

# Multiple Sequence Alignment & Phylogeny

Yazdan Asgari

2019



# Multiple Sequence Alignment

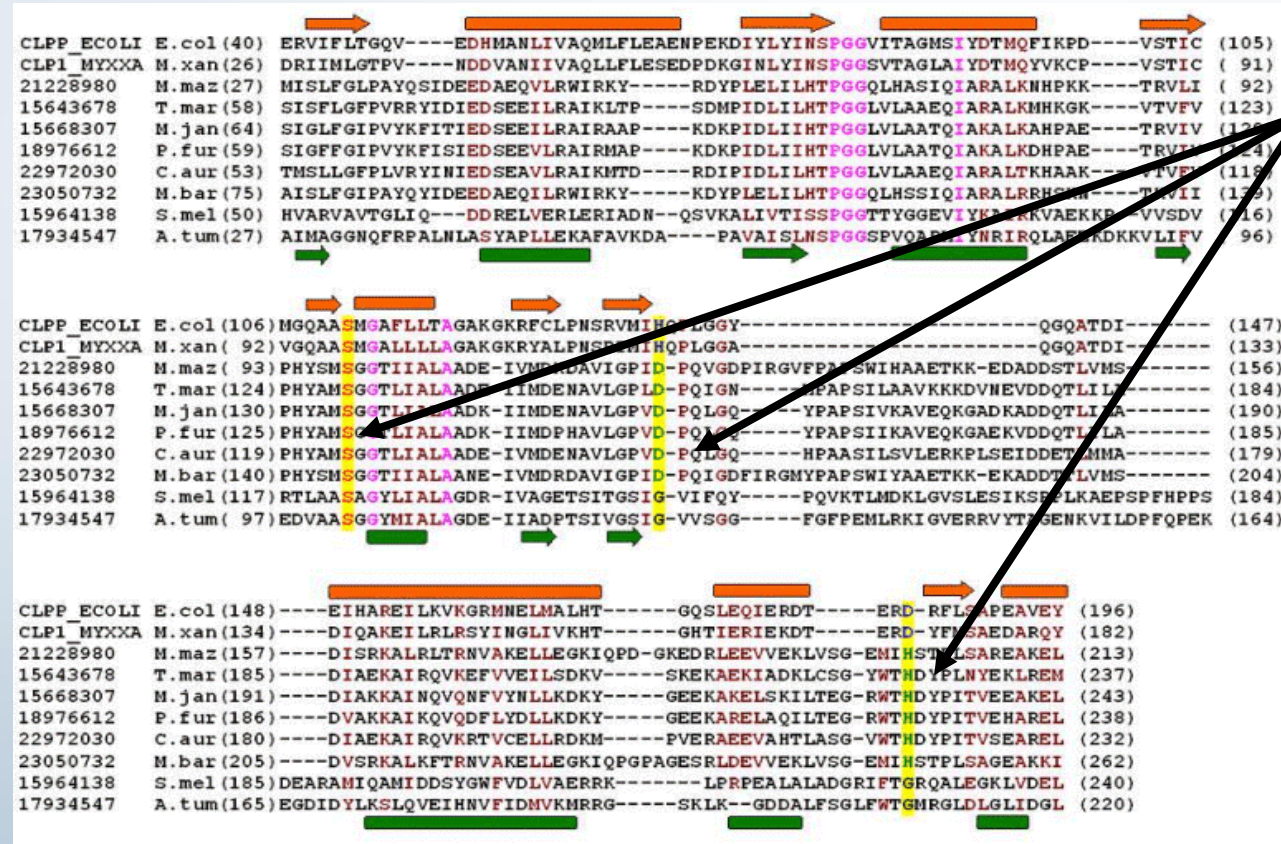
# What Is a Multiple Sequence Alignment ?

- The alignment of more than two sequences
- MSAs = multiple-sequence alignments
- The goal of an MSA is twofold:
  - Aligning corresponding regions of the sequences
  - Revealing positions that are conserved
- The main steps to a useful MSA require
  - Choosing the right sequences
  - Choosing the right MSA method
  - Interpreting the alignment

# Evolution

- Amino acids mutate randomly
- Mutations are then selected (accepted) or counter-selected (rejected)
- If a mutation is harmful, it is counter-selected
  - It disappears from the genome
  - You never see it
- Mutations of important positions (such as active sites) are almost always harmful
- You can recognize important positions because they never mutate!
- MSAs reveal these *conserved* positions

