WG2: Benchmark datasets & DREAM Challenge

1) Deliverables (& Milestone)

D2.1-2.2: Benchmark data, documentation & Web-portal

D2.3-4: Two publications/reports associated to the DREAM Challenge

DREAM Challenge announced (Y2Q2) and closed (Y2Q4)

2) August 2022 MC & WG meeting

Turku, Finland

3) Support for the other WGs

R/Bioconductor Multi-omic data science framework

MultiAssayExperiment / TreeSummarizedExperiment

HUMAN GUT MICROBIOME ATLAS

microbiomeatlas.org


Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Global and temporal state of the human gut microbiome in health and disease

Saeed Shoaie (saeed.shoaie@kcl.ac.uk)
Centre for Host-Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London https://orcid.org/0000-0001-5834-4533

WG2 data progress

Benchmarking data:

- Preprint on HGMA is out
- Continuing to process 5000 samples from the human gut health index metastudy: taxonomic information, functional genes, enzymatic reactions, metabolic pathways and predicted metabolites.
- Additional work on a coherent high-quality CRC data subset.

Processing:

- Metagenome processing:
  Biobakery Metaphlan,
  Humann & Melonnpan

- 16S processing:
  taxonomic profiles
  functional genes
  enzymatic reactions
  metabolic pathways

Processing using Metagenome ATLAS and various ML methods

Baptiste Avot & Laurent Falquet

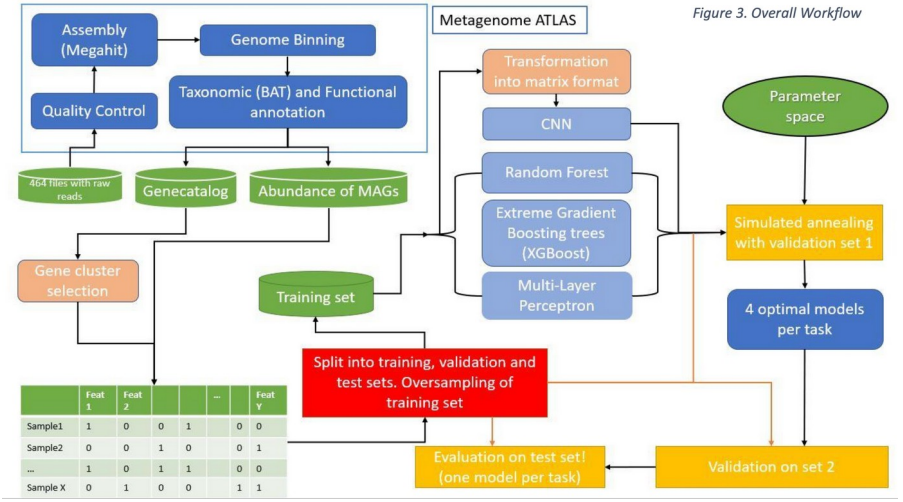
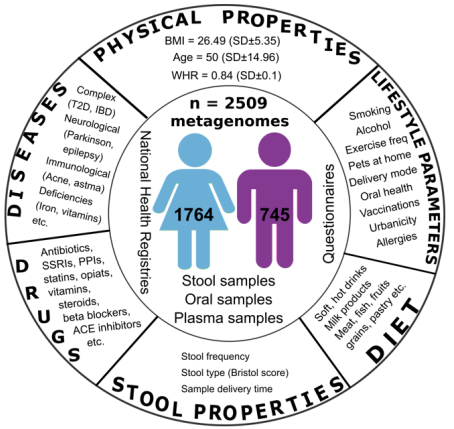
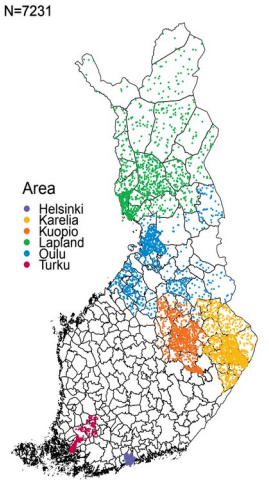


Figure 3. Overall Workflow

Elin Org – random prospective cohort of Estonia, deep sequencing
Estonian microbiome cohort (EstMB)

- Gut metagenomics data for 2509 Estonian Biobank participants (Illumina, average 20M/sample)
- Random adult population sample, age >18 (age range 23-89, mean 50 ± 14.96)
- Extensive life-style/environmental data, updated clinical data by linking to digital health registers (EHRs)
- All individuals are genotyped, overlap with other omics datasets (metabolomics, proteomics etc).





FINRISK cohort

NATIONAL INSTITUTE FOR HEALTH AND WELFARE

DATA (fastq + metadata)
Pre pre processing
HPC (Finland, Slovenia, …)

WG3

Open data
Files can be shared globally
Pre2 pre1 processing using different pipelines
Data integration

