

# Alinhamento múltiplo de sequências

# Alinhamento múltiplo de sequências. O que é ?

- Um alinhamento múltiplo de sequências é simplesmente uma extensão do alinhamento de pares de sequências para um conjunto igual ou superior a 3
- É o estabelecimento de correspondências entre resíduos de diferentes sequências
- A determinação do **alinhamento múltiplo ótimo** de um conjunto de sequências não é um problema trivial e só pode ser resolvido para um pequeno número de sequências
- A geração de alinhamentos múltiplos é normalmente feita com recurso a métodos heurísticos que não garantem a solução ótima

# Exemplo de alinhamento múltiplo

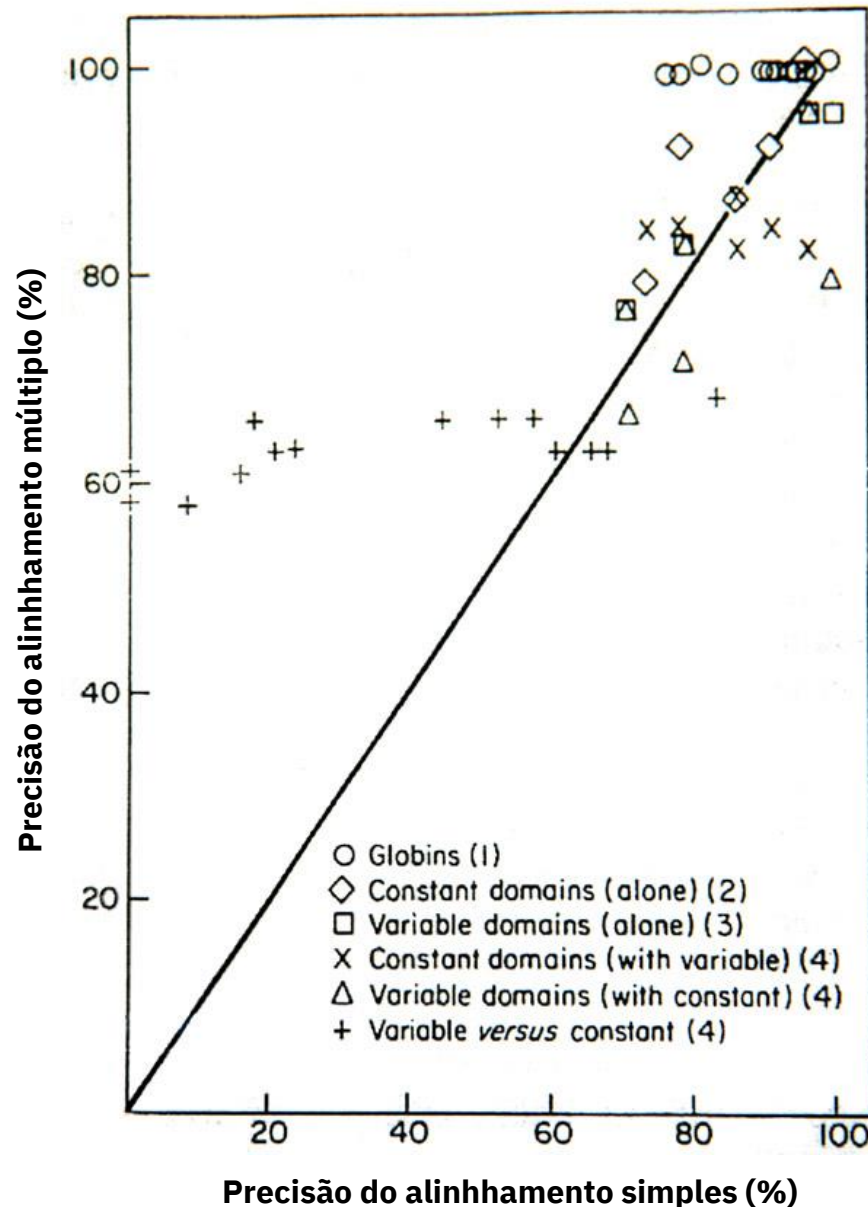
Q5E940_BOVIN	-----MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKOMQIIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE	76
RLA0_HUMAN	-----MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKOMQIIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE	76
RLA0_MOUSE	-----MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKOMQIIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE	76
RLA0_RAT	-----MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKOMQIIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE	76
RLA0_CHICK	-----MPREDRATWKSNYFMKIIQLLDDYPKCFVVGADNVGSKOMQIIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE	76
RLA0_RANSY	-----MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKOMQIIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--SALE	76
Q7ZUG3_BRARE	-----MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKOMQIIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE	76
RLA0 ICTPU	-----MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKOMQIIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE	76
RLA0_DROME	-----MVRENKAANKAQYFIKVVELFDEFKCFIVGADNVGSKOMQIIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE	76
RLA0_DICDI	-----MSGAG-SKRKKLFIEKATKLFITYDKMIVAEADFVGSQLOKIRKSIRGI-GAVLMGKNTMIRKVVIRDLADSK--PELD	75
Q54LP0_DICDI	-----MSGAG-SKRKNVFIEKATKLFITYDKMIVAEADFVGSQLOKIRKSIRGI-GAVLMGKNTMIRKVVIRDLADSK--PELD	75
RLA0_PLAF8	-----MAKLSKQQKKQMYIEKLSSLIQQYSKILIVHVDNVGSKOMQIIRMSLRGK-ATILMGKNTMIRKVVIRDLADSK--PELD	76
RLA0_SULAC	-----MIGLAVTTTKKIAKWKVDEVAELTEKLKTKTIIIANIEGFPADKLHEIRKKLRGK-ADIKVTNNLNFNIALKNAG----YDTK	79
