

Algorithms in Bioinformatics II, SS2004
Assignment sheet # 11

Daniel Huson

June 25, 2004

1 The SBH graph (3 points)

Suppose we are given the following spectrum:

$$S = \{AGG, ATG, CCG, CGG, GAT, GCC, GGA, GGC, GGT, TGG\}.$$

Please draw the SBH graph and list the possible reconstructed sequences.

2 Linkage clustering methods (4 points)

Give definitions for *single linkage* clustering, *average linkage* clustering and *complete linkage* clustering.

Present an example for which all three methods produce a different set of clusters.

Give a short characterization of the kinds of clusters that each of the three methods produce, e.g. very tight, or long chains of loose, straggly clusters etc.

3 Resequencing by DNA chips (3 points)

Research the web to find out about “resequencing using DNA chips”. Give an overview of the technique. For a fully sequenced reference genome, please discuss what kind of additional information one hopes to obtain by applying resequencing technology to the genome of other individuals of the same species? Can resequencing also be useful for investigating the genome of a closely related species?

Assignments due: **Monday, July 5, 10am**