The GlobalFungi database: the FAIR global atlas of fungi

Petr Baldrian

Institute of Microbiology, Czech Academy of Sciences

FAIR data principles & open data
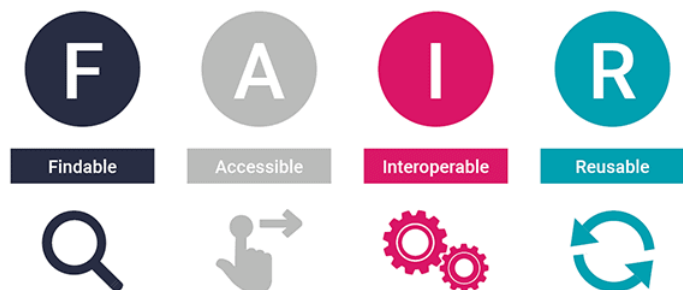
>IEH00WJ01BK2Y9 xy=533_195
ATTAGCAGTAAGATCCGTAGGTGAACCTGCGGAAGG
ATCATTATCGAAACGAATAGGAATGGGGGAGCAAGAC
GGCAAAGCAAAAGACTGTCGCTGGCTAGATTCATTTC
TGGCATGTGCACGTCCTTGCTTTTTTCGTCGACCTTTC
TCTTTCTTTCTTTCACACCTGTGCACCCGTTGTAGGTC
CTCGAAAGAGGATC
>IEH00WJ01DVNL6 xy=1473_2492
TAGTGTAGATAGATCCGTAGGTGAACCTGCGGAAGG
ATCATTATCGAATCGTAGGCGAGGGTTGTCGCTGGCC
TCTCGGGGCATGTGCACGCCCGAGCCCTTGAATCCC
AAACACCACATGTGAACCCACCGTAGGCCTTCGGGC
CTATGTCTTATCATATAATCTGAATGTCTAATAGAATG
TAAACCCATTTCGTTGC
>IEH00WJ01CDN03 xy=858_2645
CGTATGGACAGATCCGTAGGTGAACCTGCGGAGGAG
CATATCAGTAAGCGGAGGAGCATATCAATAAGCGGA
GGAGCATATCAATAAGCGGAGGAATCCGTAGTGAAC
CTGCGGAGGAGCATATCAATAAGCGGAGGAGCATAT
CAATAAGCGGAGGAGCATATCAATAAGCGGAGGAGC
ATATCAATAAGCGGAGGAGT

BioSample		BioSample ⌄	
		Advanced	

Full ⌄

Schiermonnikoog Fungal ITS

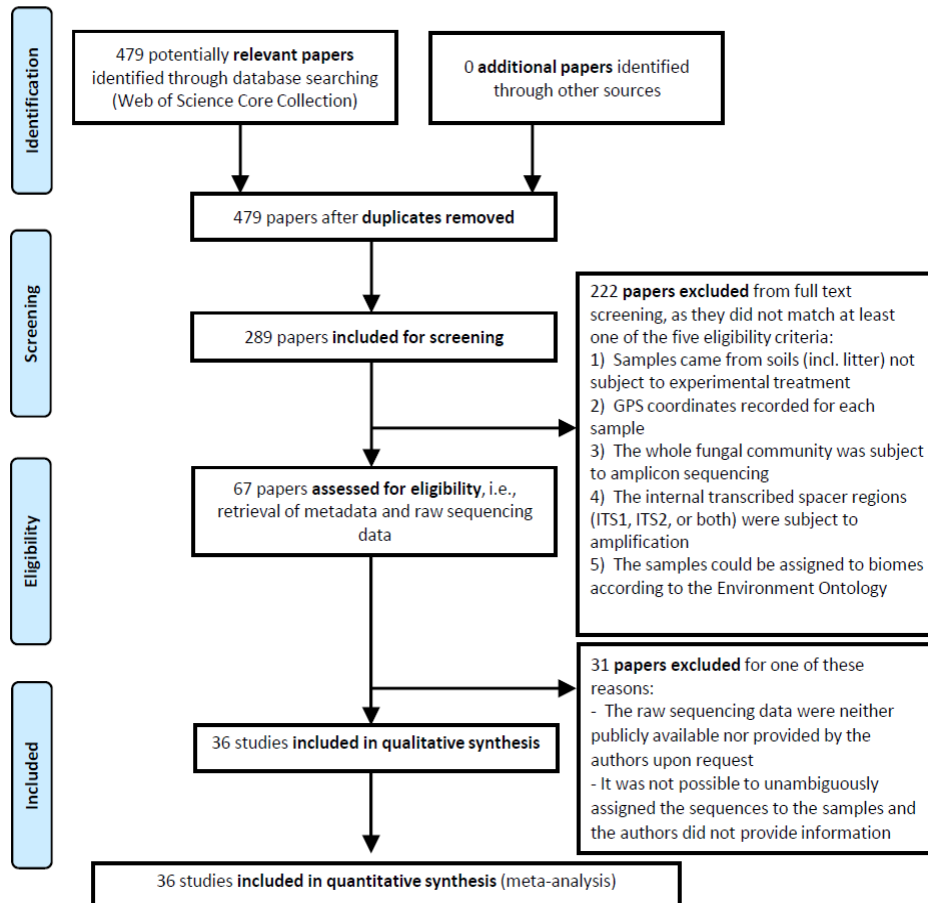| Identifiers	BioSample: SAMEA104212669; SRA: ERS1871687	
Organism	soil metagenome	
	unclassified entries; unclassified sequences; metagenomes; ecological metagenomes	
Attributes	collection date	2012
	broad-scale environmental context	Salt marsh
	local-scale environmental context	Soil chronosequence
	environmental medium	Soil
	geographic location	Netherlands
	investigation type	mimarks-survey
	project name	Schiermonnikoog Fungal ITS
	sample name	SchierFungal_ITS_59
	ENA first public	2017-08-17
	ENA last update	2017-08-14
	ENA-CHECKLIST	ERC000022
	External Id	SAMEA104212669
	INSDC center alias	UNIVERSITY OF GRONINGEN
	INSDC center name	UNIVERSITY OF GRONINGEN
	INSDC first public	2017-08-17T17:03:06Z
	INSDC last update	2017-08-14T17:15:05Z
	INSDC status	public
	Submitter Id	SchierFungal_ITS_59
	geographic location (depth)	0.1
	geographic location (elevation)	0
	geographic location (latitude)	53.4785962
	geographic location (longitude)	6.1560311
	sequencing method	Illumina MiSeq
	soil environmental package	soil

