Microbiota **D**ata **A**nalysis **Workshop**

22 - 24 January, 2020

https://www.gdc-docs.ethz.ch/MDA/site/

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

GDC Genetic Diversity Centre Zurich

Welcome

Genetic Diversity Centre Zurich

Genetic Diversity Centre
Zurich

http://www.gdc.ethz.ch/


Aria


Jean-Claude

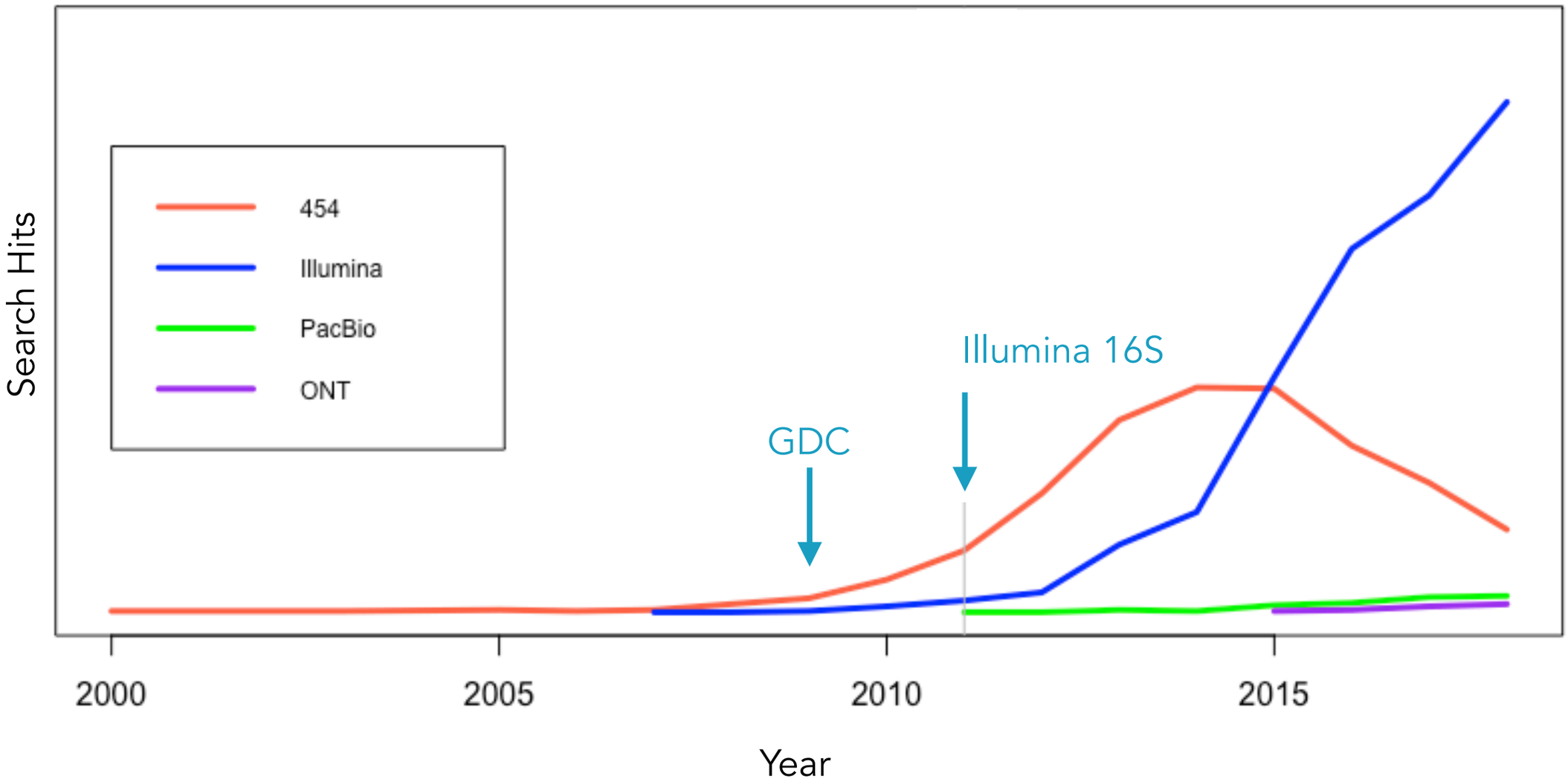

Silvia

Nik

The **Genetic Diversity Centre (GDC)** is a knowledge and technology platform of the D-USYS Department. The team GDC provides scientific and technical support for research related to genetic and genomic diversity in a wide range of non-model organisms.

Number of Publications on PubMed

(Search term: <SeqTech> and 16S)



