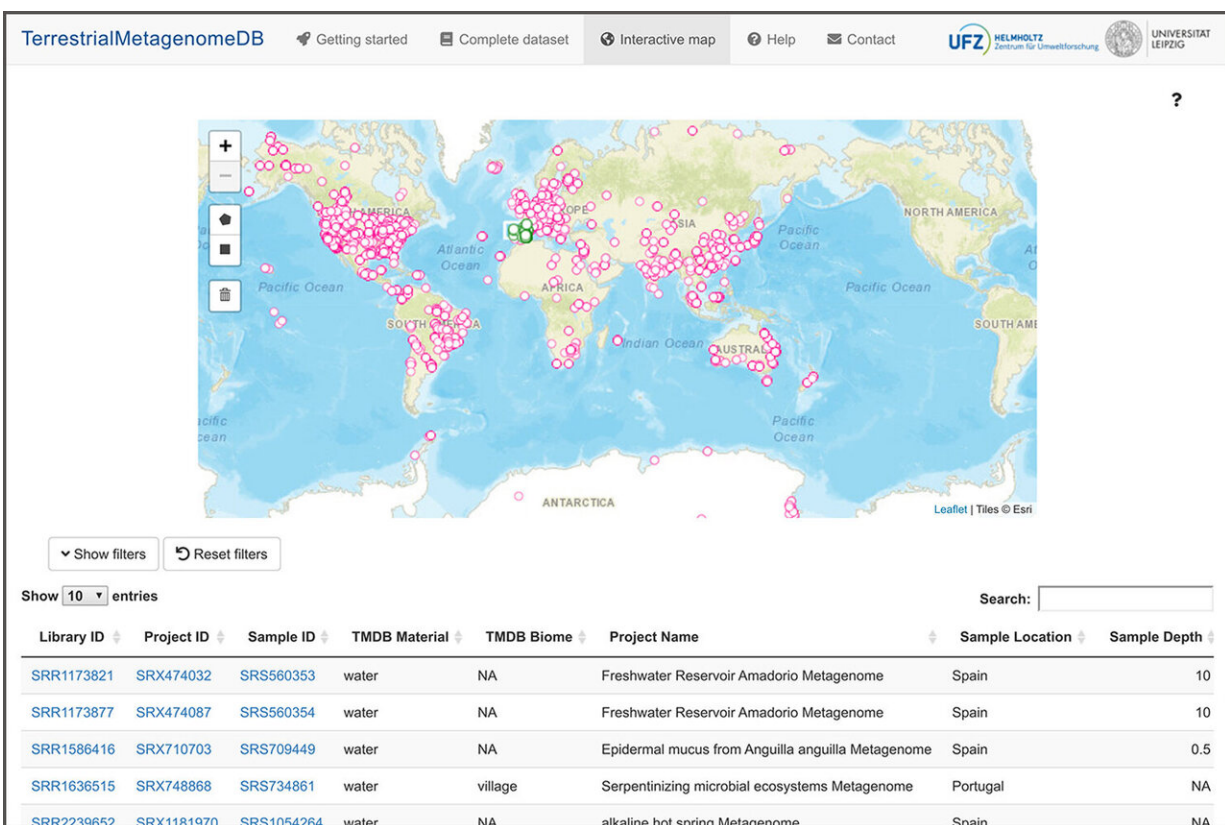


# New metadata database for terrestrial metagenomes: Helping discover the diversity in soil

February 5 2020



The screenshot shows the TerrestrialMetagenomeDB website interface. At the top, there are navigation links: Getting started, Complete dataset, Interactive map (selected), Help, and Contact. Logos for UFZ HELMHOLTZ Zentrum für Umweltforschung and UNIVERSITÄT LEIPZIG are also present. The main content area features a world map with numerous pink circular markers indicating sample locations. Below the map are controls for 'Show filters' and 'Reset filters'. A search bar is located to the right of the map. Below the map, a table displays the first five entries of the database.

