Summary paragraph:

Much remains unknown about what drives microbial community structure and diversity. Highly structured environments might offer clues. For example, it may be possible to identify metabolically similar species as groups of organisms that correlate spatially with the geochemical processes they carry out. Here, we use a 16S ribosomal RNA gene survey in a lake that has chemical gradients across its depth to identify groups of spatially correlated but phylogenetically diverse organisms. Some groups had distributions across depth that aligned with the distributions of metabolic processes predicted by a biogeochemical model, suggesting these groups performed biogeochemical functions. A single-cell genetic assay showed, however, that the groups associated with one biogeochemical process, sulfate reduction, contained only a few organisms that have the genes required to reduce sulfate. These results raise the possibility that some of these spatially-correlated groups are consortia of phylogenetically diverse and metabolically different microbes that cooperate to carry out geochemical functions.

Keywords: ecology / 16S / function / gradient / phylogeny / lake / co-occurrence

Introduction

Explaining the vast diversity of microbes found in many ecosystems [1,2] is a challenge for microbial ecology. Environments with chemical or other abiotic gradients like temperature have been a key resource for studying microbial ecology. For example, studies in Winogradsky columns [3], microbial mats [4], mine drainage sites [5], hydrothermal vents [6], and dimictic lakes [7] have provided insight about the relationships between environmental parameters, microbial diversity, and ecosystem functions. Microbial surveys with spatial scales comparable to those of the ecosystem gradients can identify groups of spatially-correlated organisms and relate the distribution of those organisms to the environmental gradients.

There are challenges to interpreting the relationship between organisms in spatially-correlated groups and environmental information. First, the relationship between an organism's spatial distribution and environmental parameters can be complicated. For example, a naïve expectation might be that sulfate-reducing organisms are abundant where sulfate concentrations are highest. In fact, the distribution of sulfate-reducing organisms also depends on the distribution of more favorable electron acceptors and the transport of sulfur compounds around the ecosystem. Even more subtly, bacterial populations may be capable of performing multiple metabolisms, and they can even be simply inactive. Thus, there is a need to develop techniques that provide quantitative expectations about factors that shape organismal distributions given observed environmental information.

A second challenge is that there are multiple experimental methods that can verify the relationships between function and phylogeny, but most of these methods are *in vitro* or perturb the environment [8]. A method that relates phylogeny and function without perturbing the natural ecosystem would clarify the *in situ* functional relationships between organisms in a spatially-correlated group. Deep metagenomic sequencing along with differential genome binning techniques can produce draft genomes from complex communities [9] but is expensive and cannot target specific functions.

A third challenge to studying spatially-correlated organisms in ecosystems with gradients is relating the groups' diversities to their environmental functions, especially if these organisms are unrelated. Organisms in these groups could use similar resources, as it is know that many traits are widespread in the tree of life [10], or could have recently exchange genes through horizontal gene transfer [11]. Unrelated organisms with similar distributions could also be found together because they are part of multispecies, symbiotic associations [12]. The challenge lies in differentiating between these or other possibilities.

In this paper, we investigate spatially-correlated organisms in an ecosystem with gradients. First, we conducted a microbial survey of a dimictic lake. Second, we constructed a quantitative, dynamic biogeochemical model that shows how bacteria can drive the creation of chemical gradients. Third, we show that there are many groups of spatially-correlated organisms in this lake and relate those groups to the biogeochemical model. Finally, we used a single-cell assay to investigate the functional capabilities of the groups of spatially-correlated bacteria related to one modeled process, sulfate reduction. We show that, taken together, these results raise the possibility that these spatially-correlated groups are multispecies, symbiotic associations of microbes, that is, consortia [13].

Results

Community structure is influenced by geochemistry. We performed our study in Upper Mystic Lake, a dimictic, eutrophic freshwater lake outside Boston, MA because this lake is seasonally stratified and supports complex microbial communities that catalyze well-characterized biogeochemical cycles [14-19]. The seasonal stratification means that the deepest three-quarters of this lake is anoxic, supporting fewer predators that can complicate microbial distribution patterns, and that the deeper parts of the lake are relatively isolated from external inputs.

To characterize microbial diversity, we conducted amplicon-based bacterial surveys (16S rRNA gene library from DNA samples) along a vertical transect in the lake, collecting samples at approximately each meter of depth. We grouped 16S rRNA gene sequences into operational taxonomic units (OTUs) using the ecologically-informed distribution-based clustering algorithm, which merges sequences from related organisms that have similar spatial distributions [20]. We also measured major geochemical parameters (temperature, specific conductivity, dissolved oxygen, nitrate, iron and sulfate; Fig. S1-S3).

The 16S rRNA gene survey showed that biogeochemistry had a major influence on the bacterial community structure (Fig. 1). Transitions in community structure lined up with the lake's major geochemical features: the thermocline, oxycline, and nitrocline (Fig. 1b). Cyanobacteria were most abundant near the surface (Fig. 1a). Bacteroidetes, Actinobacteria, and Proteobacteria were abundant across depths. δ-Proteobacteria, which include most of the known sulfate-reducing bacteria, were abundant only below the nitrocline where more favorable terminal electron acceptors were exhausted. The differences in geochemistry and bacterial community structure across depths suggest that organisms exploiting similar resources should have correlated distributions across depths.

Multiple groups of spatially-correlated OTUs in the lake. To identify groups of spatially-correlated organisms, we used hierarchical clustering to group the 536 most abundant OTUs into 49 groups based on the similarity of the OTUs' distributions across depths. We call these groups *operational ecological units* (OEUs) because they are groups of organisms that we expect have functional or ecological relationships (thus "ecological") but were defined in a purely statistical way (thus "operational" [21]).

Most OEUs contained OTUs from multiple phyla. OEUs ranged in size from 2 to 33 OTUs and contained 1 to 10 phyla. The number of phyla in each OEU (0.34 additional phyla per OTU beyond the first in the OEU; Fig. S4) was about as many as would be expected if phyla classifications had been randomly assigned to OTUs (0.35 +/− 0.01; 1000 permutations). To verify that the OEUs are robust to different bioinformatic methods, extraction methodologies, and sample years, we compared the results of the OEU analysis after varying each of these factors and found more OTUs together than would be expected by chance (Table S5).

