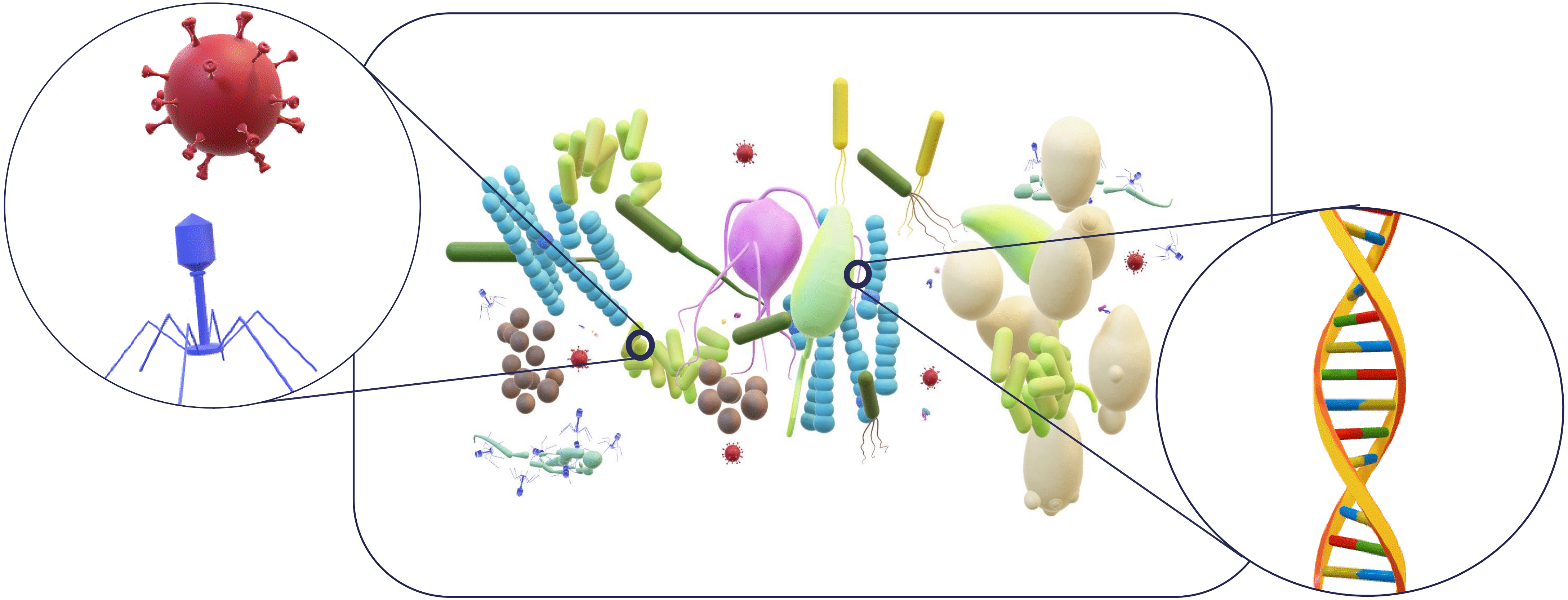
Microbiome research and the meaning of Open Science

Birgit Wassermann and Gabriele Berg
Institute of Environmental Biotechnology
TU Graz

