

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

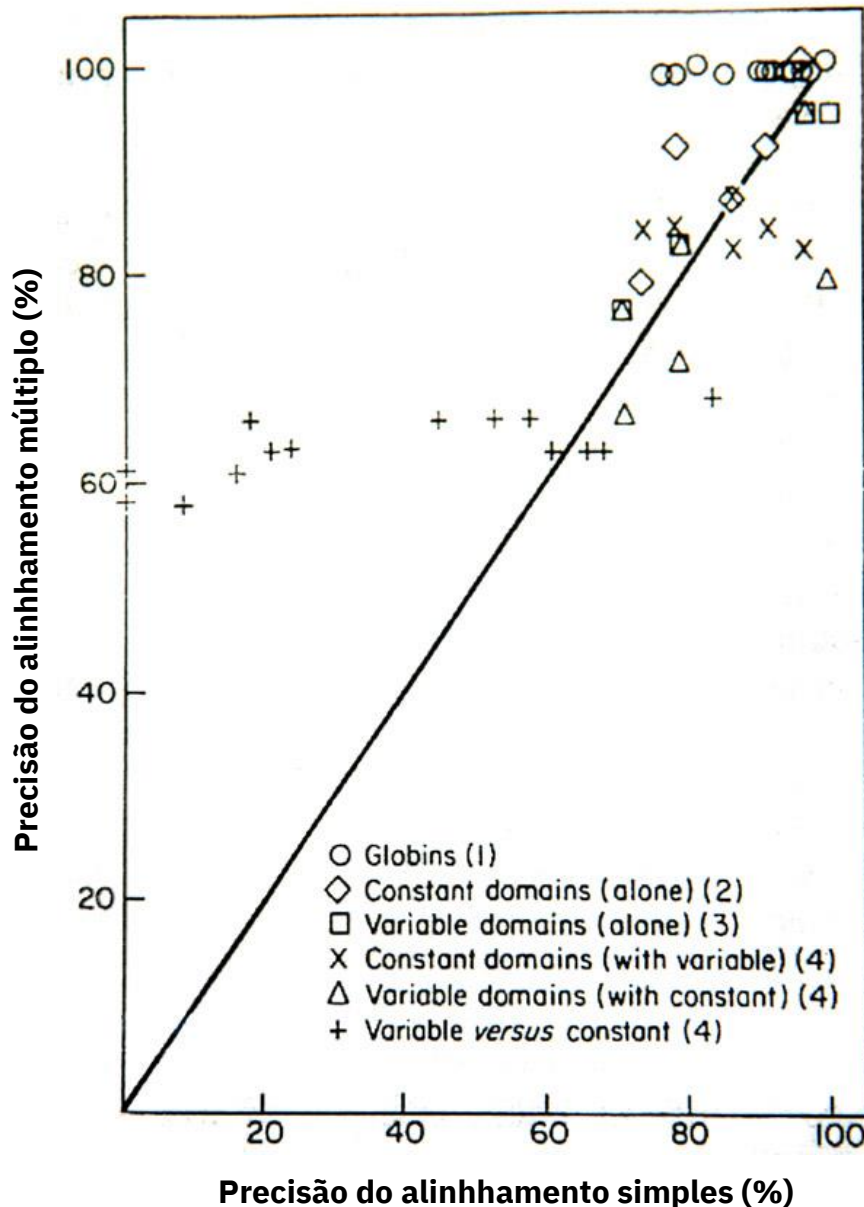
Q5E940_BOVIN	-----MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKOMQIIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE	76
RLA0_HUMAN	-----MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKOMQIIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE	76
RLA0_MOUSE	-----MPREDRATWKSNYFLKIIQLLDDYPKCFIVGADNVGSKOMQIIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE	76
RLA0_RAT	-----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	-----MVRENKAANKAQYFIKVVLFDFPKCFIVGADNVGSKOMQIIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PQLE	76
RLA0_DICDI	-----MSGAG-SKRKKLFIEKATKLFYTDKMIVAEADNVGSKOMQIIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PELD	75
Q54LP0_DICDI	-----MSGAG-SKRKNVFIEKATKLFYTDKMIVAEADNVGSKOMQIIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PELD	75
RLA0_PLAF8	-----MAKLSKQKQKQMYIEKLSSLIQQYSKILIVHVDNVGSKOMQIIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PQIE	76
RLA0_SULAC	-----MIGLAVTTTKKIAKWKVDEVAELTEKLTHTKTIIANIEGFPADKLEIRKCLRGM-ADIKVTKNNLNFNIALKNAG-----YDTK	79
RLA0_SULTO	-----MRIMAVITQERKIAKWKIEVKELEKLRHTHTIIANIEGFPADKLEIRKCLRGM-ADIKVTKNNLNFNIALKNAG-----LDVS	80
RLA0_SULSO	-----MKRLALALKQRKVASWKLIEVKELTELKNSNTILIGNLEGFPADKLEIRKCLRGM-ADIKVTKNNLNFNIALKNAG-----IDIE	80
RLA0_AERPE	MSVVS LVGQMYKREKPIPEWKTLMLELEELFSKIRVVLFADLTGTPTFVQVVRVKKLWKK-YPMVAKKRIILRAMKAAGLE---LDDN	86
RLA0_PYRAE	MMLAIGKRRYVRTRQYPARKVIVSEATELLQKYPYVFLDLHLGLSSRIHLEYRYRLRY-GVIKIIKPTLFLKIAFTKVYGG---IPAE	85
RLA0_METAC	-----MAEERHHTHEHIPQWKKDEIENIKELIQSHKVFQMGVIEGILATKMKIRRDLDKV-AVLKVSNTLTERALNQLG-----ETIP	78
RLA0_METMA	-----MAEERHHTHEHIPQWKKDEIENIKELIQSHKVFQMGVIEGILATKMKIRRDLDKV-AVLKVSNTLTERALNQLG-----ESIP	78
RLA0_ARCFU	-----MAAVRGS--PPEYKVRAVEEIKRMISSKEVVAIVSFRNVPAGOMQIIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PQLE	75
RLA0_METKA	MAVKAKGQPPSGYE PKVAEWRREVKELKELMDEVENVGLVDLEGIPAPQOEIRAKLRERDTIIRMSRNTLMRIAIEEKLDER--PELE	88
RLA0_METTH	-----MAHVAEWKKEVEELANLIKSPVIALVDVSSMPAYPLSQMRRRIRENGGLLRVSRNTLIE LAIKKAAQELGKPELE	77
RLA0_METTL	-----MITAESEHKIAPWKIEEVNKLKELLLKNGQIVALVDMMEVPAVQLQEIIRDKIR-GTMTLKMSRNTLIERAIVEVAEETGNPEFA	82
RLA0_METVA	-----MIDAKSEHKIAPWKIEEVNKLKELLLKNSVIALVDMMEVPAVQLQEIIRDKIR-DQMTLKMSRNTLIKRAVEEVAEETGNPEFA	82
RLA0_METJA	-----METKVKAHVAPWKIEEVKTLKGLIKSKPVVAIVDMMDVPAPQLQEIIRDKIR-DKVKLRMSRNTLIERALKEAAEELNPNKLA	81
RLA0_PYRAB	-----MAHVAEWKKEVEELANLIKSPVIALVDVSSMPAYPLSQMRRRIRENGGLLRVSRNTLIE LAIKKAAQELGKPELE	77
RLA0_PYRHO	-----MAHVAEWKKEVEELAKLIKSPVIALVDVSSMPAYPLSQMRRRIRENGGLLRVSRNTLIE LAIKKAAQELGKPELE	77
RLA0_PYRFU	-----MAHVAEWKKEVEELANLIKSPVIALVDVSSMPAYPLSQMRRRIRENGGLLRVSRNTLIE LAIKKAAQELGKPELE	77
RLA0_PYRKO	-----MAHVAEWKKEVEELANLIKSPVIALVDVAGVPAYPLSKMRDKLR-GKALLVSRNTLIE LAIKRAAQELGQPELE	76
RLA0_HALMA	MSAESERKTETIPEWQEQEVDVAIVEMIESYESVGVVNIAGIPSRLOQDMRRDLHGT-AELRVSNTLLE RALDDVD-----DGLE	79
RLA0_HALVO	MSESEVRQTEVIPQWKRREVDLVDVIESYESVGVVGVAGIPSRLOQDMRRDLHGT-AELRVSNTLLE RALDDVD-----DGFV	79
RLA0_HALSA	MSAEERQRTTEEVPEWKRQEAELVDLLETYSVGVVNVVTGIPSKOLODMRRDLHGT-AELRVSNTLLE RALDDVD-----DGLD	79
RLA0_THEAC	-----MKEVSQKKKEVEELANLIKSPVIALVDVAGVPAYPLSKMRDKLR-GKALLVSRNTLIE LAIKRAAQELGQPELE	76
RLA0_THEVO	-----MRKINPKKKEIVSELAQDITKSKAVAVVDIKGVRTROMODIRAKNRDK-VKIKVVKKTLFLKALDSIND-----EKLT	72
RLA0_PICTO	-----MTEPAQWKIDFVKNLENEINSRKVAAIVSIKGLRNNFQKIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE	76
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.
- Os alinhamentos múltiplos podem ser usados para a construção de *perfis*, *matriz de score de score posicionais* (PSSMs) e *modelos ocultos de Markov* (HMMs)

