REFLECT DELIVERABLE D2.5

DATABASE OF MICROBES AND ORGANIC COMPOUNDS

Summary:

This deliverable contains the raw data that constitutes the database of microbial diversity and organic compounds in geothermal fluids used for electricity production generated during the project.

Authors:

Danaé Bregnard, UNINE, PhD student; Alessio Leins, GFZ, PhD student; Andrea Vieth-Hillebrand, GFZ, Senior Scientist; Simona Regenspurg, GFZ, Senior Scientist; Pilar Junier, UNINE, Professor;

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement nº 850626.

Title:	Database of microbes and organic compounds
Lead beneficiary:	UNINE, GFZ
Other beneficiaries:	GFZ
Due date:	30th April 2023
Nature:	Public
Diffusion:	all Partners, EC and general public
Status:	Final
Document code:	Reflect_D2.5
DOI:	https://doi.org/10.48440/gfz.4.8.2023.005
License information:	CC-BY_4.0
Recommended citation:	Bregnard D., Leins A., Vieth-Hillebrand A., Regenspurg S., Junier P., The H2020 REFLECT project: Deliverable *Deliverable_2.5 – Database of microbes and organic compounds, GFZ German Research Centre ,* DOI: *https://doi.org/10.48440/gfz.4.8.2023.005*
Related data	Leins, Alessio; Vieth-Hillebrand, Andrea; Günther, Kristin; Regenspurg, Simona (2023): Dissolved organic compounds in geothermal fluids used for energy production – part II. GFZ Data Services. https://doi.org/10.5880/GFZ.4.8.2023.005 Bregnard et al., in prep. http://www.ncbi.nlm.nih.gov/bioproject/956828

Revision history	Author	Delivery date	Summary of changes and comments
Version 01	DB; AL	06.04.2023	Draft assembly
Version 02	DB; AL; AVH	13.04.2023	Draft corrections
Version 02	PJ	19.04.2023	Draft corrections
Final version	KK	27.04.2023	Included DOI and link for data publication on organis

Approval status				
	Name	Function	Date	
Deliverable responsible	P. Junier	Leader Task 2.5	20.04.2023	
Deliverable responsible	A. Vieth-Hillebrand	Task member WP 2.5		
Task Leader	P. Junier	Leader Task 2.5	20.04.2023	
WP leader	C. Boeije	Leader WP2		
Reviewer	S. Regenspurg	project coordinator	22.04.2023	
Project Coordinator	K. Kieling	Project manager	26.04.2023	

This document reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

TABLE OF CONTENTS

1 EXECUTIVE SUMMARY

This report contains the description of two datasets that were collected within the framework of the H2020 REFLECT project (Redefining geothermal properties at extreme conditions to optimize future geothermal energy extraction).

The dataset on microorganisms includes the raw sequencing files obtained after sequencing of DNA extracted from samples obtained from 7 geothermal power plants (Austria, Germany, Iceland, The Netherlands). The dataset on organic compounds comprises the compounds found after the analysis of 47 fluid samples from 11 geothermal sites (Germany, Austria, Iceland, Turkey, Netherlands, Belgium, French West Indies). Files availability and descriptions, as well as the methods used to generate each dataset, are described in the report.

The dataset on microorganisms is available on the NCBI repository under the following accession number PRJNA956828, with the following project title: “Diversity of bacteria, archaea and fungi from geothermal fluids used for electricity production”. The dataset will be available for download after the publication of the associated article (*Bregnard et al., in prep*) using the following link: http://www.ncbi.nlm.nih.gov/bioproject/956828.