RLA0_SULTO	-----MRIMAVITQERKIAKWKIEEVKELEKREYHTIIIANIEGFPADKLHDIRKKMRGM-AEIKVTNNTLFGIAAKNAG----LDVS	80
RLA0_SULSO	-----MKRLALALKQRKVASWKELEKTELKNSNTILIGNLEGFPADKLHEIRKKLRGK-ATIKVTNNTLFGIAAKNAG----IDIE	80
RLA0_AERPE	MSVVSIVGQMYKREKPIPEWKTLMLELEELFSKHVVVLFADLTGTPTFVVRVRKKLWKK-YPMVAKKRIILRAMKAAGLE---LDDN	86
RLA0_PYRAE	MMLAIGKRRYVRTQYPARKVKIVSEATELLQKYPYVFLFDLHGLSSRIHEYYRYRLRY-GVIKIKPTLFLKIAFTKVYGG---IPAE	85
RLA0_METAC	-----MAEERHHTHEIPQWKDEIENIKELIQSHKVFVMVGIEGILATKMKIRRDLDKV-AVLKVSNTLTERRALNQLG---ETIP	78
RLA0_METMA	-----MAEERHHTHEIPQWKDEIENIKELIQSHKVFVMVGIEGILATKMKIRRDLDKV-AVLKVSNTLTERRALNQLG---ESIP	78
RLA0_ARCFU	-----MAAVRGS--PPEYKVRAVEEIKRMISSEVVAIVSFRNVPAQOMQIRREFRGK-AEIKVVNTLTERALDAG---GDYL	75
RLA0_METKA	MAYKAKGQPPSGYEPKVAEWKRREVKELKELMDEVENVGLVDLEGIPAPQLOEIRAKLRERDTIIRMSRNTLMRIAEEKLDER--PELE	88
RLA0_METTH	-----MAHVAEWKKKEVQELHDLIKGYEVVGIANLADIPARQLQKMRQTLRDS-ALIRMSKKTILISLAEKAGREL--ENVN	74
RLA0_METTL	-----MITAESEHKIAPWKIEEVNKLKELLKNGQIVALVDMMEVPAVQLQEIIRDKIR-GTMTLKMSRNTLIERAIEVAEETGNPEFA	82
RLA0_METVA	-----MIDAKSEHKIAPWKIEEVNKLKELLKNSANVIALIDMMEVPAVQLQEIIRDKIR-DQMTLKMSRNTLIKRAVEEVAEETGNPEFA	82
RLA0_METJA	-----METKVKAHVAPWKIEEVKTLKGLIKSKPVVAIVDMMDVPAPQLQEIIRDKIR-DKVKLRMSRNTLIIRALKEAAEELNNPKLA	81
RLA0_PYRAB	-----MAHVAEWKKKEVEELANLIKSPVIALVDVSSMPAYPLSQMRRLIRENGGLLRVSNTLIELAIKKAAGELGKPELE	77
RLA0_PYRHO	-----MAHVAEWKKKEVEELAKLIKSPVIALVDVSSMPAYPLSQMRRLIRENGGLLRVSNTLIELAIKKAAGELGKPELE	77
RLA0_PYRFU	-----MAHVAEWKKKEVEELANLIKSPVVALVDVSSMPAYPLSQMRRLIRENGLLRVSNTLIELAIKKAAGELGKPELE	77
RLA0_PYRKO	-----MAHVAEWKKKEVEELANLIKSPVIALVDVAGVPAYPLSKMRDKLR-GKALLVSRNTLIELAIKKAAGELGKPELE	76
RLA0_HALMA	MSAESERKTETIPEWKQEEVDAIVEMIESYESVGVVNIAGIPSRLODMRRDLHGT-AELRVSNTLLEALDDVD---DGLE	79
RLA0_HALVO	MSESEVRQTEVIPQWKREEVDLVDLIESYESVGVVGVAGIPSRLODMRRDLHGT-AELRVSNTLLEALDDVD---DGFE	79
RLA0_HALSA	MSAEQRTTEEVPEWKRQEVAEVLDELLETYSVGVVNVGTGIPSKOLODMRRDLHGT-AALRVSNTLLEALDEEAG---DGLD	79
RLA0_THEAC	-----MKEVSQKKELVEELANLIKSPVIALVDVAGVPAYPLSKMRDKLR-GKALLVSRNTLIELAIKKAAGELGKPELE	72
RLA0_THEVO	-----MRKINPKKKEIVSELAQDITKSKAIVDIDKGVTRMODIRAKNRDK-VKIKVVKTLLFKALDSIND---EKLT	72
RLA0_PICTO	-----MTEPAQWKIDFVKNLENEINSRKVAAIVSIKGLRNNFQKIRNSIRDK-ARIKVSARLLRLAIENTGK---NNIV	72
ruler	1.....10.....20.....30.....40.....50.....60.....70.....80.....90	

