

RecyclesEU

recycleseu

Recycles EU

RECYCLES WORKSHOP

Metagenomics and
metabarcoding approaches to
describe ecological systems
and infer their development

5th, 6th & 7th of July 2022

Practical on metabarcoding analysis

Francesco Vitali

*Research Centre for Agriculture and Environment, Council for
Agricultural Research and Economics (CREA-AA), Florence*



GA: 872053 — H2020 - MSCA - RISE-2019

Outline



1. Introduction: different possible approaches, tools, and definitions.
2. The fundamental elements of microbial ecological analysis: metadata, otu table, taxonomy table. Importance of rich metadata annotation.
3. Setting up of working environment.
4. Sequence manipulation: pre-treatment.
5. Sequence analysis: obtaining otu table and taxonomy table (QIIME2, all R approach eventually).

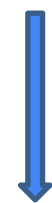
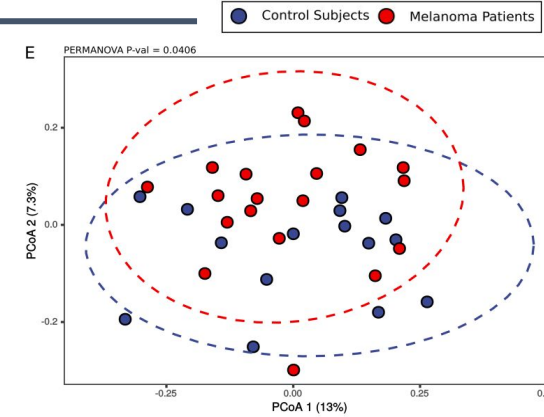
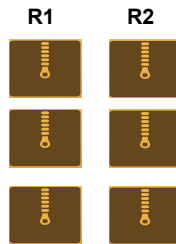
Outline



