The ISME Journal (2015) 9, 857–870

www.nature.com/ismej

ORIGINAL ARTICLE

Genomic insights into the uncultivated marine *Zetaproteobacteria* at Loihi Seamount

Erin K Field[1], Alexander Sczyrba[2], Audrey E Lyman[3], Christopher C Harris[1], Tanja Woyke[4], Ramunas Stepanauskas[1] and David Emerson[1]

[1]*Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, USA;* [2]*Center for Biotechnology, Bielefeld University, Bielefeld, Germany;* [3]*Department of Biology, Colby College, Waterville, ME, USA and* [4]*US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA, USA*

The *Zetaproteobacteria* are a candidate class of marine iron-oxidizing bacteria that are typically found in high iron environments such as hydrothermal vent sites. As much remains unknown about these organisms due to difficulties in cultivation, single-cell genomics was used to learn more about this elusive group at Loihi Seamount. Comparative genomics of 23 phylogenetically diverse single amplified genomes (SAGs) and two isolates indicate niche specialization among the *Zetaproteobacteria* may be largely due to oxygen tolerance and nitrogen transformation capabilities. Only Form II *ribulose 1,5-bisphosphate carboxylase* (*RubisCO*) genes were found in the SAGs, suggesting that some of the uncultivated *Zetaproteobacteria* may be adapted to low oxygen and/or high carbon dioxide concentrations. There is also genomic evidence of oxygen-tolerant cytochrome *c* oxidases and oxidative stress-related genes, indicating that others may be exposed to higher oxygen conditions. The *Zetaproteobacteria* also have the genomic potential for acquiring nitrogen from numerous sources including ammonium, nitrate, organic compounds, and nitrogen gas. Two types of molybdopterin oxidoreductase genes were found in the SAGs, indicating that those found in the isolates, thought to be involved in iron oxidation, are not consistent among all the *Zetaproteobacteria*. However, a novel cluster of redox-related genes was found to be conserved in 10 SAGs as well as in the isolates warranting further investigation. These results were used to isolate a novel iron-oxidizing *Zetaproteobacteria*. Physiological studies and genomic analysis of this isolate were able to support many of the findings from SAG analyses demonstrating the value of these data for designing future enrichment strategies.
The ISME Journal (2015) **9,** 857–870; doi:10.1038/ismej.2014.183; published online 10 October 2014

Introduction

Iron is the fourth most abundant element in the Earth's crust and reduced Fe(II) can be oxidized to insoluble Fe(III) oxyhydroxides through abiotic or biotic means. The *Zetaproteobacteria*, a candidate class of the *Proteobacteria* phylum, carry out Fe(II) oxidation in marine environments at low temperatures, circumneutral pH and under microaerophilic conditions. To date, the *Zetaproteobacteria* have only been identified in habitats with high ferrous iron concentrations. Within this context, they have been detected in a wide range of environments from hydrothermal vent sites, such as Loihi Seamount, where they can dominate the microbial community (Rassa *et al.*, 2009; Emerson and Moyer, 2010; McAllister *et al.*, 2011; Fleming *et al.*, 2013), to coastal environments, where they may be involved in biocorrosion of steel structures such as ships and pilings (Dang *et al.*, 2011; McBeth *et al.*, 2011).

The *Zetaproteobacteria* create distinct extracellular structures, such as helical stalks (Singer *et al.*, 2011) and tubular sheaths (Fleming *et al.*, 2013), that are composed primarily of iron oxyhydroxides and organic polymers (Chan *et al.*, 2011). These structures are a by-product of their metabolism and are capable of creating a structured habitat, both physically and chemically, within the environment. At Loihi Seamount, the focus of this study, these iron mats can be centimeters thick and are greatly affected by opposing gradients of oxygen (O_2) and Fe(II) that can result in sharp redoxclines (Glazer and Rouxel, 2009). However, there may also be gradients of other chemical constituents found in the environment that could influence the lifestyle of these organisms such as carbon dioxide (CO_2), ammonium and hydrogen sulfide (Karl *et al.*, 1988; Garcia *et al.*, 2006; Glazer and Rouxel, 2009). The only current evidence of different populations of *Zetaproteobacteria* has been from molecular-based studies evaluating the small subunit ribosomal gene, which have suggested that there is a significant

Correspondence: D Emerson, Bigelow Laboratory for Ocean Sciences, 60 Bigelow Drive, East Boothbay, ME 04544, USA.
E-mail: demerson@bigelow.org
Received 3 February 2014; revised 18 August 2014; accepted 25 August 2014; published online 10 October 2014

amount of phylogenetic diversity within the *Zetaproteobacteria* both globally and at Loihi Seamount (McAllister *et al.*, 2011). Currently available *Zetaproteobacteria* laboratory isolates represent only a small subset of this diversity as the 16S rRNA gene of these isolates are 99% similar to each other. Unfortunately, the *Zetaproteobacteria* remain difficult to grow in the lab, thus cultivation-independent methods, such as single-cell genomics, are currently the best approach to learn more about this diverse yet elusive group.

What is known about the *Zetaproteobacteria* in regards to their metabolic capabilities and role in the environment has come largely from studying cultures in the laboratory. The cultivated representatives are all closely related to *Mariprofundus ferrooxydans* strain PV-1, an obligate iron-oxidizing, chemolithoautotroph. Of these, only the closely related strains PV-1 and *M. ferrooxydans* M34 have had their genomes sequenced. Although the laboratory isolates have shed light on some of their metabolic capabilities, the question remains whether the broader phylogenetic diversity within this class is reflective of its metabolic diversity and potential niche specialization. To address this question, genomic data from 23 phylogenetically diverse, uncultivated *Zetaproteobacteria* genomes from Loihi Seamount were obtained. Comparative genomics was conducted to assess their metabolic capabilities that may lead to niche specialization and compare them with the cultivated representatives, both of which were isolated from Loihi Seamount. This information can provide a better understanding of the ecologic and metabolic role the *Zetaproteobacteria* may have in this environment.

Materials and methods

Site description and sample collection

Samples were collected from Loihi Seamount using a custom-built, manually operated syringe-based biomat sampler as described by Fleming *et al.* (2013). The sampler allowed for accurate collection of the surface layer (top cm) of the microbial mat, which can be many centimeters thick. Surface layer iron mat samples from three distinct sites at Loihi Seamount were collected and used for single-cell sorting from Hiolo North Marker 31 (AB-133, AB-137), Hiolo North Marker 39 (AB-602) and Pohaku Marker 57 (AB-604). Two additional bulk mat samples from Hiolo North Marker 39 (AC-673) and Pohaku Marker 57 (AC-675) were collected with a suction sampler collecting approximately the top 3–4 cm of the mat and were used for single-cell sorting. Upon return to the ship, all samples were immediately preserved in glycerol and stored at $-80\,^{\circ}C$ until single-cell sorting. Microscopy of corresponding samples not prepared for single-cell sorting revealed biologically produced Fe(III) oxyhydroxides (i.e. stalks and sheaths) indicative of biologic iron oxidation by *Zetaproteobacteria* in these samples. Additional site and sample details can be found in Supplementary Table S1 and Fleming *et al.* (2013).

Single-cell sorting and amplification

The semiconsolidated mat samples were pretreated before single-cell sorting using fluorescence-activated cell sorting. Each sample was aspirated with a syringe and needle to disrupt the sample and then diluted 1:10 in filter-sterilized artificial seawater before analysis. Single-cell sorting and amplification has been described previously (Stepanauskas and Sieracki, 2007; Swan *et al.*, 2011, 2013; Woyke *et al.*, 2011). Additional details can be found in Supplementary Information.

Whole genome sequencing, assembly, and quality control

Based on their identity and phylogenetic diversity, 23 unique *Zetaproteobacteria* SAGs (of the total 214 *Zetaproteobacteria* SAGs identified) were chosen for whole genome sequencing. Eight SAGs were sequenced and assembled through the Bigelow Laboratory Single Cell Genomics Center pipeline and 14 SAGs were sequenced and assembled at the Joint Genome Institute (Walnut Creek, CA, USA) (Supplementary Table S2 and Supplementary Methods). One additional SAG (AB-137-C09) was sequenced twice, once at each location, to confirm sequencing and assembly methods between locations had no effect on the resulting genomes. The sequencing at Joint Genome Institute was part of a Community Sequencing Project and followed the Joint Genome Institute's single-cell minimal draft protocol. Sequencing and assembly through the Single Cell Genomics Center pipeline followed the protocols previously described in Wilkins *et al.* (2014). Detailed methods for sequencing, assembly and thorough quality control, including contaminant identification and removal, can be found in Supplementary Information. Statistics of the final assemblies can be found in Supplementary Table S2.

