

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL - UFRGS
INSTITUTO DE INFORMÁTICA
BIOLOGIA COMPUTACIONAL - 2019 - LISTA DE EXERCÍCIOS I

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 "e1-1a.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 "e1-il.pdf" apresentando o pseudocódigo do algoritmo desenvolvido e os resultados encontrados. Arquivos corrompidos serão desconsiderados. As respostas a cada um dos itens devem obrigatoriamente ser apresentadas nos slides.
- Data de entrega: 03.09.2019 (terça-feira) até as 13:00 via Moodle (<https://moodle.ufrgs.br/login/index.php>). Não serão aceitos trabalhos atrasados.

NOME: CARTÃO:

1. O cromossomo 7 é um dos 23 pares de cromossomos do cariótipo humano. Este cromossomo representa de 5 a 5.5 % da quantidade total de DNA em células. Identificar os genes em cada cromossomo é uma grande área de pesquisa onde o número de genes presentes no cromossomo varia dependendo da técnica utilizada. O cromossomo 7 contém de 1.000 a 1.400 genes. O arquivo disponibilizado no endereço <http://www.inf.ufrgs.br/~mdorn/bio/sequence.fasta.tar.gz> contém a sequência FASTA (nt) do "Homo sapiens chromosome 7, GRCh38.p7 Primary Assembly". Implemente um conjunto de rotinas computacionais para realizar as tarefas que seguem:
 - (a) A subsequência CAATTGAATAATTG contém uma mutação. Quando não mutada esta subsequência é única no Cromossomo 7. Identifique a região no cromossomo 7 correspondente a esta subsequência (localização) e informe qual foi a mutação realizada. Entrada: sequência do cromossomo 7; Saída: sequência não mutada; região de localização da subsequência no cromossomo; mutação realizada.
 - (b) Um palíndromo é uma palavra, frase ou qualquer outra sequência de unidades (como uma cadeia de DNA; Enzima de restrição, etc) que tenha a propriedade de poder ser lida tanto da direita para a esquerda como da esquerda para a direita. Determine quantos diferentes palíndromos de tamanho 11 e 9 podem ser encontrados na sequência do cromossomo 7. Entrada: sequência do cromossomo 7; Saída: quantidade de palíndromos de tamanho 9 e 11; número de ocorrências de cada palíndromo.
 - (c) Identifique as diferentes subsequências de tamanho 37 presentes na sequência do cromossomo 7 contabilizando o número de ocorrências de cada uma destas subsequências. Entrada: sequência do cromossomo; Saída: lista das diferentes subsequências e o número de ocorrências de cada uma delas.
 - (d) Contabilize o número de ocorrências de cada um dos quatro nucleotídeos. Existe algum caracter diferente na sequência? Entrada: sequência do cromossomo; Saída: Número de ocorrências para cada nucleotídeo; sim ou não para caracter diferente, se sim identificar.
 - (e) Em genética, DNA complementar (cDNA) é o DNA sintetizado a partir de uma molécula de RNA mensageiro, cujos íntrons já foram removidos, ou seja, o mRNA já passou pelo processo de splicing, sendo uma reação catalisada pela enzima transcriptase reversa. Faça uma função que recebe a sequência do cromossomo 7 e gere a sequência complementar. Entrada: sequência do cromossomo; Saída: Sequência Complementar.