

This information has not been peer-reviewed. Responsibility for the findings rests solely with the author(s).

Deposited research article

The GRID: The General Repository for Interaction Datasets

Bobby-Joe Breitkreutz*, Chris Stark* and Mike Tyers

Address: Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto M5G 1X5, Canada.

Correspondence: Mike Tyers. E-mail: tyers@mshri.on.ca

Posted: 21 November 2002

Genome Biology 2002, **3(12)**:preprint0013.1-0013.5

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2002/3/12/preprint/0013>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

Received: 14 November 2002

This is the first version of this article to be made available publicly. A peer-reviewed and modified version is now available in full at <http://genomebiology.com/2003/4/3/R23>



deposited research

AS A SERVICE TO THE RESEARCH COMMUNITY, GENOME **BIOLOGY** PROVIDES A 'PREPRINT' DEPOSITORY TO WHICH ANY ORIGINAL RESEARCH CAN BE SUBMITTED AND WHICH ALL INDIVIDUALS CAN ACCESS FREE OF CHARGE. ANY ARTICLE CAN BE SUBMITTED BY AUTHORS, WHO HAVE SOLE RESPONSIBILITY FOR THE ARTICLE'S CONTENT. THE ONLY SCREENING IS TO ENSURE RELEVANCE OF THE PREPRINT TO GENOME **BIOLOGY**'S SCOPE AND TO AVOID ABUSIVE, LIBELLOUS OR INDECENT ARTICLES. ARTICLES IN THIS SECTION OF THE JOURNAL HAVE **NOT** BEEN PEER-REVIEWED. EACH PREPRINT HAS A PERMANENT URL, BY WHICH IT CAN BE CITED. RESEARCH SUBMITTED TO THE PREPRINT DEPOSITORY MAY BE SIMULTANEOUSLY OR SUBSEQUENTLY SUBMITTED TO GENOME **BIOLOGY** OR ANY OTHER PUBLICATION FOR PEER REVIEW; THE ONLY REQUIREMENT IS AN EXPLICIT CITATION OF, AND LINK TO, THE PREPRINT IN ANY VERSION OF THE ARTICLE THAT IS EVENTUALLY PUBLISHED. IF POSSIBLE, GENOME **BIOLOGY** WILL PROVIDE A RECIPROCAL LINK FROM THE PREPRINT TO THE PUBLISHED ARTICLE.



Software report

The GRID: The General Repository for Interaction Datasets

Bobby-Joe Breitkreutz*, **Chris Stark***, and **Mike Tyers**

Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Canada, M5G 1X5

*These authors contributed equally to this work.

Correspondence: Mike Tyers, Email tyers@mshri.on.ca

Abstract

We have developed a relational database, called the General Repository for Interaction Datasets (The GRID; <http://biodata.mshri.on.ca/grid>) to archive and display physical, genetic and functional interactions. The GRID displays data-rich interaction tables for any protein of interest, combines literature-derived and high throughput interaction datasets, and is readily accessible via the World Wide Web. Interactions parsed in the GRID can be viewed in graphical form with a versatile visualization tool called Osprey.

Rationale

Physical, genetic and functional interactions between biological molecules are being discovered at an ever-increasing rate through proteomic and functional genomic approaches [1]. As a result, large-scale datasets containing many thousands of interactions have been deposited in publicly available databases. However, none of the extant datasets are systematically linked, and most data is presented only in a rudimentary format. Thus, data analysis is often tedious and incomplete. To alleviate this bottleneck, we developed a generic interaction database called The GRID, which can be used to collate and display interactions from any data source.

Software Platform

The GRID uses MySQL version 3.23 as its underlying database [2], which is freely available from the MySQL homepage. The web-based user interface is implemented with Java Servlet technology, and the Java SDK version 1.4.0_02 [3]. These tools provide a facile interface for parsing interactions. Graphical representation of user defined interaction networks is achieved with a new visualization tool called Osprey (<http://biodata.mshri.on.ca/osprey>), which can be used to construct elaborate interaction networks from any set of interactions in the database.

Data Structure

The GRID is built on a master look-up table of all primary and secondary *Saccharomyces cerevisiae* gene names and corresponding Open Reading Frame (ORF) names. Valid name lists are compiled via the open File Transfer Protocol (FTP) provided by the *Saccharomyces* Genome Database (SGD) [4]. Currently, The GRID recognizes 6355 unique ORFs. Each gene entry in the GRID is presented in a data-rich tabular format that includes a description of gene function, Gene Ontology (GO) annotation [5], experimental system(s) on which associated interactions are based, the source of interaction data and publication links. Each row in the table represents a unique interaction, which is further divided into additional sub-sections corresponding to each experimental system in which the interaction is reported. Defined physical and genetic experimental systems currently include: affinity precipitation, affinity chromatography, two hybrid, purified complex, reconstituted complex, biochemical assay, synthetic lethality, synthetic rescue, dosage lethality, dosage suppression, chemical lethality and chemical rescue. Additional systems may be added as needed.