A biogeochemical model reproduces chemical and biological structure and dynamics. Having characterized the lake's geochemistry and identified groups of spatially-correlated organisms, we set out to design a computational framework that predicts the function of bacteria in the lake. We found no existing dynamical model that treated all the major microbial metabolic processes in a dimictic lake, so we modified and reinterpreted a model of chemical transport and microbial metabolism designed to simulate groundwater aquifers [22]. We chose to develop a model because the distribution of bacteria is the result of complex and interdependent biogeochemical cycles and hydrodynamic transport processes in the lake.

The model we developed simulates the major chemical species, redox cycles, and transport processes in the hypolimnion (Fig. S5; Tables S1-S3). We used previously published values [22] for many parameters (Table S4) and calibrated the model to match the chemical datasets (Figs. 2, S2-S3). In this lake, the water is typically well-mixed through the lake's depth in spring. During summer, warmer water sits on top of the cooler water at the bottom of the lake. Thermal resistance to mixing across this warm-cold plane (the thermocline, about 5 m deep at the time of sampling [Fig. 1b]) partially isolates water below the thermocline (the hypolimnion) from external and atmospheric influences. Heterotrophic microbes oxidize energy-rich carbon compounds as they diffuse or settle down into the hypolimnion, and the increasingly limited availability of terminal electron acceptors for these microbes leads to vertical chemical gradients. Reduced chemical species can be transported to oxidizing conditions closer to the lake's surface, fueling additional microbial activity. The model predicts the distribution of microbial metabolic processes and chemical species abundance in the lake from spring to autumn, when the thermocline breaks down and the lake mixes again.

We used the model to simulate the lake's biogeochemical dynamics for two datasets: a time series collected in 2013 and a single-time point survey collected in 2008. In both cases, the model predicted chemical dynamics (Figs. 2, S6) that were consistent with those expected from a eutrophic, dimictic lake [19]. In 2013, we had a time series covering the five months before the bacterial survey, so we initialized the model using the observed chemical parameters from the first survey in the time series. The chemical dynamics predicted by the model (Fig. 2) accorded with our measured time series, and the predicted distribution of chemical species accorded with the final survey (Fig. S2).

Because we had no initial data for the 2008 survey, we initialized the lake in a homogenous composition, as would be expected from an idealized dimictic lake that perfectly mixed throughout its entire depth in the spring. In this case, the model predicted the emergence of chemical gradients from the initially homogenous composition (Fig. S6), and the predicted distribution of chemical species accorded with the single-timepoint survey (Fig. S3).

To relate the output of the model with our biological data, we reinterpreted the modeled rates as predictions of microbial distribution. Implicit biomass models, like the one we developed, predict the rates of processes catalyzed by all microbes performing that process and assume that the biomass of the microbial community equilibrates quickly to the changing chemical environment [22]. They also assume that “everything is everywhere” and are not constrained by ecological processes like dispersion. We therefore expected that the relative rate of a modeled process should be proportional across depths to the biomass of microbes performing that process. This interpretative framework, which we call “inferred biomass”, reinterprets implicit biomass models as hypotheses about microbial community structure. Consistent with these assumptions, our model largely reproduced the distribution of key organisms known to perform the corresponding metabolisms in 2013 (Fig. 3) and 2008 (Fig. S7).

The model captures the major patterns in the lake's chemical dynamics, but there are discrepancies between the model and observation. For example, in the final 2013 survey, oxygen concentrations reach undetectable levels at about 5 meters, increase until about 8 meters, then decrease again until reaching the detection limit at 14 meters. In contrast, the model predicts that oxygen concentrations would decrease monotonically with depth.

Many groups of spatially-correlated organisms have spatial distributions that correspond to modeled processes. 63% of the OEUs from the 2013 dataset have a spatial profile that is similar (distance less than 0.25) to one or more of the biogeochemical processes simulated in the model (Fig. S8), and some of these OEU-process pairings are supported by the previously reported ecosystem functions of one or more of the OTUs in the OEU (Table S6). This spatial alignment between OEUs and modeled processes suggests that the growth of organisms in those OEUs is dependent on the energy provided by those processes. Because these OEUs are made up of OTUs that are spatially correlated, taxonomically diverse, and spatially aligned with modeled biogeochemical processes, it may be that these OEUs are consortia of organisms in syntrophic relationships.

Not all taxa corresponding to a modeled process have the same metabolism. There are other explanations for the properties of the spatially-correlated groups. Aside from consortia, they may also be groups of functionally redundant bacteria or simply groups of organisms subject to some pressure or process that only coincidentally led to spatial alignment with one of the modeled biogeochemical processes. To distinguish these explanations, we assayed the genetic capability of the OTUs in some OEUs to carry out the biogeochemical process with which they spatially align. For example, if all OTUs in an OEU have the genetic capability to perform some process, then that OEU might represent functionally redundant organisms. Conversely, if none of those OTUs have the genetic capability, then that OEU probably has little to do with biogeochemistry. If only some OTUs in an OEU can perform some process, then that OEU might represent a consortium of syntrophic organisms.

We investigated one process, sulfate reduction, in greater detail because the spatial distribution of this process had one of the best matches between the model and observation and because there was a well-studied genetic marker for this function. The three OEUs with spatial distributions that best matched this process (Fig. 5) contained 14 OTUs that were in high abundance in both this survey and the positive control for the gene fusion assay described below. Among these 14 OTUs, 6 are classified as δ-Proteobacteria, the class that contains most of the bacteria known to reduce sulfate, and one of the δ-Proteobacteria OTUs corresponds to a known sulfate-reducing organism (Table S6). Among the other OTUs, 5 are classified as Bacteroidetes, which contains no known sulfate reducers and are instead regarded as specialists in the degradation of high molecular weight organic matter [23]. Because terminal oxidation processes of organic carbon under anaerobic conditions are rarely catalyzed by a single organism, we suspected that the sulfate reducers among the δ-Proteobacteria might be in a syntrophic relationship with the Bacteroidetes organisms, which provide low-molecular weight dissolved organic carbon to sulfate reducers. Intriguingly, the reference OTU for sulfate reduction (a clone similar to *Desulfatirhabdium butyrativorans*; Table S6) and an OTU classified as Bacteroidetes (with 93% identity to the 16S rRNA of the sugar-fermenting psychrophile *Prolixibacter bellariivorans* [24]) appeared together in the same OEU in both the 2008 and 2013 datasets, suggesting that, if some OEUs do represent consortia of syntrophic organisms, some of those association might persist across years in this ecosystem.

