFQIVAKSTGRPFDRRDIEVVDNTGFAITVGLWNNTALEFDIPVGSVVAFK
GAKVQDFGGRSLSLTQSATIITNPDSPEAYQLKAWYDQQGGSNQEFKSLKNEVSSNSGLNTKQDIQSRKTILQAQ
SEELGKNDKPDYFSIKAYISYIRTENFSYPACASEGCNRKVIQQSDDTWRCEKCDVNHPKPNHRYILTLSVVDHT
GQLWVTLFDDQAQQLLGQSAGELIDLKENDMSENNHAFQQVFNRIQMKEFSFRVKASPDSYKGQTRIRYNAVSLA
KLDFALEADALADYFEGVKVN

SEQ0092

P. pastoris homolog of *S. cerevisiae* **Rad59 (YDL059C)**

Protein involved in the repair of double-strand breaks in DNA during vegetative growth via recombination and single-strand annealing; anneals complementary single-stranded DNA; homologous to Rad52p

S. cerevisiae null mutant: resistance to hydroxyurea: decreased mitotic; recombination: increased; Rad52-YFP distribution: increased; resistance to cisplatin: decreased; resistance to hydroxyurea: decreased; resistance to methyl methanesulfonate: decreased; viable

Chr 4, 0087

5' region (in bold start codon next gene)
gtaattttggaatgcagttggtttcaagaaaagtttttggttggaggcattatattcta
gaaggttgcaaatcaaagtcagacaatcttgaaccacaattgagcccctgattcctccct
ccaattcaataatttcaaaaagtttatcaaagcaattttgaaataccaccagcttttgaa
tattattattctgattgaccaagcccataagcagcaaaatgatctcgtttcgaattggct
cattagggtcgtctagcaaagacaccaataaaggcgttcccaatggtgatgatacaatgc
attcttttgatcttgatggtctacacgacaccagagcctctagtatctgtatggtgtaga
tcttgatataaaaatcatggccaatctcactatctaatatctttaataaaaggtcgaaat
tctcaggggcctgggttaattcgtcagcaactagcaaggagaactggtccattcttactt
cacctgctagaaggctgggagaggggtatttaccattttgagctcgtgattgttccgata
tcaaacttctagtttggtcttcatcagattcacctctgatgaatagtattaacaaagtct
ctaaaactgcctttacgacgtcaatatcctgtatatccttgtgtagggtggaaactaaac
ctttcaaaccattaccaataactgattccctgtaatctctactgaaattcttcaacgaaa
gaacagcggctcggcgatcagtgaggagggttgcgttggaaagacggtcacaaatcgtct
caatgacccctctgggtcaacctctgttttttgctcctaagaccagagaa**cat**agccg
tctttggaggggaaatcgaaggttttgacgagagacgtgtacttttctagttgatcgttc
ctgcccaatattatgatttgatatctccaggatattttcctactcgcgcgcgtcgtatca
gttccaacgattccagctgatcttcatctgggggtttccac

ORF
ATGGACAGTAGCGATTCAGAGTCGTTGGATGCTTGTCAAAGTGAGCTATTTCCCAAGGAGTTGAAATTTGATCTT
TTTGATGACTTTTCTATTATCAAGCCTTCAGAATGGGCGATTACAAGAATTGGCTGGCTCCAAACGAAGATCGAT
CGGCTACATAGAAGCCGTAACTCAGTGCAGAGGAGATATCTCAAGAGGGAGTTGGAAAATCAAAAACTGATCTCA
ACTGCGAACCAAGTGTTTGGGTTCAATGGCTGGTCTTCAGTTTTGGATGGGAGCTTTGAAATGTCGGAGTGTGCT
GAGGAGGATGGCAGATTTAAGATGAAAGTGCTTACCACAGTGAAACTGATATTGAAAGATGGGACTACTTGCAGT
GCTACGGGTACAGGCAGGACTCAGAATGCTTCAGGGCGGGCCCAGGCGTTGTCAAAATGTAAAAAGGAGGCGATA
AACGATGCCATAAGAAGATGCATAGAGAGTCTACCTAACTTACTATTGGAATCTCAAGACCGAAAACAAACTTTT
AAGGATTATTGA

Downstream (in bold stop codon previous gene)
ttcaaactatttacagagctatatgaatttttccagtctat**tta**ctgaaacggtccccga
agacggccatctgttgcttgatgcttatgtggaagtgccttagggtcaatttcaggttgt
gatttggtgtctatcgtatcgtagaacttttgaaagaatgatgctaacggacgtacatac
caacgttttgtcggtgattgaaaccatgccaggtgaccacctaaggatgtcgttaacatt
aatgtgaatgggttggatttgcattcctggaacggaagagcttcagaaccaacaattggg
tcatccaaagcgttgataattaatagaggtgttctaacagcagggagcctgttaaccgat

gaacccactctataatactcatatgcagtattaaacttgaacaaaggtgccgtgtacctg
ttgtcaaagtcatacagtaatttgatgccttcgtggtatctctctctattctgactgtac
aagtcactaccacgaagaatgtcatcgtgacgatctactaagttcattagcgattgagtc
attgcaggggaataaatgtacctgccaatccatgatcgatgtagcagataagatgaggcc
gctagatcccatggatttcctagaactgcagctgctgcaaagcgacactcatcaccctcc
tgtccaatgtagttagatagaatagaagctcccaaggaggcacctacagcgaaaagtgtc
ctttcaggaaaccatgtggctaagtgattcaccaagaatctcaaatcctcagtccaaaat
ccattgaacagttgaggagttgacaattgcgattgcgcacatccacgggcattcagcaca
accgcgtcgaaatcatgttcttcggtaacttccttcaacagacatctaacatatgcttca
tgactaccaccggataatccatgcaatacaaccagcatggggcgtgtatcatcagaggcc
tggttctttagctcaccctcgttcataaatctagtccgagg

AA
MDSSDSESLDACQSELFPKELKFDLFDDFSIIKPSEWAITRIGWLQTKIDRLHRSRNSVQRRYLKRELENQKLIS
TANQVFGFNGWSSVLDGSFEMSECAEEDGRFKMKVLTTVKLILKDGTTCSATGTGRTQNASGRAQALSKCKKEAI
NDAIRRCIESLPNLLLESQDRKQTFKDY

SEQ0093

P. pastoris homolog of *S. cerevisiae* **RDH54 (YBR073W)**

DNA-dependent ATPase, stimulates strand exchange by modifying the topology of double-stranded DNA; involved in recombinational repair of DNA double-strand breaks during mitosis and meiosis; proposed to be involved in crossover interference

S. cerevisiae null mutant: colony sectoring: increased; sporulation: abnormal; toxin resistance: increased; viable

Frag B, 0039

5' region (in bold stop codon next gene)
ctgtctggcgtagggatgatctccaaaccacttatgccaatctcccaaggctcaaactgc
tcttagtttccccaatctctgctatcacaaaggttcgttaagcatcatctctatacttgt
aagatccgttagagagctaggcacgtgggctttgaaactgcagcatcgatgttgtttcaa
agctgctattgccgagtactaacatgttcaatgctggtatgcgttctcaggtggaaagct
gcatgaatcaagcacagcaaattcgtctggctgatttttgccaaattagtcacgaatgcc
aggtacgatctggagctcggattgcaacttatccaaggatgggctttcaatgaagtatcc
ttaatgacataatatcatgtactatcaccgtgaattcgaggagacacacattgacattgc
cggtcctctattcttccgaaataattcacccatgcccacaacaattttgagatgtccttt
ttcgtcctgtagaagccgaataatcaaaacggatgacgagaatacgaacatcaagactgt
cacaggaaatggcggacccagacttttgcagttttcagatggagaggtgtccctgaacag
ttccaccgagaacttaaagttctaccaggtcaatgatatgtggacttttgataatattgg
agtctctaagccaacttccatggatgatgagtttagtataaccatcgatggtggttctgt
tccagtccatatagaacgttacctgacatgtgccgactgtgacaaagggccaattggaat
cgccggttcagtcaacaaagaagcattggaagacccaagcggatccaagttgatgtatta
tctttacactggctcagtgtttcaagaggtttctccttctccaattgacacatgacatgc
gactttcttcactggagtcatcttacataatcaacgcgttatctgcattaaattaacaaa
taaaatattacgtagtcaatacccgttgttgccagcccaca

ORF
ATGTACAAAAAGCCCATGAATCGGCCGTTTAAACCGCCGAGGATGCAAGAGGCCACTAATAAACCAGTAAAACCA
GCCCAGCCTGCGAAAATTTTCAAAAGAAAACCAGAAGTTTTGTCTCCTGGGACACCTGAAGTTGCACAAACAAGG
ACTACGAACTCCGTCAATGTTCTATGGAGAAAAAAGACCCTCAAGAAGAATAAGACTTGGGACGGTGATGGGATC
TTAGTGATTACGGGAGACCAGTTGATTTTCAAATGTGACCCGTTGGGAGGTTCCAACTTAAGAGAACAGGGTAGA
AGATCCAAATCTTCAAAGAGTCTTGAAGGTGTTATAGCCATAGGCAATTACGAAATAGAACTTGATGGCGAGATT
TCTGACGTGAAGCCTGCTTCGAAGAAACATAGAAAAATAGAACATGAGGTTAGTTCGGAAAAAGAATCCATAACT
TCAGCAGATAAGAAATCTTCCTCCCTTCCCTTCAAGAGCCCCATGACAAATTCAGTGACTAGACAACCACAGAAG
AAAAGCGCTCCACCATATAAATCAGATTATGAGCATTCGATTGTACTATCTCAAGAATCGGAGACTTCTAAAATG
GTAGTAGTGGATCCGTTGCTCAGCCAACATTTGAGACCTCACCAAAGGGAGGGTGTGAAATTTCTTTACAATTGT
CTGGATGGCTCCAAAGAGTTGCCACATAAAGGATGTATATTAGCAGATGAAATGGGACTTGGAAAGACATTGACC
ACAATTACCCTCATTTGGACTTTACTTAAACAAAACCAGGTTGATCACAAACGACCAGCAGTGAAGAAGGTACTG
GTTGTGTGTCCTGTTACTCTGATACACAATTGGAAAAGAGAATTTCGAAAGTGGCTTGGAATGAATAGGGTCAGT
ATTTTGGAAATGAGCAGTGCATCAAATGTAGAAGAAGACAAACGAAGTGTGATCAACTTTGGACGAACCAGAGTG
TACCAAATATTGATCTTAGGTTATGAAAAACTGCAAAAGTTGACGAATGAAATTTCCCAAATTAACGTTGATTTG

TTGGTATGTGATGAGGGACACCGGTTGAAGAACTCAAACAATAAGGTGATGAAAAGTTTGACTAGTTTCCAGATT
CCTCGTAAAATAATTTTGACAGGAACTCCAATCCAAAATGATTTGAATGAATTCTATAATATCATCAATTTTGTT
CAGCCTGGAATTGTTGGAGATTTTGCTAGTTTTAACAGGGACTACATGAGGCCGATTCTTCAAGCCAGGGAAATT
AATTGCTTGAACCGGAAAATCATTAAGGCAGGCAACGAAAAATCCAACTCCTTGGTTGAATTGACTCAAAAGTTT
ATCTTACGGAGAAAAGCTAAAGACATCAATACGAATTTCCTTCCGCCAAAAACTGAACTGATACTGATGGTTCCA
ATGACTGAATTACAACAGGAGCTATACAAGGATATAATTGAAACTAACCAAGCCAAGCTTGGCTTGATCAACGAC
AGAAACTTTTTTCTTCAAAAAATTTTGATTCTTCGTAAAATATGCAATTCACCCTCCCTGCTGAAAGACGAACCT
GATTTTGCCAGATACAATCTCGGCAATAGATTCAATAGCGGTAAGATCAAGCTAACAGTACTGCTTTTACGAAAG
CTGTTTGAAACCACCAATGAGAAGTGTGTGATTGTTTCAAACTTCACTAAAACTTTGGACGTACTTCAGCTAATC
ATAGAGCACAACAATTGGAAATACCACCGACTAGATGGTTCGAGTAAAGGACGGGACAAAATCGTACGAGATTTT
AACGAGTCGCCTCAAAAAGATCGATTCATCATGTTGCTTTCTTCCAAGGCAGGGGGAGTGGGGCTCAACTTAATT
GGAGCCTCACGCTTAATTCTTTTTGATAACGACTGGAATCCCAGTGTTGACATTCAAGCAATGGCTAGAGTGCAT
CGAGACGGGCAGAAAAGGCACACCTTTATCTATCGTTTGTATACGAAAGGCACAATTGACGAAAAGATCCTACAA
AGGCAATTGATGAAACAAAATCTGAGCGACAAATTCCTGGATGATAATGATAGCAGCAAGGATGATGTGTTTAAC
GACTACGATCTCAAAGATTTGTTTACTGTAGATCTTGACACGAATTGTAGTACACACGATTTGATGGAATGTTTA
TGTAATGGGCGGCTGAGAGATCCGACTCCCGTCTTGGAAGCAGAAGAATGCAAGACAAAACCGTTGGAGGCCGTT
GACGACACGGATGATGGTTGGATGTCAGCTCTGGATTTCAAACAGTTATCACAAAAAGAGGAGACAGGTGCTGTG
TCAACAATGCGTCAATGTCTGCTCGGATATCAACACATTGATCCAAAGATTTTGGAACCAACAGAACCTGTAGGG
GACGATTTGGTATTGGCAAACATCCTCGCGGAGTCCTCAGGCTTGGCTAAATCTGCATTGTCATCTGAAAAGAAA
CCCAAGAAACCAGTGGTGAACTTTATCTTTGTGTCAGGCCAAGACTAA

Downstream (in bold start codon previous gene)

gctggaagaacggaactttaatcgaaggaaaaattaaatgtcaaagtgggtcgatcagga
gataatccatgcttcacgtgatttttcttaataaacgccggaaaaactttctttttgtga
ccaaaattatccgatctgaaaaaaaattacgcatgcgtgaagtaggatgagagacttact
gttgaactttgtgagacgaggggaaaaggaatatcctgatcgtaaacaaaaaagttttcc
agcccaatcgggaacatctgcgaagtgttggaattcaacccctctttcgaaaatgttcca
ttttacccaaaattattgttattaaataatacatgtgttactagcaaagtctgcgctttc
catgtctcagattcggcagataacaaagttgacacgttcttgcgagatacgcatgaatct
tttggctgctttttgtgaaagagaaatggtgccatatattgcagacgcccctgaaagatt
agtgtgcggctgagtctttttttttctcaaccagcttttctttttattgggtaccatcg
cgcacgcaggactcatgctccattagacttctgaaccacctgacttaatattcatggacg
gacgcttttatccttaaattgttcatccattcctcaatttttccgtttgccctccctgta
ctattaaattacaaaagctgatctttttcaagtgtttctctttgaatcgctc**atg**catca
taaagaaagactcatagatcatatttcttcggagtctaacttctccctttccacttcgtc
gatgccgtccttcagccatgaatcaaatcaatcgccgaacccaatgctgatcgaacaagc
ttgcgattcctgtcgtaagcgcaaattaaggtgttctaaggaatacccaaagtgttccaa
atgtgtcacccacaaatggagttgcgtttattcaccaaggacagtgagatctccgttgac
gagagctcatctgaccaaagtggaaaatcgtgtacgaatgt

AA

MYKKPMNRPFKPPRMQEATNKPVKPAQPAKIFKRKPEVLSPGTPEVAQTRITNSVNVLWRKKTLKKNKTWDGDGI
LVITGDQLIFKCDPLGGSNLREQGRRSKSSKSLEGVIAIGNYEIELDGEISDVKPASKKHRKIEHEVSSEKESIT
SADKKSSSLPFKSPMTNSVTRQPQKKSAPPYKSDYEHSIVLSQESETSKMVVVDPLLSQHLRPHQREGVKFLYNC
LDGSKELPHKGCILADEMGLGKTLTTITLIWTLLKQNQVDHKRPAVKKVLVVCPVTLIHNWKREFRKWLGMNRVS
ILEMSSASNVEEDKRSVINFGRTRVYQILILGYEKLQKLTNEISQINVDLLVCDEGHRLKNSNNKVMKSLTSFQI
PRKIILTGTPIQNDLNEFYNIINFVQPGIVGDFASFNRDYMRPILQAREINCLNRKIIKAGNEKSNSLVELTQKF
ILRRKAKDINTNFLPPKTELILMVPMTELQQELYKDIIETNQAKLGLINDRNFFLQKILILRKICNSPSLLKDEP
DFARYNLGNRFNSGKIKLTVLLLRKLFETTNEKCVIVSNFTKTLDVLQLIIEHNNWKYHRLDGSSKGRDKIVRDF
NESPQKDRFIMLLSSKAGGVGLNLIGASRLILFDNDWNPSVDIQAMARVHRDGQKRHTFIYRLYTKGTIDEKILQ
RQLMKQNLSDKFLDDNDSSKDDVFNDYDLKDLFTVDLDTNCSTHDLMECLCNGRLRDPTPVLEAEECKTKPLEAV
DDTDDGWMSALDFKQLSQKEETGAVSTMRQCLLGYQHIDPKILEPTEPVGDDLVLANILAESSGLAKSALSSEKK
PKKPVVNFIFVSGQD

SEQ0094

P. pastoris homolog of *S. cerevisiae* **Polδ/Pol3 (YDL102W)**

Catalytic subunit of DNA polymerase delta; required for chromosomal DNA replication during mitosis and meiosis, intragenic recombination, repair of double strand DNA breaks, and DNA replication during nucleotide excision repair (NER)

S. cerevisiae null mutant: inviable; resistance to sulfanilamide: decreased
S. cerevisiae conditional mutant: cell cycle progression: arrested; cell cycle progression in S phase: arrested; cellular morphology: abnormal; heat sensitivity: increased
S. cerevisiae reduction of function: mutation frequency: increased
S. cerevisiae unspecified: mutation frequency: increased; resistance to phosphonoacetic acid: decreased

Chr 2-1, 0163

5' region (in bold start codon next gene)
tctcaaccccaccatacttcttccatattgtgatgaccgatagcaggtacaggaagaaca
atgaatttggaagctgaaggtagaatgagctgggacgagagaactgctccctgatgagtt
gctttgcctcatcattcaagttagctttttgaactatgttgatgaatgccattgcgaacg
tttccatatcgttgtgtctgaagttgaaagtgaattttaatacttgcaaatccaactctt
cattttctccaagtctgtcgaccaacttttccatcaaatcttttctttgaatccactcaa
tggcaaactcatgtgataaggagttcactagtttcagcataggaagaggtgatcttccct
gtttggaagtcttgaacaatacttccaccagtgcattcctgagaagaaaagaaatagagc
catgcgaggacaacagaagcttactaacgtgatccagttggacctgtaaccatgcctcat
cccgtttcaaaaccaggttcggacagaaaaatgcaaacatattcaaggctgtaacgtcgc
tcccacaatgcagccatatcttttcaaattctaagctgatatgacttcgtaataagaaag
tacggatataactctgaattatcctcacactcttatcttgcaatttggcctccttccgtc
tattcctctccaattcggctctatgaagcagttgctctctgctttcaagattattccttc
tcccactaaggtttacattccgcctcttggttgatccgtcgaagttaaa**cat**gttgccac
ggagttagtgggtcttgtgataaggatttcaggtggttaatgtgattacttaagtctaat
gttcataagcgcccttcgcttatgtaatactggagccggtgcccggcaaagcacagcgat
aaaataaaaccaaaacgcgtgcatctaaggtaacctgattcaccttaaagttgtcttctt
agcgcgtccttctctcgtttattagaaatcttcctcacact

ORF
ATGCCAGAACAGCTCAGCCAGGAGTCAGAGGCAAAGAGAGTTAAGCTGGAAGACTACCACGAGACGGCACATTCC
ATCCCAGTTGCCAATCGTGACATCAATGGCTATCAGTTGCCTGGCCACCAGAGCAACTCGTTCCATACTAAGATG
ACCTCCTCTCAAGTGGACAATGAACCTTCAACCTTCATCAAGGAGTTGGAGGACATGGACAACGAAGCCACGTCC
AAGAGTGAGGAAATCCAAATTCAGAACTGGTCTAGACCCGAATTGCCAAAATTTGATCCAAAGAAAGATGACATT
ATATTTCAGCAGTTGGATGCTGAGGAGATATCTCATGTAAATGGGCACTCTTCAGCAAGATTTTTTGGTGTCACG
GACGCGGGTAACACGGTGCTCTGTAATGTCACTGGTTTTGACCATTATTTCTATGTTGGCATGCCCAAGGCTCTC
AAGGAGTCTGATCTACCGGGGTTCACCAACTACTTGAAAGATGAATACGTTGGGGTCAACACGATTGAAATTTCC
TTCAAAGAGTCAATTTGGGGTTACAACAATAATACAAAGCTACCTTTCTTGAAGATCATAATGAGTAACTCCCGA
TTGATCTCCAGAGTGAGAACTGGGTTTGAGAATGGTAGTATTCAATACAACGATGCGTTCCCTCCCAACGGGTCC
ATCACATTTGATAATATTCAGTACCTTTTGAGAATGATGGTGGACACAAAAGTCAGCGGTATGTGCTGGATCAAG
TTGCAAGCTTCAAAATATACTCTGGTTCCACCTTCTCAACAGGTATCAACTTGCCAATTAGAGCTTTCTGTTGAC
TACAGAGATCTTGTTTCACTACCCATAGAAGGAGAGTTCAGTAAGATTGCGCCATTACGAATTTTGTCCTTTGAT
ATTGAGTGCGCTGGACGCAAAGGTATTTTCCCAGAGCCAGAGCATGACCCGGTCATTCAGATTGCTAATGTCGTT
TCTCGCACAGGTGAATCTACTCCATTCGTAAGAAACGTATTCACAGTCAATTCATGTGCCCCGATTGTGGGCTCT
CAAATATTTTCTTTTGAAAAAGAGGAGGAATTGCTTCGAAAATGGAGGGACTTTATTATCAAAGTTGACCCAGAT
GTCATCATCGGTTACAACATCGCCAACTTTGATTTCCCATATTTGATAAATAGAGCTGCTACATTAAAAGTCCAA
GATTTTCCATTTTTCAGTCGTGTTAAGACCTCAAAGCAAGAAATCAAAGACTCAACATTCAGTTCCAAAGCATAT
GGAACGAGAGAAAATAAAGTCGTCAACATTGATGGAAGGATGCAACTAGATTTGCTTCAATTTATACAAAGAGAG
TACAAGCTGAGATCGTATACCCTGAATTCTGTCTCTTCACACTTTTTGGGCGAACAAAAAGAAGATGTTCAACAT
TCCATCATTACAGATTTACAGAACGGGGACGCGGAAACCAGAAGGAGGTTGGCAGTTTATTGTTTGAAAGATGCG
TATCTTCCTTTACGACTGGCTGAGAAGTTGATGTGTTTGGTCAATTATACCGAGATGGCTAGAGTCACAGGAGTT
CCTTTCTCATTCTTGCTTAGTAGAGGTCAGCAAATCAAGGTCATTTCTCAACTGTTCCGTAAATGCTTAGAATTA
GATATCGTCATTCCAAACATGAAGAGCGATTCGGTCAATGAAGAGTACGAAGGTGCTACAGTCATTGATCCTGTT
CGTGGTTACTACGATGTTCCAGTCGCAACTTTGGATTTTGCTTCTCTGTATCCCTCCATCATGATGGCCCATAAC
TTGTGTTACACAACGTTACTGAATAAACCAACAATAGACCGGCTGGGTCTTGTCAAAGATAAGGATTACATTTTA
ACACCCAACGGAGACTTTTTTGTCAAAGAGGATCTAAGAAAAGGTATTTTGCCAACAATTTTGCAAGAACTTTTA
GCTGCTAGAAAGAAGGCCAAATATGACCTGAAAACAGAGACTGATCCGTTCAAGAAACAGGTCCTTAATGGTAGA
CAACTTGCTTTGAAGATTTCGGCAAACTCTGTCTATGGTTTCACTGGTGCTACTATTGGTAAACTGCCTTGTTTA
GCTATCTCATCCTCTGTGACCGCGTTTGGTAGAGAAATGATTGAAGCCACCAAGCAAGCTGTTCAAGAAAGATAC
ACCATTAAAAATGGTTACGATCATGATGCCCAGGTAATCTATGGTGATACTGACTCCGTCATGGTGAAGTTTGGT

GAAAATGATCTTAGAAAATGCATGAAGCTTGGAGAGGAAGCTGCAGCTTTTGTGTCAACCATCTTCAAGAACCCG
ATTAAACTAGAATTTGAAAAGGTATACTTCCCGTACCTACTGATCAATAAGAAGAGGTATGCTGGGCTTTACTGG
ACAAATCCTGACAAATATGACAAGATGGACACCAAAGGTATAGAAACTGTGCGAAGAGACAATTGCAGATTGGTT
CAGAACGTTATTACCAAAGTTTTAGAGTACATGCTGGAACTACGAGACGTCGAAGGTGCCGAAAACTTTGTCAAG
CAAACTATCGCCGACCTTTTGCAAGATAGAGTAGATTTATCTCAACTAGTGATTACAAAAGCTCTTTCAAGACAA
GACTATGCAGGAAAGCAAGCACATGTTGAGCTGGCAGAACGAATGCGCAAGAGAGATGCTGGTTCAGCTCCAACG
CTAGGTGACAGAGTGGCCTACGTCATTATAAAAACCAGCAGCACTAAGAATTATGAGAAATCTGAAGATCCCCTA
TATGTCCTGGAGCATTCGTTACCTATTGACACTAACTATTACCTCGAGAACCAGTTGGCCAAACCGTTGACCAGA
ATCTTTGAACCTATTTTAGGAGAACGACGAGCTAAGGCATTGTTGAGTGGATCACACACCCGAACTGTCTTTGTC
GCAGCTCCAAAGACTGGTGGCCTTATGAGGTTTGCAAAAAAGACCAATACTTGCAAGAATTGTAAAAGTCCGTTA
ACAAAGTCCAATTGCAAATCCCCCAATGGAAATGCTGCGCTGTGTAACAACTGCATGGATAAAGCTGGAAGCCTA
TATGGAGAAGCGCTGTCACTGATGAACTCTTTGGAAACCAAATTCAGTAAGTTATGGACCGAATGTCAACGTTGT
CAAGGTTCACTCCATCAAGAAGTGCTTTGTTCAAATAATGATTGCCCTATATTCTACATGAGAACGAAGGCCAAA
AAGGATATCATACATCAGGCAGAGGAGTTAGAAAGGTGGAATGATATGGTGTGGTAA

Downstream (in bold stop codon previous gene)

acttatttattaaattcaatacattaatgaaaataaagtcataattgggacaaactcatc
ctacgagtttgatctacttggtttct**cta**ttcgggactcgactcggcgattaacgagctt
cctgtcagttcgtcaattgatgacgcatcatagatggctgagatcgaaggtggcttgaca
acaggttgattcctggaaggcttcttggctttggcaactgtagcgcctaaggtggtagta
gtactaacagtttctcggttttccagcggtatgtgggcaaaatttgagcggccgtctaaa
aagttaccgtggatttcgttcaccctattgttaactgagtacagctgcccaccagtgctg
ttaggagacataacagcgttgatatcaaagtcttcagagaatgtggtgttcctactgcta
acgtttttcttggatagttccttgagttgattgatgttgtgaatacacttgcttattgtc
tgtaacaatacagttggcttcaaaggcttggaaagatactcatccatttgggcctttagg
cagcgttctctatcacccaacatagcgtgagctgtcaaagcaatgatcggcgtcctaaaa
ccgagtacgtcaactggattactctttctctcccagtctctgatattcgcagtagcttcg
aatcctcccatgataggcatctgcacatccataagaacaacatcatatctattctttttg
actgcattaaacgcctctagaccattttctacaacttccactttgtgtccatgtttctct
aatattctaacagcaagcttctgattgatcaaattatcctcggctaggaggatattgtat
gattgtgatccgtcggaattggatggcgaggctcttgattccaaagcaggcatcaagata
gatgctaaatcgataacggagcatggtgtgtttccgtatgatgtgataccaagatccaag
cacactctcatgttcagtgacggtactgacggattcagcaa

AA

MPEQLSQESEAKRVKLEDYHETAHSIPVANRDINGYQLPGHQSNSFHTKMTSSQVDNEPSTFIKELEDMDNEATS
KSEEIQIQNWSRPELPKFDPKKDDIIFQQLDAEEISHVNGHSSARFFGVTDAGNTVLCNVTGFDHYFYVGMPKAL
KESDLPGFTNYLKDEYVGVNTIEISFKESIWGYNNNTKLPFLKIIMSNSRLISRVRTGFENGSIQYNDAFPPNGS
ITFDNIQYLLRMMVDTKVSGMCWIKLQASKYTLVPPSQQVSTCQLELSVDYRDLVSLPIEGEFSKIAPLRILSFD
IECAGRKGIFPEPEHDPVIQIANVVSRTGESTPFVRNVFTVNSCAPIVGSQIFSFEKEEELLRKWRDFIIKVDPD
VIIGYNIANFDFPYLINRAATLKVQDFPFFSRVKTSKQEIKDSTFSSKAYGTRENKVVNIDGRMQLDLLQFIQRE
YKLRSYTLNSVSSHFLGEQKEDVQHSIITDLQNGDAETRRRLAVYCLKDAYLPLRLAEKLMCLVNYTEMARVTGV
PFSFLLSRGQQIKVISQLFRKCLELDIVIPNMKSDSVNEEYEGATVIDPVRGYYDVPVATLDFASLYPSIMMAHN
LCYTTLLNKPTIDRLGLVKDKDYILTPNGDFFVKEDLRKGILPTILQELLAARKKAKYDLKTETDPFKKQVLNGR
QLALKISANSVYGFTGATIGKLPCLAISSSVTAFGREMIEATKQAVQERYTIKNGYDHDAQVIYGDTDSVMVKFG
ENDLRKCMKLGEEAAAFVSTIFKNPIKLEFEKVYFPYLLINKKRYAGLYWTNPDKYDKMDTKGIETVRRDNCRLV
QNVITKVLEYMLELRDVEGAENFVKQTIADLLQDRVDLSQLVITKALSRQDYAGKQAHVELAERMRKRDAGSAPT
LGDRVAYVIIKTSSTKNYEKSEDPLYVLEHSLPIDTNYYLENQLAKPLTRIFEPILGERRAKALLSGSHTRTVFV
AAPKTGGLMRFAKKTNTCKNCKSPLTKSNCKSPNGNAALCNNCMDKAGSLYGEALSLMNSLETKFSKLWTECQRC
QGSLHQEVLCSNNDCPIFYMRTKAKKDIIHQAEELERWNDMVW

Other genes involved in recombination

SEQ0095

P. pastoris homolog of *S. cerevisiae* **SRS2 (YJL092W)**

DNA helicase and DNA-dependent ATPase involved in DNA repair, needed for proper timing of commitment to meiotic recombination and transition from Meiosis I to II; affects genome stability by suppressing unscheduled homologous recombination

S. cerevisiae null mutant: UV resistance: decreased; competitive fitness: decreased

resistance to cisplatin: decreased; resistance to hydroxyurea: decreased; resistance to 5-fluorouracil: decreased; resistance to methyl methanesulfonate: decreased; transposable element transposition: increased; viable

Chr 3, 0219

5' region (in bold stop codon previous gene)

aagactaagaagattgaaattattcttccgaagtatcaccaagtaaccacttatttgcca
agttctgttagaaagtacataccgacacacagagttctcatagaaaaggagatcaaaaca
gggccaagatacatatttggttatcaccctcatggggtagtttccctggggatcactgga
gcttttggcaccaatggttgtaacattggcgagttactaccaggaatcagaatatattta
ttaaccctcatcactcaattcaaacttcctctattgagagattacttaatggcattgggt
atttcttctgtttcgaaacgtaatgtgactgcactgataaaacgaaatcagtctgtctgt
attgtcattggaggtgcttcggagtctctattatccaaaccacatactattgatattgtc
ctgaaaaaaaggaaaggctttgtgaaagtcgcactagagctgggtgacactgagttagtt
ccagtatttggttttggagaaaacactgcctataatgtttttgacccaagtgtatctggc
aagtcttgctctgtcctaaattacgtgcggaagcaaatgtgtgggtttcaattatggtta
aaacaacactttggctttacctttccatttttttcatgctaggggtgttttcaatcacgac
tttggccttctaccatatcggaaacctatcaacttggtcatcggtagacccatcccggtt
ccttacattcattcaccaacccaagaacagattgaccattaccattccctatatgtcgaa
gaactgaaacgagttttgagcagaataaggagaggtttaatgctggatccttggagcta
cgaattgtcgag**tga**agttgagccagtatctttttttaattgttgtgtaagatattttcat
gtcatgtttgtagaccctaagtaaatgcatattttcattatcgcgtcacctactcgtgaa
ctttttataagtagggcattgtcaaccactttaggaaggag

ORF

ATGCTTGCATCATTAAACTCTCAGCAGTATGAGGCAGTGACTGCAGATATTCATAAGCCATTGCAAATAATTGCT
GGACCTGGAACTGGGAAGACCAAAGTTTTGACCTCACGCGTAGCTTATCTTTTGATAGAGCACAAGATCCCTCCT
TCGAATATTATTGTCACAACCTTTACAAAGAAGGCAGCCAACGAGATGGTAGAAAGACTGCAGGTTCTATTTGCA
AAAGAGAACCTTGATCCCGACAAATATCTAAAAGGTCTCTTGATAGGCACTTTTCATTCAATTTGTGTCAAGTTA
TTGAGAAGATATGGAACCCAAATTGGGTTGACTCCCAAGTTTAAAATAGCCGATGAACGGGATTCTAACACTATT
TTAAGCAAAATTTTGACCACCTTGTCAGATGATGTTGTTAGATTAAAATTTCATCCATCTTCTGTTTCAGTTGAC
ATCTATCAAAAGGATCCCAAAGAGACTGATAAATTGCATGGATATGATCTGAAAAAGGCAAAATCGCAGATCTCC
AAACTTAAATCGTTGACTACCACGCCGCAGACATATAGTACTACTTCTGCAAGATCATTTGATCTTGCTGTGATT
TACTCTTTATACCAGGAGAGTCTACGAAAGCATAATCTTATTGACTTTGATGATTGTCTCTTGTACACTTATTTG
CTTCTTTCAGAGACAAGAAAAAAGGGACGCGTTTTACTAAAGGAACTGAAGCACGTCCTGGTAGATGAATTTCAA
GACACGAACAAAATTCAAATGGAGCTGCTCTATCTTCTAGGACAGTGCACTGATGATCAAAATTTTGAGGACAAT
GTCACGGTTGTGGGAGACTCTGATCAAAGTATCTACATGTTTCGAAACGCAGTGGCTTCAAATTTTGAAAACATG
ATCATTCACTATCAAAACTGTCAGGAAATAACCTTGAAGGAGAACTATAGATCGACTTCTAATATCCTTAATTTT
TCTGAGCAGATCATGACCCAACAGAAAGACAGGAAACAAAAGTCATTAAGATCACAAATTTCACATCTAAATATC
CCGATTGTTTATGTGAGGACTCAGAATACCTTCAAAGAAGCAGACCACATAGCTTCAGAAGTAGCTAAGCTTGGA
TTAATGTCCGGTCTATTAACCAGCTATAACGATATTGCTATCCTTTTACGTTCTTCTTTTCAGTCTCGTGTCATC
GAACAGAGTTTGGTGACCAATCGCATACCTTATGAAGTCGTACATGGAAGAGCATTCTGGGAACTAAAGGAAATT
CGTGTAGCAATGGACTATTTGAGAGCAATTGTTAACAATTCAGACAACCTGGCCATTCTCCGGACTTTGAACTTC
CCCCCAAGAGGAGTAGGAAACAAAACTCTACAAAGACTAGAAGAATTGATTGACAAGTCCGATTGTCCTGTCATC
GAAACCCTTGAAATGGCAATACATGACAAGAAGAATGGATTTTTGAATGTGAAATCCGAAAGTGGAGTGAGACAA
TATATAGACATTCTAAAGGAATGTGGTAGACTACTAGATTCTACATCTTTAGATTGCGAAGCAAGTTTGGAAGCC
TGTTTTGATCAAATTGTCCAACTAGCAAACTTGAAAGAGATTAAGAAAACGCCTGAGACTTCAAAGGGAACTCAT
GATTCAGAAAATGCCCTCAAAGCTGTAGAGGAAAACTTGAATGAGTTCAAGCGACAATTAATAGAGTATAAGCCA
ATCATTGACCAGATCATTACTGCAGAAGCTGGTGATACTCCGCTCACTACTAGGGATGTCCTTTCCAGCTTTTTG
GATTCAATCGACTTGTATGAAACATCGAACAACGAGAACAAAAAAGATTCAAAAAAGGGAAGAGTGACGATATCT
ACAATTCATGGTGCGAAGGGTTTGGAATGGCCGGTAGTATTTATTCCAGGTTGCACCGACGGTGTTATTCCCTCC
AAATGGACTTTAGAAGATGATGGTCCCGATTCGTCTACAAATTTGGACGAAGATAGACGATGTTTTTATGTTGCA
CTAACAAGAGCAAAATCAATGCTATATTTATCATCATACAAAGAGAGATCAACCCCCTGGAGTGGAGAAGAATCA
AATGAAGTATCGTGCTTTTTAAAACCAACCCTTAAAATGTTTCAAAATCAGCTGGTTACTTCAATCAGCAACATG
TCCAAAAACTCTTGGAGGGATCTGTACGAAGCAATGGGTAAAGAGTACAAGCCAGTGTCAAATATTTTAACAGAA
TCAAACAGATCAAGAAATCACATCGTCTCAACCAAAGATTTTCTCAAAAATAAGCAATTCAAATCACCTTTGCTA
ACTAAAAGCACCACAGCGCCATCATATGTACCTGACAGATCAAATAACAAAAAGAGAAGACTAGGAGTCAGAAGG
CGAATATAA

Downstream (in bold stop codon next gene)

actaaa**tta**ttcaaacaccagattagcctgcttgtcgttgactttgaccaactttgctaa
ttccgttgattcaaatttaaaatcggaactagttttacagaatgggtgacagatcaatga
ctgggcagttggcctctcatcaggatcgactgtaaaacatttattgatgaaatcaattcc
atctgaagaaatatatgccttaacttcatccgaaattggcggagcagattttgaccttcc
taagttgtacattgcgcttataacttcgaagttagaccaaggacgtttgccagcaaacat
ctctaacacgacacaccctagcgaccaaacatcaactttggcactataaccctgcttttt
gtcgttgaccacattatcaattacttcaggagccatccaaaatatagacccttgcattga
catttcagcatcattcgagtagatattgcgacttctcttggatataccaaagtctgttat
tttacaggtcccatccgtttctagcaacaaattatcgcccttcaaatctctatgcaatat
accacgactgtgaatataagcaagaccttctaatacctgcttagtcaagaatctgattac
ttcctctggaaacctgccaaatactcgaagacaccaaccaactgacccccctgcaacata
ttctaagaataatgtataggtttgagctctttgttcaaatcctagatactgaacaatgtt
gtcatgatccaagtcacttagggtatcgacttcagaaagaatggcgtcgacaacttcctt
gaagacggagctgttatgcccgctcgaaggtagatcaacttgtttcacagccatcatctc
accagttgtcacgttaagagccaagaagacttttccatatgtcccttaccaatcatttc
acctcttacccaggaaaattgtttataatctccattaccatccttcaatttggatagatt
tcgggcatctccaggcttcatttctacaacctttctacccc

AA

MLASLNSQQYEAVTADIHKPLQIIAGPGTGKTKVLTSRVAYLLIEHKIPPSNIIVTTFTKKAANEMVERLQVLFA
KENLDPDKYLKGLLIGTFHSICVKLLRRYGTQIGLTPKFKIADERDSNTILSKILTTLSDDVVRLKFHPSSVSVD
IYQKDPKETDKLHGYDLKKAKSQISKLKSLTTTPQTYSTTSARSFDLAVIYSLYQESLRKHNLIDFDDCLLYTYL
LLSETRKKGRVLLKELKHVLVDEFQDTNKIQMELLYLLGQCTDDQNFEDNVTVVGDSDQSIYMFRNAVASNFENM
IIHYQNCQEITLKENYRSTSNILNFSEQIMTQQKDRKQKSLRSQISHLNIPIVYVRTQNTFKEADHIASEVAKLG
LMSGLLTSYNDIAILLRSSFQSRVIEQSLVTNRIPYEVVHGRAFWELKEIRVAMDYLRAIVNNSDNLAILRTLNF
PPRGVGNKTLQRLEELIDKSDCPVIETLEMAIHDKKNGFLNVKSESGVRQYIDILKECGRLLDSTSLDCEASLEA
CFDQIVQLANLKEIKKTPETSKGTHDSENALKAVEENLNEFKRQLIEYKPIIDQIITAEAGDTPLTTRDVLSSFL
DSIDLYETSNNENKKDSKKGRVTISTIHGAKGLEWPVVFIPGCTDGVIPSKWTLEDDGPDSSTNLDEDRRCFYVA
LTRAKSMLYLSSYKERSTPWSGEESNEVSCFLKPTLKMFQNQLVTSISNMSKNSWRDLYEAMGKEYKPVSNILTE
SNRSRNHIVSTKDFLKNKQFKSPLLTKSTTAPSYVPDRSNNKKRRLGVRRRI

4 *P. pastoris* genes with high expression levels

Based on the micro-array data set by Graf *et al.*, 2008 (BMC genomics 2008, 9:390)

55 features selected with a 20x higher signal than GAP1. Of these, 35 could be linked to a *S. cerevisiae* gene and for 32 of these a *P. pastoris* homolog could be identified.

SEQ0096

P. pastoris homolog of *S. cerevisiae* **GPM1 (YKL152C)**

Tetrameric phosphoglycerate mutase, mediates the conversion of 3-phosphoglycerate to 2-phosphoglycerate during glycolysis and the reverse reaction during gluconeogenesis

Chr 3, 0826

5' region (in bold start codon of previous gene)

ccgaaagaaggtttatctgactgttgcgcaccaccgaaaccaaaagcgggtttagctgca
gaactttcagtggccggcttgttgccaaagagactgttactgcctccatcggtcttggta
gcgcctgaaccaaacgaaaaagcaccactgggattattgtttcctccaaaagacgcggtc
gtgtttgatgctgttgaggccgcaggcgcttgaccaaaactaaacgcttttgatgtattg
tttgtaggggctgttgcgttgttagaagctggctggttgttcccaaaggaaaaagcacct
gttgatccagtgttagatccacctgcggcattcccttgagctgcggagccaaaaattgat
ccttggtttgagctctgtgtaccaaataaagatgaattggccttagctccgtcagcagga
gcgtttgcttgtccaaaggggcctccagacgacgtttgcgtgttgttggcattactagta
ttggcaggctgatttccaaacaatgagggcttggaggcgttctgggagttaccaaatgag
aaccctgtatttccttggccagagcctgagaaactgaaccctgatccaga**cat**tatcaat
accttgggttattagtagtgtccgttatttttctgtttaggttacgattttgccagattt

tttgggaggagggaaacaaaagaaccagtgctacacgacctttaagtgccatcaggcatc
ctgttttctcgacctcatctcatcacatccgtcagtctgagctttcagttctcagttttc
gattgactcttgccctgctgcgcgcacaccataccctggctccctctcatgcttctggcg
ttaccccgggaatcgtacatccatgccgcgaatcccggacaggactcagacggatttcac
tattcctagtgggcacaacctccatatataaggatacttctgccccctagcttaagtgcc
tctattttgtcagtaacaactttcaattacacaaacaaaca

ORF

ATGGTTAAGTTAATTTTGGTGCGTCACGGTCAGAGTGACTGGAATGAAAAGAACCTCTTCACCGGATGGGTCGAT
GTCAGATTGTCTCCAACTGGAGAGAAGGAGGCAGCCAGAGCCGGTGAATTGTTGAAGGAGAAGAACATTCACCCT
GACATTCTGTACACATCCAAGCTGAGCAGAGCCATCCAAACCGCCAACATCGCTTTGGACAAGGCCGACAGATTG
TCCATTCCCGTCGTGAGATCCTGGAGATTGAACGAGAGACACTACGGTGCTCTTCAAGGAAAGGACAAGGCTGAG
ACCTTGGAGAAGTACGGAGCTGAGAAGTTCCAGACCTGGAGACGTTCTTTTGACATTCCTCCACCTGAGATTGAG
GATGATTCTGAGTTCTCCCAAGCCAACGATGAGAGATACAACGACGTTGATCCAAACGTTTTGCCAAAGACTGAG
TCATTGGCCCTTGTCATTGACAGATTGTTGCCATACTGGCAAGATACCATTGCCAAGTCTCTTCTGGAGGGTAAG
GACGTCATGATCGCTGCCCACGGTAACTCCCTTAGAGCTTTGGTTAAGCATTTGGACAAGATTTCTGATGCTGAT
ATTGCTGGTTTGAACATCCCAACCGGTATTCCATTGGTCTACGAGTTGGACGAGCGTTTGAACAAGACCAAGGAA
GCTTACTACTTGGATCCAAAGGCTGCTGAGGAGGGAGCCAAGAAGGTTGCTGCTCAAGGCCAAAAGAAATAG

Downstream (in bold stop codon of next gene)

gtttcttagtaagtaataaggtgaatatattaactgttcatagtgttattgttggtgtga
ataaggggtgattttaatagaagacaacacaacacatcttggggattttaggtcatattg
gctgaattattgtgaggaaattcacctagttacttcaatgaataaccaccaaatgataaa
aaaaaagtacttttgagttatttcagaataccatacattgaccatattgctttgtggaa
caactgatattgaatttcgtttatataatagagttgttttgcaagtaataaacaactaag
aagaagaaattcacgaaaagtccgtgatagtttagtggtagaaccaccgcttgtcgcgcg
gtagaccggggttcaattccccgtcgcggagttttttttgtcattgatattccctcttac
cagacgaaatctaccaacccaaatccttctgctcttataataaataattgatgagtacaa
atcctataatctacta**cta**tgcctccaattgcttactttgctccctagcaagtctcttttttct
ccttttcttccttccggttgtacttgacatcctcagggttgaatcctccctccgatgagc
cccagtatgcccactccggagttggagctaggtaatccatatgtgtgaacttttcaccgg
attcaagtaacaggaaaaacttttttatagtcaatatgagtgtaccattggtcgttgatgt
aatatctcttcgaggcaatcaaaatacagcatgaatgagcatgttcggcagcaattccgt
actccaaacctctggaattaagctcatcattcaaagatcttacaaaattttgacattcct
catagaaggggatgttttgcattgtcaaaggatttccatttccagaggatgatccgcaga
aggttgcacctttgatttcaatcaatgagggcaatgcgctctccaccatgtcagcatatc
ccttgaaatcacccatattaaatcctttcacaagagtcaac

AA

MVKLILVRHGQSDWNEKNLFTGWVDVRLSPTGEKEAARAGELLKEKNIHPDILYTSKLSRAIQTANIALDKADRL
SIPVVRSWRLNERHYGALQGKDKAETLEKYGAEKFQTWRRSFDIPPPEIEDDSEFSQANDERYNDVDPNVLPKTE
SLALVIDRLLPYWQDTIAKSLLEGKDVMIAAHGNSLRALVKHLDKISDADIAGLNIPTGIPLVYELDERLNKTKE
AYYLDPKAAEEGAKKVAAQGQKK

SEQ0097

P. pastoris homolog of *S. cerevisiae* **TEF2 (YBR118W)**

Translational elongation factor EF-1 alpha; also encoded by TEF1; functions in the binding reaction of aminoacyl-tRNA (AA-tRNA) to ribosomes

Frag B. 0052

5' region

tggtgggtatttgacaggttggggagcaaataagtgatgatgtcccatgaaagtagaaaatggc
tagtagaaggcaaaaatttgaaattcttagagtcaaatagttagactccaagttctaatc
cacatttggtcagtttcatagcatccagagcttttgccactggtgaacatatctacccat
tgcgatgcaacaagtcactgaaagcctaaaacggagattcccctatcttacagcctcgtt
caaaaaaactgctaccgtttatctgctatggccgatgtgaggatgcgctcatgcccaaga
gtccaactttatcaaaaacttgacccgtcatacaggctctagatcaagaagcaaacttaa
tctcagcatctggttacgtaactctggcaaccagtaacacgcttaaggtttggaacaaca
ctaaactaccttgcggtactaccattgacactacacatccttaattccaatcctgtctgg
cctccttcacctttttaaccatcttgcccattccaactcgtgtcagattgcgtatcaagtg

aaaaaaaaaaattttaaaatctttaacccaatcaggtaataactgtcgcctcttttatct
gccgcactgcatgaggtgtccccttagtgggaaagagtactgagccaaccctggaggaca
gcaagggaaaaatacctacaacttgcttcataatggtcgtaaaaacaatccttgtcggat
ataagtgttgtagactgtcccttatcctctgcgatgttcttcctctcaaagtttgcgatt
tctctctatcagaattgccatcaagagactcaggactaatttcgcagtcccacacgcact
cgtacatgattggctgaaatttccctaaagaatttctttttcacgaaaatttttttttac
acaagattttcagcagatataaaatggagagcaggacctccgctgtgactcttcttttt
ttcttttattctcactacatacatttagttattcgccaac

ORF

ATGGGTAAGGAAAAGTTGCACGTTAACGTCGTTGTTATTGGACACGTCGATGCTGGTAAATCTACCACCACCGGT
CACTTGATCTACAAGTGTGGTGGTATTGACAAGCGTACCATCGAGAAGTTTGAAAAGGAGGCTGAAGAGCTCGGT
AAGGGATCTTTCAAGTACGCCTGGGTTTTGGACAAGCTTAAGGCTGAGAGAGAGAGAGGTATCACCATCGACATC
GCTTTGTGGAAGTTCGAGACTCCAAAGTACCACGTTACCGTCATTGACGCTCCAGGTCACAGAGATTTCATTAAG
AACATGATTACCGGTACTTCCCAAGCCGACTGTGCCATTTTGGTCATTGCTTCCGGTATTGGTGAGTTCGAGGCT
GGTATCTCCAAGGATGGTCAAACCAGAGAGCACGCTCTTTTGGCTTTCACCCTGGGTGTCAAGCAATTGATTGTT
GCCATCAACAAGATGGACTCCGTCAAATGGTCTCAAAAGAGATACGAGGAGATTGTCAAGGAAACTTCCAACTTC
ATCAAGAAGGTTGGTTACAACCCTAAGACTGTCCCATTCGTCCCAATTTCCGGATGGAACGGTGACAACATGATT
GAGCCATCTTCCAACTGTGACTGGTACAAGGGATGGGAGAAGGAGACCAAGGCTGGTGGTGCTACCAAGGGTAAG
ACCTTGTTGGAGGCTATTGACTCCATTGACCCACCATCCAGACCAACTGACAAGCCTCTGAGATTGCCTTTGCAG
GATGTCTACAAGATTGGTGGTATCGGAACTGTGCCAGTCGGTAGAGTTGAGACCGGTGTCATCAAGGCTGGTATG
GTCGTCACTTTCGCCCCAGCTGGTGTCACTACCGAAGTCAAGTCGGTCGAGATGCACCACGAGCAATTGGAGCAA
GGTGTCCCAGGTGACAACGTTGGATTCAACGTCAAGAACGTTTCCGTCAAGGAAATCAGAAGAGGTAACGTCTGT
GGTGACTCCAAGAACGACCCACCAAAGGCCGCTGAATCTTTCAACGCCCAGGTCATTATCTTGAACCACCCAGGT
CAAATCTCTGCTGGTTACGCTCCAGTTTTGGACTGTCACACCGCTCACATTGCTTGTAAGTTCGACGAGTTGATT
GAGAAGATTGACAGAAGAACCGGTAAGAAGACTGAGGAGAACCCTAAGTTCATCAAGTCCGGTGACGCCGCTATC
GTCAAGTTGGTCCCATCTAAGCCAATGTGTGTTGAGGCCTTCACTGACTACCCACCTTTAGGAAGATTCGCTGTC
AGAGACATGAGACAAACTGTTGCTGTCGGTGTTATCAAGTCCGTTGTCAAGACTGACAAGGCTGGTAAGGTCACC
AAGGCTGCTCAAAAGGCCGCTAAGAAATAG

Downstream (in bold stop codon of next gene)

attgcttgaagctttaatttattttattaacataataataatacaagcatgatatatttg
tattttgttcgttaacattgatgttttcttcatttactgttattgtttgtaactttgatc
gatttatcttttctactttactgtaatatggctggcgggtgagccttgaactccctgtat
tactttaccttgctattacttaat**cta**ttgactagcagcgacctcttcaaccgaagggca
agtacacagcaagttcatgtctccgtaagtgtcatcaaccctggaaacagtgggccatgt
cttttgctccttcaaaaatggcaatgggtaggctgcctcctctcttgtgtatcctctctg
ggaccactcagcgtcacttgtgctaataatatcttttaggttgtgtggggagttgtgcaa
gattgcaccatctgtttctccgttttctactttacggatttcttctctaatagagatcat
agagtcaatgaatctgtctaattcttctttgtattcggattcagttggttctaccatcaa
tgtgcccggaatagggaacgacatggtaggagcatggaatccataatcttgcagacgctt
ggccacatcaatggcctcaattccgaatttcttgaatggtctaagatcaataatgaactc
atggccacagtacttgtcatcgccacctgtgaataaaatttcataatggttctttagctt
ggctaccatatagttggcatttagaattgcaattattgacgaatactgcaaattgcttcc
acccatcattttaatataggcataactgataggcaaaatggaagcactaccaaagggagc
tgcaacaactggattaatactctgctcagtctggttggtggtagaaattatcggatgact
gggtaggaagggagtgaggtgctccttaacacaaatagggccaacacccggtcctcctcc
gccatgaggaatggcaaaagtcttatgtaaatttagatgac

AA

MGKEKLHVNVVVIGHVDAGKSTTTGHLIYKCGGIDKRTIEKFEKEAEELGKGSFKYAWVLDKLKAERERGITIDI
ALWKFETPKYHVTVIDAPGHRDFIKNMITGTSQADCAILVIASGIGEFEAGISKDGQTREHALLAFTLGVKQLIV
AINKMDSVKWSQKRYEEIVKETSNFIKKVGYNPKTVPFVPISGWNGDNMIEPSSNCDWYKGWEKETKAGGATKGK
TLLEAIDSIDPPSRPTDKPLRLPLQDVYKIGGIGTVPVGRVETGVIKAGMVVTFAPAGVTTEVKSVEMHHEQLEQ
GVPGDNVGFNVKNVSVKEIRRGNVCGDSKNDPPKAAESFNAQVIILNHPGQISAGYAPVLDCHTAHIACKFDELI
EKIDRRTGKKTEENPKFIKSGDAAIVKLVPSKPMCVEAFTDYPPLGRFAVRDMRQTVAVGVIKSVVKTDKAGKVT
KAAQKAAKK

SEQ0098

P. pastoris homolog of *S. cerevisiae* **RPS2 (YGL123W)**

Protein component of the small (40S) subunit, essential for control of translational accuracy; has similarity to E. coli S5 and rat S2 ribosomal proteins
Chr 1-4, 0589

5' region (in bold start codon of previous gene)
cctcctgagaacggacagcagcgctggaggcggcctctttaacggtggcggcgaagtcaa
caagggtagttttgattttgcctaacaactgttgcaggtccttctcagtctttgggtggc
tcttcttgaactcgtccgatgcatccttgacgaccttctcagcgtcgtcggcaatctcgt
cgacaacttgcactgatccatgagggtcttgggtttgttcgacttctggaactggttttg
gcttgccagcggaaactccttctgggggtgtggcagcttcggcgtaggtactatcagtgt
taatatcggggattggttgatgtgcgagacgaggcaggcgggagatccgtctgcctcgtc
caactgcgatagctcttacttacctcatgatgcttgtaaaaaagttaactgaaagatgga
aatgggaggggggaaaagaattgtggtcaaatccacgtcttgcgataacctcatacatttg
ctgcatgattgggggatcatcgcaatttcggtccttgtgacacacgcgagctttccgcct
ggggttagagcgcggagcaagcatactatcaagcaagaaaaaatggtatggagaaaccta
attggttaagatatatgaaaactacgacggcttcatacgtggtatctctggttgagctac
tacgaaccattttccccttcaaaccctttggcccatgc**cat**tcatgccttgcctctctc
tcaagcaactaagcaatcaagcaatttcccgccttgctgcacatgactgttcggaaatcg
gagacccaaacaccacttgttatctatgcacgtgatttttatccagggcaataaatactc
actttgcttcaaaacgcttggggcgcgcgagcggcaggctgggaaaaaaataatctcag
actttcaaaagactctcctctttaatcattgaataccgtcgatcgaaaaccaccaccat
cggctttccacgtacattccgacttttcagtgtagttaat

ORF
ATGTCTCAAGCACCAGCATCAGCTCAACAAGGATCAGCCCCACAAGGCCAAAGACAATTTGGTGGCGCCCGTGGA
GGTAGAAGAGGACCAAAGAGAGGACCAAGAAGAGGTGAGGAAGAGAAGGGTTGGACCCCAGTCACCAAGCTTGGT
AGATTGGTAAAGGCCGGTAAGATCACCTCCATCGAGGAGATTTACTTGCACTCTCTCCCAGTCAAGGAATTTCAA
ATCATCGACCACTTATTGCCTGACTTGAAGGATGAGGTCATGAAGATCAAGCCTGTTCAAAAGCAAACCAGTGCC
GGTCAAAGAACCAGATTCAAGGCTGTTGTCGTTATTGGTGACTCCAACGGTCACGTTGGTTTGGGTATCAAGACT
GCCAAGGAAGTTGCTACCGCTATTAGAGCCGCTATCATCATCGCTAAGCTGTCTATCATTCCAGTCAGAAGAGGT
TACTGGGGTTCCAACTTGGGTGCCCCTCACTCTTTGGCTTCTAAGACCTCTGGTAAGTGTGGTTCCGTTTTGGTC
AGAATCATTCCTGCCCCAAGAGGTTCCGGTATTGTTGCCTCTCCAGCCGTTAAGAAGATGATGCAACTGGCCGGT
GTTGAAGATGTTTACACTTCCTCCGGTGGTTCCACCAGAACTTTGGAGAACACTTTGAAAGCTGCCTTCATTGCC
ATTGGTAACACTTACGGTTTCTTGACCCCAACTTTGTGGGACGACTACGAGTTGTCCGAGTCTCCTCTCGACGTT
TACGCCGACGAGGCCTCTGCTTCTTACAGAAAGAGATATTAA

Downstream (in bold stop codon of next gene)
atagtcaaatattaatctatttcacctgttcaaactttacttaatgtacaaatgtggtag
tta**tta**gttttgcaacggaacttgttccataatctggtcctctgggacagcaaactgtct
ttcactagtagcgccagtttcgggagtccacacagcattagtcaccggtgcaccagcact
aatctcacgaccttctgggtgtttaaatgggcagttagggttgcggcatccagctgcaaa
cttacaatcctcatcaattggatgagtgaaaaaacagtttggtctggtacaactgttgcc
ttcacgacacagtacaggagtagttgcgtgacgtcttgggcacttgtaattacggcatga
tttaccaaatcgacattgttccaaagccctctgttgtttttgttgtttctcttcttcggt
gatcttgtgttcaggtgatcgatgagcctttggacagtccggattagagcattgcttgtc
ttctgggcaccattctagcaccgtgactcgagcagcattgttagctggagttgggtgacc
aaatggacacatttccttggaacacatattgccaaacttacatataccgatttgaccttg
ctccaaagttttcttcaactgctcctgaacttgacgttccacttgtttcatcaatttggc
ctggctaaacgtctttctttcttggaattcttttctggtcttctccagctcggcaatcag
tgcatcatcctcacctggatgaagataattacaagttccagcgggatttggacagttggg
gaatgcaaaacaggtctttgtaggatggccgaatgggcactgcttgttcttgcaatgtgg
gaagtattggcaacggtttcccttcttggaaacaaaactagatccctcgacctcctggtc
aggattggatggttgggctatgttcattgtcgaatccattgcacttaaaaaattctcctc
attcttaagggcaaaattggatctggcattattgtttctac

AA
MSQAPASAQQGSAPQGQRQFGGARGGRRGPKRGPRRGEEEKGWTPVTKLGRLVKAGKITSIEEIYLHSLPVKEFQ
IIDHLLPDLKDEVMKIKPVQKQTSAGQRTRFKAVVVIGDSNGHVGLGIKTAKEVATAIRAAIIIAKLSIIPVRRG
YWGSNLGAPHSLASKTSGKCGSVLVRIIPAPRGSGIVASPAVKKMMQLAGVEDVYTSSGGSTRTLENTLKAAFIA
IGNTYGFLTPTLWDDYELSESPLDVYADEASASYRKRY

P. pastoris homolog of *S. cerevisiae* **ENO1 (YGR254W)**

Enolase I, a phosphopyruvate hydratase that catalyzes the conversion of 2-phosphoglycerate to phosphoenolpyruvate during glycolysis and the reverse reaction during gluconeogenesis; expression is repressed in response to glucose

Chr 3, 0082

See SEQ0073

SEQ0099

P. pastoris homolog of *S. cerevisiae* **PET9 (YBL030C)**

Major ADP/ATP carrier of the mitochondrial inner membrane, exchanges cytosolic ADP for mitochondrially synthesized ATP; phosphorylated; required for viability in many common lab strains carrying a mutation in the polymorphic SAL1 gene

Chr 4, 0210

5' region (in bold stop codon of previous gene)

gttgttgttcctaaggctctgagagttctgagattgaagccaggtagaaaattcaccact
gtcggaaagttgtctacttccgtcggttggaaatacgagtctgttgttgagaagttggag
gagaagagaaaggctgaggaagctgagtaccaggagaagaagagagcttacacccagaga
ttagacgcagctagtgccgagtttgcccaaaccgaggagggaaagcagttggctgccttt
ggttac**taa**atagtaaagtagggtatcttcaagtaatagtatactaaccatctgaaataa
ccaccgtcctgtagttttttttcgatatcgaagagcctatgctagtactgtggatttgcg
ctccatccaacatctgtgcgcaaactaaaacttccgagactgacatctaccatcgctaga
ccctaagtaaaaccaatctcgcgtccgaacttttaaatttcagtccttaaaacttcagag
cattggttgtagtttccggatctgaggggtcgtattggagtcaagagacggagctgcctc
cacagcgcgaaacgtcaaccccaacaccaacctgaatttgcaatcaccatggggacaagt
ttcagcagtcaatgggcaattcagacgttgatacggtacccatttgctaagctcaatgac
gatccatccaacttcagagaaaggcctttctctggtatgctctggtattcattcgtcttt
tatcactctcgttgcacaatgcccgggtactcccggaacaagggagtcttccagccaagc
tgtacagagtgaaaaatagaaatacacctttgcaatcaagacgcgcgttggccaatcaca
agacttaatcggtgcaaagaaggattaccaaatttttttttcccaaaatcgctatataga
aataatggaggaaaaagggttaatataaaggagaattcccccgtttttctccccttttct
tttcttcttcaggctttcttacaaatctataatattccaaa

ORF

ATGGCTGACAACAACAAGTCTAACTTCTTCGTCGACTTCATGATGGGTGGTGTTTCCGCCGCCGTCTCCAAGACC
GCTGCTGCTCCAATTGAGAGAGTCAAGCTTCTTATCCAGAACCAAGATGAGATGCTTAAGCAGGGCCGTCTTGCT
AAGAAGTACGATGGTATCGCTGAATGTTTCAAGAGAACCGCTGCTGATGAGGGTATTGCTTCTTTCTGGAGAGGT
AACACTGCTAACGTTATTCGTTATTTCCCAACTCAAGCTCTGAACTTTGCTTTCAAGGACAAATTCAAGGCTATG
TTCGGTTTCAAGAAGGATGAGGGATGGTGGAAGTGGTTGGCCGGTAACCTTGCTTCCGGTGGTTTGGCTGGTGCC
ACTTCTTTGTTCTTCGTTTACTCTTTGGATTACGCCAGAACCAGATTGGCTAACGATGCCAAGGCTTCCAAGGGT
TCCGGTGAGCGTCAATTCAACGGTTTGATCGATGTCTACAAGAAGACTTTGGCTACTGATGGTATTGCTGGTTTG
TACAGAGGTTTCTTGCCTTCCGTTGTTGGTATTGTTGTTTACCGTGGTTTGTACTTCGGTTTGTACGACTCTTTG
AAGCCAATCGTTTTGGTTGGTCCTCTTGAAGGTTCTTTCTTGGCTTCCTTCCTGTTGGGTTGGACCGTTACTACT
GGTGCTTCTACTGCTTCTTACCCATTGGACACTGTTAGAAGAAGAATGATGATGACCTCTGGTCAAGCCGTTAAG
TACAACGGTGCTTTCGACGCTTTCAGAAAGATCGTTGCTGCTGAGGGTATCAAATCTTTGTTCAAGGGTTGTGGT
GCTAACATCCTGAGAGGTGTCGCTGGTGCCGGTGTTATCTCCATGTACGACCAATTGCAAATGATTATGTTCGGA
AAGAAGTTCAAATAA

Downstream (in bold start codon of next gene)

gccaattagtttgaagtgagattttatttcattcctgttaatattatatactagagtata
ttttaaattaattgttcatgaacttgccaaattatgttagttttgtgtaaacaatcttag
gctatccaatttagttctacttttggtagatttcctgttttggtaaattacaaacaacaa
tgatttgacttatattctattcggaattttacttatcaccttgtacagtttgtggggatt
tccggac**atg**aagaagtttggactggaaaaggacggagatggaatcttcaggctcaaacg
aaggtcctcctctcccaataagaagggcaaaacaacattgcctcctttctcgtcacctcc
aaatctagcccatgacagtccaacagggcccccaatattacctactctcagaaccagagt
ttctgaaactgacactttggtagatggtatcagtgatacggtcaacattccctctccaca
agctatgaatgcgagggccgggtcccccattagacccatgctaggaagccctacgaaatc
attcaatgatgaggccgatgcggaactggcccagattataatgctgctatcaaagattgt

aagactggggttcaactttgtcaaagatcttacgttggaaagcagctttatttcagtgga
gaattataccactctgcatctccccagtttcatggtgggtattgctcttgcattgttttt
cacctctggaggctcagatgaatcaagtagccgatcaagaagtcacgagtcatcaacatc
atcggcctttgggattcttcgtaagtttgtgatgggatttctggttatttccctcgtggt
gttctttgttagtttccagagtggggatcaggcatccaattccggatcggatgaactgcc
ccttgtgaactacaagcaggaaacaacagctccttcccctagcacacatccctccggtcg
gagttcaccaaagagatattcagtcttcaaggaagacctga

AA

MADNNKSNFFVDFMMGGVSAAVSKTAAAPIERVKLLIQNQDEMLKQGRLAKKYDGIAECFKRTAADEGIASFWRG
NTANVIRYFPTQALNFAFKDKFKAMFGFKKDEGWWKWLAGNLASGGLAGATSLFFVYSLDYARTRLANDAKASKG
SGERQFNGLIDVYKKTLATDGIAGLYRGFLPSVVGIVVYRGLYFGLYDSLKPIVLVGPLEGSFLASFLLGWTVTT
GASTASYPLDTVRRRMMMTSGQAVKYNGAFDAFRKIVAAEGIKSLFKGCGANILRGVAGAGVISMYDQLQMIMFG
KKFK

SEQ0100

P. pastoris homolog of *S. cerevisiae* **RPS5 (YJR123W)**

Protein component of the small (40S) ribosomal subunit, the least basic of the non-acidic ribosomal proteins; phosphorylated in vivo; essential for viability; has similarity to E. coli S7 and rat S5 ribosomal proteins

Chr 3, 0762

5' region (in bold start codon of previous gene)

aattctggaggttccaaataaattatggtacgagttcatagtcaatgggttcccactctt
gggagacacatctggtttcaacacctcactccgaagagcacaggcaaacctcaaggcgga
acgaatcaaggcagaagcacgggaagcctgggtgggggggaacaacggttttggacaacgg
gtcgtcctcaagaacgaaaaaaggatttctaggaagggtcttaccacgaagttcaccgta
agtaccaggagttgtggtgctgattccttcggcatccagatagcagtcattcttctcgta
ggatctgaccaatcgagcctggatggcccggatggtcttgctgttgatgaagtcactgtt
ctgtgccaccaacaaatttgtattcctctggtgacactataggtttcagactgtgcaata
tttgctcggcagtagaaaccaaggctgggacgggcagcttcggcagggcctgctcgtggt
caaaagttccattatc**cat**ctttgtgtgaattagcgggatgtgtcaaactgaagcggagt
tggaggagaattgcagggaagggctgaggatgcaagtatataaagggttaggtcttcccc
actatagatgatccattctcggctctagggttagtatgataagatcacgtgaggatggtc
ggattgagtgaccgatttttcgaaggatgaaggggggttggatgttgaatggatgttgga
tggaggacaacagaggagaggaagatgaggggaaaaagggagaaacaaaaataaaaaaaa
acatagctgacctccgaacatacgttgagtttgcacgtggagtttctgacgagagcagtc
accgtaaacggcaacaacagtcacgcgaaaacttgcccatccaagtccccagttttgtta
ggcgtgcgcgcgagagctccgcgacttttgcaatttttcctcaataaaaattcattttct
tcaaacatcaagtgttcatacaacaagtaggagatccacaa

ORF

ATGTCTGAAACTTACGACGAACAAGTTGTTGCCGAGCCAGTTGAGGTCCAATACGTGGAACTGTCCACCAAGATC
CCTGCCGATATCAGAGAGGGAGCTGAGGAGATCAAGCTGTTCAACAAATGGTCTTTCCAGGATGTCACCGTCAAG
GACATCTCTTTGACCGACTACATTCAAATCAGACAGCCAGTCTTTGTTTCCCACACTGCCGGAAAGTACGCCGCC
AAGAGATTCAGAAAGGCTCAGTGCCCAATCATTGAGAGATTGACTAACTCTTTGATGATGAACGGTAGAAACAAC
GGTAAGAAGCTGTTGGCCGTTAGAATCGTCAAGCATGCCTTGGAAATCATTCACGTTTTGACTGACCAGAACCCT
CTGCAAGTTCTGGTTGACGCCATTGTCAACACTGGTGCTAGAGAAGACTCCACCAGAGTTGGATCTGCTGGTGCC
GTTAGAAGACAAGCTGTTGATGTTTCTCCTTTGAGAAGAGTCAACCAAGCTATTGCTCTGTTGACCATTGGTGCT
AGAGAGTCTGCTTTCAGAAACATCAAGTCTATTGCTGAATGTTTGGCTGAGGAGTTGATCAACGCTGCTAAGGGA
TCATCTACTTCATACGCTATTAAGAAGAAGGATGAGTTGGAACGTGTTGCCAAGTCCAACCGTTAG

Downstream

gtagtgttcatttgtataatggtaatcgatactatgaaagacgtgtaggttgtgaggtga
aaggaggtgtaagatagagcggagctgtcaatgtggatgggagaacggtaaagtggccag
gacgttgtggtggttttggccttgtttaagtttcactactcgtttaacaccggctcatcc
tagcttatgttttgagtaactgttcaccggccatatttcaactctatgaattataacact
attaattacttgtttggagcgcagcaagaggttcttaaaatcataatttcgctttatata
ttaatcacaaggtgatgctaacaattaccaaaaagagaactaaactcttgcatctattat
ccactgtgcctaaagacacataaaaaatagtctagatccttctctaatacaatagaaatc

ctctttactgtgttaatctaccacaacatctgcagccgatgctttcttacctgccaagct
caaatttttcacttcaagcgccgatcaagttctcaagctggtaatcactagttgtttcca
cctcagcagctggccaatctcatctcatgcggcagtttttcctcctttctcatgcatcc
aagacaactggccccctgctaaacaataacataaaaaaaaaaaaaaaaaaaaaaaaca
accgcagaaatttgcaacaggacagaattcaataatagacgcaggtgaataaaccaagag
attcagtttattagcccaaatccttcccttgaccgccttagactttaaacaaacgctacg
aaatcgtgatacaaccctaatcgttggtgtcaccggctggttaccaagggtttacccttt
ctttttgcactgggccttttcttccaaatcgctggacacaaacttctgcaatcatgcgtc
gccctagcataaaagacgcttctttctctacatgatctcaatcacataggcaatagctta
ctaagatgcaaatggcccttattcatgctgcattggggcac

AA

MSETYDEQVVAEPVEVQYVELSTKIPADIREGAEEIKLFNKWSFQDVTVKDISLTDYIQIRQPVFVSHTAGKYAA
KRFRKAQCPIIERLTNSLMMNGRNNGKKLLAVRIVKHALEIIHVLTDQNPLQVLVDAIVNTGAREDSTRVGSAGA
VRRQAVDVSPLRRVNQAIALLTIGARESAFRNIKSIAECLAEELINAAKGSSTSYAIKKKDELERVAKSNR

SEQ0101

P. pastoris homolog of *S. cerevisiae* **PDC1 (YLR044C)**

Major of three pyruvate decarboxylase isozymes, key enzyme in alcoholic fermentation, decarboxylates pyruvate to acetaldehyde; subject to glucose-, ethanol-, and autoregulation; involved in amino acid catabolism

Chr 3, 0188

5' region (in bold stop codon of next gene)

acatattcaaataccaagcaaataaacgcaaagagcaacatattaggataaaccaacagg
ttaagtatagagctaaagcattgctgagcaatatttcggatcaacatcaacagaatagat
cgtcaccacgaagtatatccaactattccccaagaatccaggcttatccctcaagaatgg
cctctccatctccttcaatgaagatggcattgacaaactccgtcagcttgcaggatacaa
tggatgcggaactagaagcgttggcaaacgaacaatattacgtcatgctagatatcctta
agggattttgcgatctttcatttgacatggtaaacatattttctatccagatccctgagg
tattacatctcctcttcggctttggtgcagggtcactggccttaacgggagtcgtatcca
aaacgagacaagaactaataaataaacagata**taa**gggacaagcacacgattacccaatc
acttgatatgcaccaatttgttccgttgtttatgccatatttaccgaatttttcttcccag
gtttttccgaatggacatctgtagtccacttttggttatcataatcgtcccacaagtcg
tggatttaaccagaacctagtaattttaagttcgctattaatcactcagaatggtctcac
cttgctattggccaagtctggagtcgccagctaccacctcagaggctacatagacctccc
aatgtcatctcctcagtgcgctcttcaatctcgtgtcttttccgttaaaactccgttcgt
ttcaccctatactgcccctggttgtgcagctcttaccacttcgcgccgctactatccgta
gtggtcgagccgcatcaatatcacgttgaaatagaataactccctacaaaagccgcacgc
aaccatcaaatctatataaggaacctcaaatatctagcaacatcttttcaatttactaca
acatattcgttaatcatcaatcaattagctagtacacaaca

ORF

ATGGCTGAAATAACACTAGGAACTTACATCTTCGAACGTCTGAAGCAAATCGACGTCAAGACTATCTTTGGTGTT
CCAGGAGATTTCAACTTGGCTTTGTTGGATCACATCTATGAAGTTGAGGGCATGAGATGGGCAGGAAACGCCAAT
GAATTGAACGCTGCTTACGCTGCAGATGGTTATTCCAGAGTGAAGTCAATGGCTGCTTTGATCACCACCTTCGGT
GTCGGTGAGCTCAGTGCCGTCAATGGTATCGCTGGTTCCTTTGCCGAGCATGTCGGTCTGCTGCATATTGTCGGT
GTCCCAGCTATCTCTTCCCAGGAGAAGAAGCTGTTGCTTCACCACACCCTCGGTAACGGAGACTTCGGTGTCTTC
AAGCGTGTCTTCAAGAACGTCTCCAAGTCTGCCAACTTTATCTCTGATATCAACGAGGCTCAAGATATGATCGAC
GGTGCCATCCGTGAAGCTTTCATTTACCAACGTCCTATCTACCTTGGACTGCCAACTAACCTTGTCGAGATGAAG
GTTGACAGAACTCGTTTGAACACCCCAATTGACTTGAAGCCAGTTCCAAACCCAGTCGAGGCTGAGGAGGAGGCA
TTACAGTCCATCTTGGAACTGATCTCAAAGGCTTCAAAGCCAGTTATTCTGGTTGATGCCTGTGCATCCAGACAC
TTCTGTCAAAAAGAGGTTGACCAACTCATTGACGTCACCAACTTCCCAGTCTTTGTCACTCCAATGGGTAAAGGT
GGTGTTGACGAACAAAAGCCACAGTTTGGTGGTGCTTATGTGGGTTCTCTGTCTAACCCAGACATTACAGAGTTC
GTGGAGAAGGCTGATCTTGTCATCTCTGTTGGTGCTCTTCTGTCTGACTTCAACACTGGATCTTTCTCTTACTCC
CACTCCAAGAACATTGTTGAGTTCCACTCTGACTACACTCAAATCCGTAGTGCAGTGTTCCAAGGCGTTCAGATG
AAGGCCCTATTGCAAAAGCTGTTACCACTTGTCGGTAAGGCATCTAAGCACATTACTGCTCAAGTTCCTCCAAAG
ATTGCTCCTCCTATCGAAAAGGGTGCCAGCGAAGATTTGACCCAGGACTGGTTGTGGAGCAACATCAGCAAGTTC
CTCAGAGCTGGTGACGTCGTCATCACAGAGACCGGAACCTCTGCTTTCGGTATCGTTCAATCAAGATTCCCAACC
CACGTATCTGGTATTTCCCAGGTCCTGTGGGGTTCCATTGGTTTCTCCGTCGGTGCTACCCTAGGAGCTGTCTGT

GCTACTGAGGAGCTCGACCCCAACAGAAGAGTCATCCTGTTTGTTGGTGATGGTTCTTTGCAACTGACCGTCCAG
GAGATTTCCACTATGGCTCGCTGGAACTTGAAGCCATACCTGTTCATCCTTAATAACGATGGTTACACTATCGAG
AGATTGATTCACGGTGAGAAGGCTTCTTATAACGATATTCAACCATGGGACCACCTTAAGCTGCTAGACACCTTT
AAGGCCAAGAACTCTGAAAGTACTAGAGTCTCTACTGTCGGAGAGTTGACCAAATTGTTCAAGGACCAAGGTTTC
AACAAGCCTGACAAGATCCGTGTCATCGAGATCATGTTGGAGACTATGGATGCTCCAATTTCCTTAGTCAAGCAA
GCTGAAATCACTGCCAAGACCAACGCAGCTTAA

Downstream (in bold start and stop codon of previous genes)
atgtgttcacgatcataattgtttttagtctgtatatgttacaatattaacgtccattt
gaaatcttatccttcgcgtaatcttctctggagatttcttcctacttatctacttgaaca
ttgtacctacgcaccaaa**atg**gaacaccctgcttttagtctttcaggtctgtctttagtt
ggaggagtcatgggatagtaagtgtagacattgctttttgttggagtgtactaactcgtc
agtgcccgtaaaggttcccttccttccctaatagctggagtcacttttcggcaatttat
gccggagctggatatttgctgaagaataacatggaatatggggttcaaactgctttggga
gcctccactctgatgtttattgccggaatcaaccggggaattgtgtcaagatataagaaa
cccgttccaattgttctgacaattctaggagccgcttccactctgtactatgctaagaag
tatgaagaattttaccttt**taa**agctattgaaacttatctaagtcgtatgaatataaactt
ttgt**tca**ccatccatggaaaggaacccaaatatctactactgtatcttcctcaattacaa
aataccatgggaacgctgcagcagcgatgcaattatgttgggggaccactttgatcttgg
tgcctagaggtataaactcgcaattagaatcttctttaaaagggaacaagatgccgtgtt
cctggctaagctttccaacaatccaatctccataggtagaagggtcgactactattccat
gacctggaatggaactaaactctcgagccaatgcaattactccagcatttatcaactgtt
ctccgggattcctattgtttctattagggtaatgtgataccacttctgctagaacatgag
ccgacacgtcatctagtgacacgcagtttgttgcaacctgttgtaagtcacaacatggat
agtttccagcgtgtaactccagcactccctctagttggcct

AA
MAEITLGTYIFERLKQIDVKTIFGVPGDFNLALLDHIYEVEGMRWAGNANELNAAYAADGYSRVKSMAALITTFG
VGELSAVNGIAGSFAEHVGLLHIVGVPAISSQEKKLLLHHTLGNGDFGVFKRVFKNVSKSANFISDINEAQDMID
GAIREAFIYQRPIYLGLPTNLVEMKVDRTRLNTPIDLKPVPNPVEAEEEALQSILELISKASKPVILVDACASRH
FCQKEVDQLIDVTNFPVFVTPMGKGGVDEQKPQFGGAYVGSLSNPDITEFVEKADLVISVGALLSDFNTGSFSYS
HSKNIVEFHSDYTQIRSAVFQGVQMKALLQKLLPLVGKASKHITAQVPPKIAPPIEKGASEDLTQDWLWSNISKF
LRAGDVVITETGTSAFGIVQSRFPTHVSGISQVLWGSIGFSVGATLGAVCATEELDPNRRVILFVGDGSLQLTVQ
EISTMARWNLKPYLFILNNDGYTIERLIHGEKASYNDIQPWDHLKLLDTFKAKNSESTRVSTVGELTKLFKDQGF
NKPDKIRVIEIMLETMDAPISLVKQAEITAKTNAA

SEQ0102

P. pastoris homolog of *S. cerevisiae* **HNM1/CTR1 (YGL077C)**

Choline/ethanolamine transporter; involved in the uptake of nitrogen mustard and the uptake of glycine betaine during hypersaline stress; co-regulated with phospholipid biosynthetic genes and negatively regulated by choline and myo-inositol
Chr 3, 0511

5' region (in bold start codon of previous gene)
tagcatctccttccctaggtttatgaaagtactcctcaagaatcttggttgggatatgtt
ttctcaataccagatccaaatcgatatcgtctgacatgagaattaaacccttagcatcac
catagggaatgggtttgtgctccttaattgcctgatagttggccctatactgatccaacg
gaagggaatcaaa**cat**gatgaactgtttgagggaacaaccgtgtttgtaaacagacagag
ctagggtgaatacgcgttgaaatgaagcgagggtaggggaacgctggaaggaatctttgg
gaggagccctttgggaaaacgctggagggagctctgttggagatgcgaaatgcgaagcag
aaaaaccaacctgcaaactattcccaaagtggcagcagggcgctatggtaaaagcaccat
tatcaccatttaggaagtaatctttacaacataatttcaaccttacaacaagtctggtc
cacctggattatcttctagataagctttacgacctcacaccatctccccaatatctgcct
tttcgttcaaacaaaagcagtttactggaactgaccctacttcggagtacctatctaatg
tcttggtcatttggaaaaagcttcaattttctgcacagaaaacacagtcacacaagattg
gctttcccattcgggacataaccaatatattgcgacatcaccaattaggtcccctatagt
tatcaggaaattggcagtggatacttgtacaatcatgttgcacctgtaagggtctttggt
tttcccgtttcccttccaaatttcagatttttatcggaaaaggccgcctatcttggtaa
tgccttgattcagcgtttttctttttttcttgtccgaatctcgttctcgccgaaccgtt
gcacactataaatactatctcgtatcctggtttggacacctcctctaaaactatcatacc

ttgaggtttttttttaggtaacactcgttctttctacaaaa

ORF

ATGTCAAGTAGCGTGGAAAAATCCGATGTAAACGATGTTTCCCTAACAAGAAGATTCACCAAAGATTCAACTCTT
GATAAGGATGACTTAATTCTGGCTCAATTGGGTTACGAACCTGAGTTGAATAGAAACTTTTCTGTTTGGTCTGTC
TTAGGTGTTGGCTTCGGTTTGACCAACTCCTGGTTTGGTATCTCCGCTTCTTTGGTCACTGGTATCTCTTCCGGT
GGTCCAATGATGATTGTTTATGGTATCATGATTATTGCAGCAGTATCCACCTGTATTGGTATCACCTTATCTGAA
CTGGCTTCTGCTTATCCTTCTTCAGGAGGTCAATACGTTTGGGCTAAGGTATTAGCTCCTAAATACCCAGTATTG
GCTTTCCTGTGCGGTTCATTCTCATATGCTGGTTCCATCTTCTCATCAACTTCAACTACTGTGGCTACTGTTCAG
ATTGCAGTTGCATTTTACGAACTGACCCATGCTGACTATGAATTCCACCGCTGGCACGTTTTCGTAGCTTTCCAG
ATCTTGAATTTTTTCATTTTCTTTTTCAACTGTTATGCCAAGTTTCTCCCATCGATTGCAAAGTCATCGCTGTAC
ATTTCTTTGGGATCTTTTATCGTCATCACCATCACCGTACTTGCTTGCTCCAGTGGACATTTCCAAAACGCCAAA
TTCGTGTTTTCCGATTTTGACAACCAAACTGGTTGGGAATCCGCTGGAATCGCCTTCATCATTGGTTTAATTAAC
CCAAATTGGTCCTTCTCGTGTCTGGATTGTGCTACCCATATGGCTGAAGAAGTCGCCCAACCTGAAAGGGTTATT
CCAATTGCCATCATGGGAACTGTTGCCATTGGATTTTTTACTTCCTTTGTGTACTGTATTTCGATGTTCTTCTCT
ATCAGAGATCTGGATGCATTGCTGAACACTGCTACCGGTGCCCCCATTATTTCCATTTACTACCAAGCTCTTCAA
AATCAAGGAGGTGCCATTTTCCTGGGTTTCTTGTTGTTTTTGACTGCCTGTGGATGTTTAATTTCTGGTCACACC
TGGCAAATGCGACTCTGTTGGTCTTTTGCCAGAGACAATGGTCTTCCATTTTCTCCATTCCTTTCAAAGGTCGAC
AAGAGAATGGGTATTCCATTCAACGCACATTTGTTTTCCGTCTGTTTATGTTCCTTGGTCAGTTGCCTTTACTTG
GCTTCAGACTTGGCTTATAACTCTTTGGTAACCGGATGCATCACTTTCCTCCTGTTGTCTTATGCAATTCCTGTC
TCTTGTCTTATATACAATGGAAGAGACAGTATCACTCATGGTCCATTCTGGATTGGAAAGTTTGGTTATTTCACC
TCCAGTATGACCGTTGCATGGGCATTATTTGCTCTGGTATTCTACTCATTCCCATTTGTGAAGCCAGTTACTAAG
TCCAACATGAACTACGCGGCTGCAGCTATTGTTGGTTGGTTGGTTGTTTCTCTGGTCTACTGGTTTAGCTATGGA
AAGAAGATATTTATTATGAGAAACGAAGACGACGAAGACGCTGTTCTTAAAAAAATCTTTGCCAAAACCGAGTCA
CACAGAGAAAAGGACGCCACCATTCCAGTTGAAGTAACCAAAGAACTCTAG

Downstream (in bold stop codon of next gene)

aaataatgtatataaatagaaaatgcagatttaactaatacttatttaccc**tca**gctgtc
atcgattctcattttgcttattaaatcactggctttcttgaaagattttgattccatttc
atcgggtcctgaaggggtaactggttggcttcagctgtttggtccagttctttcttttg
gtctggttctaaagtatgatctaattcttcatcgtagtcgtcttcttcgtcataataatc
atctgaatactcttccccatcatcaacttcatctacgtcgtcattttcctcttcttcatc
ttcatactcgtaatcgtagtcaaagggatcgtcatcatcatccaaggtttcatcttcgtg
atcaacgtcaatgaaatcatcatgatctttcagcttttcttcatcggatgatacagaacc
aactgattttgggggtttgattgtgctaggggcaacttcagttggctttgggtcactcaa
agtcatccgttctggtgaacctaaaaggctactacttcctttgaaaaacgatgctcctgc
tgccaatattggactacgaggagcagactgtaaacgttcctcctcctttcttcttcttag
ctcatcttcactagaatcttcatcactgagctcctcaaaattgtacaaaggctctttcca
ataattgacatcagagaactttgccgtgttccgataggtatcagaaggcatcgacttaaa
tagttcagtgttcactggtggaaagaaatgatctaccagctcactaatatgtacataaga
acttctatacccagccaattcgagtagctccatgtgaacatcaccgtctgggtttatggt
aaaaattcttgatgatggtatccctactgacctataggacaaagcatcggtaattctatt
accaaaaccggcgtaaaatggcgtttgtacatcctctgtaagacttgaagatctcaaagt
gttcgtattgatatccgtgaaatctgcagaatctgtgttcg

AA

MSSSVEKSDVNDVSLTRRFTKDSTLDKDDLILAQLGYEPELNRNFSVWSVLGVGFGLTNSWFGISASLVTGISSG
GPMMIVYGIMIIAAVSTCIGITLSELASAYPSSGGQYVWAKVLAPKYPVLAFLCGSFSYAGSIFSSTSTTVATVQ
IAVAFYELTHADYEFHRWHVFVAFQILNFFIFFFNCYAKFLPSIAKSSLYISLGSFIVITITVLACSSGHFQNAK
FVFSDFDNQTGWESAGIAFIIGLINPNWSFSCLDCATHMAEEVAQPERVIPIAIMGTVAIGFFTSFVYCISMFFS
IRDLDALLNTATGAPIISIYYQALQNQGGAIFLGFLLFLTACGCLISGHTWQMRLCWSFARDNGLPFSPFLSKVD
KRMGIPFNAHLFSVCLCSLVSCLYLASDLAYNSLVTGCITFLLLSYAIPVSCLIYNGRDSITHGPFWIGKFGYFT
SSMTVAWALFALVFYSFPFVKPVTKSNMNYAAAAIVGWLVVSLVYWFSYGKKIFIMRNEDDEDAVLKKIFAKTES
HREKDATIPVEVTKEL

SEQ0103

P. pastoris homolog of *S. cerevisiae* **GLN1 (YPR035W)**

Glutamine synthetase (GS), synthesizes glutamine from glutamate and ammonia; with Glt1p, forms the secondary pathway for glutamate biosynthesis from ammonia; expression regulated by nitrogen source and by amino acid limitation

Chr 4, 0785

5' region

acggcactagtgagaatggggcaatagaaaaaaacccatacatcacattgactagtcatt
tttacttagtgtatactagatataataataaagctacgatcatcacagtgaaatatcttt
gatggtcggtcagtattaatcttgtcacatgacctacttccgtcttcccatgactccctc
ctaaggttgtgtgcgctggtgcgtcaggttcggtccccgggaggagtctcgaccaatggc
taggtagtgataacaatcaaagtttatgagaccagtagtttgacacatagtgcatttatt
ggaaatatggggacactaccttccgccaataggagacgagaatacgactagacgtatctc
gtaaagtccctccgcctcgatctagaactgccaccatatttgaggatcaaagaccagaag
aagtgtcgtgccagtttttcagataaatcattcgccgagatcagtcgatcatcagcgttc
agtccatgagaacaaggtgctggcttgacgtatcttcaggcagccagcagcttttgttt
cagcctatcatctttgaggcttatcatcttactgataagactatgggaaaaaacccaaga
aaatgagatgttttatgtgaacgaccatttacttttttagataggcaatcttctgccccg
gtgagtctcctcattgggcaatgttgtaattacacgactaacaaatcagaggcggtaaac
gctagtctcggcagcgtcatcgaggtcagtgcatactcaacaatagaaggcgcgcaatca
ttagacaacttgatccatagcccgcagcgtttcttctccagaaagttcactataattacc
cgtagaggattggttcaacatacaagaagatcagcaaatcaccttcagtattcggcgaca
agatttatttttcactgaggtttcaaagcttttcaaatttcaatataatacttccctatt
aatcatgcatgcaccctacaagatttcgcattctttttttactacaggaattactgacc
aggtgactctaaaatagctttgaggtatataaatcacaaaggcgctgttagcagacagtt
tttttctttttcgtttcttaccttcggtcaacaatattaattcgcttcaagctacttagt
tctctccattataagagaatcttaatataaacacctatacattacaattacaat

ORF

ATGTCTTCATCCGAAATCATTGAAAACACTGCTTTGCTGCAGAAATATCAGTCCTTAGACCAGAGAGGTAAGATC
ATTGCCGAGTACATCTGGATTGATGCCAGTGGAGGAATCAGATCCAAGGGTAGAACCTTGAGCAAGAAGCCTTCT
TCCATTGATGACCTTCCAGAATGGAACTTCGACGGTTCTTCCACTGGACAGGCTCCAGGTCACGACTCTGACGTC
TACCTGAGACCAGTCTCCTACTACGCCGACCCTTTCAGAAGAGGTGACAACATTGTTGTTCTGGCCGAGTGCTGG
AACAACGATGGTACTCCAAACAAGTTCAACCACAGACATGAGTGTGCCAAGCTGATGAAGGCCCACAAGGACCAG
GAGGTTTGGTTCGGTCTGGAACAGGAGTACACCCTGCTCGACAAGTATGACAATGTCTACGCTTGGCCAAAGGGC
GGTTTCCCAGCTCCTCAGGGTCCATACTACTGTGGTGTTGGTACCGGTAAGGTCTACGCTAGAGACATGATTGAA
GCTCATTACAGAGCTTGTATTCACGCCGGAATTAACATCTCCGGTATCAATGCTGAGGTCATGCCTTCCCAGTGG
GAGTTCCAGGTCGGTCCTTGCAATGGTATCGACATGGGAGACCAGCTCTGGGTTGCCCGTTACCTGTTGCAGCGT
GTTGCCGAGGAGTTCGGTATCAAGATCTCATTCCATCCTAAGCCACTGAAGGGTGACTGGAATGGTGCTGGTTGC
CATACCAATGTTTCGACTTCCGCCATGAGAGTTCCAGGTGGTATGAAATACATTGAGGCCGCTGTCGAGAAGCTG
GCCAAGAGACATTTGGAGCACATTGCTCTATATGGTTCTGACAACGAACAGAGATTGACTGGAAGACATGAAACT
GCTTCCATGACTTCTTTCTCTTCCGGTGTCGCCAACAGAGGTGCTTCTATCAGAATTCCAAGATCTGTTGCCAAG
GAGGGATACGGTTACTTTGAGGACCGTAGACCTGCTTCCAACATTGATCCATACTTGGTCACTGGTATCATGGTT
GAGACCATCTGTGGTGCCATCCCTGATGCTGACATGACCAAGGAGTACAAGAGAGAATCTGATTAA

Downstream (in bold start and stop codon of next gene)

atagtatatcagcagctcatatctagtacttcctagcttgaagtataaaaagaaatgcct
gttaagttttgtaagccaagaaatttacgtttttttagttctaaaagttcccgctgtattc
tttacataagagattcatgatccatcagccataataaattttaaatatgtaatactacaa
acgtattaagctgccaaaccgatgacaccacaagctggtcttccaccagcgttaccggtc
ttcaaagagtcggcgtgaccacccttcctagatcgtcagttccgtcgtgaataacaaca
gttctacctagaattgaggtctcaccgatcagcttaacctggttatcggttataacaccc
ttggcaacaccttcagcgtcggtcttgacgttacccaagtcaccgacgtgtctagcttcg
tcagttggagcaccgtgggtcttaccaaatgggttgaagtgaggaccagctgaagtacaa
ccattggtattgtcaccgaactggtgaatgtggaaacctctctcagcgtttggggagtta
cctttgatgtcgtaagtgatagtggtagggctgctttcggaggattgttcgaagacaacg
gtaccgccgacggtagaatcgcctcttaagactgcaactgctttgac**cat**ttttgtgtag
atgttaaatagggaagaaacacagaaaataaagaattcgaaaggtatatgaggggcagct
ttcttttatagtttgcggtcggttctagaaaataactcccccttttttcgccactgctag
tcgagaacactaataatgactaagcgtcatatcaattactgtatatgaacaaagaagaag
atgccaattgagtcaaattatgagtaaaaaacataattgaatgaggatttcaaggaacct
ggggggaaagggccccttttatagtgtctcaaaagctctttgaccctgttggtggtgaaa
ttgtagttatttactcgtttttttgctcgttttgcgtttgc

AA

MSSSEIIENTALLQKYQSLDQRGKIIAEYIWIDASGGIRSKGRTLSKKPSSIDDLPEWNFDGSSTGQAPGHDSDV
YLRPVSYYADPFRRGDNIVVLAECWNNDGTPNKFNHRHECAKLMKAHKDQEVWFGLEQEYTLLDKYDNVYAWPKG
GFPAPQGPYYCGVGTGKVYARDMIEAHYRACIHAGINISGINAEVMPSQWEFQVGPCNGIDMGDQLWVARYLLQR
VAEEFGIKISFHPKPLKGDWNGAGCHTNVSTSAMRVPGGMKYIEAAVEKLAKRHLEHIALYGSDNEQRLTGRHET
ASMTSFSSGVANRGASIRIPRSVAKEGYGYFEDRRPASNIDPYLVTGIMVETICGAIPDADMTKEYKRESD

SEQ0104

P. pastoris homolog of *S. cerevisiae* **RPP0 (YLR340W)**

Conserved ribosomal protein P0 similar to rat P0, human P0, and E. coli L10e; shown to be phosphorylated on serine 302

Chr 1-3, 0068

5' region (in bold start codon of next gene)

caaaatatctattcctgtcatcaacgttttgaatagccaattgatcgaaaccgtcatgcc
caaccccattaactgcattgactggtaggagaacaatcaaaatgagactaccaaaaaaga
aagtcaaagtcaatattcccatgtaacgtagaaataggtacccgtcaataccacagaact
tgatcaggaatttcttatctttccgtattaaatcccagatcattcttggagcaaacacag
aaactggtttgggtctgtcttctgggggcaaaagctttgtgaaactcttgggttgataga
ttcgatggaactttgctctgagcaaaataaatgctacaacgaaacctccaaacacggccg
aattcacagccaacgaagtcagcaccgtttcggtggagatgttgacctcagtcgt**cat**tg
ttggtcaaaatcaaacgggaagtgatggtgggaagagaagaaaagaagcaagggagataa
agattcaaatccagtcttgatttcgatctgtagcatatgtatgcgatatgacgcgtttag
taatcgtcatgtaagggaatgagaaatgtaacgataaccaactaaatttagaacagagca
acaaagaaggcgcttggagtcacccttgctataacatggacatgccgatcaagtagtatt
cttttaagttaaatcttggctaaacttttatgctgctgaccttagttcttacattaggc
aacgaatcatcatcctccttgctctcgcaccaaagattttcagtgcggctgatgcagtga
ttaccatagaaacattgatccagcatcagaggaaaacgtacgcctctcgaacatccagca
tcccatcgttgcaactccaaacatctcataaattctatcttcacgtgatgagggctcacg
ccttcctcatttttcatcacttttttgtcgtccaaatcgttttcttacttcttcatcc
tttaaacagcctatcctcagtatagtaataacagagtcgaa

ORF

ATGGGAGGTATCAACGAAAAGAAAGCCGAGTACTTTAACAAGCTCAGAGAGCTTTTGGAATCTTACAAGTCCATC
TTTATTGTCGGTGTTGACAATGTCTCCTCTCAGCAAATGCACGAGGTCAGACAGACTCTCCGTGGTAAAGCTGTG
ATTCTTATGGGTAAGAACACCATGGTCAGAAAGGCCTTGAGAGACTTCGTCGAGGAGCTCCCAGTCTTTGAGAAG
CTGTTGCCTTTCGTCAGAGGTAACATTGGTTTCGTCTTCACCAACGAGGACTTGAAAACCATCAGAGATGTGATC
ATCGAGAACAGAGTCGCTGCCCCAGCCAGACCAGGTGCTATCGCTCCTTTGGATGTCTTCATTCCTGCCGGTAAC
ACTGGTATGGAACCAGGTAAGACCTCTTTCTTCCAAGCTCTTGGTGTCCCAACCAAGATTTCCAGAGGTACCATT
GAGATCACCAGTGATGTTAAGGTTGTTGAGAAAGACTCCAGAGTTGGTCCTTCCGAGGCTCAATTGTTGAACATG
TTGAACATCTCTCCTTTCACCTACGGTTTGACTGTTGTTCAAGTTTTCGACGATGGACAAGTTTTCCCAGCTAAC
ATTTTGGACATCACCGATGACGAGCTGTTGTCTCACTTCACCTCTGCCATCTCTACCATTGCCCAAATCTCTTTG
GCTGCTGGTTACCCAACTTTGCCATCCGTTGGTCACTCTGTTGTCAACCACTACAAGAATGTTTTGGCTGTCTCC
ATTGCTACTGACTACTCCTTTGAGGGTTCTGAAGCTATCAAGGACAGATTGGCTAACCCTGAGGCTTACGCTGCT
GCTGCTCCTGCTGCCGGTGAGGCTTCTGCTGGTGCCGAGGAGACTGCTGCCGCTGCTGAGGAGGAAGACGAAGAG
TCTGAGGACGACGACATGGGATTCGGTTTGTTTGATTAA

Downstream (in bold stop codon of previous gene)

ataacgattaagaactaattgtagtcaatagattaagttttccgcaaca**tta**aacgtaat
taaccacaccaaaattgggttcgtagtttggtacaactttgatgggtagtctacatagaa
actctgatccactggagtagaattccactcctctgagcaaatctcctttgttgtcggaat
agatcaactctaaatcagaagattcctccactaagatgaactttattgtgaccaaatacc
ttacgttcacaatgttagttttgaactgacctggtgagttgtaaggaatgggtaaattgg
ctgatacttgctcactgttaatagtactaacaacttttcacaaaccttttcaacaagct
ggctttgttcagtcagattctcatgtatttgccctttttcatccttgaaaatgtactggt
cacttatggtctctactctttcaagtgcgactatcaaccctgtcgttttaatatgattta
ataactgaatgttaatgcgaataatatcgttgaccctaaatactgacttgtcgaggtcaa
cttgagcaaaatttttttgtatttcgtttcaacaagaacctagtttggaattttgcgggta
aaaggctatcatattcaaaaaggttcagttcttcaaaattcttttgtaggggtttgcgtt
cagtttgtctaagggaaagaacttgtgatgggttcgaaaccaaattgataaggtttaacc
tacaattttcgttcttcggagaagtgtaatctgtaagaaactctttgtgaaactcatttt

gtatttccacaacctgattaatgttggcttgacttagtttctgcataaaagccaagaact
cgattggttttaccgctatcttgtcattagaaaaatcagactcggattcaatatcgctat
gtttcaactgagacttaatggaagcaaagctggccttttgtccatcagttacctcgttaa
cttgagcctcaatgggctgacccaagagtggtttatccaag

AA
MGGINEKKAEYFNKLRELLESYKSIFIVGVDNVSSQQMHEVRQTLRGKAVILMGKNTMVRKALRDFVEELPVFEK
LLPFVRGNIGFVFTNEDLKTIRDVIIENRVAAPARPGAIAPLDVFIPAGNTGMEPGKTSFFQALGVPTKISRGTI
EITSDVKVVEKDSRVGPSEAQLLNMLNISPFTYGLTVVQVFDDGQVFPANILDITDDELLSHFTSAISTIAQISL
AAGYPTLPSVGHSVVNHYKNVLAVSIATDYSFEGSEAIKDRLANPEAYAAAAPAAGEASAGAEETAAAAEEEDEE
SEDDDMGFGLFD

P. pastoris homolog of *S. cerevisiae* **TDH3 (YGR192C)**

Glyceraldehyde-3-phosphate dehydrogenase, isozyme 3, involved in glycolysis and gluconeogenesis; tetramer that catalyzes the reaction of glyceraldehyde-3-phosphate to 1,3 bis-phosphoglycerate; detected in the cytoplasm and cell-wall
Chr 2-1, 0437
See SEQ0064

SEQ0105

P. pastoris homolog of *S. cerevisiae* **RPL17B (YJL177W)**

Protein component of the large (60S) ribosomal subunit, nearly identical to Rpl17Ap and has similarity to E. coli L22 and rat L17 ribosomal proteins
Chr 2-2, 0109

5' region (in bold start codon of next gene)
tcgtcttttctacctttgttatcagacctacatgcagctttggactttattcgtaactcc
aattgttgacggttccagccaactaattctaaaaggtctttgtcataatctccgtcttta
ctacactcaaaaacgatgtcagcagcgatagactgttttccccaataagcgcctgttaat
tgtacagtaacattgttatcattggtttcaattgagggtttgaggtttctagagaatgca
accacttcagagactactggctccttaccattgagtagcacctctgtgataccacatatc
tgggaatcatcggggcaattctccacgtccatcttatcctttccatctttatcttgagaa
tcatatatgttgagataccagactgacttggtgattgatggcggcgtttccaattcagtg
gtcacttcgtgaaggccagcaattttgttcaaatcatagggtctcaagtgtttattagat
atgtctatggcacaaaccgaagcaataaatagaatcaacctgaaatagtt**cat**tgccggt
tgaattgattaccttgttatctttgactggtggaggtatgattaagccttgcgcgcattg
gtgatggcctgttacttggtagagagcctcctagaaattccgcgaggaaaaccttggaga
gcgttaaggggatgcggtagctaatagataaaagattatgtaatacaatctgcggaacat
atattgaggaaattcactaaatatggagagtagtgaatatttagtagaagaaactcaaat
tcccagaattagaaacctattaagagtcgccactaactcaagagtctataaaatacttcg
taacagctctctctcctactagacatcatgaataacatttacatctttacatcaactatt
ttagcacgtgatccattctagggttcttagacaacctatctatcgcgatctcatcaacaa
ctttttcactcttctgttttcttcttctacatcatcctta

ORF (intron in bold)
ATGACTCGTTACAGTGCTGTCCCAGCTAACCCAGATCGTTCTGCCTCTGCTCGTGGCTCTTACCTGCGTGTTTCT
TTCAAGAACACCAGAGAGACCGCTCAAGCCGTCAACGGATGGAACCTGAAGAAGGCTCAACAGTACTTGGACCAA
GTCCTTGACCACAAGAGAGCCATTCCTTTCCGTCGTTTCAACTCTTCCATCGGAAGAACTGGACAAGGTAAGGAG
TTCGGTACTGACAAGGCTAGATGGCCTGCTAAGTCTGTCAAGTTTATCCAAGACTTATTGAAGAACGCTGAATCC
AACGCTGAG**(gtatgtatttttttttcacaaaaagtcttctggaaaaaaaaccatgaagagatcaagacagatgtt
tctaggttgagatattatggcaatatctttcttgggttcacttgtcctctgttccatgttttttttggtcttttaa
agtcttaaaggatttttttttcagttttttgattactaacatatatttag)**AGCAAGGGTTTGGACATTGACAGTTT
GATCATCTCCCACATCCAAGTTAACCAAGCTCCAAAGTTGAGAAGAAGAACTTTCAGAGCCCATGGTAGAATTAA
CGCTTACAACTCTTCTCCATGTCACATTGAGTTGATCTTGTCCGAGCCAGATGAAGCCATTGAGAAGGCCAAGGA
AGAGCCAGCCCAGTTGACTGAAGCTTAG

Downstream (in bold stop codon of previous gene)
attttgacgggtaacatatattaatacatgtaattgattgactgttctcttactttc**cta**
actatgaattcttttagttgaatgatcaggaagtggtgtgatggatcgtttgtgatctgt

ctccacatcttccgtctcaacttcttctacctcaaactccgtatctgatgccacctcgag
tggccgtgattgctgaccaatatccgcgtccgtcgtttttccctgggtagttggctcatc
tggttcaagctctgattcatactccagatagtctccctcttcaacttctgttacattgct
cggagctgccagcagcccgacattttcaagccctttacgtttaccactgtctgccatgtc
cttcaaatgctcatacacactaagcaaagttggaactaaagtaacctttatcaagatcat
attgtaggctgttttgtcagtattttccttcagcctacttaaaagttcaataaaatctct
tggagtcacatctttggtgtgaacatcatcctcaagaattttaatatcaagctcgatcat
ttccaaacagtactgtttctctggtgaataatccaaactatttgagttgcaaaaactgcg
taagtgactgaaaaatttattgaaaggttggttcaaaatagtaatatctgagaacattac
agctaaatctgaatagtcctcatcctcctctagtgcaaatggtacaagtttaaacgtatc
attggaggggagttggatgtataccggaataggcacttggtatgatgactcgtttgcagt
tgaagagttcggattttcttctttttccttttttctatttctccttctacttcttctcctgc
tgcttcttctctttcttcttctctctcttcttctctttctttttcctctccatcatcttc
ctcatattccaattcctcctggtaccccggctgtgaagagtaatccgagcccaggtcaat
taactcaatctctgcggttttgacatgtggctggacccttt

AA
MTRYSAVPANPDRSASARGSYLRVSFKNTRETAQAVNGWNLKKAQQYLDQVLDHKRAIPFRRFNSSIGRTGQGKE
FGTDKARWPAKSVKFIQDLLKNAESNAESKGLDIDSLIISHIQVNQAPKLRRRTFRAHGRINAYNSSPCHIELIL
SEPDEAIEKAKEEPAQLTEA

SEQ0106

P. pastoris homolog of *S. cerevisiae* **RPS4B (YHR203C)**

Protein component of the small (40S) ribosomal subunit; identical to Rps4Ap and has similarity to rat S4 ribosomal protein

Chr 4, 0246

5' region (in bold stop codon of previous gene)
ccattcttccagagatcgactctttggccaggatattgttgatctcacaactttgatgtc
ccctttgcacgttcaaatttgattgaaaagttgatcatgatgggcaagttccaatcgga
tgttagaaacgccaaagagaaagagcaacagcccttcttacctcatcatgagcagcatgt
tttggtcaagctgctagaatttttaattggaacaagcttacccacggatacgtaacaaa
tgacgattttggagatgaaggagacgatactaaacataagcaaatcagtttttatgaacc
tccatgcgttattcaaagctatcaaagacactacaatgatagcatatccaatataatatt
tgacttggtgataccggatgatgaagatgaagtgattgatgtcgtgttttacgactttat
agagccatacattttgtacgctttgcgacaaatcctcaatgtcgatttagacaagagatc
atgggtgcaaccaagggaacccgtatacttacaaagtgctaccaaagttggcaacactac
attatatcagagtttcttgcgatcaattcaagtttacaatgatgttgaaaatgaatggaa
cgttggagagatgctgaagagatatgtcacagagaactgggcc**taa**ataggcgtaatttt
taataacttttgtatcatttttgacttttaatctattgcttctactttaccccagctact
aaattctcatagaaatagtccaacatctcaacattttgtcacatgacaaacattagggga
gggcccaaactttgttagcagacacgggtacacttgttcaccgaaacacgaccttccata
tttgttcatcaaacggagagagagacagtgagcagcacacgacagccgcgattcatcgcg
gtaggggaaaacctctcccatgctatggtagcgctaaaaaaattcacccagagttttctc
ttttttcatcttagtagaaccattgaagtaattaattgaaa

ORF (intron in bold)
ATGGCTAGAGGACC**(gtatgttattttgttcacggattagcagaaaatggaaggagcaatatttcaagagaacag**
agatggtccatgttcagatctgaacacaaggttgtggatcaagacgtcaactttgaaaagtttttcatcatgatc
atctcttttgactgggagattgttgttttctgcaaagctattgaaggataatgaagtcgcaagtactaacaattc
ttttag)AAAGAAGCATCTAAAGCGTCTCGCAGCTCCATCCCACTGGATGCTTGACAAGTTGTCGGGTACTTATG
CCCCAAGACCATCCCCAGGTCCTCACAAGTTGCGTGAGTCCCTGCCATTGATTGTGTTCTTGAGAAACCGTCTTA
AGTATGCTTTGAACGGTCGTGAGGTCAAGGCCATTTTGATGCAGAGACACGTCCAGGTCGACGGTAAAGTCAGAA
CCGATGCTACATTCCCAGCTGGTTTCATGGACGTCATCACTTTGGAGGCCACCAACGAAAACTTCAGATTGGTCT
ACGACGTCAAGGGTAGATTTGCTGTCCACAGAATCACCGACGAGGAGGCCACTTACAAGTTGGGTAAGGTTAAGA
AGGTGCAATTAGGAAAGAAGGGTATCCCATACTTGGTCACCCACGACGGTAGAACCATCAGATACCCTGACCCAG
ATATCAAGGTTAACGACTCTATCAAGATTGACCTGGCTACCGGAAAGATTTTGGACTTTGTTAAGTTTGATGTTG
GTAAGTTAGTCATGGTTACCGGTGGTCGTAACTTGGGAAGAGTCGGTGTCATTACCCACCGTGAAAAGCACGAGG
GTGGTTTCGACTTGGTTCACATCAAGGATGCTCTGGAAAACCAGTTCGTTACTAGATTGTCCAACGTTTTCATCA
TTGGTGAAGCTAACGCCCCAATCATCTCCCTTCCAAAGGGTAAGGGTATCAAGTTGTCCATTGCTGAGGAGCGTG

ACCAGAGACGTCAACAACAGGGTCTGGATGTTGAGGCCGAAGAAGAGGTTGAGGAAGAGGCTGACGTTGATTCTG
AAGTTGATTTTGAGTAA

Downstream (in bold start codon of next gene)
atatactaggataaagtaacaccaaggtattcgtattatacagacagtagactgctaatt
tattgtgcatcattgtagaggtaggtgacttgattaggttttgggtaaatggttttgatt
acataataggctgaagtatttatccctctctgtttttggagggcatcgatggaactcctc
cgcactatcactaacc**atg**tattcattgagtaaccgagttttgccgtcatatacttcaag
aaatgtcagtaagaggattgttttgtcgtgttctagaggattgattaggagcagtaggtt
taattctaccacaacacccttacaaactccccgtcgtggcggatactggaagtactattt
gggggcatttgccttaggtggactcgcagcttctcaattgccgctttacagtatatatca
attatcagcaacgagtttgccagaggacggtgactcttttgaatacgctgagtatatgac
ttctttggaattacagctgcaaggattgccgatagtgaaagcgctaagcaacgaccccag
atatgtcaagttcagggcatggcaaaatttggaaggtaatatcttggaaaacagtcttat
tgggggtacattgaatgtaccaggaggtttcgcagtgatgcctgtagcgttcgtgaatga
gtccactgaagagatgattactgttgttcatgtgggcaggaaactctgtggatatccatt
actagttcatggaggaatactagctacaattctagatgaaaatctaaagagatgttcaat
tctagaattcaataagaaaaagaatttgaatttgaaatataatgacgagatgagctataa
tatgatccacacaaaaacacttacgatatcctacagatctccgacactggctgacaattt
catagtgataaaatctcagtgccaacctggtggggataaggattctgtcaatgtcaaggc
caacattgaaacagtaaatcaacgggttttggtgaatggtg

AA
MARGPKKHLKRLAAPSHWMLDKLSGTYAPRPSPGPHKLRESLPLIVFLRNRLKYALNGREVKAILMQRHVQVDGK
VRTDATFPAGFMDVITLEATNENFRLVYDVKGRFAVHRITDEEATYKLGKVKKVQLGKKGIPYLVTHDGRTIRYP
DPDIKVNDSIKIDLATGKILDFVKFDVGKLVMVTGGRNLGRVGVITHREKHEGGFDLVHIKDALENQFVTRLSNV
FIIGEANAPIISLPKGKGIKLSIAEERDQRRQQQGLDVEAEEEVEEEADVDSEVDFE

SEQ0107

P. pastoris homolog of *S. cerevisiae* **RPL33A (YPL143W)**

N-terminally acetylated ribosomal protein L37 of the large (60S) ribosomal subunit, nearly identical to Rpl33Bp and has similarity to rat L35a; rpl33a null mutant exhibits slow growth while rpl33a rpl33b double null mutant is inviable
Chr 1-4, 0275

5' region (in bold start and stop codon of next genes)
gaggtttacaccgttgactggagtgtagatggtaagaaggttatcagtggaggtaaagac
aaaaaggttagaatctggactcat**taa**gggtctccttacttaattcatactatttagatt
acgcgacatttcctatcagaaaaccaaaccgtg**atg**agtagaccgttcaaagtgaacatt
attgtggcagcgttgatacctgggcatgggattggcttaaaggggaaactaccttggtct
ctgaaaaatgaaatgaaatatttccgtcttctcacaacaaacacaatctcaccaacttcc
aaaaatgtcgttatcatgggcagaaatacgtgggaatcaattccctccagattcaggccg
ttacaaaacagattgaatatcgtgcttaccaggaatgccaatatttctgttgaaaatggt
gctgcggttgacgactccgtgctggtggctgattccctgaagaatggattggacattata
gaaaagacaggaaacatagaggaagtgttcttgataggaggagctgatctttataaccaa
tgccttgcatctggtcttgttgaccgtatatttttaacagaagtgaagtccttgaacgga
caaattgagactgatgtgtccatcaagattgacccttcccaatggatcaagcaggagtcc
caagtcctggagaaatacctccaagaccatggagtgaacggcttccagttagacgggaac
aaagagggagagctagtctacgactttactttgctggagcgaaga**taa**ctaatggcatac
tttacagtactatacttaataaacttcatgatagttcatcaaacattgcacgcttgcgcc
ctatgtgtactcacgtgatatcttggatttttctcaaacgcagtccattgtgcgtcacat
gatctagaagatttctgtcgtgcccgcgcgaaaccatctctactcttgtcaacgaaaaaa
acatataatcgttcttactagcactccaacagatagataaa

ORF (intron bold)
ATGGCTGAGGAAAACAGAT**(gtatgtactattaaatcgcaaggtgactggtaaatgatgaagatgatcggcttaa
tgagaagaataattgaaaacatagccagaaagttttgatttctgttaaaccaagctgaaatgatgggttgggaat
ttttttattaagccaaatgaacattcttacaactttaaccaacggatattatatatggctgagccagcgttattcc
aaaagcacaagtgggaacaggtctctaagtttttttctttactttcaagttgagttgaccgcttcatcttcatttt
ctccgactatcaccaccttctcactatatttttttagattcacgaggttactaactttgctatag)**TATACGTTAA

GGGTAAGCACGTGTCTTTCCAACGTTCCAAGAGTGTTATCCACCCAAAGACTTCCTTGATCAAAATTGAGGGTGT
CGAAAACTCAAAGGATGCTGAGTTCTACATTGGTAAGAGAATCGCTTACGTCTACAAAGGTGTAAAGGCTATCAA
CGGTACAAAGGTCAGAGTCATGTGGGGAACAGTCACCAGAACTCACGGAAACTCTGGTGTTGTCAGAGCTAAGTT
CGAACGTAACCTTCCAGGTCAGAGTTTTGGTTCCACTGTCAGAATCATGCTTTACCCATCTAACATCTAA

Downstream (in bold start and stop codon of previous genes)

attggttgttaatttatagtaatacagtcagttttttgtacaagtaatacatttgaggtt
gaaggtgttttggcaaggtttctctccttccaatggtaaaatttactctctctttctttc
tgattatgtaatcttgttgtttctttaacgtcttggtcgtgcgcccgcgacactttcagc
ttacccaatagaacatcgtccaatcttgctttaaacaccatcaact**atg**attccatttga
agctctgatcccatatgcaatcatgtttgcgggattctgtcttggtggtggtgtaatgaa
cagtgccatcaccgctgatatcaccaagagagcacgtggtccaaaccaagaaattaagag
aaacgttggcgaagatgaaggcccaaggaactacacaaagcctcgttacaacaccgacca
atgggacaaatattttgcagtcagagacctgagattgaccggctctcttagaggccaatc
tgacaacgccgttgctgaagaatccttcaaaacaaactccatccaaccttacagtaatac
tagaagaccatgggtgttgagaagacatatcatcatgaagaactctcctaccaagtgggc
tagaaaatacagcagtagaagagacgagagagagcaagagattaaagatgactacattag
aggaatgggtgaacac**tga**tttcatgtagttgtagttatctttcatttatttatttcgtg
aac**cta**tattcggacgcttcttcgcttaaagaagtccataaccttcttggctgtcgacgc
tgagttgtctttatggatgtccttagatgtagcaagacgtgtgttagtattcttcaacac
gtttacctcgttactgggtactaacggtctcctgacctctaatggagattcaatctccac
aggatccactggtgaactagcacttccgctggtatagaaacttaacaggctaaacctttt
tctttctggtttagtcttgttttccttgtcttgatgcttac

AA

MAEENRLYVKGKHVSFQRSKSVIHPKTSLIKIEGVENSKDAEFYIGKRIAYVYKGVKAINGTKVRVMWGTVTRTH
GNSGVVRAKFERNLPGQSFGSTVRIMLYPSNI

SEQ0108

P. pastoris homolog of *S. cerevisiae* **RPS26B (YER131W)**

Protein component of the small (40S) ribosomal subunit; nearly identical to Rps26Ap and has similarity to rat S26 ribosomal protein

Chr 2-1, 0362

5' region (in bold start codon of previous gene)

agacggaaccgcactcccaacatcttgcattcttctcagtggtgggcttcaaccattcga
tgttatgagaatcggctgggacaccagtgcaaccaacatagcggtagtcatcataggagt
cgatcaagattgggtccttcatggtacccttttctgacagcttccaaaggtgtttcgtcga
aaacttcgataccttccagtttacctagcaattcgtatctttctaaaccagtagcttgtt
ccaaatcggtgggaactgttcctgtcttggcaccaggaccaaccaggttctcaaggcctt
ccacctcagcaatggatgtggcagtcttgaattgctttggggtttgttgtctcaagatgg
aagcagtggagaacagacgagatgctggtctaacacgaggtaagcattgacgaagcacta
tagatgttagtaacgtgatactttaggacaaaggcctatcagattaaaagttcatgtgtg
ctgtcaacctgtggagcttggggtgggtgtggaacttactgataaataatttccttgtgg
tatcaagtgaaaggaacagttaagttgagatagcttaaacagacgaaaatcatcctattt
caaactagtgtgaatatcgcgcttcct**cat**tggttgatgctccagcagctctgctgatgt
aatatctgacgtaatcaatctgatgttcaggtgttttgactcagaggtgcaagggtgttt
atttgagattctcgaactattgtgactgttgtgatgaaatataccgcgagctccacgtaa
ttgcgagatctaaaggagtttttccttgcaaagtcttcgcgacaccttgtaaagaaaaag
gttttgttcctaatagtcgcacgtccttatttttttctctctcacctccggcacctcacg
acttagggcacacgcgcgtttttggcactccctccctcgattcctcacccaaaattcca
tcagtttgaaaaatatctttctccatcaaccaacaaccaac

