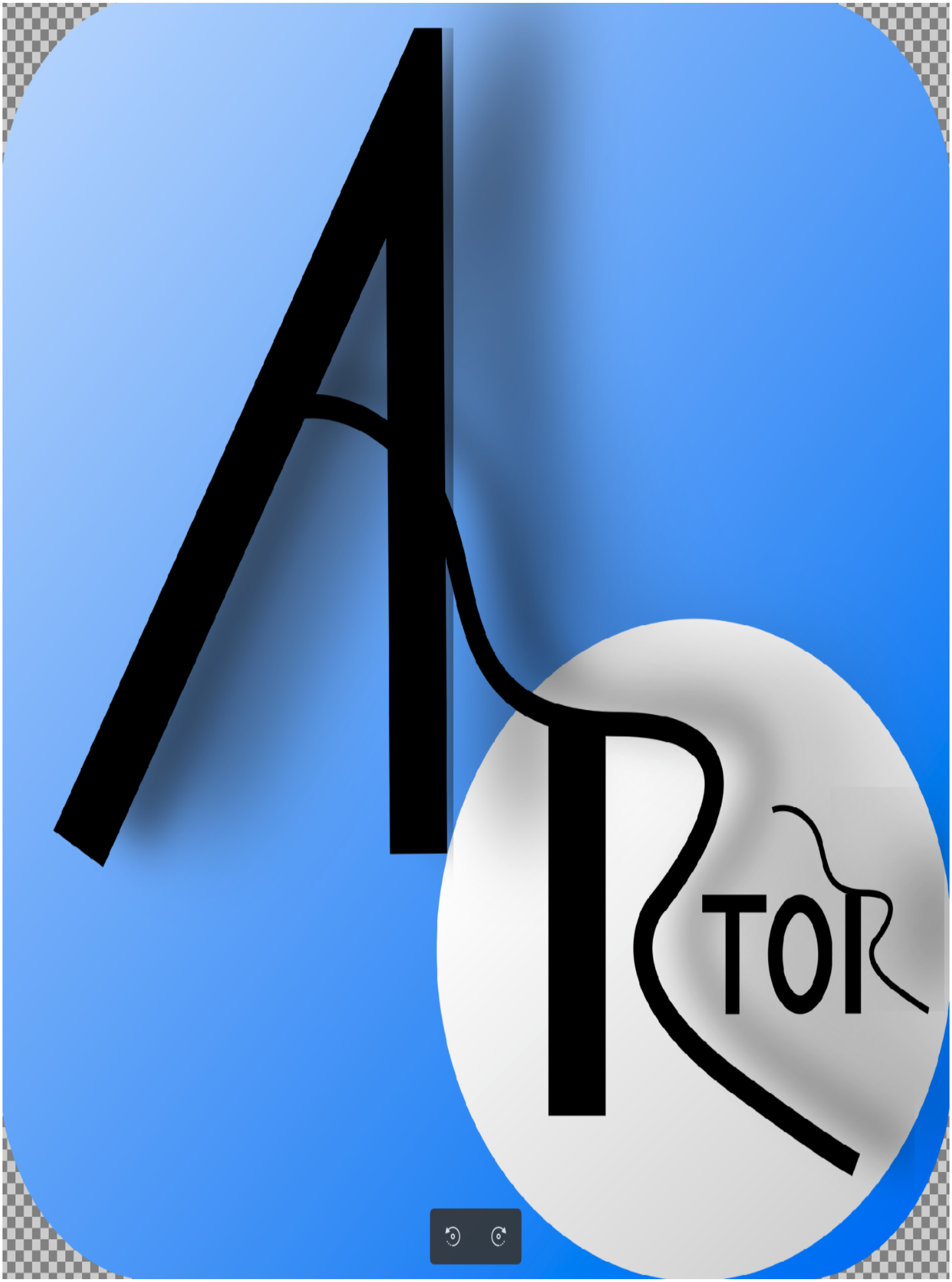

A
R
TOR

# Manual
# Annotator-RNAtor

# INDEX

## Table of Contents

# Introduction

The explosion of new sequencing technologies has brought hope to sequence genomes efficiently. Many genomes are sequenced in record time and the limits of genomes or genes research are being pushed further and further.

Comparing genes of a bacterial genus in order to highlight the evolution of homologous genes is a daunting task that required the development of Annotator-RNAtor; a bioinformatics tool that uses the power of your computer's processors to deliver efficient results to you. Annotator-RNAtor is a bioinformatics tool developed for the intra-genus study of bacterial homologous genes by comparing genes (Annotator) or gene expression (RNAtor). Annotator-RNAtor is written in python with the user-friendliness of the PyQt5 GUI. The efficiency of Annotator-RNAtor is based on running tasks in parallel thanks to multiple processors and GNU Parallel .

For the comparison and search of homologous genes, Annotator-RNAtor uses the BLAST+ suite of commands from NCBI (National Center for Biotechnology Information).

Annotator-RNAtor includes two separate methods for the dynamic study of homologous genes and the realization of graph and matrix. First method imports python's NetworkX package whereas the second method implies the GET_HOMOLOGUES tool for screening homologs and generating gene presence/absence matrix.

Annotator-RNAtor retrieves results based on user provided similarity thresholds and registers in the geneAssociation.tsv file all homologous genes with a percentage of similarity greater than or equal to the threshold of similarity validated or entered. Towards the end of the geneAssociation.tsv file, Annotator-RNAtor generates a presence (locus) or absence (-) map that allows you to finalize your analyses.

The RNAtor menu in Annotator-RNAtor implements the bwa and bwa-mem commands. Annotator-RNAtor is designed to use the annotation files of NCBI (cds_from_genomic.ffn or cds_from_genomic.fna, genomic.fna, protein.faa, genomic.gff, and feature_table.txt) and Prokka (genome.ffn, genome.fna, genome.faa and genome.gff). In the case of a Prokka annotation, the feature_table.txt or genome.tbl file should not be uploaded. Pandas and FeatureCounts of Subread are invoked by RNAtor. The Blastp and Blastn commands of Annotator are executed after invoking Bash. All Annotator-RNAtor commands can be executed stand alone or in a specific Conda environment.

## Required system and software dependencies

Annotator-RNAtor can be setup stand alone or used by setting up a conda environment. Currently, the software runs only on Linux or Ubuntu. It is not compatible with the recent ARM64 architecture supported in new MacOS.

For stand-alone installation of the pipeline, the following software need to be installed in your system

- Python3.6 or higher
- PyQt5
- Perl
- Pandas
- Parallel
- Numpy >=1.20.3
- PROKKA
- BLAST
- GET_HOMOLOGOUS
- NETWORKX
- BWA
- Subread FeatureCount

## Installation

**Manual installation using setup file**

Some of these dependencies can be installed using the setup.py file provided with the Annotator-RNAtor package on Github. Software like PROKKA and GET_HOMOLOGUES need to be installed manually by user as the required dependencies of these software may vary from system to system.

For installation via setup file follow the steps:

1. Download Annotator from github via *https://github.com/BactSymEvol/Annotator*.
2. cd Annotator
3. Type command *python3 setup.py*

The above command will install all the required dependencies except for PROKKA and GET_HOMOLOGUES.

**Installation via conda .yaml file**

In a Conda environment, it is desirable to create and work in specific environment that contain appropriate packages. This avoids uninstalling and reinstalling Conda when the system is corrupted, a waste of time.

For installation of dependencies via Conda, a meta.yaml file has been provided for setting up the environment for Annotator. Run the following command for installing dependencies via conda

*conda env create -f meta.yaml*

Once the packages are installed you can activate the environment via the following command

*conda activate Annotator*

When the specific environment is enabled, you have the name of your specific environment displayed on the left in the terminal. To disable use the following command

*conda deactivate*

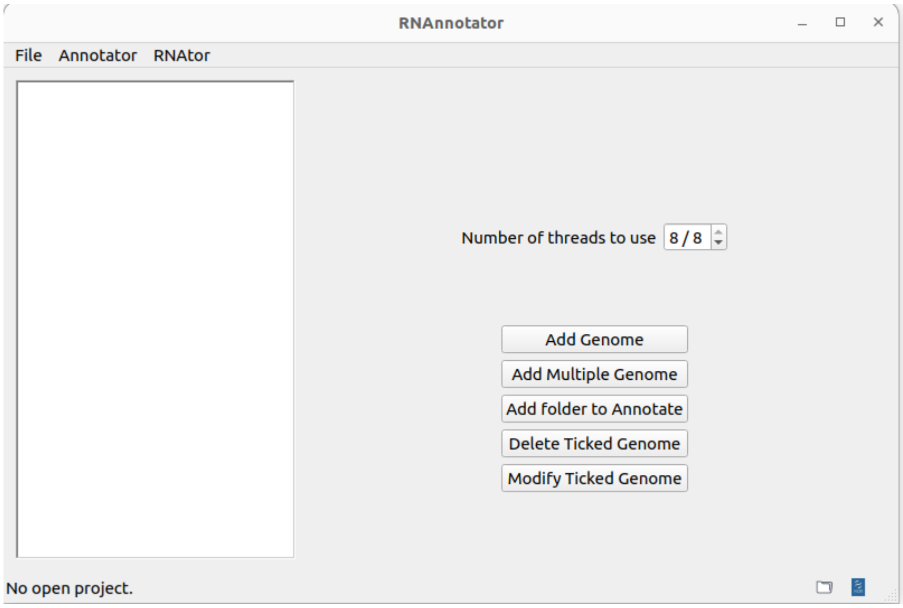
## Usage

From your terminal, go to the Annotator directory. Once in this directory, you should have the following files: **exec.py**, __init__.py, README.md, setup.py and folders gui and lib.

Another way to do this is to locate yourself on the Annotator folder, then right click and choose "Open in terminal".