Library ID	Project ID	Sample ID	TMDB Material	TMDB Biome	Project Name	Sample Location	Sample Depth
SRR1173821	SRX474032	SRS560353	water	NA	Freshwater Reservoir Amadorio Metagenome	Spain	10
SRR1173877	SRX474087	SRS560354	water	NA	Freshwater Reservoir Amadorio Metagenome	Spain	10
SRR1586416	SRX710703	SRS709449	water	NA	Epidermal mucus from Anguilla anguilla Metagenome	Spain	0.5
SRR1636515	SRX748868	SRS734861	water	village	Serpentinizing microbial ecosystems Metagenome	Portugal	NA
SRR2239652	SRX1181970	SRS1054264	water	NA	alkaline hot spring Metagenome	Spain	NA

Metadata database "TerrestrialMetagenomeDB"(screenshot) Credit: UFZ

Microbiological communities, which include bacteria, single-celled organisms and nematodes, reveal a great deal of information about the

state of soils. All around the world, a lot of research is being performed on this biodiversity at a genetic level, but third parties are not always able to put these research results to the best possible use. The reason is that the information recorded in databases varies in terms of quality. UFZ researchers have now built up a new metadata database for terrestrial metagenomes with over 15,000 datasets, which is intended to make work easier for scientists. This was published in the scientific journal *Nucleic Acids Research*.

More than 202,000 metagenomes, i.e. the entire genetic information contained in a given soil sample, can be found in the two most important databases in which microbiologists can archive [research data](#): the MG-Rast and Sequence Read Archive (SRA) repositories. Here, international researchers have recorded where they performed investigations into microbiological communities or genome sequencing on the seabed, in forests, in grassland or on rocks, and their findings. By doing so, they enable other researchers to use this data in their own research activities and compare it to their own findings. Additionally, it saves them from having to repeat time-consuming work on questions that may have already been answered.

The researchers do, however, come across obstacles to their work time and time again: The datasets are often incomplete and not uniformly marked. "This makes it more difficult for interested users to further process the data," says Dr. Ulisses Nunes da Rocha, microbiological ecologist at the UFZ and one of the study's lead authors. This starts with minor details, such as the temperature. Temperature can be recorded in different ways using Fahrenheit, Kelvin or Celsius; in addition, the way in which the units are abbreviated varies. But there is also uncertainty with regard to what may seem to be basic issues; for example, some scientists around the world have different understandings of the exact definition of a biome (the scientific term for a large-scale habitat). All this, says Dr. da Rocha, makes it more difficult to use the data

efficiently.

Dr. Ulisses Nunes da Rocha and his team have now filtered the metagenome data out of the MG Rast and SRA datasets collected by researchers in terrestrial environments. In contrast, they screened out data collected from the seas and oceans. Exactly 15,022 metagenome datasets from forests or grasslands or from the subsoil originating in 84 countries were brought together in the new metadata [database](#). They did not develop any new scientific standards for the exact description of this metadata, such as the geographical coordinates, the pH value or the temperatures involved, instead using an existing method of standardization.

"The metadata database helps researchers whose work centers on the terrestrial environment and who want to incorporate data of this kind into their own work," says the UFZ researcher. Instead of performing complex laboratory experiments for the purpose of CO<sub>2</sub> fixation or establishing the effect of pesticides on microbiological communities, to name two examples, researchers can consult the database to see if researchers somewhere else have already performed similar experiments on this topic and have made their data available.

The UFZ's freely accessible TerrestrialMetagenomeDB metadata database went online at the beginning of November. Users can initially use six filters, such as the origin of the biome, sample type or the data source to search the database and, if necessary, track down more specific data by means of a further 33 filters. Secondly, another approach provides an interactive map of the world that users can use to look for datasets according to geographical features. Three video tutorials offer additional user support on how best to conduct research and download the data.

The metadata database is automatically updated twice a year—in January

and July. As part of this process, new or corrected datasets are automatically retrieved from the MG-Rast and SRA repositories, assuming that groups of scientists have adapted the attributes of their own data to the standards of the new database. There is great potential, as there are another 100,000 or so datasets on hold containing data on terrestrial metagenomes that could not be standardized to date because the data had not been entered precisely enough. For Dr. Ulisses Nunes da Rocha and his UFZ Microbiological Systems Data Science working group, this is only the first step in a process of facilitating big data analyses of microbiological communities in terrestrial systems on a global scale.

**More information:** Felipe Borim Corrêa et al, TerrestrialMetagenomeDB: a public repository of curated and standardized metadata for terrestrial metagenomes, *Nucleic Acids Research* (2019). [DOI: 10.1093/nar/gkz994](https://doi.org/10.1093/nar/gkz994)

Provided by Helmholtz Association of German Research Centres

Citation: New metadata database for terrestrial metagenomes: Helping discover the diversity in soil (2020, February 5) retrieved 23 June 2024 from <https://phys.org/news/2020-02-metadata-database-terrestrial-metagenomes-diversity.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--