



OPEN

Carbon fixation rates in groundwater similar to those in oligotrophic marine systems

Will A. Overholt¹, Susan Trumbore^{2,3}, Xiaomei Xu³, Till L. V. Bornemann⁴, Alexander J. Probst⁴, Markus Krüger¹, Martina Herrmann^{1,5}, Bo Thamdrup⁶, Laura A. Bristow⁶, Martin Taubert¹, Valérie F. Schwab², Martin Hölzer^{7,8,10}, Manja Marz^{7,8,9} and Kirsten Küsel^{1,5}✉

The terrestrial subsurface contains nearly all of Earth's freshwater reserves and harbours the majority of our planet's total prokaryotic biomass. Although genetic surveys suggest these organisms rely on in situ carbon fixation, rather than the photosynthetically derived organic carbon transported from surface environments, direct measurements of carbon fixation in the subsurface are absent. Using an ultra-low level ¹⁴C-labelling technique, we estimate in situ carbon fixation rates in a carbonate aquifer. We find these rates are similar to those measured in oligotrophic marine surface waters and up to six-fold greater than those observed in the lower euphotic zone. Our empirical carbon fixation rates agree with nitrification rate data. Metagenomic analyses reveal abundant putative chemolithoautotrophic members of an uncharacterized order of Nitrospira that may be behind the carbon fixation. On the basis of our determined carbon fixation rates, we conservatively extrapolate global primary production in carbonate groundwaters (10% of global reserves) to be 0.11 Pg carbon per year. These rates fall within the range found for oligotrophic marine surface waters, indicating a substantial contribution of in situ primary production to subsurface ecosystem processes. We further suggest that, just as phototrophs are for marine biogeochemical cycling, such subsurface carbon fixation is potentially foundational to subsurface trophic webs.

The continental subsurface is the planet's largest carbon reservoir¹, housing up to 19% of its total biomass^{2,3} and 99% of its freshwater⁴. Despite accounting for only 6% of total stores, modern groundwater (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 water⁵. Unfortunately, modern groundwater is also the most vulnerable to anthropogenic and climatic impacts⁴. While subsurface ecosystems have long fascinated ecologists⁶, and more recently microbiologists⁷, accessibility, enormous spatial heterogeneity and challenges in interpreting process rate measurements have obscured a meaningful understanding of their contributions to global biogeochemical cycles⁸.

The widespread recognition that Earth's biosphere extends deep into the subsurface occurred only recently⁹. Historically, carbon supply in such environments was thought to be limited to the trickling of surface-produced organic matter into the shallow subsurface¹⁰ or what was stored within sedimentary rocks¹¹. By stark contrast, a wealth of compelling genetic evidence suggests that in situ carbon fixation is critical for sustaining highly diverse microbial metabolic networks in groundwater, in both the shallow and the deep subsurface^{12–19}. 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 CO₂ fixation rates resemble marine waters

Primary productivity in the shallow subsurface (groundwater wells 5–90 m deep), experimental carbon fixation rates varied from 0.043 ± 0.01 to $0.23 \pm 0.10 \mu\text{gC l}^{-1} \text{d}^{-1}$ (mean ± 1 standard deviation (s.d.); Fig. 1a, Extended Data Table 1 and Supplementary Information). The ultra-low level ¹⁴C-labelling 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 ¹⁴C-labelling approaches. Rates resulting from our new labelling technique probably approximate net primary productivity rather than gross productivity, as has been reported for marine systems^{20,21}, and we further expect them to be conservative estimates for carbon fixation as they consider only contributions from the planktonic portion of the community (Supplementary Information).

We compared these carbon fixation rates that were measured in groundwater of varying biogeochemical characteristics²² with the only other subsurface ¹⁴CO₂ assimilation measurements reported: those of a deep (830–1,078 m) groundwater borehole from crystalline

¹Aquatic Geomicrobiology, Institute of Biodiversity, Friedrich Schiller University, Jena, Germany. ²Max Planck Institute for Biogeochemistry, Jena, Germany.

³Department of Earth System Sciences, University of California, Irvine, CA, USA. ⁴Institute for Environmental Microbiology and Biotechnology, Department of Chemistry, University Duisburg-Essen, Duisburg-Essen, Germany. ⁵German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany. ⁶Department of Biology, University of Southern Denmark, Odense, Denmark. ⁷RNA Bioinformatics and High-Throughput Analysis, Faculty of Mathematics and Computer Science, Friedrich Schiller University, Jena, Germany. ⁸European Virus Bioinformatics Center, Friedrich Schiller University, Jena, Germany. ⁹FLI Leibniz Institute for Age Research, Marz Associated Research Group, Jena, Germany. ¹⁰Present address: Methodology and Research Infrastructure, MFI Bioinformatics, Robert Koch Institute, Berlin, Germany. ✉e-mail: kirsten.kuesel@uni-jena.de

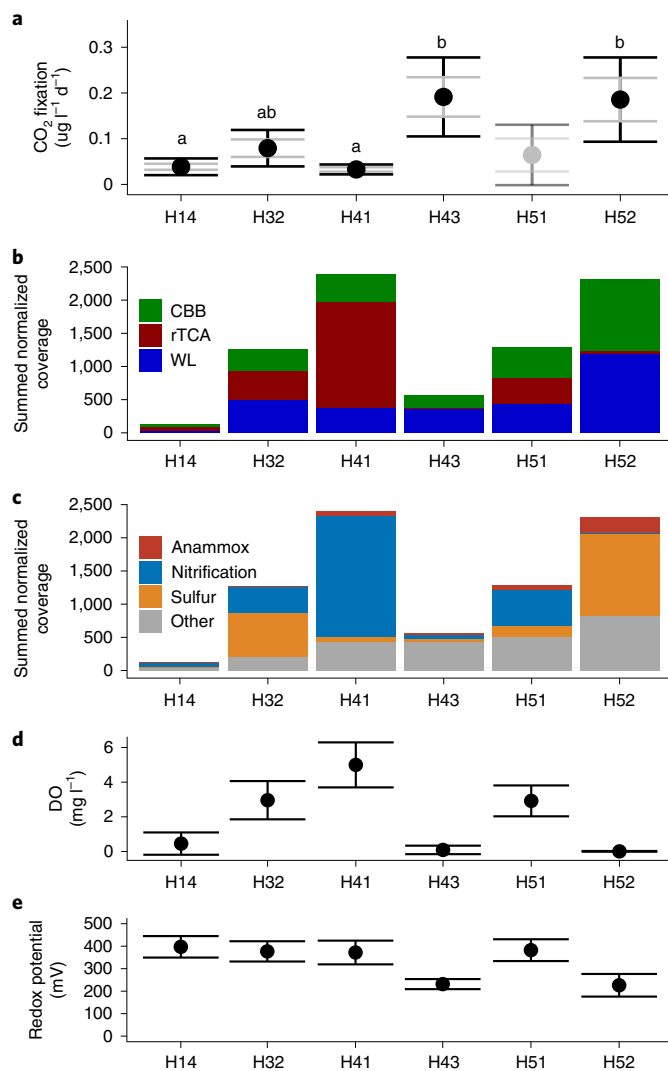


Fig. 1 | Carbon fixation rates and metagenomic potentials across the aquifer. **a**, Rates of carbon fixation. Outer error bars depict one standard deviation; inner grey bars delineate standard error of the mean. Rates for well H51 are derived from non-labelled controls (Supplementary Information). Letters denote the results of ANOVA and post hoc Tukey tests (excluding H51). **b**, Relative importance of the predicted carbon fixation pathways. **c**, Electron donor sources in each well. Values are averages from the triplicate 0.2 µm filtered fraction metagenome samples. **d**, Mean dissolved oxygen (DO) concentrations in groundwater collected in summer months (May–September 2010–2018); error bars depict standard deviation. **e**, Redox-potential measurements from identical time points; error bars depict standard deviation. While groundwater from H52 and H43 exhibited anoxic or hypoxic conditions, the positive redox potentials were due to oligotrophic conditions and mean nitrate concentrations of -4 mg l^{-1} .

bedrock in Sweden²³. To do so, we converted the published rates of isotopic incorporation to carbon equivalents, revealing the lower but overlapping range of 0.0095 to $0.0560 \text{ µg C l}^{-1} \text{ d}^{-1}$.

To better understand the relevance of the rates measured, we compared them with those of well-documented oligotrophic marine surface waters. Unlike our samples, the carbon fixed in these waters was sourced almost entirely by bacterial photoautotrophs^{24,25}. When compared directly with a comprehensive dataset compiled by ref. ²⁶, 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 ($2.65 \text{ µg l}^{-1} \text{ d}^{-1}$, interquartile range (IQR) = 1.74, 6.02) and 20% of the median for 20–140 m depths ($1.2 \text{ µg l}^{-1} \text{ d}^{-1}$, IQR = 0.6, 1.7). Comparisons with the extensively studied Sargasso Sea in the Bermuda Atlantic Timeseries Study (BATS)^{27,28} and the Hawaiian Oceanographic Timeseries (HOTs)²⁹ datasets yielded similar findings (Fig. 2). Our rate measurements ranged between 3% and 23% of the median reported net primary productivity in the upper euphotic zones (down to ~ 120 m) and between 20% and 600% of the median of the lower euphotic region (100–120 m).

