Original article

The YeastGenome app: the *Saccharomyces* Genome Database at your fingertips

Edith D. Wong, Kalpana Karra, Benjamin C. Hitz, Eurie L. Hong and J. Michael Cherry*

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

*Corresponding author: Tel: +1 650 723 7541; Email: cherry@stanford.edu

The *Saccharomyces* Genome Database (SGD) is a scientific database that provides researchers with high-quality curated data about the genes and gene products of *Saccharomyces cerevisiae*. To provide instant and easy access to this information on mobile devices, we have developed YeastGenome, a native application for the Apple iPhone and iPad. YeastGenome can be used to quickly find basic information about *S. cerevisiae* genes and chromosomal features regardless of internet connectivity. With or without network access, you can view basic information and Gene Ontology annotations about a gene of interest by searching gene names and gene descriptions or by browsing the database within the app to find the gene of interest. With internet access, the app provides more detailed information about the gene, including mutant phenotypes, references and protein and genetic interactions, as well as provides hyperlinks to retrieve detailed information by showing SGD pages and views of the genome browser. SGD provides online help describing basic ways to navigate the mobile version of SGD, highlights key features and answers frequently asked questions related to the app. The app is available from iTunes (http://itunes.com/apps/yeastgenome). The YeastGenome app is provided freely as a service to our community, as part of SGD's mission to provide free and open access to all its data and annotations.

Introduction

The *Saccharomyces* Genome Database [SGD, (1)] gathers, organizes and provides biological information about the budding yeast *Saccharomyces cerevisiae* in an integrated format via the internet (http://www.yeastgenome.org). These data include basic gene information such as gene names, descriptions, genomic sequence and location, as well as manually curated functional information such as protein–protein interaction from BioGRID (2), mutant phenotype (3), Gene Ontology annotations (4) and protein information (5). With the ever-increasing amount of biological information in SGD, it can be difficult for researchers to quickly retrieve specific data of interest, especially when using a mobile device. The increasing popularity of mobile devices has presented us with the opportunity to facilitate the retrieval of certain data types, while additionally reaching more users. SGD created the YeastGenome app for Apple mobile devices to provide a portion of our vast database in a portable, easily navigable format for users to quickly view basic gene information. A native application for the Apple platform was chosen because of the wide use of Apple products among our users, and to allow access to information when users do not have an internet connection, or if the connection is slow. This application was developed for the iPhone, iPod Touch and iPad running iOS 4.2 or newer.

Data available from the YeastGenome app

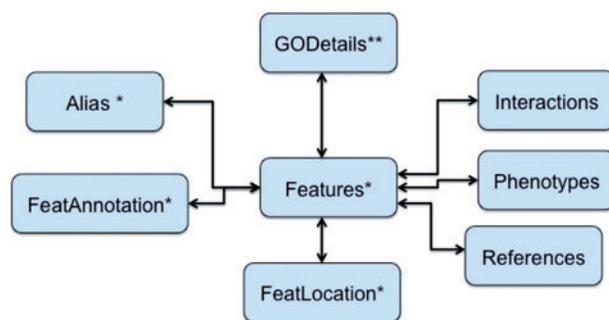### Basic feature data are stored in the local database

YeastGenome contains a small local database, initially populated when installed or updated. This database is not backed up to Apple's iCloud, as it can be easily repopulated

with the SQLite file that is packaged with the app. The SQLite file can be quickly created on a server with data retrieved directly from SGD's primary database with the use of a Perl script. The original database included with version 1.7 of the application contained only basic information about the chromosomal feature, such as feature name, gene name, feature type, feature description, genomic location and alias names (Figure 1). The local database in version 1.8 was expanded to contain Gene Ontology information. Updates of the data in the database will be released regularly, with or without software updates, through the iTunes App Store.

SGD's primary database stores data for >48 000 features, making it too large for a native mobile device application. To keep the local app database small and portable, we populated it with a subset of feature types that would be most used (Table 1). These include open reading frames (ORFs), and other features mapped to chromosomes, such as autonomously replicating sequences (ARS), transposable elements, snoRNAs, rRNAs, tRNAs, centromeres, telomeric regions, pseudogenes as well as features not found in the genomic reference sequence from strain S288c. We have restricted the local app database to storing information about only 8704 features, keeping it convenient, easily navigable and fast.