Datasets

The GRID is periodically updated to contain all published large-scale interaction datasets, as well as available curated interactions from the primary literature. At present, The GRID contains a total of 13,830 unique interactions and 21,839 total interactions, including most interactions deposited in BIND [6] and MIPS [7], as well as large-scale protein interaction datasets generated by Uetz et al [8], Ito et al [9,10],

Gavin et al [11] and Ho et al [12] and a synthetic lethal interaction dataset produced by Tong et al [13]. An upload interactions option allows new interactions to be added from a tab-delimited text file that contains the interaction pair, the experimental system and the data source. The GRID only accepts new interactions, so redundant interactions are excluded during the upload process. Details for upload format are provided at The GRID website.

Searches

Any valid gene or ORF name can be searched to yield a comprehensive list of known interactions and associated annotations in tabular format (Figure 1). The search result table provides direct links for recursive searches, PubMed citations and data-rich graphical visualization with Osprey. In addition to standard keyword searches, an advanced search option allows keywords to be combined with Boolean operators to expand or reduce the number of recovered interactions. Results from advanced searches can be displayed using Osprey, or saved as a tab-delimited text file.

Access and Software Requirements

All relevant information on The GRID can be retrieved from The GRID website at <http://biodata.mshri.on.ca/grid>. The GRID runs with the most recent versions of popular web browsers on all major platforms. An online version of the Osprey Network Visualization System is available as an add-on to The GRID that is automatically launched from a link on The GRID search result page. The Osprey add-on requires version 1.4.0_02 of the Java Plug-in [3]. A full-featured application version of Osprey is available for non-profit use at <http://biodata.mshri.on.ca/osprey> (see accompanying software report).

Private Versions

Individual laboratories at not-for profit institutions may request a private version of The GRID that can be easily customized to allow storage and manipulation of unpublished datasets, including integration and comparison with all publicly available interactions. The GRID is designed to work with any set of interactions, including those derived from other model organisms, combinations of organism systems and even social or commercial networks for which interaction data is available. For more information about hosting a private version of The GRID, please contact the authors.

Acknowledgments

We thank Lorrie Boucher, Ashton Breitzkreutz and Paul Jorgensen for suggestions on GRID features. Development of The GRID was supported by the Canadian Institutes of Health Research. MT is a Canada Research Chair in Biochemistry.

References

- 1) Vidal, M: **A biological atlas of functional maps**. *Cell* 2001, 104: 333-339.
- 2) **MySQL** [<http://www.mysql.com/>]
- 3) **Sun Microsystems Java Standard Development Kit 1.4.0_02** [<http://java.sun.com>]
- 4) Cherry et al. **Saccharomyces Genome Database** [<ftp://genome-ftp.stanford.edu/pub/yeast/SacchDB/>], June 2002.
- 5) The Gene Ontology Consortium: **Gene Ontology: tool for the unification of biology**. *Nature Genetics* 2000, 25: 25-2
- 6) Bader et al. **BIND--a data specification for storing and describing biomolecular interactions, molecular complexes and pathways**. *Bioinformatics* 16(5):465-77.
- 7) **Munich information center for protein sequences** [<http://mips.gsf.de/>]
- 8) Uetz et al. **A Comprehensive Analysis of Protein-Protein Interactions in Saccharomyces cerevisiae**. *Nature* 2000 Feb 10;403(6770):623-7
- 9) Ito et al. **A comprehensive two-hybrid analysis to explore the yeast protein interactome**. *Proc Natl Acad Sci U S A* 2001 Apr 10;98(8):4569-74
- 10) Ito et al. **Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins**. *Proc Natl Acad Sci U S A* 2000 Feb 1;97(3):1143-7

- 11) Gavin et al. **Functional organization of the yeast proteome by systematic analysis of protein complexes.** *Nature* 2002 Jan 10;415(6868):141-7
- 12) Ho et al. **Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry.** *Nature* 2002 Jan 10;415(6868):180-3
- 13) Tong et al. **Systematic genetic analysis with ordered arrays of yeast deletion mutants.** *Science* 2001 Dec 14;294(5550):2364-8

Figure 1. Search result page from The GRID. Multiple experimental systems and sources are indicated.

MOB1 / YIL106W

Search Search
(Search by Gene or ORF name)

Orf	Gene/s	Description	GO Annotations			Links
			Function	Process	Component	
YIL106W	• MOB1	Mps One Binder	• molecular_function unknown	• mitosis	• bud neck	

MOB1 was identified in association with the following 12 proteins

Orf	Gene/s	Description	GO Annotations			System	Source/s	Pubmed	Links
			Function	Process	Component				
YPR111W	• DBF2	kinase required for late nuclear division	• protein serine/threonine kinase	• cell cycle	Affinity Precipitation	MIPS	PubMed	 Visualize View Comment/s	
					Two Hybrid	MIPS	PubMed		
YGR092W	• DBF2	Kinase required for late nuclear division. C promotes exit from mitosis by directly switching on the kinase activity of Dbf2.	• protein kinase	• cell cycle	Affinity Precipitation	MIPS	PubMed	 Visualize View Comment/s	
					Synthetic Lethality	MIPS	PubMed		
					Two Hybrid	MIPS	PubMed		

Interaction with multiple sources

Interaction with multiple experimental systems