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:

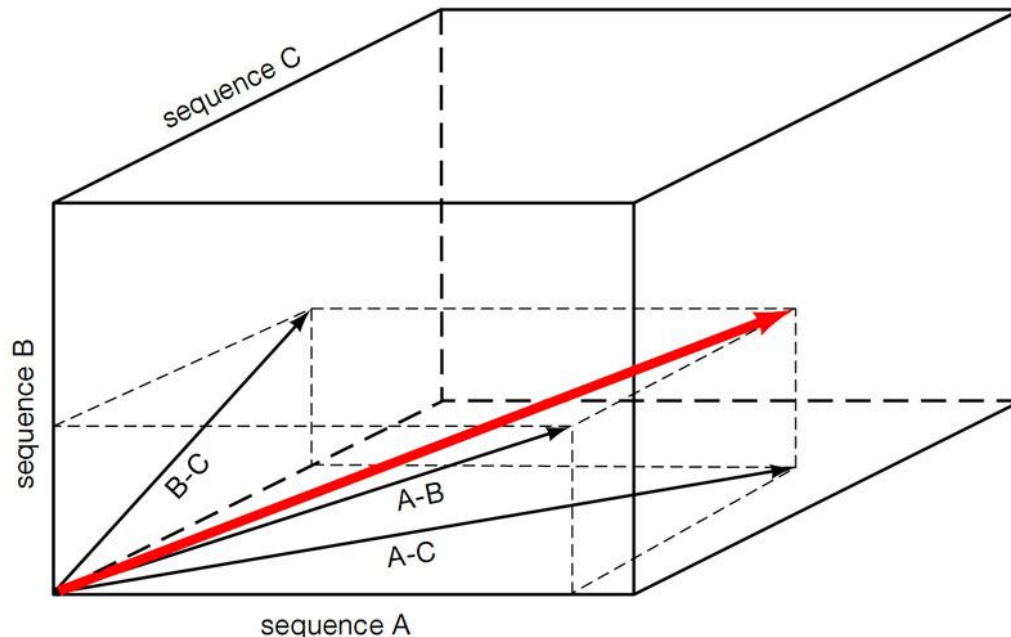
- **Extensão dos métodos ótimos para N sequências:** o algoritmo de N-W pode ser estendido para 3 ou mais sequências, mas exige o uso de matrizes multi-dimensionais e torna-se muito pesado computacionalmente (exp.: MSA)
- **Métodos progressivos (ou hierárquicos) :** baseiam-se na aplicação sucessiva 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 seqüências torna-se impraticável computacionalmente: o alinhamento óptimo é agora um caminho num cubo a N dimensões.

