

Bioinformática

Exercícios - Comparação e alinhamento de sequências

1. O programa de comparação de sequências Dotlet (<https://dotlet.vital-it.ch>) permite identificar padrões de semelhança entre sequências de DNA ou proteína:
 - a) Obtenha, a partir de base de sequências Uniprot (www.uniprot.org) a sequência da tripsina-1 ("trypsin-1") humana e da ratazana ("rat"). Faça a comparação das duas sequências no Dotlet (use uma janela de tamanho 5 e uma matriz "Blosum65"). O que observa? Use Uniprot para calcular a percentagem de identidade entre as duas sequências.
 - b) Repita o procedimento anterior, mas comparando desta vez a tripsina-1 humana com as tripsinas de salmão e da bactéria *Streptomyces griseus*. Compare com os resultados anteriores.
 - c) Faça a comparação do precursor da protrombina (código Uniprot: P00735) com a tripsina humana. Consegue interpretar o resultado obtido?
 - d) Gere uma sequência aleatória de 200 aminoácidos utilizando a página <http://expasy.org/tools/randseq.html>, e grave-a em formato FASTA num ficheiro de texto. Crie três novas sequências a partir da inicial, com respectivamente 400, 600 e 800 aminoácidos, contendo respectivamente dois, três e quatro repetições do fragmento original. Guarde essas sequências no ficheiro de texto. Faça a autocomparação de cada uma das sequências com Dotlet. Qual a relação entre o número de repetições e os padrões observados? (Nota: seleccione a opção "Identity" da lista de matrizes disponíveis para melhor evidenciar o resultado).
 - e) Faça a autocomparação da Tripsina I humana com o Dotlet. Pode concluir alguma coisa?
 - f) Faça a autocomparação do precursor da protrombina (código Uniprot: P00735), e do plasminogénio (P06868). Quantos domínios do tipo "Kringle" apresenta cada uma? (Os domínios Kringle são pequenas regiões repetidas dentro da sequência de algumas proteínas do plasma sanguíneo)
 - ~~g) Obtenha o gene da calmodulina do fungo *Aspergillus nidulans* (código GeneBank J05545) e a sequência da proteína correspondente (P60204). Use Dotlet para determinar a posição dos intrões no gene (Sugestão: use uma matriz de alta selectividade, Blosum100, e uma janela pequena, 7).~~
2. Obtenha as seguintes sequências: cadeias α e β da hemoglobina humana, leghemoglobina 1 do tremço (yellow lupin) e glutathione-S-transferase 2 (GST-2) de *Caenorhabditis Elegans*.
 - a) Usando o alinhamento local (opção "water") em <http://www.ebi.ac.uk/emboss/align>, produza os 3 alinhamentos: cadeia α com a cadeia β , cadeia α com a leghemoglobina, e a cadeia α com GST-2. Tome nota das percentagens de identidade, semelhança e gaps em cada caso. Qual das sequências, leghemoglobina e GST-2, parece ser mais aparentada com a cadeia α da hemoglobina?
 - b) Faça uma pesquisa BLAST (<http://www.ncbi.nlm.nih.gov/BLAST>) usando como "query" a sequência da leghemoglobina. Consegue encontrar as cadeias α e β da hemoglobina humana?
 - c) Compare o alinhamento local e global da cadeia α com a leghemoglobina.
 - d) Gere duas sequências aleatórias de 200 aminoácidos e alinhe-as local e globalmente, comparando os resultados obtidos com a alínea a).

3. Construa com Randseq uma sequência do tipo AB e outro do tipo BA, em que A e B são sequências de 50 aminoácidos de comprimento geradas com Randseq. Faça um alinhamento local e global destas sequências com as opções "needle" e "water" do programa Align. Comente. Use o program Lalign (http://www.ch.embnet.org/software/LALIGN_form.html) para listar alinhamentos sub-óptimos das sequências AB e BA. Compare o alinhamento obtido com Lalign com os anteriores e comente ? Terá este exemplo relevância biológica ?
4. Alinhe as sequências dos factores de coagulação IX e XII humanos local e globalmente. Use o programa Lalign (http://www.ch.embnet.org/software/LALIGN_form.html) para listar alinhamentos subóptimos. Compare os alinhamentos encontrados com as definições de domínio listadas nas entradas SWISSPROT dos factores IX e XII.