



CENTER ONE HEALTH RUHR

UA RUHR | RESEARCH ALLIANCE



Supercomputing propelling research in microbiology

Prof. Dr. Alexander Probst, Environmental Metagenomics, Research Center One Health Ruhr

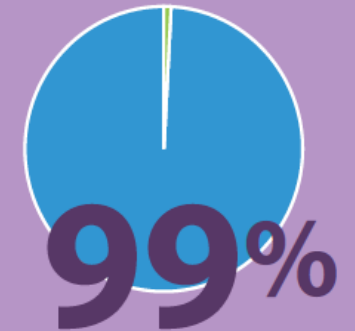
Size of microbes and their relation to humans



3 PINTS = VOLUME
of the microbiome



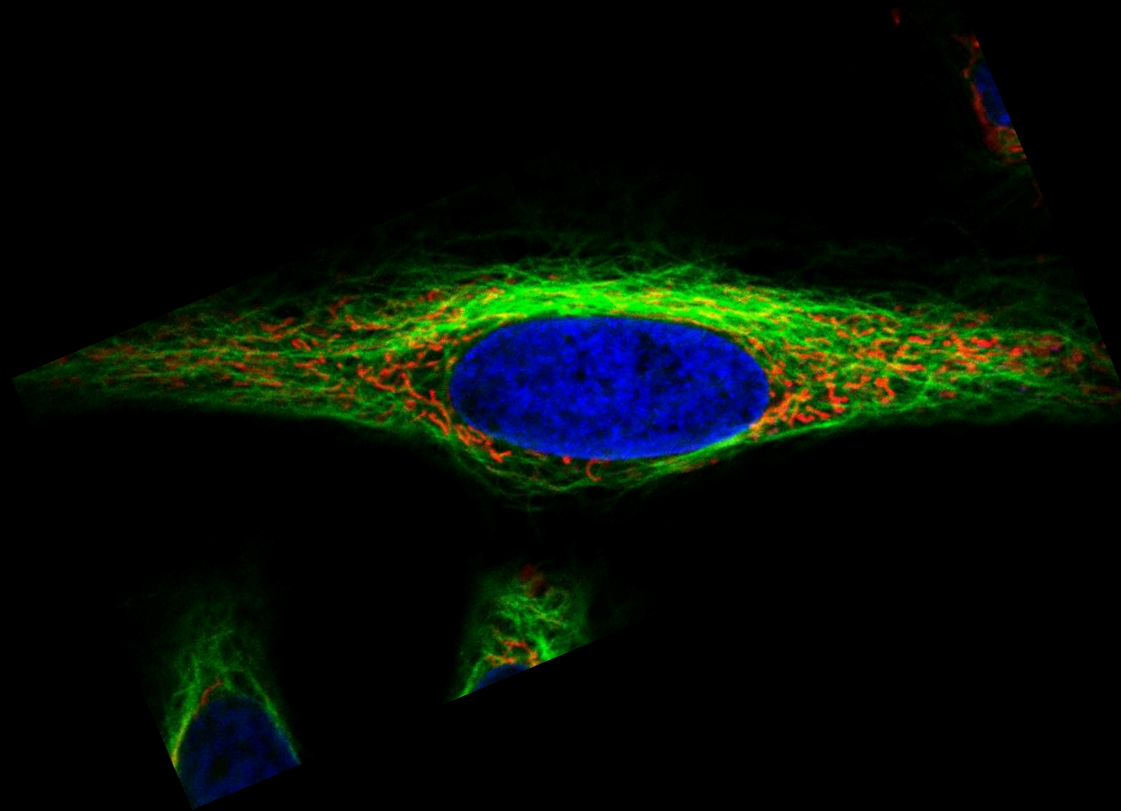
2.5 LBS = WEIGHT
of the microbiome



Microbes contribute an extra
2,000,000 genes to the 20,000 gene
human genome.

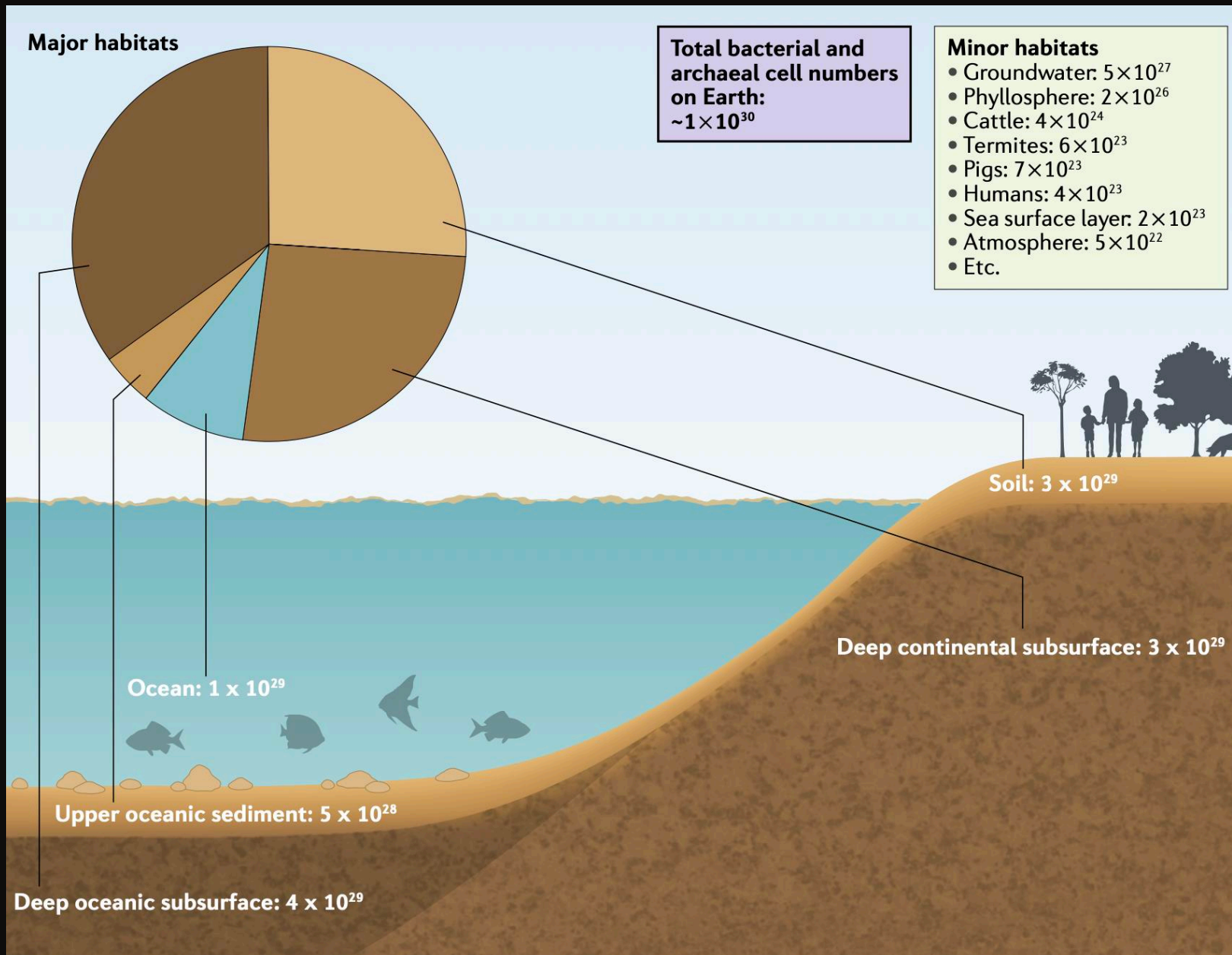
**We are as equally microbial as
we are human.**

Size of microbes and their relation to humans



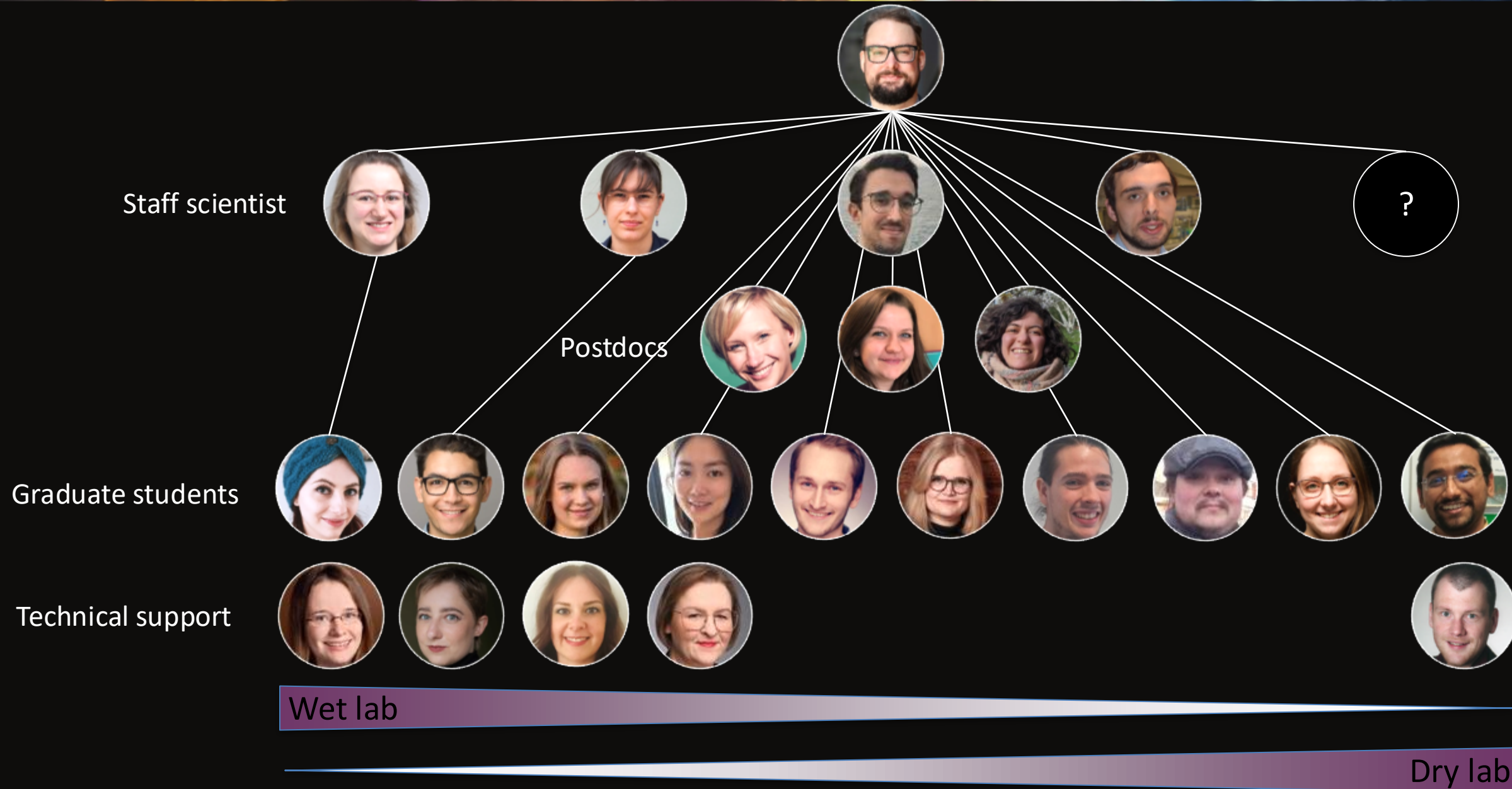
https://commons.wikimedia.org/wiki/File:Multicolor_fluorescence_image_of_a_living_HeLa_cell.jpg

