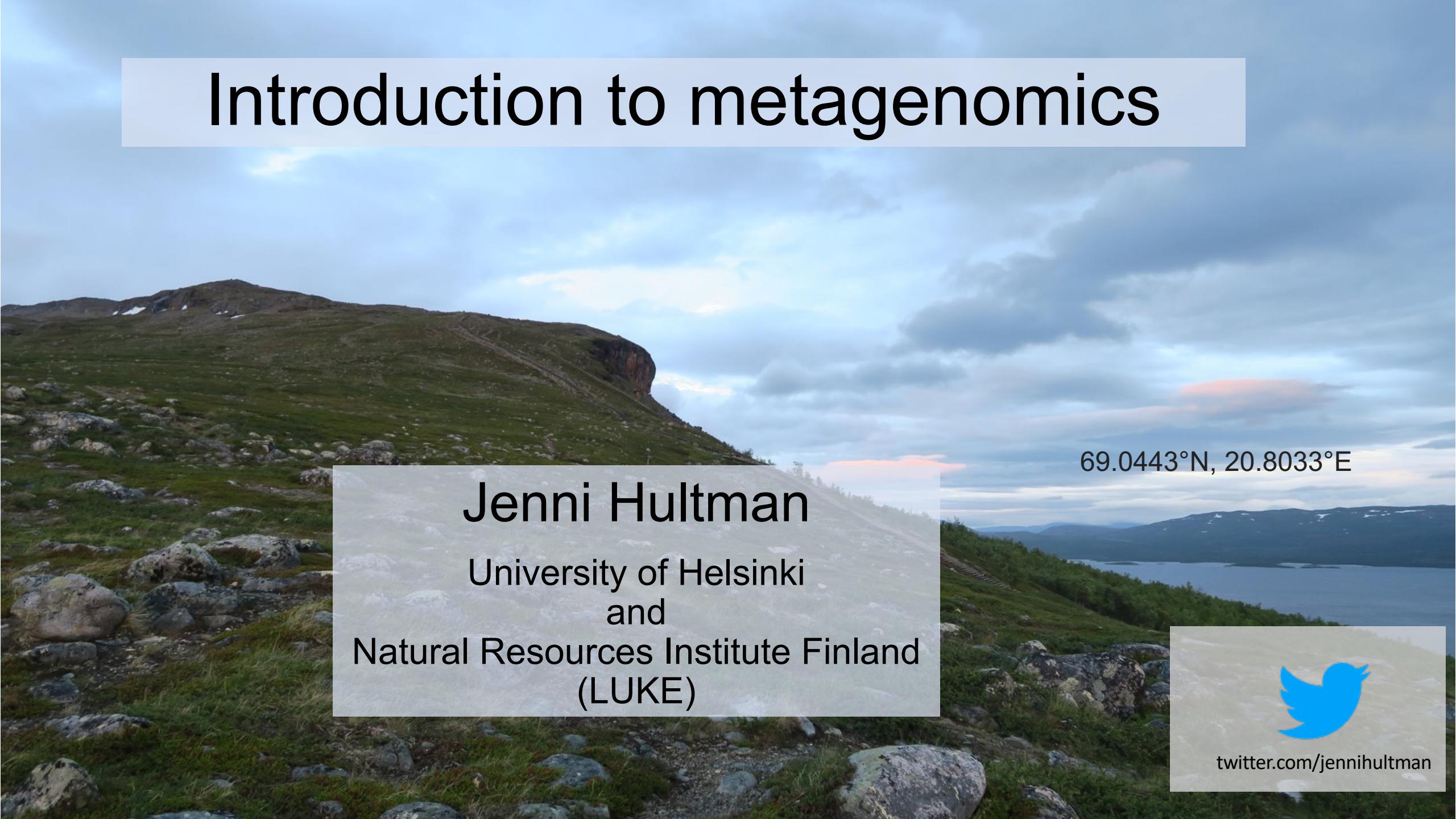


Introduction to metagenomics



Jenni Hultman

University of Helsinki
and

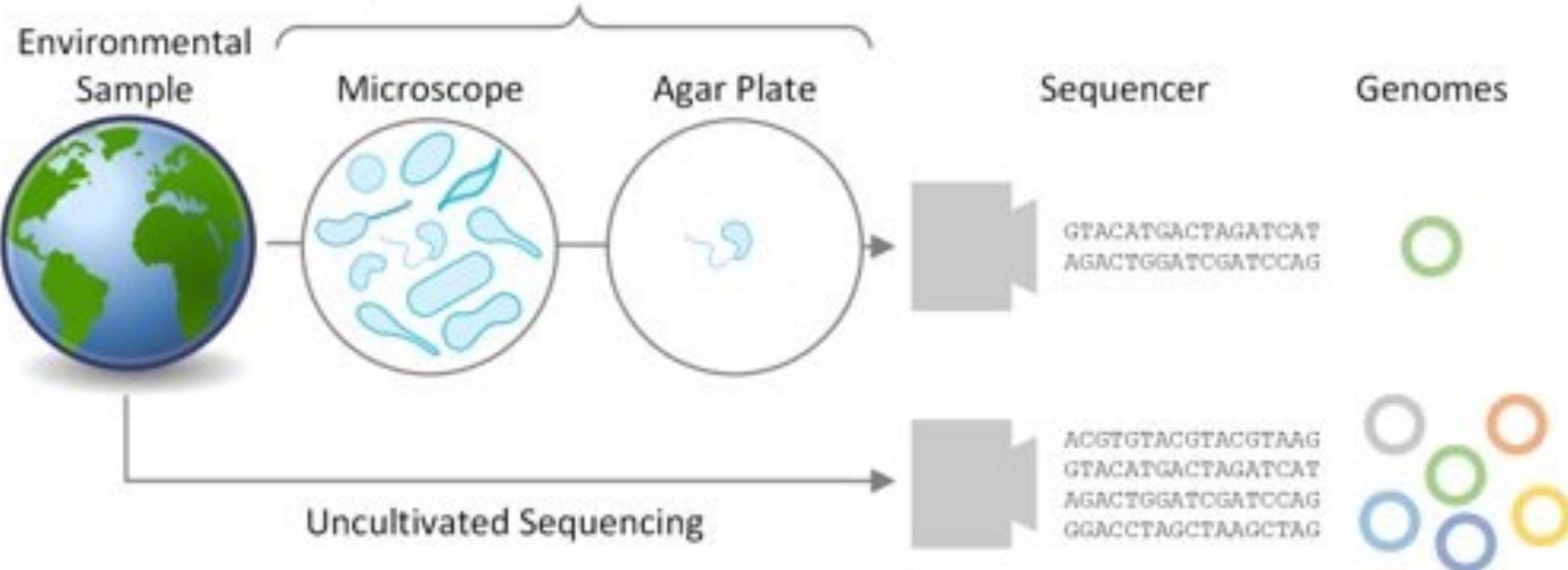
Natural Resources Institute Finland
(LUKE)

69.0443°N, 20.8033°E



twitter.com/jennihultman

Great Plate Count Anomaly
Only ~1% of Bacteria is Culturable



Metagenomics

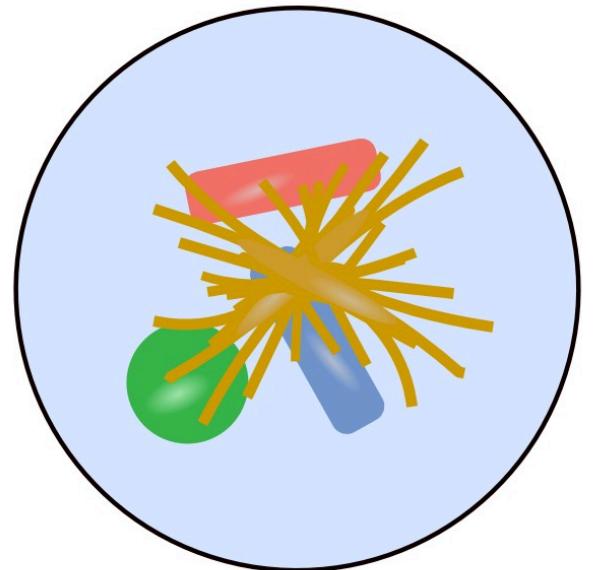
- Jo Handelsman 1998



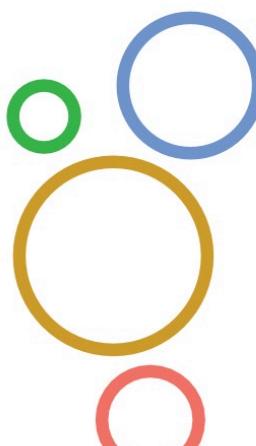
“the application of modern genomics techniques to the study of communities of microbial **organisms directly in their natural environments**, bypassing the need for isolation and lab cultivation of individual species”

16S rRNA amplicon \neq metagenome

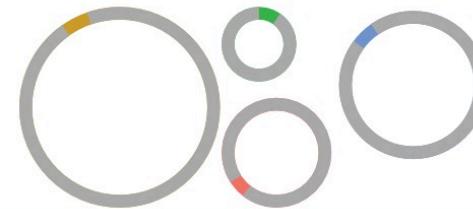
Mixed microbial community



DNA
Extraction

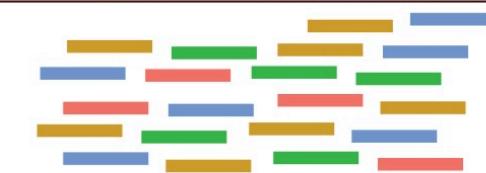
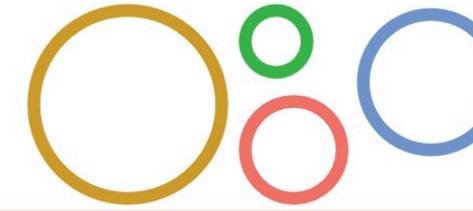


