nature geoscience

ARTICLES

https://doi.org/10.1038/s41561-021-00725-0

Effect of tectonic processes on biosphere–geosphere feedbacks across a convergent margin

Katherine M. Fullerton[1], Matthew O. Schrenk[2], Mustafa Yücel[3], Elena Manini[4], Marco Basili[1,4], Timothy J. Rogers[1], Daniele Fattorini[5,6], Marta Di Carlo[5], Giuseppe d'Errico[5,6], Francesco Regoli[5,6], Mayuko Nakagawa[7], Costantino Vetriani[8,9], Francesco Smedile[4,9], Carlos Ramírez[10], Heather Miller[2], Shaunna M. Morrison[11], Joy Buongiorno[11,18], Gerdhard L. Jessen[12], Andrew D. Steen[1,13], María Martínez[14], J. Maarten de Moor[14,15], Peter H. Barry[16], Donato Giovannelli[4,7,9,16,17] ✉ and Karen G. Lloyd[1] ✉

The subsurface is among Earth's largest biomes, but the extent to which microbial communities vary across tectonic plate boundaries or interact with subduction-scale geological processes remains unknown. Here we compare bacterial community composition with deep-subsurface geochemistry from 21 hot springs across the Costa Rican convergent margin. We find that cation and anion compositions of the springs reflect the dip angle and position of the underlying tectonic structure and also correlate with the bacterial community. Co-occurring microbial cliques related to cultured chemolithoautotrophs that use the reverse tricarboxylic acid cycle (rTCA) as well as abundances of metagenomic rTCA genes correlate with concentrations of slab-volatilized carbon. This, combined with carbon isotope evidence, suggests that fixation of slab-derived CO_2 into biomass may support a chemolithoautotrophy-based subsurface ecosystem. We calculate that this forearc subsurface biosphere could sequester 1.4×10^9 to 1.4×10^{10} mol of carbon per year, which would decrease estimates of the total carbon delivered to the mantle by 2 to 22%. Based on the observed correlations, we suggest that distribution and composition of the subsurface bacterial community are probably affected by deep tectonic processes across the Costa Rican convergent margin and that, by sequestering carbon volatilized during subduction, these chemolithoautotrophic communities could in turn impact the geosphere.

The deep-subsurface biosphere drives a wide range of key biogeochemical transformations such as reduction/oxidation and CO_2 fixation[1,2]. The composition of microbial communities in subsurface fluids of terrestrial and marine hydrothermal systems correlates with geochemical parameters such as temperature, pH, redox state and energy availability[2–6]. Freshly expressed, actively venting fluids driven by tectonic and magmatic processes provide access to subsurface microbial communities and deeply derived volatiles with varying degrees of inmixing from surficial microbes[4,7]. Subsurface microbial communities can be distinguished from microbial biofilms and planktonic communities that developed in the presence of sunlight after the fluids emerged at the surface, since the surficial mix of photosynthetic and chemosynthetic processes is recorded in the isotopic composition of the carbon pool[8]. It remains unknown whether the subsurface biosphere is connected to deep geological transitions across large (100s of kilometres) regions. A recent study showed that microbial communities in geothermal fluids across the Taupō volcanic zone vary due to dispersal and geochemistry, demonstrating the utility of combining many individual hot spring communities to form a cohesive dataset[9], similar to continental-scale assessments of soil microbial communities[10]. Here, we adapted this approach to investigate how microbial distribution is linked to changes in subsurface volatile delivery across a tectonic plate boundary.

Convergent margins connect the vast repository of carbon in the deep Earth with the planetary surface. As denser oceanic plates subduct beneath continental crust, carbon compounds and other volatiles are transferred from Earth's surface to its interior[11]. Some of these compounds are recycled back to the surface through arc volcanoes and secondary geothermal processes along subduction boundaries[11,12]. In this dynamic geologic setting, fluid release, magmatism and deformation provide diverse habitats that may be colonized by microbial assemblages with different preferences for temperature, pH, redox conditions, elemental compositions, pressure and salinity[2]. While subducting slabs can penetrate 20–150 km, the subsurface biosphere is probably limited to the upper few kilometres, where temperatures are below ~122 °C[13]. We hypothesize that the upward mobility of deeply sourced fluids may connect microorganisms to the deep tectonic processes below.

[1]Microbiology Department, University of Tennessee, Knoxville, TN, USA. [2]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI, USA. [3]Institute of Marine Sciences, Middle East Technical University, Erdemli, Turkey. [4]Institute of Marine Biological Resources and Biotechnologies, National Research Council, Ancona, Italy. [5]Dipartimento di Scienze della Vita e dell'Ambiente, Università Politecnica delle Marche, Ancona, Italy. [6]Consorzio Nazionale Interuniversitario Scienze del Mare, Rome, Italy. [7]Earth-Life Science Institute, Tokyo Institute for Technology, Tokyo, Japan. [8]Department of Biochemistry and Microbiology, Rutgers University, New Brunswick, NJ, USA. [9]Department of Marine and Coastal Science, Rutgers University, New Brunswick, NJ, USA. [10]Servicio Geológico Ambiental, Heredia, Costa Rica. [11]Earth and Planets Laboratory, Carnegie Institution for Science, Washington, DC, USA. [12]Instituto de Ciencias Marinas y Limnológicas, Universidad Austral de Chile, Valdivia, Chile. [13]Department of Earth and Planetary Sciences, University of Tennessee, Knoxville, TN, USA. [14]Observatorio Vulcanológico y Sismológico de Costa Rica, Universidad Nacional, Heredia, Costa Rica. [15]Department of Earth and Planetary Sciences, University of New Mexico, Albuquerque, NM, USA. [16]Woods Hole Oceanographic Institution, Woods Hole, MA, USA. [17]Department of Biology, University of Naples "Federico II", Naples, Italy. [18]Present address: Division of Natural Sciences, Maryville College, Maryville, TN, USA. ✉e-mail: donato.giovannelli@unina.it; klloyd@utk.edu