Size of microbes and their role on Earth



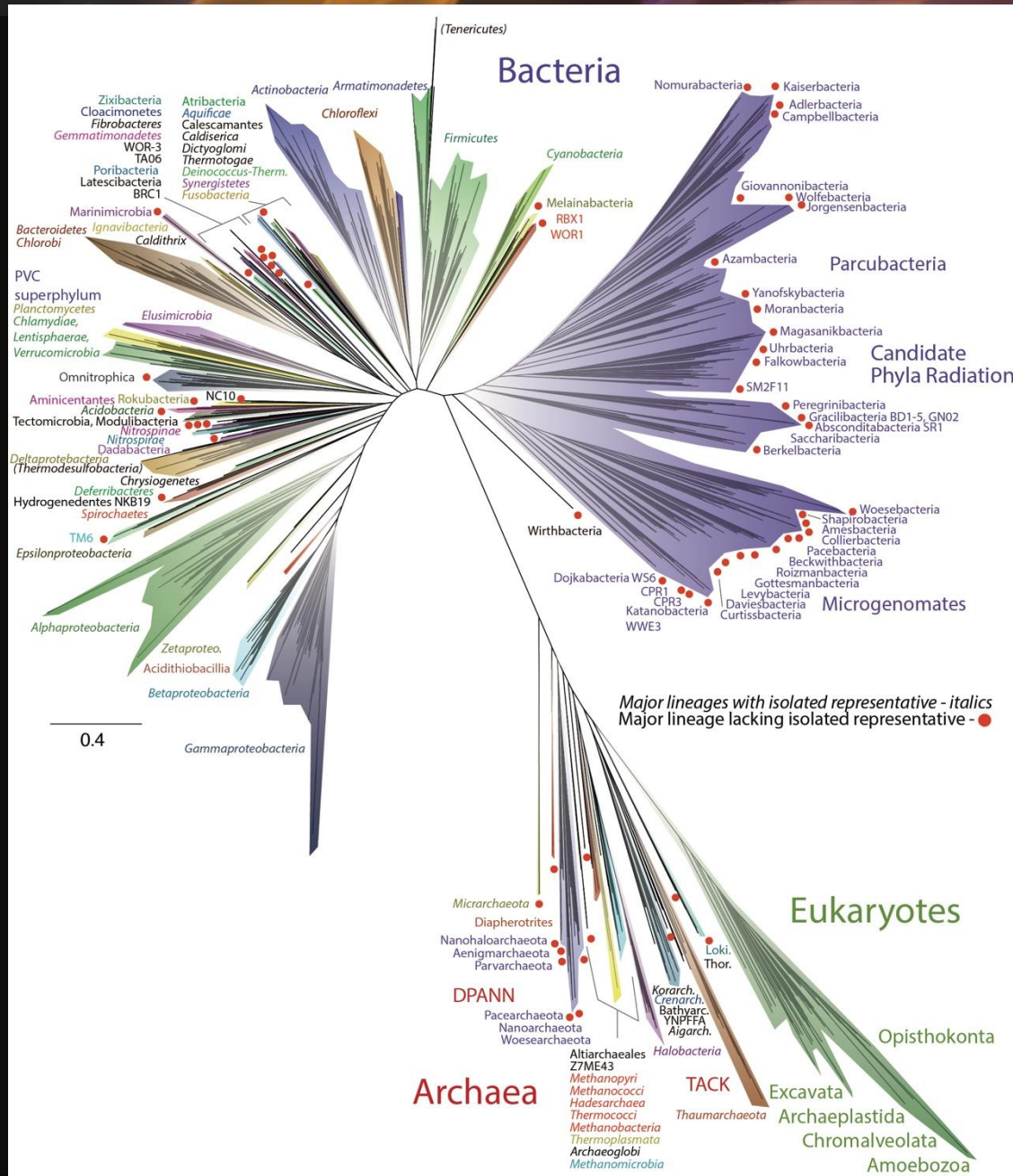
- $\sim 2/3$ of all microbes live in the deep subsurface
- $\sim 1/3$ of all microbes live in the deep continental subsurface
- Accessing samples and contamination control are tough
- Deep biosphere as the final frontier in microbiology

Nature Reviews Microbiology 2019 17, 247–260



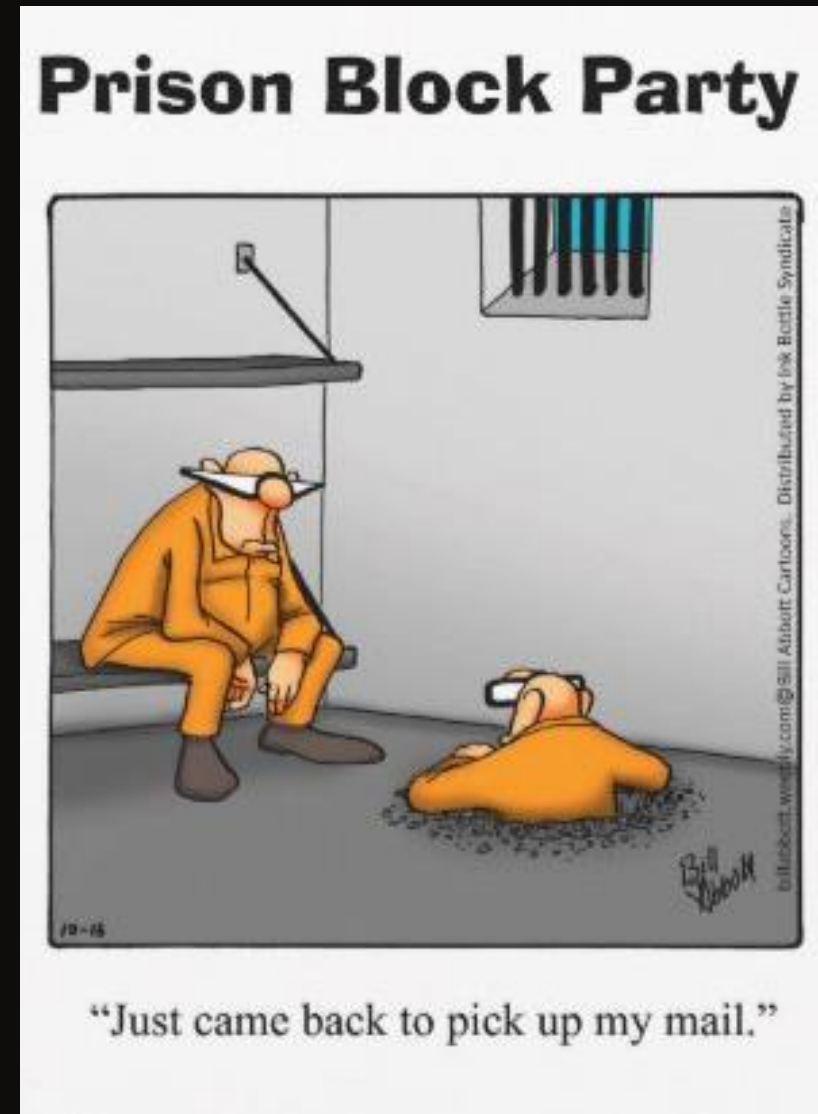
Projects in the department





- 99-99.9% of all bacteria and archaea remain uncultivated under laboratory conditions
- Study of laboratory cultures does not reflect environmental behaviour

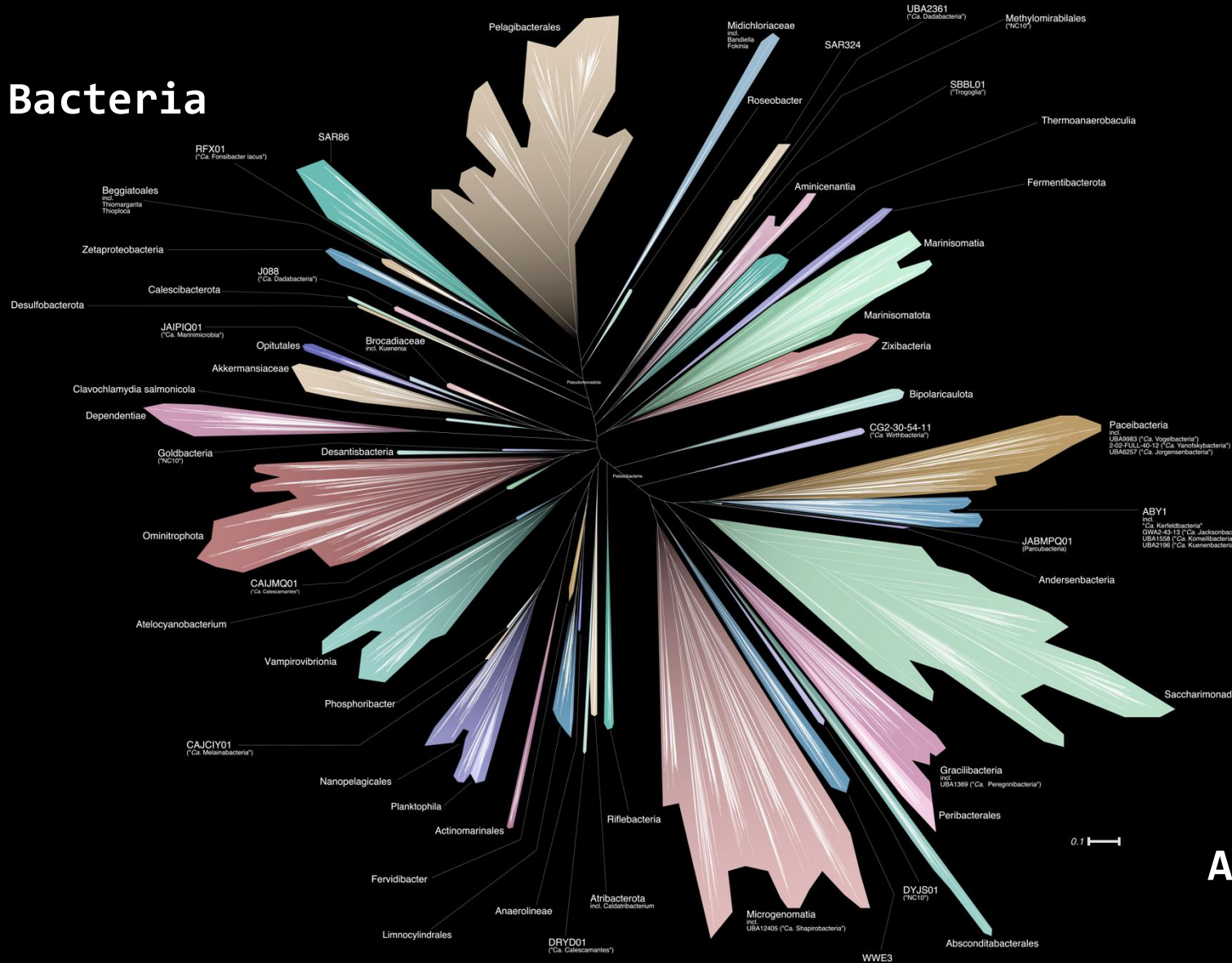
Hug et al., Nature Microbiology 2016



Aquatic ecosystems most needed microbes



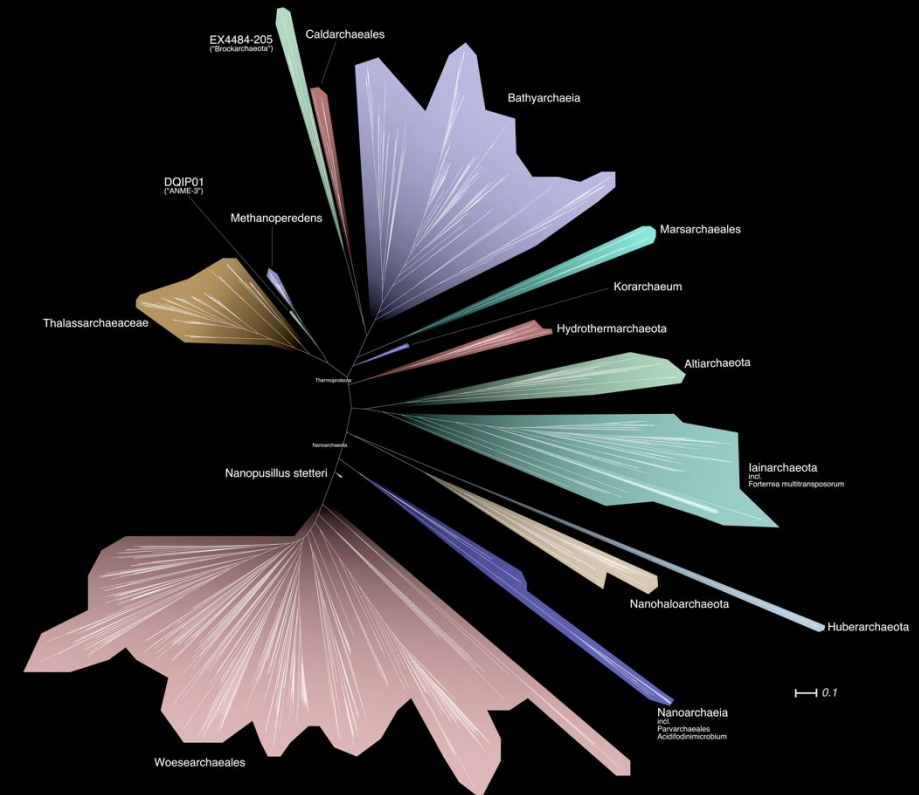
Bacteria



Less than 1% of all microbes remain cultivated.

