

Cross Database Visualization

David P. Hansen ¹

daveh@lionbio.co.uk

Thure Etzold ¹

etzold@lionbio.co.uk

Guenter Kurapkat ²

kurapkat@lionbioscience.com

¹ LION Bioscience Ltd. Shereaton House, Castle Park, Cambridge, CB3 0AX UK

² LION Bioscience AG. Im Neuenheimer Feld 515-517, D-69120 Heidelberg, Germany

1 Introduction

One of the fundamental problems in computational biology is the diversity of data types and data sources. SRS is a widely used integration tool for these databases enabling powerful cross-database queries [1, 2, 3]. The mechanism used to enable cross-database searches can also be used to combine data from multiple entries to form a “view”.

Combining data into structured objects via a CORBA interface is described in Coupaye [4]. This poster introduces a new mechanism for creating and displaying data elements from multiple, linked databases. Specific examples demonstrate views of data including public domain along with application results.

2 Discussion

SRS views take advantage of the integration of databases to enable users to design their own view of a set of data - combining as many pieces of data from as many sources as they like. This view specifies fields from a database, and linked databases, and the format to display this data. When a query is performed against a database, only data required for the selected view is retrieved and displayed.

SRS views have recently been extended so that retrieved data may be processed and presented as an image or sorted by a certain field. This allows for much more powerful views to be created, especially with large and complex genomic databases.

We have used this capability to link genome sequences, genome annotation, and ortholog databases. SRS then enables very rapid queries to, for example, find all of the members of a specific protein family in *H. pylori* that have an ortholog in *E. coli* and *H. influenzae* and show their respective localisation on the genome. Views enable the user to not only retrieve a list of ORF's, but to see on a single screen a selection of linked information from nucleotide and protein sequence databases and annotation results - for each of the orthologs.

The new features of SRS views, along with the unique linking ability between fields in different databases and the fast retrieval time of queries, make the SRS package a powerful tool for data-mining and visualisation.

The poster will demonstrate a number of graphical views and data structures that can be created by combining data from multiple, linked databases. The emphasis will be on computationally derived databases including pairwise ortholog databases generated from sequence similarity searches.

References

- [1] Etzold, T., Ulyanov, A., and Argos, P., SRS: information retrieval system for molecular biology data banks, *Methods in Enzymology*, 266:114-128, 1996.
- [2] Etzold, T. and Argos, P., SRS: an indexing and retrieval tool for flat file data libraries, *Comput. Appl. Biosci.*, 9:49-57, 1993.

- [3] Etzold, T. and Argos, P., Transforming a set of biological flat file libraries to a fast access network, *Appl. Biosci.*, 9:59–64, 1993.
- [4] Coupaye, T., Wrapping SRS with CORBA: from textual data to distributed objects, *Bioinformatics*, 15:333–338, 1999.

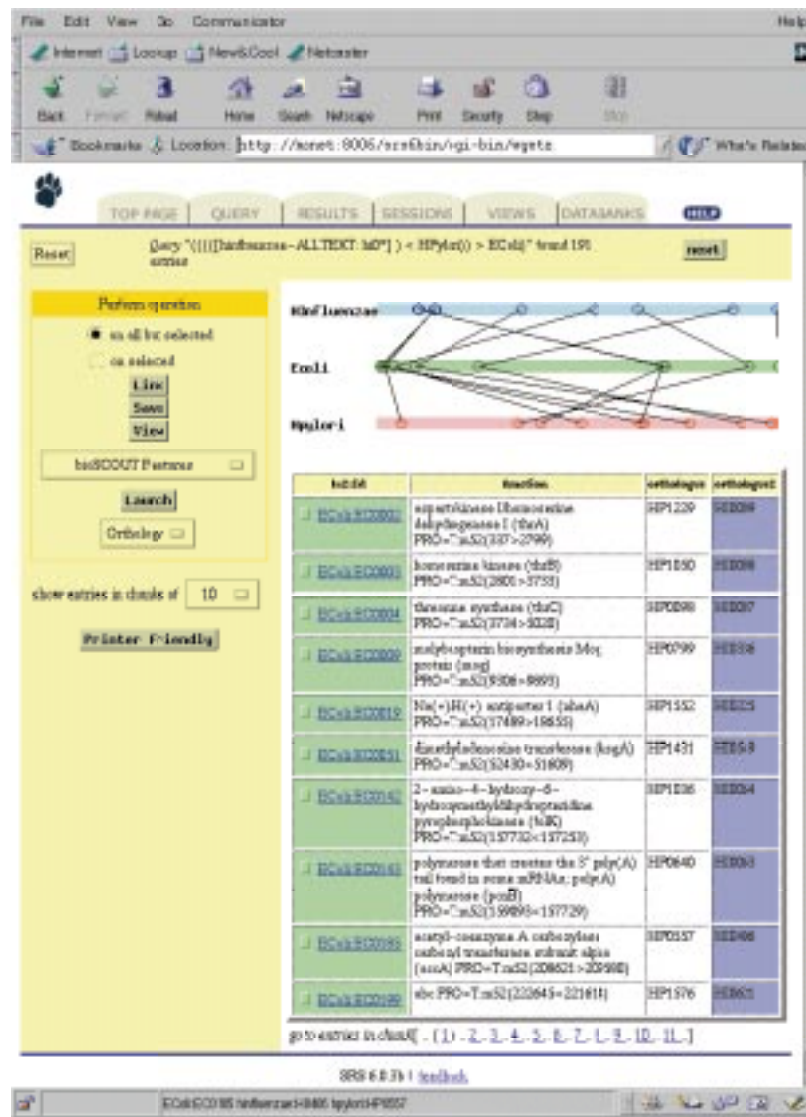


Figure 1: Orthologs from *E. coli*, *H. influenzae*, and *H. Pylori* with descriptions from the *E. coli* database and an image of the relative positions of each ORF on the respective genome with links to its orthologs.