

Algorithms in Genome Research

Winter 2018/2019

Exercises

Number 9, Discussion: 2019 January 11

1. 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, building the columns of a haplotype matrix, how many *different* haplotype vectors (rows) 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 haplotype vectors, what is the maximum number of *different* configurations at segregating sites such that the four-gametes test does not fail?
2. 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.
3. What is the difference between linkage and association?
4. Let the following haplotype matrix be given (circles represent the presence of a mutant allele):

	1	2	3	4	5	6	7	8	9	10
A	—O—	—	—	O—	O—	O—	—	—	O—	O—
B	—	—	O—	—	—	—	O—	—	—	—
C	—O—	O—	—	—	O—	O—	—	O—	O—	O—
D	—	—	O—	—	—	—	—	O—	—	—
E	—O—	O—	—	—	—	O—	—	—	O—	—
F	—	—	—	O—	—	—	O—	—	O—	—

- (a) Find the maximal regions around each segregating site, not violating the four-gametes test.
- (b) Draw the local trees of the segregating sites.
- (c) Assume that individuals C and E are the cases, the other the controls. Which of the segregating sites show highest evidence for association with the disease?