Rates of primary production in groundwater rival those in oligotrophic marine systems

Will A. Overholt[1], Susan Trumbore[2,3], Xiaomei Xu[3], Till L.V. Bornemann[4], Alexander J. Probst[4], Markus Krüger[1], Martina Herrmann[1,5], Bo Thamdrup[6], Laura Bristow[6], Martin Taubert[1], Valérie F. Schwab[2], Martin Hölzer[7,8], Manja Marz[7,8,9], and Kirsten Küsel[1,5*]

1. Aquatic Geomicrobiology, Institute of Biodiversity, Friedrich Schiller University, Jena, Germany
2. Max Planck Institute for Biogeochemistry, Jena, Germany
3. Department of Earth System Sciences, University of California, Irvine, CA, USA
4. Institute for Environmental Microbiology and Biotechnology, Department of Chemistry, University Duisburg-Essen, Germany
5. German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
6. Department of Biology, University of Southern Denmark, Odense M, Denmark
7. RNA Bioinformatics and High-Throughput Analysis, Faculty of Mathematics and Computer Science
8. European Virus Bioinformatics Center, Friedrich Schiller University Jena, Jena 07743, Germany
9. FLI Leibniz Institute for Age Research, Marz Associated Research Group, Jena 07745, Germany

*Corresponding author

Abstract

The terrestrial subsurface contains nearly all of Earth's freshwater reserves[1] and harbors upwards of 60% of our planet's total prokaryotic biomass[2,3]. While genetic surveys suggest these organisms rely on *in situ* carbon fixation, rather than the translocation of photosynthetically derived organic carbon[4–6], corroborating measurements of carbon fixation in the subsurface are absent. Using a novel ultra-low level ^{14}C-labeling technique, we show that *in situ* carbon fixation rates in a carbonate aquifer reached 10% of the median rates measured in oligotrophic marine surface waters, and were up to six-fold greater than those observed in lower euphotic zone waters where deep chlorophyll levels peak. Empirical carbon fixation rates were substantiated by both nitrification and anammox rate data. Metagenomic analyses revealed a remarkable abundance of putative chemolithoautotrophic members of an uncharacterized order of Nitrospiria – the first representatives of this class expected to fix carbon via the Wood-Ljungdahl pathway. Based on these fixation rates, we extrapolate global primary production in carbonate groundwaters to be 0.11 Pg of carbon per year.

Main

The continental subsurface is the planet's largest carbon reservoir[7], housing up to 19% of its total biomass[2,8] and 99% of its freshwater[1]. Despite accounting for only 6% of total stores, modern groundwater, *i.e.,* the fraction accrued in aquifers over the past 50 years, is the single-most significant source of potable water. Carbonate karst aquifers alone are thought to supply people with nearly 10% of their drinking water[9]. Unfortunately, modern groundwater is also the most vulnerable to anthropogenic and climatic impacts[1]. While subsurface ecosystems have long fascinated ecologists[10], and more recently microbiologists[11], accessibility, enormous spatial heterogeneity, and the complete lack of process rate measurements has obscured a meaningful understanding of their contributions to global biogeochemical cycles[12].

The widespread recognition that Earth's biosphere extends deep into the subsurface occurred only recently[13]. Historically, carbon supply in such environments was thought to be limited to the trickling of surface-produced organic matter into the shallow subsurface[5], or what was stored within sedimentary rocks[14]. In stark contrast, recent studies have shown that millimolar concentrations of dissolved H_2 amassed in deep Precambrian Shield fracture groundwaters support the proliferation of chemolithoautotrophs[15] and ultimately bacterivore nematodes[16]. A wealth of compelling genetic evidence suggests that *in situ* carbon fixation is critical for sustaining highly-diverse microbial metabolic networks in groundwater, both in the shallow and deep subsurface[4,17–23]. Despite the implications of gene-based surveys, the empirically derived activity measurements required to corroborate such inferences, constrain biogeochemical fluxes, understand system dynamics, and integrate processes into regional and global models have yet to be reported. Here we report our use of a novel radiocarbon method to derive empirical carbon fixation rates and place them in the context of global groundwater.

Groundwater carbon fixation rates resemble those of marine surface waters

In this first ever evaluation of primary productivity in the shallow subsurface (*i.e.*, groundwater wells 5 – 90 m deep), experimental carbon fixation rates varied from 0.043 ± 0.01 to 0.23 ± 0.10 μg C l^{-1} d^{-1} (mean ± SD; Fig. 1 A, Table S2, Supp. Info.). The ultra-low level ^{14}C labeling approach developed in this investigation exploits the high sensitivity of accelerator mass spectrometry, thereby minimizing impacts to groundwater hydrochemical equilibria and affording shorter incubation times. This method is particularly useful within a carbonate geological setting, where high dissolved inorganic carbon (DIC) backgrounds and a scarcity of microbes warrant greater sensitivity than is achievable via scintillation-based ^{14}C-labeling approaches. Rates resulting from our novel labeling technique likely approximate net primary productivity rather than gross productivity, as has been reported for marine systems[24,25], and we further expect them to be conservative estimates for carbon fixation (Supp. Info.).

We compared these carbon fixation rates that were measured in groundwater of varying biogeochemical characteristics[26] with the only other subsurface $^{14}CO_2$ assimilation measurements reported: those of a deep (830-1078 m) groundwater borehole from crystalline bedrock in Sweden [27]. To do so, we converted the published rates of isotopic incorporation to carbon equivalents, revealing the lower but overlapping range of 0.0095 to 0.056 μgC l^{-1} d^{-1}.

To better understand the relevance of the rates measured, we further compared them to those of well documented oligotrophic marine surface waters. Unlike our samples, the carbon

fixed in these waters was sourced almost entirely by bacterial photoautotrophs[28,29]. When compared directly to a comprehensive dataset compiled by Liang *et al.* [30], our rates overlapped with those of global marine waters at depths to 140 m, equating to roughly 10% of the reported global median for 0-20 m depths (2.65 μg l^{-1} d^{-1}, interquartile range (IQR) = 1.74, 6.02), and 20% of the median for 20-140 m depths (1.2 μg l^{-1} d^{-1}, IQR = 0.6, 1.7). Comparisons to the extensively studied Sargasso Sea in the Bermuda Atlantic Timeseries Study (BATS)[31,32] and the Hawaiian Oceanographic Timeseries (HOTs)[33] datasets yielded similar findings (Fig. 2). Our rate measurements ranged between three and 23% of the median reported net primary productivity in the upper euphotic zones (down to ~ 120 m) and 20 to 600% of the median of the lower euphotic region (100 - 120m) where deep chlorophyll levels peak.

We also considered contributions to existing particulate organic carbon (POC) stocks and new carbon inputs per microbial cell count. After normalizing for estimated total bacterial cell numbers, groundwater yielded 0.3 - 10.8 fg of fixed carbon per bacterial cell per day (Table S5), which matched estimates of 0.25 - 12.1 fg C per bacterial cell per day across the marine photic zone (5 - 150 m). However, groundwater received new daily carbon inputs of only 0.47% ± 0.22% (mean ± SD, Table S5) of its existing POC, much lower than the marine system's 2.6% ± 2.9% gain in the lower euphotic zone and 22% ± 18% at the surface [34,35]. This disparity might stem from the larger recalcitrant fraction of particulate organic carbon in groundwater compared to oligotrophic oceans, which is supported by deviations in ^{14}C and ^{13}C signatures of total groundwater POC concentrations compared to lipid signatures of resident microbes [36].

An ecosystem dominated by chemolithoautotrophs

To identify dominant microbial primary producers, a total of 4,175 metagenome-assembled genomes (MAGs) were generated from groundwater samples, 1,224 of which passed quality thresholds. Of these, 102 putative chemolithoautotrophs exhibiting at least 50% completion scores for carbon fixation pathways were identified (Fig. 3). Almost exclusively bacterial (101), these MAGs represented 17 distinct phyla, 21 classes, and 35 families (Fig. 3, Table S3). In some samples, up to 12% of metagenomic reads from a sample could be recruited to these chemolithoautotrophic MAGs (Supp. Info., Fig. S3). A single archaeal MAG of the family Nitrosopumilaceae encoded gene products for the 4-hydroxybutyrate 3-hydroxypropionate pathway and was not relatively abundant (< 5x max normalized coverage).

Three chemolithoautotrophic pathways were detected (Fig. 1 B): the Calvin-Benson-Bassham (CBB), Wood-Ljungdahl (WL), and reverse TCA (rTCA) cycles were present in 37, 50, and 15 MAGs, respectively. The summed and normalized relative coverages of MAGs equipped with these metabolic pathways aligned with the carbon fixation rates measured in wells H52, H32, and H14, while contrasting with rate data from wells H41and H43 (Fig. 1, Supp. Info.). The greatest relative abundances of chemolithoautotrophs were detected in oxic well H41 and anoxic well H52. Anoxic groundwater was dominated by putative sulfur oxidizing (53% of summed and normalized coverages of all chemolithoautotrophic MAGs) and anaerobic ammonium oxidizing (anammox [10%]) autotrophic microbes, while oxic groundwaters harbored greater abundances of potential nitrifiers (76%; Fig. S4 B, Supp. Info.).