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 fixed carbon per bacterial cell per day (Extended Data Table 2), 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 $0.47\% \pm 0.22\%$ (mean \pm s.d.; Extended Data Table 2) of its existing POC, much lower than the marine system's $2.6\% \pm 2.9\%$ gain in the lower euphotic zone and $22\% \pm 18\%$ at the surface^{30,31}. This disparity might stem from the larger recalcitrant fraction of POC in groundwater compared with oligotrophic oceans, which is supported by deviations in ^{14}C and ^{13}C signatures of total groundwater POC concentrations compared with lipid signatures of resident microbes³².

An ecosystem dominated by chemolithoautotrophs

To identify dominant microbial primary producers, a dereplicated and quality-controlled set of 1,224 metagenome-assembled genomes (MAGs) were generated from groundwater samples. 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 and Supplementary Dataset 1). In some samples, up to 12% of metagenomic reads from a sample could be recruited to these potentially chemolithoautotrophic MAGs (Supplementary Information and Extended Data Figs. 1 and 2). A single archaeal MAG of the family Nitrosopumilaceae encoded

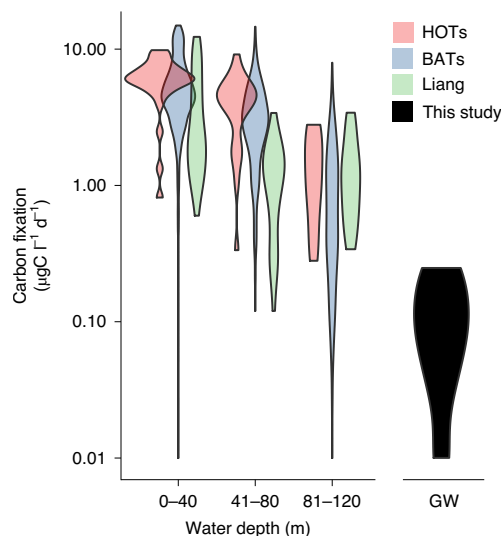


Fig. 2 | Comparison of carbon fixation rates within groundwater and the marine euphotic zone. Violin plots depicting the distribution of carbon fixation rates measured in oligotrophic marine surface waters and groundwater. HOTS, Hawaiian Oceanographic Timeseries (1999, cruises 101–110); BATS, data from 1988 to 2016 for the Bermuda Atlantic Timeseries; Liang, a collated dataset compiled by Liang et al.²⁶; GW, the range of groundwater samples shown in Fig. 1.

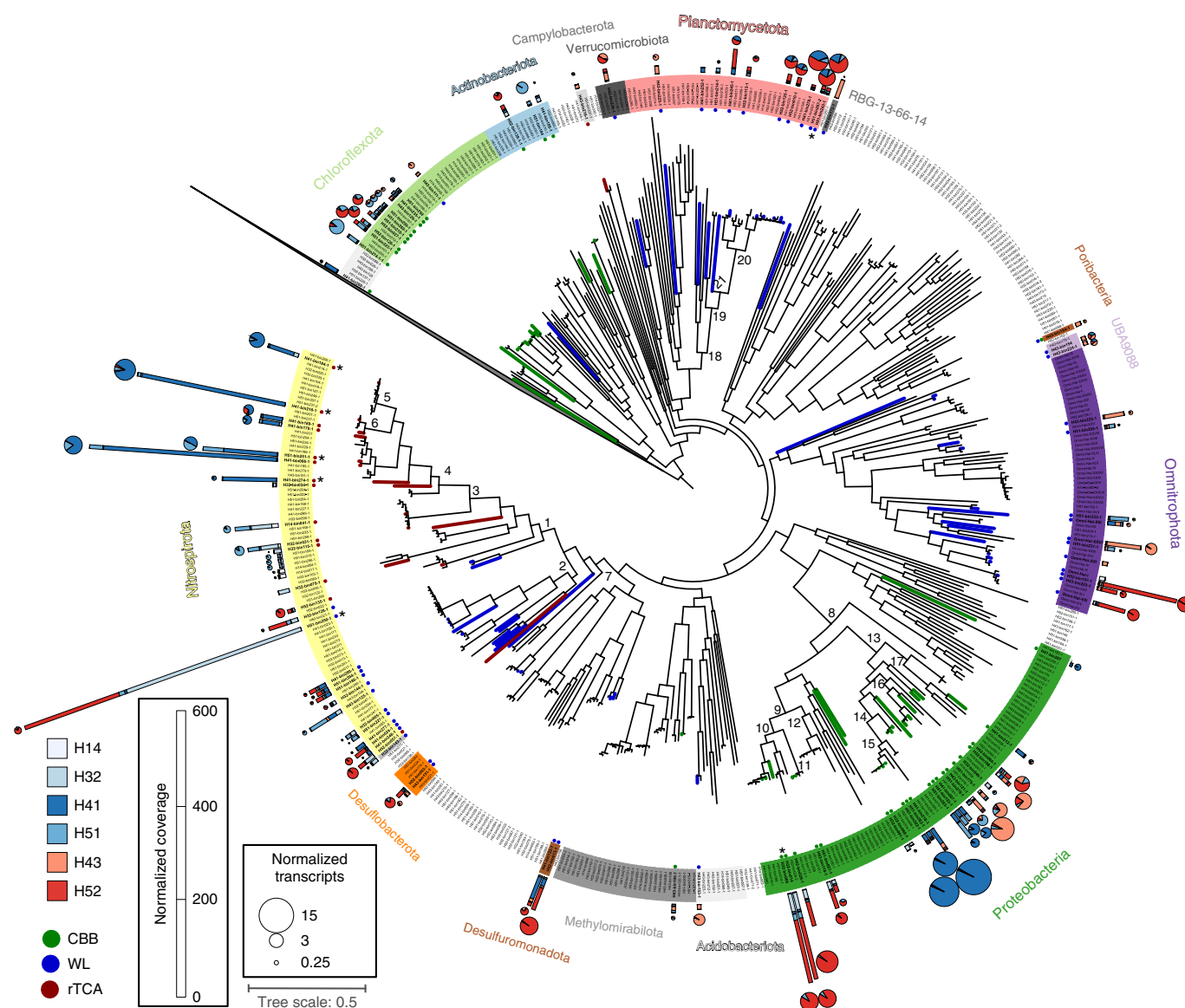


Fig. 3 | Phylogeny, relative abundances and transcriptional activities of putative chemolithoautotrophic MAGs. Approximately maximum-likelihood phylogenetic tree based on concatenated single-copy protein alignments for all bacterial MAGs considered. Branches are coloured according to the predicted carbon 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. Asterisks denote MAGs discussed in greater detail in Supplementary Information. The tree is rooted using Patescibacteria (CPR) as an outgroup, indicated by the collapsed grey leaf in the upper left. Node numbers: (1) c_Nitrospiria, (2) c_Thermodesulfobivibronia, (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_Brocadiaceae, (19) o_Brocadiales, (20) f_Brocadiaceae and (21) f_Scalinduaceae.

gene products for the 4-hydroxybutyrate/3-hydroxypropionate pathway and was not relatively abundant (<5 \times maximum normalized coverage).

Three chemolithoautotrophic pathways were detected (Fig. 1b): the Calvin–Benson–Bassham (CBB), Wood–Ljungdahl (WL) and reverse tricarboxylic acid (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 H41 and H43 (Fig. 1, Extended Data Fig. 2 and Supplementary Information). The greatest relative abundances of predicted 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 putative anaerobic ammonium-oxidizing (anammox (10%)) autotrophic microbes, while oxic groundwaters harboured greater abundances of potential nitrifiers (76%; Fig. 1c and Supplementary Information).

Uncharacterized microbes influence CO₂ 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

(Supplementary Information). 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 transcripts 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 deoxyribonucleic acid- (DNA-) based abundances, transcript data revealed far more restrictive ranges in which specific gene products were favoured (Fig. 3). Given their metabolic versatility and the results of previous cultivation-based analyses³³, these populations are expected to be mixotrophic (capable of supplementing carbon requirements with available organic matter). Overall, carbon fixation in anoxic groundwater was predicted to be fuelled by reduced sulfur, and there were three highly abundant, putative sulfur-oxidizing MAGs identified, each accounting for >2% of the total metagenomic reads in some samples (100–400× normalized coverages). Diverse reduced sulfur species fuelling these metabolisms are released from pyrite weathering of the karst rock³⁴.

The most abundant MAG (H51-bin250-1) encountered in this study belongs to a deep-branching order, 9FT-COMBO-4₂-15, of class Nitrospira and is the first representative of class Nitrospira thought to fix carbon via the WL pathway (Fig. 3, Fig. 4a and Supplementary Information). As there is precedence for autotrophic WL-utilizing bacteria within phylum Nitrospirota, and *Candidatus Magnetobacterium* was characterized with an equally flexible metabolism, these traits may be more widespread within the phylum than previously thought³⁵. In addition, two MAGs with the potential to couple sulfur oxidation to carbon fixation via the CBB cycle were identified as members of the Sulfurifustaceae family of Proteobacteria (Supplementary Information and Fig. 4b; H3₂-bin014, H3₂-bin069). These MAGs recruited tenfold more transcripts than their Nitrospirota counterparts and were among the most transcriptionally active putative chemolithoautotrophic genomes detected (Figs. 3 and 4, Extended Data Fig. 4 and Supplementary Information). 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 Genome Taxonomy Database (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^{36–38}.

