

# New data and collaborations at the *Saccharomyces* Genome Database: updated reference genome, alleles, and the Alliance of Genome Resources

Stacia R. Engel,\* Edith D. Wong, Robert S. Nash, Suzi Aleksander, Micheal Alexander, Eric Douglass, Kalpana Karra, Stuart R. Miyasato, Matt Simison, Marek S. Skrzypek, Shuai Weng, and J. Michael Cherry\*

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

\*Corresponding authors: Emails: [stacia@stanford.edu](mailto:stacia@stanford.edu) (S.R.E.); [cherry@stanford.edu](mailto:cherry@stanford.edu) (J.M.C.)

## Abstract

*Saccharomyces cerevisiae* is used to provide fundamental understanding of eukaryotic genetics, gene product function, and cellular biological processes. *Saccharomyces* Genome Database (SGD) has been supporting the yeast research community since 1993, serving as its de facto hub. Over the years, SGD has maintained the genetic nomenclature, chromosome maps, and functional annotation, and developed various tools and methods for analysis and curation of a variety of emerging data types. More recently, SGD and six other model organism focused knowledgebases have come together to create the Alliance of Genome Resources to develop sustainable genome information resources that promote and support the use of various model organisms to understand the genetic and genomic bases of human biology and disease. Here we describe recent activities at SGD, including the latest reference genome annotation update, the development of a curation system for mutant alleles, and new pages addressing homology across model organisms as well as the use of yeast to study human disease.

**Keywords:** *Saccharomyces cerevisiae*; knowledgebase; reference genome; *Saccharomyces* Genome Database; Alliance of Genome Resources

## Introduction

*Saccharomyces cerevisiae* is used to provide fundamental understanding of eukaryotic genetics, gene product function, and cellular biological processes. The published scientific literature from the yeast research community is integrated into the biomedical knowledgebase *Saccharomyces* Genome Database (SGD; [www.yeastgenome.org](http://www.yeastgenome.org)). Biocurators with expertise in genetics, cell biology, and molecular biology have collected information from more than 100,000 published papers and combined the results from diverse experiments into a comprehensive resource for researchers, educators, and students.

The *S. cerevisiae* nomenclature has been maintained by SGD since 1993 (Cherry *et al.* 1997). Soon thereafter, SGD began providing the genetic and physical maps for the 16 yeast nuclear chromosomes (Cherry *et al.* 1997), the catalog of all known yeast proteins (Chervitz *et al.* 1999; Weng *et al.* 2003), biological process and molecular function annotations using the Gene Ontology (GO) (Dwight *et al.* 2002), as well as gene expression data and tools for analysis (Ball *et al.* 2001). SGD has maintained the reference genome from strain S288C, which was the first completely sequenced eukaryotic genome (Goffeau *et al.* 1996), and its annotation since 1998 (Cherry *et al.* 1998), along with sequence analysis

and retrieval tools for studying that reference genome (Balakrishnan *et al.* 2004; Christie *et al.* 2004; Hirschman *et al.* 2006), and later broadened the reference panel by adding genomes of 11 additional highly studied strains to more fully support the work of the yeast research community (Engel and Cherry 2013). In parallel to these activities, SGD also developed principles and practices for the extraction and curation of various types of biological data (Ball *et al.* 2000; Dwight *et al.* 2004; Hong *et al.* 2008; Engel *et al.* 2010; Costanzo *et al.* 2011; Park *et al.* 2012; Balakrishnan *et al.* 2013; Skrzypek and Nash 2015). SGD then turned to further enhancing existing tools and curation practices, including the development of an automated pipeline for pan-genome analysis (Song *et al.* 2015), release of the variant viewer for analysis of the reference genome panel (Sheppard *et al.* 2016), curation of the complete set of yeast metabolic pathways (Cherry 2015), and updated curation methods and data models for the capture of post-translational modifications (Hellerstedt *et al.* 2017), regulatory relationships (Engel *et al.* 2018), macromolecular complexes (Wong *et al.* 2019), and the yeast transcriptome (Ng *et al.* 2019).

As the de facto hub of the yeast research community, SGD also engages in a wide range of outreach and communication

Received: September 15, 2021. Accepted: November 11, 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

activities to disseminate published results, promote collaboration, facilitate scientific discovery, and inform users about new tools, data, or other database developments (MacPherson et al. 2017). These activities include participating in conferences and hosting workshops, direct contact with authors of yeast research papers, the posting of online help resources, and involvement in social media, including the production of video tutorials and webinars, all hosted on SGD's YouTube channel (<https://www.youtube.com/SaccharomycesGenomeDatabase>). To increase readership and reach a broad audience, content posted on one outreach platform is often publicized or announced on other outreach platforms. We also collaborate with the Genetics Society of America to annotate online journal articles published in GENETICS and G3: Genes | Genomes | Genetics to associate yeast genes listed in these articles to their respective gene pages at SGD.

More recently, SGD and six other model organism-focused knowledgebases—Mouse Genome Database (MGD; <http://www.informatics.jax.org>, Bult et al. 2019), Rat Genome Database (RGD; <https://rgd.mcw.edu>, Laulederkind et al. 2018), Zebrafish Information Network (ZFIN; <https://zfin.org>, Ruzicka et al. 2019), WormBase (<https://wormbase.org>, Lee et al. 2018), FlyBase (<https://flybase.org>, Thurmond et al. 2019), and the GO Consortium (<http://www.geneontology.org>, The Gene Ontology Consortium 2019)—have created the Alliance of Genome Resources (“the Alliance”; <https://www.alliancegenome.org>; Alliance of Genome Resources Consortium 2019; Alliance of Genome Resources 2022). Together, we are working to create an online resource that integrates, develops, and provides harmonized data for all member projects (Alliance of Genome Resources Consortium 2020). The aim is to more broadly facilitate the use of model organisms to understand the genetic bases of human biology and disease. These efforts build on the well-established collaborations between these groups, who have long worked together to enhance data consolidation, dissemination, visualization, and the application of shared standards (i.e., “data harmonization”). Working within the Alliance has influenced SGD's curation of mutant alleles and human diseases, as well as our use of homology throughout the SGD website to highlight the greater biological context of key findings in yeast research.

## Genome version R64.3.1

### Annotation updates and additions

The *S. cerevisiae* strain S288C reference genome annotation was updated for the first time since 2014 (Table 1). The new genome annotation is release R64.3.1, dated April 21, 2021. Resequencing of S288C in 2014 indicated that genomic sequence variation was less than expected between individual laboratory copies of this strain, illustrating that the underlying genome sequence is stable and complete. As such, while SGD updated the annotation of the genomic sequence, the fundamental sequence itself remains unchanged (Cherry et al. 1998; Engel et al. 2014). All new and updated annotations are sourced from the primary literature. The R64.3.1 update included the addition of seven open reading frames (ORFs), five noncoding RNAs (ncRNAs), two upstream ORFs (uORFs), and one long terminal repeat (LTR). Three ORFs had their translation starts shifted to a different methionine, and one ORF received a newly annotated intron and had its translation stop shifted. We also changed the Sequence Ontology (SO; <http://www.sequenceontology.org/>; Eilbeck et al. 2005) term used to describe the nontranscribed spacers (NTS1-1, NTS1-2, NTS2-1, and NTS2-2) in the ribosomal DNA (rDNA) array.

