Sulfur-fueled chemolithoautotrophs replenish organic carbon inventory in groundwater

Martin Taubert[1,+], Beatrix M. Heinze[1], Will A. Overholt[1], Georgette Azemtsop[2], Rola Houhou[2], Nico Jehmlich[3], Martin von Bergen[3,4], Petra Rösch[2], Jürgen Popp[2,5], Kirsten Küsel[1,6]

[1]Aquatic Geomicrobiology, Institute of Biodiversity, Friedrich Schiller University Jena, Dornburger Str. 159, 07743 Jena, Germany

[2]Institute of Physical Chemistry and Abbe Center of Photonics, Friedrich Schiller University Jena, Helmholtzweg 4, 07743 Jena, Germany

[3]Department of Molecular Systems Biology, Helmholtz Centre for Environmental Research – UFZ, Permoserstrasse 15, 04318 Leipzig, Germany

[4]Institute of Biochemistry, Faculty of Biosciences, Pharmacy and Psychology, University of Leipzig, Brüderstraße 32, 04103 Leipzig, Germany

[5]Leibniz-Institute of Photonic Technology, Albert-Einstein-Straße 9, 07745 Jena, Germany

[6]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5E, 04103 Leipzig, Germany

[+]Corresponding author: Martin Taubert; Tel.: +49 3641 949459; Fax: +49 3641 949402; E-mail: martin.taubert@uni-jena.de

Keywords: Chemolithoautotrophy, groundwater, $^{13}CO_2$ stable isotope probing, genome-resolved metaproteomics, metagenomics, Raman microspectroscopy

Short title: Chemolithoautotrophy in shallow groundwater

Abstract

Metagenome-assembled genomes (MAGs) have revealed the existence of novel bacterial and archaeal groups and provided insight into their genetic potential. However, metagenomics and even metatranscriptomics cannot resolve how the genetic potential translates into metabolic functions and physiological activity.

Here, we present a novel approach for the quantitative and organism-specific assessment of the carbon flux through microbial communities with stable isotope probing-metaproteomics and integration of temporal dynamics in ^{13}C incorporation by Stable Isotope Cluster Analysis (SIsCA). We used groundwater microcosms labeled with $^{13}CO_2$ and D_2O as model systems and stimulated them with reduced sulfur compounds to determine the ecosystem role of chemolithoautotrophic primary production. Raman microspectroscopy detected rapid deuterium incorporation in microbial cells from 12 days onwards, indicating activity of the groundwater organisms. SIsCA revealed that groundwater microorganisms fell into five distinct carbon assimilation strategies. Only one of these strategies, comprising less than 3.5% of the community, consisted of obligate autotrophs (*Thiobacillus*), with a ^{13}C incorporation of approximately 95%. Instead, mixotrophic growth was the most successful strategy, and was represented by 12 of the 15 MAGs expressing pathways for autotrophic CO_2 fixation, including *Hydrogenophaga*, *Polaromonas* and *Dechloromonas*, with varying ^{13}C incorporation between 5% and 90%. Within 21 days, 43% of carbon in the community was replaced by ^{13}C, increasing to 80% after 70 days. Of the 31 most abundant MAGs, 16 expressed pathways for sulfur oxidation, including strict heterotrophs. We concluded that chemolithoautotrophy drives the recycling of organic carbon and serves as a fill-up function in the groundwater. Mixotrophs preferred the uptake of organic carbon over the fixation of CO_2, and heterotrophs oxidize inorganic compounds to preserve organic carbon. Our study showcases how next-generation physiology approach like SIsCA can move beyond metagenomics studies by providing information about expression of metabolic pathways and elucidating the role of MAGs in ecosystem functioning.

Introduction

Genome-resolved metagenomics has vastly expanded our knowledge about the microbial communities in Earth's ecosystems (Wrighton et al., 2012; Albertsen et al., 2013; Nielsen et al., 2014; Vollmers et al., 2017). The sequences obtained allowed unprecedented insights into microbial genetic potential and led to the discovery of novel bacterial and archaeal taxa (Brown et al., 2015; Castelle et al., 2015). Studies employing deep sequencing and genome binning approaches have recovered thousands of metagenome-assembled genomes (MAGs) that revealed a broad microbial diversity and greatly expanded the tree of life (Anantharaman et al., 2016; Hug et al., 2016; Parks et al., 2017). Such MAGs enabled researchers to highlight organisms that might play key roles in biogeochemical cycles (Long et al., 2016; Anantharaman et al., 2018), and provide a basis for the establishment of hypothesis about microbial functioning in the investigated habitats. Metagenomics studies targeting the groundwater microbiome, for example, have revealed a high abundance of organisms with the metabolic potential for CO_2 fixation (Emerson et al., 2016; Probst et al., 2018), as well as for the utilization of inorganic electron donors (Anantharaman et al., 2016; Anantharaman et al., 2018; Wegner et al., 2019). These discoveries have led to the assumption that groundwater ecosystems are dominated by chemolithoautotrophic primary production, while other carbon sources, such as the import of surface-derived products of photosynthesis, or ancient organic matter released from sedimentary rocks, play a minor role (Griebler and Lueders, 2009; Akob and Küsel, 2011; Schwab et al., 2019). The actual microbial activities and functions of the key players in the groundwater, however, are still unknown. A quantitative assessment of the flow of autochthonous CO_2-derived carbon through the groundwater microbial food web could provide an experimental validation of the dominance of chemolithoautotrophy. Approaches complementary to metagenomics that allow a determination of the active microbial physiological functions and the elucidation of the complex interactions between organisms, to show how the genetic potential is translated into a phenotype, are required to understand ecosystem community functioning.

To trace carbon fluxes through microbial communities and identify the active key players, stable isotope probing (SIP) has shown to be a valuable strategy (Neufeld et al., 2007; von Bergen et al., 2013). By introducing ^{13}C-labeled CO_2 to microbial communities, SIP offers the opportunity to illuminate the entire carbon flux from primary production through the food web. However, to pinpoint specific interactions and discern trophic fluxes between individual community members, a highly accurate and time-resolved determination of ^{13}C incorporation in biomolecules must be achieved. Even contemporary nucleic acid-based SIP approaches are typically not able to resolve such slight variations in ^{13}C incorporation patterns (Hungate et al., 2015; Starr et al., 2020). To overcome this problem, we leveraged the high sensitivity of cutting-edge Orbitrap mass spectrometric analysis in SIP-metaproteomics (Taubert et al., 2012) for a highly accurate quantitation of ^{13}C incorporation. To integrate the molecule-specific temporal dynamics in the acquired isotopologue patterns from a ^{13}C-SIP time-series experiment, we developed a PCA-based Stable Isotope Cluster Analysis (SIsCA) approach.

We employed this next-generation physiology approach to unravel the role of chemolithoautotrophy for the groundwater microbial community. Groundwater from the Hainich Critical Zone Exploratory (CZE) (Küsel et al., 2016) was supplemented with thiosulfate as electron donor in a $^{13}CO_2$ SIP microcosm experiment. Thiosulfates are common forms of reduced sulfur in the environment (Grimm et al., 2008), and occur in groundwater through pyrite oxidation (Schippers et al., 1996; Rimstidt and Vaughan, 2003; Kohlhepp et al., 2017). Microbes able to oxidize thiosulfate often possess the genetic potential for autotrophic as well as heterotrophic growth (Ghosh and Dam, 2009; Anantharaman et al., 2018; Wegner et al., 2019), hence, their preferred lifestyle *in situ* is still unknown. By amending the microcosms with thiosulfate, but no organic carbon, we provided ideal conditions for chemolithoautotrophic growth of sulfur oxidizers. We hypothesized that under these conditions, the chemolithoautotrophic activity would be the main source of organic carbon, and a unidirectional carbon flux from autotrophs to heterotrophs would occur in the groundwater microbial community. By mapping the information derived from SIsCA to MAGs, we were able to

characterize carbon utilization and trophic interactions of the active autotrophic and heterotrophic key players in the groundwater microbiome over time. This quantitative resolution of the carbon fluxes through a microbial community and the determination of taxon-specific microbial activities has the power to provide novel insights into the fundamental principles governing microbial life.

Materials and Methods

Groundwater sampling & setup of groundwater microcosms

Groundwater was obtained in June 2018 from the Hainich Critical Zone Exploratory (CZE) well H41 (51.1150842N, 10.4479713E), accessing an aquifer assemblage in 48 m depth in a trochite limestone stratum. Groundwater from this well is oxic with average dissolved oxygen concentrations of 5.0 ± 1.5 mg L^{-1} (mean ± SD), pH 7.2, < 0.1 mg L^{-1} ammonium, 1.9 ± 1.5 mg L^{-1} dissolved organic carbon, and 70.8 ± 12.7 mg L^{-1} total inorganic carbon (Kohlhepp et al., 2017; Schwab et al., 2017). The recharge area of this aquifer assemblage is located in a beech forest (*Fagus sylvatica*). In total, 120 L of groundwater was sampled using a submersible pump (Grundfos MP1, Grundfos, Bjerringbro, Denmark). To collect biomass from the groundwater, 5 L each were filtered through twenty 0.2-µm Supor filters (Pall Corporation, Port Washington, NY, USA). To replace the natural background of inorganic carbon in the groundwater by defined concentrations of ^{12}C or ^{13}C, two times 3 L of filtered groundwater were acidified to pH 4 in 5-L-bottles to remove the bicarbonate present. Following that, ^{12}C- or ^{13}C-bicarbonate was dissolved in the groundwater to a final concentration of 400 mg L^{-1}, corresponding to 79 mg C L^{-1}. The pH of the groundwater samples was adjusted to 7.2 by addition of ^{12}C- or ^{13}C-CO_2 to the headspace of the bottles.

For the ^{13}C-SIP experiment, eighteen microcosms were set up by transferring one filter each into a 500-ml-bottle with 300 ml of treated groundwater as described, 9 control microcosms with water containing ^{12}C-bicarbonate, 9 microcosms with water containing ^{13}C-bicarbonate. Additionally, two microcosms were set up by transferring one filter each into a 1-L-bottle with 350 ml untreated groundwater. One of these bottles was supplemented with 150 ml D_2O, the second bottle was supplemented with 150 ml H_2O. To all microcosms, sodium thiosulfate was added to a final concentration of 2.5 mM, and ammonium chloride was added to a final concentration of 15 µM. All microcosms were incubated in the dark at 15 °C with shaking at 100 rpm.