Tuesday



Sample1 →
Sample2 →
SampleN →



Wednesday

Real application scenarios
among research experience of
the RECYCLES partners

Outline

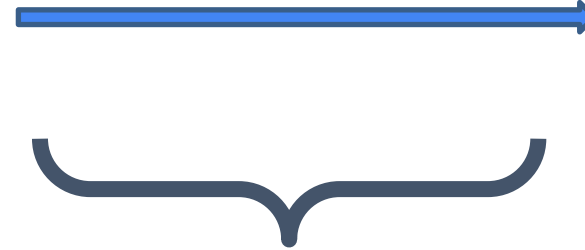


Tuesday

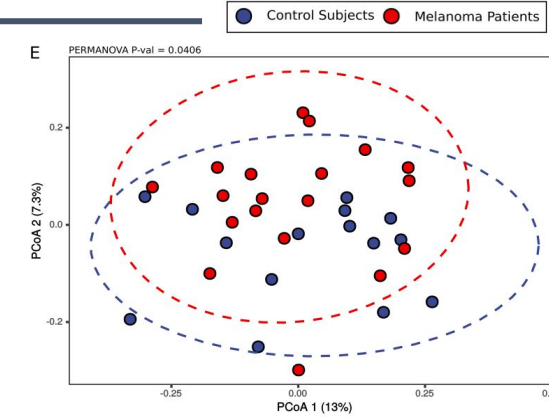


Sample1 →
Sample2 →
SampleN →

R1	R2



Today



Wednesday

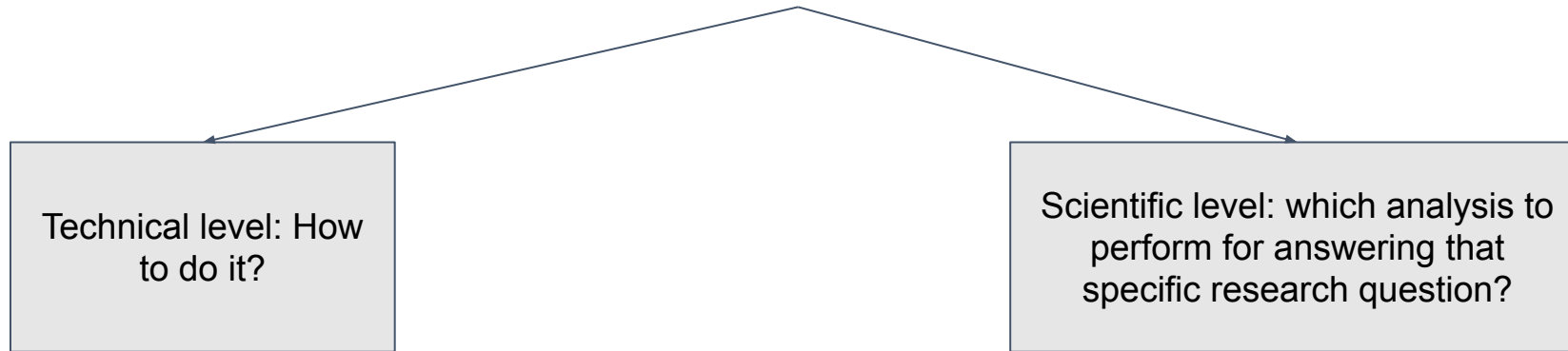
https://drive.google.com/drive/folders/1lhGO-uHo5_HCp4VRfxFny2Du-d34qikg?usp=sharing DATA and CODE is here, all should have access

Real application scenarios
among research experience of
the RECYCLES partners

Outline



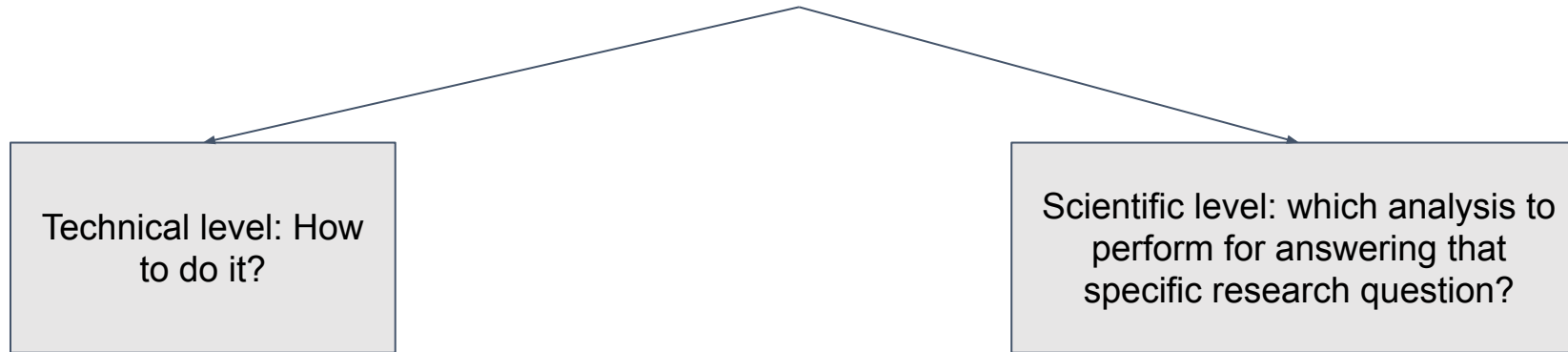
PERSONAL AIM: enable you to do an analysis on your data from Monday morning.



Outline



PERSONAL AIM: enable you to do an analysis on your data from Monday morning.



But first, do you have any experience with those kind of data?
Did any of you actually produced the data for own research?
External service or in-house sequencers?
What do you usually do?
What is you field of research/work?



PERSONAL AIM: enable you to do an analysis on your data from Monday morning.

Technical level: How to do it?

A lot of different approaches are possible. If time permits we will try to practically go over 2 of them among the most popular in my opinion: QIIME2 (hands-on) and R (demonstrated)

- We will try and set up your notebook for an analysis with QIIME2, if we see that time is running out, we will change to a full demonstration on my PC.
 - Do you all have internet connection at the moment? If not, we need one



PERSONAL AIM: enable you to do an analysis on your data from Monday morning.

Technical level: How to do it?

- Been thinking a lot about the best structure of this workshop for you.
 - The use of QIIME2 in this context seems the best option for beginners in this field (at least, it was for me with QIIME1). It is “easier” than the R approach (thanks to workflows command), even if less powerful and customizable. QIIME2 has a wide and very active community of users and developers, which prepared very detailed tutorials, and which are available for answering questions (see their forum [QIIME 2 Forum](#) here you can post your question or, even more important, find questions made by others that are already answered). Without doubt this is in my opinion the real added value of QIIME, the community they were able to build around the tool.
 - R would be my choice at this point, but is not friendly. I’ll point you to a wonderful tutorial for a full R workflow.