ELIXIR CZECH REPUBLIC

FAIR data principles & open data

How to best study fungal biodiversity and distribution on a global scale?

- Due to ecosystem diversity and enormous diversity of fungi, coordinated sampling efforts are unfeasible to cover global distribution of fungi

- **Large data accumulated** over time in high-throughput sequencing studies

- Data of sufficient quality can be used to model fungal distribution

- Meta-analysis approaches and data re-analysis have high potential

➢ Collect available data from all published papers reporting fungal community composition

➢ Use metadata on climate, vegetation, soil properties, vegetation, … to identify drivers of fungal distribution and biodiversity

Fungal distribution through meta-analysis of published records

Větrovský et al. 2019 Nature Communications

The GlobalFungi Database

All publicly available data on fungal community composition from published scientific papers
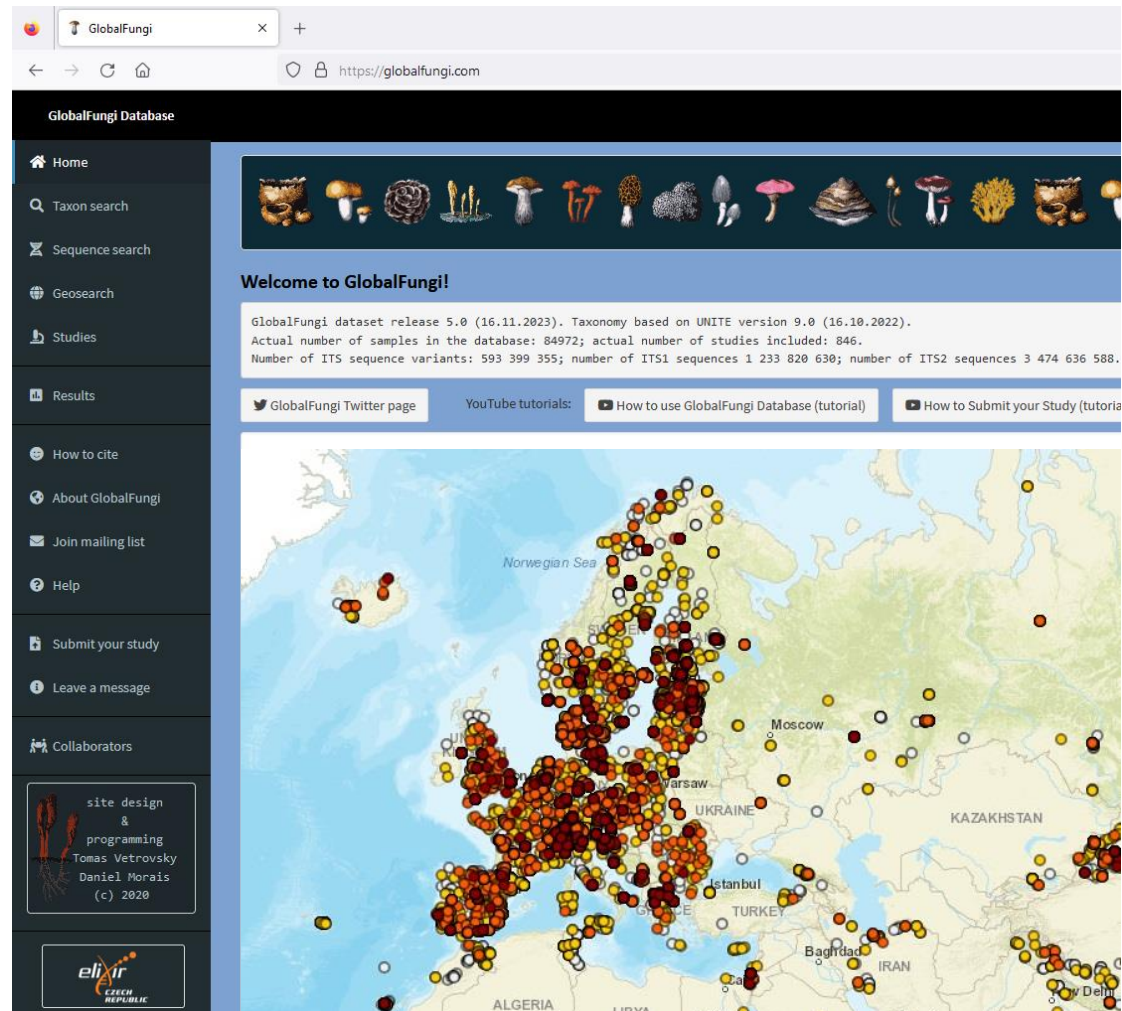
OPEN FAIR DATA