YeastMine web services are used to retrieve non-locally stored data

In addition to basic feature and Gene Ontology information, reference, genetic and physical interaction and phenotype data are available from the YeastGenome app. These data are not initially included in the app's database, but may be retrieved when internet access is available. The development of YeastMine [http://yeastmine.yeastgenome.org (6)], a *Saccharomyces*-specific installation of InterMine [http://intermine.org (7)] that was developed and is maintained at SGD in collaboration with the Intermine project, has allowed us to keep the YeastGenome app compact, with the ability to quickly access curated data from SGD's database. The YeastMine database is a mirror of the primary SGD database and is synchronized weekly. When access to this information is requested, the YeastGenome app uses the web services feature of YeastMine to quickly and seamlessly retrieve the data. A web service query

Table 1. Feature types that are available in SGD's primary and YeastGenome app databases

Feature type	SGD database	YeastGenome app database
Autonomously replicating sequence (ARS)	X	X
ARS consensus sequence	X	
CDE	X	
CDS	X	
Open reading frame (ORF)	X	X
Transcription factor-binding sites	X	
Binding site	X	
Centromere	X	X
Clone	X	
External transcribed spacer region	X	
5′ UTR	X	
Gene cassette	X	X
Insertion	X	
Internal transcribed spacer region	X	
Intron	X	
Long terminal repeat	X	X
Mating locus	X	X
Multigene locus	X	X
Non-transcribed region	X	
Non-coding exon	X	
Not in systematic sequence of S288c	X	X
Not physically mapped	X	X
Plasmid	X	
Plus 1 translational frameshift	X	
Primer	X	
Pseudogene	X	X
RNAs (ncRNA, rRNA, snRNA, snoRNA and tRNA)	X	X
Repeat region	X	
Retrotransposon	X	X
Telomere	X	X
Telomeric regions (X elements, Y elements, Z regions and repeats)	X	X
Transposable element gene	X	X
uORF	X	

Figure 1. Database schema of YeastGenome app. A schematic drawing depicting the table relationships of the YeastGenome app local database. Single asterisk indicates tables that were initially populated with data in v1.7. Double asterisk indicates an additional table that is initially populated with data in v1.8 release.

returns the resulting data as JSON (JavaScript Object Notation) objects that can be easily accessed to display the requested information. After initial retrieval, these data are stored in the local database for future viewing.

Accessing YeastGenome data

Search or browse data

Information can be discovered about a specific feature by searching or by browsing. Typically, a specific gene will be queried by whole or partial gene name, such as *ACT1* or YFL039C. Additionally, searching by gene name will also return alternative names for features. For example, searching for FIS, will return a list including YIL065C (*FIS1*) and YJL112W (*MDV1*), because YJL112W also has been referred to as *FIS2*, which is defined as an alias of *FIS1*.

Alternatively, the full-text descriptions of features prepared by SGD curators can be searched. For example, searching for 'sporulation' will return a list of 100 features, including *SPO16* and *SPO20*, genes known to be involved in sporulation, as well as YEL023C and YOR268C, genes of unknown function whose expression increases during sporulation.

Finally, browsing is provided and allows the user to explore many feature types curated by SGD. Although 21 different feature types are stored in the local database and are available for searching, we find that users typically are only interested in browsing a subset of these types. We reduced the number of feature types available from the browsing page to make the interface more manageable. For the browsing page, feature types are limited to ORFs and RNA features. ORFs are grouped by their characterization, 'Verified', 'Uncharacterized' or 'Dubious'. The ORF categories are determined by manual curation of available literature in conjugation with other available genomic and high-throughput (HTP) experimental results (8). RNA features split into and displayed as 'ncRNA', 'snoRNA', 'rRNA', 'snRNA' and 'tRNA' groupings.

How data are displayed

Once a genomic feature is selected as a result of a search or from browsing, all basic information is available at a glance (Figure 2). Feature names, descriptions, feature type, SGDID and chromosomal location are displayed on the initial feature page. Additionally, a graphical image previously generated using GBrowse [Generic Genome Browser (10)] is retrieved from the SGD server and displayed.

For additional detailed information about gene function manually curated from the literature, Gene Ontology annotations, interactions, phenotypes and references are available on subsequent pages. When internet access is available, the raw sequence, Locus Summary page providing full information at SGD, feature-specific data at

YeastMine and expression data [http://spell.yeastgenome.org (9)] can also be accessed using the links in the 'View at SGD' section, see Figure 2.

