

Poster presentation

Sequence Analysis In Promoter Regions

Jeremy Austin*

Address: Cranfield University, Department of Analytical Science and Informatics (DASI). Cranfield University, Barton Rd, Silsoe, Bedfordshire, MK45 4DT, UK.

Email: Jeremy Austin* - J.austin.s00@cranfield.ac.uk

* Corresponding author

from BioSysBio: Bioinformatics and Systems Biology Conference
Edinburgh, UK, 14–15 July 2005

Published: 21 September 2005

BMC Bioinformatics 2005, **6**(Suppl 3):P2

A gene's promoter region is a prominent factor in determining that gene's expression networks and cycle. The significant elements which determine the promoter's effect are generally discovered through laboratory methods. This project's aim is to develop high throughput sequence analysis methods for identifying the important promoter features and classifying promoters according to the occurrence of these features.

Saccharomyces Cerevisiae, as one of the first fully sequenced organisms which also has publicly available expression data, was chosen as a case study for this work. Motifs of over-represented sequence segments were sought using a bespoke implementation of the Smith-Waterman algorithm running on the Cambridge-Cranfield high performance computer facility. The algorithm identifies all pairs of sequence segments whose similarity surpasses a given threshold. The found segments are then grouped by similarity, and those groups with a many members are the motifs which are over-represented.

The motifs identified using this approach are compared to known transcription factor binding sites from the *S. Cerevisiae* Promoter Database. Most of the motifs produced are expected to correspond with binding sites; others could be new discoveries. Having validated the method on *S. Cerevisiae*, it can then be used with some confidence to analyse promoters from other species, whose characteristics are less well understood.