Amplicon sequencing



Multiple copies of fragments
from 1 target gene

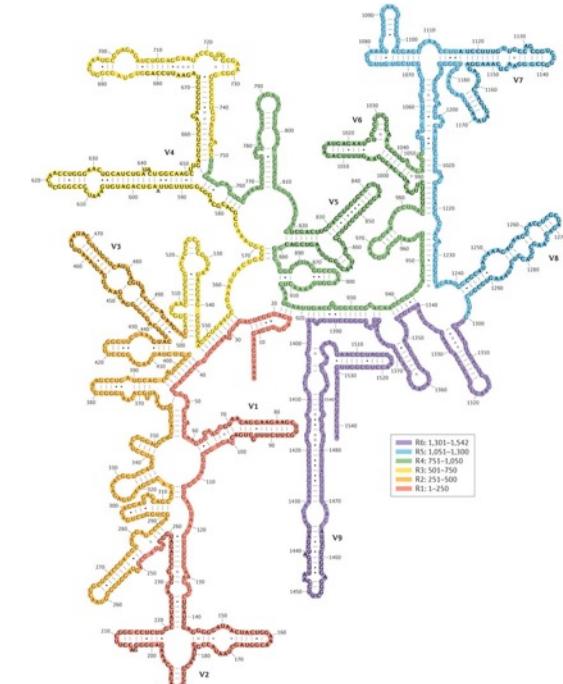
Metagenomics sequencing



Short sequence
fragments from "all" DNA

Amplicons or metagenome?

- Pros:
 - Not biased by amplicon primer set
 - Not limited by conservation of the amplicon
 - Can also provide functional information
- Cons:
 - Environmental contamination, including host
 - More expensive (150€+/sample vs 25€)
 - Complex data analysis
 - Requires high performance computing, high memory, high compute capacity



H

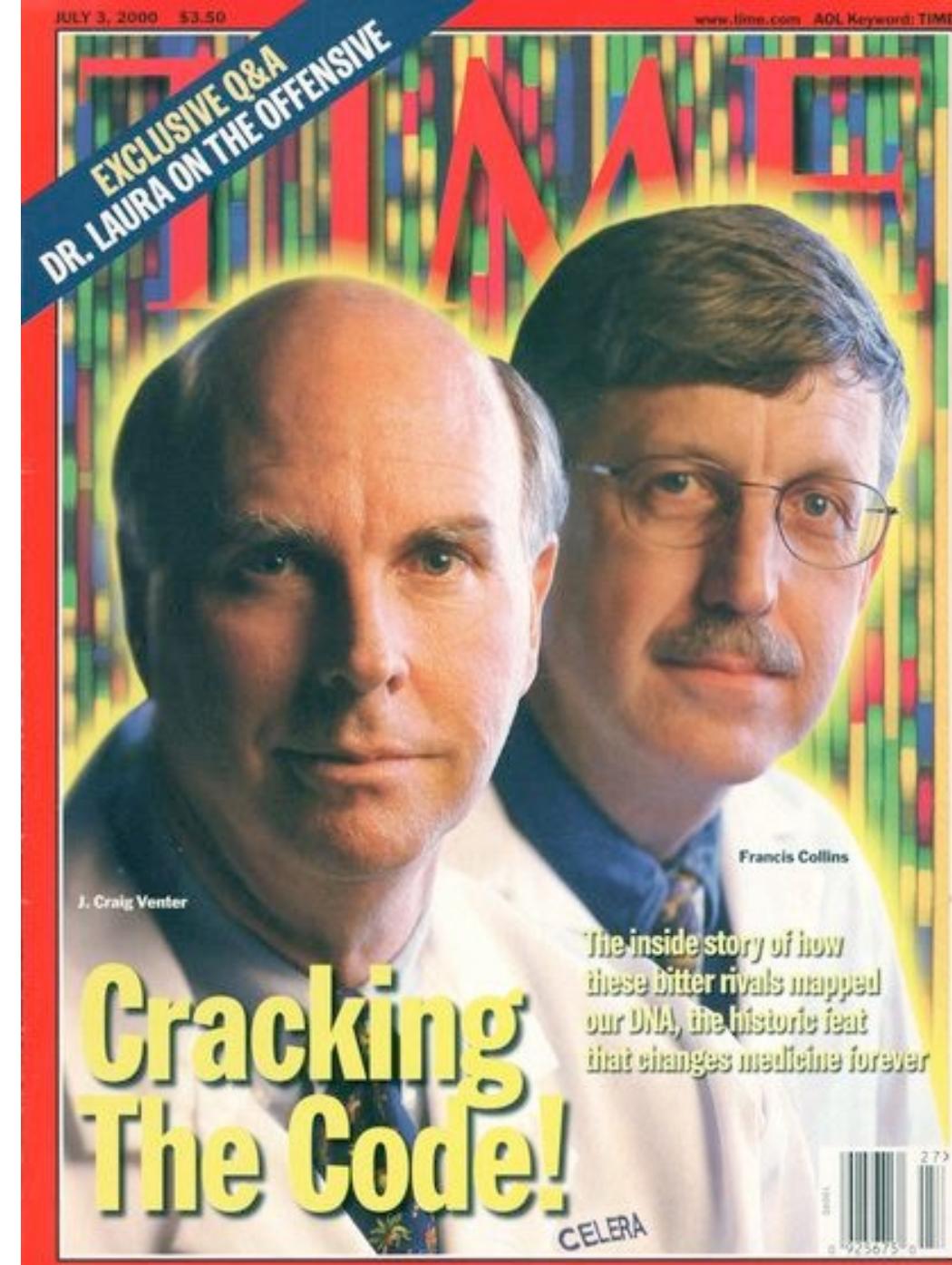
CS

ea of

A

rent

- 198
close
envir
- 199
geno
- 200
sec
virus
- 200
me



RESEARCH ARTICLE

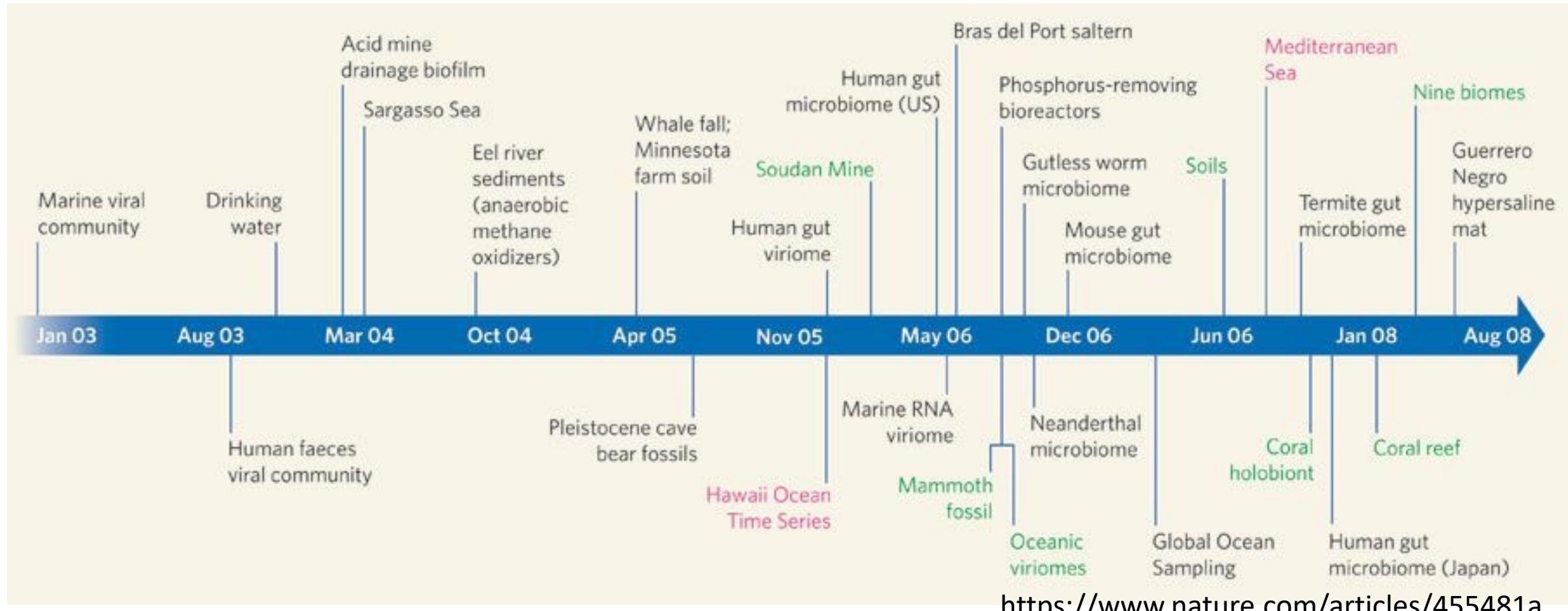
Environmental Genome Shotgun Sequencing of the Sargasso Sea

J. Craig Venter,^{1*} Karin Remington,¹ John F. Heidelberg,³
Aaron L. Halpern,² Doug Rusch,² Jonathan A. Eisen,³
Dongying Wu,³ Ian Paulsen,³ Karen E. Nelson,³ William Nelson,³
Derrick E. Fouts,³ Samuel Levy,² Anthony H. Knap,⁶
Michael W. Lomas,⁶ Ken Nealson,⁵ Owen White,³
Jeremy Peterson,³ Jeff Hoffman,¹ Rachel Parsons,⁶
Holly Baden-Tillson,¹ Cynthia Pfannkoch,¹ Yu-Hui Rogers,⁴
Hamilton O. Smith¹

We have applied "whole-genome shotgun sequencing" to microbial populations collected en masse on tangential flow and impact filters from seawater samples collected from the Sargasso Sea near Bermuda. A total of 1.045 billion base pairs of nonredundant sequence was generated, annotated, and analyzed to elucidate the gene content, diversity, and relative abundance of the organisms within these environmental samples. These data are estimated to derive from at least 1800 genomic species based on sequence relatedness, including 148 previously unknown bacterial phylotypes. We have identified over 1.2 million previously unknown genes represented in these samples, including more than 782 new rhodopsin-like photoreceptors. Variation in species present and stoichiometry suggests substantial oceanic microbial diversity.