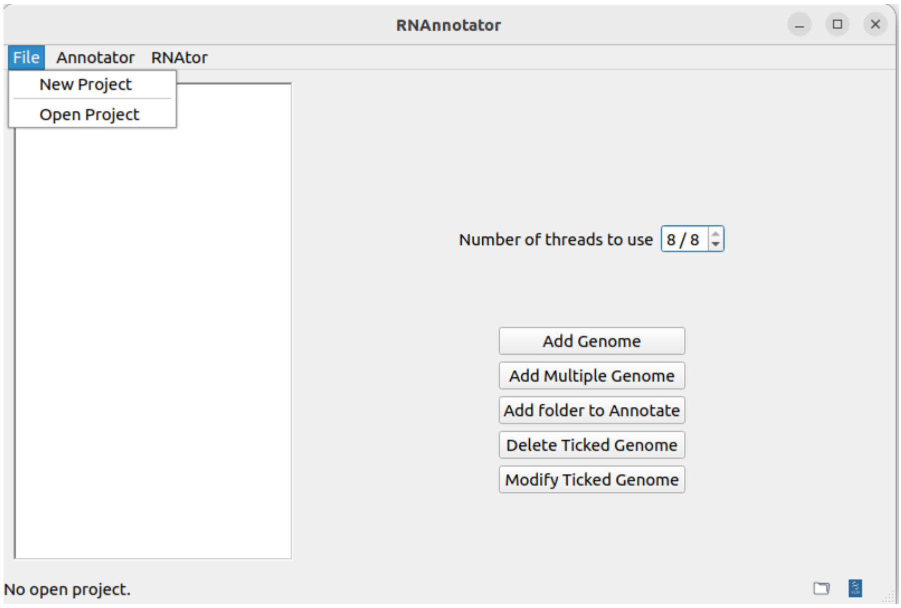
Follow the steps below:

1. Run *python3 exec.py*

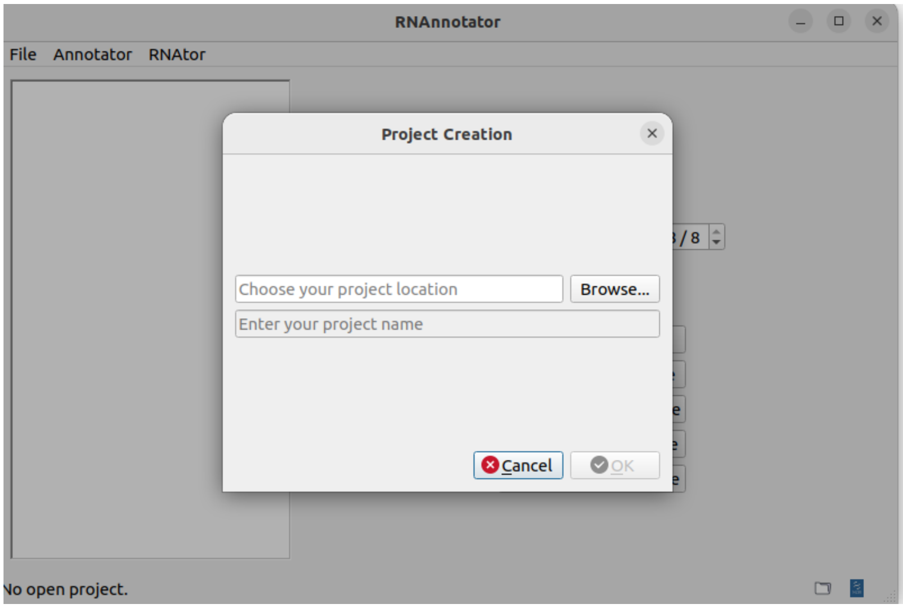


The above window will open as soon as you execute the command.

2. Go to File -> New Project, to create a new project



Once you click on New Project it will open another window as below



Choose location of your New Project and type name of the project. This will create a <projectName> folder at the location. The folder will contain sub folders as below

- Annotation
- BLASTresults
- BWA
- Cache
- Counts
- Database
- protein_files

These folders will initially be empty

3. Once the new project is created, prepare files.

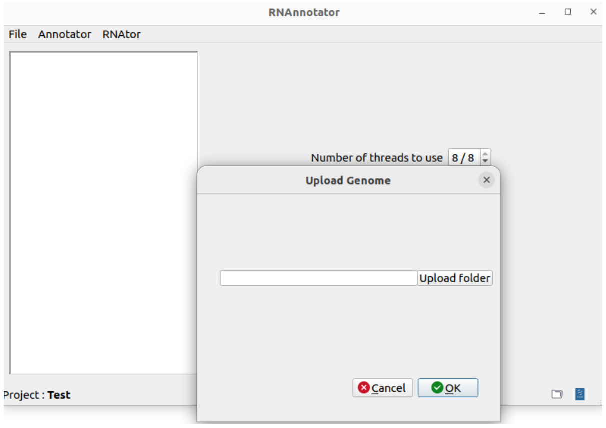
Prepare files:

a. Store all the genome .fna files in a single folder. Keep in mind the names of files should not be similar in any way.

b. Store all fastq files for the genomes in a separate folder with the same name as genome file names. For eg. If genome file name is Nelong.fna then fastq files must be named Nelong_R1.fastq and Nelong_R2.fastq. Make sure there is no difference in name. Do not keep underscore "_". If there are same species add a number or strain name like 1Nelong, 2Nelong etc…

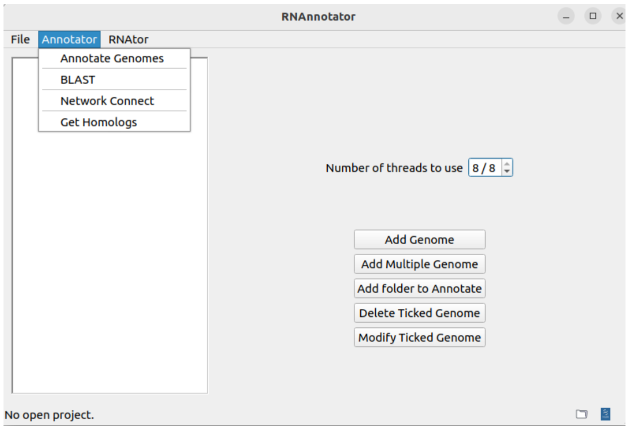
c. Click on "Add genome folder" to add genomes for annotation.

d. Click on “Annotator -> Annotate Genomes” to automatically annotate genomes using PROKKA. Once you click on Annotate Genomes it will ask for path of PROKKA installation. Paste the absolute path of PROKKA upto its executable in /bin folder in the terminal.



Number of threads to use 8 / 8

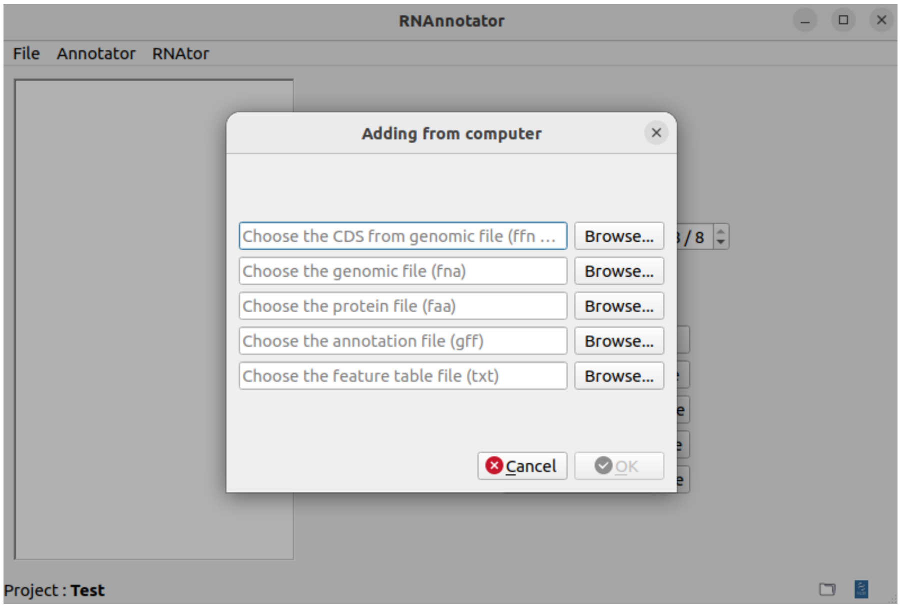
Project : Test



Number of threads to use 8 / 8

No open project.

4. Annotator requires four files for screening, namely genome.fna, genome.faa, genome.cds, and genome.gff. If you already have annotation files then you can add genomes one at a time or in bulk via <Add Genome> and <Add Multiple Genome>.

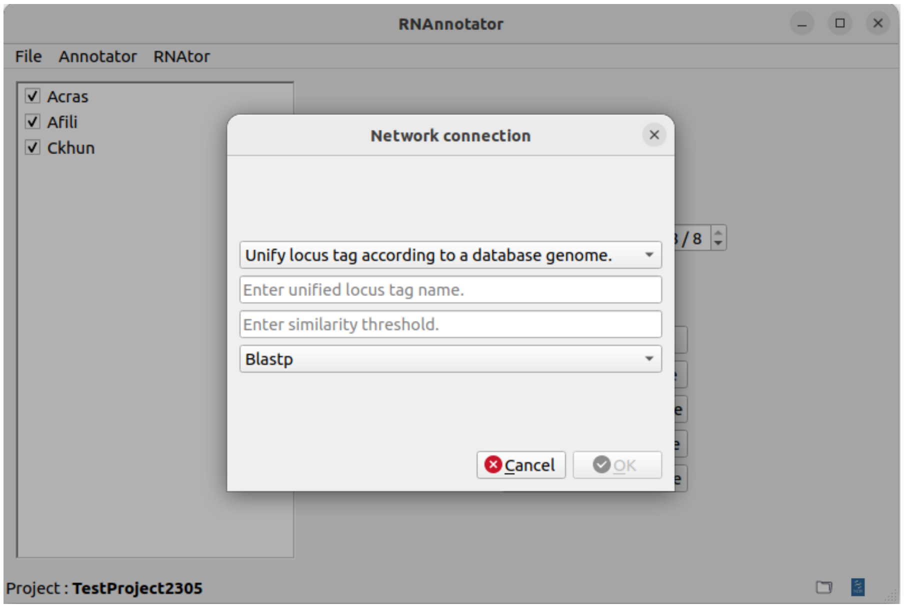

RNAnnotator
File
Annotator
RNAtor
Adding from computer
Choose the CDS from genomic file (ffn ...
Browse...
Choose the genomic file (fna)
Browse...
Choose the protein file (faa)
Browse...
Choose the annotation file (gff)
Browse...
Choose the feature table file (txt)
Browse...
Cancel
OK

Project : Test

5. Once the genome files are uploaded or PROKKA is run for annotation, the files will be displayed on the left side panel.