Genome completeness estimates

Genome completeness estimates were determined by calculating the number of conserved copy genes recovered in each genome (Woyke *et al.*, 2009; Garcia *et al.*, 2012; Kamke *et al.*, 2013; Rinke *et al.*, 2013). A total of 76 conserved single copy genes, identified based on clusters of orthologous group (COG) classifications, were found in 97% of all finished *Proteobacteria* genomes ($n = 938$) in the Integrated Microbial Genome database at the time of analysis (21 July 2013). Genome completeness estimates for each genome were calculated based on the percent of these 76 conserved single copy genes recovered in each individual genome.

Sequence classification and OTU designation
Sequences obtained from single-cell PCR screening of multiple displacement amplification products were classified using the Silva Aligner (Pruesse *et al.*, 2007) and imported into the ARB database (Ludwig *et al.*, 2004). *Zetaproteobacteria* sequences were passed through MOTHUR (Schloss and Handelsman, 2005) and clustered using a 0.03 cutoff. An operational taxonomic unit (OTU) was defined as a group of sequences with 97% sequence similarity to each other. These sequences were compared with reference sequences from McAllister *et al.* (2011) to determine which global OTU these sequences correlated with.

Whole genome comparisons
Amino-acid identities (AAIs) were calculated for all pairs of SAGs as follows: For each pair of SAGs, reciprocal best BLAST hits (E-value $<1E-3$) were used to predict orthologous genes between the two genomes using the amino-acid sequences of all predicted genes. The AAI was computed as the mean of all BLAST alignments of orthologous genes. Rows and columns in the heatmap were hierarchically clustered, resulting in the dendrogram shown. Average nucleotide identities (ANIs) were calculated using the BLAST-based algorithm for calculating average nucleotide identity (ANIb) tool in JSpecies (v.1.2.1) (Richter and Rosselló-Móra, 2009). All nucleotide ambiguities in the form of N's were removed before analysis. SAG AB-602-E04 was removed from all analyses except the heatmap because of extremely low genome recovery leading to a clear bias in the results.

Phylogenetic analyses
A maximum-likelihood phylogenetic tree was created with the 16S rRNA gene sequences obtained from assembled SAGs. In SAGs where the 16S rRNA gene was not recovered, the partially recovered 16S rRNA gene from PCR screening was used. Nucleotide sequences were aligned using the MAFFT aligner (Katoh *et al.*, 2002), and the phylogenetic tree was created in RAxML-MPI (v.2.2.3) (Stamatakis, 2006) using the GTR GAMMA nucleotide substitution model with 100 bootstrap iterations. Molybdopterin oxidoreductase and ribulose 1,5-bisphosphate carboxylase protein sequences were aligned using ClustalW (Thompson *et al.*, 1994) and further aligned manually. Maximum-likelihood phylogenetic trees were created in MEGA5 (Kumar *et al.*, 2008) using the Poisson Substitution Model with 1000 bootstrap iterations.

Isolation of Mariprofundus sp. EKF-M39
An iron mat sample was collected during a recent expedition to Loihi in March 2013 from the Marker 39 iron mat (sample: J2-674-BM1-B456). For the initial enrichment, 100 μl of mat sample was added to Balch Tubes containing sterile artificial seawater (ASW) (Emerson and Floyd, 2005) prepared under anaerobic conditions with a headspace of argon:CO_2 (80:20) mix. Artificial seawater was prepared using equivalent moles of nitrogen in the form of sodium nitrate to replace ammonium chloride, the usual N-source for artificial seawater. ATCC Trace Mineral Supplement (ATCC, Manassas, VA, USA), ATCC Vitamin Supplement (ATCC) and 500 μM $FeCl_2$ (final concentration) were also added. Trace O_2 was likely present from addition of the vitamins and mat inoculum. Cultures were transferred under the same growth conditions five times to obtain a pure culture, which was confirmed by microscopy and 16S rRNA gene sequencing.

Accession numbers
The small subunit genes obtained from multiple displacement amplification products have been submitted to GenBank under accession numbers (KM042429–KM042846). Isolate genome scaffolds can be found in GenBank under *Mariprofundus ferrooxydans* PV-1 (AATS00000000.1) and *M. ferrooxydans* M34 (NZ_ARAU00000000.1), as well as in Integrated Microbial Genome database under Taxon IDs 639857004 and 2513237158, respectively. *Mariprofundus* sp. EKF-M39 and all single amplified genomes (SAGs) have been made public in the Integrated Microbial Genome database. Strain EKF-M39 is provided under Taxon ID 2571042360, and SAG Taxon IDs are reported in Table 1.

Results

Zetaproteobacteria community structure
Microbial community composition determined by 16S rRNA gene sequencing of SAGs was in agreement with previous studies of the Loihi Seamount mats (McAllister *et al.*, 2011; Fleming *et al.*, 2013). The *Zetaproteobacteria* dominated all SAG bacterial 16S rRNA gene libraries (52–71% total SAGs) except for the bulk mat sample collected from Marker 57 (Supplementary Figures S1 and S2 and Supplementary Table S3). A total of 214 SAGs were identified as *Zetaproteobacteria*. These SAGs represented 5 of the 12 *Zetaproteobacteria* OTUs at Loihi Seamount previously reported by McAllister *et al.* (2011) including the globally ubiquitous OTUs 1 and 2 (Figure 1 and Supplementary Table S4). Interestingly, no 16S sequences from OTU 11 were recovered, indicating that none of the SAGs were closely related to strains *M. ferrooxydans* PV-1 and M34. Marker 57 samples were largely comprised of OTUs 4 and 10 members, whereas Marker 31 and Marker 39 samples were dominated by OTU 1 and 2 members, which is consistent with geographical

Table 1 *Zetaproteobacteria* SAG general statistics

SAG ID	*Collection site*	*OTU*[a] *designation*	*IMG Taxon ID*	*Size (Mb)*	*Estimated genome recovery (%)*	*GC content (%)*	*No. of predicted genes*	*No. of protein-coding genes*	*No. of RNAs*	*No. of tRNAs*	*No. of protein-coding genes w/functional prediction*
AB-137-J06	Marker 31	1	2 264 867 010	0.89	47.4	46.4	1035	1021	14	8	776
AB-602-C20	Marker 39	1	2 264 867 008	0.99	21.1	46.1	1091	1072	19	12	795
AB-133_D10	Marker 31	1	2 524 614 788	1.57	54.0	47.5	1631	1605	26	22	1267
AB-137_M18	Marker 31	1	2 524 614 790	0.45	<10	49.4	497	487	10	6	346
AB-602_L11	Marker 39	1	2 524 614 791	0.33	<10	48.5	353	345	8	3	272
AB-133-C04	Marker 31	2	2 264 867 014	1.30	73.7	43.6	1542	1514	28	17	1146
AB-137-I08	Marker 31	2	2 265 123 003	2.05	81.6	43.2	2392	2340	52	39	1790
AB-602_F03	Marker 39	2	2 524 614 781	0.99	31.6	42.9	1085	1068	17	15	838
AB-604_P22	Marker 57	2	2 524 614 792	0.61	49.1	43.2	725	716	9	6	557
AB-133_M17	Marker 31	2	2 524 614 793	1.49	77.6	42.9	1621	1589	32	25	1293
AB-137_G16	Marker 31	2	2 524 614 794	0.89	59.2	42.9	1016	995	21	15	768
AB-604-O16	Marker 57	4	2 264 867 012	0.98	46.1	50.5	1167	1141	26	19	896
AB-604_O11	Marker 57	4	2 524 614 795	0.67	26.3	48.9	764	743	21	10	502
AC-673-C02	Marker 39	4	2 528 768 162	0.88	56.6	50.3	983	971	12	6	751
AC-675-C07	Marker 57	4	2 528 768 166	0.62	15.8	49.6	724	691	33	19	435
AB-137-C09	Marker 31	6	2 264 867 009	2.45	82.9	47.6	2482	2432	50	37	1824
AB-137_C09B	Marker 31	6	2 524 614 796	2.35	82.9	48.1	2301	2239	62	44	1674
AB-602-E04	Marker 39	6	2 264 867 015	0.45	15.8	46.1	525	517	8	3	371
AB-133-G06	Marker 31	6	2 264 867 013	1.10	56.6	46.7	1247	1222	25	13	925
AB-137-L23	Marker 31	6	2 528 768 164	1.70	73.7	47.6	1751	1710	41	33	1259
AB-604-B04	Marker 57	10	2 264 867 011	0.95	39.5	50.2	1095	1080	15	11	812
AC-673-M07	Marker 39	10	2 524 614 605	1.48	68.4	50.9	1635	1610	25	20	1232
AC-673-N02	Marker 39	10	2 528 768 165	0.95	10.5	51.3	1094	1056	38	22	775
AC-673-B17	Marker 39	10	2 528 768 224	1.33	17.1	51.7	1458	1428	30	19	1076

Abbreviations: GC, Guanine Cytosine; IMG, Integrated Microbial Genome; OTU, operational taxonomic unit; SAG, single amplified genome.
[a]Comparable to OTU designations in McAllister *et al.* (2011).