To probe the genetic capability of OTUs to perform sulfate reduction, we targeted a gene, dissimilatory sulfite reductase gene (*dsrB*), whose product is a key enzyme in sulfate reduction [25]. Specifically, we used a single-cell gene-fusion technique [26] that amplifies the 16S rRNA gene only in organisms whose genomes contain *dsrB*. The technique traps cells in polyacrylamide beads during DNA extraction, isolates the extraction products in oil droplets, and amplifies a concatenation of the *dsrB* and 16S sequences using within-droplet PCR. As a control, we performed a non-specific fusion assay to verify that a wide range of taxonomic marker sequences can be amplified with this method (Fig. 5).

The single-cell assay amplified 16S-*dsrB* amplicons whose 16S rRNA gene sequences corresponded to a small number of OTUs. Only 4 OTUs that appear in any of the OEUs were amplified by this technique, and the OEUs that contain them were identified as putative sulfate-reducing groups (Fig. 5). These results imply that the genomes of the organisms corresponding to these 4 OTUs contain *dsrB* and that the genomes of the rest of the organisms in these OEUs do not. In the first of the putative sulfate-reducing OEUs, the most abundant OTUs and two other OTUs appeared to have the genetic capacity to reduce sulfate. In the second OEU, only one OTU (about one-third as abundant as the most abundant OTU in the OEU) appeared to have this capacity. In the third OEU, no OTUs appeared to be capable of reducing sulfate.

These results raise the possibility that those two OEUs represent consortia of syntrophic organisms cooperating to carry out sulfate reduction. We therefore checked if any other OEUs contained organisms with known mutualistic associations. We identified two cases where OTUs within an OEU are likely part of a consortium. In the first case, one OEU contained an OTU that was 97.7% identical to the ammonia-oxidizing bacterium *Nitrosospira briensis* [27] and an OTU 99% identical to a nitrite-oxidizing enrichment culture clone *Candidatus* Nitrotoga arctica [28]. Nitrification is a two-step process, typically carried out by different organisms [29,30], so it is likely that these two organisms interact to carry out nitrification. In the second case, one OEU that aligns with the modeled distribution of a methane oxidation metabolism contained an OTU 94.8% identical to *Methylobacter tundripaludum* (a methane oxidizer) and an OTU 98% identical to a strain of *Methylotenera versatilis* (a non-methane-oxidizer methylotroph). A study using stable isotope-labeling concluded that these organisms cooperate during methane oxidation [31].

Discussion

Our approach combined field observations, quantitative modeling, and a single-cell genetic approach to relate taxonomic diversity in survey data to ecosystem-level functions. Our results suggest that there are previously unknown consortia in the lake whose members work together to carry out major environmental processes for at least some part of the year. Previous research has studied functions of organisms containing the same functional genes or able to incorporate the same labeled compounds. In contrast, our observational approach addressed multiple processes at the same time and did not require perturbing the environment. However, because we investigated a process, sulfate reduction, that showed a strong match between the model and observations, the results we observed should not necessarily be treated as representative of what would result from studying any of the processes.

Our analysis began by defining and studying OEUs, which are a type of ecological network. Previous studies have asserted ecological network associations between organisms in surveys based on co-occurrence patterns (i.e., mutual presence or absence) in space or time [32-35]. OEUs are the result of a different measure of "co-occurrence", since we require that OTUs co-occur at similar abundances to be placed into the same OEU. Although presence-absence patterns derived from sequencing data can be affected by technical issues like sequencing depth [36], we demonstrated that our grouping of OTUs was robust to technical issues of sample preparation method or OEU calling algorithms and provides richer interpretations of survey data.

Like other studies that investigate ecological interactions between microbial taxa, however, our results provide hypotheses about such interactions but do not prove that they exist. The inferred ecological association between members of OEUs could be verified by experimental techniques like stable-isotope probing [37], FISH-NanoSIMS [38], or MAR-Fish [39]. Because these relationships may change as the conditions in the lake change throughout the season, the identification of many pairs of OTUs in the same OEU in both 2008 and 2013 is somewhat surprising. Thus, the co-occurring pairs of organisms in the same OEU in 2008 and 2013 are strong candidates to target for further investigation.

Instead of perturbative experimental techniques, we used a biogeochemical model to predict the spatial distribution of microbial metabolisms and hypothesized the function of OEUs whose spatial distributions have the same position and shape as the ones predicted by the model. Implicit biomass models in other ecosystems could be reinterpreted as inferred biomass models, which would allow survey data to be used as validation for the spatial distribution of microbial activities predicted by models or, conversely, predicted activities to be used to generate hypotheses about the function of microbes identified in surveys. The results are limited, however, because the link between OEUs and modeled biogeochemical processes was only inferential.

Despite the model's utility, there are discrepancies between its predictions and the observed chemical and biological data. We attribute these discrepancies to multiple causes. First, our model only simulates the underlying biogeochemical processes and does not account for other ecological and physiological factors that determine organismal distributions. For example, the reference OTU for methane oxidation is the most abundant methanotroph in the oxic region, but its distribution does not match the predicted distribution of methane oxidation, suggesting that the organism performs other metabolisms or that methane oxidation is not well-described by the model. Second, our model is relatively simple and does not simulate many processes that are known to be part of lake's biogeochemistry, including complex carbon substrate utilization profiles. We aimed to create a model that captured the broad spatial patterns and temporal dynamics with the minimum number of processes possible, thus balancing the model's completeness with its simplicity. For example, the observed oxygen minimum at the thermocline is unusual but not rare for dimictic lakes. There are many competing explanations for this type of minimum (e.g., changes in temperature, predator abundance, or horizontal mixing [19]) that are all beyond the scope of the model, and so it fails to predict that minimum. Third, the model was designed to model the lake's general seasonal dynamics rather than its behavior in the specific season when we conducted the survey. For the purposes of this study, understanding the lake's general organizing processes was more important than understanding the dynamics in a particular year. Although these discrepancies limit the interpretability of hypotheses about OTUs' function that are generated by the biogeochemical model, we showed that additional bioinformatic and experimental evidence can together provide a more complete picture.