6. Select the genomes you wish to screen.

7. Click on Annotator > BLAST, for generating BLAST results
8. For screening via NetworkX method click on Annotator > NetworkX. Provide a unified tag name for reannotation of the genes in the genomes. Enter similarity threshold. Recommended threshold is 60% for NetworkX. This will generate a geneAssociation.tsv file in the project folder



☑ Acras
☑ Afili
☑ Ckhun

Unify locus tag according to a database genome.

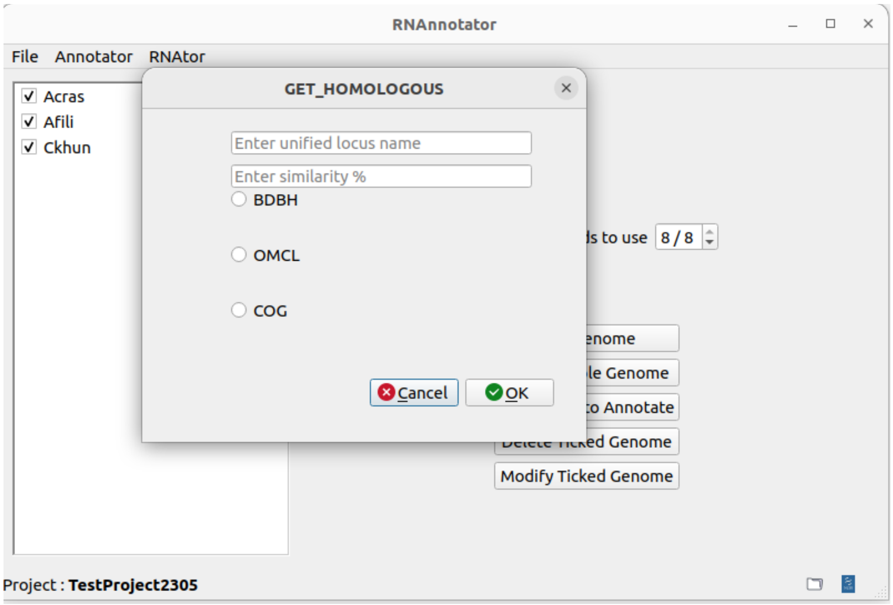
Enter unified locus tag name.

Enter similarity threshold.

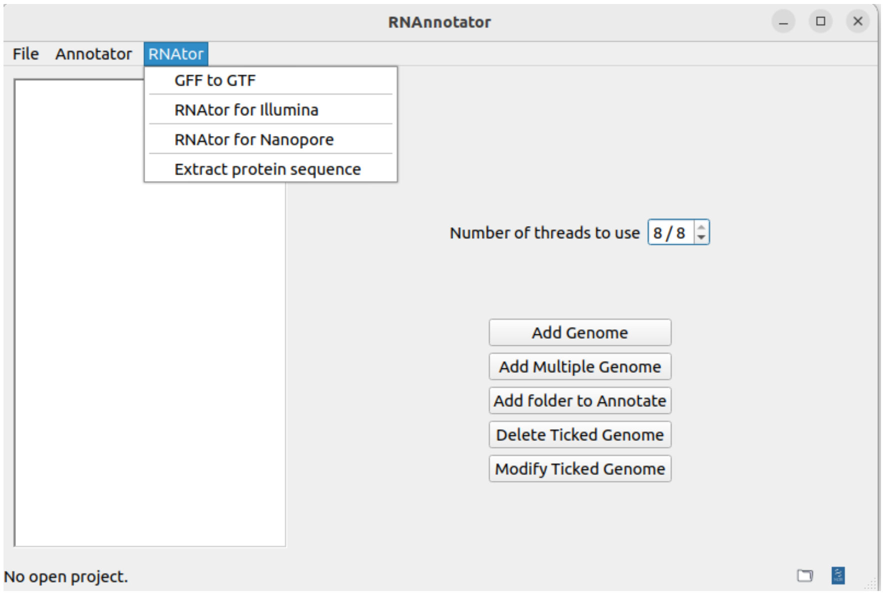
Blastp

Project : TestProject2305

9. For screening via GET_HOMOLOGUES method, click on Annotator > Get Homologs. After clicking, it will ask for the path of get_homologues.pl file. Paste the absolute path of file. For example if the file is in /home/user/Documents/get_homologoues/get_homoloues.pl, paste the complete path in terminal. GET_HOMOLOGUES uses three different methods for screening homologs which are the Bi-directional Best Hit method, OrthologousMCL method and Cluster of orthologs method. Choose method based on user preference.
10. Enter unified locus name for reannotating genes and an identity threshold.



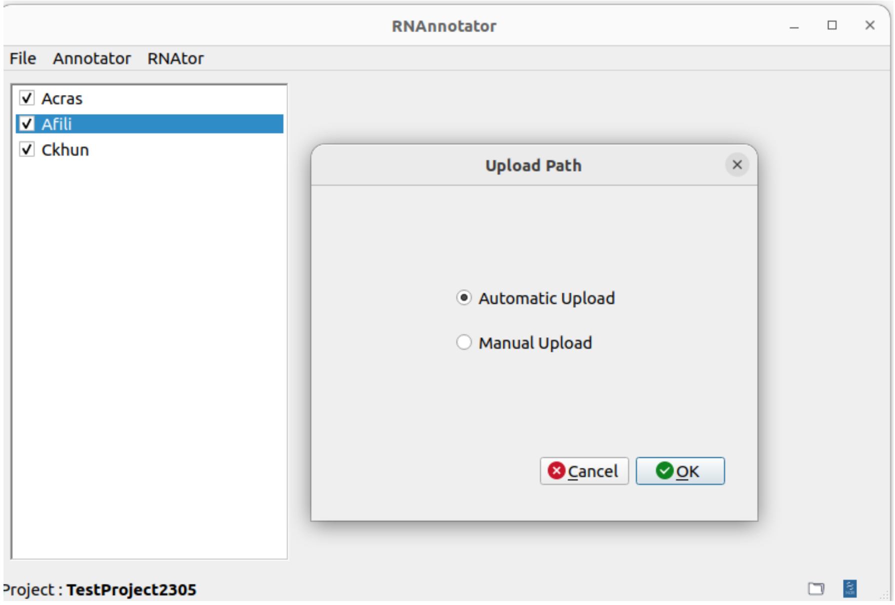
11. In the end, both methods generate geneAssociation.tsv files which show presence/absence of genes in the genomes.
12. Please note that the results of one method i.e. NetworkX will over-write results of GET_HOMOLOGUES. In order to get results from both methods it is advised to run Annotator-RNAtor for one method first and then repeat the process for second method after saving and renaming results.
13. Once the geneAssociation.tsv is generated click in RNAtor > GFF to GTF, for reannotating the gff files with the unified locus name provided. This will generate converted gtf files in the Database folder.



Number of threads to use 8 / 8

No open project.

14. Click on RNAtor for Illumina to get Counts of genes from Fastq files.



☑ Acras
☑ Afili
☑ Ckhun

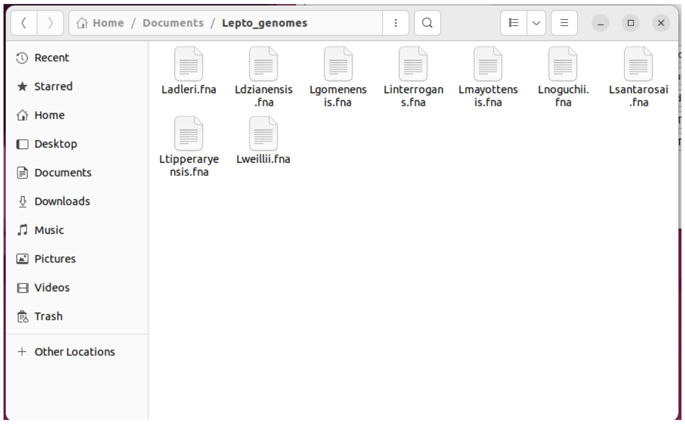
● Automatic Upload
○ Manual Upload

Project : **TestProject2305**

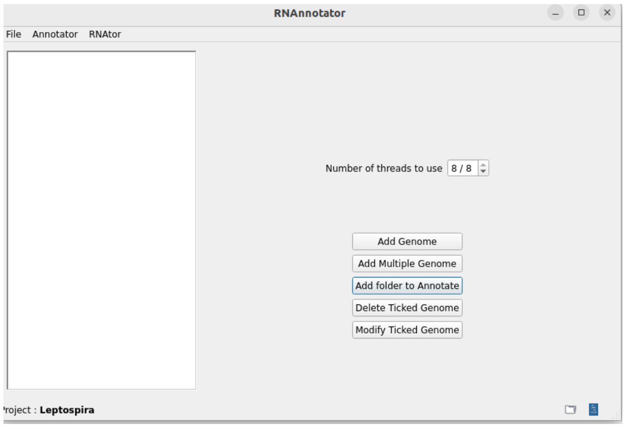
15. Click on Automatic Upload to select the Project Database folder and the fastq folder which has your fastq files as prepared in steps above.
16. Once the folders are uploaded, it will map the genomes on fastq files using bwa and generate a CountComparisonOutput.csv. This file will contain the counts of homologous genes in all the genomes.

# Usage using an example

Let us take an example of 9 genomes of *Leptospira* genus which have been used in the publication. In order to run the pipeline, we first prepared the genome files and fastq files in separate folders.



