Nonsynonymous mutation and deletion analysis: A fast and accurate pipeline

Md. Shaminur Rahman

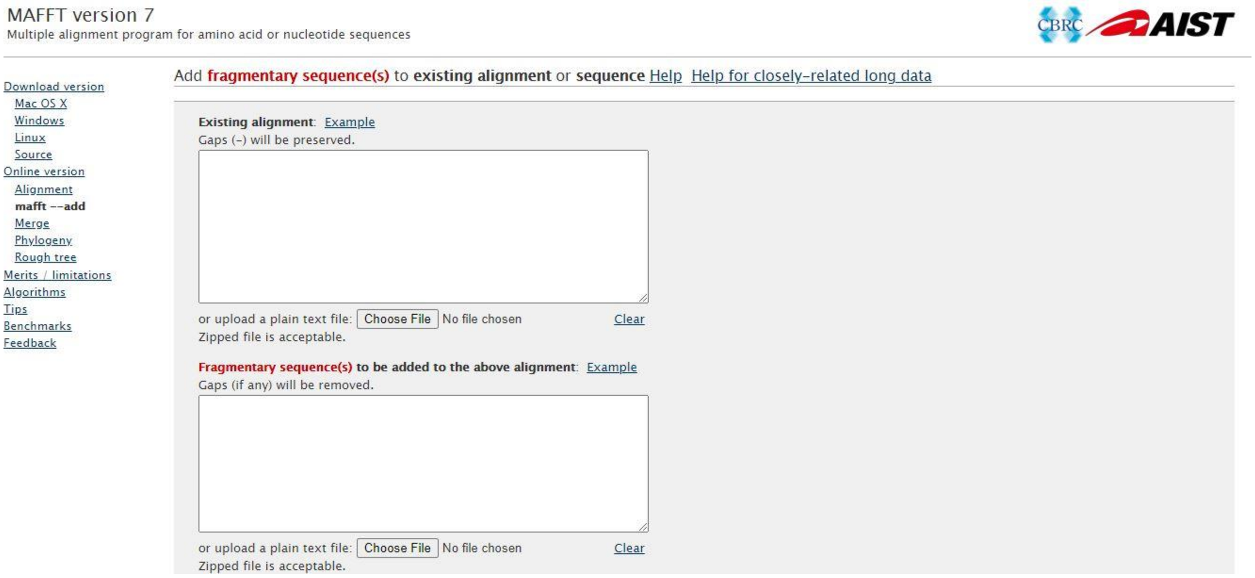
BS (Hon's) and MS, Department of Microbiology, University of Dhaka.

1. Pyfasta (https://github.com/brentp/pyfasta): For large dataset (more than 10,000 sequences), we need to differentiate the data into two or more dataset as MAFFT alignment tools has a maximum limit ≤ 10,000.

Usages (Linux):

```
pyfasta split -n 5 original.fasta  # split into 5 fasta files
```

2. MAFFT online alignment tools (https://mafft.cbrc.jp/alignment/server/)**:** MAFFT is a very user friendly multiple alignment program for amino acid or nucleotide sequences. For SARS-CoV-2, specialize version (https://mafft.cbrc.jp/alignment/server/add_fragments.html?frommanual) has been launched where the maximum limit was 10,000. Here, existing alignment or reference sequences can be selected in a menu and other menu also available for new alignment. Here, in my experiment, I have used first as reference and the second one were for aligning sequence upload.



MAFFT version 7

Multiple alignment program for amino acid or nucleotide sequences

Add fragmentary sequence(s) to existing alignment or sequence Help Help for closely-related long data

Existing alignment: Example

Gaps (-) will be preserved.

or upload a plain text file: Choose File No file chosen Clear

Zipped file is acceptable.

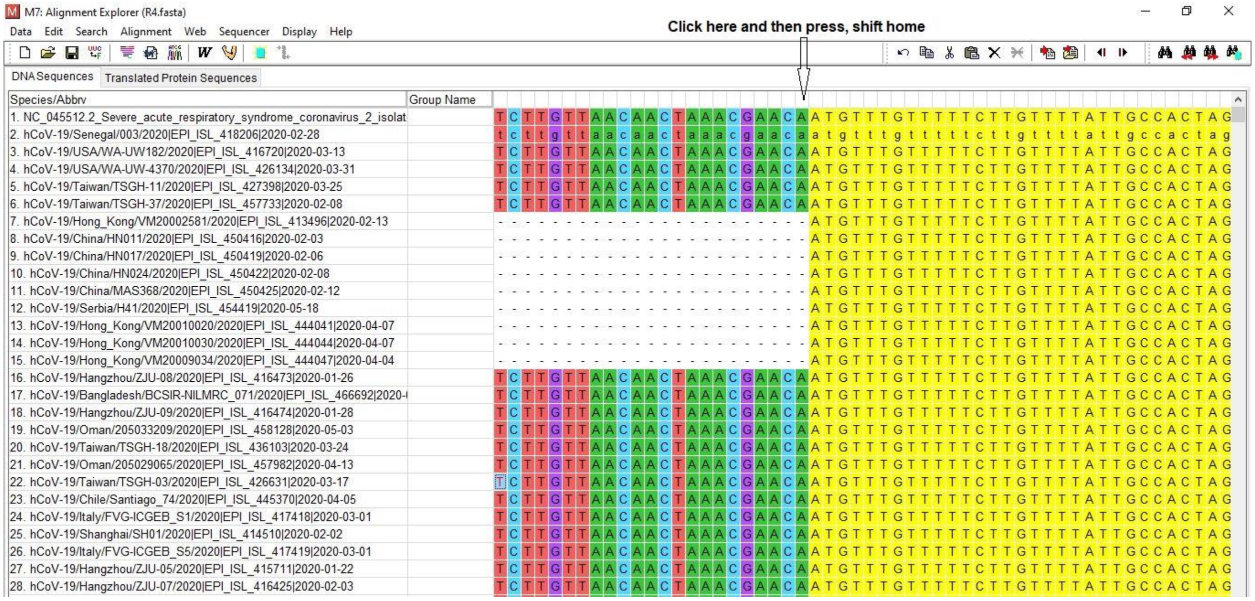
Fragmentary sequence(s) to be added to the above alignment: Example