ORF

ATGCCTAAAAAGAGAGCTTCCAACGGTAGAAACAAGAAGGGTAGAGGACACGTTAAGTTTGTCAGATGTTTGAAC
TGTGCCAGATGTGTTCCTAAGGACAAGGCCATTAAGCGTGTCACCATCAGAAACATGGTTGAGGCTGCTGCCGTC
AGAGATTTGTCTGAGGCCTCTGTTTACGAGGAGTATGCTCTTCCAAAGCTTTACAACAAGCTGCACTACTGTGTT
TCCTGTGCTATCCACACTAGAATTGTCAGAGTCAGATCCAGAAGTGACAGAAGAATCAGAGCTCCTCCTCAAAGA
CAAAGATTCAGCAATGACAGAAGAGTCAACCCAGCTGAAGCTGCCAAGAAGGCTTTGTAG

Downstream (in bold start codon of next gene)

gtgtagtgtaattcgaatatgaatcatcagttattggaaactgtagagcgtgtttctttt
aaagggaacctcacattgactcgtggtattgtaagctagctcagttctggattgttagac
agttctaatgctggtcctaacaaatgagtttttctcgcaccttcgtaggcggcctcggtg
gcccccaagcactggccctcccattccgatcctcccactacacctttctctcgcacggct
atcaccccatctctctaccgtgctccgcctttgcccttccacaaataatctccgcccccc
gcgcccaccaaggcatccaacttagcgaacgaatatttctcagacgggggagagacgtta
cctatcttctccagcaacgtaaactcaacacgaaagggagaatggaagaacacaactgga
agttttcacaatgctttggtgacaaaggtgatgtggagacaattacagaagccgacatca
tttcgacggtggaatttgaccataccggcgattttctagccaccggtgatcgaggaggcc
ggattgtgctattcgaacggaacgagagcaagaacaactgtgaatacagattctacactg
agtttcaaagccacgatgctgaatttgactatcttaagtcattggagattgaagagaaga
tcaacaagattaaatggctcaaaagaagcaacacgtcgaacatgctgctaagcacaaacg
acaagaccatcaaactgtggaagatctttgagaagcaaataaagatggtcagtgaaaaca
atctgagtgagaactctaacttaatgaacacagggaatgagacattgtcggccgccaata
taagttccaagttgttgtcgatcaacaacttgaagctgcctagattgaccttacatgata
ccatcacttcagcacaacccaagaagatttatgccaatgctcatgcctatcatatcaatt
ccatcagtgttaacagtgatggagagacattcatatccgca

AA

MPKKRASNGRNKKGRGHVKFVRCLNCARCVPKDKAIKRVTIRNMVEAAAVRDLSEASVYEEYALPKLYNKLHYCV
SCAIHTRIVRVRSRSDRRIRAPPQRQRFSNDRRVNPAEAAKKAL

SEQ0109

P. pastoris homolog of *S. cerevisiae* **RPL3 (YOR063W)**

Protein component of the large (60S) ribosomal subunit, has similarity to E. coli L3 and rat L3 ribosomal proteins; involved in the replication and maintenance of killer double stranded RNA virus

Chr 4, 0139

5' region (in bold start codon of previous gene)

tcagcagattgaactcatctctggtttttcccgatgggacaccgtagtttccaatcaggg
gttgagtgaagcagatcatttgtcccttataggacggatcagtcatcgactctgggtaac
caaccaaggaggtggtgaagacagcttctccactgatattcttgttggcaccaaaagagt
aaccgtcaaagacggggccgtttctaatggtaagagtggccctgtccagtctggcggatg
atgacaatctggtctgttgaaagaccaacgacttgaggaaatgacgattatttcttaaga
tattcatggtgaggttttccaaatgtgatctgaaatgtaatcttggcaagtttggttcat
aaaacgtcaaaagaccacctgaagataaaaaataaagtggaaggagaagaaaaagaggtt
tttttccagatgagtctatgactaagatttctcttcaaagaatttttcaaggtttgaga
gggatgaaagcgaggctgctgcacgattatcgatgtcattcaatagggaggtagaagaga
actgagggagaaaaggctacaactgaacctaaaactaactgaatttggaatgaagcctat
tacacgagaggtgtggcgcctttccaaagtcatcaggttcaagttattgctttctggagc
aataaatgatgtcaatcccatacggggtctccacccttatagaattacttatctaccggc
taggtgttgcatcgcggaccagcccaatcgggtaattttagttgaaaaggtttttagcgc
gagtctagttattacccggatagtctaatgctgaaaatttctcggacattttgatccgt
acgtatatttcagtacggagaatcatcaacaagaatctttcaatgccaccaaaaaaaaag
acatttttcaccatcgcgcattcacagtagcttgaaaaaaaaacttcaccttcacccatt
ctcttctcctccacctttttacccttactaaatttgtata

ORF

ATGTCTCACAGAAAGTTTGAAGCTCCACGTCACGGTTCCCTTGGTTTCTTGCCAAAGAAGAGAGCCTCTTCTCCA
AGAGCCAGATGTAAGGCTTTCCCTAAGGACGACAAGTCCAAGCCAGTCGCTTTGACCGCCTTCATGGGTTACAAG
GCTGGTATGTCCACCATCGTGAGAGACTTGGACCGTCCAGGTTCCAAGATGCACAAGCGTGAGGTTGTTGAGGCC
GTCTCCATTGTTGACACTCCACCTCTCGTTGTTGTCGGTGTCGTTGGTTACGTTGAGACCACCAGCGGTCTTAAG
TCTTTGACCACCGTCTGGGCTGAGCACTTGTCTGATGATGTTAGAAGAAGATTCTACAAGAACTGGTACAAGTCC
AAGAAGGTCGCTTTCAGCAAGTACTCTGCCAAGTACGCCAACGACCCAGCCTCTATCGAGACTGAGTTGGCCAGA
ATCAAGAAGTACGCTTCTGTTGTCAGAGTCCTCGTCCACACCCAGATCAGAAAGACCCCTTTGAACCAAAAGAGA
GCTCACTTGGCTGAGATCCAAATCAACGGTGGTTCCATTTCTGAGAAGGTCGACTGGGCCAGAGACCACTTCGAG
CAGACCATTGCTGTTGACACCGTCTTTGAGCAGAACGAGAACATTGACGTTATCGCTGTCACCAAGGGTCACGGT

TTCGAAGGTGTCACCCACAGATGGGGAACCAAGAAGCTTCCAAGAAAGACCCACAGAGGTTTGAGAAAGGTTGCT
TGTATTGGTGCTTGGCATCCAGCCCACGTTATGTGGACTGTTGCCAGAGCCGGTCAAAACGGTTTCCACCACAGA
ACCTCTATCAACCACAAGATTTACAGAGTTGGTAAGGGTGATGATGAGTCCAACGGTTCTACCGAGTATGACAGA
ACCAAGAAGACCGTTACTCCTCTTGGTGGTTTTGTCAGATACGGTGAGATCAGAAACGACTTCGTCATGATCAAG
GGTTCCATCCCAGGTCCAGTCAAGAGAGTTGTCACCTTGAGAAAGTCTTTGTACACTCAAACCTCCAGAAAGGCC
ACCGAGCAAGTCACCCTTAAGTGGATCGACACTGCTTCCAAATTCGGTAAGGGTAGATTCCAAACCCCAGCCGAG
AAGGCCGCCTTCTTGGGTACCCTTAAGAAGGACTTGCAGTAA

Downstream (in bold stop codon of next gene)

atacctataccatacgtaataatatagtataaagagtatatactctttttttcttttagc
tctttcaaacttagttagactctactaatgagtatatacatataatatcaccactgcaaa
tacaccactgcaaacaaaaagatatcttctttatttccataactccttccactacccac
taaaattaaaaaagaaacaccataatcaagcaccacaactgagaaacaaataatatacaa
acaaaattaataatattatcagtgtagatccacttgctgtactttaacacaacgataaat
tcacacagaaaaacgtatgctaatcaaacaatcaatagaaaaaccaagaaaaatcaaatg
agaagatgctgaataaacatgatgactgtgatgagcggttggaaagtggaatccaaccgc
aaaaactggattccagcagggtcatgctccctgttggtccaaaactatttgatacctta
taaattgataccaacaaaaaacgaaaagcaatcaaaaaaatgcctagcgacagcatcctg
tgtgtgactgactggtggt**tta**ccaccaatcagaaccagacgattgttgagtagtgttgc
tgcttggagcactaccgtagaagttagaaccaccgaagtaaccacttcctccattagccc
taccataagtttgggaacttggctgaggcgggtagttagaagatcttctgaaatcttttc
cgtaagcgtttcttccgcctcttccacctctggctccgcctctacccatagtagcttctc
tggaaagggcgctcaggaattgaggaacttcttggttagcctcggttaacagttcaatta
aacccttgatcaggttcttgttaccgcggttaacgaatgcagtggcaataccaacattac
cagcacgaccggtacgaccaatacgatggacgtaatcgtcaatgtcagctggtagatcgt
agttgacaacgtgtgtgacatttggaatgtctaaacctctg

AA

MSHRKFEAPRHGSLGFLPKKRASSPRARCKAFPKDDKSKPVALTAFMGYKAGMSTIVRDLDRPGSKMHKREVVEA
VSIVDTPPLVVVGVVGYVETTSGLKSLTTVWAEHLSDDVRRRFYKNWYKSKKVAFSKYSAKYANDPASIETELAR
IKKYASVVRVLVHTQIRKTPLNQKRAHLAEIQINGGSISEKVDWARDHFEQTIAVDTVFEQNENIDVIAVTKGHG
FEGVTHRWGTKKLPRKTHRGLRKVACIGAWHPAHVMWTVARAGQNGFHHRTSINHKIYRVGKGDDESNGSTEYDR
TKKTVTPLGGFVRYGEIRNDFVMIKGSIPGPVKRVVTLRKSLYTQTSRKATEQVTLKWIDTASKFGKGRFQTPAE
KAAFLGTLKKDLQ

P. pastoris homolog of *S. cerevisiae* **RPL2B (YIL018W)**

Protein component of the large (60S) ribosomal subunit, identical to Rpl2Ap and has similarity to E. coli L2 and rat L8 ribosomal proteins; expression is upregulated at low temperatures

2 homologs

SEQ0110

Chr 4, 0107

5' region (in bold stop codon of next gene)

tcaaacatggctgcccttatatttgcacctatttttggtgttctcattgacaaattcaac
aatcatgctcacttacttttgattgtgaacactctgagtctgttaagtgctgttggtgta
accttcattaagacaccagattcttcagcactcatctttctgttcgcaacaatgttagga
gtctcacaaattgggttcataatcatttcaatgaccatgttatctaacatttcctccaac
caagaatactacaagaacaagggatcattagttggtgtctactcattatttggatgtgtt
ggagtcttaattgttaataagtttggaggttggtttggtgactggttcatgcatggtcca
tttattctgttagctctatttaacgttttattctttattgctttatttgcattgaagggt
gttaagaagaacgaatccagagactttcagtcagttttacaaagctttactggcacccgt
gtaggacgtatgttttccaacaacagacatgttccaaactctgacgagttt**tag**tattta
taggtaaaaatttttagtatataggtcaaaatccaaattagaagtcaattagatggatta
tttctgaaagtggacctttgagccccttggaggttataattaagaagtatttcagagtg
tcatgacatactggagtataatcaattcaatatataaattagtgtaaagatgttggaagt
ctgataatacgataatgtaactctccctggcacttcaaaatagaccgaagaagaattaaa

tcattcaccctacgccactctaatcatacagacaccttcctaacgaagtcttccataagg
tacctttcaactaccaccttatgatacctcacacagcttgcccaataaggcacgtgacca
gcactagggttaacccgcgaacacaagaaccctcatactccccaaaaaaaaaaccaataa
aaaatttcaagtcctcttacagcaagaaacattaacacaac

ORF
ATGGGTAGAGTTATTCGTGCTCAAAGAAAGGGTGCTGGTTCCATCTTCACCGCTCACACTCGTCTGAGAAAGGGA
GCTGCCAAGCTTAGAGCTTTGGACTCTTCCGAACGTGAAGGACAGATCCGTGGATATGTTAAGCAGATCATTCAC
GACCCTGGTAGAGGTGCTCCTTTGGCCAAGGTTATCTTCCGTGACCCAACCAAGAAGGGTTTGCGTGAGGAGACT
TTTATTGCCAATGAGGGTTTGTACACTGGTCAATTCATCTATGCTGGTAAGGATGCTTCTTTGAATGTTGGTAAC
ATCCTGCCTTTGGGAGCCCTTCCAGAAGGTACTATTATCTCCAATGTTGAGGCCAAGCCAGGTGACAAGGGTACC
ATCGGTAGAACTTCCGGAAACTACGTTATTGTCATTGGACACAACGTTGAGGAGAACAAGACCAGAGTTAAGCTG
CCTTCCGGTGCTAAGAAGATCATTTCTTCCAACTCCAGAGGTGTCATTGGTGTTGTTGCCGGTGGAGGAAGAATC
GACAAGCCATTGTTGAAGGCTGGTAGAGCTTTCCACAAGTACAGAGTTAAGAGAAACTCTTGGCCAAGAACCAGA
GGTGTTGCCATGAACCCTGTCGACCACCCTCACGGTGGTGGTAACCATCAACATATTGGTAAGGCCTCCACCATC
TCCAGAAATGCTGTTGCTGGTCAAAAGGTTGGTTTGATTGCCGCCAGAAGAACCGGTTTGTTGCGTGGTACTCAA
AAGACCCAAGATTAA

Downstream (in bold start and stop codon of previous gene)
actatgtaactaacgaaacagcatgtactaatagaaccgtatcgagaatatttatttagg
tgagtagtaggagtgaaccagacagtcaatttagtgagctgtcccagcttttgtgcattc
cagaattgccggtcaaattggttatgggttatggggcttttccgattgaggttcagtttc
tgcggttatctctttcttgacctggtcttttacaggctgttctttctccccatgattatt
ctttagctgaagataccgc**tta**gcctgataatgtcgtcgttttgtaatcaaaatctttag
ttgggcatcgtctgaggtttcctttggcttctggggttgttagtaggaacgtaggaacca
tagtaacttttacacatacattcttatgattgcgaagtaagctgagtctgctgcttggct
cccgaagtactttctctttctctaccggttgattctccttctggtgctcctaaacgattg
tgttagaagggattgacaacagactgtggaatagtacatactcttttgaaataaaaatgc
ttgaatctagagtttcctcttttcaatggattttgtttctgatcttcctctgctttgcta
ctctctggggaagatgcctcctcctctattgtctatcaccagttagataacgtttatcta
gacatctaatccacttacaattacctcaattccttctgcgatttcgtcaatagccaaagg
tgccaggcttgagtcgtcctcgtcctgttcgatgactgtgctgattacactgtttgactt
tagatccgtcttgttcatcgcttcaactgt**cat**tctcagttgttcgttcaggtgataaat
ttccggccgaccactacacaaagctaatacgcaaacatgctgcgggggctttgtgtctac
aggacctgccctaccccagcttctcttccagtcaaatgcatgcaacacttctctgttgta
ccgttacaatctcatctcccaaccttacgtaatcaattctt

AA
MGRVIRAQRKGAGSIFTAHTRLRKGAAKLRALDSSEREGQIRGYVKQIIHDPGRGAPLAKVIFRDPTKKGLREET
FIANEGLYTGQFIYAGKDASLNVGNILPLGALPEGTIISNVEAKPGDKGTIGRTSGNYVIVIGHNVEENKTRVKL
PSGAKKIISSNSRGVIGVVAGGGRIDKPLLKAGRAFHKYRVKRNSWPRTRGVAMNPVDHPHGGGNHQHIGKASTI
SRNAVAGQKVGLIAARRTGLLRGTQKTQD

SEQ0111
Chr 1-1, 0219

5' region (in bold stop codon of previous gene)
agaagaagaaagccaacaaaaagaagacttccgctgctaagaacaatgatggaaccaccg
aaggtgttactaatactgcagtggatgggactgaggagtcccaagctgatgtcagtagca
atgccgactccgcttccgattctgttgctgattcgaaagacaaagagaacagcgctccca
atggcggcaagaagaagaagccttacagtaagaagaaacaaaatccacataaagttcccc
ttgaacagggtacgccttcagagacttcgatcttcatccgtaatttgcaccaggaggttg
tcacttccgaactgacggagttcctttctgagtatgaccctcaatgggtttcgatcccca
gaaggaatataccccgtcatgttgttcaaaaactgaggaaagagaacattccgatcaggt
cgcgtggtatggcatttgtgagatttgccgaccacgaaacgcaacaacgcgctctcaagg
aattgaacggtaaggaaattaagggcaaaccaattagcatcgatattgccattgattcct
tcgtgaaggaggaacctaaagaatctagcgtcaaagatggtgctgacgtttccatcgtta
ctgatggagctcagccgcaggatgagggagaagagggcaaggtagaggaggaagtaaagt
ccgagccgctcaaaccattggagtcaaccaccattgctccagaagtgcaaactactacag
tg**tga**ttttccaaaccgtttcgagttgctgaactctgtcattataatctataggttttgc
ttgcaactagttggcgtaataatgtatcatcattccatatttatttgttcaagacttctg

taccccgtcgcgagccctagccctgtacacgtgatggatattatactgggctcgcgacc
cacgaatacacctctacaccaacctcaaagagtggaaccttccatcttaggtgaaaaatt
cgttagactaaaggaaattctatacatttcaacaggacaaa

ORF(intron in bold)
ATGG **(gtatgtaaaccaccaaccactaaacatccagaggtcaagtactaacgtcattgtttatag)** GTAGAGTTA
TTCGTGCTCAGAGAAAGGGTGCTGGTTCTATCTTCACCGCTCACACCCGTTTAAGAAAGGGAGCCGCTAAGCTCA
GAGCTTTGGACTCTTCTGAACGTGAGGGACAAATCCGTGGGTATGTTAAGCAGATCATCCATGACCCTGGTAGAG
GTGCTCCTTTGGCCAAGGTTATCTTCCGTGACCCAACCAAGAAGGGTTTGCGTGAGGAGACTTTCATTGCCAATG
AGGGTTTGTACACTGGTCAATTCATCTATGCTGGTAAGGATGCTTCTTTGAATGTTGGTAACATCCTGCCTTTGG
GAGCCCTTCCAGAAGGTACTATTATCTCCAATGTTGAGGCCAAGCCAGGTGACAAGGGTACCATCGGTAGAACTT
CCGGAAACTACGTTATCGTCATTGGACACAACGTTGAGGAGAACAAGACCAGAGTTAAGTTGCCTTCCGGTGCTA
AGAAGATCATTTCTTCCAACTCCAGAGGTGTCATTGGTGTTGTTGCCGGTGGAGGAAGAATCGACAAGCCATTGT
TGAAGGCTGGTAGAGCTTTCCACAAGTACAGAGTTAAGAGAAACTCTTGGCCAAGAACCAGAGGTGTTGCCATGA
ACCCAGTTGACCACCCTCACGGTGGTGGTAACCATCAACATATTGGTAAGGCCTCCACCATCTCCAGAAATGCTG
TTGCTGGTCAAAAGGTTGGTTTGATTGCCGCCAGAAGAACCGGTTTGTTGCGTGGTACACAGAAGACTCAGGACT
AG

Downstream (in bold start codon of next gene)
gtacttttatagttatcaaggttaaattttgggtaatactcttgtagcttttcgtagagc
tcactgtcacctcctagttttcagcctaacctttctcagccctgacgggtgttactccct
tcgggcttgcgctcccttctaaccccgtagagcttgcttgccttttaagccgttttagt
cttcacacctacttatgacatatctttgctgaaaagcttccctttagttgaggttatcca
agctagagcttttaactccctccagttttcttgcccgccaacaggaccgtcattccttgc
cacagcgttccttgaagctttttccctgtgaagactcaattaccaggtggactcaatttt
ttcaacagttcgaagattagaagatattttctttccgttaagaagtcgtctttccttcgt
atgaattcgtacaatgagtcagccccgacgggttgcagtttctgggataatgatgatatt
tctccctgtatcagaaagtcactgctggattcgtatttgcctgctgcaattgtagtagga
tcgttactctatttactgctcataggagcacaacagatcaagactcacagaaagctctac
gccaaggatgaaacacaaccgttattagaacctgccaatggctcacctacggattattct
aatacttatggcaccatagattacgaagaggaacaatccacagcggagttgacaacctca
cagaaacattttgacatttccagactggaacccctgaaagatgatggcacaccgttagga
ctggtgaaatatgttcaaagagatggctgggagaaagttaagctaattctagagtttgtg
atattgatattccaattagttatagctgtcgtcgctcttttcgtaccttctcttaaccaa
gagtgggaaggttacaagcttactccaatagtacgagtattcgtgtggatcttcttgttt
gcacttggatccatcagggcgttgaacaagtcgggtccatt

AA
MGRVIRAQRKGAGSIFTAHTRLRKGAAKLRALDSSEREGQIRGYVKQIIHDPGRGAPLAKVIFRDPTKKGLREET
FIANEGLYTGQFIYAGKDASLNVGNILPLGALPEGTIISNVEAKPGDKGTIGRTSGNYVIVIGHNVEENKTRVKL
PSGAKKIISSNSRGVIGVVAGGGRIDKPLLKAGRAFHKYRVKRNSWPRTRGVAMNPVDHPHGGGNHQHIGKASTI
SRNAVAGQKVGLIAARRTGLLRGTQKTQD

P. pastoris homolog of *S. cerevisiae* **PGK1 (YCR012W)**

3-phosphoglycerate kinase, catalyzes transfer of high-energy phosphoryl groups from the acyl phosphate of 1,3-bisphosphoglycerate to ADP to produce ATP; key enzyme in glycolysis and gluconeogenesis
Chr 1-4, 0292
See SEQ0072

SEQ0112

P. pastoris homolog of *S. cerevisiae* **RPS14B (YJL191W)**

Ribosomal protein 59 of the small subunit, required for ribosome assembly and 20S pre-rRNA processing; mutations confer cryptopleurine resistance; nearly identical to Rps14Ap and similar to E. coli S11 and rat S14 ribosomal proteins
Chr 2-1, 0481

5' region (in bold start and stop codon of next gene)
gttgaaaaagaagttttttctcacaagtgtaagagtagaggaaaagaggtagtgtgcgc

gacaatcgcagtgggaacactgaaaaaaaagtacggtaggtgccagcaaaaggagtcatg
tgacatatgttacggctaacccctgctaacttcttctggtgtagagagtttttcagctc
gtctgccaaaaacagtctgaaaaaaaatacacagctactcttaccatcacacctatcgaa
t**atg**accagaacctctgttcttgccgacgctctgaacgctatcaacaacgccgagaagac
cggtaaacgtcaggtactcatcagacctgcttccaaggttatcgtcaagtttttgaccgt
catgcaaaatcacggctacattggagagtttgaattcattgatgaccacagatccggaaa
gattgttatccagctgaacggtagattgaacaagtgtggtgttatctccccaagattcaa
cgtcaagattggtgacattgagagatggactgacaacctgttgccagctagacaatttgg
tttcgttatcttgaccacctccgccggtattatggaccacgaggaggccagaaagaagca
cgtttctggtaagatcttgggtttcgtctac**tag**atgtgtgtacttatactaatatactg
attattcatcattaatggattgtagttacaggctctttcggtatattgatagtaccagca
ctccgcccagggagtcccattggggttcaatcctccaacaccttgtccccttgaacctc
tcccttctacctctcccccctgtaggcattttggcttttctcgcgaacgggatcctaatt
gtgactgcgttcctaatgaggcaactctcaacctccgcccaaggctcctttccccgacat
aattgggaggagtcccaattacacctccgtctcgcgaccaaaccttctcctccctctctt
tttcagtcgaaaaaagttttacctcttcatcatacaaaacc

ORF
ATGGCTTCCAAGACTGAAAACGTTACTTTGGGCCCACAAGTCCGTGATGGCTCCCAGGTGTTTGGTGTTGCTAGA
ATCTTTGCTTCTTTCAACGACACCTTCGTCCATGTCACTGATCTCTCCGGTAAGGAAACTATTGCCAGAGTCACT
GGTGGTATGAAGGTCAAGGCTGACAGAGACGAGTCTTCTCCATACGCCGCCATGTTGGCAGCTCAAGATGTTGCT
GCTAAGTGCAAAGAGGTTGGAATCACCGCCGTTCACGTTAAGCTCAGAGCTACCGGTGGTACCTCAACCAAGACC
CCAGGTCCAGGTGGACAGAGTGCTTTGAGAGCTCTGGCCAGATCTGGTTTGAGAATTGGTCGTATTGAGGATGTT
ACTCCAGTTCCATCTGACTCTACCAGAAGAAAGGGTGGTAGAAGAGGTAGACGTCGTTGA

Downstream (in bold stop codon of previous gene)
acggatctctttttactatattcatatactatatacactagtcgtgggattgaatttaat
ttcggcgtgataaattgattgtagatatttaattgaaacagacgctgagctgttttattt
gctggtattttc**cta**ctccttttgtcggtggtttgaggttggtccttggactcgattac
ggtgccatccttactgggttctttcacctcctcttcagattgcgctggagtttcgtcttt
gacctcctcgcttgcatcaacctttggttcttctttgacatgttcgttcgtatcttccct
cttctcctccacggtgtctttagtaacgtcatccttcacttcatcttctttcacctcctc
gtattcgtctagtttgatatcgtctttgattggagtggaacggcgttgtgatttactctg
ttgtctctcttcctttatcagctcatcttgttccttcaagacattcagcttcaacttatt
gtcctggtcggccagctctaacttagcaacattgtaaacctcatcaggaatcgaagaaag
tactgccaacagggcatcgtagtaattgtccaggtcatcagcgtgatctccataagtgaa
cgtgctactcaaaatcaaaagtgtcgaaggaatcttctgacgcaacctcaagtccagcca
aatcttcaaatcatccctcagtcttgcaggagaagcaccaacggttttgataccacgtga
agcacaagcactctggagctcctgaattgacagagattcaactccttcgtaatcaatggc
tcgatcatccttgatgatttgcaataacttatatctgatctgataacgtaaaatctcgtc
agttccaaacggacgtaaattcatgtaacgggccattgccatcaattgtggtcgggaaag
attgtctagcacttggtcattcttgaaacgacgggcgatacgaataatctgttctgtcga
tggtctttcacctggttcctggatactttgaaaaatgata

AA
MASKTENVTLGPQVRDGSQVFGVARIFASFNDTFVHVTDLSGKETIARVTGGMKVKADRDESSPYAAMLAAQDVA
AKCKEVGITAVHVKLRATGGTSTKTPGPGGQSALRALARSGLRIGRIEDVTPVPSDSTRRKGGRRGRRR

SEQ0113

P. pastoris homolog of *S. cerevisiae* **ATP1 (YBL099W)**

Alpha subunit of the F1 sector of mitochondrial F1F0 ATP synthase, which is a large, evolutionarily conserved enzyme complex required for ATP synthesis; phosphorylated

Chr 3, 0576

5' region (in bold start codon of next gene)
ctcagctttagcttcgatcttgtcgatcagctcttgaggtacctctgatccaatcaaacg
agcattagcaacgcctcgttcagcagcacttaattggtctagttcctctgctctcttagc
attgttgacacatttggaagcatttgtctcttttgtcactcttcgatcggcaggaatact
cgatggctccgtgtaagaggcaaaattgggcaagatcacaccttcatcctcatcaatttc
tgaatctgtatcactgatctgttcttcctcgctggagctctcatcattctcaatggattc

atctttgtcgtcatccttgctcacaagcctttcacttggatcatgaacagtaatcgacca
tttgttttcttttggattggggttccttgaatgagatctactactccgacctcttgagac
tcctctgctatgaggaccaccagaatttggagacacagaccttcccctggcaccagactc
cctccgcttttgcctttctagaatatctataatttgttttctctcatcatctaacgcagc
ctccagaga**cat**tttgcaggactgcttatcaagtagtataaaagaagggtgctttctttg
tcgtggaacctagtcaagttgcaagggaaggagcaaagtgcagcatccagtttccatcca
cgtgtgggtgctacactacttcttgtcgtgaatatttttttccgagtcgttatataagtaa
ggtcctatatatatataagtatgtttttgagttattatcttcttttataattaatagtgc
cttacttcgaatctcatccttccataaatatgggtcgtgcactacacacccccttggacc
gccccccagggtttctaaccgctcgaaattgtgaactgtggttggccaatccaagcgcgc
cctcgacactcctcaacgctgtttttatattaactaacccctttcctttccttttctctcc
tacacccagtgttaaatctttatctttaataatattcaaag

ORF (introns in bold)

ATGCTTTCAGCACGCCCCGTCATTAGAAGAGCTGCTCGTTCTGCTGTCGCCATCAATGTCGCCCGCTCTGTTCCT
CGCGTG**(gtatgtatttcctgatgatttgaatcacttttgatttggctgtcaactggctcccccatgaaaagtt**
tgtctttactcaacgttgaagcattattgatgattattttggattgtagaggatcaattccatagaatgctttaa
atatttcacacattaatgtgttttcgaacaggacaatcgcttcaaatggtgtggtgaaatctatcaaaagtgtac
ttttttctttccattgacaactactaacatatcattag)GTCCGCCCATTAGTTTCTAGAGGTTACGCTGCCGCCA
AGGCTGCTCCAACCGAAGTTTCTTCCATTTTGGAAGAGAAGATCAGAGGTGTTTCTGCCGAGGCTGACTTGAACG
AGACTGGTAGAGTCTTGTCTGTTGG**(gtaagtttttctgaatttttttttgaatatgtcatgggctgatcttagg**
gaggcatgatacaacagtgctttactccacaattgtgacgtattgattgatgtgttgtcctcagatatcttcttg
tccagatcctatttttttttcaagttttctcgacattttattggacagtcggagatgctcgacttgactccctaa
gaactacgcccttacaaacaaccttaaactgaagctcagtaaactaacattatag)TGATGGTATTGCCCGTGTC
TTCGGATTGAACAACTGTCAGGCTGAAGAGTTGGTTGAATTCTCCTCAGGAGTCAAGGGTATGGCCCTGAACTTG
GAGCCAGGTCAAGTCGGTATTGTCTTGTTCGGTTCCGACAGAGGTGTCAAGGAGGGTGATACCGTCAAGCGTACC
GGTAAGATTGTTGATGTCCCAGTTGGTCCTGAACTGTTGGGCCGTGTCATTGATGCTTTGGGTAACCCTATCGAT
GGTAAGGGACCAATTAACTCTAAGGTTACTTCCAGAGCTCAATTAAAGGCTCCAGGTATCTTGCCAAGAACATCC
GTTTTCGAGCCAATGCAGACTGGTCTGAAGTCTGTCGATGCCCTGGTGCCAGTTGGTAGAGGTCAGAGAGAGCTT
ATCATTGGTGACCGTCAAACCGGTAAGACTGCTGTCGCTCTTGACACCATTTTGAACCAAAAGAGATGGAACAAT
GGTTCCGACGAGTCCAAGAAATTGTACTGTGTCTACGTCGCTGTCGGTCAAAAACGTTCCACTGTCGCCCAGCTG
GTTCAAACTTTGGAGCAAAATGATGCTCTTAAGTACTCTATCATTGTTGCTGCCACCGCTTCCGAGGCCGCTCCA
CTGCAATACTTGGCTCCCTTCACCGCCTGTGCTGTCGGTGAATGGTTCCGTGACAACGGAAAGCATGCTTTGATC
ATCTACGATGATTTGTCCAAGCAAGCTGTTGCTTACCGTCAATTGTCTCTGTTGCTGAGACGTCCTCCTGGACGT
GAGGCTTACCCAGGTGATGTTTTCTACCTGCACTCCCGTCTGTTGGAGAGAGCCGCTAAGATGTCCGACAAATTG
GGTGGTGGTTCTTTGACCGCTTTGCCAGTCATTGAGACCCAAGGTGGTGATGTCTCTGCTTATATTCCAACCAAC
GTCATTTCTATTACTGATGGTCAAATTTTCTTGGAGGCTGAGCTGTTCTACAAGGGTATCAGACCTGCCATTAAC
GTCGGTCTGTCCGTTTCTCGTGTCGGTTCCGCCGCCCAAGTTAAGGCTATGAAGCAAGTTGCTGGTTCCTTGAAG
CTGTTCTTGGCTCAATACAGAGAAGTTGCTGCCTTCGCCCAATTCGGTTCCGACTTGGATGCTTCTACCAAGCAA
ACTTTGAACAGAGGTGAGAGATTGACTCAACTGTTAAAGCAGAAGCAATATGCTCCTTTGGCTGCTGAGGAACAG
GTTCCAGTTATTTACGCCGGTGTCAACGGTTTCTTAGACTCCGTCCCATTGGAGAGAATCGGACAATTCGAGGAG
GAATTCCTGGCTTACTTGAAGTCCAACGAGACCGAGGTCTTGGAGGCTATCAGAGTAAAGGGTCAATTGTCCGAT
GAATTGTTGGCCAAGTTGAAATCTGCTACCGAGTCCTTCGTTGCTACTTTCTAA

Downstream (in bold start codon of next gene)

acttagcagttaacgattacgatttaagcatttttgataaaagactcaaaattcgacata
caattatgatgttataaattgatatttagatttatgttttgcaatgtagagccttgtaaa
tttgatctttgatatctggagaactatgatagtagactgaggctggcgatatgctctttt
tagacacttcgtggtgttccttgctgtttcttgactgactttccctctaacatcaaattt
gtcgctaagattgtttccaagtttccttccaagtttcctcctacttattgaacc**atg**aca
tcgtctgattcaaccaaccaggaggaaactccgcaaccagagaaacattcaaagcaaatt
actaacgatgatctctaccgaaggtcgacccagttcagattttggtcatttaccgagcaa
cgattgaaggatttgaagcggtcaactaatgaaaagggagctgccaacttgaaagaaaga
atatcacaattggatccttctataccagaagtctctgagatttgggaaaaggacctcgtg
aaaccgctcactgtggaagaggaggttaaaattgtcgaatattatgcgaggaaagcccaa
ggccttgccaattatttcaaattaccaacacaatcaagggctactgccatatcatttttt
agaaagttctttctggtcaacagctgtatacaatttcacccacaatatatcatgtacaca
tgcctgtttctagcagctaaatcagacaaccactttattggaataaaggaatttagcaag
gccattccaaaaacaaccccagagtcaattttacagtacgaatttcagatactgcaaagc
ttgaagtttgctcttttatgccatcacccttacaaaccattgtatggattctttttggat
tttcaagtggaaatgaagaataatatccagtcagataaactcgctgagttgtatgaccac

gcaagagataaggtatctgaagcattattttctgatgtttc

AA
MLSARPVIRRAARSAVAINVARSVPRVVRPLVSRGYAAAKAAPTEVSSILEEKIRGVSAEADLNETGRVLSVGDG
IARVFGLNNCQAEELVEFSSGVKGMALNLEPGQVGIVLFGSDRGVKEGDTVKRTGKIVDVPVGPELLGRVIDALG
NPIDGKGPINSKVTSRAQLKAPGILPRTSVFEPMQTGLKSVDALVPVGRGQRELIIGDRQTGKTAVALDTILNQK
RWNNGSDESKKLYCVYVAVGQKRSTVAQLVQTLEQNDALKYSIIVAATASEAAPLQYLAPFTACAVGEWFRDNGK
HALIIYDDLSKQAVAYRQLSLLLRRPPGREAYPGDVFYLHSRLLERAAKMSDKLGGGSLTALPVIETQGGDVSAY
IPTNVISITDGQIFLEAELFYKGIRPAINVGLSVSRVGSAAQVKAMKQVAGSLKLFLAQYREVAAFAQFGSDLDA
STKQTLNRGERLTQLLKQKQYAPLAAEEQVPVIYAGVNGFLDSVPLERIGQFEEEFLAYLKSNETEVLEAIRVKG
QLSDELLAKLKSATESFVATF

SEQ0114

P. pastoris homolog of *S. cerevisiae* **RPL7A (YGL076C)**

Protein component of the large (60S) ribosomal subunit, nearly identical to Rpl7Bp and has similarity to E. coli L30 and rat L7 ribosomal proteins; contains a conserved C-terminal Nucleic acid Binding Domain (NDB2)
Chr 2-1, 0022

5' region (in bold start codon of previous gene)
cctacaactggagcgacactcatcagccttggatgaaagcaagatgtcgattaatgaatg
ttcaactgtaggaagaccaaaacactattgtcgcgcgaaatatccattccgtactaagat
tcactagcagcagcgttggagtactaattccgttttgatgcttgcattataatgaaatag
tttcaagaaggattttaacgacttcttacgaacgcaacaatctgcaaggcgctatgaggt
attttaacacgtagattaggatttattttgctagttacgggggagatatcattagttgaa
gctcgaaaggaaggatgcctagaagcgcaagtcgtaattacctgacggtgcggaagacct
ggcgttaagtttgaagaagcatcctaatttgtcgtccca**cat**tgtttgtgtgatcctgta
agtgaatacctggctacgatcactttctcaggtcaaaatctcctctatggtgttgcccta
gagcaaatcacatgatatattttgttctttttttttttccctttatctcgttccacgcc
catttttttgtgttcgcgcgctctcctcctccatcctgaaaaatttgttcaatctttta
cctgatattctttcccgtataccaacctattcacacaaaatgactccaaggtaagttagg
caacaagaaccctcagttcaagaatggtgactcgatcacatctttgccatttcaagatca
agaaatagctgaagtttgtcgaccagcttttgcagtaaagcttatgctctaacatgagaa
atatgtggccatgcccaatttcttctgtcgatctcattctgcgaatgacggtgtcagttc
gtagctcttcccatactaacagttttttccagagttcagaagccagagactttagtcaag
aaggagaaagtcgatgagaagactgcccaagtacaagccgagtacgctaaggctaaggct
gccgtaagtaccacaccaaactggtctaataattccttttc

ORF (intron in bold)
ATGCCTCAGAAGACCACCAGCAAGTATTTGCTGATGATACCAATATTCAATCGACTAGCTTTCGGGCAGTGTTGT
CGTTTTGTTAACCATCTTTCGGGACTGAAA**gtaagctggtagtctgagaagtggaatcgaaatttgtagattgtc**
tgaaattttgtttctttactaacatcatgtttctagTTGGACGAGGAGAAGAAGGCCATCATTTCCGAGAGAACT
GCCAAGTACTACGAGGAGTATGAGGAAGCTGCTCGTGTTCTGGAGAAGGCTAAGAGCGACGCTGCTGCCTCTGGT
TCATACTTTGTCGAGGCTGAGCCAAACTTGGTCCTTGTTGTTAGAATCAAGGGTATCAACAAGATTGCTCCAAAG
CCAAGAAAGGTTTTGCAGTTGTTGAGACTTTTGCAAATCAACGCCGCTGTTTTCGTTAGAATCACTAAGGCCACT
GCTGAGCTTTTGAGATTGGTTGAGCCATACGTCGCTTATGGTTACCCATCTCTGGAGACTGTTAGGGAGTTGATC
TACAAGAGAGGATACGGTAAGCTTAACAAACAAAGAGTTCCTTTGACCGACAACGATATCATCGAGAAGGTTTTA
GGTCAATACGGTATCATCTCCATTGAGGATTTGATCCACGAGATCTACACCGTTGGCCCAAACTTCAAGCAAGCC
AACAACTTCTTGTATCCATTCAGACTCTCCAACCCAGACGGTGGATGGGGTGTCAGAAGAAAGTTCTTCCACTAC
ATTCAAGGTGGTGCTACTGGTAACAGAGAAGAGCACATCAACAGCTTGGTCAAGAAGATGAACTAA

Downstream (in bold start and stop codon of next gene)
acatagtcttaccaataaataggccttaatttgatatgtataacactattgttatcgcat
ttatttttgggatatccaatcgttgagtccgtacttgttacgagctatctcgatggcgta
gaactgtagacggactacgaaaatagttccttcttggtcgtactctttggaactgtcagt
tctgagcaaggttccaaaattgaagtcgtcgatcttttcttcaaaaggtttcatgaactt
tctccatctctctttaccgctcagagatttcatttcgtcctcattgatcttgctggcata
ttctggattttcaaactctggaaagtccttcaataaagcctcaaaaatctcatcatcgta
aggagtaagcttgagttttgtgccaggaattttggtgagaagattccagtaagtttcagc
ttgtgtcactgccttgacagcaaattgcttctcaatatcgtccaaattttcagcttcttc

ggcgttaaaagtacttaaact**cat**agttcttctgatctggtatgaatggtgtacggaaaa
agatggaacatgagagttttttgaaaatgtggagccagtcacgtgacccagtttcaatac
ataacctctcaagttgtgtagctcaattagaatttacgacggttggggtctaagatctgt
tgatcaaaaggtttccaagtagcaagtgctatctactgaacagtggaaccaaggcagtat
gaatctgcttgctgaaatcttgaaagcctgataataagctgctgaacaacgtttggttca
ctttgagttcagctgcttcggtgcaagggtgactacagttaatattacggacgctcgtat
gctggaatgaaaggtagtcttgtaaacatcatcaaaagaattgaagaggagctgagtgtt
ttggtgaagacaagtttaaccttcggtccaagatagtttccacctggtgtagagaccact
ggatgcatgataccgttagtaagcggcaagagacggttcac

AA
MPQKTTSKYLLMIPIFNRLAFGQCCRFVNHLSGLKLDEEKKAIISERTAKYYEEYEEAARVLEKAKSDAAASGSY
FVEAEPNLVLVVRIKGINKIAPKPRKVLQLLRLLQINAAVFVRITKATAELLRLVEPYVAYGYPSLETVRELIYK
RGYGKLNKQRVPLTDNDIIEKVLGQYGIISIEDLIHEIYTVGPNFKQANNFLYPFRLSNPDGGWGVRRKFFHYIQ
GGATGNREEHINSLVKKMN

SEQ0115

P. pastoris homolog of *S. cerevisiae* **SSB2 (YNL209W)**

Cytoplasmic ATPase that is a ribosome-associated molecular chaperone, functions with J-protein partner Zuo1p; may be involved in the folding of newly-synthesized polypeptide chains; member of the HSP70 family; homolog of SSB1

Chr 3, 0731

5' region (in bold stop codon of previous gene)
ccttcataccacaactttgcgacgcctgaggattgtagaaaatggcgtagaaaggccgcc
agcctttctgactgttcccgtccttgtcaaagaacaagaagttcacagatggttagattc
tttgatggccacttatcagatccagatgcgaacaatatcagagctctctcaattccatac
agaggatgcgaactcattcatttccatcatacaattgaccattgatggcatctggttgat
ggctctggataaagaatggcttaacagcaggttgacagttgaaggagttttaaccgctgg
tagtcacgatagacacagccaattacgctttcatctcccatacaatgtagtaatttcctc
aattactcttctggacaagctcactctggcagattgccaagaggttctctatcatgtatt
caaacgagcatcttctactgatactaagactcgaagctctttccaaaacttaacaagatc
ctcagaattgtacaactgcaagaaggaactccctttacaaccattaagccaggtggatat
attggtaacgcacaatttatggctcacaaaaaggcagacttcttatctaattggctgtaa
aggttccagaattgaactcatacgaactaactccaaggctaaaatcaagatttttagtgc
actagaagaggaattaatgcctaagcaacatatttctataactggaactaagtctcaagt
tcagaaagctgtctttctgattcagttttccttggacaaactacttcagggaatcagtga
ctgg**tga**gactatttattggtgagcaaaactagaatgggctataaattccatgccacagc
cgatgaaccgggtaacctttagttttacccacaatttactgtttcgctaacctagaaaa
aaaacctctagtggttggtgaaaaatttttttaggagagccttttttttttcttgtttcct
cttctttctcaaagaaactttcaagttttcgcatatatata

ORF
ATGGCTGACGGTGTTTTCCAAGGTGCTATTGGTATCGATTTAGGTACCACCTATTCATGTGTTGCTACTTACGAC
TCCGCAGTGGAGATCATTGCCAACGAGCAAGGTAACCGTGTTACTCCTTCCTTCGTTGCTTTCACTCCTGAAGAG
AGATTGATTGGTGACGCTGCTAAGAACCAAGCTGCTTTAAACCCAAAGAACACTGTTTTTGACGCTAAGCGTTTG
ATTGGTAGAGCCTTTGATGACGAGTCTGTCCAAAAGGACATCAAGTCTTGGCCATTCAAGGTTGTCAACGACAAC
GGTAACCCATTGATTGAGGTTGAGTACTTGGGTGAGACCAAGCAATTCTCTCCTCAGGAAATTTCGTCCATGGTT
TTGACCAAGATGAAGGAGGTCGCTGAGGCTAAGATCGGTCAAAAGGTCGAGAAGGCTGTTGTCACCGTCCCAGCT
TACTTCAACGATGCTCAAAGACAAGCTACTAAAGACGCTGGTGCCATTTCTGGTCTGAACGTTTTGCGTATCATC
AACGAACCTACTGCCGCCGCCATTGCTTACGGTCTTGGTGCTGGTAAGTCTGAGGAGGAGAAGCACGTTTTGATT
TTCGATTTGGGTGGAGGAACTTTCGATGTTTCCCTGTTGCACATTGCTGGTGGTGTCTTCACTGTTAAGGCCACT
GCTGGTGACACCCACTTGGGTGGTCAGGACTTTGACACCAACTTGTTGGAGTTCTTCAAGAAGGAATTCCAGAAG
AAGACCGGTAAGGACATCTCCGACGATGCCAGAGCCTTGAGAAGATTGAGAACTGCTTGTGAGCGTGCTAAGAGA
ACTTTGTCTTCTGTTGCTCAGACCACGGTTGAAGTCGACTCCTTGTTCGATGGTGAAGACTTCACTGCTGAAATC
TCCAGAGCTAAATTCGAGGCTATCAACGCTGATTTGTTCAAGTCTACTTTGGAGCCTGTTGAACAAGTTTTGAAG
GACTCAAAGATCGAGAAATCCAAGGTCGATGACGTCGTCTTGGTTGGTGGTTCCACCAGAATTCCAAAGGTCCAA
AAGCTGTTGTCTGACTTCTTCGATGGTAAGCAATTGGAGAAATCCATCAACCCTGATGAGGCTGTCGCTTACGGT
GCCGCCGTTCAAGGTGCTATTTTGACCGGTCAATCTACCTCTGAAGAGACCAAGGACTTGTTGTTGTTGGATGTT

ATCCCTCTTTCTTTGGGTGTTGCCATGCAAGGTAACGTCTTTGCCCCAGTTGTTCCAAGAAACACCACTGTCCCA
ACTATCAAGAGAAGAACTTTCACCACTGTTGACGACCACCAAACCACTGTTCAATTCCCAGTTTACCAGGGTGAG
CGTGTCAACTGTTCTGAGAACACCTTGTTGGGTGAGTTCGACTTGAAGAACATTCCTCCTATGTCTGCTGGTGAG
CCAGTGTTGGAGGCTATCTTTGAGATTGATGCCAACGGTATCTTGAAGGTTACCGCTGTCGAGAAGTCTACCGGA
AGATCTGCCAACATCACTATTTCTAACTCTATTGGTAGATTGTCTTCTTCCGAGATCGAAAAGATGATCAACGAT
GCTGACAAGTTCAAGAAGGCTGACGAGGACTTCGCTAACCGTCACGAATCCAAACAGAAGTTGGAGGCTTACGTT
TCCTCCATTGAGTCTACCATCACTGACCCAATTTTGTCTTCCAAGCTAAAGAGAAGTGCTAAGGACAAGATCGAG
TCTGCTTTGTCTGACGCTTTGGCTGCTTTGGAGCTAGAAGATGCTTCCGGTGATGACTTCAGAAAAGCTGAATTG
GCATTGAAACGTGTGGTTACCAAGGCTATGGCAACCCGTTAA

Downstream (in bold stop codon of next gene)
gtttattagtaaaatttagggtacgcaaagttaatcttacagtatatatgtgctgttact
tcttaaatgtaaatcattatttttatttaaccgatatagttatagccatccgcagcaacga
ttgtcccttcctggatacatttgaacacccactctttctctactatccggggtactttat
ctagaacacctttgacgattggtagatcatcacccagcttttttcttagagacccaactt
catcctgtatggtcgcctcctcagtcaaccaggaatagttgaatgtggtggaagtatcca
cgattataaatgaacattggaaaatatcagttgtaagaaatcctcctagaatcactatta
cttctaacaagtatggactcataacatttgtcagaaatggagatctatagcatggtcttg
cgatataaaacttcaggcccttgaacaagttgtacagtagtggaaattttagtgggtgtt
ctaactttatcttaggtatacctttcataagcattcgcaactccttacggtcaatgggag
aggtataactatccccaaatcggtcggagtttttctttaccttttctatcaaagactcat
ccgccagtaatatgtcgttaatttgtagaggaattatagttttcagctgaagacacttca
gaatgaagcctggtcgaaagacatttagacctttattaaagtactctgtgactttgaaag
taatcttgtcactaagtaaaacaagcctctcgccatctgcaagagctacgcaatcattct
tagttattctgccaccattctcgaccagcaaagcctccagctgagcaatagagtatcgtt
tttcacccaacatgtcacttaaaacataaaagatgtaatccttgaatatgctgctcttcg
aagatacactttgtttagcgagaattggttcagtgactattgtcctctttcgtttactca
ttagactgtgtgttttctctgttttgtgtctactctttgaa

AA
MADGVFQGAIGIDLGTTYSCVATYDSAVEIIANEQGNRVTPSFVAFTPEERLIGDAAKNQAALNPKNTVFDAKRL
IGRAFDDESVQKDIKSWPFKVVNDNGNPLIEVEYLGETKQFSPQEISSMVLTKMKEVAEAKIGQKVEKAVVTVPA
YFNDAQRQATKDAGAISGLNVLRIINEPTAAAIAYGLGAGKSEEEKHVLIFDLGGGTFDVSLLHIAGGVFTVKAT
AGDTHLGGQDFDTNLLEFFKKEFQKKTGKDISDDARALRRLRTACERAKRTLSSVAQTTVEVDSLFDGEDFTAEI
SRAKFEAINADLFKSTLEPVEQVLKDSKIEKSKVDDVVLVGGSTRIPKVQKLLSDFFDGKQLEKSINPDEAVAYG
AAVQGAILTGQSTSEETKDLLLLDVIPLSLGVAMQGNVFAPVVPRNTTVPTIKRRTFTTVDDHQTTVQFPVYQGE
RVNCSENTLLGEFDLKNIPPMSAGEPVLEAIFEIDANGILKVTAVEKSTGRSANITISNSIGRLSSSEIEKMIND
ADKFKKADEDFANRHESKQKLEAYVSSIESTITDPILSSKLKRSAKDKIESALSDALAALELEDASGDDFRKAEL
ALKRVVTKAMATR

P. pastoris homolog of *S. cerevisiae* **FBA1 (YKL060C)**

Fructose 1,6-bisphosphate aldolase, required for glycolysis and gluconeogenesis; catalyzes conversion of fructose 1,6 bisphosphate to glyceraldehyde-3-P and dihydroxyacetone-P; locates to mitochondrial outer surface upon oxidative stress (2 *P. pastoris* homologs)
Chr 1-1, 0072 see SEQ0065
Chr 1-1, 0319 see SEQ0066

SEQ0116

P. pastoris homolog of *S. cerevisiae* **AGX1 (YFL030W)**

Alanine:glyoxylate aminotransferase (AGT), catalyzes the synthesis of glycine from glyoxylate, which is one of three pathways for glycine biosynthesis in yeast; has similarity to mammalian and plant alanine:glyoxylate aminotransferases
Chr 4, 0416

5' region (in bold stop codon of next gene)
gcctagaaatctcctaaagcatggctgttccttttgaaattgaaccatacgaaattcttg
gtgtgaacaatgaagctacacctgtggaaattaaaaagtcctactacaaattatgcctga
tacaccatcctgataagaaatctggatcagactcatcaaatgacgaacatttccaaaaga

tccagtttgcttattcaattctctctgactcgaaaaggaggaaaagatacgacagtacag
ggtcactagatgacactgccctagacgaagatgggttcgattggaaagagtactttgaga
cgatgaaaagccagccagtcactgaggatttgattgaagaagatagagaaaagtacaagg
gaagtgatgaagagaagcaagatattattgatgctctccaatttttatgaaatggacgttc
caaagctctttgaggccattcctcatctcgagtttgacgagagtgaagaagaacgtatct
tccacttggtaacagaattggtagacagcaaacaggttgaaactacaaataaatgggaca
agtacaaacacaataggaaatccttcataaagagacaattgcgcaaattggaaaaggagg
ctgttgaagctgaaaagcttcagaagaaactggcgaagcaaaagaaagcagatcaggata
tttcaaccttgcaaggactacagtcaattattcaggctaaacaacgatcatctaatcaaa
agttggactcgttgataaataaactggaaaccgaagaggacgcaaaggtaaaatccaaac
gaggtaagaaacggggtaccaaaagggcagcagtcccagatattgacgaagaagagttcc
aaagaatccaggcaaagctgaaacga**tag**tgggggtcaccatatactaggttgttctact
gaaaaaaaaagcttattggtcgtgtttcacacgatccatcatcaacggcatcatcgcatc
atcgcttgcagtttagagtcagacccccaaacacccgcgag

ORF

ATGATATTTCAACGACACCCGCTTCTCATCTCAATGTTCTCCAAACTTCAGCGCTTCGTACAACCTATCAGAAGC
ATGTCAACCATCACTATTCCTAATTCCAAACAACCCGCTCACAGGTTGACCATGATCCCGGGTCCAATCGAGTTC
AGCGATGATGTGTTGGACGCCATGTCAACTCCAAGTCAGGCTCATACTTCACCAGATTTTATTCCAGTTTTCCAA
GAAGTTCTGCAAAATACCCGTAAGTTGTTCAAATCTTCCAACCCAACTTCACAGCCTATAGTAATTTCTGCTTCC
GGTTCTTTGGGTTGGGATATAGTCGGTGCCAACTTCCTCAACCGAGGAGAGAAAACTCTGGTCTTTTCCACTGGT
TTTTTTTCGGACAGGCTCGCTGATGCCCTTAGTTATTACACCGAAACTGATGTCGACGTCGTTAAAGCTCCATTA
GGAGAATCTGTTCCCCTCTCCGAGGTGGAGAAGGCTCTTTCTGCAAAGTCCTACGACGTTGTCACAATTACTCAT
GTCGACACATCTTCCGCGGTTATTTCAGATGTCGAGGCTATCGCAAAAAAAAATCCGCTCCATTCAACCTGAGGCC
TTAGTTGTAGTTGATGCTGTTTGTTCAGCTGGTGTTGAAGACATTCAATTTGATAACTGGGGTTTGGACTACGTC
CTAACTGCCTCTCAGAAAGCTGTCGGAGTCCCAGCAGGTCTTTCCATCTCCATCGCCTCAGAGAGAGCAGTTAAA
AAGGCCTTCGCCAAAAAAAGGCCTACTTCCTACTACGCTAACTTGCAAAAATGGATTCCGATCATGCAAGCATAT
GAACATGGTAAGGGTGCCTATTATGCCACTCCTCCTATTCAGTTGATTCACGCATACCGTGTATCTCTCAGAGAG
ATTTTGGAGCAGGGCCTTGACAACAGATTTGCCAAGCACGCCGAAACTTCTAACAAGTTTAAGGACAACTTAGAG
TCTTTGGGACTAAAACTTGTCGCCAAGAGGGAATTTGGTGCCAATGGGCTGACCGCCGTCTACTTCCCCAAGTCC
ATCAATGGACCTGATTTGTTGAAGAAAATTTCTGAAAAAGGAGTGACCCTGTCCACTGGTATCTACGCTGGAATT
GCTACCGAATACTTTAGGGTTGGGCACATGGGTGTTTCTGCTGTTGGAAAAGATAGACACGATGTTGATATTGTT
TTCAGGCTGATCAAGGAGACCTTGGAGGAGTTAGAGAAATAG

Downstream (in bold stop codon of previous gene)

ttgtagtaatttatttatattcttttgttcactaattatatt**cta**aacagacaatgggaa
ttctcctgcccaggcgctaatctcattcttgagttgagtgatctcgtcagagccttcagc
gatcttggccttgaaatctttaagcttgttggcttcaattggtaactcgctttggacttt
cttagcatattgaacagatttgtcgatgtagttagcgatcttgacgaagtcttcctcaga
agcacctctggtggtcatagcaggagcaccgatacgaacaccacctgggaccaaagcgga
tctgtcacctggaatggagttcttattcaaagcgatattgatgttctcacaaacggtttc
aattctggcaccatcaatatccttatccttcaatgaaacaaggaccatatgggagtcagt
tccgtcagaaaccaacttgtatccaagtctcttaaactcattctccagagccttagcatt
cttcaaaacctgttcctgatactgcttgaactcaggtgtagcagcttgcttcaaagcggt
agctaaagcggcgattgtgtggttgtgtggaccaccttggtggccagggaaaacggagaa
attgatggggttttcaaggtcgtagtagatttccttaccagtctttgggttgacagaacg
gacacccttacggaagaaaatcattgcaccacgtgggcctctcaaagacttgtgggtggt
tgtggttacaatgtcggcgtattcgaaaggagatggaatgacacctgcggcaatcaaacc
agaaatgtgagccatatcaacgactaagtatgcaccaactttgtcagcaatctccctcat
tctcttgtaatcgatcaatctacagtatgcagaggtaccggcaaccaggacctttggacg
gtatagaactgcagtcttctcaagcatgtcgtagtcgataatgccagtttccagatcgac
tctgtaaggcattgtttcaaagtaagtcgagactgcagaga

AA

MIFQRHPLLISMFSKLQRFVQPIRSMSTITIPNSKQPAHRLTMIPGPIEFSDDVLDAMSTPSQAHTSPDFIPVFQ
EVLQNTRKLFKSSNPTSQPIVISASGSLGWDIVGANFLNRGEKTLVFSTGFFSDRLADALSYYTETDVDVVKAPL
GESVPLSEVEKALSAKSYDVVTITHVDTSSAVISDVEAIAKKIRSIQPEALVVVDAVCSAGVEDIQFDNWGLDYV
LTASQKAVGVPAGLSISIASERAVKKAFAKKRPTSYYANLQKWIPIMQAYEHGKGAYYATPPIQLIHAYRVSLRE
ILEQGLDNRFAKHAETSNKFKDNLESLGLKLVAKREFGANGLTAVYFPKSINGPDLLKKISEKGVTLSTGIYAGI
ATEYFRVGHMGVSAVGKDRHDVDIVFRLIKETLEELEK

SEQ0117

P. pastoris homolog of *S. cerevisiae* **ILV5 (YLR355C)**

Acetohydroxyacid reductoisomerase, mitochondrial protein involved in branched-chain amino acid biosynthesis, also required for maintenance of wild-type mitochondrial DNA and found in mitochondrial nucleoids

Chr 1-1, 0432

5' region (in bold start and stop codon of previous gene)

tatttccagataatcatgacgcatccacctcgttacaatgtacctaaactaaaagacaac
gacagccccttggttgtgcagatcatcatctcctatcaaacagcacacaaaaaactggg
taataagtttagaacgagttacaaaatgtcttcctcctttgcaattctaatctacgcgg
aatctgtcaccctcttaggtttattctc**tta**cagtacttcccctagaatcccgacaagag
ctaaacaaaaacttaggccagaaagcaaagttcccttagcatataatttacctagctttg
ttaggctatttcgaacttgattccgttcaatcgcccactccacttcatcttcgacattat
cttccatcaattctccttctacagaaacataggctgacccatctaaagaagatctttcag
taatgtcttgtttcttttgttgcagtggtgagccattttgacttcgtgaaagtttcttta
gaatagttgtttccagaggccaaacattccacccgtagtaaagtgcaagcgtaggaagac
caagactggcataaatcaggtataagtgtcgagcactggcaggtgatcttctgaaagttt
ctactagcagataagatccagtagtcatgcatatggcaacaatgtaccgtgtggatctaa
gaacgcgtcctactaaccttcgcattcgttggtccagtttgttgttatcgatcaacgtga
caaggttgtcgattccgcgtaagcatgcatacccaaggacgcctgttgcaattccaagtg
agccagttccaacaatctttgtaatattagagcactt**cat**tgtgttgcgcttgaaagtaa
aatgcgaacaaattaagagataatctcgaaaccgcgacttcaaacgccaatatgatgtgc
ggcacacaataagcgttcatatccgctgggtgactttctcgctttaaaaaattatccgaa
aaaattttcttcccttctcttccaaatatcgtctccacaaa

ORF

ATGTCCGTAAGAAATGCCACTTTTCGTTTAAGCCGTGCTGCTATTGCTAAGAGAACTTTGGTCAGAATTGCTTCT
CAAGCCTCATTAAGACCAACGACACTGAATTCTGCCACTGGAGTCGCTCGTTCGCAGTTGACTTTTAGTCGTGGT
GTGAAAACCATTAATTTTGGTGGTGTCGAAGAAGTTGTCCACGAGAGATCTGATTGGCCAAGAGAGAAACTGCTA
GAGTATTTCAAGAATGACACTTTAGCTTTAATTGGCTACGGATCTCAGGGTTATGGTCAAGGATTAAATCTAAGA
GATAATGGTCTCAATGTTATTATTGGTGTCCGAAAGAACGGCGCTTCTTGGAAGGCTGCTATTGAGGATGGCTGG
GTTCCTGGTGAAAACCTTTTTGATGTCACTGAAGCTATAAAAAAGGGTTCTATCATCATGAATCTTTTGTCTGAC
GCCGCCCAATCTGAGACTTGGTCTACCGTCAAACCCTTATTGACTGAAGGTAAAACTTTATACTTTTCTCACGGA
TTTTCTCCAGTTTTCTCAAACCTTACCCATGTAGAGCCACCTTCCAACATTGACGTTATTTTAGCTGCTCCTAAA
GGTTCAGGTAGAACTGTTCGTTCCCTTTTTAAAGAGGGACGTGGAATCAACTCTTCCTATGCTGTTTGGAACGAC
GTCACCGGAAAAGCTGAAGAAAAAGCCATTGCTTTGGCTGTAGCAGTCGGATCTGGATACGTTTATCAGACCACA
TTTGAGCGTGAAGTTAACTCCGACTTGTACGGTGAGAGAGGTTGTTTAATGGGAGGTATCCATGGTATGTTTTTG
GCACAATACGAGGTTCTAAGAGAGAATGGACATACGCCATCAGAAGCTTTCAACGAAACCGTCGAAGAGGCTACT
CAGTCTTTGTATCCCTTGGTTGGAAAGTATGGTATGGATTACATGTATGATGCCTGCTCCACTACTGCTCGTAGA
GGTGCGTTAGACTGGTATCCAATTTTCAAGGATGCCCTCAAGCCAGTGTTCGAAGATCTTTACAATTCAGTAAAG
ACAGGTAAGGAAACTCAACGTTCTCTTGACTTCAATTCACAACCTGATTATAGGGAGAAACTTGAAAAAGAACTG
GAAATTATCCGTAATATGGAGATTTGGAAGGTTGGAAAGGAAGTCCGCAAACTACGTCCAGAAAACAACTGA

Downstream (in bold start codon of next gene)

gttaaccggcttgtttgattaatgccataagttaactgtccacagaagggacaagagaga
tactgtgtatctatgtagaaagcataatgagacctatgtcttgtgaatagactgggtaac
aaatttttgacctgtagagtgtgtgaggtggcaagagtatgaaactagtagattgggttc
ttcttcgataaccatcttgcgctgctggcgcagggatttgtcacaatacaatctgatgta
ctgtgtaacccggggactgaatttgtcgtcaaaaagatatctgtttccatcaggtcacgt
gaaagtgtaccatctctatttgacctttcaggtatataactgtctggttctcccaccat
tgtccgtaagtttttaatcgtaacttgca**atg**tcaattccaaaaccagaaggttctacac
tcagattgggttctacagctcctgatttcaaggccgaaacctctcgtggaccaatctctt
tccatgacttcattggagattcctgggtagtgttgttctcacatcctgatgattttaccc
ctgtctgcacaacagaactgggagctttcgctaaactgcaaccagagtttgagaagcggg
gcgtgaaactgattggcttgtctgccaatactactgattctcatcaagcgtggatcaagg
atattgatgaagtaactggctcacatctgacttttcccattattgccgatccagagcgca
agatcgctttggcgtatgacatgattgattttcaagatgcctctaatgtggatgataaag
gagtacaattcactattcgttccgtcttcattattgatccaaaaaagaaggtcaggttga

tcctgtcttatcctgcctctactggtcgtaacactgctgaagttcttagagtcattgact
cattacagactggagatagaaaccgtgtaaccacccctattaattgggttcctggtgacg
acgttatcgttcatccatcagtaactaatgaggaagcaaaa

AA
MSVRNATFRLSRAAIAKRTLVRIASQASLRPTTLNSATGVARSQLTFSRGVKTINFGGVEEVVHERSDWPREKLL
EYFKNDTLALIGYGSQGYGQGLNLRDNGLNVIIGVRKNGASWKAAIEDGWVPGENLFDVTEAIKKGSIIMNLLSD
AAQSETWSTVKPLLTEGKTLYFSHGFSPVFSNLTHVEPPSNIDVILAAPKGSGRTVRSLFKEGRGINSSYAVWND
VTGKAEEKAIALAVAVGSGYVYQTTFEREVNSDLYGERGCLMGGIHGMFLAQYEVLRENGHTPSEAFNETVEEAT
QSLYPLVGKYGMDYMYDACSTTARRGALDWYPIFKDALKPVFEDLYNSVKTGKETQRSLDFNSQPDYREKLEKEL
EIIRNMEIWKVGKEVRKLRPENN

SEQ0118

P. pastoris homolog of *S. cerevisiae* **RPS1B (YML063W)**

Ribosomal protein 10 (rp10) of the small (40S) subunit; nearly identical to Rps1Ap and has similarity to rat S3a ribosomal protein

Chr 4, 0524

5' region (in bold start codon of next gene)
ctcctgtgtttactatttctggtatggttgcctttatgtcaaagactttactcagtatga
ctttgaatagcttttgtatgtttatagaatggcttgtggatgagaaaataagggaagctg
acatagctctggcatatttctttgcttgattggtaatctcgatttgatcctcatctggta
gatctatgaatagatcatacttggtgcccaccaaaaacgggatggctgttttattgaaac
ctcttgcttgtctataccactccttaatagagtttagagtggattttctggtcaggtcaa
acataaatattatcaccacagcgtcagttgtcaccaatggcagcatgttaatgaactctt
tctccccacccagatcccaaatggagagcgtgatctgggtatttctgatgttgatgattc
tctccatgaaattagctcctagtgtctgtgtgtaattttcatcataacagccttctacgt
actttaccattaatgatgttttccctatctgggcatctccaatcaggccaattttgacgg
cgactgagttagtttccctttgactatgtggtacattctgttgctctgaatggcttcttc
tgtgctggtttcttctcggagagtctgg**cat**ctcaatttgcgtcgattaaagactgtcta
gagacgatcttagtgtcaactgcgtgtgtgaagtaaacaagataactaccctctccataa
ctggagaccgtgcatatggaacctgatcaaatagggaaataatatctaaaccataaacac
gtgataccttttagggtgtacatttgttcaagtagtatcgtcgtcactttggtgttcaga
tgtgtttcagtttaatgggagctcgcgtactagggtatccaccaattgtggtaagaacac
gactccttttttttcgctccctcactagtggtgtctctctctaaggattgaaaattcgc
aaagaaaaaaatcctttactaaacacttcaactacataaag

ORF
ATGGCTGTCGGTAAGAACAAGAGACTTTCCAAGGGTAAGAAAGGTTTGAAGAAGAAGGTTGTCGACCCATTCCAG
AGAAAGGAATGGTACGACATCAAGGCTCCTACCACTTTTGAAAACAGAAACGTTGGTAAGACTTTGATCAACAAG
TCGACTGGTTTGAAGAACGCCGTTGATGGTTTGAAGGGCCGTGTCGTCGAAGTCTCTCTTGCTGACTTGCAAGGA
TCCGAGGACCACTCTTTCAAGAAGATTAAGTTGAGAGTCGATGAGGTTCAAGGTAAGAACTTGCTGACCAACTTC
CACGGATTCGACTTCACCTCTGACAAAGTTAGATCTTTGGTCAGAAAATGGCAATCCCTGGTTGAGGCCAACGTC
ACCGTCAAGACTGCCGATGACTACGTGCTGAGAATCTTCTGCATTGCTTTCACCAGAAGACACCAAAACCAGATC
AAGAAGACCACCTACGCTCAATCATCCCAACTGTCCGCCATCCGTAAGAAGATGGTTGAGATCATGCAAAGAGAG
GTATCCAACGTCACATTGGCTCAGTTGACCTCCAAGCTGATTCCTGAGGTTATTGGCCGTGAGATCGAGAAGGCT
GCTGAAGGTATCTACCCATTGCAGAACGTTCACATCAGAAAAGTTAAGGTGCTGAAGCAAGCTAAGTTTGACTTG
GGTGCTTTGTTGGCTCTTCACGGTGAGTCTGTCACTGAGGAGAACGGTAAGAAGGTTGGATCTGAGTTCCAAGAT
GTTGTTTTGGAGACCGTCTAA

Downstream (in bold stop codon of previous gene)
tctaagtgacagtacttatgttttatagcgaagtagaataagaaattcgtccataatagg
tgaagaaaattgttgtagtgaccttagtattctttgacttccccgagtaggtcaattcgg
atgtaggttaaaatgatttttgaggtctcgagttgagggagtctgaaaatttcaaacatg
caagtcgaattcaaattgactacacttccatgcgtctacatagccatcggcttgaggtct
taccaaatttttaatatattatatggcttatggccaatttttttt**cta**gttgtgtaaacgc
ttctcgtaagagtgcaaattgttaacaccaccttcgagttgggctgatggagttctcagt
tcaaatctgacttctccagcatcgacagctgctctgagctcatccacagacttaactcca
atgtcttggcaagaatgttgaagaccagattgcaagtatggaatgaatttggtcacagat
cctttgtcgacaacggcaccagtaacaccttgggcaaccaaaactttgtcagattcggaa

aagtatctggaggtggaagcgttggcgttggtagcagtctgctccatggcatcaatagaa
cccatacctctgtaagttttcagtctcttaccatcacggtaaaagtattcaccgggtgat
tcagtggtaccagcaagcattcctcccatcataacacaagaagctccaagggctagagcc
ttgacgatgtgaccaatattctgaacaccaccatcggcaatacatggaacaccgaatttg
ttagcaaactcagaaacggcgtaaacagcagtaccttgaggtctaccacaggccatgact
tcctgggtgatacaaatagatccggaacccataccaattctaaggccatcagctccagca
gcgatcaaagaggcagcctgctctctggtgacgacattaccagcaatgatttgcagatta
gggtgattctccttgatccatttgatcatgttcaattggaa

AA
MAVGKNKRLSKGKKGLKKKVVDPFQRKEWYDIKAPTTFENRNVGKTLINKSTGLKNAVDGLKGRVVEVSLADLQG
SEDHSFKKIKLRVDEVQGKNLLTNFHGFDFTSDKVRSLVRKWQSLVEANVTVKTADDYVLRIFCIAFTRRHQNQI
KKITYAQSSQLSAIRKKMVEIMQREVSNVTLAQLTSKLIPEVIGREIEKAAEGIYPLQNVHIRKVKVLKQAKFDL
GALLALHGESVTEENGKKVGSEFQDVVLETV

SEQ0119

P. pastoris homolog of *S. cerevisiae* **POR1 (YNL055C)**

Mitochondrial porin (voltage-dependent anion channel), outer membrane protein required for the maintenance of mitochondrial osmotic stability and mitochondrial membrane permeability; phosphorylated

Chr 2-2, 0392

5' region (in bold stop codon of previous gene)
ttatgccatcgtgaacacaaatatcacaggtgcatttggtgctatttcctggtgtctatt
agactggcgtttggagcgccgtttcagtactgttgctctatgctccggtgccatttcggg
cctcgtggcagcaactccagcctcaggtatcattcctctttgggccagtgttattcttgg
tattgtatcaggagtggtttgtaactacgcaaccaagattaaagtcatttgtcgagtcga
tgattccatggatgttctagcagagcacggtatcgctggtgttattggtctcgtcttcaa
cgcattatttgggtcggctactgtcattggttatgatggccttaccgagcacgaaggtgg
ttggatagaccacaactggaaacagttgtacaaacagattgcattcatttttgcttgtat
tggatactcgatggccatcaccgctcttatctgtttcatcctcaaccgtattccatttttt
gcaactgcgagcttcagaagaggctgaggagaaaggtatggatgaggatcagattggaga
gttcgcttatgactacgtggaagtacgtcgtgatttttggcttggggatcaggcccaaa
caatggcttcaaggagccggaagttctggatcaggtagttccggttaatgatttcagcag
tgaccagaatgtgactaatgagaccaacgaatctgagaagcag**tag**agtaaatatagaga
tgatatttagtgtattctaatgcttatgtaatgtattaagcaaaaagttgtgtttatgag
ttagcatttgtcttagcaaacataaaattatgtcgacatttgcaacccgcatgtctagtg
tttttagatcgatcttcgatgtgtagaataatagcctccacgtgatgccccgcgattttg
ttgggtctcaatgcctccaacataaacccatcacgtataaaaagccctcttaaccctccc
ccctgtttcgtttgcttcatcacttaacctgaactatcaaa

ORF
ATGGCCGTACCAGCATTTTCCGACATCTCAAAAGCATCTAATGATGTACTAGGCAAGGACTTTTATCACTTGACC
CCAGTCTCTTTGGACGTGAAGACTGTTGCTGCCAACGGTGTAACTTTCACTGCCAAAGGTAAATCTGCCGGTGAC
AAGCTATCTGGAAACCTGGAGACCAAGTACGCCGACAAAAAGAACGGTTTGACTTTGACTCAGGGCTGGAACACT
GCCAATGCCCTAGACACCAAGGTTGAGCTGGCTGACACTTTGACTCCCGGTTTGAAAGCTGAGGTTGTCGGATCT
GTTGTTCCCGACAAGAAGAAGGACGCTAAGCTGAATTTGACCTATGCTCACCAAGCATTCACTGCCCGAACATTC
TTGGATTTGTTAAAGGGCCCTACAGTGAATGCTGACTTTACCGCTGGTAAAGACGGTGTCACTTTGGGTGGAACT
GCTTCCTATGACATTAATGCCGCTTCCGTCACCAAATATGCTTTCGCTGTCGGATACAAGGCCCCAGACTACTCC
ATTTCTCTTTCTGCTTTGGATAACGTCAAACTGTTTTCTGCCGGTTATTACCACAAAGTTTCTCCTCTCGTCGAG
GTAGGTGGTAAAGCTACTTACGACTCTAAGTCCTCTATTGCAAACCCAGTTGCTTTGGAAGTAGCAACTAAGTAT
CAGGTTGATTCGACAGCTTTCGTTAAAGCCAAAATAGCTGATTCTGGTATTGCCTCATTTGCCTACTCCCAAGAT
TTAAGAAAGGGTGTTAAGTTAGGTTTAGGAGCTGCTATTGACGTTCTCAAGTTGAACGAGGCCACTCATAAGCTA
GGTGTATCTCTTTCTTTCTCTGCTTAA

Downstream (in bold start and stop codon of next gene)
gcgttctaggtagcaagtttttttaaagatgaaaaattagtaatatgatgagtactcgtat
attgctgctatgtctagcgtacttctga**tta**ccccactcggacgaactctggtttggtgt
tcttgtcgatcagtaaatcggttttgaatgtctcgatgtactccacatcgccctcccctt
ttcccccggcaaaacgtccatactcatcaaatatgccagaaatacaccacccctgaagca

gttttctcatgatacccactaatacccccaccctatgtttgcccttattcgaatgaatta
acattgggtagttttctttgtgaacaattagcttcagagcctgctgaataaccgagtcat
ccttgaacataaacggctctacacagctttgcatatttaggtaatgaaaatctatacctt
gatcgcggagccaccggtaatagtcatagttgtccgtcttgtcacccaaatatatgattg
tttttaggttcagtttttgtaaatacggaaaattgaatgcttgaggatgacctgacctgt
agataccatcctctactatagcaaaattgattgggggaacaaattcaggacaccgatcta
aaattttgagagaagtgaactcatcttgaggctcctcgtattgttctaattcaccagctg
tggaatcatagattctcagctggcggattagctgatcaccttgagagta**cat**ataacctg
acggtgatttgaaggaaggaacttactaaattcgcgatgttctactatagcatcgatgga
atttatggttcgctagttggttgaacaaatgagtaaacacattcccgaaataatgagtat
tgtacttaaatacccactaccgtcctggcatatcccaaatgcctacgctaaccaccaatg
atgcttttccccttttgccaggttcctactaacctgtctttcagcatgacatagctttcac
taatagccctcaaatacgtagatcagatcgacccgtagttg

AA

MAVPAFSDISKASNDVLGKDFYHLTPVSLDVKTVAANGVTFTAKGKSAGDKLSGNLETKYADKKNGLTLTQGWNT
ANALDTKVELADTLTPGLKAEVVGSVVPDKKKDAKLNLTYAHQAFTARTFLDLLKGPTVNADFTAGKDGVTLGGT
ASYDINAASVTKYAFAVGYKAPDYSISLSALDNVKLFSAGYYHKVSPLVEVGGKATYDSKSSIANPVALEVATKY
QVDSTAFVKAKIADSGIASFAYSQDLRKGVKLGLGAAIDVLKLNEATHKLGVSLSFSA

SEQ0120

P. pastoris homolog of *S. cerevisiae* **RPS23B (YPR123W)**

Ribosomal protein 28 (rp28) of the small (40S) ribosomal subunit, required for translational accuracy; nearly identical to Rps23Ap and similar to E. coli S12 and rat S23 ribosomal proteins; deletion of both RPS23A and RPS23B is lethal

Chr 4, 0348

5' region (in bold start codon of previous gene)

tttgaactgagatttgctgaactggatattcttaatgtctgtaaaatctattgtggtcaa
ggttacattattgtccattccaacggagaccaagtatccatgactgtcatcaatggggag
atatttcatgtcggagagaagtctaccatgcaactgaaacgactgtatttcttttgttgc
tacgtttaaaatggatagttgtccatccataccccacagagtagcagaggagaactgtc
aatgattaataccactttggagatggacttgagaacttttgaaaaccagctagcttaac
accattgagatcaaaaacctgcagttcattcgatgaggtggagatcacaatacacgaaga
gggttgatctgaaagcttcaagttcccaacagtagttgaaactactaaatcagagcccaa
cggaaccaattcttcacctatcactggatttggcaatttccatacaggcaaagctacact
attattgagcgcctgaggatcgtcctcgtttgccttttcaactctaagaacttgtatct
gtcttccacgatgtcgtagagacgctcctccaacaaatccctattgtcagagaaagttcg
attgagttcaagtttgaacgcagccaagctctccttaagccctctttcagaaagatactg
tgcgatcaacgactcgacaatttcattgataga**cat**actttcacaatcagacctaaatg
ggatctcttgtgctgtatctcatggatagataggtcatgattaagtaatctaacatacag
atctcaggcacgtgatactgtagtagctgctacgctgctacatctctacatccttacaaa
taaattcattgaattactttaccacgtgatgcgcattagggcaatgtcactgctctccct
aaagagttaacaccacctatcggttcgcgacatggatcctctcactgcccaaaaatttca
aatttctctgttcaaagttctattatcacacaacagcaaaa

ORF

ATGGGTAAAGGTAAGCCAAGAGGATTGAACGCCGCCAGAAAGTTGAGAGTGCACCGCAGAAACAACAGATGGGCC
GACCTGCAATACAAGTCCAAGTTGTTGGGAACTGCCTTCAAGTCTTCTCCATTCGGTGGTTCTTCTCACGCTAAG
GGAATCGTGTTGGAAAAGATTGGTATTGCTTCCAAGCAACCTAACTCCGCTATTAGAAAGTGTGTCAGAGTTCAG
TTGATCAAGAACGGTAAGAAGGTTACCGCTTTCGTTCCAAATGATGGTTGTCTGAACTTCGTCGATGAGAACGAC
GAAGTTTTGTTGGCCGGTTTCGGTAAGAGAGGTAAGGCTAAGGGTGATATTCCAGGTGTCAGATTCAAGGTCGTC
AAGGTCTCTGGTGTTTCTTTGCTGTCCTTGTGGAAGGAGAAGAGAGAGAAGCCAAGATCATAA

Downstream (in bold stop codon of next gene)

tggttttgtagattatcagtatttacaatacaaatattagtactatac**tta**ttgcggaag
ccctcttccttcaatggagattcctcctttcttggccgacgacttgcgtttggtccccat
agtgttgagccttgagtttaatctcttgtattggtcctctccgctcagggatgatcccac
tccggaaggcagagtccttctggctgaactcatcagcagctcggtattgacttgaggagc
gatggtgtaagcagctgttcttgcttgggggatggccgcacgctttcttctttcagctgc

ctcttcgtaaagggttttggcctcctggggtctgacactgacccttccaggcaaggattc
cgcaaactgggcaggatcaaattgctgggtcctaacacgcatgtctttgtaattatccga
tgctcccattatgggtctggtccaatcactgatcaatttctcggcagttctcttaaggtt
gggttccactctcttggatttctggtagaataccatcactttgcctaaaccactttctct
caaatgagatgtcttgatgggaagtgacttgatagcgtcaaatagcactttctgtatttg
gtaagctggcaaggatgcatctggtaatggttccagccatagtcttactgaggccaataa
gttgttgtccagtatagaatcggccaagttggctctcaaaagtatgtcctttaccttggg
aagtagtttcaatttgaaaatggcaggttttccctgttcaatgttgtttgcatcacggat
ggcggattcttccatttgtaatttaagttgttgtataagctcatcctgcatctgctctag
atccacgtcatcgttcttctttctcttgtttggcttccttttgattgcagcatccagttt
ctcttcgaacaatcttgttcttctggcggcttcgtcttccacgacatcaggctcgtattg
agtttctctggcggtttgttgctttggtcttcttctggtgg

AA
MGKGKPRGLNAARKLRVHRRNNRWADLQYKSKLLGTAFKSSPFGGSSHAKGIVLEKIGIASKQPNSAIRKCVRVQ
LIKNGKKVTAFVPNDGCLNFVDENDEVLLAGFGKRGKAKGDIPGVRFKVVKVSGVSLLSLWKEKREKPRS

SEQ0121

P. pastoris homolog of *S. cerevisiae* **RPL13B (YMR142C)**

Protein component of the large (60S) ribosomal subunit, nearly identical to Rpl13Ap; not essential for viability; has similarity to rat L13 ribosomal protein

Chr 4, 0413

5' region (in bold start codon of next gene)
tatgactggaaaatcttcttcaactcgttcttggtagactcgtcaacgaacttttgagtg
taagcgatcaagcccttagcaagagcttgtctgatggcgtaaacttgggaaacgtgacca
ccaccggtaaccttaattctgatgtcaatgttggcaaatttgtccaaaccaactaataat
agtggctcgtaaactttgtgtctcaagatctcaggttgcaccagagtgataggtgaaccg
ttaaccttgataagaccggaaccctgtttaacgtgggcaacggcagtagcggacttcttc
ttaccaaaggtctagaaatggatgttagttagattttcttcaatagtgctcccatctgat
agaaatgagtcctaccaatggaaagagcgataatacttgggaagtgaacggcgatacatt
ctgaacaccgaatccccttaatgatgaacttacctgaacactttgtacggacttagcttc
ttgaga**cat**tgttactcttacaaagaacaagttttttgttagcaatttttttcacaatcc
gagagagagcgtgagcgatcaggtagtgtgtcgcacaataagagttaacgttttcagagg
tttgactgggagggtgaaacagttagtgtgtagccattgactgttagtgcacgacttgaa
ataactgtttgcaaagaaatctaaatctcgaatttgtgactgctgtaactgaactaggaa
ggccagacctgtaaatacacgcgatttatctcccaaaactagagcgatcacgtgcattat
aagactaaattgctttgtgaaaaatagtccacctggggaaccgagcatcgtccgcccatc
tccaaagaagacccgcgcatactctcccgccagcacttttcccttcagattttggttacg
aagcccctgctcgaaatctttctcgatctgtgccttcactcacctttacaaattttcac
caagttaaaaaaagcttccaggttaactacaacaatcaaag

ORF (intron in bold)
ATGG**gtatgttctttaaatcacccaagcactatgctctcgtagtgctagcaaataccgtttccgacaaatgaagt**
actaaccaagttatagCTATCTCTAAAAATTTACCACTTTTGAAGAACCACTTCAGAAAGCACTGGCAAGAACGT
GTTAGAGTCCACTTGGACCAGGCCGGAAAGAGAACCTCCAGACGTAACGCCAGAGCTGCTAAGGCTGTCAATCTG
GCTCCAAGACCTTTGGACTCTCTGAGACCCATTGTCAGATGCCCAACCGTCAAGCACAACAGAAAAGTCAGAGCT
GGAAGAGGATTCACCATTGAAGAGCTTAAGGCTGTTGGTTTGAACGCTAACTACGCTAGAACCGTTGGTATTGCT
GTCGACACCAGAAGAAAGAACAGATCCGAGGAGGGTTTTGAGTTGAACGTCCAAAGATTGAAGGAGTACCAATCT
AAGTTGGTTGTCTTCAAGTCTCGTGCTGAGGCCAAGGATGCTAAGCAAATCTCTCTGGGTGCTTCCTTCCCAGTT
GAGCAACCAGCCATTGAGGTTGGCCCAAGAGCTGTCGTTGTTCCTGAGAAGACTGCCTTCGAGACTTTGAAGCAG
GCCAGACTGGACCAGAAGTATGCTGGTCAAGTTGCTAAGCGTGCTAAAGAGGCTGCCGAGGCTGAAGCTAACAAG
AAATAG

Downstream (in bold start and stop codon of previous genes)
acgctttgtagtgtatattgaatccgggagtaatgtaacacatagtaggctagtggtcta
catcgtagaagttcgtttggtggattgtacaaacttttcaaccaaatggagtattggcca
gatatgcttggacagtccagatatcttaagctccctacttctagatgtacatcgttacag
cttagactttacaaggagtgatcgatcgatcctagacttgtcgtcaagataaaacgaaaa
gattaaagaatcaaaaatttctgccacttaggcataggcaaggtacatcatttataattc

aaaccatgtgtattcgtatttagaatgtaac**tag**tcgaatagaccgaagcccatgtcgtc
gtcggactcttccttggcctcttcctccttctcctcagcagcctcctcagcatcagcaga
agaagcggcagcaccaccagcagcagcaccggaggcaggggcagaggaaatgttaaagaa
caaatccttcaagttcttaccgtccaaagccttggaaaaaaggtcagcgtaaatcttgtc
aaccttaacgtcagccttggtggtcaaagccaacaagttgtcggaggagatctcaatctc
agaatcagccaaaataagggcggcgtatgataaagcggtttcggtagc**cat**cttcgttta
atagagataatcggtttaagcagatattgcggactaatcccttgaagaagagagaaggac
ccactggttttttttccatgttcgaaaagttacgtggtgtaatgatgtattataa**tca**c
gtgcgtgactctggggccgatgcccttcgtcgttgcacttcgtctcgattctccataaga
ctaagacgaagagcctcctcaatcatacgctcttctagttcatgtatctcgtcgttttgt
ctagctgctcgtagtgctctcctttgactgttagatatatgctcctcctctgttatcagg
gaagtagcatgaattgctgatgcggtagccgatcgccgagc

AA

MAISKNLPLLKNHFRKHWQERVRVHLDQAGKRTSRRNARAAKAVNLAPRPLDSLRPIVRCPTVKHNRKVRAGRGF
TIEELKAVGLNANYARTVGIAVDTRRKNRSEEGFELNVQRLKEYQSKLVVFKSRAEAKDAKQISLGASFPVEQPA
IEVGPRAVVVPEKTAFETLKQARLDQKYAGQVAKRAKEAAEAEANKK

SEQ0122

P. pastoris homolog of *S. cerevisiae* **RPL12B (YDR418W)**

Protein component of the large (60S) ribosomal subunit, nearly identical to Rpl12Ap; rpl12a rpl12b double mutant exhibits slow growth and slow translation; has similarity to E. coli L11 and rat L12 ribosomal proteins

Chr 1-1, 0189

5' region (in bold stop codon of next gene)

agcattttcagaactttgcggatgaaatagctgagggtagtctacggtttaggacggaaa
gggagcgagtgcaagatatttgttttcgttgtagctctgctgttcgccaattaacaactg
aagttaaaactgcagcatcccagctgaaaaagattcatccatctcggggtcgtttcaagg
cagattacgaaaggggagtagaccaaatactccgatccagtattggcgtggctgtatttt
ccactaaagttaacaagaaaattaaagatgctgagtttgtaactcaatgcctgaaggaga
atccaagcaatcaagcatacgatgtttcattgagcagagcccaataccatcggtttttcc
gcctgcggagattaaacata**tga**ggaagtgatagttatgtagtggtctttcttgtacagc
tttggagtatgatagtgaatggccatgccagcccccaagtggagtctccgcccagtgtat
ttctttgactacggatcaaatcgttaagcagagcgaggaatcggttgactttgagccttt
gtacaatagttattatttaatgaaaggttggataagagtctaccaagatccctttgagc
acgacaattttgactaacgactttgaacacccaaccacctggaaaacaggtggttaaagc
catcggtcagtcggtcttccatccatgttccttgaaccagtacgaactatacgagctcat
cttatcaccaaaattctggtcacctggatttgtgactgtgaaattctaacacttagcgta
atgatctgaccggactcgatacttcacatgtaaactttcaaagctccttcgcgaatcata
tttccagcctcactctccttccaacaagtaagtctccctggtagttgtcttttcagtgcc
gtgagaaaaacctccccccgatttcctcgcgaagaatattctgaaaatgaaaaaatcacc
atccatttttttctttcctagtaactcctcacagaaacaag

ORF

ATGCCTCCAAAGTTCGACCCATCTGAAGTTAAGTTTATCTACCTGAGAGCCGTTGGTGGTGAAATCGGTGCTTCT
GCCGCTTTGGCCCCAAAGATCGGTCCTTTGGGTCTGTCACCAAAGAAGATCGGTGAAGACATTGCTAAGGCTACC
ACCAAGTTCAAGGGTATCAAGGTCACCGTCCAATTGAGAATCCAGAACCGTCAAGCTGTTGCCTCCGTTGTTCCA
TCCGCCTCTTCTTTAGTCATTACTGCTTTGAAGGAGCCCCCAAGAGACAGAAAGAAGGAGAAGAACGTCAAGCAC
AGTGGTAACATCCCATTGGAAGAGATCATTGACATTGCTAGACAGATGAGAGACAAGTCTTTCGGTAAGAACTTG
GCTTCCGTCACCAAGGAGATCTTGGGTACCGCTCAATCTGTTGGATGCCGTGTAGACCAAAAGAACCCTCACGAC
GTCATCGAGGCTATCGATGCTGGTGAGATTGATATCCCAGAAAACTAA

Downstream (in bold start codon of previous gene)

atgtgtaataattgagatatttaaacggttttgatgaaagatggtttgacttccagctgg
tttttgatatttgatttgactttcatttttttttccatttttttctccacctgacccgtttg
ccaacactctggtcccatttttttactctgtactgggttccaattttacaagcaaactaga
taaaaagcttatgattattacggagtttgtttagctggaagtttcgttaggggggtatca
ctattgtggtactttctggagctgtcagtatgatgaggcctcgtcgtggttgttagctga

cattgtatttaagttttgttcaggcagaaaattggaaaaaaaatcaaacaaataatcaaa
gtctccagctattaggtagtaagccagtaggttgtaagttgtagtgtcttctctcacctt
tttgctaggaagtagtcttaggttaggacgaagtctagtgagcgttaccatcgctcacga
ggttttctagcacgtatgaatggtgaccacggtccatgataccagtttcattcgttatag
accttataatgagttcgtcactgcttctgacaggttaaccaggcttctaactaaagcacc
ctggccatccgttaaacttttcagctcatcacccaaaagtcgtgcgcgggctacattata
atatttttttttgtatttctaaggctcatcaggtatactagctacagagtcttttgaac
atcaatcacgctacaacatagagtgattcagaccacagagcaagcacagcatttccacta
agactgcattttattcttacacaatttttttctcttgcatttgattgtctggttctagatt
tcact**atg**gataatcctataaggttagacctacaagactcattgaacaagttcaaagagt
acttcaacaagctggatttaccacttccggctccattcttgttagaaaatgagagcgtgc
cccaaacgttgaatacgctggccatgttgtcactaaatgac

AA
MPPKFDPSEVKFIYLRAVGGEIGASAALAPKIGPLGLSPKKIGEDIAKATTKFKGIKVTVQLRIQNRQAVASVVP
SASSLVITALKEPPRDRKKEKNVKHSGNIPLEEIIDIARQMRDKSFGKNLASVTKEILGTAQSVGCRVDQKNPHD
VIEAIDAGEIDIPEN

P. pastoris homolog of *S. cerevisiae* **PGI1 (YBR196C)**

Glycolytic enzyme phosphoglucose isomerase, catalyzes the interconversion of glucose-6-phosphate and fructose-6-phosphate; required for cell cycle progression and completion of the gluconeogenic events of sporulation
Chr 3, 0456
See SEQ0061

5' region (in bold start and stop codon of previous genes)
agttgagataagaaggatgaatctgtgtagttatccgcataatcctgtttcaaccataat
aacttcttccatggatgacgagc**cat**ttgagtggactcggcgcaaaagtgacaacgtttc
tagagggctactcgaaagcactttctctggagattagtaatcgatggtcaggatgggaac
cgaatgtctttggctggcctttccccaaatttgagaacaaaattcaattgctgaggtaat
acttttttccccatacaagaaaggccatcgtgtaattatctcttaccttctactattcga
gaagttgtaaccaa**atg**gctgataacatcgaaaccatccgtcctggaatcaactaccagg
acaaactattgaccgaaatagacattcttgatgatgttaaacaactgtcaacgcagatgg
aacaggaagggaaaggattcgtcttccctaaagatcattactccaagataaaatcactaa
aggggttgcaagtgaagctgttgaaagaaatgcaagagttggctgaaattcagcagtcta
ggcatcaagacgaaactgcatacaggcagaagcttcaagacttggaagatactacatcag
aacttagaaacgttgctttgacg**taa**taggtcacctgtgttactagatgtcggatgtctg
tgattgcagtaacacccggcattcggtcagctgtctacctttcaggatatcctcccccca
ccgagctcggacgcctttctcgcccacccaccacgcaaagttccgcccacccttatacttt
gtggaagtcttttcgccgcgcaccacacttgcccgttcgcagaaaaagaaattccaggta
gcccttattccgtcaggcaataaatatataagcgattgcagacaatctgacgtcccatcc
ccttgtctatttaaacctcctaggttgcttaaatttaaaatctattcgagtcccaacctt
tccattcttccggaataattcaactccaaccaattgataaa

ORF
ATGCCGTCTCTATTGCAAGAGGACAATGCTACTTTCAAGCTCGCATCCGAACTACCAGCTTTCGAAGAGCTAAAA
GAGCTTTATAAGTCAAAGGGAAAGAACTTTTCTGCCAAACAGGCTTTCCAAAAGGATCCAGCCAGATCTTCCAAG
TTCAGCCACACTTTCAAGAACTTCGACGGGACTGAGGTGTTTTTCGACTTTTCCAAGAACTTGATCGATGATGAG
ATTCTCGCCAAACTGTTCGACTTGGCCAGACAGGCAAACGTCGAGAAACTCCGAAACGAGATGTTTGCCGGAGAA
CATATTAATGTCACAGAGGACAGGGCTGTTTTCCACGTCGCTCTGAGAAACAGAGCTAACCGCCCAATGTACGTT
GACGGCAAGAACGTCGCCCCAGAAGTTGATAGTGTTTTGCAACATATGAAGGAGTTCTCTACGCAGGTTCGCGAT
GGTACCTGGAAGGGATACACTGGTAAGCAGATCACTGATGTGGTCAACATTGGTATCGGAGGCTCTGACTTGGGT
CCAGTCATGGTGACAGAGGCATTGAAGCCTTACGCCCAGGAAGGACTGCATGTTCACTTCGTATCCAACGTGGAC
GGTACCCATATTGCTGAGACTCTAAAATACTTGGATCCTGAGTCTACTCTTTTCTTGATTGCATCCAAGACTTTC
ACAACCGCTGAAACCATCCGTAACGCCAATACTGCTAAGGACTGGTTCCTTTCGAAAACTGGTAACAAAAGTGAG
GCAATTGCCAAGCATTTTGCTGCTTTATCCACAAATGCCGAGGAGGTCGCAAAGTTCGGTATCGACACTAAGAAT
ATGTTCGGTTTTGAAAACTGGGTTGGTGGACGTTACTCTGTGTGGTCTGCTATCGGTCTTTCAGTTGCCATCTAC
ATTGGTTTTGACAACTTTGAGGACTTCTTGAAGGGTGCCGAAGCCGTGGACAGACATTTCCTGGAAACTCCTCTG
GAGCAAAACATCCCAGTTATTGGTGGACTACTCTCCGTTTGGTATACTAACTTCTTTGGAAGTCAGACACATTTG
GTCACTCCATTTGACCAATATATGCACAGATTCCCTGCCTACTTACAACAATTGTCCATGGAATCCAACGGTAAA

TCTGTTACCAAGGGCAATGTTTTCGCCAACTACAGCACCGGCCCTGTCGTCTTTGGTGAGCCAACAACAAATGCT
CAACATTCATTCTTCCAATTGGTGCATCAAGGTACTCATTTGATCCCTGCCGATTTCATTTTGGCTGCAAAATCC
CACAACCCTGTTGCAAACAACGCTCACCAAATCTTGTTGGCATCTAACTTCTTGGCTCAAGCCGAGTCTTTATTG
CTAGGAAAGACTGAAGAGGAAGTAGCTGCTGCTGGTGCTACTGGTGGTCTAATTCCACACAAAGTATTTTCAGGT
AACAGACCAACTACATCTATTCTGACACAGAAAATCACTCCCGCAACCTTAGGTTCTTTGATCGCTTATTATGAG
CACGTCACATTCACCGAAGGAGCTATATGGAACATCAACTCATTTGATCAATGGGGTGTTGAGCTAGGAAAGGTT
CTAGCCAAGGCTGTCCAGAAAGATCTGCAGGATGACAGTGCCAACGTTGAAGAAAGCCACGACTCATCCACTGCT
CAATTGATCAAGAAGTTCAAAGCTTGGGCTTAA

Downstream (in bold start codon of next gene)
ggcgttctagttatagagatgtatattgtaatatggtatatgagtactaaagaagtgatt
atctaataaaaatttgagtaatgcactgacattttcactatcagttttggaagagggtat
gggacgcgttcattcacacgtctttggtaatt**atg**gccgaagaagttgaccccgttagtc
tgggtgaccagacgattctccgctataaaatatggaagaagaactctccatacttgtatg
attatttccaaagcaagtctctgctgtggccctctttatctgttgagttttgccagaca
ttgaacgaaatgacgaagatgagttcgattaccaaaggcttatttttggaacatttacgt
cgggagccagcaatgagtttttgaactttgggatgtttagtagacacaacgaagtctctt
tgagagagtcactgaggaactctctggacaattttgacagcgtcaaaggagaaatatcac
cactagtattaccatcttccaaagactccaaaaactctaatcgcagctgcgaaaagttga
gcatcatccaacgaatagcacataatggagaagttaataaatgcaaatatcttcctcaaa
atcccgacatcatagcgacaattaataattatgggagtgtttcgatttttgatcgaacaa
aacatccttctcaaccactaagcggcacaattaaaccagatatttactgtacatatcata
aggatgaaggttcctgtttgagttggaatcctagcgttgaaggggaactgttgtcaggct
caatggacggaacggttgtattatgggatatcaaaaagtacacgagggacaaagattctc
ttgatccatacaagatattcattgctcatgacaatggctgcaatgaccttaaattcatcc
ctagacacacatcaatttttggttctgttggagaagatggctttttaaactttgggata
ccagacagggactggatcccgttaagtcaacacgacttcat

AA
MPSLLQEDNATFKLASELPAFEELKELYKSKGKNFSAKQAFQKDPARSSKFSHTFKNFDGTEVFFDFSKNLIDDE
ILAKLFDLARQANVEKLRNEMFAGEHINVTEDRAVFHVALRNRANRPMYVDGKNVAPEVDSVLQHMKEFSTQVRD
GTWKGYIGKQITDVVNIGIGGSDLGPVMVTEALKPYAQEGLHVHFVSNVDGTHIAETLKYLDPESTLFLIASKTF
TTAETIRNANTAKDWFLSKTGNKSEAIAKHFAALSTNAEEVAKFGIDTKNMFGFENWVGGRYSVWSAIGLSVAIY
IGFDNFEDFLKGAEAVDRHFLETPLEQNIPVIGGLLSVWYTNFFGSQTHLVTPFDQYMHRFPAYLQQLSMESNGK
SVTKGNVFANYSTGPVVFGEPTTNAQHSFFQLVHQGTHLIPADFILAAKSHNPVANNAHQILLASNFLAQAESLL
LGKTEEEVAAAGATGGLIPHKVFSGNRPTTSILTQKITPATLGSLIAYYEHVTFTEGAIWNINSFDQWGVELGKV
LAKAVQKDLQDDSANVEESHDSSTAQLIKKFKAWA

SEQ0123

P. pastoris homolog of *S. cerevisiae* **RPS25A (YGR027C)**

Protein component of the small (40S) ribosomal subunit; nearly identical to Rps25Bp and has similarity to rat S25 ribosomal protein

Chr 2-2, 0326

5' region (in bold stop codon of previous gene)
agagtttttcaatgatttcaagtaatccagatactttgtgagaatgggcgtgaaatcagc
acttggatcgctagacctggcttgctcgatagcggcaacaaactgattgtttaacgctac
aatcttgtgcgtgttatcatcagccaaagaaggtgcggaattggtgctgctcactgtagg
tacaggtggatgcgaaaactttattcctgtggactgtttagaagacaacatacctctggg
ttttaagattttccttttggataagacttccgctgaagccctatgcaatcccgatgactg
ttcgtccgatgtgtcatcgtcgtgcccgagagtttccctttgaatttgggaaccagctct
cctctttcctga**cat**ttactcttaattgaggtttatttgatttgataaatgtttatagga
ggtttatgttcgcgctcaggtgtcaacaaatgaaaagggttaagccctgcacaagcatca
tcagaaaggcaaaacgccgtcctattttgtcacatgactttaaaaaactggcaggacttt
gatattttgggacgttcacactaccttgcgttggctcccgcaccctagagaaaaagagaa
aaaatcagaaccaacttttcttagtacacaggttaaagaacaggtaagttaacaagttta
gtgtgctgaaaaaagattatctatttgaaatttgaaatagtatctgctacttgctgagcc
atgtacagggagacagaaatggatagagagtcaaacgacctcatggacgaatattagttg
taatagaataaggtactcaagattcagcatggcaattacccaactctcaagtgaatttaa
aacttgatctcaagatttatgaaatagtccgtttccctgttgtctgtgctttgttgctag

atattcattttctttcccgttttcgaatgagtgctttttttctctttcactatcatcac
tgctttccgaggattatttcagtcatactaacccttttagc

ORF
ATGCCACCAAAGATTCAACAATCCAAGGCAGCCAAGGCCGCTGCCGCCATGGCCGGTAGTAAGAAGTCTAAGAAG
AAGTGGTCCAAGGGAAAGGTTAAGGACAAGGCTCAACATGCTGTCATCCTGGAACAGGACAAGTACGACAGAATC
ATGAAGGAGGTTCCAACTTACAGATACGTTTCCGTCTCCGTTTTGGTCGACAGACTAAAGATCGGTGGTTCTATG
GCTCGTGTTGCTCTGAGACAATTGGAGAACGACGGTGTCATCAAGCCAGTTCTGTTGCACTCTAAGCAACAGATC
TACACTAGAGCTACTGCTTCCGAGTAA

Downstream (in bold stop codon of next gene)
attagtgtacatctgataatatagtactaccacgtatgataatgtagagaatagtcttcc
ttgtcgagtgtgtttgcagttttcttgagtttcaaggtttaaatgctggtatattagttc
atcgaaggtttcagccaatagcaccttaaatcaatcaaactaattcgactcttacgaaag
agcctactgtgtttagtatcgaagtcgtttacctttcatgttgaatagcttcctctctga
ccctaacatttcaagatcctcctaaagttacccggattgtgaaattctaatgatccacct
gcccaatgcatttttttctttattcagtttacctttttacctaatatacgagcttgttaa
agtaagtggcactgcaatactaggcttattgttgatattatgatgaatcgttttcacaaa
cttgatttcctgtgaactcaccatgtactaaggaaaaaaacatgcatcaccatctgaata
tttgacgattagtccaaccaattagtataccccggttgtcgttactaacatcatagccag
actcagtgatcacaatgttcgcaagtatgtaggcattgtgttccgactgaaagtttaatg
atgaaaaaaaaatttcagttctgcttctgagactaatgaagataactaaatttttaccaaa
accactgaaaactttcattaggaaccgaacctattagtcgtatgtagagaagtatgctaa
tttctaccaagtttcccat**tca**gccaccagacccacgccttccaggaatctttgaacttg
ccattgaaagaaccttctttggggctccgacccccttactgacaccgcctgagattcctc
ctccaacttttgatactccagaaatcactgaattaaccagatgttccttactagacatat
gagcggaatttattgtttttttcttttgtgacgggggctgaagtttcagcagtcttgtttt
cttcctcgacgaatgcagcgtcgtagtcctgatttcttttc

AA
MPPKIQQSKAAKAAAAMAGSKKSKKKWSKGKVKDKAQHAVILEQDKYDRIMKEVPTYRYVSVSVLVDRLKIGGSM
ARVALRQLENDGVIKPVLLHSKQQIYTRATASE

SEQ0124

P. pastoris homolog of *S. cerevisiae* **ASC1 (YMR116C)**

G-protein beta subunit and guanine nucleotide dissociation inhibitor for Gpa2p; ortholog of RACK1 that inhibits translation; core component of the small (40S) ribosomal subunit; represses Gcn4p in the absence of amino acid starvation
Chr 1-4, 0063

5' region (in bold stop codon of next gene)
taaagctcgttatcctccttacacaagaaaaaatagagcgactctggagctcttagataa
aggtgtggtttggtttgaccgttccggtaacgtcccttcaagaagatcgtatttaacgaa
agttattggatactgcggattgatggttacaattgctgctgcattaattggtaaacttgt
ttttt**taa**acttttattgtttcgtctgtatcatctgttaagagatctgcatctttgataaa
ctgtttgagcttatcattggcaagattggttgccttttttgttgatatttatgggaataat
ggtaatatctgaatggtcataattggtcccagattgctctttttttttgaagagaatctcg
atggtataccccgcttatctggttgtctagtaacttgatttcctttctcaaatgtatcgt
agcatttgataagttcccataaaagtcgtgagtgtaatttttgaattcggacttgatctc
tgctaatttcggcaggtcttcatttctagaggccaactcttcaccttctcggagtttttt
ggtggcattggccaaatttttttagaatttcagctatctttgaatctatgtcatgtaagga
atcaagacgttgcttgatgtaagatagatgatcttctgacatgaacgcaagttagtaaaa
cactgattaccgaaaagtggtaattgttagctggatcttgtcttttcggtaggtgataca
atgacgtttacgttgatatgaacagtctcctctactatctgttgaaacacgtctacccca
cggctcataagtaattcgctaaccttgttaaggggcatcgcgcacttctgagttcgaggt
tcgaaaacaaagtcacataactgcctatttagtcgaactatgtcatagctcttctttctt
aacatccacacaccagccactcgttctgctcgatcgttgaaaaaaaatagatagtacata
cggatattttattctcttagtgcaatagaatagttcataaa

ORF (introns in bold)

ATGGCTGACAACAGAGAAGTTCTTGTTTTGAGAGGTACCTTGGAAG**gtaggttattttaatggagtcgacaaata**
cactttctgctgcaattttcggcgtgaaccacaacgtatactaacttttttttcagGACACAGTGGATGGGTCACT
TCCTTGGCAACTTGCCCTACTAACCCAGATATCTTACTGTCTGGATCCAGAGATAAGACCCTGATTGTCTGGCAA
TTGAACGGAGATGACCAAGCTTATGGTGTGCCAAAAAAGGCTTTAAAGGGACACTCTCACATTGTTTCTGACTGT
AAACTTTCATTGGACAGTGAATTTGCCCTATCTTCCTCCTGGGACAAGACTGTTCGTTTGTGGAACCTAAAGACC
GGAGAGATCATCAAATACGCCGGTCACACTTCCGATGTTTTGTCTGTGGACCTGTCTCAAAACCTTAGAGTTATT
GCCTCTGCTTCTAGAGATAAGACTATTAAACTTTGGAACATTGTTGGTGAAGAGGTTGCCACTTTGGAGGGTCAC
ACCGACTGGGTTACTTCTGTCAGATTTGTTCCTGGATCAGCTTCTGACATTATTTCTGCTTCATCTGACAAGTTG
GTTAAG**gtatgttttttctttgcatggttttttcaagctttcaatttttcaatcccgtcttcatttatcgcc**
aattccttctgttcacatgatgttaaaacaatattgctacttcagacagtgatctacactggatgagacttttta
caccaagatctctgatataaacaaaggaaatggatggcaattcaaaatcgggtttgaacttgaaatattgattat
ttagatatgctcttactaactcactatcttagTCCTGGGATTTGAACGAAATGAAATTGACCGCAGACTTTGTTG
GTCACACTGGTTACGTTACGGCAATTACCGTTTCTCCCGATGGATCCCTCTGTGCTTCTGCTGGAAAGGATGGTT
CCATCATACTTTGGGATTTCTCTGTTAAGAAAGTCTTGTACACTTTGGATGCAAAAGAAGAAGTTCATGCTGTTG
CTTTCTCACCTAACAGATTTTGGTTGTGCGCTGCTACTTCCAAGTCTATTAAAATTTTCCACTTGGAGAAGCGTA
AGCCAATGGATGAGCTGAAACCTGAATTTGTTTCCACTTCCACTAAGGCTAGAGACCCACACTGTATTTCCTTAG
CTTGGTCTGCTGATGGTCAGAACTTGTTTTCCGGTTATACCGACAACGTTATCCGTGTCTGGCAAGTTATGACTT
CAAGTTCGTAA

Downstream (in bold stop codon of previous gene)
gttaacacacgttcatattaatatactctgtggtttttatat**cta**ttcttctgattcttg
tagctgttcgattgctctaagggcttcttcagcatcttcaacaaaaccaagatcagcctc
taaaccccgagagcttgtcaatgcttcgttaatcaatcttcgagccgtttgagcccgacc
aaatcgtagattgactagccccaatccatacttactgatcgtagctgcctgtttgagatt
atagtcactttgtctttgtttctcaaggattttctcctcggcagcctgcggcccatcttt
gcttataactagcatgttgtttgatgatcgaaatctggaggagttattgttaactattgc
agaaaatatcgtttgggcagcattcaaagaacttttagcagccgcttcacaaacttgcaa
ctcttctttcaatttggagctttcctcagaggtgagctgaaaatcatcgccttgtcccaa
ctttttttgatgttttcactattaaactcagcaatagcggattgaagataaagaagaga
accaatagaactctgagccaaataaatatccccaataggttcccccatcgccattaagta
ctgaacaacatcttgctttactttaattgctcctccaaaatctcccatacgcagcaaaag
ggaagaaaatacttccattgcgtctaacatgttatctacaaacatgctagcagcttttcg
atttctccaaatacgtgaaatcatttccacactttcaagttccgtagggttttggccctg
tgaagctatcaactggttcatattctcttgaagagctccccccctggctttgaatggtcat
ggcggcggcttctaagctaagattaaacaatgttttgaggtcctctaattcactcggact
ggcatcctgcgacaactccgtagccttcaccaaagacatcaaacagtgttttattgcttt
aggtttgtccattagttctttttttattagttgggttaacca

AA
MADNREVLVLRGTLEGHSGWVTSLATCPTNPDILLSGSRDKTLIVWQLNGDDQAYGVPKKALKGHSHIVSDCKLS
LDSEFALSSSWDKTVRLWNLKTGEIIKYAGHTSDVLSVDLSQNLRVIASASRDKTIKLWNIVGEEVATLEGHTDW
VTSVRFVPGSASDIISASSDKLVKSWDLNEMKLTADFVGHTGYVTAITVSPDGSLCASAGKDGSIILWDFSVKKV
LYTLDAKEEVHAVAFSPNRFWLCAATSKSIKIFHLEKRKPMDELKPEFVSTSTKARDPHCISLAWSADGQNLFSG
YTDNVIRVWQVMTSSS

5 *P. pastoris* homologs of promoters used for expression of proteins in *S. cerevisiae*

SEQ0125
P. pastoris homolog of *Saccharomyces cerevisiae* **GPD1** (YDL022W) and **GPD2** (YOL059W) (glycerol-3-phosphate dehydrogenase)
NAD-dependent glycerol-3-phosphate dehydrogenase, key enzyme of glycerol synthesis, essential for growth under osmotic stress; expression regulated by high-osmolarity glycerol response pathway; homolog of Gpd2p
chr 2-2, 0111

5' region (in bold stop codon next gene)

caaagagttgggtgttcctttagacaaatgttttataagtgaaacttccacccaaagtgt
tcccaatacgtcggcaactgcagcatctgcagcctccgatttgaatggtatggcagtgaa
gaatgcttgtgacaaactcaatgaaaggctatctcctgtgaaagaaaagctaggagactc
ggccacatgggaggatattattcgaacagcttatcttgaccgaatttctctctctgcaac
tggattctacaaaactccaaagattggatatgtgttcggggatccaaaccccaagcctgc
atttttctactatacacagggctcagcaatcagtgtagtcgaagtcgatactttgactgg
agattggtcatgcctaagctcccatattaaaatggaccttgggaggccgataaaccatgc
aatcgacacatatcaaataacgggcgcatatatgcagggagtcggactttgcacaatgga
gcaaagtctctggctgcgtaacaatgggcgcctctttacaaccgggcctggcgcttacaa
agtgcctggatttagagatttgcctcaaaaattccacgtttctattctcaaagaccgcga
gttcaagcatctagacacaatatggagatccaaaggcataggtgaaccgcctctgttctt
gggattcagcgttcatttcgcactacgtgatgctatagcaactgctagaaggagtcaagg
catcgaggaagggtgcaatggcctacccttcagaagtccattgaccacagaacgtatcag
aaccatgatggctgatcctattcttttggcggctcaagtgcccgccgaaggcaatgagtg
gttcattgaggcc**tag**aagtgttgtgattattttatttatttaatttatttttgcctttc
aatttttttatttttcggctgccaccctagcgcgcgctcgcactaaaggctacttctccgt
ctccaacccagcgagacggactccaaagttccttctagtaa

ORF

ATGTACTTAACCAGCACAGTGAGAGCGCTGCCAGTACATTTTTTCAGATCAAGACATTGTATTAGAACCATGAGT
AACATTGTTGAGAAGAAACAGTCCGAATTCTCGAGAGCTGAGCCTTTCCGTGTGGCGGTTATTGGCTCCGGAAAT
TGGGGGACTACTGTTGCAAAAATTATTGCCGAGAACACCAATGAAAGACCAGAAGAATTCGTTAGAGATGTCAAT
ATGTGGGTCTACGAAGAAGACATTGAAGGCCGCAAGTTGACGGATATTATCAACGAGGAGCATGAGAATGTTAAG
TACCTTCCTGGGGTAAAATTACCCAAGAATTTGCATGCCACCCCAGACTTGTTGGCTGTTGCAAGCCCTGCAGAT
ATTCTGGTGTTCAACGTTCCTCACCAATTTTTAAGCCGAATCCTACAGCAATTGAAGGGTAAAATTAAACCGACC
GCCAGAGCCATTTCATGTCTAAAGGGCCTCAATGTAGAGAAGAACTCATGTCAACTTTTGAGCACACAAATTGAG
AACGAATTGAACATTCATTGCGGAGTGTTGTCAGGAGCTAATCTTGCACCTGAGATAGCTAGAGAATGCTGGTCT
GAAACCACGATTGCCTACGTTAAGCCAAAGGATTATCGTGGCAAGTATGACGATGTGACACCTTTCACGATCAAA
CATTTGTTCCACAGACCTGCTTATTTTCATGTGCAAGTTATTGAGGATATCGCTGGTGCATCCTTGGGTGGAGCT
CTGAAGAACGTCATTGCTCTAGCTGTTGGATTTGTGGATGGGCTTAATTGGGGAGATAATGCTAAAGGTGCCATC
ATTAGACTGGGACTTCGTGAAATGATTCATTTCGGACACACGTATTTCCCTGGTGCTAAGAGTTACACTCTAACC
TGTGAAAGTGCTGGTGCTGCCGATCTTATCACTTCCTGTGCAGGAGGCCGCAATTTTAAGGTTGGAAGAGAAATA
GCTGCCACAGGAAAGCCTGCTGAGCAGGTAGAAGCAGAGCTTCTGAATGGTCAGTCTGCCCAAGGAATTATTACT
GCTGCGGAAGTGTATGAATTTCTGTCAAGCAAGGGAGACTTAAACAGCTACCCCTTACTCATTACGGTATATCTG
ATCTTGAAGGGGGAACGGTCAGCTGAATGTATTCCAGAGTACTTCAACAAGACCGAGGACGTTAAGCATTGGGAA
GACTAA

Downstream (in bold start and stop codon previous genes)

ttagaagcctgtgtgggttcagcttgcagtactagacctttgaggcaactatcccaaccc
ccatgctaaataaacaattgtatactagcgcgaacatcgaactgttaaatgtgattatat
ttcatgtaccgtaacctactcgcttaaccacctttgatcatgtcagctcctccagaagct
ctacagaaagtatgttgcactttggttcctgcttattactatgtactaacagataggttc
tggtggaa**atg**aaccagcaaatacaaaaatcccaatcggaattggcaatgactcgtattc
aacttcaaagaaacgagaccaatcttcgcatactaaaattggcaaactctgaggtctcca
attctaaacaagatcacgtttgggaatctgttggaaagatatttctcaagacatcggttc
aagattattcatccagtttgaagaacgaaacaacttccatcaatgatgccgtgaaggctc
taaagaccaaagaaacatacttacaaactacattggaaaatacagtacaggcaattaata
agtatgtgggaaaaccctcctcc**taa**ttaggtcaat**cta**aactgcactatatccacctct
atttgagcttccaaagattttacctgccaactctctcaagaatccagggatccttgatac
gacttcaacaacggcatcgacaaactcactgaaagcctcattagaagatccccggttgct
gacattcaaccatgcagtgccaattatgtacgcagcaaagccgatcactccgagtataaa
aaaccatgtaaaccatccccaagacccgtcagaatcagagtcatgttttccgtccttatc
cttgtccttgtttttgtcgtctttctacctttgttatcagacctacatgcagctttgga
ctttattcgtaactccaattgttgacggttccagccaactaattctaaaaggtctttgtc
ataatctccgtctttactacactcaaaaacgatgtcagcag

AA

MYLTSTVRALPVHFFRSRHCIRTMSNIVEKKQSEFSRAEPFRVAVIGSGNWGTTVAKIIAENTNERPEEFVRDVN
MWVYEEDIEGRKLTDIINEEHENVKYLPGVKLPKNLHATPDLLAVASPADILVFNVPHQFLSRILQQLKGKIKPT
ARAISCLKGLNVEKNSCQLLSTQIENELNIHCGVLSGANLAPEIARECWSETTIAYVKPKDYRGKYDDVTPFTIK
HLFHRPAYFHVQVIEDIAGASLGGALKNVIALAVGFVDGLNWGDNAKGAIIRLGLREMIHFGHTYFPGAKSYTLT

CESAGAADLITSCAGGRNFKVGREIAATGKPAEQVEAELLNGQSAQGIITAAEVYEFLSSKGDLNSYPLLITVYL
ILKGERSAECIPEYFNKTEDVKHWED

SEQ0126

P. pastoris homolog of *Saccharomyces cerevisiae* **ADH1** (VOL086C) and **ADH2** (YMR303C) (alcohol dehydrogenase)

Alcohol dehydrogenase, fermentative isozyme active as homo- or heterotetramers; required for the reduction of acetaldehyde to ethanol, the last step in the glycolytic pathway

chr 2-1, 0472

5' region (in italic TATAAA signal)

cgcagcgttttctgacggtactagaggactcttaggggaaggtagaatcaataaagatca
tattaggtaagcaaattttggatggaataggagactaggtgtggatgcgcgatctcgcca
aattgcacgaccagagtggatgccggatggtggtaaaccgtttcttccttttaccaccc
aagtgcgagtgaaacaccccatggctgctctccgattgcccctctacaggcataagggtg
tgactttgtgggcttgaattttacacccccctccaacttttctcgcatcaattgatcctgt
taccaatattgcatgcccggaggagacttgccccctaatttcgcggcgtcgtcccggatc
gcagggtgagactgtagagaccccacatagtgacaatgattatgtaagaagaggggggtg
attcggccggctatcgaactctaacaactaggggggtgaacaatgcccagcagtcctccc
cactctttgacaaatcagtatcaccgattaacaccccaaatcttattctcaacggtccct
catccttgcacccctctttggacaaatggcagttagcattggtgcactgactgactgccc
aaccttaaacccaaatttcttagaaggggcccatctagttagcgaggggtgaaaaattcc
tccatcggagatgtattgaccgtaagttgctgcttaaaaaaaatcagttcagatagcgag
actttttttgatttcgcaacgggagtgcctgttccattcgattgcaattctcacccccttct
gcccagtcctgccaattgcccatgaatctgctaatttcgttgattcccacccccctttcc
aactccacaaattgtccaatctcgttttccatttgggagaatctgcatgtcgactacata
aagcgaccggtgtccgaaaagatctgtgtagttttcaacattttgtgctccccccgctgt
ttgaaaacgggggtgagcgctctccggggtgcgaattcgtgcccaattcctttcaccctg
cctattgtagacgtcaacccgcatctggtgcgaatatagcgcacccccaatgatcacacc
aacaattggtccacccctccccaatctctaatattcacaattcacctcac*tataaa*tacc
cctgtcctgctcccaaattctttttttccttcttccatcagctactagcttttatcttatt
tactttacgaaa

