Procedia Computer Science

Volume 51, 2015, Pages 713–724

ICCS 2015 International Conference On Computational Science

Comparative Visualisation of Regulatory Networks with TRNDiff

Xinyi Chua[1], Lawrence Buckingham[2], James M. Hogan[2,*] and Pavel Novichkov[3]

[1]*QFAB Bioinformatics, Institute for Molecular Biosciences, UQ, Brisbane, Australia*
[2]*School of EECS, QUT, Brisbane, Australia*
[3]*Lawrence Berkeley Laboratories, Walnut Creek, CA., USA.*

xinyi.chua@qfab.org, l.buckingham@qut.edu.au, j.hogan@qut.edu.au, psnovichkov@lbl.gov

Abstract

The advent of Next Generation Sequencing (NGS) technologies has seen explosive growth in genomic datasets, and dense coverage of related organisms, supporting study of subtle, strain-specific variations as a determinant of function. Such data collections present fresh and complex challenges for bioinformatics, those of comparing models of complex relationships across hundreds and even thousands of sequences. Transcriptional Regulatory Network (TRN) structures document the influence of regulatory proteins called Transcription Factors (TFs) on associated Target Genes (TGs). TRNs are routinely inferred from model systems or iterative search, and analysis at these scales requires simultaneous displays of multiple networks well beyond those of existing network visualisation tools [1]. In this paper we describe TRNDiff, an open source system supporting the comparative analysis and visualization of TRNs (and similarly structured data) from many genomes, allowing rapid identification of functional variations within species. The approach is demonstrated through a small scale multiple TRN analysis of the Fur iron-uptake system of *Yersinia*, suggesting a number of candidate virulence factors; and through a larger study exploiting integration with the RegPrecise database (http://regprecise.lbl.gov; [2]) - a collection of hundreds of manually curated and predicted transcription factor regulons drawn from across the entire spectrum of prokaryotic organisms.

Availability: TRNDiff is presently available in two versions, as a single client js script allowing browser based visualization of data files (here denoted V1) and as a scalable Node.js application with web service integration to RegPrecise (The QCloud Regulon Browser or V2). Information on TRNDiff ma be found at the dedicated site http://trndiff.org, which includes example data, short tutorials and links to demonstration installations of both the stand-alone system and the integrated regulon browser. Additional views from the standalone client, and support for clustering and touch-based interactivity are being incorporated into the Node.js version. Source code for both versions is freely available under a non-restrictive Apache 2.0 licence from the authors' repository at http://bitbucket.org/biovisml, and contributions and feature requests are encouraged.

Keywords: Bioinformatics, Visualisation, Transcription, Regulatory Networks

[*] Corresponding author. Support from Microsoft Research (Redmond) and NecTAR (http://nectar.org.au/) is gratefully acknowledged.

doi:10.1016/j.procs.2015.05.189

1 Introduction

Graph-based representations of biological systems have a long and successful history. Within the narrower domain of regulatory relationships, applications have included the analysis of a genome's regulatory topology; identification of global regulators; and uncovering sub-network motifs and studying cross-species regulatory evolution [3-8]. The majority of these studies can be supported by existing network visualisation tools based on the display of a single network graph, but the bulk of these systems do not readily support comparative analysis of a large collection of such networks. Our concern here is not with the complete description of a single complex network, but rather with the complexity which emerges as a very large number of simpler networks are considered together.

Transcriptional Regulatory Networks (TRNs) provide a compact description of the relationships between one or more regulatory proteins called transcription factors (TFs) and the set of target genes (TGs) whose expression they control. TRNDiff was introduced to address the challenge of exploring variations in the action of particular TFs across large numbers of organisms – whether closely or distantly related – which share a common regulatory apparatus: similar TFs, mostly similar TGs, but variations in wiring that may have considerable functional significance. As the number of genomes increases, so too does the need for multiple network displays. Yet scale brings with it constraints on the individual network. Complex displays, encoding all the information available about a particular set of relationships, are less suitable for large scale comparative analysis, and some compromise is required. A simplified representation which effectively presents a subset of the network properties is potentially more valuable than one whose scope is more ambitious.

The TRNDiff system addresses these challenges through consideration of a Transcription Factor and a universal set of regulated Target Genes (and their orthologues) drawn from across the organism set, representing these relationships through the *wagon wheel* diagram explained below. A reference organism – perhaps constructed through aggregation from actual genomes, but usually one chosen as a basis for computational inference - shows the possible associations; individual variations are seen in contrast to this superset. Positions of orthologous genes (nodes) in a particular genome are placed to correspond to those of the reference genes from the reference genome. Edges denote the presence or absence of a Transcription Factor Binding Site (TFBS) prediction. Using these visual aids, the user can rapidly assess regions of similarity or difference among the multiple structures in a study. Analysis is also supported through features such as set operations and user-defined groupings. TRNDiff is intended primarily for the comparative study of TRNs, but may readily be applied to related network visualisation tasks, showing for example the relationships in protein-protein interaction data, or constituent matches across metagenomic data sets.