DREAM challenge

GDPR
No data travel
Software delivered by participants and tested in Finland

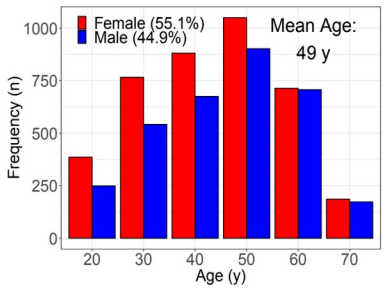
Data summary: FINRISK2002

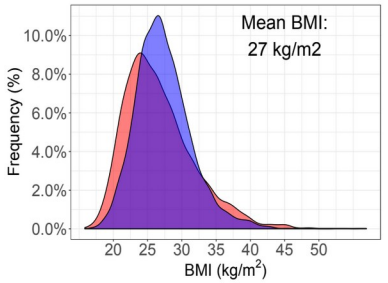
Cohort

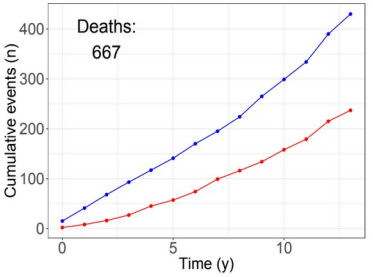
- 7231 adult stool samples (54% participation rate) *collected* in **2002**
- 18+ year follow up (in 2020)
- comprehensive health info from Finnish health registers (doctor visits, diagnoses, medication purchases..)

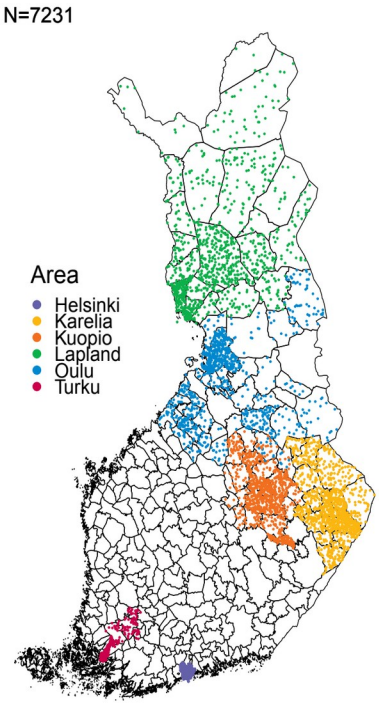
Omics and clinical measurements

- **shallow shotgun metagenome** (~1M reads/sample)
- 16S
- stool metabolome (NMR & MS)
- host genome & exome (subset of ~3000 samples)
- host & clinical parameters (BMI, Diet, BP, etc.)