- 148 previously unknown bacterial phylotypes
- 1.2 million previously unknown genes
 - 782 new rhodopsin like genes

Timeline of sequence-based metagenomic projects showing the variety of environments sampled since 2002



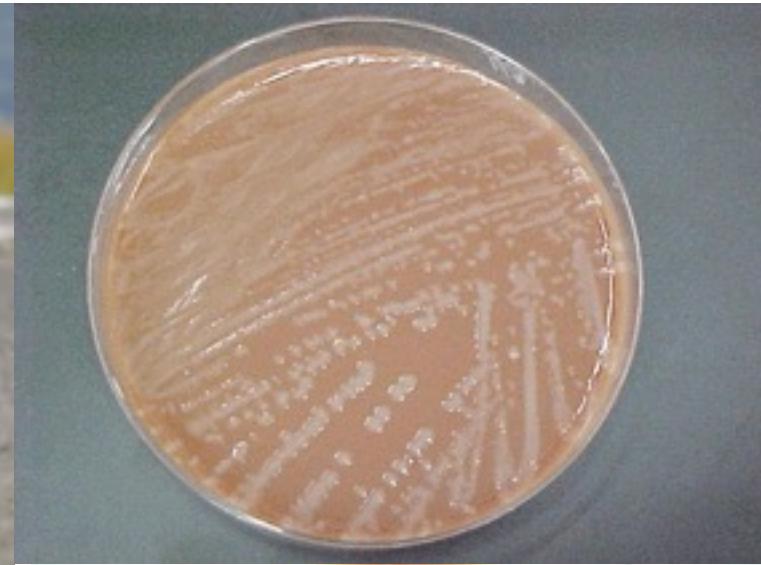
Why we need metagenomes for analysis of microbes



- Often metagenome is mostly microbial (Bacteria, Archaea, Fungi, some protist)
- Why metagenomes are high in microbes?
- Microbes are hard to study: small, diversity and numbers are high
- Phenotype

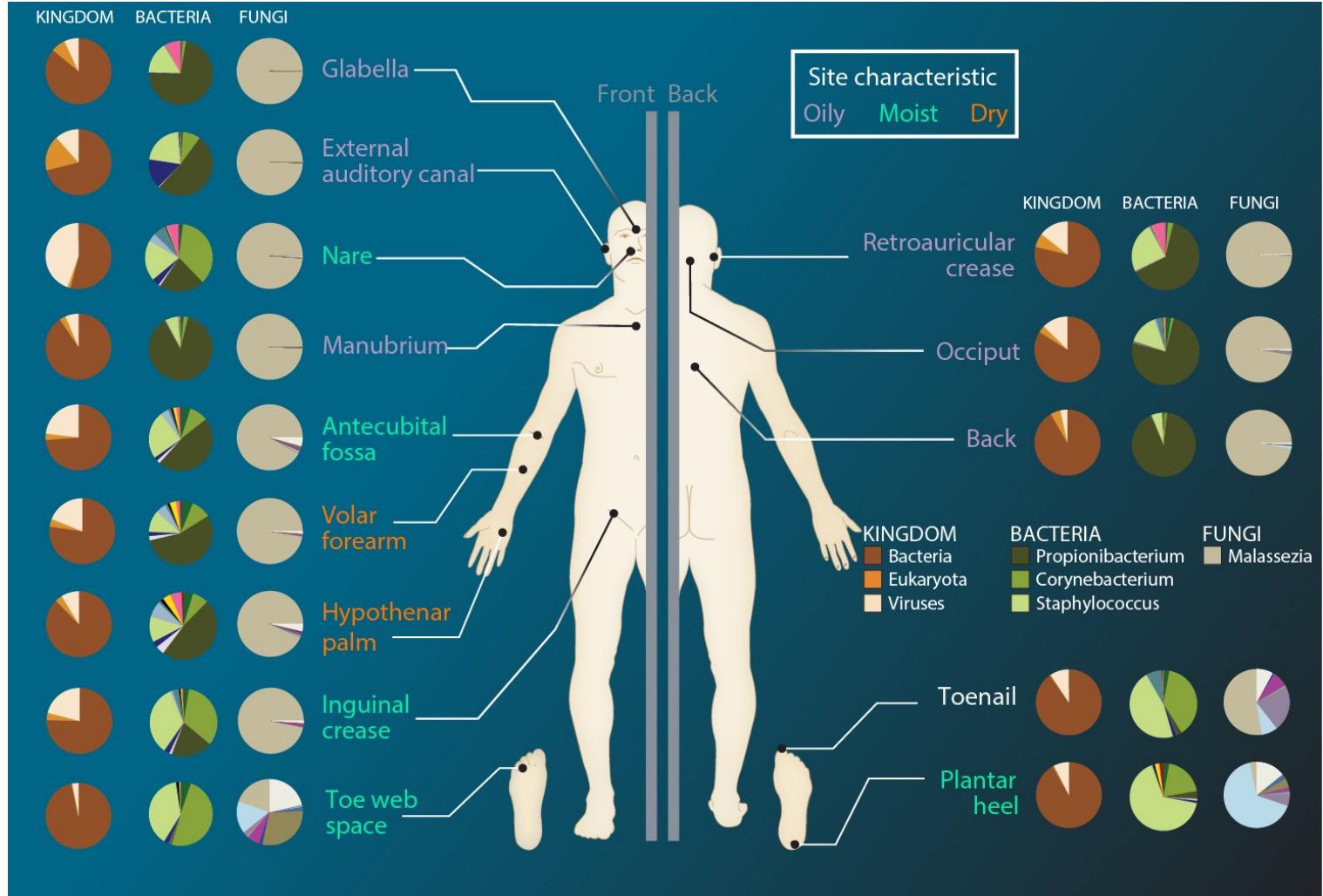
Metagenomes and microbes

Photo By Jim Floyd, Bear Viewing Client



Why study microbiome

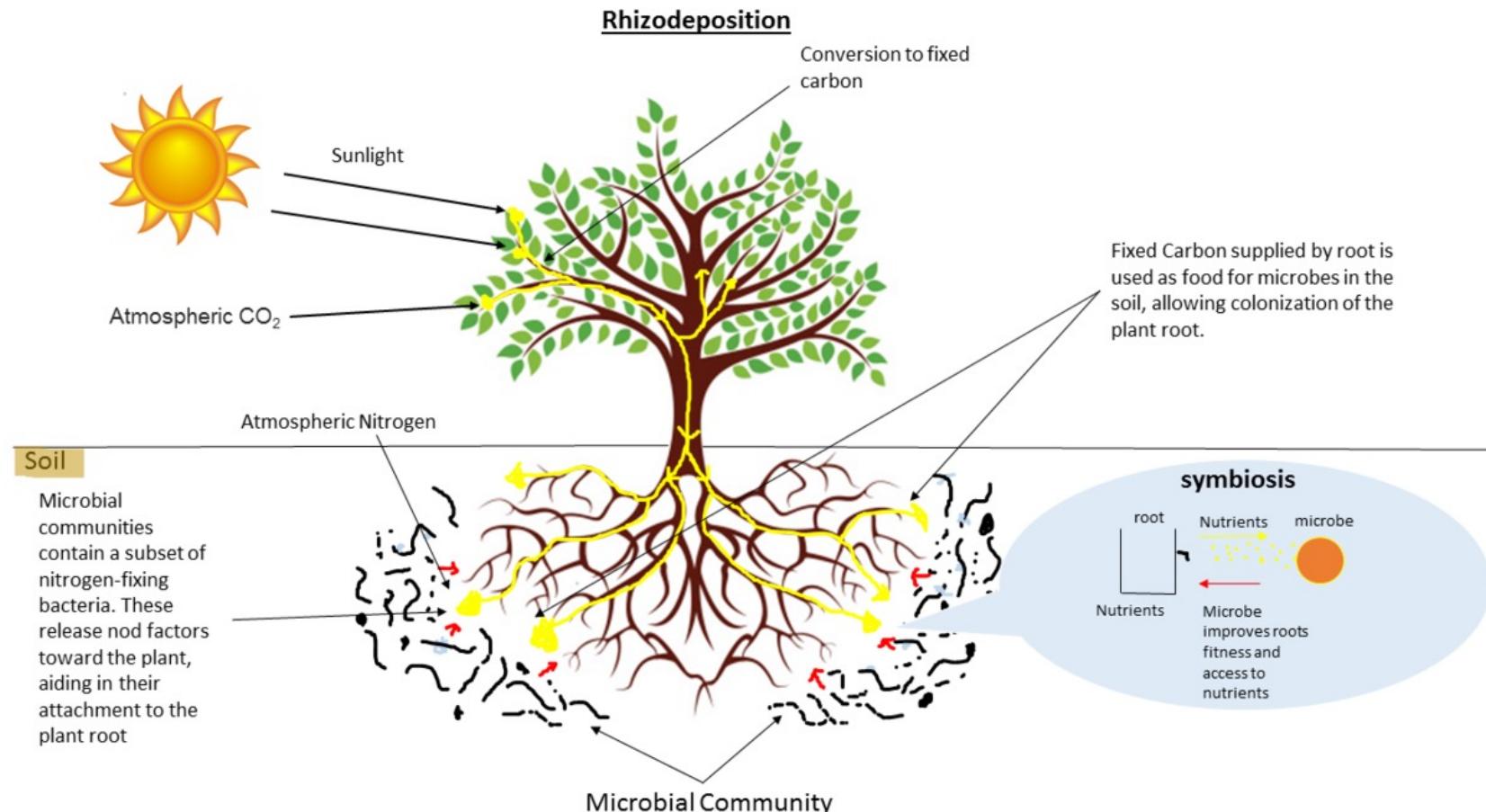
- Health care research
 - Humans are full of microorganisms
 - Skin, gut, oral cavity, nasal cavity, eyes, ..
 - Affects health, drug efficacy, etc
- Sometimes referred to as your **second genome**
 - ~10 times more cells than you
 - ~100 times more genes than you
 - ~1000s different species



