

Expanding Yeast Knowledge Online

KARA DOLINSKI¹, CATHERINE A. BALL¹, STEPHEN A. CHERVITZ¹, SELINA S. DWIGHT¹, MIDORI A. HARRIS¹, SHANNON ROBERTS¹, TAIYUN ROE¹, J. MICHAEL CHERRY* AND DAVID BOTSTEIN¹

¹Department of Genetics, Stanford University, Stanford, CA 94305-5120, U.S.A.

The completion of the *Saccharomyces cerevisiae* genome sequencing project¹¹ and the continued development of improved technology for large-scale genome analysis have led to tremendous growth in the amount of new yeast genetics and molecular biology data. Efficient organization, presentation, and dissemination of this information are essential if researchers are to exploit this knowledge. In addition, the development of tools that provide efficient analysis of this information and link it with pertinent information from other systems is becoming increasingly important at a time when the complete genome sequences of other organisms are becoming available. The aim of this review is to familiarize biologists with the type of data resources currently available on the World Wide Web (WWW). © 1998 John Wiley & Sons, Ltd.

KEY WORDS — World Wide Web; Saccharomyces Genome Database; Munich Information Center for Protein Sequences; Yeast Protein Database

CONTENTS

Introduction	1453
SGD: Saccharomyces Genome Database . . .	1453
MIPS: Munich Information Center for Protein Sequences	1457
YPD: Yeast Protein Database	1458
Systematic gene expression sites	1460
Other useful yeast sites	1462
Yeast-related and laboratory sites	1462
Useful <i>Schizosaccharomyces pombe</i> and <i>Candida albicans</i> web sites	1464
Conclusion	1468
References	1468

INTRODUCTION

There are now three major *Saccharomyces* databases that maintain web servers: Saccharomyces Genome Database (SGD), Munich Information Center for Protein Sequences (MIPS) and Yeast Protein Database (YPD).^{6,14,15} We focus this review on these three databases because they represent the most comprehensive and relevant

information for yeast researchers. In addition to SGD, MIPS and YPD, several other useful budding yeast web sites are included. Since the number of these sites is growing rapidly, we only review those that contain a significant amount of data and resources and apologize for the omission of any new sites. In addition to these *S. cerevisiae* databases, there are several useful *Schizosaccharomyces pombe* and *Candida albicans* sites, which we briefly summarize in tables. None of these descriptions is meant to be exhaustive. Instead, we highlight the general features and mention particularly useful or unique tools at each site, keeping in mind that these databases are constantly evolving. All URLs mentioned in this review can be obtained from:

http://genome-www.stanford.edu/Saccharomyces/yeast_review_URLs.html

SGD: SACCHAROMYCES GENOME DATABASE

URL: <http://genome-www.stanford.edu/Saccharomyces>

SGD was established to provide a fast, easy, and reliable method for members of the yeast community to obtain information about the *S. cerevisiae*

*Correspondence to: J. Michael Cherry, Department of Genetics, Stanford University, Stanford, CA 94305-5120, U.S.A. Tel: 1-650-723-7541; fax: 1-650-723-7016; e-mail: cherry@genome.stanford.edu

genome, the genes it contains and their interactions. SGD provides current, annotated yeast sequence through the WWW and via FTP. In addition to providing an on-line genome database, SGD also provides access to other information of interest to the budding yeast community. SGD is responsible for maintaining the official *S. cerevisiae* Gene Registry. The Gene Registry helps to maintain yeast gene names in a standardized format, and SGD mediates resolution of gene naming conflicts. On-line submission forms to register gene names are found at the SGD site. In addition, yeast researchers are encouraged to add their names and contact information to the list of colleagues at SGD, where this list can be browsed and searched. Community information such as yeast meetings are announced on the SGD page, and a 'Virtual Yeast Library' that contains hyperlinks to other useful yeast web pages is also provided.

The genome information in SGD is currently organized around a 'Locus' page for each ORF, containing a brief summary of the gene, its product and any mutant phenotypes. In addition, the 'Locus' page contains several links to the DNA sequence, protein information, a literature guide (Gene-Info), and other relevant sites. Many of the SGD features described below can be reached through hyperlinks from the 'Locus' page.

SGD: sequence analysis

Among the WWW-based yeast resources, SGD encompasses a wide array of sequence tools, including several sequence search, retrieval and analysis options. The central location on the SGD site for accessing sequence information is the Gene/Sequence Resources page. Here users can search by gene name, ORF name, chromosomal region, or raw DNA or protein sequence and select retrieve options for viewing information about the gene or sequence they entered (Figure 1). There are multiple options under the topic headings of Biology/Literature, Maps/Tables, Sequence Analysis and Sequence Retrieval. Each link allows users to perform the next logical step in their analysis. For instance, when the researcher chooses to retrieve a protein sequence, the sequence is subsequently displayed along with links to BLAST² and FASTA^{18,19} searches, making sequence retrieval and subsequent analysis very straightforward for the user.

In addition to BLAST and FASTA, SGD has developed other tools for exploring sequence

patterns and similarities and displaying the results in a user-friendly manner. The Pattern Matching program can search for short DNA or peptide sequences, allowing for ambiguous and/or degenerate patterns. Researchers may also view the results of *S. cerevisiae* chromosome-by-chromosome comparisons using the Genome-wide DNA Similarity View and protein-by-protein comparisons using the Genome-wide Protein Similarity View.

SGD: unique features

The Global Gene Hunter search program, a unique feature of SGD, is a powerful tool that quickly performs a broad survey of the information available anywhere on the WWW for a particular gene. After the user enters a gene name, the Global Gene Hunter simultaneously searches SGD, GenBank, PubMed, Sacch3D, Swiss-Prot, MIPS, YPD and PIR for information about the gene.^{3,10,22,24,26,30} Results are presented as links, providing very fast access to many different information sources.

SGD also includes a section dedicated to various lists and tables, including lists of gene names, tRNAs, commonly used auxotrophic markers, and files available for downloading via anonymous FTP. An interesting example is the Mammalian Homology to Yeast table, which allows users to view the results of a BLAST search comparing each yeast protein to all unique human, mouse, rat, cow and sheep protein sequences from a recent version of GenBank. The results are summarized in tables that contain hyperlinked entries. A similar table is also provided at MIPS.

A tool of practical use to the bench scientist is the Web Primer feature, which designs primers for users for PCR or DNA sequencing of a particular gene or raw sequence. Another useful bench resource is the Yeast Genome Restriction Analysis program. This program generates and displays a restriction map from a raw DNA sequence, gene name, ORF name, clone name, GenBank sequence name or accession number.