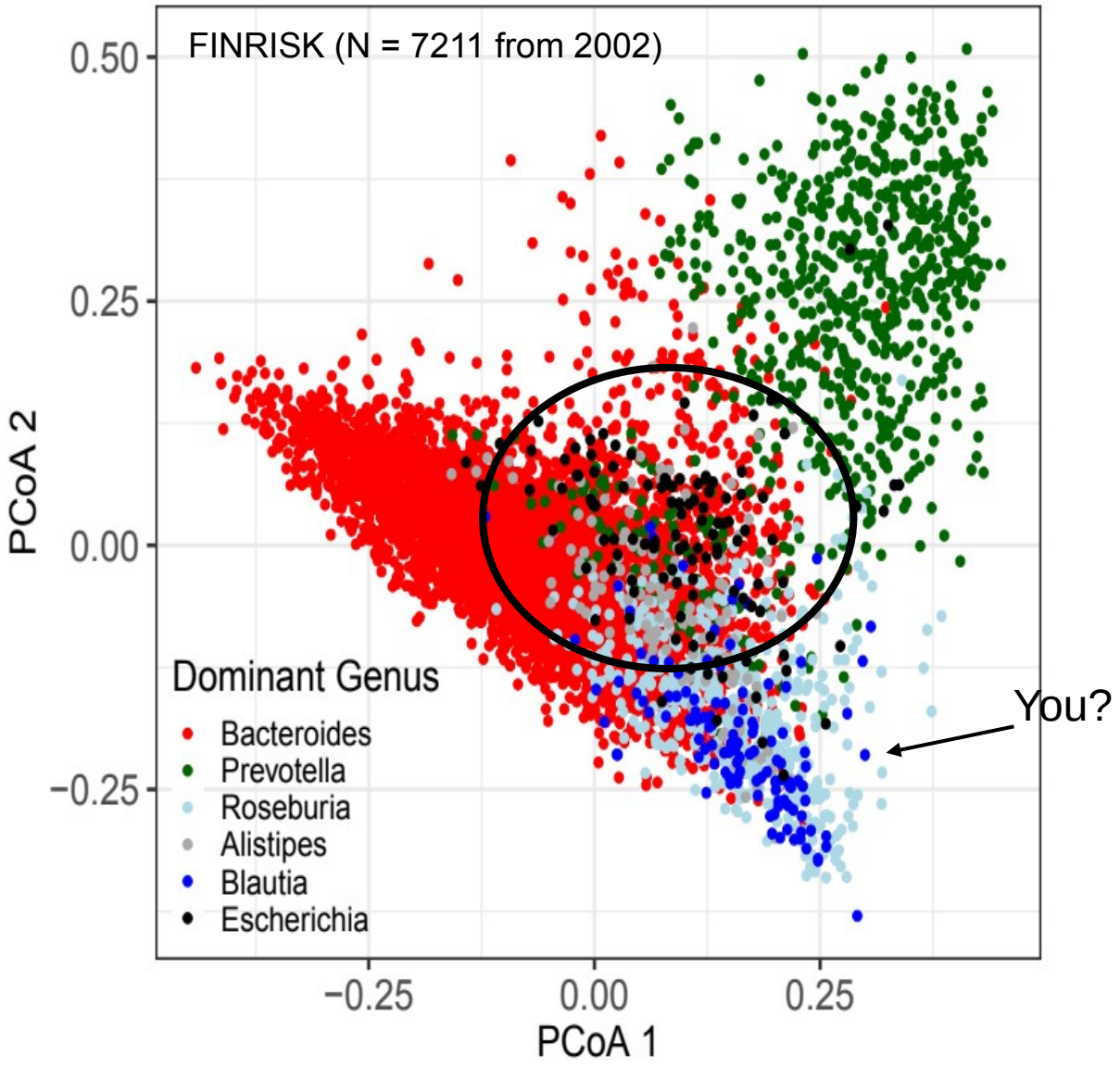


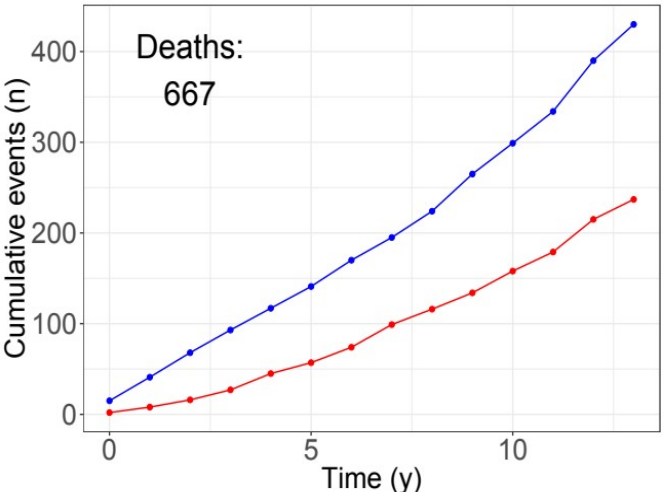


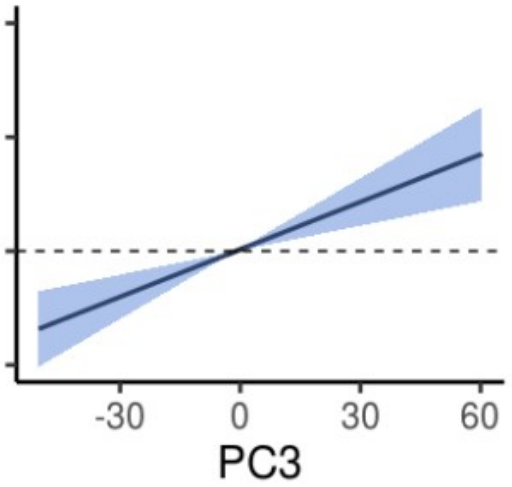


NATIONAL INSTITUTE FOR HEALTH AND WELFARE

Predicting overall mortality risk in Finnish adult population







Taxonomic signatures of cause-specific mortality risk in human gut microbiome

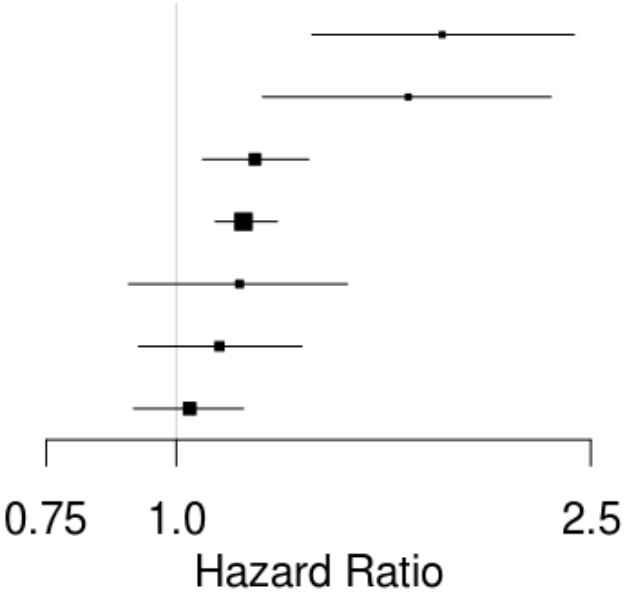
Aaro Salosensaari, Ville Laitinen, Aki S. Havulinna, Guillaume Meric, Susan Cheng, Markus Perola, Liisa Valsta, Georg Alfthan, Michael Inouye, Jeramie D. Watrous, Tao Long, Rodolfo A. Salido, Karenina Sanders, Caitriona Brennan, Gregory C. Humphrey, Jon G. Sanders, Mohit Jain, Pekka Jousilahti, Veikko Salomaa, Rob Knight, Leo Lahti ✉ & Teemu Niiranen ✉

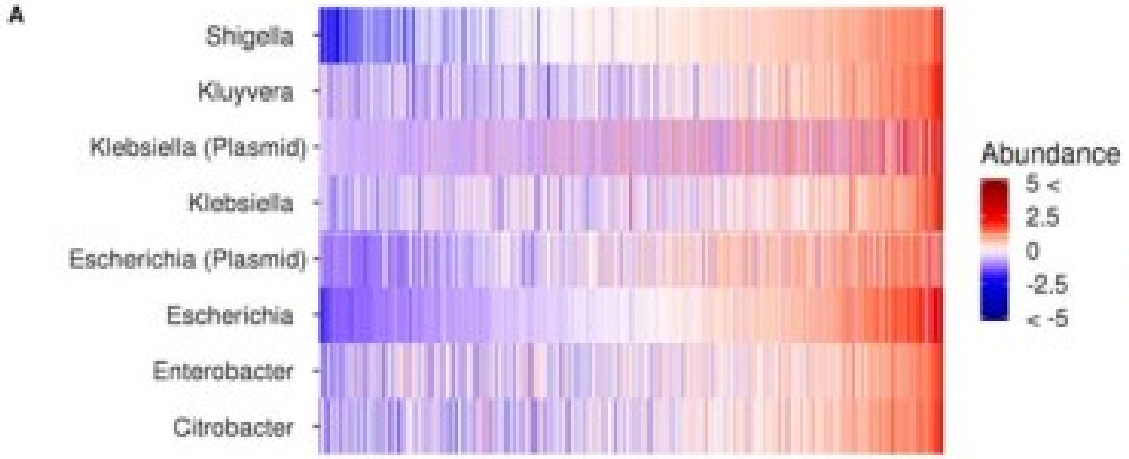
Nature Communications **12**, Article number: 2671 (2021) | Cite this article