Alinhamento de membros da família das **proteínas ribossomais L10P** de diversos organismos

# Importância do alinhamento múltiplo de sequências

- Os alinhamentos múltiplos de sequências são uma ferramenta central para a inferência da função das proteínas por comparação das suas sequências
- Os alinhamentos múltiplos são o ponto de partida para a previsão da estrutura secundária e identificação dos resíduos importantes para a especificidade
- Os alinhamentos múltiplos são a base dos métodos de pesquisa de sequências mais sensíveis de que dispomos (ex.: PSI-Blast)
- Os alinhamentos múltiplos são ainda o ponto de partida para a construção de árvores filogenéticas e determinação das relações evolutivas entre organismos
- Os alinhamentos múltiplos são uma forma conveniente de anotar as características estruturais e funcionais comuns a uma família de proteínas.

# O alinhamento múltiplo aumenta a precisão do alinhamento simples



Comparação da precisão de alinhamentos de pares de sequências quando produzidos de forma isolada ou fazendo parte de um alinhamento múltiplo.

Pode ver-se que na maior dos casos a precisão obtida com o alinhamento múltiplo é superior (valores acima da diagonal).

(A precisão é avaliada através da comparação com alinhamentos **estruturais**)

# Alinhamento múltiplo: métodos

Os métodos para a produção de alinhamentos de 3 ou mais sequências podem ser divididas em várias categorias:

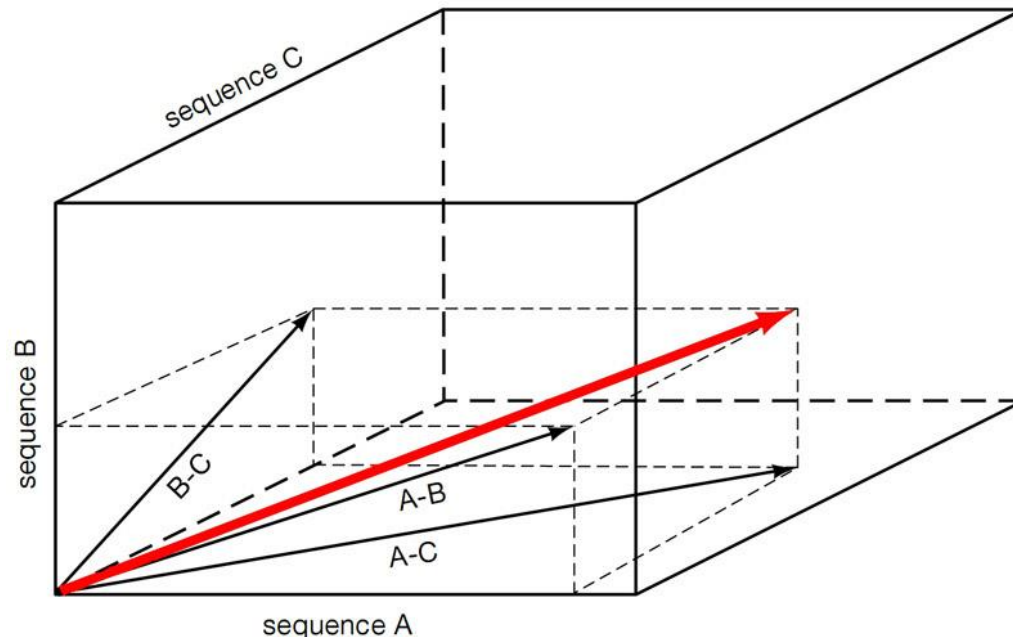
- **Extensão dos métodos ótimos para N sequências:** o algoritmo de N-W pode ser estendido para 3 ou mais sequências, mas exige o uso de matrizes multi-dimensionais e torna-se muito pesado computacionalmente (exp.: MSA)
- **Métodos progressivos (ou hierárquicos) :** baseiam-se na aplicação sucessiva de métodos ótimos a todos os pares de sequências, depois a pares de pares, etc., através de uma estrutura em árvore. São os métodos mais usados. (Exp: Clustalw, t-coffee)
- **Métodos iterativos:** geral um alinhamento global inicial de todas as sequências, que é refinado em passos sucessivos (SAGA, DIALIGN)
- **Métodos de segmentos:** comparação de “janelas” de comprimento fixo nas várias sequências (p.exp.: MACAW)

# Programação dinâmica a N dimensões

A extensão directo dos algoritmos de Needleman-Wunsch ou Smith-Waterman para N sequências torna-se impraticável computacionalmente: o alinhamento óptimo é agora um caminho num cubo a N dimensões.

Se tivermos N sequências de comprimento L, a matriz terá  $L^N$  células

Exemplo: 10 sequências de comprimento 200 -  $200^{10} = \mathbf{10^{22}}$  células !



Matriz para o alinhamento múltiplo de 3 sequências  
(a seta vermelha representa o caminho óptimo na matriz)



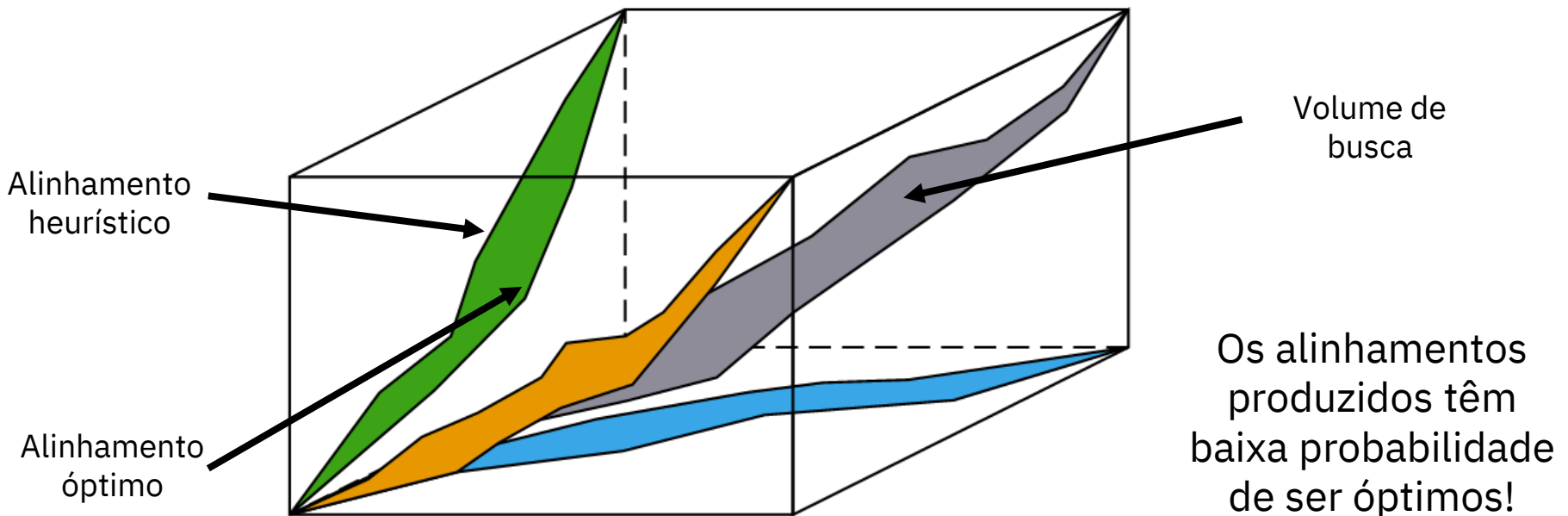
# Algoritmo de Carrillo-Lipman-Gupta

Este método é uma simplificação que reduz o espaço de busca e permite encontrar um alinhamento *próximo* do ótimo.

O método começa por definir intervalos para o alinhamento de cada par de sequências, e usa estes intervalos para definir um volume de busca dentro do hipercubo.

Implementado no programa **MSA** - demasiado pesado para ser usado com mais de 25-30 sequências com ~100 aminoácidos.

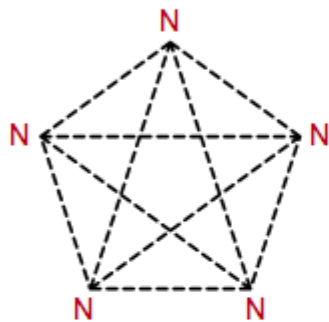
Não existem servidores de acesso livre para este programa.