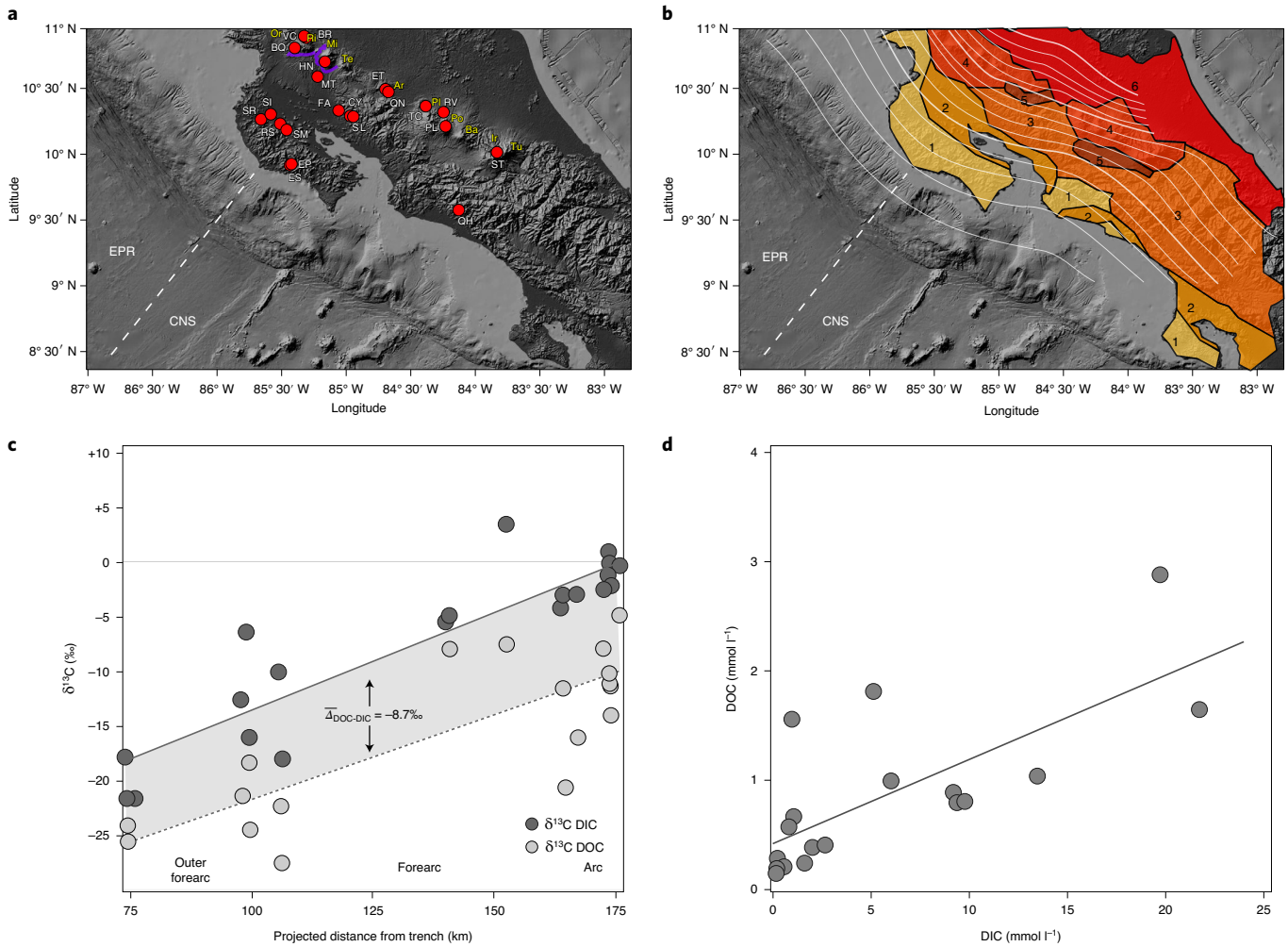

Fig. 1 | Sites span the Costa Rican convergent margin, with organic matter produced from chemolithoautotrophy of deep-slab inorganic carbon. **a**, Locations of the sampled hot springs and volcanic crater lake (red markers with white labels), volcanoes (yellow labels) and the calderas of the Guanacaste Geothermal Province (purple lines, adapted from Tassi et al.[38]). **b**, Subducting slab depth lines (thin lines every 10 km depth, thick lines every 50 km depth) and principal bedrock types: (1) Cretaceous/Tertiary ophiolites; (2) Tertiary basins; (3) Tertiary volcanic range; (4) Quaternary volcanic range; (5) intra-arc basins; (6) Caribbean coastal plain (adapted from Tassi et al.[39]). Dotted lines show the boundary between the plates of the EPR and CNS. **c**, $\delta^{13}C$ values of DIC and DOC show similar spatial trends with the projected distance from the trench (solid and dashed lines with R^2 values of 0.81 and 0.65 for DIC and DOC, respectively) with an average offset ($\Delta_{DOC-DIC}$) of -8.7 ± 1.3‰. **d**, Concentrations of DOC and DIC correlate with each other (Pearson moment correlation $R^2 = 0.53$, $P < 0.01$, $n = 18$) across the sampled sites (site SL is off the scale with 5.69 mmol l^{-1} DIC and 6.29 mmol l^{-1} DOC). Data in **c** and **d** were used in previously published calcite precipitation models[16].

Here we link bacterial community composition to aqueous and solid phase geochemistry, plate transitions and volatile emissions across the Costa Rican convergent margin (Fig. 1). In this area, the Cocos plate subducts beneath the Caribbean plate at a rate of 80–90 mm yr^{-1} (ref. [14]). The shallow subduction geometry (66–49° from north to south)[14] promotes slab dehydration in the outer forearc, forearc and arc, releasing large fluxes of carbon and reduced volatiles into the overlying plate[15]. Remarkably, the continental extension of the oceanic plate boundary between the East Pacific Rise (EPR) and Cocos Nazca spreading centre (CNS) is identifiable by a shift in the carbon-isotopic composition of arc and forearc fluid and gas emissions[16], suggesting tight coupling between deep subducting structures and near-surface fluids.

Within the Central American volcanic arc, only a few hot springs have been characterized microbiologically[17,18]. In February 2017 we sampled 21 hot springs across a >200 km section of northern and central Costa Rica. Fresh-venting fluids were collected before reaching the surface using a copper tube inserted into the source orifice. Sediments accumulated at the outflow source were also collected. The sites cover a range of subduction provinces from the outer forearc (20–40 °C, pH 8–10 and 20–40 km slab depth) to the forearc (40–60 °C, pH 4–7 and 40–100 km slab depth) and arc (>60 °C, pH 0–3 and 100–120 km slab depth)[16], as well as a wide range of geochemical parameters (Supplementary Tables 1–3 and Supplementary Figs. 1 and 2). This includes volcanoes from the Northern Guanacaste Geothermal Province, as well as Arenal, Poás and Irazú in the Central Cordillera (Fig. 1).

Fluids had 1.5×10^3 to 3.3×10^6 cells ml^{-1} (Supplementary Table 4), typical of hydrothermal systems[1,4]. Bacterial 16S ribosomal RNA gene amplicon libraries produced 1,933,379 reads after quality control, and 33,188 total amplicon sequence variants (ASVs) representing 59 phyla (Supplementary Figs. 3 and 4). At each site, helium isotope ratios ($^3He/^4He$) were enriched relative to air suggesting fluids derived from the mantle/slab with little input from surface fluids (shown by high $^4He/^{20}Ne$), and ^{13}C-enriched dissolved inorganic carbon (DIC) derived from a mixture of the

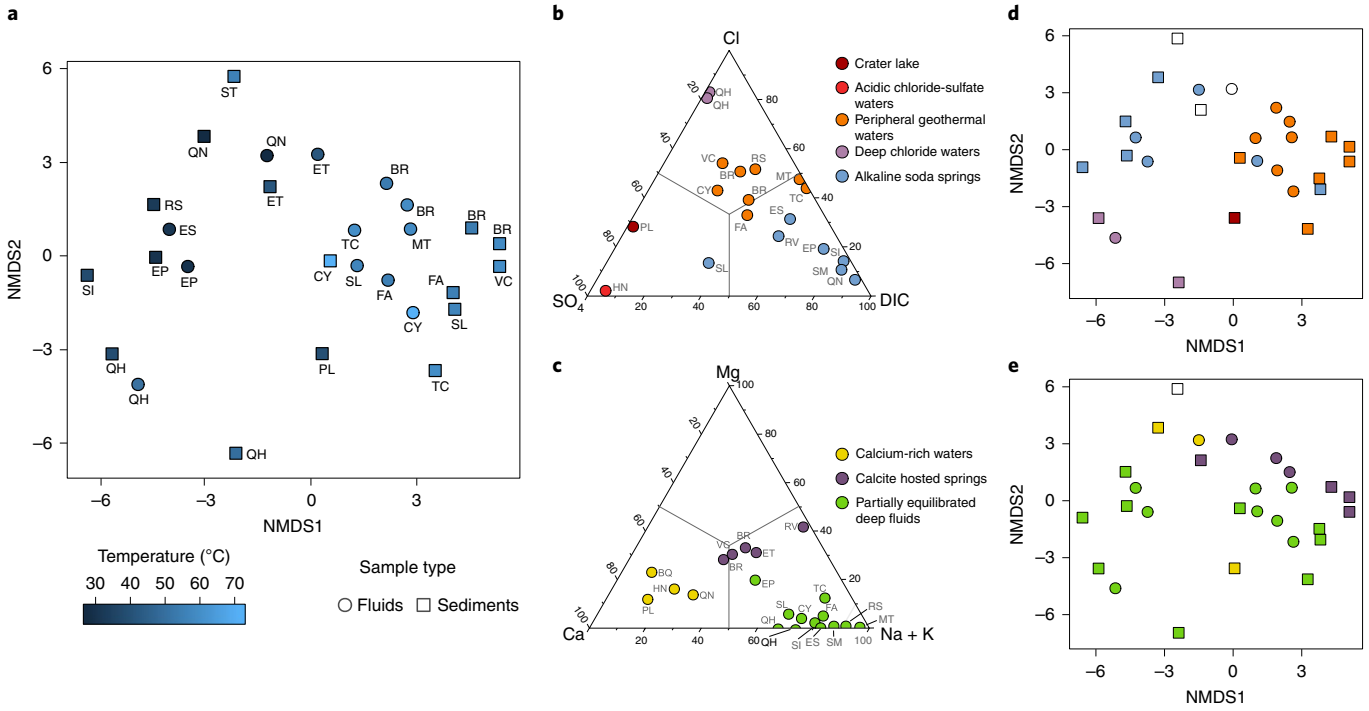

Fig. 2 | Clustering of the sites based on microbial community diversity and geochemical characteristics. a, NMDS (stress 0.17) plot of the 16S rRNA gene amplicon microbial diversity based on Jaccard dissimilarity measure in the fluids (circles) and sediments (squares), coloured by spring temperature, with sample names corresponding to those of Fig. 1. **b**,**c**, Ternary diagrams showing the clustering of the samples based on the major anions (**b**) and major cations (**c**). **d**,**e**, The same NMDS plot from **a**, but coloured according to the geochemistry-based grouping from **b** (anions, ADONIS $P<0.05$, $n=19$, Holm-corrected) and **c** (cations, ADONIS $P<0.05$, $n=21$, Holm-corrected), respectively. Empty symbols in **d** and **e** represent samples for which one or more ions were missing, and therefore have no placement in **b** and **c**.

oceanic slab and mantle rather than air[16]. The $\delta^{13}C$ values (defined as $R_{sample}/R_{standard}-1\times1000$, where R is $^{13}C/^{12}C$ normalized to the $\delta^{13}C$ value of the international standard Pee Dee Belemnite) of dissolved organic carbon (DOC) correlate positively (Pearson $R^2=0.53$, $P<0.01$, $n=18$) with those of slab/mantle-derived DIC (Fig. 1c), rather than with the degree of mixing with surface-derived organic matter[16]. Photosynthetic biomarkers are low in abundance, with $<5\,\mu g\,g^{-1}$ of total photosynthetic pigments in surface sediments and <1% and <4% chloroplast-related 16S rRNA genes in fluids and surface sediments, respectively[16]. The fluids are therefore deeply sourced and primarily contain subsurface microbes, similar to active cold springs[19] and deep-sea hydrothermal vents[4].

