



# Microbial Diversity and Function in Shallow Subsurface Sediment and Oceanic Lithosphere of the Atlantis Massif

**J. Goordial,<sup>a,b</sup> T. D'Angelo,<sup>a</sup> J. M. Labonté,<sup>a,c</sup> N. J. Poulton,<sup>a</sup> J. M. Brown,<sup>a</sup> R. Stepanauskas,<sup>a</sup> G. L. Früh-Green,<sup>d</sup> B. N. Orcutt<sup>a,e</sup>**

<sup>a</sup>Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine, USA

<sup>b</sup>University of Guelph, Guelph, Ontario, Canada

<sup>c</sup>Texas A&M University at Galveston, Galveston, Texas, USA

<sup>d</sup>ETH Zürich, Zürich, Switzerland

<sup>e</sup>Hanse Institute for Advanced Study, Delmenhorst, Germany

**ABSTRACT** The marine lithospheric subsurface is one of the largest biospheres on Earth; however, little is known about the identity and ecological function of microorganisms found in low abundance in this habitat, though these organisms impact global-scale biogeochemical cycling. Here, we describe the diversity and metabolic potential of sediment and endolithic (within rock) microbial communities found in ultrasmall amounts ( $10^1$  to  $10^4$  cells  $\text{cm}^{-3}$ ) in the subsurface of the Atlantis Massif, an oceanic core complex on the Mid-Atlantic Ridge that was sampled on International Ocean Discovery Program (IODP) Expedition 357. This study used fluorescence-activated cell sorting (FACS) to enable the first amplicon, metagenomic, and single-cell genomic study of the shallow (<20 m below seafloor) subsurface of an actively serpentinizing marine system. The shallow subsurface biosphere of the Atlantis Massif was found to be distinct from communities observed in the nearby Lost City alkaline hydrothermal fluids and chimneys, yet similar to other low-temperature, aerobic subsurface settings. Genes associated with autotrophy were rare, although heterotrophy and aerobic carbon monoxide and formate cycling metabolisms were identified. Overall, this study reveals that the shallow subsurface of an oceanic core complex hosts a biosphere that is not fueled by active serpentinization reactions and by-products.

**IMPORTANCE** The subsurface rock beneath the ocean is one of the largest biospheres on Earth, and microorganisms within influence global-scale nutrient cycles. This biosphere is difficult to study, in part due to the low concentrations of microorganisms that inhabit the vast volume of the marine lithosphere. In spite of the global significance of this biosphere, little is currently known about the microbial ecology of such rock-associated microorganisms. This study describes the identity and genomic potential of microorganisms in the subsurface rock and sediment at the Atlantis Massif, an underwater mountain near the Mid-Atlantic Ridge. To enable our analyses, fluorescence-activated cell sorting (FACS) was used as a means to concentrate cells from low biomass environmental samples for genomic analyses. We found distinct rock-associated microorganisms and found that the capacity for microorganisms to utilize organic carbon was the most prevalent form of carbon cycling. We additionally identified a potential role for carbon monoxide metabolism in the subsurface.

**KEYWORDS** Atlantis Massif, deep biosphere, oceanic crust, IODP, single-cell genomics, fluorescence-activated cell sorting, lithosphere

Much of the recent investigations into the marine deep biosphere have focused on microbial life in deep sediments, which has revealed life to be extensive, diverse, and capable of maintaining slow metabolic activity on long timescales (1–3). Less is

**Citation** Goordial J, D'Angelo T, Labonté JM, Poulton NJ, Brown JM, Stepanauskas R, Früh-Green GL, Orcutt BN. 2021. Microbial diversity and function in shallow subsurface sediment and oceanic lithosphere of the Atlantis Massif. *mBio* 12:e00490-21. <https://doi.org/10.1128/mBio.00490-21>.

**Editor** Nicole Dubilier, Max Planck Institute for Marine Microbiology

**Copyright** © 2021 Goordial et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to J. Goordial, [goordial@uoguelph.ca](mailto:goordial@uoguelph.ca), or B. N. Orcutt, [borcutt@bigelow.org](mailto:borcutt@bigelow.org).

This is Center for Dark Energy Biosphere Investigations (C-DEBI) contribution 572.

**Received** 12 April 2021

**Accepted** 21 June 2021

**Published** 3 August 2021

known about the microbial life hosted within the oceanic lithosphere. In spite of generally low biomass in the oceanic lithosphere, given the sheer volume of this habitat, crustal microbiota can impact global biogeochemical cycling, and the marine rock subsurface is potentially one of the largest biospheres on Earth (4, 5), and crustal microbiota can impact global biogeochemical cycling. For example, crustal microorganisms may be responsible for removing at least 5% of global ocean dissolved organic carbon (6). The diversity, metabolism, and potential rate of activity of microorganisms hosted within oceanic crust remains poorly constrained. The potential for both autotrophy and heterotrophy has been detected in crustal environments in activity assays in bulk samples or by accessing the fluids moving through the crust via subsurface observatories (7, 8), but these processes are poorly resolved. Currently, our understanding of the crustal deep biosphere is informed from studies of rocks at only a few sites on Earth, which revealed diverse microbial community structures, as reviewed elsewhere (9). Linking biogeochemical cycling with phylogeny and genetics of endolithic microorganisms through molecular analyses remains challenging, due to the low biomass in the crustal subsurface and difficulties in extracting DNA from these environmental samples (10–13).

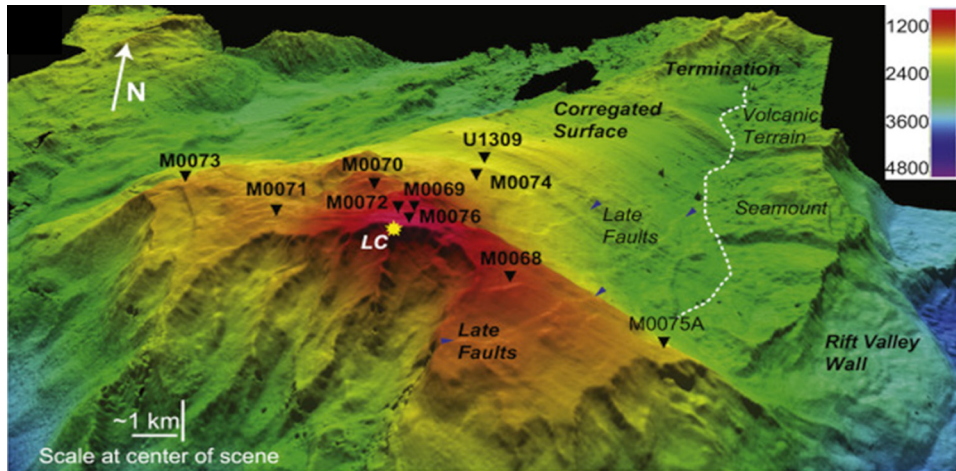
The Atlantis Massif is an underwater mountain on the flank of the Mid-Atlantic Ridge at 30°N, where ultramafic and mafic rock sequences have been uplifted to the seafloor along a major detachment fault zone in an oceanic core complex. Host to the Lost City Hydrothermal Field (LCHF) along its southern wall, the Atlantis Massif is a site of active serpentinization, where oxidized seawater interacts with olivine-rich rock to produce serpentinite in the subsurface (14). Serpentinization reactions create alkaline conditions (pH 9 to 12) and result in the abiotic production of H<sub>2</sub>, CH<sub>4</sub>, formate, and other short-chain alkanes (15–18). These by-products can be used by microorganisms for carbon and energy in the subsurface. For example, LCHF vent fluids have large amounts of methane, supporting low diversity, methane-cycling *Archaea*, sulfur-oxidizing *Bacteria*, and formate utilization (16, 19–23). While the LCHF represents a unique, high-pH microbial community, the nature of the biosphere within the crustal core of the Atlantis Massif is unknown. The central dome of the Atlantis Massif was previously drilled during Integrated Ocean Drilling Program Expeditions 304/305, but this gabbroic-rich section was likely not representative of the crustal conditions underlying Lost City (14).

To investigate the nature of the deep biosphere in the serpentinizing oceanic core complex of the Atlantis Massif, International Ocean Discovery Program (IODP) Expedition 357 drilled 17 shallow holes at nine sites across the southern wall of the massif (14) (Fig. 1). The 57 m of recovered cores were highly heterogeneous, including pervasively serpentinized ultramafic sequences as well as mafic units of gabbros, dolerites, and basalts that record multiple phases of magmatism, fluid-rock interaction, and mass transfer (14, 24). Approximately 8 m was subsampled for microbiological investigation (25, 26). Microbial cell densities in the core samples were very low, ranging from 10<sup>1</sup> to 10<sup>4</sup> cells cm<sup>-3</sup> (14).

To overcome the limitations of bulk DNA extraction from low biomass crustal samples, which is detailed extensively for this site (27), we used fluorescence-activated cell sorting (FACS) as a means to concentrate cells from the Atlantis Massif samples to enable various downstream DNA sequencing (27, 28) (Fig. 2). From these cell sorts, we queried microbial community diversity and function to examine how this actively serpentinizing, shallow, ultramafic ecosystem compares to the LCHF and other deep crustal environments. This cell sorting-based approach enabled more genomic information to be recovered from this site than has been possible with prior bulk nucleic acid extraction approaches, indicating the promise of this approach for other low biomass investigations.

## RESULTS AND DISCUSSION

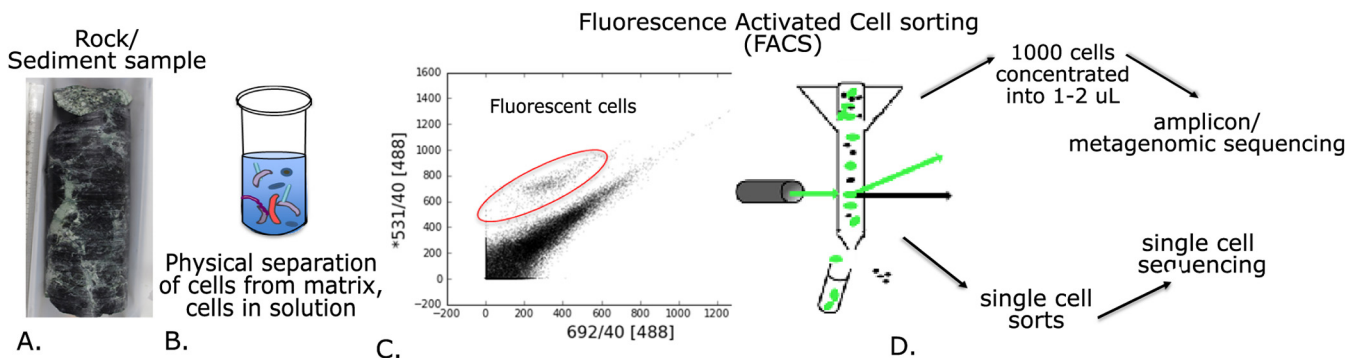
**Heterogeneous microbial community composition in the shallow subsurface of the Atlantis Massif.** A total of 29 sediment and crustal samples were screened with the bulk cell sorting approach (Table 1), with 18 yielding cell-like particles (Fig. 3 and



**FIG 1** Bathymetric map showing IODP Expedition 357 sites where samples were collected. The color legend represents water depth in meters. The figure was reproduced from Früh-Green et al. (14) with permission.

Table 1). Even with the steps taken to increase cell concentration (i.e., extracting cells from several grams of rock and concentrating to a few milliliters of fluid), samples took between 30 to 60 min of sorting to collect a sufficient number of cells (when no cells were detected, the flow cytometer was run until samples were exhausted). Sufficient DNA for amplification and amplicon sequencing of the 16S rRNA gene was successfully obtained from 8 out of the 18 samples. Three of these eight were from sediment samples, and the other five were from crust samples. Comparison of the cell detection and sequencing success of this study with the cell densities determined in parallel samples (14) indicates that the limit of this methodology does not correspond with cell density. Based on this sample set, the approach of extracting and concentrating cells via FACS from <10 g of low biomass sediment and crust samples to yield sufficient DNA for 16S rRNA gene amplicon sequencing had an overall efficiency of 8 out of 29 samples (28%). While a relatively low percentage, it is higher than many prior attempts to analyze community DNA from low biomass crustal samples (10, 11, 13, 27, 29). By comparison, a recent study described obtaining sufficient DNA for sequencing (after whole-genome amplification) from bulk nucleic acid extraction from >20 g of low biomass crustal samples with similar cell densities (30).