Aquatic microbiomes' most wanted comprise 84 taxa

Archaea



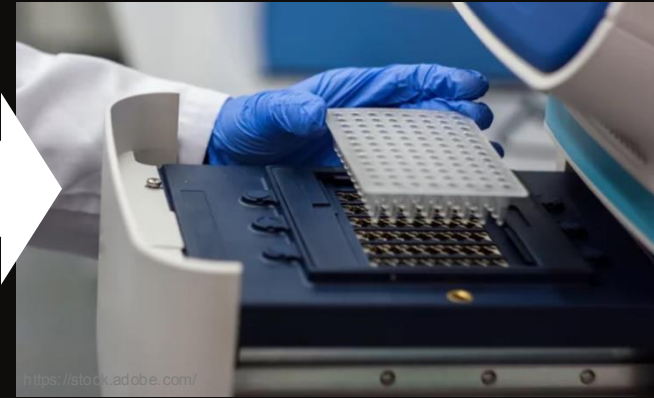
How to analyze microbial communities quickly



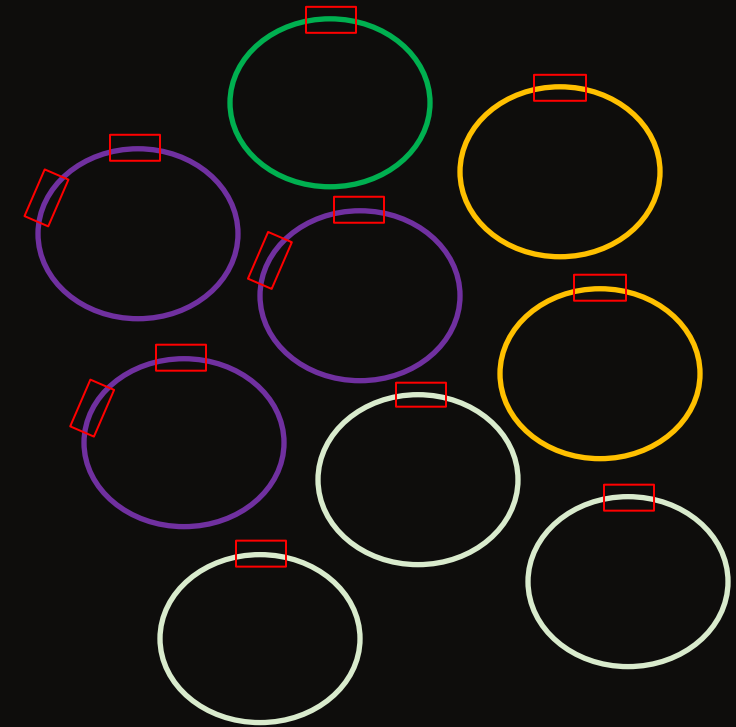
sampling



DNA extraction



PCR



Easy to apply, cheap, quick results...

Analogy to forensics



Marker gene analyses

- Not only one individual, multiple individuals at the same time
- High throughput
- Screening of hundreds of samples
- Cost effective compared to other methods
- No need to wait for microbes to grow in the lab
- Data is easy to process

Marker genes answer the question: Who of the ones we know are there?

Analogy to forensics

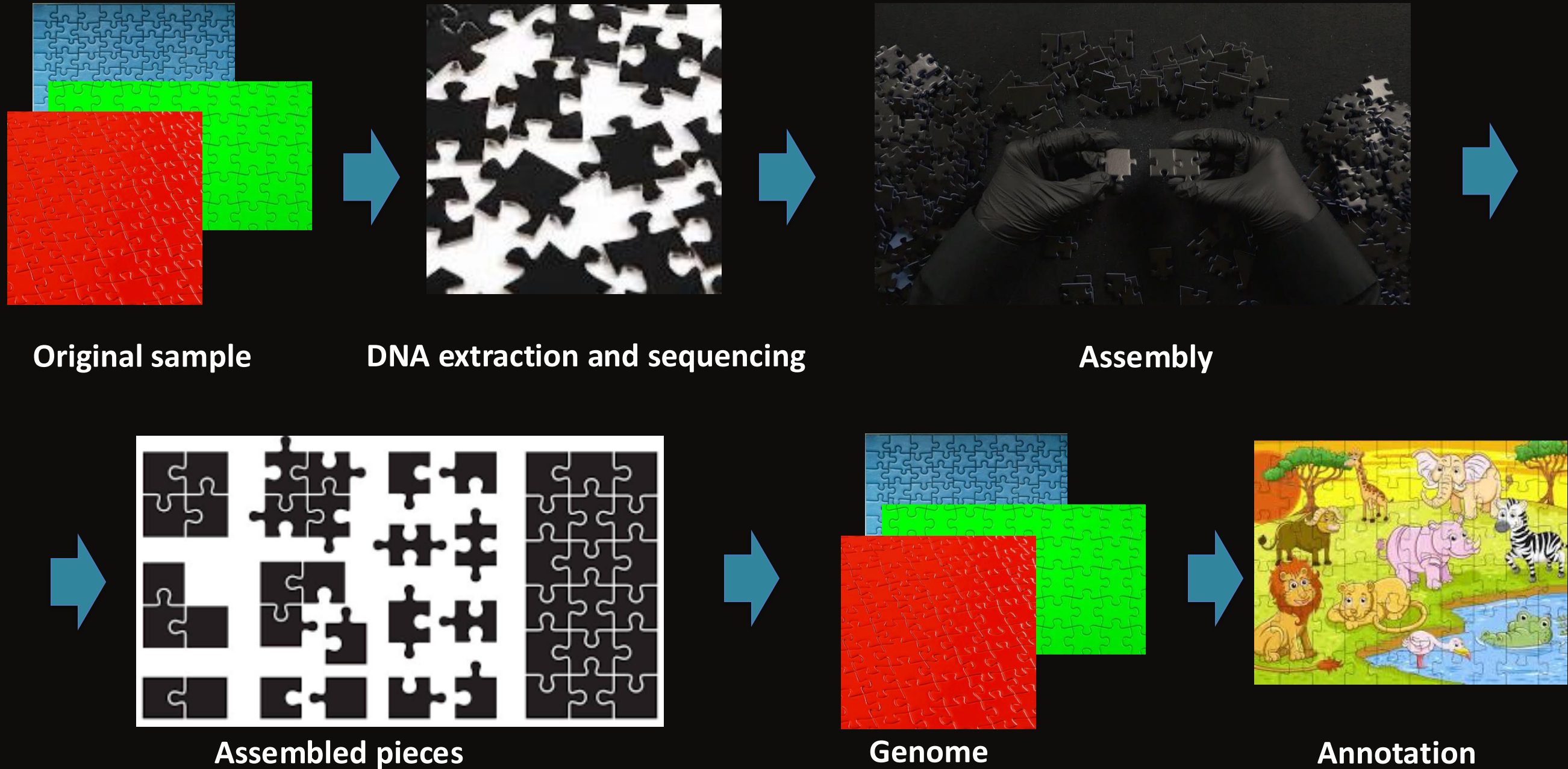


Genomics, sequencing of all genes, in analogy to forensics

- Identifying characteristics of the known and unknown
- Characteristics are:
 - sex
 - eye color
 - hair color
 - age
 - ethnicity
 - diseases
 - ...

Metagenomics answers the question: Who is there and what characteristics do these microbes have?

Metagenomes – puzzling information or a puzzle of information?



Sequencing DNA...



DNA extraction and sequencing

1st gen
Sanger



500-1000 bps

individual sequencing

high accuracy

2nd gen
Illumina



100-500 bps

standard, cheap

3rd gen
Pacific Biosciences Oxford Nanopore



<60'000 bps

massive parallel sequencing

lower accuracy



<4'700'000 bps

emerging

Human genome: 3.2 Gbps

E. coli: 4.7 Mbps

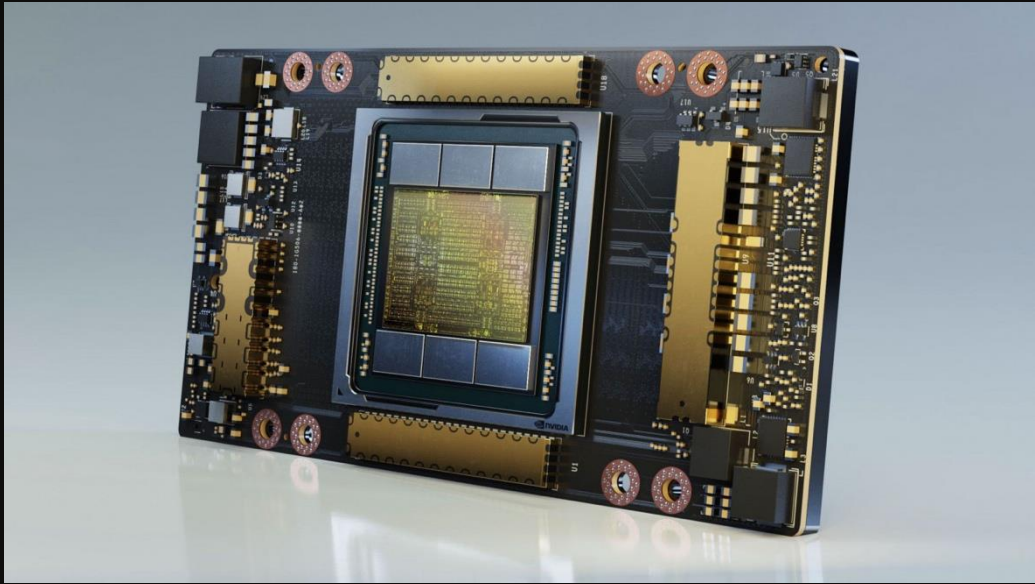
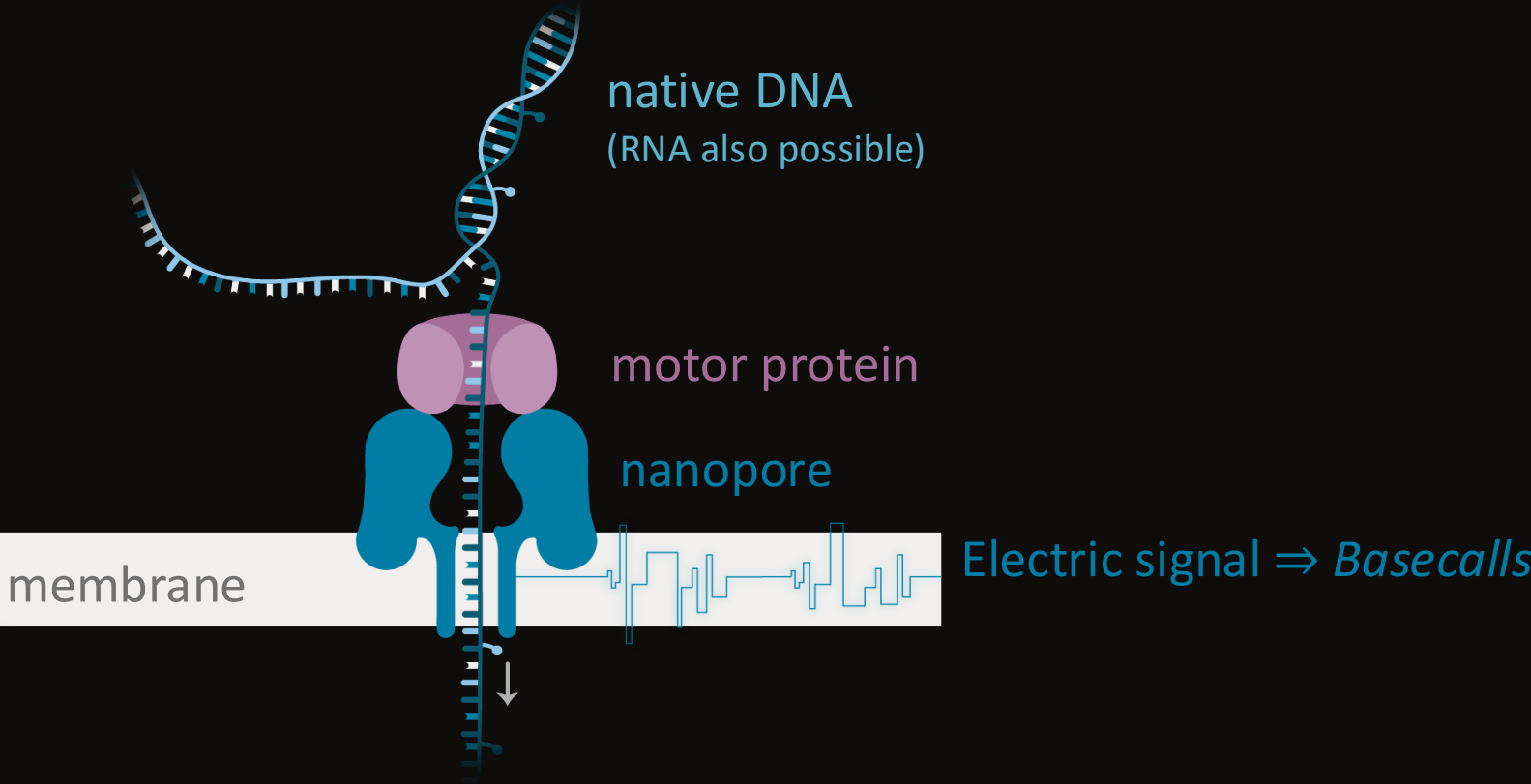
Covid-19: ~30 kbps

Result is a DNA sequence consisting of adenine (A),
guanine (G), cytosine (C), and thymine (T).



