

MSA (Multiple sequence alignment)

The first step is to get the needed sequences.

We will get the P53 nucleotide sequences using NCBI for the following accession numbers.

AB021961.1 U50395.1 D49825.1 AB082923.1

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search Advanced Help

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.coronavirus.gov>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.

Species clear Summary Sort by Default order

Animals (4)
Customize ...

Molecule types
mRNA (4)
Customize ...

Source databases
INSDC (GenBank) (4)
Customize ...

Sequence Type
Nucleotide (4)

Sequence length
Custom range...

Release date
Custom range...

Revision date
Custom range...

Clear all
Show additional filters

Items: 4
Selected: 4
Filters activated: Animals. [Clear all](#)

1. [Mus musculus mutant p53 mRNA, complete cds](#)
1,429 bp linear mRNA
Accession: AB021961.1 GI: 5421849
[Protein](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

2. [Cricetulus griseus wild type tumor suppressor P53 \(p53\) mRNA, complete cds](#)
1,829 bp linear mRNA
Accession: U50395.1 GI: 1842229
[Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

3. [Bos primigenius p53 mRNA, partial cds](#)
1,127 bp linear mRNA
Accession: D49825.1 GI: 736699
[Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

4. [Homo sapiens mRNA for P53, complete cds](#)
2,451 bp linear mRNA
Accession: AB082923.1 GI: 23491728
[Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

Summary Sort by Default order

Send to: Filters: [Manage Filters](#)

Complete Record
Coding Sequences
Gene Features

Choose Destination
File Clipboard
Collections Analysis Tool

Download 4 items.
Format
FASTA
Sort by
Default order
Show GI
Create File

Recent activity
Turn Off Clear

(p53) AND "Homo sapiens"[porgn] AND (animals[filter]) (6823) Nucleotide

p53 AND (animals[filter]) (41037) Nucleotide

p53 (95238) Nucleotide

Anti-H5N1 virus flavonoids from Capparis sinaica Veill. PubMed

Anti-H5N1 virus new diglyceride ester from the Red Sea grass Thalassodendron c PubMed

Open the downloaded fasta file and investigate each line then rename each sequence.

```
>Seq1_Homo_Sapiens AB082923.1 Homo sapiens mRNA for P53, complete cds
CGTGCTTTCCACGACGGTGACACGCTTCCCTGGATTGGCCAGACTGCCTTCCGGGTCAGTCCCATGGAGG
AGCCGAGTCAGATCCTAGCGTCGAGCCCCCTCTGAGTCAGGAAACATTTTCAGACCTATGGAACTACT
TCCTGAAAAACAACGTTCTGTCCCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCCGGACGAT
ATTGAACAATGGTTCACTGAAGACCCAGGTCCAGATGAAGCTCCCAGAATGCCAGAGGCTGCTCCCCGCG
TGGCCCCCTGCACCAGCAGCTCCTACACGGCGGGCCCCCTGCACCAGCCCCCTCTGGCCCCCTGTCATCTTC
TGTCCCTTCCCAGAAAACTACCAGGGCAGCTACGGTTTCCGTCTGGGCTTCTTGCACTTCTGGGACAGCC
AAGTCTGTGACTTGCACGTAATCCCCTGCCCTCAACAAGATGTTTTGCCAACTGGCCAAGACCTGCCCTG
TGCAGCTGTGGGTTGATTCCACACCCCCGCCCCGGCACCCGCGTCCGCGCCATGGCCATCTACAAGCAGTC
ACAGCACATGACGGAGGTTGTGAGGCGCTGCCCCCACCATGAGCGCTGCTCAGATAGCGATGGTCTGGCC
CCTCCTCAGCATCTTATCCGAGTGGAAGGAAATTTGCGTGTGGAGTATTTGGATGACAGAAACACTTTTC
GACATAGTGTGGTGGTGCCCTATGAGCGGCTGAGGTTGGCTCTGACTGTACCAATCCACTACAACTA
```

Now we are ready to perform MSA, There are many different tools with different algorithm. In this tutorial we will use MUSCLE

- 1- Open the following website <https://www.ebi.ac.uk/Tools/msa/muscle/>
- 2- Upload your fasta file or simply copy paste.
- 3- Choose output format
- 4- Press submit button

The screenshot displays the EBI MUSCLE web interface, which is organized into three main steps:

- STEP 1 - Enter your input sequences:** This section contains a text area where a multi-line DNA sequence has been pasted. Below the text area, there are links for "Use a example sequence", "Clear sequence", and "See more example inputs".
- STEP 2 - Set your Parameters:** This section features a dropdown menu for "OUTPUT FORMAT:" which is currently set to "ClustalW". Below the dropdown, a message states "The default settings will fulfill the needs of most users." and a button labeled "More options..." is provided with a note to click here for more settings.
- STEP 3 - Submit your job:** This final section includes a checkbox for "Be notified by email" and a prominent green "Submit" button.

Currently there are different output format available, the most widely used are ClustalW and Person/FASTA for more information click on the following line <https://www.ebi.ac.uk/seqdb/confluence/display/JDSAT/MUSCLE+Help+and+Documentation#MUSCLEHelpandDocumentation-format>

As Shown in the following figure the result will be generated and alignments tab will be shown by default.

Because we submit a DNA sequence you can see some letters have * because they are the same residue in all input sequences (conserved). While in the amino acid sequence case you can see *, : and .

"*" - the residues or nucleotides are identical in all sequences in the alignment.

":" - conserved substitutions are observed.

"." - semi-conserved substitutions have been observed.