In addition to amplicon sequencing of the eight samples, four metagenomes were created and assembled together from four different samples: one sediment sample



**FIG 2** Overview of fluorescence-activated cell sorting (FACS) method via flow cytometry for cell concentration of low biomass samples. (A) Low biomass environmental samples of rock and sediment from IODP Expedition 357. (B) Samples were gently sonicated and vortexed in a salt buffer to physically separate cells from the substrate and suspend in liquid solution. (C) Cells in the supernatant were stained with a fluorescent general DNA stain (SYTO green) and subjected to FACS, with cells identified via fluorescent particle properties and size. (D) Cell-like particles were either separated individually or pooled as 1,000 cells into 1 to 2  $\mu$ L of a salt buffer landing pad. Single-cell sorts and 1,000-cell sorts were subjected to downstream genomic analyses after whole-genome amplification via multiple displacement amplification.

**TABLE 1** Sample lithology and cell counts<sup>a</sup>

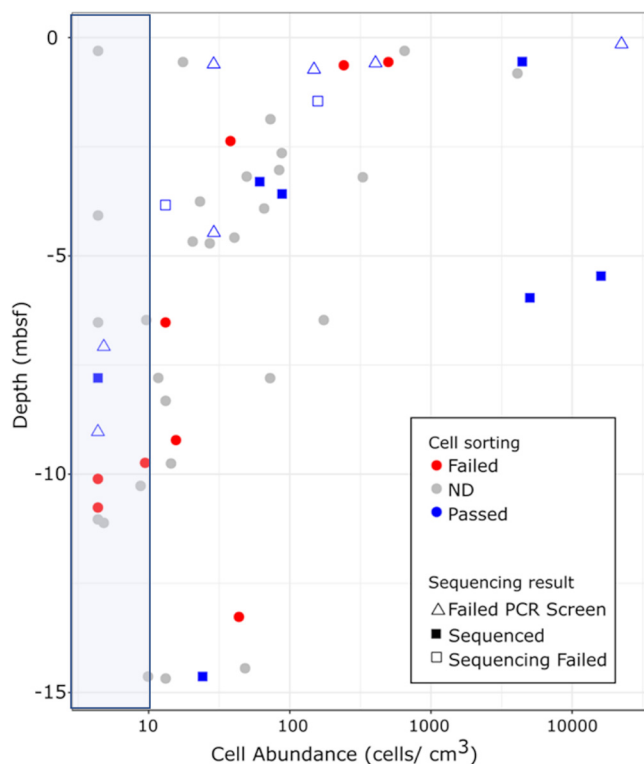
Sample <sup>b</sup>	Lithology	Cell density (cells/cm <sup>3</sup> )	PCR	Seq.	PFC INT	PFC EXT
<b>68B-1R1-1.44</b>	<b>Rubble: Metagabbro, with talc-amph overprinting</b>	<b>158</b>	+	+	<b>9 × 10<sup>4</sup></b>	<b>ADL</b>
68B-3R1-3.77	Metasomatized serpentinite (top 10 to 13 cm, possibly gabbroic domains)	13	+	–	1 × 10 <sup>5</sup>	1 × 10 <sup>5</sup>
68B-3R1-3.80	Serpentinized harzburgite	61	+	+	5 × 10 <sup>3</sup>	2 × 10 <sup>5</sup>
68B-7R1-7.73	Mixed rubble: serpentinitized harzburgite, talc-amph-chlorite schist	BMQL (4.4)	+	+	1 × 10 <sup>5</sup>	3 × 10 <sup>5</sup>
68B-8R1-9.02	Rubble: serpentinitized harzburgite, talc-amph-chlorite schist	BMQL (4.4)	–	n.a.	1 × 10 <sup>4</sup>	4 × 10 <sup>5</sup>
69A-4R1-5.41	Foraminiferal carbonate sand	16,000	+	+	BDL	BDL
<b>69A-4R1-5.91</b>	<b>Foraminiferal carbonate sand with basalt fragments</b>	<b>5,010</b>	+	+	<b>BDL</b>	<b>2 × 10<sup>2</sup></b>
69A-5R1-8.17	Metadolerite with carbonate veins	9	n.c.	n.a.	BDL	BDL
69A-7R1-11.05	Metadolerite rubble	5–16	n.c.	n.a.	2 × 10 <sup>2</sup>	BDL
<b>69A-9R2-14.61</b>	<b>Serpentinized dunite with carbonate veins</b>	<b>10–24</b>	+	+	<b>3 × 10<sup>2</sup></b>	<b>BDL-2 × 10<sup>2</sup></b>
70C-1R1-0.59	Carbonate-hosted basalt breccia	29	–	n.a.	6 × 10 <sup>2</sup>	3 × 10 <sup>4</sup>
<b>70C-3R1-3.55</b>	<b>Carbonate-hosted basalt breccia</b>	<b>88</b>	+	+	<b>1 × 10<sup>5</sup></b>	<b>1 × 10<sup>6</sup></b>
71A-1R2-0.58	Serpentinized dunite, oxidized and with carbonate veins	405	–	n.a.	BDL	7 × 10 <sup>4</sup>
71B-2R1-2.33	Metagabbro (chloride-rich)	38	n.c.	n.a.	BDL	BDL
71C-2R1-3.51	Serpentinized harzburgite	n.d.	–	n.a.	BDL	1 × 10 <sup>5</sup>
71C-3R1-5.1	Serpentinized harzburgite	n.d.	–	n.a.	BDL	6 × 10 <sup>4</sup>
71C-6R1-10.25	Metadolerite	n.d.	n.c.	n.a.	BDL	2 × 10 <sup>4</sup>
72A-1R1-0.53	Rubble: basalt-breccia; carbonate matrix	18–499	n.c.	n.a.	BDL	BDL
72B-1R1-0.62	Rubble: basalt-breccia; carbonate matrix	241	n.c.	n.a.	1	1 × 10 <sup>1</sup>
72B-3R1-3.78	Rubble of metabasalt, metagabbro, and serpentinitized harzburgite	n.d.	n.c.	n.a.	1 × 10 <sup>0</sup>	BDL
72B-5R1-6.50	Talc-amphibole-chlorite schist	13	n.c.	n.a.	1 × 10 <sup>0</sup>	6 × 10 <sup>0</sup>
72B-7R1-10.08	Metagabbro	4	n.c.	n.a.	BDL	n.d.
74A-1R1-0.1	Carbonate sand (highly disturbed during core recovery)	22,400	–	n.a.	n.a.	BDL
74A-1R1-0.5	Carbonate sand (highly disturbed during core recovery)	4,430	+	+	n.a.	6 × 10 <sup>2</sup>
75B-2R1-2.95	Rubble: metadolerite	148	–	n.a.	2 × 10 <sup>3</sup>	n.d.
76B-3R1-4.46	Metadolerite	29	–	n.a.	1 × 10 <sup>3</sup>	BDL
76B-5R1-7.08	Serpentinized harzburgite with talc-amph-chlorite alteration	BMQL (4.8)	–	n.a.	BDL	3 × 10 <sup>1</sup>
76B-7R1-10.76	Serpentinized harzburgite	BMQL (4.4)	n.c.	n.a.	BDL	2 × 10 <sup>3</sup>
76B-9R1-13.21	Serpentinized harzburgite with talc-amph-chlorite alteration and carbonate veins	44	n.c.	n.a.	1 × 10 <sup>2</sup>	8 × 10 <sup>1</sup>

<sup>a</sup>Overview of samples used in this study, including sample depth, lithology, interior cell concentrations (from Früh-Green et al. [14]), whether PCR screens of MDA products after bulk cell sorting were positive (+) or negative (–) (PCR column), whether Illumina amplicon sequencing (Seq.) of MDA product was successful, and the concentration of PFC tracer in parallel samples taken from the interior (INT) or exterior (EXT) of the core on the ship, in picograms of perfluorocyclohexane per cm<sup>3</sup> (from Orcutt et al. [66]). Samples used for metagenome and single-cell genome sequencing are highlighted in bold type. Abbreviations: ADL, above detection limit; BDL, below detection limit; BMQL, below minimum quantification limit; n.a., not applicable; n.c., no cells identified in FACS analysis; n.d., not determined.

<sup>b</sup>Samples are named as follows: hole-core section-top depth (meters below seafloor [mbsf]).

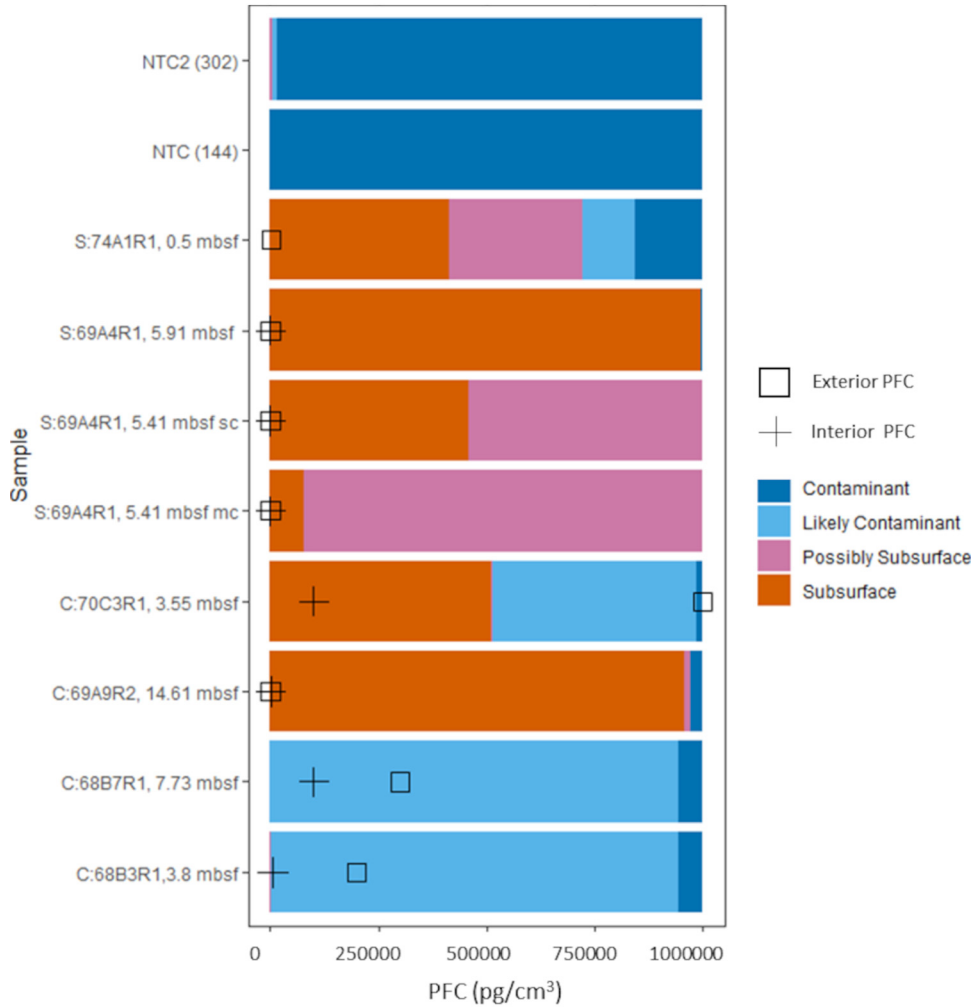
(357-69A-4R1-5.41 [sample naming method explained in Table 1, footnote b]) and three crustal samples (357-68B-1R1-1.44 mbsf, 357-69A-9R2-14.61 mbsf, and 357-70C-3R1-3.55 mbsf), resulting in 28.5 Mbp. The construction of 33 metagenome-assembled genomes (MAGs) from these metagenomes ranged in estimated completeness from <0.1% to 68% based on the presence of universal single copy genes (see Data Set S1 in the supplemental material). Finally, one sediment-basement interface sample (357-69A-4R1-5.41 mbsf) was sorted and low coverage sequencing was carried out, resulting in 227 single-cell amplified genomes (SAGs) ranging in estimated completeness from 0 to 29% (average, 6%; Data Set S1).

