Published in final edited form as:

Methods Mol Biol. 2016 ; 1356: 3–15. doi:10.1007/978-1-4939-3052-4_1.

How to use the *Candida* Genome Database

Marek S. Skrzypek, Jonathan Binkley, and Gavin Sherlock

Department of Genetics, Stanford University, Stanford, CA 94305-5120

Summary

Studying *Candida* biology requires access to genomic sequence data in conjunction with experimental information that provides functional context to genes and proteins. The *Candida* Genome Database (CGD) integrates functional information about *Candida* genes and their products with a set of analysis tools that facilitate searching for sets of genes and exploring their biological roles. This chapter describes how the various types of information available at CGD can be searched, retrieved, and analyzed. Starting with the guided tour of the CGD Home page and Locus Summary page, this unit shows how to navigate the various assemblies of the *C. albicans* genome, how to use Gene Ontology tools to make sense of large-scale data, and how to access the microarray data archived at CGD.

Keywords

Candida; genome database; expression analysis; gene ontology; GO slim

Introduction

The *Candida* Genome Database (CGD; <http://www.candidagenome.org>) started in 2004 as an online compendium of genomic, genetic and molecular biology information about *Candida albicans*. At its inception, CGD was based on the then newly assembled genomic sequence of strain SC5314 [1]. The primary goal was to couple the sequence data with the literature-derived experimental data in a single, easy to navigate web-based resource. Since then, the expanding scope of *Candida* research, facilitated by the progress in sequencing of genomes from other strains and species, has prompted the incorporation of similar data for *C. glabrata* CBS138, *C. parapsilosis* CDC317, and *C. dublinensis* CD36. CGD has also become an archive that provides access to genomic sequences for other related strains and species, including *C. albicans* WO-1, *C. guilliermondii* ATCC_6260, *C. lusitaniae* ATCC_42720, *C. orthopsilosis* Co 90-125, *C. tropicalis* MYA-3404, *Debaryomyces hansenii* CBS767, and *Lodderomyces elongisporus* NRLL YB-4239.

At the core of CGD lies human curation, a process that involves manually extracting gene-specific experimental information from the published, peer-reviewed literature and associating those annotations with the relevant genomic features. Genes and their annotations are organized in such a fashion that the information is easily browsable, searchable and retrievable for further analysis and perusal. CGD curators also make sure that every piece of information is traceable to its original source, usually a publication in a scientific journal, thus providing access to all available experimental details and their

interpretations. CGD also includes a rigorous analysis of orthology between species [2] and of protein domain structure, which allows consistent predictions of functions for genes that have not been experimentally characterized [3]. Thus, CGD provides a structured, unbiased and continuously updated collection of a large variety of experimental results and computational predictions that has become indispensable for *Candida* researchers

In order to ensure a uniform representation of biological information across different organisms, most genome databases use controlled vocabularies to annotate various attributes of genes and gene products. The most widely used vocabulary for capturing the key aspects of gene product biology is the Gene Ontology (GO; <http://www.geneontology.org/>; [4]). GO is a system of standardized terms with defined relationships that describe a primary activity of the gene product (Molecular Function), a broader cellular role the gene product is involved in (Biological Process), and the predominant localization, such as a protein complex, a subcellular structure, or an organelle (Cellular Component). CGD uses GO as the main vocabulary to annotate genes. Another data type that CGD captures, mutant phenotypes, is curated using Ascomycete Phenotype Ontology (APO), a vocabulary developed at *Saccharomyces* Genome Database (SGD) [5] that we adapted to the specific needs of *Candida* biology.

The information in CGD is organized in a system of interlinked web pages, designed with the goal of making them intuitive, easily navigable and user-friendly. However, the sheer complexity of the data presented in CGD can make it difficult for a newcomer to find the right information. This chapter provides help for navigating the site and highlights some new features of CGD. We present an overview of the main entry point, the Home Page, and the central organizing principle of the database, the Locus Summary Page. We show how to navigate between the current assembly of the *C. albicans* genome and the many historical assemblies that are still widely in use and need to be accessed frequently. We also explain how to perform some of the most common types of analysis that utilize GO annotations. Finally, we show how to access and browse large-scale datasets collected at CGD.

