

ARTICLE



Chemolithoautotroph distributions across the subsurface of a convergent margin

Timothy J. Rogers¹, Joy Buongiorno², Gerhard L. Jessen^{3,4}, Matthew O. Schrenk⁵, James A. Fordyce¹, J. Maarten de Moor^{6,7}, Carlos J. Ramírez⁸, Peter H. Barry⁹, Mustafa Yücel¹⁰, Matteo Selci¹¹, Angela Cordone¹¹, Donato Giovannelli^{11,12,13,14} and Karen G. Lloyd¹✉

© The Author(s), under exclusive licence to International Society for Microbial Ecology 2022

Subducting oceanic crusts release fluids rich in biologically relevant compounds into the overriding plate, fueling subsurface chemolithoautotrophic ecosystems. To understand the impact of subsurface geochemistry on microbial communities, we collected fluid and sediments from 14 natural springs across a ~200 km transect across the Costa Rican convergent margin and performed shotgun metagenomics. The resulting 404 metagenome-assembled genomes (MAGs) cluster into geologically distinct regions based on MAG abundance patterns: outer forearc-only (25% of total relative abundance), forearc/arc-only (38% of total relative abundance), and delocalized (37% of total relative abundance) clusters. In the outer forearc, *Thermodesulfovibrionia*, *Candidatus Bipolaricaulia*, and *Firmicutes* have hydrogenotrophic sulfate reduction and Wood-Ljungdahl (WL) carbon fixation pathways. In the forearc/arc, *Anaerolineae*, *Ca. Bipolaricaulia*, and *Thermodesulfovibrionia* have sulfur oxidation, nitrogen cycling, microaerophilic respiration, and WL, while *Aquificae* have aerobic sulfur oxidation and reverse tricarboxylic acid carbon fixation pathway. Transformation-based canonical correspondence analysis shows that MAG distribution corresponds to concentrations of aluminum, iron, nickel, dissolved inorganic carbon, and phosphate. While delocalized MAGs appear surface-derived, the subsurface chemolithoautotrophic, metabolic, and taxonomic landscape varies by the availability of minerals/metals and volcanically derived inorganic carbon. However, the WL pathway persists across all samples, suggesting that this versatile, energy-efficient carbon fixation pathway helps shape convergent margin subsurface ecosystems.

The ISME Journal (2023) 17:140–150; <https://doi.org/10.1038/s41396-022-01331-7>

INTRODUCTION

Convergent margins are regions where two or more tectonic plates collide, and the denser “subducting slab” is pushed beneath the less dense overriding plate. As the slab descends, it devolatilizes under higher temperatures and pressures, allowing dissolved inorganic carbon (DIC) and redox-active volatiles to cycle between the upper crust and Earth’s mantle [1] (Fig. 1A). These volatiles are brought to the surface by hydrothermal fluids migrating through cracks and fissures in the upper mantle and crust, possibly stimulating microbial activity [2]. In addition to providing volatiles, convergent margins offer a large potential space for a subsurface habitat since the 122 °C sterilizing isotherm (Fig. 1A) can be many kilometers deep [3, 4].

At the Costa Rican convergent margin, the Cocos Plate subducts beneath the Caribbean Plate, releasing fluids entrained with volatiles such as sulfur, iron, hydrogen, and DIC [5, 6]. These fluids fuel chemolithoautotrophy-based microbial ecosystems in the

subsurface before they are expelled at the surface at hydrothermal seeps and springs [7]. These chemolithoautotrophic ecosystems generate substantial amounts of organic matter with up to 24 mM dissolved organic carbon (DOC), that may sequester 1.4×10^9 to 1.4×10^{10} molC/year [7]. This is similar to the amount of carbon estimated to be sequestered by calcite deposition in the overriding plate [8]. However, carbon sequestration by microbial autotrophy is linked to other elemental cycles in a different way than calcite precipitation is, since chemolithoautotrophs derive energy for carbon fixation by catalyzing a wide variety of exergonic redox reactions [9].

Using an approach that combines co-located metagenome sequences and geochemical data, we investigated whether chemolithoautotrophic organisms, their encoded carbon fixation pathways (CFP), and/or the types of redox reactions they potentially use to fuel carbon fixation vary across the major geological provinces of the Costa Rican convergent margin (outer forearc,

¹University of Tennessee, Knoxville, TN, USA. ²Division of Natural Sciences, Maryville College, Maryville, TN, USA. ³Instituto de Ciencias Marinas y Limnológicas, Universidad Austral de Chile, Valdivia, Chile. ⁴Center for Oceanographic Research COPAS COASTAL, Universidad de Concepción, Valdivia, Chile. ⁵Michigan State University, East Lansing, MI, USA. ⁶National University of Costa Rica, Heredia, Costa Rica. ⁷University of New Mexico, Albuquerque, NM, USA. ⁸Servicio Geológico Ambiental, Heredia, Costa Rica. ⁹Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, USA. ¹⁰Institute of Marine Sciences, Middle East Technical University, Erdemli, Turkey. ¹¹Department of Biology, University of Naples — Federico II, Naples, Italy. ¹²Department of Marine and Coastal Science, Rutgers University, Rutgers, NJ, USA. ¹³Earth-Life Science Institute, Tokyo Institute of Technology, Tokyo, Japan. ¹⁴National Research Council – Institute of Marine Biological Resources and Biotechnologies – CNR-IRBIM, Ancona, Italy. ✉email: kloyd@utk.edu

Received: 21 March 2022 Revised: 3 October 2022 Accepted: 6 October 2022

Published online: 18 October 2022

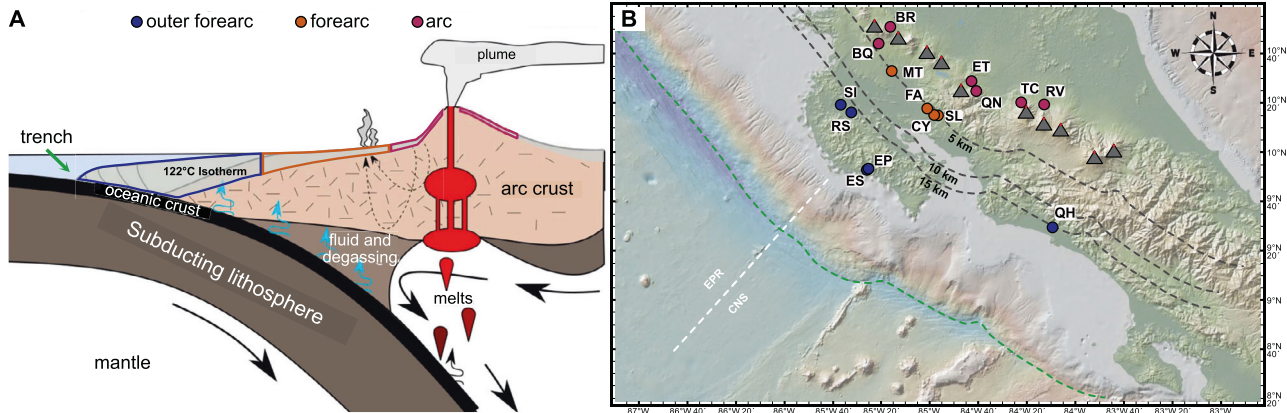


Fig. 1 General convergent margin geological structure and sample sites spanning the Costa Rican convergent margin. Geological provinces (A) and sample sites (B) are colored by their province location: outer forearc (blue), forearc (orange), and arc (red). A The green arrow indicates the location of trench, black arrows show the direction of plate movements, blue arrows show fluids and degassing throughout subduction. B Green dotted line indicates the trench while gray dotted lines mark the depth of the 122 °C isotherm in 5 km increments, based on data for this region of Costa Rica from Harris and Wang 2010. Locations of volcanoes are shown with volcano icons. Dotted white line demarcates the plates emerging from the East Pacific Rise (EPR) and the Cocos Nazca Spreading center (CNS).

forearc, and volcanic arc). The outer forearc is the province closest to the trench where the downgoing oceanic plate subducts beneath the overriding plate. Here, the subducting slab is relatively shallow and seawater along with pore fluids and serpentinization products are released from subducted material [10–12]. In the forearc, farther inland from the trench, deeply derived volatiles that flux to the surface in natural seep fluids have likely undergone some hydrothermal alteration and mixing with mantle-derived fluids. The arc consists of volcanic edifices farther inland parallel to the trench [5] typically coinciding with the point where the slab is approximately 90–110 km deep. Here, slab fluids and melts interact with the mantle wedge to generate magmas [13, 14].

Subsurface microorganisms in other convergent margins have been shown to perform carbon fixation as well as hydrogen, sulfur, and nitrogen cycling, but this has only been studied within single provinces of a convergent margin such as the backarc [15–17] or outer forearc [18]. Landscape-scale (defined here as tens to hundreds of kilometers) distributions of chemolithoautotrophic metabolisms have been observed in intraplate natural springs in Yellowstone National Park, USA [19–21], and Tengchong, Yunnan, China [22], but not across a subducting margin. Here, we expand our understanding of the geobiological underpinnings of carbon release and sequestration across a convergent margin.

We report 404 metagenome-assembled genomes (MAGs) from 14 natural deeply sourced springs (27 samples total) distributed across a ~200 km transect of the Costa Rican convergent margin spanning the outer forearc, forearc, and arc provinces (Fig. 1B and Supplementary Fig. S1). These springs are deeply sourced as their high helium-3/helium-4 values indicate substantial mantle contributions to all samples [8]. A previous study of the unassembled metagenomes from these springs indicated the presence of genes related to the reductive acetyl-CoA pathway (Wood-Ljungdahl; WL), the reductive tricarboxylic acid cycle (rTCA), and the reductive pentose phosphate cycle (Calvin Benson–Bassham Cycle; CBB) for carbon fixation [7]. However, unassembled metagenomic reads cannot be used to directly relate CFPs to the taxa that have them in their genomes, nor their redox-active metabolic pathways. Here, we show that subsurface- and surface-associated MAGs are distinguished by their distributions and metabolic pathways. Genes encoding for WL and rTCA are abundant across the margin but are encoded by MAGs with different taxonomic affiliations. In addition, these CFP are potentially powered by different redox metabolisms across different geological provinces. Together this suggests a

strong geological control over the distribution of different metabolic strategies across the convergent margin landscape.