differences observed in previous studies at Loihi Seamount (McAllister *et al.*, 2011, Singer *et al.*, 2013). Markers 31 and 39 are geographically close to each other located only a few meters apart at a water depth of 1300 m near the bottom of Pele's Pit, a crater created as a result of an eruption in 1996. At the time of sampling in 2009, the vent fluid temperature was ~42 °C, while the ambient seawater was around 4 °C. Conversely, Marker 57 is more distant at the Pohaku outcrop (400 m south of Pele's Pit), shallower (water depth of 1178 m), has cooler vent fluid temperatures (~28 °C), and a slightly higher pH (5.6–7.3 compared to 5.8–6.5) (Glazer and Rouxel, 2009). These differences in geographical location, water depth, temperature, and could all have a role in the *Zetaproteobacteria* community structures observed. The entire summit of Loihi Seamount is considered an O_2 minimum zone with seawater O_2 concentrations ranging from 50 to 80 μM while the vent fluid Fe(II) concentrations are typically several 100 μM (Glazer and Rouxel, 2009), indicating that O_2 and Fe(II) are likely not the sole contributors to the differences in community structure as they do not vary significantly between sites. At Marker 39 there were also differences between the surface layer and suction samples of bulk mat, suggesting that OTUs 2 and 6 may be more prevalent in the first centimeter of the mat, while OTUs 1, 4 and 10 may be deeper in the mat. These surface layer samples were collected in consecutive years, thus this is solely an observation, but it suggests that there may be some structuring of these organisms with depth in the mat.

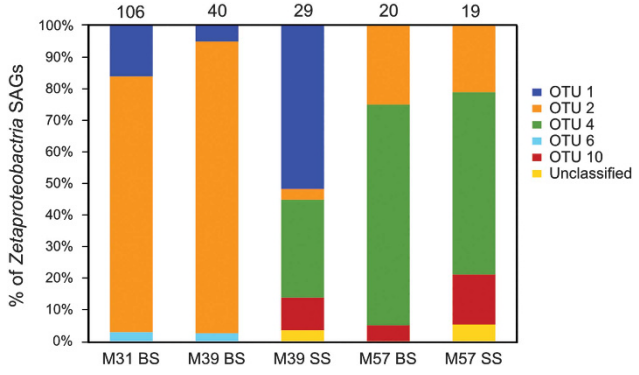

Figure 1 OTU classifications of all Loihi Seamount *Zetaproteobacteria* 16S rRNA gene sequences identified in each SAG library. Numbers at the top of each bar represent the total number of *Zetaproteobacteria* SAGs found in that sample and each bar represents a different environmental sample. BS, biomat sample; M31: Marker 31; M39: Marker 39; M57: Marker 57; SS, suction sample.

Twenty-three SAGs were chosen for sequencing from these three distinct sites within the Loihi Seamount area representing phylogenetically diverse members of the *Zetaproteobacteria*: OTU 1 (five SAGs), OTU 2 (six SAGs), OTU 4 (four SAGs), OTU 6 (four SAGs) and OTU 10 (four SAGs) (Table 1 and Figure 2) to capture as much diversity as possible. Genome recovery estimates ranged between <10% and 83% with an average of 49%.

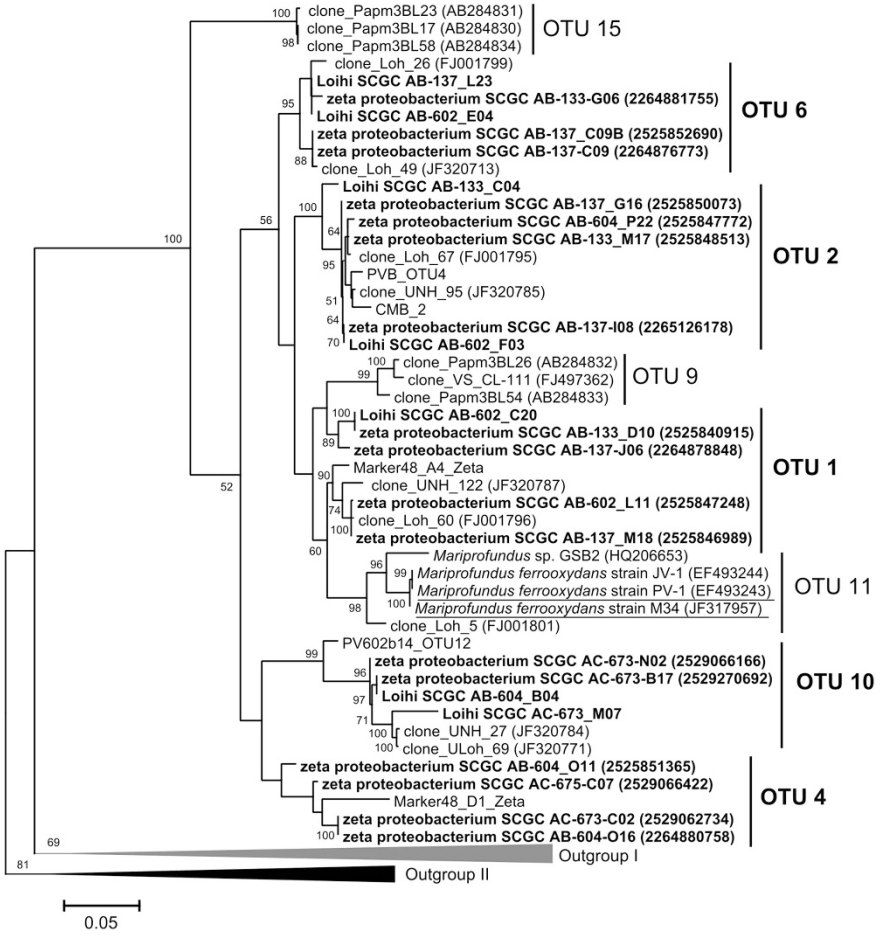

Figure 2 Maximum-likelihood phylogenetic tree of the *Zetaproteobacteria* SAG 16S rRNA genes. OTUs with representative SAGs are in bold. Sequenced isolates are underlined. Additional *Zetaproteobacteria* reference sequences were included. The tree was rooted with additional outgroups including other *Proteobacteria* classes (Outgroup I) and other phyla members (Outgroup II). The tree represents 100 iterations and bootstrap values below 50 are not reported.

General comparisons between Zetaproteobacteria genomes

Genomes were compared at both the nucleotide and protein level to determine if phylogenetic similarity correlated with overall genetic similarity. Hierarchical clustering of the genomes based on average AAIs indicated that members of the same OTU did not always cluster together (Figure 3). The OTUs that appear to be most greatly affected were OTUs 1 and 6. There was a general trend in which greater average AAIs correlated with genomes from the same OTU (average 76%) rather than genomes between OTUs (average 57%). This trend was consistent regardless of whether AAI or ANI was used (Supplementary Figure S3 and Supplementary Tables S5, S6, and S7), athough outliers were present in both instances. As expected, the number of outliers was higher for ANI than AAI comparisons. These outliers could not be explained by possible cross-contamination or chimeras as the genomes were not sequenced together and quality control measures removed chimeric sequences. It is possible that genome recovery size has a role, but only SAG-602-E04, the smallest genome recovered, is consistent with this. These results also indicate that the isolates are highly similar at the whole genome level, but quite distinct from the SAGs. Although previous studies have suggested that the 16S rRNA gene may not be the best indicator of overall genetic relatedness for many microorganism (Konstantinidis and Tiedje, 2005; Konstantinidis *et al.*, 2006; Staley, 2006; Swan *et al.*, 2013), this is the first evidence that this is also true for the *Zetaproteobacteria*.

COG comparison analyses suggested that there were no significant metabolic differences when comparing unique COGs in the *Zetaproteobacteria* genomes between geographic site, mat sample type (i.e. surface layer or bulk mat) and OTU designation. The most significant differences were between the SAGs and isolates, likely due to the phylogenetic breadth of the SAGs compared with the isolates. There were 632 unique COGs found in the SAGs compared with the 124 unique COGs found in the isolates (Supplementary Table S8), which correspond to 30% and 7% of the total unique COGs identified in each, respectively. This also indicates that the majority of COGs in both groups are shared. However, protein clustering indicated that proteins were not highly conserved between the