Our discovery of well-correlated organisms that probably cooperate to carry out biogeochemical functions raises exciting ecological questions. If some of the OEUs do represent consortia of syntrophic organisms, are the organisms that compose them physically associated because they inhabit similar particulate matter? What roles do microbes play within these consortia? Are these interspecific associations constant over time, or do they "reset" when the lake mixes in the winter? We expect that our combined framework of surveys, modeling, and single-cell genetics will be useful for *in situ*, non-perturbative identification of potential ecological interactions in other microbial ecosystems, painting a picture of a microbial world filled with complex, interlocking relationships.

Methods

Sample collection (2012-2013). Water samples were collected at Upper Mystic Lake (Medford, MA), from one location in the middle of the lake (~42 26.155N, 71 08. 961W) where the total water depth is 23 meters. Water samples were collected in 2012 (October 2) and 2013 (March 26, May 10, June 17, July17, and August 15). Water samples were collected at approximately one- to two-meter intervals through 25 meters of plastic Tygon tubing using a peristaltic pump. Two volumes of water at each depth were pumped through the tubing before 50 mL was filtered through an in-line Swinnex filter holders onto sterile 0.22 µm filters (Millipore, Billerica, MA) and the filtrate collected in a 50 mL conical tube. Filters and filtrate were placed on dry ice immediately and transported back to the laboratory where they were placed at −80 °C until processing. To determine the influence of contamination from the tubing, sampling method, and carryover from the previous depth, we collected blanks by pumping 2 L of sterile water through the tubing before and after sampling. Blanks were distinct from other samples. One mL of both filtered and unfiltered water was put into a 1.5 mL microcentrifuge tube with 43 µL of concentrated HCl and placed in the dark during transport back to the lab for ferrous (Fe^{2+}) and total iron analysis, respectively. Samples were stored at −20 °C until iron was measured.

Sample collection (2008). The methods for collecting from Upper Mystic Lake on August 13, 2008 are described elsewhere [20]. Water was collected from Upper Mystic Lake (same location) on August 13, 2008 using a peristaltic pump and plastic Tygon tubing. Tubing was lowered to a point ~1 m from the bottom, running the pump in reverse to prevent water from entering the tubing until the appropriate depth was reached. Water from depth was allowed to flow through the tubing for 5 minutes before 14 mL were collected into a 15 mL sterile falcon tube and immediately placed on dry ice. The first sample was taken from 22 meters depth and subsequent samples were taken every meter until 3 meters depth, then at 1.5 meters depth and at the surface. Samples were transported on dry ice and stored at −80 °C until processing about one year later.

Water conditions and chemistry. A Hydrolab minisonde (Hach Hydromet, Loveland, CO) was attached to the end of the tubing to record dissolved oxygen, temperature, and specific conductance during deployment. Nitrate, sulfate and chloride were measured by ion chromatography at the University of New Hampshire Water Quality Analysis Laboratory. Iron was measured by a modified ferrozine protocol [16,40]. Values for other chemical species used in the model but not directly measured were manually interpolated from previous measurements.

DNA extraction. DNA from half of the 2012 samples and all the 2013 samples was extracted with PowerWater DNA extraction kit (MoBio, USA). The manufacture's protocol was followed, except for the addition of proteinase K and alternative lysing protocol at 65 °C (MoBio Laboratories, Inc., Carlsbad, CA). Briefly, filters were sterilely transferred into the PowerWater Bead Tube and 20 μl proteinase K was added before incubating at 65 °C for 10 minutes. Tubes were vortexed on a horizontal MoBio vortex adapter. Proteins and inhibitors were removed with PW2 and PW3 before adding supernatant to Spin filter for column purification. After two washing steps, DNA was eluted with PW6 and used in PCR analyses. Samples from 2008 and the other half of the 2012 samples were extracted with the Qiagen DNeasy Blood and Tissue Kit (Qiagen, USA), as previously described [20,41]. Of the 2012 samples, 9 were prepared in duplicate, one replicate per extraction method.

Illumina library construct design. The Illumina library was created with a two-step protocol in order to add cluster binding and sequencing primer sites to the construct in the second round of PCR amplification. First step PCR amplification primers (PE16S_V4_U515_F, 5′–ACACG ACGCT CTTCC GATCT YRYRG TGCCA GCMGC CGCGG TAA–3′ and PE16S_V4_E786_R, 5′–CGGCA TTCCT GCTGA ACCGC TCTTC CGATC TGGAC TACHV GGGTW TCTAA T–3′) contain primers U515F and E786R targeting the V4 region of the 16S rRNA gene, as described previously [20,42]. Additionally, a complexity region in the forward primer (5′–YRYR–3′) was added to aid image-processing software used during Illumina next-generation sequencing. The second-step primers incorporate the Illumina adapter sequences and a 9-bp barcode for library recognition (PE-III-PCR-F, 5′-AATGA TACGG CGACC ACCGA GATCT ACACT CTTTC CCTAC ACGAC GCTCT TCCGA TCT–3′; PEIII-PCR-001-096, 5′–CAAGC AGAAG ACGGC ATACG AGATN NNNNN NNNCG GTCTC GGCAT TCCTG CTGAA CCGCT CTTCC GATCT–3′, where N indicates sample barcode position). Libraries from 2008 were created with the primer skipping protocol, as previously described [41].

Illumina library preparation and sequencing. Real-time PCR was done to ensure uniform amplification and avoid over-cycling. Both real-time and first-step PCRs were done similarly to the manufacture's protocol for Phusion polymerase (New England BioLabs, Ipswich, MA). Samples were cycled with the following conditions: denaturation at 98 °C for 30 sec annealing at 52 °C for 30 sec and extension at 72 °C for 30 sec for 40 cycles. Samples were divided into four 25-µl technical replicate reactions during both first- and second-step cycling reactions and cleaned using Agencourt AMPure XP-PCR purification (Beckman Coulter, Brea, CA). Paired-end sequencing was performed at Massachusetts Institute of Technology BioMicro Center

(BMC) on an Illumina MiSeq with 250 bases for each the forwards and reverse reads and 8 base indexing read. Non-standard Illumina indexing primers were used to initiate sequencing from just after the sequencing primer binding site for the barcode sequence (anti-reverse BMC index primer; 5′– AGATC GGAAG AGCGG TTCAG CAGGA ATGCC GAGAC CG –3′). To improve base-calling efficiency, 25% phiX control was added to the sample during sequencing.