Gaps (if any) will be removed.

or upload a plain text file: Choose File No file chosen Clear

Zipped file is acceptable.

3. Separate ORFs (S for SARS-CoV-2) from the alignment: I have downloaded the complete S (for example) from reference (https://www.ncbi.nlm.nih.gov/nuccore/1798174254) from NCBI and open it in a text editor (Notepad/Notepad++). Open the alignment in MEGA 7 or other version. Copy the first few nucleotide (at least 20) from S and the go to MEGA. Press, ctrl F, then ctrl V and finally press enter. In this, we can find the first section of S Protein in the alignment. Click on the upstream of the of the S starting and press, shift home, this will select the upstream part of S protein, then delete the upstream part. This will remove the upstream of the S protein. For downstream removal do the same but press, shift end, then delete the downstream part. Insertional sequence must be checked and delete them. If it has importance, preserve the sequence for further analysis.


M7: Alignment Explorer (R4.fasta)
Data Edit Search Alignment Web Sequencer Display Help
DNA Sequences
Translated Protein Sequences

Click here and then press, shift home

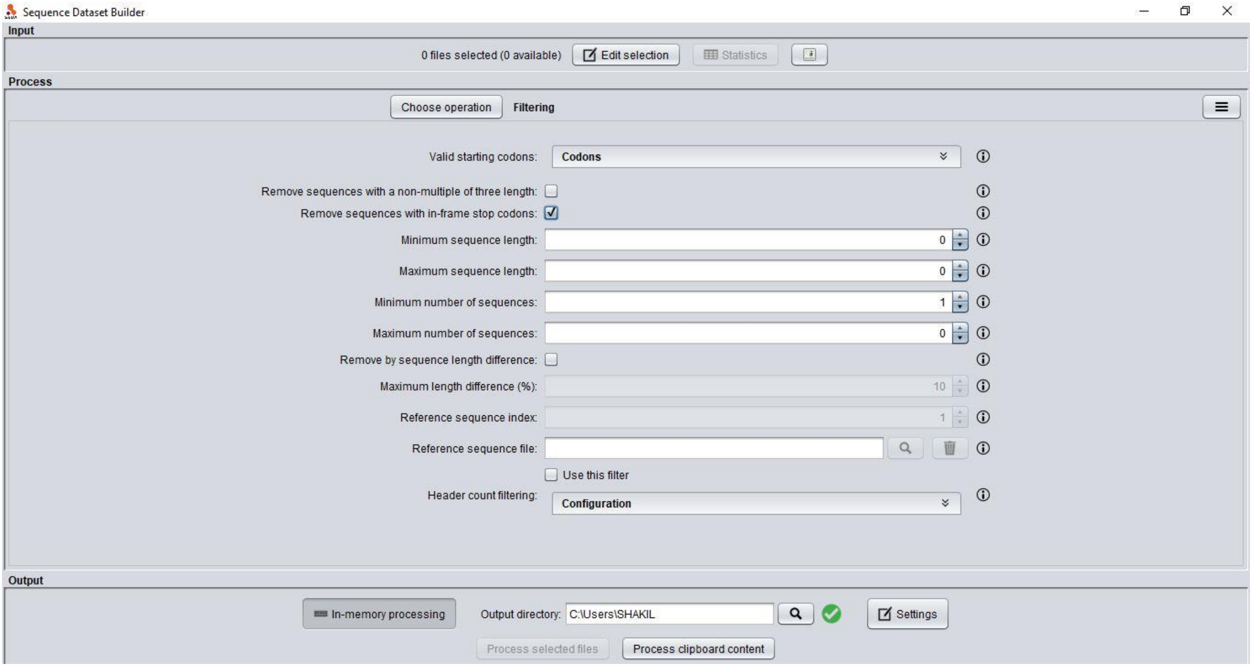
```
Species/Abbrv                                                              Group Name
1. NC_045512.2_Severe_acute_respiratory_syndrome_coronavirus_2_isolat                 TCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAG
2. hCoV-19/Senegal/003/2020|EPI_ISL_418206|2020-02-28                                 tcttgttaacaactaaacgaacaatgtttgtttttcttgttttattgccactag
3. hCoV-19/USA/WA-UW182/2020|EPI_ISL_416720|2020-03-13                                TCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAG
4. hCoV-19/USA/WA-UW-4370/2020|EPI_ISL_426134|2020-03-31                              TCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAG
5. hCoV-19/Taiwan/TSGH-11/2020|EPI_ISL_427398|2020-03-25                              TCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAG
6. hCoV-19/Taiwan/TSGH-37/2020|EPI_ISL_457733|2020-02-08                              TCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAG
7. hCoV-19/Hong_Kong/VM20002581/2020|EPI_ISL_413496|2020-02-13                        -----------------------ATGTTTGTTTTTCTTGTTTTATTGCCACTAG
8. hCoV-19/China/HN011/2020|EPI_ISL_450416|2020-02-03                                 -----------------------ATGTTTGTTTTTCTTGTTTTATTGCCACTAG
9. hCoV-19/China/HN017/2020|EPI_ISL_450419|2020-02-06                                 -----------------------ATGTTTGTTTTTCTTGTTTTATTGCCACTAG
10. hCoV-19/China/HN024/2020|EPI_ISL_450422|2020-02-08                                -----------------------ATGTTTGTTTTTCTTGTTTTATTGCCACTAG
11. hCoV-19/China/MAS368/2020|EPI_ISL_450425|2020-02-12                               -----------------------ATGTTTGTTTTTCTTGTTTTATTGCCACTAG
12. hCoV-19/Serbia/H41/2020|EPI_ISL_454419|2020-05-18                                 -----------------------ATGTTTGTTTTTCTTGTTTTATTGCCACTAG
13. hCoV-19/Hong_Kong/VM20010020/2020|EPI_ISL_444041|2020-04-07                       -----------------------ATGTTTGTTTTTCTTGTTTTATTGCCACTAG
14. hCoV-19/Hong_Kong/VM20010030/2020|EPI_ISL_444044|2020-04-07                       -----------------------ATGTTTGTTTTTCTTGTTTTATTGCCACTAG