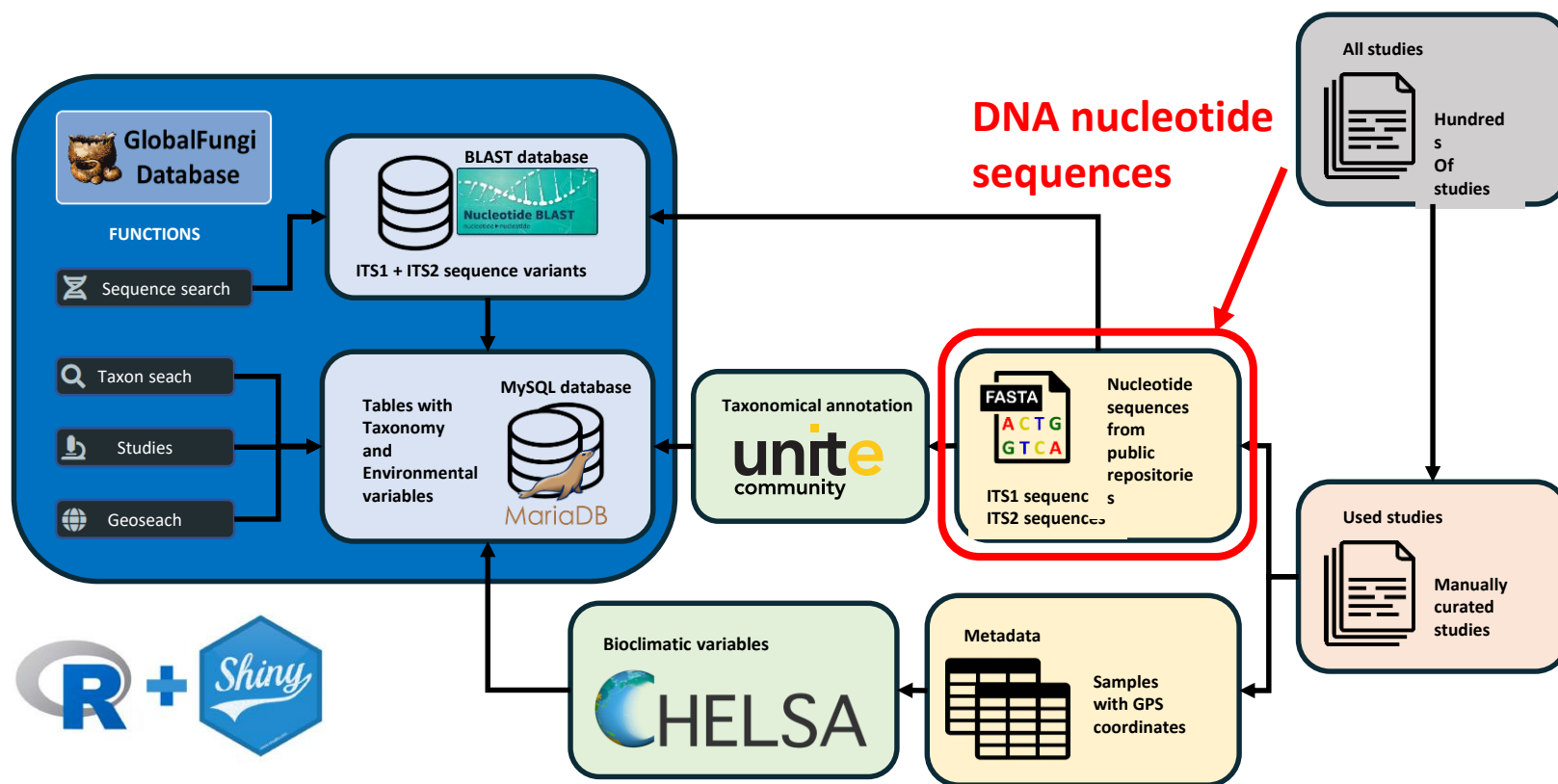**Release 5**
November 2023

84 972 samples

77 000 species of fungi

5 billion of observations of fungi globally
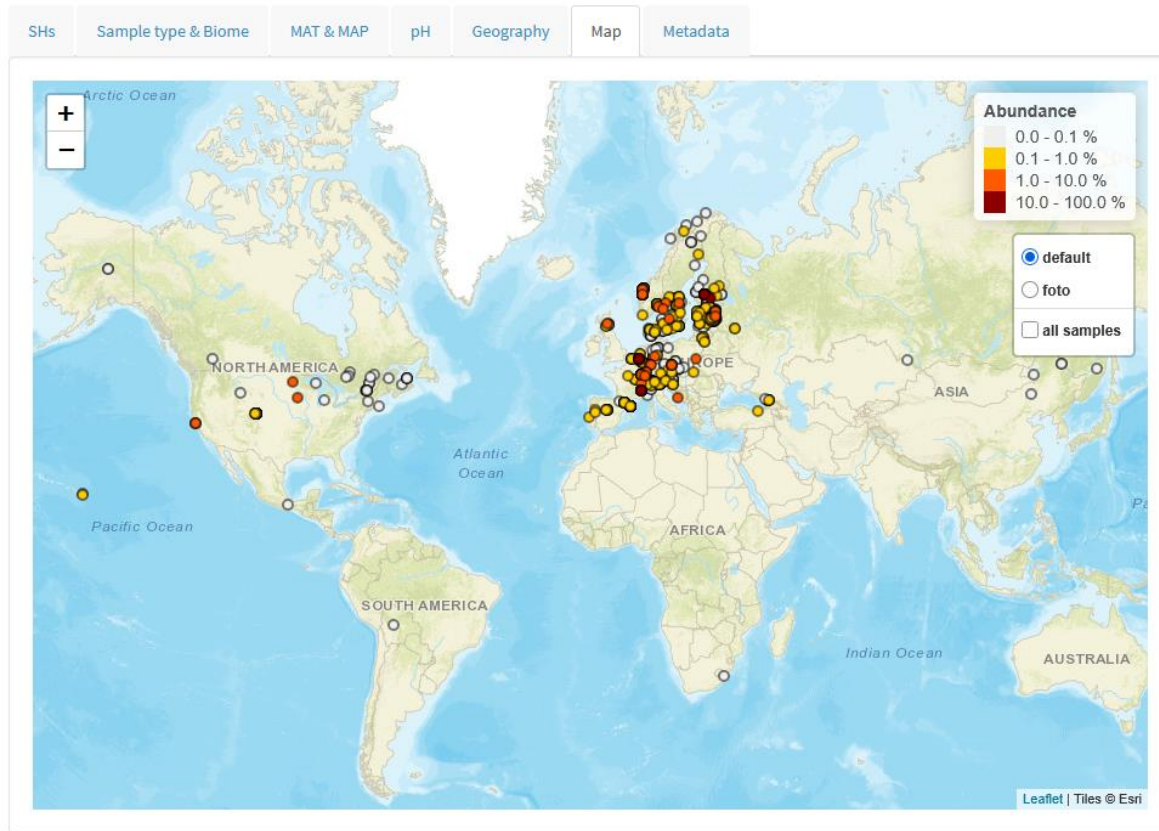