The microbiome



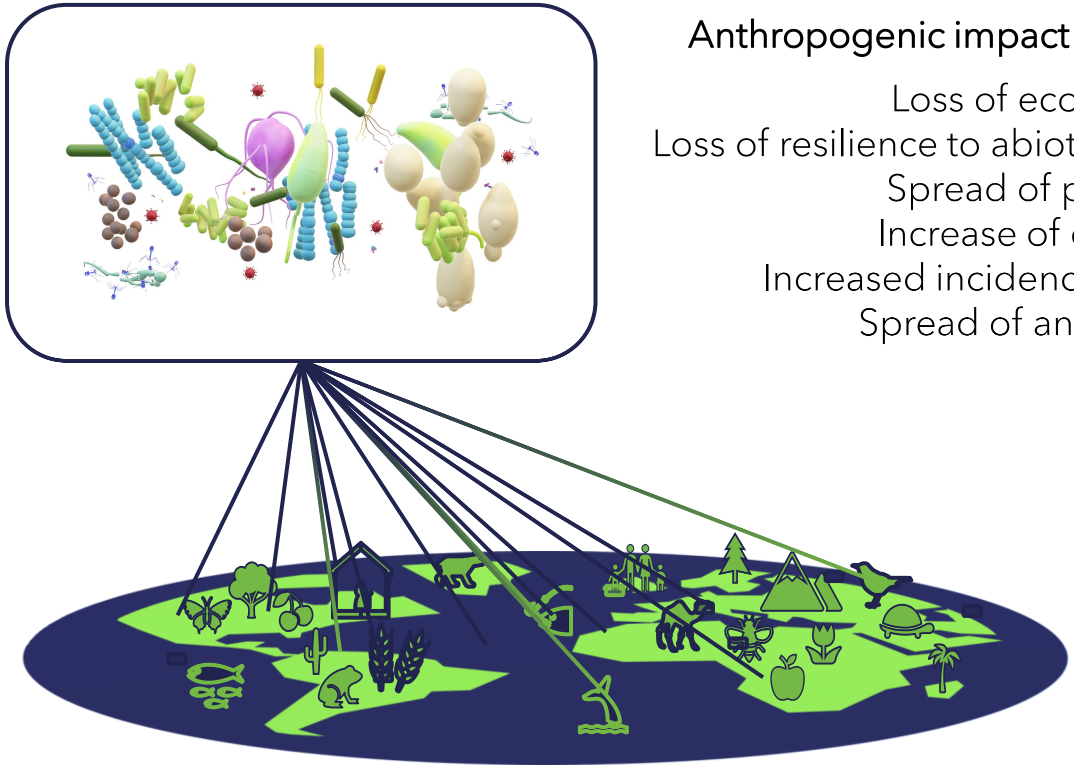
The microbiome encompasses all **microbial organisms** (Bacteria, Archaea, Fungi, Algae, Protists), relic DNA, **Phages, Plasmids, Viruses**, and the **entirety of their functions**.

MOOC Microbiome & Health, https://imoox.at/mooc/. Berg et al. 2020, Microbiome

The microbiome

and the necessity to research

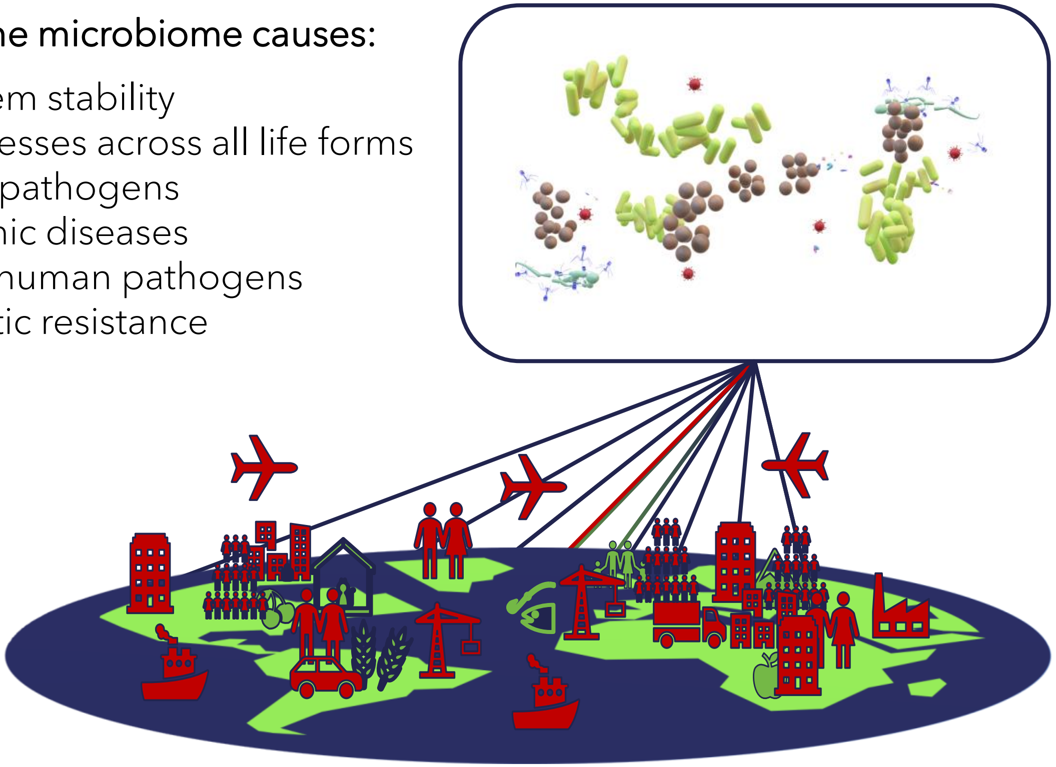
The microbiome is in control of all biogeochemical cycles and interconnects all life on earth.