Poorly characterized microbes influence carbon fixation potential

The most abundant putative chemolithoautotrophic populations represented by MAGs generated from anoxic groundwater were of poorly studied and/or uncharacterized microbial lineages. Those most abundant in oxic groundwaters, however, were phylogenetically and metabolically similar to well-characterized microbes (Supp. Info.). In both cases, metabolic reconstructions suggested that dominant subpopulations could access a diverse suite of (in)organic electron acceptors and donors. We mapped previously generated RNA-seq data [37] to these MAGs to confirm the active expression of gene products involved in energy acquisition and carbon fixation. As opposed to the broad distributions posited by DNA-based abundances, transcript data revealed far more restrictive ranges in which specific gene products were favored (Fig. 3). Given their metabolic versatility and the results of previous cultivation-based analyses [38], these populations are expected to be mixotrophic, *i.e.* capable of supplementing carbon requirements with available organic matter. Overall, carbon fixation in anoxic groundwater was predicted to be fueled by reduced sulfur and there were three highly abundant, sulfur-oxidizing MAGs identified, each accounting for > 2% of the total metagenomic reads in some samples (100-400x normalized coverages).

The most abundant MAG encountered in this study belongs to a deep-branching order, 9FT-COMBO-42-15, of class Nitrospira and is the first representative of Class Nitrospiria thought to fix carbon via the WL pathway (Fig. 3, Fig. 4A, Supp. Info.). As there is precedence for autotrophic WL-utilizing bacteria within phylum Nitrospirota, and *Ca.* Magnetobacterium was characterized with an equally flexible metabolism, these traits may be more widespread within the phylum than previously thought [39]. In addition, two MAGs capable of coupling sulfur oxidation to carbon-fixation via the CBB cycle were identified as members of the Sulfurifustaceae family of Proteobacteria (Supplemental Information, Figure 4 B). These MAGs recruited 10-fold more transcripts than their Nitrospirota counterparts and were among the most transcriptionally active chemolithoautotrophic genomes detected (Fig. 3, Fig. 4, Supp. Info.). With its closest reference genomes *Sulfuricaulis limicola* and *Ca*. Muproteobacteria (RIFCSPHIGHO2_12_FULL_60_33), the taxonomic identity of this family is under debate. Per GTDB classification nomenclature, Muproteobacteria belong to the Sulfurifustaceae family, and members of this family have been posited to oxidize sulfur in both aquatic and terrestrial environments [40–42].

Planctomycetota MAGs, predicted to couple anaerobic ammonium oxidation to carbon fixation via the WL pathway, exhibited mean transcriptional activities on par with their Sulfurifustaceae MAG counterparts (Fig.3, Fig. 4C, Supp. Info.). The elevated transcriptional activity of gene products germane to the CBB and WL pathways suggests that taxa wielding such functions play a disproportionately large role in chemolithoautotrophy relative to their DNA-based abundances. Surprisingly, all putative anammox MAGs detected were transcriptionally active in oxic groundwater (Fig. 3, Fig.4 B; wells H41 and H51). Anammox reactions are typically inhibited in the presence of oxygen [43], although microbes will still express critically important genes in low oxygen environments [44,45].

Nitrogen-based rate measurements validate carbon fixation rates

To evaluate the relationship between anammox and carbon fixation in anoxic groundwaters, we compared the rates of each in a well harboring the greatest relative abundance of anaerobic ammonium oxidizing bacteria (well H52). Well H52 exhibited anammox rates of 1.2 ± 0.5 nmol N_2 l^{-1} d^{-1}. Empirical stoichiometric data demonstrates that 1.02 moles of N_2 is produced via anaerobic ammonium oxidation for every 0.066 moles of $CH_2O_{0.5}N_{0.15}$ reduced to biomass[46]. Assuming equivalent stoichiometry, the rate of carbon fixation via anammox in groundwater would be 0.93 ± 0.39 ng C l^{-1} d^{-1}, more than 200 times lower than the 220 ng C l^{-1} d^{-1} measured. This result is corroborated by metagenomic data that suggest the high rate of carbon fixation in anoxic groundwater is more likely driven by reduced sulfur than reduced nitrogen.

Metagenomic and metatranscriptomic data predicted that nearly all of the organic carbon produced under oxic conditions in well H41 would be coupled to nitrification. To test this, we monitored the rate of aerobic ammonium oxidation in this well and recorded a mean production of 125.8 ± 5.9 nmol NO_2^- + NO_3^- l^{-1} d^{-1}. Since the most abundant nitrifiers detected were most closely related to complete ammonium oxidizing bacteria (Supp. Info.), we based our calculations on the 394 mg protein per mol of ammonia growth yields of *Nitrospira inopinata,* a comammox organism[47]. Assuming a cellular composition of $C_5H_7O_2N$[48] and 55% protein content, we estimated a rate of 48.5 ± 1.9 ng C l^{-1} -d^{-1}, which was well within the range of error for our measured rate of 43 ± 13 ng C l^{-1} d^{-1} confirming the importance of nitrification for carbon fixation at this site.

Global estimates for groundwater primary productivity

There are an estimated 22.6 million km^3 of groundwater on Earth [1], 2.26 and 12.66 million km^3 of which are housed in carbonate and crystalline aquifers, respectively. If we assume that our average rates accurately represent carbonate groundwater systems, then 0.108 ± 0.069 Pg (mean ± SD) of carbon is fixed every year in this global ecosystem (Table S4). If the values reported from crystalline aquifers[27] are representative of this environment, then another 0.15 ± 0.11 Pg C would be fixed there every year. Collectively, the net primary productivity of ~ 66% of the planet's groundwater reservoirs would total 0.26 Pg C yr^{-1}, approximately 0.5% that of marine systems and 0.25% of global NPP estimates[49]. Applying the lowest measured values from each rock type yields the more conservative estimate of 0.079 Pg C yr^{-1}, 0.076% of the global NPP. Although the total production rates in groundwater seem small, this is because of the relatively small volume of groundwater compared to the vast surface ocean - what is most surprising is that terrestrial subsurface fixation rates are approaching those of phototrophic organisms.

Conclusions

Using a novel ultra-low level ^{14}C-labeling technique to generate empirically derived estimates of primary productivity in groundwater for the first time, we showed that carbon fixation rates in a carbonate aquifer reached 10% of the median rates reported in oligotrophic marine surface waters and six-fold greater than those observed in the deep chlorophyll maximum of the lower euphotic zone. Normalizing rates according to estimated bacterial numbers revealed equivalent

carbon input (*i.e.,* 0.3 - 12 fg C per cell) for both systems, despite the fact that daily inputs of new POC were 40 times greater in marine waters than in groundwater. This disparity makes sensel, since trophic webs are simpler in the subsurface, and the export of organic matter is constrained by long water residence times within the aquifer. As the vast majority of photosynthetically derived carbon in marine systems is labile (half-life < 1 day), the findings of this study solicit new hypotheses regarding carbon cycling in the subsurface, particularly those positing newly synthesized carbon rather than surface-derived organic matter as the primary source of fuel for microbiota.

Complementary metagenomic analyses identified novel microbes capable of exploiting metabolic pathways previously unreported for their given phylotype. Carbon fixation rates coincided with the potential for chemolithoautotrophy in three of the five groundwater wells examined. Comprehensive metabolic reconstructions revealed versatile metabolisms with access to numerous sources of electron donors and acceptors, particularly in taxa detected in high abundance in anoxic and hypoxic groundwater. While these populations were widely distributed across broad biogeochemical regimes, the use of previously generated metatranscriptomes helped identify more specific activities.

Nitrogen-based transformations provided independent validation of carbon fixation rates in oxic waters and corroborated metagenomic data that hinted at the inconsequential impact of anammox on carbon fixation in anoxic groundwater. If our average rates accurately represent all carbonate groundwater systems, then 0.108 ± 0.069 (mean ± SD) Pg of carbon (0.22% of global marine NPP) is fixed every year under these geologic settings. Applying these rates of carbon fixation to ecosystem processes alters the way we think about these environments, challenges the importance of surface-derived organic matter fluxes on shallow subsurface functioning, and establishes a framework broadly applicable across groundwater systems.

Methods

Site description

Groundwater samples were sampled from the Hainich Critical Zone Exploratory (NW Thuringia, Germany) [26,50,51]. This aquifer assemblage consists of a multistory, fractured system composed of alternating layers of limestone and mudstone that developed along a hillslope of Upper Muschelkalk bedrock [26]. The primary aquifer, represented in this study by wells H41 and H51, is oxic and lies within the Trochitenkalk formation (moTK). Primarily suboxic to anoxic, mudstone-dominated overhanging strata lies within the Meissner formation (moM) and is represented here by wells H14 (moM - substory 1), H32 (moM - 5, 6, 7), H43 (moM - 8), and H52 (moM - 3, 4). Geochemically, H32 and H41 coalesce into a single cluster while each of the other wells represent distinct regimes. Consistent with previous microbiological characterizations, however, each well studied represented a distinct community state [52].

^{14}C-DIC incorporation assay

This method, similar to a sensitive methane oxidation technique previously described[53], is a modification of traditional ^{14}C-CO_2 primary productivity approaches[54] predicated upon the sensitivity offered by accelerator-based mass spectrometry. Groundwater was collected in July

2020 during sampling campaign PNK130, as described by Herrmann et al.[23]. After approximately three well volumes had been discharged and physicochemical parameters stabilized, groundwater was collected directly into nine pre-sterilized 2-liter borosilicate bottles, from the bottom up. Bottles were then overfilled with > two volumes and sealed with gas-tight rubber stoppers. Triplicate samples from each well were then subjected to three treatments. A labelling treatment consisted of 6.77×10^{-7} mmol C-$NaHCO_3$ which contained 200 Bq of activity [50 µCi; American Radiolabeled Chemicals (ARC), St. Louis, MO] diluted to 9.38 Bq/µl with sterilized milliQ water, adjusted to pH 10, and verified using a scintillation counter. An advantage of this ^{14}C technique is that the small amount of tracer added (representing 0.000006% of the total DIC) did not change the substrate concentration or influence conditions like pH that could affect microbial populations. Kill controls were prepared in the same way, except 10 ml of 50% $ZnCl_2$ (w/v; final conc. 36.7 mM) was added to inhibit microbial activity. Unamended groundwater was also used as a control. All bottles were incubated in the dark at near *in situ* temperature for ~ 24 hours. Entire volumes were acidified to pH 4 with 3M HCl, bubbled with N_2 for one h to remove DIC, and then filtered through pre-baked (550□ 8 hours) quartz fibers (47 mm, 0.3um pore size, Macherey-Nagel QN-10) using pre-baked filter stands (EMD Millipore).

