alinhamento global-alinhamento múltiplo de seqüências

Alinhamento múltiplos de seqüências

- Qual a importância de se realizar alinhamentos múltiplos em oposição a alinhamentos em pares?

Alinhamento múltiplos de seqüências

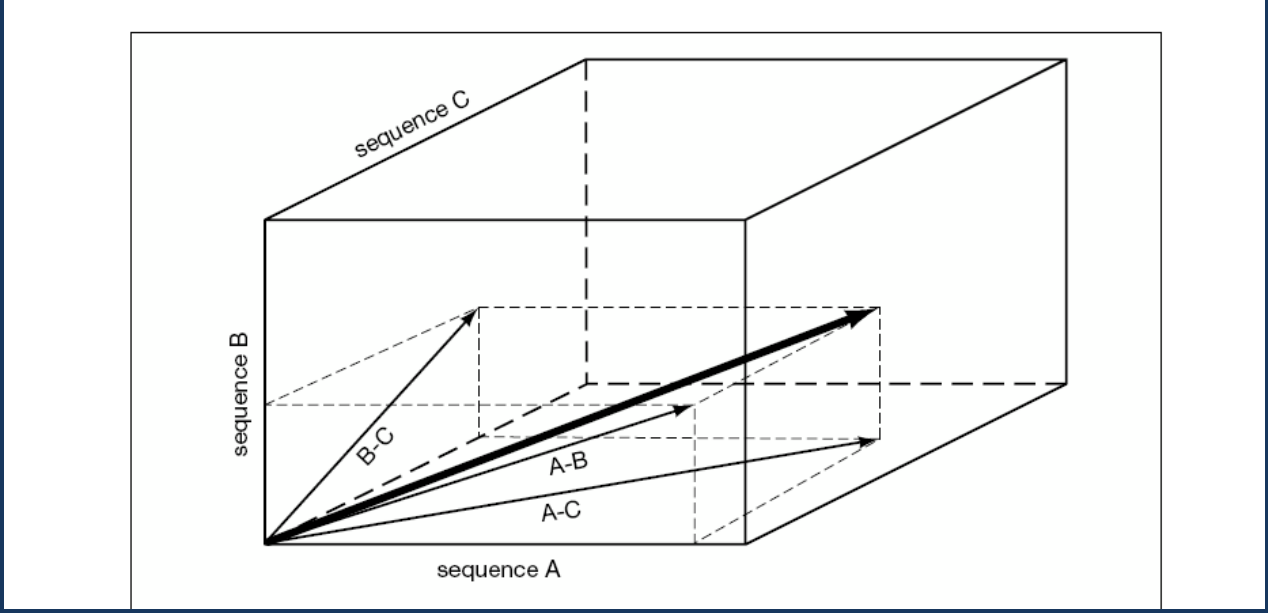
Usos do alinhamento múltiplo de seqüências

-Estabelecer relações filogenéticas entre diferentes proteínas

-Predizer domínios conservados, incluindo resíduos críticos para a função da proteína

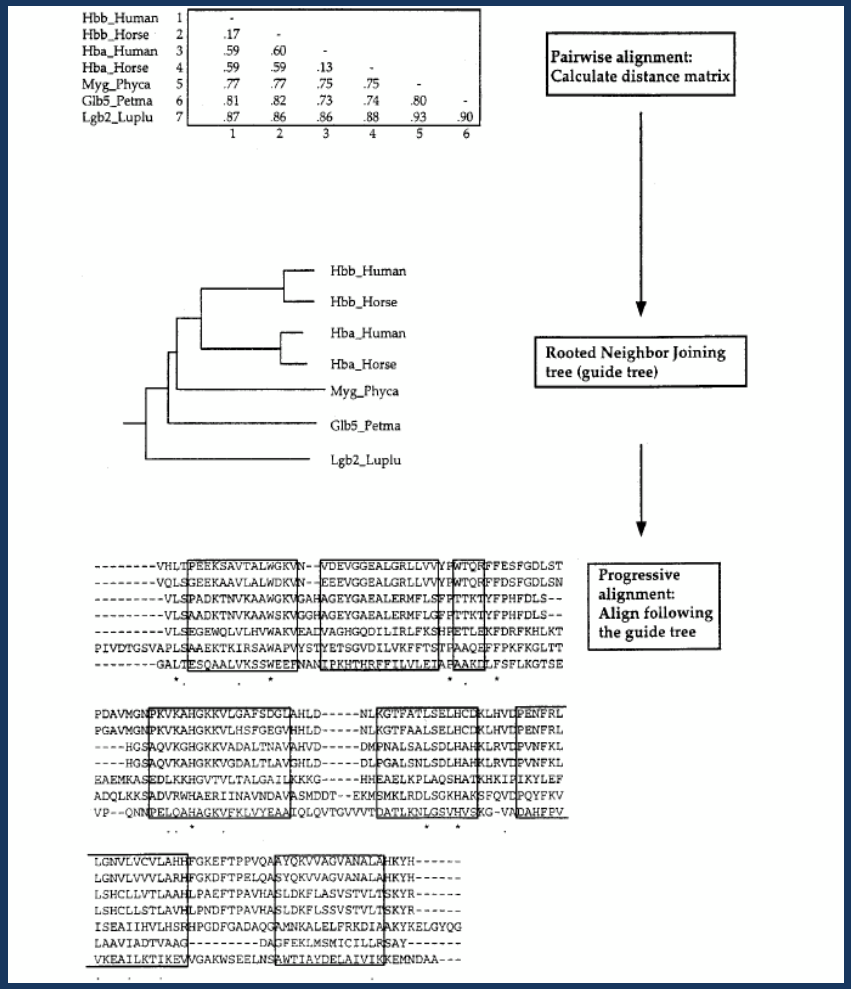
-Comparar a proteína de interesse de forma mais detalhada com membros da mesma família de proteínas

