



Multiple Sequence Alignment (MSA) di sequenze SARS-CoV2

Le risorse

- Le sequenze di **SARS-Cov2** sono reperibili nella banca dati GenBank di **NCBI** (National Center for Biotechnology Information)

<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>

- I tool di **allineamento multiplo** sono reperibili al sito di **EMBL-EBI** (European Bioinformatics Institute, parte di European Molecular Biology Laboratory)

<https://www.ebi.ac.uk/Tools/msa/>

Le risorse

- Le sequenze di **SARS-Cov2** sono reperibili nella banca dati **GenBank** di **NCBI** (National Center for Biotechnology Information)

<https://www.ncbi.nlm.nih.gov/nuccore/sars-cov-2-seqs/>

GenBank è la storica banca di sequenze nucleotidiche di NCBI

- I tool di **allineamento multiplo** sono reperibili al sito di **EMBL-EBI** (European Bioinformatics Institute, parte di European Molecular Biology Laboratory)

<https://www.ebi.ac.uk/Tools/msa/>

SARS-CoV2 a NCBI (GenBank)

<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>

BLAST against Betacoronavirus sequences

GenBank	RefSeq	Gene Region	Collection Date	Locality
MN908947	NC_045512	complete	2019-12	China
LC522350		RdRP	2020-01-26	Philippines
LC523807		N	2020-01-06	Philippines
LC523808		N		
LC523809		N		
LC528232		complete		
LC528233		complete	2020-02-10	Japan
LC529905		complete	2020-01	Japan
LC534418		complete	2020-02-14	Japan
LC534419		complete	2020-03-09	Japan
LR757995		complete	2020-01-05	China: Wuhan
LR757996		complete	2020-01-01	China: Wuhan
LR757997		complete, gapped	2019-12-31	China: Wuhan
LR757998		complete	2019-12-26	China: Wuhan

Le sequenze con campo **Gene Region** uguale a "complete" sono le sequenze complete del virus, lunghe circa 29,900 basi

SARS-CoV2 a NCBI (GenBank)

<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>

BLAST against Betacoronavirus sequences

GenBank	RefSeq	Gene Region	Collection Date	Locality
MN908947	NC_045512	complete	2019-12	China
LC522350		RdRP	2020-01-26	Philippines
LC523807			1-06	Philippines
LC523808			1-26	Philippines
LC523809			1-23	Philippines
LC528232			2-10	Japan
LC528233			2-10	Japan
LC529905			1	Japan
LC534418			2-14	Japan
LC534419			3-09	Japan
LR757995		complete	2020-01-05	China: Wuhan
LR757996		complete	2020-01-01	China: Wuhan
LR757997		complete, gapped	2019-12-31	China: Wuhan
LR757998		complete	2019-12-26	China: Wuhan

C'è un'unica sequenza che ha un campo **RefSeq** non vuoto ed è la sequenza *reference* che è stata inserita nella banca dati **RefSeq** (banca di sequenze di riferimento). MN908947 è il suo ID nella banca GenBank, mentre NC_045512 è il suo ID nella banca RefSeq.

SARS-CoV2 a NCBI (GenBank)

<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>

BLAST against Betacoronavirus sequences

GenBank	RefSeq	Gene Region	Collection Date	Locality
MN908947	NC_045512	complete	2019-12	China
LC522350		RdRP	2020-01-26	Philippines
LC523807		N	2020-01-06	Philippines
LC523808		N	2020-01-26	Philippines
LC523809		N	2020-01-23	Philippines
LC528232		complete	2020-02-10	Japan
LC528233		complete	2020-02-10	Japan
LC529905			2020-01	Japan
LC534418			2020-02-14	Japan
LC534419			2020-03-09	Japan
LR757995		complete	2020-01-05	China: Wuhan
LR757996		complete	2020-01-01	China: Wuhan
LR757997		complete, gapped	2019-12-31	China: Wuhan
LR757998		complete	2019-12-26	China: Wuhan

Per ottenere una sequenza basta cliccare sul suo identificatore GenBank

SARS-CoV2 a NCBI (GenBank)