Filters were vacuum dried, sealed in quartz tubes with cupric oxide wire under vacuum, and combusted at 900 □ for two hours. Evolved CO_2 was purified cryogenically, measured as pressure in a known volume to determine C content, and reduced to graphite for measurement by accelerator mass spectrometry at the WM Keck Carbon Cycle Accelerator Mass Spectrometry facility[55]. From the label incorporation and amount of carbon retained on the filters (Supplemental Data File), fixation rates were calculated using equation (1):

$$(1) \quad \text{carbon fixation rate}(\text{mgC l}^{-1}\ \text{d}^{-1}) = \frac{\left[\frac{^{14}C}{^{12}C}\text{sample} - \frac{^{14}C}{^{12}C}\text{control}\right] * \text{POC}_{\text{mg}}}{\frac{^{14}C}{^{12}C}\text{DICsample} * \text{V} * \text{T}}$$

The technical variation was at most 3.6% (median = 0.78%) of the biological variation for the ^{14}C measurements and was not considered in standard error of the mean calculations. Standard error of the mean was determined for both the ^{14}C-based measurements (difference between two sets of triplicates, label and control, or label and kill controls) and POC measurements (all nine bottles from each well), separately. These errors were then propagated to yield the final error estimations. Analyses of variance and post-hoc Tukey HSD tests were conducted on resulting summary statistics (mean ± SEM) using the following utility: https://acetabulum.dk/cgi-bin/anova. All ^{14}C enrichment values were calculated using the differences between the 200 Bq-labeled samples and the 200 Bq-labeled kill controls. Rates calculated based on no-label addition controls are presented in Table S2. Data from global oligotrophic marine systems was included from Supplementary Data Sheet 1[30], the Bermuda Atlantic Timeseries years 1988-2016 via FTP (http://batsftp.bios.edu/BATS/production/bats_primary_production.txt) [31], and from the Hawaiian Oceanographic Timeseries via FTP (ftp://ftp.soest.hawaii.edu/hot/primary_production). POC data from both sites was extracted from Dryad datasets generated by Martiny et al. [34,35]. Bacterial cell number estimates for HOTs were obtained from the FTP site: https://hahana.soest.hawaii.edu/FTP/hot/microscopy/EPIslides.txt.

^{15}N-isotope incubation experiments

Groundwater from wells H41 and H52 was collected in September 2018 and November 2018 to measure nitrification rates and anammox rates, respectively. Briefly, groundwater was collected into sterile glass bottles, from the bottom up, using a sterile tube. Bottles were then overfilled with three volume exchanges, and sealed headspace free with silicone septa. Each sample was collected in triplicate alongside one control bottle per well. Samples were kept at 4°C until they were processed (no more than 2 h post-collection).

For nitrification measurements, 10 ml was removed from each sampling bottle (total vol. 0.5 L) and replaced with N_2 to analyze inorganic nitrogen and pH. Groundwater from control bottles was sterile filtered through a 0.2 µm filter (Supor, Pall Corporation, USA). Sterile filtered ^{15}N ammonium sulfate solution (98%, Cambridge Isotope Laboratories, Tewksbury), serving as a substrate for ammonia oxidizing prokaryotes, was then was added to a final conc. of 50 µM. Samples were incubated at 15°C in the dark *sans* agitation for five days. Ten-mL fractions were removed and replaced with N_2 at the outset of the experiment and after 12, 24, 48, 70 and 120 hours via filtration through 0.2 µm filters, which were stored at -20°C for isotopic ratio mass spectrometry (IRMS) analyses. Additional 10 ml fractions were removed at intervals to monitor pH and inorganic nitrogen during the incubation.

For anammox rate measurements, sampling bottles (total vol. 1 L) were flushed with N_2 under sterile conditions for 30 minutes to remove all remnants of oxygen. Five-mL fractions were removed and replaced with N_2 from each sample (and control) bottle to assess background $^{14}NH_4^+$ concentrations. Subsequently, samples were spiked with either (a) 50 µM $^{15}NH_4^+$ + 5 µM $^{14}NO_2^-$ or (b) 5 µM $^{15}NO_2^-$ as previously described [56]. Control bottles, serving as abiotic controls were sterile filtered (0.2 µm filters; Supor, Pall Corporation, USA) prior to flushing and the addition of nitrogen compounds. To facilitate destructive sampling at eight time points, groundwater (30 mL; in triplicate) was dispensed into sterile serum bottles leaving ~ 8 mL of headspace. Bottles were immediately sealed with butyl septa, crimp sealed and the headspace was purged with He. All bottles were then incubated in the dark at 15°C *sans* agitation and incubations were terminated after 0, 12, 24, 36, 48, 60, 72 and 96 hours by adding 300 µl of 50% (v/w) aqueous zinc chloride solution.

Nitrification rates were determined based on $^{15}NO_2^-$ + $^{15}NO_3^-$ production in incubations with $^{15}NH_4^+$. $^{15}NO_2^-$ and $^{15}NO_3^-$ were converted to N_2 via cadmium reduction followed by a sulfamic acid addition[57,58]. The N_2 produced ($^{14}N^{15}N$ and $^{15}N^{15}N$) was analyzed on a gas chromatography isotope ratio mass spectrometer as previously described[59]. Rates were evaluated from the slope of the linear regression of ^{15}N produced with time and corrected for the fraction of the NH_4^+ pool labelled in the initial substrate pool. The production of ^{15}N labeled N_2 from anammox was analyzed on the same IRMS as for nitrification rates, and calculated as described[60]. Note, denitrification was not detected in any of the $^{15}NO_2^-$ incubations. T-tests were applied ($p < 0.05$) to assess whether rates were significantly different from zero (Fig. S2).

DNA extraction & sample preparation

Samples used to generate metagenomic libraries were collected in January 2019 during sampling campaign PNK 110. For each sample replicate, approximately 50 - 100 L of groundwater was filtered sequentially through 0.2 µm and 0.1 µm pore sized PTFE filters (142

mm, Omnipore Membrane, Merck Millipore, Germany; Table S1). With the exception of H32 (did not yield sufficient volumes), each well was sampled in triplicate. H32 was duplicated using a sample previously collected during campaign PNK108 (November 2018). Filters were frozen on dry ice and stored at -80 °C prior to extraction. DNA was extracted using a phenol-chloroform based method, as previously described[61], and resulting DNA extracts were purified using a Zymo DNA Clean & Concentrator kit. Metagenome libraries were generated with a NEBNext Ultra II FS DNA library preparation kit, in accordance with manufacturer's protocols. DNA fragment sizes were estimated using an Agilent Bioanalyzer DNA 7500 instrument with High Sensitivity kits depending on DNA concentrations and recommendations of protocols (Table S1). Sequencing of the 32 samples was performed at the Core DNA Sequencing Facility of the Fritz Lipmann Institute in Jena, Germany using an Illumina NextSeq 500 system (2 x 150bp). Resulting metagenomic library sizes ranged from 16.4 to 22.1 Gbp (mean = 19.6 Gbp; Table S1), and raw data was deposited into the ENA under project PRJEB36523.

Metagenomic assembly and binning

Adapters were trimmed and raw sequences subjected to quality control processing using BBduk v38.51 [62]. Assembly and binning were performed as previously described [63]. Briefly, all libraries were independently assembled into scaffolds using metaSPAdes v3.12 [64], all of which were taxonomically classified per Bornemann et al. [63]. For individual assemblies, open reading frames (ORFs) were identified using Prodigal v2.6.3 in meta mode [65]. To generate coverage profiles, all quality-assessed and quality-controlled (QAQC) sequences from each of the 32 metagenomic libraries were mapped back to each of the 32 scaffold databases using Bowtie2 v2.3.4.3 in the sensitive mode [66].

Scaffolds were binned using differential coverages and tetranucleotide frequencies with Maxbin2 [67]. Additionally, ESOM and abawaca [68] were used for both manual and automatic binning, based on tetranucleotide sequence signatures, using 3 kbp and 5 kbp or 5 kbp and 10 kbp as minimum scaffold sizes, respectively. DAS Tool [69] was used with default parameters to reconcile resulting bin sets. Complete sets of bins from each of the samples were dereplicated using dRep v2.4.0 [70]. All scaffolds, bin assignments, ORF predictions, and taxonomic annotations were then imported into Anvi'o v6.0 [71]. Each of the resulting 1,275 bins was manually curated in Anvi'o v6, considering both coverage and sequence compositions. In the end, 1,224 bins passed the 30% completeness [median = 61%, IQR = (49%,73%)] and 10% redundancy [median = 0%, IQR = (0%,1.4%)] quality thresholds.

Characterizations of the Metagenome-assembled Genomes

ORFs originating from all of the resulting metagenome-assembled genomes (MAGs) were annotated using kofamscan [72] with the "detail" flag, and KO annotations were filtered using a custom script (https://git.io/JtHVw). This utility preserves hits with scores of at least 80% of the kofamscan defined threshold, as well as those exhibiting a score > 100 if there is no threshold. We elected to relax the default thresholds since all MAGs representing putatively chemolithoautotrophic microbes were verified manually, and we noticed that the best reciprocal blast hits with known reference sequences routinely scored below the kofamscan thresholds, *i.e.,* we favored false positives over false negatives since we included a secondary verification step.