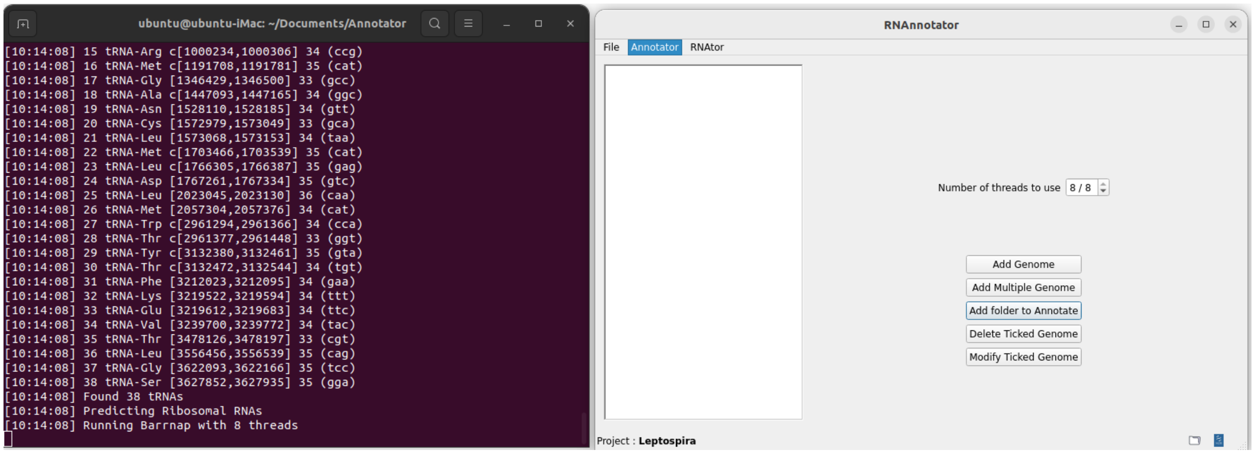
1. Go to terminal. Type "conda activate Annotator" for using conda environment.
2. cd /path/to/Annotator and run "*python3 exec.py*"
3. Create New Project from File -> New Project.
4. Add folder to Annotate.



Number of threads to use 8 / 8

roject : **Leptospira**

5. Click on Annotator -> Annotate Genomes

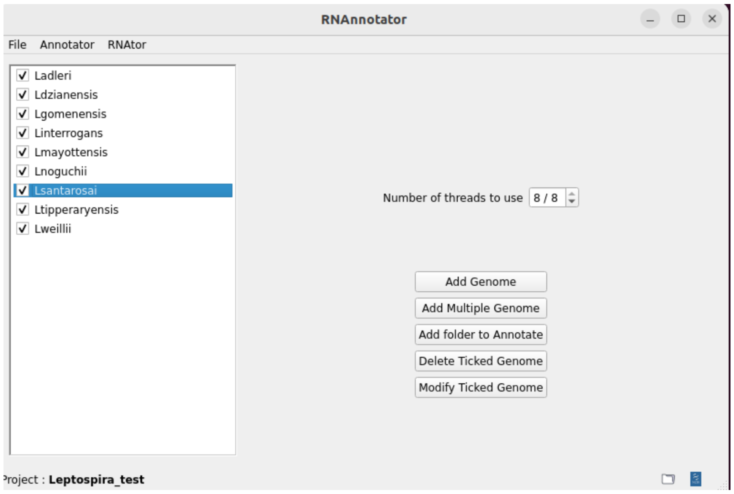


```
[10:14:08] 15 tRNA-Arg c[1000234,1000306] 34 (ccg)
[10:14:08] 16 tRNA-Met c[1191708,1191781] 35 (cat)
[10:14:08] 17 tRNA-Gly [1346429,1346500] 33 (gcc)
[10:14:08] 18 tRNA-Ala c[1447093,1447165] 34 (ggc)
[10:14:08] 19 tRNA-Asn [1528110,1528185] 34 (gtt)
[10:14:08] 20 tRNA-Cys [1572979,1573049] 33 (gca)
[10:14:08] 21 tRNA-Leu [1573068,1573153] 34 (taa)
[10:14:08] 22 tRNA-Met c[1703466,1703539] 35 (cat)
[10:14:08] 23 tRNA-Leu c[1766305,1766387] 35 (gag)
[10:14:08] 24 tRNA-Asp [1767261,1767334] 35 (gtc)
[10:14:08] 25 tRNA-Leu [2023045,2023130] 36 (caa)
[10:14:08] 26 tRNA-Met [2057304,2057376] 34 (cat)
[10:14:08] 27 tRNA-Trp c[2961294,2961366] 34 (cca)
[10:14:08] 28 tRNA-Thr c[2961377,2961448] 33 (ggt)
[10:14:08] 29 tRNA-Tyr c[3132380,3132461] 35 (gta)
[10:14:08] 30 tRNA-Thr c[3132472,3132544] 34 (tgt)
[10:14:08] 31 tRNA-Phe [3212023,3212095] 34 (gaa)
[10:14:08] 32 tRNA-Lys [3219522,3219594] 34 (ttt)
[10:14:08] 33 tRNA-Glu [3219612,3219683] 34 (ttc)
[10:14:08] 34 tRNA-Val [3239700,3239772] 34 (tac)
[10:14:08] 35 tRNA-Thr [3478126,3478197] 33 (cgt)
[10:14:08] 36 tRNA-Leu [3556456,3556539] 35 (cag)
[10:14:08] 37 tRNA-Gly [3622093,3622166] 35 (tcc)
[10:14:08] 38 tRNA-Ser [3627852,3627935] 35 (gga)
[10:14:08] Found 38 tRNAs
[10:14:08] Predicting Ribosomal RNAs
[10:14:08] Running Barrnap with 8 threads
```

Number of threads to use 8 / 8

Project : **Leptospira**

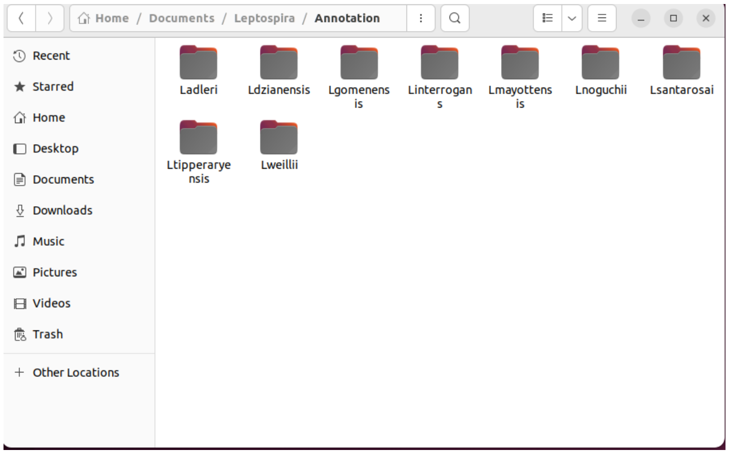
6. The genomes will be annotated using prokka. Wait until the genomes are annotated. As soon as the runs are finished, they will be seen in the left panel window



☑ Ladleri
☑ Ldzianensis
☑ Lgomenensis
☑ Linterrogans
☑ Lmayottensis
☑ Lnoguchii
☑ Lsantarosai
☑ Ltipperaryensis
☑ Lweillii

Number of threads to use 8 / 8

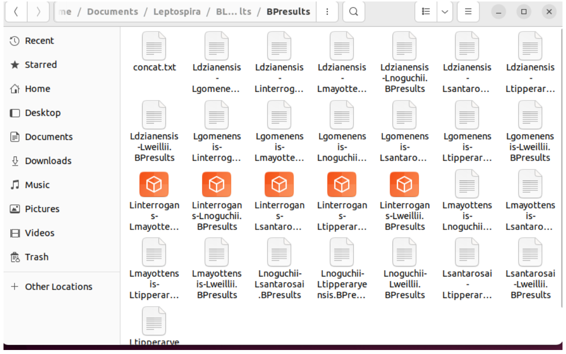
Project : **Leptospira_test**



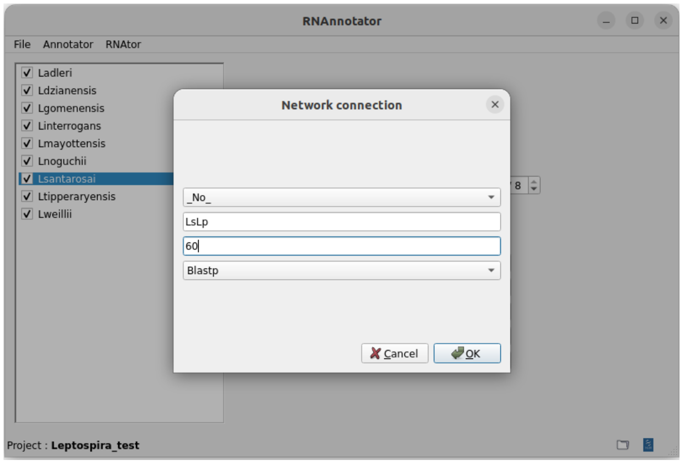
Ladleri
Ldzianensis
Lgomenensis
Linterrogans
Lmayottensis
Lnoguchii
Lsantarosai
Ltipperaryensis
Lweillii

7. You will be able to see all annotated files in the Project/Annotation folder.
8. Verify that all the genomes are annotated correctly and added to Project/Database folder

9. Click on Annotator -> Blast. It will run Blastp on all genomes and generate results as shown in the figure below


ne / Documents / Leptospira / BL… lts / BPresults
Recent
Starred
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locations
concat.txt
Ldzianensis - Lgomene…
Ldzianensis - Linterrog…
Ldzianensis - Lmayotte…
Ldzianensis -Lnoguchii. BPresults
Ldzianensis - Lsantaro…
Ldzianensis - Ltipperar…
Ldzianensis -Lweillii. BPresults
Lgomenens is- Linterrog…
Lgomenens is- Lmayotte…
Lgomenens is- Lnoguchii…
Lgomenens is- Lsantaro…
Lgomenens is- Ltipperar…
Lgomenens is-Lweillii. BPresults
Linterrogan s- Lmayotte…
Linterrogan s-Lnoguchii. BPresults
Linterrogan s- Lsantaro…
Linterrogan s- Ltipperar…
Linterrogan s-Lweillii. BPresults
Lmayottens is- Lnoguchii…
Lmayottens is- Lsantaro…
Lmayottens is- Ltipperar…
Lmayottens is-Lweillii. BPresults
Lnoguchii- Lsantarosai .BPresults
Lnoguchii- Ltipperarye nsis.BPre…
Lnoguchii- Lweillii. BPresults
Lsantarosai - Ltipperar…
Lsantarosai -Lweillii. BPresults
Ltipperarye

10. Click on Annotator > Network connection method to generate a GeneAssociation file of presence/absence. Enter details as per need.