Additional functionality

For features that are frequently examined, after the initial viewing with an internet connection, additional information, such as interaction, phenotype and reference data, is cached locally and will be available at times when internet access is slow or non-existent. These data pages must first be viewed with an internet connection before the information will be stored in the local database. Once the data are added to the database, they can be accessed both with and without an internet connection.

Features that will be repeatedly viewed can be saved to a Favorites list. This list can be created by adding features individually from the main Feature page or by selecting multiple features from a Search or Browse results list. Features can also be removed from the Favorites list as desired.

Information about a feature or multiple features can be shared by e-mail directly from the YeastGenome app. Lists resulting from gene name or description searches can be easily shared, as well as more specific feature information such as the feature page, lists of Gene Ontology annotations, interaction data, phenotype annotations, references and the URL to this information at SGD.

Additions and future enhancements

YeastGenome, a portable 10.5 MB application for Apple mobile devices, allows our user community to quickly access information with or without internet access. As of the writing of this article, the YeastGenome app has been downloaded from the iTunes store nearly 1000 times. Roughly 75% of those users have updated from version 1.7 to 1.8. As expected, YeastGenome is being used to view gene information during seminars. We are pleased to see that the app is also a useful tool for teachers, as a user reported using it to grade student assignments by being able to check answers without having to carry a laptop around. We will regularly release new versions, including database updates and software enhancements. YeastGenome v1.7 was initially released in May 2012. In August 2012, v1.8 was released with both database and software updates. The local database was updated to SGD main database content as of August 2012, including the addition of Gene Ontology annotations. Links to appropriate pages at SGD are included in the email messages sent from the app. This allows easy sharing of information from the application. Also, the favorites list is now maintained when updating the application. Enhancements, such as
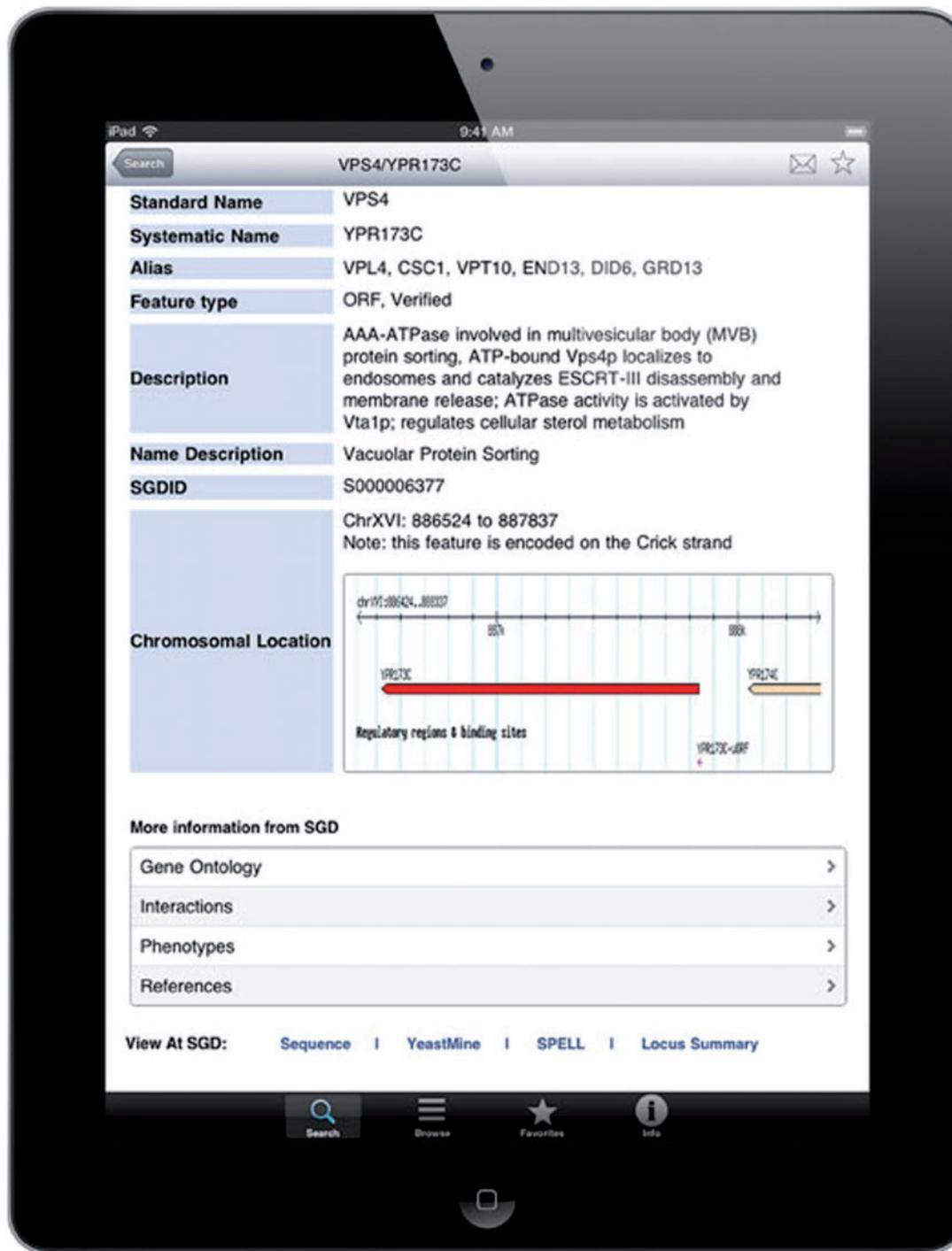
Figure 2. App presentation for budding yeast gene VPS4/YPR172C. A list of gene name aliases, free text description, SGDID and chromosomal coordinates with Gbrowse image for that region of chromosome XVI is displayed. Also provided are links to other locally stored data, such as Gene Ontology, or data retrieved from YeastMine (6), such as interactions, phenotypes and references for all these annotations. There are also hyperlinks to the SGD resource for DNA sequence in FASTA format, YeastMine for advanced queries, RNA expression data via the SPELL resource (9) and the SGD Locus Summary page for all the details about this gene.