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 (Figs. 3 and 4c, Extended Data Fig. 5 and Supplementary Information). The elevated transcriptional activity of gene products within the CBB and WL pathways suggests that these taxa play a disproportionately large role in chemolithoautotrophy relative to their DNA-based abundances. Surprisingly, all putative anammox MAGs detected were transcriptionally active inoxic groundwater (Figs. 3 and 4b; wells H41 and H51). Anammox reactions are typically inhibited in the presence of oxygen³⁹, although microbes will still express critically important genes in low oxygen environments^{40,41}.

N-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 harbouring the greatest relative abundance of putative anaerobic

ammonium-oxidizing bacteria (well H52). Here, anammox rates of $1.2 \pm 0.5 \text{ nmol l}^{-1} \text{ d}^{-1} \text{ N}_2$ were measured, similar to rates in another freshwater aquifer⁴². Empirical stoichiometric data demonstrate that 1.02 moles of N_2 is produced via anaerobic ammonium oxidation for every 0.066 moles of $\text{CH}_2\text{O}_{0.5}\text{N}_{0.15}$ reduced to biomass⁴³. Assuming equivalent stoichiometry, the rate of carbon fixation via anammox in groundwater would be $0.93 \pm 0.39 \text{ ngCl}^{-1} \text{ d}^{-1}$, more than 200 times lower than the $220 \text{ ngCl}^{-1} \text{ d}^{-1}$ 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 by reduced nitrogen.

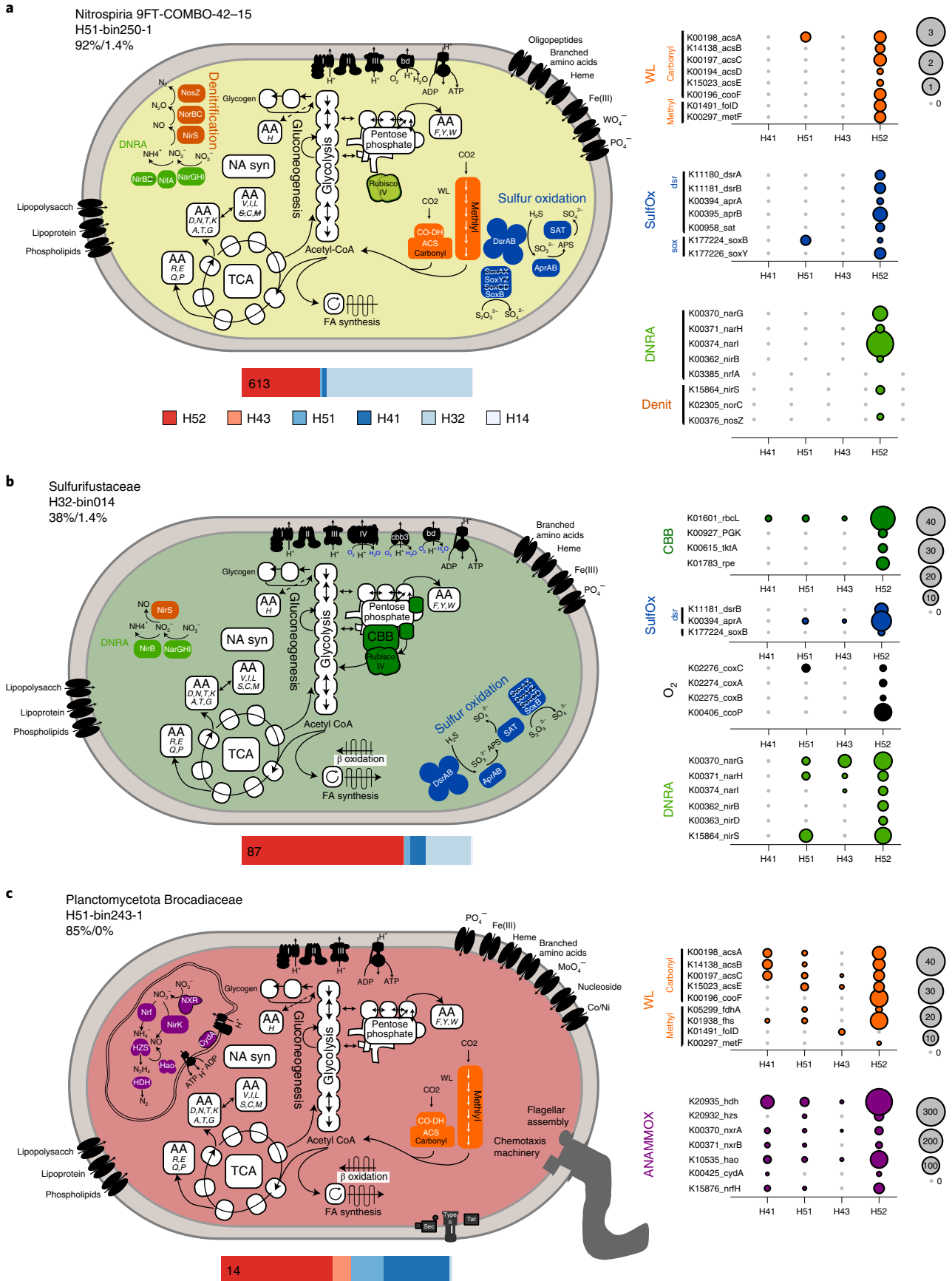
Metagenomic and metatranscriptomic data predicted that nearly all the organic carbon produced under oxic conditions in well H41 would be coupled to nitrification. To test this, we monitored the rate of aerobic ammonium oxidation in this well and recorded a mean production of $125.8 \pm 5.9 \text{ nmol NO}_2^- + \text{NO}_3^- \text{ l}^{-1} \text{ d}^{-1}$. Since the most abundant nitrifiers detected were most closely related to complete ammonium-oxidizing bacteria (Supplementary Information), we based our calculations on the 394 mg protein per mol of ammonia growth yields of *Nitrospira inopinata*, a comammox organism⁴⁴. Assuming a cellular composition of $\text{C}_5\text{H}_7\text{O}_2\text{N}$ (ref. 45) and 55% protein content, we estimated a rate of $48.5 \pm 1.9 \text{ ngCl}^{-1} \text{ d}^{-1}$, which was well within the range of error for our measured rate of $43 \pm 13 \text{ ngCl}^{-1} \text{ d}^{-1}$ and confirms the importance of nitrification for carbon fixation at this site. Furthermore, the stoichiometry determined for oligotrophic marine rTCA nitrifiers⁴⁶ of 0.0216 molC/molN matched our calculated ratio of 0.0276 ± 0.0084 , indicating they are responsible for most of the fixed carbon.

Global estimates for groundwater primary productivity

There are an estimated 22.6 million km^3 of groundwater on Earth⁴, 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 $0.108 \pm 0.069 \text{ PgC}$ (mean \pm s.d.) is fixed every year in this global ecosystem (Extended Data Table 3). If the values reported from crystalline aquifers²³ are representative of this environment, then another $0.15 \pm 0.11 \text{ PgC}$ would be fixed there annually. Collectively, the net primary productivity of ~66% of the planet's groundwater reservoirs would total 0.26 PgC yr^{-1} , approximately 0.5% that of marine systems and 0.25% of global NPP estimates⁴⁷. As these projections exclude the missing contributions from groundwaters within siliciclastic and volcanic geologic settings and activities of attached microorganisms, global contributions to the carbon cycle are expected to be many-fold higher.

We showed that conservative estimates of 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 lower euphotic zone. Within oxic groundwaters, our carbon fixation method was independently validated by nitrification rate measurements. Normalizing carbon fixation rates by estimated bacterial numbers revealed equivalent carbon input (0.3–12 fgC per cell) for both marine and groundwater systems, despite the fact that daily inputs of new POC were 40 times greater in marine waters. This disparity makes sense since trophic webs are simpler in the subsurface, and the export of organic matter is constrained by long water residence times within the aquifer. Complementary metagenomic analyses revealed that groundwater carbon fixation is not dominated by a single functional guild but rather has contributions

Fig. 4 | Metabolic reconstructions of dominant putatively chemolithoautotrophic MAGs. Bar charts below each metabolic model summarize the average normalized coverage across each sample, scaled proportionally. Values within the bar chart 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. The text information over each panel includes the predicted taxonomy (**a**, Nitrospira; **b**, Sulfurifustaceae; **c**, Brocadiaaceae), MAG identifier and estimated % completion/% redundancy. DNRA, dissimilatory nitrate reduction to ammonia.



from diverse pathways and versatile microorganisms that are setting specific. As the 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. Indeed, subsurface primary producers need to be considered as important to ecosystem processes as marine phototrophs are known to be in the surface ocean. 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.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41561-022-00968-5>.