KEGGDecoder [73] was used to assess the metabolic potential of five of the primary chemolithoautotrophic pathways: the Calvin-Benson-Bassham cycle, the Wood-Ljundahl pathway, the reverse citric acid cycle, the 4-hydroxybutyrate 3-hydroxypropionate pathway, and the 3-hydroxypropionate bicycle. MAGs were examined in greater depth if a given pathway was > 50% complete. MAGs representing potential chemolithoautotrophs were re-annotated using the online BlastKoala server [74] with essential steps verified through blast [75] against the RefSeq database. A collection of HMM models was used to determine which form of Rubisco was detected, along with potential hydrogenases [41]. Using blastp [75], dissimilatory bisulfite reductases (dsrAB) were compared to a database compiled by Pelikan *et al*. [76] to predict whether the pathway operated in an oxidative or reductive manner. Blast was used to compare gene hits for narGH/nxrAB (nitrate reductase / nitrite oxidoreductase) to a custom database based on sequences presented within Lücker *et al*.[77].

All QAQC reads were remapped to a database consisting of only contigs of dereplicated MAGs. Normalized coverages for each of the MAGs was determined by scaling the resulting Anvi'o-determined coverages based on the number of RNA polymerase B (*rpoB*) genes identified in the QAQC-filtered reads. *RpoB* sequences were identified using ROCker with the precomputed model [78]. Scaling factors were calculated by dividing the maximum number of *rpoB* identified in the 32 metagenomic libraries by the number of *rpoB* detected in each sample. Reported values represent averages of the triplicates/replicates, unless stated otherwise. The taxonomy of each MAG was evaluated using the GTDB_TK tool kit [79] in concert with the Genome Taxonomy Database (release 89) [80,81] and its associated utilities [65,82–86]. Single copy marker genes were identified and aligned bwith GTDB_TK for all bacterial MAGs, and a phylogenetic tree of the concatenated alignment was constucted using FastTree2 v2.1.10 in accordance with the JTT+CAT evolutionary model. The resulting phylogenetic tree was then imported into iToL [87] for visualization, and all MAGs were subjected to growth rate index (GriD) analysis within each metagenomic library[88].

Previously generated mRNA-enriched and post-processed metatranscriptomic libraries were procured from project PRJEB28783[37]. The groundwater source of these metatranscriptomes was collected in August and November 2015. QAQC filtered reads were mapped to MAGs using Bowtie2 v2.3.5 in sensitive mode[66], and the total number of *rpoB* transcripts from each metatranscriptomic library were determined, as described above for metagenomes. The transcriptomic coverages for each ORF from each MAG were determined using Anvi'o v6 and normalized via scaling factor calculations based on the total number of *rpoB* reads from the original metatranscriptome library (*i.e.,* the coverage of each ORF from each MAG was normalized to a community-wide estimate of the transcriptional activity of a house-keeping gene in each sample). Means were determined considering all of the metatranscriptomes generated from a given well, including different sampling timepoints. While well H32 was only sampled once, mean values from all other wells account for three to four metatranscriptome coverages each. Additionally, an average of the resulting normalized coverages for each MAG from each sample (sum of the MAG transcriptional coverage divided by the number of ORFs) was determined to estimate the relative transcriptional activity of the MAGs across the transect. Data was compiled and processed using R v.3.5.2 with Rstudio v1.1.463[89,90] and the tidyverse package[91], and color schemes were generated using the RColorBrewer utility[92]. All MAGs were deposited in project PRJEB36505's data repository.

Acknowledgments

We thank Falko Gutmann, Heiko Minkmar, Perla Abigail Figueroa Gonzalez, and the Hainich CZE site manager Robert Lehmann for their assistance with sample preparation, collection, and filtration. We also thank Martin Nowak and John Southon for advice regarding ^{14}C-bicarbonate supplementation and quantification of AMS results, respectively. Additionally, 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.
This study is part of the Collaborative Research Centre AquaDiva of the Friedrich Schiller University Jena, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1076 –Project Number 218627073. ST, XX and VFS acknowledge additional support from the European Research Council (Horizon 2020 Research and Innovation Programme, grant agreement 695101). MT gratefully acknowledges funding from the DFG under Germany's Excellence Strategy - EXC 2051 - Project-ID 390713860. TLVB and AJP were supported by the Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen ('Nachwuchsgruppe Dr. Alexander Probst'). 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). The scientific results have in part been computed at the High-Performance Computing (HPC) Cluster EVE, a joint effort of both the Helmholtz Centre for Environmental Research - UFZ (http://www.ufz.de/) and the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig (http://www.idiv-biodiversity.de/).

Data Availability

The metagenomic raw data, individual sample assemblies, and the metagenome-assembled-genomes for this study were deposited into the ENA under project PRJEB36523. All raw and summarized AMS data is available in the Supplemental Data File.

Author contributions

KK, ST, WAO and MT designed this study. WAO, KK, AJP, TLVB, MM, MH, and MT planned, designed, and conducted the metagenomic sampling approach. WAO, TLVB, and AJP performed the metagenomic analysis. WAO, ST, XX, and VFS performed the carbon fixation experiments and interpreted the results. MK, MH, BT, and LB conducted the anammox and nitrification rate measurements. WAO, KK and ST wrote the manuscript with the help of all authors.

Figures

(Vector graphics versions of all figures are attached as a separate document)

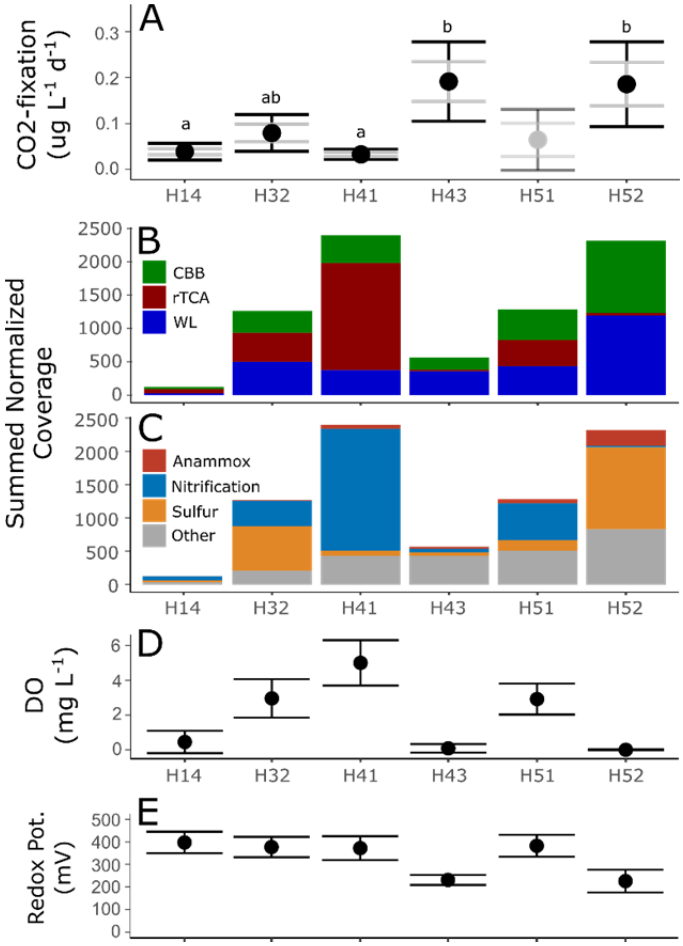


Fig. 1. (A) Rates of carbon fixation across the aquifer transect. Outer error bars depict one standard deviation while inner grey bars delineate standard error of the mean. Rates for well H51 are derived from non-labeled controls (see Supp. Info.). Letters denote the results of ANOVA and posthoc Tukey tests (H51 was excluded from testing). (B) Relative importance of the predicted carbon fixation pathways and (C) electron donor sources in each well. Values are averages from the triplicate 0.2-µm filtered fraction metagenome samples. (D) Mean dissolved oxygen concentrations in groundwater collected in summer months (May through September 2010 - 2018), error bars depict standard deviation. (E) Redox potential measurements from identical time points.

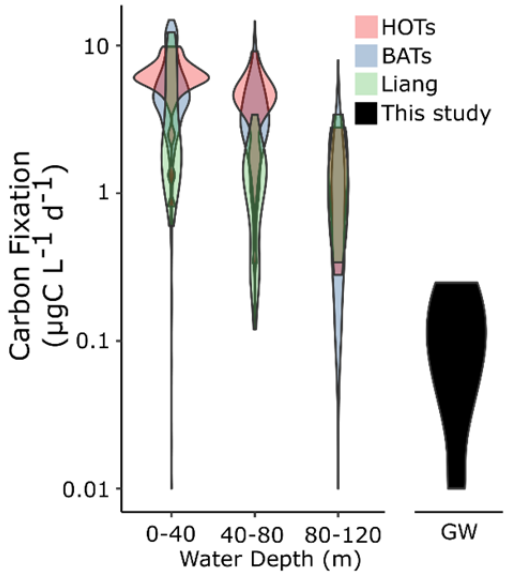


Fig. 2. Violin plots depicting the distribution of carbon fixation rates measured in oligotrophic marine surface waters and groundwater. HOTS = Hawaiian Oceanographic Timeseries (1999, cruises 101-110), BATs = data from 1988-2016 for the Bermuda Atlantic Timeseries, and Liang = a collated dataset compiled by Liang *et al*., 2017. GW represents the range of groundwater samples within the Hainich CZE.

