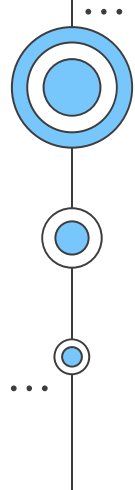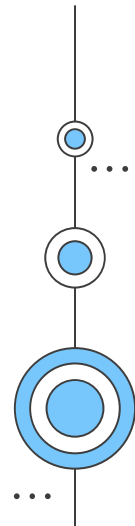# Customization of Bioinformatics Workflows for (Meta)genomics.

Ricardo Oliveira; Msc Student.
Pedro Santos; Supervisor.

01
Background

Background

Importance of identifying and characterizing microorganisms in microbiology.

Shift from traditional physical and biochemical methods to DNA-based approaches.

Advances in high-throughput sequencing technologies and their impact on metagenomics.

Explosion in bioinformatics tools for metabarcoding analysis.

Key software: QIIME, USEARCH, DADA2, DECIPHER, RDP classifier, PHYLOSEQ.

02
AIMS

Aims

01 **Evaluate the performance of metabarcoding**
Analyzing tools using different sequencing platforms. (Illumina, PacBio, Oxford Nanopore).

02 **Asses each step**
Using the following tools: DADA2, USEARCH/VSEARCH.

03 Develop customizable scripts for each analysis phase.

04 Apply scripts to a real-world case study.
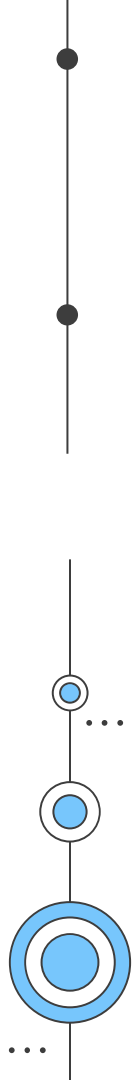
05 **Address complexity**
In selecting suitable algorithms and ensure robustness through benchmark studies.

03
Project Workflow

Project Workflow

Mock communities:
Zymo Research (D6300, D6305), ATCC.

Classification algorithms:
RDP classifier SINTAX.

Development of customizable scripts.

1   2   3   4   5

Performance analysis tools:
DADA2, USEARCH/VSEARCH.