Results for job muscle-I20200424-093237-0491-69283341-p2m

Alignments Result Summary Phylogenetic Tree Results Viewers Submission Details

Download Alignment File

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

```
Seq1_Homo_Sapiens      -----
Seq4_Bos_primigenius   -----
Seq3_Cricetulus_griseus -----TTGGGATCAGGGGGCACCTGGGATCCGGGTCCGGA---A
Seq2_Mus_Musculus      TTCCTGGNCTGTAGGTAGCGACTACAGTTAGGGGGCACCTAGCATTACGGCCCTCATCCT

Seq1_Homo_Sapiens      CGTGCTTTCCACGACGGTGACACGCTTCCTGGATT-----GGCCAGACTGCCTTCCG
Seq4_Bos_primigenius   -----
Seq3_Cricetulus_griseus TCTACTTCCCAGC-GGGTAACACGCTCCCCCTGAAGACCTGAAGCCTGGCTGACTTCTT
Seq2_Mus_Musculus      CCTCCTTCCCAGCAGGGTGTACGCTTCTCCGAAGA-----CTG

Seq1_Homo_Sapiens      GGTCACTGCCATGGAGGAGCCGAGTCAGATCCTAGCGTCGAGCCCCCTCTGAGTCAGGA
Seq4_Bos_primigenius   -----CCTCTGAGCCAGGA
Seq3_Cricetulus_griseus GGGTGCTGCCATGGAGGAGCCACAGTCAGACCTCAGCATCGAGCTCCCTCTGAGCCAGGA
Seq2_Mus_Musculus      GATGACTGCCATGGAGGAGTCACAGTCGGATATCAGCCTCGAGCTCCCTCTGAGCCAGGA
                          *****

Seq1_Homo_Sapiens      AACATTTTCAGACCTATGGAACTACTTCTTGAAACAACGTTCTGTC--CCCCTTGCCG
Seq4_Bos_primigenius   GACATTTTCCGACTTGTGGAACCTACTTCTTGAAATAACCTTCTGTC--CTCCGAGCTC
Seq3_Cricetulus_griseus GACATTTTCAGACCTGTGGAACCTACTTCTTCAAACAATGTTCTGTC--CACCTTGCCG
Seq2_Mus_Musculus      GACATTTTCAGGCTTATGGAACTACTTCTTCCAGAAGATATCTGCCATCACCTCACTG
                          *****

Seq1_Homo_Sapiens      T---CCCAAGCAATGGATGATTTGATGCTGTCCCCGACGATATTGAACAATGGTTCACT
Seq4_Bos_primigenius   T---CCGCACCCGTGGATGACCTGCTCCCGTACACAGATG---TTGCCACCTGGCT---
Seq3_Cricetulus_griseus TCATCCGATTCCATTGAAGAGCTGTTCCTGTGTCAGAGAATG---TTACAGGCTGGTT---
Seq2_Mus_Musculus      C-----ATGACGATCTGTTGCTGCCCCAGGATG---TTGAGGAGTTT---
                          * * * * *

Seq1_Homo_Sapiens      GAAGACCCAGGTCCAGATGAAGCTCCCAGAATG-----CCAGAGGCTGCTCCCCGC
Seq4_Bos_primigenius   --GGATGAATGTCCGAATGAAGCGCCCCAAATG-----CCAGAGCCTT-----
Seq3_Cricetulus_griseus --AGA---AGACTCAGGTGGAGCGCTCCAAGGGGTGGCAGCAGCAGCAGCTTCTACAGCA
Seq2_Mus_Musculus      --TGA---AGGCCAAGTGAAGCCCTCCGAGTG-----TCAGGAGCTCCTGCAGCA
                          * * * * *

Seq1_Homo_Sapiens      GTGGCCCTGCACCAGCAGCTCCTACACGGCGGCCCTGCACCAGCCCCCTCCTGGCCC
Seq4_Bos_primigenius   -----CTGCCCCAGCTGCCCGGCCACCAGCCACCCAGCACCAGCCACCTCCTGGCCC
Seq3_Cricetulus_griseus GAAGACCCTGTAACAGACTCCTGCACCGGTGGCCTCTGCACCAGCCACTCCCTGGCCC
Seq2_Mus_Musculus      CAGGACCCTGTACCGAGACCCTGGGCCAGTGGCCCTGCCCGAGCCACTCCATGGCCC
                          *** * * * * * * * * * * * * * * * *
```

Note how each sequence named in the output.

Now let's do it with command line

First, we need to install MUSCEL software:

You can do this using the following code:

```
sudo apt-get install muscle
```

Or you can go to [this](#) website and download the appropriate version.

To get all sequence we can use efetch command from entrize-direct tools

```
sudo apt install ncbi-entrez-direct
```

```
efetch -db nucleotide -format fasta -id AB021961.1,U50395.1,D49825.1,AB082923.1 >  
all.fasta
```

Using nano editor rename each sequence.

The next step is to run MUSCLE using the following basic command:

```
muscle -in sequence.fasta -out alignment.afa
```

as you can see it take a fasta file and output Fasta file and it accept DNA, RNA and protein sequences.

You can specify the input sequence type or it will be automatically detected using the -seqtype option

```
-seqtype protein
```

For this example out code will be

```
muscle -in sequence.fasta -out alignment.afa -seqtype protein
```

also, you can choose from different output format using the following options

-clw ==> clustalw format

-msf ==> MSF format

-html ==> HTML format

```
muscle -in sequence.fasta -out alignment.html -html
```

Thank you!