An overview of the bioinformatic pipelines for metabarcoding data analyses

Ali Hakimzadeh*, Sten Anslan*
*Institute of Ecology and Earth Sciences, University of Tartu, Estonia

bioinfo94
alihkz94
ali.hakimzadeh@ut.ee

SUMMIT 2022

Overview of the **22 pipelines** for metabarcoding data analysis

Since metabarcoding approach has become widely applied in biodiversity research, the **pipelines development has bloomed**
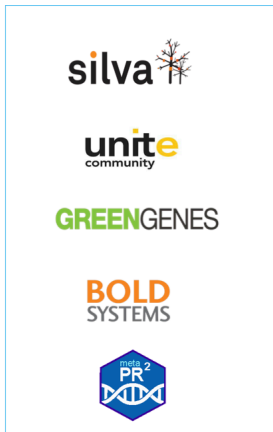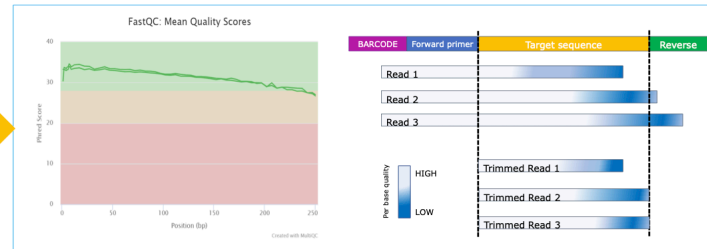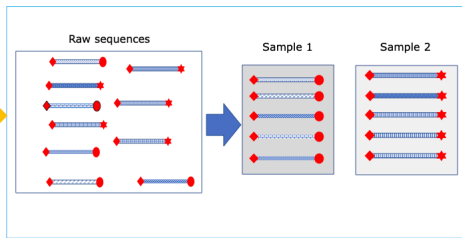
The plethora of pipelines:
Enables to select the **most appropriate one** or makes the selection even more **challenging?**

Classic workflow

Raw fastq reads
- Single-end long reads
- Paired-end short reads

Demultiplexing
Raw sequences — Sample 1 — Sample 2

QC, primer trimming, merging, quality filtering
FastQC: Mean Quality Scores

BARCODE | Forward primer | Target sequence | Reverse
Read 1, Read 2, Read 3
Trimmed Read 1, Trimmed Read 2, Trimmed Read 3
HIGH / LOW

OTUs/ASVs — **Sequences**
Clustering / Denoising
Sequencing errors

Clustering & denoising

OTU / ASV tables

OTU_id	sample1	sample2	sample3
OTU001	23	6578	0
OTU002	0	572	0
OTU003	0	8755	0
OTU004	4234	44567	0
OTU005	0	0	764
OTU006	56	2365	4532
OTU007	0	0	11235
OTU008	0	0	653
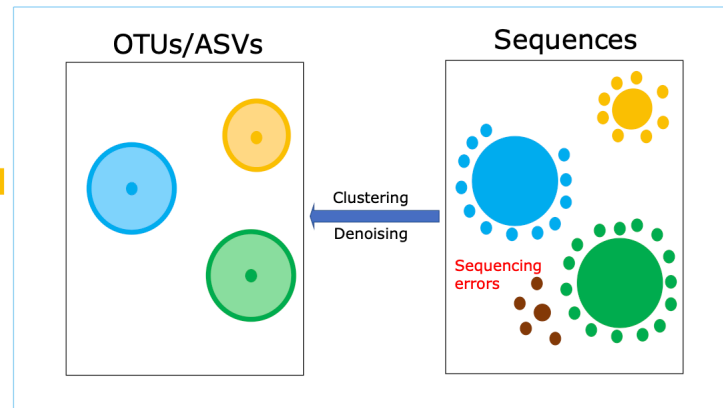
Taxonomy assignment
silva, unite community, GREENGENES, BOLD SYSTEMS, meta PR²

The framework of pipelines and compatibility with the different operating systems

| ALL STEPS IN ONE GO | STEP BY STEP EXECUTION |

AMPLISEQ, **CASCABEL**, **DADA2**, **FROGS**, **LOTUS2**, **PEMA**, **PIPECRAFT2**

ANACAPA, **APSCALE**, **COMA**, **DANIEL**, **EZMAP**, **JAMP**, **OBITOOLS**, **QIIME2**, **SEED2**, **GDAT**, **BIOCOM**

PIPTS, **SLIM**, **METAWORKS**, **BARQUE**

- 🟥 Linux / macOS / Windows
- 🟦 Linux / macOS
- ⬛ Linux
- 🟩 Linux / Web-based
- 🟪 Linux / macOS / Web-based
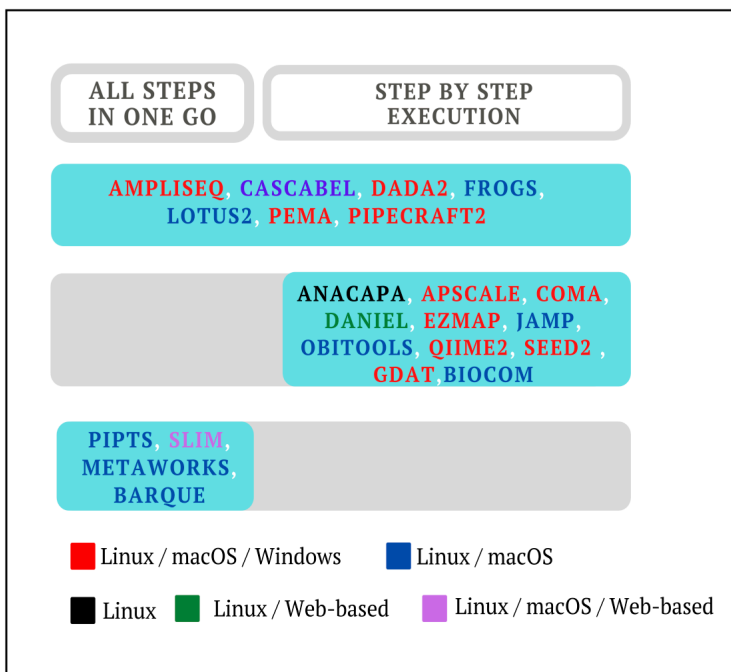
Differences between the pipelines are primarily due to the different clustering/denoising and quality trimming tools implemented (running time not considered here)

Nevertheless, pipelines designed for the same task produce similar results

The **pipeline choice** depends on the marker gene, the used sequencing platform, the operating system available for data analyses, and user preferences to use GUI (graphical user interface) or CLI (command line interface).

Pipelines ' similarity based on the implemented tools

Designed for
- 🔴 Fungi
- 🔵 Other eukaryotes
- 🟠 Prokaryotes
- ⚫ Universal

Output molecular unit
- ● ASV
- ▲ OTU
- ■ OTU/ASV