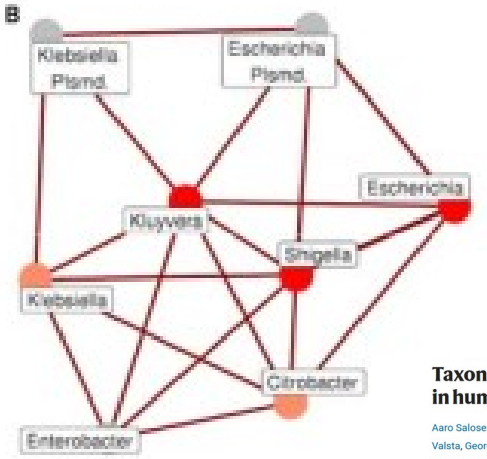
NATIONAL INSTITUTE FOR HEALTH AND WELFARE

Taxonomic signature associated & specific causes of mortality

Cause of Death	Deaths	HR	FDR
Gastrointestinal	36	1.80 (1.35-2.41)	2.2×10^{-4}
Respiratory	31	1.67 (1.21-2.29)	0.004
Cancer	238	1.19 (1.06-1.34)	0.007
All	729	1.16 (1.09-1.25)	9.5×10^{-5}
Neurological	60	1.15 (0.90-1.46)	0.333
Other	110	1.10 (0.92-1.32)	0.333
Cardiovascular	254	1.03 (0.91-1.16)	0.633