**https://globalfungi.com**

Větrovský et al. 2020 Scientific Data

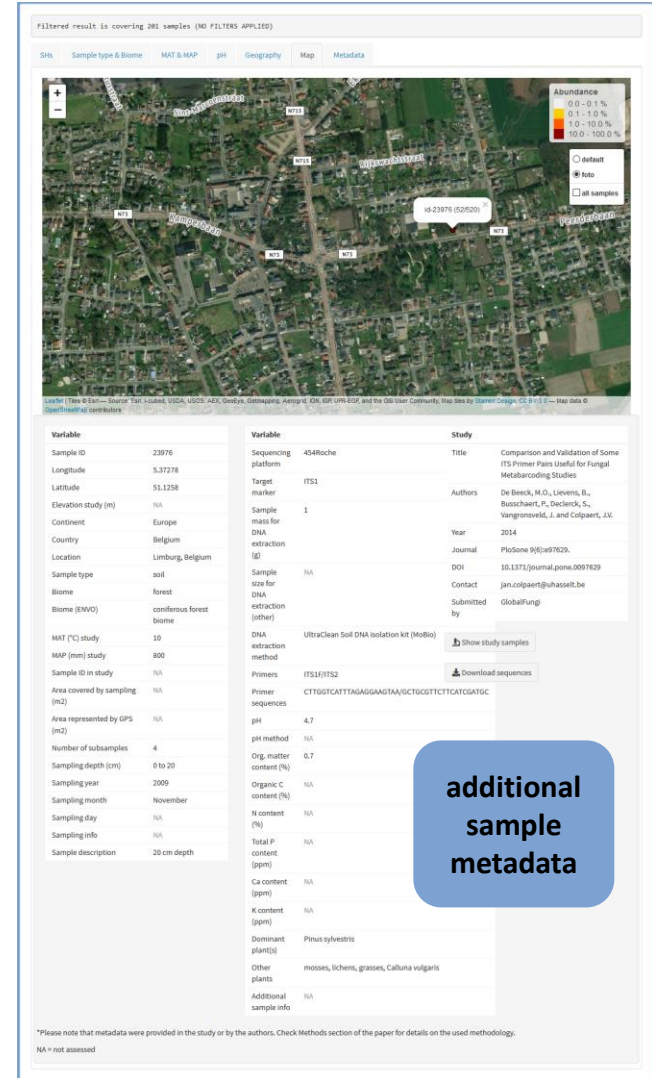# Data, metadata and annotation tools in the GlobalFungi Database

