Parallel Individual Haplotying Assembly : Xeon Phi vs. Nvidia K20x

Robert J. clucas

School of Electrical & Information Engineering, University of the Witwatersrand, Private Bag 3, 2050, Johannesburg, South Africa

Abstract:

Key words: Brach, Bound, GPU, Haplotyping, Simplex

1. INTRODUCTION

It is commonly accepted that all humans share ∼99% of the same DNA, however, the small variations cause the human beings to have different physical traits. Single nucleotide polymorphisms (SNPs), which is the variation of a single DNA base from one individual to another, and are believed to be able to address genetic differences. For diploid organisms, which have pairs of chromosomes, a *haplotype* is a sequence of SNPs in each copy of a pair of chromosomes. A *genoype* describes the conflated data of the haplotypes on a pair of chromosomes. Haplotypes are believed to contain more generic information than genotypes [1].Obtaining haplotypes correctly is a difficult problem, which is broken into two subdomains: individual haplotype assembly and haplotype inference.

Haplotype inference uses the genotype of a set of individuals. The genotype data tells the status of each allele at a position, but does not distinguish which copy of the chromosome the allele came from. CITE (HE) This negative aspects of this approach are that it cannot distinguish rare and novel SNPs, and there is no way of knowing if the inferred haplotype is completely correct.

Individual haplotype assembly uses fragments of sequences generated by sequencing technology to determine haplotypes. The fragments of a sequence come from the two copies of an individual's chromosome, the goal of the individual haplotyping problem is to correctly determine two haplotypes, where each haplotype corresponds to one of the two copies of the chromosome.

2. BACKGROUND

REFERENCES

[1] J. Stephens. "Haplotype Variation and Linkage Disequilibrium in 313 Human Genes." *Science*, vol. 293, no. 5529, pp. 489–493, Jul. 2001. URL `http://dx.doi.org/10.1126/science.1059431`.