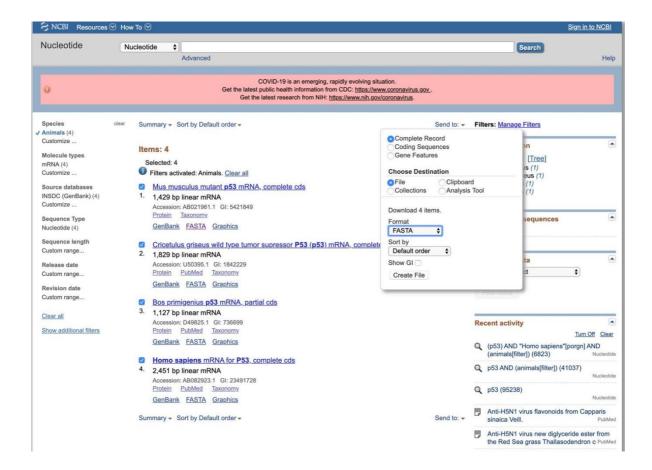
MSA (Multiple sequence alignment)

The first step is to get the needed sequences.

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

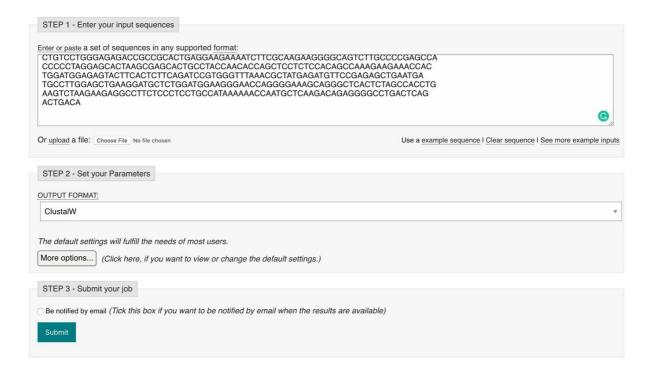
AB021961.1 U50395.1 D49825.1 AB082923.1



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

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



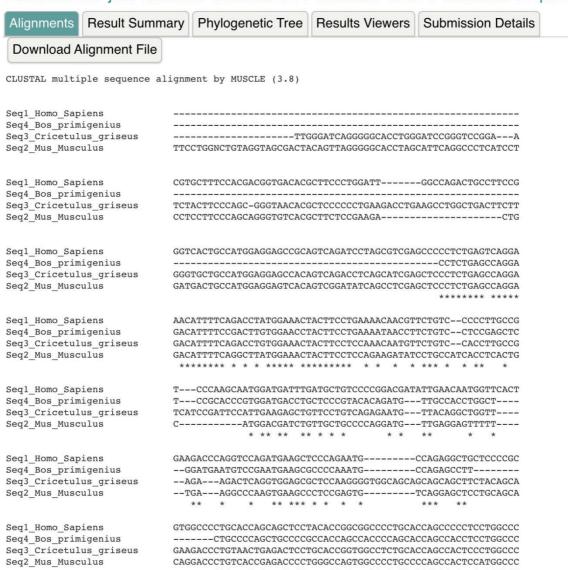
Currently there are different output format available, the most widely used are ClustalW and Person/FASTA for more information click on the following line https://www.ebi.ac.uk/seqdb/confluence/display/JDSAT/MUSCLE+Help+and+Documentation-format

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 \ast because they are the same residue in all input sequences (conserved). While in the amino acid sequence case you can see \ast , : 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



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 Person/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!