The use of numerous concurrent displays with careful alignment of node positions allows for ready comprehension of static TRN relationships. While these snapshots are an essential tool, transcriptional networks are necessarily dynamic, adjusting in response to environmental conditions – often through the action of additional transcription factors – and undergoing more fundamental topological changes over evolutionary timescales through one or more *re-wiring events*. These ideas are well explored in [9]. The wagon wheel diagrams characteristic of TRNDiff (see Fig 1) provide a natural frame for sequential animation, and this application is limited only by the availability of data. If the network is modelled for different environmental conditions, it is then a straightforward matter to construct the 'movie', although this is not yet automated. At present, our primary concern is effective comparison with increasing scale.

This paper is organized as follows. In section 2, we consider the structure of the TRNDiff diagrams and data sources in more detail, introducing in section 2.1 the V2 Node.js implementation which supports the far richer, curated data sources of the RegPrecise project. Section 3 describes two case studies – one at modest scale, illustrating the canonical usage of the system, while the other draws upon the RegPrecise dataset for the hexR regulon. In Section 4, we consider variations on the

approach, evaluation and the development of touch and very large scale deployments of the system, before concluding in section 5.

2 The TRNDiff System

We begin this section with an overview of the structure of the wagon wheel diagrams, and of the operations and general data requirements of the system. More specific discussion is given in the context of the case studies considered below, but in general terms the tool requires a simple grammar describing potential node relationships, with a single real value indicating the strength of each link. While this paper examines a number of possible applications of TRNDiff, the name itself suggests the comparison of transcriptional regulatory systems, and our discussion will focus on this problem. The approach is best understood if one assumes a well-studied model system, such as the Fur regulon in *Escherischia coli* K12, for which the set of target genes for the transcription factor has been well established through experiment. We are then concerned with the comparable regulon in other genomes, where relationships are inferred from orthology of transcription factor and target genes in the new system with those of the model, usually via the *Regulog* approach [10, 11] or some related method. Many such cases are available via integration with the RegPrecise data sources [2].

TRNDiff follows conventional regulatory diagrams in adopting a circular structure (see Fig 1) with nodes representing transcription factors and target genes. Each single 'wagon wheel' is assumed to represent a regulon governed by a single transcription factor; this single TF node is displayed at the hub of the wheel, while the associated TGs reside on the circumference, eliminating crossover of edges. It is common to visualise regulatory interactions through a directed edge between the transcription factor and the associated target gene: $TF \rightarrow TG$. Here the presence of a TG node on the wheel surface indicates that there is some evidence (usually via Regulog) that a relationship might exist; a relationship confirmed – through experiment or by computational detection of a binding site – is then represented by a solid spoke in the wheel, where the direction of the relationship is implied by the structure. Extensions of the approach to more complex systems exhibiting control by multiple TFs is in principle straightforward, but a careful balance is required: the focus on individual regulatory systems is intended as a deliberate counterpoint to spaghetti network diagrams.

Usage of the diagrams depends on the nature of the input data, but the existence of a link reflects some confidence in the relationship – based on the statistical significance of a computational binding site prediction or experimentally measured expression level. The canonical usage of TRNDiff is then to compare these 'incomplete' networks with the model, and to identify structural variation – especially that consistent across a group, such as a set of known pathogens – which may be functionally significant. This is accomplished simply, through the use of side by side, overlay and multiple network displays (best seen in the Fur case study of Fig 4 and 5) and through more complex analyses involving Boolean operations (pairwise intersection, union and XOR) (Fig 2; currently limited to V1) and (shortly) a histogram display.

At an operational level, the tool allows analysis at multiple scales, from the forest of network diagrams of Fig 4 and Fig 6 through to the pair-wise diagrams of Fig 2. In this way, trends may emerge at the group level, such as in the pathogenic vs non-pathogenic strains of the first case study, leading to targeted investigation of smaller sets of systems. Gene and transcription factor metadata is presented in response to a mouse-over event. Specific functional annotations can be highlighted from a drop-down menu: Fig 3 shows TGs from a small group of hexR systems implicated in glycolysis, and similar functionality may be observed in Fig 4 for Ybt (see section 3.1).