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

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



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

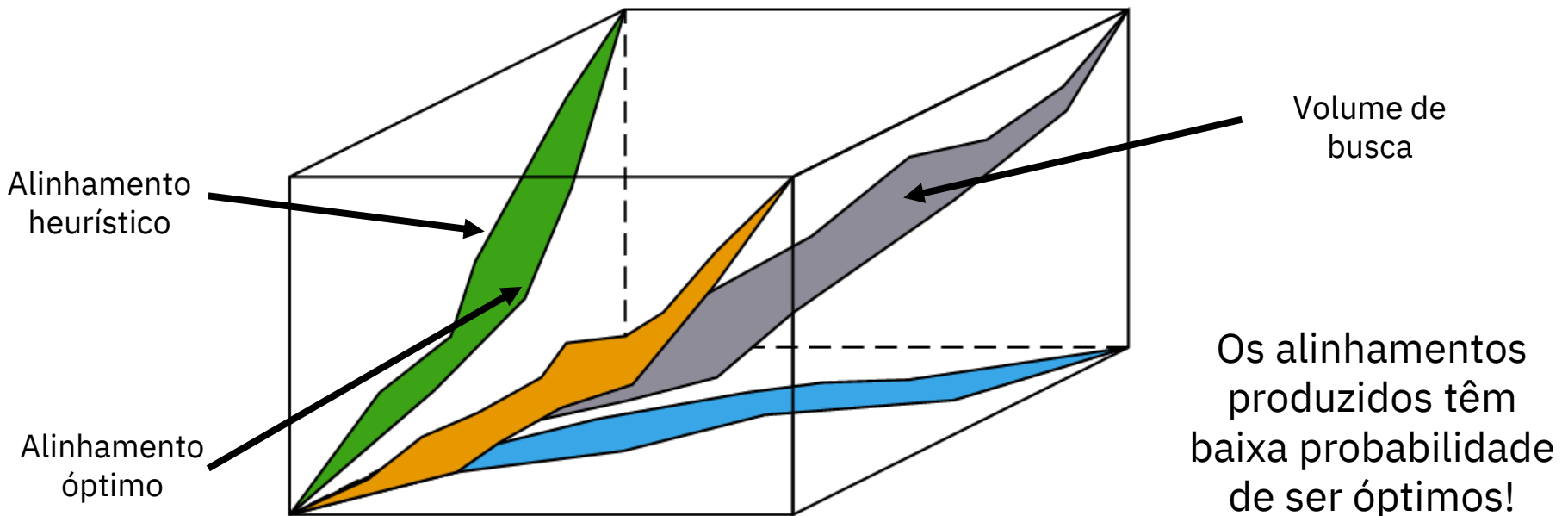
Algoritmo de Carrillo-Lipman-Gupta

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.

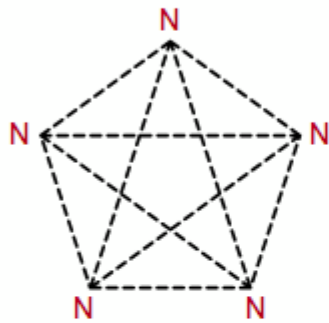
Implementado no programa **MSA** - demasiado pesado para ser usado com mais de 25-30 sequências com ~100 aminoácidos.

Não existem servidores de acesso livre para este programa.

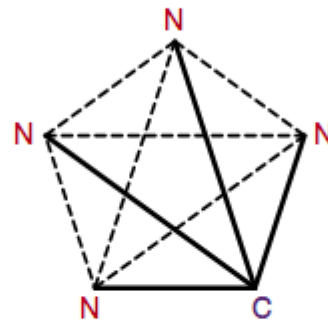


Cálculo do score num alinhamento múltiplo: o método SP (sum of pairs)

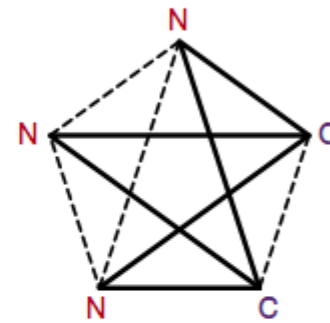
Sequence	Column A	Column B	Column C
1N.....N.....N.....
2N.....N.....N.....
3N.....N.....N.....
4N.....N.....C.....
5N.....C.....C.....



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 :

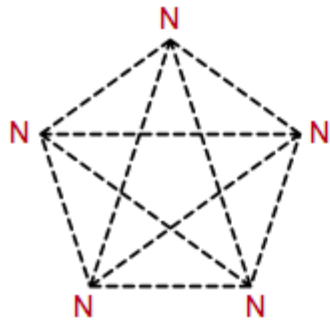
60

24

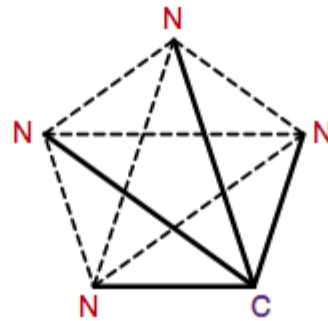
6

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

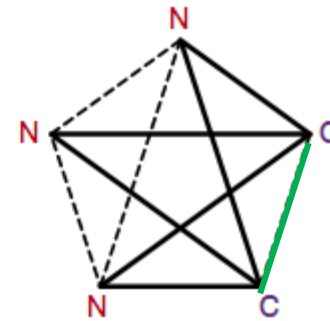
Sequence	Column A	Column B	Column C
1N.....N.....N.....
2N.....N.....N.....
3N.....N.....N.....
4N.....N.....C.....
5N.....C.....C.....



