Chapter 2

An Introduction to the Saccharomyces Genome Database (SGD)

Olivia W. Lang, Robert S. Nash, Sage T. Hellerstedt, Stacia R. Engel, and The SGD Project

Abstract

The *Saccharomyces* Genome Database (SGD) is a well-established, key resource for researchers studying *Saccharomyces cerevisiae*. In addition to updating and maintaining the official genomic sequence of this highly studied organism, SGD provides integrated data regarding gene functions and phenotypes, which are extracted from the published literature. The vast amount and variety of data housed in the database can prove challenging to navigate for the first-time user. Therefore, this chapter serves as an introduction describing how to search the database in order to discover new information. We introduce the different types of pages on the website, and describe how to manipulate the tables and diagrams therein to display, download, or analyze the data using various SGD tools.

Key words *Saccharomyces cerevisiae*, Genome database, Phenotype, Gene ontology, Yeast

1 Introduction

The *Saccharomyces* Genome Database (SGD; www.yeastgenome.org) is a public, encyclopedic resource for the budding yeast *Saccharomyces cerevisiae*. As the official keepers for the genome sequence and gene nomenclature of yeast, SGD is a core component of the yeast research community and an essential tool for experimental design and data analysis [1]. The SGD project extracts function and gene information from published literature through a process of human curation, and pools these facts with data from other repositories to create a resource for researchers that integrates sequence data with evidence-based annotations. It is this core activity of manual curation that is essential for converting the

The members of the SGD Project are listed in the acknowledgements section.

vast amount of published data into a coherent body of knowledge. All information in SGD can be traced back to its original source so that users can review the experiment details and interpretations from the authors.

The tools within SGD offer different ways for users to make both specific and broad queries. YeastMine offers a way of performing bulk queries of information from the database [2, 3] while SPELL searches SGD's curated expression datasets for genes with similar expression profiles [4]. Other tools available for exploring data at SGD include GO Term Finder [5, 6], GO Slim Mapper [5], YeastPathways [3], JBrowse genome browser [7, 8], and Variant Viewer [9]. Since users primarily use SGD as an encyclopedia for yeast genes, this chapter focuses on how to access the data at SGD using the website search and gene pages.

The majority of the manual curation efforts at SGD use controlled vocabularies to capture the main research findings from the scientific literature. SGD primarily uses the Gene Ontology (GO; www.geneontology.org; [10]) and the Ascomycete Phenotype Ontology (APO; [11]) to capture functional information about yeast genes. Three categories of Gene Ontology terms capture information on the function of a gene product: Molecular Function, Biological Process, and Cellular Component.

Many SGD users primarily use SGD to look up specific yeast genes' roles and functions. In order to successfully navigate the website and perform fruitful searches, it is important to have a basic understanding of how data are captured and displayed on the website. This chapter will explain how to use the website's search tool, and also how data are displayed in tables and network diagrams in pages across the SGD site (*see* **Note 1**).

2   Methods

2.1 Exploring SGD Through a Search Query: Finding a Gene Page

A great way to start exploring the kind of information captured by SGD is to learn to use SGD's website search tool. The search includes many features that help you narrow your results to find the information you are looking for. For this example, we will step through how to find cell-cycle-related protein-coding genes.

2.1.1 Navigating from SGD's Home Page

Open the SGD home page (www.yeastgenome.org) and browse the purple menu bar near the top of the page to find links to SGD tools and resources like the Genome Browser and SGD's instance of YeastMine (labeled "Gene Lists" in the Analyze section). The home page also includes a slideshow of images provided by members of the community, a list of upcoming meetings and conferences, and a news section for blog posts and announcements (Fig. 1).

Fig. 1 SGD home page. (1) Browse the purple menu bar for links to SGD tools and external resources. (2) Type your query into the search box. (3) Clicking "Show all results" will show all results matching the search string and will not take you directly to an SGD page. (4) For quick-links that send you directly to SGD pages, select one of the options in the quick results box. (5) Not sure what to search for? Then click the "Try this?" button to get the results from one of a set of selected biological terms and phrases

2.1.2 Enter Search

Enter a search query, like "cell cycle," into the search box within the purple bar. As you type, you will notice results showing up below the search box. Clicking on one of these results will send you directly to the Gene/Term/Reference page. Keep in mind that if your search query is a unique gene name, PubMed Identifier, or Gene Ontology Identifier (prefixed with "GO:"), you will be sent directly to the corresponding gene, reference, or GO term page. If you do not want to be sent directly to an SGD data page, click on "Show all results …" which pops up with the autofill options when you are typing in the search box (*see* **Note 2**).