# An Example of Conserved Positions: The Serine Proteases Active Site



CLPP_ECOLI	E.col (40)	ERVIFLTGQV----	EDHMANLIVAQMLFLEAENPEKDIYLYINS	PGGVITAGMSI	YDTMQFIKPD----	VSTIC (105)
CLP1_MYXXA	M.xan (26)	DRIIMLGTFV----	NDDVANIIVAQLLFLESEDPDKGINLYINS	PGGSVTAGLA	YDTMQYVKCP----	VSTIC (91)
21228980	M.maz (27)	MISLFGLPAYQSIDE	EDAEQVLRWIRKY----	RDYPLELILHT	PGGQLHASIQ	ARALKNHPPK----TRVLI (92)
15643678	T.mar (58)	SISFLGFPVRRYIDIE	DSIEILRAIKLTP----	SDMPIDLI	LHTPGGLVLA	AEQIARALKMHKGG--VTVFV (123)
15668307	M.jan (64)	SIGLFGIPVYKFITIE	DSIEILRAIRAAP----	KDKPIDLI	IHTPGGLVLA	ATQIAKALKAHPAE--TRVIV (124)
18976612	P.fur (59)	SIGFFGIPVYKFISIE	DSIEVLRAIMAP----	KDKPIDLI	IHTPGGLVLA	ATQIAKALKDHPAE--TRVIV (124)
22972030	C.aur (53)	TMSLLGFPLVRYINIE	DSIEVLRRAIKMTD----	RDIPIDLI	LHTPGGLVLA	AEQIARALTKHAAK--TVVFV (118)
23050732	M.bar (75)	AISLFGIPAYQYIDE	EDAEQILRWIRKY----	KDYPLELI	LHTPGGQLHSSIQ	ARALRRHSAH----TVVII (179)
15964138	S.mel (50)	HVARVAVTGLIQ---	DDRELVERLERIADN--	QSVKALIV	TISSPGGTT	YGGGEVYERKRVAEKKK--VVSDV (116)
17934547	A.tum (27)	AIMAGGNQFRPALNL	ASYAPLLEKAFVAKDA----	PAVAISLNS	PGGSPVQAP	YNRIRQLAEKDKKVLIFV (96)
CLPP_ECOLI	E.col (106)	MGQAAS	MGAFLLT	AGAKGKRC	CLPNSRVMI	HPGLGGY-----QQQATDI----- (147)
CLP1_MYXXA	M.xan (92)	VGQAAS	MGALLLL	AGAKGKRYALP	SPSHIHP	PLGGA-----QQQATDI----- (133)
21228980	M.maz (93)	PHYSM	SGGTII	ALA	ADE-IVMD	DAVIGPID-PQVGDPIRGVFPN--SIHAAETKK-EDADDSTLVMS----- (156)
15643678	T.mar (124)	PHYAM	SGGTII	ALA	ADE-IIMDEN	AVLGPID-PQIGN-----HPAPSILA
15668307	M.jan (130)	PHYAM	SGGTII	ALA	ADK-IIMDEN	AVLGPVD-PQLGQ-----YPAPSIVKAVEQK
18976612	P.fur (125)	PHYAM	SGGTII	ALA	ADK-IIMDP	HAVLGPVD-PQAG-----YPAPSI
22972030	C.aur (119)	PHYAM	SGGTII	ALA	ADE-IVMD	ENAVLGPVD-PQLGQ-----HPAASILSVLERKPLSEIDDET--MMA----- (179)
23050732	M.bar (140)	PHYSM	SGGTII	ALA	ANE-IVMD	RDAVIGPID-PQIGDFIRGMYPAPSWIYAAETKK-EKADDTLVMS----- (204)
15964138	S.mel (117)	RTLAAS	AGYLI	ALAGDR-IV	AGETSITGSIG-VIFQY-----PQVKTLMDKLGVSLESIKS	PELKAEPSPFHPPS (184)
17934547	A.tum (97)	EDVAAS	SGGYM	IALGDE-IA	DPTISVGSIG-VVSGG-----FGFPEMLRKIGVERRVYTSENK	VILDPPFQPEK (164)
CLPP_ECOLI	E.col (148)	----	EIHAREILKVKGRMNEIM	ALHT-----	GQSLEQIERDT-----	ERD-RFLSAPEAVEY (196)
CLP1_MYXXA	M.xan (134)	----	DIQAKEILRLRSYING	LIVKHT-----	GHTIERIEKDT-----	ERD-YFMSAEDARQY (182)
21228980	M.maz (157)	----	DISRKALRLTRNVAK	ELLEGGIKQPD-GKED	RLEEVVEKLVSG-EMI	STLSAREAKEL (213)
15643678	T.mar (185)	----	DIAEKAI	RQVKEFVVEIL	SDKV-----SKEKA	EKIADKLCSG-YWTDYPLNVEKLREM (237)
15668307	M.jan (191)	----	DIAKKAI	NQVQNFVYNLL	KDKY-----GEEKAK	ELSKILTEG-RWTDYPTVEEAKEL (243)
18976612	P.fur (186)	----	DVAKKAI	KQVQDFLYD	LLKDKY-----GEEKARE	LAQILTEG-RWTDYPTVEHAREL (238)
22972030	C.aur (180)	----	DIAEKAI	RQVKRTVCELL	RDKM-----PVERA	EEVAHTLASG-VWTDYPTVSEAREL (232)
23050732	M.bar (205)	----	DVSRKALKFTTRNV	AKELEGGIKQPG	PAGESRLDEVVEKLVSG-EMI	STPLSAGEAKKI (262)
15964138	S.mel (185)	DEARMI	QAMIDDSYGNF	VDLVA	ERRK-----LPR	PEALALADGRIFTGRQALEGKLVDEL (240)
17934547	A.tum (165)	EGDIDY	LSLQVEIHNV	FIDNV	KMRG-----SKLK--GDD	ALFSGFLFWTGHRLDLGLIDGL (220)

Active Site

# MSA – Example Applications

<i>Application</i>	<i>Procedure</i>
<b>Extrapolation</b>	Determine the <b>function</b> of your protein
<b>Phylogenetic analysis</b>	Build a Phylogenetic tree
<b>Pattern identification</b>	Discover <b>important positions</b>
<b>Domain identification</b>	Turn your alignment into a <b>domain profile</b>

# MSA – Example Applications

<b>DNA regulatory elements identification</b>	Use your alignment to discover <b>promoters</b>
<b>Structure prediction</b>	Predict the <b>secondary structure</b> of <b>proteins</b> and <b>RNA</b> molecules
<b>nsSNP analysis</b>	Discover important <b>allelic variations</b> in human and other animals (nsSNP: non-synonymous single-nucleotide polymorphisms)
<b>PCR analysis</b>	Select your <b>PCR primers</b>

# Choosing the Right Sequences

- When building an alignment, it is your job to select the sequences
- Two main factors when selecting sequences:
  - Number of sequences
  - Nature of the sequences
- A reasonable number of sequences: 20 to 50
  - Ideal for most methods
  - Small alignments are easy to display and analyze
- Types of sequences
  - Well-selected sequences  $\Leftrightarrow$  informative alignment



# Choosing Sequences that are Different Enough

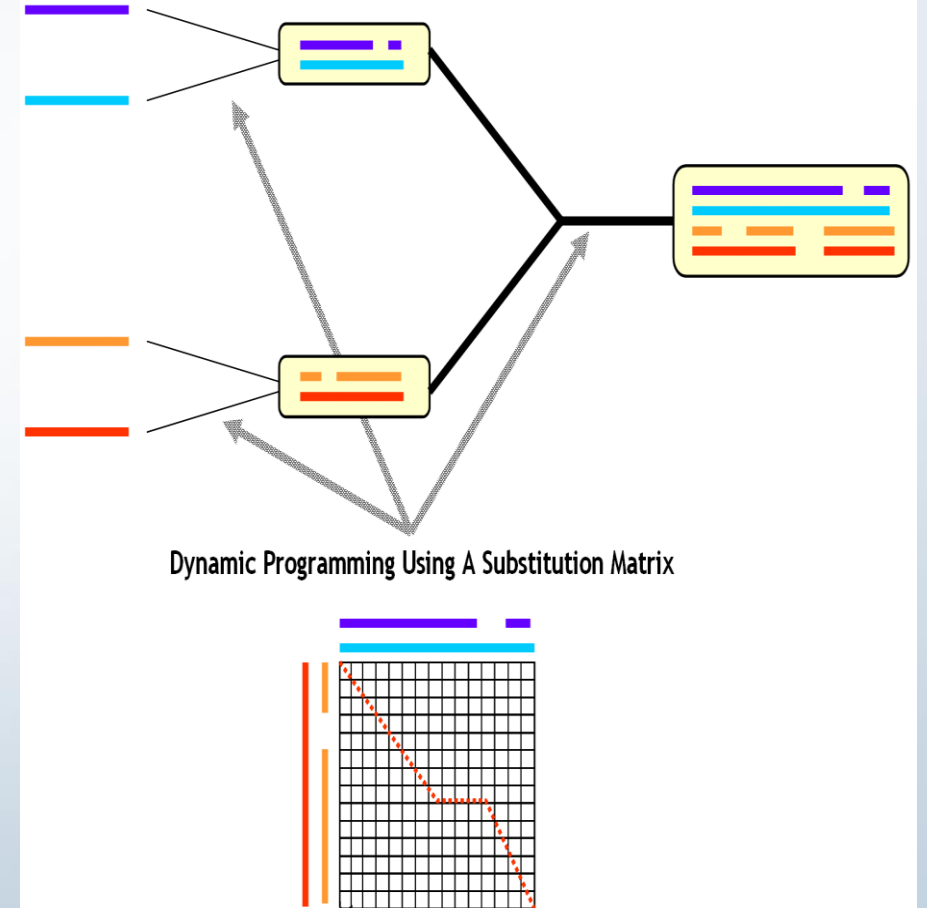
- An alignment is useful if . . .
  - The sequences are correctly aligned
  - It can be used to produce trees, profiles, and structure predictions
- To obtain this result, the sequences must be
  - Not too similar
  - Not too different
- Sequences that are very similar . . .
  - Are easy to align correctly
  - Are not informative  $\Leftrightarrow$  useless trees and profiles, bad predictions
- Sequences that are very different . . .
  - Are difficult to align
  - Are very informative  $\Leftrightarrow$  good trees and profiles, good predictions