Anthropogenic impact on the microbiome causes:

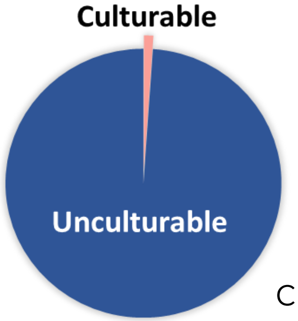
Loss of ecosystem stability
Loss of resilience to abiotic stresses across all life forms
Spread of plant pathogens
Increase of chronic diseases
Increased incidence of human pathogens
Spread of antibiotic resistance

Planetary health problems result from a massive disconnection between us and nature.

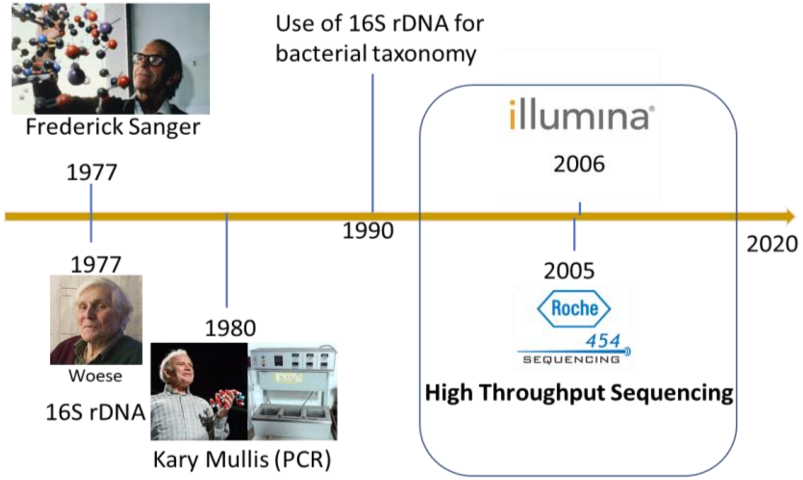


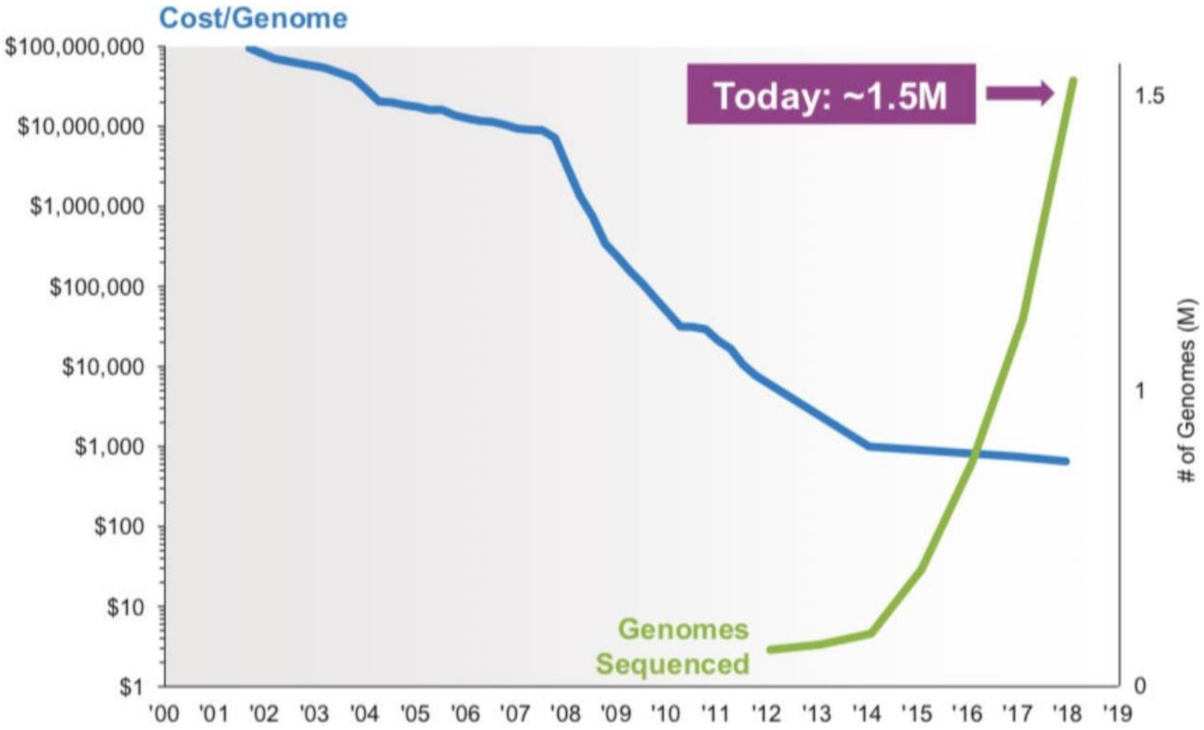
MOOC Microbiome & Health, https://imoox.at/mooc/. Blaser MJ, 2017, Nat Rev Immunol. Berg and Cernava, 2022, Microbiome

Data explosion



15 years since we can capture the *Unculturable*





https://www.forbes.com

The omics technologies:

Metagenomics: Who is there and what can they do? → DNA

Metatranscriptomics: What are they doing at the moment? → RNA

Metaproteomics: What is their metabolic potential? → Proteins

Metabolomics: What are they producing? → Metabolites

Microbiome research generates a tremendous amount of data.

