**Homework #2**

陈斯杰 Sijie Chen 2016310721

**Problem Description**

Write a program to compute dissimilarity dependent global pairwise alignment (banded DP), and output one optimal alignment with the alignment score. Your scoring function is . It should be noted that this is a distance scoring function, so the optimal alignment produces the minimum distance.

Your test examples are in HW1File1. Each student should submit the homework online (1) with a MS Word report with the idea of implementation, summary and analysis of results (the best k, alignment scores and running time), and an example alignment, (2) your source code and your binary executable file (for C/C++ users).

**Idea of Implementation**

The program implements a k-banded global alignment algorithm to compare sequences pair-wisely. The DP distance function takes matches, mismatches, and indels into consideration. The recursive relationship between sub-problem is:

dp[i][j]=min {dp[i-1][j-1]+score(a[i],b[j]),

dp[i][j-1]+indel,

dp[i-1][j]+indel,

}

After the dp matrix is constructed, the program trace back from the element dp[N][M]

Note that the program only compute the values near the left diagonal and the right diagonal, i.e. columns from to , where and  
.

The program sets every element in the DP matrix to INFINITY except for those who on the boundaries (dp[i][0]=i\*INDEL, dp[0][j]=j\*INDEL). As C++ does not support built-in INFINITY feature for int type, the program use 0x3f3f3f3f to represent INFINITY.

The program is written in C++11 with the help of Cod::Blocks IDE. The program is compiled with GCC-4.9.2 with –O3 optimization. The attached binary executable file requires MinGW environment to run.

**Result Analysis**

The program compares 3 DNA sequences pair-wisely and takes 0.825 seconds on average.

The length of 3 sequences are 1491, 1396, and 1492. Their length differences are 95, 96, and 1; therefore the K-value must be larger than 96.

We used several K to run the program and evaluated the result.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| K | Distance  (Seq1 v.s. Seq 2) | Distance  (Seq1 v.s. Seq 3) | Distance  (Seq2 v.s. Seq 3) | Time |
| 92 | NA | NA | NA | NA |
| 94 | 279 | 199 | NA | NA |
| 96 | 279 | 199 | NA | NA |
| 98 | 277 | 199 | 292 | 0.828 |
| 100 | 277 | 199 | 292 | 0.832 |
| 300 | 277 | 199 | 292 | 0.925 |
| 400 | 277 | 199 | 292 | 0.974 |

From what have been discussed above, we can draw a conclusion that K=98 is the optimal choice.

The optimal alignment scores of three comparisons are listed as the following content:

====================================================================

[S001061806 uncultured bacterium; orang2\_aai67f09; EU462707]

v.s.

[S000959403 uncultured rumen bacterium; YRC23; EU259399]

Alignment Distance=277 Alignment Length=1501

====================================================================

[S001061806 uncultured bacterium; orang2\_aai67f09; EU462707]

v.s.

[S000959403 uncultured rumen bacterium; YRC23; EU259399]

Alignment Distance=277 Alignment Length=1501

====================================================================

[S001061806 uncultured bacterium; orang2\_aai67f09; EU462707]

v.s.

[S001081077 uncultured bacterium; C89; EU622734]

Alignment Distance=292 Alignment Length=1497

**Attachment**

The attachment gives the result.

====================================================================

[S001061806 uncultured bacterium; orang2\_aai67f09; EU462707]

v.s.

[S000959403 uncultured rumen bacterium; YRC23; EU259399]

Alignment Distance=277 Alignment Length=1501

AGAGTTTGATCCTGGCTCAGGATGAACGCTAGCTACAGGCTTAACACATGCAAGTCGAGGGGCAGCATGA

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||.||.|.

