

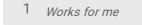


Sep 23, 2022

# fastANI analysis protocol

Jamie Harrison<sup>1</sup>, David J Studholme<sup>1</sup>

<sup>1</sup>University of Exeter





This protocol is published without a DOI.

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

**LICENSE** 

This is an open access protocol distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

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 ZZZZZZZ/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 In -s YYYYYYY/\${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 I | 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 ZZZZZZZ/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 In -s YYYYYYY/ $\{i\}$ .;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 | | 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

In -s ../reference genomes/ref list.txt .

In -s ../reference genomes/\*fasta.

In -s ../query\_genomes/query\_list.txt .

In -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.
```