The dataset on dissolved organic compounds is published as: (Leins, Alessio; Vieth-Hillebrand, Andrea; Günther, Kristin; Regenspurg, Simona (2023): Dissolved organic compounds in geothermal fluids used for energy production – part II. GFZ Data Services. https://doi.org/10.5880/GFZ.4.8.2023.005 . The dataset is available using the following link: https://dataservices.gfz-potsdam.de/panmetaworks/showshort.php?id=86c7caa8-06ca-11ed-9531-ca1f3ed77ce8)

2 DATASET OF THE MICROBIAL DIVERSITY IN GEOTHERMAL FLUIDS USED FOR ENERGY PRODUCTION

The dataset is available on the NCBI repository under the following accession number PRJNA956828, with the following project title: "Diversity of bacteria, archaea and fungi from geothermal fluids used for electricity production". The dataset will be available for download after the publication of the associated article (*Bregnard et al., in prep*) using the following link: http://www.ncbi.nlm.nih.gov/bioproject/956828.

2.1 DATA DESCRIPTION

This dataset includes the raw sequencing files obtained after sequencing of DNA extracted from samples collected in seven geothermal power plants (Austria, Germany, Iceland, The Netherlands). The samples were collected within the framework of the H2020 REFLECT project (Redefining geothermal properties at extreme conditions to optimize future geothermal energy extraction). The focus of the analysis was on the archaeal, bacterial and fungal diversity present in the fluids. Two fractions of the community were analyzed in parallel: total diversity and the diversity of cellular structures withstanding lysis.

The goal of this dataset is to provide an overview of the bacterial, fungal and archaeal diversity present in different geothermal fluids used for electricity production across Europe. This type of data is mostly absent of deep geothermal studies (exploratory and for exploitation), but could provide essential information regarding the understanding of the biogeochemical processes within the fluids and in the reservoirs. This crucial information needs to be considered in the future as part of the sustainability and mitigation strategies for the problems arising during operation. This dataset is linked to an article in preparation (*Bregnard et al., in prep*).

2.2 MATERIAL AND METHODS

2.2.1 Sampling

Geothermal fluids were taken between 18.06.2020 and 06.08.2022. Fluids were cooled down if needed for handling and either filtered directly on site or transported in aseptic containers back to the laboratory. Transported fluids were stored at 4°C upon arrival until filtration. Filtration was performed using 0.22 µm nitrocellulose filters, which were sterilized prior to their use. Filters were stored at 4°C in sterile containers until further processing.

2.3 ANALYTICAL PROCEDURE

2.3.1 Lysis-enrichment treatment and DNA extraction

Each filter was split into two. In order to enrich for cells that are highly resistant to lysis, a lysis-enrichment method was applied to one half of each filter. The lysis-resistant enrichment treatment was performed as described previously (Wunderlin *et al.*, 2016; Corona Ramirez *et al.*, 2023). DNA extraction was then done on both the total microbial community (half filter not subjected to the lysis-enrichment treatment) and the lysis enriched fraction of the

microbial community using the FastDNA®SPIN kit for soil (MP Biomedicals, USA). The standard protocol of the kit was modified to include three successive bead-beating steps and the three DNA extracts obtained from the same samples were pooled after elution, as described previously (Wunderlin *et al.*, 2016). DNA was quantified using the Qubit® dsDNA HS Assay Kit on a Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA).

2.3.2 Sequencing

For targeted amplicon sequencing, DNA extracts were sent to Fasteris (Geneva, Switzerland) and sequencing was performed using an Illumina MiSeq platform (Illumina, San Diego, CA, USA) to generate 300 bp paired-end reads. For Archaea, the V3-V4 region of the 16S rRNA gene was amplified using the 340F (5'-CCC TAY GGG GYG CAS CAG-3') and 806rB (5'-GGA CTA CNV GGG TWT CTA AT-3') primer pair (Bahram *et al.*, 2019). For Bacteria, the V3-V4 region of the 16S rRNA gene was amplified using the Bakt_341F (5'-CCT ACG GGN GGC WGC AG-3') and Bakt_805R (5'-GAC TACHVG GGT ATCTAA TCC-3') primers (Herlemann *et al.*, 2011). For Fungi, the ITS2 region was amplified using the ITS3 KYO2 (5'-GAT GAA GAA CGY AGY RAA-3') and ITS4 (5'-TCC TCCGCT TAT TGA TAT GC-3') primer pair (Toju *et al.*, 2012). Demultiplexed sequences were provided by Fasteris.

