

# UC Santa Barbara

## UC Santa Barbara Previously Published Works

### Title

Pontiella agarivorans sp. nov., a novel marine anaerobic bacterium capable of degrading macroalgal polysaccharides and fixing nitrogen.

### Permalink

<https://escholarship.org/uc/item/6418w1kb>

### Journal

Applied and Environmental Microbiology, 90(2)

### Authors

Liu, Na

Kivenson, Veronika

Peng, Xuefeng

et al.

### Publication Date

2024-02-21

### DOI

10.1128/aem.00914-23

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

# *Pontiella agarivorans* sp. nov., a novel marine anaerobic bacterium capable of degrading macroalgal polysaccharides and fixing nitrogen

Na Liu,<sup>1</sup> Veronika Kivenson,<sup>1</sup> Xuefeng Peng,<sup>2</sup> Zhisong Cui,<sup>3</sup> Thomas S. Lankiewicz,<sup>4</sup> Kelsey M. Gosselin,<sup>1</sup> Chance J. English,<sup>2,5</sup> Elaina M. Blair,<sup>4</sup> Michelle A. O'Malley,<sup>4,6</sup> David L. Valentine<sup>2,7</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 18.

**ABSTRACT** Marine macroalgae produce abundant and diverse polysaccharides, which contribute substantially to the organic matter exported to the deep ocean. Microbial degradation of these polysaccharides plays an important role in the turnover of macroalgal biomass. Various members of the *Planctomycetes-Verrucomicrobia-Chlamydia* (PVC) superphylum are degraders of polysaccharides in widespread anoxic environments. In this study, we isolated a novel anaerobic bacterial strain NLcol2<sup>T</sup> from microbial mats on the surface of marine sediments offshore Santa Barbara, CA, USA. Based on 16S ribosomal RNA (rRNA) gene and phylogenomic analyses, strain NLcol2<sup>T</sup> represents a novel species within the *Pontiella* genus in the *Kiritimatiellota* phylum (within the PVC superphylum). Strain NLcol2<sup>T</sup> is able to utilize various monosaccharides, disaccharides, and macroalgal polysaccharides such as agar and  $\iota$ -carrageenan. A near-complete genome also revealed an extensive metabolic capacity for anaerobic degradation of sulfated polysaccharides, as evidenced by 202 carbohydrate-active enzymes (CAZymes) and 165 sulfatases. Additionally, its ability of nitrogen fixation was confirmed by nitrogenase activity detected during growth on nitrogen-free medium, and the presence of nitrogenases (*nifDKH*) encoded in the genome. Based on the physiological and genomic analyses, this strain represents a new species of bacteria that may play an important role in the degradation of macroalgal polysaccharides and with relevance to the biogeochemical cycling of carbon, sulfur, and nitrogen in marine environments. Strain NLcol2<sup>T</sup> (= DSM 113125<sup>T</sup> = MCCC 1K08672<sup>T</sup>) is proposed to be the type strain of a novel species in the *Pontiella* genus, and the name *Pontiella agarivorans* sp. nov. is proposed.

**IMPORTANCE** Growth and intentional burial of marine macroalgae is being considered as a carbon dioxide reduction strategy but elicits concerns as to the fate and impacts of this macroalgal carbon in the ocean. Diverse heterotrophic microbial communities in the ocean specialize in these complex polymers such as carrageenan and fucoidan, for example, members of the *Kiritimatiellota* phylum. However, only four type strains within the phylum have been cultivated and characterized to date, and there is limited knowledge about the metabolic capabilities and functional roles of related organisms in the environment. The new isolate strain NLcol2<sup>T</sup> expands the known substrate range of this phylum and further reveals the ability to fix nitrogen during anaerobic growth on macroalgal polysaccharides, thereby informing the issue of macroalgal carbon disposal.

**KEYWORDS** polysaccharides, novel bacterium, *Kiritimatiellota*, CAZymes, sulfatases, nitrogen fixation

**Editor** Jennifer B. Glass, Georgia Institute of Technology, Atlanta, Georgia, USA

Address correspondence to David L. Valentine, valentine@ucsb.edu.

The authors declare no conflict of interest.

See the funding table on p. 19.

**Received** 7 June 2023

**Accepted** 5 December 2023

**Published** 24 January 2024

Copyright © 2024 Liu et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Marine macroalgae are important primary producers in coastal ecosystems. They sequester about 173 TgC yr<sup>-1</sup> into their biomass and are considered as part of the “blue carbon” in the ocean (1). Seaweed cultivation has been considered as one of the promising strategies to mitigate the increasing amount of anthropogenic CO<sub>2</sub> and climate change (2). A recent study shows that 24% of macroalgae will eventually reach the seafloor and thus export the fixed carbon to the deep ocean (3). Polysaccharides are important components among the fixed carbon, which includes agar, carrageenan, and fucoidan (4, 5). In contrast to terrestrial plants, marine polysaccharides are usually decorated by sulfate and other functional groups, which require specialized enzymes for removal, thereby limiting the range of microbes that can access and degrade these compounds (6).

Members of the PVC superphylum (named for *Plantomycetes*, *Verrucomicrobia*, and *Chlamydiae*) include degraders of recalcitrant glycopolymers, though much of their true functional diversity has been obscured by the lack of cultivated representatives (7–10). The PVC superphylum also consists of phyla *Kiritimatiellota* and *Lentisphaerae* as well as uncultured candidate phyla from environmental samples (11). The *Kiritimatiellota* phylum was established in 2016 (previously named as *Kiritimatiellaeota*) and was recognized as the Subdivision 5 of *Verrucomicrobia* in the PVC superphylum (12, 13). The geographic distribution of 16S rRNA gene sequences reveals that bacteria in phylum *Kiritimatiellota* are common in anoxic environments ranging from the intestines of animals to hypersaline sediments and wastewater (12). However, there are only four cultivated strains reported to date, and we know little about their metabolic capabilities and functional role in the environment. The first cultivated strain, *Kiritimatiella glycovorans* L21-Fru-AB<sup>T</sup>, is a halophilic saccharolytic bacterium isolated from an anoxic cyanobacterial mat from a hypersaline lake on the Kiritimati Atoll (14). *Pontiella desulfatans* F1<sup>T</sup> and *Pontiella sulfatireligans* F21<sup>T</sup> were isolated from Black Sea sediments and are capable of degrading sulfated polysaccharides like  $\iota$ -carrageenan and fucoidan (15, 16). *Tichowtungia aerotolerans* S-5007<sup>T</sup> was isolated from surface marine sediment and can grow under microaerobic conditions (17).

In this study, we enriched and isolated a novel anaerobic bacterial strain NLcol2<sup>T</sup> from the marine sediments offshore Santa Barbara, CA, USA, which belongs to the *Kiritimatiellota* phylum. We fed the strain with agar,  $\iota$ -carrageenan, and fucoidan as carbon substrate to test whether it is able to degrade these polysaccharides or not. Among other isolates of *Kiritimatiellota*, ammonium has been identified as the nitrogen source, but nitrogen fixation has not been observed. However, macroalgal polysaccharides are depleted in nutrients including nitrogen; therefore, we used nitrogen gas as the sole nitrogen source to test its ability of nitrogen fixation. Strain NLcol2<sup>T</sup> is characterized by phylogenomic, morphological, chemotaxonomic, and physiological traits. We further investigated its metabolic potential by analyzing carbohydrate-active enzymes (CAZymes), sulfatases, and nitrogenases in the genome in detail.

## MATERIALS AND METHODS

### Inoculum source, enrichment, and isolation of strain NLcol2<sup>T</sup>

Strain NLcol2<sup>T</sup> was enriched and isolated from microbial mats found on the surface of marine sediments at Shane Seep (34.40616 N, 119.89047 W) within the Coal Oil Point seep field offshore Santa Barbara, CA, USA. Microbial mat samples were collected at 20-m depth with an *in situ* temperature of 15°C in October 2017. The seep area is characterized by a large amount of hydrocarbon gas emissions, microbial mat coverage, and high sulfide and alkalinity in sediment porewater (18–20). The samples used for inoculum contained both microbial mats and partially decomposed macroalgae (Fig. 1a). The microbial mats were scraped off their attached surface as the inoculum source. The cultures were enriched anaerobically in semisolid agar (0.25% wt/vol, BD Difco Agar, granulated) in the top layer of the sulfide gradient media (Fig. 1b) modified from



**FIG 1** (a) Microbial mats collected on the surface of marine sediments were used as inoculum for the enrichment and isolation of bacterial cultures. (b) Anoxic enrichment culture setup with agar and sulfide gradient seawater media. (c) Isolation of culturable strains on agar plates. Colonies were highlighted in yellow circles.

Kamp et al. (2006). Cultures were maintained at room temperature in the dark and were transferred into fresh media every two to three weeks for a year.

Further isolation of strain NLcol2<sup>T</sup> was performed by streaking on agar plates in an anaerobic chamber (Coy Laboratory Products; Fig. 1c). The medium is the same as the top agar medium in enrichment cultures, except that 1.5% w/v agar (BD Difco Agar, Noble) was added as both gelling agent and substrate and 2-mM sulfide added as reducing agent. The Petri dishes were kept in the anaerobic chamber at room temperature (22°C). Single colonies formed after 3 weeks and were picked from agar plates. Streak plating was repeated for three more rounds to ensure the purity of the culture. Pure culture was subsequently maintained in liquid media with D-galactose (1 g/L) as substrate at 22°C and was transferred every other week. A fully modified medium contained the following in 1,000-mL distilled water: 28.0-g NaCl, 10.0-g  $MgCl_2 \cdot 6 H_2O$ , 3.8-g  $MgSO_4 \cdot 7 H_2O$ , 0.6-g  $CaCl_2 \cdot 2 H_2O$ , 1.0-g KCl, 37-mg  $K_2HPO_4$ , 4-mg  $Na_2MoO_4$ , 50-mg  $Na_2S_2O_5$ , 2-mg  $FeCl_3 \cdot 6 H_2O$ , 10.0-mL modified Wolin's Mineral Solution (see DSMZ medium 141), 0.5-mL Na-resazurin solution (0.1% w/v), 1.0-g D-galactose, 1.0-g  $NH_4Cl$  (optional), 0.75-g  $Na_2CO_3$ , 0.5-g  $Na_2S \cdot 9 H_2O$ , and 10.0-mL modified Wolin's vitamin solution (see DSMZ medium 141). All ingredients except carbonate, sulfide, and vitamins were dissolved under  $N_2/CO_2$  (80:20) atmosphere in Hungate tubes or serum bottles and autoclaved. Carbonate was added from a sterile anoxic stock solution prepared under  $N_2/CO_2$  (80:20) atmosphere. Sulfide and vitamins were added from sterile anoxic stock solutions prepared under 100%  $N_2$  gas. The purity of the isolate was checked by full-length 16S rRNA gene sequencing and observation of morphology under the microscope.

### Phylogenetic reconstruction by 16S rRNA gene

Full-length 16S rRNA gene of strain NLcol2<sup>T</sup> was sequenced by GENEWIZ (Azenta Life Sciences), from colonies grown on agar plates. 16S rRNA gene sequence was searched using the website tool BLASTn (21) against the 16S rRNA database and compared to the sequence identity of the other four isolated strains in the *Kiritimatiellota* phylum.

To construct a phylogenetic tree based on the 16S rRNA gene, 106 sequences over 1,200-base pair (bp) from the *Kiritimatiellales* order in SILVA Ref NR SSU r138.1 database (released August 2020, accessed November 2021; 22) were selected for alignment. The full-length 16S rRNA genes of strain NLcol2<sup>T</sup>, *Tichowtungia aerotolerans* strain S-5007<sup>T</sup>, and two *Verrucomicrobia* (ABEA03000104, AF075271 as outgroups) were also added to the alignment using SINA Aligner v1.2.11 (23). The alignment was trimmed using the "gappyout" method in TrimAl v1.4 (24) to remove ambiguous ends and columns with >95% gaps. All trimmed nucleotide sequences represent >50% of the 1,568 alignment columns. A maximum-likelihood (ML) tree was constructed using RAxML

v.8.2.9 (25) with the GTRGAMMA model of evolution. A rapid bootstrap search was stopped after 1,000 replicates with the extended majority-rule (MRE) consensus tree criterion. The best-scoring ML tree with support values was visualized in the iTOL server (26).