Announcements

GDC Genetic Diversity Centre Zurich

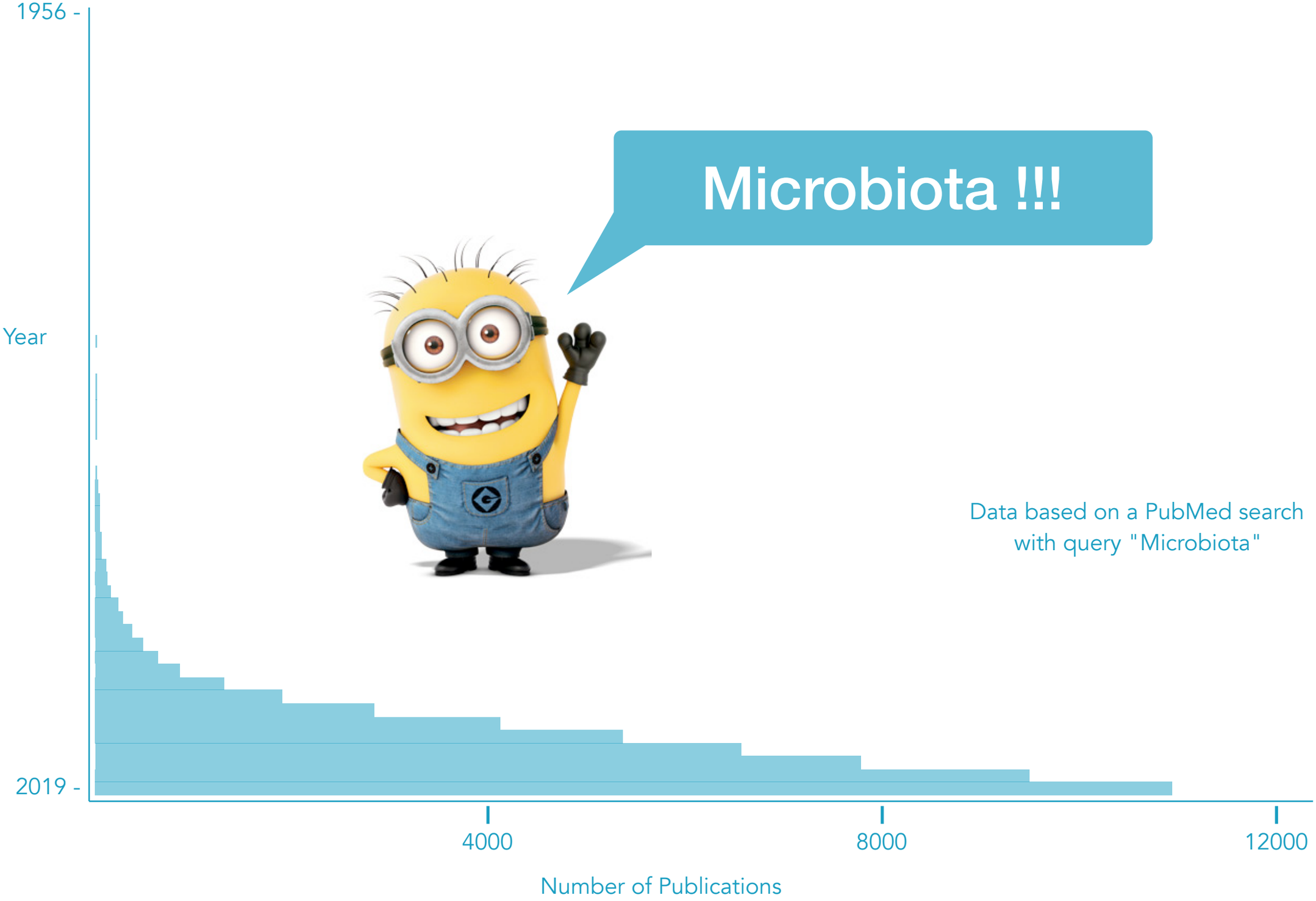
- ☞ Lecture room: key, food and drinks
- ☞ Bathrooms
- ☞ Coffee / tea breaks
- ☞ Lunch
- ☞ Social dinner - Thursday evening > Doodle

Genetic Diversity Centre Zurich

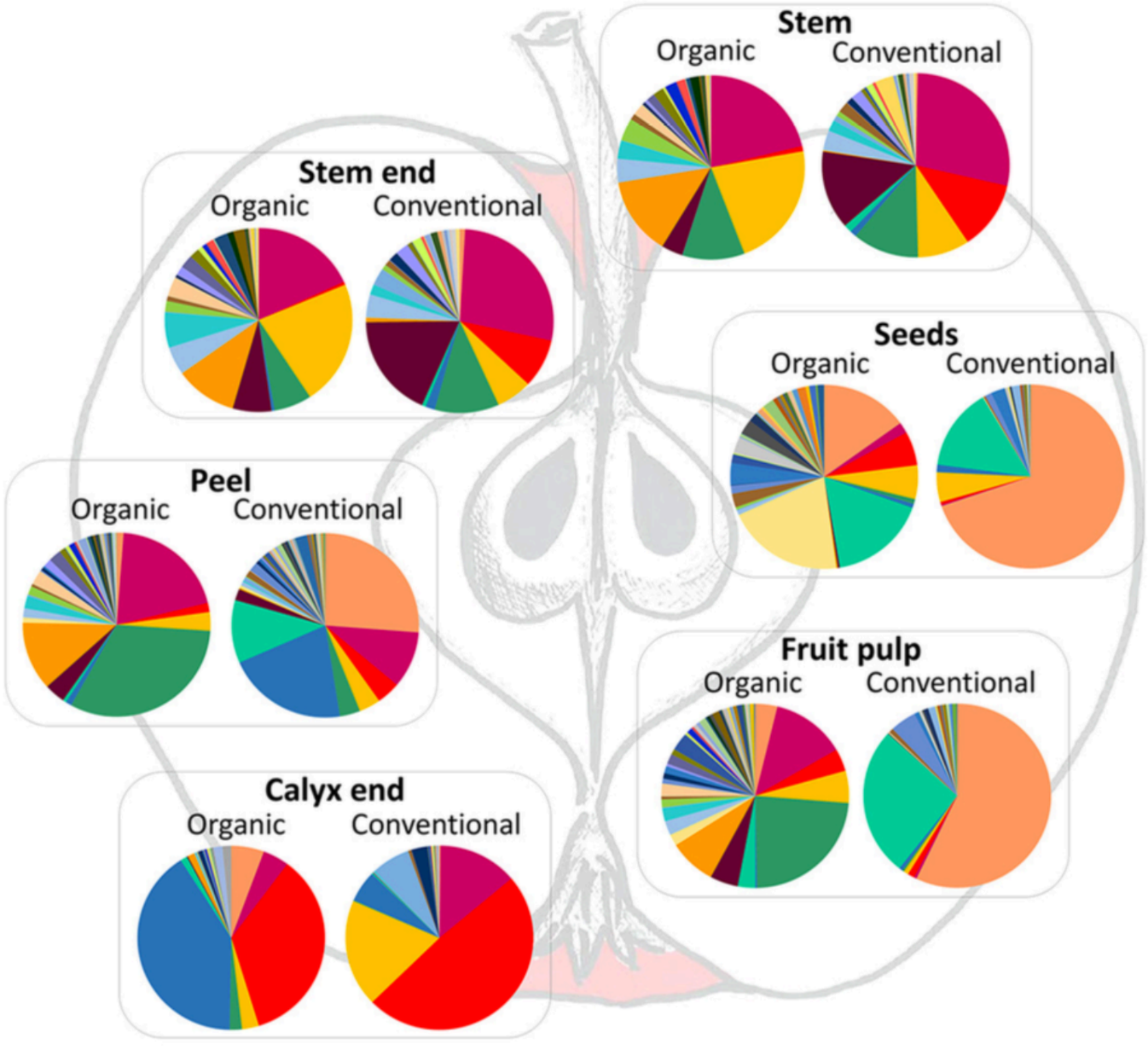
- ☞ Power plugs
- ☞ Wi-Fi connection (eduroam)
- ☞ Access to Google drive
- ☞ R libraries
- ☞ Question(s)