ORF

ATGTCTCCAACTATCCCAACTACACAAAAGGCTGTTATCTTCGAGACCAACGGCGGTCCCCTAGAGTACAAGGAC
ATTCCAGTCCCAAAGCCAAAGTCAAACGAACTTTTGATCAACGTTAAGTACTCCGGTGTCTGTCACACTGATTTG
CACGCCTGGAAGGGTGACTGGCCATTGGACAACAAGCTTCCTTTGGTTGGTGGTCACGAAGGTGCTGGTGTCGTT
GTCGCTTACGGTGAGAACGTCACTGGATGGGAGATCGGTGACTACGCTGGTATCAAATGGTTGAACGGTTCTTGT
TTGAACTGTGAGTACTGTATCCAAGGTGCTGAATCCAGTTGTGCCAAGGCTGACCTGTCTGGTTTCACCCACGAC
GGATCTTTCCAGCAGTATGCTACTGCTGATGCCACCCAAGCCGCCAGAATTCCAAAGGAGGCTGACTTGGCTGAA
GTTGCCCCAATTCTGTGTGCTGGTATCACCGTTTACAAGGCTCTTAAGACCGCTGACTTGCGTATTGGCCAATGG
GTTGCCATTTCTGGTGCTGGTGGAGGACTGGGTTCTCTTGCCGTTCAATACGCCAAGGCTCTGGGTTTGAGAGTT
TTGGGTATTGATGGTGGTGCCGACAAGGGTGAATTTGTCAAGTCCTTGGGTGCTGAGGTCTTCGTCGACTTCACT
AAGACTAAGGACGTCGTTGCTGAAGTCCAAAAGCTCACCAACGGTGGTCCACACGGTGTTATTAACGTCTCCGTT
TCCCCACATGCTATCAACCAATCTGTCCAATACGTTAGAACTTTGGGTAAGGTTGTTTTGGTTGGTCTGCCATCT
GGTGCCGTTGTCAACTCTGACGTTTTCTGGCACGTTCTGAAGTCCATCGAGATCAAGGGATCTTACGTTGGAAAC
AGAGAGGACAGTGCCGAGGCCATCGACTTGTTCACCAGAGGTTTGGTCAAGGCTCCTATCAAGATTATCGGTCTG
TCTGAACTTGCTAAGGTCTACGAACAGATGGAGGCTGGTGCCATCATCGGTAGATACGTTGTGGACACTTCCAAA
TAA

Downstream (in bold stop codon next gene)

gccgaatagtttgtatacgtc**tta**tgtaatgagtttcaatgaattacttatttttacctc
tccttttggctcaattcaactagcctctgtagcaatctgtttgcgaagaagaacttatc
taatttttcatgggttttccccacgtttttgaaaagactttgtcttaactctctgtgatc
aaaccgttgtggtggaccgctgattgattgtgattcttcttccagatctagcgactctgg
cagataagagtccccggagtcaataattaccccaagaatataccttgaaacctccgagtt
ggccagtatttgagataacagcagctccttgtatttgtaagtgccttcgttatcaattga
tccgatgtccagcattattttgatatcaggtctcccaatcttgacaaacttttctacaat

gtgtttttgaagcagttgcacaagaggagcaatatcttcaaggtatgagtcttgctggtc
gccgttcaattttaaaataagaaatgtatctgagttccctgtcgtcccaataactgctac
actccccaactttgtgacagagagaaaatgatgctcttgtgaagtggaatatatggcgtc
aatgctactttgtaatttatggttgaacgcattctcagtatctccataaacttgaaagcc
cacggggtaagagacgccagatgctaattctctatgcaattgcgactctgtgtaagtgtg
actcaaaagtccaatacagtagagatcactcacatactgaggggtcagtgtgtcggagag
ttcacctaccaatggacagtactgggctaattcggttagcaagactcgacatatcgggat
cccatactgaatctcataggacattattgcctgatcattgtctccgtttccgtactcgct
caaattggttctcattgtcactagcaagtcttttcgatctgtagacgtcaaggggttgaa
taatggagttatctctttttccaaaacaccgtccatttctc

AA

MSPTIPTTQKAVIFETNGGPLEYKDIPVPKPKSNELLINVKYSGVCHTDLHAWKGDWPLDNKLPLVGGHEGAGVV
VAYGENVTGWEIGDYAGIKWLNGSCLNCEYCIQGAESSCAKADLSGFTHDGSFQQYATADATQAARIPKEADLAE
VAPILCAGITVYKALKTADLRIGQWVAISGAGGGLGSLAVQYAKALGLRVLGIDGGADKGEFVKSLGAEVFVDFT
KTKDVVAEVQKLTNGGPHGVINVSVSPHAINQSVQYVRTLGKVVLVGLPSGAVVNSDVFWHVLKSIEIKGSYVGN
REDSAEAIDLFTRGLVKAPIKIIGLSELAKVYEQMEAGAIIGRYVVDTSK

SEQ0127

P. pastoris homolog of *Saccharomyces cerevisiae* **PHO5** (YBR093C)

Repressible acid phosphatase (1 of 3) that also mediates extracellular nucleotide-derived phosphate hydrolysis; secretory pathway derived cell surface glycoprotein; induced by phosphate starvation and coordinately regulated by PHO4 and PHO2

chromosome 2-1, 0103

P. pastoris promoter sequence partial known, ORF sequence known (Pho1 PPU28658)

5' region (in bold start codon of previous gene) (in italic TATAAA signal)

gaactagcagtttccactattgtgatgtccattttatcctccgtgtcctcttttctagca
tgttgtgttcctacttggtctttttatggagcccgtgatggggcaccttggagttgtta
ctcacgtccgtcaggacgtttcttttcagttcctgtgtcttagaccgcgtatgacgtata
gagtagacattttcgttaggctgaacattaga**cat**agctcttctagcacttgttcttagt
agaggatcttgccagctacaagtaaaaggggatagttaagggaagaaaacgcggatttgt
atatagggctagctttgtttcaactgaattgatcggattgggaaaaacaacaacaaaaaa
gggatggtgtttggatgctacttgaaatctgcttttaacgtgttgcaacatctctggtgt
tgtgatgttggatcgagttagaggaaaggaaaaaagaaacaaaaaaaaaaggcgaaaagg
gaaaagagaacaggcaaaagccgtgctatggatgtcaaggacgatggtcttccatcggtt
cacaagcccatggatctttgttcccaaatgggaaaagtaaggcttctgagtgtttttcca
tcccggatccctattgttacttttgcttaacattccaatattcttcaacggttaattgat
taacactgtaacctctgcccatgtgcttcatccaaatctggtaatctgctttctatttct
gccaaaatagttaatctatgagacatgtgccctccaattgcgcagtagatcgagtggaag
tcttctttgcgtaacactcaaagtatacccctgttagtctttattcacctgttgctgctt
ggtgcagttaccattattgtttccacttgaaaagcttgtttttttttgatagcacagaaa
cgtgggctccgataagctaaacttcaacgagaa*tataaa*agctgaaaagattcttgtcaa
gaacttgtacaacgaccaataagtctttcaaggcatcagac

ORF

ATGTTTTCTCCTATTCTAAGTCTGGAAATTATTCTCGCTTTGGCTACTCTCCAATCAGTCTTTGCGGTTGAGTTG
CAGCACGTTCTTGGAGTCAACGACAGACCCTATCCTCAGAGGACAGATGATCAGTACAACATTCTGAGACATCTG
GGAGGCTTGGGCCCCTACATCGGTTACAATGGATGGGGAATTGCTGCTGAGTCTGAAATTGAATCCTGTACGATT
GATCAGGCTCATCTGTTGATGAGACATGGAGAAAGATACCCAAGTACCAATGTGGGGAAACAACTAGAAGCTTTG
TACCAGAAACTACTAGATGCTGATGTGGAAGTCCCTACAGGACCATTGTCTTTCTTTCAAGACTATGATTACTTC
GTCTCTGACGCCGCTTGGTACGAGCAAGAAACAACTAAGGGTTTCTACTCGGGGTTAAACACCGCTTTCGATTTT
GGTACCACTTTGAGAGAACGATATGACCATTTGATAAACACTAGCGAAGAAGGAAAGAAACTTTCTGTTTGGGCT
GGCTCTCAAGAAAGAGTTGTTGACACCGCAAAGTACTTTGCTCAAGGATTTATGAAATCTAACTACACCGATATG
GTTGAAGTCGTTGCTCTAGAAGAGGAGAAATCCCAGGGACTCAACTCTCTAACGGCTCGAATTTCATGTCCAAAC

TATAACAGCCATATCTACAAAGATGGCGACTTCCCCAATGACATTGCTGAAAGAGAAGCTGACAGATTGAACACT
CTTTCTCCAGGATTTAACATTACTGCAGATGATATTCCAACAATTGCCCTATACTGTGGCTTTGAACTAAATGTA
AGAGGTGAGTCATCCTTCTGTGACGTCTTGTCAAGAGAGGCTCTACTGTACACTGCTTATCTTAGAGATTTGGGA
TGGTATTACAATGTTGGAAACGGGAACCCACTTGGAAAGACAATCGGCTACGTCTATGCCAACGCCACAAGACAG
CTGTTGGAAAACACAGAAGCTGATCCTAGAGATTATCCTTTGTACTTTTCCTTTAGTCATGATACCGATCTGCTT
CAAGTATTCACTTCACTCGGTCTTTTCAACGTGACAGATCTGCCATTAGACCAGATTCAATTCCAGACCTCTTTC
AAATCTACCGAAATAGTTCCCATGGGAGCAAGATTGCTTACCGAGAGATTATTGTGTACTGTTGAAGGTGAAGAA
AAATACTACGTTAGAACTATCCTCAACGATGCAGTCTTCCCACTGAGTGACTGTTCCTCTGGCCCTGGATTCTCT
TGTCCGTTGAACGATTATGTTTCTAGACTTGAGGCATTGAACGAGGACAGTGACTTTGCGGAAAACTGTGGAGTT
CCTAAAAATGCTTCCTACCCACTTGAACTATCATTCTTCTGGGATGACTTGTCATAA

Downstream (in bold stop codon next gene)
aaatggtaaggaatgttttgcatcagatacgagttcaaaacgattaagaagagaatgctt
ctttttttgtttctatccaattggactattttcgtttatttttaaatagcgtacaacttt
aa**cta**gatgatatcttcttcttcaaacgataccacttctctcatactaggtggaggttca
atggatcctacaaactccaacgatacaaatccaacaccaacgataaccagtaaaaaccct
atcaggccgcggaaaatcgaacctccactgatcaacactcccaaaaagatgtagaagcca
cctcttcccaaaaatgacaacaaaaacgatagatacggcgtaactttaggtaccaattcg
ggctttatttctagagctgcaatcacacctccaaatgcaattataaacagactcaagaga
aactaaaacttgggtcagtttggcgggcttgaaaaagatgtcacataatacttacagcat
tgagaccatttagcaattgtagtgctcccgcaaaaatctgtagaagaaccaaagttagta
attttccctgaatgacagttctgcttcttatcttgcttaccgtggcccctgcaacaaata
gattggcggctttgaatagagaattgtaattttgttccgtttccatctctctataagctt
caacttttactgctaggttaggtgaccaaaaataaacaccccaaaaatggtagtgtcccg
gtgaaacaccaagggcctagacgttgaaactcagttatcagatgaaattcaactatataa
ctaactgagtataagtcttctatgttaactcttagtagctaaatcactataaagttctgg
aattgaaaacggaagtcttctcagctcaaaacctaaacaaaaaaaaataataaacattctt
taggctcaaattcttgtcacctcaagccccgctccgttcacctttctcaaaaatatattc
ccgatacatatcttgtatagtttcacaaacactaatcaatc

AA
MFSPILSLEIILALATLQSVFAVELQHVLGVNDRPYPQRTDDQYNILRHLGGLGPYIGYNGWGIAAESEIESCTI
DQAHLLMRHGERYPSTNVGKQLEALYQKLLDADVEVPTGPLSFFQDYDYFVSDAAWYEQETTKGFYSGLNTAFDF
GTTLRERYDHLINTSEEGKKLSVWAGSQERVVDTAKYFAQGFMKSNYTDMVEVVALEEEKSQGLNSLTARISCPN
YNSHIYKDGDFPNDIAEREADRLNTLSPGFNITADDIPTIALYCGFELNVRGESSFCDVLSREALLYTAYLRDLG
WYYNVGNGNPLGKTIGYVYANATRQLLENTEADPRDYPLYFSFSHDTDLLQVFTSLGLFNVTDLPLDQIQFQTSF
KSTEIVPMGARLLTERLLCTVEGEEKYYVRTILNDAVFPLSDCSSGPGFSCPLNDYVSRLEALNEDSDFAENCGV
PKNASYPLELSFFWDDLS

SEQ0128

P. pastoris homolog of Saccharomyces cerevisiae **ECM17 (YJR137C)**

Sulfite reductase beta subunit, involved in amino acid biosynthesis, transcription repressed by methionine

chr 3, 1084

5' region (in bold start codon next gene)
cagttgttctttaagctgtgtcaatgtattctgataatctctcagagtaacattatattt
agtacgatcacttgtagaaacgttttgcagctctaaattgagtctctctaacaggtcata
agagtcgtctaagttttcttctacaagccttatctgccttgttctgtcaacagaattgga
cgccgatatgttagccagatttcttttagcttcagctaatgtcatctgaatatcagacga
gtatgtttcaaataatctgga**cat**agtcagcaggtattaaactggagtattgaataagag
gctaaattgtttggttctttggttggctacgcaataatggcaatatgtttggcactcagt
atatattagtggtgtaggggtgtatacaatcttgcagtaaagtctacagctatcgccaaa
gaaaagagaagattacataatgactcaaccgatgaaaactactaggtccattcccattac
tagcagaaaaaagctcaaaagttcattgtttacagtaattacaagggttcttcagattcg
tatgtttgttctcgcctattttctcgttgctttattgacttaggacttccatgatctatc
agaagttcactttgaaggtataatcacctatgattatcggctagcgagggcgagatcata
gggtcctatcatcttgcctcgcgtagctgcgtccatcatatgcaagtctaggaaagcatt
tacacttttaatttcaagggcattctgaatttggctcagaccaaacttaacaggtgatc
agagaaaactttaatcccaaagcccgaatgtagcctatagggcgtcttcgctgacataag

cgaatggaaatctggagtttgtacatcgaagtgtggctcactatcacgtgtgatagagta
actcaacagtatggagtatgaaaatgtttcctcatatattggaaatgttatcccaaagtt
gaagccgaaatactatcctacacactttggactaatacgaa

ORF

ATGACCGGTTCTGTTGGTGAGATTGTTAACGTGCTCACCCATGTGCCGAATCGTGGGCATGTTTACACCACCAAG
CTCTCAAAGGGTCTGTCAAATAATTTCAGTAAAGCTCTTGATGATGCCTCAATAGACTATCAGAGCTTAGTATCG
AATAATGACCCATTCGCTCTAGTTATTGAGACTGTCAACAAGAACTCATTTACTACCGTGCTGACTGATGCTGAA
ACACTTCTTTTCAGCGTTCCATTTCTTAGTGAGGTCAGATCAGGTGATAAGCTTATCGTCCATGCTAGATTACCG
AAGAACGATTACTCGATTGTATCTGCTTTGAAGGATCTAGATTTCGTTGTTTTGCTGTCTAGCTCAGTACAAGAA
ACACAAGATATAGCCTTGGCCGGTTATAAGTTTCTGACTCAGAATAATAAACCTGTACTTCACTTTTTCAACGAA
GACTATTCGGAACCCATAGTGGATCAGCTGTCACCTGAGGTGGTTTCTTCATATTTGAAAAGTGAACAAGAGCCA
TTCGCAGAGAAACTCTCAATTTATAACAACTTTGAGTATTCAGGAGCTAAGAATCCTTTGGATGTCGTTGTTTTG
TTGGGTGATTCTGGAATTTCACCGCCTACTCACACAATAGATTATGGTATTATTAACATCAGAGTTTTCCGACCC
TTTGACATAGAATCTTTTTTAAATGTGCTTCCTTTTTCTACTCAGAGACTTGCCATTGTTGAACAATCTCCCAAG
AAACCTTTCGCACAGTTTCAGAATTTGTTGTTGGACTTCGTTGGAGAGATCTCTCTGTTATTAGAAAGGAAGATA
TTTACAGTTGTTGCATCCCAGCTGGCCTCTCCTGATCTGTTGTCATTGCAGATGTTAGTATCAAATTTGAGACTG
AGTTCGCCGGTCCAGAACGAGGTATTCGGAGATTCATCAGAATTTCAAAATAGTTCTCTGCATCTTCGAAAGTAC
ATCACAGGTGCCCAAGCATTAGAAGATGCTTATTTGAAAATTCTTAAACAACTGCATTCTGACGGTAATTTGAGC
ATTCTAAATCAACAAGATACAGAGACAACAGGATCTTCTACTCCTGAGTTTGGCTTTGGTAGATTTCTTTTTGAA
GAGGAACAGCGTCAATCATTAGTAATTCAAGTTCATAGCATCATAAAGAACTCTGCAGTGGTCAATGAGGAGCTT
ATTCAGCTACTCACCAAATGGTTGGTATCTGCGAAGTCAAATTCCAAGTTTTTAGACACCGATAGACTGTTGCAG
CTTCTAAAAGAAAACAGTGCTCTTCCCGAAATTGAGTCTTTACTGCAGTTGGAATCTTCATTGCATCTGAAATCA
TCATGGCTGGTAGGTTCAGATTCTTGGGCATATGATCTGGGAAACTCGGGTATTCATAATGTCCTGTCTTCCCAT
AAAAATCTCAATATGCTGATAATTGATTCAGAACCTTATGGCAAAGTAAATGCTAATGGTCTTAAAAAGAAAGAT
ATTGGATTGTACGCGATGAACTTTGGAGATGTATATGTTGCATCAGTTTCAGTTTATTCTTCATACACACAAGTA
TTGCAGGCATTTCTTGAAGCAGAAAAACATGATGGTCCTTCTGTGGTGTTAGCATATCTCCCCTACCATTCGGAA
ACTGATAGTGCATTAGATGTGTTAAAAGAATCCAAAATAGCGGTAGAGACAGGTGCTTGGCCATTGTATAGATAC
AATCCGGCCGCACAATCTGAAAATGATATCTTCAAATTGGATTCTTCTGTAATAAAGCGAGAGCTGCAAAGCTTT
CTGGATAGAGAGAACAAGCTATCCATCCTAGCCAAGAAAAGTCCCTCTTTTGCTCGTAACTTAGCCTCTCAAAAT
AACAGAGCAAAGGCTCAACAAAGGTTGAAAGCCAAGGCAGCGTATGATGAGCTATTGGAAAGTTTGACCGGTGCA
TCTCTGACTGTTGCATTTGCATCTGATGGTGGAAATGCAGAATCTCTCGCTAAAAGACTAGCTAGGAGGGGCGCT
GGTCGTGGTCTTAAGACCACCGTATTATCTATGGACGATTTGTCATTGGAAGATTTTGCTCTCGAAGAGAATATT
GTTTTTATCACTTCTACTTCGGGACAAGGTGAGTTTCCCCAGAATGGTAAAGCATTCTGGGATGGGATAAAAAAT
TCTGCAGACCTAGACTTGGCCAGCGTAAAGGTATCAGTTTTTGGTTTAGGTGATTCTCTTTATTGGCCTAGAAAG
CAAGATGCGCATTATTATAATAAACCCGCCACTGATTTATGGAAGAGACTACTGTTTTTGGGAGCCCAACAATTG
GCTCCTCTAGGTTTGGGTGATGATCAGGCTTCTGATGGGTTTCAAACGGGTTATGAAGAATGGGAACCTGCTCTT
TGGATTTCCTTGGGGGTAGATGGCGCGCCCACCTTAGATGAACCTCCTCCAATAACTAACGAAGACATGAAACGT
GAATCAGATTTCTTGCGAGGTACTATTGCTGCAGGACTACAGGACACATCCACTGGCGCTATATCTGCAAGCGAT
CAACAGATGACCAAGTTCCATGGTATATATATGCAAGATGACCGCGATGTTAGAGATGAACGGAAGGCCCAAGGT
TTGGAGCCTGCTTATGCTTTCATGGCTCGGGTTAGGCTGCCAGGTGGTAATGTTACTCCAAGACAGTGGATCAAA
TTGGACGAACTGGCTGACGTGAGAGGTAATGGCACCATGAAAATTACCACGAGGGCTACATTTCAGCTGCATGGT
TTGGTCAAACACGATCTTAAAGCCGCAATTCGTGGAATGAACTCAGTATTAATGGATACCCTTGCAGCATGTGGT
GATGTCAATAGAAATGTCATGTGCTCTGCCTTACCAGGAAATGCGAAGGTGCACGCTCAAGTTGCCAAGGCCGCT
TCAGATATCTCAGAGCACCTCCTTCCTCAAACTACCGCATATTACGAAATATGGCTGGAGGGAGAAGACGAAGGT
GACTCCAGTCCCAAAGATAGATTGGTATGGGAAACTCGTAAGGAAGGGCCTAAGAAGAAGAAGGTTATGGTTGCT
GGTAACTCTCTAGTGGATGTGGAGCCATTGTATGGAAATGATTCATTGTACCTGCCCCGAAAGTTCAAGATAGTT
ATTACGGTTCCCCCATATAATGATGTCGACGTGTATGCTCATGATGTTGGTCTAATCGCAATTGTGGACGAAAGC
GATGTGGTGATTGGTTTTAACGTCTTGGCTGGGGGTGGTATGGGCACAACGCACAACAACAAGAAGACATATCCT
CGTACAGGCTCCATGATGGGATATTGTGACTACGAAGATGTGAATACCGTTTGTGAGAAGATCATGCTAGTCCAA
CGGGATAACGGGGATCGAACAAATAGAAAGCACGCAAGGCTAAAATATACTGTTGATGACATGGGGGTAGACGTT
TTCCGTTCTGCAGTGGAAGAACTGTGGGGTAAAAAATTTCAAGCTGCAAAGCCATTTGAGATAAAATCTAATATT
GACTATTTTGGATGGGTTACAGATGAGATCGGTTTACATCACTATACTTGTTTCATTGAAAACGGACGAATTGAA
GATACTCCAGATTCTCCTCATAAGACTGGTCTTCTAGAACTTGCTAAATACATGATGGAAAAGGGGGTTGGAGAA
TTCCGTTTGACTGGTAATCAACATATGCTTATCTCGAACATCACCGATAACCACCTCCAGAGAGTTAAACAATTT
CTGGTTGACTATAAGTTGGACAACACAAACTACTCCGCTTTGCGACTCTCATCTGCAGCTTGTGTCGCCTTCCCG
ACTTGCGGTTTAGCTATGGCTGAATCTGAACGTTATCTTCCTGTTCTAATTAGCAAGATTGAAGAATCTTTGGAG
GAGTATGGTTTGAGACATGATTCTGTGGTGATGAGAATGACTGGATGCCCTAATGGTTGTGCTCGACCTTGGCTA
GCCGAAGTGGCACTAGTTGGAAAAGCATATGGTGCATATAACCTATTATTAGGAGGTGGGTACCATGGACAACGT
TTGAACAAGCTTTACCGCTCGTCTATCAAAGAGGACGAGATTCTGGATATACTGAGGCCATTATTCAAACGTTGG

GCCCTGGAAAGAAACGAAGATGAGCATTTCGGAGACTTTTTGATTCGTGTAGGCATCATCAATCCTACAACTGAA
GGTAAGTATTTCCACGACGACGTTCCTGAAGAAGCCATTGCATAA

Downstream (in bold start codon previous gene)
acattatactacttatttattatttattgtttccctgagtcatttatttgctgtgggatt
cgtgtaattcaacaatcctctttccgcttcagtgacctgtaacaccttagctactgagat
cctctgctctctttcaagtgccagcacctgtttaatctttacctcccaaagctttttcag
aaccctatctagacgatgattgagtttaatatcgacttctcttaacattgactcttttg
aagctgaagtttcctagttctctctgccactacgccctgttagtaatgttacagttgaca
aaaccatttgatttctggatagtgacaataaacatttgtagtgcaaacttaccttgagct
tccattttccgtataaccaactctctttgatgtttagtgtacctcatgcttccaagggtt
gaatttgttgtgttgaaaggtacagttgagggcatcaaaaacgatacatccatacatcag
gttattccctcgagaagtctggctatctacttgccacccttaagaccccaatatctctca
aatcacatccgacg**atg**gtattggaagctacaatgataattctggacaattcagagttta
tgcgtaatggagactatcttccatcgcgtttttctgctcaattagactcggttgatttta
tttttcatgcaaagaccaattctaacccagaaaatactgtcggcctcatgtcgatgggtg
gtgatggacctcaagtactcacaactctcacagccgattttggtcgtatattggcgggta
tacatgatacaaagatttccaaaggtatacatttcagtacaggcattcaagttgctctct
tagccctgaaacacagacagaacaaggtccaccaccagcgcatcatagtttttgtgggct
cacccatcgatgaggatgaaaaagaactagaaaagctagctaagcgactgaagaaaaata
gtgttgccattgaccttattaactttggagaacatgaagta

AA
MTGSVGEIVNVLTHVPNRGHVYTTKLSKGLSNNFSKALDDASIDYQSLVSNNDPFALVIETVNKNSFTTVLTDAE
TLLFSVPFLSEVRSGDKLIVHARLPKNDYSIVSALKDLDFVVLLSSSVQETQDIALAGYKFLTQNNKPVLHFFNE
DYSEPIVDQLSPEVVSSYLKSEQEPFAEKLSIYNNFEYSGAKNPLDVVVLLGDSGISPPTHTIDYGIINIRVFRP
FDIESFLNVLPFSTQRLAIVEQSPKKPFAQFQNLLLDFVGEISLLLERKIFTVVASQLASPDLLSLQMLVSNLRL
SSPVQNEVFGDSSEFQNSSLHLRKYITGAQALEDAYLKILKQLHSDGNLSILNQQDTETTGSSTPEFGFGRFLFE
EEQRQSLVIQVHSIIKNSAVVNEELIQLLTKWLVSAKSNSKFLDTDRLLQLLKENSALPEIESLLQLESSLHLKS
SWLVGSDSWAYDLGNSGIHNVLSSHKNLNMLIIDSEPYGKVNANGLKKKDIGLYAMNFGDVYVASVSVYSSYTQV
LQAFLEAEKHDGPSVVLAYLPYHSETDSALDVLKESKIAVETGAWPLYRYNPAAQSENDIFKLDSSVIKRELQSF
LDRENKLSILAKKSPSFARNLASQNNRAKAQQRLKAKAAYDELLESLTGASLTVAFASDGGNAESLAKRLARRGA
GRGLKTTVLSMDDLSLEDFALEENIVFITSTSGQGEFPQNGKAFWDGIKNSADLDLASVKVSVFGLGDSLYWPRK
QDAHYYNKPATDLWKRLLFLGAQQLAPLGLGDDQASDGFQTGYEEWEPALWISLGVDGAPTLDEPPPITNEDMKR
ESDFLRGTIAAGLQDTSTGAISASDQQMTKFHGIYMQDDRDVRDERKAQGLEPAYAFMARVRLPGGNVTPRQWIK
LDELADVRGNGTMKITTRATFQLHGLVKHDLKAAIRGMNSVLMDTLAACGDVNRNVMCSALPGNAKVHAQVAKAA
SDISEHLLPQTTAYYEIWLEGEDEGDSSPKDRLVWETRKEGPKKKKVMVAGNSLVDVEPLYGNDSLYLPRKFKIV
ITVPPYNDVDVYAHDVGLIAIVDESDVVIGFNVLAGGGMGTTHNNKKTYPRTGSMMGYCDYEDVNTVCEKIMLVQ
RDNGDRTNRKHARLKYTVDDMGVDVFRSAVEELWGKKFQAAKPFEIKSNIDYFGWVTDEIGLHHYTCFIENGRIE
DTPDSPHKTGLLELAKYMMEKGVGEFRLTGNQHMLISNITDNHLQRVKQFLVDYKLDNTNYSALRLSSAACVAFP
TCGLAMAESERYLPVLISKIEESLEEYGLRHDSVVMRMTGCPNGCARPWLAEVALVGKAYGAYNLLLGGGYHGQR
LNKLYRSSIKEDEILDILRPLFKRWALERNEDEHFGDFLIRVGIINPTTEGKYFHDDVPEEAIA

6 *P. pastoris* genes involved in glycosylation

6.1 Nucleotide sugar synthesis and transport

SEQ0129
UDP-GlcNAc transporter
YEA4 (YEL004W) Uridine diphosphate-N-acetylglucosamine (UDP-GlcNAc) transporter required for cell wall chitin synthesis; localized to the ER
S. cerevisiae null mutant: viable; increased budding index; increased transposable element transposition
S. cerevisiae overexpression mutant: increased UDP-D-glucose excretion

Chr1-3_0163

5' region

gtttcaagggaaatctgagttttgtggagactcccaactgccgttctaattgttcatatt
gctgtaatttttctttgttgtgatcaatttgttttccttgcaataccggtacaatttttt
ccgagtttcgaaggtaatggtcgtttcccagctgagactgagctttattgagcctcctta
tcgtctcatctaaactgctctttgtgatggtgatactaagtattatgtcatccaaatctt
gcatcaaggacttgtcattggacacgacaagcttatgatttttgtgtttatcacaggttt
ttgacaatttgcgggtcattgcgatgatatcctgctgtctttctgtgctggcaacgatct
cgtcattgatgtctgaagaaaagttctcaaaaacagtgcccaagttgttgtttagttgtg
atagctcttcactgaatgaaggaaggatctgggtaagagaaaaatccatctttggtagca
gtaaatcttgttttcaggtcactggtcaatggattgagctggcccagtggacaagagaga
acatctcaagtaaattgacagcccgcgcgcagtgcccccgttaaacaaagggcagtttta
ttacgtaagcagaattggccctttttcacgttaataaagcaaaatgccccccttgcttat
cgtcctactagtcacacgacaacaataaaacgggtatattacgtaactacattgatccaa
gttctagatgacgatgatcagaattcaagttttaagacaccggtgacagtattcaaaaat
aaaacaaaatccttacatacctctcttgccagccttccaaacactacccctttaaactccc
ctacccctttgataatttttttcaaatttttccagcaatcaccgtgttccgaattaatacc
agatagcgttatctattgtcggtaaagttgaagctcgaaataccgtcgaactatatttat
ttttctcttcccttccctgttcgaacccaaaacacgtcaag

ORF

ATGATCGATATATCCCCGTTTTTGATGGTGTTCACCGGATGTTGCGGCAACGTCTTCTCCTTGGAGACCTTGGTG
ACTAAGACGACCTACAGTGTTGGTACAATGGTAACATTTTGCCAGTTCTTGTACGTAGCTCTAGTCTCGTATCTT
TTCGTGATGCTCCCAGGAGAAGGAGGCAAAAAGACAGATGATGTGCTGTTGGCAGGCCCCAAAAAGGATTGGTAC
AGCTCTTGGCTGCCAGAATCTAAGGTGCCTATGAAACGCTACTACGTGAACATTTTTTTGTTTTTCGTGACAAAT
GTCTTGAACAACTACGTTTTCGTGTTCAACATTGGCATACCCGTCCACGTCGTGTTCAGAAGTAGTAGTGTCACC
GTGACCATGTTGATTGGATATTGCTTTTTGGGCAAAACCTACAACGCAAAGCAGATTTGGGGTTCTGTGGTGCTA
TCAATTGGTGTCATTATAACAATGATTGACAACCAAATTGCCCAAAATGAAGCAAAAGGTGTTGAATACACGGGA
AATTTCATTGATTGGTCTGCTATAGACCTACATTACATGTTTGGTATATTTGTGATGACCCTGGCCAACGTTTTA
ACCTGTGTTATGGCCCTTTATACGGAGACTACTTATAAGAAGTACGGAAAGTCAACATGGATGGTGAATCTCTTC
TACCAACATGCATTTGCATTGCCATTATTTCTTTTTGTTTCCTTGAATCTGAAAAGAGAGTTTGAACACTTCACG
TCCAACCATGGATCTTTGCTCATCAACTGCCTGACGCAATTCCTCTGTGTTGCAGGTGTTAATAAAACGGCATCT
CTTTACAATGCTGTAACCCTATCCATCATTCTAATGATCCGAAAGTTGGTGTCTTTGCTGATTAGTTGTTACTTC
TTTGACAACTCCTTCAACGCAATGGGATATTTTGGAATTTCAATGGTGGTGGTCGGTACCGCTCTGTACACTATG
GGAGGCCGAAGCATTTCACCTCCACTTGCAGCAAATTCCGAGAAGCTTGATCTTGAGAGTAAGGGAAACAATCAG
GATTGA

Downstream

tatcccatcctcttttcataatatctcaaggagcctactttagcataatggtcatttttc
atagatagtagacatgtgtttaatcatcgctcaaaaagtgcgtaggagcatttaattcat
ctttctaagtaactaacccaatctttttatggatctaaaatgctcttttcttcaatgtgc
cgtaaccgttgacgcagaccatacacttacataatcatcacgacgctttcgtaccgctcg
catctcgatctcattacctgaaatagctaagcatcgatttccttctcttcccgcaccaac
tccgtgcgctctgtatactccgtcaacactgctctaattggtctctgaatgctggtatca
tggcctacttcacacttccggaagagtacctcccagacctgcaagacgccgtcaatttct
tcatttcatttgccccatttctcagctatggcaccacaatttacggcatctcccaaagcc
agtcaagtacaggcttctcgctagatatttgtgccaccatgttgctcagcgcaagtttgc
gagtgatgtactacttcaacgagaagtatgaattcaccctgttgaggcagtgcatcgtca
tgcttatcatccagatattgctcttgagaacggccttgaagtatagatcaaacacctatt
catgtgacaatcttgtacaaatgcccaattacaacgtgatatttgtgcataaaatgcagg
aattagtggcccaattatcagagagcctgatcaacgactctgaggggataccagagcgcc
tggatctgatctggaaggcatgggtcagtacttggcgcctctttttgttctcatcattt
atggcaactttgccaacttgttccttctttttgacgtcagctacgcccgcccttaaact
tttggcagtgggaggaacaagagacctattggaagttcttaggttgtttttggactacaa
tagtttccttagaattgttattatacaagaatgaacagtat

AA

MIDISPFLMVFTGCCGNVFSLETLVTKTTYSVGTMVTFCQFLYVALVSYLFVMLPGEGGKKTDDVLLAGPKKDWY
SSWLPESKVPMKRYYVNIFLFFVTNVLNNYVFVFNIGIPVHVVFRSSSVTVTMLIGYCFLGKTYNAKQIWGSVVL
SIGVIITMIDNQIAQNEAKGVEYTGNFIDWSAIDLHYMFGIFVMTLANVLTCVMALYTETTYKKYGKSTWMVNLF

YQHAFALPLFLFVSLNLKREFEHFTSNHGSLLINCLTQFLCVAGVNKTASLYNAVTLSIILMIRKLVSLLISCYF
FDNSFNAMGYFGISMVVVGTALYTMGGRSISPPLAANSEKLDLESKGNNQD

UDP-Gal
Synthesis:
SEQ0130
GAL10 (YBR019C): UDP-glucose-4-epimerase, catalyzes the interconversion of UDP-galactose and UDP-D-glucose in galactose metabolism; also catalyzes the conversion of alpha-D-glucose or alpha-D-galactose to their beta-anomers
Note: S. cer. protein has 2 enzyme activities (in 2 domains); P. pas. has only one of the 2 domains and so only 1 of the 2 activities (activity in light grey is missing)
S. cerevisiae null mutant: viable; decreased competitive fitness; decreased utilization of carbon source

Chr4_0839

5' region
attgtcagtctgtgagctatatctgtagccctcgtgtaccatgggtaagctttagacatt
ttaaaaaggtctgaatatcggttaggtaatttggaaatcagatggaacagattttagtct
tcagataaaggttggtacactacaaagtgaagtacttagttcgttaatgggctattcttg
gattgagattttgaaagtttgtgttcaattgctaacgactcgtcagcagatggaatgtaa
aactcttcatatctgcttagaatcaagactacagtagttttgtttttagattggcttctc
gtagaaacgagtttttctttaggtttcgaaggtgaaatcctagttacatctaaacggtct
ttcaaagcccgtggttactcaaaaacaagcataagggaaggtattcaagaccaaaaatgc
aggacttagcgttgctgaaccagtctttcaactgccccgtttactataagcaaagtgagg
caagttaagacgtcaatcaacaaattcctttgaaacttcctatatcactttccttcactt
cgtcgtggctccaagaggttaaagttcacaggccatagctccgaatcgatctagaacata
tcgtaagatcattaccatcgctctattaggggtagagtactacaagtatgttcaatgcga
ctccaatatatattaggcctctgctaacgagggttcaactgagtaaaacattcaacgaaa
gccgttaacaaagccagtatttgctagtaggggtttgaaagttaaagcattgaatagatt
aattagatagttgggtgccatagattccagctaattgcaccacctggcaggatctatccc
gttaaacgaggtgttcctctctgctaaaattgcgaaaattgctaaagaaaacaccctgca
tcaaggggttgaatgtcggttaattagacttggatgcacctccactttacataacttac
taatctgaaattaatccttctttattctgctttataattcg
ORF
ATGAATGGATATGTGCTAGTCACAGGTGGAGCGGGCTACATTGGATCCCATACTGTCGTTGAGTTGCTAAACAAT
GACTACCTTGTGATAGTGATTGATAATTTGTCAAACTCTTCTTATCATGTTATAAAGAGAATCGAAACTCTGACT
GGCAAATCAGTGACCTTCTTTAATATCGACCTTAGAGAAGAGAGTAAGCTCAGAAAAGTGTTCCAAGACTACACC
ATTGATTCAGTATTACACTTTGCTGCATTGAAAGCGGTTGGAGAATCTGCAAAGTCGCCACTTCAGTATTACGAT
AACAATGTCTCGGGAACAATTTCTCTTTTGAACGTGATGAAGCAAAATAATGTGAAAAAGCTTGTGTATTCGTCC
AGCGCTACAGTGTATGGTGATGCCACTAGATATGAGGACATGATTCCAATTCCTGAAATTTGCCCTGTTGTGCCC
ACCAATCCATATGGAATGTCCAAGGTTATTGTGGAGAACATTATACGGGACATATTCGCATCAGACAATGGTTGG
AAATCGGCTATCCTCAGATATTTCAACCCAATTGGAGCACATCCTTCTGGTTTGATTGGGGAGGATCCTTTTGGT
ACTCCAAACAATTTACTTCCCTATTTGGCACAGGTTGCTATCGGAAGAAGGGACAAACTATTTGTTTTTGGGAAT
GATTACACTTCAAAGGACGGTACACCGATCAGAGATTACCTTCACATTGTCGACCTGGCCCAAGGACATATTGCC
GCTCTCAAATACTTAGGTACCCGACAGGAGGGCATATGCAGAGAATGGAACTTGGGAACAGGGAATGGGTCCACT
GTTCTTGAAGTTTATGAGGCCTTTTGCAGCGCTGTAGGTAGAAAGCTGCCCTACGAAATTGTTGGAAGACGAGAG
GGTGATGTGTTGAATCTAACAGCGTTGCCCACGAGGGCCAATGAAGAGTTGGGTTGGAGGACTAAACTCACGATT
GAAGATGCTTGCATAGATCTTTGGAATTGGACTACGAACAACCCCAACGGTTACAGGATAGAAGAGCAAAAATCT
GACTGGTTGGGGAAAAAAGATTCCTTCTCTAAAAAACTATGTCGAGCAGGAGGAGTTAACACCTCTGAAGGGCCT
ATTTTGCCAGCCCGAATATTGGATGAATTTATGGAGATAGAAGTGAAAACAACGAATTAA
Downstream
tgttattcttgaagatgtcaagatacaaaaatactttatttagattagaagaaaaagtgc
ttgaagcatgctaaaataattcgaagctagtccatttctacattaactgaatgtaattgt
tgtcttgagatgctacaagtaatatagttaaaagtgacttgatgttgcttttgctttaaa
tcttctcatggcacctttatttgactgagtcaagttacaggcttgagggcggttgaccag
atcatcttcgcaactagaagtgggatccgtcaattactagaaaattccagttggtgaaga
ttcaccaaccagccaacgatcttgaaaactcttttaaaaaagagccttaccaggcaagaa

gttgctatgcttaatggttatactataaagtaactaactatagctagcatcaatgatcta
cgcagtaaaacatctctacaccatactacaccaccatgacgaattgatcaattttacgcc
atacgtgcaatacacgacgaatgcctaatccagcactctgaggcttgctctcgctagtat
gatgacattacttcaattgctagttctggtatacataatatataaagtgtttccgatact
cagaagatttttcagtactggaacaatggtgtccgaacaggttttaagcaagactaaagc
attgatcaaagatcacaaggtatttgttgcatccaagagttattgcccctactgctccca
gactaagaagttgttggagtctctaaacgccaatgcttttgtagtggaactagatactga
accagatgggtctgatattcaagccgctcttttagagctaaccggccagagaactgtacc
taacgtgttcatcaacggagagcatgttggtggaaacagcgatttgcaagcactgaacag
tgaaggaaagttaaaaactcttctcaagaactaaagcctacttcaaaacgctattatggc
ctcaaagattaatcttcggctctactgattagagaaggata

AA

MNGYVLVTGGAGYIGSHTVVELLNNDYLVIVIDNLSNSSYHVIKRIETLTGKSVTFFNIDLREESKLRKVFQDYT
IDSVLHFAALKAVGESAKSPLQYYDNNVSGTISLLNVMKQNNVKKLVYSSSATVYGDATRYEDMIPIPEICPVVP
TNPYGMSKVIVENIIRDIFASDNGWKSAILRYFNPIGAHPSGLIGEDPFGTPNNLLPYLAQVAIGRRDKLFVFGN
DYTSKDGTPIRDYLHIVDLAQGHIAALKYLGTRQEGICREWNLGTGNGSTVLEVYEAFCSAVGRKLPYEIVGRRE
GDVLNLTALPTRANEELGWRTKLTIEDACIDLWNWTTNNPNGYRIEEQKSDWLGKKDSFSKKLCRAGGVNTSEGP
ILPARILDEFMEIEVKTTN

Transport:

SEQ0131

HUT1 (YPL244C): Protein with a role in UDP-galactose transport to the Golgi lumen, has similarity to human UDP-galactose transporter UGTrel1, exhibits a genetic interaction with *S. cerevisiae* ERO1

S. cerevisiae null mutant: viable

S. cerevisiae overexpression mutant: increased UDP-D-glucose excretion

Chr2-1_0692

5' region

ccaagaagtactgaagagtctgcaatcgttgatcgagcgcttgggtcagctgacccctca
tgaacttcaaattgtggataaactgtgcctagacgtcatagaaggaagagaaatatccaa
aaaggaagttgaccagccgtggtacaaagaactgatagctgttgtggaatattccaacag
atgatgatgaaaacgtagtgggaagttgaaccaagttaaggagattgtttttctacgtta
ttggagttcccagacgatttgctgtggtgtgaataacagagcaaaatgcttctcatagcc
aagttttctatcttcgttgaagacgagcttgtgatccaaagttaatgtggtaccggcctt
taccactgcagtcttattatgtatgtacaagtatattgatttcatttccgaaacatgtaa
ctttttctaatacctccttcgcctggttaattttcattgctaaatacggactccctcct
ttatctggatgatttatcagcatacattgcctatgtcgtttctttagtttctcatgcgtt
aaagtcacaatttctgactggggaatccccaatataagcagcgcctcactttcagttatg
ggattaaaaaatcctccaggatactgggcaaactttagtttcattacgtcgttttgagtg
tgtggttttgataactgtatcccgttcaattgggctatcatagtgggggttaatttgacg
tacctataataagctctcaacccacttctaaatgtcaaagctgccacggtaagtcccaag
ccagcgattattgggagaaccatagctggagtcggagatggctggtaataatattatgaa
gaaagtaaggtagtatcacttggtctcaacctctactctaagaatgcaaacaatcgatat
agatagttcgcgagaaccttattattttcaggtaacgcactggacaggtacacttcatca
ggtatttcacgggtttctctccaaatttaagcttgcttctc

ORF

ATGAAAACATTGACTCTTTGTGTGGCAGGAATTTATGCATCGTTTATTCTCTGGGGCTATTTGCAAGAACGCCTG
TCGGAGCCCTACAATGGGCGGCCCTTCAAAGCTCCACTGATCGTTAATGCTATTCAATCGTTTTTCGCCATGATT
GTAGGCTCCATATATCTCAGTGCTAAAACTAGAAAACTCAGCACTCCAATGGACCCAATTTCGCAAAATCCCAAA
ATAATTTACCAACTGGCCCTAATTGCGGTGCTTTCTTCGGGATCGTCTGCAGTTGGGATGAAATCCTTGGAGAAT
GTCGACTACTTGACATATCTTCTTGCTAAATCATGCAAATTGATCCCTGTTATGATCGTCTCCGTCTTGGTATAC
AGAAAACGATTTCCTGTGCACAAATATTGTATTGCTTTGTGTATTAGTGCTGGCGTTATCTTGTTCACATTAAAG
CCCAAGTCGCTGAGTAATTCTATTGATTCCTCCGCTGGAAACTGGAGAGGATATCTTTGTTTGTTGATATCTTTA
TTTTTAGACGGTTTACTCAACTCCTCTCAAGATCAGCTTTTCAAGACTTTCAAAATTTCAGGAGCACAGTTGATG
GCTGCACTTAACATGTTAACATTTATCCTCATTTCATCGTACATTGTGCTATTCACAGATCAGTTGCCATATTTT
GTTTCCTTTGTACAAGTGTCTCCTCAATTGCTTCAGAATGCAATATTGTATGGTATTGCAGGAGCCATCGGACAG
ATTTTTATTTTCTTAACTCTAGAGAAATTTGACTCTATTGTCCTCACCACGGTCACGGTTACGAGAAAAATGTTC

AGCATGGTACTGAGTGTGCTTCTATTTGGACACTCATTATCCTTACAACAGCAGGTTGGAATTGGACTAGTCTTT
GGCGGTATTTGCTACGAAAGCTGGCTAAAAAGCTTCGGCTCCCAAGTCAGCCCAAAAGCAAAACTGTCGTAG

Downstream

agagaaacttagatttttaatagattagagaattacataatttacatcaaatagcattaa
ttctggtgggatgacagttgcaacaaagatttagcagctcccatggatgatggagtaaaa
atgtccctggtggttgggctggagaaaagcttgttctctctctttagcttcacgggagta
caatacctgttgaaagatgatgattcaggcaattgaatagtggtgtcattacaaatgatg
gttgaattagagttggaattgcaggcgggcaactgtgacttctttgagatcggggtggaa
gaatttgcccgttctagactatcgttatgcatatctaagtcattgtggatgctaatttta
cagttggggttccttttgatctggtcgaatgttggtggtggtagatgagccgtaatcttt
gatattggaggcaatgtcacattggctgagttctgggactgtatgcgaggtagagcagat
ttcaaagagcttctcttggtgccagcaatgtccgctagttttcttcttttgaaggatta
gacatctcatccttaagggaatcatctctcttaaagccaacttgactactgttgatcagg
tcgtcgacattcagtgataattggttggtttgaattttgaaaactgccagctgtagtcgg
gttttcaatctgtttatcttgtcattgacatcactagactcgtcttcaagttgtaacgct
tttatgtgactcttgtctggaaaggtgagaccgttgcttttgagctgctgatcaggatgg
tcagcacccaatctatgaatgtttagctgtgtctctgatatagttcccagtactgctcta
cccattagttatgctatgttttgtcccttagtttactagaactgttgttattgtctactt
cgtcagctatgtatgtattatttttttgtcaacctcaatatttacgttttactcttttcga
ttctacgcggtacatgcaatgcgatgtctggcgcgtccagc

AA

MKTLTLCVAGIYASFILWGYLQERLSEPYNGRPFKAPLIVNAIQSFFAMIVGSIYLSAKTRKLSTPMDPISQNPK
IIYQLALIAVLSSGSSAVGMKSLENVDYLTYLLAKSCKLIPVMIVSVLVYRKRFPVHKYCIALCISAGVILFTLK
PKSLSNSIDSSAGNWRGYLCLLISLFLDGLLNSSQDQLFKTFKISGAQLMAALNMLTFILISSYIVLFTDQLPYF
VSFVQVSPQLLQNAILYGIAGAIGQIFIFLTLEKFDSIVLTTVTVTRKMFSMVLSVLLFGHSLSLQQQVGIGLVF
GGICYESWLKSFGSQVSPKAKLS

SEQ0132

putative UDP-galactose transporter [Pichia stipitis CBS 6054]

chr4_0810

5' region

ggatcttgatgtgtttataaaaacgctgtttgttattgtggctgcttgtgacacttattt
tagcaacaaaggctatgggatttgaggggatccctgaagtccgaccattcaaaagggact
cttgaaatgagccagtcatcgcaagagagcttcttcttctaaaatttatgtcattattgg
tattgtgtgacaatgagtctagtttgcgctggcacttgtagcatagtcttttctgtgtgg
atgaaacaacgggattgttcaatcccggttcagcaattgtctgagacaacgacgatcggc
gtctcggttggaacatgagaaagctctgtggaacaggcatttgggcaatgacgctgaaag
atgagtatgaggcaggatttcagtcaatcagctatgatttttcgattgtaggctgtattg
agttaagaagaacgtatgataaggggcagatgatatcatattatagcgcgaaatgactaa
catcataatcgaccccagaatcactcaagtgatcatttatggttatgtaattatgaaatt
gtaatgggaaattgtcaggcaaaagtatatagcccaccgaataggtacgtacactgcata
catgggtcaaagttatcttgggccgaacaactctctgaaaagtttaatcttcaagatgaa
agtcagattgctcgattagaattgggcagacattgaaatagcctatcactctaatgaaac
tggctaattttgccaaacgaaaactcaccataaacactgccgactgacgatcaagttcag
tagcacagataatgctcccaactaacattctagaaatacaatctcattaatcagactcat
gtttcttaaatgataataatctatttatatagtattgttcttccgataccagtccttaga
gcgcgcgcccctctttaatcagctgcctctcctcctgctgaatctgaaacgttatctctc
tttcccaccagatcattcgaccatagcaagcctgttattac

ORF (in bold intron)

ATGGTCAACATTAGA**gtatgtttcgttctatcgacccctcatcattcctttcattcaattgtctaactttag**AT
TTCTCTGCTAAGTGCAGCGGTTATTGTAACAGGTTGTCTGAACTCTGTGTTCACCAAATATCAAGATAATCAATG
TGTTAGAAACTGTCAGGACCCAGATTTCTCGAAGCACGTCTACTTTGAACAGCCGGCTTTGCAAACTCTGCAGAT
GTTTGTCGGAGAATCTGGTTGCTGGGTAGTGGTCTTCTTATCTTGGCTTTCAACAGTTGTCAAATACTATTCTAA
GGGGAGAAAAAAACCACTTTACGCTCCACTGTCAGAAGGCGACTCAGATCAGACAAAGCTACCTGAGCTGACTGG
TCTTAGAAGCTTGATTTTGGCAATTCCTGCTGTTTGTGACGTGTGTGGAACGACAATGCTGAATGTAGGCTTACT
CTACACCCCTGTGAGCATTTACCAGATGGCTAGAGGGTCATTGATTTTATTTGTCGCCCTATTCTCTGTGTTTTT
TTTAAACCGAAGAATATCCAGACTTGAATGGTTGTCCCTTTTTGTGGTCACGTTTGGTGTATTCTTAGTGGGTCT
CAGTGGAGCACTTGGAAATAAAGTTCACATATCAGGGGACAAAAACACTTCATTTGCAGTCATTTTTGGTATCAC
TTTGATCCTCATTGCTGAGATTTTTGCAGCATCGCAGTTTGTGTCTGAAGAGTACATCATCTCAAGATCTTCTGT

CGATCCGTTGAGATTAGTTGGCTTCGAAGGTATTTTTGGAGGTTCTATTACGTTCCTGGTCATGGTCATAGCATA
CAGTGTTGTTGGAAGATCCAATCCTGGTGGACCATTTGATATCGTCACTTCTTTCAAGGATGTTACGGCTAACTC
ACAAGTTATCTGGTCAAGCGTGGCAATCATGATATCCATTGGAAGTTTCAATTTTTTTGGCATTTCTGTTACCGA
CGAGTTAAGTGCTACAGCCAGAAGTACTATCGACACGTGCAGAACATTACTCGTATGGCTAGTCTCCCTTTATCT
AGGTTGGGAAAGCTTCAAATGGTTACAATTAGTTGGATTTTCGCTTTTAGTGTTAGGAACACTAACCTTTAATGG
TGTTCTGAAGGCTGAGAACTGGTACTTTATTCCTGAATGGTTCAAAAAGGATGCTCCTCTACCTGGAGAAAGGGT
AATTGACGTTGTGGACGAACAAATCGAAAGATTCTAG

Downstream

tatagaattagtcattataagaagatcattgaattcatccttgtagcgagtttcccgtcc
agaactcctccgcccatctcttccctcccatcaagcattcgcgccgatggacagcgtaac
tcaacatcactcgaagactcaaagttcaaccagagacacaacaatggacgaacataggct
cttcaagttcccagtatttagacttcccaaatccgctacagttagagcaatggggatgta
cctttccggggctctctatgcgattggtttttgggttcttttagacgctgccctttactc
caaacgggtcaacgcctcattagtccatgtcactttcgttgattgggttccagcaatatg
ttcatctctaggtatgctgattattaactcgattgaaaagtctgcattattgaacaatga
tgcatcattcatgcaaccatcatctacaacaacctggcaggcccgggtggtgctgtttat
cgggttctcactacttgccacaggaattgccggttcatttattgtacttatactcaagtt
tttgatcaaaggttataatacatttcccaccttgggtatgggggtcgccaatgtggtatc
taatgctagcatcatgttaagttgtatagtgttatgggtggtacagaactttgaggatga
cgactacaattactcattggccctgtgaatttagtttgtattaggtagaatggaaggcaa
aaattgtgcactaatagacttctcatagtccgtaaagttaaatagagtgtcgctgatttc
aagtgcctgtgatgcattggtgggatacactgtgaaaccaggagcgcgcagttgatcgaa
gaagcggtgggccattatttgtgagcctgaaaacctggaggcaggctcgtaaaccaaaat
atgcatcagtaggtctatcgcatcttgtggtgaattattgaacaccttgttgaaacgggt
cgaaactttaggactgaacagtttggcggagtacgtagggc

AA

MVNIRISLLSAAVIVTGCLNSVFTKYQDNQCVRNCQDPDFSKHVYFEQPALQTLQMFVGESGCWVVVFLSWLSTV
VKYYSKGRKKPLYAPLSEGDSDQTKLPELTGLRSLILAIPAVCDVCGTTMLNVGLLYTPVSIYQMARGSLILFVA
LFSVFFLNRRISRLEWLSLFVVTFGVFLVGLSGALGNKVHISGDKNTSFAVIFGITLILIAEIFAASQFVSEEYI
ISRSSVDPLRLVGFEGIFGGSITFLVMVIAYSVVGRSNPGGPFDIVTSFKDVTANSQVIWSSVAIMISIGSFNFF
GISVTDELSATARSTIDTCRTLLVWLVSLYLGWESFKWLQLVGFSLLVLGTLTFNGVLKAENWYFIPEWFKKDAP
LPGERVIDVVDEQIERF

6.2 O-glycosylation

PMT1 (YDL095W): Protein O-mannosyltransferase, transfers mannose residues from dolichyl phosphate-D-mannose to protein serine/threonine residues; acts in a complex with Pmt2p, can instead interact with Pmt3p in some conditions; target for new antifungals

S. cerevisiae null mutant: viable; decreased protein O-mannosylation; decreased resitance to calcium dichloride and alkaline pH

PMT2 (YAL023C): Protein O-mannosyltransferase, transfers mannose residues from dolichyl phosphate-D-mannose to protein serine/threonine residues; acts in a complex with Pmt1p, can instead interact with Pmt5p in some conditions; target for new antifungals

S. cerevisiae null mutant: viable; decreased resistance to alkaline pH; abnormal budding pattern; abnormal cell shape; decreased cell size; decreased competitive fitness; decreased metal resistance

PMT3 (YOR321W):Protein O-mannosyltransferase, transfers mannose residues from dolichyl phosphate-D-mannose to protein serine/threonine residues; acts in a complex with Pmt5p, can instead interact with Pmt1p in some conditions; target for new antifungals

S. cerevisiae null mutant:viable

PMT4 (YJR143C):Protein O-mannosyltransferase, transfers mannose residues from dolichyl phosphate-D-mannose to protein serine/threonine residues; appears to form homodimers in vivo and does not complex with other Pmt proteins; target for new antifungals
S. cerevisiae overexpression mutant: decreased vegetative growth

PMT5 (YDL093W):Protein O-mannosyltransferase, transfers mannose residues from dolichyl phosphate-D-mannose to protein serine/threonine residues; acts in a complex with Pmt3p, can instead interact with Pmt2p in some conditions; target for new antifungals
S. cerevisiae null mutant: viable
S. cerevisiae overexpression mutant: abnormal cell cycle progression

PMT6 (YGR199W):Protein O-mannosyltransferase, transfers mannose from dolichyl phosphate-D-mannose to protein serine/threonine residues of secretory proteins; reaction is essential for cell wall rigidity; member of a family of mannosyltransferases
S. cerevisiae null mutant: viable

5 genes with homology

SEQ0133
chr2-1_0212 (most likely PMT1 or PMT5)
5' region
acgaacaccggaaaatcaaccctaatacgatccatcgttggtaaggattactcaaagaag
cagacagagaatggcccgggtgtctctcaccttccttcattcacaagaaaacccatgaag
ttcaaaatggacaacaacagtcttgaactcgtagatctccctggatacactgctccaaat
ggaggtgtttacaagtatcttaaggaagagaactaccgagacattttgaacgttaaacag
ttaaagccattgacatccctcaaggcatacacagaaacgttgccttcgaagccaaaacta
ttcaatggtgtgcgagtaatatgcattggtggtttagtgtacattcggcccccaaagggt
gtagtgctgaaacagtttagtctcgtcaaccttccatccttcatgtactcgtcgctaaaa
aaggccaccagtgtaatccaagcgcccccacaagccttggtgaattgcagcgtcgtcaag
gaggacagtccagatgaactggtaagatatgtgatccctccattttatggtttaattgac
ctggtcattcaaggtgttggatttatcaagcttctgcccactggagctcggaacaccaga
gaactgatagaaatttttgccccaaaagacatccagctcatggtgcgtgattccatcctc
aaatacgtctacaagacccatgccgaacacgactcaaccaataatctcctgcataaaaag
aacataaaagccagaggccaaaccatactacgaagactacccaaaaagcctgtattcaca
aagctttttcccgtaccagccaacgtaccgtctcatgaactgctcaccatggtgacggga
aaggacgacctagccgaggaagacaaagaataccgctacgatatccagtatcccaacaga
tactgggatgaaaccatctgtaaatagaatgcttatgtaatcaagcactttctgaaattc
cttagagtttcgcgtgtctccccgtcaaaaatcgcgtctc
ORF
ATGTGCCAGATATTTCTCCCGCAAAACGTAACACGTTGTTCTGTTTCCCTTTTGACAATGAGTAAAACAAGTCCT
CAAGAGGTGCCAGAAAACACTACTGAGCTTAAAATCTCAAAAGGAGAGCTCCGTCCTTTTATTGTGACCTCTCCA
TCTCCTCAATTGAGCAAGTCTCGTTCTGTGACTTCAACCAAGGAGAAGCTGATATTGGCTAGTTTGTTCATATTT
GCAATGGTCATCAGGTTCCACAACGTCGCCCACCCTGACAGCGTTGTGTTTGATGAAGTTCACTTTGGGGGGTTT
GCCAGAAAGTACATTTTGGGAACCTTTTTCATGGATGTTCATCCGCCATTGGCCAAGCTATTATTTGCTGGTGTT
GGCAGTCTTGGTGGATACGATGGAGAGTTTGAGTTCAAGAAAATTGGTGACGAATTCCCAGAGAATGTTCCTTAT
GTGCTCATGAGATATCTTCCCTCTGGTATGGGAGTTGGAACATGTATTATGTTGTATTTGACTCTGAGAGCTTCT
GGTTGTCAACCAATAGTCTGTGCTCTGACAACCGCTCTTTTGATCATTGAGAATGCTAATGTTACAATCTCCAGA
TTCATTTTGCTGGATTCGCCAATGCTGTTTTTTATTGCTTCAACAGTTTACTCTTTCAAGAAATTTCAAATTCAG
GAACCGTTTACCTTCCAATGGTACAAGACCCTTATTGCTACTGGTGTTTCTTTAGGGTTAGCAGCTTCCAGTAAA
TGGGTTGGTTTGTTCACCGTTGCCTGGATTGGATTGATAACAATTTGGGACTTATGGTTCATCATTGGTGATTTG
ACTGTTTCTGTAAAGAAAATTTTCGGCCATTTTATCACCAGAGCTGTAGCTTTCTTAGTCGTCCCCACTCTGATC
TACCTCACTTTCTTTGCCATCCATTTGCAAGTCTTAACCAAGGAAGGTGATGGTGGTGCTTTCATGTCTTCCGTC
TTCAGATCGACCTTAGAAGGTAATGCTGTTCCAAAACAGTCGCTGGCCAACGTTGGTTTGGGCTCTTTAGTCACT
ATCCGTCATTTGAACACCAGAGGTGGTTACTTACACTCTCACAATCATCTTTACGAGGGTGGTTCTGGTCAACAG
CAGGTCACCTTGTACCCACACATTGATTCTAATAATCAATGGATTGTACAGGATTACAACGCGACTGAGGAGCCA
ACTGAATTTGTTCCATTGAAAGACGGTGTCAAAATCAGATTAAACCACAAATTGACTTCCCGAAGATTGCACTCT
CATAACCTCAGACCTCCTGTGACTGAACAAGATTGGCAAAATGAGGTATCTGCTTATGGACATGAGGGCTTTGGC

GGTGATGCCAATGATGACTTTGTTGTGGAGATTGCCAAGGATCTTTCAACTACTGAAGAAGCTAAGGAAAACGTT
AGGGCCATTCAAACTGTTTTTAGATTGAGACATGCGATGACTGGTTGTTACTTGTTCTCCCACGAAGTCAAGCTT
CCCAAGTGGGCATATGAGCAACAAGAGGTTACTTGTGCTACTCAAGGTATCAAACCACTATCTTACTGGTACGTT
GAGACCAACGAAAACCCATTCTTGGATAAAGAGGTTGATGAAATAGTTAGCTATCCTGTTCCGACTTTCTTTCAA
AAGGTTGCCGAGCTACACGCCAGAATGTGGAAGATCAACAAGGGCTTAACTGATCATCATGTCTATGAATCCAGT
CCAGATTCTTGGCCCTTCCTGCTCAGAGGTATAAGCTACTGGTCAAAAAATCACTCACAAATTTATTTCATAGGT
AATGCTGTCACTTGGTGGACAGTCACCGCAAGTATTGCTTTGTTCTCTGTCTTTTTGGTTTTCTCTATTCTGAGA
TGGCAAAGAGGTTTTGGGTTCAGCGTTGACCCAACTGTGTTCAACTTCAATGTTCAAATGCTTCATTACATCCTA
GGATGGGTACTGCATTACTTGCCATCTTTCCTTATGGCCCGTCAGCTATTTTTGCACCACTATCTACCATCATTG
TACTTTGGTATATTGGCTCTCGGACATGTGTTTGAGATTATTCACTCTTATGTCTTCAAAAACAAACAGGTTGTG
TCTTACTCCATATTCGTTCTCTTTTTTGCCGTTGCGCTTTCTTTCTTCCAAAGATATTCTCCATTGATCTATGCA
GGACGATGGACCAAGGACCAATGCAACGAATCCAAGATACTCAAGTGGGACTTTGACTGTAACACCTTCCCCAGT
CACACATCTCAGTATGAAATATGGGCATCCCCTGTACAAACTTCCACTCCTAAAGAAGGAACCCACTCAGAATCT
ACCGTCGGAGAACCTGACGTTGAGAAGCTGGGAGAGACAGTCTAA

Downstream

gctgtgtttatatagccctgtacgtaaaatctatgacacaagtttatggttatttgtctt
atgtaagcaatatttggattgatgtctcgagaccatcaactccatcactgataagttgat
cggatttgtatttctgtcccctatttactaattccctttccagaaatagatcatgaatga
ggcagaatataagtgccaaagatgccggctgccgttgaccatagacggatctctggaaga
ccttagcatatcacaggccaatcttttgacgggacgaaatgggaactttacaaagaacac
aatccccttggaggatgccgtggaagaagatttacccaaggtgcctcagagccgacttaa
cctctttaaagaggtctaccagaagatggatcacgattttaccaatgccagagatgaatt
tgttgtgttgaacaagcacaatgataacagcgacgtcaatgtggagtatgattacgaaga
aaacaacactatcagtcgtagaatcaacacaatgacgaatatcttcaatatcctcagcaa
caagtacgaaattgattttccggtttgctacgaatgcgccacattgctgatggaggaatt
gaagaatgagtacgaaagggtcaatgctgataaagaagtttacgcaaagtttctatccaa
gcttcgcaaacaggacgcaggtacaaatatgaaagaaagaactgctcaactactggagca
attggagaaaactaagcaagaagagagagataaagaaaagaagctccaaggcctatatga
tgaaagagatagtttggaaaaggtattagcttctttagagaatgaaatggaacagttgaa
tattgaagagcagcaaatttttgaattagagaacaaatatgaatatgagttaatggagtt
caagaatgagcaaagcagaatggaagcaatgtatgaggatggtttgacgcaattagataa
tttaagaaaagtgaacgtctttaatgacgctttcaatatc

AA

MCQIFLPQNVTRCSVSLLTMSKTSPQEVPENTTELKISKGELRPFIVTSPSPQLSKSRSVTSTKEKLILASLFIF
AMVIRFHNVAHPDSVVFDEVHFGGFARKYILGTFFMDVHPPLAKLLFAGVGSLGGYDGEFEFKKIGDEFPENVPY
VLMRYLPSGMGVGTCIMLYLTLRASGCQPIVCALTTALLIIENANVTISRFILLDSPMLFFIASTVYSFKKFQIQ
EPFTFQWYKTLIATGVSLGLAASSKWVGLFTVAWIGLITIWDLWFIIGDLTVSVKKIFGHFITRAVAFLVVPTLI
YLTFFAIHLQVLTKEGDGGAFMSSVFRSTLEGNAVPKQSLANVGLGSLVTIRHLNTRGGYLHSHNHLYEGGSGQQ
QVTLYPHIDSNNQWIVQDYNATEEPTEFVPLKDGVKIRLNHKLTSRRLHSHNLRPPVTEQDWQNEVSAYGHEGFG
GDANDDFVVEIAKDLSTTEEAKENVRAIQTVFRLRHAMTGCYLFSHEVKLPKWAYEQQEVTCATQGIKPLSYWYV
ETNENPFLDKEVDEIVSYPVPTFFQKVAELHARMWKINKGLTDHHVYESSPDSWPFLLRGISYWSKNHSQIYFIG
NAVTWWTVTASIALFSVFLVFSILRWQRGFGFSVDPTVFNFNVQMLHYILGWVLHYLPSFLMARQLFLHHYLPSL
YFGILALGHVFEIIHSYVFKNKQVVSYSIFVLFFAVALSFFQRYSPLIYAGRWTKDQCNESKILKWDFDCNTFPS
HTSQYEIWASPVQTSTPKEGTHSESTVGEPDVEKLGETV

SEQ0134

chr2-1_0256 (PMT2 or 3)

5' region

gaagcattcttgtcgtcctttcgaaccaatcttgaaccaaatattgccatgtttgggagt
aaaactgggatcaaagaaggtaaagaatttaatgggggaaattctaatggtactttgtac
tagctgcaaatatcgattgatgagttgttctgagtttccctggataaaatcgaagatgtc
gaatatttatatgtatgctggtatgtgttagagtacggctatcccgcaattcgtggtttt
ttttttctttgatagctgccctggctcggcggcccgaatggttcttaggatctaggat
gctggttgttctctccccgccaaggcgagcgattcggtgttccatggctttcaaatgata
aggcgaacaatcggagaaagccacgctattgctagtgcacaagaggagaccgactgataa
ttctctacccacttgaatggcatgggtggtggattaatgtgaggtgattatttgaatgag
ttatatttttttactatatagttgctttttatagatgtgccacatatacttatgaaagtcg
agagtttgatccgcattttccaaagcttaacattttgagaacacgattgattggcagttt
ggttctaatttcaagataataatatgcccgggtcgaaattaaatatctaattacaaacag
gccctttaaacatcacccgtattttggttttgtaaaagaatgtcaggtgcaaaataaaga

gctataacttatttagaagtacaaaagtacctaaccttgcaaaccctggtctatatacta
actactatatactaactagtcccactttaacctttcaatcgttcaaatattatccatccc
cagtatcatcctgaatgaaatttcacccgtggcatgcatggcacaaaaacgtcgtgctcc
acccaatactcatctcatgttcccccgtcaaatttccctttaacgtgttctttcttcaaa
cacggattcaagtcctttcgattcttatttttatcttcg

ORF

ATGACAGGCCGTGTCGACCAGAAATCTGATCAGAAGGTGAAGGAATTGATCGAAAAGATCGACTCCGAATCCACT
TCCAGAGTTTTTCAGGAAGAACCAGTCACTTCGATCTTGACACGTTACGAACCCTATGTCGCCCCAATTATATTC
ACGTTGTTGTCCTTTTTCACTCGTATGTACAAAATTGGGATCAACAACCACGTCGTTTGGGATGAAGCTCACTTC
GGAAAGTTTGGCTCCTACTATCTCAGACACGAGTTCTACCACGATGTCCACCCTCCGTTGGGTAAGATGTTGGTC
GGTCTATCTGGCTACATTGCCGGTTACAATGGCTCCTGGGATTTCCCCTCCGGTCAAGAGTACCCTGACTATATT
GATTACGTTAAAATGAGGTTATTCAATGCCACCTTCAGTGCCTTATGTGTGCCATTCGCCTATTTCACCATGAAG
GAGATTGGATTTGATATCAAGACAACTTGGCTATTCACACTGATGGTCTTGTGTGAAACAAGTTATTGTACGTTA
GGAAAATTCATCTTGCTGGATTCAATGCTGCTGCTATTCACTGTGACTACGGTTTTCACCTTTGTTAGGTTCCAT
AACGAAAACAGTAAACCAGGAAACTCGTTTTCTCGCAAATGGTGGAAATGGCTTCTGCTTACTGGTATTTCCATT
GGTCTCACTTGTTCCGTCAAAATGGTGGGTTTATTTGTCACAGTATTAGTTGGAATTTACACAGTTGTTGACTTA
TGGAATAAATTTGGTGATCAATCCATTTCTCGTAAGAAATATGCTGCTCATTGGCTAGCTCGTTTCATCGGCTTG
ATTGCCATCCCAATTGGCGTTTTTCTATTGTCATTCCGTATCCATTTTGAAATATTATCCAATTCTGGTACCGGT
GATGCAAACATGTCTTCATTGTTCCAAGCTAACCTTCGTGGATCATCCGTCGGAGGAGGCCCCAGAGATGTGACC
ACTCTCAACTCTAAAGTGACCATAAAGAGCCAAGGTTTAGGATCTGGTCTGTTACATTCCCACGTTCAAACTTAT
CCTCAAGGTTCCAGCCAACAACAGATTACAACCTATTCTCACAAAGATGCCAACAATGATTGGGTGTTTCAACTT
ACGAGAGAAGACTCTCGAAACGCTTTCAAGGAAGCCCACTATGTCGTTGATGGTATGTCTGTTCGTCTCGTTCAT
TCAAACACTGGTAGAAACTTACACACTCACCAAGTTGCTGCTCCCGTCTCCTCATCCGAATGGGAAGTCAGTTGT
TATGGTAATGAAACCATTGGAGACCCGAAAGATAATTGGATTGTTGAAATTGTCGACCAGTATGGTGATGAAGAT
AAGCTGAGATTGCACCCATTGACCTCCAGTTTCCGTTTGAAATCGGCAACTCTGGGATGCTATTTGGGTACTTCG
GGTGCTTCACTGCCTCAATGGGGTTTCAGACAAGGTGAAGTTGTTTGTTACAAAAATCCGTTCCGTAGAGATAAG
CGCACCTGGTGGAACATCGAGGACCATAACAATCCTGATCTACCTAATCCTCCAGAAAATTTTGTTCTTCCCAGG
ACTCATTTTTTGAAAGACTTTGTTCAATTAAATTTAGCAATGATGGCAACAAACAACGCTTTGGTCCCAGACCCA
GATAAGGAAGATAATCTAGCTTCTTCTGCCTGGGAATGGCCCACGCTACACGTTGGTATCCGTCTGTGCGGTTGG
GGCGATGACAACGTCAAGTATTTCTTGATTGGTTCTCCCGCAACCACCTGGACTTCTTCAGTTGGTATTGTAGTA
TTCCTGTTCCTGCTGTTAATTTACTTGATCAAATGGCAACGTCAATATGTCATTTTCCCATCCGTCCAGACTCCA
CTAGAGTCAGCCGACACCAAAACAGTTGCATTGTTTGACAAGTCTGATAGCTTCAACGTCTTCCTTATGGGAGGA
TTATACCCGCTTCTGGGATGGGGTTTACATTTTGCTCCGTTTGTGATCATGTCGCGTGTTACCTACGTTCACCAT
TATCTTCCTGCATTGTACTTTGCCATGATTGTTTTCTGCTACTTGGTTTCTCTGTTGGATAAGAAACTAGGCCAC
CCAGCATTAGGATTACTGATCTATGTGGCTCTGTATTCCTTGGTCATTGGAACATTTATTTGGCTCAGCCCCGTT
GTGTTTGGTATGGACGGTCCGAACAGAAATTACAGTTACCTAAACCTTCTACCTAGTTGGAGAGTATCAGACCCA
TAG

Downstream

gtgtatatatcgtcagtacctaaatttatgatagggtaaaaccgacatcttctatctaca
attaatgcgcgactcacgctttctcttatctgcttggtcttgtttgacttcatcgttagc
gttcgcttctctcttcaggtaaaaccttaccctccatgtcgacctatagagaactggagc
gcgaggagggcgacttctctgataagaagcttctcacaggctggttctccttaaccgtgg
gacaactacaatatgctgcattcatggccgtaggaatcgcgttgctttggccttggaatt
gctttctatcagcttcagacttctttggagagcggttgcaagaacacaagtggctctctg
ctaactattcatcgtccatgatgaccatttcaacgttgacctcaacgttatgcaacgtgt
ttttgtcccaaaagcaaagtggggtagattattcaaagagactcgtcatgggccaaacaa
tcaccatagtcgtatttgcctttatgggtctgctgtgcgtatggaatacggggctagatc
caataatatttttttgtgttggtcatgattaatgttgcactgagttcagtagctgtgtcgt
tatcacaggtcggtgctatggcgatcgtgaatgtgttgggaccaatatatgctaatgctg
tggttgtggggaatgctgttgctggtgtgctaccgtccatcgctctgattattagcactg
ccctgtcgggaactcatgtggctggaaagttgcaacctaaaagagattatgcagttatgg
catacttttttgactgcctgtgtcgttagtggtattgcattagttctttttgggttggcag
agtcacatggcccaatcgacgttgttgctgcacctgttcatacttcttctactgctgatg
aggctattgaagaactaggtatccctttagaagaggaggagtatgttccctttcaactc
tgtgggccaaacttcgttttgttgcactcacaatatttac

AA

MTGRVDQKSDQKVKELIEKIDSESTSRVFQEEPVTSILTRYEPYVAPIIFTLLSFFTRMYKIGINNHVVWDEAHF
GKFGSYYLRHEFYHDVHPPLGKMLVGLSGYIAGYNGSWDFPSGQEYPDYIDYVKMRLFNATFSALCVPFAYFTMK
EIGFDIKTTWLFTLMVLCETSYCTLGKFILLDSMLLLFTVTTVFTFVRFHNENSKPGNSFSRKWWKWLLLTGISI
GLICSVKMVGLFVTVLVGIYTVVDLWNKFGDQSISRKKYAAHWLARFIGLIAIPIGVFLLSFRIHFEILSNSGTG
DANMSSLFQANLRGSSVGGGPRDVTTLNSKVTIKSQGLGSGLLHSHVQTYPQGSSQQQITTYSHKDANNDWVFQL

TREDSRNAFKEAHYVVDGMSVRLVHSNTGRNLHTHQVAAPVSSSEWEVSCYGNETIGDPKDNWIVEIVDQYGDED
KLRLHPLTSSFRLKSATLGCYLGTSGASLPQWGFRQGEVVCYKNPFRRDKRTWWNIEDHNNPDLPNPPENFVLPR
THFLKDFVQLNLAMMATNNALVPDPDKEDNLASSAWEWPTLHVGIRLCGWGDDNVKYFLIGSPATTWTSSVGIVV
FLFLLLIYLIKWQRQYVIFPSVQTPLESADTKTVALFDKSDSFNVFLMGGLYPLLGWGLHFAPFVIMSRVTYVHH
YLPALYFAMIVFCYLVSLLDKKLGHPALGLLIYVALYSLVIGTFIWLSPVVFGMDGPNRNYSYLNLLPSWRVSDP

SEQ0135

chr4_0777 (PMT6)

5' region

cacttgattatctaggaaaggcattatcaaagacgcagcacccccacccgctctacattg
ttctctttcgtaggaaccgacaggatcaaatgagtagactgctcctttaccttcttcgtc
cagtcccgcaatgatagtatgcacatagtaaggaaagaatcgttttccatatagcaaatg
ttgtatgaatcttgccgcagatttaatcgtcaagttctttttatgatcaaaatgatacca
ttttagctgattgcggaatctcttgaccaatgcgtctccatctgctgcgaatccattcgc
agtgatcagtaattcgtctcctaccgagaacacttttggctcatatctggaattgattga
atagcctgtggtatgtctggtatctcctgctaaaattgcaaagtcttctccggcaattcc
caaaactgtaccaccgttgtcttgaaattgatgttagtaactattgccattgaaatcatg
ttagcagtcttaagaccgaattataatacctaccagagtaaggattaaactggtgttcta
tggggacattggatacgcctgaattataatcagacgacactgcacttgtaacactcatta
ggaatgtagtcggtgatggtagtggcaaacctctctcaattgatatagagcgagatatct
tgaaatattgagtggtacacgggatatgtttaatttctaatgatcgttaagtgattcgaa
aaggtgctattccagaacaattggagtgaaaaatggtacctataaatgtatttgcctcat
ctattagttatagagattgtctgaagattaattttctgccattcttttggcttgtgtgga
aaacatatctgtctggctttaaaatgtggaaaaaaattacataatcaggtgagcgagaag
gggtctagcatcgagtatacctcgtcttgcataacgaacaagtggcctaccagactcata
cctgtatcttgcattgacaaattgttctctacaatccgag

ORF

ATGGCAACAGAGGAAGAGAGAAATGAACTGAGAAGTCGGATGGACGCCAATAATTCAAAAGTTTCCACGTTCACT
ACGAACAATTCAGATGATCCTTCTGTTGATAGCCAGGGTAAGGTGAAAATTAAGTCATGGGTTTGGAGCCTTGAA
TCTTTAATTGGCCCTCTGGTGATCACTGCCTTGGCAATTTTTCTTCGAGTTTACCAAATAGGAAAAGCTGATAGG
GTTGTTTGGGATGAAGCTCATTTCGGAAAGTTTGGGTCATTCTACTTGAAGCACCAGTTCTATTTTGATGTCCAT
CCTCCCCTGGGAAAACTTCTTACAGGTTTGGCTGGGCACATAGCTGGCTATAATGGGTCGTTTGAGTTCAAGAGT
GGGGTGACGTATCCAGAATATCTCGATTTCAAGGTAATGAGGATATTCAATGCTGTTTTCAGTGCACTTTGTGCC
CCTGTGGCATATTGGACTGCCAAATCATGTGGATATTCTCTACTCACGGTTTACCTTATATCATTGATGGTAGTT
TTTGAAAACTCCTACGTTGTTTTGGGGAAGTTTATTCTGTTGGATTCCATGCTTTTGTTTTTCACCACAACAACC
TTTCTGGGTTTATCAAAAGTTCATTCATTGAGACAGCAAGGAAAAGAATTAACTTACCCGTGGTGCTTCTGGTTG
ACCTTTACAGGTTTATCTATTGGATGTGTATGCAGTGTCAAACTTGTTGGACTGTTCGTAACTGCCCTTGTGGGT
CTTTATACCATTCTTGACCTTGCCGTTAAGAGATACGATGAAAACCTTAAATGGTCTAAGTACTTGACTCATTGG
GCAGTGCGCATTCTAACGTTGATTATTCTACCGTTTGCCATCTACATGTTATCCTTTAAGATCCATTTTGCTGTG
CTTTACAAGAATGGAGATGGTGCTTCTTCAATGTCAACTCTGTTCCAATCCAATTTGGAGGGAACAAAAATTCTA
ATTGATGCCCCTAGAGATGTTGCATATGGATCTGAACTTACAATAAGATCCCAAGGCCTTTCCCAAAATCTTTTG
CACTCCCATGGGTCAATTTATCCCGAAGGATCTAACCAACAACAAGTTACAACCTACGGTCATAGAGACAATAAT
AACCAATGGATTGTTCATTACCCCGTCCTCAGCAAGAAGCAAGTTAAAGAAAATGATAACTCAACAGTTCCAGAG
ATGATGAAAGATGGTGACACCATTCGTTTAAGACATCAACATACTGGTGCAAACTTGCATTCTCACAGAATTCAA
GCTCATGTTAGTAAACAATACTACGAAGTATCATGTTACGGAAATGCAAAAGTTTCAGACGGAAACGATGAGTGG
GTTGTGGAAGTTGCAGAGCAAATTCATTCCGATGACCCTAAATATGCCGCTGCCAACGAATCAGATTTAAAATTC
CAGGAACTGCTTCACCCTATATCTACTTCCTTCAGACTTCGTCATAAGAGAATTGGATGTTACTTGGCCACTACT
GGGATGGCATATCCAAGCTGGGGTTTCAAGCAGGGTGAAGTAGTCTGCCGACCATCTTGGACCTCAAGGGACAAA
TCTACGTGGTGGAACATCGAGGACCATAAGAACAAAAAGCTGCCAAATGCTACTTCATATAAAGCTCCAAAATCT
TACTTCTGGAGAGACTTTGTCATGCTGAACTATGCGATGTTAGCATCCAATAACGCCTTAGTGCCTGATCCTGAT
AAATTCGATAAACTGGCTTCTCAATGGTGGCAATGGCCTATCATTAATGTTGGACTCCGAATGTGTGGTTGGAGC
GCCTCACAATCTAGATACTTCTTGATGAGTAGCCCATTTAATACTTGGTTGTCAACTGCTTCTTTGGCTGTTTTT
TGCTTGATTGTTCTTATTTTAGTCCTACAATGGCAAAGACAAAGACTCAATCTCTCTTCTAGACAGTACTGGGAG
CTAGTCATCAAGGGTTTTGTCCCATTTTTTGGTTGGGCGTTGCATTTTGCTCCATTCATTGTGATGCAAAGGGTC
ACATACGTGCACCATTATGTTCCCGCTCTCTACTTTGCCATGTTCCTGCTGGGATTTACTGTAGACTATTTGACA
GCCAAGAGGAACTGCTACATCAAAACATTGATCTATTTTGTCTTTTATGCGGGCACTATCTACTCTTTCTACTAC
TTTTCACCTCTCAGTTTTGGCATGGATGGCCCTTTAAAAAATTATGCGTACTTGCAATGGTTCAAGAGTTGGACC
ATGGTTTGA

Downstream

ggaagaaggtgtagggcaaatattagacgatctatacagacaaaatataaatacgagttg

acgtatcactgctgccgactcccatcagtaatttttatcttgtaagtttcgactaatgg
atattctgttgactctatcaccaggaacgatcttgtaatccagctcattgatcagtacaa
attcattgttgttgttgaaagagaacacttttatactaccagactctgtgccaagtgcaa
ttacagcttttccatcaatagcagctttgaacgaatccactgccgtaacttgttcatcca
acttgagagaagcgagctgaatcgtatctttttttgatttcgtcgtaactccaagccttga
tgtgtttatctctagaaacggtgaagaacgtcgaatctgacaattctgaaggtgcccaac
aacaatcccaaataatcctcgagtgcgccttttcttgcagcttaagaagctcaaactgtt
gcaagtcacttccagttctttttccataaggaaaatcgtctatccctggacacggcgagca
agaattgattgtttggagaaaatttcaatctggtaatggtcaaatcgtgaccttcaagta
cctggggcaattgctgccacgttctggtgtcaaaaactctgactacagcatgtgcaggtg
tatttgacctacaagaagaagctaccaattgtttgtcgggagatgaatccactgccgtga
tttcataaccatgtccgtaaagcttctcaatttcgggccacaaagtatgtctttgcaggt
gatcttctaatggaggctgatcaagagcggaaacgacttcatatgaaatattcttctcgg
attggtcttcctcctcttcgtcattgttgatattttcgtgtgtaacggcagactcgttct
ctgctttgttagaaaggcctagaacaggcaacgaagcactctcaggcatttcatcaatat
gagtttgtctcatgctacagaatttttctaatagctttgca

AA

MATEEERNELRSRMDANNSKVSTFTTNNSDDPSVDSQGKVKIKSWVWSLESLIGPLVITALAIFLRVYQIGKADR
VVWDEAHFGKFGSFYLKHQFYFDVHPPLGKLLTGLAGHIAGYNGSFEFKSGVTYPEYLDFKVMRIFNAVFSALCA
PVAYWTAKSCGYSLLTVYLISLMVVFENSYVVLGKFILLDSMLLFFTTTTFLGLSKVHSLRQQGKELTYPWCFWL
TFTGLSIGCVCSVKLVGLFVTALVGLYTILDLAVKRYDENLKWSKYLTHWAVRILTLIILPFAIYMLSFKIHFAV
LYKNGDGASSMSTLFQSNLEGTKILIDAPRDVAYGSELTIRSQGLSQNLLHSHGSIYPEGSNQQQVTTYGHRDNN
NQWIVHYPVLSKKQVKENDNSTVPEMMKDGDTIRLRHQHTGANLHSHRIQAHVSKQYYEVSCYGNAKVSDGNDEW
VVEVAEQIHSDDPKYAAANESDLKFQELLHPISTSFRLRHKRIGCYLATTGMAYPSWGFKQGEVVCRPSWTSRDK
STWWNIEDHKNKKLPNATSYKAPKSYFWRDFVMLNYAMLASNNALVPDPDKFDKLASQWWQWPIINVGLRMCGWS
ASQSRYFLMSSPFNTWLSTASLAVFCLIVLILVLQWQRQRLNLSSRQYWELVIKGFVPFFGWALHFAPFIVMQRV
TYVHHYVPALYFAMFLLGFTVDYLTAKRNCYIKTLIYFVFYAGTIYSFYYFSPLSFGMDGPLKNYAYLQWFKSWT
MV

SEQ0136

chr1-1_0286 (PMT1 or 5)

5' region

aattcatttatatccaaattgctgaggcgctgttcaaagaatttaacaacgtttattgca
ccccaatcgaaatttggatttttctccattgataagtttaaggcaatgttctcaactacgg
atactggacctgagactagtttcagaaactatgaatcttgaagaattattcacctccctg
cagagactccctgtgttggaacaactgtcgtttcctcggtcttcggtctcttttgaacct
acgaagcacatttggcccaagaaactttggtaccttagattaagtggcggtatttccgat
gagtttcttatgagatccaaccttccaaagtcagtcacttatctcgagttttcccattgt
ccttcgatacacgcagatggaatttatcatgtactggacaaagttggagctaacttgacg
catctgtatattcattacccaatgcctgggttgcgaacgaattcgatggatatgatcttc
cagtattgccccaatttacagtttctctacattcacattgattacattacaagagaagtg
tttgacgaagaattgcttcctgctttggaggcaccaaggcctttgaagactctctggtta
gactcttctggcatgttgggtcaatcgtttaaaatctatcccgacgatttaaccattgct
ctctctgagggaaggttgccgtgtctgaagaccctaagaatatcccgtaaattaggatgg
aacttcaacaacgatgactgtcaggatttggtaaacgaactgcaagaccaggatggagat
gtgtatgctgtctgagataaatgtgagggatgaaactccccgagattactatattacgtg
tactggattttgagatcacctttttttaacgaattatgacattcttcttattagactgcct
agttttgtataatcttacagaaattctagctcaagccctcttacttgttcttcttctatg
tcaactgattcctcaatatatgtggttggtggcccgcgaa

ORF

ATGACTCCTGAGATATTTGGTCAAACCTACCAAAGGACACCACACCACAGTACTATAGCACAACAATACATGGCC
GCCTTTGAGTACAAAAAGGGCATTCAAAGACCCTATTTTTTTACCAAGCCATTGGTGAAACCTATAACGCTAAGC
GGCTTTGAAAAAATACAATTGGCTTTGTTTCTTGCGTTCACAGTGGCCGTGAGATTCTTCAATATTCAATACCCC
AACCAAATTGTATTTGATGAGGTCCATTTTGGAAAATATGCCCGAAACTACATCAATAGCTCATACTTCATGGAT
GTGCACCCTCCTTTAGTCAAGATGCTTTACGCCGCCATAGGCTATTTAGGTGGTTACAGAGGAGATTTTGTTTTC
AACAAGATTGGGGATAACTACATTGGTAAAGAGGGTGAAAAATTGGTACCCTACGTTTTGATGCGATCGTTTCCC
GCAATTTGTGGAGTCTTGATTGTTATTCTTTCTTACTTTATCCTTAGATACAGCGGATGCCGACATTTTATTGCA
CTTTTTGGAGCTTTACTGGTTTGTATTGAAAACTCATTGGTAGCTCAATCAAGATTTATTCTACTAGATTCTCCA
TTGCTTTTATTCATTGTTCTCACAGTATACAGTTTTGTGAGATTCAGCAATGAACCAGAACCTTTTGGCAAAGGC
TGGATAAGATATCTATTTTTCACTGGTGTGTCCTTGGGACTCAGTGTCAGTAGTAAATGGGTTGGAATATTCACA

ATTGGTTGGTTAGGAGTCATGACTGTAAACCAATTGTGGTGGTTAATTGGAGACTTAAGCGTTCCCGATCGTGAT
GTGGTAAAGCATGTCTTGTACAGAGCGTATTTTCTTATTATCCTACCAGTGATCATTTACCTTGGGGTGTTTGCA
ATCCATTTTTTGGTTCTCCATGAAGCTAGTGGCGGTTCAGGTACAGTGAGTCCTAGATTCAAAGCCAGTTTGGAC
GGAACTGATTTTTCCAATCTTTATGCTAACGTGTCTTTTGGATCCACCGTTTCGATAAGACACCTTGGTACAGGA
GAGTTTCTACACTCCCACAACCACACATATCCTAAATCGCACAACCAACAGGTAACCCTATACGGATACAAAGAC
TCCAATAATCTTTTCACTATTGAAAAGAAAGATAAGCTATCTGACAAGGAACTATTCGGCGAGGTATCCTTCCTC
CGACACAGAGATGTTATAAGATTATTTCACAAGAAAACCCAAGGATATTTGCACGTCTCTGATTCTAGACCTCCA
ATTAGTGAGCAAGAGTACAACAATGAGGTCAGTATTATAGGAGACAAAGACTATGTCCCCGATGTCAATGAAAAC
TTTGAGGTGAAGATTATCAAAGAGTACAGTGATGAAGATGCAAAGCATGAGGTTAAATCCATCGGAACTGTGTTT
CAATTATTCCATAAGGGTACCAAATGTACTCTGTTTGGTCATCGTGTGAAGCTGCCAAAAGACTGGGGATTTGGT
CAATTGGAGGTCACTTGTATCGAGTCGCCAGTCCTTAAAAATTCTCTGTGGTACATTGAAGAGAATACACACCCA
CTTTTCAACCAAACATATCCTGCAAAAGTGAAAGTCGAACCCTTAGGATTTTTTGGCAAGTTTCTTGAGCTGCAC
CAAAAAATGTGGAAAACAAATGCAGGCTTGACTGCCTCTCACAAGTATAGCTCTAGACCCGAAGATTGGCCCGTT
CTTGACAGAGGTGTGAACTATTTCAACCGATCAGGAAGGACGATCTACTTGTTAGGTAACTTGCCAATCTATTGG
GGAATTGTATTTACTATCGGAGTATTCGTTGTTTTCAAGCTTGTTCAGCTCTGGAAATGGAAGCCAAACCATGCT
CCAACAGTAACCGATGCTTCAGCTAAATATGATTCCCAATTTTTCATCTACTTTGTCGGTTGGCTATTCCATTTC
GCTCCATCTTTTTTGATGGAGCGACAGCTATTTCTGCACCACTACATACCATCTCTATGGTTTGGTATCATATCA
ATCGCTGTGCTCAGTGAATATGTTTGGGCTAAACTGGGAAAAATCGTAGGATTCTTCTACGTTATGACAATATTA
GGGCTTTCGGGTTTCTTCTTCTACTGGTATGCCCCAATCGTTTATGGGTTAGAGTGGAACAAAGACACCTGTCTG
GGTTCGAGACTATTACCAAACTGGGACATCCCTTGCGATCAATTTCAGTAG

Downstream

tatatctggcctgtcgttatctccttttgagttttgactcggcggagtgtgtatcggaga
acattacaggttttagtatctttgtttctttttgttgttgttaagataacggattaaacg
agtactatctgtacggtcaagctttttcgcacttgcggcggactcatgcagtgcgcattt
ccaccgccgactcattgggcgtttttttaatgcctactgaaccatgctagttcctgatga
ctttaccgtagtgcattataaaggtaaacgagattatcgtggtgtacatatcattaaaaa
tataattgctccaatcatggctgaacagaaccccataaggaatgcttaagtccgaaaaag
tataatctgatgtggtactattagtagatggtgttatctgatatggctggcgaggctgca
ttagatagaataagggtcagagtcatatctacagttaattagatcgtctactatcttgct
gcagagccctgaatacattcgctcgatggataggagtatataaaaccttgaggtatccct
cattgggctcaattcagtattcttacacatcacaatcaaatgactgaagaaaaaagagag
gccttcgataccaccgtaaactcggtgtcttctaatgaccaccaaattggaactggaaaa
attgttcttactgaaaaagatgcagacattgctctaggatatgccgccgagggtgttgtg
attgatgaagccaccaacaaaagaattctaaggaagattgatctctgcttgggtcctctg
atgtgtattttgtacgcctgtcagttcatggataaagtctccaactcctacgcttcagtg
atgggtttgagagaagatttgagcatggtaggtgacatgtattcctggacgggaactgca
ttctaccttggatacttggcttttgttttccctgcttcttatatgcttcaaagacttcct
gttatcaagactgttggaactttcattattatttggggtg

AA

MTPEIFGQTYQRTPHHSTIAQQYMAAFEYKKGIQRPYFFTKPLVKPITLSGFEKIQLALFLAFTVAVRFFNIQYP
NQIVFDEVHFGKYARNYINSSYFMDVHPPLVKMLYAAIGYLGGYRGDFVFNKIGDNYIGKEGEKLVPYVLMRSFP
AICGVLIVILSYFILRYSGCRHFIALFGALLVCIENSLVAQSRFILLDSPLLLFIVLTVYSFVRFSNEPEPFGKG
WIRYLFFTGVSLGLSVSSKWVGIFTIGWLGVMTVNQLWWLIGDLSVPDRDVVKHVLYRAYFLIILPVIIYLGVFA
IHFLVLHEASGGSGTVSPRFKASLDGTDFSNLYANVSFGSTVSIRHLGTGEFLHSHNHTYPKSHNQQVTLYGYKD
SNNLFTIEKKDKLSDKELFGEVSFLRHRDVIRLFHKKTQGYLHVSDSRPPISEQEYNNEVSIIGDKDYVPDVNEN
FEVKIIKEYSDEDAKHEVKSIGTVFQLFHKGTKCTLFGHRVKLPKDWGFGQLEVTCIESPVLKNSLWYIEENTHP
LFNQTYPAKVKVEPLGFFGKFLELHQKMWKTNAGLTASHKYSSRPEDWPVLDRGVNYFNRSGRTIYLLGNLPIYW
GIVFTIGVFVVFKLVQLWKWKPNHAPTVTDASAKYDSQFFIYFVGWLFHFAPSFLMERQLFLHHYIPSLWFGIIS
IAVLSEYVWAKLGKIVGFFYVMTILGLSGFFFYWYAPIVYGLEWNKDTCLGSRLLPNWDIPCDQFQ

SEQ0137

chr1-4_0033 (PMT4)

5' region

ccttctctgaaaaaaaatcttgctctccgcgtgcaatagaaactagtcggccctgtacaa
ttaaagcatactccctggttaaagtacctcctccgaacttgctcttgttgatcaaagttt
ctgaccttggggccagtccccagccaccagggccaaacgctttattgaggatacgacgat
acttaatctctggaagataaagtagtccatctggtgtgatttcgacatcttcgttgctaa
ttggttgacataatatgttactactttcattactgaaggagcaaatacctagtccatgga
acgaatccgaccaattgattccatcgccacttgtattagagattggggtgtcgtttaact
gtgaagttccaaacaaaattgataaactgctctcgttcttagcttggccacttttggag

tctcaatagtagcgttttggctctcgtgaattttctgcacagagtcggatgaagaaggtg
caaatgcttctagcattgtagagtcgaccacatagaacctttttaaagagttatgaaaat
aactcttggtagggccaaatacaacccgatatcgtcttagcataagagctgcttctttgg
aatatcgtttcttgtaagtaattacgtgttggctaaacacttagaagtcagtcgcgcatg
cggccaaaaacagactagggatagaagatgaactgacaaaaacatcaagaaggtgaagac
attcattctatgaaaactagtttttatataaaattatggtctgcatttagagagcaatga
tgtaatcaaacatcaataagtgcttgtcgcatcaatatttaataggtaatcatggagtat
tctagtctaccgccttaaaaaaagctcactcgatctagtgcagcttgattgtgtacttca
atagtattccaacgaccttaacatcttaacaccatgtaaatttaagatccacgtatacga
tacaatttctttcaatatcaattctcgttcaagccaactg

ORF

ATGATAAAATCAAGAAAGAGATCGAGAAAAGTTTCTTTGAACACTGAAAAGGAGCTGAAA
AATAGCCATATTTCTCTTGGAGATGAAAGATGGTACACTGTGGGTCTTCTCTTGGTGACA
ATCACAGCTTTCTGTACTCGATTCTATGCTATCAACTATCCAGATGAGGTTGTTTTTGAC
GAAGTTCATTTCGGAAAATTTGCTAGCTACTATCTAGAGCGTACTTATTTTTTTGATCTG
CACCCTCCGTTTGCCAAGCTCCTGATTGCGTTTGTCGGCTTTTTAGCTGGGTACAATGGT
GAGTTCAAGTTTACAACTATTGGTGAATCTTATATCAAAAACGAGGTTCCCTACGTAGTT
TACAGATCATTGAGCGCTGTGCAAGGATCTTTAACGGTGCCAATTGTTTATTTGTGTCTC
AAAGAATGCGGATATACAGTTTTGACTTGTGTTTTTGGTGCATGTATCATATTGTTTGAT
GGGGCCCACGTTGCTGAGACTAGACTAATCTTGCTGGATGCCACGTTGATTTTTTTCGTT
TCATTGTCCATCTATAGCTATATCAAATTCACAAAACAAAGATCAGAACCATTCGGCCAA
AAGTGGTGGAAGTGGCTGTTCTTTACAGGGGTGTCTTTATCTTGCGTCATAAGTACCAAG
TATGTGGGGGTGTTCACCTATCTTACAATAGGCTGTGGTGTCCTGTTTGACTTATGGAGT
TTACTGGATTATAAAAAGGGACATTCCTTGGCATATGTTGGTAAACACTTTGCTGCACGA
TTTTTCCTTCTAATACTGGTCCCTTTCTTGATATATCTCAATTGGTTTTATGTTCATTTC
GCTATTCTAAGCAAGTCTGGCCCAGGAGACAGTTTTATGAGCTCTGAATTCCAGGAGACT
CTCGGAGATTCTCCTCTTGCAGCTTTCGCAAAGGAAGTTCACTTTAACGACATAATCACA
ATAAAGCATAAAGAGACTGATGCCATGTTGCACTCACACTTGGCAAACTACCCCCTCCGT
TACGAGGACGGGAGGGTATCATCTCAAGGTCAACAAGTTACAGCATACTCTGGAGAGGAC
CCAAACAATAATTGGCAGATTATTTCTCCCGAAGGACTTACTGGCGTTGTAACTCAGGGC
GATGTCGTTAGACTGAGACACGTTGGGACAGATGGCTATCTACTGACGCATGATGTTGCG
TCTCCTTTCTATCCAACTAACGAGGAGTTTACTGTAGTGGGACAGGAGAAAGCTACTCAA
CGCTGGAACGAAACACTTTTTAGAATTGATCCCTATGACAAGAAGAAAACCCGTCCTTTG
AAGTCGAAAGCTTCATTTTTCAAACTCATTCATGTTCCTACGGTTGTGGCCATGTGGACT
CATAATGACCAGCTTCTTCCTGATTGGGGTTTCAACCAACAAGAAGTCAATGGTAATAAG
AAGCTTGCTGATGAATCAAACTTATGGGTTGTAGACAATATCGTCGATATTGCAGAGGAC
GATCCAAGGAAACACTACGTTCCAAAGGAAGTGAAAAATTTGCCATTTTTGACCAAGTGG
TTGGAATTACAAAGACTTATGTTTATTCAGAATAACAAGTTGAGCTCAGATCATCCATTT
GCGTCTGACCCTATATCTTGGCCTTTTTCACTTAGTGGGGTTTCATTTTGGACAAACAAC
GAGTCACGCAAACAGATCTATTTTGTCGGAAATATTCCTGGATGGTGGATGGAGGTTGCA
GCATTGGGATCCTTTCTAGGACTCGTGTTTGCAGATCAGTTCACGAGAAGAAGAAACAGT
CTTGTTTTGACCAATAGCGCCAGGTCTCGGTTATACAATAATTTGGGGTTCTTCTTTGTA
GGCTGGTGTTGTCATTACCTACCCTTTTTCCTAATGAGCCGTCAAAAATTTTTGCACCAT
TACTTACCTGCACATTTAATAGCAGCCATGTTCACTGCTGGTTTCTTGGAATTTATTTTT
ACTGACAACAGAACTGAAGAATTCAAGGATCAGAAAACTTCATGTGAACCTAACTCTAAT
TCTTCAAAGCCGAAAGAGCAATTGATTCTGTGGTTAAGTTTCTCGTCCTTTGTCGCTTTG
CTACTAAGCATCATTGTTTGGACTTTCTTCTTTTTTGCTCCTCTAACATATGGTAATACT
GCGCTTTCGGCGGAGGAGGTTCAGCAGCGACAATGGTTAGATATGAAGCTCCAATTCGCC
AAGTAA

Downstream

gagtatacaatgtgtagttcaacgcaaaggaaattctaactttctgtgcaatctggtgac
aatttctaaataactatcacaattggaagaagagattatcccaaatcttatcaaaaaatc
gatgattgccagtgcacaattaggcttgaatttttcttgcagcaacgaagagattacttc
agtgatgttcattagcctgaaatcttcactttcgtggtctatcggattaggaattagacc
ttgtttcatcggcaggtcgtatatgtattccacttctggttgaataaaatcttcgggtgg
tttgtttctgaacatatatgagatggctcccactggactgatatattgcgaaacatagtc
ctcattcaaccctgcctcctcgtaacattctttcaggcaagtttgcaaagtgccattagg
atattccaagcctcctgccacagtattatctaacataccgggaaatgttggtttgtgtct
gcttctcctaggtatccaaagttgaatactgttaggatcggcagaattttgcaaatatcc

attgatatgaactccataagtaacaactcccaaaatattagaaaaagccctttccaccaa
catgtacatcttatggttatcgcagtaaactgcaaaaagctcatttctccaaccgctaag
ggtttcaaagagacgctgatctctccaacgctgagctatctttgcaaacatctgcgttct
tttattttcggtatccagactaggaattatcttgacttcgtgtttttcattatttactat
cacagcctgtgtttcgaactcaaattgttttgccaccttgggaattatataccctagtaa
gatcccatcatgcgataagaatttatacacagatacttcaaattcatgaaaagatggctc
atctttatgaggaacagaatcaacagatctgactagatcaatatatggcattggttgatt
ttattcaatggttatctatctcaaacatgctataaaaata

AA

MIKSRKRSRKVSLNTEKELKNSHISLGDERWYTVGLLLVTITAFCTRFYAINYPDEVVFD
EVHFGKFASYYLERTYFFDLHPPFAKLLIAFVGFLAGYNGEFKFTTIGESYIKNEVPYVV
YRSLSAVQGSLTVPIVYLCLKECGYTVLTCVFGACIILFDGAHVAETRLILLDATLIFFV
SLSIYSYIKFTKQRSEPFGQKWWKWLFFTGVSLSCVISTKYVGVFTYLTIGCGVLFDLWS
LLDYKKGHSLAYVGKHFAARFFLLILVPFLIYLNWFYVHFAILSKSGPGDSFMSSEFQET
LGDSPLAAFAKEVHFNDIITIKHKETDAMLHSHLANYPLRYEDGRVSSQGQQVTAYSGED
PNNNWQIISPEGLTGVVTQGDVVRLRHVGTDGYLLTHDVASPFYPTNEEFTVVGQEKATQ
RWNETLFRIDPYDKKKTRPLKSKASFFKLIHVPTVVAMWTHNDQLLPDWGFNQQEVNGNK
KLADESNLWVVDNIVDIAEDDPRKHYVPKEVKNLPFLTKWLELQRLMFIQNNKLSSDHPF
ASDPISWPFSLSGVSFWTNNESRKQIYFVGNIPGWWMEVAALGSFLGLVFADQFTRRRNS
LVLTNSARSRLYNNLGFFFVGWCCHYLPFFLMSRQKFLHHYLPAHLIAAMFTAGFLEFIF
TDNRTEEFKDQKTSCEPNSNSSKPKEQLILWLSFSSFVALLLSIIVWTFFFFAPLTYGNT
ALSAEEVQQRQWLDMKLQFAK

6.3 Mannosyltransferases

6.3.1 α-MANNOSYLTRANSFERASES

6.3.1.2. α-1,6 mannosyltransferases

SEQ0138

P. pastoris homolog of *S. cerevisiae* **MNN11 (YJL183W)**

Subunit of a Golgi mannosyltransferase complex that also contains Anp1p, Mnn9p, Mnn10p, and Hoc1p, and mediates elongation of the polysaccharide mannan backbone; has homology to Mnn10p

S. cerevisiae null mutant: protein secretion: increased; resistance to Calcofluor White: decreased; chitin deposition: increased; competitive fitness: decreased; freeze-thaw resistance: decreased; metal resistance: decreased; Can1p-GFP distribution: abnormal; resistance to Calcofluor White: decreased; viable

Chr 2-2, 0125

5' region

tctcgtggcatcaactctatttttcttaattcgatccaaaaacactttctgcttctctgt
aaactcgtggataatcttcctgtgtctcaagttgaaaataccgtccccaaataaaggccc
tagaagatatcctaaaactgaagcacccatcattccaatggcagatgccataaatgggtc
aattccaaagattaattcagttggattcatttcgatttgcgaaatatatccccatgaagc
tgaagatgcaaacagagcactgaatatggaggttaccaagttgactcgtcgttggctttt
acgtaaagccaagaattgttcccaatttagcggattatctgagtgtgatttttgtggttg
tgcctgttgttgcatttgaggaacagacaaagcaaatgatctggtcaacatctgattgac
acctaaggatctgacaccgatatttctaattgaaagcattttcggccgtttgctagtttc
aagtaggtggaaatagaagggttgacttgatcgtaatcgtgagaaatttgaccttctctg
aggtagtgcatctactaaattgcagattcaagatgattggggaccatgaagacataaagg
ggttggtggggttgacagaatgataatccaaggatccaaggagacagaggcctaaagcgc
tacccaatatgtagtactacttagttttggccttgaacttgacttggctcatgactaatt

tctgggataaaacaagatcgaaaaattgagaatatagggtcttgaatactagcctagttc
acaaccattaaaaaaccacctatattggatgatccaaattacaataggtgcaccctgatt
gttagttcaagtcgttgaaccactcactatacacaatgtaataagttcgagcttacctga
aacctaattcgttatgagcgcgcgcagatgtgttatctcgaaatctaaatatctttaacg
gcgtcaagtaaatcaagaaacacgaacatctttttctaagc

ORF
ATGTATTTGGGACGGTCGGAGAAAGATGCTTTCAAACCAAAAAAGAATCGTATACTTGTGCCTCGGTCATTAGCG
ATATATTTGGTGATCAGCGTACTTTTGACAAACTTTTTGGTTTTCCAATTGTTCAACATCAACCTGATTCCATTC
CTAAACGGTCGATCTTCCAAGAATAATTTTGATTATCTGGTCAAGAAAGAGTTTTTGAACCTAGAAGAACCATAT
CGTGGTTTCTACAAAGAGGAGATTGAAAATCCTACCAATTATTTTTTCCCTACGATCGAACATGCCTCCCGATTA
AACGAAATGGGATTAGACAACTTGTTTCAGTATACTGTTGACAAGCAAACCAAGGTTCATCAATACATGCTGGAT
CCTGAGTTAAAATATGAAGACCAGCCAAAGGGGAACGAACTAGACTTTGTTCGAAGAGAATTTTTGAGCAATGGC
ATGAAGGTGCATCGTGGTGCTAATTCTCCAGAATTGGTTATTGTTACAGGTATTGATTTTGAGACCTTCGATAGT
AGCTATTTGGGCAATATTACACAGAATAGAATTGATTATGCCCAAAAATACAATTTTGGAGTCTATGTTCGTTGG
ATTCAGGAGTTTGCACCCCAGTTCAATAACTTTCAGCAAAGTAAGGATTGGACTAAGGCTTTACTTATGAGGGCT
GCAATGCTGGCCTTCCCGAATTCAAAATACTTTTGGTACATCGATTCCAACTGTTTCATTATGAACATGGAAACT
AACATCATTCCATACATTCTTGAACCACACGTGCTGGGTCCTATCATTTTAAGAGATCACCCCCTAATACTCCCT
GATGGTGGCATTAAAACCTATGCTAACATAGACCCAAAAGATGTTCATTTGATCTTGACACAGTCAGCTACTTCC
ATTAGAACTGACAGTTTTGTTATGGTCAATTCAATCTACTCTAAGGCATTGCTGGAATATCTTTCCGACCCCCTA
TTTACCAATTTTAGCACCTTTGCCCATGGATTCTATGCTGGATTGACTCATTTGCTACAATGGCATCCATATTTT
CTTTCCAAAGCGGTTCTCATTCCTCAGAGAACAATAGCTGCCGAACATTCTTCTGTAATTAGCGAAGATGAGAAT
GATATGAGATCTTACCATGAGGGAGATCTTCTGCTTCTCATGCCTGATTGTGTCAAAAATAATGATTGTGCCAAG
ATATTGGAAAGCTACACTAAGAAGAAGAAAAGCCAATCTTAG

Downstream
tttcataatagaacgaaatagatatatactcatacactttgtgcaacagagacctctgtt
ttagtagtctcttcagttttgtggttgtcaactgcttggtcctcaggggattggggttcg
tcgttaccaagacgctcttctaatacttttaaggctacctgtcgtcttctttctgcaata
tctcttgattgagaagttttattgattcgtctaatactcctcaagttgccagattctctc
tcatcgtcattgaacccttgattaggccaaaaaggattagcaaatgatagaatatgcga
cacagtggtttcaaatatttgtgaagaggagcgggaaagaactgcatgaaattgaacgta
tcagacgcatcaccttgaattgaatttttggtgttgtctggtagaacatcgttgataacg
ttgagtctttggtaaaatcgtaggtaagtccatgaaactaggaaactattccaaatcggt
agaaatggagttaacttctgcaaaacgatcgaaataaacagcgacaaaaccagtacaata
aaaggaatctgttttaccctaaggtttaggaatttcacttgatgctcaggagagtactgc
ttgagtacgactagaaaaggcatcaaaaatacaaaacttccatgattgacaggcaccaaa
ggatcgtaactactgctagcagtagccacaatgaccttgatcagacaagtgaataaattg
gtaaaaaatgctgacactagcaaatacgtgatggtttcactaaagccagcactgaaaaat
tgtataccgtcaacattctcctctaaattccactgtttctctgtgtaagcagttccaaaa
aagatattagcagcacctatgaggaattgaatgatattaacatcagtaaatgcaccagtg
ataaacgtccatgggaaaataaaactcgtctttggaactagttccaacatagggataact
atatcccttacgctttcaccactgttccccaaacttttaga

AA
MYLGRSEKDAFKPKKNRILVPRSLAIYLVISVLLTNFLVFQLFNINLIPFLNGRSSKNNFDYLVKKEFLNLEEPY
RGFYKEEIENPTNYFFPTIEHASRLNEMGLDNLFQYTVDKQTKVHQYMLDPELKYEDQPKGNELDFVRREFLSNG
MKVHRGANSPELVIVTGIDFETFDSSYLGNITQNRIDYAQKYNFGVYVRWIQEFAPQFNNFQQSKDWTKALLMRA
AMLAFPNSKYFWYIDSNCFIMNMETNIIPYILEPHVLGPIILRDHPLILPDGGIKTYANIDPKDVHLILTQSATS
IRTDSFVMVNSIYSKALLEYLSDPLFTNFSTFAHGFYAGLTHLLQWHPYFLSKAVLIPQRTIAAEHSSVISEDEN
DMRSYHEGDLLLLMPDCVKNNDCAKILESYTKKKKSQS

SEQ0139
P. pastoris homolog of *S. cerevisiae* **MNN10 (YDR254W)**
Subunit of a Golgi mannosyltransferase complex also containing Anp1p, Mnn9p, Mnn11p, and Hoc1p that mediates elongation of the polysaccharide mannan backbone; membrane protein of the mannosyltransferase family
S. cerevisiae null mutant: chitin deposition: increased; competitive fitness: decreased; freeze-thaw resistance: decreased; metal resistance: decreased; resistance to mycophenolic acid:

decreased; resistance to nickel(2+): decreased; resistance to Calcofluor White: decreased; resistance to rapamycin: decreased; resistance to wortmannin: normal; resistance to caspofungin: decreased; resistance to cisplatin: decreased; viable; protein secretion: increased; resistance to mycophenolic acid: decreased; septum formation: abnormal

S. cerevisiae overexpression mutant: abnormal cell cycle progression; nuclear morphology: abnormal; resistance to hydroxyurea: decreased; resistance to nocodazole: decreased

Chr 2-2, 0185

5' region

agcacgttgatgcttggttaaaactgggtgaggtccagacccaaaatgaaaaggagtcag
acggtattgcagctctagagaaatgcctggagttggaccccaccaatttagcagctctga
tgactcttgcaatttcttacattaatgatggttatgacaatgctgcttatgctacattgg
aaaggtggatcgagacgaaataccctgatattgcttccagggcacgctcgagtaatccag
atttggatggtggtgatcgtattgagcagaacaagcgtgtcacagagcttttcatgaagg
cagcacaattatcaccagatgttgctagcatggatgctgacgttcaaactggcttaggtg
tattgttttactcaatggaagagtttgacaagactatcgactgtttcaaggccgccattg
aggttgagcccgataaggccttgaactggaatcgactgggtgctgccttagctaattaca
acaaaccggaggaggcagtagaggcatattccagagcattgcaattgaatcctaactttg
ttagggctcgctataatcttggtgtttcattcataaacatgggcagatataaagaggctg
ttgaacacctgttgactggaattagtttgcatgaggttgaaggtgttgatgcatcagaaa
tgagtagtaatcagggccttcagaataatgcccttgtcgagaccctaaagagagcatttt
tgggtatgaataggagagacttggttgataaagtatatccaggtatgggattggcccaat
tcagaaaaatgtttgacttttagatagaattcgaaaccccacaataagtagaccacatat
atccaacttgatactagctccttctttcctcccctgaattactgttagcgaagtccggca
tccagatcacgtcgagcatctcacgaaacgtattttcaacattgatcaacatttctcata
tctctatcgccatctattttggattatagtatccattgtaa

ORF

ATGTCTAGGCAAAGGTTCCAGCTTCCTTTGAGTAAAGAGGAGGGGTTGTACTCTGATCTGGATCCAAGTTTTGAA
TTCAAGCCCTATAACGGTGAGAATTCCTACAACAACACCATCAGATGGCTATCAAGTAAAATACACAAATACAAG
AAGCTGCTTATACTACTAGTATTCTTACTTTTAGTATCTACTTTTACTTCGGCCCCTTTAGTTCCCCATTTCCAT
TCAAACAAGGATCCGAGAGTCGTAATCATCTTGGCTGCTAATGAAGGGGGTGGTGTCCAAAAATGGAAAGGTCCT
CAGGAATGGTCAGTTGAAAGAAGTTCCATCGCCAACAAAAAAAAATATGCTTCTAAACATGGTTATGGCCTAGCA
ATTAAAGACCTGACCTTAAAAAGACGGTATAGTCACGAGTGGAGAGAAGGTTGGCAAAAGGTAGATATTCTCAAA
CAAACCATGAGACAGTACCCAAATACGGAATGGTTTTGGTGGTTGGATTTACATACTCTGATAATGGATATGGAC
GTTGATTTAGAGGAATATTTGCTTAACTCTGTGGGAAGCAAAAGTTATCGAACCGTCACAAGCTTCAACCCCATT
GGTATCGAAAATCATGTTCCATACACAGACACCAGTCAACCGGTTGATTTGATAGTTGCCCAAGATTGTGGAGGT
TTCAATCTTGGATCATTTTTTGTTCGTCAATCTGAGTGGACAGAGGCTTTACTGGATGCTTGGTGGGACCCCGCT
CTTTATGAACAAATGCATATGGCTTGGGAACACAAAGAGCAGGATGCTCTAGAAACTCTTTATGATCATCAAGGC
TGGATTCGGTCTAGAACAGGTTTTATTCCCTTAAGATCCATCAATGCTTTTCCACCAGGAGCTTGTTCAGACCAA
GCAGACAATCCAAAATTTTTTTACTGTGAAAGCGACAGGGACTTTGTGGTGAATATGGCTGGCTGTGAGTATGGA
AGAGATTGCTGGAACGAGATGGAGTATTACAAAAAGTTAGCCAAACGGCATGAGGAGAGATGGTGGAAGTTTTGG
TAA

Downstream

tcattttccactttgtttattaaatagatagaccattttatcatactgtactatttcagc
ctttttcgcatcctgcgcaaggcacacaggggataggctaggaccagccaccttccagc
actccctactctagtttctatttacgcatcgcgcgcgcttctggatggtggaaagtctta
tttatttcgcttcaccgtgataaatatggctgaccagcactgggtaaatgaacaagagac
ttttaaccctagtaacggtaataatcaagatcaatatatcatgtcatatctgaatatgaa
tgagaatttaccggggaggaaagtgtgaatccaatgtccattcttgaccagtccaagca
acggccccagcaatacacaggttctaatgagtccaataaccagaacgaactaggtttgac
ggcggatccttctgtatttttcaatcagccttctaatatgcaaagttctaactctcccat
ggtatcgcaagtccagaatcagtcaccgtcatttcaacagtattctcaagcaccaactcc
tggtcaactgcaacagcagtttcaggatatgtctaatggggattccaaacaggttgttgg
agatcccaaacatcaatctcagatgcaaaaccagttaagaagacaatcgccgtctcttta
ccaagcaaattctaattcgtcccagcagcaacaactgtatcaaaaattccaacaagaaca

gggattgaactatagaagaatgcaacaacaacaacaacaacaacagcaacagcaagcaca
acttcaacaacaacagcagcagctacagcaacagcaaatgcaacaggctcaggcgcgtgc
tcaagcacaggcacaagcacaggcgcaggctcaagcccaagctcaagcccaagcccaagc
acgactacagacacagcagaaacagcaaaattctcctccacagtttcagcaaaactaccc
ggctaattctcctcaaggtgtacaggctaaccaagctcaaa

AA
MSRQRFQLPLSKEEGLYSDLDPSFEFKPYNGENSYNNTIRWLSSKIHKYKKLLILLVFLLLVSTFTSAPLVPHFH
SNKDPRVVIILAANEGGGVQKWKGPQEWSVERSSIANKKKYASKHGYGLAIKDLTLKRRYSHEWREGWQKVDILK
QTMRQYPNTEWFWWLDLHTLIMDMDVDLEEYLLNSVGSKSYRTVTSFNPIGIENHVPYTDTSQPVDLIVAQDCGG
FNLGSFFVRQSEWTEALLDAWWDPALYEQMHMAWEHKEQDALETLYDHQGWIRSRTGFIPLRSINAFPPGACSDQ
ADNPKFFYCESDRDFVVNMAGCEYGRDCWNEMEYYKKLAKRHEERWWKFW

SEQ0140
P. pastoris homolog of *S. cerevisiae* **ANP1 (YEL036C)**
Subunit of the alpha-1,6 mannosyltransferase complex; type II membrane protein; has a role in retention of glycosyltransferases in the Golgi; involved in osmotic sensitivity and resistance to aminonitrophenyl propanediol
S. cerevisiae null mutant: viable; acid pH resistance: decreased; alkaline pH resistance: decreased; glycogen accumulation: increased; chitin deposition: increased; competitive fitness: decreased; freeze-thaw resistance: decreased; lipid particle morphology: abnormal; metal resistance: decreased; resistance to hygromycin B: decreased; resistance to rapamycin: decreased; resistance to cadmium dichloride: decreased; resistance to hydroxyurea: decreased; resistance to mycophenolic acid: decreased; resistance to sulfanilamide: decreased; resistance to L-1,4-dithiothreitol: decreased; resistance to Calcofluor White: decreased
S. cerevisiae overexpression mutant: decreased vegetative growth