One of the newest features at SGD is the SAGE tag viewer (Figure 2). SAGE, or Serial Analysis of Gene Expression, is a technique that utilizes short sequences that are associated with RNA transcripts to analyse gene expression at the genome-wide level. With the SAGE tag viewer, the user can search data from a yeast SAGE study³² by entering a gene or ORF name, or by selecting a

Netscape: Gene/Sequence Resources

Location: <http://genome-www2.stanford.edu/cgi-bin/SGD/seqTools>

Gene/Sequence Resources [Help](#)

Search SGD | Gene/Seq Resources | Help | Gene Registry | Maps
BLAST | FASTA | PatMatch | Sacch3D | Primers | SGD Home

The currently selected gene/sequence is :
ACT1 / YFL039C on chromosome VI from coordinates 54696 to 53261.

[Change Entered Selection or Coordinates]

<p>Biology/Literature</p> <ul style="list-style-type: none"> Gene details Literature Summaries Sequence details Clone details Protein details 3-D Structure (Sacch3D) Global Gene Hunter Mammalian Homologs 	<p>Display Maps/Tables</p> <ul style="list-style-type: none"> Chromosomal Features Map Chromosomal Features Table Combined Physical and Genetic Map Physical Map Genetic Distance vs. Physical Distance Ratios Genomic Stripe View Genome-wide Protein Similarity View Genome-wide Protein Similarity Table SAGE Results Map 	<p>Sequence Analysis</p> <ul style="list-style-type: none"> BLAST Search FASTA Search Genome Restriction Map Design Primers
--	--	--

Sequence Retrieval

<p>DNA of Region (with introns)</p> <p>Coding Sequence of selected gene/ORF (without introns, without flanking regions)</p> <p>Protein Translation of ORF</p> <p>6-Frame Translation (with Restriction Map)</p> <p>Restriction Fragment Sizes</p>	<p>Output Format</p> <p>GCG FASTA NoHeader</p> <p>GCG FASTA NoHeader</p> <p>GCG FASTA NoHeader</p> <p>GCG</p> <p>GCG</p>
--	---

Figure 1. The Gene/Sequence Resources information retrieval Page at SGD. After searching by gene name (here, *ACT1*) at the Gene/Sequence Resources search form (<http://genome-www2.stanford.edu/cgi-bin/SGD/seqTools>), the results are displayed as shown. Each of the categories is linked to the relevant *ACT1* site. For instance, if the user clicks on BLAST Search under Sequence Analysis, the BLAST form is displayed, already containing the *ACT1* sequence.

section of a chromosome using the SGD Genomic View.

Sacch3D

Sacch3D is a component of SGD that collects and presents three-dimensional structural information for *S. cerevisiae* proteins obtained from the

Protein Data Bank (PDB).^{1,23} A user can obtain available structural information for any given gene or ORF name and display available 3D structures for a yeast protein as well as structures for proteins with significant sequence similarity to the yeast protein. A variety of information is presented for each PDB structure, including the protein descrip-