Alinhamento múltiplo de seqüências



Ao se alinhar 3 seqüências pode-se calcular o alinhamento ótimo através do dynamic programming, entretanto neste caso você passa a ter que calcular uma matriz tridimensional. Caso mais seqüências fossem adicionada seria necessário uma matriz N dimensional onde N é o numero de seqüências alinhadas
Isso cria um problema computacional pois o numero de comparações a ser feitas será o numero de aminoácidos elevado ao numero de seqüências alinhadas

Alinhamento progressivo (Clustal)



- Com a finalidade de solucionar este problema um método heurístico foi desenvolvido de modo a fazer um alinhamento progressivo de seqüências baseado na sua proximidade em uma arvore filogenética baseada em alinhamentos par a par
- Apesar deste tipo de método gerar bons alinhamentos em tempo razoável, ele não garante que o melhor alinhamento possível será obtido

Alinhamento progressivo (Clustal)

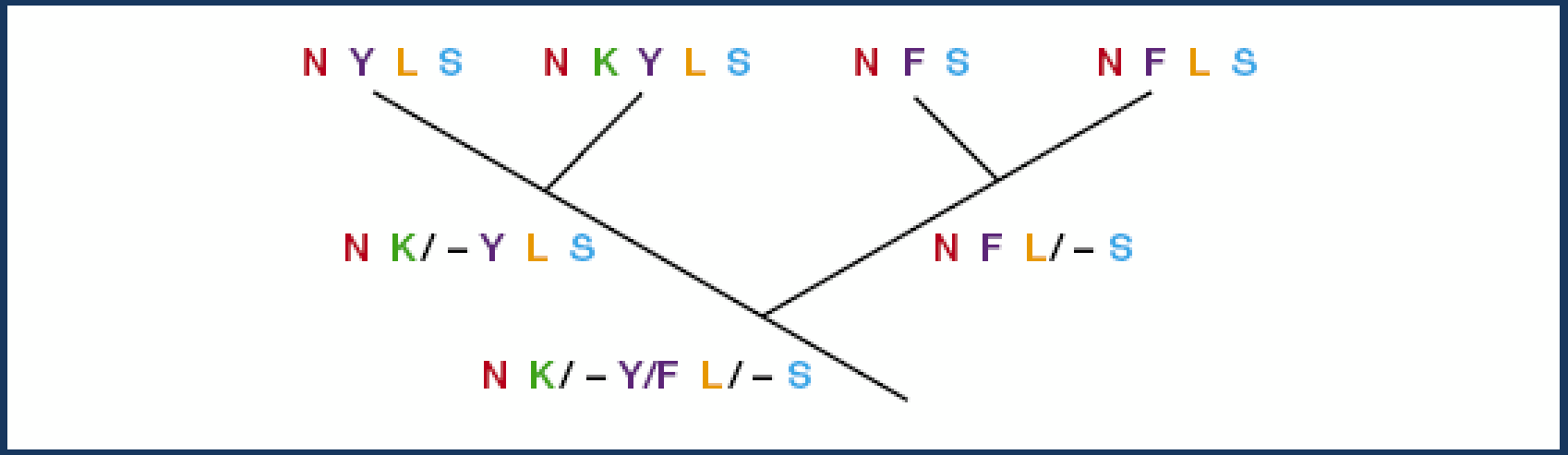
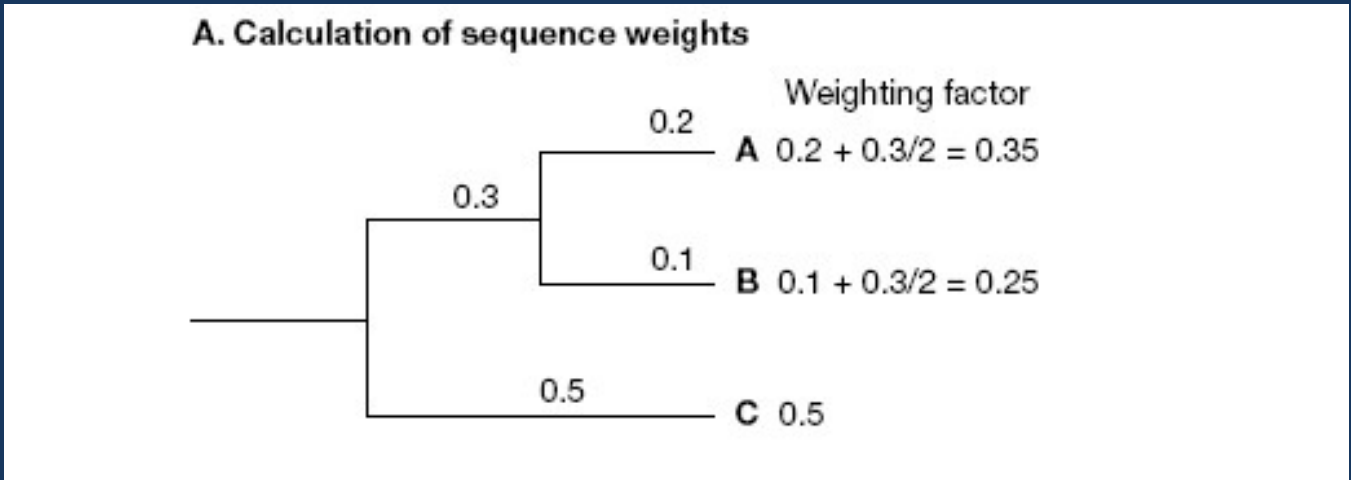


Figura mostra o principio do alinhamento progressivo. Alinhamentos são realizados seqüencialmente baseado na arvore filogenética estimada criada a partir dos alinhamentos par a par realizados.

Alinhamento progressivo (Clustal)



Ao realizar cada alinhamento são gerados pesos para cada seqüências dependendo da sua distancia em relação as outras seqüências do alinhamento. O objetivo deste procedimento é evitar a criação de vieses gerados pela escolhas de muitas seqüências próximas em oposição a poucas seqüências divergentes.