<https://www.flickr.com/photos/genomegov/27058471125/>

Environmental studies

- Microbes in the soil affect plants and animals
- Improve agriculture
- Nutrients in the environment do not cycle but are cycled by microbes
 - Greenhouse gases and climate change

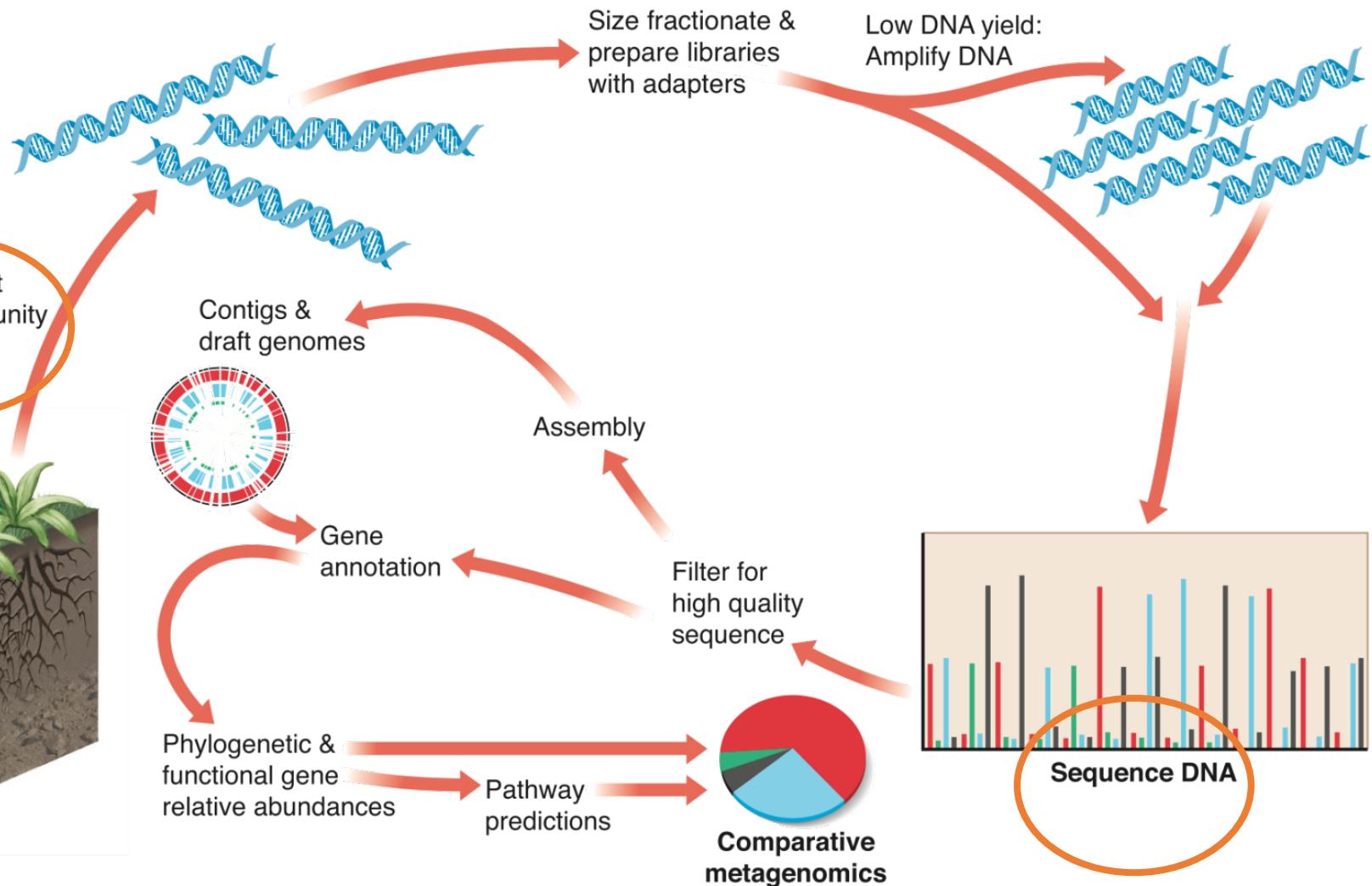
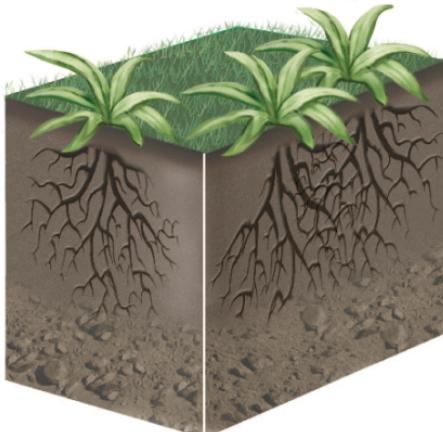


<https://upload.wikimedia.org/wikipedia/commons/d/dc/Rhizodeposition.png?20161205195949>

DNA from all
bacteria, archaea,
viruses, eukarya

Fragmented or
HMW DNA?

Extract
community
DNA



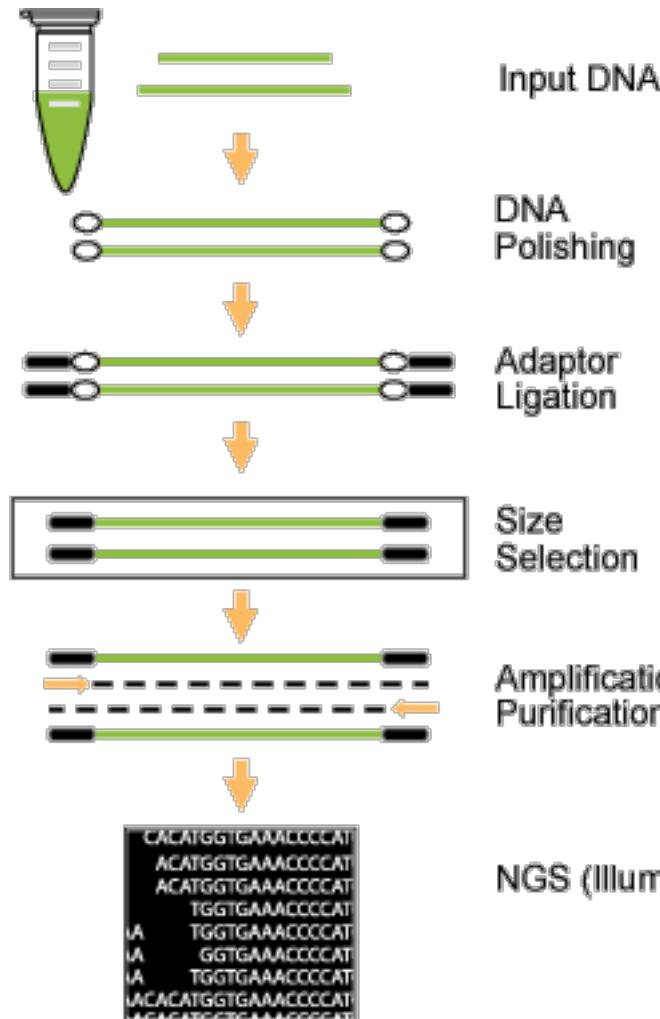
Illumina, PacBio,
Oxford Nanopore

Paired end, single
read, size of library

Sequencing depth,
replicates

Soil metagenomics methodology.

