# Ecophysiology of the cosmopolitan OM252 bacterioplankton (Gammaproteobacteria)

Emily R. Savoie[1], V. Celeste Lanclos[2], Michael W. Henson[2,*], Chuankai Cheng[2], Eric W. Getz[2], Shelby J. Barnes[2], Douglas E. LaRowe[3], Michael S. Rappé[4], and J. Cameron Thrash[2,#]

1. Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, LA, USA
2. Department of Biological Sciences, University of Southern California, Los Angeles, CA, USA
3. Department of Earth Sciences, University of Southern California, Los Angeles, CA, USA
4. Hawai'i Institute of Marine Biology, School of Ocean and Earth Science and Technology, University of Hawai‘i at Mānoa, Kāne‘ohe, HI, USA

[*]Current address: Department of Geophysical Sciences, University of Chicago, Chicago, IL, USA

[#]Correspondence:
J. Cameron Thrash
University of Southern California
3616 Trousdale Pkwy, AHF 209
Los Angeles, CA 90089, USA
thrash@usc.edu

**Abstract**

Among the thousands of species that comprise marine bacterioplankton communities, most remain functionally obscure. One key cosmopolitan group in this understudied majority is the OM252 clade of Gammaproteobacteria. Although frequently found in sequence data and even previously cultured, the diversity, metabolic potential, physiology, and distribution of this clade has not been thoroughly investigated. Here we examined these features of OM252 bacterioplankton using a newly isolated strain and genomes from publicly available databases. We demonstrated that this group constitutes a globally distributed novel genus (*Candidatus* Halomarinus), sister to *Litoricola*, comprising two subclades and multiple distinct species. OM252 organisms have small genomes (median 2.21 Mbp) and are predicted obligate aerobes capable of alternating between chemoorganoheterotrophic and chemolithotrophic growth using reduced sulfur compounds as electron donors, with subclade I genomes encoding the Calvin-Benson-Bassham cycle for carbon fixation. One representative strain of subclade I, LSUCC0096, had extensive halotolerance but a mesophilic temperature range for growth, with a maximum of 0.36 doublings/hr at 35˚C. Cells were curved rod/spirillum-shaped, ~1.5 x 0.2 µm. Growth on thiosulfate as the sole electron donor under autotrophic conditions was roughly one third that of heterotrophic growth, even though calculations indicated similar Gibbs energies for both catabolisms. These phenotypic data show that some *Ca.* Halomarinus organisms can switch between serving as carbon sources or sinks and indicate the likely anabolic cost of lithoautotrophic growth. Our results thus provide new hypotheses about the roles of these organisms in global biogeochemical cycling of carbon and sulfur.

**Importance**

Marine microbial communities are teeming with understudied taxa due to the sheer numbers of species in any given sample of seawater. One group, the OM252 clade of Gammaproteobacteria, has been identified in gene surveys from myriad locations, and one isolated organism has even been genome sequenced (HIMB30). However, further study of these organisms has not occurred. Using another isolated representative (strain LSUCC0096) and publicly available genome sequences from metagenomic and single-cell genomic datasets, we examined the diversity within the OM252 clade, the distribution of these taxa in the world's oceans, reconstructed the predicted metabolism of the group, and quantified growth dynamics in LSUCC0096. Our results generate new knowledge about the previously enigmatic OM252 clade and point towards the importance of facultative chemolithoautotrophy for supporting some clades of ostensibly "heterotrophic" taxa.

## Introduction

Marine bacterioplankton constitute $10^4$ to $10^7$ cells per milliliter in seawater (1–3), spread across hundreds to thousands of operational taxonomic units (OTUs) (2). However, many of these bacterioplankton lineages have no assigned metabolic or ecological roles and we know little more about them than their distribution in 16S rRNA gene surveys. While some of the dominant groups like SAR11 Alphaproteobacteria, *Prochlorococcus* Cyanobacteria, and SAR86 Gammaproteobacteria rightly attract considerable attention (4–7), many taxa that occur at somewhat lower relative abundances, but nevertheless are cosmopolitan microbial community members of the global oceans, have received comparably little study. One of these groups, the OM252 clade of Gammaproteobacteria, was first described over twenty years ago in clone library sequences from surface waters overlying the continental shelf off Cape Hatteras, North Carolina (8). This group is widely distributed. OM252 16S rRNA gene sequences have been reported from Sapelo Island off the coast of Georgia (9), the Gulf of Mexico (10–12), Kāneʻohe Bay in Oahu (13), the eutrophic coastal North Sea near Amsterdam (14), a lagoon in the Clipperton Atoll off the western coast of Mexico (15), and the Gulf of Lyon in the Mediterranean Sea (16). Sequences in GenBank with high percent identity to the OM252 clade also have come from near Cocos Island in the eastern tropical Pacific Ocean as well as the East China Sea. OM252 sequences occur in less saline waters, like the estuarine zone of the Jiulong River, China, and lakes with varying salinities in Tibet (17); but also hypersaline environments, like the Salton Sea in California (18), and salterns in Spain (19). There are even reports that indicate OM252 bacteria may be at least transiently associated with marine invertebrate microbiomes (20, 21). Thus, it appears that OM252 bacteria inhabit a variety of habitats and may have a euryhaline lifestyle.

Despite the widespread distribution of OM252 bacterioplankton, they remain poorly studied. The first reported isolate, HIMB30, was obtained via high-throughput dilution-to-extinction (DTE) cultivation with a natural seawater medium inoculated from Kāneʻohe Bay, Hawaiʻi (22). The ~2.17 Mbp HIMB30 genome predicted partial glycolysis, a complete TCA cycle, phototrophy via proteorhodopsin, carbon monoxide and sulfur oxidation, and $CO_2$ fixation via the Calvin-Benson-Bassham cycle (22). However, these functions have not been demonstrated experimentally, nor have growth parameters, such as temperature or salinity tolerances, been investigated. We also don’t know how representative the HIMB30 features above are for the clade. Even the phylogenetic position of OM252 within the Gammaproteobacteria remains in question. The first clone library sequence branched sister to the OM182 clone and Oceanospirillales sequences (8). The closest described organisms, *Litoricola* spp., share less than 90% 16S rRNA gene identity with HIMB30 (22). Furthermore, the gammaproteobacterial phylogeny continues to evolve, with many traditionally recognized groups no longer remaining monophyletic (23), and additional genomes from uncultivated organisms changing the topology (24). The current Genome Taxonomy Database (GTDB release 05-RS95) indicates these

organisms belong to the family Litoricolaceae, in the newly reconstituted Order Pseudomonadales (24–26).

Our previous work combining cultivation and cultivation-independent methods demonstrated that OM252 was a prominent member of the coastal northern Gulf of Mexico (10, 12). Our 16S rRNA gene amplicon data indicated at least two distinct amplicon sequence variants (ASVs) within the single observed OM252 OTU across six sampling sites and 3 different years (12). That OTU was the 25th most abundant bacterioplankton taxon in the high-salinity community (salinities > 12) observed in the three-year dataset. OM252 thus represented an important medium-abundance organism in that coastal environment. Furthermore, our artificial media facilitated ready cultivation of OM252 members, with over 30 strains isolated over the course of 17 experiments (12).

To improve our understanding of the physiology, ecology, and evolutionary relationships of the OM252 clade, we sequenced the genome of one representative isolate, LSUCC0096, and performed comparative genomic analyses with this organism, HIMB30, and 23 other publicly available environmental genomes. In parallel, we characterized physiological aspects of LSUCC0096 relevant to OM252 biology. OM252 comprised at least two subclades (I and II), both of which had a globally cosmopolitan distribution. OM252 clade members share many of the same metabolic features; however, there is subclade differentiation in the capacity for predicted sulfur-based chemolithoautotrophy. LSUCC0096 had a wide tolerance for salinity, growing from low salinity brackish water to nearly double the salinity of seawater. Furthermore, we showed that LSUCC0096 could grow under chemolithoautotrophic conditions with thiosulfate as the sole electron donor and estimate the energetic consequences of this metabolism on growth rates. Contrary to existing nomenclature in GTDB, our comparative genomic data support the designation of OM252 as a separate genus from *Litoricola*, which we propose as *Candidatus* Halomarinus, along with names for three species within the genus. These results expand our understanding of the genomic diversity, distribution, and lifestyles within the OM252 clade and provide the first cellular and physiological data for these organisms. They also raise new questions about the relationship between facultative chemolithotrophy and OM252 ecology.

## Materials and Methods

*LSUCC0096 isolation, genome sequencing, and genome assembly*

Strain LSUCC0096 was isolated and initially identified via 16S rRNA gene PCR as previously reported (10) from surface water collected in Bay Pomme d'Or near the Mississippi River Birdfoot delta on January 12, 2015 (Buras, LA) (29.348784, -89.538171). DNA was extracted from cultures of LSUCC0096 that had reached max cell density (~$10^6$ cells $mL^{-1}$) growing in JW1 medium (10) at room temperature using a MoBio PowerWater DNA Isolation kit (QIAGEN, Massachusetts, USA) following the manufacturer's protocols. Truseq DNA-seq

Library preparation and Illumina MiSeq (paired-end 250 bp reads) sequencing was completed at the Argonne National Laboratory Environmental Sample Preparation and Sequencing Facility, producing 242,062 reads (Table S1). The genome was assembled using the A5 MiSeq pipeline (version 20150522) (27) with default settings. The LSUCC0096 genome was annotated at IMG (28) (Taxon ID 2639762503). For comparative genomics we re-annotated the genome along with other analyzed genomes using Anvi'o (see below), and the scaffolds were also deposited in GenBack (see Data Availability section).

*16S rRNA gene phylogenies*

The 16S rRNA gene of the LSUCC0096 genome was searched against both the NCBI nt and refseq_rna databases (accessed August, 2018) using megablast v. 2.2.28+ with –max_target_seqs 1000 and –num_threads 16. A selection of best hits was generated from each blast search and combined with the LSUCC0096 and HIMB30 16S rRNA genes from IMG, along with those from five different *Litoricola* spp. and the original OM252 clone library sequence U70703.1 (8). Additional 16S rRNA genes from the OM252 MAGs and SAGs were obtained from the Anvi'o genome database (see above) using the command anvi-get-sequences-for-hmm-hits --external-genomes external-genomes.txt -o 16S.fna --hmm-source Ribosomal_RNAs --gene Bacterial_16S_rRNA (or --gene Archaeal_16S_rRNA). The 16S rRNA gene of TOBG-NAT-109 had best blast (megablast online, default settings) hits to Bacteroides sequences, and was removed from further 16S rRNA gene analyses. The remaining sequences were aligned with MUSCLE v3.6 (29), culled with TrimAl v1.4.rev22 (30) using the -automated1 flag, and the final alignment was inferred with IQ-TREE v1.6.11 (31) with default settings and -bb 1000 for ultrafast bootstrapping (32). Tips were edited with the nw_rename script within Newick Utilities v1.6 (33) and trees were visualized with Archaeopteryx (34). Fasta files for these trees and the naming keys are provided at https://doi.org/10.6084/m9.figshare.14036573.

*Additional taxon selection*

The HIMB30 genome (22) was downloaded from IMG (Taxon ID 2504557021). To provide a more comprehensive analysis of the OM252 clade beyond the LSUCC0096 and HIMB30 genomes, we searched for metagenome-assembled genomes (MAGs) that matched LSUCC0096 and HIMB30 using the following methods. We downloaded MAGs reconstructed from the Tara Oceans dataset (35, 36) and the northern Gulf of Mexico (37, 38). We identified all MAGs with average nucleotide identities (ANI) of > 76% to LSUCC0096 and HIMB30 using FastANI v1.1 (39) with default settings. These MAGs and the LSUCC0096 and HIMB30 genomes were then placed into the Genome Taxonomy Database (GTDB) tree [which also included additional MAGs constructed from the Tara Oceans dataset (40)] with GTDBtk v0.2.1 (24) (downloaded Feb 2019) using "classify_wf". All genomes occurred in a monophyletic group including f_Litoricolaceae. The additional genomes from GTDB in this clade were downloaded. We then searched six representative genomes (LSUCC0096, HIMB30, GCA_002480175.1, GCA_002691485.1, UBA1114, UBA12265) against the Gammaproteobacteria Single-Amplified

Genomes (SAGs) generated from GORG-Tropics collection (41) using FastANI as above. Finally, all genomes from this selection process were compared to each other with FastANI again. We then calculated the percent completion and contamination using CheckM v1.0.13 (42) using "lineage_wf". We designated genomes as redundant if they had an ANI value of ≥ 99% with another genome. If a genome had a redundant match, we kept the genome with the highest % completion and lowest % contamination. Genomes with less than 50% estimated completion were discarded. The final genome selection statistics are in **Table S1**, available at https://doi.org/10.6084/m9.figshare.14067362.

*Phylogenomics*

Based on the 16S rRNA gene phylogeny, we selected 208 genomes for a concatenated phylogenomic tree that spanned a variety of clades within the Gammaproteobacteria with members near OM252, plus 6 outgroup taxa from the Alpha- and Betaproteobacteria. These, together with 26 putative OM252 genomes (total 240), were analyzed using the Anvi'o phylogenomics pipeline through HMM assignment. Single copy marker genes that had membership in at least half the taxa in the tree (120) were selected using "anvi-get-sequences-for-hmm-hits --external-genomes external-genomes.txt -o temp.faa --hmm-source Rinke_et_al --return-best-hit --get-aa-sequences --min-num-bins-gene-occurs 120", which returned a fasta file for each of the resulting 78 gene clusters. Each of these were aligned with MUSCLE v3.6 (29), culled with TrimAl v1.4.rev22 (30) using the -automated1 flag, and concatenated with the geneStitcher.py script from the Utensils package (https://github.com/ballesterus/Utensils) as described (43). The final alignment had 29,631 amino acid positions, and the tree was inferred with IQ-TREE v1.6.11 (31) with default settings and -bb 1000 for ultrafast bootstrapping (32). Tree tips were edited with the nw_rename script within Newick Utilities v1.6 (33) and trees were visualized with Archaeopteryx (34) and FigTree v1.4.3 and edited with Adobe Illustrator. The concatenated alignment and naming key is provided at https://doi.org/10.6084/m9.figshare.14036594.

