BioAider V1.03

1. Introduction

2. Download and install

3. Functions

   3.1. SeqTools

      3.1.1. Seqformat Convertor

      3.1.2. SeqVary

      3.1.3. SequenceID Rename

      3.1.4. Split Sequence Fragmenet

      3.1.5. Combine Gene (Tandem Gene)

      3.1.6. Visual Gene Extractor

      3.1.7. Fast Annotation

   3.2. Similar Analysis

      3.2.1. Sequence Identity Matrix

      3.2.2 Remove H-Similar Sequence

   3.3. Align tools

      3.3.1. Mafft

      3.3.2. Muscle

      3.3.2. Clustal-Omeg

   3.4. Mutation Tools

      3.4.1. Mutation Analysis

      3.4.2. Site Counter

      3.4.3. Site Scree

   4. Test Datas

Manual of BioAider V1.03

A richly featured desktop platform libraries for analysis of bioinformatics datas

Home page: https://github.com/ZhijianZhou01/BioAider

Version 1.03 || August 18, 2020

1. Introduction

With the development of sequencing technology, a large amount of genomic sequenced datas has been accumulated. Analyzing these data will help us understand their genetic variation at the molecular level. However, processing a large-scale sequences is difficult for biological or clinical expert without bioinformatics and programming skills. Besides, the needs are also diverse due to different research purposes. Therefore, simplicity of operation and diversity of function are needed.

Bioinformatics Aider (BioAider) V1.03 is developed based on Python3 and PySide2, which is a user-friendly GUI-interface program. As a desktop platform for genomic sequencing data studies, BioAider is designed to simplicity of operation and high summary of analysis results, which could save a lot of time for researchers.

2. Download and install

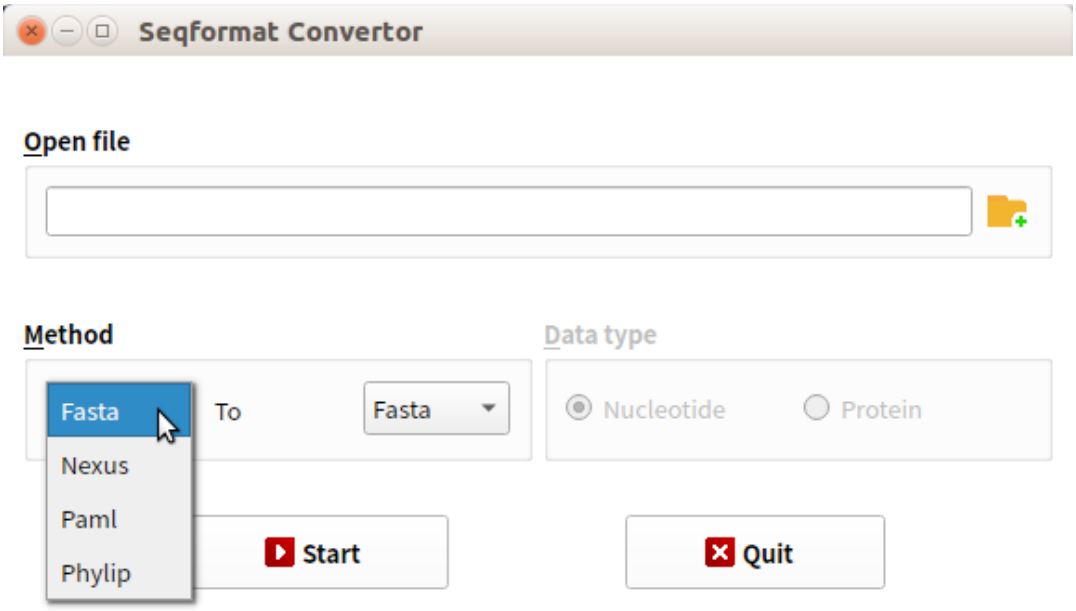
BioAider and all the updated versions is freely available for non-commercial user at https://github.com/ZhijianZhou01/BioAider/releases. After obtaining the program, users could directly run the program by clicking executable file without installing in Windows or Linux(Ubuntu 16.04 or more) systerms.

3. Functions

3.1. *SeqTools*

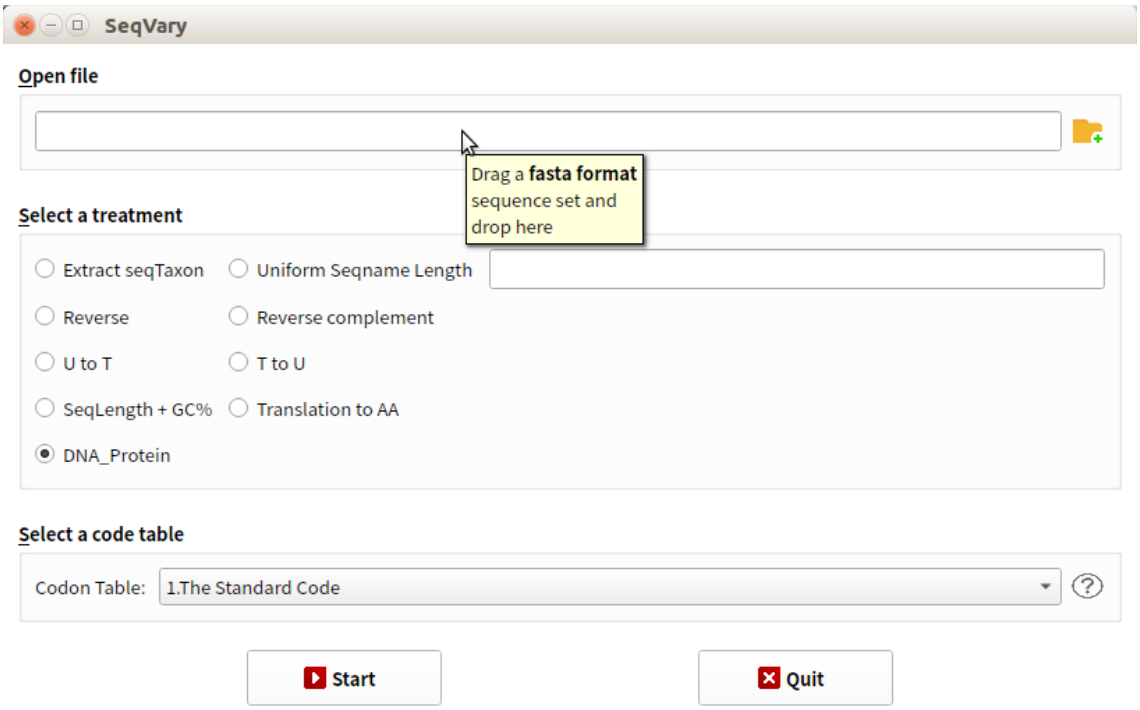
3.1.1. *Seqformat Convertor*

BioAider provides mutual conversion among several common sequence formats, which are Fasta, Nexus, Paml, and Phylip. Of note, the "***Data type***" option is only available when the target format is "Nexus".



3.1.2. *SeqVary*

The ***"SeqVary"*** option of BioAider provides some small functions for sequence preprocessing. For example, ***"SeqLength+GC%"*** is used to batch calculate sequence length and content of GC.

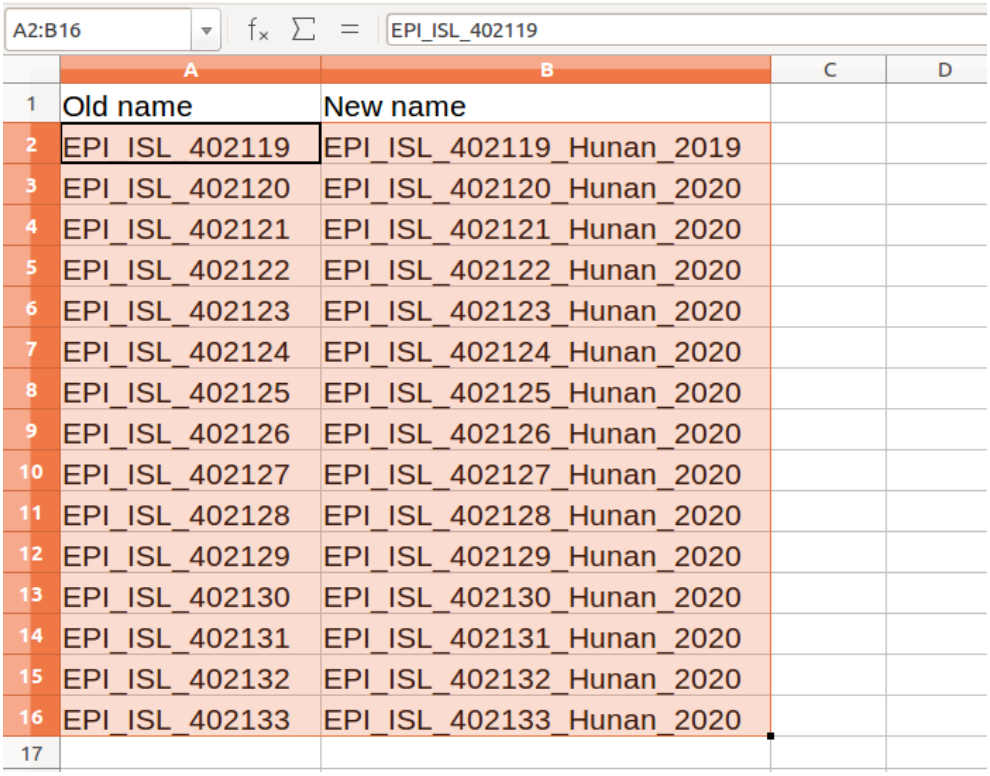


Note: the ***"DNA_Protein"*** option requires the gene sequences datas to be aligned based on codons.

3.1.3. *SequenceID Rename*

BioAider could rename the original name in **sequence datas or tree file etc**. In particular, the pictures of the evolutionary tree used for publication often require the taxons of tree to follow a uniform format, so first batch replacement in the tree file saves the trouble of using vector graphics tools to modify later.

