

**MO640 – Biologia Computacional**  
**Primeiro Semestre de 2010**  
**Terceira Lista de Exercícios**

1. Explique como usar o BLAST para remover trechos contaminados por vetor de sequenciamento.
2. Mostre um Alinhamento Estrela com pontuação pelo menos 50% maior que o melhor alinhamento múltiplo possível. Considere distância de edição para o alinhamento entre as sequências e Soma de Pares (coluna a coluna) para a pontuação do alinhamento múltiplo.
3. Considere um alinhamento múltiplo  $A$  de  $k$  sequências, todas de tamanho  $n$ . Seja  $p$  o valor do alinhamento, considerando pontuação por Soma de Pares (coluna a coluna) e distância de edição para comparação entre os caracteres. É possível modificar o alinhamento múltiplo, em tempo polinomial (em termos de  $k$  e  $n$ ), e obter um alinhamento  $A'$ , com pontuação  $p'$ , tal que  $p' < p$  (caso  $A$  não seja um alinhamento ótimo). Justifique sua resposta.
4. Considere o grafo de sobreposição  $OG(\mathcal{F}, t) = (V, E)$ , com  $V = \{a, b, c, d, e\}$ ,  $E = \{(a, b), (a, c), (a, d), (b, e), (c, e), (d, e)\}$  e  $t = 2$ . Determine um conjunto de fragmentos  $\mathcal{F}$  compatível com  $OG(\mathcal{F}, t)$ .
5. Mostre uma sequência  $S$  e um conjunto de fragmentos  $\mathcal{F}$  que cobre  $S$  completamente, mas que o modelo de Supersequência Comum Mínima (SCS) não reconstrói corretamente a sequência original.
6. Seja  $\mathcal{F} = \{ATC, TCG, AACG\}$ . Encontre o melhor alinhamento para a coleção  $\mathcal{F}$  de acordo com o modelo de reconstrução (RECONSTRUCTION), considerando  $\epsilon = 0.1$  e  $\epsilon = 0.25$ . Lembre-se de considerar também os complementos reversos dos fragmentos.
7. Seja  $\mathcal{F} = \{TCCCTACTT, AATCCGGTT, GACATCGGT\}$ . Encontre o melhor conjunto de *contigs* para a coleção  $\mathcal{F}$  de acordo com o modelo MULTICONTIG considerando  $\epsilon = 0.3$  e  $t = 5$ .
8. Para cada um dos seguintes problemas biológicos, mostre um exemplo onde o modelo SCS falha ao montar os fragmentos da sequência original:
  - a. Erros de sequenciamento
  - b. Orientação desconhecida dos fragmentos
  - c. Regiões repetidas na sequência original

Em cada caso, indique a sequência original a ser montada, os fragmentos utilizados na montagem e a sequência obtida com o modelo SCS.