METHODS

Sample collection, DNA extraction, and sequencing

Details on collection, DNA extraction, and sequencing of samples from actively venting fluids of 14 natural springs (Fig. 1B) have been described in detail previously [7, 8].

Metagenomic assembly, binning, annotation, and metabolic predictions

Raw reads were trimmed and mate-paired by Trimmomatic (version 0.38 [23]) using default settings. Reads were then assembled de novo in metaSPAdes [24] to create site-specific assemblies with a minimum contig length of 1.5 kb. Assemblies were then used to construct 404 metagenomic assembled genomes (MAGs/bins) with $\geq 70\%$ completeness and $< 5\%$ contamination (Supplementary Table S1) [25–30] through the MetaWRAP pipeline [31]. Prokka [32], FeGenie [33], GhostKOALA [34], and KEGGdecoder [35] were used for open reading frame prediction, gene annotation, KEGG orthology identifier linkage, and metabolic predictions. The genetic potential to use hydrogen, sulfur, nitrogen, and/or iron species as electron donors or acceptors was assessed based on the presence of the enzymatic components detailed in the Supplementary Methods (Supplementary Table S2). Elemental sulfur oxidation and sulfate reduction were determined by phylogenetic analyses of the dissimilatory (bi)sulfite reductase alpha subunit (*dsrA* (Supplementary Fig. S2)). In addition, a MAG was considered capable of autotrophy if 60% of the genes and all key enzymes for a CFP were present. As with *dsrA*, phylogenetic analysis was conducted on the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) large subunit to delineate the different RuBisCO forms of CBB (Supplementary Fig. S3). In-depth details for metagenomic assembly, binning, annotation, and metabolic predictions are given in Supplementary Methods.

MAG phylogenetic and statistical analysis

The Genome Taxonomy Database Toolkit (GTDB-TK [36]; version 1.4.0) was used for taxonomic identification and GTDB nomenclature has been used throughout this manuscript, although we refer to other commonly used names when it increases clarity. MAG abundance was quantified in units of GCPM (Genome Copy Per Million Reads). An in-depth description of how GCPM was calculated is given in Supplementary Methods. Abundance correlations across and within samples were determined using hierarchical clustering of Spearman rank correlation distance matrices (Fig. 2 and Supplementary Fig. S1). MAG abundance distributions across the convergent margin and their correlations with relevant environmental variables [7] were then determined via transformation-based canonical correspondence analysis (tb-CCA). See Supplementary Methods for detailed information about the statistical analysis.

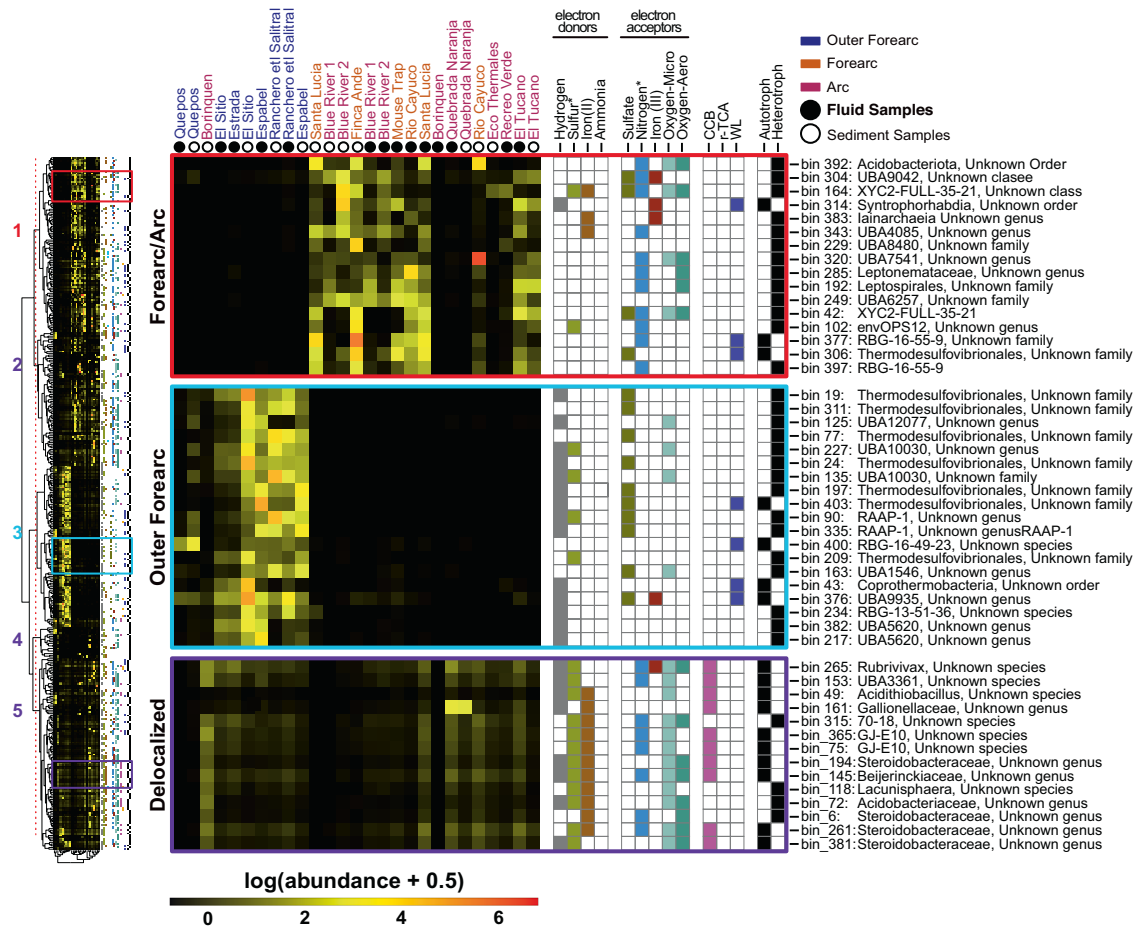


Fig. 2 Heatmap of metagenomic read recruitment to each MAG and hierarchical clustering based on Spearman rank correlations between MAGs (vertical clusters) and sample sites (horizontal clusters) shows the separation of MAGs into those found mostly in the forearc/arc, the outer forearc, or delocalized across geological provinces. Entire heatmap is shown to the left, with numbers corresponding to clusters; and representative enlarged sections are shown to the right, with metabolic pathways for redox couples and carbon fixation pathways listed. Subsections were chosen based on inclusion of the key metabolisms and taxa that characterize province-specific clusters (clusters 1 and 3) and delocalized clusters (clusters 2, 4, and 5). Sulfur* and nitrogen* refer to the metabolic capability of metabolizing multiple oxidation states. The full heatmap in high resolution is in Supplementary Fig. S1.

RESULTS

General description of the MAGs

In total, 404 medium to high-quality MAGs spanning 5 archaeal and 43 bacterial phyla were recovered from the 27 samples of hydrothermal fluids and sediment. The 372 bacterial and 32 archaeal MAGs are $\geq 70\%$ complete with $< 5\%$ multi-genomic contamination. Of these, 221 MAGs are $> 90\%$ complete, with $< 5\%$ contamination and 43 of these are high-quality based on MIMAGs criteria (25 (Supplementary Table S1)). Most MAGs belong to lineages with no cultured relatives at the family level or higher (69% of archaea and 64% of bacteria). One archaeal MAG was classifiable only at the domain level.

Most MAGs (92%) have predictions for at least one metabolic pathway (Fig. 2 and Supplementary Fig. S1). Of the MAGs with none of the genes of interest for the electron donors or acceptors, 13 are archaea (*Crenarchaeota* = 4, *Micrarchaeota* = 4, *Nanoarchaeota* = 4, and *Thermoplasmatota* = 1) and 19 are bacteria (*Actinobacteriota* = 1, *Chloroflexota* = 2, *Desulfobacterota* = 1, *Nitrospirota* = 1, *Omnitrophota* = 1, *Patescibacteria* = 8, and *WOR-3* = 4). Neither complete methanogenic pathways nor the methyl coenzyme M reductase gene are present in any of the MAGs. Photosystems are present only in the four *Cyanobacteria* MAGs, accounting for 0.2% of total GCPM.

CBB, rTCA, and WL pathways are present in about one-third of the MAGs ($n = 117$, $\sim 35\%$ of total GCPM). The most abundant is WL ($n = 67$, 19% of total GCPM), followed by CBB ($n = 42$, 8% of

total GCPM), and rTCA ($n = 12$, 7% of total GCPM) (Supplementary Fig. S4). Although three MAGs (bins 112, 253, and 91) contain $\geq 60\%$ of the 3-hydroxypropionate/4-hydroxybutyrate cycle, none contain the key enzymes for this pathway [37, 38]. Three *Anaerolineae* MAGs (bins 55, 251, and 386) contain both WL and CBB, and have RuBisCO (for CBB) and carbon monoxide dehydrogenase/acetyl-CoA synthase (CODH/ACS (for WL)). These genes are closely related to others from *Chloroflexi*, suggesting the co-occurrence of these pathways is not due to genomic contamination. Another MAG in the class *Syntrophia* (bin 351) contains genes for both WL and rTCA. Genes for the WL pathway are found in MAGs from 22 classes across 12 different phyla (*Actinobacteria*, *Bipolaricaulota*, *Chloroflexota*, *Deferrisomatota*, *Desulfobacterota*, *Firmicutes*, MBNT15, *Acidobacteriota*, *Coprothermobacterota*, *Nitrospirota*, *Omnitrophota*, and *Planctomycetota*). Genes for the rTCA cycle are found in MAGs from 4 classes across 4 phyla (*Aquificota*, *Chloroflexota*, *Desulfobacterota*, and *Nitrospirota*). Genes for the CBB cycle are found in MAGs from 8 classes across 6 phyla (*Chloroflexota*, *Proteobacteria*, *Actinobacteriota*, *Cyanobacteria*, *Deinococcota*, and *Firmicutes*), with RuBisCO Form I (25 MAGs), Form II (9 MAGs), both Form I and II (3 MAGs), or an unknown form (4 MAGs) (Supplementary Fig. S3).

