

# Algorithms in Bioinformatics II, SS2003

## Assignment sheet # 12

Daniel Huson

July 13, 2003

### 1 Preparation (2 points)

Download the following mtDNA sequences from GenBank:

- primates: gorilla, human, chimpanzee and orangutan,
- ferungulates: cat, seal, horse, blue whale and cow,
- rodents: rat and house mouse, and
- marsupials: opossum.

Download the program `GenCompress` from <http://monod.uwaterloo.ca/downloads/gencompress/>.

### 2 Compression ratios (4 points)

Using the program `GenCompress`, determine the compression ratio  $Compress(w)$  for each mtDNA genome  $w$ . Compute the matrix of all pair-wise conditional compression ratios  $Compress(w | z)$ . Compute the matrix of all pair-wise “concatenated” compression ratios  $Compress(wz)$ .

### 3 Mutual information distances (2 points)

From this information, compute the “mutual information” distance matrix using the formula

$$D(w, z) := 1 - \frac{Compress(w) - Compress(w | z)}{Compress(wz)}.$$

### 4 Phylogenetic tree (2 points)

Run the *Neighbor-Joining* algorithm on this data set to obtain a phylogeny. Discuss the tree.

Assignments due: **Monday, July 21, 10am**