MIT
Department of Mathematics
& The Theory of
Computation Group
At CSAIL

Bioinformatics Seminar

Speaker: Gabor Marth, Department of Biology, Boston College
Title: A COALESCENT COMPUTATIONAL PLATFORM TO PREDICT STRENGTH OF ASSOCIATION FOR CLINICAL SAMPLES
Date: Monday, 25 April 2005
Time & Location: **** PLEASE NOTE LOCATION****
Refreshments: 11 am in the STAR conference room at MIT's Building 32, Stata Center Room D-463
Talk: 11:30 am in the STAR conference room at MIT's Building 32, Stata Center Room D-463
URL: http://www-math.mit.edu/compbiosem/

Abstract:

The International HapMap project is genotyping millions of single-nucleotide polymorphisms (SNPs) in hundreds of individual reference DNA samples representing four different world populations. The genotype data and the annotations will collectively form a large informational resource to aid marker selection for clinical case-control association studies. Current research within the community focuses on (1) determining how best to quantify the strength of allelic association within the reference samples both in local regions of the genome and along entire chromosomes; (2) how dense a marker map is required to describe these patterns accurately; (3) what are the relationships among the patterns observed within the various HapMap reference populations; (4) whether individual SNP markers or multi-marker haplotypes are likely to carry more power for the detection of disease causing alleles in association studies; and (5) how to select an optimal set of such markers from the millions available (i.e. selection of tag SNPs).

However, the utility of the markers selected on the basis of the association patterns within the HapMap reference samples will ultimately depend on the degree to which these patterns remain constant across other sets of samples such as those from clinical populations. The consistency of the patterns may be studied experimentally, by genotyping additional individuals, and comparing the strength of association in these samples to what was measured in the HapMap samples

In this presentation we describe a computational alternative to costly genotyping of such additional samples. Using a Coalescent methodology we produce multiple, consecutive sets of simulated haplotypes that are consistent with the HapMap reference haplotype data in a given genome region. These computationally generated samples can then be used to evaluate whether the strength of association is likely to remain constant across data sets, whether tagging SNPs perform well across these sets, or the selection of a different set of tagging SNPs is necessary.

We will address technical points of our method: (i) how to generate data-relevant additional haplotypes efficiently; (ii) how to determine Coalescent model parameters that accurately represent the HapMap populations; (iii) how to use un-phased diploid genotype data in our analysis; (iv) how to proceed in the case of SNPs for which the identity of the ancestral and the mutant allele is not known. We will demonstrate how to use the simulated haplotypes for predicting allelic association strength for a future set of samples. We will show that the simulated haplotypes can be pre-computed and stored in a database, and readily updated as the HapMap project adds genotype data for additional SNP markers. This will allow us to encapsulate the algorithms in an interactive software tool that will aid study design and marker prioritization for clinical applications.

Massachusetts Institute
of Technology
77 Massachusetts Avenue
Cambridge, MA 02139

For General Questions, please contact kvdickey@mit.edu