

This workshop will show you how to

- 1) Perform sequence alignment using clustal omega.
- 2) Build UPGMA and NJ tree using R.
- 3) Construction maximum likelihood tree using IQ-tree2.
- 4) Alignment and phylogeny viewing software.

Download from NTU COOL: [data_m2_molecular_evolution.zip](#)

FASTA file format

Number of sequences =
Number of ">"

FASTA - Interleaved

>Sequence name. This can be as long or as short as you like

ATTGGACGTCTAGCTAGCAG

CATGACTAGCATCTCTCACA

CTACATATCTACTCAATCTA

>Second sequence

ATCGTCGACTGATGACTGAC

ATCAGTCATGCATGCATGCT

FASTA - Sequential

>Sequence name. Sequential format put all sequences in one line

ATTGGACGTCTAGCTAGCAGCATGACTAGCATCTCTCACACTACATATCTACTCAATCTA.....

>Second sequence

TGACGCATGTACTGACTGACTGACTGACTGACTGACTGACTAGCTAGCGACATGCTGCATGCCT.....

Part 1 - Sequence alignment

1. Go to

<http://www.clustal.org/omega/#Download>

2. Hover over

64-bit Linux binary

OR

Mac binary

3. Right click and

“Copy link address”

Download Clustal Omega

Source code

- [Source code .tar.gz \(1.2.4\)](#)

Installation instructions are available in the [INSTALL file](#), once the archive is extracted. This release is intended for use with UNIX systems. Installation on Macs is also possible. Windows users should download the binary release below.

Precompiled binaries

- [Standalone 64-bit Linux binary \(1.2.4\)](#)
- [Standalone 32-bit Linux binary \(1.2.4\)](#)
- [Standalone 64-bit FreeBSD binary \(1.2.4\)](#)
- [Standalone Mac binary \(PowerPC/64-bit\)](#)

Windows users should extract the zip file and rename the downloaded file to **clustalo** and place in the location of their choice. This file may need to be made executable.

Questionnaire

Please help us develop Clustal Omega by filling out this questionnaire. All optional, can be answered in a minute)

Distributions

A [Clustal Omega package](#) is available for Geany IDE. A [Debian package](#) is available as part of [Debian Med](#). A [wrapper](#) is available for EMBOSS.

Benchmarks

We constructed two benchmark data-sets by blending existing structural alignment reference data-sets with homologous unaligned [Pfam](#) sequences.

- [HomFam](#): 94 [HomStrad](#) single domain reference alignments with homologous Pfam sequences (version 25)
- [BaliFam](#): homologous Pfam sequences (version 26) for [BaliBASE 3](#) reference alignments
- [small trees](#)

Download program in WSL/shell

In WSL/shell, use the following commands to download the program

Paste the link from previous page

\$ wget http://www.clustal.org/omega/clustalo-1.2.4-Ubuntu-x86_64

```
cia@PC-09:~$ ls
cia@PC-09:~$ wget http://www.clustal.org/omega/clustalo-1.2.4-Ubuntu-x86_64
--2021-10-19 15:13:09-- http://www.clustal.org/omega/clustalo-1.2.4-Ubuntu-x86_64
Resolving www.clustal.org (www.clustal.org)... 137.43.93.206, 91.136.48.10, 91.136.49.10, ...
Connecting to www.clustal.org (www.clustal.org)|137.43.93.206|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 4921774 (4.7M)
Saving to: 'clustalo-1.2.4-Ubuntu-x86_64'

clustalo-1.2.4-Ubuntu-x86_64 100%[=====>] 4.69M 1.67MB/s in 2.8s
2021-10-19 15:13:14 (1.67 MB/s) - 'clustalo-1.2.4-Ubuntu-x86_64' saved [4921774/4921774]

cia@PC-09:~$ ls
clustalo-1.2.4-Ubuntu-x86_64
```

Download program in Mac/zsh

In Mac (since zsh doesn't have wget by default), use the following commands to download the program