Metrics:
Taxa identification & quantify proportions.
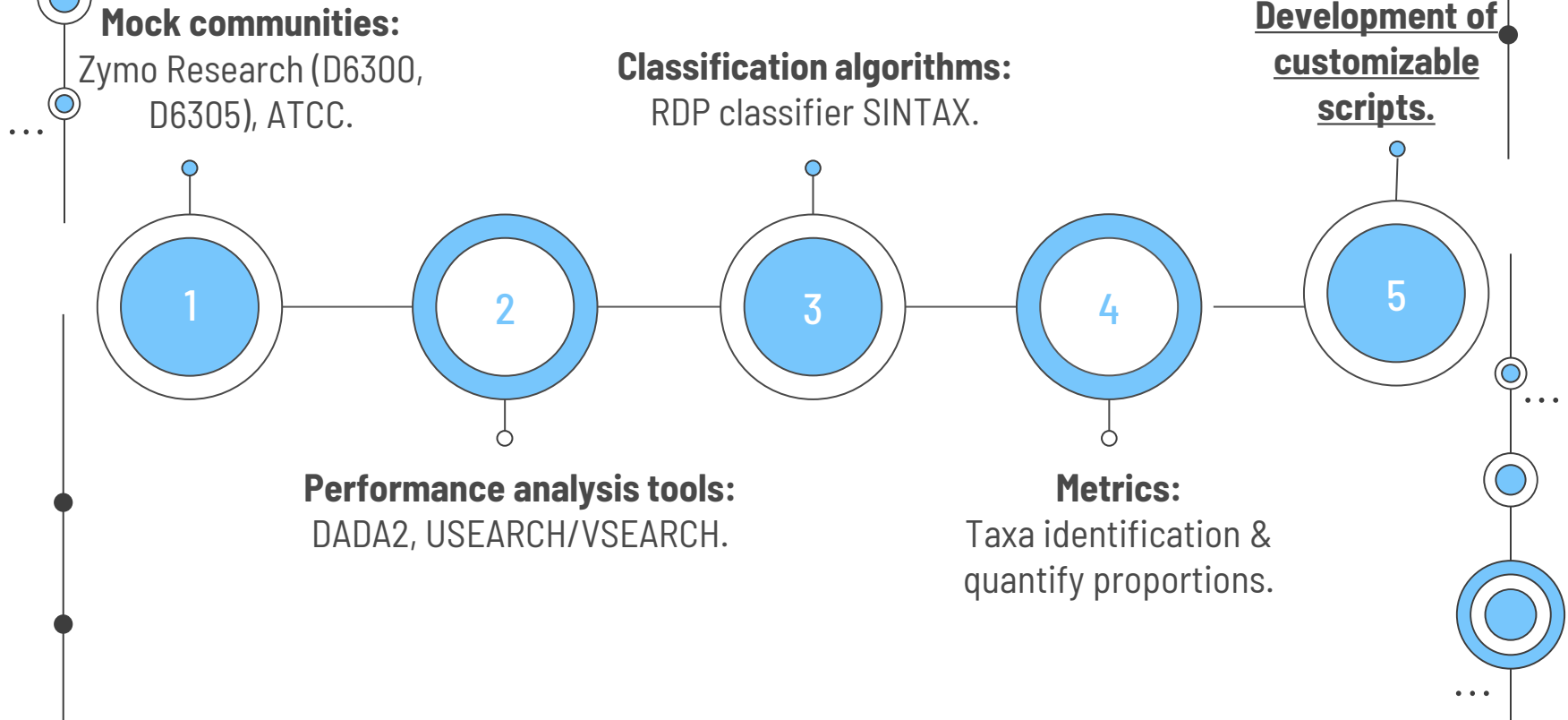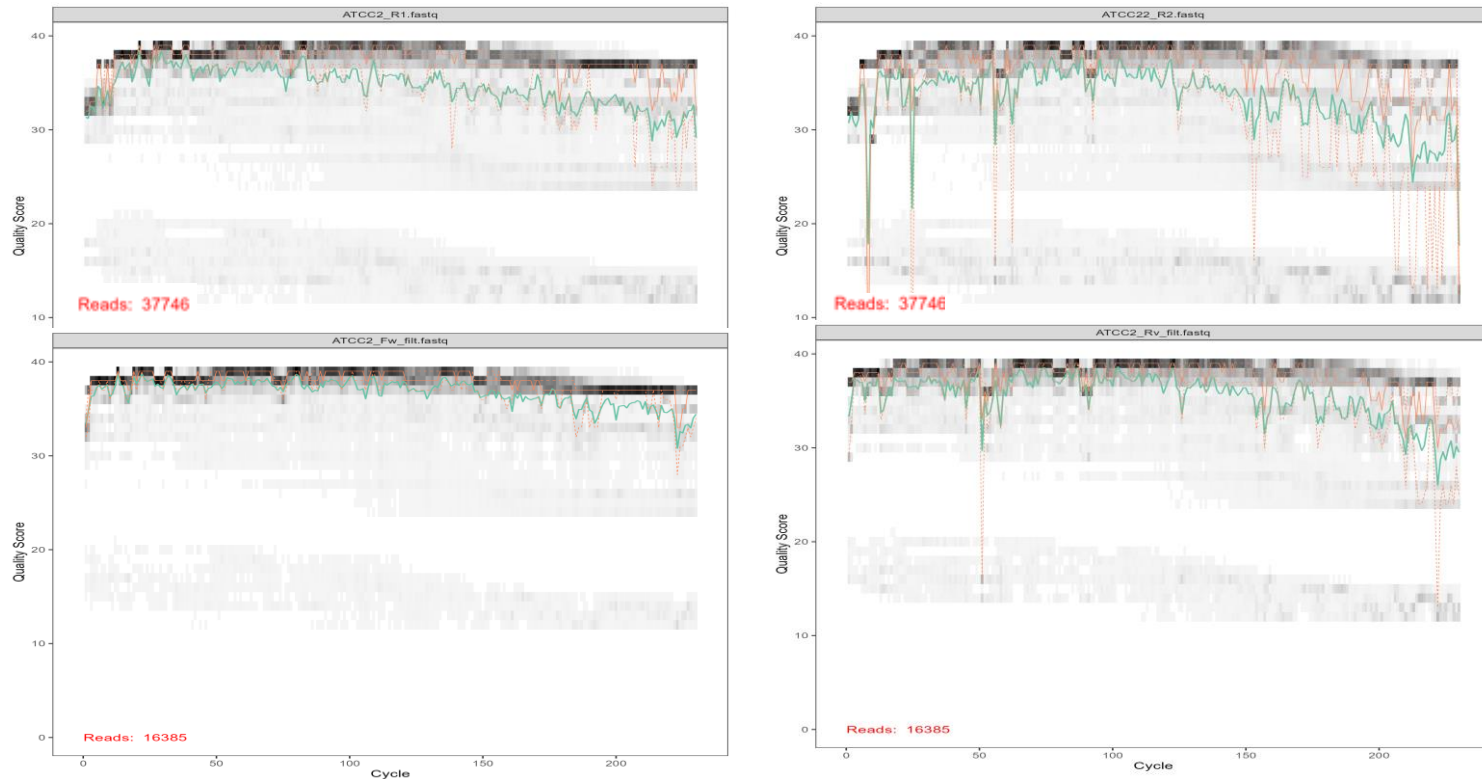
04
Results

