



Session 3: Bacterial Genomic Analysis

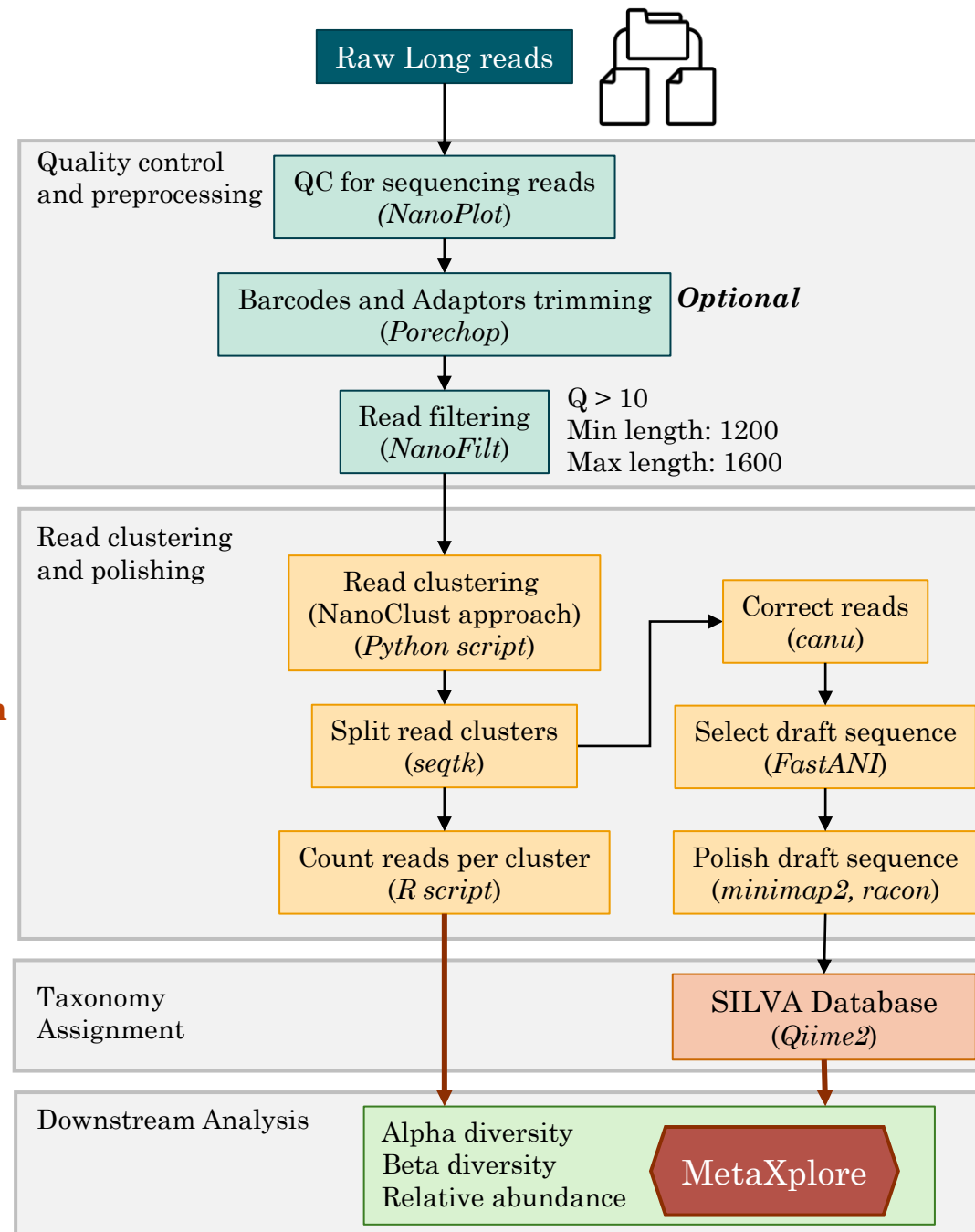
Hands-on Targeted Metagenomic Analysis

Analysis Pipeline

ProcessInput.sh

NB_ReadClust.sh

AssignTaxa.sh



Dataset Overview

Aedes aegypti

Primary vector for dengue, Zika, chikungunya, and yellow fever

Importance of studying the *Ae. aegypti* associated bacteriome

- Influences mosquito biology and behavior
- Affects vector competence (disease transmission potential)
- Potential for developing new mosquito control strategies
- Can reduce mosquito-borne disease transmission by manipulating microbial communities



Aedes aegypti

Samples Used during the Hands-On

Developmental stage	Sex	N. Samples
Larva	Larva	4
Adult	Female	4
Adult	Male	4

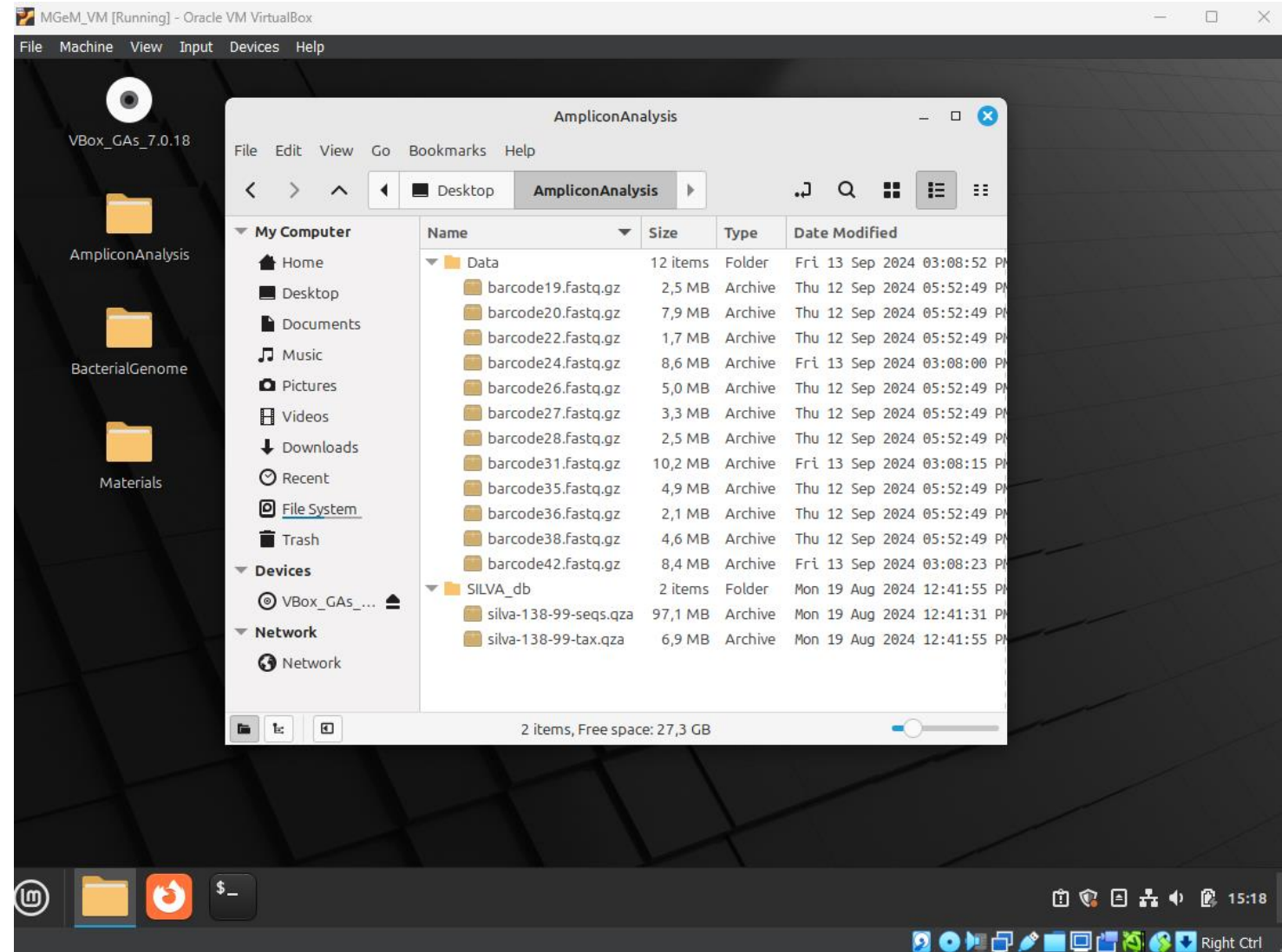


LABORATORY OF SYSTEMS
MICROBIOLOGY & APPLIED GENOMICS

Starting point

Working directory:

`/home/mgem/Desktop/AmpliconAnalysis/`



Starting point

```
mgem@mgem-VM: ~/Desktop/AmpliconAnalysis
File Edit View Search Terminal Help
mgem@mgem-VM:~$ cd Desktop/AmpliconAnalysis
mgem@mgem-VM:~/Desktop/AmpliconAnalysis$ conda info --envs
# conda environments:
#
base                /home/mgem/miniconda3
AmpliconTools       /home/mgem/miniconda3/envs/AmpliconTools
GenomeTools         /home/mgem/miniconda3/envs/GenomeTools
qiime2-amplicon-2024.5  /home/mgem/miniconda3/envs/qiime2-amplicon-2024.5

mgem@mgem-VM:~/Desktop/AmpliconAnalysis$ conda activate AmpliconTools
(AmpliconTools) mgem@mgem-VM:~/Desktop/AmpliconAnalysis$
```

Quality control: Raw data

NanoPlot

General summary:

Mean read length:	745.4
Mean read quality:	14.8
Median read length:	366.0
Median read quality:	16.4
Number of reads:	94,253.0
Read length N50:	1,440.0
STDEV read length:	587.4
Total bases:	70,256,781.0

Number, percentage and megabases of reads above quality cutoffs

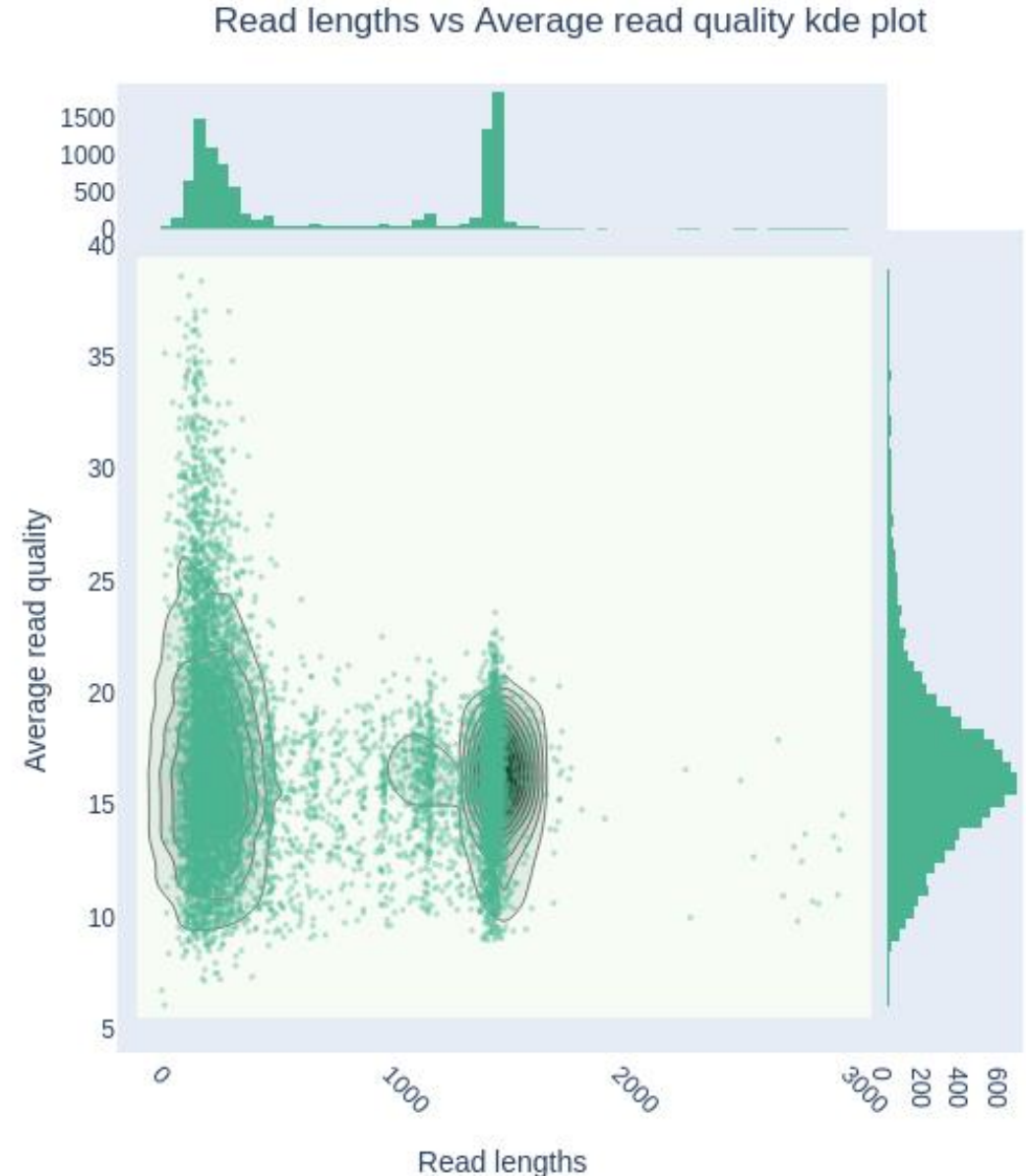
>Q10:	92427 (98.1%)	69.3Mb
>Q15:	63483 (67.4%)	47.7Mb
>Q20:	14329 (15.2%)	4.7Mb
>Q25:	4225 (4.5%)	0.8Mb
>Q30:	1386 (1.5%)	0.2Mb

Top 5 highest mean basecall quality scores and their read lengths

1:	45.4 (5)
2:	43.1 (68)
3:	42.2 (2)
4:	42.2 (62)
5:	41.5 (64)

Top 5 longest reads and their mean basecall quality score

1:	4621 (13.5)
2:	4358 (13.3)
3:	4274 (12.4)
4:	4128 (11.9)
5:	3823 (17.6)



Quality control: Raw data

ReadCount.sh

The screenshot displays a virtual machine interface with a file explorer and a terminal window. The file explorer shows a directory structure with a 'Data' folder containing various fastq files and a 'read_counts.txt' file. The terminal window shows the execution of a script named 'ReadCount.sh' which outputs the read counts for each file.

File Explorer - Data Directory

Name	Size	Type	Date
Nanoplot_raw	11 items	Folder	Fri 12
barcode19.fastq	6,9 MB	Text	Thu 12
barcode20.fastq	21,2 MB	Text	Thu 12
barcode22.fastq	4,5 MB	Text	Thu 12
barcode24.fastq	22,9 MB	Text	Fri 12
barcode26.fastq	15,2 MB	Text	Thu 12
barcode27.fastq	9,9 MB	Text	Thu 12
barcode28.fastq	7,8 MB	Text	Thu 12
barcode31.fastq	29,5 MB	Text	Fri 12
barcode35.fastq	15,3 MB	Text	Thu 12
barcode36.fastq	6,6 MB	Text	Thu 12
barcode38.fastq	14,1 MB	Text	Thu 12
barcode42.fastq	23,8 MB	Text	Fri 13
read_counts.txt	300 bytes	Text	Fri 13

