RecyclesEU

recycleseu

Recycles EU

RECYCLES WORKSHOP

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

5th, 6th & 7th of July 2022

From samples to sequences - From sequences to ecological insight with examples

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. **Why and how?** Obtaining sequences from samples
   a. Recap
   b. Library preparation and sequencing concepts

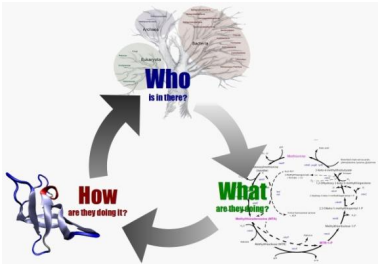
2. **What?** How does the actual data looks like
   a. Sequences in FASTQ format
   b. Metadata/Data

3. **Now, what?** Which ecological insight we usually obtain and how to interpret them
   a. OTU concept and picking
   b. Fundamental elements of ecological data analysis
   c. Community diversity evaluation and other data analysis example

1. WHY?

AIM: study the microbial community of a sample

- Evaluate who is there
- Evaluate their abundance
- Evaluate how the community change respect to condition, time, variable….
- Evaluate what, and how, the community is doing (i.e reduction of gut inflammation)

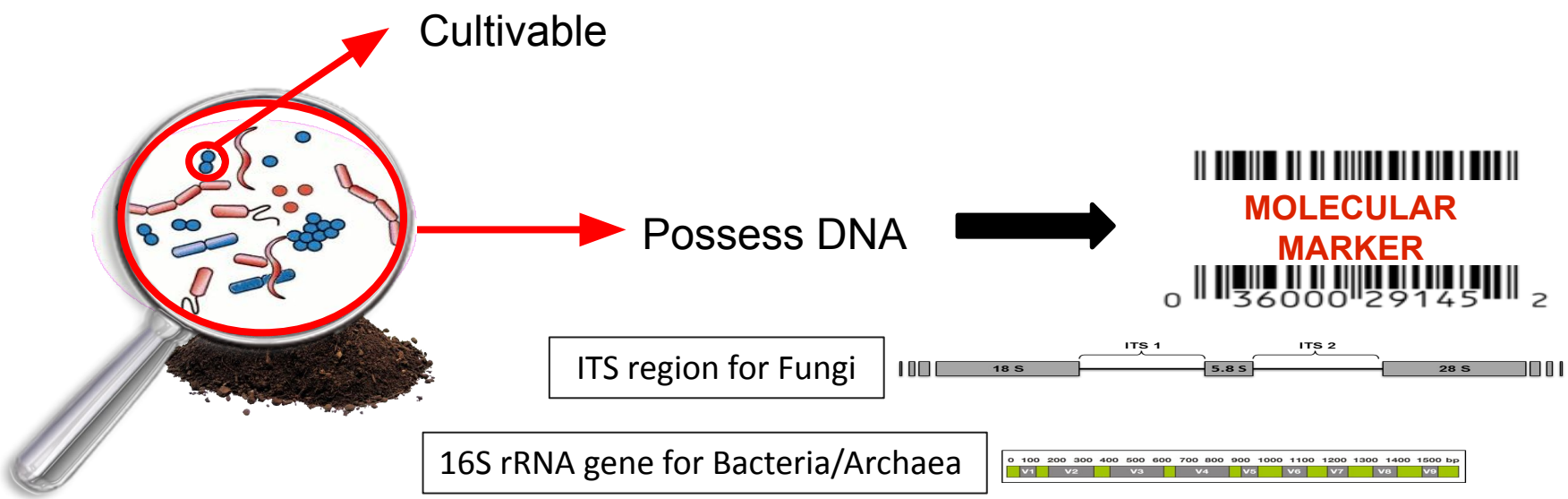


1. WHY?

PROBLEM: we (*microbial ecologist*) can't see and count microbes like ecologist do for trees, birds, sharks…

- ***First solution:*** cultivate microbes in the lab and study them, moving from the field to controlled lab environment
- ***Problem:*** only a really small fraction of the microbial diversity is actually cutivable

1. WHY?



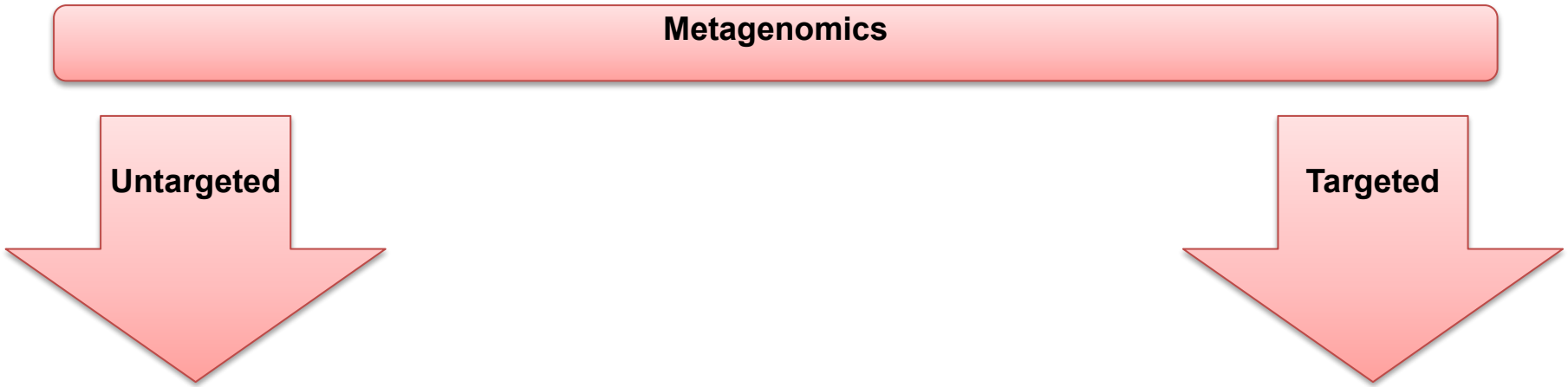
PROBLEM: we (*microbial ecologist*) can't see and count microbes like ecologist do for trees, birds, sharks…

- ***Second solution:*** cultivable or not, every microbes has DNA; we can use specific DNA region (i.e. Molecular marker) to evaluate the composition of the community
- Molecular methods are the norm in microbiology. The most recent development of such methods, is the ***METAGENOMICS***

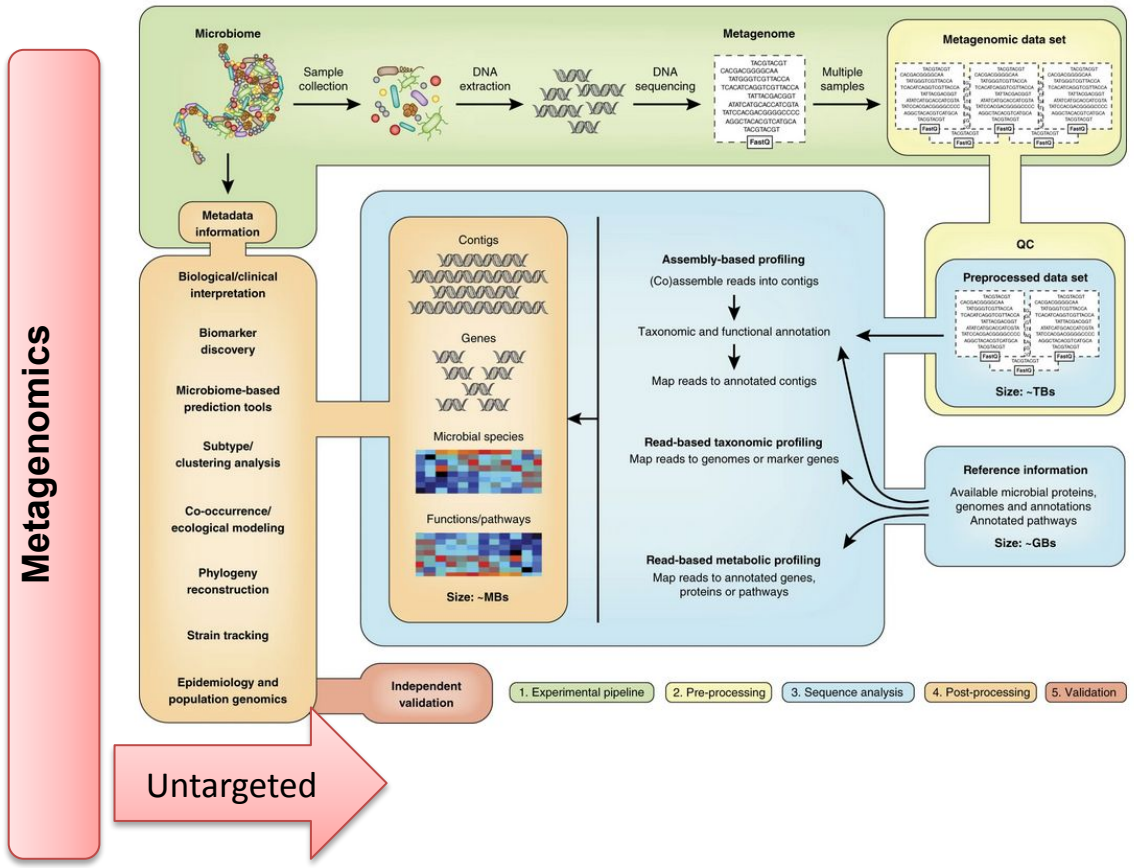
1. HOW? NGS and Metagenomics

NGS refers to the technologies but with Metagenomics we refer to the scientific topic

"Metagenomics is the study of genetic material recovered directly from environmental samples. The broad field may also be referred to as environmental genomics, ecogenomics or community genomics. (Wikipedia)"



1. HOW? NGS and Metagenomics



Application scenario:
The entire DNA is sequenced and bacterial genomes are reconstructed (MAGs).
From this huge amount of information we evaluate:
a) community composition
b) potential function of the community (evaluation of pathways)
c) virus, fagi, other organisms

nature biotechnology REVIEW

Shotgun metagenomics, from sampling to analysis

Christopher Quince[1,7], Alan W Walker[2,7], Jared T Simpson[3,4], Nicholas J Loman[5] & Nicola Segata[6]

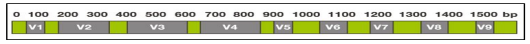
1. HOW? NGS and Metagenomics

Metagenomics

Targeted

ITS region for Fungi

16S rRNA gene for Bacteria/Archaea

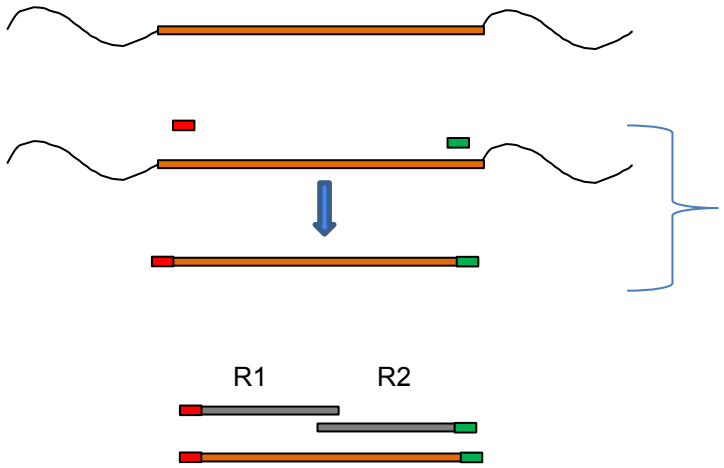


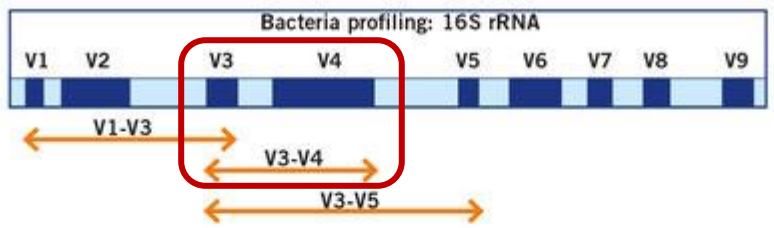
Application scenario:
Targeted metagenomics
(or Metabarcoding,
or Metataxonomic)
On 16S gene for bacterial
communities and ITS region for
fungal communities