All data are publicly available.

Microbiome research

https://mgm.duke.edu/centers-and-cores/microbiome-core-facility

Google Scholar
1,110,000 articles

THE LANCET
173 articles

Science
1,472 articles

Cell
5,393 articles

nature
12,556 articles

Publication of microbiome research requires

- ✓ Deposition of the whole dataset in a public repository that **can be accessed by everyone world-wide**
- ✓ **Provision of metadata**
- ✓ **The FAIR concept** (Findable, Accessible, Interoperable, and Reusable) is implemented
- ✓ **Upload and download is free of charge**
- ✓ Researchers can choose between **three repositories that are interlinked**

ENA
European Nucleotide Archive

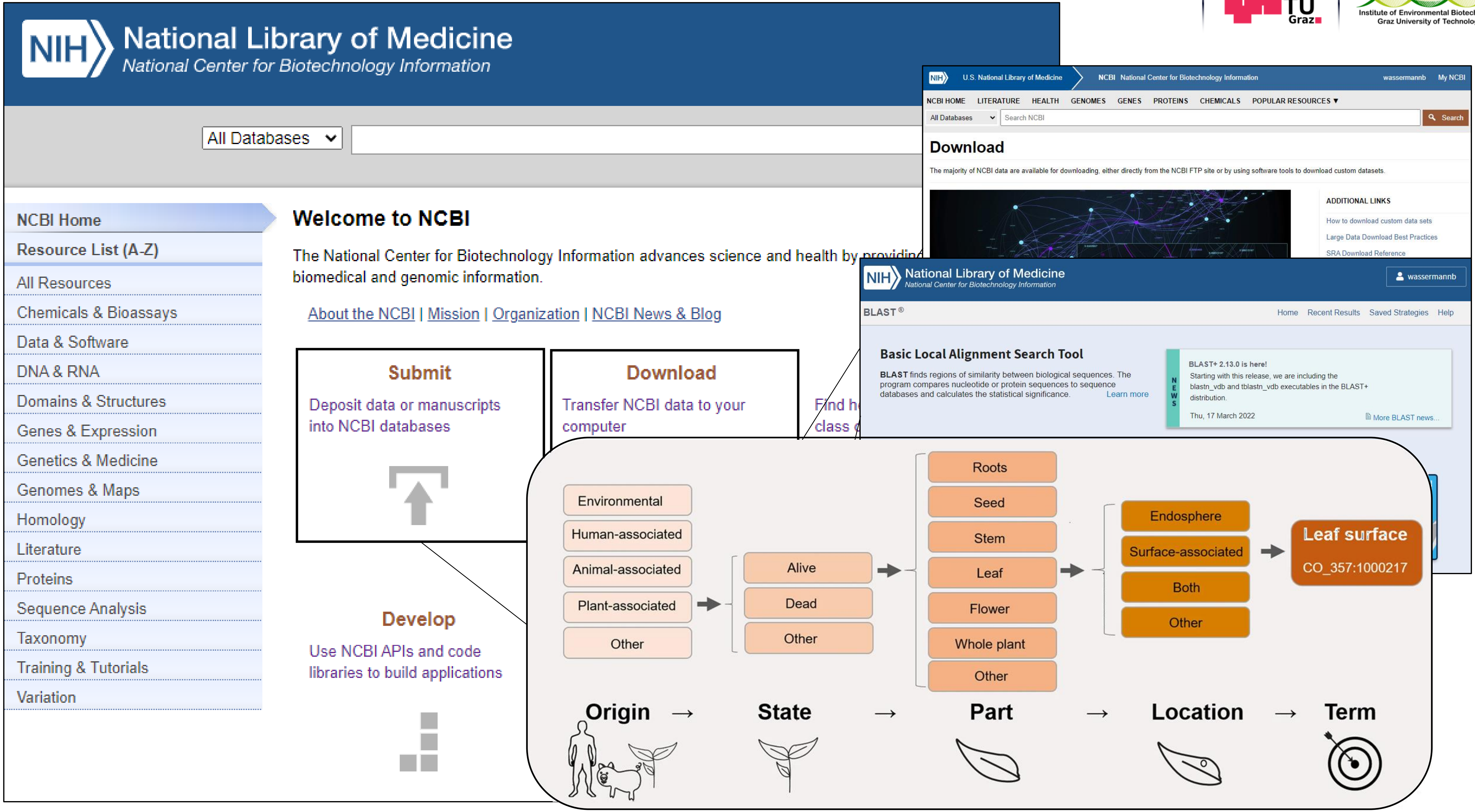
Operated by the European Molecular Biology Laboratory (EMBL) and supported by European Commission, the British Biotechnology and Biological Sciences Research Council (BBSRC) and the Wellcome Trust (WT)

NCBI

Operated by the United States National Library of Medicine (NLM), a branch of the United States National Institutes of Health (NIH) and funded by the government of the United States.