2.1.3 Search Results

Filter your results to the data type you are looking for by selecting one of the categories on the left-hand side (Fig. 2). The number of results for a category is displayed next to the category name and colored circle. For this example, select the blue "Gene" category.

2.1.4 Drilling Down

After selecting a category, the left bar will change to display further filter options. For the Genes category, you can filter by "Feature Type" or by associated Gene Ontology terms. Five of the filter values are displayed by default, but you can click "Show more" and "Show all" to see more options. For this example, select the "ORF" feature type to show only the protein-coding genes (*see* **Note 3**).

Fig. 2 SGD search results page. (1) A search using the string "cell cycle" returns 34,110 results. (2) The colored circles in the left bar indicate the types of SGD pages captured in the results and the numbers in the list of categories on the left-side bar indicate the number of results from each category. (3) Each result is an SGD page with the type of page indicated at the right and highlighted text showing how it relates to the search query

2.1.5 Editing Your Search

It is possible to edit your search by removing filters in a search query. As you progress through a search, blue rectangles are added to the top of the results page with text relating to the filter applied. These "breadcrumbs" can show you the search path to get to the current results page and by clicking the "x" you can remove the specific breadcrumb. For this example, you could remove the "ORF" breadcrumb to retrieve all the genes relating to the cell cycle.

2.1.6 Wrapped View

To download the list of genes from this search query, click the "Wrapped" button next to the highlighted "List" button. When the page is fully loaded, you will be able to download the list of genes by clicking the "Download" button next to the "Analyze" button. The function of the "Analyze" button will be explained in the section below (Fig. 3).

2.1.7 Saving Search Results in a URL

If you wish to save your search results for later, just copy and save the URL in your browser when you are on the search results page. This is especially useful for sharing search results with colleagues or bookmarking your results.

2.1.8 Selecting a Gene

To enter the SGD locus summary page, click the blue name of the gene either in the "List" or "Wrapped" view (*see* **Note 4**).

Fig. 3 With the selection of the "ORF" filter, the results of your query narrowed down to 3239 results. (1) Note the blue rectangles that appear near the number of search results with the addition of each new filter. These are called "breadcrumbs" because you can use them to trace back the filters applied to your query. You can also remove filters by clicking the "x" for each breadcrumb you wish to remove. (2) The results are displayed using the "Wrapped" view here. The "Wrapped" view includes the options of (3) downloading the results as a text file, or (4) sending the list to the "Analyze" tool

2.2 Annotation Tables at SGD

SGD displays much of its data in tabular form on various pages of the website. Tables are used to display data including ontology annotations, gene interactions, posttranslational modifications, and protein abundance. In this section, we will be exploring the annotations on the Gene Ontology Term page for "GTP-binding" (GO:0005525).

2.2.1 Navigating to the Gene Ontology Term Page

From the SGD home page, start typing in "GTP-binding" until you see the Molecular Function term pop-up in the list of quick-results. Click on the result to go straight to the GO term page (*see* **Note 5**).

2.2.2 Navigating to the GO Annotations Table

Scroll to the "Annotations" table by using the left-hand vertical menu or by scrolling past the term overview with description and past the figure of related ontology terms. Next to the "Annotations" header is a count of the number of annotations and the number of associated genes.

2.2.3 Reading GO Annotations

Each row represents an annotation that associates a gene to the "GTP-binding" molecular function term. For example, the first entry of this table was curated by SGD on March 23, 2011 to say: the study by Shin et al. (2011) [12] uses a direct assay to demonstrate that the FUN12 gene interacts selectively with guanosine triphosphate. If you need more information, the blue text within the tables often links to other SGD pages. Click on a gene if you wish to go to the SGD Locus Summary Page for that gene (Fig. 4).

Fig. 4 An SGD data table. (1) Click the up/down arrows next to the column header to sort the table by the values within a given column (2) By typing into the "Filter table" search, the table will only display entries that have matching text. You can see more annotations by (3) displaying more annotations per page or (4) flipping through the pages of annotations. (5) The "Download" button saves the contents of the table into a text file, while the (6) "Analyze" button leads to the Analyze page where you can send the list of genes into one of SGD's tools

2.2.4 Sort Feature for the Table	You can sort the table by values within each of the header sections, like "Gene" or "Source," by clicking on the up and down arrows within the header boxes. This is commonly used to see which genes have been annotated with the GO term and how many of those annotations have been made.
2.2.5 Search the Table	The tables at SGD also have a search feature in the upper right side of the table. Click the blue circle with the "?" for help information on how to use it. Typing anything within the search box will filter the table to show only annotations that contain the matching text anywhere within the entry.
2.2.6 Downloading Annotations	Most tables at SGD have a "Download (.txt)" button where you can download the current contents of the table in the order that they are being displayed. For example, if you sorted the entries by

