BIOINFORMATICS APPLICATIONS NOTE Vol. 23 no. 4 2007, pages 524–526
doi:10.1093/bioinformatics/btl523

Systems biology

OmicBrowse: a browser of multidimensional omics annotations

Tetsuro Toyoda*, Yoshiki Mochizuki, Keith Player, Naohiko Heida, Norio Kobayashi and Yoshiyuki Sakaki

Genomic Sciences Center, Yokohama Institute, RIKEN, 1-77-22, Suehiro, Tsurumi, Yokohama, Kanagawa, Japan

Received on May 24, 2006; revised on August 15, 2006; accepted on October 5, 2006
Advance Access publication October 31, 2006
Associate Editor: Alvis Brazma

ABSTRACT

Summary: OmicBrowse is a browser to explore multiple datasets coordinated in the multidimensional omic space integrating omics knowledge ranging from genomes to phenomes and connecting evolutional correspondences among multiple species. OmicBrowse integrates multiple data servers into a single omic space through secure peer-to-peer server communications, so that a user can easily obtain an integrated view of distributed data servers, e.g. an integrated view of numerous whole-genome tiling-array data retrieved from a user's in-house private-data server, along with various genomic annotations from public internet servers. OmicBrowse is especially appropriate for positional-cloning purposes. It displays both genetic maps and genomic annotations within wide chromosomal intervals and assists a user to select candidate genes by filtering their annotations or associated documents against user-specified keywords or ontology terms. We also show that an omic-space chart effectively represents schemes for integrating multiple datasets of multiple species.

Availability: OmicBrowse is developed by the Genome-Phenome Superbrain Project and is released as free open-source software under the GNU General Public License at http://omicspace.riken.jp

Contact: toyop@gsc.riken.go.jp

Genome browsers displaying annotations based on a genomic coordinate axis have become common tools in genetics and genomics (Stein *et al*., 2002; Stalker *et al*., 2004). Thanks to the physical coordinates of nucleic acid sequences elucidated from large-scale genome projects, it has become possible to share the same semantics by indexing genomic annotations based on these coordinates, and to yield profound findings of genomic functions through integrative analyses of the annotations among different datasets. The physical coordinates have also opened a way to realize a globally distributed annotation system (DAS), which is composed of multiple servers configured to return genomic annotation data upon receiving a query from a client specifying an interval on a genomic axis (Dowell *et al*., 2001). The genome browsers and DAS servers are expected to deal with data based on a one-dimensional genomic axis of index. However, recent divergence of data types in genetics and genomics requires more dimensions. Omic space is defined to have multiple orthogonal coordinate axes for heterogeneous datasets to share the same semantics by indexing their data records on the axes (Toyoda and Wada, 2004). A typical biological data record contains many attributes, such as gene names, annotations, genomic positions, ontologies, interactions, phenotypes, expression levels, experimental conditions of time series and sampled tissues. We categorize the variety of attributes into four semantic groups, and define each axis as the array of indices representing the possible instances of the attributes in each group. Herein, the omic space has four semantic dimensions: a dataset axis, an ontology axis, and two genomic axes representing interactions and evolutional correspondences (Fig. 1). Further dimensions are conceivable if necessary, but are not described here. The user's interface for configuring the dataset axis contains a menu of indices with which a user specifies the datasets to browse (Fig. 1). The ontology axis has interfaces for browsing ontology trees and a textbox, with which a user can specify terms and conditions, e.g. keywords, ontology terms, expression levels and accessions.