MOLECULAR CLOCKs

1. HOW? Practical aspects of targeted sequencing

From STEP 4 of Leandro presentation





Genomic DNA fragment (or target region amplicon) is obtained

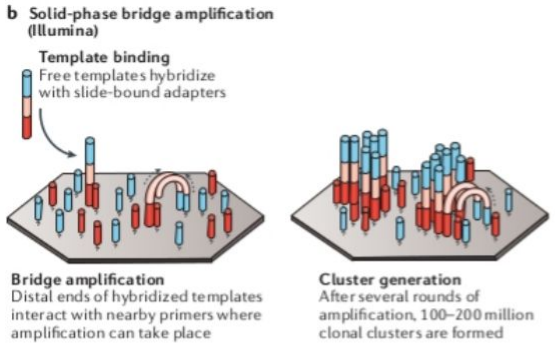
LIBRARY PREP: Target region is amplified with specific primers; adaptor are added for binding to the flow cell; fragments are added of adaptors

SEQUENCING: Two reads are produced, one starting from the forward primer (Read 1 or Read Forward) and the second from the reverse primer (Read 2 or Read Reverse)

- **Usually 2 reads of 300 base pair each are produced (2 x 300 PE protocol).**
- **Usually, the V3-V4 hypervariable regions of the 16S are targeted for sequencing (length approx 460 bp)**

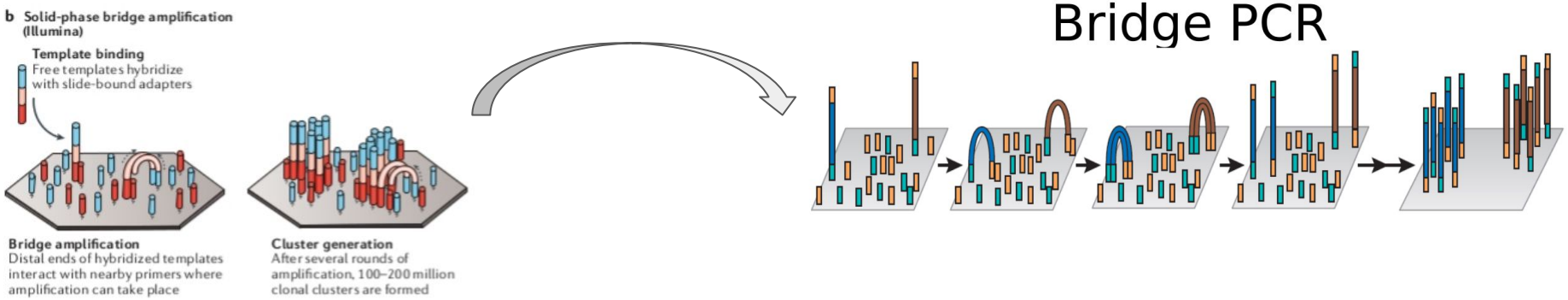
1. HOW? Practical aspects of targeted sequencing

From STEP 4 of Leandro presentation



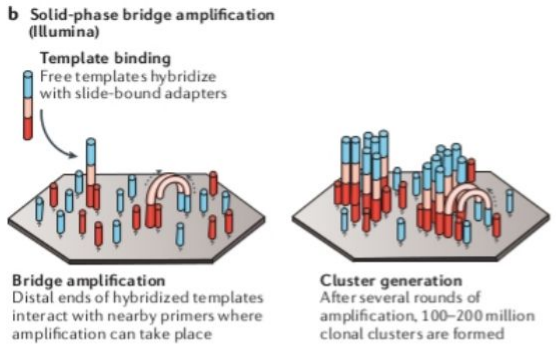
- Amplicons/fragments are attached to the flow cell thanks to the adapter sequences
- Each attached fragment is amplified to generate a cluster (BRIDGE PCR)

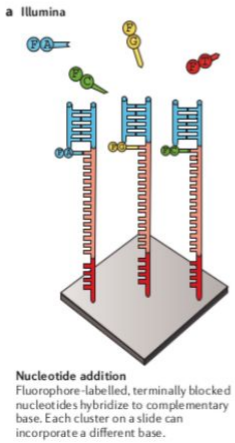
1. HOW? Practical aspects of targeted sequencing

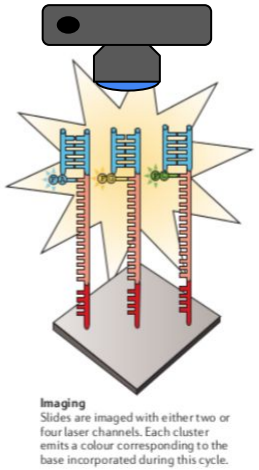


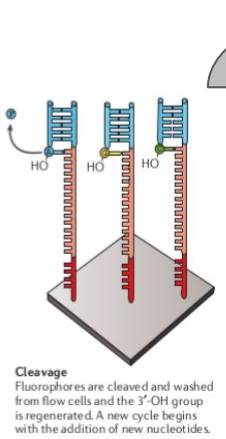
illumina®

1. HOW? Practical aspects of targeted sequencing

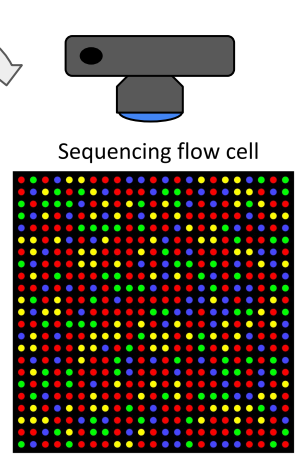




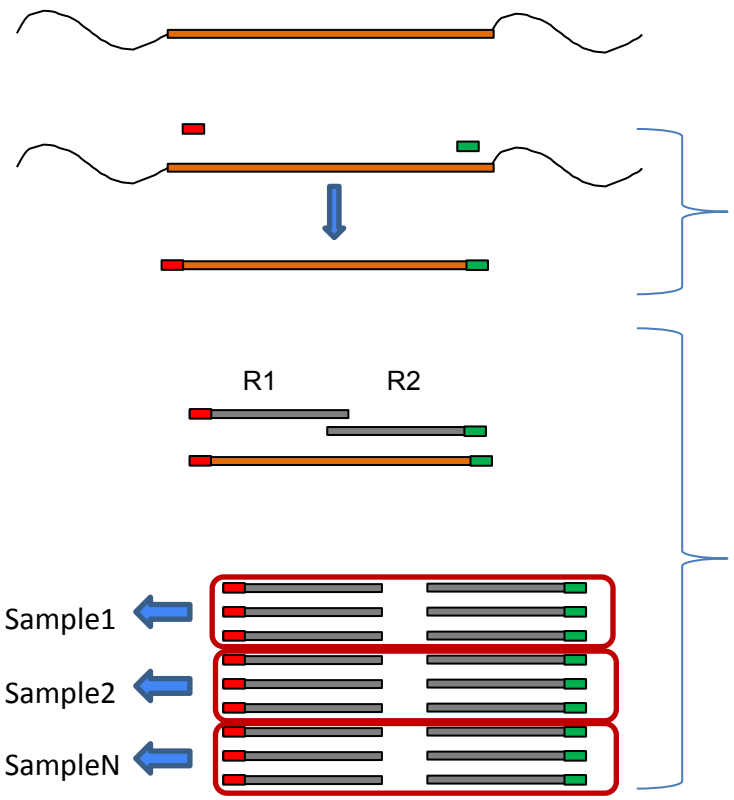




- Sequencing starts from the adapters
- At each cycle, all labeled nucleotide are added to the flow cell; only one nucleotide per cycle can react
- Image of the flow cell at each cycle allow to reconstruct the nucleotide sequence of a cluster.



1. HOW? Practical aspects of targeted sequencing



Genomic DNA fragment (or target region amplicon) is obtained

LIBRARY PREP: Target region is amplified with specific primers; adaptor are added for binding to the flow cell; fragments are added of adaptors

SEQUENCING: Two reads are produced, one starting from the forward primer (Read 1 or Read Forward) and the second from the reverse primer (Read 2 or Read Reverse)

DEMULTIPLEXING: Millions of sequences are produced for every run. By using short sequences (Barcodes or Indexes) in the adapter, produced reads are divided per sample

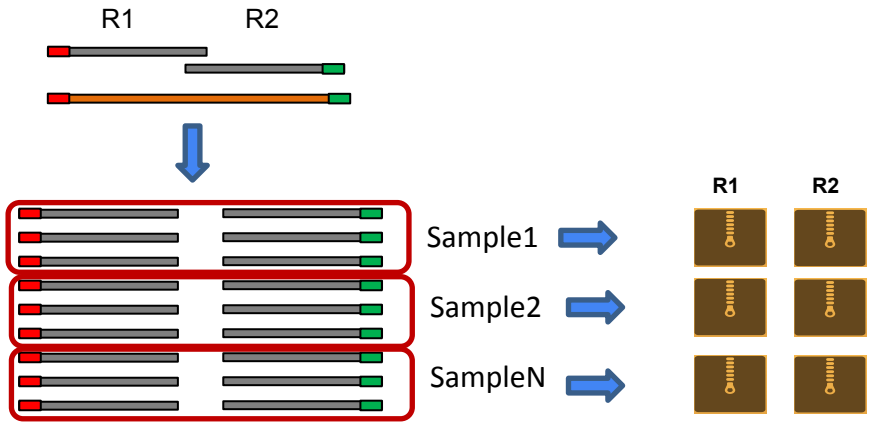
Multiplexing samples for sequencing: Sample indexing

- Multiple samples with different indices can be combined and sequenced together
- Depending on the application, may not need to generate many reads per sample (es. Environments with low biodiversity)
- Save money on sequencing costs (more samples per run), optimize use of read budget
- Sequencing center usually perform **DEMULTIPLEXING,** sequences are divided based on INDEX/BARCODE to the sample they belong
- **For each sample, two file are generated one for FORWARD and one for REVERSE read**