# DNA or Proteins?

- DNA sequences are harder to align than proteins
  - DNA-comparison models are less sophisticated
- Most methods work for both DNA and proteins
  - The results are less useful for DNA
- If your DNA is coding, work on the translated proteins
- If sequences are homologous . . .
  - Along their entire length  $\Leftrightarrow$  use progressive alignment methods
  - In terms of local similarity  $\Leftrightarrow$  use motif-discovery methods

# MSA Methods: Progressive Algorithm

- Sequences are grouped by similarity (guide tree)
- Sequences are aligned 2 by 2
- The intermediate alignments are then aligned 2 by 2
- You align 2 sequences by using dynamic programming



# MSA Methods: Progressive Algorithm

- Its main strength is its speed
- Its main weakness is its greed
  - Sequences aligned at the beginning are never realigned
  - Early mistakes cannot be corrected
- Assemble datasets with lots of intermediate sequences
- Imagine each sequence is part of a stone bridge across a river:
  - Doesn't matter how wide the river is, if the stones are close enough together
  - Doesn't matter how diverse your sequences are, if each sequence has a close relative



<https://www.tes.com>

# Selecting a Method

- Many alternative methods exist for MSAs
- Most of them use the **progressive algorithm**
- They all are approximate methods
- None is guaranteed to deliver the best alignments
- All existing methods have pros and cons
  - **ClustalW** is the most popular
  - T-Coffee/M-Coffee and ProbCons are more accurate but slower
  - MUSCLE is very fast, ideal for very large datasets

# Aligning Your Sequences Correctly

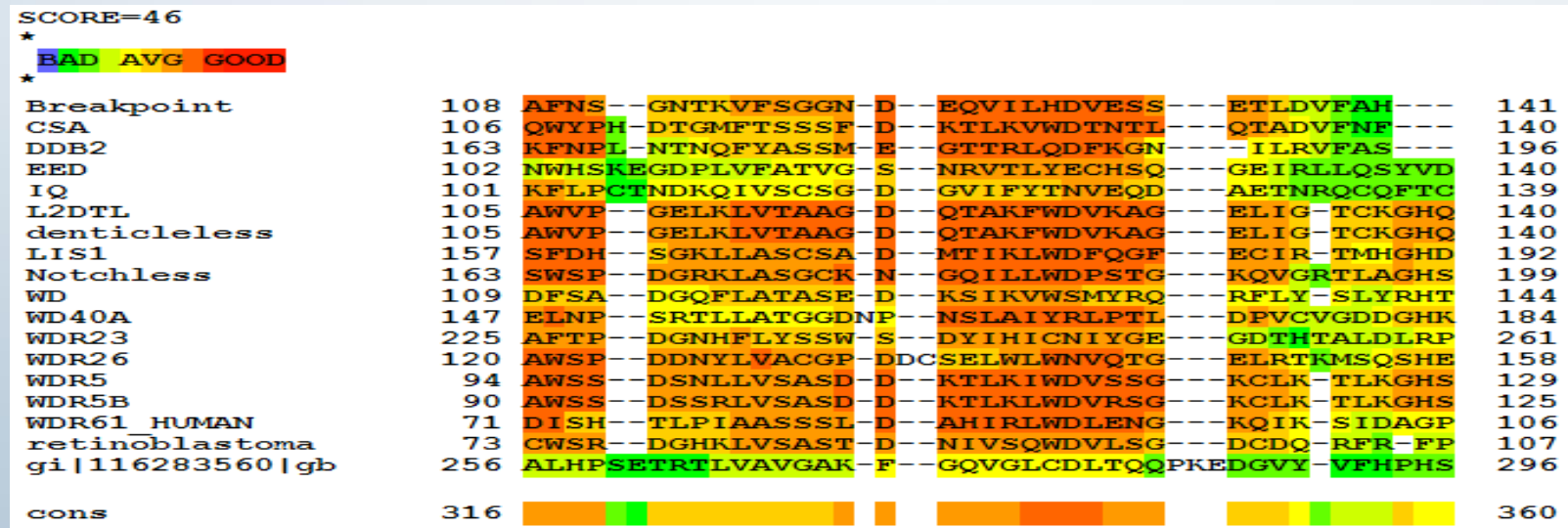
- It can be difficult to align sequences correctly
  - They evolve too fast
- For proteins, the best alternative is to use 3D structure
  - 3D Structures change slower than sequences
- Unfortunately few sequences have a known structure
- Espresso lets you find the structures that correspond to your sequences and use them to build an MSA

# Using the EXPRESSO Web server

- The EXPRESSO Web server aligns sequences using 3D information
- The EXPRESSO tool . . .
  - Looks in PDB for the structure of your sequences
  - Aligns your sequences using structural information
  - Returns a multiple-sequence alignment based on structure
- If your sequences have a known structure, EXPRESSO is the most accurate MSA method

# EXPRESSO Web server – Colored Output

- Red and Orange residues are probably well aligned
- Yellow should be treated with caution
- Green and blue are probably incorrectly aligned





# EXPRESSO Web server - Example

**TCOFFEE**[Home](#)[History](#)[Tutorial](#)[References](#)[Contacts](#)[Projects](#)[Download](#)

## Expresso

*Aligns protein sequences using structural information*

### Sequences input

Paste or upload your set of sequences in FASTA format

Sequences to align  
[Click here to use the sample file](#)

```
>1aboA
NLFVALYDFVASGDNTLSITKGEKLRVLGYNHNGEWCEAQTKNGQGWPVS
NYITPVN
>1ycsB
KGVYALWDYEPQNDDELPMKEGDCMTIIHREDEDEIEWWWARLNDKEGY
VPRNLLGLYP
>1pht
GYQYRALYDYKKEREEDIDLHLGDILTVNKGSLVALGFSDGQEARPEEIG
```

- OR - [Click here to upload a file](#)

[Show more options](#)

Your email address

[Reset](#)

<http://tcoffee.crg.cat/apps/tcoffee/do:expresso>

# EXPRESSO Web server - Example

## Expresso alignment result

### MSA

The multiple sequence alignment result as produced by T-coffee.