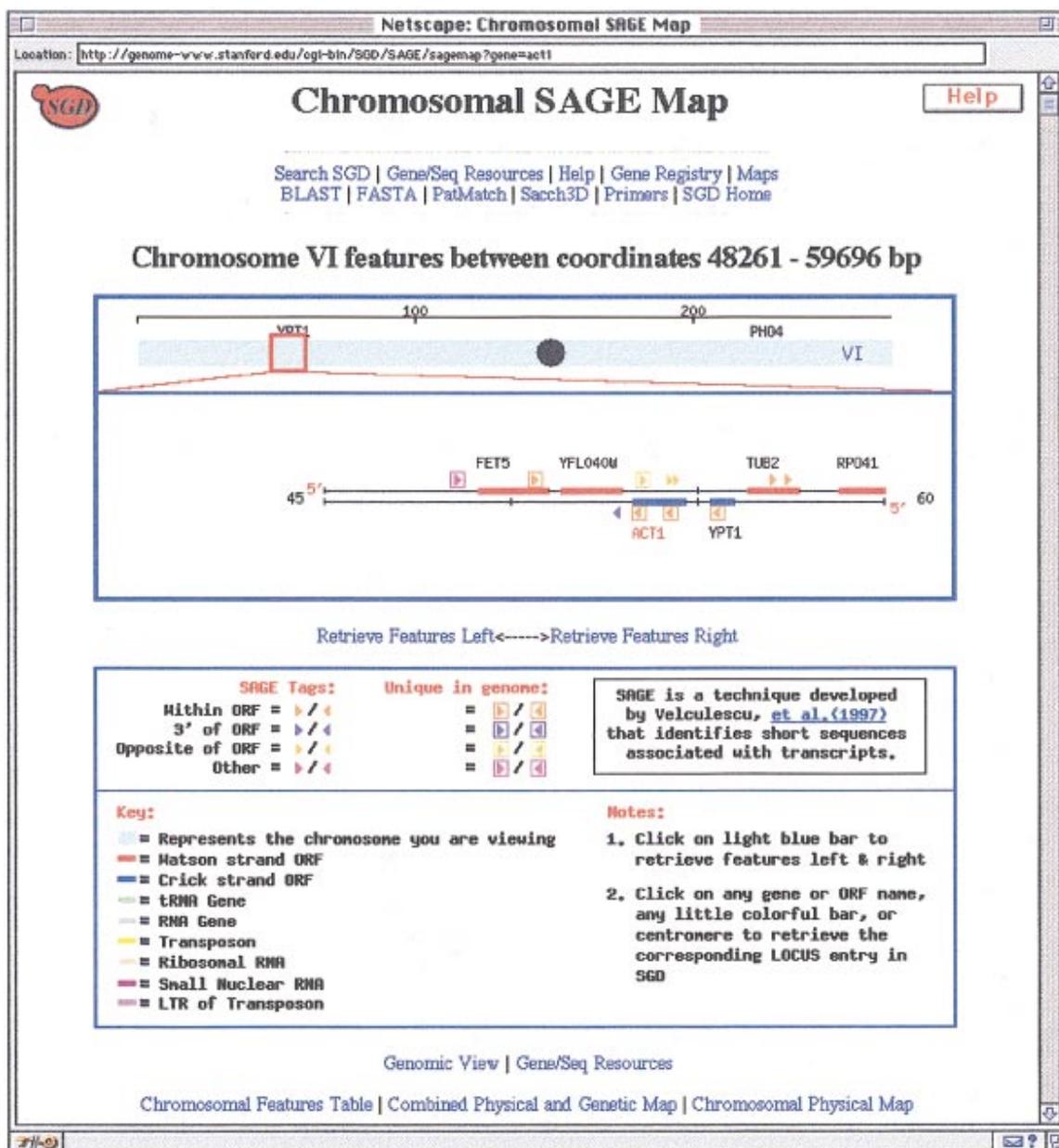


Figure 2. The SAGE tag viewer at SGD. After searching for *ACT1* at the Query SAGE data page (<http://genome-www.stanford.edu/cgi-bin/SGD/SAGE/querySAGE>), this SAGE map is displayed. By clicking on the SAGE sequence tag on the graphic display of the chromosome, the user can retrieve detailed information about that particular SAGE sequence tag, including its sequence and expression results. Links at the bottom of the page (Genomic View, Gene/Sequence Resources) provide access to other information about the chromosomal region.

tion, source organism, BLAST scores and links to interactive 3D viewers (Java/RasMol, Cn3D) and other structural databases. Also available from the structural information page are links to pre-

existing BLAST reports against the PDB and other datasets (GenBank, ESTs, mammalian), motif searches to assign functional class, secondary structure predictions and, if a PDB homologue

exists, the 3D structure for the yeast protein based on its sequence.

MIPS: MUNICH INFORMATION CENTER FOR PROTEIN SEQUENCES

URL: <http://speedy.mips.biochem.mpg.de/mips/yeast/index.htmlx>

MIPS coordinated the collaborative efforts of European groups during the *Saccharomyces* genome sequencing project and now manages a WWW server that provides the community with access to several genome databases. In addition to supplying molecular biologists with a comprehensive yeast database, MIPS also provides updates on the progress of several other genome projects, including genome data from other microorganisms and *Arabidopsis*.

The information and features at MIPS are organized into several different categories, namely Search, Chromosome Display, Tables and Graphics, and Protein Catalogues. In addition, each of these categories, along with a list of its contents, is accessible from their home page.

MIPS: sequence analysis

The MIPS yeast database contains the sequence data from the entire *Saccharomyces* genome. The mitochondrial and nuclear genomes have been analysed to extract potential ORFs, RNA genes, and genetic elements such as centromeres and telomeres. From the MIPS yeast genome page, information may be retrieved in ways similar to those used by other yeast web sites. The ORFs may be searched by name or accession number, and information about the biochemistry or physiological function of each gene is available.

Like SGD, MIPS also provides tools that allow the user to analyse the yeast genome as a whole. Entire chromosomes or chromosomal segments can be viewed graphically or in tabular form to scrutinize ORFs or genetic elements in detail. The genome can also be examined for redundancy; one chromosome can be compared against another, or against all other 15 chromosomes.

MIPS: sequence searches

Users may utilize the BLAST program at MIPS to search for homologues of their favourite gene. A precomputed set of FASTA results is also available at MIPS; it displays related proteins and their

alignments. A table describing the yeast homologues of human disease-associated genes provides descriptions of the disease, the BLAST search results and a hyperlink to the gene's entry.

MIPS: unique features

MIPS has several useful and unique features. For instance, excerpts from several recent reviews are provided at the MIPS site under a section entitled 'Selected Yeast Reviews'. An exhaustive set of *S. cerevisiae* zinc finger proteins is also available. Other information found at the review section of the web page includes a description of peroxins and peroxisome biogenesis genes, tRNA genes and retrotransposable elements, and the classification of the major facilitator superfamily. The genes listed in these tables are hyperlinked to their central locus sites, making retrieval of supplementary information about a particular gene simple and intuitive.

One of the most valuable contributions of MIPS is the yeast catalogue section. These catalogues contain lists of ORFs grouped together by several biological topics. For instance, the Functional catalogue groups genes by function, using both general (e.g. signal transduction) and specific categories (e.g. pheromone response). Other catalogues group genes by Prosite Motif,^{3,4} Enzyme Activity, Protein Class, Protein Complex, Phenotype, Physiological and Genetic Pathways, and Subcellular Localization. The Physiological and Genetic Pathways are displayed graphically to illustrate the relationships between different gene products as they work together to perform a cellular process. In addition, all of the catalogues are hyperlinked to the annotations for each gene in the database; when a gene is retrieved in a search, all relevant catalogues will appear as links on the ORF's page.

MIPS also provides a useful reference section entitled, 'Selected Tables and Graphics', which includes lists of intron-containing ORFs, essential and non-essential genes, genetic and physical protein interactions, small ORFs, and predicted transmembrane domains. Two links from this section are displayed graphically: the Yeast Centromeres page and the proteasome/protease protein table (designated 'YTA' protein table at MIPS). On the Yeast Centromere page, the centromeres are aligned, and multiple colours are effectively used to illustrate highly conserved regions (Figure 3). The user may click on a centromere

