BioE/MCB/PMB C146/246, Spring 2005

Problem Set 2

1. Actual tree relating species (not to scale):

```
           /\
          /  \
         /    \
        /\     \
       /  \     \
      /\   \    /\
     /  \  Sc  /\ \
    /   /\    Ec Hi Bs
   /\  Dm Ce
  /\ Fr
 /\ Mm
Hs Pt
```

A. a and c are clearly paralogs, because they are homologs within the same species and must be related by a duplication at some point in the past. The duplication probably occurred after the human/mouse (and human/fish) speciation. Then a and d, and a and e, are orthologs.

B. The evolutionary relationship of the three bacteria is different than the relationship of their genes, which points to a duplication event even though only one gene is seen in each. The duplication must have happened in a common ancestor of all three bacteria, before speciation. Then H. influenzae lost its copy of one gene, leaving a close relationship between j and k, and E. coli and B. subtilis lost their copies of the other gene, leaving l. So, j and k are orthologs, and j and l are paralogs.

If you just look at i, j, k, and l independent of the rest of the tree, then C. elegans has the same evolutionary relationship as its gene does, implying that it speciated before the duplication. So j and i are orthologs.

C. The overall structure implies that there was probably a duplication at the very top of the tree, with one copy lost in vertebrates after branching off from the fly/worm group, and the other lost in early prokaryotes. If so, then f and j, k, and l are paralogs.

2. Many alignments can produce the same score. Alignment matrices are found at the end of the solutions.

A. Global alignment, score -8:

```
EEHGWAGAEH
EAEHW--AP-
```

B. Local alignment, score 20:

```
EHGWA
EH-WA
```

C. Global alignment with no end gap penalties, score 17:

```
-EEHGWAGAEH
EAEH-WAP---
```

D. SSEARCH found the identical local alignment and score (20):

```
 EEHGWAGAEH
  :: ::
EAEH-WAP
```

3. Many different alignments can give the correct score. Alignment matrices at end of solutions.
A. Global alignment, score -34:

```
GCCCCGGTATATGCGTTAT
----C--TATGGGCG-CGT
```

A. Local alignment, score 18:

```
TATATGCG
TATGGGCG
```

4. Global alignment is useful when you are looking for similarity between two sequences of about the same length, such as two proteins that are suspected homologs. If the proteins are very different in length, then the necessary gaps may lower the score too much to reflect any regions of similarity.

Local alignment can help find conserved motifs or domains between proteins that may not be similar along their whole length, either because there has only been selective pressure on a part of the protein or because the domain is common to otherwise unrelated proteins.

Global alignment with no end gap penalties is useful when you are aligning sequences of significantly different lengths but with similarity along the length of the shorter sequence, such as an exon against a gene. This would have no end gap penalties in the shorter sequence, but not necessarily the longer. No end gap penalties for either sequence would be useful when aligning genome reads that overlap for much of their length.

5. The plot shows that the first halves of the two proteins are highly similar starting about 20 amino acids into the first and 30 into the second. The second half of each protein sequence contains a number of near–identical repeats of about 25 amino acids, separated by spaces of about 20 amino acids.