Raw data processing and OTU calling. Raw sequence data was demultiplexed and quality filtered using custom scripts (github.com/almlab/SmileTrain). Overlapping paired end reads were merged and sequences were pre-clustered with USEARCH [43]. Sequences were aligned to a subset of the Silva alignment [44] with mothur [45], and OTUs were called with distribution-based clustering [20] with default parameters. Reads were trimmed to 102 bases before OTU calling, which was sufficient to capture the differences between key populations while still ensuring high quality base calls. Sequences were checked for chimeras using UCHIME [46] and sequences not aligned to the Silva reference database were discarded. Sequences from 2008 were processed as previously described [20]. Sequencing of the 2008 library was not long enough for paired end reads to overlap, so only the forward read was used. Sequences were trimmed to 76 nucleotides before OTU calling.

Community analysis. OTUs were classified with RDP [47]. The phylum level assignments referenced in various analyses throughout the paper are from RDP. The dendrogram of Jensen-Shannon divergences was produced using Ward’s method (R version 3.2.2). The phylogenetic tree of OTU sequences, aligned to the Silva database as mentioned above, was calculated using

PhyML [48] with GTR model (estimated as best model), 0 invariant sites, 6 rate categories (estimated).

Calling operational ecological units (OEUs). To call OEUs, OTUs were initially filtered by abundance in all samples. OTUs making up less than 0.25% of counts across all samples were excluded from the OEU analysis, leaving 536 OTUs. Next, counts from technical replicate samples at the same depth were pooled, and OTUs that were abundant in the no-template negative control samples (i.e., more than 10% of that OTU's reads mapped to the negatives) were excluded. The sequence counts for each OTU were converted to relative abundance (i.e., the number of counts corresponding to each OTU in a sample was divided by the sum of all counts in that sample). Every OTU was then converted to a profile (i.e., the relative abundance of each OTU in a sample was divided by the sum of that OTU's relative abundances in all samples). The square of the Euclidean distance between OTU profiles was used as the dissimilarity metric in a hierarchical clustering analysis (Ward's method). The cluster dendrogram was cut to produce 50 candidate OEUs. OEUs were trimmed for quality as follows: if an OEU had at least one OTU with a mean Pearson correlation with the other OTUs in the cluster of less than 0.75, the OTU with the lowest mean correlation was removed from the cluster, and the filtering was repeated. OEUs with fewer than two member OTUs were excluded from further analysis. Every member OTU had a mean correlation of at least 0.75 with the other OTUs in the OEU. Of the original 536 OTUs and 50 candidate OEUs, 491 OTUs (92% of initial OTUs) in 49 OEUs (98% of initial OEUs) remained after quality filtering. Scripts for OEU calling are available at github.com/almlab/oeu.

Quantifying OEU reproducibility related to OEU number parameter. The OEU-calling algorithm requires a parameter: the number of initial OEUs at which to cut the cluster dendrogram. Figure S9 shows a comparison of the results produced by this OEU-calling algorithm when the dendrogram is cut to produce different numbers of candidate OEUs. In the main analysis, 50 OEUs were used because:

1. The choice of the number of initial OEUs represents a trade-off between two types of errors. A Type I error (i.e., the incorrect assertion that two OTUs are ecologically related) occurs more often with a smaller number of OEUs, while a Type II error (i.e., the incorrect assertion that two OTUs are unrelated) occurs more often with a large number of OEUs. The analyses presented here depend on OEUs correctly identifying true ecological associations, so avoiding Type I errors is more important, so a relatively large number of OEUs is appropriate.
2. Increasing the number of initial OEUs increases the number of OTUs included in the final analysis (since fewer OTUs are excluded by the final quality-control step) and decreases the within-OEU variance, so a relatively large number of initial OEUs is appropriate.
3. Increasing the number of initial OEUs decreases the size of each OEU: it causes them to contain fewer OTUs. To keep a high enough number of OTUs per OEU to generate hypotheses to test in the single-cell assay, a number of OEUs much greater than 50 would have been inappropriate.

Quantifying OEU reproducibility across time points. To compare the OEU composition across years, OEU compositions in the 2013 data, presented above, was compared to OEU compositions for corresponding 2008 data. To call OEUs on the 2008 dataset,

- OTUs that were abundant in the sample nearest the lake bottom (those with more than 5% of their reads from 23 meters depth) were excluded,
- OTUs that were abundant in the negative samples (those with more than 10% of their reads in the two blank samples) were excluded,
- low-abundance OTUs (those with less than 0.25% of all reads) were excluded, and
- the same OEU-calling methodology presented above was used.

Quantifying OEU reproducibility across DNA extraction and sequencing methodologies. To call OEUs on the duplicate 2012 samples prepared using two DNA extraction methodologies, the same exclusion criteria as for the 2008 data were used (except that OTUs with less than 0.1% of all reads were considered low-abundance). Sequences in the 76 bp dataset were matched to the 102 bp dataset by searching for exact sequence matches of the shorter sequence within the longer dataset. If multiple 76 bp OTUs matched the same 102 bp OTU, only the most abundant OTU was kept as the corresponding OTU. Not all OTUs were represented in both datasets, so some OTUs did not have a corresponding OTU in the other dataset.

Quantifying OEU reproducibility across OEU calling methodologies. To compare the effects of different OEU-calling algorithms, OEU calling was performed as described for the main 2013 dataset except replacing the Euclidean distance (i.e., L2 norm) with the L1 distance (i.e., Bray-Curtis).

Statistical methodology quantifying OEU reproducibility. To quantify the reproducibility of the OEUs between timepoints or between sample preparation methods, we computed the numbers of pairs of OTUs such that:

- both OTUs are present in both datasets (e.g., in both 2008 and 2013 datasets), and
- both OTUs were in the same OEU in both datasets (e.g., OTUs *A* and *B* were both in OEU *X* in the 2008 dataset and both in OEU *Y* in the 2013 dataset).