Received: 2 August 2021; Accepted: 17 May 2022;
Published online: 30 June 2022

References

- Falkowski, P. et al. The global carbon cycle: a test of our knowledge of Earth as a system. *Science* **290**, 291–296 (2000).
- McMahon, S. & Parnell, J. Weighing the deep continental biosphere. *FEMS Microbiol. Ecol.* **87**, 113–120 (2014).
- Magnabosco, C. et al. The biomass and biodiversity of the continental subsurface. *Nat. Geosci.* **11**, 707–717 (2018).
- 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).
- Stevanović, Z. Karst waters in potable water supply: a global scale overview. *Environ. Earth Sci.* **78**, 662 (2019).
- Poulson, T. L. & White, W. B. The cave environment. *Science* **165**, 971–981 (1969).
- Rusterholtz, K. J. & Mallory, L. M. Density, activity, and diversity of bacteria indigenous to a karstic aquifer. *Microb. Ecol.* **28**, 79–99 (1994).
- Smith, H. J. et al. Impact of hydrologic boundaries on microbial planktonic and biofilm communities in shallow terrestrial subsurface environments. *FEMS Microbiol. Ecol.* **94**, fuy191 (2018).
- Alexander, M. *Introduction to Soil Microbiology* (Wiley, 1977).
- Griebler, C. & Lueders, T. Microbial biodiversity in groundwater ecosystems. *Freshw. Biol.* **54**, 649–677 (2009).
- Krumholz, L. R., McKinley, J. P., Ulrich, G. A. & Sufliata, J. M. Confined subsurface microbial communities in Cretaceous rock. *Nature* **386**, 64–66 (1997).
- 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).
- Magnabosco, C. et al. A metagenomic window into carbon metabolism at 3 km depth in Precambrian continental crust. *ISME J.* **10**, 730–741 (2016).
- Stevens, T. O. & McKinley, J. P. Lithoautotrophic microbial ecosystems in deep basalt aquifers. *Science* **270**, 450–455 (1995).
- Tiago, I. & Verissimo, A. Microbial and functional diversity of a subterranean high pH groundwater associated to serpentinization. *Environ. Microbiol.* **15**, 1687–1706 (2013).
- Mccollom, T. M. & Amend, J. P. A thermodynamic assessment of energy requirements for biomass synthesis by chemolithoautotrophic microorganisms in oxic and anoxic environments. *Geobiology* **3**, 135–144 (2005).
- 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).
- 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).
- Herrmann, M., Rusznayk, 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. *Appl. Environ. Microbiol.* **81**, 2384–2394 (2015).
- 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).
- 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. *Front. Mar. Sci.* **2**, 73 (2015).
- 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).
- Pedersen, K. & Ekdahl, 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).
- Partensky, F. & Garczarek, L. Prochlorococcus: advantages and limits of minimalism. *Ann. Rev. Mar. Sci.* **2**, 305–331 (2010).
- 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).
- Liang, Y. et al. Estimating primary production of picophytoplankton using the carbon-based ocean productivity model: a preliminary study. *Front. Microbiol.* **8**, 1926 (2017).
- 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.* **2** **48**, 1405–1447 (2001).
- 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.* **2** **48**, 1697–1718 (2001).
- Karl, D. M. & Lukas, R. The Hawaii Ocean Time-series (HOT) program: background, rationale and field implementation. *Deep Sea Res.* **2** **43**, 129–156 (1996).
- 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).
- 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* <https://doi.org/10.5061/dryad.d702p> (2015).
- 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).
- Taubert, M. et al. Bolstering fitness via CO₂ fixation and organic carbon uptake: mixotrophs in modern groundwater. *ISME J.* **16**, 1153–1162 (2022).
- Rimstidt, J. D. & Vaughan, D. J. Pyrite oxidation: a state-of-the-art assessment of the reaction mechanism. *Geochim. Cosmochim. Acta* **67**, 873–880 (2003).
- Lin, W. et al. Genomic insights into the uncultured genus “*Candidatus Magnetobacterium*” in the phylum Nitrospirae. *ISME J.* **8**, 2463–2477 (2014).
- Kato, S. et al. Genome-enabled metabolic reconstruction of dominant chemosynthetic colonizers in deep-sea massive sulfide deposits. *Environ. Microbiol.* **20**, 862–877 (2018).
- Anantharaman, K. et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* **7**, 13219 (2016).
- Kojima, H., Watanabe, T. & Fukui, M. *Sulfuricaulis limicola* gen. nov., sp. nov., a sulfur oxidizer isolated from a lake. *Int. J. Syst. Evol. Microbiol.* **66**, 266–270 (2016).
- Strous, M., Van Gerven, E., Kuenen, J. G. & Jetten, M. Effects of aerobic and microaerobic conditions on anaerobic ammonium-oxidizing (anammox) sludge. *Appl. Environ. Microbiol.* **63**, 2446–2448 (1997).
- Ji, X., Wu, Z., Sung, S. & Lee, P.-H. Metagenomics and metatranscriptomics analyses reveal oxygen detoxification and mixotrophic potentials of an enriched anammox culture in a continuous stirred-tank reactor. *Water Res.* **166**, 115039 (2019).
- Dalsgaard, T. et al. Oxygen at nanomolar levels reversibly suppresses process rates and gene expression in anammox and denitrification in the oxygen minimum zone off northern Chile. *mBio* **5**, e01966 (2014).
- Smith, R. L., Böhlke, J. K., Song, B. & Tobias, C. R. Role of anaerobic ammonium oxidation (anammox) in nitrogen removal from a freshwater aquifer. *Environ. Sci. Technol.* **49**, 12169–12177 (2015).
- Strous, M., Heijnen, J. J., Kuenen, J. G. & Jetten, M. S. M. The sequencing batch reactor as a powerful tool for the study of slowly growing anaerobic ammonium-oxidizing microorganisms. *Appl. Microbiol. Biotechnol.* **50**, 589–596 (1998).
- Kits, K. D. et al. Kinetic analysis of a complete nitrifier reveals an oligotrophic lifestyle. *Nature* **549**, 269–272 (2017).
- Rittmann, B. E. & McCarty, P. L. *Environmental Biotechnology: Principles and Applications* (McGraw-Hill Education, 2001).
- Zhang, Y. et al. Nitrifier adaptation to low energy flux controls inventory of reduced nitrogen in the dark ocean. *Proc. Natl. Acad. Sci. USA* **117**, 4823–4830 (2020).
- Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other

third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

Methods

Site description. Groundwater samples were sampled from the Hainich Critical Zone Exploratory (NW Thuringia, Germany)^{22,48,49}. This aquifer assemblage consists of a multistage fractured system composed of alternating layers of limestone and mudstone that developed along a hillslope of Upper Muschelkalk bedrock²². 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 lie within the Meissner Formation (moM) and are 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 represents a distinct regime. Consistent with previous microbiological characterizations, however, each well studied represented a distinct community state⁵⁰.

¹⁴C-DIC incorporation assay. This method, similar to a sensitive methane oxidation technique previously described⁵¹, is a modification of traditional ¹⁴C-CO₂ primary productivity approaches⁵² predicated on the sensitivity offered by accelerator-based mass spectrometry. Groundwater was collected in July 2020 during sampling campaign PNK130, as described by ref. 19. After approximately three well volumes had been discharged and physicochemical parameters stabilized, groundwater was collected directly into nine pre-sterilized 2 l borosilicate bottles, from the bottom up. Bottles were then overfilled with greater than 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-NaHCO₃ that contained 200 Bq of activity (50 μCi; American Radiolabeled Chemicals) diluted to 9.38 Bq μl⁻¹ with sterilized milliQ water, adjusted to pH 10 and verified using a scintillation counter. An advantage of this ¹⁴C 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 such as pH that could affect microbial populations. Kill controls were prepared in the same way, except 10 ml 50% ZnCl₂ (w/v; final concentration 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 3 M HCl, bubbled with N₂ for one hour to remove DIC and then filtered through pre-baked (550 °C, eight hours) quartz fibres (47 mm, 0.3 μm 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 °C for two hours. Evolved CO₂ 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 facility⁵³. From the label incorporation and amount of carbon retained on the filters (Supplementary Data File 2), fixation rates were calculated using equation (1):

The technical variation was at most 3.6% (median = 0.78%) of the biological variation for the ¹⁴C measurements and was not considered in standard error of the mean calculations. Standard error of the mean was determined for both the ¹⁴C-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 honestly significant difference (HSD) tests were conducted on resulting summary statistics (mean ± s.e.m.) using the following utility⁵⁴. All ¹⁴C enrichment values were calculated using the differences between the 200 Bq-labelled samples and the 200 Bq-labelled kill controls. Rates calculated on the basis of no-label addition controls are presented in Extended Data Table 1. Data from global oligotrophic marine systems were included from Supplementary Data Sheet 1²⁶, the Bermuda Atlantic Timeseries years 1988–2016 via FTP^{27,55} and the Hawaiian Oceanographic Timeseries via FTP⁵⁶. POC data from both sites were extracted from Dryad datasets generated by refs. 30,31. Bacterial cell number estimates for Hawaiian Oceanographic Timeseries were obtained from the FTP site⁵⁷.

¹⁵N-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 hours post-collection).