Index

ACAGTG
ACTTGA
ATCACG
CAGATC
CGATGT
GATCAG
GCCAAT
GGCTAC
TAGCTT
TGACCA
TTAGGC

2. WHAT? Sequencing file structure that your service provider will send



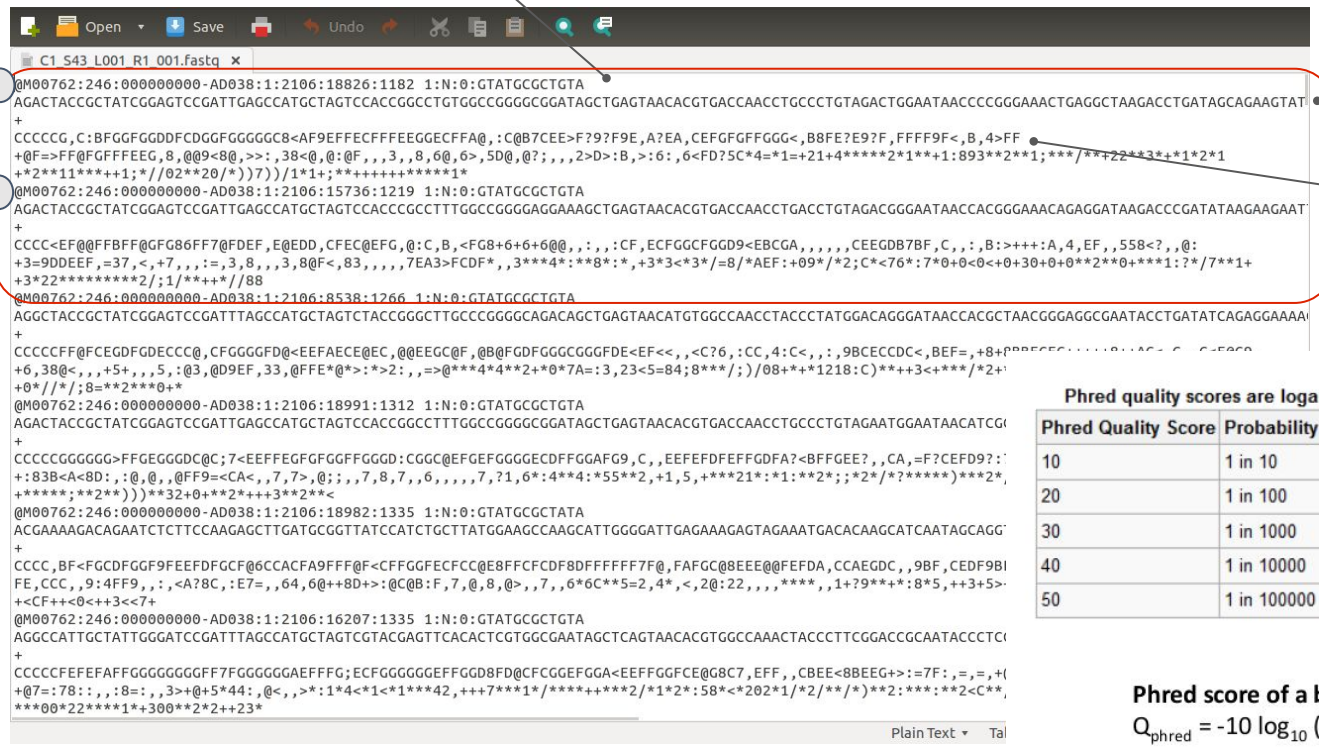
- Two compressed (often .gz format, may need 7zip for win) file for each sample, one for R1 and one for R2
- File are in FASTQ format; basically a plain text file

2. WHAT? Sequencing file structure that your service provider will send

Sequence “name” (position on the flow-cell)

Sequence base composition

Sequence per-base quality



Phred quality scores are logarithmically linked to error probabilities

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10000	99.99%
50	1 in 100000	99.999%

http://en.wikipedia.org/wiki/Phred_quality_score

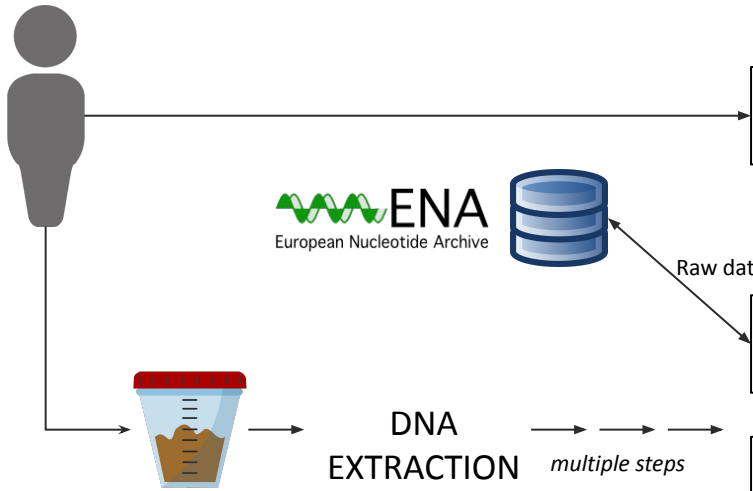
Phred score of a base is:

$Q_{phred} = -10 \log_{10} (e)$

where *e* is the estimated probability of a base being wrong

2. WHAT? Usually different assay on the same sample, metadata connect all

HOW DOES THE ACTUAL DATA LOOK LIKE?



METADATA RECORD

	AGE	BMI	Pathology
S1	35	24	
Sn			

All are plain, tab delimited, .txt files

COMMUNITY COMPOSITION DATA

	S1	S2	Sn
OTU1	500	350	
OTUn			

Count of occurrence of each OTU in all samples

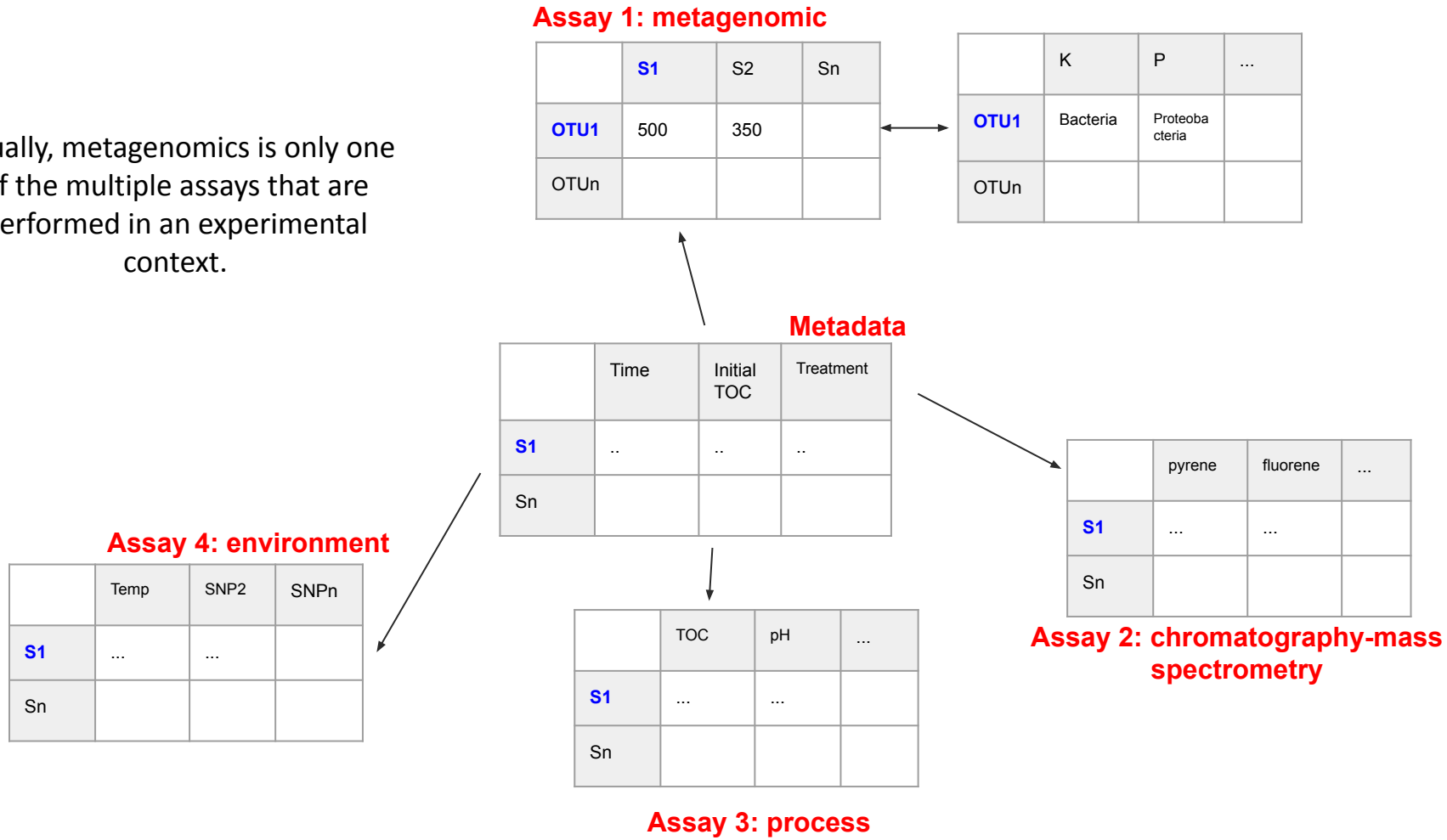
TAXONOMIC DATA

	K	P	...
OTU1	Bacteria	Proteobacteria	
OTUn			

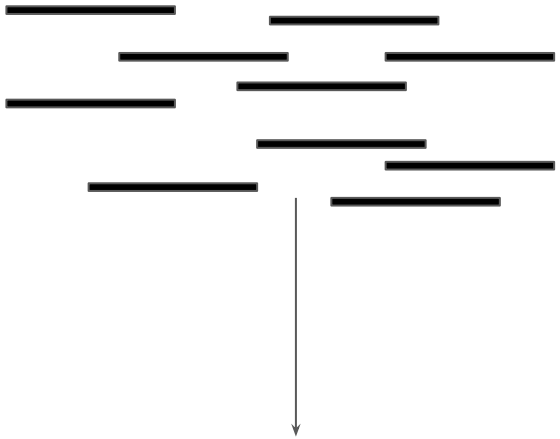
Taxonomic hierarchical classification of each OTU (*NCBITaxonomy*)

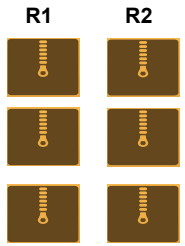
2. WHAT? Usually different assay on the same sample, metadata connect all

Usually, metagenomics is only one of the multiple assays that are performed in an experimental context.



3. Now, what?



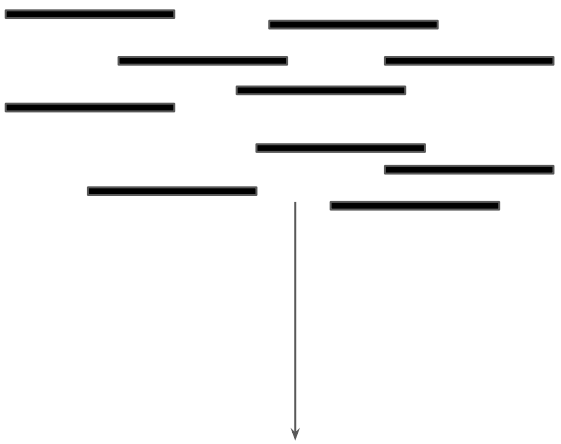


So far we obtained sequences…

UNTARGETED: millions of reads per sample
TARGETED: 50-100 k reads are obtained for each sample

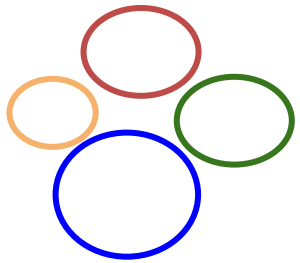
Need to group them in entities that can be counted or analyzed in other ways

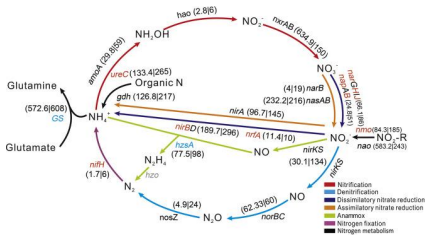
3. Now, what?



