Santhanakrishnan Ramani

SID: 105720585

**Bioinformatics & Genomics**

**Assignment – 2**

1. a) The mutation observed is presence of both **T** and **C** at the same position in the sequence of J. Deen, meaning

he can have any of the two sequence below.

Wild-type - TGAAGAACCGTTCAG**C**CAATTCTAG

J.Deen - TGAAGAACCGTTCAG**T**CAATTCTAG

75% confident based on the gel given, & will confirm the mutation using a Visualization tool like IGV.

b) Mutation alters the protein sequence.

Protein sequence encoded on the reverse complement of J.Deen’s Sequence is **LELTERFF**

1. a) Yes the sequence in general is of good quality, as 46 bases have a high quality score of 40.

b) The base A in the 43rd position has the lowest quality. The probability of this position – (1/40)

c) Under the (Phred + 64) the sequence in general is not of good quality. The worst position is remains the same as in question b.

3)

 

1. a) Maximally Scoring Alignments



b) DP matrix is generated using Needleman-Wunsch. Because the alignment here is global and didn’t

see any restart in the matrix and scores are negative in the DP matrix.

c) For this DP matrix the gap penalty is linear. We can confirm this by looking at the first column and row of the

matrix as the value increase by -2 as it moves to next column or row.

d) Scoring matrix of the above DP



1. a) (i) Hsap3 and Mmus 1 - Orthologs  
    (ii) Hsap2 and Mmus 2 - Paralogs

b) Statement “Using the BLOSSUM40 matrix, we determined that our proteins are 70% homologous.” Is Wrong

as the BLOSSUM40 matrix gives you the score for similarity between gene sequences and there is no way to

tell they are homologous from this.

1. a) BLOSUM 30 seems most appropriate as it has the highest score and highest similarity rate and lowest gap.

|  |  |
| --- | --- |
| MATRIX | SCORE |
|  |  |
| BLOSUM30 | 325 |
| BLOSUM35 | 263 |
| BLOSUM40 | 307 |
| BLOSUM45 | 198 |
| BLOSUM50 | 183 |
| BLOSUM55 | 197 |
| BLOSUM60 | 82 |
| BLOSUM62 | 84 |
| BLOSUM65 | 77 |
| BLOSUM70 | 42 |
| BLOSUM75 | 26 |
| BLOSUM80 | 138 |
| BLOSUM85 | 11 |
| BLOSUM90 | 9 |
| BLOSUM Clustered | 8 |

b)

########################################

# Program: water

# Rundate: Fri 12 Feb 2016 00:04:05

# Commandline: water

# -auto

# -stdout

# -asequence emboss\_water-I20160212-000404-0567-94890201-oy.asequence

# -bsequence emboss\_water-I20160212-000404-0567-94890201-oy.bsequence

# -datafile EBLOSUM80

# -gapopen 25.0

# -gapextend 5.0

# -aformat3 pair

# -sprotein1

# -sprotein2

# Align\_format: pair

# Report\_file: stdout

########################################

#=======================================

#

# Aligned\_sequences: 2

# 1: EMBOSS\_001

# 2: EMBOSS\_001

# Matrix: EBLOSUM80

# Gap\_penalty: 25.0

# Extend\_penalty: 5.0

#

# Length: 118

# Identity: 44/118 (37.3%)

# Similarity: 66/118 (55.9%)

# Gaps: 3/118 (2.5%)

# Score: 252.0

#

#

#=======================================

EMBOSS\_001 163 EYPVDGSLVGLQSALRVDAFIPILPLIAEMKTGSYKRDHELALAGYALAF 212

||.|||:.:|:...|.||.... .:|.:.|||:.:..|:|::.|||||.

EMBOSS\_001 200 EYRVDGTPLGMSQNLSVDVISD--SVIIDFKTGAPRDFHKLSITGYALAL 247

EMBOSS\_001 213 ESQYEIPVDFGYLCYVNVIEGKIHNNCRLIVISDTLRQEFVEVRDRALRA 262

|:.||.|.|:|.|.|:|..|.. ....:.:.||:|||:.|:|.||..:..

EMBOSS\_001 248 EAAYETPRDYGLLIYINNPEDP-RITYKPVYISNTLRRLFIEERDNIIDM 296

EMBOSS\_001 263 IDDDVDPGLAKKCSADCP 280

:.:|.:|.....|...||

EMBOSS\_001 297 LLEDAEPPKDLNCQPTCP 314

#---------------------------------------

#---------------------------------------