- The database is a service of ELIXIR-CZ Research Infrastructure
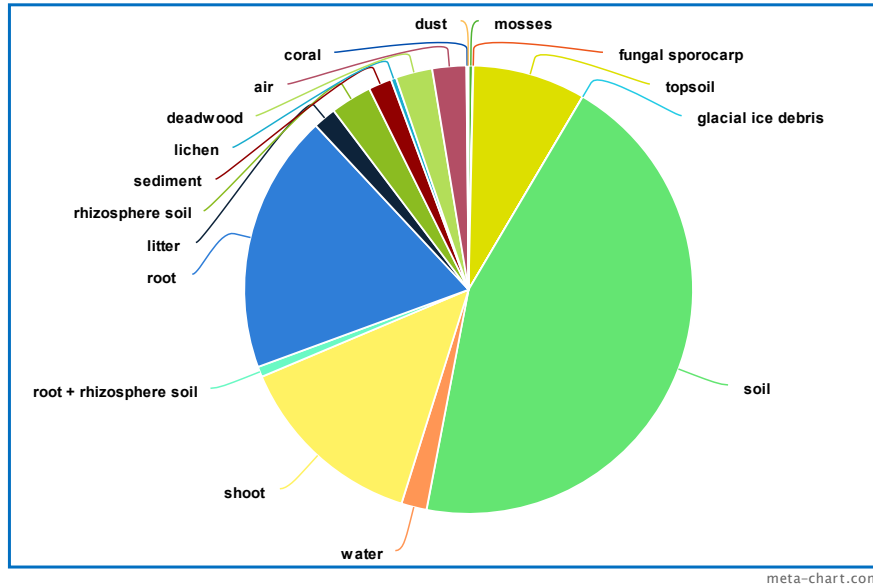- ELIXIR-CZ provides hardware and contributes to development
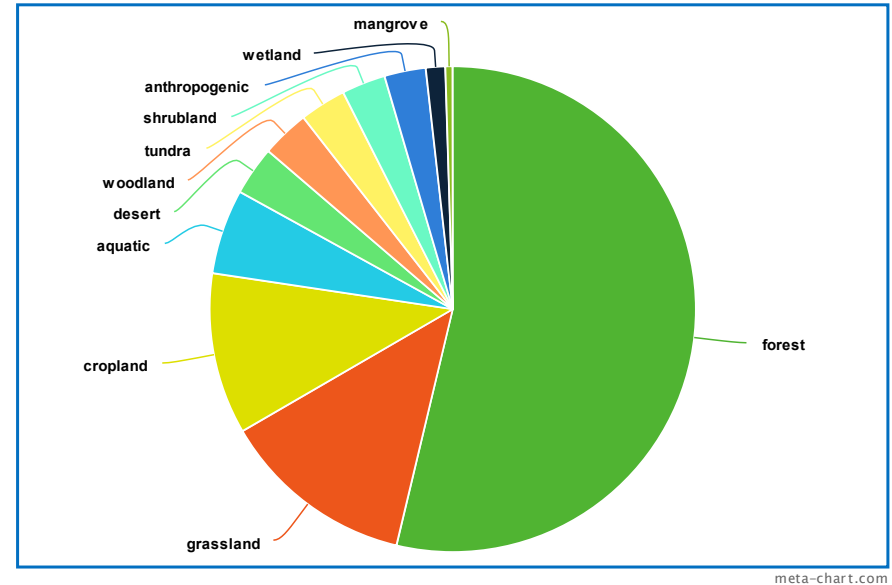
Database content – fungal species

Boletus edulis

additional sample metadata

https://globalfungi.com

Database content - samples

Sample types

Habitats

Database functions

**GlobalFungi Database**

Sequence search

>bfe0c1ab088e7e908fc275fe7b95db91
AATTCTCAACCATGTCTTGAATTAGGCATGGCTTGGACTTGGGGGTTGCTGGCACACGTTGTCAAGCTC
TCCTGAAATGCATTAGCGATGTTCAGCAAGCCTGAACGTGCACGGCCTTTTCGACGTGATAACGATCGTC
GTGGGCTGGAGCGTAGGGTGAGCAGTGAATCGCTTCTAATCTATACCTCAGTTACTAGTCGGTCGT

BLAST database
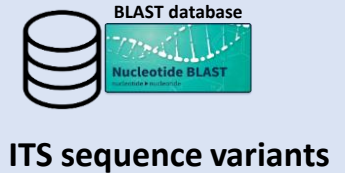
Nucleotide BLAST

ITS sequence variants

Taxon seach

SH (e.g.: SH1644963.08FU)
Species (e.g.: Boletus edulis)
Genus (e.g.: Boletus)
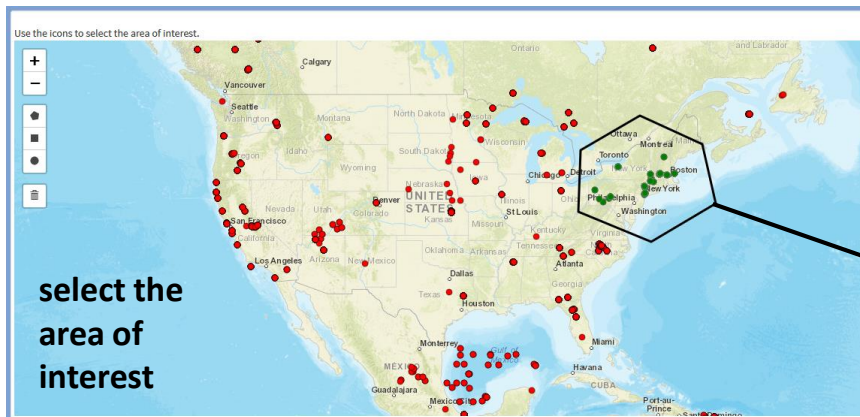
Studies

Search (e.g.: "Plant invasion impacts on fungal community")