Chr 3, 515

5' region
tcctccggcccccaaaccaagaaatcaaaagaaggagccagtcatcgaagaaaataatga
tgattcagtggtggaaagtttgaaaaaactctcaccaaagccaaagaccccagccaaacc
aagctctttacaagtgccctcaaagccaataactaccaagcctgcacctttgaaacctac
aaaacccaaagatttggcgaaaggccttgactcaaatccaccagaagctgttgtcaaggc
gaaggaaataaaatccaagcctttaatacttccgaagaaggctagtaagagtaactttaa
tcaaccgcaggaagatatattgcagtcacaattgcagaaactaagagaatccaagtctag
tggtaagaatcataatttcaacgcagctcaagaaacaatatttaacgctcaattgaatgg
actgaggtccagaaaaccagaccagaaccaatctaagggattgttaccaaaaagatcaga
aacattctcatctaagcctgttcctattccatttatggtagatccatcgtccagcaaggt
gctagagaaatccttgacaaaatcccgaaccattgatggatccgacgtaggattatcaca
agaacatcatcagagtctgtcccatccaaacatgactcgagccaaaggccccaaaaggag
acttcctaagacgctcagcaacccagaagcaatcaaacaaccacaacctacatcctctac
taccagcttgggcaagaaagtgcctcctcccaagccatccaagaagagaattgtctccaa
cagtgaactattcctatagaagtttaatgcattttagattatataatcagggagaaaccc
cgtagagaagttccacgcggaaatgtttcctgaaatccaccatcgattttcgtacaaaca
cgacacattttgtgctttccttcaaaccaacaataaatatcaacggtatcatcgattcct
tcgcactgtcgagcccaaaccattttattgtgttgtaggag

ORF
ATGTCAAAAGACTCAATAATAGCACAACCAATGGGGACATTAAACTTCAAGGTACTGAGGCGATACAACTCCACG
CTTTTGAGGAGTTCAAAACACTTTGCGACCAACCCTAAAACGATACTGGTCGTTTTAGTCGCCTTGTTTTTGATG
TTGTTTAACATGATACGACCATTCAATAACAGCCTATATGAGCCTGATTTTGAAAACACCAAGAACACACATTAT
TACGATTTGGACAACTATAAAGGGAGTGAGGATGGTTGGCAGAGAGGTGAACGCGTGGTCTTTCTGGTACCTTTG
AGAGATGCAGCTGCCCACTTGCCTATGTTCTTTAGACACATGAGAAATATGACATATCCTCATCATTTAGTAGAT
ATGGCGTTTTTGGTGTCTGACTCCTCGGACAATACAATGGGAGTCCTTAAGGACAACCTTTTAGAGTTGCAGCAT
GACCCTGATCCTAAGATCCATTTCGGGGAAATCATAATCTTCGAAAAAGATTTTGGACAAGCTATTGGACAAGGT
TTCAGTGACAGACATGGATTTGCAGCCCAGGGCCCTCGTCGTAAGTTGATGGCTCGTGCCCGTAACTGGCTCAGT
TCGGTAGCTGTCAAGCCTTACCATTCATGGGTTTATTGGAGAGATGTCGATGTGGAAACTGTTCCCACGACAATC

ATTGAGGATTTAATGCATCACAATAAAGATGTTATTGTTCCTAATGTATGGAGACCACTACCTGATTGGTTAGGT
AATCAACAGCCGTACGACTTGAACTCGTGGCAAGAGAGTGATGGTGGTCTACAACTGGCTTCAACTTTAGACGAG
GATGCAGTCATTGTTGAAGGATATGTGGAATATGCTACCTGGCGACCACATTTGGCCTACCTTCGGGACCCCTAT
GGAGATCCAGAAGCAGAAATGCCATTGGATGGTATTGGTGGTGTTTCCATTTTGAGTAAGGCCAAGGTTTTCAAG
AGTGGTAGCAATTTCCCTGCATTTTCATTTGAAAAACATGCTGAAACGGAAGGATTTGGTAAACTATCTAGAAGG
ATGGGATATGATGTTGTAGGGTTGCCTCATTACGTGGTGTGGCATATCTATGAGCCATCTTCTGACGACTTAAAG
CATATGGAATGGATGGCTAAGGAAGAAGAGAGACAGAGGCAAGAAAAAGAGATTCAGGCCGTTTACAAAAAGTAT
TGGGATCAAGGATTTGAAGACGTGAGCGATGCATGGGCAAAAGAAAGGCACTACATGATGAAGAACACAGATTTT
AGAAAACAAAGGAAAATCGAAGTAGACTGGTCGGGTGAAGATGACTACGAACCAGCGGTCAAAGCTAATGCTGTT
AAGGATCAGGATCGTGCCGCTAACTTTTATCCTCAGTAA

Downstream
taaatagtaaagttatttgtcatttaggaagggttaatacatattaataaaaaagtttta
tgcccatttaccaagcacactcatcagtatcataccagcaataagagatagaaatccaac
ggttgtgttcagccagttggcgttgactaatgctccatatatccagtcaccggcccacat
ctctaccaaaccggtccagatcaaaataccagcagagaaagagtcaagagtacccaacgc
aatcaacgttgaagagctattaccattgaatctactcaacacgccaataccaattgccat
accaataggagtaataagggcaaaagcaagagccataaggagtttataccagagtgagac
catattcagttcggcaatcttggttcccaaggcaaaaccctcaaacatttgatggaaaat
gataacgatgaacagtgtaataaagaaagaatcaccagcaactaccaatgtgacaccaag
taggatggagtggaaaataataccggcttccaacataataacggaaatttttctgctgag
tggtggaacaggcacggttccacctctttctagagaagacgaagaatcacttagctcgtt
ctcttttggattttcgtcagtcacttcgatttcttcttgggcgttttcagcatctgcttt
cttgagtctggctagagcaattcttttaatgaaatattctatagcaaaacagaggaaaat
accagccatggtaattgaggtaccagtagcttcatagtcaagtggagccatacactcatt
agaccacatcatttcagcatgggtaattaaatgcacaaatgcggtcgaaataatcacacc
tgttccaaactgtctgaaacctgtgaagatcacaccagtcaagctgaaatttaggaactt
atttgttaggataggaaggtagacacctattgcacttgtagctagaatggcaaacagaca
tccgattctaaggggaatattgtaatcatggtctggggcac

AA
MSKDSIIAQPMGTLNFKVLRRYNSTLLRSSKHFATNPKTILVVLVALFLMLFNMIRPFNNSLYEPDFENTKNTHY
YDLDNYKGSEDGWQRGERVVFLVPLRDAAAHLPMFFRHMRNMTYPHHLVDMAFLVSDSSDNTMGVLKDNLLELQH
DPDPKIHFGEIIIFEKDFGQAIGQGFSDRHGFAAQGPRRKLMARARNWLSSVAVKPYHSWVYWRDVDVETVPTTI
IEDLMHHNKDVIVPNVWRPLPDWLGNQQPYDLNSWQESDGGLQLASTLDEDAVIVEGYVEYATWRPHLAYLRDPY
GDPEAEMPLDGIGGVSILSKAKVFKSGSNFPAFSFEKHAETEGFGKLSRRMGYDVVGLPHYVVWHIYEPSSDDLK
HMEWMAKEEERQRQEKEIQAVYKKYWDQGFEDVSDAWAKERHYMMKNTDFRKQRKIEVDWSGEDDYEPAVKANAV
KDQDRAANFYPQ

SEQ0141
P. pastoris homolog of *S. cerevisiae* **HOC1 (YJR075W)**
Alpha-1,6-mannosyltransferase involved in cell wall mannan biosynthesis; subunit of a Golgi-localized complex that also contains Anp1p, Mnn9p, Mnn11p, and Mnn10p; identified as a suppressor of a cell lysis sensitive pkc1-371 allele
S. cerevisiae null mutant: glycogen accumulation: increased; competitive fitness: decreased; hyperosmotic stress resistance: increased; resistance to amiodarone: decreased; resistance to PM02734: decreased; resistance to ethanol: decreased; resistance to hygromycin B: decreased; viable

Chr 3, 0620

5' region
aaaataaatatatggtgttgctatgtaggccgtaacacggctcttcacctgtatggagca
tcggcatctaaagcgtactggatttcctgttcccatttgtcaccattagcaagcaatttt
tccttgttatacaacagctcctctctagtttcggtggcatactcaacgttaggggtttca
acaatggcatcaactggacaactttcctgacaataaccgcagtagatacacttggtcatg
tcaatgtcgtacttgtaagttcttcgggaaccgtcaatacgttcttcggcttcgatggtg

atagcctgagcaggacaaactgcttcgcataacttacaggcaatgcaacgttcttctccg
gatggatatcttctcaaagcgtgttctcctctgaaacggggagaaactggacctttttca
aaagggtagtagattgtatatggggcacgaaagtacatttccaaagtgatataaaggcct
ctaaagatctcagacagtaggtaccatttggtggccttggaaagagcgctttcactgctt
tcttcccacgtctttggtcttgggagtctgaatccttctggactgtgtgccttggatcga
accagaagttaaagccaatgtgggtttctgtgcaatgagtcccgtcgaaaacctgcaaac
tggtattctgtttatggcacggcaggaaggcttcaatgctagtggtctaaacattgtgtt
ataaatgtgtaggtagaaacctataagtttgtgttggggttcgataatcaaacttgcctt
gaaggtaattggctaggtgcttgttgagtgtcgcacgtctcggtatttttttacttgtcg
attctctcttactgtccaagatgtacgaaagatgtgaagtttctttgttggcgattttct
tgcagcgcgatcgcggtcggtatgatttcccttcagaccaaaaacagatatagatattca
attgaaactcgcccttccaacaccctgaacaccacccccta

ORF
ATGCAACAACAGCTGTTTAGAAAAGTAATTTGGCTCACGGTTGGACTCATAACGGTGATATTGGTAATAATAAAG
ATTTCATCCAGCAAGTCCACCGCCACAGACCTTCAGAAAGTATTGAAGAATGCCAATATTTTACCGCAGGACGTG
ATCAATTACAATAGCAGAAAAGTAACGGATGAACTCGCTTCAAAGCTGGACGAGATTCAAAAGAAGTATCTCTCG
AAGCAAGATGATAGAATTAGCAAGCTCGAAGCTGAACGGGCAGATCTACTGGAACAGGTTAGATTTCTAAGGAAC
CCCCCTGCAGGATCAAGTTTAAGAGAAAAATTGGCTTATCTGTTTCCTTATAATGAAAACGGCAAATTCCCTGCT
TATATATGGCAAACATGGAAGTATGGCTTGAATGACGATCGGTTTGGAGAAAAGTTCAAAGAAGGCGAAACTCAG
TGGGCTTCGAAGAATCCTGGTTTCGTTCATGAGTTGTTTAACGATGATACTTCCGGTGTGTTTATTCACCATCTG
TATATCAATGTTCCAGAAGTGATCAAAGCATACGAGCTGCTTCCCAACATAATCTTGAAAATGGACTTCTTCAGA
TATTTGGTTTTATACGCCAAGGGAGGTGTCTATGCAGACGTTGATACTATGCCTCTTCAGCCTGTACCAAACTGG
ATTCCTGAAAATGTCTCCCCAAAAAGCATCGGGATGATCATTGGAATACAAAACGATGCTAACAATCCAGATTGG
AAAAAAGATTACGTACATCGTCTACAATTTTCCAATTGGTGTATTCAAGCGAAGCCTGGTCACCCAATCTTGAGA
GAATTAATCGCCAAAATCACAGAGGATACACTGCAACGAGCCGAGTCAAACTCACTGGAACTAGCTGACATTAGC
GAAGAAGGCGGCCTGTCTGATAAGAATTTGTCCATTATGCAATGGACGGGTACTGGTATTTTTACAGATGCCATA
TTTACCTATTTTAATGACTACATTCAAAGTAGTATCTATACCAAAGTTACTTGGAAAGAATTCTCCAAATTGAGA
AAGCCCAAGCTTGTCAGTGATGTATTGGTACTGCCGATTATCAGCTTCTCGGCCGGTGCAGGTAGTGGAAAATCG
ACTGAACTGAACGATCCCTTAGCATTCGTACAACATTATTTTGAAAGATTACATAACGACAACCACTAA

Downstream
gggtcagaaccatttagattgtctggatctatcattatggccttgtttatagacaaagaa
ttgtatcctggactgaagggaagtttatagagtaataccccctctgacaaccaactcgaat
gggtattttgagataatttcccatatattattgtcttcactgggagacttacttcttctt
tgtagttccggtatcggcagcaaacttatggtggtcacatcttccaaacttctcggtata
gtagaactatttgtggttgatacggtctgggtatcggtgataatgcttgtaataatggat
tttgtgggtgtaactgaagctggtgtattagtggtagtggtagtggtggttgttactgtg
ctatctgtaggttctgggagtggtttctgagccgggtttaaaattttgatctctgtggta
tcttcgtctggatttcgactaaaaaatccgccttgaaatacgagtggaacaggtaaacga
ttgatggagcccagaagcactgtgtccaaggaagattcatcaaaagtgaaaggagtgcca
tcttcgtttccatcgttaagatgagaacttttggcaaaaaaatcaatcagaacgtcatct
atttctatggcgacaaatccgggattgaatccttttaccgttaaactgaaaactatctct
tccttagttagtaagatattgtctatccggtccacagtgaagtttgttaaaggcttggtg
atggcgataaaaaacccaaataaaaaccccattatgagaaacgctgatatcaagctcgca
gtgtaacaaaaatgttttacttgacgtaatcttgtcatcttggtagatgtgaaatcatgc
ggcgagtaataatagtggtctggtttaatattctggctgtgtctttttaatttatgtcta
ttgttattagatacggaggttgaaccgtattccaagggacattcatgatggtctaacaaa
agcggttcttgataaccgtatgcaggatagtcatataatcc

AA
MQQQLFRKVIWLTVGLITVILVIIKISSSKSTATDLQKVLKNANILPQDVINYNSRKVTDELASKLDEIQKKYLS
KQDDRISKLEAERADLLEQVRFLRNPPAGSSLREKLAYLFPYNENGKFPAYIWQTWKYGLNDDRFGEKFKEGETQ
WASKNPGFVHELFNDDTSGVFIHHLYINVPEVIKAYELLPNIILKMDFFRYLVLYAKGGVYADVDTMPLQPVPNW
IPENVSPKSIGMIIGIQNDANNPDWKKDYVHRLQFSNWCIQAKPGHPILRELIAKITEDTLQRAESNSLELADIS
EEGGLSDKNLSIMQWTGTGIFTDAIFTYFNDYIQSSIYTKVTWKEFSKLRKPKLVSDVLVLPIISFSAGAGSGKS
TELNDPLAFVQHYFERLHNDNH

SEQ0142
P. pastoris homolog of *S. cerevisiae* **MNN9 (YPL050C)**

Subunit of Golgi mannosyltransferase complex also containing Anp1p, Mnn10p, Mnn11p, and Hoc1p that mediates elongation of the polysaccharide mannan backbone; forms a separate complex with Van1p that is also involved in backbone elongation
S. cerevisiae null mutant: septum formation: abnormal; glycogen accumulation: increased; chitin deposition: increased; competitive fitness: decreased; resistance to gentamycin C1: decreased; resistance to L-1,4-dithiothreitol: decreased; resistance to sulfanilamide: decreased; respiratory growth rate: decreased; viable

Chr 4, 0103

5' region
tctggagactattttcgacaaggcatggatttccaagcaccaggacatcttcagatggtt
cagcactgtgtcagcccacccagttctgtcttggagatacgttgacttcaagccaagaga
caagcctgtcgaatacgtccctcctaagaaggaaaagaaagaaaagaagaaggaggagaa
accaaagcaagagaagaagcctgctcctgctgctgctgagcaatctgatgagccacctgt
tgagaagaaggcaaaacacccattggaagctcttggaaaaccaaagttggccattgacga
atggaagagattctactccaatgaggagaccagagagtctgcaattccatggttctggga
acactttgacgccgatgaatggtccttgtggaaggtcacttacaaatacaatgatgaatt
gaccttgaccttcatgtcaaacaacttgattggaggtttctttgccagattgaccggttc
gatcaagtacatgttcggatgtgctgtagtgtacggagagaacaacaacaacggtatcat
cggtggtttcatggtcagaggacaagactacaagccagcttttgaggtcgctccagactg
ggagtcttactcttacgagaagttggaccctaacagtgaggaagaccgtgagttcttcaa
caacatgttggcttgggataagccaattgttgtcaacggtgaagagaaggagattgctga
tggtaaggttttgaaataggataagttagcttagtaattttaaggtaaacgaggccccag
tgatactggtattaattggtactttctttcctggtagtctggtagcccgtccaagttttt
cttgtcggcggcggagaacggtttttttttatctttcacccactctgatcggctcttct
taagctttttttttctcctttcgcgactaccgaggaattagtttcaatcaaaatgcctga
ggttcagcgtgttgtttcctcccactcactagaccattcca

ORF
ATGTTGGTGCCCAATAAAAACCATTTGTTCCACCAGATCAGAAGGAAGCCGTCGCAAATAGTGGCGCCCATCTTG
TTTTTGGTCCTCGTTGTGTATTTACTATTTGGAATCGGGCAATCTTCGTCCAAGAAGCCAAAGTATTCATACAAA
GATAAAACACAAGGATGGCTATTGTCAAGCAAAATCCCTCCTGGGCCGGACGAAATGGCCAAAAATCACATTATG
CATTATAATATGAACTTACTTGAGACTACTGTCATGCCTGCGTTCAACAAAGAACAAGTCTTGGTGCTGACTCCC
ATGTCCAAGTTTTATCCCGAATATTGGGATAATTTATTGGCCCTCTCCTATCCAAGAGAACTAATGGAGCTAGGA
TTTATTGTTCCACGTGGAAAGGCTGGTGATGAAGTCGTCAAGAAATTGGAAACCGCAGTCAAGGCCGTTCAGAGA
TCTTCCACCGAATCGGCAAGATTCGCTCGTGTAACCATACTCAGACAGGATAGTGAATCCGTGGAATCACAGTTA
GAGAAAGATAGACACGCTTTAAAGGCTCAAAAAGAACGTCGATCAAAGATGGCTTTGGCAAGAAATTCGTTATTG
TTCTCAACCTTAGGCCCCTTCACTTCGTGGGTGTTATGGCTCGATGCTGACGTTGTAGAAACGCCAAAAACTCTG
ATTCAAGATTTAGCCACCCACAACAAAGCAGTGATAGCTGCTAACTGTTACCAAAGATATTACAACGAGGAGACC
AAGAAAAATGACATTAGACCCTACGATTACAACAACTGGGTAGAATCAGAGGAGGGTCTACGGTTGGCTTCTACT
TTGGGACCAGATGAAATCATCGTAGAAGGCTACGCAGAGCTTCCCACCTTCAGAGCATTGATGGCACATTTCTAT
GATCCACATGGAGATTTGAGTACAGAAATGCAACTGGATGGTGTTGGAGGTACTGCAGTGCTAGTTAAAGCAGAA
GTGCATAGAGATGGTGCCATGTTCCCAAGTTTCCCATTTTACCATCTTATTGAGACTGAAGGATTTGCTAAGATG
GCCAAACGTCTGGGCTACCAAGTCTTTGGTCTTCCAAACTACCTGGTTTACCACTATAACGAGTAG

Downstream
atggggcagtagtgttaagtatcgcgacacttgcagcatcaaagtcgtgcttcttatcaa
tcattaatcctctttcatactgataaggagtagttatccctacccacttcttaggcaaca
atagagcaaggccttatcaataagtatcgcaaaatagtagaataaatccaatctataaag
atagttccttcacctttcatggcaaagtctcccgaaacaagaatgcggcctcccctggag
gcttcatttcactcctcctccacaagacggtttagcgggcctaaaaatagtcttgccgag
tccaccagcggattggcctccccattggctaccaaagagcttaaaatttcgaccagcaaa
caaaatcagactccagtgaccatcagcaacgatatgctgaacataaatcctaatctcaag
tcaattgtggagggtggaacgtctcctatagcagaatcgaatgtcggtgtggcagcgttt
gcagctacagcaacggcagcatctttacaactgagtggctccaactcctctaaaagtccc
acatcaaggaatttgtcgcatgttccctgcaaattctaccgtcaaggggcctgtcaagct
ggttcttcatgtccctttcccacactttaacccaaacaagccaggcagcaacttgcaaa
tattttcagaaaggaacatgcaaattcggatcgaaatgtgcactggtgcatatttctccg

gagggaaaaaaggttaatctgaaggctctaaatcaagcttatcaggtgccaaaacatgaa
cgatctctgtcctcttcccagggtttaagtcagcagatacacaaagatccagaaacatct
gatgatcaagcttccagcatgtcctctcctactcgaacagtaacatcacagatggcatcg
tctcccttcagcaatacaaactccatttggtcaggcacaggttcaaaggaaaggcgattg
tctggaccatctttttcgacagggtctttctctagaggttc

AA
MLVPNKNHLFHQIRRKPSQIVAPILFLVLVVYLLFGIGQSSSKKPKYSYKDKTQGWLLSSKIPPGPDEMAKNHIM
HYNMNLLETTVMPAFNKEQVLVLTPMSKFYPEYWDNLLALSYPRELMELGFIVPRGKAGDEVVKKLETAVKAVQR
SSTESARFARVTILRQDSESVESQLEKDRHALKAQKERRSKMALARNSLLFSTLGPFTSWVLWLDADVVETPKTL
IQDLATHNKAVIAANCYQRYYNEETKKNDIRPYDYNNWVESEEGLRLASTLGPDEIIVEGYAELPTFRALMAHFY
DPHGDLSTEMQLDGVGGTAVLVKAEVHRDGAMFPSFPFYHLIETEGFAKMAKRLGYQVFGLPNYLVYHYNE

6.3.1.3. α-1,2 mannosyltransferases

SEQ0143
P. pastoris homolog of *S. cerevisiae* **KTR1 (YOR099W)**
Alpha-1,2-mannosyltransferase involved in O- and N-linked protein glycosylation; type II membrane protein; member of the KRE2/MNT1 mannosyltransferase family
S. cerevisiae null mutant: viable
S. cerevisiae overexpression mutant: increased filamentous growth

Chr1_3, 0138

5' region
gtttttgtcttcgctgatatcatggtgaaacatcaaagtggtatggggcttagcatattc
cttgtgaagtttgtactcggccaacgaagtgaacatacagtgggtggtatttggtcctgg
gaaagcccattgaatgaaatgaatttccacaccttcgtcttcctttggtagtggtaaaat
gaaagaagggttctccctagcgttggtgtatagtttttgataaacgtctttgggaataac
agcaaccaacacattttccttctcctggaatctggctctccacaagaattcaacctgatc
cggcgaaaggtctttgatcttctccaactttaggaaatcatccagtactttataaggctt
gattggcgcatcaggcgatctaggagatcttaattttgtgaccttatgtccatttttgt
agctttcctttcatactcttccagatctgtcaaggggtccatcctgttcaactctatctt
ggttttctctattttatccttgtatttttcttgaagtttttcaattgactctgcaccttc
ttccttcatctttttctcaagttttttctttatatcgattcagcacgtcctctttagcagc
tttggtttggacatatctaatgttatccaaagctacggcatttcggtactgaatggggag
gccaacacacctgttgaataggttgaagtttttgatatatacaaagagcggtatgacat
gttctcgttacttccagaattcataggggtgaatccaaaccaaaaactcagtaggaagat
atttcgaagatgttagttgatacttttccacgcacttctccagagaataaaactgaaaa
tgtccgacggcacctatgtccagagggttatattgcgccaagttgattacgtaactaata
aggtcataaaatacagaccatacaggtctataaataaattgggatagggtgcttcggaag
ttggttgatctgattcagttcctgaatgtgccgctgtaggg

ORF
ATGATGCGAGCAAGATTAAGCCTTGAACGAGTTAACTTGAGCTTTATTACGTCCGTATTTTTGGCTTCAGTTGCA
GTTCTTTTCATCTCTTTGGAGATGCCGAAAGTTCTTGCAAGGGACCGCCAAATACTCAAACTAAAACTTGGTTTC
ATGGGCAGCGGACTCCAAAAAGGGTCCCTTGAGACAAGTGGAAATATTGAAAACACAGAATCAAACATAAATTCA
CAGACAACTCAACACATAGGCACCATAGGCGCATCAAATGAACGAGCAAATGCCACCTTCTACACACTTTGTCGA
AATGAAGAGCTCTATCAAATGTTGGAGACAGTTCAGAACTATGAAGACAGATTTAACAGTAAATTCAAGTATGAT
TGGGTATTTCTGAATGACTACCCGTTCACTGATGAGTTCAAAAGGGTAATCTCCCATGCAATAAGTGGTGAGGCA
AAATTTGGTCAAGTACCGGCCTCTCATTGGAGATTTCCGGATCATATTGATCAGCAGAAAGTATACGAGTCGATG
GATAAAATGGACTCAGACAACACAACAGGCGATTATCTGGGGCTCCCTATTCCTTATGCTAAAAGCATAAGCTAT
AGACACATGTGCCGCTATCAATCTGGTTTTTTTTATAAGCATGGACTTCTTCAAGGTTACAAATACTTCTGGAGG
GTTGAACCTGATGTCAAGTTATACTGTGATATTGACTACGATGTGTTCAAATCAATGGAGCAAAATGGCAAGAGA
TACGGTTTTGTCATCTCAATGATGGAATTTGAAAAGACAATCGAATCATTGTTTAAAGAAGTGAAAAATTACCTC
CAAATGAAGGGGGTGTCACGATTGCTTGAGGATACTGATAACTTATCTGATTTTGTATACGATGAGCTATCGGGT
GACTATACATTGTGTCATTTCTGGTCCAATTTTGAAATTGGAGACCTGGATTTTTTCAGAGGTCGAGAGTACAAT
GAATTCTTTGACTATTTGGATAGCAAAGGAGGTTTCTACTATGAAAGATGGGGGGATGCTCCGATTCATTCCATA

GCAGTAAGTTTATTCATGCAATGGAACGATGTCAAGTGGTTTTCAGACATTGGATATAGGCATCCACCGTACCTT
TCCTGCCCGCTTTCTGAAGAAGTGAGGTTAGAGAAAAAATGCTCATGTGATCCAAAGCAGGACTTTACCATGGAC
GCTTATAGCTGTACAAGGTTCTATCAGGACATTATAAGAGATAAACAGAAATCTCAGGGAAGCAACCCCTGA

Downstream

ttttgggcatttagtttcaggacagtttagatgaggatacaaatattacataatcaacct
acttccaggtagcatatccgcactctctgtatttagatctctgctacctcagcttgggcc
gcccacacattagtagtacacagacaatggccatcaagtatttgattatgctttcgaagc
agggaaaagtgaggctttccaaatggtataccgtaatttctcagaaggaaaagcttcaat
taattaaagaattaacagctatagttctgaatagaaaggcaaagatgtgcaatgtcttgg
aatataaaggtgagttttgcgtttgaggcctgtgtctttagaaactttactaacaggatt
gcatagatcacaaaattgtttatagaaggtatgcttctttatttttcattgctggcatag
atgttgacgataatgagttattgacacttgaaattatccatagatacgtggaacaaatgg
acaaagcatatggaaatgtctgcgaactggatattatcttcaatttccagaaagcatatc
acatcttagacgaattactgttggacggtacattacaagagagctccaaaagggaggtcc
tgagaagggtgggacaacaggatcaatacgagttagaggacgagcttgatactaagttct
aatagatgtagctaaagtgatttgattgatcttaaaacgtagaggttggtactatgtcca
acacggagtaacaaaaaatgaaaaaattaaagtttgacacggacaggaattgaacctgca
acccttcgattgcactgtattccaactggagtcgaaagctctaccattgagccaccgcat
caactataaacaagaaagcaaaaaagctgtttaagatgcaaatttccgattaaaaaatac
tattataatgaaccgatggaaaagaagatctttctattaaattggtgtgggttattaaag
gtaacgcaattaagtaattgactgttttgtaggaactcata

AA

MMRARLSLERVNLSFITSVFLASVAVLFISLEMPKVLARDRQILKLKLGFMGSGLQKGSL
ETSGNIENTESNINSQTTQHIGTIGASNERANATFYTLCRNEELYQMLETVQNYEDRFNS
KFKYDWVFLNDYPFTDEFKRVISHAISGEAKFGQVPASHWRFPDHIDQQKVYESMDKMDS
DNTTGDYLGLPIPYAKSISYRHMCRYQSGFFYKHGLLQGYKYFWRVEPDVKLYCDIDYDV
FKSMEQNGKRYGFVISMMEFEKTIESLFKEVKNYLQMKGVSRLLEDTDNLSDFVYDELSG
DYTLCHFWSNFEIGDLDFFRGREYNEFFDYLDSKGGFYYERWGDAPIHSIAVSLFMQWND
VKWFSDIGYRHPPYLSCPLSEEVRLEKKCSCDPKQDFTMDAYSCTRFYQDIIRDKQKSQG
SN

P. pastoris homolog of *S. cerevisiae* **MNN2 (YBR015C)**

Alpha-1,2-mannosyltransferase, responsible for addition of the first alpha-1,2-linked mannose to form the branches on the mannan backbone of oligosaccharides, localizes to an early Golgi compartment

S. cerevisiae null mutant: budding pattern: abnormal; cell shape: abnormal; cell size: decreased; glycogen accumulation: increased; chitin deposition: increased; competitive fitness: decreased; resistance to hygromycin B: decreased; resistance to Calcofluor White: decreased; viable

3 homologs

SEQ0144

Chr 1-4, 0037

5' region

aaggtcgtgcacgacaacggttgttaatgaggttagttacttccgtcctatgaggatgca
actggtctaacaaaatctatcctcacagttcgtgtcctttctttcattaagtgatcaata
catcatgtcctttaggttagccagaacatttttttttttctctctttggatacgaactgaa
ggtttttttagtcttcaatagcactggctgacttcgaaacaccccatttttccactacaaa
ggagcttttttatttcgtttatagtctctggtagaatccttttggatatagcatgttttag
attttcttgatttagtttcttgtgaatcgacttttgcttaacagaaagcaaatcttcgac
tgagtaattcacaaaacttttaatgcatattcgaccgttcaataaatccagtatgaaggc
tatatcttgatttttgattgctattaaattcttcctaaaggttgttctatagttttttag
atctcgattgtgctttgagtataattctcgttcgtatgctgttgcgattccagtaatgtc
gtagcttaccccaccaagcgaaccttcatttttctttttttcaatattagcctcttctgagc
atccgtaattctagtccgcaaggcctcttctatcttgtggctacaggcacttctcaccct

tgccacttcaatttccaattctgcatcagtcttaacggatgaattatttgagtttctttt
cgccaattttcgcttgcgaattacccccttcaagcctttcccagtcgtgggacttttcgc
cataattgaaatgatcaagctgactagtgtacactcaaacttgaagtttatggcttgacc
aatacatcaacccaaaaatgttctgtgtggtcactacaccattatgttacatgttcgcgc
gtctgctatcataaccagaatcgtagtttctgcactttggaaacctacgcgtttcataaa
tcagtcctttcattaaattcaaagcgagatcaccgaagata

ORF

ATGCTCTTCGGACTAATTAGGCACTCACGTAGACAGTTGTTGTTTTTGGGTGCGCTGGTAACCGTGATAGTCTTA
ATTTTTACCTTGCCCAACACATCACCAATAGAAGCAAATGGAGTAAAATCAGAAGAAGGGTCCATTACACCTATT
ATTCCAGTTTTGGAAAGTCCAGCAAACTCATTGGAAAAAATTGTTGATACTGCCTCAGAGGAACGAATTGGAGGT
GCAACTTTGGAAGAAGGACATGAAAACAATAAGGAAGAACAAGCTTTGGAAAATGCTGAACGAGCTAAAGAGAAA
GAAAAAACTGAGGCGATTGCGGCAGAAGAGGAAAAACTGAAAGCAGCTGAGTTATTGAGACAGCAAGAAACTACG
AGGGAGAAGGAAGCGGCGAAAGAGGACGATAGCAAGAAACCAAACCAGGAACTTGTGGAACAAGACACTTACCTT
GACGATATTCCAGATGACGTAGAAGATAATATTATAATCTCAGAACAAGATCGCAAGAAAATTATATTACCGTCC
TACACTCCTAAAACAGACCCTGCCTATTCCAAGCGGGCCACTGCTTTAAAAATTTTTTATAATGATTTTTCATA
AAGGTGGCTGATTCAGGACCTAATACAGCGCCCATAACCAAGAAGACAAGAAAGAAAGGAAAAAGTAAGTTAAAG
GGTGACGTTTCCTCAGGTGATAAATATGAAGGTCCTGTATTGACGGAAGATTTTCTACGTTTCATGGAGATCTAT
TCTGACGAGTTCATCGATGCTGTATCAGAATCCCACTCAAAAATTGTCAATTTGATGCCCGAATCTTTCCCCAAA
GGAATGTATCAAGGCGACGGTATTGTCATTATTGGAGGAGGGGTGTACAGTTGGTATGGACTGTTAGCAATACGC
AACTTAAGAGATGGTGGGAATACTCTGCCAGTGGAGTTAATGTTGCCAAGCGACAATGAATATGAGCCTCAATTA
TGTGAGCAGATTTTACCGTCTTTGAATGCCAAATGTATTATGCTCTCTGATATTGTTGATCAAGATGTGTTAAAG
AAGCTTGATTTCAAGGGATATCAGTTTAAGGCATTATCTTTGTTGGCCTCAAGCTTTGAGAATGTTCTTTCATTA
GACTCTGATAATATCCCAGTTGCAAATGTTAGCCATTTATTTGATCATGAACCATTCAGCGAAACTGGGCTTGTA
TCATGGCCTGATTTTTGGAGAAGAACTACAAACCCTCGGTATTATGAAGCGGCAGGAATCAAGATAGGAGAATAC
CAAGTCAGAAATTGCCTTGATGGCTTTGTTCCTGAGTCTGATTTTGTTCATATTGGACTGAAAGATATTCCTCTT
CATGATCGGAATGGGACAATACCTGATGCGTCAACTGAGTCTGGACAATTATTGGTAAACAAGAATAAGCATGCA
AAAACACTTATGCTAATGTTCTACTATAATTTTTACGGCCCAGGTTACTATTATCCATTATTAAGTCAGGGAATG
GCTGGTGAAGGAGACAAAGAAACTTTTTTGGCTGCGGCAAATTTTTTTGGGCTTCCCTTCTACCAAGTCAAGGCA
GGTCCAGGGATTTTGGGCCATCATGACTCAACGGGTGCGTTTACAGGCGTCGCAATTGTGCAATACGATCCAATC
GCTGATTACGAGCTTACAAAAGAAAATTTTGTGGGTGAAAAGAGAAAGGGAATTGAGGCACCAAAGGCATTTTAT
GGTAACAATAACAAATCTCCATTGTTCCATCATTGTAATTTCCCAAAATTGGACCCTGTAAAACTAATCAAAGAG
AAGAAGTTGATCGACAACAAAACTCACAAATTCAACAGAATGTATGGCCCAAACACCAAATTGAAGTACGATTTT
GAAGAAAGACAATGGAAGTATACCAAAGAGTACTTGTGTGAAAAAAAGTACAACCTACTATATTTTACTGAACAA
TACAAAAATTACGGACAGGGATACTCCCAGGAGCGTATTTGTAAGTTTTCTGACCGATTTTTGAAATTTTTGAGT
GACAATCCAATTAGGATCGAGGGCTAA

Downstream

tttctatcgtgcttcaaaacgatttccactggtatgccagtatgatcgcgacaattttga
tgacttcaaacagggaacacaatttcaccattctgtcatactgtgggcttaggtatatag
ttaacttaggagcaacgaatgcaggaagtagatagttcactaccacaagtgatcacgtat
ccgtgtgaatacgttactcaacaaattcagggtaattacttctctttaatttttctgata
tgcgcttaaataatggccaacaccaatagtatacccccgtcctatgttagcaatcgtact
cgtttctgttctaatttgttattagtttagagaaattgttatttttttagggagcgagtca
caaaaaaaaaaattaattttctcattaagagttcgttacacgatcctcttactaccaaat
ctcgtctttttttttagatcgttttggttctccaaaacatcttttccaatccatgggtagt
cacagctctacaactggaccgtaagttttttgtgaacattacaaatcgccgtaaagtccc
atcctattaacactattttcacatagtcaactcaatttgcatgacttttatgatggaaat
gaactgtcctaccaacataaatacactaagctgacatcagttatgtccctagatgactgg
tcatctaacacatcaagatggtctcgatctagtgtctactattatagcaatttagtgtta
gataaatttgggaggaaccaagtgggcctttacaattctaaaggaaaatacggacctaaa
aaccggaacgtcaagcaccttattcacccctttcagtatgaaggtacctggaatgtggcg
attctgctgcgaaaaactgcagtcaggggagggatagatgtagacaactcgcatgcattc
tccgagctgggaaaagagactaaagtatcagttggggaactggtactgatcaattgcacc
cctaatgaccccaggtcggaattcccagttgatgtttgttt

AA

MLFGLIRHSRRQLLFLGALVTVIVLIFTLPNTSPIEANGVKSEEGSITPIIPVLESPANSLEKIVDTASEERIGG
ATLEEGHENNKEEQALENAERAKEKEKTEAIAAEEEKLKAAELLRQQETTREKEAAKEDDSKKPNQELVEQDTYL
DDIPDDVEDNIIISEQDRKKIILPSYTPKTDPAYSKRATALKIFYNDFFIKVADSGPNTAPITKKTRKKGKSKLK
GDVSSGDKYEGPVLTEDFLRFMEIYSDEFIDAVSESHSKIVNLMPESFPKGMYQGDGIVIIGGGVYSWYGLLAIR

NLRDGGNTLPVELMLPSDNEYEPQLCEQILPSLNAKCIMLSDIVDQDVLKKLDFKGYQFKALSLLASSFENVLSL
DSDNIPVANVSHLFDHEPFSETGLVSWPDFWRRTTNPRYYEAAGIKIGEYQVRNCLDGFVPESDFVHIGLKDIPL
HDRNGTIPDASTESGQLLVNKNKHAKTLMLMFYYNFYGPGYYYPLLSQGMAGEGDKETFLAAANFFGLPFYQVKA
GPGILGHHDSTGAFTGVAIVQYDPIADYELTKENFVGEKRKGIEAPKAFYGNNNKSPLFHHCNFPKLDPVKLIKE
KKLIDNKTHKFNRMYGPNTKLKYDFEERQWKYTKEYLCEKKYNLLYFTEQYKNYGQGYSQERICKFSDRFLKFLS
DNPIRIEG

SEQ0145
Chr 3, 0370

5' region
gtagtatctagaaaaatgaaacgaagaaccccaaccgtactcgtagaaatcagtaaccag
gttgtagtaatgcttggtcagctccgaataatcactctttcttgagtgttcagcttcctg
atcagtcttaccattccagtgcttgaagtagccatcgaccgctacttcttgggctgaggc
gtcctttgtgacccaagctccgagaccggttttcttagcggcatccttgccgtgcaaagc
ctttgcaaattctctatcttgctcatagttctttggggcaagtggctggctttcttgttc
agttgtagaggtagtcattttcggtaataaacgagtgttgagtagagtaaagttgacctt
ctaaaagctgaatattataggttttcactatctaaacgatgggaacgacaaatctaccag
gcgcgcgtatcaatttttagagtgtaaacgagatgtacgaccgattcaattcgagtgcga
caactgtcaaattgtcacattggacaaaaaacgcgcttgacacgatggtttcaatactga
caggctgttcttatatatactttacgaactgaccgaagcaattgcgccagaaaggcgcga
gatgaaatttcgaaatccatgttaaagccgcatccgatcaagggaccctcatttcagttg
gtttggttccacatgtttactgttaacgagcgaaggcctcttgaacgctctcaacagaaa
aacacgtgctgtcgtagaacacaattgttattcacctaggttgccgtaatcgtctttgta
acgccctcgttggggtaaactttccatggactgtgcacgttgtttagcttgctctctgct
ctactccacatcttctccaccccaattgtcgtctcttacccaatttgtatggaagtctta
acaccatttccacatctcagaaccatccgatcccatataagttgacgtatacggctccac
taatttccatccgcattctgcattcacaataacttcccaat

ORF
ATGTTTGGAAAGAGGAGGCAGGTAAGAAAATTACTTATATGGGTTGTCTTACTGTTGATAGTTTACTTTTTTGGT
TTGCAATTTAGAGCCAAAAACTCAGCTCATCAGTCTTCCATCAGGAGCTTTTATGCGGATAACAAAGAGTTTTTT
GATCGGCAATACAGCAGATATGACGAGTATGATATCATAGACAACATGAATAGCCATAACGAGTTACTACAGGAA
CAGTTTCGCAATGGAAAGCTCGCTGCGGGACTCAGAGGTGTAGCCGAAGAACCTAATTCCGATGAGGTTACAGAT
GACACTGCCATCGAAGAAGACGAGCAAGCTGCAATGATTAATTTCCCGAAGAGATCTCCCCAGAGAGAGAAGAGT
CTAGTTGAGTTACGCAAGTTTTACAAGAACGTGCTTTCGATTATAATCAACAACAAACCTGCTATGCCAATTGAA
AATCCTCGTGATCCCACACCTAACGAAAACGCACTCAAGAGAAAGTTCGGTAAAAGTGGAATCATCAACATTGCT
TTACATGACACCGATCCTTCGCTGCCTATTCTTTCTGAAGCGTATCTCCGGGACTCGTTGCAGTTGAGCCCCTCT
TTTATTGCCTCATTGTCCAAATCCCACAGCGCCGTGGTAAAGGCGTTTCCCCCTTCCTTTCCTGCGAATGCATAT
AATGGGACAGGAATTGTATTTATTGGTGGTCAAAAGTTTTCCTGGTTATCATTACTTTCTATCGAAAACCTTAGG
AAAACAGGATCCAAAGTACCTGTCGAGTTGATAATCCCATTTGCACATGAATATGAACCTCAATTATGTGAAGAG
ATACTACCAAAGTTGAATGCTACTTGCGTTCTTTTGCAGGAGACTGTGGGAATTGATCTGCTTAAATCTGGACAT
CTCAAAGGGTACCAGTTTAAGTCATTAGCGCTTTTGGCCTCCTCTTTTGAGCAGGTTCTACTAGTAGACTCTGAT
AACATCATTGTTGAAAATCCAGACCCCATATTTGATTCAGAGGTTTTTCAACGGACGGGATTAGTGTTGTGGCCT
GATTTCTGGAGGCGCGTTACCCATCCTGATTACTATAAGATTGCTGGCATCAAGTTGGGCTCTGAACGAGTAAGG
CATGTTGTTGACTCCTACACTGACCCCTCGTTGTATACCTCTAGTTCGGAAGATCCCTTTACCGATATTCCTTTG
CATGATAGGGAAGGGGCAATTCCAGATGGTTCCACAGAATCTGGCCAAATACTAATATCCAAGACCAAACATTGC
CAAACGATCCTTCTATCTTTGTATTATAACTTTTTTGGACCAGATTACTACTACCCTCTGTTCACCCAAGGAGCA
AGCGGAGAAGGAGATAAAGAGACCTTTTTAGCTGCTGCTAACTACTATAAACTACCATTTTACAACATCAAGAAA
GGTGTGGACGTCATTGGATATTGGAAGCCTGACCAATCTGCATATCAGGGTTGCGGTATGCTGCAATACGACCCA
ATCGTAGATTATCAAAACTTACAAACATTTTTAAAGACTCATAAAGGCTCAAGGGTCAACAAACTTGAACAGTCG
GAGCTAGACAAACCCGGATTACTATCCAGATTAATACCAAAATTTTTCTTTCGCAAAACTTTTGATGAGCATCAG
CTTCAGAGTCATTTCACCAAGGACAGATCTAAGATCATGTTTATACACTCAAACTTTCCAAAACTAGACCCATTT
GGATTGAAGCTTCACAACTATCTTTTTGTAGATCAAGACACTCATAAACCAAGGATAAGAATGTATGCGGACCAA
ACAGGACTTTCATTTGACTTTGAGCTCCGTCAATGGATTATCATTCATGAATACTTTTGTGAGTACCCAGATTTC
AACTTGAAATATTTGGAGAACGCTAACGTGAAGCCTCAAGATTTATGTATGTTCATCAAGGAAGAGCTAAACTTT
TTACAGAATAACCCAATACAATTGACGTGA

Downstream
cttgctactggtttatcacactaatgacgaaaagaagagaaataaagaacaggtaaatga
ctgagaatgtccttttttattcgaaagacaagaatcaaaacaaaatgaccgtaacacgaa

aaataaatacacataccgaattggaaaccatgattgactattttgaactgccaataaaaa
ttgtacgatataatgaaacagataattacttgtaataagaacattggttatggtgaacca
ttggatgaaacatgttctttaaacgtatggatactttcaaagatcgagaatatcatcttt
aaatgtccccctgtagatatgctatactttaccacccaaaatatgctcgttacaaaaaga
gagccctcagctgtcacttttatatcaagcagttattcgaaattcttgccactaccataa
ctagttgagattagagtagccatcaaaaactaataaacacatttcaaaggaatggacaac
aattgtcgtcccgtaatgtttataatgcaatgtgatgtaacataattgatacatataaca
atacagatcatttgatacaagctagccaggatacaccgaaaagtttcgaaaataaacagg
atacataatgctacagtgtcgtgtctatcaacataacggttttggttcctggaaataggt
tagtagagaacagactcgttgtcaaaaatggcagatccttgtcaagtcaaaaaagatgtg
tcatcctattctttatctttcttttcttcattgatacagagggaagaatacaaatagagt
tgtgcggtgtttctttgtagtggtaggacacaaatgaagaaaaaagtttcatgtggggta
aatgggttctgtactggtgggagcgcctgaaatatgtctcaccactcaacgaataacaat
gttaaaatacagaagcagaccagataacgatgcaaggaatagttcatagcctatcagaac
aacaatagctttccatcagtgcaatccaaagtgttcaatg

AA
MFGKRRQVRKLLIWVVLLLIVYFFGLQFRAKNSAHQSSIRSFYADNKEFFDRQYSRYDEYDIIDNMNSHNELLQE
QFRNGKLAAGLRGVAEEPNSDEVTDDTAIEEDEQAAMINFPKRSPQREKSLVELRKFYKNVLSIIINNKPAMPIE
NPRDPTPNENALKRKFGKSGIINIALHDTDPSLPILSEAYLRDSLQLSPSFIASLSKSHSAVVKAFPPSFPANAY
NGTGIVFIGGQKFSWLSLLSIENLRKTGSKVPVELIIPFAHEYEPQLCEEILPKLNATCVLLQETVGIDLLKSGH
LKGYQFKSLALLASSFEQVLLVDSDNIIVENPDPIFDSEVFQRTGLVLWPDFWRRVTHPDYYKIAGIKLGSERVR
HVVDSYTDPSLYTSSSEDPFTDIPLHDREGAIPDGSTESGQILISKTKHCQTILLSLYYNFFGPDYYYPLFTQGA
SGEGDKETFLAAANYYKLPFYNIKKGVDVIGYWKPDQSAYQGCGMLQYDPIVDYQNLQTFLKTHKGSRVNKLEQS
ELDKPGLLSRLIPKFFFRKTFDEHQLQSHFTKDRSKIMFIHSNFPKLDPFGLKLHNYLFVDQDTHKPRIRMYADQ
TGLSFDFELRQWIIIHEYFCEYPDFNLKYLENANVKPQDLCMFIKEELNFLQNNPIQLT

SEQ0146
Chr 3, 0787

5' region
gactctggctccagaaggagaaagacgaagaaaaaaacgaaaaaatatatgactccccct
ctcatctttccaacagactacaatgacaggagaacataaacgaagttccctcattaagag
gatcacaaacgaaaccaaaatacaaatagctctcagtctcgacggtggacctgtgagtct
ggctcaatcacttttcaaagataaggactattctgcagaacatgcagcccaggcaacatc
atcccagttcatctctgtgaacacaggaataggattcctggaccatatgttacacgcact
tgctaagcacggcggctggtctgtcattatcgaatgtgtaggtgatttgcacattgatga
ccatcattcagcagaagatactggaatcgcattggggatggcattcaaagaagccttggg
ccatgttcgtggtatcaaaagattcgggtccggatttgctccactagacgaagctctcag
tcgggctgttattgatatgtctaacaggccctatgctgttgtcgatctgggtttgaaaag
agagaagattggagacctatcgtgtgagatgattccccatgttttggaaagttttgccca
aggagcccatgtaaccatgcacgtagattgtttgcgaggtttcaacgaccatcatcgtgc
cgagagtgcattcaaagctttggctatagctatcaaagaggccatttcaagcaacggcac
ggacgacattccaagtacgaagggtgttcttttctgagtctggaaggtgtctacatctgt
gaaatccgtatttatttaagtaaaacaatcagtaatataagatcttagttggtttaccac
atagtcggtaccggtcgtgtgaacaatagttcaatgcctccgattgtgccttattgttgt
ggtctgcattttcgcggcgaaatttctacttcagatcggggctgagatgaccttagtact
cacatcaaccagctcgttgaaagttcccacatgaccactca

ORF
ATGTTTAATAGCTTGGCACCCATGAGGTTGAAGAAACTACTTAAGGTGTTTTGTGCCTCAGTAGTGCTGTTAGCG
GCGACATCTGTGGTGTTATTTTTCCACTTTGGAGGTCAGATCATAATCCCCATACCGGAACGCACTGTGACCTTA
AGTACTCCTCCCGCAAACGATACTTGGCAGTTTCAACAGTTCTTCAACGGCTATTTAGACGCCCTGTTAGAGAAT
AACCTGTCGTATCCGATACCAGAAAGGTGGAATCATGAAGTTACAAATGTAAGATTCTTCAATCGCATAGGTGAA
TTGCTCTCGGAGAGTAGGCTACAGGAGCTGATTCATTTTAGTCCTGAGTTCATAGAGGATACCAGTGACAAATTC
GACAATATTGTTGAACAAATTCCAGCAAAATGGCCTTACGAAAACATGTACAGAGGAGATGGATACGTTATTGTT
GGTGGTGGCAGACACACCTTTTTGGCACTGCTGAATATCAACGCTTTGAGAAGAGCAGGCAATAAACTGCCAGTT
GAGGTCGTGTTGCCAACTTACGACGACTATGAGGAAGATTTCTGTGAAAATCATTTTCCACTTTTGAATGCAAGA
TGCGTAATCTTAGAAGAACGATTTGGTGACCAAGTTTATCCCCGGTTACAACTAGGAGGCTACCAGTTTAAAATA
TTTGCGATAGCAGCAAGTTCATTCAAAAACTGCTTTTTGTTAGATTCAGATAATATACCCTTGCGAAAGATGGAT

AAGATATTCTCAAGCGAACTATACAAGAATAAGACAATGATTACTTGGCCAGACTTTTGGCTTCGAAGTACCAGC
CCGCACTATTATCATAACATCACCAAGACCCCTATTGGTGATAAAAGAGTACGATATTTTAACGATTTCTACACG
AATCCAAACGAGTATTACTACGGAGATGAGGATCCCAGATCAGAAATACCGTTTCACGATAGGGAGGGCACGATA
CCTGACTGGACCACAGAAAGTGGACAATTGGTTATAAATAAAGAAGTTCATTTTCCTGCAATTCTTCTAGGTCTT
TTCTACAACTTTAACGGGCCGATGGGCTTCTACCCTCTTCTCTCCCAAGGTGGAGCCGGAGAGGGAGATAAAGAC
ACATTCGTTGCAGCTTCCCATTACTACAATCTCCCGTATTACCAAGTTTATAAGAACTGTGAGATGTTGTATGGC
TGGGTTGATCATGCCAATAGTGGAAGAATTGAACACTCTGCAATCGTCCAGTACAATCCCATCGTGGACTATGAG
AATCTCCAAAGTGTCAAGGCGAAAGCGGAGATTATACTGAAAAACCACGAGCCAGATTCAAGGAAGAAAAGTTCG
AAACCAAAGTCTTACAGCAAAACAAGGCTTTCCACTCATGTCAAGGGTTCAATATACTCATATAGACGGCTCTTC
AGAGACTCGTTCAATAAAGCCAACTCTGATGAAATGTTCCTACATTGTCATACCCCTAAGATTGAACCTTACAGG
ATAATGGAAGATGATTTGACACTTGGGAGGAACAAAGAAGCTAAACAACGATGGTATGGAGGACGAAAGAATAGA
GTGCGGTTTGGATATGATGTGGAACTTTACATCTGGGAGCTAATTGACCAATATATCTGTGACAAGAATATCCAA
TACAAAATTTTTGAAGGTAAGGACAGAGATGCTCTTTGTGGGAGTTTCATGAGGGAACAGTTGGGATTTTTACGC
TCCACGGGGGATTAG

Downstream

catagagtatgtatataatccagtattattgaatattcataaaaaaagaggtcgtacccg
gattcgaaccggggttgttcggatcagaaccgaaagtgataaccactacactatacaacc
aatttttctttagaatttgtattttctgtttttactcaaattagatgactctcttgatgg
atacttcaagttttattcagagttacacggtcttaccatgtagtcagcccttctttaatg
tgcactgtaaactttacctgcactatttatgacaatatcattgcaacaaccagttatttg
ttgattagaaagtataatatcggagtttacagtgtattggtctgtgttttatcatgtttt
gttattctcgcaaatcgttacgctggtgatccatttgtcgtgttcctcaaattaccggag
ttaaagctggaaaatcaaacagccatatcatcattgttgtagaaagtctcctggacgttg
agtaatgacatttacacaccaaaaaagtatattcgtatttcatatcctagggtataggct
aagaactacactcgtcttcgaactcgtccttattagctcagttggtagagcgttcggctt
ttaagtcaacacaaggaaccgaaatgtccagggttcgagcccctgatgaggagttctttt
ttgaatttttaaacttcaattacgggccgtctattatattattccctatttacccttct
cttgcccatcttagcctcaattagaagattacttcctttcttcaccgaccttgagatttg
ttcttcgtcactcgagtcgctggcatcattctgatttcgttgtgagttctgtttctccaa
gttttcttgacgaaacctccttctatcgtcgtttctcatcttgtttttcaaagattgtat
caccttactgtcatctttggaaactctgtggcgattgttgattctggccaatttgtcatc
cacttgacttttcatcttctgtttttggtcatggtagaaa

AA

MFNSLAPMRLKKLLKVFCASVVLLAATSVVLFFHFGGQIIIPIPERTVTLSTPPANDTWQFQQFFNGYLDALLEN
NLSYPIPERWNHEVTNVRFFNRIGELLSESRLQELIHFSPEFIEDTSDKFDNIVEQIPAKWPYENMYRGDGYVIV
GGGRHTFLALLNINALRRAGNKLPVEVVLPTYDDYEEDFCENHFPLLNARCVILEERFGDQVYPRLQLGGYQFKI
FAIAASSFKNCFLLDSDNIPLRKMDKIFSSELYKNKTMITWPDFWLRSTSPHYYHNITKTPIGDKRVRYFNDFYT
NPNEYYYGDEDPRSEIPFHDREGTIPDWTTESGQLVINKEVHFPAILLGLFYNFNGPMGFYPLLSQGGAGEGDKD
TFVAASHYYNLPYYQVYKNCEMLYGWVDHANSGRIEHSAIVQYNPIVDYENLQSVKAKAEIILKNHEPDSRKKSS
KPKSYSKTRLSTHVKGSIYSYRRLFRDSFNKANSDEMFLHCHTPKIEPYRIMEDDLTLGRNKEAKQRWYGGRKNR
VRFGYDVELYIWELIDQYICDKNIQYKIFEGKDRDALCGSFMREQLGFLRSTGD

SEQ0147

P. pastoris homolog of *S. cerevisiae* **KTR4 (YBR199W)**

Putative mannosyltransferase involved in protein glycosylation; member of the KRE2/MNT1 mannosyltransferase family

S. cerevisiae null mutant: viable, decreased budding index

Chr 1-4, 0050

5' region

tttattgcgccgttgccaagaggcgagagtggaaactatatttgatatcatgtccatgga
agacgatgagcgagacaatttgttgaagctatccaatgcgcagctccagaaagtagctga
attcgtgaacaaattccccaatgttgatattgattacgagttggatatcactgaaggtac
tataatagttgacgaggagagagaaataatagttacacttaccagagatgaacaacctga
ggatctcactgttatttcttcggtatatccctatacaaaaacagagaactggtggctcgt
tattggttgctcccattcaaaagagctttatggtatcaaacgtacaagaatttctaaaca

gcaagagcaagttaaagttactttctcggttccttccccaggatcccacgagataacact
ctggtgcatgtgtgattcttatatggatgcagataaggaagtaagcttcgagcttagagt
ggaagataatcaaatccgtatatgaataaggattgccaatacttttatttttcgctctat
ttatttaattcatgaaaattcgtaaattgaacccagctatttcagtagatatatatgaat
aatttgccggcaaagttttgcgttcgctgcgtgagttcaacttgagaatcttggaactag
ctgcaaactaaaaatcctggttcagaaaaaaataatgcgtatgcccggagtcgaaccggg
ggcccgacgatggcaacgtcgaattttaccgctaaactacatacgctttttgtatacaaa
agttcgttcagtatatatcatatattcataacatattagccaatttccctgaaacaggaa
aatgagtgaaaaggcaatactgcggtttattcggtatttgtgccgctgctaaagcttttg
agtgtctcatcccgacccgaaggcgataacataccaagaaagataattcaaacacgcgaa
cgtgctccttatttgacaaggtgcaacttgttattcttctg

ORF
ATGATGATCTCCCTTACGAAAAGGTTTACCAAGTTGGCAATCTTTGGAAGTTTATCTTTCATACTTACTACAGCA
GGATTATGGCTGTACTGGGATGCCATCCAGTATATGATGACATCTGGCAAAATTCCAACTTTAGACTTTCAATTC
GAAGACTTTATGAATAGACATGATGATATTGTTGATGATATGATGAAGAAATATGACAAAATCATGAAAGCGGAG
GTCAAAGAACCTAATGTTGGCAATTTGGTTTATGCACCAGAAAGTCTGGTTGATTATGGACGTGAGAATGCCACA
TTGCTAATGTTGGTTAGAAACAAAGAACTCAGGACTGCTCTACAGGCAATTGAAACGGTGGAAAGCCAATTTAAT
CACAAATTTCAGTATCCATACGTTTTTTTGAATGATAAAGAATTTACTGACAAGTTCAAAAGCACCATAACTGAA
AAAGTTTCAGGGCAAGTATTCTTTGAAACGATTGACAAAGTAACCTGGGACAGACCTGACTGGATAGATTCTGCC
AAGGAGTCAGAGAGAATTAAGGTCATGAGAAAATACAATGTTGGCTACGCTGATAAGTTGTCCTACCATAACATG
TGTCGATACTATAGTCGGGGTTTTTATAACCATCCTAGGTTGCAACAATTCAAATACTACTGGAGATTTGAACCT
GGCACACATTATCATACTTCCATCGATTATGATGTATTCAAATTTATGTCAGCAAATGATAAGACGTATGGGTTT
GTTATTAGTTTGTACGATACAGAACGATCGATCGAGACATTATGGCCTGAGACATTGAAGTTTATTGAACAGAAT
CCTCAATTTGTCAACAAAAACGCTGCTTGGGACTGGTTAACAGAAAAGAAACAAAATCCTCAAAAAACTAGAATT
GCGAATGGCTACTCTACCTGTCATTTCTGGTCCAACTTCGAAATAGGCGATATGGACTTTTTTCGAAGTGAAGCG
TACACCAAATGGGTGAACCACCTCGATGCAACAGGAGGATTCTATTACGAAAGATGGGGAGATGCACCAGTGCAT
TCAATAGGGGCGACTCTATTTCAAGACAAATCCAAAGTTCACTGGTTCCGAGATATCGGTTATTACCATGCTCCC
TATTATCAATGTCCCAATAGCCCTCAGTCGGATGGAAAATGCGAGGTTGGCAAATTCAGTTTCCCAAACCTCAGT
GATCAAAATTGTCTGATCAATTGGATTGAAATGGTAGCGGACAATGAACTAAGCATGTATTAA

Downstream
gacgaacacggaggtgctttctcaatcatttagtattgttcagcttgatatttacagaag
ttactaataccacgtgacaatacgagaaattgccaggcaattcttggttttcttctcacc
ttttagttcaagtaaactactattctccttatgtcccctgcaatcaacatcgccattatc
ggagcaggtattgtcggttctgcgttcattaatcagttgaggagcgtatcgtctttgttc
gactttaacgttatttatttggctaaaagcaaaagtgccttgatctcttatgatttcaaa
ccgttgaatctagccaattggaaagcagacttggacaatagctcattaaatcccttgacg
gtggaagaacttttgaaattcttgaaaaattcacctttaccagcaattttagttgataac
acttctagtgctgacgttgcgtttgcgtacccccaatttgtctctaacggtatttcagtt
gcaactcctaacaagaaagcatttccggggacctaaatttatggagtgaaatattcaaa
gcttctggagaacccaatggtgggctagtatatcacgaggctactgtcggagccggtttg
ccggttattagtactatcaatgatctggtcaacactggagataagattgaaaagattgaa
ggaatttttccggatctctatcctatatcttcaatcgcttctctactacaaagccaaat
gatattaagttttctgaaattgtgtccaaagccaaggagctgggttacactgagccagac
cctagagacgacctctccgggttggatgttgctagaaaagtcaccattttagccaggtta
tctggctttaatgtgaaatctcctgattcattccctgttgactctttagttccccaggag
cttgaaactttggagacgagcgatgaatttctttcaagattgcatgagtttgactcccaa
gtggggcacctaaaggaacaagctgcttccgagggcaaggt

AA
MMISLTKRFTKLAIFGSLSFILTTAGLWLYWDAIQYMMTSGKIPTLDFQFEDFMNRHDDIVDDMMKKYDKIMKAE
VKEPNVGNLVYAPESLVDYGRENATLLMLVRNKELRTALQAIETVESQFNHKFQYPYVFLNDKEFTDKFKSTITE
KVSGQVFFETIDKVTWDRPDWIDSAKESERIKVMRKYNVGYADKLSYHNMCRYYSRGFYNHPRLQQFKYYWRFEP
GTHYHTSIDYDVFKFMSANDKTYGFVISLYDTERSIETLWPETLKFIEQNPQFVNKNAAWDWLTEKKQNPQKTRI
ANGYSTCHFWSNFEIGDMDFFRSEAYTKWVNHLDATGGFYYERWGDAPVHSIGATLFQDKSKVHWFRDIGYYHAP
YYQCPNSPQSDGKCEVGKFSFPNLSDQNCLINWIEMVADNELSMY

SEQ0148
P. pastoris homolog of *S. cerevisiae* **KTR2 (YKR061W)**

Mannosyltransferase involved in N-linked protein glycosylation; member of the KRE2/MNT1 mannosyltransferase family
S.cerevisiae null mutant: decreased competitive fitness, viable

Chr 2-2, 0105

5' region
gaattgacaggcccatcgcgtctgctgatgtaatcataacagagaaaaaaaagtttagct
caacttgatacaatatcaggcgtctcaagatcaaccaccgatgagctttgagttatcagc
ttccgatattaaacagactcagaagtcaatagctgcactgatagcccacaccaagaaaac
caaagattcaaaactcaatattaatctggttattggcctggaagaagaccagactttacc
agaaaatgacaatactccaagaatagtccctgtcaggaatcgactagaaaagagtactga
aaaatcaatcctgttgatcacaaaagactcttctgagccatatgtccatgcgcttaaaga
aaagggagcacctacagaggatacgttcgcaagaataatttcctatcacaagctcaaatc
cctatcaggcaaaccacaggaaatgaagaaattataccacgaatacgatcttattctagc
cgactataggatttatccctttctcagaagaacactagggtccaaattttatagctcgaa
ggagaagattcccttcatgattcaaatggccaaaccttccaagaaaatcgattattcaag
aaccgaggggagccaccagaagacaaacgatgcgatcctatatatgtcaagaaacaggt
catttcaatcgttaagaacacttttttcatcccttcaaacaaatctcatttcatcagtgt
agttgttggagacactgacaaggatccaaaaaaactcgctgaaaatgtatcagacatact
gtctttcttgaccgatattaagcgcaaaccaatcggaggcttgctagctgctgacggaat
caaatcacttcacgtgaggacgtctgaaagcattcccttaccagtgagagaaaatgatga
aatatagaactcagaaaccatgtattatgtaatcatatattagattgtccaaggcttatc
acgctaagttatttcgttgcacctatcgatggcgactactt

ORF
ATGAAAGTTGTCTGGTTAGCATGTTTCATAATACTGGCAGCTATATGGTATAAAGACTACCAAAGTTTAAGAGGA
TTCATGGACGACAGAGTGTCGAAGACCTTGCCAATTAATTTTAACGCACTAAAGTTATCTACCAATTCTTATATA
CCCGTGGATGAGCACCTTATAAAGCCTAACAGAGAACCGAACCCAAAATTTGTTAAAGAAAATGCTACTTTGCTG
ATGTTATGCCGAAATTGGGAACTAGAAGAAGTACTGCAATCAATGAGATCACTCGAAGATAGGTTCAATGGTAGG
TATCAGTATACTTGGACATTCTTGAATGATGTGCCCTTTGAGAAACAGTTCATACAGGAAACAACTTTGATGGCC
TCGGGGAAAACTCAATATGCGTTAATTTCTAGCACAGATTGGAATAGACCTTCTTTTATAAACGAAACGAGGTTC
GAACAGAATTTAATCCAATCTGAGAAGGATGATATCATCTACGGTGGAAGTCCATCTTATCGAAATATGTGCCGT
TTCAATTCAGGATTCTTTTACAAACAAAAGATACTCGACCAATATGATTATTATTTTAGAGTGGAACCCGGAGTG
GAGTACTTTTGTGATTTAGAAGAGGATCCATTCAGATATATGAGACTTCACGATAAAAAGTATGGGTTTGTCATT
TCACTTTACGAGTATGAAAACACCATTCCAACGTTATGGCAGACCGTTGAAAAATTCATAGAAAATCATCCAGAG
TACATCCACCCAAACAACTCCTACGAATTTTTAACTGACAAGGAAGTTGTTGGCCCTTTAGGCTTAGTTGCATTA
ACTGAACAGACTTATAATCTGTGCCATTTTTGGTCTAATTTTGAAATTGGAGATTTGAACTTTTTTAGAAGTGAA
AAGTACGAGGCTTTTTTCCAGTTTCTGGATCAGGCTGGAGGATTTTATTATGAAAGATGGGGAGATGCCCCAGTG
CATTCGATAGCAGTTGGCATTTTGTTAGATAAGAGACAGATACACCATTTTGAGAATATTGGTTATTACCATCTT
CCATTCAGCACATGTCCCCAGTCATATTGGTCATACAAGTGCAATAGATGCATTTGCAAGCGGAATGAATCCATC
GATTTAGTTCCACATAGTTGTCTTAGCAAATGGTGGAAATATGGAGGTAAAACATTCCTCCAATAG

Downstream
aaaaatgatcatacatagtatatagtacagggtctacttttatttatagggatcttctaa
cctgttcagttcatctaaaatcctttgtaagtttacctctttttcgcagccaaacttgg
aattaatgttatagcttcgtcaacatcttcacaccccagagaacttaattgggcaatttc
aaaaggatgtaacacagagcaatctgaagaattatgcaatagctgatcaacggcagtaca
agtctcttcatctttgaatcttgcaaacgtgttcagatagtctaatgttttcttcactac
tccgttggctacagcacctgaagtcacttttgcaagctcgtcatcgtcgatttccccatt
gctggactcaatatcaactccgccattacgtgcccgtgatctagctttcagagcttccct
aattaacaaccttgattctgataaactaagagctattaagtcatgttcattaccatcatg
atcatactgtttcaaggcaaactctggccctagccttaataatgtagcgttctcctcatc
gtctacctgctgctttgctctcctacgacgagctcccaccgtggatgttgaaacattcat
agtgacgtacgagtgtttagcaagggctaggaagaaactgaaaaagtactcaaacaaact
gtttttgtcgatttccgcagtaacagctagaaggaatagcgaactccctctggcccacg
aactctctgggcaaagtttgattattaaccaaaagcttcactaatgtagcgaattgcttt
cccaaaaccttccttctaccacagattgccgtttccttccagaaatccgcctagcatgag
ccgcccgcggatagttacctttcgtgttccgtgccatatcacctccccacgcttatttac
agcaaacacacgaatatttagctcacacatttccatataccatcatgagttttaatgaca

ggcaaggatggaagcaaagggcgcagagtagcaatcaagga

AA
MKVVWLACFIILAAIWYKDYQSLRGFMDDRVSKTLPINFNALKLSTNSYIPVDEHLIKPNREPNPKFVKENATLL
MLCRNWELEEVLQSMRSLEDRFNGRYQYTWTFLNDVPFEKQFIQETTLMASGKTQYALISSTDWNRPSFINETRF
EQNLIQSEKDDIIYGGSPSYRNMCRFNSGFFYKQKILDQYDYYFRVEPGVEYFCDLEEDPFRYMRLHDKKYGFVI
SLYEYENTIPTLWQTVEKFIENHPEYIHPNNSYEFLTDKEVVGPLGLVALTEQTYNLCHFWSNFEIGDLNFFRSE
KYEAFFQFLDQAGGFYYERWGDAPVHSIAVGILLDKRQIHHFENIGYYHLPFSTCPQSYWSYKCNRCICKRNESI
DLVPHSCLSKWWKYGGKTFLQ

SEQ0149
P. pastoris homolog of *S. cerevisiae* **KTR5 (YNL029C)**
Putative mannosyltransferase involved in protein glycosylation; member of the KRE2/MNT1 mannosyltransferase family
S.cerevisiae null mutant: viable

Chr 2-2, 0201

5' region
aagaaggacaaaaaaagacttcaaatccagaacaaattgataaaaatggaggaaaagttt
caagaagaaaaggatttcatttatagaagcaatcttcatgatttgcagtatgaaaggagt
actctccacgccgacgtcaatgaggcatacttaacagagatccgtgatcttcaagaaatc
agagatgaagagctggtaaggctacgtctttgggaagattatcaagtgagttgcgtcacg
agacagtttaatgaaggctatgaaaaggccaacaaagagtatgacagattggtcactatt
ttcaaaacaaaactgaaagagagaattgagcacaaaataaaacaacttaaagaagataag
gtactaatggactttagtagccatttgccaaagacgaacactcgttctcataataacgtc
aatggtggaatatttggccttacccccggtaacatttctgatagagaggtcaattctaac
ggctattactcttccacggatagaagatcacgccgtaggaagtttgatcataccacagat
aacggtgaagactcatacgattcaaatgataacacttctggatctgcctcgaacatttcc
agaagatccaagatgaagaaaactggtggtaatacatctgcaagatcttcatctgacact
gacgacactttcttgaccgacgatcatgctttgaaattgttatttggttcagattacgag
aagccgaaagataaaccctctactagacattcttcaaaaagctttaatggagtttcaagt
cttaaaccagaggaattaaacgaagacctggagatcctaaggagtgctatccacactata
tcgcatacaaagaagatggctagtaagtaatgaagttacacctgattctctgtatgtctt
ccttttttttttttttttttttgtactctgtttccaatacctcgctgttagggtgtttga
cgaaagatgaatgttccacaaatacaaaacctcaggcaata

ORF
ATGTCATTCCGATTGGGGTATATACAGGCCATTGTACTAGGCCTAGTTCTACTCTCTGTTTGTTGGACGATTGTA
ATTAGGCCTGACCCATCGTCTGCTATTGATTTAGCCTCCCCTGTCACAATCGATTTAGAAAACTCACTTACCAAT
TTGAAAAGTTTTCCTATTTCCTCCCGTAGGATATCATCAAATATTGATCATGTGTTTCAAACTGGGTGTAGGAAT
GTTTTCAAGAACAAAAAGAAAGCCAACGCTGCCCTCGTTGTTTTGGCAAGAAATTCCGAACTGGAAGGTGTTCAA
AAGTCAATGTTTTCTATGGAGAGACATTTCAATCAATGGTTTAATTATCCATGGATATTTTTAAACGACGAAGAA
TTCACAGAGTCATTTAAGGATGGTGTGATGAATATGACTTCCAGTGGTGTTTCCTTCGGCGTAATATCTAAGCCA
GACTGGAATTTTTCAGAAGAAAAAGACAGGGGTAGTACAGAGTTTTTGAGGTTCAATGAATTCATCCAGAATCAA
GGGGATAGAGGCATTATGTACGGAGCGTTACCCTCCTATCACAAGATGTGTAGGTTTTATTCAGGATATTTCTTT
AAACATCCCTTAGTTGCCAAATTGAGCTGGTATTGGAGAGTTGAACCGGATGTTGAGTTTTTTTGTGATTTGACC
TATGATCCGTTTTTGGAGATGGAGGCGTCAGGTAAAAAATACGGCTTTGCAGTGATTATAAAGGAGTTGTCTAAT
ACTGTACCAAATCTGTTTCGCCATACTCAGAGTTTCATTGAGAAATATGGTATCTCAGTGGACGAGAAAGCATGG
TCGATTTTTACTAACCGGAGGTCTTTCGGAGAAAAAGAATCTATGAAGCTAATTGACAAGATACGAATTAATCAT
TTGCTTTCTAATTTCAGTGGAGGTATTGGCACCAGGTTATTGTCCAGTCTCTCTAGAATGAACCTTCCTACTTCG
TTCTCTTCGAAAAAGCCTTTTTTCTACGGCGAAGAGTATAATTTGTGTCATTTTTGGTCAAATTTTGAGATTGCG
AGTACCGATTTGTTTTCATCACCAGAATACGAATCGTACTTTCAGTTTCTGGAAGAAAAGAAGGGCTTTTATCAG
GAAAGATGGGGGGATGCCCCTGTCCATAGTTTGGCTGTAGCGATGTTTCTCAACATCAGTGAAATCCATTACTTC
AGAGATATTGGTTACAGGCATTCCAACCTAGTACACTGCCCAAAGAATGCTCCTGATGAATTACAACTTCCTTAC
GTTCCAGCATCACCAGAGTATGCTTCCAGCGCCAAACCTGACAAACCACCGAGAGTTTCGGTTAGGGATGTTTTT
AGGAGCGGTCGGCAAACGGAAGGAGTCAACAACTTGAATCGTGGATCTGGATGTAGATGCAACTGCCCGAAAAAA
TACAAAGAATTAGAAGATAGTCCCAGCTGCTGTATTGGCAGATGGATGGTATTAACTAACGATAAGTATAAAGGC

GAAAAATACCTAGATAAATACTCAATGGCCGAAGAGGTCAAGCAGACTCTAAGCAAGGGTGAAAAACTAAACGTT
AAAGAGATATTGAAGAGGCACCACAAGTATCCAACATGA

Downstream
gaagagaacccagaaagagtcaaaatacaaactcgaaaacataccgattggcagaatacc
agccatacattacatttgatactaatatacataaaatacactaacctccgaaaccataca
aggtacgtccttgtcttttcaaagcgtagacgacatccagagaagtaacagtctttctct
tggcatgctcagtgtaagtaacagcatctctaatgacgttctctaagaaagtcttcagca
cagctctgacttcttcgtagatcaaagcggaaatacgcttaacaccacctcttctggcca
atcttctgatagctggctttgtgataccttgaatgttgtctctaagaatctttctgtgac
gcttagcaccggattttcctagacctttccacctttcctctaccagacatatttattg
attatttgtttatgggtgagtctagaaaaggacgcactcgtcttgtatttatagatgaaa
agagttaaggtggacaatgcagtgccaaaccgtaatgttgatacgacacggtacgcgatg
aactgagccgtactcctagaggcagagcgtgcgtaaatttgaaactgtaagaacatgtcc
catttctatgctaccagaactcataataaacttgcatcccatttcagtcgggagtgtcgc
agtgcgaaatactgttggcactgatggcaaaaagtgccaaatcgttatctatgggagact
cttataaatatccagatcctccccctcctgcttcttcttgtgtctatcgtagtaaaaatg
gctagaactaaacaaacagcaagaaaatccactggtggtaaagccccaagaaagcaattg
gcttccaaggctgccagaaaatccgctccatctgctgctggaggtgtcaagaagccccac
agatataagccaggtaccgttgctctgagagaaatcagaagattccaaaagtctactgag
ttgcttatcagaaagctgcctttccaaagattggtcagaga
AA
MSFRLGYIQAIVLGLVLLSVCWTIVIRPDPSSAIDLASPVTIDLENSLTNLKSFPISSRRISSNIDHVFQTGCRN
VFKNKKKANAALVVLARNSELEGVQKSMFSMERHFNQWFNYPWIFLNDEEFTESFKDGVMNMTSSGVSFGVISKP
DWNFSEEKDRGSTEFLRFNEFIQNQGDRGIMYGALPSYHKMCRFYSGYFFKHPLVAKLSWYWRVEPDVEFFCDLT
YDPFLEMEASGKKYGFAVIIKELSNTVPNLFRHTQSFIEKYGISVDEKAWSIFTNRRSFGEKESMKLIDKIRINH
LLSNFSGGIGTRLLSSLSRMNLPTSFSSKKPFFYGEEYNLCHFWSNFEIASTDLFSSPEYESYFQFLEEKKGFYQ
ERWGDAPVHSLAVAMFLNISEIHYFRDIGYRHSNLVHCPKNAPDELQLPYVPASPEYASSAKPDKPPRVSVRDVF
RSGRQTEGVNNLNRGSGCRCNCPKKYKELEDSPSCCIGRWMVLTNDKYKGEKYLDKYSMAEEVKQTLSKGEKLNV
KEILKRHHKYPT

SEQ0150
P. pastoris homolog of *S. cerevisiae* **KRE2 (YDR483W)**
Alpha1,2-mannosyltransferase of the Golgi involved in protein mannosylation
S.cerevisiae null mutant: viable; increased heat sensitivity; abnormal protein N-glycosylation; decreased resistance to sulfanilamide

Chr 3, 0215

5' region
tcacgtattcaagcaacgatggatcttctttatctctcagaaaatgccaaaattgttcgt
tatcccaaatacgaagattccaacctatgatattactgtacgtttctagttgattgcgct
ctgtagctgcaactttcatttttctgccctcacatagatttccttcagaagatattcaa
aaacttcgttttggtgttgatctagggcgatgattaatggggatatggtgtccagcacca
tgtatgaacgattatcgaacgtcgtcattttgcaatgacgtgaagcttttgtcaagattt
tagcatatttggcaagagtcactgtgttgatattatcctgcttattgcctagcagggcag
taagtgcaaattctaagcttcgaactcctccaaatagactaatcaatttttttcaattgtt
caattatattggctttgttagacagttctaaatccaactggaaaagcaatggataatctt
cgggattttcaatagaagagattctgaatctatctatgatctctatcaggagggatagac
tcatgaatttgtacttgcagacctctgcgtctaataccaaatgaaattcctcactagaat
ctgaatcacatttttggcttataaaactcttggagattactttgtagattaaaaaatccc
tttctttttttggtgcagagtttaagttgagaacactgtagaaatgacgattcagggcgt
agagcggagccaacttactgaaggaaaatacttcatagattagctctggaggcaaatctt
caaaacctatctttcttgccttcttactgggtacctttctaagatatttcttgattcgtt
tatgagttcttccagggatgaggaaccccattagggacgaagtgtattaaggttttttc
ttcctagtgtttttttttttttgtctccaagaaccagcgaccgaggggggctaagctatt
taacatttctccaggctcgtctttcttagctatctagtaaa

ORF

ATGGTACACATAGGGTTCAGAAGCTTGAAAGCGGTGTTCATTTTGGCCCTTTCGTCATTGATTCTGTACGGGATT
GTCACGACCTTTGACGGGAGTAGGGCAAGCCGTTATCAACCACCTTACGTGAACCACTCACAGGATCCACTTTAT
CATTCTGGGAATAGCTACAACAGAGAAAACGCCACTTTTGTCACTCTCTGTAGAAATGAAGACCTTTACTCAATC
ATCCAATCAATAAAGAAAGTAGAGGACAGATTCAATAACAAATTTGCTTATGACTGGGTGTTTTTGAATGAAGTT
CCTTTCACCGATGAGTTCAAGGAGAGAACGAGCGTGTTAATCAGTGGTCAAGCAAAGTACGGATTGATCCCCAAA
GAGCATTGGTCGTATCCTGATTACATTGATCAAGAAAGAGCTGCAGAGTCAAGAAGGCAGTTAGAAGACCAGCAC
GTAGTATATGGAGGTCTGGAAAGCTACAGACACATGTGTCGATTCAACTCTGGATTCTTCTACAAACACCCGCTG
ATGCTGGACTATCGATATTACTGGAGAGTGGAGCCTGAAATCGAAATTTTATGTGACGTTGAGACAGACTTGTTC
AGATACATGCGAGAAAACAATAAGACGTATGGCTTCACTATATCAATCCATGAGTTTGAGAAGACTATCCCCACA
CTTTGGGAGACTACTAAAGAGTTCATGAAACAAAATCCCAGCTACATTGCTGAAAACAACCTGATGAATTTTATC
TCCGATGATAACGGTAAGACTTATAACTTGTGCCATTTCTGGTCGAATTTTGAAGTGGCAGACATGGATTTTTGG
AGATCTGACGTTTACGAGAAGTATTTCAAATTTTTGGACGATACCGGGAAATTTTTCTATGAAAGATGGGGAGAT
GCTCCTGTTCATTCCCTTGCTGTCTCGCTATTTCTACCGAAAGAAAAAGTCCATTTTTTCAATGAAGTGGGCTAC
AAACATAGTGTTTACTCCATGTGTCCAATTGACAAAGACATATGGAAGAACCGTAAATGCTACTGTGATCCCAAC
ACGGATTTCACTTTTAGGGGATATTCATGTGGAAGACAATACTACAAAGCTACTGGTCTGACCAGACCTTCAAAC
TGGAAAGATTATGACTAA

Downstream

aattgaaagaaggatacaataaaaaaaaagaaactcaaaaccgtttttcatttggtctact
taaaaaacgcctgcagtatataatatttaaaaagaacacctggtacccatcactatgcgt
ggaacatgtaccgtacccaaagacaatagaaacatctccgcgacatctacctattcctac
atatgtgagtgtgaaattccatggaattagtgcgcctggccaatcttgtcaacgtcaacc
acccttcgagcaaagcaatatatatcgcgttccacttttcttccttctctcaactacca
gaccagacaggacaacggtacaaatggcaggtgcaactaggatcaattcacgagtagttc
ggtttgctattttcgcatcaatcctggtactgttaggattcatcctatcaagagggtctg
ccacatcgtactctcttccctcagggttgactagtgacacgtcacaatcaactggctcat
ctcccaaatctgagtctaaaccttcttcccaaggcagcagtggtgcaactgaactgaaga
agacttataccacggacggaaaggaaaaggctactttcgtttctcttgctagaaatagtg
acgtatggagtttggctagctctatcagacacgtcgaagatagattcaaccacaagttcc
actatgattgggtcttcttgaatgacgaggaattcagtgacgagttcaagcgtgtcacct
ctgctttaacttctggaaaagctaaatacggtctgattccaaaggaacattggtcgttcc
cggagtggatcgacaaggagcgtgctgctaagaccaggaaggaaatggctgctaagaagg
tcatctacggtgactccatttcttacagacatatgtgtcgttttgaatcgggtttcttct
ttagacacgaattgatgcaagagtatgaatggtactggcgtgtggagcctgatatcaaga
tttactgtgatatcgactacgatgtgttcaagttcatgaag

AA

MVHIGFRSLKAVFILALSSLILYGIVTTFDGSRASRYQPPYVNHSQDPLYHSGNSYNRENATFVTLCRNEDLYSI
IQSIKKVEDRFNNKFAYDWVFLNEVPFTDEFKERTSVLISGQAKYGLIPKEHWSYPDYIDQERAAESRRQLEDQH
VVYGGLESYRHMCRFNSGFFYKHPLMLDYRYYWRVEPEIEILCDVETDLFRYMRENNKTYGFTISIHEFEKTIPT
LWETTKEFMKQNPSYIAENNLMNFISDDNGKTYNLCHFWSNFEVADMDFWRSDVYEKYFKFLDDTGKFFYERWGD
APVHSLAVSLFLPKEKVHFFNEVGYKHSVYSMCPIDKDIWKNRKCYCDPNTDFTFRGYSCGRQYYKATGLTRPSN
WKDYD

SEQ0151
P. pastoris homolog of P. stipitis CBS 6054
A-1,2-mannosyltransferase
Chr 3, 1162

5' region

ctggaaagctacagacacatgtgtcgattcaactctggattcttctacaaacacccgctg
atgctggactatcgatattactggagagtggagcctgaaatcgaaattttatgtgacgtt
gagacagacttgttcagatacatgcgagaaaacaataagacgtatggcttcactatatca
atccatgagtttgagaagactatccccacactttgggagactactaaagagttcatgaaa
caaaatcccagctacattgctgaaaacaacctgatgaattttatctccgatgataacggt
aagacttataacttgtgccatttctggtcgaattttgaagtggcagacatggatttttgg
agatctgacgtttacgagaagtatttcaaatttttggacgataccgggaaatttttctat
gaaagatggggagatgctcctgttcattcccttgctgtctcgctatttctaccgaaagaa
aaagtccatttttttcaatgaagtgggctacaaacatagtgtttactccatgtgtccaatt

gacaaagacatatggaagaaccgtaaatgctactgtgatcccaacacggatttcactttt
aggggatattcatgtggaagacaatactacaaagctactggtctgaccagaccttcaaac
tggaaagattatgactaaaattgaaagaaggatacaataaaaaaaagaaactcaaaaccg
tttttcatttggtctacttaaaaaacgcctgcagtatataatatttaaaaagaacacctg
gtacccatcactatgcgtggaacatgtaccgtacccaaagacaatagaaacatctccgcg
acatctacctattcctacatatgtgagtgtgaaattccatggaattagtgcgcctggcca
atcttgtcaacgtcaaccacccctttcgagcaaagcaatatatatcgcgttccacttttct
tccttctctcaactaccagaccagacaggacaacggtacaa

ORF
ATGGCAGGTGCAACTAGGATCAATTCACGAGTAGTTCGGTTTGCTATTTTCGCATCAATCCTGGTACTGTTAGGA
TTCATCCTATCAAGAGGGTCTGCCACATCGTACTCTCTTCCCTCAGGGTTGACTAGTGACACGTCACAATCAACT
GGCTCATCTCCCAAATCTGAGTCTAAACCTTCTTCCCAAGGCAGCAGTGGTGCAACTGAACTGAAGAAGACTTAT
ACCACGGACGGAAAGGAAAAGGCTACTTTCGTTTCTCTTGCTAGAAATAGTGACGTATGGAGTTTGGCTAGCTCT
ATCAGACACGTCGAAGATAGATTCAACCACAAGTTCCACTATGATTGGGTCTTCTTGAATGACGAGGAATTCAGT
GACGAGTTCAAGCGTGTCACCTCTGCTTTAACTTCTGGAAAAGCTAAATACGGTCTGATTCCAAAGGAACATTGG
TCGTTCCCGGAGTGGATCGACAAGGAGCGTGCTGCTAAGACCAGGAAGGAAATGGCTGCTAAGAAGGTCATCTAC
GGTGACTCCATTTCTTACAGACATATGTGTCGTTTTGAATCGGGTTTCTTCTTTAGACACGAATTGATGCAAGAG
TATGAATGGTACTGGCGTGTGGAGCCTGATATCAAGATTTACTGTGATATCGACTACGATGTGTTCAAGTTCATG
AAGGATAATAATAAGATGTACGGTTTTACTGTTTCTTTGCCTGAATACGTTGCTACCATTGAGACTTTATGGGAC
ACTACTCGTGCATTCATTAAGGAGAACCCTCAATACTTACCTGAGGATAACATGATGGACTTTATTTCTGATGAC
GATGGCTTGTCTTATAATGGATGCCATTTCTGGTCCAATTTTGAAGTTGGTTCCCTCAGCTTATGGCGTTCTGAG
GCTTACCTCAAGTATTTTGACCACTTGGATAAAGCTGGTGGATTCTTCTATGAGCGTTGGGGTGATGCTCCAGTA
CACTCAATTGCGGCTGCTTTGTTTTTACATCGTGATCAAATTCACTTCTTTGATGACGTTGGATACTTCCACAAT
CCATTCAACAATTGTCCAGTAGACGCCGATCTCAGAGAGGAAAGGAGATGTATGTGTAATCCTAAGGACGATTTC
ACTTGGAAGGGATACTCTTGTGTGCCAGAGTTCTTCACTGTGAATAACATGAAGAGGCCCAAGGGTTGGGAAGCT
TTCAGTGGGTAA

Downstream
ttttagttttaactatggtcttacataggcgcgatgtgtgctatctagaaaaatgtatag
tcgaaacaatcttgtactcgatttcagttctataaactttcaaaagattcttgagggagt
ttacgtttcatccttgctaacgttgcatcagtttgcttatgcagatcatcaaacttccat
ggctctgtcatcaagaagagcagaggaagttcgaggtttacacaatttccgtttctgctg
atggtaaaagactagcttctggaggtttagatggtaagatcaaaatatggtctatggact
caatttatcaatacaagaggagcgataaagagaatgatcttaggtacttacaaggttcag
ttcagtcggatgaatcaaaaaatgcattgaaagttagtgagaatatctgtcgtcctctct
gttctatgtcaagacacactggagctgtcacttgtcttcgattttctccaaacaatagat
tcctggcttctggctcggacgataaaatagtgttaatatgggaacaagatgaagagtatg
agtacgattcttctgttatggagggcatgaacccagtttttccaacgggtcatctggag
atcaaatggacatggagcgctggactgtccgtaaaagattagtggctcatgataatgaca
ttcaggacatggcttgggcgcccgatggcagcatccttgttaccgtgggtctcgatagat
caattatcatatggaacggacaaacatttgaaaaaatgaaaaggtatgacattcacaatt
ctcacgtaaagggaattgtgtttgatcctgcaaacaaatatttcattacatcttctgacg
atagaacttgccgtgtcttcagatatcacaaaacttcacccactgagatgatttttagtg
tggaacatgtgattacagaaccgttcaccaaatctccaatgacgacttattttcgaagat
tgtcgtggtcaccagatggtttaagtattgctattccaaac

AA
MAGATRINSRVVRFAIFASILVLLGFILSRGSATSYSLPSGLTSDTSQSTGSSPKSESKPSSQGSSGATELKKTY
TTDGKEKATFVSLARNSDVWSLASSIRHVEDRFNHKFHYDWVFLNDEEFSDEFKRVTSALTSGKAKYGLIPKEHW
SFPEWIDKERAAKTRKEMAAKKVIYGDSISYRHMCRFESGFFFRHELMQEYEWYWRVEPDIKIYCDIDYDVFKFM
KDNNKMYGFTVSLPEYVATIETLWDTTRAFIKENPQYLPEDNMMDFISDDDGLSYNGCHFWSNFEVGSLSLWRSE
AYLKYFDHLDKAGGFFYERWGDAPVHSIAAALFLHRDQIHFFDDVGYFHNPFNNCPVDADLREERRCMCNPKDDF
TWKGYSCVPEFFTVNNMKRPKGWEAFSG

6.3.2 X-MANNOSYLTRANSFERASES

SEQ0152
P. pastoris homolog of P. stipitis CBS 6054
Mannosyltransferase

Chr 4, 0536

5' region

ggtgatcaatggattcaaagtattgcatgacatgacattgttgtttccagaaattgacca
gctcctggaactccatttgagccttcatatataggacatttttcagtggtagatctttga
cgaatctctgtctttcaaattcttgtctgatctttgttctgacttgggaagctggaatat
ccaagttgtaattgttgacaaatcctctagagtgtcttaagtatctacggtataaattga
gcaccagggttctcatttccgtggaatccttgacataccttgtagtctgggcaaatgcag
ttgggttggtcatcttatgttgaagactcctctatgtgtatgatggttcaggtgcagtaa
tcattcgtcgctaattcgatgagaggatattgggtgttgaactacttatcgaatagttag
gaattttcaggaggaggtaggttgattaaatacgttctattaattggaatgttgttcgtt
tcggtcgtattacaatcgaaagctcctttagcaagtagagtctttgaaaaagggtttgac
taataaggctccctttcattctatttgacggtagggcacacttcattctcagatctaacg
tgactccaccaattttgaacgacgaccactgaacacagttgtaagaacgaatgacaatga
caggccacttacctaaaagcgctactcgcaagagctgacgatcctatctggactacagat
gtgatccttacaaacttaagatgtctcgcaaaaccatttgaaggcacactatacgatgaa
aagatgctcttgccttgagactcaaagggttactacttttaaaggctaaaaaggtcagaa
ctcttcaaacacggtcacaacatcaactggtgctcagtatctattgcattcctaacctct
atcagccttttctgcctttcgagcttccgctacaacgtggtgttaggattcgttatttgg
cttccatctccatctttaactccatttctgcttagctgtct

ORF

ATGGAGTGTTTAGTGTATAAAGATGCATCGGTTGTTGGATTTGTCCTGGCGACGTTGATGGCGACGGGGATTTTT
CTCTCATATGTCCCTCAACACAGTAGAATCATGAAACGGAGAACCTCAGAAGGGTTATCCCCATTCTTCCTACTA
CTGGGTACAGTCTCCGCTATCTCTGCCTTTGCCAATCTTTTACTCTACAGTAATGATGGAAGAATTTGTTGCCGG
TTGGGACTGAGCGACTTTCAATGTATCAACTCACAAATTGGAATGATTCAAGTGGGATTGCAGACCTTGGGCTAC
TCATTGATCTTGGTCCTATGTGTCTATTATACCAGGGATTCATTGACGGAAAACCAAAGAGAATACGCGCAATTG
CTGAAGGTTTATCGTTTCTTTCTCTTTTATGCTTTTCTGAACCTCTTTGTCGTTGTTTACCTCATAAAAAGGAAA
CAAGATACGGAAATATTTGTTCAGTTTGCAAACCTTAGTGGAGTCTTTTCTTCCTTGGTGGGTGGGTTTCAATAT
CTACCCCAAATATACACTACTTACAAACTCAAACATCCAGGTTCGTTATCCATAAAAATGATGTCGATTCAAACC
CCCGGTGGGTTTGTCTGGACTGCTAGTTTATTCCTCCAACCAAACTCTAAATGGAGCTCTTGGTTGCCGTATTTC
ACTGCTGCCATGGCCCAAGGCATACTCCTTATAATGTGCATCTACTACAAACACACCTACAATATTGACGAGCTG
GAAAGAGAGCAATTGGCTAGAATCACAGAGGAAAACTTAGCCTATACGTCTTTGGAAACCACTGATGATGGTCTT
CTTCAATAA

Downstream

agtaaataaagagaagagttatatatgtctaaatctatacggagattgtcttccaaacat
tcagtccatcagatcctgtgctgaccaaagtacctggtatttggctgtgccatctcacat
ccttcacatctctttgccaatgaacaaaaagtagttgtggtggaatatcattcagttctt
tagtttcttgttgctgttgtttgatttcctcgtcatcggcttccactgccaaatcccaca
atgtcactgtgttgtcctcggatgaaacagcaacaatagactcgtcaagcgggttgaaag
aaatagaagttatggcagatttgtgaaaatcatagttaacaacaggtgctggagccgggg
aagagccgaagcttctcagatcccatataccccagtttccatcgtcatgaccagaagcca
gcaaatagtcaaccttggagcaccaggaaatcacattgacatctgtgtcggaggccttca
cactgatagcaggcttgttgtttttagaacgagtatcccagattcgaacatatccatcgg
tgcctgcagttgcaaacacagtttttctgaagtagaccattgaatgtcttcgatggagg
catcagacgcaacaaatggggtgccttctgttgtccagctggatgtagttctcgaagtga
agtaaacattaccattaacatcaccggataaaagggcaccggtactgatcaatggagacc
aatccagtccatatccttcgaccttgccatgatttcttatagtgtaaataggagctctgg
attgcttggggataacaaatccgggtgtggtgaacgctctatattgggaggcaacatcaa
atagttggacctcaccgttttctagcattgatgctgtcagatactccccagttttagcag
catgaggagagactctgagtctattcgttgtactctttagtggtatgttctcgttttcca
agatggggtcagaagtgctgccctcattatcctcatcgtcg

AA

MECLVYKDASVVGFVLATLMATGIFLSYVPQHSRIMKRRTSEGLSPFFLLLGTVSAISAFANLLLYSNDGRICCR
LGLSDFQCINSQIGMIQVGLQTLGYSLILVLCVYYTRDSLTENQREYAQLLKVYRFFLFYAFLNLFVVVYLIKRK
QDTEIFVQFANLSGVFSSLVGGFQYLPQIYTTYKLKHPGSLSIKMMSIQTPGGFVWTASLFLQPNSKWSSWLPYF
TAAMAQGILLIMCIYYKHTYNIDELEREQLARITEENLAYTSLETTDDGLLQ

6.4 ER glycosylation pathway

SEQ0153
SEC59 (YMR013C): Dolichol kinase, catalyzes the terminal step in dolichyl monophosphate (Dol-P) biosynthesis; required for viability and for normal rates of lipid intermediate synthesis and protein N-glycosylation
S. cerevisiae null mutant: inviable
S. cerevisiae conditional mutant: decreased protein O-mannosylation; decreased protein N-mannosylation; decreased GPI anchor protein modification

chr2-1_0498

5' region

```
ttttcctatgcttaagcactctgtaaaaagattgatcctccttcgaacgatgtctaacta
ccctaaagcgactgaagaaagatctaaagagcttttagagaatgttaatcatatattgga
ccagataaccagtgtaaagtcgcccaacgggaggcctgtcaacttggtttgtgtgtctaa
gctgaaaccatcaagtgatattatggcactttatgatgctggctttcgccattttggaga
aaactacgttcaggaactggtttcaaaagcacaagagcttcccaaagacatcaagtggca
ctttattggtggtctccaaaccaataaatgtaaggatttagctaccaggattgataattt
gtatgcagtggagacaatcgatagcatcaagaaggcgaataagttgaacagcagccgtga
tgcatccaaacctaaaattaacgtcttcatccaggtgaatacttccggtgaagaacaaaa
gagcggcatatcctcttatgacgatcttctcgcattggccaaggtgatcaagaatgattg
ccccaatctcactttgaagggtttaatgaccattgggtcaattcaacaatctacagcagc
cggtgaaaggaacaaagattttgaccagttgacggcacacaatgaaaaactagaatctga
tttaggaaccacgctggagctctccatgggtatgagctcagacttcgagcaggctataag
gcaagggtctacaagcgttcgtgttggtagcagcatatttggtgcaaggcctcctagaaa
tgggcattagaacgctacatagagtgtacatatttggtaggactgttcaagatcaggtaa
tatttcaagggctctagcccgtttttccccggtggtggatgttccctaccagtcgattt
tgttccagcctcgaaaggaggtacctaccatatttttttatcaaagccagccgccataat
ctgatccacctgtcaaactattccttctgcttattccaaa
```

ORF

```
ATGGCAAAGAAAAAAGCGGCTGCTAAGCGTCCCAAGAAGACCCTGCCGCCTGCATCGTCGAAGCCTTTAAACTCG
GGGGAGAATCCTGACGAGAGCCAGGCAGAATATTCCGGCCTCCCTCCTCCACAACTCAGCTTCAAATCAGCCATT
GTGGCCGAGATGACCTCGGTGATAAGTAATGGAATTACCATTAACAAAGTTATCCAAGTTCTTATCTTGGCCTAC
ATTGCGAACCTGCTCTATGTGAAACTCGATTTGAACTTTGACTACGATTCCAAGGTTGAAATTGGGATTGGATTT
CTAAGTATTCTGTCTTCAATTTTGATTATTTACGTTTCCAGAAAAAGAAGATGGAATACAGGAGTGGAATTGGGC
TCTATCGACAAGGACGATCCAGAAAAATGGGTTGAATTGCCCGAGTTTAACCTTATTTACGCAATTGTTCTGCCC
ATTTTTGTCACGTACATAGTTGATAAGCGATATTTGGGGGTCAACTTGGTCAATATACTGTTGGTGATCGACATC
CCAATCGCCTTTAAGATCCTTTTGGCTCTTTCAATGGAACAGCAGGTTCTTGATGAAAGTGGGTATCTGAATCAG
GGGTTTGTGACTACCAGGAACAACTTGTTGATACCAGTAGCTCATGTTGGATTTAGGGAATTACTAGGATATTAT
GGCGGATCATCATTTTCAAATACAGAAAGTGAACTCTTCTCTACTCTGTTTGTCAATTTGGTGTTATTTGTTGAC
AAATCCAACATCGAGCTGTTTATTTTGGCCAAATTGACTCTTGCATTCATTGGTTCATTGGTCATTTGCACTGCA
TTGTACCGAGTTCTCATACTTTTTTCACCAAAATATGTTATTTCGTTGATACTGCACCCGGTTTTCACTGCTTTG
GTGTACTATGGATCGGTATATCAATTAGCTCCGGTTCTTGGTAAACACCCATGGGACTACTTGGTGAATGATATC
ATGAATGAGCAGAGGATTCATTTCCTTTATTCTTGGTTGATTCTACTGGTGGTAGTGGTTCCTCCAGTTTTCTAC
TTTGTGGATGCGATTCCTTTAGATTTTAGAAGAAAAATATGGCATTTCACGATCCTTATCATTTTGAGTTACCCC
TTAAGTATCGATCCTTCATTTTGTGTTATAGCACTGGGGGGTGTCTTTGGGCTGTTCTTGGTGGTTGAAATTTTA
AGGAGCACCGAAATGCCCCCCTTTGGACAAATTCTGAGCAGAAATCTTGAGAAATTTCAGGATGAAAGGGATAAG
AGGGGCAACATAACAATCTCCTACTTATACTTGGTGCTAGGGATTGTTCTTCCGGTGATGTTTGATGGATCCAGT
TGTGCGGGATTGGTTAGTTTGGGATTGGGAGACTCCATGGCCTCTATGGTGGGTAAACGGTATGGTTTGGTGAAG
TGGCCTGGCTCTAACAAATCGGTTGAGGGAACGTTTGCTTTTATCGTTGTGACCTTTTTGGGACTTTTAGCGGCC
AGAACCTTCTTTGGCTATCAATTCTCTTGGGAGATCTCCTTCATCGCAGCCGCCCTGGCAGGCGTGTTAGAGGGT
ATAAGTGATTTCAATGACAACATCATCATCCCACTAGTCGTGTTTACAATATTGCATGGTCAATAG
```

Downstream

```
aatggttatagtctcttaatatatagaaggatgtatatataaaaagttgactaactatat
agagacgttttatactaaaactgatatctaaagtactactcattcgctaccccaaactaa
accttgttggtggagcaagttgtgtgaatggatgactttgccaagctcaaattcatcgac
ttgtctaccttatgcgtttccaggaataaacacctgaagcttcccgagcaaaactctggg
caactgaaatttttgagaagaaccagcacttctttgatgtgcgagtgtatgatgacatc
```

gcgtggaacaattatctcctagaaacaattaaatcaaacaccgacggagtagttatacta
tttgactgctttttagaatggcacaagattcagccgatgtttctgaggggctcccaatct
aatcatttatttgtcgtttcggagtttactgcaattgaggaattgtcgaacaaactctcc
aagtttgaagatatcttgctcaaactggatgtcgaggcaaggaagttaccacttaagggc
attgtgctaaccaatatgtcagtttattactggcaatcaaaagataacggtaaaaagacc
tggttttctgtggttgaccaattggaacgattgcaggccaaatacgggtgcacgattata
cgtggaatgtgggactttacctctaaaagggaataaacgctactgcctgcaattaacaat
gtcttttgacccttgttcattcaaaagctaaggctacacatccttcttgctacctttaaa
ctagaagacactatttccaaatcaggaaataacatgtccacttttttcttcattcttctc
aaatttttttttttccctttaaattcttctctctaagaaagatccctaaacccccgatca
atgtctgaagaacaatcaaaaaaggagatcctggagcaagtggaattttacttttccgac
tccaacctaccaaaggacagatttttatacactacagcgc

AA

MAKKKAAAKRPKKTLPPASSKPLNSGENPDESQAEYSGLPPPQLSFKSAIVAEMTSVISNGITINKVIQVLILAY
IANLLYVKLDLNFDYDSKVEIGIGFLSILSSILIIYVSRKRRWNTGVELGSIDKDDPEKWVELPEFNLIYAIVLP
IFVTYIVDKRYLGVNLVNILLVIDIPIAFKILLALSMEQQVLDESGYLNQGFVTTRNNLLIPVAHVGFRELLGYY
GGSSFSNTESELFSTLFVNLVLFVDKSNIELFILAKLTLAFIGSLVICTALYRVLILFSPKYVISLILHPVFTAL
VYYGSVYQLAPVLGKHPWDYLVNDIMNEQRIHFLYSWLILLVVVVPPVFYFVDAIPLDFRRKIWHFTILIILSYP
LSIDPSFCVIALGGVFGLFLVVEILRSTEMPPFGQILSRNLEKFQDERDKRGNITISYLYLVLGIVLPVMFDGSS
CAGLVSLGLGDSMASMVGKRYGLVKWPGSNKSVEGTFAFIVVTFLGLLAARTFFGYQFSWEISFIAAALAGVLEG
ISDFNDNIIIPLVVFTILHGQ

SEQ0154

ALG7 (YBR243C): UDP-N-acetyl-glucosamine-1-P transferase, transfers Glc-Nac-P from UDP-GlcNac to Dol-P in the ER in the first step of the dolichol pathway of protein asparagine-linked glycosylation; inhibited by tunicamycin

S. cerevisiae null mutant: inviable

S. cerevisiae repressible mutant: decreased lipid-linked oligosaccharide accumulation; abnormal mitochondrial genome maintenance; decreased carboxypeptidase Y modification; decreased resistance to tunicamycin

chr2-1_0727

5' region

gttattaattgaaggcacaggtgtgcatgaacgccagttcctctgttgtacaggatttct
agtgtctcgacccatttagcatcggtaaatgagtctgcaaacgctgtcgctttcactttg
ttctttggagtaatagttactttactatctccctgcttctgactttgtggaccaacttgc
cgaatataacccagtagcaaggatagaatctctaaatcatcattagcaaacttgcctgcc
gtaataatcttttccaaaaaaacatgcagtgtcttgagagttatcccttcgtctctagtg
acatggttcttgtctgctgataagagcatttgttgaattatgggtagacttaagaagttt
gagttgaccttggagtttgcataggctaacttcacatatatcttcgtaaaatcccataat
gtttgccctgataaattaaagttaacaagtccataaactctcaaacaattaagaacaacc
aacttcaagattttcctgttggaaataatgtcagggttcacacccccaagactaaaaatc
ttcgatgactcttcagctgtttccctcaaatagaccttgagaaaattttgaagctgctga
aaagatttggagcttaatctttgcgtggattcatcttctcgttttgataagagaaccgtt
actttctgtattagtgcaatgtaggtctgtaacagatggacccatatcacggggtcctta
tcggttagcgacgaacgtaatttgaagggtgggtaagctacaatggggatgtttaataca
tttgaattaacttcttcactggaaggctggcaatttttctactattaccggtgccatttac
tggttgactgaaatgactatatcgatagtctagataccctttgatagtgatgcaaggatg
tagttttccacgagcgctcataaaatcctaatcgatcttacgtgtctctaacggttgatt
gattctgaatggtaataaatagatacgttcactttctaat

ORF

ATGCAACAATTACCCAAAATAGGACTATTGGCAGTATCAATGGCATTGATTTGCCACACATATTCTCCATTGCAA
CCTATACAGTCCTCAATTGGTTTCGCAGTTCTGGGATATTTGTTATCAGACTACCTCATCCCTGCCACGGCGCCC
TATTTCATCAAGATTGGTCTATTTGGAAAGGACTTGAGTAAGAAAGATAAACCAGTCATTCCGGAAACGATTGGA
ATTATTCCTGCTGTAGTCTATCTTTTTATCATGTTCTCATTTATCCCATTCATGTTTTTCAAGTTTCTAGTCGTG
GATACTAGTGGAGGAGGATCTAGGGACACAGGTGTCGATTTGGAGACAGCTGACTCCAATTATTTTCCACACAAC
AAACTATCTTCATATCTAAGTGGTATACTTTCGTTGGAAAGCATGGTATTGTTAGGACTGTTAGATGATCTCTTT
GATATAAGATGGAGGCACAAGTTCTTTCTTCCCGCTATTGCAGCTATACCTTTATTGATCGTCTATTATGTTGAC

TTTGGTGTTACACATATCCTTATCCCTACTTTCATCAAGAACATCTTTGGATTTGAAGCCGTTTCCATAGATTTA
GGTGCTCTGTATTATGGATATATGGCAGCCGTCGCCATATTTTGTCCGAACTCTATAAACATTCTTGCTGGAATC
AATGGTCTTGAGGTTGGGCAGAGTGTGGTTTTGGCAGTTCTTTTACTACTGAATGATTTCTGCTATCTAATTCCA
GCGTCTCTTAGATCAACTCCCGCATATGAAACACATCTGATGTCAACGTGTATTTTAGTGCCCTTTCTAGGAGTC
TCCCTTAGCCTTTTGAAATTCAACTGGTACCCCGCTAAAGTATTTGTAGGTGATACTTTCTGTTATTTTAGTGGT
ATGGTCTTTGCATTTGTTGGTATTTCAGGTCATTTTTCAAAGACACTGTTACTGTTTTTTCTTCCCCAGATTTTC
AACTTTGTTTACTCCATTCCACAATTGTTTGGCCTTGTAGAATGCCCAAGACATAGGCTGCCCCGGTTCAATGAA
GAAGATAACATGATGTATCCTTCTCATGCTGTGTTTAAAAAACGATTGCCCAAACTTATAGAAAAGGGCATGCTA
ATTTTGGAAGCTTTAGGACTATTAGAGGTTGTTAAAGAAGAAATGAAGACTGACAATAAGACCGAGATAATAATC
AAAGAGTGTTCCAATTTTACCCTTATCAACCTTGTACTGGTTTGGTTTGGGCCAATGAGAGAAGATAGGTTGTGC
TTTGTCATTCTACTACTGCAATTCAGCATTGGATTGATAAGTCTCGTGGCAAGACATACAATTGCAGCACTTCTA
TTTGGATACGATAATCTTAGTATTTTCAATCTCAGCATACTGAGTTCCAACTAA

Downstream

tttacatttttatttatgaatctccaaaattcgactcgttggtgaagaatttcatctgcc
aatcggcccacttcatcggtcttgatttgctgacgggtcccctcaaagaattcatccaga
cgcttagatcctttgtatcgtttctcatattgcacagcttggaccaggaactcatatttg
tcaatgtccttcacgattcgagattcaaagtttcgttgttcctcgtaatcaagccacaga
ttgaccatctcctttgcaaaagcactattgtagggatctatcaagcttcctagatacttt
atggcggccaattctctttcgtatttttgctgtttggtaacagtggtgtcttttggggtg
atatctccaaccaaagcttcagctatgtcgtggaccaacgaaatcttgatacattgtgat
aaatctataggttccttttgagctgtggataagtttgagttgtctttgaaattggcggtg
ttgagagacatagatatgattgacattctatacatatgatctgaaatagactctgcattg
tcaatccccatattcaaccacccagttctcttctgtgtcttcagtagttcgacaacttgt
ataaatgcactaatgtacgtgtaagggcctacttcagcattctgaagaagatccttaacc
tcgctgggtacgtgatcctccgggttccaatcattggtggacatcggacagtataatgta
aaatgtaaacagtagaataggatatttactaagcgaaaatttgaccccaagaaaaaaatt
aagccaaagcagatatagagtgaaatatggttaattggtgaaatatcactcattcgtttt
cagatcatgagatgctcagaaagtacttcaaaaacagatatcttcaaagaaggtctacaa
tgcagtttagctatggaaatttttttgtttaaatatacaaaagagagagtctcaaagtcga
catttaagcaactaattcaacaacaccaccaacagctctg

AA

MQQLPKIGLLAVSMALICHTYSPLQPIQSSIGFAVLGYLLSDYLIPATAPYFIKIGLFGKDLSKKDKPVIPETIG
IIPAVVYLFIMFSFIPFMFFKFLVVDTSGGGSRDTGVDLETADSNYFPHNKLSSYLSGILSLESMVLLGLLDDLF
DIRWRHKFFLPAIAAIPLLIVYYVDFGVTHILIPTFIKNIFGFEAVSIDLGALYYGYMAAVAIFCPNSINILAGI
NGLEVGQSVVLAVLLLLNDFCYLIPASLRSTPAYETHLMSTCILVPFLGVSLSLLKFNWYPAKVFVGDTFCYFSG
MVFAFVGISGHFSKTLLLFFLPQIFNFVYSIPQLFGLVECPRHRLPRFNEEDNMMYPSHAVFKKRLPKLIEKGML
ILEALGLLEVVKEEMKTDNKTEIIIKECSNFTLINLVLVWFGPMREDRLCFVILLLQFSIGLISLVARHTIAALL
FGYDNLSIFNLSILSSN

SEQ0155

ALG13 (YGL047W)**:** Catalytic component of UDP-GlcNAc transferase, required for the second step of dolichyl-linked oligosaccharide synthesis; anchored to the ER membrane via interaction with Alg14p; similar to bacterial and human glycosyltransferases

S. cerevisiae null mutant: inviable; increased cell size, decreased resistance to 2-phenyl-3-nitroso-imidazol(1,2-a)pyridine

Chr1-4_0448

5' region

tctgaaaacacttttcgaaaacagcctcaaatcgaacgggattgatgcttctgtggagag
catcataaatgactcagttgcagttttcgttgctgggaattacttggaaagctgcgtaac
gggattagttgctggcactggcgtcaattgtgctttggttgtcgataggaaaatgattgg
attacaaaagtcacaggaacagattcggcctttacgaaagggaaacttcgacaaagtagt
tctgaactccgaactaggtttctttggtggttttctgtgtcaacatagtaccatctggga
taaaatggtgcatgagaactgggcactacatcactcagttcaaacgaaccaaccacatat
gactcttgctcttggatgttttcagccgattgaaatgatgtgttctggccgatatattgg
agagttgacaaggctagctattctgtcgctaattgactctaaacagctattcaaaggaag
tcggcaaaatatccctagtgcatttaacgttccgtatagcctaggaggagcaattgttag
tgaaatctacgaatcagaaggacatctacctgagctctgggctcccttgccccttcaca

agaagattaccagatattgtttaaggttgttgattcgataattgaaagagcgtctgtagt
tttggggtgtgctatcattgctcttctgaagctcagcgagtcttcagctacaattgaccc
gactggaagtggttcatgtcgaccttaccgaatcgggtatgttggttcagtgctaaagtg
tttcaagccttatcgggaaaaagttgaaatggttttgaatttggcctcaaaacttcaaat
catcaatcgaccagcctccctcatttacatagatgatagtaatctagttggagcggcagt
ttcagtcttcacaccagaatgattagattagttagtatgcatcgcgcgatcagaatttac
ctaccttactccataagcgcagagaaatatcatccattct

ORF

ATGAAGAGTGTACTTGTAACAGGAGGAGCCACTGTCACATTTGTTGCGCTACTCCAATTGACTCTCAATGAAAAG
TTTATTTCTGCCTTGAAGAAAAACAATTTCGATAAACTTGTTGTTCAGTATGGAACTCAACCCACAGGTGAATCA
CTATTTCTGTCGCTGATCAACAAACTAACCGATGAAGATTATAAGAAGTCTCAATTAGGTCAGCTATACAACATA
AAGCTAAAGGATGGCTTTTTAATCCAAGGATTGAGCTTTGATACTGATTTCGTCAAGAATTACACGTCCAAAGTT
GATCTCGTGATCTCACATGGAGGAACTGGCTCTATTCTAGATACTCTGAGAGCAGGTAAAAAGTTGGTGGTTGTT
GTAAACGATACATTGGCCGATAATCACCAACTAGAGCTCACTGAAAAATTTGCAGAAAGAGAAGTTTTGGGCTAT
TGTCGGAAAAACACGGTAGAGGAGCTGATTGATCAAGTGAATCAAGCCGAATCAAGAGAGTTTAAAAGATTGGCG
CCTAGCAAGGGAAGAATCATACAAGATGTTATATTGAATAATATTATTAAGCCATAA

Downstream

aagcacaaggttgagtctaaagttacattgtgagtgtacttccgtcgcttatcacatcaa
gggaatcactcagccacttttggaaccaaaaatgttattcgataagccttctggcaact
cggctttttcaagttgatgctatcttttatggatattaagccagtgaaacttagagtta
gcagtatcttatcaagagtgaaaaagttgtgtttcttttcatttgaattgtgcttggtca
ttgatgaaatcagagtcattctcaagatgtataaccatatcgatctataagtcgcagttg
cttccaagtttgactcttgctcaatatccagatctatggaatcttgagcaggtcttttgg
aataaaatgcgactaaaaacccagaaagtagcccaattatatgcagtctgaacatgagtg
gtactttggtgagtgacctccatatccatgacatggatggatttcgccctttttcttgtgt
aatatgacatcaacaacgacgtggatgacacagtaacaacagtcaaggagagtttgagac
tttcttttacgctttttatgactatctgtttgtaatacttccatttgctagccgctttca
gctgttccaattcttccgtgctaagtctcaagttcataaagaagaaaaatggaaagaggt
attcaaggactaccgtgtattttcttggcaaatatcgcaacagaaagtttctcagatcaa
atgcaaatcgatttttcatgctattcttaccaattatgctttccagttcatagaaagatt
tgaccatatcaccagatgaaaccatgcgagaagttcctcttttgactaataggccttcac
ccataaagtttaagatgttcctgaaatatactggacagttctcgtaatccatgataaacg
acttgaaaatctgcgagtaacataatgggaatagataccatgaacgtaagagtttgtctc
tctttggaacacttttagcgctttgagcctacgaatgaa

AA

MKSVLVTGGATVTFVALLQLTLNEKFISALKKNNFDKLVVQYGTQPTGESLFLSLINKLTDEDYKKSQLGQLYNI
KLKDGFLIQGLSFDTDFVKNYTSKVDLVISHGGTGSILDTLRAGKKLVVVVNDTLADNHQLELTEKFAEREVLGY
CRKNTVEELIDQVNQAESREFKRLAPSKGRIIQDVILNNIIKP

SEQ0156

ALG14 (YBR070C): Component of UDP-GlcNAc transferase required for the second step of dolichyl-linked oligosaccharide synthesis; anchors the catalytic subunit Alg13p to the ER membrane; similar to bacterial and human glycosyltransferases
S. cerevisiae null mutant: inviable
S. cerevisiae conditional mutant: increased heat sensitivity

chr3_0944

5' region

ggatatctggttggctgttttgaataaatattcagtgacaggtgttggactgaaggctca
catgtacatcgaagaggcagtcgatcctgatttctctgcctctacgccgactaaagaggg
agatattgactccacgggactaaacattgcacagttgacgtcatccgttaaccctgaaat
tgcagctcgtgttgctgagttgttaaaagaccctgaatttgctgttcatgtaggcaaaga
tactgaaaccaagaagcctacagaaacaccagctactccaccagaaacttcataatcatg
ccctcagaatattcaatgttttcttttgtacatagaaatatacaatactataactcagc
catacatgtcgtcaaaatcatcttcctcctcactatcagagggtagtactatctccttcc
ttttcgtcacgtcaacacgacacaacgggcaggtcttgttcaaatgcaaccatggtccta
tgcagtctaagtcaaaatagtgattcttgttgcatggcaacctcacaaccaaagggtatg
gatcgttgtgaaattcattagtgcaaatgggacatgtatcgtcaggagataaatgcttga
cgttaactctatccaattgatctaaataggcatcatcaactcccttggcagtatctgtgt

ctatttgatcccccagagagttcagcaaatcaaccatagaatctaatctgtttgaatcta
gactggaactactctggagctggttgagaagcatggatactacggaataaggatcatcat
taaaactgcttcgagactgactggtgggtctattttccaggtacgtctctaccatttgtc
tgagagtgggccttcgacggccattgttagcctgagatgacatccagctgtatctggttg
aatgagaggcaattgattatcaagttacaatattttcgctaaagccattgcacgtttggg
gtaggacaagtatactattgttggttcgtcctacaacgag

ORF

ATGACCACTTCAATTTACTGTCTTTTGGCTTTATTGATATCATTGACTCTTGTTGTCATACGTGTAATTTTCTGT
GTCCCATTTTGTCGTCTTCAGGATGTCAGTCCAGTATTAGACAAACCGTTAAGTGTTTTGATCCTTTTGGGATCT
GGGGGTCATACCGGGGAAATGTTGAACATCCTTTCTCAATTAGACCACAAGTTCAAATATTCATTCATTGTTCAA
AGCAATGACGAGTCTTCGGTTCTTAGGTTGGAAAAGAGTCAGGTTAAAGGAACTGTGTATACCGTCCCGAGGGCC
AGAAACGTAGGTGATGGCCTACTCAGAAGCATTCAAGGAACGCTAAAGTGTTGGCTGGGAACCATGAAGGTGCTT
GTTTTCGATAAGAAGTGGAAAGAAGGAAACATACCGTCCGTCCTTCTGGTGAATGGACCTGGGTCCTGTGTGCCC
TTAGCGTACAGTATAGTGGTTCTGAACATTCTGGGACTAGCAAGCGCACGTATCATTTATATGGAATCACTAACC
AGGGTCAATGAATTGAGCTTGTCAGGAAAGTTGCTGTATTTGGTTGCCGACAGATTTGTTGTTCAATGGCCAGAA
TTGGCCCAAAAGTATAGACGTACCGAGTACCATGGAATCCTTGTCTAA