<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>

GenBank ▾

Send to: ▾

Change region shown

Severe acute respiratory syndrome coronavirus 2 SARS-CoV-2/Hu/DP/Kng/19-031 RNA, complete genome

GenBank: LC534418.1

[FASTA](#) [Graphics](#)

Go to: ▾

LOCUS LC534418 29878 bp RNA linear VRL 28-MAR-2020
DEFINITION Severe acute respiratory syndrome coronavirus 2 SARS-CoV-2/Hu/DP/Kng/19-031 RNA, complete genome.
ACCESSION LC534418
VERSION LC534418.1
KEYWORDS .
SOURCE Severe acute respiratory syndrome coronavirus 2 (SARS-CoV2)
ORGANISM [Severe acute respiratory syndrome coronavirus 2](#)
Viruses; Riboviria; Nidovirales; Cornidovirinae; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus.
REFERENCE 1
AUTHORS Hishiki,T., Suzuki,R., Sakuragi,J., Usui,K., Tanaka,Y., Kawai,J., Kogo,Y., Matsuki,Y., An,T., Hayashizaki,Y. and Takasaki,T.
TITLE SARS-CoV-2 isolation from COVID-19 patients and asymptomatic individual
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 29878)
AUTHORS Hishiki,T., Suzuki,R., Sakuragi,J., Usui,K., Tanaka,Y., Kawai,J., Kogo,Y., Matsuki,Y., An,T., Hayashizaki,Y. and Takasaki,T.
TITLE Direct Submission
JOURNAL Submitted (19-MAR-2020) Contact:Takayuki Hishiki Kanagawa Prefectural Institute of Public Health; 1-3-1 Shimomachiya, Chigasaki, Kanagawa 253-0087, Japan

Send to: ▾

Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Si apre la pagina con la sequenza nel tipico formato GenBank che riporta una serie di informazioni (piuttosto dettagliate) sulla sequenza.

Taxonomy

PubMed (Weighted)

RefSeq Genome for Species

RefSeq Genome Sequences

SARS-CoV-2 sequences.

SARS-CoV2 a NCBI (GenBank)

<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>

GenBank ▾

Send to: ▾

Change region shown

Severe acute respiratory syndrome coronavirus 2 SARS-CoV-2/Hu/DP/Kng/19-031

RNA, complete genome

GenBank: LC534418

FASTA

Graph

Go to: ▾

LOCUS LC534418 29878 bp RNA linear VRL 28-MAR-2020

DEFINITION Severe acute respiratory syndrome coronavirus 2
SARS-CoV-2/Hu/DP/Kng/19-031 RNA, complete genome.

ACCESSION LC534418

VERSION LC534418.1

KEYWORDS .

SOURCE Severe acute respiratory syndrome coronavirus 2 (SARS-CoV2)

ORGANISM [Severe acute respiratory syndrome coronavirus 2](#)

Viruses; Riboviria; Nidovirales; Coronidovirinae; Coronaviridae;

Orthocoronavirinae; Betacoronavirus; Sarbecovirus.

REFERENCE 1

AUTHORS Hishiki,T., Suzuki,R., Sakuragi,J., Usui,K., Tanaka,Y., Kawai,J.,
Kogo,Y., Matsuki,Y., An,T., Hayashizaki,Y. and Takasaki,T.

TITLE SARS-CoV-2 isolation from COVID-19 patients and asymptomatic
individual

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 29878)

AUTHORS Hishiki,T., Suzuki,R., Sakuragi,J., Usui,K., Tanaka,Y., Kawai,J.,
Kogo,Y., Matsuki,Y., An,T., Hayashizaki,Y. and Takasaki,T.

TITLE Direct Submission

JOURNAL Submitted (19-MAR-2020) Contact:Takayuki Hishiki Kanagawa
Prefectural Institute of Public Health; 1-3-1 Shimomachiya,
Chigasaki, Kanagawa 253-0087, Japan

Cliccare su "FASTA" per ottenere la
sequenza in formato FASTA

Send to: ▾

Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

NCBI Virus

Retrieve, view, and download SARS-CoV-2
coronavirus genomic and protein sequences.

