Manual

fasta_feature.pl

This script is designed to investigate some common features of a sequence file. After installing Perl in your local computer, the script can be executed by entering "perl -w fasta_feature.pl" in linux command line, and the main menu will show up like (Figure 1).

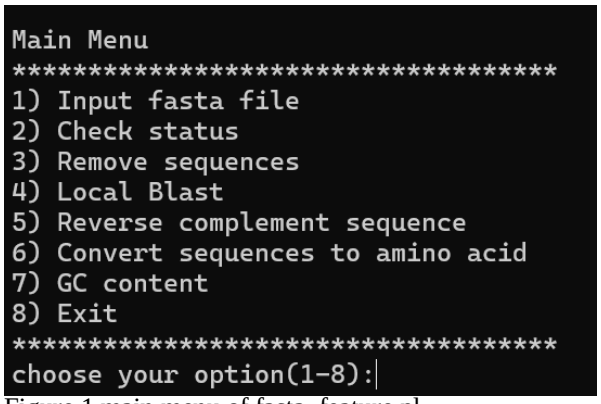

Figure 1 main menu of fasta_feature.pl

The script has the following functions:

1. Input fasta file:

To run the rest function, inputting fasta file is the first thing to do. You have to put the fasta file in the same directory as the script, and you can type the name of the fasta file (no need to type ".fasta"). Once you finishing this step, enter "F" to end the process and go back to main menu. What's worth noting is that if you reenter to input data again, you can decide if you want to clear the previous data or not.

2. Check status:

After inputting data, you can enter 2 to check what you have input. It will show the name and the length of each sequence.

3. Remove sequences:

If you want to remove certain sequences, you can input a .txt file with the name of sequence you want to delete, and those sequence will be delete from the input data.

4. Local blast:

This function requires the installation of BLAST+ (Index of /blast/executables/blast+/LATEST (nih.gov)). You can enter the index number of the query sequence, and enter the file name of reference fasta file in order to make the local database. You will get the result in a new directory called blast_output_dir.

5. Reverse complement sequence:

The reverse complement sequences are important for designing primers and the aligning reads process. Input sequences will be transformed with the rule of A→T, T →A, C→G, G→C. Then, the sequence will be reversed.

6. Convert sequences to amino acid:

You can type in the number -3 ~ 3 to decide which opening reading frame (ORF) to convert. The results will show up in command line.

7. GC content:

GC content of a sequence may reflect its stability. The function will return the GC content of each sequence.

fasta_transformation.pl

the script is design to adjust the fasta files. Some sequence will have multiple lines, and this script can allow one to group multiple lines of a single sequence together in a single line for the convenience of data analysis. The -f parameter represents the file name, and the -o parameter is the output file name of the new fasta file.