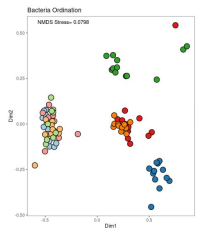
Alignment and assembly

Metagenome-assembled genomes (MAGs)





Reconstruction of pathways



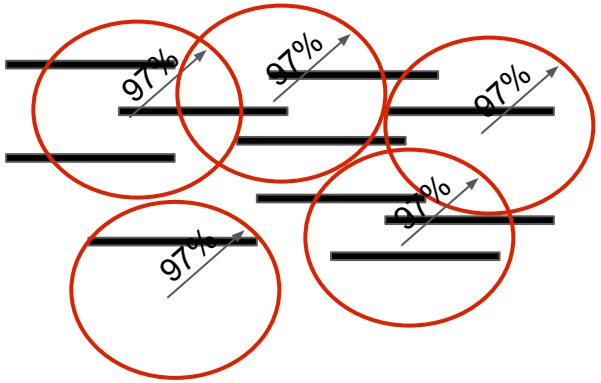
Diversity & composition

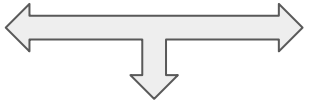
Study ARGs "resistome"

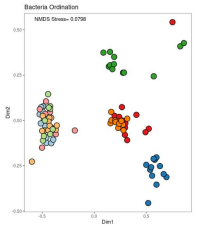
Untargeted Metagenomics

3. Now, what? Operational Taxonomic Unit and Amplicon Sequence Variant

Group seqs in OTUs







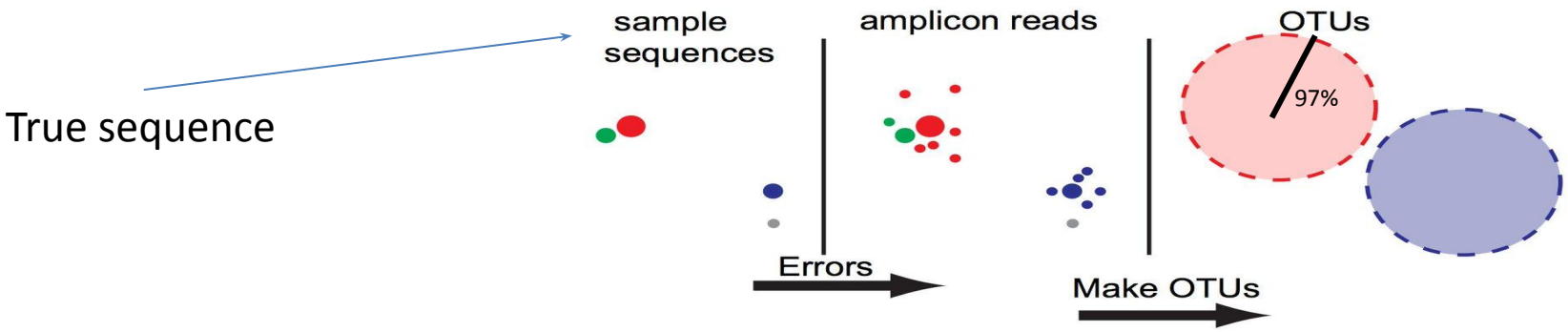
Diversity & composition

Targeted Metagenomics

3. Now, what? Operational Taxonomic Unit and Amplicon Sequence Variant

Bias with similarity based OTU picking

Sequencing error generate «Child» of true seq



PROBLEM: distinguish a real sequence variant (SV) from point sequencing error

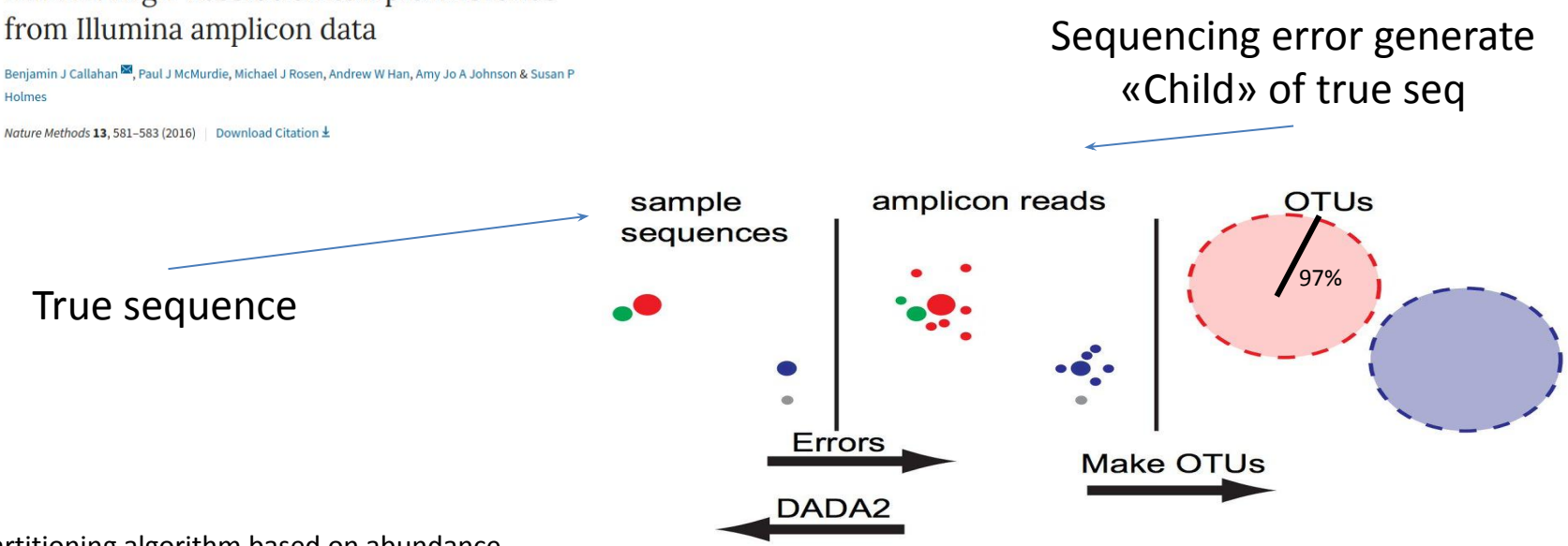
3. Now, what? Operational Taxonomic Unit and Amplicon Sequence Variant

Brief Communication | Published: 23 May 2016

DADA2: High-resolution sample inference from Illumina amplicon data

Benjamin J Callahan, Paul J McMurdie, Michael J Rosen, Andrew W Han, Amy Jo A Johnson & Susan P Holmes

Nature Methods **13**, 581–583 (2016) | Download Citation



- Partitioning algorithm based on abundance
- All sequences start in the same partition, sequences with equal base composition are grouped in unique sequences (DEREPLICATION)
- Each unique sequence has associated abundance
- Assuming that if the abundance (and its p-value) of reads in an unique sequence is higher than a treshold, the abundance cannot be explained by sequencing errors alone
- If this is true, form another partition with the unique sequence as centroid
- Repeat

3. Now, what? Operational Taxonomic Unit and Amplicon Sequence Variant

UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing

Robert C. Edgar
Independent Investigator
Tiburon, California, USA.
robert@drive5.com

DENOISER: infer accurate biological template sequences from noisy reads.

1. correcting point errors to obtain an accurate set of amplicon sequences (denoising) and
2. filtering of chimeric amplicons.

The result is a set of predicted biological sequences that I call ZOTUs (zero-radius OTUs). ZOTUs are valid operational taxonomic units that are superior to conventional 97% OTUs for most purposes because they provide the maximum possible biological resolution given the data while using 97% identity may merge phenotypically different strains with distinct sequences into a single cluster(Tikhonov et al., 2015; Callahan et al., 2016).

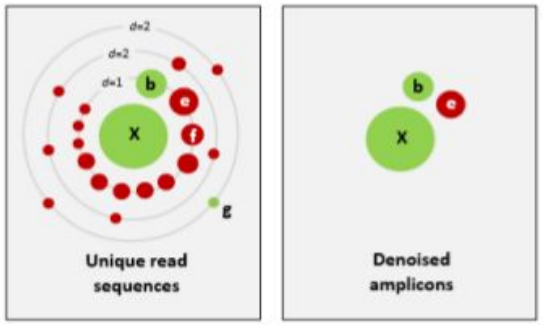


Figure 1. Schematic of the UNOISE2 denoising strategy. The left panel shows the neighborhood close to a high-abundance unique read sequence **X**, grouped by the number of sequence differences (*d*). Dots are unique sequences, the size of a dot indicates its abundance. Green dots are correct biological sequences; red dots have one or more errors. Neighbors with small numbers of differences and small abundance compared to **X** are predicted to be bad reads of **X.** The right panel shows the denoised amplicons. Here, **X** and **b** were correctly predicted, **e** is an error with anomalously high abundance that was wrongly predicted to be correct, **f** is an error that was correctly discarded but has an abundance almost high enough to be a false positive, and **g** is a low-abundance correct amplicon that was wrongly discarded. The abundances of **b**, **e**, and **f** are similar, illustrating the fundamental challenge in denoising: how to set an abundance threshold that distinguishes correct sequences from errors.

3. Now, what? Operational Taxonomic Unit and Amplicon Sequence Variant

- **Operational taxonomic unit (OTUs)** are formed based on sequence identity, sequences are clustered together if they are more similar than a user-defined identity threshold, presented as a percentage traditionally set it at 97% of sequence similarity

1. The *De Novo* approach groups sequences based on sequence similarity
2. The closed-reference approach matches sequences to an existing database of reference sequences, if a sequence fails to match the database, it is discarded
3. The open-reference approach also starts with an existing database and tries to match the sequences against them, if a sequence does not match the data-base, it is added to the database as a new reference sequence

- **Amplicon sequence variants (ASV**) New methods control errors sufficiently such that amplicon sequence variants (ASVs) can be resolved exactly, down to the level of single-nucleotide differences over the sequenced gene region
  - https://www.nature.com/articles/ismej2017119

3. Now, what? Operational Taxonomic Unit and Amplicon Sequence Variant

At present, there is some terminology chaos:

- **Operational taxonomic unit (OTUs)** is usually used for referring to grouping sequences on a similarity base and using a threshold of 97% sequence similarity
- **Amplicon sequence variants (ASV)** (sometimes referred also as ESV: Exact sequence variant or zero-radius OTU (zOTU) is usually used for referring to grouping sequences on the base of single nucleotide difference (in a sense, OTU defined by 100% sequence similarity)

- Overall, they refer to the *same general concept* of Operational Taxonomic Unit:
  - an entity in which grouping similar DNA sequences, which has a taxonomic meaning
  - taxonomic meaning is usually at the genus/species level; 400-500 bp of the sequenced region are hardly enough for species level classification