Alinhamento progressivo (Clustal)

B. Use of sequence weights

Column in alignment 1

Sequence A (weight a) .....K.....

Sequence B (weight b) .....I.....

Column in alignment 2

Sequence C (weight c) .....L.....

Sequence D (weight d) .....V.....

Score for matching these two column in an msa =

[a x c x score (K,L) +
a x d x score (K,V) +
b x c x score (I,L) +
b x d x score (I,V)] / 4

Exemplo de calculo de escore em alinhamento progressivo a partir de dois pares de seqüências alinhadas. Note que cada seqüência possui um peso especifico mesmo já estando alinhada com uma outra seqüência. Escores de substituição ou identidade entre seqüências já alinhadas (Seq A e B, no exemplo acima) não são computados visto que o alinhamento entre A e B não vai mais variar.

Problemas do método de alinhamento progressivo

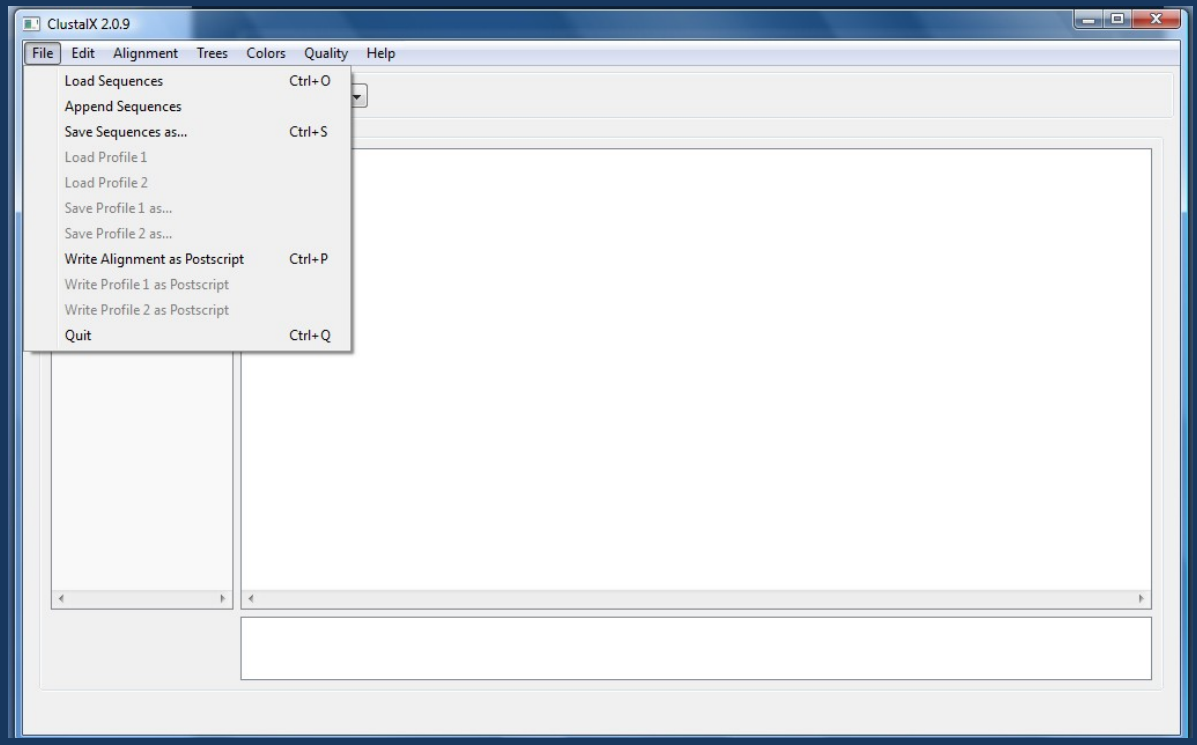
- Este tipo de algoritmo é altamente dependente da qualidade dos alinhamentos iniciais, visto que ao longo do processo eles não serão mais alterados
- Dificuldade é maior quando mesmo as seqüências mais próximas são distantemente relacionadas
- Dificuldade em estabelecer parâmetro ótimos de alinhamento

Uso de aproximação sucessiva para refinar alinhamentos

- Devido a limitações impostas devido ao alinhamento progressivo foi implementado nas ultimas versões do clustal um algoritmo utilizando aproximação sucessiva (iteration) para minimizar este tipo de problema

