

Finding Friends Outside the Species: Making Sense of Large Scale BLAST Results with SilverMap

Peter Ansell, Lawrence Buckingham, Xin-Yi Chua, James M. Hogan¹, Scott Mann, Paul Roe

Microsoft QUT eResearch Centre, Brisbane, AUSTRALIA

j.hogan@qut.edu.au

INTRODUCTION

Bioinformatics is increasingly dominated by comparative investigations of coding and regulatory function across genome and species boundaries. Among the most important tasks in this enterprise is the discovery of similarity between fragments from different sequences, and this is normally accomplished through the use of BLAST, the well-known alignment search tool provided by NCBI [1]. Regardless of the tool employed, the *n vs. n-1* comparisons characteristic of such studies may produce literally thousands of numeric similarity scores and candidate alignments, making comprehension of all but the most clear-cut results time consuming and error-prone, and discovery of more subtle relationships all but impossible. Some visualisation options are available, but the traditional, alignment-oriented views are severely limited in their capacity to handle complex relationships, and the approach plainly cannot scale to accommodate the datasets now routinely being processed.

SilverMap is a web-based tool for visualising the results of comparative studies in genomics and proteomics, showing the relationships between items of interest based on the distance between them - essentially the inverse of their similarity - as though the genes or proteins are cities on a map. In this way, researchers may discover and explore relationships through the familiar visual metaphor of maps and navigation.

The system presents a radar-style view which displays query results according to similarity to a centrally located primary feature. Animation facilities allow an almost instant re-organisation of the map around another feature of interest, enabling rapid clustering by eye, and exploration of groupings based on distance threshold. Semantic information is provided through integration with the Bio2RDF project [2], and our group provides a partial mirror of this system. User tagging of genes and relationships is permitted, with naming and authentication provided through OpenID.

AVAILABILITY AND USAGE

The present beta release of SilverMap is freely available for public access at [4]. The system executes in the Microsoft Silverlight presentation environment [5], and runs on all commonly available browsers. Figure 1 (below) shows SilverMap running in the Google Chrome web browser. The tool currently allows researchers to surf a database containing approximately 51 million BLAST results computed from a collection containing a representative sample of approximately 180,000 proteins: ~60,000 proteins from the human genome; ~50,000 from the house mouse *Mus musculus*; with the remaining ~70,000 drawn from the genomes of 14 strains of *E. coli*.

Alternative entry points have been created to give users a quick start with SilverMap by accessing preconfigured queries. A query over mammalian data may be accessed directly via hyperlink [6], while hyperlink [7] provides access to a query over bacterial data. Upon opening the hyperlink, the user's browser will display SilverMap and query results will be loaded automatically. On loading the link at [7], for example, the display is centred on an *E. coli* gene (b3067) strongly implicated in regulation of standard or housekeeping cellular processes. Other nodes are positioned in the map according to their distance from b3067, a distance calculated by exponentiating and normalising the BLAST e-value. The red circle defines a radius threshold for grouping purposes, and the *Select* and *Deselect* buttons at bottom right toggle selection (indicated by green colouring) of the nodes within this boundary. The text box appears after clicking on the label for any of the nodes. The display shows publicly available data and a BLAST alignment between the central gene and the node of interest.

Links at top left allow users to specify target and comparison organisms and gene lists for load; a spreadsheet-like view of the details of all genes shown on the map; and a snapshot tool that supports interactive scientific collaboration and publication. This works in two ways: an HTML `<div>` component which encapsulates the current view may be pasted into a wiki or html page; alternatively, a HTML query string (expressed as a URL) may be copied and sent directly to a collaborator who is then able to navigate immediately to the specified view. These facilities allow researchers to record, publish and share views into the data. Collaborators may then begin a browsing session by opening this shared view, perusing the query results, selected items and tags, and subsequently explore the dataset dynamically.

¹ Author to whom correspondence should be addressed. The authors would like to acknowledge support from Microsoft Research (Redmond) and the Queensland Government under the Smart State programme.