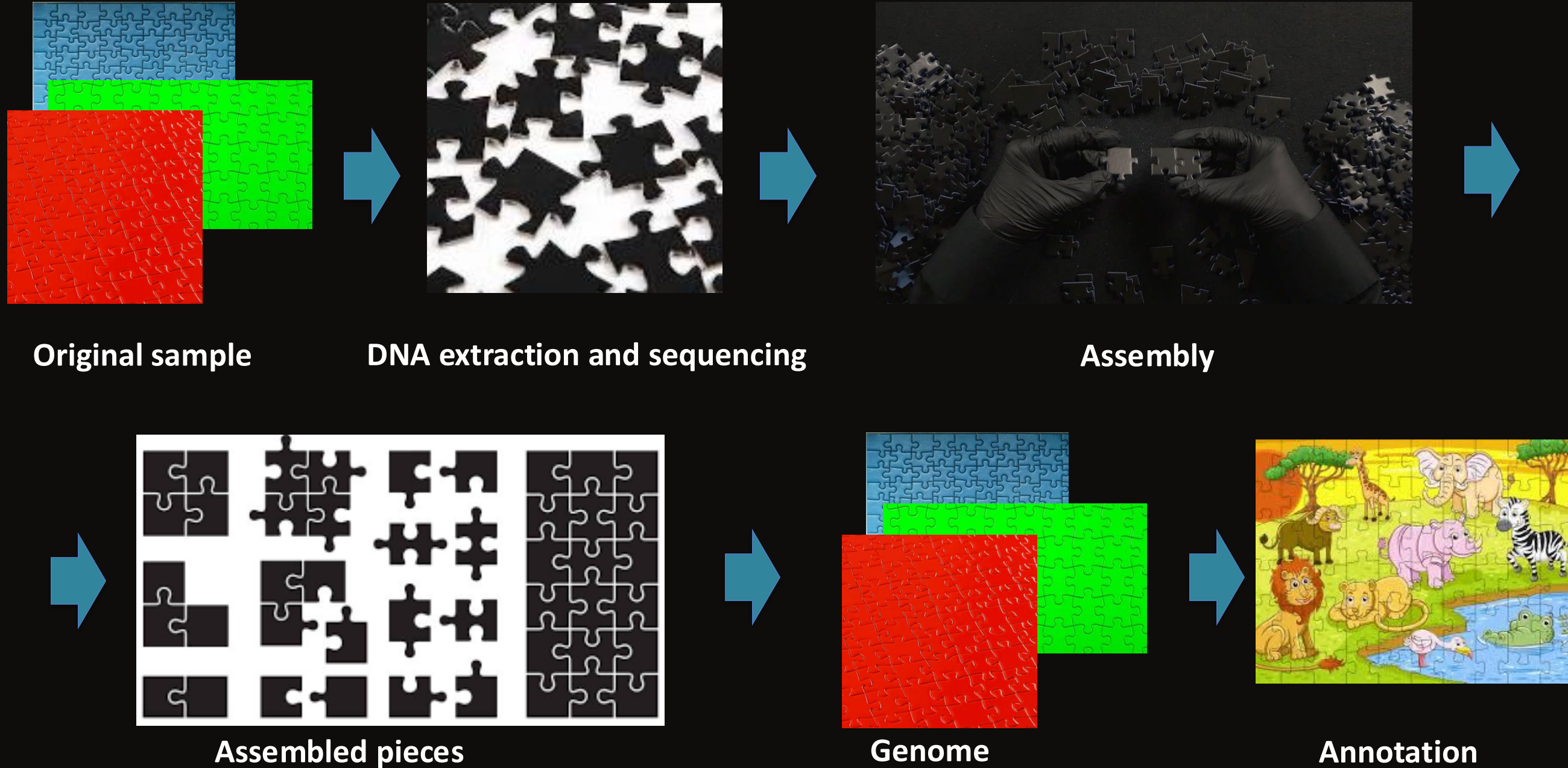
© ONT, nanoporetech.com/about/for-the-media [20.08.2024]

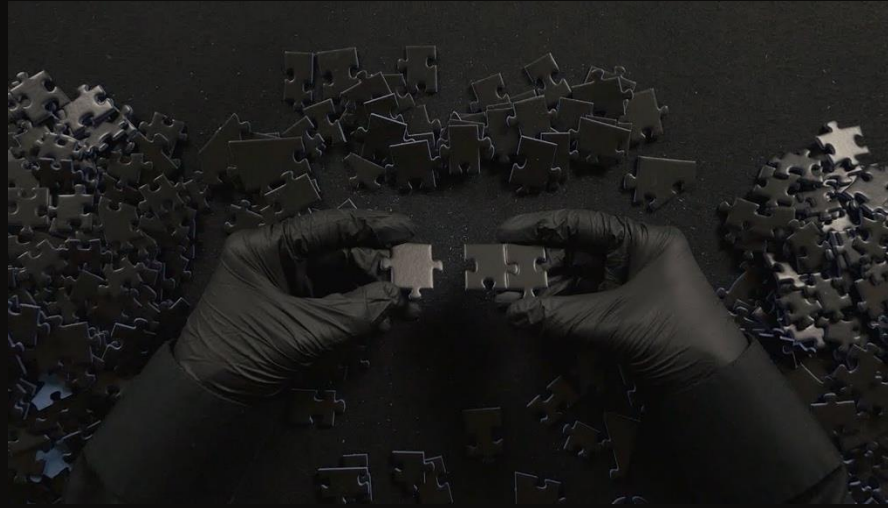
Calling bases for Oxford Nanopore Technology



© ONT; nanoporetech.com/about/for-the-media [20.08.2024]

Metagenomes – puzzling information or a puzzle of information?





Assembly

it was the best of times it was the worst of times

it was the best

was the best of

the best of times

best of times it

of times it was

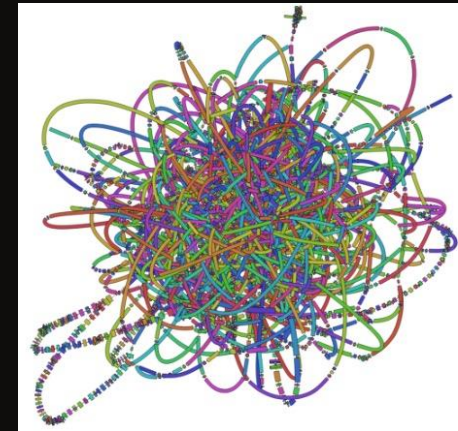
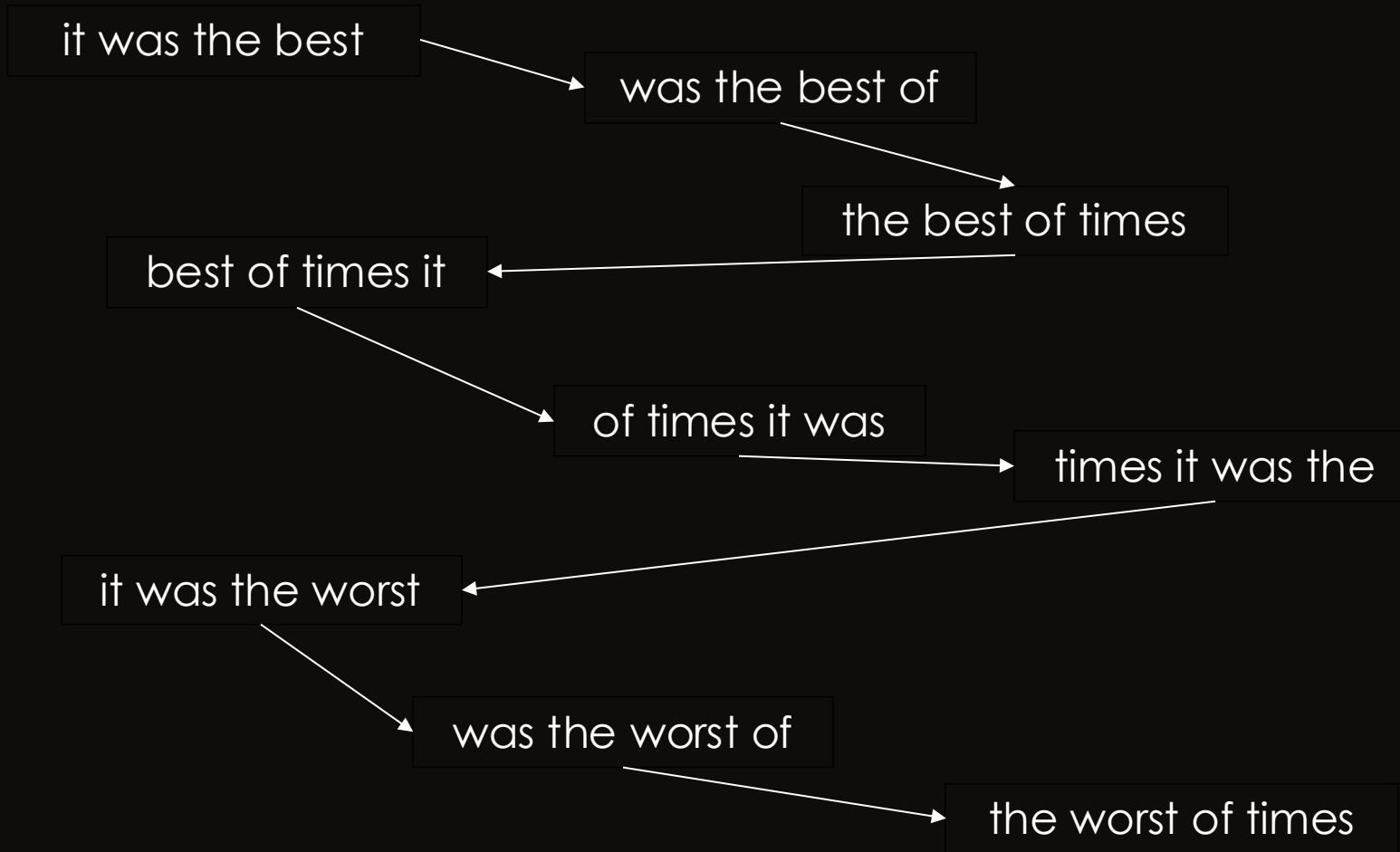
times it was the

