

The next level of abstraction: a meta-metagenomic analysis

Aleksandra Perz
Research Trainee at Wren Lab
Oklahoma Medical Research Foundation

Metagenomics

- Genetic material taken directly from a sample
- Is a better representation of taxonomic diversity than cultured samples
- 16S rRNA gene used for phylogenetic classification

Browse projects

By selected biomes



Soil (136)



Marine (41)



Forest (16)



Non-human
host (61)



Engineered
(21)



Freshwater
(15)



Grassland
(10)



Human gut
(30)



Air (1)

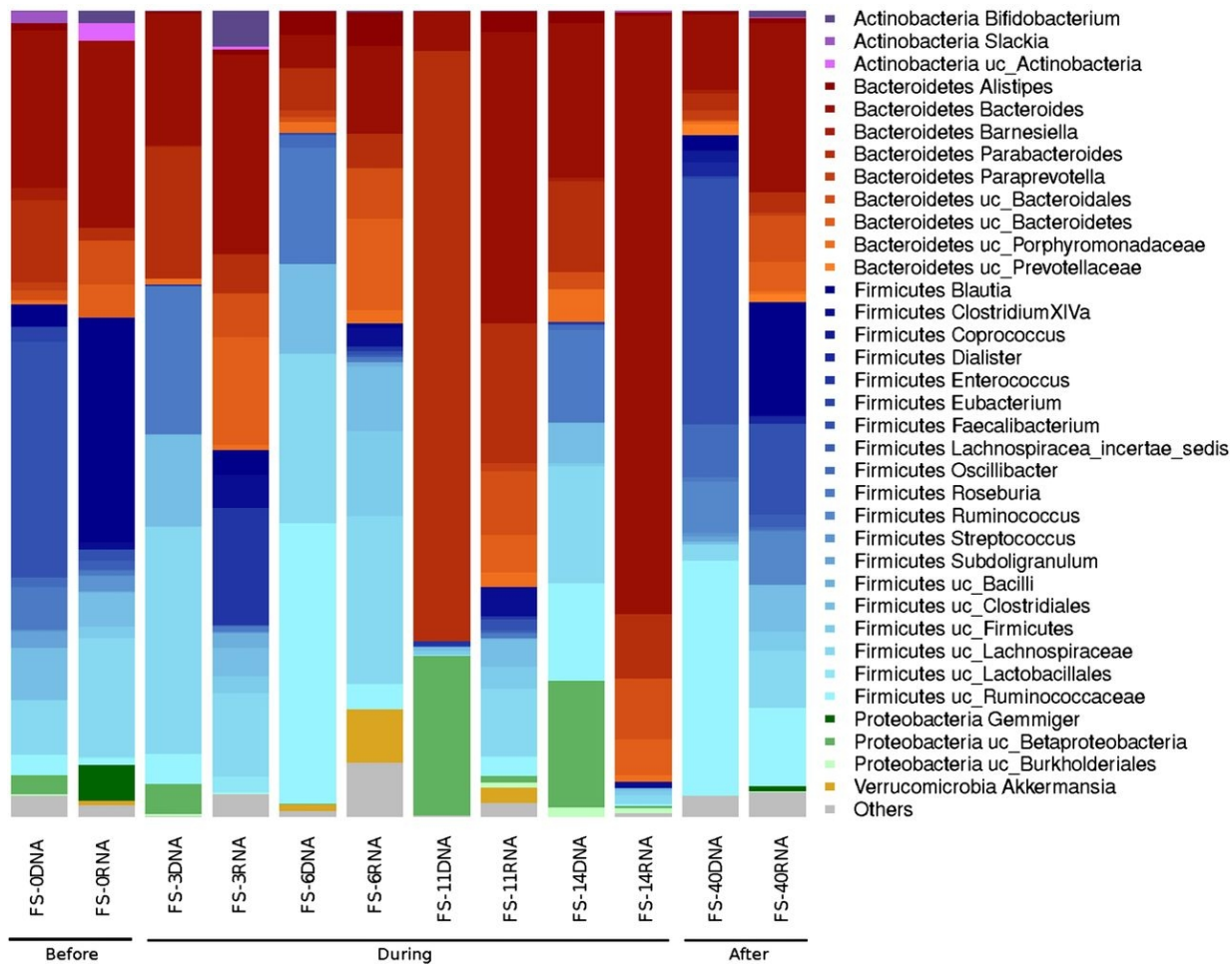


Wastewater
(3)

[View all biomes](#)

<https://www.ebi.ac.uk/metagenomics/>

Inspiration



Gut bacterial composition.
FS-0, FS-3, FS-6, FS-11, FS-14:
respective days of antibiotic
treatment.
FS-40: 40 days after end of
treatment.