DDBJ
DNA Data Bank of Japan

Funded by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT).



National Library of Medicine
National Center for Biotechnology Information

All Databases

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providin
biomedical and genomic information.

About the NCBI | Mission | Organization | NCBI News & Blog

Submit
Deposit data or manuscripts into NCBI databases

Download
Transfer NCBI data to your computer

Find h
class c

Develop
Use NCBI APIs and code libraries to build applications

Download

The majority of NCBI data are available for downloading, either directly from the NCBI FTP site or by using software tools to download custom datasets.

ADDITIONAL LINKS
How to download custom data sets
Large Data Download Best Practices
SRA Download Reference

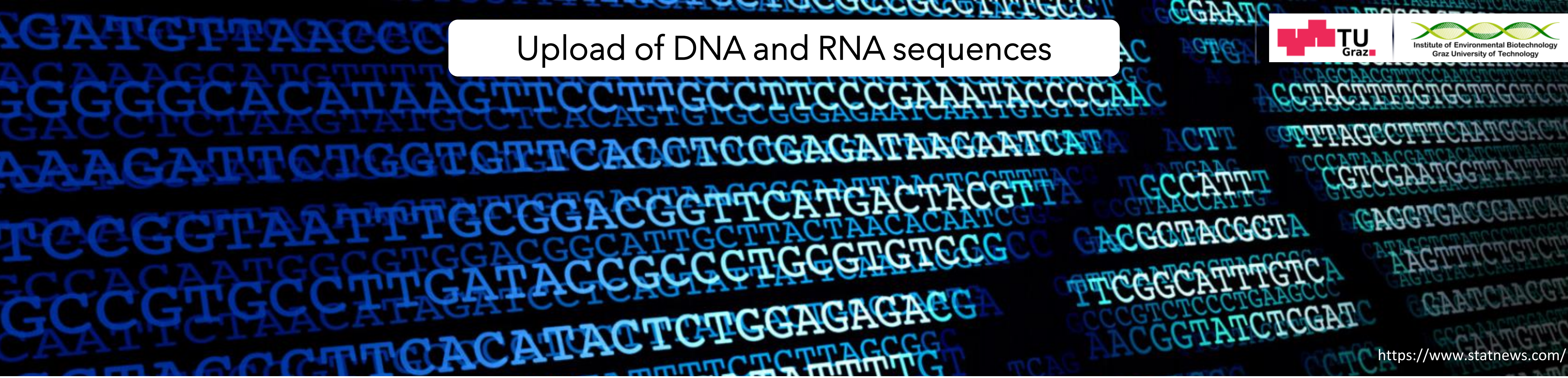
Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. Learn more

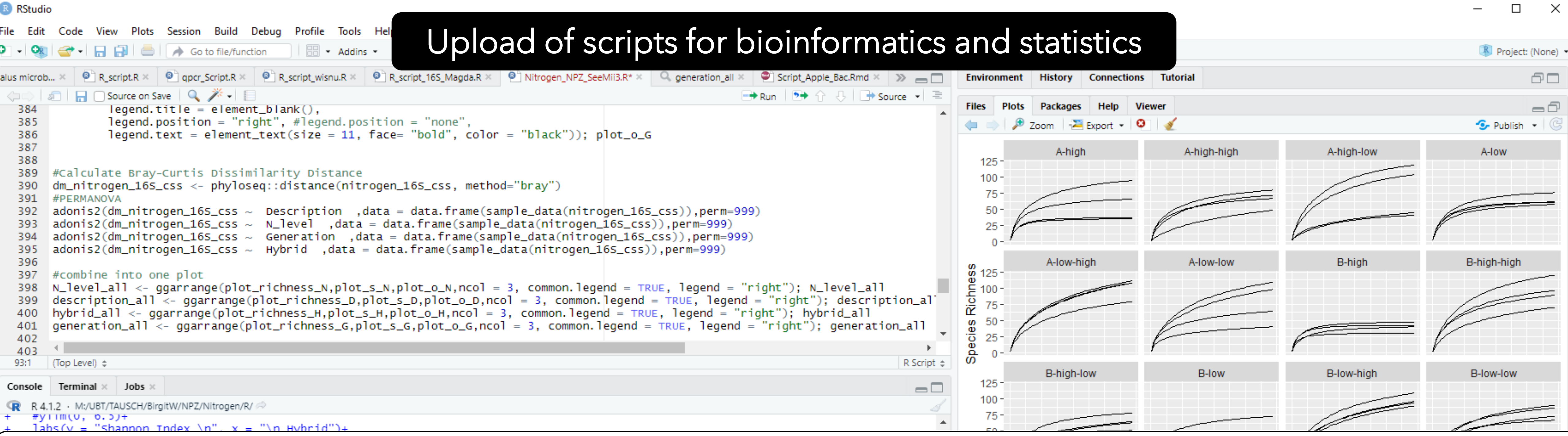
NEWS
BLAST+ 2.13.0 is here!
Starting with this release, we are including the blastn_vdb and tblastn_vdb executables in the BLAST+ distribution.
Thu, 17 March 2022
More BLAST news...