Column A



Column B

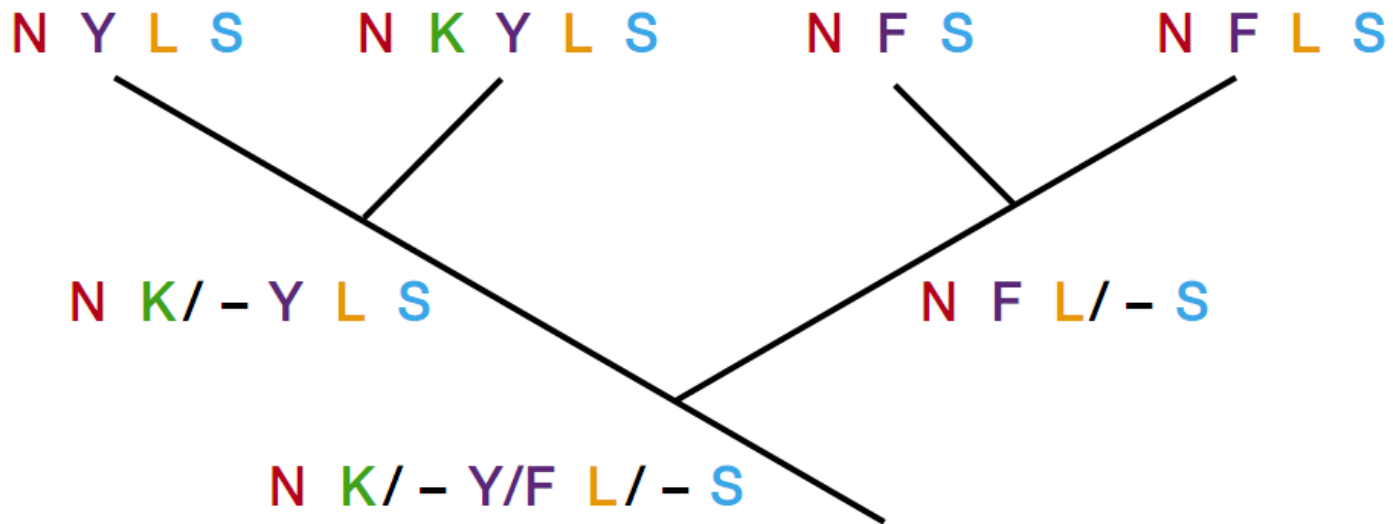


Column C

Pares (N,N)	10	6	3	$S_{Bl62}(N,N) = 6$
Pares (N,C)	0	4	6	$S_{B62}(N,C) = -3$
Pares (C,C)	0	0	1	$S_{Bl2}(C,C) = 9$
Score	$10*6+0*(-3)+0*9=0$	$6*6+4*(-3)+0*9=24$	$3*6+6*(-3)+1*9=9$	

Métodos de alinhamento progressivo

Os métodos de alinhamento progressivo usam o algoritmo de programação dinâmica para 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, 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, incorpora informação de vários métodos de alinhamento tanto locais como globais. 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

- Criado em 1994 por Thompson, Higgins and Gibson.
- Alinhamento de todos os possíveis pares de sequências com um método de programação dinâmica convencional
- Árvore de guia (*guide tree*) construída a partir das similaridades entre todos os pares calculadas no passo anterior é usada para conduzir a adição sucessiva de sequências ao alinhamento múltiplo
- As sequências mais próximas são alinhadas primeiro
- Os gaps nos alinhamentos de pares são conservados até ao fim do alinhamento
- Esquema de penalização de gaps ajustável
- Sequências são pesadas de acordo com seu grau de similaridade

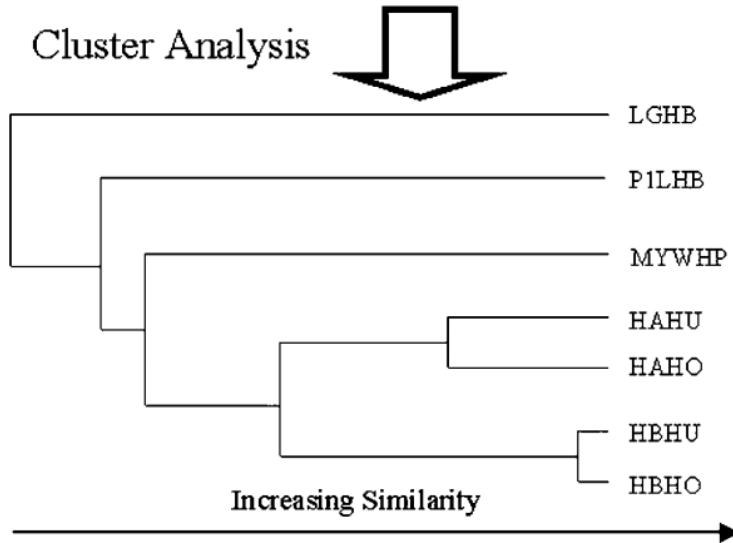
Vantagens do método: rápido, suporta um grande número de sequências ; facilmente configurável

Principal problema: os erros no alinhamento de pares propagam-se ao alinhamento (o alinhamento de cada par não tem em consideração a relação de cada sequência as demais sequências).

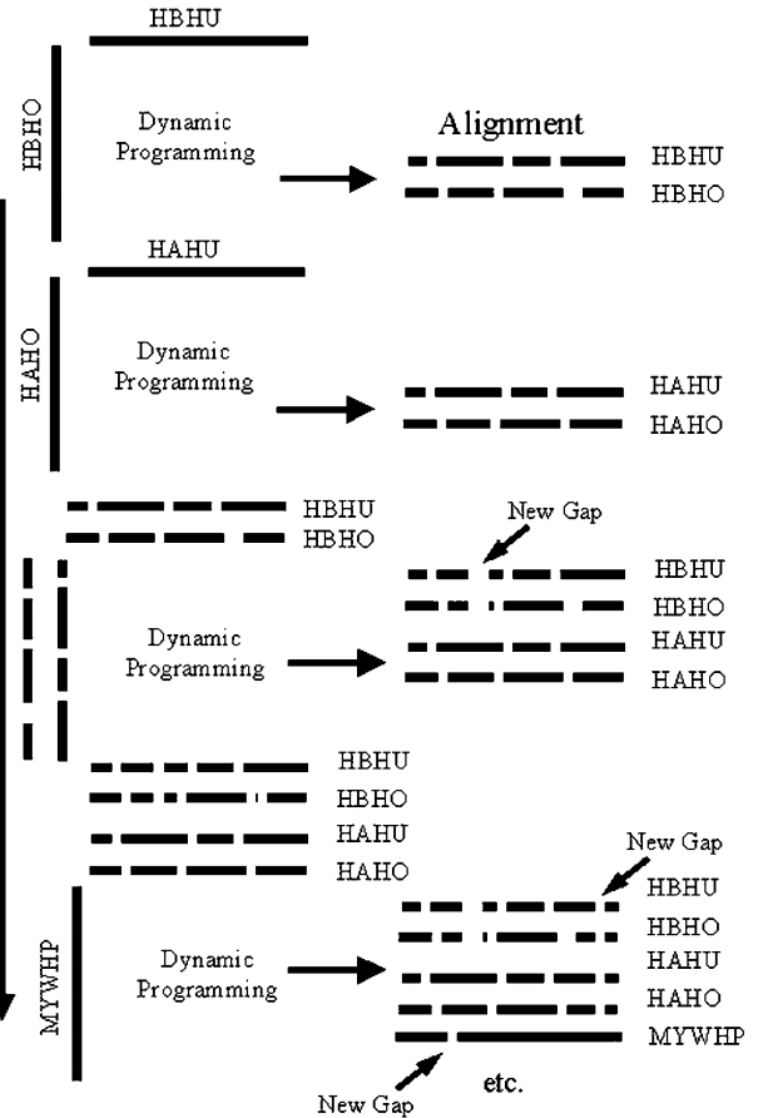
CLUSTALW

	HAHU	HBHU	HAHO	HBHO	MYWHP	P1LHB	LGHB
HAHU							
HBHU	21.1						
HAHO	32.9	19.7					
HBHO	20.7	39.0	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