Possibilities, tools, definition, and gold resources



Gold source of information:

	Site	URL	Topics
1	The QIIME forum	https://forum.qiime2.org/	All topics QIIME related, workflow and suggestion, lots of guide, microbial ecology methods
2	Happy Belly Bioinformatic	https://astrobiomike.github.io/	BASH, R, QIIME, complete pipeline example
3	Video by Patrick Schloss	https://www.youtube.com/c/RiffomonasProject	Mainly R (and MOTHUR), microbial ecology methods
4	Stack Overflow	https://stackoverflow.com/	If you search how to do something with R, you end here 99% of the time
5	Gusta Me	https://mb3is.megx.net/gustame	Multivariate methods for microbial ecology
6	Biostars	https://www.biostars.org/	General bioinformatics Q/A

Possibilities, tools, definition, and gold resources



Different possible approaches:

1. QIIME2
 - a. Dada2
 - b. Deblur
 - c. Open-Reference OTUs
2. R
 - a. Dada2 <https://bioconductor.org/packages/release/bioc/html/dada2.html>
3. MOTHUR (own method, OTU clustering based and ASV) <https://mothur.org/>
4. VSEARCH (implements UNOISE/USEARCH) <https://github.com/torognes/vsearch>
5. MICCA (implements UNOISE or clustering based methods) <https://micca.readthedocs.io/en/latest/index.html>

Tools installation: where are we?



OS in use		WSL	
Windows	IOS	yes	no
27	7	3	31

	Participants
TOTAL	35

- Most of you work on Windows.
- None of you is using a tablet or a chromebook
- Hardware side, all of your machine should be able to perform analysis on targeted metagenomics (we will use a very reduced dataset). No information of RAM for some of you, but should all be around or over 8GB.
 - RAM here is maybe among the main limiting factors, as could cause crashed/freezes (use of paging)
- None of you seems to use alternatives to Excel (but may know of some names); most of you know of .csv or .tsv files

BASH			R		
yes	blank	no	yes	blank	no
9	13	18	16	13	11

- Not many know of BASH, but around 50% know R, both of them are really useful for the analysis in general

Tools installation: where are we?



TSV/CSV file format for spreadsheets; always better than xsl for data analysis...

Basically a plain text file, with a specific character that tells a program where each cell starts and finishes