For nitrification measurements, 10 ml was removed from each sampling bottle (total volume 0.5 l) and replaced with N₂ to analyse inorganic nitrogen and pH. Groundwater from control bottles was sterile filtered through a 0.2 μm filter (Supor, Pall Corporation). Sterile filtered ¹⁵N ammonium sulfate solution (98%, Cambridge Isotope Laboratories), serving as a substrate for ammonia-oxidizing prokaryotes, was then added to a final concentration of 50 μM. Samples were incubated at 15 °C in the dark sans agitation for five days. Ten-millilitre fractions were removed and replaced with N₂ at the outset of the experiment and after 12, 24, 48, 70 and 120 hours via filtration through 0.2 μm filters; these fractions were stored at -20 °C for isotopic ratio mass spectrometry 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 volume 1 l) were flushed with N₂ under sterile conditions for 30 minutes to remove all remnants of oxygen. Five-millilitre fractions were removed and replaced with N₂ from each sample (and control) bottle to assess background ¹⁴NH₄⁺ concentrations. Subsequently, samples were spiked with either (1) 50 μM ¹⁵NH₄⁺ + 5 μM ¹⁴NO₂⁻ or (2) 5 μM ¹⁵NO₂⁻ as previously described⁵⁸. Control bottles, serving as abiotic controls, were sterile filtered (0.2 μm filters; Supor, Pall Corporation) before 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 and 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 50% (v/w) aqueous zinc chloride solution.

Nitrification rates were determined on the basis of ¹⁵NO₂⁻ + ¹⁵NO₃⁻ production in incubations with ¹⁵NH₄⁺, ¹⁵NO₂⁻ and ¹⁵NO₃⁻ were converted to N₂ via cadmium reduction followed by a sulfamic acid addition^{59,60}. The N₂ produced (¹⁴N¹⁵N and ¹⁵N¹⁵N) was analysed on a gas chromatography isotope ratio mass spectrometer as previously described⁶¹. Rates were evaluated from the slope of the linear regression of ¹⁵N produced with time and corrected for the fraction of the NH₄⁺ pool labelled in the initial substrate pool. The production of ¹⁵N-labelled N₂ from anammox was analysed on the same isotopic ratio mass spectrometer as for nitrification rates and calculated as described⁶². Note, denitrification was not detected in any of the ¹⁵NO₂⁻ incubations. *T* tests were applied (*P* < 0.05) to assess whether rates were significantly different from zero (Extended Data Fig. 3).

DNA extraction and 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 polytetrafluoroethylene (PTFE) filters (142 mm, Omnipore Membrane, Merck Millipore; Supplementary Data File 3). With the exception of H32 (which 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 before extraction. DNA was extracted using a phenol-chloroform-based method, as previously described⁶³, 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 (Supplementary Data File 3). Sequencing of the 32 samples was performed at the Core DNA Sequencing Facility of the Fritz Lipmann Institute using an Illumina NextSeq 500 system (2 × 150 bp). Resulting metagenomic library sizes ranged from 16.4 to 22.1 Gbp (mean = 19.6 Gbp; Supplementary Data File 3), and raw data were deposited into the European Nucleotide Archive under project PRJEB36523.

Metagenomic assembly and binning. Adaptors were trimmed and raw sequences subjected to quality control processing using BBduk v.38.51⁶⁴. Assembly and binning were performed as previously described⁶⁵. Briefly, all libraries were independently assembled into scaffolds using metaSPAdes v.3.12⁶⁶, all of which were taxonomically classified per ref. 65. For individual assemblies, open reading frames (ORFs) were identified using Prodigal v.2.6.3 in meta mode⁶⁷. 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 v.2.3.4.3 in the sensitive mode⁶⁸.

Scaffolds were binned using differential coverages and tetranucleotide frequencies with Maxbin2⁶⁹. In addition, ESOM and abawaca⁷⁰ 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 Tool⁷¹ was used with default parameters to reconcile resulting bin sets. Complete sets of bins from each of the samples were dereplicated using dRep v.2.4.0⁷². All scaffolds, bin assignments, ORF predictions and taxonomic annotations were then imported into Anvi'o v.6.0⁷³. Each of the resulting 1,275 bins was manually curated in Anvi'o v.6, 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 MAGs. ORFs originating from all of the resulting MAGs were annotated using kofamsan⁷⁴ 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 kofamsan 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 kofamsan thresholds; that is, we favoured false positives over false negatives since we included a secondary verification step.

KEGGDecoder⁷⁵ was used to assess the metabolic potential of five of the primary chemolithoautotrophic pathways: the CBB cycle, the WL pathway, the reverse citric acid cycle, the 4-hydroxybutyrate 3-hydroxypropionate pathway and the 3-hydroxypropionate bi-cycle. MAGs were examined in greater depth if a given pathway was >50% complete. The MAGs representing potential chemolithoautotrophs were re-annotated using the online BlastKoala server⁷⁶ with essential steps verified through blast⁷⁷ against the RefSeq database. A collection of HMM models was used to determine which form of Rubisco was detected, along with potential hydrogenases⁷⁷. Using blastp⁷⁷, dissimilatory bisulfite reductases (dsrAB) were compared with a database compiled by ref. ⁷⁸ 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) with a custom database based on sequences presented within ref. ⁷⁹.

All QAQC reads were remapped to a database consisting of only contigs of dereplicated MAGs. Normalized coverages for each of the MAGs were determined by scaling the resulting Anvi'o-determined coverages on the basis of the number of RNA polymerase B (*rpoB*) genes identified in the QAQC filtered reads. *RpoB* sequences were identified using ROCKER with the precomputed model⁸⁰. 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 kit⁸¹ in concert with the Genome Taxonomy Database (release 89)^{82,83} and its associated utilities^{67,84–88}. Single-copy marker genes were identified and aligned with GTDB_TK for all bacterial MAGs, and a phylogenetic tree of the concatenated alignment was constructed using FastTree2 v.2.1.10 in accordance with the JTT + CAT evolutionary model. The resulting phylogenetic tree was then imported into iTOL⁸⁹ for visualization, and all MAGs were subjected to growth rate index analysis within each metagenomic library⁹⁰.

Previously generated mRNA-enriched and post-processed metatranscriptomic libraries were procured from project PRJEB28783⁹¹. The groundwater source of these metatranscriptomes was collected in August and November 2015. QAQC filtered reads were mapped to MAGs using Bowtie2 v.2.3.5 in sensitive mode⁶⁸, and the total number of *rpoB* transcripts from each metatranscriptomic library was determined, as described in the preceding for metagenomes. The transcriptomic coverages for each ORF from each MAG were determined using Anvi'o v.6 and normalized via scaling-factor calculations based on the total number of *rpoB* reads from the original metatranscriptomic library (the coverage of each ORF from each MAG was normalized to a community-wide estimate of the transcriptional activity of a housekeeping gene in each sample). Means were determined considering all of the metatranscriptomes generated from a given well, including different sampling time points. While well H32 was sampled only once, mean values from all other wells account for three to four metatranscriptome coverages each. In addition, 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 were compiled and processed using R v.3.5.2 with Rstudio v.1.1.463^{92,93} and the tidyverse package⁹⁴, and colour schemes were generated using the RColorBrewer utility⁹⁵. All MAGs were deposited in project PRJEB36505's data repository.

Data availability

The metagenomic raw data for this study was uploaded to the European Nucleotide Archive (ENA) under project PRJEB36505, while the individual sample assemblies and the metagenome-assembled-genomes were deposited into the ENA under project PRJEB36523. To improve ease of access, we also uploaded all MAGs and their associated data to the Open Science Framework (OSF) Project <https://osf.io/4ceqs/>. All raw and summarized accelerator mass spectrometry (AMS) data are available in Supplementary Data File 2, which was also uploaded to the same OSF project as the MAGs.

References

- Lehmann, R. & Totsche, K. U. Multi-directional flow dynamics shape groundwater quality in sloping bedrock strata. *J. Hydrol.* **580**, 124291 (2020).
- Küsel, K. et al. How deep can surface signals be traced in the Critical Zone? Merging biodiversity with biogeochemistry research in a central German Muschelkalk landscape. *Front. Earth Sci.* **4**, 32 (2016).
- Yan, L. et al. Environmental selection shapes the formation of near-surface groundwater microbiomes. *Water Res.* **170**, 115341 (2019).
- Pack, M. A. et al. A method for measuring methane oxidation rates using low levels of ¹⁴C-labeled methane and accelerator mass spectrometry: methane oxidation rates by AMS. *Limnol. Oceanogr. Methods* **9**, 245–260 (2011).
- Nielsen, E. S. The use of radio-active carbon (C¹⁴) for measuring organic production in the sea. *ICES J. Mar. Sci.* **18**, 117–140 (1952).
- 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).
- Merser, S. Acetabulum online interactive statistical calculators. Accessed Feb, 2021. <https://acetabulum.dk/anova.html>
- Bermuda Oceanographic Timeseries, accessed 21 Oct 2020, http://batsftp.bios.edu/BATS/production/bats_primary_production.txt
- Hawaiian Oceanographic Timeseries, accessed 21 Oct 2020, ftp://ftp.soest.hawaii.edu/hot/primary_production
- Hawaiian Oceanographic Timeseries, accessed 21 Oct 2020, <https://hahana.soest.hawaii.edu/FTP/hot/microscopy/EPISlides.txt>
- 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).
- Füssel, J. et al. Nitrite oxidation in the Namibian oxygen minimum zone. *ISME J.* **6**, 1200–1209 (2012).
- 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).
- Dalsgaard, T., Thamdrup, B., Farias, L. & Revsbech, N. P. Anammox and denitrification in the oxygen minimum zone of the eastern South Pacific. *Limnol. Oceanogr.* **57**, 1331–1346 (2012).
- Thamdrup, B. et al. Anaerobic ammonium oxidation in the oxygen-deficient waters off northern Chile. *Limnol. Oceanogr.* **51**, 2145–2156 (2006).
- Taubert, M. et al. Tracking active groundwater microbes with D₂O labelling to understand their ecosystem function. *Environ. Microbiol.* **20**, 369–384 (2018).
- Bushnell, B. BBMap (SourceForge, 2014); <http://sourceforge.net/projects/bbmap>
- Bornemann, T. L. V. et al. Geological degassing enhances microbial metabolism in the continental subsurface. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.03.07.980714> (2020).
- Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
- Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
- Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357 (2012).
- 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).
- Brown, C. T. et al. Unusual biology across a group comprising more than 15% of domain bacteria. *Nature* **523**, 208–211 (2015).
- Sieber, C. M. K. et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* **3**, 836–843 (2018).
- 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).
- Murat Eren, A. et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**, e1319 (2015).
- Aramaki, T. et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- Parks, D. H. et al. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-020-0501-8> (2020).
- 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).

85. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90 K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).
86. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
87. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
88. Ondov, B. D. et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).
89. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).
90. Emiola, A. & Oh, J. High throughput in situ metagenomic measurement of bacterial replication at ultra-low sequencing coverage. *Nat. Commun.* **9**, 4956 (2018).
91. Wegner, C.-E. et al. Biogeochemical regimes in shallow aquifers reflect the metabolic coupling of the elements nitrogen, sulfur, and carbon. *Appl. Environ. Microbiol.* **85**, e02346-18 (2019).
92. R: A Language and Environment for Statistical Computing (R Core Team, 2018).
93. RStudio: Integrated Development Environment for R (RStudio Team, 2016).
94. Wickham, H. et al. Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
95. Neuwirth, E. RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. <https://CRAN.R-project.org/package=RColorBrewer> (2014).

Acknowledgements

We thank F. Gutmann, H. Minkmar, P. A. Figueroa-Gonzalez and R. Lehmann for their assistance with sample preparation, collection and filtration. We also thank M. Nowak and J. Southon for advice regarding ¹⁴C-bicarbonate supplementation and quantification of AMS results, respectively. In addition, we thank I. Görlich and M. 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. S.T., X.X. and V.F.S. acknowledge additional support from the European Research Council (Horizon 2020 Research and Innovation Programme, grant agreement 695101). M.T. gratefully acknowledges funding from the DFG under Germany's Excellence Strategy—EXC 2051—project ID 390713860. T.L.V.B. and

A.J.P. 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/>).

Author contributions

K.K., S.T., W.A.O. and M.T. designed this study. W.A.O., K.K., A.J.P., T.L.V.B., M.M., M. Hölzer and M.T. planned, designed and conducted the metagenomic sampling approach. W.A.O., T.L.V.B. and A.J.P. performed the metagenomic analysis. W.A.O., S.T., X.X. and V.F.S. performed the carbon fixation experiments and interpreted the results. M.K., M. Herrmann, B.T. and L.B. conducted the anammox and nitrification rate measurements. W.A.O., K.K. and S.T. wrote the manuscript with the help of all authors.

Funding

Open access funding provided by Friedrich-Schiller-Universität Jena.

Competing interests

The authors declare no competing interests.

Additional information

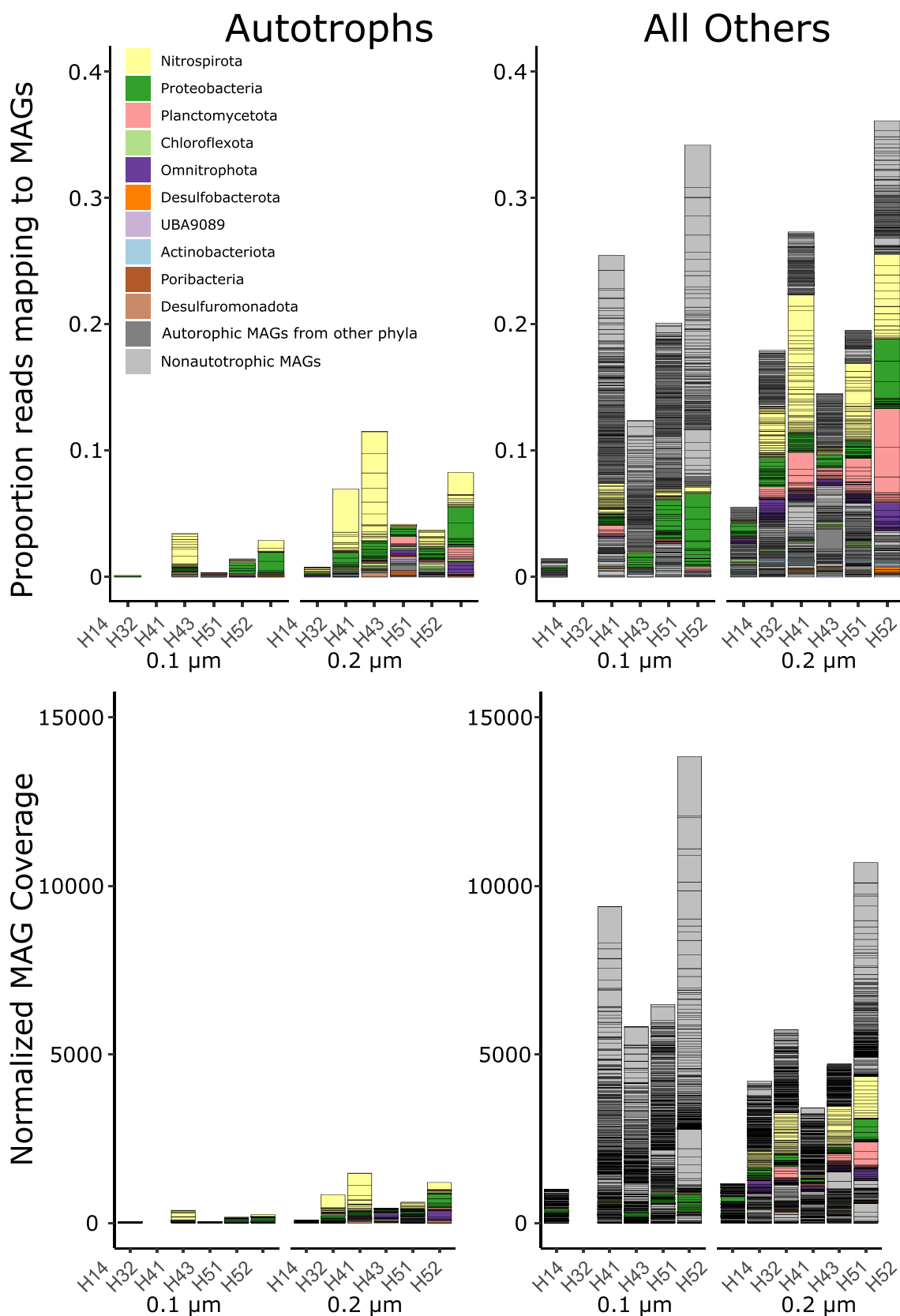
Extended data is available for this paper at <https://doi.org/10.1038/s41561-022-00968-5>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41561-022-00968-5>.