it was the worst

was the worst of

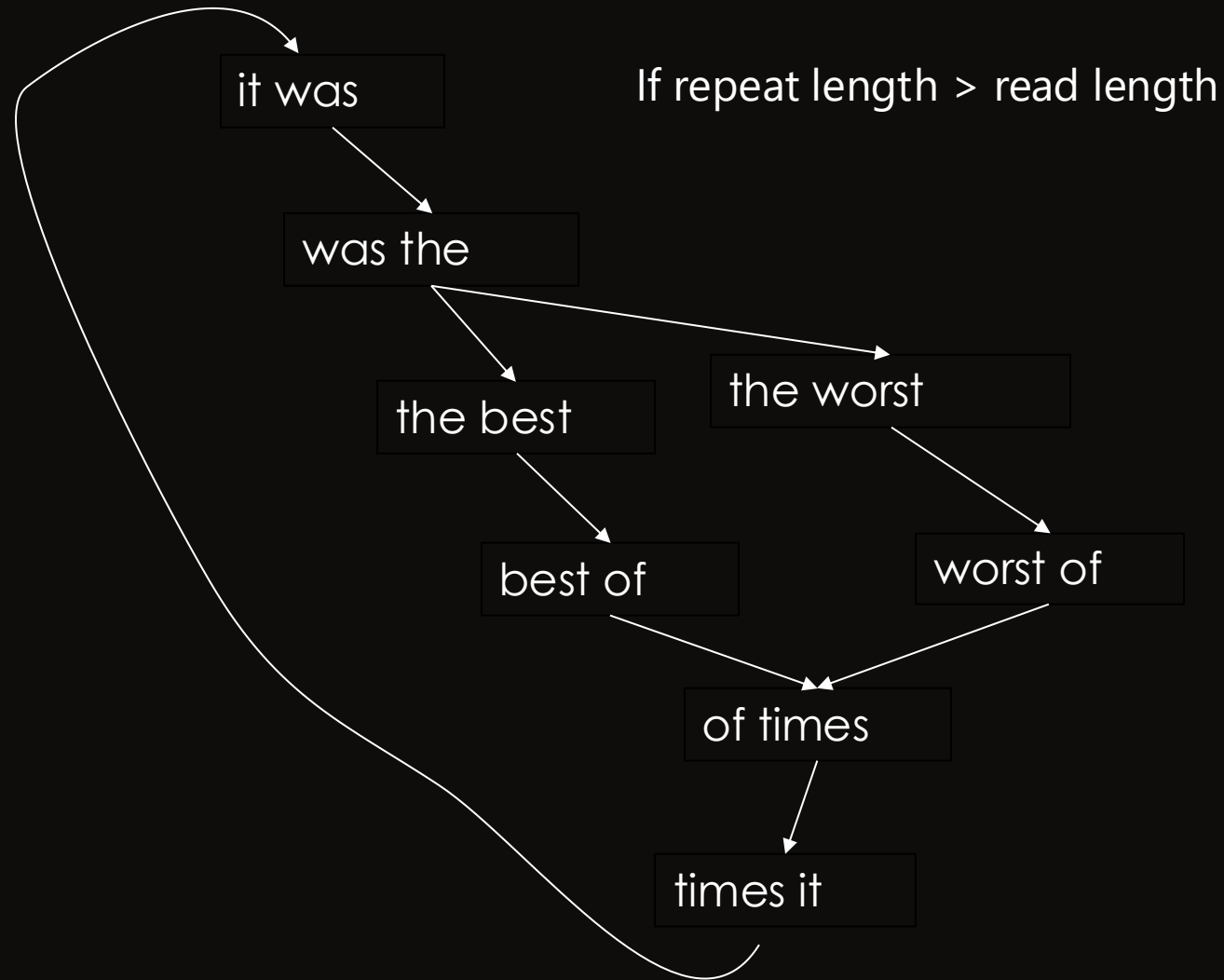
the worst of times

Unambiguous ways of assembly graphs needed



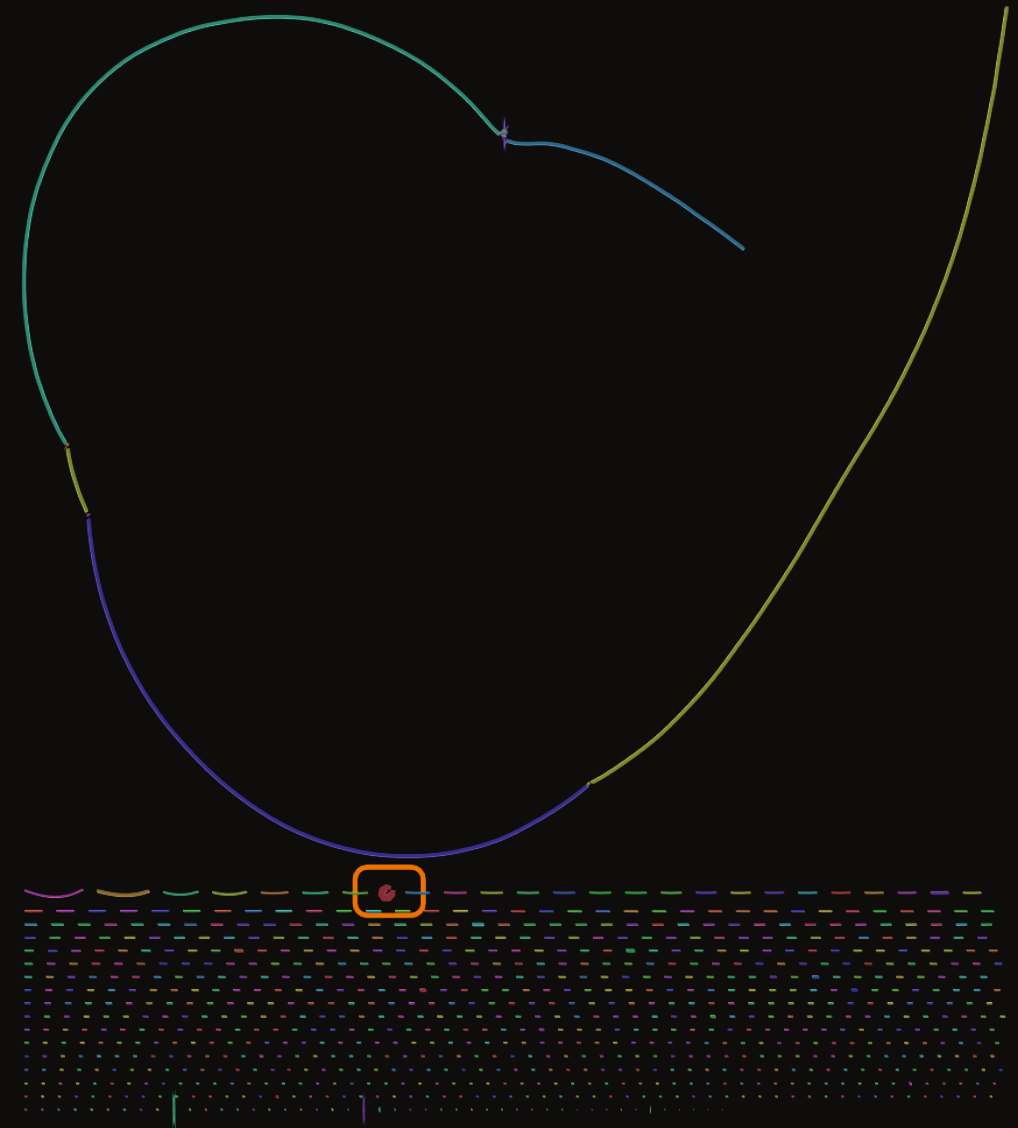
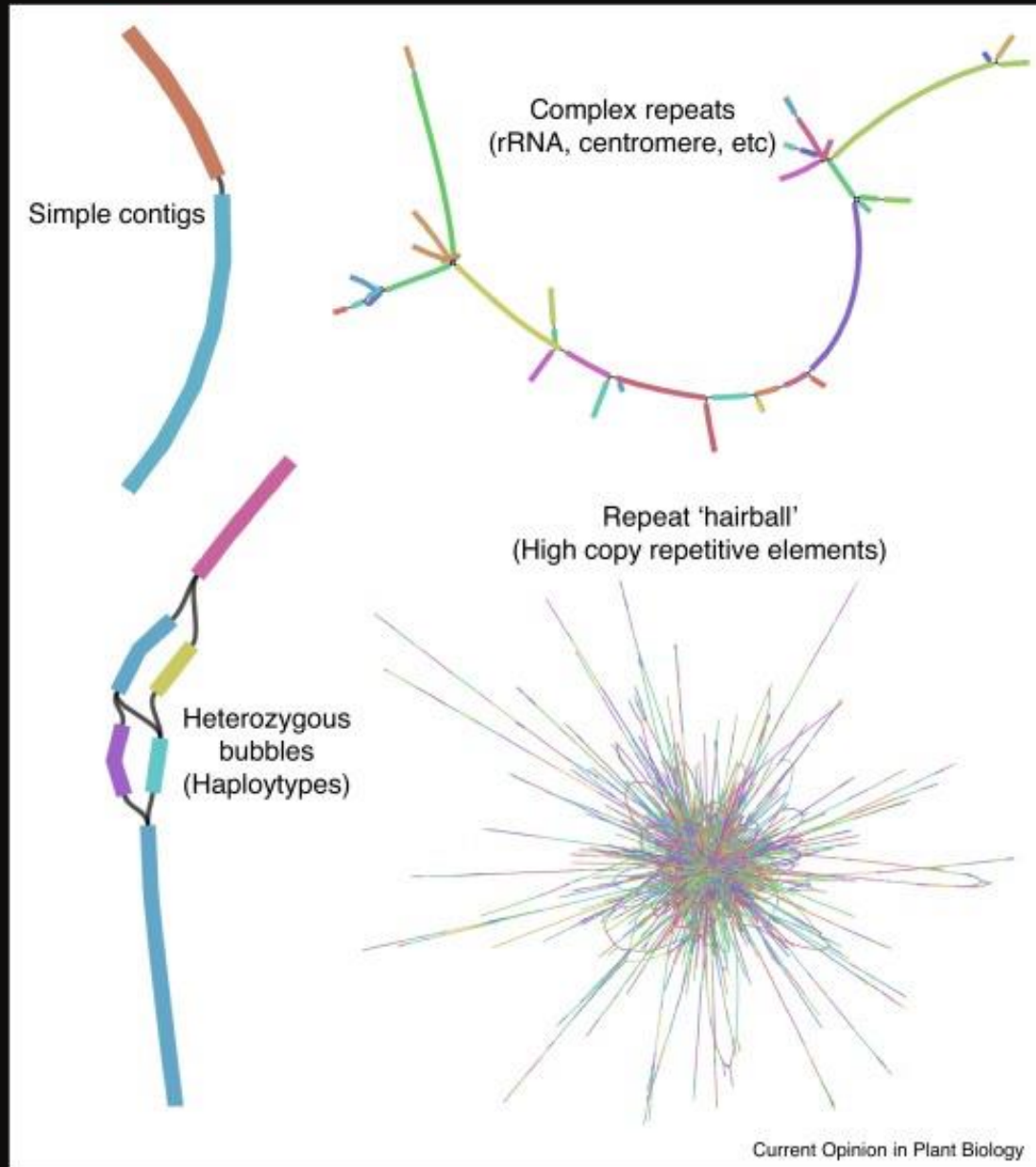
"assembly graphs"

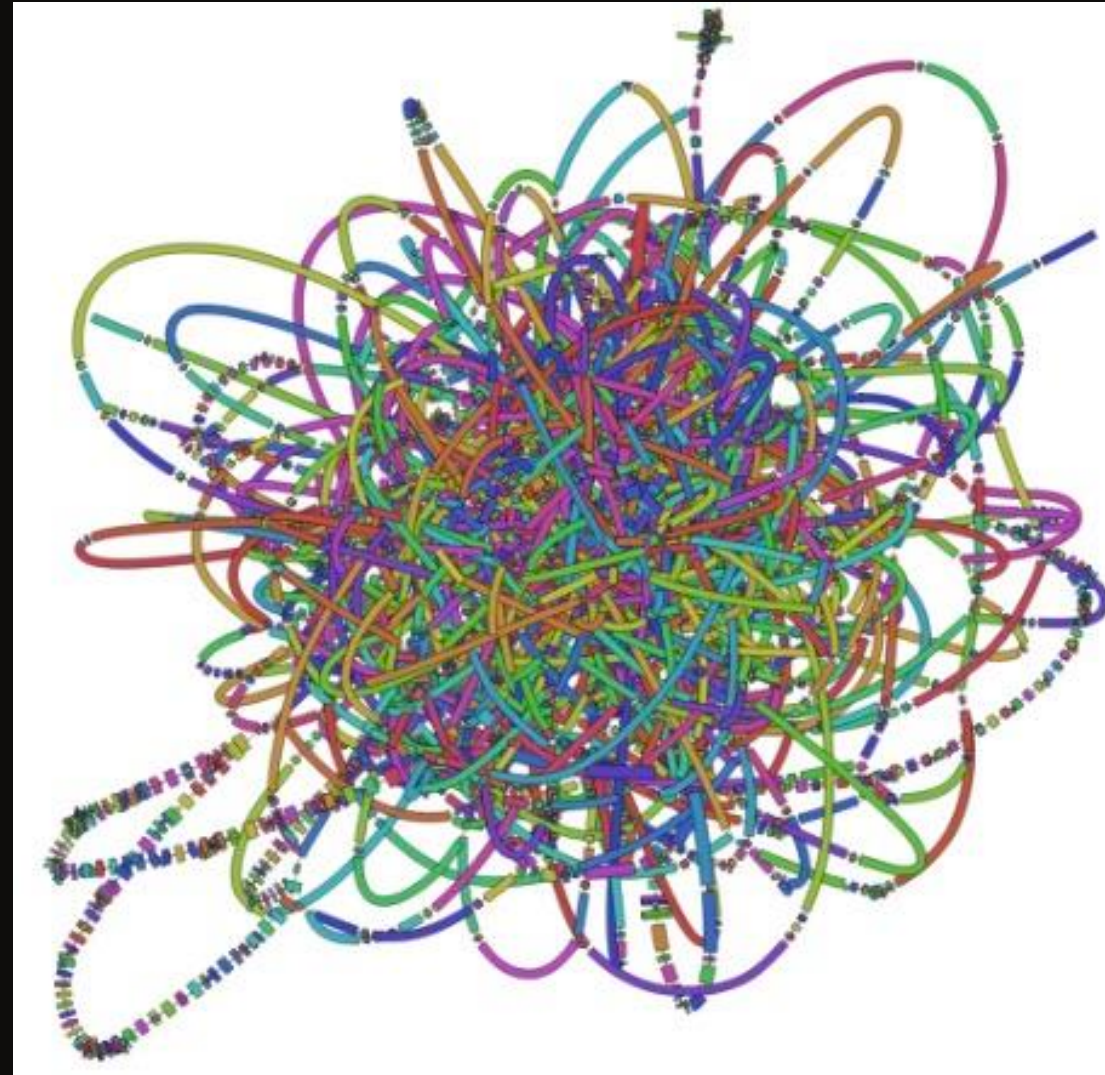
Size of metagenomes for de novo assembly is ~30-100 Gbps. Standard is short read assemblies with 150 bps per read resulting in 200 Mio DNA pieces to be assembled!