Together, WL-containing MAGs have the most diverse metabolic predictions, containing all the metabolic pathways that were identified in this study in some combination. Among the

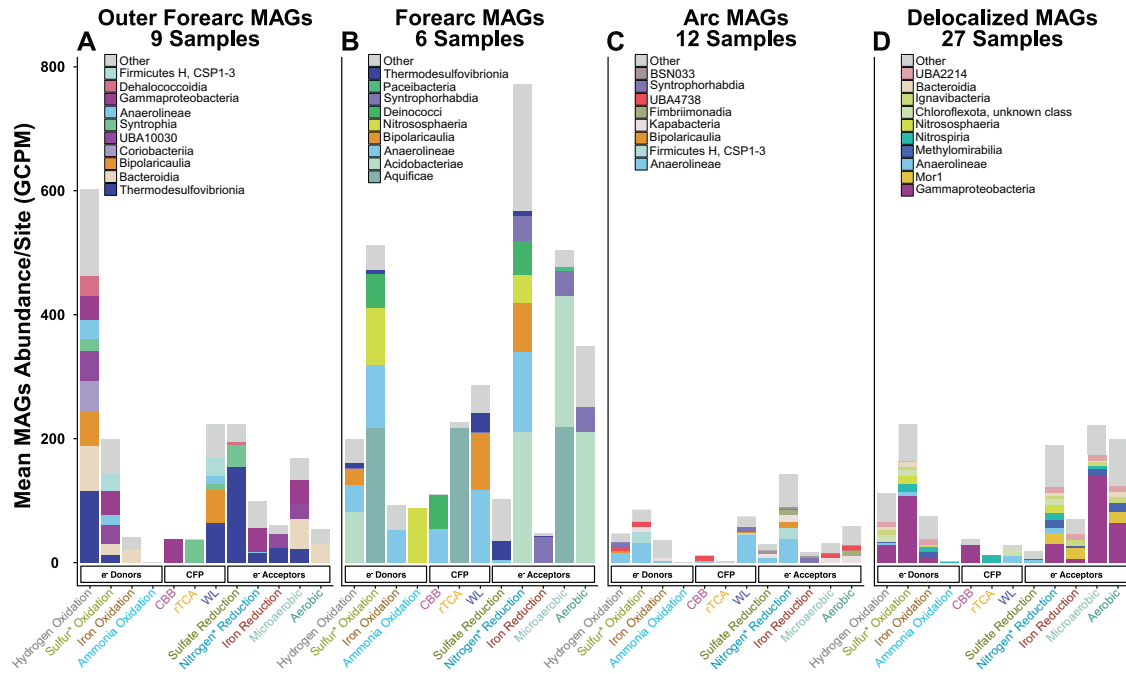


Fig. 3 Average read abundance (GCPM) per site of MAGs with metabolic pathways for redox reactions and carbon fixation pathways. **A–C** show the province-specific average read recruitment of MAGs with these metabolic pathways at all **A** outer forearc sites, **B** forearc sites, and **C** arc sites, after removing MAGs in the delocalized cluster, which are in panel **D**. Colors correspond to the top 10 most abundant classes for each province and the delocalized community. Panels **B** and **C** show top 9 and 8 most abundant classes, respectively, since MAGs lacking redox genes and carbon fixation pathways were in the top 10 most abundant classes. Bars represent the GCPM mean for MAGs at each province specific site.

WL-containing MAGs, the most abundant have the genetic potential for the oxidation of hydrogen and sulfide, and some possess *bd* and *aa₃*-type cytochromes. *rTCA*-containing MAGs have the genetic potential for every redox pathway except nitrogen cycling and iron (III) reduction. In this category, the most abundant MAGs have the genetic potential for sulfide and thiosulfate oxidation and the reduction of oxygen. *CBB*-containing MAGs have the genetic potential for every redox pathway except sulfate reduction. These MAGs have the genetic potential for hydrogen, sulfide, and thiosulfate oxidation and the reduction of oxygen (microaerophilic and/or aerobic) and nitrate.

Cluster and multivariate analysis

Hierarchical clustering of sites based on Spearman correlations of MAG abundance resulted in all the outer forearc sites clustering together and most of the forearc and arc sites clustering together (Fig. 2 and Supplementary Fig. S1, horizontal sorting). The only exception is the sediment sample from Borinquen, as discussed below. Of the 10 sites where metagenomes were sequenced from both fluids and sediments, four have their fluid and sediment samples clustering together (El Tucano, Quebrada Naranja, Rancho el Saltral, and Quepos). Four other sites in the forearc/arc (Blue River 1 and 2, Santa Lucia, and Finca Asociación Nacional de Educadores y Educadoras (ANDE)), have their fluid MAGs clustering together and their sediments clustering adjacent to the fluid MAGs (Fig. 2 and Supplementary Fig. S1). This suggests that fluids and sediments from the same site are generally more similar to each other than to fluids and sediments in different pools, in agreement with 16S rRNA gene results [7]. MAGs cluster into five groups, based on their distribution across the sites, with a linkage height of 10 (Fig. 2 and Supplementary Fig. S1, vertical sorting). Cluster 1 MAGs are mainly in the forearc/arc, cluster 2 MAGs are abundant in the forearc/arc but are also common elsewhere, cluster 3 MAGs are mainly in the outer forearc, and clusters 4 and 5 MAGs are either not localized to any geologic province or are

present in only a few sites with no clear distribution pattern across the convergent margin. Based on these clustering patterns, MAGs separate into three groups: two province-specific groups (cluster 1 from forearc/arc and cluster 3 from outer forearc) and delocalized (clusters 2, 4, and 5).

Metabolic and taxonomic descriptions of the clusters

Outer forearc (cluster 3). Cluster 3 comprises 94 MAGs, 19 phyla, and 25% of total GCPM, almost exclusively from outer forearc sites. *Bacteroidota*, *Nitrospirota*, and *Chloroflexota* dominate numbers of MAGs ($n = 21, 16,$ and 13 respectively), and 19%, 21%, and 9% of cluster 3 GCPM, respectively. The majority of cluster 3 MAGs (73%) are from uncultured families or higher taxonomic levels (Supplementary Fig. S5). Cluster 3 MAGs metabolic pathway predictions are dominated by hydrogen (79% of cluster 3 GCPM) and sulfur (26% of cluster 3 GCPM) oxidation, as well as sulfate reduction (29% of cluster 3 GCPM), microaerophilic respiration (22% of cluster 3 GCPM) and nitrogen reduction (13% of cluster 3 GCPM) (Fig. 3). Genomic material from the top 10 classes is predominantly found in geothermal areas, aquifers, and other deep subsurface systems (Supplementary Table S3).

WL is the most abundant and diverse CFP, encompassing 23% of MAGs and 29% of cluster 3 GCPM. *Thermodesulfovibrionia* MAGs form the dominant WL-containing group and have the genetic potential for hydrogenotrophic sulfate reduction. A single MAG classified as *Ca. Bipolaricaulia* (bin 327) is the second most abundant WL-containing member which also has the genetic potential for hydrogenotrophic growth [39]. Four MAGs classified as CSP1-3 (in phylum *Armatimonadetes*), the third most abundant WL-containing group, have the genetic potential for sulfur oxidation (Fig. 4). Together, these cluster 3 WL-containing members encompass 11 MAGs and 19% of cluster 3 GCPM. In this cluster, genes for *rTCA* and *CBB* are in lower abundance (5% each of cluster 3 GCPM). *Syntrophia* MAGs dominate the *rTCA*-containing MAGs and have the genetic potential for sulfate