*Pangenomics*

Upon inspection of the phylogenomic tree, one putative genome, GCA_002408105.1, branched outside of the OM252 group (Fig. 1). We therefore excluded it from pangenomic analyses. The final 25 OM252 genomes were processed via Anvi'o v5.3 (44) using the pan-genomics workflow (45). This approach ensured that all genomes were subject to the same gene-calling and annotation workflow. Single copy marker genes via Anvi'o-provided HMMs and NCBI COGs were assigned, and Anvi'o-based gene calls were used for additional external annotation via Interproscan v5.33-72.0 (46) and KEGG assignments with GhostKOALA (47). All annotations are provided in Table S1 (https://doi.org/10.6084/m9.figshare.14067362) as part of the pangenome summary generated via Anvi'o: "anvi-summarize -p OM252/OM252pang-PAN.db -g OM252-GENOMES.db -C DEFAULT -o PAN_SUMMARY". Metabolic reconstruction was completed using the KEGG annotations from GhostKOALA and a custom set of HMMs

deployed with the KEGG-decoder, KEGG-expander, and Order_Decode_and_Expand scripts used previously (48). HMM searches for this workflow were completed using HMMER3.1b1 (49). Gene function enrichments based on annotations and pan-genome distribution were also calculated with Anvi'o. Using the phylogenomic clade structure (see below) that was also supported by ANI values, the subclades were imported into the Anvi'o database as layers using anvi-import-misc-data. Functional enrichments were then quantified for all the various annotation sources via the following command "for i in COG_CATEGORY Hamap ProSiteProfiles KeggGhostKoala SMART Gene3D TIGRFAM COG_FUNCTION SFLD PANTHER Coils CDD Pfam MobiDBLite ProSitePatterns PIRSF PRINTS SUPERFAMILY ProDom; do anvi-get-enriched-functions-per-pan-group -p OM252/OM252pang-PAN.db -g OM252-GENOMES.db --category subclade --annotation-source $i -o $i.enriched-subclade.txt; done". ProgressiveMauve v2.4.0 (50) was used to align the HIMB30 and LSUCC0096 genomes using default settings and genbank files supplied from IMG. The OM252 Anvi'o pangenomic summary, including all annotations, is available in Table S1 (https://doi.org/10.6084/m9.figshare.14067362). The enriched function files are available at https://doi.org/10.6084/m9.figshare.14036579 and the ProgressiveMauve alignments are available https://doi.org/10.6084/m9.figshare.14036588.

*Metagenomic read recruitment and analyses*

Competitive recruitment of the metagenomic reads from Tara Oceans (2), BIOGEOTRACES (51), the Malaspina Global Expedition (52), the Southern California Bight near Los Angeles (53), the San Francisco Bay Estuary (with permission - C. A. Francis), and the northern Gulf of Mexico hypoxic zone (38) to the OM252 genomes was completed using the protocol available at http://merenlab.org/data/tara-oceans-mags/. Reads were cleaned using illumina utils v2.6 (54) implementing the method described in (55). Mapping used Bowtie 2 v2.3.2 (56), processing with SAMtools v0.1.19-44428cd (57), and read filtering with BamM v1.7.3 (http://ecogenomics.github.io/BamM/) to include only recruited hits with an identity of at least 95% and alignment length of at least 75%. The count table for each sample was generated using the get_count_table.py script from (https://github.com/edamame-course/Metagenome). Reads per kilobase per million (RPKM) calculations were performed using RPKM_heater (https://github.com/thrash-lab/rpkm_heater) and $\log_{10}$-transformed to improve visualization of recruitment across wide variations in abundance. RPKM calculations are available in Table S1 (https://doi.org/10.6084/m9.figshare.14067362). Visualization of the data for individual genome recruitment was completed in R (https://github.com/thrash-lab/metaG_plots). OM252 community diversity was assessed using the skbio.diversity algorithmic suite v0.5.6 (http://scikit-bio.org/docs/latest/diversity.html). Recruited OM252 reads from 588 metagenomic samples including TARA (2), BIOGEOTRACES (51), and Malaspina Global Expedition (52) were normalized to transcripts per million (TPM) values and used analogously for count dissimilarities. TPM calculations were performed using (https://github.com/thrash-lab/counts_to_tpm). The □-diversity algorithm (within http://scikit-

bio.org/docs/latest/diversity.html) was retrofitted to interpret phylogenetic relationships using weighted-unifrac distances in place of the traditional Bray-Curtis dissimilarity. Retrofitting was performed *via* (https://github.com/thrash-lab/diversity_metrics). Sampling metadata for latitude, ocean region, depth, salinity, and temperature were collected to qualitatively assess the dissimilarity matrix in relation to designated intervals. ANOSIM correlation statistics were calculated for each metadata analysis treating absolute similarity of each OM252 community between samples as the null hypothesis to the alternative where community recruitment varies strongly with environmental metadata, such that: ANOSIM = 0 ≅ absolute similarity ≅ ubiquitous and even distribution across samples, and ANOSIM = 1 ≅ absolute dissimilarity ≅ highly varied distribution strongly related to an environmental factor (e.g., temperature). Three-dimensional ordination plots were constructed to visualize the principal coordinate analysis (PCoA) along the three foremost axes. The PCoA plots for latitude, ocean region, depth, salinity, and temperature are available in Fig. S6.

*RuBisCO phylogeny*

The predicted large subunit for both the LSUCC0096 and HIMB30 RuBisCO genes were searched against ncbi nt via the web (March 2019). The top 100 hits from each were screened for redundant sequences and combined with the 8 additional near-full-length homologs in the OM252 group identified via Anvi'o (anvi-get-sequences-for-gene-clusters -p OM252/OM252pang-PAN.db -g OM252-GENOMES.db --gene-cluster-id GC_00001582 -o GC_00001582.faa) and a set of RuBisCO reference type genes (58). Genes were aligned with MUSCLE v3.6 (29), culled with TrimAl v1.4.rev22 (30) using the -automated1 flag, and the final alignment was inferred with IQ-TREE v1.6.11 (31) with default settings and -bb 1000 for ultrafast bootstrapping (32). The final tree was visualized with Archaeopteryx (34). The fasta file, script, and tree file are provided at https://doi.org/10.6084/m9.figshare.14036597.

*Growth experiments*

Cell concentrations for all physiological experiments were measured via a Guava EasyCyte 5HT flow cytometer (Millipore, Massachusetts, USA) as previously reported (10, 59), and cells were grown in acid-washed polycarbonate flasks. The growth temperature range was tested in the isolation medium, JW1 (10), in triplicate, at 4˚C, 12˚C, 25˚C, 30˚C, 35˚C, and 40 ˚C; using a refrigerator, Isotemp cooling incubator (Fisher), and benchtop heating incubators (Fisher) (all non-shaking). Salinity tolerance was tested using two different methods. First, we only altered the concentration of NaCl in the JW1 medium from 0-5%, producing a range of salinities from 8.66 to 63.5 (calculated from chlorinity according to S ‰ = 1.80655 Cl ‰ (60)). In the second method, we altered the concentration of all major ions proportionally, producing a salinity range of 0.36 to 34.8 as previously reported (61). In both approaches, all other media components (carbon, iron, phosphate, nitrogen, vitamins, and trace metals) were unaltered. All salinity growth experiments were conducted in triplicate and at room temperature.

We tested whether LSUCC0096 could grow under chemolithoautotrophic conditions by growing cells in base JW1 medium with 100 µM thiosulfate and all organic carbon sources excluded (aside from that possibly obtained via vitamins) for four consecutive growth cycles to eliminate the possibility of carryover from the seed culture grown in JW1. As a positive control, LSUCC0096 was grown in JW1 medium with the normal suite of carbon compounds (10). For a negative control, LSUCC0096 was grown in JW1 media with no added organic carbon aside from vitamins. During the fourth and final growth cycle, a second negative control was added where organic carbon and vitamins were both excluded. Cells for each condition were grown in triplicate. Cultures were counted once a day for the first two cycles to ensure transfers could be completed at the end of log phase. For the third and fourth growth cycles, cells were counted every 12 hours. The fourth growth cycle is depicted in Figure 5. Strain purity and identity were verified at the end of experiments using PCR of the 16S rRNA gene as previously reported (10). Growth rates for all experiments were calculated with the sparse-growth-curve script (https://github.com/thrash-lab/sparse-growth-curve). In brief, for each individual growth curve (cell density vs. time), the exponential phase is extracted and cell densities are transformed in natural log. Linear regression is performed on the ln(cell density) vs. time in the exponential phase. The specific growth rate is therefore the slope of the linear regression result. The doubling rate equals the specific growth rate divided by ln(2). The sparse-growth-curve script can do the exponential phase extractions automatically. It calculates the numerical differentiations (Δ ln(Cell density)/ Δdt) between two time points as the instantaneous growth rates. The different phases (lag, exponential, and stationary) are differentiated by performing decision tree regression. The exponential phase is the period with maximum orders of magnitudes converted to cell densities.

*Electron microscopy*

To preserve cells for microscopy, we fixed 100mL of mid-exponential LSUCC0096 culture with 3% glutaraldehyde (Sigma Aldrich) and stored at 4˚C overnight. We filtered the cells onto a 25mm diameter 0.2µm pore-sized isopore polycarbonate membrane filter (MilliporeSigma) via vacuum filtration and performed ethanol dehydration by soaking the filter for 25 minutes in each the following ethanol concentrations: 30%, 50%, 70%, 80%, 90%, 95%, and 100%. The filter was then put into a Tousimis 815 critical point dryer and sputter coated for 45s in a Cressington 108 Manual Sputter Coater. Cells were imaged using the JSM-7001F-LV scanning electron microscope at the University of Southern California Core Center of Excellence in Nano Imaging (http://cemma.usc.edu/) with a working distance of 6.8mm and 15.0kV.

*Energetics*

Overall Gibbs energies, $\Delta G_r$, of thiosulfate and organic carbon oxidation,

$$S_2O_3^{2-} + 2O_2 + H_2O \rightarrow 2SO_4^{2-} + 2H^+ \quad (1)$$

and

$$C_{3.46}H_{7.5}O_{2.04}N_{0.84}S_{0.0073} + 3.7\ O_2 + 0.97\ H_2O \rightarrow 2.6\ H^+ + 3.46\ HCO_3^- + 0.0073\ HS^- + 0.84\ NH_4^+, \quad (2)$$

were calculated using

$$\Delta G_r = \Delta G_r^0 + RTlnQ_r \quad (3)$$

where $\Delta G_r^0$ and $Q_r$ refer to the standard molal Gibbs energy and the reaction quotient of the indicated reaction, respectively, $R$ represents the gas constant, and $T$ denotes temperature in Kelvin. Values of $\Delta G_r^0$ were calculated using the revised-HKF equations of state (62–64), the SUPCRT92 software package (65), and thermodynamic data taken from a number of sources (66–68). Values of $Q_r$ are calculated using

$$Q_r = \prod_i a_i^{\nu_i} \quad , \quad (4)$$

where $a_i$ stands for the activity of the $i$th species and $\nu_i$ corresponds to the stoichiometric coefficient of the $i$th species in the reaction of interest. Because standard states in thermodynamics specify a composition (69, 70) values of $Q_r$ must be calculated to take into account how environmental conditions impact overall Gibbs energies. In this study we use the classical chemical-thermodynamic standard state in which the activities of pure liquids are taken to be 1 as are those for aqueous species in a hypothetical 1 molal solution referenced to infinite dilution at any temperature or pressure.

Activities are related to concentration, $C$, by

$$a_i = \gamma_i \left(\frac{C_i}{C_i^{\theta}}\right) \quad (5)$$

where $\gamma_i$ and $C_i$ stand for the individual activity coefficient and concentration of the $i$th species, respectively, and $C_i^{\theta}$ refers to the concentration of the $i$th species under standard state conditions, which is taken to be equal to one molal referenced to infinite dilution. Values of $\gamma_i$ were computed using an extended version of the Debye-Hückel equation (71)). Concentration of the species shown in Reactions (1) and (2) are those used in the media ($[O_2]$ = saturation in seawater at 25°C (205 µmol); $[C_{org}]$ = 66.6 µM; $[HCO_3^-]$ = seawater (2 mmol); $[HS^-]$ = oxic seawater (0.5 nM)).

Because it is not clear which organic compounds are being oxidized for energy, we calculated values of $\Delta G_r$ for this process by representing DOC as a weighted average of all of the organic compounds in the media recipe, shown on the left side of reaction (2). This composite formula was used to calculate the standard state Gibbs energy of reaction (2), $\Delta G_r^0$, according to the algorithm given in LaRowe and Van Cappellen (2011), which relates the nominal oxidation of carbon, NOSC, in organics to their Gibbs energy of oxidation (the average weighted NOSC in the media is -0.26).

## Results

*Isolation and genome sequencing*

LSUCC0096 was isolated as part of a series of DTE experiments using water samples from across the southern Louisiana coast (10). The specific sample from which we obtained LSUCC0096 came from surface water in the Bay Pomme d'Or near the Mississippi River Birdfoot delta; salinity 26, 7.7˚C, pH 7.99. LSUCC0096 was grown for genome sequencing in JW1 medium (10). Illumina MiSeq PE 250bp sequencing generated 242,062 reads. Assembly with the A5 MiSeq pipeline resulted in 4 scaffolds with a total length of 1,935,310 bp, N50 of 1,442,657 bp, 30x median coverage, and a GC content of 48.5% (Table 1). Annotation by IMG predicted 2,001 protein-coding genes and 46 RNA genes- one copy of the 5S, 16S, and 23S rRNA genes and 36 predicted tRNA genes. The genome was estimated to be 96.17% complete with 0.37% contamination and a coding density of 95% (via CheckM (42), Table 1).

*Taxonomy*

Initial blast searches of the 16S rRNA gene sequence to GenBank identified LSUCC0096 as a gammaproteobacterium, with the closest cultivated representative being the OM252 clade organism HIMB30 (22). To better understand the phylogenetic breadth of this group we identified 23 non-redundant good or high-quality MAGs and SAGs closely related to HIMB30 and/or LSUCC0096 based on average nucleotide identity (ANI) and monophyletic grouping within the Genome Taxonomy Database (GTDB) (Table S1). Phylogenetic inference using 16S rRNA gene sequence phylogenies produced different results depending on taxon selection (Figs. S1 and S2). OM252 clade sequences branched sister to the *Litoricola* genus in the RefSeq tree (Fig. S1), but with substantial evolutionary distance between them. However, with added diversity contributed by clones and other non-RefSeq sequences, this relationship did not hold (Fig. S2). *Litoricola* branched in a completely different part of the tree, whereas the OM252 clone library sequence (U70703.1) remained in a monophyletic group containing the genomes in this study (Fig. S2).

To improve the placement of the OM252 clade within the Gammaproteobacteria and test the sister relationship with *Litoricola*, we created a phylogenomic tree using concatenated single copy marker genes from OM252 and other Gammaproteobacteria genomes selected based on the 16S rRNA gene trees. Consistent with the RefSeq 16S rRNA gene tree, the OM252 clade branched sister to *Litoricola*, which together were sister to the SAR86 clade (Figs. 1, S3). This group branched between the Moraxellaceae and the remainder of the newly recircumscribed Pseudomonadales Order in GTDB (Figs 1, S3). Even though the tree contains all the major families designated by GTDB in the Pseudomonadales, the relationship of the Litoricolaceae does not match their current topology (https://gtdb.ecogenomic.org/, accessed February 2021), which places Litoricolaceae sister to the Saccharospirillaceae and Oleiphilaceae. However, the bootstrap support for the relationships with SAR86 and the branch leading to the rest of the Pseudomonadales was poor (Figs. 1, S3). Subclade structure within the OM252 clade (subclades I and II, Fig. 1) corresponded to subgroups circumscribed via FastANI during our taxon selection

(Table S1). Within subclade I, a monophyletic group of genomes (UBA12265 to UBA8357 - Fig. 1) represented a single species according to the 95% ANI species cutoff (39), with multiple additional species in both subclades (Table S1). The isolates HIMB30 and LSUCC0096 shared only 80.3% ANI, making them distinct species.