MySQL database

MariaDB

Tables with taxonomy and Environmental variables

Geoseach

select the area of interest
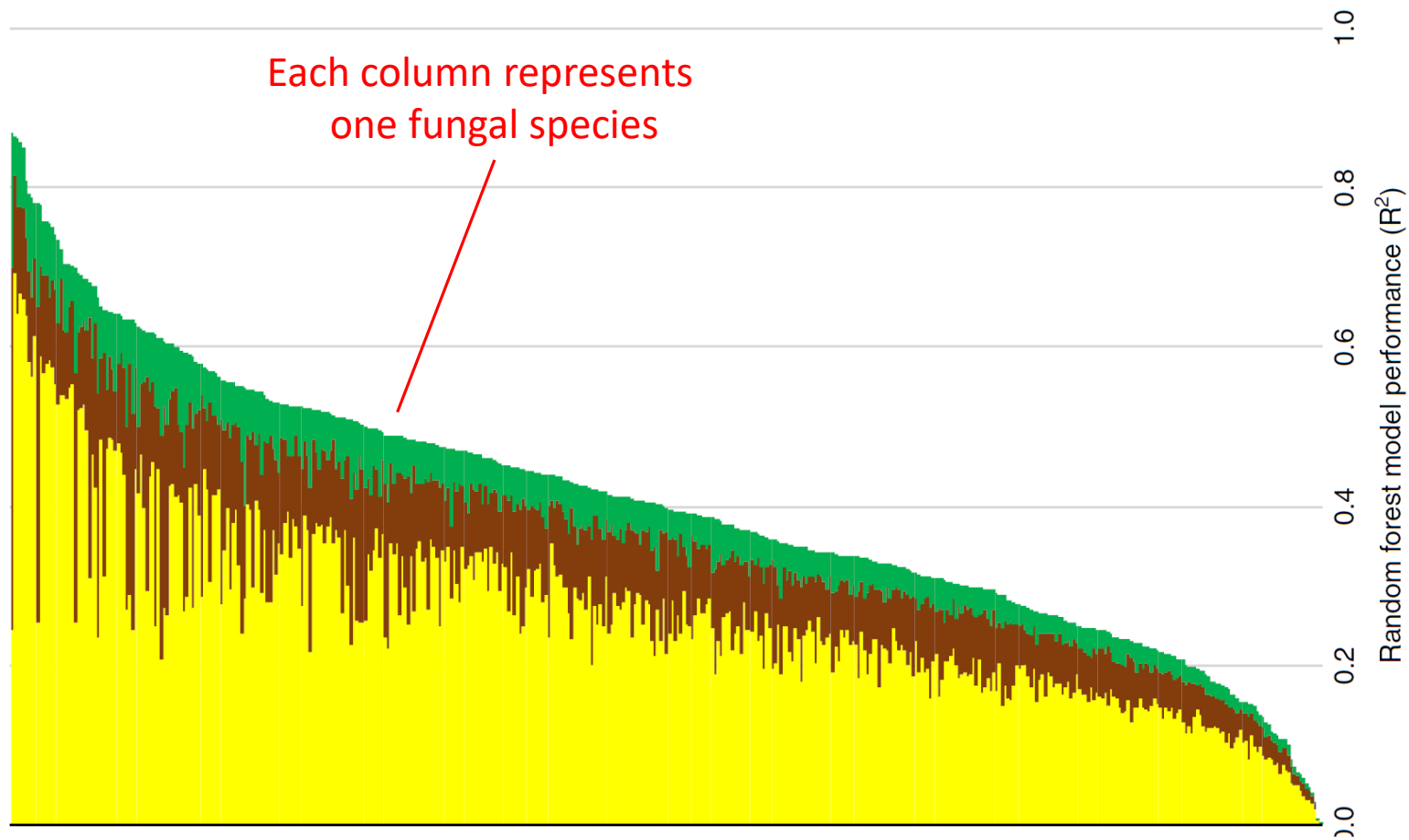
get observed species

Submit your study!

Both databases invites participation of the scientific community in that it encourages submission of data by the authors of studies that are not yet covered.

The science: Drivers of fungal distribution

Each column represents one fungal species

Random forest model performance (R^2)

- Climatic factors explain most of fungal distribution.
- The most important climatic variables are complex, indicating, e.g., aridity