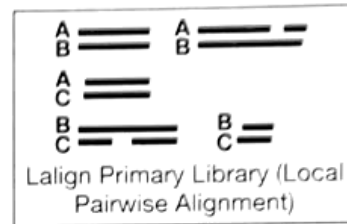
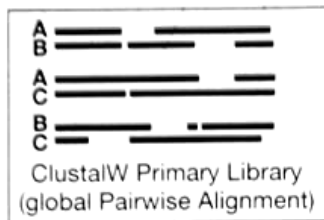
T-Coffee

T-Coffee – **T**ree-based **C**onsistency **O**bject **F**unction **F**or Alignment **E**valuation

- Criado por Cédric Notredame e colaboradores em 2000.
- Procura obviar a principal limitação dos métodos de alinhamento progressivos: propagação dos erros nos alinhamentos de pares para o alinhamento múltiplo final
- Combina a informação de alinhamentos locais e globais entre todos os pares de sequências para a geração de uma *livraria*.
- Os pares de resíduos da livraria são *pesados* de acordo com a fiabilidade do alinhamento de pares de onde provêm.
- A livraria de pares é *extendida* corrigindo os pesos de cada par de alinhamentos com a geração de todos os tripletos de sequências incluindo cada par.
- Os pesos são usados para gerar alinhamentos de pares pelo método de *programação dinâmica*. As similaridades obtidas permitem gerar uma árvore guia e alinhar progressivamente todas as sequências (tal como CLUSTALW)

Vantagens do método: mais rigoroso que CLUSTALW, melhor para sequências com regiões mais divergentes

Principal problema: a precisão diminui à medida que o número de sequências aumenta. Não é fiável para mais de ~100 sequências.



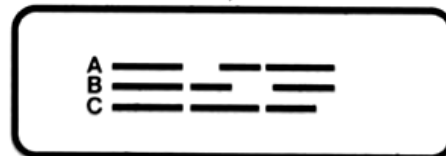
Pesagem

Biblioteca
primária

Extensão

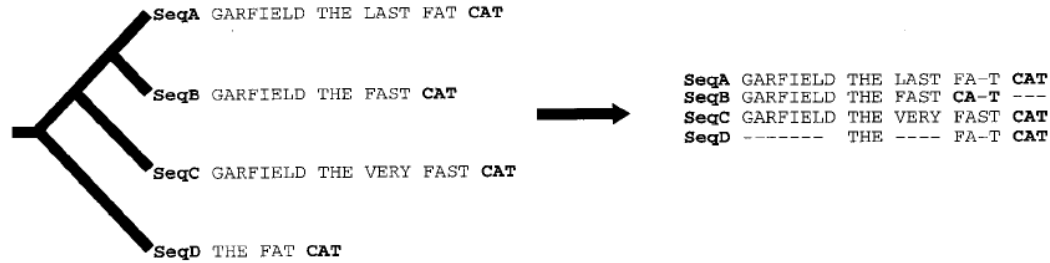
Biblioteca
extendida

Alinhamento
progressivo



T-Coffee

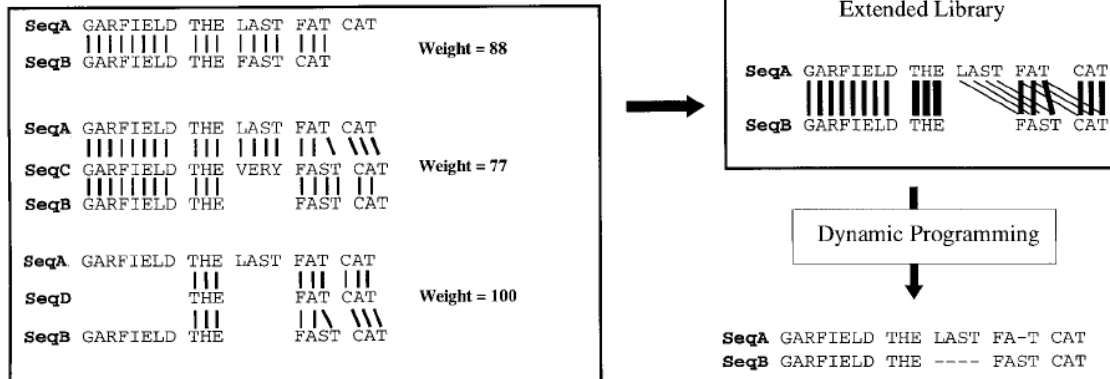
a) Regular Progressive Alignment Strategy



b) Primary Library

SeqA GARFIELD THE LAST FAT CAT SeqB GARFIELD THE FAST CAT ---	Prim. Weight = 88	SeqB GARFIELD THE ---- FAST CAT SeqC GARFIELD THE VERY FAST CAT	Prim Weight = 100
SeqA GARFIELD THE LAST FA-T CAT SeqC GARFIELD THE VERY FAST CAT	Prim. Weight = 77	SeqB GARFIELD THE FAST CAT SeqD ----- THE FA-T CAT	Prim. Weight = 100
SeqA GARFIELD THE LAST FAT CAT SeqD ----- THE ---- FAT CAT	Prim. Weight = 100	SeqC GARFIELD THE VERY FAST CAT SeqD ----- THE ---- FA-T CAT	Prim. Weight = 100

c) Extended Library for SeqA and SeqB



CLUSTAL OMEGA



- Publicado em 2011
- Optimizações no algoritmo reduzem a memória e tempo de computação necessárias possibilitando o alinhamento de um número muito maior de sequências (> 100 000)
- Utilização de *perfis* baseados em modelos de Markov (HMMs) permite um alinhamento mais rigoroso, particularmente nas regiões de fraca similaridade
- Para além de mais rápido, é muito mais preciso do que ClustalW
- Permite adicionar sequências a um alinhamento previamente calculado
- Permite usar um perfil HMM para uma determinada família de sequências como input adicional para o cálculo do alinhamento múltiplo, com aumento significativo da precisão

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).