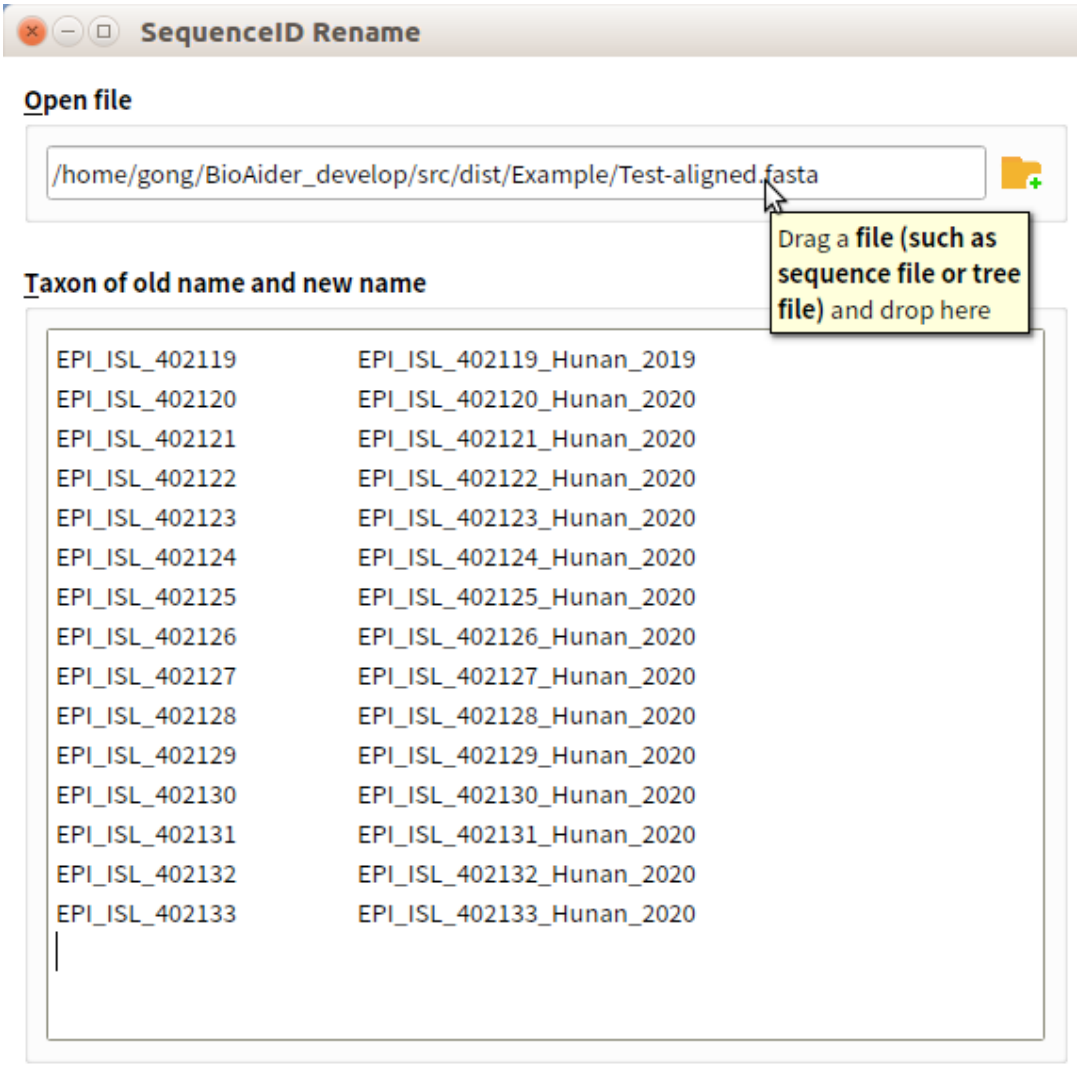
First, make a table of **new and old names** in a table editor :



A2:B16 | EPI_ISL_402119

	A	B	C	D
1	Old name	New name		
2	EPI_ISL_402119	EPI_ISL_402119_Hunan_2019		
3	EPI_ISL_402120	EPI_ISL_402120_Hunan_2020		
4	EPI_ISL_402121	EPI_ISL_402121_Hunan_2020		
5	EPI_ISL_402122	EPI_ISL_402122_Hunan_2020		
6	EPI_ISL_402123	EPI_ISL_402123_Hunan_2020		
7	EPI_ISL_402124	EPI_ISL_402124_Hunan_2020		
8	EPI_ISL_402125	EPI_ISL_402125_Hunan_2020		
9	EPI_ISL_402126	EPI_ISL_402126_Hunan_2020		
10	EPI_ISL_402127	EPI_ISL_402127_Hunan_2020		
11	EPI_ISL_402128	EPI_ISL_402128_Hunan_2020		
12	EPI_ISL_402129	EPI_ISL_402129_Hunan_2020		
13	EPI_ISL_402130	EPI_ISL_402130_Hunan_2020		
14	EPI_ISL_402131	EPI_ISL_402131_Hunan_2020		
15	EPI_ISL_402132	EPI_ISL_402132_Hunan_2020		
16	EPI_ISL_402133	EPI_ISL_402133_Hunan_2020		
17				

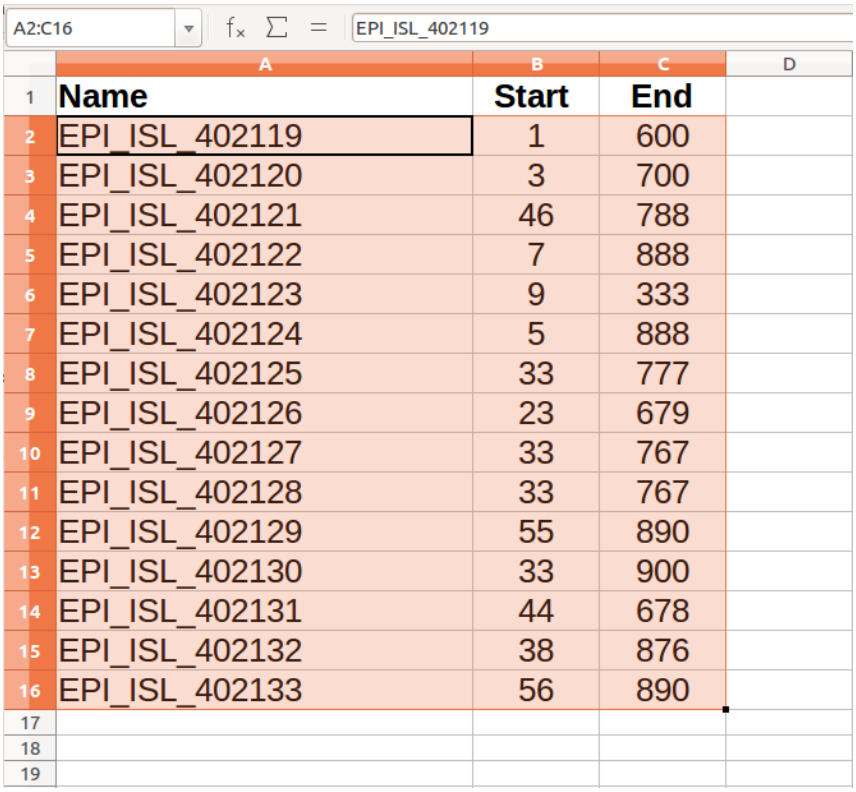
Then copy and paste them into BioAider:



3.1.4. *Split Sequence Fragmenet*

This function can batch intercept the specified range of gene fragments, two different modes are available: specified different range (***"Different range"***) for each sequence, equal range for all sequences (***"Equal range"***).

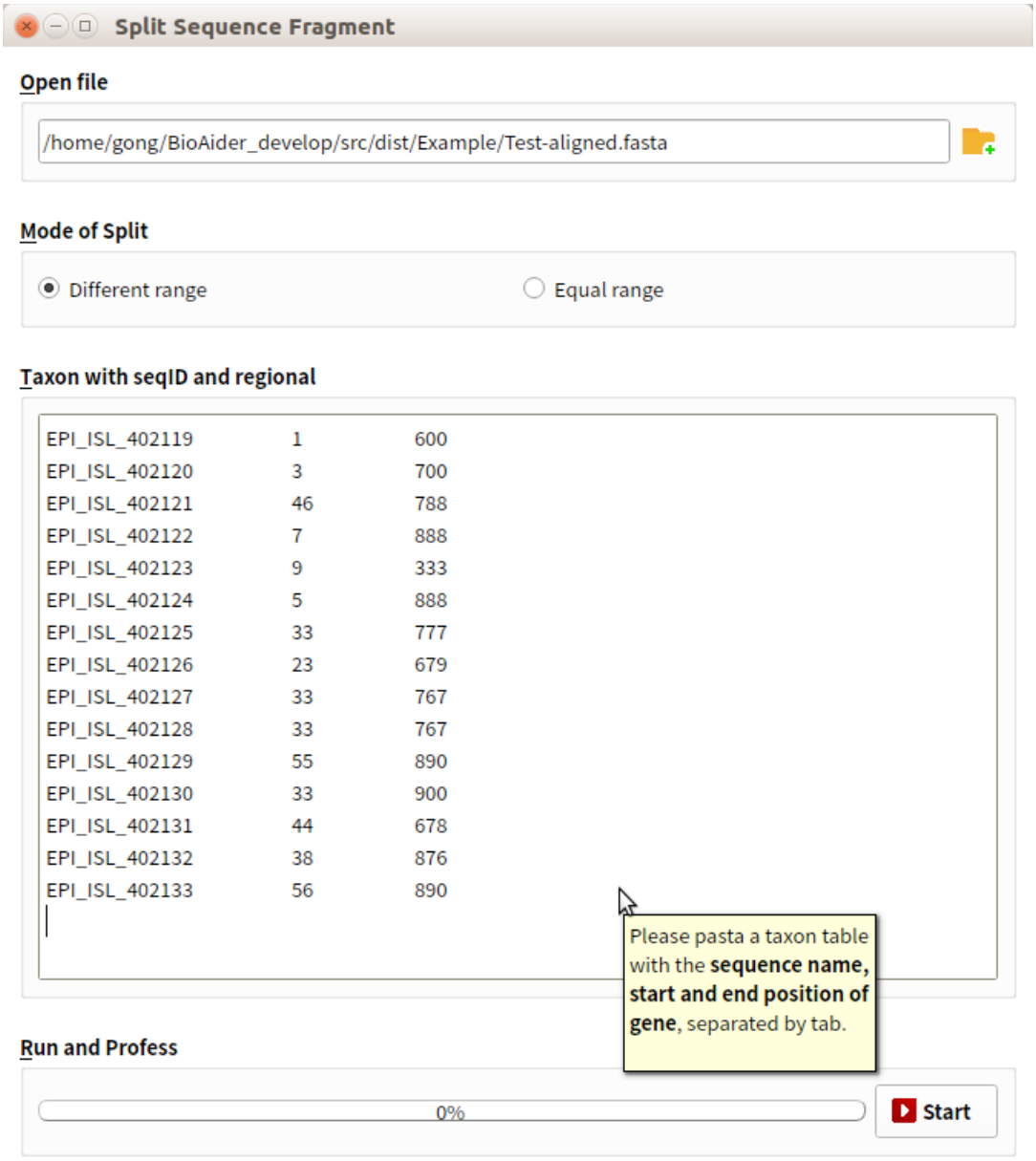
If you want to use the ***"Different range"*** to split for each sequence , make a table of start and end location firstly:



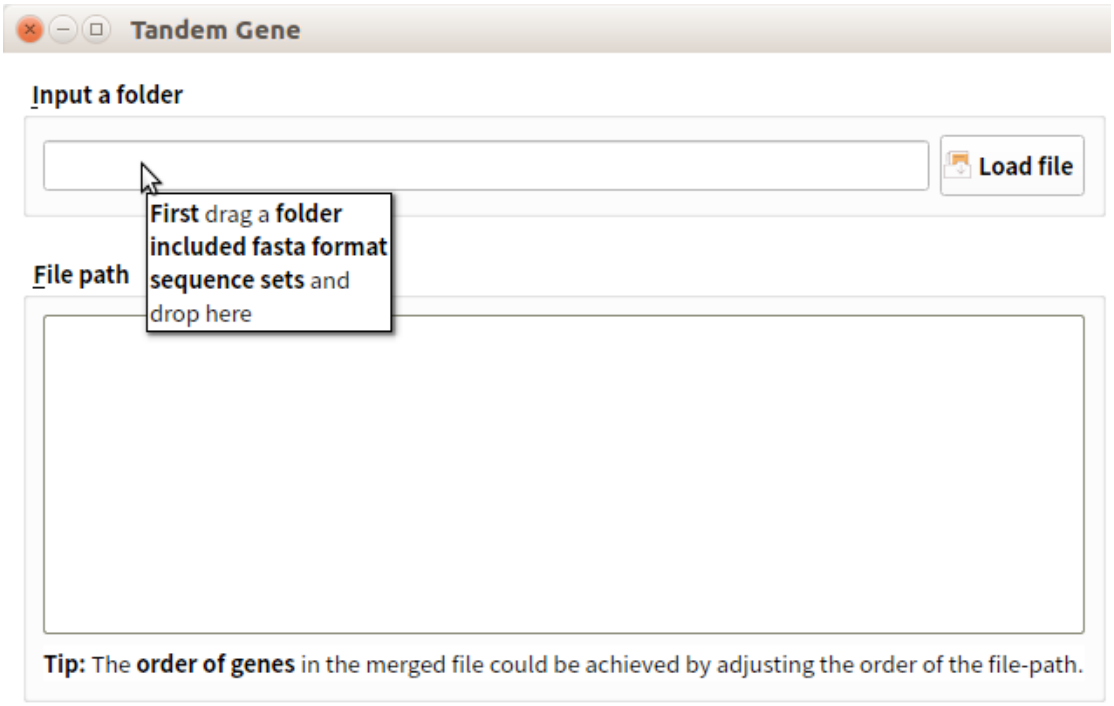
A2:C16 | EPI_ISL_402119

	A	B	C	D
1	**Name**	**Start**	**End**	
2	EPI_ISL_402119	1	600	
3	EPI_ISL_402120	3	700	
4	EPI_ISL_402121	46	788	
5	EPI_ISL_402122	7	888	
6	EPI_ISL_402123	9	333	
7	EPI_ISL_402124	5	888	
8	EPI_ISL_402125	33	777	
9	EPI_ISL_402126	23	679	
10	EPI_ISL_402127	33	767	
11	EPI_ISL_402128	33	767	
12	EPI_ISL_402129	55	890	
13	EPI_ISL_402130	33	900	
14	EPI_ISL_402131	44	678	
15	EPI_ISL_402132	38	876	
16	EPI_ISL_402133	56	890	
17				
18				
19				