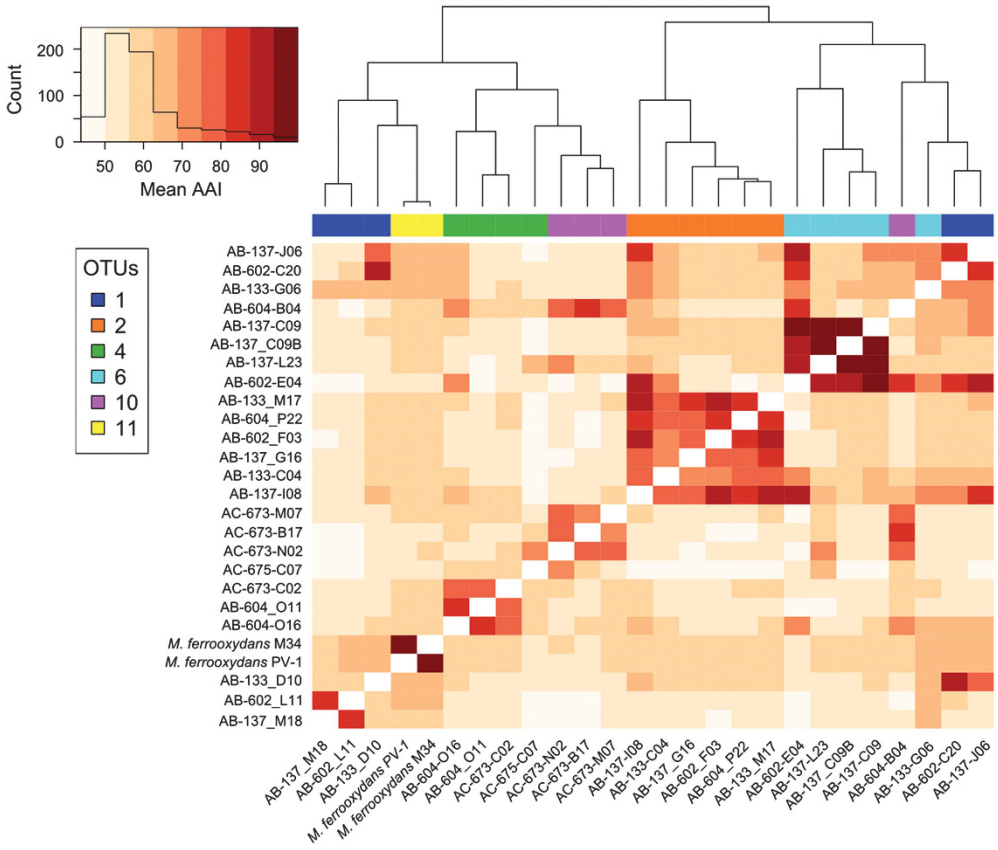

Figure 3 Hierachical clustering heatmap and dendogram of average amino acid identities (AAI) between all *Zetaproteobacteria* genomes in relation to their designated OTU. Genome self-comparisons are presented in white.

isolates and SAGs, regardless of COG category (Supplementary Figure S4 and Supplementary Table S9). Overall, these results suggest that the genomic variation observed by ANI and AAI comparisons may be due to subtle variations such as environmental responses or evolutionary influences. The subsequent sections will address some of these differences.

Carbon utilization

Studies of *Zetaproteobacteria* cultures that were isolated using autotrophic media have indicated that they are obligate autotrophs (Singer *et al.*, 2011). Strains PV-1 and M34 have complete pathways for carrying out carbon fixation using the Calvin–Benson–Bassham cycle. Eleven of the 23 SAGs analyzed here also show evidence of the Calvin–Benson–Bassham cycle, based on the presence of either the *ribulose 1,5-bisphosphate carboxylase* (*RubisCO*) gene or the phosphoribulokinase gene, both of which are specific to this pathway (Berg, 2011) (Supplementary Table S10). The detection of these genes did not correlate to genome size or OTU designation. There was no evidence of other carbon fixation mechanisms in any of the *Zetaproteobacteria* genomes. Both strains PV-1 and M34 contain a Form I and Form II large subunit *RubisCO* gene. Large subunit *RubisCO* genes were recovered from 8 of the 23 SAGs. Three additional SAGs contained *RubisCO* activation genes *cbbQ* and *cbbO*. Interestingly, all eight of the *RubisCO* genes found in the SAGs were of the Form II type (Figure 4). Form II *RubisCO* genes were notably absent from OTU 4 genomes, but at least one of these genomes had other genes from the Calvin–Benson–Bassham pathway, suggesting that they may still be capable of autotrophy.

The probability that Form I may have been present but not detected in any of the 23 analyzed SAGs because of incomplete genome recovery is extremely low (estimated 7.49E − 08%, Supplementary Methods) if it is assumed that multiple displacement amplification biases and subsequent genome coverage is random (Lasken and Stockwell, 2007). Therefore, while both isolates encode Form I and Form II *RubisCO* genes, it is unlikely that all *Zetaproteobacteria* do. The presence of Form II rather than Form I is interesting as it is indicative of adaptations to high CO_2 and very low O_2 environments (Badger and Bek, 2008; Tabita *et al.*, 2008), which is consistent with the O_2 profiles presented by Glazer and Rouxel (2009), in which the O_2 levels at Loihi Seamount iron mats were often below the detection limit (3 μM) even at the mat surface. These findings suggest that the broad

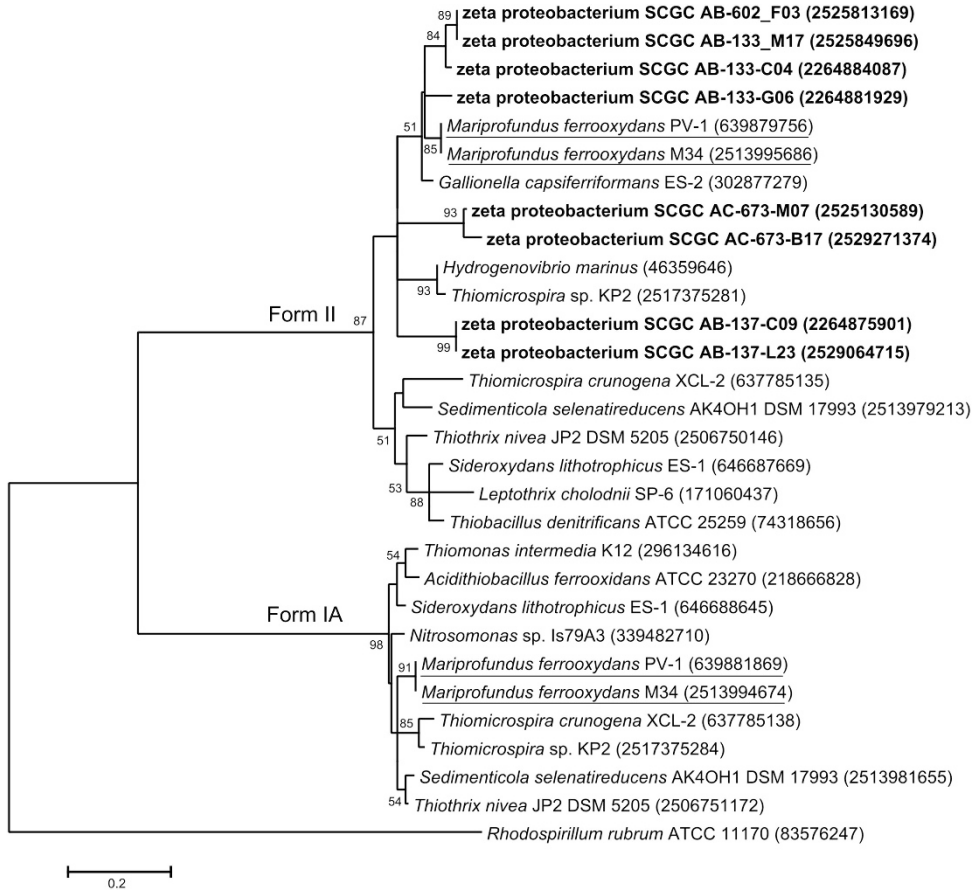

Figure 4 Maximum-likelihood phylogenetic tree of the RubisCO proteins of all *Zetaproteobacteria* SAGs (bold), *M. ferrooxydans* PV-1, and M34 (underlined) and other reference organisms. Numbers within parentheses are gene identification numbers. Bootstrap values are representative of 1000 iterations and those values below 50 are not shown.

phylogenetic spectrum of *Zetaproteobacteria* represented by SAGs analyzed in this study have a lower O_2 tolerance for fixing carbon compared with their cultivated counterparts. Decreased O_2 and/or increased CO_2 may represent favorable conditions for future cultivation efforts.

Nine of the 23 SAGs, as well as strains PV-1 and M34, show the potential for carbohydrate uptake through the phosphotransferase system (Supplementary Table S11). These genes were found across all OTUs with the exception of OTU 4 and include the mannose/fructose-specific component IIA and the Ntr-type mannitol/fructose-specific component IIA genes. As the phosphotransferase system genes may have a complex functional role in the cell (Deutscher *et al.*, 2014), the specific role of these genes in the *Zetaproteobacteria* remains uncertain. Seven SAGs from three OTUs (1, 4 and 10) have endoglucanases, β-glucosidases, xylanases, and chitin deacetylases, suggesting that there is a genetic potential for degradation and utilization of complex carbohydrates. The presence of these genes suggests that there is a potential for heterotrophy or possibly mixotrophy among some of the *Zetaproteobacteria*, although there is no evidence to date that either strain PV-1 nor M34 can grow using any carbon source other than CO_2.