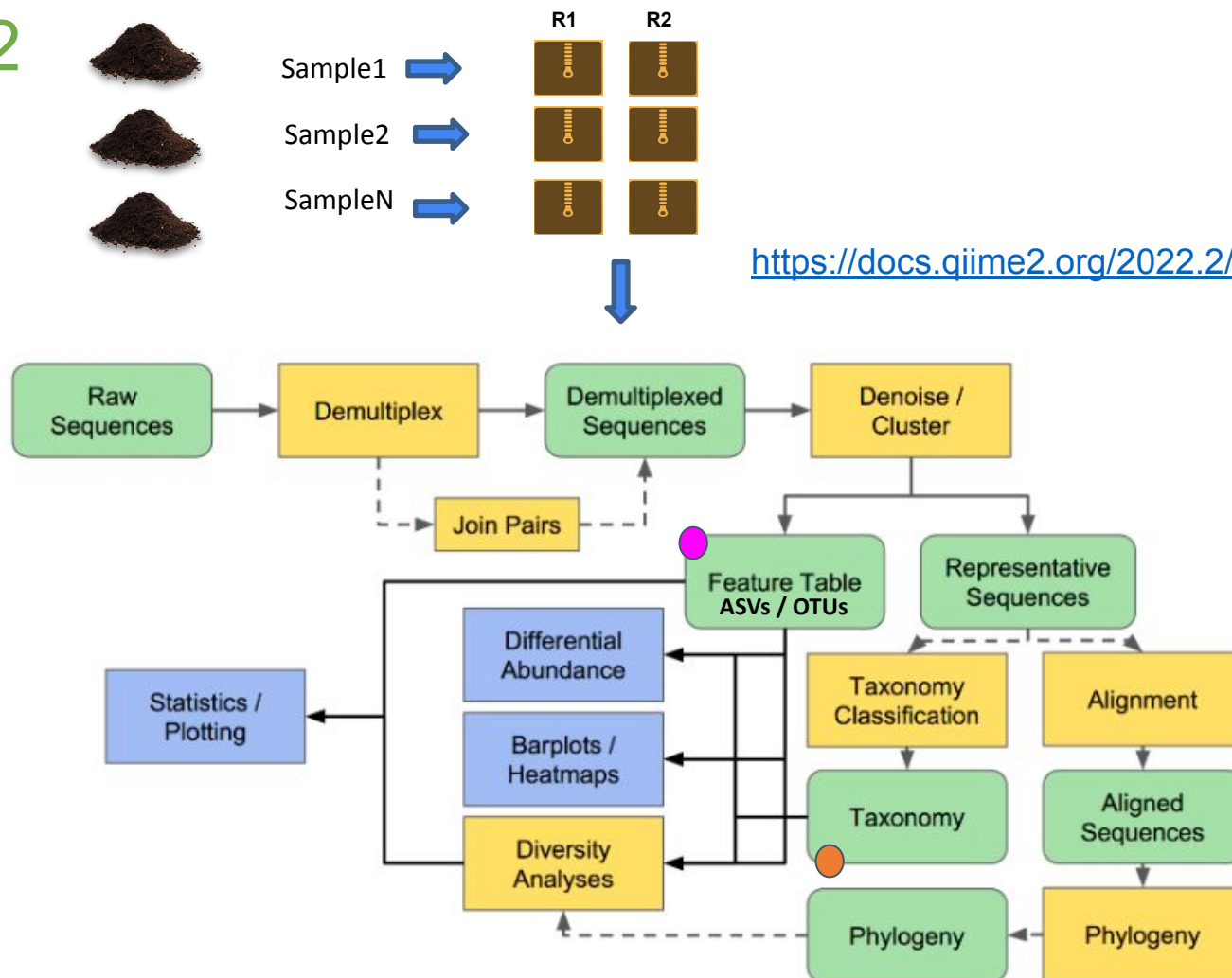
PRO: any software and any operating system will read the file in the same way

CONS: it is plain text (no color, no formatting, no joined cells, no formulas)

Tools installation: QIIME2 pipeline



QIIME2



<https://docs.qiime2.org/2022.2/tutorials/overview/>

	S1	S2	Sn
OTU1	500	350	
OTUn			

Count of occurrence of each OTU in all samples

	K	P	...
OTU1	Bacteria	Proteobacteria	
OTUn			

Taxonomic hierarchical classification of each OTU

The user define and provide the third element, the METADATA table



QIIME2

Please, open those links while we talk!

- QIIME2 installation instruction [Installing QIIME 2 — QIIME 2 2022.2.0 documentation](#)

Installing QIIME 2

QIIME 2 can be installed natively or using virtual machines. The following pages describe how to install the [QIIME 2 Core 2022.2 distribution](#) in either scenario.

- [Natively installing QIIME 2](#)

- Miniconda
- Install QIIME 2 within a **conda** environment
- Activate the **conda** environment
- Test your installation
- Next steps
- [How do I update to the newest version of QIIME 2?](#)
- (Re-)Activating QIIME 2

We will use this method at the end...

- [Installing QIIME 2 using Virtual Machines](#)

- [Installing QIIME 2 using VirtualBox](#)
- [Installing QIIME 2 using Amazon Web Services](#)
- [Installing QIIME 2 using Docker](#)
- [Installing QIIME 2 using WSL](#)

This could be an alternative, but generally worst. It lowers your PC hardware (the VM is running as a host operating system inside and already running operating system, which is Windows). Setting up a VM and moving files from host to VM, set up the connections or USB and so on, is not easy.

...first we need to setup this



WSL

Please, open those links while we talk!

- WSL : I don't have direct experience, but it seems a very good solution to avoid setting up a dual boot system

WHAT IS IT? <https://docs.microsoft.com/en-us/windows/wsl/about>

The Windows Subsystem for Linux lets developers run a GNU/Linux environment -- including most command-line tools, utilities, and applications -- directly on Windows, unmodified, without the overhead of a traditional virtual machine or dualboot setup.