2.4 DATA PROCESSING

No data treatment was applied.

2.5 REPOSITORY

The dataset is available on the NCBI repository under the following accession number PRJNA956828, with the following project title: "Diversity of bacteria, archaea and fungi from geothermal fluids used for electricity production". The dataset will be available for download after the publication of the associated article (*Bregnard et al., in prep*) using the following link: http://www.ncbi.nlm.nih.gov/bioproject/956828.

2.6 FILE INVENTORY

The file inventory, in the NCBI BioProject (PRJNA956828), consists of the following files:

- Meta_ENV_Reflect_DATA.xlsx
- 6 raw sequencing files per sample (.fastq.gz)
- SRA_metadata.xlsx

2.7 DESCRIPTION OF DATA TABLES

2.7.1 Meta_ENV_Reflect_DATA.xlsx

The file Meta_ENV_Reflect_DATA.xlsx includes relevant metadata information linked to each sample (see Table 1 for description).

Table 1 : Description of the metadata included in the file Meta_ENV_Reflect_DATA.xlsx.

Column header	Unit or possibilities	Description
sample_name		Name of the sample
organism	Archaea; Bacteria; Fungi	Target organism (i.e. Archaea, Bacteria or Fungi)
host		Not applicable
isolation_source	Geothermal fluids	Source of the sample
collection_date		Sampling date (year – month – day)
geo_loc_name		Country: power plant name
lat_lon		GPS coordinates
altitude	meters	Altitude of the power plant
Treatment	Total; Lysis	Treatment of the sample (Total or lysis-resistant enrichment)
Well type	Production; Middle; Injection	Type of well sampled
Reservoir_depth_m	meters	Depth of the reservoir or the well
Well ID		Name of the well
Pair		An assigned number for each pair of samples (total or lysis from the same sample)

2.7.2 Sequencing files (.fastq.gz)

Each half filter is assigned to a sample name. For each sample name, 6 raw sequencing files are provided, following the sample_name column of the Meta_ENV_Reflect_DATA.xlsx file. Two of these files belong to archaeal sequences (samplename_Archae_R1.fastq.gz and samplename_Archae_R2.fastq.gz), two belong to bacterial sequences (samplename_16S_R1.fastq.gz and samplename _16S_R2.fastq.gz), and two belong to fungal sequences (samplename_ITS3_R1.fastq.gz and samplename _ITS3_R2.fastq.gz). The treatment of the sample combines the sample name, with either an R (reverse primer) or F (forward primer), the organism targeted (16S, Archae, ITS3) and the treatment (Total or Lysis).

2.7.3 SRA_metadata.xlsx

The SRA_metadata.xlsx file contains the list of all .fastq.gz files as well as related information on the methodology.

Column header	Possibilities	Description
sample_name		Name of the sample
Library_ID	Ex. 10_16S	Name of the sample combined to the targeted gene (16S or ITS)
Title	Bacterial / Archaeal / Fungal diversity (metagenomic 16S rRNA seq data) in geothermal fluid	Title of the project
library_strategy	AMPLICON	Method
library_source	METAGENOMIC	Method
library_selection	PCR	Method
library_layout	paired	Type of sequences
platform	ILLUMINA	
instrument_model	Illumina MiSeq	
design_description	Metagenomic 16S rRNA seq data in geothermal fluids	
filetype	fastq	
filename	Ex. 10_16S_R1.fastq.gz	
Filename 2	Ex. 10_16S_R2.fastq.gz	

2.8 ACKNOWLEDGEMENTS

The authors would like to thank all project partners involved in the sampling campaigns (ISOR, TNO, Hydroisotop GmbH, Natürlich Insheim GmbH, Neustadt-Glewe GmbH, Landsvirkjun, Spa Therme Blumau Betriebs GmbH, B.V., Floricultura Heemskerk,). This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement nº 850626 (REFLECT).

