REVIEW

Biocuration at the *Saccharomyces* Genome Database

Marek S. Skrzypek,* and Robert S. Nash

Department of Genetics, Saccharomyces Genome Database, Stanford University, Stanford, California

Received 3 April 2015; Revised 12 May 2015; Accepted 13 May 2015

Summary: Saccharomyces Genome Database is an online resource dedicated to managing information about the biology and genetics of the model organism, yeast (*Saccharomyces cerevisiae*). This information is derived primarily from scientific publications through a process of human curation that involves manual extraction of data and their organization into a comprehensive system of knowledge. This system provides a foundation for further analysis of experimental data coming from research on yeast as well as other organisms. In this review we will demonstrate how biocuration and biocurators add a key component, the biological context, to our understanding of how genes, proteins, genomes and cells function and interact. We will explain the role biocurators play in sifting through the wealth of biological data to incorporate and connect key information. We will also discuss the many ways we assist researchers with their various research needs. We hope to convince the reader that manual curation is vital in converting the flood of data into organized and interconnected knowledge, and that biocurators play an essential role in the integration of scientific information into a coherent model of the cell.

Key words: yeast; database; biocuration

INTRODUCTION

Saccharomyces Genome Database (SGD) was established in 1993 to collect, organize, and make easily accessible the rapidly growing knowledge about genes of the model organism *Saccharomyces cerevisiae*, also known as brewer's yeast, baker's yeast, or informally as yeast (Cherry et al., 1998). Humans have been using yeast since Neolithic times to make such essential elements of civilization as wine, beer and bread, but the career of yeast as a model organism in biomedical research did not begin until the 1930s when genetic crosses and biochemical methods were being developed (Mortimer, 2000). In the decades that followed, our understanding of biochemical pathways and other key aspects of cell biology, such as cell cycle control, differentiation, and DNA repair were greatly informed by yeast research (Mortimer and Johnston, 1986). The further development of molecular techniques, along with the ease and relatively low cost of doing yeast research, resulted in the coining of the phrase "The Awesome Power of Yeast Genetics" and elevated the status of yeast as the premier model organism for the study of eukaryotic molecular and cellular biology (Duina et al., 2014). As a consequence of these advances in technology and the vast amount of knowledge that had accumulated, yeast became the best-studied Eukaryote even before the dawn of the genomic age. It was therefore no accident that *S. cerevisiae* was chosen to become the first eukaryotic organism with a completely sequenced genome (Clayton et al., 1997).

At the time of the yeast genome-sequencing project, it became apparent that making sense of the ensuing flood of information would require the creation of a specialized database and development of a specialized process of data collection and management. The task for SGD was to store the newly determined genome sequence and link it to the wealth of data coming from the fields of genetics, biochemistry and cell biology. Since then, the mission of SGD has evolved far beyond

Abbreviations: EBI, European Bioinformatics Institute; GEO, Gene Expression Omnibus; GO, gene ontology; OMIM, Online Mendelian Inheritance in Man; SGD, *Saccharomyces* Genome Database.

*Correspondence to: Marek S. Skrzypek, Saccharomyces Genome Database, Department of Genetics, Stanford University, Stanford, CA 94305-5477. E-mail: marek.skrzypek@stanford.edu

Contract grant sponsor: National Human Genome Research Institute, Contract grant number: U41 HG001315

Published online 3 July 2015 in
Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/dvg.22862

that of being just a collection of facts and software tools. In the past 20 years, SGD has become a central resource serving the greater research community by educating students, fostering the exchange of information among researchers, and facilitating scientific discovery. These days, when multiple answers of varying veracity for any single question can be found on the internet, the goal of SGD is to provide the Gold Standard of Knowledge - a high-quality compendium of scientific facts that are validated, traceable, up-to-date, and integrated in a coherent system that researchers can trust and rely upon when conducting their research and writing their publications and grant proposals. This expert knowledgebase can only be built through the participation of highly dedicated, scholarly scientists: biocurators. Like the monastic scribes of the middle ages, biocurators preserve data and wisdom for future generations of researchers, but they also actively participate in building the system of knowledge. In the Information Age, it is biocurators who act as stewards of information for the benefit of the research community and for the common good.