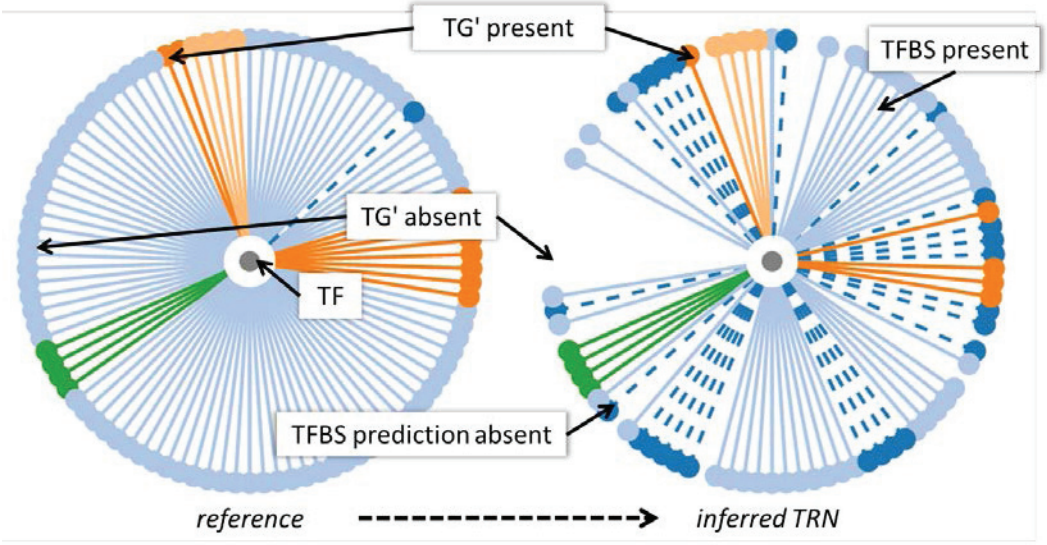

Figure 1: TRNDiff (V2) display showing the reference transcriptional regulatory network on the left and the inferred regulatory network on the right. The transcription factor resides in the centre of each graph and is surrounded by the regulated target genes. A solid line in the inferred network denotes an identified transcription factor binding site, a dotted line its absence.

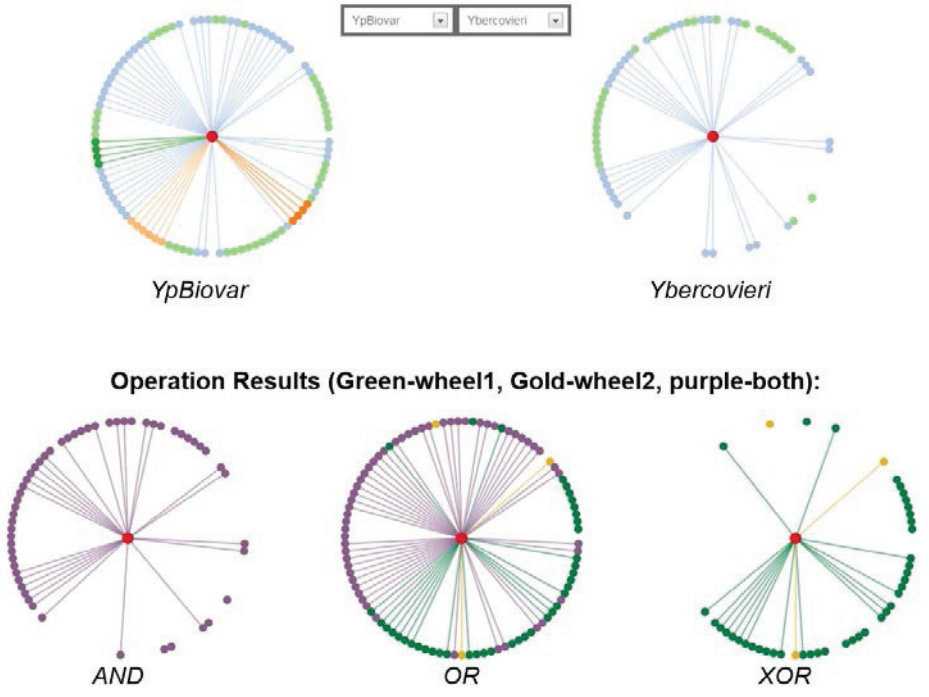

Figure 2: Direct Boolean comparison of two *Yersinia* strains, showing AND, OR and XOR operations (V1).

2.1 Integration with RegPrecise

TRNDiff uses a novel wagon-wheel visualisation to provide a highly effective mechanism for biologists to visually explore the evolutionary relationships between gene regulatory networks across species. However, in order to use V1 and earlier versions of the tool, a researcher must create an input document which uses JavaScript Object Notation (JSON) to express the hierarchical relationship between organism, transcription factors and the genes regulated by each transcription factor. While the JSON format employed in the original is moderately convenient as a machine-readable serialization mechanism, the notation is inconvenient and prone to error when used by non-programmers without access to a JavaScript-aware text editor. A further obstacle confronting would-be users of the tool is the need to either manually or programmatically transform data from other sources to the prescribed TRNDiff format prior to use.

Recently, TRNDiff has been comprehensively re-engineered to achieve compatibility with the Lawrence Berkeley Laboratories RegPrecise data set [2] 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 [12] and sequence comparison tools such as MEME [13], BioPatML [14] and MUMmer [15]. 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 [16] 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.

We now consider the case studies chosen to illustrate the capabilities of TRNDiff. The first is a novel analysis of the Fur iron uptake system in *Yersinia*, part of a more extensive study of this regulon across a number of species [17]. This is intended to show the core visualisations of TRNDiff and to illustrate their utility in identifying candidate virulence factors. The second of these studies demonstrates the integration of the system with RegPrecise [2] as described above.