We compared this number of pairs against the number of pairs satisfying the same criteria that would arise at random, specifically, if we randomly shuffled the abundances of OTUs in each sample before computing the Euclidean distance between OTUs.

Reference OTU selection. Reference OTUs were selected by matching the Illumina OTUs to Sanger clone sequences. Only exact matches between the 77 bp Illumina OTUs and Sanger clones were considered. Three Illumina OTUs matched multiple Sanger clones with nucleotide distances between clones larger than 0.1 and resulted in OTU distributions that were the product of two distinct organismal signals. These were corrected by aligning Sanger clone sequences to identify discriminating bases 5′ of the Illumina OTU sequence end point. One or two differentiating bases were identified for each of the three cases and the length of sequence required to differentiate between the two sequences was determined. Once a unique sequence was identified to discriminate the different clones in the Illumina data, a count of the discriminating sequence across libraries was generated from the raw data expressed as a percent of total reads. This replaced the previously merged OTU for populations 16, 141 and 125 (OTU IDs).

To gain functional information for the most abundant OTUs, we generated a Sanger-sequenced clone library to provide more phylogenetic information for the shorter Illumina OTUs. To make the Sanger-sequenced clone library, 16S rRNA sequences were amplified from DNA extracted from the 6 meter and 21 meters samples with Phusion polymerase (New England Biolabs, Ipswich, MA) and 27F and 1492R primers [42]. PCR products were cloned into the pCR Blunt II plasmid with the Zero Blunt TOPO PCR cloning kit (Invitrogen, Carlsbad, CA) and sequenced in at least one direction with Sanger sequencing (Genewiz, South Plainfield, NJ). Longer Sanger sequences were assigned the functional capabilities of the best BLAST hit [49] to a type strain or genome sequence. To verify expected profiles for each process in the biogeochemical model, we selected a set of nine reference OTUs involved in the modeled biogeochemical processes. These reference OTUs were among the 100 most abundant OTUs and had sequences that matched longer 16S rRNA sequences from clone libraries sequenced with Sanger sequencing developed from the lake samples.

Functions, OTU IDs, full genome matches, and accessions for reference OTUs are shown in Table S6. The matching Sanger sequences for the reference OTUs corresponded to organisms with metabolisms characterized by genomic analysis or *in vitro* experiments. Reference OTUs had spatial distributions in the lake that were consistent with their purported metabolism.

Biogeochemical model. The biogeochemical model (almlab.mit.edu/mystic.html), inspired by Hunter *et al.* [22], was run with Matlab (version 8) and supporting Python scripts (version 2.7). Details on the mechanics, implementation, and parameter values are in the Supplementary

Information. Briefly, the water under the thermocline is modeled as 17 linked compartments, one per meter depth. Within each compartment, a minimal set of abstracted chemical species interconvert through a minimal set of modeled primary and secondary redox reactions (Fig. S5, Tables S1-S3). Primary oxidation rates follow a formulation informed by the relative favorability of electron acceptors. Secondary oxidation rates follow simple mass action rate forms. Chemical species are transported between adjacent compartments via bulk diffusion (for all species) and settling (for biomass and oxidized iron). The outside world is modeled by constant source terms: oxygen and biomass are added in the uppermost compartment (at the thermocline), while methane is added in the lowermost compartment (at the sediment). The resulting set of ordinary differential equations is solved numerically.

We intended to model the general distribution of chemical and biological species in the lake. Because the model is conceptual, it includes many simplifications compared to the aquifer model. First, transport is modeled compartment-by-compartment, using ordinary differential equations rather than partial differential equations. We greatly reduced the number of simulated chemical species (from 25 to 9). Many simulated chemical species consist of multiple chemical species found in nature (e.g., the modeled oxidized sulfur species includes hydrogen sulfide H_2S, bisulfide HS^-, and sulfide S^{2-}; there is only one modeled carbon species). Other chemical species found in nature are not treated in the model (e.g., elemental sulfur S^0 and all manganese compounds) because less is known about their importance to the lake's biogeochemistry. The primary redox reactions are borrowed almost exactly from the aquifer model (excepting some parameter changes and the removal of manganese as an electron acceptor). The secondary redox reactions are similar to those in the aquifer model (excepting some parameter changes, the removal of some reactions, and the addition of iron oxidation on nitrate). Precipitation-dissolution, acid dissolution, and adsorption reactions relevant in the groundwater system were part of the original aquifer model but were not simulated here. Further details about these alterations of the original model are included in the Supplementary Information.

Comparing OEUs and biogeochemical processes. We asserted that certain OEUs are related to certain modeled biogeochemical processes by comparing the spatial distributions of OEUs and modeled processes and by manual bioinformatic inference. Average Euclidean distance to each process for all OTUs within an OEU was calculated (Fig. S8). OEUs containing the reference OTUs were chosen to represent each modeled process because existing literature about the reference OTU suggests that those OTUs perform that modeled process. To assign an OEU to a process, we further required that the imputed process be one of the two processes least distant from that OEU, as described above, except for methane oxidation on oxygen, which is not within the lowest two distances for that OEU (see Discussion).

Linking taxonomic marker sequences with a functional gene. Data for 16S rRNA gene fusion products with both selective (*dsrB*) and non-selective (barcode) sequences experiments was obtained from a previous analysis [26]. Briefly, seven mL of water from both the 2 meter and 21 meter samples on August 12, 2013 was added to 7 mL of 50% glycerol (25% final concentration) to preserve membrane integrity for single-cell techniques, immediately placed on dry ice and stored at −80 °C. 16S rRNA gene sequences were fused to a 20 base pair droplet barcode to control for effects of the protocol on limiting diversity. In a separate reaction, 16S rRNA gene sequences were fused to a portion of the diagnostic gene for dissimilatory sulfite reduction (*dsrB*) to probe for functional information. Cells are trapped in 10 μm diameter polyacrylamide beads [50]. Poisson statistics predict that only 0.45% of beads will contain more than one cell. The DNA trapped within the beads is used as the template for PCR inside an emulsion [51]. The first set of primers for 16S rRNA gene amplification include U515F and 1492R, and the fusion reaction is nested within the 16S gene using E786R. The *dsrB* gene primers were adapted from Wagner *et al.* [52] and slightly modified to fit the needs of the molecular construct. Sequences and the results of a traditional *dsrB* survey are provided in the original publication. The *dsrB* gene is highly conserved across known sulfate reducers [53], but it is possible that there are variants of the gene that are prevalent in the lake that these primers did not amplify, in which case the following analysis would contain false negatives (i.e., OTUs that can reduce sulfate but did not produce 16S-*dsrB* amplicons). Comparison of the *dsrB*-16S rRNA gene fusion assay to a bulk *dsrB* gene survey in the original previous analysis demonstrate significant overlap [26], showing the fusion PCR assay targets a wide variety of reductive *dsrB* genes from the *δ-Proteobacteria dsrB* supercluster.