Amplicon sequencing negative controls yielded a total of 147 sequences, dominated by *Halomonas* (110 sequences [seq]) and *Shewanella* (21 seq). As the numbers of sequenced reads in the negative control were very small in relation to sample data



**FIG 3** Success of cell sorting and DNA amplification/sequencing in Atlantis Massif shallow subsurface samples as a function of sample cell concentration and depth in meters below seafloor (mbsf). Cell abundance data are from Früh-Green et al. (14) on parallel subsamples, and the shaded band represents the limits of cell quantification (<10 cells per cubic cm) in that study. Red filled circles labeled “Failed” in the symbol legend show that no cells were detected on the flow cytometer. Gray filled circles labeled “ND” show parallel samples for which cell density data exist in Früh-Green et al. (14) but that were not analyzed in this study. Open blue triangles labeled “Failed PCR screen” represent samples that had detectable cells in the flow cytometer but amplified DNA extracts from sorted cells did not have an observable PCR product when run on an agarose gel. Open blue squares labeled “Sequencing Failed” represent samples that had detectable cells in the flow cytometer and amplified DNA extracts from sorted cells had an observable PCR product, but amplicon sequencing failed. Filled blue squares labeled “Sequenced” represent samples that had detectable cells on the flow cytometer, and 16S rRNA genes were successfully sequenced from amplified DNA extracts from pooled sorted cells. The observation that both successful cell sorting (all blue symbols) and sequencing success (blue filled squares) did not correlate with cell density indicates that the gentle cell extraction approach used in the study was not biased by cell density.

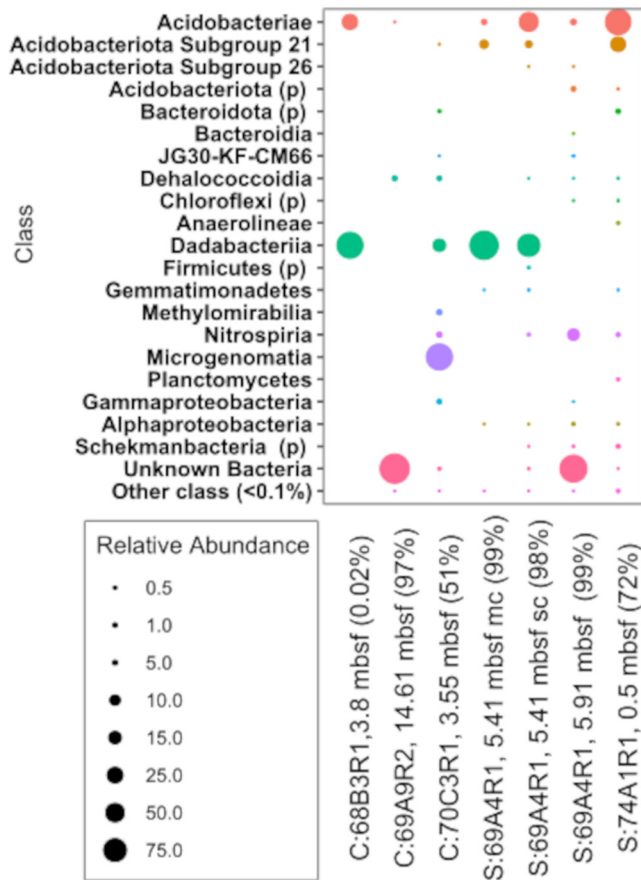
sets (see Table S1 in the supplemental material), there was a possibility of falsely removing sequences due to contamination of negative controls from sample carryover (31). To understand the potential for sequenced taxa to represent contamination from molecular analyses, the drilling process, or the surface environment, we compared the 16S rRNA gene amplicon data in the present study to those obtained by Motamedi et al. (27) that included 14 drilling fluid samples and 212 water column and serpentinite samples from IODP Expedition 357. To further explore whether the sequences in the subsurface-origin categories reflected that habitat origin, representative amplicon sequence variant (ASV) sequences from abundant taxa (taxa with >5,000 reads) were searched against the GenBank database to determine percent similarity to existing sequences and the environments of the closest matches. Finally, we carried out phylogenetic analyses on selected MAGs and SAGs using concatenated conserved genes in order to assess similarity to taxonomic relatives (and the environments these sequences were derived from; see Fig. S4a-c in the supplemental material). These methods are described in detail in Text S1 in the supplemental material. We used these data to classify sequences as originating from “Subsurface, Possibly Subsurface, Likely Contaminant, and Contaminant” (Fig. 4).



**FIG 4** Relative abundance of 16S rRNA sequence origin categories in Atlantis Massif shallow subsurface crustal and sediment samples and no-template controls (NTC, top x axis) compared with perfluorocarbon (PFC) tracer concentrations (bottom x axis). Sequence categories were defined as sequences originating from the subsurface, possibly subsurface, likely contamination, and contamination, as defined in the text and the supplemental material. The concentration of PFC tracer was measured in parallel samples from both the interior (circles) or exterior (black squares) of the core as described in Orcutt et al. (66).

Samples with a higher percentage of amplicon reads identified as subsurface in origin generally corresponded with samples with low drilling fluid tracer (i.e., perfluorocarbon [PFC]) concentrations (Table 1). Suspected contaminant sequences were removed from all subsequent amplicon analyses, resulting in the removal of 20% of sequence data from seven of the eight samples and the complete exclusion of one crustal sample amplicon data set (i.e., 357-68B-3R1-7.73 mbsf), which was composed entirely of suspected contaminants. While three samples used for metagenomic sequencing did have detectable PFC contamination tracer concentrations, the sequence grouping categorization described above indicates that the 16S rRNA gene sequences from these samples are of subsurface origin. Thus, we feel confident in describing the bulk metabolic functional potential interpretation from these data sets.

Based on 16S rRNA gene amplicon sequences that passed contamination screening, the bacterial community composition of the Atlantis Massif shallow subsurface sediment and crust samples was heterogeneous and low in diversity (Fig. 5). Caution should be taken when evaluating amplicon-based relative diversity levels between samples, as the cell extraction step did not recover all cells (Fig. S1), the cell concentration steps may not have recovered all cell types equally, and the whole-genome amplification



**FIG 5** Relative abundance of amplicon 16S rRNA gene (v4v5) sequences in Subsurface/Possibly Subsurface category taxa from bulk sorts of ~1,000 cells from (at family level) crust (C) and sediment (S) samples of the shallow subsurface of Atlantis Massif. Families which had less than 0.1% relative abundance across all samples were grouped into the “other family” category. Sediment sample 69A4R1-5.41 mbsf was processed and sequenced with two whole-genome kits: the REPLI-g single-cell kit (sc) and the REPLI-g minikit (mc). ASVs that were considered to be contaminants or likely contaminants are not shown. (p) denotes where taxa could not be resolved beyond the phylum level.

applied to concentrated bulk cell sort DNA is likely to overamplify some groups (32). However, presence-absence comparisons are useful. For example, most abundant taxa were present in both sediment and crustal libraries (Fig. 5), such as *Acidobacteria*, “*Candidatus Dadabacteria*,” *Dehalococcoidia/Chloroflexi*, and taxa classified as unknown bacteria.

Within the *Acidobacteria*, there were several clades observed. In addition, further phylogenetic analysis of the unknown bacteria 16S rRNA gene operational taxonomic units (OTUs) revealed that 30 of 39 of these OTUs formed a clade most closely related to *Acidobacteriae* subgroup 2 sequences from other terrestrial subsurface environments (33, 34) (Fig. S3). The 16S rRNA gene similarity of this clade of Atlantis Massif OTUs with other *Acidobacteria* ranges from 82.4 to 85.8%, which passes the threshold for suggesting that these organisms may be candidates of a novel class or phylum-level lineage (35). These OTUs were distinct from *Acidobacteriae* subgroup 2 ASVs identified in amplicon analysis (Fig. 5), which formed a separate related clade. Phylogenetic trees of concatenated conserved genes in assembled MAGs similarly revealed clades of *Acidobacteria* (Fig. S4A). We identified one clade most similar to genomes from subsurface metal-contaminated sediments (36) and clades related to deep seawater. By comparing sequences generated in this study at the ASV level with those from sequenced seawater samples obtained during the drilling expedition, we could identify only one acidobacterial ASV (subgroup 2) that could be a potential contaminant from seawater.

However, this ASV was also identified in higher abundance in serpentinites that seawater in the Motamedi study and was considered to be likely indigenous (27). The remaining *Acidobacteria* subgroup 2 sequences unique to this study were all found to be most similar to those from other terrestrial and marine subsurface environments (Fig. S3). Taken together, these results suggest these *Acidobacteria* are true subsurface inhabitants.

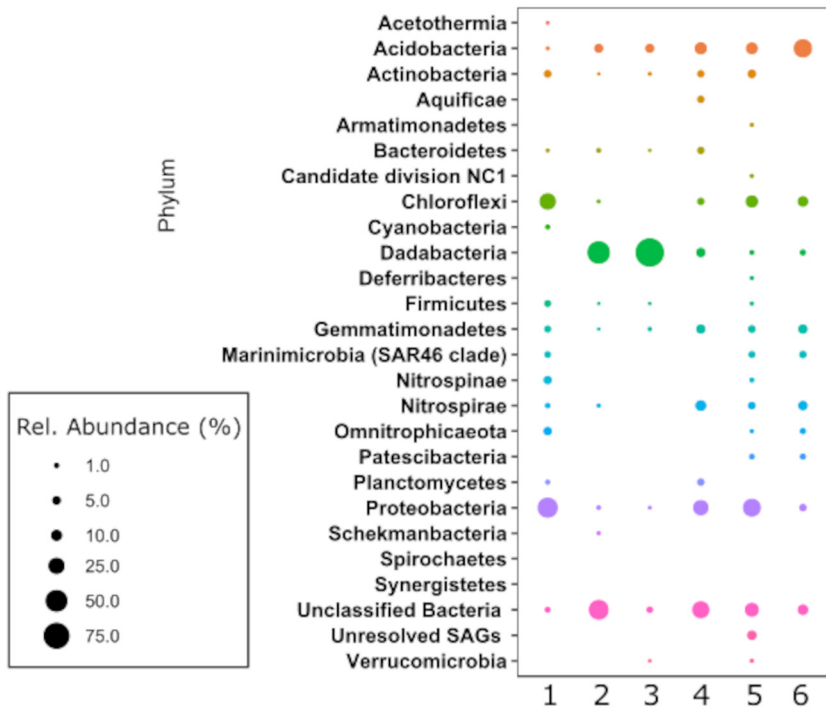
*Dehalococcoides (Chloroflexi)* were found in both crust and sediment amplicon and metagenome samples. MAGs and SAGs formed separate clades with most similarity to other sediments, which included Arctic Ocean Mid-Ocean Ridge sediment and metal contaminated aquifer sediments (Fig. S4B). “*Ca. Dadabacteria*” sequences comprised a large proportion of reads from two crust samples and a foraminiferal carbonate sand and were most closely related to environmental sequences from hot springs, hydrothermal vents, and the terrestrial subsurface (Fig. S4C). Microgenomatia (“*Candidatus Patescibacteria*”), which are widely observed in subsurface environments (37, 38), were abundant only in one crustal sample (Fig. 5). Methylomirabilia, an organism associated with anaerobic oxidation of methane (39) was identified in the 16S rRNA amplicon sequences of one crustal sample (357-70C-3R1-3.55 mbsf).