Total bacterial community and geological features

Bacterial community composition varies significantly with geologic province across the subduction zone (outer forearc, forearc and arc), between the upper plate overlying the EPR and CNS subducted crusts, as well as with the dominant bedrock types (Fig. 1, Supplementary Table 5 and Supplementary Fig. 5; ADONIS, permutational multivariate analysis of variance, weighted Jaccard, $P<0.05$, $n=28$). Phyla with thermophilic isolates, such as Thermotogae, Aquificae and candidate phyla Atribacteria and Hydrothermae, increase in relative 16S rRNA gene abundances with increasing temperature and acidity. The opposite trend is observed for Proteobacteria, a few uncultured phyla, *Candidatus* Cloacimonetes and others (Fig. 2a and Supplementary Figs. 5 and 6; Jaccard similarity in Supplementary Tables 6 and 7). Most phyla, however, do not vary systematically with temperature and pH (Supplementary Fig. 6), suggesting that their distribution across this convergent margin is not a simple function of these two parameters.

A ternary distribution of aqueous anions (Cl^-, SO_4^{2-} and DIC, Fig. 2b)[20] distinguishes between: (1) acidic (pH 0 to 3) chloride-sulfate waters with direct absorption of magmatic gases at arc sites, (2) sulfate-poor peripheral geothermal fluids, intermediate in composition between deeply derived chloride-rich waters and soda springs characteristic of flank volcanic sites and forearc locations and (3) alkaline outer forearc sites, relatively poor in both sulfate and chloride[21]. Aqueous major cations (Fig. 2c; Mg^{2+}, Ca^{2+} and $Na^+ + K^+$) distinguish: (1) Ca^{2+}-rich acidic arc sites (for example, ref. [22]), (2) volcanic-flank geothermal sites and forearc springs often associated with travertine deposition[16] and (3) deep fluids partially equilibrated with feldspars and clays in the outer forearc[20]. Both anion and cation fluid compositions correlate well with bacterial community compositions in both fluids and sediments (Fig. 2d,e; Supplementary Tables 5–7). This suggests that deep, tectonically controlled geochemistry of basement rocks and the hydrothermal fluids resulting from subsurface water–rock interactions are well-correlated with bacterial community composition (Fig. 2).

Iron and sulfur cycling microbes

The extent to which geochemical differences influence bacterial distribution and composition across the Costa Rican convergent margin is demonstrated by the stark differences in the distribution of likely sulfur- and iron-oxidizing bacteria. Genera related to iron-oxidizing isolates (such as *Gallionella*, *Geobacter* and *Ferritrophicum*) (Supplementary Fig. 7) and the presence of their mineral product, twisted stalks of iron hydroxide[23] (Supplementary Fig. 8), are only found at central sites. In contrast, genera of sulfide-oxidizing isolates (such as *Sulfurihydrogenibium*) are found mainly at northern forearc/arc sites (Supplementary Fig. 7). The northern sites are mostly located within the Guanacaste Geothermal

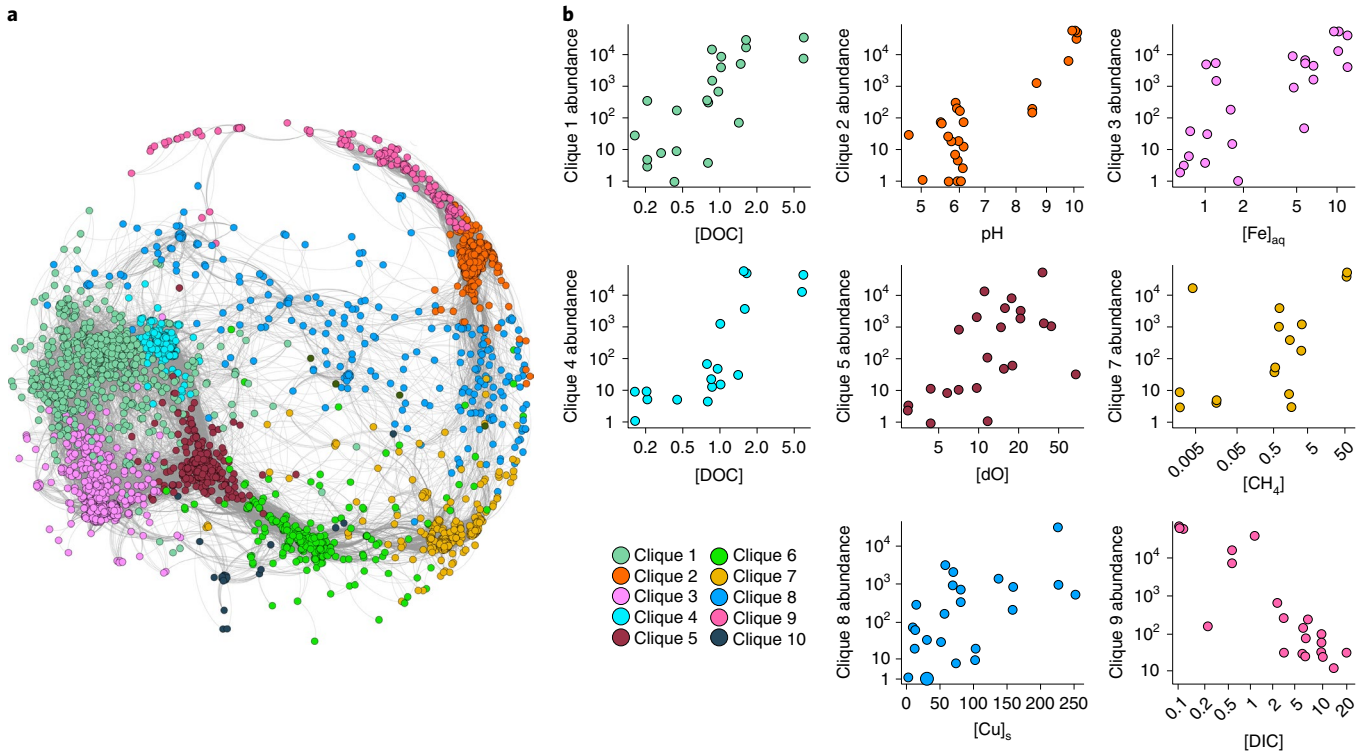

Fig. 3 | Bacterial cliques each have different relationships to subduction zone geochemistry. a, Co-occurrence network analysis of the dominant ASVs across all sampling sites, coloured by clique, with positive Spearman correlations >0.7 plotted as edges. **b**, Sum of ASVs in each clique plotted against the subduction zone geochemical variable identified by RF, Spearman correlation and Pearson correlation as the major explanatory variable for that clique ($P < 0.01$, Holm-corrected, $n = 16$–25). Cliques 6 and 10 were not significantly correlated with any geochemical variable. Concentrations are indicated by square bracket notation. dO, dissolved oxygen.

Province (Fig. 1a), where the subducting slab angle is steeper than in the central region (66° versus 49°)[14], resulting in an extensional stress regime and favourable conditions for caldera formation[24], more permeability in the upper crust and extensive gas–fluid–rock interactions leading to sulfide generation. In the central region, the shallower subduction angle results in compressional tectonics favoring the formation of large stratovolcanoes with limited permeability[25]. Here, hydrothermal systems are shallower, and sulfur is mostly oxidized (SO_2, sulfuric acid, native sulfur)[26] and concentrated toward the vents of active volcanoes[26]. The springs surrounding the Central Cordillera stratovolcanoes are enriched in total dissolvable iron leached from acidic water–rock interactions, of which Fe(II) probably dominates and provides electrons for microbial oxidation (Supplementary Table 3). In the northern sites, sulfide may limit iron availability for microbial respiration, since sulfide reacts with Fe(II) to form iron-sulfide and pyrite, as suggested by the presence of pyrite framboids (Supplementary Fig. 8). The relatively lower Fe/Cl ratio (average 0.36 μmol mmol^{-1}) in the northern sites compared with central sites (average 1.17 μmol mmol^{-1}) is also consistent with iron in the dissolved fraction being limited by trapping in pyrite (Supplementary Tables 2 and 3). This combination of 16S rRNA gene and mineralogical data suggests that the along-arc crustal structural differences between the two regions, ultimately driven by dip angles and extensional versus compressional local stress regimes[25], dictate the distribution of sulfur- and iron-oxidizing microbes.

Co-occurring microbial cliques and geochemistry

We sought emergent patterns in the responses of these communities to tectonically driven geochemical variables across the convergent margin through a co-occurrence network of ASVs coupled to random forest (RF) variable ranking and correlations with Spearman and Pearson statistics (Fig. 3 and Supplementary Tables 8–13)[27]. This approach allows for the identification of microbial groups, hereafter called cliques, that cohesively respond to variation in geochemical parameters across the dataset. In this way, cliques sharing similar spatial patterns of abundance can be correlated to subsurface conditions as a group rather than as individuals, accounting for functional redundancies and ecological succession. Using this approach, we identify ten cliques of co-occurring ASVs (Fig. 3 and Supplementary Fig. 9, Supplementary Tables 8 and 9 and Supplementary Discussion), most of which contain multiple taxa (Supplementary Fig. 10), comprising >99% of the reads in the network analysis (Supplementary Table 9).