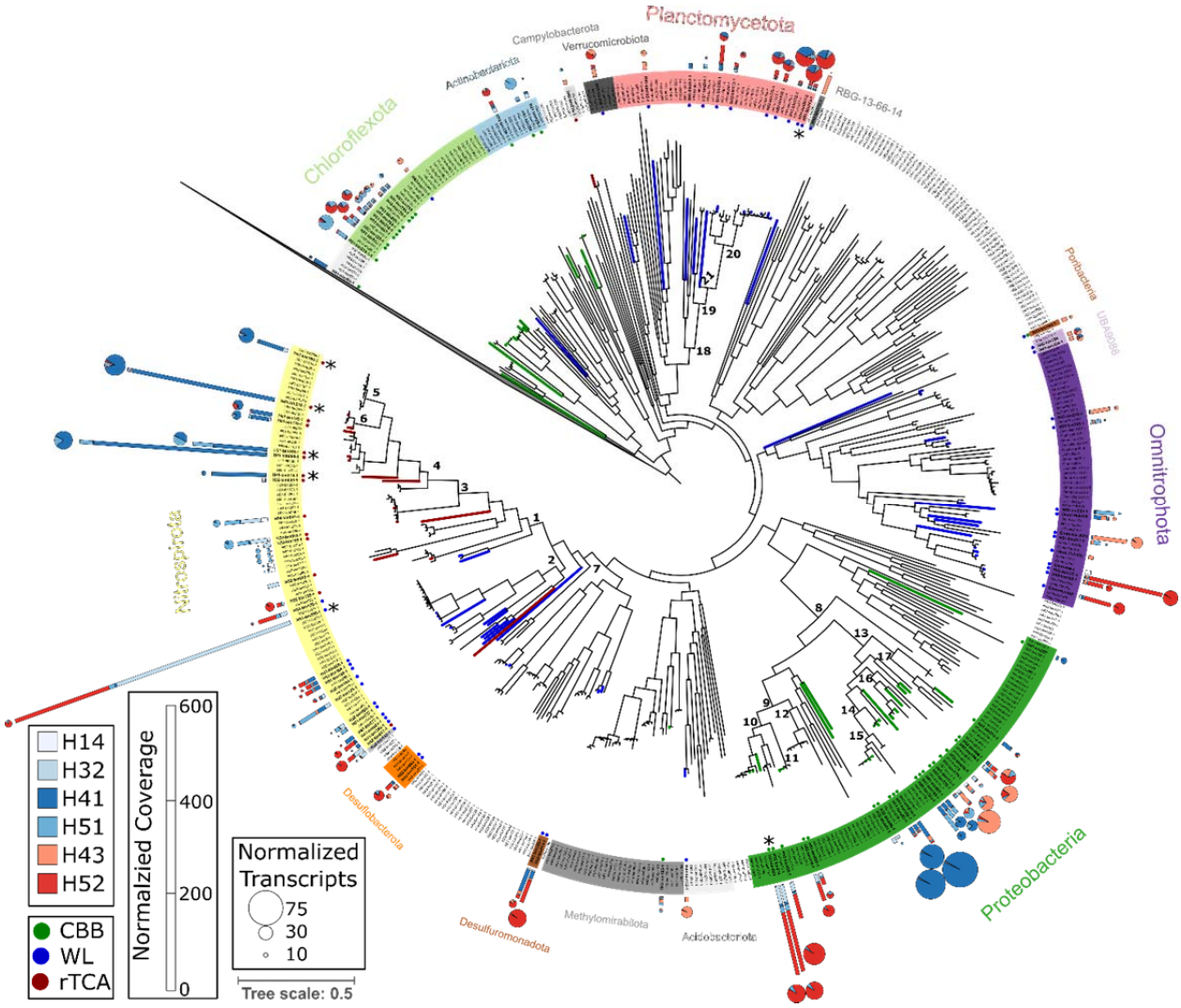


Figure 3. Approximately-maximum-likelihood phylogenetic tree based on concatenated single-copy protein alignments for all bacterial MAGs considered. Branches are colored according to the predicted carbon fixation pathway, and the matching leaf is indicated by a point. Bar charts present average normalized metagenomic coverages within each well from triplicate 0.2-µm filtered fractions. Pie charts show the coverage of mRNA transcripts recruited, normalized to gene number and library size. CBB = Calvin-Benson-Basham cycle, WL = Wood Ljungdahl pathway, rTCA = reverse TCA cycle. Asterisks denote MAGs discussed in greater detail in Supp. Info. The tree is rooted using Patescibacteria (CPR) as an outgroup, indicated by the collapsed grey leaf in the upper left. Node numbers represent: (1) c_Nitrospiria, (2) c_Thermodesulfovibrionia, (3) o_Nitrospirales, (4) f_Nitrospiraceae, (5) g_Nitrospira_D, (6) g_Nitrospira, (7) p_Nitrospinota, (8) c_Gammaproteobacteria, (9) o_Acidiferrobacterales, (10) f_Sulfurifustaceae, (11) g_SM1-46, (12) f_UBA6901, (13) o_Burkholderiales, (14) f_Nitrosomonadaceae, (15) g_Nitrosomonas, (16) f_SG8-41, (17) f_SG8-39, (18) c_Brocadiae, (19) o_Brocadiales, (20) f_Brocadiaceae, and (21) f_Scalinduaceae.

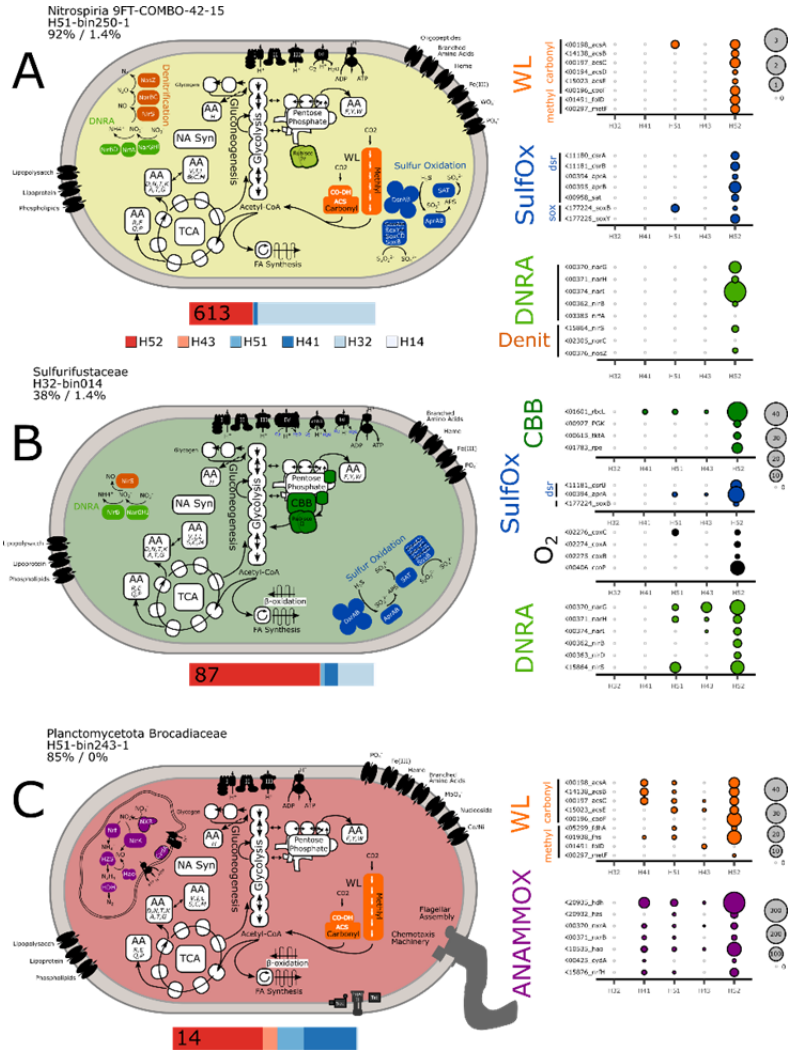


Fig. 4. Metabolic reconstructions of dominant populations recovered from wells exhibiting both elevated rates of carbon fixation and relatively high abundances of MAGs representing chemolithoautotrophs. Bar charts below each metabolic model summarize the average normalized coverage across each sample, scaled proportionally. Values indicate the sum normalized coverage for each MAG. Balloon plots depict normalized transcript coverages for genes affiliated with each pathway. If multiple copies were present, only the most active copy was plotted.