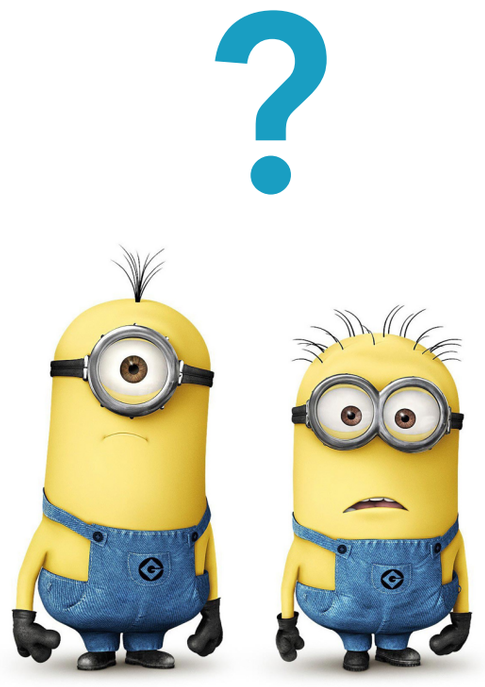
Overview

GDC Genetic Diversity Centre Zurich



Data based on a PubMed search with query "Microbiota"





Wassermann et al. (2019) An Apple a Day: Which Bacteria Do We Eat With Organic and Conventional Apples? Frontiers in Microbiology. Volume 10 | Article 1629.

MDA ▷ Introduction

GDC Genetic Diversity Centre Zurich

Wednesday 22.01.20	
Time	Subject
8:00 - 9:00	Registration & Installation
9:00-9:30	Introduction
9:30-10:30	Warm Up
10:30 - 11:00	Tea / Coffee Break
11:00-11:45	Data Preparation
11:45-12:30	eDNA
12:30 - 13:30	Lunch
13:30-15:00	Reproducible Science
15:00 - 15:30	Tea / Coffee Break
15:30-16:15	Reality Check
16:15-17:00	Q&A
17:00	End

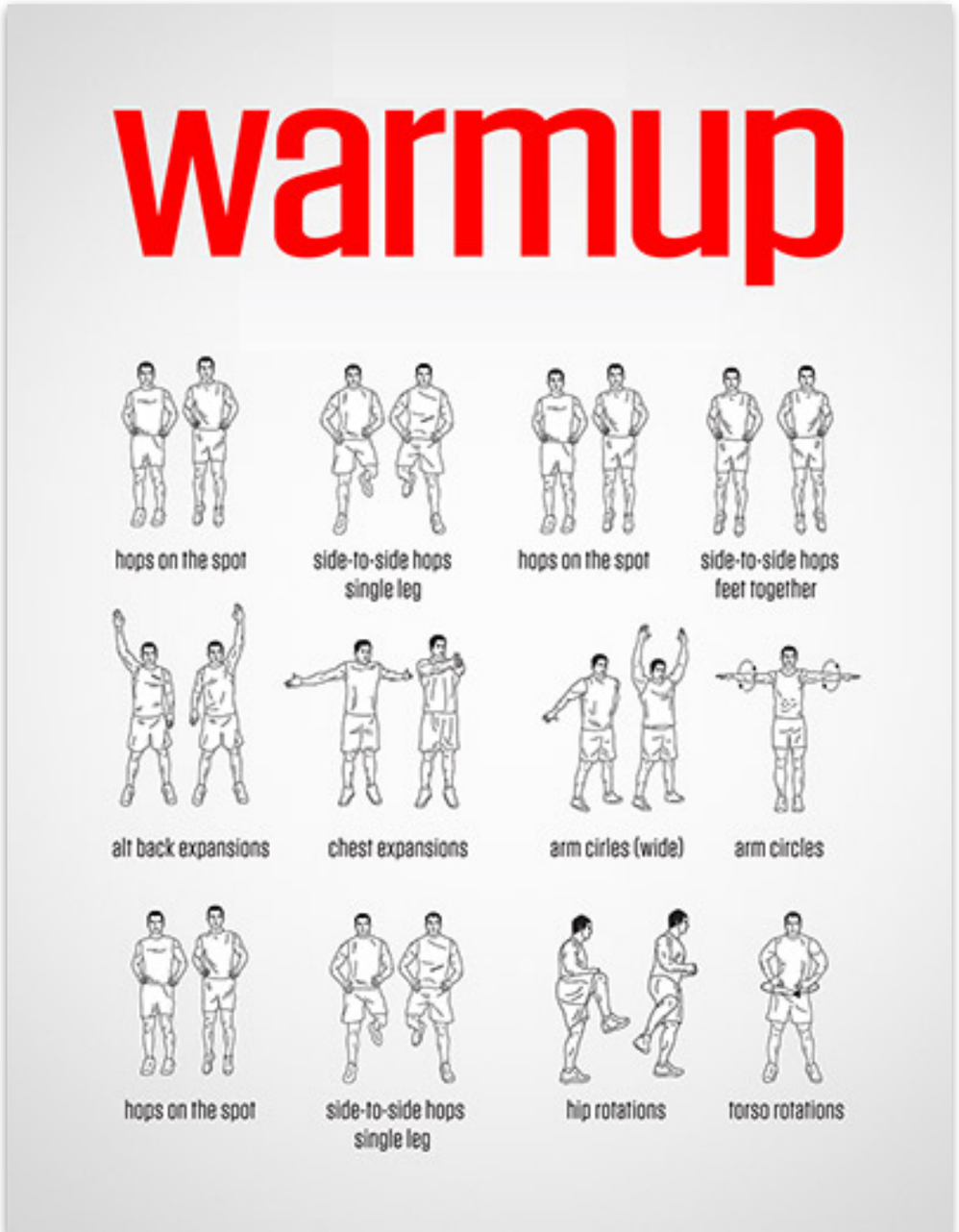
Thursday 23.01.20	
Time	Subject
8:00 - 9:00	Installation
9:00-10:30	Diversity Analysis
10:30 - 11:00	Tea / Coffee Break
11:00-12:30	Diversity Analysis
12:30 - 13:30	Lunch
13:30-15:00	Diversity Analysis
15:00 - 15:30	Tea / Coffee Break
15:30-17:00	Diversity Analysis
17:00	End

Friday 24.01.20	
Time	Subject
8:00 - 9:00	Installation
9:00-10:30	Microbial Network
10:30 - 11:00	Tea / Coffee Break
11:00-12:30	Microbial Network
12:30 - 13:30	Lunch
13:30-15:00	Data Submission
15:00 - 15:30	Tea / Coffee Break
15:30-17:00	Random Forest
17:00	End