BIOCURATION PROCESS

From SGD's inception, the primary goal of biocurators was to associate individual genes with various pieces of information about them, including name, structure, and function. Initially, almost all data were derived from small-scale experiments aimed at the characterization of a single gene or a small group of functionally related genes. Later advances in large-scale experimental methods led to an explosion of genome-wide data from studies that involved the analysis of large numbers of genes, or even the entire genome. As a result, it has become even more important to validate and contextualize such data. This process of biocuration, i.e., manually extracting, collating, analyzing, and disseminating information, is performed by biocurators: Ph.D. level scientists with years of scientific research experience. SGD biocurators read published, peer-reviewed journal articles relevant to yeast genetics, biochemistry, and cell biology, then locate and extract key information, and add the pertinent data to the database using specialized curation tools developed at SGD (Hirschman et al., 2010).

The literature curation workflow at SGD is outlined in Figure 1. The initial step is performed by automated scripts that search titles and abstracts of the published articles in Pubmed for specific keywords that identify those articles as potentially relevant to SGD. The subsequent steps are performed by biocurators who manually review, prioritize and triage the selected articles, on average nearly 100 per week, in preparation for deep curation.

It is important to note that curation involves much more than data extraction. The synthesis of a coherent base of knowledge from a large quantity of data, often of varying quality, requires perceptive and insightful biocurators. Every day, biocurators call on their own judgment to identify the entities associated with results presented in papers, to assess the scientific soundness of those results, to reconcile conflicting conclusions, and to track down missing or incomplete data. The task of assimilating these data into a body of expertly curated information frequently involves communication with researchers, authors, and biocurators from other database groups. Small-scale data are considered in the context of the information already available for a given gene and are entered into the database individually, whereas large-scale data are carefully evaluated, processed programmatically, and loaded in bulk. As the focus of research in biology increasingly shifts from individual genes or proteins to the behavior of entire networks of genes and proteins in the context of the cell, SGD biocurators stay attuned to these shifts by curating information regarding biochemical pathways, protein complexes, and regulatory pathways.

For many well-studied genes, researchers can face a bewildering maze of individual observations that often do not add up to an easily discernable "big picture". It is for this reason that expert biocuration goes far beyond merely collecting and displaying all available data. SGD biocurators also organize and summarize the conclusions from functional, phenotype, and regulation data into clear, human-readable synopses in order to sift through the noise and provide a succinct, accurate assessment of what is currently known about a gene.

Curation consistency is an issue often raised in discussions about manual curation. To ensure that the curated content does not differ significantly regardless of who enters the data into the database, SGD biocurators follow elaborate and time-tested curation procedures. Their experience in both wet-laboratory research and as highly educated scientists is a significant factor contributing to the quality and consistency of curation. SGD biocurators often discuss among themselves the best ways of capturing particular types of information and periodically perform curation consistency exercises focused on specific data types to ensure uniformity in data extraction and entry. However, the ultimate quality control is provided through collaborative interaction with the active community of SGD users, who never shy away from calling our attention to errors and omissions. Promptly responding to such comments is always our utmost priority, and we routinely make improvements to the database and its contents based on the valuable feedback we receive from students, educators, and researchers who use SGD.

CURATING VARIOUS DATA TYPES

The main types of data collected by SGD curators are shown in Table 1, which also shows the locations of the