References

1. Gleeson, T., Befus, K. M., Jasechko, S., Luijendijk, E. & Cardenas, M. B. The global volume and distribution of modern groundwater. *Nat. Geosci.* **9**, 161–167 (2016).
2. Magnabosco, C. *et al.* The biomass and biodiversity of the continental subsurface. *Nat. Geosci.* **11**, 707–717 (2018).
3. Flemming, H.-C. & Wuertz, S. Bacteria and archaea on Earth and their abundance in biofilms. *Nat. Rev. Microbiol.* **17**, 247–260 (2019).
4. Probst, A. J. *et al.* Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat Microbiol* **3**, 328–336 (2018).
5. Griebler, C. & Lueders, T. Microbial biodiversity in groundwater ecosystems. *Freshw. Biol.* **54**, 649–677 (2009).
6. Probst, A. J. *et al.* Lipid analysis of CO2-rich subsurface aquifers suggests an autotrophy-based deep biosphere with lysolipids enriched in CPR bacteria. *ISME J.* **14**, 1547–1560 (2020).
7. Falkowski, P. *et al.* The global carbon cycle: a test of our knowledge of earth as a system. *Science* **290**, 291–296 (2000).
8. McMahon, S. & Parnell, J. Weighing the deep continental biosphere. *FEMS Microbiol. Ecol.* **87**, 113–120 (2014).
9. Stevanović, Z. Karst waters in potable water supply: a global scale overview. *Environ. Earth Sci.* **78**, 662 (2019).
10. Poulson, T. L. & White, W. B. The cave environment. *Science* **165**, 971–981 (1969).
11. Rusterholtz, K. J. & Mallory, L. M. Density, activity, and diversity of bacteria indigenous to a karstic aquifer. *Microb. Ecol.* **28**, 79–99 (1994).
12. Smith, H. J. *et al.* Impact of hydrologic boundaries on microbial planktonic and biofilm communities in shallow terrestrial subsurface environments. *FEMS Microbiol. Ecol.* **94**, (2018).
13. Alexander, M. *Introduction to Soil Microbiology*. (Wiley, 1977).
14. Krumholz, L. R., McKinley, J. P., Ulrich, G. A. & Suflita, J. M. Confined subsurface microbial communities in Cretaceous rock. *Nature* **386**, 64–66 (1997).
15. Lollar, B. S. *et al.* Hydrogeologic Controls on Episodic H2 Release from Precambrian Fractured Rocks—Energy for Deep Subsurface Life on Earth and Mars. *Astrobiology* **7**, 971–986 (2007).
16. Borgonie, G. *et al.* Nematoda from the terrestrial deep subsurface of South Africa. *Nature* **474**, 79–82 (2011).
17. Magnabosco, C. *et al.* A metagenomic window into carbon metabolism at 3 km depth in Precambrian continental crust. *ISME J.* **10**, 730–741 (2016).
18. Stevens, T. O. & McKinley, J. P. Lithoautotrophic Microbial Ecosystems in Deep Basalt Aquifers. *Science* **270**, 450–455 (1995).
19. Tiago, I. & Veríssimo, A. Microbial and functional diversity of a subterrestrial high pH groundwater associated to serpentinization. *Environ. Microbiol.* **15**, 1687–1706 (2013).
20. Mccollom, T. M. & Amend, J. P. A thermodynamic assessment of energy requirements for biomass synthesis by chemolithoautotrophic micro-organisms in oxic and anoxic

environments. *Geobiology* **3**, 135–144 (2005).
21. Momper, L., Jungbluth, S. P., Lee, M. D. & Amend, J. P. Energy and carbon metabolisms in a deep terrestrial subsurface fluid microbial community. *ISME J.* **11**, 2319–2333 (2017).
22. Jewell, T. N. M., Karaoz, U., Brodie, E. L., Williams, K. H. & Beller, H. R. Metatranscriptomic evidence of pervasive and diverse chemolithoautotrophy relevant to C, S, N and Fe cycling in a shallow alluvial aquifer. *ISME J.* **10**, 2106–2117 (2016).
23. Herrmann, M., Rusznyák, A. & Akob, D. M. Large fractions of CO2-fixing microorganisms in pristine limestone aquifers appear to be involved in the oxidation of reduced sulfur and nitrogen compounds. *Applied and* (2015).
24. Peterson, B. J. Aquatic Primary Productivity and the 14C-CO2 Method: A History of the Productivity Problem. *Annu. Rev. Ecol. Syst.* **11**, 359–385 (1980).
25. Viviani, D. A., Karl, D. M. & Church, M. J. Variability in photosynthetic production of dissolved and particulate organic carbon in the North Pacific Subtropical Gyre. *Frontiers in Marine Science* **2**, 73 (2015).
26. Kohlhepp, B. *et al.* Aquifer configuration and geostructural links control the groundwater quality in thin-bedded carbonate–siliciclastic alternations of the Hainich CZE, central Germany. *Hydrol. Earth Syst. Sci.* **21**, 6091–6116 (2017).
27. Pedersen, K. & Ekendahl, S. Assimilation of CO2 and introduced organic compounds by bacterial communities in groundwater from southeastern Sweden deep crystalline bedrock. *Microb. Ecol.* **23**, 1–14 (1992).
28. Partensky, F. & Garczarek, L. Prochlorococcus: advantages and limits of minimalism. *Ann. Rev. Mar. Sci.* **2**, 305–331 (2010).
29. Karl, D. M., Hebel, D. V., Björkman, K. & Letelier, R. M. The role of dissolved organic matter release in the productivity of the oligotrophic North Pacific Ocean. *Limnol. Oceanogr.* **43**, 1270–1286 (1998).
30. Liang, Y. *et al.* Estimating Primary Production of Picophytoplankton Using the Carbon-Based Ocean Productivity Model: A Preliminary Study. *Front. Microbiol.* **8**, 1926 (2017).
31. Steinberg, D. K. *et al.* Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): a decade-scale look at ocean biology and biogeochemistry. *Deep Sea Res. Part 2 Top. Stud. Oceanogr.* **48**, 1405–1447 (2001).
32. Gundersen, K., Orcutt, K. M., Purdie, D. A., Michaels, A. F. & Knap, A. H. Particulate organic carbon mass distribution at the Bermuda Atlantic Time-series Study (BATS) site. *Deep Sea Res. Part 2 Top. Stud. Oceanogr.* **48**, 1697–1718 (2001).
33. Karl, D. M. & Lukas, R. The Hawaii Ocean Time-series (HOT) program: Background, rationale and field implementation. *Deep Sea Res. Part 2 Top. Stud. Oceanogr.* **43**, 129–156 (1996).
34. Martiny, A. C., Vrugt, J. A. & Lomas, M. W. Concentrations and ratios of particulate organic carbon, nitrogen, and phosphorus in the global ocean. *Sci Data* **1**, 140048 (2014).
35. Martiny, A. C., Vrugt, J. A. & Lomas, M. W. Data from: Concentrations and ratios of particulate organic carbon, nitrogen, and phosphorus in the global ocean. *Dryad Dataset* (2015) doi:10.5061/dryad.d702p.
36. Schwab, V. F. *et al.* 14C-Free Carbon Is a Major Contributor to Cellular Biomass in Geochemically Distinct Groundwater of Shallow Sedimentary Bedrock Aquifers. *Water Resour. Res.* **55**, 2104–2121 (2019).

583 37. Wegner, C.-E. *et al.* Biogeochemical Regimes in Shallow Aquifers Reflect the Metabolic
584 Coupling of the Elements Nitrogen, Sulfur, and Carbon. *Appl. Environ. Microbiol.* **85**,
585 (2019).
586 38. Taubert, M. *et al.* Sulfur-fueled chemolithoautotrophs replenish organic carbon inventory in
587 groundwater. *Cold Spring Harbor Laboratory* 2021.01.26.428071 (2021)
588 doi:10.1101/2021.01.26.428071.
589 39. Lin, W. *et al.* Genomic insights into the uncultured genus 'Candidatus Magnetobacterium' in
590 the phylum Nitrospirae. *ISME J.* **8**, 2463–2477 (2014).
591 40. Kato, S. *et al.* Genome-enabled metabolic reconstruction of dominant chemosynthetic
592 colonizers in deep-sea massive sulfide deposits. *Environ. Microbiol.* **20**, 862–877 (2018).
593 41. Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected
594 biogeochemical processes in an aquifer system. *Nat. Commun.* **7**, 13219 (2016).
595 42. Kojima, H., Watanabe, T. & Fukui, M. Sulfuricaulis limicola gen. nov., sp. nov., a sulfur
596 oxidizer isolated from a lake. *Int. J. Syst. Evol. Microbiol.* **66**, 266–270 (2016).
597 43. Strous, M., Van Gerven, E., Kuenen, J. G. & Jetten, M. Effects of aerobic and microaerobic
598 conditions on anaerobic ammonium-oxidizing (anammox) sludge. *Appl. Environ. Microbiol.*
599 **63**, 2446–2448 (1997).
600 44. Ji, X., Wu, Z., Sung, S. & Lee, P.-H. Metagenomics and metatranscriptomics analyses
601 reveal oxygen detoxification and mixotrophic potentials of an enriched anammox culture in
602 a continuous stirred-tank reactor. *Water Res.* **166**, 115039 (2019).
603 45. Dalsgaard, T. *et al.* Oxygen at nanomolar levels reversibly suppresses process rates and
604 gene expression in anammox and denitrification in the oxygen minimum zone off northern
605 Chile. *MBio* **5**, e01966 (2014).
606 46. Strous, M., Heijnen, J. J., Kuenen, J. G. & Jetten, M. S. M. The sequencing batch reactor
607 as a powerful tool for the study of slowly growing anaerobic ammonium-oxidizing
608 microorganisms. *Appl. Microbiol. Biotechnol.* **50**, 589–596 (1998).
609 47. Kits, K. D. *et al.* Kinetic analysis of a complete nitrifier reveals an oligotrophic lifestyle.
610 *Nature* **549**, 269–272 (2017).
611 48. Rittmann, B. E. & McCarty, P. L. *Environmental biotechnology: principles and applications*.
612 (McGraw-Hill Education, 2001).
613 49. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the
614 biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).
615 50. Lehmann, R. & Totsche, K. U. Multi-directional flow dynamics shape groundwater quality in
616 sloping bedrock strata. *J. Hydrol.* **580**, 124291 (2020).
617 51. Küsel, K. *et al.* How Deep Can Surface Signals Be Traced in the Critical Zone? Merging
618 Biodiversity with Biogeochemistry Research in a Central German Muschelkalk Landscape.
619 *Front Earth Sci. Chin.* **4**, 32 (2016).
620 52. Yan, L. *et al.* Environmental selection shapes the formation of near-surface groundwater
621 microbiomes. *Water Res.* **170**, 115341 (2019).
622 53. Pack, M. A. *et al.* A method for measuring methane oxidation rates using lowlevels of 14C-
623 labeled methane and accelerator mass spectrometry: Methane oxidation rates by AMS.
624 *Limnol. Oceanogr. Methods* **9**, 245–260 (2011).
625 54. Nielsen, E. S. The Use of Radio-active Carbon (C14) for Measuring Organic Production in
626 the Sea. *ICES J. Mar. Sci.* **18**, 117–140 (1952).