3. Now, what? Operational Taxonomic Unit and Amplicon Sequence Variant

Resulting OTU/ASV table

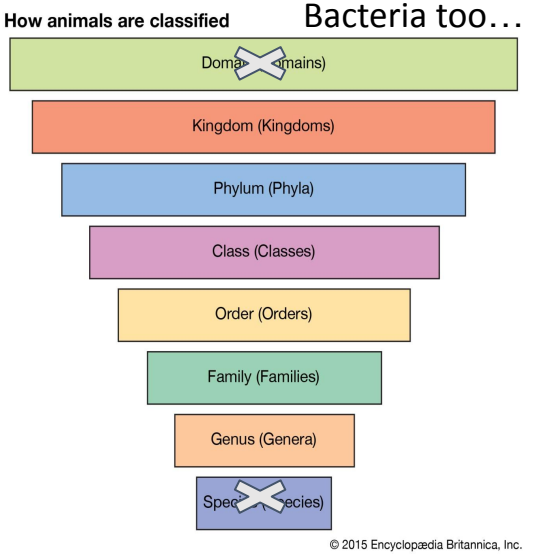
	A	B	C	D	E
1	OTU	Taxonomy	ED1-GKBN-A	ED10-SVAN-T1	ED11-OOBED-T0
2	OTU1	Bacteria;Actinobacteria;Actinobacteria;Actinomycetales;Corynebacteriaceae;Corynebacterium	1	8	0
3	OTU2	Bacteria;Actinobacteria;Actinobacteria;Actinomycetales;Micrococcaceae;Rothia	0	0	6
4	OTU3	Bacteria;Actinobacteria;Actinobacteria;Bifidobacteriales;Bifidobacteriaceae;Bifidobacterium	4937	430	151
5	OTU4	Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Asaccharobacter	0	2	0
6	OTU5	Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Collinsella	0	76	55
7	OTU6	Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Eggerthella	315	0	0
8	OTU7	Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Enterorhabdus	0	0	2
9	OTU8	Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Gordonibacter	24	7	0
10	OTU9	Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Olsenella	0	0	0
11	OTU10	Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Slackia	0	0	8
12	OTU11	Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Unclassified	5	13	33
13	OTU12	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides	7262	15661	4382
14	OTU13	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella	68	0	1448
15	OTU14	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Butyricimonas	0	635	430
16	OTU15	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Coprobacter	0	134	0
17	OTU16	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Odoribacter	349	0	2041
18	OTU17	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Parabacteroides	3163	1272	389
19	OTU18	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Unclassified	222	229	57
20	OTU19	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Alloprevotella	0	0	103
21	OTU20	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Paraprevotella	100	0	0
22	OTU21	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Prevotella	0	0	6586
23	OTU22	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes	2423	2902	6625
24	OTU23	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Unclassified;Unclassified	387	0	0
25	OTU24	Bacteria;Bacteroidetes;Unclassified;Unclassified;Unclassified;Unclassified	0	0	219
26	OTU25	Bacteria;Firmicutes;Bacilli;Bacillales;Staphylococcaceae;Staphylococcus	9	5	6
27	OTU26	Bacteria;Firmicutes;Bacilli;Lactobacillales;Aerococcaceae;Facklamia	2	2	0

3. Now, what? Fundamental elements of ecological data analysis

ENTITIES (Species, Genes, **OTUs** …)

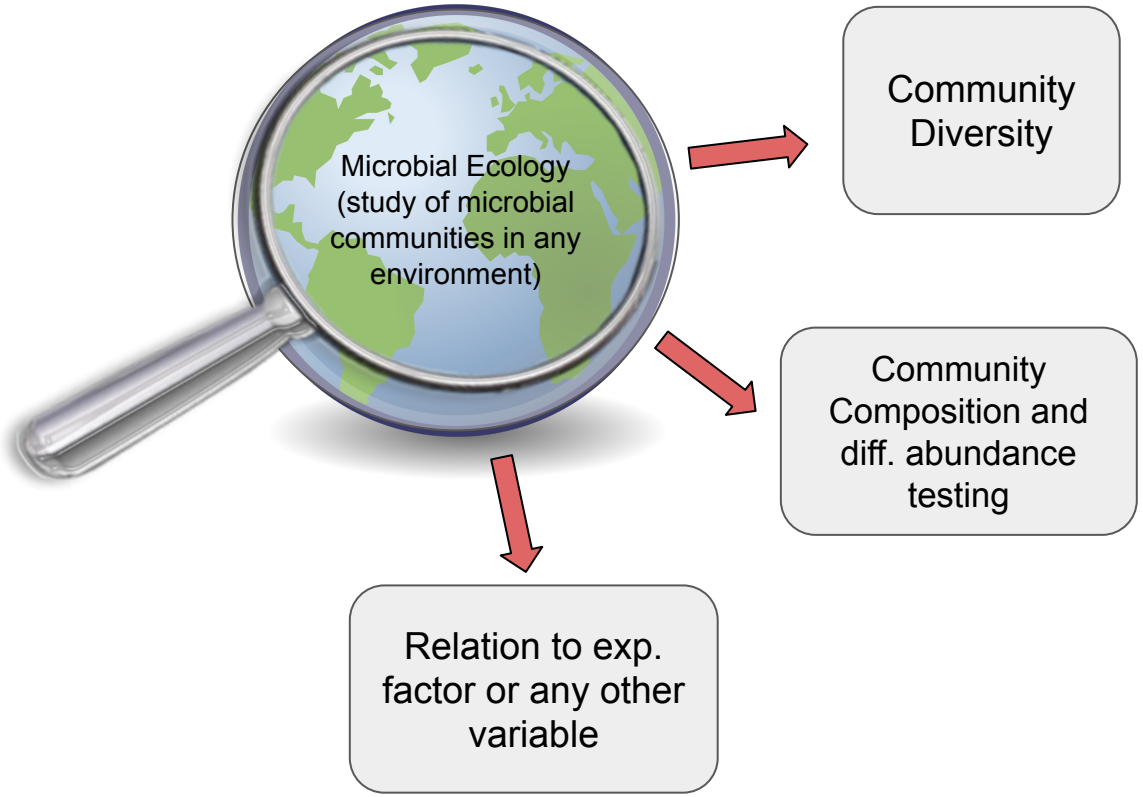
Resulting OTU/ASV table

	A	B	C	D	E
1	OTU	Taxonomy	ED1-GKBN-A	ED10-SVAN-T1	ED11-OOBED-T0
2	OTU1	Bacteria;Actinobacteria;Actinobacteria;Actinomycetales;Corynebacteriaceae;Corynebacterium	1	8	0
3	OTU2	Bacteria;Actinobacteria;Actinobacteria;Actinomycetales;Micrococcaceae;Rothia	0	0	6
4	OTU3	Bacteria;Actinobacteria;Actinobacteria;Bifidobacteriales;Bifidobacteriaceae;Bifidobacterium	4937	430	151
5	OTU4	Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Asaccharobacter	0	2	0
6	OTU5	Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Collinsella	0	76	55
7	OTU6	Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Eggerthella	315	0	0
8	OTU7	Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Enterorhabdus	0	0	2
9	OTU8	Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Gordonibacter	24	7	0
10	OTU9	Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Olsenella	0	0	0
11	OTU10	Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Slackia	0	0	8
12	OTU11	Bacteria;Actinobacteria;Actinobacteria;Coriobacteriales;Coriobacteriaceae;Unclassified	5	13	33
13	OTU12	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Bacteroidaceae;Bacteroides	7262	15661	4382
14	OTU13	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Barnesiella	68	0	1448
15	OTU14	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Butyricimonas	0	635	430
16	OTU15	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Coprobacter	0	134	0
17	OTU16	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Odoribacter	349	0	2041
18	OTU17	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Parabacteroides	3163	1272	389
19	OTU18	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Unclassified	222	229	57
20	OTU19	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Alloprevotella	0	0	103
21	OTU20	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Paraprevotella	100	0	0
22	OTU21	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Prevotellaceae;Prevotella	0	0	6586
23	OTU22	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Rikenellaceae;Alistipes	2423	2902	6625
24	OTU23	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Unclassified;Unclassified	387	0	0
25	OTU24	Bacteria;Bacteroidetes;Unclassified;Unclassified;Unclassified;Unclassified	0	0	219
26	OTU25	Bacteria;Firmicutes;Bacilli;Bacillales;Staphylococcaceae;Staphylococcus	9	5	6
27	OTU26	Bacteria;Firmicutes;Bacilli;Lactobacillales;Aerococcaceae;Facklamia	2	2	0



ABUNDANCE (i.e. Counts of sequences) of each COUNTED ENTITY in the SAMPLES
Each entity is identified in comparison to known sequence database
and is characterized by a TAXONOMIC classification

3. Now, what? Community diversity evaluation and other data analysis



3. Now, what? Community diversity evaluation and other data analysis

	S1	S2	S3	S4
S1	0	-	-	-
S2	10	0	-	-
S3	8	1	0	-
S4	2	7	8	0

Dissimilarity calculation
(Bray-Curtis, UniFrac)

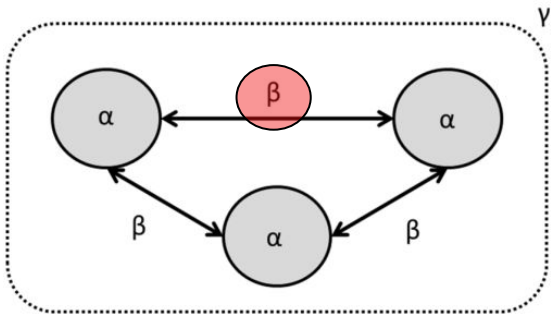
Beta-Diversity

- ordination analysis
- relation to var.
- clustering

- **Beta-diversity:** measure the diversity between couples of samples
- Information of abundance of OTUs in the samples are used to calculate dissimilarity matrix, which is the base for ordination analysis (and other)
- The ordination can be related to variable to have statistic and visual

	OTU1	OTU2	OTU3	OTU4	OTU5	OTU6
S1	..	..	..	..	..	..
S2	..	..	..	..	..	..
S3	..	..	..	..	..	..
S4	..	..	..	..	..	..

	SEX	AGE	PATH
S1	M	24	CRC
S2	M	30	CTR
S3	M	25	CTR
S4	M	25	CRC



3. Now, what? Community diversity evaluation and other data analysis

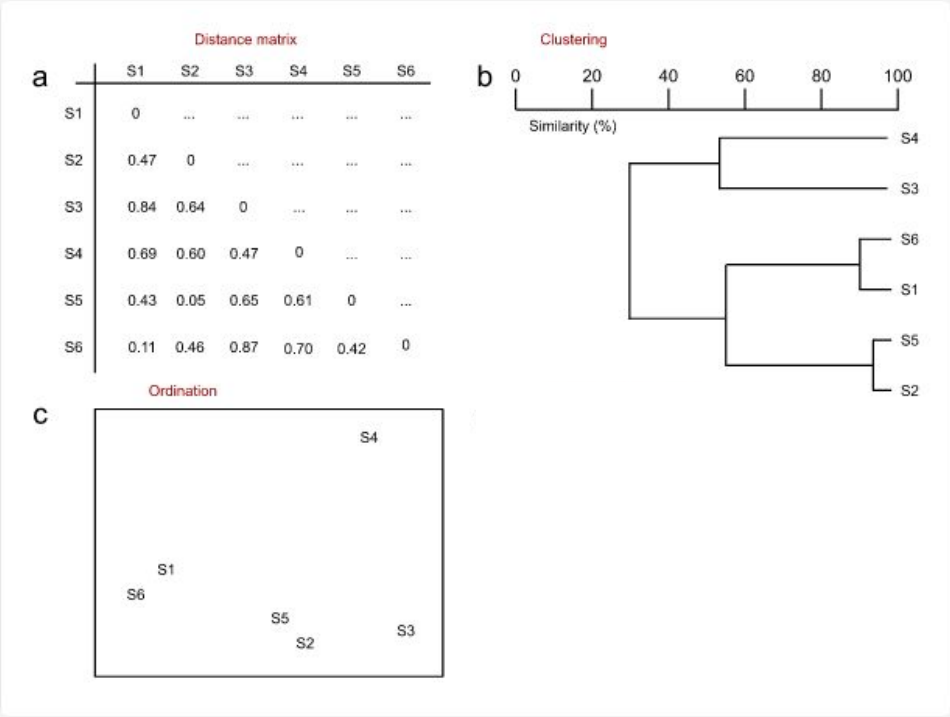
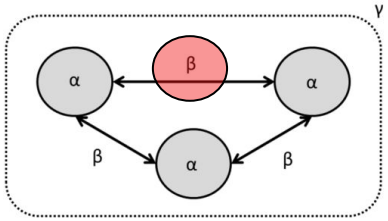


Figure 1: A distance matrix (a) provides input for both b) hierarchical cluster analysis and c) non-metric dimensional scaling. The results of the cluster analysis may be superimposed on the ordination (d) to validate that each solution corroborates the other. Adapted from Ramette 2007, originally adapted from Legendre & Legendre 1998.

