

Leipzig School of Human Origins Course on  
'Phylogeny Reconstruction' (04.-08.05.2007)

Final Exam - 08.06.2007

Name:	
Vorname:	

1. What is the alignment score of the following alignment using the scoring function:

**2 P**

$$s(a, b) = \begin{cases} +5 & \text{if } a = b \text{ (match)} \\ -2 & \text{if } a \neq b \text{ (mismatch)} \\ -6 & \text{if } a \text{ or } b \text{ indel (gap)} \end{cases}$$

sequence 1: G C T C G T A  
sequence 2: - - T C A T A

2. Why can the inclusion of paralogous sequences make phylogenetic analysis for the relationship of species fail? (2-4 sentences)

**1 P**

3. What is the difference between transitions and transversions? (1-2 sentences)

**1 P**

4. Order the following evolutionary models by their complexity (i.e., number of parameters)?

**1 P**

- GTR+ $\Gamma$  (General-time-reversible+ $\Gamma$ -distributed rate heterogeneity)
- HKY85 (Hasegawa, Kishino, Yano)
- GTR
- JC69 (Jukes-Cantor)

5. Which of the following tree-reconstruction methods are using (explicit) models of sequence evolution?

**1 P**

- Maximum Parsimony
- Neighbor Joining
- Least Square
- Maximum Likelihood

6. Does the following distance matrix representable as a tree?

**3 P**

Hint: four-point condition:  $d(uv) + d(xy) \leq d(ux) + d(vy) = d(uy) + d(vx)$

	A	B	C	D
A	0	8	5	11
B		0	9	5
C			0	12
D				0

7. For the alignment below do the following:

**5 P**

- Mark all informative sites.
- Draw the maximally parsimonious tree.
- Compute the 'length' of the tree, i.e., the number of mutations needed.

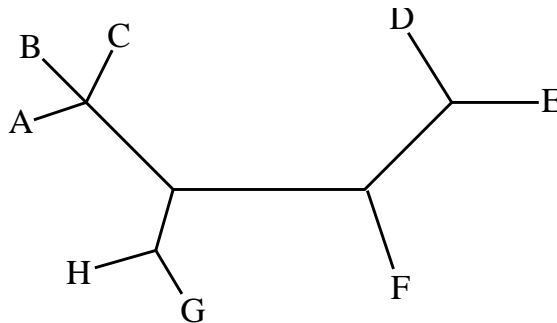
All mutations are counting 1, equally.

```
seq1:  G  G  A  A  C  G  A  T  T  A
seq2:  G  A  A  A  C  T  C  T  C  A
seq3:  G  A  A  C  C  T  C  T  C  A
seq4:  G  G  A  C  C  T  A  T  T  A
```

8. Assign the following terms to their proper locations in the tree.

**2 P**

- leaf
- external branch
- root
- multifurcation
- inner node



9. From gene A you reconstruct a tree with log-likelihood  $-2534.3$ . Phylogenetic inference from gene B produced a tree with log-likelihood  $-753.7$ . Are these two likelihoods comparable? **1 P**

10. How many mutations are expected to have occurred in a sequence of 200nt along a branch of length 0.15? **1 P**

11. What is the main differences between the maximum likelihood approach and Bayesian inference in phylogeny reconstruction? (2-5 sentences) **2 P**

12. Does a higher  $\alpha$ -value (shape parameter) in a  $\Gamma$ -distribution mean more rate homogeneity or more rate heterogeneity. **1 P**

13. A bootstrap analysis produced the following splits in the according percentages of trees: **3 P**

$ABC DEF$	100%
$AB CDEF$	60%
$ABCF DE$	45%
$AC BDEF$	40%
$ABCD EF$	35%
$ABCE DF$	25%

Construct and draw the the 50%-majority rule consensus and the strict consensus.

14. The following plot shows the log-likelihoods of the sampled trees from a Bayesian analysis. Mark the minimum length of the burn-in in the diagram. Is this a definite decision? (1 sentence)

**2 P**

