**Universität Bielefeld**
**Technische Fakultät**

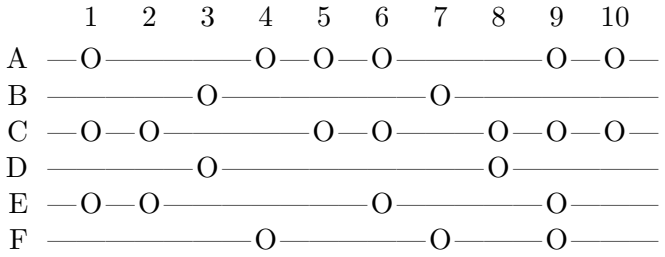
**AG Genominformatik**
**Prof. Dr. Jens Stoye**
**Dr. habil. Marília D. V. Braga**

# Algorithms in Genome Research
## Winter 2025/2026

## Exercises

**Number 8, Discussion: 2026-January-9**

1. Investigate: What is the difference between linkage and association?
2. Let the following haplotype matrix be given (circles represent the presence of a mutant allele):




   (a) Find the maximal regions around each segregating site, not violating the four-gametes test.
   (b) Draw the local trees for these regions.
   (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?

3. Discuss: Can pedigree (family) information help in the analysis of whole-genome association studies?
4. Give general formulas for the following questions. If this is difficult, enumerate the solutions for small examples.

   (a) For $n$ biallelic sites, represented by the columns of a $k \times n$ binary haplotype matrix, how many *different* haplotype sequences (rows) are at most possible?
   (b) If the haplotypes come in blocks of 10 sites each, 2 variants each, how does this decrease the number of *different* haplotype vectors?
   (c) For $l$ founder sequences and $m$ recombination hot spots, how many haplotype vectors are possible (under the assumption that recombinations occur only at hot spots)?
   (d) For $k$ haplotype sequences, what is the maximum number $n_{\max}(k)$ of *different* configurations at segregating sites (columns) such that the four-gametes test does not fail?