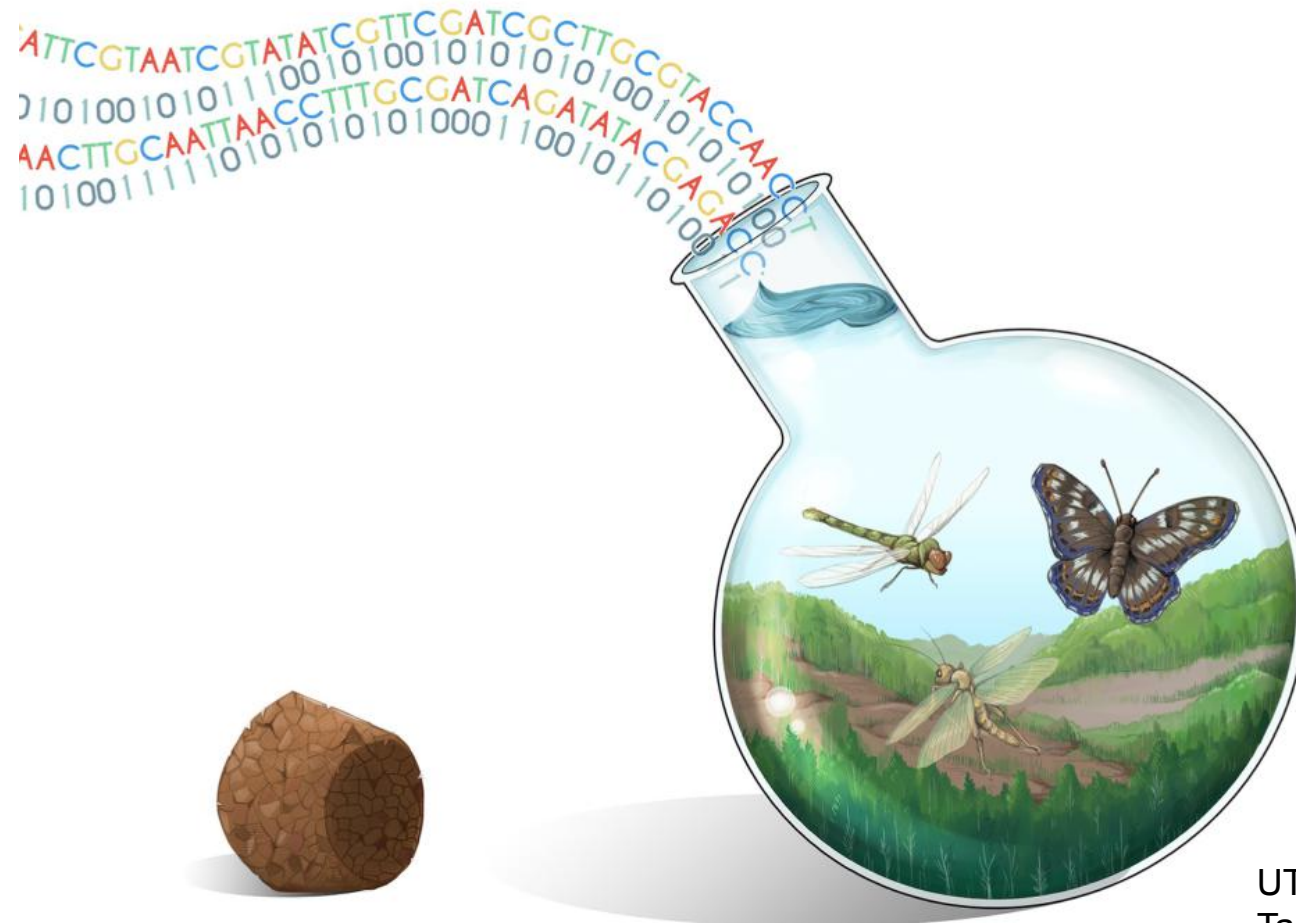


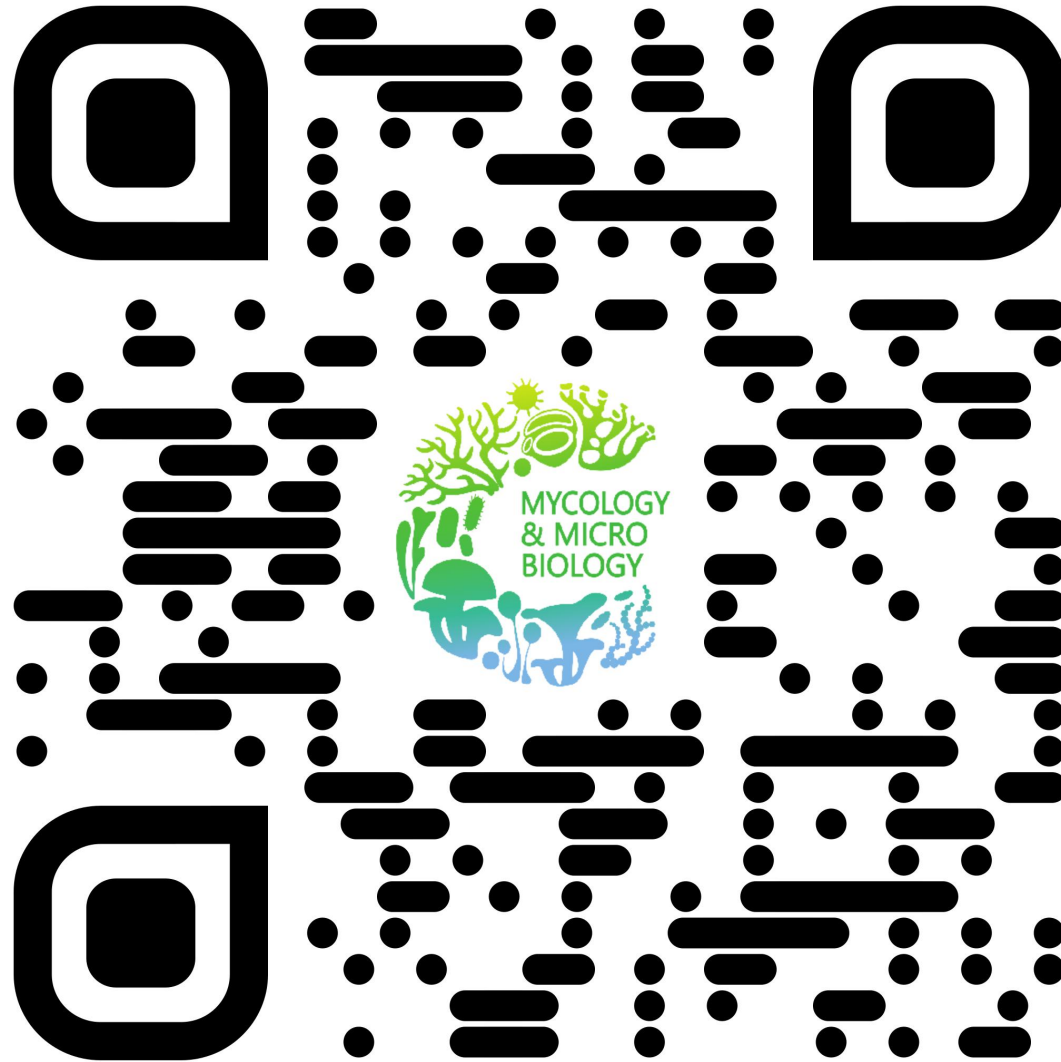


'Expert mode'



Vladimir Mikryukov

UT International Summer University,
Tartu, August 01-04 2022



<https://mycology-microbiology-center.github.io/Metabarcoding2022/>

<https://t.ly/dvZS>

Metabarcoding: from Lab to Bioinformatics

Metabarcoding: from Lab to Bioinformatics (UT International Summer University, 2022)