Energy & CFP Coupling

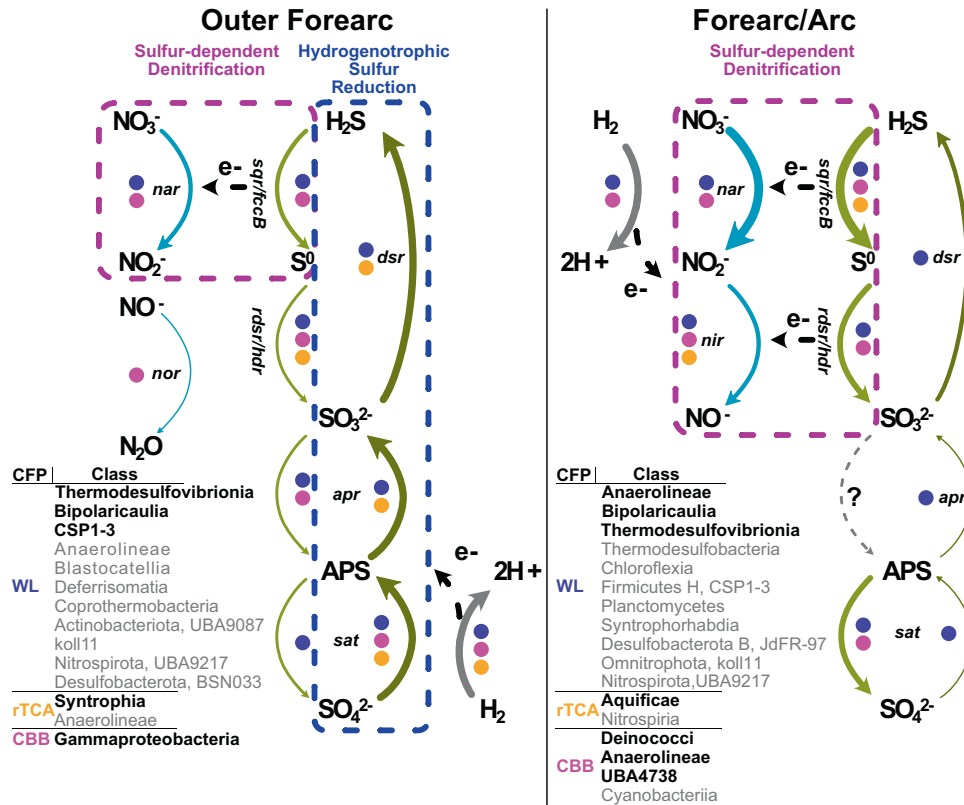


Fig. 4 Nitrogen, sulfur, and hydrogen cycling pathways in MAGs with carbon fixation pathways from the outer forearc and the forearc/arc. Arrow width corresponds to abundance of MAGs with that gene and dots show which CFP pathways were present in the MAGs that had that gene, with blue for WL, yellow for rTCA, and pink for CBB. Class names are listed from most to least abundant (top to bottom), and MAGs in the top 3 CFP-containing classes are in bold. Pink dashed box highlights sulfur-associated chemoautotrophic nitrogen reduction. Blue dashed box highlights hydrogenotrophic sulfur reduction. Gray dotted line with a question mark represents missing *aprBA* predictions for all sulfide-oxidizing MAGs.

reduction (Fig. 4). Most CBB-containing MAGs in cluster 3 classify as *Gammaproteobacteria* and encode RuBisCO Form IA. This form is the primary form in chemolithoautotrophs from a diversity of subsurface ecosystems [40]. These MAGs have the genetic potential for sulfide oxidation and reduction of nitrate/nitrite [41] and/or oxygen (microaerophilic) (Fig. 4).

Forearc/arc (cluster 1). Cluster 1 comprises 102 MAGs, 34 phyla, and 38% of total GCPM, mainly in forearc/arc sites. *Chloroflexota*, *Acidobacteriota*, and *Aquificota* dominate this cluster (~15%, 14%, and 12% of cluster 1 GCPM, respectively). *Chloroflexota* MAGs are the most diverse of these ($n = 15$), then *Acidobacteriota* MAGs ($n = 8$), while there is only one MAG classified as *Aquificota* (bin 303, *Sulfurihydrogenibium* sp.). This cluster has most of the archaeal MAGs ($n = 19$) and 88% of the MAGs are taxonomically identified as uncultured families or higher taxonomic levels (Supplementary Fig. S5). The cluster contains MAGs with many redox metabolisms, predominantly nitrogen (60% of cluster 1 GCPM) and oxygen reduction (microaerophilic = 32% and aerobic = 26% of cluster 1 GCPM), as well as sulfur (39% of cluster 1 GCPM) and hydrogen (16% of cluster 1 GCPM) oxidation. Forearc sites contain the only MAGs capable of ammonia oxidation ($n = 4$; 5% of cluster 1 GCPM) within the archaea class *Nitrososphaeria* (Figs. 2 and 3). DNA sequences from previously sequenced representatives of the top 10 classes are predominantly found in geothermal areas, aquifers, and other deep subsurface systems as well as soils (Supplementary Table S3).

As with cluster 3, WL is the most abundant CFP in cluster 1 ($n = 24$, 25% of cluster 1 GCPM), and is mostly found in *Anaerolineae*

MAGs with the genetic potential for sulfur-dependent reduction of nitrate and/or nitrite. *Ca. Bipolaricaulia* MAGs, the second most abundant group of WL-containing MAGs, have the genetic potential for hydrogen oxidation and the reduction of nitrogen species. All *Thermodesulfovibrionia* MAGs, the next most abundant group of WL-containing MAGs, contain all genes for sulfate reduction (*dsrAB*, *aprAB*, and *sat*). Together, the WL-containing members of *Anaerolineae*, *Ca. Bipolaricaulia*, and *Thermodesulfovibrionia* account for 16 MAGs and 20% of cluster 1 GCPM. A single *Sulfurihydrogenibium* MAG accounts for the prominence of rTCA as the second most abundant CFP (12% of cluster 1 GCPM). This MAG has the genetic potential for sulfur oxidation and microaerophilic reduction of oxygen (Fig. 4), similar to cultured relatives [42, 43]. It recruited no reads from the outer forearc and <0.05% total GCPM found in the arc sites, suggesting that it is specialized to the forearc (Supplementary Fig. S1). CBB is present in 5 MAGs from the *Deinococci*, *Anaerolineae*, UBA4738 (in phylum *Actinobacteriota*), and *Cyanobacteria* classes with only 8% of cluster 1 GCPM. The *Deinococci* and *Anaerolineae* MAGs contain the genetic potential to oxidize sulfur and reduce nitrogen species (Fig. 4).

Delocalized (clusters 2, 4, and 5). MAGs in clusters 2 (78 MAGs, 24% of total GCPM), 4 (18 MAGs, 1% of total GCPM), and 5 (112 MAGs, 12% of total GCPM) are common across multiple provinces. Cluster 2 is more concentrated in the forearc and arc, but these MAGs are much more common in other provinces than those of cluster 1, and they share the characteristics discussed below with clusters 4 and 5, so we include them in the delocalized cluster. The

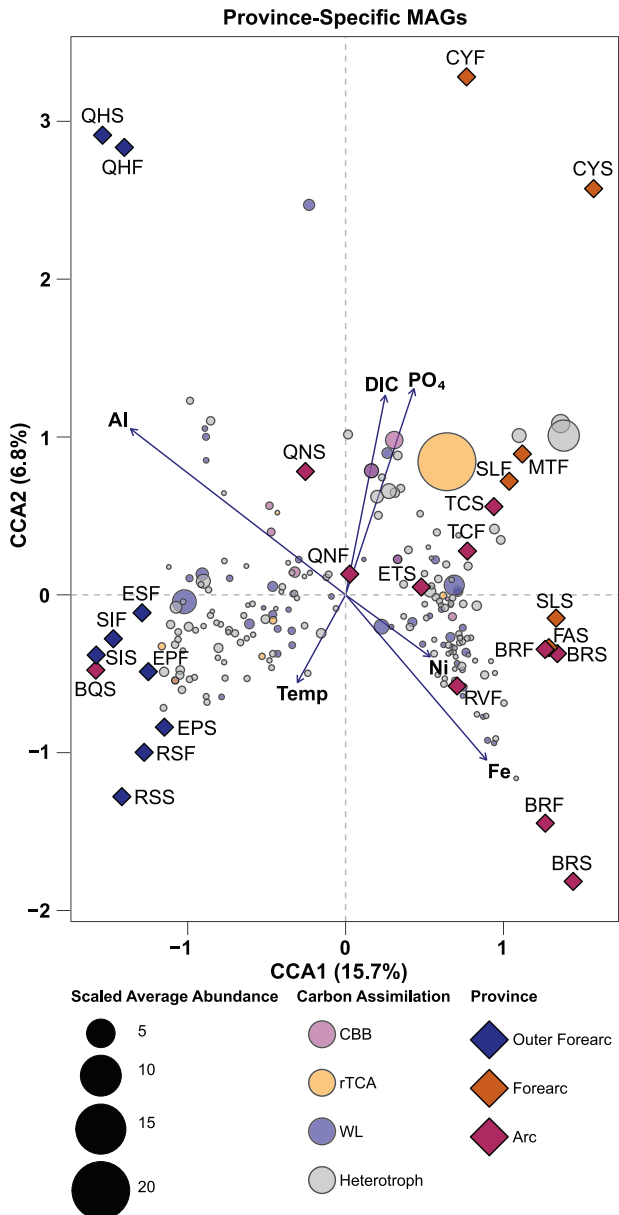


Fig. 5 Transformation-based canonical correspondence analysis showing variables most closely correlated with MAG abundance distribution at each site and site distribution of each MAG (scaling = 2). Distances between MAGs are chi-square distances. A 90° projection of a MAG marker on an environmental vector represents the maximum abundance of that MAG along that vector. Site markers are distributed around the weighted centroid of the MAGs. The model explains 36% of the total variation in community structure and the axes that are displayed contributed 62% of the total explained variation. The four MAGs with multiple carbon fixation pathways have those two colors overlain.

top 10 classes in the delocalized clusters are commonly found in humans, animals, and/or plants [44]; human-influenced environments [45]; soil [46–51]; marine/freshwater [39, 47, 48, 50, 52–56]; and hot spring microbial mats [57]. These clusters contained three of the only four *Cyanobacteria* MAGs in the whole dataset, all of the common lab contaminant genera (*Sphingomonas*, *Pelomonas*, and *Aquabacterium*) in the whole dataset [58], and high abundances of human commensals (*Chromobacterium haemolyticum*, *Acinetobacter junii*, and *Acinetobacter indicus* [59, 60]) in hot spring spas at resort sites (Blue River and Recreo Verde).

Borinquen is the only site dominated by MAGs from clusters 4 and 5. This could be because the low biomass at Borinquen allowed laboratory and human contaminants to comprise a larger proportion of the extracted and amplified DNA, causing it to cluster away from other forearc/arc sites. Most delocalized MAGs have the genetic potential for sulfur and/or hydrogen oxidation as well as oxygen and/or nitrate reduction. CBB dominates the CFPs in these clusters, mostly with Form I RuBisCO, which is typically found in organisms that respire atmospheric levels of oxygen (Supplementary Fig. S1).