3. Methods

3.1 Exploring CGD Home and Locus Summary pages

The CGD home page (<http://www.candidagenome.org>) serves as a place for database-related announcements, community news, and upcoming meetings of interest to the *Candida* community. It also provides a starting platform for many of the features and tools available at CGD. The search box is the primary gateway to a variety of data types available in the database and is present on most CGD pages (see Note 1). The Locus Summary page (Figure 1) offers an up-to-date summary of what is known about a particular gene (see Note 2).

1. Open the home page (<http://www.candidagenome.org>) and explore the Banner at the top of the page (see Figure 1A). This Banner, present on most CGD pages, provides quick links to multiple data search and analysis tools, literature tools including Textpresso, a full-text searching tool [6], as well as bulk download tools and various community-related information. Hover the mouse over each item to reveal a drop-down menu with available options.

2. Enter your query into the “Search our site” box above the Banner. If you enter a unique gene name, such as a systematic name, clicking on the result will jump directly to the Locus Summary page for that gene (see Note 3). If your query produces multiple hits, for instance, a gene name that is used in several *Candida* species represented in CGD, you will get a “CGD Quick Search Result” page that lists the type and number of hits, general and broken down by species. Positive hits are hyperlinked to either their respective Locus Summary pages, or to an intermediate list of individual hits. Select an individual hit to open its Locus Summary page.
3. The Locus Summary page contains several tabs, with the Summary tab open by default. Typically, for a protein-coding gene, the other tabs are Locus History, Literature, Gene Ontology, Phenotype, Homologs, and Protein. Flip back and forth through the tabs to see specific types of information they include.
4. The Basic Information section at the top of the Summary tab (see Figure 1A) contains a description that summarizes the most significant features of the locus in a concise, headline-like format. It also lists all the names associated with the locus including the standard name - typically the genetic name under which the gene was first published. If there are other names by which the gene has been referred to in the literature, they are listed as aliases. The identifier assigned during genome sequencing is listed as Systematic Name along with the name of the reference strain used in the sequencing project (see Note 4). The Basic Information section allows easy navigation to genes in other organisms. Click on any hyperlinked name listed among “Orthologous genes in *Candida* species” to open its Locus Summary page in CGD. You can also click on the “View ortholog cluster” link, which will show the orthologs from 15 *Candida* species in their genomic context, a report produced by the Candida Gene Order Browser [2]. In order to explore orthologs from other fungal species, select a gene from “Ortholog(s) in non-CGD species” and you will access information at other resources: AspGD, Broad Institute, PomBase, and SGD, for genes from *A. nidulans*, *N. crassa*, *S. pombe*, *S. cerevisiae*, respectively (see Note 5). At the bottom of this section, there is a thumbnail showing the chromosomal location of the gene, hyperlinked to GBrowse2 [7], which provides a graphical interface to inspect the genomic context of the gene.
5. The GO Annotations and Mutant Phenotypes sections (Figure 1B) show current information about the function of the gene. For more detailed information, including the references on which these annotations are based, open the Gene Ontology or Phenotype tab, respectively. GO annotations on the Summary tab are divided by the GO aspect (Molecular Function, Biological Process, Cellular Component) and by the annotation method. Manually curated annotations are assigned by a CGD curator on the basis of published, experimental (in most cases) results. Computational annotations are produced by transferring experiment-based GO annotations from orthologous genes in other species, or by predictions based on domain structure. Each annotation is accompanied by an evidence code that indicates the reason behind the annotation, for instance, IMP means Inferred from Mutant Phenotype. Click on any evidence code to see a table of all the evidence

codes and their definitions. Annotations based on sequence similarity, genetic or physical interactions also list the source gene(s) and the organism. Click on any gene name to see a report for that gene in CGD or in an external database. Each GO term itself is hyperlinked to another CGD page that provides more information, including term definitions, a diagram depicting the relevant segment of the ontology, and a list of other genes in CGD that are also annotated with that term. Similarly, each mutant phenotype is hyperlinked to a page that lists all other genes in CGD that display the same phenotype.