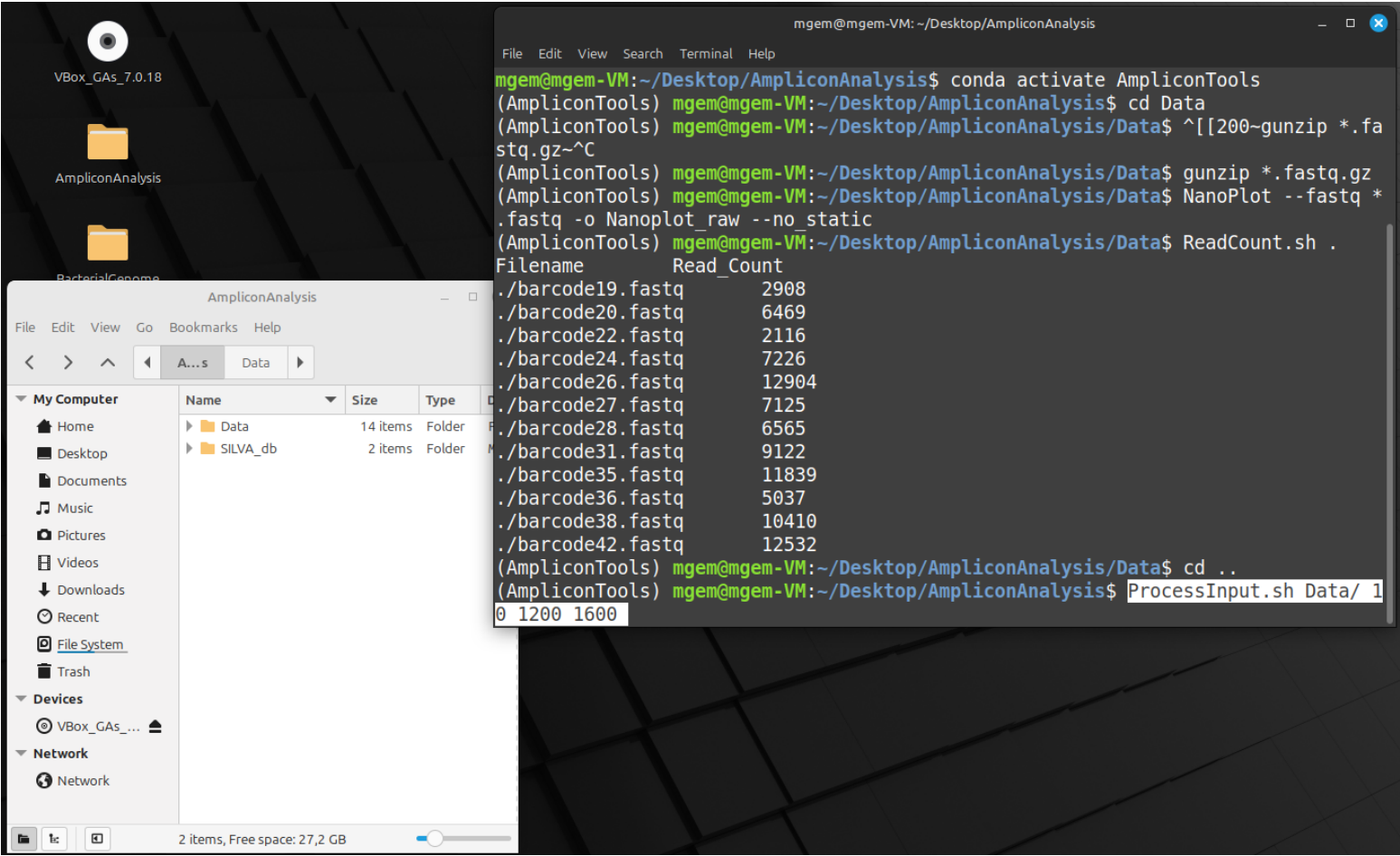
Terminal Output

```
mgem@mgem-VM: ~/Desktop/AmpliconAnalysis/Data
qiime2-amplicon-2024.5 /home/mgem/miniconda3/envs/qiime2-amplicon-2024.5

mgem@mgem-VM:~/Desktop/AmpliconAnalysis$ conda activate AmpliconTools
(AmpliconTools) mgem@mgem-VM:~/Desktop/AmpliconAnalysis$ cd Data
(AmpliconTools) mgem@mgem-VM:~/Desktop/AmpliconAnalysis/Data$ ^[[200~gunzip *.fa
stq.gz~^C
(AmpliconTools) mgem@mgem-VM:~/Desktop/AmpliconAnalysis/Data$ gunzip *.fastq.gz
(AmpliconTools) mgem@mgem-VM:~/Desktop/AmpliconAnalysis/Data$ NanoPlot --fastq *
.fastq -o Nanoplot_raw --no_static
(AmpliconTools) mgem@mgem-VM:~/Desktop/AmpliconAnalysis/Data$ ReadCount.sh .
Filename      Read_Count
./barcode19.fastq 2908
./barcode20.fastq 6469
./barcode22.fastq 2116
./barcode24.fastq 7226
./barcode26.fastq 12904
./barcode27.fastq 7125
./barcode28.fastq 6565
./barcode31.fastq 9122
./barcode35.fastq 11839
./barcode36.fastq 5037
./barcode38.fastq 10410
./barcode42.fastq 12532
(AmpliconTools) mgem@mgem-VM:~/Desktop/AmpliconAnalysis/Data$
```

Reads filtering

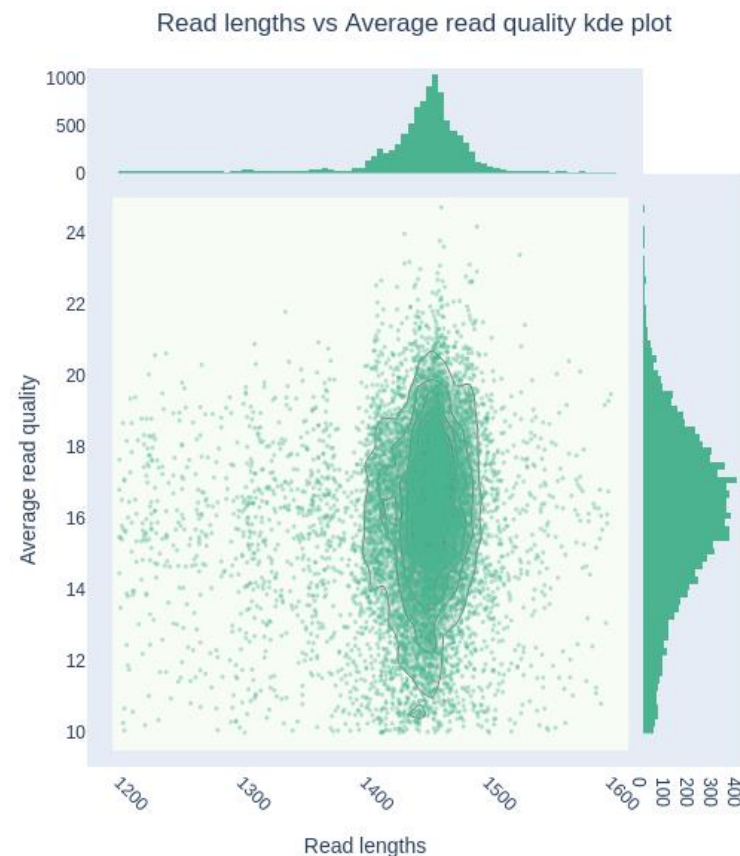
ProcessInput.sh



Name	Size	Type	Dat
▶ Data	14 items	Folder	Fri
▼ Results	1 item	Folder	Fri
▶ Processed_fq	8 items	Folder	Fri
▶ SILVA_db	2 items	Folder	Mon

Reads filtering: Quality Check

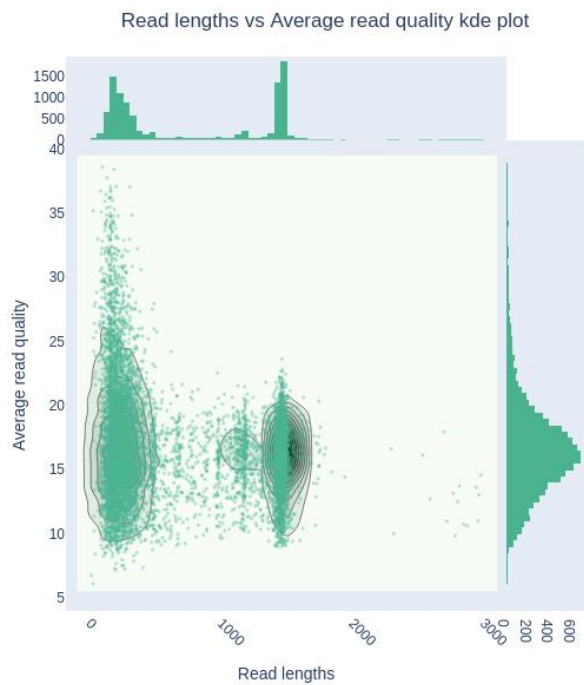
Name	Size	Type
▶ Data	14 items	Folder
▼ Results	1 item	Folder
▶ Processed_fq	14 items	Folder
▶ Nanoplot_filt	10 items	Folder
barcode19.fastq	4,2 MB	Text
barcode20.fastq	18,7 MB	Text
barcode22.fastq	2,7 MB	Text
barcode24.fastq	17,8 MB	Text
barcode26.fastq	2,0 MB	Text
barcode27.fastq	3,5 MB	Text
barcode28.fastq	2,4 MB	Text
barcode31.fastq	23,7 MB	Text
barcode35.fastq	5,1 MB	Text
barcode36.fastq	2,0 MB	Text
barcode38.fastq	4,7 MB	Text
barcode42.fastq	12,8 MB	Text
read_counts.txt	292 bytes	Text
▶ SILVA_db	2 items	Folder