Transformation-based canonical correspondence analysis

The tb-CCA of MAGs from province-specific clusters (clusters 1 and 3) explains 36.3% of the variation in the samples and recovers the same province-specific clusters predicted in Spearman-based hierarchical clusters. The important variables are, in order of most marginal effect significance: concentrations of sediment-derived aluminum, aqueous iron concentrations, DIC, temperature, aqueous phosphate concentrations, and aqueous nickel concentrations (Fig. 5 and Supplementary Table S4). Based on strong correlations, the concentration of aqueous iron is also a proxy for aqueous zinc concentrations. MAGs at outer forearc sites correlate with sediment-associated aluminum; MAGs at forearc sites correlate with DIC, phosphate concentrations, and to a lesser extent aqueous iron and nickel concentrations, while arc sites are tightly associated with iron and nickel concentrations. Distribution of WL-containing MAGs is well-explained by concentrations of aluminum, nickel, and iron, whereas rTCA- and CBB-containing MAGs are more influenced by DIC and phosphate concentrations. Heterotrophs are distributed alongside autotrophs, with the highly abundant heterotrophs co-occurring with rTCA- and CBB-containing MAGs. For MAGs in the delocalized clusters, a tb-CCA plot explains only 14.2% of the between-site variation, and results in no discernable clustering (Supplementary Fig. S6), suggesting that local geochemical variables are not major determinants of the distribution of the delocalized MAGs.

DISCUSSION

Each site contains MAGs with multiple carbon fixation and metabolic pathway predictions, suggesting taxonomically diverse populations with carbon fixation and metabolic pathways that differ systematically across the three geological provinces (outer forearc, forearc, and arc) (Fig. 6). Approximately one-third of the metagenomic reads map to MAGs containing CFPs, consistent with the proportion of chemolithoautotrophs found in other subsurface lithoautotrophic ecosystems [27, 61, 62]. Even though temperatures are low enough for photosynthesis (<70 °C at >90% of sites), only four MAGs contain photosystems, and all are *Cyanobacteria* (<1% total GCPM). Only one was found in cluster 1 (bin 379) (0.04% of total GCPM) while the other three were found in the delocalized group. These findings support the conclusions reached by Barry et al. (2019) and Fullerton et al. (2021) that chemolithoautotrophy is the main driver of primary productivity in the subsurface of these natural spring communities.

While most MAGs (63% of total GCPM) separate by Spearman-based hierarchical clustering into those abundant in outer forearc sites and those abundant in forearc/arc sites (Fig. 2 and Supplementary Fig. S1), forearc and arc sites only differentiate from each other in the tb-CCA analysis (Fig. 5). The tb-CCA analysis shows the separation of sites by province (outer forearc, forearc, and arc), based on relationships between their associated MAGs' correlations with metals, slab-derived carbon, and phosphate concentrations as well as temperature. Differences in community composition are significantly correlated with changes in concentrations of aqueous iron, nickel (with zinc by proxy), and sediment-associated aluminum ($p < 0.01$). Correlations with aqueous iron and nickel concentrations are strongest in the more acidic forearc/

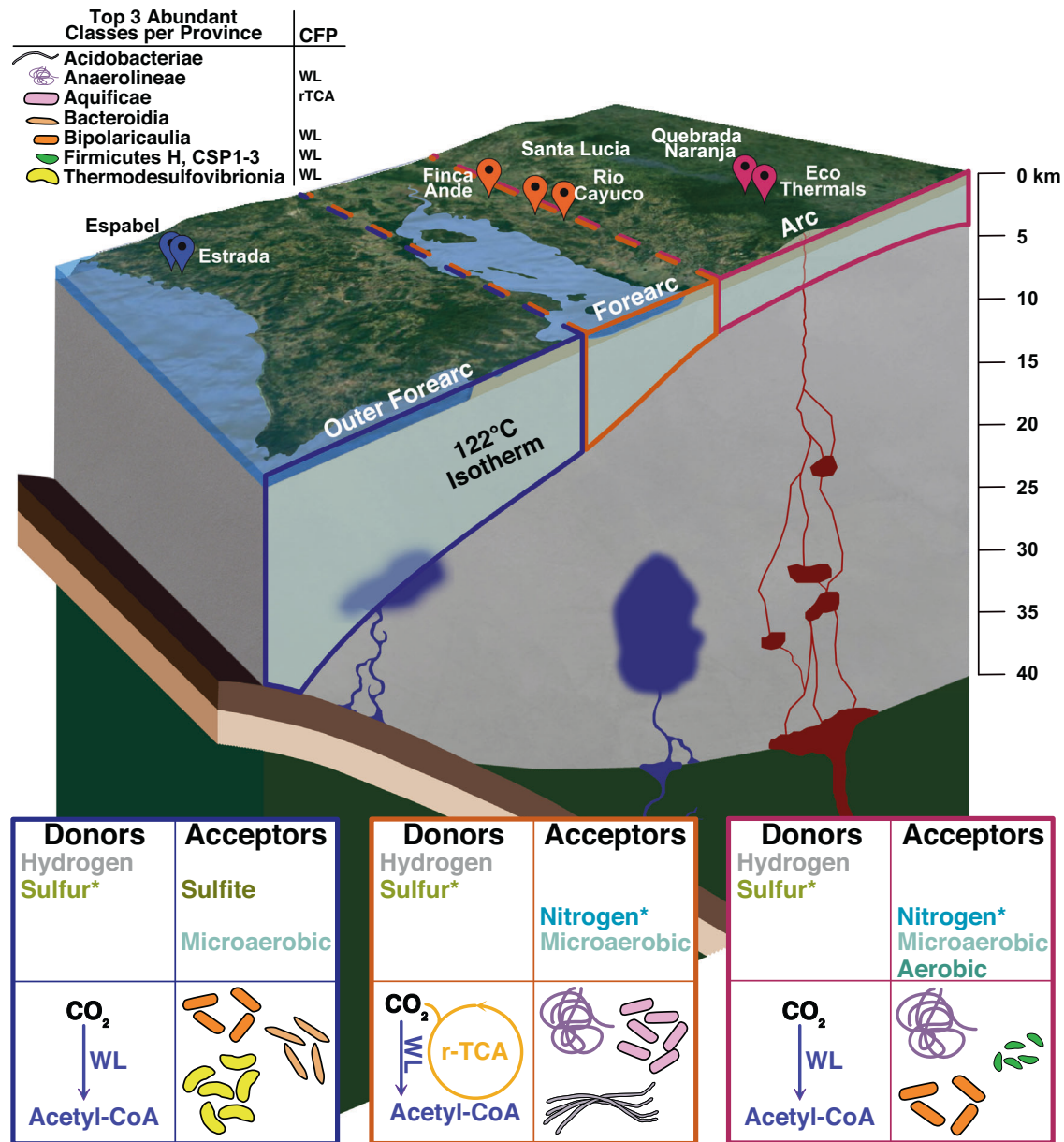


Fig. 6 Schematic cross-section (to scale) of the 3D distribution of subsurface microbial communities and their metabolisms across the Costa Rica convergent margin, showing a shallowing of potential habitable area from the outer forearc to the arc. A subset of sites is shown. Crustal thickness, angle of subduction, and 122 °C isothermal depth are from Harris and Wang [3], while aquifer depth, location, and size are from Worzewski et al. [63]. Top 3 classes per province and their associated metabolisms are listed. Cell morphology displayed is based on morphologies of cultures from each class. Site colors follow those of Fig. 1.

arc samples where reactions between the magmatic fluid and the underlying mafic rocks lead to higher concentrations of dissolved iron and iron-sulfides, also yielding pyrite framboids [7]. In particular, MAGs from Blue River (arc) and Finca ANDE (forearc) have the greatest association with aqueous iron and nickel concentrations, likely reflecting the greater influence of metals from the large hydrothermal system.

MAGs in outer forearc sites correlate with lower concentrations of aqueous phosphate and higher concentrations of sediment-associated aluminum. These fluids contain the lowest concentrations of aqueous phosphate (<0.001 mmol/L) and are partially equilibrated with feldspars and clays [7], likely derived from the dehydration of mineral-bound water in the slab material 12–15 km below the surface [63, 64]. The isostructural similarity of aluminum phosphate with polymorphs of silica

allows for extensive aluminum phosphate substitution of silica in feldspar [65, 66]. Microorganisms in soil and environments low in aqueous phosphate can liberate this phosphate through the biological weathering of feldspar [67, 68], increasing the concentration of solid and aqueous aluminum species [69, 70]. Therefore, the correlation of outer forearc MAGs with aluminum concentrations may reflect an increase in the liberation of phosphate from feldspars. Relative to the other provinces, the outer forearc has a higher abundance of *Thermodesulfovibrionia*, *Actinobacteriota*, *Desulfobacterota*, and *Gammaproteobacteria*; the forearc has a higher abundance of *Sulfurihydrogenibium* sp., *Thaumarchaeota*, and *Acidobacteriota*; and the arc has a higher abundance of *Firmicutes*. *Chloroflexi* and *Ca. Bipolaricaulota* are abundant everywhere, although their individual MAGs differ between provinces.

Distribution of carbon fixation pathways across geological provinces

WL is the dominant CFP across the outer forearc, forearc, and arc, while rTCA is abundant only in the forearc, and CBB is in very low abundance throughout the provinces (Fig. 2 and Supplementary Figs. S1 and S4). The WL pathway is highly oxygen-sensitive, the rTCA cycle is moderately oxygen-sensitive, and both tolerate high temperatures [38, 71], making them suitable for subsurface hydrothermal systems. In addition to being more abundant and widespread, WL-containing MAGs are from more diverse taxa and contain a wider range of redox pathway predictions than rTCA and CBB. Since WL has the lowest energy requirements of all known CFPs [38], it may provide autotrophs with a selective advantage. rTCA is abundant only in the forearc, mostly in *Sulfurihydrogenibium* sp. The four MAGs (class *Anaerolineae*: bins 55, 251, and 386 and class *Syntrophia*: bin 351) that contain genes encoding two CFPs each may be adapted to variable carbon dioxide and oxygen concentrations [27, 38, 72].

