



# MSA using EMBL-EBI Clustal omega

# AIM of the experiment-



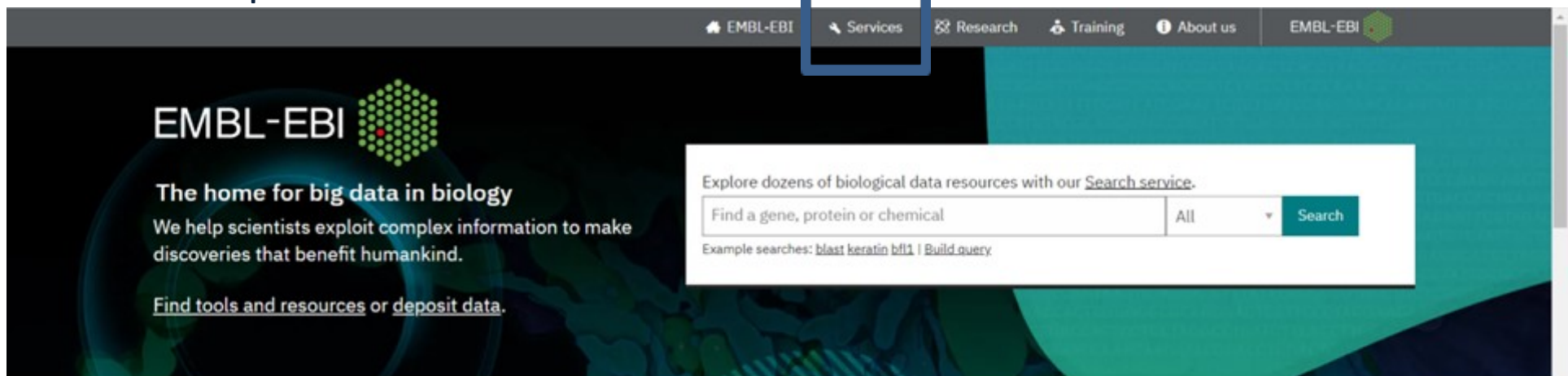
- To generate alignments between **three or more** sequences (multiple sequence alignment-MSA) by using clustal omega tool of EMBL-EBI.

## Principle-

- The aim of EMBL's research is to achieve a fundamental understanding of biological processes. We need to study and understand biological phenomena and systems in their full complexity.
- EMBL has developed an integrative, interdisciplinary structure that is ideally suited to tackling this challenge.
- Clustal Omega is a multiple sequence alignment program for aligning **three or more** sequences together in a computationally efficient and accurate manner.
- It produces biologically meaningful multiple sequence alignments of divergent sequences.
- Evolutionary relationships can be seen via viewing Cladograms or Phylograms.

# Procedure

- Google EMBL-EBI
- EMBL-EBI home page
- In the top task bar click on services



The screenshot shows the EMBL-EBI website home page. The top navigation bar includes links for EMBL-EBI, Services, Research, Training, and About us. A blue arrow points to the 'Services' link, which is highlighted with a blue box. Below the navigation bar, the main content area features the EMBL-EBI logo and the text 'The home for big data in biology'. A search bar is visible with the text 'Find a gene, protein or chemical' and a 'Search' button. Below the search bar, there are example searches: 'blast keratin bfl1 | Build query'.

## EMBL-EBI response to COVID-19

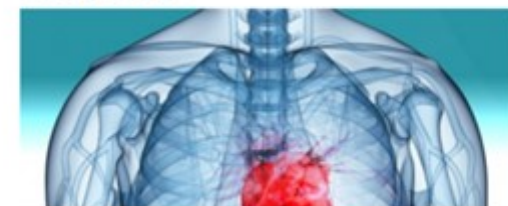
To help protect staff and visitors from the coronavirus outbreak, EMBL-EBI closed its premises on 18 March at 17:00. Our data resources and tools will continue to function as normal. [Read more >](#)

### Featured topic



The featured topic banner displays the 'COVID-19 Data Portal' with a red coronavirus particle illustration. It includes a 'Watch later' button and a 'Share' button. A dark grey box at the bottom of the banner contains the text 'SARS-CoV-2'.

### Latest news



# Services

Overview

A to Z

Data submission

Research infrastructure development programme

Support

The European Bioinformatics Institute (EMBL-EBI) maintains the world's most comprehensive range of freely available and up-to-date molecular data resources.

Developed in collaboration with our colleagues worldwide, our services let you share data, perform complex queries and analyse the results in different ways. You can work locally by downloading our data and software, or use our [web services](#) to access our resources programmatically.