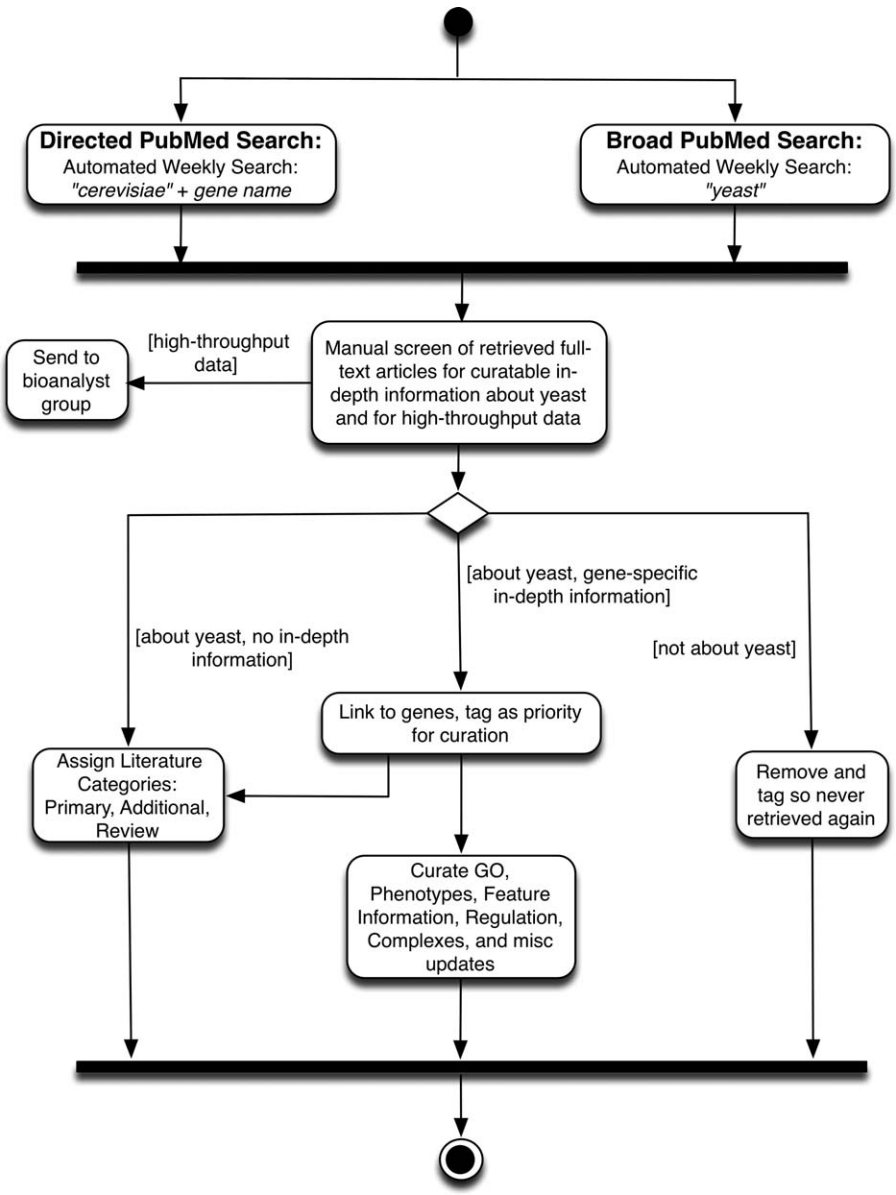

FIG. 1. SGD literature curation workflow. Automated weekly searches of Pubmed produce lists of articles to be manually screened by biocurators for relevant information and triaged for high-throughput data retrieval, for assignment of literature categories, and for deep curation that includes curation of GO, phenotypes, feature information (gene names, aliases, descriptions), regulation, complex memberships, and other updates.

data files within the SGD website and the references where the sources, curation procedures and navigation of the data have been previously described.

From its very beginning, SGD has been the official repository of the genome sequence for the budding yeast *Saccharomyces cerevisiae*. The original genome-sequencing project was performed by an international consortium and resulted in the first completely sequenced eukaryotic genome (Goffeau, et al., 1997). Since then, many incremental updates have been made, both to the sequence and to gene models, recently leading to a complete re-sequencing of the entire genome from the reference strain S288C with the use of modern sequencing methods (Engel et al., 2014). The responsibility falls upon SGD biocurators to ensure that the annotation of the reference sequence is accurate and accessible, as this sequence is vital for ongoing research in genetics, genomics, and molecular biology of budding yeast, involving both small- and large-scale wet lab experiments, as well as computational and comparative studies.

In SGD information about the biological functions of gene products is organized in the form of gene ontology

Table 1
Types of Data Available at SGD With Links to Corresponding Files and References