Library preparation

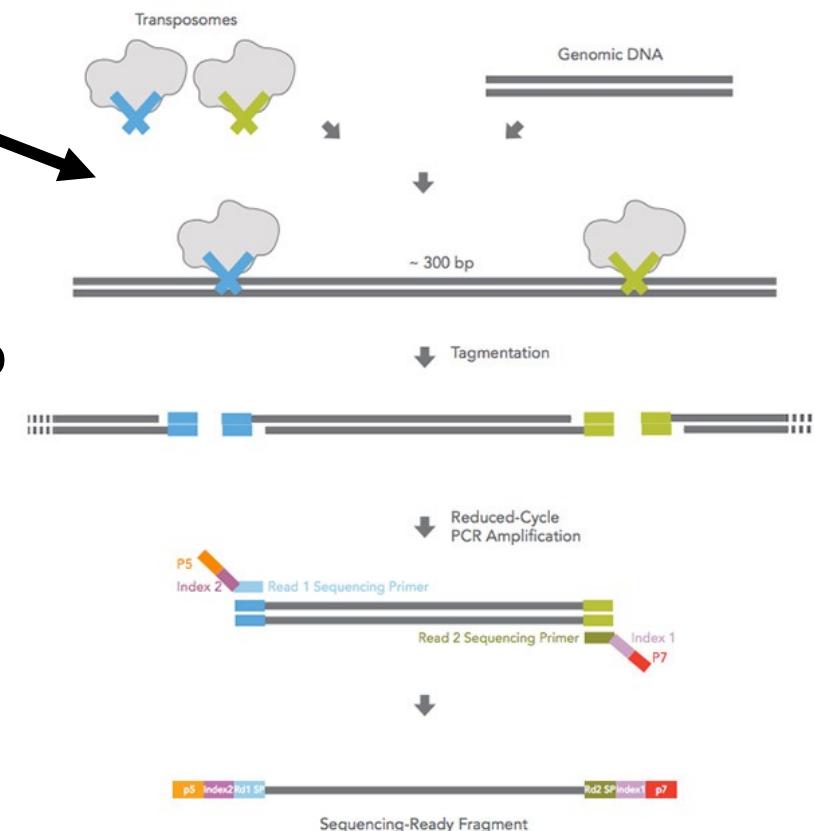


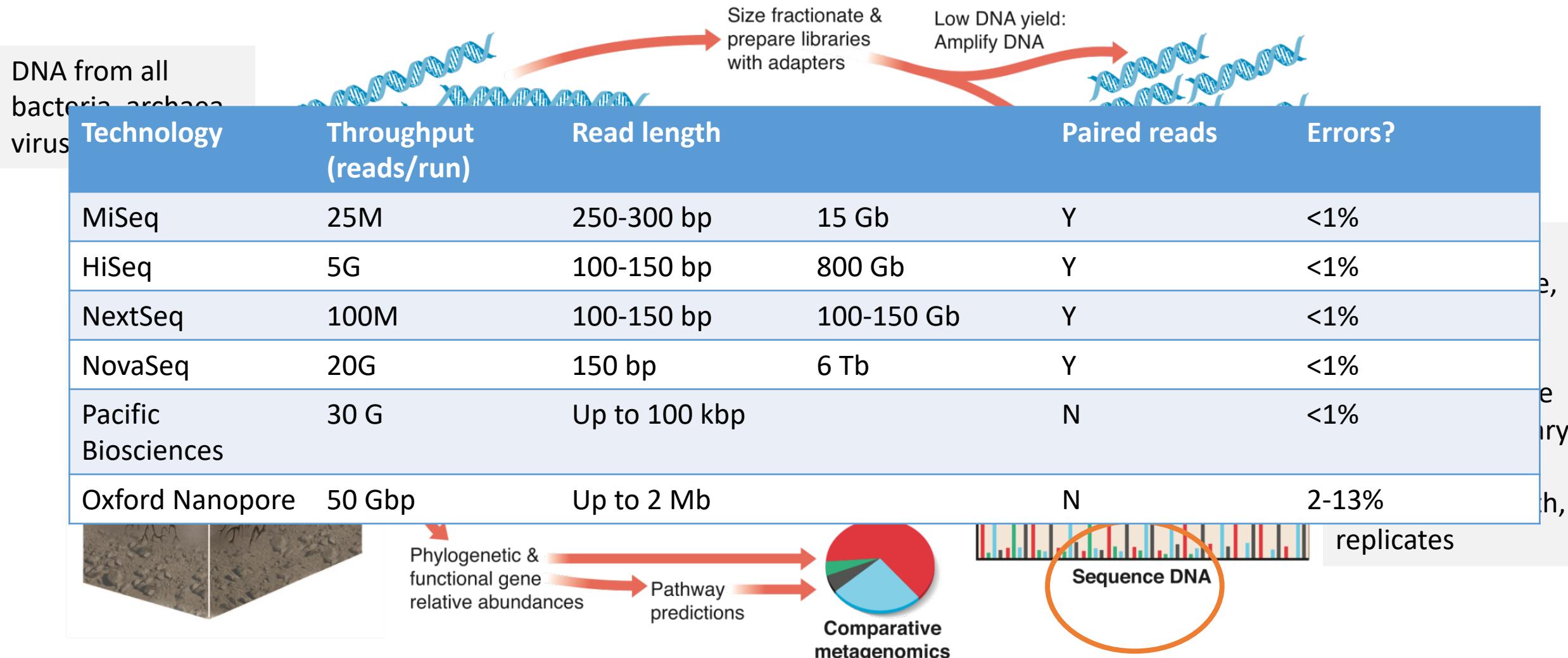
DNA shearing and adapter ligation

- DNA quantity

NexTera transposases

- Inhibitors (in soil)
- 1-10 ng of DNA
- Transposase:DNA ratio





Soil metagenomics methodology.

DNA from all
bacteria, archaea,
viruses, eukarya



Soil metagenomics methodology.

Jansson 2011: Towards tera-terra. Microbe June 2011

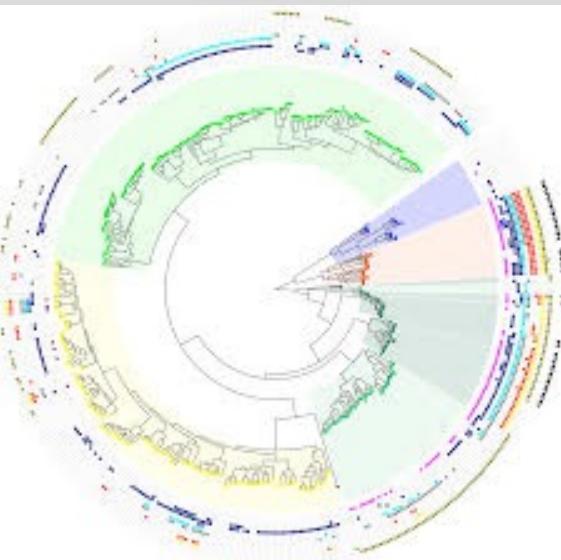
Illumina, PacBio,
Oxford Nanopore,
Ion Torrent

Paired end, single
read, size of library

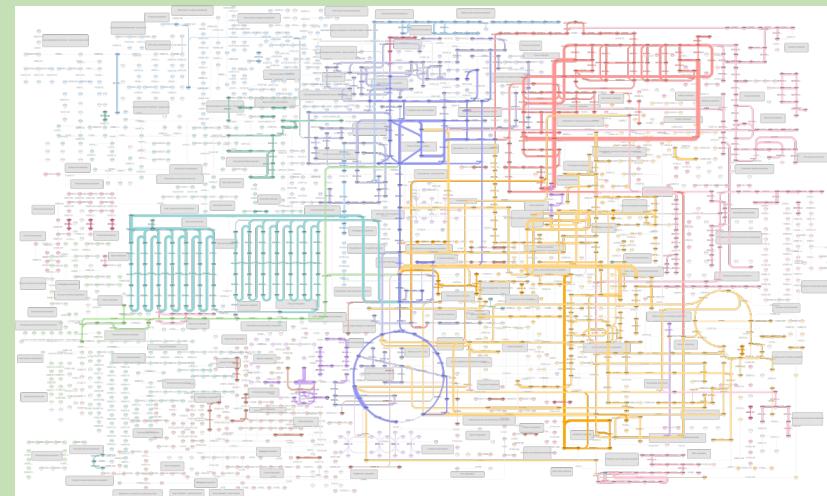
Sequencing depth,
replicates

With metagenomics

- Microbial community composition – who is there
16S/18S, ITS



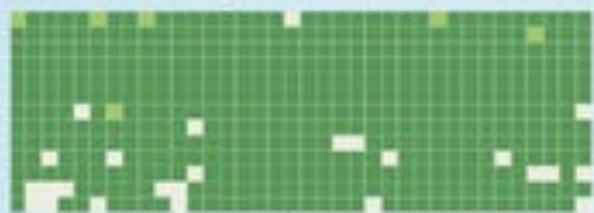
- Microbial community function – what can they do
Protein coding sequences



Assembly-based profiling

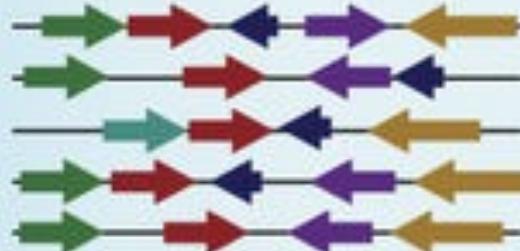
Quality control of MAGs

MAGs



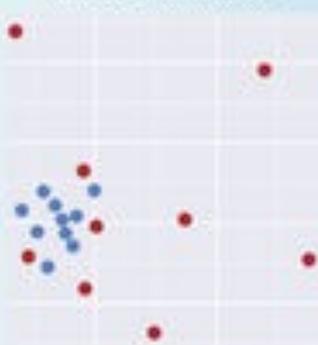
Genome-level characterization for abundant MAGs