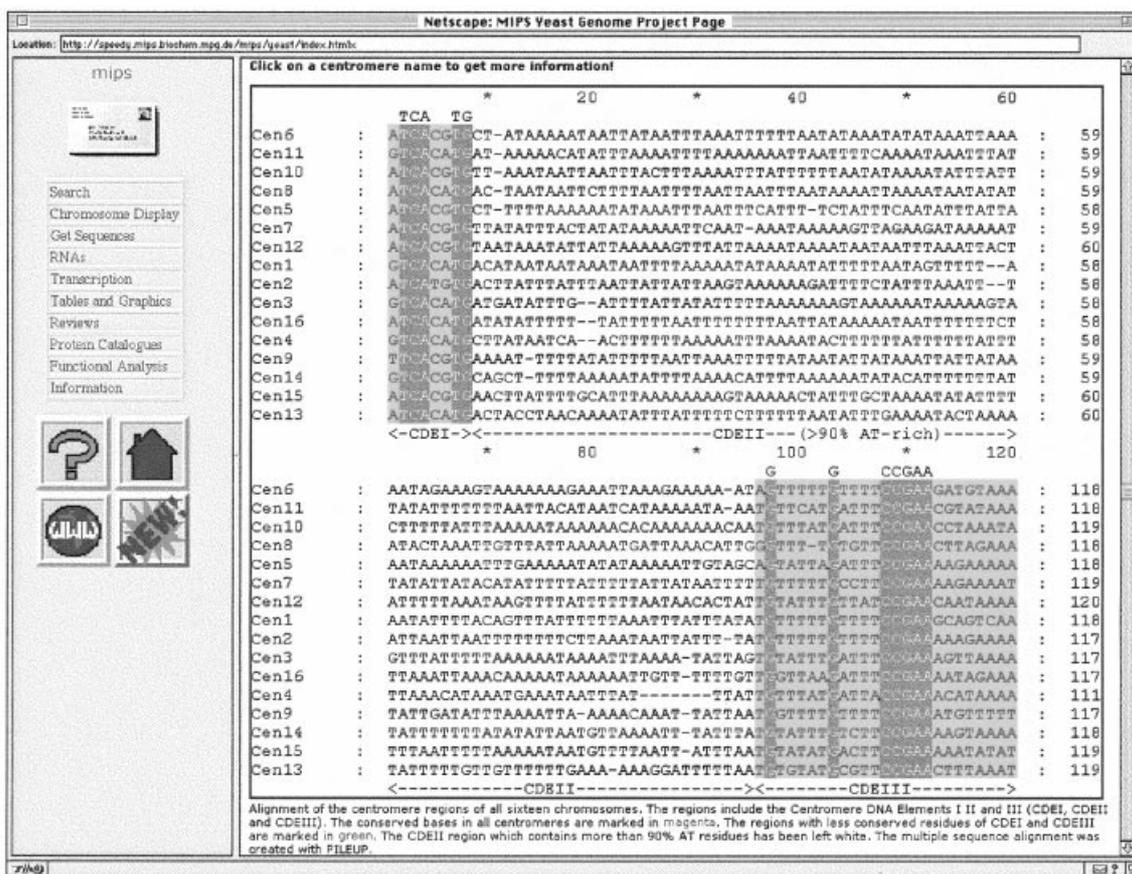


Figure 3. Alignment of the yeast centromeres at MIPS. This multiple sequence alignment of all the yeast centromeres was created with PILEUP. Each centromere name is linked to the MIPS page containing detailed information about it. In addition, there is a link at the bottom of the page to literature characterizing the centromeres.

name to retrieve more information about that particular centromere. The YTA protein link goes to a list of related topics that each lead to a schematic diagram (Figure 4). The protein names in the schematic are linked to information about their respective genes; the graphics effectively convey information about the localization, functional relationships and homologues of these proteins.

Pedant

Pedant,^{9,20} another service offered by MIPS, is an automated genome analysis resource focused on assigning functions to protein sequences. The results from many different sequence analysis techniques are integrated and made accessible via an intuitive user interface. Users can browse the set of yeast ORFs based on the MIPS or bacterial functional categories, PIR keywords, PIR superfamily

names or Prosite pattern categories. For any given ORF, Pedant provides a PileUp²¹ multiple sequence alignment against proteins from the major sequence databases, a list of sequence motifs from the BLOCKS¹³ and Prosite databases, a list of PDB structural homologues, a raw BLAST report, and a general report page that contains functional information, the amino acid sequence and secondary structure prediction. Pedant also contains information from the analysis of many other genomes and is regularly updated.

YPD: YEAST PROTEIN DATABASE

URL: <http://www.proteome.com/YPDhome.html>

YPD began as a protein database rather than a genome database; thus, its information is centred

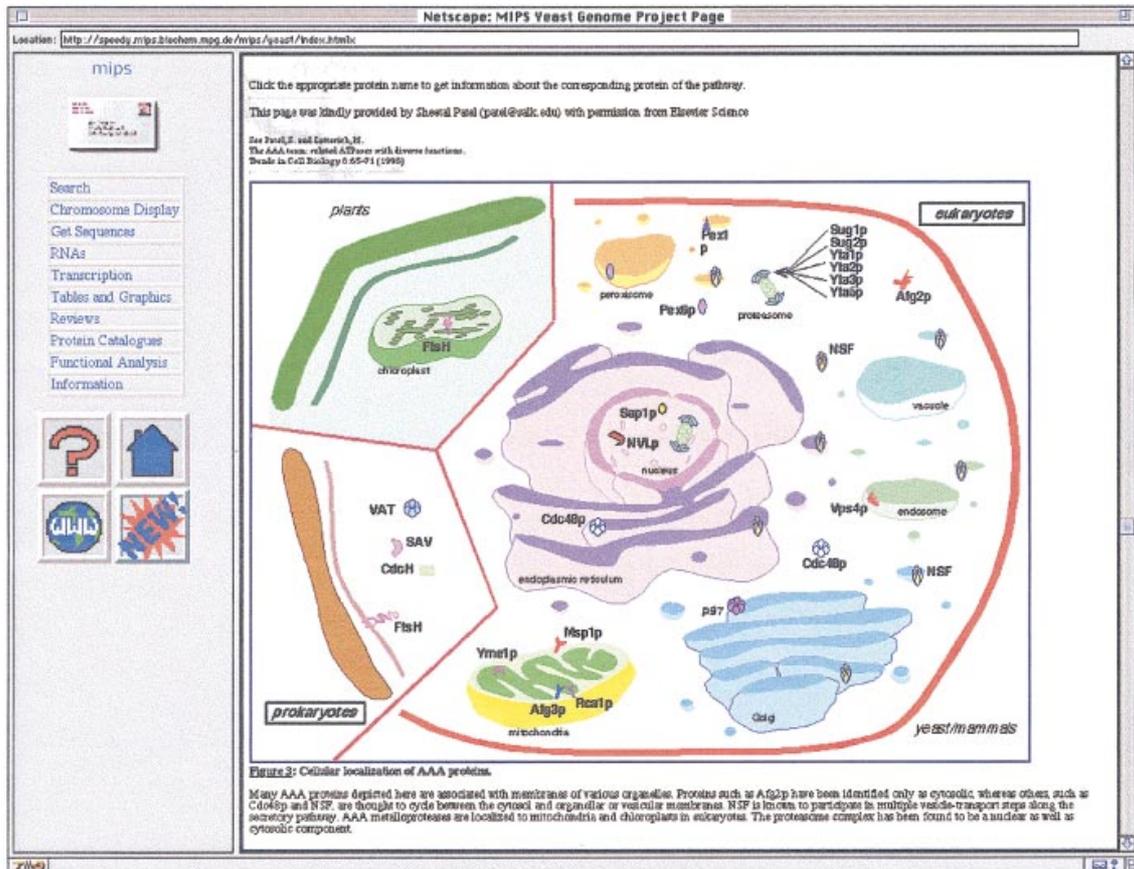


Figure 4. Cellular localization of YTA proteins at MIPS. Each yeast protein name is linked to the MIPS page describing the characteristics of the protein and the gene encoding it.

around amino acid rather than nucleic acid sequences. In this regard, it is different conceptually from MIPS and SGD. For instance, there is no access to DNA sequences through YPD. Instead, emphasis is placed on providing detailed information about the *Saccharomyces* proteins themselves. Although much of YPD's protein reports are included in MIPS and SGD, YPD excels at presenting its information in a very readable, compact form. Moreover, most of the data for a given protein is usually available with minimal scrolling and within one click from the yeast 'protein report'. YPD is a product of Proteome, Inc; note that, while YPD is free for academic use, non-academic users must pay a fee to access this site.