Paste the link from previous page

```
$ curl http://www.clustal.org/omega/clustal-omega-1.2.3-macosx  
-o clustalo
```

```
[steven@localhost ~]$ curl http://www.clustal.org/omega/clustal-omega-1.2.3-macosx -o clustalo
```

% Total	% Received	% Xferd	Average Speed	Time	Time	Time	Current				
			Dload	Upload	Total	Spent	Left	Speed			
100	1907k	100	1907k	0	0	204k	0	0:00:09	0:00:09	--:--:--	232k

Change permission

Add x (execute) to everybody

```
$ chmod +x clustalo-1.2.4-Ubuntu-x86_64
```

```
cia@PC-09:~$ chmod +x clustalo-1.2.4-Ubuntu-x86_64
```

```
cia@PC-09:~$ ls -l
```

```
total 4808
```

```
-rwxrwxr-x 1 cia cia 4921774 Dec 21 2016 clustalo-1.2.4-Ubuntu-x86_64
```

```
cia@PC-09:~$ ./clustalo-1.2.4-Ubuntu-x86_64
```

```
FATAL: No sequence input was provided. For more information try: --help
```

```
cia@PC-09:~$ ./clustalo-1.2.4-Ubuntu-x86_64 --help
```

```
Clustal Omega - 1.2.4 (AndreaGiacomo)
```

```
If you like Clustal-Omega please cite:
```

Execute the program with correct permission

```
$ ./clustalo-1.2.4-Ubuntu-x86_64  
$ ./clustalo-1.2.4-Ubuntu-x86_64 --help
```

```
cia@PC-09:~$ chmod 764 clustalo-1.2.4-Ubuntu-x86_64  
cia@PC-09:~$ ls -l  
total 4808  
-rwxrw-r-- 1 cia cia 4921774 Dec 21 2016 clustalo-1.2.4-Ubuntu-x86_64  
cia@PC-09:~$ ./clustalo-1.2.4-Ubuntu-x86_64  
FATAL: No sequence input was provided. For more information try: --help  
cia@PC-09:~$ ./clustalo-1.2.4-Ubuntu-x86_64 --help  
Clustal Omega - 1.2.4 (AndreaGiacomo)
```

If you like Clustal-Omega please cite:

Running alignment program with raw data

```
$ ./clustalo-1.2.4-Ubuntu-x86_64 -i INFILE -o OUTFILE
```

```
$ ./clustalo-1.2.4-Ubuntu-x86_64 -i SARS_CoV_2_human_only.fasta -o  
SARS_CoV2.aln # COVID_19 dataset
```

```
$ ./clustalo-1.2.4-Ubuntu-x86_64 -i coronavirus_others.fasta -o  
coronavirus.aln # corona virus dataset
```

```
$ ./clustalo-1.2.4-Ubuntu-x86_64 -i  
COXII_Cytochrome_c_oxidase_subunit_II.fasta -o COXII.aln # Cytochrome C  
dataset for animals
```

```
$ ./clustalo-1.2.4-Ubuntu-x86_64 -i three_kingdoms.fasta -o  
three_kingdoms.aln # AA dataset for 3 kingdoms
```


Part 2 - Distance method with R

```
> library(ape)
# Install the package as needed
# install.packages("ape")

# read.FASTA: Read alignment file.
# Make sure you are in the right working directory.
> aln <- read.FASTA("SARS_CoV2.aln")

# View the alignment in R - Generally it's a bad idea
> alview(aln) # Horrible way to view an alignment
> image(aln)  # Slightly better, but still pretty bad.
# Solution. Use a third party program.
```

Part 2 - Distance method with R

```
# Calculate the pairwise distance
```

```
> data_dist <- dist.dna(aln)
```

```
# UPGMA
```

```
> tree_upgma <- as.phylo(hclust(data_dist, method="average"))
```

```
> plot(tree_upgma) # view tree
```

```
> write.tree(tree_upgma, file="tree_upgma.nexus") # save phylogeny
```

```
# Neighbour joining
```

```
> tree_nj <- nj(data_dist) # Basic NJ method
```

```
> tree_bionj <- bionj(data_dist) # Improved NJ method
```

```
> write.tree(tree_bionj, file="tree_bionj.nexus") # save phylogeny
```