15. hCoV-19/Hong_Kong/VM20009034/2020|EPI_ISL_444047|2020-04-04                       -----------------------ATGTTTGTTTTTCTTGTTTTATTGCCACTAG
16. hCoV-19/Hangzhou/ZJU-08/2020|EPI_ISL_416473|2020-01-26                            TCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAG
17. hCoV-19/Bangladesh/BCSIR-NILMRC_071/2020|EPI_ISL_466692|2020-(                    TCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAG
18. hCoV-19/Hangzhou/ZJU-09/2020|EPI_ISL_416474|2020-01-28                            TCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAG
19. hCoV-19/Oman/205033209/2020|EPI_ISL_458128|2020-05-03                             TCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAG
20. hCoV-19/Taiwan/TSGH-18/2020|EPI_ISL_436103|2020-03-24                             TCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAG
21. hCoV-19/Oman/205029065/2020|EPI_ISL_457982|2020-04-13                             TCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAG
22. hCoV-19/Taiwan/TSGH-03/2020|EPI_ISL_426631|2020-03-17                             TCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAG
23. hCoV-19/Chile/Santiago_74/2020|EPI_ISL_445370|2020-04-05                          TCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAG
24. hCoV-19/Italy/FVG-ICGEB_S1/2020|EPI_ISL_417418|2020-03-01                         TCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAG
25. hCoV-19/Shanghai/SH01/2020|EPI_ISL_414510|2020-02-02                              TCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAG
26. hCoV-19/Italy/FVG-ICGEB_S5/2020|EPI_ISL_417419|2020-03-01                         TCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAG
27. hCoV-19/Hangzhou/ZJU-05/2020|EPI_ISL_415711|2020-01-22                            TCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAG
28. hCoV-19/Hangzhou/ZJU-07/2020|EPI_ISL_416425|2020-02-03                            TCTTGTTAACAACTAAACGAACAATGTTTGTTTTTCTTGTTTTATTGCCACTAG
```

4. Sequence cleaner (https://github.com/metageni/Sequence-Cleaner): Sequence cleaner will assist to remove all sequences which contain ambiguous ('M', 'D', 'R', 'N', 'K', 'Y', 'S', 'B', 'H', '-', 'V', 'W') characters. -ml 3822 means minimum length less than remove and -mn 0 means remove all N containing sequences.

Usages (Linux)

```
sequence_cleaner -q /path/to/directory/of/FASTA -o clean.fasta -ml 3822 -mn 0 --keep_all
```

5. Remove internal stop codon using SEDA (http://www.sing-group.org/seda/download.html): Internal stop codon must remove before mutation analysis and SEDA v1.1 (https://www.sing-group.org/seda/manual/operations.html#id5) will assist to remove this. Windows 10 defender takes the file as corrupted, so before using it just make sure off the real-time protection.



Sequence Dataset Builder

Input

0 files selected (0 available) | Edit selection | Statistics

Process

Choose operation | Filtering

Valid starting codons: Codons

Remove sequences with a non-multiple of three length: ☐

Remove sequences with in-frame stop codons: ☑

Minimum sequence length: 0

Maximum sequence length: 0

Minimum number of sequences: 1

Maximum number of sequences: 0

Remove by sequence length difference: ☐

Maximum length difference (%): 10

Reference sequence index: 1

Reference sequence file:

☐ Use this filter

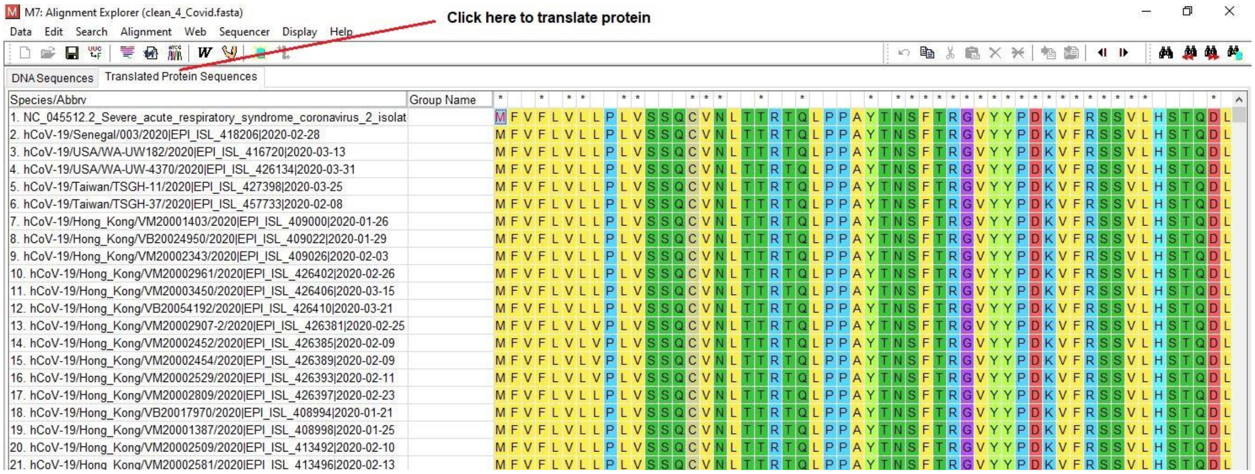
Header count filtering: Configuration

Output

In-memory processing | Output directory: C:\Users\SHAKIL | Settings

Process selected files | Process clipboard content

6. DNA to Protein in MEGA (https://www.megasoftware.net/dload_win_gui): After removal of all low quality and in frame stop codon sequences, we need to translate the DNA to protein. MEGA 7 or other version will aid us to do it. Just open the multiple sequence alignment and press the translate to protein options. Then export the protein alignment to fasta format.



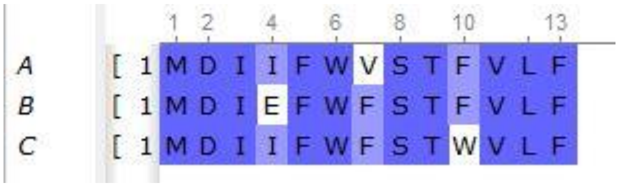
M7: Alignment Explorer (clean_4_Covid.fasta)

Click here to translate protein

Species/Abbrv	Group Name	* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
1. NC_045512.2_Severe_acute_respiratory_syndrome_coronavirus_2_isolat		M F V F L V L L P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
2. hCoV-19/Senegal/003/2020|EPI_ISL_418206|2020-02-28		M F V F L V L L P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
3. hCoV-19/USA/WA-UW182/2020|EPI_ISL_416720|2020-03-13		M F V F L V L L P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
4. hCoV-19/USA/WA-UW-4370/2020|EPI_ISL_426134|2020-03-31		M F V F L V L L P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
5. hCoV-19/Taiwan/TSGH-11/2020|EPI_ISL_427398|2020-03-25		M F V F L V L L P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
6. hCoV-19/Taiwan/TSGH-37/2020|EPI_ISL_457733|2020-02-08		M F V F L V L L P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
7. hCoV-19/Hong_Kong/VM20001403/2020|EPI_ISL_409000|2020-01-26		M F V F L V L L P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
8. hCoV-19/Hong_Kong/VB20024950/2020|EPI_ISL_409022|2020-01-29		M F V F L V L L P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
9. hCoV-19/Hong_Kong/VM20002343/2020|EPI_ISL_409026|2020-02-03		M F V F L V L L P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
10. hCoV-19/Hong_Kong/VM20002961/2020|EPI_ISL_426402|2020-02-26		M F V F L V L L P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
11. hCoV-19/Hong_Kong/VM20003450/2020|EPI_ISL_426406|2020-03-15		M F V F L V L L P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
12. hCoV-19/Hong_Kong/VB20054192/2020|EPI_ISL_426410|2020-03-21		M F V F L V L L P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
13. hCoV-19/Hong_Kong/VM20002907-2/2020|EPI_ISL_426381|2020-02-25		M F V F L V L V P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
14. hCoV-19/Hong_Kong/VM20002452/2020|EPI_ISL_426385|2020-02-09		M F V F L V L V P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
15. hCoV-19/Hong_Kong/VM20002454/2020|EPI_ISL_426389|2020-02-09		M F V F L V L V P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
16. hCoV-19/Hong_Kong/VM20002529/2020|EPI_ISL_426393|2020-02-11		M F V F L V L V P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
17. hCoV-19/Hong_Kong/VM20002809/2020|EPI_ISL_426397|2020-02-23		M F V F L V L L P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
18. hCoV-19/Hong_Kong/VB20017970/2020|EPI_ISL_408994|2020-01-21		M F V F L V L L P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
19. hCoV-19/Hong_Kong/VM20001387/2020|EPI_ISL_408998|2020-01-25		M F V F L V L L P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
20. hCoV-19/Hong_Kong/VM20002509/2020|EPI_ISL_413492|2020-02-10		M F V F L V L L P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L
21. hCoV-19/Hong_Kong/VM20002581/2020|EPI_ISL_413496|2020-02-13		M F V F L V L L P L V S S Q C V N L T T R T Q L P P A Y T N S F T R G V Y Y P D K V F R S S V L H S T Q D L

7. Pairwise mutation analysis (https://github.com/SShaminur/Mutation-Analysis): Here the mutation results come from pairwise alignment with respect to reference genome (reference amino acid:position:strain amino acid). Before running, we must to ensure that there are no stop codon (*) in the last position of the ORF. We must delete the last stop codon in MEGA analysis.

