Aug 26, 2024

Annotation for Fungi

RESERVED DOI:
10.17504/protocols.io.e6nvw14nwlmk/v1

Sebastian Bassi[1], Claudia V. Gonzalez[1], Tristan Yang[1]

[1]Toyoko

Sebastian Bassi
Toyoko

OPEN ACCESS

Protocol Citation: Sebastian Bassi, Claudia V. Gonzalez, Tristan Yang 2024. Annotation for Fungi. **protocols.io** **https://protocols.io/view/annotation-for-fungi-dj6n4rde**

Protocol status: In development
We are still developing and optimizing this protocol

Created: August 26, 2024

Last Modified: August 26, 2024

Protocol Integer ID: 106414

Keywords: docker, bioinformatics, dnalinux, fungi

Abstract

Protocol to annotate a fungi genome

Setup

1 Install Docker

If you don't have Docker already, install it. There are two versions, Docker Engine (also known as CE) and Docker Desktop. The Desktop version is more user friendly but since may require commercial license for large enterprise, this tutorial is based on the Docker engine. Both version will work in this protocol. Linux users can install both Docker CE and Desktop, while macOS and Windows users should install Docker Desktop.

Follow the installation instructions from **https://docs.docker.com/engine/install/**

2 Get your data

You will need fastq data (long read, called ID.fastq), short reads, and the assembly data.

Run sspace_longread

3 Run the following command (replace /your_dir for the base directory where you have your data

```
docker run -it -v /your_dir:/ftmp dnalinux/sspace_longread:latest
perl SSPACE-LongRead.pl -c /ftmp/assembly.fasta -p /ftmp/ID.fastq
-b /ftmp/outputperlID
```

Run Gapcloser

4 Run the following command (replace /your_dir for the base directory where you have your data

```
docker run -it -v /your_dir:/ftmp dnalinux/lr_gapcloser:latest
bash /LR_Gapcloser/src/LR_Gapcloser.sh -i
/ftmp/outputperlID/scaffolds.fasta -l /ftmp/ID.fastq -o
/ftmp/ID_lr-gapcloser
```

Run BWA Index

5 Run the following command (replace /your_dir for the base directory where you have your data

```
docker run -it -v /your_dir:/ftmp dnalinux/bwa:0.7.17-3-deb bwa
index /ftmp/ID_lr-gapcloser/iteration-1/gapclosed.fasta
```

Run fastp

6 Run the following command (replace /your_dir for the base directory where you have your data)

```
docker run -it -v /your_dir:/ftmp dnalinux/fastp:0.23.4 fastp --
in1  /ftmp/shortreads/ID_R1.fastq.gz --in2
/ftmp/shortreads/ID_R2.fastq.gz --out1
/ftmp/shortreads/ID_R1_trim.fastq.gz --out2
/ftmp/shortreads/ID_R2_trim.fastq.gz
```

Run BWA mem

7 Run the following command (replace /your_dir for the base directory where you have your data). Replace CPU for your CPU count.

```
docker run -it -v /your_dir:/ftmp dnalinux/bwa:0.7.17-3-deb
/bin/bash -c "bwa mem -t CPU /ftmp/ID_lr-gapcloser/iteration-
1/gapclosed.fasta /ftmp/shortreads/ID_R1_trim.fastq.gz
/ftmp/shortreads/ID_R2_trim.fastq.gz > /ftmp/ID_aligned_reads.sam"
```

Run SAMTOOLS

8 SAMTOOLS View, Sort and Index

Run the following command (replace /your_dir for the base directory where you have your data).

```
docker run -it -v /your_dir:/ftmp dnalinux/samtools:1.20-3-deb
/bin/bash -c "samtools view -Sb /ftmp/ID_aligned_reads.sam >
/ftmp/ID_aligned_reads.bam"

docker run -it -v /your_dir:/ftmp dnalinux/samtools:1.20-3-deb
/bin/bash -c "samtools sort /ftmp/ID_aligned_reads.bam -o
/ftmp/ID_sorted_aligned_reads.bam"

docker run -it -v /your_dir:/ftmp dnalinux/samtools:1.20-3-deb
/bin/bash -c "samtools index /ftmp/ID_sorted_aligned_reads.bam"
```

Pilon

9 Run the following command (replace /your_dir for the base directory where you have your data).

```
docker run -it -v /your_dir:/ftmp dnalinux/pilon:1.24-3-deb pilon
--genome /ftmp/ID_lr-gapcloser/iteration-1/gapclosed.fasta --frags
/ftmp/ID_sorted_aligned_reads.bam --output /ftmp/ID_polished
```

Funannotate

10

Funannotate Clean and Sort

Run the following command (replace /your_dir for the base directory where you have your data).

```
docker run -it -v /your_dir:/ftmp dnalinux/funannotate:latest
/bin/bash -c "funannotate clean -i /ftmp/ID_polished.fasta -o
/ftmp/funannotate_prep/ID_polished_clean.fasta"

docker run -it -v /your_dir:/ftmp dnalinux/funannotate:latest
/bin/bash -c "funannotate sort -i
/ftmp/funannotate_prep/ID_polished_clean.fasta -o
/ftmp/funannotate_prep/ID_polished_clean_sort.fasta --minlen 1000"
```

RepeatMasker

11 Run the following command (replace /your_dir for the base directory where you have your data).

```
docker run -it -v /your_dir:/ftmp dnalinux/repeatmasker:latest
/usr/local/RepeatMasker/RepeatMasker -s -species Fungi
/ftmp/funannotate_prep/ID_polished_clean_sort.fasta -xsmall
```

Fuannotate Predict

12 Run the following command (replace /your_dir for the base directory where you have your data). Replace CPU for your CPU count.

```
docker run -it -v /your_dir:/ftmp dnalinux/funannotate-gmes-
dikarya:latest /bin/bash -c "funannotate predict -i
/ftmp/funannotate_prep/ID_polished_clean_sort.fasta.masked -s ID -
o /ftmp/funannotate --cpus CPU"
```

Interproscan

13 Run the following command (replace /your_dir for the base directory where you have your data). Replace CPU for your CPU count.

```
docker run -it -v /your_dir:/ftmp -v
/home/sebastian/ips/interproscan-5.69-
101.0/data:/opt/interproscan/data -v /tmp:/temp
dnalinux/interproscan:5.69-101.0 --input
/ftmp/funannotate/predict_results/ID.proteins.fa --disable-precalc
--output-dir /ftmp/funannotate/ipsout --cpu CPU
```

Funannoate annotate

14 Run the following command (replace /your_dir for the base directory where you have your data)

```
docker run -it -v /your_dir:/ftmp dnalinux/funannotate-gmes-
dikarya /bin/bash -c "funannotate annotate -i /ftmp/funannotate --
fasta /ftmp/funannotate/predict_results/ID.proteins.fa --species
ID --out /ftmp/FA_results --iprscan
/ftmp/funannotate/ipsout/ID.proteins.fa.xml"
```