#!/bin/usr/python3

# Requires Biopython, pandas

from Bio import SeqIO

import pandas as pd

from glob import glob

import os

inp\_folder="./Mackay\_MAGs/"

# Pattern that doesn't include the summary file!

file\_pattern="abricate\*0.tsv"

op\_folder="./Mackay\_MAGs\_abricate\_all\_gff\_folders/"

if os.path.exists("%s/" % op\_folder)==False:

os.mkdir("%s/" % op\_folder)

for inp\_file in glob("%s\*%s" % (inp\_folder,file\_pattern)):

print(inp\_file)

subop\_folder=inp\_file.split('/')[-1].split('.')[0]

if os.path.exists("%s/%s" % (op\_folder, subop\_folder))==False:

os.mkdir("%s/%s" % (op\_folder, subop\_folder))

abr\_df=pd.read\_csv(inp\_file, sep='\t')

filename\_list=list(set(abr\_df['#FILE'].to\_list()))

for file in filename\_list:

sample\_name=file.split("/")[-1].split(".")[0]

sample\_op\_file="%s/%s/%s.gff" % (op\_folder,subop\_folder,sample\_name)

sample\_df=abr\_df[abr\_df['#FILE']==file]

#open(sample\_op\_file, 'w')

sample\_handle=open(sample\_op\_file, 'w')

#new\_record=SeqRecord('', id=

for row in sample\_df.itertuples(index=False):

contig=str(row[1])

source=str(row[11])

feature=str(row[13])

start=str(row[2])

end=str(row[3])

score="."

strand=str(row[4])

phase="."

attribute="Name=%s;ID=%s;product=%s;coverage=%i;identity=%i" % (row[5],row[12],row[14],row[9],row[10])

new\_row="\t".join([contig,source,feature,start,end,score,strand,phase,attribute])

sample\_handle.write(new\_row)

sample\_handle.write('\n')

sample\_handle.close()