T-COFFEE, Version\_11.00.d625267 (2016-01-11 15:25:41 - Revision d625267 - Build 507)

Cedric Notredame

SCORE=91

\*

BAD AVG GOOD

\*

laboA	:	92
lvcsB	:	92
lpht	:	88
lvie	:	91
lihvA	:	91
cons	:	9

laboA	-NLFVALYDFVASGDNTLSITKGEKLRVL	-----GYNHN---	GEWCEAOTK---	NGOGWVPSNYITPVN
lvcsB	KGVIYALWDYEPONDDELPMKEGDCMTII	-----HRED-EDEIEWWWARLN---	DKEGYVPRNLLGLYP	
lpht	GYOYRALYDYKKEREEDIDLHLGDILTVNKGSLVALGFSD	GOEARPEEIGWLN	GYNETT	-GERGDFPGTYVEYIG
lvie	-DRVRKKSG-----AAWOGOIV	-----GWYCTNLTPEGYAVESEAH	PSVOIYPVAALERIN	
lihvA	NFRVYYRDS-----RDPVWKGPAKLL	-----WKG-----EGAVVIQDN---	SDIKVVPRRKAKIIR	

cons

		.				:					.	*
--	--	---	--	--	--	---	--	--	--	--	---	---

# Local Multiple Comparison Methods

- Most MSA programs assume your sequences are related along their whole length
- When this assumption is not true, the progressive approach will not work
- The only alternative is to compare multiple sequences locally

# Local Multiple-Comparison Methods

- Gibbs Sampler
  - Will make a local multiple alignment
  - Will ignore unrelated segments of your sequences
  - Ideal for finding DNA patterns such as promoters
- Motif discovery methods
  - Will look for motifs conserved in your sequences
  - The sequences do not need to be aligned
- The most popular motif-discovery methods:
  - TEIRESIAS, MEME, SMILE, PRATT

# MSA - Summary

- Assembling MSAs is a bit of an art
- Experience is a key factor
- Most methods are now available online
- Make sure you know which method to use:
  - ClustalW-like method to align homologous sequences
  - Motif method to look for conserved regions



# Phylogeny

# Why Build a Phylogenetic Tree ?

- Phylogenetic trees reconstruct the evolutionary history of your sequences
- They tell you who is closer to whom in the big tree of life
- Phylogenetic trees are based on sequence similarity rather than morphologic characters

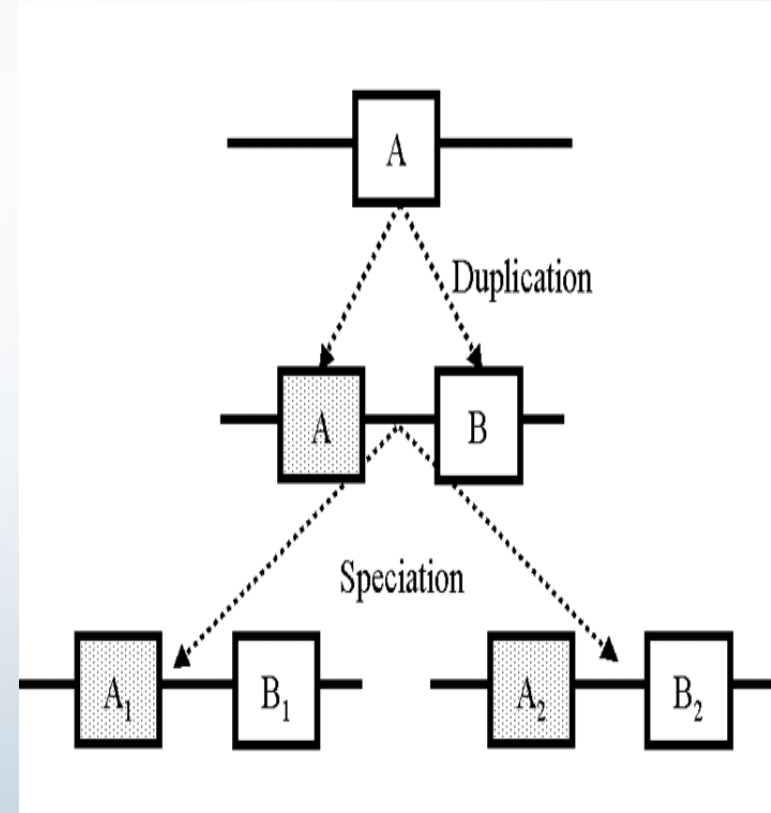
# 3 applications to use your constructed Tree

- Finding the closest relative of your organism
  - Usually done with a tree based on the ribosomal RNA
- Discovering the function of a gene
  - Finding the orthologues of your gene
- Finding the origin of your gene
  - Finding whether your gene comes from another species



# Orthology and Paralogy

- Orthologous genes
  - Separated by speciation
  - Often have the same function
- Paralogous genes
  - Separated by duplications
  - Can have different functions
- In the graph:
  - A is paralogous with B
  - A1 is orthologous with A2



# Working on the Right Data

- Garbage in  $\Rightarrow$  Garbage out
- The quality of your tree depends on the quality of the data
- Your first task is to assemble a very accurate MSA

# Again: DNA or Proteins

- Most phylogenetic methods work on **Proteins** and **DNA** sequences
- If possible, always compute a multiple-sequence alignment on the protein sequences
  - Translate the sequences if the DNA is coding
  - Align the sequences
  - Thread the DNA sequences back onto the protein MSA
- If your DNA sequences are coding and have more than 70% identity . . .
  - Compute the **tree** on the **DNA** multiple-sequence alignment
- If your DNA sequences are coding and have less than 70% identity . . .
  - Compute the **tree** on the **protein** multiple-sequence alignment

# Which Sequences ?

- Orthologous sequences
  - Produce a species tree
  - Show how the considered species have diverged
- Paralogous sequences
  - Produce a gene tree
  - Show the evolution of a protein family