Potential for iron oxidation

Neither the *Zetaproteobacteria* SAGs nor isolates contain homologs to well-characterized genes involved in iron oxidation identified in other microorganisms including *iro* (Amouric *et al.*, 2011), *rus* (Cox and Boxer, 1978), *mtoAB* (Shi *et al.*, 2012), *pioABC* (Jiao and Newman, 2007), *foxEYZ* (Croal *et al.*, 2007) and *cyt579* (Singer *et al.*, 2008). Previous studies have suggested that the *Zetaproteobacteria* may use a molybdopterin oxidoreductase and alternative respiratory complex III mechanism to carry out iron oxidation (Singer *et al.*, 2011). Molybdopterin oxidoreductase genes were found in 12 of the *Zetaproteobacteria* SAGs and in strain M34 forming two phylogenetic clusters (Figure 5a). Cluster 1 contains 11 SAG proteins predicted to be nitrate reductases. Cluster 2 contains the molybdopterin oxidoreductase proteins associated with alternative respiratory complex III homologs from strains PV-1 and M34, two SAGs as well as freshwater iron-oxidizing bacteria *Gallionella capsiferriformans* ES-2 and *Sideroxydans lithotrophicus* ES-1. The alternative respiratory complex III syntenous operon structure in members of Cluster 2 is conserved and in agreement with previous findings (Singer *et al.*, 2011) (Figure 5b).

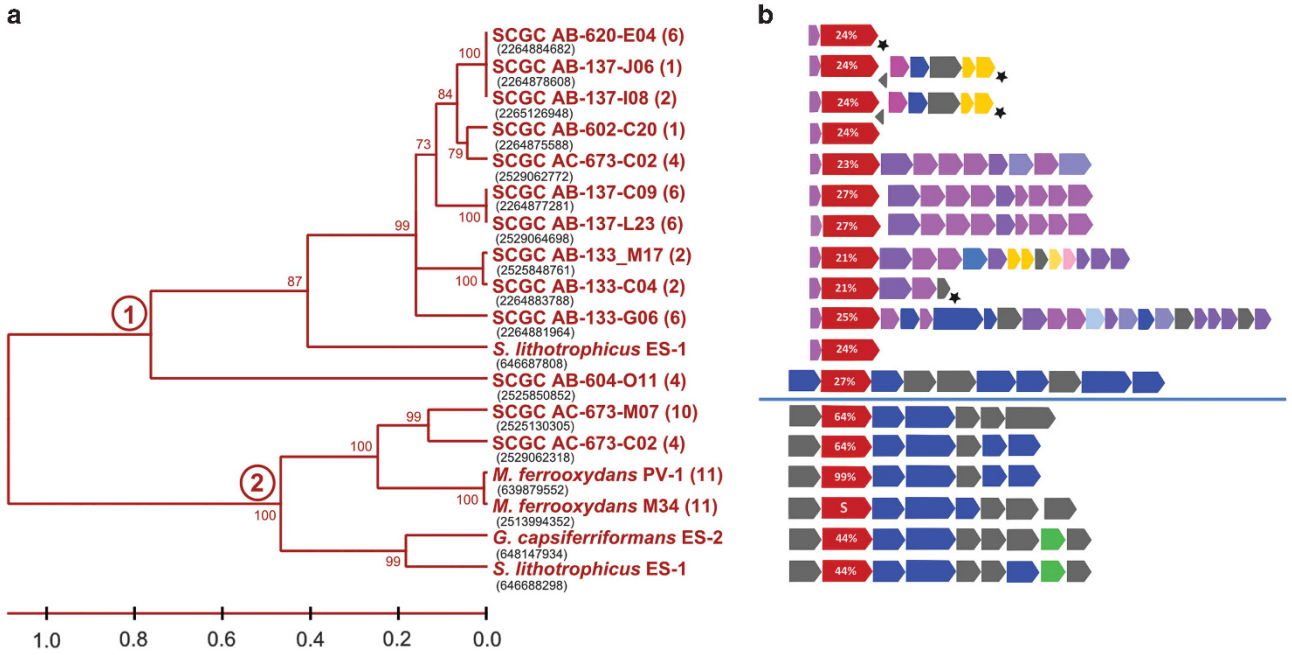

Figure 5 Protein phylogeny (**a**) and gene operon synteny (**b**) of the molybdopterin oxidoreductase protein of *M. ferrooxydans* PV-1, *M. ferrooxydans* M34 and the SAGs with homologs. Molybdopterin oxidoreductase protein (designated in red) phylogeny is presented in the linearized maximum-likelihood phylogenetic tree. BLAST similarities to *M. ferrooxydans* PV-1 (designated with an S) are presented on the gene of interest. OTU designations are listed within parentheses and gene ID numbers are in black. The blue line separates the two distinct clusters identified in the maximum-likelihood tree in (**a**). The black stars represent the end of the gene operon due to the end of a contig. Gene colors are indicative of COG functional category colors except for the molybdopterin oxidoreductase genes (in red), which correlate with the proteins in (**a**). Color Code: green, category O; dark blue, category C; light blue, category E; pale blue, category G; dark purple, category H; light purple, category P; violet, category Q; fuchsia, category L; light pink, category J; yellow, category T; light yellow, category D; gray, no category designated.

In contrast, the associated Cluster 1 operon in the SAGs is comprised of genes from different COG categories than those associated with Cluster 2. Many of the genes in Cluster 1's operon are unique to SAGs, with no homologs in either strain PV-1 or M34. These genes include nitrite reductases, molybdenum co-factor biosynthesis genes, molybdenum transporters, molybdopterin-converting factors, and lactate dehydrogenase. SAG AB-604_O11 encodes the only member of Cluster 1 that is a predicted ferredoxin-type nitrate reductase. While this protein is phylogenetically similar to those in Cluster 1, the gene operon structure is much more similar to Cluster 2's alternative respiratory complex III homologs. In all, these results pose intriguing questions about the functional role and evolution of these molybdopterin oxidoreductase genes and corresponding operons in the *Zetaproteobacteria* and specifically what role, if any, they may actually have in iron oxidation.

Cytochromes and electron transport

Strains PV-1 and M34 as well as fourteen SAGs encode a cbb_3-type cytochrome *c* oxidase, which has a high affinity for O_2 (Pitcher and Watmough, 2004) and is used under microaerophilic conditions. Notably, cbb_3 oxidase genes were not recovered from any of the OTU 4 SAGs and only from one of the OTU 10 SAGs (Supplementary Table S12). Instead, these SAGs have a more O_2-tolerant cytochrome *c* oxidase, suggesting that they are adapted to higher O_2 conditions. Strain M34 and eight SAGs, including members of OTU 10, have the bd-I ubiquinol oxidase, which is also used under microaerophilic conditions (Borisov *et al.*, 2011). The presence of both high O_2 and low O_2 adapted cytochromes suggests that OTU 10 members may be exposed to both low and high O_2 conditions. Therefore, this presents intriguing questions regarding the conditions under which they are exposed to higher O_2 concentrations, as well as what would constitute a higher O_2 concentration. This is also surprising as it was observed that OTU 10 members were more prevalent in the bulk mat samples, suggesting that they may be found deeper within the mats, which is not where higher O_2 concentrations would be expected.

There is an operon of ~20 genes that appears to be highly conserved among the *Zetaproteobacteria*. Emerson *et al.* (2013) pointed out this large cluster, composed of two distinct subclusters, is conserved between strain PV-1 and *S. lithotrophicus* ES-1. Acidophilic iron-oxidizing *Leptospirillum* spp. also show similarities at the amino-acid level to many of these genes (Emerson *et al.*, 2013). Homologs to these genes were identified in 10 of the SAGs (Figure 6). Interestingly, no homologs were found in any OTU 4 or 10 SAGs. Most of the genes in this