Pairwise blast of the 16S rRNA gene from all nine OM252 subclade I members for which the gene was recovered with five *Litoricola* representatives from multiple species (Figs. S1, S2) corroborated the phylogenetic separation of these groups: no *Litoricola* sequence had greater than 89.8% identity with any OM252 genome, whereas the range of identity within OM252 subclade I was ≥ 98.5% (Table S1; no legitimate 16S rRNA genes were recovered from subclade II). Thus, OM252 subclade I constitutes a distinct genus from *Litoricola* based on pairwise 16S rRNA gene identity alone (72). The monophyletic relationship of subclade I and II in the phylogenomic tree, the presence of multiple distinct species within both subclades based on ANI, as well as the comparative branch length distances between subclades I and II vs. *Litoricola*, support inclusion of subclade I and II into the same group. Finally, there is a considerable difference in the GC content of all OM252 genomes (both subclades) compared with *Litoricola* (47-51% vs. 58-60% (26, 73), respectively). Thus, we propose the provisional genus name *Candidatus* Halomarinus for the OM252 clade. Since HIMB30 was the first reported isolate from OM252, this would be the type strain. However, since it is not currently deposited in international culture collections, we propose the species name as *Candidatus* Halomarinus kaneohensis, sp. nov.. We also propose *Candidatus* Halomarinus pommedorensis, sp. nov., for strain LSUCC0096, and *Candidatus* Halomarinus estuariensis, sp. nov., for the species cluster comprising UBA12265 to UBA8357 in Figure 1. We provide genus and species descriptions below.

*Distribution*

We previously reported the distribution of the OM252 clade within the 16S rRNA gene amplicon data associated with three years of sampling in support of DTE experiments from the Louisiana coast (12). The single OM252 OTU was moderately abundant (relative abundance up to ~1%) at salinities > 5, regardless of site. We also identified two amplicon sequence variants (ASVs) associated with the OM252 clade- one which was generally much more abundant than the other. LSUCC0096 matched the more abundant, and more frequently cultivated, ASV5512 (12), representative of subclade I. ASV5512 was found across a range of salinities, but was more prevalent in salinities above 12, where it was one of the top 50 most abundant ASVs in the three-year dataset.

We expanded our assessment of OM252 genome abundance and distribution using metagenomic read recruitment for the global oceans. OM252 members from both subclades recruited reads

from metagenomic samples across the globe (Fig. S4). The two most abundant taxa were represented by the subclade II MAGs TOBG-NAT-109 and GCA_002480175 (Fig. S5). The two most abundant subclade I taxa were represented by the SAGs AG-905-C17 and AG-900-B21. Genomes from the isolate strains LSUCC0096 and HIMB30 were the third and sixth ranking by median recruitment values for subclade I. Recruitment to the *Ca.* H. estuariensis cluster genomes was generally lower than to other genomes in subclade I (Fig. S5). Assessment of recruitment-based abundance patterns to the entire OM252 clade revealed no strong relationships with latitude, salinity, temperature, region, or depth (Fig. S6).

Individual genome recruitment was not influenced by salinity, but the salinity variation in the tested samples was quite limited (Fig. S7). Some genomes showed trends consistent with recruitment based on temperature (Fig. S8). For example, HIMB30 and LSUCC0096 RPKMs had significant negative relationships with temperature (linear regression, P-values 0.00168 and 0.00289, respectively). However, there was no consistent pattern for the genomes within a given subclade, and many genomes had no significant recruitment relationship with temperature (Fig. S8). The vast majority of samples from the dataset were from the epipelagic, and recruitment to OM252 genomes predominated in surface waters (Fig. S9). We observed very high relative recruitment to HIMB30 and AG-898-O07 in bathy- and abyssopelagic waters, and intermediate relative recruitment in deep ocean samples to other genomes from both subclades (Fig. S9), suggesting that some strains of OM252 may either preferentially or transiently inhabit the ocean interior. We also examined latitudinal distribution in recruitment. The dataset had a bimodal distribution with the majority of sites occurring in the mid-latitudes (Fig. S10). Most genomes did not show recruitment patterns consistent with the sample distribution alone, nor did we observe any clear relationships between subclade genomes with latitude (Fig. S11).

Separately, we recruited metagenomes from two coastal and one estuarine site- namely, the northern Gulf of Mexico “dead zone” (38); samples from the San Pedro shelf, basin and Catalina Island in the Southern California Bight (53); and a transect of samples from the San Francisco Bay estuary (Fig. S12). In contrast to the recruitment in the global oceans, the *Ca.* H. estuariensis cluster of genomes were the most abundant across these coastal/estuarine sites (Fig. S13), and in particular, were the highest recruiting members in the most saline sites of the San Francisco Bay estuary (Fig. S14). Subclade II was nearly absent in that same system. Thus, we hypothesize that the *Ca.* H. estuariensis group represents an estuarine-adapted species within OM252. Within the coastal/estuarine dataset, subclade I members generally showed increasing relative abundance with decreasing salinities, whereas subclade II members showed the opposite trends, although the data was sparse (Fig. S15).

*General genome characteristics*

The shared and variable gene content and corresponding metabolic functions of the OM252 genomes are shown in Table 1. Estimated genome completion spanned 51.6-96.2%, with

LSUCC0096 the most complete. Median estimated complete genome size was 2.21 Mbp, median GC content was 49% (47-51%), and median coding density was 92% (82-96%) (Table S1). Genome sizes are comparable to the *Litoricola lipolytica* IMCC1097 genome, the nearest phylogenomic neighbor. However, the IMCC1097 genome has a much higher GC content (58.8%) than that of *Ca.* Halomarinus.

*Electron transport and energy conservation*

*Ca.* Halomarinus bacteria were predicted to be aerobic chemotrophs (Fig. 2, Table S1). No alternative terminal electron accepting processes were identified in either subclade (Fig. S16). A second, high-affinity (*cbb*$_3$-type) cytochrome c oxidase was additionally present in seven subclade I genomes (Fig. 2, Table S1). Both subclades had sodium-translocating respiratory NADH dehydrogenases and Na+/H+ F-type ATPases, indicating the likely use of a sodium-motive force (Fig. 2). Most *Ca.* Halomarinus genomes contained proteorhodopsin (18/25) and retinal biosynthesis (19/25), with one notable exception being LSUCC0096. The proteorhodopsin gene in HIMB30 is located in an indel region with neighboring sections syntenic to the second largest contig in the LSUCC0096 genome (see ProgressiveMauve alignment in Supplementary Information). Therefore, it appears that the gene is truly absent from LSUCC0096 rather than missing as a result of the genome being incomplete.

*Carbon*

Both subclades had predicted genes for the Entner-Doudoroff pathway, the TCA cycle, gluconeogenesis, most of the genes of the pentose-phosphate pathway, and fructokinase (*scrK*) for fructose utilization (Fig. 2, Table S1). Six subclade I genomes, including HIMB30 and LSUCC0096, had an annotated mannose-6-phosphate isomerase for mannose utilization. No genome had an annotated phosphofructokinase gene for glycolysis through the Embden-Meyerhof-Parnas pathway. The *coxMSL* aerobic carbon monoxide dehydrogenase genes originally reported in HIMB30 (22) were conserved among most genomes (e.g., *coxL* in 16/25) in *Ca.* Halomarinus, except LSUCC0096 (Table S1). Similar to proteorhodopsin, this deletion in LSUCC0096 occured with flanking regions of conservation to the HIMB30 genome and were not near any contig boundaries, making it likely this is a true gene deletion in the LSUCC0096 genome. Subclade I, but not II, had predicted genes for the glyoxylate bypass. Eighteen genomes also contained a predicted beta-N-acetylhexosaminidase (Fig. 2), a glycoside hydrolase of the CAZyme GH-20 family that may confer chitin-degradation capabilities on *Ca.* Halomarinus bacteria (74, 75).

Ten of the *Ca.* Halomarinus genomes in subclade I had a ribulose-1,5-bisphosphate carboxylase (RuBisCO) gene and the associated Calvin-Benson-Bassham pathway for carbon fixation (Fig. 2), making these organisms predicted facultative autotrophs. Phylogenetic analysis of the large RuBisCO subunit demonstrated that all were Type I RuBisCO genes; however, the LSUCC0096 large subunit grouped away from that of HIMB30 and the other *Ca.* Halomarinus genomes for

which a sequence was recovered (Fig. S17). The LSUCC0096 RuBisCO genes were located directly upstream of likely *cbbQ* and *cbbO* activase genes (76), whereas the HIMB30 RuBisCO genes were located upstream from a suite of alpha-carboxysome genes (*csoS2, csoSCA, ccmL, ccmK*). Although we found carboxysome genes in other *Ca.* Halomarinus genomes (Fig. 2), none were annotated in the LSUCC0096 genome. Conversely, the LSUCC0096 *cbbQ* and *cbbO* genes had no matching orthologs in any of the other *Ca.* Halomarinus genomes. Thus, the LSUCC0096 RuBisCO is in a unique gene neighborhood and likely has a separate evolutionary history from the other *Ca.* Halomarinus RuBisCO genes.

Multiple pathways for phosphoglycolate salvage have recently been investigated in the model chemolithoautotrophic organism *Cupriavidas necator* H16 (77). We found annotated genes supporting the presence of the C2 cycle (*glyA*, *hprA*, and associated aminotransferases) and the malate cycle (*aceB*, *maeB*, pyruvate dehydrogenase *aceEF*) in *Ca.* Halomarinus genomes, but genes for the oxalyl-CoA decarboxylation route, as well as the *gcl* gene for the glycerate pathway, appear to be missing.

*N, P, and S*

*Ca.* Halomarinus uses the PII nitrogen response system and 24 of 25 genomes had the *amtB* ammonia transporter (14 genomes had two copies), with ten genomes containing complete urea transporter genes (*urtABCDE*), and others with partial transporters (Fig. 2, Table S1). Urease alpha, beta, and gamma subunit genes were conserved in HIMB30 and LSUCC0096 and five other genomes across both subclades (Fig. 2, Table S1), with partial urease genes found in more genomes. Complete or partial phosphate transporter genes (*pstABC*) were conserved across 17 genomes, and ten in both subclades were predicted to transport phosphonate (*phnCDE*) as well (Fig. 2). However, the *phn* C-P lyase genes were present exclusively in a subset of seven subclade I genomes. Both subclades had predicted genes for sulfide oxidation, and the sulfite dehydrogenase had variable distribution across the subclades as well (Fig. 2). We also found *sox* genes for thiosulfate oxidation in both subclades, with the exception of the species cluster *Ca.* H. estuariensis in subclade I (Fig. 2), and all genomes contained at least one copy of a sulfite exporter (*tauE*) (Table S1). Thus, subclades I and II were predicted to carry out sulfide- and thiosulfate-based chemolithotrophy. DMSP demethylation and synthesis genes were missing from all genomes, although three genomes had predicted DMSP lyases (Fig. 2).

*Other features*

The majority of *Ca.* Halomarinus genomes contained biosynthesis pathways for the bulk of essential amino acids, but none of the genomes contained genes for phenylalanine biosynthesis (Fig. 2). Thus, it appears this auxotrophy is conserved across the clade. Branched-chain and polar amino acid ABC transporters were present in the majority of genomes, as was a glycine betaine/proline ABC transporter (Table S1). B vitamin biosynthesis was limited. Thiamin (B1) and riboflavin (B2) biosynthesis pathways were partially complete in genomes from both

subclades (Fig. 2). Most genomes had predicted *ribBAHE* genes for riboflavin synthesis from ribulose-5P. Fourteen genomes contained the *thiXYZ* transporter for hydroxymethylpyrimidine (HMP) and 24 and 23 contained *thiD* and *thiE*, respectively (Fig. 2, Table S1). Thus, *Ca.* Halomarinus may synthesize thiamin from imported HMP. *Ca.* Halomarinus genomes appeared auxotrophic for biotin (B7), but possessed the biotin transporter component *bioY*. These organisms additionally had only partial pathways for pantothenate (B5), pyridoxine (B6), and folate (B9). No genes were present for nicotinamide/nicotinate (B3) biosynthesis, although NAD+ biosynthesis was intact. LSUCC0096, the most complete genome, was the only genome with a predicted *btuB* transporter component for cobalamin (B12). Most also contained genes for transport of ferric iron (*afuABC* - 23/25), copper (*copA* - 14/25), tungstate (*tupABC* - 15/25), zinc (*znuABC* - 12/25), and chromate (*chrA* - 12/25) (Fig. 2, Table S1). A small subset of genomes, including HIMB30 and LSUCC0096, contained all the genes for molybdate (*modABC* - 4/25), and iron complex transport (*fhuDBC* - 2/25), although all but three genomes had a predicted *fhuC* (Fig. 2, Table S1). None of the genomes with ureases contained annotated *nik* transporters for nickel, despite it being the required cofactor for urease. Thus, nickel may be obtained by promiscuous activity from one of the other ABC transporters in the genome, or they may be mis-annotated (78).

Eighteen genomes from both subclades had genes for flagellar biosynthesis, and so we predict *Ca.* Halomarinus cells to be motile (Fig. 2). Consistent with a sodium-motive force in OM252 cells, many of genomes contained sodium symporters for phosphate (8/25), acetate (15/25), and melibiose (25/25), as well as sodium antiporters for calcium (*yrbG* - 23/25) and protons (*nhaA* - 15/25) (Fig. 2, Table S1). The latter may provide a useful means for converting the proton motive force generated by proteorhodopsin to a sodium motive force in some strains. There was also a proton-chloride antiporter (*clcA*) conserved in 23 genomes. We found peroxiredoxin in 21 genomes and catalase (*katG*) in six genomes spanning both subclades as well (Fig. 2, Table S1). Finally, almost all genomes had *phbBC* genes to synthesize (and degrade) poly-ß-hydroxybutyrate (PHB), and two associated phasin genes were found in many genomes as well.