# Building the Right Tree

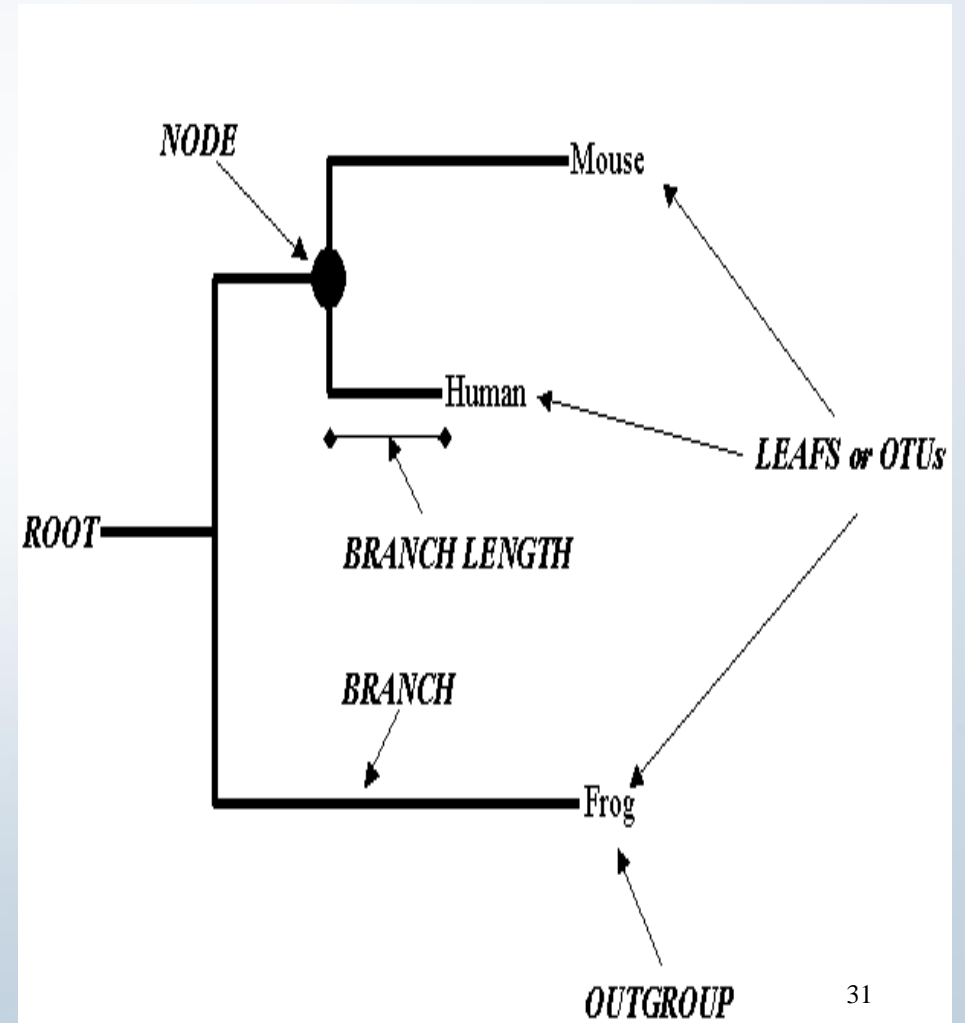
- There are two types of tree-reconstruction methods
  - Distance-based methods
  - Statistical methods
    - Statistical methods are the most accurate
      - Maximum likelihood of success
      - Parsimony
    - Statistical methods take more time
    - Limited to small datasets

# Distance-based Methods for Tree Reconstruction

- Distance-based methods are the most popular
  - Neighbor Joining (NJ)
  - UPGMA
- Distance-based methods involve 2 steps:
  - Measure the distances between pairs of sequences in the MSA
  - Transform the distance matrix into a tree
- The two most popular packages for making trees are
  - **Clustalw**: very simple, not very sophisticated
  - **Phylip**: very powerful, less convivial

# Reading Your Tree

- There's a lot of vocabulary in a tree
- **Nodes** correspond to common ancestors
- **Root** is the oldest ancestor
  - Often artificial
  - Only meaningful with a good outgroup
- Trees can be un-rooted
- Branch lengths are only meaningful when the tree is scaled
  - **Cladograms** have the same branch lengths
  - **Phylograms** have real branch lengths



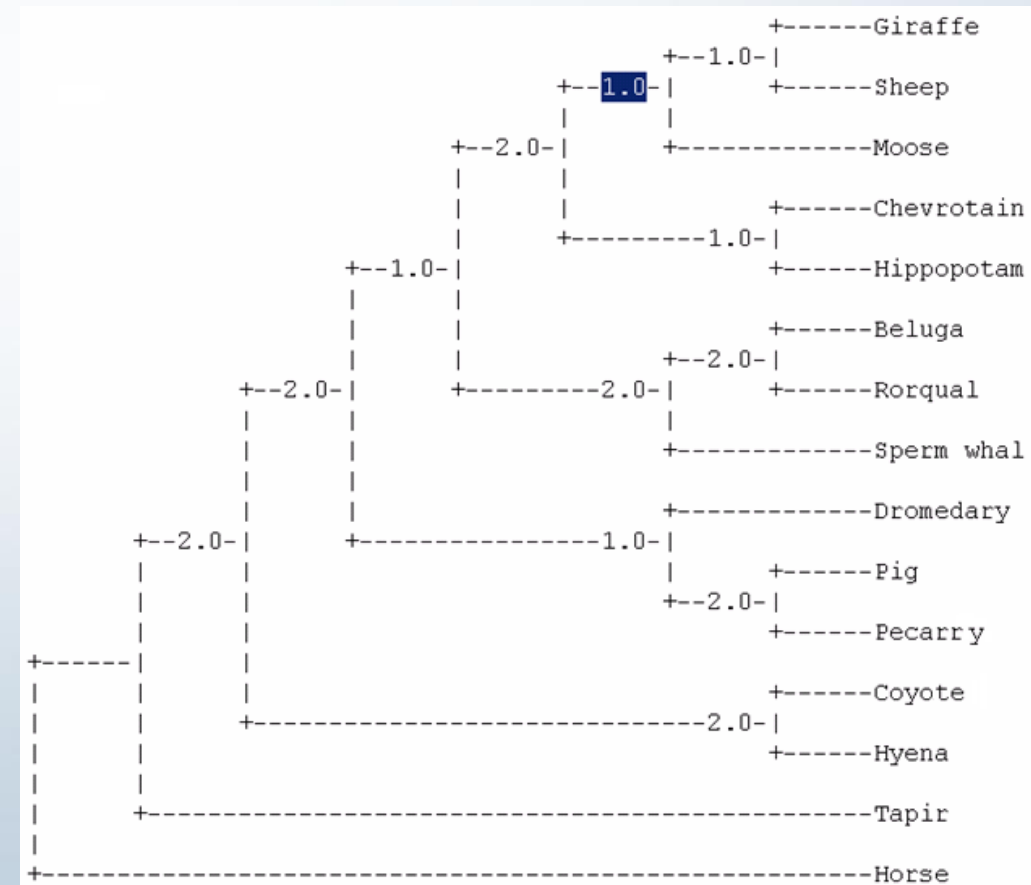
# Bootstrapping

- Use bootstrapping to verify the solidity of each node
- ClustalW and Phylip do bootstrap operations automatically
- Bootstrapping involves these steps:
  - Select a subset of your MSA
  - Redo the tree
  - Repeat this operation N times (100 or 1000 times if you can)
  - Compute a consensus tree of the N trees
  - Measure how many of the N trees agree with the consensus tree on each node
- Each node gets a bootstrap figure between 0 and N
- High bootstrap  $\Leftrightarrow$  good node



# A Bootstrapped Tree

- This tree was produced with 2 bootstrap cycles
- It shows some nodes as more robust than others
- In practice, always use more than 100 cycles





# Practical Session

# EMBL-EBI

## Multiple Sequence Alignment

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

[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](#)