Apparently you can have a full LINUX inside a Windows system, without wasting resources (Virtual Machines) or without particular installations (dual-boot)

Can't really help you on this, as I have a machine with only Ubuntu...



WSL

Please, open those links while we talk!

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



MINICONDA

Please, open those links while we talk!

INSTALLATION INSTRUCTION <https://educ-ubc.github.io/conda.html> (not my tutorial, found it online)

MINICONDA is a light version of CONDA.
CONDA is a package manager and environment manager for python.

Lots of software we regularly use (QIIME2 included) are a big collection of Python/R scripts, often developed from external researcher respect to the QIIME team, which are collected in the same place in a coordinated way (i.e. a pipeline).
[this is a very, very, simple recap of what QIIME2 is on the core]

With CONDA/MINICONDA we download a self-contained environment which has all the required software pre-installed and pre-compiled (like an image of another computer in a sense) and install this on our local machine. In this way, the installation is easy and fast, and no conflict of software versioning arises with what is already installed on local machine.



QIIME2

Please, open those links while we talk!

INSTALLATION INSTRUCTION <https://educe-ubc.github.io/qiime2.html> (not my tutorial, found it online)

Example pipeline



QIIME2

Please, open those links while we talk!

See here for all the available official tutorial <https://docs.qiime2.org/2022.2/tutorials/>

We will follow the “Atacama soil microbiome” tutorial (<https://docs.qiime2.org/2022.2/tutorials/atacama-soils/>)

It uses paired-end sequencing data (2x150) which comes in as multiplexed file (i.e. 1 file for R1 and one file for R2 with all samples in it).

<https://view.qiime2.org/> to view artifacts

OPEN THE TERMINAL AND THE SCRIPT
atacama soil demo.sh

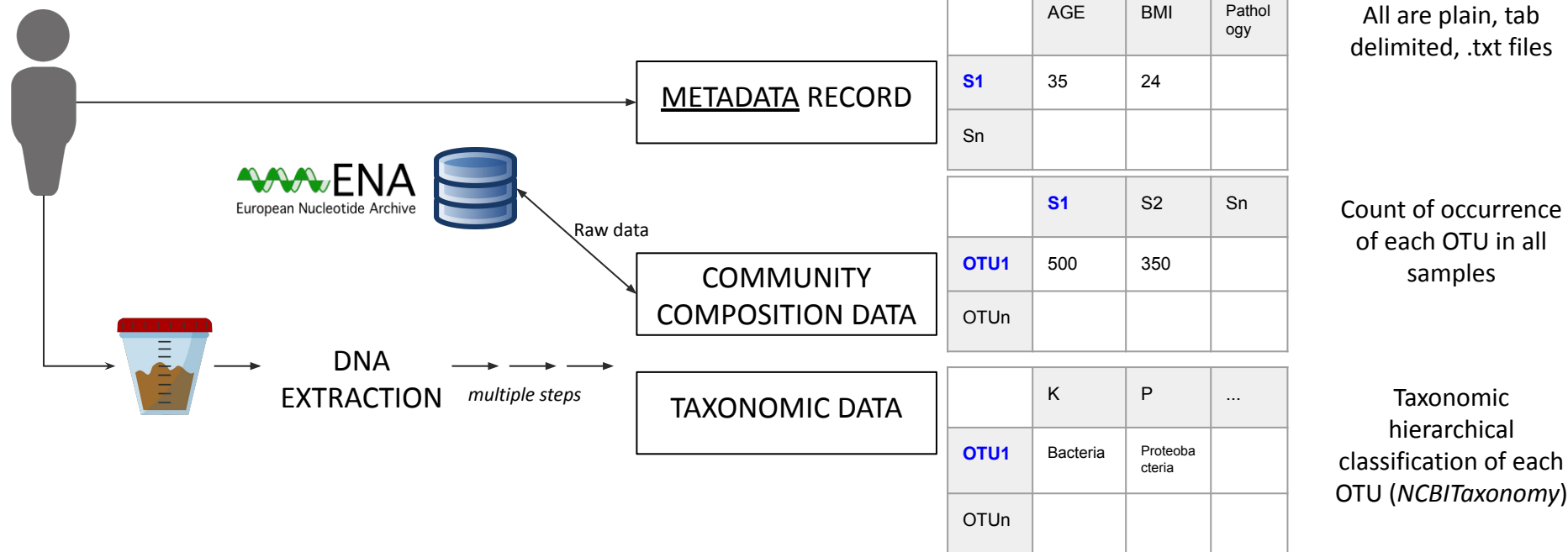
Consider this as a plain text file, and open with wordpad or notepad. Optimally, install a text editor capable of syntax highlighting, for example Sublime text, or Atom

