Comparação de sequências II -Alinhamentos pareados

Matrizes de substituição

No arquivo EPAM250.txt (disponível no e-Disciplinas) encontrará a matriz de substituição PAM250

Quem e como criou a família de matrices de substituição PAM?

Onde encontra as pontuações (scores) maiores? Por quê?

Qual a substituição com a maior pontuação (scores)?

Por que as identidades (um aminoácido substituído por ele mesmo) nem sempre tem as mesmas pontuações?

Alinhamento Global

O objetivo no alinhamento global é comparar duas sequências ao longo de todo seu comprimento. Ele é apropriado quando esperamos semelhança entre as duas sequências inteiras.

A aplicação needle do EMBOSS implementa o algoritmo de Needleman & Wunsch (1970), que é capaz de achar o alinhamento global ótimo (exato) entre duas sequencias usando o algoritmo de programação dinâmica. Está implementação pode tomar muito tempo se as sequencias comparadas são muito compridas.

Quais outras aplicações no EMBOSS permitem obter alinhamentos globais? Quais as diferenças delas com o needle?

Faz um alinhamento global entre as sequências cDNA ANAC092 e genômico ANAC092 que estão disponíveis no e-Disciplinas.

Qual matriz de substituição e custo de gaps usou? Por quê?

Qual é a pontuação do alinhamento, seu comprimento, e porcentagens de identidade e semelhança? Explique a diferença entre semelhança e identidade

Qual é o significado dos símbolos: “:”, “.”, e “|”?

Nos arquivos ANAC092\_pep.fasta e PpNAC\_e\_gw1.5.134.1.fasta encontra as sequências de amino ácidos de dois genes da família de fatores de transcrição NAC em *Arabidosis thaliana* e o musgo *Physcomitrella patens*, respectivamente.

Faz um alinhamento global entre essas duas sequências.

Qual matriz de substituição e custo de gaps usou (abrir e extensão) ? Por quê?

Qual é a pontuação do alinhamento, seu comprimento, e porcentagens de identidade e semelhança? Pode melhorar o alinhamento escolhendo outros parâmetros? Quais?

12.3. Alinhamentos locais

Como foi falado anteriormente o alinhamento global supõe semelhança ao longo de toda a sequência. Será sempre essa estratégia a mas adequada sempre?

Que acontecerá se você compara duas proteínas multidominio que só compartilham um domínio entre elas?

O objetivo do alinhamento local é encontrar as regiões de semelhança local, e por tanto não é necessário incluir a sequência inteira no alinhamento. Este tipo de alinhamento é muito útil para fazer buscas em bancos de dados, ou quando não tem uma ideia clara com relação a semelhança da sequência query com as sequencias do banco de dados.

No EMBOSS pode encontrar a aplicação water que implementa o algoritmo de Smith & Waterman (1981) para gerar alinhamentos locais ótimos por programação dinâmica.

Quais outras aplicações no EMBOSS permitem obter alinhamentos locais? Quais as diferenças delas com o water?

Faz uma alinhamento local entre as sequências de amino ácidos das proteínas NAC de *A.thaliana* y *P. patens*.

Qual matriz de substituição e custo de gaps usou (abrir e extensão)? Por quê?

Qual é a pontuação do alinhamento, seu comprimento, e porcentagens de identidade e semelhança?

Pode melhorar o alinhamento escolhendo outros parâmetros? Quais?

Que diferenças existem entre o alinhamento global e o local para essas duas sequencias?

Significância dos alinhamentos

Sem importar quais sequências você use nos programas de alinhamento, eles sempre vão gerar o alinhamento.

Peque a sequência de amino ácidos ANAC092 e use o programa shuffleseq do EMBOSS, crie duas sequencias randômicas com a mesma composição monomérica do ANAC092. Faz alinhamentos globais e locais com essas duas sequencias randômicas.

Qual matriz de substituição e custo de gaps usou (abrir e extensão)? Por quê?

Qual é a pontuação do alinhamento, seu comprimento, e porcentagens de identidade e semelhança?

Pode melhorar o alinhamento escolhendo outros parâmetros? Quais?

Agora faz alinhamentos locais e globais entre a sequência original de amino ácidos ANAC092 e uma das sequencias randômicas.

Qual matriz de substituição e custo de gaps usou (abrir e extensão)? Por quê?

Qual é a pontuação do alinhamento, seu comprimento, e porcentagens de identidade e semelhança?

Pode melhorar o alinhamento escolhendo outros parâmetros? Quais?