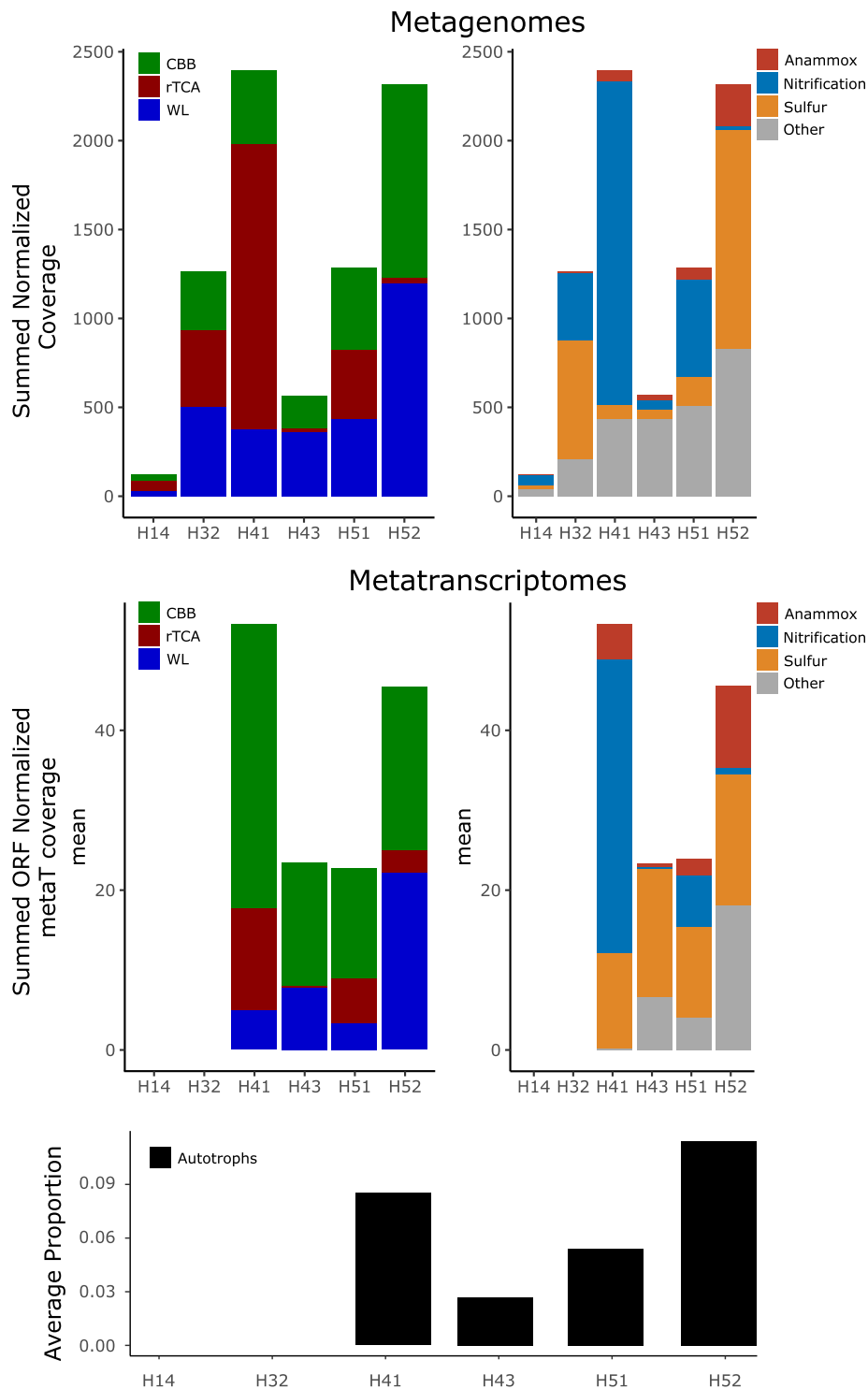
Correspondence and requests for materials should be addressed to Kirsten Küsel.

Peer review information *Nature Geoscience* thanks Michael Wilkins and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Kyle Frischkorn and Tamara Goldin, in collaboration with the *Nature Geoscience* team.

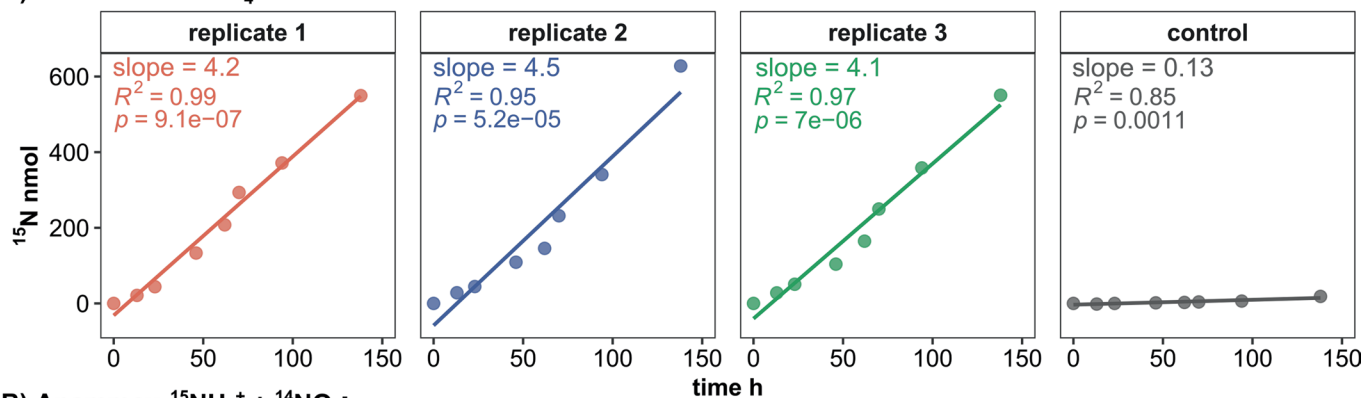
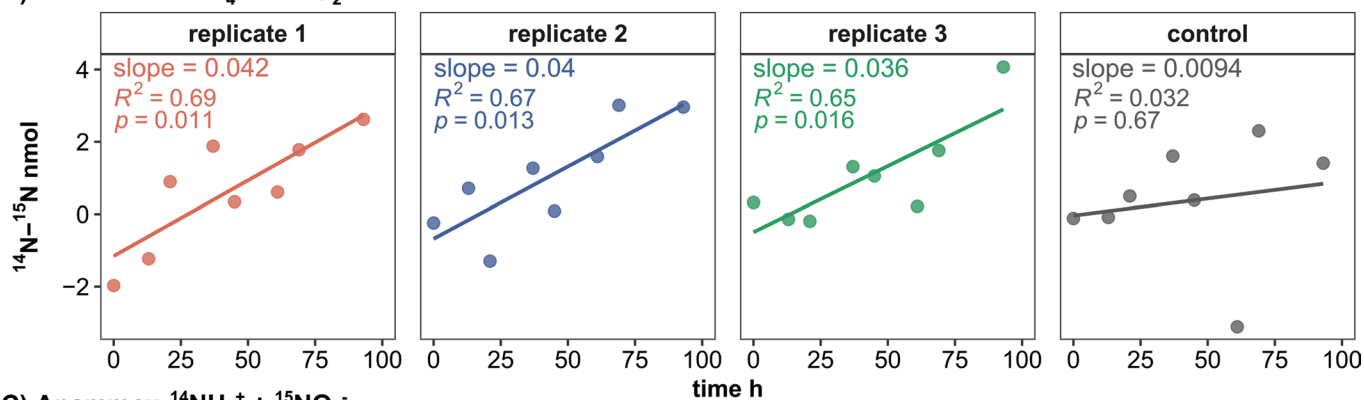
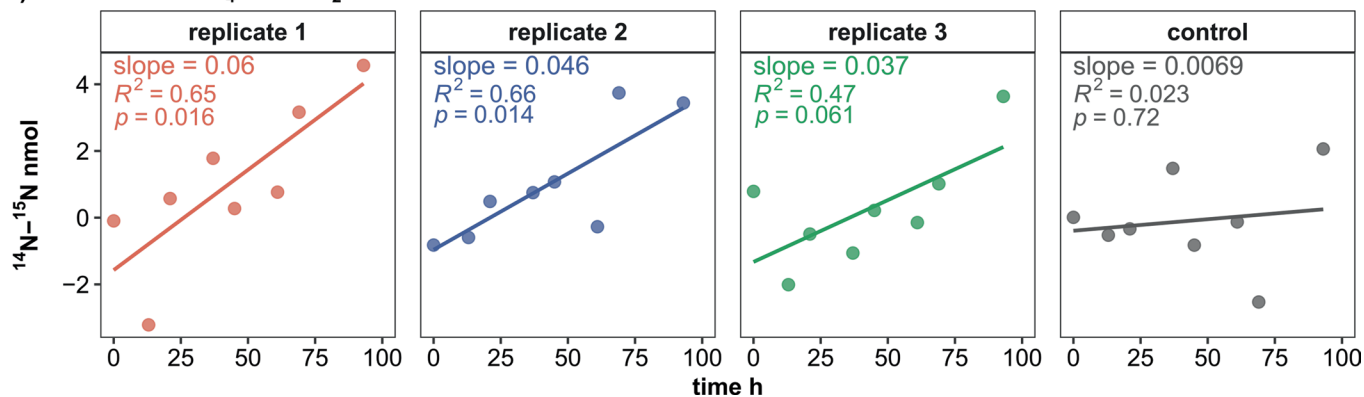
Reprints and permissions information is available at www.nature.com/reprints.



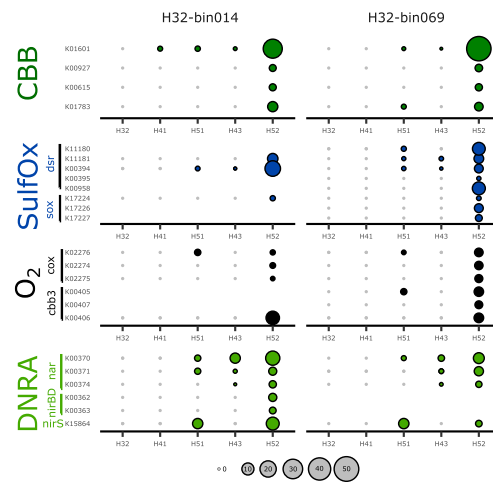
Extended Data Fig. 1 | Relative abundances of putative chemolithoautotrophic MAGs. (Top) Proportion of metagenomic short reads that mapped to the putatively chemolithoautotrophic MAGs. Colors indicate the phylum and each bar indicates a specific MAG. (Bottom) The normalized coverages assigned to the MAGs, which unlike the proportional representation is independent of MAG size. For each plot, 0.1 μm and 0.2 μm represent the filter fraction from which the DNA was extracted from.