Filename	Read_Count
./barcode19.fastq	1446
./barcode20.fastq	6451
./barcode22.fastq	928
./barcode24.fastq	6213
./barcode26.fastq	694
./barcode27.fastq	1208
./barcode28.fastq	818
./barcode31.fastq	8120
./barcode35.fastq	1749
./barcode36.fastq	685
./barcode38.fastq	1600
./barcode42.fastq	4377

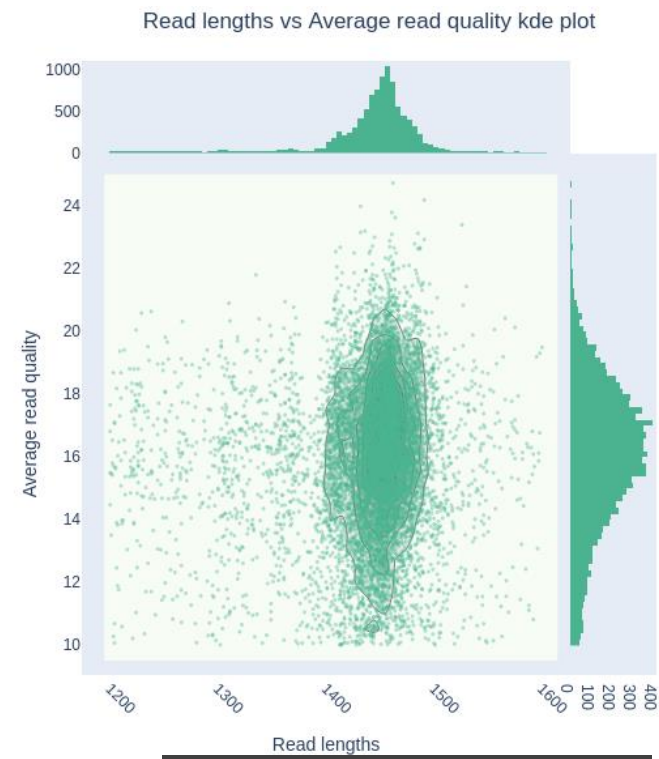
Reads filtering: Quality Check

Before



Filename	Read_Count
./barcode19.fastq	2908
./barcode20.fastq	6469
./barcode22.fastq	2116
./barcode24.fastq	7226
./barcode26.fastq	12904
./barcode27.fastq	7125
./barcode28.fastq	6565
./barcode31.fastq	9122
./barcode35.fastq	11839
./barcode36.fastq	5037
./barcode38.fastq	10410
./barcode42.fastq	12532

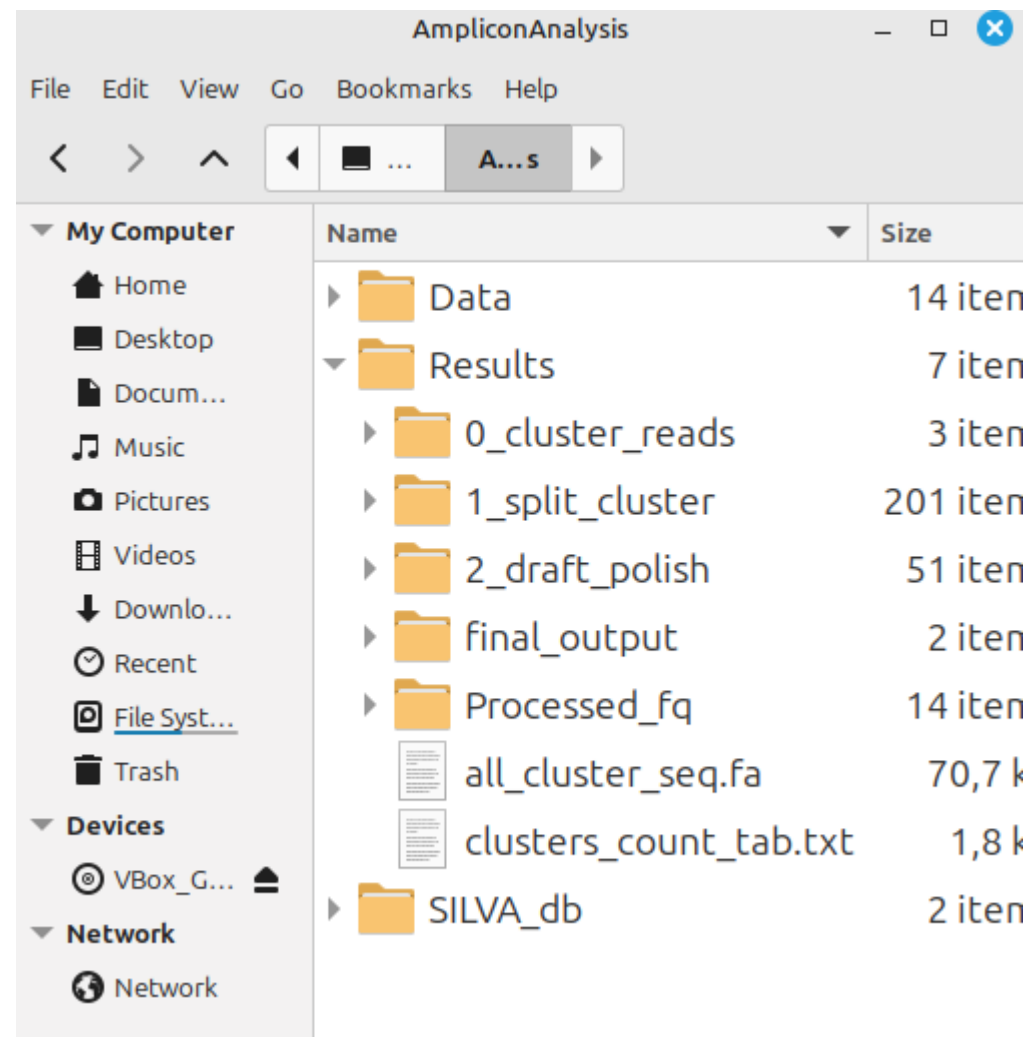
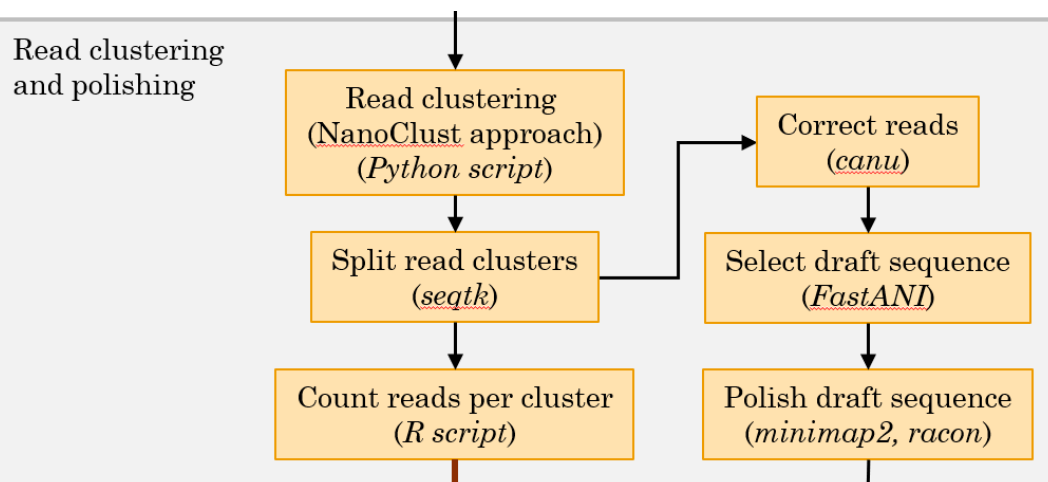
After