YPD: searching the protein database

YPD provides a single search page (in a long or short version) for accessing the data. With the long

form, users have the ability to do simple yet powerful Boolean (true/false) searches for proteins that match criteria in 14 different categories, including items such as subcellular localization, viability or inviability of the knockout mutant, molecular weight and number of potential transmembrane domains (Figure 5). For instance, the user could search the YPD for all essential proteins localized to the nucleus that contain transmembrane domains. YPD, with its long search form, sets an example of new developments in search capabilities that will greatly benefit researchers.

YPD: protein reports

YPD's protein reports are presented in tabular form, with most of the information organized so that it is either on this report or just one click away (Figure 6). A brief yet informative description of the protein is provided at the top of the table. In

The screenshot shows a Netscape browser window titled "Netscape: Proteome, Inc." with the address bar containing "http://www.proteome.com/search1.html". The main content area is titled "Subcellular localization categories" and is divided into three sections:

- 1. Major Localization:**
 - Cytoplasmic
 - Nuclear
 - Mitochondrial
 - Cytoskeletal
 - Peroxisomal
 - Endoplasmic reticulum
 - Golgi
 - Vacuolar
 - Vesicles, secretory system
 - Plasma membrane
 - Extracellular
 - Membrane, unspecified
 - Lipid particles
 - Unknown localization
- 2. Minor Localization:**
 - Mitochondrial inner membrane
 - Mitochondrial outer membrane
 - Nucleolus
 - Nuclear pore
 - Spindle pole body
 - Nuclear transport
 - Cell wall
- 3. Molecular Environment:**
 - Soluble
 - Integral membrane
 - Peripheral membrane
 - Actin-associated
 - Tubulin-associated
 - Ribosomal
 - Protein synthesis factor
 - DNA-associated (may be indirect)
 - RNA-associated (may be indirect)

At the bottom of the form, there are two buttons: "Submit Search" and "Clear Form". Below the buttons are several search criteria in brackets: [LOCALIZATION], [FUNCTION], [MODIFICATIONS], [GENETICS], [CALCULATED PROPERTIES], [PREDICTED MOTIFS], [GENE NAME SELECTION], and [KEYWORD SELECTION].

Figure 5. Long search form at YPD (<http://www.proteome.com/search1.html>). The user can search for sets of proteins that belong to several categories; the subcellular localization parameters are shown here.

this table, YPD also lists synonyms for the gene encoding the protein, links to various databases (e.g. SGD, GenBank, PIR, Swiss-Prot), physical characteristics of the protein [e.g. molecular weight, PI, CAI (codon adaptation index²⁸)] and information about protein modifications and/or motifs. Another nice feature of the protein report is that the user can click on 'Related Genes' to obtain pre-computed protein sequence comparisons of that ORF against other yeast, *Drosophila* and human proteins.

One of the most impressive aspects of YPD is the excellent literature summary that accompanies each protein report. These annotations are divided into different categories, such as 'Phenotypes' and 'Related to'. Each annotation is associated with one or more numbers that are linked to the reference(s) that support the annotation. The references are conveniently located at the bottom of the report, and the references are linked to PubMed. In addition to these annotations, the protein reports also contain links to information about 'Interactions', 'Regulations' and 'Modifications'. For instance, by clicking on 'Interactions', users find a list of proteins (that are linked to their protein reports) with which their protein of interest genetically or physically associates.

YPD: unique features

There are other notable features at the YPD site. A spreadsheet version of their database is available

upon request (for a fee to commercial users). For those who may still prefer the convenience of using a book rather than a computer terminal, YPD also sells a printed version of their database, called the *Yeast Proteome Handbook*. The web site includes an extensive user manual that is yet another feature that makes YPD friendly and easy to use. Finally, the YPD site includes images of theoretical two-dimensional gels for several classes of proteins.

SYSTEMATIC GENE EXPRESSION SITES

Here we briefly summarize the web sites that present data from systematic gene expression projects.

Yale Genome Analysis Center

URL: <http://ycmi.med.yale.edu/YGAC/home.html>

The investigators at the Yale Genome Analysis Center are using a transposon-based system to construct DNA libraries that generate random fusions of genes to lacZ or GFP as well as full-length, HA-tagged proteins in yeast. The transformants can then be screened to identify genes that are expressed in particular conditions (for instance, during sporulation), and to determine the localization of proteins and the phenotypes of disruption mutants. This project is being pursued

Netscape: YPD<tm> Protein Report for CTT1

Location: <http://www.proteome.com/YPD/CTT1.html>

YPD™ Protein Report for CTT1

This report is from the Yeast Protein Database, maintained by Proteome, Inc. YPD is updated daily and can be searched by gene name, keywords, and protein properties.

CATALASE T (CYTOSOLIC)

YPD Name	CTT1		
Swiss-Prot Name	CTT1		
SGD Name	CTT1 (Saccharomyces Genome Database)		
Synonym List	CTT1/G4628/YGR088W		
GenBank #	1323129	Chromosome	VII
PIR #	CSBYT	Introns	no
Swiss-Prot #	P06115	Knockout	viable
YEPD #			
PI	6.410	Subcellular Localization	cyt
Molecular Weight	65733	Molecular Environment	
Codon Bias	0.306	Function Category	hsp
CAI	0.241		
Length	573		
N-term Modif.		CDC28 Sites, potential	0
C-term Modif.		CKII Sites, potential	0
Phosphorylation	unknown	PKA Sites, potential	1
Glycosylation	unknown	N-glc Sites, potential	2
Pre-peptide Length		TM Domains, potential	0
N-term Seq. (precursor)	mpirsis	Interactions	
N-term Seq. (mature)	mpirsis	Regulations	more...
C-term Seq. (mature)	wsiasan	Modifications	
Full Sequence	view		
Related Genes	more...		