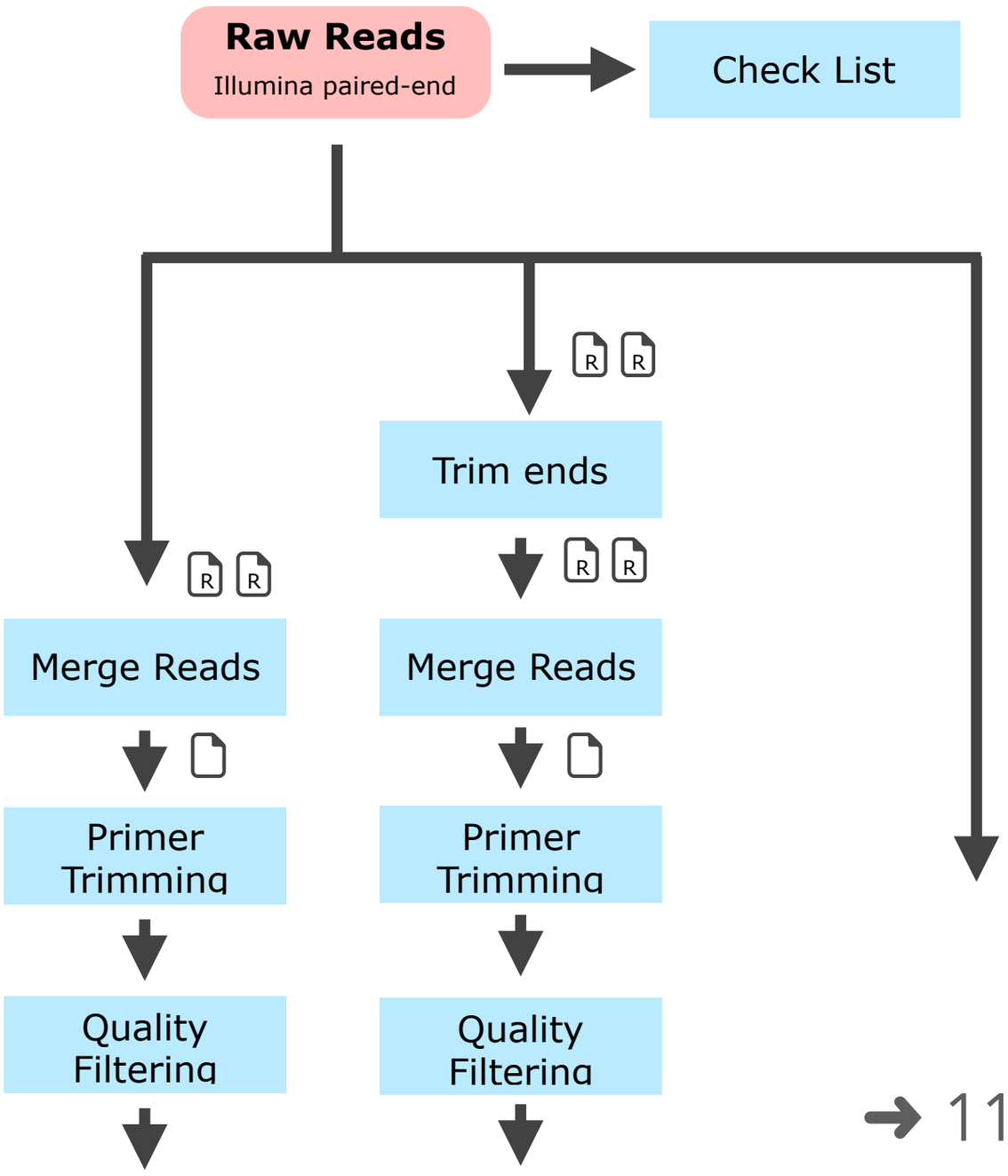
Mahendra Mariadassou (INRA)
Robert Edgar (drive5)
Sam Holt (EMBL-EBI)
Klaus Schläppi (University of Bern)
Kristy Deiner (ETH Zürich)
Jean-Claude Walser (GDC)
Nik Zemp (GDC)

Dinner*

* Dinner with the speakers but everybody is welcome to join.