6. The Sequence Information section (Figure 1C) shows the basic data about the gene (chromosomal coordinates, intron-exon structure), but also provides easy access to sequences and sequence analysis tools. Open the drop-down menu next to Retrieve Sequences to see available options that include retrieval of DNA sequence in several configurations and, for protein-coding genes, for the predicted protein sequence as well. Similarly, open the Sequence Analysis Tools drop-down menu to start BLAST searches, restriction analysis or primer design tools. More analysis tools are available from the Banner on top of the page; click on Search, Sequence, Tools, or Download to see the options. In addition, at the bottom of the Sequence Information section there are links to sequence data available from external resources, such as GenBank, UniProt and others.

3.2 Navigating between the current *C. albicans* assembly and historical assemblies

Clinical isolates of *C. albicans* (including the sequenced reference strain, SC5314) are diploids, and exhibit a high level of heterozygosity: on average, a single nucleotide polymorphism (SNP) occurs once every 237 bases between homologous chromosomes [1]. The first genome sequence published for *C. albicans* (Assembly 19) included both haplotypes – that is, the sequence from both members of every homologous chromosome pair was represented. However, full chromosome sequences were not available: rather, the assembly consisted of several hundred contigs (contiguous segments inferred from overlapping sequencing reads). These ranged in size from roughly 1,000 to 420,000 base pairs, but were not mapped to, or ordered along their parent chromosomes. The next major sequence release for the reference strain (Assembly 21) was assembled to the full chromosome level [8]. However, this assembly presented a pseudo-haploid genome (also known as reftigs), consisting of ordered, non-redundant sequences, but with any given contiguous region not necessarily derived from the same haplotype. The latest *C. albicans* assembly (Assembly 22; [9]) addresses both issues: it includes both haplotypes, and is assembled to the chromosome level. Since much research was conducted that referred to the older assemblies, it is important to be able to easily navigate between the current and past assemblies, and to relate them to each other. CGD provides tools and mapping files to accomplish this. Here we describe two ways to relate information between chromosomal features from different *C. albicans* assemblies.

3.2.1. Navigate between current and historic features using GBrowse

1. From the Locus Summary Page of a feature of interest, scroll down to the section entitled “Chromosomal Location”. Two small map images are displayed, showing

the immediate genomic region around the feature in each of the two Assembly 22 haplotypes. Click on one of the images to open the GBrowse tool.

2. A new window opens, displaying a browser with a map of the feature plus the regions 10,000 bases upstream and downstream on the chromosome. Above the browser window on the left side, click on the tab labeled “Select Tracks”. Choosing this tab reveals several groups of check-boxes, which control what is displayed in the browser. One group is titled “Historic assemblies”. Choose to display features from Assemblies 21 and 19 by checking “Assem21 ORF” and “Assem19 ORF (AWG)”, respectively, then click “Back to Browser” above the check-boxes to return to the browser window.
3. The browser now displays the same Assembly 22 region and features as before, with the selected historic tracks displayed above. Most of the historic features will correspond to current features, but the comparison may reveal features that have been added or deleted in the current assembly relative to the older assemblies. Clicking on the icons for the historic features opens a page from which files in GFF format, containing location information for the historic features, can be downloaded.

3.2.2. View the Assembly 21 Locus Summary Page

1. From the Locus Summary Page of a feature of interest, scroll down to the section entitled “Assembly 19/21 Identifier”. Click on the link to the right, which is labeled with the old identifier (beginning with “orf19”).
2. A new page opens, displaying a summary of information for the same feature in the context of Assembly 21. Much of the information on this page is redundant with or superseded by information on the current (Assembly 22) page, but of interest are the sections labeled “Chromosomal Location” and “Sequence Information”.
 - a. The Chromosomal Location section shows several small maps. The top map shows the feature and its immediate genomic region for Assembly 21. Clicking on the map opens a GBrowse window, from which the user may navigate to other Assembly 21 features. The lower maps, if available, show the feature for both haplotypes in Assembly 19.
 - b. The Sequence Information section has two pull-down menus. From the “Retrieve Sequences” menu the user can display or download sequences for the feature: genomic DNA (with or without flanking DNA), coding DNA (introns removed), or protein sequence (see Figure 1C). Sequences are also available from this menu for the “allele” – the feature from Assembly 19 corresponding to the opposite haplotype. The second pull-down menu, labeled “Sequence Analysis Tools”, allows users to perform BLAST searches, design primers, or create restriction-fragment maps using sequences from the Assembly 21 feature or the allele.