OmicBrowse is a system composed of a client and a server. The client is a Flash application which automatically runs in an ordinary internet browser with Adobe Flash plug-in, and retrieves XML-formatted data from the server composed of Java servlets. Upon receiving a HTTP request sent by a client specifying multiple conditions on every axis, the server gathers records from those multiple datasets that satisfy the query and returns them as a single HTTP response. A user can configure a server to subscribe to other OmicBrowse servers and to return a response of merged data thereof, so that the user can browse his/her private datasets along with public servers' datasets easily, and search multiple datasets across every server simultaneously and genome-widely. The subscription relationships are inheritably shared among user-trusted servers automatically, so as to establish secure peer-to-peer (P2P) communications among the servers. P2P databases are known to have so-called heterogeneity of information sources problem preventing semantic integration of distributed data (Gribble *et al*., 2001). We propose here our protocols termed Semantic-Space-based Distributed Annotation System (SDAS), which we expect will overcome the difficulties and facilitate global semantic integration of heterogeneous data. Optionally, a configuration tool, sequence server, DAS1 adapter, manuals, tutorials, and OmicBrowse demo databases for human, mouse, *Caenorhabditis elegans* and *Arabidopsis* are available at our website. Also, we

*To whom correspondence should be addressed.

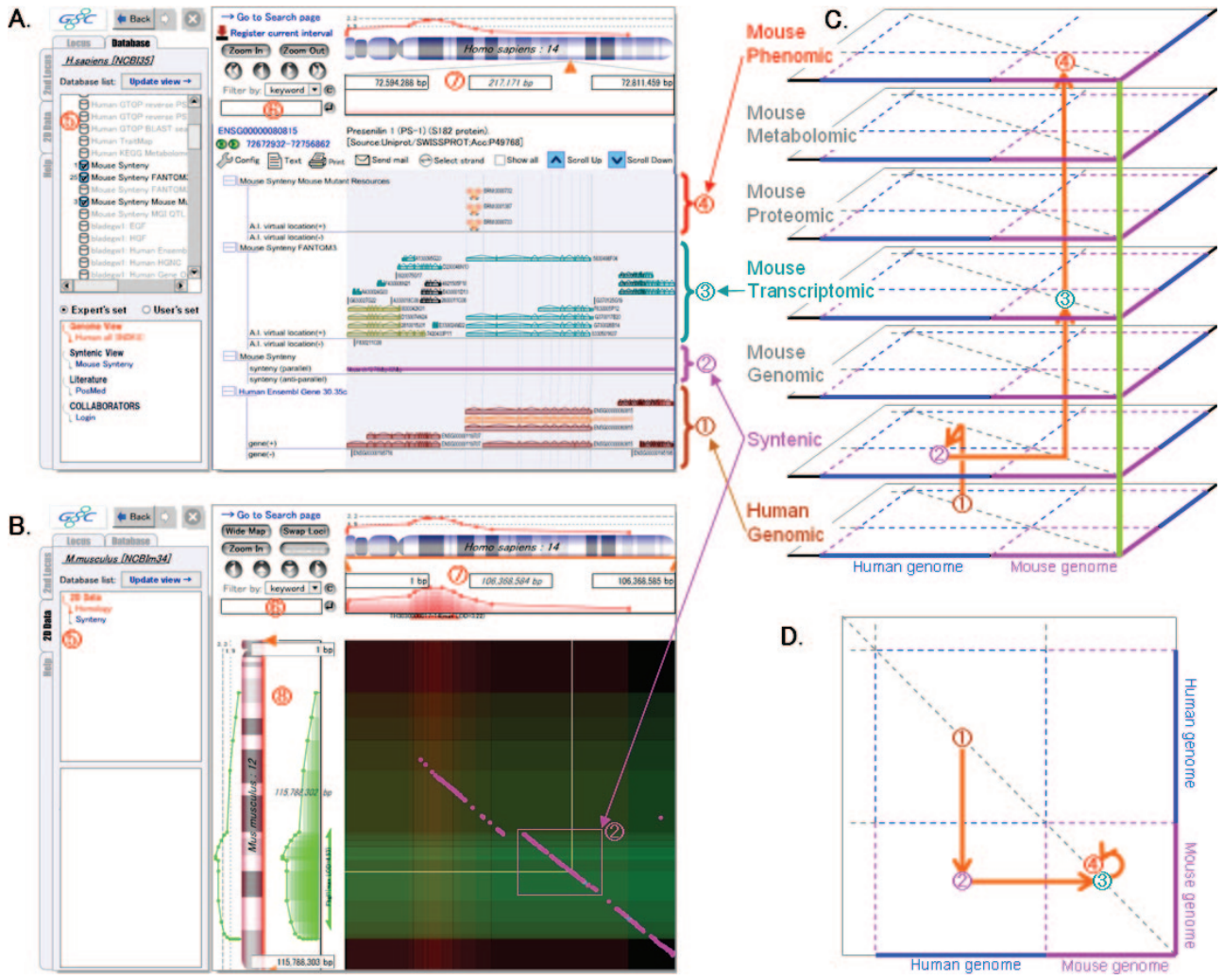

Fig. 1. (**A**) One-dimensional human-chromosomal-interval view integrating human genes (1), a syntenic region estimated based on the homologous gene pairs between human and mouse (2), mouse transcripts mapped on the mouse syntenic region corresponding to the human interval (3), and available mutant mice with knocked-out genes at the syntenic region (4), and displaying only those data records that are selected in the dataset-axis menu (5) that contain the terms specified in the textbox for the ontology-axis (6). (**B**) Two-dimensional view displaying the evolutional correspondences or homologous gene pairs as dots (2) between the horizontal human genomic axis (7) and the vertical mouse genomic axis (8). The graphs of genetically mapped trait loci are shown along the chromosomes. (**C**) Omic-space chart in which the cascades of orange arrows represent the procedures