Hydrochemical analyses

During the incubation, in the 18 microcosms supplemented with ^{12}C- or ^{13}C-bicarbonate, concentrations of oxygen, thiosulfate and sulfate were monitored at regular intervals. Oxygen concentrations were determined using the contactless fiber-optic oxygen sensor system Fibox 4 trace with SP-PSt3-SA23-D5-YOP-US dots (PreSens Precision Sensing GmbH, Regensburg, Germany). Measurements were taken in three ^{12}C microcosms and three ^{13}C microcosms every two days for the first three weeks, and every week thereafter. Thiosulfate concentration was determined in a colorimetric titration assay using iodine. Samples from all microcosms were measured every 4 to 7 days. For each measurement, 1 ml of sample was mixed with 2 mg potassium iodide, then 10 µl of zinc iodide-starch-solution (4 g L^{-1} starch, 20 g L^{-1} zinc chloride and 2 g L^{-1} zinc iodide) and 10 µl of 17% (v:v) phosphoric acid were added. Titration was performed using 0.005 N iodine solution in steps of 5 µl until formation of a faint blue color. The concentration of thiosulfate ($c_{thiosulfate}$ in mg L^{-1}) was calculated according to equation (I), where V_{iodine} is the volume of iodine solution added and V_{sample} is the sample volume:

$$c_{thiosulfate} = \frac{V_{iodine} \times 561}{V_{sample}} \quad \text{(I)}$$

The concentration of sulfate was determined using a turbidimetric assay (Tabatabai, 1974) in samples from all microcosms every 4 to 7 days. For each measurement, 1 ml of sample, standard (50 µM to 1000 µM potassium sulfate) or blank (dH_2O) was mixed with 0.4 ml 0.5 M HCl and 0.2 ml $BaCl_2$-gelatin reagent (0.5 g gelatin and 8 g $BaCl_2$ in 200 ml dH_2O). After incubation for 1 h in the dark, the absorbance at 420 nm was measured in a DR3900 spectrophotometer (HACH, Düsseldorf, Germany).

Detection of cellular activity by Raman microspectroscopy

The two microcosms supplemented with D_2O or H_2O were sampled regularly during the first 7 weeks of incubation for single cell Raman analysis, to quantify the incorporation of deuterium in the biomolecules of active cells as carbon-deuterium-(C-D)-bonds. For Raman spectroscopic analysis, 1 ml sample was filtered through a 0.5-µm-filter and subsequently washed three times by centrifuging

at 10,000g for 2 min (Hettich Rotina 380R), discarding the supernatant and re-suspending the pellet in 1 ml of sterile ddH_2O. Afterwards, the pellet was re-suspended in 50 µl ddH_2O, and 10 µl of the final bacterial suspension were placed on a nickel foil (Raman substrate) and air-dried at room temperature. Microbial cells were located by dark field microscopy. The measurements were performed with a Raman microscope (BioParticleExplorer 0.5, rap.ID Particle Systems GmbH) with an excitation wavelength of 532 nm from a solid-state frequency-doubled Nd:YAG module (Cobolt Samba 25 mW) and a laser power of 13 mW at the sample. The laser light was focused with an x100 objective (Olympus MPLFLN 100xBD) and with a spot size <1 µm laterally. The 180°-backscattered light was diffracted by a single-stage monochromator (Horiba Jobin Yvon HE 532) with a 920 line mm^{-1} grating and then registered with a thermoelectrically cooled CCD camera (Andor DV401-BV) resulting in a spectral resolution of about 8 cm^{-1}. An integration time of 5 s was used per Raman spectrum (-57 to 3203 cm^{-1}).

Pre-processing and analysis of Raman data

The data pre-processing and statistical analysis were carried out in the software GNU R (R Core Team, 2019). First, cosmic spikes were detected and removed from the spectra (Ryabchykov et al., 2016). Subsequently, a wavenumber calibration was applied using 4-acetamidophenol standard spectra (Dörfer et al., 2011), while an intensity calibration was performed using the SRM2242 standard material (Bocklitz et al., 2015; Guo et al., 2017). The fluorescence contribution was removed from the spectra using the asymmetric least-squares (ALS) baseline correction method (Liland et al., 2020). Finally, the spectra were vector normalized and subjected to dimensionality reduction using principal component analysis (PCA). Five principal components were used to build a linear discriminant analysis (LDA) classification model to differentiate between deuterium labeled and non-labeled bacterial cells. The deuterium uptake was expressed as the C-D ratio A(C-D) / [A(C-D) + A(C-H)], which was calculated by integrating the areas of the C-H (2800 - 3100 cm^{-1}) and C-D (2040 - 2300 cm^{-1}) stretching vibration bands. The detection and quantification of deuterium incorporation

from D_2O in microbial cells was used to assess metabolic activity and decide on time points for sampling of microcosms.

Sampling, DNA and protein extraction

After 21, 43 and 70 days of incubation, microbial biomass from the microcosms was recovered by filtration of the water phase through 0.2-µm Supor filters (Pall Corporation). The filters used for biomass enrichment before incubation were retrieved from the microcosms as well and combined with the filters used for the water phase. Combined DNA and protein extraction was performed using a phenol/chloroform-based protocol as previously described (Taubert et al., 2018).

Amplicon sequencing

For taxonomic characterization of the bacterial community in the microcosms, amplicon sequencing of the bacterial 16S rRNA gene, region V3 to V5, was done. Polymerase chain reaction was performed using primer pair Bact_341F/Bact_805R (Herlemann et al., 2011) and HotStarTaq Mastermix (Qiagen, Hilden, Germany) as described previously (Kumar et al., 2018). Amplicons were purified using NucleoSpin Gel & PCR Clean-Up Kit (Macherey-Nagel, Düren, Germany). The NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs, Frankfurt, Germany) was used to prepare libraries for amplicon sequencing, following the manufacturer's instructions. Amplicons were purified using AMPure XP Beads (Beckman Coulter, Krefeld, Germany), and amplicon sequencing was then carried out in-house on a MiSeq Illumina platform (Illumina, Eindhoven, The Netherlands) with v3 chemistry.

Raw sequence data was analyzed using mothur (v.1.39) (Schloss et al., 2009), according to the mothur standard operating procedures (Kozich et al., 2013) as previously described (Taubert et al., 2018). OTU binning with a 3% identity cutoff was performed, followed by OUT classification using the SILVA reference database release SSU 132 (Quast et al., 2013). Raw Illumina MiSeq sequencing data have been deposited in the Sequence Read Archive (SRA) of NCBI under accession numbers SRR11805453 to SRR11805472.

Metagenomic analysis

To obtain genomes of the key organisms in the microcosms, metagenomic sequencing was performed on DNA samples from 4 selected ^{12}C microcosms, one replicate after 21 days of incubation and 43 days of incubation, and two replicates after 70 days of incubation. These samples were selected to cover the majority of taxonomic diversity, based on principal component analysis of the 16S rRNA gene amplicon sequencing data. DNA sizing, quantitation, integrity, and purity were determined using the Agilent 2100 Bioanalyzer system (Santa Clara, CA, USA). Library preparation was performed with a NEBNext Ultra II DNA Lib Prep Kit (New England Biolabs, Ipswich, MA, USA) as recommended by the manufacturer, followed by multiplexed sequencing on one flow cell of an Illumina NextSeq 500 system (300 cycles) to generate 150-bp paired-end reads.

Raw sequencing data was quality filtered using BBDuk (Bushnell, 2014) following assembly using metaSPAdes v3.13.0 (Nurk et al., 2017). Using contigs with more than 1000 bp, three different binning algorithms with default parameters were used to obtain genomic bins: MaxBin 2.0 v2.2.7 (Wu et al., 2016), MetaBAT 2 v2.12.1 (Kang et al., 2019), and BinSanity v0.2.7 (Graham et al., 2017). Bin refinement was performed using the MetaWRAP pipeline v1.1.3 (Uritskiy et al., 2018). Only bins with more than 50% completeness and less than 10% contamination were considered. Bins were classified using GTDB-Tk v0.3.2 (Chaumeil et al., 2019) and completeness parameters were assessed using CheckM v1.0.12 (Parks et al., 2015). Bins from different samples were dereplicated using FastANI v1.0 (Jain et al., 2018). The Prokka pipeline v1.13.3 (Seemann, 2014) was used for calling and functional annotation of gene sequences, and for translation into amino acid sequences for metaproteomics analysis. Selected bins of key organisms covered by the metaproteomics analysis were manually refined using Anvi'o v6.1 (Eren et al., 2015) to obtain the final metagenome-assembled genomes (MAGs). Normalized coverage values for all MAGs were calculated by dividing raw coverage values by the relative abundance of *rpoB* genes of each metagenome. The *rpoB* gene abundances were determined using ROCker with the precomputed model (Orellana et al., 2017). Raw Illumina NextSeq metagenome sequencing data have been deposited in the Sequence Read

Archive (SRA) of NCBI under accession numbers SRR12964611 to SRR12964614. Curated MAGs have been deposited in the Whole Genome Shotgun (WGS) database of NCBI under accession numbers JADMJD000000000 to JADMKH000000000 as well as under BioSample accessions SAMN16635721 to SAMN16635751.

Metaproteomics analysis

Proteins extracted from the microcosms were first subjected to SDS polyacrylamide gel electrophoresis, followed by in-gel tryptic cleavage as previously described (Taubert et al., 2018). After reconstitution in 0.1% formic acid (v:v), LC-MS/MS analysis was performed on a Q Exactive HF instrument (Thermo Fisher Scientific, Waltham, MA, USA) equipped with a TriVersa NanoMate source (Advion Ltd., Ithaca, NY, USA) in LC chip coupling mode. Raw data files were analyzed with Proteome Discoverer (v1.4.1.14, Thermo Fisher Scientific, Waltham, MA, USA) using the Sequest HT search algorithm. Amino acid sequences derived from translation of the genes from all contigs of the metagenomes were used as reference database for protein identification, to allow the assignment of identified proteins to MAGs. Within Proteome Discoverer, the enzyme specificity was set to trypsin, with two missed cleavages allowed. A 5 ppm peptide ion tolerance and a 0.05 Da MS/MS tolerance were used. Oxidation (methionine) and carbamidomethylation (cysteine) were selected as modifications. Peptides were considered identified when they scored a q-value < 1% based on a decoy database and when their peptide rank was 1. Functional classification of peptides was done according to the gene annotation by Prokka, and taxonomic classification based on the dereplicated and refined MAGs described above.