2.9 REFERENCES

Bahram, M. *et al.* (2019) 'Newly designed 16S rRNA metabarcoding primers amplify diverse and novel archaeal taxa from the environment', *Environmental Microbiology Reports*, 11(4), pp. 487–494. Available at: https://doi.org/10.1111/1758-2229.12684.

Corona Ramirez, A. *et al.* (2023) 'Assessment of fungal spores and spore-like diversity in environmental samples by targeted lysis', *BMC Microbiology*, 23(1), p. 68. Available at: https://doi.org/10.1186/s12866-023-02809-w.

Herlemann, D.P. *et al.* (2011) 'Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea', *The ISME Journal*, 5(10), pp. 1571–1579. Available at: https://doi.org/10.1038/ismej.2011.41.

Toju, H. *et al.* (2012) 'High-Coverage ITS Primers for the DNA-Based Identification of Ascomycetes and Basidiomycetes in Environmental Samples', *PLOS ONE*, 7(7), p. e40863. Available at: https://doi.org/10.1371/journal.pone.0040863.

Wunderlin, T. *et al.* (2016) 'Physical Isolation of Endospores from Environmental Samples by Targeted Lysis of Vegetative Cells', *JoVE (Journal of Visualized Experiments)*, (107), p. e53411. Available at: https://doi.org/10.3791/53411.

3 DISSOLVED ORGANIC COMPOUNDS IN GEOTHERMAL FLUIDS USED FOR ENERGY PRODUCTION – PART II

This chapter is a data publication from the GFZ publication service (Leins et al. 2023, https://dataservices.gfz-potsdam.de/panmetaworks/showshort.php?id=86c7caa8-06ca-11ed-9531-ca1f3ed77ce8). and is a continuation of a first data publication of organic compounds in geothermal fluids (Leins et al. 2022).

3.1 DATA DESCRIPTION

This dataset comprises 47 fluid samples from 11 geothermal sites (Germany, Austria, Iceland, Turkey, Netherlands, Belgium, French West Indies). The samples were collected within the REFLECT project (Redefining geothermal properties at extreme conditions to optimize future geothermal energy extraction). The focus of these analyses was on the dissolved organic compound composition of the fluids, since they are rarely included in the regular analyses of fluids taken from geothermal power plants. Understanding the dissolved organic compound composition of geothermal fluids might help to better understand chemical reactions within the fluids and might help to mitigate problems that arise with the operation of a geothermal power plant such as mineral precipitation (scaling) and corrosion of the casing and pipes.

Liquid chromatography organic carbon detection (LC-OCD) was conducted to quantify and characterize the dissolved organic carbon (DOC). Additionally, ion chromatography (IC) was conducted to detect organic acid anions (formate, acetate, propionate, butyrate, valerate, oxalate) as well as inorganic anions such as chloride (Cl^-), bromide (Br^-), fluoride (F^-), sulfate (SO_4^{2-}), phosphate (PO_4^{3-}) and nitrate (NO_3^-). When there is no value given, this can be regarded as “not detected”.

The data includes fluids from low- to high temperature geothermal sites (80-346°C) with chloride concentrations ranging from 0.03-155 g/L. DOC concentrations vary from 0.17-27.7 mg C/L. Moreover, sulfite and thiosulphate were detected in the Icelandic samples. However, as our sampling procedures and sample preparation methods are not adequate for reliable detection of these compounds, there are no concentrations reported.

3.2 MATERIAL AND METHODS

3.2.1 Sampling

The samples were taken by the project partners and sent to GFZ for the analyses. Generally, the fluid samples were taken at the surface installations. Sampling bottles were rinsed and flushed with the fluids prior to filling them up. Samples were stored at 4°C until analyses. Sampling volumes ranged from 250-500 ml per sample.

3.3 ANALYTICAL PROCEDURE

All analytical work was done at GFZ Helmholtz Centre Potsdam German Research Centre for Geosciences in Section 3.2 Organic Geochemistry.