Cernava et al., 2022, Environmental Microbiome

Upload of DNA and RNA sequences


https://www.statnews.com/

Upload of scripts for bioinformatics and statistics



```r
        legend.title = element_blank(),
        legend.position = "right", #legend.position = "none",
        legend.text = element_text(size = 11, face= "bold", color = "black")); plot_o_G


#Calculate Bray-Curtis Dissimilarity Distance
dm_nitrogen_16S_css <- phyloseq::distance(nitrogen_16S_css, method="bray")
#PERMANOVA
adonis2(dm_nitrogen_16S_css ~  Description  ,data = data.frame(sample_data(nitrogen_16S_css)),perm=999)
adonis2(dm_nitrogen_16S_css ~  N_level  ,data = data.frame(sample_data(nitrogen_16S_css)),perm=999)
adonis2(dm_nitrogen_16S_css ~  Generation  ,data = data.frame(sample_data(nitrogen_16S_css)),perm=999)
adonis2(dm_nitrogen_16S_css ~  Hybrid  ,data = data.frame(sample_data(nitrogen_16S_css)),perm=999)

#combine into one plot
N_level_all <- ggarrange(plot_richness_N,plot_s_N,plot_o_N,ncol = 3, common.legend = TRUE, legend = "right"); N_level_all
description_all <- ggarrange(plot_richness_D,plot_s_D,plot_o_D,ncol = 3, common.legend = TRUE, legend = "right"); description_al
hybrid_all <- ggarrange(plot_richness_H,plot_s_H,plot_o_H,ncol = 3, common.legend = TRUE, legend = "right"); hybrid_all
generation_all <- ggarrange(plot_richness_G,plot_s_G,plot_o_G,ncol = 3, common.legend = TRUE, legend = "right"); generation_all
```

```
+    #ylim(0, 6.5)+
+    labs(y = "Shannon Index \n", x = "\n Hybrid")+
```

Re-use of mutiple datasets increases the value of the study

The added value of sharing data

Loss of gut microbial diversity is significantly associated with chronic diseases in industrial populations.

498 genomes from human paleofaeces compared to **789 publicly available** present-day gut samples

Ancient human stool samples (1000–2000 years old)

✓ are more similar to present-day non-industrialized human gut samples and distinct from industrialized samples.

✓ contain important health-beneficial microorganisms

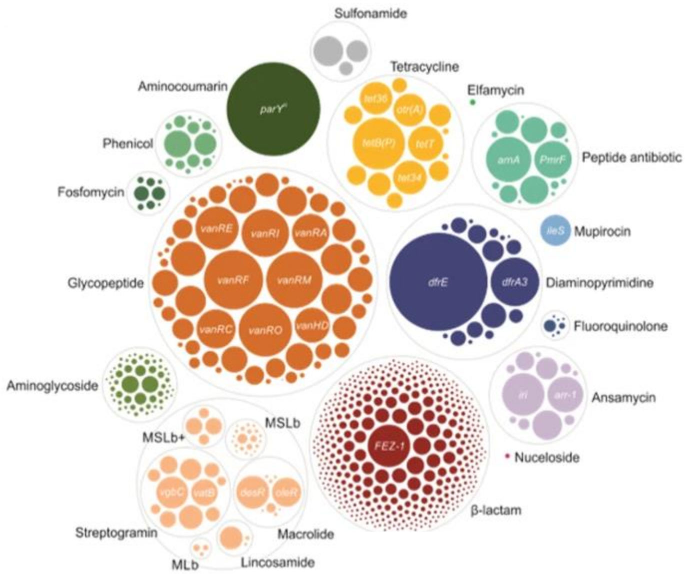
✓ contain less antibiotic-resistance genes