Most cliques correlate with subsurface geochemistry (Fig. 3 and Supplementary Tables 10–13), suggesting a strong link with the underlying geological settings. The exceptions were clique 10, which is not present in enough sites to make cross-site correlations significant, and clique 6, with <1% of the total reads and dominated by *Aquabacterium* sp. and *Alishewanella* sp., common heat-tolerant bacteria from soils[28] and freshwater[29]. These cliques probably represent shallow subsurface organisms that contrast with the rest of the cliques, which correlate with deep-subsurface parameters.

Carbon availability is the most important corollary for the greatest number of cliques. Cliques 9 and 2 are dominated by *Thiothrix* sp. (65% of clique 9) and *Hydrogenophilaceae* (12% of clique 2), and correlate best with decreasing DIC and increasing pH, respectively (Fig. 3b). Isolates from these clades include chemolithoautotrophic sulfur oxidizers[30], suggesting their distribution across the convergent margin may be due to an ability to tolerate limiting DIC concentrations and high pH in the outer forearc. Clique 4 correlates best with increasing DOC concentrations (Fig. 3b), which increase toward the arc, and is dominated by *Sulfurihydrogenibium* sp. (55%), which,

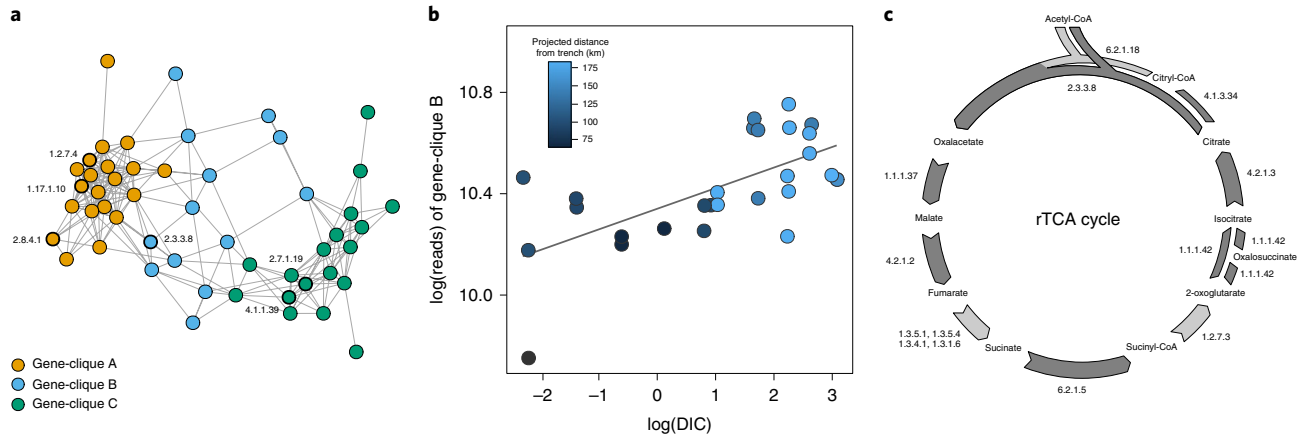

Fig. 4 | Metagenome-derived genes from the same carbon-fixation pathway correlate with each other and with subduction zone geochemistry. **a**, Co-occurrence network analysis of the genes involved in all central carbon-metabolism pathways, with edges representing Spearman correlations with a $\rho > 0.5$ ($P < 0.01$, $n = 49$, Holm-corrected). The key genes for the major carbon-fixation pathways were recovered in discrete gene cliques: Wood–Ljungdahl key genes in gene-clique A, rTCA genes in gene-clique B and Calvin–Benson–Bassham in gene-clique C. The key genes for each pathway are highlighted with thicker outlines and have the Enzyme Commission (EC) number plotted. **b**, Sum of gene-clique B genes (normalized read abundance for each enzyme) plotted against DIC concentrations (in mmolC l^{-1}) with blue colour saturation corresponding to the distance from the trench (in km) for each site (Spearman correlation of $\rho = 0.66$; $P < 0.001$, $n = 27$, Holm-corrected). **c**, Diagram of the rTCA cycle with genes recovered in gene-clique B highlighted in dark grey. These include the key genes ATP citrate (pro-S) lyase and the citryl-CoA lyase necessary for the functioning of the two alternative versions of the rTCA cycle[33]. The two light grey rTCA genes were present in the samples but did not correlate as tightly with the other genes to be included in gene-clique B, suggesting they were also used in other pathways. EC numbers are shown.

along with the other unclassified *Hydrogenothermacea* in this clique, are related to facultative autotrophic sulfur- and hydrogen-oxidizers with the reverse tricarboxylic acid (rTCA) pathway[31]. Clique 1 correlates more loosely with DOC, and includes diverse heterotrophic genera able to utilize a variety of anaerobic terminal electron acceptors. Together these cliques contain the majority (57.3%) of reads in the network, suggesting that even though different electron donors and acceptors are important for defining the community, availability of carbon volatilized during subduction is the overall limiter of community composition[16].

To investigate direct genetic evidence for connections between putative chemolithoautotrophs and carbon geochemistry, we constructed a co-occurrence network analysis of metagenomic DNA sequences related to all types of central carbon metabolism across all sites. Enzymes are present from three carbon-fixation pathways (rTCA, Calvin–Benson–Bassham and Wood–Ljungdahl[32]) and they cluster in statistically robust gene cliques around the key gene in each pathway (Fig. 4a, Supplementary Fig. 12 and Supplementary Table 14). Gene-clique B, containing all the genes necessary for rTCA, including the key genes encoding for adenosine triphosphate (ATP) citrate lyase (as well as its alternative, citryl-CoA lyase[33]) and 2-oxoglutarate synthase (Fig. 4c), is the only gene clique that correlates consistently with subsurface geochemical parameters (Fig. 4b, Supplementary Figs. 11 and 12 and Supplementary Tables 15–17). Across all statistical tests (Supplementary Tables 15–17), gene-clique B correlates best with DIC, suggesting a strong relationship between DIC and rTCA-based chemolithoautotrophy.

Chemolithoautotrophy across the convergent margin

Our microbial community analyses show that: (1) cliques containing known rTCA-utilizing chemolithoautotrophic genera correlate best with DOC concentrations, consistent with these organisms contributing to the DOC pool; (2) members of known heterotrophic genera also correlate with DOC, suggesting that chemosynthetically derived organic matter stimulates secondary consumers; (3) all necessary genes in the rTCA pathway, including those encoding key enzymes, form statistically significant gene-clique B, consistent with them being used together in the rTCA pathway and (4) the rTCA-gene gene-clique, B, correlates best with DIC concentrations, consistent with higher DIC concentrations stimulating more rTCA-based chemolithoautotrophy. Our geochemical analyses show that: (1) the entire DIC pool was initially derived from a deep-subsurface slab/mantle source, which was subsequently modified by varying amounts of CO_2 loss from deep calcite deposition[16]; (2) the entire DOC pool has a $\Delta^{13}C_{DIC-DOC}$ value ($-8.7 \pm 1.3‰$) in the range of the measured fractionation factors for rTCA (-2 to $-12‰$)[32] (Fig. 1) and (3) concentrations of DOC correlate with concentrations of DIC[16], suggesting that more DOC is produced when more DIC is available. Together, these findings are consistent with the ecosystem being supported primarily by chemolithoautotrophic biomass production through the rTCA cycle, dependent on the supply of DIC from deeply sourced fluids rising from the slab/mantle after subsequently undergoing calcite precipitation. Given reasonable estimates of cell turnover time and carbon content (Supplementary Methods), we estimate that 44 to 442 years would be required to produce the entire DOC pool from chemolithoautotrophy. Our estimated fluid residence times of 8.8×10^4 to 8.8×10^6 years, which are in the range measured for other systems[34], are sufficiently long for this to occur. Other chemolithoautotrophy-based ecosystems have been described on the basis of metagenomic, metatranscriptomic and metaproteomic data in oligotrophic (DOC < 30 μM) deep-subsurface aquifers[5,6]. Our work shows that such chemolithoautotrophically based ecosystems can fuel production of DOC to 100-fold higher concentrations (>1 mM) across an active convergent margin.