Multiple sequence alignment (Top one will be selected as reference)



```
          1 2   4   6   8   10    13
A  [ 1  M D I I F W V S T F V L F
B  [ 1  M D I E F W F S T F V L F
C  [ 1  M D I I F W F S T W V L F
```

Results (Reference A)

```
B I4E
B V7F
C V7F
C F10W
```

Usages (Linux): (Requirements: Python ≥3.7, Biopython)

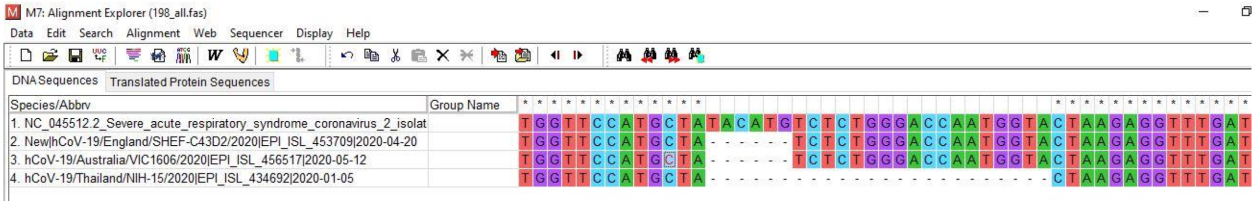
```
./pairwise_mutation.py input_align.fasta > Output.txt
```

8. Deletion analysis (https://bioinf.shenwei.me/seqkit/usage/**):** SeqKit tools used to arrest all the sequences containing gap (-). From there in frame deletion should be carefully find out. To ensure the in frame deletion with MEGA, remove the reference strain, then remove the gap of the strain/s and translate it to protein. Export the protein sequence/s in fasta format. Usually this translation finds the stop codon in last position. If not, then there are sequencing error (if virus not evolve too much). Then again align the protein sequence to the reference genome protein sequences. We can find the deleted amino acid position. Triplet codon deletion should be screen for deletion analysis. Then finally visualize the deletion data in a suitable software (Jalview, Unipro-UGENE, BioEdit etc.)

Usages (Linux):

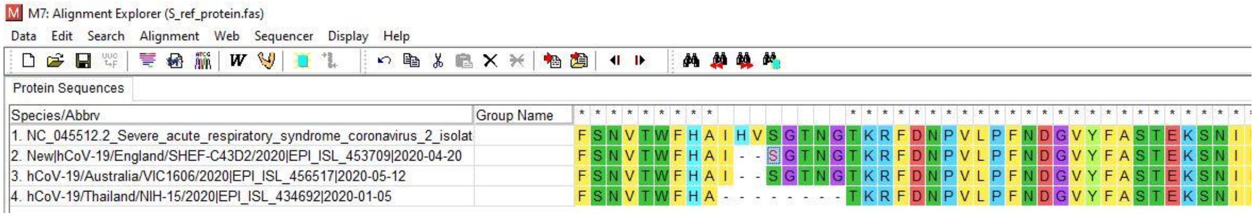
```
seqkit grep -s -p  "-" in.fa > out.fa
```

<u>*DNA deletion*</u>


M7: Alignment Explorer (198_all.fas)
Data
Edit
Search
Alignment
Web
Sequencer
Display
Help
DNA Sequences
Translated Protein Sequences

```
Species/Abbrv                                                  Group Name  * * * * * * * * * * * * *                                                 * * * * * * * * * * * * * *
1. NC_045512.2_Severe_acute_respiratory_syndrome_coronavirus_2_isolat      T G G T T C C A T G C T A T A C A T G T C T C T G G G A C C A A T G G T A C T A A G A G G T T T G A T
2. New|hCoV-19/England/SHEF-C43D2/2020|EPI_ISL_453709|2020-04-20           T G G T T C C A T G C T A - - - - - - T C T C T G G G A C C A A T G G T A C T A A G A G G T T T G A T
3. hCoV-19/Australia/VIC1606/2020|EPI_ISL_456517|2020-05-12                T G G T T C C A T G C T A - - - - - - T C T C T G G G A C C A A T G G T A C T A A G A G G T T T G A T
4. hCoV-19/Thailand/NIH-15/2020|EPI_ISL_434692|2020-01-05                  T G G T T C C A T G C T A - - - - - - - - - - - - - - - - - - - - - - - - C T A A G A G G T T T G A T
```

