

LISTA DE EXERCÍCIOS II - Parte 2

Instruções:

- A resolução do exercício deve ser feita **individualmente**. Cópias evidentes entre trabalhos não serão aceitas.
- A entrega deve ser online via Moodle (exclusivamente), somente até a data especificada.
- Para cada uma das tarefas deve-se entregar o código fonte. O nome do arquivo deve identificar a tarefa, exemplo "e2-1a.py" referente ao item "1a" da tarefa. Arquivos corrompidos serão desconsiderados.
- Além do código fonte deve-se entregar um único arquivo PDF com o nome "e2-i1.pdf" apresentando o pseudocódigo do algoritmo desenvolvido e os resultados encontrados.
- Não serão aceitos trabalhos atrasados.
- Data de entrega: 19.09.2019 (quinta-feira) até as 23:55 via Moodle (<https://moodle.ufrgs.br/login/index.php>).

NOME: CARTÃO:

1. Considere as seguintes seqüências:

```
>gi|110319967|emb|CAJ44466.1| haemoglobin type 1 [Biomphalaria glabrata]
MFVLKGSVVQAFVLLSIVCLEITIADDGVRVYVNAEWKRPEQSQEGRHSCTARRLEDNSEEACSTEVKFR
QRAPAEYANKIKKAKDKLRRLESQFDDCCQENDRKDRILQLQANLTDTHRLVTDSDIQALRSSWATLTA
GADGRNNGFNNFVLLNTIPNIRERFEKFNHQSDEALKNDNEFVKQVKLIVGGLQSFIDNLENPGQLQ
ATIERLASVHLKMRPTIGLEYFRPLQENIAQYVASALGVGADDAAPKAWERLLNANFNEVLNSFANYNIGL
SDTDKVALQSSWSRLTAGADGKRNAGVRLVLWMFNVPNMRRERFTKFNAQSDEALKTDAEFLKQVDAII
GGFETLINLNDADLLNRLSLADEHLEKKPAISSNYFGPLQKNIHLFIEGTLNFGSDSDEARAWTHLV
GALNKVIKDHAIHNLGLSDIDRDALVSSWNQLTGRAGGSRNAGTNLVLWMLNVPNMRRDQFSKFNAQSD
DNLKDAEFVRQVDLITGGLESVDNVNPIFLQEALVRLADAHLNLKPRVGLEYFGPLQRYIHAYIEKA
LGVSADSAAPRAWDLLTAFNNVLKDRFTFLRIVSDDRRALQSSWSRLQSQAGNKEAGIKLVTLFDNV
PNMRDRFTKFNAHSSDEALRANNEFLRQVDVIVGGLDSLNNVDNSDNFQAAIERLVDVHLHMSFVSGLE
YFGPLQQNIRSYIQNALGVAADSAEARSWTNLFTAFNEFLADHTIQIKLSPTDRKVLDRTWKQLTSGGK
QEAGVKLVLMFQVPMNRDQFSKFDARKSEAEALRNDAEFINQVNNIVGGLDSILNNDKPGQLQAALER
LADYHLDHKPRIGLEFFGGLQKYIHLIESALNVAVGSEESRAWTDLLTALNKVIRDHAIIDRLGLSDNDR
EADSSWKKLRSGAGGRRNAGIKLVQWMLRTIPNMRRDRFNKFDKQSDAALQRDPEFLAQVDRLGGVES
LVNNVDDPVALKAAIDRLADAHLSFDPRVGLDYFGPLQTYIHDYIEEALGVGADSDEAKGWTDLFAAFNK
VLKERTVLKIVSDNERAALRSSWDSLKSAAGGTQEAGVRLVLWMLQNVPMRRERFNKFNALQGDDALRAD
AEFVKQVERIAGGLESINNVDNQGKLQAAIDRLVNAHLNFRPSVGLYFQPLQENIYKYLESAALIVAAD
SNEAKAWTHLLSAFNTVLREHSLEKIGLSDVDRKALESSWKKLTDAAGGSENAGTNLVFWLLDNVPMNRD
RFTKFNAHQSNAALQDDDEFNRQVRAITRGIESFVNNVNNPAALQSSIENLVDAHLNFQPSIGLSYFGSV
QQYIHLIYIAKALGVASNSDEAKSWTNLFAAFNKVLKEHSLEKIGISDSKRALVSSWKKLTAGGRQNFV
DLVLWMFNVPNMREQFTKFDKQSDADLRDPNFKQVNNIINGVGLVDVSDNPGKLQANLERLTDVH
LHFVPSVGPEFFGGLQKNIHTFIEQALGVGADSDEPKAWTDLIGAFNKVLNDHAIQIHLSETDRRALDS
SWKRLTAGENGQKAGVNLVLWFNNIPNMRRERFTKFDANQADDALRADPEFQKQVNVIVGGLKSFLDSV
NDPIALQANMDRVAEHLSDMPVVGVPYFALSQNIHRFIEISLGVTADESQAWTDLLAGFTRVVRNR
AVLRKVSDDSKSAFVSSWNLIRKAASRRNAGVNLVLWLFNNVPNMRRNHTKFNQNPDAALRNDQEFNL
QVDRIAGGLESVKNVNNPARFLDALERLSSAHLNMKPSIGLEYFGPLQKNIHTYIESALGVAAGSDEAN
```

