BIOVERSE: enhancements to the framework for structural, functional and contextual modeling of proteins and proteomes

Jason McDermott, Michal Guerquin, Zach Frazier, Aaron N. Chang and Ram Samudrala*

Department of Microbiology, University of Washington, Seattle, WA, USA

ABSTRACT

We have made a number of enhancements to the previously described Bioverse web server and computational biology framework (http://bioverse. compbio.washington.edu). In this update, we provide an overview of the new features available that include: (i) expansion of the number of organisms represented in the Bioverse and addition of new data sources and novel prediction techniques not available elsewhere, including network-based annotation; (ii) reengineering the database backend and supporting code resulting in significant speed, search and ease-of-use improvements; and (iii) creation of a stateful and dynamic web application frontend to improve interface speed and usability. Integrated Java-based applications also allow dynamic visualization of real and predicted protein interaction networks.

INTRODUCTION

We described the web-based interface to the Bioverse framework previously (1), which provides objected-oriented representations of biological components and relationships between them, along with associated confidence values, at the single molecule as well as the genomic/proteomic levels. Since then, a number of improvements, detailed below, have been made to the Bioverse database and web interface to increase its utility to the life sciences community.

DATA IMPROVEMENTS

The number of organisms represented in the Bioverse has grown to >50, including >400 000 protein sequences. Network-based functional annotation has been performed for all genomes, providing novel annotations for ~4000 proteins without existing annotations. This method is based on the integration of functions from neighboring proteins in real or predicted protein interaction networks and has been previously shown to provide accurate predictions (2–5). Other new features include Superfamily (6) and CATH (7) sequence to structural classification and evolutionary information content for all proteins. Confidence values for all predictions are dynamic and are constantly being refined against experimental data. Detailed explanations of the derivation of confidence values for each type of prediction are provided on the web server.

FRAMEWORK IMPROVEMENTS

A relational database backend implemented in MySQL and an object-relational mapping layer with an XMLRPC interface have been implemented to facilitate data interchange internally and with other databases. These modifications result in better speed, stability and accessibility compared with the previous implementation.

WEB INTERFACE IMPROVEMENTS

The web server component of the Bioverse is now a stateful and dynamic web application, which provides a more intuitive interface. Web server operations, such as performing a search, dynamically update information in the current browser page using client-directed server requests and content updates. This decreases the time required to render complicated data representations and allows emulation of familiar behaviors of desktop applications. Users can now customize the behavior of the interface using an options page and compile and annotate lists of proteins with a user history manager. The range of search options has also been significantly enhanced and more detailed information about each matched protein is given. A much broader range of protein characteristics is searchable and searches for proteins with particular relationships, e.g. evolutionary similarity and predicted functional interactions, are now possible.

To allow dynamic visualization of predicted and experimental protein interaction networks, we developed a

*To whom correspondence should be addressed. Tel: +1 206 732 6122; Fax: +1 206 732 6055; Email: ram@compbio.washington.edu

Java-based interaction viewer (8) that was capable of only handling networks of limited size. We have developed a second version of this viewer, called the Integrator, that communicates with the Bioverse object layer and enables exploration of arbitrarily large networks (A. N. Chang, Z. Frazier, M. Guerquin, J. McDermott and R. Samudrala, manuscript submitted). In addition, the Integrator can be used to upload user-supplied data, such as gene expression data, and with our predicted networks, visually search for interacting clusters of proteins corresponding to differentially expressed genes.

CONCLUSION

The Bioverse has been used by biologists to annotate and analyze large-scale genome sequencing projects (9,10). The new features described here enhance the value of the resource by providing a rich feature set, intuitive interface and tight integration with visual and algorithmic tools for exploring single molecules and interactomes.

ACKNOWLEDGEMENTS

Conflict of interest statement. None declared.

REFERENCES

1. McDermott,J. and Samudrala,R. (2003) Bioverse: functional, structural and contextual annotation of proteins and proteomes. *Nucleic Acids Res.*, **31**, 3736–3737.
2. McDermott,J. and Samudrala,R. (2004) Enhanced functional information from predicted protein networks. *Trends Biotechnol.*, **22**, 60–62.
3. Schwikowski,B., Uetz,P. and Fields,S. (2000) A network of protein–protein interactions in yeast. *Nat. Biotechnol.*, **18**, 1257–1261.
4. Deng,M., Tu,Z., Sun,F. and Chen,T. (2004) Mapping gene ontology to proteins based on protein–protein interaction data. *Bioinformatics*, **20**, 895–902.
5. Vazquez,A., Flammini,A., Maritan,A. and Vespignani,A. (2003) Global protein function prediction from protein-protein interaction networks. *Nat. Biotechnol.*, **21**, 697–700.
6. Gough,J. and Chothia,C. (2002) SUPERFAMILY: HMMs representing all proteins of known structure. SCOP sequence searches, alignments and genome assignments. *Nucleic Acids Res.*, **30**, 268–272.
7. Pearl,F.M., Bennett,C.F., Bray,J.E., Harrison,A.P., Martin,N., Shepherd,A., Sillitoe,I., Thornton,J. and Orengo,C.A. (2003) The CATH database: an extended protein family resource for structural and functional genomics. *Nucleic Acids Res.*, **31**, 452–455.
8. Chang,A.N., McDermott,J. and Samudrala,R. (2004) An enhanced Java graph applet interface for visualizing interactomes. *Bioinformatics*, in press.
9. Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H. *et al.* (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. *Science*, **301**, 376–379.
10. Yu,J., Wang,J., Lin,W., Li,S., Li,H., Zhou,J., Ni,P., Dong,W., Hu,S., Zeng,C. *et al.* (2005) The genomes of Oryza sativa: a history of duplications. *PLoS Biol.*, **3**, e38.