- early reduction in Gram-negative organisms (day 6)
- overall collapse in diversity

Gut microbiota disturbance
during antibiotic therapy: a
multi-omic approach (Pérez-
Cobas et al., Gut, 2013)

There is knowledge out there

Bring in the already
available knowledge
to interpret a new
finding

Find samples that are similar:
learn something new about
the finding based on the literature

Compare shifts observed
between conditions:
find samples that exhibit similar
changes in taxonomic abundance

Predict characteristics
of the finding:
confirm the finding

The data



METADATA

contains
information
about the
sample and the
studies

TAXA
ABUNDANCE
DATA

a representation
of how many of
each taxonomic
unit there is in
each sample

Obtained from European Bioinformatics Institute

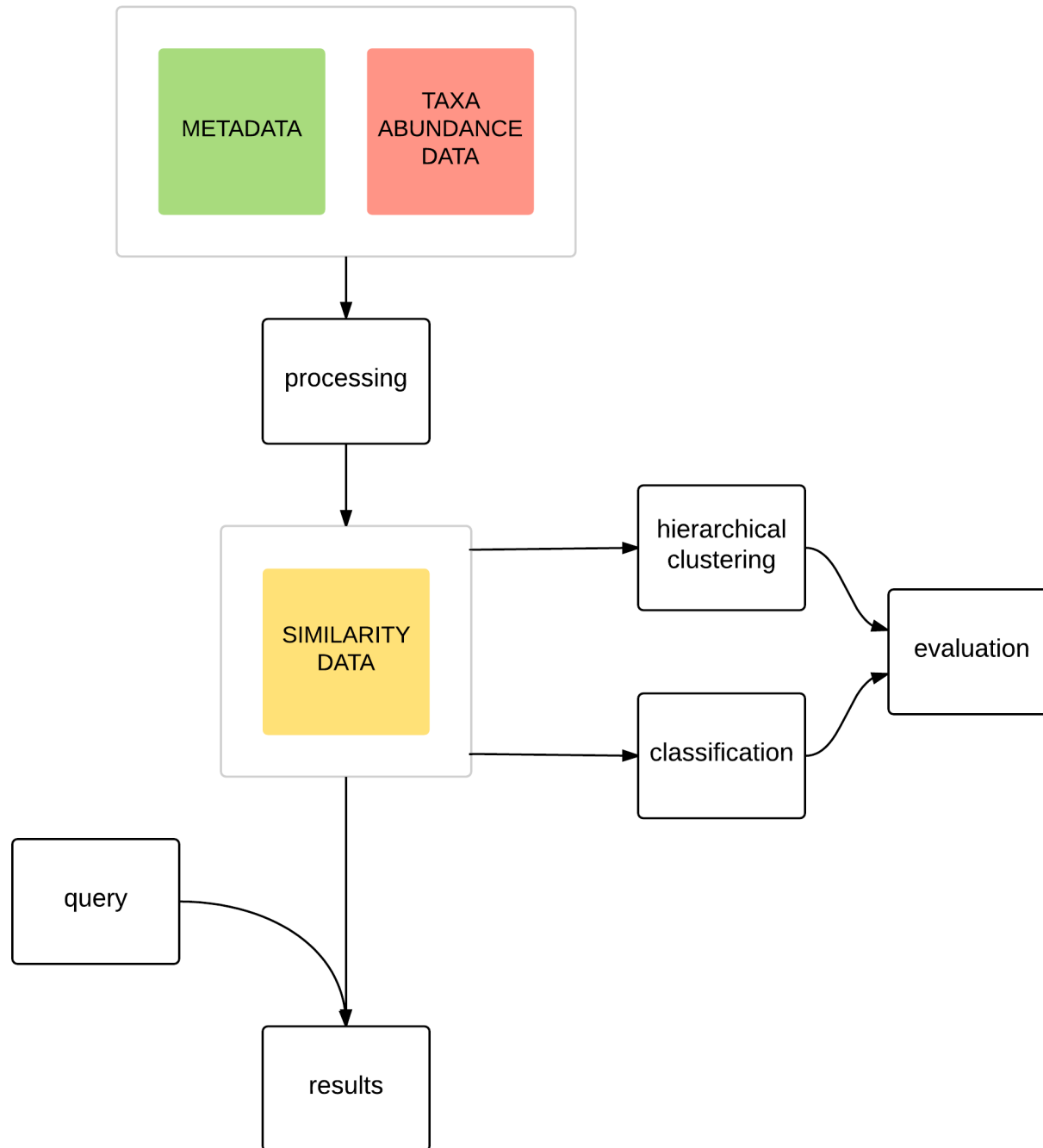
Metadata preparation

Study ID	ERP001506
Study title	Beta Lactam Antibiotics and Human Gut Microbiota
Sample ID	ERS149134
Sample classification	Host-associated:Human > Digestive system > Large intestine > Fecal
Sample description	Total DNA was extracted from human faecal samples using QIAamp® DNA Stool Kit (Quiagen) following the manufacturer's instructions
Sample name	Human gut metagenome, 40 days after antibiotics
Run IDs	ERR139700
Keywords	gut intestine faecal fecal human

- Erroneous or incomplete metadata:
automated extraction of labels
developed
- Keywords extracted
(e.g. "human", "soil", "diabetic")