Repetitive sequences are the core problem but biology loves repetition: Copying (and altering) is easier than invention.

Complexity of assembly graphs





environmental
microbiology



Environmental Microbiology (2010) 12(7), 1806–1810

doi:10.1111/j.1462-2920.2010.02201.x

Opinion

Replicate or lie

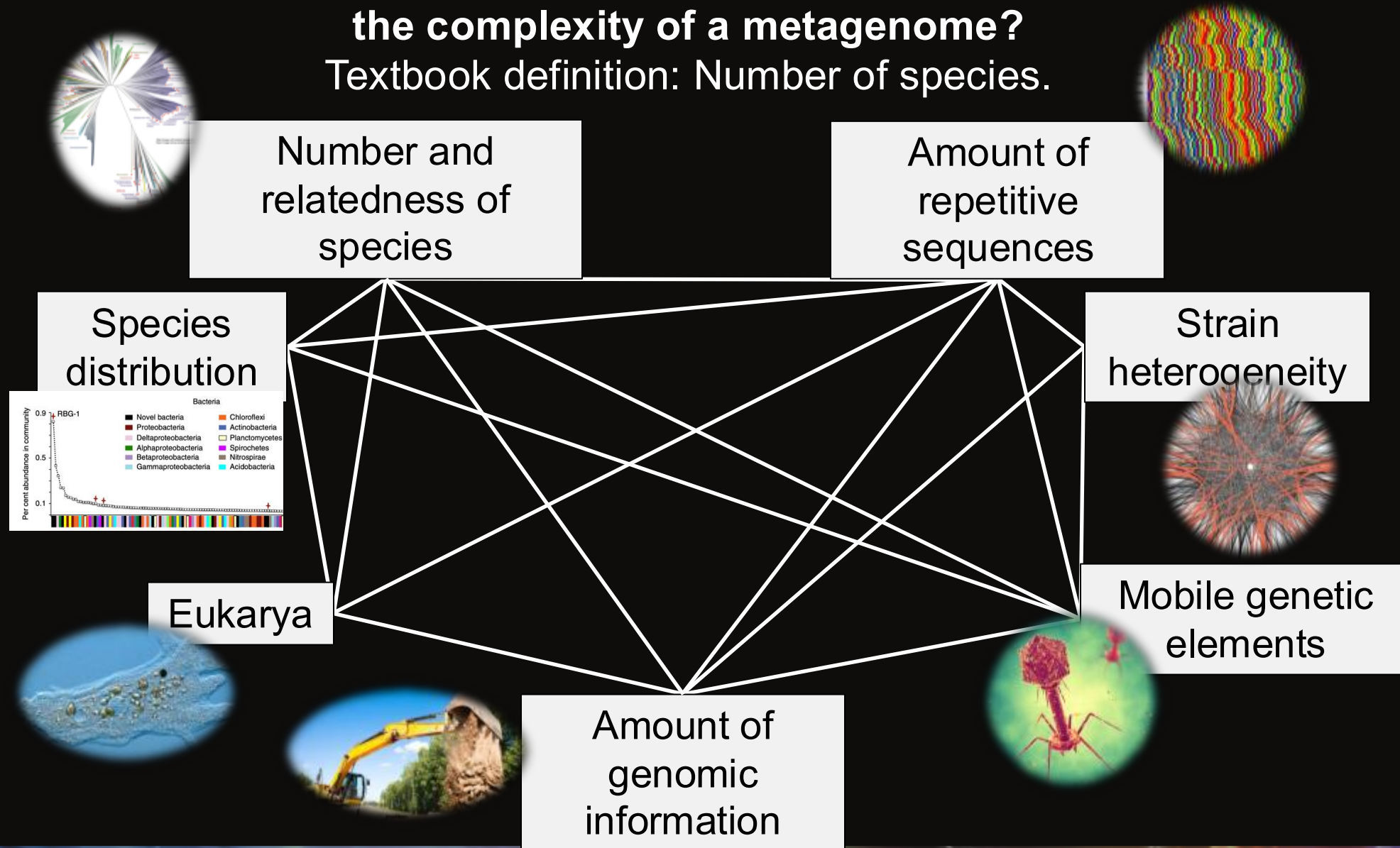
James I. Prosser*
*Institute of Biological
University of Aberdeen
Machar Drive, Aberdeen*

Introduction

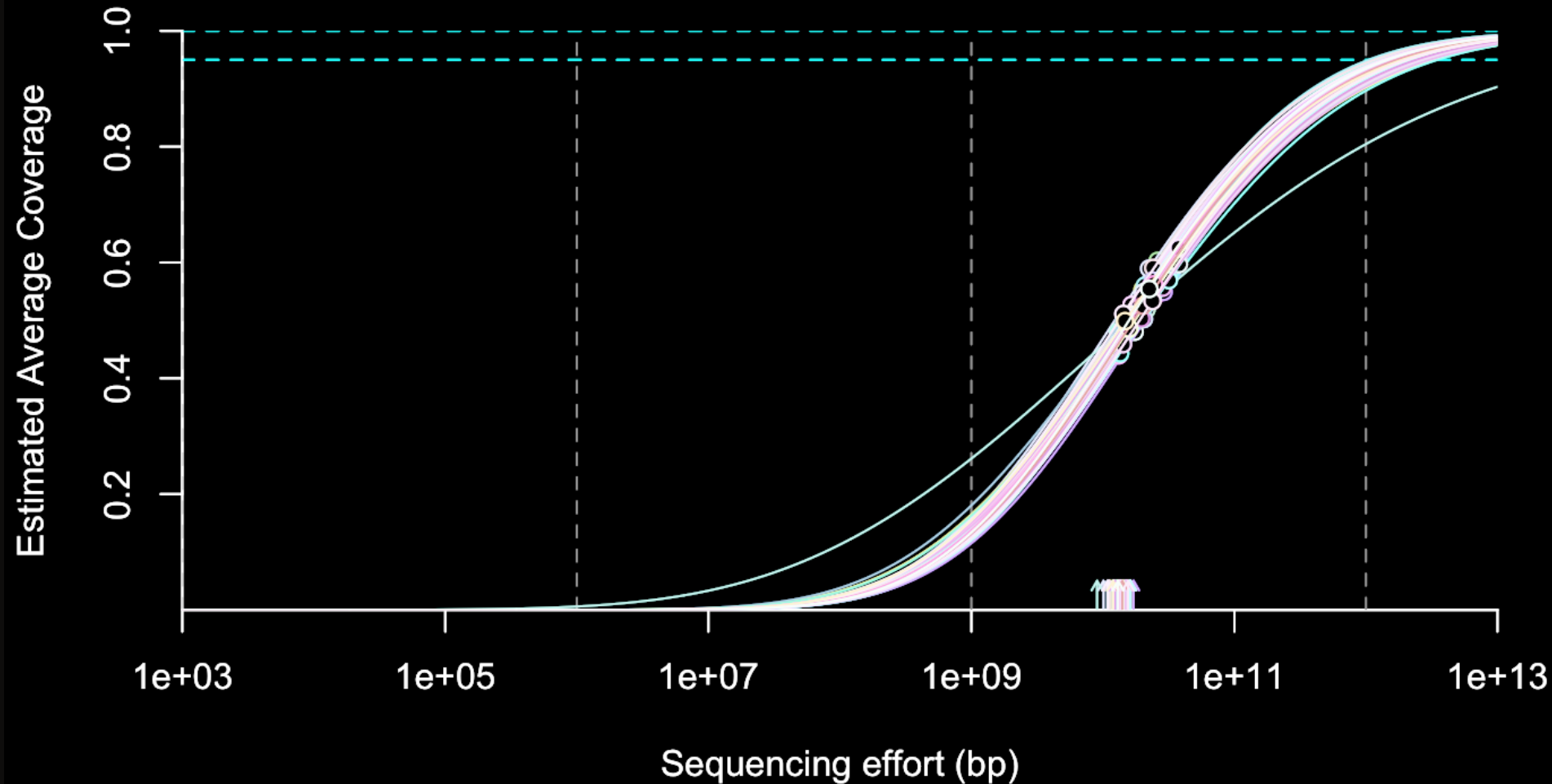
Andr n and colleagues' paper reminding us of the importance of replication and learnt during our undergraduate years is a frequently forgotten. This



What are determinants of the complexity of a metagenome? Textbook definition: Number of species.



Nonpareil Curves



Currently, we need 2-3 TB of RAM to assemble 100 Gbps of data according to cutting edge standards.

If we need 1 Tbps of sequencing data to cover all microbial information in a sample, we hit the limit of current supercomputing.

Projects in the department





Important ecosystem services of rivers associated with One Health:

- Carbon turnover (leaf litter degradation, carbon fixation)
- Nitrogen cycling and particularly denitrification
- Antibiotics and antimicrobial resistance degradation
- Pathogen and pollutant attenuation
- Drinking water provision
- ...

Open questions:

- Are these services maintained during anthropogenic stress exposure of rivers?
- How is the microbial biodiversity and its functions structured in rivers?
- How do anthropogenic stressors affect these microbial communities at the cellular level?

Catchments as model ecosystems: Emscher and Boye



The Emscher - a river in the Ruhr area
→ from floodplain idyll to epidemic hotspot - and back again



© EGLV

© EGLV

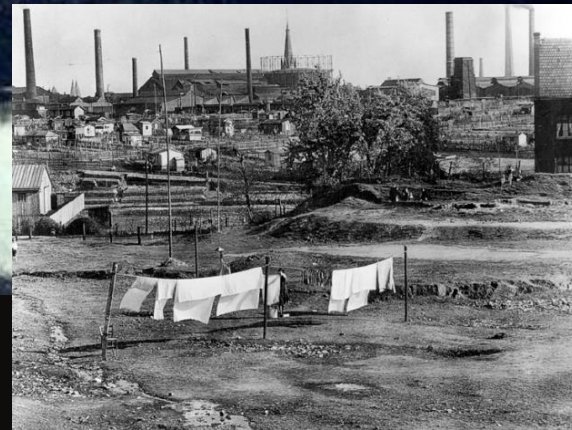


© Daniel Grabner

© Daniel Grabner

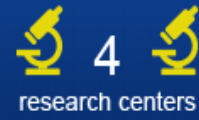


Herne um 1900.