## New ncRNA systematic nomenclature

Included in the R64.3.1 update is the establishment of a new systematic nomenclature for yeast ncRNAs (430 genes; Supplementary Table S1). For many years, a widely adopted systematic nomenclature has existed for yeast protein-coding genes, or ORFs, as many yeast researchers call them (Cherry et al. 1998). With this most recent genome version, we have established a systematic nomenclature for yeast ncRNAs that is similar to, but distinct from, that used for ORFs. All annotated *S. cerevisiae* ncRNAs are designated by a symbol consisting of four uppercase letters, a four-digit number, and another letter, as follows: Y for “Yeast,” NC for “noncoding,” A-Q for the chromosome on which the ncRNA gene resides (where “A” is chromosome I, “B” is chromosome II, etc., up to “P” for chromosome XVI, and lastly “Q” for the mitochondrial chromosome), a four-digit number corresponding to the sequential order of the ncRNA gene on the chromosome (starting from the left telomere and counting toward the right telomere), and W or C indicating whether the ncRNA gene is encoded on the “Watson” or “Crick” strand (where “Watson” runs 5' to 3' from left telomere to right telomere, and “Crick” runs 3' to 5'). For example, YNCP002W is the second ncRNA gene from the left end of chromosome XVI and is encoded on the Watson strand. Going forward, when evidence is published pointing to new ncRNA genes, they will be added to the annotation using the next sequential number available for the specific chromosome on which the ncRNA gene resides. In cases in which more than one ncRNA gene is added to any particular chromosome during the same annotation update (i.e., same genome revision), they will be named using the next sequential number starting with the leftmost ncRNA gene and proceeding to the right of the chromosome.

## Nomenclature updates—“legacy” gene names

SGD has long been the keeper of the official *S. cerevisiae* gene nomenclature (Cherry 1998). Robert Mortimer handed over this responsibility to SGD in 1993 after maintaining the yeast genetic map and gene nomenclature for 30 years (Hawthorne and Mortimer 1960; Mortimer and Schild 1980). The accepted format for gene names in *S. cerevisiae* comprises three uppercase letters followed by a number. The letters typically signify a phrase (referred to as the “Name Description” in SGD) that provides information about a function, mutant phenotype, or process related to that gene, for example, “ADE” for “ADENine biosynthesis” or “CDC” for “Cell Division Cycle.” Gene names for many types of chromosomal features follow this basic format regardless of the type of feature named, whether an ORF, a tRNA, another type of noncoding RNA, an ARS, or a genetic locus. Some *S. cerevisiae* gene names that pre-date the current nomenclature standards do not conform to this format, such as MRLP38, RPL1A, and OM45. A few historical gene names predate both the nomenclature standards and the database, and are less computer-friendly than more recent gene names, due to the presence of punctuation. SGD recently updated these gene names to be consistent with current standards and to be more software-friendly by removing punctuation (Table 2).

Although nonstandard historical names are maintained in SGD, any new names for yeast genes must conform to the standard format. The SGD Gene Registry ([https://www.yeastgenome.org/reserved\\_name/new](https://www.yeastgenome.org/reserved_name/new)) is the agreed-upon system used by the *S. cerevisiae* community and SGD to reserve standard names for newly characterized genes. We have developed a set of Gene Naming guidelines and procedures in order to assist researchers

**Table 1** The *Saccharomyces cerevisiae* strain S288C reference genome annotation was updated

Chromosome	Feature	Description of change	Reference
II	YNCB0008W aka GAL10-ncRNA	New ncRNA antisense to GAL10: coordinates 276,805–280,645	Houseley et al. (2008); Pinskaya et al. (2009); Geisler et al. (2012)
II	YNCB0014W aka TBRT/XUT_2F-154	New ncRNA antisense to TAT1: coordinates 376,610–378,633	Awasthi et al. (2020)
III	RE/RE301	New recombination enhancer: coordinates 29,108–29,809	Wu and Haber (1996)
V	YELWdelta27	New Ty1 LTR: coordinates 449, 274–449,626	Nene et al. (2018)
V	HPA3/YEL066W	Moved translation start to Met19: old coordinates 26,667–27,206; new coordinates 26,721–27,206	Sampath et al. (2013)
VII	OTO1/YGR227C-A	New ORF: coordinates 949052–949225	Makanae et al. (2015)
VII	ROK1/YGL171W	Two new uORFs: coordinates uORF1 182,286–182,407; coordinates uORF2 182,291–182,329	Jeon and Kim (2010)
VIII	YHR052C-B	New ORF: coordinates 212,519–212,692	He et al. (2018)
VIII	YHR054C-B	New ORF: coordinates 214,517–214,690	He et al. (2018)
VIII	SUT169/YNCH0011W	New ncRNA: coordinates 378, 254–379,237	Xu et al. (2009); Geisler et al. (2012); Huber et al. (2016); Bunina et al. (2017)
X	YJR012C	Moved start to Met76: old coordinates 459,795–460,418 Crick; new coordinates 459,795–460,193 Crick	Sadhu et al. (2018)
X	YJR107C-A	New ORF: coordinates 628,457–628,693	Yagoub et al. (2015); He et al. (2018)
XI	YKL104W-A	New ORF: coordinates 245,032–245,286	He et al. (2018)
XII	YLR379W-A	New ORF: coordinates 877,444–877,716	Internal reanalysis of results from Song et al. (2015) to find and annotate missing S288C ORFs
XII	NTS1-2, NTS2-1, NTS2-2	Change feature_type/SO_term from SO:0001637 rRNA_gene to SO:0000183 non_transcribed_region	
XIII	LDO45/YMR147W	Shift stop to be same as LDO16/YMR148W, add intron: old coordinates 559,199–559,870; new coordinates 559,199–559,780, 560,156–560,812	Eisenberg-Bord et al. (2018)
XIII	YMR008C-A	New ORF: coordinates 283,081–283,548	Internal reanalysis of results from Song et al. (2015) to find and annotate missing S288C ORFs
XIII	YNCM0001W aka PHO84 lncRNA	New ncRNA: coordinates 23,564–26,578	Camblong et al. (2007)
XIV	LTO1/YNL260C	Move start to Met37: old coordinates 156,859–157,455 Crick; new coordinates 156,859–157,347 Crick	Paul et al. (2015)
XVI	YNCP0002W aka GAL4 lncRNA	New ncRNA: coordinates 79,562–82,648	Geisler et al. (2012)

The new genome annotation is release R64.3.1, dated April 21, 2021.

**Table 2** “Legacy” gene names which predate the database have been updated to be more software-friendly by removing unnecessary punctuation