3.3.1 IC (Ion Chromatography)

The quantification of organic anions (formate, acetate, propionate, butyrate, valerate, oxalate) and inorganic anions (F^-, Cl^-, Br^-, SO_4^{2-}, PO_4^{3-}, NO_3^-) was conducted via ion chromatography (ICS 3000, Thermo Fisher Scientific) using an AS-AP autosampler, AS11 HC column and a conductivity detector. KOH solutions with varying concentrations over time were used as eluent for the samples. The initial KOH concentration was 1.4 mM and stepwise increased towards 60 mM within 32 min. After 32 min the concentration was reduced to the initial value of 1.4 mM and equilibrated for 12 min. The flow-rate was 0.38 ml/min. 10 µL of sample were injected for each run. The quality of the measurements was daily verified using standards that contain the analytes in different concentrations. The concentrations were 0.02; 1.0; 10 and 100 mg/L. For samples with chloride concentrations exceeding 1 g/L, the chloride was reduced prior to the analysis of the organic anions using OnGuard II AG/H cartridges (Thermo Fisher Scientific).

3.3.2 LC-OCD (Liquid Chromatography Organic Carbon Detection)

The characterization and quantification of the dissolved organic carbon (DOC) and its fractions were determined by size-exclusion-chromatography (SEC) with subsequent UV (λ = 254 nm) and IR detection by a LC-OCD system. Phosphate buffer (pH 6.85; 2.7 g/L KH_2PO_4, 1.6 g/L Na_2HPO_4) was used as mobile phase set to a flow of 1.1 mL/min (Huber et al., 2011). The samples passed a 0.45 µm membrane syringe filter before entering the chromatographic column. With the LC-OCD the organic matter can be separated into different fractions according to their molecular mass: Macro.1/Biopolymers (> 10,000 Da), Macro.2/Humic Substances (~ 1000 Da), Macro.3/Building Blocks (350-500 Da), low molecular weight acids/LMWA (< 350 Da), and low molecular weight neutrals/LMWN (< 350 Da; Zhu et al., 2015). The hydrophobic organic carbon (HOC) is calculated from the difference between DOC and sum of the chromatographically derived fractions listed before. The DOC was quantified by IR-detection of the released CO_2 after UV-oxidation (λ = 185 nm) in a Gräntzel thin-film reactor. Humic and fulvic acids standards of the Suwannee River, provided by the International Humic Substances Society (IHSS), were used for molecular mass calibration. Fluids with chloride concentrations exceeding 1 g/L were diluted prior to the measurement to avoid disturbances during chromatographic separation (Henne 2011, Sachse & Mierke, 2014).

3.4 DATA PROCESSING

No data treatment was applied.

3.5 FILE INVENTORY

The file inventory is summarized in one document „2023-005_Leins_et-al_Data.txt“

3.6 DESCRIPTION OF DATA TABLE

3.6.1 2023-005_Leins_et-al_Data.txt

Table 2: The File 2023-005_Leins_et-al_Data.txt contains results of organic and inorganic compound analyses of fluid samples from geothermal sites. Note: In the case of the Bad Blumau samples, the columns Makro.1 to Makro.3 are given as one Makro fraction. In these samples, only one peak showed up in the retention time frame of all 3 Makro fractions, making the differentiation into Makro.1 to Makro.3 impossible.