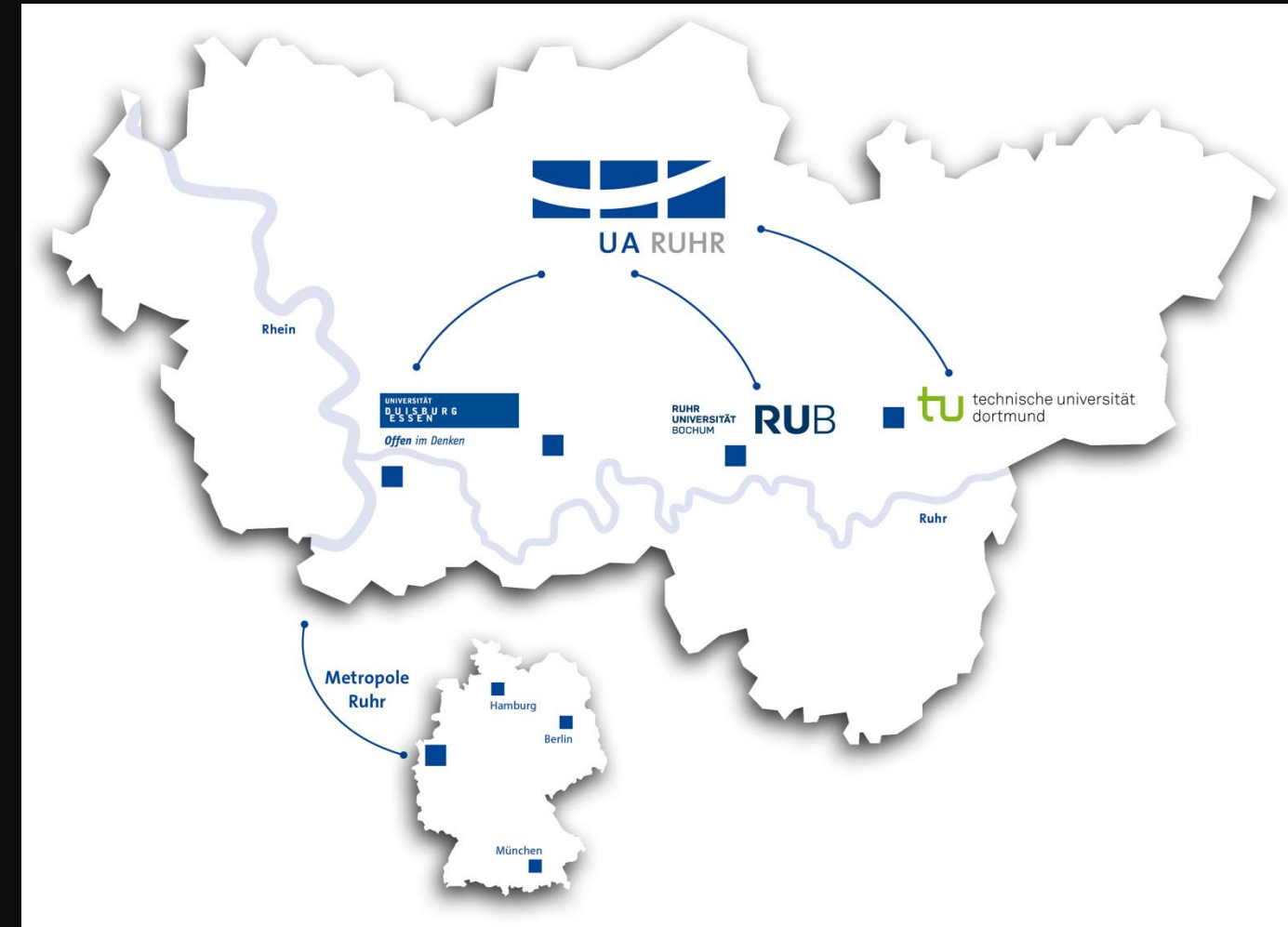


Around 1900

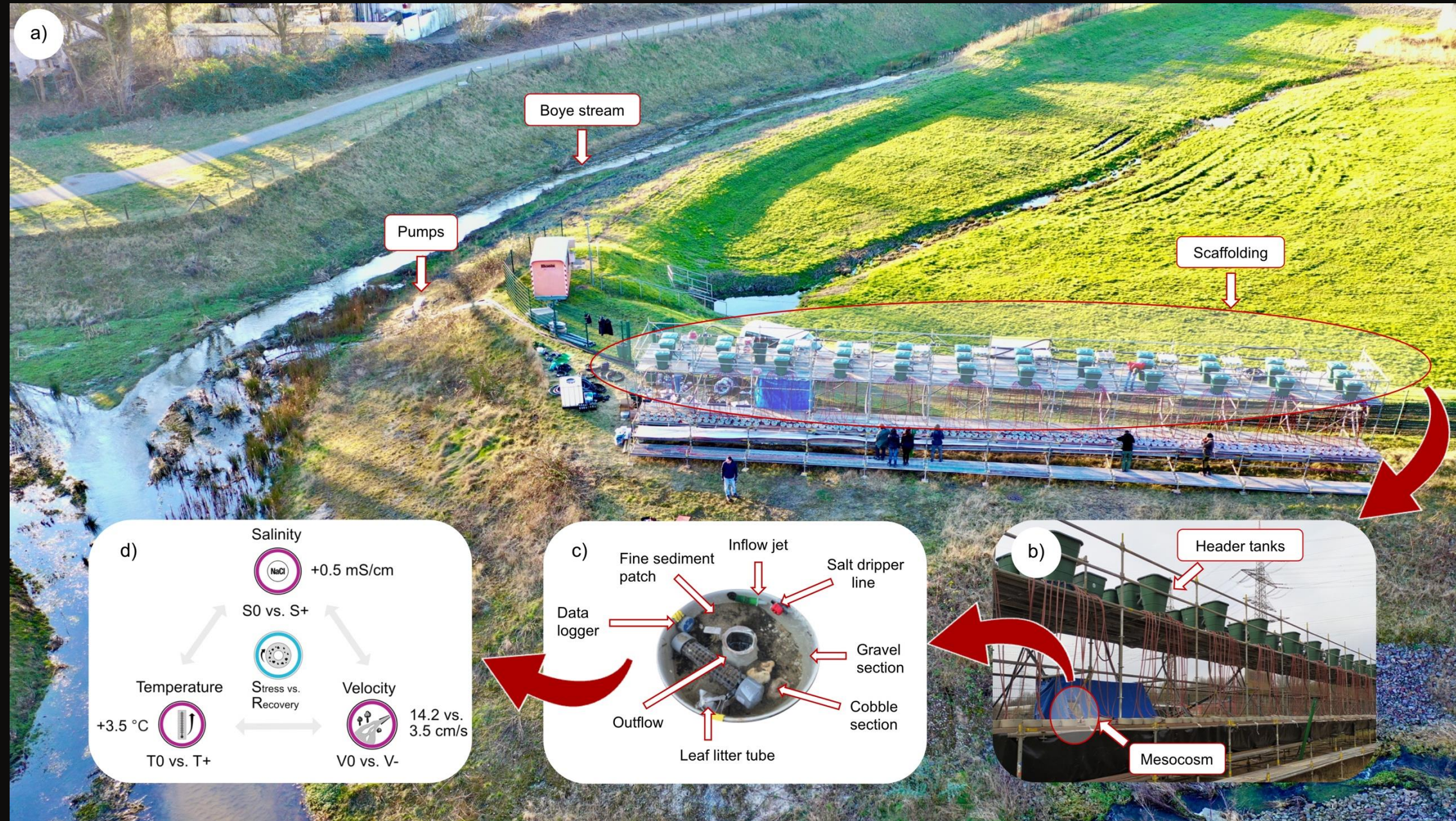
Around 2000



- **One Health Ruhr – from Molecules to Systems**
- Chemical Sciences and Sustainability
- Trustworthy Data Science and Security
- Future Energy Materials and Systems
- College for Social Sciences and Humanities



Mesocosm approach: ExStream

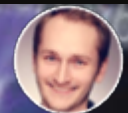




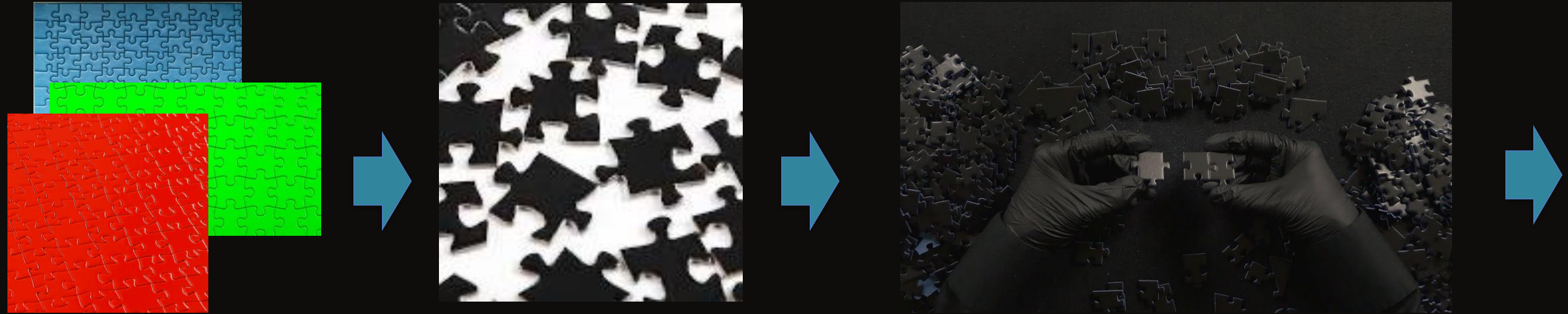
ExStream details



One of 64 mesocosm setups



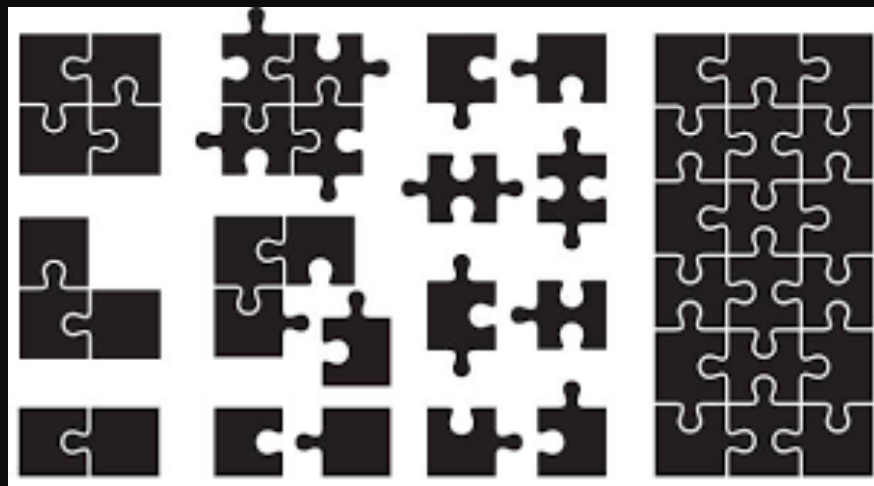
Metagenomes – puzzling information or a puzzle of information?