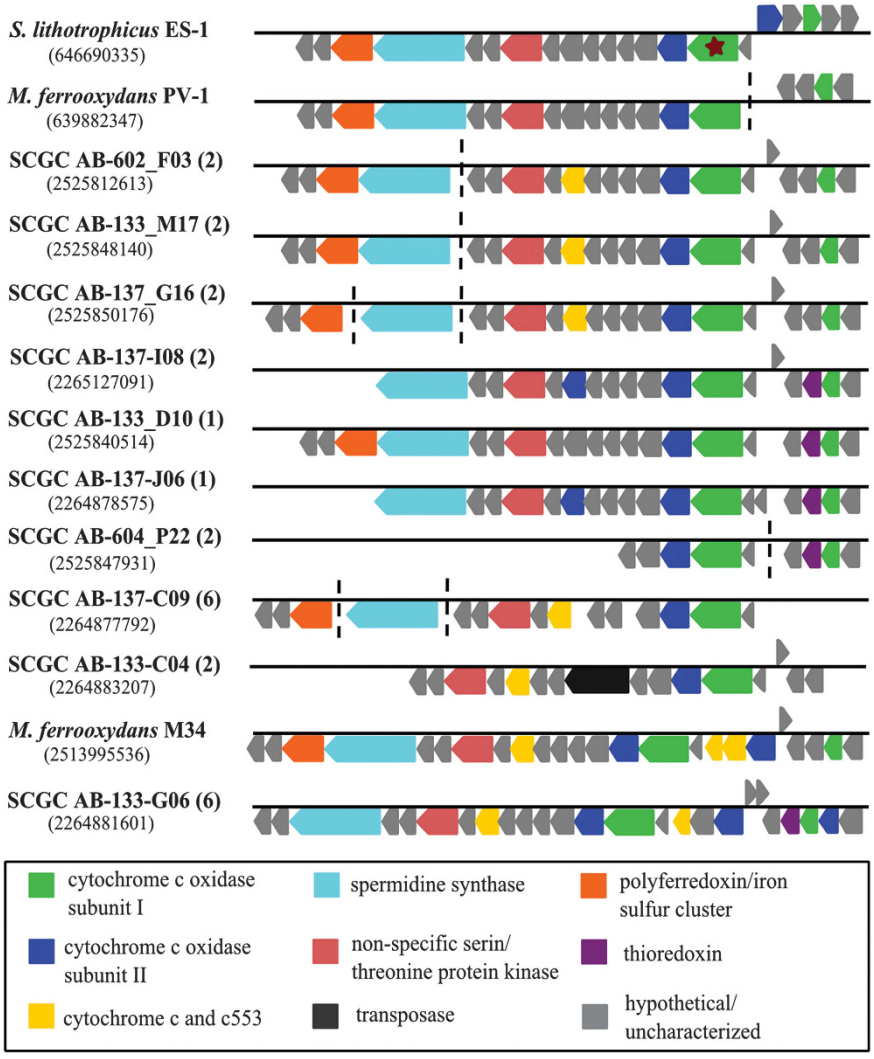

Figure 6 Gene operon synteny of conserved gene cluster homologs between freshwater iron oxidizer *S. lithotrophicus* ES-1, and *Zetaproteobacteria*. Gene ID numbers corresponding to homologs of *S. lithotrophicus* cytochrome *c* oxidase, cbb_3-type gene (labeled with a star) are provided under genome names. OTU designations are provided within parentheses. Dashed lines designate non-contiguous gene operons and directionality corresponds to that of *S. lithotrophicus* ES-1.

cluster are hypothetical; however, the few that have annotations indicate that they are likely involved in electron transport. The presence of this large gene cluster in many of the *Zetaproteobacteria*, *S. lithotrophicus* ES-1 and homologs in *Leptospirillum* spp. further suggests that they may have a common function in electron transport under microaerophilic conditions and possibly have a role in iron oxidation. Although the specific function of these genes remains unclear, their conservation among the *Zetaproteobacteria* and other iron oxidizers indicates that they warrant further investigation.

Oxidative stress and protection

Production of reactive O_2 species such as superoxide and hydroxyl radicals can occur under the high Fe(II) and fluctuating O_2 conditions (Cabiscol *et al.*, 2010) that the *Zetaproteobacteria* are exposed to. Compared with other SAGs, members of OTU 10 appear to have greater protection against reactive O_2 species. Superoxide dismutase is largely absent in the SAGs being present only in SAGs from OTU 10 and the isolates. There are also only two SAGs that encode catalase genes, one of which is OTU 10 SAG AC-673-M07 (Supplementary Table S13). Additionally, two OTU 10 members have bacterioferritin, which is absent in all other *Zetaproteobacteria* genomes. Bacterioferritin is intriguing as it has been shown to be involved in both iron storage and protection against oxidative damage created by the reaction of ferrous iron with O_2 (Carrondo, 2003). This increased repertoire of genes involved in mitigating oxidative stress suggests that OTU 10 may be adapted to fluctuating O_2 conditions compared with other uncultivated *Zetaproteobacteria*. It is currently unclear whether the depth within the mat, geographical location, or other factors have a role in the development of these attributes and should be evaluated further.

Nitrogen metabolism

Previous studies have indicated that Loihi Seamount vent fluids are enriched in ammonium and depleted in nitrate and nitrite compared with ambient seawater (Sedwick *et al.*, 1992; Karl *et al.*, 1988; Garcia *et al.*, 2006). Strains PV-1 and M34 were both isolated using medium in which ammonium was the sole nitrogen source, and genomic analysis indicates that they, as well as 13 SAGs, encode the genes necessary to utilize ammonium (Figure 7 and Supplementary Table S14). In addition, 11 SAGs have nitrate reductases, all of which show structural similarities to assimilatory nitrate reductases (Moreno-Vivián *et al.*, 1999; Kraft *et al.*, 2011). Nine SAGs contain nitrite reductases (both large and small subunits). This indicates that almost half of the analyzed cells have the potential to assimilate nitrate. These same SAGs also contain nitrate/nitrite transporters as well as the nitrogen assimilation regulatory genes *ntrB/ntrC* (Luque-Almagro *et al.*, 2011) directly upstream. Strain M34 encodes nitrogenase genes, although their functionality remains to be verified. Nitrogenase genes were surprisingly absent in all of the SAGs. Eleven SAGs as well as both isolates have nitrilase genes, which have been found to hydrolyze organic nitrogen compounds and produce ammonia (Podar *et al.*, 2005). One SAG (AB-137_G06) also contains a nitronate monooxygenase gene that may be involved in carbon–nitrogen compound degradation through the use of O_2 and forming nitrite as one of its degradation products (Gadda and Francis, 2010). There is also genetic evidence for potential urea utilization through the presence of transporters and urea carboxylase in both strains PV-1 and M34, as well as a urease gene in one of the SAGs (AC-673-M07).

There appears to be niche specialization among the *Zetaproteobacteria* in the type of nitrogen source they can use, whether it is nitrate, ammonium, organic compounds, nitrogen gas or some intermediate. Most have the potential to take up ammonium, but it appears as though some have acquired additional mechanisms for obtaining nitrogen. The partial recovery of genomes does not allow us to speculate on completeness of pathways as some of the genes may not have been recovered, but overall the detection of these genes in a high proportion of the SAGs suggest that they may have functional importance in the *Zetaproteobacteria*. Nitrogen availability, competition between cells, or fluctuating O_2 conditions may lead to the switch between nitrogen sources. It is surprising to find nitrate assimilation genes in so many of the SAGs, as the nitrate concentrations previously measured in the vent fluids were extremely low. The source of nitrate remains unknown, but it is possible that it comes from ambient seawater or that there are ammonia-oxidizing or other nitrifying microbes within the mats that are producing nitrate. The *Zetaproteobacteria* may also compete with these organisms for ammonia or O_2 leading to the development of these alternative nitrogen pathways. It appears as though *Zetaproteobacteria* use nitrate directly for a nitrogen source, although it cannot be ruled out that nitrate may be used as an alternative electron acceptor under certain conditions. Although there are a number of nitrate-dependent iron-oxidizing bacteria that grow under anaerobic conditions (Weber *et al.*, 2006), thus far there is no direct evidence that the *Zetaproteobacteria* are capable of this metabolism. These results indicate that the *Zetaproteobacteria* may have more diverse systems for acquiring nitrogen and therefore have a more significant role in nitrogen transformations than previously believed. The interactions between the *Zetaproteobacteria* living within the mats and different forms of nitrogen may be complex, dynamic, and represent a significantly understudied process in these environments.

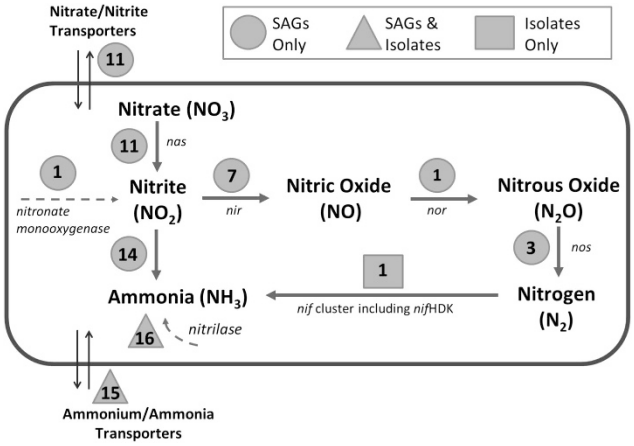

Figure 7 Potential nitrogen transformations by *Zetaproteobacteria* based on genomic analysis of 23 SAGs and two sequenced isolates (*M. ferrooxydans* PV-1 and *M. ferrooxydans* M34). Symbols represent in which genomic source genes were found. Numbers within symbols represent the number of genomes in which that gene was found. Broken lines denote unknown source compounds.