AGAGTTTGATCCTGGCTCAGGATGAACGCTAGCTACAGGCTTAACACATGCAAGTCGAGGGGCATCAGGG

GAGGA-CTTCGGTCCTTTCGATGGCGACCGGCGAATGGGTGAGTAACACGTATCCAACCTGCCCCTGA-C

|||.| ||| |.|..|..||.||||||||||||.|.||||||||||||||||||||||||| ||||.| |

GAGTAGCTT-GCTATTCCCGCTGGCGACCGGCGCACGGGTGAGTAACACGTATCCAACCTG-CCCTCATC

CGGGGGATAGCCCTCCGAAAGGAGAATTAACACCCCATGGGTGTCCGTGC-GG-CATCGCGCGGCCATGA

|.|||||||||||||.||||||||.|||||.||||.|||.| |||..|.| || ||||..|.....|..|

CCGGGGATAGCCCTCTGAAAGGAGGATTAATACCCGATGCG-GTCATTTCGGGACATCCTGTTATGACTA

AAGGTT-A-CGG-TCAGGGATGGGGATGCGTCCGATTAGCTTGCTGGCGGGGTAACGGCCCACCAGGGCG

|||.|| | ||| |.| |||||||||||||||.||||||||||.|||||||||||||||||||||.|||.

AAGATTCATCGGATGA-GGATGGGGATGCGTCTGATTAGCTTGTTGGCGGGGTAACGGCCCACCAAGGCA

TCGATCGGTAGGGGTTCTGAGAGGAAGGTCCCCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGG

||||||.||||||||||||||||||||||||||||||..||||||||||||||||||..|||||||||||

TCGATCAGTAGGGGTTCTGAGAGGAAGGTCCCCCACATAGGAACTGAGACACGGTCCTAACTCCTACGGG

AGGCAGCAGTGAGGAATATTGGTCAATGGGCGGAAGCCTGAACCAGCCAAGTAGCGTGAAGGATGACGGC

||||||||||||||||||||||||||||||||..|||||||||||||||||||||||||.|||.|||.||

AGGCAGCAGTGAGGAATATTGGTCAATGGGCGTGAGCCTGAACCAGCCAAGTAGCGTGAGGGACGACTGC

CCTACGGGTTGTAAACTTCTTTTATGCGGGAACAAAGTGCGCCACGCGTGGCGTTTTGCGCGTACCGCAG

||||.|||||||||||.|||||||||||||||.||.|..|||.|||.||.|||...|||..||||||||.