— You can read more about our services in the journal *Nucleic Acids Research*

## Tools & Data Resources

Search all tools & data resources

### Tools

#### Clustal Omega

Multiple sequence alignment of DNA or protein sequences. Clustal Omega replaces the older ClustalW alignment tools.

Multiple sequence alignment



#### InterProScan

InterProScan searches sequences against InterPro's predictive protein signatures.

Protein feature detection Sequence motif recognition

### Data resources

#### Ensembl

Genome browser, API and database, providing access to reference genome annotation



#### UniProt

A comprehensive resource for protein sequence and functional annotation.



#### PDBe

The European resource for the collection,



### Browse by type

DNA & RNA	Gene Expression	Proteins
Structures	Systems	Chemical biology
Ontologies	Literature	Cross domain

### Programmatic access

EMBL-EBI web services allow you to query our large biological data resources programmatically, so that you can develop data analysis pipelines or integrate public

- Click on Clustal Omega tool for MSA(multiple sequence alignment).

# Clustal Omega

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

Tools > Multiple Sequence Alignment > Clustal Omega

## 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

A →

Enter or paste a set of

PROTEIN

D →

sequences in any supported format:

Or, upload a file:  No file chosen

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



- A. Set the data base DNA, RNA, Protein according to your query sequence which you have.
- B. Put atleast 3 sequence for comparision, and at a time we can put multiple number of sequence for MSA.
- C. Here I have protein sequence of RUBISCO (ribulose-1,5-bisphosphate carboxylase/oxygenase) large subunit from 6 different phylum and sub-phylum of kingdome plante.
  1. *Chlamydomonas raudensis* (algae)
  2. *Riccia fluitans* (Bryophyta)
  3. *Lycopodium clavatum* (Pteridophyta)
  4. *Cycas szechuanensis* (Gymnosperm)
  5. *Oryza sativa Japonica* Group (Angiosperm-Monocot)
  6. *Arabidopsis thaliana* (Angiosperm-Dicot)
- D. Paste the sequences.
- E. Click on submit below.

# UniProtKB 2020\_05 results

## Filter by<sup>i</sup>

[BLAST](#) [Align](#) [Download](#) [Add to basket](#) [Columns](#) [↔](#)

◀ 1 to 25 of 167,967 ▶ Show 25 ▾

 Reviewed (1,008)  
Swiss-Prot

 Unreviewed (166,959)  
TrEMBL

## Popular organisms

[A. thaliana \(62\)](#)
[Rice \(39\)](#)
[B. subtilis \(4\)](#)
[Bovine \(4\)](#)
[Human \(4\)](#)

Other organisms

 [Go](#)

## Search terms

Filter "large" as:

[gene name \(12\)](#)
[gene ontology \(11\)](#)
[organism \(49\)](#)

Quote terms: "rubisco large"

<input type="checkbox"/>	Entry ▾	Entry name ▾	Protein names ▾	Gene names ▾	Organism ▾	Length ▾	
<input type="checkbox"/>	Q43088	RBCMT_PEA	Ribulose-1,5 bisphosphate carboxyla...	RBCMT	Pisum sativum (Garden pea)	489	
<input type="checkbox"/>	P00875	RBL_SPIOL	Ribulose bisphosphate carboxylase l...	rbcl	Spinacia oleracea (Spinach)	475	
<input type="checkbox"/>	P00876	RBL_TOBAC	Ribulose bisphosphate carboxylase l...	rbcl rbcA	Nicotiana tabacum (Common tobacco)	477	
<input type="checkbox"/>	P00880	RBL_SYNP6	Ribulose bisphosphate carboxylase l...	cbbL rbcA, rbcl, syc0130_c	Synechococcus sp. (strain ATCC 27144 / PCC 6301 / SAUG 1402/1) (Anacystis nidulans)	472	
<input type="checkbox"/>	P00877	RBL_CHLRE	Ribulose bisphosphate carboxylase l...	rbcl	Chlamydomonas reinhardtii (Chlamydomonas smithii)	475	
<input type="checkbox"/>	O03042	RBL_ARATH	Ribulose bisphosphate carboxylase l...	rbcl AtCg00490	Arabidopsis thaliana (Mouse-ear cress)	479	
<input type="checkbox"/>	P0C512	RBL_ORYSJ	Ribulose bisphosphate carboxylase l...	rbcl LOC_Osp1g00420, Nip064	Oryza sativa subsp. japonica (Rice)	477	
<input type="checkbox"/>	P0C2C2	RBL1C_CUPNE	Ribulose bisphosphate carboxylase l...	cbbL1 cbbL, cbxLC, cfxLC	Cupriavidus necator (Alcaligenes eutrophus) (Ralstonia eutropha)	486	
<input type="checkbox"/>	P11383	RBL_WHEAT	Ribulose bisphosphate carboxylase l...	rbcl	Triticum aestivum (Wheat)	477	
<input type="checkbox"/>	P21240	CPNB1_ARATH	Chaperonin 60 subunit beta 1, chlor...	CPN60B1 Cpn60-B(3), LEN1, At1g55490, T5A14.11	Arabidopsis thaliana (Mouse-ear cress)	600	
<input type="checkbox"/>	Q31NB3	RBL_SYNE7	Ribulose bisphosphate carboxylase l...	cbbL rbcl, Synpcc7942_1426	Synechococcus elongatus (strain PCC 7942 / FACHB-805) (Anacystis nidulans R2)	472	