AWTDVFGAFNEILKYSSVEKIGLSDSDKQALTSSWSSLIAEGKDTAGEKLVLMFNDVNPMMRARFTKFDA
 TQSDEALRNDKEFRNQVNVIVGGLELLINSVNEPGQLQANLEKLVDDHLMVPSVGLLEYFGPLQKYIHLF
 IEKALGVSSNSVESKSWTNLLRAFNKVLKEHSVKKIGLSEEDRKAIVSSWRKLVGRAGGRDNAGTNLVLW
 MFENVPNMRDRFTKFNAYQPDALRQDREFGAQVDRTSGLESLVNNVENPGQFQAALERLSTLHKNKTP
 SVGLQYFGPLQRYIHLIYIEQNLNVASDSVESRAWTNLFAFNEVLKKA

>gi|19549331|ref|NP599030.1| cytoglobin [Homo sapiens]
 MEKVPGEMEIERRERSEELSEAERKAVQAMWARLYANCEDVGVAAILVRFFVNFPSAKQYFSQFKHEDPL
 EMERSPQLRKHACRVMGALNTVVENLHDPDKVSSVLALVGKAHALKHKVEPVYFKILSGVILEVVAEEFA
 SDFPPETQRAWAKLRGLIYSHVTAAYKEVGWVQQVPNATTPPATLPSSGP



Figure 1: Biomphalaria Glabrata (esquerda) e Homo Sapiens (direita).

Implemente um conjunto de rotinas computacionais para realizar as tarefas que seguem:

- Desenvolva um algoritmo que leia duas seqüências n e m .
- Implemente e utilize o algoritmo de Smith-Waterman para alinhar estas seqüências.
- O programa deve imprimir na tela a tabela final de alinhamento, as duas cadeias alinhadas e identidade do alinhamento.
- Regras:
 - Valor para GAP: -2
 - Valor para Match: +1
 - Valor para Mismatch: -1
 - $S_{i,j} = \text{MAX} \begin{bmatrix} 0 \\ S_{i-1,j-1} + s(a_i, b_j) \\ S_{i,j-1} + \text{gap}(\text{seq1}) \\ S_{i-1,j} + \text{gap}(\text{seq2}) \end{bmatrix}$
- O que podemos concluir a partir do alinhamento?

2. Considere as seguintes seqüências:

Organismo MusMusculus – PKp2.

ATAGCAGAGTCTGGGAGGCTGTCCGCTTTTCAGTTTGTCTAAAGGTTAGGGCAGTTGATGTTTGCAT
 ACTTGAAAAGTAAAGAGGTGTTTAATGATTGCAAGCTCATGTAAGTGAGTGCCTGTAACGTCCCCTAA
 TTAGGCTATGTGGTTTCGCACTTTATATTATCTTCTTAAATGAAGTGTGTATTTTTTCTTTCTCT
 TTCCATATGTGTAGATATAGATACGATTTTAAAGATGCAGCCTCAAACCTCCCGTCTCCGTCTCCCA
 TGTGCTGAGATTACAGACATGGGCTGGGTTTACTGTGGATCCAAGCTTAGAGACAAGGAACTGAGC
 AAATAAAACAGTCAAGATTCCAGACAGAAAGTGGCATAGGTGTTATATAAAAGCCAAAATGTGGAGT
 TTTAAGTCGATCATGAGTTAACATATTAAGTGTGTTGATTACTAACACAAAATGAAATTACTTGGG

AACTAAACTTCCCTGTAACGACTCCTTTTCCTCTAAGTAAACTAAGGAAGTTCTAAGAGGATGG
AAAGGTTTACACAGGAAAGAACTGCTTTAGTTCCTGGTGAGAAGAAAAAGCAAACCTTAGGGTTA
GGA

Organismo MusMusculus – PKp2 com mutacao.

ATAGCAGAGTCTGGGAGGCTGTCCGCTTTTCAGTTTGTCTAAAGGTTAGGGCAGTTGATGTTTGCA
TCTTGAAAAGTAAAGAGGTGTTTAATGATTGCAAGCTCATGTAAGTGAGTGCCTGTAACGTTAATT
AGGCTATGTGGTTCGCACTTTTATATTATCTTCTTAAATGAACTGTGTATTTTCCTTTCTCTTTCC
ATATGTGTAGATATAGATACGATTTTAAAGATGCAGCCTCAAACCTCCCGTCTCCGTCTCCCATGT
GCTGAGATTACAGACATGGGCTGGGTTTACTGTGGATCCAAGCTTAGAGACAAGGAACTGAGCAA
ATAAACAGTCAAGATTCCAGACAGAAAGTGGCATAGGTGGTATATAAAAGCCAAAATGTGGAGTT
TTAAGTCGATCATGAGTTAACATATTAAGTGTAAAGATTACTAACACAAAATGAAATTACTTGGGAA
CTAAACTTCCCTGTAACGACTCCTTTTCCTCTAAGTAAACTAAGGAAGTTCTAAGAGGATGGAAA
GGTTTACACAGGAAAGAACTGCTTTAGTTCCTGGTGAGAAGAAAAAGCAAACCTTAGGGTTAGG

Implemente o método de matriz de pontos para alinhar as duas sequências. É possível identificar quantas e quais foram as mutações realizadas na sequência "Organismo MusMusculus – PKp2 com mutacao"? É possível identificar a inserção de GAPS?