Data type	Content	Link and file	Reference
Reference Genome	*S. cerevisiae* reference genome and other strains, liftOver files, *Saccharomyces sensu stricto* species, sequences of *S. cerevisiae* viruses	http://www.yeastgenome.org/download-data/sequence	Engel et al., 2014
Gene Ontology	Molecular Function, Biological Process and Cellular Component annotations for *S. cerevisiae* genes with evidence codes and references	http://www.yeastgenome.org/download-data/curation (gene_association.sgd.gz)	Hong et al., 2008
Phenotype	Curated phenotypes with references and associated data	http://www.yeastgenome.org/download-data/curation (phenotype_data.tab)	Engel et al., 2010
Interaction	Genetic and physical interaction data for yeast incorporated from BioGRID (http://www.thebiogrid.org/)	http://www.yeastgenome.org/download-data/curation (interaction_data.tab)	Chatr-Aryamontri et al., 2014
Expression	Data files from microarray experiments used to populate SPELL (Serial Pattern of Expression Levels Locator)	http://www.yeastgenome.org/download-data/expression	Skrzypek and Hirschman, 2011
Biochemical Pathway	Yeast biochemical pathways with the names of the enzymes, E.C. numbers, genes that encode the enzymes and references	http://www.yeastgenome.org/download-data/curation (biochemical_pathways.tab)	Caspi et al., 2014
Comparative Genomic	Sequence comparisons between *S. cerevisiae* and other species	http://www.yeastgenome.org/download-data/genomics	
Literature	Literature information for references in SGD	http://www.yeastgenome.org/download-data/curation (gene_literature.tab)	Cherry et al., 2012
Much of the data in SGD are available for search and retrieval with YeastMine. This includes the data listed above, and the following:		http://yeastmine.yeastgenome.org	Balakrishnan et al., 2012
Protein	Yeast proteins, sequences, physical properties, modifications, domains		
Homology	Homologs in humans, fungi, other yeast		
Regulation	Transcriptional regulators and targets, transcription factor binding sites		

(GO) annotations. GO is a set of controlled vocabularies that describe three main aspects: molecular function, biological process and cellular component (Ashburner et al., 2000). Precisely defined terms are applied to gene products based on consistent, standardized curation procedures and provide a common language for functional annotations of genes regardless of species of origin. All the GO terms are organized in structured hierarchies (ontologies), so that their relationships to other GO terms are unambiguously traceable. Thus, GO provides a uniform, precise system that organizes knowledge about genes across all branches of life.

Biocurators assign most of the GO annotations in SGD based on information published in peer-reviewed research papers. Selecting the GO term that best fits the underlying observation requires both an extensive familiarity with the GO vocabularies and annotation guidelines, and a deep understanding of biology and experimental techniques. Each GO annotation is associated with an evidence code that indicates the type of data that supports the annotation, and the reference from which the annotation is derived. Additional optional fields provide more specific details about substrates, interaction partners, or the particular biological circumstances under which an observation is made, such as the cell cycle phase or cellular location. SGD biocurators are responsible for maintaining an accurate and complete set of GO annotations, and biocurators from other databases often leverage annotations from well-characterized yeast genes to predict the functions of homologous genes from organisms that are less well studied experimentally.

SGD biocurators manually extract mutant phenotypes from the published literature and categorize them with the use of a controlled vocabulary termed the Yeast Phenotype Ontology (Costanzo *et al.*, 2009). Descriptions of phenotypes in publications are often very detailed and therefore pose a significant challenge for biocurators trying to capture overall biological sense of an observation in a clear and searchable terminology without losing sight of all the significant details. The phenotype statement itself is composed of two elements: an observable that indicates a specific feature of the mutant cells, colonies, or cultures, and a qualifier that reveals the direction of the change in the mutant cells relative to wild-type cells. Several additional elements further extend the annotation: the experimental conditions under which the phenotype is assayed, the genetic background (strain), the type of experiment performed to construct and/or observe the effect of the mutated gene (classical genetics approach, or large-scale analyses) and the impact of the mutation on the gene (null, conditional, reduction-of-function, overexpression etc.). In addition, for phenotypes involving treatment with a chemical compound, the Chemical Entities of Biological Interest (ChEBI) developed by the European Bioinformatics Institute (EBI) is the dictionary of choice (Degtyarenko et al., 2008). In ChEBI, compounds are ontologically classified and ordered, so that biocurators can better ensure a higher degree of phenotype-to-phenotype uniformity. The role of the biocurator in phenotype curation is to understand the experiment, then accurately extract and present the information so that it makes sense to the biologist, and can also be data-mined by those who are computationally minded. As is the case with GO annotations, each phenotype annotation is associated with a reference, which is the source of information and provides users with access to all the details that SGD does not collect, such as strain origin, detailed experimental procedures etc.

References cited in SGD are linked to the relevant genes, and to both the GO and phenotype annotations extracted from the paper. They are categorized depending on the biocurator's assessment of the papers focus where primary literature is used to indicate that a gene or genes is the primary focus in this paper, while additional literature may describe a gene or its homologs, but not as the principal focus.

