

Generating Multiple Sequence Alignments with ClustalW

- ## ClustalW: creating a multiple sequence alignment
- You identify sequences to align
 - Get Fasta sequences and copy and paste into window or put all sequences in one document and attach (example: copy and paste into one Word document)

ClustalW: Creating the initial alignment

The screenshot shows the ClustalW web interface. The main heading is 'ClustalW2 - Multiple Sequence Alignment'. Below it, there are four steps:

- STEP 1 - Enter your input sequences:** 'Enter or paste a set of (from 2) sequences in any standard format.' An arrow points to the input field.
- STEP 2 - Set your Pairwise Alignment Options:** Includes 'Alignment Type: Slow Fast' and 'The default settings will fulfil the needs of most users and, for that reason, are not visible.'
- STEP 3 - Set your Multiple Sequence Alignment Options:** Similar to Step 2, with default settings.
- STEP 4 - Submit your job:** Includes a checkbox 'Be notified by email' and a 'Submit' button. An arrow points to the 'Submit' button.

ClustalW: The Alignment Report

The screenshot shows the 'Alignment' report. It displays a ClustalW 2.0.10 multiple sequence alignment of four sequences:

- g118895923 ref|NP_571479.1|
- g112248117 db|BA024244.1|
- g111347790 ref|NP_474905.1|
- g11980379 db|AAC95182.1|

 The alignment shows gaps (indicated by dashes) and asterisks (*) indicating positions where all sequences have identical residues.

ClustalW: The Alignment Report

CONSENSUS SYMBOLS:

An alignment will display by default the following symbols denoting the degree of conservation observed in each column:

- *** means that the residues or nucleotides in that column are identical in all sequences in the alignment.
- ** means that conserved substitutions have been observed, according to the COLOUR table above.
- * means that semi-conserved substitutions are observed.

ClustalW: The Alignment Report

This screenshot is identical to the one above, but includes an arrow pointing to the 'Show Colors' button at the top of the alignment report, indicating that the alignment is color-coded based on the consensus symbols.

Alignment

Hide Colors (New Alignment File)

CLUSTAL 2.0.10 Multiple sequence alignment

```

#11889223 ref|NP_571479.1| MRFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGVGGV 50
#11224911 db|BA02424.1| MRFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 50
#11249117 db|BA02424.1| MRFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 50
#13980379 g|AAC95182.1| MRFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 50
#11889223 ref|NP_571479.1| TLKADQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 100
#11224911 db|BA02424.1| TLKADQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWV 100
#11249117 db|BA02424.1| TLKADQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWV 100
#13980379 g|AAC95182.1| TLKADQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWV 100
#11889223 ref|NP_571479.1| LVTFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 150
#11224911 db|BA02424.1| LVTFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 150
#11249117 db|BA02424.1| LVTFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 150
#13980379 g|AAC95182.1| LVTFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 150
#11889223 ref|NP_571479.1| LQVQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 200
#11224911 db|BA02424.1| LQVQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 200
#11249117 db|BA02424.1| LQVQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 200
#13980379 g|AAC95182.1| LQVQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 200
#11889223 ref|NP_571479.1| EPVQLTALVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWV 250
#11224911 db|BA02424.1| EPVQLTALVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWV 250
#11249117 db|BA02424.1| EPVQLTALVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWV 250
#13980379 g|AAC95182.1| EPVQLTALVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWV 250
#11889223 ref|NP_571479.1| LAKVQDSGGG-----262
#11224911 db|BA02424.1| LAKVQDSGGG-----262
#11249117 db|BA02424.1| LAKVQDSGGG-----262
#13980379 g|AAC95182.1| LAKVQDSGGG-----262
*****

```

ClustalW: The Alignment Report

• COLOUR

Show Colors
A button labeled 'Show Colors' will be displayed in the Alignment section of results page. If you press this button the alignment will be shown in color according to the table below.
NOTE: This option only works when you have chosen ALN or DCG as the output format.