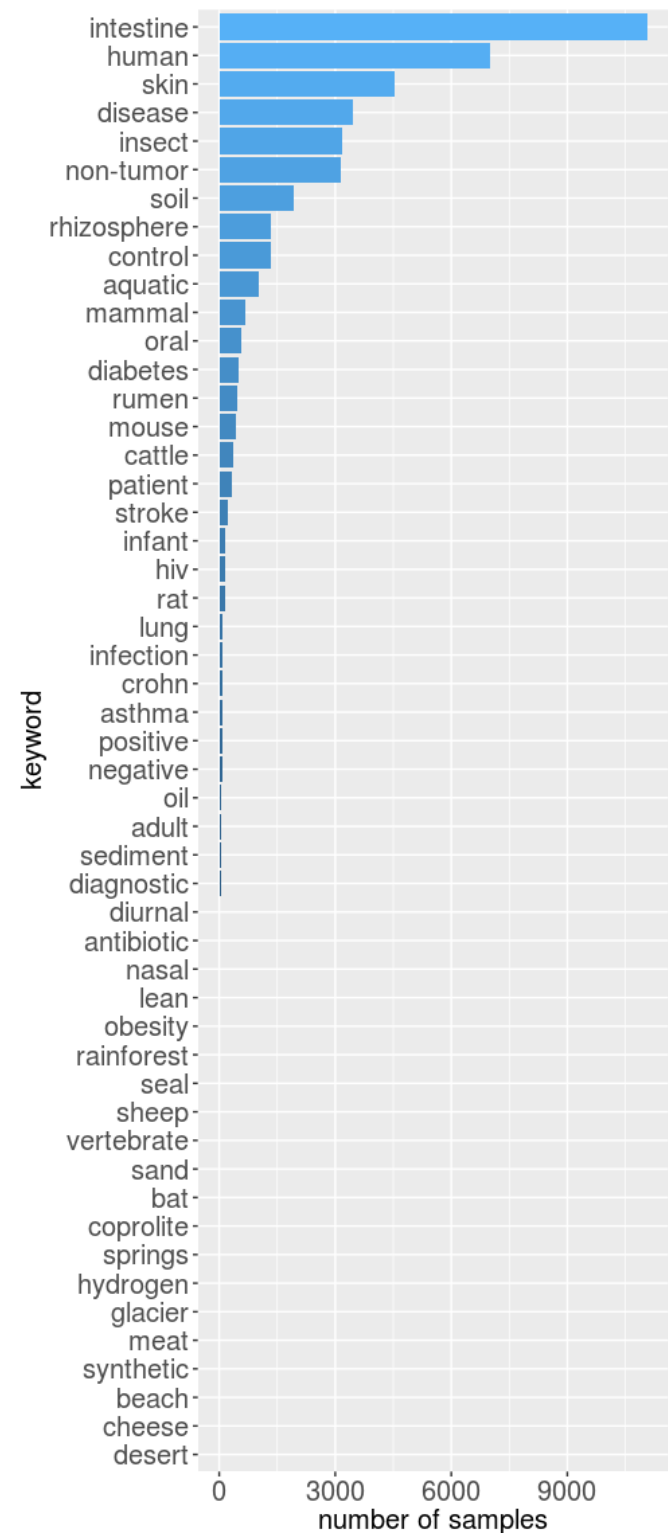


METADATA



Classification

- How well the algorithm predicts sample labels based on the taxonomic abundance?
- Which of the labels are well predicted, which are not?
- ML approach
- k Nearest Neighbors
- Cross-validated