### 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.

### 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.

# Example: Human TNF-alpha orthologous

## TNF tumor necrosis factor [ *Homo sapiens* (human) ]

Gene ID: 7124, updated on 6-Dec-2016

### Summary

Official Symbol	TNF <small>provided by <a href="#">HGNC</a></small>
Official Full Name	tumor necrosis factor <small>provided by <a href="#">HGNC</a></small>
Primary source	<a href="#">HGNC:HGNC:11892</a>
See related	<a href="#">Ensembl:ENSG00000232810</a> <a href="#">MIM:191160</a> <a href="#">Vega:OTTHUMG00000031194</a>
Gene type	protein coding
RefSeq status	REVIEWED
Organism	<a href="#">Homo sapiens</a>
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	DIF; TNFA; TNFSF2; TNLG1F; TNF-alpha
Summary	<p>This gene encodes a multifunctional proinflammatory cytokine that belongs to the tumor necrosis factor (TNF) superfamily. This cytokine is mainly secreted by macrophages. It can bind to, and thus functions through its receptors TNFRSF1A/TNFR1 and TNFRSF1B/TNFR. This cytokine is involved in the regulation of a wide spectrum of biological processes including cell proliferation, differentiation, apoptosis, lipid metabolism, and coagulation. This cytokine has been implicated in a variety of diseases, including autoimmune diseases, insulin resistance, and cancer. Knockout studies in mice also suggested the neuroprotective function of this cytokine. [provided by RefSeq, Jul 2008]</p>
Orthologs	<a href="#">mouse</a> <a href="#">all</a>

# Clustal Omega / DNA

EMBL-EBI

Services

Research

Training

About us

Clustal Omega

Input form

Web services

Help & Documentation

Share

Feedback

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](#).

STEP 1 - Enter your input sequences

Enter or paste a set of DNA sequences in any supported format:

>NM\_001114778.1 Xenopus laevis tumor necrosis factor L homeolog (tnf.L), mRNA  
ATTTCAAAGAAGAAAAAGAACTGGAGCAGAGCAGTGAAGACGTGAACCAAGTGGAAATATAAAATGAAT  
AGTGTAGAGTCACAGATGGAGAATGGGCTCCTAATTGTCAGGCAGGAAAGAAGCAACAGGAATTCCACCT  
GGCGCTTTGTCAGTATCTGTGCATTCTGTTATTACTGGGATCCACCATCCTCTTTGCTCTGCTGCATT  
TCAGATCATCCCAAACCTTTGCTGAAAAGGATGAGAGCAAGATGCCGGAAATATTGGCTATCAAGCTTGT  
TTAGAGTCTGTACCAAGTAGCCCAAGCAATGAAAGGGGAAAAGCTGGCAGCTCATTTTACAGGTTTTAAAG  
AAAGTGACAAATTATTTGGGAGCCATTTGATTCCGATTCATCTGTTGCTGAAAAAATGCTCAAGGATAA

Or, upload a file: Choose File No file chosen

STEP 2 - Set your parameters

OUTPUT FORMAT Clustal w/o numbers

The default settings will fulfill the needs of most users and, for that reason, are not visible.

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

# Results

Results for job clustalo-l20161215-140559-0800-79720144-pg

Alignments

Result Summary

Phylogenetic Tree

Submission Details

Download Alignment File

Send to Simple\_Phylogeny

CLUSTAL O(1.2.3) multiple sequence alignment

```
NM_001024447.1 -----
NM_001200172.1 -----
NM_001114778.1 ATTTCAAAGAAGAAAAAGAACTGGAGCAGAGCAGTGAAGACGTGAACCAAGTGGAAATA
NM_001113671.1 --AAAGAAAGAAAGAAAGAAAGAAAGCAGAGCGGTGAAGCGTGAACCAAGTGAATA
NM_214022.1 -----CCGAGAGTGAGGACACAGGGGACCCAGCCAGGAGAGA
NM_001286442.1 -----
NM_173966.3 -----AACAGAAGCTCCAGAGCGGGGACACAGGGGACCCAGGAGAGA
NM_012675.3 -----CACCAAGGGACCCAGCCAGGAGGGA
NM_001082263.1 -----AAGCTCCCTCAGTGAG--GACACGGGACCCAGTAGGAGGGA
NM_001003244.4 -----
NM_001257261.1 -----A
NM_001285277.1 -----
NM_000594.3 -----CAGACGCTCCCTCAGCAAGGACAGCAGAGGACCCAGCTAAGAGGGA
NM_001045511.1 -----
```

Results for job clustalo-l20161215-140559-0800-79720144-pg

Alignments

Result Summary

Phylogenetic Tree

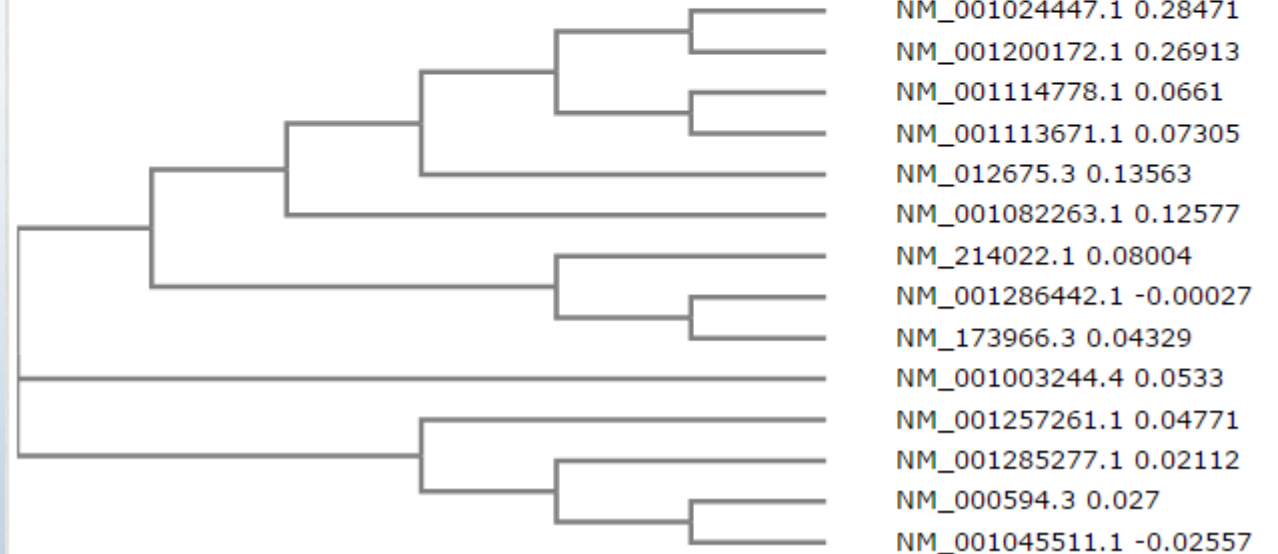
Submission Details

## Phylogenetic Tree

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

Download Phylogenetic Tree Data

Branch length: ☒ Cladogram ☐ Real



# Clustal Omega / Protein

EMBL-EBI

Services

Research

Training

About us

## Clustal Omega

Input form

Web services

Help & Documentation

Share

Feedback

[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](#).