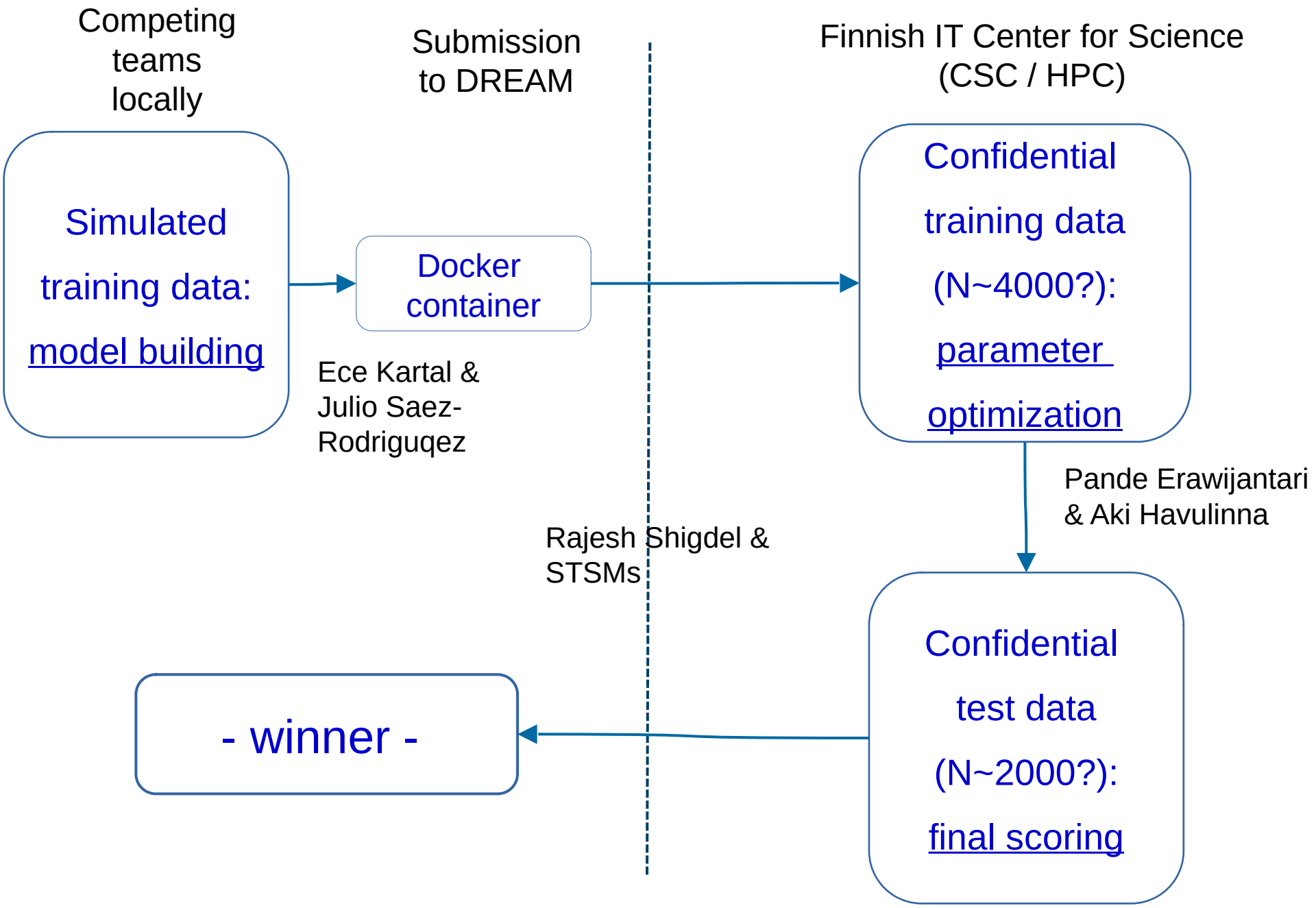


Taxonomic signatures of cause-specific mortality risk in human gut microbiome

Aaro Salosensaari, Ville Laitinen, Aki S. Havulinna, Guillaume Meric, Susan Cheng, Markus Perola, Liisa Valsta, Georg Alfthan, Michael Inouye, Jeramie D. Watrous, Tao Long, Rodolfo A. Salido, Karenina Sanders, Caitriona Brennan, Gregory C. Humphrey, Jon G. Sanders, Mohit Jain, Pekka Jousilahti, Veikko Salomaa, Rob Knight, Leo Lahti ✉ & Teemu Niiranen ✉

Nature Communications **12**, Article number: 2671 (2021) | Cite this article

DREAM Status & logistics



Remaining tasks on DREAM challenge

0. Get the STSMs
1. Finalize simulations
2. Finalize containers
3. Finalize the submission workflow
4. Finalize the challenge agreements & materials (wiki, website, flyers..)
5. Finalize the approval by data owners (THL & partners)
6. Launch the competition (summer 2022?)
7. Publish two *reports* on DREAM challenge

Notes
https://docs.google.com/presentation/d/1YC79LYEwiZvoPiYSrqfg8ijji2TIzTEUFH-8cruWNlM/edit?usp=sharing

Support for other WGs

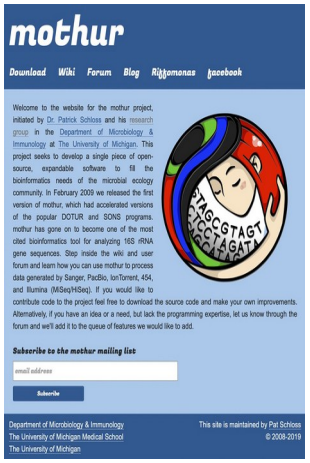
WG1 State-of-the-art evaluation and update

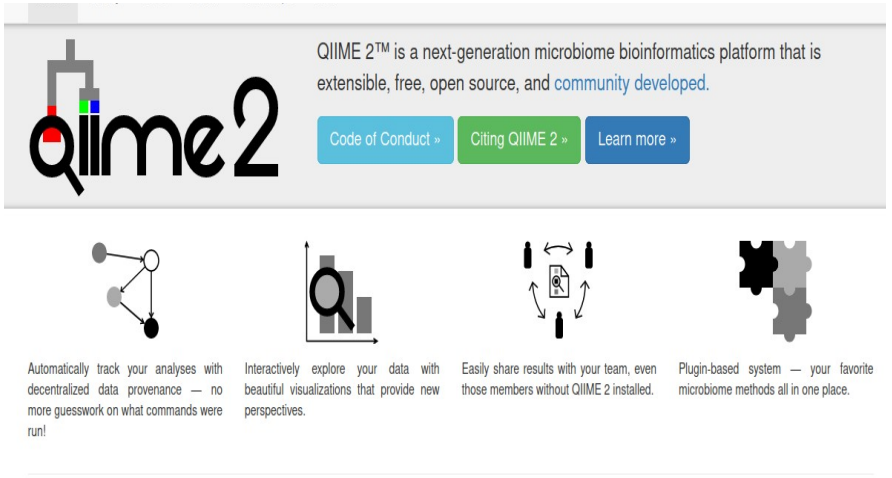
WG2 Benchmark data & DREAM Challenge

WG3 Optimisation and standardisation

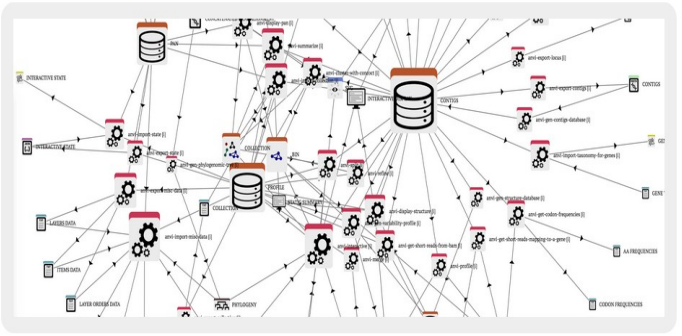
WG4 Dissemination and training
- ML4microbiome training schools 2021&2022