Quality control

Fig 1: Reads filtration for the sample ATCC2 (Top: unfiltered; Bottom: filtered)

ASV formation

	ATCC1	ATCC2	ATCC3	Zymo1	Zymo2
ASV1	2582	922	9962	19675	27727
ASV2	0	0	0	17079	30098
ASV3	2640	1444	2388	22623	16044
ASV4	1556	663	4027	8864	23051
ASV5	0	0	0	22941	12222
ASV6	704	387	1341	9923	22553
ASV7	0	0	0	15156	17897
ASV8	0	0	0	16951	8369
ASV9	3633	1629	15847	0	0
ASV10	3260	1436	13705	0	0
ASV11	2908	1142	12588	0	0
ASV12	3132	1545	8396	0	0
ASV13	1851	702	9223	0	0
ASV14	0	0	0	2729	8617
ASV15	2483	1416	4783	0	70
ASV16	1325	545	6799	0	0
ASV17	2102	996	5074	0	0
ASV18	614	279	5460	0	0
ASV19	1485	654	3324	0	0
ASV20	1512	684	2975	0	0
ASV21	1151	503	2147	0	0
ASV22	847	343	2588	0	0
ASV23	573	283	2003	0	0
ASV24	444	170	1003	0	0
ASV25	228	117	1174	0	0
ASV26	333	172	922	0	0
ASV27	142	108	192	0	0
ASV28	135	0	0	0	0
ASV29	0	9	58	0	0
ASV30	50	0	0	0	0
ASV31	0	0	0	0	48
ASV32	0	0	0	0	37
ASV33	0	0	0	34	0

```
>ASV1
GCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCAGATGT
GAAATCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGG
GGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATAACCGGTGGCGAA
GGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGG
>ASV2
GCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTCTGTCAAGTCGGATGT
GAAATCCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTTGTAGAGG
GGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATAACCGGTGGCGAA
GGCGGCCCCCTGGACAAAGACTGACGCTCAGGTGCGAAAGCGTGGGG
>ASV3
GCAAGCGTTATCCGGAATTATTGGGCGTAAAGCGCGCGTAGGCGGTTTTTTAAGTCTGATGT
GAAAGCCCACGGCTCAACCGTGGAGGGTCATTGGAAACTGGAAAACTTGAGTGCAGAAGAGG
AAAGTGGAATTCCATGTGTAGCGGTGAAATGCGCAGAGATATGGAGGAACACCAGTGGCGAA
GGCGACTTTCTGGTCTGTAACTGACGCTGATGTGCGAAAGCGTGGGG
>ASV4
GCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTCAGCAAGTTGGATGT
GAAATCCCCGGGCTCAACCTGGGAACTGCATCCAAAACTACTGAGCTAGAGTACGGTAGAGG
GTGGTGGAATTCCTGTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCAGTGGCGAA
GGCGACCACCTGGACTGATACTGACACTGAGGTGCGAAAGCGTGGGG
>ASV5
GCAAGCGTTGTCCGGAATTATTGGGCGTAAAGGGCTCGCAGGCGGTTCCTTAAGTCTGATGT
GAAAGCCCCCGGCTCAACCGGGGAGGGTCATTGGAAACTGGGGAACTTGAGTGCAGAAGAGG
AGAGTGGAATTCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAACACCAGTGGCGAA
GGCGACTCTCTGGTCTGTAACTGACGCTGAGGAGCGAAAGCGTGGGG
>ASV6
GCAAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAGCGCAGGCGGTTTCTTAAGTCTGATGT
GAAAGCCCCCGGCTCAACCGGGGAGGGTCATTGGAAACTGGGAGACTTGAGTGCAGAAGAGG
AGAGTGGAATTCCATGTGTAGCGGTGAAATGCGTAGATATAGGAGGAACACCAGTGGCGAA
GGCGGCTCTCTGGTCTGTAACTGACGCTGAGGTGCGAAAGCGTGGGG
>ASV7
GCAAGCGTTATCCGGATTTATTGGGCGTAAAGAGAGTGCAGGCGGTTTTCTAAGTCTGATGT
GAAAGCCTTCGGCTTAACCGGAGAAGTGCATCGGGAAACTGGATAACTTGAGTGCAGAAGAGG
GTAGTGGAACTCCATGTGTAGCGGTGGAATGCGTAGATATATGGAAGAACACCAGTGGCGAA
GGCGGCTACCTGGTCTGCAACTGACGCTGAGACTCGAAAGCATGGGT
>ASV8
GCAAGCGTTGTCCGGATTTATTGGGCGTAAAGCGCGCGCAGGCGGTCTTTTAAGTCTGATGT
GAAAGCCCCCGGCTTAACCGGGGAGGGTCATTGGAAACTGGAAGACTGGAGTGCAGAAGAGG
AGAGTGGAATTCCACGTGTAGCGGTGAAATGCGTAGATATGTGGAGGAACACCAGTGGCGAA
GGCGACTCTCTGGTCTGTAACTGACGCTGAGGCGCGAAAGCGTGGGG
```

