[10pt,twocolumn]witseiepaper

KJN textcomp url amsfonts amsmath algorithm

k m n d p q

document

# Parallelizing the Individual Haplotying Assembly Problem

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

Brach, Bound, GPU, Haplotyping, Simplex

emptyempty

## INTRODUCTION

It is commonly accepted that all humans share ∼99% of the same DNA, however, small variations cause human beings to have different physical traits. Single nucleotide polymorphisms (SNPs), which are variations of a single DNA base from one individual to another, 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 genotype describes the conflated data of the haplotypes on a pair of chromosomes. Haplotypes are believed to contain more generic information than genotypes stephens:2001, however, obtaining haplotypes correctly is a difficult problem, which is broken into two subdomains: haplotype assembly and haplotype inference.

## HAPLOTYPE ASSEMBLY PROBLEM sec:hap

This section will provide a brief overview of the haplotype assembly (HA) problem, and define the notation used through the rest of the paper. The input to the problem is a set of reads from a given genome sequence, where each read contains fragments from each of the two chromosomes which make up the genome sequence. These characters of a read consist of elements from a ternary string, where a ternary string has characters from the set $\{0, 1, -\}. Avalueof0referstothemajoralleleatasite, avalueof1totheminorallele, andavalueof-$
$tothelackofareadatthesite, andisreferredtoasagap. Thesereadsarethencombinedtoformamatrix, M, whereeachrowofth$

itemize

n polynomial time.

G being bipartate.

ized schwartz:2010.

lution for all cases.