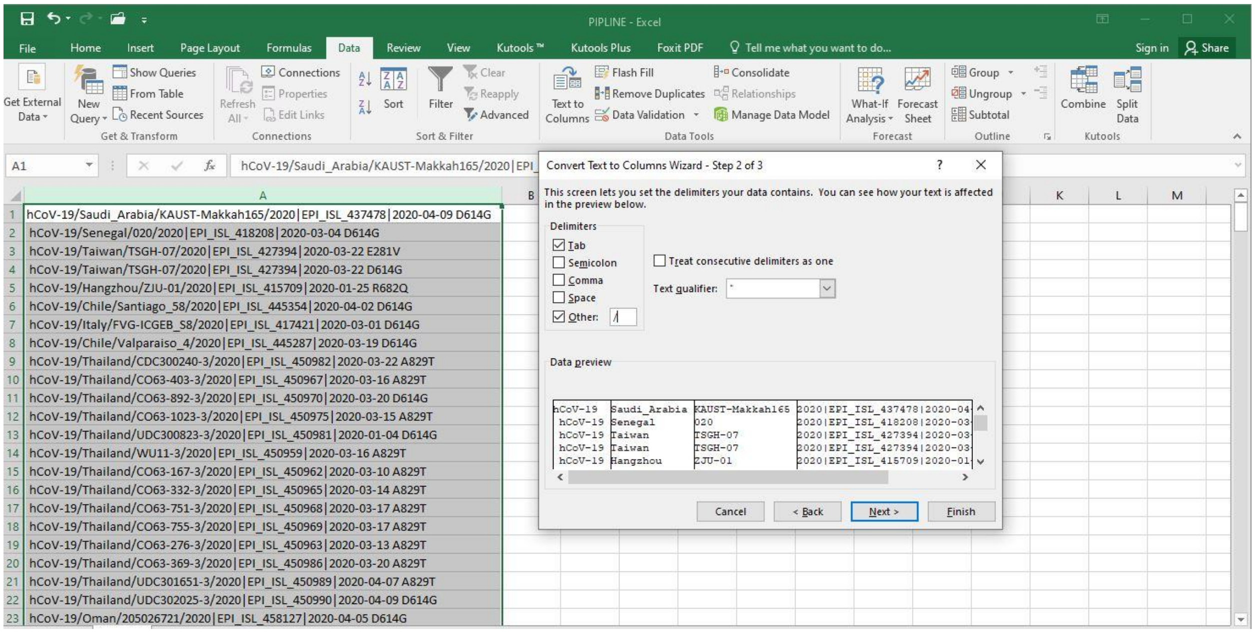
<u>*Respective protein deletion*</u>



```
Species/Abbrv                                                  Group Name  * * * * * * * * *                 * * * * * * * * * * * * * * * * * * * * * * * * *
1. NC_045512.2_Severe_acute_respiratory_syndrome_coronavirus_2_isolat      F S N V T W F H A I H V S G T N G T K R F D N P V L P F N D G V Y F A S T E K S N I
2. New|hCoV-19/England/SHEF-C43D2/2020|EPI_ISL_453709|2020-04-20           F S N V T W F H A I - - S G T N G T K R F D N P V L P F N D G V Y F A S T E K S N I
3. hCoV-19/Australia/VIC1606/2020|EPI_ISL_456517|2020-05-12                F S N V T W F H A I - - S G T N G T K R F D N P V L P F N D G V Y F A S T E K S N I
4. hCoV-19/Thailand/NIH-15/2020|EPI_ISL_434692|2020-01-05                  F S N V T W F H A - - - - - - - - T K R F D N P V L P F N D G V Y F A S T E K S N I
```

9. Mutational analysis using Microsoft Excel:

I. **Text to column:** Select the column, Then, **Data >Text to Columns > Delimited > Next > Space, Other (/) > Next and Finish.** This will differentiate country name Mutation and others. If we want to separate accession number, the just put Other **(**|**)**.



	A
1	hCoV-19/Saudi_Arabia/KAUST-Makkah165/2020|EPI_ISL_437478|2020-04-09 D614G
2	hCoV-19/Senegal/020/2020|EPI_ISL_418208|2020-03-04 D614G
3	hCoV-19/Taiwan/TSGH-07/2020|EPI_ISL_427394|2020-03-22 E281V
4	hCoV-19/Taiwan/TSGH-07/2020|EPI_ISL_427394|2020-03-22 D614G
5	hCoV-19/Hangzhou/ZJU-01/2020|EPI_ISL_415709|2020-01-25 R682Q
6	hCoV-19/Chile/Santiago_58/2020|EPI_ISL_445354|2020-04-02 D614G
7	hCoV-19/Italy/FVG-ICGEB_S8/2020|EPI_ISL_417421|2020-03-01 D614G
8	hCoV-19/Chile/Valparaiso_4/2020|EPI_ISL_445287|2020-03-19 D614G
9	hCoV-19/Thailand/CDC300240-3/2020|EPI_ISL_450982|2020-03-22 A829T
10	hCoV-19/Thailand/CO63-403-3/2020|EPI_ISL_450967|2020-03-16 A829T
11	hCoV-19/Thailand/CO63-892-3/2020|EPI_ISL_450970|2020-03-20 D614G
12	hCoV-19/Thailand/CO63-1023-3/2020|EPI_ISL_450975|2020-03-15 A829T
13	hCoV-19/Thailand/UDC300823-3/2020|EPI_ISL_450981|2020-01-04 D614G
14	hCoV-19/Thailand/WU11-3/2020|EPI_ISL_450959|2020-03-16 A829T
15	hCoV-19/Thailand/CO63-167-3/2020|EPI_ISL_450962|2020-03-10 A829T
16	hCoV-19/Thailand/CO63-332-3/2020|EPI_ISL_450965|2020-03-14 A829T
17	hCoV-19/Thailand/CO63-751-3/2020|EPI_ISL_450968|2020-03-17 A829T
18	hCoV-19/Thailand/CO63-755-3/2020|EPI_ISL_450969|2020-03-17 A829T
19	hCoV-19/Thailand/CO63-276-3/2020|EPI_ISL_450963|2020-03-13 A829T
20	hCoV-19/Thailand/CO63-369-3/2020|EPI_ISL_450986|2020-03-20 A829T
21	hCoV-19/Thailand/UDC301651-3/2020|EPI_ISL_450989|2020-04-07 A829T
22	hCoV-19/Thailand/UDC302025-3/2020|EPI_ISL_450990|2020-04-09 D614G
23	hCoV-19/Oman/205026721/2020|EPI_ISL_458127|2020-04-05 D614G

Convert Text to Columns Wizard - Step 2 of 3

This screen lets you set the delimiters your data contains. You can see how your text is affected in the preview below.