We estimate 7.9×10^{10} to 7.9×10^{11} molC is trapped as cell biomass in the Northern Costa Rican convergent margin subsurface. This is an update of our previous estimate[16] of 2.8×10^9, due to a refinement of the shape of the 122 °C isotherm[35]. An independent calculation using an arc-specific power-law fit with depth[1] resulted in similar numbers of 1.8×10^{11} to 6.4×10^{11} molC. Even if only 30% of the cells are chemolithoautotrophs[5,6], the biological fixation of slab-derived CO_2 would result in 1.4×10^9 to 1.4×10^{10} molC yr^{-1} of microbial cells. This is much greater than the amount of carbon accounted for by DOC alone (4.1×10^5 to 4.4×10^8 molC yr^{-1})[16], accounting for an additional removal of 1–13% of the carbon inputs to the subduction zone and

a further reduction of 2% to 22% in the amount of carbon assumed to be transported to the deep mantle. This carbon sequestration is in addition to the carbon sequestered as calcite, DIC and DOC[16,26]. Our metagenomic results suggest that these chemolithoautotrophs are limited by DIC availability, since calcite deposition dominates the forearc subsurface carbon sink. If other convergent margins lack such a strong calcite sink, chemolithoautotrophs would no longer be limited by DIC availability, suggesting that the biological carbon sink could compensate for the calcite carbon sink in other systems. Additionally, this biological carbon sink intersects with other biologically required elements' biogeochemical cycles[36,37], making the entire system react differently to changes in chemical compositions than if calcite alone were the major carbon sink.

We conclude that the microbial community of an ~200 km subduction segment in Costa Rica correlates cohesively to geochemical signals from deep tectonic processes (Fig. 2). This subsurface biosphere landscape varies along the convergent margin, through changes in supply of redox-active substrates, as well as across the convergent margin, through changes in the supply of DIC to a chemosynthetically based ecosystem. Collectively, our work shows that volatiles and elements mobilized from the descending slab and mantle can be altered by interaction with the deep-subsurface biosphere on their trek back to the surface, resulting in a coupling between geological and biological feedbacks affecting the volatile budget in a convergent margin. This has implications for the understanding of carbon reservoir changes in deep time, the coupling of carbon sequestration to the mantle with crustal carbon sequestration and, ultimately, the planetary redox balance and long-term climate stability.

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/s41561-021-00725-0.

Received: 6 September 2020; Accepted: 3 March 2021;
Published online: 22 April 2021

References

1. Magnabosco, C. et al. The biomass and biodiversity of the continental subsurface. *Nat. Geosci.* **11**, 707–717 (2018).
2. Merino, N. et al. Living at the extremes: extremophiles and the limits of life in a planetary context. *Front. Microbiol.* **10**, 780 (2019).
3. Colman, D. R. et al. Geobiological feedbacks and the evolution of thermoacidophiles. *ISME J.* **12**, 225–236 (2018).
4. Reveillaud, J. et al. Subseafloor microbial communities in hydrogen-rich vent fluids from hydrothermal systems along the Mid-Cayman Rise. *Environ. Microbiol.* **18**, 1970–1987 (2016).
5. Lau, M. C. Y. et al. An oligotrophic deep-subsurface community dependent on syntrophy is dominated by sulfur-driven autotrophic denitrifiers. *Proc. Natl Acad. Sci. USA* **113**, 7927–7936 (2016).
6. 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).
7. Brazelton, W. J. et al. Metagenomic identification of active methanogens and methanotrophs in serpentinite springs of the Voltri Massif, Italy. *PeerJ* **5**, e2945 (2017).
8. Havig, J. R., Raymond, J., Meyer-Dombard, D. R., Zolotova, N. & Shock, E. L. Merging isotopes and community genomics in a siliceous sinter-depositing hot spring. *J. Geophys. Res. Biogeosci.* **116**, G01005 (2011).
9. Power, J. F. et al. Microbial biogeography of 925 geothermal springs in New Zealand. *Nat. Commun.* **9**, 2876 (2018).
10. Lauber, C. L., Hamady, M., Knight, R. & Fierer, N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* **75**, 5111–5120 (2009).
11. Kelemen, P. B. & Manning, C. E. Reevaluating carbon fluxes in subduction zones, what goes down, mostly comes up. *Proc. Natl Acad. Sci. USA* **112**, E3997–E4006 (2015).
12. Brovarone, A. V. et al. Subduction hides high-pressure sources of energy that may feed the deep subsurface biosphere. *Nat. Commun.* **11**, 3880 (2020).
13. Plümper, O. et al. Subduction zone forearc serpentinites as incubators for deep microbial life. *Proc. Natl Acad. Sci. USA* **114**, 4324–4329 (2017).
14. Syracuse, E. M. & Abers, G. A. Global compilation of variations in slab depth beneath arc volcanoes and implications. *Geochem. Geophys. Geosyst.* **7**, Q05017 (2006).
15. Shaw, A. M., Hilton, D. R., Fischer, T. P., Walker, J. A. & Alvarado, G. E. Contrasting He–C relationships in Nicaragua and Costa Rica: insights into C cycling through subduction zones. *Earth Planet. Sci. Lett.* **214**, 499–513 (2003).
16. Barry, P. H. et al. Forearc carbon sink reduces long-term volatile recycling into the mantle. *Nature* **568**, 487–492 (2019).
17. Arce-Rodríguez, A. et al. Thermoplasmatales and sulfur-oxidizing bacteria dominate the microbial community at the surface water of a CO_2-rich hydrothermal spring located in Tenorio Volcano National Park, Costa Rica. *Extremophiles* **23**, 177–187 (2019).
18. Crespo-Medina, M. et al. Methane dynamics in a tropical serpentinizing environment: the Santa Elena ophiolite, Costa Rica. *Front. Microbiol.* **8**, 916 (2017).
19. Probst, A. J. & Moissl-Eichinger, C. "*Altiarchaeales*": uncultivated Archaea from the subsurface. *Life* https://doi.org/10.3390/life5021381 (2015).
20. Giggenbach, W. F. Geothermal solute equilibria, derivation of Na-K-Mg-Ca geoindicators. *Geochim. Cosmochim. Acta* **52**, 2749–2765 (1988).
21. Giggenbach, W. F. & Soto, R. C. Isotopic and chemical composition of water and steam discharges from volcanic–magmatic–hydrothermal systems of the Guanacaste Geothermal Province, Costa Rica. *Appl. Geochem.* **7**, 309–332 (1992).
22. Rodríguez, A. & van Bergen, M. J. Superficial alteration mineralogy in active volcanic systems: an example of Poás volcano, Costa Rica. *J. Volcanol. Geotherm. Res.* **346**, 54–80 (2017).
23. Chan, C. S., Fakra, S. C., Emerson, D., Fleming, E. J. & Edwards, K. J. Lithotrophic iron-oxidizing bacteria produce organic stalks to control mineral growth: implications for biosignature formation. *ISME J.* **5**, 717–727 (2011).
24. Lücke, O. H. & Arroyo, I. G. Density structure and geometry of the Costa Rican subduction zone from 3-D gravity modeling and local earthquake data. *Solid Earth* **6**, 1169–1183 (2015).
25. Protti, M., Gündel, F. & McNally, K. The geometry of the Wadati–Benioff zone under southern Central America and its tectonic significance: results from a high-resolution local seismographic network. *Phys. Earth Planet. Inter.* **84**, 271–287 (1994).
26. de Moor, J. M. et al. A new sulfur and carbon degassing inventory for the Southern Central American Volcanic Arc: the importance of accurate time-series data sets and possible tectonic processes responsible for temporal variations in arc-scale volatile emissions: new volatile budget for Central America. *Geochem. Geophys. Geosyst.* **18**, 4437–4468 (2017).
27. Delgado-Baquerizo, M. et al. A global atlas of the dominant bacteria found in soil. *Science* **359**, 320–325 (2018).
28. Kim, M. S., Jo, S. K., Roh, S. W. & Bae, J. W. *Alishewanella agri* sp. nov., isolated from landfill soil. *Int. J. Syst. Evol. Microbiol.* **60**, 2199–2203 (2010).
29. Chen, W. M. et al. *Aquabacterium limnoticum* sp. nov., isolated from a freshwater spring. *Int. J. Syst. Evol. Microbiol.* **62**, 698–704 (2012).
30. Garrity, G. M. & Bell, J. A. *Bergey's Manual of Systematics of Archaea and Bacteria* (Bergey's Manual Trust, 2015).
31. Hayashi, N. R., Ishida, T., Yokota, A., Kodama, T. & Igarashi, Y. *Hydrogenophilus thermoluteolus* gen. nov., sp. nov., a thermophilic, facultatively chemolithoautotrophic, hydrogen-oxidizing bacterium. *Int. J. Syst. Evol. Microbiol.* **49**, 783–786 (1999).
32. Berg, I. A. et al. Autotrophic carbon fixation in archaea. *Nat. Rev. Microbiol.* **8**, 447–460 (2010).
33. Giovannelli, D. et al. Insight into the evolution of microbial metabolism from the deep-branching bacterium, *Thermovibrio ammonificans*. *eLife* **6**, e18990 (2017).
34. Yokochi, R. et al. Noble gas radionuclides in Yellowstone geothermal gas emissions: a reconnaissance. *Chem. Geol.* **339**, 43–51 (2013).
35. Harris, R. N. & Wang, K. Thermal models of the Middle America Trench at the Nicoya Peninsula, Costa Rica. *Geophys. Res. Lett.* **29**, 6-1–6-4 (2010).
36. Jelen, B. I., Giovannelli, D. & Falkowski, P. G. The role of microbial electron transfer in the coevolution of the biosphere and geosphere. *Annu. Rev. Microbiol.* **70**, 45–62 (2016).
37. Falkowski, P. G., Fenchel, T. & Delong, E. F. The microbial engines that drive Earth's biogeochemical cycles. *Science* **320**, 1034–1039 (2008).
38. Tassi, F. et al. The geothermal resource in the Guanacaste region (Costa Rica): new hints from the geochemistry of naturally discharging fluids. *Front. Earth Sci.* **6**, 69 (2018).
39. Tassi, F., Vaselli, O., Barboza, V., Fernandez, E. & Duarte, E. Fluid geochemistry and seismic activity in the period 1998–2002 at Turrialba Volcano (Costa Rica). *Ann. Geophys.* **47**, 4 (2004).

