#!/usr/bin/python3

“””

Convert antismash (v6) full gbk files to gff

This script puts the gff files into the folder that it found the gbk file,

so remember to copy to a combined gff folder if required (bash):

Eg.for item in \*\_antismash\_all\_options; do cp ${item}/\*.gff Mackay\_MAGs\_antismash\_gffs/;done

“””

from Bio import SeqIO

from Bio.SeqRecord import SeqRecord

from BCBio import GFF

from glob import glob

antismash\_folder="./"

# If the antismash output folders are named something other than “sample/”

# Assumes form <sample><label>

# Can be empty string

antismash\_label="\_antismash\_all\_options"

for sample\_folder in glob("%s/\*%s\*" % (antismash\_folder,antismash\_label)):

if sample\_folder[-1]=="/":

sample\_name=sample\_folder.replace(antismash\_label, '').split('/')[-2]

elif sample\_folder[-1]!="/":

sample\_name=sample\_folder.replace(antismash\_label, '').split('/')[-1]

full\_antismash\_gbk="%s/%s.gbk" % (sample\_folder, sample\_name)

inp\_handle=open(full\_antismash\_gbk, "rU")

op\_handle=open("%s/%s.gff" % (sample\_folder, sample\_name), "w")

print("Now working on %s...\n" % (sample\_name))

new\_data=[]

for rec in SeqIO.parse(inp\_handle, "genbank"):

new\_rec = SeqRecord(rec.seq, rec.id)

clusters=[]

for feat in rec.features:

# This may depend on AS version

if feat.type == "region":

clusters.append(feat)

new\_rec.features=clusters

if len(new\_rec.features) >= 1:

new\_data.append(new\_rec)

#GFF.write(SeqIO.parse(inp\_handle, "genbank"), op\_handle)

#recs=[rec for rec in SeqIO.parse(inp\_handle, "genbank")]

#feats=[feat for feat in rec.features if feat.type == "protocluster"]

GFF.write(new\_data, op\_handle)