Example pipeline



QIIME2

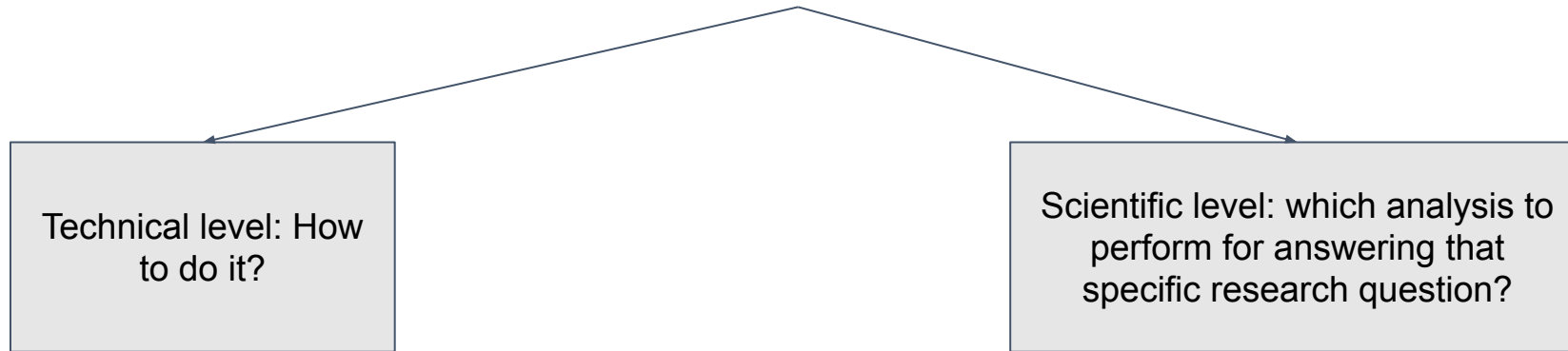
AFTER THE ATACAMA SOIL DEMO, WE SHOULD HAVE ALL THE THREE ELEMENTS OF MICROBIAL ECOLOGY ANALYSIS



Outline



PERSONAL AIM: enable you to do an analysis on your data from Monday morning.





PERSONAL AIM: enable you to do an analysis on your data from Monday morning.

Scientific level: which analysis to perform for answering that specific research question?

4. Ecological Insight: QIIME2



- QIIME2 also offer a complete set of tools for data analysis beyond the production of OTU/ASV table and taxonomy.