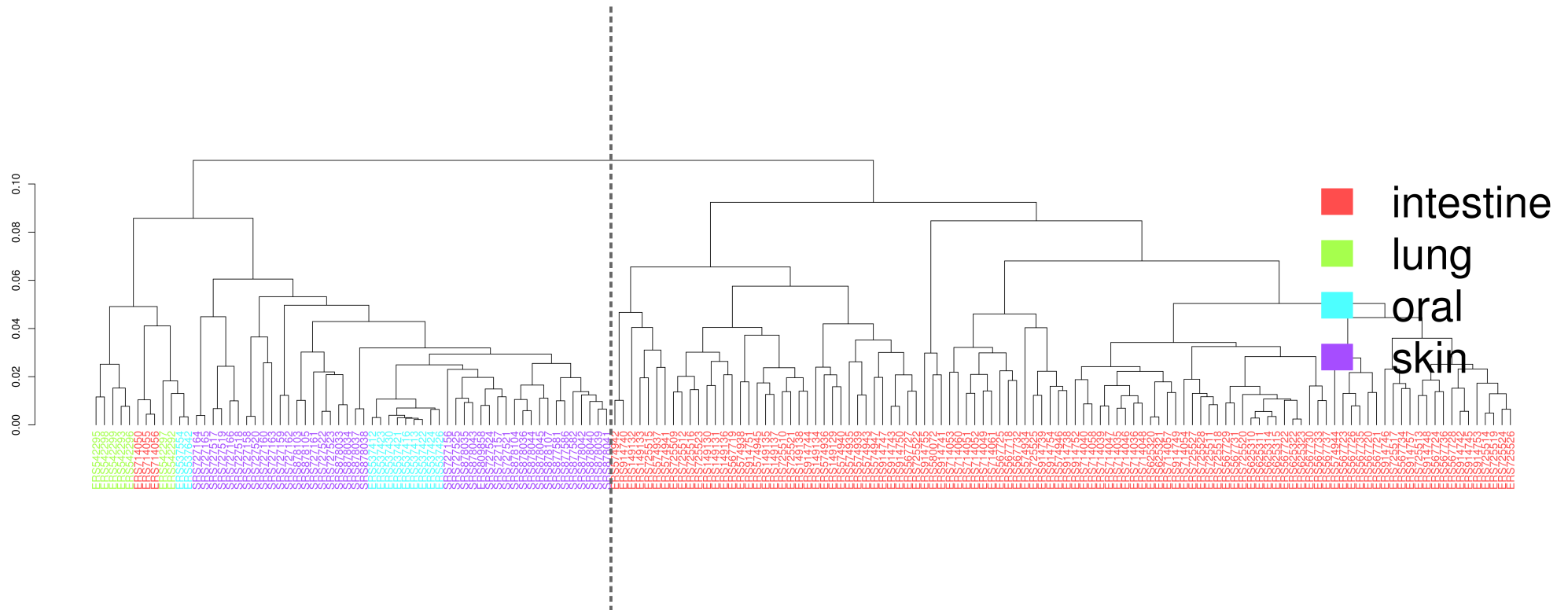
f1.score	samples belonging to class	class name
0.94	9398	intestine
0.93	342	mouse
0.91	212	cattle
0.86	1659	soil
0.83	4487	skin
0.81	5250	human
0.19	592	aquatic
...
0	72	positive
0	74	rat
0	1258	rhizosphere
0	14	seal
0	18	sheep
0	233	patient
0	12	vertebrate

