

Bringing TRNDiff to the Cloud: Supporting large scale visualisation of transcriptional regulatory networks for biological researchers

Xin-Yi Chua¹, Lawrence Buckingham², James M. Hogan³

1. University of Queensland, Brisbane, Australia, x.chua2@uq.edu.au

2. Queensland University of Technology, Brisbane, Australia, l.buckingham@qut.edu.au

3. Queensland University of Technology, Australia, j.hogan@qut.edu.au

INTRODUCTION

We describe the outcomes of a QCIF-funded infrastructure project under which an established bioinformatic analysis tool – TRNDiff [1], described below – has been migrated to the QCloud platform to enable private off-the-shelf deployment of the toolset by bioinformatics researchers without requiring them to fund IT support and software development.

In the current project, TRNDiff has been comprehensively re-engineered to achieve compatibility with the Lawrence Berkeley Laboratories RegPrecise data set [2-4] and to provide improved mechanisms for preparation and maintenance of local datasets. In addition to these improvements, infrastructure has been put in place to support the integration of TRNDiff with other standard bioinformatics tools such as NCBI BLAST [5] and sequence comparison tools such as MEME [6], BioPatML [7] and MUMmer [8].

SYSTEM OVERVIEW

TRNDiff [1] is a visual analysis tool which enables bioinformatics and system biology researchers to analyse the evolution and conservation of gene regulatory networks across groups of related bacterial species. TRNDiff has been used by microbiology researchers at QUT and has generated significant interest from researchers in the United States. The tool employs a novel wagon-wheel visualisation (see Figure 1) to provide a highly effective mechanism for biologists to visually explore the evolutionary relationships between gene regulatory networks across species.

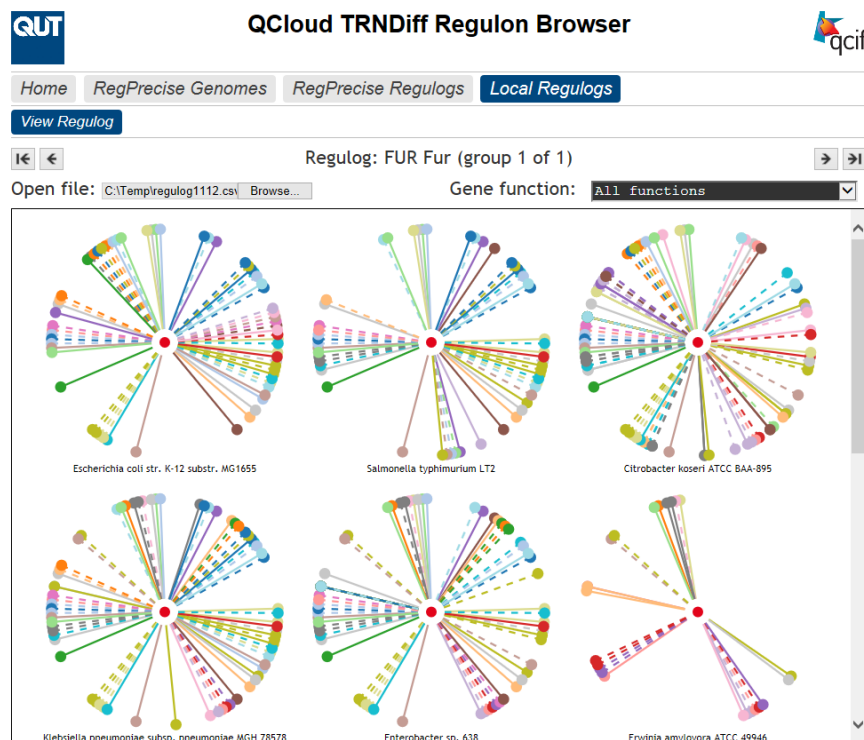


Figure 1: Ferric ion uptake regulation (FUR) regulog viewed in TRNDiff/QCloud

TRNDiff/QCloud has been adapted to source data from the RegPrecise database [2-4], a repository of high-quality reference data which captures the regulatory relationships between genes in 390 prokaryote genomes. The database contains manually curated details of genes, transcription factors and transcription factor binding sites which have been obtained by careful comparative genomic analysis, experimentation and literature review [3]. Integration with RegPrecise simplifies the tasks involved in the analysis of regulatory network conservation to help researchers efficiently identify significant factors which explain or predict variation in the properties and behaviour of bacteria.