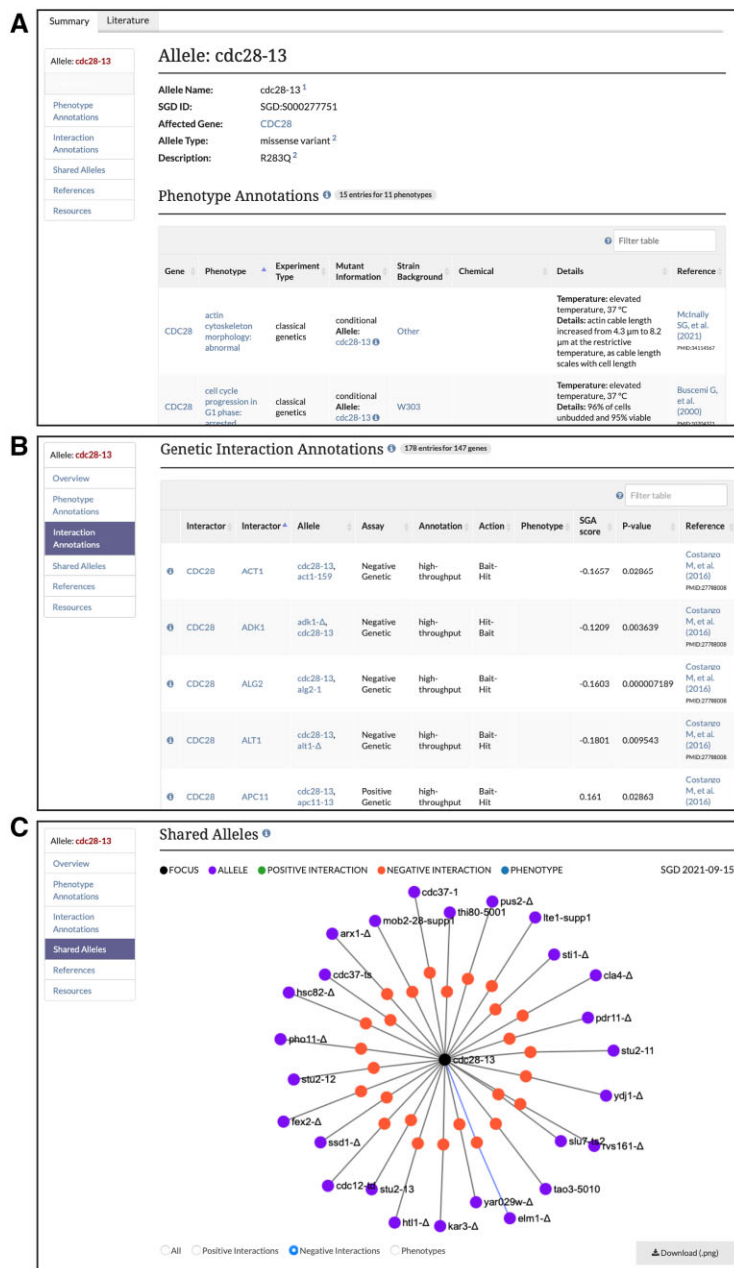
ORF	Old gene name	New gene name
YGL234W	ADE5,7	ADE57
YER069W	ARG5,6	ARG56
YBR208C	DUR1,2	DUR12
YIL154C	IMP2'	IMP21

in gene naming. We urge researchers to contact SGD prior to the publication of a new gene name (even if they have previously reserved it) to ensure that the gene name they wish to use is still appropriate. To date, 1363 currently annotated *S. cerevisiae* genes still have no standard gene name and are identified solely via their systematic names (ex. YFL019C). All *S. cerevisiae* ORFs are categorized into one of three groups: “verified” ORFs are those for which there is clear experimental evidence for the presence of

expression of a protein-coding gene; “uncharacterized” ORFs are likely, but not yet established, to encode a protein; and “dubious” ORFs are unlikely to encode a protein (Fisk et al. 2006). Of the 1363 currently annotated *S. cerevisiae* ORFs without a standard gene name, 131 are verified, 571 are uncharacterized, and the remaining 661 are classified as dubious.

## Alleles

To serve the needs of the yeast community more completely, and influenced by our active participation in the Alliance, SGD has recently revamped the way we curate mutant alleles of genes and has released new allele web pages. For over a decade, SGD has been recording mutant allele information as descriptive properties of phenotype annotations (Costanzo et al. 2009). We have recently elevated alleles in the database from phenotype properties to primary data objects, i.e., standalone entities to which other



**Figure 1** SGD allele pages include (A) Overview with name, affected gene, type of allele, and a description of sequence change and/or domain mutated; phenotype annotations; (B) genetic interaction annotations; and (C) shared alleles network diagram depicting shared phenotypes and interactions with other alleles.

annotations, such as phenotypes or genetic interactions, can be attached, and that have their own detailed web pages.

Previously, as part of phenotype curation, only allele names, mutant types, and free-text descriptions were collected. We now have expanded the information gathered and displayed for alleles to include affected gene, alias names, description in free text including relevant nucleotide and/or amino acid changes and correspondence to human alleles when known, and allele type. Allele type, which is captured using the *structural\_variant* (SO:0001537) branch of the Sequence Ontology, describes the change to the relevant sequence feature with terms like *missense\_variant* (SO:0001583), *stop\_gained* (SO:0001587), and *frameshift\_variant* (SO:0001589). To further describe the functional effect of the structural change within the context of phenotype annotations, SGD uses the *functional\_effect\_variant* (SO:0001536) branch

of the SO, which contains terms like *null\_mutation* (SO:0002055), *loss\_of\_function\_variant* (SO:0002054), and *dominant\_negative\_variant* (SO:0002052).

## New allele pages

The new allele pages at SGD contain different types of information divided into several sections: Overview, Phenotypes, Genetic Interactions, Shared Alleles Network, External Resources, and Relevant Literature (Figure 1). The Allele Overview provides general information about the allele, including its name, the affected gene, the type of allele (e.g., *missense\_variant*, as described above), and a description of sequence change and/or domain mutated. The phenotype annotations for an allele include an observable feature (e.g., “cell shape”), a qualifier (e.g., “abnormal”), a mutant type (e.g., null), strain background, and a reference. In addition,

annotations are classified as classical genetics or high-throughput (e.g., large-scale survey, systematic mutation set). The Genetic Interactions for an allele are defined as experimentally observed genetic interactions between that allele and another of a different gene. All interactions listed in SGD are curated by BioGRID (Oughtred et al. 2021). The Shared Alleles Network displays positive genetic interactions, negative genetic interactions, and phenotypes that are shared between the given allele and other alleles (Figure 1). The resources section provides links to allele-related information, such as mutant strain repositories, external phenotype and interaction databases, and information about the yeast phenotype ontology. In addition, all literature associated with an allele can be found on its literature page (Figure 1).

## Updates to interactions data/pages

To further accommodate the enhanced focus on alleles and for improved clarity, SGD has made recent updates to the way we present Interaction data. Previously, genetic and physical interaction annotations were combined in one table. These annotations are now displayed in separate annotation tables on Interactions pages (ex. <https://www.yeastgenome.org/locus/S000003424/interaction>), Reference pages (ex. <https://www.yeastgenome.org/reference/S000305076>), and in YeastMine (<https://yeastmine.yeastgenome.org/yeastmine/templates.do>) to allow for the listing of allele designations, synthetic genetic analysis (SGA) scores, and *P*-values from the global genetic interactions paper by Costanzo et al. (2016).

## Searching for alleles

All alleles for a specific gene can be accessed via the Alleles section on the Locus Summary Page or downloaded from YeastMine using the Genes -> Alleles template ([https://yeastmine.yeastgenome.org/yeastmine/template.do?name=Gene\\_Alleles&scope=all](https://yeastmine.yeastgenome.org/yeastmine/template.do?name=Gene_Alleles&scope=all)). The alleles have also been added to SGD's Elasticsearch, with 'facets' for publications, allele types, affected genes, and phenotypes, which allow for the browsing, partitioning, and viewing of the approximately 14,000 yeast alleles annotated so far in SGD.

## Homology and the alliance of genome resources

SGD and six other model organism focused knowledgebases—MGD, RGD, ZFIN, WormBase, FlyBase, and the GO Consortium—have recently created a new knowledgebase for model organisms. These major community-driven model organism projects have taken up the mission of harmonizing common data types and creating an integrated web resource. The Alliance of Genome Resources has brought together biocurators and software engineers from all the MODs to build a new central resource. Teams have been created to define data models for the major data types provided by the MODs. These teams also define how the information should be presented. Software engineers from the MODs work together to create the cloud computational environment, component tools, and web pages that define such an important resource. Gene products, proteins, ncRNAs, and pseudogenes are connected by their homology, molecular functions, biological processes, cellular component location, anatomical expression, and association with disease. Human gene details are provided by RGD.