55. Xu, X. *et al.* Modifying a sealed tube zinc reduction method for preparation of AMS graphite targets: Reducing background and attaining high precision. *Nucl. Instrum. Methods Phys. Res. B* **259**, 320–329 (2007).
56. Kumar, S. *et al.* Nitrogen Loss from Pristine Carbonate-Rock Aquifers of the Hainich Critical Zone Exploratory (Germany) Is Primarily Driven by Chemolithoautotrophic Anammox Processes. *Front. Microbiol.* **8**, 1951 (2017).
57. Füssel, J. *et al.* Nitrite oxidation in the Namibian oxygen minimum zone. *ISME J.* **6**, 1200–1209 (2012).
58. McIlvin, M. R. & Altabet, M. A. Chemical conversion of nitrate and nitrite to nitrous oxide for nitrogen and oxygen isotopic analysis in freshwater and seawater. *Anal. Chem.* **77**, 5589–5595 (2005).
59. Dalsgaard, T., Thamdrup, B., Farías, L. & Revsbech, N. P. Anammox and denitrification in the oxygen minimum zone of the eastern South Pacific. *Limnol. Oceanogr.* **57**, 1331–1346 (2012).
60. Thamdrup, B. *et al.* Anaerobic ammonium oxidation in the oxygen-deficient waters off northern Chile. *Limnol. Oceanogr.* **51**, 2145–2156 (2006).
61. Taubert, M. *et al.* Tracking active groundwater microbes with D2O labelling to understand their ecosystem function. *Environ. Microbiol.* **20**, 369–384 (2018).
62. Bushnell, B. BBTools software package. *URL http://sourceforge. net/projects/bbmap* (2014).
63. Bornemann, T. L. V. *et al.* Geological degassing enhances microbial metabolism in the continental subsurface. *Cold Spring Harbor Laboratory* 2020.03.07.980714 (2020) doi:10.1101/2020.03.07.980714.
64. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
65. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
66. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357 (2012).
67. Wu, Y.-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2016).
68. Brown, C. T. *et al.* Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208–211 (2015).
69. Sieber, C. M. K. *et al.* Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* **3**, 836–843 (2018).
70. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).
71. Murat Eren, A. *et al.* Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**, e1319 (2015).
72. Aramaki, T. *et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
73. Graham, E. D., Heidelberg, J. F. & Tully, B. J. Potential for primary productivity in a globally-distributed bacterial phototroph. *ISME J.* **12**, 1861–1866 (2018).

74. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J. Mol. Biol.* **428**, 726–731 (2016).
75. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
76. Pelikan, C. *et al.* Diversity analysis of sulfite- and sulfate-reducing microorganisms by multiplex dsrA and dsrB amplicon sequencing using new primers and mock community-optimized bioinformatics. *Environ. Microbiol.* **18**, 2994–3009 (2016).
77. Lücker, S., Nowka, B., Rattei, T., Spieck, E. & Daims, H. The Genome of Nitrospina gracilis Illuminates the Metabolism and Evolution of the Major Marine Nitrite Oxidizer. *Front. Microbiol.* **4**, 27 (2013).
78. Orellana, L. H., Rodriguez-R, L. M. & Konstantinidis, K. T. ROCker: accurate detection and quantification of target genes in short-read metagenomic data sets by modeling sliding-window bitscores. *Nucleic Acids Res.* **45**, e14 (2017).
79. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2020).
80. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
81. Parks, D. H. *et al.* A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat. Biotechnol.* (2020) doi:10.1038/s41587-020-0501-8.
82. Matsen, F. A., Kodner, R. B. & Armbrust, E. V. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* **11**, 538 (2010).
83. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).
84. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS One* **5**, e9490 (2010).
85. Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
86. Ondov, B. D. *et al.* Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).
87. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).
88. Emiola, A. & Oh, J. High throughput in situ metagenomic measurement of bacterial replication at ultra-low sequencing coverage. *Nat. Commun.* **9**, 4956 (2018).
89. R Core Team. R: A Language and Environment for Statistical Computing. (2018).
90. RStudio Team. RStudio: Integrated Development Environment for R. (2016).
91. Wickham, H. *et al.* Welcome to the Tidyverse. *Journal of Open Source Software* vol. 4 1686 (2019).
92. Neuwirth, E. RColorBrewer: ColorBrewer palettes [Software]. (2014).

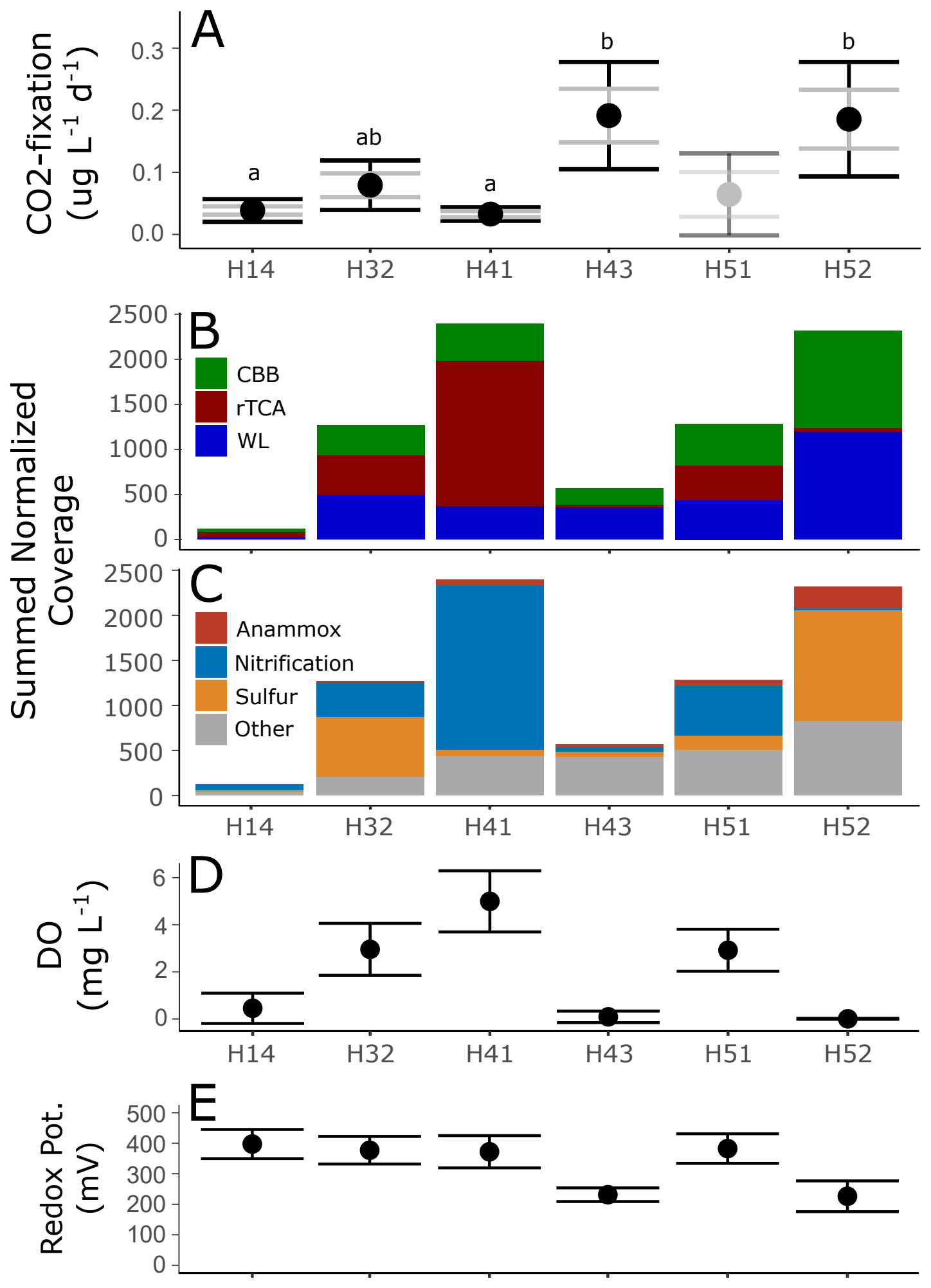

A
CO2-fixation (ug L-1 d-1)
0.3
0.2
0.1
0.0
a
ab
a
b
b
H14
H32
H41
H43
H51
H52
B
Summed Normalized Coverage
2500
2000
1500
1000
500
0
CBB
rTCA
WL
C
Anammox
Nitrification
Sulfur
Other
D
DO (mg L-1)
6
4
2
0
E
Redox Pot. (mV)
500
400
300
200
100
0