Column header	unit	Description
GFZ Sample ID		Sample Identifier in section 3.2 (Organic Geochemistry) at GFZ German Research Centre for Geosciences.
Site		Name of the sampled geothermal site
Country		Country in which the site is located
Latitude	Northing	Coordinate
Longitude	Easting	Coordinate
Sampling Date	YYYY-mm-dd	Date on which the sampling was conducted
Sampling Well		Well at which the samples were taken. Included are also additional sampling points on the surface in the installed power plant.
Local Well ID		Local ID of the well from which samples were taken.
Remark		Column with additional remarks for the sample
Well Head Temperature	°C	Temperature of the fluids at the well head
Bottomhole Temperature	°C	Temperature at the bottom of the well
Well Depth t.v.d	m	True vertical depth of the sampled well
Cl	g/L	Chloride concentration
Br	mg/L	Bromide concentration
SO4	mg/L	Sulphate concentration
F	mg/L	Fluoride concentration
Acetate Acetate_C	mg/L mg C/L	Acetate concentration Acetate concentration in mg carbon/L
Propionate Propionate_C	mg/L mg C/L	Propionate concentration Propionate concentration in mg carbon/L
Formate Formate_C	mg/L mg C/L	Formate concentration Formate concentration in mg carbon/L
Butyrate Butyrate_C	mg/L mg C/L	Butyrate concentration Butyrate concentration in mg carbon/L
Valerate Valerate_C	mg/L mg C/L	Valerate concentration Valerate concentration in mg carbon/L
Oxalate Oxalate_C	mg/L mg C/L	Oxalate concentration Oxalate concentration in mg carbon/L

DOC	mg C/L	Dissolved organic carbon concentration in mg carbon/L
HOC	mg C/L	Hydrophobic organic carbon (Fraction of the DOC)
Makro 1	mg C/L	Biopolymers (Fraction of the DOC)
Makro 2	mg C/L	Humic substances (Fraction of the DOC)
Makro 3	mg C/L	Building Blocks (Fraction of the DOC)
LMWA	mg C/L	Low Molecular Weight Acids (Fraction of the DOC)
LMWN	mg C/L	Low Molecular Weight Neutrals (Fraction of the DOC)

3.7 ACKNOWLEDGEMENTS

We would like to thank all project partners and operators of geothermal plants (ISOR, IZTECH, TNO, BRGM, Hydroisotop GmbH, Natürlich Insheim GmbH, Neustadt-Glewe GmbH, Landsvirkjun, Spa Therme Blumau Betriebs GmbH, Ammerlaan Geothermie B.V., Floricultura Heemskerk, Gümüsköy Jeotermal Elektrik Üretim A.S., Tuzla Jeotermal Enerjy A.S.) who helped with the sampling and shipping. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement nº 850626 (REFLECT).

3.8 REFERENCES

Henne, S. 2011. Geochemie von geothermischen Fluiden. Diploma Thesis. Technische Universität Bergakademie Freiberg. Fakultät für Geowissenschaften, Geotechnik und Bergbau.

Huber, S.A. et al. (2011) 'Characterisation of aquatic humic and non-humic matter with size-exclusion chromatography – organic carbon detection – organic nitrogen detection (LC-OCD-OND)', Water Research, 45(2), pp. 879–885. Available at: https://doi.org/10.1016/j.watres.2010.09.023.

Leins, A., Vieth-Hillebrand, A., Bregnard, D., Günther, K., Junier, P., Regenspurg, S. (2022): Dissolved organic compounds in geothermal fluids used for energy production.https://doi.org/10.5880/GFZ.4.8.2022.001

Sachse, A. & Mierke, W. 2014. Einfluß von Salzen auf die quantitative und qualitative Analyse mit der LC-OCD. Unpublished report.

Zhu, Y. et al. (2015) 'Characterization of water-soluble organic compounds released from black shales and coals', International Journal of Coal Geology, 150–151, pp. 265–275. Available at: https://doi.org/10.1016/j.coal.2015.09.009.

4 CONCLUSION

These two databases provide an overview of the diversity of microorganisms and organic compounds present in geothermal fluids used for electricity production in several geothermal power plants across Europe. The raw datasets are available at the NCBI database (PRJNA956828; http://www.ncbi.nlm.nih.gov/bioproject/956828) and at the GFZ data repository (https://dataservices.gfz-potsdam.de/panmetaworks/showshort.php?id=86c7caa8-06ca-11ed-9531-ca1f3ed77ce8).