3.3. GO Term finder and GO Slim Mapper, tools for GO analysis

The GO vocabularies are constructed as hierarchies, where more general terms, so-called “parents”, encompass more specific “child” terms. When making annotations curators are required to assign the most specific (granular) term that is supported by the evidence presented in the publication. Within the hierarchical structure of the ontology, annotation to a child term implies that annotation to its parent term is also correct. This feature of GO has to be taken into account during analysis of large-scale data, when it is necessary to identify common biological features in a set of genes that are, for instance, co-regulated in a microarray experiment. Finding statistically significant similarities in GO annotations for multiple genes requires a tool, GO Term Finder [10], that is able to determine if there are any GO terms that annotate genes in a list (either directly, or indirectly via the GO hierarchy) at a rate greater than would be expected by chance. Given the complexity of GO, it can also be desirable to group genes into broad categories using only high-level terms. The GO Slims are such lists of high-level terms from each ontology branch (Molecular Function, Biological Process, and Cellular Component), carefully selected to cover most of the curated GO information in CGD. The tool, GO Slim Mapper, is able to categorize (map) a large set of genes to user-selected GO Slim terms.

3.3.1. Using GO Term Finder

1. Open the GO Term Finder page by selecting it from the options in the Search or GO pull-down menus in the Banner (see Note 6).
2. Select the species in Step 1; the default species is *Candida albicans*, but you can run this analysis for other CGD-curated species: *C. glabrata*, *C. parapsilosis* or *C. dublinensis*.
3. In Step 2, enter a list of gene names. You can either type the name of the genes in the input box or upload a file that contains the list. Either genetic names (CGD Standard Names, e.g., ‘AAF1’) or systematic names (‘C3_06470W_A’, or orf19 identifiers, e.g., ‘orf19.7436’) may be used (see Note 7).
4. In Step 3, select one of the three branches of GO (biological process, molecular function, or cellular component) by checking the boxes. The tool only searches one of the three branches at a time.
5. Click the Search button after Step 2 to use the default settings or go further down to Steps 3 and 4 to specify and customize your background set and/or refine the types of annotations in your background set.
6. You may change your background set in Step 4. The default background set includes all the genes in the database that have at least one GO annotation. You can also customize the background set by choosing which feature type(s) it should include.
7. In Step 5 you can deselect specific types of GO annotations that should not be used for calculations. By default, annotations collected by all methods and with all types of evidence are included,

8. The results page displays the significant shared GO terms (or their parents) in both graphic and table form, within the set of genes entered on the previous page.
9. The graphic shows the GO tree that includes terms used directly or indirectly in annotations for the genes in your list. The terms are color-coded to indicate their statistical significance (p-value score). Genes associated with the GO terms are shown in gray boxes, with links to their respective Locus Summary pages.
10. The table below the graph lists each significant GO term, the number of times the GO term is used to annotate genes in the list and the number of times that the term is used to annotate genes in the background set (see Note 8).
11. Additional columns list the p-value, the False Discovery Rate (FDR), and a list of all the genes annotated, either directly or indirectly, to the term. FDR is an estimate of the percent chance that a particular GO term might actually be a false positive. It represents the fraction of the nodes with p-values as good or better than the node with this FDR that would be expected to be false positives.
12. The statistical significance of the association of a particular GO term with a group of genes in the list is indicated by the p-value: the probability of seeing at least x number of genes out of the total n genes in the list annotated to a particular GO term, given the proportion of genes in the whole genome that are annotated to that GO Term. The closer the p-value is to zero, the more significant the particular GO term association with the group of genes is (i.e. the less likely to occur by chance).