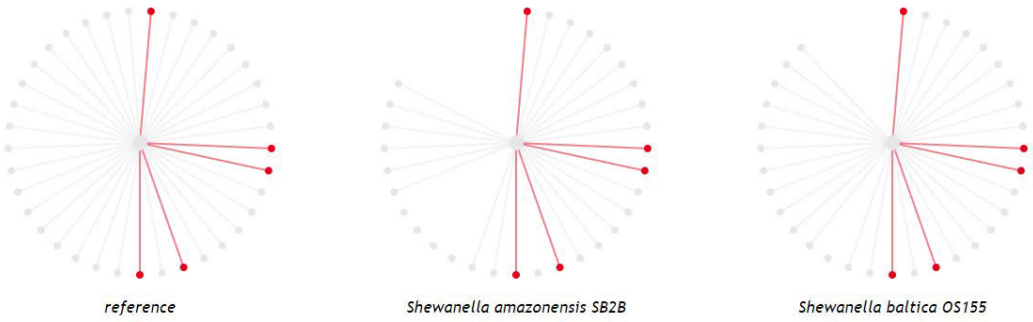

Figure 3: HexR systems highlighting genes implicated in glycolysis (V1). Annotations in V2 are sourced automatically from the RegPrecise database (see Fig 4A).

3 Case Studies

3.1 Determining putative virulence factors in the *Yersinia* Fur regulon

Regulation of iron uptake is a critical system in bacteria: iron starvation is a classical immune response from an infected host; yet high iron levels are also toxic. Successful pathogens have developed a range of iron-uptake systems to assimilate iron from different environmental conditions. For instance, *Yersinia pestis*, which is the causative agent of plague, has at least eight known iron-uptake systems: Ybt, Yfe, Yfu, Hmu, Has, Yiu, Iuc and Fhu. Two of these, the Ybt and Yfe systems, have been identified as prerequisites for full virulence in bubonic plague [18, 19]. Many studies have focused on the three human pathogens, overlooking the eight pathogenic to other species [20].

Iron uptake systems regulated by the Fur transcription factor were here investigated across 20 *Yersinia* genomes from 11 species, comprising 12 pathogenic and 8 non-pathogenic strains. The experimentally determined *Yersinia pestis* CO92 Fur regulon, established through microarray and electrophoretic mobility shift assays (EMSA) [18], was taken as the reference TRN. The *Y. pestis* CO92 network was used to infer putative relationships using the classical regulog method [10, 11]. Orthology was established through a bi-directional best hit (BDBH) approach using BLAST [12] searches with an e-value $< 1.0\times10-5$ and a coverage of 65%[†]. Operon structures were predicted based on a combination of the intergenic distances – similar to the approach used in [21,22] – and confidence scores from OperonDB [23], where the score indicates the probability that two adjacent genes belong to the same operon. Network inference is completed through de novo search using MEME [13] upstream of the first gene in the identified operons containing the target gene orthologue, accepting the most statistically significant predictions consistent with TFBS search parameters as confirmation of the binding site. In the absence of a significant binding site prediction, the relationship is held to be unconfirmed and a dotted link is shown.

Substantial variations between pathogens (Fig 4A) and non-pathogens (Fig 4B) are immediately observable in the TRNs inferred from the model CO92 system. Even within these groups, slight variations may be observed, suggesting detailed investigations of the role of Ybt (unexpectedly absent in three of the pathogenic species), Yiu (arguably a candidate virulence determinant), and of the yfeABCD operon in the non-pathogenic strains[‡].

Overall, we conclude that the pathogens show substantially greater conservation with respect to the (pathogenic) model network CO92 than the non-pathogenic set. While the non-pathogen TRNs appear noticeably less connected, many interactions are conserved, suggesting these regulatory interactions may be involved in the basic survival of these organism, and are unlikely to be implicated as a determinant of virulence. The more limited connectivity of the non-pathogenic strain TRNs also suggests an alternative set of iron-uptake systems, which can be confirmed by a genome wide scan for putative Fur binding sites. Such hypotheses arose directly from consideration of TRNDiff diagrams, and would not easily have emerged from more conventional network views.

[†] Given a query gene, q, from genome X that has a best scoring hit, p, from genome Y : if the best hit for p in genome X is q, then (q; p) is a BDBH pair for the genomes X; Y .

[‡] See [17] for a detailed discussion of these issues.

