Algorithms in Genome Research
Winter 2025/2026

Exercises

Number 7, Discussion: 2025-December-19

1. Given the following instance of the haplotype inference problem (HIP): (1) 100111, (2) 100121,
   (3) 220102, (4) 020221, (5) 120102, (6) 120122, (7) 200121.

   (a) Apply Clark's algorithm.

   (b) Solve the pure parsimony variant of the HIP.

   (c) Solve the perfect phylogeny haplotyping problem (PHP), if possible.

2. In class we have discussed the Haplotype Assembly Problem and its solution by an Integer
   Linear Program (ILP). Details and some more discussion can be found in the textbook "Integer Linear Programming in Computational and Systems Biology" by Dan Gusfield (2019),
   Section 20.2. There you find also the following two exercises (Exercise 20.2.5):

   (a) Explain how to modify the ILP formulation to the haplotype assembly problem, when
       a column could contain more than two characters.

   (b) What changes are needed in the ILP if the reads can be of unequal length?

3. Molecular haplotyping modeled by the (weighted) minimum error correction problem focuses
   on assembling SNP haplotypes from reads of a sequenced genome. To fully characterize an
   individual genome, however, haplotyping must produce exhaustive lists of both SNPs and
   non-SNPs, that is, larger variants (indels, block replacements, inversions, ...). Discuss any
   ideas how non–SNP variants could be integrated in the analysis.