Nucleotide sequence accession numbers. All clone sequences were submitted to GenBank (accession no. KC192376 to KC192544). Illumina data were submitted to the Sequence Read Archive under study accession number PRJNA217938.

Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1122374 and by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0008743.

References

1 Pace, N. R. A Molecular View of Microbial Diversity and the Biosphere. *Science* **276**, 734-740, doi:10.1126/science.276.5313.734 (1997).

2 Wilmes, P., Simmons, S. L., Denef, V. J. & Banfield, J. F. The dynamic genetic repertoire of microbial communities. *FEMS Microbiol Rev* **33**, 109-132, doi:10.1111/j.1574-6976.2008.00144.x (2009).

3 Rundell, E. A. *et al.* 16S rRNA Gene Survey of Microbial Communities in Winogradsky Columns. *PLoS ONE* **9**, e104134, doi:10.1371/journal.pone.0104134 (2014).

4 Ward, D. M., Ferris, M. J., Nold, S. C. & Bateson, M. M. A Natural View of Microbial Biodiversity within Hot Spring Cyanobacterial Mat Communities. *Microbiol Mol Biol Rev* **62**, 1353-1370 (1998).

5 Bier, R. L., Voss, K. A. & Bernhardt, E. S. Bacterial community responses to a gradient of alkaline mountaintop mine drainage in Central Appalachian streams. *ISME J* **9**, 1378-1390, doi:10.1038/ismej.2014.222 (2015).

6 Schrenk, M. O., Kelley, D. S., Delaney, J. R. & Baross, J. A. Incidence and diversity of microorganisms within the walls of an active deep-sea sulfide chimney. *Appl Environ Microbiol* **69**, 3580-3592, doi:10.1128/aem.69.6.3580-3592.2003 (2003).

7 Pinel-Alloul, B. & Ghadouani, A. in *The Spatial Distribution of Microbes in the Environment* (eds Rima B. Franklin & Aaron L. Mills) 203-310 (Springer Netherlands, 2007).

8 Neufeld, J. D., Wagner, M. & Murrell, J. C. Who eats what, where and when? Isotope-labelling experiments are coming of age. *ISME J* **1**, 103-110, doi:10.1038/ismej.2007.30 (2007).

9 Albertsen, M. *et al.* Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* **31**, 533-538, doi:10.1038/nbt.2579 (2013).

10 Martiny, A. C., Treseder, K. & Pusch, G. Phylogenetic conservatism of functional traits in microorganisms. *ISME J* **7**, 830-838, doi:10.1038/ismej.2012.160 (2013).

11 Klein, M. *et al.* Multiple lateral transfers of dissimilatory sulfite reductase genes between major lineages of sulfate-reducing prokaryotes. *J Bacteriol* **183**, 6028-6035, doi:10.1128/Jb.183.20.6028-6035.2001 (2001).

12 Orphan, V. J. *et al.* Comparative analysis of methane-oxidizing archaea and sulfate-reducing bacteria in anoxic marine sediments. *Appl Environ Microbiol* **67**, 1922-1934, doi:10.1128/aem.67.4.1922-1934.2001 (2001).

13 Madigan, M. T., Martinko, J. M., Dunlap, P. V. & Clark, D. P. *Brock Biology of Microorganisms*. 12th edn, (Pearson/Benjamin Cummings, 2009).

14 Varadharajan, C. *Magnitude and spatio-temporal variability of methane emissions from a eutrophic freshwater lake* PhD thesis, Massachusetts Institute of Technology, (2009).

15 Varadharajan, C. & Hemond, H. F. Time-series analysis of high-resolution ebullition fluxes from a stratified, freshwater lake. *J Geophys Res Biogeosci* **117**, doi:10.1029/2011jg001866 (2012).

16 Senn, D. B. *Coupled arsenic, iron, and nitrogen cycling in arsenic-contaminated Upper Mystic Lake* PhD thesis, Massachusetts Institute of Technology, (2001).

17 Senn, D. B. & Hemond, H. F. Nitrate controls on iron and arsenic in an urban lake. *Science* **296**, 2373-2376, doi:10.1126/science.1072402 (2002).

18 Peterson, E. J. R. *Carbon and electron flow via methanogenesis, SO42-, NO3- and Fe3+ reduction in the anoxic hypolimnia of Upper Mystic Lake* Masters thesis, Massachusetts Institute of Technology, (2005).

19 Wetzel, R. G. *Limnology*. 3rd edn, (Academic Press, 2001).

20 Preheim, S. P., Perrotta, A. R., Martin-Platero, A. M., Gupta, A. & Alm, E. J. Distribution-Based Clustering: Using Ecology To Refine the Operational Taxonomic Unit. *Appl Environ Microbiol* **79**, 6593-6603, doi:10.1128/aem.00342-13 (2013).

21 Jax, K. Ecological Units: Definitions and Application. *Q Rev Biol* **81**, 237-258, doi:10.1086/506237 (2006).

22 Hunter, K. S., Wang, Y. F. & Van Cappellen, P. Kinetic modeling of microbially-driven redox chemistry of subsurface environments: coupling transport, microbial metabolism and geochemistry. *J Hydrol* **209**, 53-80, doi:10.1016/S0022-1694(98)00157-7 (1998).

23 Thomas, F., Hehemann, J.-H., Rebuffet, E., Czjzek, M. & Michel, G. Environmental and gut Bacteroidetes: the food connection. *Front Microbiol* **2**, doi:10.3389/fmicb.2011.00093 (2011).

