

Genotyping arrays produce diploid genetic data in which maternally and paternally derived haploid chromosomes are combined into a total allele count at each site. Inferring haploid phase from diploid data -- "phasing" for short -- is a fundamental question in human genetics and a key step in genotype imputation. Most existing methods for computational phasing apply hidden Markov models; these algorithms are statistically precise but computationally challenging. Long-range phasing (LRP) is an alternative, much faster approach that harnesses long identity-by-descent (IBD) tracts shared among related individuals; in such IBD regions, phase inference is straightforward. However, because of its reliance on long IBD, LRP has previously only been successfully applied in data sets containing close relatives. I will describe a new LRP-based method, Eagle, that leverages distant relatedness -- ubiquitous in very large data sets such as the UK Biobank -- along with fast approximate HMM decoding to achieve a 1-2 order of magnitude speedup over existing methods.