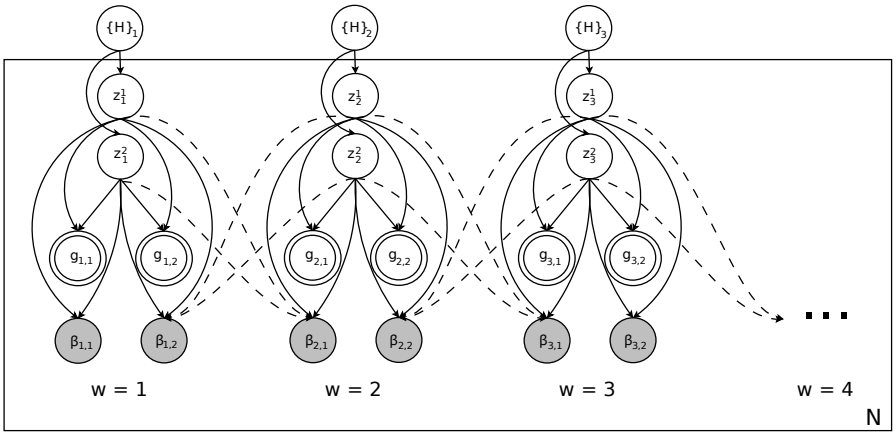
Supplementary Materials

GeneImp: Fast Imputation to Large Reference Panels Using Genotype Likelihoods from Ultra-Low Coverage Sequencing

Athina Spiliopoulou, Marco Colombo, Peter Orchard, Felix Agakov, and Paul McKeigue



Supplementary Figure 1. Probabilistic graphical model of GeneImp for 3 consecutive windows. Each target sample is modelled independently of other target samples (denoted by the plate notation with N plates). $\{H\}_w$ denotes the set of reference haplotype sequences at sites within window w. Double-circles denote deterministic nodes (Equation 3 in the main text). Gray nodes denote observed variables (e.g. sequencing reads). Dashed edges denote dependencies that are dropped in the factorised approximation of the joint distribution over hidden states (Equation 4 in the main text). We depict only two loci with reads ($\beta_{w,1}$ and $\beta_{w,2}$) in each window; in practice, this number is higher and varies among windows.