HYPERLINK	RED	Small (small hydrophobic (not aromatic) *)
DE	BLUE	Acidic
IK	MAGENTA	Basic
STYCHING	GREEN	Hydroxy + Amino + Basic + Q
Others	Grey	

Figure 2. A color alignment report is generated for the alignment of multiple sequences. The alignment is shown in color according to the legend in the table below.

ClustalW: Jalview

ClustalW2 Results

Alignment | Result Summary | Submission Details | Submit Another Job

Multiple Sequence Alignment

```

#11889223 ref|NP_571479.1| MRFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 50
#11224911 db|BA02424.1| MRFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 50
#11249117 db|BA02424.1| MRFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 50
#13980379 g|AAC95182.1| MRFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 50
#11889223 ref|NP_571479.1| TLKADQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 100
#11224911 db|BA02424.1| TLKADQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 100
#11249117 db|BA02424.1| TLKADQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 100
#13980379 g|AAC95182.1| TLKADQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 100
#11889223 ref|NP_571479.1| LVTFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 150
#11224911 db|BA02424.1| LVTFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 150
#11249117 db|BA02424.1| LVTFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 150
#13980379 g|AAC95182.1| LVTFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 150
#11889223 ref|NP_571479.1| LQVQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 200
#11224911 db|BA02424.1| LQVQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 200
#11249117 db|BA02424.1| LQVQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 200
#13980379 g|AAC95182.1| LQVQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 200
#11889223 ref|NP_571479.1| EPVQLTALVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWV 250
#11224911 db|BA02424.1| EPVQLTALVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWV 250
#11249117 db|BA02424.1| EPVQLTALVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWV 250
#13980379 g|AAC95182.1| EPVQLTALVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWV 250
#11889223 ref|NP_571479.1| LAKVQDSGGG-----262
#11224911 db|BA02424.1| LAKVQDSGGG-----262
#11249117 db|BA02424.1| LAKVQDSGGG-----262
#13980379 g|AAC95182.1| LAKVQDSGGG-----262
*****

```

ClustalW: Jalview

EBI > Tools > Multiple Sequence Alignment > ClustalW

ClustalW2 Results

Alignments | Result Summary | Guide Tree | Submission Details | Submit Another Job

Result files

Input Sequences
Multiple Sequence Alignment
Phylogeny

clustalw2-i20110221-201650-0418-22091680-oy-input
clustalw2-i20110221-201650-0418-22091680-oy-output
Alignments in CLUSTAL format
clustalw2-i20110221-201650-0418-22091680-oy-clustalw
Guide Tree
clustalw2-i20110221-201650-0418-22091680-oy.dnd

Start Jalview

Scores Table

Seq#	Name	Length	Seq#	Name	Length	Score
1	g 3980379 g AAC95182.1	212	2	g 17136790 ref NP_476905.1	312	62.0
1	g 3980379 g AAC95182.1	212	3	g 18856223 ref NP_571479.1	261	62.0
1	g 3980379 g AAC95182.1	212	4	g 12249117 dbj BA02424.1	262	63.0
2	g 17136790 ref NP_476905.1	312	3	g 18856223 ref NP_571479.1	261	67.0
2	g 17136790 ref NP_476905.1	312	4	g 12249117 dbj BA02424.1	262	67.0
3	g 18856223 ref NP_571479.1	261	4	g 12249117 dbj BA02424.1	262	67.0

