*Image credit:*
PhD Comics

# Intro to Metagenomics

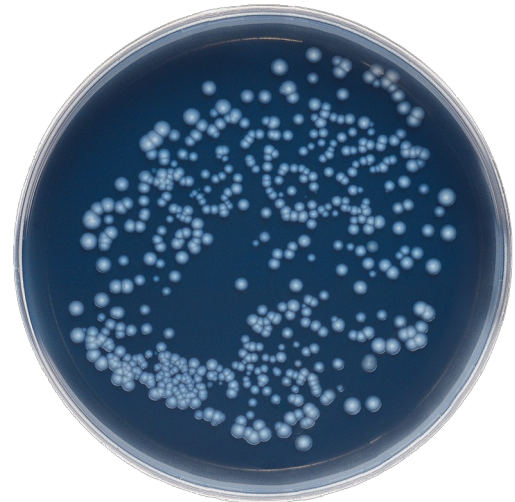SLU --- Intro to Metagenomics 2020

@metamoritz
@moritzbuck
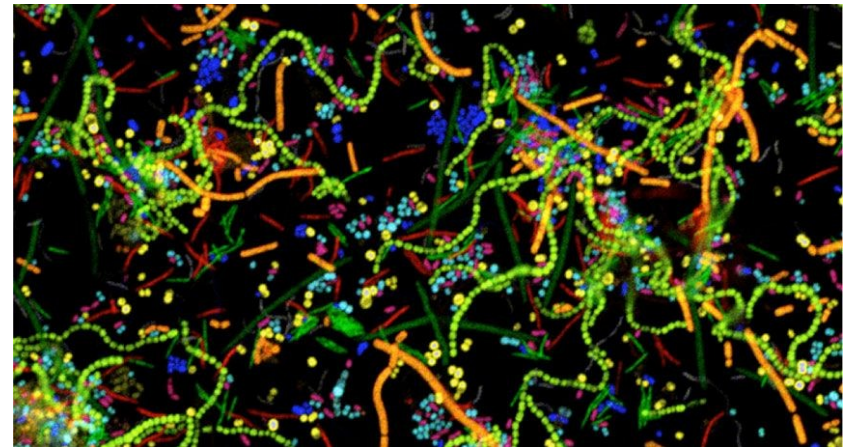@metamoritz

# typical genomics

Single organism
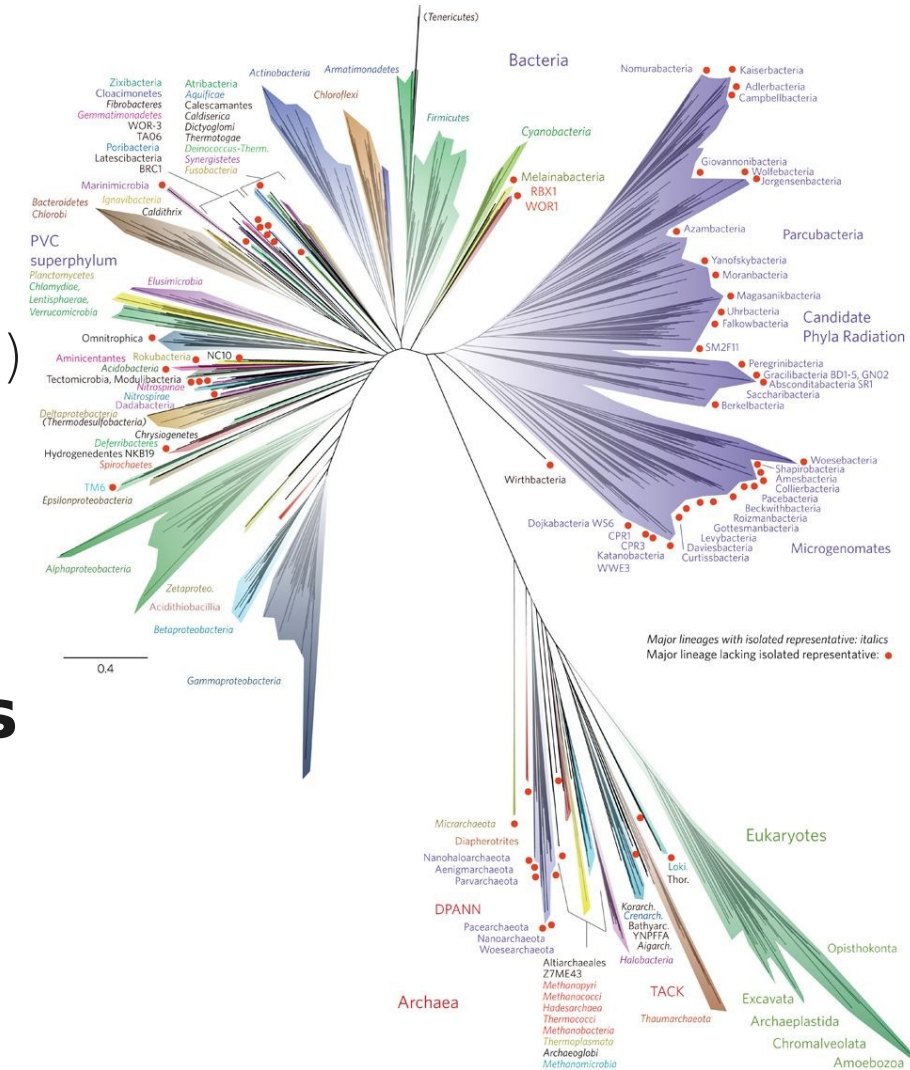Single genome

# Meta- genomics



Many organisms
A Metagenome

# Why make our life complicated

- **It is hard/impossible to separate the organisms:**
  - Isolating (e.g. making a pure culture) microbes is hard
- **We want to study interaction**
- **We want to study organisms *in situ***
- **It's mainly about the microbes**



Hug et al 2016 (10.1038/nmicrobiol.2016.48)

# A human gut



- 10*10^13 (1.3 x human cells)
- About 0.5% of dry weight
- ~150 species in every gut
- 300x human genes
- But diverse from one to the other! More than 10 that many known gut species and genes
- But only ~20% cultivated

# A lake



- 1.35*10^17 cells (a milion per mL)
- About 135kg of dry weight
- ~700 species in the water
- 3.000.000 encoded genes
- Only a couple of handfull are cultivated.

# It all maters

- **They are everywhere**
- **They do a lot of stuff:**
  - 90% of disease in humans can be related to microbiomes
  - Most nutrients that end up in your blood stream have been somewhat metabolized by some microbe
  - Microbes do 2/3rds of carbon fixation in aquatic environments
  - But also the majority of carbon emissions!