Determination of ^{13}C incorporation in peptides

To quantify the incorporation of ^{13}C, peptide identifications from samples of unlabeled microcosms were mapped to mass spectra of the corresponding ^{13}C-labeled samples, by comparison of expected peptide mass, chromatographic retention time and MS/MS fragmentation pattern. The molecular masses of peptides were calculated based on their amino acid sequence. Isotopologue patterns of

253 labeled peptides were extracted manually from mass spectral data using the Xcalibur Qual Browser
254 (v3.0.63, Thermo Fisher Scientific, Waltham, MA, USA), and the ^{13}C incorporation was calculated as
255 previously described (Taubert et al., 2011).

256 To determine and compare carbon utilization of the organisms represented by the MAGs, Stable
257 Isotope Cluster Analysis (SIsCA) was performed using R (R Core Team, 2019). Measured isotopologue
258 patterns were compared with 21 predicted isotopologue patterns with varying ^{13}C relative isotope
259 abundance (RIA) in 5% intervals, from 0% to 100% ^{13}C RIA. To determine the most probable RIA, the
260 coefficient of determination (R^2) for each comparison was calculated. We expected that microbes
261 fixing CO_2 would exhibit a high (> 90%) ^{13}C RIA, while microbes utilizing organic carbon should feature
262 a lower ^{13}C RIA with an increasing trend over time. To differentiate lifestyles of different microbes, R^2
263 values were averaged between samples obtained from replicate microcosms and peptides assigned
264 to the same MAG. The resulting datasets of 21 R^2 values per time point for each MAG were
265 compared using principal component analysis in the software package vegan (Dixon, 2003).

266 Generation times of individual taxa were calculated by comparing the relative intensity of unlabeled
267 and labeled peptide signals in mass spectrometric data as previously described (Taubert et al., 2012).
268 First, the number of doubling was calculated according to equation (II), where n is the number of
269 doublings, I_{12_C} is the signal intensity of the unlabeled peptide and I_{13_C} is the signal intensity of the
270 labeled peptide.

271 $$n = \log_2 \frac{I_{12_C} + I_{13_C}}{I_{13_C}} \quad \text{(II)}$$

272 If the mass spectrometric signals of labeled and unlabeled peptides were overlapping, the
273 monoisotopic peak was used to determine the total abundance of unlabeled peptide based on the
274 natural distribution of heavy isotopes, as previously described (Taubert et al., 2011). The generation
275 time t_d was subsequently calculated following equation (III), with Δt being the incubation time.

276 $$t_d = \frac{\Delta t}{n} \quad \text{(III)}$$

Assessment of DNA density shifts

To evaluate the ^{13}C incorporation on nucleic acid level, DNA samples extracted from the triplicate groundwater microcosms supplemented with ^{13}C and ^{12}C bicarbonate after 21 and 70 days of incubation were subjected to SIP ultracentrifugation in CsCl gradients as previously described (Neufeld et al., 2007). Ultracentrifugation was carried out in an NVT 90 rotor (Beckman Coulter, Brea, CA, USA) in a Sorvall Discovery 90SE ultracentrifuge (Thermo Fisher Scientific, Waltham, MA, USA) at 40,900 rpm and 20 °C for 60 h. The gradients were separated into 12 to 14 fractions, covering a buoyant density range from 1.77 g ml^{-1} to 1.68 g ml^{-1}. The buoyant density of each fraction was determined using a Reichert AR200 digital refractometer (Reichert Analytical Instruments, Depew, NY, USA). To account for tube-to-tube and spin-to-spin variations of the density gradient (Sieradzki et al., 2020), gradients were normalized by centering the density of ^{12}C DNA onto 1.70 g ml^{-1}. The DNA in the density fractions was purified by precipitation with NaCl-PEG as previously described (Taubert et al., 2019) and quantified fluorometrically using the Qubit dsDNA broad-range assays (Thermo Fisher Scientific). Subsequently, amplicon sequencing of the bacterial 16S rRNA gene was performed as described above. OTU-wise DNA buoyant density profiles over the gradients were obtained as previously described (Taubert et al., 2017). Only OTUs that were represented by at least 10 reads in one fraction of each ^{12}C and each ^{13}C replicate were included in the analysis. This was done separately for both time points. From the DNA buoyant density profiles of each OTU, the DNA density of that OTU ($\overline{\rho}_{OTU}$) was determined by performing least-squares regression of the OTU abundance A_{OTU} per fraction f to the fraction density ρ_f using a nonlinear model (IV), where α^2 represents the variance of the OTU DNA density.

$$A_{OTU,f} = e^{-\frac{\left(\rho_f - \overline{\rho}_{OTU}\right)^2}{2\alpha^2}} \qquad \text{(IV)}$$

The average ^{12}C and ^{13}C OTU DNA density was determined as the mean of the respective triplicates. The density shift between ^{12}C and ^{13}C samples was determined separately for each time point. The significance of the shift was assessed using Student's *t*-test based on the respective triplicates.

Results

Sulfur oxidation by active groundwater microbes

The addition of thiosulfate as an electron donor to the groundwater microcosms was expected to result in an activation of chemolithoautotrophic microbes. To determine the onset of metabolic activity, single-cell Raman microspectroscopic analysis was performed in real time on the microcosm labeled with D_2O. The majority of single cell Raman spectra acquired showed a distinct C-D-band at a wavelength position between 2,100 and 2,200 cm^{-1} within 12 days of incubation (Figure 1). Hence, most cells in the microcosms had become metabolically active, synthesizing new biomolecules and incorporating deuterium from D_2O into new carbon-deuterium bonds. The relative C-D-band intensity increased further over time, from 18.3% after 12 days to 25.7% after 47 days of incubation (median values), indicating continued microbial growth and cross-feeding of deuterium-labeled organic carbon.

Thiosulfate consumption rates in the microcosms with ^{12}C- and ^{13}C-bicarbonate started at 1.7 ± 1.9 µmol d^{-1} during the first three weeks, and increased significantly to 7.2 ± 2.0 µmol d^{-1} in the final four weeks (p = 0.0006, Student's t-test) (Figure S1). Oxygen consumption increased significantly from 5.5 ± 2.0 µmol d^{-1} to 12.8 ± 3.2 µmol d^{-1} in this period (p = 0.0013). Sulfate was produced with rates of 8.1 to 9.6 µmol d^{-1} during the duration of the experiment. The approximated stoichiometry observed over the entire time course of incubation was 2.8:1:2.6 (oxygen:thiosulfate:sulfate), close to the theoretical values of 2:1:2 for oxygen-dependent thiosulfate oxidation. The hydrochemical analyses thus demonstrated sulfur oxidation in the microcosms, providing evidence for chemolithoautotrophic primary production. No differences in consumption or formation rates were observed between the ^{12}C- and ^{13}C-bicarbonate supplemented microcosms, indicating that the isotope treatment had no effect on microbial activity.

Organism-specific ^{13}C incorporation reveals distinct lifestyles

To identify the microbial key players responsible for chemolithoautotrophic primary production, we performed metagenomic sequencing of DNA from selected microcosms after 21, 43 and 70 days to obtain MAGs of the groundwater organisms. We then conducted genome-resolved SIP-metaproteomics based on these MAGs using all ^{12}C- and ^{13}C-bicarbonate-supplemented microcosms.

Our newly developed SIsCA (Stable Isotope Cluster Analysis) approach revealed five clusters of MAGs with distinct carbon utilization patterns among the 31 most abundant MAGs (Figure 2, Figure S2). The organisms represented by the MAGs in cluster I exhibited a ^{13}C RIA of 95% during the entire 70 days of the experiment (Figure 2). This stable and high ^{13}C incorporation can have only resulted from exclusive assimilation of CO_2, and hence these organisms, related to *Thiobacillus* (*Betaproteobacteriales*), displayed a strictly autotrophic lifestyle. The generation times of these strict autotrophs were shorter than our two-day detection limit within the first 21 days of incubation (Figure 3). Hence, the organisms of cluster I were rapidly producing new, ^{13}C-labeled biomass from $^{13}CO_2$.

Organisms in cluster II displayed a ^{13}C RIA of 65% on average after 21 days of incubation (Figure 2). After 43 and 70 days, their ^{13}C RIA had increased substantially to 91%. The intermediate ^{13}C incorporation at the first time point suggested that these organisms were using organic carbon produced by the autotrophs from cluster I, in addition to organic carbon from the groundwater. Potentially due to limitation of this carbon, they switched to chemolithoautotrophic growth later on. The members of cluster II displaying such a mixotrophic lifestyle were related to *Methyloversatilis*, *Polaromonas* and *Dechloromonas* (all *Betaproteobacteriales*), and showed generation times from 2 to 4 days (Figure 3).

In cluster III, average ^{13}C RIAs increased from 65 to 76% and in cluster IV from 18 to 53% from 21 to 70 days of incubation (Figure 2). The intermediate levels of ^{13}C incorporation indicated the assimilation of organic carbon from chemolithoautotrophic primary production of clusters I and II.

The observed increase in ^{13}C RIA reflects the increasing labeling of the organic carbon in the microcosms due to $^{13}CO_2$ fixation. Variations in the ^{13}C RIAs between species suggested different extents of cross-feeding on the chemolithoautotrophically produced organic carbon, potentially due to preferences of different organic carbon compounds. Most of the organisms in these clusters, related to various *Betaproteobacteriales* and other bacterial groups, exhibited generation times of 3 to 4 days (Figure 3). Members of cluster III affiliated with *Hydrogenophaga*, *Vitreoscilla* and *Rubrivivax*, however, showed faster growth, comparable with cluster I. In cluster V, average ^{13}C RIAs reached 6% after 21 days of incubation and did not change at the later time points, suggesting that the organisms were active only in this first period, where they displayed a heterotrophic lifestyle. Generation times in this period were slightly longer and displayed more variability than for clusters III and IV, ranging from 3.5 days for one *Acidivorax* species to eight days for *Aquabacterium* (Figure 3).

Approximately 43% of carbon in the microbial community was replaced by ^{13}C after 21 days of incubation, based on the corresponding peptide RIAs of all analyzed MAGs. This portion increased further to 68% after 43 days and 80% after 70 days of incubation. This increase of ^{13}C incorporation across the entire microbiome demonstrated the extensive cross-feeding and recycling of organic carbon. SIsCA revealed the carbon flux from autotrophic cluster I to mixotrophic cluster II. On both of these clusters, further cross-feeding by the heterotrophic cluster III to V occurred. Recycling of organic carbon likely played a role within cluster II as well as among the heterotrophic clusters. In agreement, the number of ^{13}C-labeled taxa increased from 21 OTUs after 21 days of incubation to 65 OTUs after 70 days of incubation, observable by a significant shift of DNA buoyant density (Figure S3). While after 21 days, these OTUs were mainly affiliated with *Betaproteobacteriales* such as *Thiobacillus*, *Hydrogenophaga* and *Polaromonas*, after 70 days, various other *Alpha*- and *Gammaproteobacteria* were included. The average buoyant density shift of these OTUs likewise increased significantly from 0.021 ± 0.010 g ml^{-1} to 0.028 ± 0.013 g ml^{-1} (p = 0.017, Student's t-test) in this period. This highlights the increasing role of $^{13}CO_2$-derived carbon introduced by

chemolithoautotrophic activity into the microbial carbon pool in the groundwater incubations, and the flux of ^{13}C through the microbial food web.