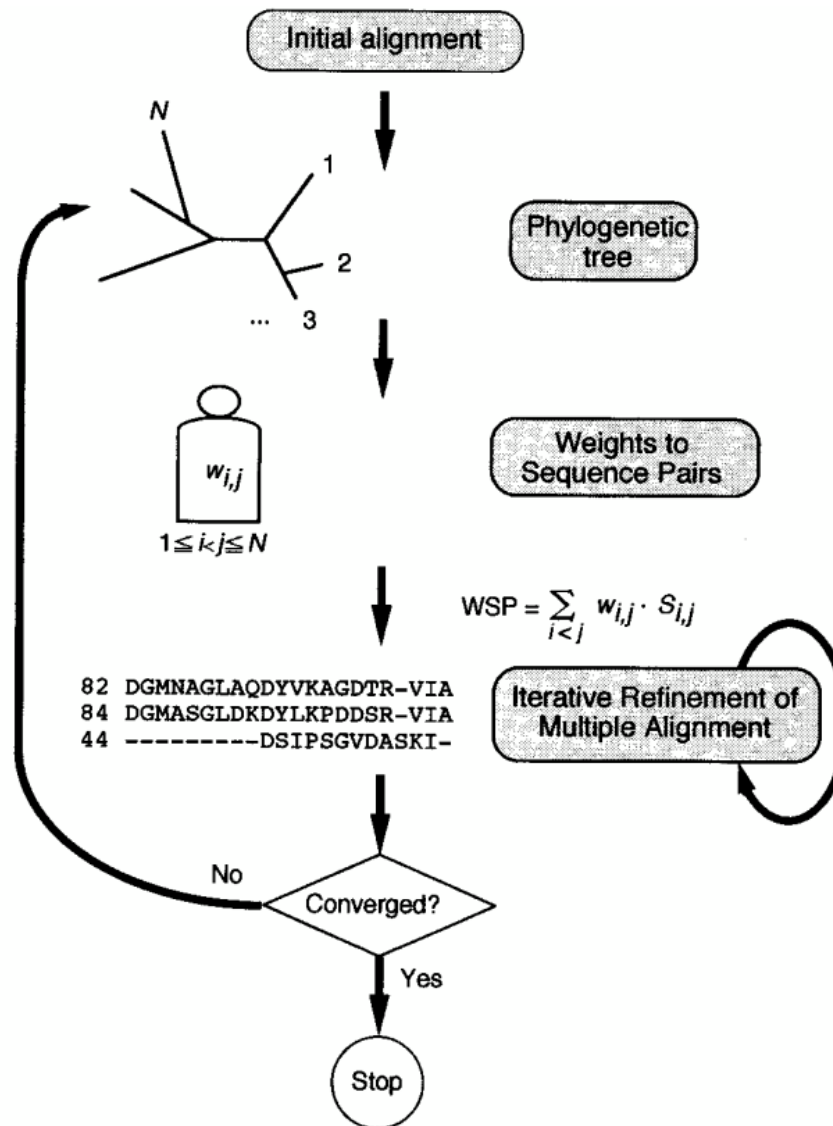
- **DIALIGN:** pesquisa de alinhamentos locais sem gaps em pares de sequências, pesados para o cálculo e otimização do alinhamento final.
<https://dialign.gobics.de/anchor/submission.php>

- **PRRP/PRRN:** refinamento iterativo de um alinhamento progressivo com construção de árvore e uso de pesos no alinhamento de pares.
<https://www.genome.jp/tools-bin/prrn>

- **SAGA:** método iterativo baseado num algoritmo genético. Não está disponível na forma de serviço “on-line”. É bastante pesado computacionalmente.

- <https://tcoffee.org/Projects/saga/index.html>

Alinhamento iterativo com PRRP/PRRN



Multiple Sequence Alignment

[Feedback](#)[Tools > Multiple Sequence Alignment](#)

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 (MSA) is generally the alignment of three or more biological sequences (protein or nucleic acid) of similar length. From the output, homology can be inferred and the evolutionary relationships between the sequences studied.

By contrast, Pairwise Sequence Alignment tools are used to identify regions of similarity that may indicate functional, structural and/or evolutionary relationships between two biological sequences.

Clustal Omega

New MSA tool that uses seeded guide trees and HMM profile-profile techniques to generate alignments. Suitable for medium-large alignments.

[Launch Clustal Omega](#)

Cons (EMBOSS)

EMBOSS Cons creates a consensus sequence from a protein or nucleotide multiple alignment.

[Launch EMBOSS Cons](#)

Kalign

Very fast MSA tool that concentrates on local regions. Suitable for large alignments.

[Launch Kalign](#)

MAFFT

MSA tool that uses Fast Fourier Transforms. Suitable for medium-large alignments.

[Launch MAFFT](#)

MUSCLE

Accurate MSA tool, especially good with proteins. Suitable for medium alignments.

[Launch MUSCLE](#)

MView

Transform a Sequence Similarity Search result into a Multiple Sequence Alignment or reformat a Multiple Sequence Alignment using the MView program.

[Launch MView](#)

T-Coffee

Consistency-based MSA tool that attempts to mitigate the pitfalls of progressive alignment methods. Suitable for small alignments.

[Launch T-Coffee](#)

WebPRANK

The EBI has a new phylogeny-aware multiple sequence alignment program which makes use of evolutionary information to help place insertions and deletions. Try it out at [WebPRANK](#).

The tools described on this page are provided using [Search and sequence analysis tools services from EMBL-EBI in 2022](#)

Please read the provided Help & Documentation and FAQs before seeking help from our support staff. If you have any feedback or encountered any issues please let us know via [EMBL-EBI Support](#). If you plan to use these services during a course please [contact us](#). Read our [Privacy Notice](#) if you are concerned with your privacy and how we handle personal information.



Services

Data resources and tools
Data submission
Support and feedback
Licensing

Research

Publications
Research groups
Postdocs and PhDs

Training

Live training
On-demand training
Support for trainers
Contact organisers

Industry

Members Area
Contact Industry team

About EMBL-EBI

Contact us
Events
Jobs
News
People and groups

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](#)

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

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

If you use this service, please consider citing the following publication: [Search and sequence analysis tools services from EMBL-EBI in 2022](#)

Please read the provided Help & Documentation and FAQs before seeking help from our support staff. If you have any feedback or encountered any issues please let us know via [EMBL-EBI Support](#). If you plan to use these services during a course please [contact us](#). Read our [Privacy Notice](#) if you are concerned with your privacy and how we handle personal information.

