

Universidade Federal do Rio Grande do Sul
Instituto de Informática
Departamento de Informática Teórica
INF05018 – Biologia Computacional (2017/2)
LISTA DE EXERCÍCIOS 01

Instruções:

- A resolução do exercício deve ser feita **individualmente**. Cópias evidentes entre trabalhos não serão aceitas.
 - Para cada uma das tarefas, deve-se entregar o código-fonte em linguagem Python. **Não utilize módulos externos (somente built-in)**. O nome do arquivo deve identificar a tarefa (ex.: "e1-1a.py" referente ao item "1a" da tarefa).
 - Além do código-fonte deve-se entregar um único arquivo PDF com o nome "e1-i1.pdf", apresentando o pseudocódigo do algoritmo desenvolvido e os resultados encontrados. As respostas a cada um dos itens devem obrigatoriamente ser apresentadas nos slides.
 - Os arquivos devem ser enviados em um arquivo compactado com nome no formato "IDA_e1", sendo IDA as iniciais do nome do aluno.
 - Arquivos corrompidos serão desconsiderados.
 - Data de entrega: 19/09/2017 (terça-feira) até as 12h via Moodle.
 - **Não serão aceitos trabalhos atrasados.**
-

- 1) O cromossomo 21 é um dos 22 pares de cromossomos somáticos do cariótipo humano. Identificar os genes em cada cromossomo é uma grande área de pesquisa onde o número de genes identificados no cromossomo varia dependendo da técnica utilizada. Este cromossomo contém aproximadamente 750 genes, codificando em torno de 1.200 proteínas. Uma duplicação anormal desse cromossomo, gerando três cópias, causa a Síndrome de Down (trissomia do cromossomo 21), associada a complicações de nível físico e neurológico. A sequência a ser utilizada é de acesso [NC 000021.9](https://www.ncbi.nlm.nih.gov/nuclot/NC_000021.9), referente à montagem GRCh38.p7 do cromossomo 21 de *Homo sapiens*, que deverá ser baixada em formato FASTA (nt). Implemente um conjunto de rotinas computacionais para realizar as tarefas que seguem:
 - a) A subsequência ACCTGGTGTCCCA contém uma mutação. Identifique o(s) potencial(is) candidatos a sequência original, informando a localização da subsequência, bem como a mutação identificada. **Entrada:** sequência do cromossomo 21. **Saída:** sequência não mutada, região de localização da subsequência, mutação existente.
 - 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 da mesma maneira da direita para a esquerda e vice-versa. Determine quantos diferentes palíndromos de tamanho 9 podem ser encontrados. **Entrada:** sequência do cromossomo 21. **Saída:** quantidade de palíndromos de tamanho 9, número de ocorrências de cada palíndromo.
 - c) Identifique as diferentes subsequências de tamanho 37, contabilizando o número de ocorrências de cada uma delas. **Entrada:** sequência do cromossomo 21. **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. Estime o GC%. Existe um caractere diferente na sequência? **Entrada:** sequência do cromossomo 21. **Saída:** Número de ocorrências para cada nucleotídeo, GC%, “sim” ou “não” para caractere diferente e identificar, caso exista.