Related information

Protein

Taxonomy

PubMed (Weighted)

RefSeq Genome for Species

RefSeq Genome Sequences

SARS-CoV2 a NCBI (GenBank)

<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>

Severe acute respiratory syndrome coronavirus 2 SARS-CoV-2/Hu/DP/Kng/19-031 RNA, complete genome

GenBank: LC534418.1

[GenBank](#) [Graphics](#)

>LC534418.1 Severe acute respiratory syndrome coronavirus 2 SARS-CoV-2/Hu/DP/Kng /19-031 RNA, complete genome

```
AAGGTTTATACTTCCCGTAACAAACCAACCAACTTGCATCTCTGTAGATCTGTTCTCAAACGAA  
CTTTAAAATCTGTGGCTGTCACTCGCTGCATGCTTAGTGCACTCACGCACTATAATTAAACTAAT  
TACTGCTTGACAGGACACGAGTAACCTGCTATCTCTGCAGGCTGTTACGGTTCTGGCTGTTGC  
AGCGCATCATCAGCACATCTAGGTTCTGCCGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCCTG  
GTTCAACGAGAAAACACACGTCAACTCAGTTGCTGTTACAGGTTGCGACGTCCTGTCAGTGG  
CTTGGAGACTCCGTGGAGGAGGTCTTATCAGAGGCACGTCAACATCTAAAGATGGACTTGTGGCTTA  
GTAGAAGTTGAAAAAGCGTTTGCCTCAACTGTAACAGCCCTATGTGTTCATCAAACGTCGGATGCTC  
GAACGTGACCTCATGTCATGTTATGTTGAGCTGGTAGCAGAAACTCGAACGGCATTCACTACGGTCGTAG  
TGTTGAGACACTTGGTGTCTTGTGCCCTCATGTCGGCAAATACCAAGTGGCTTACCGCAAGGTTCTTCTT  
CGTAAGAACGGTAATAAGGAGCTGGCCATAGTTACGGGCCGATCTAAAGTCATTGACTTAGGGC  
ACGAGCTGGCACTGATCCTTATGAAGATTTCAAGAAAACCTGGAACACTAAACATAGCAGTGGTGTAC  
CCGTGAACCTCATGGTGTGAGCTTAACGGAGGGCATACTCGCTATGCGATAACAACCTTGTGGCCCT  
GATGGCTACCCCTCTGAGTGCATTAAGACCTTCTAGCACGTGCTGGTAAAGCTTCACTGCACCTTGTCCG  
AACAACTGGACTTTATTGACACTAACAGGGGTGTATACTGCTGCCGTGAACATGAGCATGAAATTGCTT  
GTACACGGAACGTTCTGAAAAGAGCTATGAATTGCGACACCTTTGAAATTAAATTGGCAAAGAAATT  
GACACCTCAATGGGAATGTCCAATTGGTATTTCCCTAAATTCCATAATCAAGACTATTCAACCAA  
GGGTTGAAAAGAAAAGCTTGTGGCTTATGGTAGAATTGATCTGTCTATCCAGTGGTCACCAAA  
TGAATGCAACCAAATGTGCCCTTCACACTCATGAAGTGTGATCATTGTTGAAACTTCATGGCAGACG  
GGCAGTTGTTAAAGCCACTTGCGAATTGTTGCACTGAGAATTGACTAAAGAAGGTGCCACTACTT  
GTGGTTACTTACCCAAAATGCTGTTAAAATTATTGTCAGCATGTCAACATTGAGAAGTAGGACC  
TGACCATAGTCTGCCAATACCCATAATGAATCTGGCTGAAACACATTCTCGTAAGGGTGGTCGCACT  
ATTGCCCTTGGAGGCTGTGTTCTTATGTTGGTGCCATAACAAAGTGTGCCATTGGGTTCCACGTG  
CTAGCGCTAACATAGGGTGTAAACCATACAGGTGGTAGAAGGGTCCGAAGGTCTTAATGACAACCT  
TCTTGAATAACTCCAAAAGAGAAAAGTCAACATCAATATTGTTGGTAGCTTAAACTTAATGAAGAGATC  
GCCATTATTTGGCATTTCTGCCACAAGTGTGAAAGCTTGTGAAAGTGTGAAAGTTGGATTATA
```

[Customize view](#)

[Analyze this sequence](#)

[Run BLAST](#)

[Pick Primers](#)

[Highlight Sequence Features](#)

[Find in this Sequence](#)

SARS-CoV-2 sequences.