Functional characterization of MAGs reveals putative mixotrophs

All putative autotrophs detected employed the Calvin-Benson-Bassham (CBB) cycle for CO_2 fixation (Figure 4). Subunits of the key enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) were detected for 15 of the 31 MAGs. In 14 of these, further enzymes of the CBB cycle were present. Interestingly, proteins of the CBB cycle were not only present in strict or facultative autotrophs related to *Thiobacillus* from cluster I or *Methyloversatilis*, *Polaromonas* and *Dechloromonas* from cluster II, but also in heterotrophic organisms affiliated with genera *Hydrogenophaga*, *Rhodoferax, Paucibacter* and *Rubrivivax*, from cluster III and IV. Hence, the ability for mixotrophic growth was present in a large number of organisms in the groundwater microcosms also beyond cluster II. No enzymes allowing a complete reconstruction of other CO_2 fixation pathways, like the Wood-Ljungdahl pathway (no carbon monoxide dehydrogenase/acetyl-CoA synthase (CODH/ACS) or no accessory enzymes), the reductive tricarboxylic acid cycle, or the 3-hydroxypropanoate cycle, were detected.

Functional composition of the whole microbial community

Mapping the functional information obtained to the corresponding taxa in 16S rRNA gene profiles allowed us to classify the lifestyle of up to 50% of the microbial community. Only 3.2 ± 3.1% (mean ± sd) of the community were composed of strict autotrophs, primarily affiliated with *Thiobacillus*. Mixotrophs comprised 17.6 ± 4.3% of the total community and were dominated by *Rhodoferax* and *Hydrogenophaga*, but the largest fraction of the total community, with 20.1 ± 7.8%, consisted of heterotrophs, primarily affiliated to *Sediminibacterium* (*Bacteroidetes*), *Pseudomonas* (*Gammaproteobacteria)*, *Sericytochromatia* (*Cyanobacteria*) and *Microbacterium* (*Actinobacteria*) (Figure S4, Figure S5).

Based on the metagenomics dataset, cluster I containing the strict autotrophs had the lowest normalized coverage, making up 11% of all five clusters. The normalized coverage of the mixotrophic cluster II and the heterotrophic/mixotrophic cluster III were more than twice as high, with 26% and 28% of all clusters, respectively. The clusters IV and V also had higher normalized coverages than cluster I, representing 20% and 15% of all clusters, respectively. This demonstrates that the strict autotrophs represented a minor part of the microbial community, suggesting that mixotrophs were of higher importance for the fixation of CO_2 in the microcosms.

MAGs express pathways for the utilization of reduced sulfur compounds

Sixteen of the MAGs obtained were expressing proteins for sulfur oxidation, belonging to the Sox or the Dsr enzyme system (Figure 4). Organisms from clusters II to IV, affiliated with the genera *Methyloversatilis*, *Dechloromonas*, *Hydrogenophaga*, *Rhodoferax* and other *Betaproteobacteriales*, were exclusively using the Sox system, which allows the oxidation of both the sulfane group and the sulfone group of thiosulfate to sulfate, without any free intermediates (Ghosh and Dam, 2009). The corresponding MAGs exhibited gene clusters sharing a conserved *soxCDYZAXB* gene order (Figure S6), featuring the core components of the Kelly-Friedrich pathway (Kelly et al., 1997; Friedrich et al., 2001). Further genes, including *soxVW*, *soxEF*, *soxTRS*, and *soxH*, were typically found in different regions of the MAGs, showing variable arrangements. Hence, the clusters differ from the canonical operon structure *soxVWXYZABCDEFGH* described for the *Alphaproteobacterium Paracoccus pantotrophus* (Friedrich et al., 2001; Ghosh and Dam, 2009). Organisms from cluster I, affiliated with *Thiobacillus*, were found to produce enzymes from both the Sox and the Dsr system. The corresponding MAGs contained a truncated *soxXYZAB* gene cluster, lacking the genes *soxCD* required for oxidation of the sulfane group of thiosulfate, and hence likely utilized the branched thiosulfate oxidation pathway. In this pathway, the reverse activity of the Dsr enzyme system is used to oxidize the sulfane-derived sulfur atom to sulfite, via elemental sulfur as an intermediate (Ghosh and Dam, 2009). The MAGs in cluster I contained the conserved operon structure *dsrABEFHCMKLJOPNR*, thus including the genes *dsrEFH* and *dsrL* described to be typical for sulfur-oxidizers, but lacking the gene

dsrD indicative of sulfate-reducers (Grimm et al., 2008). In addition, these organisms also expressed the *aprAB* and *sat* genes encoding Adenosine-5'-phosphosulfate reductase and ATP sulfurylase, which can likewise act in reverse to oxidize sulfite to sulfate (Beller et al., 2006). Hence, facultative chemolithoautotrophs in the groundwater microcosms used the Sox system to oxidize thiosulfate to sulfate, whereas obligate chemolithoautotrophs seemed to have utilized an incomplete version of the Sox system to oxidize the sulfone group and the Dsr/Apr/Sat system to oxidize the sulfane group of thiosulfate.

Sulfur oxidizers can use different electron acceptors and donors

For 15 of the MAGs of sulfur oxidizers, cytochrome c oxidase and other enzymes of the respiratory chain were detected, and for 12 of them, enzymes for nitrate reduction (nitrate reductase, nitrite reductase, and nitric oxide reductase) were found (Figure 4). Multiple organisms such as *Dechloromonas* and *Rhodoferax* had both pathways expressed simultaneously. Proteins for ammonia oxidation, ammonia monooxygenase and hydroxylamine oxidoreductase, were produced by various organisms in clusters I to IV, such as *Thiobacillus* and *Methyloversatilis*. Furthermore, MAG_77 (*Thiobacillus*), MAG_55 (*Dechloromonas*) and MAG_7 (*Hydrogenophaga*) expressed [NiFe]-hydrogenase genes. This indicates that the sulfur oxidizers were able to perform both aerobic respiration and denitrification, and that also nitrogen compounds and hydrogen could be used as electron donor in the microcosms.

MAGs salvage a variety of organic carbon sources

For organic carbon utilization, various pathways, targeting simple sugars (Glycolysis, pentose phosphate pathway), amino acids (TCA cycle), fatty acids (beta-oxidation), C_1 compounds and aromatic compounds, were found throughout clusters II to V (Figure 4). The obligate autotrophs in cluster I only expressed pathways for degradation of simple sugars. The TCA cycle was present as one of the most abundant metabolic modules in the organisms of clusters II to V. Pathways for toluene and ethylbenzene utilization were expressed by organisms affiliated with *Dechloromonas* and

Rhizobacter (*Betaproteobacteriales*), respectively. Enzymes for naphthalene and catechol utilization were detected, e.g., for organisms related to *Hydrogenophaga* and *Pseudomonas*. The utilization of C_1 compounds was primarily detected for organisms related to *Methyloversatilis*, which possessed methanol dehydrogenase, formate dehydrogenase and various enzymes for tetrahydromethanopterin-dependent C_1-cycling. Enzymes for degradation of complex carbohydrates such as starch and chitin were produced by organisms related to *Microbacterium* and *Sediminibacterium*. Reflecting this high versatility for organic carbon utilization, a wide range of proteins with import functions was observed (Figure 4). In mixotrophs and heterotrophs from clusters II to V, peptides of import systems for amino acids and carboxylic acids (e.g. alpha-keto acids, C_4-dicarboxylates, lactate) were highly abundant. The microorganisms from clusters III to V, which were exclusively growing heterotrophically in our incubations, possessed further transport systems for carbohydrates and nucleotides, and showed the broadest variety of import functions. For the obligate autotrophs in cluster I, only transporters targeting cations (mostly iron transporters) and phosphate were found. This suggests an increasing ability to utilize organic carbon from cluster I to clusters III - V.

Discussion

An understanding of the function of microbial communities in the environment cannot be achieved from genomic data alone, as information about their activity is required. The combination of ^{13}C- and D_2O-SIP with genome-resolved metaproteomics and Raman microspectroscopy allowed us to determine the activity of organisms on the level of specific taxa as well as on the level of single cells. SIsCA revealed the carbon flux between the active taxa qualitatively and quantitatively. By comparing the ^{13}C incorporation patterns after different periods of incubation, we were able to follow the metabolic processes of the organisms over time (Figure 5). We expected that sulfur-oxidizing chemolithoautotrophs would be activated first. Later on, organic carbon provided through chemolithoautotrophic primary production by this dominant population would be released, for

example through dying cells releasing necromass into the water and fuel a heterotrophic satellite community. The genomic makeup of the organisms identified supported this classical separation of autotrophic and heterotrophic lifestyles on first glance. Raman analysis, however, showed that most microbial cells were active in a short period, questioning our concept of a strict temporal and functional separation. We observed that most organisms were combining an autotrophic and a heterotrophic lifestyle to establish individual, optimized strategies for survival, adapted to the conditions they experienced at the time. This gave rise to a high functional diversity in the groundwater microcosms. We could show exactly how the diverse organisms leveraged their genomic potential to come up with strategies to succeed in the groundwater microcosms, moving beyond a separation of different lifestyles based on the genomic potential of organisms.