Part 3 - Maximum likelihood with IQ-Tree (WSL/Shell)

1. Go to

<http://www.iqtree.org/#download>

2. Under COVID-19 release

2.1.3. Hover over

64-bit multicore Linux

OR

64-bit multicore macOS

3. Right click and

“Copy link address”



Download program in WSL/shell

WSL/shell. Paste the link from previous page

\$ wget

<https://github.com/iqtree/iqtree2/releases/download/v2.1.3/iqtree-2.1.3-Linux.tar.gz>

```
cia@PC-08:~$ wget https://github.com/iqtree/iqtree2/releases/download/v2.1.3/iqtree-2.1.3-Linux.tar.gz
--2021-11-04 15:10:16-- https://github.com/iqtree/iqtree2/releases/download/v2.1.3/iqtree-2.1.3-Linux.tar.gz
Resolving github.com (github.com)... 52.192.72.89
Connecting to github.com (github.com)|52.192.72.89|:443... connected.
HTTP request sent, awaiting response... 302 Found
Location: https://github-releases.githubusercontent.com/272864622/2181d500-a2b1-11eb-8ab0-26e1ea54673d?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIWNJYAX4CSVEH53A%2F20211104%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20211104T151016Z&X-Amz-Expires=300&X-Amz-Signature=f67ef120d4102af91cd9dfc10f0969fdae2363a44541729b394e64a5e319bf8f&X-Amz-SignedHeaders=host&actor_id=0&key_id=0&repo_id=272864622&response-content-disposition=attachment%3B%20filename%3D%20iqtree-2.1.3-Linux.tar.gz

```

Download program in Mac/zsh

```
# Mac/zsh
```

```
# Make sure you have -L here!
```

```
$ curl -L
```

```
https://github.com/iqtree/iqtree2/releases/download/v2.1.3/iqtree-2.1.3-MacOSX.zip -o iqtree_mac.zip
```

```
cia@PC-08:~$ curl -L https://github.com/iqtree/iqtree2/releases/download/v2.1.3/iqtree-2.1.3-Linux.tar.gz -o iqtree.tar.gz
```

% Total	% Received	% Xferd	Average	Speed	Time	Time	Time	Current
			Dload	Upload	Total	Spent	Left	Speed
100	633	100	633	0	0	4164	0	--:--:-- --:--:-- --:--:-- 4164
100	4572k	100	4572k	0	0	6655k	0	--:--:-- --:--:-- --:--:-- 11.6M

Uncompress file (tar.gz)

```
# *.tar.gz is one of the most common compression formats.  
# -x extract      # -v verbose      # -f filename  
$ tar -xvf iqtrees-2.1.3-Linux.tar.gz
```

```
cia@PC-08:~$ ls  
iqtrees-2.1.3-Linux.tar.gz  
cia@PC-08:~$ tar -xvf iqtrees-2.1.3-Linux.tar.gz  
iqtrees-2.1.3-Linux/bin/  
iqtrees-2.1.3-Linux/bin/iqtrees2  
iqtrees-2.1.3-Linux/example.phy  
iqtrees-2.1.3-Linux/example.nex  
iqtrees-2.1.3-Linux/models.nex  
iqtrees-2.1.3-Linux/example.cf  
cia@PC-08:~$
```

Uncompress file (zip)

Mac/zsh. Unteste on mac!