The microbial genera identified in the shallow subsurface of the Atlantis Massif by amplicon sequencing were distinct from those described previously from LCHF fluids and chimneys, which are dominated by methane-oxidizing *Methylomonas*, sulfate-reducing *Thiomicrospira*, and the archaeal methanogens and methanotrophs *Methanosarcinales* and ANME-1 (20, 21, 23). This indicates that endolithic communities in the shallow subsurface of the Atlantis Massif are distinct from those in the actively venting LCHF system. While the shallow Atlantis Massif cores have clear evidence of serpentinization in the highly altered lithologies recovered, it is likely that the LCHF system is sourced from deeper fluids than would have been accessed during the shallow drilling of IODP Expedition 357 (14, 24). Previous work in shallow gabbroic rock samples from the Atlantis Massif suggested that the microbiota are composed of widespread and ubiquitous generalists (29), with similar taxa also reported in deep gabbros from the Atlantis Bank in the Indian Ocean (30). The new results from Atlantis Massif contrast to the results of the prior studies, which may reflect the different approaches used and how potential contamination was tracked. Moreover, recent studies have documented differences between rock-hosted biofilm microbial communities and those within the fluids circulating through subsurface crust (40), which is consistent with our results. Regardless, the results of phylogenomic analysis of the shallow subsurface Atlantis Massif samples (Fig. S4a to e, Fig. S3, and Table S2) support a subsurface origin for these groups.

No archaeal sequences were observed in amplicon libraries, though one MAG classified as a *Nanoarchaeota* (51% complete) and two SAGs classified as *Crenarchaeota* (18 to 23% complete) were obtained from sample 357-69A-4R1-5.41 mbsf. The pattern of low to no archaeal reads found in the genomic data sets are similar to 16S rRNA gene amplicon and metagenomic surveys in fractures from the deep terrestrial subsurface (41) and in lower gabbroic crust from the Southwest Indian Ridge (30) but are in contrast to abundant *Archaea* found in hot anoxic crustal fluids at the Juan de Fuca Ridge flank (42, 43) and to the *Archaea*-dominated (*Methanosarcinales*) carbonate chimneys of the LCHF (20, 21). If we assume that there is no preferential bias against archaeal cells with the cell extraction approaches used in this study, our results again point to a rock-associated community in the Atlantis Massif shallow subsurface distinct from communities in the high pH fluids of LCHF.

**Different approaches yield different subsurface community compositions.** One Atlantis Massif subsurface sample—sample 357-69A-4R1-5.41 mbsf from the sediment-basement interface—was sequenced with amplicon (two replicates with different multiple displacement amplification [MDA] kits), metagenomic and single-cell genomic approaches, resulting in an amplicon library with 199 to 386 OTUs, 20 metagenome-assembled genomes, and a plate of 227 single-cell amplified genomes. In addition, we





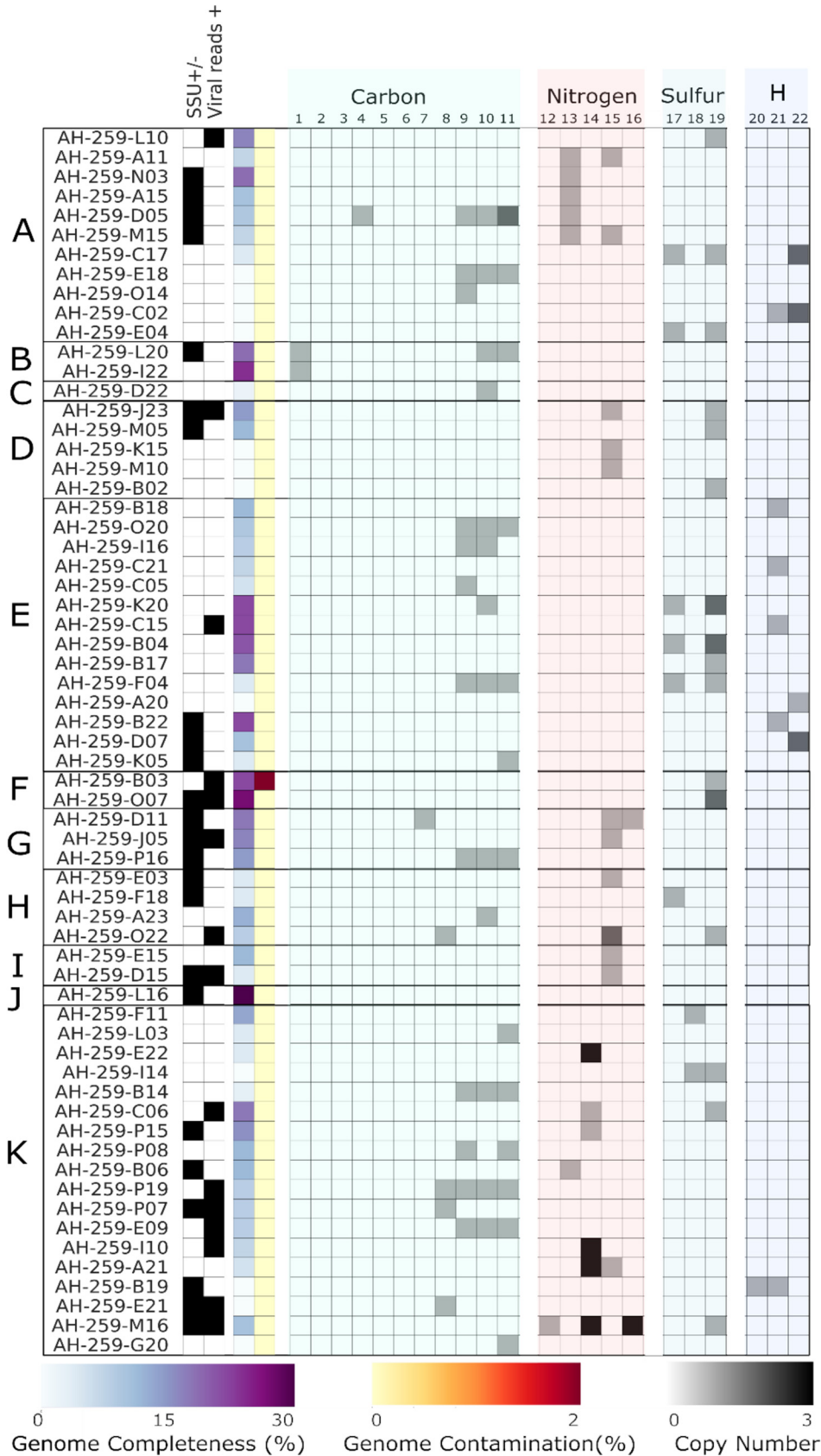
**FIG 6** Comparison of -omics approaches employed in the present study. Relative abundance (as a percentage) of bacterial sequences (phylum level) for different DNA extracts generated by different approaches from cells concentrated and sorted from one subsurface sample (from sediment sample 357-69A-4R1-5.41 mbsf). Lanes: 1, amplicon sequences from bulk extractions as in Motamedi et al. (27), total 44,443 sequences (seqs); 2 and 3, 16S rRNA gene V4 hypervariable region amplicon sequences from ~1,000 cells sorted together using two different MDA kits, the Repli-G single-cell kit (total 49,920 seqs) (lane 2) and Repli-G minikit (total 59,649 seqs) (lane 3); lane 4, bacterial metagenome-assembled genomes (MAGs) from metagenomic DNA sequences generated from ~1,000 cells sorted together, total of 20 MAGs; lanes 5 and 6, bacterial single-cell amplified genomes (SAGs) from cells sorted individually, with taxonomic identity indicated by either from checkM ID of SAG DNA after low-coverage (LoCoS) sequencing (lane 5) (total of 227 SAGs) or from Sanger sequencing of the 16S rRNA gene (lane 6) (total of 93 seqs).

compared these data with amplicon sequencing results obtained from a bulk DNA extraction protocol (27). The microbial diversity suggested by each of the sequencing approaches differed substantially (Fig. 6), highlighting the influence that amplification bias has on community structure analysis from low biomass samples. For example, cell sort amplicon data indicated the dominance of candidate phylum Dadabacteria (formally SBR1093); however, only one MAG and one SAG belonging to Dadabacteria was identified in the metagenome and single-cell genome data, respectively. The highest diversity of phyla observed was from the single-cell genomics data set of 227 SAGs. Within this plate, the best taxonomic classifications resulted from Sanger sequencing of the 16S rRNA gene (which was successful for 94 of the 227 SAGs), as the partial genomes obtained by low-coverage genome sequencing often did not contain sufficient marker genes for identification. The SAG-based community structure was most similar to that generated from bulk DNA extraction and sequencing techniques in terms of relative abundance of phyla (27). Despite the fact that the “gentle” extraction technique did not quantitatively recover all cells, our results indicate that the relative community structure was still maintained and not biased by the cell extraction technique. Thus, SAG-level microbial community structure analysis of low biomass samples, even with low-coverage genome sequencing, is likely a more accurate representation of relative diversity compared to the other cell sorting-based methods that pool cells and amplify DNA, reflecting the bias introduced from multiple displacement amplification and primer bias in the bulk cell sorts. We highlight here that the SAG-based approach required only a few grams of sample material, whereas bulk DNA extraction

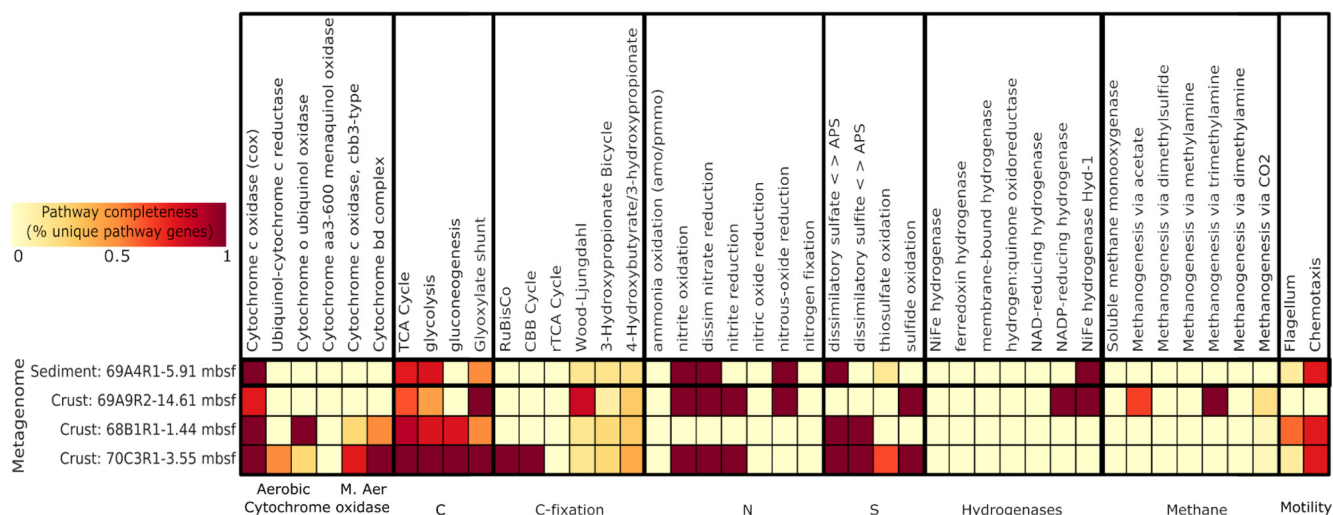
techniques in other studies required concentration of DNA from several dozens of grams of material (27, 30).