WL-containing MAGs contribute most to the correlations between the microbial composition of the sites and higher iron and nickel concentrations (Fig. 5). WL has high metal requirements, especially iron and nickel, since nickel is required for CODH and iron is required for ferredoxin [38]. In addition, higher iron, nickel, and zinc concentrations are proxies for more reducing, low pH, and high-temperature environments, which would be favored by WL-containing organisms. Abundances of CBB- and rTCA-containing MAGs, on the other hand, are more associated with DIC and phosphate in the arc and forearc, suggesting that the availability of carbon and phosphorous may be limiting for these chemolithoautotrophic communities, rather than pH, redox metabolic elements, or other nutrients. This is supported by a positive correlation between DOC and DIC concentrations across these sites, suggesting that more DOC can be produced when more DIC is available [7]. DIC in this system is almost entirely derived from the deep slab/mantle mixture and it increases in concentration at the arc due to higher volcanic inputs and less removal of carbon (through the formation of calcite) in the crust [8]. The low abundance rTCA-containing autotrophs contain a low carbon dioxide affinity pyruvate synthase [73] that may be disadvantageous in the outer forearc where alkaline conditions favor calcite precipitation and low carbon dioxide concentrations.

Outer forearc WL-containing MAGs have a higher association with aluminum and a lower association with nickel and iron than their counterparts in the arc and forearc. In the outer forearc, iron and nickel concentrations remain constant through the sites (0.7 mg/g and 0.8 $\mu\text{mol/L}$ respectively), except for the Quepos samples (1.83 mg/g and 2.5 $\mu\text{mol/L}$ respectively). However, as mentioned above, aqueous phosphate here is limited. Nutrient-limited environments, such as those limited in phosphate [70], have been shown to encourage lithoautotrophic communities to enhance the weathering of minerals [69, 70, 74, 75], such as feldspar, leaving behind an aluminum-rich residual [69, 70]. The distribution of heterotrophic MAGs is similar to that of autotrophic MAGs, suggesting that they are either similarly influenced by the geochemical variables, or that autotroph distribution drives heterotroph distribution.

Metabolic potential in the outer forearc

In the outer forearc, hydrogenotrophic sulfate reduction is the most abundant predicted metabolic pathway and occurs in the widest range of MAGs, whether or not they have CFPs. In the outer forearc, dominant WL-containing MAGs classifying as *Thermodesulfovibronia* and *Ca. Bipolaricaulia* have the genetic potential for hydrogen oxidation and *Thermodesulfovibronia* MAGs also have genes for sulfate reduction, in agreement with other studies of this strictly anaerobic class [76]. Two of the four dominant rTCA-containing MAGs classify as *Syntrophia* and contain hydrogenases that have been demonstrated to perform hydrogen uptake (Supplementary methods), while all four contain *dsrAB* genes for sulfate reduction. The genetic propensity to perform carbon

fixation via hydrogenotrophic sulfate reduction and hydrogen oxidation has been found in subsurface organisms associated with analogous submarine serpentinization sites (e.g., the Lost City marine hydrothermal field) terrestrial hydrothermal systems [77, 78], and a ~3 km deep borehole in South Africa [79].

Serpentinization-sourced hydrogen and low dissolved oxygen concentrations in the outer forearc create amenable conditions for WL and rTCA pathways, hydrogen oxidation, and sulfate reduction. These results contrast sharply with the high prevalence of methanogens, absence of sulfate-reducers, and evidence for very limited biogenic sulfate reduction just to the north in the Santa Elena serpentinizing system [80–82]. The Santa Elena system is ophiolite-hosted [81], whereas outer forearc sites in the Nicoya peninsula are more influenced by the mantle wedge [83]. Santa Elena also has lower sulfate concentrations, which may allow methanogens to compete for electron donors [81]. This suggests that differences in underlying geology, or possibly different degrees of inmixing of subducted sulfate-rich seawater, can drive differences in microbial metabolic pathways, even though both systems are serpentinizing and have similar pH and temperatures.

Metabolic potential in the forearc and arc

The most abundant and widespread predicted metabolic pathways in the forearc are sulfur oxidation and the reduction of oxygen and oxidized nitrogen compounds. Arc MAGs have fewer genes encoding for fully aerobic oxygen reduction (i.e., aa3-and/or bo-type cytochromes) and lack ammonium oxidation genes. Acidic water-rock interactions in the arc and forearc enable higher concentrations of aqueous carbon dioxide, which may alleviate DIC limitation in these chemosynthetic communities. In addition, the high pressure of deep magmatic degassing, cooling as these gases rise, and gas-fluid-rock interactions in local hydrothermal systems favor the more reduced sulfide over oxidized sulfate [84, 85] creating the ideal conditions for sulfide oxidation with high carbon dioxide availability as an autotrophic substrate. In the forearc and arc, the most abundant WL-containing MAGs (class *Anaerolineae*) and the lower abundant CBB-containing MAGs (classes *Deinococci* and *Anaerolineae*) have the potential to power carbon fixation with sulfur-dependent denitrification. The second most abundant WL-MAGs (class *Ca. Bipolaricaulia*) have the genetic potential to power carbon fixation with hydrogen oxidation using nitrate/nitrite as terminal electron acceptors. In the forearc, the rTCA-containing MAG (genus *Sulfurihydrogenibium* sp.) has the genetic potential for the aerobic oxidation of sulfide and thiosulfate. While sulfide oxidation is the most abundant energy pathway in the arc and forearc, the resulting oxidized sulfate can be used by the third most abundant WL-containing MAGs (class *Thermodesulfovibronia*) to power carbon fixation with sulfate reduction. Although CBB is less prevalent, the CBB-containing MAGs (class *Deinococci*) contain form IC of RuBisCO, which may be advantageous in medium to high carbon dioxide suboxic environments [86]. MAGs classified as *Anaerolineae* in the forearc/arc contain an unknown form of RuBisCO that may be the recently discovered Form I' or Form I- α RuBisCO which have been proposed to be the ancestral form of all RuBisCO forms [87].

As the distance increases from the trench toward the arc, the geothermal gradient to the 122 °C isotherm steepens [3], potentially compressing the overall volume of available subsurface habitat. The greater presence of respiratory pathways in the forearc and arc may therefore reflect their greater proximity to atmospheric recharge and faster recharge with oxygenated meteoric water driven by the steep geothermal gradient and high rainfall in these mountainous areas.

Delocalized clusters are surface associated

We explored the possibility that the delocalized MAGs in clusters 2, 4, and 5, constitute a low abundance surface-associated community. The tb-CCA analysis shows that the distribution of the delocalized MAGs across sites is poorly explained by local

geochemical variables (Supplementary Fig. S6) and their ten most abundant classes are common soils or animal inhabitants. The delocalized clusters also contain the only MAGs from the genera *Aquabacterium*, *Alishewanella*, *Rheinheimera*, and *Exiguobacterium* whose 16S rRNA genes were previously found to lack correlation with geological parameters [7]. Most MAGs in clusters 2, 4, and 5 (51%) are in cultured family to species (Supplementary Fig. S5), a feature common in human-influenced environments [88]. The presence of common soil inhabitants and human commensals, prevalence of fully aerobic metabolisms including CBB, and lack of correlations with local geochemical and geological parameters suggest that the delocalized clusters are composed mostly of surface-associated MAGs and a few laboratories/in situ contaminants.

CONCLUSIONS

We conclude that the distribution of province-specific MAGs across the subsurface of the Costa Rican convergent margin correlates with aluminum, iron, nickel, DIC, and phosphate concentrations, which reflect the underlying geology, fluid sources, and fluid-rock interactions. Hydrogenotrophic sulfate reduction dominates the outer forearc, whereas more oxidized substrates such as nitrate and oxygen can oxidize sulfide in the forearc and arc. This suggests that, unlike individual volcanic systems (i.e., Yellowstone National Park [19, 89]), pH and temperature are not the driving variables that control the microbial distribution at a landscape scale in a convergent margin. MAGs with chemolithoautotrophic pathways comprise a third of this subsurface ecosystem. The WL pathway is found in a wide diversity of taxa, which collectively also contain genes encoding for all available metabolic pathways, in every part of the subsurface landscape, with metal availability associated distributions. rTCA, CBB, and sulfur oxidation are more prominent in the forearc and arc which have high carbon dioxide concentrations. We conclude that these subsurface chemolithoautotrophic ecosystems support a diverse set of metabolic processes and associated autotrophic pathways, tightly connected to deep subsurface geological and geochemical parameters such as dissolved metal content and the degree to which deep volcanic inorganic carbon and other volatile components are available for microbial usage.

DATA AVAILABILITY

Metagenomic data are available in the NCBI SRA with project ID PRJNA627197. All MAGs analyzed in this study are also included under the same BioProject. BioSample accessions for each MAG are included in Supplementary Table S1. The full environmental dataset along with the R scripts for this analysis is available at <https://github.com/TJrogers86/Costa-Rica-2017.git> and released as a permanent version (v1.0.0) using Zenodo under <https://doi.org/10.5281/zenodo.6259294>.