Clustal Omega

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
MSALLLALV GAAVAFPLED DDKIVGGYTC PEHSVPYQVS LNSGYHFCGG SLINDQWVVS
AAHCYKSRIQ VRLGEHNINV LEGDEQFINA AKIHKHPNYS SWTLNNDIML IKLSSPVKLN
ARVAPVALPS ACAPAGTQCL ISGWGNTLSN GVNNPDLLQC VDAPVLSQAD CEAAYPGEIT
SSMICVGFLE GGDSCQGDS GGPVVCNGQL QGIVSWGYGC ALPDNPGVYT KVCNFBVGIQ
DTIAAN
```

```
>sp|P07477|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](#)

Service Annou

The new Job Disp
webpages!

```
>sp|P35035|TRY1_ANOGA Trypsin-1 OS=Anopheles gambiae OX=7165 GN=TRYP1 PE=2 SV=3
MSNKIAILLAVLVAVVACAEQAQNRHRLVRPSPSFSRPRRYAVGQRIVGGFEIDVSDAP
YQVSLQYNKRHNCGGSVLSSKWVLTAAHCTAGASPSLTVRLGTSRHASGGTVVRVARVV
QHPKYDSSSIDFDYSLLELEDELTFSDSVQPVGLPKQDETVKDGTTTVSGWGNTQSAAE
SNAVLRAANVPTVNQKECNKAYSEFGGVTDRLMLCAGYQQGGKDACQGDSSGGPLVADGKLV
GVVSWGYGCAQAGYPGVYSRVAVVRDWWRENSGV
```

ack about the new

Multiple

Clustal Omega is a
or more sequences

```
>sp|P35031|TRY1_SALSA Trypsin-1 OS=Salmo salar OX=8030 PE=1 SV=1
MISLVFVLLIGAAFATEDDDKIVGGYECKAYSQTHQVSLNSGYHFCGGSLVNNWVWSAAH
CYKSRVEVRLGEHNIKVTEGSEQFISSSRVIRHPNYSSYNIDNDIMLIKLSKPATLNTYV
QPVALPTSCAPAGTMCVSGWGNTMSSTADSNKLQCLNIPILSYSDCNSYPGMITNAMF
CAGYLEGGKDSCQGDSSGGPVVVCNGELQGVVSWGYGCAEPGNPGVYAKVICFNDWLTSTMA
SY
```

alignments between **three**

Important note: Th

STEP 1 - Enter

```
>sp|P07477|TRY1_HUMAN Serine protease 1 OS=Homo sapiens OX=9606 GN=PRSS1 PE=1 SV=1
MNPILLILTFVAAALAAPFDDDDKIVGGYNCEENSVPYQVSLNSGYHFCGGSLINEQWVVS
AGHCYKSRIQVRLGEHNIIEVLEGNQFINAAKIIRHPQYDRKTLNNDIMLIKLSRAVIN
ARVSTISLPTAPPATGTCKLISGWGNTASSGADYPDELQCLDAPVLSQAKCEASYPGKIT
SNMFCVGFLEGGKDSCQGDSSGGPVVVCNGQLQGVVSWGDGCAQKNKPGVYTKVYNYVKWIK
NTIAANS
```

Enter or paste a se

PROTEIN

sequences in any

```
>sp|P00762|TRY1_RAT Serine protease 1 OS=Rattus norvegicus OX=10116 GN=Prss1 PE=1 SV=1
MSALLLILALVGA AVAFLEDDDKIVGGYTCEHSVPYQVSLNSGYHFCGGSLINDQWVVS
AAHCYKSRIQVRLGEHNIINVLGDEQFINAAKIIRHPNYSSWTLNNDIMLIKLSPPVKLN
ARVAPVALPSACAPAGTQCLISGWGNTLSNGVNNPDLLQCVDPVLSQADCEAAYPGEIT
SSMICVGFLEGGKDSCQGDSSGGPVVVCNGQLQGVVSWGYGCALPDNPGVYTKVCFVWGIQ
DTIAAN
```

>sp|P35031|TR

```
>sp|P35049|TRYP_FUSOX Trypsin OS=Fusarium oxysporum OX=5507 PE=1 SV=1
MVKFASVVALVAPLAAAAPQEIPNIVGGTSASAGDFPFIVSISRNGGPWCGGSLNANTV
LTAACHCVSGYAQSGFQIRAGSLRSTSGGITSSLSRVVHPSYSGNNDLAILKLSSTIPS
GGNIGYARLAASGSDPVAGSSATVAGWGATSEGGSSTPVNLKVTVPVIVSRATCRAQYGT
SAITNQMFCAVSSGGKDSCQGDSSGGPIVDSNTLIGAVSWGNGCARPNYSYGVYASVGL
RSFIDTYA
```

Or, upload a file:

STEP 2 - Set yo

OUTPUT FORMAT

nce | See more example inputs

T-Coffee

Input form | Web services | Help & Documentation | Bioinformatics Tools FAQ

Feedback

Tools > Multiple Sequence Alignment > T-Coffee

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

T-Coffee is a multiple sequence alignment program. Its main characteristic is that it will allow you to combine results obtained with several alignment methods.

Important note: This tool can align up to 500 sequences or a maximum file size of 1 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 an example sequence](#) | [Clear sequence](#) | [See more example inputs](#)

STEP 2 - Set your Parameters

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

MATRIX

None

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

If you use this service, please consider citing the following publication: [Search and sequence analysis tools services from EMBL-EBI in 2022](#)

Please read the provided [Help & Documentation](#) and [FAQs](#) before seeking help from our support staff. If you have any feedback or encountered any issues please let us know via [EMBL-EBI Support](#). If you plan to use these services during a course please [contact us](#). Read our [Privacy Notice](#) if you are concerned with your privacy and how we handle personal information.