Regulon collections may be exported to a tabular format and saved as CSV-formatted spreadsheet documents which may be retained as-is to provide a record of an investigation or edited to define new collections which can then be opened in TRNDiff for further exploration. Users may add custom annotations in the form of additional columns in the tables; when the dataset is opened these annotations become available for viewing within the tool.

IMPLEMENTATION OVERVIEW

The system is implemented as a client-server web application with a HTML5/JavaScript user interface which targets modern web browsers running on desktop machines and tablets. The server presence is provided by a compact, self-contained web server which runs under Node.js [9] and is capable of deployment into environments where a web server such as Apache or Internet Information Services is not installed. TRNDiff can be installed on a web server as a traditional web application or copied to a user's desktop computer and executed as a local desktop application. The latter scenario is supported by inclusion of a stand-alone copy of the Node.js executable as part of the distribution. The use of JavaScript to implement both client and server permits substantial sharing of code between client and server.

CONCLUSIONS

In conclusion, the QCloud deployment of TRNDiff provides a robust and carefully architected and engineered system for interactive exploration of novel and curated data describing bacterial and other regulatory systems. The tool allows configurable access to local data sources and persistence of data of interest. The underlying server readily supports QCloud VM and local deployment, while providing a firm basis for an integrated environment for bioinformatics visualisation. The system has benefitted from user engagement and interaction, and from extensive automated system testing. The system remains an active focus for further development through student projects and contributions from a developing open source community at QUT, and this work will be ongoing from the beginning of 2014.

REFERENCES

1. Xin-Yi Chua (2012). *"Prediction of transcriptional regulatory interactions in bacteria: a comparative genomics approach."* PhD thesis, Queensland University of Technology.
2. RegPrecise. <http://regprecise.lbl.gov/RegPrecise/>.
3. Pavel S. Novichkov, Olga N. Laikova, Elena S. Novichkova, Mikhail S. Gelfand, Adam P. Arkin, Inna Dubchak, and Dmitry A. Rodionov (2010) "RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes." *Nucleic Acids Res.* 2010 January; 38(Database issue): D111–D118.
4. P.S. Novichkov, T.S. Brettin, E.S. Novichkova, P.S. Dehal, A.P. Arkin, I. Dubchak, D.A. Rodionov (2012). "RegPrecise web services interface: programmatic access to the transcriptional regulatory interactions in bacteria reconstructed by comparative genomics." *Nucleic Acids Res.* 2012 Jul; 40(Web Server issue):W604-8.
5. Stephen Altschul, Warren Gish, Webb Miller; Eugene Myers, David Lipman (1990). "Basic local alignment search tool." *Journal of Molecular Biology* 215 (3): 403–410.
6. Michael W. Towsey, James M. Hogan, Sarah A. Mathews, Peter Timms. (2007) "The In Silico Prediction of Promoters in Bacterial Genomes." In Ng, See-Kiong, Mamitsuka, Hiroshi, & Wong, Limsoon (Eds.) *Genome Informatics Workshop 2007, 3rd - 5th December 2007, Singapore.*
7. Timothy L. Bailey and Charles Elkan (1994). "Fitting a mixture model by expectation maximization to discover motifs in biopolymers." *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, pp. 28-36, AAAI Press, Menlo Park, California.
8. Stefan Kurtz, Adam Phillippy, Arthur L Delcher, Michael Smoot, Martin Shumway, Corina Antonescu and Steven L Salzberg. (2004) "Versatile and open software for comparing large genomes." *Genome Biology* 5:R12.
9. Node.js. <http://nodejs.org/>.

ABOUT THE AUTHOR(S)

For each author, please include a short bio of 150-200 words. This information can extend onto a third page if needed. The information about the topic should take up no more than two pages.