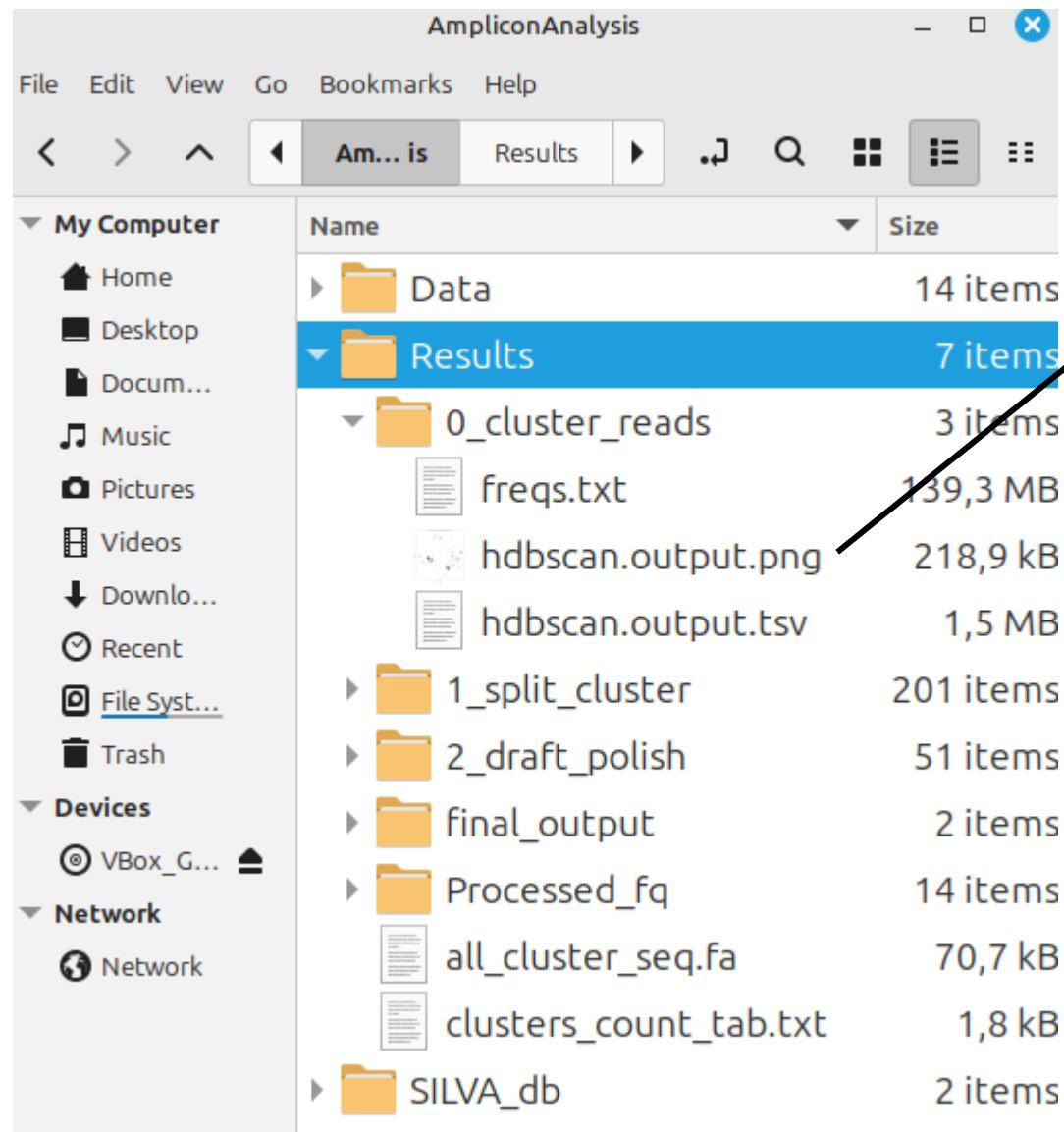
Filename	Read_Count
./barcode19.fastq	1446
./barcode20.fastq	6451
./barcode22.fastq	928
./barcode24.fastq	6213
./barcode26.fastq	694
./barcode27.fastq	1208
./barcode28.fastq	818
./barcode31.fastq	8120
./barcode35.fastq	1749
./barcode36.fastq	685
./barcode38.fastq	1600
./barcode42.fastq	4377