MAGs

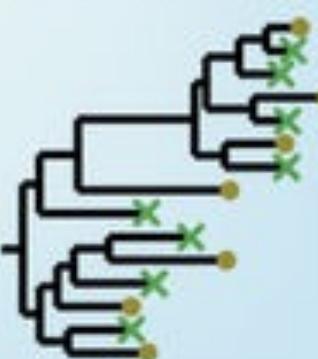


Genome annotations

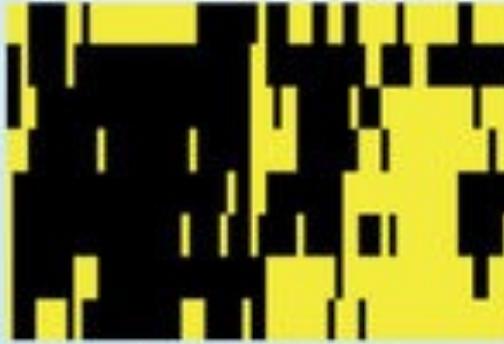
Ordination on MAG abundance



MAG phylogenies



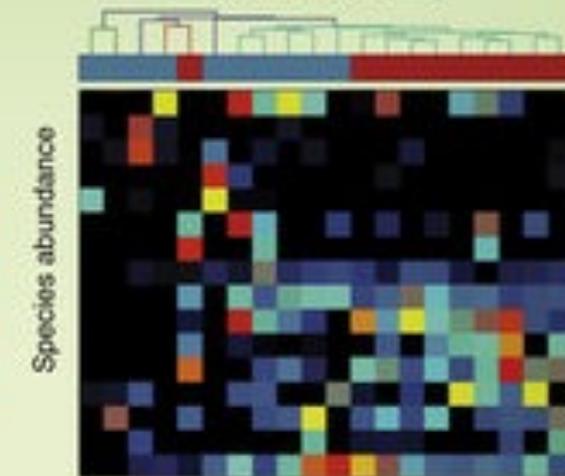
MAGs



Functional modules

Read-based profiling

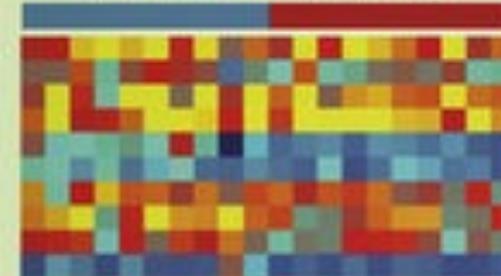
Taxonomic profiling



Species abundance

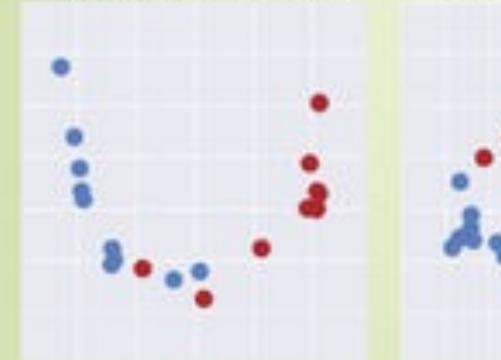
Sample

Functional potential profiling

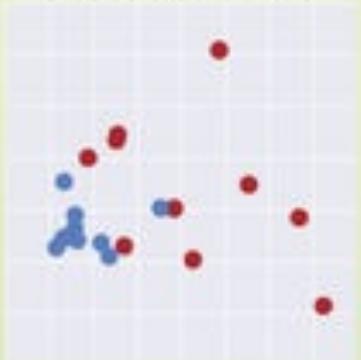


Samples

Ordination on functions

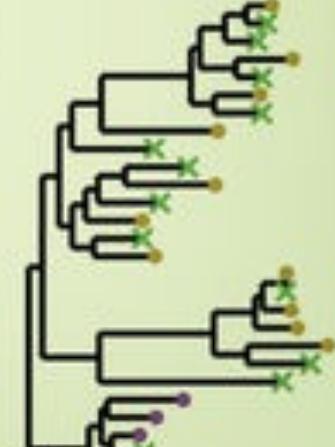


Ordination on taxa



■ Case samples