open microbiome data science

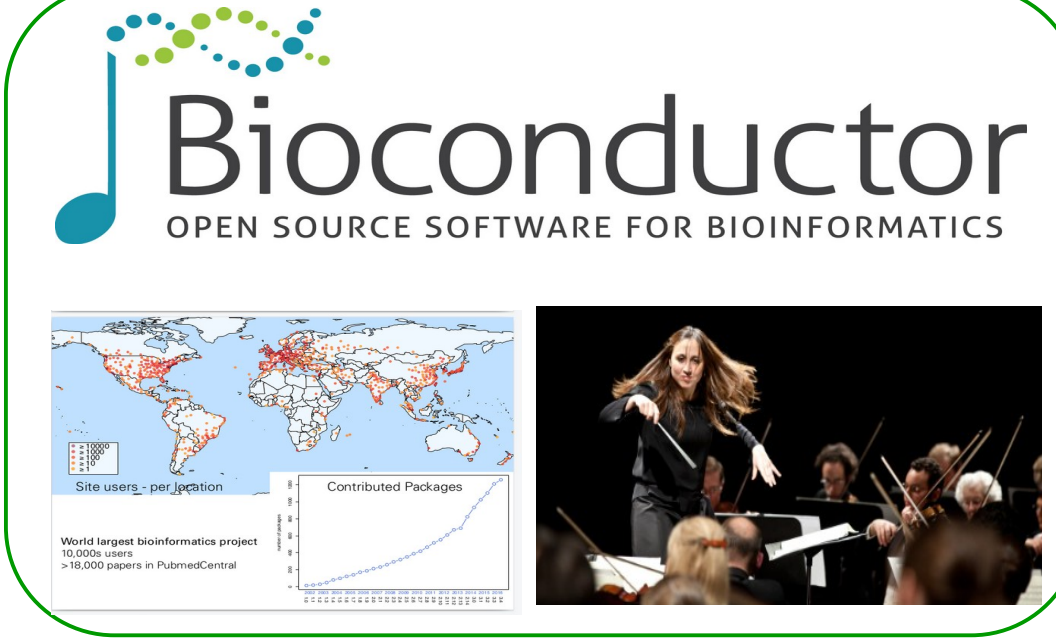




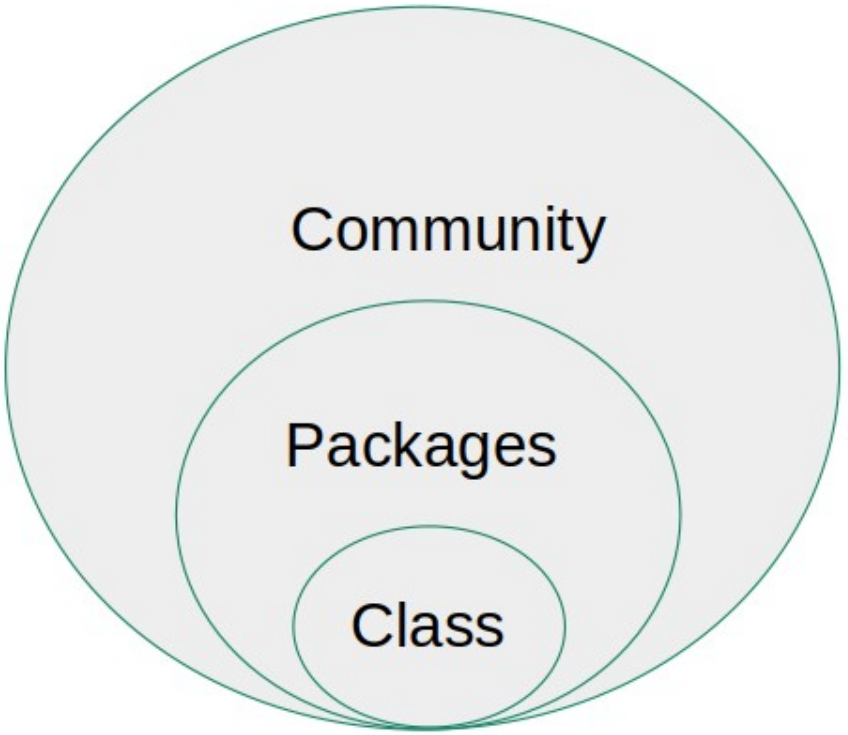
Anvi'o in a nutshell



Anvi'o is an open-source, community-driven analysis and visualization platform for 'omics data.

