

Algorithms in Genome Research
Winter 2017/2018

Exercises

Number 04, Discussion: 2017 November 17

1. Remember and explain a few of the terms mentioned in class:
 - genotype, haplotype
 - segregating site
 - genealogy
 - infinite sites model
 - four-gametes test
2. Give general formulas for the following questions. If this is difficult, enumerate the solutions for small examples.
 - (a) For k markers of two alleles each, how many different haplotype vectors are at most possible?
 - (b) If the haplotypes come in blocks of 10 sites each, how does this decrease the number of different haplotype vectors?
 - (c) For m founder sequences and l recombination hot spots, how many haplotype vectors are possible (under the assumption that recombinations only occur at hot spots)?
 - (d) For k individuals, what is the maximum number of different segregating sites such that the four-gametes test does not fail?
3. Given the following instance of the haplotype inference problem (HIP): (1) 102212, (2) 212210, (3) 222010, (4) 101110, (5) 122212, (6) 121110, (7) 222210.
 - (a) Apply Clark's algorithm.
 - (b) Solve the pure parsimony variant of the HIP.
 - (c) Solve the perfect phylogeny haplotyping problem (PHP), if possible.