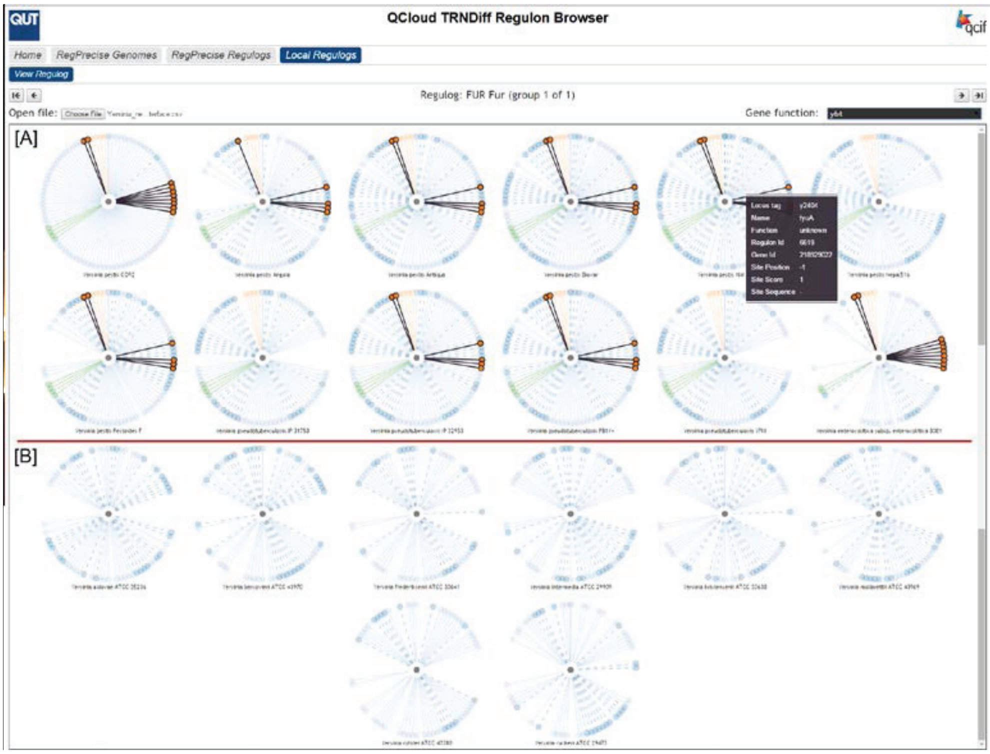

Figure 4: The reconstructed Fur regulon across 20 Yersinia strains over 11 species. The experimentally verified Fur regulon from *Y. pestis* CO92 [13] was used as the model network. [A] There were 11 pathogenic strains in three species and [B] 8 non-pathogenic strains. At this resolution, we may only observe the general differences in connectivity pattern, but nevertheless, the distinction across groupings is clear.

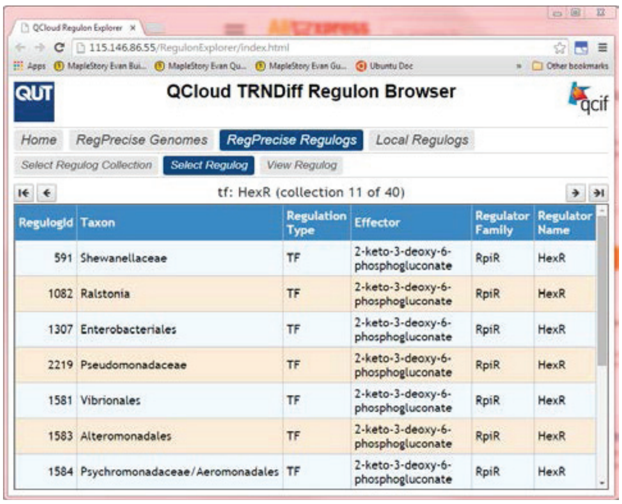
QCloud TRNDiff Regulon Browser
Home
RegPrecise Genomes
RegPrecise Regulogs
Local Regulogs
Select Regulog Collection
Select Regulog
View Regulog
tf: HexR (collection 11 of 40)
RegulogId
Taxon
Regulation Type
Effector
Regulator Family
Regulator Name
591 Shewanellaceae TF 2-keto-3-deoxy-6-phosphogluconate RpiR HexR
1082 Ralstonia TF 2-keto-3-deoxy-6-phosphogluconate RpiR HexR
1307 Enterobacteriales TF 2-keto-3-deoxy-6-phosphogluconate RpiR HexR
2219 Pseudomonadaceae TF 2-keto-3-deoxy-6-phosphogluconate RpiR HexR
1581 Vibrionales TF 2-keto-3-deoxy-6-phosphogluconate RpiR HexR
1583 Alteromonadales TF 2-keto-3-deoxy-6-phosphogluconate RpiR HexR
1584 Psychromonadaceae/Aeromonadales TF 2-keto-3-deoxy-6-phosphogluconate RpiR HexR

Figure 5: The QCloud Regulon Browser (TRNDiff V2) showing regulon selection according to taxonomic group. Note the alternatives of browsing according to genome and loading from an external file.