*Morphology and growth characteristics of LSUCC0096*

Cells of LSUCC0096 were curved-rod/spirillum-shaped, approximately 1.5 µm long and 0.2-0.3 µm wide (Figs. 3, S18). We also found evidence of a flagellum (Figs. 3 inset, S18BD), corroborating genomic predictions (above). Coastal Louisiana experiences dramatic shifts in salinity owing to a large number of estuaries in the region and tidal forcing through barrier islands, marshes, and different delta formations (79). Our previous 16S rRNA gene data suggested that OM252, and specifically the ASV5512 that matched LSUCC0096, had a euryhaline lifestyle, being found across a range of salinities but having greater prevalence in salinities above 12 (12). Therefore, we examined the salinity tolerance of LSUCC0096 through two complimentary methods: first, by altering only the concentration of NaCl in the medium, and second, by changing all major ion concentrations proportionally (Fig. 4A). LSUCC0096 grew in

salinities between 5.79 to 63.5, with a maximum growth rate of 0.23 (+/- 0.01) doublings/hour at 11.6 under the proportional scheme. We detected no growth at salinities of 0.36 or 1.45 (Fig. S19A). Although there was an overlap of salinities from 8.66 to 34.8 between the two experiments, the growth rates were higher when the ion concentrations were altered proportionally compared to when only the concentration of NaCl was altered (Fig. 4A). LSUCC0096 also grew between 12˚C and 35˚C in the isolation medium JW1 with a maximum growth rate of 0.36 (+/- 0.06) doublings/hour at 35˚C and a "typical" growth rate of 0.19 (+/- 0.03) at 25˚C (Fig. 4B). We did not detect growth at 4˚C or 40˚C (Fig. S19B).

*Thiosulfate-dependent chemolithoautotrophic growth*

We tested the ability of LSUCC0096 to grow under chemolithoautotrophic conditions with thiosulfate as the sole electron donor. We measured growth of LSUCC0096 across four consecutive transfers in modified JW1 medium with no added organic carbon (other than trace quantities of vitamins) and 100 µM thiosulfate. Inorganic carbon was present as bicarbonate (10 mM), used as the medium buffer (10). Growth curves from the fourth growth cycle are presented in Figure 5. When grown under strict chemolithoautotrophic conditions, LSUCC0096 increased in cell density more than two orders of magnitude in a typical logarithmic growth pattern, albeit more slowly, and to a lower cell density, than when grown in chemoorganoheterotrophic conditions (Fig. 5). The positive controls from the fourth transfer had an average growth rate of 0.20 +/- 0.01 doublings/hour, which is similar to growth rates found under normal growth conditions (Fig. 4), whereas the experimental replicates had a much slower average growth rate of 0.07 +/- 0.01 doublings/hour (Fig. S19C). Growth yields under chemolithoautotrophic conditions were roughly 68% of that under chemoorganoheterotrophic conditions (1.26 +/- 0.85 x $10^6$ vs. 1.85 +/- 1.29 x $10^6$ cells/ml), although the variance overlapped. The overall Gibbs energies, $\Delta G_r$, of organic carbon and thiosulfate oxidation under the experimental conditions were -113.2 kJ (mol $e^-$)$^{-1}$ and -100.9 kJ (mol $e^-$)$^{-1}$, respectively. We observed limited growth in the negative controls, but these were inconsistent and at much lower rates than for the experimental conditions (Fig. S19C). It is possible that storage compounds like PHB were not fully exhausted after four successive transfers, thus supplying the necessary energy and carbon for limited additional growth (80). Nevertheless, in conjunction with the genomic data, our experimental results provide strong evidence that LSUCC0096 is capable of oxidizing thiosulfate as a facultative chemolithoautotroph.

## Discussion

We comprehensively examined the distribution, genomic diversity, and taxonomy of OM252 bacterioplankton using 25 genomes from two pure cultures (including our recently isolated strain LSUCC0096), 7 SAGs, and 16 MAGs. These organisms were generally characterized by genomes in the 2.2 Mb range, with ~ 49% GC content and coding densities ~ 92%, although the two most complete genomes, from isolates HIMB30 and LSUCC0096, had 94 and 95% coding

density, respectively. Thus, OM252 genomes are slightly larger and less streamlined than SAR11 genomes, but smaller than most Roseobacter spp. (81–83).

Images of strain LSUCC0096 indicate these cells are curved-rod/spirillum-shaped (Figs. 3, S18) and somewhat larger than typical SAR11 cells (84). LSUCC0096 cells were also narrower and longer than *Litoricola marina* and *Litoricola lipolytica*, which were described as short rods, with no mention of curvature (26, 73). It remains to be seen if LSUCC0096 morphology is conserved throughout the OM252 clade. *L. marina* and *L. lipolytica* were also reported to be non-motile (26, 73), whereas OM252 genomes contain flagellar genes and we found evidence of a polar flagellum in LSUCC0096 (Figs. 3 inset, S18BD).

Phylogenetics with 16S rRNA genes and concatenated single-copy marker genes, as well as ANI comparisons, corroborated the sister relationship of OM252 with the genus *Litoricola*, and defined two major subclades and several species boundaries within OM252. Currently, HIMB30 and several MAGs used in this study are classified as a *Litoricola* species in the Genome Taxonomy Database (GTDB). However, based on the depth of branching between *Litoricola* and OM252 within our trees, the ANI values among OM252 genomes, the pairwise 16S rRNA gene identities within the group and between OM252 and *Litoricola* (72), and the substantial difference in GC content between OM252 and *Litoricola* (47-51% vs. 58-60% (26, 73), respectively), we argue here for distinguishing OM252 as a separate genus, which we propose as *Candidatus* Halomarinus, gen. nov.. Our whole genome phylogeny is consistent with the current placement of *Ca.* Halomarinus within the Litoricolaceae, and for the Litoricolaceae within the Pseudomonadales, as currently defined in GTDB (https://gtdb.ecogenomic.org/, accessed February 2021). Poor branch support at the internal nodes grouping Litoricolaceae with SAR86, and that group within the remainder of the Pseudomonadales, precludes us from commenting on the likely position of Litoricolaceae in that Order (Figs. 1, S3).

*Ca.* Halomarinus bacteria are globally distributed in marine surface waters, and some strains can be found in bathy- and abyssopelagic depths. The *Ca.* H. estuarensis species cluster was predominant in the San Francisco Bay estuary (Fig. S14), but recruited poorly from open ocean samples (Fig. S5). Thus, certain species may have more restricted biogeography than others. Increases in taxon selection and additional coastal and estuarine metagenomic sampling will improve these types of assessments. We demonstrated that strain LSUCC0096 can grow over a wide range of salinities (5.77-63.6), although it appears to be adapted for brackish conditions with an optimal growth salinity of 11.6 (Fig. 4A). Our coastal 16S rRNA gene data previously demonstrated that the ASV matching LSUCC0096 had maximum abundances in salinities between 12-21 (12) and recruitment data from the San Francisco Bay Estuary system (Fig. S12) supports the hypothesis that LSUCC0096 represents a brackish water specialist, but is incapable of growth in fresh water (Fig. S18). Nevertheless, this organism also recruited reads from all over the globe (Fig. S4) and displayed considerable halotolerance. Although we measured

growth at up to 5% NaCl (calculated salinity 63.5), we did not actually find the maximum salinity beyond which the cells could not grow (Fig. 4A). The extensive salinity tolerance of strain LSUCC0096 corroborates culture-independent detection of the OM252 clade in very salty environments like the Salton Sea (18) and Spanish salterns (19), as well as their cosmopolitan distribution in the global oceans (Fig. S4). Future work on additional *Ca.* Halomarinus isolates will expand our understanding of the halotolerance and optimal salinities for the various other species in the group. Thermal tolerances were more pedestrian, with LSUCC0096 exhibiting a mesophilic temperature growth range.

Comparative genomics predicted that *Ca.* Halomarinus spp. are obligate aerobes with the capacity for both chemoorganoheterotrophic and sulfur-oxidizing chemolithotrophic metabolism. In support of these predictions, both existing isolates, strains HIMB30 and LSUCC0096, were isolated under aerobic, chemoorganoheterotrophic growth conditions (10, 22). We predict that *Ca.* Halomarinus spp. can utilize TCA cycle intermediates, some sugars, and possibly amino acids as carbon and energy sources, although direct characterization of the suite of compounds that can be used needs further investigation. Given the possibility for utilization of the storage compound PHB, either as both an energy and carbon source or as an energy source in conjunction with RuBisCO-based carbon fixation (80), future experiments would require a $CO_2$-free headspace, an alternative buffer to the bicarbonate used in JW1, and probably five or more successive growth cycles to eliminate storage compounds.

*Ca.* Halomarinus appears to subsist on ammonia or urea as nitrogen sources, with phosphate as the primary source of phosphorous, and phosphonates as possible substitutes for some strains, similarly to SAR11 (81, 85). However, many Pelagibacterales lack the PII system (86) that is present in OM252. *Ca.* Halomarinus spp. can likely synthesize most amino acids except phenylalanine. We predicted B vitamin synthesis is limited to riboflavin and thiamin via *thiDE* after import of HMP. Thiamin biosynthesis after HMP import is also similar to the Pelagibacterales (87), with the major difference being the presence of the *thiXYZ* HMP transporter in OM252. We also predict *Ca.* Halomarinus spp. utilize ferric iron and may additionally interact with copper, tungstate, zinc, and chromate. Thus, they have a similarly restricted set of metal transporters as SAR11, and far fewer than many Roseobacter spp. (88).

Most *Ca.* Halomarinus genomes had predicted genes for oxidation of reduced sulfur compounds, and subclade I organisms additionally had genes for autotrophy via the CBB cycle and RuBisCO. Corroborating these predictions, we demonstrated that the subclade I representative LSUCC0096 could grow for successive transfers under strict chemolithoautotrophic conditions with thiosulfate as the sole electron donor and bicarbonate as the sole available carbon source (excepting vitamins). The ability to switch between autotrophic and heterotrophic metabolism also has biogeochemical relevance because it means that these organisms can switch between serving as inorganic carbon sources and sinks. This behavior has implications for modeling

marine carbon cycling since these organisms cannot be simply classified as heterotrophic. The pervasiveness of facultative lithoautotrophy among putative "heterotrophic" lineages deserves further investigation.

The relevance of facultative lithoautotrophy to both the carbon and sulfur cycles also places emphasis on understanding what may control these different lifestyles in nature. The experimental data from strain LSUCC0096 (growth rates and lag times after repeated transfers) suggest heterotrophic growth will always be favored to lithoautotrophic growth in *Ca*. Halomarinus subclade I strains, and calculations indicate that this probably arises due to the energetics of anabolism rather than catabolism. The energy available from organic carbon oxidation was only about 12% greater than thiosulfate oxidation, yet the growth rate was nearly three times greater in the heterotrophic experiment. This divergence could be explained by the much larger difference in the energetics of biomolecule synthesis when the starting materials are inorganic compounds such as $CO_2$ and $NH_4^+$ versus a suite of organic compounds. The energetics of protein synthesis provides an illustrative example since bacterial cells are approximately 50% protein (89). If an environment is replete with amino acids, then microorganisms need only to obtain and polymerize them to build proteins. The Gibbs energy of peptide bond formation is ~40 kJ per $mol^{-1}$ (90). However, if the organisms must first synthesize amino acids *de novo* before polymerization, the cost is much greater. For instance, the $\Delta G_r$ of alanine synthesis from $CO_2$ and $NH_4^+$ in an oxidizing environment is 1,380 kJ per $mol^{-1}$, and for more complex amino acids such tryptophan, it is 5,320 kJ per $mol^{-1}$ (91). Therefore, the LSUCC0096 cells in the thiosulfate experiment had to dedicate a larger *flux* of their catabolic energy to biomolecule synthesis than the heterotrophs who were essentially given all of the starting materials. These conclusions suggest that thiosulfate-based chemolithoautotrophy is utilized in nature when organic compound concentrations become limiting. However, more research is required to understand whether there is a complete metabolic switch to autotrophy, and if so, whether it is strictly controlled by the relative availability of growth substrates or if some additional regulation is involved.

Another intriguing mystery is the source and temporal availability of thiosulfate and other reduced sulfur compounds in natural marine systems. Abiotic oxidation of sulfide results in the generation of stable thiosulfate in seawater, an effect that was enhanced by the presence of trace metals like Fe, Pb, and Cu (92). Thus, any source of sulfide could theoretically lead to production of thiosulfate if the sulfide is not first consumed by other microbes. Alternatively, thiosulfate may occur as a transient intermediate as a direct result of microbial metabolism, a process labeled the "thiosulfate shunt" in sedimentary systems (93). If it could escape into the oxic water column from systems with low or no oxygen, organisms like *Ca.* Halomarinus may be able to harvest energy from thiosulfate originating from cryptic sulfur cycling processes near OMZs and sinking particles (94–96), or possibly also shallow sediments, where thiosulfate can persist in micromolar concentrations (97, 98). Researchers have been isolating thiosulfate-

utilizing bacteria from seawater for over a century (99), and it has long been known that thiosulfate-oxidizers occupy the marine water column and can connect this metabolism to carbon-fixation activity (100–102). More recently, we have also learned that *sox* genes are common in the oxic marine water column (94, 103, 104). Indeed, many marine prokaryotes contain these genes and thiosulfate oxidation can be used to stimulate growth under mixotrophic conditions and even anapleurotic carbon-fixation (see, for example, (105–108)). Thus, a variety of microorganisms from oxic marine waters are poised for thiosulfate-based chemolithotrophy and sometimes autotrophy. The circumstances and controls on reduced inorganic sulfur compound use by obligate aerobes like *Ca.* Halomarinuin the oxic water column requires further study.

**Description of *Halomarinus*, gen. nov.**

*Halomarinus* (Ha.lo.ma.ri.nus G. masc. n. *halo* salt, sea; L. masc. adj. *marinus*, of the sea; N.L. masc. n. *Halomarinus* salty, seagoing, in reference to the marine habitat and high salinity tolerance of the organisms).

Aerobic, chemoorganoheterotrophic and chemolithotrophic, with sodium-translocating NADH dehydrogenases, capable of glycolysis, gluconeogenesis, and possessing a complete TCA cycle. Has genes for motility via flagella. Possesses the PII-dependent nitrogen response system and genes for ammonia, phosphate, ferric iron, tungstate, copper, zinc, chromate transport. Has genes for synthesizing histidine, arginine, lysine, serine, threonine, glutamine, cysteine glycine, proline, methionine, isoleucine, leucine, tryptophan, tyrosine aspartate, glutamate, but are phenylalanine auxotrophs. Genes for synthesis of riboflavin (vitamin B2) and thiamine (vitamin B1) from HMP. Auxotrophic for vitamins B3, B5, B6, B7, B9, B12. Has genes for poly-ß-hydroxybutyrate production and degradation and peroxiredoxin. Estimated complete genome sizes between 1.49 - 2.68 Mbp, GC content between 47 - 51%, coding densities between 82-96%.

The type species is *Candidatus* Halomarinus kaneohensis.

**Description of *Candidatus* Halomarinus kaneohensis, sp. nov.**

*Candidatus* Halomarinus kaneohensis (ka.ne.o.hen.sis N.L. n. *Kāne*‘*ohe*, a bay on the island of Oahu, HI, USA, from where the strain was isolated).

