Sep 23, 2022

fastANI analysis protocol

Jamie Harrison[1], David J Studholme[1]

[1]University of Exeter

Jamie Harrison

ABSTRACT

this is the protocol to conduct ani analysis between groups of genomes using fastANI and produce the heatmap figure in R using the pheatmap package

PROTOCOL CITATION

Jamie Harrison, David J Studholme 2022. fastANI analysis protocol. **protocols.io** https://protocols.io/view/fastani-analysis-protocol-cgritv4e

CREATED

Sep 16, 2022

LAST MODIFIED

Sep 23, 2022

PROTOCOL INTEGER ID

70154

1 Create directory structure and link files

create directories

1.1

For this there there needs to be two files, one listing query genomes and one listing reference genomes, each with a single filename per line. Suggested names for these files = query_list.txt, reference_list.txt

1.2 mkdir analysis query_genomes reference_genomes

Create directories

mkdir analysis query_genomes, reference_genomes

create dirs for each of analysis, query genomes and reference genomes

1.3 move to query dir

move to reference genomes dir

cd ../reference_genomes

change dir to reference_genomes dir

1.4

Copy reference_list.txt

cp ZZZZZZZZ/reference_list.txt .

Copy reference_list.txt to reference_genomes dir, where ZZZZZZZZ is location of reference_list.txt

1.5

link reference genomes to reference_genomes dir

while read i; do ln -s YYYYYYYY/${i} .;done < reference_list.txt

loop through reference_list.txt and link each reference genomes to reference_genomes dir. YYYYYYYY is the location of the genome fasta files

1.6 check soft links and remove bad links

check softlinks and remove any that don't work

find . -xtype l | xargs rm

check softlinks and remove any that don't work

1.7

create query genome list

ls *fasta > query_list.txt

1.8 move to reference dir

move to reference genomes dir

cd ../reference_genomes

change dir to reference_genomes dir

1.9

Copy reference_list.txt

cp ZZZZZZZZ/reference_list.txt .

Copy reference_list.txt to reference_genomes dir, where ZZZZZZZZ is location of reference_list.txt

1.10

link reference genomes to reference_genomes dir

while read i; do ln -s YYYYYYYY/${i} .;done < reference_list.txt

loop through reference_list.txt and link each reference genomes to reference_genomes dir. YYYYYYYY is the location of the genome fasta files

1.11

check softlinks and remove any that don't work

find . -xtype l | xargs rm

check softlinks and remove any that don't work

1.12

create query genome list

ls *fasta > query_list.txt

1.13

move to analysis dir

cd ../analysis

move to directory to be used for the analysis step

1.14

softlink all necessary files to analysis directory

ln -s ../reference_genomes/ref_list.txt .
ln -s ../reference_genomes/*fasta .
ln -s ../query_genomes/query_list.txt .
ln -s ../query_genomes/query_list.txt .

softlink all necessary files to the analysis directory to be used in the fastANI analysis

2

Perform fastANI analysis

fastANI --rl reflist.txt --ql querylist.txt --matrix -o fastANI_out

perform the fastANI analysis of the query genomes vs the reference genomes
-rl specifies reference list of genomes
-ql specifies query genome list
-matrix outputs a bottom half triangular matrix of results
-o specifies output file prefix

this step can also be submitted to job queue on HPC cluster.

3

reformat fastANI output

git clone https://github.com/jh288/fastANI_reformatter.git

the output of fastANI is not in a suitable format to produce figure but this is addressed with a simple script available from github

3.1

reformat fastANI output for figure prep

fastANI_reformater.pl fastANI_out > fastANI_out_reformat.tab

reformat fastANI output for use in the R package pheatmap to create the figure

4

produce heatmap figure in R

4.1

edit matrix tab file

sed 's/.fasta//g' fastANI_out_reformat.tab > fastANI_out_reformat_ed.tab

sed -i 's/_/ /g' fastANI_out_reformat_ed.tab

remove ".fasta" and substitute spaces for underscores in taxa names

4.2

produce heatmap in R

```
##load libraries

library("pheatmap")

library("RColorBrewer")

##load matrix into dataframe
b1<-read.delim("fastANI_out_reformat_ed.tab", header =T, row.names=1, check.names=F)

##set output parameters
png("fastANI_out_reformat_ed.png", height=1500, width=750)

###run pheatmap
pheatmap(t(as.matrix(b1)), color = brewer.pal(n = 7, name ="Blues"), display_numbers=T, number_format = "%.2f", number_color="black")

dev.off()
```

r code to produce heatmap of fastANI results.