ClustalW: Jalview

```

#11889223 ref|NP_571479.1| MRFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 50
#11224911 db|BA02424.1| MRFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 50
#11249117 db|BA02424.1| MRFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 50
#13980379 g|AAC95182.1| MRFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 50
#11889223 ref|NP_571479.1| TLKADQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 100
#11224911 db|BA02424.1| TLKADQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 100
#11249117 db|BA02424.1| TLKADQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 100
#13980379 g|AAC95182.1| TLKADQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 100
#11889223 ref|NP_571479.1| LVTFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 150
#11224911 db|BA02424.1| LVTFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 150
#11249117 db|BA02424.1| LVTFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 150
#13980379 g|AAC95182.1| LVTFAKAVQVLLKPVVLAALKLTETAKVWVGGVGGVGGVGGVGGV 150
#11889223 ref|NP_571479.1| LQVQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 200
#11224911 db|BA02424.1| LQVQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 200
#11249117 db|BA02424.1| LQVQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 200
#13980379 g|AAC95182.1| LQVQVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWVWVWV 200
#11889223 ref|NP_571479.1| EPVQLTALVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWV 250
#11224911 db|BA02424.1| EPVQLTALVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWV 250
#11249117 db|BA02424.1| EPVQLTALVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWV 250
#13980379 g|AAC95182.1| EPVQLTALVDFVFCVGLKAVLWVWVWVWVWVWVWVWVWVWVWV 250
#11889223 ref|NP_571479.1| LAKVQDSGGG-----262
#11224911 db|BA02424.1| LAKVQDSGGG-----262
#11249117 db|BA02424.1| LAKVQDSGGG-----262
#13980379 g|AAC95182.1| LAKVQDSGGG-----262
*****

```

Conservation

Quality

Conservation

ClustalW: creating a phylogenetic tree

- ClustalW can only assemble a phylogenetic tree from multiple sequence alignments generated by ClustalW

ClustalW: Making a Phylogenetic Tree

EMBL-EBI EBI Search All Databases Enter Text Here Go Reset Advanced Search Give us Feedback

Databases Tools EBI Groups Training Industry About Us Help

EBI > Tools > Multiple Sequence Alignment > ClustalW2

ClustalW2 - Multiple Sequence Alignment

ClustalW2 is a general purpose multiple sequence alignment program for DNA or proteins.

Use this tool

STEP 1 - Enter your input sequences
Enter or paste a set of (Protein) sequences in any supported format:

Or, upload a file: no file selected

STEP 2 - Set your Pairwise Alignment Options
Alignment Type: Slow Fast
The default settings will suit 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.)

STEP 3 - Set your Multiple Sequence Alignment Options
The default settings will suit 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.)

STEP 4 - Submit your job
 Be notified by email (Tick this box if you want to be notified by email when the results are available)

ClustalW: Making a Phylogenetic Tree

EMBL-EBI EBI Search All Databases Enter Text Here Go Reset Advanced Search Give us Feedback

Databases Tools EBI Groups Training Industry About Us Help

EBI > Tools > Phylogeny

Phylogeny

ClustalW2 Phylogenetic tree generation using the ClustalW2 program.

EMBOSS phylogeny tools are available at the EBI via [SOAPLAB](#).

The tools described on this page are provided using our [new bioinformatics analysis tools framework](#). If you have any feedback or encounter any issues please let us know via [EBI support](#).

ClustalW: Making a Phylogenetic Tree

EMBL-EBI EBI Search All Databases Enter Text Here Go Reset Advanced Search Give us Feedback

Databases Tools EBI Groups Training Industry About Us Help

EBI > Tools > Phylogeny > ClustalW2

ClustalW2 - Phylogeny

Commonly used phylogenetic tree generation methods provided by the ClustalW2 program.

Use this tool

STEP 1 - Enter your multiple sequence alignment
Enter or paste a multiple sequence alignment in any supported format:

Or, upload a file: no file selected

STEP 2 - Set your Phylogeny options
TREE FORMAT: DISTANCE CORRECTION: EXCLUDE GAPS: CLUSTERING METHOD: P.I.M.:

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)

ClustalW: Making a Phylogenetic Tree

EMBL-EBI EBI Search All Databases Enter Text Here Go Reset Advanced Search Give us Feedback

Databases Tools EBI Groups Training Industry About Us Help

EBI > Tools > Phylogeny > ClustalW2

ClustalW2 - Phylogeny

Commonly used phylogenetic tree generation methods provided by the ClustalW2 program.