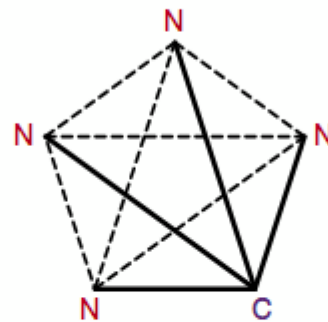


# Cálculo do score num alinhamento múltiplo: o método SP (sum of pairs)

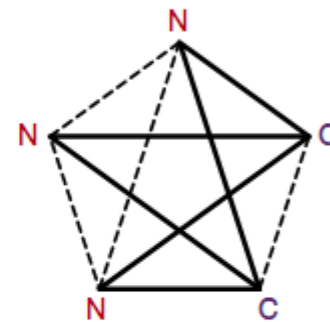
Sequence	Column A	Column B	Column C
1	.....N.....	.....N.....	.....N.....
2	.....N.....	.....N.....	.....N.....
3	.....N.....	.....N.....	.....N.....
4	.....N.....	.....N.....	.....C.....
5	.....N.....	.....C.....	.....C.....



Column A



Column B



Column C

No. of N-N matched pairs (each scores 6):

10

6

4

No. of N-C matched pairs (each scores -3):

0

4

6

BLOSUM62 score :

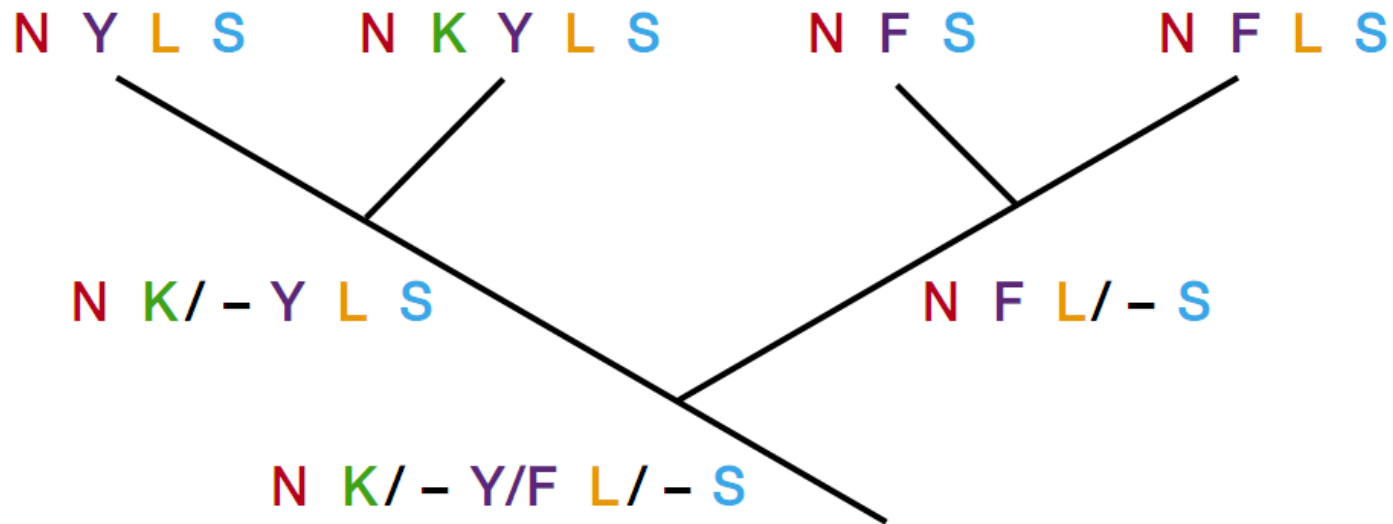
60

24

6

# Métodos de alinhamento progressivo

Os métodos de alinhamento progressivo usam o algoritmo de programação dinâmica calcular distâncias entre pares de sequências. As distâncias são usadas para construir uma árvore que serve de guia para criação do alinhamento múltiplo.



# Software para alinhamento múltiplo progressivo

- **CLUSTALW:**

Um dos softwares mais usados, existe também como um programa que pode ser instalado e executado no PC.

**DEIXOU DE SER SUPORTADO PELO EBI**

- **CLUSTAL OMEGA:**

Nova versão do programa clustalw que usa modelos HMM em vez de matrizes de score. Recomendado para proteínas.

<http://www.ebi.ac.uk/Tools/msa/clustalo/>

- **T-COFFEE:**

Mais rigoroso, mas mais lento que CLUSTALW, usa uma combinação de vários métodos, incluindo a geração de alinhamentos de pares seu óptimos com o programa LALIGN. Recomendado para alinhamentos pequenos.

<http://www.ebi.ac.uk/Tools/msa/tcoffee/>

- **MUSCLE:**

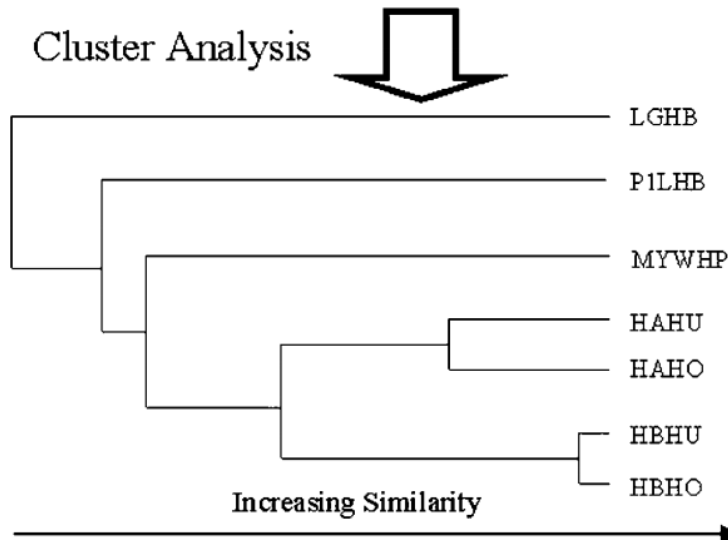
Recomendado para alinhamentos de DNA/RNA