Mixotrophs, and not obligate chemolithoautotrophs, were the most abundant active organisms in our groundwater microcosm. These mixotrophs strongly preferred heterotrophic growth. Their ^{13}C incorporation patterns, raising from 65 up to 91% RIA, clearly demonstrated the uptake of organic carbon. This is astonishing as thiosulfate and oxygen were available in sufficient amounts throughout the whole experiment, providing conditions that should have favored chemolithoautotrophic growth. A heterotrophic lifestyle was likely favored due to the higher metabolic cost of fixing CO_2 via the CBB cycle during autotrophic growth (Berg, 2011). Consequently, CO_2 fixation only occurred when insufficient organic carbon was available. The organisms of cluster II switched from heterotrophic growth to CO_2 fixation at later time points, likely due to such limitations of suitable organic carbon compounds. The only strict autotrophs were the Thiobacillus-related organisms from cluster I. *Thiobacillus* species are known to be obligate autotrophs, featuring an incomplete TCA cycle that prevents heterotrophic growth (Boden et al., 2017). Performing thiosulfate and hydrogen-driven denitrification, *Thiobacillus* constituted up to 50% of a previous enrichment culture obtained from the groundwater investigated here (Kumar et al., 2018), but only made up a minor fraction of the total community in our experiment. Nevertheless, cluster I represented the first entry point of CO_2-derived ^{13}C. Their low abundance despite the short generation times of less than 2 days suggested a

high turnover of biomass, allowing rapid cross-feeding by other organisms. Within 21 days, 43% of the microbial biomass was ^{13}C-labeled. The further increase of ^{13}C-labeling in the microbial biomass, up to 80% after 70 days, was driven through a cascade of recycling reactions by organisms with a preference for heterotrophic growth. Similar continuous recycling of microbial biomass is also characterizing the transport of organic matter from soils (Stevens, 1997; Roth et al., 2019). When no suitable organic carbon was available, these organisms turned to autotrophic growth instead, and further increased the ^{13}C-labeling in the community. This opportunistic strategy to use autotrophy to fill up their carbon demand allowed them to dominate the microbial community, with generation times almost as short as the strict autotrophs, with 4 days to less than 2 days. Considering the overall low concentrations but high diversity of organic carbon compounds in the groundwater (Benk et al., 2019; Schwab et al., 2019), such a versatile and adaptive strategy seemed to be superior to focusing on one particular lifestyle. In contrast to organisms restricted to a purely heterotrophic lifestyle in cluster IV and V with generation times of up to 8 days, the ability to use CO_2 fixation as a fill-up function for organic carbon proved highly advantageous in the oligotrophic groundwater.

Sulfur oxidation was not restricted to strict chemolithoautotrophs in our groundwater microcosms. Mixotrophs and heterotrophs expressed sulfur oxidation pathways as well. The use of such inorganic electron donors as energy sources is an alternative to the oxidation of organic compounds. Especially when the concentrations of organic carbon present are low, as in pristine groundwater, microorganisms might rather use it as carbon source than to gain energy, as the re-fixation of CO_2 is associated with higher energy costs (Berg, 2011). Under such conditions, the ability to use inorganic electron donors would be an advantageous part of the survival strategy. For various facultative autotrophic taxa from clusters II to IV, including *Polaromonas*, *Dechloromonas*, *Hydrogenophaga*, and *Rhodoferax*, the ability for sulfur oxidation has been suggested based on genomic evidence (Kämpfer et al., 2005; Mattes et al., 2008; Salinero et al., 2009; Jin et al., 2020). However, chemolithoautotrophic growth on reduced sulfur compounds has not been observed for these genera in pure culture so far. Members of these taxa frequently occur in ecosystems where the

oxidation of thiosulfate plays an important role, e.g., linked to denitrification (Kumar et al., 2017; Kumar et al., 2018; Sun et al., 2020). Our study demonstrates that these organisms can use reduced sulfur compounds as energy source, and *Polaromonas*, *Dechloromonas* and potentially *Hydrogenophaga* used it to fuel autotrophic growth. Interestingly, for most of the observed taxa the growth on H_2 and CO_2 was experimentally demonstrated (Willems et al., 1989; Shrout et al., 2005; Sizova and Panikov, 2007), and other H_2 oxidizing organisms are known to grow chemolithoautotrophically also on reduced sulfur compounds (Sano et al., 2010). This suggests a close link between the hydrogen and the sulfur cycle, potentially via the production of H_2 from reduced sulfur compounds, in chemolithoautotrophic sulfur oxidizers.

The ability to use organic carbon compounds showed an inverse, gradual relationship to the CO_2 fixation activity of the organisms in our groundwater microcosms. Organisms with a strictly autotrophic strategy had no transport functions for organic carbon. Their metabolic capabilities were restricted to simple sugars, likely to allow the utilization of organic carbon assimilated via the CBB cycle (Berg, 2011). Organisms in cluster II with a strategy that combined autotrophic and heterotrophic lifestyle possessed importers for amino acids, and highly expressed the TCA cycle to facilitate their assimilation. This suggests that amino acids were central compounds in the carbon transfer between microbes. As amino acids are major components of microbial necromass, it is plausible that lysis of autotrophic cells made these compounds available in the microcosms. This release of amino acids by the strict autotrophs of cluster I was limited, so cluster II adapted its strategy by switching to fixation of CO_2 later on. The success of this strategic mix of chemolithoautotrophy with utilization of easily available organic carbon compounds allowed them to become 2.5-fold as abundant as cluster I. The organisms of cluster III were able to utilize a wider range of organic compounds, including carbohydrates, methyl compounds, nucleotides and aromatic compounds. The strategy of diversifying their organic carbon diet allowed them to become the most abundant cluster in the microcosms, being more independent of autotrophic CO_2 fixation until the end of the experiment. Whether they grew entirely heterotrophic or obtained a minor part of carbon

553 from CO_2 via the expressed CBB cycle is difficult to determine. Also during purely heterotrophic
554 growth, a small part of CO_2-derived carbon would be incorporated into biomass due to anaplerotic
555 reactions (Alonso-Sáez et al., 2010; Spona-Friedl et al., 2020). The strict heterotrophs in cluster IV
556 expressed most of the organic carbon transport and assimilation pathways, allowing them to succeed
557 without autotrophic CO_2 fixation capabilities by relying entirely on the organic carbon in the
558 groundwater microcosms. The organisms in the groundwater microcosms hence relied on various
559 strategies combining chemolithoautotrophy and heterotrophy to different extends, by utilizing no
560 organic carbon, only the most easily available compounds, a diverse range, or as many compounds as
561 possible. The more diverse the metabolic potential of an organism is, the higher the tendency to use
562 organic carbon and the higher the independence from CO_2 fixation. In its entirety, the community
563 was highly efficient in using the diverse carbon compounds available in the groundwater microcosms,
564 despite facing a diverse mixture of organic carbon compounds.

565 Metagenomics studies of multiple groundwater sites suggested a diverse metabolic potential, e.g. for
566 using multiple electron donors and acceptors, to be a common trait of chemolithoautotrophic sulfur
567 oxidizers (Anantharaman et al., 2016; Jewell et al., 2016; Wegner et al., 2019). Based on the
568 strategies observed in our microcosms, sulfur oxidizers should have two main advantages in shallow
569 aquifers of mixed siliciclastic-limestone alternations. (1) Reduced sulfur compounds are released
570 autochthonously from weathering of interspersed pyrite minerals (Schippers et al., 1996; Rimstidt
571 and Vaughan, 2003; Kohlhepp et al., 2017) so the energy source for sulfur oxidizers is independent
572 from surface input dynamics. (2) Sulfur oxidizers can tailor their organic carbon requirements
573 dependent on the availability. Within the subsurface, carbon can be sourced from surface based
574 primary production and translocated with groundwater recharge. However, this surface carbon has
575 typically already been recycled through the microbial loop and incorporated into labile microbial
576 carbon (Roth et al., 2019). Second, microbial cells are transported down from the soils, representing
577 another source of microbial carbon (Herrmann et al., 2019). Third, subsurface organic carbon can
578 also be released from sedimentary rocks, which is often characterized by a higher fraction of

aromatic and aliphatic functions, similar to organic carbon from marine sediments rather than soil (Benk et al., 2019). The ability of sulfur oxidizers to access all of these carbon sources, and additionally fix CO_2, should allow them to thrive independent of variations in the organic carbon supply.

The majority of the active sulfur oxidizers observed in our groundwater microcosms were affiliated with the class *Gammaproteobacteria*. In the groundwater of well H41, *Gammaproteobacteria* are the second most abundant class after *Cand.* Parcubacteria (Yan et al., 2020). Organisms affiliated with *Betaproteobacteriales* were found to have the highest abundance of RuBisCO-encoding transcripts (Herrmann et al., 2015), highlighting their importance for CO_2 fixation in this well. The rapid replacement of organic carbon by ^{13}C in our microcosms likewise suggested a high importance of CO_2 fixation for carbon cycling in the groundwater. Indeed, genes for pathways of CO_2 fixation, such as the CBB cycle, are widely distributed: In 85% of 48 groundwater samples obtained from various locations in Germany and Austria, RuBisCO genes were detected (Alfreider et al., 2009), and at the site investigated here, RuBisCO transcripts made up 4% of the total metatranscriptomic reads (Wegner et al., 2019). The recent finding that the metabolic potential for chemolithoautotrophy coincides with the complexity of subsurface food webs (Herrmann et al., 2020) demonstrates that this essential function has consequences for the entire ecosystem.

Conclusions

Stable Isotope Cluster Analysis utilizing high-resolution ^{13}C isotopologue patterns of peptides obtained by genome-resolved SIP-metaproteomics allowed us to provide MAG-specific information of CO_2 fixation, organic carbon utilization and activity of the microorganisms in our groundwater model system. We observed that mainly mixotrophs were abundant in the groundwater microcosms, and largely preferred the uptake of organic carbon to CO_2 fixation. Chemolithoautotrophy is an essential function for the recycling of organic carbon, and seems to serve as a “fill-up function” in the groundwater, to replenish organic carbon that has been lost or in situations where not enough

organic carbon is available for the organisms. A variety of autotrophic, mixotrophic and heterotrophic organisms were able to utilize reduced sulfur compounds as energy sources. These organisms thus do not rely solely on the oxidation of the precious organic carbon to CO_2 for energy gain, and instead can focus on employing it as carbon source. For this, the groundwater microbes show a high metabolic versatility to utilize as much as possible of the scarce organic carbon compounds present in the oligotrophic groundwater. The presented next generation physiology approach provides information about the metabolic phenotype of individual taxa as well as the carbon fluxes between them. This information is essential for the confirmation of hypotheses established in metagenomics studies and to understand the role of microbial communities in ecosystem functioning.

Acknowledgements

We are grateful to Robert Lehmann, Falko Gutmann, Heiko Minkmar, Jens Wurlitzer and Lena Carstens for help with field and lab work, and sampling of groundwater. We thank Ivonne Görlich and Marco Groth from the Core Facility DNA sequencing of the Leibniz Institute on Aging - Fritz Lipmann Institute in Jena for their help with Illumina sequencing. We further thank Daniel Desirò and Martin Hölzer for assistance with sequencing organization. This study is part of the Collaborative Research Centre AquaDiva (CRC 1076 AquaDiva, 218627073) of the Friedrich Schiller University Jena, funded by the Deutsche Forschungsgemeinschaft, and was financially supported by the German Center for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig. Climate chambers to conduct experiments under controlled temperature conditions were financially supported by the Thüringer Ministerium für Wirtschaft, Wissenschaft und Digitale Gesellschaft (TMWWDG; project B 715-09075).