Methods

Location and sample collection. At each sampling site, 0.5 to 1.5 litres of hydrothermal fluids were filtered through Sterivex 0.22 µm filter cartridges (MilliporeSigma) and 15 ml falcon tubes were filled with surficial sediments that were constantly bathed in spring fluids. Both filters and sediment-filled tubes were immediately frozen onsite at −196 °C in a cryogenic dry shipper (Thermo Fisher Scientific, Arctic Express 20) for transport back to the home laboratory. Sites and their GPS coordinates were described previously[16,40]. Field sampling for trace metals in the fluids was carried out by fixing a filtered (0.22 µm) subsample in 5% HNO_3; filtered (0.22 µm) subsamples were also taken for major ion measurements. Samples for the determination of trace elements in the solid-sediment fraction were sampled in 15 ml falcon tubes and stored frozen. Sediments were also sampled for mineralogical analyses, while samples for scanning electron microscopy and cell counts were fixed in 3% formaldehyde and kept at 4 °C.

DNA extraction. High-quality bacterial 16 S rRNA gene libraries were produced for 18 sites; 10 sites worked for both fluids and sediments, 6 were successful only in sediments and 2 more only in fluids. DNA extractions from Sterivex filters were performed using a modified phenol-chloroform extraction optimized for low biomass samples based on previously published methods[41], with additional modifications for use with Sterivex filters as described previously[42]. Briefly, extractions were performed via chemical lysis with lysozyme, proteinase-K and sodium dodecyl sulfate (SDS) treatment, then purified with phenol-chloroform extractions and precipitated with sodium acetate and isopropyl alcohol. Initial extractions from sediment samples were performed using the Qiagen DNeasy PowerSoil HTP 96 Kit, with additional extractions performed using the modified phenol-chloroform extraction described above, followed by concentration using the Zymo Genomic DNA Clean and Concentrator Kit. Extracted DNA was quantified using a NanoDrop 2000c (Thermo Fischer Scientific) with additional polymerase chain reaction screening performed using universal bacterial primers[43,44].

Sedimentary organic matter. Total protein, carbohydrate, lipid, chlorophyll-a and phaeopigments were determined as previously described[45]. Concentrations were calculated using standard curves and normalized to sediment dry weight after desiccation (60 °C, 24 h). Protein, carbohydrate and lipid concentrations were converted into carbon (C) equivalents using the conversion factors of 0.49, 0.40 and 0.75 $\mu gC\,\mu g^{-1}$ dry weight, respectively. Chloroplastic pigment equivalents are defined here as the sum of the chlorophyll-a and phaeopigment concentrations.

Geochemistry. Data for the carbon isotope analysis of DIC and DOC and the concentration of sedimentary aliphatic hydrocarbons and polycyclic aromatic hydrocarbons were previously reported[16,40], along with the methods used to quantify them. Concentrations of anions were determined using a Dionex AS4A-SC separation column, sodium hydroxide eluent and ASRS-I suppressor. For cations, a Dionex CS12-SC separation column was used, with methane sulfonic acid eluent and CSRS-I suppressor. Trace metal concentrations were determined in aqueous and acid-digested solid samples with a NexIon 350X inductively coupled plasma mass spectrometry instrument. Total acid digestion included microwave-assisted digestion of dry sediments with nitric acid (16 N HNO_3) and suprapure hydrofluoric acid and boric acid. The calibration standards were prepared by using PerkinElmer multi-element calibration standard solution of metals (including Fe, Al, As, Mn, Mg, K, V, Cr, Co, Ni, Ca, Mg, Se, Sr, Ga, Ba, Be, Pb and Cs) in 5% HNO_3 with concentrations of 10 $\mu g\,ml^{-1}$ of each element. Internal yttrium standard was added in each sample before analysis to correct the intensity deviations during measurement with inductively coupled plasma mass spectrometry. The molar concentration of each element was calculated by a standard calibration curve of each element with multiplying by volume, dilution and dividing by molar mass. Geochemical data are listed in Supplementary Tables 1–3.

Flow cytometry. Fluids (1 ml) obtained directly from source fluids were placed into a 2 ml plastic tube with a rubber O-ring screwcap (to prevent evaporation) containing 500 µl of 3% paraformaldehyde solution in phosphate-buffered saline pre-filtered at 0.22 µm. Cell-count samples were kept at 4 °C during return to the University of Tennessee and were weighed upon returning to the lab. Cell counts were determined on a Guava easyCyte 6HT-2L (Millipore) flow cytometer. Triplicate aliquots of each sample (200 µl) were stained with 5X Sybr Green before analysis. Gating strategy was optimized using stained, unstained and filtered controls.

Scanning electron microscopy. Scanning electron microscopy micrographs of the hydrothermal sediments were obtained on a Phenom ProX scanning electron microscope at 10 and 15 kV and using a charge-reduction sample holder at Rutgers University. Samples were previously dried at 40 °C for 24 hours before imaging and mounted using conductive carbon tape on a sample pin. The same instrument was used at 15 kV to perform energy-dispersive X-ray spectroscopy for elemental analysis of particles in the samples.

X-ray diffraction and Raman spectroscopy. Each sediment sample was dried at 50 °C for 24 hours. A representative portion of each sample was ground to <10 um grain size with an alumina mortar and pestle. The ground sample material was analysed with a Bruker D8 powder X-ray diffractometer, with a Cu source (1.5406 nm) and 2-theta range of 5° to 70° at 0.01° increments. Bruker EVA software was used to identify mineral phases with pattern search-match performed on the RRUFF database (http://www.rruff.info/) and American Mineralogist Crystal Structure Database (AMCSD) (http://rruff.geo.arizona.edu/) pattern libraries. Raman spectroscopy was performed on selected samples. The crystals were randomly oriented and the Thermo Almega micro Raman system was set at 100% power, using a 532 nm solid-state laser and a thermoelectrically cooled charge-coupled device (CCD) detector. The laser was partially polarized with 4 cm^{-1} resolution and a spot size of 1 µm. Phase identification was performed using the search-match routines available in the Thermo Almega OMNIC and CrystalSleuth software against the RRUFF database Raman spectra library. Trimming and background removal was performed with CrystalSleuth software.

Sequence processing and statistical analysis. Extracted DNA was sequenced for analysis of the bacterial diversity after amplifying the bacteria-specific V4-V5 region of the 16S rRNA gene using primers 518F (CCAGCAGCYGCGGTAAN) and 926R (CCGTCAATTCNTTTRAGT and CCGTCAATTTCTTTGAGT and CCGTCTATTCCTTTGANT) (https://vamps.mbl.edu/resources/primers.php). The same extracted DNA was used for shotgun metagenomic sequencing without amplification. Sequencing was performed as part of the Census of Deep Life initiative within the Deep Carbon Observatory and performed at the Marine Biological Laboratory sequencing facility (https://www.mbl.edu/) on an Illumina MiSeq platform for amplicons and an Illumina NextSeq platform for metagenomes. Amplicon sequences were screened for quality, including chimera checking with UCHIME, by the MBL as previously described, and high-quality merged sequences were published on the Visualization and Analysis of Microbial Population Structures (VAMPS) website[46,47]. Obtained reads were processed using mothur[48], following the MiSeq standard operating procedure using ASVs rather than operational taxonomic units. Taxonomy was assigned using the Ribosomal Database Project (RDP) naïve Bayesian classifier against the SILVA version 132 release[49]. Metagenomic reads were processed with Trimmomatic[50], and then metagenomic short reads were annotated with Mifaser[51] using the Gold Standard Plus (GS+) database (https://bromberglab.org/project/mifaser/) that includes gene sequences from biogeochemically relevant pathways.