3.2 Analysis of hexR Systems

This second case study, based on the HexR regulon [24] across multiple species, is intended to illustrate the TRNDiff V2 integration with RegPrecise, and to set the scene for the discussions of sections 4 and 5. Fig 5 shows the TRNdiff V2 web interface for browsing and selecting regulon collections, organized on a taxonomic basis. As can be seen from the highlighted buttons, we are here selecting from the RegPrecise regulog collections. The alternative of browsing and selecting according to genome is also available, as is the option of local file display, the approach used to create Fig 4. Fig 6 shows the wagon wheel display of HexR regulons for 16 Shewanella species. Once again, we are able immediately to examine substantial variations such as the loss of the respiratory chain function in *Shewanella denitrificans* OS217 and *Shewanella woodyi* ATCC 51908 (obscured). Although orthologues of genes involved in glycine cleavage were identified in both species, there appears to be a loss of direct transcriptional regulation, a functional re-wiring which is not apparent from examining phylogenetic similarity. A similar analysis for the metJ system over some 192 species [25] (not included in this paper for reasons of clarity) shows that the system allows the researcher to partition the strains into distinct groupings according to function, as identified through their regulatory systems.

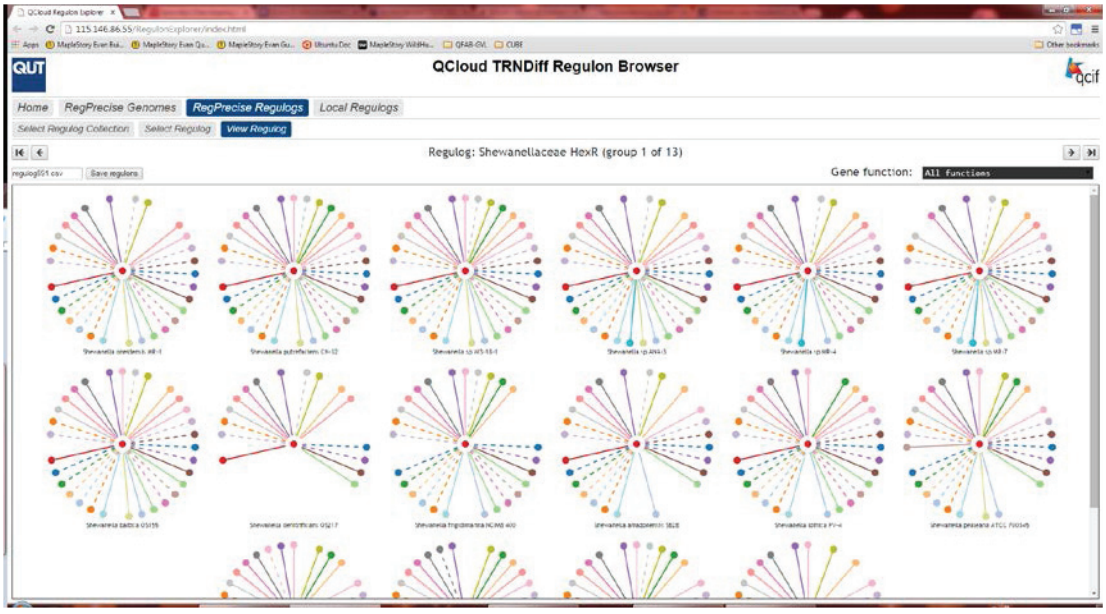

Figure 6: The HexR regulog across the Shewanellaceae. Note in particular the deletions immediately apparent in *Shewanella denitrificans* OS217 (second from left, row 2).

As can be seen from the regulon browser of Fig 5, the user now has the capacity to examine these diagrams for any of numerous taxonomic groups for which data is available, and to identify the resulting similarities and differences as they occur within and across taxonomic groups. However, the richness of the data sources available, and developments in the scale and interactivity of display hardware, offer very profitable avenues for investigation and development, and these are considered in the following section.

4 Evaluations and Enhancements

TRNDiff was initially developed as a browser-based system implemented in the Microsoft Silverlight environment, and subsequently ported to Javascript to enable broader uptake of the system. A number of features discussed earlier – notably the use of pairwise Boolean operations, and display of annotations – were included with this revision and are gradually being migrated to the integrated V2 system. Two additional approaches have been explored: regulatory tree displays, in which phylogenetic relationships are organized according to the Hamming distance between each wagon wheel[§] (Fig 7); and animations of network state through wagon wheels based on observations under different experimental condition (Fig 8). These are explained briefly in the associated captions.