**Microbial functional potential in the Atlantis Massif subsurface.** The four metagenomes and 227 low-coverage SAGs were screened for essential genes in known respiratory and carbon fixation pathways, methane metabolism, and sulfur and nitrogen cycling (Fig. 7 and 8; see also Data Set S1 in the supplemental material). Overall, heterotrophic metabolisms in the Atlantis Massif shallow subsurface appears common, as the tricarboxylic acid (TCA) cycle was observed in all metagenomes, and it was prevalent in low-coverage SAGs and MAGs, as were genes associated with glycolysis/gluconeogenesis. In contrast, key genes for autotrophic carbon fixation pathways were rare. For example, essential genes for the reductive TCA cycle (ATP citrate lyase, 2-oxoglutarate:ferredoxin oxidoreductase) were not found in any data set. Some genes related to the 3-hydroxypropionate bicycle (3-HP bicycle) and the 3-hydroxypropionate/4-hydroxybutyrate cycle were detected, but only one acetyl coenzyme A (acetyl-CoA)/propionyl-CoA carboxylase (key gene of 3-HP bicycle) was detected in all four metagenomic libraries. Putative RuBisCO genes and a nearly complete Calvin-Benson-Bassham (CBB) cycle were found in crustal metagenome 357-70C-3R1-3.55 mbsf; however, it could not be associated with a MAG (Fig. S5). In the SAGs, the potential for carbon fixation via the CBB was identified in *Crenarchaeota* (AH-259-L20 and AH-259-I22) based on the presence of RuBisCO (Fig. 7). The presence of CBB and reductive TCA (rTCA) autotrophic pathway genes have previously been associated with ultramafic-hosted vent fluids and chimneys from the Lost City and Logatchev hydrothermal fields on the Mid-Atlantic Ridge (19, 44). The Wood-Ljungdahl (WL) pathway for anaerobic autotrophy is thought to potentially have emerged in serpentinizing systems, such as the Atlantis Massif, and may be one of the oldest metabolisms on Earth (45, 46). One *Dehalococcoidetes* MAG (MAG 22; 51% complete, 0.99% estimated genome contamination) encoded a nearly complete WL pathway, missing two enzymes in the pathway (methylene tetrahydrofolate reductase, *metF*, and 5-methyltetrahydrofolate methyltransferase, *ascE*). This MAG contained the only anaerobic carbon monoxide dehydrogenase (CODH)/acetyl-CoA synthase (ACS) complex found in all MAGs and SAGs, as well as key genes for nitrite reduction, sulfite reduction, and hydrogenases (Ni-Fe-Se/Fe-Fe; Fig. S5). An incomplete WL pathway occurs in *Dehalococcoides mccartyi* strain 195, which also lacks *metF*, and nonetheless can incorporate both formate and CO into biomass via an alternative pathway for methyl-tetrahydrofolate production (47). Reads for *Dehalococcoidetes* MAG22 were recruited entirely from the crustal metagenome from sample 357-69A-9R2-14.61 mbsf, a serpentinized dunite that was the deepest sample obtained from the site closest to the LCHF (Fig. 1).

While there was limited evidence for autotrophy in samples in this study, aerobic carbon monoxide dehydrogenases were widespread in diverse phyla in the SAG and MAG data sets, suggesting CO metabolism in the Atlantis Massif shallow subsurface may be important for energy conservation. CO assimilation and metabolism are thought to potentially be important in serpentinizing systems, where CO<sub>2</sub> is limited due to alkaline conditions (16). Aerobic carbon monoxide dehydrogenase subunits were found in five MAGs (from the *Acidobacteria*, *Chloroflexi*, and *Proteobacteria*) and in 21 diverse SAGs (11% of SAGs ranging from 0 to 21% completeness; Fig. 7; see also Fig. S5 and Data Set S1). The crenarchaeal SAGs (AH-259-L20 and AH-259-I22) with aerobic CO dehydrogenases also contained RuBisCO coding sequences (CDSs) associated with the CBB (Fig. 7). Aerobic CO-oxidizing microorganisms are widespread in both marine (48) and terrestrial settings (49). However, previous metagenomic surveys of Lost City chimneys indicate this gene was rare, with only one shotgun read recovered (50). CO oxidation can be used to conserve energy supporting aerobic heterotrophic organisms in oligotrophic or nutrient-stressed conditions (49, 51), while some aerobic carboxydrotrophs can use the energy conserved to support carbon fixation through the CBB cycle (49, 52). The importance of energy and carbon conservation in the Atlantis Massif subsurface is also suggested by the presence of isocitrate lyase and malate



**FIG 7** Metabolic functional potential of selected SAGs from sediment sample 357-69A-4R1-5.41 mbsf. Organization of y axis: bacterial or archaeal phyla of each SAG: *Acidobacteria* (A), *Crenarchaeota* (B), (Continued on next page)



**FIG 8** Metabolic potential in subsurface sediment and crustal microbial communities from Atlantis Massif IODP Expedition 357 samples based on metagenomes constructed from bulk cell sorts. Pathway completeness is a measure of the percentage of genes unique to that pathway as in Graham et al. (76). M. Aer oxidase, microaerobic cytochrome oxidase.

synthase, key genes in the glyoxylate shunt pathway (Fig. 7 and 8 and Data Set S1). Six of the MAGs and nine SAGs had at least one of these key genes. This pathway enables the CO<sub>2</sub> releasing steps of the TCA cycle to be bypassed, conserving carbon in an oligotrophic environment. The pathway enables microorganisms to use acetate as the sole carbon source (53). Acetate has been observed in micromolar concentrations in LCHF fluids (54).

While autotrophic metabolisms common in LCHF fluids were absent in the metagenomic data sets, the ability for formate utilization was present in our data (Data Set S1), which is consistent with prior observations of formate-dependent metabolisms at the LCHF (17, 23). Formate is associated with serpentinizing systems and has been detected in deeply sourced Atlantis Massif hydrothermal vent fluids (16, 54, 55). Several types of formate dehydrogenases and associated transporters (including oxalate:formate antiporters) were found throughout all data sets. One acidobacterial MAG (MAG 31) had multiple putative types: the alpha subunit of formate dehydrogenase (FDH), formate dehydrogenase N (FDN), and formate dehydrogenase O (FDO). FDO allows the use of formate as an electron donor with oxygen as a terminal electron acceptor, while FDN is induced by nitrate during anaerobic growth (56). Seven SAGs contained formate dehydrogenase H (FDH), part of the fermentative hydrogenlyase complex (57) (six *Alphaproteobacteria* and an *Acidobacteria* subgroup 21). Formate dehydrogenase, as well as formate-tetrahydrofolate ligase, a key enzyme that acts upon formate in the WL pathway, were identified in *Dehalococcoidetes* MAG 22. As autotrophic pathways are largely absent in our data, it is unlikely that formate plays a role in anabolic processes in most organisms. Rather, available formate may be used as an

**FIG 7 Legend (Continued)**

*Armatimonadetes* (C), unclassified bacteria (D), *Chloroflexi* (E), “Ca. Dadabacteria” (F), *Gemmatimonadales* (G), “Ca. Marinimicrobia” (H), *Nitrospira* (I), “Ca. Patescibacteria” (J), and *Proteobacteria* (K). Organization of x-axis columns: presence (filled)/absence (unfilled) of ribosomal small subunit (SSU) and/or viral reads in SAG, estimated completeness of SAG based on single copy marker genes (scaled from 0 to 30% per legend), estimated percentage of contamination of genome from checkM (scaled from 0 to 2% per legend), and copy number of metabolic function genes (scaled from 0 to 3 copies per legend) grouped as RuBisCO (1), phosphoribulokinase (2), carbon monoxide dehydrogenase (CoSH)/acetyl-CoA synthase (ACS) (3), acetyl-CoA synthase (4), 4-hydroxybutyryl-CoA dehydratase (5), methane monooxygenase (6), isocitrate lyase (7), malate synthase (8), CODH large chain (9), CODH medium chain (10), CODH small chain (11), ammonia monooxygenase (12), periplasmic nitrate reductase (13), respiratory nitrate reductase (14), nitrite reductase (15), nitric oxide reductase (16), adenylylsulfate reductase (17), phosphoadenosine phosphosulfate reductase (18), sulfite reductase (19), periplasmic [NiFe] hydrogenase (20), periplasmic [NiFeSe] hydrogenase (21), and NADP-reducing hydrogenase [FeFe] (22).

electron donor for energy conservation in the subsurface, and/or it might also be assimilated by heterotrophs.

No essential genes for methane metabolisms, such as methane monooxygenase (*mmoX*, *pmoA*) or methyl coenzyme M reductase associated with methane oxidation/methanogenesis, were found in any MAGs or SAGs generated in this study, suggesting that methane cycling is not a dominant metabolism in the shallow serpentinite subsurface of the Atlantis Massif. This is in contrast to methane cycling observed to support LCHF venting fluids and chimney microbiota (17, 21, 22, 58) and the observation of elevated methane in Atlantis Massif samples during drilling (2 to 48 nM in samples compared to below detection limit [0.7 nM] in bottom water) (14), though consistent with cultivation attempts in parallel samples, which did not identify any methanogens (59). Acetyl-CoA decarboxylase/synthase complex genes associated with methanogenesis via acetate (via the WL pathway), and trimethylamine methyltransferase associated with methane production via trimethylamine, were found in the deepest crustal metagenome from sample 357-69A-9R2-14.61 mbsf (Fig. 8). Alkanes produced in the Atlantis Massif subsurface and found in hydrothermal fluids were thought to be a potential C source in this environment (16). Alkane degradation genes associated with known aerobic and anaerobic pathways were rare, with only one propane and butane monooxygenase recovered from crustal metagenomes (Data Set S1). Thus, the use of alkanes as a major nutrient source for crustal microbiota in the shallow subsurface of Atlantis Massif is not supported by metagenomic evidence in our study.

Beyond carbon cycling potential, we also examined terminal electron acceptor pathways in the Atlantis Massif subsurface. Aerobic-type cytochrome oxidases were the most prevalent in all metagenomes (Fig. 8), with subunits for cytochrome c oxidase found in 22% of SAGs and 51% MAGs generated. Oxygen-sensitive cytochromes Cbb and Bd type were found in 4% of SAGs and 6% of the MAGs, and recruited from only crustal metagenomes (Fig. 8). Key genes associated with denitrification (nitrate reductase, nitrite reductase, nitric oxide reductase) were recovered, though respiratory nitrate reductase could not be binned to a MAG. Respiratory nitrate reductases were found in six proteobacterial SAGs (Data Set S1). The metabolic potential for ammonia oxidation or nitrogen fixation was not observed in the metagenomic sequences (Fig. 8). This is in contrast to the diversity of nitrogen fixation genes found in biofilms from carbonate chimneys at the LCHF (58) but similar to the observation of nitrogen cycling potential in deep gabbros of the Atlantis Bank (30). S-cycling genes as sulfate and sulfite reductases were found in all data sets (Fig. 7 and Fig. S5). Sulfate reduction is an important metabolism in LCHF vent fluids and likely in the subsurface, while chimney exteriors are associated more with sulfur oxidation (17). Potential metal-respiring and metal-associated CDSs were found in all metagenomes (Data Set S1) and indicate that detoxification and resistance to metals such as mercury and arsenic may additionally be important in the Atlantis Massif subsurface. Thus, overall, diverse aerobic and anaerobic metabolic strategies appear to be supported throughout the Atlantis Massif shallow subsurface.

Virus-like sequences were identified in 9 (of 33) MAGs and in 60 (of 227) SAGs, resulting in ~26% of the partially complete genomes with evidence of past or current viral interactions (Fig. 7 and Data Set S1). The virus-like sequences discovered in this study may be coming from either active infections or prophages. As the genome completeness estimates were low for the MAGs and SAGs, cells with evidence of viral interactions may be higher than observed (60). Viral sequences were found associated with sediment SAGs across a diversity of phyla, including *Acidobacteria*, *Chloroflexi* (*Dehalococcoides*), "*Ca. Dadabacteria*," *Gemmatimonadetes*, *Gammaproteobacteria* (*Halomonas*), *Deltaproteobacteria* (*Deferrisoma*), "*Candidatus Marinimicrobia*" (SAR406), *Nitrospirae*, *Nitrospinae*, and *Alphaproteobacteria*. Alphaproteobacterial SAGs classified as *Rhodospirillales* had the highest number of associated viral reads, representing 8 of the 39 SAGs with viral sequences in category 1 and category 2. Because multiple displacement amplification (MDA) amplifies only double-stranded DNA (dsDNA) and single-stranded

DNA (ssDNA), RNA viruses were not detected in this study. For comparison, warm, anoxic crustal fluid flowing through the basaltic basement at the Juan de Fuca Ridge flank identified abundant viruses in the crustal fluids, with archaeal viruses the most common, and only 20 to 36% of viruses predicted to infect bacterial cells (61). This is the second documentation of the host phylogenetic affiliation of viruses in the crustal biome, and the first to include cells from rock as opposed to solely in the crustal fluids, though infected cells from fluids would have been captured in our cell extraction methods as well. These data in rocks and sediments add to growing evidence that viral processes extend to the deep biosphere (61), and may play an important role in evolution and ecological functioning, even in low biomass environments.