Downstream

acgacatacaaaaacacaaaattgaaatagattaaaggaatgaattataatattacaaac
tgaagtataaatacgcgttccatttcactagtaggaggttagttgttgacaatttcttct
tttatttaaaggatcatgagatgaaaaacttgggtgtcagggttaaaaagctctgtttta
ctgaatccacataatatttcaggtgaagaaattcatgttgtaaaaatgttcatcatccaa
ccaggttattcagcgatccttactcctaaaaacttacccatggctgaggccaaataaaat
agaagacattgctactaacaaaaccgatagctagtaaatttttcttcccagcatcccaca
aggttgagaactgtgtgatacacgaccacgcggttgaacggatttttattcagggttgtc
ccgaaagtataaatgtgggctgtgtaattttacccttaatcacattctgcgcggaacata
ctgcatgattagaaatatttatgctctcaattaattgatcggcatttttttatgattcagc
ctgtttttccggtgggggcgtttttttattgttcgtttcaagttgaaaaagctaaatttca
ggtagttacctttttactttaatccctaactattgagggcaaagcctccttgcaggccta
ttataagagctattaatactgtttaattactaatttgctcctatatcatgaactctggaa
ttctacgtcatatagcgtcgcatgcagctgtgcctattaccccccttttgagagtgccgt
ccaatcattgtcggtcacacgtctcgccgaaaaagctttttttttgttgtttctcggacgt
tgccaagattaacgtcgcaacccttgcagagtataggccctgaaaccagaattttatcag
ccaggtttgccacatcaaaaatcactcaactctatttttactaccgtgttcgatttttgg
ccagagatatatatgccctttggcccccagaatagctgtt

AA

MTTSIYCLLALLISLTLVVIRVIFCVPFCRLQDVSPVLDKPLSVLILLGSGGHTGEMLNILSQLDHKFKYSFIVQ
SNDESSVLRLEKSQVKGTVYTVPRARNVGDGLLRSIQGTLKCWLGTMKVLVFDKKWKEGNIPSVLLVNGPGSCVP
LAYSIVVLNILGLASARIIYMESLTRVNELSLSGKLLYLVADRFVVQWPELAQKYRRTEYHGILV

SEQ0157

P. pastoris homolog of *S. cerevisiae* **ALG1 (YBR110W)**

Mannosyltransferase, involved in asparagine-linked glycosylation in the endoplasmic reticulum (ER); essential for viability, mutation is functionally complemented by human ortholog

S. cerevisiae null mutant: inviable

Chr 2-1, 0759

5' region

tacaggcatctcatcacaagtccacgtagaagagtcatatggaactagaatggttgttgt
tgatatcactactaatagattgcgatacagaacttgtcacaaataggatatcaatgagcg
gcaaaatgactacctaaatatagcgtctcttgtacagatgattcataaacctgcgaatcg
tcatactcagcagcagttgcaactgtaacagcctggctgttttccgtccattatcgctaa
ttgggtggtcgactttatcactcaattagttaggaccagtcagcggatctattttctgag
ctagtccatatttacggactttgtattcggatggtgtctctagagattatcaacaactaa
attaactaccgaactatttattaactaggagattggaaaattccttgatgtcaataacac
catcattgttagtatcagcttctctaagcatctcgtcaacctcctcctctgtcaacttct
caccgatggaagtcagaacatgtttcagttcacccttgtcgatttttccatcgccatcct
tgtcaaatactcggaaagcctcgaagatttcggcttgactatcagagtctctcattttcc

tagccatcattgttaaaaactcggggaaatctatgcttccatcggtattgctatcaatct
ctcggattaaatcgttcaattcactctctgttggagtttgacccaaagagcgcatcacaa
ttcctagctctttggaagtaatcttgccatcttgatcttgatcaaaaagggaaaatgctt
ctttgaattcagaaatttgggcttcggacaatttatcggactaagagttgttagtagaca
attggctagaaacaatttatcaatacttaccattatggcagtaattacacggtacaaagt
gaggagcaggttattttgagattcaagagaaagtatctctttaggttaggacacgacgag
catgttgtcatgaattttgtctctataagatctctatttgc

ORF
ATGTCTCAGTTGAAGGAAGGCTTATTTAAGTTATTCGATTTGCCAACATGGTGCTGGTTATTGGTATTCGTCTAC
GCCACAATTCCGCTGACATTTTACTACTTTTTGCCAATGTTAGGTAGTTGTCTTAGAAAGAGATCGTGCATTGCA
GACAAAGCCGTGATCCTGGTTATGGGAGACTTAGGCCATTCTCCTAGGATGAATAACCACGCGTTGTCTTTCAGT
AGGATTGATTATCAAGTTGAATTGTGTGGATATATTGATTCCAAACTAGCTTTTGATATTATGCACAGCGACAAT
ATTTCAATTAACCAGATTCAAGCCTTAAGGAATACTGTAGGTTTACCATACTTTCTTTTTGCCATGTGGAAAATT
GTTTATCAACTTCTCCAGCTACTAAAGTTACTGATGAGGGTTATGCAAGATTCTCGGTATATTTTAGTACAAAAC
CCGCCTGGGATACCTTCAATATTAGTGATTGTTATTCTCAAAAAAGTATTGTTTCCACATTGCAAGTTGATTATA
GACTGGCATAACCTGAATTACACGATTCTAAATTTGAAGTATCAAAATTTACACCATCCATTTGTCAGATTCTTG
AGATTCTACGAATTCCAGATGTCAAAATATAGTGACTTGAACTTAACAGTAAGTGAATCAATGAATAAATTTCTT
CAGACAGAGTTCGGAATCTCTTCAGCAAAATTGGTTACACTCTATGATAGGGCTCCTACACAATTCAAGCCTATT
TTAGATCCATTTCAAAAGGAATCAGTCATGAAAGCTCATCCACGTTTGTTCCAACACAGTATTTTCTTCAGTAAG
ATTTTAGTCACAAGTACGTCTTTCACTCAAGACGAAGATCTCCCATCATTTTTGAAAGCTTTGAAGTTTATTGAT
TGCTCTTTGAAGTGCAAAATTCTGGTGATCGTCACAGGAAAAGGGCCTCTAAAAAATAATTTTGAGGAGCAGTGT
AATGCTCTCACTTTCTCCAATATTTTGGTCAAAACTTGCTGGTTATCTCCTGAAGATTACCCCAAGATCCTAGCG
ATTTCTGATCTTGGAGTATCATTGCATGTATCGTCAAGCGGAATTGATCTTCCAATGAAGGTAGTCGACCTCTTT
GGATGTGGGGTGCCTGTGGCTTCTTTGAAGTTTGATGCAATTACCGAGCTCATTAAGGAAGATGTCAACGGTGTT
TTGTGCGACGATGCTTTCAGCCTAGGAACTACGATACAACGTCTTTTCGAAAATGCCGCAGAACTACAAGCGTTG
AAACAAGGTGCTCTAGAAGAAAGTTTAAAGGAGTGGAACGTGGAATGGAATAAAAAACTAGGTAGTTTGTTGACT
TAG

Downstream
cggagatatggtacttctgaacgatgacaatgttgatcaccattgtccactattgggata
cagaatgaactcttttcatcttgaacgcatgagtggaaatatatcataaggcgtcctagt
aaggaaactctgtgattcgatctcgaccatcagaattattgggcatatcttatgtattat
ttgtatatatagctgtatatctgctattgttctccgtttggctcttacggaacgtgccaa
aaagtatctataccaatccattagcaatatgcaaccaattaccttgatacaccggtcgaa
tgaacaggaactgtaggaatgcttagaactcgtttattattcaagggtaataccaagcaa
ccaggcacgagtaaagaccaaagatagtcaagtcagaactcttgatcgttactgagttag
tactcagtaaattgatatcacgaaccataaaggccgacattagactataccaagagagaa
tcatcacctaaaagtccaaagcatctctagcccaagtcggcaaaatgaaagatgcttcgt
gtatttttttgttgtagtaacgcaagtgctctttctcaaatgtgtcggaaaaggatctca
aagggaacttcaaactagcgttaggatctttgctacacaccataaacccaatctgacccg
atggatatgttggaatggtacagtaagcgtattcaacaacaggaaaaacctctaaacaag
ctttcttaaggttcttgatgatctccatatgaagccaaatattctcaccttgagtagtga
taactcccttctcagtcagtgcgtcatttaacaatttgaaatacggtttttcaaacaatg
aagctgctggtccttctggatcagaggagtcagtaataatgacgtcaaaagtgtttttgt
actcgtccaaaaatttgaacccgtcgccaatatgaagtgtgactttgggatgcttatatg
atttagccatttctggaaggtagattttggatagacgggga

AA
MSQLKEGLFKLFDLPTWCWLLVFVYATIPLTFYYFLPMLGSCLRKRSCIADKAVILVMGDLGHSPRMNNHALSFS
RIDYQVELCGYIDSKLAFDIMHSDNISINQIQALRNTVGLPYFLFAMWKIVYQLLQLLKLLMRVMQDSRYILVQN
PPGIPSILVIVILKKVLFPHCKLIIDWHNLNYTILNLKYQNLHHPFVRFLRFYEFQMSKYSDLNLTVSESMNKFL
QTEFGISSAKLVTLYDRAPTQFKPILDPFQKESVMKAHPRLFQHSIFFSKILVTSTSFTQDEDLPSFLKALKFID
CSLKCKILVIVTGKGPLKNNFEEQCNALTFSNILVKTCWLSPEDYPKILAISDLGVSLHVSSSGIDLPMKVVDLF
GCGVPVASLKFDAITELIKEDVNGVLCDDAFSLGTTIQRLFENAAELQALKQGALEESLKEWNVEWNKKLGSLLT

SEQ0158
P. pastoris homolog of Candida glabrata **ALG2**
Alpha-1,3-mannosyltransferase
S. cerevisiae null mutant: inviable

C121, 0002

5' region

attgtcttgtgagttggttcactgggaagagaatacgctgtcagctgtgcgcttgcgcgg
gttttccaatatctgtgcagaccctaaaagtttaaaatttggtctctagatagaactctt
ccacttatcaatcgccttttgtgccatttttttcattcccctctgatggactttgtggaga
aggcagagtaagtgcatcgtctttttagcaatgacttaactatctaacaatccaggaaaa
cactcgccacgtattcaacaactagagacctgtgttcggcagtagtatctaatgagtaca
accctggattaccccatttcacaaggagtttagtgtggaagcttactttgattacagaat
atttcaagtccaaaccacatttgccttcatcgatgaaagttgattattcggttttgactc
agactcgaaacacttatcagaagtacagtgagaaatacccagtaccgtggagtttactgc
agcaaaatgatgaattttaccaagaacaggatgcttctgttgatccccttttctaacacag
acgatttagccaaactgcagactttggtgacagattgtaaacgcctgtttccaagttatc
ccaatttatttatcaaggacaaggagaactctagaaatatcatcacatctttatttatat
ggtggaagctcaacgacgatgtaaatgatacgaactctgggtatcgccaaggcatgcatg
agatctttggactgatcatgatcagcttgatcaaagaacagctaccacaagggaaaccag
agcaattctatacgaagcatccggtggcccaattgtacggtcccaaatttacacatattg
actcgtttacgatattcaatcgcatgatgtataagctggctccatcatttttcaggaag
acaatttcatccaagaatccattaaatttgacattctatttcacaactgtgaccgatacc
accatcaatatctgacgaagacgttgaagattgactcttca

ORF (underlined intron)

ATGAACATTGCATTCGTTCATCCCGATCTGGGCATAGgtaagctttgtcatgaatgtagaacaatagaaagaggc
ttactaacatttgtttaagGTGGTGCTGAACGTCTCGTTGTGGATGCTGCCTGTTCTCTGCAGAAAATTGAGAAC
CAAGTCATCATATACACATCGCACTGTGACAAAACTCACTGTTTTGAAGAGATCAAGAACGATGAAATCGAGTCA
ATGGTTTTTGGAGATTTCTTGCCAACCCAAATTGCTGGCAGATTTAGCATACTGTGTGCTATAATAAGGCAAGCC
TGGTTGATTCTTCGTTTGGCATTGTCTGGAAAAATCAAAGAGCATGACGTGTTCATTGTTGACCAGTTATCATTT
TGTCTTCCCATATTGCATTATCTAAAGCGAAGTGACGCTAGAATCCTATTCTACTGTCATTTTCCTGATCTGAAG
CTTGCTTCTCGAGACACCACTCTACGAAGTATATATCGGAAACCGTTTGATTTACTGGAGCAATATACAACCATC
TGCGCAGACAGAGTGGTGGTGAACAGCAATTTTACGAAAGGTATTTTCAAAGAAACGTTTCCCATTATTTCCAAA
TATTACGAGCCCACAGTGCTGTATCCATGTGTCGATACGGAAGCTTCTTCGATCGACGACACAACTGAAGAGGAG
ATGAAAGAGTTTTTTCATAGTAATGAACGATTCTTTCTATCTGTCAACAGATTTGAAAGAGCCAAAAACATTTCA
CTCGCAATCAGGGTGTTTGCAAAGCTTAAGAAAGAGAATCTCCAATTTTTCAAGGAGAATCGCCTTAAGCTTATA
GTGGCAGGTGGGTATGATTCCAGAGTGAGGGAAAACGTGGAACATCTAATTGAACTGGAAGATTTAGCTAGATCT
CTCGGTTTGAAGGTTATATCAATACGAGGCCGTCTCTTTTCATATCCAGCTAGTGACGTTATTTTCCTGCCCTCG
GTAAGTTCCGATATAAAGGATTATTTAATTTCCAAAGCTGAAGCCCTACTCTACACCCCAGGGTTTGAACATTTT
GGAATTGTGCCTTTGGAAGCAATGAAATTTGGAACTCCAGTTATTGCGGTCAATCACGGTGGCCCCACGGAGACT
GTGGTAGATGACGCCCAACCTGAACCAACTGGCTACCTAAGATCAAATGAAGTTGACCCATGGTACCAGGCTTGC
TCTCAGGTAGTCAGATTATCTGACGAAGAAAGGCAGAAGCTCTCTGCCAATTCTAAGAAGCGCGTTGAAACATAT
TTTTCAAGAGAGGCAATGGGTAAAGCATTTTCTGACAACATCGATCAAATGTTAACTGAACCTACAAGTGATAGA
TTCAGTTATGAAAAACTTATAGACATCTTCTTCGTATTGGGACTTGTACTACTCAATGTCATACCAATCGTAGTT
GCAGGGTATACATTCAAGCGAATGAGAGGCGATAATTAA

Downstream

agttactaatggctacataaatagagtagacaaaaatataatacaattaattctccctgt
aatttggttgatactgattaaattgaccttggaattgatttcccatgagcggatttgaca
attgctcatttggtagctgggaggtttgattcggcccttgattggccaaatttgatggaa
ttgcaagaggtgagctcatagaagttactgggctggttggagaattttgcatctgtcttt
gagcagcctgaagcacttgtgcccttgactgggcttttagctgtgcctgttgaaattgag
tctgtgcttgtgattgtacatgtgatggctgttgttgctgctgctgcatctgattttgaa
ttgcttgatttttatctgctgttggagctgaactaattgcatcctttgaaattgagctc
ttttctcgggtgtaaggtttgcaaactgttgttgctgttgcggagtcatacgcatgtgga
gattcctcagatgctgttgctgctgcagtcgaaattgttctttctgcttttgttgtaact
gcaattgttgctgaagctgcaattgttgcggcgtcatctgagaatactgttgcggcagtt
gattcgaaggacggttgtcggggattggtgattggtgggtaatcggtggtgactgttgct
gttgagcttgttgctgttgctgttgttgtaagagtagcatttgctgttgc

AA

MNIAFVHPDLGIGGAERLVVDAACSLQKIENQVIIYTSHCDKTHCFEEIKNDEIESMVFG

DFLPTQIAGRFSILCAIIRQAWLILRLALSGKIKEHDVFIVDQLSFCLPILHYLKRSDAR
ILFYCHFPDLKLASRDTTLRSIYRKPFDLLEQYTTICADRVVVNSNFTKGIFKETFPIIS
KYYEPTVLYPCVDTEASSIDDTTEEEMKEFFHSNERFFLSVNRFERAKNISLAIRVFAKL
KKENLQFFKENRLKLIVAGGYDSRVRENVEHLIELEDLARSLGLKVISIRGRLFSYPASD
VIFLPSVSSDIKDYLISKAEALLYTPGFEHFGIVPLEAMKFGTPVIAVNHGGPTETVVDD
AQPEPTGYLRSNEVDPWYQACSQVVRLSDEERQKLSANSKKRVETYFSREAMGKAFSDNI
DQMLTEPTSDRFSYEKLIDIFFVLGLVLLNVIPIVVAGYTFKRMRGDN

SEQ0159

P. pastoris homolog of *S. cerevisiae* **ALG11 (YNL048W)**

Alpha-1,2-mannosyltransferase, catalyzes sequential addition of the two terminal alpha 1,2-mannose residues to the Man5GlcNAc2-PP-dolichol intermediate during asparagine-linked glycosylation in the ER

Chr 1-4, 0417

5' region

aggttgttggattagattaaacctagttcttgaagaaatgctatcagcgcgcggcgataa
ccgcgatggccggaaacgctatcaagattcatattataatcacagataataccaggtgat
gaaagaaaggagattattcattgtggtgaagagcatgaattatacataaaacactactta
ctggtttgaacactttaatattggtcaaattcttcttgaaaaagattttaaaacaccggt
gaatgaactaaaaatattatgttgccaaatgacaggtaaataattatgtacccaaagcag
taatcgatcagtaggcctttcaagagtcatccactagaatgttgttagtggagacaggtc
tagcatccttgtccgagtcaacctttaatccattcacttttggagttaaagtgtttaaag
tatcttcagttgtgcttcgaatatggtctggatctgttaagtttggattggaggcagtct
ccaaagatctacgatcatttcctgactcgtcgtcgacctgagtcgagattccaagagcgc
tcacttcaggggcgtaattgaacccttgtggcacctgagcagcatcttggttgggtaatt
gctgagcgtaggtggtgtagtcgttaccagaagcttgcgctccggatcgtggagattggg
gaccattcctttgtgagttctcaaagtatggagcttgattgccctctgagccaggaacat
tctgaggcaaacttagaactgtgttcagctggtcgtattgagtagaatgctggtgcccct
tatctgatggtacagttgaagatgaaagaagagaaccagctttagagtggcgctttgatt
gtttcacctgtactctctgtgactctctttgggcggcaatagaactacgggcttgaggag
taggtagcgcaattaattgcttcatatcgttgtaccatttcatcatctcctggtaaccct
cagcacggaacacccagttatgacccctgtttatgatacc

ORF (frameshift present, possible around region in bold)

ATGGTTTTTTCAGACAAACTTACATTTCCAGATATGCCCGTTGAAGAGACCCAGACGGGCTCCAAGGTGGCACTG
TTCGGACTAGTTCGTTGGTTTCTTTTGCCAGGGATTCTACTTTTTTTTATCTGTTATTTAGTAGAGTTTATCTAC
GATGGAGAGTGGTGACACCTAACAACTGGAAGAGGTATGTGACGAAGGCCGTAGCAACAGGAGAGCCAATAAATA
CCACTCCGTTGCATGCTAAGGCTGGTGCTGTTAGACGGAGGCTAATTTTAGGCTCTAGAAATCCTTTCTTTTATT
CTACGGCTGCGGATCCCACGAAGGACAAATTTATTGAAATCGCAAAGGAGGATAGGGCTAATGCCAAGGAATTCA
TAAACGAAATTCTTCCCAAGACTGCGTATGATGTCTCTAGAAAGATCATCTTTGGA**TTTTTT**CACCC**TTTTT**GCA
ATGCTGGCGGTGGTGGAGAGAGAGTCCTATGGCAAGGTGTTAAGACCACTTTAAAGGTTGATGATAAGAACATTT
GCTTGATCTACACCGGAGATTTAGTCACCGGCGATGAAATTTTAGACAACGTTGCAAGGAATTTTGGTATTTCTT
TCACTGATGTAGAACGTCCCAGATTGGTGTTCATCCATCTTTACAAGAGAAAGTACGTTGAGGCGCAGTATTGGC
CTAGACTAACTCTAGTTGGACAGGCTTTCGGATCCATGTTACTCACATTAGAGGCATTACAAAAAGTTATTCCAG
ACGTCTGGGTAGACACTATGGGTTACCCATTTTCGTATCCTTTGGTCAGAGTTGGTGTTAAGATTCCAATTGTCA
CCTACACGCATTATCCTATCATTTCAACAGACATGCTCAACAAGATCAAGATCAATAGCATTAAGACTCTAGCCA
AGAAGATCTACTGGTGGGGGTTTATGTTAGCCTACAAATTTGTAGGACATTTTGTGGATATCGCATTTGTCAATT
CTAGTTGGACTTCTAATCATATGAAGAAAATCTGGGCTTCTGTGTCCACAGTCAAGACACTATATCCTCCTATTG
GAATTGACAAGAAAATAACTCTAGACTCAGTCAGCCCTAGAGAGAAAATAATTTTATATATCGCTCAATTCAGGC
CTGAAAAGAGACACGAGTTATTGATCAATGAGTTTTCCAAATTTTTCCAGCTAGATAAATCTTACAAACTAGTCA
TGATCGGTTCGATACGTTCGAAAGACGACGAGGATAGAGTTGATTTTTTGAAAGACCTGGCTTATAAGAAGAAAG
GAATTCCTGAGGAAAACTTGGTTATTATCAAGGATGCTAGCTATTCAGTCATTCAAGACTACCTTTCAAAGGCAG
AAATTGGTGTCAATGCCATGTGGAATGAACATTTCGGAATTGCAATAGTGGAGTACATGTTCAATGGACTTCTTC
CGTTGGCCCACGCTAGTTCTGGACCATATTTGGATATTTGTGTGCCTTGGAGCCTCGATAAGGGTGAACAGCTAT
CCGTCAACTCATCTAAACCTACAAGCGATGATTCTAATAGAACGGGGTATCTCTTTATATCCGAAGAAGACCCCG
ATTTTGAAGCCCGTACTGAACCTTTTCCTAGTTTACATGAGGCGCTTCTGGATATCTCAAAGCTCAGCCAGGAGC
AAAAATTGGAGATTTTGAAACGTGGAGCTTCAGTGGTTAATAGCAAATTCCAAGACAATGTTTTCAATGAAAATT
GGAAAAAAGCCATCACTGATGCTGCAAAACTAGAGCAGGTCTACCGGATATCTAAACGAGATGGGGAAGTTGAGC
AATTATTCTAA

Downstream

ttttgaaataatgtagtgtcgtttatagggtatggtccttaagttgttccgaatgaattt
tcacttgtattatctggatttattatccgcaagacttctctttcatcctcaaaatcgtca
tcttgatcaaaccttctaacctgtccgctcccaaatggacctgttgaatttgtatctagt
gaggaaatcacggactgggaagatgatgacaaaatgcttccattttgatggtagttttgg
acatagttcctaacaaattcagaaacataatcgtctctttctgcaacagaagactcatct
tctggtaattgcgcacttttaatgttcctccttcttctgacaaagtagcatattacaacg
aagagtaccaagagcgaagaacagccaacggatattcctacgatagcaccagtgctcaaa
acttgtccgttattctcaatgtatgaggccgttgccgaagatggagatggagtcggtata
aactgggaagaagactggagttcgtcggatgaagtgtcaacagtatttacctgagtggct
tcattggctgcagggactgccggagccactgtttcagctacttggctttcaatggaagtc
gtctctgaccattctgaagagtcttctgacgatgatgttatgttggccaaagtgttccaa
gagaaaaatgtctggggagatgatgatgaggaaagagaaaaggtaaaattctgaactccc
gcagtagttaatagtgttgaacttccttccaaaaatagGtaactcccatcacttattgtt
gaagtatcattactgtcagatggcaaagtagtagaaggatactcactggtggaagcggat
gaagagattgagtttgccagatcctcgtggagaaaaacctggaaactgaatgtcccacca
caagtttgcgtagtgttaccagcacatctaacttcacactcgatgctcctttcgctttca
acggcgaaggaatccgaacaataacactctgttccattga

SEQ0160

P. pastoris homolog of *S. cerevisiae* **ALG9 (YNL219C)**

Mannosyltransferase, involved in N-linked glycosylation; catalyzes the transfer of mannose from Dol-P-Man to lipid-linked oligosaccharides; mutation of the human ortholog causes type 1 congenital disorders of glycosylation

S. cerevisiae null mutant: carboxypeptidase Y modification: decreased; resistance to brefeldin A: decreased; resistance to tunicamycin: decreased; resistance to mercaptoethanol: decreased; resistance to L-1,4-dithiothreitol: decreased; viable; inositols excretion: increased; competitive fitness: decreased; sporulation: abnormal

Chr 2-2, 0036

PATENTED BUT PATENT CONTAINS ONLY PARTIAL SEQUENCE (DD175842.1)

5' region

agcacaccgtcgcagtcaaagagaaaagtgtcataggccgctaatagtccaacagctgac
tcctttgagtcaatcttggtagccataatcggtaaaatagagagaaaggtgcctattttg
cgcgaacaatctctcaataaggtggaatctgataaggtacagtaccatcgtcttcaacat
ccgccatttaacatcgctcaattaactaccactctcgaaaaacaaaaactccattggtca
aatccagaaccacaactaacgatatggctacccaagacttaccacccatctccgggtatg
cacccatccagtggaagagaaatttaccctcacgtggtttcaggcctgttatttggtttc
ttggcctctgtaccgctaccggatacggtttctatcagatttccttagcaaacagagaga
agatcgaattgaaaagagaaaagctgtgggccagagctcatttgatgcctctactgcaag
ccgaacaggacaggaatgtagttagaagaacgtttgcctactacaagagagagggagaaa
tcatgaaggacgtcccatggtgggaggtcaaatcaacttactccaacaaagatattttcc
atcctcctcaaacggtgttattcggtaaacacattgacagggacagcggactgatgtacg
caaactcacgtaacaggtttgggcctgatgaggaggagaccaaatgaagtatttatttat
taagactagagcaatatattatttttattgtccattaatggtgttcatgcataaagtttc
gatatggagctgaggagtgtcgatcacattgaccacaatcgattcactacaactactcgt
tacgctgataccatttgtagactgttggactaattccccgttcatttcattttattcatt
ttattcattgttgtttcaatcacccgcccaacggttcccactaacctcctctccggtggc
cttccgtcgctcgattcttccttttacagtaaccaacatac

ORF

ATGTTCTCCAACAGGCTCTTGTTGTTTGGCCTTTTCTATCTGAGACTGATATCAACCTTCTATGGTATTATTTCA
GACTGTGATGAAGTGTTCAACTACTGGGAGCCACTCAACTTCATGCTTAGAGGGTTTGGAAAACAGACTTGGGAG
TATTCTCCAGAGTATGCCATCCGATCTTGGTCCTATCTAGTGCCACTTTGGATAGCAGGCTATCCACCATTGTTC

CTGGATATCCCTTCTTACTACTTTTTCTACTTTTTCAGACTACTGCTGGTTATTTTTTCATTGGTTGCAGAAGTC
AAGTTGTACCATAGTTTGAAGAAAAATGTCAGCAGTAAGATCAGTTTCTGGTACCTTCTATTTACAACCGTTGCT
CCAGGAATGTCTCATAGCACGATAGCCTTATTACCATCCTCTTTTGCTATGGTTTGTCACACTTTTGCCATTAGA
TACGTCATTGATTACCTACAATTACCAACATTAATGCGCACAATCAGAGAGACTGCTGCCATCTCACCAGCTCAC
AAACAACAACTAGCCAACTCTCTCAACAACTCCAGTCAATATTTGATAGTAGCAAAGGCAGTTTTTTGGTACACA
GTGGGTGGATTACTAGGATGGCCTTTCGCATTAGCTTTGGCTCTGCCCTTTGCTATCACCATTTTTGTTCGAAAA
GTTTACTACAAGGAATTGCATCAATTGGCTGTTATTGGCGCTCTAAGCGTGTTCATAGTCCTTCTTATCCTTGCA
TATGTGGTCCAAATTGACTCAATATGTTACAACAAAACAGAATTAATTCCTCTTAACATTGTTCTTTACAACGTC
TTGAATACCGATGACTCTGTTGGTCCTGATATTTTTGGCACTGAGCCAGTCTCATACTACATCCTCAATTTGCTG
CTAAACTTCAATTTCTTAGCATTGTTGGGATATGCAAGTTTGCCCTTGCTGCTAATATTTAACTTTCTCCCCGAT
TTCAATTTGAGGAGTTCCAATGTTAAACTTTTTGGGCATGAGGGCGGTTCTAATAAACTTATAACTTTGTTTGCC
CCACTATACCTTTGGTCTGTCATATTTTTTACCCAACCACACAAAGAGGAAAGGTTTCTTTACCCCATTTATCCA
CTTATCACTTTGGGTGCTGCATTTGCTACTCATCAATTGATTCGAATTCCATCATTATTATCAACTGTATTTATT
CCAAACAAAAGGGTTCTTCACAAAGCAATTAATCTCACGATTATCATCTTGTTTTCATCATTTATTGTTGTTATC
TCCGTCTTAAGAGTTTTTGCTCTTATTTCGCATTATAGTGCTCCGTTATTTGTCTATCAACATCTGGACCAGTTT
GCTTCAGATACTCCAAAAAATGTATGTGTTGGTAGAGAATGGTATCACTATCCTTCTTCATTTTTCCTACCTCCT
TCTATGCGTCTACGCTTTGTTAGGTCAGGCTTTAGTGGGATGTTACCAGGTGATTTTGACGAATCCGTTTCAAGG
CTTTCTTCCATGTCTAATATACCAAGAGAACTGTATAATAACAAAAATCTTTTTGAATCCGACAAGGTCATCCCC
TTTGAGGAGTGTGATTTTTACGTTGACATTTCCAAAGCTGTGGATTCTGAACAAAAAGAAATAGCAATATTGGCC
CCATCCTACTCAGATCAAGATCCCGCAGTTTTATTAGATGATTGGTCCTTGGTACACTGTGATAAGTTCATAGAT
CAGGATCACAGCAGAGGCCTGGGAAGAATCCTTTATCTTCCAAAACGCTTCCACAGAAAAATGAAGACAAATCTG
AGATACCAGCATTACTGTCTATTGAAAAAGAAAAGTAGTTAA

Downstream
aatattaggaattgtacatatgtttcaatacgctcaaccaaactgtttcaactcaatgtg
actttcttgttcgtccgtatcggaagaccagtcattgttgactcgtaacttgactaattt
ttcagtcttttctttctcttcttcactcaacgtgttgaccaaatccttaaataaggccgg
gttaaattgtattaaatccaatagcttagaatggcagtagtttcttactttcacatcgtc
atctccaaccttatcgtccacccagagtatcaacggcaaagtcagacccaatgctgattt
caaattgtcttcgtctgtcagttgcttggtgaagacaatgaaaatctcaagaatcagtct
aggatcttcaattttgtctgagtcgtttttttctcaagttattcttcagcaagcttatcaa
tgcaggaagcaacttgcgcacaatgtttttagcaccatcaacgtttccagctgttgcttg
tttgacgacatttctaacaaactgtgtcgtaaaggttagttgttgcttggagatcaaact
tttacacaaatcttgtgcaaacacctcacagtcctcgttggtgaaaaaaaaggtgccttc
cgccgacagtaaaatcattcttatcagagttgagtgtttgttgctcaacaggccaggttt
tatctcattgtagaacttgatcaaaatatccgatcttcctttcgaattcagtgatactac
taccagctttagtaatggcaaagtgactgaaatgatatctgtattgttcaactcgaagca
taagcttataacatataacattgaagatagcaaatcaatttggatcatttcgggaagatg
gctgattaacttatgaagcaagctaaatgatttttttcaacaaagtcagatccttatcaga
aaatggtccaagtttctctctttctatagtcttctctgacacaaacggaagaatttgaat
aacaggtaaatagctcaatctcaatagttcgaatagcttgt

AA
MFSNRLLLFGLFYLRLISTFYGIISDCDEVFNYWEPLNFMLRGFGKQTWEYSPEYAIRSWSYLVPLWIAGYPPLF
LDIPSYYFFYFFRLLLVIFSLVAEVKLYHSLKKNVSSKISFWYLLFTTVAPGMSHSTIALLPSSFAMVCHTFAIR
YVIDYLQLPTLMRTIRETAAISPAHKQQLANSLNNSSQYLIVAKAVFWYTVGGLLGWPFALALALPFAITIFVRK
VYYKELHQLAVIGALSVFIVLLILAYVVQIDSICYNKTELIPLNIVLYNVLNTDDSVGPDIFGTEPVSYYILNLL
LNFNFLALLGYASLPLLLIFNFLPDFNLRSSNVKLFGHEGGSNKLITLFAPLYLWSVIFFTQPHKEERFLYPIYP
LITLGAAFATHQLIRIPSLLSTVFIPNKRVLHKAINLTIIILFSSFIVVISVLRVFALISHYSAPLFVYQHLDQF
ASDTPKNVCVGREWYHYPSSFFLPPSMRLRFVRSGFSGMLPGDFDESVSRLSSMSNIPRELYNNKNLFESDKVIP
FEECDFYVDISKAVDSEQKEIAILAPSYSDQDPAVLLDDWSLVHCDKFIDQDHSRGLGRILYLPKRFHRKMKTNL
RYQHYCLLKKKSS

SEQ0161
P. pastoris homolog of *S. cerevisiae* **ALG8 (YOR067C)**
Glucosyl transferase, involved in N-linked glycosylation; adds glucose to the dolichol-linked oligosaccharide precursor prior to transfer to protein during lipid-linked oligosaccharide biosynthesis; similar to Alg6p
S. cerevisiae null mutant: Bud8p-GFP distribution: abnormal; carboxypeptidase Y (Prc1p) modification: decreased; resistance to mercaptoethanol: decreased; resistance to tunicamycin:

normal; resistance to L-1,4-dithiothreitol: decreased; budding pattern: abnormal; cell shape: abnormal; competitive fitness: decreased; resistance to hygromycin B: decreased; resistance to ethanol: decreased; viable

Chr 3, 0999

5' region

gttgattctcttttcacctggtgccatttgcttaggtagattatgcctatcgattttgta
ggtcaaaagaaatacaggctactttgctttctttcccctgaagtgaagggaaaagacgct
ctatctatcccactatcaggatacagctttctacaacgtattggattctgatcctagctc
tgattacttgaaagacagcatcagaattggggttggaacccagttacaatataagacagg
tcgcaagttaggttcaattgacacttcacaatggagattctactggaagaagaaaggtga
gggtgttttgtatccactaaggacgttaaacgactttgttagatttgtcgatagccttag
tgagttctcattcaatttaataagactagttgtgagcactgatgatcctgttaatgaatg
tagaggggagagacgaaaggaaaaacgaaaggcaaagaaggaaaagaaaaagaaagataa
gagaaaagataagaaaagaaaagaccacaaacttgggacggaactgccacaatttgacat
gggagaggaaatcaatgaccctaaatggaacgtgaatgaggacctctctttagcaggttt
tgattgtctagcaattgcgacttggaataatgaatccagagagaaaccccctggatggctt
atcgcagggcgaatctgattctgagtgcgaacagaaaaacagttccatcgtaatgaagcc
agatgaagcgttatctgaaactgtacgagagcggtataaagagttgttgctccttgacga
ggaagtggcgaattgcatgtcagaggacacttctggaggggaggaggactacgcgccaag
atctcctttatttatcgaatcaatgttgtcggattctgactttgagcaagtcttttgacg
ctaaccatggtaatagcctgtctttgtttattcaaggggccacctgtcctttcccaacct
tttttattttagttccattatttgtattacttaatgtttaat

ORF

ATGTTTGATCTCATTTCTCTGCACTTTTTCTTACATCAGCAAGCGATCGATACCTATTATCCGATTTTGTTTTTT
GCTTCCAGTCCATTCACCAAGCCCTCCCATATGAGCGGGAAACGTAATTTTTCTCTATGGAATATCTGGGTAGCA
AGCACCCTTCTCAAAGTGCTTCTATATCCGGCATACCATAGCACGGATTTTGACGTACATCGGAATTGGCTCGCA
ATCACAAACAAGTTACCGTTGAAGGAATGGTATCTGGAGAACACTAGCCAATGGACTCTGGATTACCCTCCATTC
TTTGCTTATTTCGAATGGTTGTTGAGTCAGTTCGTTCCAGCATCTGTTGCTGACGACGGTTGTCTTGATATAGTT
GATGTGGGCAACTACGGTTGGCCTACAGTAGTTTTCCAAAGAAGCACAGTGATTTTGAGCGAGATCGTACTCTTC
TTAGCATTACAGAAATACATCAACATTTCCGCCGGCAAAGAAAAGGCTAGAAGCTTTGTTGTTGCTAGTTCAATT
GCTCTTTCGCCAGGTCTTCTGATAGTGGATCATATCCACTTTCAGTACAATGGCATGATGTTTGGAATCCTAATT
TTCTCTCTACTAGCAGCCAAGCAAAAAAAATATCTACAATGTGGAGCACTGTTCAGTGTATTGCTATGTTTTAAA
CACATATTTCTTTATATAGCCCCGGCATATTTTGTGTTTTTGCTCAGAGTTTATTGTCTGGATATTCATGAGACC
TCCTTCAAGACACCACGATCACTATTGAAGTCTGTGCGATGGTCCAATTTGTTTAAACTGGGATTCACAGTTATT
ACCGTATTTGTCATTGCATTTGCTCCCTTCGCCTATTACGGCGTGATTCCTAACCTAATCTCAAGACTGTTTCCG
TTCTCAAGAGGGTTAACTCACGCTTATTGGGCTCCAAATATTTGGGCCCTATATTCTTTCCTGGATAGGGTGTTG
GTTCAGTTATACCTGCATGTTCCAGGTATTTCCAGCATCATCTCAAGATACATTGGACCAGAAAGTCTCGTTTCA
AGGCTGAAAAACTCTACAGCATCTACTAAGGGCTTAGTGGGAGATGTGGAATTTTTTATAGTTCCAACTATAACT
CCCAAAATGTCGTTTATTCTAACTCTTTTCTATCAAATTTTGGCCGTACTACCTTTACTGCTCTTCCCGACGTTC
AAGAAATTTTTAGGTTCCATTACATTGTGTGCATTTGCAAGTTTCTTATTTGGATGGCATGTACACGAGAAAGCC
ATCATGCTTGTCATTATACCATTTTCATTTATAGCCGTGAGTGATAGAAGGTTATTGTACCCTTTCCATACTTTA
GTCAACGCAGGCTACGTGTCTTTGTTTCCACTATTATTCAAGTCTCCAGAATGGCTTGTAAAGGTCCTGTACACT
TTGGTGTGGTGCATTATTTATTTCCTCTCCTTCAATGAGGTATCCAAGCTGTCAAAGTCATTATCTCGACGAGTC
TTTTACATGGATAGAACAAATCTGATATATATTCTTGGACTGATCCCATTGACATTAACCATTGGTGTATTGGAT
GTGCTATCTGAAAACTTTGAATTATTGAAGAGACTTGAGTTTCTAAGGCTGATGATGTACAGCGTTTACTGTGCA
CTGGGAATAATCGGCAGTTGGAACGGATTAAGCTGGATCTATTTCTTGGATGATACACTGAATGGAGACTCGAGT
GAGGTCGCATAG

Downstream

gacggaatgcgcagcaattaatatgaacatctcgcgtatcataatagacgtatcattttc
accatcgttttcctcgtctcttctttttgcgtaatcggacaacgctgcactctggacaaa
accacttcccctcgggtgccctcgtcaagttcacacatttgtaatgaaaccattcatact
tgcaaccgggattgtcacaggcaaccatctttcctaaacctggtcctctacataggcaat
agacgtcctcggcgggtggactgggcgctggtattggtgaagggagggctggaggggtta
taataattttcttagaaggtttgacgactttggacttttgttttccattttgcacttttc
ttttagtttttgtagcatttattgatttgcgattttgtttagggagggatatgtgaattt
taatggcatcccgattctctacagggagggtattaggatgcagatcccctcttttcaacc

catgttctttcaaaaactgtttcttgaactgttcaaatgacgctcgtgtgttattctgct
cgttgatttttagctgcttcaagcgctccatacagagagaatcactattcaagtagtctt
gatggttatctagtttatccttgagatgttgggactctgcaacaatctcctttgagtatt
tgcttgccaaatttcgaatacttacaatctttgctaacacctggggatcattaaggtccg
agtctttcagtatatgatctagctggttcatgtactgatccacccgtaaattgagaaatt
ggattagccataatgatctaaccacatcgcaaggtaaatgatcaagcgctgaaatgttag
tacaatgtcattatcccaataacgtgacaaatataagtggtactcactggacacaaaaga
gttgctcaagtttacgatattcgaagttaacatgacattatttgatatttagtgttatga
ggaggtcttgatgaatttgatatcaaccgcgagaattgtt

AA
MFDLISLHFFLHQQAIDTYYPILFFASSPFTKPSHMSGKRNFSLWNIWVASTLLKVLLYPAYHSTDFDVHRNWLA
ITNKLPLKEWYLENTSQWTLDYPPFFAYFEWLLSQFVPASVADDGCLDIVDVGNYGWPTVVFQRSTVILSEIVLF
LALQKYINISAGKEKARSFVVASSIALSPGLLIVDHIHFQYNGMMFGILIFSLLAAKQKKYLQCGALFSVLLCFK
HIFLYIAPAYFVFLLRVYCLDIHETSFKTPRSLLKSVRWSNLFKLGFTVITVFVIAFAPFAYYGVIPNLISRLFP
FSRGLTHAYWAPNIWALYSFLDRVLVQLYLHVPGISSIISRYIGPESLVSRLKNSTASTKGLVGDVEFFIVPTIT
PKMSFILTLFYQILAVLPLLLFPTFKKFLGSITLCAFASFLFGWHVHEKAIMLVIIPFSFIAVSDRRLLYPFHTL
VNAGYVSLFPLLFKSPEWLVKVLYTLVWCIIYFLSFNEVSKLSKSLSRRVFYMDRTNLIYILGLIPLTLTIGVLD
VLSENFELLKRLEFLRLMMYSVYCALGIIGSWNGLSWIYFLDDTLNGDSSEVA*

SEQ0162
P. pastoris homolog of *S. cerevisiae* **ALG10/DIE2 (YGR227W)**
Dolichyl-phosphoglucose-dependent alpha-1,2 glucosyltransferase of the ER, functions in the pathway that synthesizes the dolichol-linked oligosaccharide precursor for N-linked protein glycosylation, has a role in regulation of ITR1 and INO1
S. cerevisiae null mutant: Bud8p-GFP distribution: abnormal; budding pattern: abnormal; cell shape: abnormal; resistance to ethanol: decreased; viable

Chr1-4, 0475

5' region
taacgtcgatagttgaattttgtatgaaatagagattagagtcccgctccaattcatagc
tcgtatcttcatacatattattctcaagaataatcggtatcactctgatccctctactct
gcaagagctctcctaaaattgggtagccattgggatcgaattctaaaccctgatatgtgc
tattggatagctgagcaacccgtaacgaagcaggggtcaaaatttgtacttgatcgatgc
tcctaactaatggtcttgataaggttcccggctgggaaactggagttgggttttgggtgc
caattgaatgcaaactcatataaatagtaatcaaatttaaggatacggatctagctgagg
ccctgatatgttcatcaattctcgtgaaaatggaagccaaatccagtcttctgtcacagg
tgtactgcttattatcccattccagttgtaggatcgtactaccatcagcaattggagtaa
cattagcactaaagggagcaggacataattgccaattaaatcctccctcatcccaatagg
cgtcaatcaggatgtttcctacagacgcgtttaaaaggttgtccacccgattcaaatacg
gttgagtgtatgtgtctctttcaaaaatgactgttgacagccggaccccaacatcaacta
actggtcaacgctgacattgtaggagacatccctttgagacctcagagcgacttcattat
caacagacaaaggcggccaaatgacacaatagacgcattgtatcaggagaaaaaaggcta
tgatcataagtctaaagagacacaggcggcgagcaacatttgggctttaacaatacagg
cctcgaatgcttagttttctgatatctcagatctcaactggcaaggttcaccatcactga
aataggctgggggattccgcaacggtgctgcaactcagaggaacctagagcttgaccaat
taacagcgtaatttacatacaaaatcattctatcaaccaga

ORF
ATGGCAGGAAGTAATAGCATTTCGCTATTGTCTCATTTGACAGTTATTGTAATACGTAGCTGCTTTCAAAAGGTG
ACTAACGTGGTAAGGAAACCATTCATCGATGAAATATTCCACATTCCTCAGGCAAGACAGTATTGCAGAGGGAGA
TTTGACGTTTGGGATAATAAAATCACGACTCCACCGGGTTTATACTGGTTGGGTTATGTCTGGGTAAAAATTCTA
GCAGTGCTCAATGGCGGCGAGTTTAAATGTGATACCAATACTCTGAGGGATATCAACTTCGTAGGATTTGTTGTG
TTACAGCTTCTGATTTTCTACCTTCAGAAGGGGACTACGGGTAATAGCTACTCTACAAGTTCCATATCATTGAAT
CCGTTGATCACATTATACTATTCGTTATTCTACACTGATGTTTGGTCTACAGTTTTTATTGTGGCCTCTTATGTT
GTTATTGTCAAACAGCCGTTTGGTAAATATAGAAGTGCGACGATTTCAGCGTTCATAGGACTGGCAAGTGTGACC
TTCAGACAGACAAACATTATATGGAATGCCCTGATTCTAGCAACCTTCATTGATCAACAGATCGACCCGAAAGAT
CGAACGAATTCATTCTCCGACATAAAACTCTTTATCGCTGAAACATGGAGAAACATCCTGGGAGTACTTCCGTTT

GCCATCAACTTTGGACTCTTTTTGGCATTTGTTTACACCAATGGAGGAATAACATTAGGTGATAAGCAGAATCAT
GTTTTCAGTGTCCATATTGCTCAATTATTTTACTTTACCTCTTTTGTTGCAATGCTTTCCATTCCATTGTGGATA
TCGCCCTCTTTTTTCCTGGGATACTTGAAACTTTTGAGACAAAACATAATCTCGACCATCATCTCGTGGGCTGTG
ATAGCTTTACTTGTGCATTACTTTACTGTTGTGCATCCATTTTTACTTGCCGATAACCGTCATTACACATTTTAT
ATTTGGAGAAGAATCATCAACTTGACAGCCTATTCCAGATATATGCTAGCCCCCGCCTACCATTTCTCCATTTAT
GTCACCTTCAAAATGCTTGCTGATAACATTCTCTCGCTCCCTAACGAGCAAGAGATACAACAGCAAGAGACAGAG
TACGAGCAGAAGGAAAGGTTTACAGAGGAGGATTTTGCTGATGAAGGGGAAGATCCCATCAAGCCCACCTTCATT
ACCATTGCTGCACTTTCAATATGCACTGCTCTTACTTTAGTGCCGTCTCCCCTTTTTGAGCCACGATATTTCATT
ATTCCATTCACTTTCTGGAGGCTTCTAGTCCGTCCTTCGGACAGCACTTTGTTTGAGAATGAGTCTTTGAAAAAA
ACCAACAATCGCACAAGATTATTGTTTGAGTTAGCATACTTTGTCGCTTTCACGTATCTTCTTTACGAAGTATTC
ATCCAATACACTTTCAAATGGCCCAGTGAGCCCCTGGAATATCAGAGAATCATCTGGTGA

Downstream
tgttgacaatgttttccttatgctgttccaacacccgtttcttcaaaacatcaaagtaac
ttacctcttgcttagtaagtgggtcaatgtatgctaaaccagcaaaagtcccatacatat
cattctgataatccaacttccagttcttcgactcgagatctccagctacgagtaataaaa
aatctccccaattgcgagcaaccagatatttggtatcgaaatctctcccaaacagtatca
cctgacccactgttcctttctcatctggagcaagatccagagccacatggttccccatat
aatccgttagcaccggaatccagttcgagtgactgtaactaggaatgattgtcttttctg
gtatagaacgctgtggtggtagagttagcggtttgatctgtgtcgagtcgtcgaagggag
aacttgacaagtggccggatttgtgattctcattcttcaaataccctaattgtggctggg
agatgttattaacgctgtgggcgctcttagctttcgccaggtccatttcgattttgtttg
tgatttttctccaatttttggtcatagagataattgagtccaaggacatgagttctaacc
cgaatatcagcccactggaatcctgactagattcggtatactcttgaccatcatggatgg
aaatggatttgacgaagcaaggggggaaactcaattcccaaatctttttgagcacaaatta
tgtcggggactgtcacaggtttagcaaatctggaggatagttcgtgtttcgaacaccact
cgtcaagcaattgccatgcaaaaacgacttcctcagtgccttctgctccttcagcagctt
tgactggatccactgatagtttattgaatagtgtgtcgtcaggatatgatgtcaatgagg
ctttgacaaatgtaccggagttattcttagccagaaaatcctccgatttaatcgttgttt
ggttcgatttgtaatcagcatatctatcgttggtggtagtg

AA
MAGSNSISLLSHLTVIVIRSCFQKVTNVVRKPFIDEIFHIPQARQYCRGRFDVWDNKITTPPGLYWLGYVWVKIL
AVLNGGEFKCDTNTLRDINFVGFVVLQLLIFYLQKGTTGNSYSTSSISLNPLITLYYSLFYTDVWSTVFIVASYV
VIVKQPFGKYRSATISAFIGLASVTFRQTNIIWNALILATFIDQQIDPKDRTNSFSDIKLFIAETWRNILGVLPF
AINFGLFLAFVYTNGGITLGDKQNHVFSVHIAQLFYFTSFVAMLSIPLWISPSFFLGYLKLLRQNIISTIISWAV
IALLVHYFTVVHPFLLADNRHYTFYIWRRIINLTAYSRYMLAPAYHFSIYVTFKMLADNILSLPNEQEIQQQETE
YEQKERFTEEDFADEGEDPIKPTFITIAALSICTALTLVPSPLFEPRYFIIPFTFWRLLVRPSDSTLFENESLKK
TNNRTRLLFELAYFVAFTYLLYEVFIQYTFKWPSEPLEYQRIIW

SEQ0163
STT3 (YGL022W): Subunit of the oligosaccharyltransferase complex of the ER lumen, which catalyzes asparagine-linked glycosylation of newly synthesized proteins; forms a subcomplex with Ost3p and Ost4p and is directly involved in catalysis
S. cerevisiae null mutant: inviable

chr1-4_0685
5' region
tttgagttgtgtttaccacaattaaaaaggcttccgctgatggaatccgacatatttccg
tcccttcattttcttttaaaacttgtacaagagatagagagtgaagaggtgcaaaagaag
ttggactcattggtaagaagattttcggatacaatttcagaagctttgcagtatgagttg
agtttcactgttgataaatccaaccttcccctgccagcaagttgggtatttgagggcgag
agcatttcgtatggttcaacgcagtctactcaaccattaagagcaacacctcctcgtatg
aaattgagcaaagaaaaacaatacagcttctctcccaagacaccgctttggaaaaggtgg
ttcaattcaccaattaagaaacctcaaagactaaagggaaagaaagtccgatggagcatg
ccctacactaacgaaaatatacatgaagacggggagaaggattttaacgaagatgatgaa
gagcaacagcatatggaaaccagtcttgaagattcagacgtagacatggataaatttgtt
gacgctatggatatttcaccgttgccagatgccgcagattcgtcattttctacggttaaa
gcttccagacagagttcactaaccacaagaaaactaattccgtccaaacaaagtaagagc

cttctaagctctttgaagaacgctgaagctcaaccagatgaaacagaaatagttccaccc
ttaggtgcaccctcacgaatgaatttggtagaacctaatatggtgcttgaagataataat
aaatagatcaatcaactcaccgaacaaatgattatataattgggctctcctttcctgcta
gcccttgcacttcccttccctagtaaatacatccgagagcatccttcgcgaataccttcc
aacacataaacagtacactactccgccgaaaaagacacgttggagcgactagcttaaaat
actctccaccgccaaatcctccctcaacggatctccaaca

ORF

ATGGTTACTATTAACGATCAGGGGTATATCACCGTTAACGATAGGGTGCTAAAGCTCATAAAGAGCCTTTTGATC
GTGCTTATTTTCATAAGCATAACAATTGCAGCAGTTTCATCACGATTGTTCTCAGTGATTCGGTTTGAGTCAATT
ATCCATGAATTCGACCCATGGTTCAACTTTAGAGCCACAAAGTATTTGGTTCACAATGGATTTTACAAGTTCTTG
AACTGGTTTGATGACAAGACTTGGTATCCTCTAGGACGAGTCACTGGTGGGACTCTTTATCCAGGTTTGATGGTC
ACTAGTGCCGTCATTCACAACCTGTTGGCCAAGATCGGACTTCCAATTGATATTAGAAACATCTGTGTCATGTTA
GCTCCTGCTTTCAGTTCATTGACTGCCATTGCTATGTACTTTTTGACCTTGGAACTGACCAATGACTCTGAATCC
ATCGCCAATGGGACTGCAAAGGCAACTGCCGCTTTGTTCAGTGCAATTTTCATGGGAATTACTCCAGGTTATATT
TCTCGTTCGGTGGCAGGCTCTTATGATAATGAAGCTATCGCTATCACTTTGTTGATGGTGACTTTTTACTTCTGG
ATTAAAGCTGTTAAATTAGGATCCATTTTTTACAGTTCCGTCACTGCTTTGTTTTACTTTTACATGGTCAGTGCT
TGGGGTGGATATGTTTTCATCACAAACCTGATTCCTCTTCACGTGTTTGTTCTACTTTTGATGGGAAGGTTCACT
CACAAGATTTACGTCAGTTATACTACTTGGTACGTCCTTGGAACTCTTATGAGTATGCAGATCCCCTTCGTCGGC
TTTTTGCCCATTCGTTCCAATGATCATATGGCTCCATTAGGTGTTTTCGGCCTCATCCAACTTGTTTTGATTGGA
GATTTCTTCAAATCTCAACTGTCCCGTAAAGTTTTTATCAAGCTGGCGATTGCTTCTGGTGTCGTTATTGGTATC
CTAGGTGTTGTTGGCTTGGTTCTTGCTACTAAAATTGGTCTTATCGCTCCATGGACCGGAAGATTCTACTCATTA
TGGGATACCAATTACGCAAAGATTCATATTCCAATCATCGCATCGGTCTCTGAACATCAACCTACCCCATGGGCT
TCATTTTTCTTTGATTTAAACTTCTTGATTTGGTTGTTTCCAGTAGGTGTGTGGTTCTGTTTCCAAGAATTGACT
GATGGAGCAGTGTTTGTGATCATCTACTCTGTTCTTGCCAGTTACTTTGCTGGAGTCATGGTTAGACTAATATTG
ACCTTAGCTCCAATTGTTTGTGTATGTGGTGCTATTGCCATTACTAAACTTTTCGAAGTGTATTCTGATTTCACG
GACGTTGTCAAAGGCAAATCAGGAAACTTCTTTACCTTATTCTCCAAGCTGGCTGTTTTGGGCTCTTTTGGATTT
TATCTTTTCTTTTACGTCAAGCACTGCACATGGGTCACTGAGAACGCGTACTCATCCCCTTCGGTTGTTCTTGCC
TCTCACGCAGCAGATGGTTCTCAAATTTTAATCGATGACTATCGTGAAGCATATTACTGGCTTCGTATGAATACT
CCGGAAGATGCCAAGGTAATGGCCTGGTGGGATTACGGTTATCAAATCGGAGGAATGGCTGATAGAACTACGTTT
GTTGATAACAATACCTGGAACAACACACACATAGCAACTGTTGGTAAGGCAATGGCTGTATCTGAAGAAAAATCA
GAAGTCATCATGAGACAACTCGGAGTTGATTACATTTTGGTCATTTTCGGGGGGTGTTCTTGGATATTCTGGCGAT
GACATAAACAAATTCTTATGGATGGTCAGAATTTCCGAAGGTATCTGGCCCGAGGAAGTCAGCGAAAGAGGATAC
TTTACCCCCAGAGGAGAATATAAAATTGATGACAATGCAGCTCAAGCCATGAAAGACTCAATGCTGTACAAGATG
TCATTTTACAGATTCGGTGAACTTTTCCCCTCTGGAGATGCCATAGACCGTGTCAGAGGTCAAAGATTGAGTCGC
TCTTACGCTGAGTCTATAGATCTTAACATTGTTGAAGAAGTATTCACAAGTGAGAACTGGCTGGTTCGTCTTTAC
AAGCTCAAAGAGCCAGACAACTTGGGAAGAAGCCTACTGACCCTGAAGGATAACGAGAAAAAGTTGGCCACTAAA
AAAGGTCGCAGATTGAGGGTCAACAAGAAACCCTCATTGGATCTACGGGTGTAG

Downstream

tgtagttgtgtattttagatagacagagctctggggttagaatgataacatattgcgtgc
ttatgtaatgatgttcacctttcgtaggtctcaaaagttagcctagtcttagcccgtaac
ttgctaatgactgcaccaaaacctcttcctacgctaaatgaagaccgccaatctgaattg
gtcaataatctgagccattgggcgctaggaaatggtttagcgatgtaccccaccaatttt
gagcttcataacactggctatgcaccagttacccttttccaacaccatttccaagagga
caatttgagaaggcattggccgttcaggaagactttaatgagctctatgctcaggtagtc
aaaaatcaggaatggctgggctcaattctggaggacttatctcaatttgatagggatttt
actggtaagttatgggaaatttacaaggaggccaagaagattggaattgttcaaccagtc
tcattgggcttgtttcgatccgattacatggtggatacgttatctgctccctcatttgat
ggtataaatggaaaaatcaaacaaattgaattcaacactgttagtgtatcctttgggggg
ttatcacccaaggttgcccagcttcacaggtacttaaatgagaatggtaattaccacaat
gagggtaagcttcattttgaagatgatgagcttccaaattctgaatcgacagtttctttg
gccgatggtttggcaaaagctgccgcgtattataatactagtgagtctgttgagtcctct
gttgtgctagtggtagtgcaaaataatgaacgaaacgtttttgaccaaagggcccttgaa
ttcgaactgttaaaacgtcataaagttagatctattcgattgcgcatggaagatatatct
gcaagtattgatattgatgagcaaacgagaagaattcgactgaaaacgacgggggaagaa
gtttcagttatctattatcgttcagcttatgcaccaagtg

AA

MVTINDQGYITVNDRVLKLIKSLLIVLIFISITIAAVSSRLFSVIRFESIIHEFDPWFNFRATKYLVHNGFYKFL
NWFDDKIWYPLGRVTGGTLYPGLMVTSAVIHNLLAKIGLPIDIRNICVMLAPAFSSLTAIAMYFLTLELTNDSES
IANGTAKATAALFSAIFMGITPGYISRSVAGSYDNEAIAITLLMVTFYFWIKAVKLGSIFYSSVTALFYFYMVSA
WGGYVFITNLIPLHVFVLLLMGRFTHKIYVSYTTWYVLGTLMSMQIPFVGFLPIRSNDHMAPLGVFGLIQLVLIG
DFFKSQLSRKVFIKLAIASGVVIGILGVVGLVLATKIGLIAPWTGRFYSLWDTNYAKIHIPIIASVSEHQPTPWA

SFFFDLNFLIWLFPVGVWFCFQELTDGAVFVIIYSVLASYFAGVMVRLILTLAPIVCVCGAIAITKLFEVYSDFT
DVVKGKSGNFFTLFSKLAVLGSFGFYLFFYVKHCTWVTENAYSSPSVVLASHAADGSQILIDDYREAYYWLRMNT
PEDAKVMAWWDYGYQIGGMADRTTFVDNNTWNNTHIATVGKAMAVSEEKSEVIMRQLGVDYILVIFGGVLGYSGD
DINKFLWMVRISEGIWPEEVSERGYFTPRGEYKIDDNAAQAMKDSMLYKMSFYRFGELFPSGDAIDRVRGQRLSR
SYAESIDLNIVEEVFTSENWLVRLYKLKEPDNLGRSLLTLKDNEKKLATKKGRRLRVNKKPSLDLRV

SEQ0164

GLS1 (YGL027C):Processing alpha glucosidase I, ER type II integral membrane N-glycoprotein involved in assembly of cell wall beta 1,6 glucan and asparagine-linked protein glycosylation; also involved in ER protein quality control and sensing of ER stress
S. cerevisiae null mutant: viable; decreased resistance to hygromycin B and PM02734

chr1-1_0215

5' region

atgattcaagtcagccctcccagaaataccttatggaaaccgtgaccattaaagacttgg
aaagtggtcaattggatatagacctacttttatcacatatctccctattgaaggaaaaac
tgtctcatttacggtaccagatggtttcaagtcttaaactgctagcctcagttgacagta
gtaccggcccaagtgggtttttccagacggtctcaactcaggtttcagacataaaacggg
atatcgaagagtatcaccaagaactgcaaaaagttttaccggtggtacgattttgcaaga
tcaaaatgggcctgagccccaatgattctgtgaaagtcacgaagcatgaggttaaaatcc
agtaccccaatggttctggggctctagtcggctctgtaccagggtctaacgacaaacaag
atagaagggttctacagcagaaaaagactgctgtgaggaagccaagaaaaaatattaaaa
agactccaaatataaaccaatcaggtcatactccttccagtctaggcaatactaacggaa
atgctcaaacgccaatgtaccagatgaatttacagtcacaacagttacaacaacctccac
aaatccaacagcagctcccacaacagccgcctttacaacaatcacaacaacaacacccccc
aagcacagcaacaaccacaacaaattcaatatgaacaatacgctcagcaggcccagcaag
attacggctctgctaatcagcccattctactgtgattgttccgtgcgagtctataataag
gaagtagtatggaaaaaccattttattgtgatcctctcaatgtaaacaacgagctattca
tcttagcattagcaccatcagacactctctctgctgttgtctcgaacatctctttttgct
ccatgggttggttctccttgggcctcaacccgaatctctgttcagtaccctcgactgttt
taagatctaccgagctccaccgtttcatctttttcactc

ORF

ATGAACCTCTTCAACCGCCGGATCCTAAGCCTCATTTCGGTTTTGGTCGTTGCCATTGCTTTTGAAAATGTCCCT
TCTGATGAATTGTCCCTACCGGAGGCTTTTCAGAAGATCTCCGATCAATCGCTCTTGTGGGGTCCTTACAGATCT
AACTTGTATGTCGGAATCAAACCAAAGATTCCCCATAGTTTCCTTTCAGGTCTCATGTGGTTCAATGCTGACGAT
CCTGAAGGAATCGTAAAGCTGAGGCATTCTTGTGAACAGGACCCAGAGATTCAAAGTTTCGGTTGGGTAAAGTAC
GATGCAAGGGTTGGAGGAAGACACATCATCAAGGACAGAGGCTGTAAAGTGCTGATAAAATCTGACTTTGTTAAG
ACCAGTGATGGTAATTGGGCCCTGAAAATTACTGGAGTGCCAAAGAAAGGACAGGAAAATGTCAAAACGTCTCTG
ATATTCTATGCTGGAACTGAGGAAGACCAGGATAACATGTTAATGTTTGCAGGCAATAAAGATGAATTTGGGAAT
GTTAACAATGATCTGACAAGATTGACTGGCGTCTCCAAAGTTTTGGGAGGCGCTTTTGAATTATTAGTCGAAGAT
GGTCCCTCCAACAGGTACCCGTCTGCTACTGTTTTGGCTGCCCCTGATTTAGATCCTTCTCTAACGCATCATTTA
TCTTTGCATGTCCCCGACGAACATTTGTGGCAGGCAAAAGAGGTGTTTATTTCATTACTTCAAGAATCTGTTGGA
AAAATTAGAGAAGATCCCTCTTTGAATGTGAGTGAAATACCAGTTGATCAATTGACAACCTTGAGGAATATTAAC
AATTTTGAAGGTAATTTACACTTCATTCAAAAGACATTTCAAGGTAAGTTTGAGTTGAACATTATTTTCAACTTA
GAAGATGCTCCTGAAAAACTTTCCTCTTCCAACATCGATTCTTATGTGGATAGAGCATTAACCCACTTTGATGAA
AAATTCTCTTCTCAATTACAGTTTCAAGCCCCATTTCATACCAAAGAATATTTGAACTTTGGTAAGGAATTCTTA
TCAAATCTAGCAGGGGGAATCGGATACTTTTATGGTGAACAGTTAGTCGATAGGAATGCATTTGTTGACGACGAC
TCCTTTGACAATGTCAAACTTGTTGGGAAACCGGAAGGTCCGTCTGAGTTATTTACTTCCGTTCCAAGCAGACCA
TTTTTTCCTCGTGGATTCTATTGGGATGAAGGATTTCATCTTTTATCCATTTTGGATTATGACAGTGATCTGGCA
CTCGAAATTCTAAAAAGTTGGTTTGCATTGATAGATGACGACGGTTGGATTGCTAGGGAACAAATCTTGGGCCCT
GAAGCCAGAAGTAGGGTACCGGCAGAGTTTCAAGTTCAAAATCCTAATATCGCCAACCCTCCTACATTAATGTTA
GTCTTTGCTAAGCTACTAAATATGGCTCATGATTCTCCAAAGTCGGAAGACGATTTTGTCTCCATCCAGGATCTT
TCCGAAAATATGGGCCATGTTCATTTAGACAATCCAGAAATACTAGTTGACTACGCAGAAGACATTTATCCTAGA
TTGAAAAAACACTTCGAATGGTTTGCCACGACTCAAAAGGGAGAGACCTTGGGGCTCAGCAGAGAGTCCAAGTAT
CCCAACCAGCTTTACAGATGGATAGGAAGAACCAAAGATTTGTGCCTACCAAGTGGACTAGATGATTATCCAAGA
GCATCAGAACCAGACTCCGGTGAACTCAACATTGATCTTCTTTCATGGATGGGGCTAATGTCTCGATCAATGAAG
AGCATTGCGAAGTTGTTGAATAAAGCCGACGATGTTCACTACTTTGAGACTATTGAACAGGGAGTGCTATATAAC
TTGGATCAGATCCATTGGTCTGAAGAAGAGAAGAGTTACTGCGACATAACGGTTGATGACGACGACAATGATGTT
TTTGAATGTCATAAGGGATACGTGACTCTTCTACCGTTTGCTTTGAAACTTATTCCAGAAGATTCTGAACGGTTG

ATTCATGTATTACAAGATCTGAGAGATCCAGATATATGTTGGTCTCAATTCGGCATTCGATCTTTATCCAAATCG
CACCCTCTTTTCCACAGTGGAGAAGACTACTGGAGGGGAAATATTTGGTTGAACATTAACTATCTAATCTTGGAT
GCTTTGAAACATTACGCAGAGTCTTCCGGTGCTCTTCCAGATGTAAAGGAAATGGCCAAGCCTATTTACAAAGAA
TTGCGAGAGATTCTAGTGACCAATGTTTACAATGAATTTAAAAGAACGGGCTATGCCTGGGAACAGTACAATGAA
GCAACCGGAAAGGGCCAAAGAACGAGACACTTCCTTGGTTGGACTTCCTTAGTCATCCCAATAATGAAGATGCCA
GAGGAACTAGCATAA

Downstream

gagaagggagaaacacatagcgtagcgggaaaagaatatttctgcacacatcccaccaga
tcaaagttagaagaatttctccatggacaattcatcatcatttagaaccaaaacaatgag
ctgtttttattctttttagatcatgaaaagtggagaatagaaggctagaccttcgcata
aaagttgatgactttctcaactcaaccccacattggtataaagaaaatttgagttttatc
tcaaattacaaaaatttgaaccaaaacacatcatctaaattctgaacttttggtaatcat
cttttctaaagaaataagtactaaaaatgaaatccaaaggttctcttgaccacttatctt
ctattctccagtctatagttttatttatacgcatgatatagtttttgttgttttctatga
atagaaaagtaccgtcattcaaatgtcttgactcagttatttcatacatgcttcctgttc
accttcagctttaaaatatgataagtatccaagtctattctcttttcgtctgattggtta
tataatctcactttctgttcaacctgatgatcatcaggtggatttcttgatgatgttgct
tcttctggcgaattgttactaatttcaatgtcatgcatatattcttccatagactcatat
tctcctagtgcatactctctagaatgccatatttttttgctttcaaaactacatctttttc
gttccttaatacaaaacaatatggattcatagctgtatctagattgcatagataaagttt
tatatcttctttatttttgactttaacaaaacataatctgcaatagctattggtacctgt
gtcaggacagtaaaaaaaattctacatctggtatttctttattcactaggaacattaaat
acatcttaatctagtctacttgccttgaccttagcatatacctccaaattaaattatttg
caccctacaccataaacaaaccctatagacaatcacatg

AA

MNLFNRRILSLISVLVVAIAFENVPSDELSLPEAFQKISDQSLLWGPYRSNLYVGIKPKIPH
SFLSGLMWFNADDPEGIVKLRHSCEQDPEIQSFGWVKYDARVGGRHIIKDRGCKVLIKSDFV
KTSDGNWALKITGVPKKGQENVKTSLIFYAGTEEDQDNMLMFAGNKDEFGNVNNDLTRLTGV
SKVLGGAFELLVEDGPSNRYPSATVLAAPDLDPSLTHHLSLHVPDEHLWQAKEVFISLLQES
VGKIREDPSLNVSEIPVDQLTTLRNINNFEGNLHFIQKTFQGKFELNIIFNLEDAPEKLSSS
NIDSYVDRALTHFDEKFSSQLQFQAPFHTKEYLNFGKEFLSNLAGGIGYFYGEQLVDRNAFV
DDDSFDNVKLVGKPEGPSELFTSVPSRPFFPRGFYWDEGFHLLSILDYDSDLALEILKSWFA
LIDDDGWIAREQILGPEARSRVPAEFQVQNPNIANPPTLMLVFAKLLNMAHDSPKSEDDFVS
IQDLSENMGHVHLDNPEILVDYAEDIYPRLKKHFEWFATTQKGETLGLSRESKYPNQLYRWI
GRTKDLCLPSGLDDYPRASEPDSGELNIDLLSWMGLMSRSMKSIAKLLNKADDVHYFETIEQ
GVLYNLDQIHWSEEEKSYCDITVDDDDNDVFECHKGYVTLLPFALKLIPEDSERLIHVLQDL
RDPDICWSQFGIRSLSKSHPLFHSGEDYWRGNIWLNINYLILDALKHYAESSGALPDVKEMA
KPIYKELREILVTNVYNEFKRTGYAWEQYNEATGKGQRTRHFLGWTSLVIPIMKMPEELA

SEQ0165

P. pastoris homolog of *S. cerevisiae* **ALG12 (YNR030W)**

Alpha-1,6-mannosyltransferase localized to the ER; responsible for the addition of the alpha-1,6 mannose to dolichol-linked Man7GlcNAc2, acts in the dolichol pathway for N-glycosylation

Chr 4, 0544

S. cerevisiae null mutant: viable; increased lifespan; decreased resistance to brefeldin A

5' region

tgtgctcaaatttttgcaagctggtgaaatcgaagaggggcaaatcgatgcaaattgtgg
aaattctagaatcatgaaaacttggcttgatcgattgtatgctttctatgtctctgtttt
catatgcgtgaaggtctaagatcttggataagtcaatgtccccaaattgcgtttcatata
tcggagcaatgctgttgatcgttttaatccgtttcttgatgcgagcctcgtcttgttcgt
tgattctatcacgtttattcaggagaatgacatcggcgagtgcaatctgtaaatgagcgg
tcgtcgtcccgtctgtttctgttgatatatcctctccgtgccaatgtcctccaacatcgt
caaggcaggattcaacgtgctcactatccaaaacacacacaaccccgtcaatatacacat
tactcgctaatccgtcatctacccaaaacatattaacaataggacctggatcagcaacac

ctgttgtttcaagcagaatgtagtcgaaaccggccttcttggctatcaatctctctattg
ctgcaactccattatccttgacagagcagcataaacatccatttcccagatccaaccatt
cttcgtaactttctcccttatcttgaatggttagactcttttcaatttctacagaatctc
caaattcgttcagtatgactgctagttttctatctccttttttctgctatacttttcagta
gggtggatttacctgatcctaagtagccagtgattatcgatatgggcaccttaactgatt
cgtctacagcatttcttataacggagacaggattcgaagggtcaacaatcaactctccct
tgatttcagaaacatcttctaccaactgtgggatatcctcatccattcttgctgacgaaa
aagattcctgtgcgatgctccatgatatcctggtatcgcgaacagaagtatgcattttcc
agctcttccactgttagtgtcaagacgggcctatctaagta

ORF
ATGATCAATCTCAAACTATTGGACTTAGCATTGCTATTCCTGATCGGGCTACACCTATTTATATCTCCTTTTACC
AAAGTGGAGGAATCGTTCAACATCCAAGCCTGTCATGATATGATAGTTTACGGCTTTGATGATCTCTCTCATTAT
GATCATACTCAATTTCCAGGTGCAATTCAAAGATCTTTCTGGGGAGCCGGACTTTTAAGTATAGCTATCAGGCCT
TTCAAAGGCTTGCTGTCCAGCTTCTTTGCCAATTGGCCAACAAGACTATTCTACCAGTATTTGGTCAGGGGACTA
TTAGGCTTGTTCAACGGGCTAGGGTTGATCAGAATCAGGAGAGTTCTGAGTCGCAATATCTCAAAGGAAACAGCA
TTTTGGTATATGATACTCCAAATGTGTCAATTCCACATCGTTTATTACAGTAGTAGAACATTGCCCAACTTCATA
GCGTTACCGTTAGTTTCAAACGCCTTTGCGTTATACCTAACGCACGAGAATGTATCATTTGCTGTCCTGGCCTTT
TCAGGAGTGGTTCTTAGAGGAGAAATTGGTGTCTTTGCCGTAATTCTGGCGATTGTTAGTGTCCTTCAAACCCAC
AATTTCGATAATTTCCTCCATATTTTTAAGAATGGTGTCGTTGGCACCTTGGCAGGTGCCATTTGCAGCTACGTA
TTCGACTCCTATTTCTGGGGATACAAGGTAATCCCTGAGATCAATAGTTTCGTCTTCAATGTTTTAGAAGGGCAA
TCAAGTATTTGGGGAGTCGAACCATGGTATGCTTATCTAATGAAGTATCTGCCAAACTTGTTTAACAAATCTCCT
TTACTAATCTTAGTGATACCAGGACTTTTTCTAAAGAATGATAAACTGAAGAACTCAAAATCTCTCACACTTTCA
TCGTTACTGTACTTGGCAATCATTGCATTTCAGCCTCATAAAGAATGGAGATTTATAGTTTACATTGTTCCACCA
CTTGTCATCACCATATCTACAGTACTTGCACAACTACCCAGGAGATTCACAATCGTCAAAGTTGCTGTTTTTCTC
CTAAGTTTCGGCTCTTTGCTCATATCCCTGTCGTTTCTTTTCATCTCATCGTATAACTACCCTGGGGGTGAAGCT
TTACAGCATTTGAACGAGAAACTCCTTCTACTGGACCAAAGTTCCCTACCTGTTGATATTAAGGTTCATATGGAT
GTCCCTGCATGCATGACTGGGGTGACTTTATTTGGTTACTTGGATAACTCAAAATTGAACAATTTAAGAATTGTC
TATGATAAAACAGAAGACGAGTCGCTGGACACAATCTGGGATTCTTTCAATTATGTCATCTCCGAAATTGACTTG
GATTCTTCGACTGCTCCCAAATGGGAGGGGGATTGGCTGAAGATTGATGTTGTCCAAGGCTACAACGGCATCAAT
AAACAATCTATCAAAAATACAATTTTCAATTATGGAATACTTAAACGGATGATAAGAGACGCAACCAAACTTGAT
GTTGGATTTATTCGTACGGTCTTTCGATCCTTCATAAAATTTGATGATAAATTATTCATTTATGAGAGGAGCAGT
CAAACCTGA

Downstream
aaatatatacctcatttgttcaatttggtgtaaagagtgtggcggatagacttcttgtaa
atcaggaaagctacaattccaattgctgcaaaaaataccaatgcccataaaccagtatga
gcggtgccttcgacggattgcttactttccgacccttttgtcgtttgattcttctgccttt
ggtgagtcagtttgtttcgactttatatctgactcatcaacttcctttacggttgcgttt
ttaatcataattttagccgttggcttattatcccttgagttggtaggagttttgatgatg
ctactaaacacttcaaacccatcaaccaattggccaaacacaacgtgatgtccatctaac
caagaagtagctttggcagtaataaagaattgagatccattggtatcctttccagagttt
gccatgcttactcttcctaatctgtcatgcttcaaaacaaaattctcatcaggaaaatct
cctttgttctcaccatatatcgaacgaccacctttcccgtctccagatgtaaaatctcct
ccttgaatcataaaatcctcaatgactctatggaataaggatccatcgtatccaaacccg
ggtgtcatagcagcgagctggtagaagttcttcacggtcttaggaacgacagtaccgtac
agtcccagcaccagatcacccaaaggctgatcgtcctgggtgatgtgaaacgtcacgaaa
tgcgtataaggaggttcttctatcggtgcactccaaacaccaagcaaaaggctgatcagg
gtgaagatcaaagttaaattcatttctcaagtgttcggtgggggaagcagtcaatacgtt
gtgctcttcgaaagatgtaactttcgcgcgatgtatttttcaatcgaaaaattatctga
agtgcaaagcttacgagcctaacgtctccctaaacgcttcccaacaatcactagtcgatt
ctttggtggaatgaccaactttgtcaagcaaacatcgtttg

AA
MINLKLLDLALLFLIGLHLFISPFTKVEESFNIQACHDMIVYGFDDLSHYDHTQFPGAIQRSFWGAGLLSIAIRP
FKGLLSSFFANWPTRLFYQYLVRGLLGLFNGLGLIRIRRVLSRNISKETAFWYMILQMCQFHIVYYSSRTLPNFI
ALPLVSNAFALYLTHENVSFAVLAFSGVVLRGEIGVFAVILAIVSVLQTHNFDNFLHIFKNGVVGTLAGAICSYV
FDSYFWGYKVIPEINSFVFNVLEGQSSIWGVEPWYAYLMKYLPNLFNKSPLLILVIPGLFLKNDKLKNSKSLTLS
SLLYLAIIAFQPHKEWRFIVYIVPPLVITISTVLAQLPRRFTIVKVAVFLLSFGSLLISLSFLFISSYNYPGGEA
LQHLNEKLLLLDQSSLPVDIKVHMDVPACMTGVTLFGYLDNSKLNNLRIVYDKTEDESLDTIWDSFNYVISEIDL

DSSTAPKWEGDWLKIDVVQGYNGINKQSIKNTIFNYGILKRMIRDATKLDVGFIRTVFRSFIKFDDKLFIYERSS
QT

6.5 Remaining genes of the glycosylation pathway

MNN4 (YKL201C): Putative positive regulator of mannosylphosphate transferase (Mnn6p), involved in mannosylphosphorylation of N-linked oligosaccharides; expression increases in late-logarithmic and stationary growth phases
S. cerevisiae null mutant: viable
Homology not so high

SEQ0166 PATENTED elimination of mannosylphosphorylation
chr1-4_0409
5' region
cgcatagaacaaaaggcaaaggattaaacttcattgatgcaagattcataaatgttgaaa
caggcctttatatcgatatcactggattaagtaccagtcagtcagctcgaccgccaaggt
ttagtaacgcttcgaagaaagatcctatttacaattgcaggaataatcatttctactctc
ataacaatatagcacctctcaaatacacgttgatggaggggttcccagtttcattcctc
aacagtatgaagaaatattgagagaggagtatacaactggtttgacttcgaaacactaca
acggcaacttttttatgactcaattgaatttgtggcttgaaagagatccaatgctagcac
ttgtgccttcatccaaatacgaaattgaaggtggaggggtggaccataacaagattatca
agtctattcttgaactttccaacatcaaaaaattggaattgttggatgataatcccgata
tattagaggaggtgatcaggacatacgaactgacttccattcaccataaagagatgcagt
atctttccagtgtcaaaccagatggggacaggtccatgcagtcaaatgacataaccagtt
cttaccaggagtttctagcaagtctgaagaaattccagcctttacgcaaagatttgttcc
aatttgagcggatagacctttctaagcatagaaaacagtgagcagccgttttgcctaaaa
tgttccagaaactataggataaatatatacagtaatgaattaggtgatgttagcatttag
tccccaaaaatacctcgaatctccagctccatagcgcaaaatctccaaatctacttcaag
acgcactcatatgaaattcctgatgatctcgcgccaaaaggggaaaaatacattgtgtcg
tgtgatcagactttgttttcctttggactctgtcagattttggttatcggcacacaaaca
gtagttttttttgcctcaagctaaacttacagactcaagg
ORF
ATGAAAGTATCAAAGCGGTTGATACCGAGGAGATCTCGTCTCCTCATTATGATGATGCTACTGGTTGTTTACCAG
CTGGTGGTTTTGGTCCTAGGATTGGAGAGCGTCTCTGAAGGAAAATTAGCAAGCTTGCTTGACTTGGGCGATTGG
GATCTAGCTAACTCCTCGCTATCTATATCCGATTTCATAAAGCTGAAGCTCAAAGGCCAAAAGACTTATCACAAA
TTTGATGAACATGTCTTCGCCGCAATGGCAAGAATTCAAAGTAATGAGAATGGCAAGTTGGCGGATTACGAGTCT
ACTTCATCGAAGACTGACGTAACCATTCAAAATGTTGAACTTTGGAAGAGATTGAGCGAAGAAGAATACACTTAC
GAACCGCGGATAACTTTGGCTGTGTATCTGAGCTACATTCATCAGAGGACTTATGACAGGTACGCGACTAGTTAC
GCTCCTTATAACTTGCGGGTGCCTTTTTCGTGGGCTGACTGGATAGATCTGACGGCCCTAAATCAATACTTGGAT
AAAACGAAAGGCTGCGAGGCAGTTTTCCCTAGAGAAAGTGAGGCAACTATGAAGCTTAACAATATCACTGTTGTG
GACTGGCTTGAGGGCCTTTGCATAACTGATAAATCACTTCAAAATTCCGTAAACTCCACATATGCGGAAGAGATT
AATAGTCGGGACATCTTGTCTCCTAACTTCCATGTGTTTGGTTATTCTGATGCTAAAGATAATCCTCAGCAAAAA
ATCTTTCAATCTAAATCTTATATCAACTCAAAGCTGCCGCTCCCAAAAAGTTTGATATTTTTAACAGATGGAGGT
AGTTACGCTTTGACAGTCGACCGAACTCAAAATAAAAGAATTCTAAAATCTGGCCTGCTTTCACACTTTTTCTCA
AAGAAAAAGAAGGAACACAATCTGCCTCAAGACCAAAAAACTTTCACGTTTGACCCCGTATACGAATTCAATAGA
CTGAAATCTCAGGTCAAGCCCCGTCCAATATCTTCAGAACCTAGTATTGATTCTGCTTTGAAGGAAAATGACTAC
AAGCTTAAACTGAAAGAGTCGTCGTTTATTTTTAATTACGGAAGGATTCTTTCGAACTATGAAGAGCGGCTTGAG
AGTCTAAATGACTTCGAGAAATCGCACTACGAGTCCTTAGCTTATTCCTCCTTGTTAGAGGCAAGAAAGTTGCCC
AAGTATTTTGGCGAAGTTATATTGAAGAACCCACAAGATGGTGGAATTCATTATGATTACAGATTCTTCAGCGGA
CTCATTGATAAAACTCAGATAAATCATTTTGAGGATGAGACTGAAAGAAAGAAAATAATCATGCATAGACTTCTT
CGAACTTGGCAGTACTTCACGTATCACAATAACATTATCAATTGGATCTCGCACGGTTCTTTACTGTCATGGTAT
TGGGATGGACTTTCTTTTCCATGGGACAATGACATTGACGTACAAATGCCCATAATGGAGCTGAATAACTTCTGC
AAACAGTTCAACAATTCTCTGGTCGTGGAGGATGTTTCTCAAGGGTTTGGCAGATACTACGTTGATTGCACGAGC
TTCCTGGCCCAGAGAACGCGAGGTAATGGTAACAACAACATTGATGCCCGTTTTATTGATGTGTCGTCTGGTCTC
TTCATTGACATTACGGGTTTGGCTTTGACTGGATCAACAATGCCCAAAAGATACTCCAATAAGCTGATAAAACAA
CCGAAAAAATCTACCGACTCAACAGGATCGACTCCTGAGAACGGACTCACTAGAAACTTGAGGCAAAATTTGAAT
GCACAAGTTTACAACTGTAGAAACGGTCATTTTTACCAATACTCGGAGCTATCTCCTTTGAAGTTGTCGATAGTA

GAAGGTGCACTCACCCTAATACCCAACGATTTTGTTACTATATTGGAAACTGAGTACCAAAGGAGAGGTCTTGAA
AAGAACACATATGCGAAGTATCTCTACGTTCCAGAGCTTCGACTTTGGATGTCATACAATGACATCTATGATATC
TTGCAAGGTACTAATAGTCATGGCCGTCCTTTATCTGCAAAGACAATGGCGACTATCTTTCCTCGGTTAAACTCT
GACATTAATCTAAAAAAGTTTTTGCGCAATGATCATACTTTTAAGAACATTTATTCTACTTTCAACGTGACACGA
GTGCACGAGGAGGAACTGAAGCATTTGATAGTAAACTATGACCAAAATAAACGGAAGTCGGCTGAGTACAGGCAG
TTCTTGGAAAACTTGCGGTTTATGAATCCAATCAGAAAAGATCTGGTGACTTACGAGAGTAGGTTGAAGGCTCTT
GATGGATACAATGAGGTCGAAGAATTAGAAAAGAAGCAAGAGAATAGGGAAAAAGAAAGAAAGGAGAAGAAGGAA
AAGGAGGAAAAAGAGAAGAAGGAAAAGGAGGAAAAAGAGAAGAAGGAAAAGGAAGAAAAGGAAAAGAAGGAAAAG
GAAGAAAAGGAGAGGAAAGAGAAGGAAGAAAAGGAAGAATATGAAGAAGACGATAATGAGGGCGAACAACCAACA
GAACAAAAGAGCCAGCAGGAGGCTAAAGAATAG

Downstream

cggaggaatgcaaataataatctccttaattacccactgataagctcaagagacgcggtt
tgaaaacgatataatgaatcatttggattttataataaaccctgacagtttttccactgt
attgttttaacactcattggaagctgtattgattctaagaagctagaaatcaatacggcc
atacaaaagatgacattgaataagcaccggctttttttgattagcatataccttaaagcat
gcattcatggctacatagttgttaaagggcttcttccattatcagtataatgaattacat
aatcatgcacttatatttgcccatctctgttctctcactcttgcctgggtatattctatg
aaattgcgtatagcgtgtctccagttgaaccccaagcttggcgagtttgaagagaatgct
aaccttgcgtattccttgcttcaggaaacattcaaggagaaacaggtcaagaagccaaac
attttgatccttcccgagttagcattgactggctacaattttcaaagccagcagcggata
gagccttttttggaggaaacaaccaagggagctagtacccaatgggctcaaaaagtatcc
aagacgtgggattgctttactttaataggatacccagaaaaaagtttagagagccctccc
cgtatttacaacagtgcggtacttgtatcgcctcagggaaaagtaatgaacaactacaga
aagtccttcttgtatgaagctgatgaacattggggatgttcggaatcttctgatgggttt
caaacagtagatttattaattgaaggaaagactgtaaagacatcatttggaatttgcatg
gatttgaatccttataaatttgaagctccattcacagacttcgagttcagtggccattgc
ttgaaaaccggtacaagactcattttgtgcccaatggcctggttgtcccctctatcgcct
tccattaaaaaggatcttagtgatatagagaaaagcagac

AA

MKVSKRLIPRRSRLLIMMMLLVVYQLVVLVLGLESVSEGKLASLLDLGDWDLANSSLSISDFIKLKLKGQKTYHK
FDEHVFAAMARIQSNENGKLADYESTSSKTDVTIQNVELWKRLSEEEYTYEPRITLAVYLSYIHQRTYDRYATSY
APYNLRVPFSWADWIDLTALNQYLDKTKGCEAVFPRESEATMKLNNITVVDWLEGLCITDKSLQNSVNSTYAEEI
NSRDILSPNFHVFGYSDAKDNPQQKIFQSKSYINSKLPLPKSLIFLTDGGSYALTVDRTQNKRILKSGLLSHFFS
KKKKEHNLPQDQKTFTFDPVYEFNRLKSQVKPRPISSEPSIDSALKENDYKLKLKESSFIFNYGRILSNYEERLE
SLNDFEKSHYESLAYSSLLEARKLPKYFGEVILKNPQDGGIHYDYRFFSGLIDKTQINHFEDETERKKIIMHRLL
RTWQYFTYHNNIINWISHGSLLSWYWDGLSFPWDNDIDVQMPIMELNNFCKQFNNSLVVEDVSQGFGRYYVDCTS
FLAQRTRGNGNNNIDARFIDVSSGLFIDITGLALTGSTMPKRYSNKLIKQPKKSTDSTGSTPENGLTRNLRQNLN
AQVYNCRNGHFYQYSELSPLKLSIVEGALTLIPNDFVTILETEYQRRGLEKNTYAKYLYVPELRLWMSYNDIYDI
LQGTNSHGRPLSAKTMATIFPRLNSDINLKKFLRNDHTFKNIYSTFNVTRVHEEELKHLIVNYDQNKRKSAEYRQ
FLENLRFMNPIRKDLVTYESRLKALDGYNEVEELEKKQENREKERKEKKEKEEKEKKEKEEKEKKEKEEKEKKEK
EEKERKEKEEKEEYEEDDNEGEQPTEQKSQQEAKE

SEQ0167

chr1-4_0410 PATENTED

5' region

gggagagttagctagcatacaagataatgaaggatcaatagcggtagttaaagtgcacaa
gaaaagagcacctgttgaggctgatgataaagctccaattacattgccacagagaaacac
agtaacagaaataggaggggatgcaccacgagaagagcattcagtgaacaactttgccaa
attcataaccccaagcgctaataagccaatgtcaaagtcggctactaacattaatagtac
aacaactatcgattttcaaccagatgtttgcaaggactacaaacagacaggttactgcgg
atatggtgacacttgtaagtttttgcacctgagggatgatttcaaacagggatggaaatt
agatagggagtgggaaaatgtccaaaagaagaagcataatactctcaaaggggttaagga
gatccaaatgtttaatgaagatgagctcaaagatatcccgtttaaatgcattatatgcaa
aggagattacaaatcacccgtgaaaacttcttgcaatcattatttttgcgaacaatgttt
cctgcaacggtcaagaagaaaaccaaattgtattatatgtggcagagacactttaggagt
tgctttaccagcaaagaagttgtcccaatttctggctaagatacataataatgaaagtaa
taaagtttagtaattgcattgcgttgactattgattgcattgatgtcgtgtgatactttc
accgaaaaaaaacacgaagcgcaataggagcggttgcatattagtccccaaagctattta
attgtgcctgaaactgtttttttaagctcatcaagcataattgtatgcattgcgacgtaac
caacgtttaggcgcagtttaatcatagcccactgctaagccagaattctaatatgtaact

acgtacctttccttttaataaatgatctgtattttccacctagtagcagatcaaattgtt
caactttaagtctttggtccctcaagcgagagaacttgcg

ORF

ATGACACTCAGGAGTGCCATAAAAGCCAGAACCTCAAAAGGACTGATCGGAGCTGTTATTATAGCCTCAATAATA
TTTTTCACCACAGTAACCTTCTACGATGAAAGCAAAATTGTCGGCATAATAAGAGTTTCTGATACTTATACAGGC
CATAGCGCTGTATCTTCAACTTTCAATGCTTCTTCCGTTGTTAGTGACAACAAGATCAACGGATATGGACTTCCT
TTGATTGACACGGAATCAAATAGCCGTTATGAGGATCCAGACGATATTTCCATTGAAAACGAATTGCGCTATAGA
ATTGCCCAATCTACCAAAGAGGAAGAAAACATGTGGAAACTCGATACCACTCTCACGGAAGCAAGCTTGAAAATC
CCCAACATACAGTCGTTTGAGCTGCAGCCGTTCAAAGAAAGACTTGATAATTCACTTTACAATTCTAAGAACATA
GGAAACTTTTACTTCTATGACCCAAGGCTTACATTCTCAGTTTACTTGAAGTATATCAAGGATAAATTGGCCTCT
GGAAGCACAACAAATCTTACAATACCCTTCAACTGGGCACATTTTAGAGATTTATCGTCACTGAATCCTTATTTG
GACATAAAACAAGAAGATAAGGTCGCATGTGATTACTTTTATGAATCAAGTAATAAAGACAAACGAAAACCCACG
GGTAACTGTATTGAGTTTAAAGATGTTCGTGATGAGCACCTGATACAGTATGGGATTTCATCAAAAGACCATCTA
CCTGGTCCTTTTATTTTAAAGTCACTTGGAATTCCCATGCAGCATACAGCCAAGCGACTGGAATCAAATCTTTAT
CTATTAACCGGTGCGCCAGTTCCACTTTCATTAAGTTTCATGACTAAAAAGGGATTATACCAAGTTGGAGTTGAC
CAAACAGGAAAACTTGATCCAAACATTGCTCGTACTGAACTATGGGAGTTTTACAAAAATGGGAAAGAAAACCTT
CAATTTAATGCACAAGAGGAGCTATCTCACCTGATAGAGACAGTCCCTTCATCTAGTAACTCATCCAGTGGAGAA
GGCTATTTCACTACTGAATTAAAGGAGAACAACTTTGAGTTGCCCCTGAGTAAGAATGATTTTACCTTTGATGAT
TCCGAGGTCGAGTCATTGATTAAAGGTTTATCTGAACAGGACTTGGACCTTCATACCCAGAGATACAAAGAATCT
TTGCAGTACTCTTTTGCGACTCGAGAGAATGACGTGAAGAAATACTTTTATGAGGCCAGAATGATTATCAATACT
GTTAACAAAGAAGGTGGAGCGCATTATGACTGGAGGTTTTTCAATGGAGCCATGAATCATGAAAGTTCCGGTTTT
ACTGAGGAAGAAAGACAACTGAGAAAGAGATCTGTTTTGCATCGTTTATTGCGAAACTGGCTTGTATTCAATTAC
CAGCAAGGATCTCCCACTTGGTTGGCTCATGGAACTTTACTTTCTTGGTATTGGAATTCATTGATGTTCCCTTGG
GATTATGATATTGATGTGCAAATGCCAATCAAGAGTTTGAACAATCTATGTGCTAACTTCAACCAATCATTAATA
ATTGAGGATCTTACTGAAGGATATTCTTCTTTTTTCTTGGATTGCGGATCAAGTATCACGCATAGAACAAAAGGC
AAAGGATTAAACTTCATTGATGCAAGATTCATAAATGTTGAAACAGGCCTTTATATCGATATCACTGGATTAAGT
ACCAGTCAGTCAGCTCGACCGCCAAGGTTTAGTAACGCTTCGAAGAAAGATCCTATTTACAATTGCAGGAATAAT
CATTTCTACTCTCATAACAATATAGCACCTCTCAAATACACGTTGATGGAGGGGGTTCCCAGTTTCATTCCTCAA
CAGTATGAAGAAATATTGAGAGAGGAGTATACAACTGGTTTGACTTCGAAACACTACAACGGCAACTTTTTTATG
ACTCAATTGAATTTGTGGCTTGAAAGAGATCCAATGCTAGCACTTGTGCCTTCATCCAAATACGAAATTGAAGGT
GGAGGGGTGGACCATAACAAGATTATCAAGTCTATTCTTGAACTTTCCAACATCAAAAAATTGGAATTGTTGGAT
GATAATCCCGATATATTAGAGGAGGTGATCAGGACATACGAACTGACTTCCATTCACCATAAAGAGATGCAGTAT
CTTTCCAGTGTCAAACCAGATGGGGACAGGTCCATGCAGTCAAATGACATAACCAGTTCTTACCAGGAGTTTCTA
GCAAGTCTGAAGAAATTCCAGCCTTTACGCAAAGATTTGTTCCAATTTGAGCGGATAGACCTTTCTAAGCATAGA
AAACAGTGA

Downstream

gcagccgttttgcctaaaatgttccagaaactataggataaatatatacagtaatgaatt
aggtgatgttagcatttagtccccaaaaatacctcgaatctccagctccatagcgcaaaa
tctccaaatctacttcaagacgcactcatatgaaattcctgatgatctcgcgccaaaagg
ggaaaaatacattgtgtcgtgtgatcagactttgttttcctttggactctgtcagatttt
ggttatcggcacacaaacagtagttttttttgcctcaagctaaacttacagactcaagga
tgaaagtatcaaagcggttgataccgaggagatctcgtctcctcattatgatgatgctac
tggttgtttaccagctggtggttttggtcctaggattggagagcgtctctgaaggaaaat
tagcaagcttgcttgacttgggcgattgggatctagctaactcctcgctatctatatccg
atttcataaagctgaagctcaaaggccaaaagacttatcacaaatttgatgaacatgtct
tcgccgcaatggcaagaattcaaagtaatgagaatggcaagttggcggattacgagtcta
cttcatcgaagactgacgtaaccattcaaaatgttgaactttggaagagattgagcgaag
aagaatacacttacgaaccgcggataactttggctgtgtatctgagctacattcatcaga
ggacttatgacaggtacgcgactagttacgctccttataacttgcgggtgccttttcgt
gggctgactggatagatctgacggccctaaatcaatacttggataaaacgaaaggctgcg
aggcagttttccctagagaaagtgaggcaactatgaagcttaacaatatcactgttgtgg
actggcttgagggcctttgcataactgataaatcacttcaaaattccgtaaactccacat
atgcggaagagattaatagtcgggacatcttgtctcctaa

AA

MTLRSAIKARTSKGLIGAVIIASIIFFTTVTFYDESKIVGIIRVSDTYTGHSAVSSTFNASSVVSDNKINGYGLP
LIDTESNSRYEDPDDISIENELRYRIAQSTKEEENMWKLDTTLTEASLKIPNIQSFELQPFKERLDNSLYNSKNI
GNFYFYDPRLTFSVYLKYIKDKLASGSTTNLTIPFNWAHFRDLSSLNPYLDIKQEDKVACDYFYESSNKDKRKPT
GNCIEFKDVRDEHLIQYGISSKDHLPGPFILKSLGIPMQHTAKRLESNLYLLTGAPVPLSLSFMTKKGLYQVGVD
QTGKLDPNIARTELWEFYKNGKENLQFNAQEELSHLIETVPSSSNSSSGEGYFTTELKENNFELPLSKNDFTFDD
SEVESLIKGLSEQDLDLHTQRYKESLQYSFATRENDVKKYFYEARMIINTVNKEGGAHYDWRFFNGAMNHESSGF
TEEERQLRKRSVLHRLLRNWLVFNYQQGSPTWLAHGTLLSWYWNSLMFPWDYDIDVQMPIKSLNNLCANFNQSLI

IEDLTEGYSSFFLDCGSSITHRTKGKGLNFIDARFINVETGLYIDITGLSTSQSARPPRFSNASKKDPIYNCRNN
HFYSHNNIAPLKYTLMEGVPSFIPQQYEEILREEYTTGLTSKHYNGNFFMTQLNLWLERDPMLALVPSSKYEIEG
GGVDHNKIIKSILELSNIKKLELLDDNPDILEEVIRTYELTSIHHKEMQYLSSVKPDGDRSMQSNDITSSYQEFL
ASLKKFQPLRKDLFQFERIDLSKHRKQ

SEQ0168

chr2-1_0718 PATENTED

5' region

tgcggtgttggtgatgctaattacgtaccagaacccaatctgtggactcaggaccagctc
agcttgacaaaccaagacttgcactccaatgtgcacaacccagtgattgagcagatcgaa
acctcatcaggagtcagattgtagtatggaaaactttgtattctctatgtacttaaacac
tggtttatttttttattgatcgttatattgaacagtttacactggaacatcttcagggtc
gatgtccttaatccagtgttgaccaaagattgggatcttctcgaagaaagtcttttggaa
caaaggccagttttcagtgaaagtgaagacggcaaacagaccggcacctccccagaaagc
caaaattggaaaagtagttttaatttgggttggagtcaatccagcaattttcttgacggt
ggtatacttgggacctttaacgtactgaaatgttagaacatgtttgtaaaaatcaaatca
tcactgcagaaacggtttgtgtgcctgcaccggagggttatcataatgccacttacgttg
accattttggaggtgtttgactaagttcaaatatgaatctctaagaaaactaataatcaa
tatggtgcgagcattgattggttggacagctagtttggagaagtacacgacttagatgaa
tctgcaataaggaatagtccaatctgattatgtaagctctccttttttggttttcatttcc
atcagctcaagcttatcatagctcaggtcccctccagcttatgatggaataggccattat
ttttgccctaaaaagtggaagtccacaagaagaaatacaaatactcaaaattcaaaagt
cttcctttgagtggatgcaattttacgtagtttactgtatgacgtaactaatgaaccctt
ccgacacaaagattgaggtgcctcacttaacgtcattcttctatacccacgagtgcaact
gactaggtcttattttgttaattgcctcagtttctccgaa

ORF

ATGAGTGGCAATCCTTTTCTGTTCTCTCCTTCAAATTTTGACTTTTCTGGTTTGGATCATTATAGATCCACTGAT
AAAGATCACTTAGCTCTAGATGTTCTCGATTATGACAAAAATCACTTCTTCTCCAGAAACTCCCCCAGTTTGAAA
TCTCGTATTCACTTTTATCGACATAAATTGACCACTAGAAAGCAAATTGGACTTTTCAGCGGCAGACTGAAGCTT
TTTGTGCTTGCTCTCTTTGTGTTGATCACATTTTCTGCAATCCACATTCCAATCCCTTTCTCTTTGGATATTCTA
GGTTCCCATGTCAAATACCTGCCCTTACGAGAGAAAGTCGATCCGGAAGAGGCACTCCATCTGCACGGACTGGAT
CTCTCGGTAGCAGAGCTACCTTTTTTCAATGATGACATGATGTCTGAATTTAACTACGATCCTAGACTACCCACC
GCTTTGATTTTGAAGTTAGTGTTAGATCATATAAGTGTGCGTAATGGAACGTTTGATGCTAAGTTTAAGGTCCCC
TTTAACTGGAAACTTTGGGTGGATTTGCATTCAAGGTTAGTTCCATCTAATAGTTGGTATAATCGATTTCGATTA
CCCTCAGGTCGTTTCGAAACATGCGATGAATTTAAGAGGTTTTTCGGAATCACTAAGAATCACTTTGGAACAGAC
CTTGATAATTGCGTTGATATCGAGTATGATACTCCGGAAGGTTATCCAAAGTTCAAAGTTTTGCATGCGGAAGAT
AAAGCTCTTCCTTATGAAGCACGTATCATTTATGGTGCTTCTTACCTTTACCACGAAGCACAGAATCCTAAAAGG
TTGATATTTTTAGGATTGGGCAAGTCCAATGAGTCTTTGATCTTACCAGTTGAGGCAAATGACAGTTCCAACTTA
ATGCAATTCAACCACGAATATGCAAGAAGCTTTAACGATCAACCTTTCGTTTCTCTTGAGGAACTTGTCAAGAAG
GTTTCACTGACCTTGAATTTGAATAGTGATAAGGTGCTACCAATCAATGAACTGGACGTTATCAAAGACACCCCG
CGCTTAATGAATCACAACAACCAGGGACTGAGCATAGACAAGAGCTCATTTCAATGGGATCTGGAAAGGGAATTA
CAGTTGTTAGAACATAGAACCAGTCAAGTTAATGACGTGGAAGGCCTTGATGCGGGTATTTATTCAACAATTCAA
TGTGAAATGCGCTCTATGTACGATTTTTCAAAATACTTCCATGAATCAAAAGTCTCTGGTAAATATCTTCCTTCT
GGAGAGCACTATGACTGGCGATTTTTTAATGGTTTTTACCTTTCTCAGCAGGAGAATCTAGCTGTCCTGCACAGG
TTAGGAAGAGCATGGCTACGCTTTTCTCGTGCTGCTGGTTTACATACATGGATTGCTCACGGGACACTGTTGGGT
TGGTATTGGAATGGTCTGATTCTGCCGTGGGATCAGGATCTTGATGTTCAAATGACTGTACAATCATTGTATCTG
TTGGGAAGGAATTTCAACAGCTCTCTTGTAACTGATGTTAGTATTGAAGATGGCTACAGCTCAGCATTGGGACAT
TACTATATTGACGTTGGATCCTCCTTCTTTGTTAGGGATAAACTAAATGGTAACAATGCTATAGATGCACGTTTC
GTTGATACTGAGACCGGGTTGTATGTTGATATAACTGCATTGGCTTTTACAGATCACTTAAAACTAAAACTCACT
ACCAAAGAGAAAGTTGAGCTACAGAAGGTTATGGATCCAAATGTAAAGGAAAAATTGCAGTGGATCAAAAATAAA
TATTCAACGGCCACGCTACCGGGTGTGATAGAAACAGATAGGAATAAAGTATCTGATGCGCTAGAGAAGCAATTT
CATGATTTCAAGTTCGACAATTTTGTCAACAAAGAGTTGTTTCACTGTCGAAATAACCATTTCTACAAATATGGA
GAGGTTGGCCGATTACGGAGCACTATGTTTGAGGGCGTTCCTGCCCTTATACCATTTGAATTTGAGTCCATACTG
AAACGAGAATATCCTAAAGGTCTAACTTTGAAGCATTTCTCCAATCATTTTTGGGATCCAGTGAACCGATTGTGG
GTACCAGAAAAGAAGAAAAAAATTAGACACATAGAGTTTTCACTTACGAAGGAAGTTACAGAAAGCCACAAGAAA
GAACTTGCACAGATCCATGGGAACGAAACGGGTATAACCTCCGACTTCGCATATTCTCCTTTCAGAATAGATCCC
TGGCTGTCTCGATACAGGAAAAAAATGACTAGGAGCCAATAA

Downstream

ttatgcatatactgaaacaacagaaggaactacagtaaattcataaaaagcttaattctt
actttcatctcggcactgtaaattaactcaagttggggcaacattgtgtgtatactctta

ctggcatcttttcatctgaagtcatcttctactactcttctcttctgtatgacgtaatca
gctcggcagctgtggcatcgaacaaaaaaatgaacagccatccgtcatatctcatgactg
actgagcaagaactaagtcaacaggaaacctaaaataagctttccatttcttttgcgctg
aagccaaccactccccacacagttgatgagtggacgcaaaaccagctcctataccttgac
agaagagtcgccggaatcaacctcaacaattcaggatatacgagaggaagaccaagttgc
tccagggccccagcaagaagcacctaaacagtcacatatccaaaaatggttaagcgatca
tcctaaagtatacgcagttttatcttggatatggaaattttggttgaaacagtggtttct
catatgtttgggccctgcggttgctctagctcatgcatacccaaattttgccagacatga
tggaaccattaggtcagagtatactatcaactacggagccgtagctatcatattcttcat
ctctggtcttactatgaaaaccaaagacttttgaagaactttggacactggagagccca
tttcacagtgttaagctgctcgtttctacttacttcttctatcatttacggtatagcgtg
cggtataagagctgctcatgattccaatatcgatgactggatgttagcaggacttattgt
taccgcatgttgtccaaccactgtgagtagtaacgttgtcatgacagaacaagctcatgg
taatgttttcttgagtatttgcgaagttattattggcaatgttttaggggggtttcatcac
acctgctctggtgcaaatgtatctgacaggtagctgggat

AA

MSGNPFLFSPSNFDFSGLDHYRSTDKDHLALDVLDYDKNHFFSRNSPSLKSRIHFYRHKLTTRKQIGLFSGRLKL
FVLALFVLITFSAIHIPIPFSLDILGSHVKYLPLREKVDPEEALHLHGLDLSVAELPFFNDDMMSEFNYDPRLPT
ALILKLVLDHISVRNGTFDAKFKVPFNWKLWVDLHSRLVPSNSWYNRFRLPSGRFETCDEFKRFFGITKNHFGTD
LDNCVDIEYDTPEGYPKFKVLHAEDKALPYEARIIYGASYLYHEAQNPKRLIFLGLGKSNESLILPVEANDSSNL
MQFNHEYARSFNDQPFVSLEELVKKVSLTLNLNSDKVLPINELDVIKDTPRLMNHNNQGLSIDKSSFQWDLEREL
QLLEHRTSQVNDVEGLDAGIYSTIQCEMRSMYDFSKYFHESKVSGKYLPSGEHYDWRFFNGFYLSQQENLAVLHR
LGRAWLRFSRAAGLHTWIAHGTLLGWYWNGLILPWDQDLDVQMTVQSLYLLGRNFNSSLVTDVSIEDGYSSALGH
YYIDVGSSFFVRDKLNGNNAIDARFVDTETGLYVDITALAFTDHLKLKLTTKEKVELQKVMDPNVKEKLQWIKNK
YSTATLPGVIETDRNKVSDALEKQFHDFKFDNFVNKELFHCRNNHFYKYGEVGRLRSTMFEGVPALIPFEFESIL
KREYPKGLTLKHFSNHFWDPVNRLWVPEKKKKIRHIEFSLTKEVTESHKKELAQIHGNETGITSDFAYSPFRIDP
WLSRYRKKMTRSQ

MNN6 (YPL053C): Probable mannosylphosphate transferase involved in the synthesis of core oligosaccharides in protein glycosylation pathway; member of the KRE2/MNT1 mannosyltransferase family
S. cerevisiae null mutant: viable
Chr3_1162
Chr3_0215
Chr2-2_0105
Chr1-3_0138
Chr1-4_0050
These have homology, but are assigned to other mannosyltransferase genes

7 Methanol metabolism

AOX: alcohol oxidase
AOX1: chr4_0152
AOX2: chr1-1_0226
genes known; promoters known

SEQ0169
CAT: catalase
chr2-2_0131
gene known; promoter unknown

5' region

ctatgttgacggagagtgttgggtctaacatcatggcacatggtagaggaggtgttaaga
catctagagagcgtgatagagaggaagagatcagagaatacgaggagaccaatttcacta
gattaccaacttcggttactgagaagtcaaagaaacaaaagaaagatcatagattgaaca
cctttgcaggagaggattggtcattctttggcaaggacagagatgaagacatgaagaaaa

gtgcgaggaagaataaaaatactgcttcctccgcctgggaaagagcaaaaagacgcagag
gaaac**taa**agtgtgtaatcatatatataataaatgaggaataataattgaatagagattt
aacgagtcgaagtttctgaaatatacgcacagtttatatttatgattttgatatctaact
acagtcttctccatatatttaactataaataataaagtatataactcttatgaaactgtt
tcaccacatttttttctacgtaatcgaactccgaatgcggttctcctgtaaccttaattg
tagcatagatcacttaaataaactcatggcctgacatctgtacacgttcttattggtctt
ttagcaatcttgaagtctttctattgttccggtcggcattacctaataaattcgaatcga
gattgctagtacctgatatcatatgaagtaatcatcacatgcaagttccatgataccctc
tactaatggaattgaacaaagtttaagcttctcgcacgagaccgaatccatactatgcac
ccctcaaagttgggattagtcaggaaagctgagcaattaacttccctcgattggcctgga
cttttcgcttagcctgccgcaatcggtaagtttcattatcccagcggggtgatagcctct
gttgctcatcaggccaaaatcata*tataa*gctgtagacccagcacttcaattacttgaaa
ttcaccataacacttgctctagtcaagacttacaattaaa

ORF

ATGTCTCAACCACCTAAATGGACAACATCAAATGGTGCTCCTGTCTCCGATGTATTTGCTACAGAGAGAGCCACC
TTCGACAACGCTAATCACGCTAACAATGCCCCCAAAGTTGGGCCTCTGCTGCTACAAGACTTTCAATTGATTGAT
TCTCTTGCCCATTTCGACAGAGAACGTATTCCAGAGAGAGTTGTGCATGCCAAAGGTGCGGGAGCTTTCGGTGAG
TTTGAGGTGACTGACGACATATCTGACGTTTGTGCTGCCAAATTCTTGGATACTATTGGGAAGAAAACCCGTATC
TTCACTCGATTCTCCACCGTTGGAGGTGAGAAAGGTTCTGCCGACAGTGCCCGTGACCCAAGAGGGTTTTCCACC
AAGTTTTACACAGAGGAAGGTAACTTGGACCTTGTGTACAACAACACTCCCATCTTCTTTATCAGAGATCCATCC
AAGTTCCCACATTTCATCCACACCCAAAAGAGGAATCCTGCTACTAACCTGAAGGACGCCAATATGTTCTGGGAT
TACCTTGTTAACAACCAAGAGTCCATTCATCAGGTGATGTACTTGTTCTCAGATCGTGGAACTCCTGCTTCTCTT
AGAAAAATGAACGGGTACTCTGGTCACACTTACAAATGGTACAACAAGAAAGGTGAGTGGGTTTACGTACAAGTT
CACTTCAAGTCGGACCTGGGTGTCGTAAACTTCAACAATGAAGAGGCTGGAAAGCTTGCTGGCGAGGATCCAGAC
TACCACACCGGTGATCTTTTCAACGCCATTGAGAGAGGCGAGTATCCTTCATGGACTTGTTACATTCAGACCATG
ACTCAAGAACAAGCTGCCAAGCAACCATTTTCCGTGTTTGACTTGACCAAGGTTTGGCCACACAAAGACTTCCCA
TTGAGGAGATTTGGTAAATTCACTTTGAACGAAAACCCAAAGAACTACTTTGCCGAAGTTGAGCAAGCCGCCTTC
TCTCCATCTCATACTATCCCTAGTATGCAGCCATCTGCTGATCCAGTCCTTCAATCTCGACTGTTCTCCTACCCA
GATACTCACCGTCATAGATTGGGTGTTAACTATCAGCAGATTCCTGTCAATTGTCCAGTGGCTCCTGTTTTTACC
CCTCAGATGCGTGATGGTTCCATGACTGTCAATGGCAACCTTGGGTCTACTCCCAACTACAAGTCTTCTTTCTGT
CCATTCTCTACAGAGGCACAGATCCAGACCAATTCACACACACCAGAAGAAGTTTTGGCTGCTCACACTGAGAAA
TTCCACTGGGGAGGTATCCTTGACAGTAAGAGCTACGATTTTGAACAGCCAAGAGCGCTTTGGAAAGTATTCGGC
AAAACTCCTGGTCAGCAACGTAACTTCTGCCACAACGTTGCCGTTCACGTTGCGGCTGCCAACCATGAAATTCAA
GACAGGGTTTTTGAATACTTCTCCAAGGTTTACCCTGAAATTGGTGACCAGATTCGTAAAGAAGTATTGCAGCTA
TCTCCAAGAGGTGACTCTGCAGCAAGATTGTAG

Downstream

gctaactatatttattattaattaaagattctttaacttcggtaatttgtagaacagaaa
gagggattgtcattctcggttcattatcctttccagcacatcattttgtgggggactcta
gccttcgcgctgcttctactactggtcaaataacaattctccataacggtattcgcttac
ctcatcggttacgagcgttgtccgatgtaggccaatcaacgatgatcgaaacatcatcag
aaaactaaactctcacatagagacaaaggtcacaatc**atg**ctaagacaacaagtgggtag
aaagcttgtatccaagcgtttcttcagtggcttgaaggatagtgaccgtatcttccaaaa
tgtgtacagtaaatatggtgaagacctgaagagttcacagcaacgtggtgactggtacaa
gaccaaggaaatcatcctcaagggacacgaatggcttatcaacgaattgaaggcttctgg
tctacgtggaagaggtggtgccggtttcccatctggccttaagtatagtttcatgcctcc
taacccaaacagagaaccacaatacctggtggtcaatgctgatgaaggtgaaccaggaac
atgtaaagatagagaaatcattagaaaggagccccacaaattggtagagggctgtctgct
ggccggaagagccatgaatgccactgcggcctatatctatattagaggggagttttacaa
cgaagctgttgctttgcagacagctattaacgaggcttacaaagctggactactgggaaa
gaacgcctgtggttccggatacgacttcgacgtttacattcaccgtggtatgggtgctta
tatttgtggtgaagaaactgcacttattgagtccattgagggcaaagctggtaagccaag
actgaagcctcctttccccgcaggtgttggtttgttcggtcgtccaacaacagtaacaaa
tgttgaaaccgtcgctgtcactccaacaatctttagaaga

AA

MSQPPKWTTSNGAPVSDVFATERATFDNANHANNAPKVGPLLLQDFQLIDSLAHFDRERI
PERVVHAKGAGAFGEFEVTDDISDVCAAKFLDTIGKKTRIFTRFSTVGGEKGSADSARDP
RGFSTKFYTEEGNLDLVYNNTPIFFIRDPSKFPHFIHTQKRNPATNLKDANMFWDYLVNN
QESIHQVMYLFSDRGTPASLRKMNGYSGHTYKWYNKKGEWVYVQVHFKSDLGVVNFNNEE
AGKLAGEDPDYHTGDLFNAIERGEYPSWTCYIQTMTQEQAAKQPFSVFDLTKVWPHKDFP
LRRFGKFTLNENPKNYFAEVEQAAFSPSHTIPSMQPSADPVLQSRLFSYPDTHRHRLGVN
YQQIPVNCPVAPVFTPQMRDGSMTVNGNLGSTPNYKSSFCPFSTEAQIQTNSHTPEEVLA

AHTEKFHWGGILDSKSYDFEQPRALWKVFGKTPGQQRNFCHNVAVHVAAANHEIQDRVFE
YFSKVYPEIGDQIRKEVLQLSPRGDSAARL

FLD: formaldehyde dehydrogenase
chr3_1028
Gene known, promoter known

SEQ0170
FGH: S-formylgluthatione hydrolase
chr3_0867
gene known; promoter unknown

5' region

tggttccctctcggtccaataccaaaaatattatcaccatacaggtctcccttcgatacc
agtgcaaagttgaaccgtgggattaccttggaatctacaaaaatagtgtcactcacaagt
ttgtcatcaaccacgctgccgcttgcaaaggagaactgaacatgaaggttgttagggttt
gttatattggaataagtggtggatttgttgaaggcgaacgcaccaaagctacatccgtcc
tgagcacactgtgaatttgtcacggaattgaccaagaggtcagacgatcctgtatcccat
tgagccgttatgctttgtgggggaaaccctatttctatcgtactaagaaaaccaatggtg
aactcatattcggtatcaatggcgacgattccagcatagcctgtagacagtaacaacact
agggcaacagcaactaa**cat**atcttcattgatgaaacgttgtgatcggtgtgacttttat
agtaaaagctacaactgtttgaaataccaagatatcattgtgaatggctcaaaagggtaa
tacatctgaaaaacctgaagtgtggaaaattccgatggagccaactcatgataacgcaga
agtcccattttgccatcttctcttggtatgaaacggtagaaaatgatccgagtatgccaa
ttgatactcttgattcatgccctatagtttgcgtagggtttaattgatctcctggtctat
cgatctgggacgcaatgtagaccccattagtggaaacactgaaagggatccaacactcta
ggcggacccgctcacagtcatttcaggacaatcaccacaggaatcaactacttctcccag
tcttccttgcgtgaagcttcaagcctacaacataacacttcttacttaatctttgattct
cgaattgtttacccaatcttgacaacttagcctaagcaatactctggggttatatatagc
aattgctcttcctcgctgtagcgttcattccatctttcta

ORF

ATGTCATCAATTACTACTTCAATCTTCAAGGTAACAGCTGAAATCCAAAGTTTTGGGGGAAAGCTAGTCAAACTT
CAACACAAGTCCGATGAGACGAAGACTGACATGGATGTGAACGTCTACCTTCCAGCTCAATTCTTTGCCAATGGA
GCCAAGGGAAAATCATTACCAGTTCTACTTTATTTGAGTGGTCTGACTTGCACTCCCAACAATGCCTCAGAGAAG
GCATTTTGGCAACCATATGCAAATAAGTACGGTTTTGCTGTGGTTTTCCCGGATACTTCACCCAGAGGGCTCAAC
ATCGAAGGAGAGCACGACTCTTATGATTTTGGATCCGGTGCCGGGTTCTACGTGGATGCCACTACTGAGAAATGG
AAGGATAATTATAGAATGTACAGTTATGTTAACTCGGAATTGCTACCCAAATTGCAGGCTGACTTCCCAATTCTA
AACTTTGACAATATTTCAATCACGGGCCACTCCATGGGAGGTTACGGAGCTTTACAGTTATTCTTGAGAAACCCG
GGAAAATTCAAGTCGGTTTCCGCATTTTCTCCAATCTCCAACCCCACTAAAGCCCCATGGGGTGAGAAGTGCTTC
TCTGGATACCTGGGACAGGACAAGTCCACTTGGACTCAGTACGACCCAACCGAATTGATTGGAAAATACCAAGGC
CCCTCAGATTCCAGCATTTTGATTCACGTTGGAAAGAGTGATTCGTTCTACTTCAAGGACCACCAGCTGCTACCT
GAGAACTTCTTGAAGGCTTCAGAGAACTCTGTGTTCAAGGGAAAAGTGGACTTGAACTTGGTAGATGGCTATGAC
CATTCTTACTACTTTATCTCTTCATTCACAGACGTTCATGCTGCTCACCATGCAAAGTATTTGGGGTTAAACTAG

Downstream

aactatgtaagttatataggctttaagtcaaacgagg**tta**acatgacctatgagaaatga
agagcaggcacatgataaccgtgaaagatgccaggaacggagctagccgtctagactttc
gttgcttgttgtttggatgagtcgcaggtgctggaaggctagaatcctgggtcatggaag
ccttatcagaaatgttttggattgtggtcacagtggttgtcagttcatcgtacgaaggga
gatcctcatgttgttccaaggcgttcaaagtgacaactttgtagattttggggttgtatt
ggtccatattgtccagatccactttagcgtcgaactctacgtgatcgaaagctggcaaaa
tggtatagttcgaaagctctgtgtttctattttgttgtgattgtgggtgaaccttattaa
tattgacgttagtgataactattttgctctgctttgcgccgttgatttggacgctgacat
tctggttgtctatgatctctggcgaactcgacgaagtcgctgaaaccgtttcaatgtatc
ttggtgggtt**cat**aaaagggtggtggttgcaagattagaattgtaagatctctactgcga
caatagccagggagttaaccgccgcgaacaataggccttaggcacctctgcgtaaataaa
tttgctgataagggaaagaatcgcgagatagaacaccttaccctcggtatgctctcaccg
acccgagagtttaccgagagttccgatattgagttggggtgagcagaaccagaattcagc
atgcttttgatttattgggtcaaaatgggttccgctccgccaaggttcgtatctttgagc
cttcgccaagcgggcgggcgtcccgtttgacaccacgctcgtaattcctttaatgcaggc

cgaaaacgtcacaccccacaactcacccccttcaattgttgttttgattcacagctgaaac
taagcctttattacgtaagccacaaatgggggactagaag