In addition to the characteristics for the genus, it has the following features. Has proteorhodopsin and retinal biosynthesis genes. Has a predicted *cbb*$_3$-type cytochrome c oxidase, genes for the glyoxylate shunt, urease, and the *coxMSL* aerobic carbon monoxide dehydrogenase genes. Is predicted to be capable of thiosulfate and sulfide oxidation, as well as autotrophy via the Calvin-Benson-Bassham cycle.

The type strain, HIMB30$^T$, was isolated from seawater collected in Kāneʻohe Bay, Oahu, HI, U.S.A. (21.460467, -157.787657) (22). The genome sequence for HIMB30$^T$ is available under NCBI BioProject PRJNA47035. Estimated complete genome size of 2.26 Mbp, GC content of 50% from genome sequencing. The culture is maintained in cryostocks at the University of Hawaiʻi at Mānoa by M.S. Rappé. We provide the *Candidatus* designation since the culture has not been deposited in two international culture collections and therefore does not satisfy the naming conventions of the International Code of Nomenclature for Prokaryotes (ICNP) (109). However, the characterization here is more than sufficient for naming recognition via genomic type material (110, 111).

**Description of *Candidatus* Halomarinus pommedorensis, sp. nov.**

*Candidatus* Halomarinus pommedorensis (pomme.d.or.en.sis N.L. n. *Pomme d'Or*, a bay in southern Louisiana, U.S.A., from where the strain was isolated).

In addition to the characteristics for the genus, it has the following features. Cells are curved-rod/spirillum-shaped, ~ 1.5 µm x 0.2-0.3 µm. Halotolerant, being capable of growth in salinities between 5.8 and at least 63.4, but not at 1.5 or below. Mesophilic, being capable of growth at temperatures between 12 and 35˚C but not at 4 or 40˚C. Has a maximum growth rate at 35˚C in the isolation medium JW1 of 0.36 (+/- 0.06) doublings/hour. Has a predicted *cbb*$_3$-type cytochrome c oxidase and genes for the glyoxylate shunt, urease, and cobalamin transport. Has predicted genes for thiosulfate and sulfide oxidation, as well as autotrophy via the Calvin-Benson-Bassham cycle. Grows under thiosulfate-oxidizing chemolithoautotrophic conditions at 0.07 (+/- 0.01) doublings/hour.

The type strain, LSUCC0096$^T$, was isolated from seawater collected at Bay Pomme d'Or, Buras, LA, U.S.A. (29.348784, -89.538171) (10). The GenBank accession number for the 16S rRNA gene of LSUCC0096$^T$ is KU382366.1. The genome sequence is available under BioProject PRJNA551315. The culture is maintained in cryostocks at the University of Southern California by J.C. Thrash. Estimated complete genome size of 2.01 Mbp, GC content of 49% from genome sequencing. We provide the *Candidatus* designation since the culture has not been deposited in two international culture collections and therefore does not satisfy the naming conventions of the ICNP (109). However, the characterization here is more than sufficient for naming recognition via genomic type material (110, 111).

**Description of *Candidatus* Halomarinus estuariensis, sp. nov.**

*Candidatus* Halomarinus estuariensis (es.tu.ar.i.en.sis L. masc. adj. *estuarine*, based on its relative abundance in an estuary).

In addition to the characteristics for the genus, it has the following features. Has proteorhodopsin and retinal biosynthesis genes. Has an additional *cbb*$_3$-type cytochrome c oxidase. Is predicted to

be capable of sulfide oxidation, as well as autotrophy via the Calvin-Benson-Bassham cycle, but not thiosulfate oxidation. Has predicted genes for the glyoxylate shunt, as well as D-galacturonate epimerase. Some strains have C-P lyase and DMSP lyase genes. Estimated complete genome sizes between 1.98 and 2.68 Mbp, GC content between 47 - 49% from genome sequencing.

We provide the Candidatus designation since this species has not yet been cultivated. Genomes were reconstructed from metagenomic sequencing.

**Data Availability**

The LSUCC0096 genome is available on NCBI under BioProject number PRJNA551315, and IMG under Taxon ID 2639762503. The raw reads from which LSUCC0096 was assembled are available at the NCBI SRA under accession number SRR9598636. Metagenomic sequences from the San Francisco Bay estuary were used with permission from Dr. Christopher A. Francis and are available on NCBI under the following SRA accessions: SRR7130817, SRR7130819, SRR7130820, SRR7130903, SRR7131305, SRR7131306, SRR7132116, SRR7132117. Cryostocks and/or live cultures of strains LSUCC0096 and HIMB30 are available upon request.

**Acknowledgements**

This work was supported by a Louisiana Board of Regents grant [LEQSF(2014-17)-RD-A-06], a Simons Early Career Investigator in Marine Microbial Ecology and Evolution Award, and NSF Biological Oceanography Program grants (OCE-1747681 and OCE-1945279) to JCT. The authors also thank Dr. Christopher A. Francis for use of the San Francisco Bay metagenomic dataset.

## Table 1. Genome characteristics

| Bin | SC | Est. Compln (%) | Est. Contam (%) | Strain het (%) | Genome size (bp) | Est. genome size (bp) | N50 scaffolds | Scaffolds | Longest scaffold | GC fract. | Coding density fract. | Type | Source | Study |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AG-359-J14 | I | 63.37 | 0 | 0 | 1319993 | 2082994 | 41035 | 55 | 132304 | 0.486 | 0.918 | SAG | BATS | Pachiadaki et al. 2019 |
| AG-900-B21 | I | 53.83 | 0 | 0 | 1259049 | 2338936 | 36080 | 65 | 106558 | 0.486 | 0.909 | SAG | BATS | Pachiadaki et al. 2019 |
| AG-918-E15 | I | 71.08 | 0.37 | 0 | 1671916 | 2352161 | 62227 | 52 | 215445 | 0.488 | 0.927 | SAG | BATS | Pachiadaki et al. 2019 |
| TOBG-MED-814 | I | 90.2 | 6.51 | 84 | 1967603 | 2181378 | 39535 | 74 | 144762 | 0.498 | 0.937 | MAG | Meditterene an Sea | Tully et al. 2018 |
| AG-898-O07 | I | 56.73 | 0 | 0 | 1067362 | 1881477 | 29177 | 60 | 110676 | 0.486 | 0.917 | SAG | BATS | Pachiadaki et al. 2019 |
| AG-915-K04 | I | 51.64 | 0 | 0 | 1290806 | 2499624 | 53286 | 58 | 157605 | 0.486 | 0.920 | SAG | BATS | Pachiadaki et al. 2019 |
| HIMB30 | I | 95.99 | 0.49 | 0 | 2168870 | 2259475 | 223986 | 36 | 638152 | 0.499 | 0.941 | Isolate | Kāne‘ohe Bay, HI | This study |
| AG-905-C17 | I | 68.02 | 0 | 0 | 1505956 | 2213990 | 89387 | 53 | 220750 | 0.483 | 0.911 | SAG | BATS | Pachiadaki et al. 2019 |
| TOBG-MED-759 | I | 89.74 | 3.21 | 93.33 | 1788781 | 1993293 | 84972 | 31 | 211681 | 0.506 | 0.940 | MAG | Meditterene an Sea | Tully et al. 2018 |
| LSUCC0096 | I | 96.17 | 0.37 | 0 | 1935310 | 2012384 | 1442657 | 4 | 1442657 | 0.487 | 0.952 | Isolate | Bay Pomme d'Or, LA | This study |
| UBA12265 | I | 70.95 | 1.91 | 25 | 1444999 | 2036644 | 8081 | 220 | 28363 | 0.510 | 0.856 | MAG | Adriatic Sea | Parks et al. 2017 |
| UBA9605 | I | 69.58 | 1.7 | 11.11 | 1221408 | 1755401 | 6540 | 213 | 25083 | 0.511 | 0.912 | MAG | Inner Oslofjord | Parks et al. 2017 |
| UBA9601 | I | 77.59 | 0 | 0 | 1349581 | 1739375 | 8473 | 186 | 32481 | 0.511 | 0.913 | MAG | Outer Oslofjord | Parks et al. 2017 |
| UBA11194 | I | 68.83 | 0.75 | 0 | 1212584 | 1761709 | 6292 | 216 | 24783 | 0.511 | 0.916 | MAG | Adriatic Sea | Parks et al. 2017 |
| TOBG-MED-626 | I | 80.2 | 2.1 | 73.33 | 1193441 | 1488081 | 23640 | 59 | 71869 | 0.509 | 0.958 | MAG | Meditterene an Sea | Tully et al. 2018 |
| AG-896-J04 | I | 70.54 | 0.46 | 0 | 1622917 | 2300705 | 96547 | 37 | 166779 | 0.492 | 0.923 | SAG | BATS | Pachiadaki et al. 2019 |
| UBA8357 | I | 65.76 | 1.23 | 40 | 1312486 | 1995873 | 8992 | 180 | 28717 | 0.510 | 0.823 | MAG | Sapelo Island, GA | Parks et al. 2017 |
| TOBG-RS-469 | II | 69.27 | 2.04 | 16.67 | 1799083 | 2597204 | 14438 | 138 | 66997 | 0.494 | 0.932 | MAG | Red Sea | Tully et al. 2018 |
| TOBG-SAT-133 | II | 75.37 | 1.19 | 0 | 1870654 | 2481961 | 27755 | 73 | 67858 | 0.491 | 0.929 | MAG | South Atlantic | Tully et al. 2018 |
| TOBG-NAT-109 | II | 85.36 | 2.9 | 7.69 | 2285316 | 2677268 | 25595 | 107 | 89471 | 0.487 | 0.925 | MAG | North Atlantic | Tully et al. 2018 |
| GCA_002480175.1 | II | 83.79 | 2.75 | 27.27 | 2207349 | 2634382 | 10539 | 273 | 37539 | 0.485 | 0.851 | MAG | South Western Atlantic | Parks et al. 2017 |
| TOBG-NP-1444 | II | 83.86 | 3.15 | 11.11 | 2099531 | 2503614 | 26349 | 109 | 70809 | 0.475 | 0.911 | MAG | North Pacific | Tully et al. 2018 |
| TOBG-SP-353 | II | 53.45 | 0 | 0 | 1212569 | 2268604 | 23702 | 51 | 50076 | 0.482 | 0.924 | MAG | South Pacific | Tully et al. 2018 |
| UBA11144 | II | 72.06 | 1.95 | 20 | 1667084 | 2313467 | 8361 | 246 | 24773 | 0.483 | 0.857 | MAG | South Eastern Pacific | Parks et al. 2017 |
| TOBG-NP-1472 | II | 76.68 | 1.79 | 0 | 1521562 | 1984301 | 19361 | 94 | 46684 | 0.488 | 0.929 | MAG | North Pacific | Tully et al. 2018 |

Genomes shaded according to subclade (SC)

## Figures

**Figure 1. Phylogenomic tree of the Pseudomonadales and OM252.** Maximum-likelihood tree based on 78 concatenated single-copy genes within the Pseudomonadales (as designated by GTDB) and selected other Gammaproteobacteria, with Alpha- and Betaproteobacteria outgroup. Final alignment = 29,631 amino acid positions. Families designated by GTDB within the Pseudomondales are indicated, with shading for the OM252 clade. Species designated in this study are highlighted in red and orange text. Values at nodes indicate bootstrap support (n=1000), scale indicated changes per position.

**Figure 2. Metabolic reconstruction of the OM252 clade.** Heatmap displays gene and pathway content according to the scale on the right. Subgroups of processes and key metabolic pathways are highlighted for ease of viewing. Subclades and species designations follow that in Figure 1.

**Figure 3. Scanning electron micrographs of LSUCC0096.** Main- 25,000x magnification of two cells on a 0.2 µm filter, scale bar = 1 µm. Inset- 40,000x magnification of a dividing cell to focus on possible polar flagellum in the upper pole, scale bar = 100 nm.

**Figure 4. Salinity and temperature growth ranges for LSUCC0096.** A) Specific growth rates and doubling times according to variable salinity based on proportional dilution of major ions (orange) or changing only NaCl concentration (blue) within the media. B) Specific growth rates and doubling times according to temperature.

**Figure 5. Thiosulfate-based chemolithoautotrophic growth in LSUCC0096.** Cell numbers plotted against time for growth in chemolithoautotrophic conditions with thiosulfate as the sole electron donor (green), compared with typical heterotrophic medium (yellow), and no carbon (orange) and no carbon/vitamin (blue) controls. Curves depict growth after four consecutive transfers.

## References Cited


1. Azam F, Fenchel T, Field G, Gray JS, Meyer-Reil LA, Thingstad F. 1983. The Ecological Role of Water-Column Microbes in the Sea. Vol 10:257–263.

2. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, d'Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmento H, Vieira-Silva S, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Tara Oceans coordinators, Bowler C, de Vargas C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S, Stemmann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG, Bork P. 2015. Ocean plankton. Structure and function of the global ocean microbiome. Science 348:1261359.

3. Arístegui J, Gasol JM, Duarte CM, Herndld GJ. 2009. Microbial oceanography of the dark ocean's pelagic realm. Limnol Oceanogr 54:1501–1529.

4. Giovannoni SJ. 2017. SAR11 Bacteria: The Most Abundant Plankton in the Oceans. Ann Rev Mar Sci 9:231–255.

5. Hoarfrost A, Nayfach S, Ladau J, Yooseph S, Arnosti C, Dupont CL, Pollard KS. 2019. Global ecotypes in the ubiquitous marine clade SAR86. ISME J https://doi.org/10.1038/s41396-019-0516-7.

6. Partensky F, Hess WR, Vaulot D. 1999. Prochlorococcus, a marine photosynthetic prokaryote of global significance. Microbiol Mol Biol Rev 63:106–127.

7. Dupont CL, Rusch DB, Yooseph S, Lombardo M-J, Richter RA, Valas R, Novotny M, Yee-Greenbaum J, Selengut JD, Haft DH, Halpern AL, Lasken RS, Nealson K, Friedman R, Venter JC. 2012. Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. ISME J 6:1186–1199.

8. Rappé MS, Kemp PF, Giovannoni SJ. 1997. Phylogenetic diversity of marine coastal picoplankton 16S rRNA genes cloned from the continental shelf off Cape Hatteras, North Carolina. Limnol Oceanogr 42:811–826.

9. Poretsky RS, Sun S, Mou X, Moran MA. 2010. Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. Environ Microbiol 12:616–627.

10. Henson MW, Pitre DM, Weckhorst JL, Lanclos VC, Webber AT, Thrash JC. 2016. Artificial Seawater Media Facilitate Cultivating Members of the Microbial Majority from the Gulf of Mexico. mSphere 1.

11. Redmond MC, Valentine DL. 2012. Natural gas and temperature structured a microbial community response to the Deepwater Horizon oil spill. Proc Natl Acad Sci U S A 109:20292–20297.

12. Henson MW, Lanclos VC, Pitre DM, Weckhorst JL, Lucchesi AM, Cheng C, Temperton B, Thrash JC. 2020. Expanding the Diversity of Bacterioplankton Isolates and Modeling Isolation Efficacy with Large-Scale Dilution-to-Extinction Cultivation. Appl Environ Microbiol 86.