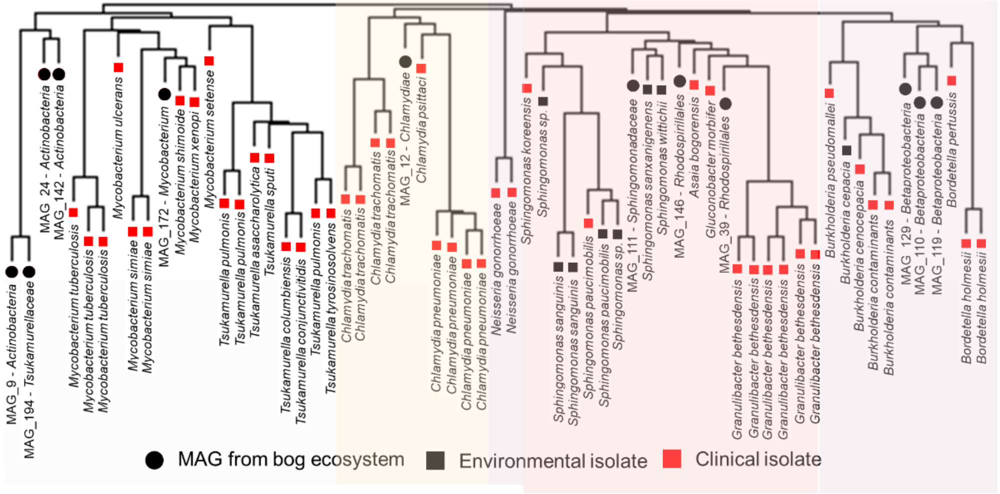
Wibowo et al. 2021. Nature

The added value of sharing data

Bog ecosystems store 30% of global carbon. Ongoing global warming will make them one of the largest CO_2 sources.

Austrian bog microbiomes and **publicly available datasets from Norway, Germany, and Sweden**





Evolutionary old bog ecosystems harbor

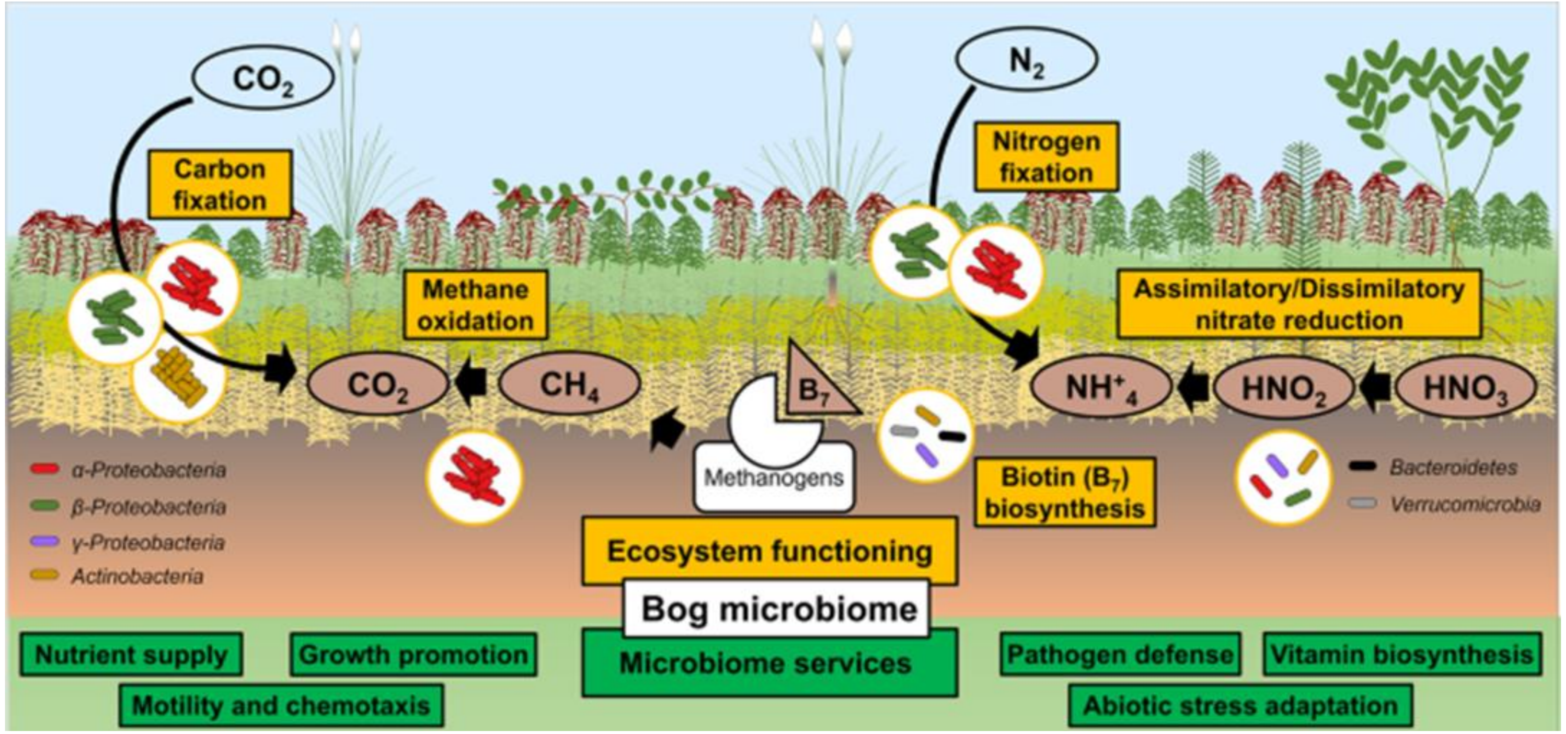
✓ a high diversity of antibiotic resistance genes