Si apre la pagina con la sequenza in formato FASTA.

[RefSeq Genome for Species](#)

[RefSeq Genome Sequences](#)

[LinkOut to external resources](#)

SARS-CoV2 a NCBI (GenBank)

<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>

Severe acute respiratory syndrome coronavirus 2 SARS-CoV-2/Hu/DP/Kng/19-031 RNA, complete genome

GenBank: LC534418.1

[GenBank](#) [Graphics](#)

>LC534418.1 Severe acute respiratory syndrome coronavirus 2 SARS-CoV-2/Hu/DP/Kng /19-031 RNA, complete genome

AAGGTTTATACTTCCCGAGTAACAAACCAACCAACTTGCATCTCTGTAGATCTGTTCTCTAAACGAA
CTTAAATCTGTGGCTGTCACTGGCTGCATGCTAGTGCACTCACGCACTATAATTAAACTAAT
TACTGCTTGACAGGACACGAGTAACCTGCTATCTCTGCAGGCTGTTACGGTTCTGCTGTTGC
AGCGATCATCAGCACATCTAGGTTCTGCCGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCCTG
GTTCAACGAGAAAACACACGTCAACTCAGTTGCTGTTACAGGTTCCGACGTCCTCGTACGTTG
CTTGGAGACTCCGTGGAGGAGGTCTTATCAGAGGCACGTCAACATCTAAAC
GTAGAAGTTGAAAAAGCGTTTGGCTCAACTTGAACAGCCCTATGTTCA
GAAC TGCAACCTCATGTCATGTTATGTTGAGCTGGTAGCAGAACCTGAAGG
TGTTGAGACACTTGGTGTCCCTGCTCATGTTGGCGAAATACCAAGTGGCTT
CGTAAGAACGGTAATAAGGAGCTGGCCATAGTTACGGGCCGATCTAA
ACGAGCTGGCACTGATCCTTATGAAGATTTCAAGAAAATCTGAAACACTAA
CCGTGAACACTCATGGTGGCTTAAACGGAGGGGCATACACTCGCTATGTCGAT
GATGGCTACCCCTCTGAGTCATTAAAGACCTTCTAGCACGTGCTGGTAAAGCTTCTGCACTTTGTCCG

Per scaricare le sequenze è però più
comodo cliccare qui

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

NCBI Virus

Retrieve, view, and download SARS-CoV-2
coronavirus genomic and protein sequences.

Related information

Protein

Taxonomy

PubMed (Weighted)

RefSeq Genome for Species

RefSeq Genome Sequences

LinkOut to external resources

SARS-CoV2 a NCBI (GenBank)

<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>

i Severe acute respiratory syndrome coronavirus 2 data hub
Search, retrieve, and analyze SARS-CoV-2 GenBank data.

Attenzione: non c'è in questo caso un campo Gene Region, e quindi le sequenze complete sono quelle che hanno lunghezza di circa 29,900 basi

Med Download Align Build Phylogenetic Tree

Select Columns

<input type="checkbox"/> Accession	Release Date	Species	Length	Geo Location	H
MT050414	2020-02-12	Severe acute respiratory s...	562	Australia	H
MT050415	2020-02-12	Severe acute respiratory s...	562	Australia	H
MT050417	2020-02-12	Severe acute respiratory s...	562	Australia	H
MT050416	2020-02-12	Severe acute respiratory s...	562	Australia	H
MT111896	2020-02-26	Severe acute respiratory s...	569	Australia: Queensland	H
MT111895	2020-02-26	Severe acute respiratory s...	770	Australia: Queensland	H
MT007544	2020-01-31	Severe acute respiratory s...	29893	Australia: Victoria	H

Feedback

Virus +
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), taxid:2697049 X
Accession +
Sequence Length +
Sequence Type +
Nucleotide Completeness +
Provirus +

Expand Tab

SARS-CoV2 a NCBI (GenBank)

<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>

i Severe acute respiratory syndrome coronavirus 2 data hub
Search, retrieve, and analyze SARS-CoV-2 GenBank data.