Using genomic insights to isolate novel Zetaproteobacteria

The results presented here pose new enrichment strategies through which novel *Zetaproteobacteria* may be isolated. Enrichments targeting a wider range of O_2 conditions may lead to the isolation of some of these higher and lower O_2-tolerant *Zetaproteobacteria*. For example, higher O_2-tolerant *Zetaproteobacteria* appear to be largely from OTU 10, which are more prevalent at Marker 57 and could be targeted from this site in future sampling efforts. Similarly, they were also found in both bulk mat samples, suggesting that suction samples may be more appropriate starting material for inoculating enrichments specific for these organisms. The use of a variety of available nitrogen sources as well as the addition of heterotrophic or complex carbon sources may be advantageous for isolating some of these novel groups as well.

Although we cannot demonstrate that the genomes obtained for this study are from cells that are carrying out iron oxidation, there is surprisingly little genomic evidence in these 23 SAGs of other lithotrophic metabolisms, including sulfur oxidation, ammonia oxidation, and methane oxidation. The possible exception being hydrogen oxidation because of the presence of Type I and Type III Ni-Fe hydrogenases and cytochrome *b*561 genes that have been implicated in hydrogen oxidation processes (Yoon *et al*., 2008) (Supplementary Table S15). Another finding gleaned from the SAGs is the presence of aerobic and anaerobic cobalamin (vitamin b_{12}) biosynthesis genes and corresponding cobalt transporters (Supplementary Tables S16 and S17), indicating that enrichments lacking B vitamins may be a relatively easy way to enrich novel strains.

We used the data presented here to isolate *Mariprofundus* sp. EKF-M39, a novel stalk-forming, iron-oxidizing *Zetaproteobacteria* strain from Loihi Seamount using medium containing nitrate as the sole nitrogen source, under very low O_2 conditions. Based on 16S rRNA gene comparisons, this isolate is most closely related to strain PV-1 and strain M34, but at only 96% gene similarity it is the most phylogenetically distinct *Zetaproteobacteria* isolate obtained from Loihi Seamount to date (Supplementary Figure S5). The genome of strain EKF-M39 was sequenced (2.7 Mb) and whole genome comparisons indicate that it is even more distant from these two isolates at the ANI and AAI levels (71% and 69%, respectively). Compared with the SAGs, this isolate has an average similarity of 93% based on 16S, 67% ANI and 59% AAI.

Physiologic studies with strain EKF-M39 were conducted in an effort to verify some of the predicted SAG functions. This organism carries out lithotrophic iron oxidation under microaerophilic conditions. Strain EKF-M39 has phosphotransferase system genes, but did not show evidence for heterotrophic growth on mannose, fructose, acetate, or cellobiose when these were provided as the sole carbon sources (Table 2 and Supplementary Methods). It also is unable to carry out vitamin B_{12} auxotrophy under iron-oxidizing conditions or grow with hydrogen as the electron donor. Most interestingly, strain EKF-M39 confirms that the *Zetaproteobacteria* are versatile in their nitrogen utilization capabilities as this strain can grow using either nitrate or ammonium as the sole nitrogen source. It can also grow in nitrogen-deplete medium, which is consistent with nitrogen fixation. Strain EKF-M39 has the genetic capacity to carry out different modes of nitrogen assimilation and has a full suite of nitrogen fixation genes confirming the physiologic tests. Strain EKF-M39 is not capable of anaerobic nitrate-dependent iron oxidation (Table 2 and Supplementary Methods), confirming that it uses nitrate solely as a nitrogen source.

The genomic potential of this isolate supports additional findings introduced by the analysis of the SAGs. It represents the first *Zetaproteobacteria* isolate to contain a Cluster 1 molybdopterin oxidoreductase gene (Supplementary Figure S6) and lack the Cluster 2 molybdopterin oxidoreductase. The Cluster 1 molybdopterin oxidoreductase gene and corresponding operon in strain EKF-M39 are the only nitrate assimilation genes in the near-complete genome, indicating that these are likely responsible for the growth by strain EKF-M39 when nitrate was the sole nitrogen source. This possibly indicates that strain EKF-M39 has a different mechanisms for iron oxidation compared with strains PV-1 and M34, and/or the Cluster 2 molybdopterin oxidoreductase genes are not essential for iron oxidation. Strain EKF-M39 also has the conserved set of cytochrome genes found in strains PV-1, M34, and *S. lithotrophicus* ES-1, supporting the potential importance of this gene operon in electron transport and possibly in iron oxidation. Last, it is the first *Zetaproteobacteria* isolate to have only a form II *RubisCO* gene (Supplementary Figure S7), which is consistent with the very low O_2 enrichment conditions used to isolate this organism, demonstrating that this strategy can be successful for finding unique isolates. Additional in-depth physiological and genomic analyses of this isolate are currently being conducted and will be described elsewhere.

Table 2 Physiological characteristics of the iron-oxidizing isolate *Mariprofundus* sp. EKF-M39

Oxygen conditions	*Type of growth*	*Substrate*	*Growth*
Microaerophilic	Autotrophy	CO_2	Yes
	Heterotrophy	Mannose	No
		Fructose	No
		Sodium acetate	No
		Cellobiose	No
	Nitrogen utilization	Ammonium chloride	Yes
		Sodium nitrate	Yes
		N-deplete (N_2 only)	Yes
	Vitamin B_{12} auxotrophy	No vitamins	No
		Riboflavin + Vit B_5 only	No
	Hydrogen oxidation	H_2	No
Anaerobic	Nitrate-dependent iron oxidation	Sodium nitrate	No

Conclusions

Partial genome sequences from 23 uncultivated *Zetaproteobacteria* from Loihi Seamount significantly expand our understanding of the genetic diversity, metabolic potential, and ecologic roles of this elusive taxonomic group. This study suggests that O_2 tolerances and nitrogen metabolisms may have a significant role in niche specialization among the *Zetaproteobacteria*. Future analyses of fine-scale sampling efforts within these mats will be necessary to better understand the extent of these niche differentiations. Aside from Fe-oxidation, there was a surprising lack of other lithotrophic or anaerobic metabolisms in the *Zetaproteobacteria* SAGs at Loihi Seamount. This may explain why previous studies have identified these organisms almost exclusively in high iron environments. Genomic comparisons with *Zetaproteobacteria* from non-hydrothermal vent sites and other geographic locations will also help determine the true breadth of their metabolic versatility on a global scale. Overall, these findings lay the groundwork for future laboratory and *in situ* studies as they provide insight into the potential lifestyle of the *Zetaproteobacteria*.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank Brian Thompson, Nicole Poulton, and Elizabeth Fergusson for their technical expertise and support in single-cell sorting and sequencing, as well as Anna Leavitt for her assistance with the physiologic studies. We also thank the captains, crew and operators of the R/V *Kilo Moana*, R/V *Thomas N Thompson*, and *Jason II* ROV without whom this work would not have been possible. Last, we appreciate the intellectual input provided by the members of the Single Cell Genomics Center and Geomicrobiology labs at Bigelow. This work was financially supported by the NASA EPSCoR program administered through the Maine Space Consortium Grant 10-EPSCoR-0005 and the National Science Foundation Grants OCE-0821374 and OCE-1155754. The work conducted by the US Department of Energy Joint Genome Institute is supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231. Additional sequencing support for Bigelow Laboratory was provided through the Joint Genome Institute's Community Sequencing Program, Project 560.

References

Amouric A, Brochier-Armanet C, Johnson DB, Bonnefoy V, Hallberg KB. (2011). Phylogenetic and genetic variation among Fe(II)-oxidizing acidithiobacilli supports the view that these comprise multiple species with different ferrous iron oxidation pathways. *Microbiology* **157**: 111–122.

Badger MR, Bek EJ. (2008). Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO_2 acquisition by the CBB cycle. *J Exp Bot* **59**: 1525–1541.

Berg IA. (2011). Ecological aspects of the distribution of different autotrophic CO_2 fixation pathways. *Appl Environ Microbiol* **77**: 1925–1936.

Borisov VB, Gennis RB, Hemp J, Verkhovsky MI. (2011). The cytochrome *bd* respiratory oxygen reductases. *BBA-Bioenergetics* **1807**: 1398–1413.

Cabiscol E, Tamarit J, Ros J. (2010). Oxidative stress in bacteria and protein damage by reactive oxygen species. *Int Microbiol* **3**: 3–8.

Carrondo MA. (2003). Ferritins, iron uptake and storage from the bacterioferritin viewpoint. *EMBO J* **22**: 1959–1968.

Chan CS, Fakra SC, Emerson D, Fleming EJ, Edwards KJ. (2011). Lithotrophic iron-oxidizing bacteria produce organic stalks to control mineral growth: implications for biosignature formation. *ISME J* **5**: 717–727.

Cox JC, Boxer DH. (1978). The purification and some properties of rusticyanin, a blue copper protein involved in iron(II) oxidation from *Thiobacillus* ferro-oxidans. *Biochem J* **174**: 497–502.

Croal LR, Jiao Y, Newman DK. (2007). The *fox* operon from *Rhodobacter* strain SW2 promotes phototrophic Fe(II) oxidation in *Rhodobacter capsulatus* SB1003. *J Bacteriol* **189**: 1774–1782.