<http://www.ebi.ac.uk/Tools/msa/muscle/>

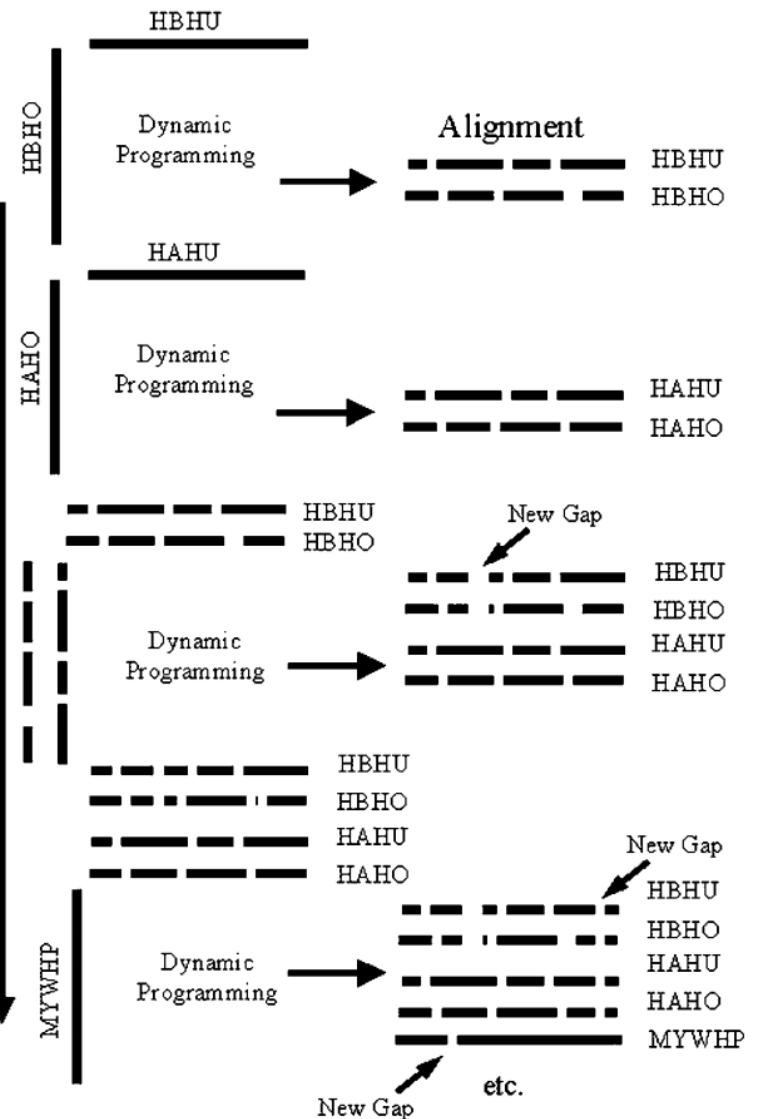
# CLUSTALW

	HAHU	HBHU	HAHO	HBHO	MYWHP	P1LHB	LGHB
HAHU							
HBHU	21.1						
HAHO	32.9	19.7					
HBHO	20.7	<b>39.0</b>	20.4				
MYWHP	11.0	9.8	10.3	9.7			
P1LHB	9.3	8.6	9.6	8.4	7.0		
LGHB	7.1	7.3	7.5	7.4	7.3	4.3	