All statistical analyses, data processing and plotting were carried out in the R statistical software version 3.6[52], using the phyloseq[53], vegan[54], ggtern[55], missForest[56], VSURF[57] and ggplot2[58] packages. A complete R script containing all the step to reproduce our analysis is available at https://github.com/dgiovannelli/SubductCR_16S-diversity.git and is released as a permanent version (version 1.0) using Zenodo at https://doi.org/10.5281/zenodo.4553845. Briefly, the obtained count table, taxonomy assignment and phylogenetic tree were combined together with the environmental variables into a phyloseq object. Low prevalence ASVs, mitochondria, chloroplast-related sequences and potential contaminants were removed (described in more detail below; Supplementary Fig. 3). In both fluids and sediments, common laboratory contaminants from DNA processing, faeces and skin[59] were largely absent (<0.04% in the entire dataset and less than 0.01% in any individual library), and no ASV was shared by all samples. *Acinetobacter* sp., a group containing hospital-acquired pathogens as well as environmental clades[60], was in high abundance (between 20 and 60% of the reads) in three samples of hydrothermal water collected in spas/resorts and was removed from further analysis. The remaining ASVs represented ~81% of the original reads. ASVs with fewer than five reads when summed across all samples were removed. In total, 1,933,379 reads and 33,188 ASVs were retained after the preprocessing steps. Read counts were normalized to a common scale by transforming them to relative abundance within a sample and then multiplying this proportion by the median library size across all samples. This simplistic normalization is preferable to rarifying data for many of our downstream analyses[61], and does not suffer from high false-discovery rates when library sizes are similar (<10 times different) across all sites[62]. Our library sizes ranged from 47,553 to 272,991 (<5.7 times different) and were all sequenced in the same sequencing run.

Obtained results were used for diversity plots and multivariate analysis. The basic approach involved non-metric multidimensional scaling (NMDS) using Jaccard and Unifrac distances (the latter both weighted and unweighted) on the normalized count data to identify similarity in bacterial diversity community composition across the sampled stations as previously described[45,63]. NMDS ordinations were used to identify potential environmental explanatory variables using linear correlations of environmental vectors with the envfit function in vegan. The roles of different sampling factors in influencing the observed community patterns were tested using a permutation distance-based approach using the ADONIS function of the vegan package. All *P* values were adjusted using the Holm correction for multiple hypothesis testing which reduces the probability of false positives[64]. Tested factors included the sample matrix type (type: fluids or sediments), the subducting plate (plate: EPR or CNS), the location of the sampling site along the volcanic arc (province: outer forearc, forearc or arc), the geological province based on the map reproduced in Fig. 1 (geol_prov: 1, 2, 3 or 4), dominant basement rock type obtained from the USGS Mineral Resources GIS maps[65] (rocks) and the volcanic area the sampling sites are located in (volcano: forearc

and the name of the major Costa Rican volcanoes). Additionally, two factors (anions and cations) were obtained on the basis of the ternary plot of the aqueous geochemical composition of the fluids at each site as presented in Fig. 2b,d. The sampled sites were classified using their position in these plots, based on their geochemical composition, their interaction with different underlying basement rocks and the degree of equilibration[20,21]. Fluids and sediments were analysed separately before combining them into a single analysis. Within-site similarity between fluids and sediments was much higher than between-sample similarity as determined by NMDS analysis (Supplementary Fig. 5), thus only the combined analysis is used below.

Dominant ASVs were obtained by adding a further step of prevalence filtering, removing all the ASVs with a global abundance of less than 20 reads and present in less than 3 samples, following an established approach[27]. This step reduced the number of ASVs to about 12% of the original variants, while retaining ~73% of the total reads. The diversity plots were inspected to ensure that no major changes in the dominant phylotypes and taxonomic groups were introduced (Supplementary Fig. 4). At this step site PL was removed from the dataset because it was the only sample representing a hyperacidic volcanic crater lake and was therefore an outlier in the NMDS analysis (Fig. 2). A co-occurrence network was constructed on the basis of pairwise Spearman rank correlations among the ASVs across the entire dataset. Only positive correlations with a Spearman's correlation coefficient (ρ) >0.7 were retained, as they provide information on microbial phylotypes that may respond similarly to environmental conditions. A Spearman rank correlation was chosen over microbiome-specific correlation metrics[66–68] because, with $n = 27$, a Spearman coefficient equal to or above 0.608 is sufficient[69] to retain relationships that have a $P < 0.001$. This strict alpha level controlled for false-discovery rates while maintaining good power. Furthermore, after prevalence filtering, our dataset violates the underlying assumption of high sparsity in microbiome-specific correlation metrics. The network we recovered included 3,935 nodes with 339,803 edges. The topology of the obtained network was investigated and a modularity analysis using a number of clustering algorithms built in the R package igraph[70] was performed (random walks, label propagation and Louvain clustering algorithms). While the total number of clusters changed, the main clusters identified by the tested algorithms converged and the 10 clusters identified by the Louvain clustering algorithms were retained for downstream analysis. Identified clusters represented ecological cliques of phylotypes showing a cohesive distribution across the sampled hot springs. The relationship between each clique's cumulative abundance and environmental predictors was investigated using an RFs regression analysis. The analysis was carried out using the VSURF package. Clique abundances were re-scaled using z-scores before the RFs analysis and missing environmental observations were imputed using the missForest package. To further test the validity of the identified environmental predictors the cumulative abundance of each clique was correlated using both Pearson moment correlation and Spearman rank correlation against all the environmental predictors. All P values were adjusted using the Holm correction for multiple hypothesis testing[64]. Correlations were also manually inspected using scatterplots to identify possible nonlinear relationships and confirm correlations identified with statistical testing. The variables identified by each different approach (RFs analysis, scatterplot inspection, Pearson moment correlation and Spearman rank correlation) overlapped consistently, and in many cases showed strong collinearity in our dataset (Supplementary Fig. 2). The most biologically informative environmental variable (identified with the above approach and presented in Supplementary Table 13) associated with the distribution of each clique was selected for plotting in Fig. 3b.

For the metagenomic reads, the relative abundance of DNA reads corresponding to each enzyme was calculated by dividing the raw abundance value by the total number of enzyme counts for each sample. Enzyme abundance was then normalized to library size by multiplying the relative abundance by the median library size (65,709,491 bp). Cliques were obtained from pairwise spearman correlations and investigated for their relationship with environmental parameters using the same approach described for the ASV cliques.

Carbon flux calculations. Estimates of total residence time required for the total cellular biomass to be converted to the observed DOC were made with the following:

$$\frac{t_2 \times [\mathrm{DOC}]}{[\mathrm{cells}] \times f_{\mathrm{autotroph}} \times \mathrm{C}_{\mathrm{cells}} \times f_{\mathrm{attached}}} \quad (1)$$

where t_2 is the turnover time of the population (17 years, a conservative estimate because it was calculated for marine sediments[71] which may have slower turnover times since they lack advective fluid/nutrient fluxes), [DOC] is the mean concentration of DOC across all sites from the values listed in Supplementary Table 1 (1.15 mmolC l^{-1}), [cells] is the mean concentration of cells in fluids across all sites (7.7×10^5 cells ml^{-1}), $f_{\mathrm{autotroph}}$ is the fraction of cells that are autotrophs (30% based on transcriptomic and proteomic data from the deep terrestrial subsurface[5,6]) and $\mathrm{C}_{\mathrm{cells}}$ is the carbon content of microbial cells (23 fgC per cell, a conservative estimate since it is based on extremely energy-starved marine sediment populations[71] and is only 25% of values used in other biomass calculations[72]. f_{attached} is a multiplication factor to account for the fact that many more cells are likely to be attached to particles than are planktonic in the fluids that were sampled. The values for f_{attached} were determined through a meta-analysis[73] of many published estimates to range from 100 to 1,000.

Residence time of fluids in Costa Rican forearc springs was calculated with the following:

$$\frac{V \times \theta}{N_{\mathrm{springs}} \times J_{\mathrm{springs}}} \quad (2)$$

Where V is the habitable (that is, <122 °C) volume of the Costa Rican subsurface forearc from trench to volcanic arc. The area of the forearc, 20,511 km^2, was calculated with Google Earth from the Nicoya coastline to the volcanic line in the northern/central section of Costa Rica that has arc volcanoes. The depth of the 122 °C isotherm was determined from heat-flow modelling of this region[35]. To account for the low fraction of habitable pore space in in crystalline bedrock, we assumed an average pore fraction, θ, equal to 0.05, or 5% (assuming reference values of 4–11% for basalt, 4–10% for granite and 18–51% for ignimbrite). This is probably an underestimate, considering that the majority of Costa Rica's forearc is constituted by heavily altered igneous rocks. The number of hot springs in this area of Costa Rica (N_{springs}) ranges from 50 to 500, and the flow rate from these springs, J_{springs}, ranges from 1–10 $l\,s^{-1}$, with both ranges obtained from ref. [16].

Total microbial cellular biomass carbon, $\mathrm{C}_{\mathrm{biomass}}$, in the Costa Rican forearc was calculated with the following:

$$V_{2\mathrm{km}} \times \theta \times [\mathrm{cells}] \times f_{\mathrm{attached}} \times \mathrm{C}_{\mathrm{cells}} \quad (3)$$

following the method of McMahon and Parnell[73], who determined that the average cell abundance in fluids applied only over the upper 2 km accounts for a decrease with depth. $V_{2\mathrm{km}}$ denotes the volume of the subsurface of this section of the arc down to 2 km. However, since the 122 °C isotherm is shallow at the arc volcanoes, the full depth should not be used uniformly across the region. Therefore, we decreased the value logarithmically from 2 km at the Nicoya coastline to 0 km depth at the arc volcanoes, which matches the heat-flow models for this region[35].