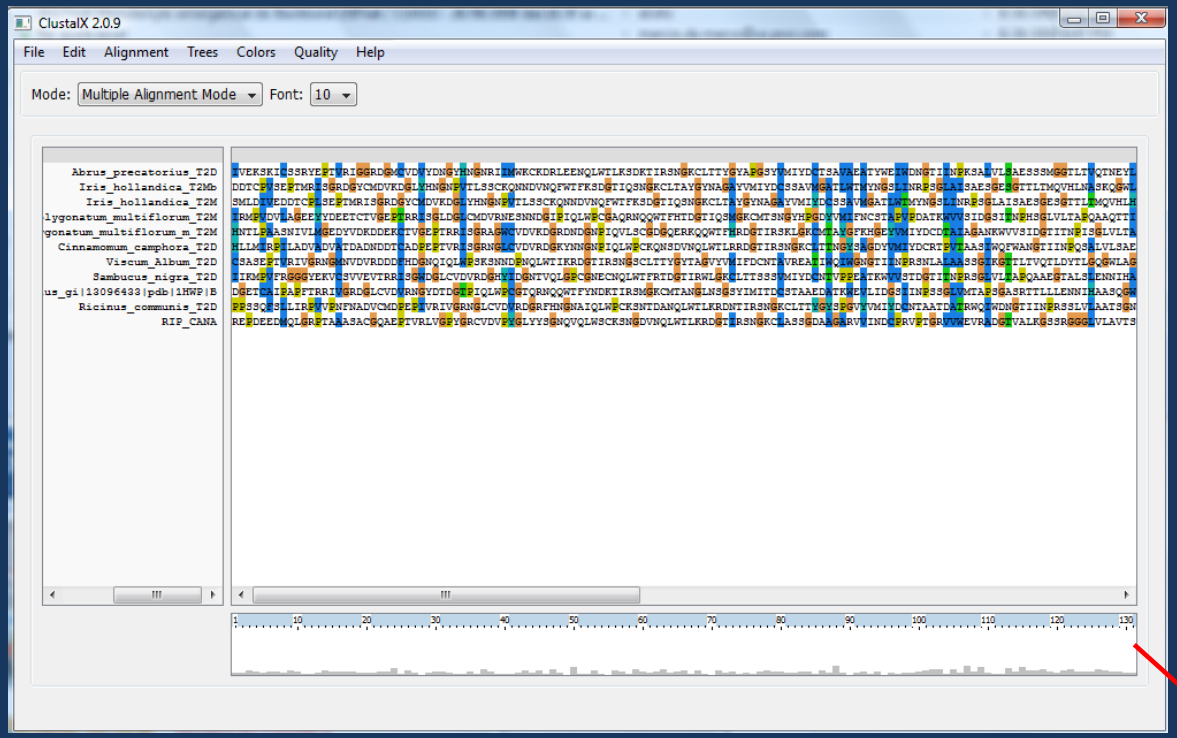
- O algoritmo realiza após cada alinhamento um procedimento no qual ele seleciona uma seqüência e realinha esta com o resto do alinhamento, caso o escore resultante for melhor que o inicial o novo alinhamento é mantido. Isto e realizado sucessivamente com todas as seqüências do alinhamento.

ClustalX



- Seqüências em formato multi-fasta ou aln pode ser carregadas no clustalX através do comando “load sequences”