## Genome sequencing and analyses

Genomic DNA was extracted from the isolate cultures using FastDNA Spin Kit for Soil (MP Biomedicals, OH). Genomic DNA library preparation and sequencing were performed at the University of California Davis Genome Center on the Illumina HiSeq 4000 platform with 150-bp paired-end reads. Trimmomatic v.0.36 (27) and Sickle v.1.33 (28) were used to remove the adapter and low-quality or short reads. Trimmed reads were assembled into contigs using MEGAHIT v.1.1.1 (29). Contigs longer than 2,500 bp were kept, and the trimmed reads were mapped back to those contigs using Bowtie2 v.2.3.4.1 (30) and Samtools v.1.7 (31). Contigs were visualized using the Anvi'o v.3 interactive interface (32), and manual binning was performed based on coverage, DNA G+C content, and tetranucleotide frequency signatures. Completion and redundancy of the reconstructed genome were determined using CheckM v.1.0.7 (33).

Open reading frame (ORF) features and protein-coding gene sequences were predicted using Prodigal v.2.6.3 (34). Annotation was assigned to proteins using HMMER v.3.1b2 (35) hmmscan searching against the Pfam v.32.0 (36) and TIGRFAMs v.15.0 (37) databases with a maximum e-value of  $1 \times 10^{-7}$ , corresponding to a bit score of >30 to balance the tradeoffs between false positives and missed matches. Information on protein family, domain, and conserved site were confirmed using InterProScan5 (38). The amino acid sequences of protein-coding genes were further searched against NCBI's Conserved Domain Database (CDD; 39) using the RPS-BLAST program v.2.7.1. The cdd2cog script (40) was used to assign cluster of orthologous group (COG) categories (41) to each protein-coding gene. Protein sequences were also submitted to the BlastKOALA server (42) for KEGG Orthology (KO) ID assignments. Ribosomal RNA (rRNA) genes were determined by RNAmmer v.1.2 (43). Transfer RNA (tRNA) genes were predicted by the tRNAscan-SE 2.0 server (44). Metabolic pathways were reconstructed using the KEGG Mapper (45) and MetaCyc database (46).

For phylogenomic analyses, high-quality genomes in the *Kiritimatiellales* order from NCBI's GenBank database and the Genome Taxonomy Database (GTDB) r95 were selected (accessed on 1 February 2021). *Opitutus terrae* PB90-1 from the *Verrucomicrobia* phylum was selected as the outgroup. All genomes meet the GTDB quality criterion based on completeness and redundancy from CheckM: completeness –  $5 \times$  redundancy > 50. A total of 120 single-copy genes were searched and aligned using GTDB-Tk v1.4.0 (47). The concatenated alignment was further trimmed using TrimAl v1.4 (24) with "gappymout" parameter, which resulted in a final alignment with 4,488 amino acid columns. A maximum-likelihood phylogenetic tree was calculated using RAXML v.8.2.9 (25) with the PROTGAMMALG model of evolution. A rapid bootstrap search was stopped after 350 replicates with MRE-based criterion. The best-scoring ML tree with support values was visualized in the iTOL server (26). The average nucleotide identity (ANI) and average amino acid identity (AAI) between genomes were calculated using the ANI/AAI calculator (48).

Carbohydrate-active enzymes were predicted using dbCAN2 meta server (49). In brief, uploaded protein sequences were searched against the dbCAN CAZyme domain HMM database v.7, CAZy database ([www.cazy.org](http://www.cazy.org)), and PPR library using HMMER, DIAMOND, and Hotpep programs, respectively (49). Only genes predicted by no less than two programs were defined as CAZymes for further analysis. To classify sulfatases into families and subfamilies, gene sequences with an annotated sulfatase domain (PF00884) were searched and classified by the SulfAtlas database v.1.1 (50) using the BLASTp program (21). Additionally, SignalP v.5 (51) was used to predict signal peptides for the translocation of sulfatases into the periplasmic space and outside of the cells.

To better understand the evolution of nitrogen fixation in the *Kiritimatiellota* phylum, reannotation and phylogenetic analysis of the *nifH* gene were performed for all 52 genomes in this phylum from NCBI's GenBank database (accessed on 3 March 2020). The same annotation pipeline described above was used to keep consistency and allow better comparison. *nifH* gene sequences were aligned with 879 full-length *nifH* genes from the genomes of cultivated diazotrophs (<https://www.zehr.pmc.ucsc.edu/Genome879/>) using MUSCLE v.3.8 (52). Two light-independent protochlorophyllide reductases were included as outgroups: ChL from *Trichormus variabilis* ATCC 29413 (WP\_011320185.1) and BchL from *Chlorobium limicola* DSM 245 (WP\_012467085.1). The alignment was trimmed in Jalview v.2.10.5 (53) to remove ambiguous ends and the columns with >95% gaps. All trimmed amino acid sequences represent >81% of the alignment columns. A maximum-likelihood tree was constructed using RAxML v.8.2.9 (25) with the PROTGAMMALG model of evolution. A rapid bootstrap search was stopped after 350 replicates with MRE-based criterion. The best-scoring ML tree with support values was visualized in the iTOL server (26).

## Microscopy

To obtain high-resolution images, cell morphology was examined under the transmission electron microscope (TEM). For TEM imaging, cells grown on the agar plates were fixed with modified Karnovsky's fixative (2% paraformaldehyde and 2.5% glutaraldehyde in 0.1-M sodium phosphate buffer) and spun down into a cell pellet. Cells were rinsed in 0.1-M sodium phosphate buffer and fixed again with 1% osmium tetroxide in the same buffer. After another rinse, they were dehydrated in 50% EtOH, 75% EtOH, 95% EtOH, 100% EtOH, and propylene oxide twice. Cells were pre-infiltrated in 1:1 propylene oxide:resin (Epon/Araldite mixture) overnight, infiltrated in 100% resin, and embedded in fresh resin at 60°C overnight. Ultrathin sections were cut using a Diatome diamond knife. Sections were picked up on copper grids and imaged in a FEI Talos 120C transmission electron microscope at the Biological Electron Microscopy Facility, University of California Davis.

## Chemotaxonomic analysis

The cellular fatty acid composition of strain NLcol2<sup>T</sup> was determined from cells grown at 22°C to late-log phase in a liquid medium with 1.0-g/L D-galactose as carbon source and nitrogen gas as nitrogen source. Cells were centrifuged down at 10,000 × *g* for 10 min and were frozen at −80°C.

Cellular fatty acids were extracted twice using a modified Folch method (54) with a chloroform/methanol mixture (2:1) and tridecanoic acid as an internal standard. The samples were partitioned, and the organic phase containing the total lipid extract (TLE) was retained. Transesterification of the TLE was performed by adding toluene and 1% sulfuric acid in methanol to the TLE after it was brought to complete dryness under N<sub>2</sub>. The acidic methanol/toluene TLE was heated at 90°C for 90 min to produce fatty acid methyl esters (FAMES). The FAMES were extracted from the acidic methanol by adding hexane and water, vortexing, centrifuging, and removing the top (hexane) fraction to a new vial twice. The combined transesterified hexane extracts were dried under N<sub>2</sub> to a final volume of 300 μL. Each extract was spiked with methyl heptadecanoate to calculate the recovery of the internal standard and analyzed by gas chromatography with flame ionization detection (GC-FID).

Concentration analysis was performed with an HP 5890 Series II GC-FID. Chromatography was performed with a 30-m × 0.25-mm internal diameter (ID), 0.25-μm pore size, fused silica Omegawax capillary column with a splitless 1-μL injection. The initial oven temperature was set at 50°C and held for 2 min, followed by a 10°C min<sup>−1</sup> ramp to 150°C, then a 5°C min<sup>−1</sup> ramp to the final temperature of 265°C. A certified reference material (FAME 37, Supelco) was run to calculate retention times and identify peaks. Peak identification was further confirmed by their mass spectra.

Analyses of catalase, oxidase, and API ZYM assay for semiquantitation of enzymatic activities (e.g.,  $\beta$ -galactosidase) were carried out by DSMZ Services, Leibniz-Institut DSMZ—Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany.

## Physiology

The bacterial growth of strain NLcol2<sup>T</sup> was monitored by measuring the optical density (OD) of liquid cultures at 600-nm wavelength. Growth at different temperature (4, 10, 14, 22, 26, 31, 37, 55°C), salinity (0%, 1%, 2%, 2.5%, 3%, 4%, 5%, 6% NaCl), and pH (4.0, 5.0, 5.5, 6.0, 6.5, 7.0, 8.0, 9.0) conditions were determined in triplicates when growing on D-galactose with ammonium supplied. Growth was tested on various substrates (1 g/L) in triplicates at optimum temperature, salinity, and pH conditions with ammonium supplied: D-glucose, D-galactose, D-fructose, L-fucose, L-rhamnose, D-mannose, D-mannitol, meso-inositol, D-arabinose, D-xylose, D-cellobiose, lactose, sucrose, maltose, xylan from corn core (TCI), starch (Sigma-Aldrich), cellulose (Sigma-Aldrich), alginic acid (Acros Organics), agarose (Sigma-Aldrich), agar (BD Difco Agar, Noble),  $\iota$ -carrageenan (TCI), fucoidan from *Macrocystis pyrifera* (Sigma-Aldrich), commercially bought dried red algae (*Porphyra* spp.), commercially bought dried brown algae (*Saccharina japonica*), and the giant kelp (*Macrocystis pyrifera*) harvested from offshore Santa Barbara, CA, USA.

To test the utilization of several nitrogen sources by strain NLcol2<sup>T</sup>, we cultured them with sodium nitrate (1 g/L) and ammonium chloride (1 g/L) and without any nitrogen species supplemented in the liquid media. Two sets of tubes with headspace gases of nitrogen gas or helium gas were made as experimental and control groups, respectively. Triplicate cultures were supplied with 1-g/L D-galactose as substrate and incubated at room temperature (22°C) for 14 days. Growth was monitored by OD (600 nm) measurements.

## Metabolite analysis from galactose fermentation

To quantify the metabolic products of strain NLcol2<sup>T</sup> from galactose fermentation, cultures were grown in triplicates at room temperature (22°C) with D-galactose as the carbon source for 10 days. No ammonium was added to the media, and N<sub>2</sub> gas served as the sole nitrogen source. Growth was monitored by measuring OD at 600-nm wavelength. In addition, 2 mL of culture was subsampled each day (twice a day during the exponential phase) for the quantification of metabolites.

The chromatography protocols used in this study are similar to those previously described (55, 56). Galactose, acetate, succinate, and malate concentrations were measured on an Agilent Infinity 1260 (Agilent Technologies, Santa Clara, CA, USA) high-performance liquid chromatograph (HPLC) using an Aminex HPX-87H analytical column (part no. 1250140, Bio-Rad, Hercules, CA, USA) protected by, first, a 0.22- $\mu$ m physical filter, followed by a Coregel USP L-17 guard cartridge (Concise Separations, San Jose, CA, USA). Separations were performed at 60°C with a flow rate of 0.6-mL/min and a 5-mM sulfuric acid (H<sub>2</sub>SO<sub>4</sub>) mobile phase. Acetate, succinate, and malate were measured using a variable wavelength detector set to 210 nm, while galactose was measured using a refractive index detector set to 35°C. Samples and standards for HPLC were acidified to a concentration of 5-mM H<sub>2</sub>SO<sub>4</sub>, incubated for 5 min at room temperature, and spun at maximum speed in a tabletop centrifuge for 5 min to pellet bacterial cells. The samples were removed from above the cell pellet, and 0.22- $\mu$ m filtered through a polyethersulfone (PES) membrane into HPLC vials with 300- $\mu$ L polypropylene inserts. Standard curves for each compound of interest were constructed using triplicate standards of 0.1, 0.5, and 1.0 g/L. Peaks were integrated using OpenLab CDS analysis software (version 2.6, Agilent Technologies).