13. Yeo SK, Huggett MJ, Eiler A, Rappé MS. 2013. Coastal bacterioplankton community dynamics in response to a natural disturbance. PLoS One 8:e56207.

14. Sintes E, Witte H, Stodderegger K, Steiner P, Herndl GJ. 2013. Temporal dynamics in the free-living bacterial community composition in the coastal North Sea. FEMS Microbiol Ecol 83:413–424.

15. Galand PE, Bourrain M, De Maistre E, Catala P, Desdevises Y, Elifantz H, Kirchman DL, Lebaron P. 2012. Phylogenetic and functional diversity of Bacteria and Archaea in a unique stratified lagoon, the Clipperton atoll (N Pacific). FEMS Microbiol Ecol 79:203–217.

16. Schäfer H, Servais P, Muyzer G. 2000. Successional changes in the genetic diversity of a marine bacterial assemblage during confinement. Arch Microbiol 173:138–145.

17. Zhang R, Wu Q, Piceno YM, Desantis TZ, Saunders FM, Andersen GL, Liu W-T. 2013. Diversity of bacterioplankton in contrasting Tibetan lakes revealed by high-density microarray and clone library analysis. FEMS Microbiol Ecol 86:277–287.

18. Dillon JG, McMath LM, Trout AL. 2009. Seasonal changes in bacterial diversity in the Salton Sea. Hydrobiologia 632:49–64.

19. Fernandez AB, Ghai R, Martin-Cuadrado AB, Sanchez-Porro C, Rodriguez-Valera F, Ventosa A. 2013. Metagenome sequencing of prokaryotic microbiota from two hypersaline ponds of a marine saltern in santa pola, Spain. Genome Announc 1.

20. Giles EC, Kamke J, Moitinho-Silva L, Taylor MW, Hentschel U, Ravasi T, Schmitt S. 2013. Bacterial community profiles in low microbial abundance sponges. FEMS Microbiol Ecol 83:232–241.

21. Dishaw LJ, Flores-Torres J, Lax S, Gemayel K, Leigh B, Melillo D, Mueller MG, Natale L, Zucchetti I, De Santis R, Pinto MR, Litman GW, Gilbert JA. 2014. The gut of geographically disparate Ciona intestinalis harbors a core microbiota. PLoS One 9:e93386.

22. Huggett MJ, Rappé MS. 2012. Genome sequence of strain HIMB30, a novel member of the marine Gammaproteobacteria. J Bacteriol 194:732–733.

23. Williams KP, Gillespie JJ, Sobral BWS, Nordberg EK, Snyder EE, Shallom JM, Dickerman AW. 2010. Phylogeny of gammaproteobacteria. J Bacteriol 192:2305–2314.

24. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol https://doi.org/10.1038/nbt.4229.

25. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, Hugenholtz P. 2020. A complete domain-to-species taxonomy for Bacteria and Archaea. Nat Biotechnol https://doi.org/10.1038/s41587-020-0501-8.

26. Kim H, Choo Y-J, Cho J-C. 2007. Litoricolaceae fam. nov., to include Litoricola lipolytica gen. nov., sp. nov., a marine bacterium belonging to the order Oceanospirillales. Int J Syst Evol Microbiol 57:1793–1798.

27. Coil D, Jospin G, Darling AE. 2015. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics 31:587–589.

28. Chen I-MA, Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, Huntemann M, Varghese N, White JR, Seshadri R, Smirnova T, Kirton E, Jungbluth SP, Woyke T, Eloe-Fadrosh EA, Ivanova NN, Kyrpides NC. 2018. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. Nucleic Acids Res https://doi.org/10.1093/nar/gky901.

29. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797.

30. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973.

31. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274.

32. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol 35:518–522.

33. Junier T, Zdobnov EM. 2010. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. Bioinformatics 26:1669–1670.

34. Han MV, Zmasek CM. 2009. phyloXML: XML for evolutionary biology and comparative genomics. BMC

Bioinformatics 10:356.

35. Delmont TO, Quince C, Shaiber A, Esen ÖC, Lee ST, Rappé MS, McLellan SL, Lücker S, Eren AM. 2018. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. Nat Microbiol 3:804–813.

36. Tully BJ, Graham ED, Heidelberg JF. 2018. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Sci Data 5:170203.

37. Thrash JC, Baker BJ, Seitz KW, Temperton B, Campbell LG, Rabalais NN, Henrissat B, Mason OU. 2018. Metagenomic Assembly and Prokaryotic Metagenome-Assembled Genome Sequences from the Northern Gulf of Mexico "Dead Zone." Microbiol Resour Announc 7:e01033–18.

38. Thrash JC, Seitz KW, Baker BJ, Temperton B, Gillies LE, Rabalais NN, Henrissat B, Mason OU. 2017. Metabolic Roles of Uncultivated Bacterioplankton Lineages in the Northern Gulf of Mexico "Dead Zone." MBio 8:e01017–17.

39. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114.

40. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol 2:1533–1542.

41. Pachiadaki MG, Brown JM, Brown J, Bezuidt O, Berube PM, Biller SJ, Poulton NJ, Burkart MD, La Clair JJ, Chisholm SW, Stepanauskas R. 2019. Charting the Complexity of the Marine Microbiome through Single-Cell Genomics. Cell 179:1623–1635.e11.

42. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055.

43. Ballesteros JA, Hormiga G. 2016. A New Orthology Assessment Method for Phylogenomic Data: Unrooted Phylogenetic Orthology. Mol Biol Evol 33:2117–2134.

44. Murat Eren A, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3:e1319.

45. Delmont TO, Eren AM. 2018. Linking pangenomes and metagenomes: the Prochlorococcus metapangenome. PeerJ 6:e4320.

46. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240.

47. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol 428:726–731.

48. Tully BJ. 2019. Metabolic diversity within the globally abundant Marine Group II Euryarchaea offers insight into ecological patterns. Nat Commun 10:271.

49. Eddy SR. 2011. Accelerated Profile HMM Searches. PLoS Comput Biol 7:e1002195.

50. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147.

51. Biller SJ, Berube PM, Dooley K, Williams M, Satinsky BM, Hackl T, Hogle SL, Coe A, Bergauer K, Bouman HA, Browning TJ, De Corte D, Hassler C, Hulston D, Jacquot JE, Maas EW, Reinthaler T, Sintes E, Yokokawa T, Chisholm SW. 2018. Marine microbial metagenomes sampled across space and time. Sci Data 5:180176.

52. Acinas SG, Sánchez P, Salazar G, Cornejo-Castillo FM, Sebastián M, Logares R, Sunagawa S, Hingamp P, Ogata H, Lima-Mendez G, Roux S, González JM, Arrieta JM, Alam IS, Kamau A, Bowler C, Raes J, Pesant S, Bork P, Agustí S, Gojobori T, Bajic V, Vaqué D, Sullivan MB, Pedrós-Alió C, Massana R, Duarte CM, Gasol JM. 2019. Metabolic Architecture of the Deep Ocean Microbiome. Cold Spring Harbor Laboratory.

53. Sieradzki ET, Cesar Ignacio-Espinoza J, Needham DM, Fichot EB, Fuhrman JA. 2019. Dynamic marine viral infections and major contribution to photosynthetic processes shown by spatiotemporal picoplankton metatranscriptomes. Nat Commun 10:1169.

54. Eren AM, Vineis JH, Morrison HG, Sogin ML. 2013. A filtering method to generate high quality short reads using illumina paired-end technology. PLoS One 8:e66643.

55. Minoche AE, Dohm JC, Himmelbauer H. 2011. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. Genome Biol 12:R112.

56. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.

57. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

58. Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC, Wilkins MJ, Hettich RL, Lipton MS, Williams KH, Long PE, Banfield JF. 2012. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. Science 337:1661–1665.

59. Thrash JC, Weckhorst JL, Pitre DM. 2015. Cultivating fastidious microbes, p. 57–78. *In* McGenity, TJ, Timmis, KN, Nogales, B (eds.), Hydrocarbon and Lipid Microbiology Protocols. Springer, Berlin, Germany.

60. Lewis E. 1980. The practical salinity scale 1978 and its antecedents. IEEE J Oceanic Eng 5:3–8.

61. Henson MW, Lanclos VC, Faircloth BC, Thrash JC. 2018. Cultivation and genomics of the first freshwater SAR11 (LD12) isolate. ISME J 12:1846–1860.

62. Helgeson HC, Kirkham DH, Flowers GC. 1981. Theoretical prediction of the thermodynamic behavior of aqueous electrolytes by high pressures and temperatures; IV, Calculation of activity coefficients, osmotic coefficients, and apparent molal and standard and relative partial molal properties to 600 degrees C and 5kb. Am J Sci 281:1249–1516.

63. Tanger JC, Helgeson HC. 1988. Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures; revised equations of state for the standard partial molal properties of ions and electrolytes. Am J Sci 288:19–98.

64. Shock EL, Oelkers EH, Johnson JW, Sverjensky DA, Helgeson HC. 1992. Calculation of the thermodynamic properties of aqueous species at high pressures and temperatures. Effective electrostatic radii, dissociation constants and standard partial molal properties to 1000 °C and 5 kbar. J Chem Soc Faraday Trans 88:803–826.

65. Johnson JW, Oelkers EH, Helgeson HC. 1992. SUPCRT92: A software package for calculating the standard molal thermodynamic properties of minerals, gases, aqueous species, and reactions from 1 to 5000 bar and 0 to 1000°C. Comput Geosci 18:899–947.

66. Shock EL, Helgeson HC. 1988. Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures: Correlation algorithms for ionic species and equation of state predictions to 5 kb and 1000°C. Geochim Cosmochim Acta 52:2009–2036.

67. Shock EL, Helgeson HC, Sverjensky DA. 1989. Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures: Standard partial molal properties of inorganic neutral species. Geochim Cosmochim Acta 53:2157–2183.

68. LaRowe DE, Van Cappellen P. 2011. Degradation of natural organic matter: A thermodynamic analysis. Geochim Cosmochim Acta 75:2030–2042.

69. LaRowe D, Amend J. 2020. Energy limits for life in the subsurface, p. 585–619. *In* B.N. Orcutt, I. Daniel, and R. Dasgupta (ed.), Whole Earth Carbon: Past to Present. Cambridge University Press, Cambridge.

70. Amend JP, LaRowe DE. 2019. Mini-Review: Demystifying Microbial Reaction Energetics. Environ Microbiol https://doi.org/10.1111/1462-2920.14778.

71. Helgeson HC. 1969. Thermodynamics of hydrothermal systems at elevated temperatures and pressures. Am J Sci 267:729–804.

72. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Microbiol 12:635–645.

73. Choi A, Lee K, Oh H-M, Feng J, Cho J-C. 2010. Litoricola marina sp. nov. Int J Syst Evol Microbiol 60:1303–1306.

74. Tews I, Perrakis A, Oppenheim A, Dauter Z, Wilson KS, Vorgias CE. 1996. Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease. Nat Struct Biol 3:638–648.

75. Kulik N, Slámová K, Ettrich R, Křen V. 2015. Computational study of β-N-acetylhexosaminidase from Talaromyces flavus, a glycosidase with high substrate flexibility. BMC Bioinformatics 16:28.

76. Tsai Y-CC, Lapina MC, Bhushan S, Mueller-Cajar O. 2015. Identification and characterization of multiple rubisco activases in chemoautotrophic bacteria. Nat Commun 6:8883.

77. Claassens NJ, Scarinci G, Fischer A, Flamholz AI, Newell W, Frielingsdorf S, Lenz O, Bar-Even A. 2020. Phosphoglycolate salvage in a chemolithoautotroph using the Calvin cycle. Proc Natl Acad Sci U S A https://doi.org/10.1073/pnas.2012288117.

78. Bennett BD, Redford KE, Gralnick JA. 2018. MgtE Homolog FicI Acts as a Secondary Ferrous Iron Importer in Shewanella oneidensis Strain MR-1. Appl Environ Microbiol 84.

79. Schroeder WW, Wiseman WJ. 1998. Geology and Hydrodynamics of Gulf of Mexico Estuaries, p. 3–27. *In* Bianchi, TS, Pennock, JR, Twilley, RR (eds.), Biogeochemistry of Gulf of Mexico Estuaries. John Wiley & Sons, Inc.

80. Anderson AJ, Dawes EA. 1990. Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. Microbiol Rev 54:450–472.

81. Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, Giovannoni SJ, Rappé MS. 2012. Streamlining and core genome conservation among highly divergent members of the SAR11 clade. MBio 3.

82. Luo H, Csuros M, Hughes AL, Moran MA. 2013. Evolution of divergent life history strategies in marine alphaproteobacteria. MBio 4.

83. Giovannoni SJ, Thrash JC, Temperton B. 2014. Implications of streamlining theory for microbial ecology. ISME J 8:1553–1565.

84. Zhao X, Schwartz CL, Pierson J, Giovannoni SJ, Richard McIntosh J, Nicastro D. 2017. Three-Dimensional Structure of the Ultraoligotrophic Marine Bacterium “Candidatus Pelagibacter ubique.” Appl Environ Microbiol 83:e02807–16.

85. Carini P, Steindler L, Beszteri S, Giovannoni SJ. 2013. Nutrient requirements for growth of the extreme oligotroph “Candidatus Pelagibacter ubique” HTCC1062 on a defined medium. ISME J 7:592–602.

86. Smith DP, Thrash JC, Nicora CD, Lipton MS, Burnum-Johnson KE, Carini P, Smith RD, Giovannoni SJ. 2013. Proteomic and transcriptomic analyses of "Candidatus Pelagibacter ubique" describe the first PII-independent response to nitrogen limitation in a free-living Alphaproteobacterium. MBio 4:e00133–12.

87. Carini P, Campbell EO, Morré J, Sañudo-Wilhelmy SA, Thrash JC, Bennett SE, Temperton B, Begley T, Giovannoni SJ. 2014. Discovery of a SAR11 growth requirement for thiamin's pyrimidine precursor and its distribution in the Sargasso Sea. ISME J 8:1727–1738.

88. Hogle SL, Thrash JC, Dupont CL, Barbeau KA. 2016. Trace Metal Acquisition by Marine Heterotrophic Bacterioplankton with Contrasting Trophic Strategies. Appl Environ Microbiol 82:1613–1624.

89. Neidhardt FC. 1987. Chemical composition of Escherichia coli, p. 3–6. *In* Neidhardt, FC (ed.), Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology. American Society for Microbiology, Washington, D.C.

90. Amend JP, LaRowe DE, McCollom TM, Shock EL. 2013. The energetics of organic synthesis inside and outside the cell. Philos Trans R Soc Lond B Biol Sci 368:20120255.