Reads clustering and polishing

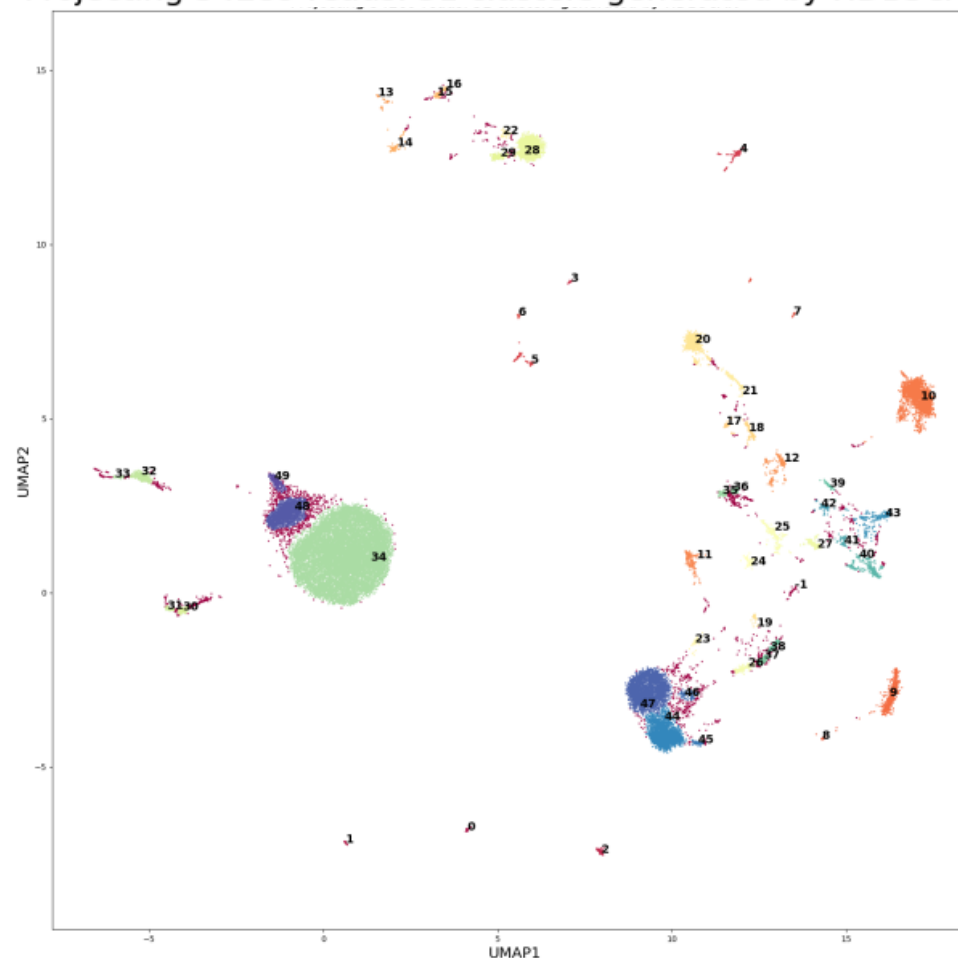
NB_ReadClust.sh



Reads clustering and polishing

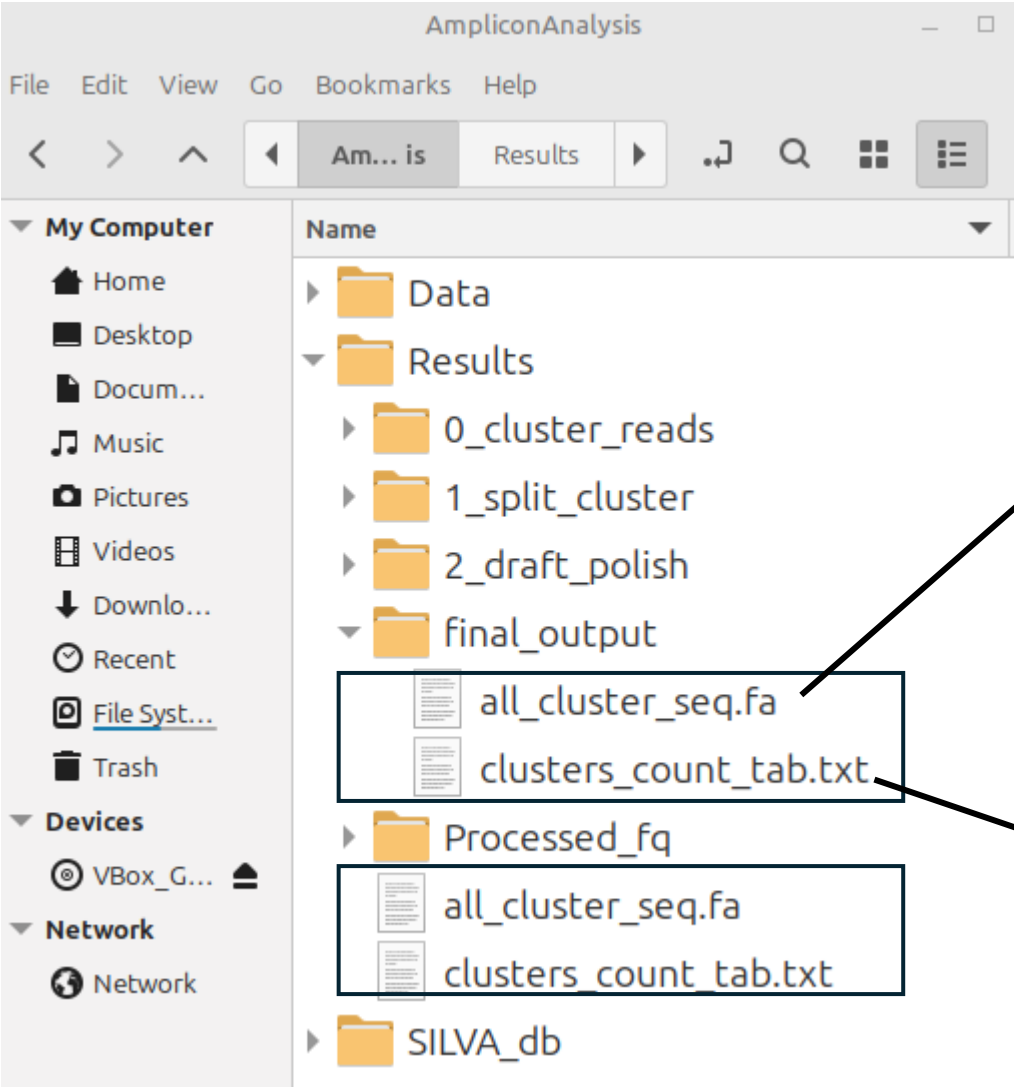


Projecting 34289 reads. 51 clusters generated by HDBSCAN



Rodríguez-Pérez H, Ciuffreda L, Flores C. NanoCLUST: a species-level analysis of 16S rRNA nanopore sequencing data. *Bioinformatics*. 2021 Jul 12;37(11):1600-1601

Reads clustering and polishing

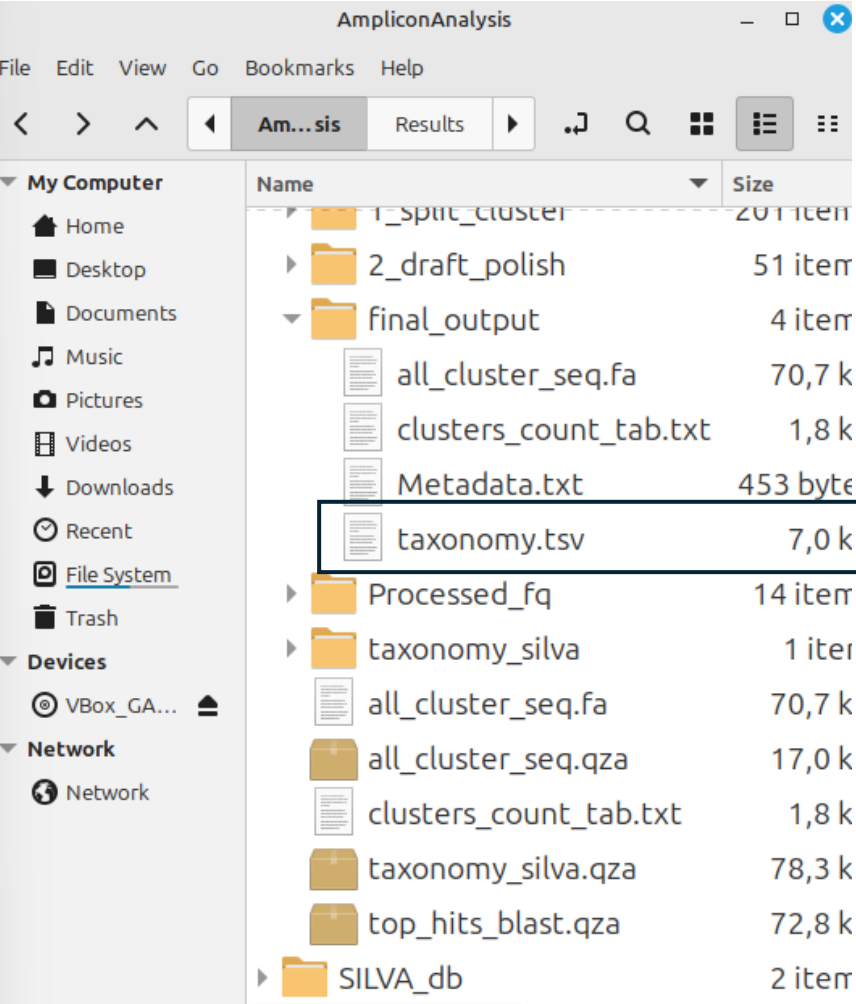


```
all_cluster_seq.fa (~/Desktop/AmpliconAnalysis/Results)
File Edit View Search Tools Documents Help
all_cluster_seq.fa x
>C0
GGCTCAGGATGAACGCTAGCGGGAGGCCTAACACATGCAAGCCGAGCGGTAGAGATTCTTCGGAATCTTGAGAGCGGCGT/
>C1
TTAGCCCTAGTTACTAGTTTTACCCTAGGCAGCTCCTTTTACGGTCACCGACTTCAGGTACCCCAAGCTTCCATGGCTTG/
>C2
GGTTACCTTGTTACGACTTCACCCAGTCGCTGACCCTACCGTGGTCGCCTGCCTCCCTTGCGGGTTAGCCCAGCGCCTTC/
>C3
CGGCCACACCGTGGCAAGCGCCCTCCCGAAGGTTAAGCTACCTGCTTCTGGTGCAACAACTCCCATGGTGTGACGGGCG/
>C4
CAACCTGGCTCATTACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAT/
```

<u>clusterID</u>	barcode22	barcode20	barcode36	barcode38	barcode27
C0	8	43	1	0	0
C1	0	0	0	1	4
C2	13	55	0	0	0
C3	0	3	0	0	1
C4	0	1	6	124	1
C5	0	2	6	9	7
C6	4	3	1	0	0
C7	0	4	1	6	1

Assign taxonomy

AssignTaxa.sh



Feature ID	Taxon
C0	d_Bacteria; p_Bacteroidota; c_Bacteroidia; o_Flavobacteriales; f_Weeksellaceae; g_Chryseobacterium; s_Chryseobacterium_culicis
C1	d_Bacteria; p_Bacteroidota; c_Bacteroidia; o_Flavobacteriales; f_Weeksellaceae; g_Elizabethingia; s_Elizabethingia_meningoseptica
C10	d_Bacteria; p_Proteobacteria; c_Alphaproteobacteria; o_Rhizobiales; f_Xanthobacteraceae; g_Xanthobacter; s_uncultured_Xanthobacter
C11	d_Bacteria; p_Proteobacteria; c_Alphaproteobacteria; o_Rhizobiales; f_Xanthobacteraceae; g_Xanthobacter; s_Xanthobacteraceae_bacterium
C12	d_Bacteria; p_Proteobacteria; c_Alphaproteobacteria; o_Caulobacteriales; f_Caulobacteraceae; g_Brevundimonas; s_uncultured_bacterium