REFERENCES

- Kelemen PB, Manning CE. Reevaluating carbon fluxes in subduction zones, what goes down, mostly comes up. *Proc Natl Acad Sci USA*. 2015;112:E3997–4006.
- Vitale Brovarone A, Sverjensky DA, Piccoli F, Ressico F, Giovannelli D, Daniel I. Subduction hides high-pressure sources of energy that may feed the deep subsurface biosphere. *Nat Commun*. 2020;11:1–1.
- Harris RN, Wang K. Thermal models of the middle America trench at the Nicoya Peninsula, Costa Rica. *Geophys Res Lett*. 2002;29:6–1.
- Plümper O, King HE, Geisler T, Liu Y, Pabst S, Savov IP, et al. Subduction zone forearc serpentinites as incubators for deep microbial life. *Proc Natl Acad Sci USA*. 2017;114:4324–9.
- Lee H, Fischer TP, de Moor JM, Sharp ZD, Takahata N, Sano Y. Nitrogen recycling at the Costa Rican subduction zone: the role of incoming plate structure. *Sci Rep*. 2017;7:1–10.
- Stern RJ. Subduction zones. *Rev Geophys*. 2002;40:3–38.
- Fullerton KM, Schrenk MO, Yücel M, Manini E, Basili M, Rogers TJ, et al. Effect of tectonic processes on biosphere–geosphere feedbacks across a convergent margin. *Nat Geosci*. 2021;14:301–6.
- Barry PH, de Moor JM, Giovannelli D, Schrenk M, Hummer DR, Lopez T, et al. Forearc carbon sink reduces long-term volatile recycling into the mantle. *Nature*. 2019;568:487–92.
- Moore EK, Jelen BI, Giovannelli D, Raanan H, Falkowski PG. Metal availability and the expanding network of microbial metabolisms in the Archaean eon. *Nat Geosci*. 2017;10:629–36.
- Barnes JD, Cullen J, Barker S, Agostini S, Penniston-Dorland S, Lassiter JC, et al. The role of the upper plate in controlling fluid-mobile element (Cl, Li, B) cycling through subduction zones: Hikurangi forearc, New Zealand. *Geosphere*. 2019;15:642–58.
- Clift P, Vannucchi P. Controls on tectonic accretion versus erosion in subduction zones: Implications for the origin and recycling of the continental crust. *Rev Geophys*. 2004;42:1–31.
- Rüpke LH, Morgan JP, Hort M, Connolly JA. Serpentine and the subduction zone water cycle. *Earth Planet Sci Lett*. 2004;223:17–34.
- Carr MJ, Feigenson MD, Bennett EA. Incompatible element and isotopic evidence for tectonic control of source mixing and melt extraction along the Central American arc. *Contrib Miner Pet*. 1990;105:369–80.
- Gazel E, Carr MJ, Hoernle K, Feigenson MD, Szymanski D, Hauff F, et al. Galapagos-OIB signature in southern Central America: mantle refertilization by arc-hot spot interaction. *Geochem Geophys Geosyst*. 2009;10:1–32.
- Trembath-Reichert E, Butterfield DA, Huber JA. Active seafloor microbial communities from Mariana back-arc venting fluids share metabolic strategies across different thermal niches and taxa. *ISME J*. 2019;13:2264–79. <https://doi.org/10.1038/s41396-019-0431-y>.
- Power JF, Carere CR, Lee CK, Wakerley GL, Evans DW, Button M, et al. Microbial biogeography of 925 geothermal springs in New Zealand. *Nat Commun*. 2018;9:1–2.
- Acocella V, Spinks K, Cole J, Nicol A. Oblique back arc rifting of Taupo Volcanic zone. *NZ Tecton*. 2003;22:1–18.
- Curtis AC, Wheat CG, Fryer P, Moyer CL. Mariana forearc serpentinite mud volcanoes harbor novel communities of extremophilic archaea. *Geomicrobiol J*. 2013;30:430–41.
- Inskeep WP, Jay ZJ, Herrgård MJ, Kozubal MA, Rusch DB, Tringe SG, et al. Phylogenetic and functional analysis of metagenome sequence from high-temperature archaeal habitats demonstrate linkages between metabolic potential and geochemistry. *Front Microbiol*. 2013;4:1–21.
- Colman DR, Lindsay MR, Amenabar MJ, Boyd ES. The intersection of geology, geochemistry, and microbiology in continental hydrothermal systems. *Astrobiology*. 2019;19:1505–22.
- Inskeep WP, Jay ZJ, Tringe SG, Herrgård MJ, Rusch DB, YNP Metagenome Project Steering Committee and Working Group Members. The YNP metagenome project: environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. *Front Microbiol*. 2013;4:1–15.
- Hou W, Wang S, Dong H, Jiang H, Briggs BR, Peacock JP, et al. A comprehensive census of microbial diversity in hot springs of Tengchong, Yunnan Province China using 16S rRNA gene pyrosequencing. *PLoS One*. 2013;8:1–15.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
- Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res*. 2017;27:824–34.
- Bowers RM, Kyrpidis NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TB, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol*. 2017;35:725–31.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 2015;25:1043–55.
- Probst AJ, Castelle CJ, Singh A, Brown CT, Anantharaman K, Sharon I, et al. Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO₂ concentrations. *Environ Microbiol*. 2017;19:459–74.
- Probst AJ, Ladd B, Jarett JK, Geller-McGrath DE, Sieber CM, Emerson JB, et al. Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat Microbiol*. 2018;3:328–36.
- He C, Keren R, Whittaker M, Farag IF, Doudna J, Cate JH, et al. Genome-resolved metagenomics reveals site-specific diversity of epibiotic CPR bacteria and DPANN archaea in groundwater ecosystems. *Nat Microbiol*. 2021;6:354–65.
- Grettenberger CL, Hamilton TL. Metagenome-assembled genomes of novel taxa from an acid mine drainage environment. *Appl Environ Microbiol*. 2021;87:e0077221. <https://doi.org/10.1101/2020.07.02.185728>.
- Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome*. 2018;6:1–3.
- Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30:2068–9.