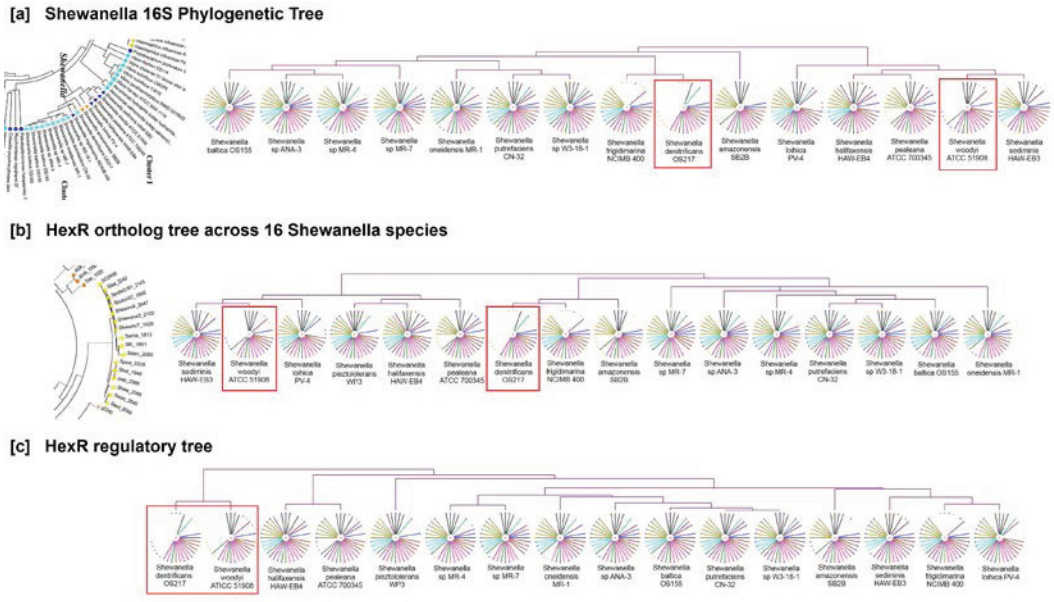

Figure 7: Comparison of the phylogenetic and regulatory trees for the HexR regulons in 16 *Shewanella* species using the original TRNDiff system. Variations correspond to those discussed above but use different node ordering. [A] Phylogenetic tree generated using 16S rDNA from [24]; [B] shows the phylogenetic tree based on the HexR orthology and [C] shows the HexR regulatory tree generated based on the degree of shared interactions present in the regulatory networks. Deviations from the standard phylogenetic structure may suggest rewiring events.

The system is currently under evaluation as part of an ongoing PhD project through open ended interview based assessments with students and active researchers in molecular biology. While these results are preliminary, there is a strong interest in greater interactivity, in particular through the ability to select actively from the displays and to group sets of wagon wheels according to observations. A touch based version of the system implementing this functionality, together with the option of guided automatic grouping based on Hamming similarity, is under active development and will be trialed in the first half of 2015. The other main limitation is imposed by the display hardware available to most researchers. We are presently developing a version of this system to run on QUT's Cube visualization facility (www.thecube.qut.edu.au) a very large facility comparable to an IMAX in size, and coupled

[§] Each wagon wheel is here represented as a high dimensional binary vector, with bit values of one for each realised connection, and zero otherwise. Hamming distances then take the place of scores resulting from sequence comparison.

with large multitouch screens to support interactivity. Experiments in a student project indicate that commercial browsers such as Chrome can support in excess of 800 wagon wheels of this type without significant degradation of performance, and it is expected that we will be able to trial this system at very large scale in the coming months.

5 Conclusion

In this paper we have introduced a novel, modular approach to the visualisation of transcriptional regulatory networks, where the focus lies on the direct comparison of individual regulons across populations. The approach is carefully selected to address the problem of very large scale comparison of regulatory systems across organisms. The approach is general and supports the visualisation of datasets with similar characteristics, such as those from protein-protein interactions and some types of metagenomic data. In the present work, we have shown its effectiveness in identifying candidate virulence factors in the Fur regulon of the *Yersiniae*, and in elucidating the regulatory evolution of hexR (and metJ) in proteobacteria. Integration with the RegPrecise respository provides a compelling use case for the system, and datasets of sufficient richness and scale as to justify very large scale investigations. The system is presently being evaluated using structured task based open ended interviews, allowing active biologists to assess its usability and our project team to measure their effectiveness in uncovering relationships within the data. Interactive and display wall versions of the system will be developed over the course of 2015.

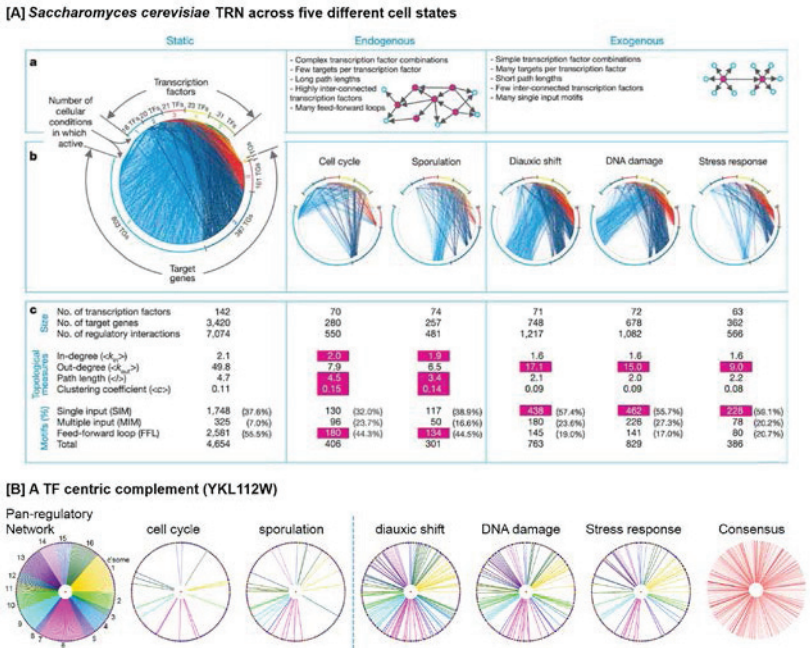

Figure 8: [A] The dynamic genome-scale transcriptional regulatory network of *Saccharomyces cerevisiae* [9]. Edges are coloured, based on the number of different conditions in which the transcription factor is found active. [B] A TRNDiff view providing a TF-centre complement of the Abf1p transcription factor starting with the static network; sub-networks across the five different conditions and a summary graph indicating the degree of activeness of the interactions. Edges are coloured according to the target gene chromosome location.