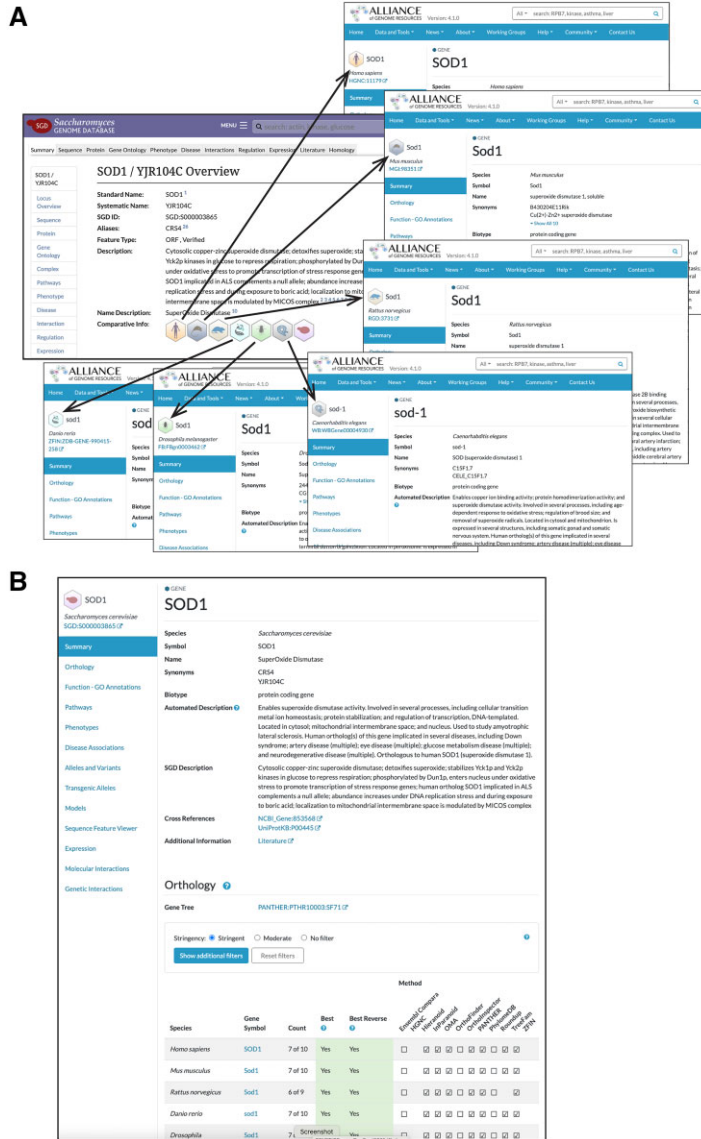
One of the first work products to come out of the Alliance was a consolidated set of orthologs, using data from several different

computational and manually curated sources (Howe et al. 2018). Many aspects of data integration presented at the Alliance require a common set of orthology relationships among genes for the organisms represented, including human. The Alliance provides the results of all methods that have been benchmarked by the Quest for Orthologs Consortium (QfO; <https://questfororthologs.org/>, Linard et al. 2021). The homolog inferences from the different methods have been integrated using the DRSC Integrative Ortholog Prediction Tool (DIOPT; Hu et al. 2011), which integrates a number of existing methods including those used by the Alliance: Ensembl Compara (Vilella et al. 2009), HUGO Gene Nomenclature Committee (HGNC; Povey et al. 2001), Hieranoid (Kaduk and Sonnhammer 2017), InParanoid (O'Brien et al. 2005), the Orthologous Matrix project (OMA; Schneider et al. 2007), OrthoMCL (Li et al. 2003), OrthoFinder (Emms and Kelly 2015), OrthoInspector (Linard et al. 2011), PANTHER (Mi et al. 2019), PhylomeDB (Huerta-Cepas et al. 2008), Roundup (Deluca et al. 2006), TreeFam (Li et al. 2006), and ZFIN (Westerfield et al. 1997). This set of assertions (the "orthology set") is key because it allows the inference of function from one species to another, and is especially helpful because work using a highly studied and experimentally tractable organism such as yeast can be informative regarding the biology of other organisms in which targeted experiments are not possible, such as human (Dolinski and Botstein 2007). Currently included in the Alliance and in the orthology set are five other leading model organisms in addition to *S. cerevisiae* (yeast): *Caenorhabditis elegans* (worm), *Drosophila melanogaster* (fruit fly), *Danio rerio* (zebrafish), *Rattus norvegicus* (rat), and *Mus musculus* (mouse).

SGD takes advantage of the homology data from the Alliance to provide easy access to information about homologous genes in just one click. At SGD, we have used the orthology set to provide links between SGD gene pages and those for orthologous genes at the Alliance (Figure 2). On SGD gene pages, users will find hexagonal icons representing each model organism (human, mouse, rat, zebrafish, fly, worm, and yeast) for which there is homologous gene information at the Alliance. Clicking on the icon immediately directs the user to the gene page for the selected model organism on the Alliance website. The Alliance gene pages present a variety of data types for all of the various organisms, including functional annotations using the GO, phenotypes, disease associations, alleles, variants, sequence features, expression information, and both physical and genetic interactions. Data for individual genes can be downloaded directly from Alliance gene pages, and bulk downloads are available from the Alliance Downloads site (<https://www.alliancegenome.org/downloads>).

## Homology pages

We have also used the Alliance orthology set to establish new Homology pages at SGD for protein-coding genes. The information displayed on the Homology Pages is divided into several sections: Homologs, Functional Complementation, Fungal Homologs, and External Identifiers (Figure 3). The Homologs section lists information about genes in the Alliance orthology set, including species, gene ID, and gene name, with links to the corresponding gene page at the Alliance. The Functional Complementation section presents data about cross-species complementation between yeast and other species, as curated by SGD and including legacy data from the Princeton Protein Orthology Database (P-POD; Heinicke et al. 2007). These curated Functional Complementation data are also displayed on SGD Reference pages. The Fungal Homologs section shows curated



**Figure 2** (A) SGD provides links between SGD gene pages and those for orthologous genes at the Alliance of Genome Resources. Hexagonal icons represent each organism (human, mouse, rat, zebrafish, fly, worm, and yeast) for which there is gene information at the Alliance. (B) Curated information from SGD is presented on yeast gene pages at the Alliance of Genome Resources. The various types of yeast data that can be found at the Alliance can be viewed using the menu on the left side of the gene page, and include orthologs, functional annotations using the GO, cellular pathways, phenotypes, disease associations, alleles, variants, sequence features, expression, and interactions. Data can be downloaded in bulk from <https://www.alliancegenome.org/downloads>.

information including species, gene ID, and database source for orthologs in 24 additional species of fungi, such as those in *Candida*, *Neurospora*, and *Aspergillus*, among others. The External Identifiers section lists identifiers for the protein from various database sources.