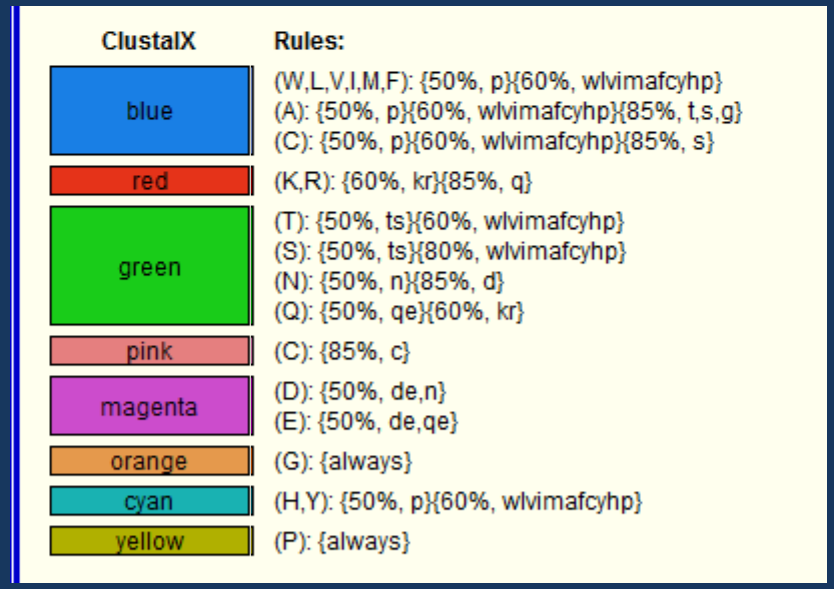
ClustalX



Cada linha representa uma seqüência

Gráfico com Indicação do grau de conservação

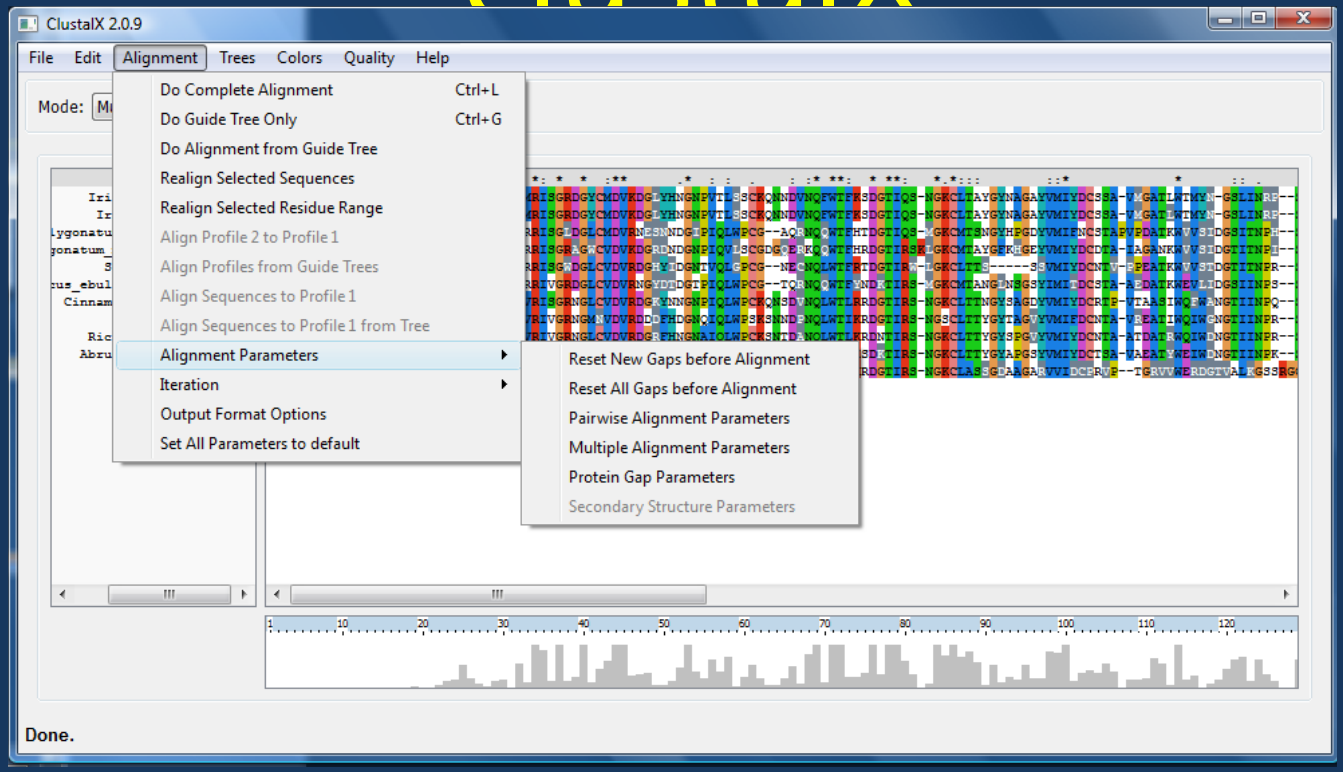
Sistema de cores do clustal



ClustalX	Rules:
blue	(W,L,V,I,M,F): {50%, p}{60%, wlvimafcyhp} (A): {50%, p}{60%, wlvimafcyhp}{85%, t,s,g} (C): {50%, p}{60%, wlvimafcyhp}{85%, s}
red	(K,R): {60%, kr}{85%, q}
green	(T): {50%, ts}{60%, wlvimafcyhp} (S): {50%, ts}{80%, wlvimafcyhp} (N): {50%, n}{85%, d} (Q): {50%, qe}{60%, kr}
pink	(C): {85%, c}
magenta	(D): {50%, de,n} (E): {50%, de,qe}
orange	(G): {always}
cyan	(H,Y): {50%, p}{60%, wlvimafcyhp}
yellow	(P): {always}

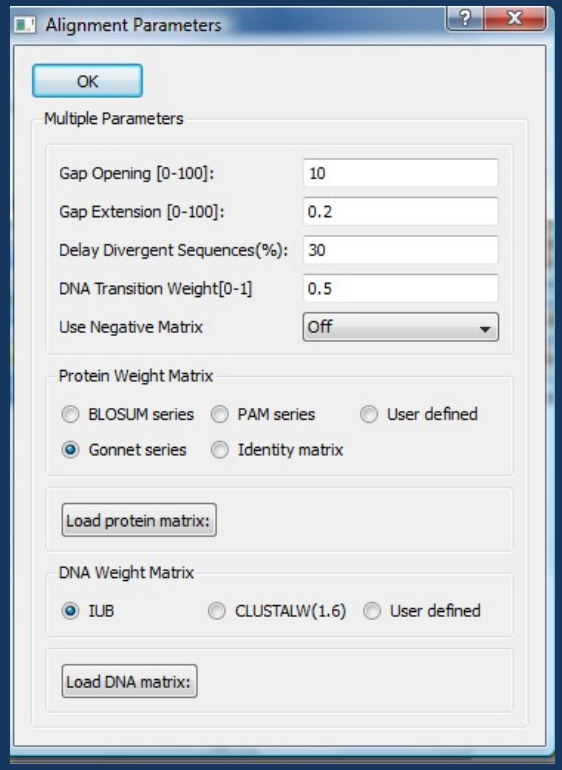
- O fundo colorido do clustal tem duas funções: chamar a atenção de regiões conservadas e para o caráter dos aminoácidos

Alinhando seqüências no clustalX



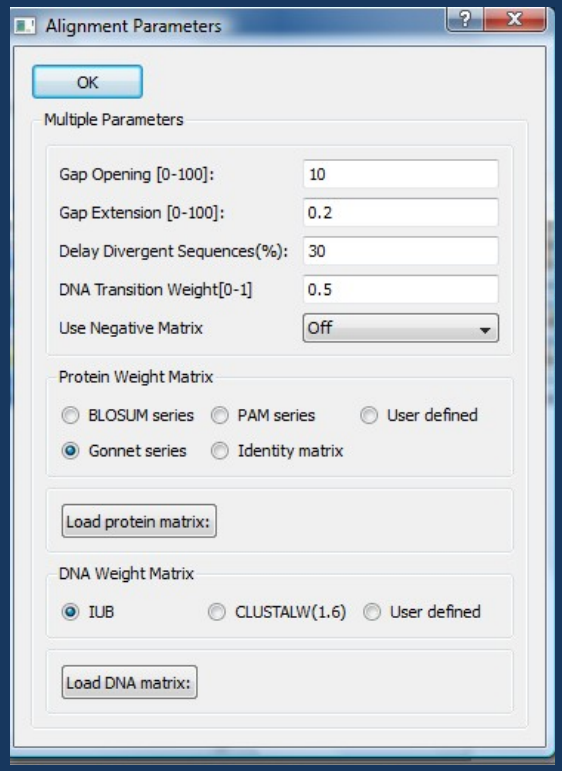
Menu “alignment” permite acessar diversos comandos incluindo menus de ajuste de parâmetros de alinhamentos

Alinhando seqüências no clustalX



- Clicando no item “multiple aligment parameters” é possível acessar um menu onde é possivel ajustar os seguintes parametros
- Gap opening- Penalização de escore para iniciar uma região de “gap”
- Gap extension- Penalização de escore para extender uma região de “gap” (normalmente menor que o Gap opening)
- Delay divergent sequences- Atraso de alinhamento de seqüências divergentes que somente serão alinhadas após as outras sequencias (porcentagem de identidade abaixo da qual que faz a seqüência seja considerada divergente)