Selected Results: 0

PubMed Download Align Build Phylogenetic Tree

Nucleotide (932) Protein (6,901) Select Columns

Accession MT050414

MT050414 2020-02-12 Severe acute respiratory s... 562 Australia

MT050417 2020-02-12 Severe acute respiratory s... 562 Australia

MT050416 2020-02-12 Severe acute respiratory s... 562 Australia

MT111896 2020-02-26 Severe acute respiratory s... 569 Australia: Queensland

MT111895 2020-02-26 Severe acute respiratory s... 770 Australia: Queensland

MT007544 2020-01-31 Severe acute respiratory s... 29893 Australia: Victoria

Geo Location

Australia

Selezionare le sequenze da scaricare e cliccare su "Download"

Feedback

Refine Results Reset

Virus +

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), taxid:2697049

Accession +

Sequence Length +

Sequence Type +

Nucleotide Completeness +

Provirus +

Expand Table

SARS-CoV2 a NCBI (GenBank)

<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>

Severe acute respiratory syndrome coronavirus 2 data
Search, retrieve, and analyze SARS-CoV-2 GenBank data.

Selected Results: 1

PubMed Download Align Build Phylogenetic Tree

Refine Results Reset

Virus +

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), taxid:2697049 X

Accession +

Sequence Length +

Sequence Type +

Nucleotide Completeness +

Provirus +

Geographic Region +

Download Results X

Step 1 of 3: Select Data Type

Sequence data (FASTA Format) Accession List Current table view result

Nucleotide Nucleotide CSV format

Coding Region Protein XML format

Protein

Next

MT007544 2020-01-31 Severe acute respiratory s... 29893 Australia: Victoria Feedback

Selezionare "Nucleotide" sotto Sequence data (FASTA format)

SARS-CoV2 a NCBI (GenBank)

<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>

i Severe acute respiratory syndrome coronavirus 2 data
Search, retrieve, and analyze SARS-CoV-2 GenBank data.

Selected Results: 1

Refine Results Reset

Virus

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), taxid:2697049

Accession

Sequence Length

Sequence Type

Nucleotide Completeness

Provirus

Geographic Region

PubMed Download Align Build Phylogenetic Tree

Select Columns

Geo Location

Australia

Australia

Australia

Australia

Back Next

MT050410 2020-02-12 Severe acute respiratory s... 562 Australia

MT111896 2020-02-26 Severe acute respiratory s... 569 Australia: Queensland

MT111895 2020-02-26 Severe acute respiratory s... 770 Australia: Queensland

MT007544 2020-01-31 Severe acute respiratory s... 29893 Australia: Victoria

Feedback

Scaricare solo i record selezionati

Download Results

Step 2 of 3: Select Records

Download Selected Records

Download All Records

SARS-CoV2 a NCBI (GenBank)

<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>

i Severe acute respiratory syndrome coronavirus 2 data
Search, retrieve, and analyze SARS-CoV-2 GenBank data.

Refine Results Reset

Virus

- Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), taxid:2697049

Accession

Sequence Length

Sequence Type

Nucleotide Completeness

Provirus

Geographic Region

Selected Results: 1

PubMed Download Align Build Phylogenetic Tree

Select Columns

Step 3 of 3: Select FASTA definition line

Use default : Accession GenBank Title

Build custom

Back Download

MT050410 2020-02-12 Severe acute respiratory s... 562 Australia

MT111896 2020-02-26 Severe acute respiratory s... 569 Australia: Queensland

MT111895 2020-02-26 Severe acute respiratory s... 770 Australia: Queensland

MT007544 2020-01-31 Severe acute respiratory s... 29893 Australia: Victoria

Feedback

Selezione "Build custom" per personalizzare l'header FASTA

SARS-CoV2 a NCBI (GenBank)

<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>

i Severe acute respiratory syndrome coronavirus 2 data
Search, retrieve, and analyze SARS-CoV-2 GenBank data.