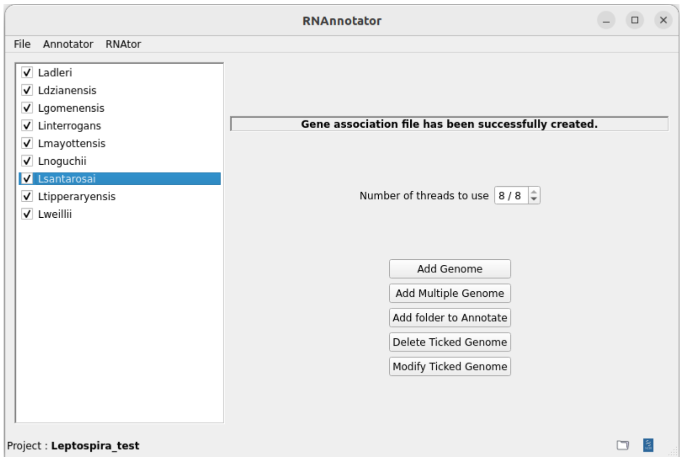
_No_

LsLp

60

Blastp

Project : **Leptospira_test**

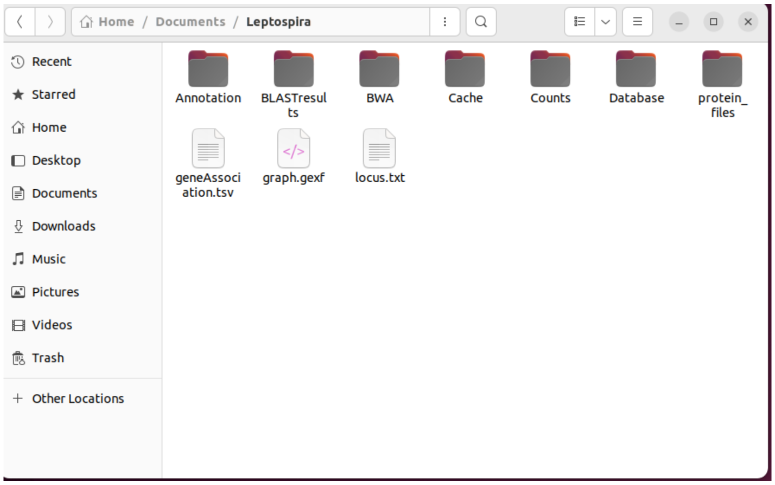


**Gene association file has been successfully created.**

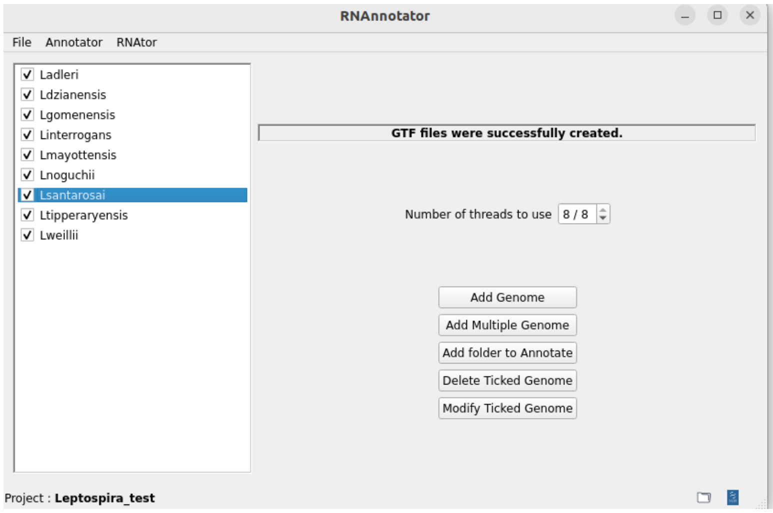
Number of threads to use 8 / 8

Project : **Leptospira_test**

11. Once the run is finished you will be able to see a geneAssociation.tsv file in the Project folder as below



12. Click on RNAtor -> GFF to GTF to reannotate gene names as per user provided locus



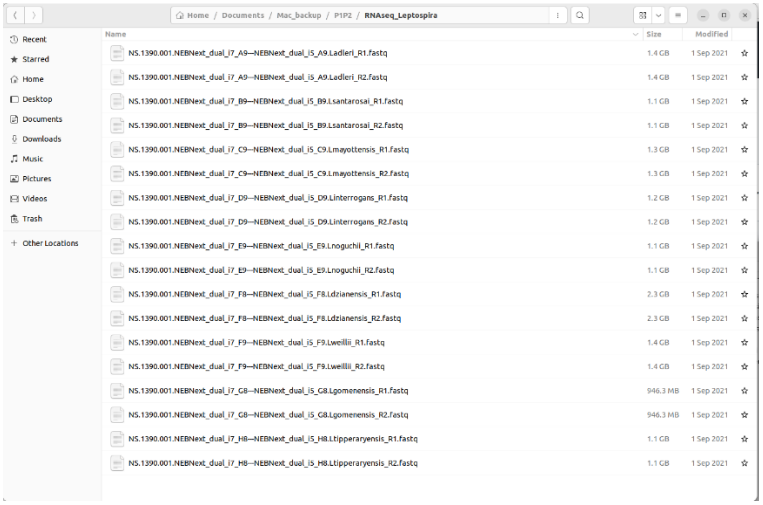
☑ Ladleri
☑ Ldzianensis
☑ Lgomenensis
☑ Linterrogans
☑ Lmayottensis
☑ Lnoguchii
☑ Lsantarosai
☑ Ltipperaryensis
☑ Lweillii

GTF files were successfully created.

Number of threads to use 8 / 8

Project : **Leptospira_test**

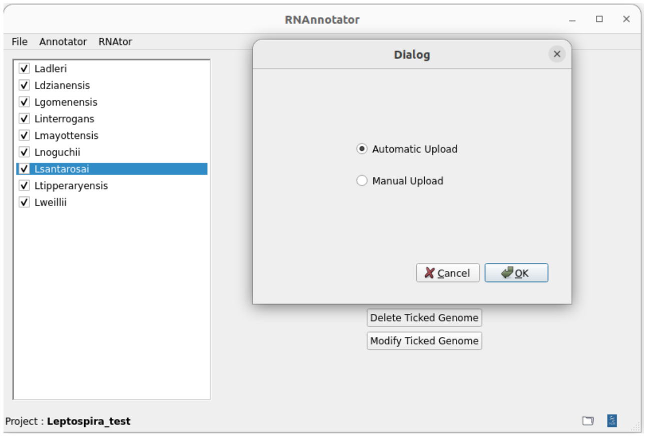
13. Click on RNAtor for Illumina to run RNAtor. Before this make sure you have your Fastq folder prepared. For this example it can be seen as below



| Name | Size | Modified |
|---|---|---|
| NS.1390.001.NEBNext_dual_i7_A9--NEBNext_dual_i5_A9.Ladleri_R1.fastq | 1.4 GB | 1 Sep 2021 |
| NS.1390.001.NEBNext_dual_i7_A9--NEBNext_dual_i5_A9.Ladleri_R2.fastq | 1.4 GB | 1 Sep 2021 |
| NS.1390.001.NEBNext_dual_i7_B9--NEBNext_dual_i5_B9.Lsantarosai_R1.fastq | 1.1 GB | 1 Sep 2021 |
| NS.1390.001.NEBNext_dual_i7_B9--NEBNext_dual_i5_B9.Lsantarosai_R2.fastq | 1.1 GB | 1 Sep 2021 |
| NS.1390.001.NEBNext_dual_i7_C9--NEBNext_dual_i5_C9.Lmayottensis_R1.fastq | 1.3 GB | 1 Sep 2021 |
| NS.1390.001.NEBNext_dual_i7_C9--NEBNext_dual_i5_C9.Lmayottensis_R2.fastq | 1.3 GB | 1 Sep 2021 |
| NS.1390.001.NEBNext_dual_i7_D9--NEBNext_dual_i5_D9.Linterrogans_R1.fastq | 1.2 GB | 1 Sep 2021 |
| NS.1390.001.NEBNext_dual_i7_D9--NEBNext_dual_i5_D9.Linterrogans_R2.fastq | 1.2 GB | 1 Sep 2021 |
| NS.1390.001.NEBNext_dual_i7_E9--NEBNext_dual_i5_E9.Lnoguchii_R1.fastq | 1.1 GB | 1 Sep 2021 |
| NS.1390.001.NEBNext_dual_i7_E9--NEBNext_dual_i5_E9.Lnoguchii_R2.fastq | 1.1 GB | 1 Sep 2021 |
| NS.1390.001.NEBNext_dual_i7_F8--NEBNext_dual_i5_F8.Ldzianensis_R1.fastq | 2.3 GB | 1 Sep 2021 |
| NS.1390.001.NEBNext_dual_i7_F8--NEBNext_dual_i5_F8.Ldzianensis_R2.fastq | 2.3 GB | 1 Sep 2021 |
| NS.1390.001.NEBNext_dual_i7_F9--NEBNext_dual_i5_F9.Lweillii_R1.fastq | 1.4 GB | 1 Sep 2021 |
| NS.1390.001.NEBNext_dual_i7_F9--NEBNext_dual_i5_F9.Lweillii_R2.fastq | 1.4 GB | 1 Sep 2021 |
| NS.1390.001.NEBNext_dual_i7_G8--NEBNext_dual_i5_G8.Lgomenensis_R1.fastq | 946.3 MB | 1 Sep 2021 |
| NS.1390.001.NEBNext_dual_i7_G8--NEBNext_dual_i5_G8.Lgomenensis_R2.fastq | 946.3 MB | 1 Sep 2021 |
| NS.1390.001.NEBNext_dual_i7_H8--NEBNext_dual_i5_H8.Ltipperaryensis_R1.fastq | 1.1 GB | 1 Sep 2021 |
| NS.1390.001.NEBNext_dual_i7_H8--NEBNext_dual_i5_H8.Ltipperaryensis_R2.fastq | 1.1 GB | 1 Sep 2021 |