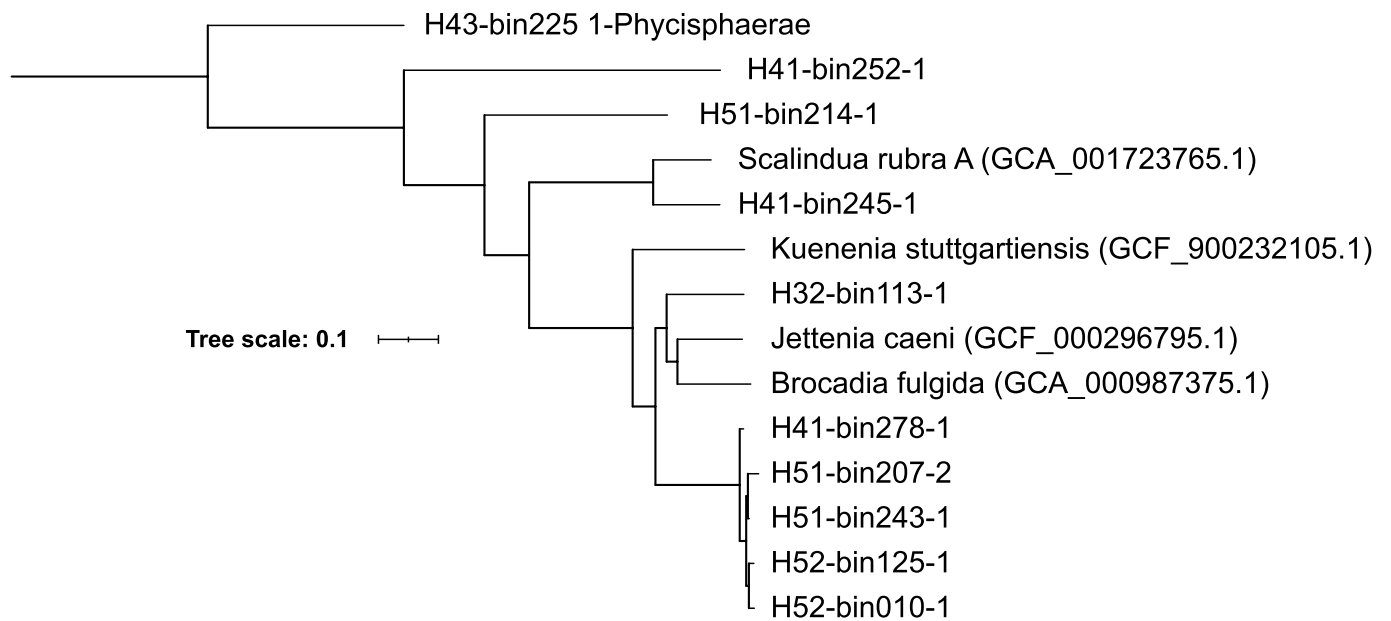
Extended Data Fig. 2 | Relative abundances of transcripts mapping to putative chemolithoautotrophic MAGs. Top panels are reproduced from Fig. 1 B, C. Middle panels are the corresponding values for the RNA-based reads. These values are derived from the ORF-based average normalized transcriptional activity of all putatively chemolithoautotrophic MAGs, as described in the methods and represent the sample-wise sum of the pie charts shown in Fig. 3. Bottom panel shows the average proportion of metatranscriptomic reads that map to the putatively chemolithoautotrophic MAGs. We did not have metatranscriptomes from H14 and the single H32 sample did not recruit reads.

A) Nitrification $^{15}\text{NH}_4^+$ B) Anammox $^{15}\text{NH}_4^+ + ^{14}\text{NO}_2^-$ C) Anammox $^{14}\text{NH}_4^+ + ^{15}\text{NO}_2^-$ 

Extended Data Fig. 3 | Nitrification and anammox linear regressions. Assessment of nitrogen transformation process rates in groundwater incubations by linear regression of measured increase in ^{15}N over time for (a) nitrification with $^{15}\text{NH}_4^+$ label, (b) anammox with $^{15}\text{NH}_4^+$ label and (c) anammox with $^{15}\text{NO}_2^-$ label. Shown are the slopes of the regression line for each replicate and control sample from which nitrification and anammox rates were calculated. R^2 values describe the accuracy of the linearity and p -values the significance of results for $p < 0.05$.



Extended Data Fig. 4 | Transcriptional activity of Sulfurifustaceae MAGs. Transcriptional activity of both dominant and putatively chemolithoautotrophic MAGs within the family Sulfurifustaceae. These bins exhibited nearly identical distributions throughout the aquifer transect and a > 94% AAI. H3₂-bin014 is also shown in Fig. 4b. Bubble size is relative to the normalized ORF transcriptional coverage.



Extended Data Fig. 5 | Phylogenetic tree of putative anammox MAGs. Approximately maximum-likelihood phylogenetic tree of Wood Ljungdahl containing anammox MAGs and reference genomes. The multiple protein sequence alignment from GTDB_TK (5040 positions) was used with the JTT + CAT model, manually rooted with a MAG classified to the sister class Phycisphaerae within the phylum Planctomycetota.

Extended Data Table 1 | CO₂ fixation rates

	CO ₂ fixation rate						Rate Difference (%)
	vs. kill controls			vs. no label controls			
	mean_ug/L/d	SD	SEM	mean_ug/L/d	SD	SEM	no-label vs. kill
H52	0.227	0.111	0.057	0.237	0.112	0.058	4.4
H51	-	-	-	0.099	0.097	0.053	-
H43	0.234	0.104	0.052	0.256	0.107	0.053	9.3
H41	0.043	0.013	0.006	0.050	0.014	0.005	17.7
H32	0.099	0.048	0.023	0.103	0.049	0.023	3.7
H14	0.050	0.022	0.008	0.055	0.024	0.010	9.8

ANOVA

	SS	DoF	MS	F	P
Between	0.1055	4	0.026365	6.676	0.00696
Within	0.0395	10	0.003949		
Total	0.1449	14			

PostHoc Tukey

H14 vs H43	[0.05, 0.234]	0.184	0.0151	0.3529	0.03163	*
H14 vs H52	[0.05, 0.227]	0.177	0.0081	0.3459	0.03909	*
H41 vs H43	[0.043, 0.234]	0.191	0.0221	0.3599	0.02561	*
H41 vs H52	[0.043, 0.227]	0.184	0.0151	0.3529	0.03163	*

CO₂ fixation rates, calculated from both the kill controls and the no-label addition controls. Significance testing uses the mean and SEM values and results from the posthoc Tukey test that were significant are reported.

Extended Data Table 2 | Predicted global rates

Groundwater

	Carbon fixation (ug/L/d)	Particulate organic carbon (µg / l)	16S gene-copies / L *	% POC	fgC fixed/bacterial_cell/d
H52	0.227	48.48	2.04E+08	0.47	1.11
H51	-		2.57E+07		
H43	0.234	44.70	2.16E+07	0.52	10.81
H41	0.043	51.84	2.91E+07	0.08	1.48
H32	0.099	25.55	6.73E+07	0.39	1.47
H14	0.05	24.62	1.51E+08	0.20	0.33

Oligotrophic Marine Waters (HOTS)

Water Depth Intervals	Carbon fixation (ug/L/d)**	Particulate organic carbon (µg / l)**	bacterial cell counts / l****	% POC	fgC //d/bacterial_cell
5	7.06	25.64	6.04E+08	27.51	11.68
25	6.48	26.48	6.23E+08	24.46	10.39
45	5.51	26.64	6.54E+08	20.67	8.42
75	3.27	25.48	6.25E+08	12.84	5.24
100	1.40	22.93	5.14E+08	6.09	2.72
125	0.43	18.67	3.58E+08	2.28	1.19
150	0.09	13.05	3.27E+08	0.66	0.26
175	0.01	9.94	3.73E+08	0.07	0.02

* Assuming each bacterial cell has one 16S gene

** Primary productivity rates from 1999 available from the HOTS FTP site (see methods)

Estimated based on the average values within the Martiny et al., 2015 dataset for year 1999 from the HOTS dataset, binned into the same depth ranges as given for the primary productivity depths.

**** Cells counts based on epifluorescent microscopy data available from the HOTS FTP site (see methods), from year 1999. Estimated by the sum of Prochlorococcus, heterotrophs, and Synechococcus counts.

Estimations of global subsurface carbon fixation rates.

Extended Data Table 3 | Normalized rates

Carbonate Aquifers						
total volume (Mkm ³)	2.26	Pg C / yr	sd	% Marine NPP (48.5 Pg yr ⁻¹)	% Land NPP (56.4 Pg yr ⁻¹)	% Global NPP (104.9 Pg yr ⁻¹)
mean rate (ug C / l / d)	0.13	0.107	0.069	0.221	0.190	0.102
sd	0.084					
lowest rate (ug C / l / d)	0.043	0.035		0.073	0.063	0.034

Crystalline Aquifers						
total volume (km ³)	12.66	Pg C / yr	sd			
mean rate (ug C / l / d)	0.0328	0.151	0.107	0.312	0.268	0.144
sd	0.0233					
lowest rate (ug C / l / d)	0.0095	0.044	0.107	0.091	0.078	0.042

Total		Pg C / yr				
mean		0.259		0.533	0.458	0.246
lowest		0.079		0.164	0.141	0.076

Normalizing the rate of carbon fixation to standing particulate organic carbon concentrations and estimated bacterial cell numbers.