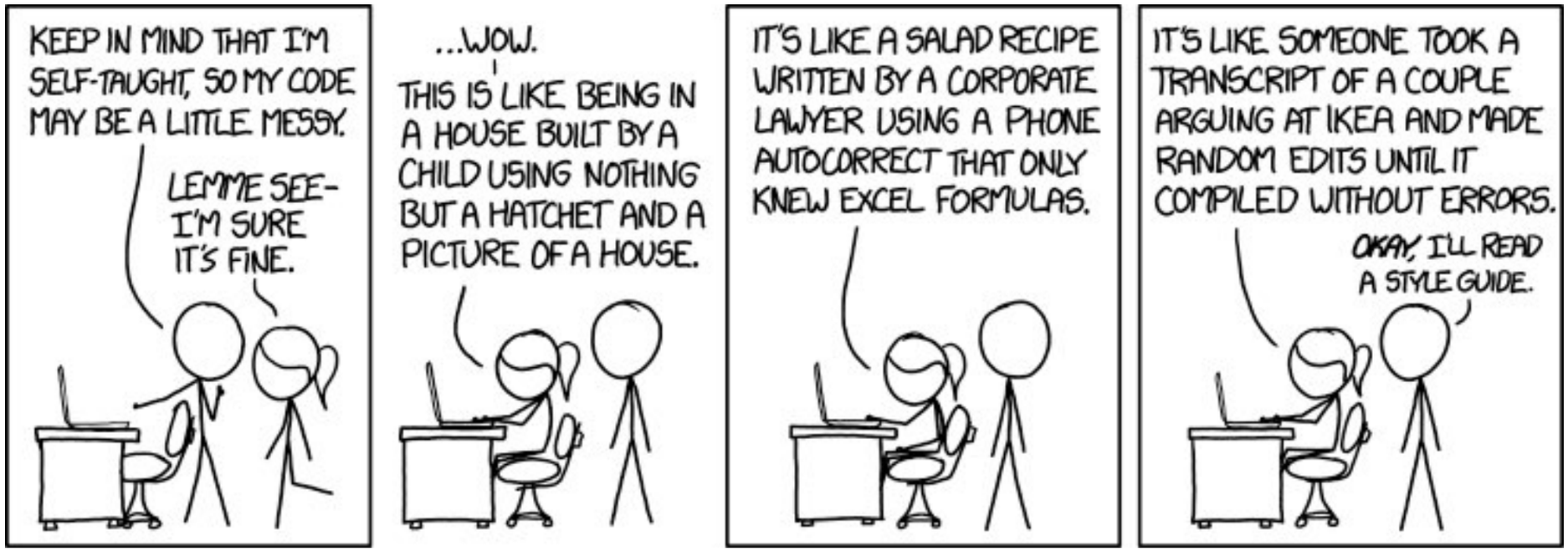
➜ 9:30-10:30 Jean-Claude Walser



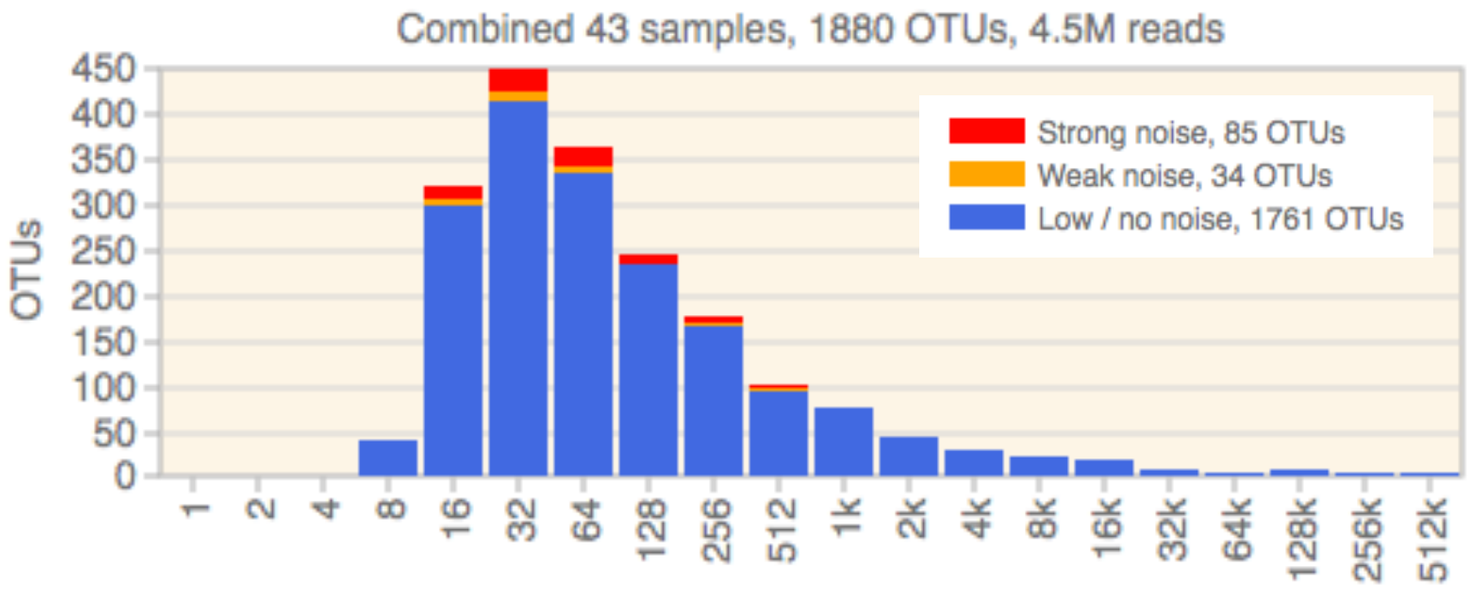
➜ 11:00-11:45 Jean-Claude Walser

➜ 11:45 - 12:30 Kristy Deiner

Reproducible Science



➜ 13:30 - 15:00 Nik Zemp



➜ 15:30 - 16:15 Robert Edgar

➜ 16:15 - 17:00 Robert Edgar

GDC Genetic Diversity Centre Zurich

Count / OTU Table

	Sample1	Sample2	Sample3	Sample4	Sample5	Sample6
OUT1	2519	2354	1074	2	452	1233
OUT2	2520	1158	1287	1	34	3184
OUT3	106	4	2	0	0	0
OUT4	490	82	148	0	0	22
OUT5	120	73	111	0	0	133
OUT6	13	1	6	0	0	0
OUT7	9	0	0	0	0	0
OUT8	813	415	142	0	0	808
OUT9	45	2	7	0	0	0

GDC Genetic Diversity Centre Zurich

Microbiome data are …

(a) **compositional** (multiple parts of nonnegative numbers).

(b) **high dimensional** (few data points and many features)

and **underdetermined** (the number of OTUs much greater than the number of samples).

(c) **overdispersed** (variance of the counts of read is larger than expected).

(d) often spares with **many zeros** (zero-inflated).

Zurich Centre Genetic Diversity

Metadata / Mapfile

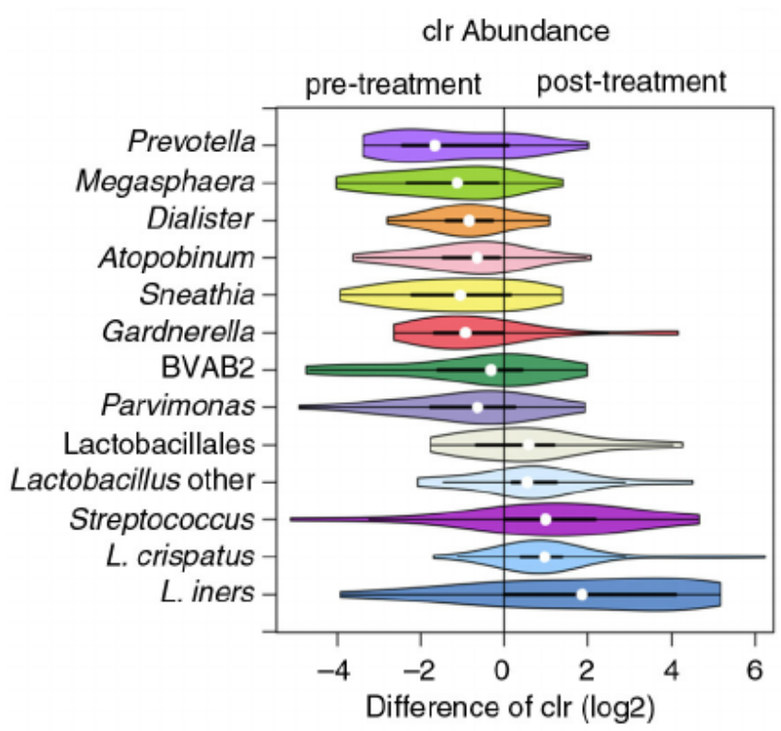
	pH	TP	NO2	NO3	NH4	Shannon
Sample1	8.3	13.3	17.7	1.6	67.4	3.5
Sample2	8.2	13.3	17.8	1.6	67.9	3.3
Sample3	8.2	13.2	17.7	1.6	68.6	3.3
Sample4	8.2	14.3	17.3	1.5	75.2	3.4
Sample5	9.1	14.2	16.6	1.5	75.1	3.5
Sample6	9.1	14.1	17.5	1.6	94.2	3.4
Sample7	9.1	15.3	22.3	1.5	11.8	3.6
Sample8	7.9	14.3	22.3	1.4	11.3	3.6
Sample9	7.9	16.0	22.4	1.4	9.3	3.7