**Conclusion.** The marine crustal subsurface is an undersampled enigmatic habitat, yet a deep biosphere in lithosphere has consequences for global biogeochemical cycling. In addition to the logistical obstacles with obtaining samples suitable for microbial analysis, low cell numbers make molecular approaches that identify function and diversity inherently difficult. By using FACS to concentrate cells from grams of low biomass ( $10^1$  to  $10^4$  cells  $g^{-1}$ ) subsurface crust and sediment samples to enable downstream 'omic analyses, this study represents the first metagenomic and single-cell genomic analysis of ultramafic rock and sediment in a marine serpentinizing system. This approach revealed that microbial communities in the shallow subsurface of Atlantis Massif are predominantly heterotrophic. This supports the interpretation that, as biologically available  $CO_2$  is scarce due to low dissolved inorganic carbon (DIC) and high alkalinity, it is likely that heterotrophic organisms are foundational in serpentinizing systems (62). Such organisms would have to be poised to metabolize abiotically produced small organic molecules like formate and  $CH_4$  or would need access to circulating seawater carrying dissolved organic carbon sourced from elsewhere. Additionally, we identified a potential role for carbon monoxide oxidation as a supplementary energy source in the shallow subsurface. The insights generated by this new cell sorting-based approach increase our understanding of microbial metabolism and composition in low biomass oceanic crust environments. These results are also relevant to the field of astrobiology, as serpentinizing systems are of interest both for the origins of life on Earth and the habitability of other planetary bodies such as Enceladus, a moon of Saturn, or Mars, where serpentinization reactions are thought to have occurred (63–65).

## MATERIALS AND METHODS

**Sample collection.** Samples of ultramafic and mafic rocks and sediment (Table 1) from the southern wall of the Atlantis Massif were collected during IODP Expedition 357 (October–December 2015, RRS *James Cook*), as detailed elsewhere (14) (Fig. 1). The seabed drills were equipped with tracer delivery systems to allow for evaluation of the potential for contamination of the samples (66). For this study, approximately 5-cm<sup>3</sup> subsamples of core were extracted by chisel from the whole round cores after flame sterilization of the exterior in a shipboard clean air system (KOACH model T-500-F; Koken, Ltd., Tokyo, Japan) to avoid potential airborne contaminants (e.g., dust). Sediment subsamples were collected via cut-end syringes inserted into the core interior. The subsamples were immediately mixed with a buffer ( $1 \times$  GlyTE [5% glycerol {final concentration}] in  $1 \times$  Tris-EDTA buffer {pH 8}) in 15- or 50-ml centrifuge tubes and frozen at  $-80^\circ C$  until shore-based analysis. Parallel subsamples from these same cores were collected for contamination tracer, cell density, and bulk community DNA extraction analysis, as presented elsewhere (14, 27, 66).

**Cell extraction, sorting, and DNA amplification from environmental samples.** The FACS cell concentration approach to enable downstream sequencing required cells to be extracted from the rock and sediment matrix into liquid buffer (Fig. 2). Following testing of various cell extraction protocols, we developed a “gentle” method for cell extraction that employed only physical disruption in salt buffers (described in Text S1 in the supplemental material). While this “gentle” method does not result in complete liberation of all cells from test rock samples, acids, detergent, and density gradient reagents commonly used for quantitative cell extraction (“harsh” methods) (67, 68) were deemed to interfere with downstream DNA amplification (see Text S1 and Fig. S1 in the supplemental material). To create a cell extract with the “gentle” method, 10 mM EDTA (final concentration) was added to the sample tubes containing thawed rock pieces in  $1 \times$  GlyTE buffer followed by physical disruption by shaking on a vortex at setting 7 (model Vortex-Genie-2) for 5 min and 1 min of sonication of the sample while on ice. Rock particles were pelleted by centrifuging at 1,000 rpm, and the remaining supernatant centrifuged at 10,000 rpm in order to pellet cells suspended in liquid. Pelleted cells were resuspended in 1 ml of GlyTE. A detailed step-by-step protocol is found on Protocols.io with DOI number [dx.doi.org/10.17504/protocols.io.bvrnm546](https://doi.org/10.17504/protocols.io.bvrnm546).

The cell extracts were thawed, mixed with SYTO 9 Green Fluorescent Nucleic Acid stain (Thermo Fisher Scientific; 5  $\mu$ M final concentration) and passed through a 40- $\mu$ m-mesh nylon filter prior to FACS. Cells were sorted in a clean room within the Single Cell Genomics Center (SCGC) located at the Bigelow Laboratory for Ocean Sciences using a Becton Dickinson (formerly Cytopeia) Influx Mariner with sterile 15 ppt NaCl solution as sheath fluid (69). For bulk concentration of cells for screening, collections of 1,000 cell-like particles were sorted into a 1.5-ml centrifuge tube containing a 1- $\mu$ l landing pad of 1  $\times$  GlyTE using a blue solid-state laser (488 nm) line for excitation. The sort gate was manually set based on red (692/40) and green (531/40) relative fluorescence (example flow cytogram provided in Fig. S2). Sorted cells were physically lysed (five freeze-thaw cycles) before chemical lysis, neutralization, and multiple displacement amplification (MDA) with the REPLI-g minikit (Qiagen, Valencia, CA) according to the manufacturer's protocol. Dilutions of amplified DNA (0, 1:10, 1:100) were used to verify a 16S rRNA gene product via PCR with universal 518F and 800R primers (518F, CCAGCAGCCGCGGTAATACG; 800R, TACCAGGGTATCTAATCC) (70) prior to sequencing, with the most successful dilution of DNA treated with the REPLI-g minikit used for downstream amplicon sequencing. Negative REPLI-g amplification controls were included with every batch of samples.

**DNA sequencing and analysis.** Illumina-based amplicon sequencing of the 16S rRNA gene V4V5 hypervariable region from the bulk cell sort amplified DNA was done at the Marine Biological Laboratory (MBL) (Woods Hole, MA, USA) as part of the Deep Carbon Observatory's Census of Deep Life, as described in the supplemental material. Additional amplicon sequencing was done at the Integrated Microbiome Resource at Dalhousie University. REPLI-g amplification-negative controls were pooled and included with every sequencing batch. The sequence data sets were analyzed using DADA2 (71) for classification and generation of amplicon sequence variants (ASVs) and *mothur* v.1.39.5 (72) for grouping sequences into operational taxonomic units (OTUs) of  $\geq 97\%$  sequence similarity, as described in the supplemental material. Additionally, comparative analyses of ASVs with other sequence data from IODP Expedition 357 were carried out to determine the proportion of sequences that may have sourced from contamination (supplemental material). We compared our amplicon sequence data to those of Motamedi et al. (27) which explicitly examined the potential for contamination in Atlantis Massif subsurface samples through amplicon analyses. This included 212 amplicon data sets derived from water column samples collected during the IODP X357 drilling expedition, as well as 14 drilling grease samples. The analyses carried out in Motamedi et al. (27) generated a data set of 684 ASVs considered to be "Likely Indigenous" microbial sequences by carrying out differential abundance and simple overlap analyses on their larger data set. The cores from the Motamedi study were collected during the same research expedition, and subsamples split for analyses by separate lab. ASV sequences that were identified as 99% similar to those in the no-template controls, or in the grease and water samples from the Motamedi study were considered to be contaminants. An exception to this was if such reads were also identified in the Motamedi study as likely to be indigenous via their methods and larger data set. In such instances, we examined the environmental source of the closest BLAST match (see Table S2 in the supplemental material). Based on this information, the ASV was manually assigned as either possibly subsurface or likely contaminant. This resulted in grouping sequences in four categories of sequence origin: subsurface, possibly subsurface, likely contamination, and contamination.

Based on diversity in 16S rRNA gene amplicon sequence screens, one sediment and three crustal samples were chosen for additional metagenomic sequencing at the MBL (highlighted in bold in Table 1). DNA sequencing, analysis, and metagenome assembly methods are described in detail in the supplemental material. Metagenome-assembled genomes (MAGs) were annotated using Prokka (73), the Joint Genome Institute (JGI) Integrated Microbial Genomes & Microbiomes (IMG/M) pipeline (74), and GhostKoala (75). Genes of interest in the MAGs were searched for using the python script `dna_analysis.py` (<https://github.com/JackieGOO/dna-profiling>). A heatmap of completeness of pathways (based on proteins unique to that pathway) and proteins of interest was made using KEGG annotated genes from GhostKoala, and the python script `KEGG-decoder.py` (76) with some modification; the altered python script and all other scripts used can be found at <https://github.com/JackieGOO/AMRocks>.

The generation and low-coverage genomic sequencing of single-cell amplified genomes (SAGs) from one sediment-basement interface sample (sample 357-69A-4R1-5.41 m below seafloor [mbsf]) was performed at the SCGC with a protocol described previously (77). Briefly, cell-like particles were individually sorted into 384-well plates, followed by cell lysis and whole-genome amplification using the WGA-X DNA polymerase, and whole-genome sequencing, genome assembly, and annotation using Prokka carried out as described in reference 77. Multiple approaches were used to classify the taxonomy of the resulting SAGs and was similar to the steps carried out as for MAGs as described in the supplemental material.

Phylogenomic trees were created using concatenated genes from abundant taxa grouping in the Subsurface and Possibly Subsurface categories to further examine the potential for these taxa to represent contaminants as described in the Supplemental Methods. SAG and MAG contigs were run through VirSorter (78) (version 1.0.5, using the "--db 2" option) to identify potential viruses. Contigs and contig regions falling into VirSorter categories 1 and 2 were considered viral.

**Data availability.** 16S rRNA gene amplicon data sets are available in NCBI SRA under project numbers [PRJNA610843](#) and [PRJNA590516](#). Metagenomic data sets are available at the Joint Genome Institute's IMG/MER database under accession numbers [3300031850](#) (sample 357-70C-3R1-3.55 mbsf), [3300031906](#) (sample 357-69A-4R1-5.41 mbsf), [3300031907](#) (sample 357-69A-9R2-14.61 mbsf), and [3300022577](#) (sample 357-68B-1R1-1.44 mbsf). Sanger sequenced 16S rRNA genes from single-cell genomes from sediment sample 357-69A-4R1-5.41 mbsf are available on SRA under accession numbers [MN715239](#) to [MN715312](#). SAG and MAG data sets are available at the Joint Genome Institute's IMG/MER under GOLD Study identifier (ID)

Gs0136008 and Gs0133404, respectively. Consensus ID, genome statistics for all MAGs and SAGs, and their accession numbers on IMG are provided in Data Set S1.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**DATA SET S1**, XLSX file, 0.2 MB.

**TEXT S1**, DOCX file, 0.03 MB.

**FIG S1**, DOCX file, 0.1 MB.

**FIG S2**, DOCX file, 2.1 MB.

**FIG S3**, DOCX file, 1.1 MB.

**FIG S4**, DOCX file, 0.5 MB.

**FIG S5**, DOCX file, 0.2 MB.

**TABLE S1**, DOCX file, 0.01 MB.

**TABLE S2**, DOCX file, 0.01 MB.

## ACKNOWLEDGMENTS

This project would not have been possible without the support of the entire International Ocean Discovery Program (IODP) Expedition 357 scientific party and support staff from the European Consortium for Ocean Research Drilling (ECORD) Science Operator. In particular, we thank Expedition Project Managers Carol Cotterill and Sophie Green and Offshore Operations Manager David Smith; the microbiology/geochemistry team for assistance with sample processing (W. Brazelton, K. Hickok, S. Lang, M. Lilley, Y. Morono, M. Quéméneur, M. Schrenk, C. Thornton, K. Twing, and N. Xiao), and the entire core description team. Additional thanks to B. Brazelton, Y. Morono, and S. Motamedi for sharing data during the project. This research used samples and data provided by the International Ocean Discovery Program. We are grateful to the staff of the Single Cell Genomics Center for their support of the flow cytometric sorting and DNA amplification. Some DNA sequencing was done at the Marine Biological Laboratory through the Census of Deep Life, and we are grateful to Mitch Sogin and Hilary Morrison. We are grateful to the Joint Genome Institute for their free IMG resource.