■ Control samples



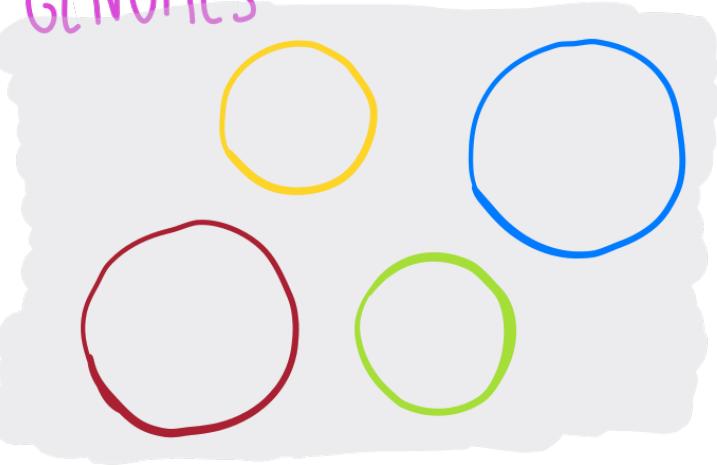
Sample

Strain-level phylogenies



Reconstructing genomes from metagenomes

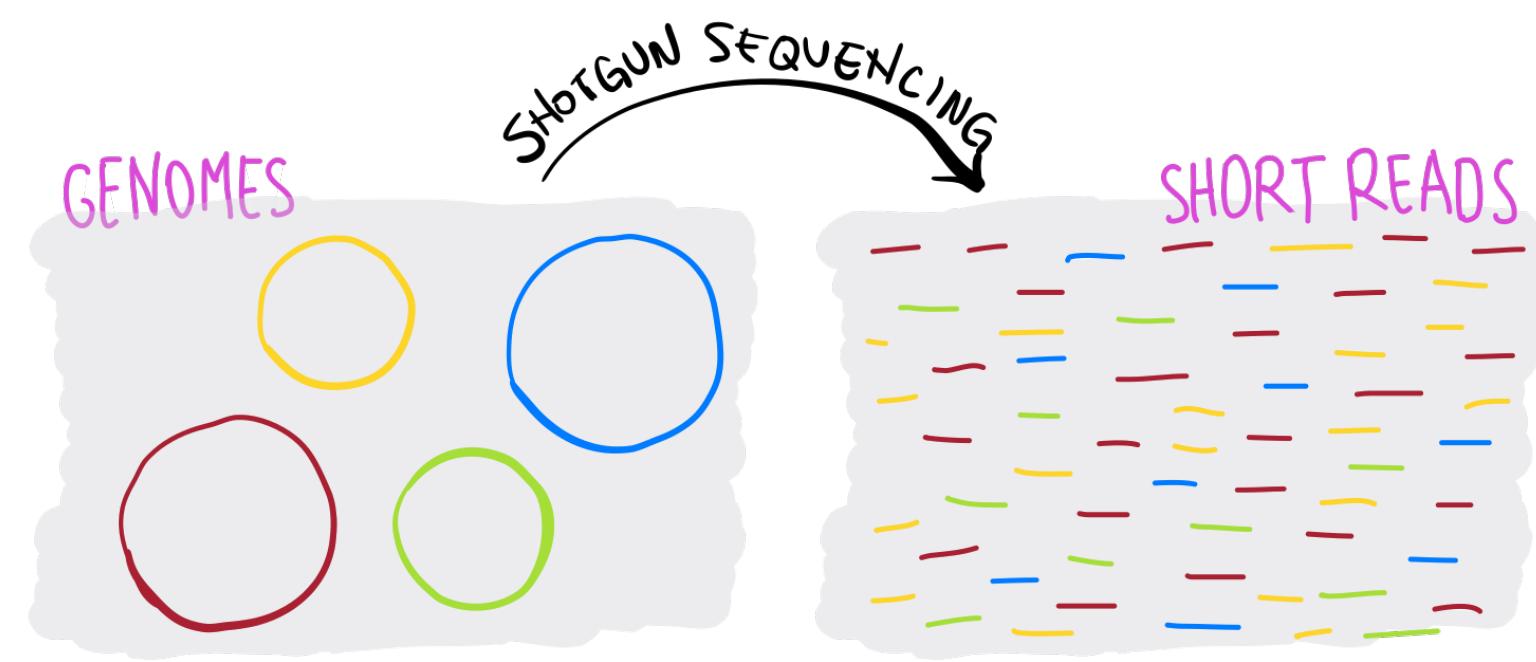
GENOMES



GENOMES

SHOTGUN SEQUENCING

SHORT READS



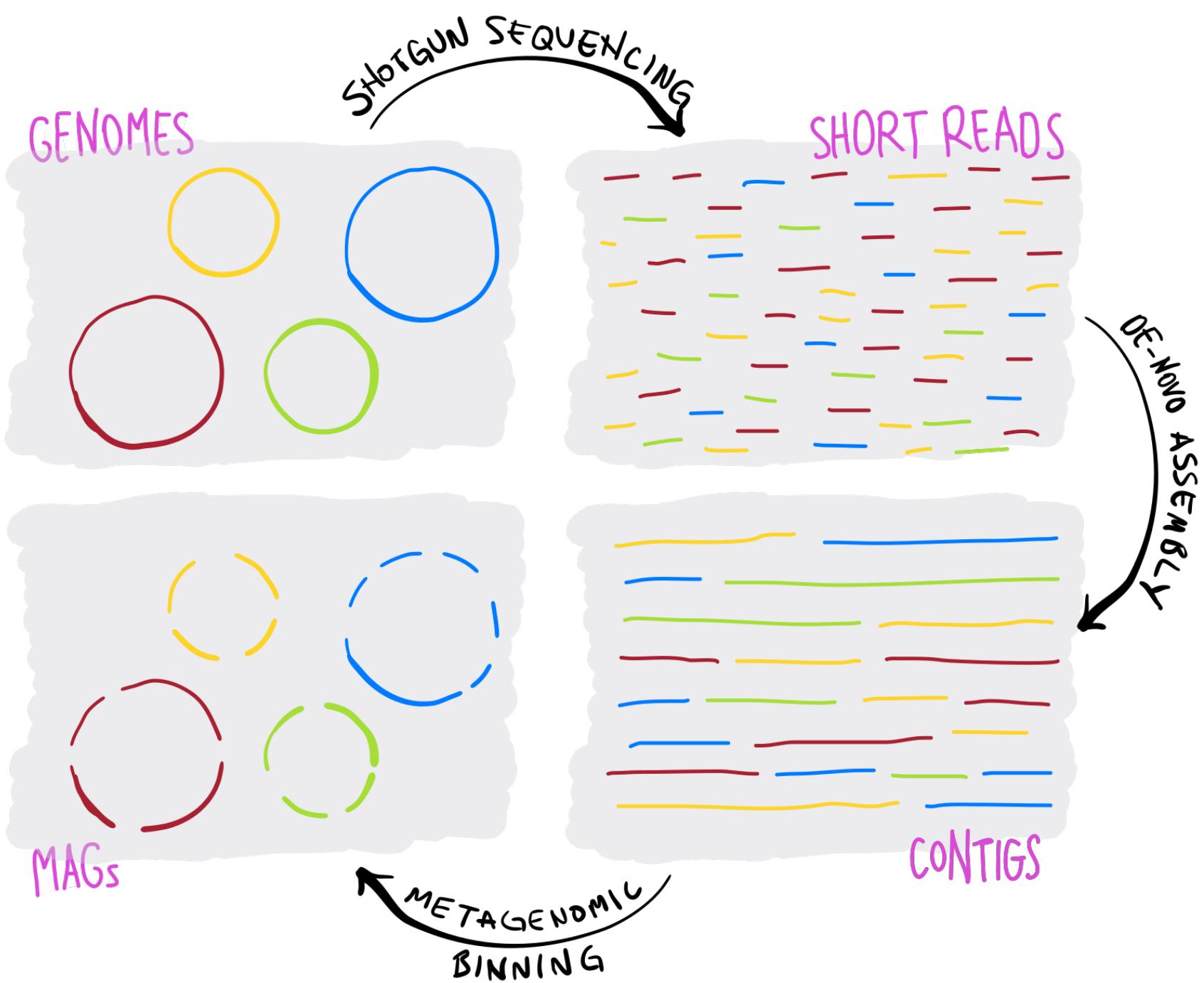
GENOMES

SHOTGUN SEQUENCING

SHORT READS

DE-NOVO ASSEMBLY

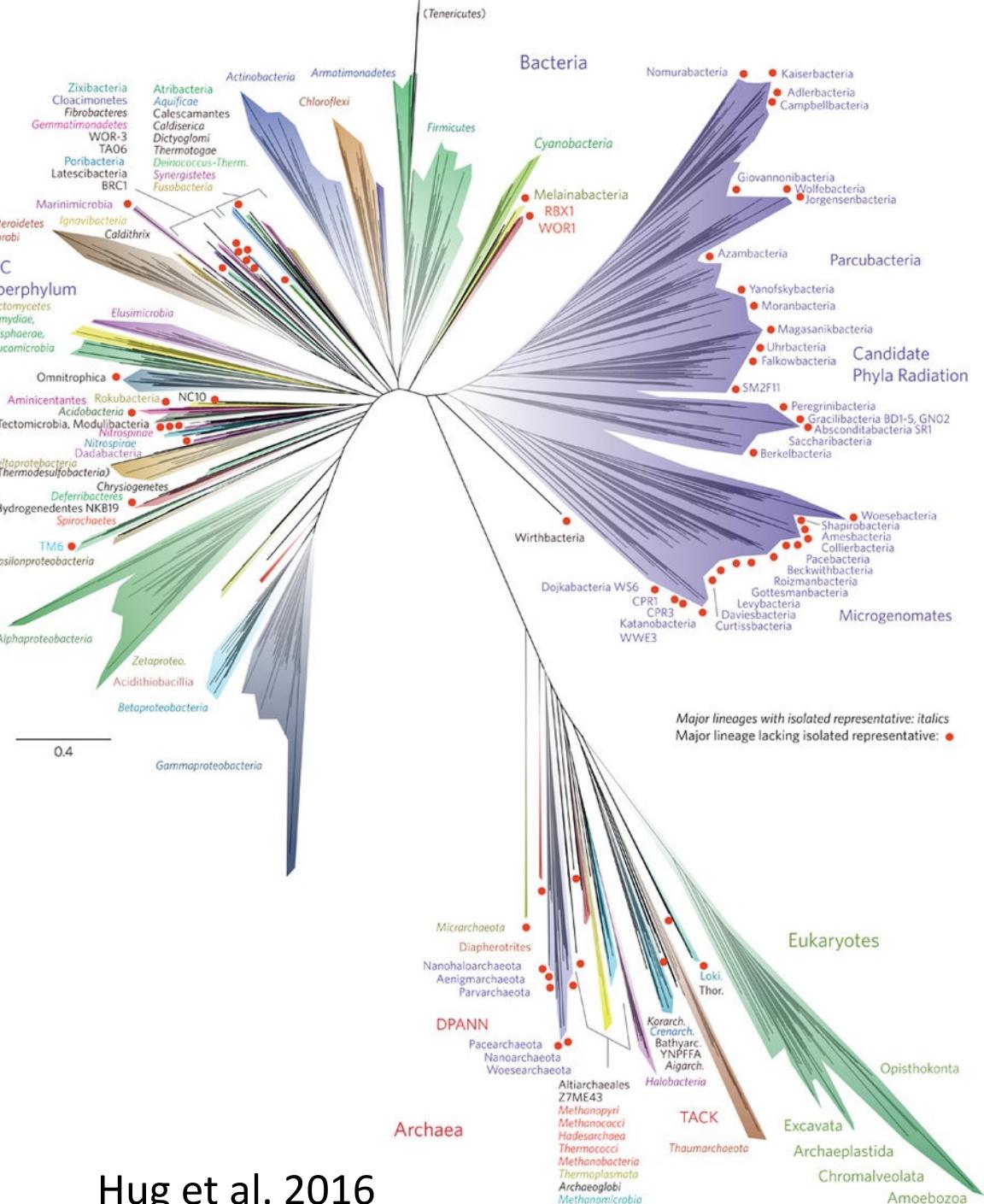
CONTIGS



Same
community
without
cultivation of
the members

Omics

- Metagenomics
- Metatranscriptomics
- Metaproteomics
- Meta-metabolomics

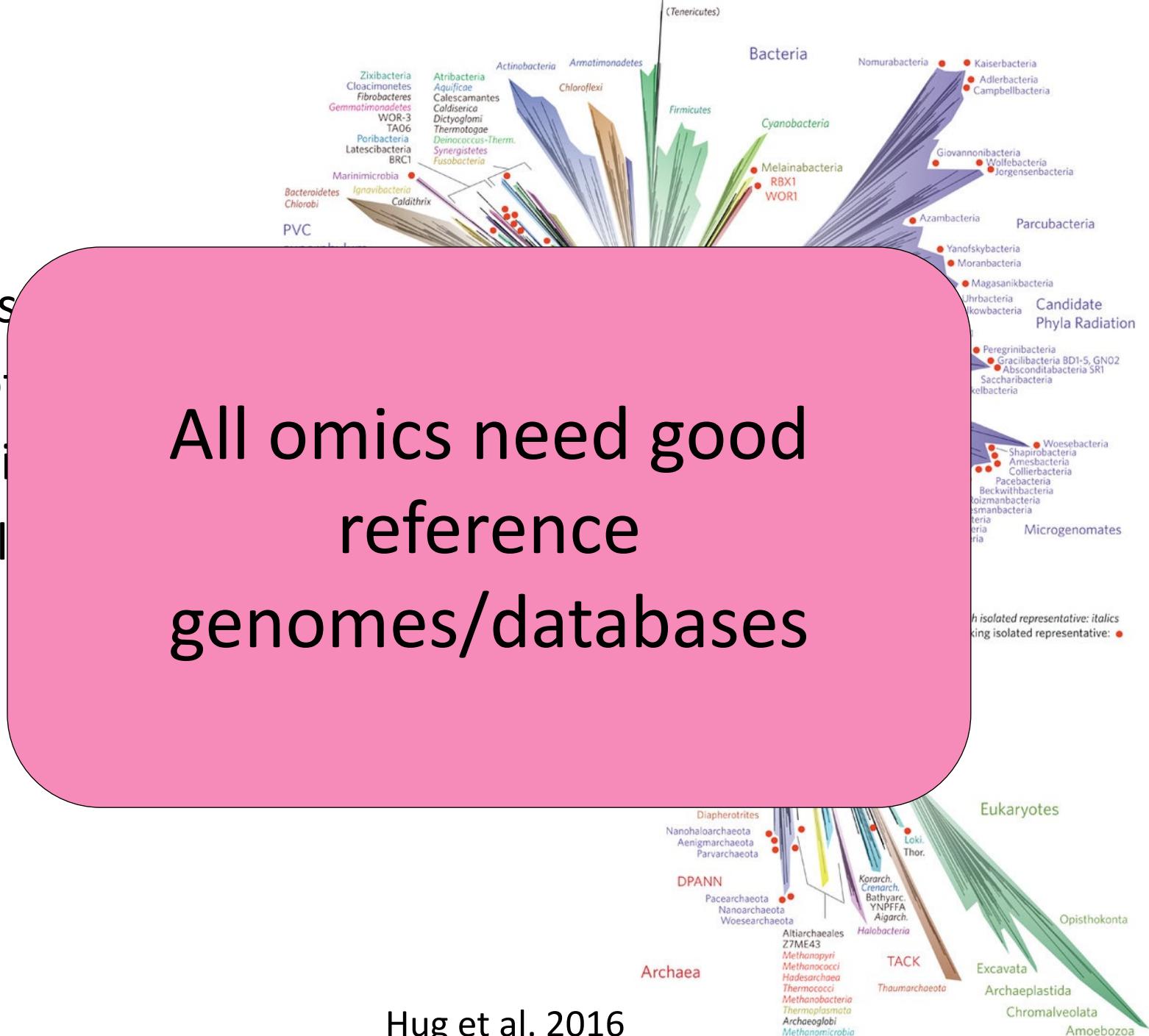


Omics

