

LISTA DE EXERCÍCIO V

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 "e5-1a.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 "e5.pdf" apresentando o pseudocódigo do algoritmo desenvolvido e os resultados encontrados.
- Não serão aceitos trabalhos atrasados.
- Data de entrega: 05.12.2019 (quinta-feira) até as 13:00 via Moodle (<https://moodle.ufrgs.br/login/index.php>).

NOME: CARTÃO:

1. Realizar a leitura dos seguintes artigos:

- Artigo 1: Das, M. and Dai, H-K. A survey of DNA motif finding algorithms. BMC Bioinformatics, 2007, Vol. 8 (Suppl 7). *PDF disponibilizado no Moodle.*
- Artigo 2: Hertz, G.Z. and Hartzell, G.W. and Stormo, G.D. Identification of consensus patterns in unaligned DNA sequences known to be functionally related. Comput Appl Biosci. 1990, Vol. 6 (2). *PDF disponibilizado no Moodle.*

2. Considere as sequências de nucleotídeos a seguir. Implemente o algoritmo de *consensus* apresentado em sala de aula para buscar motivos estruturais. Os seguintes motivos devem ser buscados:

(a) Conjunto de Teste II:

- Encontrar o motivo de tamanho 8 aceitando 2 mutações;
- Encontrar o motivo de tamanho 5 aceitando 3 mutações;
- Encontrar o motivo de tamanho 3 aceitando 1 mutação;
- O código fonte do algoritmo deverá ser entregue.
- A análise dos experimentos devem ser entregues em PDF.
- Os resultados serão apresentados em sala de aula.

Seq1 :

cccctgatagacgctatctggctatccacgtacataggtcctctgtgcgaatctatgcgtttccaacct

Seq2 :

agtttactggtgtacatttgatacgtacgtacaccggcaacctgaaacaaacgctcagaaccagaagtgc

Seq3 :

aaaggagtccgtgcaccctctttcttcgtggctctggccaacgagggtgatgtataagacgaaaatttt

Seq4 :

agccccctccgatgtaagtcatactgtaactattacctgccacccctattacatcttacgtacgtataca
Seq5 :
ctgggttataacaacgcgtcatggcggggtatgcgttttggctcgtcgtacgctcgatcggttaacgtaggtc

(b) Conjunto de Teste II:

- Encontrar o motivo de tamanho 3 aceitando 1 mutações;
- Encontrar o motivo de tamanho 5 aceitando 2 mutações;

Seq1 :
gtcacgcttctgcataccatcctgactactcgtggcgaatacggttcgtctcagaacattgacgagtagga
cctccatgtacacgtgagttcgccagtagagggcagaactagaggcccgagctcgttaccagtatatgta
ctcggcacacactgggatataatactacacgggataactaatagtggtcatatcacgccg

Seq2 :
atccctctaacaagttgttttgacggaccgtatttccaaatgtgctcggcttcagaaacaacctttctgcc
ctctactggcgacgtcacaacgacgacaacagaccatatggagtggaccctactcatgtaattgagaccgt
cgcatgtagttgatattatgtaaacatatggctctagtttcaggccctgtaaaggtaa

Seq3 :
ttacataggttccttcacgtcactccttgcgcgatatctcctcttacccttactaccaagcgtttcctg
aaaggcaatgaaaagtgtccatgcgctgtcgccagtagagggcagaataccaaggcgttcagacaactgt
cgctgttcgtgggtgggagggttgtatctataatataggtatgtcgtatcgaaaaa

Seq4 :
ttatcgaccgccacttttctgccagtagagggcagaaccacaaagtgactccccgagcaatggctgaccta
tagttatccggcatcacatcgggcacatatacgggcgagaccgagccctctccgtaaccaccagtcctact
acttcacaggcatatcctgtatcaatgaaatcacaaacgttcgcatgaagataatcgt

Seq5 :
gacggcacattttaacggcccaggttggcagacaggaaatctacgatggtgctactgctttcccgagctct
gccacgatgccacacagcacaaattctgccctctactggcgatcacctcgaataaaaccgaatgcaagaccg
agtaacagcggctggtaacatgcgggaggacgcgctttccgcaagtatattaataggt

Seq6 :
tgcacataggttagtaagagttataaatcttcgatccctaagtgtggtgcactactcggctcgacctcgca
ttgacacaacgcgagagtcgccagtagagggcagaagccggcacttttgacctcttctatagaaggtagac
cgtgagatcgcgccgaaggggcccgcggtctccaaggtggaacgtattaggtaatc

Seq7 :
gccgtgtataggcctccgatcgtgcggttctgccctctactggcgaaaggggcatttgctattccaatcgc
atagattaccaataaaaaacgaaagaaggccgtccttgcaaagcttagtccttaactgagatgcttggc
gaccggccataagctccactcgcttgagcacatcaccaagaatcaaagtagcaaaccc

Seq8 :
tgtccgctctgcagacgtccgggatctacgttggtgttctctctagtaacagtacggcagttctttttcgg
tccgaagcgaatccacccgccaaggttacataagcattatctgaaggcaaccatacgaactctcattggc
tcgccagtagagggcagaaggacatcgtgtcatagcacatgccacagaggagattcg

Seq9 :
ccggtctcaatagccgaacaggatcgactggtaggcgtgtcgggtgtgtgtggaccggcctttggaagaac
cacacttctctggccctcaattggccaaaagtcatttaaggatcggttggccggcagccctcgtgacta
cgataaccccagggttctgccctctactggcgaccttgacagagcacttaccactgta

Seq10 :
aaaagagtagtgatgagttagaagaatttaaggacatcctcttgatttggacggctatccccaggaaat
cgtaggcgggggtgcacatggatatctttaggtattaattccccattccctctcgttctgccctctact
ggcgaatgttgctcgcaatactacagcctcctcaatacaggtagggtattttacatat