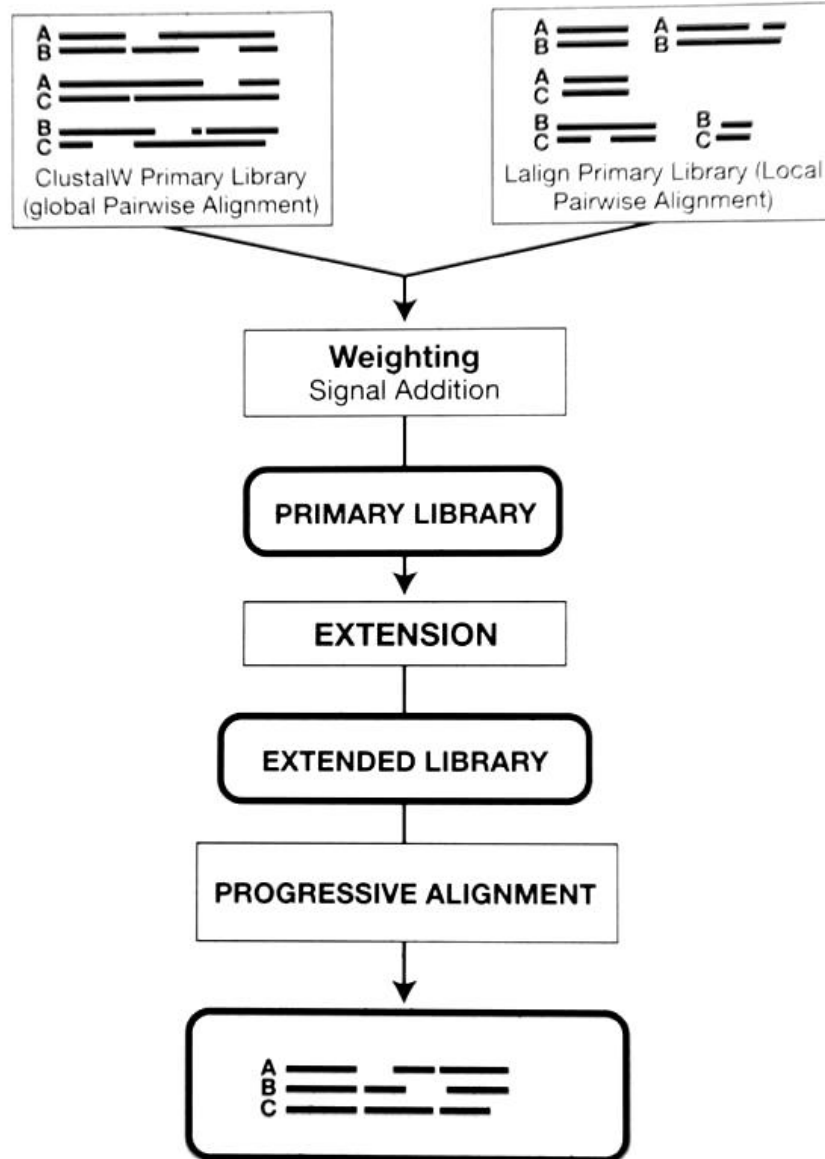
Cluster Analysis



Multiple Alignment



# T-Coffee



# Métodos iterativos alinhamento

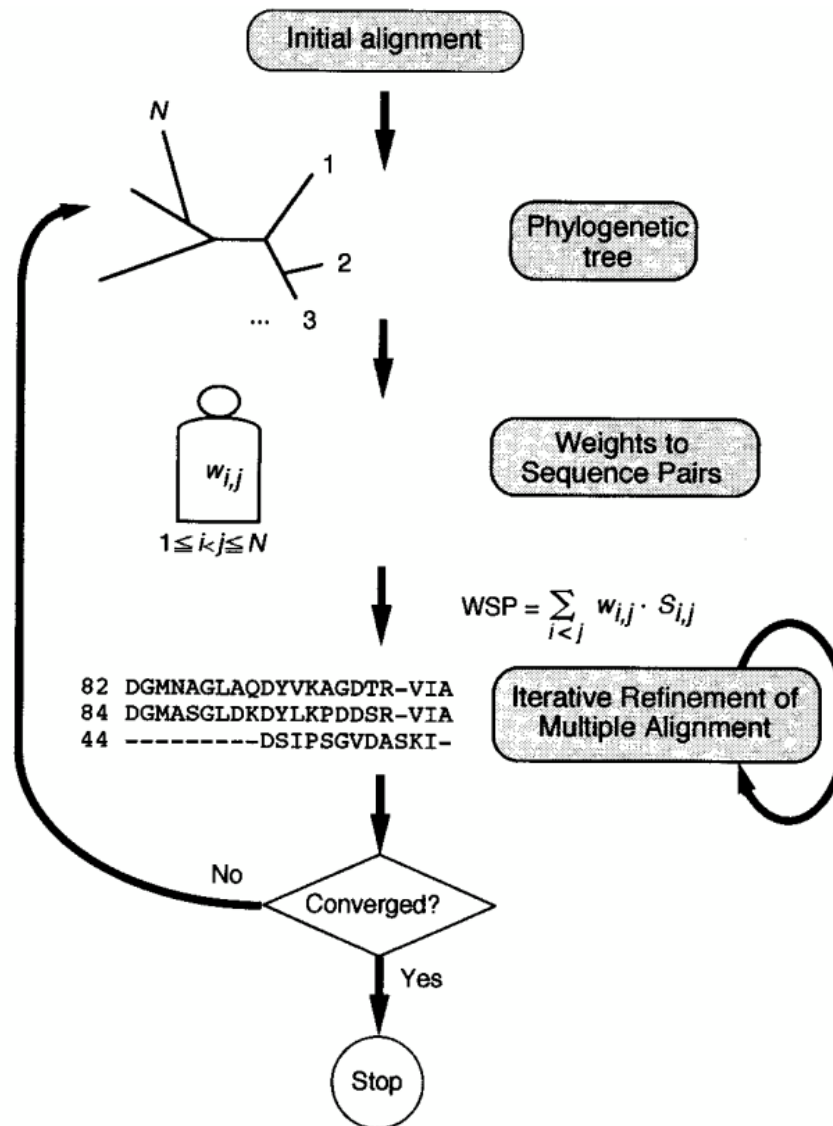
Os métodos de alinhamento progressivo têm como principal problema a propagação dos erros nos alinhamentos iniciais para o alinhamento final. Os métodos iterativos obviam esta situação através de repetidos passos de alinhamento global, com vista à otimização do score (por exemplo SP).

- **DIALIGN:** pesquisa de alinhamentos locais sem gaps em pares de sequências, pesados para o cálculo e otimização do alinhamento final.  
<http://bibiserv.techfak.uni-bielefeld.de/dialign>

- **PRRP/PRRN:** refinamento iterativo de um alinhamento progressivo com construção de árvore e uso de pesos no alinhamento de pares.  
<http://prrn.hgc.jp>

- **SAGA:** método iterativo baseado num algoritmo genético. Não está disponível na forma de serviço “on-line”. É bastante pesado computacionalmente.  
[http://www.tcoffee.org/Packages/SAGA\\_V0.95.tar.gz](http://www.tcoffee.org/Packages/SAGA_V0.95.tar.gz)