"Source" and entered text within the search box to filter the annotations displayed, the downloaded file will save the subset of features selected by the search all ordered by source.

|	
2.2.7 Analyze Toolbox	Clicking on the "Analyze" toolbox below the table will send you to the analyze page where you can use the selected genes from the table as input for several of SGD's tools. You can send the list of genes to the "GO Term Finder" tool, the "GO Slim Mapper" tool, the gene expression tool, "SPELL," or the popular "YeastMine" tool (*see* **Note 1**).
2.3 Network Diagrams at SGD	Some SGD pages have interactive Network Diagrams that offer a visual approach to studying the data. These diagrams are used on several different pages (Gene Ontology, Phenotype, Interaction, Regulation, and Literature), and also for displaying the ontologies on GO term and APO term pages.
2.3.1 Go to a Protein Page	Go to the SGD gene page of your favorite gene and select the "Protein" tab at the top. The first section includes overview information about the protein product and the subsequent "Experimental Data" section includes information from published datasets.
2.3.2 Protein "Domains and Classification" Section	Scroll to the "Domains and Classifications" section to view the table, the "Domain Locations" diagram, and the "Shared Domains" network diagram. For more information on any of these figures or tables, click the blue "i" icon for details on how the data were generated or how the data are displayed (Fig. 5).
2.3.3 "Shared Domains" Network Section	Scroll to the "Shared Domains" network diagram and click on the blue "i" icon for information about the network and the blue "?" for information on how to interact with the network. You can find these blue icons all across the SGD website pages next to headers and within tables. Clicking on these icons will give you important information about how the data at SGD is captured and displayed. The date of the last update to the diagram is displayed in the network diagram on the lower right (Fig. 5).
2.3.4 Clicking on the Nodes	Clicking on the nodes will send you to an SGD page. There are two kinds of nodes: domain nodes (square) that send you to SGD domain pages, and gene nodes (round) that send you to the SGD locus summary page for the gene. The domain nodes are connected to genes that contain that domain. The gene nodes are connected to nodes representing domains that are found within that gene. The legend is displayed at the bottom left (Fig. 5).
2.3.5 Dragging Nodes/ Groups of Nodes Around	You can move around the nodes of the graph by clicking on a node and dragging it to where you want it to be. For particularly large networks, it is useful to highlight a group of nodes and move the
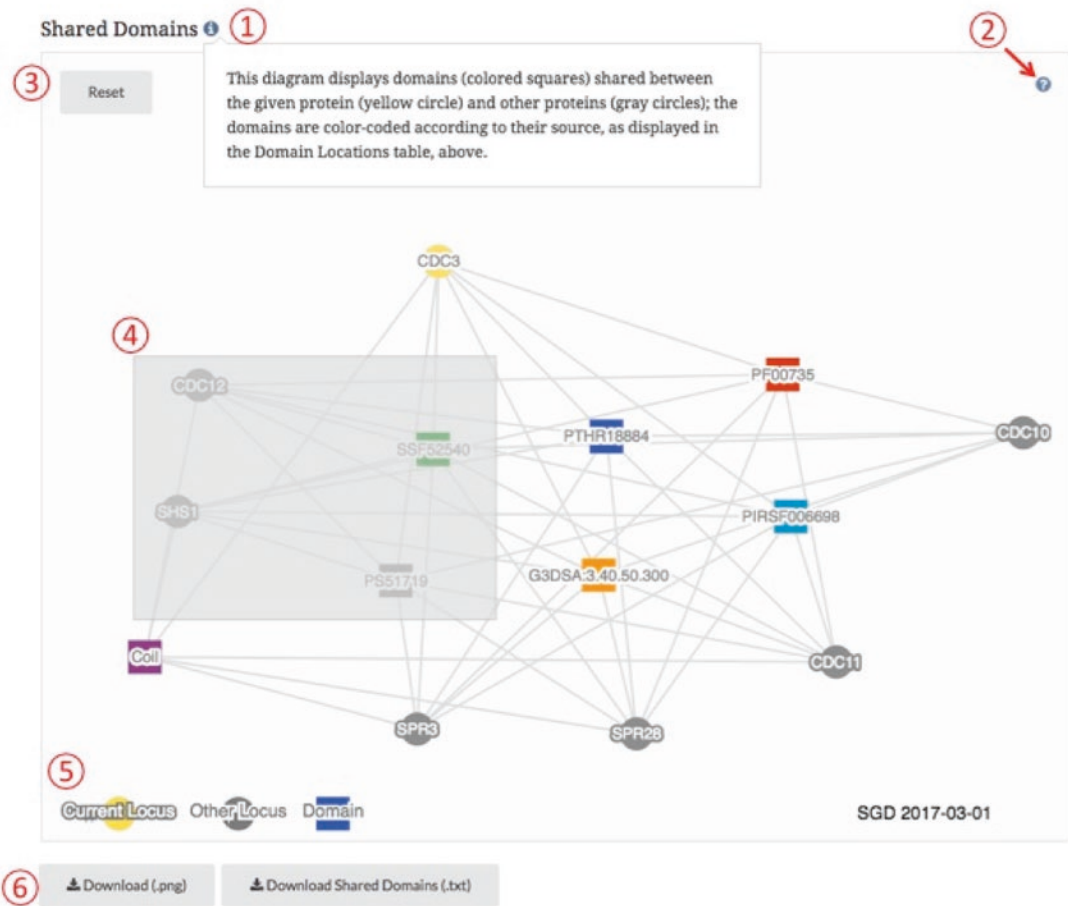
Fig. 5 An SGD network diagram. (1) Clicking the blue info icon displays an explanation of how the data are displayed. (2) Clicking the blue question mark displays help descriptions for how to interact with the diagram. (3) Clicking the "Reset" button returns the diagram to its original state. (4) Dragging your mouse across the figure will select a group of genes that you can manipulate together. (5) The legend describes the differences between the domain and gene nodes. (6) The Download options allow download of a .png or .txt file of the figure