A dotplot of a protein against itself can show domain repeats in the protein as long diagonals running parallel to the main alignment. Shorter parallel diagonals might show repeated motifs. Also, a high density of dots scattered around the main diagonal show low complexity regions in the protein sequence. For DNA, a dotplot can show inverted sequences.

```
Problem 2 global alignment
                E      A      E      H      W      A      P
         0     -8    -16    -24    -32    -40    -48    -56
  E     -8      5     -3    -11    -19    -27    -35    -43
  E    -16     -3      4      2     -6    -14    -22    -30
  H    -24    -11     -4      4     10      2     -6    -14
  G    -32    -19    -11     -4      2      8      2     -6
  W    -40    -27    -19    -12     -6     13      5     -2
  A    -48    -35    -23    -20    -14      5     17      9
  G    -56    -43    -31    -25    -22     -3      9     15
  A    -64    -51    -39    -32    -27    -11      1      8
  E    -72    -59    -47    -34    -32    -19     -7      0
  H    -80    -67    -55    -42    -26    -27    -15     -8

    E A E H W A P
  0 - - - - - - -
E | \ - \ - - - -
E | \ \ \ - - - -
H | | | \ \ - - -
G | | \ | \ \ \ -
W | | | | \ \ \ \
A | | \ \ \ | \ -
G | | | \ \ | | \
A | | \ \ \ | \ \
E | \ | \ \ | | \
H | | | | \ | | |

local alignment
                E      A      E      H      W      A      P
         0      0      0      0      0      0      0      0
  E      0      5      0      5      0      0      0      0
  E      0      5      4      5      5      0      0      0
  H      0      0      3      4     13      5      0      0
  G      0      0      0      1      5     11      5      0
  W      0      0      0      0      0     16      8      1
  A      0      0      4      0      0      8     20     12
  G      0      0      0      2      0      0     12     18
  A      0      0      4      0      0      0      4     11
  E      0      5      0      9      1      0      0      3
  H      0      0      3      1     17      9      1      0

    E A E H W A P
  0 0 0 0 0 0 0 0
E 0 \ 0 \ \ 0 0 0
E 0 \ \ \ \ 0 0 0
H 0 \ \ \ \ - 0 0
G 0 0 \ \ | \ \ 0
```

```
W 0 0 0 0 0 \ \ \
A 0 0 \ 0 0 | \ -
G 0 0 \ \ 0 | | \
A 0 0 \ 0 \ 0 \ \
E 0 \ 0 \ - 0 0 \
H 0 \ \ | \ - - 0
```

```
no end gaps
                E      A      E      H      W      A      P
         0      0      0      0      0      0      0      0
  E      0      5     -1      5      0     -3     -1     -1
  E      0      5      4      4      5     -3     -4     -2
  H      0      0      3      4     12      4     -4     -6
  G      0     -2      0      1      4     10      4     -4
  W      0     -3     -5     -3     -1     15      7      0
  A      0     -1      1     -6     -5      7     19     11
  G      0     -2     -1     -1     -8     -1     11     17
  A      0     -1      2     -2     -3     -9      3     10
  E      0      5     -2      7     -1     -6     -5      2
  H      0      0      3     -1     15      7     -1     -6
```

```
    E A E H W A P
  0 0 0 0 0 0 0 0
E 0 \ \ \ \ \ \ \
E 0 \ \ \ \ \ \ \
H 0 \ \ \ \ - - \
G 0 \ \ \ | \ \ -
W 0 \ \ \ \ \ \ \
A 0 \ \ \ \ | \ -
G 0 \ \ \ \ | | \
A 0 \ \ \ \ | \ \
E 0 \ \ \ - \ | \
H 0 \ \ | \ - - |
```

```
problem 3 global alignment
            C    T    A    T    G    G    G    C    G    C    G    T
       0   -8  -16  -24  -32  -40  -48  -56  -64  -72  -80  -88  -96
G     -8   -4  -12  -18  -26  -28  -36  -44  -52  -60  -68  -76  -84
C    -16   -4   -6  -14  -20  -28  -32  -40  -40  -48  -56  -64  -72
C    -24  -12   -6  -10  -16  -24  -32  -36  -36  -44  -44  -52  -60
C    -32  -20  -14  -10  -12  -20  -28  -36  -32  -40  -40  -48  -54
C    -40  -28  -22  -18  -12  -16  -24  -32  -32  -36  -36  -44  -50
G    -48  -36  -30  -24  -20   -8  -12  -20  -28  -28  -36  -32  -40
G    -56  -44  -38  -32  -28  -16   -4   -8  -16  -24  -32  -32  -36
T    -64  -52  -40  -40  -28  -24  -12   -8  -10  -18  -26  -34  -28
A    -72  -60  -48  -36  -36  -30  -20  -14  -12  -12  -20  -28  -36
T    -80  -68  -56  -44  -32  -38  -28  -22  -16  -16  -14  -22  -24
A    -88  -76  -64  -52  -40  -34  -36  -30  -24  -18  -20  -16  -24
T    -96  -84  -72  -60  -48  -42  -38  -38  -32  -26  -20  -24  -12
G   -104  -92  -80  -68  -56  -44  -38  -34  -40  -28  -28  -16  -20
C   -112 -100  -88  -76  -64  -52  -46  -42  -30  -36  -24  -24  -18
G   -120 -108  -96  -84  -72  -60  -48  -42  -38  -26  -32  -20  -26
T   -128 -116 -104  -92  -80  -68  -56  -50  -44  -34  -28  -28  -16
T   -136 -124 -112 -100  -88  -76  -64  -58  -52  -42  -36  -32  -24
A   -144 -132 -120 -108  -96  -84  -72  -66  -60  -50  -44  -38  -32
T   -152 -140 -128 -116 -104  -92  -80  -74  -68  -58  -52  -46  -34
```

```
local alignment
             C    T    A    T    G    G    G    C    G    C    G    T
        0    0    0    0    0    0    0    0    0    0    0    0    0
  G     0    0    0    0    0    4    4    4    0    4    0    4    0
  C     0    4    0    0    0    0    0    0    8    0    8    0    2
  C     0    4    2    0    0    0    0    0    4    4    4    4    0
  C     0    4    2    0    0    0    0    0    4    0    8    0    2
  C     0    4    2    0    0    0    0    0    4    0    4    4    0
  G     0    0    0    0    0    4    4    4    0    8    0    8    0
  G     0    0    0    0    0    4    8    8    0    4    4    4    4
  T     0    0    4    0    4    0    0    4    6    0    2    0    8
  A     0    0    0    8    0    2    0    0    0    4    0    0    0
  T     0    0    4    0   12    4    0    0    0    0    2    0    4
  A     0    0    0    8    4   10    2    0    0    0    0    0    0
  T     0    0    4    0   12    4    6    0    0    0    0    0    4
  G     0    0    0    2    4   16    8   10    2    4    0    4    0
  C     0    4    0    0    0    8   12    4   14    6    8    0    2
  G     0    0    0    0    0    4   12   16    8   18   10   12    4
  T     0    0    4    0    4    0    4    8   14   10   16    8   16
  T     0    0    4    0    4    0    0    0    6   10    8   12   12
  A     0    0    0    8    0    2    0    0    0    4    6    6    8
  T     0    0    4    0   12    4    0    0    0    0    2    2   10
```