Universität Bielefeld
Technische Fakultät

AG Genominformatik
Prof. Dr. Jens Stoye

Algorithms in Genome Research
Winter 2009/2010

Exercises

Number 1, Discussion: 2009 October 30

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 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 segregating sites such that the four-gametes test does not fail?
3. Let the following haplotype matrix be given (circles represent the presence of a mutant allele):

```
1 —O————————O—O—O————————O—O—
2 ————————O————————O————————
3 —O—O————————O—O———O—O—O—
4 ————————O————————O————————
5 —O—O————————O————————O—
6 ————————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 3 and 5 are the cases, the other the controls. Which of the segregating sites show highest evidence for association with the disease?