# Before the computer

- **Get sample**
- **Extract DNA**
- **Prep library**
- **Send to Sequencing facility**
- **????**
- **PROFIT**

# Dream-quest of unknown metagenome

**From an environment with the power of genomics we want to know:**

- Who is in the environment? (taxonomy)
- How many of each? (abundance)
- What can they do they? (genetic potential)
- What are they doing? (expression analysis)
- What do they eat? (metabolism)
- Where do they come from? (evolution and ecology)

# Sequencing strategies

- **Targeted, a.k.a. amplicon /metabarcoding/edna**
- **Shotgun sequencing:**
  - Short reads (HiSeq, NovaSeq, all Illumina basically)
  - Long reads (PacBio, Nanopore, …)
- **Metatranscriptomic**
- **Functional metagenomics**
- **…**



*Image credit: xkcd*

# Shotgun and amplicon

**Two main players:**

- **Shotgun:**
  - The obvious approach
  - Sequence all : make sense of it
  - Computationally diverse methodologies
- **Amplicon:**
  - Pick a gene, sequence it
  - Snapshot of a function or taxonomy
  - Much cheaper, more depth



*Image credit:* xkcd

# This week

- **First let's do some 16s rRNA-gene amplicon!**
  - Some mouse poop samples!
  - Using dada2 and phyloseq in R
- **Then let's do some shotgun!**
  - The king-discipline (I am biased)
  - Some simulated data based on data from the TARA ocean stuff
  - Assembly with megahit, binning with metabat, and much more!