Buttigieg PL, Ramette A (2014) *A Guide to Statistical Analysis in Microbial Ecology: a community-focused, living review of multivariate data analyses*. ***FEMS Microbiol Ecol. 90: 543–550.***

https://sites.google.com/site/mb3gustame/home

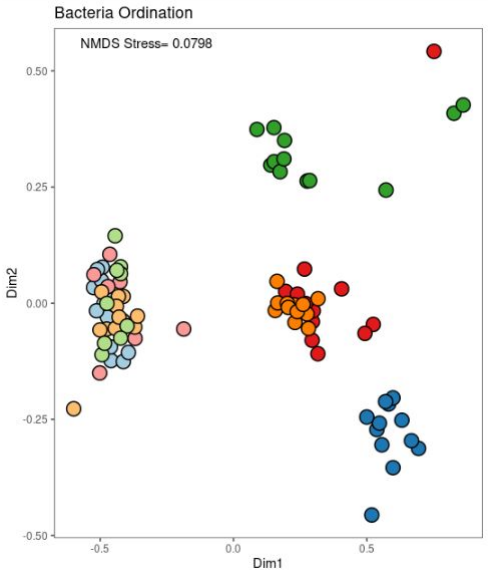


- The distance **a)** matrix (Bray-Curtis) can be used for different kind of analysis aimed at visualizing/measuring the similarity between different samples (i.e. various ways of analysing beta-diversity).

- The **c)** ordination analysis is defined UNSUPERVISED (or **UNCONSTRAINED**) as there is no included information about variables, just the community composition (i.e. OTU table). Some example of methods:
  - PCA
  - **PCoA**
  - NMDS

3. Now, what? Community diversity evaluation and other data analysis

- expectation…..



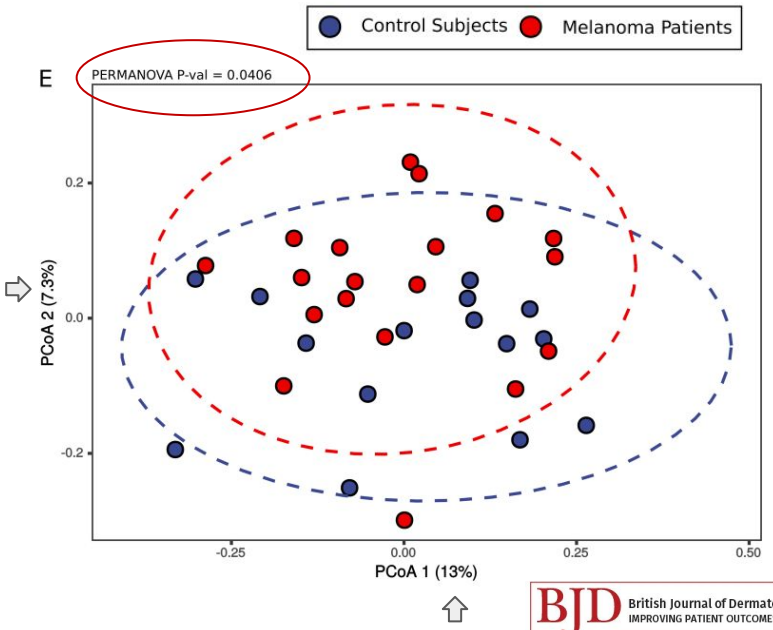
- Color of point represents an experimental factor with various levels
- If the experimental factor has a clear effect on microbiota composition, we expect to be able to see discrete samples clusters
- We may conclude that the experimental factor has a real biological effect in shaping microbiota composition.

If an experimental factor has a real biological effect on the community, we expect to see samples belonging to a level of the experimental factor to be closer in the ordination (i.e. low within group dissimilarity) respect to samples from other level of the experimental factor (i.e. high between group dissimilarity).

This generate discrete cluster with the same color

3. Now, what? Community diversity evaluation and other data analysis

- reality…..



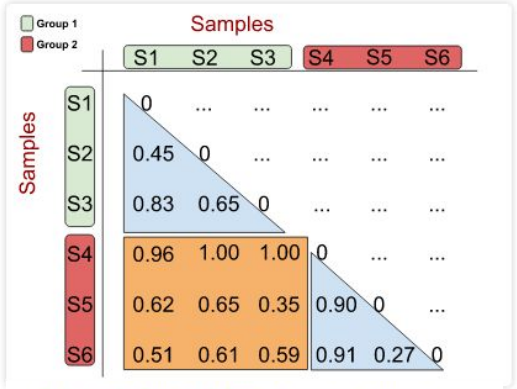
BJD British Journal of Dermatology
IMPROVING PATIENT OUTCOMES IN SKIN DISEASE WORLDWIDE
ORIGINAL ARTICLE | Open Access

Early melanoma invasivity correlates with gut fungal and bacterial profiles

F. Vitali, R. Colucci, M. Di Paola, M. Pindo, C. De Filippo, S. Moretti, D. Cavalieri

First published: 06 July 2021 | https://doi.org/10.1111/bjd.20626

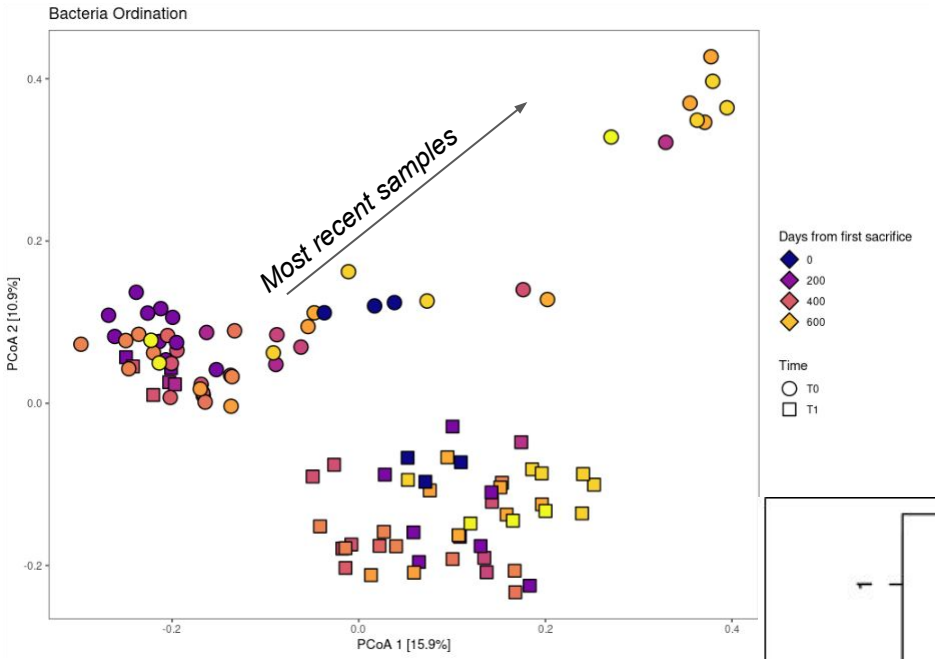
- Microbiota data have high variability/noise and often the picture is not defined. We need other methods than simple observation on the ordination graph
- PERMANOVA is a statistical test used for that purpose; compares within group dissimilarities (blue) to between group dissimilarities (orange) in an ANOVA-like analysis of one or multiple factors/variables



Buttigieg PL, Ramette A (2014) A Guide to Statistical Analysis in Microbial Ecology: a community-focused, living review of multivariate data analyses. FEMS Microbiol Ecol. 90: 543–550.

https://sites.google.com/site/mb3gustame/home

3. Now, what? Community diversity evaluation and other data analysis



- Not only categorical variables; ordination can be inspected also considering continuous numerical variables

- Relations can be simply visualized as point color
- Or statistical testing can be performed with PERMANOVA (so, relation of the distance matrix to the variable in ANOVA type of analysis) or with fitting methods, such as ENVFIT/ORDISURF (which search the best fit of coordinates of points in ordination, to a numerical variable)

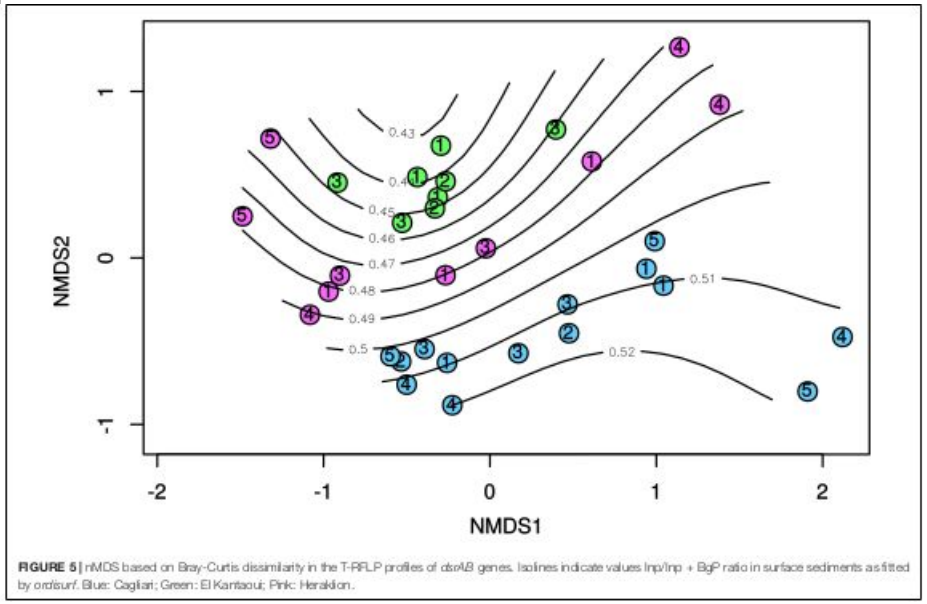


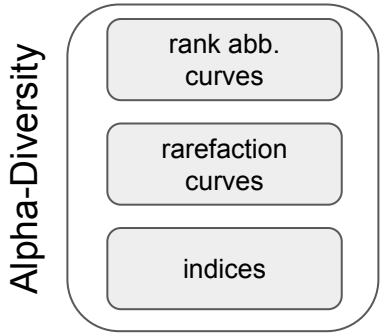
FIGURE 5 | nMDS based on Bray-Curtis dissimilarity in the T-RFLP profiles of *dsrAB* genes. Isolines indicate values Inp/Inp + BgP ratio in surface sediments as fitted by *ordisurf*. Blue: Cagliari; Green: El Kantaoui; Pink: Heraklion.

