

[SCC000271] Introdução à Bioinformática
Lista de exercícios 1

Professor: Dilvan de Abreu Moreira
Estagiário PAE: Pedro Shiguihara
Monitor: Vicente Mattos

14 de setembro de 2011

Exercício 1. Por que é utilizada a modelagem das sequências de DNA, através da análise estatística:

1. A modelagem precisa imitar fielmente as sequências de DNA.
2. A modelagem precisa capturar suficiente das propriedades de sequências enquanto ainda está sendo eficiente computável.
3. A modelagem precisa reproduzir as sequências exatamente sem importar a eficiência em termos de computabilidade.

Exercício 2. Considerar as cadeias de Markov de primeiro ordem com probabilidades iniciais: $p_a = 0.1, p_c = 0.2, p_g = 0.4, p_t = 0.3$, e a seguinte matriz de transição:

$$M = \begin{vmatrix} 0.1 & 0.2 & 0.1 & 0.6 \\ 0.2 & 0.1 & 0.3 & 0.4 \\ 0.3 & 0.3 & 0.2 & 0.2 \\ 0.1 & 0.2 & 0.3 & 0.4 \end{vmatrix}.$$

Exercício 3. Calcular o conteúdo GC da sequência: *GATCACAGGTCTATCACCCCTATTAACCACTCACGGGAGCTCTCCATGCATTTGGTATTTTCGTCTGGGGG*, com tamanho de janela $k=5$.

Exercício 4. Quantas janelas de tamanho $k = 15$ são produzidas ao analisar o genoma da *M. genitalium* com 580074 bases.

Exercício 5. Tendo em conta que há 15 códons possíveis. Qual é o comprimento mínimo dum ORF (considerar codon de início e de parada), se apenas 6% são ORFs significativos.

Exercício 6. Os códons de parada: TAA, TAG, TGA, são utilizados para determinar ORFs na sequência: 3'ATGCCCAAGCTGAATAGCGTAGAGGGGTTTTTCATCATTTGAGGAC5'. Indique pelo menos 2 ORFs existentes (o códon de início pode ser qualquer triplaleta exceto dos códons de parada).