Then copy and paste them into BioAider:



3.1.5. *Combine Gene (Tandem Gene)*



This function is used to concatenate multiple gene sequences into one. Users can first put different genes dataset files into the same folder, and then drag the folder into the ***inputbox***, then click the ***"Load file button"*** import the file path of each genes datasets into ***textbox***.

Note, users can modify the order of genes in tandemy sequence by adjusting the sort of inputfile path in the ***textbox.*** Of note, all the sequences which are used for combined should be fasta format.

3.1.6. *Visual Gene Extractor*

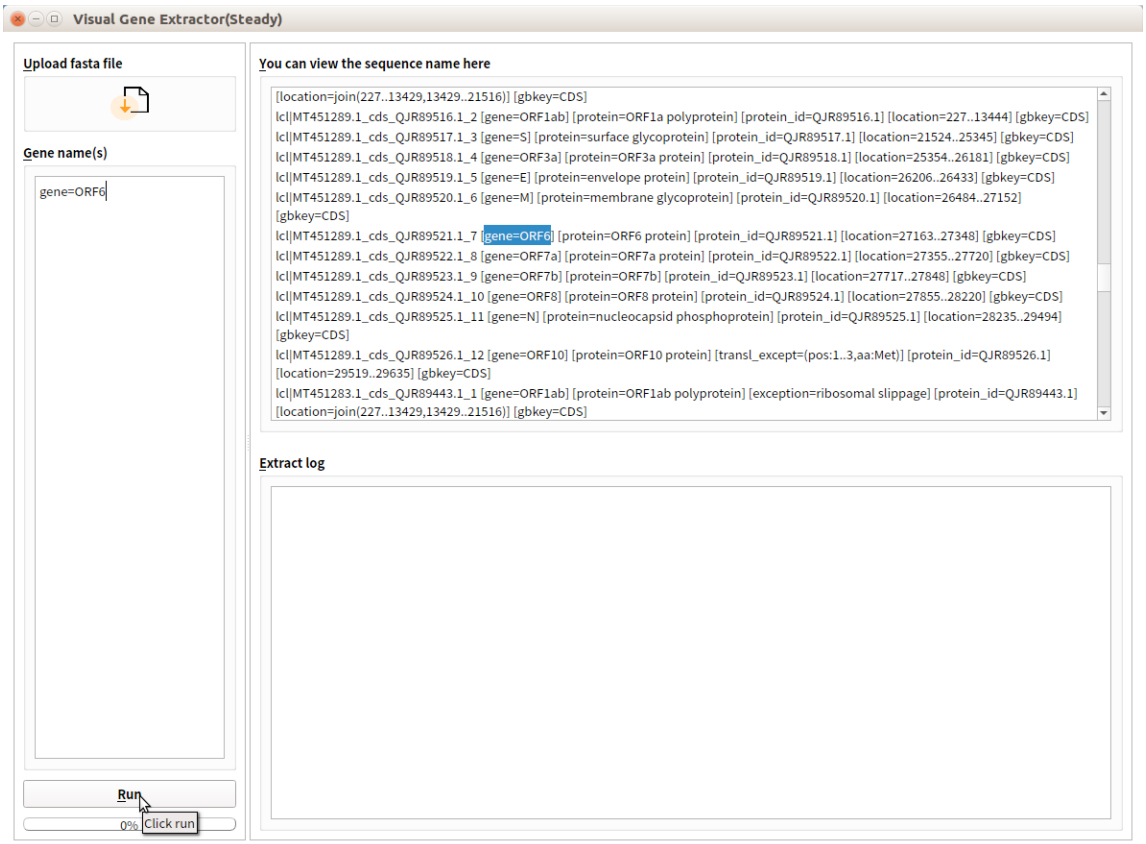
This function is used to extract the sequences included specified gene from mixed coding gene sequence set , especially when these sequences datas are downloaded from NCBI database. Given that the same gene may have different manifestations in different studies, the textbox of ***"Gene name"*** could enter multiple names, and BioAider will extract the corresponding gene sequence which contain these gene names.

BioAider providers two versions, ***Visual Gene Extractor(Streamlit)*** and **Visual Gene Extractor(Steady)**.The versions of ***Visual Gene Extractor(Streamlit)*** was developed based on ***Streamlit***, so if you use this version, you need to install Streamlit on your

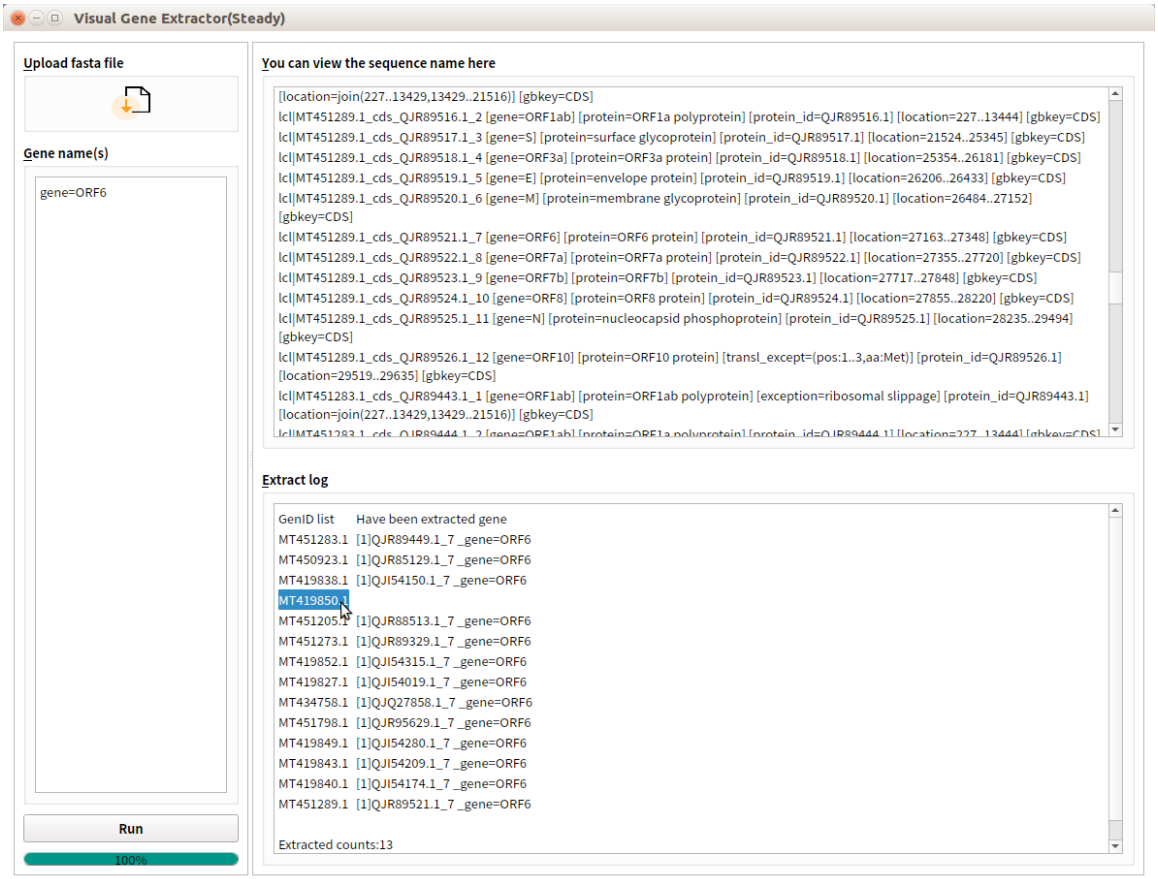
computer first. On the other hand, the ***Visual Gene Extractor(Steady)*** does not need any other environment.

Next, we demonstrate the use of ***Visual Gene Extractor(Steady)***.

Example 1 (The sequences datas are directly downloaded from NCBI database, including some gene fragments of SARS-CoV-2):

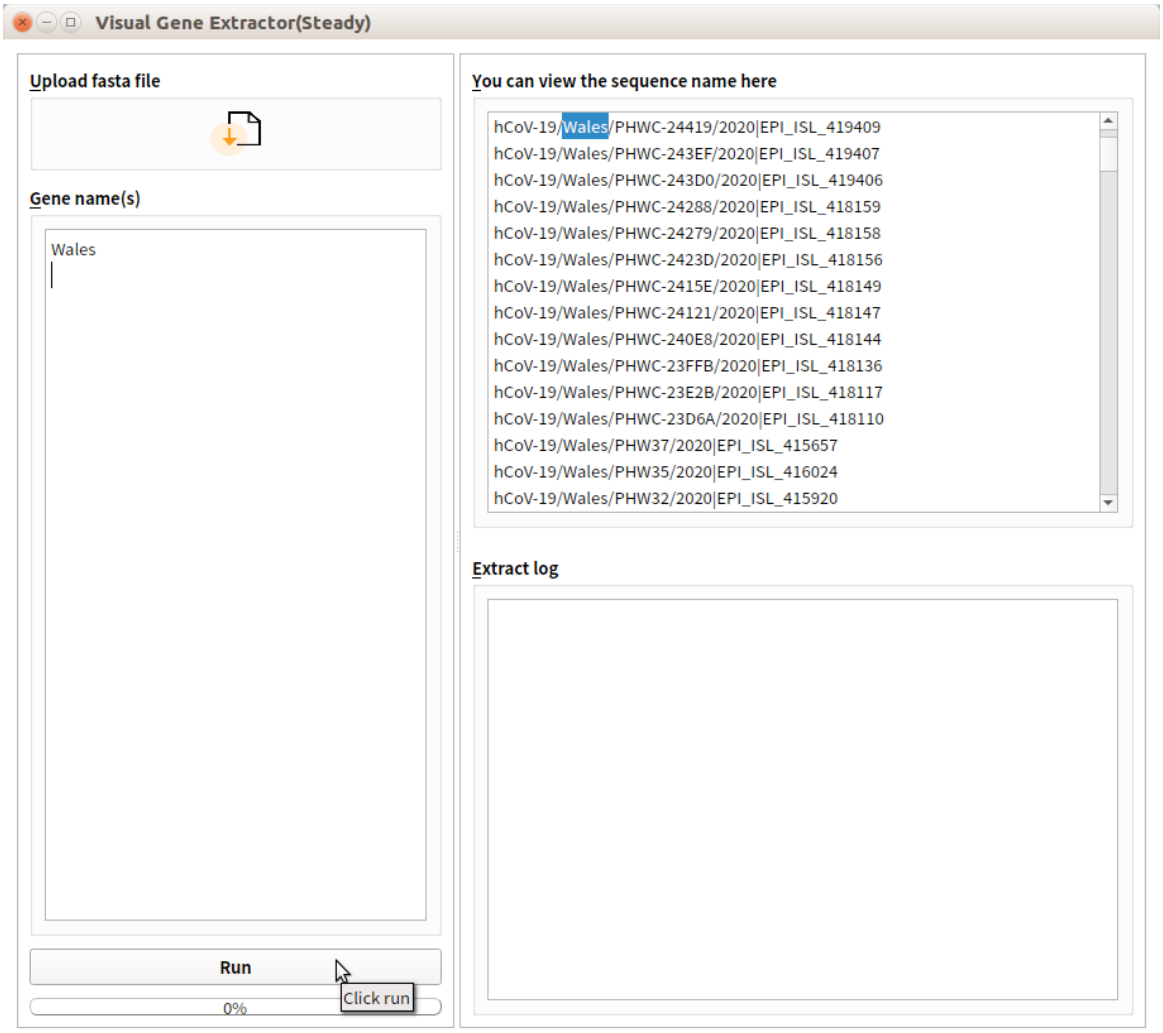


After uploading the sequence to BioAidrs as above, then we extract ORF6 gene sequence of SARS-CoV-2. Input a string **containing at least the gene name** to textbox of ***"Gene name(s)"***, then click button of ***"Run"***, run log as follows:

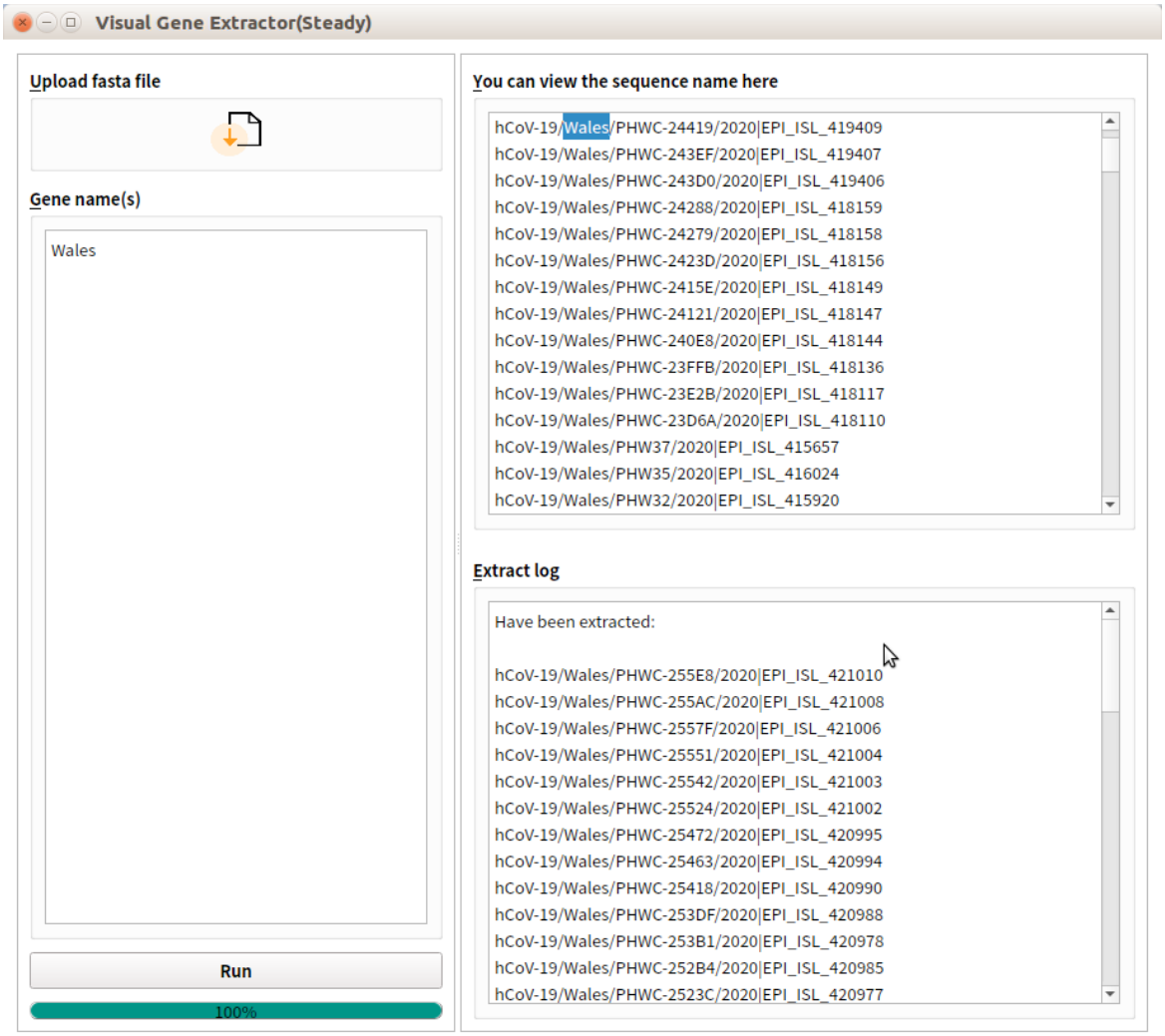


If some gene fragments are not extracted (as shown by the arrow), the possible reason is that the strain does not contain this gene fragment or the gene owns other names in some sequences. If it is the second case, you can append other names of this gene to the next line of ***"Gene name(s)"*** textbox.

Example 2 (Arbitrary fasta sequence):

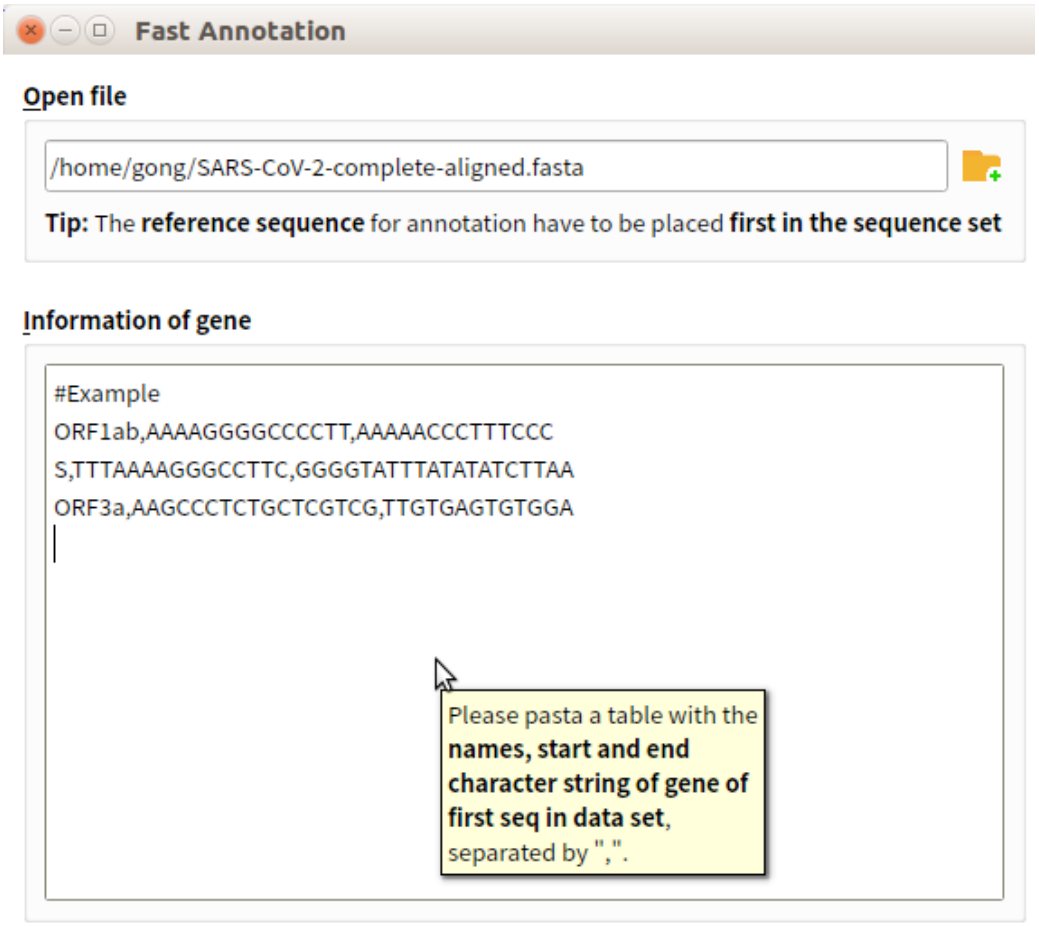


As shown above, if you want to extract these sequences which containing the tags of ***"Wales"***, BioAider could accomplish it well.



3.1.7. *Fast Annotation*

For these strain sequences from the same or highly related species, their nucleotide identity is usually relatively higher. Therefore, the sequences annotation could be based on the gene information of the reference sequence after multi-sequence alignment.

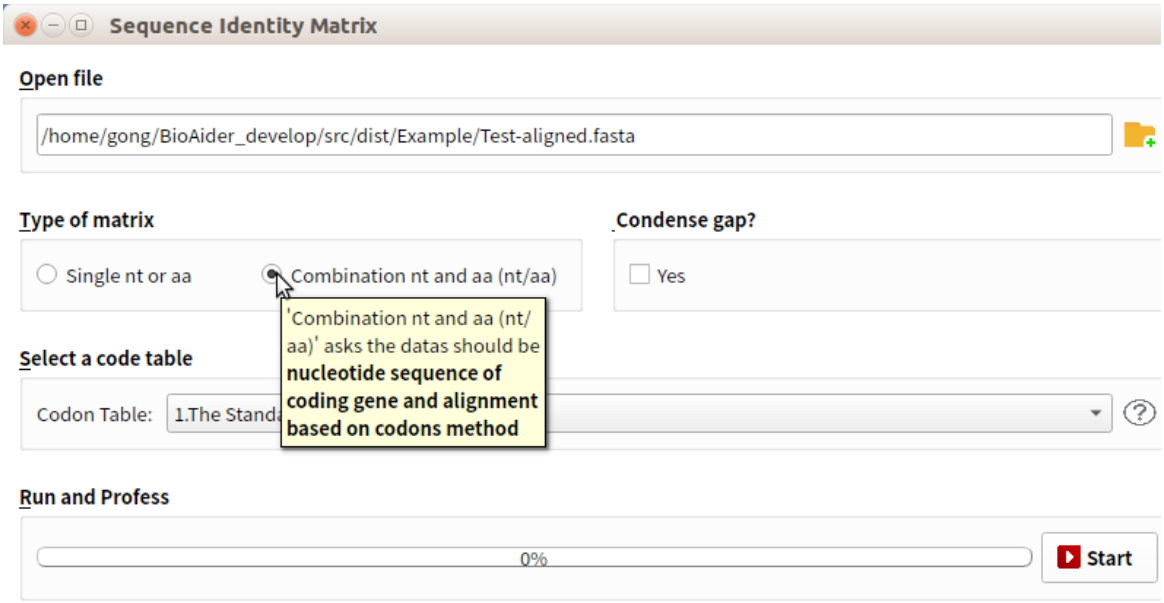


BioAider provides a quickly sequence annotation function, users can import the aligned complete genome sequence set (fasta format file), and adjust the reference sequence for annotation to the forefront of the file. Paste the gene information of reference sequence in aligned sets, name, starting string and end string into the ***textbox***, separated by ***","***. Then batch abstract genes.

