Efficient Haplotype Inference on Pedigrees and Haplotype Based Disease Gene Mapping as presented by Jing Li

Christopher E. Davis - chris2d@cs.unm.edu
University of New Mexico
Computer Science Colloquia Spring 2004

April 30, 2004

Abstract

In spite of all the recent "breakthroughs" in genetic science, much of the information that genomes hold is still a mystery to us. In fact, one of the most basic questions, "Where is the gene for this disease?" remains largely unanswered. Mr. Jing Li presented several potential approaches to inferring haplotypes then using the inferred data to try to construct a gene mapping.

Haplotype inference is the process of taking a parent's genotype and inferring the offspring's haplotype. A haplotype is the organism's actual genetic code. We as humans have two haplotypes (diploid). Mendelian law gives us a reference and potential solution checker, but does not provide an efficient method to arrive at a solution.

The inference problem has several confounding factors. The first confounding point is that parental genotype data may be missing. This makes the predictions weaker and results in a larger solution space to search. The second confounding factor is that parents may have identical genomes. This means that we can not be certain of the lineage of a gene. And lastly recombination may occur. Recombination is the process by which diploid or polyploid organisms are able to mutate their genomes slightly by recombining strands of their parent's DNA. The process of recombination occurs at inception. Basically a small section

of genetic code may be swapped into the chromosome from the other parent's genes.

Parsimony suggests that the recombination problem can be dealt with. By applying the rule of parsimony researchers are able to simplify their problem while increasing their assumptions. Basically the idea is that the fewer the recombinations needed to produce a haplotype, the more likely the haplotype.

All of this research is aimed at the goal of better understanding of what our genes actually code for. By examining the data produced by the haplotype inference methods, researchers hope to gain some insight into which genes are involved in certain genetically linked diseases.

Mr. Li presented several algorithms for the inference and the mapping, however none of them were covered in enough depth to make them worth presenting here. This topic is interesting and worth more investigation.

1 Applicant Commentary

I generally liked the presentation. I felt that Mr. Li handled the questions well, and gave nice simple answers. I felt that his formula slides were a bit dense and presented a bit too quickly. I really believe that if you intend to present formulas that quickly then you should just leave them out of your presentation. I know he can do the math, he doesn't need to show me. Several of his graphs were too small to read anything. As such they should have been left out or enlarged. I felt that his clustering methods were simplistic at best and his similarity measure was only ok. It was too bad that Professor Moret missed this presentation as he could have most likely asked some meaningful questions.