**Disease pages**

To promote the use of yeast as a catalyst for biomedical research, SGD utilizes the Disease Ontology (DO; Schriml et al. 2019) to describe human diseases that are associated with yeast homologs. Disease Ontology annotations to yeast genes are now available through SGD’s new Disease pages, each of which corresponds to a Disease Ontology term, such as *amyotrophic lateral sclerosis* (DOID: 332), and lists out all yeast genes annotated to the term by SGD. Yeast genes with one or more human disease associations also have a new Disease tab accessible from the genes’ respective

locus pages (Figure 4). The Disease tab shows all manually curated, high-throughput, and computational disease annotations for the yeast gene. In addition, these pages feature a network diagram that depicts shared disease annotations for other yeast genes and their human homologs.

While the Alliance will provide much of the information researchers will want to explore, there will continue to be data that cannot be harmonized across organisms because of its unique characteristics, such as an aspect of biology only found in one of the species or the nuance of an experimental assay that is unique to one community. In time, these unique bits of information will be provided directly from the Alliance site. However, until then, the main knowledgebases are essential. All of the MODs look forward to the common user interface and ease of discovering common associations between genes in other organisms. This is an important step to allow researchers, educators, and

**A**

Summary Sequence Protein Gene Ontology Phenotype Disease Interactions Regulation Expression Literature **Homology**

**CDC28 / YBR160W Homology**

CDC28 / YBR160W

Homology Overview

Homologs

Functional Complementations

Fungal Homologs

External Identifiers

Resources

**Standard Name:** CDC28<sup>1,2</sup>

**Systematic Name:** YBR160W

**SGD ID:** SGD:S000000364

**Aliases:** CDK1, HSL5, SRM5<sup>63</sup>

**Feature Type:** ORF, Verified

**Description:** Cyclin-dependent kinase (CDK) catalytic subunit; master regulator of mitotic and meiotic cell cycles; alternately associates with G1, S, G2/M phase cyclins, which provide substrate specificity; regulates metabolism, basal transcription, chromosome dynamics, growth and morphogenesis; transcript induction in osmostress involves antisense RNA; human homologs CDK1, CDK2, CDK3 can complement yeast conditional *cdc28* mutants; human CDK1, CDK2 can complement yeast *cdc28* null mutant<sup>3 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20</sup>

**Name Description:** Cell Division Cycle<sup>4</sup>

**Comparative Info:**

**B Homologs** 12 entries for 12 homologs

Filter table

Species	Gene ID	Gene name	Source
<i>Caenorhabditis elegans</i>	WB:WBGene00000405	cdk-1	Alliance
<i>Danio rerio</i>	ZFIN:ZDB-GENE-010320-1	cdk1	Alliance
<i>Danio rerio</i>	ZFIN:ZDB-GENE-040426-2741	cdk2	Alliance
<i>Drosophila melanogaster</i>	FB:FBgn0004106	Cdk1	Alliance
<i>Drosophila melanogaster</i>	FB:FBgn0004107	Cdk2	Alliance
<i>Homo sapiens</i>	HGNC:1722	CDK1	Alliance
<i>Homo sapiens</i>	HGNC:1771	CDK2	Alliance
<i>Homo sapiens</i>	HGNC:1772	CDK3	Alliance
<i>Mus musculus</i>	MGI:88351	Cdk1	Alliance
<i>Mus musculus</i>	MGI:104772	Cdk2	Alliance

**C Functional Complementation** 12 entries for 12 complements

Filter table

Species	Gene ID	Strain background	Direction	Details	Source	Reference
<i>Homo sapiens</i>	HGNC:1722	S288C	other complements yeast	Human CDK1 complements the thermosensitivity of the yeast <i>cdc28-4</i> mutant as well as the inviability of the <i>cdc28</i> null mutant.	SGD	Elledge SJ and Spottswood MR (1991) PMID:1714386
<i>Homo sapiens</i>	HGNC:1722	S288C	other complements yeast	Human CDK1 complements the thermosensitivity of the yeast <i>cdc28-13</i> mutant as well as rescuing the inviability of the null mutant.	SGD	Ninomiya-Tsuji J, et al. (1991) PMID:1717994
<i>Homo sapiens</i>	HGNC:1722	S288C	other complements yeast	Human CDK1 complements the inviability of the <i>S. cerevisiae cdc28</i> null mutant.	SGD	Mayi T, et al. (2015) PMID:25541464
<i>Homo sapiens</i>	HGNC:1722	S288C	other complements yeast	Human CDK1 complements the inviability of the <i>S. cerevisiae cdc28</i> null mutant.	SGD	Hamza A, et al. (2015) PMID:26354769

**Figure 3** SGD homology pages (A) include an overview of general information about the yeast gene, with links to homologous gene pages at the Alliance; (B) information about known homologs including species corresponding Gene ID, and name of the homolog; and (C) data about cross-species functional complementation between yeast and other species. Also included on the page, but not shown here, are Fungal Homologs, gene identifiers in other databases, and links to external resources.

students to have access to the gold standard of expertly curated information on each of these foundations of biological science.

## Other updates

### Supplementary data and published datasets on reference pages

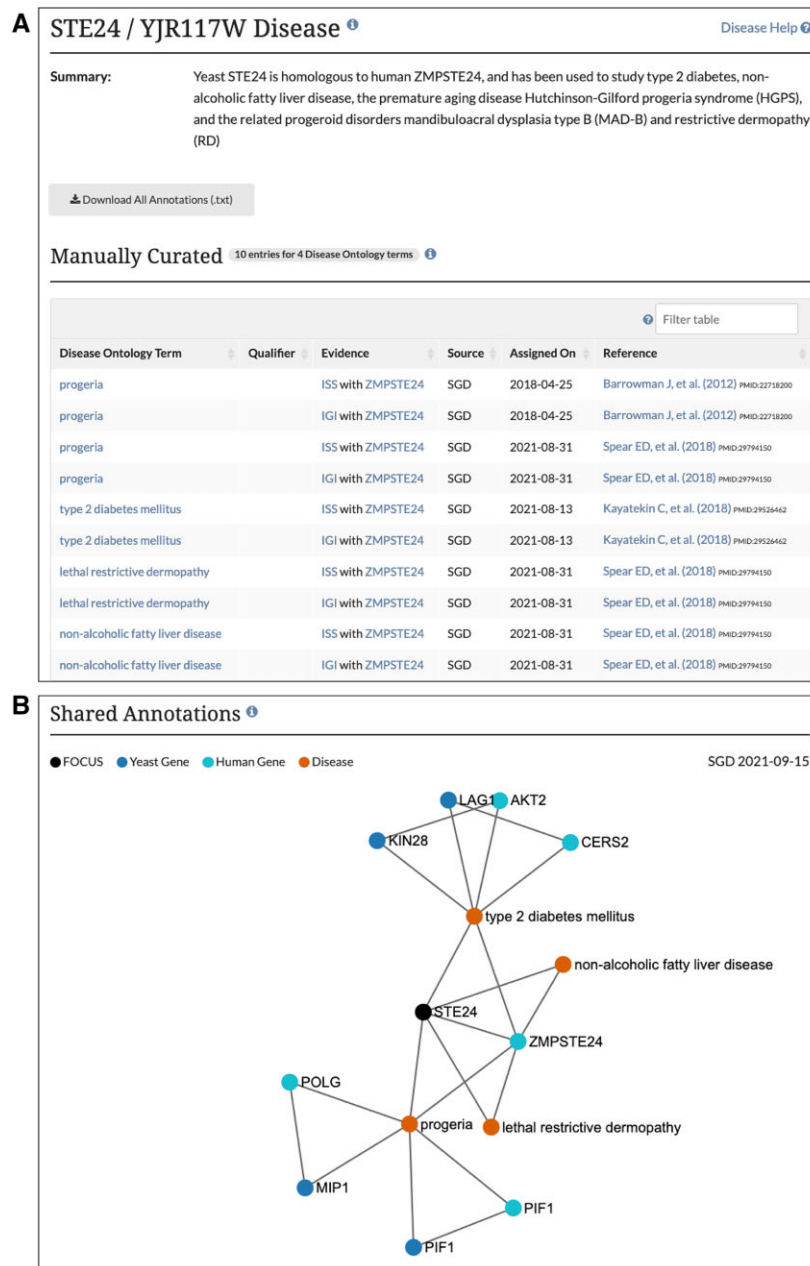
SGD procures and displays Supplementary materials for references stored in our database. We are hosting data from past, present, and future papers on our literature pages. To access these data, simply search SGD with the PubMed ID and then look for the “Downloadable Files.” In addition, published large-scale datasets from gene expression omnibus (GEO) are displayed in the “Published Datasets” section. The dataset title is linked to a dataset-specific page and a controlled vocabulary of terms is used to bin similar datasets into broad categories.