multiple gene name searches and local storage of GBrowse images, are being developed for future releases of YeastGenome. New feature requests are welcomed at the SGD helpdesk (sgd-helpdesk@lists.stanford.edu). SGD continues to provide novel tools, resources and enhanced curation of existing and new datatypes that meet our community's evolving needs.

Source code and support online

The source code is freely available at https://github.com/edwong57/YeastGenome-app. A web page is provided that provides basic app information and how to obtain help, 'http://www.yeastgenome.org/yeastgenome-app-information'.

Acknowledgements

Funding

Conflict of interest. None declared.

References

1. Cherry,J.M., Hong,E.L., Amundsen,C. *et al*. (2012) *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.*, **40**, D700–D705.

2. Stark,C., Breitkreutz,B.J., Chatr-Aryamontri,A. *et al*. (2010) The BioGRID interaction database: 2011 update. *Nucleic Acids Res.*, **39**, D698–D704.

3. Costanzo,M.C., Skrzypek,M.S., Nash,R. *et al*.; the Saccharomyces Genome Database Project. (2009) New mutant phenotype data curation system in the *Saccharomyces* Genome Database. *Database (Oxford)*, **2009**, bap001.

4. Gene Ontology Consortium. (2011) The Gene Ontology: enhancements for 2011. *Nucleic Acids Res.*, **40**, D559–D564.

5. Nash,R., Weng,S., Hitz,B. *et al*. (2007) Expanded protein information at SGD: new pages and proteome browser. *Nucleic Acids Res.*, **35**, D468–D471.

6. Balakrishnan,R., Park,J., Karra,K. *et al*. (2012) YeastMine–an integrated data warehouse for *Saccharomyces cerevisiae* data as a multipurpose tool-kit. *Database (Oxford)*, **2012**, bar062.

7. Smith,R.N., Aleksic,J., Butano,D. *et al*. (2012) InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data. *Bioinformatics*, **28**, 3163–3165.

8. Fisk,D.G., Ball,C.A., Dolinski,K. *et al*. (2006) *Saccharomyces cerevisiae* S288C genome annotation: a working hypothesis. *Yeast*, **23**, 857–867.

9. Hibbs,M.A., Hess,D.C., Myers,C.L. *et al*. (2007) Exploring the functional landscape of gene expression: directed search of large microarray compendia. *Bioinformatics*, **23**, 2692–2699.

10. Stein,L.D., Mungall,C., Shu,S. *et al*. (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.