\$ unzip iqtrees_mac.zip

```
cia@PC-08:~$ unzip iqtrees.zip
Archive:  iqtrees.zip
  creating: iqtrees-2.1.3-MacOSX/bin/
  inflating: iqtrees-2.1.3-MacOSX/bin/iqtrees2
  inflating: iqtrees-2.1.3-MacOSX/example.nex
  inflating: iqtrees-2.1.3-MacOSX/example.cf
  inflating: iqtrees-2.1.3-MacOSX/models.nex
  inflating: iqtrees-2.1.3-MacOSX/example.phy
cia@PC-08:~$
```

Note: unzip is NOT installed in WSL (and many other shells) by default

Locate the executable

```
$ ls
```

```
# You should see one of these two folders
```

```
# iqtree-2.1.3-Linux/ OR iqtree-2.1.3-MacOSX/
```

```
# go into the folder and have a look inside
```

```
$ cd iqtree-2.1.3-Linux/
```

```
$ ls bin/
```

```
# Look under the bin folder, you should see "iqtree2"
```

```
[bioinfo@localhost ~]$ ls iqtree-2.1.3-Linux  
bin  example.cf  example.nex  example.phy  models.nex  
[bioinfo@localhost ~]$ cd iqtree-2.1.3-Linux/  
[bioinfo@localhost iqtree-2.1.3-Linux]$ ls bin  
iqtree2  
[bioinfo@localhost iqtree-2.1.3-Linux]$
```


Locate the executable

```
# check file permissions. Make sure it got 'x' (executable)
$ ls -l bin/iqtree2
```

```
# In this case, it got 'x' already
# chmod +x iqtree2 # If needed.
```

```
[bioinfo@localhost iqtree-2.1.3-Linux]$ ls -l bin
total 10304
-rwxr-xr-x 1 bioinfo bioinfo 10548168 Apr 21  2021 iqtree2
[bioinfo@localhost iqtree-2.1.3-Linux]$
```

Build the phylogeny

```
# NOTE: Think about the location of two files.  
# They might not be in the same folder.  
# Where is the alignment file?  
# Where is the executable?  
# They need to be able to see each other!
```

```
# Potentially failed example  
$ ./iqtree2 -s ALIGNMENT_FILE
```

```
# Note. This probably will NOT work, unless you sort out the  
location of these files first.
```

Build the phylogeny

Assuming the following folder/file structure

/home/cia/

| - **ALIGNMENT_FILE**

| - iqtree-2.1.3-Linux/

 | - bin/

 | - **iqtree2**

Where are you?

\$ pwd

iqtree2 and ALIGNMENT_FILE are NOT in the same folder, they can "see" each other.

Build the phylogeny

Copy the alignment file to the bin folder. Build the phylogeny in the bin file

```
$ cp ALIGNMENT_FILE iqtrees-2.1.3-Linux/bin/
```

```
$ cd iqtrees-2.1.3-Linux/bin/
```

```
$ ./iqtrees2 -s ALIGNMENT_FILE
```

You are here -->

```
/home/cia/  
|- ALIGNMENT_FILE  
|- iqtrees-2.1.3-Linux/  
    |- bin/  
        |- iqtrees2  
        |- ALIGNMENT_FILE
```

Build the phylogeny

Build the phylogeny in the bin file, with relative path to the alignment file

```
$ cd iqtrees-2.1.3-Linux/bin/
```

```
$ ./iqtrees2 -s ../../ALIGNMENT_FILE
```

```
# ../ means 1 level up.
```

```
# ../../ means 2 levels up.
```

```
# ./iqtrees2 -s PATH_TO_ALIGNMENT_FILE
```

You are here -->

```
/home/cia/  
|- ALIGNMENT_FILE  
|- iqtrees-2.1.3-Linux/  
    |- bin/  
        |- iqtrees2
```

Build the phylogeny

Build the phylogeny in same folder as the alignment file, with relative path to iqtree2 executable

```
$ ./iqtree-2.1.3-Linux/bin/iqtree2 -s ALIGNMENT_FILE
```

You are here -->

```
/home/cia/  
|- ALIGNMENT_FILE  
|- iqtree-2.1.3-Linux/  
    |- bin/  
        |- iqtree2
```