✓ microorganisms with sequence similarity to human pathogens.

Obermeier et al. 2020, ISME J., Wicaksono et al. 2021, Microbiome

The added value of sharing data

Bog ecosystems store 30% of global carbon. Ongoing global warming will make them one of the largest CO_2 sources.



Evolutionary old bog ecosystems harbor

✓ microbiomes that are essential for carbon capture

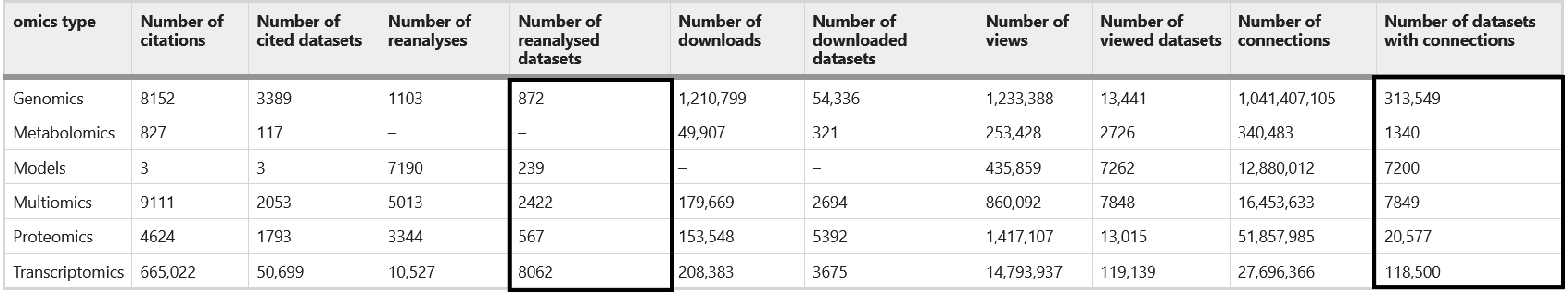
✓ plant-beneficial microorganisms that can be used for sustainable agriculture.

Wicaksono et al. 2021, Microbiome

The issue with ‚data about data'

- ✓ Compared to the massive amount of data, only few products from microbiome research were patented (8,000 patents worldwide).
- ✓ Generalization of findings by performing integrative analyses is difficult.
- ✓ Metadata should provide all information to repeat a study and to reuse data in a broader context.
- ✓ A common solution is necessary for this fast-developing research field.

Number of citations, reanalyses, downloads, views, and connections (April 2019)



omics type	Number of citations	Number of cited datasets	Number of reanalyses	Number of reanalysed datasets	Number of downloads	Number of downloaded datasets	Number of views	Number of viewed datasets	Number of connections	Number of datasets with connections
Genomics	8152	3389	1103	872	1,210,799	54,336	1,233,388	13,441	1,041,407,105	313,549
Metabolomics	827	117	–	–	49,907	321	253,428	2726	340,483	1340
Models	3	3	7190	239	–	–	435,859	7262	12,880,012	7200
Multiomics	9111	2053	5013	2422	179,669	2694	860,092	7848	16,453,633	7849
Proteomics	4624	1793	3344	567	153,548	5392	1,417,107	13,015	51,857,985	20,577
Transcriptomics	665,022	50,699	10,527	8062	208,383	3675	14,793,937	119,139	27,696,366	118,500

Perez-Riverol et al. 2019, Nature Communications. Cernava et al. 2022, Environ Microbiome

The availability of datasets led to the most important discoveries in microbiome research for medicine, society and planetary health.

"One of the biggest contributions scientists leave is the potential of their research being reused in the future!"

Tweet by Ulisses Nunes da Rocha (UFZ, Leipzig)

Thank you for your attention!