3.3.2. Using GO Slim Mapper

1. Select GO Slim Mapper from the options in the Search or GO pull-down menus in the Banner. In Step 1 you can type or paste your list of genes or upload them as a file.
2. In Step 2 use the pull-down menu to select the GO Set Name: GO Slim Component, GO Slim Function, or GO Slim Process. The list of terms from the selected set appears in the window in Step 4.
3. In Step 4 you can specify which particular term, or terms, you want to use. The default setting is "Select ALL Terms.", but you can highlight only terms you are interested in.
4. In optional Step 5 you can exclude certain types of annotations, eg. Computational and High-throughput.
5. Click Search to start the process; note that long lists of genes will take significant time to analyze.
6. Results appear in a table with three columns: the GO Slim terms chosen, with a link to graphical depiction of that branch of GO, the percentage of genes in your list annotated to each term, and the genes from your your list that are annotated to that term. You can also download the results in a tab-delimited file.

3.4 Browsing large-scale data and GeneXplorer

CGD has been collecting large-scale datasets from its inception. The datasets from authors' websites or from supplementary materials that accompany publications are archived in the Datasets page accessible from the Download menu in the Banner, thus offering a convenient one-stop shop for datasets pertaining to *Candida* that can be then imported into one's own tools for analysis. The microarray results are displayed by GeneXplorer [11], a web tool that allows browsing and searching of the expression data at the CGD site, with the benefit of an instant access to functional annotations.

1. To access the archived datasets, click on the Download pull-down menu in the top Banner and select Datasets. The resulting page lists all the publications from which the datasets are archived in CGD, ordered by year and author. All the references have icons that lead to the abstract, full-text articles and to the original data website, as well as to the download directory at CGD. The microarray-containing references have a pair (or pairs) of green-blue icons. Click the one on the right that is labeled “.pcl” and you will get the entire datasets in the pcl format, a standard format that makes the dataset amenable to analysis by many popular software suites. Find a publication of interest and click the icon on the left to see the data in GeneXplorer (Figure 2A).
2. The GeneXplorer page (Figure 2B) shows a heat map image of clustered expression profiles from the selected publication, where increased or decreased gene expression is shown in shades of red or green, respectively. Zoom in and out using the + or – buttons above the cluster. Select any particular region by dragging over it. Click on a profile for a particular gene to see the expression patterns of the most similarly or dissimilarly expressed genes. Find a gene of interest by entering its name into the “Search for” box. Click on any gene name to go to its Locus Summary page.

4. Notes

1. The query entered into the search box may be a *Candida* gene or protein name (standard or systematic name, or an alias), author or colleague name, PubMed ID, or any keyword (such as a functional term or phenotype). It can even be a name of an ortholog from one of the non-CGD species. When there are multiple hits, a list of matches is displayed.
2. In addition to protein-coding genes, several other entities have their Locus Summary pages in CGD, including various RNA genes (tRNA, rRNA, snRNA, etc.), as well as a variety of chromosomal features, such as centromeres, telomeres, repeated sequences, and others.
3. “Search our site” box accepts a wildcard character *. For example, enter “act*” to retrieve any piece of data starting with “act”. Also, the search box has an autocomplete feature, which provides suggestions when you start typing your query.

4. For *C. albicans*, the systematic name shown on the Locus Summary page is always the name of the haplotype A allele, as denoted by an “A” suffix, with the corresponding haplotype B allele listed below. Since the systematic names from previous assemblies of *C. albicans* genome (so called “orf19” names) continue to be widely used, the Assembly 19/21 identifier is also shown.
5. The ortholog mappings among *Candida* strains, and between *Candida* strains and *S. cerevisiae*, are derived from the curated syntenic groupings at the Candida Gene Order Browser (CJOB) [12]. The ortholog mappings between *Candida* strains and *S. pombe*, *A. nidulans*, and *N. crassa* are made by pairwise comparisons using the InParanoid software [13].
6. Links to both GO Term Finder and GO Slim Mapper appear at the bottom of many pages with results of searches that produce a list of genes. Selecting those links open the respective tool with the gene list already pre-loaded into the input box.
7. When a name is entered that is an alias for one gene or feature, the program will map the name to that gene. If the name is an alias for more than one gene but not a Standard or Systematic name for any genes, the program will present a list of possible mappings. The user can decide which gene was intended and edit the input.
8. Because the frequency of any given annotation within the background set is compared against the frequency of the annotation within the query set (input), the choice of background set affects the significance of the results that are returned by the tool. Please note that the specific background set of genes that was used in the absence of any user-defined set (the default background set) has changed over time.