Hydrogen gas production was measured on a Fisher Scientific TRACE 1300 gas chromatograph (Thermo Fisher Scientific, Waltham, MA, USA) using a TRACE TR-5 GC Column (part no. 260E113P, Thermo Fisher Scientific) at 30°C, with an Instant Connect Pulsed Discharge Detector (PDD; part no. 19070014, Thermo Fisher Scientific) at 150°C,

and ultrahigh purity He as a carrier gas. All injections of samples and standards were 100  $\mu\text{L}$  in volume. Supplier-mixed standards of 50 ppm, 500 ppm, and 1% hydrogen were run before and after injecting samples, and hydrogen peaks were integrated using Chromeleon Chromatography Data System (CDS) Software (version 6.8, Thermo Fisher Scientific).  $\text{CO}_2$  was not considered due to the carbonate-buffered medium and  $\text{N}_2/\text{CO}_2$  atmosphere.

A tentative fermentation balance was formulated based on the concentrations of galactose, succinate, acetate, malate, and hydrogen measured above. The changes in concentrations in mmol/L were taken as coefficients for these compounds.

### Metabolite analysis from agar and $\iota$ -carrageenan degradation

Cultures were grown at 33°C with 1-g/L agar (BD Difco Agar, Noble) and 1-g/L  $\iota$ -carrageenan (TCI) as carbon sources and ammonium was supplied in the media. In addition, 10 mL of culture was sampled and filtered through 0.22- $\mu\text{m}$  PES membrane (Millipore Millex) both on Day 0 immediately after inoculation, as well as on Day 9 and Day 7 for agar and carrageenan incubations, respectively. Growth was confirmed by OD (600 nm) measurements.

Agar and carrageenan concentrations were quantified as polymeric galactose, the main sugar component of the two polymers. Polymeric galactose was quantified as the difference between total galactose and free galactose. To measure total galactose, 5 mL of 0.22- $\mu\text{m}$ -filtered media was acid hydrolyzed to cleave glycosidic linkages and release galactose. Samples were hydrolyzed in 1-M HCl at 100°C for 20 h. Following hydrolysis, samples were neutralized by  $\text{N}_2$  evaporation and diluted 1:1,000 with ultrapure water. Galactose was quantified using high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) on a DIONEX ICS5000+ equipped with a CarboPac PA10 column using an isocratic elution of 18-mM NaOH for 20 min (57). Free galactose was measured by HPAEC-PAD before acid hydrolysis. Incubation media was 1:25 diluted with ultrapure water to reduce the salt concentration and quantified using the same gradient program described above.

Acetate and succinate concentrations were measured on the Agilent Infinity 1260 HPLC using a similar protocol described above in the "Metabolite analysis from galactose fermentation" section, except using a refractive index detector.

### Acetylene reduction assay

To test the nitrogenase activity of strain NLcol2<sup>T</sup> when growing with nitrogen gas as the sole nitrogen source, an acetylene reduction assay was performed following Hardy et al. (1968). In short, acetylene ( $\text{C}_2\text{H}_2$ ) can be reduced to ethylene ( $\text{C}_2\text{H}_4$ ) when nitrogenases actively fix nitrogen gas at the same time. Cultures were grown on D-galactose in triplicates at 22°C, and triplicate media bottles without inoculation were used as controls. In addition, 1.2 mL of acetylene was injected into all culture and control bottles, which contained 80 mL of liquid and 80 mL of headspace pressurized at 150 kPa at the beginning. Gas concentrations and  $\text{OD}_{600}$  were measured at six time points during the 18-day incubation. Acetylene and ethylene concentrations were resolved on a Shimadzu 8A gas chromatograph with a flame ionization detector (GC-FID). Furthermore, 1.5-mL samples and standards were injected and then carried by  $\text{N}_2$  at a flow rate of 20 mL/min through an n-octane on Res-Sil C packed column (Restek, Centre County, PA, USA) set at 25°C. Moreover, 0.5% and 1.0% GASCO calibration gas mixtures of acetylene and ethylene (Cal Gas Direct Incorporated, Huntington Beach, CA, USA) were used for the standard curves.



## RESULTS AND DISCUSSION

### Phylogenetic analyses

Phylogenetic placement of strain NLcol2<sup>T</sup> was determined by comparing full-length 16S rRNA gene, single-copy genes, and whole-genome similarity metrics including ANI and AAI.

Strain NLcol2<sup>T</sup> was classified within the R76-B128 clade (*Pontiellaceae* family in GTDB database) of the *Verrucomicrobia* phylum under the current SILVA taxonomy (SILVA Ref NR SSU r138.1). Full-length 16S rRNA gene of the isolate shares 84.1%, 88.9%, 92.9%, and 94.5% identity with the four reported cultivated strains in the *Kiritimatiellota* phylum: *Kiritimatiella glycovorans* strain L21-Fru-AB<sup>T</sup>, *Tichowtungia aerotolerans* strain S-5007<sup>T</sup>, *Pontiella sulfatireligans* strain F21<sup>T</sup>, and *Pontiella desulfatans* strain F1<sup>T</sup>, respectively. Strain NLcol2<sup>T</sup> is more closely related to *P. desulfatans* and *P. sulfatireligans* than *K. glycovorans* and *T. aerotolerans*. The 16S rRNA gene identities compared to *P. desulfatans* and *P. sulfatireligans* were absolutely higher than the 86.5% threshold for family level but fall on the edge of the threshold for a new genus as 94.5% (58). A maximum-likelihood tree of 16S rRNA gene sequences from the *Kiritimatiellota* phylum was reconstructed by RAxML (Fig. S1). The R76-B128 clade (*Pontiellaceae* family) formed a monophyletic group with the MSBL3 clade (*Tichowtungiaceae* family) as the sister group, both of which are in a different cluster from the *Kiritimatiellaceae* family. It is clear that strain NLcol2<sup>T</sup> is not affiliated with *K. glycovorans* within the *Kiritimatiellaceae* family nor with *T. aerotolerans* within the *Tichowtungiaceae* family (MSBL3 clade) but belongs to the *Pontiellaceae* family (R76-B128 clade) within the *Kiritimatiellales* order as do *P. desulfatans* and *P. sulfatireligans* (15).

To resolve the phylogeny of strain NLcol2<sup>T</sup> in detail, we further performed genome-level phylogenetic analyses using the Genome Taxonomy Database Toolkit (47). A concatenated phylogenomic tree was reconstructed from 120 bacterial single-copy genes of genomes in the *Kiritimatiellales* order (Fig. 2). Here, strain NLcol2<sup>T</sup> falls within the *Pontiellaceae* family with a bootstrap value of 100. Additionally, the AAI values of the genomes between strain NLcol2<sup>T</sup> and *P. desulfatans* and *P. sulfatireligans* are 69.94% and 68.51%, which are slightly above the threshold of 65% for the same genus (59). However, within *Pontiella* genus, it represents a different group from *P. desulfatans* and *P. sulfatireligans*. Moreover, the ANI values of the genomes between strain NLcol2<sup>T</sup> and *P. sulfatireligans* and *P. desulfatans* are 72.73% and 73.71% respectively, which was much lower than the 95% ANI criterion for the same species (60, 61). Therefore, we propose that strain NLcol2<sup>T</sup> represents a novel species within the *Pontiella* genus according to the phylogenetic analyses above.

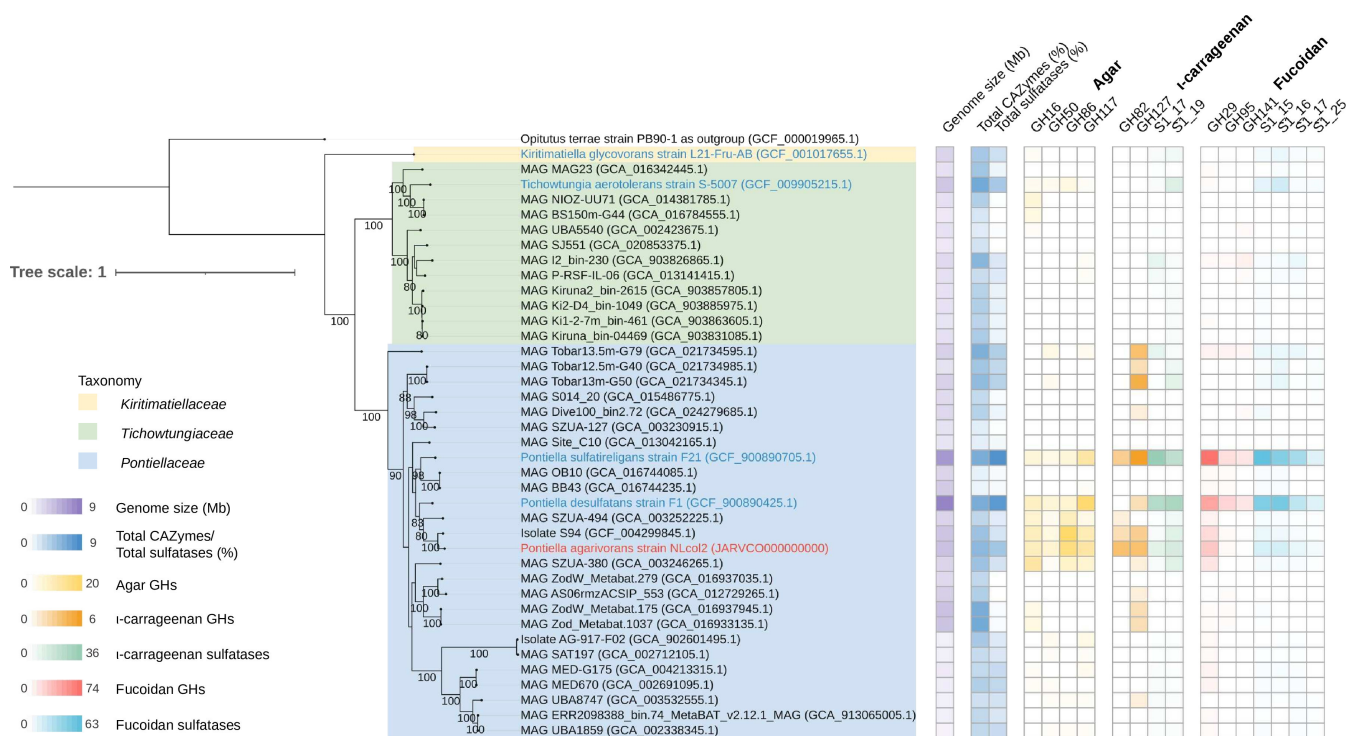
### General features of the genome

The draft genome of strain NLcol2<sup>T</sup> is 95% complete with 4% redundancy. The genome consists of 12 contigs (N50 is 1,265,434 bp) with a total length of 4,436,865 bp, and the mean coverage is 593x. DNA G + C content is 52.4 mol%. 5S, 16S, and 23S rRNA genes and 50 tRNA genes were found in the genome.

A total of 3,611 open reading frame (ORF) features were predicted by Prodigal, among which 2,757 proteins in the genome were assigned with COG functional category codes. The number of genes in each functional category is shown in Figure S2. More genes are involved in carbohydrate ( $n = 260$ ) and amino acid ( $n = 188$ ) transport and metabolism than those of nucleotides ( $n = 62$ ) and lipids ( $n = 61$ ), which is similar to that in *Kiritimatiella glycovorans* (12). A further detailed analysis of genes involved in macroalgal polysaccharide degradation and nitrogen fixation is presented in the "Anaerobic degradation of macroalgal polysaccharides" and "Nitrogen fixation" sections below.

