



Downloading Data from *Saccharomyces* Genome Database



Felix Gondwe, Gail A. Binkley, Travis K. Sheppard, Pedro H. R. de Assis, Kalpana Karra, Ajay Shrivatsav Vichanthangal Prathivadhib, Shuai Weng, J. Michael Cherry and the SGD Project

Stanford University, School of Medicine, Department of Genetics, Stanford, CA

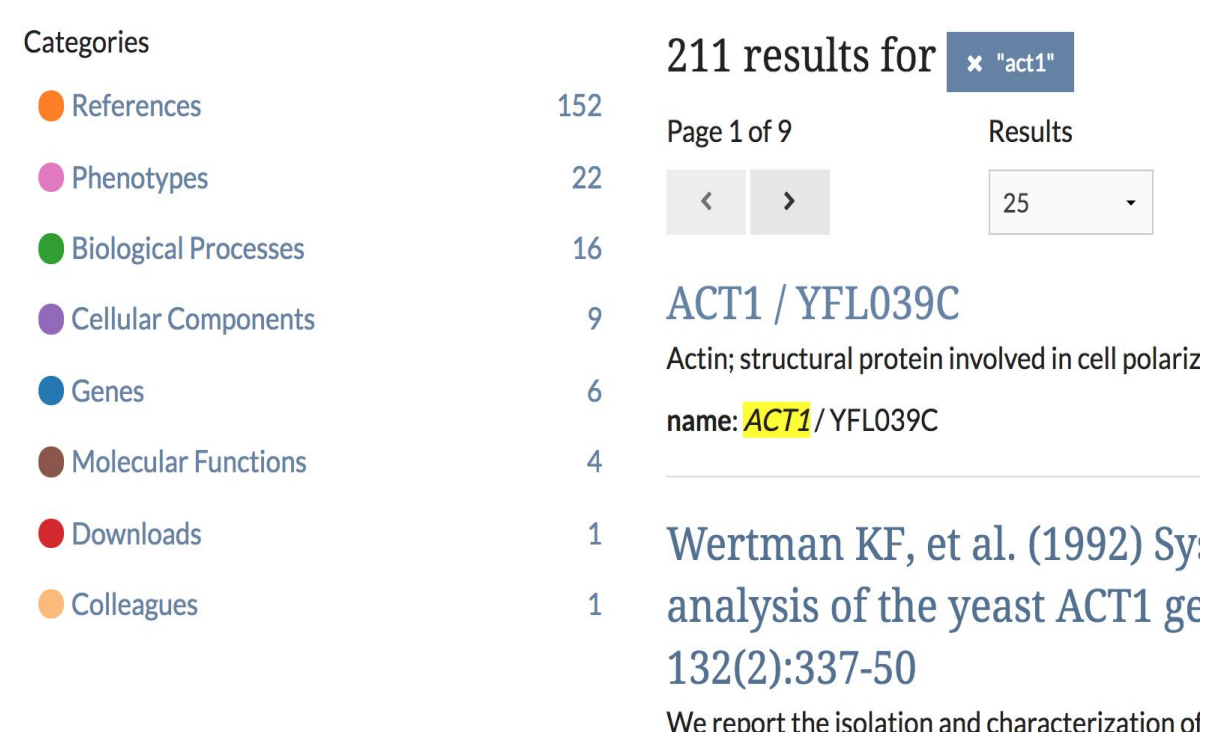
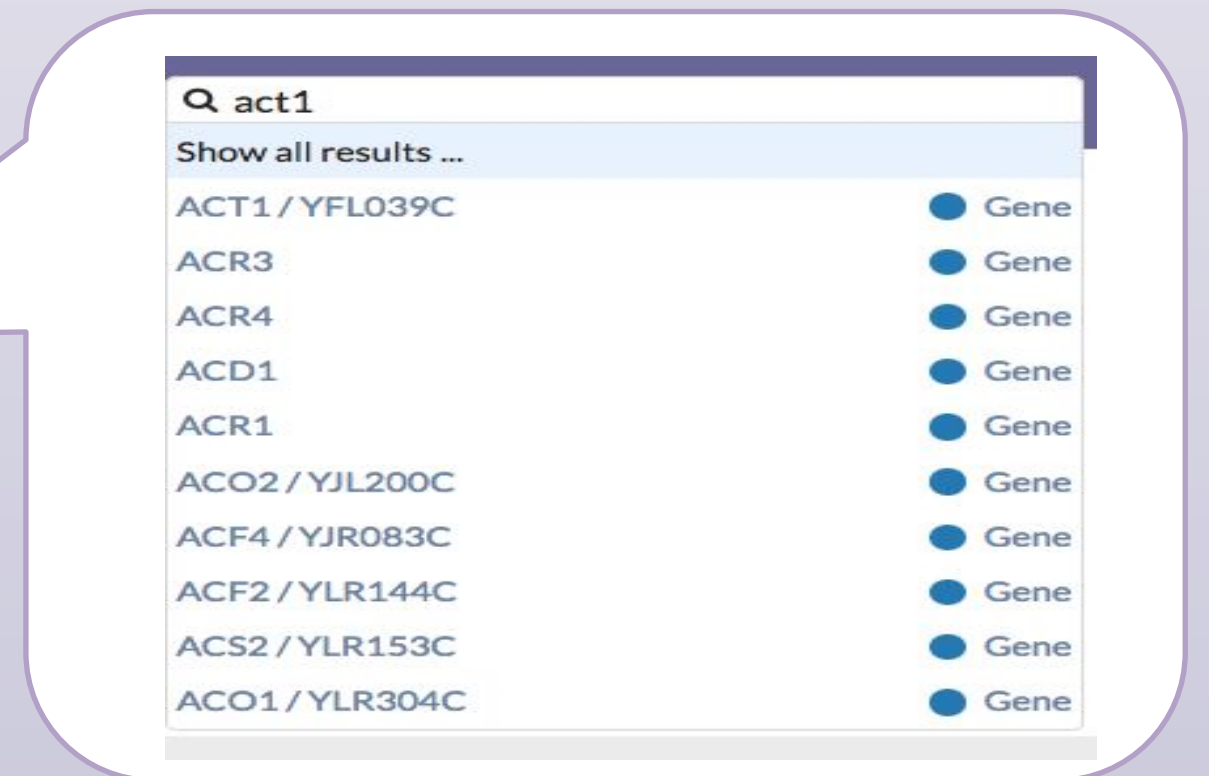
Data at the *Saccharomyces* Genome Database (SGD; <http://www.yeastgenome.org>) is accessed through a dynamic faceted search backed by Elasticsearch technology, a full-text search and analytics engine. Elasticsearch provides a powerful application programming interface (API) into SGD's complex data. Traditionally, SGD users used the File Transfer Protocol (FTP) server to download data files, but we are moving towards a more flexible approach by leveraging our search tool. File metadata has been loaded into the database and integrated into the greater SGD search. As a result, users can search for files the same way they might search for a gene or phenotype data. Results are matched based on metadata, including the file name, description, keywords, and the PMID if the file is associated with a published reference. The files themselves are stored using the AWS (Amazon Web Services) S3 storage service. Our new approach to downloading SGD data files provides our users with a more customized, end-to-end experience. SGD is funded by the NIH-NHGRI [U41HG001315].