Delimiters

☑ Tab
☐ Semicolon
☐ Comma
☐ Space
☑ Other: /

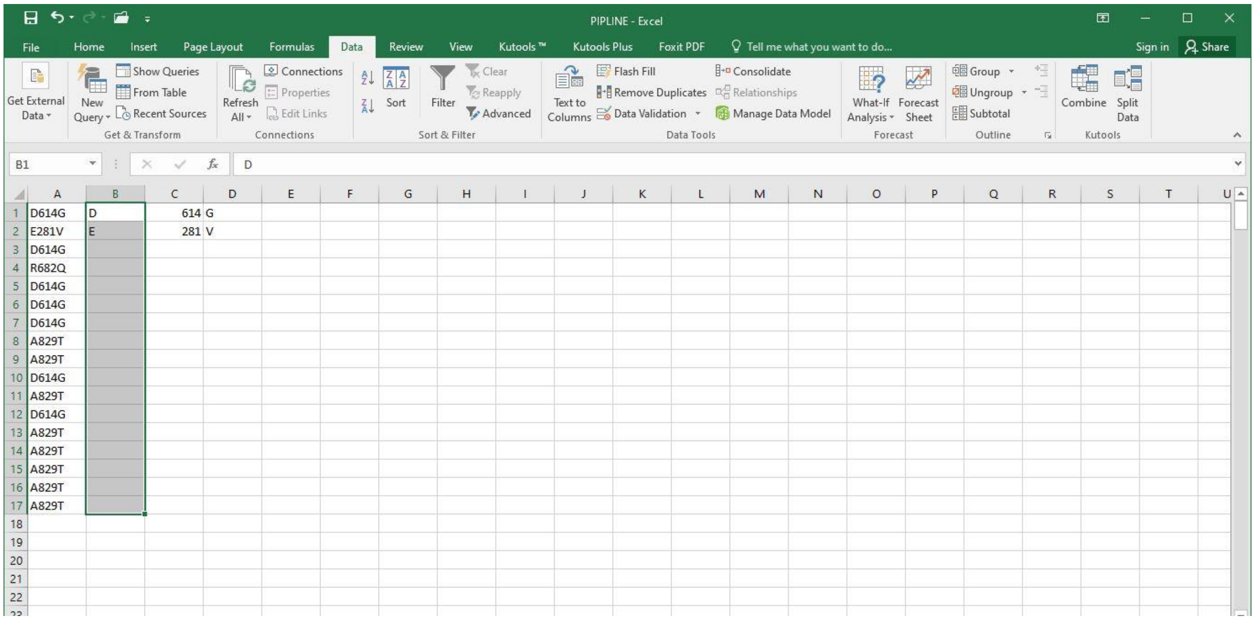
☐ Treat consecutive delimiters as one

Text qualifier: "

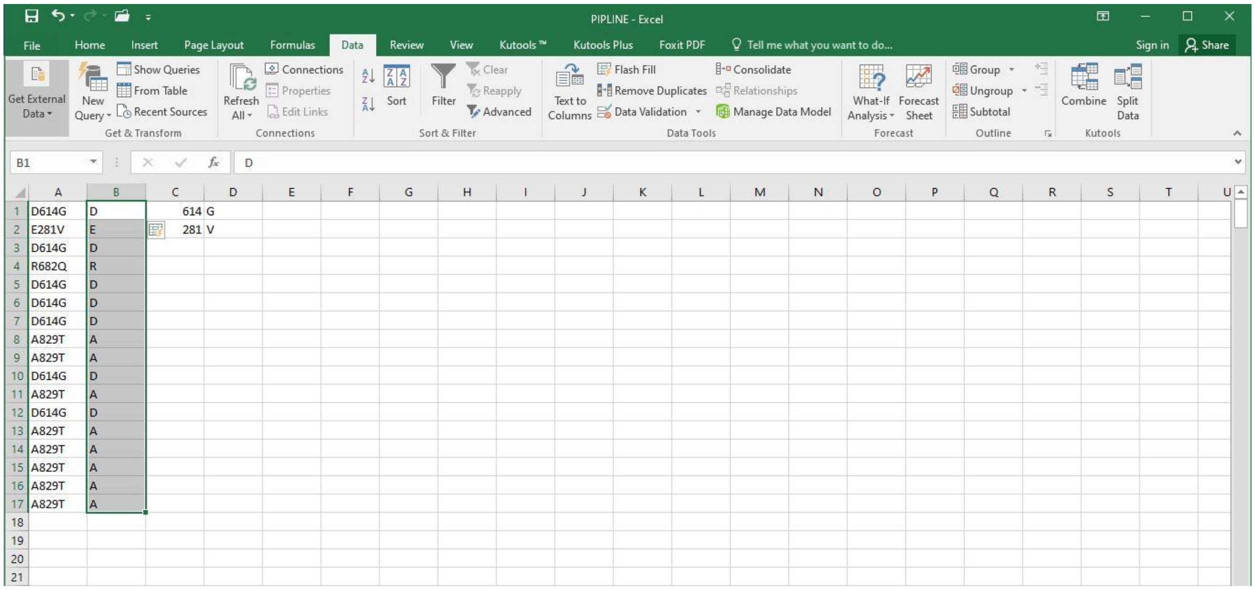
Data preview

```
hCoV-19   Saudi_Arabia  KAUST-Makkah165  2020|EPI_ISL_437478|2020-04
 hCoV-19  Senegal       020              2020|EPI_ISL_418208|2020-03
 hCoV-19  Taiwan        TSGH-07          2020|EPI_ISL_427394|2020-03
 hCoV-19  Taiwan        TSGH-07          2020|EPI_ISL_427394|2020-03
 hCoV-19  Hangzhou      ZJU-01           2020|EPI_ISL_415709|2020-01
```