Větrovský et al. 2019 Nature Communications

The science: The future of pathogens
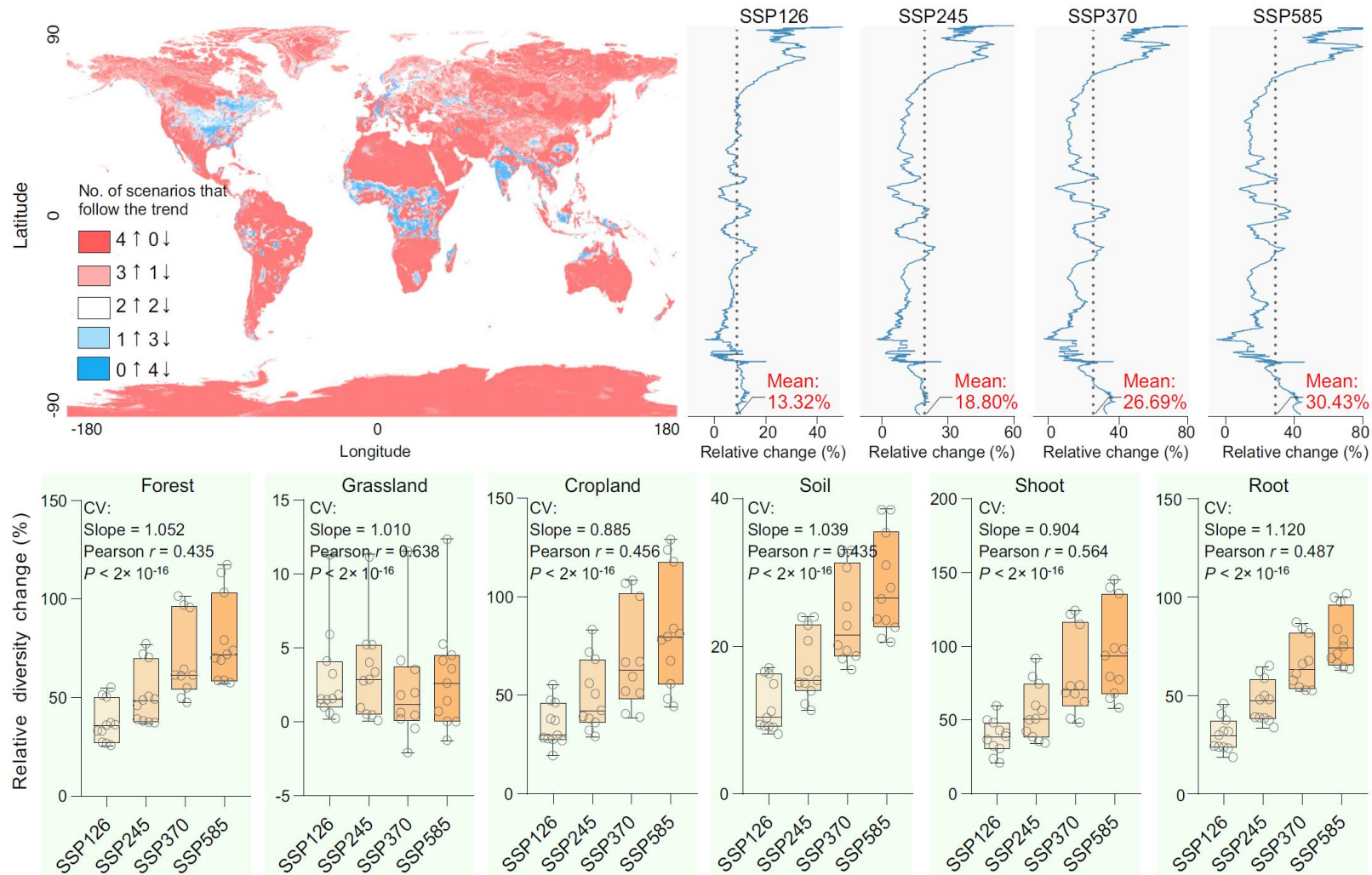
- Increase in pathogen diversity is predicted under most climate change scenarios