Metabarcoding: from Lab to Bioinformatics

University of Tartu, 2022

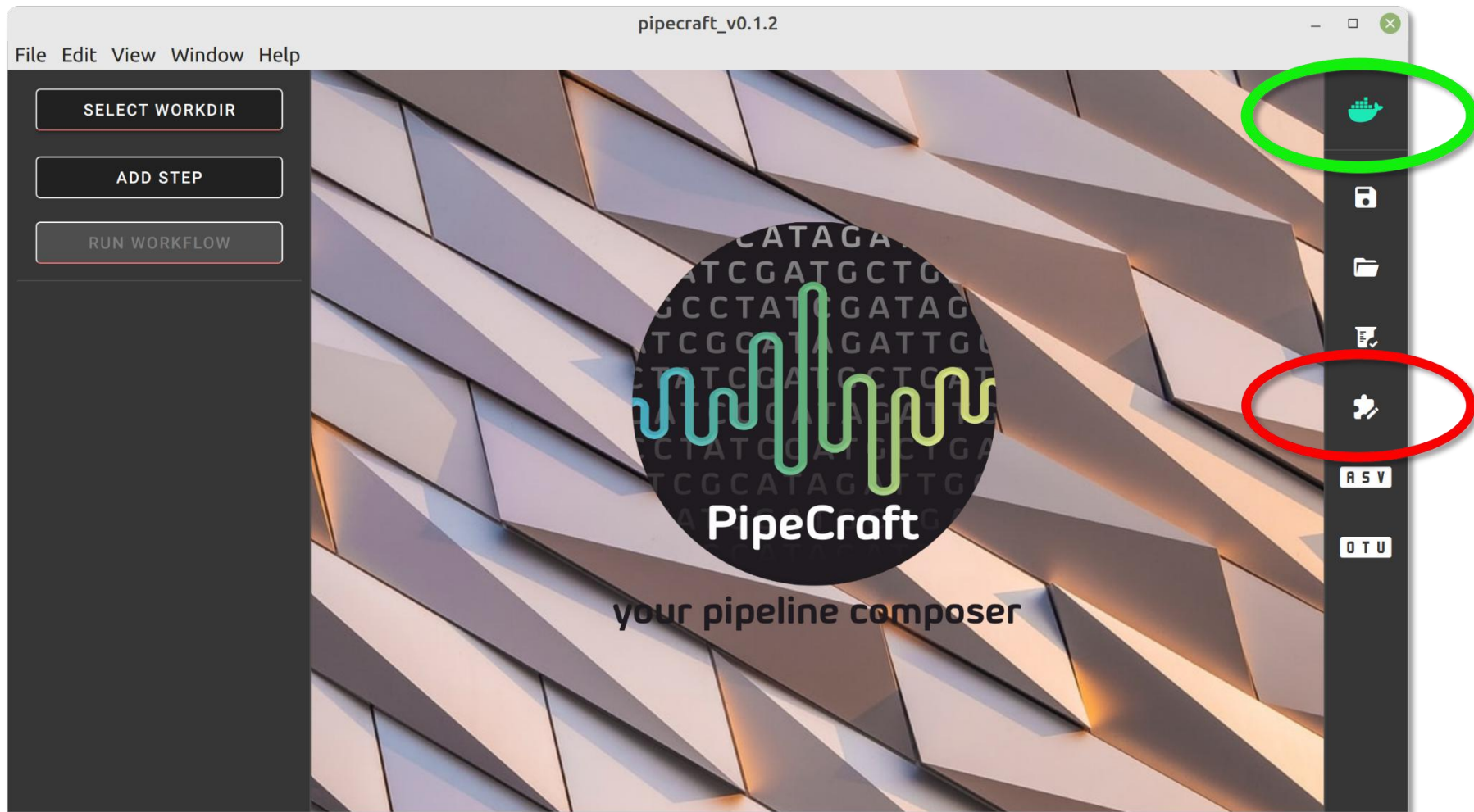
~~Data used during the course~~

~~"Expert mode" commands~~

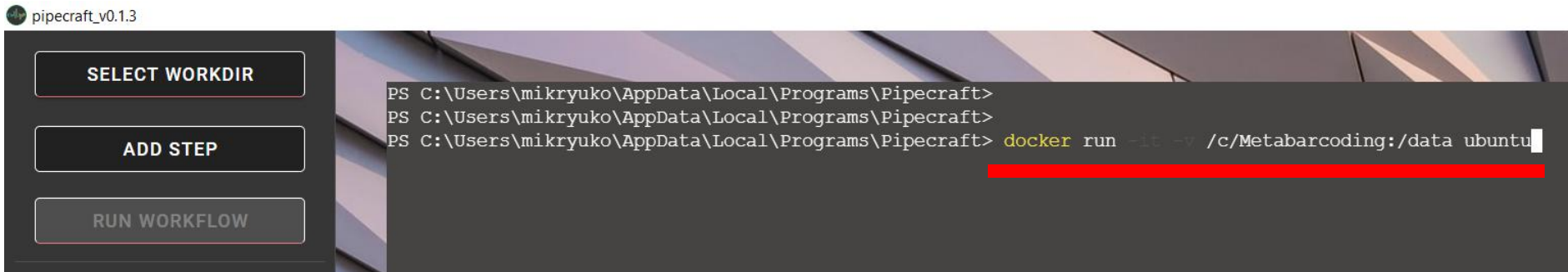
~~Individual projects~~

Slides (will be released after the course)