Note that the start string or end string of the gene is not limited in length, but it is required to be unique in the reference sequence. Besides, the higher of similarity among sequences, the higher accuracy of the annotation.

3.2. *Similar Analysis*

3.2.1. *Sequence Identity Matrix*

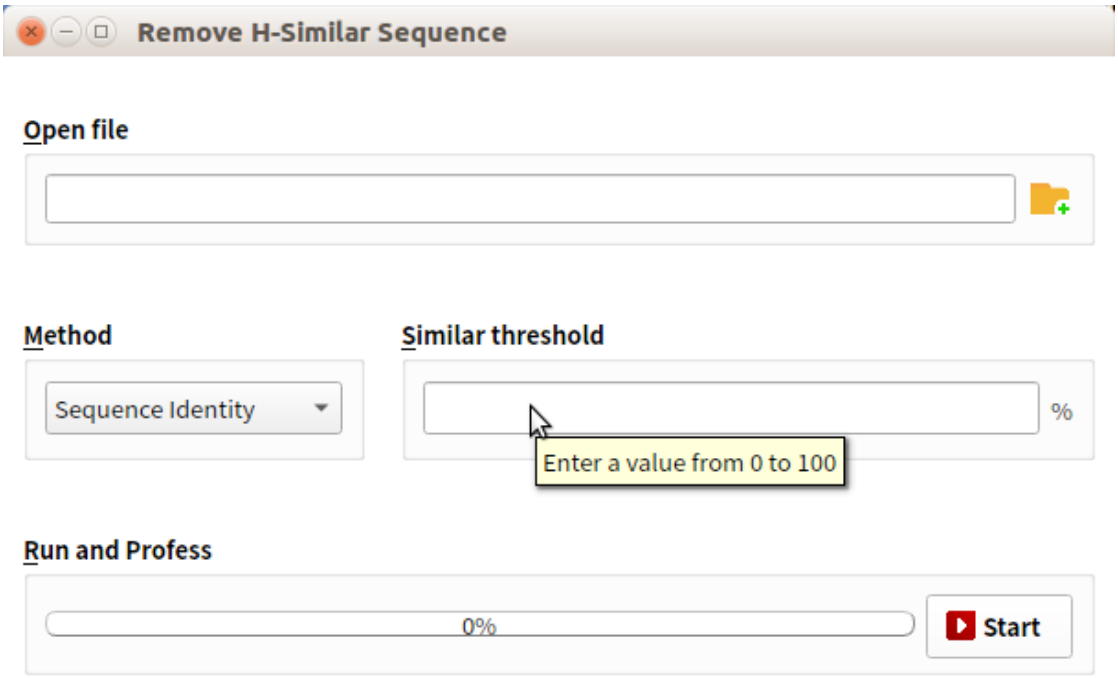


By inputting the aligned sequence datasets in fasta format, and a pairwise sequence identity matrix can be generated. This function contains two different modes: nucleotide or amino acid sequence identity matrix (***"Single nt or aa"***), nucleotide plus amino acid sequence identity matrix (***"Combination nt and aa"***).

It should be noted that if the ***"Combination nt and aa"*** is selected, the inputed sequences should be aligned based on codon method in advance. In order to better fit the variation characteristics , BioAider provides the ***"Condense gap"*** function. If the option was selected, the program will treat every three consecutive inserted or deleted bases as one.

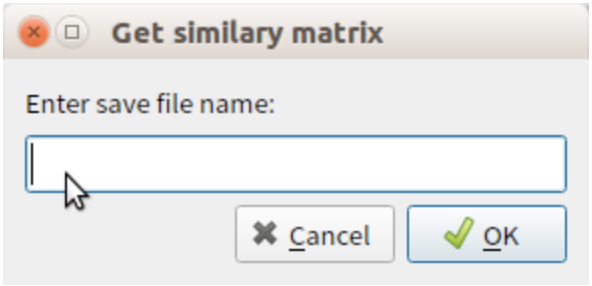
3.2.2 *Remove H-Similar Sequence*

This function could remove highly similar sequences and keep one by specifing the threshold of similarity (***"Similar threshold"***). BioAider provides 6 different methods for calculating the similarity of sequences.



It should be noted that the ***"Sequence Identity"*** and ***"Hamming"*** methods require the input sequences data are aligned, and we suggest that the sequences datasets for remaining 4 methods better not be pre-aligned, because these algorithm own alignment function. If the Similar threshold is seted to 100, the function of excluding repeated sequences will be turned on. "***Similar threshold"*** is seted to 100, the function of eliminateing duplicate sequences repeated sequences will be turned on.

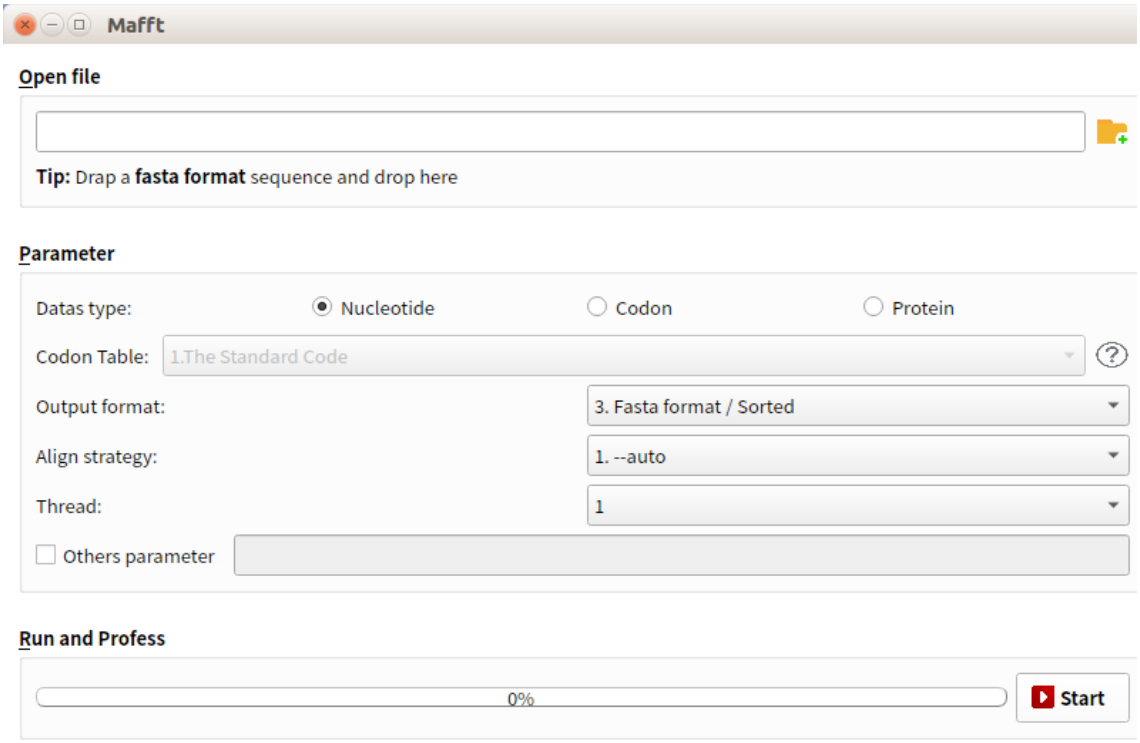
If you want to obtain the sequence similarity matrix calculated by the above 6 methods, you can click ***the right button of mouse*** in any region of the program interface to call up the functional menu.



3.3. *Align tools*

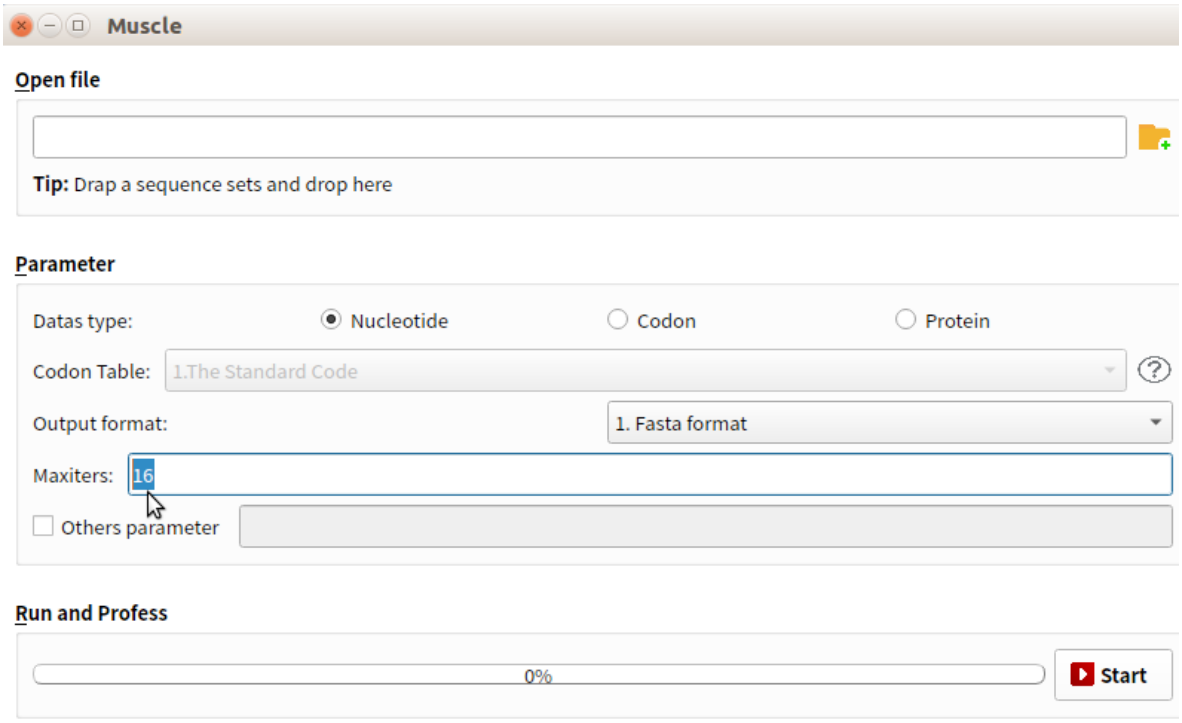
Multiple-Sequence-Alignment (MSA) is the most common analysis in sequence processing, most classic MSA software runs as a command symbol. It is very inconvenient for non-bioinformatics analysts. **BioAider packed three MSA software** (Mafft, Muscle and Clsutal-Omega) in the graphical interface, **and provided translation-alignment additionally based on multiple sets of codon tables.**