CCTATGGGTTGTAAACCTCTTTTATGCGGGAATAACGGTCGCTACGTGTAGCGGCGTGCATGTACCGCAT

GAAAAAGCACCGGCTAATTCCGTGCCAGCAGCCGCGGTAATACGGAAGGTGCGAGCGTTATCCGGATTCA

|||.|||||.||||||||||||||||||||||||||||||||||||||.|||||||||||||||||||.|

GAATAAGCATCGGCTAATTCCGTGCCAGCAGCCGCGGTAATACGGAAGATGCGAGCGTTATCCGGATTTA

TTGGGTTTAAAGGGAGCGTTGGCGGAGCGC-CAAGTCAGCTGTGAAAT-CCCGCGGCTCAACCGTGGAAC

|||||||||||||||||||.|||||| ||| .|||.|.||.||.|||| .|.|| ||.||||..||| .|

TTGGGTTTAAAGGGAGCGTAGGCGGA-CGCTTAAGCCTGCGGTAAAATGTCAGC-GCCCAACGCTGG-CC

TGCAGTTGAAACTGGCGCCCTTG--TGTGCACATGAGGATGGTGGAATTCGTGGTGTAGCGGTGAAATGC

.||.||.|.||||||....|||| |||||||| ||||.||||||||||||||||||||||||||||||

CGCCGTGGGAACTGGGTGTCTTGAATGTGCACA--AGGAAGGTGGAATTCGTGGTGTAGCGGTGAAATGC

TTAGATATCACGAAGAACTCCGATTGCGAAGGCAGCTGTCTGGGGTGCCACTGACGCTGAGGCTCGAAAG

||||||||||||||||||||||||||||||||||||..|||||||....|.|||||||||||||||||||

TTAGATATCACGAAGAACTCCGATTGCGAAGGCAGCCTTCTGGGGCATGATTGACGCTGAGGCTCGAAAG

TGCGGGTATCAAACAGGATTAGATACCCTGGTAGTCCGCACAGTAAACGATGGATACTCGTGGTCGGCGA

||||||||||.|||||||||||||||||||||||||||||||||||||||||..|.|.||...|||||||

TGCGGGTATCGAACAGGATTAGATACCCTGGTAGTCCGCACAGTAAACGATGAGTGCCCGCTCTCGGCGA

CACACTGCCGGTCACCAAGCGAAAGCGATAAGTATCCCACCTGGGGAGTACGCCGGCAACGGTGAAACTC

||.||.|.|||...||||||||||||..||||.|..||||||||||||||||||||||||||||||||||

CAGACAGTCGGGGGCCAAGCGAAAGCATTAAGCACTCCACCTGGGGAGTACGCCGGCAACGGTGAAACTC

AAAGGAATTGACGGGGGCCCGCACAAGCGGAGGAACATGTGGTTTAATTCGATGTTACGCGAGGAACCTT

||||||||||||||||||||||||||||||||||||||||||||||||||||||.|||||||||||||||

AAAGGAATTGACGGGGGCCCGCACAAGCGGAGGAACATGTGGTTTAATTCGATGATACGCGAGGAACCTT

ACCCGGGCTTGAACTGCA-AGGAC---CGTCCGAGA-GATCGGTCTTCCCTTCGGGGCCTTGGCGGAGGT

||||||||||||||||.| |.||| .|.|.|||| |..||.|| ||||....||||.|...|||||||

ACCCGGGCTTGAACTGGACATGACGTTGGGCAGAGACGCCCGATC-TCCCGCAAGGGCATGCTCGGAGGT

GCTGCATGGTTGTCGTCAGCTCGTGCCGTGAGGTGTCGGCTTAAGTGCCATAACGAGCGCAACCCCTGTC

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||..|.

GCTGCATGGTTGTCGTCAGCTCGTGCCGTGAGGTGTCGGCTTAAGTGCCATAACGAGCGCAACCCCCCTG

TTCGGTTGCCATCGGGTAATGCCGGGCACTCCGGAGATACTGCCATCGCAAGATGTGAGGAAGGTGGGGA

|...|||||||||.||||||||.||||||||.|.|.||||||||||||||||||||||||||||||||||

TGTAGTTGCCATCAGGTAATGCTGGGCACTCTGCACATACTGCCATCGCAAGATGTGAGGAAGGTGGGGA

TGACGTCAAATCAGCACGGCCCTTACGTCCGGGGCTACACACGTGTTACAATGGGAGGTACAGAAGGCTG

|||||||||||||||||||||||||||||||||||||||||||||||||||||||.||.|||||..||||

TGACGTCAAATCAGCACGGCCCTTACGTCCGGGGCTACACACGTGTTACAATGGGGGGCACAGAGAGCTG

CTACCCGGCGACGGGATGCCAATCCCCAAATCCTCTCTCAGTTCGGATCGGAGTCTGCAACCCGACCCCG

||.|..||||||..|..||.||||...||||||.|||||||||||||..|||||||||||||||||.||.

CTGCATGGCGACATGCGGCGAATCTTGAAATCCCCTCTCAGTTCGGACTGGAGTCTGCAACCCGACTCCA

TGAAGCTGGATTCGCTAGTAATCGCGCATCAGCCATGGCGCGGTGAATACGTTCCC--G----G------

.||||||||||||||||||||||||||||||||||||||||||||||||||||||| | |

CGAAGCTGGATTCGCTAGTAATCGCGCATCAGCCATGGCGCGGTGAATACGTTCCCGGGCCTTGTACACA

--G-CC-T-------T-----------G---C---A--CT-----CA------CC------G---C----

| || | | | | | || || || | |

CCGCCCGTCAAGCCATGAAAGCCGGGGGCACCTGAAGGCTGCGGCCATCGCGGCCTAGGGTGAAACTGGT

====================================================================

[S000959403 uncultured rumen bacterium; YRC23; EU259399]

v.s.

[S001081077 uncultured bacterium; C89; EU622734]

Alignment Distance=199 Alignment Length=1503

AGAGTTTGATCCTGGCTCAGGATGAACGCTAGCTACAGGCTTAACACATGCAAGTCGAGGGGCATCAGGG

.||.||||||||||||..|||||||||||||||||||||||||||||||||||||||||||||.||.|.

