**change GCA filename**

This script is for change the filename downloaded from NCBI database, it can process multiple files in one command line. first it will unzip all .gz file in all directories, and change all GCA filename with its organism name and isolate respectively, and it’s also create a table to record the accession number, organism name and the isolate.

**reliance**

pandas

fnmatch

gzip

subprocess

shutil

**parameter**

It’s just need you provide an absolute path of directory contain all folder downed from NCBI database.

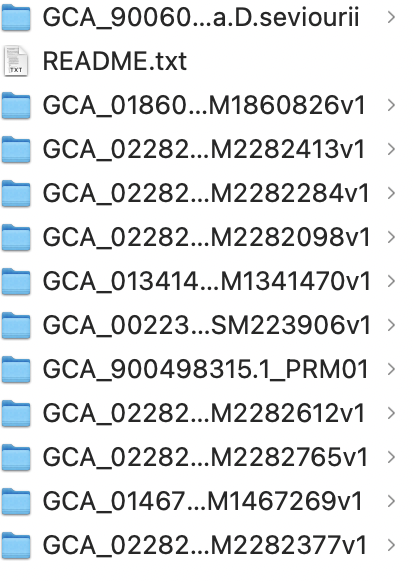
please use parameter --input /User/\*

**input example**

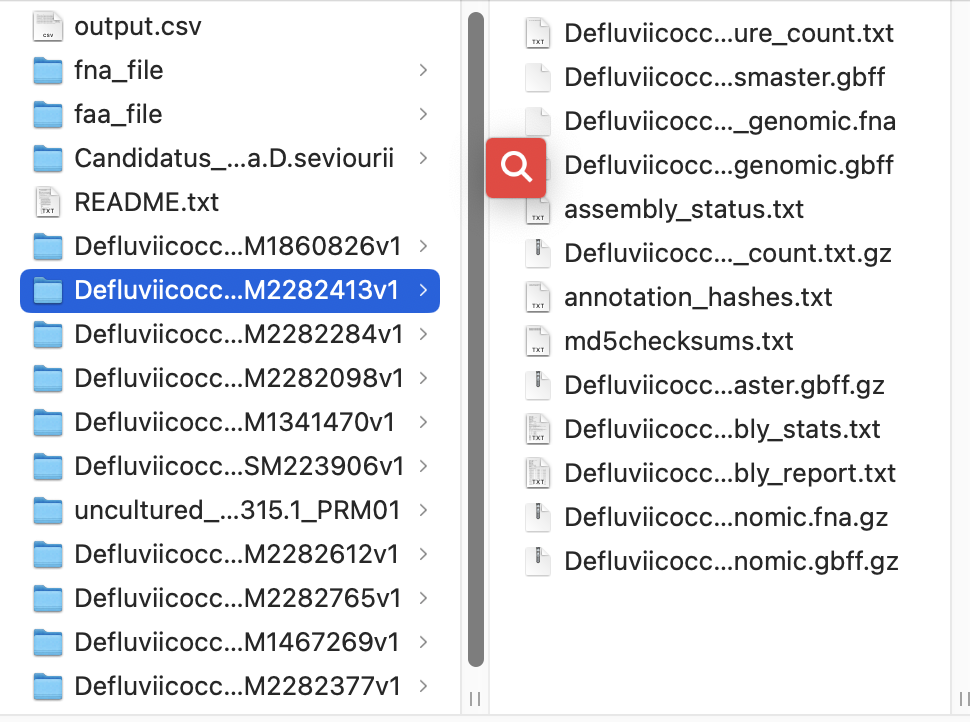
python rename.py --input /User/Xuhan\_Deng/test