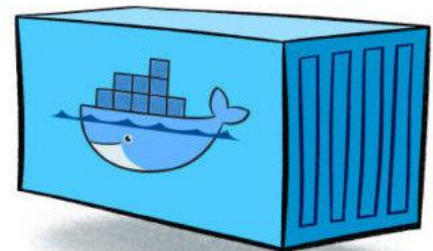
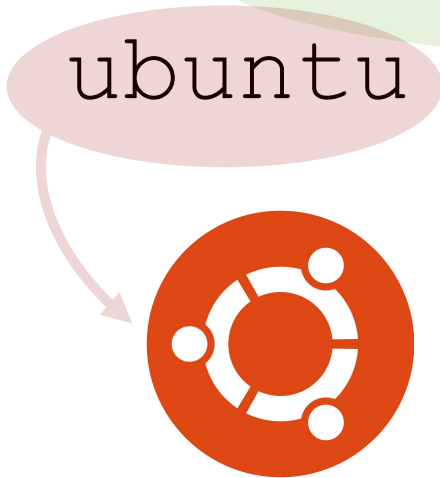
Expert mode in PipeCraft2



Docker

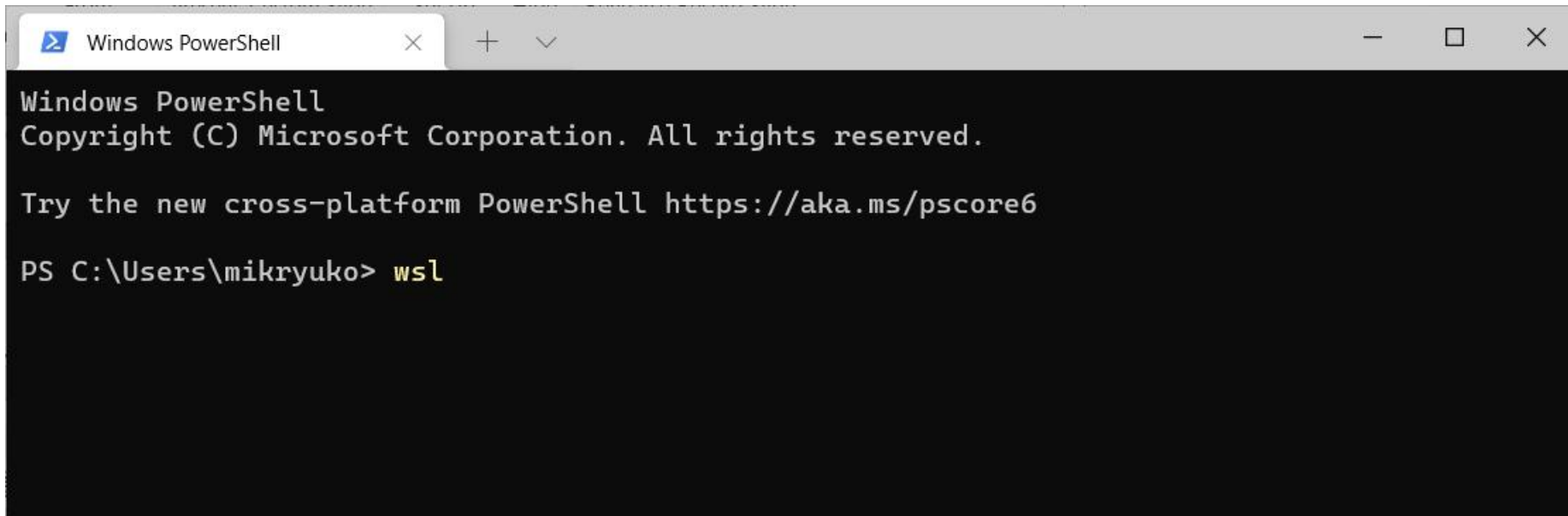


`docker run -it`
`-v c:\Metabarcoding:/data`
`ubuntu`



Windows Terminal

+ Windows Subsystem for Linux (WSL)