ORIGINAL RESEARCH article
Front. Mar. Sci., 20 September 2019 | https://doi.org/10.3389/fmars.2019.00590

Benthic Prokaryotic Community Response to Polycyclic Aromatic Hydrocarbon Chronic Exposure: Importance of Emission Sources in Mediterranean Ports

Francesco Vitali[1,2], Manolis Mandalakis[3], Eva Chatzinikolaou[3], Thanos Dailianis[3], Giuliana Senatore[2], Enrico Casalone[2], Giorgio Mastromei[2], Simona Sergi[2], Raffaela Lussu[1], Christos Arvanitidis[3] and Elena Tamburini[1*]

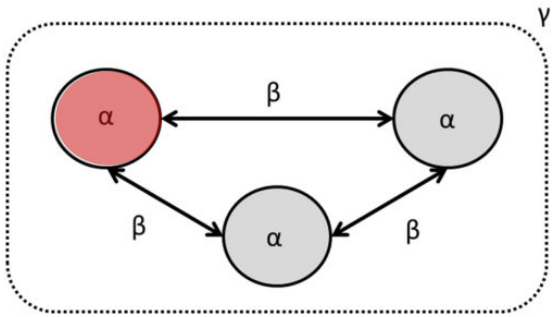
3. Now, what? Community diversity evaluation and other data analysis



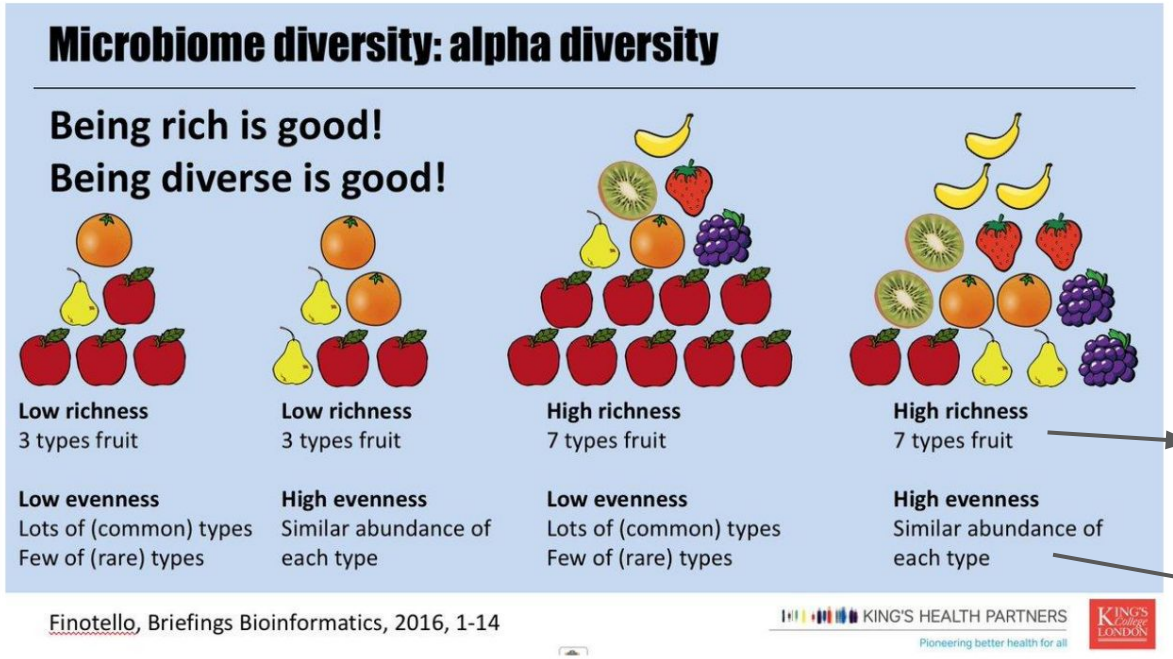
- **Alpha-diversity:** measure the diversity of the community in each sample
- Uses the information of abundance of OTUs in the samples (and information from metadata)
- Indices: we obtain a value of the indices for each sample, then we can compare the values between i.e. groups

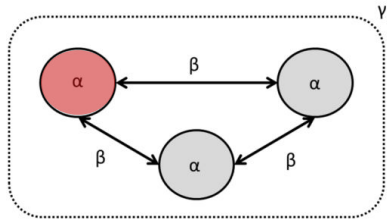
	OTU1	OTU2	OTU3	OTU4	OTU5	OTU6
S1	..	..	..	..	..	..
S2	..	..	..	..	..	..
S3	..	..	..	..	..	..
S4	..	..	..	..	..	..

	SEX	AGE	COND
S1	M	24	CRC
S2	M	30	CTR
S3	M	25	CTR
S4	M	25	CRC

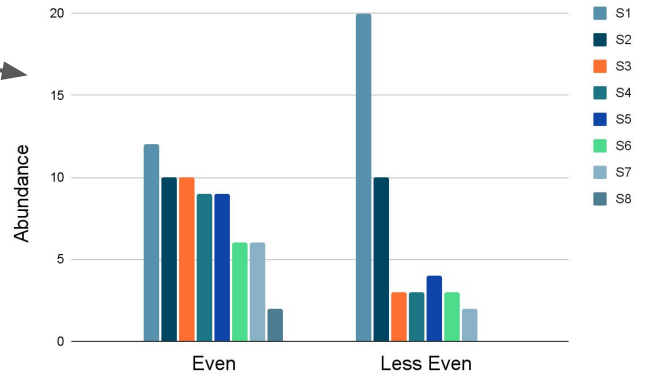


3. Now, what? Community diversity evaluation and other data analysis





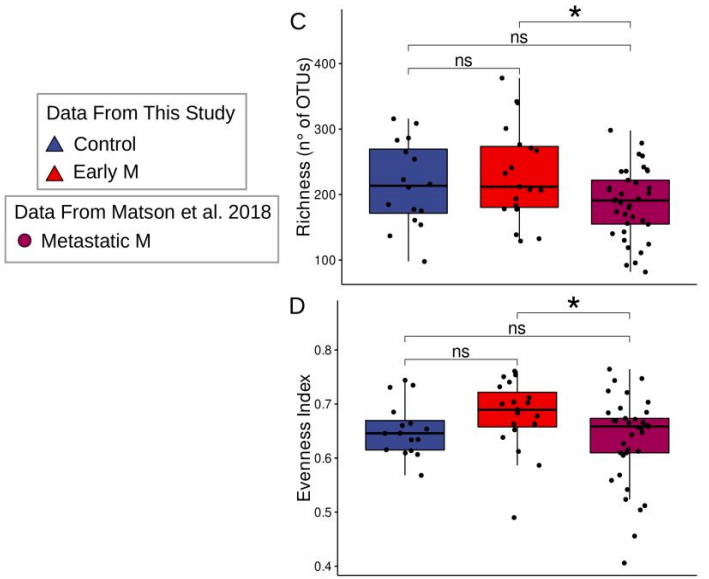
- **Richness**: number of different OTUs in a sample



- **Evenness**: evaluation of equitability of the distribution of abundance of different OTUs in a sample
- **Shannon's index**: a diversity index increasing both with richness and evenness

3. Now, what? Community diversity evaluation and other data analysis

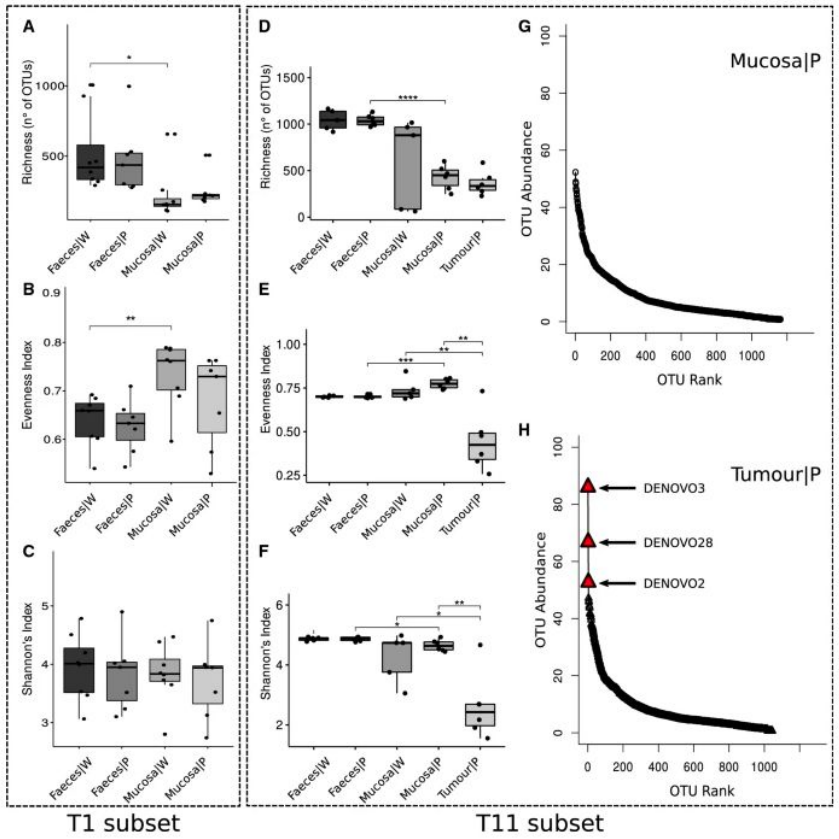
- Indices values in different samples group can be explored with a Boxplot representation, coupled with statistical testing (i.e Wilcoxon test)
- In the example, metastatic M had lower richness and evenness than early M and control, suggesting a decline in community diversity along the control - early - metastatic melanoma axis



3. Now, what? Community diversity evaluation and other data analysis

A more “dynamic” view of alpha diversity

- Rank abundance curves can suggest specific OTUs dominance in the community (Dominant = abundance higher, stands out from the rest of the curve)



scientific reports

OPEN

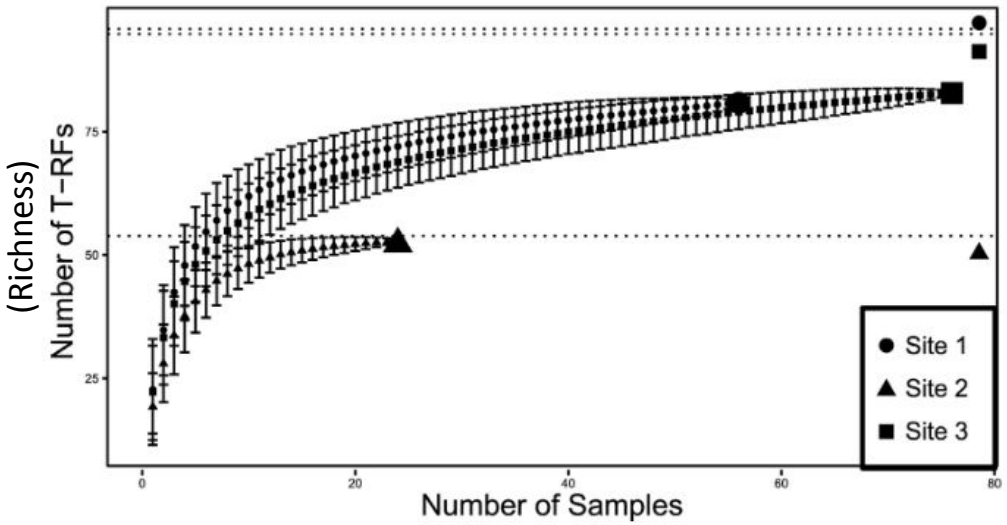
Intestinal microbiota profiles in a genetic model of colon tumorigenesis correlates with colon cancer biomarkers