AA

MSSITTSIFKVTAEIQSFGGKLVKLQHKSDETKTDMDVNVYLPAQFFANGAKGKSLPVLL
YLSGLTCTPNNASEKAFWQPYANKYGFAVVFPDTSPRGLNIEGEHDSYDFGSGAGFYVDA
TTEKWKDNYRMYSYVNSELLPKLQADFPILNFDNISITGHSMGGYGALQLFLRNPGKFKS
VSAFSPISNPTKAPWGEKCFSGYLGQDKSTWTQYDPTELIGKYQGPSDSSILIHVGKSDS
FYFKDHQLLPENFLKASENSVFKGKVDLNLVDGYDHSYYFISSFTDVHAAHHAKYLGLN

SEQ0171

FDH: formate dehydrogenase (=FMD promoter, used in H. polymorpha)
Chr3-0932
gene known; promoter unknown

5' region

aaatggcagaaggatcagcctggacgaagcaaccagttccaactgctaagtaaagaagat
gctagacgaaggagacttcagaggtgaaaagtttgcaagaagagagctgcgggaaataaa
ttttcaatttaaggacttgagtgcgtccatattcgtgtacgtgtccaactgttttccatt
acctaagaaaaacataaagattaaaaagataaacccaatcgggaaactttagcgtgccgt
ttcggattccgaaaaactttggagcgccagatgactatggaaagaggagtgtaccaaaa
tggcaagtcgggggctactcaccggatagccaatacattctctaggaaccagggatgaat
ccaggtttttgttgtcacggtaggtcaagcattcacttcttaggaatatctcgttgaaag
ctacttgaaatcccattgggtgcggaaccagcttctaattaaatagttcgatgatgttct
ctaagtgggactctacggctcaaacttctacacagcatcatcttagtagtcccttcccaa
aacaccattctaggtttcggaacgtaacgaaacaatgttcctctcttcacattgggccgt
tactctagccttccgaagaaccaataaaagggaccggctgaaacgggtgtggaaactcct
gtccagtttatggcaaaggctacagaaatcccaatcttgtcgggatgttgctcctcccaa
acgccatattgtactgcagttggtgcgcatttagggaaaatttaccccagatgtcctga
ttttcgagggctacccccaactccctgtgcttatacttagtctaattctattcagtgtgc
tgacctacacgtaatgatgtcgtaacccagttaaatggccgaaaaactatttaagtaagt
ttatttctcctccagatgagactctccttcttttctccgctagttatcaaactataaacc
tattttacctcaaatacctccaacatcacccacttaaaca

ORF

ATGAAAATCGTTCTCGTTTTGTACTCCGCTGGTAAGCACGCCGCCGATGAACCAAAGTTGTATGGTTGTATCGAA
AATGAATTGGGTATTAGACAATGGCTTGAGAAGGGCGGCCATGAATTGGTTACTACATCAGACAAAGAGGGTGAA
AACTCTGAGTTAGAAAAGCACATTCCTGACGCTGATGTGATTATTTCCACTCCATTCCATCCAGCCTACATCACG
AAGGAGAGAATCCAAAAAGCCAAGAAGCTGAAGTTGTTGGTCGTTGCTGGTGTCGGTTCCGACCACATTGACTTG
GACTACATTGAACAAAATGGCCTAGATATTTCGGTCCTAGAGGTTACTGGTTCCAACGTTGTTTCAGTGGCTGAG
CATGTCGTTATGACTATATTGAACTTGGTGAGAAACTTTGTTCCAGCTCACGAGCAAATTGTTAACCACGGCTGG
GACGTTGCTGCCATCGCCAAGGACGCCTACGATATCGAAGGTAAGACCATCGCAACAATTGGTGCTGGAAGAATT
GGTTACAGAGTCTTAGAGAGACTTGTGGCTTTCAACCCTAAGGAATTGTTGTACTACGACTACCAAGGTCTTCCA
AAAGAGGCCGAGGAAAAAGTTGGTGCCAGAAGAGTCGACACTGTCGAGGAGCTGGTTGCTCAAGCCGATGTTGTT
ACCGTCAATGCCCCACTGCACGCAGGTACTAAGGGTTTAGTTAACAAGGAGCTTCTGTCCAAGTTCAAGAAGGGT
GCTTGGTTGGTTAACACAGCCAGAGGTGCCATCTGCAATGCTCAAGATGTCGCTGATGCCGTTGCATCTGGTCAA
TTGAGAGGTTACGGTGGTGACGTCTGGTTCCCTCAGCCAGCTCCAAAGGACCATCCATGGAGAGATATGAGAAAC
AAGTACGGATACGGAAACGCCATGACTCCTCATTACTCAGGTACCACTTTGGACGCCCAGGTCAGATATGCCGAA
GGTACCAAGAACATCTTGAACTCATTCCTTACCAAGAAGTTTGACTACAGACCTCAAGATGTCATTCTTTTGAAC
GGTAAGTACAAGACCAAGGCTTATGGTAATGACAAAAAGGTCGCATAA

Downstream

ttgaaatgtatttaatttgatattaagtaaatgaatgattatgactttatgaattcgcaa
tgttttctccttgattatttctgtattgtattggaatgattatagaatactcatatattg
attatagtattagcacataaaacgtttgttgttaaactcacttccgtacgcaaccatttc
tatttctagctatcttgataaggt**tca**ctgctcctgtactttgagtttctccttcaaact
tctcacctcctccttcagaagaccgatctcaatgtctttctgtttagaccgtccgtaact
cagcatttttgccatttcttcattttctttagtcaacgaacgtagtttagtcaacacctt
gtgagcgatctcctgctcatcgtaccgtccatctttcgtaattgcttgatattgtccttg
tactggattcatcgcaattgcggttaaaacgcactctaatcgttctttggtgtcttcaag
ctcagattgtacatcgttcagtttagtagacatttctgtgtacaattgctcgtagttacc
aatgtgcgacgatccggacattgcatcgtttttttccttggctctccatttctagcttctc
ttggttcaattgctcaacctctttctctaaagcttcaatctgggagttttgtattgcag

ctctagagctagaagttgagcggaatcttcacgcttcttgtggccattgtcgttgctatt
tgagttaaagggctcttgatacaccgacatgtaagtgtaattgttgccaaacttactaac
tccaggattcagagactgtctgtatggggtgttgtctcccataaatgtcattccactgtt
attaagaggcgaggtctgcggttgttgttgttggccgactccggtcataaaatcattgtg
aaagctagtattggttgaaggaaacaatgaagccttatagttgccttttgtttgcccaac
cgtagaagagctgccaattgttgaggctactgtatccaat

AA

MKIVLVLYSAGKHAADEPKLYGCIENELGIRQWLEKGGHELVTTSDKEGENSELEKHIPD
ADVIISTPFHPAYITKERIQKAKKLKLLVVAGVGSDHIDLDYIEQNGLDISVLEVTGSNV
VSVAEHVVMTILNLVRNFVPAHEQIVNHGWDVAAIAKDAYDIEGKTIATIGAGRIGYRVL
ERLVAFNPKELLYYDYQGLPKEAEEKVGARRVDTVEELVAQADVVTVNAPLHAGTKGLVN
KELLSKFKKGAWLVNTARGAICNAQDVADAVASGQLRGYGGDVWFPQPAPKDHPWRDMRN
KYGYGNAMTPHYSGTTLDAQVRYAEGTKNILNSFLTKKFDYRPQDVILLNGKYKTKAYGN
DKKVA

DAS: dihydroxyacetone synthase
2 genes; gene unknown; promoter unknown

SEQ0172

1. Chr3-0832

5' region

attactgttttgggcaatcctgttgataagacgcattctagagttgtttcatgaaagggt
tacgggtgttgattggtttgagatatgccagaggacagatcaatctgtggtttgctaaac
tggaagtctggtaaggactctagcaagtccgttactcaaaaagtcataccaagtaagatt
acgtaacacctgggcatgactttctaagttagcaagtcaccaagagggtcctatttaacg
tttggcggtatctgaaacacaagacttgcctatcccatagtacatcatattacctgtcaa
gctatgctaccccacagaaataccccaaaagttgaagtgaaaaaatgaaaattactggta
acttcaccccataacaaacttaataatttctgtagccaatgaaagtaaaccccattcaat
gttccgagatttagtatacttgcccctataagaaacgaaggatttcagcttccttacccc
atgaacagaaatcttccatttacccccccactggagagatccgcccaaacgaacagataat
agaaaaaagaaattcggacaaatagaacactttctcagccaattaaagtcattccatgca
ctccctttagctgccgttccatccctttgttgagcaacaccatcgttagccagtacgaaa
gaggaaacttaaccgataccttggagaaatctaaggcgcgaatgagtttagcctagatat
ccttagtgaagggttgttccgatacttctccacattcagtcatagatgggcagctttgtt
atcatgaagagacggaaacgggcattaagggttaaccgccaaattatataaagacaacat
gtccccagtttaaagtttttctttcctattcttgtatcctgagtgaccgttgtgtttaat
ataacaagttcgttttaacttaagaccaaaaccagttacaacaaattataacccctctaa
acactaaagttcactcttatcaaactatcaaacatcaaaa

ORF

ATGGCTAGAATTCCAAAAGCAGTATCGACACAAGATGACATTCATGAATTGGTCATCAAAACCTTCCGTTGTTAC
GTTCTCGACTTAGTCGAACAGTATGGTGGTGGTCACCCTGGTTCTGCCATGGGTATGGTCGCCATTGGTATCGCT
CTGTGGAAGTACCAGATGAAGTACGCTCCAAATGATCCAGACTACTTCAACAGAGATCGTTTTGTCTTGTCAAAC
GGTCACGTCTGTCTGTTCCAATACTTGTTCCAGCACTTAACTGGTTTGAAGGAGATGACTGTCAAGCAACTTCAA
TCTTACCACTCTTCCGATTATCACTCATTGACTCCTGGACACCCTGAAATTGAGAACCCTGCTGTTGAGGTTACC
ACTGGTCCCCTGGGACAAGGTATCTCTAACGCTGTCGGTATGGCCATTGGTTCAAAGAACCTGGCCGCTACTTAC
AACAGACCTGGCTTCCCTGTCGTTGACAACACTATCTATGCTATTGTTGGTGATGCTTGTTTGCAAGAGGGACCT
GCTTTGGAATCGATTTCCTTAGCCGGTCACTTGGCCTTGGACAACCTTATTGTGATCTACGACAACAACCAGGTT
TGTTGTGATGGTTCCGTCGATGTTAACAACACCGAAGACATCTCCGCAAAGTTCAGAGCTCAGAACTGGAATGTT
ATCGACATTGTAGACGGTTCTAGAGATGTCGCTACCATTGTCAAGGCTATCGATTGGGCCAAGGCTGAGACTGAG
AGACCAACTCTGATCAACGTTAGAACTGAAATTGGACAGGATTCTGCTTTCGGTAACCACCACGCTGCTCACGGT
TCTGCTCTAGGTGAGGAAGGTATCCGGGAGTTGAAGACTAAGTACGGTTTTAACCCTGCCCAAAAGTTCTGGTTC
CCTAAAGAAGTATACGACTTCTTTGCTGAGAAACCAGCTAAAGGTGACGAGTTAGTAAAGAACTGGAAAAAGTTA
GTTGATAGCTATGTCAAAGAGTACCCTCGTGAGGGACAAGAGTTCCTTTCTCGTGTTAGAGGTGAGCTTCCAAAG
AACTGGAGAACTTACATTCCTCAAGACAAGCCTACCGAACCAACCGCCACCAGAACCTCTGCTAGAGAAATTGTT
AGGGCCCTTGGAAAGAACCTTCCTCAAGTTATTGCCGGTTCCGGTGACTTATCTGTCTCAATTCTTTTGAACTGG
GACGGAGTGAAGTACTTCTTCAACCCTAAGTTACAGACTTTCTGTGGATTAGGTGGTGACTACTCTGGTAGATAT
ATTGAGTTTGGTATCAGAGAACACTCTATGTGTGCTATTGCCAACGGTTTGGCTGCATACAACAAGGGTACTTTC
TTGCCTATTACCTCTACCTTCTACATGTTCTACCTGTATGCAGCACCTGCCTTGCGTATGGCTGCTCTTCAAGAG
TTGAAAGCGATTCACATTGCTACACACGACTCTATTGGAGCTGGTGAAGATGGTCCAACCCACCAGCCTATTGCT
TTGTCTTCATTATTCAGAGCTATGCCCAACTTCTACTACATGAGACCAGCCGATGCTACCGAAGTTGCAGCTCTG

TTTGAAGTGGCTGTTGAGCTTGAACACTCCACATTGCTTTCTCTGTCCAGACACGAGGTTGACCAATACCCAGGT
AAGACTTCTGCCCAAGGAGCCAAAAGAGGTGGTTACGTTGTTGAAGACTGCGAAGGAAAGCCAGATGTGCAACTG
ATCGGAACTGGTTCCGAGTTGGAATTCGCTATTAAGACTGCTCGTTTGCTAAGACAACAGAAGGGATGGAAGGTC
AGAGTTCTGTCATTCCCATGTCAGAGATTGTTTGACGAGCAGTCTATTACTTACAGACGTTCCGTCCTTAGAAGA
GGAGAAGTTCCAACTGTCGTTGTTGAGGCCTATGTCGCATACGGATGGGAGAGATACGCCACTGCTGGTTACACC
ATGAACACCTTCGGTAAGTCTCTTCCTGTTGAGGATGTCTACAAATACTTCGGATACACTCCTGAGAAGATTGGT
GAGAGAGTGGTTCAATATGTCAACTCTATCAAGGCTAGTCCTCAAATCCTTTACGAATTCCACGACTTGAAGGGA
AAACCAAAGCATGACAAGTTGTAA

Downstream

acgggaagtctttacagttttagttaggagcccttatatatgacagtaatgctagtacgt
tttgttttgtttaattaataacttagtttatgttagcctagtatagactccatcaatttt
ttttgttattacgtaagccgcgatgataatatctgatgaaaaattcctatcagaaaataa
tttatcaaaagtttcatgcgat**atg**agactaagtagaatagggactcccaaagtgtcagt
cacaagggtcattcccgttcgtaatgtggtgatagcgaggagaaaacctgtcagagcaag
taacaccgacgcaaagacatggctaatgaaagaagagcagagaagaataagacagaagga
gcaggagatgaaacaaaggctagaggaactagaaaggttcaaaacaaaagtacagaaatc
atatataaggaaagaggataggcatttggcacaagagatagaaaaggatcttgacataat
cactgatgattacaatttggacagcgatgtagacctggtctttggagaactcatgcaagc
agaagagcagcccaaatcgttgaagagtttaccaggagccagtgatgcaagtgataacac
agataaatcggtggaactgttttcctttccatctccgaatctgacgctacctgaaaaggt
gatacaccatattgggccactggtgaagcacatcagtaatcctgaaaacattcaatgggg
gagacttttactggatttggaaaaaaatcaggggtttaacggtctatctgccgtagatgt
taccagactgattcaaaatatacccaaagaggaaaaatatcagcatatgtctttgattca
tgaaatgatgtttaacagtgggatcagtcctgatcggtacttgacagatttgatgatgac
tgccttctctgaaaggagctactatgaaccattggttgaggctcttttccaagactatga
catcaatggctgggctccaacagattatacttttggagca

AA

MARIPKAVSTQDDIHELVIKTFRCYVLDLVEQYGGGHPGSAMGMVAIGIALWKYQMKYAP
NDPDYFNRDRFVLSNGHVCLFQYLFQHLTGLKEMTVKQLQSYHSSDYHSLTPGHPEIENP
AVEVTTGPLGQGISNAVGMAIGSKNLAATYNRPGFPVVDNTIYAIVGDACLQEGPALESI
SLAGHLALDNLIVIYDNNQVCCDGSVDVNNTEDISAKFRAQNWNVIDIVDGSRDVATIVK
AIDWAKAETERPTLINVRTEIGQDSAFGNHHAAHGSALGEEGIRELKTKYGFNPAQKFWF
PKEVYDFFAEKPAKGDELVKNWKKLVDSYVKEYPREGQEFLSRVRGELPKNWRTYIPQDK
PTEPTATRTSAREIVRALGKNLPQVIAGSGDLSVSILLNWDGVKYFFNPKLQTFCGLGGD
YSGRYIEFGIREHSMCAIANGLAAYNKGTFLPITSTFYMFYLYAAPALRMAALQELKAIH
IATHDSIGAGEDGPTHQPIALSSLFRAMPNFYYMRPADATEVAALFEVAVELEHSTLLSL
SRHEVDQYPGKTSAQGAKRGGYVVEDCEGKPDVQLIGTGSELEFAIKTARLLRQQKGWKV
RVLSFPCQRLFDEQSITYRRSVLRRGEVPTVVVEAYVAYGWERYATAGYTMNTFGKSLPV
EDVYKYFGYTPEKIGERVVQYVNSIKASPQILYEFHDLKGKPKHDKL

SEQ0173

2. chr3_0834

5' region

aataaaaaaacgttatagaaagaaattggactacgatatgctccaatccaaattgtcaaa
attgaccaccgaaaaagaacaattggaatttgacaagaggaacaactcactagattctca
aacggagcgtcacctagagtcagtttccaagtcaattacagaaagtttggaaacagaaga
ggagtatctacaattgaattccaaacttaaagtcgagctgtccgaattcatgtcgctaag
gctttcttacttggaccccatttttgaaagtttcattaaagttcagtcaaaaattttcat
ggacatttatgacacattaaagagcggactaccttatgttgattctctatccaaagagga
ttatcagtccaagatcttggactctagaatagataacattctgtcgaaaatggaagcgct
gaaccttcaagcttacattgatgat**tag**agcaatgatataaacaacaattgagtgacagg
tctactttgttctcaaaaggccataaccatctgtttgcatctcttatcaccacaccatcc
tcctcatctggccttcaattgtggggaacaactagcatcccaacaccagactaactccac
ccagatgaaaccagttgtcgcttaccagtcaatgaatgttgagctaacgttccttgaaac
tcgaatgatcccagccttgctgcgtatcatccctccgctattccgccgcttgctccaacc
atgtttccgccttttcgaacaagttcaaatacctatctttggcaggacttttcctcctg
ccttttttagcctcaggtctcggttagcctctaggcaaattctggtcttcatacctatat
caacttttcatcagatagcctttgggttcaaaaaagaactaaagcaggatgcctgatata
taaatcccagatgatctgcttttgaaactattttcagtatcttgattcgtttacttacaa
acaactattgttgattttatctggagaataatcgaacaaa

ORF

ATGGCTAGAATTCCCAAAGCAGTTTCTTACAATGATGACATCCATGACTTGGTCATCAAAACCTTCCGTTGTTAC
GTTCTCGACTTAGTCGAACAGTATGGTGGTGGTCACCCTGGTTCTGCCATGGGTATGGTCGCCATTGGTATCGCT
CTGTGGAAGTACCAGATGAAGTACGCTCCAAATGATCCAGACTACTTCAACAGAGATCGTTTTGTCTTGTCAAAC
GGTCACGTCTGTTTGTTCCAATACTTGTTCCAGCACTTAACTGGTTTGAAGGAGATGACTGTCAAGCAACTTCAA
TCTTACCACTCTTCCGATTATCACTCATTGACTCCTGGACACCCTGAAATTGAGAACCCTGCTGTTGAGGTTACC
ACTGGTCCCCTGGGACAAGGTATCTCTAACGCTGTCGGTATGGCCATTGGTTCAAAGAACCTGGCCGCTACTTAC
AACAGACCTGGCTTCCCTGTCGTTGACAACACTATCTATGCTATTGTTGGTGATGCTTGTTTGCAAGAGGGACCT
GCTTTGGAATCGATTTCCTTAGCTGGTCACTTGGCCTTGGACAACCTTATTGTGATCTACGACAACAACCAGGTT
TGTTGTGATGGTTCCGTCGATGTTAACAACACCGAAGACATTTCTGCTAAGTTTAGAGCTCAGAACTGGAATGTC
ATTGAAGTCGAGAATGGTTCTAGAGATGTTGCTACCCTTGTCAAGGCCATCGAATGGGCCAAGGCTGAGAATGAG
AGACCAACTCTGATCAACGTTAGAACTGAAATTGGACAGGATTCTGCTTTCGGTAACCACCACGCTGCTCACGGT
TCTGCTCTTGGTGAGGAAGGTATCCGGGAGTTGAAGGCCAAGTACGGTTTCGATGTCGCTAGAAAGTTCTGGTTC
CCACAGGAGGTCTATGATTTCTTTGCTGAAAAACCAGCCGAGGGTGATCAACTAGTTGCTAACTGGAAGAAACTT
TTGGATGAGTACGTTAAGAACTATCCTCAAGAAGGTGAGGAATTAAAGGCCCGTATTAGAGGTGAACTTCCAAAG
AACTGGAAGAGTTTCATTCCACAGGACAAACCAACCGAGCCAACTGCTACCAGAACCTCTGCTAGAGAAATTGTT
AGATCTCTGGGACAAAACCTTCCTCAGGTTATTGCTGGTTCTGGTGACTTGTCCGTGTCCATTCTTTTGAACTGG
GGAGGAGTTAAGTACTTCTTCAACCCTAAGTTACAAACTTTCTGTGGATTGGGTGGTGACTACTCTGGTAGATAT
ATTGAGTTTGGTATCAGAGAACACTCTATGTGTGCTATTGCCAATGGTTTGGCTGCATACAACAAGGGTACTTTC
TTGCCTATTACCTCAACTTTCTACATGTTCTACCTGTATGCAGCACCTGCCTTGCGTATGGCTGCACTTCAAGAG
TTGAAAGCAATTCACATTGCTACACACGACTCCATCGGAGCTGGTGAAGATGGTCCAACGCACCAGCCTATTGCT
TTGTCTTCATTATTCAGAGCTATGCCCAACTTCTACTACATTAGACCAGCCGATGCTACCGAGGTTGCAGCTCTG
TTTGAAGTAGCTGTTGAGCTCGAGCACTCCACCTTGTTCTCTCTGTCCAGACACGAGGTTGAGCAATACCCAGGT
AAGACTTCGGCTGAGGGAGCCAAAAGAGGTGGTTACGTCGTTGAAGACTGTGAGGGTAAGCCAGACGTCCAATTA
ATTGGTGCTGGTTCCGAATTGGAGTTTGCCGTCAAAACTGCTCGTTTGCTAAGACAACAGAAGGGATGGAAGGTC
AGAGTTCTGTCATTCCCATGTCAGAGACTGTTTGACCAACAATCCCTGGCATACAGACGTTCTGTCCTTAGAAGA
GGAGAGGTTCCAACTGTCGTTGTTGAGGCCTATGTCGCATACGGATGGGAGAGATACGCCACTGCTGGTTACACC
ATGAACACCTTTGGTAAGTCTCTTCCTGTTGAGGATGTCTACAAATACTTCGGATACACTCCTGAGAAGATTGGT
GAGAAGGTTGCTGCATACGTCAACTCTATTAAGGCTAGTCCTCAAATCCTTTACGAATTCACCGATTTGAAGGGA
AAACCAAAGCACGACAAACTATAA

Downstream

gtagatttggccactaacgggttagtagttgtgtaagtctattaaatttgatttttgttt
atggatgatcatcgtagtggctatctgtttacctgtaggacatcctagggtgggatggtg
atgtacacccccctcaatcttcagatgcaacactatgtggtaggtcattgacataaggttt
aggaaagacctgttttttgaccaataaatggaacaggaaggaaaggaggaaccagtttac
gaaccccgtcggacatacaaaaccaaaaaatgagagcatgtagctcctcttgatggcttt
gcaggaaacactgctgcactcagaaatggcaacgcccggatattgaagggacgaagattg
tgggggagaaacctgtcactgcagatatgaggagtttgatttataggaaacatttttcaa
tgtgggtgaatcctctgcgacctacatactatgtgcataatacgtgcataaaaagacaac
aagataaatatagtaaaatatggggtaatagagagggctcaaccgaacgttagtggtagg
attggggaaacagtattggagagagtggaatagatacacagctagtgttggattgtatag
tagtttcgtgtttataggtcgctgattctagtactgcgtacagctgaagagagaaaaaaa
aatctgaaaactggggtaaagaactccttaaaaatttttcatatgtatgatgaaagactg
ctgcaggagacgttcacacttgaaaaatttgatcgaagcgagctacaagtccatatcgtg
ttggaacagaatctggaggtgtaaacaccgtcattccctaatctaagtttctgtccagcc
gtttagaaagttcctcttcgtcaataaagtttaactttgaacgtgagagttgagctaatc
ggaaatactcccgagagcacacccgagcagtgaagcggggcatcatcaaaagaattaagg
ggtgctgggtgcggattgcgcgcagtagctggcatcaatg

AA

MARIPKAVSYNDDIHDLVIKTFRCYVLDLVEQYGGGHPGSAMGMVAIGIALWKYQMKYAP
NDPDYFNRDRFVLSNGHVCLFQYLFQHLTGLKEMTVKQLQSYHSSDYHSLTPGHPEIENP
AVEVTTGPLGQGISNAVGMAIGSKNLAATYNRPGFPVVDNTIYAIVGDACLQEGPALESI
SLAGHLALDNLIVIYDNNQVCCDGSVDVNNTEDISAKFRAQNWNVIEVENGSRDVATLVK
AIEWAKAENERPTLINVRTEIGQDSAFGNHHAAHGSALGEEGIRELKAKYGFDVARKFWF
PQEVYDFFAEKPAEGDQLVANWKKLLDEYVKNYPQEGEELKARIRGELPKNWKSFIPQDK
PTEPTATRTSAREIVRSLGQNLPQVIAGSGDLSVSILLNWGGVKYFFNPKLQTFCGLGGD
YSGRYIEFGIREHSMCAIANGLAAYNKGTFLPITSTFYMFYLYAAPALRMAALQELKAIH
IATHDSIGAGEDGPTHQPIALSSLFRAMPNFYYIRPADATEVAALFEVAVELEHSTLFSL
SRHEVEQYPGKTSAEGAKRGGYVVEDCEGKPDVQLIGAGSELEFAVKTARLLRQQKGWKV
RVLSFPCQRLFDQQSLAYRRSVLRRGEVPTVVVEAYVAYGWERYATAGYTMNTFGKSLPV

EDVYKYFGYTPEKIGEKVAAYVNSIKASPQILYEFTDLKGKPKHDKL

SEQ0174

TPI: triosephosphate isomerase (= also a gene of the glycolysis)
gene unknown; promoter unknown

chr3_0951

5' region

ttcaacgagacactcttccgtcagttccaaaaccataagtttgccgatgtgttggtcctt
gtaacgcatggaatttgggccagggtatttttgatgaaatggttcagatggtctgtggag
gagtttgaaggcttacgaaatataccacattgccagtttatacagatggttaagggtgaa
aatcaacgttacaccttgacgaccccattattacgatggcgtgaaggagatgaagaccgg
gtagaagaaataagaaaagcggtacagtttaggtccggagatctagggaaggaggcctta
gcttatattgtagctgctgagagagaggcagctgctggaagatctgaaggccctatcacg
tatgatgatggtgatgaccat**tag**agaacgcccagagattgatagccagttcttggacaa
caattcggaactttattcacggtgcaaacatgatttgtgtggatagcttcaagtcagaca
tttcatctcatccccctttttactgctgctaatcaccgttagtccgacagttactctaat
caatatttattagtgttttagttgcgcaaaactcgagcctcttttccttatctcttgaca
cttcctggagtcgaagtttttcagcgcaaattcactctacaatgtctaccgatactagac
cgcctatcttccccctctaaatagcctattggaagggtgcaataaggtatataaatctgg
cgcgattcccccggacttttatgatccacatcacctcatcttactgccctcactctcttt
cctgatcctcccaggtccaccgatttcctcactatcgtcggatttctccttccagcgccc
tagagaattccgtaaccaccgcaaaaatagcagcccccccctcacccatttttttattta
aaagaacaccttactggcccgttttcgtttctcctttactacaattgatttttaattttc
agtttttttcattgatatacaagatctatcacaaacaca

ORF (intron in bold and between brackets)

ATG**(GTACGTGCATTGAGTCTCATAAGTGCCATCCAGTCACACCTGGCAGGTGGAGGCAATGGGAGGTCGGAGAA
ACAGAACCCACATCAGAAGGACCATACTAACTCTTCCCAG)**GCTAGAACATTTTTCGTAGGAGGAAACTTCAAAA
TGAATGGCTCCAAGAAGTCAATTCACGAGATTATCGAGAGATTGAACAACACCAAGTTACCAGAGAACGTAGAGG
TTGTGATTGCTCCTCCTGCTCCCTACTTGCAGCAGGCTGTCACCGAGAACAAGCAGAAGACTGTTTACGTCTCCG
CACAAAACAGTTTCGACAAAGCTTCTGGTGCCTACACAGGTGAAGTTTCTGTCGAGGCTTTGAAGGACTTGGGTG
TCCCTTACGTTATCTTGGGTCACTCCGAACGTAGAACTATTAACAAGGAAGACGATGCTTTCATTGCTTCCAAGA
CCAAATTTGCTCTTGACCAGGGTTTGAAGGTTATCCTCTGTATTGGTGAGACACTCGAGGAGAAGCAAGCCAACA
TCACGCTAGACGTTGTTAAAAGACAACTGCAAGCTGTTGTCGATGTTGTTTCCGACTGGACTAACATTGTCGTTG
CTTACGAGCCTGTCTGGGCTATCGGTACTGGTCTTGCTGCCACTCCAAGCGATGCCCAAGACGTTCACAAGCAAA
TTAGAGATTTCCTTGCTACTGTCATTGGTAAGGACCAAGCTGAAAAAGTCCGTATCTTGTACGGTGGATCCGTCA
ACGGCAAGAACGCTGTCGAGTTCCGTGACAAGGCCGATGTTGATGGATTCTTGGTTGGTGGAGCTTCTCTGAAAC
CTGAATTCGTGGACATTATTAATTCCAGAAACTAG

Downstream

tttacatatgaacatattactactctatattcgggacagcctcgattatttctctttctc
ttcgtctcttgtttaaagtcttctt**tca**tatcgttccttttcatcctctcgttccgctc
gatccttcaacgttgaaagagccagaggtgtcatattgggaagagaaccaccaacgggat
ctaagtttggagattctgaccctatatgctgtatgacggattgttgctgtatctgtaacc
caaggggtggagtttgtagccacatctttgatggtgacccgacgactgagactggcatca
agccaggtgagtcaggggtgcggaccgtacttccagagccgtgttgcaccagtgttattg
caggaccatgagtgggcaccgccgtataatgctgatgcctttggagcatgagatggttgc
tggaattagctcttagatgctgtggcggtagccggatagcctttacaacttgttcctctt
catcgtccactactgcactagaagtgctatcattactattgttctgttggttgtgtatat
tcaatgaaagtaaactaatagtagaggagctgggtactagacggccctccacatctttac
caacagacgaggccagaagtggtggtgacggatgttcatcagtgttgttagtccatctac
ttgaatttgtattgagcccgcttgctggaatttgaatgggcatgccctgcatgctttcca
ttagctcaggctctgaaagtactgagcctctaaaagatggagg**cat**gtctaagtcgaagt
ggattggttttgataagtttgctgtaaactaatttttcgtatagagttcgcgaaaattct
caccaccaccaagggtcaagtcactgatcgcttcaacccacaaaacgacgata**tta**aagt
tcaacatattgttctaggtacgccagtccgcgatccaagtttgaaatctccacatctaca
ggtgcgtactgtagtagttcatccaaagctgggatcttgc

AA

MARTFFVGGNFKMNGSKKSIHEIIERLNNTKLPENVEVVIAPPAPYLQQAVTENKQKTVY
VSAQNSFDKASGAYTGEVSVEALKDLGVPYVILGHSERRTINKEDDAFIASKTKFALDQG
LKVILCIGETLEEKQANITLDVVKRQLQAVVDVVSDWTNIVVAYEPVWAIGTGLAATPSD

AQDVHKQIRDFLATVIGKDQAEKVRILYGGSVNGKNAVEFRDKADVDGFLVGGASLKPEF
VDIINSRN

DAK: dihydroxyacetone kinase

gene known; promoter unknown
SEQ0175
chr3_0841:
5' region
tgaacatgatatccaaaccttttggagtaggttgtaaaagggtatttaggtgtttttcgt
ccagggaaatcagtgttctttctcttggacaatgaattagcctagctctcaggaacatgt
tcagcagtgctagacaggagtccaaatccatagaacatgaaatacagaggttcttataaa
cggactttatgttcaaagacaatccattcaacttgactattgcatgagatgccttgaaac
aatagaggttgttatggagttggagtttgaatatgccaaacaatcccgattcacttgggt
ctccagcttccaattggagagacaagatcaacaggctaaagaaatacttgatctctttag
aattgaacccgatagactgttcatttgccattgt**cat**ctgctgatgctgtgagggagaaa
gaagtaggggtgatacatggtttataggcaaagcatgtttgtttcagatcaaagattagc
gtttcaaagttgtggaaaagtgaccatgcaacaatatgcaacacattcggattatctgat
aagtttcaaagctactaagtaagcccgtttcaagtctccagaccgacatctgccatccag
tgattttcttagtcctgaaaaatacgatgtgtaaacataaaccacaaagatcggcctccg
aggttgaacccttacgaaagagacatctggtagcgccaatgccaaaaaaaaatcacacca
gaaggacaattcccttccccccccagcccattaaagcttaccatttcctattccaatacgt
tccatagagggcatcgctcggctcattttcgcgtgggtcatactagagcggctagctagt
cggctgtttgagctctctaatcgaggggtaaggatgtctaatatgtcataatggctcact
atataaagaacccgcttgctcaaccttcgactcctttcccgatcctttgcttgttgcttc
ttcttttataacaggaaacaaaggaatttatacactttaa
ORF
ATGTCTAGTAAACATTGGGATTACAAGAAAGATCTGGTACTCAGCCATCTGGCTGGATTATGTCAGTCCAACCCC
CATGTAAGGCTAATTGAGTCCGAAAGAGTCGTCATCTCCGCCGAGAACCAGGAGGACAAGATTACCTTGATCTCT
GGTGGAGGTTCAGGACACGAGCCTCTACATGCTGGCTTTGTCACCAAGGACGGTCTTTTGGATGCCGCTGTGGCC
GGGTTCATCTTTGCCTCGCCATCTACTAAACAGATATTTAGCGCAATCAAAGCCAAACCTTCGAAGAAGGGAACT
TTGATCATCGTCAAGAACTATACAGGAGACATTCTTCACTTCGGTCTTGCCGCCGAAAAGGCCAAGGCCGAAGGT
CTCAACGCAGAACTGCTGATTGTTCAGGACGATGTCTCTGTTGGTAAAGCCAAGAATGGCCTGGTAGGGCGTAGA
GGTTTGGCTGGAACTTCTTTAGTGCACAAGATTTTGGGTGCCAAAGCTTATTTACAAAAGGACAACCTGGAATTG
CATCAGCTGGTCACTTTCGGTGAGAAGGTCGTTGCCAACCTGGTTACAATTGGTGCTTCCTTGGACCATGTCACC
ATTCCAGCCAGAGCAAACAAACAGGAGGAAGATGATTCAGATGACGAGCACGGCTACGAAGTGCTAAAGCACGAT
GAATTTGAAATCGGAATGGGTATTCACAACGAACCAGGTATCAAGAAGTCGTCGCCAATTCCAACAGTAGACGAA
TTGGTGGCTGAGCTACTGGAGTACCTTCTTTCGACTACCGATAAGGACCGAAATTACGTCCAATTTGACAAGAAT
GATGAAGTTGTGCTACTGATCAATAACTTGGGAGGAACCTCAGTGTTGGAATTGTACGCTATTCAGAACATTGTT
GTCGACCAACTGGCCTCCAAATATTCCATTAAACCAGTCCGCATTTTCACTGGCACTTTCACCACATCACTGGAT
GGGCCCGGTTTTTCTATCACTTTATTGAATGCTACCAAGACAGGAGATAAAGATATCCTAAAGTTTTTAGACCAC
AAGACTAGTGCCCCTGGCTGGAACTCCAACATATCAGACTGGTCTGGCAGGGTTGACAATTTTATTGTGGCTGCC
CCAGAAATAGACGAAGGAGACAGTTCATCTAAGGTATCTGTTGATGCTAAGCTGTATGCTGACCTTCTGGAGTCT
GGTGTGAAGAAGGTCATTTCAAAAGAGCCAAAGATCACATTGTATGACACTGTTGCAGGAGATGGTGATTGTGGT
GAAACGTTGGCGAATGGTTCCAATGCCATTTTAAAGGCTCTGGCTGAAGGAAAGCTTGATCTCAAAGACGGTGTT
AAATCACTGGTGCAAATTACAGACATCGTGGAAACTGCCATGGGTGGAACATCTGGAGGTCTGTACTCCATTTTC
ATCAGCGCTTTAGCAAAGTCTTTGAAAGAGAAGGAACTTTCCGAGGGTGCCTACACACTGACCCTTGAAACGATC
TCTGGATCCTTACAGGCAGCACTTCAGTCTCTATTCAAATACACAAGAGCACGTACTGGAGACAGAACTCTTATT
GACGCTTTGGAGCCATTTGTAAAGGAATTCGCCAAGTCTAAGGATCTGAAACTTGCAAACAAAGCTGCCCATGAT
GGTGCTGAAGCCACAAGAAAGCTCGAAGCCAAGTTTGGAAGAGCATCTTATGTTGCCGAAGAAGAGTTCAAACAA
TTCGAAAGTGAAGGTGGACTACCTGATCCTGGTGCTATTGGTCTTGCCGCTTTGATTTCAGGAATTACTGATGCA
TACTTCAAATCTGAAACCAAGTTGTAG
Downstream
caatccaatcccaagagcaagttatgaatgagtaacgttgaggagtacctctaatgatta
cttacggctaacggccatatctagtttttataaacatttaatgataagcacggttctgaa
agttgatcaaattgtaagtgtgacttgtgatctatacagtcaaacttgtctagcaagtta
actgacaattgagagcactaataattttcatgtggtctcttagcaagtagaccatagaaa
tacctgtttcatctcttctggaggtgaccctcaccgaatcattggggtcaaatgcacttg
gactacagtaaacctactccttttgtgactgtaccctaacaacgaataaccccataacaa
tgcaatcaatatttcagaggttaggatgtattacagacgtaaagagtcagttcatggaag

actatcggtcggtaggtagtcatttccccattaggaagtctgtttagtcgtgactcagcg
aatacgttttcttatttaagttgtaagatcccgaattgcgagaggctcatctcggacaac
gctggtctctatccttac**atg**accaaccaatcaacagtggtggatttacgcctttcatcc
aagagagttgttggcaaaccagtcaagttgcccacagtcctagcgtgctcagggtcagat
tcttccggtggtgcagggatcgaagcagatatcaaatccatcacggcttttgggtgctat
gcgctaacagcaattacatctttaactgcccagaataccaaaggtgtcaccagtatagaa
aacaccgacccaaagtttttcgaagagattttagaggcaaattttgaggacattgaaatc
gatgtggtgaaaactggactgttaaaccctgagtcatctcgtttattgctgaaatttttta
gataaataccacaaaggaaagccatttgtcctggatccggtcttagtggctacgtctggt
tcaatgcttgcagatcaacacgaattagggttcaccattg

AA

MSSKHWDYKKDLVLSHLAGLCQSNPHVRLIESERVVISAENQEDKITLISGGGSGHEPLH
AGFVTKDGLLDAAVAGFIFASPSTKQIFSAIKAKPSKKGTLIIVKNYTGDILHFGLAAEK
AKAEGLNAELLIVQDDVSVGKAKNGLVGRRGLAGTSLVHKILGAKAYLQKDNLELHQLVT
FGEKVVANLVTIGASLDHVTIPARANKQEEDDSDDEHGYEVLKHDEFEIGMGIHNEPGIK
KSSPIPTVDELVAELLEYLLSTTDKDRNYVQFDKNDEVVLLINNLGGTSVLELYAIQNIV
VDQLASKYSIKPVRIFTGTFTTSLDGPGFSITLLNATKTGDKDILKFLDHKTSAPGWNSN
ISDWSGRVDNFIVAAPEIDEGDSSSKVSVDAKLYADLLESGVKKVISKEPKITLYDTVAG
DGDCGETLANGSNAILKALAEGKLDLKDGVKSLVQITDIVETAMGGTSGGLYSIFISALA
KSLKEKELSEGAYTLTLETISGSLQAALQSLFKYTRARTGDRTLIDALEPFVKEFAKSKD
LKLANKAAHDGAEATRKLEAKFGRASYVAEEEFKQFESEGGLPDPGAIGLAALISGITDA
YFKSETKL

FBA: fructose 1,6-bisphosphate aldolase
2 genes; genes unknown; promoters unknown
SEQ0176 See SEQ0065
chr1-1_0072:
SEQ0177 See SEQ0066
chr1-1_0319:

SEQ0178

FBP: fructose 1,6-bisphosphatase
gene unknown; promoter unknown

chr3_0868:

5' region

tttgatgatgatttgttgtgaacgttccatgtgtttgaatctttggacggccggtaaaaa
ccgactgtttttaatgattgctggttagccttgtcgacaaatgatttgtatgaagctctg
accactgaaaggatttcagatcggttgaaatgatcgacttttgctgtttgagtgttggta
ttaggagttgtagtggcaccagtcggggtgctacttcgtgatgagctgggattctttgcg
ttcacagccttcgaagtggcatgcttgtcatttcttatcttatccttaccagcagccagg
ttcgcccactcaccttgaactggtacaaggatatttctctgtttttgctcggtggttcc**c**
attgctcaaacgagtggagagggaaatcgattcagcagttaaatcaatgctggaaaatat
tcgagattacctaatcggatctggaacttacttcgacctgacattttcttgcctggggag
ccacgatcgattatgtaatcaagaatatggacagagggaaacagatttagctgtcaaaag
cccaagagaagctaccgatcaatggatgcggatagataaagaaaagtccttttttttttca
ttagccatccgagttgtccaatcaaatgtctgcctgctacgctggagaggaatcacgcgt
gtttaacattcggattgtcgcctaaaataagcctattacctacacagtaaaacccggggg
gtgctttggtatcaatgaccccgggattttatccaccagttttttttctttctggcaagag
tgcattgcatccccgtacaaatagtagcaacctccacaagaggaatcccctatgagcgag
aagtccatagtaataccccccgcggaaaagagatattttgtttccgtgttgcccttgaact
tcagtttccccccatcagtttatatagtagccgggttcccaatctctagcccttctttcct
cctatttcattcctctcttcttacgttatcttacattagc

ORF

ATGTCCAATAACACCACCCAAAACTTGGCTGAACAGAAGGGTATTCAGACCGACTTGGTCACTCTGACTCGTTTC
ATCTTGGACGAGCAGAAGAAATCTGCCCCAAACGCTACCGGTGAGCTGACTTTGCTTCTGAACTCCCTGCAATTT
GCCTTCAAATTCATTGCTCACACCATCAGAAGATCCGAATTGGTCAACTTAATTGGATTGGCCGGTGTGACCAAC
GCCACTGGTGACGATCAGAAAAAGCTGGATGTTATTGGAGATGAGATCTTCATCAATGCCATGAAGGGATCCGGT

AACGTTAAGTTACTTGTTTCCGAAGAGCAAGAAGACCTTATTGTGTTTGAGTCCTCCAAGGGAAACTATGCTGTT
GTCTGCGACCCAATTGACGGTTCATCCAACTTGGACGCTGGTGTCTCAGTTGGTACCATTTTTGGTGTATACAAA
CTGTTGCCAGGCTCAGCTGGTTCTATTAAGGATGTTTTGAGATCAGGAACAGAGATGGTGGCTGCCGGTTACACC
ATGTACGGTGCATCTTCTCACTTGATGCTTACTACCGGAAATGGAGTCAACGGTTTCACCTTGGACACCGACTTG
GGAGAGTTCATTCTCACGTATCCATCCTTAAAAATTCCGCACACCAGGGCTATCTACTCTATCAATGAAGGTAAC
TCGCACTACTGGACCGATGGAGTCAATGAGTACATTGCTTCTTTGAAGAAACCCCAAGCAAACGGAAAACCATAC
AGTGCCCGTTATATCGGATCAATGGTTGCTGATGTCCACAGAACTCTGTTATATGGAGGTATCTTTGGCTATCCA
GCAGACTCCAAGTCTAAGTCCGGTAAACTGAGAGTCTTGTACGAGTGTTTCCCAATGGCCCTCCTGTTGGAGCAG
GCTGGTGGTGAGGCAGTCAACGATAAGGGTGAGAGAATTCTAAACTTGGAGCCAAAACAAGTGCATGAGAGATCT
GGTATCTGGTTGGGTTCCAAGGGAGAAGTAGAGAGATTACTACCCTATCTAACAAAAAAAATTAAGATCCAATCT
GTGAATCTCTAG

Downstream

tgcgttcgattggcactgtttccgagatttgatactttgtacacgatgttatatgaggtt
atatacttgataagagggtttttacgtttgcaattagcacaatttcggagtagcactggc
gggagtgaaccttgagtagtctggatcaatgtaatcttcgtataggctagacaccccgga
ttgggagtgctgacgaaatgatgttgggatgtgatgacatcatagggataatagaaaagt
aaggttccgcgtgagccgttgaacgcgcactggaatggatggtctgtgacgtagccagac
tgaacttgaaattccttccaagaaagtacatttttatttcattcattcattcgaaaggga
ggcttgtgggggaaccccccaatcaaatacctaactactacttacaatatccaaacctaac
aacgagctcctcattcccga**tca**ctccttctttctaccaaataatcctctcttcttcttg
cccttaccaccacttcccgttggtgaattaacatcctgttcaaaagtattctgatctcct
gaaaacgagatcacggggttttgaccagcaatgggggacgaggtctcattctgtcgtga
tattcgtagtttctgccggactcataacctggaaaatcaggacgtaccctccaacctaaa
gtcggggtctccagtagttgtttataagtagtatctccagtgatggcatattcaaaggaa
cggatcgtatctaaaggcctctcgtctctcgatctggcaggattggaaacgtcggggtta
ttgattggttgtccaaaaatgtcacggatgtactctccatcgcccgacttattactaaac
gattgttgtctggcatgagtgtagtccgttgattgttcgaacggctgagcctcgttgaca
gccatcaagatggggtcgtgaatgttgtttgatctgtacttgagtctgtccgtgtcgtct
ttggtgacgtgaacttttggtactttaccaacatcagttttc

AA

MSNNTTQNLAEQKGIQTDLVTLTRFILDEQKKSAPNATGELTLLLNSLQFAFKFIAHTIR
RSELVNLIGLAGVTNATGDDQKKLDVIGDEIFINAMKGSGNVKLLVSEEQEDLIVFESSK
GNYAVVCDPIDGSSNLDAGVSVGTIFGVYKLLPGSAGSIKDVLRSGTEMVAAGYTMYGAS
SHLMLTTGNGVNGFTLDTDLGEFILTYPSLKIPHTRAIYSINEGNSHYWTDGVNEYIASL
KKPQANGKPYSARYIGSMVADVHRTLLYGGIFGYPADSKSKSGKLRVLYECFPMALLLEQ
AGGEAVNDKGERILNLEPKQVHERSGIWLGSKGEVERLLPYLTKKIKIQSVNL

8 Annotation of homologues of *S. cerevisiae proteases*

8.1 Serine-type peptidases

SEQ0179

NMA111 homologue: Protein of unknown function which may contribute to lipid homeostasis and/or apoptosis; sequence similarity to the mammalian Omi/HtrA2 family of serine proteases

S. cerevisiae null mutant: viable; increased heat sensitivity; increased lifespan; normal sporulation

S. cerevisiae overexpression mutant: increased apoptosis; decreased vegetative growth

chr2-2_0367 (69% homologous)

5' region

ccacattgtagcttatgttctgatatttttcagggataggttgaaacttgagaccttgag
ttgaatttctaggtgctcgagctgttcgatttacgtggtttcttatagaaccatttggtc
tgtccgcatttggtctagggttattgtagatgtaacttgaaggtctggagccatagcgtc
tagaatgtggatcttccttctcccaagagaaatcatcccaatcactgtactcattgtacg
agggcctcatattagaggtacgttctcgtccataaccgtcatgataagaacggttatacg
a**cat**attgctaaagatgctatcaaacgttgccctcctaatagaagtctggtaaatagtaa

aactgaagaagaatgcgcgttgaacaattacataatcacgaaacttacaatcctccactt
tgagaatcctcttaactaagtatactatttctttggcaagcagtagaagagagtgcatg
catcgccgatctaagttcagaatccgtcttgttgtatccgctgccatgtccaagaacttc
ccctttgttatttagacatttagagataaattttgaaccattgaaggcagtagcgtatct
catgggacccattaacttatgaagagaatccttagcggatctgtcaacgaggttttcaac
ccgcaacttgtaaattttctctgaagcaagtttcaagatccactgagtaagtaaagtaag
taggttatagtcggaccccgtctcgattttgagagccgcaacgtacgccatgaaacactc
tgccattactttgggtcggaactgctttgctgggaacttgctttccagtagctctctgaa
atttaaacgaagcgcccaaatatggacattgttaaagtgtaaaacttcttgcacggcaca
atctaacagctcagcatctagcgaaggatgtagttcattcaatttgaaactgactaagca
atgacaaaactttgaacccaaataaaccagagagtcattcgtttgaacactggaatcatt
cgatacgagtttttatgaagaagagctttgttttcaaaggtctcattctttaacttagg
aagacggtcaatatcttggaaagtaactctataatcgtcttgaatgctgatttcatttt
atggttgtcagaatcaactgaacctgcagtgttctcatttctttcagtctcatgactatg
cacaatttcttcctcagtggtttcgactaaagatctctgaaactggtttatagcatcaat
caggccgttcactctatgttttttgtttgaatcaggttcactcacagtatcatccaaaag
agagtcatggccccgtttggaagatctctcattttgccgctgaacctctatttgggcatt
gtgtattatattctctaccagaccctctatatcttcggagttttcttgcgagttcaatag
agaagtggaaaaaagttgtctagaatgagggcacggttcattcttagctttcatagatgt
cgccccctagttaattgaaatcttaagttcttgaatgagattgtactcttaattgtgaca
gatattagacgaagtaaaccactaatgacgaatatgagataacctctgtgattcttctgg
aagaacgtagcatttgcagtgaaagagcagacctgttccattacgtaatccaaatatgga
agatactcaagtctagaattacgacctgtacttactcattccaccaaaggtgttagggct
gacaaccagtgaacctcccattcatccagtagtcaaacatatccaatgtttaccataatt
agtaattctttaaacaatctacttagttcggatatgcttcaatggccctttagagtggta
aaagttggcagagtaatccttctttacaaaccgtctaactttgggttcttgaatatagta
gtcccagccttgtcctttagcgaaacggatagctccttctctagtatcaaacttcaagat
agttccttgcatgtagtcagaggaggactggtaacccattaagtcattttcccatctgtt
tcctttattcaggatatcccagtctattttccacgtttcagtattgtgctgaccactctg
ggtggcagctttggcttcttggtaaatcctggccactctctcggtaataagttctttggg
agccccactaaccaactcctttgctaagtctgtatggccttcatttattgcgggagtaat
attggtattcttggcagctgatacaacacggccacacgagaaccctcttacttgtacttt
cagcaaacttctcaacattttgtgatggacgttacaacttcgtagaagttggagattgga
ttataaccattggctcgcgacttcaaaaaccatgatgacataagctaagattggtatgct
cgaaatgatttttttttttttcactccaaatcaagatttgcaagttactcattacaggga
ggtcttaactttggctccgcttgatatagtaactactcac

ORF

ATGACACAACATGTCTCCAAGAAAAGACGGACCGATGAACTTTCCGATTCTGATAATTTTAACACTAGTGACCAT
GAACAAGAAATTAAATTGAATAGCATTCCTTACGTTGCAAATGGCTCTTCAAACCAATCATGGCAAACTACTATA
GAAAAGGTTGTGCAATCTGTGGTTAGTATTCACTTTTGTCAGGTGGCGTCTTTTGATACTGAAGACGCAGTGGTC
AGTCAAGCAACTGGATTTGTGGTCGATAGTGTCAATGGATACATCTTAACCAATAGACATGTTGTTGGACCCGGA
CCTTTCGTAGGCTATGCTGTGTTTGACAATCACGAAGAATGTGATGTGAAGCCTATTTACAGAGATCCAGTGCAC
GATTTTGGTGTATTACAGTTTGAACCCAAGAACATAAAATATATGAAAGTATCAGAATTAACTCTCAGACCAGAC
TTAGCAAAAGTCGGGTGCGAAATTAGAGTGGTAGGAAACGACGCTGGTGAAAAGTTGTCAATTTTAAGTGGTTTT
ATTTCCAGATTAGACAGGAACGCTCCTGAATATGGATCACTTACTTATAACGATTTTAATACCGAGTATATTCAG
GCAGCTGCCAGTGCAACTGGGGGCTCATCTGGTTCCCCAGTAGTGGACATCGACGGATATGCAGTTGCTCTTCAA
GCTGGAGGATCTACGGAATCATCTACTGACTTTTTTTTTCCAGTTTATAGAGCTTTACGTGCTCTTAGATGCATC
CAAAACGGTGAGCCAATTTCAAGAGGCACCATACAAGTTCAATGGATATTGAGACCATTCGATGAATGTAGAAGA
TTGGGTATTCGTTCTGATAATGAGAAGACAATGCGAGATAAGTTTCCTTCTATTCACGGATTACTTGTAGCAGAA
GTTGTCCTTCCAGAGGGTCCGGCCGATGGGTCTTTAAGAGAGGGGGATACGTTGATCTCCATTAATGGTGAACTA
GTCTCTTCATTTGTCAAAGTAGATGAAGTACTTGATTCTTCTGTAGGACAAACAGTCGAACTAGTGGTTGATAGA
AACGGGAAGGATCTTCATTTTACTATTCCTGTCAATGACTTACATGCGATCACTCCAGCTAGGTATTTGGAAGTT
TGTGGTGCCTCCTTTAATGATCTTAGTTACCAAATGGCACGCTTGTATGCTATTCCAGTAAGAGGTGTGTTTGCT
AATAGAGCATCAGGAAGCTTCACATTAGACACGAGAGATAGGTGTGGTTGGATAATTGACTCCTTAGATCACAAA
GATACACCCGACTTAGATACCTTTATCGAGGTTCTCAAGTCTATTCCCGACTGCTCTCGAGTTCAAATTAGTTTC
AGGCATATTTCCGACCTTCATACTATTGAAGAAAGCGTTGTGTACATTGATCGACACTGGTACTCTGAATTCAGG
CTGGCAACTAGAAACGATGAGACTGGTTTGTGGGATTTTACTAATCTTCAAGAAAAACCTCTACCTCCCAAACCA
CTGAGACCCTTGCATGCTAAATTCATTGATATTCCCACAGAGAAACCGGGGTGTTCCAAACTAGGTCACAGTATG
GTACTTGTTACTGCTCACTTCCCGTTAGTGATGGATGGGTACAAAGACAATACCTCAAGAGGATATGGTGTAATT
GCTAGCGCTGAAAATGGATATGTCATCATTTCACGTAAGGTCGTACCACACGATCTCATCGAAGTATTCGTCACT
GTTGCCGAGTCGATCATTGTACCTGCAAAGGTTAAATTTTTGCATCCTTTGCACAACTATGCAATCGTTAAGTAT

GACCCAAGCTTAGTGGAGGCTGACGTTTTGACTCCCAATTTTTCTTCGAACCCTCTGAAAAGGGGTGACAAAGTC
GTCTTTGTTGGATTCAATCAAAATATGAGAGCTGTTTCAGATATCACAAAGGTCAGCGATATACCAGTACTGAAT
GTGCCTATCAATCCATTGAGTCCCCGTTATAGGGCATGTAACTTTGAAGGTATTCAAGTTGACTCAAGTGTTGCC
AATCCTGCACCCTCAGGTGTACTCGCTGATGAGGATGGAACTATACGTGCATTATGGTTAACATATCTAGGATCG
GTGACCGATGAGGGCTACGATCGTGTATTCGGAATGGGATTCGATACCAGTCATATCAATCAAATTGTGTCCCGA
TGCATAGCAAACGAAGGGCATTTAACTGACTTGCGTATTATTGATAGTGAATTTTACGCTTTACCTGTAATTCAA
GCCCGGTTGAGAGGGGTTCCAGAAGCATGGATCGAAAGAGTAGAATCACGAAAAGATGTTGAGAGACCTCAATTT
TTTACCGTTTTAAGAACCAGTACCCCAGCCATTGGAATGGAAGCTTCGCCTTTAAAAGTTGGTGATATTGTATTG
TCTCTAAATGGACATGCCGTGAACAAAATGGCTGACCTTGACGACATGTATACTCAAACTGAGCTGCAAGTTGAA
ATTCTAAGAAAAAAACAAATTATAAAGGTGACCGTTCCCACCGTTTCGACGGAAAAGTTCAACACATCCCATCTA
GTTTATTGGTCTGGTGCGTTATTACAGCCCCCTCACCAGTCCGTGCGCCAGGTTATGAAGAATTTGCCTTCCTCA
ATTTATATCATGAGCAGAAATCAAGGATCACCAGCTACCCAGTACGGTTTAAATTCTACACAATTTGTGACACAT
GTAAATGAGCAAGAGACTCCTGATCTTGAAGCGTTTATCAATGTCGTTCGTGGAATCCCAGACAACACTTATTGT
AAGTTGCGTCTAGTCTCATTTGACAACATTCCTTCAGCATTAACACTGAAAACTAATTATCACTATTTTCCAACG
TCGGTGATTCGGAAAGATGAAGCTGAGGACAAATGGATAGAAATAGATTTTACAAAGGACGGGCCAGTTCGTCTT
GAATTAAACTGA

Downstream

atacagtcataaatactaattgacttcattcattgtctctaatatatacacgtagcgact
cc**cta**tagattgggcaagagtttagctaatctagctttttgattacctgctcctctagac
gttgcctgacacgtaccaaccagctttcccaacactcggttagaccacccagcattgcca
tgttttcttccagagagacctctcttgtatccattgcctgtttttctccatctcttataa
gttgcccggtgagtcttcagcaagtttcgagtaattgtaaggaccggagaagaatgtgaa
attgtggaggcggaaagagccatccgacctgtcatcggggtcttgttaaacaatatgtga
aacatgtccaaagaggtgaaaagttgaaatggaagatattcgatacagtt**cat**tctatta
tattttctagctactgttcgcgacatataatgagactatgctactatcaacccttgctca
gaacttgtaatttactaacttacggattcaatacttcaatatgtggtgacttatgagagt
actaaccatagttaaataatacaatcactgaaaaaaaaagaaatatatgctacaaaaaac
cctgtatttgaagtacgcttctggaaaatatcaacttcaattttgtaaaatttttgaaac
gtgaactgtgacgcgtgatggtattcaaaagtggccactacatattgaatcaatcagtcg
agttacaacaagctgagtaaattattctgg**tca**tatatcttcatgggctttctcgtcctc
gtcatcatcagaactactgtaccccactatatttcctcttgaagaattgtctaactttct
tttcttaagcgtaggtttagttttggcagaagtgataaggctggagacctttaaatgttg
cattgcatttaggtttggacccatatcttttagctttcgatactcttctactcgttgccc
gagaaagcgggcctttatggattcttgctccttttgttttcgttgtagttcattgtaaaa
cgtagaatctttttcattaagcctgtaggcagaattggtatgttctagctgttgttggta
ttcttgatatttttccaacttgtgcttctccagttgctgaactaaagtctttccctctct
ttgacgttgttcttcagcatgctttgcagactcttcattttccctttcaatttgtttttg
aaatgtttctgaatctatctgccctgcctgcacaaatctatttga**cat**gtactaccatca
actactcgttcgtaacttgtgaatagtatcatatttcatgcttagtatgtcaagtatgta
atcgcgaactatgtatgtcaaactatgttaaccattattaacagtatacatatataactc
tatttacataggaggagg**tta**agcagaagcggcagtagcttcattctggtgaccttgggc
ttgagcaattatttcttcaactctcttactaagagctggatctttttcacccttgcttg

AA

MTQHVSKKRRTDELSDSDNFNTSDHEQEIKLNSIPYVANGSSNQSWQTTIEKVVQSVVSIHFCQVASFDTEDAVV
SQATGFVVDSVNGYILTNRHVVGPGPFVGYAVFDNHEECDVKPIYRDPVHDFGVLQFEPKNIKYMKVSELTLRPD
LAKVGCEIRVVGNDAGEKLSILSGFISRLDRNAPEYGSLTYNDFNTEYIQAAASATGGSSGSPVVDIDGYAVALQ
AGGSTESSTDFFFPVYRALRALRCIQNGEPISRGTIQVQWILRPFDECRRLGIRSDNEKTMRDKFPSIHGLLVAE
VVLPEGPADGSLREGDTLISINGELVSSFVKVDEVLDSSVGQTVELVVDRNGKDLHFTIPVNDLHAITPARYLEV
CGASFNDLSYQMARLYAIPVRGVFANRASGSFTLDTRDRCGWIIDSLDHKDTPDLDTFIEVLKSIPDCSRVQISF
RHISDLHTIEESVVYIDRHWYSEFRLATRNDETGLWDFTNLQEKPLPPKPLRPLHAKFIDIPTEKPGCSKLGHSM
VLVTAHFPLVMDGYKDNTSRGYGVIASAENGYVIISRKVVPHDLIEVFVTVAESIIVPAKVKFLHPLHNYAIVKY
DPSLVEADVLTPNFSSNPLKRGDKVVFVGFNQNMRAVSDITKVSDIPVLNVPINPLSPRYRACNFEGIQVDSSVA
NPAPSGVLADEDGTIRALWLTYLGSVTDEGYDRVFGMGFDTSHINQIVSRCIANEGHLTDLRIIDSEFYALPVIQ
ARLRGVPEAWIERVESRKDVERPQFFTVLRTSTPAIGMEASPLKVGDIVLSLNGHAVNKMADLDDMYTQTELQVE
ILRKKQIIKVTVPTVSTEKFNTSHLVYWSGALLQPPHQSVRQVMKNLPSSIYIMSRNQGSPATQYGLNSTQFVTH
VNEQETPDLEAFINVVRGIPDNTYCKLRLVSFDNIPSALTLKTNYHYFPTSVIRKDEAEDKWIEIDFTKDGPVRL
ELN

8.2 Serine-type endopeptidase inhibitor

SEQ0180

Homologue of **TFS1** (YLR178C): Carboxypeptidase Y inhibitor, function requires acetylation by the NatB N-terminal acetyltransferase; phosphatidylethanolamine-binding protein involved in protein kinase A signaling pathway

S. cerevisiae null mutant: viable; decreased resistance to amiodarone

chr3_0640

5' region

tggtacgtttccctttctcttcgtctcgaggcacaatatcaatcagaaaatcgaacatat
cagatttctttaaggccgcggcaatatcagatcgttgtaaagttcgccttttgttctctt
ctgcgtgtatccacgctctcattgttagctctgtgatgaaaatatcacatcccttggcaa
acaaaataggggcttcggcactaatcattttcacctcttcatcagttttcattactttct
tgattctggccaaaggtagctgatggttcttaaaatcatgatcgtcgtgttcaatagagt
tgatggtttcctgccaatattgcatcatcatgttcttggatcggcctgtcagcccttgac
ctacattctggaaggcactgccggctccgaactcatcgtcttcatcctcttcttcttctt
gtccttctgattggtcatcctcggcagcgttttcctcgttctgctcgtactcttcgtagt
actcctgctcttg**cat**tgtattctattaatcaacgagtatagagtctccctgtggggatt
gaaggatggaagagtttgatagggaggttcgcgacggctattagggaaggtgtgccaagg
ctaaagcttcaggacacgcagaaaaggaaagacccgtccgttttaaatcctccgcctacc
gcccacgaaactgcggactcaacgggcaaaaaaaaggtaatctgtcttcttgttcccctt
tcatcaacaccttggactgcaaaggtcctaacaacggctgatagctttttcgtacacac
atcccattaaac**atg**cccagtccatttcagtcatttcaatcaaacctgtctccttctaag
catacgcagtcctttcaagagctctacggcgaaccagagaactttcttgaaattgaagta
ataaaccccataacacacggatcaggttctagcatgtacacggactatgaaatcgtttgc
agagtaagtacaaaaactcaattgccataaagaacgaattactaacaatacagaccaaca
taccgatgtttaaatttaaggaatcaagggtgcggagaaaatactcagattttgactcgt
tcagaaaggtactagagtcccagacaaataacgtcgtgatccctaaacttcctgagaaat
cgttcttcaactaccatcgcttcaatgatgacttcattgaggagcgcagacagggtttac
aacagttcttaaaggtgattgccggccatccacttcttcaaacgggctctaaagctttaa
cttcatttgtgcaggatgaacattggaataaatctaagttctta**taa**tcaacggttaaag
ggagaataaaaggggtgagagcgtcatttttctagaacatggtgaggcataagtacctcg
atcttatggggaaacacataatctcgaactaataattaatttgcaaacttgcttttgctg
tttcccttctattttctcttcccacttttttcttttgagtcttcaccagctcaatcgat
tg

ORF

ATGTTATTGGGATTTCAAAGATTCCCAAAATTCCATACACTTACTAGAAAGAGTCCAATCAATTCCATCAGATCC
CTCCGCTACAAGTTTCAACGACCTTTGCTTACCCCCCTATTTCTTTCTCCAATTATTAAACAGAGCATGAACCTT
GTCACCATTTCAGACTCAATCAGAGAAAACCTCATCAAACACGAGGTCATTCCGGATGTTATCAAGGACAAATCT
TTTGTTCCTAAAGGTCTGTTGATCATTTCCTATGGAAGTCCAGATAAGGAAGTCGTATTGGGCAACACTCTTAAA
GTTGAGGATGCTCAATCAATTCCTAAGATAACTTTCACCGTTAATCTCAATGAGGAACAGGAGATTGCTAGTTTC
TTTGACGAAAAGTTCACTCTTGTTGTTACGGACCCTGACGCTCCATCTCGTACTGATAACAAGTGGTCCGAATTC
TGTCATTATGTCGTCTCCGACTTGTCATTGAGTACAAAGTCCGGAACTACTGACGCTGAGGATCAGGTGAATTTT
ACTACCGACTTGAAGGTTGATACCATTCCAGAATCGGATTCTAAAACACTGGTTCCCTACCTCGGACCAGGTCCG
CCTCCAAAGACAGGACTGCATCGTTATGTTTTCCTCTTATATAAGCAGAAACCTGGTGTTTCTCTAGAAGGCCCT
GATCCAAAGAACAGACCCAATTGGGGTACAGGAATTCCGGGCTCTGGTGTCTCAGACTGGGCTGCCAAAAATTCC
TTGAGTCTTCTTGCTGTGAACTTTTTCTTTGCCCAGAACAAGGACAATTAG

Downstream

tacttaagtagtttttattctctccttaaaatttcgttatgtaagcattttctttcaac
ctcgaaatctgccttttcttcgttggtgaatcctctttggcataggaacggaacaacttg
gattagattcgatacccgccttcctctctttctttcgcctccctaaaaactaaagcact**a**
tggagactcagattagcagtctgatttctgctagagaggttgcattccagaaaccgcaac
tgtatacagacatctttcacaaccttcttccgttcatctacaatcaaaatactactttac
aaaccttggttattgactttctcaaagaatcttttgtggatgaaaaattggataaagagc
ttcagcttcagctggctgatgaactgcttcaacctttgagttacctaatttcctcaaagt
ctcatctagaattgccaagggccccttcacctgctaattttgtcaagattattcaattga
ccaaccatatatatccgatcattttcaaaaagtgtctgtccagcccccagatgcaggact
tggagtccaaatgggaccaattacagagtttgagggacaatttgattcaaaaatttggta
ctgcatatcctttactcccactgaatcctgaatcggacgccagtcgctcaaccagatcta
acataacgctagttatgttcatcacaaatgtgttgtcaatacttagtattgctccaccaa
aggatccaagagcagctagggctagaaaattgaagaacacaccggaggatctatctatag

tgagtatcccaagtgatcacaaatttcttaacatgactcgtctggaggttgaaggaagac
aaattttagacgtcctctttgatgctttaaatgattccaggataataaactccccgttat
tttgtgtcattacaaaccgtctgttaacattgttcaagaaaagaccaggtttagtaagcg
tcaaactgtttgatttcatgatagcctatgaatcaatgaataaggaagacccactctttg
agtcgaaccagttgaaaatgcgccttattcgaaggttcaacgatcgtattctgaaagtaa
tgataggatattgtctaaataagagttttgtgaaagatacgaagttactcaaaaggttcc
agaacaaattcacttactaccagggcatgtttgaagatcaacgtaagaggggtctgttga
aaccagatagtcatgaaattaagagagaagaaaagagggctgaatttaagaagaaaaaaa
ttaccggaaatgaaagtttggaagagcaacttcagaacaacccttccgaatcacccctttt
tttactacaattcaagcccaataacgcctgatttgtcttactcatcattatactcattaa
tatcgccaaacaataaactggctagcttcaatatgagtgaactaccacagaacattttga
ttgacatgattgcatattctttacaaaatacaactactagaaaactagtaaaaggactag

AA

MLLGFQRFPKFHTLTRKSPINSIRSLRYKFQRPLLTPLFLSPIIKQSMNLVTISDSIRENLIKHEVIPDVIKDKS
FVPKGLLIISYGSPDKEVVLGNTLKVEDAQSIPKITFTVNLNEEQEIASFFDEKFTLVVTDPDAPSRTDNKWSEF
CHYVVSDLSLSTKSGTTDAEDQVNFTTDLKVDTIPESDSKTLVPYLGPGPPPKTGLHRYVFLLYKQKPGVSLEGP
DPKNRPNWGTGIPGSGVSDWAAKNSLSLLAVNFFFAQNKDN

8.3 Aspartic-type endopeptidases

Homologue of **YPS2** of *P. pastoris*: aspartic protease yapsin 2
S. cerevisiae null mutant: viable

SEQ0181
chr3_1157
5' region
agccggggatgcgatgcacttttttcaccttataaccatatcgtttagcttcaggttgt
ttctattcccaccgtttccggccactatttagctagtagaaagatcagaa**atg**tccgtgt
ttacgccaatagaggacgccctgaaggcttacagtatgttactcttcctgctacttgaaa
acgaatcatatgattactaacgcatctagaaaatggagaatttctcatcgtgatggacga
tgaagacagagaaaacgaaggcgacctgattgttgccgcagaattgataacgacacagaa
aatggccttcctcgtgaggtattcctctggatatgtttgcgttccactttccaccaagag
agctgcagagttagaactaccaccgatgatcaagaacagaacagacaggcagggcactgc
gtacacacacactgtggacgccagagatggcactaccactgggatttccgcctctgacag
agcgttgacctgccaggttttggccaaccccaggtcaaaacccgaagagttgatgagacc
gggacatatttgccctctgattgcacgagaaggcctgttgaaagagagaagagggcacac
ggaagctgctgttcaattgtgtgaattgaccggattacaacctgctggagtgataggaga
gctggttcgtgacgaggacggctctatgatgcgattagacgactgtgttcagtttggtct
ccgccacaacgtaaaaattatcaaccttgaccagatcattgaatacatggattccaagaa
cagc**tag**atacgatggataggaatacagagatatcatgattgaggaacgtaagagctttt
tcgaaagtgtgagtttgtggtgagggccaggcggtggggaggtggtggggagcctccttg
gtcgaatgtagatatagtaagcaagacacaagagcgcgcgaagtcttcaacgaggcggcg
ttgggtcttgtacgcaacgtaatgactacacagttgagcttgtcgcgaaccggtcgacat
tttgatcatgcatactatgttgagacaccatctcgtactattgcggcaaccagctgtaaa
tttgactaattaaagctgatgaaggatgcagggcgtcgtcaatttttgattgattgcat
ttaattgtttgagccattcaaggctgaatgcccggcaccctagacccttcttgtgagtac
tataaacccgcaggcagggtacccttggccttctgcgagactaccagtcataacgtatat
ccacaatgtactagtaatagccccggaaaactctaatcccacagaacgtctaacgcctcc
tatgtcatcgatacccattcgcactactgccatggccccccttacgtgatcatttcactt
actcccgcctaagcttcgcccacatgcctgcgttttgccaagatttactgacgagtttgg
tttactcatcctctatttataactactagactttcaccattcttcaccaccctcgtgcca

ORF
atgatcatcaaccacttggtattgacagccctcagcattgcactagcaagtgcgcaactc
caatcgcctttcaaggctaacaagttgccattcaaaaaagtttatcattccaacgacccaa
aggaccgtttaattaagagagatgactacgagtccctcgacttgagacacatcggagtct
tgtacactgcagagatccaaattggatctgacgaaactgaaattgaggtcattgtcgaca
ctggttctgccgacttgtgggtcatcgattccgacgctgccgtctgtgagttatcctacg
atgagattgaggccaatagcttttcctcggcttctgccaaattcatggacaagatagctc

ctccatcacaagagctcctggatgggctgagtgagtttggatttgctctcgatggtgaaa
tttctcaatacctagccgataaatctggacgtgtttcgaaaagagaggaaaatcaacaag
atttcaacattaaccgtgacgagcctgtgtgtgaacagtttggttccttcgattctagtt
cttccgacactttccaaagcaacaattcagcttttggtattgcttaccttgatggaacca
ctgctaacggaacttgggtcagggacacagtccgcatcggcgactttgccatcagccaac
agagttttgccttagtcaacatcacagataactacatgggaatcttgggtctcggtcctg
ctacccaacaaaccaccaatagtaacccaattgcagcaaacagatttacttatgatggtg
ttgtggattcattgcggtcccaaggatttatcaattcagcatcgttttctgtttacttgt
ctccagatgaagataacgagcacgacgaattcagcgacggagaaattttatttggtgcta
ttgatagggccaagatagacgggccatttagacttttcccatatgtcaatccttacaaac
cagtttaccccgatcaatatacttcctacgttacagtgtccacaattgcggtgtcttcgt
cagatgaaactctcattattgaaagacgtcctcgtttggcattaatcgatacaggtgcca
ccttctcctatttgccaacctacccattgattcgtttagcgttttccatccatggaggct
ttgaatatgtttctcaattgggactatttgtcattcgtacaagttctctgtctgttgcta
gaaataaggtgattgagttcaagtttggtgaagacgttgtgatccaatccccagtttctg
atcatctattggacgtctcaggccttttactgatggccaacaatactccgcattaactg
tacgtgaaagtcttgacggactttccattctaggtgatacattcatcaaatcggcctact
tattctttgacaatgaaaacagccagctgggtattggtcagatcaacgtcactgatgacg
aggatattgaggtggtcggtgatttcactattgaacgagacccagcctactcctctactt
ggtctagcgatttacctcatgaaacacccactagggctttgagtactgcttcaggggggag
gccttggtaccggaataaacacggccacaagtcgtgcaagttctcgttccacatctggct
ctacttcacgaacttcttctacatctggctctgcttctggtacttcttcaggtgcatctt
ctgctactcaaaatgacgaaacatccactgatcttggagctccagctgcatctttaagtg
caacgccatgtctttttgccatcttgctgctcatgttgtag

Downstream

tagacttttttttcactgagttttatgtactactgattacattgtgtaggtgtaatga
tgtgcactataatactaatatagtcaaaatgctacagaggaaagtgcaggttgcctgtgg
tggtttttcttattagcaccctctgaacactctttacctctaacatcctcagcc**atg**cta
atcgcgcataaaataaatcttcgaacttttttccattttatgctcataaagcttccttac
tgtcaccttatcaaaagagcttttgccactaaagtagtcacacccagaattgctcccgaa
tatcgtccaacaatgctaggatctgtggaaagtttgacaaataatttgaacaccttgagc
ttgaagcttcctgaagttaatatccaaggctcctttccagaaagtaacccagtggacctt
ttgagaaactacatcactcaagaacttagtaaaatttctggagttgacaaagaattgatt
ttcccagccttggaatggggtaccacactggaaaaaggtgatcttttgatcccagttcct
cgtctgagaataaagggtgctaatcctaaagatttagccgaacaatgggctgctgcattc
ccaaagggtggatatcttaaagacgttattgcgcaaggacctttcttgcagttctttttt
aacacatcggttctgtacaagttggtgatatctgatgctctggagagaggcgatgacttt
ggtgcacttcctctaggaaagggacaaaaagttatagtggagttttcttctccaaatatt
gccaaacctttccacgctggccatcttagaagtacaatcatcggtggttttatttccaat
ctgtatgaaaagctgggtcatgaagttatgaggatgaattatttgggagactggggaaaa
caatttggtgttcttgcagtaggatttgagcgttacggtgatgaggcaaaattaaagact
gatccaatcaaccatttgtttgaggtctatgttaaaatcaaccaagatattaaggctcaa
tcagagtctactgaggagattgcagaagggcaatcattagatgaccaggcaagagctttt
ttcaagaaaatggaaaatggcgacgaatcggctgtaagcttgtggaaaagattccgtgag
ttatccattgagaagtacattgatacttatgcccgcctcaacatcaaatatgatctttat
tctggtgagtcacaagtcccccaggaaattatggacgaagccaccaagatctttgaagaa
aaaggcctgcttagcgaagatcagggtgctaaactgatcgatctaacaaagttcaacaag
aaattaggaaaagctatcgttcagaagaaggatggtacttcattatatttaactcgtgat
gttggtgctgccattgatcgttacagaaagtataaattcgataaaatgatctacgtgatt
gccagtcaacaagatctgcacactgctcagttctttgaaattctgaaacggatggatttt

AA

MIINHLVLTALSIALASAQLQSPFKANKLPFKKVYHSNDPKDRLIKRDDYESLDLRHIGV
LYTAEIQIGSDETEIEVIVDTGSADLWVIDSDAAVCELSYDEIEANSFSSASAKFMDKIA
PPSQELLDGLSEFGFALDGEISQYLADKSGRVSKREENQQDFNINRDEPVCEQFGSFDSS
SSDTFQSNNSAFGIAYLDGTTANGTWVRDTVRIGDFAISQQSFALVNITDNYMGILGLGP
ATQQTTNSNPIAANRFTYDGVVDSLRSQGFINSASFSVYLSPDEDNEHDEFSDGEILFGA
IDRAKIDGPFRLFPYVNPYKPVYPDQYTSYVTVSTIAVSSSDETLIIERRPRLALIDTGA
TFSYLPTYPLIRLAFSIHGGFEYVSQLGLFVIRTSSLSVARNKVIEFKFGEDVVIQSPVS
DHLLDVSGLFTDGQQYSALTVRESLDGLSILGDTFIKSAYLFFDNENSQLGIGQINVTDD
EDIEVVGDFTIERDPAYSSTWSSDLPHETPTRALSTASGGGLGTGINTATSRASSRSTSG
STSRTSSTSGSASGTSSGASSATQNDETSTDLGAPAASLSATPCLFAILLLML*

Homologue of **YPS3** (YLR121C): Aspartic protease, attached to the plasma membrane via a glycosylphosphatidylinositol (GPI) anchor
S. cerevisiae null mutant: viable

2 homologs

SEQ0182
chr3_0303
5' region
accgcttgatcaaaagaaagaatggtattttcagggaacaatttcctatagcctctatga
gctccaacagtgatagccattcttcggtgcagtaggaagcgatgagatgagatgaggttt
tttgacgg**cat**tgcgcgaaaaaaagaaagccgggaggagagcgaatccatggagggacta
ggcccaggagactttcttctgctgacgaacagcttgaagaatatgtgagagatagggccc
ttcttgcacccaaggcggattaggcggacattggtcacatcctcagctgaaatcaacaag
aataaagtcaaagctctatgttctgcgttacaatggtcaatcaagtgattgtagattcct
tctagctcttgttattctctgttcccgcaggtagctatgactctggccagaaaactgata
gacaaacttaacctgactcatatagtctagggtaaaacctcgaagatgacttcattcaac
ctctatgattgtctagatgttatcagtccggagtatccaagcattcgtgaatttatcatt
gatcgcccagtcaaaatcattataactacagctgaaacgctgcgtgtttaaatgatcgtc
gatcgttggtgtgtgacgcgtgcaagctcgaagtgcttgcgcttgacaacttcaataagg
gctgggacctttgaagcttgaccacggtgatcaatccacaactcaaaagcagatctgaat
cctgtaccaagatataaactcatatactctccgtaaaactatgcgtttattccattttga
gtcgtgtgatattgggccttccgaatttgctggctcaatttcttcgtatacgggacaagt
aacctaaaaaagtgacaacgctttgaaatctcaaataagtcaaagcacctttttcggggc
gaatgcacactttcggtccgatgagaagagaactagtgggctgcaaggacgtctgcattg
cctctgatcgggctaaacaggtggttaggctttcctatgcagttcgccggcaaataaccg
ctaatcgtgcaaagggcttcaaggacactgttgtttcatttaacccaaaagtgcattatg
tctaactctgaacaataaatcctgctcactcaatagatgtatttggattcagtgcgggaa
tggacgtttctctaacttcaacaataagctcactcactcttttatccttcaggcaacata
ttgcagcattgtggcaattttccgtctgttgaggtcttggttacccaaatttttccttcg
aattgaacccttaaggcctatttgctcaagaaaaattcactctcacttcgtttcagtgaa
gaaaaaaaaagctgagaaagcttgcctcaattggtcttggatgcacaacagatgtgcata
tagaagcttaagcaataaacgcatgaccccacccaaattaactccagcaattgggctagg
tctgcaaattgtaaataaagccggcctcctcttcttctatctgttgacagggtcgtcaag
ORF
ATGTTGCCCATCCGCTTATCCAAACTTCTGCTTTTGCTCTCCTTAAAGTTGAAATTGGGTACAGCTGAAGAAAAA
TACCAAAAGTTGGATTTAAAAAGAATTGACAAAGACTATTATGCCGTCGATGTCAAAGTCGGCTCCGATGAGCAG
GAGATCAAAGAGGTACTAATAGATACGGGTTCATCTGATTTCTGGATCTTGGACAAATCGTTCTGTAATTCTCCA
ACATCAGAGGAAGAAGAGAACAGTAACGGGCGTAGCAACAAGGAAAGCTGTGGAGTCTATGGCTCGTTCGACTCC
AACAAGTCAGAGACATTTCAGGCAACTGGCCAAGTATTTGACGCTGCTTACGGTGACACCACAGCCGAGTCGACA
GGATCTTCAGGAGTTCGAGGAATTGATCAGCTACGGGTAGGAGATATTCATATAGAAGAACTCTATTTTGGACTA
GTGACAAACACTACAAGTTTACCACCCGTTTTAGGAATTGCCCAGCTTTCCGAAGAGTTCAGCAACAACTCTTAT
CCTAACTTTCCATACCAGATGAAAGAGGAAGGTCTGATTGATGTTGTTGCATACTCTCTCTCCTTGGGCCAAAGT
AAAGGTGAACTACTGTTCGGGGCTATGGACCACTCAAAATATAATGGAACACTATTGAAAGCCCCTATATTGCAG
GCGGGCACACCAGGAATGCAAGTTCTTTTAACTGGAGTGGCCCTTACAAATGGTTCATCAAGCGTCTTCAATGAG
ACAGACAATAAAGGTTTTATCTACTTTGACAGTGGGACTACTGCTTCCACTCTGCCATCAGAGCACTTTGATGAT
CTTTTCAACCATCACGGATGGGCGTACGATGGTGATACATTGACATATTCGATTCAATGCGATAGTGAGGGAGAA
AAATCTTTACTTGACTTCACTTTAGAATATACCATTGCTGGTAATATTGTCATCAAAGTACCATTTGAAGACATT
ATTATGAAGAATGAAAATGATGGAGAATGCCTCTCAACCGTAATGGTGTCGAACCAGACTTCTTTTTCATATTCC
GATGACACACCCTTTTTCGTTGCTGGAGACGAAGTTCTGTTGAACGCTTATGTTGTTTACAACCTAGAAACACAA
GAGCTGGCCATTGCTCCAGCAGTGGATAATCCAGAAGATACTGAAGAAGATATTGAGATTATCTCCGCAGACTTT
GATATTTCAGAAGCCAGAGATTATAGCGTTGGATTAGAGTTCAGAAATACCACAATTCCAGCTACAACTGATTAC
TTGCCTTCCTCGATGTCGTCAGGTTCAGTCAGCGAAGAGACTGGTTCCAAGTCTGAGAGCTCTACTTCTGAGGAC
TTTGCTGCAGCCACGTTGAAACCATTTACATTTTGGGGTTTCGTCCTTTTTTTCTTTCACTTTTTGATTTGA
Downstream
gaacatagcagaagctaaggtacttttaaatctacatacctactaatgtaattgagatat
attgtagtgtaagcattattctagagaagttttcaaatctattatcatgagcacatacta
atacggaattaaatagaagctatcggttgcggccacatctagcagaaagcaccgtttccc
gtccgatcaaccgtagttaagctgctaagagcctgaccgagtagtgtagtgggagaccat

acgcgaaactcaggtgctgcaatctttttttgtctctcctagctgtaaagtaactcttgc
ccatacgcttttgactggtgcagtgtagtaatcgaataggggcctgacattcaatggttc
acctttcattcgttttgagttttctggtttccagatcatgtcaattgtagttaaacatat
attttactgtgaaggctttggaaagcctcaatcaagaagtccccaccacttagttggtcc
tgcggtttcttctaggtatttcattccacactacgacaatagagagtggcccaatactaa
ttagcaacataccgagagagctcggagtgacttgccagttttagaaagattcatccacga
tgttatcttcatccacgttctgaacccttatatgggtcatgtctagagtttatcactcct
ttatcgtgttgagccaagcgcaccgctataagcttccgtcagttaccgaactgataacca
tcaaggaccaatatggtggatcatcacacaacagttattaagcgacacccgcgatgttaa
tgtgaagcgaggctttcagggtctatgacgtagtacaggggttaaactatgtggtcctta
gctgacctgcatgcactttgactatcatcgaccttataatgcctttattatccgaaggat
tgaggacggttagaacagctattacgtcactcaaaagacaagtgataattttcagttaaa
gaactaattctgataagccttttgctgcagtatgatgtagtaagtaagtacgttccgct
gttatccgctcacacatcctgttagactgaaagagtcacacatttctccttggatctacc
tgactgtcagtgtctgctcagccttgttgaggacaatatattaatcttatttgaaaccgg
ttgaactttactcatcaatctatctaca**atg**gctcattttaggataaacatgcttcaaaa
acgatccgtccgtcgtcaaattgttatcgttggtgctggcagttatgggatcgctctagc
taaagaattggcaccgctggtagaatctatggccgatgttgtgctggtttcagactctga
acaagttatctttctaccatcattggttcgattggatgctcttaaaaacgttgctaatct
attcgttggtctcgagggtctttttgctaaaaccaaggtaaagctggtcgtggaccatgt
tgtcaacatcacgaaaaggggagtgcaattggagcatacaggattactccaatatgatta

AA

MLPIRLSKLLLLLSLKLKLGTAEEKYQKLDLKRIDKDYYAVDVKVGSDEQEIKEVLIDTGSSDFWILDKSFCNSP
TSEEEENSNGRSNKESCGVYGSFDSNKSETFQATGQVFDAAYGDTTAESTGSSGVRGIDQLRVGDIHIEELYFGL
VTNTTSLPPVLGIAQLSEEFSNNSYPNFPYQMKEEGLIDVVAYSLSLGQSKGELLFGAMDHSKYNGTLLKAPILQ
AGTPGMQVLLTGVALTNGSSSVFNETDNKGFIYFDSGTTASTLPSEHFDDLFNHHGWAYDGDTLTYSIQCDSEGE
KSLLDFTLEYTIAGNIVIKVPFEDIIMKNENDGECLSTVMVSNQTSFSYSDDTPFFVAGDEVLLNAYVVYNLETQ
ELAIAPAVDNPEDTEEDIEIISADFDISEARDYSVGLEFRNTTIPATTDYLPSSMSSGSVSEETGSKSESSTSED
FAAATLKPFTFWGFVLFFFHFLI

SEQ0183

chr3_0866

5' region

taacctcgtttgacttaaagcctatataacttacatagtt**cta**gtttaaccccaaatact
ttgcatggtgagcagcatgaacgtctgtgaatgaagagataaagtagtaagaatggtcat
agccatctaccaagttcaagtccacttttcccttgaacacagagttctctgaagccttca
agaagttctcaggtagcagctggtggtccttgaagtagaacgaatcactctttccaacgt
gaatcaaaatgctggaatctgaggggccttggtattttccaatcaattcggttgggtcgt
actgagtccaagtggacttgtcctgtcccaggtatccagagaagcacttctcaccccatg
gggctttagtggggttggagattggagaaaatgcggaaaccgacttgaattttcccgggt
ttctcaagaataactgtaaagctccgtaacctcccatggagtggcccgtgattgaaatat
tgtcaaagtttagaattgggaagtcagcctgcaatttgggtagcaattccgagttaacat
aactgtacattctataattatccttccatttctcagtagtggcatccacgtagaacccgg
caccggatccaaaatcataagagtcgtgctctccttcgatgttgagccctctgggtgaag
tatccgggaaaaccacagcaaaaccgtacttatttgcatatggttgccaaaatgccttct
ctgaggcattgttgggagtgcaagtcagaccactcaaataaagtagaactggtaatgatt
ttcccttggctccattggcaaagaattgagctggaaggtagacgttcacatccatgtcag
tcttcgtctcatcggacttgtgttgaagtttgactagctttcccccaaaactttggattt
cagctgttaccttgaagattgaagtagtaattgatga**cat**tagaaagatggaatgaacgc
tacagcgaggaagagcaattgctatatataaccccagagtattgcttaggctaagttgtc
aagattgggtaaacaattcgagaatcaaagattaagtaagaagtgttatgttgtaggctt
gaagcttcacgcaaggaagactgggagaagtagttgattcctgtggtgattgtcctgaaa
tgactgtgagcgggtccgcctagagtgttggatccctttcagtgtttccactaatggggt
ctacattgcgtcccagatcgatagaccaggagatcaattaaaccctacgcaaactatagg
gcatgaatcaagagtatcaattggcatactcggatcattttctaccgtttcataccaaga
gaagatggcaaaatgggacttctgcgttatcatgagttggctccatcggaattttccaca
cttcaggtttttcagatgtattacccttttgagccattcacaatgatatcttggtatttc
aaacagttgtagcttttactataaaagtcacaccgatcacaacgtttcatcaatgaagat

ORF

ATGTTAGTTGCTGTTGCCCTAGTGTTGTTACTGTCTACAGGCTATGCTGGAATCGTCGCCATTGATACCGAATAT
GAGTTCACCATTGGTTTTCTTAGTACGATAGAAATAGGGTTTCCCCCACAAAGCATAACGGCTCAATGGGATACA
GGATCGTCTGACCTCTTGGTCAATTCCGTGACAAATTCACAGTGTGCTCAGGACGGATGTAGCTTTGGTGCGTTC
GCCTTCAACAAATCCACCACTTATTCCAATATAACAAACCCTAACAACCTTCATGTTCAGTTCTCCTTTGCAAGC
GGCAGCGTGGTTGATGACAAACTTGTGAGTGACACTATTTTTGTAGATTCCAAGGTAATCCCACGGTTCAACTTT
GCACTGGTATCGAAGGGAGACCTGTATGGTGATAATATTTTTGGTATTGGACCGAGAGGGAACCAGGGAACATTC
GATTCCAATGGAACTCCAGCTTTCTATGATAGCTTTCCTTATCACTTGAAGGCCCTCGGTTTAATCAAACGACTG
GCTTACTCATTTTACACTGGGCCCACCCAGGGAAAGGTAGTATTTGGAGGGGTGGATCATGGAAAGTACGATGGG
TGCCTGGAGAAACTCGAGATTGTCCATGACAGTGCTTTTTACACACTGCTTGAGGCAATTGATGCTGATGATACT
TCCGTCTTGGATGAGCAAATTCATGTTTTGTTTGATACTGGTACCGCCTTGACACTTTTTCCCAGCTTTATTGCT
GAACAACTGGCTGATTTTTTGAAAGCTACATATTCGGACGAATACAATACGTTTGTAGTTCCCTGCGACCAAGAT
TTTGATTTTGAATACCTTCATTTTGGTTTTCGAAACATTAAGTTGTCGGTGCGCTTTAAGGATCTGTTTTTAGTC
ATTGACGATAGTGTTTGTGCTGTGGGGTTTGATCAAGGGGCAGATGCAAACAAGATAACCTTTGGGTCTTCACTT
TTAAGAAACTACTACACGCTTTATGATCTAGATTCCAAAGAAATTTTGATTGCTGACGTCAAGCCTGATGGTCCA
GACGATATTGAAATATTATCGGGTCCAGTTCAACGAATTTGTGATGAAAAGGGTGTCAGTAGCACTTCATTATGG
AGTAGTCTGAGTATAGAGTCCACGATAGAACCAGACACTTTTACCACTAAGCCTTCTATTTCCCAGACACGGTAT
TCGACTAGCTCCATTGGACCTCAAAACATTTCTAACTCTTTAGGTGAATATCCTTCAGTTTCCGTCACTCTTTCT
GAACACCATAACACTACTTCCATAGCCTCAAATTCCTCATTAGAAGGGAAACCAGCAACTCCAACTGTTACAGAC
CAGTCGTACCAGAATAATAAGACTACCTCTACCGTAATTGCTGTGAATTTGATTACCCATTCAACCACTCATTCA
ACCACTCATTCACCCACCTATTCAACCACTCATTCTAGTAATGGATCACGCTCAACTTTAGAGTACACTTCAACC
AAGGAATCCTCGGTGAAAATGCCCTGTGCGTTGATCATCTCCGACACAATTCCGTACAATGCTTCCGGTGGGAAT
AGTAGTTATGGATCGTTAATTTCAACATCTACGGTTAACAATGTTGAAGAGAATAATTCAAACACTGTTAGACCA
AGAAAAAGACAGACCTTCGTTTCGGGAACCACTTCCACGATACTACTCTATTCCTCAACTACGACCCAAGCATAT
CAGATGTTGTCCTCAACTTCAATCCCCCGACCATCCATAAAAGCCAGTTCAAATGCTGGTAGCCGCAAAACTTCA
AAGACATTATTAACATTTATCATATTGTATATTTTTTAG

Downstream

ttcgtaaatcaataaatcaaaacgcattgttcattctatatatcaacaaagattaaatca
ctctaatt**cta**agcagagccggaatacgtccccgcaacaagggtctttaactcatggact
ggtttccattctccactagagacattatcggaagaagaaaacggatcattggtaacctct
tggtattgcaaagtaaatcctgaatcaaccaacctggttccaaaatcatcattaatcaca
aacttcaccaaacatccagtgtctgcctcttgagccaatcctgaagtcataaatcttgct
atcagtcgctcttcgttagtagctgtcagcgttatagggttggggaaagtccaagcaatt
ctgttcttttccttattgaaagagccttgtggtttagtggctgctgaagttgcttgcgca
cccttgattgaaacggataacgagaagttactaagggtgagccctttgaagtcttctggc
aacgtggccaacaattctggggagagtctgactgagattatgacacttgcctgatgttcc
tcaaacttccacgtaggaataacaacaattggagcagcaacgtcattcagcagatacttc
acagcacctaaagtctggaattgaatctgttgcacatcggttagtgtaaacgagttctct
gcatgttgtctcagaatgtcattgtttagtaagtatttgtcagcctgctttttgctattt
tgtaacgtaaagttgatattcgaaggcaaggtattatttgcaggatccgttatatatgag
aacgcaacttcaccaactatactggaacgagtcttaacaccatctttgaatgaagcgttg
attacttcagcaatacttgcattcaatccaggtgcccttagcgaaggatgttggaaagga
gaacctgcagaagtacctgtcaagttaggtttaatactggaagaacctgtgttttgtaca
gcaattccagcgccaccaatacgtacactgctaacatcttcgttgtttgttgccacggta
gacgcaaaacttggttgtctagcctgtggaaggtcatggaacagatgggattgcactctt
tggttatttctaggcacaggaggtggtggaacagcaccatgacttgaacgaggaggattc
aacgcagattcgctggtataagtgtgcctcgatgttggaggagctggtggttgagcaact
ccatacgacagctgagcggcattttctttctgttgttgttgttgctcttgttgtggtggt
tgtgcattctgttgttgttttttggttgtaagggttcgaaagaataaagagaagaagaggtt
tttcatactgtggcgatgattgttggagtttgataggggaaccaggcgaagaagtctcca
cagaagacacataattatcttgagcagtctttggtatctgtggttgaatgcccagctgct
taatttcggatttcggagcagcaacattattctccaaagaattttgaatactggaaggcc

AA

MLVAVALVLLLSTGYAGIVAIDTEYEFTIGFLSTIEIGFPPQSITAQWDTGSSDLLVNSVTNSQCAQDGCSFGAF
AFNKSTTYSNITNPNNLHVQFSFASGSVVDDKLVSDTIFVDSKVIPRFNFALVSKGDLYGDNIFGIGPRGNQGTF
DSNGTPAFYDSFPYHLKALGLIKRLAYSFYTGPTQGKVVFGGVDHGKYDGCLEKLEIVHDSAFYTLLEAIDADDT
SVLDEQIHVLFDTGTALTLFPSFIAEQLADFLKATYSDEYNTFVVPCDQDFDFEYLHFGFRNIKLSVRFKDLFLV
IDDSVCAVGFDQGADANKITFGSSLLRNYYTLYDLDSKEILIADVKPDGPDDIEILSGPVQRICDEKGVSSTSLW
SSLSIESTIEPDTFTTKPSISQTRYSTSSIGPQNISNSLGEYPSVSVTLSEHHNTTSIASNSSLEGKPATPTVTD
QSYQNNKTTSTVIAVNLITHSTTHSTTHSPTYSTTHSSNGSRSTLEYTSTKESSVKMPCALIISDTIPYNASGGN
SSYGSLISTSTVNNVEENNSNTVRPRKRQTFVSGTTSTILLYSSTTTQAYQMLSSTSIPRPSIKASSNAGSRKTS
KTLLTFIILYIF

SEQ0184

Homologue of **YPS7** (YDR349C): Putative GPI-anchored aspartic protease, located in the cytoplasm and endoplasmic reticulum

S. cerevisiae null mutant: viable; abnormal budding pattern; abnormal cell shape; decreased cell size; decreased resistance to hygromycin B

chr3_0394

5' region

caaaaccttccaagctcaagtaattcgaaggatggtagtcaggatcacagttctttggcc
gaatcgtccaaggagcttaacattcataaaaggaacagaagaaagcgtgtctctagtgta
gtgacttcagttaccaatggagaaacctactaccattgctccgaatgtccttctaaattt
aagtttcgtggatacttaacgcggcattccaagaagcattcacaacgtaaggcttatcat
tgtcctttctatgaatcaggccatggcaaatgcagtttaggtgagggtttttccagacgt
gatacctacaaggtacatttgaaagctatgcacttcaagtaccctaaaggtgtgaaatgc
gctgatcgaactggtatgatggggtggtgtggatcatgtggtcagagctttcagaacaat
gaaatctgggttgagaggcatatcgagaagggtgtctgtcgaggtcttccagaaggttac
agtagtcgtatcatgacacgaaatcgaaagaaaacaggaaagcagtcatctctattgaat
gtcgctaacaacggtgtcaacgaaaatccaatgctcaaactagttatggatggagaagta
gtttcagcccccaagttggaggatgtgacacatctacttggtgagtctggctccacgaac
acatggactgaaaactcagaaaatgcaagccgaacttcttcgatacacgacttaagtgat
cactcctctgctagtgactctttaagcggacatacgagtggggaatcgcttgcaggctca
actaacgattctagctctcagtctttagatgaagcgataaatcccattggacacccagtg
gatcggtcgttattcgcccatccaaattccaatggggaaatgtttgggaggtcttggacg
gcaagagagtcaagtttgcactcagttgttgatccttttggtggacaggtgcaagcacag
gtgcaggcacagatttacatgaagtcccaggcagattcacaggtgtctcagagcgactcc
cagaaacaggtaaactttcaaactaactttcagcagattccccaaaactttgagaattct
gtaccactagctatgccaataggcgatgatgattttcctgcattggatggagaaaatgcg
tatattgccctattacaacaacatcggcagctccctgattcaaggccagagcgcgccaag
cctgctcacttaggccaagatccattatttcatccaaacatgcatgtttcgtggcttagc
caacctgatcttacattt**tag**ttgtgtaaatgatgaattgaacgattgaacgattagaag
agatctcatgtacgtttactctggagttttttcttgatccttgcatgacctgaatttttc
agtgatcgtatttccctcattcttttcttaggctcgcataaccctcgaattgcaatatt
tgtcatcaaaaatttatagtgacttgttcccttgattaatcccagaccaggccgaaaagc

ORF

ATGTACCAGGCGTTGTTGGTTTTGTCTCTGATATGCTTTTCGTCGGCTAATTTTGTTAAGCTGCGAAGCAACGCT
GGTATGTTTTATGATACTATGGCTGGAGTTCCACGTTCAGATGAAGAGTTCTGGTTGCGTTTGGATATTAACCAA
GGTCTCTCTTGGACTCTGGATAGTAGCTACTACTCCTGTAATGGCTCAAATGTTTCGTCTTCCCTGTGTTTCAAT
TCTGCTCAAAACGTTTACGATGCTTCCAATAGTCCAACTGCAGATTTCGTTGATGTCTACGCAAACACAACTGTA
AACAATACAGATGAGGCATCGGCCGAGAGAGTAAATCTTACAAACAACTTATTTGCTGATGGCGTTTATATGGAA
GACAATTTTTACGTCACATTGAATAATGGAGCAAGAATGACTGCTACAGATCTGAAATTTTTGAATGCCCACAAT
AGTAGCGCCGCTGTGGGGTCTTTGGCGTTGGGGAGTTACACCTCACAGGACGTGCCAACTTTCTTACAAAGACTC
CAAAGCGGTGGTCTTATTGAATCCAACTCGTTTTCATTGGCATTAAACGAAATCGATTCTTCATATGGAGAGCTC
TATTTGGGGACAATAAACTCTACCAAGTATGTCGAGCCTCTGGTAGAATTCGATTTTATTCCGGTGTCAGATCCC
AATGGAGTTTTTGGATTCGATTGGGAAGATACATTCCCTACAGTTCCGATCAGCGGATTAAGCATGTCTTCGAAT
GACAAACAGAGAACTGTCTTTTTCCCCAATGAGTGGAACAACACGGTCTTAACGGGAACATACCCACTTCCAATG
ATGTTAGATTCAAGAAACATCTTTATCCATCTTCCATTCTCTTCAATCATACATATAGCAGTGCAGCTTAATGCA
CTGTATCTTGATACACTTCATAAATGGGCCGTGAACTGTTCTGTTGGTCAACTGGACGCAACTTTAAACTTTCAC
ATGGGTAACCTTACCGTTCATGCTCCTATCAAGGAGTTGATTTATCCAGCATACCAAGGAGACAAAAGGCTGAGC
TTTGCTAATGGAGAAGATGTTTGTATTCTTGCCATGGCTCCTGATGTTTACATTGGTTATCCACTGCTAGGAACC
CCCTTTTTAAGGAATGCAGTGGTTGCCGTTAATCATGATTCAAAAAAGGTCGCCGTTGCCAATCTTAATAGAGAT
AGCATTCCTCCCGCTTCGAACGTTTCTGTTTCGGAATCAATGGGAGTTTATGTTCCTCCACCTGTTTCAACTTCA
AGAACATCGGAGAGACCGTCCACACTAGATGAGACTAGTACAGCCAATTTTGACAAAAGGGAAGAGTCTGCAATA
TCATCAAGTTCAGTCACTAACAGCTCGTCTAGAAATTCTTCAACCATAACTTCTTCAGGAACTCAAACCGAGCAA
ACATCAGGCATAGCTACCATCGAAACAGATAGCATACCAGGAGCTCTAGGGAATAATTTAACTGATTATTCAACG
CTGACTCTAACAATATACACCAATTCCGAAGTGGACGAACTCAATCCTAACATAGCAACAGCATTCATTTCCAAT
GGTTCTATTTATTCAGAGCCTTACCCCTTTTCCGGAACTGCAGTTGCTGAATCATTCAGTGCATCACCTTCACAG
GCTGAAGGATCGAACTCATCGTCCTCAGGATCTTCTTTAGTTTTGTGTTTCTTTACATCATTGGCCAGTCTGTTG
ACTGTGAGCTGTCTACTACTGTAA

Downstream

attaattgaatagaaagtaataagtaaaatgtcatgattcaaaattacatatatatatat
atgaagttaa**tca**attatagttcccactgggggccatacttctccagatacaatcactt
cggtcatccattcccaattaagaagctgaatcctagtccttttagattcgggcaagtttc
ttgcttgaagcttcgaaacgctttttgacctttttgattttatttctggttcatcgtcac
tgtcttcattataattgattttcttaactgtccgtcgtcgtcgggtggtaatcaaatctt
gcgagtcaatccccctcttctttgtggtcgccccggaagcaagctcttctgtttgatcgt
taggggtttgattgtcaattttatcgtatgccaaattgagtgctccagttatctccttga
tggtcttgtcgtcatacaaggcacattcacatttatcttcccttccattagaggggtggt
atttttctatgatgctcaaagcttcctcgtagttggcaggtttgtccaacgatttaacac
caaaacaataagctagaaattcaaccacaccggctttcacctggtcaccatcctggatga
ctaagagtctcatgccgttcattttgtaattatttgcccgcaactgggagactaagcttt
ttccagcctcaaagttgtttataaagcgggaaatctctatactcttcaaccctgatacca
gtttggtctttcctgcaggcaaaagatatgctttccacctagaaaaagttttggaatcct
ttaagcaatcaccaataaattcatgccctaatacaggccaacccaacgccaacgcttcca
aatattttggtgttcttacaactccgtttgaaattatggcagcaaaagaaaaagagtcta
aactgtcctttggtgtcaattgtaaattatgttcgctgtcaataaattgggtcctcgagt
ctagaagaaagggtctaaacccattctccaaaatttttccgccatttgaatcgatttcag
ccttcaaatccaaaatctctttacctttttttatctgtggtgcctgtgatcgcaaataaac
aaccggaaaacacaccatggtcattatttaaatatttcgaagttaggccgagtgattcct
ttacctggcataaataatctttgaagaagtctgtgtcattaaatagtttcgacctagtag
ggcaacgaaaaagagtacctgacaagtaaatttcattcaagttaactttcttttcttcat
cctttcaagatccactcccttcttatgttccttaagatatactgtattgaatccacgag
cacattgaatatcagtgttatccgatgcaatgcttcccttaaacgctaatcccgtgacaa
tgcatgttctacgcttcagacttttccacttaatgacgtctccgattttgatatccaaga
tgccaagtacgttcatttccctaatccggaagacatccacaccatctgaaacatcaactg

AA

MYQALLVLSLICFSSANFVKLRSNAGMFYDTMAGVPRSDEEFWLRLDINQGLSWTLDSSYYSCNGSNVSSSLCFN
SAQNVYDASNSPTADFVDVYANTTVNNTDEASAERVNLTNNLFADGVYMEDNFYVTLNNGARMTATDLKFLNAHN
SSAAVGSLALGSYTSQDVPTFLQRLQSGGLIESNSFSLALNEIDSSYGELYLGTINSTKYVEPLVEFDFIPVSDP
NGVFGFDWEDTFPTVPISGLSMSSNDKQRTVFFPNEWNNTVLTGTYPLPMMLDSRNIFIHLPFSSIIHIAVQLNA
LYLDTLHKWAVNCSVGQLDATLNFHMGNLTVHAPIKELIYPAYQGDKRLSFANGEDVCILAMAPDVYIGYPLLGT
PFLRNAVVAVNHDSKKVAVANLNRDSIPPASNVSVSESMGVYVPPPVSTSRTSERPSTLDETSTANFDKREESAI
SSSSVTNSSSRNSSTITSSGTQTEQTSGIATIETDSIPGALGNNLTDYSTLTLTIYTNSEVDELNPNIATAFISN
GSIYSEPYPFSGTAVAESFSASPSQAEGSNSSSSGSSLVLCFFTSLASLLIVSCLLL

SEQ0185

Homologue of **MKC7** (YDR144C): GPI-anchored aspartyl protease (yapsin) involved in protein processing; shares functions with Yap3p and Kex2p

S. cerevisiae null mutant: viable

chr1-1_0379

5' region

gccttggtgcaatcaaatggtccatttgaatacactttgataagaaagctcgttttatgc
ctgaggattgtaagtctgctatgatatcctctgtgtaatagtacaccgcaactgggaaat
cgtaaaagagatttttgggaactcatatctatcacattcaagacattctaagttttgtca
attcaagatggctatgtatgtcaggctacaggttgtcgattgcattgttcccgtacactt
gatgtgctcccatctaaaataatagaggatatagttaaccactgttgactcaagtttggt
caaacaacttgctcagcagataagcttcattatttgcaatagaaaatcaactttagtcac
ttacaataaggggaatattcgaaggttccgtttctgtattatagctgctacctgagttag
aacggtctttaggctagttgcaatatatccaacagaatgggatctggacgacgttttcga
aggttaagatcattcgtctttctctcgtagaaattttgtcaatacattactcccatatga
gtgccggaattcttcataaactcaattaatcctttgacgagagatatgtcttagctttac
ataatggggattacttttgatacataatacaaggagcattgaaagaaaaaaaaaaaaat
taaaacattgacccttgagttgttactcggtccctttttacagccagcgaccaaaatatg
ctatctgacatcccggtcaccggttaaatagcattaggctgtgagtctggctggcccggg
cgaacatgaacatattggctgcaatcagtcactatttgattcaacattactcttgtttca
gatacttaatggccaaaggactgcggttttgataaatctgacatgagcccaagctcatct
ttcagaatggtattgaatgaaattttcctggtaagttaaaacatggatatatgcgtttct
aggagttgagatagttctacgaaaccgttattcagtgtctgcaaatatgtccaggcaacc

tatcgtggacaaagacgatggatgtcggctctatgatgaaatcagcggtcgccctctttt
tggatactgacgttcttgatgtcatgcatttttgagtgagacacatagcttgggtgtcac
acaataataaaaacgcagggctgtgtacagctgcgtctataatccacacatttgactagt
agcagcatggcagctaacactgttgataataaccatgatgtacgcaaatgcatttcattc
agtcttcgaaggggcaaggcaaaggaggtagataatgagggcctcgagtgaagtgttttt
aaatttgtaaaattggaaaaccttttcttcctgagaaaccttactgttcagcttacgctag
aaatattttgcctatatgtatttccgccacgccaagaattgaattacttcgtctgtatct
gttttctctccaagactataaataaacaaactcgtccaacttgatcaagtgccttaccga

ORF (frameshift present in bold region)

atgtttgtgatccagctggcattcctatgtctaggcgtcagcctaaccactgcacaacct
agttcacctttcaaggcaaataagtttccttttaaaaaggttcactactcatcaaaccct
agcgatcgccttattaagcgagacaactataagaagcttgacttgagacatcttggcgtc
ttgtatactgcggaaattgaaattggttcaggcaaaactgaaatcgaagttattgttgac
accggatctgcagatttgtgggtaattga**ctcaaatgcagccgtatgcgattgtcctatc**
ttgagatacaaggtacaagtgtttccacccttagtcaaactgccaacgtaacacccctat
caggtaaacttttgaatggacttcaagaaattggcattgtaactgatggcaaaatttcca
aaaagtttcaggaaaaccatcttttgaagagaaacgaggccttgaattttgatgtcgatc
tgaataagcccatttgtgatcaatttggatccttcaatccacagtcatcaagaacttttc
aaagcaacgacacagcatttagtatcagatatctggacaactcttttgccaatggatcgt
gggtgagggatacggtttatgttggtgattttgaaattgaccagcaaagttttgcattgg
ttgatatcacaaataactacatgggaattctgggccttggtccttctagtcagcagacaa
ccaatagtgatcctacagataacagtttcacttatcttggtattctggattctttgcggg
cccaaggattcattaattcagcctcgtactcggtttatctggccccagatggtaagactg
atgatactgatcacgatgatggtgagatcctgtttggtgctatcgacgaggctaaaatta
atggacagttgaagttgtttccatatgtcaatccttataaatcggtataccctgaccaat
acgcttcatacatcaccgtttccagtattactgtagccagttattttagtagccgcttgg
ttgaaagaatccctcaattagctcttttagacactggtgccacattttcttacttgccaa
cttatacgctgatacgtctcgcctatgccatccatcctggttttgagtatgtccgacaac
tgggtttatttattatagagtcaaacgtactctccagtgcgagacaaagtaccattgact
tccggtttggcaaagacgtagtaattcgatccaatgtttcagaccatctactcgacgtat
cacaatacttcacatctggacattatcttgcacttaccatccatgaaagtgtcgatgggc
ttctcattttgggtgacacgtttatcaagtccacctacttatttttcgacaatgataaca
gtgaattgggtattggtcagatcaaaattaccaatgacgaggatattcaagaagttggtg
aattcaccttagaacgcgattcagactattcttctacatggtccatttactcttatgaaa
cttctttggatcccttaagcactggcactggtacggggtcaacctattctcctactcgca
gtactacagctagaagcgaaccgactacgtctcgacgctccaccacccttcaacccagaa
caactgtgattccttctattgacaggctttcattgaacagcataactagtcatggttcct
ctactaacggaacctccccaactaatgagacttcttttgctgaggatggaggaactttga
cacccgaagaagcttctttgacaacttcactaaattctgctactatttctgagactactt
ttgtcgatgttgaaacttctactaccaatggtgcttcagttgtatctttgagtgttggtc
cctgcattattgccttcctactactcatctcttaa

Downstream

aatacgcgtttatgactttcaattgacttgcaactattagatttcatgcgtaattcaagt
tttcgaagcccacagagatctttattaccgtgttcaaatgatacaagtaattggttccaa
gacggtaccagttagttggattcttactccgggtcactaagacatcggatactcaagaat
agtcttgagctacttagagctcagttgtaccgttttctgatgttcatagaaaaacagatt
cacacttatgttgcttttatgacccttgcttttccctatacggctggacaaagtaactg
ctttttgcgattgctttatgatctatttttatggatgattgccaatgattgaagaacct
agaacttgtttttgaaaaaagattacaaattttcatgacaaggatggtttttgaagcac
ataactgactacgccctttgatcacgattatttctggaatacaaggagctgagctgaga
gcataatagcaactgaattcattaaaattaaatagatttctcg**tta**gacataaaaacctg
cttgattataatcaggagacctagctgatgacggcattagcagccattttgtttttggaa
caggaatataactgttacgttggtttttggtcattccgacatatccagcatagaaagcgc
acatagcagcaagaaaacaaaatacaccactagcagtttcacatccagtatgatttccaa
agatggcaatatcatacaaaaggaaaaacatagctgtgaagaacgtaagtaagaacaacg
gcaaggtcgatttaaaagtagccgtccaaaggacaatggcaaagacagtccaggctgcaa
cgtatagacccaacgcgttagtcaattcatccgcatcggtataggcatctactcccatcc
cgttctgaattgcaccaaatgaggcaaagaaaccaccataacatagcagcagtactgaac
caaagacgttctcaaggacaatactccagataccacaaatgatttggattaaacctgaag

aaaacaaaaaggaacctgttaaaactgctggacttgatacatttctggttccggcaagaa
taagacccaaagtgaacagagaagtacccagtgaaaacacgcctgctgggactggatttg
caaacttacgagaaggtctgacttggtagccgcttatcaagcccccagtgaaaggggcca
agtaatcttgttgagttgtaacaaactcacccttatgtatttcgtcgtgggggttatggg
ctgtttctgctgatgactcgatatcttgatgagtagacattaaaagttatattttcttaa
gaatggagtgtaattcggatgaattgagctttcttatatacatttctgatcgttattaat
gtgctcatgagtgtgaaggtcgatcaactaaaaaaccacgcagaaaactggtgattatgg
tccagttgagaagcttcagtggatcggaaattagacggaaaaggattgaagacgtattct

SEQ0186

Homologue of **PEP4** (YPL154C): vacuolar aspartyl protease (proteinase A), required for the posttranslational precursor maturation of vacuolar proteinases; important for protein turnover after oxidative damage; synthesized as a zymogen, self-activates

S. cerevisiae null mutant: viable; abnormal autophagy; decreased glycogen accumulation; absence of mitophagy; decreased aminopeptidase I modification; abnormal polyphosphate accumulation; decreased Pho8 modification; decreased Cps1 modification

S. cerevisiae reduction of function mutant: increased total cellular protein accumulation

chr3_1087

5' region

gaagcgactaaaaagaatattataaaaaagtcgagagagtcatattcttcaaaagtcgat
atgtggtctattggttgcttggtgtatgttatcttgacagcacatttaccgtttagcggt
tcgacacaagattcgttgacaagaaacatactagaaggtaattatcactactcattgcta
aaggagaacggaatctcaaacaaggccaaagatttccttgataggctcttgaatgttgat
ccaacaattcggttaggtgtcaaacaagctttaactcattcatggatatctgaattgtcc
gatgaaagccaagtcagcctttcacaaagtcaatcacatcaaagaagaattgaaagtcag
gaatctcattcattcatcacgggcatgaatggtgatgcatacgagcgaattgaagaagca
gatgaggtagatgagcagtgcgaagatcaagaaggagaacaccctcatgaggcccaaggt
tcgaaactgcagctagcaaaagaagcgagtttctccgtatctaatctttctcgctccccg
tacgttaagaatgaaatttctacttccattatagaaaatagtgtatcactgccagcatct
tttactcacaagcaattaaacaaagtaacaatggtctctaagcaattggaatcaccacag
gggacctttatcacgttgaatctagttgaaaattcagtgtccaagttcggtgcagtacac
ataccacaaggaaaaaccccatttgttgttggtagagattcatcttgtgactggttgatc
aaagaagaaagaatttccaaaatacactgcatgattgccaaaaaaaggcatcctactgct
aatccttccatatttgagtcacctgctttagggctggaagatatttggttactagatttt
agtacaaactcttgctttgtcaatgacattaaaataggcaagaatcgcaaaactcaaata
tttcatggagatgagatatgcttgttcaaagatgcccagaaaaaagagcaactcgtttat
agggttcatattgatgatggaacaggccttttccagggaggtgaaagaacccaagccaat
tctgatgacattctggatattgatgaggttgatgaaaagttaagagaactattgacaaga
gcctcaaggaaacggcatatcacccctgcattggaaactcctgataaacgtgtaaaaaga
gcttatttgaacagtattactgataactcttgatggaccttaaagatgtataatagtaga
cagaattcataatggtgagattaggtaatcgtccggaataggaatagtggtttggggcga
ttaatcgcacctgccttatatggtaagtaccttgaccgataaggtggcaactatttagaa
caaagcaagccacctttctttatctgtaactctgtcgaagcaagcatctttactagagaa
catctaaaccattttacattctagagttccatttctcaattactgataatcaatttaaag

ORF

ATGATATTTGACGGTACTACGATGTCAATTGCCATTGGTTTGCTCTCTACTCTAGGTATTGGTGCTGAAGCCAAA
GTTCATTCTGCTAAGATACACAAGCATCCAGTCTCAGAAACTTTAAAAGAGGCCAATTTTGGGCAGTATGTCTCT
GCTCTGGAACATAAATATGTTTCTCTGTTCAACGAACAAAATGCTTTGTCCAAGTCGAATTTTATGTCTCAGCAA
GATGGTTTTGCCGTTGAAGCTTCGCATGATGCTCCACTTACAAACTATCTTAACGCTCAGTATTTTACTGAGGTA
TCATTAGGTACCCCTCCACAATCGTTCAAGGTGATTCTTGACACAGGATCCTCCAATTTATGGGTTCCTAGCAAA
GATTGTGGATCATTAGCTTGCTTCTTGCATGCTAAGTATGACCATGATGAGTCTTCTACTTATAAGAAGAATGGT
AGTAGCTTTGAAATTAGGTATGGATCCGGTTCCATGGAAGGGTATGTTTCTCAGGATGTGTTGCAAATTGGGGAT
TTGACCATTCCCAAAGTTGATTTTGCTGAGGCCACATCGGAGCCGGGGTTGGCCTTCGCTTTTGGCAAATTTGAC
GGAATTTTGGGGCTTGCTTATGATTCAATATCAGTAAATAAGATTGTTCCTCCAATTTACAAGGCTTTGGAATTA
GATCTCCTTGACGAACCAAAATTTGCCTTCTACTTGGGGGATACGGACAAAGATGAATCCGATGGCGGTTTGGCC
ACATTTGGTGGTGTGGACAAATCTAAGTATGAAGGAAAGATCACCTGGTTGCCTGTCAGAAGAAAGGCTTACTGG
GAGGTCTCTTTTGATGGTGTAGGTTTGGGATCCGAATATGCTGAATTGCAAAAAACTGGTGCAGCCATCGACACT

GGAACCTCATTGATTGCTTTGCCCAGTGGCCTAGCTGAAATTCTCAATGCAGAAATTGGTGCTACCAAGGGTTGG
TCTGGTCAATACGCTGTGGACTGTGACACTAGAGACTCTTTGCCAGACTTAACTTTAACCTTCGCCGGTTACAAC
TTTACCATTACTCCATATGACTATACTTTGGAGGTTTCTGGGTCATGTATTAGTGCTTTCACCCCCATGGACTTT
CCTGAACCAATAGGTCCTTTGGCAATCATTGGTGACTCGTTCTTGAGAAAATATTACTCAGTTTATGACCTAGGC
AAAGATGCAGTAGGTTTAGCCAAGTCTATTTAG