A screenshot of a Windows PowerShell terminal window. The title bar shows 'Windows PowerShell' with a blue icon, a close button, and window controls. The terminal content is as follows:

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell https://aka.ms/pscore6

PS C:\Users\mikryuko> wsl
```

<https://docs.microsoft.com/en-us/windows/wsl/install>

Command line

cd

~

cd /data

ls

ls

-l

ls

-l

-S

Command line

```
cd /data/QC
```

```
cat QC.md5
```

```
md5sum -c QC.md5
```

```
md5sum *.fq.gz > Checksum.md5
```

```
md5sum MiSeq_R{1,2}.fq.gz
```


Command line

```
less QC.md5
```

```
grep "MiSeq" QC.md5
```

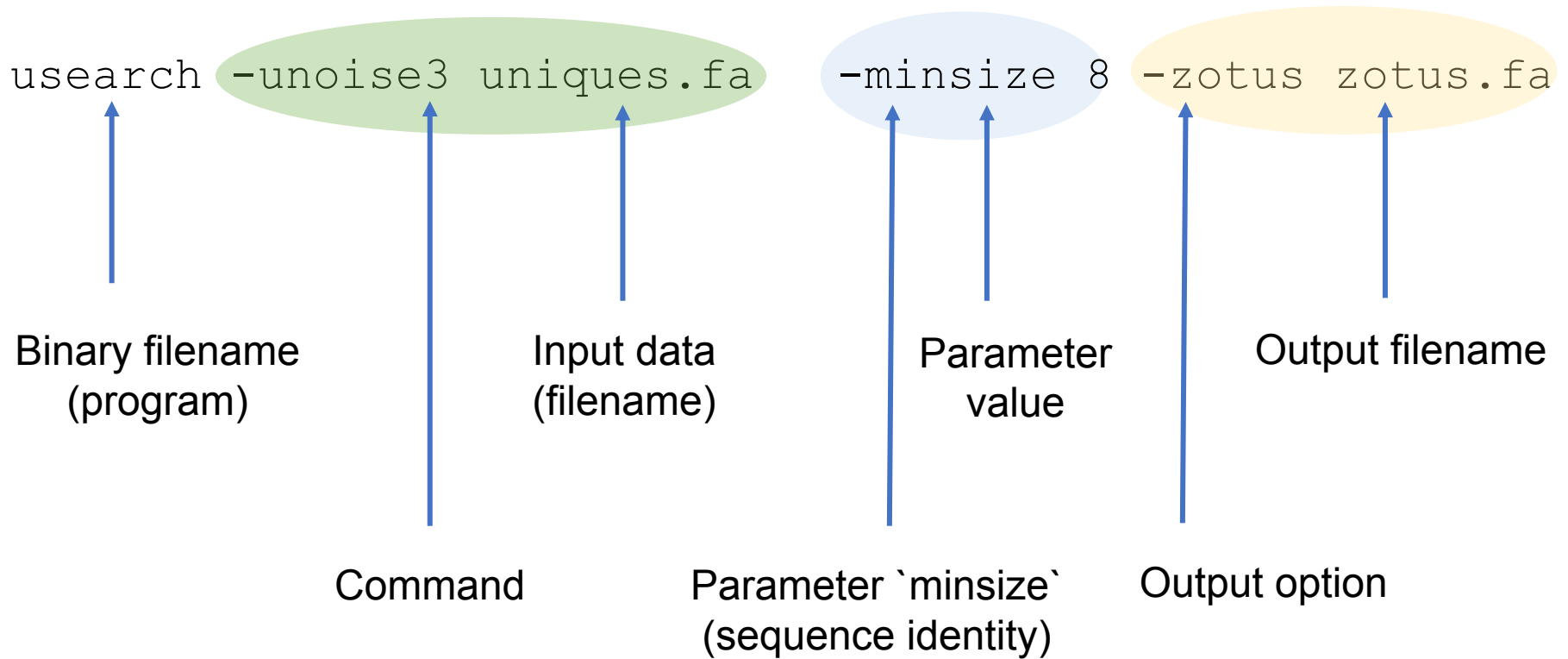
```
grep "MiSeq" QC.md5 | wc -l
```

Command line

```
zgrep "GTGAATCATCGAATCTTTG" \
  --color=always \
  "MiSeq_R1.fq.gz" \
  | less -R
```