Use this tool

STEP 1 - Enter your multiple sequence alignment
Enter or paste a multiple sequence alignment in any supported format:
CLUSTAL 2.1 multiple sequence alignment

Must use Clustal aligned sequences & include "Clustal" line

```
gi|18859223|ref|NP_571479.1| MFEARLVQGSILKVKLEAKDLITEACWDVSSGSLQSMDSHSHVSLVQL 50
gi|12249117|db|JBAR20424.1| MFEARLVQGSVLRVLEAKDLITEACWDLGCGSLQSMDSHSHVSLVQL 50
gi|17136790|ref|NP_476905.1| MFEARLQATILKILDAIKDLINATTCDCSDGCLQAMDNHSHVSLVSL 50
gi|3980379|gb|AACQ3182.1| MLELRVQGSLLKVKLEAKDLINDANFDCTTFCFLQAMDSHSHVSLVSL 50
```

Or, upload a file: no file selected

STEP 2 - Set your Phylogeny options
TREE FORMAT: DISTANCE CORRECTION: EXCLUDE GAPS: CLUSTERING METHOD: P.I.M.:

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)

ClustalW: Making a Phylogenetic Tree

EMBL-EBI EBI Search All Databases Enter Text Here Go Reset Advanced Search Give us Feedback

Databases Tools EBI Groups Training Industry About Us Help

EBI > Tools > Phylogeny > ClustalW2

ClustalW2 - Phylogeny

Commonly used phylogenetic tree generation methods provided by the ClustalW2 program.

Use this tool

STEP 1 - Enter your multiple sequence alignment
Enter or paste a multiple sequence alignment in any supported format:
CLUSTAL 2.1 multiple sequence alignment

```
gi|18859223|ref|NP_571479.1| MFEARLVQGSILKVKLEAKDLITEACWDVSSGSLQSMDSHSHVSLVQL 50
gi|12249117|db|JBAR20424.1| MFEARLVQGSVLRVLEAKDLITEACWDLGCGSLQSMDSHSHVSLVQL 50
gi|17136790|ref|NP_476905.1| MFEARLQATILKILDAIKDLINATTCDCSDGCLQAMDNHSHVSLVSL 50
gi|3980379|gb|AACQ3182.1| MLELRVQGSLLKVKLEAKDLINDANFDCTTFCFLQAMDSHSHVSLVSL 50
```

Or, upload a file: no file selected

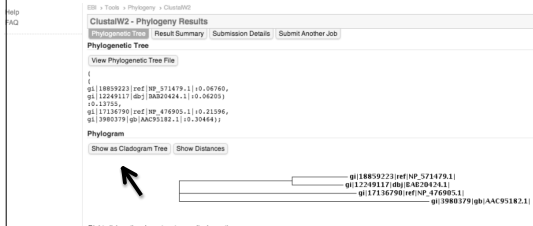
STEP 2 - Set your Phylogeny options
TREE FORMAT: DISTANCE CORRECTION: EXCLUDE GAPS: CLUSTERING METHOD: P.I.M.:

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)

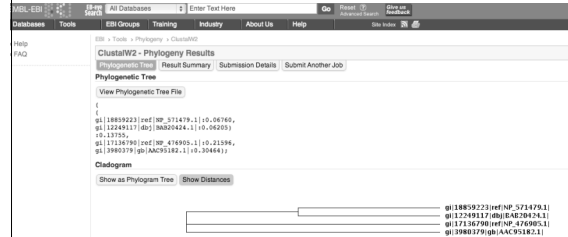
ClustalW: Making a Phylogenetic Tree

- Phylogram: tree that provides an estimate of a phylogeny, where the branch lengths are proportional to the amount of evolutionary divergence.
- Cladogram: tree that provides an estimate of phylogeny where the branches are of equal length: therefore cladograms indicate common ancestry, but do not reflect amount of evolutionary time separating sequences.

ClustalW: The Phylogram Tree



ClustalW: The Cladogram Tree



ClustalW: Practice Weblem

- Align the refseq protein sequences of *Arabidopsis thaliana* RDR6, *Schizosaccharomyces pombe* Rdp1, *Caenorhabditis elegans* rrf-2, *Caenorhabditis elegans* rrf-3, and *Aspergillus fumigatus* Sad-1 using ClustalW. Print out this alignment.
- Generate and print out a phylogram and cladogram using this alignment.
- Given what you know about the evolutionary relationships between these organisms, does the phylogram and cladogram make sense? Why or why not?