References

Akob, D.M., and Küsel, K. (2011) Where microorganisms meet rocks in the Earth's Critical Zone. *Biogeosciences* **8**: 3531-3543.

628 Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., and Nielsen, P.H. (2013)
629 Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple
630 metagenomes. *Nat Biotechnol* **31**: 533-+.

631 Alfreider, A., Vogt, C., Geiger-Kaiser, M., and Psenner, R. (2009) Distribution and diversity of
632 autotrophic bacteria in groundwater systems based on the analysis of RubisCO genotypes. *Syst Appl*
633 *Microbiol* **32**: 140-150.

634 Alonso-Sáez, L., Galand, P.E., Casamayor, E.O., Pedrós-Alió, C., and Bertilsson, S. (2010) High
635 bicarbonate assimilation in the dark by Arctic bacteria. *Isme J* **4**: 1581-1590.

636 Anantharaman, K., Hausmann, B., Jungbluth, S.P., Kantor, R.S., Lavy, A., Warren, L.A. et al. (2018)
637 Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. *Isme J* **12**: 1715-
638 1728.

639 Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J. et al. (2016) Thousands
640 of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system.
641 *Nat Commun* **7**.

642 Beller, H.R., Chain, P.S.G., Letain, T.E., Chakicherla, A., Larimer, F.W., Richardson, P.M. et al. (2006)
643 The genome sequence of the obligately chemolithoautotrophic, facultatively anaerobic bacterium
644 *Thiobacillus denitrificans*. *J Bacteriol* **188**: 1473-1488.

645 Benk, S.A., Yan, L.J., Lehmann, R., Roth, V.N., Schwab, V.F., Totsche, K.U. et al. (2019) Fueling diversity
646 in the subsurface: Composition and age of dissolved organic matter in the Critical Zone. *Front Earth*
647 *Sci* **7**.

648 Berg, I.A. (2011) Ecological aspects of the distribution of different autotrophic CO_2 fixation pathways.
649 *Appl Environ Microb* **77**: 1925-1936.

650 Bocklitz, T.W., Dörfer, T., Heinke, R., Schmitt, M., and Popp, J. (2015) Spectrometer calibration
651 protocol for Raman spectra recorded with different excitation wavelengths. *Spectrochim Acta A* **149**:
652 544-549.

653 Boden, R., Hutt, L.P., and Rae, A.W. (2017) Reclassification of *Thiobacillus aquaesulis* (Wood & Kelly,
654 1995) as *Annwoodia aquaesulis* gen. nov., comb. nov., transfer of *Thiobacillus* (Beijerinck, 1904) from
655 the *Hydrogenophilales* to the *Nitrosomonadales*, proposal of *Hydrogenophilalia* class. nov within the
656 '*Proteobacteria*', and four new families within the orders *Nitrosomonadales* and *Rhodocyclales*. *Int J*
657 *Syst Evol Micr* **67**: 1191-1205.

658 Brown, C.T., Hug, L.A., Thomas, B.C., Sharon, I., Castelle, C.J., Singh, A. et al. (2015) Unusual biology
659 across a group comprising more than 15% of domain *Bacteria*. *Nature* **523**: 208-U173.

660 Bushnell, B. (2014) BBTools software package. *URL http://sourceforge net/projects/bbmap*.

661 Castelle, C.J., Wrighton, K.C., Thomas, B.C., Hug, L.A., Brown, C.T., Wilkins, M.J. et al. (2015) Genomic
662 expansion of domain *Archaea* highlights roles for organisms from new phyla in anaerobic carbon
663 cycling. *Curr Biol* **25**: 690-701.

664 Chaumeil, P.A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2019) GTDB-Tk: a toolkit to classify
665 genomes with the Genome Taxonomy Database. *Bioinformatics*.

666 Dixon, P. (2003) VEGAN, a package of R functions for community ecology. *J Veg Sci* **14**: 927-930.

667 Dörfer, T., Bocklitz, T., Tarcea, N., Schmitt, M., and Popp, J. (2011) Checking and improving calibration
668 of Raman spectra using chemometric approaches. *Zeitschrift Für Physikalische Chemie* **225**: 753-764.

669 Emerson, J.B., Thomas, B.C., Alvarez, W., and Banfield, J.F. (2016) Metagenomic analysis of a high
670 carbon dioxide subsurface microbial community populated by chemolithoautotrophs and bacteria
671 and archaea from candidate phyla. *Environ Microbiol* **18**: 1686-1703.

672 Eren, A.M., Esen, O.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., and Delmont, T.O. (2015)
673 Anvi'o: an advanced analysis and visualization platform for 'omics data. *Peerj* **3**: e1319.

674 Friedrich, C.G., Rother, D., Bardischewsky, F., Quentmeier, A., and Fischer, J. (2001) Oxidation of
675 reduced inorganic sulfur compounds by bacteria: Emergence of a common mechanism? *Appl Environ*
676 *Microb* **67**: 2873-2882.

677 Ghosh, W., and Dam, B. (2009) Biochemistry and molecular biology of lithotrophic sulfur oxidation by
678 taxonomically and ecologically diverse Bacteria and Archaea. *Fems Microbiol Rev* **33**: 999-1043.

679 Graham, E.D., Heidelberg, J.F., and Tully, B.J. (2017) BinSanity: unsupervised clustering of
680 environmental microbial assemblies using coverage and affinity propagation. *Peerj* **5**.

681 Griebler, C., and Lueders, T. (2009) Microbial biodiversity in groundwater ecosystems. *Freshwater*
682 *Biol* **54**: 649-677.

683 Grimm, F., Franz, B., and Dahl, C. (2008) Thiosulfate and sulfur oxidation in purple sulfur bacteria. In
684 *Microbial Sulfur Metabolism*: Springer, pp. 101-116.

685 Guo, S.X., Heinke, R., Stöckel, S., Rösch, P., Bocklitz, T., and Popp, J. (2017) Towards an improvement
686 of model transferability for Raman spectroscopy in biological applications. *Vib Spectrosc* **91**: 111-118.

687 Herlemann, D.P.R., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J.J., and Andersson, A.F. (2011)
688 Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *Isme J* **5**:
689 1571-1579.

690 Herrmann, M., Geesink, P., Yan, L., Lehmann, R., Totsche, K.U., and Küsel, K. (2020) Complex food
691 webs coincide with high genetic potential for chemolithoautotrophy in fractured bedrock
692 groundwater. *Water Res* **170**.

693 Herrmann, M., Rusznyak, A., Akob, D.M., Schulze, I., Opitz, S., Totsche, K.U., and Küsel, K. (2015)
694 Large fractions of CO_2-fixing microorganisms in pristine limestone aquifers appear to be involved in
695 the oxidation of reduced sulfur and nitrogen compounds. *Appl Environ Microb* **81**: 2384-2394.

696 Herrmann, M., Wegner, C.E., Taubert, M., Geesink, P., Lehmann, K., Yan, L.J. et al. (2019)
697 Predominance of *Cand.* Patescibacteria in groundwater is caused by their preferential mobilization
698 from soils and flourishing under oligotrophic conditions. *Front Microbiol* **10**.

699 Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J. et al. (2016) A new
700 view of the tree of life. *Nat Microbiol* **1**.

701 Hungate, B.A., Mau, R.L., Schwartz, E., Caporaso, J.G., Dijkstra, P., van Gestel, N. et al. (2015)
702 Quantitative Microbial Ecology through Stable Isotope Probing. *Appl Environ Microb* **81**: 7570-7581.

703 Jain, C., Rodriguez, R.L., Phillippy, A.M., Konstantinidis, K.T., and Aluru, S. (2018) High throughput ANI
704 analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* **9**: 5114.

705 Jewell, T.N.M., Karaoz, U., Brodie, E.L., Williams, K.H., and Beller, H.R. (2016) Metatranscriptomic
706 evidence of pervasive and diverse chemolithoautotrophy relevant to C, S, N and Fe cycling in a
707 shallow alluvial aquifer. *Isme J* **10**: 2106-2117.

708 Jin, C.Z., Zhuo, Y., Wu, X.W., Ko, S.R., Li, T.H., Jin, F.J. et al. (2020) Genomic and metabolic insights
709 into denitrification, sulfur oxidation, and multidrug efflux pump mechanisms in the bacterium
710 *Rhodoferax sediminis* sp. nov. *Microorganisms* **8**.

711 Kämpfer, P., Schulze, R., Jäckel, U., Malik, K.A., Amann, R., and Spring, S. (2005) *Hydrogenophaga*
712 *defluvii* sp. nov. and *Hydrogenophaga atypica* sp. nov., isolated from activated sludge. *Int J Syst Evol*
713 *Micr* **55**: 341-344.

714 Kang, D.W.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019) MetaBAT 2: an
715 adaptive binning algorithm for robust and efficient genome reconstruction from metagenome
716 assemblies. *Peerj* **7**.

717 Kelly, D.P., Shergill, J.K., Lu, W.P., and Wood, A.P. (1997) Oxidative metabolism of inorganic sulfur
718 compounds by bacteria. *Anton Leeuw Int J G* **71**: 95-107.

719 Kohlhepp, B., Lehmann, R., Seeber, P., Küsel, K., Trumbore, S.E., and Totsche, K.U. (2017) Aquifer
720 configuration and geostructural links control the groundwater quality in thin-bedded carbonate-
721 siliciclastic alternations of the Hainich CZE, central Germany. *Hydrol Earth Syst Sc* **21**: 6091-6116.

722 Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013) Development of a
723 dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the
724 MiSeq Illumina sequencing platform. *Appl Environ Microb* **79**: 5112-5120.

725 Kumar, S., Herrmann, M., Blohm, A., Hilke, I., Frosch, T., Trumbore, S.E., and Küsel, K. (2018)
726 Thiosulfate- and hydrogen-driven autotrophic denitrification by a microbial consortium enriched
727 from groundwater of an oligotrophic limestone aquifer. *Fems Microbiol Ecol* **94**.

728 Kumar, S., Herrmann, M., Thamdrup, B., Schwab, V.F., Geesink, P., Trumbore, S.E. et al. (2017)
729 Nitrogen loss from pristine carbonate-rock aquifers of the Hainich Critical Zone Exploratory
730 (Germany) is primarily driven by chemolithoautotrophic anammox processes. *Front Microbiol* **8**.

731 Küsel, K., Totsche, K.U., Trumbore, S.E., Lehmann, R., Steinhäuser, C., and Herrmann, M. (2016) How
732 deep can surface signals be traced in the Critical Zone? Merging biodiversity with biogeochemistry
733 research in a Central German Muschelkalk landscape. *Front Earth Sci* **4**.