Search

Dynamic full text search backed by Elasticsearch allows fast execution of complex data queries



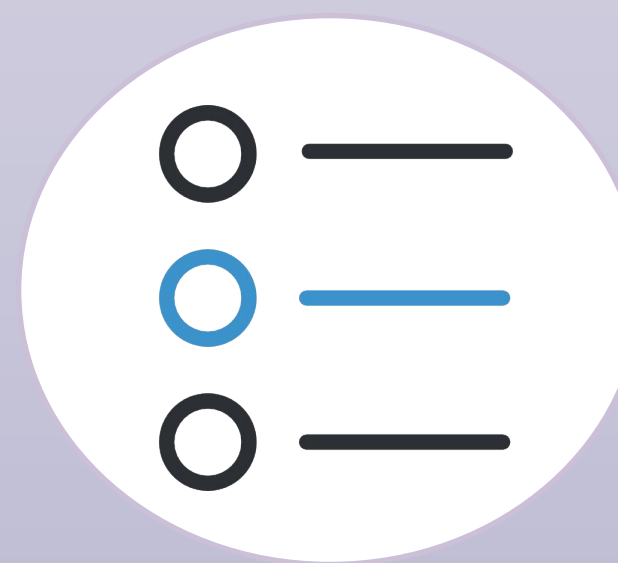
01



02

Faceted Results

Faceted search results allows granular filtering and classification of downloadable data files

