

## **EFFICIENT COMPUTATION OF MINIMUM RECOMBINATION WITH GENOTYPES (NOT HAPLOTYPES)**

Yufeng Wu and Dan Gusfield  
Department of Computer Science  
University of California, Davis  
Davis, CA 95616, U.S.A.

A current major focus in genomics is the large-scale collection of genotype data in populations in order to detect variations in the population. The variation data are sought in order to address fundamental and applied questions in genetics that concern the haplotypes in the population. Since almost all the collected data is in the form of genotypes, but the downstream genetics questions concern haplotypes, the standard approach to this issue has been to try to first infer haplotypes from the genotypes, and then answer the downstream questions using the inferred haplotypes. That two-stage approach has potential deficiencies, giving rise to the general question of how well one can answer the downstream questions using genotype data without first inferring haplotypes, and also giving rise to the goal of computing the range of downstream answers that would be obtained over the range of possible inferred haplotype solutions. This paper provides some tools for the study of those issues, and some partial answers. We present algorithms to solve downstream questions concerning the minimum amount of recombination needed to derive given genotypic data, without first fixing a choice of haplotypes. We apply these algorithms to the goal of finding recombination hotspots, obtaining as good results as a published method that first infers haplotypes; and to the case of estimating the minimum amount of recombination needed to derive the true haplotypes underlying the genotypic data, obtaining weaker results compared to first inferring haplotypes using the program PHASE. Hence our tools allow an initial study of the two-stage versus one-stage issue, in the context of specific downstream questions, but our experiments certainly do not fully resolve the issue.