Statistical analysis and Visualization

MetaXplore



```
MGem_VM [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help

VBox_GAS_7.0.18
AmpliconAnalysis
BacterialGenome
Materials

mgem@mgem-VM: ~
File Edit View Search Terminal Help

rotate

The following object is masked from 'package:ape':
  rotate

Attaching package: 'shinyalert'

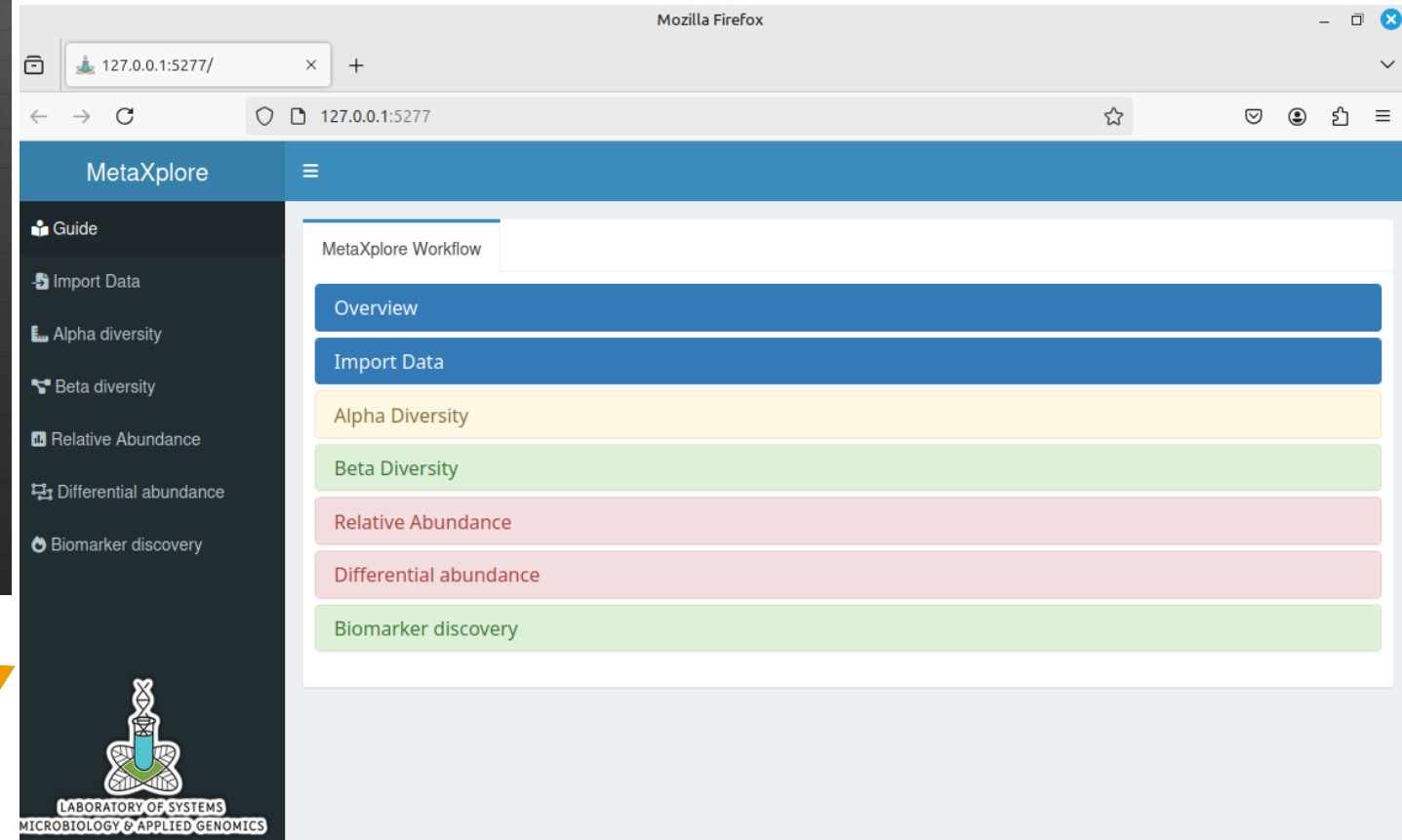
The following object is masked from 'package:shinyjs':
  runExample

The following object is masked from 'package:shinyBS':
  closeAlert

The following object is masked from 'package:shiny':
  runExample

Listening on http://127.0.0.1:5277
```

Ctrl + Click



Statistical analysis and Visualization

MetaXplore

Import Data

MetaXplore

Load Data

Load cluster table: Browse... clusters_c Upload complete

Load Taxa file: Browse... taxonomy: Upload complete

Load Metadata file: Browse... Metadata.t Upload complete

Load tree file (optional): Browse... No file selected

Select normalization: None

Minimum frequency (%): 0

Read Input

NUMBER OF SAMPLES: 12 / 12

NUMBER OF TAXA: 50 / 50

NUMBER OF FACTORS: 3

OTUs table Taxonomy table Metatable Visualize Input

Show 10 entries

Search:

	barcode19	barcode20	barcode22	barcode24	barcode26	barcode27	barcode28	bar
C0	12	43	8	17	0	0	0	
C1	0	0	0	0	6	4	1	

Statistical analysis and Visualization

Import Data

MetaXplore

Guide

Import Data

Alpha diversity

Beta diversity

Relative Abundance

Differential abundance

Biomarker discovery

Load Data

Load cluster table

Browse... clusters_c
Upload complete

Load Taxa file

Browse... taxonomy.
Upload complete

Load Metadata file

Browse... Metadata.t
Upload complete

Load tree file (optional)

Browse... No file sel
Upload complete

Select normalization

OTUs Norm

Minimum frequency (%)

0.01

Read Input

NUMBER OF SAMPL...
12 / 12

NUMBER OF TAXA (...)
39 / 50

NUMBER OF FACTO...
3

OTUs table Taxonomy table Metatable

Show 10 entries

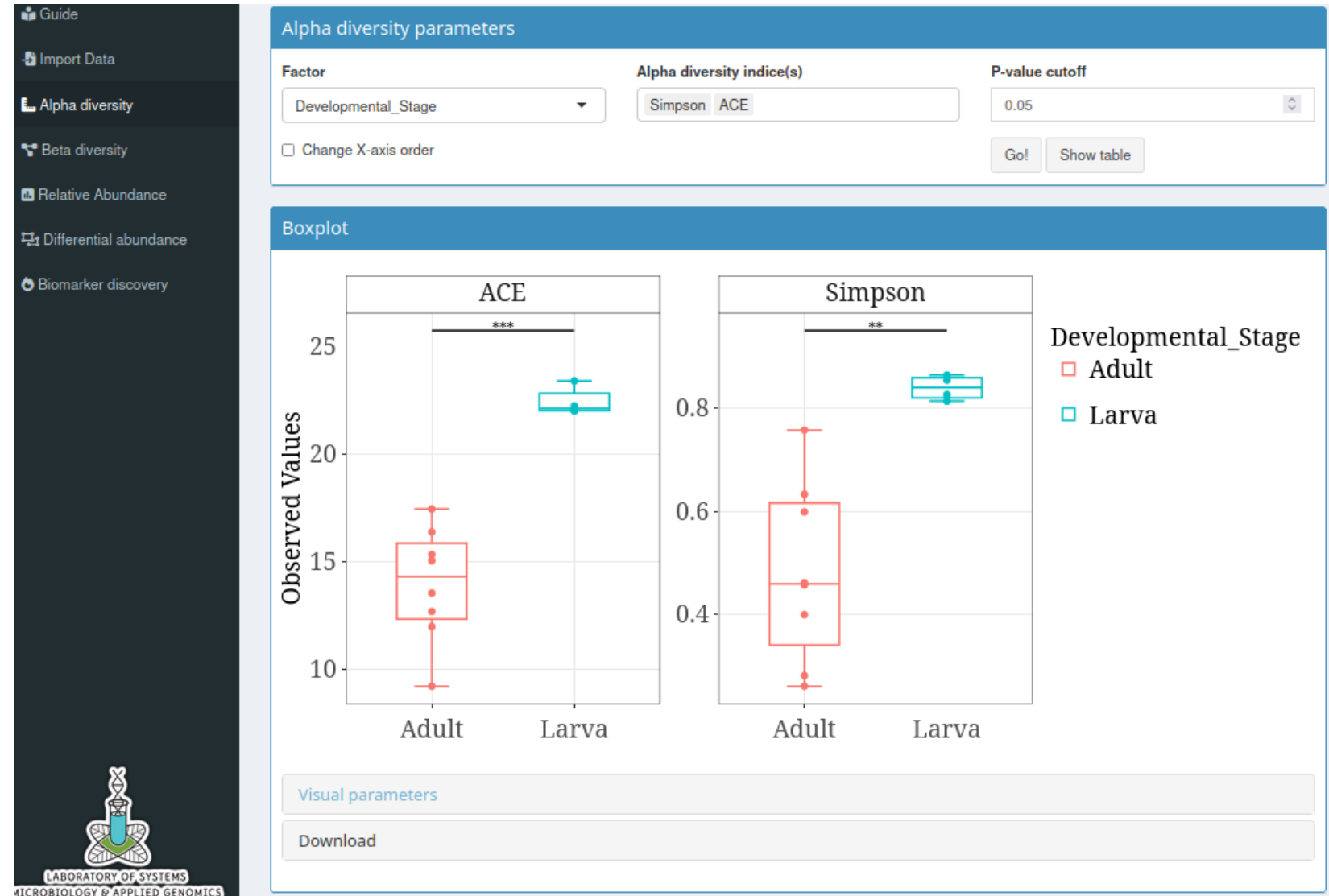
Search:

	barcode19	barcode20	barcode22	barcode24	barcode26	barcode27	barcode28	bar
C10	64	61	66	78	0	0	1	
C11	3	7	12	8	0	0	0	