Alinhando seqüências no clustalX

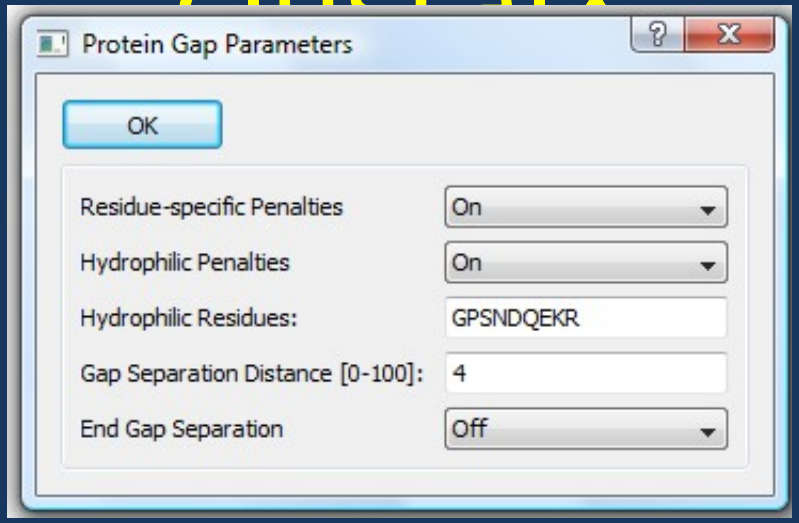


- Transition weight (somente DNA)- da a transições (A<->G, T<->C) um escore diferente de 0.

- Use negative matrix- permite o uso de matrizes negativas, importante quando as seqüências forem relacionadas somente em uma pequena porção. Em condições normais prejudica um pouco o alinhamento

- Protein Weight Matrix- Matriz de substituição a ser utilizada (note que voce só escolhe a serie de matriz: Blosum, PAM, etc... O tipo de matriz dentro desta serie (por exemplo Blosum 62, blosum 80) é escolhido automaticamente pelo programa

Alinhando seqüências no clustalX



- Residue specific penalty- Considera a vizinhança de alguns resíduos como mais ou menos favoráveis para abertura de “gaps”
- Hydrophilic penalties- aumentqa a chance de gaps em regiões ricas em resíduos hidrofílicos, que usualmente representam regiões menos estruturadas.
- Hydrophilic residues- especifica que residuos são considerados hidrofilicos
- Gap separation distance- numero de resíduos de distancia de uma região com gap na qual é penalizada uma nova abertura de gap
- Considera o gap no final da seqüência normalmente para o parâmetro acima

Alinhando seqüências no clustalX

- Como realiza um alinhamento global o clustal tentará realizar o alinhamento de toda a proteína não “jogando fora pedaços” como ocorre no caso de alinhamento por blast. Em adição não há a geração de um parâmetro de confiança que permita você avaliar a significância do seu alinhamento

- Deste modo o clustal é uma ferramenta muito pobre para identificação de função de proteínas. O seu principal uso é para uma caracterização mais fina após a determinação da identidade e domínios presentes na seqüência.

Alinhando seqüências no clustalX

- É muito importante que haja uma seleção criteriosa de que seqüências e quais regiões destas seqüência serão alinhadas

- Muitas vezes ao invés de selecionarmos proteínas inteiras, que podem conter um mosaico de regiões com diversas origens evolutivas, é preferível alinhar somente regiões de domínios em comum

Alinhando seqüências no clustalX

- Alinhamento de múltiplas seqüências são bastante influenciados pelas penalizações de abertura e extensão de gaps e ao fazer alinhamentos o usuário normalmente deve ajustar estes parâmetros de modo a obter um bom alinhamento

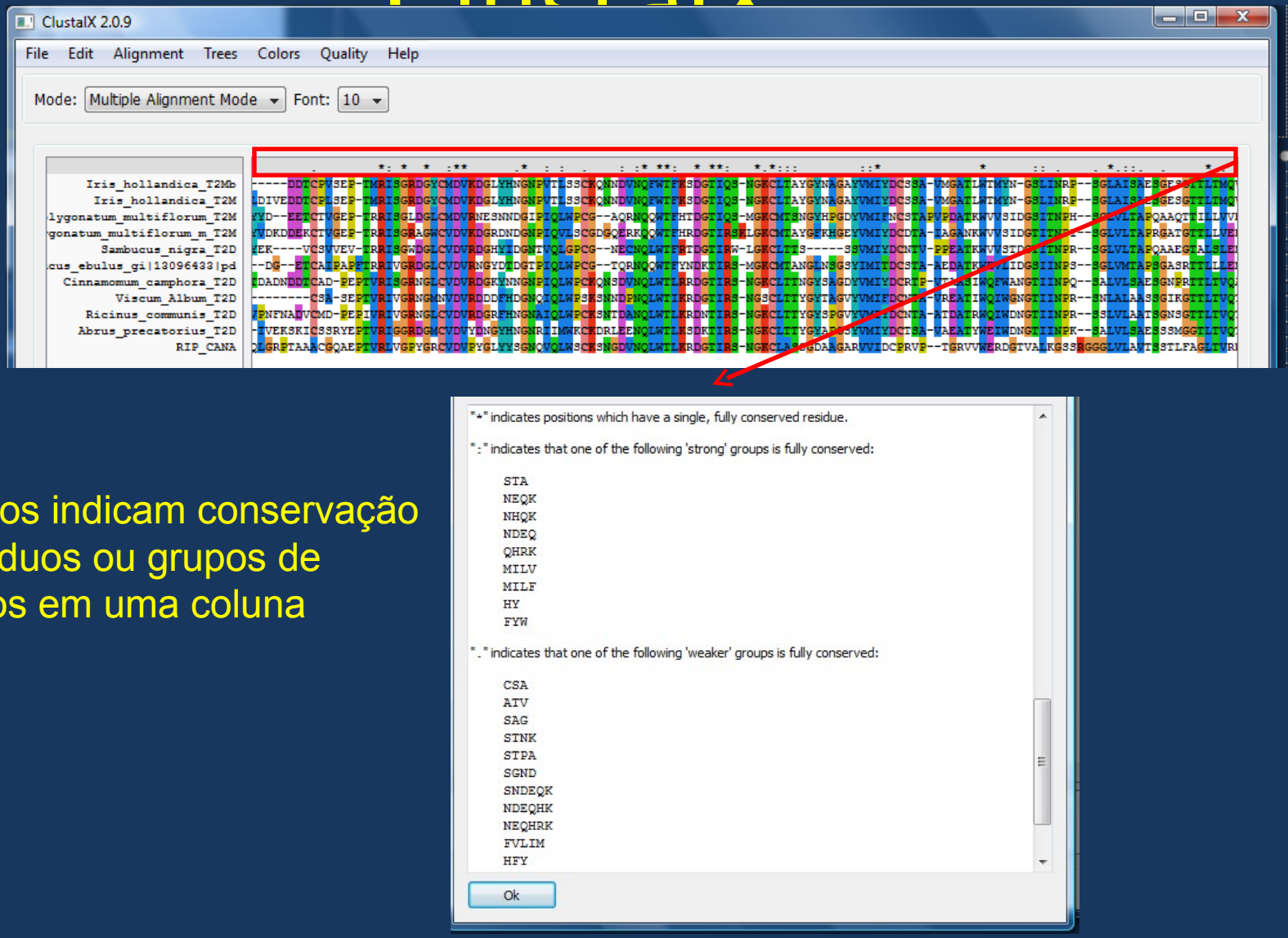
- Para o alinhamento de seqüências mais divergentes é necessário a utilização de penalizações menores para a abertura de gaps