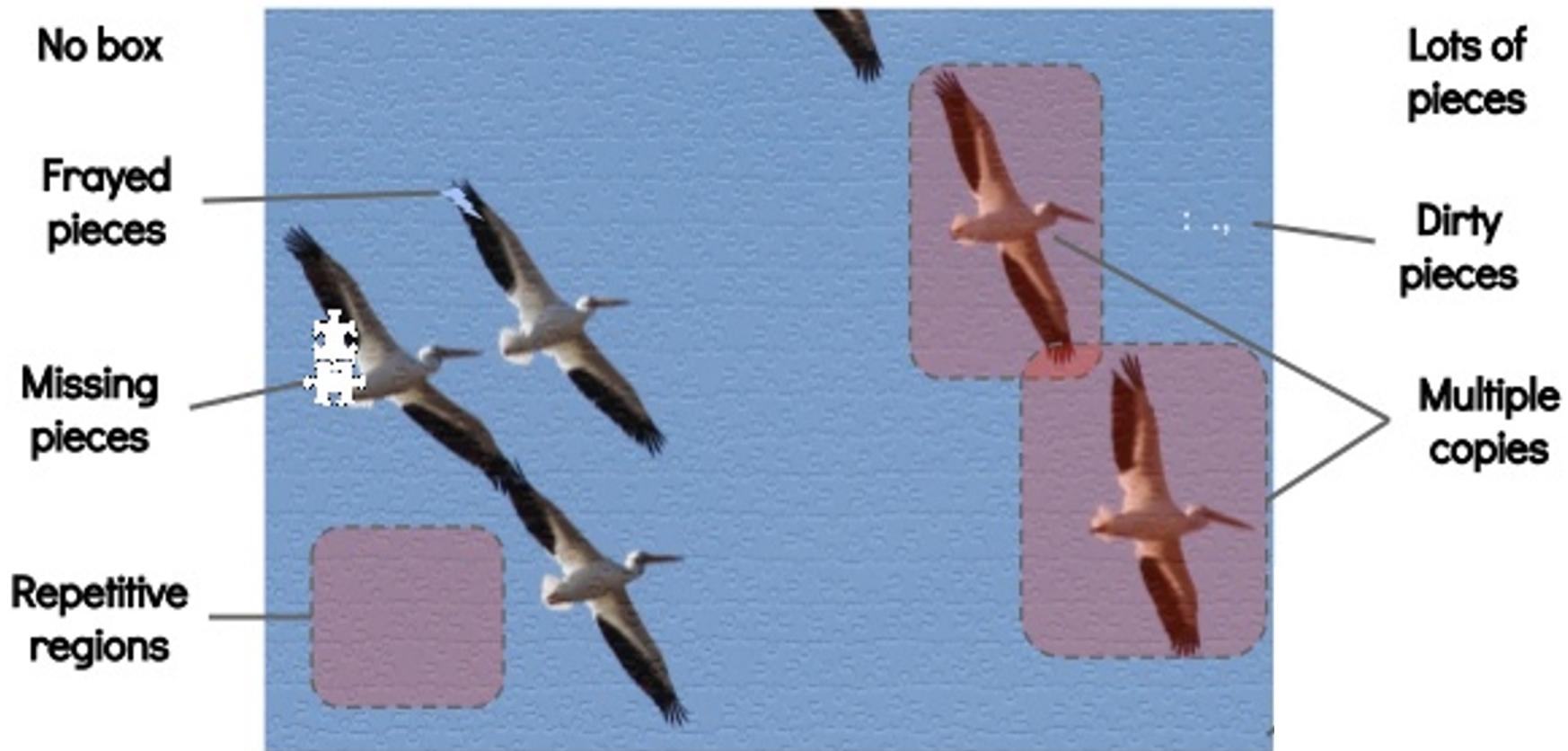
- Metagenomics
- Metatranscriptomics
- Metaproteomics
- Meta-metabolomics

All omics need good reference genomes/databases

Hug et al. 2016



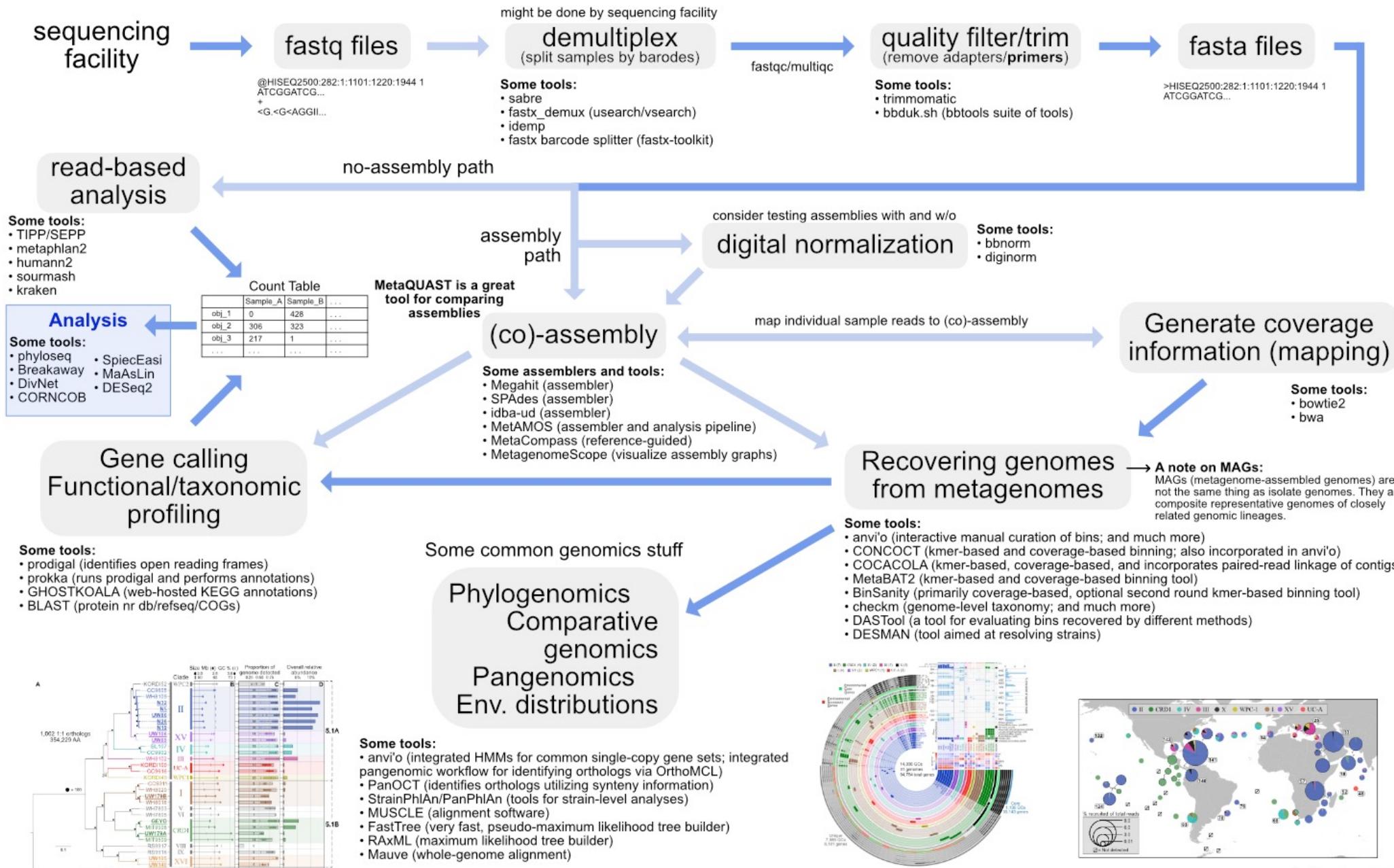
What makes a puzzle hard?



Overview of generic* metagenomics workflow

*This is generic; specific workflows can vary on the order of steps here and how they are done.

When working with your own data you should never follow any pipeline blindly. There can be critical differences based on your data.



Thank you!

jenni.hultman@luke.fi



twitter.com/jennihultman