### Morphologic and chemotaxonomic characterization of strain NLcol2<sup>T</sup>

Single colonies on agar plates were white or ivory, circular, and smooth after growing anaerobically for 2 weeks at 22°C. Bacterial cells of strain NLcol2<sup>T</sup> have a round to ovoid



**FIG 2** Concatenated maximum-likelihood phylogenomic tree of 120 bacterial single-copy genes from 39 genomes in the *Kiritimatiellales* order. Strain NLcol2<sup>T</sup> is labeled in red, and other cultivated strains are labeled in blue. A *Verrucomicrobia* genome was selected as an outgroup, and the tree was rooted there. Bootstrap values over 80 are shown on the nodes. Genome size (Mb), total CAZymes (%), and total sulfatases (%) as a percentage of all protein-coding genes in each genome are presented as reference. The number of glycoside hydrolase (GH) and sulfatase homologs involved in the degradation pathways of agar, l-carrageenan, and fucoidan is presented in the heatmap.

shape with a size of 1  $\mu\text{m}$  in diameter observed under the microscope (Fig. 3). Cells divided by binary fission and genes of bacterial cell division complex including the FtsZ family were present. No motility or flagella were observed, although a full set of genes coding for flagellar assembly was present in the genome. No spore formation was observed. A Gram-negative cell wall structure of the outer membrane, periplasmic space, and cytoplasmic membrane was shown by electron microscopy (Fig. 3). There are also genes coding for proteins involved in lipopolysaccharide export and peptidoglycan synthesis in the genome. Some bacteria in the PVC superphylum exhibit compartments inside the cells (62), but like other strains in the *Kiritimatiellota* phylum, no compartmentalization of the cytoplasm was observed in strain NLcol2<sup>T</sup>. There were unknown inclusions or granules present inside the cells, and genes involved in the synthesis and utilization of polyphosphate and glycogen were found in the genome, which may serve as phosphate and energy storage materials, respectively.

Major cellular fatty acids (>10% of total) of strain NLcol2<sup>T</sup> include C18:0, *i*-C12:0, *i*-C18:0, and *i*-C14:0, in order of abundance. The major cellular fatty acid profile is quite different from *K. glycovorans* and *T. aerotolerans* but almost the same as that in *P. desulfatans* and *P. sulfatireligans*, except that *P. sulfatireligans* also has *i*-C16:0 as one of the major components (Table 1). Again, this agrees with the phylogenetic placement of strain NLcol2<sup>T</sup> in the *Pontietta* genus, being more closely related to *P. desulfatans* than *P. sulfatireligans*. However, strain NLcol2<sup>T</sup> can be further distinguished by a relatively higher composition of *i*-C18:0 than *i*-C14:0, while *P. desulfatans* has more *i*-C14:0 than *i*-C18:0 (Table S1). Other cellular fatty acids detected in strain NLcol2<sup>T</sup> include C16:0, *i*-C16:0, C20:0, and *i*-C20:0 (Table S1).

TABLE 1 Comparison of phenotypic characteristics between strain NLcol2<sup>T</sup> and four other isolated strains in the *Kiritimatiellota* phylum<sup>a</sup>

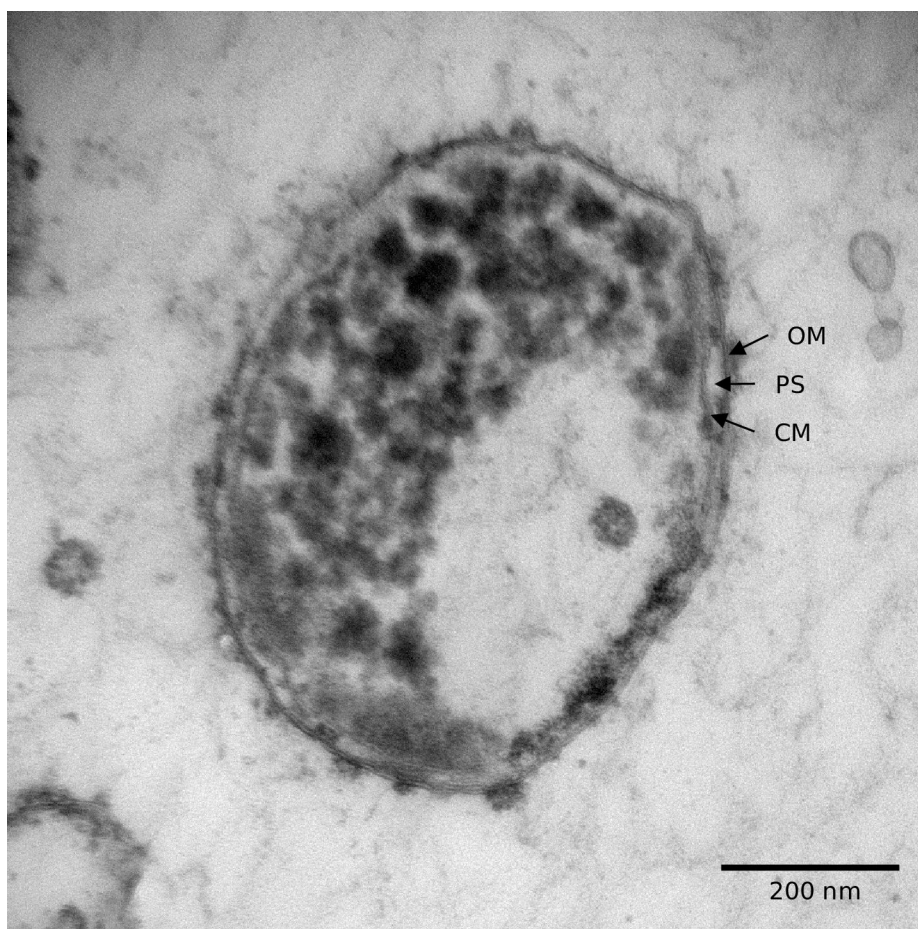
Strains	<i>P. agarivorans</i> NLcol2 <sup>T</sup>	<i>P. desulfatans</i> F1 <sup>T(a)</sup>	<i>P. sulfatireligans</i> F21 <sup>T(a)</sup>	<i>K. glycovorans</i> L21-Fru-AB <sup>T(b)</sup>	<i>T. aerotolerans</i> S-5007 <sup>T(c)</sup>
Isolation source	Microbial mat on marine sediment	Anoxic marine sediment	Anoxic marine sediment	Hypersaline microbial mat	Marine sediment
Cell shape	Spherical	Spherical	Spherical	Spherical	Spherical
Cell size (µm)	1.0	0.5–1.2	0.5–1.0	1.0–2.0	0.5–0.8
Motility	-	-	-	-	-
Genome size (Mbp)	4.44	8.66	7.40	2.95	3.88
DNA G + C content (mol%)	52.4	56.0	54.6	63.3	53.1
Major cellular fatty acids (>10% of total)	C18:0, <i>i</i> -C12:0, <i>i</i> -C18:0	C18:0, <i>i</i> -C12:0, <i>i</i> -C14:0	C18:0, <i>i</i> -C12:0, <i>i</i> -C18:0	<i>i</i> -C14:0, C18:0	C18:0, <i>i</i> -C12:0, <i>i</i> -C18:0, <i>i</i> -C16:0
Catalase activity	-	-	-	-	Weak
Oxidase activity	-	-	+	-	-
Growth temperature (°C)					
Range	10–37	10–30	0–25	20–40	15–45
Optimum	31	25	25	28	33–35
Growth salinity (g/L NaCl)					
Range	10–60	10–31	10–50	20–180	5–80
Optimum	25–30	23	23	60–70	30–40
Growth pH					
Range	6.0–9.0	6.5–8.5	6.0–8.5	6.5–8.0	6.0–8.5
Optimum	8.0	7.5	7.5	7.5	7.0–7.5
Substrate utilization					
Glucose	+	+	+	+	+
Galactose	+	+	+	+/-	+
Fructose	+	+	+	-	+
Fucose	-	+	+	-	NA
Rhamnose	-	+	+	+/-	+
Mannose	+	-	+	+	-
Mannitol	+	-	+	-	-
Arabinose	-	+	-	-	+
Xylose	+	+	+	+	-
Lactose	+	+	+	-	NA
Cellobiose	+	+	+	-	+
Sucrose	+	+	+	-	-
Maltose	+	+	+	-	-
Fucoidan	-	+	+	+/-	NA
<i>l</i> -Carrageenan	+	-	+	+/-	NA
Xylan	+	-	-	NA	NA
Agar	+	-	-	-	-

(Continued on next page)

TABLE 1 Comparison of phenotypic characteristics between strain NLcol2<sup>T</sup> and four other isolated strains in the *Kiritimatellota* phylum<sup>a</sup> (Continued)

Strains	<i>P. agarivorans</i> NLcol2 <sup>T</sup>	<i>P. desulfatans</i> F1 <sup>T(a)</sup>	<i>P. sulfatireligans</i> F21 <sup>T(a)</sup>	<i>K. glycovorans</i> L21-Fru-AB <sup>T(b)</sup>	<i>T. aerotolerans</i> S-5007 <sup>T(c)</sup>
Major nongaseous fermentation products*	Succinate, acetate, malate	Acetate, ethanol, lactate	Acetate, ethanol, lactate	Ethanol, acetate	Acid (maybe acetate**)
Nitrogen sources	N <sub>2</sub> , NH <sub>4</sub> <sup>+</sup>	NH <sub>4</sub> <sup>+</sup>	NH <sub>4</sub> <sup>+</sup>	NH <sub>4</sub> <sup>+</sup>	NH <sub>4</sub> <sup>+</sup>

<sup>a</sup>Notations: NA, data not available. +, positive; -, negative; +/-, unstable, ceasing growth upon the second transfer. Data for strains other than NLcol2<sup>T</sup> were referenced from literatures: (a) (16); (b) (12); (c) (17). \*Substrates were D-galactose for strain NLcol2<sup>T</sup> and D-glucose for other strains. \*\*For strain S-5007<sup>T</sup>, acetate production was predicted from genomic data.



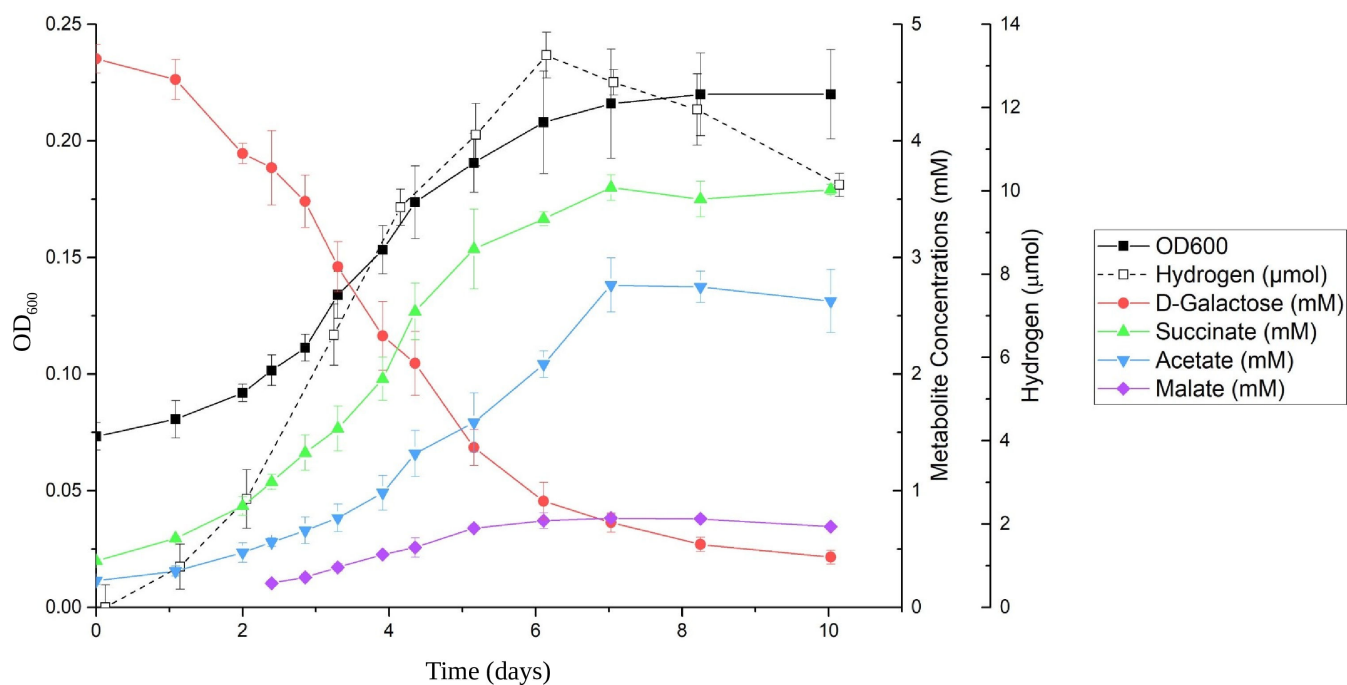
**FIG 3** Transmission electron microscopy (TEM) image of strain NLcol2<sup>T</sup> shows the cell of ~1  $\mu\text{m}$  in diameter with a Gram-negative cell wall structure. OM, outer membrane; PS, periplasmic space; CM, cell membrane.