Francesco Vitali, Katia Tortora, Monica Di Paola, Gianluca Bartolucci, Marta Menicatti, Carlotta De Filippo & Giovanna Caderni

3. Now, what? Community diversity evaluation and other data analysis

A more “dynamic” view of alpha diversity

- Rank abundance curves can suggest specific OTUs dominance in the community (Dominant = abundance higher, stands out from the rest of the curve)
- Species accumulation curves can suggest differences in how an indices evolves with different sampling efforts (n° of samples, or n° of reads)

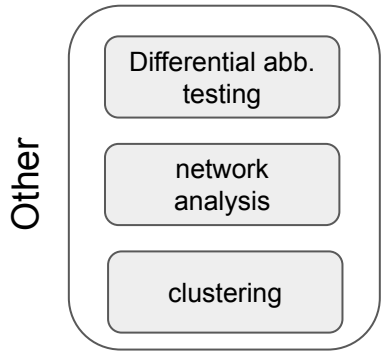


Contents lists available at ScienceDirect
Microbiological Research
journal homepage: www.elsevier.com/locate/micres

Long lasting effects of the conversion from natural forest to poplar plantation on soil microbial communities

Francesco Vitali, Giorgio Mastromei, Giuliana Senatore, Cesarea Caroppo, Enrico Casalone*
Department of Biology, University of Florence, Via Madonna del Piano 6, Sesto Fiorentino, 50019 Florence, Italy

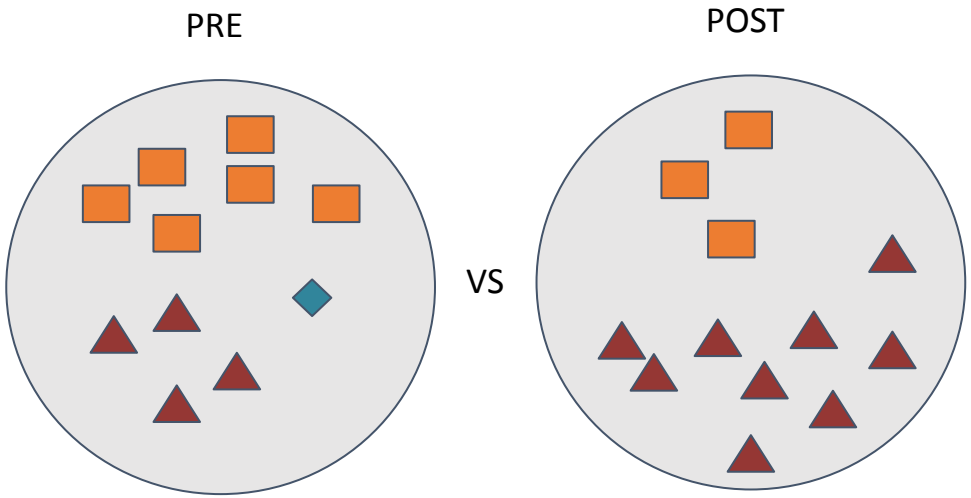
3. Now, what? Community diversity evaluation and other data analysis



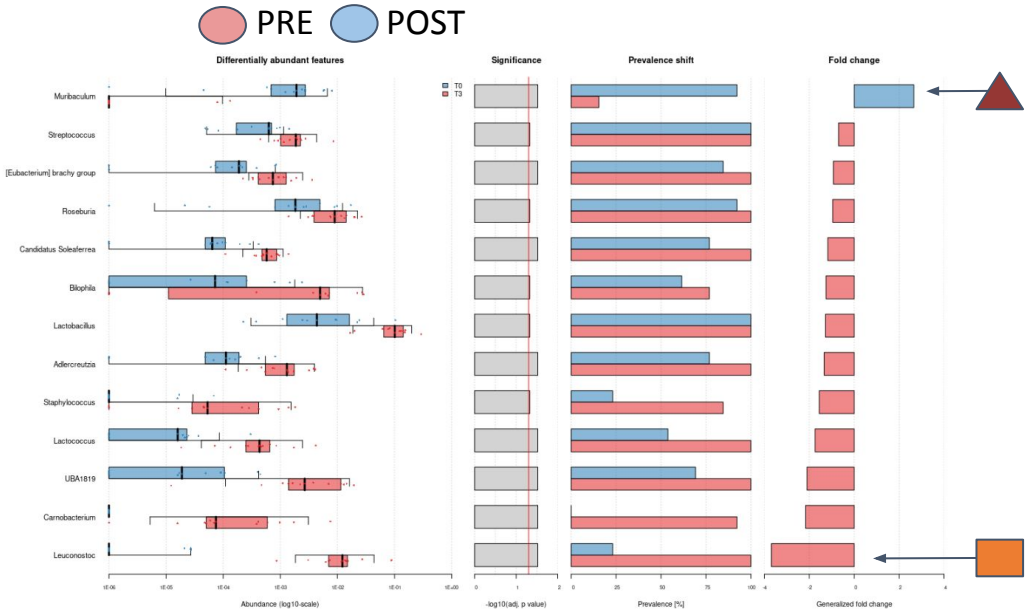
- Possibilities for **OTHER** kind of data analysis are many
- The third “great classic” of microbiota analysis is differential abundance testing
- Many methods (even much pitfalls) are available

	OTU1	OTU2	OTU3	OTU4	OTU5	OTU6
S1	..	..	..	..	..	..
S2	..	..	..	..	..	..
S3	..	..	..	..	..	..
S4	..	..	..	..	..	..

	SEX	AGE	PATH
S1	M	24	CRC
S2	M	30	CTR
S3	M	25	CTR
S4	M	25	CRC

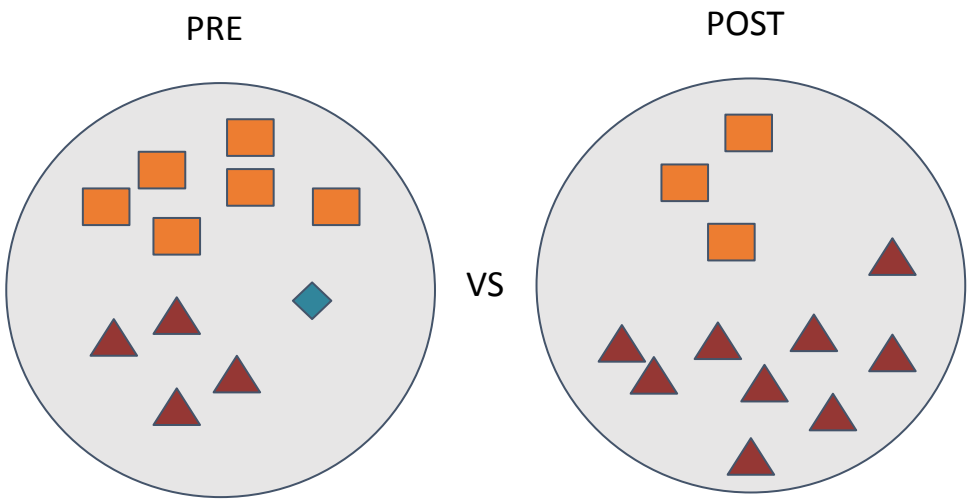


3. Now, what? Community diversity evaluation and other data analysis

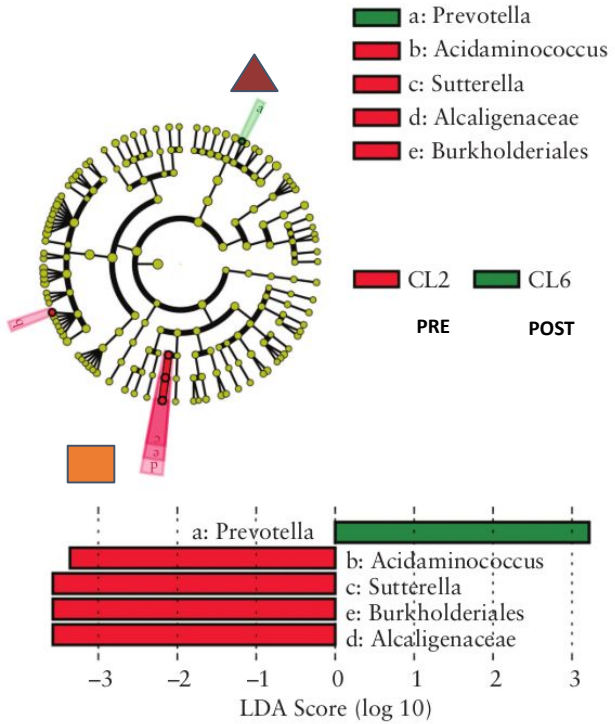


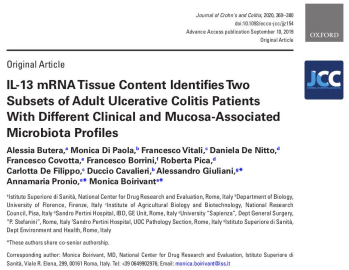
SIAMCAT function; Paired wilcoxon with fdr correction for multiple comparisons

- Possibilities for **OTHER** kind of data analysis are many
- The third “great classic” of microbiota analysis is differential abundance testing
- Many methods (even much pitfalls) are available



3. Now, what? Community diversity evaluation and other data analysis

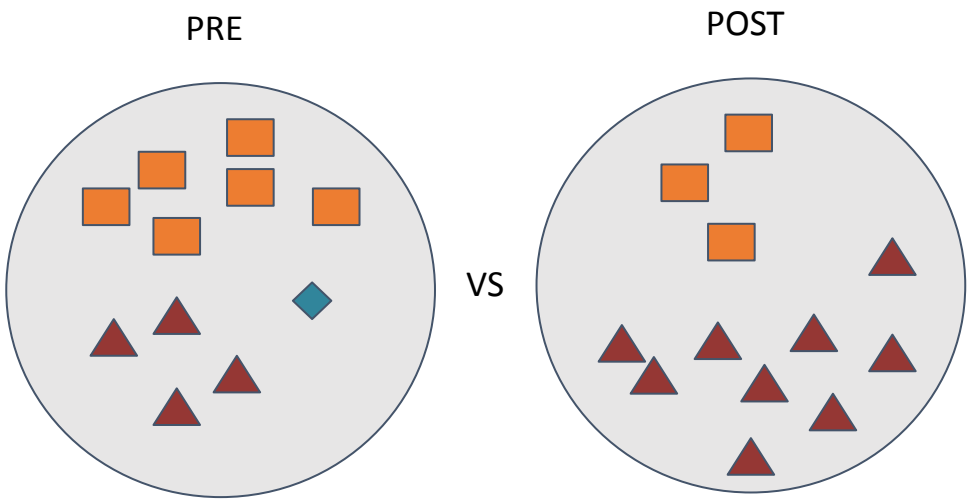




Original Article

IL-13 mRNA Tissue Content Identifies Two Subsets of Adult Ulcerative Colitis Patients With Different Clinical and Mucosa-Associated Microbiota Profiles

- Possibilities for **OTHER** kind of data analysis are many
- The third “great classic” of microbiota analysis is differential abundance testing
- Many methods (even much pitfalls) are available



RecyclesEU

recycleseu

Recycles EU

RECYCLES WORKSHOP

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

5th, 6th & 7th of July 2022

Thank you!

Francesco Vitali

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

@svito_fi

FrancescoVit

GA: 872053 — H2020 - MSCA - RISE-2019

European Commission