- **These are the 6 protein sequences that I have downloaded for doing MSA.**

>AEM75128.1 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial (chloroplast) [*Chlamydomonas raudensis*]

MVPQTETRAGAGFKAGVKDYRLTYYPDYVVKDTDILAAFRMTPQPGVPAEECGAAVAESSTGTWTTVWTDGLTSLDRYKGRICYDIEPVP  
EDNQYIAYVAYPIDLFEEGSVTNLFTSIVGNVFGFKALRALRLEDLRIPPAYVKTFSGPPHGIQVERDKINKYGRGLLGCTIKPKLGLSAKNYGRAVY  
CLRGGLDFTKDDENVNPQPFMRWRDRFLFCAEAIYKAQSETGEVKGHYLNATAATSEEMIKRVVCAKEFGVPIIMHDYLTGGFTANTSLSNYCR  
DHGLLLHIHRAMHAVIDRQRNHGIHFRVLAKALRLSGGDHLHSGTVVGKLEGEREVTLGFDLMDRDNYVEKDRSRGIYFTQDWCSMGGVMP  
VASGGIHVWHMPALVEIFGDDACLQFGGGTLGHPWGNAPGAVANRVALEACTQARNEGRDLAREGGDV

>YP\_009659421.1 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (chloroplast) [*Riccia fluitans*]

MSPQTETKAGAGFKAGVKDYRLTYYPDYETKDTDILAAFRMTPQPGVPAEEAGAAVAESSTGTWTTVWTDGLTNLDRYKGRICYDIEPVAGE  
QNQYIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPPAYTKTFQGPPHGIQVERDKLNKYGRPLLCTIKPKLGLSAKNYGRAVY  
ECLRGGLDFTKDDENVNSQPFMRWRDRFLFVAEALFKAQAETGEIKGHYLNATAGTCEEMLKRAACARELGVPIVMHDYLTGGFTANTTLAFY  
CRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKGLRLSGGDHIHAGTVVGKLEGDREVTLGFDLLRDDYIEKDRSRGVYFTQDWVSLPGVL  
PVASGGIHVWHMPALTEIFGDDSVLQFGGGTLGHPWGNAPGAVANRVALEACVQARNEGRDLARQGNIEIREAAKWSPELAAACEVWKEIK  
FEFDIIDL

>BAJ40178.1 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial (chloroplast) [*Lycopodium clavatum*]

YTPNYKTKDTDILAAFRMTPQPGVPEEAGAAVAESSTGTWTTVWTDGLTSLDRYKGRICYDIEPVVGEKNQYIAYVAYPLDLFEEGSVTNLFTS  
IVGNVFGFKALRALRLEDLRIPPAYSKTFIGPPHGIQVERDKLNKYGRPLLCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNSQPFMRWR  
DRFLFVAEALNKAQVETGEIKGHYLNATAGTYEEMLKRAQCAKELGVPIIMHDYLTGGFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGI  
HFRVLAKALRMSGGDHIHAGTVVGKLEGERDVTLGFDLLRDDYIEKDRSRGIYFTQDWVSMPGVLPVASGGIHVWHMPALTEIFGDDSVLQ  
FGGGTLGHPWGNAPGAVANRVALEA