734 Liland, K.H., Mevik, B.H., and Canteri, R. (2020). Baseline correction of spectra. R package v. 1.3-0.

735 Long, P.E., Williams, K.H., Hubbard, S.S., and Banfield, J.F. (2016) Microbial metagenomics reveals
736 climate-relevant subsurface biogeochemical processes. *Trends Microbiol* **24**: 600-610.

737 Mattes, T.E., Alexander, A.K., Richardson, P.M., Munk, A.C., Han, C.S., Stothard, P., and Coleman, N.V.
738 (2008) The genome of *Polaromonas* sp. strain JS666: Insights into the evolution of a hydrocarbon-
739 and xenobiotic-degrading bacterium, and features of relevance to biotechnology. *Appl Environ*
740 *Microb* **74**: 6405-6416.

741 Neufeld, J.D., Vohra, J., Dumont, M.G., Lueders, T., Manefield, M., Friedrich, M.W., and Murrell, J.C.
742 (2007) DNA stable-isotope probing. *Nat Protoc* **2**: 860-866.

743 Nielsen, H.B., Almeida, M., Juncker, A.S., Rasmussen, S., Li, J.H., Sunagawa, S. et al. (2014)
744 Identification and assembly of genomes and genetic elements in complex metagenomic samples
745 without using reference genomes. *Nat Biotechnol* **32**: 822-828.

746 Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017) metaSPAdes: a new versatile
747 metagenomic assembler. *Genome Res* **27**: 824-834.

748 Orellana, L.H., Rodriguez-R, L.M., and Konstantinidis, K.T. (2017) ROCker: accurate detection and
749 quantification of target genes in short-read metagenomic data sets by modeling sliding-window
750 bitscores. *Nucleic Acids Res* **45**.

751 Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing
752 the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome*
753 *Res* **25**: 1043-1055.

754 Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.A., Woodcroft, B.J., Evans, P.N. et al. (2017)
755 Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life.
756 *Nat Microbiol* **2**: 1533-1542.

757 Probst, A.J., Ladd, B., Jarett, J.K., Geller-McGrath, D.E., Sieber, C.M.K., Emerson, J.B. et al. (2018)
758 Differential depth distribution of microbial function and putative symbionts through sediment-
759 hosted aquifers in the deep terrestrial subsurface. *Nat Microbiol* **3**: 328-336.

760 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P. et al. (2013) The SILVA ribosomal
761 RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**:
762 D590-D596.

763 R Core Team (2019). R: A language and environment for statistical computing. R Foundation for
764 Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

765 Rimstidt, J.D., and Vaughan, D.J. (2003) Pyrite oxidation: A state-of-the-art assessment of the
766 reaction mechanism. *Geochim Cosmochim Ac* **67**: 873-880.

767 Roth, V.N., Lange, M., Simon, C., Hertkorn, N., Bucher, S., Goodall, T. et al. (2019) Persistence of
768 dissolved organic matter explained by molecular changes during its passage through soil. *Nat Geosci*
769 **12**: 755-+.

770 Ryabchykov, O., Bocklitz, T., Ramoji, A., Neugebauer, U., Foerster, M., Kroegel, C. et al. (2016)
771 Automatization of spike correction in Raman spectra of biological samples. *Chemometr Intell Lab*
772 **155**: 1-6.

773 Salinero, K.K., Keller, K., Feil, W.S., Feil, H., Trong, S., Di Bartolo, G., and Lapidus, A. (2009) Metabolic
774 analysis of the soil microbe *Dechloromonas aromatica* str. RCB: indications of a surprisingly complex
775 life-style and cryptic anaerobic pathways for aromatic degradation. *Bmc Genomics* **10**.

776 Sano, R., Kameya, M., Wakai, S., Arai, H., Igarashi, Y., Ishii, M., and Sambongi, Y. (2010) Thiosulfate
777 oxidation by a thermo-neutrophilic hydrogen-oxidizing bacterium, *Hydrogenobacter thermophilus*.
778 *Biosci Biotech Bioch* **74**: 892-894.

779 Schippers, A., Jozsa, P.G., and Sand, W. (1996) Sulfur chemistry in bacterial leaching of pyrite. *Appl*
780 *Environ Microb* **62**: 3424-3431.

781 Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B. et al. (2009)
782 Introducing mothur: Open-source, platform-independent, community-supported software for
783 describing and comparing microbial communities. *Appl Environ Microb* **75**: 7537-7541.

784 Schwab, V.F., Herrmann, M., Roth, V.N., Gleixner, G., Lehmann, R., Pohnert, G. et al. (2017)
785 Functional diversity of microbial communities in pristine aquifers inferred by PLFA- and sequencing-
786 based approaches. *Biogeosciences* **14**: 2697-2714.

787 Schwab, V.F., Nowak, M.E., Elder, C.D., Trumbore, S.E., Xu, X.M., Gleixner, G. et al. (2019) ^{14}C-free
788 carbon is a major contributor to cellular biomass in geochemically distinct groundwater of shallow
789 sedimentary bedrock aquifers. *Water Resour Res* **55**: 2104-2121.

790 Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068-2069.

791 Shrout, J.D., Scheetz, T.E., Casavant, T.L., and Parkin, G.F. (2005) Isolation and characterization of
792 autotrophic, hydrogen-utilizing, perchlorate-reducing bacteria. *Appl Microbiol Biot* **67**: 261-268.

793 Sieradzki, E.T., Koch, B.J., Greenlon, A., Sachdeva, R., Malmstrom, R.R., Mau, R.L. et al. (2020)
794 Measurement error and resolution in quantitative stable isotope probing: implications for
795 experimental design. *bioRxiv*.

796 Sizova, M., and Panikov, N. (2007) *Polaromonas hydrogenivorans* sp. nov., a psychrotolerant
797 hydrogen-oxidizing bacterium from Alaskan soil. *Int J Syst Evol Micr* **57**: 616-619.

798 Spona-Friedl, M., Braun, A., Huber, C., Eisenreich, W., Griebler, C., Kappler, A., and Elsner, M. (2020)
799 Substrate-dependent CO_2 fixation in heterotrophic bacteria revealed by stable isotope labelling. *Fems*
800 *Microbiol Ecol* **96**.

801 Starr, E.P., Shi, S., Blazewicz, S.J., Koch, B.J., Probst, A.J., Hungate, B.A. et al. (2020) Stable isotope
802 informed genome-resolved metagenomics uncovers potential trophic interactions in rhizosphere soil.
803 *bioRxiv*.

804 Stevens, T. (1997) Lithoautotrophy in the subsurface. *Fems Microbiol Rev* **20**: 327-337.

805 Sun, S.S., Liu, J., Zhang, M.P., and He, S.B. (2020) Thiosulfate-driven autotrophic and mixotrophic
806 denitrification processes for secondary effluent treatment: Reducing sulfate production and nitrous
807 oxide emission. *Bioresource Technol* **300**.

808 Tabatabai, M. (1974) A rapid method for determination of sulfate in water samples. *Environmental*
809 *Letters* **7**: 237-243.

810 Taubert, M., Baumann, S., von Bergen, M., and Seifert, J. (2011) Exploring the limits of robust
811 detection of incorporation of ^{13}C by mass spectrometry in protein-based stable isotope probing
812 (protein-SIP). *Analytical and bioanalytical chemistry* **401**: 1975-1982.

813 Taubert, M., Stähly, J., Kolb, S., and Küsel, K. (2019) Divergent microbial communities in groundwater
814 and overlying soils exhibit functional redundancy for plant-polysaccharide degradation. *Plos One* **14**.

815 Taubert, M., Stöckel, S., Geesink, P., Girnus, S., Jehmlich, N., von Bergen, M. et al. (2018) Tracking
816 active groundwater microbes with D_2O labelling to understand their ecosystem function. *Environ*
817 *Microbiol* **20**: 369-384.

818 Taubert, M., Vogt, C., Wubet, T., Kleinsteuber, S., Tarkka, M.T., Harms, H. et al. (2012) Protein-SIP
819 enables time-resolved analysis of the carbon flux in a sulfate-reducing, benzene-degrading microbial
820 consortium. *Isme J* **6**: 2291-2301.

821 Taubert, M., Grob, C., Howat, A.M., Burns, O.J., Pratscher, J., Jehmlich, N. et al. (2017) Methylamine
822 as a nitrogen source for microorganisms from a coastal marine environment. *Environ Microbiol* **19**:
823 2246-2257.

824 Uritskiy, G.V., DiRuggiero, J., and Taylor, J. (2018) MetaWRAP - a flexible pipeline for genome-
825 resolved metagenomic data analysis. *Microbiome* **6**.

826 Vollmers, J., Frentrup, M., Rast, P., Jogler, C., and Kaster, A.K. (2017) Untangling genomes of novel
827 planctomycetal and verrucomicrobial species from Monterey Bay kelp forest metagenomes by
828 refined binning. *Front Microbiol* **8**.

829 von Bergen, M., Jehmlich, N., Taubert, M., Vogt, C., Bastida, F., Herbst, F.A. et al. (2013) Insights from
830 quantitative metaproteomics and protein-stable isotope probing into microbial ecology. *Isme J* **7**:
831 1877-1885.

832 Wegner, C.E., Gaspar, M., Geesink, P., Herrmann, M., Marz, M., and Küsel, K. (2019) Biogeochemical
833 regimes in shallow aquifers reflect the metabolic coupling of the elements nitrogen, sulfur, and
834 carbon. *Appl Environ Microb* **85**.

835 Willems, A., Busse, J., Goor, M., Pot, B., Falsen, E., Jantzen, E. et al. (1989) *Hydrogenophaga*, a new
836 genus of hydrogen-oxidizing bacteria that includes *Hydrogenophaga flava* comb. nov. (formerly
837 *Pseudomonas flava*), *Hydrogenophaga palleronii* (formerly *Pseudomonas palleronii*),
838 *Hydrogenophaga pseudoflava* (formerly *Pseudomonas pseudoflava* and *Pseudomonas*
839 *carboxydoflava*), and *Hydrogenophaga taeniospiralis* (formerly *Pseudomonas taeniospiralis*). *Int J*
840 *Syst Bacteriol* **39**: 319-333.

841 Wrighton, K.C., Thomas, B.C., Sharon, I., Miller, C.S., Castelle, C.J., VerBerkmoes, N.C. et al. (2012)
842 Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* **337**:
843 1661-1665.

844 Wu, Y.W., Simmons, B.A., and Singer, S.W. (2016) MaxBin 2.0: an automated binning algorithm to
845 recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**: 605-607.

