

Algorithms in Genome Research
Winter 2019/2020

Exercises

Number 8, Discussion: 2020 January 17

1. Give general formulas for the following questions. If this is difficult, enumerate the solutions for small examples.
 - (a) A given genotype g has k *biallelic* sites, i.e., sites with value 2. How many different haplotype pairs could have given rise to g ?
 - (b) If the haplotypes come in blocks of 10 sites each, how does this decrease the number of *different* haplotype pairs?
 - (c) For l founder sequences and m recombination hot spots, how many haplotype pairs are possible (under the assumption that recombinations only occur at hot spots)?
2. In the lecture, we discussed a simplified ILP solution to the Haplotype Assembly Problem, where each column was constrained to contain at most two distinct characters. Explain how to modify the ILP formulation to eliminate this constraint, while still assuming that the phased genome sequence is diploid.
3. Solve the following instance P of the perfect-phylogeny haplotype (PPH) problem, i.e., check whether a haplotype matrix exists that gives rise to P and report such matrix if it exists: (optional)
$$P = \begin{pmatrix} 1 & 0 & 2 & 2 \\ 2 & 1 & 2 & 2 \\ 2 & 2 & 2 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 2 & 2 & 2 \end{pmatrix}$$
4. Based on the strategy discussed in the lecture, develop an abstract ILP that solves the PPH problem. (optional)