>YP\_009647987.1 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (chloroplast) [*Cycas szechuanensis*]  
MSPKTETKASVGFKAGVKDYRLTYTPEYQTKDLDILAAFRVTPQPGVPPEEAGAAVAESSTGTWTTVWTDGLTSLDRYKGRCYDIEPVPGEE  
NQFIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRVPPAYSKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYE  
CLRGGLDFTKDDENVNSQPFRWRDRFCFAEAIYKAQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTGGFTANTSLAHYC  
RDNGLLLHIHRAMHAVIDRQRNHGMHFRVLAKALRMSGGDHIHAGTVVGKLEGERDVTLGFDLLRDDFIERDRSRGIYFTQDWVSM PGVL  
PVASGGIHVWHMPALTEIFGDDSVLQFGGGTLGHPWGNAPGAVANRVALEACVQARNEGRDLAREGNEVIREASKWSPELAAACEVWKEIK  
FEFEAMDVL

>ADD62840.1 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (chloroplast) [*Oryza sativa Japonica* Group]  
MSCREGRMSPQTETKASVGFKAGVKDYKLTYYTPEYETKDDILAAFRVTPQPGVPPEEAGAAVAESSTGTWTTVWTDGLTSLDRYKGRCYHI  
EPVVGEDNQYIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPPTYSKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKN  
YGRACYECLRGGLDFTKDDENVNSQPFRWRDRFVFCFAEAIYKSAQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTGGFTAN  
TSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRMSGGDHIHAGTVVGKLEGEREMTLGFVDLLRDDFIEKDRARGIFFTQDW  
VSMPGVIPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLGHPWGNAPGAAANRVALEACVQARNEGRDLAREGNEIIRSACKWSPELAAAC  
EIWKAIKFEFEPVDKLDS

>NP\_051067.1 ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (chloroplast) [*Arabidopsis thaliana*]  
MSPQTETKASVGFKAGVKEYKLTYYTPEYETKDDILAAFRVTPQPGVPPEEAGAAVAESSTGTWTTVWTDGLTSLDRYKGRCYHIEPVPGEE  
TQFIAYVAYPLDLFEEGSVTNMFTSIVGNVFGFKALALRLEDLRIPPAYTKTFQGPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYE  
CLRGGLDFTKDDENVNSQPFRWRDRFLFCFAEAIYKSAQAETGEIKGHYLNATAGTCEEMIKRAVFARELGVPIVMHDYLTGGFTANTSLSHYCR  
DNGLLLLHIHRAMHAVIDRQKNHGMHFRVLAKALRLSGGDHIHAGTVVGKLEGDRESTLGFVDLLRDDYVEKDRSRGIFFTQDWVSLPGVLPV  
ASGGIHVWHMPALTEIFGDDSVLQFGGGTLGHPWGNAPGAVANRVALEACVQARNEGRDLAVEGNEIIREACKWSPELAAACEVWKEITFN  
FPTIDKLDGQE

# Results



1. I got the MSA of the sequences which we have used.

## Clustal Omega

[Input form](#)

[Web services](#)

[Help & Documentation](#)

[Bioinformatics Tools FAQ](#)

[Feedback](#)

[Share](#)