Funding for this work was provided by the Center for Dark Energy Biosphere Investigations (C-DEBI) funded by the U.S. National Science Foundation (NSF) award OCE-0939564 (subawards to J.G. and B.N.O.), NSF awards DEB-1441717 and OCE-1335810 to R.S., the U.S. Science Support Program to the IODP [NSF award OCE-1450528, subaward 16(GG009393-01) to B.N.O.], the European Consortium for Ocean Research Drilling (ECORD), the Swiss National Science Foundation (SNSF) project 200021\_163187 (to G.L.F.-G.), the Hanse Institute for Advanced study (fellowship to B.N.O.), and the Deep Carbon Observatory funded by the Alfred P. Sloan Foundation (Sloan-G-2015-14084-Deep Life Community subaward to B.N.O. and for Census of Deep Life sequencing).

## REFERENCES

- Orcutt BN, LaRowe DE, Biddle JF, Colwell FS, Glazer BT, Reese BK, Kirkpatrick JB, Lapham LL, Mills HJ, Sylvan JB, Wankel SD, Wheat CG. 2013. Microbial activity in the marine deep biosphere: progress and prospects. *Front Microbiol* 4:189. <https://doi.org/10.3389/fmicb.2013.00189>.
- Trembath-Reichert E, Morono Y, Ijiri A, Hoshino T, Dawson KS, Inagaki F, Orphan VJ. 2017. Methyl-compound use and slow growth characterize microbial life in 2-km-deep seafloor coal and shale beds. *Proc Natl Acad Sci U S A* 114:E9206–E9215. <https://doi.org/10.1073/pnas.1707525114>.
- Jørgensen BB, Marshall IPG. 2016. Slow microbial life in the seabed. *Annu Rev Mar Sci* 8:311–332. <https://doi.org/10.1146/annurev-marine-010814-015535>.
- Flemming H-C, Wuertz S. 2019. Bacteria and archaea on Earth and their abundance in biofilms. *Nat Rev Microbiol* 17:247–260. <https://doi.org/10.1038/s41579-019-0158-9>.
- Magnabosco C, Lin L-H, Dong H, Bomberg M, Ghiorse W, Stan-Lotter H, Pedersen K, Kieft TL, van Heerden E, Onstott TC. 2018. The biomass and biodiversity of the continental subsurface. *Nat Geosci* 11:707–717. <https://doi.org/10.1038/s41561-018-0221-6>.
- Shah Walter SR, Jaekel U, Osterholz H, Fisher AT, Huber JA, Pearson A, Dittmar T, Girguis PR. 2018. Microbial decomposition of marine dissolved organic matter in cool oceanic crust. *Nat Geosci* 11:334. <https://doi.org/10.1038/s41561-018-0109-5>.
- Orcutt BN, Sylvan JB, Rogers DR, Delaney J, Lee RW, Girguis PR. 2015. Carbon fixation by basalt-hosted microbial communities. *Front Microbiol* 6:904. <https://doi.org/10.3389/fmicb.2015.00904>.
- Meyer JL, Jaekel U, Tully BJ, Glazer BT, Wheat CG, Lin H-T, Hsieh C-C, Cowen JP, Hulme SM, Girguis PR, Huber JA. 2016. A distinct and active bacterial community in cold oxygenated fluids circulating beneath the western flank of the Mid-Atlantic ridge. *Sci Rep* 6:22541. <https://doi.org/10.1038/srep22541>.
- Orcutt B, D'Angelo T, Jungbluth SP, Huber JA, Sylvan JB. February 2020. Microbial life in oceanic crust. *OSF Preprints*. <https://doi.org/10.31219/osf.io/2wxe6>.



10. Labonté JM, Lever MA, Edwards KJ, Orcutt BN. 2017. Influence of igneous basement on deep sediment microbial diversity on the eastern Juan de Fuca Ridge flank. *Front Microbiol* 8:1434. <https://doi.org/10.3389/fmicb.2017.01434>.
11. Lever MA, Alperin M, Engelen B, Inagaki F, Nakagawa S, Steinsbu BO, Teske A. 2006. Trends in basalt and sediment core contamination during IODP Expedition 301. *Geomicrobiol J* 23:517–530. <https://doi.org/10.1080/01490450600897245>.
12. Lever MA, Torti A, Eickenbusch P, Michaud AB, Šantl-Temkiv T, Jørgensen BB. 2015. A modular method for the extraction of DNA and RNA, and the separation of DNA pools from diverse environmental sample types. *Front Microbiol* 6:476. <https://doi.org/10.3389/fmicb.2015.00476>.
13. Santelli CM, Banerjee N, Bach W, Edwards KJ. 2010. Tapping the subsurface ocean crust biosphere: low biomass and drilling-related contamination calls for improved quality controls. *Geomicrobiol J* 27:158–169. <https://doi.org/10.1080/01490450903456780>.
14. Früh-Green GL, Orcutt BN, Rouméjon S, Lilley MD, Morono Y, Cotterill C, Green S, Escartin J, John BE, McCaig AM, Cannat M, Menez B, Schwarzenbach EM, Williams MJ, Morgan S, Lang SQ, Schrenk MO, Brazelton WJ, Akizawa N, Boschi C, Dunkel KG, Quéméneur M, Whattam SA, Mayhew L, Harris M, Bayrakti G, Behrmann J-H, Herrero-Bervera E, Hesse K, Liu H-Q, Ratnayake AS, Twing K, Weis D, Zhao R, Bilenker L. 2018. Magmatism, serpentinization and life: insights through drilling the Atlantis Massif (IODP Expedition 357). *Lithos* 323:137–155. <https://doi.org/10.1016/j.lithos.2018.09.012>.
15. Proskurowski G, Lilley MD, Seewald JS, Früh-Green GL, Olson EJ, Lupton JE, Sylva SP, Kelley DS. 2008. Abiogenic hydrocarbon production at Lost City hydrothermal field. *Science* 319:604–607. <https://doi.org/10.1126/science.1151194>.
16. Schrenk MO, Brazelton WJ, Lang SQ. 2013. Serpentinization, carbon, and deep life. *Rev Mineral Geochem* 75:575–606. <https://doi.org/10.2138/rmg.2013.75.18>.
17. Lang SQ, Früh-Green GL, Bernasconi SM, Brazelton WJ, Schrenk MO, McGonigle JM. 2018. Deeply-sourced formate fuels sulfate reducers but not methanogens at Lost City hydrothermal field. *Sci Rep* 8:755. <https://doi.org/10.1038/s41598-017-19002-5>.
18. Kelley DS. 2005. A serpentinite-hosted ecosystem: the Lost City hydrothermal field. *Science* 307:1428–1434. <https://doi.org/10.1126/science.1102556>.
19. Perner M, Kuever J, Seifert R, Pape T, Koschinsky A, Schmidt K, Strauss H, Imhoff JF. 2007. The influence of ultramafic rocks on microbial communities at the Logatchev hydrothermal field, located 15°N on the Mid-Atlantic Ridge. *FEMS Microbiol Ecol* 61:97–109. <https://doi.org/10.1111/j.1574-6941.2007.00325.x>.
20. Brazelton WJ, Ludwig KA, Sogin ML, Andreishcheva EN, Kelley DS, Shen C-C, Edwards RL, Baross JA. 2010. Archaea and bacteria with surprising microdiversity show shifts in dominance over 1,000-year time scales in hydrothermal chimneys. *Proc Natl Acad Sci U S A* 107:1612–1617. <https://doi.org/10.1073/pnas.0905369107>.
21. Brazelton WJ, Schrenk MO, Kelley DS, Baross JA. 2006. Methane- and sulfur-metabolizing microbial communities dominate the Lost City hydrothermal field ecosystem. *Appl Environ Microbiol* 72:6257–6270. <https://doi.org/10.1128/AEM.00574-06>.
22. Schrenk MO, Kelley DS, Bolton SA, Baross JA. 2004. Low archaeal diversity linked to seafloor geochemical processes at the Lost City Hydrothermal Field, Mid-Atlantic Ridge. *Environ Microbiol* 6:1086–1095. <https://doi.org/10.1111/j.1462-2920.2004.00650.x>.
23. McGonigle JM, Lang SQ, Brazelton WJ. 2020. Genomic evidence for formate metabolism by *Chloroflexi* as the key to unlocking deep carbon in Lost City microbial ecosystems. *Appl Environ Microbiol* 86:e02583-19. <https://doi.org/10.1128/AEM.02583-19>.
24. Rouméjon S, Früh-Green GL, Orcutt BN, the IODP Expedition 357 Science Party. 2018. Alteration heterogeneities in peridotites exhumed on the southern wall of the Atlantis Massif (IODP Expedition 357). *J Petrol* 59:1329–1358. <https://doi.org/10.1093/ptrology/egy065>.
25. Früh-Green G, Orcutt BN, Green SL, Cotterill C, the Expedition 357 Scientists. 2016. Atlantis Massif Serpentinization and Life-Expedition 357 of the mission-specific drilling platform from and to Southampton, United Kingdom, Sites M0068–M0076, 26 October–11 December 2015. Proceedings of the International Ocean Discovery Program, vol 357. International Ocean Discovery Program, College Station, TX.
26. Früh-Green GL, Orcutt BN, Green SL, Cotterill C, Morgan S, Akizawa N, Bayrakti G, Behrmann J-H, Boschi C, Brazelton WJ, Cannat M, Dunkel KG, Escartin J, Harris M, Herrero-Bervera E, Hesse K, John BE, Lang SQ, Lilley MD, Liu H-Q, Mayhew LE, McCaig AM, Menez B, Morono Y, Quéméneur M, Rouméjon S, Sandaruwan Ratnayake A, Schrenk MO, Schwarzenbach EM, Twing KI, Weis D, Whattam SA, Williams M, Zhao R. 2017. Expedition 357 methods. In Früh-Green GL, Orcutt BN, Green SL, Cotterill C, the Expedition 357 Scientists, Atlantis Massif Serpentinization and Life. Proceedings of the International Ocean Discovery Program, vol 357. International Ocean Discovery Program, College Station, TX.
27. Motamedi S, Orcutt BN, Früh-Green GL, Twing KI, Pendleton HL, Brazelton WJ. 2020. Microbial residents of the Atlantis Massif's shallow serpentinite subsurface. *Appl Environ Microbiol* 81:e00356-20. <https://doi.org/10.1128/AEM.00356-20>.
28. Sheik CS, Reese BK, Twing KI, Sylvan JB, Grim SL, Schrenk MO, Sogin ML, Colwell FS. 2018. Identification and removal of contaminant sequences from ribosomal gene databases: lessons from the census of deep life. *Front Microbiol* 9:840. <https://doi.org/10.3389/fmicb.2018.00840>.
29. Mason OU, Nakagawa T, Rosner M, Van Nostrand JD, Zhou J, Maruyama A, Fisk MR, Giovannoni SJ. 2010. First investigation of the microbiology of the deepest layer of ocean crust. *PLoS One* 5:e15399. <https://doi.org/10.1371/journal.pone.0015399>.
30. Li J, Mara P, Schubotz F, Sylvan JB, Burgaud G, Klein F, Beaudoin D, Wee SY, Dick HJB, Lott S, Cox R, Meyer LAE, Quémener M, Blackman DK, Edgcomb VP. 2020. Recycling and metabolic flexibility dictate life in the lower oceanic crust. *Nature* 579:250–255. <https://doi.org/10.1038/s41586-020-2075-5>.
31. Karstens L, Asquith M, Davin S, Fair D, Gregory WT, Wolfe AJ, Braun J, McWeeney S. 2019. Controlling for contaminants in low-biomass 16S rRNA gene sequencing experiments. *mSystems* 4:e00290-19. <https://doi.org/10.1128/mSystems.00290-19>.
32. Yilmaz S, Allgaier M, Hugenholtz P. 2010. Multiple displacement amplification compromises quantitative analysis of metagenomes. *Nat Methods* 7:943–944. <https://doi.org/10.1038/nmeth.1210-943>.
33. Hathaway JJM, Garcia MG, Balasch MM, Spilde MN, Stone FD, Dapkevicius MDLNE, Amorim IR, Gabriel R, Borges PAV, Northup DE. 2014. Comparison of bacterial diversity in Azorean and Hawaiian lava cave microbial mats. *Geomicrobiol J* 31:205–220. <https://doi.org/10.1080/01490451.2013.777491>.
34. Lin X, Kennedy D, Fredrickson J, Bjornstad B, Konopka A. 2012. Vertical stratification of subsurface microbial community composition across geological formations at the Hanford Site. *Environ Microbiol* 14:414–425. <https://doi.org/10.1111/j.1462-2920.2011.02659.x>.
35. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* 12:635–645. <https://doi.org/10.1038/nrmicro3330>.
36. Hug LA, Thomas BC, Brown CT, Frischkorn KR, Williams KH, Tringe SG, Banfield JF. 2015. Aquifer environment selects for microbial species cohorts in sediment and groundwater. *ISME J* 9:1846–1856. <https://doi.org/10.1038/ismej.2015.2>.
37. Herrmann M, Wegner C-E, Taubert M, Geesink P, Lehmann K, Yan L, Lehmann R, Totsche KU, Küsel K. 2019. Predominance of Cand. Patescibacteria in groundwater is caused by their preferential mobilization from soils and flourishing under oligotrophic conditions. *Front Microbiol* 10:1407–1407. <https://doi.org/10.3389/fmicb.2019.01407>.
38. Tian R, Ning D, He Z, Zhang P, Spencer SJ, Gao S, Shi W, Wu L, Zhang Y, Yang Y, Adams BG, Rocha AM, Detienne BL, Lowe KA, Joyner DC, Klingeman DM, Arkin AP, Fields MW, Hazen TC, Stahl DA, Alm EJ, Zhou J. 2020. Small and mighty: adaptation of superphylum Patescibacteria to groundwater environment drives their genome simplicity. *Microbiome* 8:51. <https://doi.org/10.1186/s40168-020-00825-w>.
39. Graf JS, Mayr MJ, Marchant HK, Tienken D, Hach PF, Brand A, Schubert CJ, Kuypers MMM, Milucka J. 2018. Bloom of a denitrifying methanotroph, 'Candidatus Methylospirillum limnetica', in a deep stratified lake. *Environ Microbiol* 20:2598–2614. <https://doi.org/10.1111/1462-2920.14285>.
40. Ramírez GA, Garber AI, Lecoeuvre A, D'Angelo T, Wheat CG, Orcutt BN. 2019. Ecology of seafloor crustal biofilms. *Front Microbiol* 10:1983. <https://doi.org/10.3389/fmicb.2019.01983>.
41. Magnabosco C, Ryan K, Lau MCY, Kuloyo O, Sherwood Lollar B, Kieft TL, van Heerden E, Onstott TC. 2016. A metagenomic window into carbon metabolism at 3 km depth in Precambrian continental crust. *ISME J* 10:730–741. <https://doi.org/10.1038/ismej.2015.150>.
42. Jungbluth SP, Grote J, Lin H-T, Cowen JP, Rappé MS. 2013. Microbial diversity within basement fluids of the sediment-buried Juan de Fuca Ridge flank. *ISME J* 7:161–172. <https://doi.org/10.1038/ismej.2012.73>.
43. Jungbluth SP, Lin H-T, Cowen JP, Glazer BT, Rappé MS. 2014. Phylogenetic diversity of microorganisms in seafloor crustal fluids from Holes 1025C and 1026B along the Juan de Fuca Ridge flank. *Front Microbiol* 5:119. <https://doi.org/10.3389/fmicb.2014.00119>.