GDC Genetic Diversity Centre Zurich

Thursday, 23.01.2020

	Sample1	Sample2	Sample3	Sample4	Sample5	Sample6
OUT1	2519	2354	1074	2	452	1233
OUT2	2520	1158	1287	1	34	3184
OUT3	106	4	2	0	0	0
OUT4	490	82	148	0	0	22
OUT5	120	73	111	0	0	133
OUT6	13	1	6	0	0	0
OUT7	9	0	0	0	0	0
OUT8	813	415	142	0	0	808
OUT9	45	2	7	0	0	0

	pH	TP	NO2	NO3	NH4	Shannon
Sample1	8.3	13.3	17.7	1.6	67.4	3.5
Sample2	8.2	13.3	17.8	1.6	67.9	3.3
Sample3	8.2	13.2	17.7	1.6	68.6	3.3
Sample4	8.2	14.3	17.3	1.5	75.2	3.4
Sample5	9.1	14.2	16.6	1.5	75.1	3.5
Sample6	9.1	14.1	17.5	1.6	94.2	3.4
Sample7	9.1	15.3	22.3	1.5	11.8	3.6
Sample8	7.9	14.3	22.3	1.4	11.3	3.6
Sample9	7.9	16.0	22.4	1.4	9.3	3.7



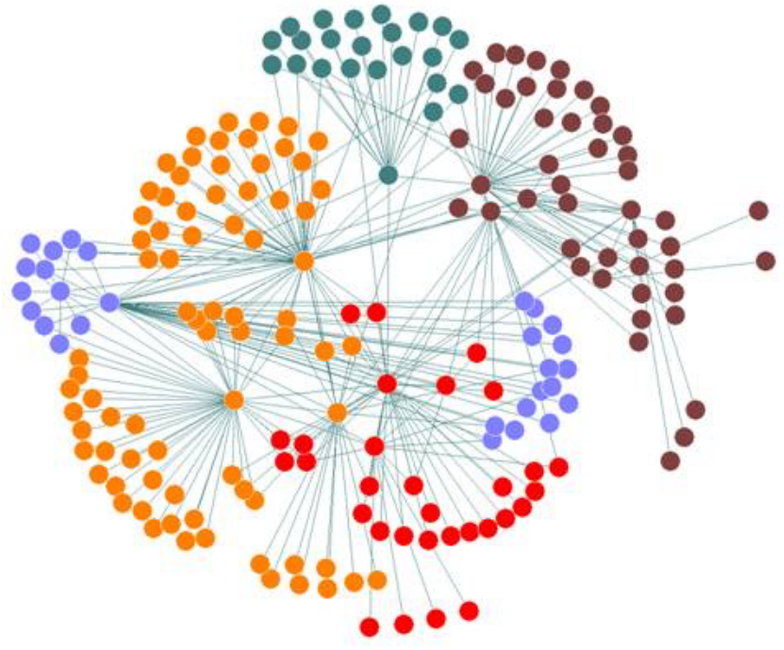
➜ Tomorrow: 9.00 - 17.00 Mahendra Mariadassou

Genetic Diversity Centre Zurich

Friday, 24.01.2020

	Sample1	Sample2	Sample3	Sample4	Sample5	Sample6
OUT1	2519	2354	1074	2	452	1233
OUT2	2520	1158	1287	1	34	3184
OUT3	106	4	2	0	0	0
OUT4	490	82	148	0	0	22
OUT5	120	73	111	0	0	133
OUT6	13	1	6	0	0	0
OUT7	9	0	0	0	0	0
OUT8	813	415	142	0	0	808
OUT9	45	2	7	0	0	0

	pH	TP	NO2	NO3	NH4	Shannon
Sample1	8.3	13.3	17.7	1.6	67.4	3.5
Sample2	8.2	13.3	17.8	1.6	67.9	3.3
Sample3	8.2	13.2	17.7	1.6	68.6	3.3
Sample4	8.2	14.3	17.3	1.5	75.2	3.4
Sample5	9.1	14.2	16.6	1.5	75.1	3.5
Sample6	9.1	14.1	17.5	1.6	94.2	3.4
Sample7	9.1	15.3	22.3	1.5	11.8	3.6
Sample8	7.9	14.3	22.3	1.4	11.3	3.6
Sample9	7.9	16.0	22.4	1.4	9.3	3.7



➜ Friday 9:00 - 10:30 Jean-Claude Walser

➜ Friday 11:00 - 12:30 Klaus Schlaeppi

Friday, 24.01.2020

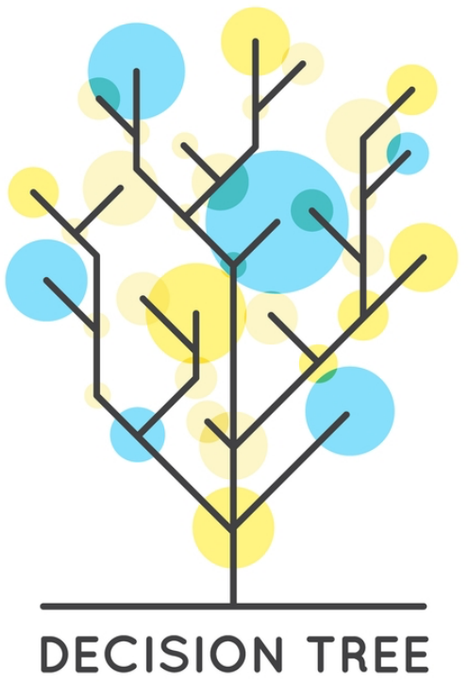
➜ Friday 13:30 - 15:00 Sam Holt

Genetic Diversity Centre Zurich

Friday, 24.01.2020

	Sample1	Sample2	Sample3	Sample4	Sample5	Sample6
OUT1	2519	2354	1074	2	452	1233
OUT2	2520	1158	1287	1	34	3184
OUT3	106	4	2	0	0	0
OUT4	490	82	148	0	0	22
OUT5	120	73	111	0	0	133
OUT6	13	1	6	0	0	0
OUT7	9	0	0	0	0	0
OUT8	813	415	142	0	0	808
OUT9	45	2	7	0	0	0

	pH	TP	NO2	NO3	NH4	Shannon
Sample1	8.3	13.3	17.7	1.6	67.4	3.5
Sample2	8.2	13.3	17.8	1.6	67.9	3.3
Sample3	8.2	13.2	17.7	1.6	68.6	3.3
Sample4	8.2	14.3	17.3	1.5	75.2	3.4
Sample5	9.1	14.2	16.6	1.5	75.1	3.5
Sample6	9.1	14.1	17.5	1.6	94.2	3.4
Sample7	9.1	15.3	22.3	1.5	11.8	3.6
Sample8	7.9	14.3	22.3	1.4	11.3	3.6
Sample9	7.9	16.0	22.4	1.4	9.3	3.7