STEP 1 - Enter your input sequences

Enter or paste a set of **PROTEIN** sequences in any supported format:

>NP\_001187101.1 tumor necrosis factor [Ictalurus punctatus]  
MASDSQVLDVDGPRVTIVREKASWSSSGVWRTCGVLLAVALCAAAAVCFSSQNKTHNKPDETQEIKHSLR  
QISQTAKAAIHLSGHYNPQVSSVSMQWFDNADQSFSSGLKLEDNEIKILRDGLYFVYSQASYRLLCKAEG  
DETEGEVMHMSHKVSRWSDSYSSWKPLLSATRSACKKTTEEYQKYWYGAVYLGAAFNLKAGDRLRTVMDE  
KLLPKVESAGGKTFFGTFSL

Or, upload a file: 

Choose File

 No file chosen

STEP 2 - Set your parameters

OUTPUT FORMAT

Clustal w/o numbers

The default settings will fulfill the needs of most users and, for that reason, are not visible.

More options...

 (Click here, if you want to view or change the default settings.)



# Results

Results for job clustalo-l20161215-140736-0483-2746817-pg

Alignments Result Summary Phylogenetic Tree Submission Details

Download Alignment File Show Colors Send to Simple\_Phylogeny

CLUSTAL O(1.2.3) multiple sequence alignment

```
NP_001187101.1 MASDSQVVLVDVG---PRVTIVREKASWSSSGVWRTCGVLLAVALCAAAVCFSQN----
NP_001108250.1 -----MNSVESQMENGLLIVRQER-SNRNSTWR-FVSICAFLLLLGSTILFALLHFQI
NP_001107143.1 -----MNTVESQMENGLLIVRHER-SNRDSTWR-CVSICAFLLLLGSTILFALLHFQI
NP_036807.1 MST-ESMIRDVEL--AEEALPKKMGG-LQNSRRCL-CLSLFSFLLVAGATTFLCLLHFGV
NP_001273371.1 MST-KSMIRDVEL--AEEVLKKAGG-PQGSRRSCV-CLSLFSFLLVAGATTFLCLLHFGV
NP_999187.1 MST-ESMIRDVEL--AEEALAKKAGG-PQGSRRCL-CLSLFSFLLVAGATTFLCLLHFEV
NP_001244190.1 MST-ETMIQDVEL--AEEALPKT-RG-PQGSRRCL-FLSLFSFLLVAGATTFLCLLHFGV
NP_001003244.4 MST-ESMIRDVEL--AEEPLKKAGG-PQGSRRCF-CLSLFSFLLVAGATTFLCLLHFGV
NP_000585.2 MST-ESMIRDVEL--AEEALPKKTGG-PQGSRRCL-FLSLFSFLLVAGATTFLCLLHFGV
NP_001038976.1 MST-ESMIRDVEL--AEEALPKKTGG-PQGSRRCL-FLSLFSFLLVAGATTFLCLLHFGV
: *: . : : : *
```

```
NP_001187101.1 ----KTHNKPDQTQ--EIKHSL-----RQIS--QTAKAAIHLSGHYNPQVSSVSMQWFDN
NP_001108250.1 IPNFAEKDESKMPEILAIKSCLESVPVA--QAMKGEKLAH----FTGFKESDKLFWEPF
NP_001107143.1 IPNFANKDESKMPEILAIKTYLESVPVARAQARKGDKLAH----FTGAKENDKLWNKY
NP_036807.1 IGPNNKEKFPNGLPL--ISSMAQTLTLRSSSQNSSDKPVAVH---VANHQAEQLEWLSQ
NP_001273371.1 IGPQREEQSPAGPSF--NRPLVQT--LRSSSQASSNKPVAHV---VANISAPGQLRWGDS
NP_999187.1 IGPQKEE-FPAGPLS--INPLAQG--LRSSSQ-TSDKPVAVH---VANVKAEGQLQWQSG
NP_001244190.1 IGPQKDE-LSKDFSL--ISPLALA--VRSSSRIPSDKPVAVH---VANPQAEGLQWLNR
NP_001003244.4 IGPQREE-LPNGLQL--ISPLAQT--VKSSSRTPSDKPVAVH---VANPEAEGQLQWLSR
NP_000585.2 IGPQREE-FPRDLSL--ISPLAQA--VRSSSRTPSDKPVAVH---VANPQAEGLQWLNR
NP_001038976.1 IGPQREE-FPRDLSL--ISPLAQA--G-SSSRTPSDKPVAVH---VANPQAEGLQWLNR
```

Results for job clustalo-l20161215-140736-0483-2746817-pg

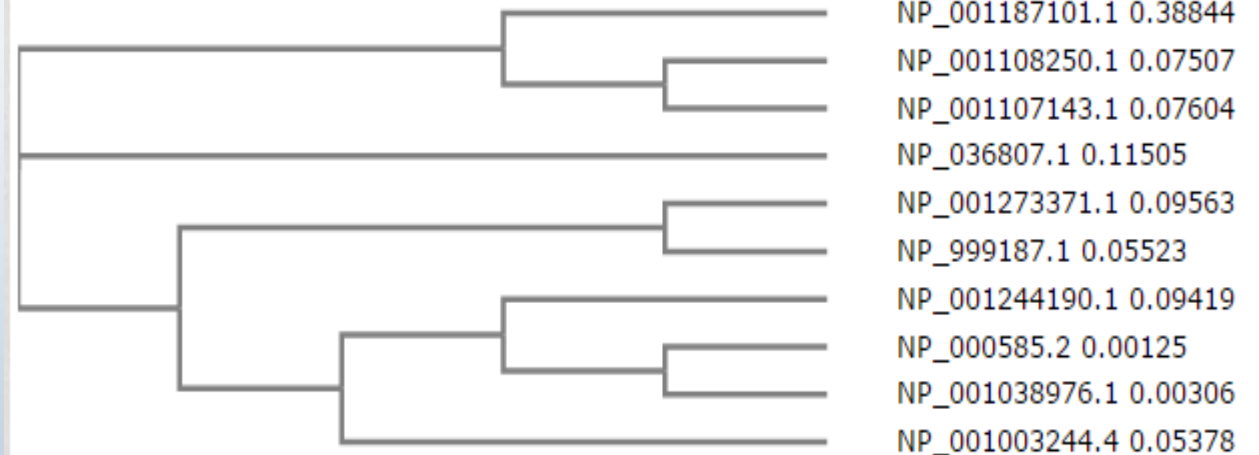
Alignments Result Summary **Phylogenetic Tree** Submission Details

## Phylogenetic Tree

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

Download Phylogenetic Tree Data

Branch length: ☒ Cladogram ☐ Real





# NCBI



U.S. National Library of Medicine

NCBI National Center for Biotechnology Information

[Sign in to NCBI](#)

## COBALT

### Constraint-based Multiple Alignment Tool

[Home](#) [Recent Results](#) [Help](#)

COBALT computes a multiple protein sequence alignment using conserved domain and local sequence similarity information. [?](#)

[Reset page](#)

#### Enter Query Sequences

Enter at least 2 protein accessions, gis, or FASTA sequences [?](#)

[Clear](#)

```
>NP_001187101.1 tumor necrosis factor [Ictalurus punctatus]  
MASDSQVLDVDGPRVTIVREKASWSSSGVWRTCGVLLAVALCAAAVCFSQNKTHNKPDETQEIKHSLR  
QISQTAKAAIHLSGHYNPQVSSVSMQWFDNADQSFSSGLKLEDNEIKILRDGLYFVYSQASYRLLCKAEG  
DETEGEVMHSHKVSRWSDSYSSWKPLL SATRSACKKTEEYQKYWYGAVYLGAAFNLKAGDRLRTVMDE  
KLLPKVESAGGKTFFGTFSL
```

Or, upload FASTA file

Choose File

No file chosen

Job Title

**Align**



Show results in a new window

[Advanced parameters](#)

BLAST is a registered trademark of the National Library of Medicine.

