

## 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.
- Não serão aceitos trabalhos atrasados.
- Data de entrega: 21.08.2018 (terça-feira) até as 15:00 via Moodle (<https://moodle.ufrgs.br/login/index.php>).

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. A identificação de genes é uma tarefa nem sempre fácil, onde o número de genes encontrados em um mesmo cromossomo pode variar 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) Variações de um único nucleotídeo, chamados de SNPs, são extremamente comuns ao longo de genomas. Localizar e identificar estas trocas únicas são tarefas de extrema importância, pois sabe-se que podem estar diretamente relacionados a patologias. A subsequência CAGGAGATCTTCGTGCCAC possui uma mutação em um único nucleotídeo e é responsável pelo desenvolvimento de *diabetes mellitus* neonatal. Quando não mutada, esta sequência aparece apenas 1 vez 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) Uma das bases da engenharia genética é a utilização de enzimas de restrição, responsáveis por "cortar" o DNA em locais específicos, os chamados *sítios de restrição*. Estas enzimas, geralmente, atuam em regiões palíndromicas e complementares, ou seja, sequências de nucleotídeos que, quando reversas, são complementares a si mesmas (GAATTC, pois a complementar CT-TAAG, é idêntica ao reverso dela mesma). Determine quantos diferentes palíndromos de tamanho 9 podem ser encontrados na sequência do cromossomo 7. Entrada: sequência do cromossomo; Saída: quantidade de palíndromos de tamanho 9; número de ocorrências de cada palíndromo.

- (c) Sabe-se que regiões repetitivas são comumente achadas ao longo do genoma de plantas e animais, muitas vezes constituindo a maior parcela de subsequências dentro do genoma. Portanto, 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) A partir do seu número de matrícula completo (8 dígitos), gere uma sequência de nucleotídeos, onde: 0 - A, 1 - T, 2 - G, 3 - C, 4 - C, 5 - G, 6 - T, 7 - A, 8 - A, 9 - C; A partir disto, gere a sequência complementar e 3 sequências mutando o nucleotídeo correspondente a quinta posição da sequência. Contabilize o número de ocorrências de cada uma das 5 sequências no cromossomo 7. Entrada: sequência do cromossomo e número de matrícula; Saída: Cada sequência gerada e o número de ocorrências para cada uma delas.