Li 2023 Nature Communications

The science: New species

a

b

c

e

① A. rodenticum
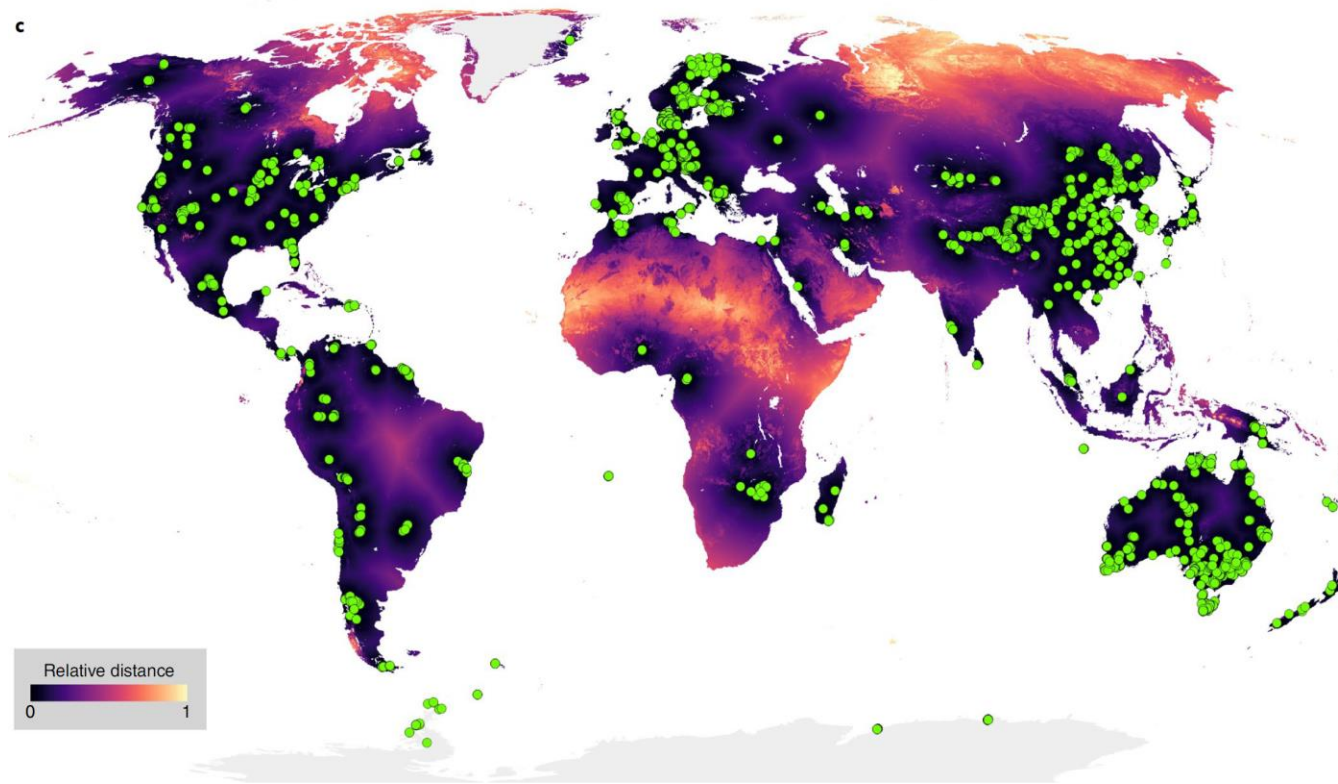② A. simile
③ A. zoogenum
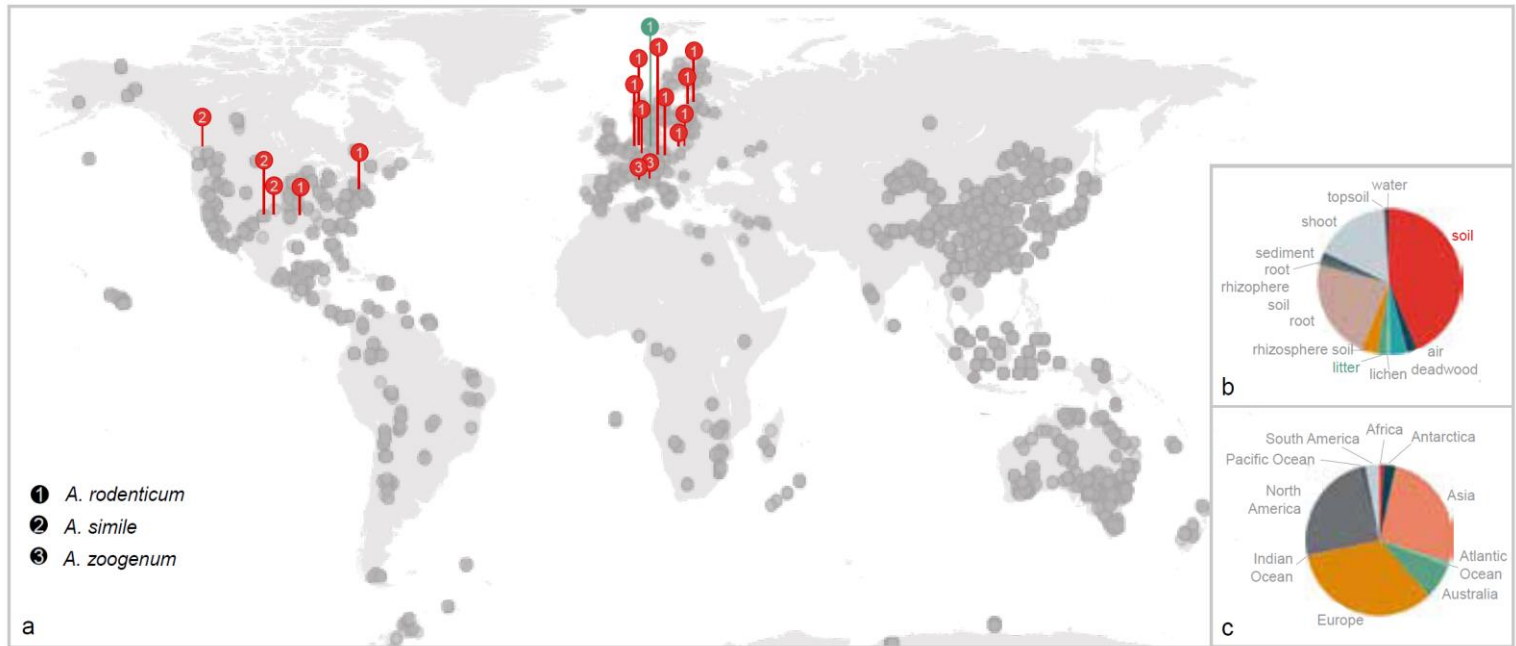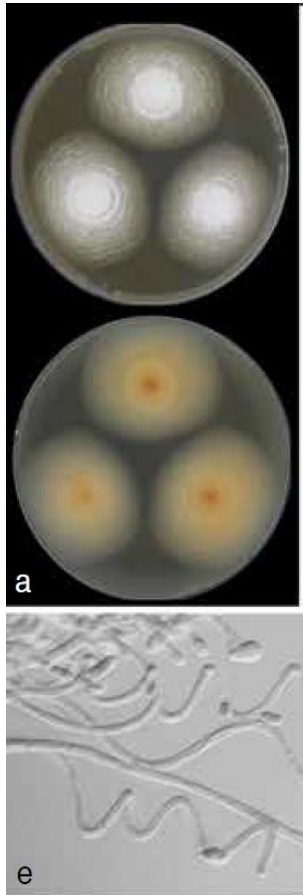
- Newly described, potentially pathogenic *Arthroderma* species from wild rodents. Where are their reservoirs?

Moulíková 2023 Persoonia

The science: Directing future research

Directing future research priorities

Relative distance
0   1

- Unexplored areas of global biodiversity

https://www.spun.earth

THANK YOU FOR YOUR ATTENTION

Petr Baldrian

baldrian@biomed.cas.cz