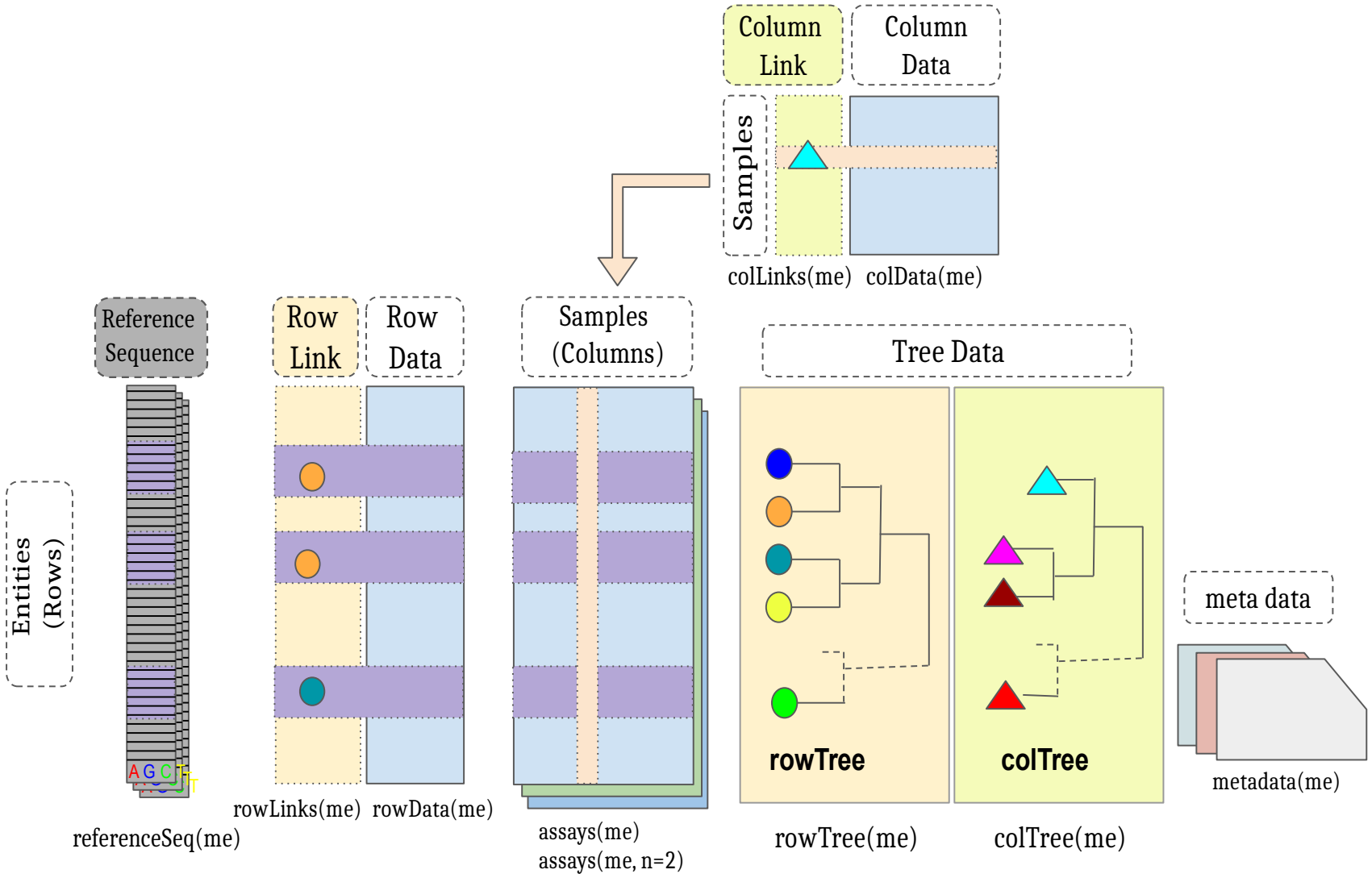


Standardized data containers are central for the R/Bioc ecosystem



TreeSummarizedExperiment

by Ruizhu @fiona Huang



Seamless conversion from *phyloseq* & other raw data types

Optimal container for microbiome data?

Multiple assays
seamless interlinking

Hierarchical data
supporting samples & features

Side information
extended capabilities & data types

Optimized
for speed & memory

Integrated
with other applications & frameworks

Reduce overlapping efforts, improve interoperability, ensure sustainability.