3.3.1. *Mafft*



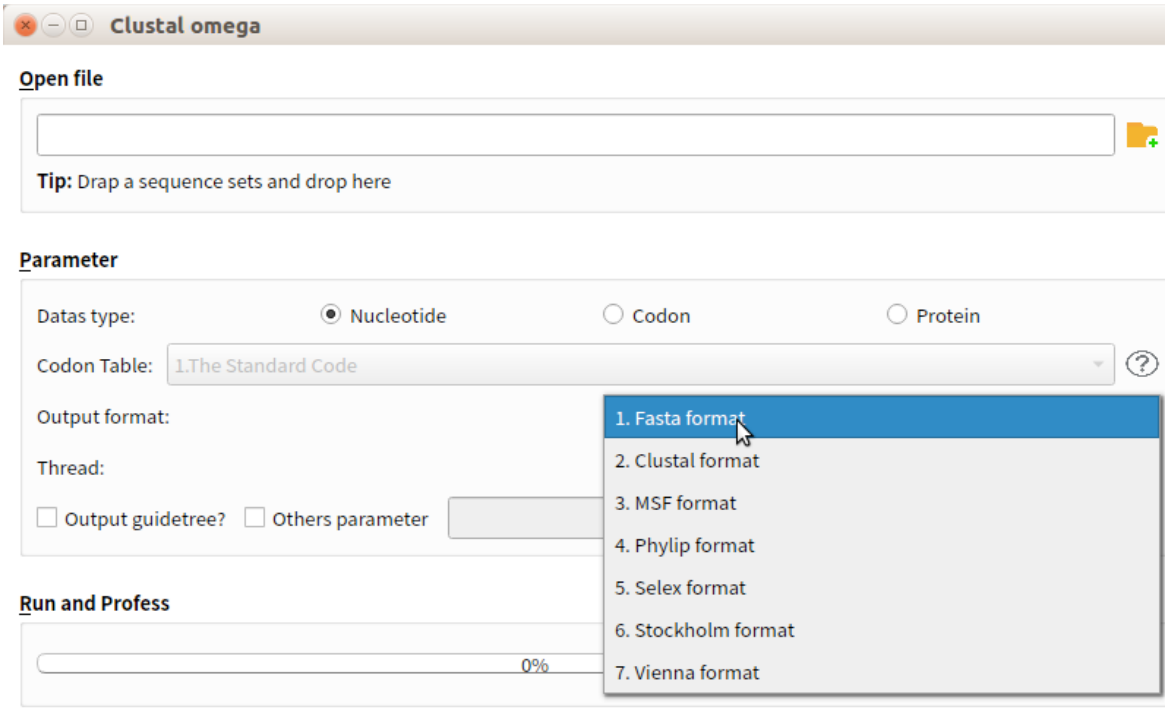
Mafft is a very popular MSA software with higher comparison accuracy, and its comparison speed is also relatively good. Some common parameter are encapsulated into the graphical interface in BioAider, and other parameters also could be add flexibly. More detailed information about Mafft could be got from https://mafft.cbrc.jp/alignment/software/.

3.3.2. *Muscle*



The comparison rate and accuracy of Muscle are good, according to the instruction manual of Muscle, setting ***"Maxiters"*** to 1 or 2 will significantly speed up the operation.
(Reference: http://petrov.stanford.edu/software/src/muscle3.6/muscle3.6.html)

3.3.2. *Clustal-Omeg*



As a relatively classic MSA software, *Clustal* has a broad user base. As the latest addition to the Clustal family. Clustal–Omega offers a significant increase in scalability over previous versions, more detailed reference http://www.clustal.org/omega/.

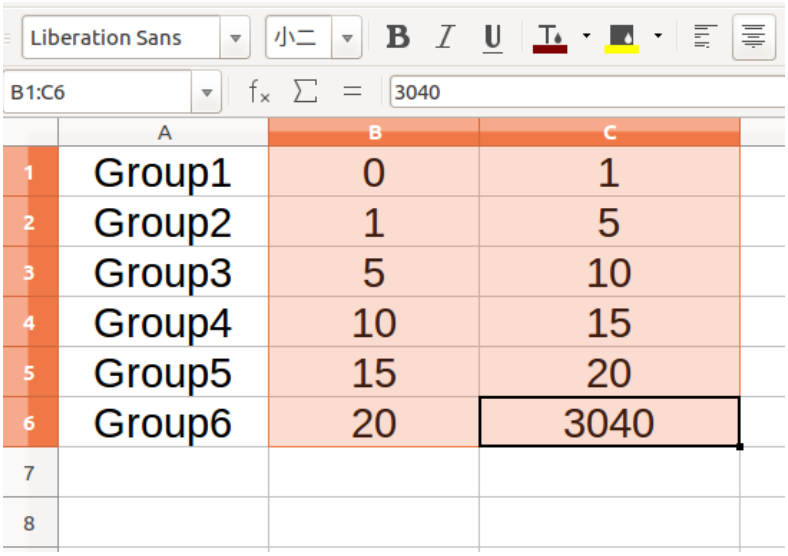
3.4. *Mutation Tools*

3.4.1. *Mutation Analysis*

This function could be used for analysis of the **mutations characteristicson on large numbers of sequenced strains.** The sequence datas for analysis needs to be aligned in advance, and they could be nucleotides, proteins (amino acid) sequences or simply coding gene fragments. For nucleotides and proteins sequences, BioAider could summarizes all the mutation sites with corresponding frequency and strains.

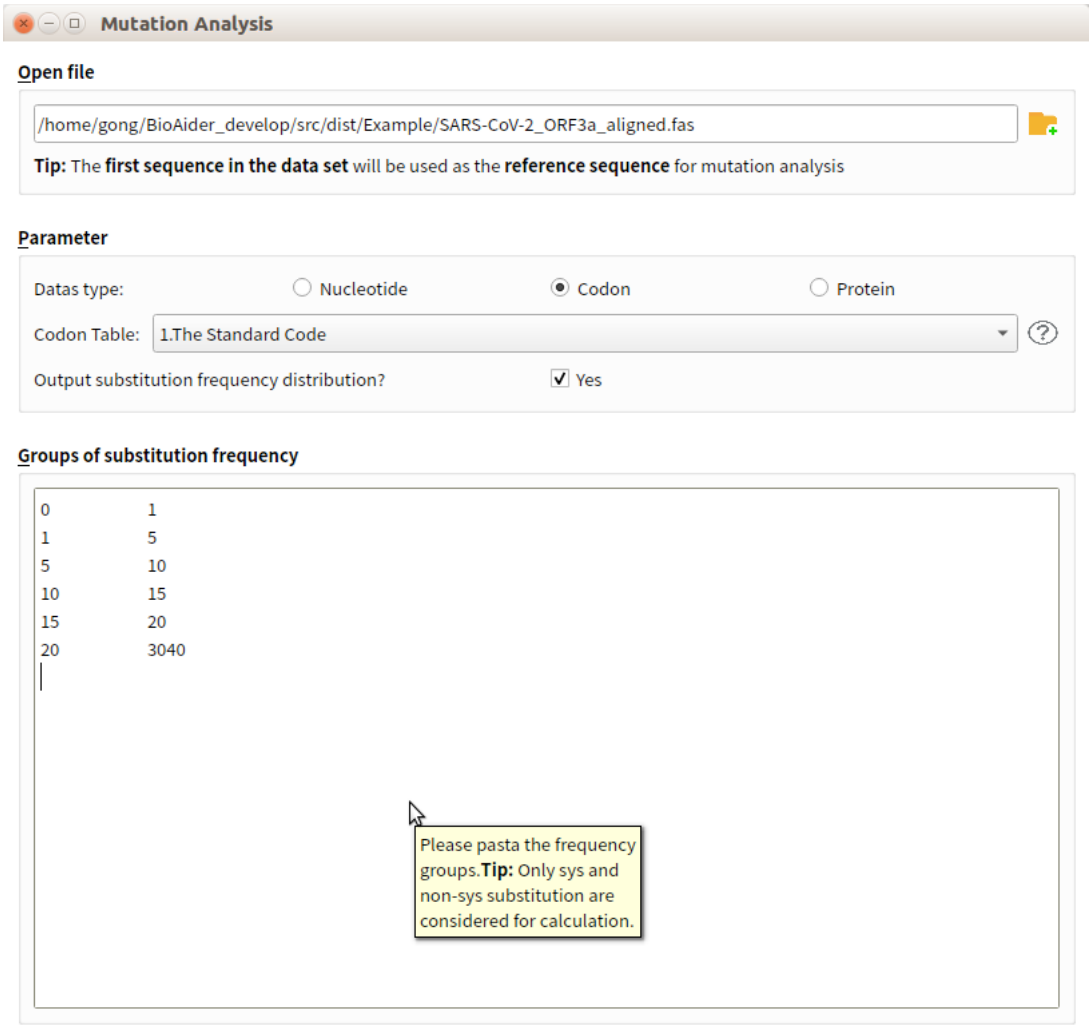
Of course, if the datas is codon gene, BioAider provides multiple sets of different codon tables for users, and could scan each condon sites in aligned sequence datasets, and identifies the type of mutation, including synonymous, non–synonymous, insertions and deletions and early termination. Finally, BioAider will automatically summarize and output the relevant analysis results.

Note: The codon gene sequences for mutations analysis have to be aligned by translation-alignment methon in advance, It is worth mentioning that BioAider packed three multiple-sequence-alignment software (mafft, muscle and clsutal-omega) in the graphical interface, and provided translation-alignment additionally.
Whether it's nucleotides or amino acids or coding genes, BioAider could plot the frequency distribution graph for mutation sites through specifing groups of substitution frequencey in custom.
Eaxmple of mutations analysis for aligned SARS-CoV-2 ORF3a gene sequences. First, create frequency grouping in a table editor:

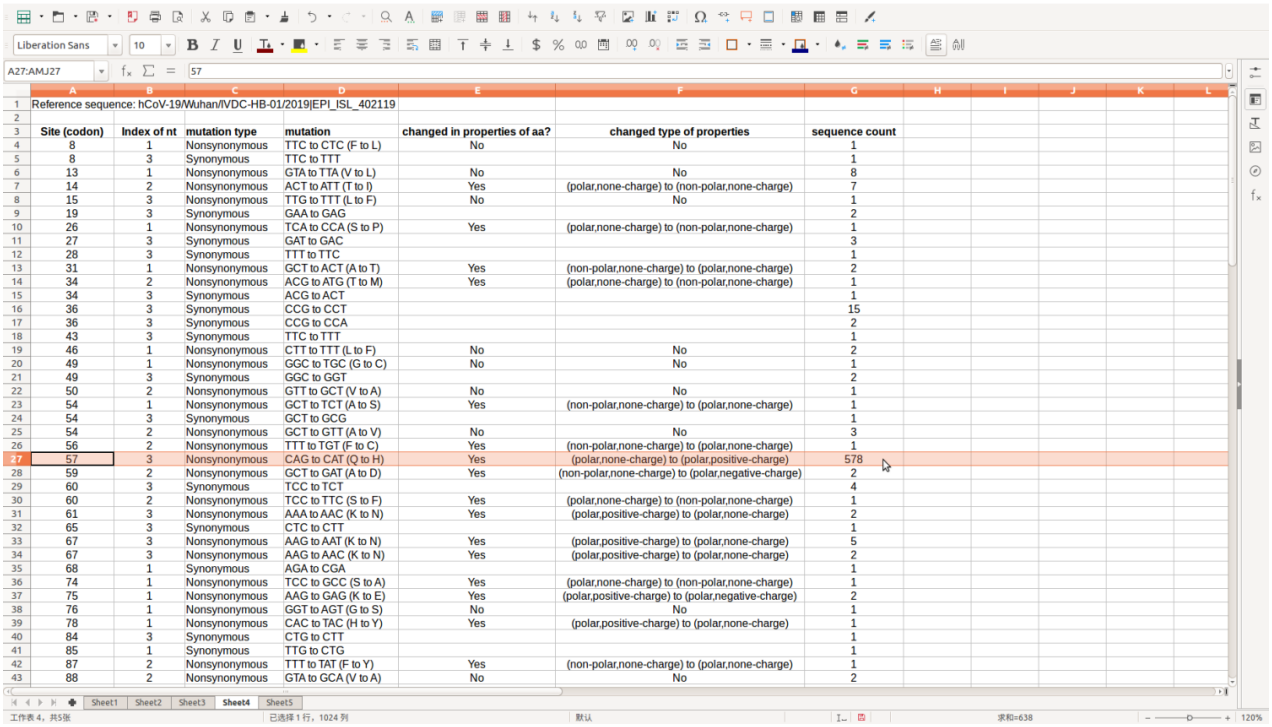


	A	B	C
1	Group1	0	1
2	Group2	1	5
3	Group3	5	10
4	Group4	10	15
5	Group5	15	20
6	Group6	20	3040
7			
8			