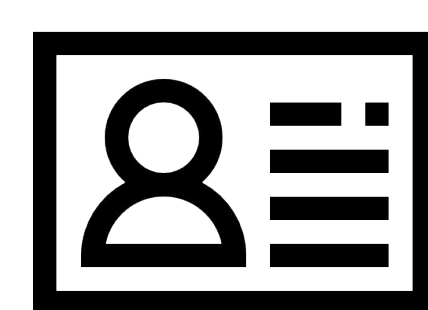
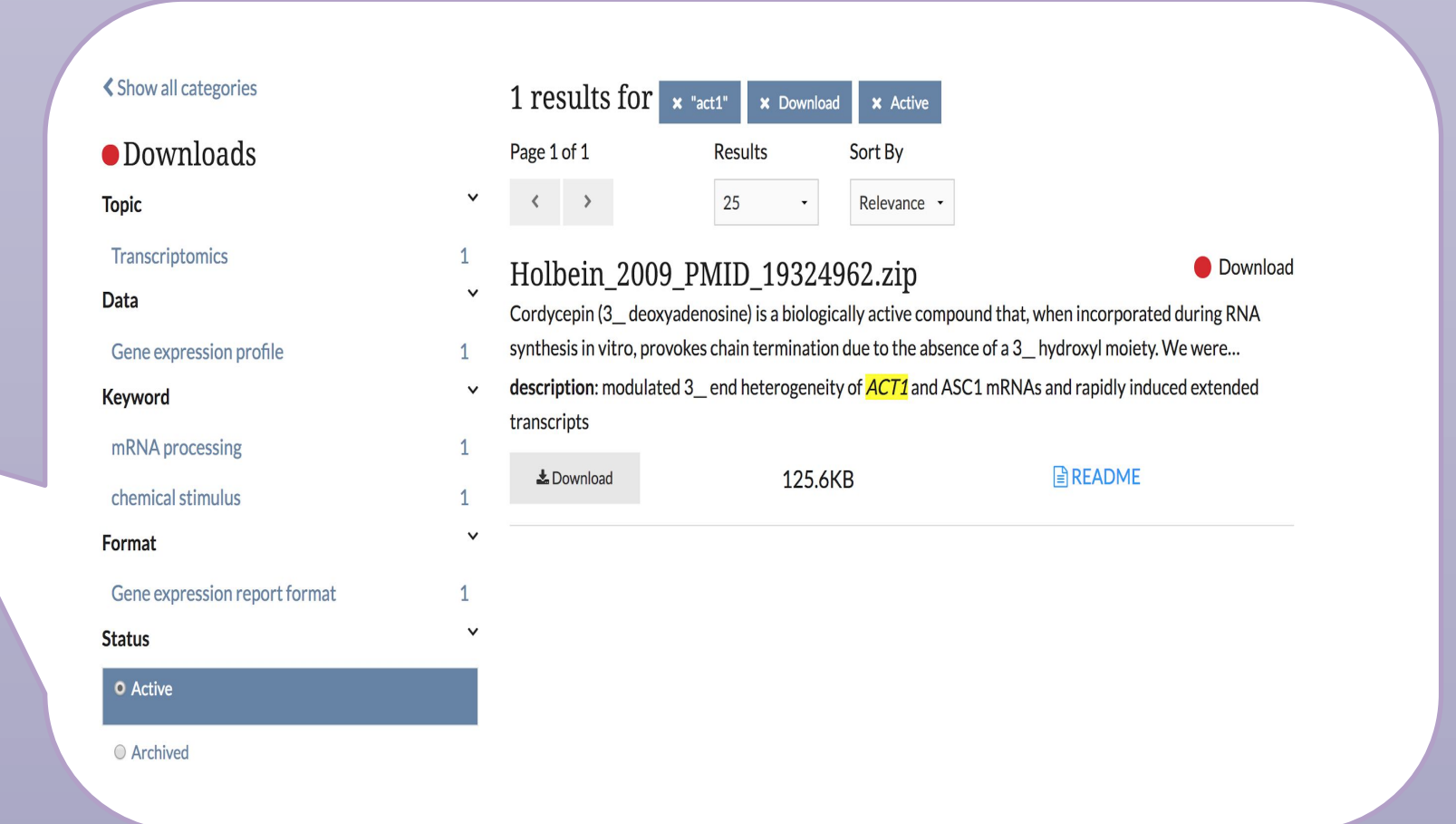


Download

Download data files hosted in AWS S3 bucket for high availability and almost 0 downtime



03



Contact Us

- @yeastgenome
- @yeastgenome
- sgd-helpdesk@lists.stanford.edu

04