II. **Flash Fill:** This will separate the mutation (ref:position:strain) into three different columns. From there, we can sort largest to smallest or vice versa. For Flash Fill, first we need to fill up at least two raw then, **Data** > select the 1st desired column > **Flash Fill.** For rest of the two columns, do the same thing.



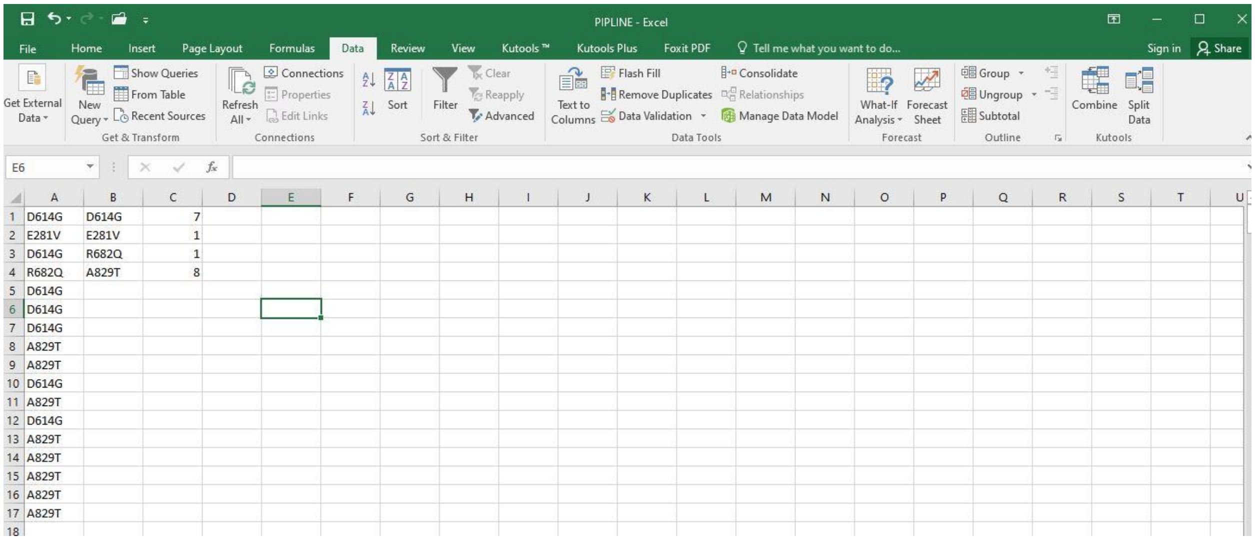
	A	B	C	D
1	D614G	D	614	G
2	E281V	E	281	V
3	D614G			
4	R682Q			
5	D614G			
6	D614G			
7	D614G			
8	A829T			
9	A829T			
10	D614G			
11	A829T			
12	D614G			
13	A829T			
14	A829T			
15	A829T			
16	A829T			
17	A829T			



	A	B	C	D
1	D614G	D	614	G
2	E281V	E	281	V
3	D614G	D		
4	R682Q	R		
5	D614G	D		
6	D614G	D		
7	D614G	D		
8	A829T	A		
9	A829T	A		
10	D614G	D		
11	A829T	A		
12	D614G	D		
13	A829T	A		
14	A829T	A		
15	A829T	A		
16	A829T	A		
17	A829T	A		

III. **Remove Duplicates:** Unique mutation, Unique position mutation can be found by removing duplicates. Select the column (where duplicates need to remove) then, **Data** > **Remove Duplicates.**

IV. **Frequency count:** For frequency count, 1st column contains the original data and second column contains the duplicate removal data. Then select a cell in third column and then, =COUNTIFS(A1:A17,B1).



	A	B	C
1	D614G	D614G	7
2	E281V	E281V	1
3	D614G	R682Q	1
4	R682Q	A829T	8
5	D614G		
6	D614G		
7	D614G		
8	A829T		
9	A829T		
10	D614G		
11	A829T		
12	D614G		
13	A829T		
14	A829T		
15	A829T		
16	A829T		
17	A829T		
18			

10. Arrest sequence through sequence ID:

Usages (Linux):

```
grep -A1 -w -f ID.txt data_with_fasta.fas > ID.fasta
```