Services

Data resources and tools
Data submission
Support and feedback
Licensing

Research

Publications
Research groups
Postdocs and PhDs

Training

Live training
On-demand training
Support for trainers
Contact organisers

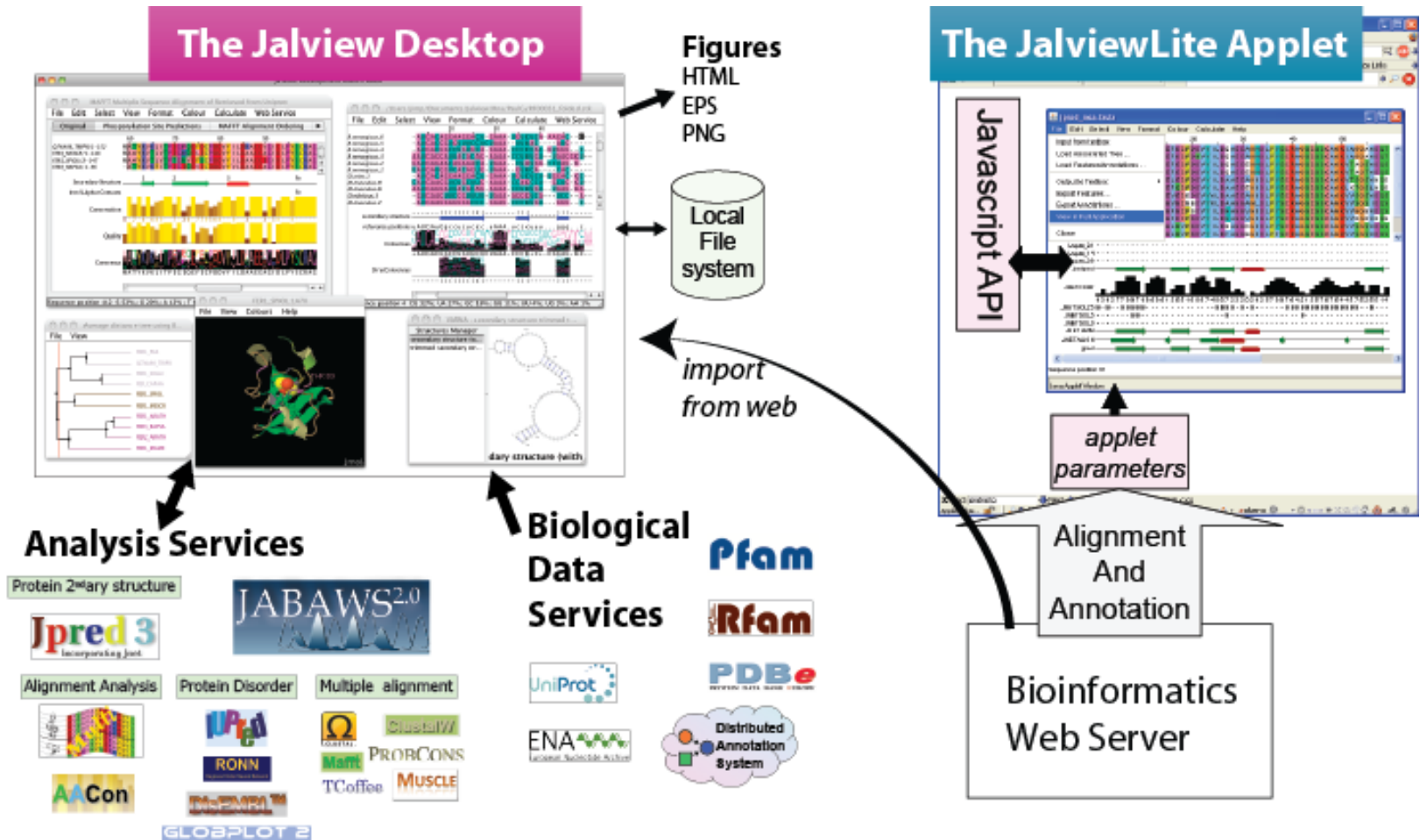
Industry

Members Area
Contact Industry team

About EMBL-EBI

Contact us
Events
Jobs
News
People and groups

Edição interactiva de alinhamentos com Jalview



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

Jalview desktop

The image displays the Jalview 2.11.3.0 desktop application interface. The main window, titled "MAFFT Multiple Sequence Alignment of Retrieved from Uniprot", shows a sequence alignment of various FER1 protein variants. The alignment is color-coded by amino acid type and includes a secondary structure diagram below it. The secondary structure diagram shows alpha-helices (green arrows) and beta-strands (red bars) across the sequence. Below the alignment are conservation and quality plots, and a consensus sequence.

On the right, a smaller window titled "Average distance tree using BLOSUM..." displays a phylogenetic tree of the sequences. The tree shows relationships between sequences such as FER1_PEA, Q7XA98_TRIPR, FER1_SOLLC, FER_CAPAA, FER1_SPIOL, FER1_MESCR, FER2_ARATH, FER3_RAPSA, FER1_ARATH, and FER1_MAIZE.

At the bottom right, a Jmol window titled "Jmol view for FER1_SPIOL:1A70" shows a 3D ribbon model of the protein structure. Key residues are highlighted: GLU 29, THR 29, and GLN 76. The protein is shown in a yellow and green color scheme.

JalviewJS web app

The screenshot displays the JalviewJS web application interface. At the top, the browser window shows the URL <https://www.ebi.ac.uk/Tools/services/rest/tcoffee/result/tcoffee-l20231119-202604-0519-71624238-p1m/aln-clustalw>. The main content area shows a sequence alignment of four proteins: *TRY1_RAT/1-246*, *TRY1_HUMAN/1-247*, *TRY1_SALSA/1-242*, and *TRY1_ANOGA/1-274*. The alignment is color-coded by amino acid type. Below the alignment are four plots: Conservation (yellow bars), Quality (brown bars), Consensus (black bars), and Occupancy (grey bars). The Conservation plot includes numerical values: 535-740-6-798776-044356-75723374777-84-5336-999468-98-84-4769-7-5264-754-6856585044-5-7442288-78-+7-8437347549. The Consensus plot shows the sequence: MS+L I + I L V + A L V + A A V + A E A Q A N Q R H R L V R P S P S F S P R P R + + D D K I V G G Y E C + + S V P Y Q V S L N S N G Y H F C G G S L + N E + W V V S A A H C Y K G + + + S R I Q V R L G E H N I + V + E G + E Q F I + A A + V I R H P N Y S S + T + N N D I M L I K L S S P A T L N A R V. The bottom of the image shows the Windows taskbar with the search bar and various application icons.

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