Complete workflows

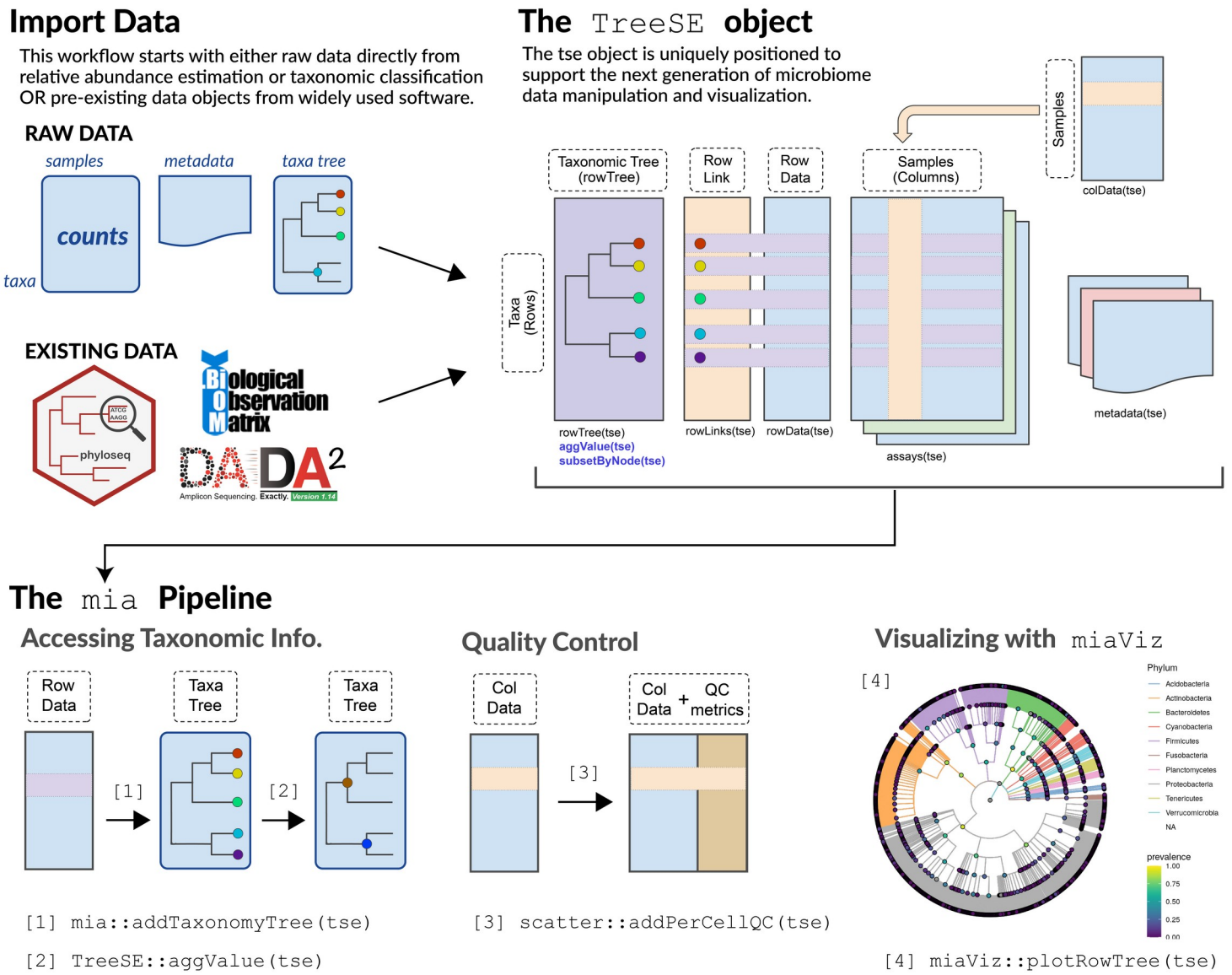


Figure: Poster F1000 / EuroBioC 2020

Special thanks
Felix G.M. Ernst
Sudarshan A. Shetty
Tuomas Borman
Ruizhu Huang
Domenick J. Braccia
Héctor Corrada Bravo
Leo Lahti

Orchestrating Microbiome Analysis

Authors: *Leo Lahti [aut], Sudarshan Shetty [aut], Felix GM Ernst [aut, cre]*
Version: *0.98.9*
Modified: *2021-04-10*
Compiled: *2021-07-29*
Environment: *R version 4.1.0 (2021-05-18), Bioconductor 3.14*
License: *CC BY-NC-SA 3.0 US*
Copyright:
Source: *https://github.com/microbiome/OMA*

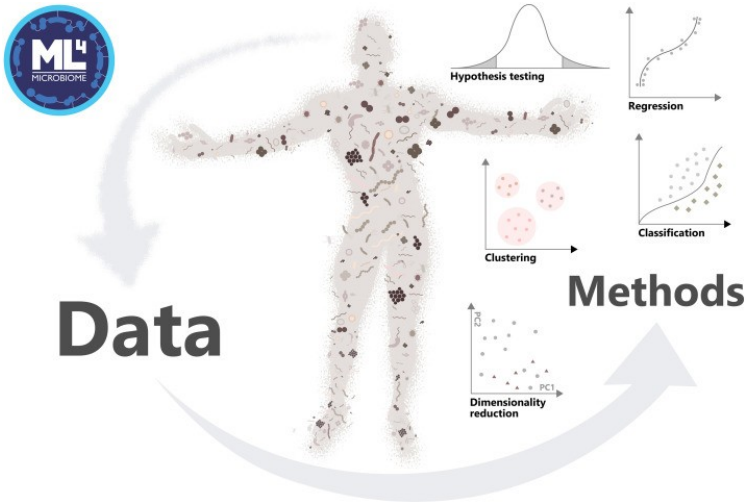


Figure source: Moreno-Indias et al. (2021) Statistical and Machine Learning Techniques in Human Microbiome Studies: Contemporary Challenges and Solutions. Frontiers in Microbiology 12:11.

Online tutorial (beta)

microbiome.github.io

Forthcoming courses & events:

June 20-23, Oulu, Finland

July 11-15, Nijmegen, The Netherlands

August 8-12, Pune, India (online / hybrid?)

Sep 26-30, CSC, Finland (online)

Oct 5-7, Barcelona, Spain (ML4microbiome)

Nov, Turku, Finland?

- Linking more tightly with ML4microbiome COST action?
- Gut Microbiome Atlas as a demonstration data set?

Next MC & WG meeting
August 29-31, 2022 - Turku, Finland

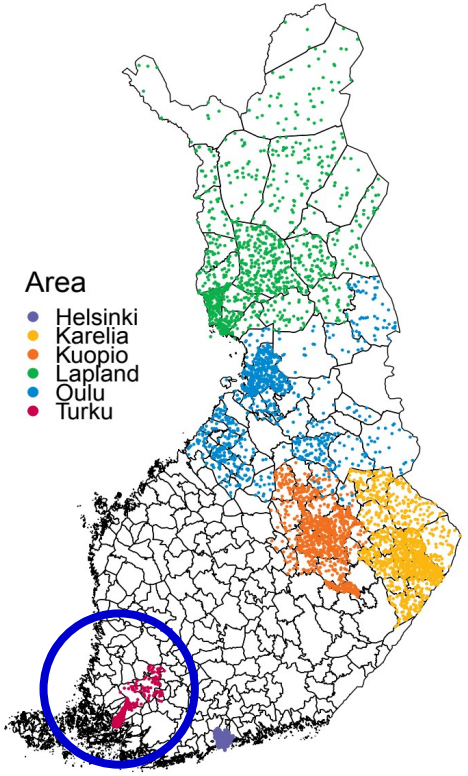
<u>Tentative program</u>

Monday
- Arrival
- Afternoon session
- Optional evening program

Tuesday
- MC meetings
- WG meetings
- Dinner

Wednesday
- Morning session
- Departure

→ Other needs?
→ Schedules?
→ Public seminar?



<u>Travel?</u>
- Fly to Helsinki
- Airport Bus/Train to hotel