### Submit data form

Authors can submit data and information about their publications, including novel results, datasets (including accession IDs), or other important information, using SGD’s simple “Submit Data” form (<https://www.yeastgenome.org/submitData>).

### Explore button

SGD introduced a new “Explore SGD” button on our homepage, which allows users to peruse SGD data and pages without an initial search query. After selecting the “Explore SGD” button, users are redirected to a search results page where they can browse all of the information in SGD. The tool is designed for both new and veteran users alike, as new users are provided a glimpse into the warehouse of information SGD contains, while seasoned users may discover something new. After clicking on the “Explore SGD” button, use the categories on the left to navigate through the various pages and examine areas of interest. An “Explore” link has



**Figure 4** SGD disease pages use the Disease Ontology to describe human diseases that are associated with yeast homologs, and include (A) disease annotation summary and individual annotations; and (B) shared annotation network that depicts shared disease annotations for other yeast genes and their human homologs.

also been added to the selection of links available in the black bar at the top of every SGD webpage, giving users the ability to access the search results page from anywhere on the SGD website.

### microPublication Biology

SGD has partnered with *microPublication Biology* (<https://www.micropublication.org>) to promote the growing model of rapid publication of research results. *microPublication* further expands this rapid-release model by coupling publication with curation and deposition into community-directed authoritative databases such as SGD and the other Alliance member databases mentioned above. Each publication is peer-reviewed, assigned a digital object identifier (DOI), published online, indexed in PubMed (<https://pubmed.ncbi.nlm.nih.gov>), and then integrated into the relevant databases, speeding information

dissemination and scientific discovery. SGD encourages authors to utilize *microPublication* for new research findings or reagents, experimental results that do not fit into a larger existing or future narrative, negative results, successful replications of recent work, cautionary findings regarding unsuccessful attempts at replication of recent work, “data not shown” from other publications, and data from student projects. All published yeast articles are immediately available in SGD (<https://www.yeastgenome.org/search?category=reference&journal=microPublication.%20Biology>).

### Sequence accessions

NCBI RefSeq accession numbers corresponding to the *Saccharomyces cerevisiae* S288C reference genome version R64.3.1.



GSF_000146045.2	RefSeq Assembly
NC_001133.9	Chromosome I
NC_001134.8	Chromosome II
NC_001135.5	Chromosome III
NC_001136.10	Chromosome IV
NC_001137.3	Chromosome V
NC_001138.5	Chromosome VI
NC_001139.9	Chromosome VII
NC_001140.6	Chromosome VIII
NC_001141.2	Chromosome IX
NC_001142.9	Chromosome X
NC_001143.9	Chromosome XI
NC_001144.5	Chromosome XII
NC_001145.3	Chromosome XIII
NC_001146.8	Chromosome XIV
NC_001147.6	Chromosome XV
NC_001148.4	Chromosome XVI
NC_001224.1	Mitochondrion

## Data availability

All information and materials provided by SGD are available with the Creative Commons Attribution 4.0 International (CC BY 4.0) license. This license allows others to distribute, remix, adapt, and build upon the information or materials, even commercially, as long as credit to the source is provided.

[Supplementary material](#) is available at GENETICS online.

## Acknowledgments

Thanks to Jodi Lew-Smith for help preparing this manuscript.

## Funding

Research reported in this publication was supported by the National Human Genome Research Institute of the National Institutes of Health under Award Numbers U24HG001315, U24HG010859, and U41HG002273. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Conflicts of interest

The authors declare that there is no conflict of interest.

## Literature cited

Alliance of Genome Resources Consortium. 2019. The Alliance of Genome Resources: building a modern data ecosystem for model organism databases. *Genetics*. 213:1189–1196. doi:10.1534/genetics.119.302523.

Alliance of Genome Resources Consortium. 2020. Alliance of Genome Resources Portal: unified model organism research platform. *Nucleic Acids Res*. 48:D650–D658. doi:10.1093/nar/gkz813.

Alliance of Genome Resources. 2022. Genetics the same issue as this paper.

Awasthi A, Nain V, Srikanth CV, Puria R. 2020. A regulatory circuit between lncRNA and TOR directs amino acid uptake in yeast. *Biochim Biophys Acta Mol Cell Res*. 1867:118680. doi:10.1016/j.bbamcr.2020.118680.

Balakrishnan R, Christie KR, Costanzo MC, Dolinski K, Dwight SS, et al. 2004. Fungal BLAST and model organism BLASTP best hits:

new comparison resources at the *Saccharomyces* Genome Database (SGD). *Nucleic Acids Res*. 33:D374–D377.

Balakrishnan R, Harris MA, Huntley R, Van Auken K, Cherry JM. 2013. A guide to best practices for Gene Ontology (GO) manual annotation. *Database (Oxford)*. 2013:bat054. doi:10.1093/database/bat054.

Ball CA, Dolinski K, Dwight SS, Harris MA, Issel-Tarver L, et al. 2000. Integrating functional genomic information into the *Saccharomyces* Genome Database. *Nucleic Acids Res*. 28:77–80.

Ball CA, Jin H, Sherlock G, Weng S, Matese JC, et al. 2001. *Saccharomyces* Genome Database provides tools to survey gene expression and functional analysis data. *Nucleic Acids Res*. 29:80–81.

Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE; Mouse Genome Database Group. 2019. Mouse genome database (MGD) 2019. *Nucleic Acids Res*. 47:D801–D806. doi:10.1093/nar/gky1056.

Bunina D, Štefl M, Huber F, Khmelinskii A, Meurer M, et al. 2017. Upregulation of SPS100 gene expression by an antisense RNA via a switch of mRNA isoforms with different stabilities. *Nucleic Acids Res*. 45:11144–11158. doi:10.1093/nar/gkx737.

Camblong J, Iglesias N, Fickentscher C, Dieppois G, Stutz F. 2007. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell*. 131:706–717. doi:10.1016/j.cell.2007.09.014.

Cherry JM. 1998. Genetic nomenclature guide, *Saccharomyces cerevisiae*. *Trends Genet*. 14:S10–S11.

Cherry JM. 2015. The *Saccharomyces* Genome Database: exploring biochemical pathways and mutant phenotypes. *Cold Spring Harb Protoc*. 2015:pdb.prot088898.

Cherry JM, Adler C, Ball CA, Chervitz SA, Dwight SS, et al. 1998. SGD: *Saccharomyces* Genome Database. *Nucleic Acids Res*. 26:73–80.

Cherry JM, Ball C, Weng S, Juvik G, Schmidt R, et al. 1997. Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature*. 387:67–73.