Tools > Multiple Sequence Alignment > Clustal Omega

### Results for job clustalo-I20200929-094231-0878-75372199-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

```
AEH75128.1      -----MNPQTETRAGAGFKAGVKDYRLTYTTPDYVVKDIDILAAFRHTPQPGVPAEEC      53
YP_009659421.1 -----MSPQTEKAGAGFKAGVKDYRLTYTTPDYETKDDTDILAAFRHTPQPGVPAEEA      53
BAJ40178.1      .....YTPNYKTKDDTDILAAFRHTPQPGVPPEEA      29
NP_051067.1      -----MSPQTEKASVGFKAGVKDYRLTYTTPDYETKDDTDILAAFRHTPQPGVPPEEA      53
YP_009647987.1 -----MSPKTEKASVGFKAGVKDYRLTYTTPDYQTKDDTDILAAFRHTPQPGVPPEEA      53
ADD62840.1      HSCREGHMSPTETKASVGFKAGVKDYRLTYTTPDYETKDDTDILAAFRHTPQPGVPPEEA      60
                * * * * *
                * * * * *

AEH75128.1      GAAVAAESSTGTMTTWTVDGLTSLDRYKGRCYDIEPVVGEDNQYIAYVAYPLDLFEEGSV      113
YP_009659421.1 GAAVAAESSTGTMTTWTVDGLTSLDRYKGRCYDIEPVVAGEQNYIAYVAYPLDLFEEGSV      113
BAJ40178.1      GAAVAAESSTGTMTTWTVDGLTSLDRYKGRCYDIEPVVGERNQYIAYVAYPLDLFEEGSV      89
NP_051067.1      GAAVAAESSTGTMTTWTVDGLTSLDRYKGRCYHIEPVVGEETQFIAYVAYPLDLFEEGSV      113
YP_009647987.1 GAAVAAESSTGTMTTWTVDGLTSLDRYKGRCYDIEPVVGEENQFIAYVAYPLDLFEEGSV      113
ADD62840.1      GAAVAAESSTGTMTTWTVDGLTSLDRYKGRCYHIEPVVGEDNQYIAYVAYPLDLFEEGSV      120
                * * * * *
                * * * * *

AEH75128.1      TNLFTSIVGNVFGFKALRALRLEDLRIPPAYVYKTFSGPPHGIQVERDKLNKYGRPLLGCT      173
YP_009659421.1 TNLFTSIVGNVFGFKALRALRLEDLRIPPAYKTFQGPPIHGIQVERDKLNKYGRPLLGCT      173
BAJ40178.1      TNLFTSIVGNVFGFKALRALRLEDLRIPPAYSKTFIGPPHGIQVERDKLNKYGRPLLGCT      149
NP_051067.1      TNLFTSIVGNVFGKALRALRLEDLRIPPAYKTFQGPPIHGIQVERDKLNKYGRPLLGCT      173
YP_009647987.1 TNLFTSIVGNVFGFKALRALRLEDLRVPPAYSKTFQGPPIHGIQVERDKLNKYGRPLLGCT      173
ADD62840.1      TNLFTSIVGNVFGFKALRALRLEDLRIPPYYSKTFQGPPIHGIQVERDKLNKYGRPLLGCT      180
                * * * * *
                * * * * *

AEH75128.1      IKPKLGLSAKNIYGRAVYECLRGGLDFTKDDEINVPQPFMRWRDRFLFCAEAITYKQSETG      233
YP_009659421.1 IKPKLGLSAKNIYGRAVYECLRGGLDFTKDDEINVSQPFMRWRDRFLFVAEALFKAQSETG      233
BAJ40178.1      IKPKLGLSAKNIYGRAVYECLRGGLDFTKDDEINVSQPFMRWRDRFLFVAEALNKAQVETG      209
NP_051067.1      IKPKLGLSAKNIYGRAVYECLRGGLDFTKDDEINVSQPFMRWRDRFLFCAEAITYKQSETG      233
YP_009647987.1 IKPKLGLSAKNIYGRAVYECLRGGLDFTKDDEINVSQPFMRWRDRFCFCAEAITYKQSETG      233
ADD62840.1      IKPKLGLSAKNIYGRACYECLRGGLDFTKDDEINVSQPFMRWRDRFVFCAEAITYKQSETG      240
                * * * * *
                * * * * *
```

# 2. Guide tree

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

Results for job clustalo-l20200929-094231-0878-75372199-p1m

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

## Phylogram

Branch length:  Cladogram  Real



AEM75128.1 0.0597645  
YP\_009659421.1 0.0298507  
BAJ40178.1 0.0298507  
NP\_051067.1 0.0330173  
YP\_009647987.1 0.0294737  
ADD62840.1 0.0294737

## Guide Tree

```
(  
  AEM75128.1:0.0597645  
  )  
(  
(  
  YP_009659421.1:0.0298507  
  BAJ40178.1:0.0298507  
):0.00873429  
(  
  NP_051067.1:0.0330173  
  )  
(  
  YP_009647987.1:0.0294737  
  ADD62840.1:0.0294737  
):0.00354357  

```

# 3. Phylogenetic tree



## Clustal Omega

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

Tools > Multiple Sequence Alignment > Clustal Omega

Results for job clustalo-l20200929-094231-0878-75372199-p1m

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

### Phylogenetic Tree

*This is a Neighbour-joining tree without distance corrections.*

Branch length:  Cladogram  Real



AEM75128.1 0.08078  
BAJ40178.1 0.02121  
YP\_009659421.1 0.03819  
NP\_051067.1 0.03103  
ADD62840.1 0.03187  
YP\_009647987.1 0.02578

### Tree Data

```
{
{
AEM75128.1:0.08078,
BAJ40178.1:0.02121)
:0.00148,
YP_009659421.1:0.03819,
{
```

we can also download the alignment file, guide tree data and phylogenetic tree data.

# Conclusion



- Multiple sequence alignment is a tool used to study closely related genes or proteins in order to find the evolutionary relationships between genes and to identify shared patterns among functionally or structurally related genes.