-TAGATTGATCCTGGCTAGGGATGAACGCTAGCTACAGGCTTAACACATGCAAGTCGAGGGGCAGCATGA

GAGTAGCTTGCTATTCCCGCTGGCGACCGGCGCACGGGTGAGTAACACGTATCCAACCTGCCCTCATCCC

...||||||||||.....|.||||||||||||||||||||||||||.||||||||||.||.|||...|.|

TCTTAGCTTGCTAAGGTTGATGGCGACCGGCGCACGGGTGAGTAACGCGTATCCAACTTGGCCTTTACTC

GGGGATAGCCCTCTGAAAGGAGGATTAATACCCGATGCGGTCAT--TTCGGGACATCCTGTTATGACTAA

.|||||||||.|..|||||.|.|||||||||..||||...|..| ||| .| |||....||||.|.|||

CGGGATAGCCTTTCGAAAGAAAGATTAATACGGGATGGTTTTGTGGTTC-CG-CATGGGATTATAAATAA

AGATTCATCGGATGAGGATGGGGATGCGTCTGATTAGCTTGTTGGCGGGGTAACGGCCCACCAAGGCATC

|||||.|||||...|||||||||||||||...|||||...||||||||||||||||||||||||..||||

AGATTTATCGGTAAAGGATGGGGATGCGTTCCATTAGGCAGTTGGCGGGGTAACGGCCCACCAAACCATC

GATCAGTAGGGGTTCTGAGAGGAAGGTCCCCCACATAGGAACTGAGACACGGTCCTAACTCCTACGGGAG

|||...||||||||||||||||||||||||||||||.|||||||||||.||||||.||||||||||||||

GATGGATAGGGGTTCTGAGAGGAAGGTCCCCCACATTGGAACTGAGACGCGGTCCAAACTCCTACGGGAG

GCAGCAGTGAGGAATATTGGTCAATGGGCGTGAGCCTGAACCAGCCAAGTAGCGTGAGGGACGACTGCCC

||||||||||||||||||||||||||||||.||||||||||||||||||||||||||.|||.||||||||

GCAGCAGTGAGGAATATTGGTCAATGGGCGAGAGCCTGAACCAGCCAAGTAGCGTGAAGGAAGACTGCCC

TATGGGTTGTAAACCTCTTTTATGCGGGAATAACGGTCGCTACGTGTAGCGGCGTGCATGTACCGCATGA

||||||||||||||.||||||||..||||||||.|..|||.||||||.|.|...||.|||||||..||||

TATGGGTTGTAAACTTCTTTTATAAGGGAATAAAGAGCGCCACGTGTGGTGTGTTGTATGTACCTTATGA

ATAAGCATCGGCTAATTCCGTGCCAGCAGCCGCGGTAATACGGAAGATGCGAGCGTTATCCGGATTTATT

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

ATAAGCATCGGCTAATTCCGTGCCAGCAGCCGCGGTAATACGGAAGATGCGAGCGTTATCCGGATTTATT

GGGTTTAAAGGGAGCGTAGGC-GGACGCTTAAGCCTGCGGTAAAATGTCAG-CGCCCAACGCTGGCCCGC

||||||||||||||||||||| |||.| |||||.|.|||||.||| ..|.| .||.||||..|.|..|||

GGGTTTAAAGGGAGCGTAGGCGGGAAG-TTAAGTCAGCGGTCAAA-CACGGTTGCTCAACAATCGTTCGC

CGTGGGAACTGGGTGTCTTGAATGTGCACAAGGAAGGTGGAATTCGTGGTGTAGCGGTGAAATGCTTAGA

|||.|.|||||..|.||||||||||...|||||.||||||||||||||||||||||||||||||||||||

CGTTGAAACTGACTTTCTTGAATGTAGTCAAGGCAGGTGGAATTCGTGGTGTAGCGGTGAAATGCTTAGA

TATCACGAAGAACTCCGATTGCGAAGGCAGCCTTCTGGGGCATGATTGACGCTGAGGCTCGAAAGTGCGG

|||||||||||||||||||.|||||||||||||.||||.|.|||||||||||||||||||||||||||||

TATCACGAAGAACTCCGATAGCGAAGGCAGCCTGCTGGAGTATGATTGACGCTGAGGCTCGAAAGTGCGG

GTATCGAACAGGATTAGATACCCTGGTAGTCCGCACAGTAAACGATGAGTGCCCGCTCTCGGCGACAGAC

|.|||.||||||||||||||||||||||||||||||.|||||||||||.|||.||..||..||||.|.|.

