

Poster presentation

Open Access

## Predicting gene co-expression from CIS-regulatory regions

Jamie Allen\*<sup>1,2</sup>, Mark Shirley<sup>1</sup>, Steve Rushton<sup>1</sup>, Ajay Kohli<sup>1</sup> and Ehsan Mesbahi<sup>2</sup>

Address: <sup>1</sup>Institute for Research on Environment and Sustainability (IRES), Devonshire Building, Newcastle University, Newcastle, NE1 7RU, UK and <sup>2</sup>School of Marine Science and Technology, Armstrong Building, Newcastle University, Newcastle, NE1 7RU, UK

Email: Jamie Allen\* - [jamie.allen@ncl.ac.uk](mailto:jamie.allen@ncl.ac.uk)

\* Corresponding author

from BioSysBio 2007: Systems Biology, Bioinformatics and Synthetic Biology Manchester, UK. 11–13 January 2007

Published: 8 May 2007

BMC Systems Biology 2007, 1(Suppl 1):P56 doi:10.1186/1752-0509-1-S1-P56

This abstract is available from: <http://www.biomedcentral.com/1752-0509/1?issue=S1>

© 2007 Allen et al; licensee BioMed Central Ltd.

### Background

Single Nucleotide Polymorphisms (SNPs) affecting phenotype are frequently found in the CIS regulatory factors of a gene. However current prediction tools ignore positional data and thus over estimate the co-expression of genes. In order to reduce this error, this work intends to produce a method of correlating quantitative trait loci (QTLs) simultaneously to their CIS element and the expressed phenotype. To decipher the cryptic code in this non-coding DNA autoassociative neural network tools and multidimensional self-organising maps (Kohonen maps) are being used.

Preliminary work produced a tool using a 2 dimensional Kohonen map has produced some promising results. This work aims to further improve the prediction rate by increasing the dimensions used in the self-organising map so more positional data is incorporated. Once an effective prediction tool has been developed it will be trained using existing genomic data from Arabidopsis pollen and its effectiveness assessed. Once trained the prediction tool can then be used on novel data and the predictions, if not available in published literature can be tested *in vitro* using Transcript and Locus profiling techniques within the university.

The ultimate aim being to use the prediction tool on regions of the rice genome mapped for high density heterosis QTLs to predicted co-regulated genes. Then using in-

house genomics techniques meaningful biological information may be assigned to the co-expression.

### Methods

1000 base upstream sequence of redox genes and isogenes in bacteria and plants were obtained from the NCBI <http://www.ncbi.nlm.nih.gov/> and TIGR <http://www.tigr.org/databases>.

Motifs in bacteria were scanned using the Prokaryotic Database of Gene Regulation's virtual footprint tool <http://prodoric.tu-bs.de/>. Output was filtered using customised Perl scripts. For rice and arabidopsis, motif lists were obtained from the Plant CIS-acting regulatory DNA elements (PLACE) database [1] and are scanned using a custom written Perl script.

### Results

Bacterial data has been initially analysed using multivariate techniques and indicates greater variation between isogenes than between genes or even between isogenes of the same gene in different species. The future is to develop the neural network tools to refine prediction.

### References

1. Higo K, Ugawa Y, Iwamoto M, Korenaga T: **Plant cis-acting regulatory DNA elements (PLACE) database**. *Nucleic Acids Res* 1999, **27**:297-300.