14. Make sure the fastq file names contain genome names exactly as genome1_R1.fastq and genome1_R2.fastq. If they are not named correctly the pipeline will not run properly and show an error in terminal

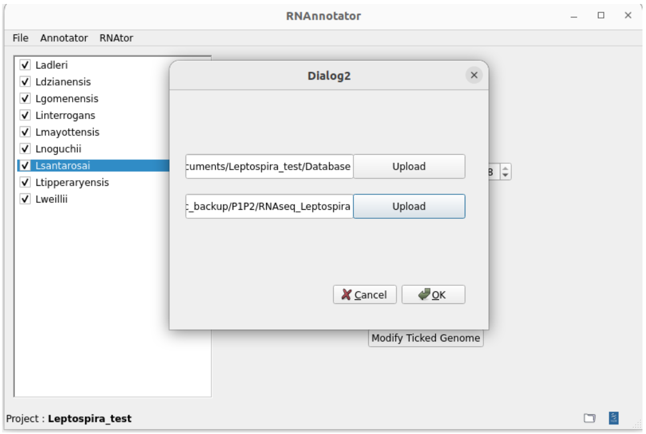
15. Select "Automatic Upload" and click OK. Select the Project/Database folder and Fastq folder and click OK.



● Automatic Upload

○ Manual Upload

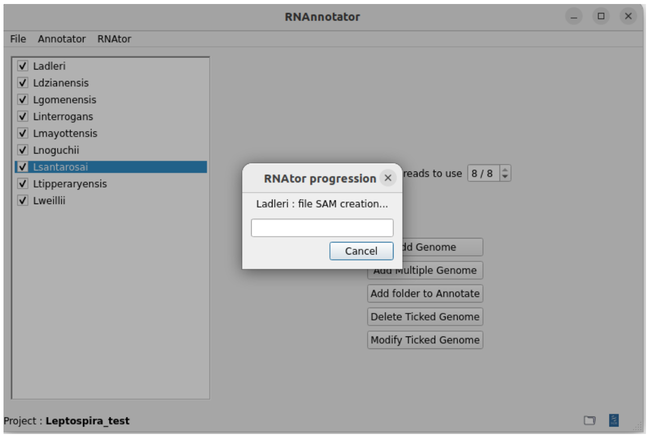
Project : **Leptospira_test**



:uments/Leptospira_test/Database

:_backup/P1P2/RNAseq_Leptospira

Project : **Leptospira_test**



Ladleri : file SAM creation...

reads to use 8 / 8

Project : **Leptospira_test**

16. Once it is completed, you will see a CountsComparisonOutput.csv file in the project folder.

## Output Files

For a sample project of 9 genomes the following output files will be generated in a project folder