GAATCAAACAGGATTAGATACCCTGGTAGTCCGCACCGTAAACGATGAATGCTCGTCCTTTGCGATACAA

AGTCGGGGGCCAAGCGAAAGCATTAAGCACTCCACCTGGGGAGTACGCCGGCAACGGTGAAACTCAAAGG

.||..|||.||||||||||||||||||||.||||||||||||||||||||||||||||||||||||||||

TGTAAGGGACCAAGCGAAAGCATTAAGCATTCCACCTGGGGAGTACGCCGGCAACGGTGAAACTCAAAGG

AATTGACGGGGGCCCGCACAAGCGGAGGAACATGTGGTTTAATTCGATGATACGCGAGGAACCTTACCCG

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

AATTGACGGGGGCCCGCACAAGCGGAGGAACATGTGGTTTAATTCGATGATACGCGAGGAACCTTACCCG

GGCTTGAACTGGACATGACGTTGGGCAGAGACGCCCGATCTCCCGCAAGGGCATGCT-CGGAGGTGCTGC

||||||||.||.|..||||| .|..||||||.|..| .| ||||....|||| .||| .|.|||||||||

GGCTTGAATTGCAGCTGACG-GGTCCAGAGATGGAC-CT-TCCCTTCGGGGC-GGCTGTGAAGGTGCTGC

ATGGTTGTCGTCAGCTCGTGCCGTGAGGTGTCGGCTTAAGTGCCATAACGAGCGCAACCCCCCTGTGTAG

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||...|...|||

ATGGTTGTCGTCAGCTCGTGCCGTGAGGTGTCGGCTTAAGTGCCATAACGAGCGCAACCCTTGTCCATAG

TTGCCATCAGGTAATGCTGGGCACTCTGCACATACTGCCATCGCAAGATGTGAGGAAGGTGGGGATGACG

||||||||.||||||||.||||||||||...|.|||||||||||||||||||||||||||||||||||||

TTGCCATCGGGTAATGCCGGGCACTCTGTGGAGACTGCCATCGCAAGATGTGAGGAAGGTGGGGATGACG

TCAAATCAGCACGGCCCTTACGTCCGGGGCTACACACGTGTTACAATGGGGGGCACAGAGAGCTGCT-GC

|||||||||||||||||||||||||||||||||||||||||||||||||||||.||||||.||.||| ||

TCAAATCAGCACGGCCCTTACGTCCGGGGCTACACACGTGTTACAATGGGGGGTACAGAGGGCCGCTCGC

ATGGCGACATGC-GGCGAATCTTGAAATCCCCTCTCAGTTCGGACTGGAGTCTGCAACCCGACTCCACGA

.||.||| .|| .||.||||||.|||.||||||||||||||||||||||||||||||||||||||||||

-CGGTGAC-GGCATGCCAATCTTTAAAACCCCTCTCAGTTCGGACTGGAGTCTGCAACCCGACTCCACGA

AGCTGGATTCGCTAGTAATCGCGCATCAGCCATGGCGCGGTGAATACGTTCCCGGGCCTTGTACACACCG

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

AGCTGGATTCGCTAGTAATCGCGCATCAGCCATGGCGCGGTGAATACGTTCCCGGGCCTTGTACACACCG

CCCGTCAAGCCATGAAAGCCGGGGGCACCTGAAGGCTGCGGCC--ATCG--CGGCCTAGGGTGAAACTGG

||||||||||||||||||||.|.||.||||||||.|.|.|.|| |..| |||||||||||.||.|.||

CCCGTCAAGCCATGAAAGCCTGTGGTACCTGAAGTCCGTGACCGTAAGGATCGGCCTAGGGTAAAGCCGG

====================================================================

[S001061806 uncultured bacterium; orang2\_aai67f09; EU462707]

v.s.