Refer to final steps (after DADA/DEBLUR of the “Moving Picture” tutorial

<https://docs.qiime2.org/2022.2/tutorials/moving-pictures/>

Data analysis and visualization are collected in workflow command

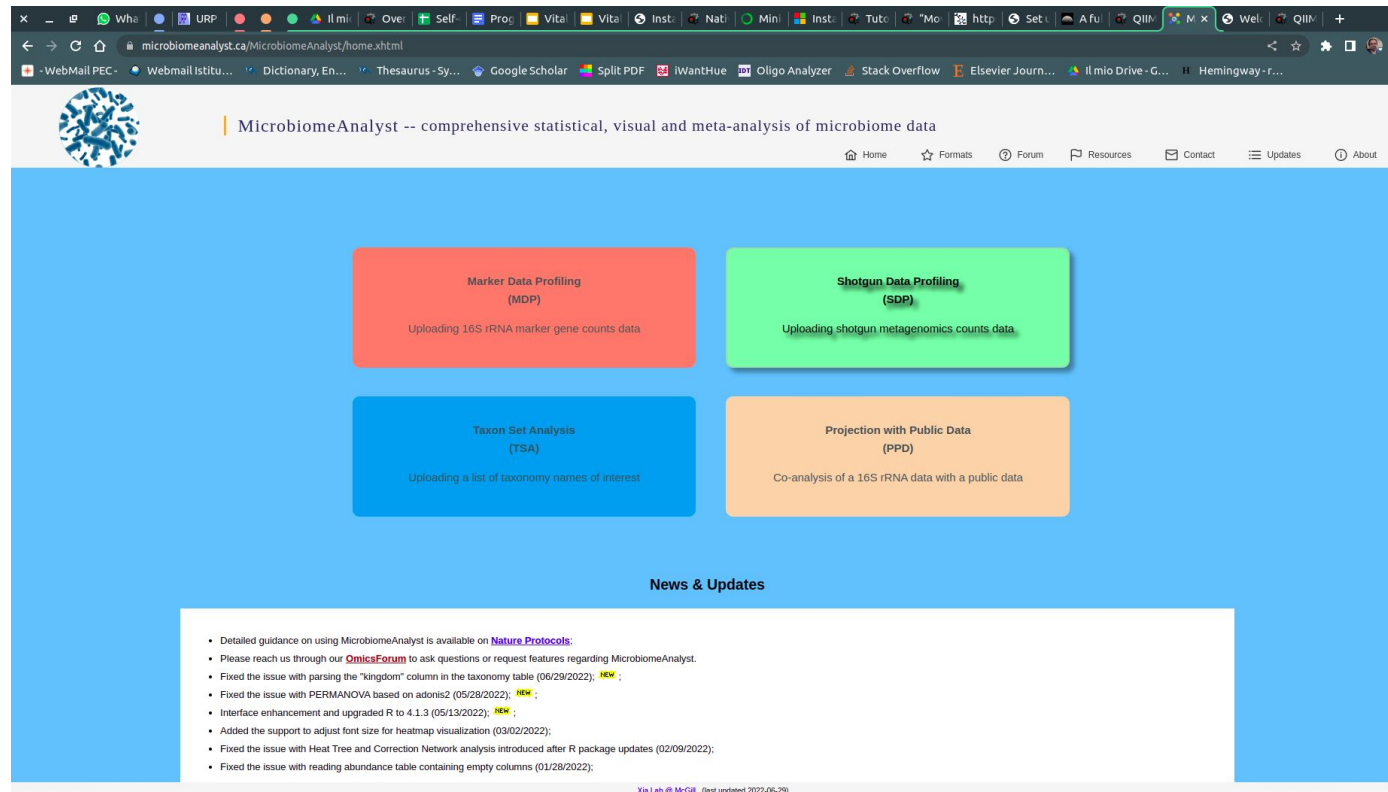
- qiime diversity core-metrics [calculates metrics and indices, plot beta diversity]
- qiime diversity alpha-group-significance [plot comparisons with metadata for alpha diversity]
- qiime diversity beta-group-significance [run PERMANOVA analysis]

LET'S TRY WITH THE ATACAMA RESULTS
OBTAINED

4. Ecological Insight: MicrobiomeAnalyst



- A good alternative for starting, if you do not already know R
- Useful also if you know R, as it release the code for the analysis



LET'S TRY WITH THE ATACAMA RESULTS
OBTAINED

<https://www.microbiomeanalyst.ca/MicrobiomeAnalyst/home.xhtml>

4. Ecological Insight: R



- R is maybe the most powerful and versatile solution. Some packages that will be greatly useful

Data analysis related and specific for the microbiome field

- Phyloseq (and Speedyseq)
- Vegan
- Microbiome
- Metagenomeseq
- dada2
- psych
- factoextra
- factomineR

Visualization

- ggpubr
- ggplot2
- ggsci
- microshades
- Pheatmap
- patchwork

Utilities for data manipulation

- tidyverse

LET'S TRY WITH THE ATACAMA RESULTS
OBTAINED



RecyclesEU

recycleseu

Recycles EU

RECYCLES WORKSHOP

Metagenomics and
metabarcoding approaches to
describe ecological systems
and infer their development

5th, 6th & 7th of July 2022



GA: 872053 — H2020 - MSCA - RISE-2019