**image example**

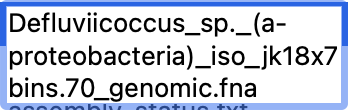
input folder just like this



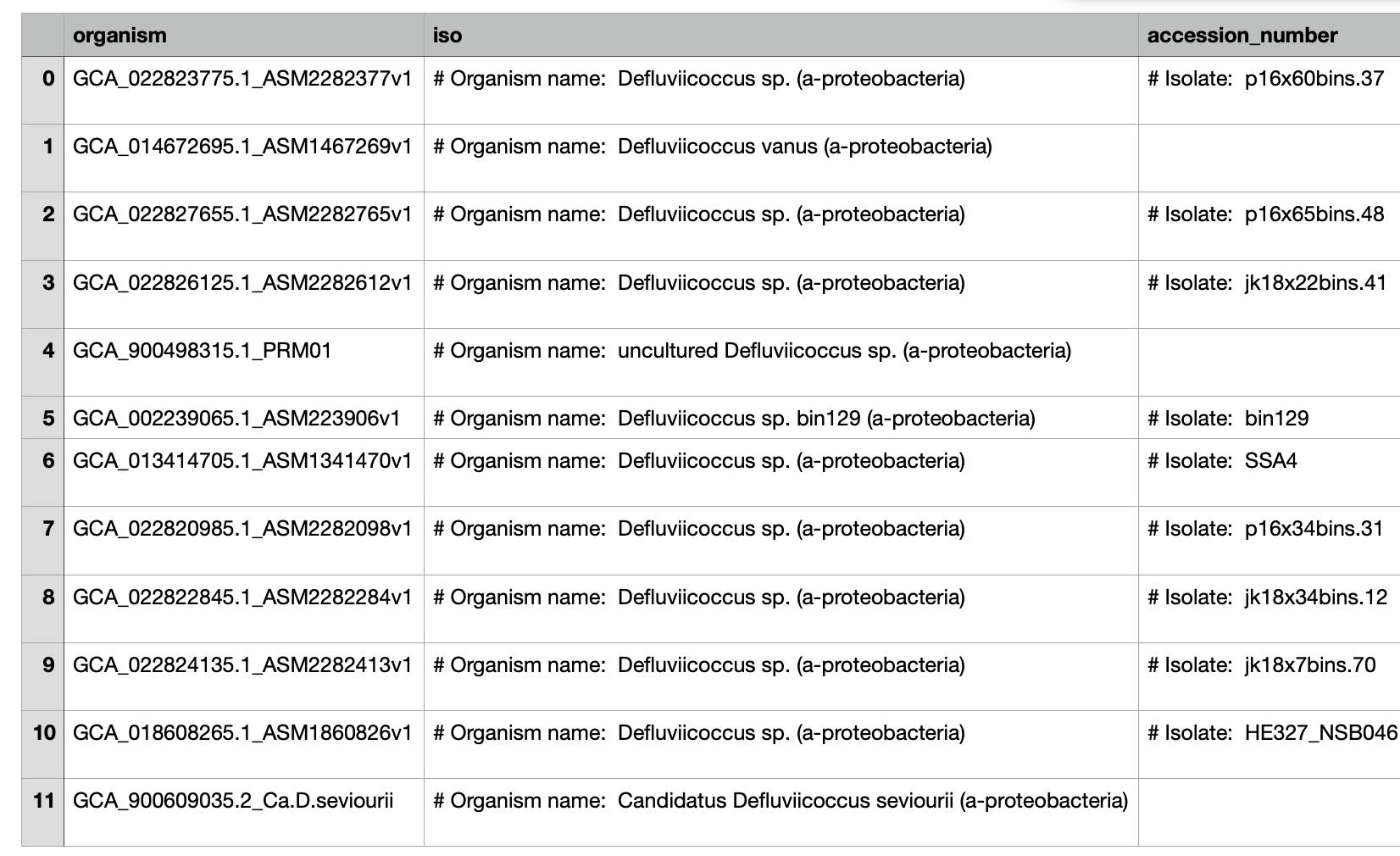
After processing it will generate these



the file name is changed to this pattern



it also generates a table to record the accession number, isolate, organism



for next step analysis, it also retrieve all fna sequence file to faa\_file directory and faa protein file to faa\_file

Graphical user interface, text, application, email

Description automatically generated