Chervitz SA, Hester ET, Ball CA, Dolinski K, Dwight SS, et al. 1999. Using the *Saccharomyces* Genome Database (SGD) for analysis of protein similarities and structure. *Nucleic Acids Res*. 27:74–78.

Christie KR, Weng S, Balakrishnan R, Costanzo MC, Dolinski K, et al. 2004. *Saccharomyces* Genome Database (SGD) provides tools to identify and analyze sequences from *Saccharomyces cerevisiae* and related sequences from other organisms. *Nucleic Acids Res*. 32:D311–D314.

Costanzo M, VanderSluis B, Koch EH, Baryshnikova A, Pons C, et al. 2016. A global genetic interaction network maps a wiring diagram of cellular function. *Science*. 353:aaf1420. doi:10.1126/science.aaf1420.

Costanzo MC, Park J, Balakrishnan R, Cherry JM, Hong EL. 2011. Using computational predictions to improve literature-based Gene Ontology annotations: a feasibility study. *Database (Oxford)*. 2011:bar004. doi:10.1093/database/bar004.

Costanzo MC, Skrzypek MS, Nash R, Wong E, Binkley G, et al.; The *Saccharomyces* Genome Database Project. 2009. New mutant phenotype data curation system in the *Saccharomyces* Genome Database. *Database (Oxford)*. 2009:bap001. doi:10.1093/database/bap001.

Deluca TF, Wu IH, Pu J, Monaghan T, Peshkin L, et al. 2006. Roundup: a multi-genome repository of orthologs and evolutionary distances. *Bioinformatics*. 22:2044–2046.

Dolinski K, Botstein D. 2007. Orthology and functional conservation in eukaryotes. *Annu Rev Genet*. 41:465–507. doi:10.1146/annurev.genet.40.110405.090439.

Dwight SS, Harris MA, Dolinski K, Ball CA, Binkley G, et al. 2002. *Saccharomyces* Genome Database (SGD) provides secondary gene

- annotation using the Gene Ontology (GO). *Nucleic Acids Res.* 30: 69–72.
- Dwight SS, Balakrishnan R, Christie KR, Costanzo MC, Dolinski K, et al. 2004. *Saccharomyces* genome database: underlying principles and organisation. *Brief Bioinform.* 5:9–22.
- Eilbeck K, Lewis SE, Mungall CJ, Yandell M, Stein L, et al. 2005. The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol.* 6:R44. doi:10.1186/gb-2005-6-5-r44.
- Eisenberg-Bord M, Mari M, Weill U, Rosenfeld-Gur E, Moldavski O, et al. 2018. Identification of seipin-linked factors that act as determinants of a lipid droplet subpopulation. *J Cell Biol.* 217:269–282. doi:10.1083/jcb.201704122.
- Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157. doi:10.1186/s13059-015-0721-2.
- Engel SR, Cherry JM. 2013. The new modern era of yeast genomics: community sequencing and the resulting annotation of multiple *Saccharomyces cerevisiae* strains at the *Saccharomyces* Genome Database. *Database (Oxford)*. 2013:bat012. doi:10.1093/database/bat012.
- Engel SR, Balakrishnan R, Binkley G, Christie KR, Costanzo MC, et al. 2010. *Saccharomyces* Genome Database provides mutant phenotype data. *Nucleic Acids Res.* 38:D433–D436.
- Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, et al. 2014. The reference genome sequence of *Saccharomyces cerevisiae*: then and now. *G3 (Bethesda)*. 4:389–398. doi:10.1534/g3.113.008995.
- Engel SR, Skrzypek MS, Hellerstedt ST, Wong ED, Nash RS, et al. 2018. Updated regulation curation model at the *Saccharomyces* Genome Database. *Database (Oxford)*. 2018:bay007. doi:10.1093/database/bay007.
- Fisk DG, Ball CA, Dolinski K, Engel SR, Hong EL, et al.; *Saccharomyces* Genome Database Project. 2006. *Saccharomyces cerevisiae* S288C genome annotation: a working hypothesis. *Yeast*. 23:857–865.
- Geisler S, Lojek L, Khalil AM, Baker KE, Coller J. 2012. Decapping of long noncoding RNAs regulates inducible genes. *Mol Cell.* 45: 279–291. doi:10.1016/j.molcel.2011.11.025.
- Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, et al. 1996. Life with 6000 genes. *Science*. 274:546, 563–567. doi:10.1126/science.274.5287.546.
- Hawthorne DC, Mortimer RK. 1960. Chromosome mapping in *Saccharomyces*: centromere-linked genes. *Genetics*. 45: 1085–1110.
- He C, Jia C, Zhang Y, Xu P. 2018. Enrichment-based proteogenomics identifies microproteins, missing proteins, and novel smORFs in *Saccharomyces cerevisiae*. *J Proteome Res.* 17:2335–2344. doi: 10.1021/acs.jproteome.8b00032.
- Heinicke S, Livstone MS, Lu C, Oughtred R, Kang F, et al. 2007. The Princeton Protein Orthology Database (P-POD): a comparative genomics analysis tool for biologists. *PLoS One*. 2:e766. doi: 10.1371/journal.pone.0000766.
- Hellerstedt ST, Nash RS, Weng S, Paskov KM, Wong ED, et al. 2017. Curated protein information in the *Saccharomyces* Genome Database. 2017:bax011. doi:10.1093/database/bax011.
- Hirschman JE, Balakrishnan R, Christie KR, Costanzo MC, Dwight SS, et al. 2006. Genome snapshot: a new resource at the *Saccharomyces* Genome Database (SGD) presenting an overview of the *Saccharomyces cerevisiae* genome. *Nucleic Acids Res.* 34: D442–D445.
- Hong EL, Balakrishnan R, Dong Q, Christie KR, Park J, et al. 2008. Gene Ontology annotations at SGD: new data sources and annotation methods. *Nucleic Acids Res.* 36:D577–D581.
- Houseley J, Rubbi L, Grunstein M, Tollervey D, Vogelauer M. 2008. A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. *Mol Cell.* 32:685–695. doi:10.1016/j.molcel.2008.09.027.
- Howe DG, Blake JA, Bradford YM, Bult CJ, Calvi BR, et al. 2018. Model organism data evolving in support of translational medicine. *Lab Anim (NY)*. 47:277–289. doi:10.1038/s41684-018-0150-4.
- Hu Y, Flockhart I, Vinayagam A, Bergwitz C, Berger B, et al. 2011. An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics*. 12:357. doi: 10.1186/1471-2105-12-357.
- Huber F, Bunina D, Gupta I, Khmelinskii A, Meurer M, et al. 2016. Protein abundance control by non-coding antisense transcription. *Cell Rep.* 15:2625–2636. doi:10.1016/j.celrep.2016.05.043.
- Huerta-Cepas J, Bueno A, Dopazo J, Gabaldon T. 2008. PhylomeDB: a database for genome-wide collections of gene phylogenies. *Nucleic Acids Res.* 36:D491–D496.
- Jeon S, Kim J. 2010. Upstream open reading frames regulate the cell cycle-dependent expression of the RNA helicase Rok1 in *Saccharomyces cerevisiae*. *FEBS Lett.* 584:4593–4598. doi:10.1016/j.febslet.2010.10.019.
- Kaduk M, Sonnhammer E. 2017. Improved orthology inference with Hieranoid 2. *Bioinformatics*. 33:1154–1159.
- Laulederkind SJF, Hayman GT, Wang SJ, Smith JR, Petri V, et al. 2018. A primer for the rat genome database (RGD). *Methods Mol Biol.* 1757:163–209. doi:10.1007/978-1-4939-7737-6\_8.
- Lee RYN, Howe KL, Harris TW, Arnaboldi V, Cain S, et al. 2018. WormBase 2017: molting into a new stage. *Nucleic Acids Res.* 46: D869–D874. doi:10.1093/nar/gkx998.
- Li H, Coghlan A, Ruan J, Coin LJ, Heriche JK, et al. 2006. TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res.* 34:D572–D580.
- Li L, Stoeckert CJ, Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13: 2178–2189. doi:10.1101/gr.1224503.
- Linarid B, Ebersberger I, McGlynn S, Glover N, Mochizuki T, et al.; QFO Consortium. 2021. Ten years of collaborative progress in the Quest for Orthologs. *Mol Biol Evol.* 38:3033–3045. doi:10.1093/molbev/msab098.
- Linarid B, Thompson JD, Poch O, Lecompte O. 2011. OrthoInspector: comprehensive orthology analysis and visual exploration. *BMC Bioinformatics*. 12:11. doi:10.1186/1471-2105-12-11.
- Nene RV, Putnam CD, Li BZ, Nguyen KG, Srivatsan A, et al. 2018. Cdc73 suppresses genome instability by mediating telomere homeostasis. *PLoS Genet.* 14:e1007170. doi:10.1371/journal.pgen.1007170.
- MacPherson KA, Starr B, Wong ED, Dalusag KS, Hellerstedt ST, et al. 2017. Outreach and online training services at the *Saccharomyces* Genome Database. 2017:bax002. doi:10.1093/database/bax002.
- Makanaka K, Kintaka R, Ishikawa K, Moriya H. 2015. Small toxic protein encoded on chromosome VII of *Saccharomyces cerevisiae*. *PLoS One*. 10:e0120678. doi:10.1371/journal.pone.0120678.
- Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 47:D419–D426.
- Mortimer RK, Schild D. 1980. Genetic map of *Saccharomyces cerevisiae*. *Microbiol Rev.* 44:519–571.
- Ng PC, Wong ED, MacPherson KA, Aleksander S, Argasinska J, et al. Transcriptome visualization and data availability at the *Saccharomyces* Genome Database. *Nucleic Acids Res.* 2020; 48(D1):D743–D748. doi:10.1093/nar/gkz892