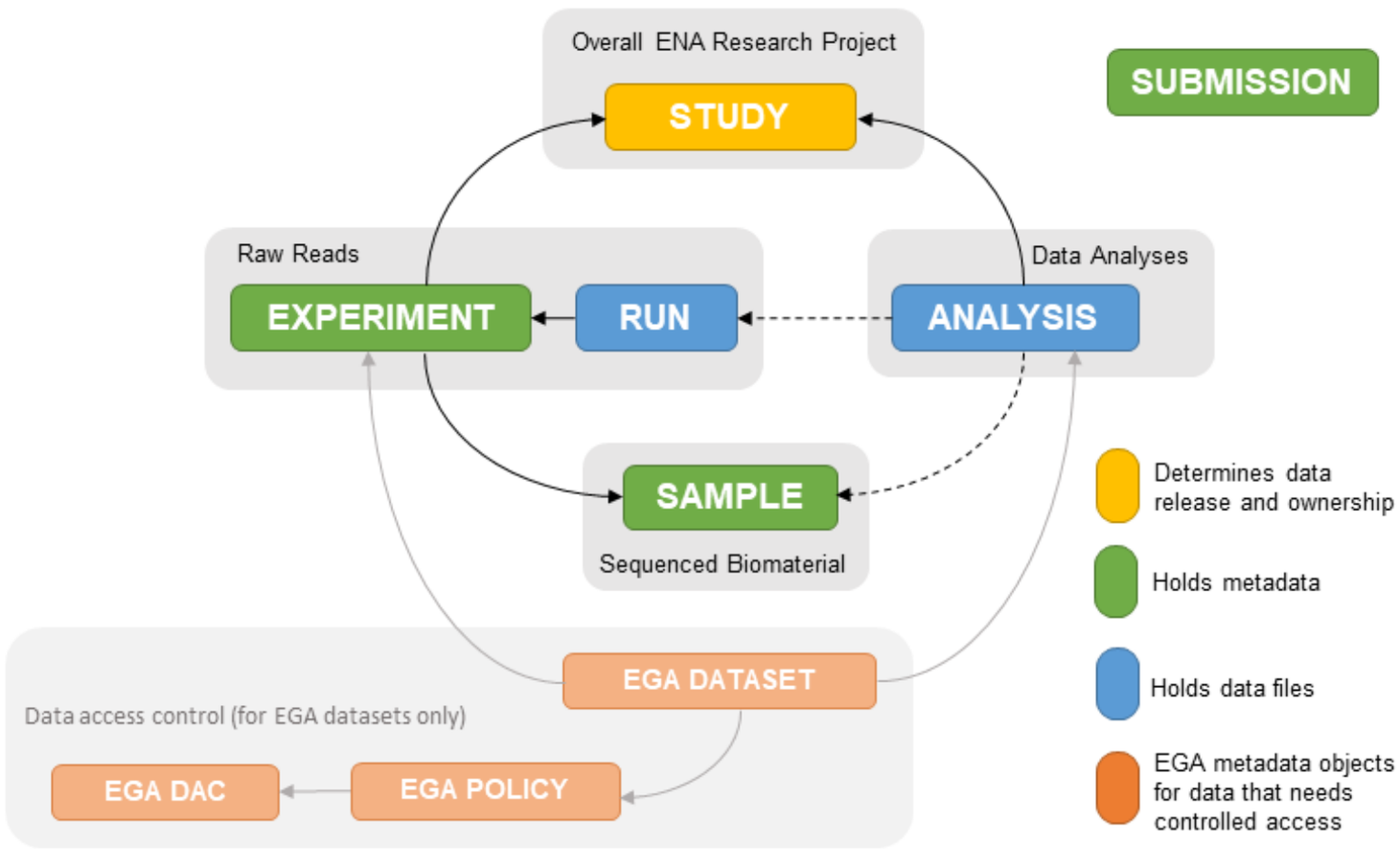
➜ Friday 15:30 - 17:00 Jean-Claude Walser

Introduction

Cluster-Game

ENA
European Nucleotide Archive

European Nucleotide Archive (ENA) accepts sequence reads and associated analyses. Once public, data submitted to ENA is exchanged between International Nucleotide Sequence Database Collaboration (INSDC) partners: NCBI and DDBJ.



Genetic Diversity Centre Zurich

MIxS – Minimum information about any (x) sequence

MIxP – Minimum information about any (x) participants

***The **MIxS** is a unified standard developed by the Genomic Standards Consortium (GSC) for reporting of minimum information about any (x) nucleotide sequence. It consists of MIGS (Minimum Information about a Genome Sequence), MIMS (Minimum Information about a Metagenome Sequence) and **MIMARKS** (Minimum Information about a MARKer gene Sequence) standards and describes fourteen environments.

Questions:

1. What **data type** are you working with?
2. From which biome(s) do your samples come from?
3. What species-diversity are you interested in?
4. What sequencing platform/technology are you using?
5. What is the read / sequence length?
6. Do you prepare the library yourself or do you outsource it?