Just as genome-wide mutant screens have become commonplace, so have studies of gene expression on a genome-wide scale. SGD biocurators have collected and processed a large compendium of microarray data generated by members of the yeast community to analyze gene expression across the genome. Data are downloaded from the Gene Expression Omnibus (GEO), processed, and used to populate the expression analysis tool known as SPELL (Serial Pattern of Expression Levels Locator). SPELL is a query-driven search engine useful for identifying relevant and informative datasets for a small set of query genes supplied by the user (Hibbs et al., 2007). To enhance the utility of the SPELL search engine even further, SGD biocurators developed a set of category tags designed to cover the range of biological perturbations employed in typical microarray experiments. These tags aid in the selection and grouping of datasets by filtering the results based on the perturbation—for example, "drug response" or "heat shock." The SPELL expression analysis tool at SGD currently contains close to 500 datasets, representing over 10,000 total arrays from almost 350 studies. All of these datasets have been individually examined by biocurators and categorized using the appropriate tags based on the biological context of the associated experiments. These data are accessible from multiple locations, including the Expression option in the Function pull-down menu on the SGD home page.

COMMUNITY RESOURCES

At its inception, SGD was designed not only to store biological information about *S. cerevisiae* but also to

provide a forum for the community of yeast researchers, facilitating interactions between its members. To that end, we have provided for our users several opportunities to exchange information. Researchers have the ability to display their names, contact information, website link(s), laboratory members, associates and collaborators, as well as research interests, so that they can be readily found and contacted by other colleagues. SGD also provides a list of yeast labs from around the world that can be searched using a variety of keywords. Biocurators promptly process the newly submitted colleague entries, but it is up to the colleagues to make sure they stay up to date.

Part of the legacy of decades of genetic research in yeast is the profusion of gene names. Many genes are referred to by several different names, and often the same name is used to refer to two or more genes. This ambiguity can be a source of great confusion in the literature, and cannot be remedied without human intervention. All gene names are meticulously captured by SGD biocurators and linked to the appropriate genes, either as primary names or as aliases (synonyms). It is equally important to avoid adding to the naming confusion as newly discovered gene functions are published. Early on, after discussions with the yeast community, SGD assumed authority to mediate budding yeast genetic nomenclature. In this role, SGD biocurators work with researchers to ensure that proposed gene names are unique, have the correct format, and are consistent with the names of related genes. A web-based gene registry submission form is available for users who would like to reserve a gene name prior to publication. Biocurators review and accept these submissions, then, once a reserved gene name has been published in a peer-reviewed journal article, SGD biocurators will promote the reserved name as the standard name for that gene.

As keepers of the genetic nomenclature, biocurators also assist in the resolution of conflicts, and mediate community-based gene name change requests. The rationale for the request is first summarized, and is then communicated to researchers who have worked on the gene. Responses are collected and evaluated, and changes are made based on community consensus. SGD also helps with other nomenclature-related requests received from the community. These include a wide range of naming challenges such as defining the best way to name intergenic regions, developing the appropriate nomenclature for new chromosomal features (such as ncRNAs), and coming up with systematic names for genes that are either not present in the reference strain (S288C) or are newly discovered. Finally, SGD biocurators provide advice to researchers with respect to the correct format for gene, mutant allele and protein nomenclature as they relate to manuscript preparation, and to researchers developing nomenclature guidelines for use in other species.

SGD has always employed a wide variety of methods to communicate with and engage our user community. We continue to use home page announcements, a quarterly newsletter, posters and presentations at meetings, and video tutorials to disseminate information about new features, software, data and data visualization, and future scientific meetings and courses. Since January 2012, SGD has highlighted one current research article each week in the form of a blog post on our homepage. These blog posts use real-world analogies to communicate biology and informally spotlight papers of general interest to students and seasoned researchers, as well as anyone interested in science, with particular emphasis on those articles that demonstrate the awesome power of yeast. The blog is becoming increasingly popular and has received a very positive response.

With the advent of social media, SGD has developed ways to utilize these forms of communication, and now has a presence on both Twitter and Facebook. Through Facebook, SGD is able to reach a broader audience and keep users informed about upcoming meetings, highlight items of interest to scientists, including our blog posts, and provide information about our educational outreach activities. SGD uses Twitter to highlight new features and activities at SGD, as well as promote new papers that may be of general interest. At scientific meetings and conferences, Twitter is used to share highlights of talks for users who are unable to attend. In addition, social media have provided users with more avenues to contact and interact with us. SGD is also developing a presence on LinkedIn and Google+, and will continue to expand its reach into social media as new platforms are developed.

