Midterm Exam #1
MB 451
Microbial Diversity

Honor pledge: "I have neither given nor received unauthorized aid on this test."

Signed: ____________________________________________________________  Date: __Feb 5, 2007__________

Name: ___KEY__________________________________________________________

1. What are the three primary evolutionary branches of life? (5 points)
   *Archaea, Bacteria, and Eukarya*

Multiple choice (2 points each)

2. ___D___ Which of the following is not an aspect of a taxonomy?
   A. Grouping
   B. Naming
   C. Identifying
   D. Isolating
   E. None of the above

3. ___E___ Which describes a phylogeny?
   A. a genealogy
   B. a tree
   C. evolutionary pathways
   D. natural relationships between organisms
   E. All of the above

4. ___A___ Species evolve by ...
   A. diversification
   B. progression
   C. linear advancement
   D. magic
   E. none of the above

5. ___D___ Which of these evolutionary representations is both objective and quantitative?
   A. the 5-Kingdom tree
   B. the Evolutionary Ladder
   C. the Prokaryote/Eukaryote dichotomy
   D. the 3-Domain molecular phylogenetic tree
   E. None of the above

6. ___D___ Which of the following is not an algorithm for generating phylogenetic trees from molecular data?
   A. Neighbor-joining
   B. Parsimony
   C. Maximum likelihood
   D. Jukes & Cantor
   E. all of the above

7. ___B___ A photoheterotroph gets its energy from X and its carbon from Y.
   A. X= light ; Y= carbon dioxide
   B. X= light ; Y= organic compounds
   C. X= inorganic compounds ; Y= carbon dioxide
   D. X and Y both = organic compounds
   E. none of the above
Short answer & fill in the blank (points as indicated)

8. Usually the best single molecular sequence for phylogenetic analysis is __ssu-rRNA_____________________. (2 points)

9. A 2-dimensional matrix of molecular sequences is called a(n) ___sequence alignment_____________________. (2 points)

10. Molecular sequences may be alignable using ______structural superimposition___________________________ even if they are so different from one another that they cannot be reliably aligned based on the sequence alone. (2 points)

11. ___bootstrapping_____________________________ is a way to judge the reliability of the branches in a tree. (2 points)

12. Why is an understanding of phylogeny important? (4 points)

_Both to make predictions about the properties of an organism, and to prevent inappropriate comparisons between unrelated organisms._

13. What are the main steps of a molecular phylogenetic analysis? (4 points)

_1. Decide on organisms and sequences_
_2. Obtain this data_
_3. Identify homologous residues_
_4. perform phylogenetic analysis_

14. Why is raw sequence similarity an underestimate of evolutionary distance? (5 points)

_Because of the occurrence of multiple substitutions at a single site that are only counted as a single difference, or in the case of reversion, as no difference at all._
Tree interpretation: answer the following questions based on this ssu-rRNA-based phylogenetic tree. The scale bar (0.10) represents evolutionary distance. Sequences/organisms without names (i.e., pBB, Nak_9, SRI-240, and VC2.1bac27) are from uncultivated species. (2 points each)

15. Is this a dendrogram or a phenogram? __ phenogram

16. Which sequence(s) is/are presumably the outgroup? __ Methanococcus jannashii

17. Which sequence is most closely related to Bacillus subtilis? __ Escherichia coli

18. What is the evolutionary distance between Persephonella marina and Hydrogenobacter thermophilus? __ 0.2

19. Circle (on the tree above) the last common ancestor of Bacillus subtilis and Thermus thermophilus.

20. Except for the outgroup, which two sequences in this tree are the most distantly related? __ Flexibacter flexilis __ and __ pBB

21. Which branch(es) would you have the least confidence in? 
   The one between the 2 Persephonella species, and the one between E.coli and B.subtilis (the one between pBB and Nak_9 is also right up there).
Problem-solving (points as indicated)

22. Align the following sequences: (8 points)

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Alignment 1</th>
<th>Alignment 2</th>
<th>Alignment 3</th>
<th>Alignment 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence 1</td>
<td>G G G G U U C G C U C A</td>
<td>- G G G G U U C G C U C A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequence 3</td>
<td>C G A G G C U G C U C</td>
<td>C G A G G C U - G C U C -</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequence 4</td>
<td>U G A G G C U C G C U C A</td>
<td>U G A G G C U C G C U C A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

23. Generate a similarity matrix from the following alignment: (10 points)

<table>
<thead>
<tr>
<th>Sequence 1</th>
<th>G A U C G A U C G A</th>
<th>Seq1</th>
<th>XX</th>
<th>XX</th>
<th>XX</th>
<th>XX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence 2</td>
<td>G A U A G A U C G A</td>
<td>Seq2</td>
<td>0.9</td>
<td>XX</td>
<td>XX</td>
<td>XX</td>
</tr>
<tr>
<td>Sequence 3</td>
<td>G A C C G A C C G A</td>
<td>Seq3</td>
<td>0.8</td>
<td>0.7</td>
<td>XX</td>
<td>XX</td>
</tr>
<tr>
<td>Sequence 4</td>
<td>G A G U G A C C G A</td>
<td>Seq4</td>
<td>0.7</td>
<td>0.7</td>
<td>0.8</td>
<td>XX</td>
</tr>
</tbody>
</table>

24. Convert the following similarity matrix into a distance matrix using the Jukes & Cantor curve: (10 points)

<table>
<thead>
<tr>
<th>Similarity matrix</th>
<th>Distance matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq1 Seq2 Seq3 Seq4</td>
<td>Seq1 Seq2 Seq3 Seq4</td>
</tr>
<tr>
<td>Seq1 XX XX XX XX</td>
<td>Seq1 XX XX XX XX</td>
</tr>
<tr>
<td>Seq2 0.90 XX XX XX</td>
<td>Seq2 0.12 XX XX XX</td>
</tr>
<tr>
<td>Seq3 0.60 0.75 XX XX</td>
<td>Seq3 0.55 0.30 XX XX</td>
</tr>
<tr>
<td>Seq4 0.50 0.45 0.35 XX</td>
<td>Seq4 0.77 0.90 1.30 XX</td>
</tr>
</tbody>
</table>

JUKES & CANTOR CURVE
25. Convert the following phenogram into a dendrogram. (10 points)

```
A
B
C
D
```

26. Generate a tree using the neighbor-joining method from the following distance matrix. SHOW YOUR WORK! You need only solve the structure of the tree; you do not need to solve the branch lengths. (10 points)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>B</td>
<td>0.3</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>C</td>
<td>0.6</td>
<td>0.5</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>D</td>
<td>1.7</td>
<td>1.6</td>
<td>1.7</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>E</td>
<td>1.0</td>
<td>0.9</td>
<td>1.0</td>
<td>1.7</td>
<td>X</td>
</tr>
</tbody>
</table>

```
red matrix
A  B  C  D  E
A/B X X X X X
A/B X X X X X
C 0.55 X X X
D 1.65 1.7 X X
E 0.95 1.0 1.7 X
```

The nearest neighbors are A and B, so these are joined:

```
A
B
```

Average distances in the matrix to A/B; the nearest neighbors are then between A/B and C, so these are joined:

```
A
B
C
```

The nodes are completely resolved - you’re done!