Answers:

1. Amplicon ⇨ 16S
2. Sediment ⇨ Deep sea sediment
3. Bacteria ⇨ Cyanobacteria
4. Illumina ⇨ MiSeq
5. paired-end 2 x 300
6. NO (of course not)

GDC Genetic Diversity Centre Zurich

1 Data Type

Amplicon

Assemblies

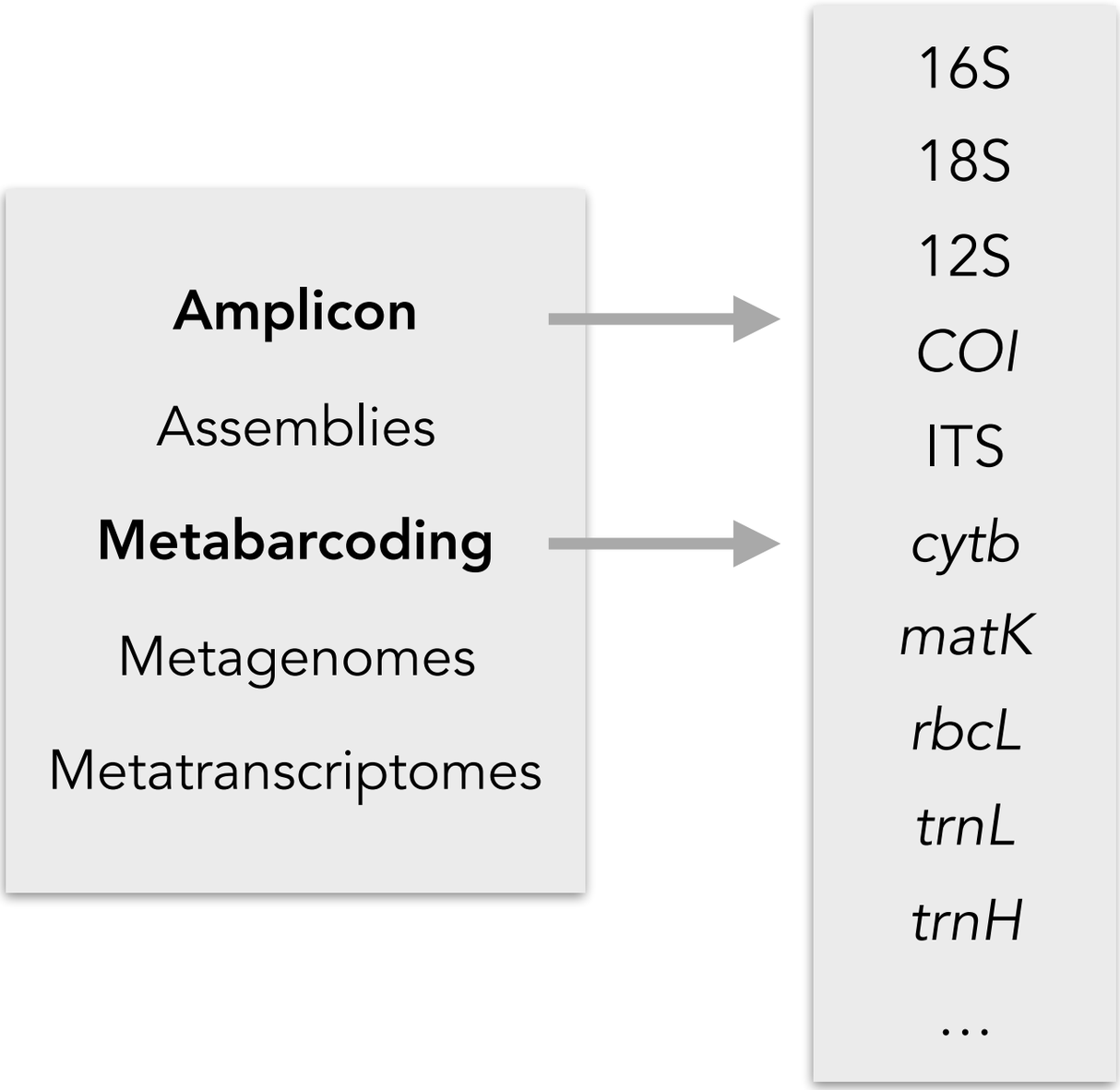
Metabarcoding

Metagenomes

Metatranscriptomes

GDC Genetic Diversity Centre Zurich

1b Data **Sub**-Types



GDC Genetic Diversity Centre Zurich

2 Environment-specific Descriptors

(Genomics Standards Consortium (GSC) MixS Checklists)

Air

Host-associated

Human-associated

Microbial mat/biofilm

Plant-associated

Sediment

Soil

Wastewater / Sludge

Water

GDC Genetic Diversity Centre Zurich

2b Environment-specific **Sub**-Descriptors

(Genomics Standards Consortium (GSC) MixS Checklists)

Air

Host-associated

Human-associated

Microbial mat/biofilm

Plant-associated

Sediment

Soil

Wastewater / Sludge

Water

Air ⇨ AC filter system

Host ⇨ daphnia

Human ⇨ baby gut

Biofilm ⇨ rubber ducks

Plant ⇨ leave

Sediment ⇨ deep sea sediment

Soil ⇨ farm land

Sludge ⇨ WWTPs

Water ⇨ river systems

GDC Genetic Diversity Centre Zurich

3 Target Species

Bacteria

Archaea

Fungi

Plants

Eukaryote

...

4 NGS Data Formats

Illumina

Roche LS454

PacBio

IonTorrent

Oxford Nanopore

GDC Genetic Diversity Centre Zurich

4b NGS Data Formats - Read Length

Illumina

LS454

PacBio

IonTorrent

Oxford Nanopore

→

MiSeq PE300

MiSeq PE250

NextSeq PE150

HiSeq 2500 PE125

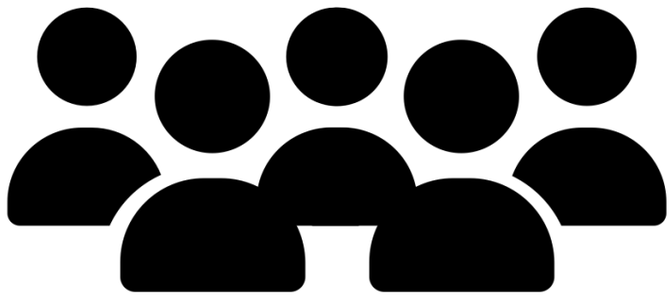
…

Genetic Diversity Centre Zurich

5 Library Prep

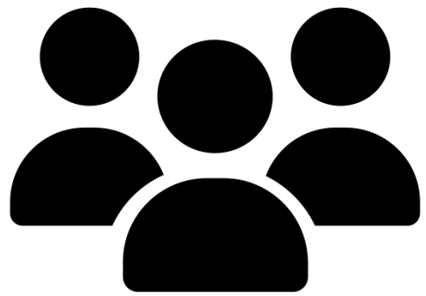
Yes / No

Genetic Diversity Centre Zurich



Answers:

1. **Amplicon** ⇨ 16S



Answers:

1. Metagenomics

Lets Cluster!