RECENT DEVELOPMENTS

SGD biocurators and software developers have recently embarked on a series of new projects aimed to capture information about macromolecular complexes, human homologs and associated diseases. The goal of these new projects is to integrate additional data types of interest to our users based on activity in the scientific literature and on feedback that SGD has received at meetings and through online surveys. These new data are examined in detail by SGD biocurators to determine the key pieces of information to capture and how to best present them to users.

To pursue the curation of yeast macromolecular complexes, SGD recently began collaborating with biocurators from EBI's IntAct Molecular Interaction Database (http://www.ebi.ac.uk/intact/). This project involves collecting information on named, curated complexes including the participating subunits (proteins, nucleic acids and cofactors), properties of the complex (stoichiometry and molecular weight), and the function of the complex using GO annotations specific to the complex.

Information about these curated complexes is available through the Complex Portal database at EBI where complexes from more than ten different organisms can be viewed. There are currently 349 annotated yeast complexes available for viewing in the Complex Portal database (Meldal *et al*., 2015).

Data specific to yeast macromolecular complexes have been loaded into YeastMine, SGD's powerful search and retrieval tool, where they can be queried by the name of the complex or of a subunit (Balakrishnan et al., 2012). In the future, biocurators will use these data to design macromolecular complex pages on the SGD website, which will enable users to explore complexes and their interconnections with other data at SGD.

Based on community driven requests for access to human homolog and disease relationships, SGD has begun actively exploring this area of biology and is incorporating such information into YeastMine. Several queries have been created in YeastMine to facilitate retrieval of homology and human disease-related information. Starting with a yeast gene(s), users can access predicted homologs, orthologs, and paralogs from species including human, rat, mouse, worm, fly and yeast. In the future we hope to extend this feature to other eukaryotic model systems. Homology data are retrieved periodically from TreeFam, PantherDB, and Homologene to keep this information current. More recently, data from Online Mendelian Inheritance in Man (OMIM) have been incorporated so that users can access both OMIM-derived human homologs and any associated disease phenotypes using their favorite yeast gene(s) as query. Alternatively, users can start with a human gene and retrieve yeast homologs and disease associations, or start with an OMIM disease phenotype to obtain associated human gene(s) and their yeast homologs. Although at this time we cannot directly link mutant phenotypes with OMIM entries, it is our hope that in the future we will be able to comprehensively cross-reference yeast phenotypes with phenotypic manifestations of human disease, particularly in cases where cross-species complementation has been used to study disease-associated alleles from the human ortholog. Biocurators are currently reviewing the literature to assemble a subset of homology-derived yeast-human gene pairs where partial or full cross-species complementation has been documented. Individual studies are examined in detail to identify the specific mutations studied and possible disease relevance. This information will be available on SGD web pages and through YeastMine, as part of SGD's effort in promoting the ways in which yeast and yeast research can inform genetic medicine.

FINAL THOUGHTS

SGD continues to explore and support the needs of not only the yeast community but also the broader scientific research community moving forward. As new experimental methods are developed and new data types become available, SGD works with, and for, researchers to make sense of these data, and make them available as rapidly as possible. Biology is firmly in the era of Big Data, in which hundreds and thousands of genes and proteins are studied simultaneously, generating literally millions of data points. While this abundance of data is awe-inspiring, it does not automatically translate into usable knowledge. Any analysis that gives data biological meaning can only be based on a solid foundation of verified biological facts that are expertly collected from multiple sources, carefully organized and synthesized into a coherent system. This foundation must be continuously updated, maintained, and broadened as new information and new technologies emerge. Continued stewardship of such a system of knowledge is SGD's ongoing goal, and is the challenge for all expert biocurators.

ACKNOWLEDGMENTS

The authors thank Drs Maria Costanzo, Stacia Engel, and J. Michael Cherry for their helpful suggestions during the writing of this manuscript; the authors also acknowledge the tireless effort of all the system administrators, programmers and biocurators in building and improving SGD.

LITERATURE CITED

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25–29.

Balakrishnan R, Park J, Karra K, Hitz BC, Binkley G, Hong EL, Sullivan J, Micklem G, Cherry JM. 2012. YeastMine–an integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose tool-kit. Database (Oxford) 2000:bar062.

Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, Holland TA, Keseler IM, Kothari A, Kubo A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Subhraveti P, Weaver DS, Weerasinghe D, Zhang P, Karp PD. 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res 42:D459–D471.

Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, Stark C, Breitkreutz A, Kolas N, O'Donnell L, Reguly T, Nixon J, Ramage L, Winter A, Sellam A, Chang C, Hirschman J, Theesfeld C, Rust J, Livstone MS, Dolinski K, Tyers

M. 2014. The BioGRID interaction database: 2015 update. Nucleic Acids Research 43:D470–D478.

Cherry JM, Adler C, Ball C, Chervitz SA, Dwight SS, Hester ET, Jia Y, Juvik G, Roe T, Schroeder M, Weng S, Botstein D. 1998. SGD: Saccharomyces Genome Database. Nucleic Acids Res 26:73–79.

Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M, Weng S, Wong ED. 2012. Saccharomyces Genome Database: The genomics resource of budding yeast. Nucleic Acids Res 40; D700–D705.

Clayton RA, et al. 1997. The first genome from the third domain of life. Nature 387:459−462.

Costanzo MC, Skrzypek MS, Nash R, Wong E, Binkley G, Engel SR, Hitz B, Hong EL, Cherry JM; the Saccharomyces Genome Database Project. 2009. New mutant phenotype data curation system in the Saccharomyces Genome Database. Database (Oxford) 2009:bap001.

Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, Alcántara R, Darsow M, Guedj M, Ashburner M. 2008. ChEBI: A database and ontology for chemical entities of biological interest. Nucleic Acids Res 36:D344–D350.

Duina AA1, Miller ME, Keeney JB. 2014. Budding yeast for budding geneticists: A primer on the Saccharomyces cerevisiae model system. Genetics 197:33–48.

Engel SR, Balakrishnan R, Binkley G, Christie KR, Costanzo MC, Dwight SS, Fisk DG, Hirschman JE, Hitz BC, Hong EL, Krieger CJ, Livstone MS, Miyasato SR, Nash R, Oughtred R, Park J, Skrzypek MS, Weng S, Wong ED, Dolinski K, Botstein D, Cherry JM. 2010. Saccharomyces genome database provides mutant phenotype data. Nucleic Acids Res 38: D433–D436.

Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, Costanzo MC, Dwight SS, Hitz BC, Karra K, Nash RS, Weng S, Wong ED, Lloyd P, Skrzypek MS, Miyasato SR, Simison M, Cherry JM. 2014. The reference genome sequence of *Saccharomyces cerevisiae*: Then and now. G3 (Bethesda) 4:389−398.

Goffeau A, et al. 1997. The yeast genome directory. Nature 387(6632 Suppl):5.

Hibbs MA, Hess DC, Myers CL, Huttenhower C, Li K, Troyanskaya OG. 2007. Exploring the functional landscape of gene expression: Directed search of large microarray compendia. Bioinformatics 23: 2692−2699.

Hirschman J, Berardini TZ, Drabkin HJ, Howe D. 2010. A MOD(ern) perspective on literature curation. Mol Genet Genomics 283:415–425.

Hong EL, Balakrishnan R, Dong Q, Christie KR, Park J, Binkley G, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Krieger CJ, Livstone MS, Miyasato SR, Nash RS, Oughtred R, Skrzypek MS, Weng S, Wong ED, Zhu KK, Dolinski K, Botstein D, Cherry JM. 2008. Gene Ontology annotations at SGD: New data sources and annotation methods. Nucleic Acids Res 36:D577–D581.

Meldal BH, Forner-Martinez O, Costanzo MC, Dana J, Demeter J, Dumousseau M, Dwight SS, Gaulton A, Licata L, Melidoni AN, Ricard-Blum S, Roechert B, Skrzypek MS, Tiwari M, Velankar S, Wong ED, Hermjakob H, Orchard S. 2015. The complex portal–an encyclopaedia of macromolecular complexes. Nucleic Acids Res 43:D479–D484.

Mortimer RK. 2000. Evolution and variation of the yeast (Saccharomyces) genome. Genome Res 10:403–409.

Mortimer RK, Johnston JR. 1986. Genealogy of principal strains of the yeast genetic stock center. Genetics 113:35–43.

Skrzypek MS, Hirschman J. 2011. Using the Saccharomyces Genome Database (SGD) for Analysis of Genomic Information. Curr Protoc Bioinformatics Chapter 1:Unit1.20.