Alinhando seqüências no clustalX

- Devido aos problemas do algoritmo de alinhamento progressivo quando alinhamos seqüências distantes é sempre recomendável inserir no alinhamento múltiplo seqüências adicionais que sejam mais próximas das seqüências analisadas

- De modo geral , ter um numero razoável de seqüência ajuda o programa de alinhamento múltiplo e facilita a interpretação dos dados

Resultado de alinhamento ClustalX

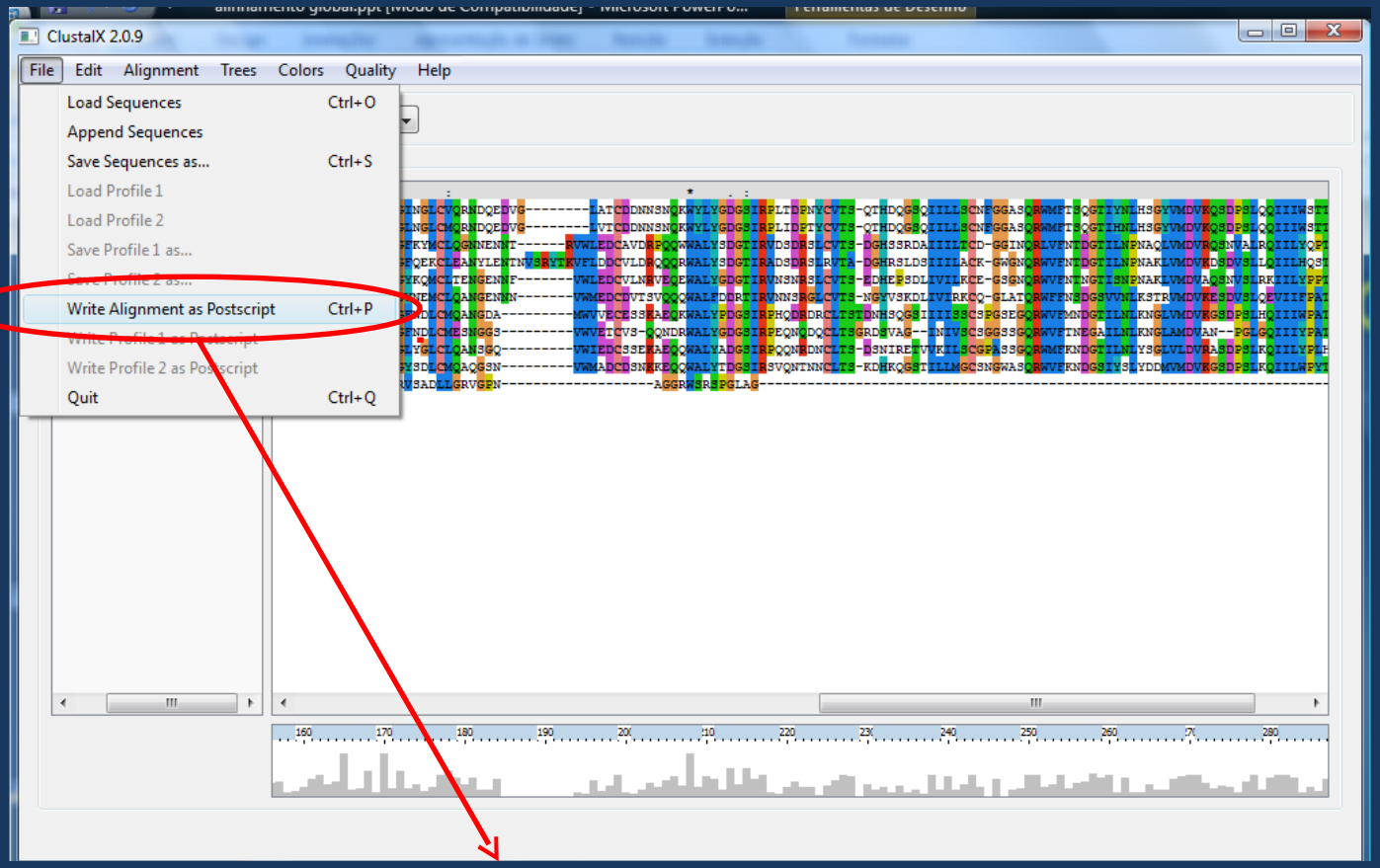


Símbolos indicam conservação de resíduos ou grupos de resíduos em uma coluna

Resultado de alinhamento ClustalX

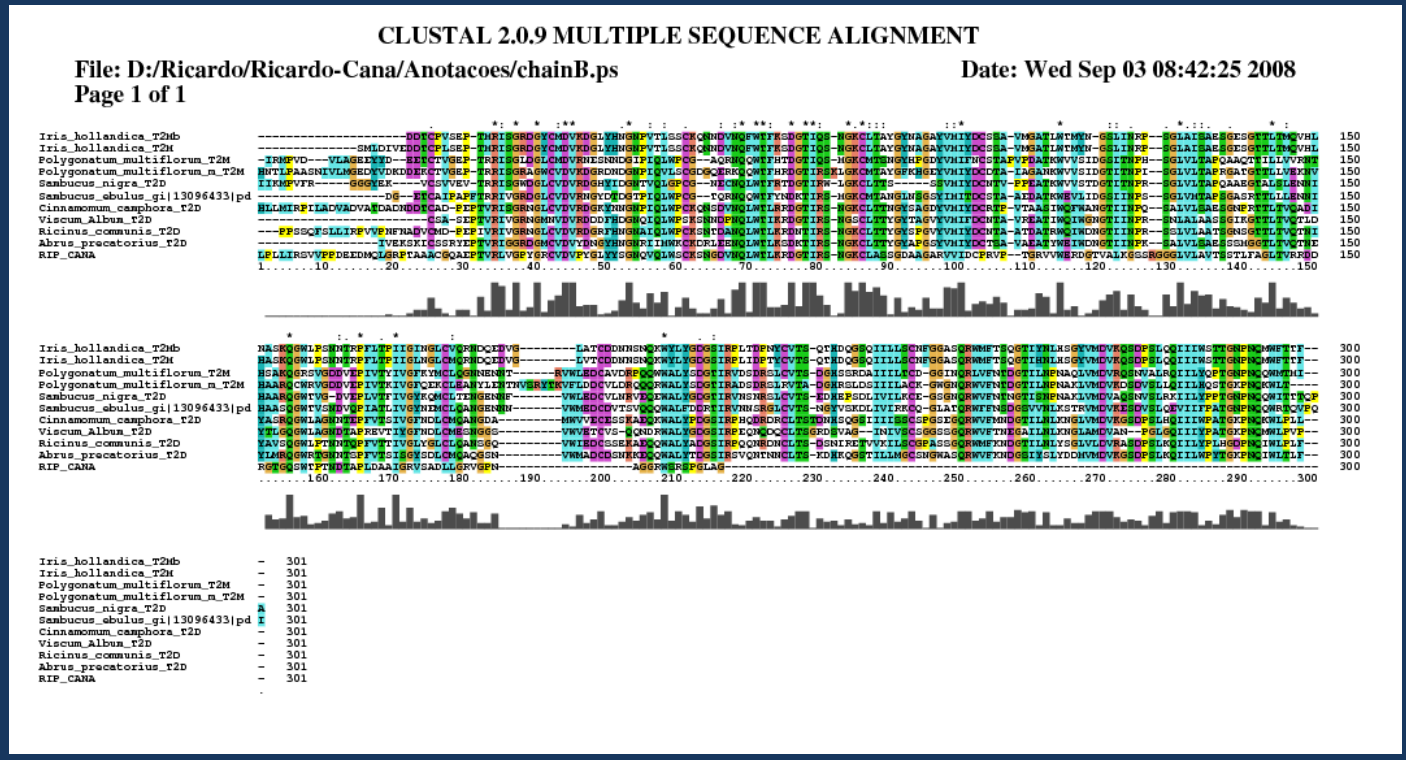