The each groups of substitution frequencey contains start value and end value which are separated by tab symbol. Note, ***the start value*** of each group is not included in the range of frequency.
Then copy them to the textedit box of BioAider,and select "Codon" single button in ***"Datas type"***:

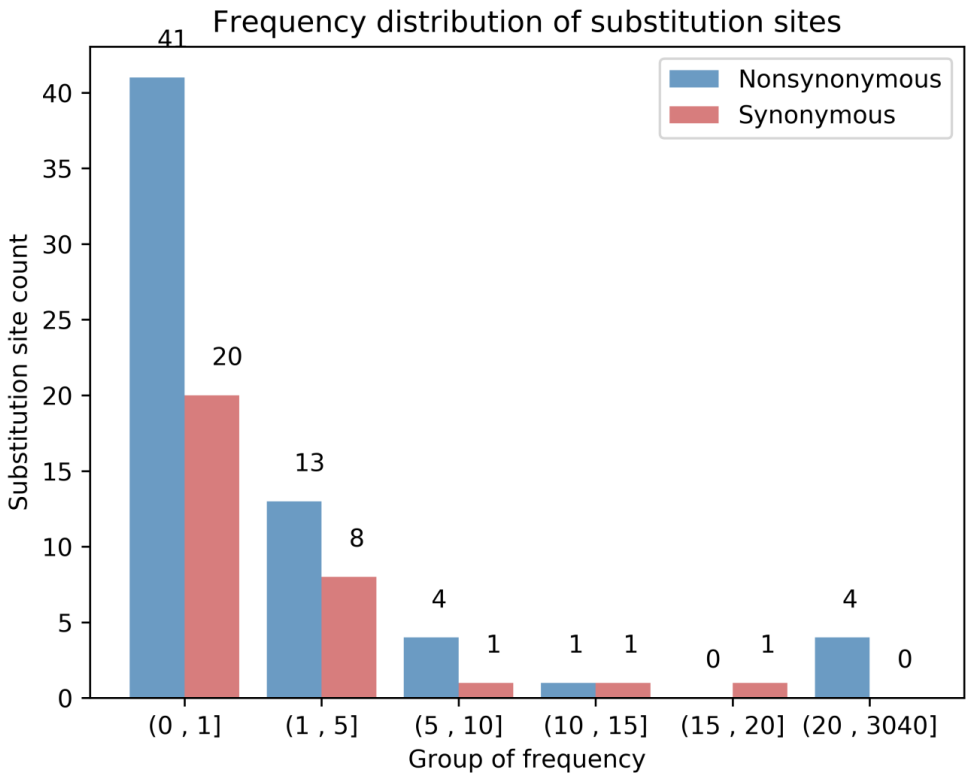


After the run is over, these analysis result could be found in the directory where the source file is located, you could scan the ****_mutation site summary*** file then know the overall variation and mutation hotspots.



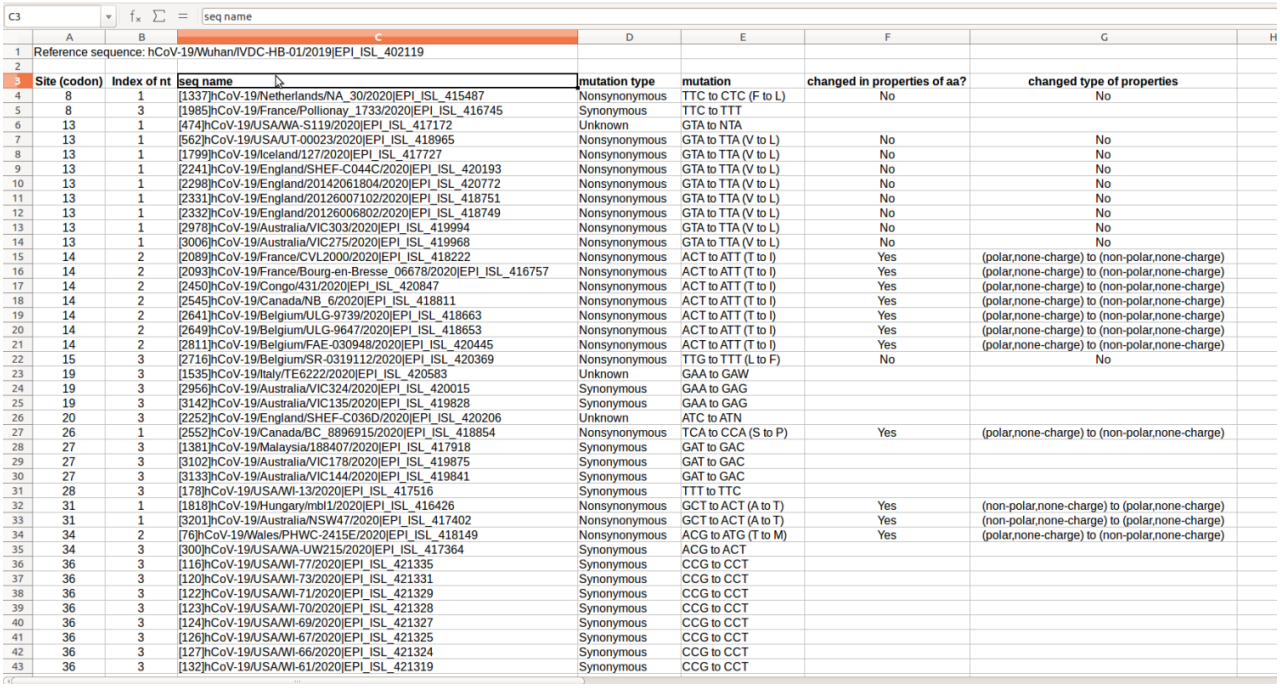
	A	B	C	D	E	F	G
1	Reference sequence: hCoV-19/Wuhan/IVDC-HB-01/2019|EPI_ISL_402119						
2							
3	Site (codon)	Index of nt	mutation type	mutation	changed in properties of aa?	changed type of properties	sequence count
4	8	1	Nonsynonymous	TTC to CTC (F to L)	No	No	1
5	8	3	Synonymous	TTC to TTT			1
6	13	1	Nonsynonymous	GTA to TTA (V to L)	No	No	8
7	14	2	Nonsynonymous	ACT to ATT (T to I)	Yes	(polar,none-charge) to (non-polar,none-charge)	7
8	15	3	Nonsynonymous	TTG to TTT (L to F)	No	No	1
9	19	3	Synonymous	GAA to GAG			2
10	26	1	Nonsynonymous	TCA to CCA (S to P)	Yes	(polar,none-charge) to (non-polar,none-charge)	1
11	27	3	Synonymous	GAT to GAC			3
12	28	3	Synonymous	TTT to TTC			1
13	31	1	Nonsynonymous	GCT to ACT (A to T)	Yes	(non-polar,none-charge) to (polar,none-charge)	2
14	34	2	Nonsynonymous	ACG to ATG (T to M)	Yes	(polar,none-charge) to (non-polar,none-charge)	1
15	34	3	Synonymous	ACG to ACT			1
16	36	3	Synonymous	CCG to CCT			15
17	36	3	Synonymous	CCG to CCA			2
18	43	3	Synonymous	TTC to TTT			1
19	46	1	Nonsynonymous	CTT to TTT (L to F)	No	No	2
20	49	1	Nonsynonymous	GGC to TGC (G to C)	No	No	1
21	49	3	Synonymous	GGC to GGT			2
22	50	2	Nonsynonymous	GTT to GCT (V to A)	No	No	1
23	54	1	Nonsynonymous	GCT to TCT (A to S)	Yes	(non-polar,none-charge) to (polar,none-charge)	1
24	54	3	Synonymous	GCT to GCG			1
25	54	2	Nonsynonymous	GCT to GTT (A to V)	No	No	3
26	56	2	Nonsynonymous	TTT to TGT (F to C)	Yes	(non-polar,none-charge) to (polar,none-charge)	1
27	57	3	Nonsynonymous	CAG to CAT (Q to H)	Yes	(polar,none-charge) to (polar,positive-charge)	578
28	59	2	Nonsynonymous	GCT to GAT (A to D)	Yes	(non-polar,none-charge) to (polar,negative-charge)	2
29	60	3	Synonymous	TCC to TCT			4
30	60	2	Nonsynonymous	TCC to TTC (S to F)	Yes	(polar,none-charge) to (non-polar,none-charge)	1
31	61	3	Nonsynonymous	AAA to AAC (K to N)	Yes	(polar,positive-charge) to (polar,none-charge)	2
32	65	3	Synonymous	CTC to CTT			1
33	67	3	Nonsynonymous	AAG to AAT (K to N)	Yes	(polar,positive-charge) to (polar,none-charge)	5
34	67	3	Nonsynonymous	AAG to AAC (K to N)	Yes	(polar,positive-charge) to (polar,none-charge)	2
35	68	1	Synonymous	AGA to CGA			1
36	74	1	Nonsynonymous	TCC to GCC (S to A)	Yes	(polar,none-charge) to (non-polar,none-charge)	1
37	75	1	Nonsynonymous	AAG to GAG (K to E)	Yes	(polar,positive-charge) to (polar,negative-charge)	2
38	76	1	Nonsynonymous	GGT to AGT (G to S)	No	No	1
39	78	1	Nonsynonymous	CAC to TAC (H to Y)	Yes	(polar,positive-charge) to (polar,none-charge)	1
40	84	3	Synonymous	CTG to CTT			1
41	85	1	Synonymous	TTG to CTG			1
42	87	2	Nonsynonymous	TTT to TAT (F to Y)	Yes	(non-polar,none-charge) to (polar,none-charge)	1
43	88	2	Nonsynonymous	GTA to GCA (V to A)	No	No	2

You could also konw the number of mutation sites under each mutation frequency group through view ****_substitution frequency distribution.png***:



It is not difficult to find that more than half of the mutation sites only appear in a single strain, although there are many mutation sites in ORF3a gene.