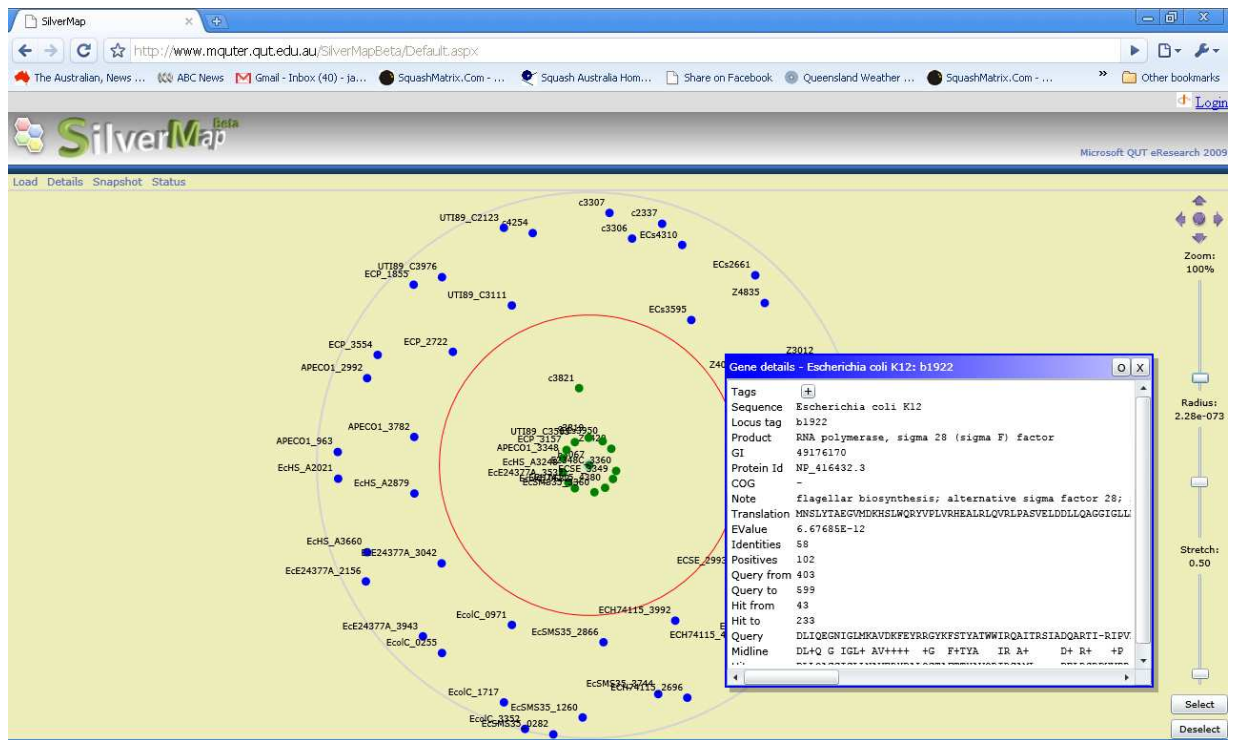


Figure 1: SilverMap showing relationships among a subset of 70000 *E. Coli* genes.

ONGOING WORK

With SilverMap it has become feasible for researchers to query and surf very large collections of precomputed BLAST results. The present beta version is not tightly coupled to its underlying data source, making it easy to plug in any data source of a similar nature. It is envisaged that SilverMap will develop as a standard addition to large datasets of this type, both for laboratories active in sequencing, and for bioinformatics researchers and service providers wishing to offer pre-computed BLAST results. A large scale deployment of this type is presently being examined.

While SilverMap has been created primarily for comparative genomics, the tool is architected to support visualisation for any collection of data on which a suitable similarity measure may be defined, and the emerging SilverMap API allows users and contributors to implement a measure tailored to their specific problem. For example, our group is presently exploring gene relationships based on common occurrences in publications, and node separation in defined bioinformatic ontologies such as GO [3].

In addition, several specific enhancements are presently in development or under consideration. The highly intuitive visual navigation and query experience will be augmented by exploiting biological cross-reference systems such as Bio2RDF [2] to provide further query refinement and filtering. A mechanism will be put in place to allow users to customise and extend the tool by defining custom similarity measures and filters that can be incorporated in the system. A desktop executable version of SilverMap that leverages Microsoft Windows Presentation Foundation [8] is under active development at the time of writing. This will be released together with an Open Source software library that will enable bioinformaticians and developers to use SilverMap in their programs. Recognising that many important results cannot be obtained from existing collections of precomputed BLAST results, a free-standing implementation is under development to provide a visual interactive BLAST query interface and result explorer to complement present-day web-based BLAST services.

REFERENCES

1. *Basic Local Alignment Search Tool (BLAST)*. Available from <http://blast.ncbi.nlm.nih.gov/>, accessed June 2009.
2. Belleau, F. et al., *Bio2RDF: Towards a mashup to build bioinformatics knowledge systems*. Journal of Biomedical Informatics, 2008, **41**: pp 706-716.
3. The Gene Ontology Project (GO). Available from <http://www.geneontology.org/>, accessed June 2009.
4. Microsoft QUT e-Research Centre, <http://www.mquter.qut.edu.au>, accessed June 2009.
5. Microsoft Silverlight. Available from <http://silverlight.net/>, accessed June 2009.
6. SilverMap mammalian example, <http://www.mquter.qut.edu.au/SilverMapBeta/MammalianExample.aspx>, accessed June 2009
7. SilverMap prokaryote example, <http://www.mquter.qut.edu.au/SilverMapBeta/Default.aspx>, accessed June 2009.
8. Windows Presentation Foundation, <http://windowsclient.net/>, accessed June 2009.