```
zgrep "GTGAATCATCGAATCTTTG" \
  "MiSeq_R1.fq.gz" \
  | grep -v "TAAGCGGGCGGACT" \
  | sort | uniq \
  > some_sequences.txt
```

UNOISE



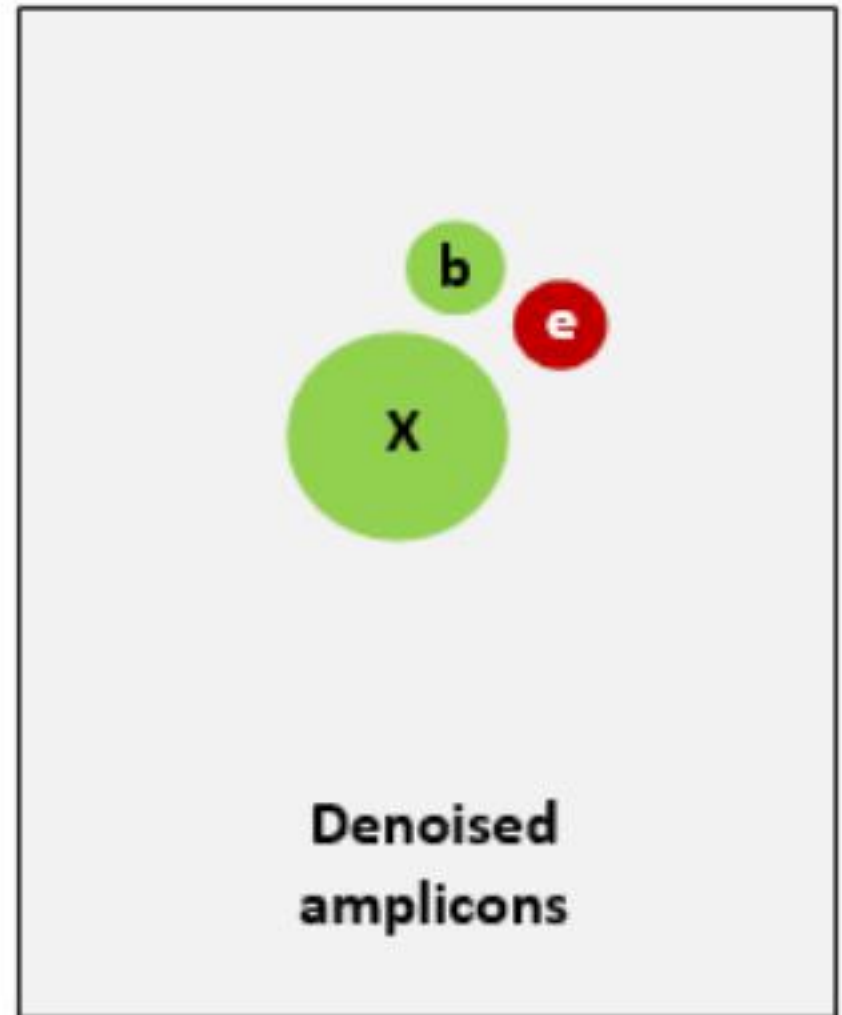
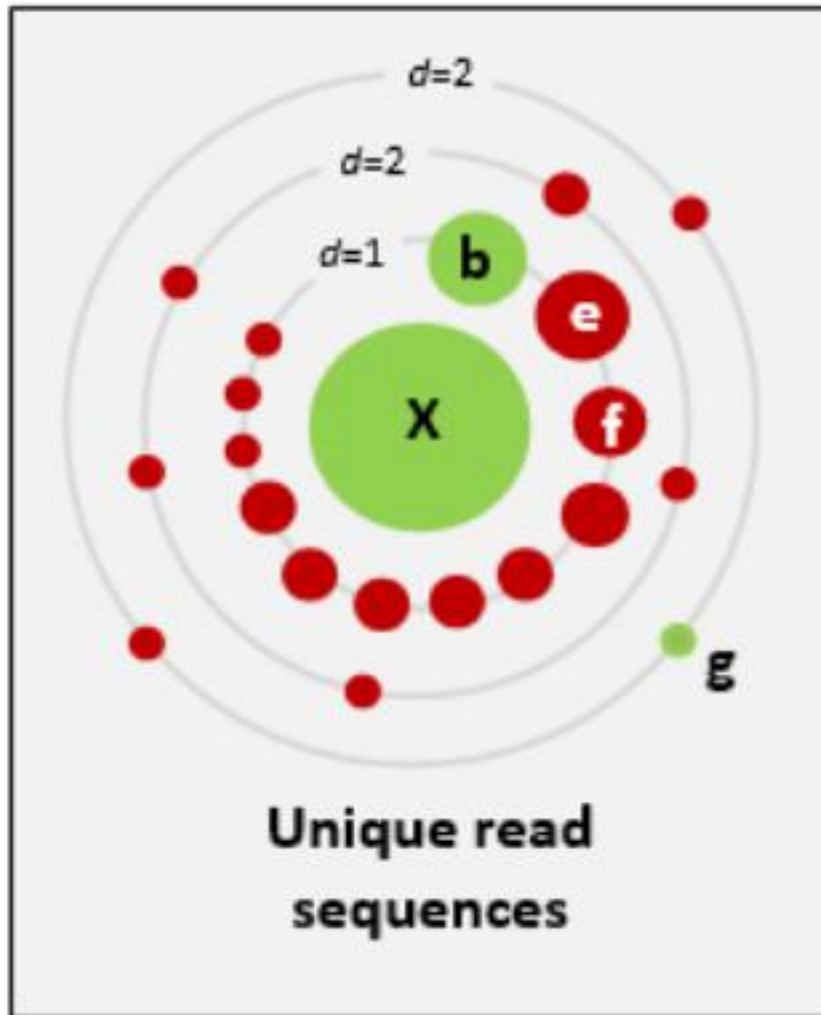
USEARCH

<https://drive5.com/usearch/>



Robert Edgar

UNOISE

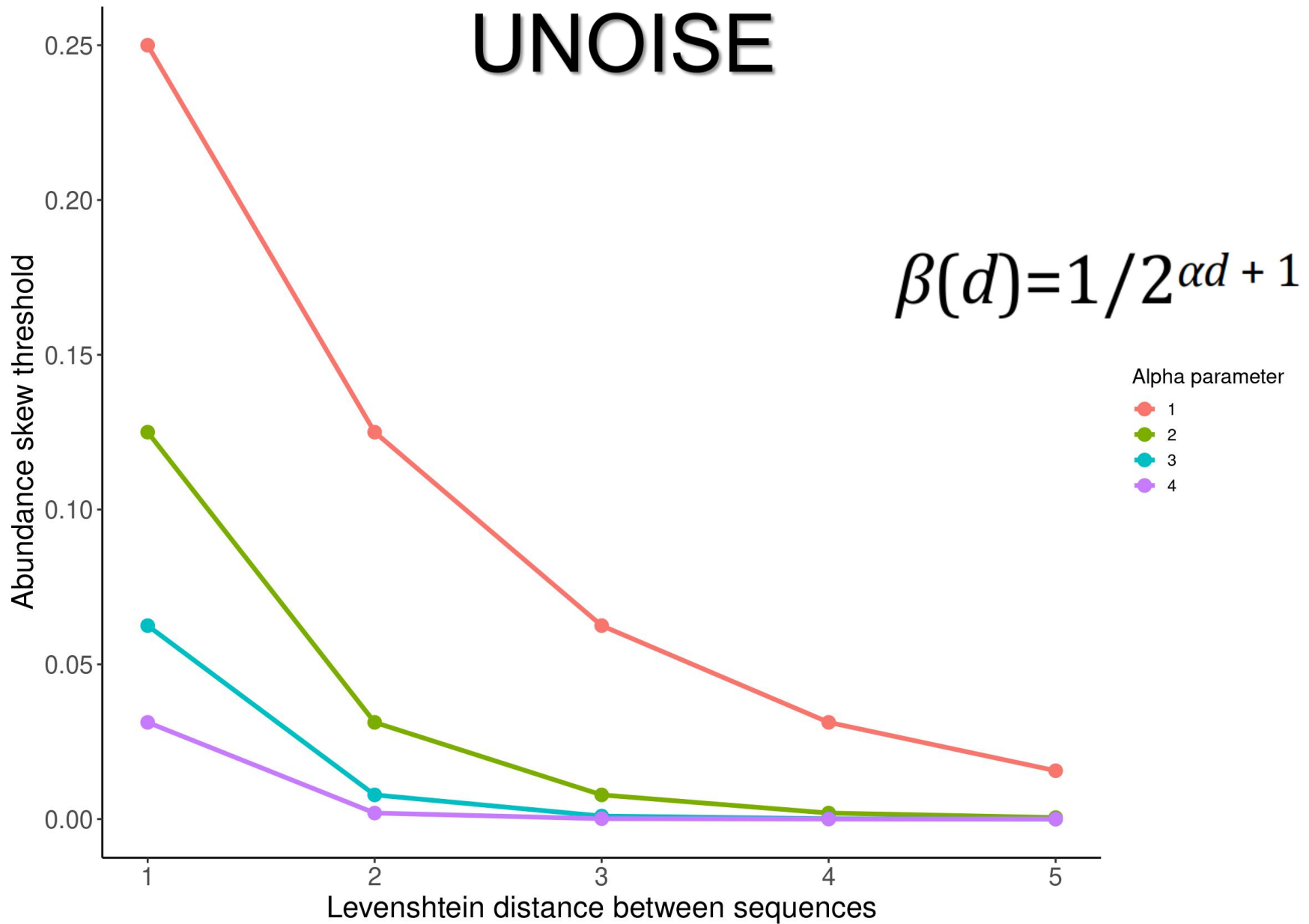


Edgar (2016) DOI:10.1101/081257

https://drive5.com/usearch/manual10/unoise_algo.html

UNOISE

$$\beta(d) = 1/2^{\alpha d + 1}$$



Metabarcoding: from Lab to Bioinformatics

Metabarcoding: from Lab to Bioinformatics (UT International Summer University, 2022)

Metabarcoding: from Lab to Bioinformatics

University of Tartu, 2022

~~Data used during the course~~

~~“Expert mode” commands~~

~~Individual projects~~

~~Slides (will be released after the course)~~

Docker images

Working docker images