Function

- not involved in resistance to copper (21695)
- murine MEKK rescues bck1 mutant cells, but does not rescue ste11 mutant cells, nor induce phosphorylation of Hog1p, nor induce expression of CTT1 (21331)

Overproduction

- overproduction is more resistant to lethal heat shock (25859)

Genetic interactions

- transcription activation driven by STRE sequence is enhanced in ras2 mutant and lost in bey1 mutant (2529)

Figure 6. Protein report at YPD. After searching for the gene name *CTT1* with the YPD short search form (<http://www.proteome.com/search0.html>), the Protein Report is displayed; shown here is the first page. By scrolling down, the user can view an extensive summary of the literature, which contain links to the PubMed abstracts.²⁴

with several innovations that will greatly improve the rate at which interesting transformants can be studied and made publicly available.⁵

While the tools for searching and examining data are still being developed and it is a little difficult to find the links containing background information, this web site provides a convenient form for requesting reagents, links to useful and detailed protocols, and maps and descriptions of the libraries and other DNA reagents.

Exploring the metabolic and genetic control of gene expression on a genomic scale

URL: <http://cmgm.stanford.edu/pbrown/explore/index.html>

This Web site is associated with a *Science* paper of the same title⁸ and is maintained by Patrick Brown's laboratory at Stanford University. This site describes the use of microarrays to investigate the temporal expression of yeast genes during the shift from fermentation to respiration. The Brown laboratory does a great job in describing this new technology with both graphics and text. There is even a protocol with instructions on how to build your own microarrayer.

Saccharomyces Cell Cycle Expression Database

URL: <http://genomics.stanford.edu/yeast/cellcycle.html>

This site, maintained by Ron Davis's laboratory at Stanford University, describes a project to characterize cell cycle-dependent gene expression in the yeast genome using DNA chip technology. The experiments identify 422 ORFs whose mRNA levels show periodic variation during the cell cycle. The methods are described at the site, and the raw data can be downloaded. Some figures summarizing the data are also available. Genes are organized into functional categories, and mammalian homologues are identified. Most of this information is available to the general public although one link, the Access Expression Database, requires a password.

OTHER USEFUL YEAST SITES

GeneQuiz

URL: <http://www.sander.ebi.ac.uk/genequiz/genomes/sc>

GeneQuiz²⁷ is a system for large-scale biological sequence analysis that uses a variety of sequence

search and expert system analysis methods to assign function to protein sequences. In January 1997, GeneQuiz analysed all ORFs in the yeast genome and assigned function based on sequence similarity against the major DNA and protein databases. All ORFs were classified into five groups: '3D' shows clear similarity to a protein of known 3D structure; 'clear function' has strong similarity to a protein with functional annotation; 'tentative function' means there is weak similarity to a protein with functional annotation; 'homologue' denotes strong similarity to a protein without functional annotation; and 'no homologue'.

Genome Navigator

URL: <http://www.mpimg-berlin-dahlem.mpg.de/~andy/GN/S.cerevisiae/>

The Genome Navigator¹² provides an interactive Java-based map viewer ('DerBrowser') for browsing features along yeast chromosomes. Users can view ORFs along a chromosome based on functional category and obtain information from several external databases for a selected ORF (SGD, YPD, MIPS, Pedant, GeneQuiz, Gene duplications). A separate interactive Java applet is also available at the Genome Navigator site for viewing duplications in the yeast genome. This applet is integrated with the yeast gene duplication website,³³ discussed below.

Yeast Gene Duplications

URL: <http://acer.gen.tcd.ie/~khwolfe/yeast/nova/>

The yeast gene duplication website permits access to data for duplicated chromosomal regions in the *Saccharomyces* genome.³³ The duplicated regions are contained in 55 blocks that together occupy half of the genome. The different blocks, including the genes contained within them, can be browsed. For each gene, users can obtain BLASTP database search and Smith-Waterman^{18,29} pairwise comparison results of a protein sequence against all yeast proteins. Pairwise alignments for all hits are available. Dot matrix plots comparing the duplicated chromosomal segments are provided. The duplication map was in the process of being updated as this manuscript was being prepared in August 1998.

YEAST-RELATED AND LABORATORY SITES

Several yeast laboratories maintain helpful sites that provide on-line protocols (Table 1) and other

Table 1. *Saccharomyces cerevisiae* methods and protocols sites.

Laboratory	Description	URL
Boeke	Information of use of the selectable/counter-selectable colour marker <i>MET15</i>	http://www.welch.jhu.edu/~gregory/MET15.html
Botstein	Protocols for immunofluorescence and staining; also nice graphics displaying examples of actin staining	http://genome-www.stanford.edu/group/botlab/protocols.html
Breeden Brenner	Procedures for FACS and differential display History and uses of yeast transformation; links to sequences of commonly used vectors	http://www.fhrc.org/science/basic/labs/breeden/Methods_BreedenLab.html http://asterix.jci.tju.edu/brenner/bi532.html
Brent	Resource centre for those performing two-hybrid screens	http://xanadu.mgh.harvard.edu/brentlabweb/interactiontrap.html
Brown	Includes links to genetic footprinting and microarray guides	http://cmgm.stanford.edu/pbrown/
Cooper	Antibody preparation and staining protocols, instructions for preparing Web movies	http://www.cooperlab.wustl.edu/
Dohlman Fangman and Brewer	Several common yeast protocols DNA replication-related protocols	http://info.med.yale.edu/bcmm/dohlman/protocol.html http://fangman-brewer.genetics.washington.edu/#AnchorMethod
Gietz	An extensive list of protocols and tips for transforming yeast; description of useful set of shuttle vectors	http://www.umanitoba.ca/faculties/medicine/human_genetics/gietz/
Gottschling	Several general yeast and bacteria protocols; also links to abstracts from the Gottschling laboratory	http://www.fhrc.org/~gottschling/yeast.html
Herskowitz	An excellent primer for those interested in yeast molecular biology, includes an introductory reading list, protocols, and tetrad dissection and scoring templates	http://www.sacs.ucsf.edu/home/HerskowitzLab/protocols/protocol.html
Sherman	An excellent introduction to yeast genetics and molecular biology	http://www.urmc.rochester.edu/smd/biochem/yeast

useful information; lists of laboratory sites can be obtained from either of the following URLs:

<http://genome-www.stanford.edu/Saccharomyces/yeastlabs.html>

<http://genome-www.stanford.edu/Saccharomyces/VL-yeast.html>

In addition, there are a number of excellent sites that do not specialize in *S. cerevisiae per se* but instead emphasize comparative genomics studies, which may prove useful to yeast researchers. We list a few such sites here.