44. Brazelton WJ, Baross JA. 2010. Metagenomic comparison of two Thiomicrospira lineages inhabiting contrasting deep-sea hydrothermal environments. *PLoS One* 5:e13530. <https://doi.org/10.1371/journal.pone.0013530>.
45. Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, Martin WF. 2016. The physiology and habitat of the last universal common ancestor. *Nat Microbiol* 1:16116. <https://doi.org/10.1038/nmicrobiol.2016.116>.
46. Preiner M, Igarashi K, Muchowska KB, Yu M, Varma SJ, Kleinermanns K, Nobu MK, Kamagata Y, Tüysüz H, Moran J, Martin WF. 2020. A hydrogen-dependent geochemical analogue of primordial carbon and energy metabolism. *Nat Ecol Evol* 4:534–542. <https://doi.org/10.1038/s41559-020-1125-6>.
47. Zhuang W-Q, Yi S, Bill M, Brissin VL, Feng X, Men Y, Conrad ME, Tang YJ, Alvarez-Cohen L. 2014. Incomplete Wood-Ljungdahl pathway facilitates one-carbon metabolism in organohalide-respiring *Dehalococcoides mccartyi*. *Proc Natl Acad Sci U S A* 111:6419–6424. <https://doi.org/10.1073/pnas.1321542111>.
48. Zafriou OC, Andrews SS, Wang W. 2003. Concordant estimates of oceanic carbon monoxide source and sink processes in the Pacific yield a balanced global “blue-water” CO budget. *Global Biogeochem Cycles* 17:1015. <https://doi.org/10.1029/2001GB001638>.
49. Cordero PRF, Bayly K, Man Leung P, Huang C, Islam ZF, Schittenhelm RB, King GM, Greening C. 2019. Atmospheric carbon monoxide oxidation is a widespread mechanism supporting microbial survival. *ISME J* 13:2868–2881. <https://doi.org/10.1038/s41396-019-0479-8>.
50. Brazelton W, Nelson S, Schrenk M. 2012. Metagenomic evidence for H<sub>2</sub> oxidation and H<sub>2</sub> production by serpentinite-hosted subsurface microbial communities. *Front Microbiol* 2:268. <https://doi.org/10.3389/fmicb.2011.00268>.
51. Robb FT, Techtman SM. 2018. Life on the fringe: microbial adaptation to growth on carbon monoxide. *F1000Res* 7:1981. <https://doi.org/10.12688/f1000research.16059.1>.
52. Tiquia-Arashiro SM. 2014. Microbial CO metabolism, p 5–9. *In* Thermophilic carboxydrotrophs and their applications in biotechnology. Springer, Cham, Switzerland.
53. Ensign SA. 2006. Revisiting the glyoxylate cycle: alternate pathways for microbial acetate assimilation. *Mol Microbiol* 61:274–276. <https://doi.org/10.1111/j.1365-2958.2006.05247.x>.
54. Lang SQ, Butterfield DA, Schulte M, Kelley DS, Lilley MD. 2010. Elevated concentrations of formate, acetate and dissolved organic carbon found at the Lost City hydrothermal field. *Geochim Cosmochim Acta* 74:941–952. <https://doi.org/10.1016/j.gca.2009.10.045>.
55. Proskurowski G, Lilley MD, Kelley DS, Olson EJ. 2006. Low temperature volatile production at the Lost City Hydrothermal Field, evidence from a hydrogen stable isotope geothermometer. *Chem Geol* 229:331–343. <https://doi.org/10.1016/j.chemgeo.2005.11.005>.
56. Berg BL, Li J, Heider J, Stewart V. 1991. Nitrate-inducible formate dehydrogenase in *Escherichia coli* K-12. I. Nucleotide sequence of the *fdnGHl* operon and evidence that *opal* (UGA) encodes selenocysteine. *J Biol Chem* 266:22380–22385. [https://doi.org/10.1016/S0021-9258\(18\)54583-X](https://doi.org/10.1016/S0021-9258(18)54583-X).
57. McDowell JS, Murphy BJ, Haumann M, Palmer T, Armstrong FA, Sargent F. 2014. Bacterial formate hydrogenlyase complex. *Proc Natl Acad Sci U S A* 111:E3948–E3956. <https://doi.org/10.1073/pnas.1407927111>.
58. Brazelton WJ, Mehta MP, Kelley DS, Baross JA. 2011. Physiological differentiation within a single-species biofilm fueled by serpentinization. *mBio* 2:e00127-11. <https://doi.org/10.1128/mBio.00127-11>.
59. Quéménéur M, Erauso G, Frouin E, Zeghal E, Vandecasteele C, Ollivier B, Tamburini C, Garel M, Ménez B, Postec A. 2019. Hydrostatic pressure helps to cultivate an original anaerobic bacterium from the Atlantis Massif seafloor (IODP Expedition 357): *Petrocella atlantisensis* gen. nov. sp. nov. *Front Microbiol* 10:1497. <https://doi.org/10.3389/fmicb.2019.01497>.
60. Munson-McGee JH, Peng S, Dewerrf S, Stepanauskas R, Whitaker RJ, Weitz JS, Young MJ. 2018. A virus or more in (nearly) every cell: ubiquitous networks of virus–host interactions in extreme environments. *ISME J* 12:1706–1714. <https://doi.org/10.1038/s41396-018-0071-7>.
61. Nigro OD, Jungbluth SP, Lin H-T, Hsieh C-C, Miranda JA, Schvarcz CR, Rappé MS, Steward GF. 2017. Viruses in the oceanic basement. *mBio* 8:e02129-16. <https://doi.org/10.1128/mBio.02129-16>.
62. Lang SQ, Brazelton WJ. 2020. Habitability of the marine serpentinite subsurface: a case study of the Lost City hydrothermal field. *Philos Trans R Soc A* 378:20180429. <https://doi.org/10.1098/rsta.2018.0429>.
63. Russell MJ, Hall AJ, Martin W. 2010. Serpentinization as a source of energy at the origin of life. *Geobiology* 8:355–371. <https://doi.org/10.1111/j.1472-4669.2010.00249.x>.
64. Schulte M, Blake D, Hoehler T, McCollom T. 2006. Serpentinization and its implications for life on the early Earth and Mars. *Astrobiology* 6:364–376. <https://doi.org/10.1089/ast.2006.6.364>.
65. Jones RM, Goordial JM, Orcutt BN. 2018. Low energy subsurface environments as extraterrestrial analogs. *Front Microbiol* 9:1605. <https://doi.org/10.3389/fmicb.2018.01605>.
66. Orcutt BN, Bergenthal M, Freudenthal T, Smith D, Lilley MD, Schnieders L, Green S, Früh-Green GL. 2017. Contamination tracer testing with seabed drills: IODP Expedition 357. *Sci Drill* 23:39–46. <https://doi.org/10.5194/sd-23-39-2017>.
67. Kallmeyer J, Smith DC, Spivack AJ, D'Hondt S. 2008. New cell extraction procedure applied to deep subsurface sediments. *Limnol Oceanogr* Methods 6:236–245. <https://doi.org/10.4319/lom.2008.6.236>.
68. Morono Y, Terada T, Kallmeyer J, Inagaki F. 2013. An improved cell separation technique for marine subsurface sediments: applications for high-throughput analysis using flow cytometry and cell sorting. *Environ Microbiol* 15:2841–2849. <https://doi.org/10.1111/1462-2920.12153>.
69. Stepanauskas R, Sieracki ME. 2007. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc Natl Acad Sci U S A* 104:9052–9057. <https://doi.org/10.1073/pnas.0700496104>.
70. Park HJ, Kim D. 2015. Isolation and characterization of humic substances-degrading bacteria from the subarctic Alaska grasslands. *J Basic Microbiol* 55:54–61. <https://doi.org/10.1002/jobm.201300087>.
71. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581–583. <https://doi.org/10.1038/nmeth.3869>.
72. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 79:5112–5120. <https://doi.org/10.1128/AEM.01043-13>.
73. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
74. Chen I-MA, Markowitz VM, Chu K, Palaniappan K, Szeto E, Pillay M, Ratner A, Huang J, Andersen E, Huntemann M, Varghese N, Hadjithomas M, Tennessen K, Nielsen T, Ivanova NN, Kyrpides NC. 2017. IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res* 45:D507–D516. <https://doi.org/10.1093/nar/gkw929>.
75. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 428:726–731. <https://doi.org/10.1016/j.jmb.2015.11.006>.
76. Graham ED, Heidelberg JF, Tully BJ. 2018. Potential for primary productivity in a globally-distributed bacterial phototroph. *ISME J* 12:1861–1866. <https://doi.org/10.1038/s41396-018-0091-3>.
77. Stepanauskas R, Fergusson EA, Brown J, Poulton NJ, Tupper B, Labonté JM, Becraft ED, Brown JM, Pachiadaki MG, Povilaitis T, Thompson BP, Mascena CJ, Bellows WK, Lubys A. 2017. Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. *Nat Commun* 8:84. <https://doi.org/10.1038/s41467-017-00128-z>.
78. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3:e985. <https://doi.org/10.7717/peerj.985>.