

Clustering of Anopheles Gene Expression Profiles

David Stephens

Malaria represents one of the major worldwide challenges to public health. A recent breakthrough in the study of the disease follows the annotation of the genome from the malaria parasite *Plasmodium falciparum* and the mosquito vector *Anopheles*. Of particular interest is the molecular biology underlying the immune response system of *Anopheles*, which actively fights against *Plasmodium* infection.

This talk reports a statistical analysis of gene expression profiles in mosquitoes that have been infected with bacterial agents; such an experiment gives insight into the biological mechanisms of immune defence in the mosquito *in vivo*.

One of the most important features of the analysis is the sample size; typically large numbers of genes (several thousand) are required to be analyzed simultaneously. In general, the sheer magnitude of the expression data sets concerned makes formal Bayesian inference (specifically, via simulation methods) infeasible.

This talk describes the use of a tailored Bayesian hierarchical model-based clustering algorithm to investigate variation and potential coregulation in the gene profiles; measures of cluster integrity, expression summaries, and heterogeneity are given.

We demonstrate that that our approach is readily implementable, and illustrate its use by reference to several large mosquito data sets, and explore current methods for high dimensional clustering.