Entrez Genome Page

URL: <http://www3.ncbi.nlm.nih.gov/Entrez/Genome/org.html>

The Entrez Genomes Division is supported by the National Center of Biotechnology Information (NCBI) and provides access to the genomic sequence of *S. cerevisiae* as well as other organisms. In addition, it has physical and genetic mapping data for organisms whose genomic sequence is still incomplete. The *S. cerevisiae* genomic sequence supplied by the Entrez Genome Division is compiled from assorted GenBank files into a single DNA sequence for each chromosome.

The *S. cerevisiae* page can be reached from the Entrez Genome Page. The NCBI offers both text and graphic means of surveying the sequence of all 16 *S. cerevisiae* chromosomes. Each chromosome can be seen as a GenBank text file of an 'Overview' display, which illustrates the entire chromosome graphically along with labelled ORFs. Different colours in the graphical overview illustrate where different GenBank files were required to assemble the virtual chromosome sequence. By clicking on an ORF name on the chromosome, the user can zoom in on that particular region of the chromosome to view a more detailed map of that chromosomal region. Additionally, the Genome Query, at the top of the page, allows the user to search the chromosome for an ORF name. The resulting display highlights the ORF along with neighbouring ORFs. All of the ORF names displayed are linked to GenBank text files describing the ORF in more detail. By selecting a button beside the graphic displays, the user can utilize the 'ORF Finder' program to search for annotated as well as un-annotated ORFs. Using a minimum size as low as 50 nucleotides, 'ORF Finder' displays all possible ORFs in six reading frames for a 10 kb region. Any ORF identified by ORF Finder can then be

seen in various sequence formats or used in a BLAST search. The positions of start, stop and alternative start codons can be seen in the 'Six Frame' view. On every page, the Genome Query pages contain links to the home pages of NCBI, Entrez, Entrez FTP sites, SGD and YPD.

Clusters of Orthologous Groups

URL: <http://www.ncbi.nlm.nih.gov/COG/>

The 'Clusters of Orthologous Groups' (COGs) project is part of NCBI's comparative genome efforts.³¹ Protein sequences from seven sequenced organisms, including *S. cerevisiae*, were compared to find proteins that contained sequence similarity and therefore represent possible orthologs, i.e. putative homologous proteins between different species.

What Is There?

URL: <http://wit.mcs.anl.gov/WIT2>

The What Is There? (WIT) project is attempting to model the metabolism of about 30 organisms, including *S. cerevisiae*, with sequenced or partially sequenced genomes. Their metabolic reconstructions consist of diagrams of metabolic pathways, function and pathway classification schemes, and assignments of ORFs to functional roles.

Kyoto Encyclopedia of Genes and Genomes

URL: <http://www.genome.ad.jp/kegg/kegg2.html>

The Kyoto Encyclopedia of Genes and Genomes (KEGG) project in Japan has several features that may prove useful to yeast researchers. At their 'Identify gene clusters in two genomes' page, the user can choose two genomes from a list of sequenced genomes (including *S. cerevisiae*) and perform a search for homologous gene clusters. The KEGG site also contains many well-drawn metabolic and regulatory diagrams that are organized within hierarchical classification schemes. The user can choose to see what portions of those diagrams have been identified in a particular organism (Figure 7).

USEFUL SCHIZOSACCHAROMYCES POMBE AND CANDIDA ALBICANS WEB SITES

Summarized here are the contents of a handful of *S. pombe* Web sites. Other *S. pombe* and

Table 2. A selection of *S. pombe* web sites.

Site	Description	URL
Cold Spring Harbor <i>S. pombe</i> sequencing project	Home page for CSH <i>S. pombe</i> genomic sequencing project	http://nucleus.cshl.org/pombeweb
National collection of yeast cultures	Source of strains of <i>S. pombe</i> and other yeast	http://www.ifrn.bbsrc.ac.uk/hcyc/Default.html
Pombe resources at ATCC	Strains, plasmids, other reagents from ATCC	http://www.atcc.org/highlights/sp_info.html
1st International Fission Yeast Meeting	Information about the 1st International Fission Yeast Meeting, 23–30 September 1999	http://www.ed.ac.uk/~ebmv26/pombemeeting.html
Functional categories of <i>S. pombe</i>	Categorization of <i>S. pombe</i> metabolic processes; part of the PUMA project at Argonne National Laboratory	http://www-c.mcs.anl.gov/home/compbio/PUMA/Production/NodeObjects/Schizosaccharomyces_pombe.html
<i>S. pombe</i> mitochondrial genome	Fungal mitochondrial genome project (FMGP), funded by MRC Canada	http://megasun.bch.umontreal.ca/People/lang/species/spo/spombe.html
<i>S. pombe</i> expression systems	Includes links to general pombe information and lists of commercially available expression systems; under construction	http://www.jeton.or.jp/users/hthohda

Table 3. A selection of *Candida albicans* web sites.

Site	Description	URL
The <i>Candida</i> web site, University of Minnesota	Very nice site with maps, genetic information, sequence, strains, methods and other resources	http://alces.med.umn.edu/Candida.html
<i>Candida</i> Adherence Mycology Research Unit	Reference for projects about <i>Candida</i> adherence	http://www.unr.edu/mycology/
<i>Candida albicans</i> sequencing project	Information on the <i>Candida</i> sequencing project and access to the sequence database	http://candida.stanford.edu/
Swiss-Prot <i>Candida</i> sequences	<i>Candida</i> sequence database at Swiss-Prot	http://www.expasy.ch/cgi-bin/lists?calbican.txt
<i>Candida albicans</i> pilot sequencing project	Information about the pilot sequencing project at the Sanger Centre	http://www.sanger.ac.uk/Projects/C_albicans

Candida albicans sites are summarized in Tables 2 and 3. There is also a *Candida* news mailing list; to subscribe, send e-mail to: listproc@stonebow.otago.ac.nz

The Sanger Centre, Cambridge, UK
Schizosaccharomyces pombe genome sequencing project

URL: http://www.sanger.ac.uk/Projects/S_pombe/

Begun in 1995, the *S. pombe* genome sequencing project is predicted to be completed within two years. The Sanger Centre page reports the current status of the project, organized by chromosome, and provides some raw sequences via FTP. The cosmids used in systematic sequencing are listed for each chromosome, and sources are identified. PomBase, a compilation of protein and DNA sequences, genetic and physical maps, references, and gene designations, is also available by FTP. Other useful features include a BLAST server and a link to Pedant, which provides an automated analysis of protein translations. The introductory page and BLAST forms are well-designed and easy to use; cosmid clones and sequencing progress are thoroughly documented.