Downstream

gcaagaataaaagttgctcagctgaacttatttggttacttatcaggtagtgaagatgta
gagaatatatgtttaggtatttttttttagtttttctcctataactcatcttcagtacgtg
attgcttgtcagctaccttgacaggggcgcataagtgatatcgtgtactgctcaatcaag
atttgcctgctccattgataagggtataagagacccacctgctcctctttaaaattctct
cttaactgttgtgaaaatcatcttcgaagcaaattcgagtttaaatctatgcggttggta
actaaaggtatgtc**atg**gtggtatatagtttttcatttacctttttactaatcagtttta
cagaagaggaacgtctttctcaagatcgaaataggactaaatactggagacgatggggtc
cttatttgggtgaaaggcagtgggctacagtaagggaagactattccgatgatggagatg
cttggtctgcttttccttttgagcaatctcatttgagaacttatcgctggggagaggatg
gactagctggagtctcagacaatcatcaactaatttgtttctcaatggcactgtggaatg
agaatgatgatattttgaaggagcgattatttggggtcactggagaggctgcaaatcatg
gagaggatgttaaggagctttattattatcttgataatacaccttctcactcttatatga
aataccttacaaatatccacaatcgaaatttccttacgaagaattgatttcagagaacc
gtaaacgttccagattagaaagagagtacgagattactgactctgaagtactgaaggata
acagatattttgatgtgatctttgaaatggcaaaggacgatgaagatgagaatgaacttt
actttagaattaccgcttacaaccgaggtcccacccctgccccttacatgtcgctccac
aggtaacctttagaaatacctggtcctggggtatagatgaggaaaaggatcacgacaaac
ctatagcttgcaaggaataccaagacaacaactattctattcggttagatagttggaagt
atggctcaaatagattggtatttgctccttcgcccgccttcagtgaagaatcatcagaca
ttgaacctaagcttcttttcactaataatgaaagtaataaacaaaaactatggaatcaaa
aaaatgcgtcaccctacaccaaggatgctttccatgagtacattgtcaatgaagatagct
ctgcaattaacccacaacaaaaggtacgaaggcctgtgcctggttttcttttgatgaaa
atggtggtgttccccctggtgattatgtaactatccgttacagattcacaagaacagata
gtgaccagcctgtaattgatgaagaggctttcgataaagtatttgctcgtagacaactag
aggctgacgagttttattggaggatttctcccttgccaatcagtgatgagctcagaagtg

AA

MIFDGTTMSIAIGLLSTLGIGAEAKVHSAKIHKHPVSETLKEANFGQYVSALEHKYVSLFNEQNALSKSNFMSQQ
DGFAVEASHDAPLTNYLNAQYFTEVSLGTPPQSFKVILDTGSSNLWVPSKDCGSLACFLHAKYDHDESSTYKKNG
SSFEIRYGSGSMEGYVSQDVLQIGDLTIPKVDFAEATSEPGLAFAFGKFDGILGLAYDSISVNKIVPPIYKALEL
DLLDEPKFAFYLGDTDKDESDGGLATFGGVDKSKYEGKITWLPVRRKAYWEVSFDGVGLGSEYAELQKTGAAIDT
GTSLIALPSGLAEILNAEIGATKGWSGQYAVDCDTRDSLPDLTLTFAGYNFTITPYDYTLEVSGSCISAFTPMDF
PEPIGPLAIIGDSFLRKYYSVYDLGKDAVGLAKSI

9 Chaperones

SEQ0187

Homologue of *Saccharomyces cerevisiae* **ROT1** (YMR200W):

Essential ER membrane protein; may be involved in protein folding; mutation causes defects in cell wall synthesis and in lysis of autophagic bodies, suppresses tor2 mutations, and is synthetically lethal with kar2-1 and with rot2 mutations

S. cerevisiae null mutant: inviable; cell shape: abnormal; chitin deposition: increased; killer toxin resistance: increased; liquid culture appearance: abnormal; resistance to hygromycin B: decreased; resistance to sodium dodecyl sulfate: decreased; viable

S. cerevisiae reduction of function mutant: growth rate in exponential phase: decreased; resistance to tunicamycin: decreased; vegetative growth: decreased

S. cerevisiae conditional mutant: increased heat sensitivity

FragB_0048

5' region

ggcccaggtggcagcttcaattaatgacatttacgacaaactacaatcacaaggacacag
tcaaggacgcagattaatgtcacagcaaagcaaacaatcaaccctatatacaacaactcc
atctcatcaaatagcatctctaaacccaaacctaccattgttgtctcctcctcatacatc

ttcacatcccaaactggaagagccagccatcaataagaacactactatcactagttttac
aatgcatagtcgatgctactgatactctagtatccacttactgtgcacactctccagtca
aagagtttcaaaagatggcccagatgcaattggtgtcattgaaatccagacaatttagtg
cacactctccattctctttaacaggacaagaagaagacatattctccgaagacacctact
gggagtctggagacgagtcattcgattctatttttgaacctcccaccatgaaattcaaga
accattcactcagctctgtgagctcatcacccacagaacaccattcaccaacaatttcga
ttacaaagccttggatatcacaatctcctcgcaagagccctaccttatcacaagtagctt
ctcgttaccgcaaggtttcataattgtaccagataccccgtatatccaaataaattgttc
acttgcttctgccctcctgtgaggtgccaaacttttcttacataatctactgcccgactt
taaatagccagacgcgaaacatcctggcaaaaattcagtttctgtctttccttgcataac
attctttttcctctttccactatctgccaagtgttctggaacgtgttgctacccaagccc
cttcgtttattactgtccttttaactggcgtatacaactttctggtaccactcattgttc
agcttcaggtccctttttcattttgttttccgcctgtttttttatcactatcccacagccac
attcctacatctgtgatcagtcagccacaaataatacaat

ORF

ATGGTTTTGATACAAAATTTTTTGCCGCTGTTTGCGTACACTTTGTTCTTCAACCAGCGGGCCGCCCTTGCTGAT
GACAGCGTCCCAGAATCCGAGCTTACTATCGTCGGAACATGGTCATCCAAATCCAACACAGTGTTCACAGGATCA
GGCTTTTATGATCCAGTGGACGAGCTATTGATTGAACCAGATTTGCCTGGGATCTCCTACTCTTTCACTGATGAC
GGATACTTTGAAGAGGCCCTTTACCAGGTTGCGGGAAATGCGAAGGATCATCATTGTCCCACAGCTGTTTTAATT
TTCCAACACGGTACCTACAGAGAGTTAGATAACGGAACTCTCGTTTTGGAGCCATATGATGTCGACGGAAGACAG
CTTTTATCGCAACCATGCGAGGATAAAGGGATTTCAACCTATTCTCGTTACAACCAAACAGAGGTGTTCCGTAAC
TATGAAGTGTCCCTGAATACGTACCACGGTCGTATCCAGCTGCAGCTATATGCCAGTGATGGTGTGAAACAGCCT
CCCCTCTACTTAGCGTACAGACCTCCCAACATGTTGCCTACGATCGTACTAAACCCCACTGCAGCAAGTGATGAG
GCCCAAGCAACCGCCACTGGTGCCTCTGCCAAGATAAAAAGATCACTAGAGAACAGATACAGGACCAACGCTGTC
AAGGAATCTTCTTTGAATTACACTCTATGGTGGTGGCTTGGGGCAATTCTGATGGGTGCTGGATCCACCATCTAC
TTCCTGTATTAA

Downstream

gtaatacgtatcttagataatcacacggtcttacctacaatcgatgttcgcgcgttcgcc
atcctcctactgtttaactgaaccttgtaatttggcttttccccgataacagtcccatac
ttctaacgaatgtcctggttcagtggaaagaactcagttcccttggaaaacaaaattaat
gaggcaacatccgagtttatacccgacggagaaatcgacctagaggtttctctggagatc
accgatataatacgatccaaacaggttactcctagggatgcaatgagagctctgaaacgt
cgattcatgggttccaacaatcccaatatacagaaatccagtatcaaacttattgacttt
tgtatcaaaaatgggggaatacattttgtccaagaaatcagtaccaaggagtttctggat
cccattgtattgaaactacatgataagagcctaaactctgaggtgaaggctctaatttta
gactccatccaaaactggtcgattttattcagcactaaccccaaattggagtacgtgaca
accatttacaataagctccaagatgagaagatattcgaatttccctcaatatatcacacc
gagaccattggagcaagttttattgaatctgaggttgcaccagaatggatggactcagat
gcttgtatgatatgctccgatttattcaccatgataaacaggaaacaccattgccgttca
tgtgggggagtattctgtggacaacattctgcaaagagatgcaagctgccaaaactgggc
atcaccctaccggtcagagtatgtgataactgctatgaccagcataagtcaagaaaacaa
cgccacaagaattcaaactcagtgacgaccgctgccgcaccttctgacgcggatatggat
gctgatttgaaactggcaatcgaactctcattaaaggatagcggtggttcacaatatcct
gtgcctgttgctggtccaaaagtatcttccaccgttaaag

AA

MVLIQNFLPLFAYTLFFNQRAALADDSVPESELTIVGTWSSKSNTVFTGSGFYDPVDELLIEPDLPGISYSFTDD
GYFEEALYQVAGNAKDHHCPTAVLIFQHGTYRELDNGTLVLEPYDVDGRQLLSQPCEDKGISTYSRYNQTEVFRN
YEVSLNTYHGRIQLQLYASDGVKQPPLYLAYRPPNMLPTIVLNPTAASDEAQATATGASAKIKRSLENRYRTNAV
KESSLNYTLWWWLGAILMGAGSTIYFLY

SEQ0188

Homologue of **LHS1** (YKL073W):

Molecular chaperone of the endoplasmic reticulum lumen, involved in polypeptide translocation and folding; member of the Hsp70 family; localizes to the lumen of the ER; regulated by the unfolded protein response pathway

S. cerevisiae null mutant: viable; competitive fitness: decreased; fermentative growth rate: decreased; resistance to rapamycin: decreased; resistance to L-1,4-dithiothreitol: decreased;

resistance to wortmannin: normal; resistance to tunicamycin: decreased; resistance to amiodarone: decreased; resistance to ethanol: decreased

Chr1-3_0063

5' region

gatagtttggctctggaagcttgaacttttttcatatcccctgggtttgacgtcccaaaaa
ctaatatccctttttgacctcattcaagggtagcacattcgtaagatctacttttctagat
tgattttgtctggagtgactccttggggcgtgttcctgggagcttgtatatctgtcatta
gggtactctgatgtatttcttaatctttgttttgagggtccatttctttcaaactgctgt
ggctgataatcagaagatctttctctttcccttgcactcaccattccattgtgtggtcta
taactatgataactatcctgcacttcttctggtgcataatctctgtaggaccttctctcg
tggcgctgattgcctcgtccacgagagtagtcattatacccccctatctctggagaaattt
ctggagaatccggtatcattataattctgtgaaaaatctttagctcctctctctccatct
ctcctataccctctgtagtcattgttgcgttctcttggtgggcctctagaaattgcaccc
ggagattcagtgcgataattccttcgagtatttgttacatcaatattttgatctgagtaa
agagcttgattatctcgggcataccctgcccttgctgttgtaccctaggatggtggttc
ggttcaatttcttcttgcttctcactagcaggcgcatcgggcatcaacttgttattcctt
cgagaagtttttaacagatcttggactttcagtgccccatagccattggaataaactggt
tttccattctctcgttggccagttaatctttcatcatggaaagctccgttatctttattt
tc**cat**ctgctagagcttatatcaaagaattcaaactaagaaaaaaagtctagggttagcg
ttccgtcgcgatttttcaccatctctcaagaggcaaggctcgcttcgctttctagaaaaa
accctgtagttggaaaagacacttgttcagatcaattgtt

ORF

ATGAGAACACAAAAGATAGTAACAGTACTTTGTTTGCTACTAAATACTGTGCTTGGAGCTCTGTTGGGCATCGAT
TATGGTCAAGAGTTTACTAAGGCTGTCCTAGTGGCTCCTGGTGTCCCTTTTGAAGTTATCTTGACTCCAGACTCC
AAACGTAAAGATAATTCAATGATGGCCATCAAGGAAAATTCCAAAGGTGAAATTGAGAGATATTATGGATCCTCA
GCTAGTTCTGTTTGTATCAGAAACCCTGAAACTTGCTTGAATCATCTGAAGTCATTGATAGGTGTTTCAATTGAT
GACGTTTCAACTATAGATTACAAGAAGTACCATTCAGGTGCTGAGATGGTTCCATCCAAAAATAACAGGAACACG
GTTGCCTTTAAGTTGGGCTCTTCTGTATATCCTGTAGAAGAGATACTTGCTATGAGTTTAGATGACATTAAATCT
AGAGCTGAAGATCATTTAAAACACGCGGTGCCAGGTTCCTATTCAGTTATCAGTGATGCTGTCATCACAGTACCC
ACTTTTTTTACCCAATCGCAAAGACTGGCCTTGAAAGATGCTGCCGAAATTAGTGGCTTAAAAGTCGTTGGCTTG
GTTGATGACGGTATATCTGTGGCCGTTAACTATGCCTCTTCAAGGCAGTTCAATGGAGACAAACAATATCATATG
ATCTATGACATGGGGGCTGGTTCTTTACAGGCGACTTTGGTTTCTATATCTTCCAGTGATGATGGTGGAATTGTT
ATTGATGTAGAGGCTATTGCCTATGACAAGTCGCTGGGAGGCCAGTTGTTCACACAATCTGTTTATGACATCCTT
TTGCAGAAGTTCTTGTCTGAGCATCCTTCCTTTAGCGAGTCCGACTTCAACAAGAATAGTAAATCTATGTCAAAA
CTTTGGCAAGCGGCTGAAAAGGCAAAGACAATTTTGAGTGCAAACACTGACACAAGAGTTTCCGTTGAATCCTTA
TACAATGACATTGACTTTAGAGCCACAATAGCAAGAGACGAATTCGAAGATTACAATGCAGAGCATGTTCATAGG
ATCACTGCTCCTATCATCGAGGCCTTAAGTCATCCATTGAATGGGAATCTGACGTCACCTTTTCCACTGACCAGT
TTAAGTTCAGTAATTCTCACAGGCGGGTCAACAAGAGTGCCGATGGTGAAAAAGCACCTAGAATCTTTGCTAGGA
TCTGAATTGATTGCAAAGAATGTTAACGCTGATGAGTCAGCCGTTTTTGGTTCTACTCTCCGTGGTGTAACTTTA
TCGCAAATGTTCAAAGCGAAACAGATGACCGTAAATGAAAGAAGTGTATATGACTATTGCCTAAAAGTTGGTTCT
TCAGAGATAAACGTGTTCCCAGTTGGCACCCCTCTTGCTACTAAGAAAGTGGTCGAGCTGGAAAATGTAGACAGT
GAGAACCAGCTCACGATTGGGCTCTACGAGAACGGACAATTGTTTGCCAGTCATGAGGTTACAGACCTCAAGAAG
AGTATCAAATCTCTAACTCAAGAAGGTAAAGAGTGTTCTAATATTAATTACGAGGCTACAGTCGAGTTATCTGAG
AGCAGATTGCTTTCTTTAACTCGTCTGCAGGCCAAATGTGCTGACGAGGCTGAATATTTACCTCCTGTGGACACA
GAGTCTGAAGACACTAAATCTGAAAACTCAACTACTAGTGAGACTATTGAAAAACCAAACAAGAAGCTATTCTAT
CCTGTGACTATACCTACTCAACTGAAATCCGTTCACGTGAAACCAATGGGGTCCTCTACCAAGGTATCTTCATCT
TTGAAAATCAAGGAGTTGAACAAGAAGGATGCTGTAAAGAGATCGATCGAAGAATTGAAGAATCAGCTGGAATCG
AAATTATACCGCGTGCGCTCGTATTTAGAGGATGAGGAAGTGGTTGAAAAAGGGCCAGCATCACAAGTTGAGGCT
TTGTCAACACTGGTTGCTGAGAATCTTGAGTGGTTGGACTATGATAGCGACGATGCATCAGCAAAAGATATCAGG
GAAAAACTAAATTCTGTGTCAGATAGTGTTGCCTTCATCAAGAGCTACATTGATCTGAACGATGTCACTTTTGAT
AATAATCTTTTCACTACGATTTACAACACTACTTTAAACTCCATGCAAAATGTTCAAGAACTAATGTTAAACATG
AGTGAGGATGCTCTGAGTTTAATGCAGCAGTATGAGAAGGAAGGTTTAGACTTCGCCAAAGAAAGTCAAAAGATC
AAAATAAAATCTCCTCCTTTATCAGACAAAGAGCTTGATAATCTCTTTAACACTGTTACCGAAAAGTTAGAGCAT
GTCAGAATGTTGACTGAAAAGGACACTATAAGTGATTTGCCTAGAGAGGAGCTTTTTAAGCTGTATCAAGAATTG
CAGAACTACTCTTCCCGATTTGAAGCAATCATGGCCAGTTTGGAAGATGTACACTCTCAAAGAATCAACCGTTTG
ACAGACAAGTTACGCAAACATATTGAAAGGGTGAGCAATGAAGCATTGAAGGCAGCTCTCAAGGAAGCTAAACGT
CAACAAGAGGAGGAAAAAAGCCACGAGCAGAATGAGGGAGAAGAGCAAAGTTCTGCTTCCACTTCTCACACTAAT
GAAGATATAGAGGAACCATCAGAATCGCCTAAGGTTCAAACATCCCATGATGAGTTGTAG

Downstream

tatcccacagttatatatagaccacataattcattagagatatttacgcgcgattaccac

ccctcaaagataaaaacgacttgaacctcagatatcaccataccactcgaaacaactacc
aa**atg**tttggactagggatattattttacacagttctattccttgtgaatgccatagcga
tcttatcagaagacagatttctaaatcgaattggatggggctcatatgccccagcaagta
ctccgcaagctagcttccaaaccactacatttgacaactacaacgccaacccagaagttt
ctatgaaaacgaaaattgtaacattaataagtgctacaagaactttgcttagactgccat
taattgctatcaacagcgtagtcataatctacgaattggtgctagga**tag**atgaatacac
attaaaaattatctatacactaaccttaatacacagttcgatgtcatgattcatacc**cta**
gtattccgcttcaaaggggggttgctcctttcttttagtctcaacgaagttctcaacgat
ggatattcgttcgcctgagtaatcgcaatgatcatgaactacatttgagtccttgtcctt
gagatacaatcgaaggaataagatgatggtatcataaaccagcagatattgtctcaaagt
ttggaccatgaaaacacgctgctttctaaactcttctactattctccaaactgggtcata
ggtggcgctacagtttttcgatgtcaaactgctatcctttaacagatgttccacaactgc
gtcgattgaacaaaatgttccggttcttccacagcccgcagagcaatgaacgactattgg
gacattatgttctttatctcgacaattcctggcaatataccttctgagataaattaaatt
caagaggtcattcatcaatagtgtcccgaaatctggccaggaagaaacttgcaactgtaa
tacttctctgctgatattagcagtactgacacgtagacgt

AA

MRTQKIVTVLCLLLNTVLGALLGIDYGQEFTKAVLVAPGVPFEVILTPDSKRKDNSMMAIKENSKGEIERYYGSS
ASSVCIRNPETCLNHLKSLIGVSIDDVSTIDYKKYHSGAEMVPSKNNRNTVAFKLGSSVYPVEEILAMSLDDIKS
RAEDHLKHAVPGSYSVISDAVITVPTFFTQSQRLALKDAAEISGLKVVGLVDDGISVAVNYASSRQFNGDKQYHM
IYDMGAGSLQATLVSISSSDDGGIVIDVEAIAYDKSLGGQLFTQSVYDILLQKFLSEHPSFSESDFNKNSKSMSK
LWQAAEKAKTILSANTDTRVSVESLYNDIDFRATIARDEFEDYNAEHVHRITAPIIEALSHPLNGNLTSPFPLTS
LSSVILTGGSTRVPMVKKHLESLLGSELIAKNVNADESAVFGSTLRGVTLSQMFKAKQMTVNERSVYDYCLKVGS
SEINVFPVGTPLATKKVVELENVDSENQLTIGLYENGQLFASHEVTDLKKSIKSLTQEGKECSNINYEATVELSE
SRLLSLTRLQAKCADEAEYLPPVDTESEDTKSENSTTSETIEKPNKKLFYPVTIPTQLKSVHVKPMGSSTKVSSS
LKIKELNKKDAVKRSIEELKNQLESKLYRVRSYLEDEEVVEKGPASQVEALSTLVAENLEWLDYDSDDASAKDIR
EKLNSVSDSVAFIKSYIDLNDVTFDNNLFTTIYNTTLNSMQNVQELMLNMSEDALSLMQQYEKEGLDFAKESQKI
KIKSPPLSDKELDNLFNTVTEKLEHVRMLTEKDTISDLPREELFKLYQELQNYSSRFEAIMASLEDVHSQRINRL
TDKLRKHIERVSNEALKAALKEAKRQQEEEKSHEQNEGEEQSSASTSHTNEDIEEPSESPKVQTSHDEL

SEQ0189
Homologue of **CNE1** (YAL058W):
Calnexin; integral membrane ER chaperone involved in folding and quality control of glycoproteins; chaperone activity is inhibited by Mpd1p, with which Cne1p interacts; 24% identical to mammalian calnexin; Ca+ binding not yet shown in yeast
S. cerevisiae null mutant: viable; decreased competitive fitness; decreased resistance to hygromycin B; increased resistance to mitoxantrone

chr2-1_0322
5' region

aagagcgcgatggagtgtagcgtgattacatcatcagatgctacattgattctctgatat
gaatggtgatggaactttctagaggttccttgaagaaataaatacatttacaagcagaac
tccacttttt**tca**cggagaatcatctaagttaggcatacgaaggatctcgccttcgttgtt
tgcactcatctcctgtagtttagcgagaatcttggagtccttccacttttcaggcaatgg
ggtaacctcgtagtttttcacggcccagtaataaatatcccaatccaactcgtttaataa
gtcatcatactcttccatctcttccacactcattgtcggtagatagcgttttgcgaaacg
agacagaagaaggtctgtttccaagattcctctctttcttgactgataaaccagacggcg
tctcttgacatcttccgactcgttatcacgtttcagaggttcaactttcagtatcagctc
ctgcctcaagaaggggagagaatgaaaagatttcgaaaacacccttggacaagtcttgct
accttgaaactgagttctttggaaaagccggagcataatgggtgaattaagcagaaagaa
ggtaactgatttgctgagacccaaatcatctacagtttcgcgaagcataaagttcacact
gattttctggggaagaactggtaaaccacatgttgtctc**cat**tccacgataaaccgttca
agcaaggccgtcttagaatgcacaagacaatttaggtaaactacctttcctggaagcgaa
agcagacgttacaatctgtttcatcccccaactgcactcctctctcctctgctagccaag
acgatctttcatagaatttgatggaatttacgcgaaatcgccacgtaatcatatttcgaa
cagatctaccaccctatcctactgcatctatatatactcacctagaaccccttttttgaa
cggactacttctcttgttactcatcagttcttttgccacgtc

ORF

ATGAAGATCTCTACCATTGCAAGTTCTACGTTGTTCGCTGTTGGTGCTTTAGCCGAATCCGAACCCGCTGAGTTC
AGACCCTTGGAAGCTCAGTTGGACAAGTCATCTTTCTTTGAACAATTCGACAAGGAACCGAAACTCGGCGACACC
TGGAAGATCTCCCATGCCGTTAAGAATGAAGAATTCACTTATGTTGGAGAATGGGCCATTGAGGAACCTGTTGTC
TATCCTGGATTCAAGAAGGACAGGGGTCTGGTTGTGAAATCTGAGGCAGCTCACCACGCAATATCTGCCCAATTA
CCACAGGTATTTGACAACACTGACAATACGTTGGTCTTGCAATACGAAGTCAAGCTTCAACAAGGATTGAACTGT
GGAGGTGCTTATGTTAAATTATTGAGTGCTGAGGGTCTGAACAAGAATGAGTTCTCTAACGAGACCCCTTATCAA
GTCATGTTTGGTCCTGATAAATGTGGAACCACGAATAAAGTGCACTTGATTATTAAGAGGAAGAACCCAGCCACC
GGCGAATATGAGGAACATCAATTGGCTACTCCTCCAATGGGTAGAATCGTCAAGACTACTTCTCTATACACCCTG
ATTATCAAGCCCAATAATGACTTTGAAATCAGAATCAACGGTGAGGTTGCTAAAGCTGGTAACTTGTTGAACGAG
AAGTTGATAAAGCCACCATTTGGCGCTCCGAAGGAGATTGACGATCCGGAAGACCAAAAACCCGAAGATTGGGTT
GATGAAGACATGATCCCAGATCCAGATGCTGTCAAGCCTGAAGATTGGGACGAGTCCGAGCCATTGCGAATCGTC
GATCCGGAAGCTGTGAAACCAGAAAACTGGAACGAAGATGCTGAATTGTACATCCCTGATCCAGAGGCCACCAAG
CCCGAAGACTGGGACGATGAAGAGGATGGCGAATGGGTTGCTCCTGTTATTCCAAATCCAGAATGCGCAGATATT
GGATGTGGCCCTTGGAAGGCTCCATATATTACCAATCCTAACTACAAGGGTAAGTGGAGTCCTCCTCTTATTGAG
AACCCTGATTACAAGGGACCTTGGAGCCCTAAGAAGATTGCCAACCCTAACTACTTTGAAGACAAGAAACCAGCA
AACTTGGAACCTATTGGAGGTCTTGGTTTTGAGTTGTGGACCATGAATGCAGATATTCTCTTTGACAACATCTAC
CTAGGCCATTCTATTAAGGAAGCTGAATTCATTGGAAACGAAACTTTTGTTCCAAAATTGGAGTTGGAGGAAGCA
GAATCTGCAAAGAATGCTCCAAAGCCAGATTTTGAGCCTGAAACCCCACCTGAAACTGGTTTAGACTCTACTGGA
AACATTTTTACAGATGTGTTCGACACTGTTTTCCAAACTGTATTGGAGTATTACGTCAGTGCTAACGCCTTTTTG
AATGATGTTGTTCAAGAACCCTCAGTTTTGTTGGAACGTCCGGGCGAAGCAGTTTACTACTTGGTCACATTTTTT
TCGGGTTTCACCTTCCTTGTAGCCGTTTGGTCAGGTTTGATTTTTGCATTGACCGGTGCTGGCAAGAAGCCAGAG
ACTTCTGCAAAGACACCAACAAAGTCTACCCAGAAGATCGAGGAGGTTACTGAAGATGAGACTGAGAAGACTGAT
TCCTCGTCTGCTGCTAAGAACTCGACTAAAGCTACAAAGAGAACCTAG

Downstream

atatgtttataatgtgtacacactaattaaaaattgcctgtttgacgtattgcttgctgt
ttttggagttgggtatatctatctcgcctagattgctttgctcgttatcgattttcaacc
actaagtaaacccagag**atg**aagaaatttttgttgaaattcatcactaaaaagttctttg
gagagaccaagctgaaccggtttggctttgatgatccttattacgaggaggtgcaactcg
atcctaccgaccccaattccaaagtgattcgcagttatatccagccacctgtggggcttt
cagataacgacaccaaactgttccaaaagattcaaagaagatcctataggctagactact
tattcaatatttgcggagtcaagttcggttggaatggagttatatctgtggttcctgttg
ttggagatgtgttagctgcaatcaactcattgctaatctacaggctaatttttcaattgg
acaatccagttccatgggatatccatttcttcatactgctgaatattatagccgattttc
tgtttaagctgattccaattgtgggcagctttgtggacattggattcaaagccaacagtc
gtaatgctgcgttaattgagaaacatttggcaaccgttgctaaagcgaatcaatttttca
acgatgtggattctgaatcagcgtcaactaaatcactggtatcctggcagacctccagtg
tcaccgggatacccgatattgttggttcgttgaggtccagaatctttgggaggggagcaa
aagtagaagaaaaaccgattgagtccatcgacttgggggtcagctccaaaagtgttctca
ataagcctaggccaagggtgaactccacgggaggaatttcaactgccagggcggtagatt
ctcctcaaaagaccatgctgaaa**tga**gtaggccgttcgataggtcgatgatgtaagtaaa
ggaagtatacagcgcgatgtaataaaatttgttccaacat

AA

MKISTIASSTLFAVGALAESEPAEFRPLEAQLDKSSFFEQFDKEPKLGDTWKISHAVKNEEFTYVGEWAIEEPVV
YPGFKKDRGLVVKSEAAHHAISAQLPQVFDNTDNTLVLQYEVKLQQGLNCGGAYVKLLSAEGLNKNEFSNETPYQ
VMFGPDKCGTTNKVHLIIKRKNPATGEYEEHQLATPPMGRIVKTTSLYTLIIKPNNDFEIRINGEVAKAGNLLNE
KLIKPPFGAPKEIDDPEDQKPEDWVDEDMIPDPDAVKPEDWDESEPLRIVDPEAVKPENWNEDAELYIPDPEATK
PEDWDDEEDGEWVAPVIPNPECADIGCGPWKAPYITNPNYKGKWSPPLIENPDYKGPWSPKKIANPNYFEDKKPA
NLEPIGGLGFELWTMNADILFDNIYLGHSIKEAEFIGNETFVPKLELEEAESAKNAPKPDFEPETPPETGLDSTG
NIFTDVFDTVFQTVLEYYVSANAFLNDVVQEPSVLLERPGEAVYYLVTFFSGFTFLVAVWSGLIFALTGAGKKPE
TSAKTPTKSTQKIEEVTEDETEKTDSSSAAKNSTKATKRT

SEQ0190

Homologue of **YDJ1** (YNL064C):

Protein chaperone involved in regulation of the HSP90 and HSP70 functions; involved in protein translocation across membranes; member of the DnaJ family

S. cerevisiae null mutant: colony shape: abnormal; gamma ray resistance: decreased; heat sensitivity: increased; resistance to daunorubicin: decreased; resistance to camptothecin: decreased; resistance to cisplatin: decreased; resistance to calcofluor white: decreased; vegetative growth: decreased; competitive fitness: decreased; resistance to rapamycin:

decreased; resistance to L-1,4-dithiothreitol: decreased; resistance to hydroxyurea: decreased; resistance to doxorubicin: decreased; resistance to wortmannin: normal; resistance to cisplatin: decreased; resistance to 4-(N-(S-glutathionylacetyl)amino) phenylarsenoxide (GSAO): decreased; resistance to 1,2-bis(2-aminophenoxy)ethane-N,N,N',N'-tetraacetic acid: decreased; resistance to amiodarone: decreased; respiratory growth rate: decreased; respiratory growth rate: absent; viable

chr2-2_0066

5' region

tctggatcacccacaaatgctcccttagctagtgcttctgctcctgttgctggaggtccc
ccgccacccctccacctcctcctgctgatgtttttcagactgatgattcgaaagggaag
aaaccacaaggaatgggagccgtgtttgctgacttaaacaaaggagaagctattacatca
ggtctgcgtaaagtagacaaatcagaaatggttcataagaatcctgaactgcgccaatct
aatgttgttccagctaataaggaaaaaccaggaaagcctcaacctccaaagaaaccgtct
ggtttgtctacaaagaaacctgcgcgtaaggaattgcaagacaacaaatggctggtcgaa
aattacgatggtgaacaagagattgtcattgaaggagagatgcagcagagtgtattcatc
agcaagtgtgacaactcgacaatcacaattcgtggcaaagttgctgctgtctcattgact
caatgctccaaggttggtctagtagtagactccttggtgtctggaattgatgttatcaaa
agtcgtggattcggcgttcaggtactagggaatgtccccaccatcaccattgaccaatct
gataacggtcaaatatatttgggagcagaaagcctgggaaccgagatttacgcatcatct
accacatcgttgaacataaatgtcccagatgaagacggagatttccgagagatggcagcc
cctgaacagttcaaacatacggtggctggaggcaatctacaaaccacagttgtagagcac
gtcggt**tag**caaaagacaccaaccctaccttgactctggaactgatctgtttataacatt
taagtgtatttacgtatttactaagagatcagatcgcgaatgaaaataggttttccagaa
acttctctcactggtacacgactgtttttcttgagaaaaaaaaaatttagaggcaaaact
tccttactggacatacacctagagatatatcaatcgtgag

ORF

ATGGTTAGAGAAACAAAGTTATATGACATCTTAGGCGTGTCTCCAGACGCTACCGATGCACAACTTAAAAAAGCA
TACAGAGTCGGAGCACTGAAGAATCATCCTGACAAGAATCCATCTCCAGAGGCAGCTGAAACGTTCAAGGGAATG
TCTCACGCATACGAGGTTCTTTCCGATCCTCAAAAGAGAGAAATATATGACCAATACGGAGAAGAAGGCTTGAAC
GGTGGTGGTGCTGGACCAGGTGGCATGGGTGAGGACATCTTTTCACAATTCTTCGGAGGAATGTTCCCAGGAGGT
GGCCAACCTACAGGACCACAGAGAGGTAAAGATATCAAGCACTCCATCTCTTGTACATTAGAGGAACTGTACAAG
GGCCGTACAGCTAAGCTTGCTCTCAATAAGACTGTCCTATGTAAGGAGTGTGACGGTAAGGGTGGTAAGAATGTC
AAGAAATGTTCCGCCTGTAACGGTCAAGGTTTGAGATTCGTAACAAGACAAATTGGCCCCATGATCCAAAGAGCC
CAAGTGAGATGTGATGTTTGTAACGGAGAAGGTGATATAATCAGCGGAGCTGACAGATGTAAGGCATGTTCTGGT
AAAAAGATTACCAACGAACGTAAAATTTTGGAGGTCAACATTGAACGTGGTATGCGTCATGGGCAAAAAGTTGTT
TTCAGTGGTGAATCGGACCAGGCGCCCGATGTTATTCCTGGAGATGTTATCTTTGTTGTTGATGAGAAACCTCAC
AAAGACTTTTCTCGTAAGGGCGATGACCTTTACTACGAAGCCAAAATTGATCTCTTAACTGCCTTGGCTGGAGGA
GAGCTTGCTATCAAACACATCAGTGGTGAATATCTTAAGATCACTATCATTCCAGGTGAAGTTATTTCTCCAGGT
TCCGTCAAGGTTATTGTTGGTAAGGGAATGCCTGTACGCAAATCATCCAGCTACGGTAACCTTTATGTCAAGTTT
GAAATTGACTTCCCACCTAAAAACTTTACAACTGCTGAGAACCTGCAACTATTAGAACAAGTCCTTCCGGCTAGA
ACTCCAGTTTCTATCCCAGCCGATGCAGAAGTTGATGAGGTGGTCTTAGCGGATGTTGACCCTACCCAGCAACAA
CGCCAAGGAGGACGTGGTGGCCAGTCGTATGATTCAGATGATGAAGAGCAAGGTGGCCAAGGTGTCCAATGTGCT
TCTCAGTAA

Downstream

ttgtatttgatagtgaacataatgtaaatctgattatctgcgttcgtgtctatgaattat
ttatggtagtttatttctat**cta**cagtttgtattttgttcccctcctagccacatagag
aacagggatcatggcaaccaatgagatacctgcaaagaaagtgtagcaaccaccgatggt
cattgctgcaatcatcttggttagcactcctacgaatagggcgctcattaggcatcttgt
gatgttcacgcaagaagtggcggtggatgactgtcctgggaacatgtctatcaaaagagt
ggtcgacatagtcattgttcccagacatgcccatgtaaagaaaaatattgtgatgatgat
gggggctggatgagtccttttatctaaagtccacccaaatgttaactgggaagccacaca
gatgagggtgtagacccagtaaattgagagccttgccttgtaaatggaaaactctgactt
caatttctggcaacgataagatctgtcaaggatttttccagttgcaaaagatcctaaaac
acatcccactcctgatgatatgaatactaatccgatttttaaagtcgaataatggtaatg
attctcaagctcagtagcaagagtagttaaacaacaagtccaacttgcaaaaagaattga
gcacggtatcagcgttaagtatactggtggcgcaatcaagatcctaaaggtgtccagcag
attgtaatgagggtgtgggtctgtctgggcagagtcatccagtcgtctctggtacgatgg
taggagcaagataggacaataattccaaagtctgcgtggttccttggacatgtttccaac

gatagaacgattcgtctctggtaagaaagagaaatccagcgtgaataccaacatacccag
gactgctaagaaccaaaagacagctctccatgattggaatccttcatcgacggcacctcc
taagagcggaccaatggcctggcccattaaaagacagcca

AA

MVRETKLYDILGVSPDATDAQLKKAYRVGALKNHPDKNPSPEAAETFKGMSHAYEVLSDPQKREIYDQYGEEGLN
GGGAGPGGMGEDIFSQFFGGMFPGGGQPTGPQRGKDIKHSISCTLEELYKGRTAKLALNKTVLCKECDGKGGKNV
KKCSACNGQGLRFVTRQIGPMIQRAQVRCDVCNGEGDIISGADRCKACSGKKITNERKILEVNIERGMRHGQKVV
FSGESDQAPDVIPGDVIFVVDEKPHKDFSRKGDDLYYEAKIDLLTALAGGELAIKHISGEYLKITIIPGEVISPG
SVKVIVGKGMPVRKSSSYGNLYVKFEIDFPPKNFTTAENLQLLEQVLPARTPVSIPADAEVDEVVLADVDPTQQQ
RQGGRGGQSYDSDDEEQGGQGVQCASQ

SEQ0191

Homologue of **EPS1** (YIL005W):

ER protein with chaperone and co-chaperone activity, involved in retention of resident ER proteins; has a role in recognizing proteins targeted for ER-associated degradation (ERAD), member of the protein disulfide isomerase family

S. cerevisiae null mutant: resistance to rapamycin: decreased; resistance to wortmannin: decreased; resistance to amiodarone: decreased; viable

chr2-1_0421

5' region

aggaagaagtcaggactcataaagactataaagagcatatatttgaaggaaggtattcgt
gggttctattctggtttcgtcattaatttaaccagaaccttaccctccagtgcagtgacg
ctagtctcttttgagtacattaagaactaccttgacaagattgcttctcccgcccta**tag**
tcatcacatttgtagcatataatattcggttttatagagtttggttcgattccaaagaat
tgcaccaaggagacattaaaggtggttaagataagtgcctgttaatcaatacagcgtgaa
tcttggtgccttgaatggactcaacaaagtatttaatttttgatttttcagtgaataccg
catcgcatggcaccagcgaagcagactgcttattgtattagatcgaaaatcataaacgag
ccattcacttgagcaaacaactgtatacagtcagacaaaagagtgttagcctgtagctat
tttcactcaactcgttagctcaaaatcccacaaaactactgattgacttctcatgacata
aacatgttcatatgcacaataccaggaccagacacgaaaaactagacacgagacatctat
ctgttcgcgattggctttgccttccgtagattactaccgttctacagttacaaacccttg
ataaggacagcatagcacaaaccgtttcattctgacaacagtcaccatgaaagttgcaaa
atctatcgtggccagtcttttggcaatagcttttgctaaagcagccgatccagccgaggg
ctccggtaatgggggtggagaatgggccgactgtcatcagtatcaactccatccagctgc
cagaacttttgacacccgatactttcaatgagactcttgcccaaggaatccactttgtag
agtttttctccctactgttatcattgcaaacaactggagcccatttggaaagaaactta
tcaccgcttcttgaaagaaagtgcgatatcagggatcagt

ORF

ATGAGACAAGTTGATTGTGCCGCAAACGGTGATTTCTGTGGAAAACAGTCAATAAACTCATATCCCACTTTGAGA
CTGTATGGCCCTAGTGAAGATAACACCTCGTACAAATTGATAACAACTTACCCGAGTGGAGGCAAGAGAACGCCG
CAGAATTTCTTGAAGTTTTTGAGATCTCAATATGATGATCTGAAAAACGACCGTTTCAACCTTCCCGTTAAAGGA
GAGGTTTTGACTGAGGAAAAAATGTTGAAACTAATGAATGGAGATATTGAAGAACCAGTATTGGTTTCATTCTGG
CCTACCACTGATAAAGAAACAACTTTGAAATATTTTGAAGATGTTCACAGCAACCCAATCAGTTTCAAGAACTGC
TATTCGTGCTTTAACTTGGCAGTTTTATGGTCCAGAATAACCAATCGTTTGCCTGACCTGGAGACTGCCGTCTTC
AACTGTGGAGGGACTAATTCTCGAGTATGCCAAGCTCTCAATCTACCCTTTAACGTAAACAGAGCTCAAGTTTCG
CCTGAGATATACATGTTCTTGCCAAACTCTCATGGAGGCATTCGTGTGAAATATAATCATGACCTGGTAATTTCT
GACATTGTAGATTGGTCAGAAAGATTGTTAGCCAATGCTCAAGCTGAGGAGGTCACTTTAGATTCATTGGCGGAA
AAGATGACTTTATTGAATCCAGCTAAGGGTCTAGACTTTTTCAAGGGTCCAGATCCAAACCAGAAGCAGGTATTT
GTCTACTATTATGAAGAAGGCTCTGACGCACCTGAGGATTTTGAAATATGGCCTCATCTCCTTCAGCCTATCATG
GATTTGACTACGAATACTTATATCTATCGATCCAAAGATTCTCAGCTAGAAGATTTGCTTGATAAAAAGTACAAG
AAGCTTATTGATTATATCAATCAACCTGACTTTGAGCCGGAGAGACCTTTAAATAGGGAGACCTATCTGGCAAGA
ACAATCACATCCGTTCCAACATTCCTCGTTTTCAAAGATAATAATATGATTCCAACAGTCTACCAAAACTTTTCA
CCCAATGAGATCAGAGATGTCAAGAAAGTCGTCAATTTTATTGAACAGAATCAGTTCCCATTAGTGGACAGATTG
ACAACTGATAATTACGAATCGTACTTCCCAAAGTTTAACCCACAAATTCACGATAAAGACGAAAAGATTGTCATT
TCTTTCTTCTCTTCTGACAATAAAACACAGACTACTAATGATTACAACCAGCTTTTGTTTGTTCAGCATTCGTAT
GACTATATCAAACACCAAAATAGATTTGATTTGATCGAAAAGGCTCGAGCTGACAAGACTGACAAGTCAGAGCAG
TTGAAGGGGGAGGGCATGGATCCAAAAGAGATCATCAAAGTTTTATCCAAGAAGATTGACCATTTAAACAACGTT
GGAAATGTGTTGACGGTATACGTGGATTTAGCTGATGATCGGAAACTTCTTGAGAAGACAGGATGGATATCGAGT

AGAACAAAGTACAAACCAGGTGAATCTATAATAGTGACCAGATTCGGCAAACATTATTGGGATAGAGATGTTTTT
GGACACAAGCTAATGACCACTGCTCAAAGCTTGCGAGAAACGCTCTCTTACCTTTTGTTTCCATCCACATATATT
CCACCAGAAAAGAAGACTACAGTGAGACCATCTCCTAGAGTTGCAGGGTCACCATATCCCGATGCATTTGCGTTC
GTAGATATCATTCATCAGTACGGGTTGTTTGGTTATCTGTTAATTATTACATCTGTTATTGGGGTGTACATGCTT
ATCAAGAGTCATCGGAGAAGGTCCCGCAAGGCCAAATTTCTTTCGCACAGAAGCCACAAGGAAGGATTTCAAGAT
GCTTACATCGAATTGAGCAAACACGATTAG

Downstream

tagatagagtctcacaccaataggccatcagtaggcgacaatcgcttaaagtgtgtgggt
atcggtgcgcgatatagccttacataatacattcagatcacacatgttgtgagccacagt
ttaactcttcttcatacttgtatcgacatccctaagaacata**atg**cagctttactatcat
gatggtttagacaccccagataatttctgtgaagatcacaattctggggaagaggtgtca
gttgaagagctgtccaaaattggagtgatatactaccatttggacgaagagagtcaagtt
gagaagattgccagggaaagagattacaagaacagagattttgtggttctatctgatgag
actttccctggtgggccagatgctttgcaagagaaactgaaagtcttctacaaggagcat
cttcacgaagatgaggagattagattcatcctggatggagaaggctactttgacgtgaga
tcacagaacgacaggtggatccgaagcaagcttttcaaaggtgatttactgattcttcct
gctgggatttaccacagattcacgttgtccactaagaagtatgtcaagacaatgaggttg
ttcaaggaggagccaaaatggcaagcctataacagaccggctgatgacaaaggagctaga
ttggagtacttgaaaagtattcaggtg**tag**gtttttgtagttaacgaatatatgtgtatt
tattactttcttgggttgaaagatgcagagatcatttggaggtgacattggtgaaaaaaa
tgaagaactagagtgttcagaatataatttttttcgacagctgacattgcaggtaccagcc
tctagtttcctaatgatcaattttggtagctaaagatcatctttgcttgctaacgtatc
gatggactcaaaatgtcttcttttaagtgccaacctggattgtcttaaccagttcgctaa
cattgaaatattcgaattccctactcatcccaccattatg

AA

MRQVDCAANGDFCGKQSINSYPTLRLYGPSEDNTSYKLITTYPSGGKRTPQNFLKFLRSQYDDLKNDRFNLPVKG
EVLTEEKMLKLMNGDIEEPVLVSFWPTTDKETTLKYFEDVHSNPISFKNCYSCFNLAVLWSRITNRLPDLETAVF
NCGGTNSRVCQALNLPFNVNRAQVSPEIYMFLPNSHGGIRVKYNHDLVISDIVDWSERLLANAQAEEVTLDSLAE
KMTLLNPAKGLDFFKGPDPNQKQVFVYYYEEGSDAPEDFEIWPHLLQPIMDLTTNTYIYRSKDSQLEDLLDKKYK
KLIDYINQPDFEPERPLNRETYLARTITSVPTFLVFKDNNMIPTVYQNFSPNEIRDVKKVVNFIEQNQFPLVDRL
TTDNYESYFPKFNPQIHDKDEKIVISFFSSDNKTQTTNDYNQLLFVQHSYDYIKHQNRFDLIEKARADKTDKSEQ
LKGEGMDPKEIIKVLSKKIDHLNNVGNVLTVYVDLADDRKLLEKTGWISSRTKYKPGESIIVTRFGKHYWDRDVF
GHKLMTTAQSLRETLSYLLFPSTYIPPEKKTTVRPSPRVAGSPYPDAFAFVDIIHQYGLFGYLLIITSVIGVYML
IKSHRRRSRKAKFLSHRSHKEGFQDAYIELSKHD

SEQ0192

Homologue of *S. cerevisiae* **SHR3** (YDL212W)

Endoplasmic reticulum packaging chaperone, required for incorporation of amino acid permeases into COPII coated vesicles for transport to the cell surface

S. cerevisiae null mutant: inviable; normal vegetative growth

Chr1-3_0116

5' region

cgtggacctcctgaaaaatattctctctcctggtgaaaaaaaacctctaacgttcactaa
ttagtttcttctccccattaaccaacaacatccacaaat**atg**ccagaagttttgatctcc
aagaaaagaaagctcgtagctgacggtgtcttctacgctgaattgaacgagtacttcact
agagaactcgctgaggacggatactccggtgtcgaagtcagagtcaccccaaccactact
gagatcatcatcagagctacccagaccaactctgtcttgggtgagaacggtagaagaatc
agagaattgactgctttggtccaaaagagattcaagatcgatgagggtaaggttatcctc
tacgctgagagagtccaagaccgtggtttgtccgctgttacccaatgtgagtctttgaga
tacaagctgttgaacggtctggccatcagaagagccgcttacggtgtcgtcagattcgtc
atggactctggtgccaagggtgttgaggttgtcgtttccggtaagctaagagctgccaga
gccaagtctatgaagttcgctgatggtttcatgatccactccggtcaaccagccaaggat
ttcattgagactgccaccagacacgttctgttgagacagggtgttttgggtatcaaggtc
aagatcatgagagaccctgctgctgccagaaccggtcctaaggcccttcctgactttgtc
aagatccacgaaccaaaggatgagagagaggttccagctcctttcgtcaagtcttatgtt
gctgagccagtcgttgaggctgaggaggctgaggagcaagctactgct**taa**actgtttaa
ttctgtaggataaacgaaataactgtatctttataatcatccttaagatagccctggact

cgtcagacattagttcgcgcacattattttcctccactgtctgcaaacagccacacggca
tttccctctctagtttcatttatcaaagtagcacaacatc

ORF

ATGATCAGCTACCAAGAATTGCTTCCTCTGGGAACTGGTCTTATTGTGACAT**(GTAAGTGATTTCCCTGCTTTCA**
ACAAACTATTGATCTAGTACTAACAATTGCTCTTCAG)CCGCATCATTTTGCTTGGGTCTTGCCTATTCCAACTG
GCCTTTTGATTATTTCACTCTATACTCTGACGGAGGGGATGAGGCATTCAAACAGTCATTGGACCATCTCCTTCT
GTGGTCCAAGGTTCCACGTTTCGTACACTGGACTTTACATGGAGTTATTGGTCTCGGTTACATTGGTTGTTTTAT
CAAGATCTTCAAGCCAGACCCAGATCAATCATTGTTCGAATACGGAACTTTGGTTCTATTTGTAGTGGCCACAGT
CATGTATTTGACCAATATTCGTTTGGGTATCTTGAGTGCTGAGAAGGGTGAATGGGGCGACGTGGACGAACAGAC
AGGAATCAACGTCATTGCTGCCTCTACTGGGCTTCTAGTTTTCGTCCTTATCGGTATTCTGCTTTTACAATTTGG
TTTGTGGTACGCAGTACACGCAGACAGGGTACTCAAGGAGAAGTATTTGAAGGAGGAGGCTGCAAAAGAGGCCAA
ACTGGAGAAGGAAAGCGAAAAAACTGAAGATAAGGCAGCAGAAGAATCTGAAGCAAAGCCAGATTCGTACAAATC
TAAGAGTAAGAAAAGCTCCACTTCTGGATCTAAATCAACTGCTACTACCGCTCAGAAGAGAAAGGCTTAA

Downstream

ctgttaaccttattcattattgtattatctatgtttatttagcttaagtcgtgtactctt
cccaatatatttccacatagttttttatccaccaatagcgatcgaatcccattccaagt
cacatcaaactacacatctatcatattgtccgaa**atg**gctattggtgactatccagtgta
ttataagcagcccattagagtaagtcgttaattctgtttaccaagtacttcattttaaac
tccaactaactcggattttcaaagtggctgagttatcatgctcacactcgtccccatgtc
ttttacgctattgctgttgctgcgctaggccctgttttcatgatggccacccctcttaga
agaaagtttttgtacgatgatcatgaacctttgcctgccaatggttacccacttccaaat
agatcaagagacaagacactgacgggttacgacgat**taa**gattagaacaggtattccctt
ttatttatatttattctctatttacagcatatttgataaaa**tta**ccatttgtagaatcca
gaaggttttgatctgatcttctgatcaatcacggaccctggaattttgttgttctttggc
tgttgggatttttcccatgtaattctacgcaacgtttggtcagaacgtagcacgaatttg
acgttgacatgacgatgggtaataataccagttctacccccttcctttacattccactctc
ttcgacttatgtgaatctgagcctacccaggcctcagagatgtacagttcttctgggtta
tatcctaaagtttcagcatcttttaaacctctttctagtaaaatataaagatctgtggca
atcttcttacgcatgaaatgaagctgggtaatggcttcttttaatctaagtcctttaaga
gcacggagcacaatgttagccttttgggagtcgattttattctgtaagactctaaataa
acactgggttctaacgccttaacctcttcgggtgtcaatttc

AA

MISYQELLPLGTGLIVTSASFCLGLAYSNWPFDYFTLYSDGGDEAFKQSLDHLLLWSKVPRFVHWTLHGVIGLGY
IGCFIKIFKPDPDQSLFEYGTLVLFVVATVMYLTNIRLGILSAEKGEWGDVDEQTGINVIAASTGLLVFVLIGIL
LLQFGLWYAVHADRVLKEKYLKEEAAKEAKLEKESEKTEDKAAEESEAKPDSYKSKSKKSSTSGSKSTATTAQKR
KA

SEQ0193

Homologue of *S. cerevisiae* **ZUO1** (YGR285C):

Cytosolic ribosome-associated chaperone that acts, together with Ssz1p and the Ssb proteins, as a chaperone for nascent polypeptide chains; contains a DnaJ domain and functions as a J-protein partner for Ssb1p and Ssb2p

S. cerevisiae null mutant: cold sensitivity: increased; gamma ray resistance: decreased; ionic stress resistance: decreased; resistance to daunorubicin: decreased; resistance to camptothecin: decreased; resistance to N-Methyl-N'-Nitro-N-Nitrosoguanidine (MNNG): decreased; resistance to gentamycins: decreased; resistance to paromomycin: decreased; resistance to cisplatin: decreased; acid pH resistance: decreased; budding pattern: abnormal; cell shape: abnormal; competitive fitness: decreased; fermentative growth rate: decreased; metal resistance: decreased; resistance to gentamycin C1: decreased; resistance to doxorubicin: decreased; resistance to calcium dichloride: decreased; resistance to rapamycin: decreased; resistance to cisplatin: decreased; sporulation: decreased; viable

Chr2-1_0282

5' region

tgcaagtatctatctgtttataatcaatatggttttaatttatctaaaaaatactttgac
aaatccttctttgttgacgaggtataagtactgttgctcgtttgtttctaagtttttcgc
ctcatcttcgaacgatattgatcccttcgaaactgttattgtgcctgttacgtcgtctgc

tcgtcccgtctccaagggtcgtatcgatattatcccctggcttctatgatacaatttggt
gagaaactcagtttgtctatgtgtaatggttgaagagtctgtttcattcaggtctatcag
atttgcatcagcaaaagtcgaaacaagtacgttcgccaatgagttgatataagatatttg
gctcatgaattcgttgatagaacattgatttgtcaaaaagtatagttctggatgatccag
aattaatgtgcagggtaaggctgagctcttcaaagcagcaattacaggcttgaataactt
gtccacaggagcctctgccgttgaaagatcgaaaaaactaaggttttcgtcaccttgaaa
tgaagattccaccccaagatattttacccacgtcttgttgtagaagcttatcccatgcaa
gaagctgacaatgacaactctgcttctgggatttttgagtgaaaagttggaagttttgaa
ggaatttgcatccttattgatattacatgtgccaaaaattagattctccacaaagcaagt
ggtcagccatgatgacgacgttccttgagtgtccgtgtaggtaaaacatgatctaaaagc
ttgtgagtcgatgagagcttgcggtatgacctcccctgacctgaaaagagagaaatcttg
agcttggtggga**cat**gatggatttggcccttgattaggttggattctcgttattgtccta
aatttgatatttcctcgaactgtggcacgtgattagttggaaaaaaaaatcatcaataat
ccagaaaatctttcccttcactagaatacctatagttaac

ORF

ATGTCCTTGACTTTGCCTGCATTGCCATCTAGCGTTTCAACTGCTGCCAGTTTAGGTAAATACACTGCCGCTGTT
AGAAAACCAATTGAGCCAGTTGGCCGTCACTTCTTGGCACTTTATTTGTGGGGACACACATGGTCAGAATTCGAA
AAGATTCAGGCTGAAACTGATGTCAAGAAAGCTGAAGACGTCAACGAAGATGCCGAATTTGACATTGAGCAAAGC
GAAGAGCTTATTGAGCACGACCCAAGAGACTGGAAGTCTGCTGATTTGTACGCCGTTTTGGGTCTGAACAACTAC
CGTTACAAAGCCACAGAAGACCAGATTAGAAGAGCCTACAGAGTCCAGGTTTTGAAACATCATCCTGATAAGATG
GGTGAAAAGGGAGGCCTGGATAAGGATGGATTTTTCAAGATCATCCAGAAAGCTCACGAAACTTTGATGGACCCT
GTTAAGAGGCAACAATTTGATTCTGTTGATGAAGAGGCCAACATTGAACCACCTGCACCAAAGAGCAAGTACGAC
TTTTTCGAAGCTTGGGGTCCACTTTTCGCTTCTGAGGGACGTTTTTCTAAGAAGCAGCCCGTTCCTGAACTAGGA
AATGCTGATACTCCAAAGGAAGAGGTTGATGCTTTTTACAAGTTTTTCTCCAACTTAGACTCATGGAGAACCTTC
GAATTTTTGGATGAGGACGTTCCAGATGATACTTCCAATAGATACCACAAGAGAATGATTGAGAAGAAGAATAAA
GCTGCTAGACAGAAGAGGAAGACAGAGGATAACAAGCGTCTTTCTGAATTGGTTAGAAGAGCTGCTAATGAGGAT
CCCCGTCTTAAGGCCTTCAGAGATAAAGAGAAGGAAGAGAAGAACAAGAAAAAGCAGGAGAGGGAGGCTGAAGCT
GCTAAGGCCGCTGAAGAAGTTAAAGCCAAGAGGGAGGCTGTTGAAAGGGCTAGAGCTGAAGCCGCCAAGGCTAGA
GCCAAGTCCAAGGCCAAACCTTCCTATCAACCAATCCAAACTCAAAGCAAAGGTAAAGGTAAGGGTAAAGGTAAA
GGTAAAGGAAAGAAGAGTACTGTCAGCAAAGAACAGCGAACTGAATCCAAAATCAACATTAGAAACTCTGTTAAA
GATGCTAACTACTTTGGTACTGAGCTGCAATCTGAGATTGACCATGATATTAATGTCTTCCTTGATAACTTTGAG
GATGAAAAGTCTGTGGAAGTTTCTAATGCCATCGCTGGTGCCACTGCTGATGCTGTAAAGGAGATCTTCGATGCT
GCTGCCAAAGAGTACATTGAGCAAGGTTTCATTACAGCTGACGACTTGAAGTTCTTTGCAAAATAA

Downstream

atagataggcaacattatttttct**tca**ccagaatcgataccatgatttttgctgtcttt
gcgttgtaaatccaaatgctcgtctataagttgttgtatatccctttcgagtttcgaaac
cagtatttcattttcttggataattttgtctctttccaagtcccaccacaagtactgtag
gccgaagtagactagtgatgccaataggaaagtattcaacagaggccttgaaaaatgtga
tattctgactttggaatcttccttctgttgctccttttttttctccaaggatttcagttc
ctgttgtttgatctggtcgtttatctggtcgattcttgattgaataatgtgtcgttggga
aggatattgtagcggatcttgagggacatattgagtagtagattttggtaagtgtgtgct
ctgttttcgtatctgtctccagaccagacttgatttacaaatttgcggcct**cat**ggtgaa
atttgcttgtagggaagctgaccactttcagagactaacgattgaaagaccttggatgga
ctccacgtagatattccatgtgatttaggtttcgcggttattatatattggtgatacgtc
aacaaccaagcctaagtatatatatacagtgtacactattatcaatcaccactgttcttt
tgctcc**tca**tttgacgttcaaaacctcctgtagattgaacctcttgcatttgtcatatat
ctttccgtccaacactgacatgctgatattatttttgaaacactcaaccaatatcagtac
atcgtaatcttttaggtcaaccttgtggcctccctcgttaccaacgcaactggccacggt
tttgatgacatccagcttctctagcacgctgaagtttctcctattgacttgaatggcaaa
tttgagtggtttctgaacctctttgtcttgatgaaaatggggttttaacaatcgttgtgc
caacttggttaactctggtaaagtgcctgagcatgaacca

AA

MSLTLPALPSSVSTAASLGKYTAAVRKPIEPVGRHFLALYLWGHTWSEFEKIQAETDVKKAEDVNEDAEFDIEQS
EELIEHDPRDWKSADLYAVLGLNNYRYKATEDQIRRAYRVQVLKHHPDKMGEKGGLDKDGFFKIIQKAHETLMDP
VKRQQFDSVDEEANIEPPAPKSKYDFFEAWGPLFASEGRFSKKQPVPELGNADTPKEEVDAFYKFFSNLDSWRTF
EFLDEDVPDDTSNRYHKRMIEKKNKAARQKRKTEDNKRLSELVRRAANEDPRLKAFRDKEKEEKNKKKQEREAEA
AKAAEEVKAKREAVERARAEAAKARAKSKAKPSYQPIQTQSKGKGKGKGKGKGKKSTVSKEQRTESKINIRNSVK
DANYFGTELQSEIDHDINVFLDNFEDEKSVEVSNAIAGATADAVKEIFDAAAKEYIEQGFITADDLKFFAK

SEQ0194

Homologue of **CPR6** (YLR216C):
Peptidyl-prolyl cis-trans isomerase (cyclophilin), catalyzes the cis-trans isomerization of peptide bonds N-terminal to proline residues; binds to Hsp82p and contributes to chaperone activity
S. cerevisiae null mutant:viable

chr3_0567
5' region

caagagctttcaaagtggcatcatcagcttgctcaagcaacgacttttgataatatataga
tcgtgattccagctaatccaatatttttttgagcaccggccattatcaatccatatttac
tgacatccagcctctttgataagatgttggatgacatatctgcaactacttcaacaccag
gaaaactttcaaatgggaactccttgaactcaacaccgttcacagtttcattatcacaat
aatacacataagaagtgtttgatagatcagtggggatgctccatttgtcaactgggggta
tggatccaaattttccatcaacttttttactatccaaaacaatgtcaacgtccaaaccca
aacgctgggcttcttcaacagatttcttagaccaagttccagtgattgcataagcagcac
ggcccttccttccagttttcttaacaaaggaggcagtcaagttggaagcaatagaagaaa
atcctgttgttcctccccccttgtaagaaaaaaacctcatgtgtatcagggatgttcaaca
aactaactagattggccttcgtgttatcaatgacttcactagcttgcttagaacgatgag
aaatttcaccaatacctataccaacaccctggtagttgaccaagtcataggcagcctctt
gcaatactgaagtaggcagtagcgctggcccagccccaaaataatgtggttcttctcttt
ccaaagtttttgc**cat**tgtattatatggttagttcaagatgaagaagtaagatgagacga
agagtcaaaactggcaattgcttagagtcagatgttggcgcgattcagttgatctaggag
gatagccacgcagaggtcgcacccttcgaaactgccttacaatgtgccctgtgtctcaga
aatttctagacaatcgcgccttacttttcaaacaaggggaccgaatccgtataaatcaga
cacccgtccatagtgcccccccatccaatccttttcagta

ORF

ATGACTCCCCGTTCTCATATTTTCTTTGACATCTCCATCAACAACCAGCCAGCTGGCAGAATAATCTTTGAGCTC
TTCAATGACATTGTTCCTAAGACAGCAGAGAATTTTAGAGCTTTATCTACTGGTGAGAAAGGTATAGGTAAGTCT
GGGAAACCATTGCACTACAAGGGTTCTACTTTCCATAGGATCATTAAGGATTTTATGGTACAAGGTGGTGACTTT
ACCAACGGTAACGGTACTGGAGGCGAATCCATATATGGAGAAAAATTTGAAGATGAAAATTTCCAATTGACTCAT
GACAAACCGTTCCTTCTCTCTATGGCAAACGCTGGACCAGGAACTAATGGATCCCAGTTTTTTATCACCACCGTT
CCTACTCCTCATCTGGATAACAAGCATGTAGTGTTTGGAAAAGTAATTGCTGGTAAAGCCACAGTTAGAAAGATT
GAAAGAAACTCCGAAGGTGAAGCTCCAATTGAACCCGTTGTCATTGAGGACTGTGGTGAACTTCCAGAAGACGCA
GATTTGACCATCTCCGACGAGACTGGAGACAAGTATGAGGAAGTTCTGAAAGATAATGAGAACATAGACATCGAT
GACTTTGAACAGGTCTACCAGGCCATCACTGAAATCAAAGAATTGGGTACAAAGTATTTCAAAAATGGTGACACC
AAAATCGCCTTCGAAAAGTATCAAAAGGCTGCTAATTATTTGCTGGAATACATACCATCAGATTTATCAGAGGAA
CAGAGCTCTAAGTTGGAGCTGCTAAAAACATCTGTCTTCTCCAACGTGGCATTGGCTGGACTGAAAGTTTCCAAG
TTCAAAGACACGATTAAGTATGCCACATTGGTCATTGAGGATGAATCTGCGGATGCAAAGGCCAAGTCCAAAGGC
TACTACCGTAGAGGAAGTGCTTACAGCTCACTGAAAGACGAAGATTCAGCCATCTCAGATTTCCAGAAAGCACTT
GAATTATCCCCAGGTGATCCTGCAATTAGCCAATCTCTACAAAGAACCACGAAGGCCAGAAAAGACCGTCTTGCC
AAAGAGAAAGCTGCTTTGTCTAAGTTCTTTGAGTAG

Downstream

agtcatataaacacatagacaactaataattcatcccagcgatggcacgtaaacatcaag
acaggttagatagtacccgcgcatccgcgtatcgtggttccacctaccgtcccgaagcgt
cgtagcacctccccccaaccctggtgcgtaattatctggccatctccgtaatttcgcgcc
cctctcttttgcctccgacgaaaaagcatgaaaaacaacgagaacaccctacacacaagc
ctctccagctggcaaccatcttcctgagaatcaatctgattagacattgtattaacagtt
atattaatccagttgtcttaacgtcttattgtgtgaaaatagaattggcactcatttggt
tgttgttgttaatcagctcattagacgtgaatttgatttgaaagggattttgttaatttt
tgctaaacccataatttccacttgcaagtgatacaagagtaactgacacatctagttaac
ccccgtaaactataatagttcccaaacaaaaacagaaaaaaaaagcttagaaattagcaa
gat**atg**agtcagtttccttattcaagtgccccactgaggtccgtcaaggaagtgcagttt
ggcttgttgtctcccgaggaaattagagctatttcagttgttaagattgaatatcctgaa
atcatggatgagagtcgccaaagaccgagggaaggtggattaaatgatcccaagttgggt
tccatagaccgtaatttcaaatgtcaaacttgtggtgagggaatggcagaatgtcccggt
cattttggacacatggagctagctaagccagtctttcatattggtttcattccgaagatt
aaaaaagtctgcgaatgtatctgtatgaactgtggaaagcttcttcttgacgaaacaaac
cctaccatggctcaggctatccgtattagagatccaaaaaaacgatttaacgctgtgtgg
cagctgtgtaaaactaagatggtatgtgaagctgatgccc

AA

MTPRSHIFFDISINNQPAGRIIFELFNDIVPKTAENFRALSTGEKGIGKSGKPLHYKGSTFHRIIKDFMVQGGDF
TNGNGTGGESIYGEKFEDENFQLTHDKPFLLSMANAGPGTNGSQFFITTVPTPHLDNKHVVFGKVIAGKATVRKI
ERNSEGEAPIEPVVIEDCGELPEDADLTISDETGDKYEEVLKDNENIDIDDFEQVYQAITEIKELGTKYFKNGDT
KIAFEKYQKAANYLLEYIPSDLSEEQSSKLELLKTSVFSNVALAGLKVSKFKDTIKYATLVIEDESADAKAKSKG
YYRRGSAYSSLKDEDSAISDFQKALELSPGDPAISQSLQRTTKARKDRLAKEKAALSKFFE

SEQ0195

Homologue of **CPR7** (YJR032W):

Peptidyl-prolyl cis-trans isomerase (cyclophilin), catalyzes the cis-trans isomerization of peptide bonds N-terminal to proline residues; binds to Hsp82p and contributes to chaperone activity

S. cerevisiae null mutant: growth rate in exponential phase: decreased; viable; competitive fitness: decreased; mitochondrial genome maintenance: abnormal; resistance to amiodarone: decreased; resistance to hygromycin B: decreased; respiratory growth rate: decreased; transposable element transposition: decreased

chr3_0378

5' region

ctgttcctgtgacaaatcatgcttggccatcaagtccaacggcaggggaatatgattttg
tgttgtagcataatactttaatcctatcaacatagtggcaacccctgtcgcttgacctag
atgggccgccacttcagttaaatttgtctctaggaacggggcctgctccaatagttcaac
agcagatggggagatgttaggtgataatagtaatcgctgtgttaaataattgctctgtga
ataaatgccttcgccataggaacaaacttcatccatgttcttgaaccctaacttagaagt
gttcttcaagtaatgcaatctagatgtcaatatctgttgaaacatgccaatgtccaagtt
taactctttgttgataccatcccgtaacaaagtagccaccggttcacccaaaggcttctg
gtacccgcctagaaacacttttgaagtaattcctcccaaaatttgaatttgacgtcaag
taggccaacccccacattagctcggaattgactgttggctcttgtattggaatctgatat
cttgttcacttccaatagaaagcatcgtatcgccatccaggtatctcttgcaggttctgg
aatgtaatgagccaataggtaacttgacttgtcaaacttgttcaacatctcggtgcaatc
tgtccttgctttcatcacagattgtctatattgtagatctggtggggacgtgaactgttg
gtggacagcatgaacattcagcctccttgcagcggcgtgaaaccttctccttcccaa**cat**
gctaggatatttctgggtaggaatgtctagcgttgataggtaatcaagtaaaatatactc
aagacggagaaacgagtctctccgaagttgcttacttttcgcttccttactggctagagt
ctggaaagaaatccgatacgggggctaccaggaaatcttgcgcgagctgaaggatcatc
tctctctacaaggattgtagattgaagggtttaatcgcaa

ORF

ATGTTTGCCTTTCTAGATGTCACTATCGGAGGCAGCAAGATTGGCAGAATAGTTTTGCAGCTTGACAATGTTAAC
TGTCCATTGGCTTCCACTAATTTTTTGAATTTGTGCAACCAAGTGGGAAAGACTTTTCATTCGAAAGAGTATGGA
GATATAAAGTTAACCTATCAAGGAATAAGGTTCCACAGAATCATGAAAAATTTTATTGTCCAGGCCGGTGATCTT
GTTTTTGGCAGGGACGATTCATTCAATCAGGAACAGGTGGGTGGTGGAGGGGTTTCCACTTACATTAATGATTTG
AGTAACAACGAAAAGGGCCCTATCTATGGAAAGTTTGATGACGAAGCGCTGGACAATAATCCAGTAAGTAAATTT
GACCGTCCTTTCGTTTTGGCAATGGCCAACAGCGGCCCCAATAGTAACAGTTCTCAGTTTTTCATTACAACTTAC
CCCTGCCCACATCTTCAAGGGAAGCACACCGCATTTGGATCTGTAGTTCATGGGAAATCTGTAGTGAGAGACATT
GAAAAAGTTGACACGACCAAGGAAAACAACAGCCCGCTTCAGCCAGTGGTCATAGTGGCGTGTGGAGAGTGGAAA
GAAGGAGATCCTGTTCCTGTTTTCAATTGCTCTAACAGTCAAATCGGGGGAGATGTCTTTGAGGAATATCCTGAC
GATGACGACAATTTTGACAAAGATTCTCTTGAAAAGGCATTGGAAGCTGTCGAGATAATAAAAGAATCTGGAAGT
CTGCTTTTCAAACAAAATCAGATCCAAAATGCTTTATTCAAGTACAAAAAGTCTTATCGTTATCTCAATGAATTC
TCCCCCGACCAAGATCAAAGCCCCCAGTTATACAACAGGTTCCTAGCATTGAAAAAGAAATTGTACTTGAACCTT
AGTCTTGTTTATTACCGACTAGGACAGTTCTCTAGAGCGAGAGATTATGCCGAATATTTATTGGAGATGGAGGAC
AACCAAGCGATAACTGATAAGGATAGAGCTAAAGCATTTTTCAGAAAAGGATTGGCTCAAATTGCATTGAATGAT
TTTGAGAATGGTTTACACGCTCTGAAAGCTGCTCAAGACTTGACTCCTACTGATTCTATGATCTCAAATGAAATT
TTGAAGGCATCGAAATCTTTTGAACTATTCAAGAAGAGAGAGAAAAATAAACTGGAAGCTTTTTTTAAGTAA

Downstream

gatttattggttgaaaaggatttcagaaagaaacgcacttttggcattggttaaactat
aggtcctctgtttgctgaggttataaagaacgtgcagcgggctgatcgcttggattgttg
aagcattaattagtctgtggactagaggattctagctgttaaagatcttttctgtgtctc
gcattaatttctaagcagaagcggtggattacaaatgcaaaatccgaaaaattgtcgatt

gggctccattcgtatctaatcagtgctattcctgttctgatatgtaccgcttaccgaaaa
agccactactagaggtggctattgtttgcacatgacaaagtcacagaaatgtccacagat
aatctttgtctagatatagcgtaggagcatcaaataagctgggaatgggcatcgctaagc
cactattcacgatgcacgaattgaccgggctttcagtgttgattactaagtttaagttta
ggtttcaagtgcatgcaaatagacttcatgcagggaattaaagttacatatgcgccatct
ttaaacggtttacgatttggccaagtaggaggaaatcgaacagtgtgcaaatttttattt
acagccaacgtctttcatcattctttcaataagggagcttagtatatgtcatagttaaca
ctcagtatccttggaaaatagatcaagctctgcatttttttcaattgattaataatacta
aaggacgcacctaagataatactaacagcttttgattactgcacgcgcccgcgcacagtc
tattaacaaaataacgcaactgaaaggcattattcagatcaccttattgttaaggtcgac
cccctgaattattgtggaagtatcgcctgaaacgaagcctcataggttgttccgtttgaa
cttggagtctctatctctagccatcatcgccctttgatctcccaaatcatggacctcggt
aaaaagatttattacgtaatcattttattttgggaacgtg

AA

MFAFLDVTIGGSKIGRIVLQLDNVNCPLASTNFLNLCNQVGKTFHSKEYGDIKLTYQGIRFHRIMKNFIVQAGDL
VFGRDDSFNQEQVGGGGVSTYINDLSNNEKGPIYGKFDDEALDNNPVSKFDRPFVLAMANSGPNSNSSQFFITTY
PCPHLQGKHTAFGSVVHGKSVVRDIEKVDTTKENNSPLQPVVIVACGEWKEGDPVPVFNCSNSQIGGDVFEEYPD
DDDNFDKDSLEKALEAVEIIKESGSLLFKQNQIQNALFKYKKSYRYLNEFSPDQDQSPQLYNRFLALKKKLYLNL
SLVYYRLGQFSRARDYAEYLLEMEDNQAITDKDRAKAFFRKGLAQIALNDFENGLHALKAAQDLTPTDSMISNEI
LKASKSFELFKKREKNKLEAFFK

SEQ0196

Homologue of **PIH1** (YHR034C): Protein of unresolved function; may function in protein folding and/or rRNA processing, interacts with a chaperone (Hsp82p), two chromatin remodeling factors (Rvb1p, Rvb2p) and two rRNA processing factors (Rrp43p, Nop58p)
S. cerevisiae null mutant: heat sensitivity: increased; resistance to 5-fluorouracil: decreased; competitive fitness: decreased; fermentative growth rate: decreased; resistance to rapamycin: decreased; resistance to 5-fluorouracil: decreased; resistance to wortmannin: increased; viable

Chr3_0296-1

5' region (in bold start codon next gene)

cttccacattaagcatcaaatttgacttcacttcattcatgaaagtccataattgtgagc
tcatagaattctgtgttttattcgattcgttagttgcgtgagagttgggtgcggataact
cgtccatgagatttagtttctcaaattcatgttgtatcaatgtatcaaactgttggagat
agctcaacggtttggataaccaggaatttggggtatccttttgcttgccatctaccgatt
ttaaaccttcattaggattcgttgtccttaagggatctttgtgcacacctacttccacat
ctagtttgtggggtctttggctagatgaaggcgatctcgttacgactggaggtggatttg
aggatatttttccacactttggacctggtcatgtacttgcttgtataaggatctttcaa
tttcagttttaagcttgttaacgcggtgttctttttgtctctttttctctagtaatgcct
taatttgttccttgatctccaggatctccatttccgttgtggccaaaactctcaatttat
ccgacaaatgattatcattattcgaactcaaagaaggagaatcaactgaggaattgccac
tactttctttccttggtgtggatgatgcctttgtaactttggaggagtttggagcagggc
ttcgacggcgggtctcaaggactgatattcccagattcaagccggatttgcgcggcgaat
ttgtagatgatgtctgaaattcatc**cat**tactgattctgctatcaaataaaagacgcaat
cttttggtaagaatagagaaggatgacctactctctacagggttgctggaactagtagtt
tcgtttctgtaaatcctatgaaattgatgtctatcgcgcgacttacgatgcgcgcgtgtc
aaactgtactagtcatctctgcgatcagtatcaattcaagcccatctttgtttccaccct
gtaaatattccggtaatagagactgataccttgcttttca

ORF

ATGGTCAGTCAGAAAATTACCCCAAAACCTCACTTGGTCCTTAAAACAAAACTTTTGGCT
CCAATAAGAGGCCTTCCAGTGGGAACCAAACTCTTTATCAACGTTTGCACTAATGATCTA
GTTCCGTTGCCACCGCCTTTAACACCTTTCACCCCAGCTGAGGATGATGTGTTTGATTTG
TCGGTGATTTTACCAAAGGTTATGGATGATAAGTGGGAAATCCCCATACTTACATCTCCA
GACCTAAGAATGGATAAAGACAAGTCAGGAAACCTTTCATTGGTAACAGATTGCGTAATC
AATGACAAACCTATGAATTGGTGTCTAATAAGTAAGGACTTGATGGACATTCTTATTGAA
TGGTGTTTTGAGTCCTTGGAATTTCAATATGATTCAGAGGTGTTGGTGGTTGATCATATC

AAGTACTCAATTCCTAAAATGTCTTATAAAGGCAAAGAAGTACAGGAAATGGAAATTGAT
CTTGAATCAAACGAACTACAACGTAAGGAGTTTGAGCGAATAAAGACAACCCTAACCAAG
AACGAACCACTTGGTATTTTAGAATCGAAGAGGGCTCAAGATGAAGACGATCTTGATACG
GATAAACAGGAATTAACCTCATTAGATAGTCTATTGGCCCGTCCAGTAGCAAAACCAGTT
GGAAATGGACCTTTGATTCAAGAGATCAATGACATGACAATTTCACAATCTCGCCCCAAA
CCAAAACTTGAAACTTCACAGAACGCTACAAAATCCCCTCTCCAGAAATCTAAACCAAAA
CTGTCGCTCGACATAAAAGTTGTGAAGCTATCTCCCGAGGATACTGAAACCAAATACAAG
CTTCTGCTACAAATCTCATCCAATCTCACTGACAAATCAGCTTACAAACTTGATCTAAAT
AAGGAGAACTCTACCCTGTTAATATCTACGACCGATTCAAAGTATGACTTTTCCAAGTCT
CAGGTTCATTTTCCTCTACCTATCGATACAGAGCAAATAAGGGTGATTAATGTAAACTCG
GAGGCTTTTCTACGTGTCTACATCAAGTAG

downstream (in bold start codon previous gene)

actctgatctaagaccatatccataagatagtgtagcctattacccctttcggacttaaa
cgagactaaaatttttgaaaagaaaaaaagaatcttttcttacatatatacagccagaca
aataaaatccaaaggattgatgtattatcttgatatagatttgactctttgaaatctttt
cacacatagcaatttttttattcatacattggcg**atg**tcggatacggatggtttatctttg
actccgacctcaagcagaaaatacggcagcacgaatgatactgaggcaagtatagataat
gagattctggtccaacttccagtgactacgttcaaatctgaggctaaaatctactacgaa
tcggcttttcctttagtaatatcatttttcttacagtatctattggctgccactcctata
tttattgtagggcatattggagctgttgagttgggatctgtaagtttagctatcatgact
ttcaatattacgggtgcaggtttggttcaaggatgcgtcactgctttagacacctttgcc
gcccaagcttatggagcacaaaaatatcatttagtgggagagttttgtcaaaaagcttgg
atgttgataaatattctttccattccagtgctcatgacttggtggttcatggaacctatc
ctcggatttttacaaccagatagacagatagccatgctagccacacagtatttgcgaatt
ttaattttgggattaccagcagttattatatttgaagttgcaaaacgatttttgcaatgc
cagaaaatattcgacgcttccaccaaagtcttatttttttggggcgccattcggttttttt
atttcatatctactggttttcaagtttgagcttggatatattggagctcctattgcagtc
gttatttcgtattggacaattgccattttactcatattctacgtgatcctcatcgatgga
aaacaatgctggggaggattttgtaagaatgctcttttcca

AA

MVSQKITPKPHLVLKTKLLAPIRGLPVGTKLFINVCTNDLVPLPPPLTPFTPAEDDVFDL
SVILPKVMDDKWEIPILTSPDLRMDKDKSGNLSLVTDCVINDKPMNWCLISKDLMDILIE
WCFESLEFQYDSEVLVVDHIKYSIPKMSYKGKEVQEMEIDLESNELQRKEFERIKTTLTK
NEPLGILESKRAQDEDDLDTDKQELTSLDSLLARPVAKPVGNGPLIQEINDMTISQSRPK
PKLETSQNATKSPLQKSKPKLSLDIKVVKLSPEDTETKYKLLLQISSNLTDKSAYKLDLN
KENSTLLISTTDSKYDFSKSQVHFPLPIDTEQIRVINVNSEAFLRVYIK

Homologue of **SSB1** (YDL229W) and **SSB2** (YNL209W):

SSB1: Cytoplasmic ATPase that is a ribosome-associated molecular chaperone, functions with J-protein partner Zuo1p; may be involved in folding of newly-made polypeptide chains; member of the HSP70 family; interacts with phosphatase subunit Reg1p

S. cerevisiae null mutant: viable, petite-negative

S. cerevisiae overexpression mutant: increased growth rate in exponential phase; increased resistance to FK506.

SSB2: Cytoplasmic ATPase that is a ribosome-associated molecular chaperone, functions with J-protein partner Zuo1p; may be involved in the folding of newly-synthesized polypeptide chains; member of the HSP70 family; homolog of SSB1

SEQ0197

chr3_0731 (*P. pastoris* known gene)

5' region

cttcataccacaactttgcgacgcctgaggattgtagaaaatggcgtagaaaggccgcca
gcctttctgactgttcccgtccttgtcaaagaacaagaagttcacagatggttagattct
ttgatggccacttatcagatccagatgcgaacaatatcagagctctctcaattccataca
gaggatgcgaactcattcatttccatcatacaattgaccattgatggcatctggttgatg
gctctggataaagaatggcttaacagcaggttgacagttgaaggagttttaaccgctggt

agtcacgatagacacagccaattacgctttcatctcccatacaatgtagtaatttcctca
attactcttctggacaagctcactctggcagattgccaagaggttctctatcatgtattc
aaacgagcatcttctactgatactaagactcgaagctctttccaaaacttaacaagatcc
tcagaattgtacaactgcaagaaggaactccctttacaaccattaagccaggtggatata
ttggtaacgcacaatttatggctcacaaaaaggcagacttcttatctaattggctgtaaa
ggttccagaattgaactcatacgaactaactccaaggctaaaatcaagatttttagtgca
ctagaagaggaattaatgcctaagcaacatatttctataactggaactaagtctcaagtt
cagaaagctgtctttctgattcagttttccttggacaaactacttcagggaatcagtgac
tggtgagactatttattggtgagcaaaactagaatgggctataaattccatgccacagcc
gatgaaccgggtaacctttttagttttacccacaatttactgtttcgctaacctagaaaaa
aaacctctagtggttggtgaaaaatttttttaggagagccttttttttttcttgtttcctc
ttctttctcaaagaaactttcaagttttcgcatatatata

ORF

ATGGCTGACGGTGTTTTCCAAGGTGCTATTGGTATCGATTTAGGTACCACCTATTCATGTGTTGCTACTTACGAC
TCCGCAGTGGAGATCATTGCCAACGAGCAAGGTAACCGTGTTACTCCTTCCTTCGTTGCTTTCACTCCTGAAGAG
AGATTGATTGGTGACGCTGCTAAGAACCAAGCTGCTTTAAACCCAAAGAACACTGTTTTTGACGCTAAGCGTTTG
ATTGGTAGAGCCTTTGATGACGAGTCTGTCCAAAAGGACATCAAGTCTTGGCCATTCAAGGTTGTCAACGACAAC
GGTAACCCATTGATTGAGGTTGAGTACTTGGGTGAGACCAAGCAATTCTCTCCTCAGGAAATTTCGTCCATGGTT
TTGACCAAGATGAAGGAGGTCGCTGAGGCTAAGATCGGTCAAAAGGTCGAGAAGGCTGTTGTCACCGTCCCAGCT
TACTTCAACGATGCTCAAAGACAAGCTACTAAAGACGCTGGTGCCATTTCTGGTCTGAACGTTTTGCGTATCATC
AACGAACCTACTGCCGCCGCCATTGCTTACGGTCTTGGTGCTGGTAAGTCTGAGGAGGAGAAGCACGTTTTGATT
TTCGATTTGGGTGGAGGAACTTTCGATGTTTCCCTGTTGCACATTGCTGGTGGTGTCTTCACTGTTAAGGCCACT
GCTGGTGACACCCACTTGGGTGGTCAGGACTTTGACACCAACTTGTTGGAGTTCTTCAAGAAGGAATTCCAGAAG
AAGACCGGTAAGGACATCTCCGACGATGCCAGAGCCTTGAGAAGATTGAGAACTGCTTGTGAGCGTGCTAAGAGA
ACTTTGTCTTCTGTTGCTCAGACCACGGTTGAAGTCGACTCCTTGTTCGATGGTGAAGACTTCACTGCTGAAATC
TCCAGAGCTAAATTCGAGGCTATCAACGCTGATTTGTTCAAGTCTACTTTGGAGCCTGTTGAACAAGTTTTGAAG
GACTCAAAGATCGAGAAATCCAAGGTCGATGACGTCGTCTTGGTTGGTGGTTCCACCAGAATTCCAAAGGTCCAA
AAGCTGTTGTCTGACTTCTTCGATGGTAAGCAATTGGAGAAATCCATCAACCCTGATGAGGCTGTCGCTTACGGT
GCCGCCGTTCAAGGTGCTATTTTGACCGGTCAATCTACCTCTGAAGAGACCAAGGACTTGTTGTTGTTGGATGTT
ATCCCTCTTTCTTTGGGTGTTGCCATGCAAGGTAACGTCTTTGCCCCAGTTGTTCCAAGAAACACCACTGTCCCA
ACTATCAAGAGAAGAACTTTCACCACTGTTGACGACCACCAAACCACTGTTCAATTCCCAGTTTACCAGGGTGAG
CGTGTCAACTGTTCTGAGAACACCTTGTTGGGTGAGTTCGACTTGAAGAACATTCCTCCTATGTCTGCTGGTGAG
CCAGTGTTGGAGGCTATCTTTGAGATTGATGCCAACGGTATCTTGAAGGTTACCGCTGTCGAGAAGTCTACCGGA
AGATCTGCCAACATCACTATTTCTAACTCTATTGGTAGATTGTCTTCTTCCGAGATCGAAAAGATGATCAACGAT
GCTGACAAGTTCAAGAAGGCTGACGAGGACTTCGCTAACCGTCACGAATCCAAACAGAAGTTGGAGGCTTACGTT
TCCTCCATTGAGTCTACCATCACTGACCCAATTTTGTCTTCCAAGCTAAAGAGAAGTGCTAAGGACAAGATCGAG
TCTGCTTTGTCTGACGCTTTGGCTGCTTTGGAGCTAGAAGATGCTTCCGGTGATGACTTCAGAAAAGCTGAATTG
GCATTGAAACGTGTGGTTACCAAGGCTATGGCAACCCGTTAA

Downstream

gtttattagtaaaatttagggtacgcaaagttaatcttacagtatatatgtgctgttact
tcttaaatgtaaatcattatttttattaaccgatatagttatagccatccgcagcaacga
ttgtcccttcctggatacatttgaacacccactctttctctactatccggggtactttat
ctagaacacctttgacgattggtagatcatcacccagctttttttcttagagacccaactt
catcctgtatggtcgcctcctcagtcaaccaggaatagttgaatgtggtggaagtatcca
cgattataaatgaacattggaaaatatcagttgtaagaaatcctcctagaatcactatta
cttctaacaagtatggactcataacatttgtcagaaatggagatctatagcatggtcttg
cgatataaaacttcaggcccttgaacaagttgtacagtagtggaaattttagtgggtgtt
ctaactttatcttaggtatacctttcataagcattcgcaactccttacggtcaatgggag
aggtataactatccccaaatcggtcggagtttttctttaccttttctatcaaagactcat
ccgccagtaatatgtcgttaatttgtagaggaattatagttttcagctgaagacacttca
gaatgaagcctggtcgaaagacatttagacctttattaaagtactctgtgactttgaaag
taatcttgtcactaagtaaaacaagcctctcgccatctgcaagagctacgcaatcattct
tagttattctgccaccattctcgaccagcaaagcctccagctgagcaatagagtatcgtt
tttcacccaacatgtcacttaaaacataaaagatgtaatccttgaatatgctgctcttcg
aagatacactttgtttagcgagaattggttcagtgactattgtcctctttcgtttactca
ttagactgtgtgttttctctgttttgtgtctactctttga

AA

MADGVFQGAIGIDLGTTYSCVATYDSAVEIIANEQGNRVTPSFVAFTPEERLIGDAAKNQAALNPKNTVFDAKRL
IGRAFDDESVQKDIKSWPFKVVNDNGNPLIEVEYLGETKQFSPQEISSMVLTKMKEVAEAKIGQKVEKAVVTVPA
YFNDAQRQATKDAGAISGLNVLRIINEPTAAAIAYGLGAGKSEEEKHVLIFDLGGGTFDVSLLHIAGGVFTVKAT
AGDTHLGGQDFDTNLLEFFKKEFQKKTGKDISDDARALRRLRTACERAKRTLSSVAQTTVEVDSLFDGEDFTAEI

SRAKFEAINADLFKSTLEPVEQVLKDSKIEKSKVDDVVLVGGSTRIPKVQKLLSDFFDGKQLEKSINPDEAVAYG
AAVQGAILTGQSTSEETKDLLLLDVIPLSLGVAMQGNVFAPVVPRNTTVPTIKRRTFTTVDDHQTTVQFPVYQGE
RVNCSENTLLGEFDLKNIPPMSAGEPVLEAIFEIDANGILKVTAVEKSTGRSANITISNSIGRLSSSEIEKMIND
ADKFKKADEDFANRHESKQKLEAYVSSIESTITDPILSSKLKRSAKDKIESALSDALAALELEDASGDDFRKAEL
ALKRVVTKAMATR

SEQ0198

chr4_0552

5' region

cgttcttctacaacgctgtatttgtgggcgctgtcgagtctttcggtgtccttttgttcc
aaagctataccaacaagactttgtcgttgaaaccactggacctttttggctttggagaata
ttcagtactttgccattgctgctttgttcttgactatcctcccccatattgcaattgctg
ttgttccatttaccttgttcagtgtgtttcacgctctgtcatatacccgcagcaacattc
taccattattgaatatcccagcaaccgccaccgttagctccaaaattgcggactttgtta
agacttataacgacaagtctttatttttggtgtcaaatttcgaatttttttaccctggcca
ttattttagttaaggctctgctattcagaaaaggttactggattgctgccgctatttaca
ccgtcttctttaagctacgattcgaaaaatctatctttttgaggtcagtcgtaaagtctt
gggaagcaagaattgatggtttagtctcacatccttccctcaacgttgctgtcaaatcta
actgggtgcaggtgaagaaactgattagaacctatggaggaaagccaatcctggccacta
cggcttccggttcgtcatcttccaacccatccactccacagaagaatgaa**taa**attttttt
tttgaagcaaacctctaggtatgacttttgcttccttttgtttgtgttattattattcct
agtatgcatttttataacacctagttttcaatagttacaaattataccctatccagctta
tctaacggtgtcgtcatagcacagatcgctagctaatgttgtcgtgctctactcaatatc
atggatcttgcatcgaagcctttagaacttcatggaatttacaggcatgaaaaattggcg
agaccaagatcctataaattaccataatcccctaccatgcatttcttctctccttcttta
caactgtcagttttttcacaaatagacataataacataat

ORF

ATGCCAGCTGTCGGTATTGATTTAGGAACTACATACTCATGTGTGGCACATTTCGCCAATGATCGTGTTGAGATC
ATTGCAAATGACCAGGGTAACAGAACTACCCCATCATTCGTCGCTTTCACTGACACCGAGAGACTGATTGGTGAC
GCTGCTAAGAACCAAGCTGCTATGAACCCAGCCAACACCGTTTTCGACGCCAAGCGTTTGATCGGTAGAAAGTTC
TCAGATGCTGAGACCCAAGCTGATATCAAGCACTTCCCATTCAAGGTCGTTGACAAGGGTGGAAAGCCAAACATT
CAAGTTGAATTCAAAGGTGAGACTAAGGTCTTCACTCCAGAAGAGATCTCTTCCATGGTTTTGACCAAGATGAAG
GACACTGCTGAGCAATTCCTTGGTGACAAGGTTAACGATGCCGTTGTCACTGTCCCAGCTTACTTCAACGACTCT
CAGAGACAGGCCACTAAGGATGCTGGTCTCATTGCCGGTTTGAACGTTATGAGAATCATCAACGAGCCAACCGCT
GCTGCCATTGCCTACGGTTTGGACAAAAAAGCTGAGGGTGAGAAGAACGTTCTGATCTTCGACTTGGGTGGAGGA
ACTTTCGATGTTTCTTTGCTGTCCATTGAAGACGGTATTTTCGAGGTCAAGGCTACTGCTGGTGACACTCACTTA
GGTGGTGAGGACTTTGACAACAGATTGGTCAACCACTTCATCGCTGAATTCAAGAGAAAGAACAAGAAGGACTTG
TCTTCAAACCAAAGAGCTTTGAGAAGATTGAGAACCGCCTGTGAGCGTGCTAAGAGAACTTTGTCTTCTTCTGCT
CAAACTTCCATCGAGATCGACTCTTTGTTTGAAGGTGTTGACTTTTACACTTCTTTGACCAGAGCCAGATTCGAG
GAACTGTGTGGTGATCTATTCAGATCTACCATTGAGCCTGTGGAGAAGGTTTTGAAGGACGCTAAGTTGGACAAG
TCCCAAGTCAACGAGATTGTCTTGGTTGGTGGTTCCACCAGAATTCCAAAGGTCCAAAAGTTGGTGTCTGACTTC
TTCAACGGTAAGGAACCAAACAGATCTATCAACCCAGATGAGGCTGTTGCCTACGGTGCCGCTGTGCAAGCTGCT
ATTCTTTCCGGTGACACTTCCTCCAAGACCCAAGATTTGTTGCTGTTGGACGTTGCTCCTCTGTCCCTGGGTATT
GAGACTGCTGGTGGTATCATGACCAAGCTGATCCCAAGAAACTCTACCATCCCAACCAAGAAGTCCGAGACTTTC
TCCACTTACGCTGACAACCAACCAGGTGTGTTGATTCAGGTATACGAGGGTGAGAGAGCCAAGACTGCCGACAAC
AACTTATTGGGTAAGTTCGAGCTGTCTGGTATTCCTCCAGCTCCAAGAGGTGTTCCTCAGATCGAGGTCACTTTC
GACATGGATGCCAATGGTATCTTGAACGTCTCTGCCGTTGAGAAGGGTACCGGTAAGGCTCAACAGATCACTATC
ACCAACGACAAGGGTAGATTGTCCAAGGAAGACATTGAGGCCATGATCTCTGAGGCTGAGAAGTACAAGGATGAA
GATGAGAAGGAGGCTGCCAGAATCCAAGCTAGAAACGCTCTGGAGTCTTACTCATTCTCTCTGAAGAACACCTTG
AACGAGAAGGAAGTCGGTGAGAAATTGGATGCTGCCGACAAGGAGTCTTTGACCAAGGCTATTGACGAGACTACC
TCATGGATCGACGAGAACCAAACTGCTACCACCGAAGAGTTCGAGGCTAAGCAAAAGGAATTGGAGGGTGTTGCT
AACCCAATCATGACTAAGTTCTACCAGGCCAACGGAGGTGCCCCAGGTGGTGCTGCTCCAGGTGGCTTCCCAGGT
GCTGCCGGTGCTGGTGCCGAGGCCCCAGGTGCTGACGGACCAACTGTTGAGGAAGTCGATTAG

Downstream

atgatatgactccctatcctgaatcttttagatgttgtacattacatagaagagcttaat
tctaatttctccatatttacctttcattgtggtaaaaataaaacgtaaacaataaggcac
gaaacgcgacacattctaatgtgacggatgcgattttctgtaattcccccatgtaacaat
ccctttgtttgacgaggcggagtatataaaacaccaaaaattctcgtgtgtttcgtgttg
tcttgaacccattgacctaattttac**atg**tcattttccattgtatcctctagggactccc
gtgtgatgagtgacatggaatcattgaccttgaatcaacaaaagaataagcgattgtcat

ttgaagtggtaagctgcagaagctcacaacacgttgtcagccctgcttatgaacactttg
tagaggggtatttgaaagacgttgatcctgctaaggaggctgatacaccagaagacgcca
acaaattggaaggtgatacagtccttgaagacgcaaatgaaagcccaaaaagtcagaaaa
gtagggctcatagcaatcatagccatagtagtgtgagttttcatgaggcacaatccaatg
aagaagagttgtctcaagacttgcaaattaaccaggacatgactgaaagcaatccatctc
tagatgataaccaagaagtggtctccttggagagtctaccaccaaaggctagtctaactg
atctgccctccaaattactggatattatcgccagtttccttccccaacagtcagctataa
atttgtcacaatgtgcaaagaaattgattcccagcacaaggaaacgtttgctgcgaaaag
tggtgtttgtcaattcgtggaatgaagcatgtagtttgcagagagaactcaactcaaata
tccctagcaccattgtttcaatggaaaaaatgattccctttttcagtaccattagactga
gtagcggcaacttgggttcattggtcaaagaattgtggtg

AA

MPAVGIDLGTTYSCVAHFANDRVEIIANDQGNRTTPSFVAFTDTERLIGDAAKNQAAMNPANTVFDAKRLIGRKF
SDAETQADIKHFPFKVVDKGGKPNIQVEFKGETKVFTPEEISSMVLTKMKDTAEQFLGDKVNDAVVTVPAYFNDS
QRQATKDAGLIAGLNVMRIINEPTAAAIAYGLDKKAEGEKNVLIFDLGGGTFDVSLLSIEDGIFEVKATAGDTHL
GGEDFDNRLVNHFIAEFKRKNKKDLSSNQRALRRLRTACERAKRTLSSSAQTSIEIDSLFEGVDFYTSLTRARFE
ELCGDLFRSTIEPVEKVLKDAKLDKSQVNEIVLVGGSTRIPKVQKLVSDFFNGKEPNRSINPDEAVAYGAAVQAA
ILSGDTSSKTQDLLLLDVAPLSLGIETAGGIMTKLIPRNSTIPTKKSETFSTYADNQPGVLIQVYEGERAKTADN
NLLGKFELSGIPPAPRGVPQIEVTFDMDANGILNVSAVEKGTGKAQQITITNDKGRLSKEDIEAMISEAEKYKDE
DEKEAARIQARNALESYSFSLKNTLNEKEVGEKLDAADKESLTKAIDETTSWIDENQTATTEEFEAKQKELEGVA
NPIMTKFYQANGGAPGGAAPGGFPGAAGAGAEAPGADGPTVEEVD

SEQ0199

chr3_0230

5' region

gggttgtatccattcactatttactctttgtttcatttcttgaattatttggatactact
ctgctggcaactctaccagtctcaaacgcagaccaggttcgcaatttgattagaatgttc
gtgagctcttacaatgaaaagtccatgtaccttgcggctagttgtgaattatttttagtt
ccttctttgttgctatcctctttgaagtcgattatattgctggaatggtatagggctccc
ttttcatttatcaggcaattaatcgtggtattctccgtgatctcgtttctgagattaaga
tatcaacagaatgtttacatgaaacaattagttgatagttatgatttgaagatcagtcaa
ctcttataccatccccaacttcctcaaggattcaggttgggatatttacgatttaagagt
ctattaacaagcacgctaggatacttagaattggaaaaaaagaccaga**taa**tgagattga
actcgaaatttaggatcacccatatgacgaagaattcatttagattattgaaggtgtttt
catgtttacctccatgagaccatttctgtcacagcaaatacaggcaacgcttttcaccag
agcttgttggtacaacttttcagatgacgccaaattctcacgcgcctcactttgtgcggc
gctaacaataggccatttttttgtacctcccggatggttcagctcaatcactcgattgag
aggtttttgttccgcgatttttgttcaccccacacttttctcgaaggttctagcaatcaa
gataaacaccgcaaagagagccgcaggaaccatatgtggtaccacaagtggtcttaaaca
actctggtagaattcgatggaattcgatggaagccgatcgactccgatcgaattgaagca
attcgtatatataaggagaacctagttccacccctttactcgaccattagtttacaagact
aacttcacagaagcatagaaattaaacaaagttaaacatt

ORF

ATGGGTAAATCAATTGGAATTGATTTGGGTACCACATACTCTTGTGTGGCACATTTTGCTAATGATCGTGTTGAG
ATCATAGCTAACGACCAAGGTAACAGGACGACTCCATCGTTCGTCGCCTTTACCGACACTGAAAGATTGATTGGT
GATGCTGCAAAGAACCAAGCTGCCATGAATCCAGCTAACACTGTTTTCGATGCCAAACGTTTAATCGGTAGAAAA
TTCGACGACCCGGAAACTCAGGCCGATATTAAGCACTTCCCTTTCAAAGTTATCAACAAGGGGGGAAAGCCTAAT
ATCCAAGTCGAATTTAAGGGTGAGACTAAGGTTTTCAGCCCCGAAGAGATTTCCTCCATGGTTCTAACAAAAATG
AAGGATACTGCTGAGCAGTATTTGGGTGAGAAAATCAACGATGCAGTTGTCACTGTTCCTGCTTACTTCAATGAC
TCTCAAAGACAAGCCACCAAGGATGCTGGTTTGATTGCTGGTTTGAACGTTCAAAGAATCATTAATGAGCCCACC
GCTGCCGCAATTGCTTACGGGTTGGACAAGAAGGATGCAGGCCACGGTGAGCACAACATTCTAATCTTCGATCTA
GGTGGAGGAACTTTCGATGTTTCTCTACTATCTATTGATGAGGGTATTTTCGAAGTCAAGGCCACCGCAGGTGAC
ACCCACTTGGGTGGTGAGGACTTCGATAACAGATTAGTCAACCACTTTATCGCCGAGTTCAAGAGAAAGACCAAG
AAAGATCTTTCTACAAACCAGAGATCCCTTAGAAGACTAAGAACCGCTTGTGAGCGTGCAAAGAGAACTTTGTCT
TCTTCTGCTCAGACCTCCATCGAGATTGATTCTTTGTTCGAGGGTATCGACTTCTACACCTCGATCACTAGAGCT
AGATTCGAGGAGCTCTGTGCCGACTTGTTCAGATCCACCATCGAGCCTGTTGAGAGAGTCTTGAAAGACTCCAAG
TTGGACAAATCTCAAGTTCATGAGATTGTTTTGGTTGGTGGTTCTACCAGAATTCCAAAGGTTCAGAAATTAGTT
TCTGACTTTTTCAATGGTAAGGAGCCAAACAAGTCCATCAACCCAGACGAAGCCGTTGCATATGGTGCTGCTGTC
CAAGCAGCTATTTTGTCTGGAGATACTTCTTCCAAGACACAAGACTTGTTATTGCTGGATGTTGCTCCTCTATCT
TTGGGTATTGAAACCGCTGGTGGTATCATGACCAAGCTGATCCCAAGAAACTCCACAATCCCAGCCAAAAAGTCA

GAAATCTTTTCGACATATGCTGACAACCAACCAGGTGTTTTGATTCAAGTCTTTGAAGGTGAGAGAACTAGAACC
AAGGACAACAACCTGTTGGGTAAGTTTGAACTTTCTGGTATTCCTCCTGCTCCAAGAGGTGTTCCTCAAATTGAG
GTCACCTTCGATATGGATGCCAACGGTATTTTGAATGTATCTGCTGTTGAGAAGGGTACCGGTAAGACTCAAAAG
ATTACTATTACCAACGATAAGGGAAGATTGTCCAAGGAAGACATCGAGAGAATGGTTTCTGAAGCTGAAAAATTC
AAGGATGAAGACGAGAAGGAAGCCGAGAGAGTTGCTGCCAAGAATGGCTTGGAATCATATGCTTACTCTCTGAAG
AACTCTGCAGCTGAATCTGGATTCAAGGACAAGGTTGGAGAGGATGATCTTGCCAAGTTGAACAAGTCAGTTGAA
GAGACAATATCTTGGTTAGATGAGTCACAATCTGCTTCCACAGACGAGTACAAGGACAGGCAAAAGGAATTGGAA
GAAGTTGCTAACCCAATAATGAGCAAGTTCTATGGAGCTGCTGGTGGAGCTCCTGGTGGAGCTCCTGGTGGCTTC
CCTGGAGGTTTCCCTGGCGGAGCTGGCGCAGCTGGCGGTGCCCCAGGTGGTGCTGCCCCAGGCGGAGACAGCGGA
CCAACCGTGGAAGAAGTCGATTAA

Downstream

gcaattcaacggataaattctggttaatatatataacgtgaataggaaattaaggaaatt
ttggatctaataatgtgctgtatgccgacatcgggcatcgtagattgtatagtatcgctg
acactataataagccagccaaaacccctaaaccagttgccctccactaattagtgtacta
cccaatcttgcctcttcgggtgtcttttataaggacagattcacaagctcttgttgccca
atacacacatacacacagagataatagcagtc**atg**aacagcgagtacgactacttattca
agttgctgcttattggggactctggagtgggaaaatcttgtctcttgctgagatttgccg
aggatacttacacccaggactatatttcgacgatcggagttgatttcaagattagaacta
tagagctagatggtaagacaatcaagttacaaatctgggacacggctggtcaagaaagat
tccgcactattacatcctcctattatcgaggagctcatggtatcatcatcgtatatgatg
ttaccgaccaagagtcatttaataacgtcaaccagtggttgcaagagattgaccgatacg
ccactggcggagtgatgaagttgttggttgggaacaagtccgatctcaaggataagaagg
ttgtggattacactgttgccaaggagtttgcagatgctctggagatcccattaatagaga
catcagccttggactctacaaatgttgagcaagcattcttcaacatggctcgtcaaatca
aatctcagatggtgaacaaccaagctggtagtggtaccgagaaggcatcaatcaacctga
ggggtcaatccgttggaggcagtcaaaactctagttgctgt**tag**ttagtttaagataatt
gttaattggtaaatcaaagagtaagagtatatacattctgaggcgttgccgttcttctga
gtagcaagtatttgccatttggtgttgttgcctgctcgtt

AA

MGKSIGIDLGTTYSCVAHFANDRVEIIANDQGNRTTPSFVAFTDTERLIGDAAKNQAAMNPANTVFDAKRLIGRK
FDDPETQADIKHFPFKVINKGGKPNIQVEFKGETKVFSPEEISSMVLTKMKDTAEQYLGEKINDAVVTVPAYFND
SQRQATKDAGLIAGLNVQRIINEPTAAAIAYGLDKKDAGHGEHNILIFDLGGGTFDVSLLSIDEGIFEVKATAGD
THLGGEDFDNRLVNHFIAEFKRKTKKDLSTNQRSLRRLRTACERAKRTLSSSAQTSIEIDSLFEGIDFYTSITRA
RFEELCADLFRSTIEPVERVLKDSKLDKSQVHEIVLVGGSTRIPKVQKLVSDFFNGKEPNKSINPDEAVAYGAAV
QAAILSGDTSSKTQDLLLLDVAPLSLGIETAGGIMTKLIPRNSTIPAKKSEIFSTYADNQPGVLIQVFEGERTRT
KDNNLLGKFELSGIPPAPRGVPQIEVTFDMDANGILNVSAVEKGTGKTQKITITNDKGRLSKEDIERMVSEAEKF
KDEDEKEAERVAAKNGLESYAYSLKNSAAESGFKDKVGEDDLAKLNKSVEETISWLDESQSASTDEYKDRQKELE
EVANPIMSKFYGAAGGAPGGAPGGFPGGFPGGAGAAGGAPGGAAPGGDSGPTVEEVD

SEQ0200

Homologue of **SIL1** (YOL031C): Nucleotide exchange factor for the endoplasmic reticulum (ER) lumenal Hsp70 chaperone Kar2p, required for protein translocation into the ER; homolog of Yarrowia lipolytica SLS1; GrpE-like protein

S. cerevisiae null mutant: resistance to tunicamycin: decreased; resistance to mercaptoethanol: normal; resistance to L-1,4-dithiothreitol: decreased; viable

chr1-1_0237

5' region

gaaacttcctccgttgccgtctcaatccaatcttataatgcagctgaggaaagcagcttt
acacccattactgtttagaggaaatttcactgacaaagttattaaggagatgtccgttaa
aatcatgaaagaacctgtctacgctgatgccaacctggaatacatttacgaagatatggc
cataatgaacgactatgagctaaatgaactgtgcttaaagtatcctcgtacattatcaaa
gtacaagttagaagaaaaatcttttatgcagagcggaaaagttcaaaagcttaaggaatt
actagacaagattatttttgagagagaggaaaaggtattaatcttttccttattcactca
ggtattggacattattgaggtggttctctccgttcttaaaatcaagtttttgagacttga
cggccaaacttctgttgatatcagacaagatatcatcgacaaattctatgaagatgaaac
aattcccgtttttctttatctaccaaagctggtggattcggaattaacttggtatgtgc
caataatgtaattatcttcgatcaatcgtttaacccacacgatgacaagcaagctgaaga

tagagcacacagggtgggtcagaccaaagaagttaacgtttatcgaatggttaccaagaa
tacaatagatgagagcattctgcagttggccttgaacaaactacaattggatagtaccgt
cagtgatgacaacgataaaagcaccagggcatttgaggaacagacagccaaggtcattga
acaaatgctttttgaaagggaaagtagcacagacgttccagaagagaagaatgcg**taa**ag
ctgaagagtgcttctggtgtacaaatctcctagactgaatttccttgaataatttttatag
ttttagacacttagattggcattaccacgatagagggtacgggcgccattacataatcaa
tactagatcatctctcgaaataccttagaaaccttcaagc

ORF

ATGAAAGTGACATTATCTGTGTTAGCTATTGCCTCCCAATTGGTTAGAATCGTTTGTTCGGAAGGAGAAAATATC
TGCATAGGTGACCAGTGCTATCCGAAGAATTTTGAACCTGACAAGGAGTGGAAACCTGTTCAGGAAGGCCAGATT
ATCCCTCCAGGATCACACGTAAGAATGGACTTTAATACACACCAGAGAGAGGCAAAACTGGTGGAAGAGAATGAG
GATATAGACCCCTCATCATTGGGAGTGGCTGTAGTGGATTCCACCGGTTCGTTTGCTGATGATCAATCTTTGGAA
AAGATTGAGGGACTTTCCATGGAACAACTAGATGAGAAGTTAGAAGAACTGATTGAGCTTTCCCATGACTACGAG
TACGGATCAGACATAATCTTGAGTGATCAGTATATTTTTGGAGTAGCCGGGCTAGTTCCTACTAAGACAAAGTTT
ACTTCTGAGTTGAAGGAAAAGGCCTTGAGAATTGTCGGATCATGCTTGAGAAACAATGCCGATGCGGTAGAGAAA
CTACTGGGAACTGTTCCAAATACTATAACCATACAATTCATGTCAAACCTAGTGGGTAAAGTAAATTCCACTGGA
GAGAATGTTGACTCTGTTGAACAGAAACGAATCCTTTCAATTATTGGAGCTGTTATTCCTTTCAAAATTGGAAAG
GTATTGTTTGAAGCTTGTTCGGGAACGCAGAAGCTATTACTATCCTTGGATAAACTGGAAAGTTCAGTTCAACTG
AGAGGATACCAAATGTTGGACGACTTCATTCATCACCCTGAAGAGGAACTTCTCTCTTCATTGACAGCAAAGGAA
CGATTAGTAAAGCATATTGAGTTGATTCAATCATTTTTTGCATCAGGAAAGCATTCTCTTGATATAGCAATAAAT
CGTGAGTTATTCACTAGGCTGATTGCCTTACGAACCAATTTAGAATCTGCCAATCCAAATCTATGTAAACCATCA
ACTGACTTTTTGAACTGGCTGATCGACGAAATTGAAGCTACGAAAGATACCGATCCACACTTTTCAAAAGAGCTT
AAACATTTACGTTTTGAACTTTTTGGGAACCCATTGGCATCTAGGAAAGGTTTCTCCGATGAGTTATAG

Downstream

ataattggggatccctatctcattacaacattctccatgttgtaggatttttttcaagga
tgcatgcgaaagttaacgtaaattaccaaccggacccagatattcgacatcgctttcgtc
tatgaaaaattttcgggagactagttcttgcttcctatatatcccctaatctttaatata
ga**atg**aagtacgttttatccgagcaagttcttactgtcccagaagatgtctccgtgtcta
tcaaggccagagtcatcaaagttactggaccaaggggtgagctgaccaaagacttgaagc
acatcaacgttgcctttgagaaatctggcgacaacgagatcaagatcattgtgcatcacg
gtaacagaaagcacgttgctgctctgagaactgtcaagtcattaatctctaacatgatca
ctggtgtcaccaagggttacaagtacaagatgagattggtttatgcgcatttcccaatta
atgtcaacttcctcgagagagatggtcaccagtacgttgagatcagaaacttcttgggtg
agaagagagtcagagaggtcaaggtttacgatggtgtcactgcatccaactcttctgctt
tgaaggatgagctgatctttgagggtaactccattgagaacgtctctcaaacttgtgctg
atgtccaacagatttgccgtgttagaaacaaggatatccgtaaatttttagacggtatct
acgtttctgagaagggaactatcgttcaagaagaa**taa**attctggtatatctcaaaagat
ttatgtacattagttaggctttccacactgtgtgtttcctacccccattcccaaccccgt
gaatcgttacggcccttaaagagaaggcccccttacgattacacacccgtaaaataacccc
ttattatatctggaattttgttgtgatagttactgattctttggaggattaacctccagt
ggcctgggtttagttgtccgaatgaattattaacttcata

AA

MKVTLSVLAIASQLVRIVCSEGENICIGDQCYPKNFEPDKEWKPVQEGQIIPPGSHVRMDFNTHQREAKLVEENE
DIDPSSLGVAVVDSTGSFADDQSLEKIEGLSMEQLDEKLEELIELSHDYEYGSDIILSDQYIFGVAGLVPTKTKF
TSELKEKALRIVGSCLRNNADAVEKLLGTVPNTITIQFMSNLVGKVNSTGENVDSVEQKRILSIIGAVIPFKIGK
VLFEACSGTQKLLLSLDKLESSVQLRGYQMLDDFIHHPEEELLSSLTAKERLVKHIELIQSFFASGKHSLDIAIN
RELFTRLIALRTNLESANPNLCKPSTDFLNWLIDEIEATKDTDPHFSKELKHLRFELFGNPLASRKGFSDEL

10 5S ribosomal RNA gene

21 copies spread over the 4 chromosomes

SEQ0201

Ggttgcggccatatctagcagaaagcaccgtttcccgtccgatcaaccgtagttaagctgctaagagcctgaccg
agtagtgtagtgggagaccatacgcgaaactcaggtgctgcaatct

11 Xylose metabolism

SEQ0202

XYL1: chr3_0744 (80% homoloog with P. stipitis; 84% with C. albicans)
Aldose reductase: converts xylose to xylitol

5' region

atatgtattcccaaactgccgaacacttgcttagacacgtcgaaaaaccagatcagccat
ggtctctggtggggatactcttggtatcttttaaaaagaagggaactcagtgacagaagc
cctaggatcgcttgtacaaacaaggagaatggacctagtagcttacactgatcagacatt
gctgagtaacgaagaaactttaacagaaaaggatgacagaaggggaacacatagggaaag
tttcgcgtatgataaacccatgtgggaggagacaataattctttagacggagtaagtggt
gtaatgctttttcgtccggggaaagctctttcaaaagccaaccgcgagcctcagttattt
tctccttcccaaagatcacaccgacgatgtccgaaccagctcaacaaccaaaagcagaat
gtaagtaggcttttaagcttcagaagaattagggattaagcgtactaacggcgttgttag
ctaagcccaagccatgttgtgtctgcaagccagagaaagaggcccgtgatcaatgcttat
taataaatggtcaagactcaggaaaatgtgatgacttgatttcccaatacaaggtctgca
tgaaaggattcgggtttgacactaactagctctcagatgcgtgaatatctcttgtaaata
ggtcaaaattgacaggcattcatccatacggagatccaactgtggatacgctgaacatct
tgattttcggtgaaaaatcgtaacatctctggccgtctaggcagagtaacttgatagcta
tcacagtgcgtcccaaccattcttgtgaagctattttattttgacttctctcagcatgtt
tgtccgtttagcctcccaaacttcaaagatatagtttcaccatagtataaccgcatttcg
taatcaatcgcggttaaagcaagtatcatgcgctaatatataagaatctcctaagttgcc
tcccgaattcagattattctcccaacaattaacttatcggc

ORF

ATGGCTACTCTATTAAAATTGAACAATGGTCTCAAACTTCCCCAAGTAGGACTGGGTGTCTGGAAGATTCCCAAC
GAGTTGACAGCTGAGACTGTATACAATGCCATTAAGCAAGGCTACCGTCTTTTTGATGGTGCGGAAGATTATGGT
AATGAAAAGGAAGTAGGACAGGGTGTCCGAAGGGCAATTGATGAAGGACTGGTCAAAAGAGAAGATCTTTTCATC
GTATCCAAATTGTGGAACAACTACCATCACCCTGACAACGTCGGAAAAGCTCTTGACAGAACTCTTTCCGATTTG
GGTTTAGATTATCTGGATCTATTCTACATCCATTTCCCAATTGCTTTCAAGTTCGTACCATTAGAGGAGAAGTAT
CCTCCTGCATTTTACTGTGGTGATGGGAACAATTTCCATTATGAAGATGTGCCACTGCTCGACACTTACCGTGCG
CTAGAGAGGTTGGTGGATGCTGGCAGGATCAAATCCCTGGGTGTTTCGAATTTCAATGGTGCCTTGTTGCAGGAT
CTTTTAAGAGGTGCTCGTATCAAGCCTGTCGCATTGCAGATTGAACATCATCCTTATCTAGTGCAACAGAAACTG
ATTGAGTACGCTCAATCTGAAGACATTGTTGTTGTGGCTTACTCGTCATTTGGACCCCAGTCGTTTTTGGAACTG
AAGGTCAACAAGGCCTTGACGGCTGTATCTTTGTTTGAGCACGATGTCATCAAGAAGATCGCCCAAGCTCATAAC
AGATCTGCTGGTGAAGTGTTGTTGAGATGGGCAACTCAAAGAGGTCTGGCTATCATTCCCAAGTCCAGTAAACCA
GAGCGTTTGAGCAGTAACCTGCATATCAATTCCTTTGATCTTACGAAGGAAGATTTAGAAACCATTTCGTCGCTT
GACCTCGGCTTGAGATTCAACGATCCATGGGACTGGGACAAGATCCCTATATTTGCCTAG

AA:

MATLLKLNNGLKLPQVGLGVWKIPNELTAETVYNAIKQGYRLFDGAEDYGNEKEVGQGVRRAIDEGLVKREDLFI
VSKLWNNYHHPDNVGKALDRTLSDLGLDYLDLFYIHFPIAFKFVPLEEKYPPAFYCGDGNNFHYEDVPLLDTYRA
LERLVDAGRIKSLGVSNFNGALLQDLLRGARIKPVALQIEHHPYLVQQKLIEYAQSEDIVVVAYSSFGPQSFLEL
KVNKALTAVSLFEHDVIKKIAQAHNRSAGEVLLRWATQRGLAIIPKSSKPERLSSNLHINSFDLTKEDLETISSL
DLGLRFNDPWDWDKIPIFA

SEQ0203

XYL2: chr1-1_0490 (69% homoloog with P. stipitis)
Xylitol dehydrogenase: converts xylitol to xylulose

5' region:

cctgttcaagcgtttttctttacactttagatccaaaccctggcattgttcttcaagcta
ccgatttctctcatgtgaaaacgggtaatgacagaggtgacaatctgccctttgatgatg
tattcttctatgaacagaacgcgttgagtaagctagtggttgctcatgttggagatacaa
gagtagtgttgtgtgataaaaacggactggcccacccactaacacatagccaccaccctt
ctaatccaatcgaaagcagacgtctttcaaggttttctgcagggttaatgatgacggatt
cttttggagaagaaaggttcatcaactttgccaatactagatcgtttggagatcttacag
ccaaaaacaaggggggtgtccgcagagcctgaaatatcccaatacctcataggaaattcga
ctaaactgcaagcatacaaagatcaaaattctgatttgattgcaagtaaagagatacgag
actttggagggaatgagtgctttttagcacttgtaaccgatggagtcacagatcacttat

cagaccaggagattgtagatttggtcatgacaactaccaataataaaggacttttgagag
gaacacctcatgaggctgccaaagaagtcattaaattcgtggagtgcgttggaggtgatg
acaatgccacatgtaacataatcaaattgaacggttggggtaattggccgatgatggacc
gaacgggtaagctaagagaggacaaaatgatgcaaggcggaagagcaaggttagataggt
agtgttaaaaagttgtatattattaatggcggggcagaatgctctcagtctcaacttttg
atgattgcgcgggattaaaaaatcatgcacaccggaaaaacttatcccaaatgtatgtac
cccacgtaatccgttgtagttccaccaaagaatttaaaacgtacttcccgcccaataatc
tgttccctttcccgcttcattacactcactattatccaac