Part 4 - Alignment viewer

Cross platform

Aliview - <http://www.ormbunkar.se/aliview>

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

Browser

NCBI - <https://www.ncbi.nlm.nih.gov/projects/msaviewer/>

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

Aliview

Install instructions and Downloads:

[Version history here \(version_history.txt\)](#)

[Some demo datasets can be found here](#)

Mac OS X (all versions) [download here](#)

The program is distributed as an Application (**AliView-app.zip**), that only needs to be dropped in your App-folder (sometimes you need to unpack it manually)

OBS! On Mavericks or Sierra you might get an error message "This application is damaged and can't be opened. You should move it to the Trash." - The fix is to need to change to be able to download software from other places than the App-store.

How to solve (on Mavericks): <http://answers.uchicago.edu/page.php?id=25481>

How to solve (on Sierra): <https://www.tekrevue.com/tip/gatekeeper-macos-sierra>

Windows (all versions) [download here](#)

Download and run **AliView-Setup.exe**

- the setup program will default installation to directory "C:\Program Files (x86)\AliView"

Alternatively you can download the program **AliView.exe** or **aliview.jar** from sub-directory /without_installer_version/

NOTE: some virus scanners might block the installation and need to be temporarily disabled.

Linux (all versions) [download here](#)

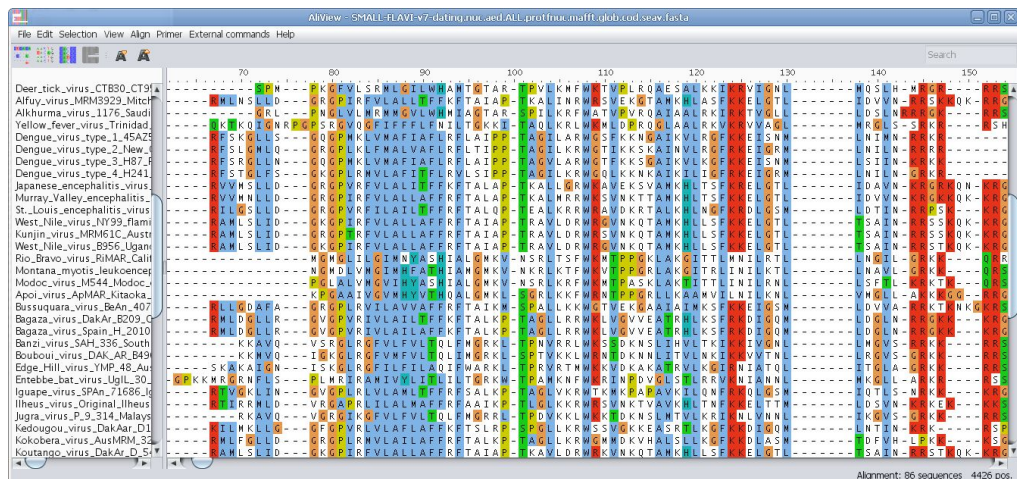
The simplest install is to download the file: **aliview.install.run**

(this is a executable archive that will copy the files to the /usr/bin/ and /usr/share/aliview/, and will also install a ".desktop" link to the start-menu on compatible systems.)

- after downloading you will likely need to change the execution rights of the install file:

```
chmod +x aliview.install.run
```


Aliview



Phylogeny viewer

Cross platform

Figtree: Official website <http://tree.bio.ed.ac.uk/software/figtree/>

Download link: <https://github.com/rambaut/figtree/releases>

Browser


ETE: <http://etetoolkit.org/treeview/>


FigTree v1.4.4


Latest


Fixed issues with FigTree on Java 9 (on any platform) and Mac OS X High Sierra and later.


▼ Assets 5

 [FigTree.v1.4.4.dmg](#)

 [FigTree.v1.4.4.zip](#)

 [FigTree_v1.4.4.tgz](#)

 [Source code](#) (zip)

 [Source code](#) (tar.gz)

Phylogeny viewer

