

Phylogenetic Reconstruction from Microarray Data.

David Hoyle

Bioinformatics Group, Dept. of Computer Science
University of Manchester
Kilburn Building
Oxford Rd
Manchester
M13 9PL

Microarrays have become a powerful research tool in molecular biology and are routinely used to probe mRNA expression levels on a genomic scale. They can also be used to assess the genetic similarity of two species by hybridizing extracted genomic DNA from species B to a microarray containing genomic DNA from species A.

This suggests the possibility of using them to study phylogenetic (evolutionary) relationships between species by study genetic variation on a genome wide scale.

The application of this technology to phylogenetics raises statistical issues at each stage of the process, for example:

- i) The experimental design.
- ii) Initial processing of the data - How should subtraction of background intensities be performed, to provide raw estimates of homologue content?
- iii) Normalization of the microarray data to remove bias.
- iv) Which standard models or new generative models should be used to infer phylogenies?
- v) The possibility of applying machine learning algorithms to training data sets where the phylogeny is already confidently known.

We shall report some preliminary results from our research on the sensu stricto set of *Saccharomyces* species, and attempt to touch upon all the issues above. We shall highlight the potential power and limitations of this approach to phylogenetic reconstruction.