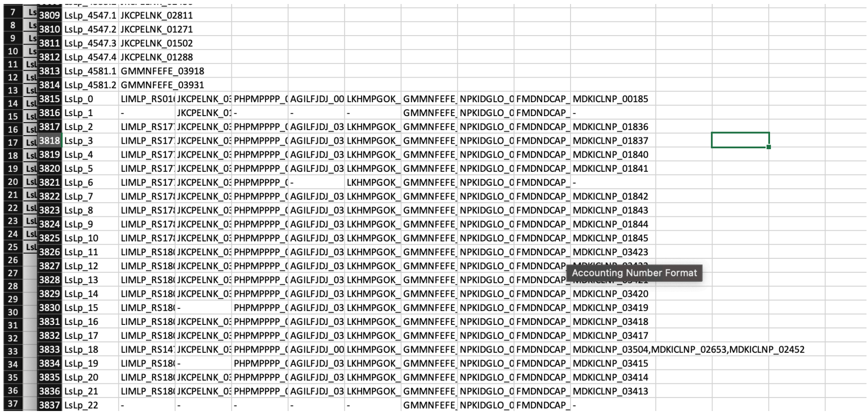
```
.
├── BLASTresults
│   ├── BPresults
│   │   ├── concat.txt
│   │   ├── Ladleri-Ldzianensis.BPresults
│   │   ├── Ladleri-Lgomenensis.BPresults
│   │   ├── Ladleri-Linterrogans.BPresults
│   │   ├── Ladleri-Lmayottensis.BPresults
│   │   ├── Ladleri-Lnoguchii.BPresults
│   │   ├── Ladleri-Lsantarosai.BPresults
│   │   ├── Ladleri-Ltipperaryensis.BPresults
│   │   ├── Ladleri-Lweillii.BPresults
│   │   ├── Ldzianensis-Lgomenensis.BPresults
│   │   ├── Ldzianensis-Linterrogans.BPresults
│   │   ├── Ldzianensis-Lmayottensis.BPresults
│   │   ├── Ldzianensis-Lnoguchii.BPresults
│   │   ├── Ldzianensis-Lsantarosai.BPresults
│   │   ├── Ldzianensis-Ltipperaryensis.BPresults
│   │   ├── Ldzianensis-Lweillii.BPresults
│   │   ├── Lgomenensis-Linterrogans.BPresults
│   │   ├── Lgomenensis-Lmayottensis.BPresults
│   │   ├── Lgomenensis-Lnoguchii.BPresults
│   │   ├── Lgomenensis-Lsantarosai.BPresults
│   │   ├── Lgomenensis-Ltipperaryensis.BPresults
│   │   ├── Lgomenensis-Lweillii.BPresults
│   │   ├── Linterrogans-Lmayottensis.BPresults
│   │   ├── Linterrogans-Lnoguchii.BPresults
│   │   ├── Linterrogans-Lsantarosai.BPresults
│   │   ├── Linterrogans-Ltipperaryensis.BPresults
│   │   ├── Linterrogans-Lweillii.BPresults
│   │   ├── Lmayottensis-Lnoguchii.BPresults
│   │   ├── Lmayottensis-Lsantarosai.BPresults
│   │   ├── Lmayottensis-Ltipperaryensis.BPresults
│   │   ├── Lmayottensis-Lweillii.BPresults
│   │   ├── Lnoguchii-Lsantarosai.BPresults
│   │   ├── Lnoguchii-Ltipperaryensis.BPresults
│   │   ├── Lnoguchii-Lweillii.BPresults
│   │   ├── Lsantarosai-Ltipperaryensis.BPresults
│   │   ├── Lsantarosai-Lweillii.BPresults
│   │   ├── Ltipperaryensis-Lweillii.BPresults
│   │   └── niceConcat.txt
│   └── DBp
│       ├── Ladleri.phr
│       ├── Ladleri.pin
│       ├── Ladleri.psq
│       ├── Ldzianensis.phr
│       ├── Ldzianensis.pin
│       ├── Ldzianensis.psq
│       ├── Lgomenensis.phr
│       ├── Lgomenensis.pin
│       ├── Lgomenensis.psq
│       ├── Linterrogans.phr
│       ├── Linterrogans.pin
│       ├── Linterrogans.psq
│       ├── Lmayottensis.phr
│       ├── Lmayottensis.pin
│       ├── Lmayottensis.psq
│       ├── Lnoguchii.phr
│       ├── Lnoguchii.pin
│       ├── Lnoguchii.psq
│       ├── LOG.txt
│       ├── Lsantarosai.phr
│       ├── Lsantarosai.pin
│       ├── Lsantarosai.psq
│       ├── Ltipperaryensis.phr
│       ├── Ltipperaryensis.pin
│       ├── Ltipperaryensis.psq
│       ├── Lweillii.phr
│       ├── Lweillii.pin
│       └── Lweillii.psq
├── BWA
│   ├── Ladleri.sam
│   ├── Ldzianensis.sam
│   ├── Lgomenensis.sam
│   ├── Linterrogans.sam
│   ├── Lmayottensis.sam
│   ├── Lnoguchii.sam
│   ├── Lsantarosai.sam
│   ├── Ltipperaryensis.sam
│   └── Lweillii.sam
├── Cache
│   └── project.cache
├
```

```
├── CountComparisonOutput.csv
├
├── Counts
│   ├── counts_Ladleri.txt
│   ├── counts_Ladleri.txt.summary
│   ├── counts_Ldzianensis.txt
│   ├── counts_Ldzianensis.txt.summary
│   ├── counts_Lgomenensis.txt
│   ├── counts_Lgomenensis.txt.summary
│   ├── counts_Linterrogans.txt
│   ├── counts_Linterrogans.txt.summary
│   ├── counts_Lmayottensis.txt
│   ├── counts_Lmayottensis.txt.summary
│   ├── counts_Lnoguchii.txt
│   ├── counts_Lnoguchii.txt.summary
│   ├── counts_Lsantarosai.txt
│   ├── counts_Lsantarosai.txt.summary
│   ├── counts_Ltipperaryensis.txt
│   ├── counts_Ltipperaryensis.txt.summary
│   ├── counts_Lweillii.txt
│   ├── counts_Lweillii.txt.summary
│   ├── featoutput_Ladleri.txt
│   ├── featoutput_Ldzianensis.txt
│   ├── featoutput_Lgomenensis.txt
│   ├── featoutput_Linterrogans.txt
│   ├── featoutput_Lmayottensis.txt
│   ├── featoutput_Lnoguchii.txt
│   ├── featoutput_Lsantarosai.txt
│   ├── featoutput_Ltipperaryensis.txt
│   └── featoutput_Lweillii.txt
├── Database
│   ├── Ladleri
│   │   ├── FAA
│   │   │   ├── Ladleri_converted.faa
│   │   │   └── Ladleri.faa
│   │   ├── FeatureTable
│   │   │   └── Ladleri.txt
│   │   ├── FFN
│   │   │   ├── Ladleri_converted.ffn
│   │   │   └── Ladleri.ffn
│   │   ├── FNA
│   │   │   ├── Ladleri.fna
│   │   │   ├── Ladleri.fna.amb
│   │   │   ├── Ladleri.fna.ann
│   │   │   ├── Ladleri.fna.bwt
│   │   │   ├── Ladleri.fna.pac
│   │   │   │
│   │   │   └── Ladleri.fna.sa
│   │   ├── GFF
│   │   │   ├── Ladleri_ANO.gff
│   │   │   └── Ladleri.gff
│   │   └── GTF
│   │       └── Ladleri_ANO.gtf
│   ├── Ldzianensis
│   │   ├── FAA
│   │   │   ├── Ldzianensis_converted.faa
│   │   │   └── Ldzianensis.faa
│   │   ├── FeatureTable
│   │   │   └── Ldzianensis.txt
│   │   ├── FFN
│   │   │   ├── Ldzianensis_converted.ffn
│   │   │   └── Ldzianensis.ffn
│   │   ├── FNA
│   │   │   ├── Ldzianensis.fna
│   │   │   ├── Ldzianensis.fna.amb
│   │   │   ├── Ldzianensis.fna.ann
│   │   │   ├── Ldzianensis.fna.bwt
│   │   │   ├── Ldzianensis.fna.pac
│   │   │   └── Ldzianensis.fna.sa
│   │   ├── GFF
│   │   │   ├── Ldzianensis_ANO.gff
│   │   │   └── Ldzianensis.gff
│   │   └── GTF
│   │       └── Ldzianensis_ANO.gtf
│   ├── Lgomenensis
│   │   ├── FAA
│   │   │   ├── Lgomenensis_converted.faa
│   │   │   └── Lgomenensis.faa
│   │   ├── FeatureTable
│   │   │   └── Lgomenensis.txt
│   │   ├── FFN
│   │   │   ├── Lgomenensis_converted.ffn
│   │   │   └── Lgomenensis.ffn
│   │   ├── FNA
│   │   │   ├── Lgomenensis.fna
│   │   │   ├── Lgomenensis.fna.amb
│   │   │   ├── Lgomenensis.fna.ann
│   │   │   ├── Lgomenensis.fna.bwt
│   │   │   ├── Lgomenensis.fna.pac
│   │   │   └── Lgomenensis.fna.sa
│   │   ├── GFF
│   │   │   ├── Lgomenensis_ANO.gff
│   │   │   └── Lgomenensis.gff
│   │   │
│   │   └── GTF
│   │       └── Lgomenensis_ANO.gtf
│   ├── Linterrogans
│   │   ├── FAA
│   │   │   ├── Linterrogans_converted.faa
│   │   │   └── Linterrogans.faa
│   │   ├── FeatureTable
│   │   │   └── Linterrogans.txt
│   │   ├── FFN
│   │   │   ├── Linterrogans_converted.ffn
│   │   │   └── Linterrogans.ffn
│   │   ├── FNA
│   │   │   ├── Linterrogans.fna
│   │   │   ├── Linterrogans.fna.amb
│   │   │   ├── Linterrogans.fna.ann
│   │   │   ├── Linterrogans.fna.bwt
│   │   │   ├── Linterrogans.fna.fai
│   │   │   ├── Linterrogans.fna.pac
│   │   │   └── Linterrogans.fna.sa
│   │   ├── GFF
│   │   │   ├── Linterrogans_ANO.gff
│   │   │   └── Linterrogans.gff
│   │   └── GTF
│   │       └── Linterrogans_ANO.gtf
│   ├── Lmayottensis
│   │   ├── FAA
│   │   │   ├── Lmayottensis_converted.faa
│   │   │   └── Lmayottensis.faa
│   │   ├── FeatureTable
│   │   │   └── Lmayottensis.txt
│   │   ├── FFN
│   │   │   ├── Lmayottensis_converted.ffn
│   │   │   └── Lmayottensis.ffn
│   │   ├── FNA
│   │   │   ├── Lmayottensis.fna
│   │   │   ├── Lmayottensis.fna.amb
│   │   │   ├── Lmayottensis.fna.ann
│   │   │   ├── Lmayottensis.fna.bwt
│   │   │   ├── Lmayottensis.fna.pac
│   │   │   └── Lmayottensis.fna.sa
│   │   ├── GFF
│   │   │   ├── Lmayottensis_ANO.gff
│   │   │   └── Lmayottensis.gff
│   │   └── GTF
│   │       └── Lmayottensis_ANO.gtf
│   ├── Lnoguchii
│   │   ├── FAA
│   │   │   ├── Lnoguchii_converted.faa
│   │   │   └── Lnoguchii.faa
│   │   ├── FeatureTable
│   │   │   └── Lnoguchii.txt
│   │   ├── FFN
│   │   │   ├── Lnoguchii_converted.ffn
│   │   │   └── Lnoguchii.ffn
│   │   ├── FNA
│   │   │   ├── Lnoguchii.fna
│   │   │   ├── Lnoguchii.fna.amb
│   │   │   ├── Lnoguchii.fna.ann
│   │   │   ├── Lnoguchii.fna.bwt
│   │   │   ├── Lnoguchii.fna.pac
│   │   │   └── Lnoguchii.fna.sa
│   │   ├── GFF
│   │   │   ├── Lnoguchii_ANO.gff
│   │   │   └── Lnoguchii.gff
│   │   └── GTF
│   │       └── Lnoguchii_ANO.gtf
│   ├── Lsantarosai
│   │   ├── FAA
│   │   │   ├── Lsantarosai_converted.faa
│   │   │   └── Lsantarosai.faa
│   │   ├── FeatureTable
│   │   │   └── Lsantarosai.txt
│   │   ├── FFN
│   │   │   ├── Lsantarosai_converted.ffn
│   │   │   └── Lsantarosai.ffn
│   │   ├── FNA
│   │   │   ├── Lsantarosai.fna
│   │   │   ├── Lsantarosai.fna.amb
│   │   │   ├── Lsantarosai.fna.ann
│   │   │   ├── Lsantarosai.fna.bwt
│   │   │   ├── Lsantarosai.fna.pac
│   │   │   └── Lsantarosai.fna.sa
│   │   ├── GFF
│   │   │   ├── Lsantarosai_ANO.gff
│   │   │   └── Lsantarosai.gff
│   │   └── GTF
│   │       └── Lsantarosai_ANO.gtf
│   ├── Ltipperaryensis
│   │   ├── FAA
│   │   │   ├── Ltipperaryensis_converted.faa
│   │   │   └── Ltipperaryensis.faa
│   │   ├── FeatureTable
│   │   │   └── Ltipperaryensis.txt
│   │   ├── FFN
│   │   │   ├── Ltipperaryensis_converted.ffn
│   │   │   └── Ltipperaryensis.ffn
│   │   ├── FNA
│   │   │   ├── Ltipperaryensis.fna
│   │   │   ├── Ltipperaryensis.fna.amb
│   │   │   ├── Ltipperaryensis.fna.ann
│   │   │   ├── Ltipperaryensis.fna.bwt
│   │   │   ├── Ltipperaryensis.fna.pac
│   │   │   └── Ltipperaryensis.fna.sa
│   │   ├── GFF
│   │   │   ├── Ltipperaryensis_ANO.gff
│   │   │   └── Ltipperaryensis.gff
│   │   └── GTF
│   │       └── Ltipperaryensis_ANO.gtf
│   └── Lweillii
│       ├── FAA
│       │   ├── Lweillii_converted.faa
│       │   └── Lweillii.faa
│       ├── FeatureTable
│       │   └── Lweillii.txt
│       ├── FFN
│       │   ├── Lweillii_converted.ffn
│       │   └── Lweillii.ffn
│       ├── FNA
│       │   ├── Lweillii.fna
│       │   ├── Lweillii.fna.amb
│       │   ├── Lweillii.fna.ann
│       │   ├── Lweillii.fna.bwt
│       │   ├── Lweillii.fna.pac
│       │   └── Lweillii.fna.sa
│       ├── GFF
│       │   ├── Lweillii_ANO.gff
│       │   └── Lweillii.gff
│       └── GTF
│           └── Lweillii_ANO.gtf
├
├── geneAssociation.tsv
├── genomeTag.txt
├── graph.gexf
├── locus.txt
└── protein_files
    ├── Ladleri_converted.faa
    ├── Ldzianensis_converted.faa
    ├── Lgomenensis_converted.faa
    │
```

```
├── Linterrogans_converted.faa
├── Lmayottensis_converted.faa
├── Lnoguchii_converted.faa
├── Lsantarosai_converted.faa
├── Ltipperaryensis_converted.faa
└── Lweillii_converted.faa
```

Once the output files are generated, make sure to save them by renaming the geneAssociation.tsv and CountsComparisonOutput.csv along with Project/GFF/*ANO_converted.gff. These files need to be saved for each method as running the second method will over-write the files.