33. Garber AI, Neelson KH, Okamoto A, McAllister SM, Chan CS, Barco RA, et al. FeGenie: a comprehensive tool for the identification of iron genes and iron gene neighborhoods in genome and metagenome assemblies. *Front Microbiol.* 2020;11:37. <https://doi.org/10.3389/fmicb.2020.00037>.
34. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol.* 2016;428:726–31.
35. Graham ED, Heidelberg JF, Tully BJ. Potential for primary productivity in a globally distributed bacterial phototroph. *ISME J.* 2018;12:1861–6.
36. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics.* 2020;36:1925–27.
37. Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hügler M, et al. Autotrophic carbon fixation in archaea. *Nat Rev Microbiol.* 2010;8:447–60.
38. Berg IA. Ecological aspects of the distribution of different autotrophic CO₂ fixation pathways. *Appl Environ Microbiol.* 2011;77:1925–36.
39. Youssef NH, Farag IF, Hahn CR, Jarett J, Becraft E, Eloë-Fadrosch E, et al. Genomic characterization of candidate division LCP-89 reveals an atypical cell wall structure, microcompartment production, and dual respiratory and fermentative capacities. *Appl Environ Microbiol.* 2019;85:1–19.
40. Nigro LM, King GM. Disparate distributions of chemolithotrophs containing form IA or IC large subunit genes for ribulose-1, 5-bisphosphate carboxylase/oxygenase in intertidal marine and littoral lake sediments. *FEMS Microbiol Ecol.* 2007;60:113–25.
41. Aminuddin M, Nicholas DJ. Electron transfer during sulphide and sulphite oxidation in *Thiobacillus denitrificans*. *Microbiology.* 1974;82:115–23.
42. Giovannelli D, Sievert SM, Hügler M, Markert S, Becher D, Schweder T, et al. Insight into the evolution of microbial metabolism from the deep-branching bacterium, *Thermovibrio ammonificans*. *eLife.* 2017;6:1–31.
43. Nakagawa S, Shataih Z, Banta A, Beveridge TJ, Sako Y, Reysenbach AL. *Sulfurihydrogenibium yellowstonense* sp. nov., an extremely thermophilic, facultatively heterotrophic, sulfur-oxidizing bacterium from Yellowstone National Park, and emended descriptions of the genus *Sulfurihydrogenibium*, *Sulfurihydrogenibium subteraneum*. *Int J Syst Evol Microbiol.* 2005;55:2263–8.
44. Leclercq A, Kleespies RGA. *Rickettsiella* bacterium from the hard tick, *Ixodes woodi*: molecular taxonomy combining multilocus sequence typing (MLST) with significance testing. *PLoS One.* 2012;7:e38062. <https://doi.org/10.1371/journal.pone.0038062>.
45. Quatrini R, Johnson DB. *Acidithiobacillus ferrooxidans*. *Trends Microbiol.* 2019;27:282–3.
46. Spang A, Pöehlein A, Offre P, Zumbärgel S, Haider S, Rychlik N, et al. The genome of the ammonia-oxidizing *Candidatus Nitrososphaera gargensis*: insights into metabolic versatility and environmental adaptations. *Environ Microbiol.* 2012;14:3122–45.
47. Chen CY, Chen PC, Weng FC, Shaw GT, Wang D. Habitat and indigenous gut microbes contribute to the plasticity of gut microbiome in oriental river prawn during rapid environmental change. *PLoS One.* 2017;12:e0181427. <https://doi.org/10.1371/journal.pone.0181427>.
48. Garcia R, Müller R. The family *Myxococcaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. *The prokaryotes: Deltaproteobacteria and Epsilonproteobacteria*. Berlin: Springer; 2014. p. 191–212.
49. Garcia R, Müller R. *Simulacricoccus ruber* gen. nov., sp. nov., a microaerotolerant, non-fruiting, myxospore-forming soil myxobacterium and emended description of the family *Myxococcaceae*. *Int J Syst Evol Microbiol.* 2018;68:3101–10.
50. Iino T. The family *Ignavibacteriaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. *The prokaryotes: other major lineages of bacteria and the archaea*. New York, NY: Springer Science + Business Media; 2014. p. 701–3.
51. Petrie L, North NN, Dollhopf SL, Balkwill DL, Kostka JE. Enumeration and characterization of iron (III)-reducing microbial communities from acidic subsurface sediments contaminated with uranium (VI). *Appl Environ Microbiol.* 2003;69:7467–79.
52. Fincker M, Huber JA, Orphan VJ, Rappé MS, Teske A, Spormann AM. Metabolic strategies of marine subsurface *Chloroflexi* inferred from genome reconstructions. *Environ Microbiol.* 2020;22:3188–204.
53. Chen YJ, Leung PM, Wood JL, Bay SK, Hugenholtz P, Kessler AJ, et al. Metabolic flexibility allows bacterial habitat generalists to become dominant in a frequently disturbed ecosystem. *ISME J.* 2021;15:2986–3004.
54. Flieder M, Buongiorno J, Herbold CW, Hausmann B, Rattai T, Lloyd KG, et al. Novel taxa of *Acidobacteriota* implicated in seafloor sulfur cycling. *ISME J.* 2021;15:3159–80.
55. Kim M, Wilpiseski RL, Wells M, Wymore AM, Gionfriddo CM, Brooks SC, et al. Metagenome-assembled genome sequences of novel prokaryotic species from the mercury-contaminated East Fork Poplar Creek, Oak Ridge, Tennessee, USA. *Microbiol Resour Anounc.* 2021;10:e00153–21. <https://doi.org/10.1128/MRA.00153-21>.
56. Santos-Júnior CD, Logares R, Henrique-Silva F. Microbial population genomes from the Amazon River reveal possible modulation of the organic matter degradation process in tropical freshwaters. *Mol Ecol.* 2022;31:206–19.
57. Yamada T, Sekiguchi Y. Cultivation of uncultured *Chloroflexi* subphyla: significance and ecophysiology of formerly uncultured *Chloroflexi* 'subphylum I' with natural and biotechnological relevance. *Microbes Environ.* 2009;24:205–16.
58. Sheik CS, Reese BK, Twing KI, Sylvan JB, Grim SL, Schrenk MO, et al. Identification and removal of contaminant sequences from ribosomal gene databases: lessons from the census of deep life. *Front Microbiol.* 2018;9:840. <https://doi.org/10.3389/fmicb.2018.00840>.
59. Doughari HJ, Ndakidemi PA, Human IS, Benade S. The ecology, biology and pathogenesis of *Acinetobacter* spp.: an overview. *Microbes Environ.* 2011;26:101–12.
60. Han XY, Han FS, Segal J. *Chromobacterium haemolyticum* sp. nov., a strongly haemolytic species. *Int J Syst Evol Microbiol.* 2008;58:1398–403.
61. Lau MC, Kieft TL, Kuloyo O, Linage-Alvarez B, Van Heerden E, Lindsay MR, et al. An oligotrophic deep-subsurface community dependent on syntrophy is dominated by sulfur-driven autotrophic denitrifiers. *Proc Natl Acad Sci USA.* 2016;113:E7927–36.
62. Momper L, Jungbluth SP, Lee MD, Amend JP. Energy and carbon metabolisms in a deep terrestrial subsurface fluid microbial community. *ISME J.* 2017;11:2319–33.
63. Worzewski T, Jegen M, Kopp H, Brasse H, Taylor Castillo W. Magnetotelluric image of the fluid cycle in the Costa Rican subduction zone. *Nat Geosci.* 2011;4:108–11.
64. Hensen C, Wallmann K, Schmidt M, Ranero CR, Suess E. Fluid expulsion related to mud extrusion off Costa Rica—a window to the subducting slab. *Geology.* 2004;32:201–4.
65. Simpson DR. Aluminum phosphate variants of feldspar. *Am Miner.* 1977;62:351–5.
66. London DA, Cerny P, Loomis J, Pan JJ. Phosphorus in alkali feldspars of rare-element granitic pegmatites. *Can Miner.* 1990;28:771–86.
67. Petrillo C, Castaldi S, Lanzilli M, Selci M, Cordone A, Giovannelli D, et al. Genomic and physiological characterization of *Bacilli* isolated from salt-pans with plant growth promoting features. *Front Microbiol.* 2021;12:715678. <https://doi.org/10.3389/fmicb.2021.715678>.
68. Ghiorse WC, Wilson JT. Microbial ecology of the terrestrial subsurface. *Adv Appl Microbiol.* 1988;33:107–72.
69. Barker WW, Welch SA, Chu S, Banfield JF. Experimental observations of the effects of bacteria on aluminosilicate weathering. *Am Miner.* 1998;83:1551–63.
70. Bennett PC, Rogers JR, Choi WJ, Hiebert FK. Silicates, silicate weathering, and microbial ecology. *Geomicrobiol J.* 2001;18:3–19.
71. Hügler M, Sievert SM. Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. *Ann Rev Mar Sci.* 2013;3:261–89.
72. Markert S, Arndt C, Felbeck H, Becher D, Sievert SM, Hügler M, et al. Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila*. *Science.* 2007;315:247–50.
73. Bar-Even A, Noor E, Milo R. A survey of carbon fixation pathways through a quantitative lens. *J Exp Bot.* 2012;63:2325–42.
74. Stevens TO, McKinley JP. Lithoautotrophic microbial ecosystems in deep basalt aquifers. *Science.* 1995;270:450–5.
75. Barker WW, Welch SA, Banfield JF. Biogeochemical weathering of silicate minerals. *Rev Miner Geochem.* 1997;35:391–428.
76. Frank YA, Kadnikov VV, Lukina AP, Banks D, Beletsky AV, Mardanov AV, et al. Characterization and genome analysis of the first facultatively alkaliphilic *Thermodesulfobacterium* isolated from the deep terrestrial subsurface. *Front Microbiol.* 2016;7:2000. <https://doi.org/10.3389/fmicb.2016.02000>.
77. Woycheese KM, Meyer-Dombard DA, Cardace D, Argayosa AM, Arcilla CA. Out of the dark: transitional subsurface-to-surface microbial diversity in a terrestrial serpentinizing seep (Manleluag, Pangasinan, the Philippines). *Front Microbiol.* 2015;6:1–12.
78. Brazelton WJ, Morrill PL, Szponar N, Schrenk MO. Bacterial communities associated with subsurface geochemical processes in continental serpentinite springs. *Appl Environ Microbiol.* 2013;79:3906–16.
79. Moser DP, Gihring TM, Brockman FJ, Fredrickson JK, Balkwill DL, Dollhopf ME, et al. *Desulfotomaculum* and *Methanobacterium* spp. dominate a 4-to 5-kilometer-deep fault. *Appl Environ Microbiol.* 2005;71:8773–83.
80. Schwarzenbach EM, Gill BC, Gazel E, Madrigal P. Sulfur and carbon geochemistry of the Santa Elena peridotites: comparing oceanic and continental processes during peridotite alteration. *Lithos.* 2016;252:92–108.
81. Sánchez-Murillo R, Gazel E, Schwarzenbach EM, Crespo-Medina M, Schrenk MO, Boll J, et al. Geochemical evidence for active tropical serpentinization in the Santa Elena Ophiolite, Costa Rica: an analog of a humid early Earth? *Geochem Geophys Geosyst.* 2014;15:1783–800.
82. Crespo-Medina M, Twing KI, Sánchez-Murillo R, Brazelton WJ, McCollom TM, Schrenk MO. Methane dynamics in a tropical serpentinizing environment: the Santa Elena Ophiolite, Costa Rica. *Front Microbiol.* 2017;8:916. <https://doi.org/10.3389/fmicb.2017.00916>.
83. DeShon HR, Schwartz SY. Evidence for serpentinization of the forearc mantle wedge along the Nicoya Peninsula, Costa Rica. *Geophys Res Lett.* 2004;31. <https://doi.org/10.1029/2004GL021179>.
84. Delmelle P, Stix J. Volcanic gases. In: Sigurdsson H, Houghton B, McNutt S, Rymer H, Stix J, editors. *Encyclopedia of volcanoes*. New York, NY: Elsevier; 2000. p. 803–15.

85. Kharaka YK, Mariner RH. Geothermal systems. In: Sigurdsson H, Houghton B, McNutt S, Rymer H, Stix J, editors. *Encyclopedia of volcanoes*. New York, NY: Elsevier; 2000. p. 817–34.
86. Badger MR, Bek EJ. Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO₂ acquisition by the CBB cycle. *J Exp Bot*. 2008; 59:1525–41.
87. West-Roberts JA, Carnevali PB, Scholmerich MC, Al-Shayeb B, Thomas A, Sharrar AM, et al. The *Chloroflexi* supergroup is metabolically diverse and representatives have novel genes for non-photosynthesis based CO₂ fixation. *bioRxiv* [Preprint]. 2021. Available from: <https://doi.org/10.1101/2020.05.14.094862>.
88. Lloyd KG, Steen AD, Ladau J, Yin J, Crosby L. Phylogenetically novel uncultured microbial cells dominate earth microbiomes. *mSystems*. 2018;3:1–12.
89. Colman DR, Lindsay MR, Boyd ES. Mixing of meteoric and geothermal fluids supports hyperdiverse chemosynthetic hydrothermal communities. *Nat Commun*. 2019;10:1–3.

ACKNOWLEDGEMENTS

The authors acknowledge the Biology Meets Subduction Project, funded by the Alfred P. Sloan Foundation and the Deep Carbon Observatory (G-2016-7206) to PHB, JMdM, DG, and KGL, with DNA sequencing from the Census of Deep Life. Additional support came from NSF FRES (Award# 21211637) to PHB, JMdM and KGL. U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program (DE-SC0020369 to KGL) and FONDECYT Grant 11191138 (ANID Chile) and COPAS COASTAL ANID FB210021 to GLJ. DG, AC, and MS were partially supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program Grant Agreement No. 948972—COEVOLVE—ERC-2020-STG.

AUTHOR CONTRIBUTIONS

Conceptualization and funding acquisition were performed by PHB, JMdM, DG and KGL. Investigations and data acquisition were performed by TJR, KGL, DG, JB, GLJ, MOS, JAF, JMdM, CR, PHB, MY, MS, and AC. TJR performed data analysis, visualization, and drafted the paper with edits and feedback of from KGL, DG, JB, GLJ, MOS, JAF, JMdM, CR, PHB, MY, MS, and AC.

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/s41396-022-01331-7>.

Correspondence and requests for materials should be addressed to Karen G. Lloyd.

Reprints and permission information is available at <http://www.nature.com/reprints>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.