of the above-displayed data integration. The ontology axis is not drawn here, but the dataset axis (green) is drawn vertically. Each horizontal plane defined by two orthogonal axes representing both human (blue) and mouse (pink) genomes represents each dataset integrated in the space. (**D**) The omic space chart viewed from above. The data attributed to single genes of human (1) or mouse (3,4) are plotted on the diagonal line of human or mouse genomes, respectively, whereas the syntenic relationship attributed to both human and mouse homologous genes (2) is plotted so as to indicate the correspondence between the genes.

propose the omic-space charts as convenient models to express schemes for integrating multiple datasets (Fig. 1).

For positional-cloning purposes, a user can select candidate responsible genes from numerous genes existing within wide genetic loci by filtering only those whose annotations or associated documents meet the condition specified on the ontology axis. As a genome browser OmicBrowse can color-code gene annotations by text-matching or gene expression levels. OmicBrowse also assists users to share views by e-mail, to print out views, to register frequently viewed chromosomal intervals and to define groups of often browsed datasets as menu items. The utilities of OmicBrowse have also been highly evaluated through analyses of multiple genetic maps (Heida *et al*., 2004) and the analysis of whole-genome tiling arrays (Toyoda and Shinozaki, 2005). Funding to pay the Open Access publication charges for this article was provided by RIKEN.

Conflict of Interest: none declared.

REFERENCES

Dowell,R.D. *et al*. (2001) The distributed annotation system. *BMC Bioinformatics*, **2**, 7.
Gribble,S. *et al*. (2001) What can peer-to-peer do for databases, and vice versa? *WebDB*, 2001, 31–36.

Heida,N. *et al*. (2004) TraitMap: an XML-based genetic-map database combining multigenic loci and biomolecular networks. *Bioinformatics*, **20** (Suppl. 1), I152–I160.

Toyoda,T. and Shinozaki,K. (2005) Tiling array-driven elucidation of transcriptional structures based on maximum-likelihood and Markov models. *Plant J*., **43**, 611–621.

Toyoda,T. and Wada,A. (2004) Omic space: coordinate-based integration and analysis of genomic-phenomic interactions. *Bioinformatics*, **20**, 1759–1765.

Stalker,J. *et al*. (2004) The Ensembl web site: mechanics of a genome browser. *Genome Res*., **14**, 951–955.

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