Figure 2

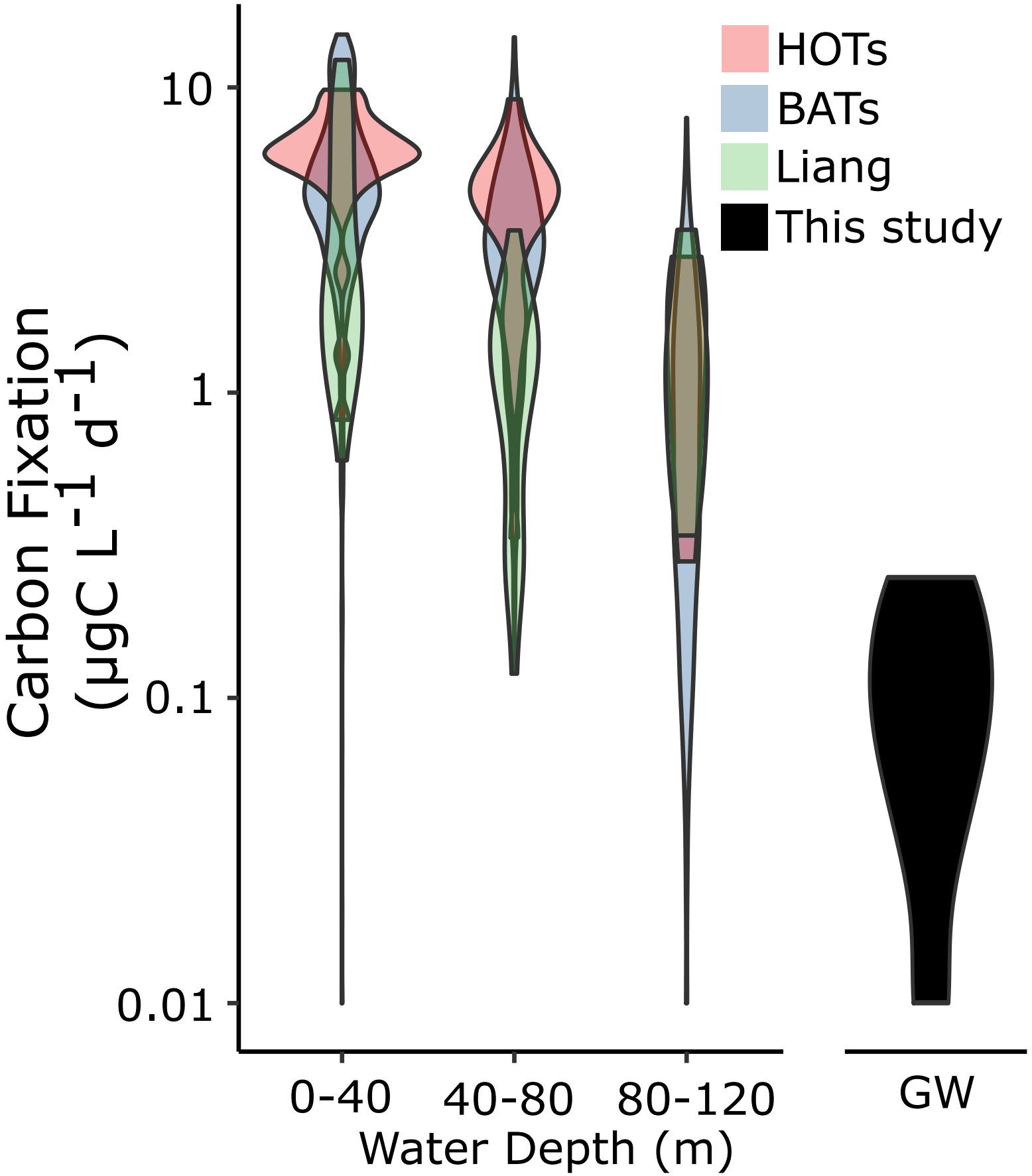

HOTs
BATs
Liang
This study
10
1
0.1
0.01
Carbon Fixation
(μgC L-1 d-1)
0-40
40-80
80-120
GW
Water Depth (m)

Figure 3

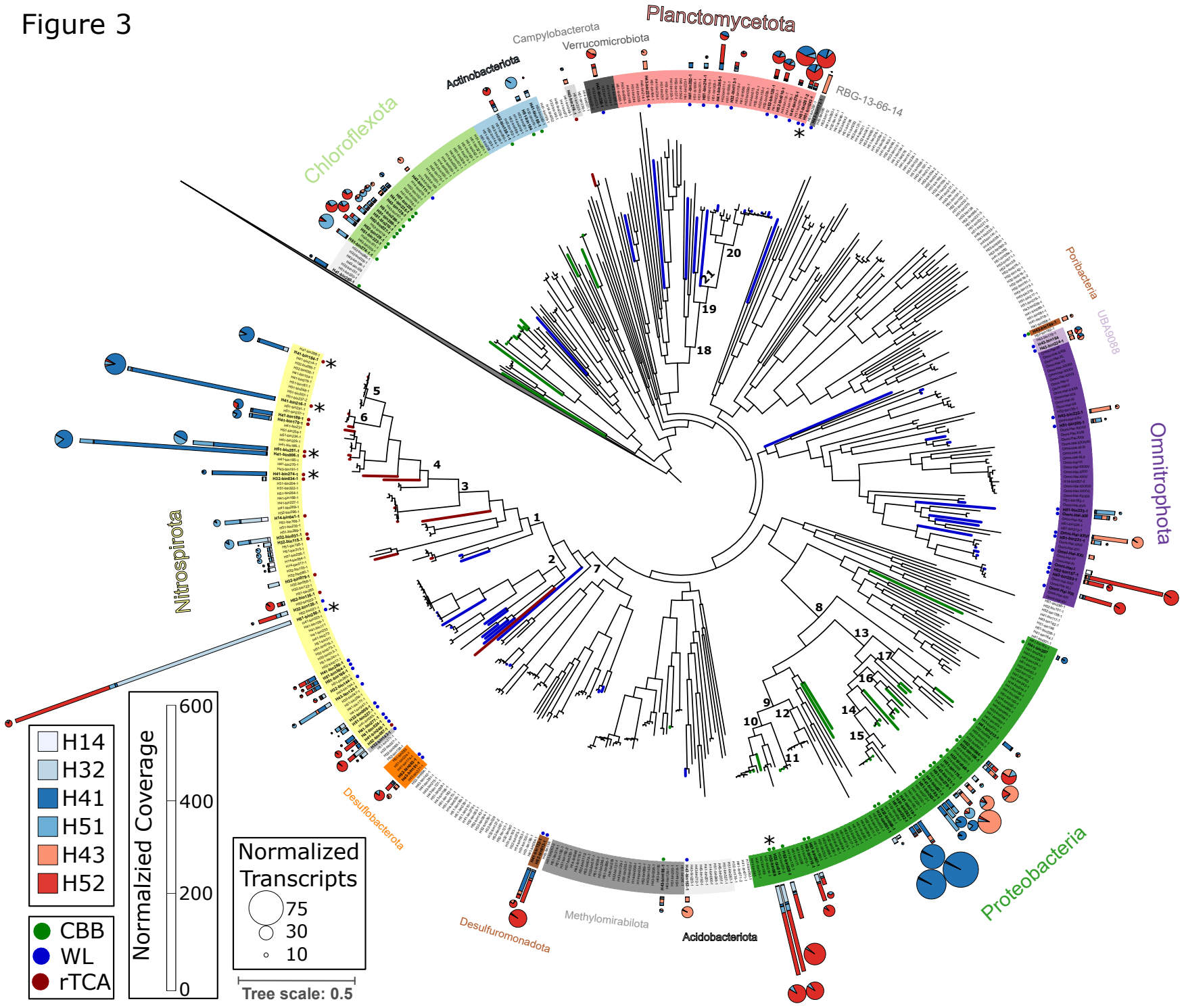
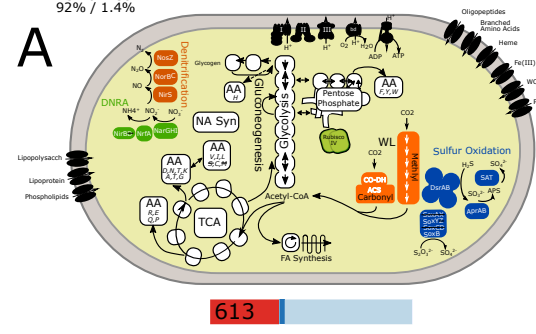
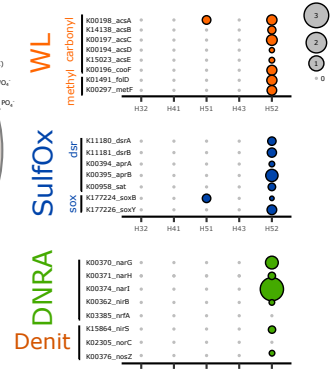

Planctomycetota
Campylobacterota
Verrucomicrobiota
Actinobacteriota
Chloroflexota
RBG-13-66-14
Poribacteria
UBA9088
Omnitrophota
Nitrospirota
Desulfobacterota
Desulfuromonadota
Methylomirabilota
Acidobacteriota
Proteobacteria
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
H14
H32
H41
H51
H43
H52
CBB
WL
rTCA
Normalzied Coverage
600
400
200
0
Normalized Transcripts
75
30
10
Tree scale: 0.5

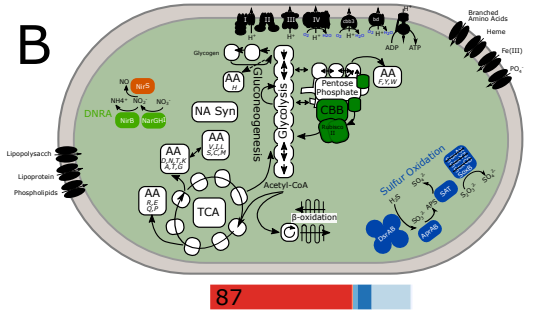
Figure 4

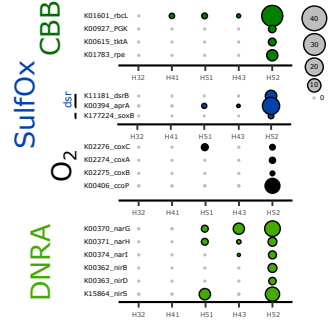
Nitrospiria 9FT-COMBO-42-15
H51-bin250-1
92% / 1.4%





Sulfurifustaceae
H32-bin014
38% / 1.4%





Planctomycetota Brocadiaceae
H51-bin243-1
85% / 0%