Refine Results Reset

Virus +
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), taxid:2697049 X

Accession +

Sequence Length +

Sequence Type +

Nucleotide Completeness +

Provirus +

Geographic Region +

Download Results

Step 3 of 3: Select FASTA definition line

Use default : Accession GenBank Title

Build custom : Accession GenBank Title Geo Location Species Nucleotide Completeness
Host

Add ➤ Remove

Accession
GenBank Title
Geo Location
Species
Nucleotide Completeness

Back **Download**

Aggiungere/Rimuovere i campi che si vogliono inserire nell'header, e poi scaricare le sequenze

Build Phylogenetic Tree	
Select Columns	
Geo Location	H
australia	H
australia: Queensland	H
australia: Queensland	H
australia: Victoria	H

fppt.com

MSA Tools EMBL-EBI

<https://www.ebi.ac.uk/Tools/msa/>

EMBL-EBI mette a disposizione una serie di tools di allineamento multiplo.



Services

Research

Training

Industry

About us



EMBL-EBI Hinxton ▾

Multiple Sequence Alignment

Feedback

Share

Tools > Multiple Sequence Alignment

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](#)

EMBOSS Cons

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) and [Terms of Use](#).

[I agree, dismiss this banner](#)

MSA Tools EMBL-EBI

<https://www.ebi.ac.uk/Tools/msa/>

EMBL-EBI mette a disposizione una serie di tools di allineamento multiplo.

- Clustal Omega
- EMBOSS Cons
- Kalign
- MAFFT
- MUSCLE
- MView
- T-Coffee
- WebPRANK

MSA Tools EMBL-EBI

Ogni tool permette di trovare l'allineamento multiplo di un set di sequenze (DNA o proteine), e produce:

- l'output in diversi formati, tra cui quello che mette in evidenza la matrice di allineamento
- un albero filogenetico
- un file **Jalview** (estensione **jvl**) per visualizzare l'allineamento ed evidenziare le sue features tramite colorazione

Il tool di visualizzazione **Jalview** è scaricabile all'indirizzo <http://www.jalview.org/getdown/release/>

Esempio con Clustal Omega

Multiple Sequence Alignment

[Feedback](#)[Share](#)

Tools > Multiple Sequence Alignment

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](#)

EMBOSS Cons

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) and [Terms of Use](#).

[I agree, dismiss this banner](#)

Esempio con Clustal Omega

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

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

DNA

Selezionare il tipo delle sequenze

sequences in any supported format:

Or, upload a file: Nessun f

Caricare il file in formato FASTA che contiene le sequenze da allineare

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

Esempio con Clustal Omega

Input form Web services Help & Documentation Bioinformatics Tools FAQ Feedback Share

Or, upload a file: Nessun file selezionato.

Use a example sequence | Clear sequence | See more example inputs

STEP 2 - Set your parameters

OUTPUT FORMAT

ClustalW with character counts

The default settings will fulfill the needs of most users.

More options... (Click here, if you want to change the settings)

STEP 3 - Submit your job

Be notified by email (Tick this box if you want to receive an email when your job is available)

Selezionare il tipo di output.

“ClustalW with characters counts” è il formato testuale che mette in evidenza l'allineamento.

NB: scegliere il formato FASTA se si vuole visualizzare l'allineamento con il tool Jalview

Sottomettere il job

Esempio con Clustal Omega

Clustal Omega

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

Tools > Multiple Sequence Alignment > Clustal Omega

Results for job clustalo-l20200415-223239-0349-61683497-p2m

[Alignments](#)[Result Summary](#)[Download Alignment File](#)

CLUSTAL O(1.2.4) multiple sequence alignment

Per scaricare l'output
dell'allineamento

[Session Details](#)

MT077125	-----	ATCT	4
MT066156	ATTAAGGTTTACCTCCAGGTAAACAAACCAACTTCGATCTTGATGATCT	60	****
<hr/>			
MT077125	GTTCTCTAACGAACCTTAAATCTGTGTTGGCTGTCACTCGGCTGCATGCTTAGTGCAC	64	
MT066156	GTTCCTAACGAACCTTAAATCTGTGTTGGCTGTCACTCGGCTGCATGCTTAGTGCAC	120	*****
<hr/>			
MT077125	CACGCAGTATAATTAAACTAAATTACTGTCGTTGACAGGACACGAGTAACTCGTCTATC	124	
MT066156	CACGCAGTATAATTAAACTAAATTACTGTCGTTGACAGGACACGAGTAACTCGTCTATC	180	*****
<hr/>			
MT077125	TTCTGCAGGCTGCTTACGGTTCGTCCGTGTTGCAGCCGATCATCAGCACATCTAGTTT	184	
MT066156	TTCTGCAGGCTGCTTACGGTTCGTCCGTGTTGCAGCCGATCATCAGCACATCTAGTTT	240	*****