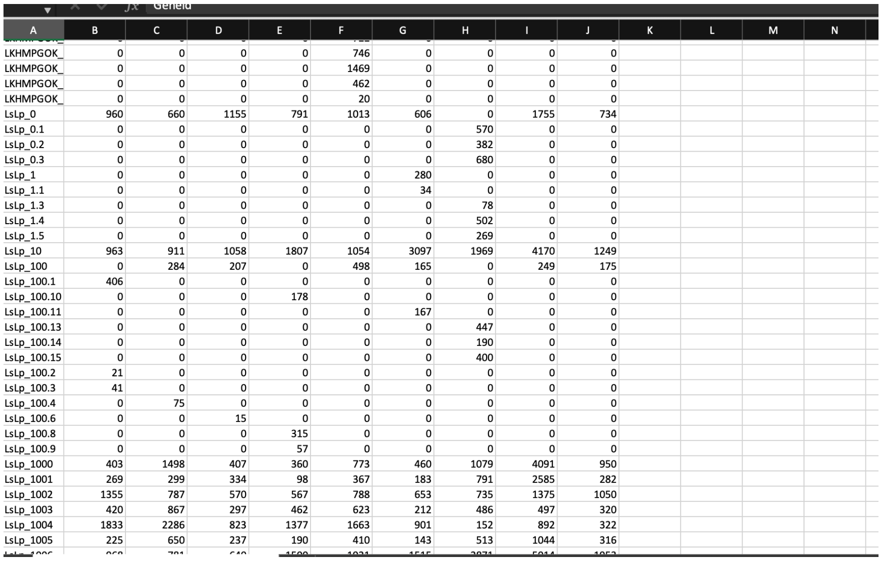
You can open geneAssociation.tsv to check presence/absence of homologous genes in Excel.



| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | 3809 | LsLp_4547.1 | JKCPELNK_02811 | | | | | | | | |
| 8 | 3810 | LsLp_4547.2 | JKCPELNK_01271 | | | | | | | | |
| 9 | 3811 | LsLp_4547.3 | JKCPELNK_01502 | | | | | | | | |
| 10 | 3812 | LsLp_4547.4 | JKCPELNK_01288 | | | | | | | | |
| 11 | 3813 | LsLp_4581.1 | GMMNFEFE_03918 | | | | | | | | |
| 12 | 3814 | LsLp_4581.2 | GMMNFEFE_03931 | | | | | | | | |
| 13 | 3815 | LsLp_0 | LIMLP_RS01 | JKCPELNK_0 | PHPMPPPP_ | AGILFJDJ_00 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | MDKICLNP_00185 |
| 14 | 3816 | LsLp_1 | - | JKCPELNK_0 | - | - | - | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | - |
| 15 | 3817 | LsLp_2 | LIMLP_RS17 | JKCPELNK_0 | PHPMPPPP_ | AGILFJDJ_03 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | MDKICLNP_01836 |
| 16 | 3818 | LsLp_3 | LIMLP_RS17 | JKCPELNK_0 | PHPMPPPP_ | AGILFJDJ_03 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | MDKICLNP_01837 |
| 17 | 3819 | LsLp_4 | LIMLP_RS17 | JKCPELNK_0 | PHPMPPPP_ | AGILFJDJ_03 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | MDKICLNP_01840 |
| 18 | 3820 | LsLp_5 | LIMLP_RS17 | JKCPELNK_0 | PHPMPPPP_ | AGILFJDJ_03 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | MDKICLNP_01841 |
| 19 | 3821 | LsLp_6 | LIMLP_RS17 | JKCPELNK_0 | PHPMPPPP_ | - | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | - |
| 20 | 3822 | LsLp_7 | LIMLP_RS17 | JKCPELNK_0 | PHPMPPPP_ | AGILFJDJ_03 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | MDKICLNP_01842 |
| 21 | 3823 | LsLp_8 | LIMLP_RS17 | JKCPELNK_0 | PHPMPPPP_ | AGILFJDJ_03 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | MDKICLNP_01843 |
| 22 | 3824 | LsLp_9 | LIMLP_RS17 | JKCPELNK_0 | PHPMPPPP_ | AGILFJDJ_03 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | MDKICLNP_01844 |
| 23 | 3825 | LsLp_10 | LIMLP_RS17 | JKCPELNK_0 | PHPMPPPP_ | AGILFJDJ_03 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | MDKICLNP_01845 |
| 24 | 3826 | LsLp_11 | LIMLP_RS18 | JKCPELNK_0 | PHPMPPPP_ | AGILFJDJ_03 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | MDKICLNP_03423 |
| 25 | 3827 | LsLp_12 | LIMLP_RS18 | JKCPELNK_0 | PHPMPPPP_ | AGILFJDJ_03 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | [illegible] |
| 26 | 3828 | LsLp_13 | LIMLP_RS18 | JKCPELNK_0 | PHPMPPPP_ | AGILFJDJ_03 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | [illegible] |
| 27 | 3829 | LsLp_14 | LIMLP_RS18 | JKCPELNK_0 | PHPMPPPP_ | AGILFJDJ_03 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | MDKICLNP_03420 |
| 28 | 3830 | LsLp_15 | LIMLP_RS18 | - | PHPMPPPP_ | AGILFJDJ_03 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | MDKICLNP_03419 |
| 29 | 3831 | LsLp_16 | LIMLP_RS18 | JKCPELNK_0 | PHPMPPPP_ | AGILFJDJ_03 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | MDKICLNP_03418 |
| 30 | 3832 | LsLp_17 | LIMLP_RS18 | JKCPELNK_0 | PHPMPPPP_ | AGILFJDJ_03 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | MDKICLNP_03417 |
| 31 | 3833 | LsLp_18 | LIMLP_RS14 | JKCPELNK_0 | PHPMPPPP_ | AGILFJDJ_00 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | MDKICLNP_03504,MDKICLNP_02653,MDKICLNP_02452 |
| 32 | 3834 | LsLp_19 | LIMLP_RS18 | - | PHPMPPPP_ | AGILFJDJ_03 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | MDKICLNP_03415 |
| 33 | 3835 | LsLp_20 | LIMLP_RS18 | JKCPELNK_0 | PHPMPPPP_ | AGILFJDJ_03 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | MDKICLNP_03414 |
| 34 | 3836 | LsLp_21 | LIMLP_RS18 | JKCPELNK_0 | PHPMPPPP_ | AGILFJDJ_03 | LKHMPGOK_ | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | MDKICLNP_03413 |
| 35 | 3837 | LsLp_22 | - | - | - | - | - | GMMNFEFE_ | NPKIDGLO_0 | FMDNDCAP_ | - |

As depicted in the figure above, the locus tags of proteins present in the genomes are shown. Singletons can be seen at the beginning of the file generated using Network connection. A "-" symbol indicates the absence of the protein in the genome.

Similarly, in the CountComparisonOutput.csv generated you will be able to see the featureCounts of the reannotated genes as shown in below figure.



| A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LKHMPGOK_ | 0 | 0 | 0 | 0 | 746 | 0 | 0 | 0 | 0 | | | | |
| LKHMPGOK_ | 0 | 0 | 0 | 0 | 1469 | 0 | 0 | 0 | 0 | | | | |
| LKHMPGOK_ | 0 | 0 | 0 | 0 | 462 | 0 | 0 | 0 | 0 | | | | |
| LKHMPGOK_ | 0 | 0 | 0 | 0 | 20 | 0 | 0 | 0 | 0 | | | | |
| LsLp_0 | 960 | 660 | 1155 | 791 | 1013 | 606 | 0 | 1755 | 734 | | | | |
| LsLp_0.1 | 0 | 0 | 0 | 0 | 0 | 0 | 570 | 0 | 0 | | | | |
| LsLp_0.2 | 0 | 0 | 0 | 0 | 0 | 0 | 382 | 0 | 0 | | | | |
| LsLp_0.3 | 0 | 0 | 0 | 0 | 0 | 0 | 680 | 0 | 0 | | | | |
| LsLp_1 | 0 | 0 | 0 | 0 | 0 | 280 | 0 | 0 | 0 | | | | |
| LsLp_1.1 | 0 | 0 | 0 | 0 | 0 | 34 | 0 | 0 | 0 | | | | |
| LsLp_1.3 | 0 | 0 | 0 | 0 | 0 | 0 | 78 | 0 | 0 | | | | |
| LsLp_1.4 | 0 | 0 | 0 | 0 | 0 | 0 | 502 | 0 | 0 | | | | |
| LsLp_1.5 | 0 | 0 | 0 | 0 | 0 | 0 | 269 | 0 | 0 | | | | |
| LsLp_10 | 963 | 911 | 1058 | 1807 | 1054 | 3097 | 1969 | 4170 | 1249 | | | | |
| LsLp_100 | 0 | 284 | 207 | 0 | 498 | 165 | 0 | 249 | 175 | | | | |
| LsLp_100.1 | 406 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | |
| LsLp_100.10 | 0 | 0 | 0 | 178 | 0 | 0 | 0 | 0 | 0 | | | | |
| LsLp_100.11 | 0 | 0 | 0 | 0 | 0 | 167 | 0 | 0 | 0 | | | | |
| LsLp_100.13 | 0 | 0 | 0 | 0 | 0 | 0 | 447 | 0 | 0 | | | | |
| LsLp_100.14 | 0 | 0 | 0 | 0 | 0 | 0 | 190 | 0 | 0 | | | | |
| LsLp_100.15 | 0 | 0 | 0 | 0 | 0 | 0 | 400 | 0 | 0 | | | | |
| LsLp_100.2 | 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | |
| LsLp_100.3 | 41 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | |
| LsLp_100.4 | 0 | 75 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | |
| LsLp_100.6 | 0 | 0 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | | | | |
| LsLp_100.8 | 0 | 0 | 0 | 315 | 0 | 0 | 0 | 0 | 0 | | | | |
| LsLp_100.9 | 0 | 0 | 0 | 57 | 0 | 0 | 0 | 0 | 0 | | | | |
| LsLp_1000 | 403 | 1498 | 407 | 360 | 773 | 460 | 1079 | 4091 | 950 | | | | |
| LsLp_1001 | 269 | 299 | 334 | 98 | 367 | 183 | 791 | 2585 | 282 | | | | |
| LsLp_1002 | 1355 | 787 | 570 | 567 | 788 | 653 | 735 | 1375 | 1050 | | | | |
| LsLp_1003 | 420 | 867 | 297 | 462 | 623 | 212 | 486 | 497 | 320 | | | | |
| LsLp_1004 | 1833 | 2286 | 823 | 1377 | 1663 | 901 | 152 | 892 | 322 | | | | |
| LsLp_1005 | 225 | 650 | 237 | 190 | 410 | 143 | 513 | 1044 | 316 | | | | |

## Frequent warnings and errors

The software pops up messages if there are errors in the files.