Strain NLcol2<sup>T</sup> tested negative for both catalase and oxidase, which is common in strict anaerobes (Table 1).  $\beta$ -galactosidase was tested positive with ~5 nanomoles of substrate hydrolyzed in the API Zym assay.

### Physiology of growth

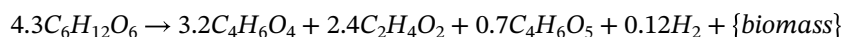
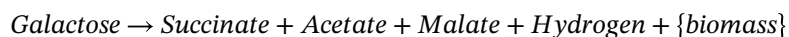
Strain NLcol2<sup>T</sup> exhibited consistent growth between 10–37°C (optimum 31°C), with NaCl concentration between 10 and 60 g/L (optimum 25–30 g/L) and with pH 6.0–9.0 (optimum pH 8.0) when D-galactose was utilized as the substrate. It was determined as a mesophilic and neutrophilic bacterium, which is similar to the other four isolated strains from the *Kiritimatiellota* phylum (Table 1). Growth with ammonium supplied in the medium was faster than when dependent on nitrogen fixation. The doubling times were 15 h and 65 h when growing with and without ammonium, respectively, at room temperature (22°C). Strain NLcol2<sup>T</sup> was considered as obligately anaerobic, being unable to grow with the presence of oxygen and even in a non-reduced medium lacking sulfide as the reducing agent.

Strain NLcol2<sup>T</sup> was able to grow on various carbohydrate substrates under optimal conditions with ammonium supplied, which includes D-glucose, D-galactose, D-fructose, D-mannose, D-mannitol, D-xylose, D-cellobiose, lactose, sucrose, maltose, xylan, agarose, agar, and  $\iota$ -carrageenan (Fig. S3). No growth was observed when supplied with L-fucose, L-rhamnose, D-arabinose, meso-inositol, starch, cellulose, alginic acid, or fucoidan from *Macrocystis pyrifera*.



**FIG 4** Time series of major metabolites from D-galactose fermentation by strain NLcol2<sup>T</sup>. The growth curve was measured by optical density at 600 nm (OD<sub>600</sub>). D-Galactose decreased while fermentation products of succinate, acetate, malate, and hydrogen gas were produced during anaerobic bacterial growth on D-galactose.

When growing on D-galactose, major fermentation products formed were succinate and acetate, with small amounts of malate and hydrogen gas also detected during the incubation (Fig. 4). Initially, the culture was supplied with  $4.71 \pm 0.12$ -mM D-galactose, and only  $0.43 \pm 0.06$  mM D-galactose remained after the 10-day incubation period. Taking all fermentation products into consideration, the fractional electron recovery for galactose fermentation by strain NLcol2<sup>T</sup> was about 75%. The remaining electrons could be shunted to and utilized by nitrogen fixation and biomass formation. A tentative fermentation balance was formulated as below, including measured fermentation products:



## Anaerobic degradation of macroalgal polysaccharides

### CAZyme analyses

Microbial degradation of macroalgal polysaccharides involves complex metabolic pathways and requires a large number of enzymes during the process (63–66). CAZymes, especially glycoside hydrolases (GHs) and polysaccharide lyases, can break down polysaccharides into oligosaccharides (67). In the genome of strain NLcol2<sup>T</sup>, 202 genes (5.6% of predicted ORFs) were predicted to be CAZymes and associated carbohydrate-binding modules (CBM) by dbCAN2 meta server (49; Table S2). Among these, 164 genes were annotated to be in the GH families. GH2, GH29, GH86, and GH117 are the most abundant families mainly represented by  $\beta$ -galactosidase,  $\alpha$ -L-fucosidase,  $\beta$ -agarase, and  $\alpha$ -1,3-L-neoagarooligosaccharide hydrolase. Furthermore, 100 GHs were predicted with signal peptide sequences indicating that 61% of GHs target the cell membrane or can be exported outside of the cell. Extracellular and membrane-associated GHs may hydrolyze large extracellular polymers that cannot otherwise enter the cell. Four porins and nine

sugar transporters of the major facilitator superfamily were also present in the genome, which may help with the acquisition of carbohydrate molecules by the cell.

### Sulfatase analyses

As most marine polysaccharides are sulfated, another group of enzymes called sulfatases is needed in the degradation pathway, which can cleave sulfate ester groups off the carbohydrate backbone (50). It has been shown that *Kiritimatiellota* and PVC superphylum have large numbers of copies of sulfatase genes in their genomes (15), and it is the same case in strain NLcol2<sup>T</sup>. We found 165 sulfatase genes (PF00884), comprising 4.6% of predicted ORFs in the genome.

Sulfatases are activated via posttranslational modification by other enzymes before functioning. The most common one is the formylglycine-generating enzyme (FGE), which transforms a cysteine or serine residue into a catalytic formylglycine (68). These fGly-sulfatases are classified as type I sulfatases (family S1), which contain all carbohydrate sulfatases and are the largest sulfatase family (6). Sulfatases were classified into 22 subfamilies in the SulfAtlas database (50), all of which belong to family S1 fGly-sulfatases (Table S3). The most abundant subfamilies (>5% of total sulfatases) in strain NLcol2<sup>T</sup> are S1\_16, S1\_7, S1\_15, S1\_24, S1\_8, S1\_19, S1\_17, and S1\_20. Homologous sulfatases with known enzymatic activities within these subfamilies include D-galactose-6-sulfate 6-O-sulfohydrolase, endo-/exo-xylose-2-sulfate-2-O-sulfohydrolase, endo-/exo-galactose-4-sulfate-4-O-sulfohydrolase, endo-3,6-anhydro-D-galactose-2-sulfate-2-O-sulfohydrolase, and exo-fucose-2-sulfate-2-O-sulfohydrolase, and the known substrates of these sulfatases include alpha-/iota-/kappa-carrageenan, fucan, and ulvan (Table S3). These results imply that strain NLcol2<sup>T</sup> has the potential to target a vast variety of sulfated polysaccharides, similar to isolates *K. glycovorans* (12), *P. desulfatans*, and *P. sulfatireligans* (15, 16). However, due to the limited number of characterized fGly-sulfatases, there are still many unknowns about the specific substrates and/or reactions catalyzed by sulfatases in each subfamily (50). In addition, 128 sulfatases have the best match genes from organisms in the PVC superphylum and 32 from *Bacteroidota*, and 96% of sulfatases (n = 158) were predicted to have a signal peptide sequence, indicating most sulfatases could be membrane-anchored or exported outside of the cell.

Although less well studied, the anaerobic sulfatase-maturing enzyme can mature either cysteine or serine sulfatases under anaerobic conditions (69, 70). There are also five genes encoding formylglycine-generating enzyme and one encoding anaerobic sulfatase-maturing enzyme, which are essential for the activation of sulfatase by posttranslational modification (68, 69).

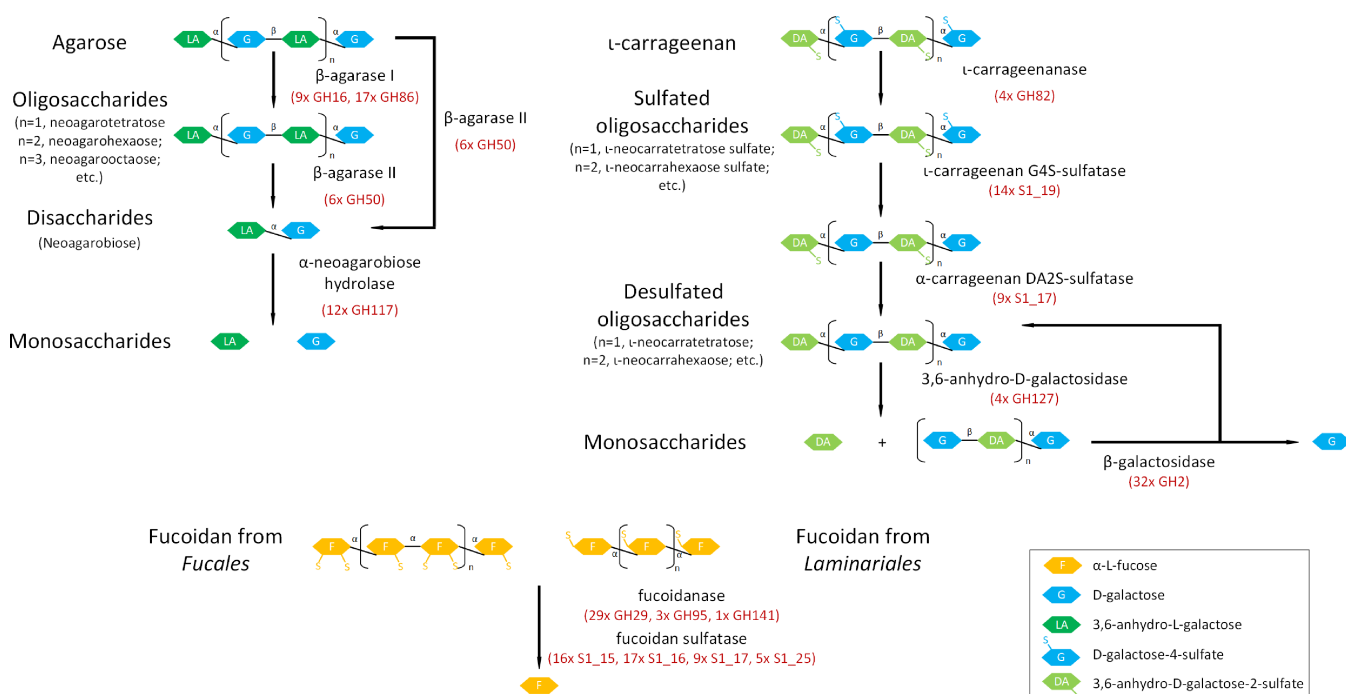
### Growth on macroalgal polysaccharides

We further confirmed the ability of strain NLcol2<sup>T</sup> to grow on different macroalgal polysaccharides in live cultures. Bacterial growth was observed in anaerobic cultures with agarose, agar, and ι-carrageenan, but not fucoidan. Many commercially bought algal polysaccharides are contaminated with co-extracted impurities, so we took direct measurements of polysaccharides to confirm the degradation of agar and ι-carrageenan by strain NLcol2<sup>T</sup>. Agar and ι-carrageenan concentrations were quantified as polymeric galactose after acid hydrolysis. Polymeric galactose of agar and ι-carrageenan decreased by 88% and 91%, respectively, while the fermentation products of succinate and acetate increased by 88%–93% along with bacterial growth (Fig. S4). The carbon recovery rates are 78% and 87% for agar and ι-carrageenan degradation, respectively. This indicates that strain NLcol2<sup>T</sup> is able to degrade agar and ι-carrageenan and their growth was mainly fueled by these polysaccharides but not the impurities. This is the first strain reported with the ability to utilize agar as substrate in the *Kiritimatiellota* phylum. We further tested their growth on seaweeds and cells also exhibited consistent growth on dried red algae (*Porphyra* spp.) and dried brown algae (*Saccharina japonica*), but not on the giant kelp (*Macrocystis pyrifera*). Since agar, porphyran, and carrageenan are all

sulfated polysaccharides extracted mainly from red algae with a similar structure (5, 71), it is not surprising that cells can grow on *Porphyra* spp. directly.