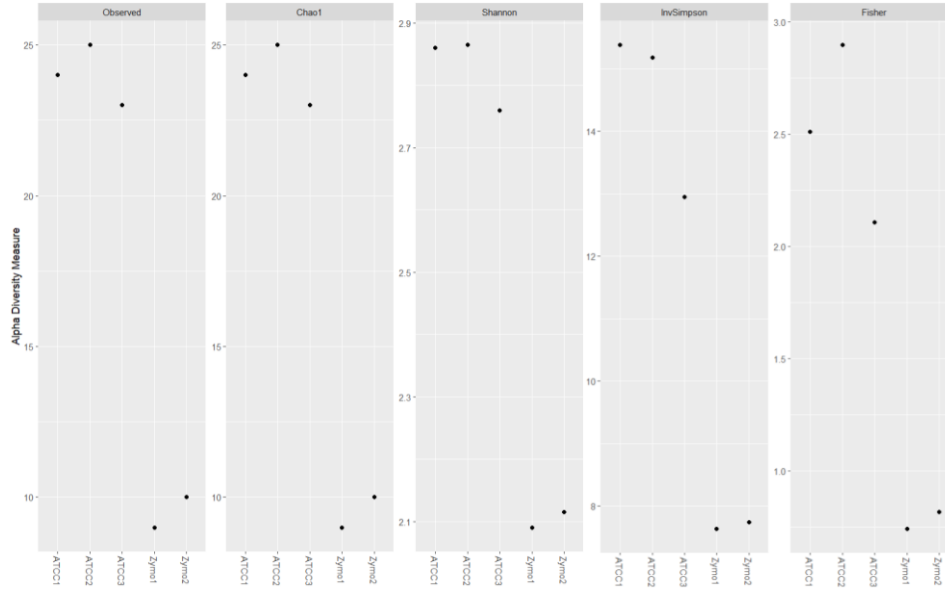Fig 2: Results of the ASV formation using DADA2