WWW information on Schizosaccharomyces pombe

URL: <http://www.bio.uva.nl/pombe/>

Frans Hochstenbach at the University of Amsterdam maintains this useful pombe page. This site contains information about genomic sequencing efforts, including the Sanger Centre's project described above, as well as other sequencing projects. Hochstenbach's site also contains summaries of *S. pombe* phylogeny and early research and links to many other relevant web sites. Of particular interest is the online *Fission Yeast Handbook*, which covers numerous research protocols. The home page contains a large amount of text, but is reasonably well organized. Some of the highlighted items, such as an animation of cell division, require a password.

The Forsburg Lab Pombe Pages

URL: <http://pingu.salk.edu/users/forsburg/>

Susan Forsburg maintains her laboratory page at the Salk Institute for Biological Studies. This site includes an introduction to *S. pombe* biology, with emphasis on DNA replication and the cell

cycle. An impressive array of technical information is available, including sources of strains and plasmids, technical references, vectors and selectable markers, and more. There is also a list of *S. pombe* laboratory home pages. Forsburg's pages are remarkably well designed and organized, combining clear text with well-chosen graphics. The site contains a wealth of information and is easy and fun to use.

CONCLUSION

The completion of the *S. cerevisiae* genomic sequence has provided a staggering wealth of information available on-line. Sequence information and new biological techniques have changed the strategies of bench scientists. The on-line resources described in this review both reflect these changes and foreshadow the revolution to come. Systematic functional analysis projects are on the cusp of facilitating another quantum leap in the data available to describe the yeast cell. The yeast community can confidently expect that the computer resources at their disposal now will be available to organize, analyse and disseminate the functional analysis data.

REFERENCES

1. Abola, E. E., Sussmann, J. L., Prilusky, J. and Manning, N. O. (1997). Protein Data Bank archives of three-dimensional macromolecular structures. *Meth. Enz.* **277**, 556–571.
2. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
3. Bairoch, A. and Apweiler, R. (1997). The SWISS-PROT protein sequence data bank and its supplement TrEMBL. *Nucl. Acids Res.* **25**, 31–36.
4. Bairoch, A., Bucher, P. and Hofmann, K. (1997). The PROSITE database, its status in 1997. *Nucl. Acids Res.* **25**, 217–221.
5. Burns, N., Grimwade, B., Ross-Macdonald, P. B., Choi, E.-Y., Finberg, K., Roeder, G. S. and Snyder, M. (1996). Large-scale characterization of gene expression, protein localization, and gene disruption in *Saccharomyces cerevisiae*. *Genes Dev.* **8**, 1087–1105.
6. Cherry, J. M., Adler, C., Ball, C., Chervitz, S. A., Dwight, S. S., Hester, E. T., Jia, Y., Juvik, G., Roe, T., Schroeder, M., Weng, S. and Botstein, D. (1998). SGD: *Saccharomyces* Genome Database. *Nucl. Acids Res.* **26**, 73–79.
7. Cn3D: <http://www.ncbi.nlm.nih.gov/Structure/cn3d.html>

8. DeRisi, J. L., Iyer, V. R. and Brown, P. O. (1997). Exploring the metabolic and genetic control on a genomic scale. *Science* **278**, 680–686.
9. Frishman, D. and Mewes, H. W. (1997). PEDAN Tic genomic analysis. *Trends Genet.* **13**, 415–416.
10. GenBank: <http://www.ncbi.nlm.nih.gov/Web/Genbank/>
11. Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M. *et al.* (1996). Life with 6000 genes. *Science* **274**, 546.
12. Grigoriev, A. (1997). Genomes with a view. *Trends Genet.* **13**, 499.
13. Henikoff, J. G. and Henikoff, S. (1996). Blocks database and its applications. *Meth. Enz.* **266**, 88–105.
14. Hodges, P. E., Payne, W. E. and Garrels, J. I. (1998). The Yeast Protein Database (YPD): a curated proteome database for *Saccharomyces cerevisiae*. *Nucl. Acids Res.* **26**, 68–72.
15. Mewes, H. W., Hani, J., Pfeiffer, F. and Frishman, D. (1998). MIPS: a database for protein sequences and complete genomes. *Nucl. Acids Res.* **26**, 33–37.
16. Murzin, A. G., Brenner, S. E., Hubbard, T. and Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540.
17. NCBI: <http://www.ncbi.nlm.nih.gov/>
18. Pearson, W. R. (1991). Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith–Waterman and FASTA algorithms. *Genomics* **11**, 635–650.
19. Pearson, W. R. and Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA* **85**, 2444–2448.
20. Pedant: <http://pedant.mips.biochem.mpg.de/>
21. PileUp: part of the Wisconsin Package, Genetics Computer Group (GCG): <http://www.gcg.com>
22. PIR: <http://www-nbrf.georgetown.edu/pir/>
23. Protein Database (PDB): <http://www.pdb.bnl.gov/>
24. PubMed: <http://www.ncbi.nlm.nih.gov/PubMed/>
25. RasMol: <http://www.umass.edu/microbio/rasmol>
26. Sacch3D: <http://genome-www.stanford.edu/Sacch3D/>
27. Scharf, M., Schneider, R., Casari, G., Bork, P., Valencia, A., Ouzounis, C. and Sander, C. (1994). GeneQuiz: a workbench for sequence analysis. In Altman, R., Brutlag, D., Karp, P., Lathrop, R. and Searls, D. (Eds), *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*. AAAI Press, Menlo Park, CA, pp. 348–353.
28. Sharp, P. and Li, W. H. (1987). The Codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucl. Acids Res.* **15**, 1281–1295.
29. Smith, T. F. and Waterman, M. S. (1981). Identification of common molecular subsequences. *J. Molec. Biol.* **147**, 195–197.
30. Swiss-Prot: <http://www.expasy.ch/sprot/>
31. Tatusov, R. L., Koonin, E. V. and Lipman, D. J. (1997). A genomic perspective on protein families. *Science* **278**, 631–637.
32. Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett, D. E. Jr, Hieter, P., Vogelstein, B. and Kinzler, K. W. (1997). Characterization of the yeast transcriptome. *Cell* **88**, 243–251.
33. Wolfe, K. H. and Shields, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708–713.