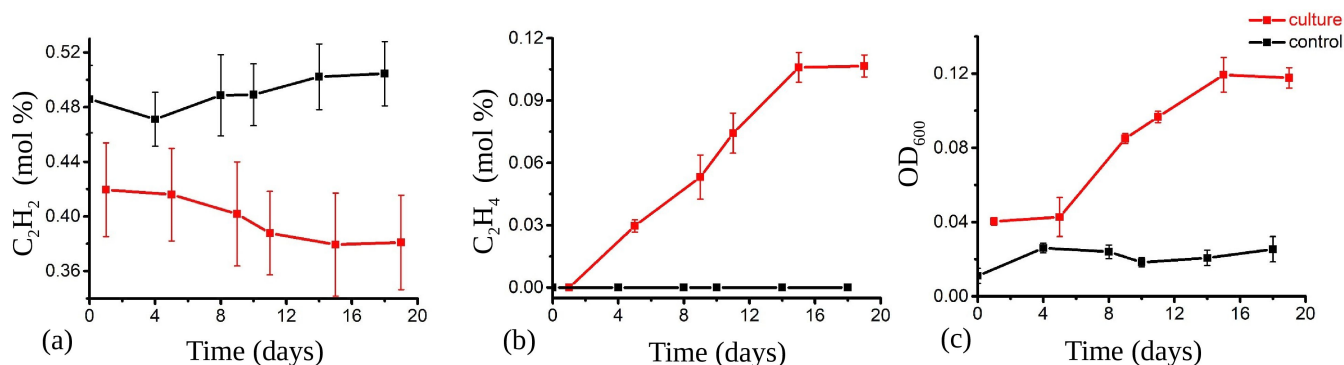
Agar is a mixture of agarose and agaropectin, which is commonly used as a solidifying agent for culture media. Agarose is composed of alternating  $\alpha$ -1,3-linked D-galactose and  $\beta$ -1,4-linked 3,6-anhydro- $\alpha$ -L-galactose with little sulfate modification, while agaropectin is heavily modified with sulfate (72–74). Carrageenan is structurally related to agarose, except the  $\beta$ -linked unit is D-galactose-6-sulfate (73). Fucoidan is also a sulfated polysaccharide composed mainly of L-fucose units adorned with sulfate esters, while minor xylose, galactose, mannose, and glucuronic acid can be present too (75). Algal polysaccharide degradation has been well studied in *Zobellia galactanivorans* Dsij<sup>T</sup>, the marine *Bacteroidota* model for the discovery of agarases, porphyranases, and carrageenases (64, 65, 76). We found not only potential genes involved in the degradation pathways of agar and  $\iota$ -carrageenan but also fucoidan (Fig. 5; Table S4). Homologous genes encoding for potential  $\beta$ -agarases,  $\iota$ -carrageenases, and associated sulfatases were found in the genome of strain NLcol2<sup>T</sup> and could be involved in degrading agar and  $\iota$ -carrageenan into galactose and anhydrogalactose, which then can be directed to the central metabolism for energy. Potential fucoidanases were also found in the genome, but this contrasts with the experimental observation that cells did not grow with fucoidan from *M. pyrifer* as the sole carbon source. However, bacterial growth was not supported by L-fucose either (see the “Physiology of growth” section above), indicating that strain NLcol2<sup>T</sup> may house potential fucoidanases only to remove fucose from fucoidan, but cannot further metabolize fucose and cannot gain energy from fucoidan degradation to support its growth. Alternatively, these genes may not encode fucoidanases to degrade fucoidan but may encode enzymes for other purposes, for example, removing the fucose cap from mucin-like molecules (77).

A neighborhood analysis of the genome shows that GHs and sulfatases are often located nearby (within the distance of five genes), suggesting certain sulfatases and glycoside hydrolases could be regulated together to degrade sulfated polysaccharides



**FIG 5** Homologs of key enzymes involved in the metabolic pathways of agar,  $\iota$ -carrageenan, and fucoidan degradation were found in the genome. The number of gene copies encoding homologs of enzymes in the pathway is also listed. GH, glycoside hydrolase; S1\_#, sulfatase classified in S1 family\_subfamily. The GHs and sulfatases listed are families/subfamilies including putative agarases, carrageenans, and fucoidanases. These annotations need further subfamily-based annotation and biochemical verification.





**FIG 6** Nitrogenase activity of strain NLcol2<sup>T</sup> detected by acetylene reduction assay. Acetylene (a), ethylene (b) percentage, and OD<sub>600</sub> (c) change with time in liquid culture with D-galactose as substrate and nitrogen gas as the sole N source. Cultures with bacteria are labeled in red, and media without inoculum (control) are labeled in black. The decrease of acetylene and the increase of ethylene indicated active nitrogen fixation during bacterial growth.

(78). In some cases, histidine kinase (PF07730, PF02518), response regulator (PF00072), and TonB-dependent transporters (PF00593, PF03544) are in the neighborhood too. The histidine kinase and response regulator together form a two-component signal transduction system that may help bacteria sense available substrates and respond to the changing environments (79). There are cases when sulfatases themselves cluster together, for example, four or six copies in a row. A complete pathway for assimilatory sulfate reduction is also present in the genome and the cells may utilize the cleaved sulfate group for biosynthesis of reduced sulfur compounds.

A comparative study of GHs and sulfatases in selected genomes of the *Kiritimatiellales* order revealed that not all genomes harbor enzymes involved in degradation pathways of agar, ι-carrageenan, and fucoidan, and some bacteria don't have any GHs or sulfatases at all (Fig. 2; Table S5). However, certain genomes in the *Pontiella* genus show a relatively larger component of GHs and sulfatases. This indicates that these bacteria may adopt the lifestyle of utilizing macroalgal polysaccharides like agar, carrageenan, and fucoidan as carbon and energy sources, while other clades in the *Kiritimatiellales* order may specialize on other substrates available in their living environments. Some genomes in the *Pontiellaceae* family do not have a high number of GHs or sulfatases either. This may indicate that these carbohydrate-related genes could be laterally transferred into the *Pontiella* genus, but some were lost during evolution while living in environments where other available substrates were preferred. For example, such phenomenon was reported that the lateral gene transfer of porphyranases was from the marine *Bacteroidota*, *Zobellia galactanivorans*, to the human gut bacterium *Bacteroides plebeius* (80). Another possibility would be that these metagenome-assembled genomes (MAGs) were incomplete, and the GHs or sulfatases investigated were not easy to capture.

## Nitrogen fixation

We further tested nitrogen-fixing ability in live cultures of strain NLcol2<sup>T</sup>. The strain was able to grow on nitrogen gas as the sole N source in a nitrogen-free medium with D-galactose as the carbon source. No growth was observed when nitrogen was replaced by helium in the headspace. Bacterial growth was also supported by ammonium but not nitrate (Fig. S5), and neither assimilatory nor dissimilatory nitrate reductase was present in the genome. Nitrogenase activity was detected by acetylene reduction assay. The production of ethylene from acetylene during bacterial growth on nitrogen gas as the sole nitrogen source showed that the cultures expressed active nitrogenases and could fix nitrogen gas into bioavailable forms to support their growth (Fig. 6). This nitrogen-fixing ability may give them the advantage to survive in nitrogen-limiting environments.

Mo-dependent nitrogenase is the most common and widely studied enzyme that performs nitrogen fixation. It contains two components: an Fe protein as the reductase

(*nifH*) collecting and transferring electrons and a MoFe protein (*nifDK*) binding dinitrogen (N<sub>2</sub>) and converting it to ammonia (NH<sub>3</sub>; 81). Genes encoding both nitrogenase iron protein (*nifH*, PF00142) and nitrogenase molybdenum-iron protein alpha and beta subunits (*nifDK*, PF00148) are present in the genome, which together form a complete pathway of nitrogen fixation. No alternative vanadium-iron nitrogenase or iron-only nitrogenase was found. In addition to *nifHDK*, both *nifB* and *nifE* involved in the biosynthesis of nitrogenase MoFe cofactor are present in the genome. Two genes coding for nitrogen regulatory protein PII were present, which are important for the regulation of nitrogen fixation in response to nitrogen source availability (82). The rop-like protein is uncharacterized but often found in nitrogen fixation operons and may play a role in regulation (83). There are various other *nif* genes present in other parts of the genome including *nifA*, *M*, *S*, *U*, and *V*, which together may help regulate the function of nitrogenase (Table S6).

Nitrogenases are highly oxygen-sensitive, but even though there are diverse anaerobes in the PVC superphylum, only a few studies demonstrated nitrogen fixation in this superphylum (84–87), and no reports in the *Kiritimatiellota* phylum. Moreover, we have little knowledge as to where *nif* genes were acquired by the nitrogen-fixing members in the PVC superphylum. We found five genomes in this phylum housing a *nifH* gene. Three were from *P. desulfatans*, *P. sulfatireligans*, and isolate S94, and two were from the marine sediments at the hydrothermal vent of South Mid-Atlantic Ridge (SZUA-380 and SZUA-494). All *nifH* genes in this clade were classified as cluster III, which is dominated by distantly related obligate anaerobes (88). All six *nifH* genes from the *Kiritimatiellota* phylum form a monophyletic clade with a bootstrap value of 89 (Fig. S6). They also cluster together with sequences from *Chlorobi*, *Bacteroidota*, and *Delataproteobacteria* (mainly the *Desulfovibrio* genus), *Spirochaetes*, and some *Verrucomicrobia* to form a monophyletic clade with a bootstrap value of 85. This suggests that there could be lateral gene transfer between the *Kiritimatiellota* phylum and other phyla in this clade, but some bacteria in the *Kiritimatiellota* phylum may have lost *nif* genes during evolution. Nitrogen fixation genes in a methanotrophic verrucomicrobial isolate *Methylacidiphilum fumariolicum* strain SolV resemble those from the *Gammaproteobacteria*, which supports their acquisition of *nif* genes through lateral gene transfer (84).

## Conclusion

In this study, we reported a novel anaerobic bacterial strain NLcol2<sup>T</sup> isolated from microbial mats in marine sediments as the representative of a novel species in the *Pontiella* genus, which is the fifth cultivated strain in the *Kiritimatiellota* phylum. It represents the first strain to utilize agar as a substrate with nitrogen-fixing ability in the *Kiritimatiellota* phylum. An extensive list of CAZymes and sulfatases shows its potential to degrade diverse macroalgae-derived sulfated polysaccharides in marine environments.

## Description of *Pontiella agarivorans* sp. nov.

*Pontiella agarivorans* (a.ga.ri.vo'rans. N.L. neut. n. *agarum* agar, algal polysaccharide; L. pres. part. adj. *vorans* devouring, consuming; N.L. part. adj. *agarivorans* agar-devouring).

Cells are Gram-negative, anaerobic, nonmotile cocci with a diameter of 1 μm. No spore formation was observed. Cells divide by binary fission. Colonies on agar plates are milky or ivory, circular, and smooth. Growth occurs at 10–37°C (optimum 31°C), with NaCl concentration between 10 and 60 g/L (optimum 25–30 g/L) and with pH 6.0–9.0 (optimum pH 8.0) when D-galactose was utilized as the substrate. The following substrates support growth: D-glucose, D-galactose, D-fructose, D-mannose, D-mannitol, D-xylose, D-cellobiose, lactose, sucrose, maltose, xylan, agarose, agar, ι-carrageenan, and fucoidan. The following compounds do not support growth under laboratory conditions: L-fucose, L-rhamnose, D-arabinose, meso-inositol, starch, cellulose, or alginate. The nongaseous fermentation products from D-galactose are succinate, acetate, and malate (traces). Both ammonium and nitrogen gas can be utilized as nitrogen sources, but

nitrate and nitrite were not utilized. Major cellular fatty acids are C18:0, *i*-C12:0, and *i*-C18:0.

The type strain NLcol2<sup>T</sup> (= DSM 113125<sup>T</sup> = MCCC 1K08672<sup>T</sup>), was isolated from microbial mats on the surface of marine sediments offshore Santa Barbara, CA, USA. The genome of the type strain is 4.4 Mbp in size, and DNA G + C content is 52.4 mol%. The GenBank accession number for the full-length 16S rRNA gene sequence of strain NLcol2<sup>T</sup> is [OQ749723](https://www.ncbi.nlm.nih.gov/nuclot/OQ749723). The genome of strain NLcol2<sup>T</sup> has been deposited as the Whole Genome Shotgun project at DDBJ/ENA/GenBank under the accession number [JARVCO000000000](https://www.ncbi.nlm.nih.gov/nuclot/JARVCO000000000). The version described in this paper is version [JARVCO010000000](https://www.ncbi.nlm.nih.gov/nuclot/JARVCO010000000).

## ACKNOWLEDGMENTS

We would like to thank Frank Kinnaman and Christoph Pierre at UC Santa Barbara for collecting sediment samples.

We thank the Biological Electron Microscopy Facility at UC Davis for TEM imaging.