Percentage of mantle input sequestered as biomass in the Costa Rican forearc subsurface was therefore:

$$\frac{\mathrm{C}_{\mathrm{biomass}} \times f_{\mathrm{autotrophs}}}{\mathrm{C}_{\mathrm{mantle}} \times t_2} \quad (4)$$

Where $\mathrm{C}_{\mathrm{mantle}}$ is the carbon flux to the mantle determined for this forearc[16] of 6.3×10^{10} molC yr^{-1} (ref. [16]).

The total microbial cellular biomass carbon, $\mathrm{C}_{\mathrm{biomass}}$, in the Costa Rican forearc was estimated a second way, using published decay rates of cell densities with depth in the subsurface of arc environments[1] with the general form of log(cells) $= A + B \times \log(\mathrm{depth})$. Values for arc crust ($A = 14.16$ cells cm^{-3}, $B = -0.94$) were used. Magnabosco et al.[1] report 95% prediction intervals for that fit of $A = 8.00$ to 8.45 and $B = -1.01$ to -0.81. Since the definite integral of $10^A \times \mathrm{depth}^B$ is not defined for $B \leq -1$, we converted the 95% intervals (median $\pm 2\sigma$) to 68% intervals (median $\pm \sigma$). Cell density was integrated from the surface to the depth of the 122 °C isotherm, the depth of which was calculated by linearly interpolating the isotherms reported for the same region by Harris and Wang[35]. An R notebook describing the calculation in more detail is available at the permanent repository: https://doi.org/10.5281/zenodo.4553845.

Data availability

This Targeted Locus Study project has been deposited at DDBJ/EMBL/GenBank under the accession KEBJ00000000, with project ID PRJNA579365. The version described in this paper is the first version, KEBJ01000000. Metagenomic data are in the NCBI SRA with project ID PRJNA627197. The full environmental dataset is available at https://github.com/dgiovannelli/SubductCR_16S-diversity.git and released as a permanent version (v1.0) using Zenodo under https://doi.org/10.5281/zenodo.4553845. Source data are provided with this paper.

Code availability

A complete R script containing all the steps to reproduce our analysis, including the full environmental dataset, is available at https://github.com/dgiovannelli/SubductCR_16S-diversity.git and released as a permanent version (v1.0) using Zenodo under https://doi.org/10.5281/zenodo.4553845, https://doi.org/10.5281/zenodo.3483104.

References

40. Barry, P. H. et al. Helium, inorganic and organic carbon isotopes of fluids and gases across the Costa Rica convergent margin. *Sci. Data* https://doi.org/10.1038/s41597-019-0302-4 (2019).
41. Vetriani, C., Jannasch, H. W., MacGregor, B. J., Stahl, D. A. & Reysenbach, A.-L. Population structure and phylogenetic characterization of marine benthic archaea in deep-sea sediments. *Appl. Environ. Microbiol.* **65**, 4375–4384 (1999).

42. Wright, J. J., Lee, S., Zaikova, E., Walsh, D. A. & Hallam, S. J. DNA extraction from 0.22 µm Sterivex filters and cesium chloride density gradient centrifugation. *JOVE* https://doi.org/10.3791/1352 (2009).
43. Teare, J. M. et al. Measurement of nucleic acid concentrations using the DyNA Quant™ and the GeneQuant™. *Biotechniques* **22**, 1170–1174 (1997).
44. Simbolo, M. et al. DNA qualification workflow for next generation sequencing of histopathological samples. *PLoS ONE* **8**, e62692 (2013).
45. Giovannelli, D. et al. Diversity and distribution of prokaryotes within a shallow-water pockmark field. *Front. Microbiol.* **7**, 941 (2016).
46. Huse, S. M. et al. VAMPS: a website for visualization and analysis of microbial population structures. *BMC Bioinformatics* **15**, 41 (2014).
47. Huse, S. M. et al. Comparison of brush and biopsy sampling methods of the ileal pouch for assessment of mucosa-associated microbiota of human subjects. *Microbiome* **2**, 5 (2014).
48. Schloss, P. D. et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
49. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).
50. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
51. Zhu, C. et al. Functional sequencing read annotation for high precision microbiome analysis. *Nucleic Acids Res.* **46**, e23 (2018).
52. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2013).
53. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).
54. *vegan* (CRAN, 2019).
55. Hamilton, N. E. & Ferry, M. ggtern: ternary diagrams using ggplot2. *J. Stat. Softw.* **87**, 1–17 (2018).
56. Stekhoven, D. J. & Buhlmann, P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28**, 112–118 (2012).
57. Genuer, R., Poggi, J.-M. & Tuleau-Malot, C. VSURF: an R package for variable selection using random forests. *R J.* **7**, 1–19 (2015).
58. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).
59. Sheik, C. S. et al. Identification and removal of contaminant sequences from ribosomal gene databases: lessons from the Census of Deep Life. *Front. Microbiol.* **9**, 840 (2018).
60. Sugimori, K. et al. Microbial life in the acid lake and hot springs of Poas Volcano, Costa Rica. In *Proc. Colima Volcano International Meeting* (2002).
61. Mcmurdie, P. J. & Holmes, S. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* **10**, e1003531 (2014).
62. Weiss, S. et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 27 (2017).
63. Giovannelli, D. et al. Large-scale distribution and activity of prokaryotes in deep-sea surface sediments of the Mediterranean Sea and the adjacent Atlantic Ocean. *PLoS ONE* **8**, e72996 (2013).
64. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70 (1979).
65. Schruben, P. G. *Geology and Resource Assessment of Costa Rica* DDS-19-R (USGS, 1987).
66. Friedman, J. & Alm, E. J. Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* **8**, e1002687 (2012).
67. Kurtz, Z. D. et al. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput. Biol.* **11**, e1004226 (2015).
68. Schwager, E., Mallick, H., Ventz, S. & Huttenhower, C. A Bayesian method for detecting pairwise associations in compositional data. *PLoS Comput. Biol.* **13**, e1005852 (2017).
69. Zar, J. H. Significance testing of the spearman rank correlation coefficient. *J. Am. Stat. Assoc.* **67**, 578–580 (1972).
70. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal* **1695**, 1–9 (2006).
71. Braun, S. et al. Microbial turnover times in the deep seabed studied by amino acid racemization modelling. *Sci. Rep.* **7**, 5680 (2017).
72. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: the unseen majority. *Proc. Natl Acad. Sci. USA* **95**, 6578–6583 (1998).
73. McMahon, S. & Parnell, J. Weighing the deep continental biosphere. *FEMS Microbiol. Ecol.* **87**, 113–120 (2013).

Acknowledgements

This work is part of the Biology Meets Subduction project, a collaboration of 46 researchers from 19 institutions from 9 nationalities. We thank P. Barcala Dominguez for assistance with figure illustrations, and T. Hoehler for advice. Principal support came from the Alfred P. Sloan Foundation and the Deep Carbon Observatory (G-2016-7206) to P.H.B., J.M.d.M, D.G. and K.G.L., with DNA sequencing from the Census of Deep Life. Additional support came from NSF OCE-1431598, NASA Exobiology NNX16AL59G and Simons Foundation 404586 to K.G.L., NSF 1144559 to P.H.B., NSF 1850699 to J.M.d.M., NSF MCB 15–17567 to D.G. and C.V., ELSI Origins Network (EON) Research Fellowship from the John Templeton Foundation to D.G., Deep Life Modeling and Visualization Fellowship from the Deep Carbon Observatory to D.G., FONDECYT Grant 11191138 (ANID Chile) to G.L.J., ENIGMA (NASA Astrobiology Institute cycle 8, 80NSSC18M0093) to D.G., S.M.M. and J.B, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (DE-SC0020369) to A.D.S. and K.G.L., JSPS KAKENHI grants JP17K14412, JP17H06105 and JP17H02989 to M.N. and DEKOSIM grant BAP-08-11-DPT.2012K120880, financed by the Strategy and Budget Ministry of Turkey, to M.Y. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

Author contributions

K.M.F. and K.G.L. wrote the original draft and performed initial data analyses. Conceptualization and funding acquisition were performed by P.H.B., J.M.d.M., D.G. and K.G.L. Formal analysis and visualization were performed by K.M.F. and D.G. Investigations and data acquisition were performed by K.M.F., M.Y., E.M., G.D., D.F., M.D.C., F.R., M.N., F.S., H.M., S.M.M., T.J.R., M.B., J.B., A.D.S. and D.G. Writing and editing of the final draft was performed by M.O.S., M.Y., M.N., C.V., C.R., G.L.J., H.M., T.J.R., M.M., J.B., J.M.d.M., P.H.B., A.D.S., D.G. and K.G.L.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41561-021-00725-0.

Correspondence and requests for materials should be addressed to D.G. or K.G.L.

Peer review information Primary Handling Editor(s): Rebecca Neely. *Nature Geoscience* thanks Jan Amend and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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