References

[1] Gehlenborg N, O'Donoghue SI, Baliga NS, Goesmann A, Hibbs MA, Kitano H, Kohlbacher O, Neuweger H, Schneider R, Tenenbaum D, Gavin AC: Visualization of omics data for systems biology. Nature Methods 2010, 7(3):S56.

[2]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

[3] Albert R: Scale-free networks in cell biology. Journal of Cell Science 2005, 118:4947–4957.

[4] Balazsi G, Barabasi AL, Oltvai ZN: Topological units of environmental signal processing in the transcriptional regulatory network of Escherichia coli. PNAS 2006, 102(22):7841–7846.

[5]Babu MM, Teichmann SA, Aravind L: Evolutionary Dynamics of Prokaryotic Transcriptional Regulatory Net- works. J. Mol. Biol. 2006, 358:614–633.

[6]Ma HW, Buer J, Zeng AP: Hierarchical structure and modules in the Escherichia coli transcriptional regulatory network revealed by a new top-down approach. BMC Bioinformatics 2004, 5(199).

[7] Ma HW, Kumar B, Ditges U, Gunzer F, Buer J, Zeng AP: An extended transcriptional regulatory networks of Esherichia coli and analysis of its hierarchical structure and network motifs. Nucleic Acids Research 2004,32(22):6643–6649.

[8] Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U: Network Motifs: Simple Building Blocks of Complex Networks. Science 2002, 298:824–827.

[9] Luscombe NM, Babu MM, Yu H, Snyder M, Treichmann SA, Gerstein M: Genomic analysis of regulatory network dynamics reveals large topological changes. Nature 2004, 431:308–312.

[10] Babu MM, Teichmann SA, Aravind L: Evolutionary Dynamics of Prokaryotic Transcriptional Regulatory Networks. J. Mol. Biol. 2006, 358:614-633.

[11] Lozada-Chavez I, Janga SC, Collado-Vides J: Bacterial regulatory networks are extremely flexible in evolution. Nucleic Acids Research 2006, 34(12):3434-3445.

[12] Stephen Altschul, Warren Gish, Webb Miller; Eugene Myers, David Lipman (1990). "Basic local alignment search tool." Journal of Molecular Biology 215 (3): 403–410.

[13] 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.

[14] 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.

[15] 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

[16] Node.js. http://nodejs.org/

[17] Chua XY: Prediction of transcriptional regulatory interactions in bacteria: a comparative genomics approach. PhD thesis, Queensland University of Technology 2012.

[18] Gao H, Li Y, Guo Z, Han Y, Song Y, Zhai J, Du Z, Wang X, Lu J, Zhou D, Yang R: The iron-responsive Fur regulon in Yersinia pestis. J. Bacteriol. 2008, 190(8):3063-3075.

[19] Carniel E: The Yersinia high-pathogencity island: an iron-uptake island. Microbes and Infection 2001, 3(7):561-569.

[20] Sulakvelidze A: Yersiniae other than Y. enterocolitica, Y. pseudotuberculosis, and Y. pestis: the ignored species. Microbes and Infection 2000, 2(5):497-513.

[21] Hoon MJLD, Imoto S, Kobayashi K, Ogasawara N, Miyano S: Predicting the Operon Structure of Bacillus subtilis Using Operon Length, Intergene Distance, and Gene Expression Information. Pac Symp BioComput 2004, 9:276-287.

[22] Salgado H, Moreno-Hagelsieb G, Smith TF, Collado-Vides J: Operons in Escherichia coli: Genomic analyses andpredictions. PNAS 2000, 97(12):6652-6657.

[23] Pertea M, Ayanbule K, Smedingho_ M, Salzberg SL: OperonDB: a comprehensive database of predicted operonsin microbial genomes. Nucleic Acids Research 2008, 37(Database issue):D479-D482.

[24] Leyn SA, Li X, Zheng Q, Novichkov PS, Reed S, Romine MF, Fredrickson JK, Yang C, Osterman AL, Rodionov DA: Control of Proteobacterial Central Carbon Metabolism by the HexR Transcriptional Regulator. Journal of Biological Chemistry 2011, 286(41):35782–35794.

[25] Augustus AM, Spicer LD: The MetJ regulon in gamma proteobacteria determined by comparative genomics methods. BMC Genomics 2011, 12(588).