24 Holmes, D. E., Nevin, K. P., Woodard, T. L., Peacock, A. D. & Lovley, D. R. Prolixibacter bellariivorans gen. nov., sp. nov., a sugar-fermenting, psychrotolerant anaerobe of the phylum Bacteroidetes, isolated from a marine-sediment fuel cell. *Int J Syst Evol Microbiol* **57**, 701-707, doi:10.1099/ijs.0.64296-0 (2007).

25 Leloup, J., Quillet, L., Oger, C., Boust, D. & Petit, F. Molecular quantification of sulfate-reducing microorganisms (carrying dsrAB genes) by competitive PCR in estuarine sediments. *FEMS Microbiol Ecol* **47**, 207-214, doi:10.1016/S0168-6496(03)00262-9 (2004).

26 Spencer, S. J. *et al.* Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *ISME J* **10**, 427-436, doi:10.1038/ismej.2015.124 (2016).

27 Purkhold, U., Wagner, M., Timmermann, G., Pommerening-Roser, A. & Koops, H. P. 16S rRNA and amoA-based phylogeny of 12 novel betaproteobacterial ammonia-oxidizing isolates: extension of the dataset and proposal of a new lineage within the nitrosomonads. *Int J Syst Evol Microbiol* **53**, 1485-1494, doi:10.1099/ijs.0.02638-0 (2003).

28 Alawi, M., Lipski, A., Sanders, T., Pfeiffer, E.-M. & Spieck, E. Cultivation of a novel cold-adapted nitrite oxidizing betaproteobacterium from the Siberian Arctic. *ISME J* **1**, 256-264, doi:10.1038/ismej.2007.34 (2007).

29 Canfield, D. E., Kristensen, E. & Thamdrup, B. in *Advances in Marine Biology* Vol. 48 (eds D. E. Canfield, E. Kristensen, & B. Thamdrup) 205-267 (Academic Press, 2005).

30 Costa, E., Perez, J. & Kreft, J. U. Why is metabolic labour divided in nitrification? *Trends Microbiol* **14**, 213-219, doi:10.1016/J.Tim.2006.03.006 (2006).

31 Beck, D. A. C. *et al.* A metagenomic insight into freshwater methane-utilizing communities and evidence for cooperation between the *Methylococcaceae* and the *Methylophilaceae*. *PeerJ* **1**, e23, doi:10.7717/peerj.23 (2013).

32 Faust, K. & Raes, J. Microbial interactions: from networks to models. *Nat Rev Micro* **10**, 538-550, doi:10.1038/nrmicro2832 (2012).

33 Eiler, A., Heinrich, F. & Bertilsson, S. Coherent dynamics and association networks among lake bacterioplankton taxa. *ISME J* **6**, 330-342, doi:10.1038/ismej.2011.113 (2012).

34 Gies, E. A., Konwar, K. M., Beatty, J. T. & Hallam, S. J. Illuminating Microbial Dark Matter in Meromictic Sakinaw Lake. *Appl Environ Microbiol* **80**, 6807-6818, doi:10.1128/aem.01774-14 (2014).

35 Koenig, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci USA* **108**, 4578-4585, doi:10.1073/pnas.1000081107 (2011).

36 Horner-Devine, M. C. *et al.* A comparison of taxon co-occurrence patterns for macro- and microorganisms. *Ecology* **88**, 1345-1353, doi:10.1890/06-0286 (2007).

37 Radajewski, S., Ineson, P., Parekh, N. R. & Murrell, J. C. Stable-isotope probing as a tool in microbial ecology. *Nature* **403**, 646-649, doi:10.1038/35001054 (2000).

38 Dekas, A. E. & Orphan, V. J. Identification of diazotrophic microorganisms in marine sediment via fluorescence in situ hybridization coupled to nanoscale secondary ion mass spectrometry (FISH-NanoSIMS). *Methods Enzymol* **486**, 281-305, doi:10.1016/S0076-6879(11)86012-X (2011).

39 Nielsen, J. L. & Nielsen, P. H. in *Handbook of Hydrocarbon and Lipid Microbiology* (ed Kenneth N. Timmis) (Springer-Verlag, Berlin, 2010).

40 Stookey, L. L. Ferrozine–a new spectrophotometric reagent for iron. *Anal Chem* **42**, 779-781, doi:10.1021/ac60289a016 (1970).

41 Blackburn, M. C. *Development of new tools and applications for high-throughput sequencing of microbiomes in environmental or clinical samples* Masters thesis, Massachusetts Institute of Technology, (2010).

42 Lane, D. J. in *Nucleic acid techniques in bacterial systematics* (eds E. Stackebrandt & M Goodfellow) 125-175 (John Wiley & Sons, 1991).

43 Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Meth* **10**, 996-998, doi:10.1038/nmeth.2604 (2013).

44 Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**, D590-D596, doi:10.1093/nar/gks1219 (2013).

45 Schloss, P. D. *et al.* Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl Environ Microbiol* **75**, 7537-7541, doi:10.1128/aem.01541-09 (2009).

46 Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194-21200, doi: 10.1093/bioinformatics/btr381 (2011).

47 Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**, 5261-5267, doi:10.1128/aem.00062-07 (2007).

48 Guindon, S. *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Biol* **59**, 307-321, doi:10.1093/sysbio/syq010 (2010).

49 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2 (1990).

50 Tamminen, M. V. & Virta, M. P. J. Single gene-based distinction of individual microbial genomes from a mixed population of microbial cells. *Front Microbiol* **6**, doi:10.3389/fmicb.2015.00195 (2015).

51 Turchaninova, M. A. *et al.* Pairing of T-cell receptor chains via emulsion PCR. *Eur J Immunol* **43**, 2507-2515, doi:10.1002/Eji.201343453 (2013).

52 Wagner, M. *et al.* Functional marker genes for identification of sulfate-reducing prokaryotes. *Method Enzymol* **397**, 469-489, doi:10.1016/S0076-6879(05)97029-8 (2005).

53 Wagner, M., Roger, A. J., Flax, J. L., Brusseau, G. A. & Stahl, D. A. Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration. *J Bacteriol* **180**, 2975-2982 (1998).