We thank Professor Alex Sessions at Caltech who helped with identifying fatty acid profiles on gas chromatography-mass spectrometry (GC-MS).

We thank Professor Craig Carlson at UC Santa Barbara for his assistance with polysaccharide measurements.

The sequencing was carried out at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center. The genomic analysis work used Bridges and Bridges-2 at Pittsburg Supercomputing Center through allocation DEB190007 from the Extreme Science and Engineering Discovery Environment (XSEDE; 89) and the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program (90), which was supported by National Science Foundation grant number #1830033.

This research was funded by the Army Research Office (Grant No. W911NF-19-1-0010), the National Science Foundation (Grant No. 1830033), and the National Science Foundation of China (Grant No. 42076165).

## AUTHOR AFFILIATIONS

<sup>1</sup>Interdepartmental Graduate Program in Marine Science, University of California Santa Barbara, Santa Barbara, California, USA

<sup>2</sup>Marine Science Institute, University of California Santa Barbara, Santa Barbara, California, USA

<sup>3</sup>Marine Bioresource and Environment Research Center, Key Laboratory of Marine Eco-Environmental Science and Technology, First Institute of Oceanography, Ministry of Natural Resources of China, Qingdao, China

<sup>4</sup>Department of Chemical Engineering, University of California, Santa Barbara, California, USA

<sup>5</sup>Department of Ecology Evolution, and Marine Biology, University of California, Santa Barbara, California, USA

<sup>6</sup>Biological Engineering Program, University of California, Santa Barbara, California, USA

<sup>7</sup>Department of Earth Science, University of California Santa Barbara, Santa Barbara, California, USA

## PRESENT ADDRESS

Veronika Kivenson, Innovative Genomics Institute, University of California, Berkeley, California, USA

Xuefeng Peng, School of Earth, Ocean and Environment, University of South Carolina, Columbia, South Carolina, USA

Thomas S. Lankiewicz, Allonnia, Boston, Massachusetts, USA

## AUTHOR ORCIDs

Na Liu  <http://orcid.org/0000-0002-2071-4129>

David L. Valentine  <http://orcid.org/0000-0001-5914-9107>

## FUNDING

Funder	Grant(s)	Author(s)
<a href="#">DOD   USA   AFC   CDC   Army Research Office (ARO)</a>	W911NF-19-1-0010	Michelle A. O'Malley David L. Valentine
<a href="#">National Science Foundation (NSF)</a>	1830033	David L. Valentine
<a href="#">National Science Foundation of China</a>	42076165	Zhisong Cui

## AUTHOR CONTRIBUTIONS

Na Liu, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review and editing | Veronika Kivenson, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing – review and editing | Xuefeng Peng, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing – review and editing | Zhisong Cui, Conceptualization, Funding acquisition, Methodology, Validation, Writing – review and editing | Thomas S. Lankiewicz, Investigation, Methodology, Validation, Writing – original draft, Writing – review and editing | Kelsey M. Gosselin, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review and editing | Chance J. English, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review and editing | Elaina M. Blair, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review and editing | Michelle A. O'Malley, Funding acquisition, project administration, resources, Supervision, Validation, Writing – review and editing | David L. Valentine, Conceptualization, Funding acquisition, Supervision, Validation, Writing – original draft, Writing – review and editing

## DATA AVAILABILITY

Strain NLcol2<sup>T</sup> has been deposited at Leibniz-Institut DSMZ (= DSM 113125<sup>T</sup>) and Marine Culture Collection of China (= MCCC 1K08672<sup>T</sup>). The GenBank accession number for the full-length 16S rRNA gene sequence of strain NLcol2<sup>T</sup> is [OQ749723](#). The genome of strain NLcol2<sup>T</sup> has been deposited as the Whole Genome Shotgun project at DDBJ/ENA/GenBankNCBI under the accession number [JARVCO00000000](#). The version described in this paper is version [JARVCO01000000](#).

## ADDITIONAL FILES

The following material is available [online](#).

### Supplemental Material

**Supplemental Figures S1-S6 (AEM00914-23-s0001.pdf).** Supplemental figures.

**Supplemental Table S1 (AEM00914-23-s0002.xlsx).** FAMES analyses.

**Supplemental Table S2 (AEM00914-23-s0003.xlsx).** CAZymes.

**Supplemental Table S3 (AEM00914-23-s0004.xlsx).** Sulfatases.

**Supplemental Table S4 (AEM00914-23-s0005.xlsx).** Potential genes involved in macroalgal polysaccharide degradation.

**Supplemental Table S5 (AEM00914-23-s0006.xlsx).** Comparative genomic data for heatmap.

**Supplemental Table S6 (AEM00914-23-s0007.xlsx).** nif genes.

## REFERENCES

- Krause-Jensen D, Duarte CM. 2016. Substantial role of macroalgae in marine carbon sequestration. *Nature Geosci* 9:737–742. <https://doi.org/10.1038/ngeo2790>
- National Academies of Sciences, Engineering, and Medicine. 2022. A research strategy for ocean-based carbon dioxide removal and sequestration. The National Academies Press, Washington, DC. <https://nap.nationalacademies.org/catalog/26278/a-research-strategy-for-ocean-based-carbon-dioxide-removal-and-sequestration>.
- Ortega A, Geraldi NR, Alam I, Kamau AA, Acinas SG, Logares R, Gasol JM, Massana R, Krause-Jensen D, Duarte CM. 2019. Important contribution of macroalgae to oceanic carbon sequestration. *Nat Geosci* 12:748–754. <https://doi.org/10.1038/s41561-019-0421-8>
- Popper ZA, Michel G, Hervé C, Domozych DS, Willats WGT, Tuohy MG, Kloareg B, Stengel DB. 2011. Evolution and diversity of plant cell walls: from algae to flowering plants. *Annu Rev Plant Biol* 62:567–590. <https://doi.org/10.1146/annurev-arplant-042110-103809>
- Wei N, Quarterman J, Jin Y-S. 2013. Marine macroalgae: an untapped resource for producing fuels and chemicals. *Trends Biotechnol* 31:70–77. <https://doi.org/10.1016/j.tibtech.2012.10.009>
- Helbert W. 2017. Marine polysaccharide sulfatases. *Front Mar Sci* 4. <https://doi.org/10.3389/fmars.2017.00006>
- Glöckner FO, Kube M, Bauer M, Teeling H, Lombardot T, Ludwig W, Gade D, Beck A, Borzym K, Heitmann K, Rabus R, Schlesner H, Amann R, Reinhardt R. 2003. Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. *Proc Natl Acad Sci U S A* 100:8298–8303. <https://doi.org/10.1073/pnas.1431443100>
- Martinez-Garcia M, Brazel DM, Swan BK, Arnosti C, Chain PSG, Reitenga KG, Xie G, Poulton NJ, Gomez ML, Masland DED, Thompson B, Bellows WK, Ziervogel K, Lo C-C, Ahmed S, Gleasner CD, Detter CJ, Stepanauskas R. 2012. Capturing single cell genomes of active polysaccharide degraders: an unexpected contribution of *Verrucomicrobia*. *PLoS One* 7:e35314. <https://doi.org/10.1371/journal.pone.0035314>
- Kim JW, Brawley SH, Prochnik S, Chovatia M, Grimwood J, Jenkins J, LaButti K, Mavromatis K, Nolan M, Zane M, Schmutz J, Stiller JW, Grossman AR. 2016. Genome analysis of planctomyces inhabiting blades of the red Alga *Porphyra umbilicalis*. *PLoS One* 11:e0151883. <https://doi.org/10.1371/journal.pone.0151883>
- Cardman Z, Arnosti C, Durbin A, Ziervogel K, Cox C, Steen AD, Teske A. 2014. *Verrucomicrobia* are candidates for polysaccharide-degrading bacterioplankton in an arctic fjord of Svalbard. *Appl Environ Microbiol* 80:3749–3756. <https://doi.org/10.1128/AEM.00899-14>
- Rivas-Marín E, Devos DP. 2018. The paradigms they are a-changin': past, present and future of PVC bacteria research. *Antonie Van Leeuwenhoek* 111:785–799. <https://doi.org/10.1007/s10482-017-0962-z>
- Spring S, Bunk B, Spröer C, Schumann P, Rohde M, Tindall BJ, Klenk H-P. 2016. Characterization of the first cultured representative of *Verrucomicrobia* subdivision 5 indicates the proposal of a novel phylum. *ISME J* 10:2801–2816. <https://doi.org/10.1038/ismej.2016.84>
- Oren A, Garrity GM. 2021. Valid publication of the names of forty-two phyla of prokaryotes. *Int J Syst Evol Microbiol* 71:005056. <https://doi.org/10.1099/ijsem.0.005056>
- Spring S, Brinkmann N, Murrja M, Spröer C, Reitner J, Klenk H-P. 2015. High diversity of culturable prokaryotes in a lithifying hypersaline microbial mat. *Geomicrobiology J* 32:332–346. <https://doi.org/10.1080/01490451.2014.913095>
- van Vliet DM, Palakawong Na Ayudthaya S, Diop S, Villanueva L, Stams AJM, Sánchez-Andrea I. 2019. Anaerobic degradation of sulfated polysaccharides by two novel *Kiritimatiellales* strains isolated from black sea sediment. *Front Microbiol* 10:253. <https://doi.org/10.3389/fmicb.2019.00253>
- van Vliet DM, Lin Y, Bale NJ, Koenen M, Villanueva L, Stams AJM, Sánchez-Andrea I. 2020. *Pontiella desulfatans* gen. nov., sp. nov., and *Pontiella sulfatireligans* sp. nov., two marine anaerobes of the *Pontiellaceae* fam. nov. producing sulfated glycosaminoglycan-like exopolymers. *Microorganisms* 8:920. <https://doi.org/10.3390/microorganisms8060920>
- Mu D-S, Zhou L-Y, Liang Q-Y, Chen G-J, Du Z-J. 2020. *Tichowtungia aerotolerans* gen. nov., sp. nov., a novel representative of the phylum *Kiritimatiellaota* and proposal of *Tichowtungiaceae* fam. nov., *Tichowtungiales* ord. nov. and *Tichowtungia* class. nov. *Int J Syst Evol Microbiol* 70:5001–5011. <https://doi.org/10.1099/ijsem.0.004370>
- Ding H, Valentine DL. 2008. Methanotrophic bacteria occupy benthic microbial mats in shallow marine hydrocarbon seeps, Coal Oil Point, California. *J Geophys Res* 113. <https://doi.org/10.1029/2007JG000537>
- Eichhubl P, Greene HG, Naehr T, Maher N. 2000. Structural control of fluid flow: offshore fluid seepage in the Santa Barbara Basin, California. *J Geochem Explor* 69–70:545–549. [https://doi.org/10.1016/S0375-6742\(00\)00107-2](https://doi.org/10.1016/S0375-6742(00)00107-2)
- Washburn L, Clark JF, Kyriakidis P. 2005. The spatial scales, distribution, and intensity of natural marine hydrocarbon seeps near Coal Oil Point, California. *Mar Pet Geol* 22:569–578. <https://doi.org/10.1016/j.marpetgeo.2004.08.006>
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41:D590–D596. <https://doi.org/10.1093/nar/gks1219>
- Pruesse E, Peplies J, Glöckner FO. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28:1823–1829. <https://doi.org/10.1093/bioinformatics/bts252>
- 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. <https://doi.org/10.1093/bioinformatics/btp348>
- Stamatakis A. 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Letunic I, Bork P. 2019. Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* 47:W256–W259. <https://doi.org/10.1093/nar/gkz239>
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Joshi NA, Fass JN. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (version 1.33). <https://github.com/najoshi/sickle>.
- Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. *Bioinformatics* 31:1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
- 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. <https://doi.org/10.1093/bioinformatics/btp352>
- Eren AM, 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. <https://doi.org/10.7717/peerj.1319>
- 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. <https://doi.org/10.1101/gr.186072.114>
- Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>
- Eddy SR. 2011. Accelerated profile HMM searches. *PLOS Comput Biol* 7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>
- El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladini L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res* 47:D427–D432. <https://doi.org/10.1093/nar/gky995>
- Haft DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT, White O. 2001. TIGRFAMs: A protein family resource for the functional