91. LaRowe DE, Amend JP. 2016. The energetics of anabolism in natural settings. ISME J 10:1285–1295.

92. Zhang J-Z, Millero FJ. 1993. The products from the oxidation of H2S in seawater. Geochim Cosmochim Acta 57:1705–1718.

93. Jørgensen BB. 1990. A thiosulfate shunt in the sulfur cycle of marine sediments. Science 249:152–154.

94. Canfield DE, Stewart FJ, Thamdrup B, De Brabandere L, Dalsgaard T, Delong EF, Revsbech NP, Ulloa O. 2010. A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean coast. Science 330:1375–1378.

95. Callbeck CM, Lavik G, Ferdelman TG, Fuchs B, Gruber-Vodicka HR, Hach PF, Littmann S, Schoffelen NJ, Kalvelage T, Thomsen S, Schunck H, Löscher CR, Schmitz RA, Kuypers MMM. 2018. Oxygen minimum zone cryptic sulfur cycling sustained by offshore transport of key sulfur oxidizing bacteria. Nat Commun 9:1729.

96. Raven MR, Keil RG, Webb SM. 2020. Microbial sulfate reduction and organic sulfur formation in sinking marine particles. Science https://doi.org/10.1126/science.abc6035.

97. Jørgensen BB, Bak F. 1991. Pathways and microbiology of thiosulfate transformations and sulfate reduction in a marine sediment (kattegat, denmark). Appl Environ Microbiol 57:847–856.

98. Zopfi J, Ferdelman TG, Fossing H. 2004. Distribution and fate of sulfur intermediates—sulfite, tetrathionate, thiosulfate, and elemental sulfur—in marine sediments, p. 97–116. *In* Sulfur Biogeochemistry - Past and Present. Geological Society of America.

99. Nathansohn A. 1902. Über eine neue Gruppe von Schwefelbakterien und ihren Stoffwechsel. Mittheilungen aus der Zoologischen Station zu Neapel 15:655–680.

100. Tuttle JH, Holmes PE, Jannasch HW. 1974. Growth rate stimulation of marine pseudomonads by thiosulfate. Arch Microbiol 99:1–14.

101. Tuttle JH, Jannasch HW. 1977. Thiosulfate stimulation of microbial dark assimilation of carbon dioxide in shallow marine waters. Microb Ecol 4:9–25.

102. Tuttle JH, Jannasch HW. 1972. Occurrence and types of Thiobacillus-like bacteria in the sea. Limnol Oceanogr 17:532–543.

103. Moran MA, Buchan A, González JM, Heidelberg JF, Whitman WB, Kiene RP, Henriksen JR, King GM, Belas R, Fuqua C, Brinkac L, Lewis M, Johri S, Weaver B, Pai G, Eisen JA, Rahe E, Sheldon WM, Ye W, Miller TR, Carlton J, Rasko DA, Paulsen IT, Ren Q, Daugherty SC, Deboy RT, Dodson RJ, Durkin AS,

Madupu R, Nelson WC, Sullivan SA, Rosovitz MJ, Haft DH, Selengut J, Ward N. 2004. Genome sequence of Silicibacter pomeroyi reveals adaptations to the marine environment. Nature 432:910–913.

104. Poretsky RS, Hewson I, Sun S, Allen AE, Zehr JP, Moran MA. 2009. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. Environ Microbiol 11:1358–1375.

105. Muthusamy S, Baltar F, González JM, Pinhassi J. 2014. Dynamics of metabolic activities and gene expression in the Roseobacter clade bacterium Phaeobacter sp. strain MED193 during growth with thiosulfate. Appl Environ Microbiol 80:6933–6942.

106. Kelly DP, Wood AP. 2013. The Chemolithotrophic Prokaryotes, p. 275–287. *In* Rosenberg, E, DeLong, EF, Lory, S, Stackebrandt, E, Thompson, F (eds.), The Prokaryotes: Prokaryotic Communities and Ecophysiology. Springer Berlin Heidelberg, Berlin, Heidelberg.

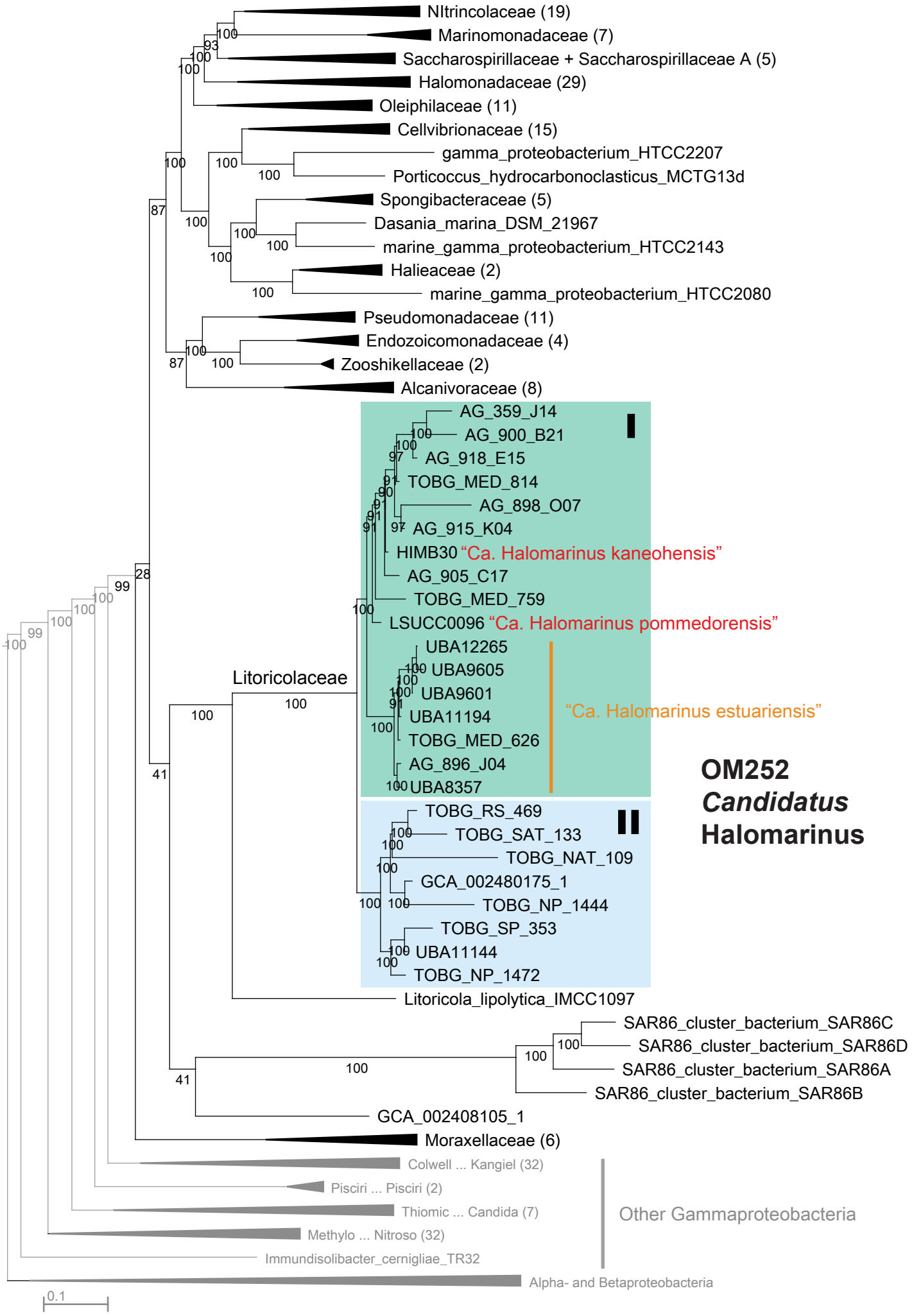
107. Ghosh W, Dam B. 2009. Biochemistry and molecular biology of lithotrophic sulfur oxidation by taxonomically and ecologically diverse bacteria and archaea. FEMS Microbiol Rev 33:999–1043.

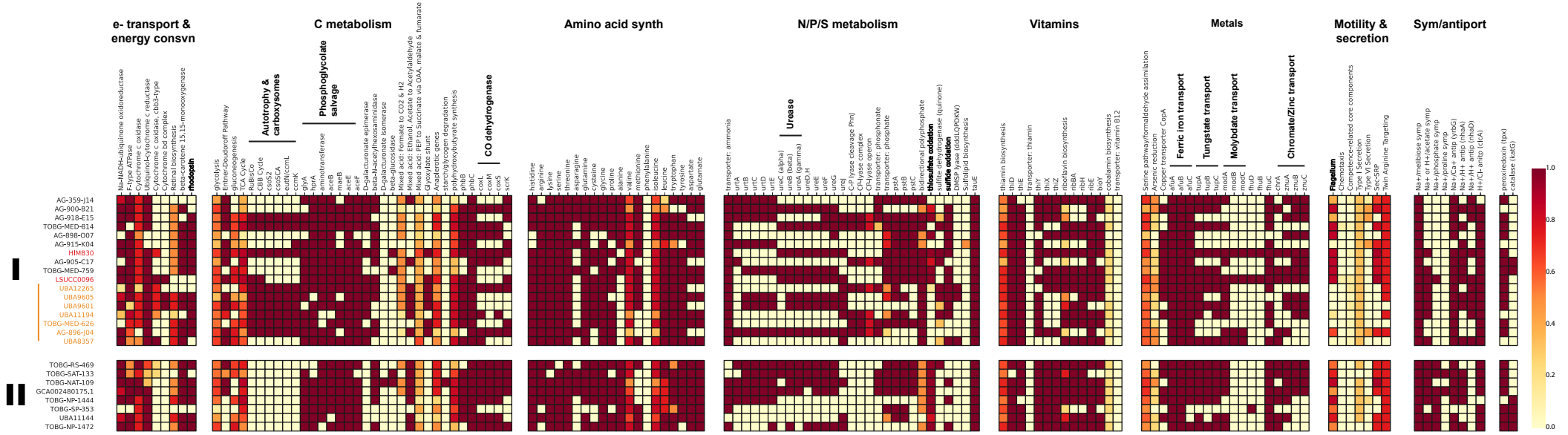
108. van Vliet DM, von Meijenfeldt FAB, Dutilh BE, Villanueva L, Sinninghe Damsté JS, Stams AJM, Sánchez-Andrea I. 2020. The bacterial sulfur cycle in expanding dysoxic and euxinic marine waters. Environ Microbiol https://doi.org/10.1111/1462-2920.15265.

109. Parker CT, Tindall BJ, Garrity GM. 2015. International Code of Nomenclature of Prokaryotes. Int J Syst Evol Microbiol https://doi.org/10.1099/ijsem.0.000778.

110. Murray AE, Freudenstein J, Gribaldo S, Hatzenpichler R, Hugenholtz P, Kämpfer P, Konstantinidis KT, Lane CE, Papke RT, Parks DH, Rossello-Mora R, Stott MB, Sutcliffe IC, Thrash JC, Venter SN, Whitman WB, Acinas SG, Amann RI, Anantharaman K, Armengaud J, Baker BJ, Barco RA, Bode HB, Boyd ES, Brady CL, Carini P, Chain PSG, Colman DR, DeAngelis KM, de Los Rios MA, Estrada-de Los Santos P, Dunlap CA, Eisen JA, Emerson D, Ettema TJG, Eveillard D, Girguis PR, Hentschel U, Hollibaugh JT, Hug LA, Inskeep WP, Ivanova EP, Klenk H-P, Li W-J, Lloyd KG, Löffler FE, Makhalanyane TP, Moser DP, Nunoura T, Palmer M, Parro V, Pedrós-Alió C, Probst AJ, Smits THM, Steen AD, Steenkamp ET, Spang A, Stewart FJ, Tiedje JM, Vandamme P, Wagner M, Wang F-P, Hedlund BP, Reysenbach A-L. 2020. Roadmap for naming uncultivated Archaea and Bacteria. Nat Microbiol 5:987–994.

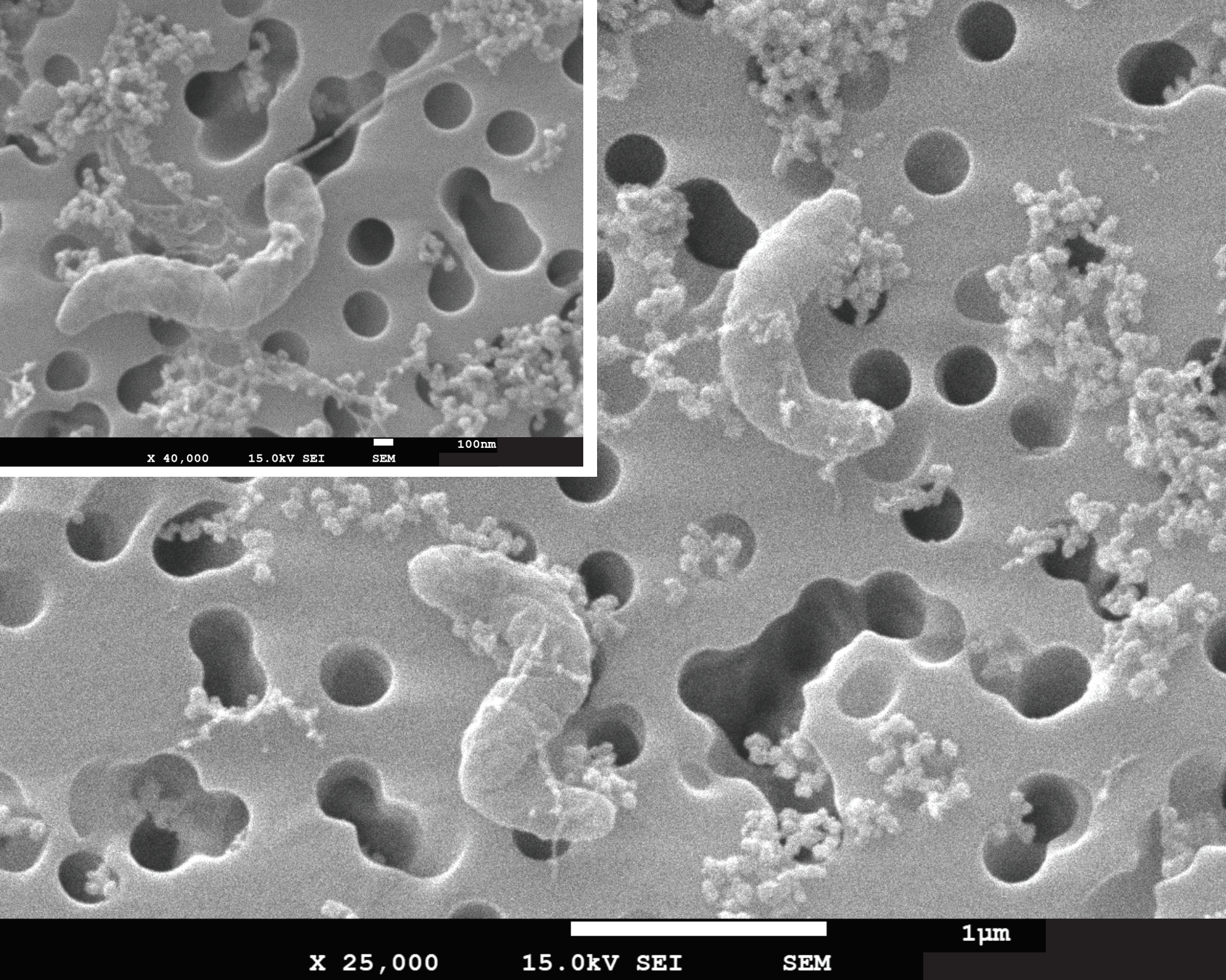
111. Whitman WB. 2015. Genome sequences as the type material for taxonomic descriptions of prokaryotes. Syst Appl Microbiol 38:217–222.