Statistical analysis and Visualization

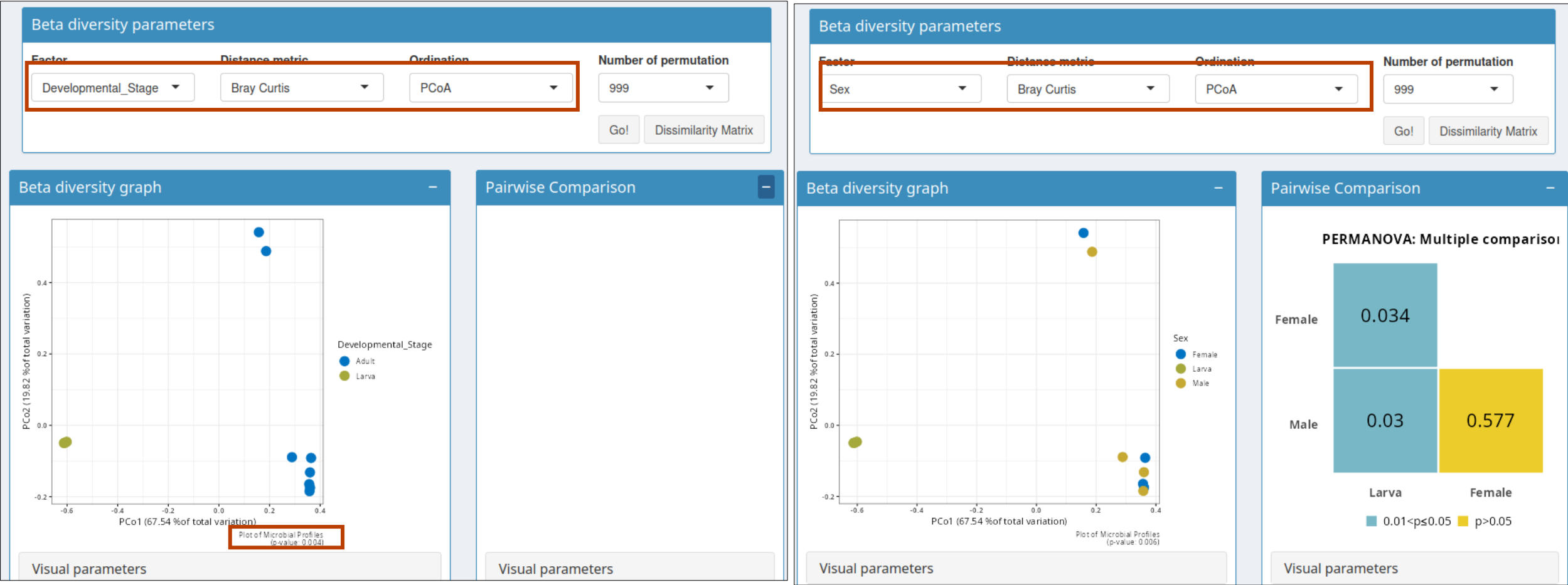
MetaXplore

Alpha diversity



Statistical analysis and Visualization

Beta diversity



Developmental stage

Sex

Statistical analysis and Visualization

Relative abundance

Visual parameters

☒ Labels☐ Color☐ Cluster☐ Re-size☒ Body

Rows labels size

101424

columns labels size

101224

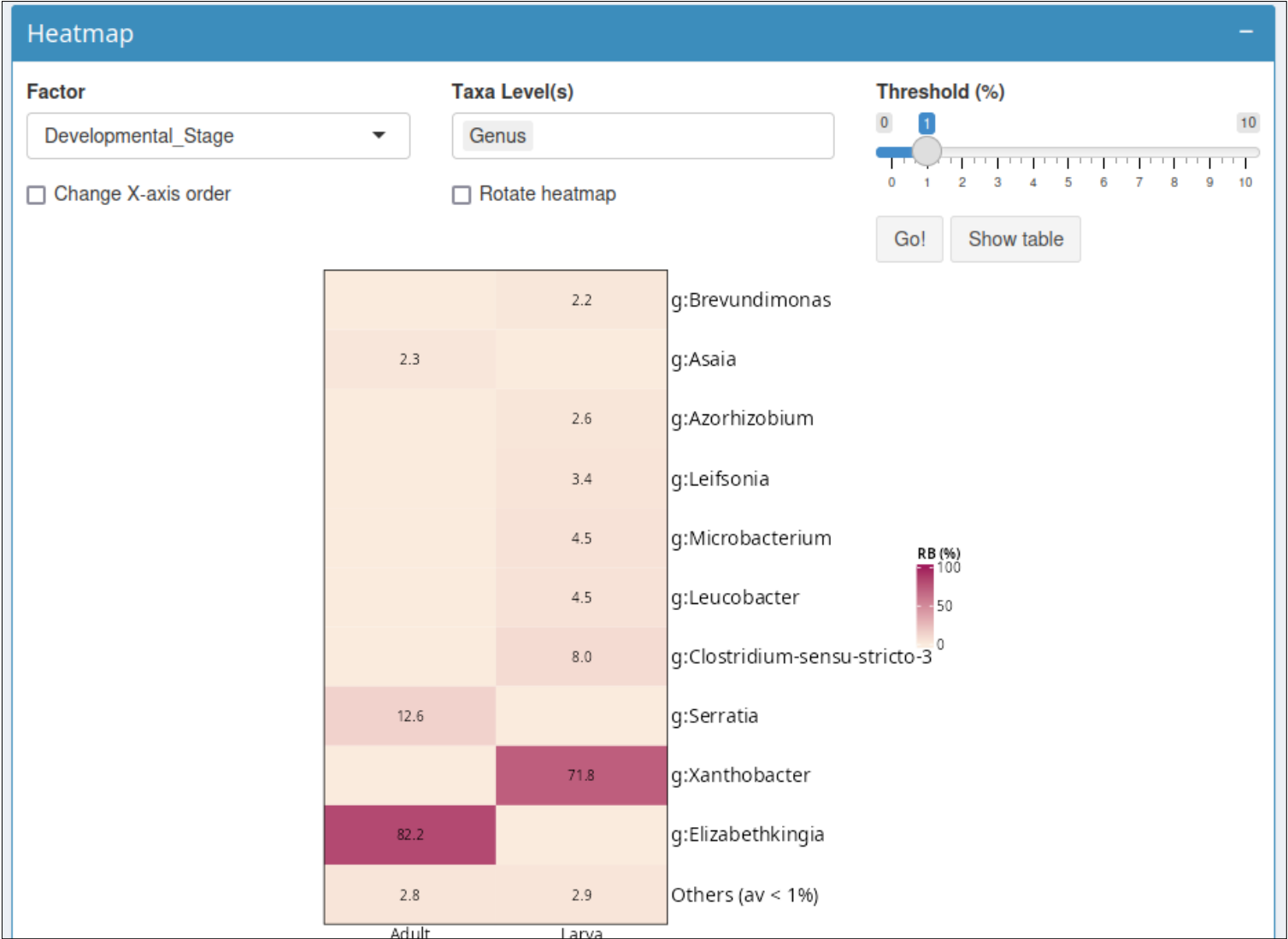
Rows labels rotation

090

columns labels rotation

090

☒ Cells values



Statistical analysis and Visualization

Differential Abundance

