

BioInformática - Trabalho 1

Algoritmos de Alinhamento de Nucleotídeos e Aminoácidos

Introdução e objectivos:

O alinhamento de sequências, quer de proteínas quer de DNA, é importante para identificar regiões semelhantes em espécies diferentes. Desta forma, quando alinharmos duas sequências que têm o mesmo ancestral comum, o mismatch pode ser interpretado como uma mutação pontual e os gaps como indels, introduzidos numa das espécies ou em ambas as espécies. No alinhamento de proteínas, a similaridade entre os aminoácidos que estão numa certa posição na sequência pode ser interpretada como uma medida aproximada do quão conservada está a região.

A proteína que escolhemos foi a do capsídeo do vírus da imunodeficiência do humano (HIV) e do gato (FIV). Esta proteína é responsável por proteger o material genético do vírus enquanto este é copiado num ambiente agressivo. Relativamente ao gene, escolhemos o gag do HIV e do FIV. Este gene é responsável pela produção de várias proteínas estruturais do vírus, como por exemplo a proteína do capsídeo que escolhemos acima.

O trabalho consiste na implementação de um alinhamento (global e local) de proteínas e genes por programação dinâmica usando o nosso programa e comparar os resultados através de um acesso a um servidor web. O objetivo é alinhar 2 proteínas reais e 2 genes reais, da mesma família mas de espécies diferentes, usando custo linear.

Descrição do trabalho efectuado:

Usando uma linguagem orientada a objetos (java) foram implementados os algoritmos de programação dinâmica de Needleman-Wunsch (alinhamento global) e de Smith-Waterman. No alinhamento de proteínas usamos a tabela de substituição BLOSUM62 e no alinhamento de nucleotídeos foram utilizadas as seguintes pontuações: -2 para gap, +1 para match e -1 para mismatch.

Após executar o algoritmo selecionado nas sequências passadas como input, o nosso programa permite o envio de um pedido HTTP/Rest ao servidor do EBI para os respectivos algoritmos.

Instruções para uso:

O nosso programa é bastante intuitivo, sendo que as instruções para uso são:

- 1º - Alinhamento Global (1) ou Alinhamento Local (2)?;
- 2º - Alinhar Genes (1) ou Alinhar Proteínas (2)?;
- 3º - Input File;;
- 4º - Fazer pedido ao servidor do EBI? y or n ?.

Organização do Programa:

O programa está organizado em várias funções que desempenham papéis diferentes mas igualmente importantes.

As funções “lineUpProteinsGlobal()”, “lineUpProteinsLocal()”, “lineUpGenesGlobal()” e “lineUpGenesLocal()” são responsáveis pelo preenchimento da matriz bidimensional através dos procedimentos de programação dinâmica, utilizando BLOSUM62 para proteínas e score(-2 gap, -1 mismatch, +1 match) para genes.

As funções “goBackGlobal()” e “goBackLocal()” são recursivas e responsáveis por encontrar o caminho da sequência mais vantajosa, de acordo com a técnica escolhida.

A função “printAlignment()” imprime os resultados do processo de alinhamento.

Por fim, a função “askEBI()” efectua a sequência de pedidos/resposta ao/do servidor do EBI com a finalidade de imprimir o resultado do alinhamento das mesmas sequências realizado pelos algoritmos falados com scores o mais semelhantes quanto possível.

Exemplo:

Global(1) or Local(2) Alignment?

1

Compare Genes(1) or Proteins(2)?

2

Input file name?

Proteins.fasta // ficheiro contendo duas sequências de aminoácidos.

Output do programa

EBI Request - Do you wanna see the result for the same sequences from EBI website?

y or n ?

Output do servidor

Nota (exemplo de input - input/Proteins.fasta):

>NP_579880.1 capsid [Human immunodeficiency virus 1]

PIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALESEGATPQDLNTMLNTVGGHQA
AMQMLKETINEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTNNPPIPVG
EIYKRWILGLNKIVRMYSPTSILDIRQGPKPEFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQN
ANPDCKTILKALGPAATLEEMMTACQGVGGPGHKARVL

>PDB|5NA2|A Capsid protein (p24)

TIQTVNGVPQYVALDPKMVSIFMEKAREGLGGEEVQLWFTAFSANLTPTDMATLIMAAPGCAAD
KEILDESLKQLTAEYDRTHPPDAPRPLPYFTAAEIMGIGLTQEQQAEARFAPARMQCRAWYLEA
LGKLAAIAKASPRAVQLRQGAKEDYSSFIDRLFAQIDQEQNTAEVKLYLKQSLSIANANADCKKA
MSHLKPESTLEEKLRACQEI

Porto, 09 de Abril de 2017

Trabalho realizado por:

Cristiano Barbosa up201402990

Rui Pinto up201408076

Bárbara Ferreira up201001552



DEPARTAMENTO DE CIÊNCIA DE COMPUTADORES
FACULDADE DE CIÊNCIAS DA UNIVERSIDADE DO PORTO