Phyloseq analysis

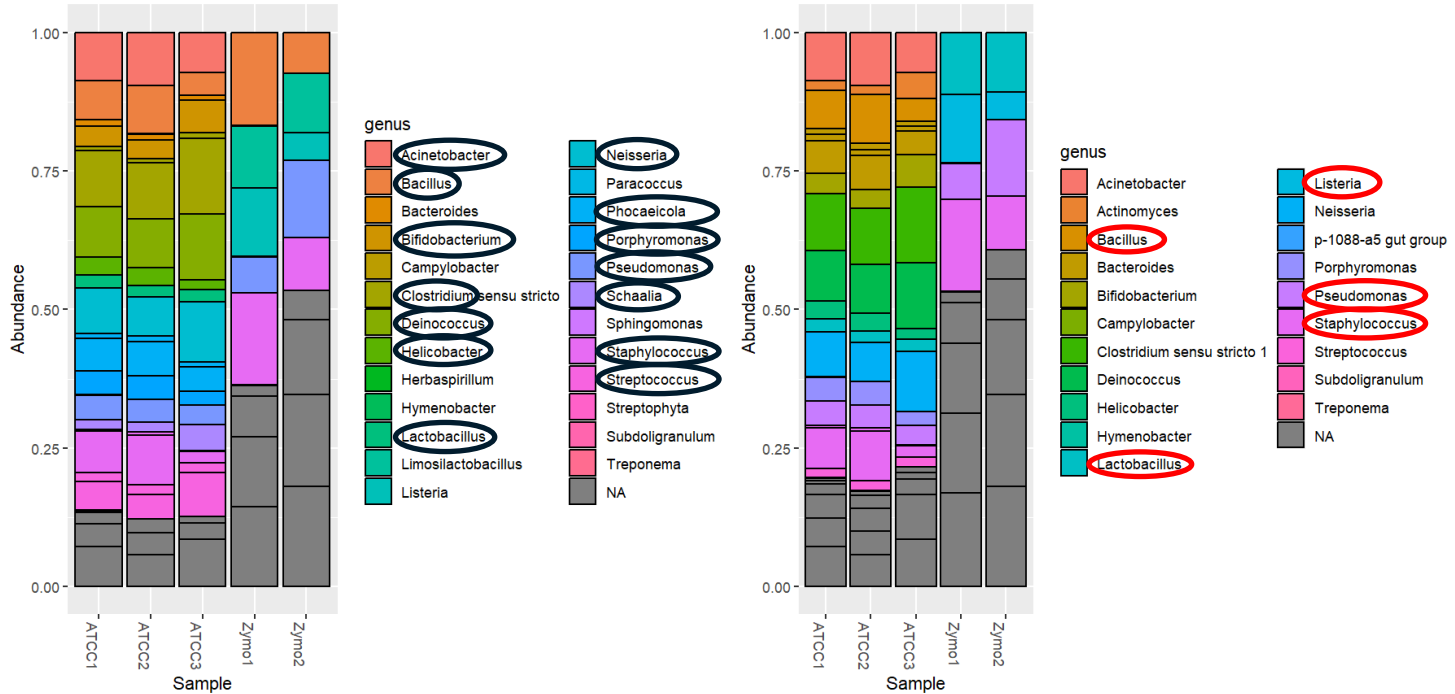Fig 3: Biodiversity analysis

Phyloseq analysis

Fig 4: Genus classification (left: RDP; right: SILVA; blue: ATCC; red: Zymo)

05
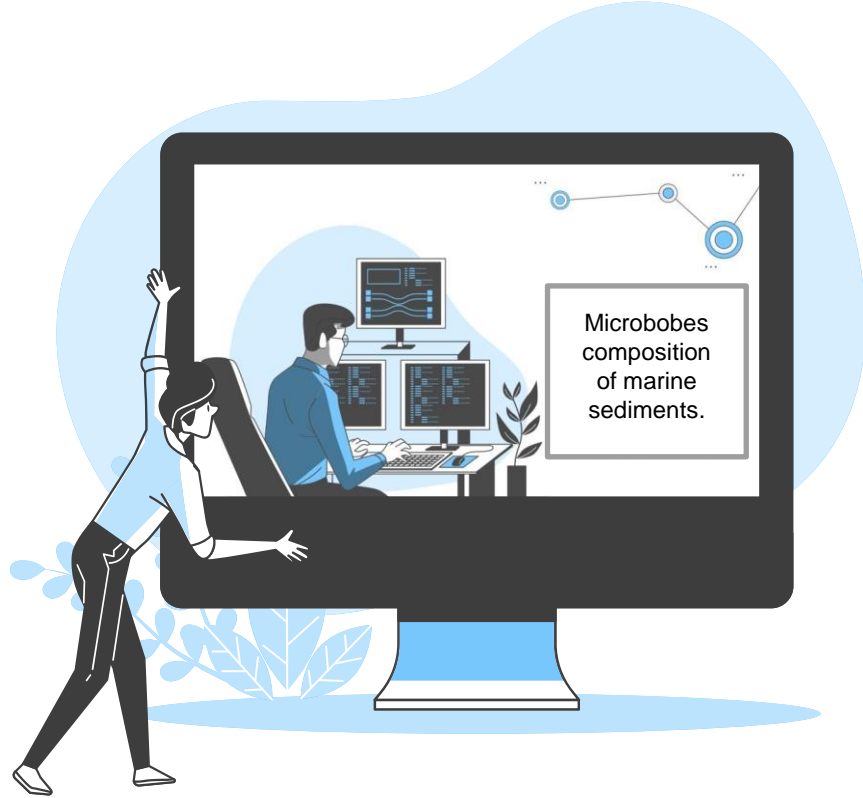## Upcoming Goals

Upcoming Goals

Application of the develop scripts to real case study.

Expansion of the software used, and the sequencing technologies supported. Allow a personalized analysis

Testing & validation of the develop workflow.

Microbobes composition of marine sediments.

Thanks!