

Refinement and Validation of Algorithms for Deduction of Gene Content from CGH Microarrays.

Doctor Of Philosophy

Robert Paul Davey

Institute of Food Research,
Norwich Research Park,
Colney Lane,
Norwich,
NR4 7UA

September 2005

©This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that no quotation from the thesis, nor any information derived therefrom, may be published without the author's prior, written consent.



ABSTRACT

Comparative Genomic Hybridisation (CGH) microarray technology is a high-throughput data generation technique that offers a new route to understanding sources of fine-detail genome variation. Despite the fact that genome sequencing is becoming cheaper and quicker, it is not reasonable to assume that all organisms and strains will ever be sequenced. CGH can provide a method of whole genome investigation without the need for the slower and more expensive sequencing process. Furthermore, it has the potential to resolve areas of phylogenetic uncertainty by providing a robust platform for evolutionary analysis. However, this type of microarray data is prone to high false positive and false negative error rates through experimental noise and systematic bias. Therefore, new approaches for data analysis and subsequent validation of results are required.

The Microarray-to-Phylogeny Pipeline (MPP), presented in this thesis, is a new application that implements the novel `mppLM` algorithm. MPP is able to facilitate the necessary data processing, statistical and analytical steps required to produce a reliable gene content estimation from CGH datasets via `mppLM`. MPP is then able to utilise this genomic footprint to infer phylogenies based on a statistical model of evolutionary gene loss. In addition, MPP provides a mechanism for the verification of this gene content information using published whole genome sequence information, allowing precise validation of algorithm accuracy and array quality. The success of MPP as a novel two-colour DNA hybridisation microarray analysis platform is demonstrated by re-analysis of published CGH datasets and validation of results.

Written in the Java programming language, MPP inherits the speed, modularity and stability associated with this object-oriented software development environment. New data processing and analysis modules are easily incorporated into MPP by means of a dedicated Java plugin framework. Based on the results of this project, recommendations are made for the future use of MPP and CGH as tools for comparative genomics and phylogenetics.

UEA Print Thesis

The full text of this thesis is not currently available
in the repository

Please contact

lle.helpdesk@uea.ac.uk

to request access to the print copy