

Universidade Federal do Rio Grande do Sul – UFRGS  
Biologia Computacional – 2014

Prof. Dr. Márcio Dorn  
12 de Agosto de 2014

DESAFIO COMPUTACIONAL I  
OPERAÇÕES COM *STRINGS*

Objetivos: manipulação de *strings*, implementação de algoritmos básicos, operações com arquivos.

1. Acesse o endereço <http://www.inf.ufrgs.br/~mdorn/bio/T1/> e baixe o arquivo *\*.fasta*. Este arquivo contém a sequência *W* de nucleotídeos do "*Homo sapiens chromosome 7*". Implemente um conjunto de rotinas computacionais para realizar as tarefas que seguem:
  - (a) Calcule a quantidade de diferentes subsequências de tamanho 5 (cinco) que estão presentes na sequência *W*.
    - Versão 1: NÃO utilize a estrutura de dados do tipo Dicionário.
    - Versão 2: Solução Livre.
  - (b) Considere a subsequência *X* = "TTT". Assumindo que *X* representa uma região de corte (terminação/início) de outras subsequências determine:
    - Quantidade de subsequências separadas por *X*.
    - Existe alguma alteração no número de subsequências obtidas ao processarmos a sequência *W* da esquerda para a direita e da direita para a esquerda?
  - (c) Considere um resíduo coringa *U*. Quantas diferentes subsequências do tipo *sUsUs* podem ser encontradas na sequência *W*? *s* representa qualquer uma das bases da sequência *W*.
  - (d) Um palíndromo é uma palavra, frase ou qualquer outra sequência de unidades (como uma cadeia de DNA; Enzima de restrição) 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 7 podem ser encontrados na sequência *W*.

REGRAS:

*Data de entrega e apresentação:* 14.08.2013 via moodle.

*Formato de entrega:* para cada um dos itens acima (a,b,c,d) entregar arquivo com código fonte.

*Apresentação:* máximo 5 lâminas apresentando o pseudocódigo de cada algoritmo desenvolvido.

*Modalidade:* individual.

ALGUMAS SUGESTÕES:

Use linguagem Python.

Revise o material de manipulação de *strings* disponível no Moodle.