Dang H, Chen R, Wang L, Shao S, Dai L, Ye Y *et al.* (2011). Molecular characterization of putative biocorroding microbiota with a novel niche detection of *Epsilon*- and *Zetaproteobacteria* in Pacific Ocean coastal seawaters. *Environ Microbiol* **13**: 3059–3074.

Deutscher J, Aké FMD, Derkaoui M, Zébré AC, Cao TN, Bouraoui H *et al.* (2014). The bacterial phosphoenolpyruvate: carbohydrate phosphotransferase system: regulation by protein phosphorylation and phosphorylation-dependent protein–protein interactions. *Microbiol Mol Biol Rev* **78**: 231–256.

Emerson D, Floyd MM. (2005). Enrichment and isolation of iron-oxidizing bacteria at neutral pH. *Methods Enzymol* **397**: 112–123.

Emerson D, Moyer CL. (2010). Microbiology of seamounts: common patterns observed in community structure. *Oceanography* **23**: 148–163.

Emerson D, Field EK, Chertkov O, Davenport KW, Goodwin L, Munk C *et al.* (2013). Comparative genomics of freshwater Fe-oxidizing bacteria: implications for physiology, ecology, and systematics. *Front Microbiol* **4**: 254.

Fleming EJ, Davis RE, McAllister SM, Chan CS, Moyer CL, Tebo BM *et al.* (2013). Hidden in plain sight: discovery of sheath-forming, iron-oxidizing *Zetaproteobacteria* at Loihi Seamount, Hawaii, USA. *FEMS Microbiol Ecol* **85**: 116–127.

Gadda G, Francis K. (2010). Nitronate monooxygenase, a model for anionic flavin semiquinone intermediates in oxidative catalysis. *Arch Biochem Biophys* **493**: 53–61.

Garcia MO, Caplan-Auerbach J, De Carlo ÉH, Kurz MD, Becker N. (2006). Geology, geochemistry and earthquake history of Loihi Seamount, Hawaii's youngest volcano. *Chem Erde-Geochem* **66**: 81–108.

Garcia SL, McMahon KD, Martinez-Garcia M, Srivastava A, Sczyrba A, Stepanauskas R *et al.* (2012). Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. *ISME J* **7**: 137–147.

Glazer BT, Rouxel OJ. (2009). Redox speciation and distribution within diverse iron-dominated microbial habitats at Loihi Seamount. *Geomicrobiol J* **26**: 606–622.

Jiao Y, Newman DK. (2007). The *pio* operon is essential for phototrophic Fe(II) oxidation in *Rhodopseudomonas palustris* TIE-1. *J Bacteriol* **189**: 1765–1773.

Kamke J, Sczyrba A, Ivanova N, Schwientek P, Rinke C, Mavromatis K *et al.* (2013). Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *ISME J* **7**: 2287–2300.

Karl D, McMurtry G, Malahoff A, Garcia M. (1988). Loihi Seamount, Hawaii: a mid-plate volcano with a distinctive hydrothermal system. *Nature* **335**: 532–535.

Katoh K, Misawa K, Ki Kuma, Miyata T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**: 3059–3066.

Konstantinidis KT, Tiedje JM. (2005). Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* **187**: 6258–6264.

Konstantinidis KT, Ramette A, Tiedje JM. (2006). The bacterial species definition in the genomic era. *Philos Trans Roy Soc Ser B* **361**: 1929–1940.

Kraft B, Strous M, Tegetmeyer HE. (2011). Microbial nitrate respiration–genes, enzymes and environmental distribution. *J Biotechnol* **155**: 104–117.

Kumar S, Nei M, Dudley J, Tamura K. (2008). MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* **9**: 299–306.

Lasken RS, Stockwell TB. (2007). Mechanism of chimera formation during the Multiple Displacement Amplification reaction. *BMC Biotechnol* **7**: 19.

Ludwig W, Strunk O, Westram R, Richter L, Meier H. (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.

Luque-Almagro VM, Gates AJ, Moreno-Vivián C, Ferguson SJ, Richardson DJ, Roldán M. (2011). Bacterial nitrate assimilation: gene distribution and regulation. *Biochem Soc T* **39**: 1838–1843.

McAllister SM, Davis RE, McBeth JM, Tebo BM, Emerson D, Moyer CL. (2011). Biodiversity and emerging biogeography of the neutrophilic iron-oxidizing Zetaproteobacteria. *Appl Environ Microbiol* **77**: 5445–5457.

McBeth JM, Little BJ, Ray RI, Farrar KM, Emerson D. (2011). Neutrophilic iron-oxidizing '*Zetaproteobacteria*' and mild steel corrosion in nearshore marine environments. *Appl Environ Microbiol* **77**: 1405–1412.

Moreno-Vivián C, Cabello P, Martínez-Luque M, Blasco R, Castillo F. (1999). Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases. *J Bacteriol* **181**: 6573–6584.

Pitcher RS, Watmough NJ. (2004). The bacterial cytochrome cbb_3 oxidases. *Biochim Biophys Acta Bioenerget* **1655**: 388–399.

Podar M, Eads JR, Richardson TH. (2005). Evolution of a microbial nitrilase gene family: a comparative and environmental genomics study. *BMC Evol Biol* **5**: 42.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J *et al.* (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**: 7188–7196.

Rassa AC, McAllister SM, Safran SA, Moyer CL. (2009). Zeta-proteobacteria dominate the colonization and formation of microbial mats in low-temperature hydrothermal vents at Loihi Seamount, Hawaii. *Geomicrobiol J* **26**: 623–638.

Richter M, Rosselló-Móra R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA* **106**: 19126–19131.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F *et al.* (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431–437.

Schloss P, Handelsman J. (2005). Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**: 1501–1506.

Sedwick PN, McMurtry G, Macdougall J. (1992). Chemistry of hydrothermal solutions from Pele's vents, Loihi Seamount, Hawaii. *Geochim Cosmochim Acta* **56**: 3643–3667.

Shi L, Rosso KM, Clarke TA, Richardson DJ, Zachara JM, Fredrickson JK. (2012). Molecular underpinnings of Fe(III) oxide reduction by *Shewanella oneidensis* MR-1. *Front Microbiol* **3**: 50.

Singer E, Emerson D, Webb EA, Barco RA, Kuenen JG, Nelson WC *et al.* (2011). *Mariprofundus ferrooxydans* PV-1 the first genome of a marine Fe(II) oxidizing Zetaproteobacterium. *PLoS One* **6**: e25386.

Singer E, Heidelberg JF, Dhillon A, Edwards KJ. (2013). Metagenomic insights into the dominant Fe(II) oxidizing *Zetaproteobacteria* from Lo'ihi, Hawai'I. *Front Microbiol* **4**: 52.

Singer SW, Chan CS, Zemla A, VerBerkmoes NC, Hwang M, Hettich RL *et al.* (2008). Characterization of cytochrome 579, an unusual cytochrome isolated from an iron-oxidizing microbial community. *Appl Environ Microbiol* **74**: 4454–4462.

Staley JT. (2006). The bacterial species dilemma and the genomic–phylogenetic species concept. *Philos Trans R Soc Ser B* **361**: 1899–1909.

Stamatakis A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.

Stepanauskas R, Sieracki ME. (2007). Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc Natl Acad Sci USA* **104**: 9052–9057.

Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D *et al.* (2011). Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* **333**: 1296–1300.

Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, González JM *et al.* (2013). Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc Natl Acad Sci USA* **110**: 11463–11468.

Tabita FR, Satagopan S, Hanson TE, Kreel NE, Scott SS. (2008). Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *J Exp Bot* **59**: 1515–1524.

Thompson JD, Higgins DG, Gibson TJ. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting,

position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.

Weber KA, Achenbach LA, Coates JD. (2006). Microorganisms pumping iron: anaerobic microbial iron oxidation and reduction. *Nat Rev Microbiol* **4**: 752–764.

Wilkins MJ, Kennedy DW, Castelle CS, Field EK, Stepanauskas R, Fredrickson JK *et al.* (2014). Single-cell genomics reveal metabolic strategies for growth and survival in an oligotrophic aquifer. *Microbiology* **160**: 362–372.

Woyke T, Xie G, Copeland A, Gonzalez JM, Han C, Kiss H *et al.* (2009). Assembling the marine metagenome, one cell at a time. *PLoS One* **4**: e5299.

Woyke T, Sczyrba A, Lee J, Rinke C, Tighe D, Clingenpeel S *et al.* (2011). Decontamination of MDA reagents for single cell whole genome amplification. *PLoS One* **6**: e26161.

Yoon KS, Tsukada N, Sakai Y, Ishii M, Igarashi Y, Nishihara H. (2008). Isolation and characterization of a new facultatively autotrophic hydrogen-oxidizing Betaproteobacterium, *Hydrogenophaga* sp. AH-24. *FEMS Microbiol Lett* **278**: 94–100.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)