Acknowledgments

This work was supported by National Institute of Dental and Craniofacial Research at the US National Institutes of Health (grant no. R01 DE015873).

References

1. Jones T, Federspiel NA, Chibana H, Dungan J, Kalman S, Magee BB, Newport G, Thorstenson YR, Agabian N, Magee PT, Davis RW, Scherer S. The diploid genome sequence of *Candida albicans*. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101:7329–7334. [PubMed: 15123810]
2. Fitzpatrick DA, O’Gaora P, Byrne KP, Butler G. Analysis of gene evolution and metabolic pathways using the Candida Gene Order Browser. *BMC genomics*. 2010; 11:290. [PubMed: 20459735]
3. Inglis DO, Arnaud MB, Binkley J, Shah P, Skrzypek MS, Wymore F, Binkley G, Miyasato SR, Simison M, Sherlock G. The Candida genome database incorporates multiple *Candida* species: multispecies search and analysis tools with curated gene and protein information for *Candida albicans* and *Candida glabrata*. *Nucleic acids research*. 2012; 40:D667–674. [PubMed: 22064862]
4. 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. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*. 2000; 25:25–29. [PubMed: 10802651]
5. 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. *Saccharomyces Genome*

Database provides mutant phenotype data. *Nucleic acids research*. 2010; 38:D433–436. [PubMed: 19906697]

6. Muller HM, Kenny EE, Sternberg PW. Textpresso: an ontology-based information retrieval and extraction system for biological literature. *PLoS biology*. 2004; 2:e309. [PubMed: 15383839]
7. Stein LD. Using GBrowse 2.0 to visualize and share next-generation sequence data. *Briefings in bioinformatics*. 2013; 14:162–171. [PubMed: 23376193]
8. van het Hoog M, Rast TJ, Martchenko M, Grindle S, Dignard D, Hogues H, Cuomo C, Berriman M, Scherer S, Magee BB, Whiteway M, Chibana H, Nantel A, Magee PT. Assembly of the *Candida albicans* genome into sixteen supercontigs aligned on the eight chromosomes. *Genome biology*. 2007; 8:R52. [PubMed: 17419877]
9. Muzzey D, Schwartz K, Weissman JS, Sherlock G. Assembly of a phased diploid *Candida albicans* genome facilitates allele-specific measurements and provides a simple model for repeat and indel structure. *Genome biology*. 2013; 14:R97. [PubMed: 24025428]
10. Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM, Sherlock G. GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*. 2004; 20:3710–3715. [PubMed: 15297299]
11. Rees CA, Demeter J, Matese JC, Botstein D, Sherlock G. GeneXplorer: an interactive web application for microarray data visualization and analysis. *BMC bioinformatics*. 2004; 5:141. [PubMed: 15458579]
12. Maguire SL, OhEigeartaigh SS, Byrne KP, Schroder MS, O’Gaora P, Wolfe KH, Butler G. Comparative genome analysis and gene finding in *Candida* species using CGOB. *Molecular biology and evolution*. 2013; 30:1281–1291. [PubMed: 23486613]
13. Ostlund G, Schmitt T, Forslund K, Kostler T, Messina DN, Roopra S, Frings O, Sonnhammer EL. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic acids research*. 2010; 38:D196–203. [PubMed: 19892828]

Figure 1a

GO Annotations

Molecular Function

Manually curated

Computational

Biological Process

High-throughput

Computational

Cellular Component

Computational

Pathways

Mutant Phenotype

Classical genetics

heterozygous null

null

unspecified

Large-scale survey

heterozygous null

repressible

● C-8 sterol isomerase activity (IMP, IGI with *S. cerevisiae*: [ERG2](#))

● C-8 sterol isomerase activity (IEA with *S. cerevisiae*: [ERG2](#))

● cellular response to drug (IMP)

● ascospore formation (IEA with *S. pombe*: [erg2](#))

● ergosterol biosynthetic process (IEA with *S. cerevisiae*: [ERG2](#), *S. pombe*: [erg2](#))

● endoplasmic reticulum (IEA with *S. pombe*: [erg2](#))