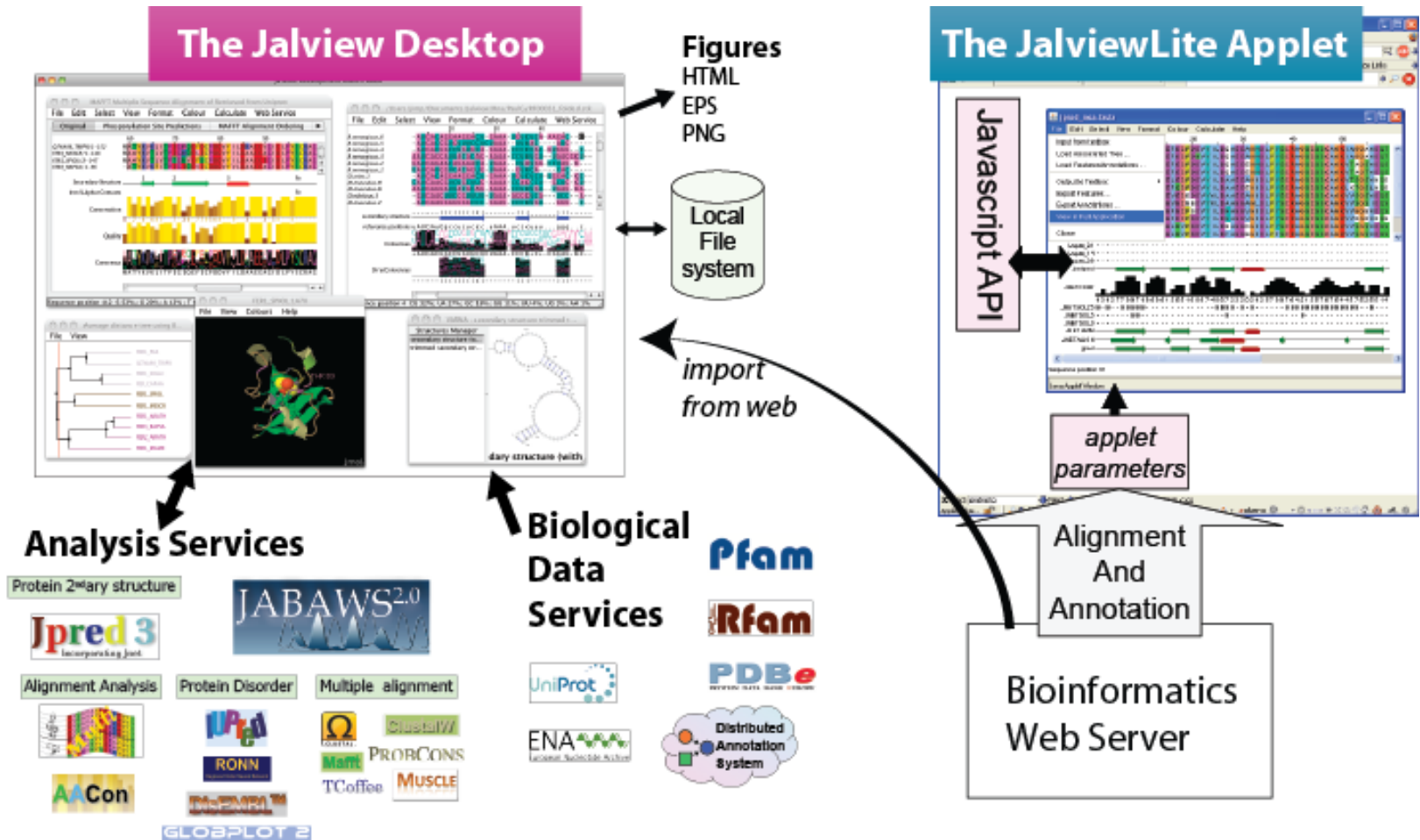
# Alinhamento iterativo com PRRP/PRRN





## Alinhamento de sequências da tioredoxina

# Edição interactiva de alinhamentos com Jalview



<http://www.jalview.org/>

# Clustal Omega

Input form

Web services

Help & Documentation

Bioinformatics Tools FAQ

Feedback

Tools > Multiple Sequence Alignment > Clustal Omega

## Service Announcement

The new Job Dispatcher Services beta website is now available at <https://wwwdev.ebi.ac.uk/Tools/jdispatcher>. We'd love to hear your [feedback](#) about the new webpages!

## Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment tools](#).

**Important note:** This tool can align up to 4000 sequences or a maximum file size of 4 MB.

### STEP 1 - Enter your input sequences

Enter or paste a set of

PROTEIN

sequences in any supported format:

Or, upload a file:  No file chosen

[Use a example sequence](#) | 
 [Clear sequence](#) | 
 [See more example inputs](#)

### STEP 2 - Set your parameters

OUTPUT FORMAT

ClustalW with character counts

[DEALIGN INPUT SEQUENCES](#)


[MBED-LIKE CLUSTERING GUIDE-TREE](#)

[MBED-LIKE CLUSTERING ITERATION](#)

[NUMBER of COMBINED ITERATIONS](#)

# Clustal Omega

[Input form](#)
[Web services](#)
[Help & Documentation](#)
[Bioinformatics Tools FAQ](#)


[Feedback](#)

Tools > Multiple Sequence Alignment > Clustal Omega

## Service Announcement

The new Job Dispatcher Services beta website is now available at <https://wwwdev.ebi.ac.uk/Tools/jdispatcher>. We'd love to hear your [feedback](#) about the new webpages!

## Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment tools](#).

**Important note:** This tool can align up to 4000 sequences or a maximum file size of 4 MB.

### STEP 1 - Enter your input sequences

Enter or paste a set of

PROTEIN

sequences in any supported format:

```
>sp|P00762|TRY1_RAT Serine protease 1 OS=Rattus norvegicus OX=10116 GN=Prss1 PE=1 SV=1
MSALLLILALV GAAVAFPLED DDKIVGGYTC PEHVSVPYQVS LNSGYHFCGG SLINDQWVVS
AAHCYKSRIQ VRLGEHNINV LEGDEQFINA AKIHKHPNYS SWTLNNDIML IKLSSSPVKLN
ARVAPVALPS ACAPAGTQCL ISGWGNTLSN GVNNPDLLQC VDAPVLSQAD CEAYPGEIT
SSMICVGFLF GGDSCQGDG GGPVVCNGQL QGIVSWGYG C ALPDNPGVYT KVCNFGVGIQ
DTIAAN
```

```
>sp|P07473|TRY1_HUMAN Serine protease 1 OS=Homo sapiens OX=9606 GN=PRSS1 PE=1 SV=1
```

Or, [upload a file](#):  No file chosen

[Use a example sequence](#) | 
 [Clear sequence](#) | 
 [See more example inputs](#)

### STEP 2 - Set your parameters

OUTPUT FORMAT

ClustalW with character counts

[DEALIGN INPUT SEQUENCES](#)

[MBED-LIKE CLUSTERING GUIDE-TREE](#)

[MBED-LIKE CLUSTERING ITERATION](#)

[NUMBER of COMBINED ITERATIONS](#)

<a href="#">Input form</a>	<a href="#">Web services</a>	<a href="#">Help &amp; Documentation</a>	<a href="#">Bioinformatics Tools FAQ</a>
----------------------------	------------------------------	--	--

Feedback

## Service Announcement

The new Job Dispatcher Services beta website is now available at <https://www.dev.ebi.ac.uk/Tools/jdispatcher>. We'd love to hear your feedback about the new webpages!

