

Algorithms in Genome Research
Winter 2005/2006

Exercises

Sheet 2, Discussion: 17.11.2005

1. Discuss the main experimental problems that make sequence assembly difficult.
2. Find the shortest common superstring of the following sequences:

ATAGCC
ATATAT
ATATCG
CGGGAC
GACATA
GACTAT
GCCGGT
GGTATA
TATATA
TATCGG

Is the coverage uniform? If not, find a layout with a more uniform coverage.

3. In the overlap phase, prefix-suffix “local alignments” are sought. Work out the details of an algorithm.
4. What are mate pairs? Do they simplify the assembly problem?