**Программа для создания csv-файла по записям из полей CDS, genes записей GenBank**

Ссылка на GitHub: <https://github.com/xfluffymuffin/astrovirus.git>

Описание:

Программа довольно проста как по задумке, так и по исполнению; использует всего один модуль subprocess.

Последовательно запускаются две функции:  
1) *launch\_gene\_cds\_field\_distribution()*

Функция запускает скрипт *gene\_cds\_field\_distribution.py* с помощью следующей внутренней команды:

python genecds\_field\_distribution.py -input Mamastrovirus\_1\_complete\_genome\_records.gb -odir alignment\_of\_orfs\_output -oname Mamastrovirus\_1\_gene\_product\_output.txt

Выходной файл такой же, как и у запускаемого изнутри скрипта: *Mamastrovirus\_1\_gene\_product\_output.txt*

2) *gene\_cds\_field\_distribution\_output\_handle()*

Функция непосредственно создает необходимый csv-файл на основании выходного файла предыдущей функции

Если строка поля CDS/gene – это слово orf1/orf1a, либо в строке имеется “unknown”, “protease”, “NS”, то после запятой добавляется “1A”.

Строка – “orf1b” или в строке “polymerase” – добавляется “1B”/

В строке “Non” – добавляется “1AB”.

В строке “non” – добавляется “1A”, “1B” и “1AB”.

Иначе добавляется “2”.

Скрипт *genecds\_field\_distribution.py* и файл с GenBank-записями *Mamastrovirus\_1\_complete\_genome\_records.gb* должны находиться в той же директории, что и сам *genecds\_create\_csv.py*.