Esempio con Clustal Omega

Clustal Omega

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

Tools > Multiple Sequence Alignment > Clustal Omega

Results for job clustalo-I20200415-223239-0349-61683497-p2m

[Alignments](#) [Result Summary](#) [Phylogenetic Tree](#) [Results Viewers](#) [Submission Details](#)

[Download Alignment File](#)

CLUSTAL O(1.2.4) multiple sequence

MT077125
MT066156

ATTAAAGGTTTACCTTCC

MT077125
MT066156

GTTCTCTAACGAACCTTAA
GTTCCTAACGAACCTTAA

MT077125
MT066156

CACGCAGTATAATTAAACTAATTACTGCGTGACAGGACACGAGTAACCTCGTCTATC
CACGCAGTATAATTAAACTAATTACTGCGTGACAGGACACGAGTAACCTCGTCTATC

124
180

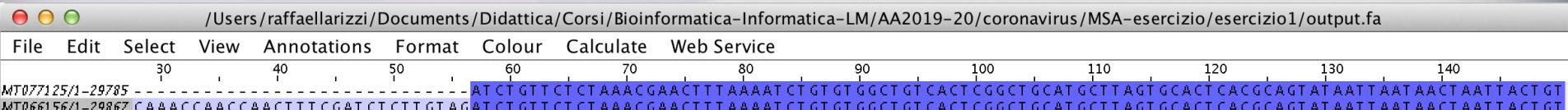
MT077125
MT066156

TTCTGCAGGCTGCTTACGGTTCGTCCGTGTTGCAGCCGATCATCAGCACATCTAGTTT
TTCTGCAGGCTGCTTACGGTTCGTCCGTGTTGCAGCCGATCATCAGCACATCTAGTTT

184
240

Per scaricare il file da
visualizzare con Jalview

Visualizzazione con Jalview



Consensus

CAAACCCAACCAACTTTCGATCTCTTGTAGATCTGTTCTCTAACCGAACCTTAAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCACGCAGTATAATTAAATAACTAATTACTGT

Occupancy

Progetto

- Formare gruppi di al più tre persone
- Scaricare da NCBI la sequenza RefSeq (*reference sequence*)
- Scaricare da NCBI un set di sequenze che si ritiene interessante allineare
- Scegliere da EMBL-EBI i tool di allineamento che si ritiene interessante usare
- Allineare le sequenze scaricate (*reference compresa*) con i tool scelti
- Per ognuno dei tool scelti, produrre in output le variazioni delle sequenze scaricate rispetto al *reference*, in un formato progettato *ad-hoc* e descritto in apposita documentazione
- Produrre un report conclusivo su quanto fatto e sui risultati ottenuti

NOTA

Produrre in output le variazioni significa elencare (in un formato definito opportunamente) tutte le variazioni delle sequenze allineate rispetto alla sequenza *reference*, riportate sulla sequenza *reference*.

Ad esempio, nel seguente allineamento:

reference	gataggat--ggg-catacgataggccgtgatgcc
seq1	gatatgatttaggcatacgtataggccgt---gcc
seq2	gataggatt-gggccatacga-aggcccgtatgcc
seq3	gataggatttaggcatacgtataggccgtatgcc

si notano le seguenti variazioni:

- in posizione=5, sostituzione **g** → **t** in seq1
- dopo la posizione=8, inserimento di **ta** in seq1 e seq3
- dopo la posizione=8, inserimento di **t** in seq2

NOTA

...

- dopo la posizione=11, inserimento di **c** in seq1 seq2 e seq3
- in posizione=19, cancellazione di **t** in seq2
- in posizione=28, cancellazione di **gat** in seq1

Il formato è da definire in modo da codificare queste variazioni, ed eventualmente altre più complicate che si possono verificare.