Image	Software
ewels/multiqc:latest	mutliqc v1.12
staphb/fastqc:0.11.9	fastqc v0.11.9
pipecraft/cutadapt:3.5	cutadapt v3.5, seqkit v2.0.0, python3, biopython
pipecraft/dada2:1.20	dada2 v1.20, seqkit v2.0.0, lulu v0.1.0, R
pipecraft/reorient:1	fqgrep v0.4.4, seqkit v2.0.0, mothur v.1.43.0
pipecraft/trimmomatic:0.39	trimmomatic 0.39, seqkit v2.0.0
pipecraft/vsearch:2.18	vsearch v2.18, seqkit v2.0.0, GNU parallel
pipecraft/itsx:1.1.3	ITSx v1.1.3, seqkit v2.0.0, mothur v1.46.1
pipecraft/blast:2.12	BLAST v2.12.0+
pipecraft/deicode:0.2.4	DEICODE v0.2.4, qiime2-2002.2
pipecraft/fastp:0.23.2	fastp v0.23.2
pipecraft/blast:2.12	BLAST 2.12.0+, biopython, python3, gawk

https://pipecraft2-manual.readthedocs.io/en/latest/user_guide.html#working-docker-images

Run VSEARCH container

```
docker run -it  
-v /c/Metabarcoding:/data  
pipecraft/vsearch:2.18
```

UNOISE

I. Dereplicate sequences

```
cd /data
```

```
vsearch
```

```
--derep_fulllength MC_1_.fasta  
--output MC_1_dereplicated.fasta  
--fasta_width 0  
--relabel_sha1  
--sizeout
```

UNOISE

II. UNOISE clustering

vsearch

```
--cluster_unoise MC_1_dereplicated.fasta  
--strand both  
--minsize 4  
--unoise_alpha 2  
--fasta_width 0  
--sizein --sizeout  
--centroids MC_1_unoise.fasta
```

UNOISE

Count sequence numbers

```
grep -c "^>" MC_1_dereplicated.fasta
```

```
grep -c "^>" MC_1_unoise.fasta
```

UNOISE

Loop through all samples

```
for i in  
chimera_Filtered_out/*.fasta; do  
  
    echo ${i}  
  
done
```

UNOISE

Loop through all samples

```
vsearch
```

```
--derep_fulllength ${i}
```

```
--output -
```

```
| vsearch
```

```
--cluster_unoise -
```

```
--centroids $(basename ${i})
```