[Copyright](#) | [Disclaimer](#) | [Privacy](#) | [Accessibility](#) | [Contact](#) | [Send feedback](#)

[NCBI](#) | [NLM](#) | [NIH](#) | [DHHS](#)

[https://www.ncbi.nlm.nih.gov/tools/cobalt/re\\_cobalt.cgi](https://www.ncbi.nlm.nih.gov/tools/cobalt/re_cobalt.cgi)

# Results

▼ **Descriptions** ☒ Select All [Re-align](#) [Alignment parameters](#)

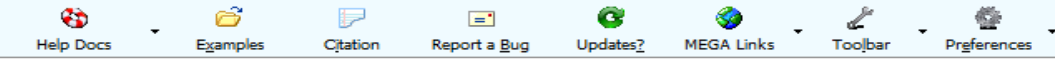
Legend for links to other resources: [U](#) UniGene [E](#) GEO [G](#) Gene [S](#) Structure [M](#) Map Viewer

Accession	Description
<input checked="" type="checkbox"/> Icl Query_10001	NP_001108250.1 tumor necrosis factor L homeolog [Xenopus laevis]
<input checked="" type="checkbox"/> Icl Query_10002	NP_001107143.1 tumor necrosis factor [Xenopus tropicalis]
<input checked="" type="checkbox"/> Icl Query_10003	NP_000585.2 tumor necrosis factor [Homo sapiens]
<input checked="" type="checkbox"/> Icl Query_10004	NP_001273371.1 tumor necrosis factor [Capra hircus]
<input checked="" type="checkbox"/> Icl Query_10005	NP_001244190.1 tumor necrosis factor [Callithrix jacchus]
<input checked="" type="checkbox"/> Icl Query_10006	NP_001003244.4 tumor necrosis factor [Canis lupus familiaris]
<input checked="" type="checkbox"/> Icl Query_10007	NP_001038976.1 tumor necrosis factor [Pan troglodytes]
<input checked="" type="checkbox"/> Icl Query_10008	NP_999187.1 tumor necrosis factor [Sus scrofa]
<input checked="" type="checkbox"/> Icl Query_10009	NP_036807.1 tumor necrosis factor [Rattus norvegicus]
<input checked="" type="checkbox"/> Icl Query_10010	NP_001187101.1 tumor necrosis factor [Ictalurus punctatus]

▼ **Alignments** ☒ Select All [Re-align](#) [Mouse over the sequence identifier for sequence title](#)

View Format: [Compact](#) [?](#) Conservation Setting: [2 Bits](#) [?](#)

<input checked="" type="checkbox"/> Query_10001	1	MNSvESQMENGLL----IVRQERSNRNSTWR--FVSICAFLLLLLGSTILFALLHFQIIpnfAEKDESKMPEILAIKSCLE	74
<input checked="" type="checkbox"/> Query_10002	1	MNTvESQMENGLL----IVRHESNRDSTWR--CVSICAFLLLLLGSTILFALLHFQIIpnfANKDESKMPEILAIKTYLE	74
<input checked="" type="checkbox"/> Query_10003	1	MST-ESMIRDVELAE-EALPKKTGGPQGSRRCLFSLFSFLIVAGATTLFCLLHFGVI---GPQREE-FPRDLSLISPLA	74
<input checked="" type="checkbox"/> Query_10004	1	MST-KSMIRDVELAE-EVLSKKAGGPQGSRSWCLSLFSFLLVAGATTLFCLLHFGVI---GPQREEQSPAGPSFNRPLV	75
<input checked="" type="checkbox"/> Query_10005	1	MST-ETMIQDVELAE-EALPK-TRGPQGSKRRLFLSLFSFLLVAGATALFCLLHFGVI---GPQKDE-LSKDFSLSISPLA	73
<input checked="" type="checkbox"/> Query_10006	1	MST-ESMIRDVELAE-EPLPKKAGGPPGSRRRCFLSLFSFLLVAGATTLFCLLHFGVI---GPQREE-LPNGLQLISPLA	74
<input checked="" type="checkbox"/> Query_10007	1	MST-ESMIRDVELAE-EALPKKTGGPQGSRRCLFSLFSFLIVAGATTLFCLLHFGVI---GPQREE-FPRDLSLISPLA	74
<input checked="" type="checkbox"/> Query_10008	1	MST-ESMIRDVELAE-EALAKKAGGPQGSRRCLCLSLFSFLLVAGATTLFCLLHFEVI---GPQKEE-FPAGPLSINPLA	74
<input checked="" type="checkbox"/> Query_10009	1	MST-ESMIRDVELAE-EALPKKMGGQLNSRRCLCLSLFSFLLVAGATTLFCLLNFGVI---GPNKEEKFPNGLPLISSMA	75



# Results

M7: Pairwise Distances (C:\Users\yazdan\AppData\Local\Temp\PhyloAnalysis.m...)

File Display Average Caption Help

(A,B) 0.0 0.00 [Save] [XL] [CSV] [MEGA] [TXT] [Copy]

	1	2	3	4	5
1. NP 001108250.1 tumor necrosis factor L homeolog Xenopus laevis					
2. NP 001107143.1 tumor necrosis factor Xenopus tropicalis	0.182				
3. NP 000585.2 tumor necrosis factor Homo sapiens	1.224	1.224			
4. NP 001273371.1 tumor necrosis factor Capra hircus	1.175	1.159	0.231		
5. NP 001244190.1 tumor necrosis factor Callithrix jacchus	1.207	1.191	0.125	0.321	
6. NP 001003244.4 tumor necrosis factor Canis lupus familiaris	1.241	1.207	0.092	0.224	0.1
7. NP 001038976.1 tumor necrosis factor Pan troglodytes	1.241	1.241	0.005	0.231	0.1
8. NP 999187.1 tumor necrosis factor Sus scrofa	1.191	1.159	0.148	0.171	0.2
9. NP 036807.1 tumor necrosis factor Rattus norvegicus	1.224	1.159	0.243	0.328	0.3
10. NP 001187101.1 tumor necrosis factor Ictalurus punctatus	1.447	1.386	1.367	1.329	1.3

[1,1] (NP 001108250.1 tumor necrosis factor L homeolog Xenopus laevis-NP 001108250.1 tun