● ergosterol biosynthesis

● resistance to clotrimazole: decreased

● resistance to amphotericin B: increased

● sterol accumulation: abnormal

● resistance to 1,10-phenanthroline: decreased

● resistance to 4-nitroquinoline N-oxide: decreased

● resistance to fluconazole: decreased

● resistance to itraconazole: decreased

● resistance to ketoconazole: decreased

● resistance to terbinafine: decreased

● resistance to nystatin: increased

● viable

● resistance to dyclonine: decreased

● viable

● virulence: decreased

View all [ERG2](#) GO evidence and references

View all [ERG2](#) Phenotype details and references

Figure 1b

Sequence Information

Last Update

Subfeature Details

Coordinates: 2014-06-24 | Sequence: 2014-06-24

	Relative Coordinates	Chromosomal Coordinates	Most Recent Update	
			Coordinates	Sequence
CDS	1 to 654	155,285 to 154,632	2014-06-24	2014-06-24

Retrieve Sequences

-- C. albicans SC5314 Assembly 22 --

View

Sequence Analysis Tools

-- C. albicans SC5314 Assembly 22 --

View

Maps & Displays

Flanking Features Table

View

Allele Location

Allele C1_00800C_B

Last Update

Subfeature Details

Coordinates: 2014-06-24 | Sequence: 2014-06-24

	Relative Coordinates	Chromosomal Coordinates	Most Recent Update	
			Coordinates	Sequence
CDS	1 to 654	155,291 to 154,638	2014-06-24	2014-06-24

External Links

Broad | CGOB | Entrez Gene (1, 2) | Fungal Orthologs | GenBank (1, 2, 3, 4, 5, 6) | Non-periodic Expression | Pedant | PhylomeDB | TrEMBL

Primary CGDID

CAL0000177093

Ca22chr1A_C_albicans_SC5314:155285 to 154632 | GBrowse

Note: this feature is encoded on the Crick strand.

Ca22chr1B_C_albicans_SC5314:155291 to 154638 | GBrowse

Note: this feature is encoded on the Crick strand.

Figure 1c

Figure 1. Locus Summary page for *ERG2*. (A) Basic Information section; red arrows point to the Banner and search box present on most CGD pages. (B) GO Annotations and Mutant Phenotype sections. (C) Sequence Information section; red arrow indicates sequence retrieval and analysis tools.

Methods Mol Biol. Author manuscript; available in PMC 2016 June 07.

Browse the archive directory here

2014 | 2013 | 2012 | 2011 | 2010 | 2009 | 2008 | 2007 | 2006 | 2005 | 2004 | 2003 | 2002 | 2001

Note: the strains used in some of the experiments listed below may have chromosomal abnormalities such as aneuploidy, which introduces chromosome-specific bias into microarray and other results. For more details, see Arbour et al. (2009).

2014:

- Delgado-Silva Y, et al. (2014) Participation of Candida albicans Transcription Factor RLM1 in Cell Wall Biogenesis and Virulence. *PLoS One* 9(1):e86270

[COPURATED](#) [PubMed](#) [Access Full Text](#) [Access Full Text](#) [Access Full Text](#) [Access Full Text](#) [Web Supplement](#) [Download Datasets](#)

- Khamooshi K, et al. (2014) The Rbf1, Hfl1 and Dbp4 of Candida albicans regulate common as well as transcription factor-specific mitochondrial and other cell activities. *BMC Genomics* 15(1):56

[COPURATED](#) [PubMed](#) [Access Full Text](#) [Web Supplement](#) [Download Datasets](#)

- Sellam A, et al. (2014) Modeling the Transcriptional Regulatory Network That Controls the Early Hypoxic Response in Candida albicans. *Eukaryot Cell* 13(5):675-90

[COPURATED](#) [PubMed](#) [Access Full Text](#) [Web Supplement](#) [Download Datasets](#)

Figure 2a

Figure 2b

Figure 2.

GeneXplorer page. (A) Example entries in the Dataset Archive page; red arrow points to icons that lead to the GeneXplorer browser or pcl file download. (B) GeneXplorer page: red arrows indicate the zoom buttons, the gene search box, and the individual expression profile to click on to see genes similarly expressed to *SAP4*.