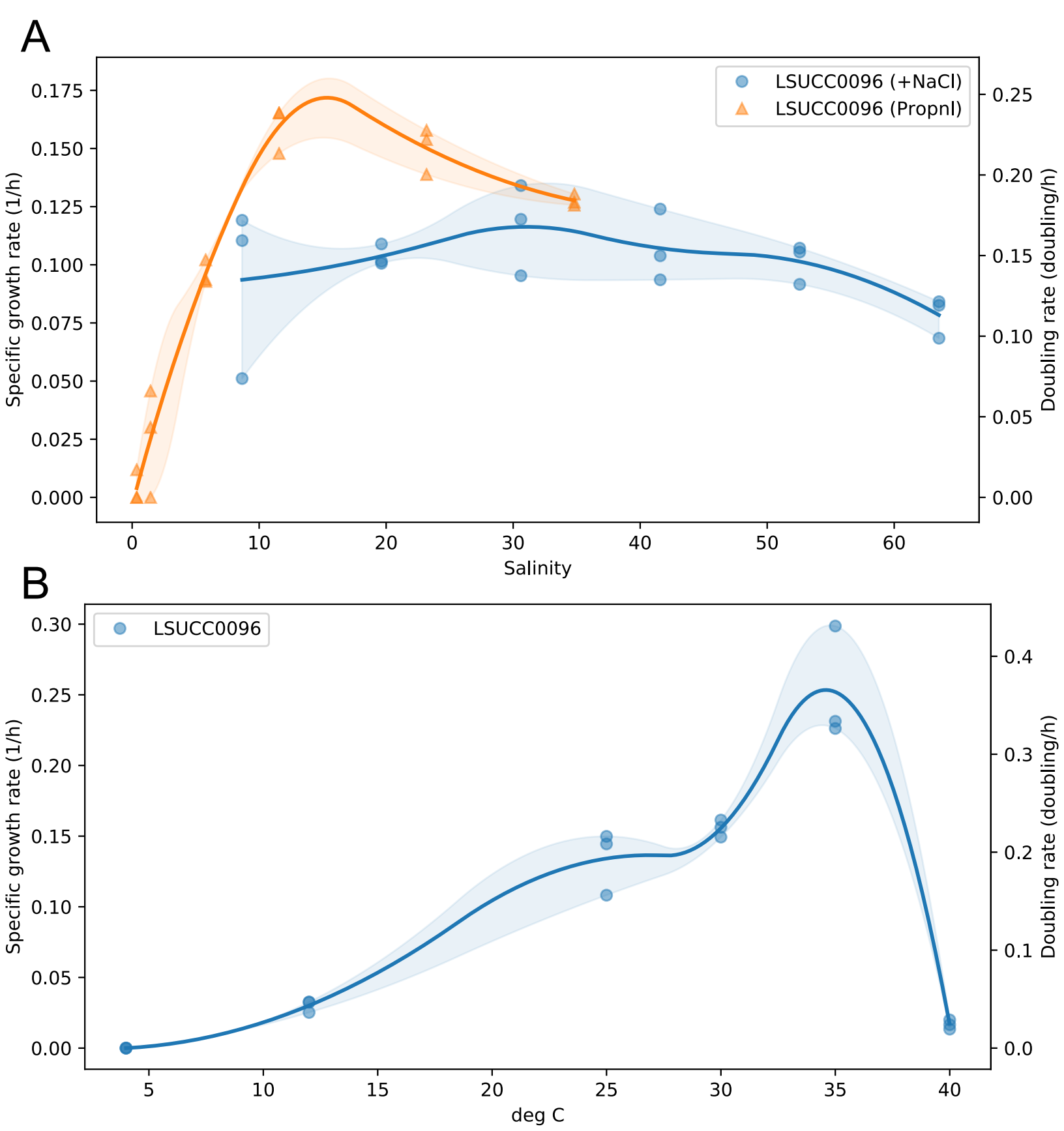

NItrincolaceae (19)
Marinomonadaceae (7)
Saccharospirillaceae + Saccharospirillaceae A (5)
Halomonadaceae (29)
Oleiphilaceae (11)
Cellvibrionaceae (15)
gamma_proteobacterium_HTCC2207
Porticoccus_hydrocarbonoclasticus_MCTG13d
Spongibacteraceae (5)
Dasania_marina_DSM_21967
marine_gamma_proteobacterium_HTCC2143
Halieaceae (2)
marine_gamma_proteobacterium_HTCC2080
Pseudomonadaceae (11)
Endozoicomonadaceae (4)
Zooshikellaceae (2)
Alcanivoraceae (8)
I
AG_359_J14
AG_900_B21
AG_918_E15
TOBG_MED_814
AG_898_O07
AG_915_K04
HIMB30 "Ca. Halomarinus kaneohensis"
AG_905_C17
TOBG_MED_759
LSUCC0096 "Ca. Halomarinus pommedorensis"
UBA12265
UBA9605
UBA9601
UBA11194
TOBG_MED_626
AG_896_J04
UBA8357
"Ca. Halomarinus estuariensis"
Litoricolaceae
II
TOBG_RS_469
TOBG_SAT_133
TOBG_NAT_109
GCA_002480175_1
TOBG_NP_1444
TOBG_SP_353
UBA11144
TOBG_NP_1472
Litoricola_lipolytica_IMCC1097
OM252
*Candidatus*
Halomarinus
SAR86_cluster_bacterium_SAR86C
SAR86_cluster_bacterium_SAR86D
SAR86_cluster_bacterium_SAR86A
SAR86_cluster_bacterium_SAR86B
GCA_002408105_1
Moraxellaceae (6)
Colwell ... Kangiel (32)
Pisciri ... Pisciri (2)
Thiomic ... Candida (7)
Methylo ... Nitroso (32)
Immundisolibacter_cernigliae_TR32
Other Gammaproteobacteria
Alpha- and Betaproteobacteria
0.1

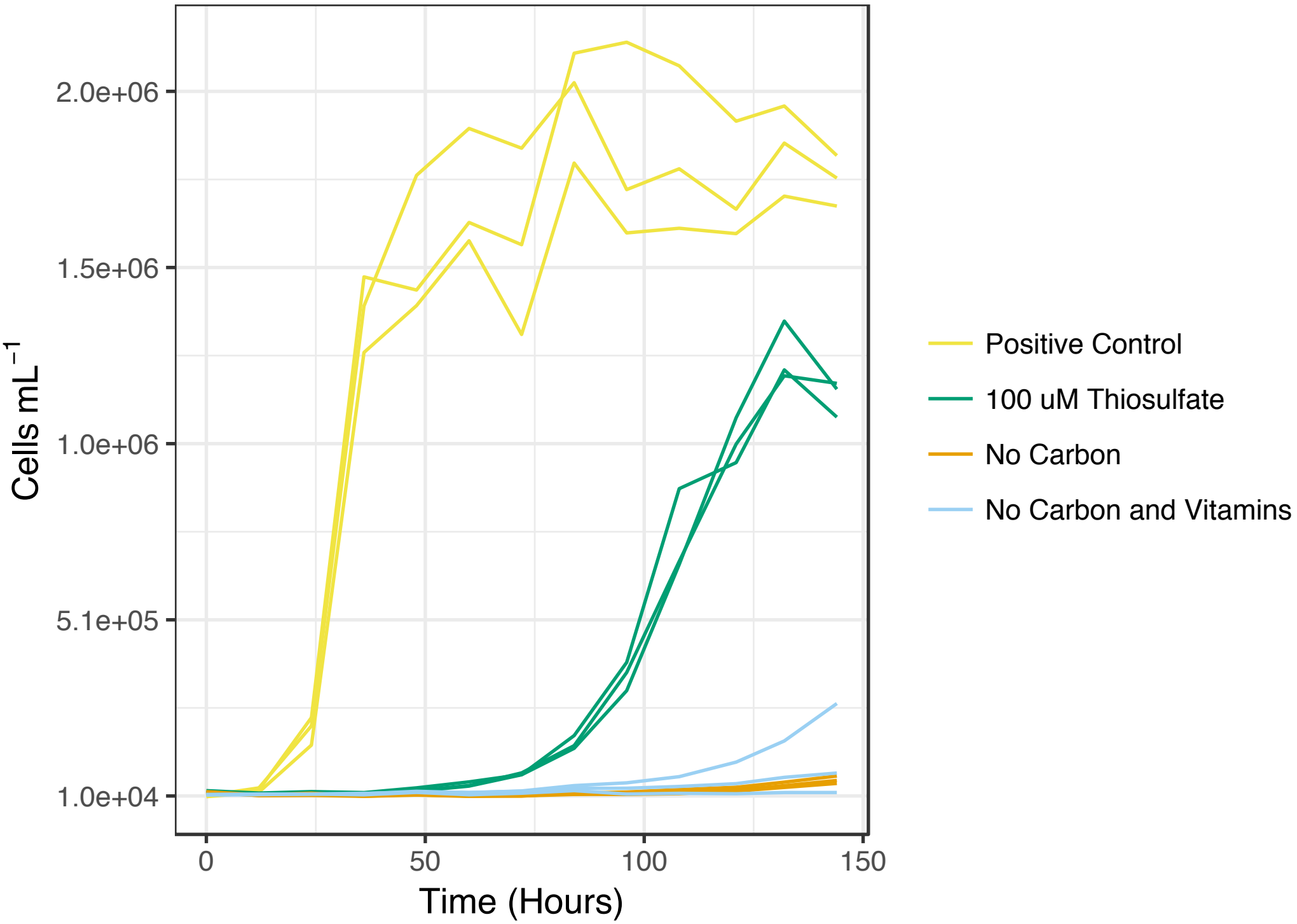


e- transport & energy consvn
C metabolism
Amino acid synth
N/P/S metabolism
Vitamins
Metals
Motility & secretion
Sym/antiport
Autotrophy & carboxysomes
Phosphoglycolate salvage
CO dehydrogenase
Urease
Ferric iron transport
Tungstate transport
Molybdate transport
Chromate/Zinc transport
Flagellum
Na-NADH-ubiquinone oxidoreductase
F-type ATPase
Cytochrome c oxidase
Ubiquinol-cytochrome c reductase
Cytochrome c oxidase, cbb3-type
Cytochrome bd complex
Retinal biosynthesis
beta-carotene 15,15-monooxygenase
rhodopsin
glycolysis
Entner-Doudoroff Pathway
gluconeogenesis
TCA Cycle
RuBisCo
CBB Cycle
cosS2
csoSCA
eutN/ccmL
ccmK
glyA
hprA
aminotransferase
aceB
maeB
aceE
aceF
D-galacturonate epimerase
beta-N-acetylhexosaminidase
D-galacturonate isomerase
beta-glucosidase
Mixed acid: Formate to CO2 & H2
Mixed acid: Ethanol, Acetate to Acetylaldehyde
Mixed acid: PEP to Succinate via OAA, malate & fumarate
Glyoxylate shunt
Anaplerotic genes
starch/glycogen degradation
polyhydroxybutyrate synthesis
phbB
phbC
coxL
coxM
coxS
scrK
histidine
arginine
lysine
serine
threonine
asparagine
glutamine
cysteine
glycine
proline
alanine
valine
methionine
phenylalanine
isoleucine
leucine
tryptophan
tyrosine
aspartate
glutamate
transporter: ammonia
urtA
urtB
urtC
urtD
urtE
ureC (alpha)
ureB (beta)
ureA (gamma)
ureD,H
ureE
ureF
ureG
ureJ
C-P lyase cleavage PhnJ
CP-lyase complex
CP-lyase operon
transporter: phosphonate
transporter: phosphate
psiA
pstB
pstC
bidirectional polyphosphate
thiosulfate oxidation
sulfite dehydrogenase (quinone)
sulfide oxidation
DMSP lyase (ddd,LQPDKW)
Sulfolipid biosynthesis
tauE
thiamin biosynthesis
thiD
thiE
transporter: thiamin
thiY
thiX
thiZ
riboflavin biosynthesis
ribBA
ribH
ribE
bioY
cobalamin biosynthesis
transporter: vitamin B12
Serine pathway/formaldehyde assimilation
Arsenic reduction
Copper transporter CopA
afuA
afuB
afuC
tupA
tupB
tupC
modA
modB
modC
fhuD
fhuB
fhuC
chrA
znuA
znuB
znuC
Chemotaxis
Competence-related core components
Type I Secretion
Type VI Secretion
Sec-SRP
Twin Arginine Targeting
Na+/melibiose symp
Na+ or H+/acetate symp
Na+/phosphate symp
Na+/proline symp
Na+/Ca+ antip (yrbG)
Na+/H+ antip (nhaA)
Na+/H+ antip (nhaD)
H+/Cl- antip (clcA)
peroxiredoxin (tpx)
catalase (katG)
AG-359-J14
AG-900-B21
AG-918-E15
TOBG-MED-814
AG-898-O07
AG-915-K04
HIMB30
AG-905-C17
TOBG-MED-759
LSUCC0096
UBA12265
UBA9605
UBA9601
UBA11194
TOBG-MED-626
AG-896-J04
UBA8357
TOBG-RS-469
TOBG-SAT-133
TOBG-NAT-109
GCA002480175.1
TOBG-NP-1444
TOBG-SP-353
UBA11144
TOBG-NP-1472
1.0
0.8
0.6
0.4
0.2
0.0

100nm
X 40,000 15.0kV SEI SEM


1µm
X 25,000 15.0kV SEI SEM


A
LSUCC0096 (+NaCl)
LSUCC0096 (PropnI)
Specific growth rate (1/h)
Doubling rate (doubling/h)
Salinity
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0.25
0.20
0.15
0.10
0.05
0.00
0
10
20
30
40
50
60
B
LSUCC0096
Specific growth rate (1/h)
Doubling rate (doubling/h)
deg C
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.4
0.3
0.2
0.1
0.0
5
10
15
20
25
30
35
40


2.0e+06
1.5e+06
1.0e+06
5.1e+05
1.0e+04
Cells mL$^{-1}$
0
50
100
150
Time (Hours)
Positive Control
100 uM Thiosulfate
No Carbon
No Carbon and Vitamins