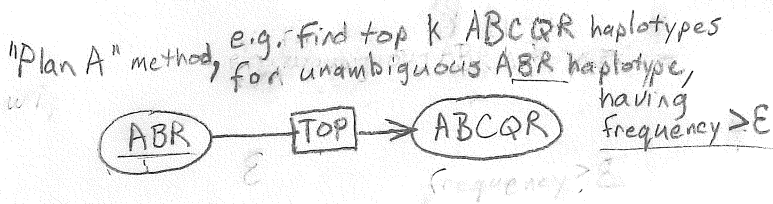


Graph-Based Imputation

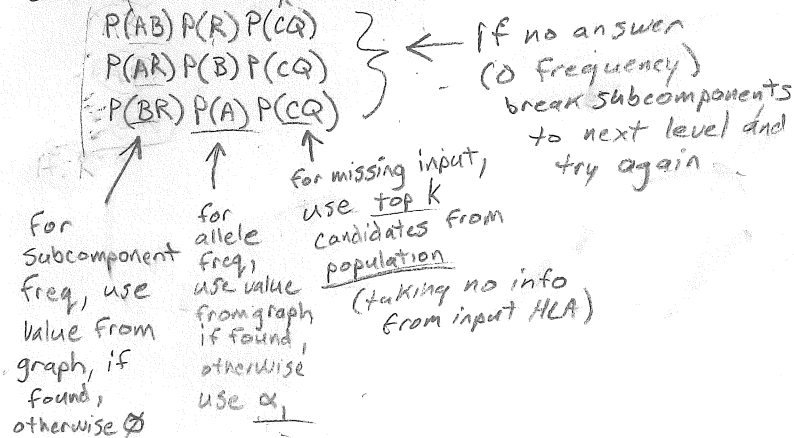
Input: 1) Target locus set, e.g. ABCQR
 2) ambiguous HLA data, possibly with missing loci (as compared to target locus set)
 Output: 1) multilocus genotype list: list of unambiguous haplotype pairs

Handling ambiguous HLA input data

- 1) Convert ambiguous HLA into a list of unambiguous unphased multilocus genotype.
- 2) Convert each ^{unambiguous} unphased multilocus genotype into a list of unambiguous haplotype pairs representing each of its (2^{n-1}) phases ^{$n = \text{number of input loci}$}
- 3) Use unambiguous haplotype pairs (possibly with missing loci as compared to target locus set) as input to "Plan A" imputation method, using an initial ϵ -value, e.g. $\emptyset.\emptyset1$.
- 4) If "Plan A" imputation method fails when using the initial ϵ value, try again using a lower ϵ value, e.g. $\emptyset.\emptyset\emptyset1$. If that fails, try a few more times, using lower and lower ϵ values (e.g. $\emptyset.\emptyset\emptyset\emptyset1$, $\emptyset.\emptyset\emptyset\emptyset\emptyset1$). If that fails, try again using $\epsilon = \emptyset$.
- 5) ~~only~~ "Plan A" imputation method yields no genotype frequencies (via haplotype pairs) for the target locus set, use "Plan B" imputation ^{method} to generate haplotype candidates and frequency estimates for all input haplotypes which could not be imputed using "Plan A".



"Plan B" method, e.g. for ABR haplotype not found in graph
 Consider all ABR ^{direct} subcomponents



Priors	
α_0 = unobserved ABCQR haplotype frequency	α_1 = unobserved allele frequency
α_i = unobserved allele frequency	
Z = normalization factor, e.g. 1.5	

Populate graph using \tilde{P}_i , where (for haplotype)

$$\tilde{P}_i = \frac{P_i + \alpha_0}{Z}$$

So, $Z = \sum \tilde{P}_i$

For unobserved haplotypes,

$$\tilde{P}_i = \frac{\alpha_0}{Z}$$