An Algorithm Based on Perfect phylogeny for Identifying Haplotype Block Boundaries and Tag SNP's

S. N. Afzaly¹, H. Pezeshk²,³, M. Sadeghi¹,²,⁴ and Ch. Eslahchi¹,²

¹- Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran
²- Institute for Studies in Theoretical Physics and Mathematics, Tehran, Iran.
³- Center of Excellence in Biomathematics, School of Mathematics, Statistics and Computer Sciences, College of Science, University of Tehran, Tehran, Iran
⁴- National Institute for Genetic Engineering and Biotechnology, Tehran, Iran

Abstract

The study based on haplotype block partitioning has drawn wide attraction from the researchers. There are some commonly used methods to identifying blocks in human genome. However, it is a challenging problem to identify block boundaries. In this paper we describe a new method based on the perfect phylogeny for finding haplotype blocks. We give a rigorous definition of the quality of the segmentation of haplotype region into blocks and describe a dynamic programming algorithm and greedy algorithm for finding the proper segmentation with respect to this measure. It is shown that the minimum number of Tag SNP's in a perfect phylogeny block with \( m \) haplotype and \( n \) SNP's is \( m-1 \) provided that there is no missing data. To identify the Tag SNP's, a polynomial time algorithm is presented. Using this algorithm, the number of Tag SNP's in blocks with missing data is approximated. Introducing a distance function, we finally apply the algorithm to published SNP data of human chromosome 21 and compare the results with the method used by Patil's et al 2001.