Good predictions

Useless predictions
or not predicted at all

Hierarchical clustering

Cluster samples based on a similarity metric




Example with some REAL data

Sample_id	control	intestine	human
ERS149129	0	0.2	0.2
ERS149130	0	1	1
ERS149131	0	1	1
ERS149132	0	1	1
ERS149133	0	1	1
ERS149134	0	0.8	0.8
ERS149135	0	1	0.8
ERS149136	0	1	1
ERS149137	0	1	1
ERS149138	0	1	1
ERS149139	0	1	1
ERS149140	0.8	1	1

Probabilities that
the sample belongs
to the class

Hierarchical clustering

query



Sample_id	Study_title	Sample_description	Sample_name
ERS149132	Beta Lactam Antibiotics and Human Gut Microbiota	Total DNA was extracted from human faecal samples (...)	Human gut metagenome, 11 days after antibiotics
ERS574934	Changes imposed by Clostridium difficile infection on the human gut microbiome	16S rRNA	16S rRNA
ERS915363	American Gut Project	American Gut Project Stool sample	10317

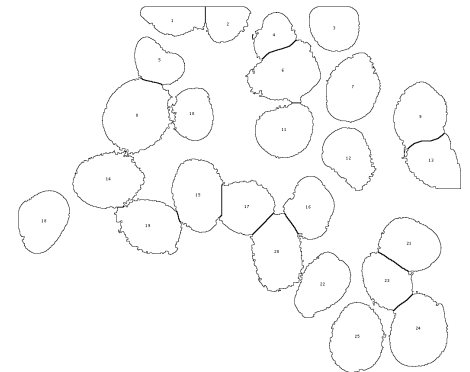
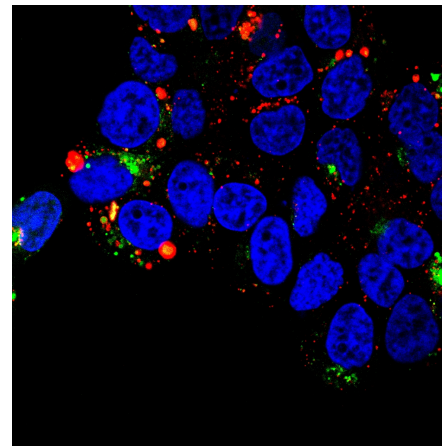
Other things I've been doing

GAMMA-seq

- Functional prediction for lncRNAs
- Data: expression levels for transcripts

Automated confocal microscope image analysis

- Measure intensity of each dye, colocalization, count cells, etc.



Jonathan Wren

Constantin Georgescu

Cory Giles

Christopher Reighard

Xiavan Roopnarinesingh

Thank them.
They are quite awesome.

Thank you for attention.

Supplementary material

Gut microbiota disturbance during antibiotic therapy: a multi-omic approach

Ana Elena Pérez-Cobas, María José Gosalbes, Anette Friedrichs, Henrik Knecht,
Alejandro Artacho

2013 Gut 62(11), 1591-1601

EBI



Unique studies: 180

Unique samples: 17269

Human samples: 12331

American gut: 8037

Classes: 152

- https://github.com/ProteinsWebTeam/ebi-metagenomics/blob/master/memi/memi-web/developers/python/mgportal_bulk_download.py

Confocal image analysis

cell_line	treatment	cell_count	red_area_fraction	red_mean	red_stdev	yellow_area_fraction	yellow_mean_per_cell	green_in_nuclei_area_fraction	yellow_area_um
cell line 1	control	8.4	0.00	0.12	3.37	0.00	0.01	0.03	34.60
cell line 2	treatment 1	19.8	0.06	4.33	26.09	0.01	0.06	0.02	180.30
cell line 1	treatment 2	13.8	0.02	1.08	12.93	0.00	0.03	0.00	69.93
cell line 2	treatment 1	20	0.15	12.10	44.16	0.07	0.51	0.01	1199.43