[S001081077 uncultured bacterium; C89; EU622734]

Alignment Distance=292 Alignment Length=1497

AGAGTTTGATCCTGGCTCAGGATGAACGCTAGCTACAGGCTTAACACATGCAAGTCGAGGGGCAGCATGA

.||.||||||||||||..|||||||||||||||||||||||||||||||||||||||||||||||||||

-TAGATTGATCCTGGCTAGGGATGAACGCTAGCTACAGGCTTAACACATGCAAGTCGAGGGGCAGCATGA

GAGGA-CTTCGGTCCTTTCGATGGCGACCGGCGAATGGGTGAGTAACACGTATCCAACCTGCCCCTGACC

....| ||| |.|....|.||||||||||||||.|.|||||||||||.||||||||||.||.||.|.||.

TCTTAGCTT-GCTAAGGTTGATGGCGACCGGCGCACGGGTGAGTAACGCGTATCCAACTTGGCCTTTACT

GGGGGATAGCCCTCCGAAAGGAGAATTAACACCCCATGGGTGT-CCGTGCGGCATCGCGCGGCCATGAAA

..|||||||||.|.||||||.|..|||||.||...||||.|.| ..||.|.||||.|.......|..|||

CCGGGATAGCCTTTCGAAAGAAAGATTAATACGGGATGGTTTTGTGGTTCCGCATGGGATTATAAATAAA

G-GTTA-CGGTCAGGGATGGGGATGCGTCCGATTAGCTTGCTGGCGGGGTAACGGCCCACCAGGGCGTCG

| .||| ||||.|.||||||||||||||.|.|||||...|.|||||||||||||||||||||...|.|||

GATTTATCGGTAAAGGATGGGGATGCGTTCCATTAGGCAGTTGGCGGGGTAACGGCCCACCAAACCATCG

ATCGGTAGGGGTTCTGAGAGGAAGGTCCCCCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGG

||.|.|||||||||||||||||||||||||||||.||||||||||||.|||||||.||||||||||||||

ATGGATAGGGGTTCTGAGAGGAAGGTCCCCCACATTGGAACTGAGACGCGGTCCAAACTCCTACGGGAGG

CAGCAGTGAGGAATATTGGTCAATGGGCGGAAGCCTGAACCAGCCAAGTAGCGTGAAGGATGACGGCCCT

|||||||||||||||||||||||||||||..|||||||||||||||||||||||||||||.|||.|||||

CAGCAGTGAGGAATATTGGTCAATGGGCGAGAGCCTGAACCAGCCAAGTAGCGTGAAGGAAGACTGCCCT

ACGGGTTGTAAACTTCTTTTATGCGGGAACAAAGTGCGCCACGCGTGGCGTTTTGCGCGTACCGCAGGAA

|.||||||||||||||||||||..|||||.||||.||||||||.||||.||.|||...|||||..|.|||

ATGGGTTGTAAACTTCTTTTATAAGGGAATAAAGAGCGCCACGTGTGGTGTGTTGTATGTACCTTATGAA

AAAGCACCGGCTAATTCCGTGCCAGCAGCCGCGGTAATACGGAAGGTGCGAGCGTTATCCGGATTCATTG

.|||||.||||||||||||||||||||||||||||||||||||||.|||||||||||||||||||.||||

TAAGCATCGGCTAATTCCGTGCCAGCAGCCGCGGTAATACGGAAGATGCGAGCGTTATCCGGATTTATTG

GGTTTAAAGGGAGCGTTGGCGGAGCGCCAAGTCAGCTGTGAAATCCCGCGGCTCAACCGTGGAACTGCAG

||||||||||||||||.|||||...|..||||||||.||.|||..|.|..|||||||..|.|..| ||.|

GGTTTAAAGGGAGCGTAGGCGGGAAGTTAAGTCAGCGGTCAAACACGGTTGCTCAACAATCGTTC-GCCG

TTGAAACTGGCGCCCTTGTGTGCACATGAGGATGGTGGAATTCGTGGTGTAGCGGTGAAATGCTTAGATA

|||||||||.|...||||..||.|....|||..|||||||||||||||||||||||||||||||||||||

TTGAAACTGACTTTCTTGAATGTAGTCAAGGCAGGTGGAATTCGTGGTGTAGCGGTGAAATGCTTAGATA

TCACGAAGAACTCCGATTGCGAAGGCAG-CTGTCTGGGGTGCCACTGACGCTGAGGCTCGAAAGTGCGGG

|||||||||||||||||.|||||||||| ||| ||||.||...|.|||||||||||||||||||||||||

TCACGAAGAACTCCGATAGCGAAGGCAGCCTG-CTGGAGTATGATTGACGCTGAGGCTCGAAAGTGCGGG

TATCAAACAGGATTAGATACCCTGGTAGTCCGCACAGTAAACGATGGATACTCGTGGTCGGCGACACACT

.||||||||||||||||||||||||||||||||||.||||||||||.||.|||||..|..||||.|||.|

AATCAAACAGGATTAGATACCCTGGTAGTCCGCACCGTAAACGATGAATGCTCGTCCTTTGCGATACAAT

GCCGGTCACCAAGCGAAAGCGATAAGTATCCCACCTGGGGAGTACGCCGGCAACGGTGAAACTCAAAGGA

|...|..|||||||||||||..||||.||.||||||||||||||||||||||||||||||||||||||||

GTAAGGGACCAAGCGAAAGCATTAAGCATTCCACCTGGGGAGTACGCCGGCAACGGTGAAACTCAAAGGA

ATTGACGGGGGCCCGCACAAGCGGAGGAACATGTGGTTTAATTCGATGTTACGCGAGGAACCTTACCCGG

||||||||||||||||||||||||||||||||||||||||||||||||.|||||||||||||||||||||

ATTGACGGGGGCCCGCACAAGCGGAGGAACATGTGGTTTAATTCGATGATACGCGAGGAACCTTACCCGG

GCTTGAACTGCA-AGGAC-CGTCCGAGAGATCGGTCTTCCCTTCGGGGCCTTGGCGGAGGTGCTGCATGG

|||||||.|||| ..||| .|||| ||||||.|..||||||||||||||....|.|.|||||||||||||

GCTTGAATTGCAGCTGACGGGTCC-AGAGATGGACCTTCCCTTCGGGGCGGCTGTGAAGGTGCTGCATGG

TTGTCGTCAGCTCGTGCCGTGAGGTGTCGGCTTAAGTGCCATAACGAGCGCAACCCCTGTCTTCGGTTGC

||||||||||||||||||||||||||||||||||||||||||||||||||||||||.||||....|||||

TTGTCGTCAGCTCGTGCCGTGAGGTGTCGGCTTAAGTGCCATAACGAGCGCAACCCTTGTCCATAGTTGC

CATCGGGTAATGCCGGGCACTCCGGAGATACTGCCATCGCAAGATGTGAGGAAGGTGGGGATGACGTCAA

||||||||||||||||||||||.|..||.|||||||||||||||||||||||||||||||||||||||||

CATCGGGTAATGCCGGGCACTCTGTGGAGACTGCCATCGCAAGATGTGAGGAAGGTGGGGATGACGTCAA

ATCAGCACGGCCCTTACGTCCGGGGCTACACACGTGTTACAATGGGAGGTACAGAAGGCTGCTACCCGGC

||||||||||||||||||||||||||||||||||||||||||||||.||||||||.|||.|||..||||.

ATCAGCACGGCCCTTACGTCCGGGGCTACACACGTGTTACAATGGGGGGTACAGAGGGCCGCTCGCCGGT

GACGGGATGCCAATCCCCAAATCCTCTCTCAGTTCGGATCGGAGTCTGCAACCCGACCCCGTGAAGCTGG

|||||.|||||||||...|||.||.|||||||||||||..|||||||||||||||||.||..||||||||

GACGGCATGCCAATCTTTAAAACCCCTCTCAGTTCGGACTGGAGTCTGCAACCCGACTCCACGAAGCTGG

ATTCGCTAGTAATCGCGCATCAGCCATGGCGCGGTGAATA------C--G---T-T---------CC--C

|||||||||||||||||||||||||||||||||||||||| | | | | || |

ATTCGCTAGTAATCGCGCATCAGCCATGGCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTC

--G----G---GCCT-T-G---C---A---C-T---C------A----CC---------G-C--------

| | |||| | | | | | | | | || | |

AAGCCATGAAAGCCTGTGGTACCTGAAGTCCGTGACCGTAAGGATCGGCCTAGGGTAAAGCCGGTGATTG