846 Yan, L.J., Herrmann, M., Kampe, B., Lehmann, R., Totsche, K.U., and Küsel, K. (2020) Environmental
847 selection shapes the formation of near-surface groundwater microbiomes. *Water Res* **170**.

848

849

850

Figure and Table legends

Figure 1: Quantification of deuterium incorporation by single cell Raman microspectroscopy. Boxplots depict the relative intensity of the Raman C-D-band determined as A(C-D) / [A(C-D) + A(C-H)] from single cell Raman spectra. Spectra were obtained from samples of groundwater microcosms with 30% D_2O (shaded) or with H_2O only (empty) at different time points of incubation. Boxes show median, first and third quartile. Whiskers indicate 5^{th} and 95^{th} percentile. Outliers are depicted as dots. At least n=147 spectra were obtained at each time point.

Figure 2: Clustering of selected MAGs based on carbon utilization. Left panel: Stable isotope cluster analysis based on PCA of ^{13}C incorporation profiles over incubation time obtained by SIP-metaproteomics of samples from the ^{13}C-microcosms. Each point represents an organism associated with one MAG. Clusters of MAGs are indicated with Latin numbers. Ellipses depict 95% confidence intervals. Only MAGs are shown where at least two replicates of ^{13}C incorporation patterns per time point could be acquired. Right panel: Representative ^{13}C incorporation profiles of MAGs marked with an asterisk are given for each cluster. Heatmaps depict the extent of ^{13}C incorporation in peptides of the corresponding MAG after 21 (T1), 43 (T2) and 70 days (T3) of incubation in 5%-intervals, ranging from 0 to 100% ^{13}C relative isotope abundance.

Figure 3: Generation times of groundwater organisms in the microcosms. Given values were determined for the first 3 weeks of incubation, based on the relative abundance of ^{12}C and ^{13}C peptides. Shown are mean and standard deviation based on $n \geq 4$ replicate determinations. Colored horizontal lines indicate average generation time for each cluster. bdl: generation time was below the detection limit of 2 days. na: no quantification of the generation time was possible.

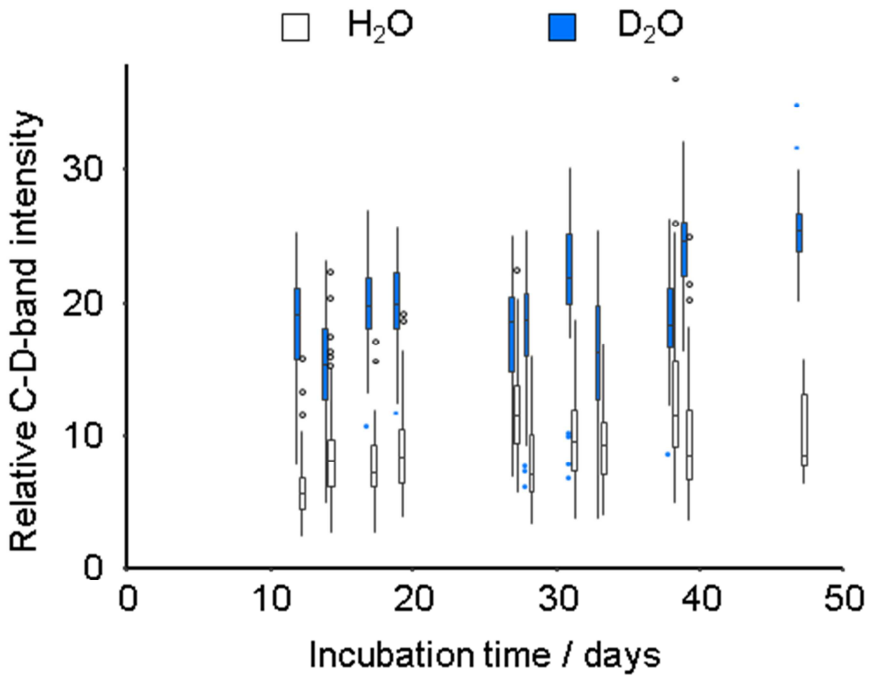
Figure 4: Metabolic functions of selected MAGs. The sizes of the bubbles correspond to the total number of peptides for each MAG and each functional category identified at any time point. Metabolic functions are grouped into CO_2 fixation (red), sulfur cycling (yellow), nitrogen cycling (green), aerobic respiration and ATP synthesis (blue) and organic carbon utilization (black). The

876 taxonomic categories “others” include peptides that were assigned to multiple MAGs affiliated with
877 the same genus. Only MAGs included in the stable isotope cluster analysis are shown. RuBisCO:
878 ribulose-1,5-bisphosphate carboxylase/oxygenase, CODH/ACS: carbon monoxide
879 dehydrogenase/acetyl-CoA synthase, TCA cycle: tricarboxylic acid cycle.

880 **Figure 5: Carbon flux between microbial clusters.** Red arrow inlays illustrate the portion of $^{13}CO_2$-
881 derived carbon assimilated by each microbial cluster after 21, 43 and 70 days. Overall arrow width
882 scales with total amount of carbon assimilated based on the relative abundance of the respective
883 microbial cluster in the metagenomics analysis. Fading grey arrows indicate uptake of unlabeled
884 organic carbon from the groundwater. Checkboxes highlight the presence and activity of metabolic
885 functions for CO_2 fixation, utilization of organic carbon and sulfur oxidation.

886

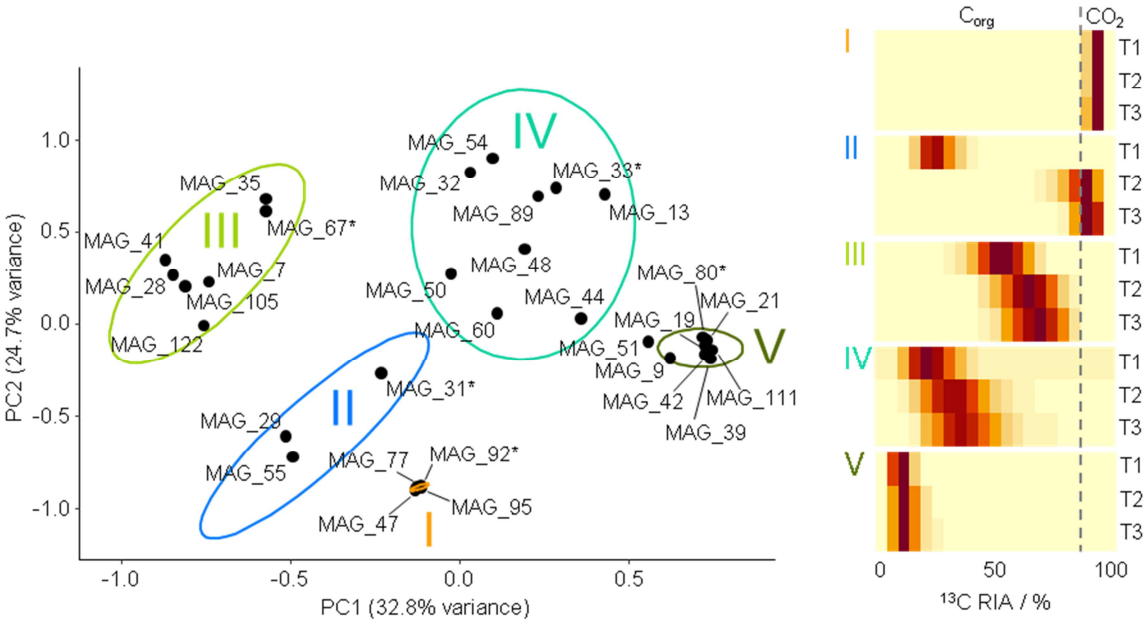
887 **Figure 1**



888

889

890 **Figure 2**



891

892

Figure 3

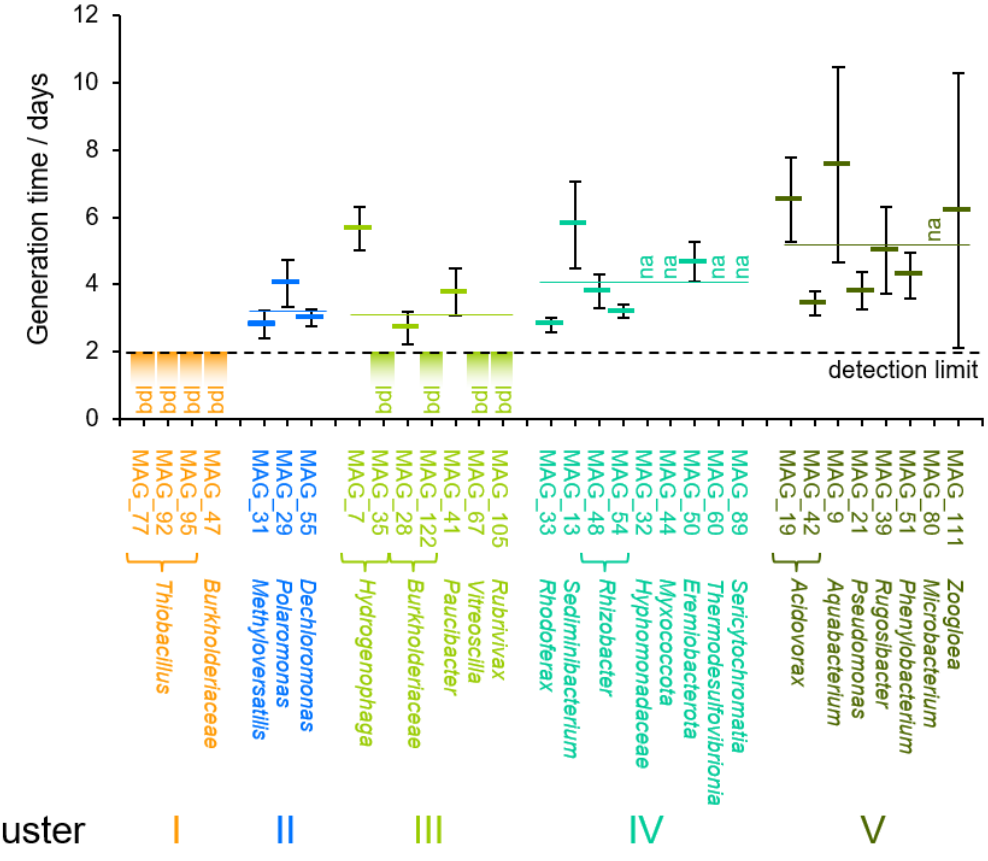


Figure 4

Figure 5