Or could obtain the corresponding mutant strain of these variant sites in the detailed ***_log.txt*** file:

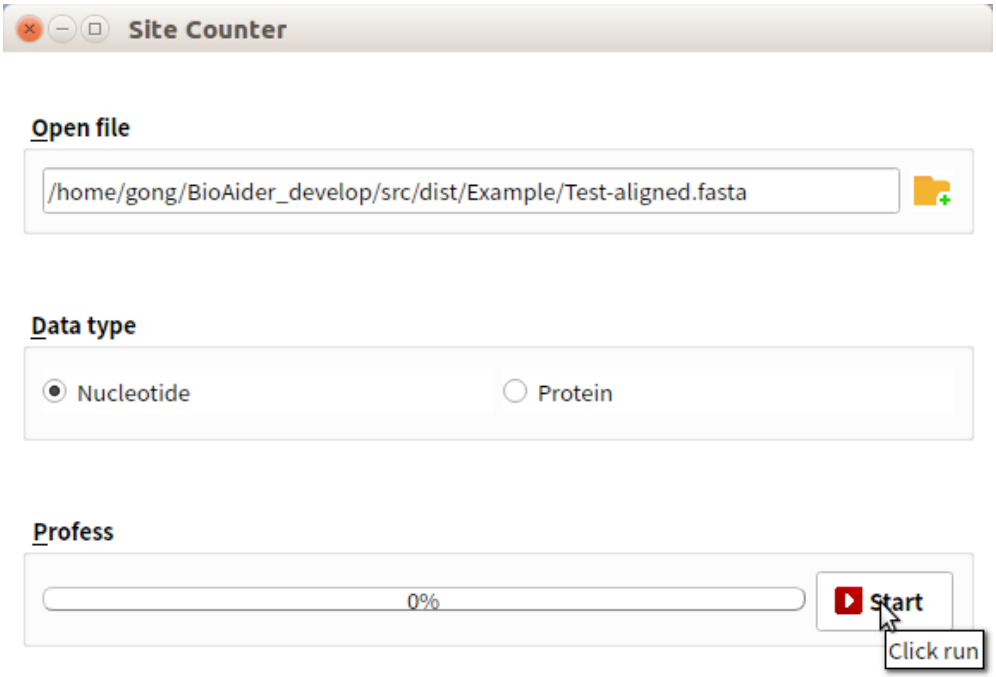


C3 | seq name

	A	B	C	D	E	F	G
1	Reference sequence: hCoV-19/Wuhan/IVDC-HB-01/2019|EPI_ISL_402119						
2							
3	**Site (codon)**	**Index of nt**	**seq name**	**mutation type**	**mutation**	**changed in properties of aa?**	**changed type of properties**
4	8	1	[1337]hCoV-19/Netherlands/NA_30/2020|EPI_ISL_415487	Nonsynonymous	TTC to CTC (F to L)	No	No
5	8	3	[1985]hCoV-19/France/Pollionay_1733/2020|EPI_ISL_416745	Synonymous	TTC to TTT		
6	13	1	[474]hCoV-19/USA/WA-S119/2020|EPI_ISL_417172	Unknown	GTA to NTA		
7	13	1	[562]hCoV-19/USA/UT-00023/2020|EPI_ISL_418965	Nonsynonymous	GTA to TTA (V to L)	No	No
8	13	1	[1799]hCoV-19/Iceland/127/2020|EPI_ISL_417727	Nonsynonymous	GTA to TTA (V to L)	No	No
9	13	1	[2241]hCoV-19/England/SHEF-C044C/2020|EPI_ISL_420193	Nonsynonymous	GTA to TTA (V to L)	No	No
10	13	1	[2298]hCoV-19/England/20142061804/2020|EPI_ISL_420772	Nonsynonymous	GTA to TTA (V to L)	No	No
11	13	1	[2331]hCoV-19/England/20126007102/2020|EPI_ISL_418751	Nonsynonymous	GTA to TTA (V to L)	No	No
12	13	1	[2332]hCoV-19/England/20126006802/2020|EPI_ISL_418749	Nonsynonymous	GTA to TTA (V to L)	No	No
13	13	1	[2978]hCoV-19/Australia/VIC303/2020|EPI_ISL_419994	Nonsynonymous	GTA to TTA (V to L)	No	No
14	13	1	[3006]hCoV-19/Australia/VIC275/2020|EPI_ISL_419968	Nonsynonymous	GTA to TTA (V to L)	No	No
15	14	2	[2089]hCoV-19/France/CVL2000/2020|EPI_ISL_418222	Nonsynonymous	ACT to ATT (T to I)	Yes	(polar,none-charge) to (non-polar,none-charge)
16	14	2	[2093]hCoV-19/France/Bourg-en-Bresse_06678/2020|EPI_ISL_416757	Nonsynonymous	ACT to ATT (T to I)	Yes	(polar,none-charge) to (non-polar,none-charge)
17	14	2	[2450]hCoV-19/Congo/431/2020|EPI_ISL_420847	Nonsynonymous	ACT to ATT (T to I)	Yes	(polar,none-charge) to (non-polar,none-charge)
18	14	2	[2545]hCoV-19/Canada/NB_6/2020|EPI_ISL_418811	Nonsynonymous	ACT to ATT (T to I)	Yes	(polar,none-charge) to (non-polar,none-charge)
19	14	2	[2641]hCoV-19/Belgium/ULG-9739/2020|EPI_ISL_418663	Nonsynonymous	ACT to ATT (T to I)	Yes	(polar,none-charge) to (non-polar,none-charge)
20	14	2	[2649]hCoV-19/Belgium/ULG-9647/2020|EPI_ISL_418653	Nonsynonymous	ACT to ATT (T to I)	Yes	(polar,none-charge) to (non-polar,none-charge)
21	14	2	[2811]hCoV-19/Belgium/FAE-030948/2020|EPI_ISL_420445	Nonsynonymous	ACT to ATT (T to I)	Yes	(polar,none-charge) to (non-polar,none-charge)
22	15	3	[2716]hCoV-19/Belgium/SR-0319112/2020|EPI_ISL_420369	Nonsynonymous	TTG to TTT (L to F)	No	No
23	19	3	[1535]hCoV-19/Italy/TE6222/2020|EPI_ISL_420583	Unknown	GAA to GAW		
24	19	3	[2956]hCoV-19/Australia/VIC324/2020|EPI_ISL_420015	Synonymous	GAA to GAG		
25	19	3	[3142]hCoV-19/Australia/VIC135/2020|EPI_ISL_419828	Synonymous	GAA to GAG		
26	20	3	[2252]hCoV-19/England/SHEF-C036D/2020|EPI_ISL_420206	Unknown	ATC to ATN		
27	26	1	[2552]hCoV-19/Canada/BC_8896915/2020|EPI_ISL_418854	Nonsynonymous	TCA to CCA (S to P)	Yes	(polar,none-charge) to (non-polar,none-charge)
28	27	3	[1381]hCoV-19/Malaysia/188407/2020|EPI_ISL_417918	Synonymous	GAT to GAC		
29	27	3	[3102]hCoV-19/Australia/VIC178/2020|EPI_ISL_419875	Synonymous	GAT to GAC		
30	27	3	[3133]hCoV-19/Australia/VIC144/2020|EPI_ISL_419841	Synonymous	GAT to GAC		
31	28	3	[178]hCoV-19/USA/WI-13/2020|EPI_ISL_417516	Synonymous	TTT to TTC		
32	31	1	[1818]hCoV-19/Hungary/mbl1/2020|EPI_ISL_416426	Nonsynonymous	GCT to ACT (A to T)	Yes	(non-polar,none-charge) to (polar,none-charge)
33	31	1	[3201]hCoV-19/Australia/NSW47/2020|EPI_ISL_417402	Nonsynonymous	GCT to ACT (A to T)	Yes	(non-polar,none-charge) to (polar,none-charge)
34	34	2	[76]hCoV-19/Wales/PHWC-2415E/2020|EPI_ISL_418149	Nonsynonymous	ACG to ATG (T to M)	Yes	(polar,none-charge) to (non-polar,none-charge)
35	34	3	[300]hCoV-19/USA/WA-UW215/2020|EPI_ISL_417364	Synonymous	ACG to ACT		
36	36	3	[116]hCoV-19/USA/WI-77/2020|EPI_ISL_421335	Synonymous	CCG to CCT		
37	36	3	[120]hCoV-19/USA/WI-73/2020|EPI_ISL_421331	Synonymous	CCG to CCT		
38	36	3	[122]hCoV-19/USA/WI-71/2020|EPI_ISL_421329	Synonymous	CCG to CCT		
39	36	3	[123]hCoV-19/USA/WI-70/2020|EPI_ISL_421328	Synonymous	CCG to CCT		
40	36	3	[124]hCoV-19/USA/WI-69/2020|EPI_ISL_421327	Synonymous	CCG to CCT		
41	36	3	[126]hCoV-19/USA/WI-67/2020|EPI_ISL_421325	Synonymous	CCG to CCT		
42	36	3	[127]hCoV-19/USA/WI-66/2020|EPI_ISL_421324	Synonymous	CCG to CCT		
43	36	3	[132]hCoV-19/USA/WI-61/2020|EPI_ISL_421319	Synonymous	CCG to CCT		

3.4.2. *Site Counter*

This function could summary the type, count and proportion of bases (or amino acids) at each site for the aligned sequence datasets. In addition, BioAider will output a consensus sequence based on the highest proportion base (or amino acid) in each site.

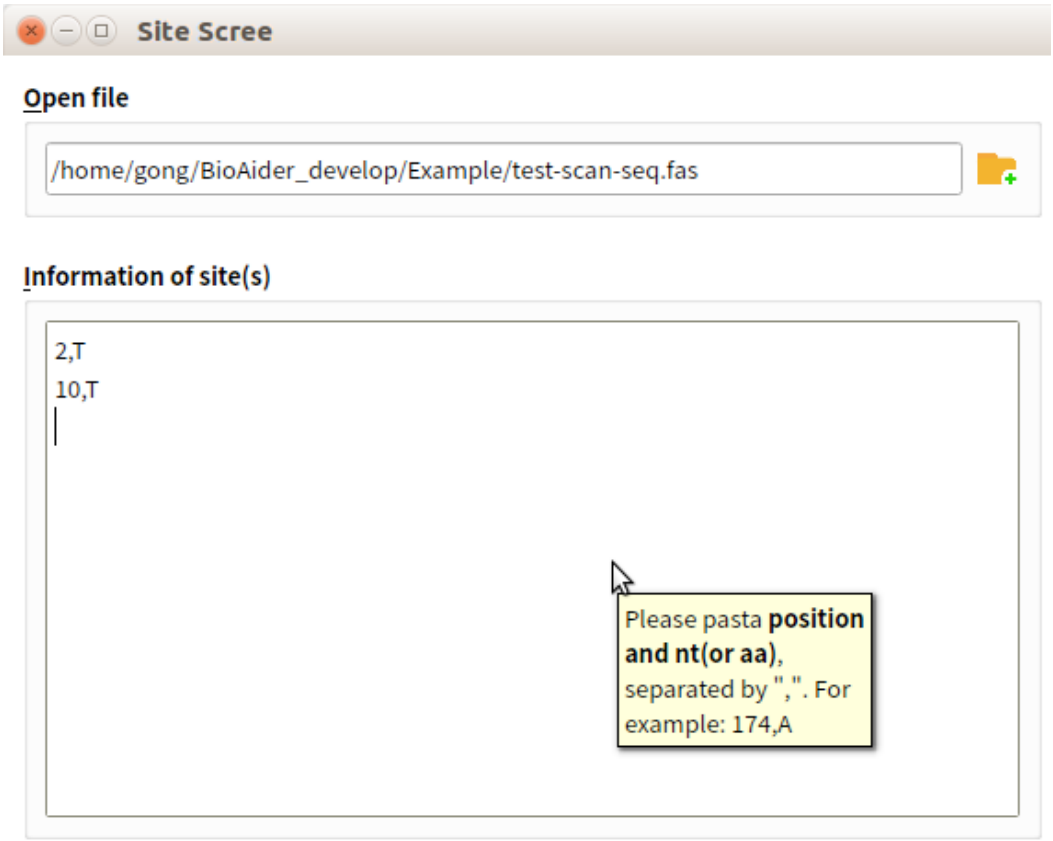


For DNA sequence datasets, the one of results (***_site_count.csv***)was as follows:

I12

	A	B	C	D	E	F	G	H
1		site1	site2	site3	site4	site5	site6	site7
2	A	29	0	0	0	0	0	0
3	G	0	0	29	0	0	0	29
4	C	0	0	0	0	0	0	0
5	T	0	29	0	29	29	29	0
6	-	0	0	0	0	0	0	0
7								
8								

3.4.3. *Site Scree*



This function is used to extract the sequences with corresponding base (or amino acid) in ***specified one or more*** site. It is very useful for studying whether there is linkage inheritance among different gene sites.

4. Test Datas

Examples and test are available at https://github.com/ZhijianZhou01/BioAider/tree/master/Example.