group together. To do this, click in the empty space of the graph and drag the mouse to expand a rectangle that will enclose the nodes you wish to group. This selects the group of genes so you can click and drag any one of the highlighted nodes to move the entire group (Fig. 5).

2.3.6 Resetting the Diagram

At any time you wish to reset the network diagram to the original conformation, click the "Reset" button in the upper left corner of the diagram.

2.3.7 Downloading the Diagram

To download the network as an image, click "Download (.png)," or download the data displayed in the network diagram as a file using the "Download (.txt)" button.

3   Notes

1. For help on how to use specific tools, or other aspects of the SGD, please visit the SGD YouTube channel (youtube.com/c/SaccharomycesGenomeDatabase) for video tutorials or send an e-mail to the SGD help desk (sgd-helpdesk@lists.stanford.edu).

2. If you are searching for a specific phrase, you can use quotations as in a Google search to group words into a specific order (ex., "cell cycle").

3. When the "Genes" category is selected, the left bar changes to display the subcategory options for further filtering. The "Show more" button displays more filter options. You can return to view the other categories by clicking the "Show all categories" button above the left bar.

4. By clicking any of the blue gene names in the search results, you will be sent to an SGD Locus Summary page. The summary page includes an overview description and highlights from various topics of information. From here, you can look at the details of the information SGD captures by clicking any of the tabs at the top or any of the links to the right of the section header. For example, clicking the "Phenotypes" tab will take you to a page with phenotype annotations for the gene of interest.

5. Alternatively, you can enter in the term ID (GO:0005525).

Acknowledgments

References

1. 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 (2013) The reference genome sequence of Saccharomyces cerevisiae: then and now. G3 (Bethesda) 4:389–398

2. Engel SR, MacPherson KA (2016) Using model organism databases (MODs). Curr Protoc Essential Lab Tech 13:11.4.1–11.4.22

3. Cherry JM (2015) The Saccharomyces Genome Database: advanced searching methods and data mining. Cold Spring Harb Protoc. https://doi.org/10.1101/pdb.prot088906

4. 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(20):2692–2699

5. Skrzypek MS, Binkley J, Sherlock G (2016) How to use the Candida Genome Database. Methods Mol Biol 1356:3–15

6. Cherry JM (2015) The Saccharomyces Genome Database: gene product annotation of function, process, and component. Cold Spring Harb Protoc. https://doi.org/10.1101/pdb.prot088914

7. Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH (2009) JBrowse: a next-generation genome browser. Genome Res 19(9):1630–1638

8. Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G, Goodstein DM, Elsik CG, Lewis SE, Stein L, Holmes IH (2016) JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biol 17:66

9. Sheppard TK, Hitz BC, Engel SR, Song G, Balakrishnan R, Binkley G, Costanzo MC, Dalusag KS, Demeter J, Hellerstedt ST, Karra K, Nash RS, Paskov KM, Skrzypek MS, Weng S, Wong ED, Cherry JM (2016) The Saccharomyces Genome Database variant viewer. Nucleic Acids Res 44(D1):D698–D702

10. 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(1):25–29

11. 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. https://doi.org/10.1093/nar/gkp917

12 Shin B-S, Acker MG, Kim J-R, Maher KN, Arefin SM, Lorsch JR, Dever TE (2011) Structural integrity of alpha-helix H12 in translation initiation factor eIF5B is critical for 80S complex stability. RNA 17(4):687–696