ORF:
ATGTCCGATAACCCAAGTGTTATTCTTAAAAGGATTAATGAGATTGTCATAGAAGATAGACCAATTCCAGCCATT
GAGGATCCTCACTATGTGAAAATAGCAATCAAGAAGACCGGAATTTGTGGTAGTGATGTCCATTTCTACACAGAT
GGATGCTGTGGATCATTCAAACTTGAGTCACCGATGGTTCTTGGGCACGAGTCTGCGGGAATTGTCGTTGAAGTT
GGCTCCGAGGTTAAGTCGCTTAGGGTGGGTGACAAAGTTGCTTGTGAACCGGGTATCCCGTCTCGTTACAGCAAT
GCCTATAAATCAGGACACTACAATTTGTGTCCAGAGATGGCGTTTGCTGCTACTCCACCTATTGATGGTACCCTA
TGTCGCTACTTTTTACTTCCTGAAGATTTCTGCGTCAAGCTTCCTGAACATGTGTCATTAGAGGAAGGTGCATTA
GTAGAACCTTTGAGTGTTGCTGTCCATGCTGCAAGACTTGCGAAGATTACCTTTGGAGACAGTGTAGTAGTTTTT
GGAGCTGGNCCTGTTGGTCTGCTGGTTGCTGCTACGGCTAGAGCTTACGGCGCAACCAACGTTCTCATCGTGGAT
ATTTTCGATGATAAACTGACACTCGCAAAGGACACCTTACAAGTGGCCACGCATAGTTTCAACTCAAAGAATGGT
ATGGATAATCTTTTGGAATCATTTGAAGGAAAGCATCCAAACGTTTCCATTGATTGTACTGGAGTTGAATCGTGT
ATTGCGGCAGGTATCAATGCACTGGCTCCAAGGGGAGTGCACGTTCAAGTCGGAATGGGAAAATCCGAATATAAC
AACTTCCCATTGGGACTTATATGTGAGAAGGAATGTATCGTAAAGGGTGTTTTCAGATACTGTTACAACGATTAC
AACTTAGCAGTTGAACTGATAGCTTCAGGAAAAGTCGAAGTGAAAGGATTAGTAACCCACAGGTTCAAATTTACC
GAAGCAGTAGATGCCTATGATACTGTTAGGCAAGGTAAGGCTATCAAGGCTATCATTGACGGCCCAGAGTAA

AA:
MSDNPSVILKRINEIVIEDRPIPAIEDPHYVKIAIKKTGICGSDVHFYTDGCCGSFKLESPMVLGHESAGIVVEV
GSEVKSLRVGDKVACEPGIPSRYSNAYKSGHYNLCPEMAFAATPPIDGTLCRYFLLPEDFCVKLPEHVSLEEGAL
VEPLSVAVHAARLAKITFGDSVVVFGAGPVGLLVAATARAYGATNVLIVDIFDDKLTLAKDTLQVATHSFNSKNG
MDNLLESFEGKHPNVSIDCTGVESCIAAGINALAPRGVHVQVGMGKSEYNNFPLGLICEKECIVKGVFRYCYNDY
NLAVELIASGKVEVKGLVTHRFKFTEAVDAYDTVRQGKAIKAIIDGPE*

SEQ0204

XYL3 (XKS1): chr1-1_0280 (66 % homology with P. stipitis)
Xylulokinase: converts D-xylulose and ATP to xylulose 5-phosphate and ADP

5' region:
ggtgcattagtagaacctttgagtgttgctgtccatgctgcaagacttgcgaagattacc
tttggagacagtgtagtagtttttggagctggncctgttggtctgctggttgctgctacg
gctagagcttacggcgcaaccaacgttctcatcgtggatattttcgatgataaactgaca
ctcgcaaaggacaccttacaagtggccacgcatagtttcaactcaaagaatggtatggat
aatcttttggaatcatttgaaggaaagcatccaaacgtttccattgattgtactggagtt
gaatcgtgtattgcggcaggtatcaatgcactggctccaaggggagtgcacgttcaagtc
ggaatgggaaaatccgaatataacaacttcccattgggacttatatgtgagaaggaatgt
atcgtaaagggtgttttcagatactgttacaacgattacaacttagcagttgaactgata
gcttcaggaaaagtcgaagtgaaaggattagtaacccacaggttcaaatttaccgaagca
gtagatgcctatgatactgttaggcaaggtaaggctatcaaggctatcattgacggccca
gagtaaacggagacatatataataataaaattgtaaggttctcttagcattgcaccccac
cctaggtttctttttagactgttatgcactcgaaatttaggtgggtatgtgtcgatactc
aattctgcacacaccaccaaaatctgtcgcctttctcttatatctccaacccgagtccaa
tgtcattttcttaccttgaaggcaacatcgatttaaaggacaggaacttgcaaacagga
tcactaaaaaactaatcacatttggtgcaattattagttttctggtaggatttttgagtg
acaacatcttatacactgtatacactttcgcagcttttggtttattgactgcttctttgg
ttattcccccttttagcttctacaaaaagaaccctgtaac

ORF:
ATGGTTACCAAAGAAATCCAAAATAGAGATTCAGCATTGACCGAATCAGTTCCGAATGACCTTTACCTTGGGTTT
GATCTAAGTACCCAACAATTGAAAATCACCTCTTTCGAAGGGCGTTCATTGACACATTTCAAGACTTATAGAGTT
GATTTCGATGAGGAGTTGAGTGTTTATGGAATCAATAATGGAGTTTATGTCAATGAAGAGACTGGAGAGATCAAC
GCTCCAGTGGCAATGTGGGTCGAAGCGTTAGATCTGATCTTTTCTAAGATGCAGAAGGACAAGTTTCCATTTGGC
ATCGTGAAAGGGATGTCTGGATCTTGTCAACAGCATGGCTCAGTTTATTGGTCTAAAGATGCCCCAGATCTCTTA

TCATCTCTCTCTCCATCGAAAGATTTGAAGTCCCAACTGTGTCCGAAGGCTTTCACATTCGAAAAGTCTCCCAAT
TGGCAAGATCACTCAACCGGAGAAGAGCTGGAAATATTTGAAAGAAAAGCAGGATCTCCTGAGAATCTGTCTAAG
ATTACAGGATCCAGGGCTCATTATCGATTTACGGGATCTCAGATTAGAAAATTGGCCAAACGCGTAAATCCAGAA
CTTTATAAGGAAACATACAGAATTTCTTTAATCAGTAGTTTTCTAAGTTCTCTACTATGCGGAAGAATAACTAAA
ATTGAAGAGAGTGATGGATGTGGAATGAATATCTACGATATTCAGAACTCGAGGTATGATGAAGACTTGTTAGCT
GTTACAGCAGCTGTTGATCCAGAGATCGACGGGGCCACTGAACATGAACGTCAAGAAGGCGTAGCTAGGCTGAAA
GATAAACTGCAAGATTTGGAACCAGTCGGGTATCGTTCAATTGGCACTATAGCTGCATATTTTGTGGAAAAATAT
GGTTTCAGTGAAGATAGCAAAGTATTCTCCTTTACAGGTGACAACTTGGCTACCATTCTTTCCTTACCACTGCAT
AACGATGACATTCTTGTTTCTTTGGGTACATCTACAACAGTCCTTTTGGTTACAGAGACTTATTGGCCTAATTCC
AATTACCATGTTTTCAAACATCCAACAGTTCCGGGCTCGTATATGGTGATGCTTTGTTACGTAAATGGAGCATTA
GCTAGAAACCAGATAAAGACTAGCCTGGATAAAAAATACAACGTGTCTGATCCTAACGACTGGACTAAATTCAAT
GAAATCTTAGACAAATCCAAACCTCTTCATGGCAAGGAGGAACTTGGTGTTTACTTTCCCAAAGGTGAGATTATT
CCTAACTGTGTTGCCCAAACAAAACGTTTTTCTTATGATGCAAAAAGTAAGAAGTTAGTTACAGCTAATTGGGAC
ATTGAAGATGATGTTGTTTCCATTGTTGAGTCTCAAGCCCTATCATGCCGTCTTCGATCGGGTCCACTCTATCAT
GGATCTGATGAAACTGACCAAGAGGAAGAGAGTGAAGTTATCCAGAGATTATCAAATTTCCCCAAAATCTCTGCT
GATGGTAAAGATCAACGTCTTCCCGACTTGATCTCTCATCCAAAGAAAGCGTTTTACGTAGGAGGAGCTTCACAG
AATGTGTCAATTGTGCGCAAGTTTTCTGAAGTTCTAGGAGCAAAAGAGGGAAACTATCAAATCAACCTTGGTGAT
GCTTGTGCCATTGGCGGAGCTTTCAAAGCCGTTTGGTCGGATCTCTGCGAAACTGAAAAAGCTATCCCGTATAGT
GACTTCTTGAGGAAAAACTTCCATTGGAAAGAAAATGTCAAGCCAGTGGAAGCTGACTCTTCCCTGTGGCTGCAG
TATGTCGATGGAGTTGGAATCTTGAGTGAAATAGAACAAACTCTAGAGAAGTAA

AA:

MVTKEIQNRDSALTESVPNDLYLGFDLSTQQLKITSFEGRSLTHFKTYRVDFDEELSVYGINNGVYVNEETGEIN
APVAMWVEALDLIFSKMQKDKFPFGIVKGMSGSCQQHGSVYWSKDAPDLLSSLSPSKDLKSQLCPKAFTFEKSPN
WQDHSTGEELEIFERKAGSPENLSKITGSRAHYRFTGSQIRKLAKRVNPELYKETYRISLISSFLSSLLCGRITK
IEESDGCGMNIYDIQNSRYDEDLLAVTAAVDPEIDGATEHERQEGVARLKDKLQDLEPVGYRSIGTIAAYFVEKY
GFSEDSKVFSFTGDNLATILSLPLHNDDILVSLGTSTTVLLVTETYWPNSNYHVFKHPTVPGSYMVMLCYVNGAL
ARNQIKTSLDKKYNVSDPNDWTKFNEILDKSKPLHGKEELGVYFPKGEIIPNCVAQTKRFSYDAKSKKLVTANWD
IEDDVVSIVESQALSCRLRSGPLYHGSDETDQEEESEVIQRLSNFPKISADGKDQRLPDLISHPKKAFYVGGASQ
NVSIVRKFSEVLGAKEGNYQINLGDACAIGGAFKAVWSDLCETEKAIPYSDFLRKNFHWKENVKPVEADSSLWLQ
YVDGVGILSEIEQTLEK*

SEQ0205

RPE1: chr3_0441 (81 % homology with S. cerevisiae; 86% homology with D. hanseii)
D-ribulose-5-phosphate 3-epimerase: converts Xylulose-5-P to Ribulose-5-P

5' region (bold and underlined stopcodon previous gene)

gcgatggacagactgcctcacaacagttcgtcaaacagggaaaagacgccttgagagtct
actggttgatgcttatctacatgattatcatgatggccgcattcaactttagttctcact
cctcccaagatctttaccctactcttctaacagttcaatatgaatttggtaaagacagat
ctacagttactaacgtcgtggcaaatcttggagccattgtcggaggtattttctggggtc
atatgtccaacttcattggtcgacgtcttgcagttttgctctgttgtatagtaggtggag
ctttcatttacccttgggcatttataaatggttcagggatcaatgctggtgctttcttct
tacaatttgctgtccaaggtggatggggtgtatgtccagttcatcttgctgagatgtctc
cgccccatttccgagcttttgttacaggaaccacctaccagctaggaaacctggcatcat
ctggatcatctactatcctggccgatattggagaaagattccccatttacgacgagaatg
gagtgcggaaagagggtgtttatgactatgccaaagtaatggctatatttatgggagctg
tgtttgctttcctgtttattgtgatgttgctgggtccagaaagaagaggagcctccaatc
ttgatgatctccaggattacattgacgattgggagcaaaaagatctggacaacaaaaaag
gaattgaaactgagttcatcgaagatgttcagttaatggaggctgtctctaatgaggacc
tttctgagaaaggcaatgagaaagtagaggcttttgaaggaaataag**taa**gattagggta
cttcttagaaattactaatactacacctctacattcgttatcttatcaggcttgtcgcga
tgggctaaactatagcctatgagtgggggtcgtagttaaagaaatatatactatatgcg
aagatgggaaagtaaaactagttgaatactgtgtgagatc

ORF:

ATGGTTAAAACAATTATTGCTCCTTCAATCCTGTCGGCAGATTTTTCAAACTTGGGATGT
GAATGTCACAAGATGTTTAAGCTACGGGCTGACTGGGTCCACATTGACGTCATGGACGGC
CATTTCGTCCCTAATTTGACCATGGGTCCTCCAATTATCTCGTCCCTTAGAAAAGCTGTT
CCGAGGGGAGAGAAAGATGGACAGACGACACACTTCTTTGACTGCCATATGATGGTGGCC
AATCCCGAACAATGGGTCCCAGAGGTTGCAAAGGCAGGTGGTGATCAGTACACTTTTCAC

TACGAAGCAACTAAGGATCCAGTGAAGTTAGTGGAACTTATCAAGAGCCATGGACTGAAA
GCTGCATGTGCTATCAAGCCCGGAACATCAGTCGATGTCTTGTACCCTCTAGCAGACAAG
CTTGACATGGCTCTGGTAATGACAGTTGAGCCAGGGTTTGGTGGACAGAAGTTTATGGCT
GATATGATGCCCAAAGTTGAAGCATTGCGGGCCAAATTTCCAAACTTAGATATTCAGGTT
GACGGTGGTCTTGGAAAGGAAACCATAGGAGTAGCAGCCGATGCTGGGGCAAACGTAATT
GTTGCTGGATCTTCTGTTTTTGGTGCTAAAGACCCCGGTGAGGTTATCCAATTTTTGCGT
GACACTGTCCAAGATAGTCTCAAAAAGAAAGGTCTTTTAGATGAATAG

AA:
MVKTIIAPSILSADFSNLGCECHKMFKLRADWVHIDVMDGHFVPNLTMGPPIISSLRKAV
PRGEKDGQTTHFFDCHMMVANPEQWVPEVAKAGGDQYTFHYEATKDPVKLVELIKSHGLK
AACAIKPGTSVDVLYPLADKLDMALVMTVEPGFGGQKFMADMMPKVEALRAKFPNLDIQV
DGGLGKETIGVAADAGANVIVAGSSVFGAKDPGEVIQFLRDTVQDSLKKKGLLDE

12 Arabinose metabolism

Not all enzymes could be detected

LAD1: L-arabinitol-4-dehydrogenase of Trichoderma reesei has homology with P. pastoris XYL2 (chr1-1_0490)

LXR1: L-xylulose reductase of Trichoderma reesei has homology with peroxisomal 2,4- dienoyl-CoA reductase (chr4_0754 & chr4_0988)

SEQ0206

chr2-2_0305 putative kinase (possibly ribulokinase, araB), 72% homology with possible ribitol kinase or glycerol kinase of Pichia stipitis and homology with hypothetical proteins of P. guillermondi & D. hanseii

5' region (bold & underlined stopcodon previous gene)

acattggtaattcgttagcgctggagccttacttcagcgagagatggccccagaaccacg
ttctactgggggtcatggacaagatacacaagatttactcaacagactttgcaccattgg
ttgcagcccgttcatttggtgttgatgttttgaattcactgggaccaatcaaggatttta
tggtgggaaggatcagtggaaattccaaaagatgagatgtcttctgttctcccttcttgt
ttagcagagtagaacagctttacctgaatggctgatgtgcattgcgattaggctaatcag
taactcaagcttttccagtcagatcgatcgactactcgaaagccccagtcctcatccctc
tactatcggcgcccgtgtatataaagagatagaagtaaactctgcgcgcgttttgcctct
attcttctgctagaaaagaaaaaaaaaaacgtttctcatatactaagcactaagaatgag
tgaagaagactcccaggaattagaagagacttttaacagaataaaaacacatttcccctc
tgctaggataaaaaaattgatgcagagtgatgatgatatcggaaaagttgcccaggccac
cccagtggtggtgggaagagcattggaactgtttttatgttcactggtagacaagtcatt
ggaggtggctcgtgagagcggatcacgccgtatacaacctgcacatttgcgtaaggctgt
ggcagagaatgagcaattcgacttttgccagtcgattctggatggtgagccaagggaaga
g**tga**ttgactacaaaacaataaatttcatgcttctgcgtcttctccccctcccatttcct
ccgcgcaaaaagagaaactagcttacccaatagtaacacaagggaatttgtttttttttgc
cccttccacgtcctttagcatcgccttcttatcgcaaccatttcccattctagtctcttt
cttctccactaaccctagagagcattttatttagataca

ORF:

ATGTCGAAAAATAATTTAAAGTCGTCCCGGTCCCGGTCCTTCACTAATCCAAATTTTCATCTCAACATCTCCGAG
TTACACCAGAACTTCGCCACTCCGGTCTATTACGTTGGTGTTGACGTTGGATCTGGATCTGCCAGGGCTGCAGTA
GTCGATCAGGCCGGTGCAATATTGGGGCTAGCAGAAAAGCCTATCTCCAAGTATACCCCTAAAGCAGATTATGTC
AATCAATCTTCCACAGAGATTTGGGAAGCCGTATCTTACTGTGTAAAGACTGCCCTGACGCAGTCCCACATTGAT
CCTGCTCTTGTCATGGGAATTGGGTTTGACGCGACTTGTTCTTTGGTTGTATTGGATGAAGAAACTGACGAACCA
ATTGCTGTTGGTCCTGATTTCACGGAATCAGAACAGAACATTATCATGTGGATGGACCACAGGGCACATGAAGAA
ACCAAAGCCATCAACCGTACTGGTGACAAATGTTTGAAATATGTTGGTGGTCAAATGTCGATAGAGATGGAGCTC
CCCAAGATGAAATGGCTCAAACATCATCTTCCTCGTGATGAAACTGGAAAGTCACTTTTTGAACGTTGCAAGTTT
TACGATTTGGCAGACTTTTTAACACACAAGGCTACAAACACGGAAACTAGGTCTTATTGCTCAGTTACTTGCAAG
CAAGGTTTTGTCCCTCAAGGTGTAGATGGTTCAGTAGACGGCTGGTCCAAGGAATTCCTCGCACAGGTAGAGCTT
CCAGAGTTAGCAGCAAACGATTTCCAAAAACTTGGAGGTATCCCTGGAAAAAATGGAAAGTACCTAAGTGCTGGT
GACTCCGTTGGACCTCTCAGTGCTGATGCTGCAGAGCAATTGGGCTTGACAACTGCTTGTTGGGTTGCTTCGGGT
GTTATTGATTGTTACGCTGGGTGGGTTGGAACAATCGCTGCAAAAACTGAAATCCCGCTTCCAGATTTGGTTGAA
CAAGACAATAATTTTTCTGGTATTGACAAAGCCTGTGGCCGTCTTGCTGCTGTTGCAGGTACATCCACTTGTCAC
TGTGTCATGTCTAAGGATCCAATTTTTGTGCATGGTGTTTGGGGTCCTTATAGGGATGTGATGGCTAAAGATTAC
TGGCTTGCAGAAGGTGGCCAGTCATGCACAGGCGCTCTTTTGGCCCATGTTCTCACGACCCATCCTGCCTACACT
GAGTTAGGAAAAGCTTCGGAATCCTCTGGGCTGTCTCGATTTGATTTTTTGAACAACAGACTCGAGAATTTGAAA
AGATCTAGGAAAGAACGCTCAGTTCTTGCCCTAGGAAAGGATTTATTTTTTTACGGTGACTATCATGGTAACAGA
TCACCATTGGCTGATCCTGATATGAAAGCCTCCATCATTGGCCAGTCCATGGATACTTCTTTAGACTCTTTGGCT
ATAGAGTATCTTGGTGCTTGTGAATTCATTGCACAACAGACCCGTCAAATTGTGGAGAAGATGGAAAAGTCAGGA
CACAATATTTCTTGTATTTTCCTTTCAGGGGGACAGTGTCGTAACGGTCTCTTGATGAGACTGTTAGCAGATTGT
ACTGGTCTTCCAATCATCATCCCAAGATATATTGAGGCATCGGTAGTTTTTGGATCAGCACTGTTAGGTGCGGTT
GCAGCTGATGACGCCATTTTGGCACAGAACAAAGGCAGGACAGGGAGAACAGTTAGAAGCAGTTCCAATGTTTCA
GAGCCTCCTCATGAAATGCCACCCTCTCCTTATACTGCACCTACCGCAACAGCGTCGATGACCTCAATTTCAGCT

ATGAATGCTCCAAACTTGAGCGGTGCACCTGGTGGCTTCCCATTTCCTCTTATGACACCAATCGAGGATGAGTCG
AAGCAATTGGGTTTCGAAGATGGGGTTGAGGATAACACCTCGTCAGACGAAGAAACTTTATCTTTTGGCACCAAG
AACAAAAACCAAGCCAATGTTCAGACTAACTTCGTTCAACAGCGCACAAGTCAAAAGCCACTAGTAAGCTCTAAG
CCATCGAAGTTAGGCGAGGAACGTTCTGGAGATCGTCTGTGGAAAGTCATGGAGTCTCTGAGTGGAGTAGGAAAG
GTCATTATGCCTTCATCTCCTAATGAACCTGACCGTAAGCTATTGAACACCAAGTACAAAGTCTTCCTTGATCAA
GCGGAGTCACAAATGCGCTACCGTAAGCTAGTAGCTGACACTGAAGAGGCTATTAAAGCTTACATGAATATTTAG

AA:
MSKNNLKSSRSRSFTNPNFHLNISELHQNFATPVYYVGVDVGSGSARAAVVDQAGAILGLAEKPISKYTPKADYV
NQSSTEIWEAVSYCVKTALTQSHIDPALVMGIGFDATCSLVVLDEETDEPIAVGPDFTESEQNIIMWMDHRAHEE
TKAINRTGDKCLKYVGGQMSIEMELPKMKWLKHHLPRDETGKSLFERCKFYDLADFLTHKATNTETRSYCSVTCK
QGFVPQGVDGSVDGWSKEFLAQVELPELAANDFQKLGGIPGKNGKYLSAGDSVGPLSADAAEQLGLTTACWVASG
VIDCYAGWVGTIAAKTEIPLPDLVEQDNNFSGIDKACGRLAAVAGTSTCHCVMSKDPIFVHGVWGPYRDVMAKDY
WLAEGGQSCTGALLAHVLTTHPAYTELGKASESSGLSRFDFLNNRLENLKRSRKERSVLALGKDLFFYGDYHGNR
SPLADPDMKASIIGQSMDTSLDSLAIEYLGACEFIAQQTRQIVEKMEKSGHNISCIFLSGGQCRNGLLMRLLADC
TGLPIIIPRYIEASVVFGSALLGAVAADDAILAQNKGRTGRTVRSSSNVSEPPHEMPPSPYTAPTATASMTSISA
MNAPNLSGAPGGFPFPLMTPIEDESKQLGFEDGVEDNTSSDEETLSFGTKNKNQANVQTNFVQQRTSQKPLVSSK
PSKLGEERSGDRLWKVMESLSGVGKVIMPSSPNEPDRKLLNTKYKVFLDQAESQMRYRKLVADTEEAIKAYMNI*

SEQ0207

Chr4_0280: Putative methylthio-ribulose-1-phosphate dehydratase (homology with E. coli araD); acts in the methionine salvage pathway; potential Smt3p sumoylation substrate; expression downregulated by caspofungin and deletion mutant is caspofungin resistant

5' region (bold & underlined: stopcodon of previous gene)
aagctcgtaatcagggcttggacagatcattgtttaagattttgagtgagaagcatcctc
aaagtgtcgtctctttgagctatcaatatcgcatgtgcgcagacgtgatgcttctatcca
acaaactaatttatgaaaaccggttaaaatgtggtagcgataacgttgcccaacggtctt
taagaattccaatgaaagagagactggatcacttactgcttgatccaagtacaccgaaaa
atcgattgtggcttcaccacatctttgtagaagagaacagggttatcttcttggaccatg
atgacgttccagcaaaagaaagaatgtcgacagatagagtggacaacccaactgaagcag
aactggttggccaaattgttgacagtcttatcaattgtgggtgtgacagcaaggagattg
gtgtgatgtcgttttatcgagcccaggtaaaattgatgaaaacaagacttgtttctcatg
tgcaaaatgtggaggtcatgacagcagatcagtttcaaggaagggataaagaatgcatta
taatatcactagttcgatccaacgctaccaactcagtgggtaatctactaaaagagtgga
gacgaatcaacgttgccattaccagagccaaatcaaagttactaattctcggcagcagga
agacattggaacaacttgaaactctgaaagccttttgtcactattgaaaagtagaggat
ggatttacagactgccgcttggagctgattccgtttacaaaggcctcaagcaagataacc
ataagacagagccagccagaaaagacaattccaaaatgtcattgaaaatgataaaatcca
ggccaatattgaaggacattttcaatgacttggtaaac**taa**tattattatcatacatcgt
gtatatagacgatcgattagcattttatttagtaggcccacgagcccagtgatttctcga
tttaataatttgacatttgatccattcggctgtcaactcca
ORF
ATGAGTAGTTGTTTCTGTTCAAAGAGCCAAGAGACTGAAAAGCGAGGAGCTGTATCAGAAGACCAGCTAGTGATT
TCTGATGATTCCACTCATCCTGCAAACTTGATCTGTGAACTCTGTAAGGTTTTCTATAACAACGGATGGGTGACA
GGTACTGGTGGTGGAATTTCCATCAGAGAAGGATCTAAGATCTACATAGCGCCCAGCGGTGTACAGAAGGAAAGA
ATGGTGCCAGACAACATGTTTGTGATGGACCTAGAATCAGAGAATTATTTGCGAACTCCTCTAACACTCAAGCCT
TCTGCCTGCACTCCATTATTTCTCAGTGCTTACAAGATGAGAGACGCAGGTGCATGTATCCATACCCACTCTCAG
GCGGCTGTCATGGTGACATTGCTGTATGATAAGGTGTTTGAAATTTCATCTATAGAGCAAATCAAAGCAATTCCA
AAAGTCACTGAGAAGGGAAACTTGATGTACTCTGATAGGTTGGTGATTCCCATCATAGAAAACACTGAGAGAGAG
GAAGATCTGACTGATTCTCTTCAGCAGGCCATCGAGGAATATCCTGGAACAACAGCTGTATTGGTAAGAAGACAT
GGTATCTACGTTTGGGGAGAAACCGTCTGGAAGGCCAAAGTTTACAATGAGGCCATTGACTATTTGTTGGAGCTA
GCTCTCAAGATGCATCAACTTGGTATTCCCACTGTAAAACAGAATTAG
AA
MSSCFCSKSQETEKRGAVSEDQLVISDDSTHPANLICELCKVFYNNGWVTGTGGGISIREGSKIYIAPSGVQKER
MVPDNMFVMDLESENYLRTPLTLKPSACTPLFLSAYKMRDAGACIHTHSQAAVMVTLLYDKVFEISSIEQIKAIP
KVTEKGNLMYSDRLVIPIIENTEREEDLTDSLQQAIEEYPGTTAVLVRRHGIYVWGETVWKAKVYNEAIDYLLEL
ALKMHQLGIPTVKQN

araA (L-arabinose_isomerase): no homology

13 Threhalose metabolism

SEQ0208

ATH1 (YPR026W)

Acid trehalase required for utilization of extracellular trehalose

Chr4_0342

Upstream (in bold start of next gene)

cgtgagatcgacctttccgttgcttaaatgtttgacagtttctagatcttgtatacttct
agccccaccagcatatgtaactggaagatcactccattcacttaatttccgtaccagctc
ttggtcaattccttgacaaagtccctccacatctgcagcatgcactaagaactcatcgca
atactgggacagcagttggaaaaactctctatctagcacggatgaggtcaatgtttgcca
tttgttcatagcaaccacccattggggacttccatctatctcttgtctacgacaactcaa
gtccactaccagtttctcttttccaataaggctggacaggtgttgaagtctttctataga
aaactgaccctcagggaataaccatgaagtgactatgacatgcgatgcaccttgatctat
ccattcttgagcattggaatccgttattccaccgccaatctgcaaacccccctctccacgc
acccagagcttctcttgcagcgtcgtcattagcctttaaagagcccaatttgatgacatg
tgttccatgaacctggttctgcttatataactgtgcataataggacgagggctgttcact
gacaaaatttgtttctacttcatcagattcagtgtcatccttcaccagcttgccgcctac
aatctgtttgactttgccggaatggatatcgatgcaacctctgaagacagt**cat**acggac
tatttttctatggtggagcacttactccaaaacctctcgatctggcgagcaagatcccaa
agacggaaaaaaaaaagtcatacattccccatgtttaatgcacctaaatttctaatttagg
aactcgcaaactttcaggtgccagatagagcaatcaaatttctccgtggaccgatgcgga
acatttggcttttacttcactagattagggcaatgccgtatggatccatatataatagca
gaataccaaagaaacctcctcccacctctcagactagagaa

ORF

ATGCTGAATAGAGTGTTGTTGGTTGCCCTTTCTTGTGTGGTTTTCTTTCACTTGGTTACAACCTTTCCTGTTGGC
ACTTCAAGTGATTCCTTGCAGATTAGGAATCTTCTGAGTCACAACTTCACCAGAGCTAATATCTCTGAAGGTCTC
AGTTCAGGAGCCACTTATTTTGTCGATGAAGACACTGAGACATATTACGACAAAGAGCTAAAGGTCCTGAGAACT
ACAAGATTCCCTCGCTATAACAATTACCAGCTGCAACCTTATGTTGCCAACGGTTATATTGGTTCCCGAATCCCC
CGAGTTGGGTCAGGATTTACCTACGACACTAGCGATAATAAGACTTCCGAGAATCTGAAAAACGGCTGGCCCTTA
TTCAACAAGAGGTACTCAGGGGCCTTTATTGCTGGATTTTTCAACTCTCAACCGACCGTACCTGAGACCAACTTC
GAAGAGTTGGAAAAAGACGGTTATGAGAGTATTATTGCCTCTATTCCTCAATGGACTAGCCTAGAGTTAACCGTT
AATGTGAATGGAACCAACCAAACTCTCAAAGCAGATGATGTCGATATAACCCACATCAGTGACTATTCTCAACAG
TTATCGTTGTTAGATGGAATTGTCACTACTAATTACACATGGTTGGGACTGGTCAATGTTTCAATCTCTGTCCTA
GCCCATAGAGACATTGTAAGTTTGGGATTCGTGTCTCTGGAGCTATCATCGCAGAAGAATATAACTGTTTCCGTC
ACTGATATTCTTGATTTCGCAACCAGCACCAGGTGTTCTTATTTGGACTCAGGAGTAAATGAACAGAGTATTTTC
ATGAAGGTTCAACCGTCAAATGTTCCGACGAATGCTACGATTTACTCTTCATTGATGAGCTCTAATTCGACCTCG
TCCCTTTTAAAGCAAAATCAAACGGTATCTCAAACACTTCGGGTAAATCTATCGAAGAACCAAGCTGCATCGTTT
CAAAAATACGTCGGCGTTGTCTCTGACGACTATCTAGATAGCATTGAAACTAACTTGACCTCTTACCAATTTGCT
AGAGAAACTGCTAAGTTTGCTGAAATAAAAGGGCGTTCGTGGATCTTGAAATCTCATAAAGAGGCATGGAACGAA
TTGTTAAATGGAAAGTCTATTGTCTTTCACGACAATGATTTTTTGACATTGGCTTCCGATTCAAGCATTTACCAC
TTGATGGCCAATACAAGATCAGAAGCCAATGGAGGCACTTCTGCTTTAGGGGTTTCCGGGCTGTCTTCGGATTCA
TACGGAGGAATGGTGTTTTGGGACACTGACTTTTGGATGCTACCGAGTGTCCAAGCCTTTAGTCCCCGACATGCC
GTTTCTTTATCCAAATTTAGAGACCATACCCATGATCAAGCTAAGAAGAATGCCCAAACTCGTGACATGAATGGG
GCAGTCTACCCCTGGACTAGTGGTAGATTTGGAAATTGTACCTCAACTGGTCCGTGTTATGATTACGAGTATCAT
ATCAATATAGACATTGCCTTCATGTTCTGGAAATTATATTTGGGAGGTGCAATTGATGATGATTATATGAAGGAA
TTTGGATACCCAATAATTGAAGATGTTGCCTCTTTTTTCGTGGATTACGTTGACTACAATTCAACTTTAGATAAA
TATACGACCAGAAATCTGACAGACCCAGATGAGTATGCTGAGTTCAAAAATAACGCAGCATTTACCAACGTCGGT
ATTTCTCAGCTCATGAAATGGGCTCTCATCCTTGGAAAGCATCTTAAAGTCGGAAATGAAAGGAGTTACGACAAA
TGGGAAGATATCATGACCAAAATGTATTTACCTGTCAACCATGCAGGAGATGTTACGTTGGAATATACAGGGATG
AACAACTCTATCGAAGTCAAACAAGCTGATGTGGTTCTTATTTCTTATCCTTTAGATGATGAGGATGGTGCCTTA
CAGGAATATTTCGATTACGATGAGGATAGGGCTATCAGCGACGTTAGATATTACTCTGATAAACAAACAGATGAA
GGCCCGGCTATGACTTTCTCTGTTTACTCTGCGGTCAATGCGAAGTTTAACAAAGAAGGATGCTCTTCTCAAACC
TATCTCTTGAAATCTGTAGAACCTTATTTCAGGTTTCCATTTGGACAGATGAGTGAACAGTCAACGGATCAGTAT
GATACCAATGGAGGAACTCACCCTGCCTTCCCATTTTTAACCGGTCATGGAGCTTTCTTGCAAAGTTCTATTTAT
GGTTTGACCGGATTAAGATTTAGCTATATTTACAACGATACAGACAAATCTATCAAGAGAAGGCTGGCCTTCGAT
CCCCTTCAACTACCGTGTTTACCGGGCGGATTCTCTATAAATGGATTTGTCTATATGAACCAGACGTTGGACATC

ACTGTGAACGATACGTATGCTACAATTGCACATCGGGGAAATGCTACCACCATCAATGTTTATGTGGATTCTAGA
AATGAAATGGGAGGTAAGGAACACAAGATCCAACCAGGAAAGAGCCTTTCCATCCCATTATACCAAACAGAGCAG
AATATTCCAGGATCATTCATTGAATGCACGGTAAAGAATGTCACTGCATTGCAACCTGGCGTGGTAGGAGACCCT
ATCCAGGCAGTTGCTGACGGAGATAACTCCACTATCTGGAAAATTGAATCTCGTGAAGAACCGACTCATCTGATA
TTTGATCTTGGAGATGAGCTTGACATTGAAGGAGGCTTGGTTGTATGGGGAACATATCCTGCAGAGAGCTTTTCA
GTCTCGGTTTTGAGAGATTTTAATTCAACGAACTACAGAGTCATCAACAACGTTGAAAACTACGATTTGATCTAC
GAATCTGGGAATGTCACAGCATCCAGCCCTTTTGACGAATCTCACATCAAGAAAGTCCAAATCTTGCCTCACAAT
TGTACGAATTTTACGTTTTCCGAACTGACAGCCTCTAGGTATGTCTTATTTGAGTTCACAGATGTTCTAGGTTAC
CCCCAAGATTATTCATATGGAGCCCAAGTAGCCGAAGTTGTTCTCTACTAA

Downstream (in bold start of previous gene)
cgaccttacgtaacgaaatatccccccatctttagattaaatgaaaaagttttgtttgtta
ttacataatctcagtaggttgtcgcatcatccggggtaatgtcgcaccgaaatctgggga
ttccctctgtaacccaccagtataaataatgaacagttctaccagcttcccggttactcc
aacca**atg**tctccagtacctcctgctccaggaatttacactccagtaccaacattcttta
agaacgacggttataccatcgatttcgacgctaatgtcaagcatgcgaagtttttgaaag
acaacggtattgctggtttagttatcatgggttccaccggtgaaggtgtccatttgacta
aagaggagagagctgcaactataaaagctgttcatgacgctcttccagatttccccatca
ttggtggtgtggttcaaaacagtgtacaagatgccctggatgagatcgattccatcaaag
ctgcaggagccacccatgctgtagtcctttcaagtaattactacggtgctggtatcaagc
aacaaggtatcattgactggtttactgccgttgcagacaaagcttctctcccaattttgg
tgtacgtttacccaggagtttccaataacttgttcaccgaccctgccacagtaatcaagc
tatctgggcacccaaacattgttggcactaaaatttcacacggtgacgtatcccatcaca
caataattgccactgataagaatgtccagcagaataacttcaacacttttactggccttg
gacaattgctagttccaactcttaccatcggctgtaaaggaaccatcgatgctctttctg
gtgcctttcctaaaatttacgttaaaatctttcaattggtgcaggacggcaagattgagg
aggcagtaaaactgcagaacactgtatctagaggcgaggagatcgtggttaattttggcg
tcattggaatcaagaaggctattcatttgggagccggaata

AA
MLNRVLLVALSCVVFFHLVTTFPVGTSSDSLQIRNLLSHNFTRANISEGLSSGATYFVDEDTETYYDKELKVLRT
TRFPRYNNYQLQPYVANGYIGSRIPRVGSGFTYDTSDNKTSENLKNGWPLFNKRYSGAFIAGFFNSQPTVPETNF
EELEKDGYESIIASIPQWTSLELTVNVNGTNQTLKADDVDITHISDYSQQLSLLDGIVTTNYTWLGLVNVSISVL
AHRDIVSLGFVSLELSSQKNITVSVTDILDFATSTRCSYLDSGVNEQSIFMKVQPSNVPTNATIYSSLMSSNSTS
SLLKQNQTVSQTLRVNLSKNQAASFQKYVGVVSDDYLDSIETNLTSYQFARETAKFAEIKGRSWILKSHKEAWNE
LLNGKSIVFHDNDFLTLASDSSIYHLMANTRSEANGGTSALGVSGLSSDSYGGMVFWDTDFWMLPSVQAFSPRHA
VSLSKFRDHTHDQAKKNAQTRDMNGAVYPWTSGRFGNCTSTGPCYDYEYHINIDIAFMFWKLYLGGAIDDDYMKE
FGYPIIEDVASFFVDYVDYNSTLDKYTTRNLTDPDEYAEFKNNAAFTNVGISQLMKWALILGKHLKVGNERSYDK
WEDIMTKMYLPVNHAGDVTLEYTGMNNSIEVKQADVVLISYPLDDEDGALQEYFDYDEDRAISDVRYYSDKQTDE
GPAMTFSVYSAVNAKFNKEGCSSQTYLLKSVEPYFRFPFGQMSEQSTDQYDTNGGTHPAFPFLTGHGAFLQSSIY
GLTGLRFSYIYNDTDKSIKRRLAFDPLQLPCLPGGFSINGFVYMNQTLDITVNDTYATIAHRGNATTINVYVDSR
NEMGGKEHKIQPGKSLSIPLYQTEQNIPGSFIECTVKNVTALQPGVVGDPIQAVADGDNSTIWKIESREEPTHLI
FDLGDELDIEGGLVVWGTYPAESFSVSVLRDFNSTNYRVINNVENYDLIYESGNVTASSPFDESHIKKVQILPHN
CTNFTFSELTASRYVLFEFTDVLGYPQDYSYGAQVAEVVLY

SEQ0209

NTH1 (YDR001C)

Neutral trehalase, degrades trehalose; required for thermotolerance and may mediate resistance to other cellular stresses; may be phosphorylated by Cdc28p Chr4_0816

Upstream
gaaaaaccaatctaatcagtaatgttctcaagaatatagcaaaacaaattcgcaggaaaa
gggaaataggataatacacaattgttagcattcattatttggtcttgttaatctcttgag
ctttctcaaattccttcttgaatccgtcagcattttccttgttaccaaaacggatggcaa
aagtttgggcttctggcttaccctcagaaacatcggcggttactgagtaaacccatgaac
ggtctgaaccgacgtttggcttcaaggtgaagtcaggggaaatgaagtgattggcgcata
ctttgagaaccttgtctctacgcatcaaaagacgaacttttccagtttccttgttcttta
agaacttagcatctccaacacctctttccttccattccttaccttcttggtcaaaacgga
acaacttagcacgaactttgtacaatacctcctcattctcttcgagtgttttcacttcca

cttttctagtttgacaataggatcgaaatgaatatctggtgattctggctcctcttctt
tgtcagcctttggctcgctggaggcttcagattctttggagacctcagcttccttgtctt
ccttagaagtttcagtttccttcttctttccaccgaacatagagaagacattggaagtag
gaggagtccgagttccagaaaaggaattcaaagtggaagaagtctgagtttcggctggct
tggactcttcagctggcttgttttcttgcttttgttcagacattgactctgttatgcgtg
ttttcaagaaaaatttgttcgttttggttctctttgacattttgatcccttcgcgaagga
gcttctgggggaaggaactttgctattttgcttggtggcttctcctgccccccacatac
gcgaactgcctgacctgacttcaaaaattaagaaacttgcttttgcctcttatctatttc
cattttacagccttataacataccgcatacaacattcatca

ORF
ATGCAGGACGGCAACGCAATTGAACGCCCACACAGCGTAGCCGTCGACGTCCCAGACGGCTTGCAACAACCTCAT
CATACCAGAACTTTCTCTACTGGATCCAACTCGGGGTTCATTGATCCATTCTCTGAGCCCTCTCAATACTATGGA
CCCGCCACTGATATTGCACGCGCAAGTACTAGGTCAAAGATTGGTCGTACAAGAACACTTTCAGCTGTCGAGAAC
AGATTTAATCGGCCAATGGTGCACGAAGATGACGCCTTCAGCAAGGTTCACCTTCAAAGAAGAGGATCTAATGAT
ACTCTGGTTAACAGAAGGTTTTTCATTAGCGATATTGATGAGACATTACGAGATCTCTTGAAGAATGAAGATAGC
GATTCCAATTGTCAGATCACTATCGAGGACACTGGACCGAAGGTTTTGAGAGTTGGAACTGCCAATTCAAACGGA
TACCGACAAACTGATATTCGAGGAACATATATGCTATCAAACTTATTGCAGGAACTAACTATCGCGAAACGTTTC
GGCCGAAATCAGATGGTGCTAGATGAAGCTAGATTGAATGAGAACCCAGTCAAGAGACTTCGAAGGTTGATTTCT
CATTCTTTCTGGAAAAACTTGACCCGTCAAGTTGGAAGTGGCAATGTGGCCGAGATCGCTCAAGACACCAAAATT
GACACCCCCGATGGCCGGATCCCTCGTATCTACGTTCCACATGATGACCCGGAGCAATATGAGTTCTATCTAGAA
ATTGCACGAAAAAATGAGGCCGTCGGAGGGGCCCGGCTAGATGTTAGGTATTTGAATGAAGTGATTGATCCGGAG
TATGTCAAATCTATTAACGGTACTCCGGGACTTCTTGCTCTCGCAACCCATAAGGTGCCTGATGGCCAGGGTGGA
ATTACCCTAGAAGGATATCCATATGTGGTTCCTGGAGGCAGATTCAACGAATTATATGGCTGGGATTCTTACATG
ATGGCATTAGGTTTACTCAAGGACGGAATGTTGGATCTTGCTCGAGGAATGGTGGAGAATTTCATCTATGAAATC
CGTCATTATGGAATGATCCTCAATGCTAACAGGTCCTACTACTTGGGCCGCTCACAGCCCCCATTTTTAACTGAT
TTTGGAATCATGGTTTACGAGGCCATGGTTGAGCATGCCCATCAGAATGAAGAAGATACTGAAAAATCTTTACAA
GATGCCGAGTACTTTCTTAGACGTACATTTTGCGCGGCCATATGCGAATATAAGACCGTATGGTGTTGCGAGCCT
AGATTAGACAAGACTACAGGATTGTCATGTTACCACCCACAGGGAGTAGGTATTCCTCCTGAAACCGAATCAACA
CACTTTACTACACTTTTGACTCCATATGCTAAGAAGCATGGAGTCTCTATAGAAGAGTTCTATGAGCTATACAAC
AATCATACCATAGTCGAGCCTGAACTCGATGAATACTTCCTTCATGACCGTGCAGTGAGAGAAAGTGGACACGAT
ACTTCTTATAGATTGGAAGGAATTTGTGCCAGTCTGGCAACTGTGGACTTGAACGCACTACTTTATAAGTACGAG
GTAGACATTGCCAAGATTATCAAAGAGCATTTCAATGACGAACTGGTCACTGAAAACAGCGTCGAGCACAGCGCT
GAATGGACCAAGCGTGCTGAACTTCGAAAGGAAAGGATGGACAAGTACCTTTGGAACGAAGAAGAGGGTATTTAT
TTTGATTATAATCTCAAGTTGAAAAAGCAACATCGTTATGAGAGTGTAACCACTTTTTGGCCAATGTGGGCCGGA
TGCAGTTCTCAAGAACAAGCTCAGTTAATGGTGGATAAGTCTCTATCTAAATTTGAGGAATTTGGTGGTCTTGTC
GCTGGTACTCTGGCATCCCGAGGACGTGTGGGATTAGAAAGGCCCAGTAGACAATGGGACTACCCTTATGGATGG
GCGCCACAGCAAATCCTAGCTTGGGTAGGACTTGACAAGTACGGCTTCCGTGGTCATACCAAACGCCTTTGTTAC
CGTTGGCTATACCTAATGACCAAGGCATTTGTGGATTATAACGGTATCGTAGTTGAAAAATATGACGTAACCACT
GGTACTAGCCCCCACAAGGTTGATGCTGAATACGGGAATCAAGGTGCCGATTTTAAAGGAGTAGCCACTGAAGGG
TTTGGCTGGGTCAATGCCTCTTTTATTTTGGGATTAACCTACTTAGATGTCCAAGGCATCCGTGCGATAGGCGCC
GTTACGTCCCCTGATGTATTTTTCAGGAAGCTCAAACCTTGGGAACGCGCAAGTTATGGCCTAAGGCCATGTAAC
GAGATAGTCAAGTCAAACTAG

Downstream (in bold start codon previous gene)
aagtatacggtttccccgcagaaatagcagaaataggcgacaaatacatacaacattttc
attgtgataggggcggcggttcctaggagggacaacccccagaaaccttgtagactacg
ttttcacgacgatgggttattactgtaaaggaagaatatactacccaccagttgaatgtt
tgaacggatcaaaggtcgaagggagtacacggcccaaccaacgtagctaccggagaaagc
aagactttcccaaaccaaatagctccgggtttcttctccggcaacccgtcagtttttgtg
tggccggacaaaaattcgcaccctcagtctaattgaaaggtcgggctccgagctctaggc
gtttgcgcatgtaatattgcatcccctcccatagataatactgcgcgaacacagggtgca
aattatgatgaccacacatgccagtgaccaaaacagtttttagtctttaaaaaccctcg
gaacttctgagtatataaaggcttctcatttcctacaagcaaacaaagaagaaacttcca
ctttctaacttttttatctatagactttagagttacaaccaacgaacaataacaaa**atg**gt
taaagtcacagtttgcggagcagccggaggtattgggcagccattatctctgatgttcaa
attgaacccatatgttaccactttggcactttacgatgtcgtaaatgttccaggagtcgg
taaggatctatcccacattgacacagacactaaacttgagtcttacttgcccgagaacga
cggactggagaaggccttgacgggttctgacttggtcatcatcccagctggtgttccaag
aaagcctggtatgaccagagacgatttgtttgccattaacgctggtatcatcagagactt
ggctaatggaatcgctcagtttgccccatcggctttcgttcttgtcatttctaaccctgt

caactccacggttcctattgtcgcagagattcttaagaaga

AA
MQDGNAIERPHSVAVDVPDGLQQPHHTRTFSTGSNSGFIDPFSEPSQYYGPATDIARASTRSKIGRTRTLSAVEN
RFNRPMVHEDDAFSKVHLQRRGSNDTLVNRRFFISDIDETLRDLLKNEDSDSNCQITIEDTGPKVLRVGTANSNG
YRQTDIRGTYMLSNLLQELTIAKRFGRNQMVLDEARLNENPVKRLRRLISHSFWKNLTRQVGSGNVAEIAQDTKI
DTPDGRIPRIYVPHDDPEQYEFYLEIARKNEAVGGARLDVRYLNEVIDPEYVKSINGTPGLLALATHKVPDGQGG
ITLEGYPYVVPGGRFNELYGWDSYMMALGLLKDGMLDLARGMVENFIYEIRHYGMILNANRSYYLGRSQPPFLTD
FGIMVYEAMVEHAHQNEEDTEKSLQDAEYFLRRTFCAAICEYKTVWCCEPRLDKTTGLSCYHPQGVGIPPETEST
HFTTLLTPYAKKHGVSIEEFYELYNNHTIVEPELDEYFLHDRAVRESGHDTSYRLEGICASLATVDLNALLYKYE
VDIAKIIKEHFNDELVTENSVEHSAEWTKRAELRKERMDKYLWNEEEGIYFDYNLKLKKQHRYESVTTFWPMWAG
CSSQEQAQLMVDKSLSKFEEFGGLVAGTLASRGRVGLERPSRQWDYPYGWAPQQILAWVGLDKYGFRGHTKRLCY
RWLYLMTKAFVDYNGIVVEKYDVTTGTSPHKVDAEYGNQGADFKGVATEGFGWVNASFILGLTYLDVQGIRAIGA
VTSPDVFFRKLKPWERASYGLRPCNEIVKSN

14 β-mannosyltransferases

3 Pichia pastoris homologs of the AMR2 gene (chr4_0450)
Of all 3 the amino acid sequence is patented (US 7465577) and published (Identification of a new family of genes involved in beta-1,2-mannosylation of glycans in Pichia pastoris and Candida albicans. Mille C, Bobrowicz P, Trinel PA, Li H, Maes E, Guerardel Y, Fradin C, Martínez-Esparza M, Davidson RC, Janbon G, Poulain D, Wildt S.)

SEQ0210

5' region
taccatgaacattcgtttgggcttaccattgttgccgataactgttgcattctgtaagat
ggtgatttccagcgttcgaaaagcctgtagtacccagaattttgatcaaagtgtcctctt
agccgtggcatccgcggctgtcatcttgatccttcttcgcatgagcacttacgtgatgct
gataagatgggcacgccacaacactagggccaagacctcgcctagtggtggcgactatct
aaagggctcacctaacacctctttagcggggattgaagacattcatacacgaagtctgct
gtatgggtcagacgagaaaatgccgcaatctaaagaagaacaaaggtcccacaatgcgaa
tcgtgatggtgataaatcattggctcaggtcagtagatttgaaatgcatgataaacgcat
atggtgagagccgttctgcacaactagatgttttcgagcttcgcattgtttcctgcagct
cgactattgaattaagatttccggatatctccaatctcacaaaaacttatgttgaccacg
tgctttcctgaggcgaggtgttttatatgcaagctgccaaaaatggaaaacgaatggcca
tttttcgcccaggcaaattattcgattactgctgtcataaagacagtgttgcaaggctca
catttttttttaggatccgagataaagtgaatacaggacagcttatctctatatcttgta
ccattcgtgaatcttaagagttcggttaggggactctagttgagggttggcactcacgt
atggctgggcgcagaaataaaattcaggcgcagcagcacttatcgatgcatgcaaggcga
gaaaaataaagaacaaaaacacaccttgctaagaccacaacctttcaaattttgagattg
ttattctttcatcctaaaacaccaccgtcctatctcttgggaacgtacatatcattgagc
ttggttcattgatacatcactgtatctaactctccttttt
ORF
ATGGTCGATCTGTTTCAGTGGTTGAAATTTTATAGCATGAGAAGACTCGGTCAGGTTGCTATAACCTTGGTTCTT
TTGAACCTGTTTGTATTTTTGGGCTACAAGTTCACTCCTAGCACCGTGATAGGTTCACCCTCCTGGGAACCTGCA
GTGGTGCCCACAGTTTTCAATGAATCATATTTGGATTCGCTTCAGTTCACGGACATAAACGTTGATTCGTTCCTT
TCGGATACTAATGGTAGAATTTCCGTGACGTGTGACAGCTTAGCTTATAAAGGGCTAGTCAAGACTAGTAAGAAG
AAGGAACTGGACTGCGATATGGCCTATATCAGGAGGAAAATTTTTTCATCCGAGGAATACGGAGTACTGGCCGAT
TTGGAGGCTCAGGACATCACAGAGGAACAGAGAATAAAAAAGCACTGGTTTACCTTCTATGGAAGTTCTGTCTAT
CTTCCCGAACACGAAGTCCACTATCTTGTTCGTCGTGTTTTGTTCAGCAAGGTGGGAAGAGCAGACACGCCGGTG
ATCTCGCTTTTAGTTGCCCAACTTTACGATAAAGACTGGAATGAGTTAACGCCTCATACGCTGGAGATAGTGAAC
CCAGCCACAGGCAATGTCACGCCCCAAACATTCCCTCAGCTCATACACGTCCCCATTGAGTGGAGCGTTGATGAT
AAATGGAAAGGTACGGAGGATCCTAGAGTTTTCCTCAAGCCTTCAAAGACAGGCGTATCAGAACCCATTGTGCTA
TTTAACTTGCAGTCCTCTTTGTGTGACGGTAAGCGTGGAATGTTTGTGACTTCTCCTTTTCGGTCCGACAAAGTA
AATCTTCTTGATATTGAAGATAAGGAAAGACCAAATTCAGAAAAGAATTGGTCCCCATTTTTCTTGGACGATGTT
GAAGTTTCCAAATACAGCACTGGATATGTGCACTTTGTGTACTCTTTCAATCCTTTGAAAGTTATAAAATGCAGC
TTGGACACAGGAGCCTGTCGTATGATCTACGAAAGTCCAGAGGAAGGCCGTTTTGGCTCAGAGTTGCGAGGTGCC

ACACCTATGGTGAAACTACCTGTTCACCTTTCGCTTCCTAAAGGTAAGGAGGTCTGGGTTGCTTTCCCCAGAACT
AGACTACGAGACTGTGGATGTTCAAGAACAACCTATCGACCAGTATTGACGCTTTTTGTCAAGGAGGGCAATAAA
TTCTACACAGAATTGATTTCCAGTTCCATTGACTTTCACATTGACGTATTATCATACGATGCAAAGGGAGAGTCA
TGTTCTGGATCGATTTCAGTGTTGATCCCCAATGGTATTGACAGCTGGGATGTAAGCAAGAAGCAAGGAGGAAAG
TCTGATATCTTGACTCTGACCTTGAGTGAGGCAGACAGAAACACGGTGGTTGTTCATGTAAAGGGATTGTTAGAT
TACTTGCTAGTGCTAAACGGTGAGGGACCTATACATGATAGTCACTCTTTCAAGAATGTGCTGTCTACAAATCAT
TTTAAAAGTGACACCACTCTGTTGAACTCAGTCAAGGCTGCTGAATGTGCTATTTTTTCGAGTAGGGACTACTGC
AAGAAGTACGGGGAGACCCGTGGTGAGCCTGCCAGATATGCAAAACAGATGGAGAATGAAAGGAAGGAAAAGGAG
AAAAAAGAGAAAGAAGCCAAGGAGAAGCTGGAAGCTGAGAAAGCTGAAATGGAAGAAGCTGTAAGAAAGGCACAG
GAAGCCATTGCACAGAAGGAGAGAGAAAAAGAGGAAGCCGAACAGGAGAAGAAAGCACAGCAGGAAGCAAAGGAG
AAGGAAGCAGAGGAAAAAGCGGCAAAGGAAAAGGAAGCAAAGGAGAATGAAGCAAAGAAGAAGATAATAGTGGAG
AAACTTGCAAAGGAGCAGGAGGAGGCTGAAAAGTTAGAGGCAAAGAAGAAACTGTATCAATTGCAGGAAGAAGAA
AGAAGTTAA

Downstream

cgggcagaaagactgaactaccataaaacgtaaaccagtgtttttcaacctttgtttta
cagtcaaggctttatcttctaagtcttttatggcggagtaaacgtcaatgtccaaaatct
gagctcttatgaactccaaatcgcaagctaagtcatgctttggggtgactttcaacaatc
catgataagacaaatccgtacatcctagcactggagcattgtcgtagatgttggtgacat
aggagtcgacctccatgtcgttgaagcttcccaatgatagcagatcatcataggcagaat
ccttgcctccgtttagggagattccgccaagattattatcattccaagtaaggaggactc
cgatccaaatcacagacaaaacactggcaatacagagcagcaggaagttgagtcgtgttc
tcatccaacagtatcaggcacggtatagaagcggagatgaaatagaccaaccttgcttga
ggctaagggaaaacaaccaagaaaaaaggcgcctgataaattgggccctcaaaaacaata
caccaaggcgcacaggcacctgcattcaattgcgacacactttctcgtttaagcggaaag
aggaaggtttcattccgttgtagggaagtagccacataagtgcacagcataccgtgttgt
gcctcggcaatttgttggtacggtcctatacaatcttatgaagcctcctgctacgctcag
aaatgccaaattgctctctgagacaataaatccgcgtagttttgaaatttgtagtgaaaa
tctaggcccgctcgcaggcgagattgacattagtagctgcatagggctgaagcatactgt
gcttgcatcacataatcttggaattacattcatttattagacaaaaatttaaattaagtt
attttgtggagcccaaacttttgagtttttgtttttatttataactgtgaattctacga
caatattaccgttttcaggccattacggtaactctagctttt

AA

MVDLFQWLKFYSMRRLGQVAITLVLLNLFVFLGYKFTPSTVIGSPSWEPAVVPTVFNESYLDSLQFTDINVDSFL
SDTNGRISVTCDSLAYKGLVKTSKKKELDCDMAYIRRKIFSSEEYGVLADLEAQDITEEQRIKKHWFTFYGSSVY
LPEHEVHYLVRRVLFSKVGRADTPVISLLVAQLYDKDWNELTPHTLEIVNPATGNVTPQTFPQLIHVPIEWSVDD
KWKGTEDPRVFLKPSKTGVSEPIVLFNLQSSLCDGKRGMFVTSPFRSDKVNLLDIEDKERPNSEKNWSPFFLDDV
EVSKYSTGYVHFVYSFNPLKVIKCSLDTGACRMIYESPEEGRFGSELRGATPMVKLPVHLSLPKGKEVWVAFPRT
RLRDCGCSRTTYRPVLTLFVKEGNKFYTELISSSIDFHIDVLSYDAKGESCSGSISVLIPNGIDSWDVSKKQGGK
SDILTLTLSEADRNTVVVHVKGLLDYLLVLNGEGPIHDSHSFKNVLSTNHFKSDTTLLNSVKAAECAIFSSRDYC
KKYGETRGEPARYAKQMENERKEKEKKEKEAKEKLEAEKAEMEEAVRKAQEAIAQKEREKEEAEQEKKAQQEAKE
KEAEEKAAKEKEAKENEAKKKIIVEKLAKEQEEAEKLEAKKKLYQLQEEERS*

SEQ0211

chr4_0471

5' region

atgcagccccaggcgcccgttctgatggcttgatgaccgttgtattgcctgtcactatag
ccaggggtagggtccataaaggaatcatagcagggaaattaaaagggcatattgatgcaa
tcactcccaatggctctcttgccattgaagtctccatatcagcactaacttccaagaagg
acccccttcaagtctgacgtgatagagcacgcttgctctgccacctgtagtcctctcaaaa
cgtcaccttgtgcatcagcaaagactttaccttgctccaatactatgacggaggcaattc
tgtcaaaattctctctcagcaattcaaccaacttgaaagcaaattgctgtctcttgatga
tggagacttttttccaagattgaaatgcaatgtgggacgactcaattgcttcttccagct
cctcttcggttgattgaggaacttttgaaaccacaaaattggtcgttgggtcatgtacat
caaaccattctgtagatttagattcgacgaaagcgttgttgatgaaggaaaaggttggat
acggtttgtcggtctctttggtatggccggtggggtatgcaattgcagtagaagataatt
ggacagccattgttgaaggtagagaaaaggtcagggaacttgggggttatttataccatt
ttaccccacaaataacaactgaaaagtacccattccatagtgagaggtaaccgacggaaa
aagacgggcccatgttctgggaccaatagaactgtgtaatccattgggactaatcaacag
acgattggcaatataatgaaatagttcgttgaaaagccacgtcagctgtcttttcattaa
ctttggtcggacacaacattttctactgttgtatctgtcctactttgcttatcatctgcc
acagggcaagtggatttccttctcgcgcggctgggtgaaaacggttaacgtgaaatgaaa

ttagacacacaacaaatctctcatctcctctcacgtcaa
ORF
ATGTATCATTTAGCTCCACGGAAGAAGCTCCTGATTTGGGGTGGCAGTCTGGGATTTGTACTACTTCTGCTCATA
GTGGCATCATCTCACCAGAGGATCAGATCCACAATTCTCCATCGAACGCCAATTTCAACTTTGCCTGTCATTTCC
CAAGAAGTTATCACTGCGGATTATCATCCAACACTATTGACCGGGTTCATACCGACAGACTCCGATGATAGTGAT
TGTGCTGATTTCAGTCCAAGCGGTGTGATTTACTCGACTGACAAGCTCGTGCTGCACGATTCTCTGAAAGATATC
AGAGATTCTCTTTTGAAAACACAATACAAAGACCTCGTAACTCTTGAAGATGAGGAAAAAATGAATATAGACGAT
ATTTTAAAACGTTGGTATACTCTCTCTGGTTCAAGTGTTTGGATTCCCGGAATGAAAGCGCATTTGGTTGTTTCC
AGAGTTATGTATTTAGGAACCAATGGTAGAAGTGACCCATTAGTATCGTTTGTTAGAGTACAATTATTTGACCCA
GATTTTAATGAGTTGAAGGACATAGCGCTCAAGTTTTCTGATAAGCCAGATGGAACGGTCATATTCCCATATATT
CTGCCTGTAGATATACCGAGAGAAGGAAGTCGATGGTTAGGTCCTGAAGATGCAAAGATTGCTGTAAATCCAGAA
ACGCCAGATGATCCAATAGTGATATTCAATATGCAAAACAGTGTGAACAGAGCCATGTACGGATTTTACCCTTTC
AGACCTGAAAACAAACAAGTCTTGTTTTCTATAAAAGATGAAGAGCCAAGAAAGAAGGAGAAGAACTGGACTCCG
TTTTTTGTCCCTGGCTCACCAACTACCGTTAACTTCGTATATGACCTACAGAAGTTAACAATTCTAAAGTGCTCC
ATCATAACCGGAATCTGTGAAAAAGAATTTGTTAGTGGAGACGATGGGCAGAATCATGGTATCGGAATATTTAGA
GGTGGATCAAACCTTGTCCCCTTTCCTACCAGCTTTACTGATAAAGATGTTTGGGTGGGCTTCCCTAAAACACAC
ATGGAAAGCTGTGGATGCAGCAGTCATATTTACAGGCCTTACTTAATGGTTTTAGTGAGGAAAGGAGATTTCTAC
TACAAGGCGTTCGTCAGTACACCGCTAGATTTTGGAATCGACGTTCGATCGTGGGAAAGTGCAGAGAGCACTAGT
TGCCAAACAGCCAAGAACGTGCTAGCAGTCAACTCCATCTCTAATTGGGATTTACTAGACGATGGTCTGGACAAA
GATTATATGACAATCACTCTCAGCGAGGCAGACGTTGTCAACTCAGTATTGAGAGTCCGTGGTATCGCCAAGTTT
GTAGACAACCTCACAATGGACGATGGATCCACAACCCTTTCCACTTCAAACAAGATTGATGAATGTGCAACTACT
GGCTCCAAACAATACTGTCAAAGGTATGGTGAGCTTCACTAA
Downstream
gccttgggggacttcaagtctttgctagaaactagatgaggtcaggccctcttatggttg
tgtcccaattgggcaatttcactcacctaaaaagcatgacaattatttagcgaaataggt
agtatattttccctcatctcccaagcagtttcgtttttgcatccatatctctcaaatgag
cagctacgactcattagaaccagagtcaagtaggggtgagctcagtcatcagccttcgtt
tctaaaacgattgagttcttttgttgctacaggaagcgccctagggaactttcgcacttt
ggaaatagattttgatgaccaagagcgggagttgatattagagaggctgtccaaagtaca
tgggatcaggccggccaaattgattggtgtgactaaaccattgtgtacttggacactcta
ttacaaaagcgaagatgatttgaagtattacaagtcccgaagtgttagaggattctatcg
agcccagaatgaaatcatcaaccgttatcagcagattgataaactcttggaaagcggtat
cccattttcattattgaagaactacgataatgaagatgtgagagacggcgaccctctgaa
cgtagacgaagaaacaaatctacttttggggtacaatagagaaagtgaatcaagggaggt
atttgtggccataatactcaactctatcattaatgtggttcttttggtagcaaaaatctt
tgttgttttgttcagttcctcactctcattgatggcttcgttagttgactccgtgatgga
tttcttatctactttgatcatatatgtttctaactcttttgctgggaaaagagacaagaa
tgagtatccagttggaaggtcaaggttggagcccttaggagttcttgtcttttccgtaat
cataattgtctctttcatccaggtcggaaatgaatctttgaaaaagctaatcagtggcga
ccgtgatgttgtttctttagataaaaccactatcagcgtc
AA
MYHLAPRKKLLIWGGSLGFVLLLLIVASSHQRIRSTILHRTPISTLPVISQEVITADYHPTLLTGFIPTDSDDSD
CADFSPSGVIYSTDKLVLHDSLKDIRDSLLKTQYKDLVTLEDEEKMNIDDILKRWYTLSGSSVWIPGMKAHLVVS
RVMYLGTNGRSDPLVSFVRVQLFDPDFNELKDIALKFSDKPDGTVIFPYILPVDIPREGSRWLGPEDAKIAVNPE
TPDDPIVIFNMQNSVNRAMYGFYPFRPENKQVLFSIKDEEPRKKEKNWTPFFVPGSPTTVNFVYDLQKLTILKCS
IITGICEKEFVSGDDGQNHGIGIFRGGSNLVPFPTSFTDKDVWVGFPKTHMESCGCSSHIYRPYLMVLVRKGDFY
YKAFVSTPLDFGIDVRSWESAESTSCQTAKNVLAVNSISNWDLLDDGLDKDYMTITLSEADVVNSVLRVRGIAKF
VDNLTMDDGSTTLSTSNKIDECATTGSKQYCQRYGELH*

SEQ0212

chr1-4_0696
5' region
cgctgctttctttatcatggtgcagggtcccacacagaccaaatccaagttattccctaa
gggatggttttcagagagagaagaaactataatggtaaaccggttgttgcgggatgaccc
ctctaagagtgatatgcacaatagacagtctgttacgccaaaagaactgtggaaagtgtt
gggagactacgatctatggcctatatacgcattggccatggtattttccattccccagat
accgataaagagataccttactcttactttgagggcattagagttcactaccactgagat
caatctcttaacaatccctgcttcgtttctggcgggaatcatgtcaattgctatttcatt
agtcagtgagttcttcaatgaaggtttgattatcggtatattgtgtcaattctggttgct
tattatggttatcattgaatacacctctgtggaaaagatatctccctggggacaatatgt
gttgcaactgttcgttgttggtgccccagtcccccaaccggtactaatcggtctatgttc

ccgtaactcatattcggttagaactagaacaataagtgcatcattgttcaacattgtggt
tcaattgtcgaacattgctggtgcttatatctacagggaagacgataagcctttgtacaa
gagaggtaacagacagttaattggtatttctttgggagtcgttgccctctacgttgtctc
caagacatactacattctgagaaacagatggaagactcaaaaatgggagaagcttagtga
agaagagaaagttgcctacttggacagagctgagaaggagaacctgggttctaagaggct
ggacttttgttcgagagttaaactgcataatttttctaagtaaatttcatagttatga
aatttctgcagcttagtgtttactgcatcgtttactgcatcaccctgtaaataatgtgag
ctttttccttccattgcttggtatcttccttgctgctgttt

ORF

ATGAGAATACGGTCGAACGTGTTGCTCCTTAGCACAGCGGGAGCCTTAGCCTTAGTCTGGTTTGCGGTAGTGTTT
TCCTGGGACGACAAGAGTATTTTTGGAATCCCCACACCAGGCCATGCCGTGGCCTCTGCTTACGACAGTTCGGTT
ACCTTGGGGACTTTCAATGACATGGAGGTTGACTCTTATGTCACCAACATCTATGATAATGCTCCAGTGCTAGGT
TGTTACGATTTGTCTTATCATGGCTTGTTGAAAGTTAGTCCTAAGCATGAAATACTTTGCGATATGAAATTTATA
AGAGCCCGAGTCTTGGAGACCGAAGCTTATGCTGCTTTGAAGGATTTAGAGCACAAGAAATTGACTGAAGAAGAA
AAAATTGAGAAACATTGGTTTACGTTTTATGGTAGTTCAGTATTTCTGCCCGACCACGACGTGCATTACTTGGTT
CGACGGGTGGTCTTTTCAGGAGAAGGAAAAGCAAATAGACCAATAACGTCTATTCTGGTTGCTCAAATATACGAT
AAAAACTGGAACGAACTGAATGGCCATTTTCTGAACGTTCTGAATCCAAATACTGGAAAGCTCCAGCACCATGCA
TTTCCGCAAGTTCTGCCTATTGCAGTAAATTGGGACAGAAATAGCAAGTATCGTGGACAGGAAGATCCTAGAGTT
GTTTTGAGAAGGGGTCGTTTTGGTCCTGATCCTTTGGTGATGTTCAACACTTTGACGCAAAATAATAAACTTAGA
AGACTTTTCACTATCTCTCCATTTGACCAATACAAAACAGTCATGTACAGAACTAACGCCTTTAAGATGCAGACC
ACTGAAAAGAATTGGGTCCCATTTTTCTTGAAAGACGACCAGGAATCTGTCCATTTTGTTTACTCGTTCAATCCT
CTGAGAGTACTCAACTGCAGTCTTGATAACGGTGCATGTGATGTTCTATTTGAGTTACCACATGATTTTGGCATG
TCTTCCGAGCTACGTGGTGCCACTCCTATGCTCAATCTTCCTCAGGCAATCCCGATGGCAGACGACAAAGAAATT
TGGGTTTCATTCCCAAGAACGAGAATATCAGATTGCGGGTGTTCTGAAACAATGTACAGGCCAATGTTAATGCTT
TTTGTTAGAGAAGGAACAAACTTTTTTGCTGAGCTTCTCTCCTCTTCTATTGATTTTGGTTTGGAGGTTATACCG
TATACCGGAGACGGATTACCATGCAGTTCAGGCCAGTCTGTTTTAATTCCCAATAGTATTGATAACTGGGAGGTT
ACGGGTAGCAATGGAGAAGATATTTTGTCATTGACATTTAGTGAGGCGGATAAAAGTACCTCTGTGGTTCACATC
AGAGGATTATACAAATACCTGTCTGAATTGGACGGTTATGGAGGTCCAGAAGCGGAAGATGAACACAATTTTCAA
CGTATCCTTAGTGATTTGCATTTTGACGGTAAGAAAACGATAGAAAACTTCAAAAAAGTACAATCATGTGCATTG
GATGCTGCCAAAGCCTATTGTAAAGAATATGGTGTCACTCGTGGGGAGGAAGATAGACTAAAGAACAAAGAGAAG
GAAAGGAAAATAGAAGAAAAAAGAAAGAAGGAGGAAGAGAGGAAGAAGAAGGAAGAGGAAAAGAAGAAGAAGGAA
GAAGAGGAAAAGAAGAAAAAGGAAGAGGAAGAGGAAGAGGAAAAGAGGCTTAAAGAATTGAAAAAGAAGCTTAAA
GAACTACAAGAAGAGCTTGAAAAGCAGAAAGATGAAGTGAAGGATACAAAGGCGAAGTAG

Downstream

gccttaataacacttgagtttttggtatgaccccatcctgatcgccaaatctataatga
ttttcgtggatgtgaaaagtcaaagagggagggaatttttcgtacttgtgaagtatctcc
tcactcgtttcggcaaatttgtaatttttgagtttctgatgcagtactttctgttgcaat
aatctttgttgttgtggagtcaagttctgagggtttatcctctggccattgctcagggtc
acagacccattgattggtctttgttgagcctgaggctgtgcggtttgcatatttttcgcc
cccacttggcttgactggatttgactcatttcggagtatttattagatgggggccagtgc
ttcaactctgaggatttatgatacggagccttcgtgttgtgaaacttagtttaccttgag
tatgtgacagacgaaaaaagatcgcgaagtgttggttccggagatgttcagcggtcacat
gatcacaacgttttacctctttctatgtaaaacggtttttaccccggttcttgaaaagac
tctatgtagaaaccgagctgctgtatgcgtaaggttgtatggcggaaggtaaacggagat
ttcatgtttaagaggccttgcgacatctcatgtttttgatccttatattatggactatgc
aaaaacggaaatgtcaatcaaatactggaaaccccatgctatctgtaagacctgtgtgtt
gaaatacagtcgatgtcaataatcgtgacagcgatattagcagattaaattataaaagga
ttatagacagcctagaaagatatagtagcaagaaaccgattacggaatcaatctgcccta
tggtgtttaatttctgcattctttttcggtctttcagtttcagtttcgttaaatctatt
ctactattttcttaatctctccagaataccttgctcttccctttgcttctctctcccttc
gttcttcgttctccgttctcctttctccgttttccgtctc

AA

MRIRSNVLLLSTAGALALVWFAVVFSWDDKSIFGIPTPGHAVASAYDSSVTLGTFNDMEVDSYVTNIYD
NAPVLGCYDLSYHGLLKVSPKHEILCDMKFIRARVLETEAYAALKDLEHKKLTEEEKIEKHWFTFYGSS
VFLPDHDVHYLVRRVVFSGEGKANRPITSILVAQIYDKNWNELNGHFLNVLNPNTGKLQHHAFPQVLPI
AVNWDRNSKYRGQEDPRVVLRRGRFGPDPLVMFNTLTQNNKLRRLFTISPFDQYKTVMYRTNAFKMQ
TTEKNWVPFFLKDDQESVHFVYSFNPLRVLNCSLDNGACDVLFELPHDFGMSSELRGATPMLNLPQAIP
MADDKEIWVSFPRTRISDCGCSETMYRPMLMLFVREGTNFFAELLSSSIDFGLEVIPYTGDGLPCSSGQS
VLIPNSIDNWEVTGSNGEDILSLTFSEADKSTSVVHIRGLYKYLSELDGYGGPEAEDEHNFQRILSDLHFD
GKKTIENFKKVQSCALDAAKAYCKEYGVTRGEEDRLKNKEKERKIEEKRKKEEERKKKEEEKKKKEEE
EKKKKEEEEEEEKRLKELKKKLKELQEELEKQKDEVKDTKAK*

WHAT IS CLAIMED IS:

1. An isolated nucleic acid which encodes a protein as set forth in any one of SEQ0001-0025, 0027-0126, 0128-0165, 0172-0174, 0176-0200, or 0202-0212, or a protein substantially homologous thereto.

2. The isolated nucleic acid of claim 1, comprising a nucleotide sequence that is identical or substantially homologous to the coding sequence of an ORF as set forth in any one of SEQ0001-0025, 0027-0126, 0128-0165, 0172-0174, 0176-0200, or 0202-0212.

3. An isolated protein comprising an amino acid sequence as set forth in any one of SEQ0001-0025, 0027-0126, 0128-0165, 0172-0174, 0176-0200, or 0202-0209, or an amino acid sequence substantially homologous thereto.

4. An isolated nucleic acid, which encodes a signal peptide as set forth in any one of SEQ0001-0025 or 0027-0054.

5. An isolated peptide comprising an amino acid sequence identical with a signal peptide as set forth in any one of SEQ0001-0025 or 0027-0054.

6. A library of expression vectors, wherein each vector of the library encodes one and only one of the signal peptides identified in SEQ0001-0025 and 0027-0054, and each of the signal peptides identified in SEQ0001-0025 and 0027-0054 is encoded by one vector in the library, each vector comprising from 5' to 3', a promoter, the coding sequence of a signal peptide, and an intron sequence comprising a cloning site for insertion of a coding sequence of a heterologous protein.

7. A library of expression vectors, wherein each vector of the library encodes one and only one of the signal peptides identified in SEQ0001-0025 and 0027-0054, and each of the signal peptides identified in SEQ0001-0025 and 0027-0054 is encoded by one vector in the library, each vector comprising from 5' to 3', a promoter, the coding sequence of a signal peptide, fused in frame to a coding sequence of a heterologous protein.

8. An isolated promoter, comprising a promoter sequence identified in any one of SEQ0060-0085, SEQ0096-0124, SEQ0125-0128, or SEQ0169-0178.

9. An expression vector comprising the isolated promoter of claim 8, operably linked to a heterologous coding sequence.

10. An expression vector for enhanced expression of a glycosylation precursor synthesis enzyme or transporter, comprising, from 5' to 3', a promoter, operably linked to a coding sequence as set forth in any one of SEQ0129-0132, and a transcription termination sequence.

11. The expression vector of claim 10, wherein the protein encoded by the coding sequence is modified to include an ER or Golgi localization signal.

12. A genetically engineered methylotrophic yeast strain, capable of overexpressing a glycosylation precursor synthesis enzyme or transporter as set forth in any one of SEQ0129-0132.

13. A *Pichia pastoris* strain, in which at least one of the genes encoding a mannosyl transferase as set forth in SEQ0133-0137 or 0210-0212 has been inactivated.

14. A *Pichia pastoris* strain, in which at least one of the genes as set forth in SEQ0153-0165 has been inactivated.

15. A genetically engineered methylotrophic yeast strain, capable of overexpressing any one of the genes set forth in SEQ0166-0168.

16. A *Pichia pastoris* strain, in which at least one of the protease-encoding genes as set forth in SEQ0179 and SEQ0181-0186 has been inactivated.

17. A genetically engineered methylotrophic yeast strain, capable of overexpressing the protease inhibitor as set forth in SEQ0180.

18. An expression vector for overexpression of chaperones, comprising, from 5' to 3', a promoter, operably linked to a coding sequence as set forth in any one of SEQ0187-SEQ0200, and a transcription termination sequence.

19. A collection of expression vectors, each vector in said collection comprising from 5' to 3', a promoter, operably linked to a different coding sequence as set forth in any one of SEQ0187-SEQ0200, and a transcription termination sequence.

20. A genetically engineered methylotrophic yeast strain, capable of overexpressing at least one chaperone set forth in any one of SEQ0187-SEQ0200.

21. The strain of claim 20, capable of overexpressing multiple chaperons.

22. An isolated nucleic acid molecule comprising the nucleotide sequence as set forth in SEQ0200.

23. An vector for mediating multi-copy integration of a heterologous coding sequence, comprising the nucleotide sequence as set forth in SEQ0200, operably linked to an expression cassette, wherein said expression cassette comprises a promoter placed in operable linkage to said heterologous sequence and a selectable marker gene.

24. A *Pichia pastoris* strain, comprising an expression cassette integrated in multiple copies at native 5SrRNA loci, wherein said expression cassette comprises a promoter in operable linkage to said heterologous sequence and a selectable marker gene.

A

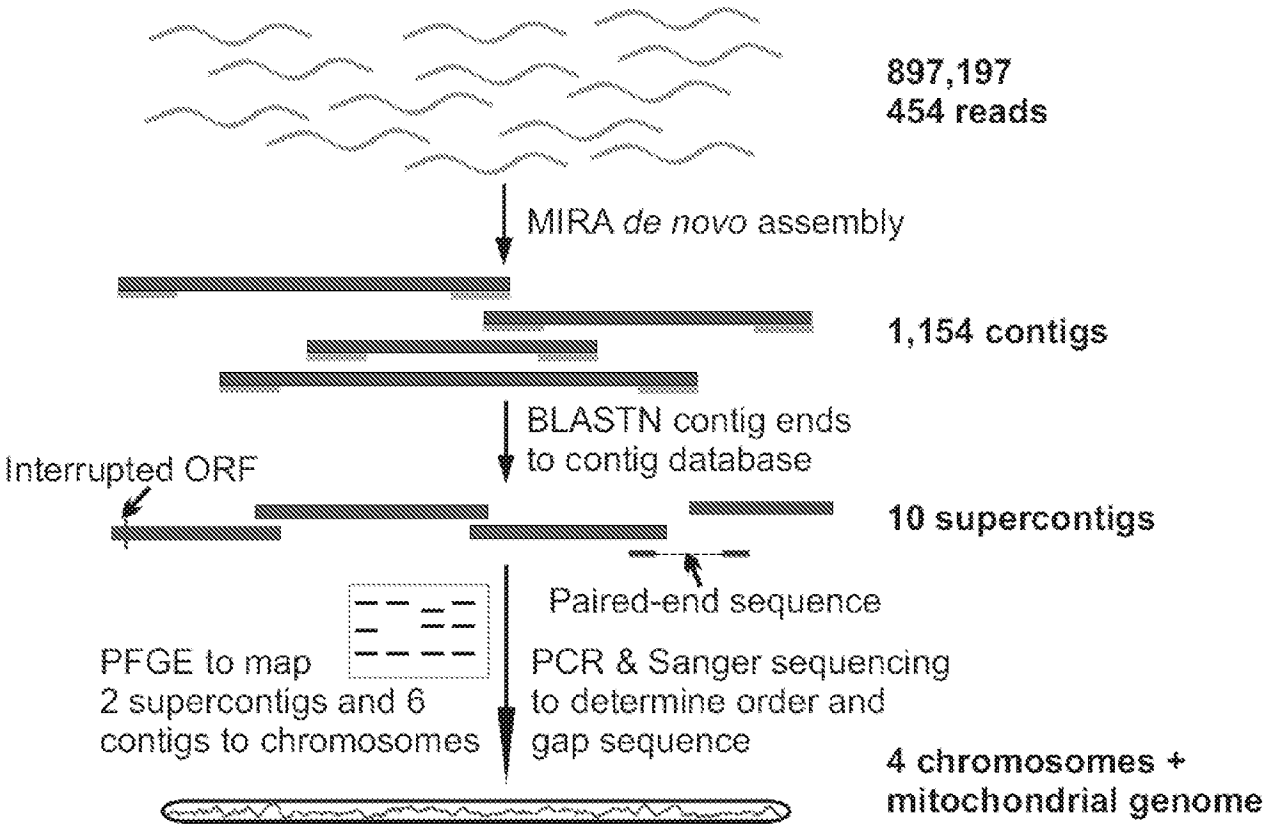
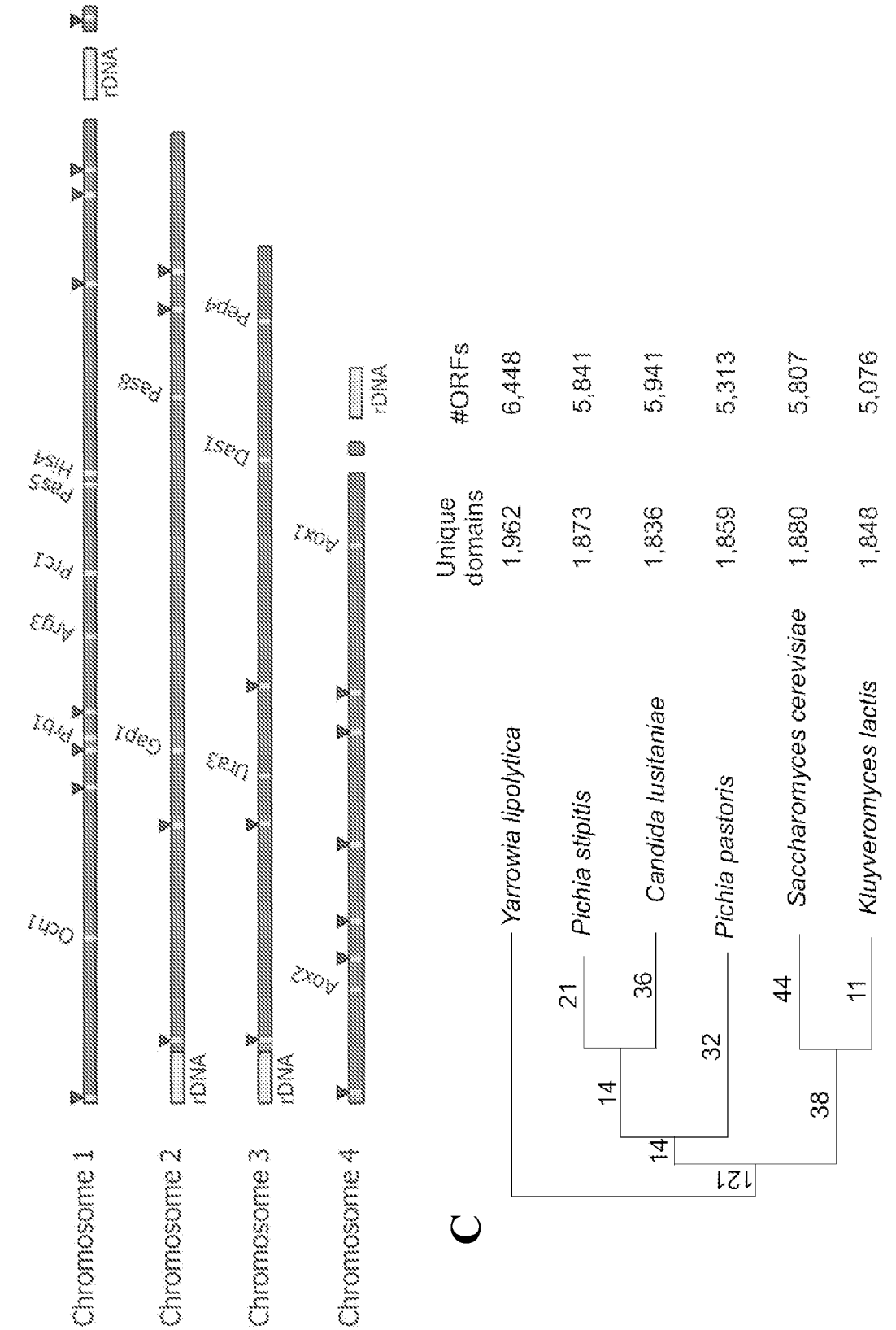

897,197
454 reads
MIRA *de novo* assembly
1,154 contigs
BLASTN contig ends
to contig database
Interrupted ORF
10 supercontigs
Paired-end sequence
PFGE to map
2 supercontigs and 6
contigs to chromosomes
PCR & Sanger sequencing
to determine order and
gap sequence
4 chromosomes +
mitochondrial genome

FIGURE 1A

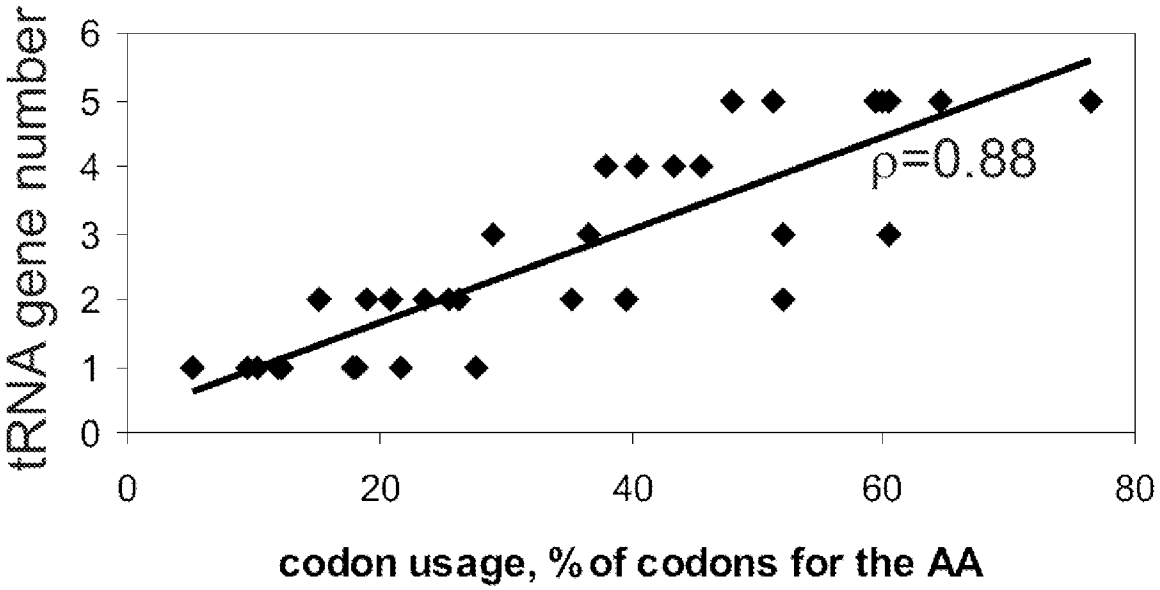


FIGURES 1B-1C

A

	U	C	A	G	
U	Phe 57,76 42,24	Ser 25,98 17,41 20,84 9,46	Tyr 51,41 48,59	Cys 62,19 37,81	U
	Leu 17,86 29,06 17,07 8,42 12,32 15,27		Stop 40,04 34,02	25,94 Trp 100,00	C A G
C		Pro 33,89 18,05 37,85 10,22	His 61,77 38,23	Arg 14,05 5,06 11,96 5,25	U
			Gln 60,36 39,64		C A G
A	Ile 46,65 29,75 23,60	Thr 36,25 23,77 27,73 12,26	Asn 53,46 46,54	Ser 15,87 10,44	U C
	Met 100,00		Lys 51,97 48,03	Arg 45,55 18,13	A G
G	Val 38,62 21,69 17,79 21,69	Ala 40,12 24,49 27,54 7,85	Asp 61,74 38,26	Gly 35,76 15,57 36,47 12,19	U
			Glu 59,49 40,51		C A G

B



FIGURES 2A-2B

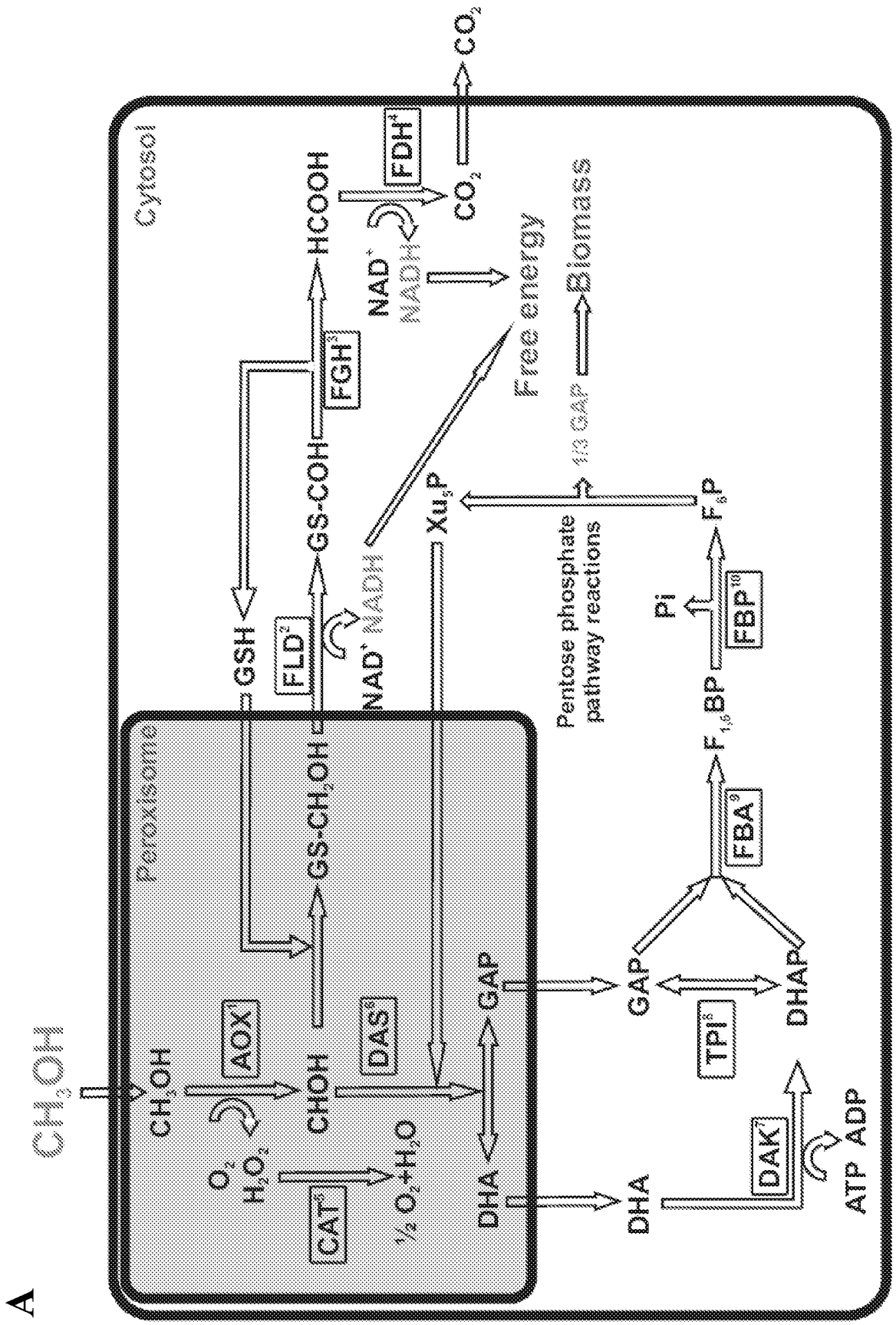

A
CH_3OH
Cytosol
Peroxisome
CH_3OH
AOX[1]
O_2
H_2O_2
CAT[2]
$\frac{1}{2}O_2+H_2O$
$CHOH$
DAS[4]
DHA
GAP
GSH
FLD[3]
$GS\text{-}CH_2OH$
$GS\text{-}COH$
NAD` NADH
Xu_5P
HCOOH
FGH[3]
FDH[4]
CO_2
CO_2
NAD+ NADH
Free energy
Biomass
1/3 GAP
Pentose phosphate pathway reactions
F_6P
Pi
FBP[10]
$F_{1,6}BP$
FBA[9]
GAP
TPI[5]
DHAP
DHA
DAK[7]
ATP ADP

FIGURE 3A

B

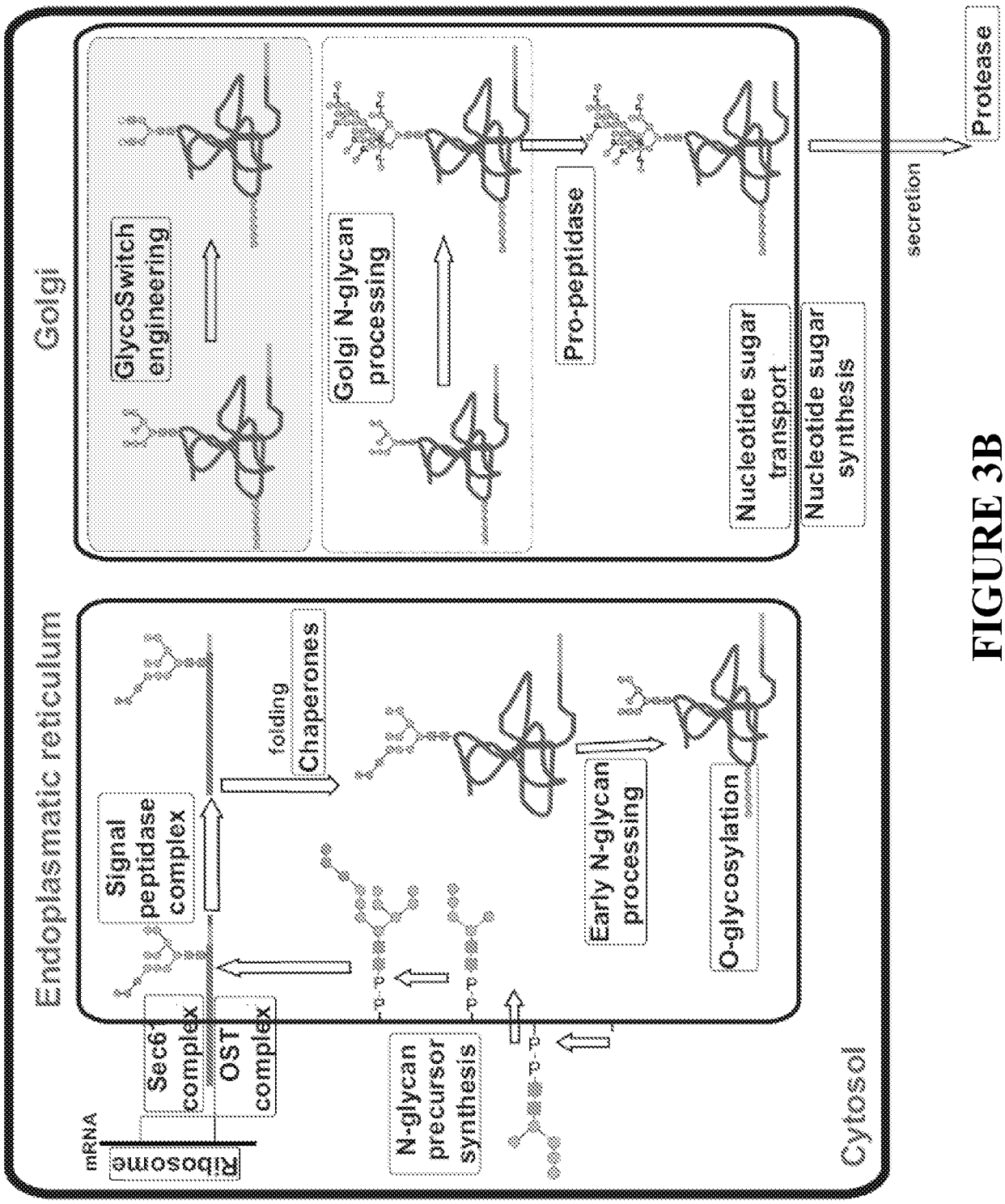

Golgi
GlycoSwitch engineering
Golgi N-glycan processing
Pro-peptidase
Nucleotide sugar transport
Nucleotide sugar synthesis
secretion
Protease
Endoplasmatic reticulum
Signal peptidase complex
folding
Chaperones
Early N-glycan processing
O-glycosylation
Sec6 complex
OST complex
mRNA
Ribosome
N-glycan precursor synthesis
Cytosol

FIGURE 3B

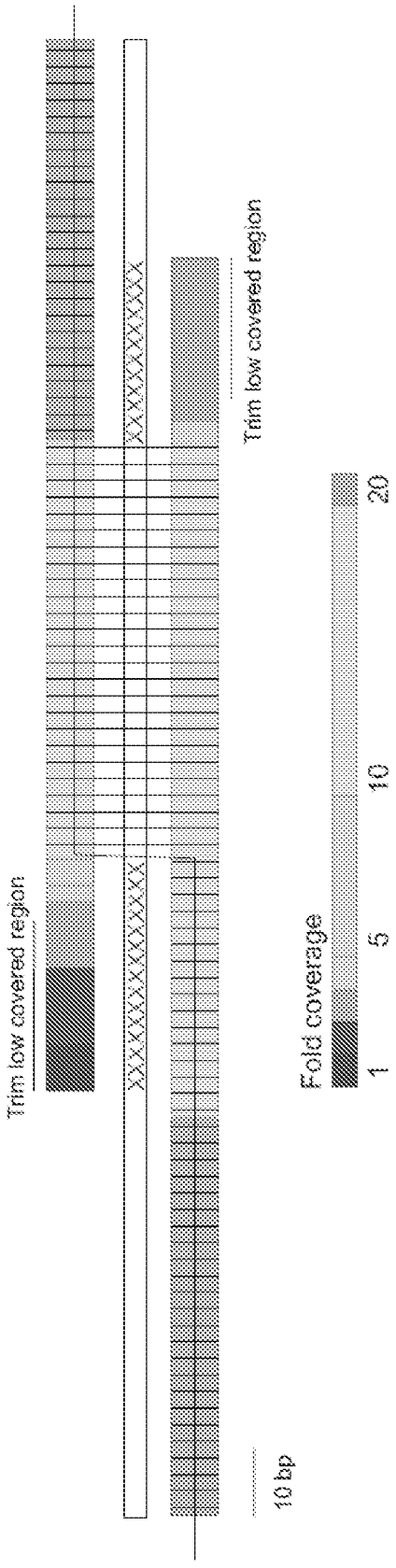
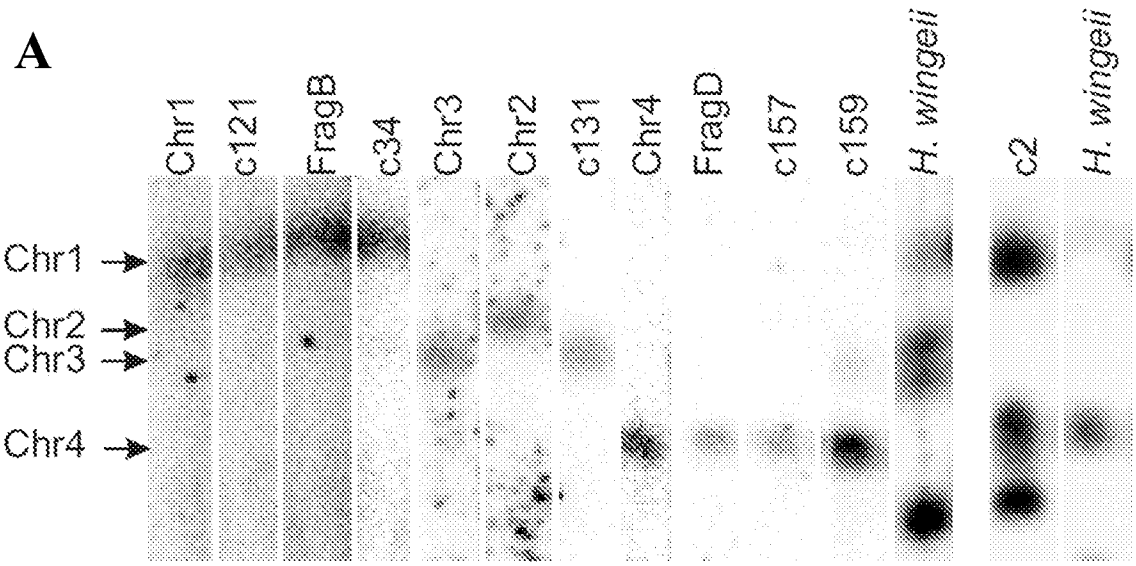
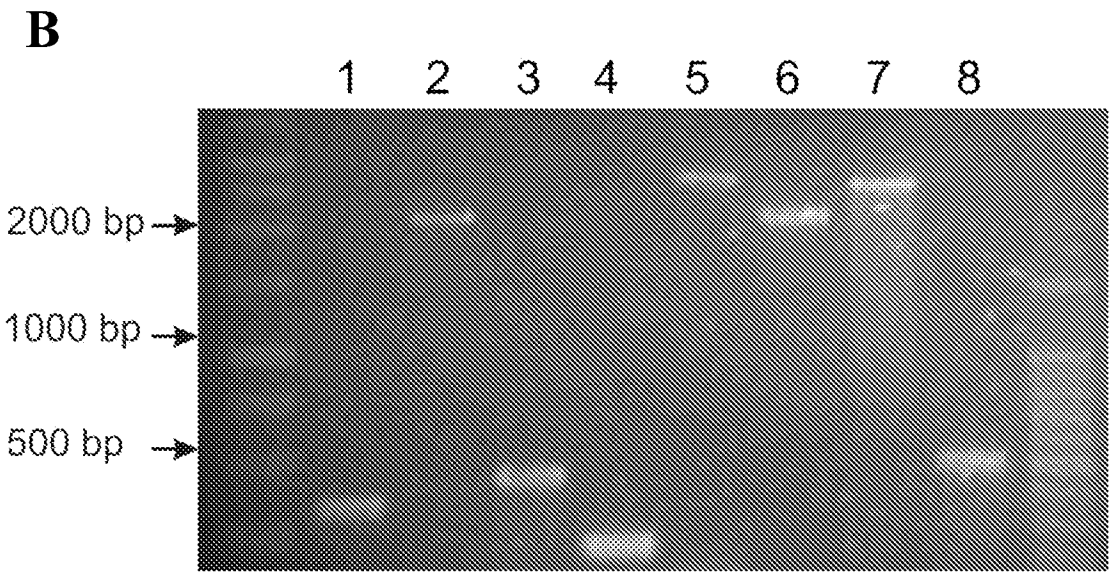

Trim low covered region
Trim low covered region
Fold coverage
1
5
10
20
10 bp

FIGURE 4





FIGURES 5A-5B

C

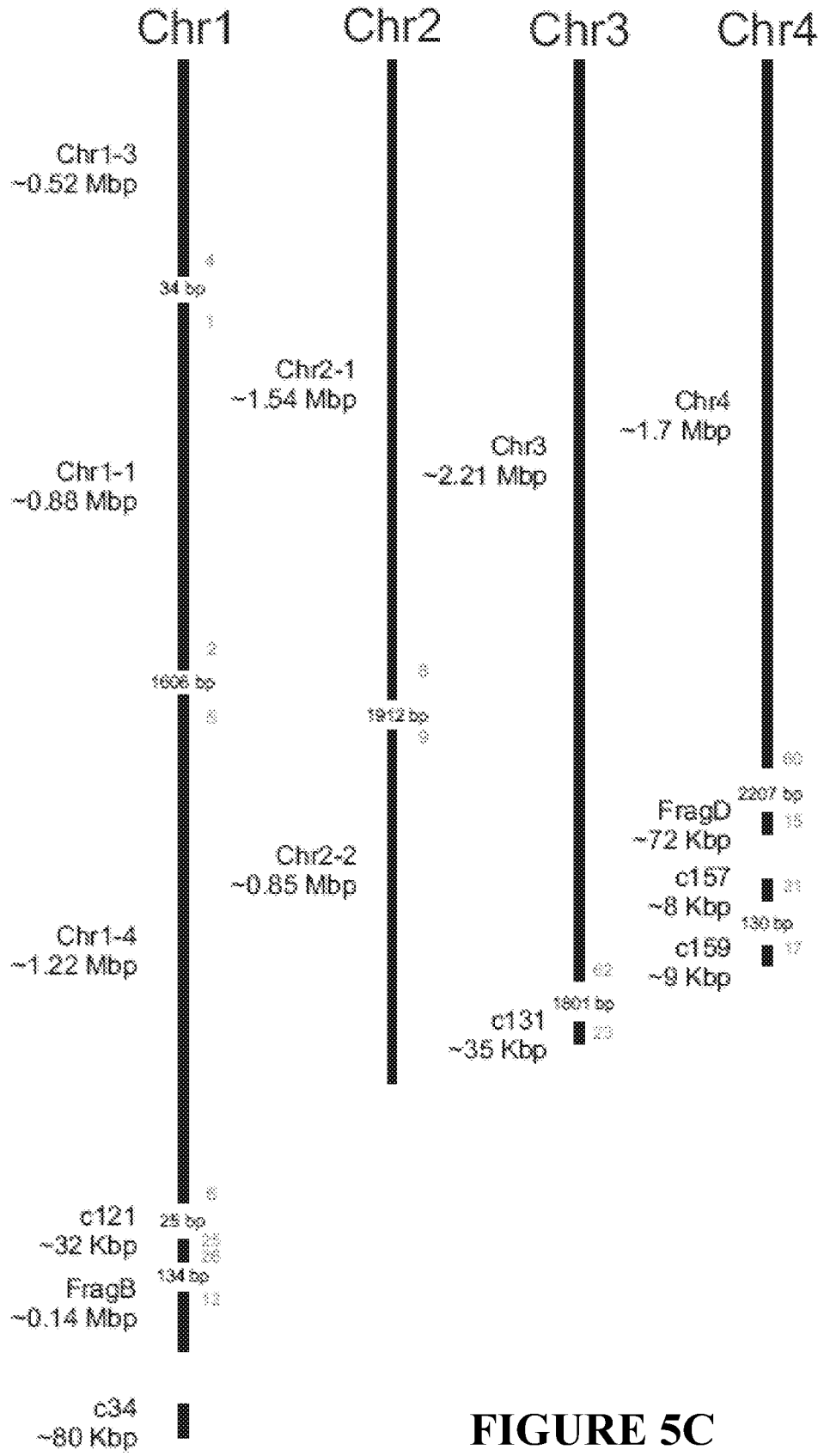

Chr1
Chr2
Chr3
Chr4
Chr1-3
~0.52 Mbp
34 bp
Chr2-1
~1.54 Mbp
Chr3
~2.21 Mbp
Chr4
~1.7 Mbp
Chr1-1
~0.88 Mbp
1608 bp
1912 bp
2207 bp
FragD
~72 Kbp
Chr2-2
~0.85 Mbp
c157
~8 Kbp
130 bp
c159
~9 Kbp
Chr1-4
~1.22 Mbp
1801 bp
c131
~35 Kbp
c121
~32 Kbp
25 bp
134 bp
FragB
~0.14 Mbp
c34
~80 Kbp

FIGURE 5C

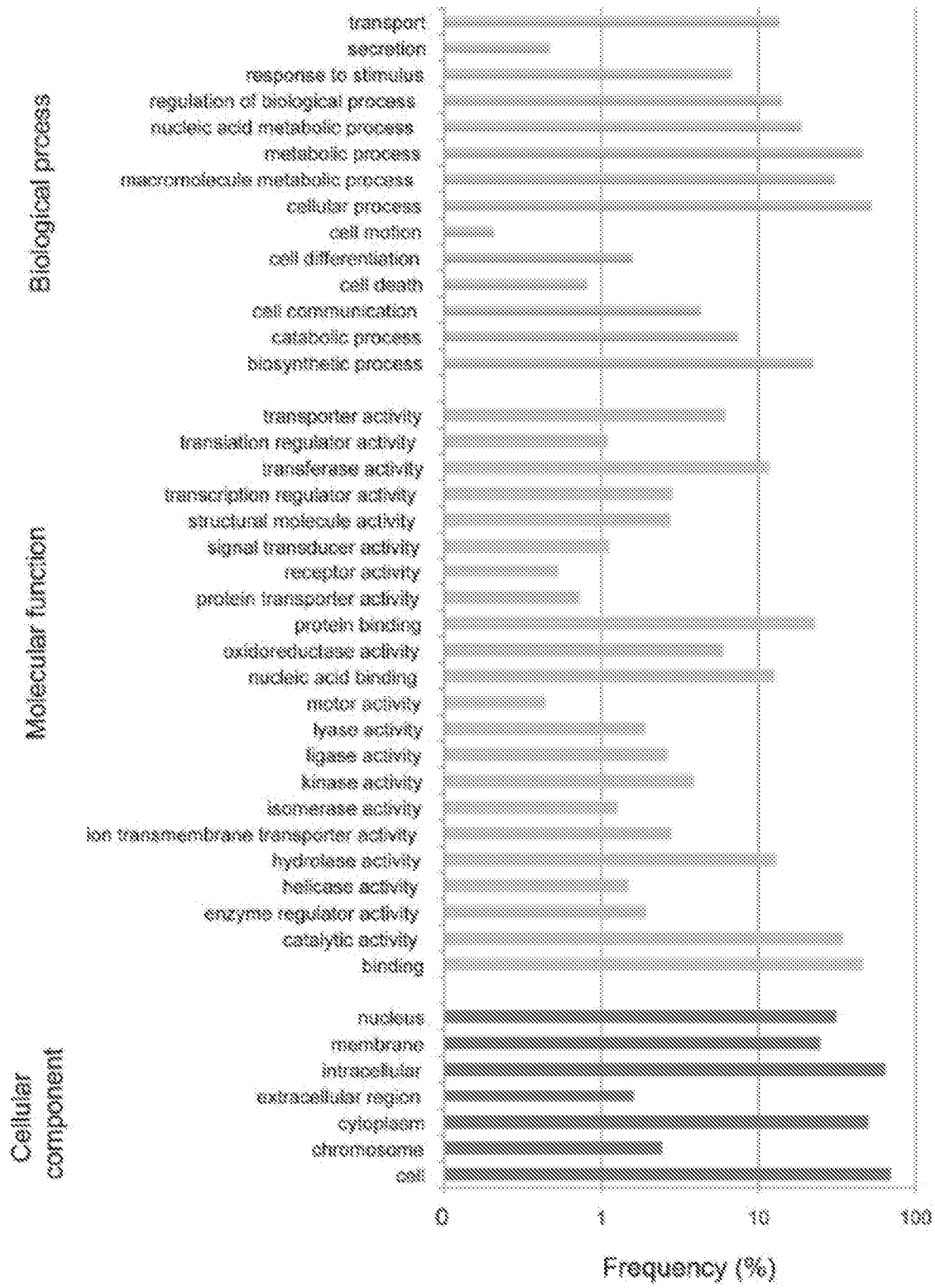

Biological process
transport
secretion
response to stimulus
regulation of biological process
nucleic acid metabolic process
metabolic process
macromolecule metabolic process
cellular process
cell motion
cell differentiation
cell death
cell communication
catabolic process
biosynthetic process
Molecular function
transporter activity
translation regulator activity
transferase activity
transcription regulator activity
structural molecule activity
signal transducer activity
receptor activity
protein transporter activity
protein binding
oxidoreductase activity
nucleic acid binding
motor activity
lyase activity
ligase activity
kinase activity
isomerase activity
ion transmembrane transporter activity
hydrolase activity
helicase activity
enzyme regulator activity
catalytic activity
binding
Cellular component
nucleus
membrane
intracellular
extracellular region
cytoplasm
chromosome
cell
0
1
10
100
Frequency (%)

FIGURE 6

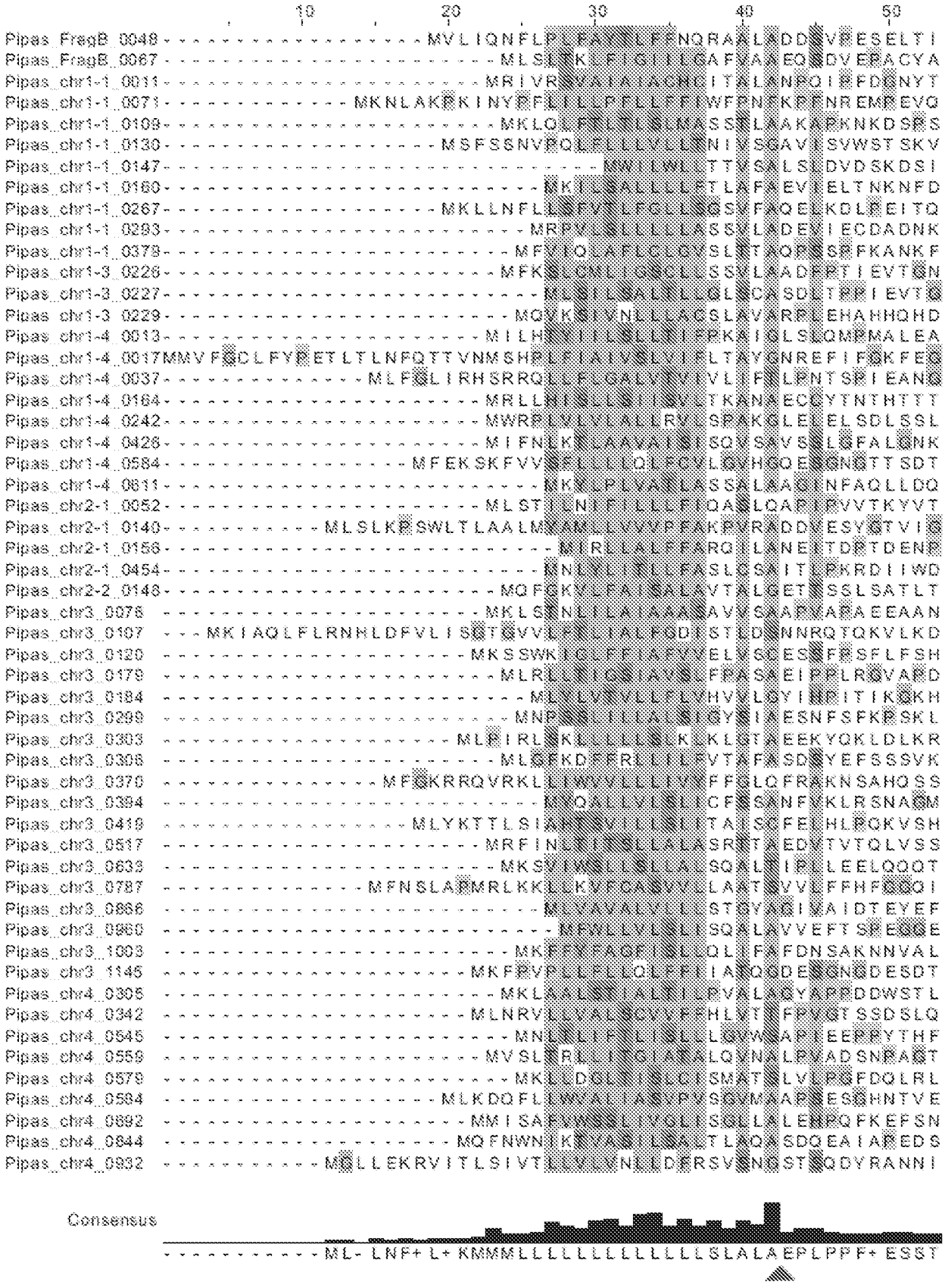

10
20
30
40
50
Pipas_FragB_0048
Pipas_FragB_0067
Pipas_chr1-1_0011
Pipas_chr1-1_0071
Pipas_chr1-1_0109
Pipas_chr1-1_0130
Pipas_chr1-1_0147
Pipas_chr1-1_0160
Pipas_chr1-1_0267
Pipas_chr1-1_0293
Pipas_chr1-1_0378
Pipas_chr1-3_0226
Pipas_chr1-3_0227
Pipas_chr1-3_0229
Pipas_chr1-4_0013
Pipas_chr1-4_0017
Pipas_chr1-4_0037
Pipas_chr1-4_0164
Pipas_chr1-4_0242
Pipas_chr1-4_0426
Pipas_chr1-4_0584
Pipas_chr1-4_0611
Pipas_chr2-1_0052
Pipas_chr2-1_0140
Pipas_chr2-1_0156
Pipas_chr2-1_0454
Pipas_chr2-2_0148
Pipas_chr3_0076
Pipas_chr3_0107
Pipas_chr3_0120
Pipas_chr3_0179
Pipas_chr3_0184
Pipas_chr3_0299
Pipas_chr3_0303
Pipas_chr3_0306
Pipas_chr3_0370
Pipas_chr3_0394
Pipas_chr3_0419
Pipas_chr3_0517
Pipas_chr3_0633
Pipas_chr3_0787
Pipas_chr3_0866
Pipas_chr3_0960
Pipas_chr3_1003
Pipas_chr3_1145
Pipas_chr4_0305
Pipas_chr4_0342
Pipas_chr4_0545
Pipas_chr4_0559
Pipas_chr4_0579
Pipas_chr4_0594
Pipas_chr4_0692
Pipas_chr4_0844
Pipas_chr4_0932
Consensus

FIGURE 7

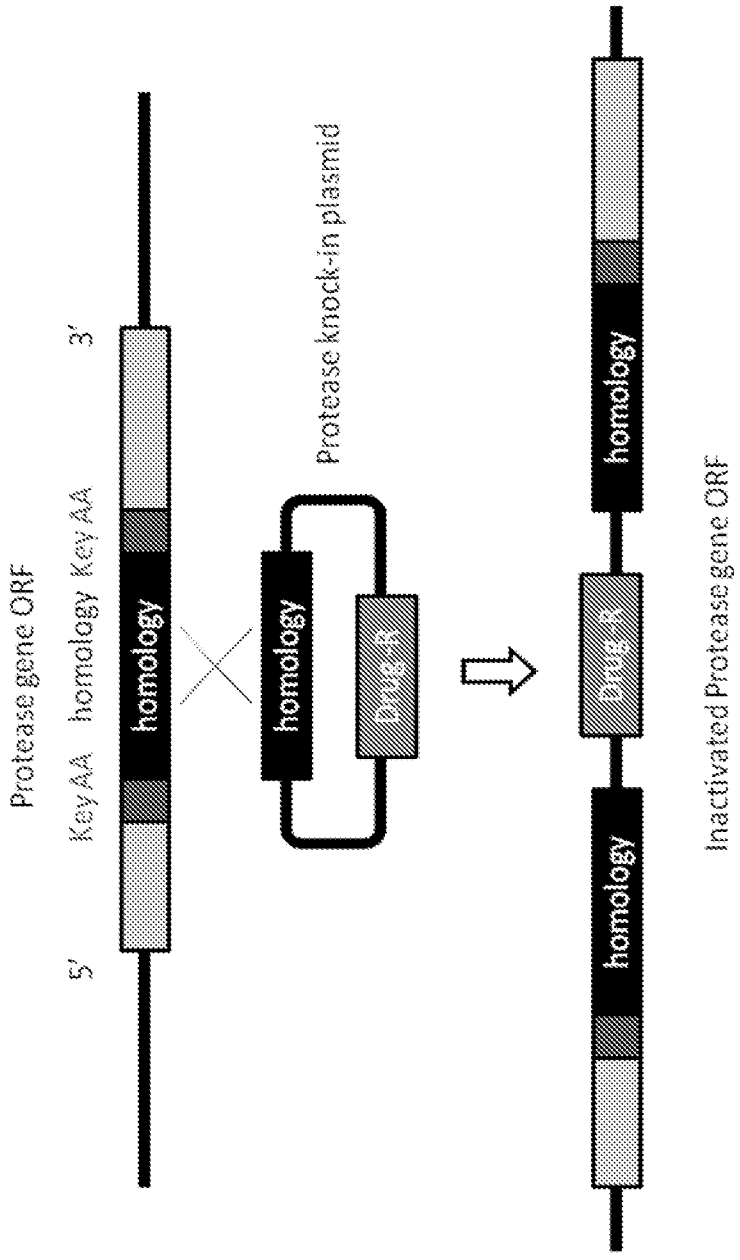

Protease gene ORF
5'
Key AA
homology
Key AA
3'
homology
Protease knock-in plasmid
homology
Drug R
homology
Drug R
homology
Inactivated Protease gene ORF

FIGURE 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 10/35825

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12N 1/19 (2010.01)
USPC - 435/69.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - C12N 1/19 (2010.01)
USPC - 435/69.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
435/254.23

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWEST - DB=PGPB,USPT,USOC,EPAB,JPAB; PLUR=NO; OP=ADJ; Google Scholar
search terms: 5S, rRNA, ribosomal RNA, recombin$, expression, cassette, heterolog$, endogen$, exogen$, enginee$, transfec$, copy, copies, multiple, Pichia, pastoris, marker, yeast, vector, plasmid, host, ribosomal, rna,

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 6,638,735 B1 (RHEE et al.) 28 October 2003 (28.10.2003) col. 2, ln 6-30; col. 4, ln 10-35; col. 5, ln 18-25.	24
Y	US 2009/0042264 A1 (FATLAND-BLOOM et al.) 12 February 2009 (12.02.2009) para [0034]; [0058]; [0065]; [0116]; [0125]; Fig. 10.	24

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 August 2010 (16.08.2010)	25 AUG 2010
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 10/35825

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 1-23
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

Applicant failed to respond to the ISA/225 mailed on 27 May 2010. Accordingly, the USPTO cannot supply a search for the sequences listed in this application. Claims requiring a sequence are therefore held as unsearchable

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.