- O'Brien KP, Remm M, Sonnhammer EL. 2005. Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res.* 33: D476–D480.
- Oughtred R, Rust J, Chang C, Breitkreutz B-J, Stark C, et al. 2021. The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* 30: 187–200.
- Park J, Costanzo MC, Balakrishnan R, Cherry JM, Hong EL. 2012. CvManGO, a method for leveraging computational predictions to improve literature-based Gene Ontology annotations. *Database (Oxford)*. 2012:bas001. doi:10.1093/database/bas001.
- Paul VD, Mühlenhoff U, Stümpfig M, Seebacher J, Kugler KG, et al. 2015. The deca-GX3 proteins Yae1-Lto1 function as adaptors recruiting the ABC protein Rli1 for iron-sulfur cluster insertion. *Elife*. 4:e08231. doi:10.7554/eLife.08231.
- Pinskaya M, Gourvenec S, Morillon A. H3 lysine 4 di- and tri-methylation deposited by cryptic transcription attenuates promoter activation. *EMBO J.* 2009;28:1697–1707.
- Povey S, Lovering R, Bruford E, Wright M, Lush M, et al. 2001. The HUGO Gene Nomenclature Committee (HGNC). *Hum Genet.* 109: 678–680.
- Ruzicka L, Howe DG, Ramachandran S, Toro S, Van Slyke CE, et al. 2019. The Zebrafish Information Network: new support for non-coding genes, richer Gene Ontology annotations and the Alliance of Genome Resources. *Nucleic Acids Res.* 47:D867–D873. doi:10.1093/nar/gky1090.
- Sadhu MJ, Bloom JS, Day L, Siegel JJ, Kosuri S, et al. 2018. Highly parallel genome variant engineering with CRISPR-Cas9. *Nat Genet.* 50: 510–514. doi:10.1038/s41588-018-0087-y.
- Sampath V, Liu B, Tafrov S, Srinivasan M, Rieger R, et al. 2013. Biochemical characterization of Hpa2 and Hpa3, two small closely related acetyltransferases from *Saccharomyces cerevisiae*. *J Biol Chem.* 288:21506–21513. doi:10.1074/jbc.M113.486274.
- Sheppard TK, Hitz BC, Engel SR, Song G, Balakrishnan R, et al. 2016. The *Saccharomyces* Genome Database variant viewer. *Nucleic Acids Res.* 44:D698–D702. doi:10.1093/nar/gkv1250.
- Schriml LM, Mitraka E, Munro J, Tauber B, Schor M, et al. 2019. Human Disease Ontology 2018 update: classification, content and workflow expansion. *Nucleic Acids Res.* 47:D955–D962. doi: 10.1093/nar/gky1032.
- Schneider A, Dessimoz C, Gonnet GH. 2007. OMA Browser—exploring orthologous relations across 352 complete genomes. *Bioinformatics.* 23:2180–2182. doi:10.1093/bioinformatics/btm295.
- Skrzypek MS, Nash RS. 2015. Biocuration at the *Saccharomyces* Genome Database. *Genesis.* 53:450–457. doi:10.1002/dvg.22862.
- Song G, Dickins BJ, Demeter J, Engel S, Dunn B, et al. 2015. AGAPE (Automated Genome Analysis Pipeline) for pan-genome analysis of *Saccharomyces cerevisiae*. *PLoS One.* 10:e0120671. doi:10.1371/journal.pone.0120671.
- The Gene Ontology Consortium. 2019. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47:D330–D338. doi:10.1093/nar/gky1055.
- Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, et al.; FlyBase Consortium. 2019. FlyBase 2.0: the next generation. *Nucleic Acids Res.* 47:D759–D765. doi:10.1093/nar/gky1003.
- Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, et al. 2009. EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* 19:327–335.
- Weng S, Dong Q, Balakrishnan R, Christie K, Costanzo M, et al. 2003. *Saccharomyces* Genome Database (SGD) provides biochemical and structural information for budding yeast proteins. *Nucleic Acids Res.* 31:216–218.
- Westerfield M, Doerry E, Kirkpatrick AE, Driever W, Douglas SA. An on-line database for zebrafish development and genetics research. *Semin Cell Dev Biol.* 1997;8:477–488.
- Wong ED, Skrzypek MS, Weng S, Binkley G, Meldal BHM, et al.; the SGD Project. 2019. Integration of macromolecular complex data into the *Saccharomyces* Genome Database. *Database.* 2019:baz008. doi:10.1093/database/baz008.
- Wu X, Haber JE. 1996. A 700 bp cis-acting region controls mating-type dependent recombination along the entire left arm of yeast chromosome III. *Cell.* 87:277–285. doi:10.1016/s0092-8674(00)81345-8.
- Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, et al. 2009. Bidirectional promoters generate pervasive transcription in yeast. *Nature.* 457:1033–1037. doi:10.1038/nature07728.
- Yagoub D, Tay AP, Chen Z, Hamey JJ, Cai C, et al. 2015. Proteogenomic discovery of a small, novel protein in yeast reveals a strategy for the detection of unannotated short open reading frames. *J Proteome Res.* 14:5038–5047. doi:10.1021/acs.jproteome.5b00734.

Communicating editor: V. Wood