- identification of proteins. *Nucleic Acids Res* 29:41–43. <https://doi.org/10.1093/nar/29.1.41>
38. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang H-Y, El-Gebali S, Fraser MI, et al. 2019. Interpro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res* 47:D351–D360. <https://doi.org/10.1093/nar/gky1100>
  39. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 45:D200–D203. <https://doi.org/10.1093/nar/gkw1129>
  40. Leimbach A. 2016. Bac-Genomics-scripts: Bovine *E. coli* Mastitis comparative Genomics edition. Zenodo. <https://doi.org/10.5281/zenodo.215824>
  41. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. <https://doi.org/10.1186/1471-2105-4-41>
  42. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 428:726–731. <https://doi.org/10.1016/j.jmb.2015.11.006>
  43. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100–3108. <https://doi.org/10.1093/nar/gkm160>
  44. Lowe TM, Chan PP. 2016. tRNAscan-SE on-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* 44:W54–W57. <https://doi.org/10.1093/nar/gkw413>
  45. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research* 40:D109–D114. <https://doi.org/10.1093/nar/gkr988>
  46. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, Ong WK, Paley S, Subhraveti P, Karp PD. 2020. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Res* 48:D445–D453. <https://doi.org/10.1093/nar/gkz862>
  47. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH, Hancock J. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>
  48. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Preprints*. <https://doi.org/10.7287/peerj.preprints.1900v1>
  49. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* 46:W95–W101. <https://doi.org/10.1093/nar/gky418>
  50. Barbeyron T, Brillet-Guéguen L, Carré W, Carrière C, Caron C, Czjzek M, Hoebeker M, Michel G. 2016. Matching the diversity of sulfated biomolecules: Creation of a classification database for sulfatases reflecting their substrate specificity. *PLOS One* 11:e0164846. <https://doi.org/10.1371/journal.pone.0164846>
  51. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 37:420–423. <https://doi.org/10.1038/s41587-019-0036-z>
  52. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>
  53. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>
  54. Folch J, Lees M, Sloane Stanley GH. 1957. A simple method for the isolation and purification of total lipides from animal tissues. *J Biol Chem* 226:497–509.
  55. Gilmore SP, Lankiewicz TS, Wilken SE, Brown JL, Sexton JA, Henske JK, Theodorou MK, Valentine DL, O'Malley MA. 2019. Top-down enrichment guides in formation of synthetic microbial consortia for biomass degradation. *ACS Synth Biol* 8:2174–2185. <https://doi.org/10.1021/acssynbio.9b00271>
  56. Peng X, Wilken StE, Lankiewicz TS, Gilmore SP, Brown JL, Henske JK, Swift CL, Salamov A, Barry K, Grigoriev IV, Theodorou MK, Valentine DL, O'Malley MA. 2021. Genomic and functional analyses of fungal and bacterial consortia that enable lignocellulose breakdown in goat gut microbiomes. *Nat Microbiol* 6:499–511. <https://doi.org/10.1038/s41564-020-00861-0>
  57. Engel A, Händel N. 2011. A novel protocol for determining the concentration and composition of sugars in particulate and in high molecular weight dissolved organic matter (HMW-DOM) in seawater. *Marine Chemistry* 127:180–191. <https://doi.org/10.1016/j.marchem.2011.09.004>
  58. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* 12:635–645. <https://doi.org/10.1038/nrmicro3330>
  59. Konstantinidis KT, Rosselló-Móra R, Amann R. 2017. Uncultivated microbes in need of their own taxonomy. *ISME J* 11:2399–2406. <https://doi.org/10.1038/ismej.2017.113>
  60. Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A* 102:2567–2572. <https://doi.org/10.1073/pnas.0409727102>
  61. 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. <https://doi.org/10.1038/s41467-018-07641-9>
  62. Santarella-Mellwig R, Franke J, Jaedicke A, Gorjánác M, Bauer U, Budd A, Mattaj IW, Devos DP. 2010. The compartmentalized bacteria of the Planctomycetes-Verrucomicrobia-Chlamydiae superphylum have membrane coat-like proteins. *PLOS Biol* 8:e1000281. <https://doi.org/10.1371/journal.pbio.1000281>
  63. Sichert A, Corzett CH, Schechter MS, Unfried F, Markert S, Becher D, Fernandez-Guerra A, Liebecke M, Schweder T, Polz MF, Hehemann J-H. 2020. Verrucomicrobia use hundreds of enzymes to digest the algal polysaccharide fucoidan. *Nat Microbiol* 5:1026–1039. <https://doi.org/10.1038/s41564-020-0720-2>
  64. Hehemann J-H, Correc G, Thomas F, Bernard T, Barbeyron T, Jam M, Helbert W, Michel G, Czjzek M. 2012. Biochemical and structural characterization of the complex agarolytic enzyme system from the marine bacterium *Zobellia galactanivorans*. *J Biol Chem* 287:30571–30584. <https://doi.org/10.1074/jbc.M112.377184>
  65. Ficko-Blean E, Préchoux A, Thomas F, Rochat T, Larocque R, Zhu Y, Stam M, Génicot S, Jam M, Calteau A, Viart B, Ropartz D, Pérez-Pascual D, Correc G, Matard-Mann M, Stubbs KA, Rogniaux H, Jeudy A, Barbeyron T, Médigue C, Czjzek M, Vallenet D, McBride MJ, Duchaud E, Michel G. 2017. Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. *Nat Commun* 8:1685. <https://doi.org/10.1038/s41467-017-01832-6>
  66. Reisky L, Préchoux A, Zühlke M-K, Bäumgen M, Robb CS, Gerlach N, Roret T, Stanetty C, Larocque R, Michel G, Song T, Markert S, Unfried F, Mihovilovic MD, Trautwein-Schult A, Becher D, Schweder T, Borscheuer UT, Hehemann J-H. 2019. A marine bacterial enzymatic cascade degrades the algal polysaccharide ulvan. *Nat Chem Biol* 15:803–812. <https://doi.org/10.1038/s41589-019-0311-9>
  67. Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. 2022. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Research* 50:D571–D577. <https://doi.org/10.1093/nar/gkab1045>
  68. Appel MJ, Bertozzi CR. 2015. Formylglycine, a post-translationally generated residue with unique catalytic capabilities and biotechnology applications. *ACS Chem Biol* 10:72–84. <https://doi.org/10.1021/cb500897w>
  69. Berteau O, Guillot A, Benjdia A, Rabot S. 2006. A new type of bacterial sulfatase reveals a novel maturation pathway in prokaryotes. *J Biol Chem* 281:22464–22470. <https://doi.org/10.1074/jbc.M602504200>
  70. Wagner M, Horn M. 2006. The *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae* and sister phyla comprise a superphylum with biotechnological and medical relevance. *Curr Opin Biotechnol* 17:241–249. <https://doi.org/10.1016/j.copbio.2006.05.005>

71. Zhang Q, Li N, Liu X, Zhao Z, Li Z, Xu Z. 2004. The structure of a sulfated galactan from *Porphyra haitanensis* and its *in vivo* antioxidant activity. *Carbohydr Res* 339:105–111. <https://doi.org/10.1016/j.carres.2003.09.015>
72. Araki C. 1956. Structure of the agarose constituent of agar-agar. *BCSJ* 29:543–544. <https://doi.org/10.1246/bcsj.29.543>
73. Anderson NS, Dolan TCS, Rees DA. 1965. Evidence for a common structural pattern in the polysaccharide sulphates of the Rhodophyceae. 4976. *Nature* 205:1060–1062. <https://doi.org/10.1038/2051060a0>
74. Duckworth M, Yaphe W. 1971. The structure of Agar: Part I. Fractionation of a complex mixture of polysaccharides. *Carbohydr Res* 16:189–197. [https://doi.org/10.1016/S0008-6215\(00\)86113-3](https://doi.org/10.1016/S0008-6215(00)86113-3)
75. Cong Q, Chen H, Liao W, Xiao F, Wang P, Qin Y, Dong Q, Ding K. 2016. Structural characterization and effect on anti-angiogenic activity of a fucoidan from *Sargassum fusiforme*. *Carbohydr Polym* 136:899–907. <https://doi.org/10.1016/j.carbpol.2015.09.087>
76. Hehemann J-H, Kelly AG, Pudlo NA, Martens EC, Boraston AB. 2012. Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proc Natl Acad Sci U S A* 109:19786–19791. <https://doi.org/10.1073/pnas.1211002109>
77. Glover JS, Ticer TD, Engevik MA. 2022. Characterizing the mucin-degrading capacity of the human gut microbiota. *Sci Rep* 12:8456. <https://doi.org/10.1038/s41598-022-11819-z>
78. Grondin JM, Tamura K, Déjean G, Abbott DW, Brumer H. 2017. Polysaccharide utilization loci: fueling microbial communities. *J Bacteriol* 199:00860–16. <https://doi.org/10.1128/JB.00860-16>
79. Skerker JM, Prasol MS, Perchuk BS, Biondi EG, Laub MT. 2005. Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis. *PLoS Biol* 3:e334. <https://doi.org/10.1371/journal.pbio.0030334>
80. Hehemann J-H, Correc G, Barbeyron T, Helbert W, Czjzek M, Michel G. 2010. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* 464:908–912. <https://doi.org/10.1038/nature08937>
81. Seefeldt LC, Hoffman BM, Dean DR. 2009. Mechanism of Mo-dependent nitrogenase. *Annu Rev Biochem* 78:701–722. <https://doi.org/10.1146/annurev.biochem.78.070907.103812>
82. Arcandéguy T, Jack R, Merrick M. 2001. PII signal transduction proteins, pivotal players in microbial nitrogen control. *Microbiol Mol Biol Rev* 65:80–105. <https://doi.org/10.1128/MMBR.65.1.80-105.2001>
83. Buchko GW, Robinson H, Addlagatta A. 2009. Structural characterization of the protein cce\_0567 from *Cyanothece 51142*, a metalloprotein associated with nitrogen fixation in the DUF683 family. *Biochim Biophys Acta* 1794:627–633. <https://doi.org/10.1016/j.bbapap.2009.01.002>
84. Khadem AF, Pol A, Jetten MSM, Op den Camp HJM. 2010. Nitrogen fixation by the verrucomicrobial methanotroph '*Methylacidiphilum fumariolicum*' SolV. *Microbiology* 156:1052–1059. <https://doi.org/10.1099/mic.0.036061-0>
85. Wertz JT, Kim E, Breznak JA, Schmidt TM, Rodrigues JLM. 2012. Genomic and physiological characterization of the *Verrucomicrobia* isolate *Geminisphaera colitermitum* gen. nov., sp. nov., reveals microaerophily and nitrogen fixation genes. *Appl Environ Microbiol* 78:1544–1555. <https://doi.org/10.1128/AEM.06466-11>
86. Cabello-Yeves PJ, Ghai R, Mehrshad M, Picazo A, Camacho A, Rodriguez-Valera F. 2017. Reconstruction of diverse verrucomicrobial genomes from metagenome datasets of freshwater reservoirs. *Front Microbiol* 8:2131. <https://doi.org/10.3389/fmicb.2017.02131>
87. Delmont TO, Quince C, Shaiber A, Esen ÖC, Lee ST, Rappé MS, McLellan SL, Lückner S, Eren AM. 2018. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat Microbiol* 3:963. <https://doi.org/10.1038/s41564-018-0209-4>
88. Zehr JP, Jenkins BD, Short SM, Steward GF. 2003. Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environ Microbiol* 5:539–554. <https://doi.org/10.1046/j.1462-2920.2003.00451.x>
89. Towns J, Cockerill T, Dahan M, Foster I, Gaither K, Grimshaw A, Hazlewood V, Lathrop S, Lifka D, Peterson GD, Roskies R, Scott JR, Wilkins-Diehr N. 2014. XSEDE: accelerating scientific discovery. *Comput Sci Eng* 16:62–74. <https://doi.org/10.1109/MCSE.2014.80>
90. Boerner TJ, Deems S, Furlani TR, Knuth SL, Towns J. 2023. ACCESS: Advancing innovation: NSF's advanced cyberinfrastructure coordination ecosystem: Services & support. *PEARC*:173–176. <https://doi.org/10.1145/3569951.3597559>