- Low scoring segments- existe a opção de mostrar em cinza regiões com baixo escore e que portanto não seriam muito confiáveis

Visualização do resultado de alinhamento ClustalX



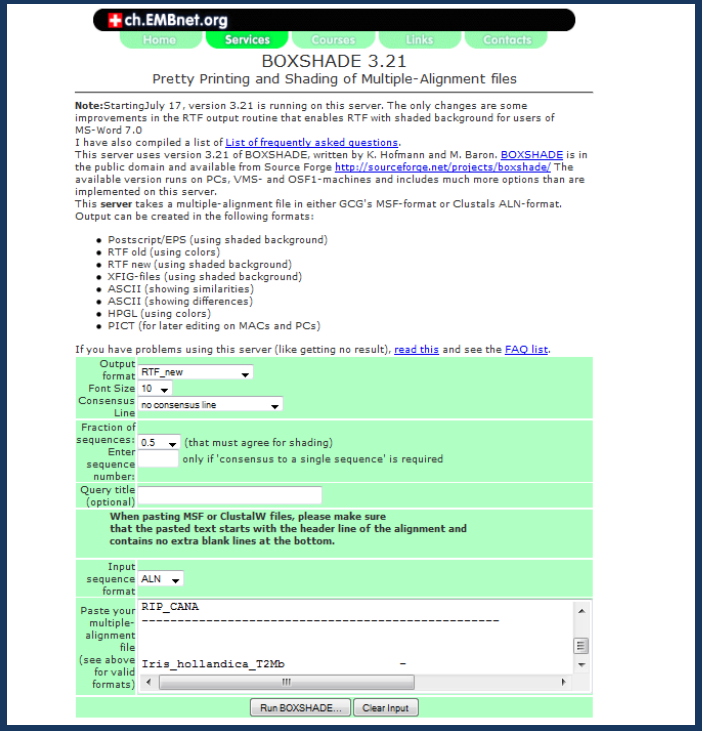
Exporta o alinhamento em formato postscript

Visualização do resultado de alinhamento ClustalX



- Resultado de alinhamento em arquivo postscript

Visualização do resultado de alinhamento ClustalX



Box shade- ferramenta para visualização de alinhamentos. Aceita formato aln (clustal). Pode exportar em formato rtf

http://www.ch.embnet.org/software/BOX_form.html

Visualização do resultado de alinhamento ClustalX