Original sample

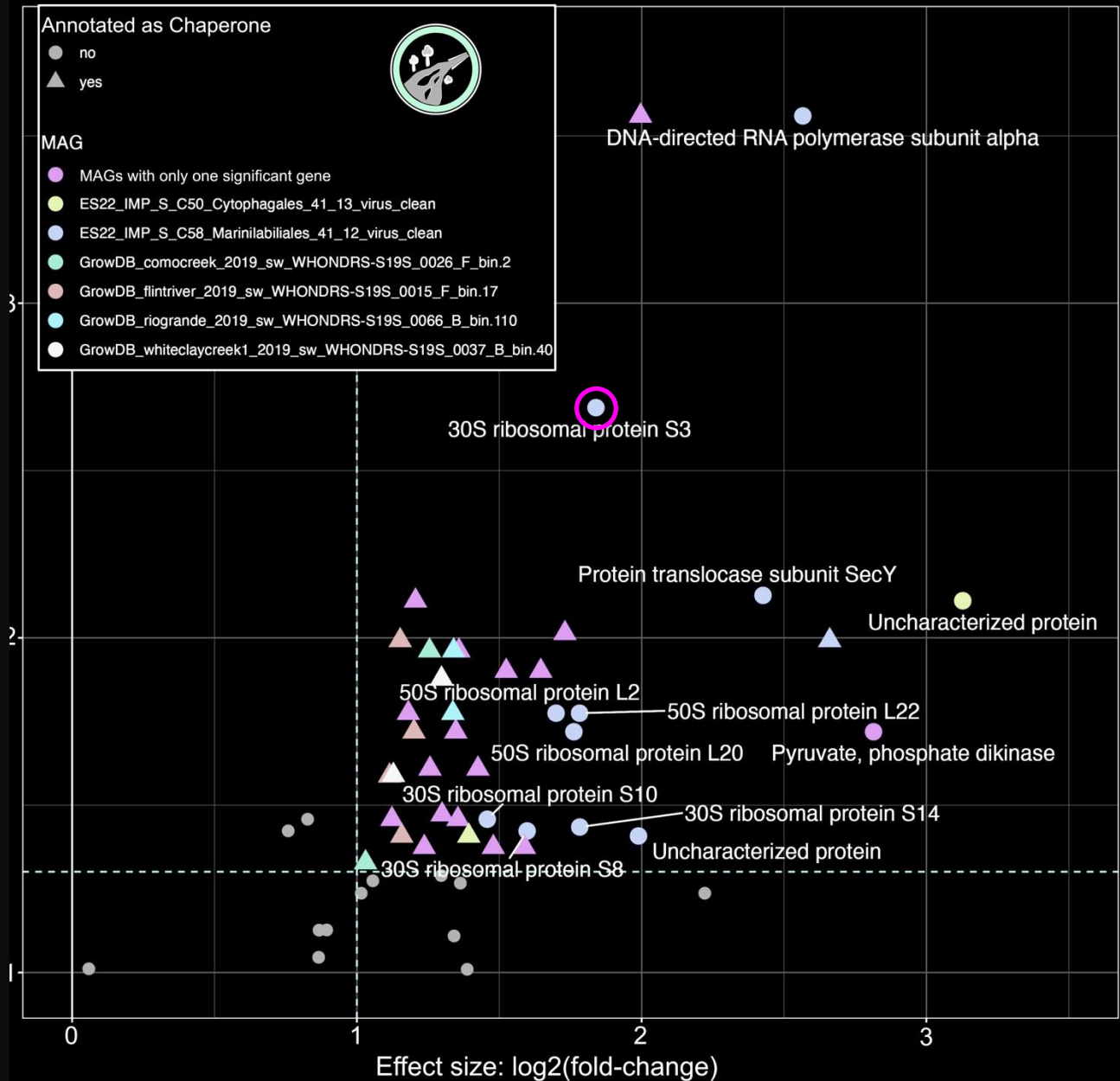
DNA extraction and sequencing

Assembly



Assembled pieces

Transcriptomic response of MAGs

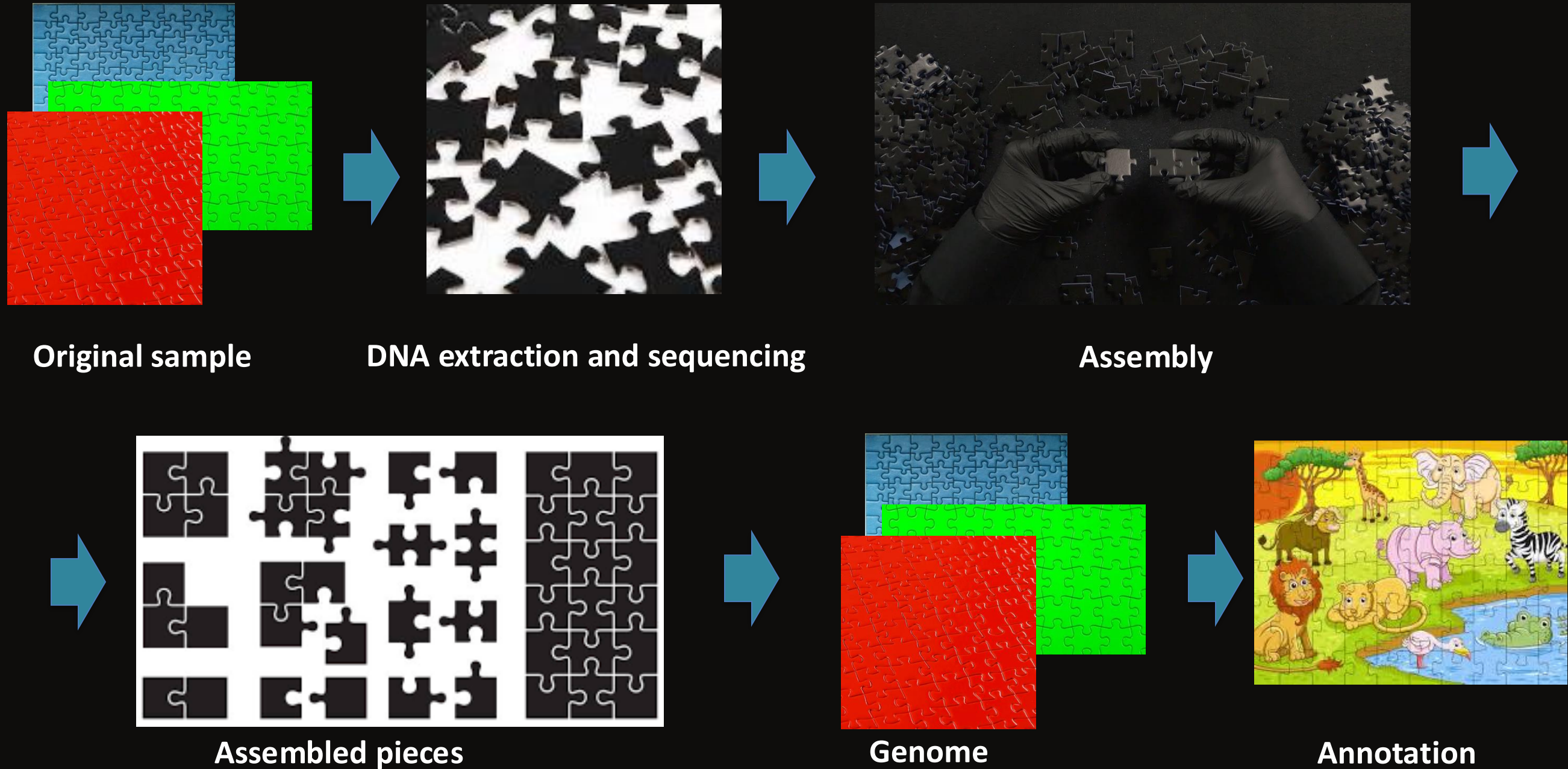


How do organisms with MAGs respond to reduced flow velocity?

Massive upregulation of ribosomal protein / gene expression.

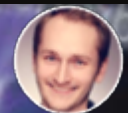
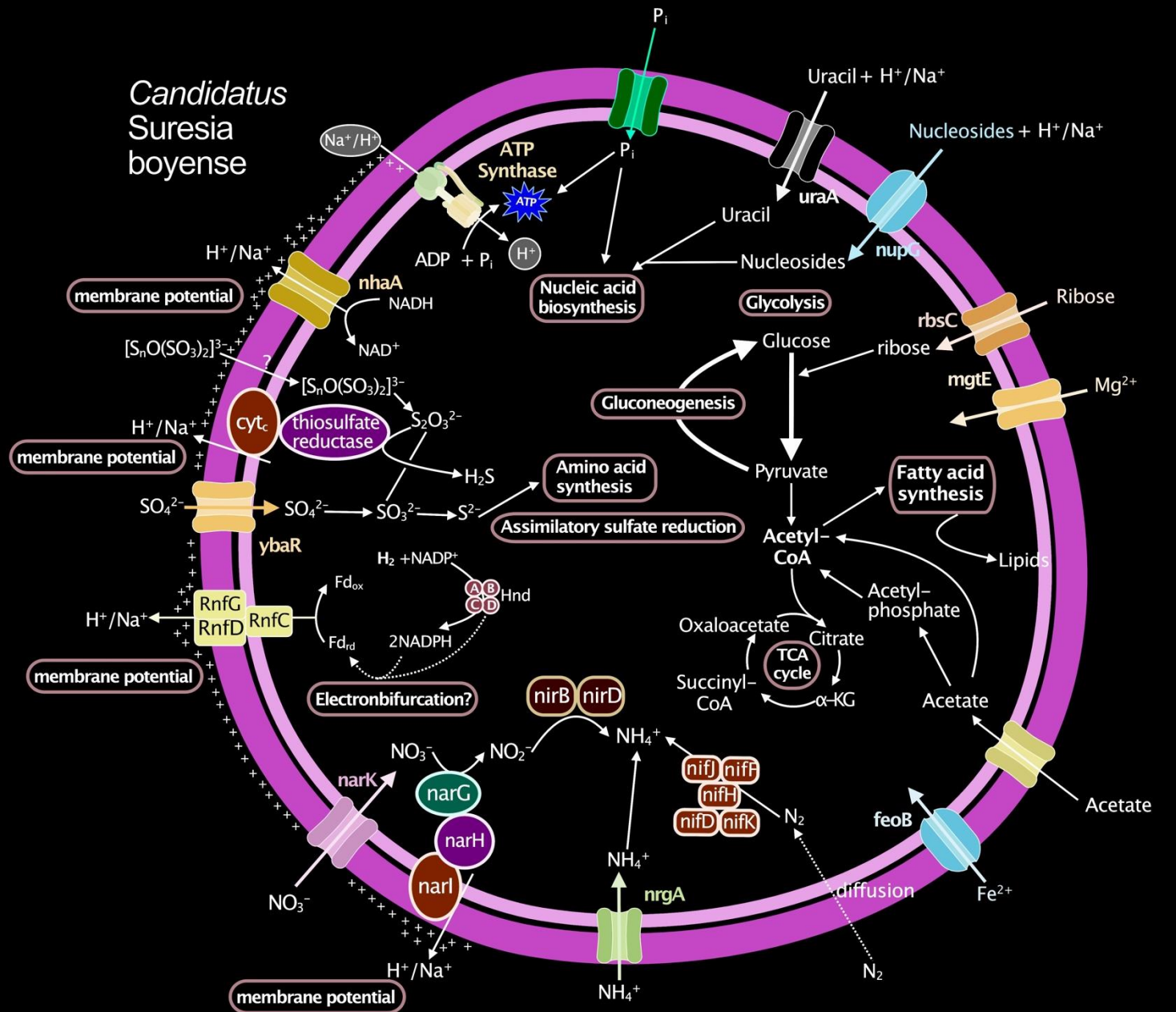


Metagenomes – puzzling information or a puzzle of information?

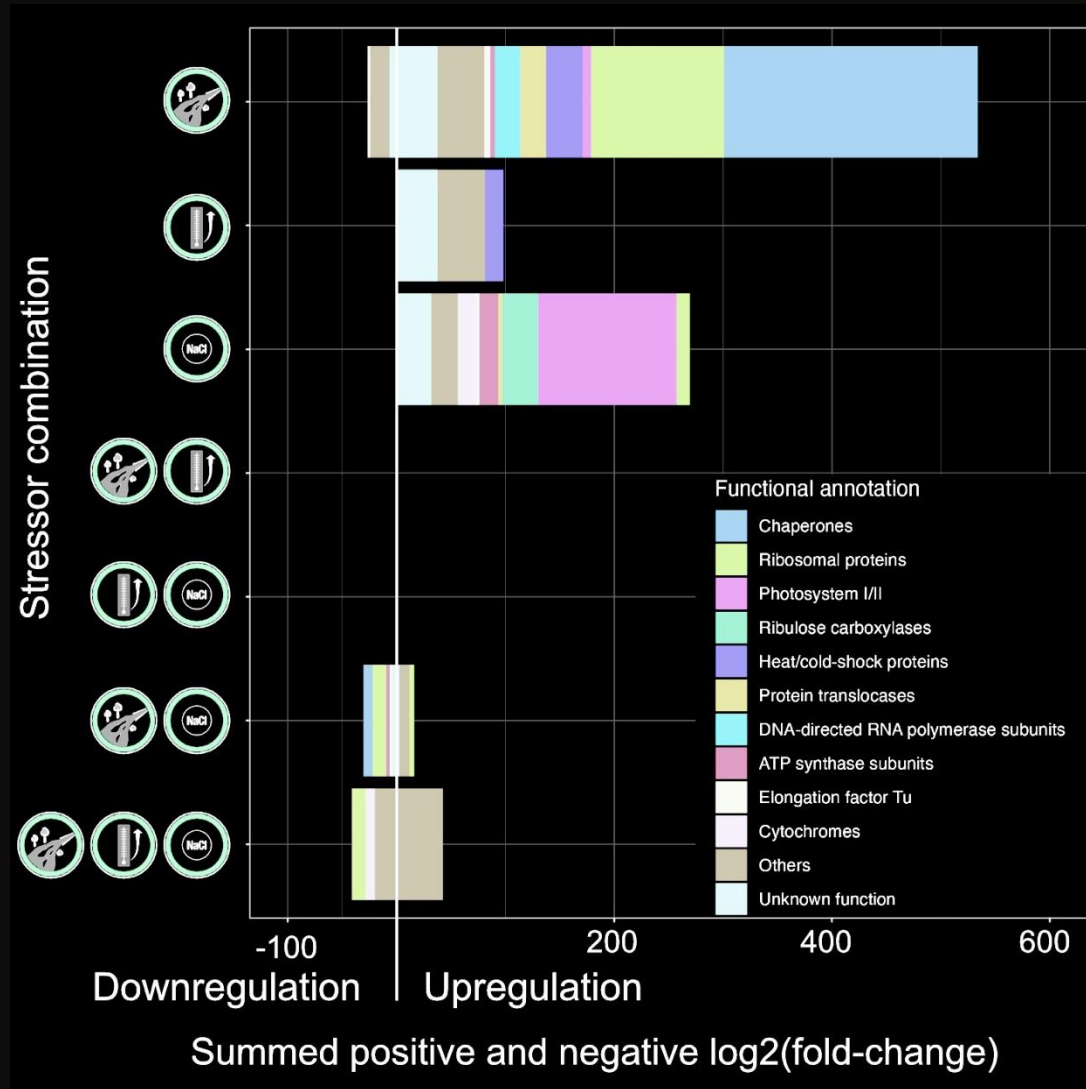


Candidatus "Suresia boyense" – a keystone species

- Reduced flow velocity revealed a new keystone species
- Low abundant but highly active new family of bacteria
- Named "Suresia boyense" after Bernd Sures and the river Boye
- Anaerobic metabolism
- Versatile electron acceptors for anaerobic respiration (nitrate reduction, thiosulfate reduction, ...)
- Nitrogen fixation



Substantial transcriptomic response by reduced flow velocity



Thank you for listening



The Probst Lab

Dr. Andre Soares
Dr. Cristina Moraru
Dr. Till Bornemann
Dr. Indra Banas
Dr. Julia Nuy
Dr. Sarah Esser
Dr. Ferial Bouderk
Julia Plewka
Sophie Anja Simon
Tom Lennard Stach
Jörn Starke
Julian Künkel
Ismael Hayani Aji
Pengyao Zhang
Ken L. Dreger
Manan Shah
Sabrina Eisfeld
Ines Pothmann
Michelle Moll
Maximiliane Ackers
Lea Griesdorn
Noel Kleber
Katharina Sures
Marie Dieckmann
Kai Rogmann
Jannis Becker
Alina Helbrecht
Lars Hermann
Flora Mai
Komal Preet Kaur