Results for job clustalo-l20231113-121354-0410-67345595-p1m

Alignments Result Summary Guide Tree Phylogenetic Tree Results Viewers Submission Details

Download Alignment File Show Colors

CLUSTAL O(1.2.4) multiple sequence alignment

sp P51580 TRVP_SARBU	-----MRLFTAV--FAL--	10
sp P35031 TRV1_SALSA	-----MRLSVLE-----	7
sp P07477 TRV1_HUMAN	-----MRLPLLI-----	7
sp P35030 TRV1_HUMAN	MGDPDRC <sup>1</sup> PAHPGPGRAVCKGKLAARPGVRERGGAGGAGLEL <sup>1</sup> LGRTVAAR <sup>1</sup>	60
sp P07146 TRV2_MOUSE	-----MSALLLI-----	7
sp P08762 TRV1_RAT	-----MRLPLLI-----	7
sp P08763 TRV2_RAT	-----MRLPLLI-----	7
sp P08760 TRV1_BOVIN	-----MRLPLLI-----	7
sp P08761 TRV1_PIG	-----MRLTFL-----	7

[illegible]

sp155181 TRV1_SARSA	AMGYSVTSQVIGLGVSTYY--HGGELGVYSKAFKFGHGYMDTMDVALLKSLTPA	123
sp155081 TRV1_SALSA	AMCHYS--RVELRGHGMIVLGELEGFVSSSRVIRHGYMDTMDVALLKSLTPA	124
sp150747 TRV1_HUMAN	GMCHYS--RIGRLGHGMIVLELEGEGFVNAIKTRHPQVDTMDLMDLTKSSRA	125
sp150747 TRV1_HUMAN	AMCHYT--RIGRLGHGMIVLELEGEGFVNAIKTRHPQVDTMDLMDLTKSSRA	126
sp150747 TRV1_HUMAN	AMCHYT--RIGRLGHGMIVLELEGEGFVNAIKTRHPQVDTMDLMDLTKSSRA	127
sp150747 TRV1_HUMAN	AMCHYT--RIGRLGHGMIVLELEGEGFVNAIKTRHPQVDTMDLMDLTKSSRA	128
sp150762 TRV1_RAT	AMCHYS--RIGRLGHGMIVLELEGEGFVNAIKTRHPQVDTMDLMDLTKSSRA	129
sp150763 TRV1_RAT	AMCHYS--RIGRLGHGMIVLELEGEGFVNAIKTRHPQVDTMDLMDLTKSSRA	130
sp150760 TRV1_BOVIN	AMCHYS--GIQRLGHGMIVLELEGEGFVSSKSTVSYSHSLMDLMDLTKSSAA	131
sp150761 TRV1_PIG	AMCHYS--RIGRLGHGMIVLELEGEGFVNAIKTRHPHGFDTMDLMDLTKSSA	132

sp155181 TRV1_SARSA	RESSKIFVRIALDRITDTPGPAWGTGCTFVCLTFLQVEQVEVQKQASG--	181
sp155080 TRV1_SALSA	LITNVYPAALDPTSCAPATGCTVSGMGTHSN--DSD--SHKQLCKVTSYDQ--H	182
sp156747 TRV1_HUMAN	VZVNAVSTSLTTPAPATGCTLKSGMTGAS--SGADYPEQLQDAPLVSQACE--	173
sp156748 TRV1_HUMAN	VZVNAVSTSLTTPAPATGCTLKSGMTGAS--FGADYPEQLQDAPLVTAQCE--	173
sp156749 TRV1_HUMAN	VZVNAVSTSLTTPAPATGCTLKSGMTGAS--FGADYPEQLQDAPLVTAQCE--	173
sp156750 TRV1_HUMAN	VZVNAVSTSLTTPAPATGCTLKSGMTGAS--FGADYPEQLQDAPLVTAQCE--	173
sp156761 TRV1_RAT	KLNAVPAALPSCAPATGCTLKSGMTGAS--GVNDFVQLQDAPLVLSQACE--	173
sp156763 TRV2_RAT	KLNAVPAALPSCAPATGCTLKSGMTGAS--GVNDFVQLQDAPLVLPQACE--	173
sp156760 TRV1_BOVIN	SLHRSVASTSLTSCASAGCTGCTLKSGMTGAS--GVTSVPVQLCKAPLVSDSKC--	173
sp156761 TRV1_PIG	SLHRSVASTSLPSCAGATGCTLKSGMTGAS--GVTSVPVQLCKAPLVSDSKC--	158

[illegible]

sp P51580 TRYP_SARBU	SVRSNKTAKEL-	254
sp P35031 TRY1_SALSA	IFNHNLTSMASY-	242
sp P07477 TRY1_HUMAN	NYVNIQNTIAAMS	246
sp P35030 TRY1_HUMAN	NYVNIQNTIAAMS	304
sp P07146 TRY1_MOUSE	NYVNIQNTIAAM-	247
sp P00762 TRY1_RAT	NYVNIQNTIAAM-	246
sp P00763 TRY2_RAT	NYVNIQNTIAAM-	246
sp P00760 TRY1_BOVIN	NYVNIQNTIAAM-	246
sp P00761 TRYB_PIG	NYVNIQNTIAAM-	231

PLEASE NOTE: Showing colors on large alignments is slow.



## STEP 2 - Set your parameters

### OUTPUT FORMAT

ClustalW with character counts

### DEALIGN INPUT SEQUENCES

no

### MBED-LIKE CLUSTERING GUIDE-TREE

yes

### MBED-LIKE CLUSTERING ITERATION

yes

### NUMBER of COMBINED ITERATIONS

default(0)

### MAX GUIDE TREE ITERATIONS

default

### MAX HMM ITERATIONS

default

### DISTANCE MATRIX

no

### GUIDE TREE

yes

### ORDER

aligned

## STEP 3 - Submit your job

☐ Be notified by email *(Tick this box if you want to be notified by email when the results are available)*

Submit

# Protocolo de alinhamentos múltiplos

1. Encontrar as sequências a alinhar, através de pesquisas em bases de dados ou por outra via
2. Definir as regiões de cada sequência a incluir no alinhamento (não tentar alinhar regiões demasiado diferentes!)
3. Avaliar o grau de semelhança das sequências através dos alinhamentos de pares
4. Começar por alinhar as sequências mais semelhantes, adicionar em seguida as mais distantes
5. Inspeccionar o alinhamento obtido, procurando problemas: regiões com demasiados gaps, baixa conservação, conflito com outras fontes de informação (p.exp. localização do centro activo). Corrigir manualmente com um editor de alinhamento (p.exp. Seaview ou Jalview)
6. Remover as sequências que “destroem” o alinhamento, re-alinhar as restantes
7. Usar os resíduos-chave conservados no sub-alinhamento como guia para a adição de novas sequências